Dynamics and Social Networks

by

Fang-Yi Yu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Computer Science and Engineering)
in the University of Michigan
2019

Doctoral Committee:

Assistant Professor Grant Schoenebeck, Chair
Assistant Professor Danai Koutra
Professor Mark E. Newman
Professor Seth Pettie

Fang-Yi Yu

fayu@umich.edu

ORCID iD: 0000-0002-3697-8807

ACKNOWLEDGMENT

I have so many people to thank for this thesis. First, I cannot express enough gratitude to Grant Schoenebeck. He is the best advisor I could have hoped for in graduate school. He has gone far out of his way to make sure that I had every opportunity: to learn, travel, speak at conferences, and so on.

In addition to my advisor, I thank Jie Gao, who has taught me so much about how to be a researcher. I also like to thank all of my other coauthors and collaborators throughout grad school: Biaoshuai Tao, Yuqing Kong, Bo Li, and Golnaz Ghasemiesfeh. Finally, I thank the other members of my thesis committee, Seth Pettie, Mark E. Newman, and Danai Koutra, for their helpful feedback on this dissertation and throughout graduate school.

I thank all of my friends and family. Thanks to Mom, Dad, and Kevin for your support while I delay adulthood. And, thank you to Frances, my friend and the love of my life, for everything.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

Social networks are everywhere in our everyday lives. We aggregate information, make decisions, and form opinions through these interactions on social networks. This thesis aims to improve our understanding of social network structures and social network dynamics, including the spread of social contagions, opinion formation, and myopic routing.

We first consider complex contagions where a node requires several infected neighbors before becoming infected itself, and we give a theoretical analysis of which properties of social networks— small-world properties, power-law degree distribution, time evolving, and community structure— can affect the spread of contagions. Finally, we consider the influence maximization problem on social networks with community structure when the contagions are complex.

For social network structures, we begin with the role of strong and weak ties. Exploiting the idea of strong ties we propose a Sybil detection algorithm which prevents an adversary from creating a large number of identities to attack a recommendation system. Later, we study the role of weak ties in echo chambers and bubble filters.

Finally, we focus on opinion formation and community structures. We propose a family of general rich-get-richer dynamics which includes several well-studied models as special cases. This family of dynamics reaches consensus fast on dense Erdös-Rényi graphs. In contrast to this result, we prove a dichotomy theorem about community structures and these richer-get-richer dynamics.

CHAPTER 1

Introduction

Social networks are omnipresent in our everyday lives. For example, we learn the latest news from friends' post on social media. We buy phones based on neighbors' choices. We form our opinions on politics under our family's effect. We derive information, make decisions, and form opinions through these interactions on social networks. To better understand society we need to improve our understanding of social networks.

The study of social networks starts from descriptive characterization form collected data [211, 87, 70, 227, 193]. In the early 2000s, the widespread availability of electronic databases and the development of the internet enable large-scale empirical studies. These observations (e.g., power-law degree distribution, clustering, small-world property [8, 43, 89, 11]) provided convincing evidence that real social networks have unique structures that are not captured by previous simple mathematical (Erdös-Rényi) models of networks [87], and triggered a massive surge of interest in using statistical models to quantify and understand this structure (e.g., small-world model, time evaluating model [169, 234, 178, 24, 133]). This line of works discovers many interesting properties of social networks. However, these should be our first step toward the ultimate goal of understanding the function of systems built on social networks— society. We want to ask: how do network properties (small-world, power-law degree distribution, time evolving, community structure, and strong/weak ties) affect the function of systems (contagions and other dynamics) on social networks? In particular, we study

- which properties of social networks— small-world, power-law degree distribution, time evolving— enable infections spread.
- how the role of ties in social networks— strong/weak ties and inter-communities connection— affects the global dynamics or character of social networks.

Before diving into the details, the main contributions are outlined in this section, and each part will be partitioned into background followed by our contributions. In section 1.1, beginning with social contagions, we introduce basic models for social networks. Then in section 1.2 and 1.3, we study characteristics of ties which consists of two parts— strong/weak ties and community structure.

1.1 Social Contagions

Our everyday life is embedded in a network of social interactions. Through these social interactions, we *spread* and *receive* information, beliefs, diseases, technologies, and behaviors. A better understanding of these social interactions promises a better understanding of and the ability to influence a wide range of phenomena— financial practices, healthy/unhealthy habits, and voting practices [23, 60, 161, 4]. However, these successful models, which apply to diseases, are missing key ingredients that severely restrict their application to other social phenomena. One missing ingredient is that many social contagions are *complex*. We call contagions simple when the influence of an additionally infected neighbor is marginally decreasing, and call contagions complex when this fails to hold (e.g., contagions that require activation from multiple neighbors). Many examples of complex contagions have been reported, including pricey technology innovations, the change of social behaviors, the decision to participate in a migration, etc [156, 213]. In this work, we consider two complex contagion models: *k-complex contagions (bootstrap percolation)* [49, 50], and the *general threshold model* [95].

1.1.1 Model of Contagions

Given initially infected agents, a contagion propagates through local interactions. Therefore a model of contagion can be specified with three aspects: The initially infected nodes, graph (the local interactions), and the dynamics of the contagions (how do contagions propagate through local interactions?). We focus on contagions with a constant number of initially infected vertices and say a contagion *spreads* if there are a constant fraction of vertices infected with high probability.

In this section, we begin with one of the most basic model of complex contagions considered by Centola and Macy [49] which is defined formally in Definition 2.1. Given the set of initially infected vertices $Z(0)$, at each round a node becomes active if at least k neighbors are active. This single threshold model is also called a *k-complex*

contagion. If $k = 1$, a simple contagion, all nodes connected to an infected node is infected eventually. When $k > 1$, a single active neighbor is not enough to trigger activation. Thus captures the idea of complex contagions and also the essence of the theoretical challenges of studying complex contagions.

We will begin with discussing k-complex contagions on (1) the Erdös-Rényi models as an exercise and demonstration of the difference between simple and complex contagions which result is done by Janson et al. [119]. Later, we will present our results for k-complex contagions on (2) a generalized Kleinberg's small world [97] and (3) configuration models with power-law degree distribution [205]. These results show the behaviors of complex contagions on social networks which have unique structures (small world property and power-law degree distribution) are different from complex contagions on simple mathematical models— the Erdös-Rényi model.

Finally, one of the limitations of this k-complex contagion model is the dependency on the fixed threshold k for all nodes in the network. In practice, some people like to try out new things and are more risk driven while others are risk averse. Therefore the threshold function is not necessarily uniform. As a result, we also consider one step of generalizing the k-complex contagion model by considering the threshold coming from a distribution $\mathcal{D}$ on positive integers— called *general threshold models* [95]. The initial adoption barrier can still exist which makes the contagion complex. We provide analysis of the spreading behaviors on a broad family of networks that grow over time.

1.1.2 Erdös-Rényi model

Here we consider k-complex contagions on one of the simplest mathematical models— the Erdös-Rényi model and demonstrate the difference between simple and complex contagions. An *Erdös-Rényi model* [87], $G(n, p)$, is a distribution over graphs with vertex set $[n] := \{1, 2, \ldots, n\}$ where every edge appears independently with probability $p = p(n)$. If $k = 1$, the fraction of contagions is related to the giant connected component problem, and the 1-complex contagion spreads if and only if the expected degree of each vertex is greater than 1. On the other hand, when $k > 1$, Janson et al. [119] shows k-complex contagions spread only if the expected degree of each vertex is $\Omega(n^\gamma)$ where the constant γ depends on k and p. It is unrealistic for real human to have a super constant number of friends; however, as we will see shortly, if the network has certain common social network properties, the complex contagions can spread even with a constant degree.

1.1.3 Small world model

The Erdös-Rényi model is simple, but it does not capture many unique structures in real social networks. A primary example is called the small-world phenomenon. Influenced by de Sola Pool and Kochen [70], Milgram devised an experiment to test the small-world conjecture [227]. After selecting a target person and a group of starting individuals, the experimenter gives each starter an envelope with basic information about the target and ask him/her to forward the envelope to someone he/she knew so that the envelope might reach the target. The process repeated until the envelope reached the target. Over many trials, the average number of intermediate steps in a successful chain was between five and six, a quantity that has since entered popular culture as the "six degrees of separation" principle. This experiment triggered a surge of interest in modeling the small world property.

One observation from this experiment is that social networks are not only highly clustered but also have small diameter— most pairs of individuals are connected by short paths. Motivated by this observation, the Watts-Strogatz network [234] was proposed to explain the small-world phenomenon. Furthermore, Kleinberg made another insightful observation—instead of *having* short routes, the experiment also supports that most pairs of individuals can *find* such routes, and he proposed the Kleinberg's small world model [133]. Kleinberg's small world model simulates social networks whose ties have two characters strong and weak ties which are parameterized by γ.

One recent work [99] on social influence by k-complex contagion models discovered that the generation of weak-ties also crucially impacts the spreading rate on Kleinberg's small world model: If a k-seed cluster is infected initially, the contagion spreads to the entire network in $O(\text{polylog}(n))$ rounds if $\gamma \in [2; \frac{2(k+1)}{k})$, and in $\Omega(\text{poly}(n))$ rounds otherwise. The parameter of $\gamma = 2$ is particular: when γ is anything but 2 the k-complex contagions spread slowly for large enough k.

Contributions [95] In Chapter 5, I propose a natural generalization of Kleinberg's small world model to allow node heterogeneity: each node has a personalized parameter γ sampled from a shared distribution, and prove that this model enables k-complex contagions on a broad range of the parameter space, improving the robustness of the model. Moreover, we empirically show that this generalization is supported by real-world data.

1.1.4 Power-law degree distribution

Another fundamental characteristic of a social network is its degree distribution which is a description of the relative frequencies of nodes that have different degrees. A graph has power-law degree distribution, if the number of nodes having degree d is proportional to $d^{-\alpha}$, for a positive constant α, and the graph is also called *scale-free network*. In 1965, Price [193] showed that the number of citations to papers follows a power law distribution. Later, studies of the World Wide Web reported that the network of web pages also has a power law degree distribution [8, 43, 89]. Observations of many different types of social networks also found power-law degree distributions, as well as biological, economic and semantic networks [218, 6, 177].

Although random graph models $G(n, p)$ do not have a power-law degree distribution, we can "impose" this property on random graph model by introducing a generalization called the *configuration model graphs with degree sequence* k where k is a vector, and k_i is the degree of vertex i. Given a degree sequence, we may construct a random graph by choosing a uniformly random matching on the degree "stubs"(half-edges), and the histogram of k would be the degree distribution of resulting graph. Therefore, we can construct a configuration model with power-law degree distribution by taking degree sequence k such that the corresponding histogram satisfies power-law distribution.

Contribution [205] With this minimal modification, a natural question is: do k-complex contagions spread on configuration model with power-law degree distribution. In Chapter 4, I study this problem and tried to answer if power-law degree distributions alone are sufficient to spread the contagions. I theoretically show the contagions spread on configuration model graphs with power-law degree distribution with the exponent between 2 and 3 which also have a constant average degree: the single seed of the highest degree node will infect a constant fraction of the graph with high probability. This result shows a distinction between random graphs $G(n, p)$ and configuration model with power-law degree distribution in the sense that although both models have a constant average degree, k-complex contagions do not spread on $G(n, p)$ but do spread on on configuration model with a power-law degree.

1.1.5 Time evolution models

In addition to using the configuration model with power-law degree distribution to model social network with power-law degree distribution, Barabási and Albert [24]

proposed an elegant model— preferential attachment model— which also has power-law degree distribution. Moreover, they attributed the power-law degree distribution to two generic properties: first the network is time evolving, and second, nodes attach to the previous nodes preferentially (nodes gain new edges in proportion to the number they already have). These two properties further stimulate a large number of follow-up work [187, 77].

It has been shown that a k-complex contagion is generally slower and more delicate than simple contagion $k = 1$ [99, 84]. One of the limitations of this k-complex contagion model is the dependency on the fixed threshold k for all nodes in the network. In practice, some people who like to try out new things and are more risk driven while others are risk averse. Therefore the threshold function is not necessarily uniform.

Contribution [95] In Chapter 5, I consider a *general threshold model* (Definition 5.1): the threshold k for each node sampling from a distribution $\mathcal{D}$ identically and independently which reduces to k-complex contagions when $\mathcal{D}$ is a constant distribution, and provide analysis on the spreading behaviors on a general family of time evolving networks which contains preferential attachment model. Theoretically, I show if the set of initial infected seed contains the early arrival agents, in the directed case, the fraction of nodes infected converges and will approach to a fixed point of a function determined by the parameters of the contagion and networks. In the undirected case, I show with high probability, the total number of infected nodes will always be a constant fraction higher than the highest stable fixed point of the above polynomial, when a non-zero stable fixed point exists. Furthermore, via simulations this work also shows, using a co-authorship network derived from DBLP databases and the Stanford web network, that our theoretical results can be used to predict the infection rate.

1.2 Characters of ties— Strong and weak ties

Besides the global structure of social networks— degree distribution, small-world property or time evolving, real social networks also exhibit unique local structures which can often be checked locally, e.g., strong/weak ties and community structure. In this section, we want to ask how do these characters of ties affect the global behavior of dynamics on social networks.

The connection of social networks may have different strengths and are often

classified into strong ties and weak ties. Strong ties represent close relationships such as family members, close friends, relationships that people invest time and energy to maintain actively; while weak ties capture relationships with acquaintances, friends from the old days, relationships that are relatively less invested. Due to the limited time and energy, an individual often has only a small number of strong ties but may have a much more significant number of weak ties.

1.2.1 Strong ties and Sybil detection

Besides the relational meanings, strong ties and weak ties also have structurally different meanings. Strong ties often exhibit the transitive closure or triadic closure property [195]— if A and B are close friends, and B and C are close friends, then it is very likely that A and C are also close friends. The existence of strong ties in social networks separates them from "typical" graphs, say the Erdö-Rènyi model which does not have any transitive closure with high probability when the expected degree is constant. How can we exploit this special property of social networks and detect *anomaly*?

Contribution [207] In Chapter 7, we exploit the idea of strong ties and low-dimensional properties of social networks, and propose a *Sybil Detection algorithm* which prevents an adversary from creating a large number of identities in social networks to attack a recommendation system.

Previous Sybil detection algorithms which use networks structure typically assume that it is difficult for an adversary to create edges to any real identities in the network which seems to fail in real-world settings. Instead, based on the observation— though Sybils may make a large number of weak ties, it is hard for them to form strong ties which require time and energy to maintain. This work makes a much weaker assumption: creating edges from Sybils to significant fraction of real identities is difficult, yet allowing that Sybils can connect to a random constant fraction of real identities. Moreover, by iteratively removing identities without "strong ties" our algorithm provably detects Sybil (anomaly) in social networks.

1.2.2 Weak ties and echo chambers

In the seminal paper "The Strength of the Weak Ties" Granovetter [105] showed how information spreads through weak ties. While strong ties connect people who are more similar to each other (due to homophily), weak ties tend to bring fresh information to

a social group, which can be extremely valuable, for example, in the case of looking for new jobs. However, weak ties, especially those discovered on a social platform, are a lot easier to form or break, making it convenient to block opinions that one does not like and stay in a comfortable "echo chamber" [20, 223]. How do weak ties affect the consensus of society given two roles of weak ties: the strength of weak ties— that weak ties are useful for spreading information; and the idea of echo chambers or filter bubbles, that people are typically bombarded by the opinions of like-minded individuals.

Contribution In Chapter 8, I look at opinion formation and the role of weak ties. There are two phenomena both of which promote consensus between agents connected by ties: *influence*, agents changing their opinions to match their neighbors; and *selection*, agents re-wiring to connect to new agents when the existing neighbor has a different opinion. In my agent-based model, I consider that only weak ties can be rewired and strong ties do not change. The network structure as well as the opinion landscape thus co-evolve with two important parameters: the probability of influence versus selection; and the fraction of strong ties versus weak ties. Using empirical and theoretical methodologies we discovered that on a two-dimensional spatial network:

- With a no/low selection rate the presence of weak ties enables fast consensus. It conforms with the classical theory that weak ties are helpful for quickly mixing and spreading information, and strong ties alone act much more slowly.
- With a high selection rate, too many weak ties inhibit any consensus at all—the graph partitions. The weak ties reinforce the opinions rather than mixing them. However, sufficiently many strong ties promote convergence, though at a slower pace.

1.3 Characters of ties— community structure

Similar to strong and weak ties, community structure is also ubiquitous in social networks. Social networks may have different communities, and each agent belongs to different communities— occupation, sex, or geographical region. Agents within the same community tend to have more connection, and fewer when they are in different communities. Conversely, a person has many friends in a specific community may also in the community— the membership of the community can be checked locally. A natural question is how does the community structure affect the global behavior

of dynamics on social networks. In particular, we want to understand the synergy between opinion formulation and community structure.

In addition to spreading information, social networks also interplay with the formation of opinion— agents' beliefs and opinions evolve as they interact with others on social networks. The opinion formulation plays a role in nearly every social, political and economic process.[175] A more preliminary question we might ask is *whether the society can form a consensus, if so how long does it take to reach a consensus?* If a social network has community structure, people interact more with those in the same community than the other. Therefore with community structure, it is likely to form local agreement within each community and prevent global consensus. We present two models for opinion formulation and examine the relationship between these dynamics and social networks with and without community structure.

1.3.1 Node Dynamics

Due to the complexity of the real world, it is unlikely that any specific simple mathematical model can accurately describe the exchange of opinions among people. Thus we seek to define general classes of dynamics, and study the holistic behavior of these family rather than particulars. In Chapter 9 and 10, we proposed a general model for opinion formulation on a fixed graph with a finite set of opinions—called *Node dynamics*. Given a fixed graph and an initial opinion each node has, each node's opinion is updated over time by randomly selecting a node to update, and then updating the node's opinion according to some function of its neighbor's opinions. This model is general enough to capture a variety of well-studied models: the voter model [65, 114, 150, 151, 32, 47, 46, 212, 221, 222, 64], iterative majority [136, 34, 125, 176, 224], and iterative k-majority processes [75, 66, 63, 31, 1].

Contribution for Node Dynamics on Erdös-Rènyi models [206] In Chapter 9, I prove that the Node Dynamics converge to consensus in time $\Theta(n \log n)$ on complete graphs and dense Erdös-Rènyi model with n nodes when the update function is from a large family of majority-like functions. This family of majority-like node dynamics tries to be the minimal model to capture two important aspects of opinion formulation: (1) People update their opinions based on the relative frequency of each opinion in their neighbors, instead of the absolute number of each opinions. (2) People are (disproportionally) more likely to updates to the majority opinion—majority-like. Our results support that people converge to consensus fast when there is no community structure (complete graphs and Erdös-Rènyi model).

Contribution for Node Dynamics on block models In Chapter 10, I study majority-like Node Dynamics on networks with two communities. The networks we consider are weighted graphs comprised of two equally sized communities where intracommunity edges have weight p, and intercommunity edges have weight q. Thus q and p parameterize the connectivity between the two communities.

I prove a dichotomy theorem about the interaction of the two parameters: 1) the "majority-like" update function, and 2) the level of inter-community connectivity. For each setting of parameters, we show that either: the system quickly converges to consensus with high probability in time $\Theta(n \log(n))$; or, the system can get "stuck" and take time $2^{\Theta(n)}$ to reach consensus. Technically, I achieve this fast convergence result by exploiting the connection between a family of reinforced random walks and dynamical systems literature. One main result shows if the systems are a reinforced random walk with a gradient-like function, it converges to an arbitrary neighborhood of a local attracting point in $O(n \log n)$ time with high probability. This result adds to the recent literature on saddle-point analysis and shows a large family of stochastic gradient descent algorithm converges to local minimal in $O(n \log n)$ when the step size $O(1/n)$.

1.3.2 Networks with Community Structure and Naming game

Another stylized model of opinion formulation we have analyzed in depth is the Naming Game [26, 214], in which agents negotiate conventions through local pairwise interactions. The Naming Game captures the generic and essential features of an agreement process over a *non-finite* set of names in networked agent-based systems. Briefly speaking, when two agents wish to communicate, one agent, the speaker, randomly selects one from her list of current names and uses this name to initiate communication with the listener. If the listener recognizes that name, both the speaker and listener purge their lists of current name to only include that "successful" convention. If the listener does not recognize it, she adds it to her list of known names.

Contribution [96] In Chapter 11, I study the Naming Game to analyze how agents reach global consensus, and empirically show how community structure (asymmetric topology) and heterogeneous agents affect the convergence rates, and how consensus get manipulated when (adversarial) stubborn nodes are introduced at different points of the process. In this work, with stability analysis in control theory, we show the process with binary opinions converges of segregation when the community structure

exists which supports that community structure indeed hinders the dynamics reaching consensus.

1.4 Technique Overview

In addition to improving understanding of social networks, the tools used and developed in this thesis might be useful in its own right in Markov chain, dynamical system theory, and random graph theory.

1.4.1 Mean field approximation

Most of the dynamics on social networks we studied are large and complex stochastic models—a large number of individual agents which interact with each other. To understand this complex stochastic system, a generic approach partitions the problem into two steps as follows:

1. First, approximate these dynamics by simpler models (e.g., use the number of infected nodes to represent the contagions process of interest or the number of nodes with given opinion) which are often discrete stochastic dynamical systems over low dimensional geometrical spaces.

2. Based on the observation that, in most of the cases, as the size of the systems increases, the time evolution become smooth, in the second step, we usually can show the temporal behavior process of interest "converges" to the behavior of a continuous dynamical system, and use existing tools to analyze the continuous dynamical system.

Though in the first step, techniques to reduce complex systems to simpler models are mostly ad hoc, for the second step, there are two dominant tools to connect discrete stochastic systems with continuous dynamical systems: *Wormold's method* [239], and *Pemantle's reinforced random walk* [189]. However, these two methods yield different convergence guarantee under different conditions. If a discrete stochastic system has certain nice conditions, Pemantle's method guarantees the limit behavior of the system and corresponding continuous deterministic system are similar. On the other hand, if a discrete stochastic system satisfies some stronger conditions, Wormold's method proves the system is close to the corresponding continuous deterministic system when the time span is comparable to the size of the dimension. In this thesis, we give two examples of relaxing the above limitations.

To apply Wormold's method requires a detailed understanding of the continuous deterministic behavior. However, for example, in [205], the number of infected nodes does not have a clean closed form, and the process does not meet the condition for Wormold's method. We give a new analysis of the process and provide a lower bound for the number of infected nodes due to the monotonic property of contagions.

On the other hand, Pemantle's method shines when we want to understand the limit behavior of a process. In [95] in the directed case the fraction of infected nodes possesses nice limit behaviors and using Pemantle's method off-shelf yields the fraction of infected nodes converges. However, to prove the fraction of infected nodes in undirected is constant larger the one in the directed case, the rate of convergence matters. In the paper, we give an asymptotic convergence rate for the fraction of infected nodes in the directed case which overcome the limitation of Pemantle's method.

1.4.2 Hitting time and potential function

A large volume of literature is devoted to bounding the hitting time on different Markov process and achieves fast convergence. The techniques typically employed are:

1. Showing the Markov chain has fast mixing time [171]
2. Reducing the dimension of the process into a small set of parameters (e.g., the frequency of each opinion) and using a mean-field approximation and concentration property to control the behavior of the process [31]
3. Using handcrafted potential functions [176]

In a recent work [206], we propose a general framework that upper bounds the consensus time which circumvent some limitations of the above three approaches.

In Chapter 9 I combine (2) and (3): I use (2) to derive a rough characterization of a process of interest, and then based on this characterization adaptively construct a potential function to fit the process. Additionally, in Chapter 10, I exploit the literature of dynamical systems and show the existence of global potential function of our system by phase portrait.

1.5 Structure of the thesis

Chapter 2 introduces notions for social networks and related dynamics which serve as vocabularies of most of our results. Readers may skip most of them and come back

to this chapter later.

In the first part, we collects results about social contagions which follows the outline in Section 1.1. There is little dependence between chapters. Chapter 3 presents a general Kleinberg's small world model, and discusses k-complex contagions and myopic routing on this model. This chapter focuses more on the new model, general Kleinberg's small world model, and less on complex contagions. Chapter 4 studies k-complex contagions on configuration model graphs with power-law degree distributions. This chapter mostly concentrates on a threshold property of complex contagions on the configuration model. Chapter 5 introduces a general threshold contagions model and considers the contagions on time evolving models. Additionally, the chapter also discuss more about the difference in complex contagions and simple contagions. Finally, Chapter 6 investigate influence maximization problem on graphs with community structures.

The next part is a bridge to the next part about opinion formations. In this part, Chapter 7 demonstrates an algorithm use strong ties to detect anomaly in social networks. Chapter 8 tries to understand the role of weak ties in opinion formations, and the model studied is related to the next chapter.

The last part focuses on opinion formation and community structures. The technical tools used are quite different than the previous chapters, Chapter 9 and 10 work on hitting time of Markov chains on Euclidean space which involves Markov chain theory, dynamical system theory, and spectral graph theory. Chapter 9 shows a family general rich-get-richer dynamics reaches consensus fast on graphs close to complete graph, dense Erdös-Rényi Graphs. Chapter 10 proves a dichotomy theorem about community structures and richer-get-richer dynamics. Finally, Chapter 11 studies opinion formation with a variety of interventions: 1) asymmetry in network topology, 2) adding uniform random edges, 3) the presence of stubborn nodes.

CHAPTER 2

Preliminaries

In this chapter, we introduce several notions related to social networks and dynamics on them. Section 2.3 presents concepts for dynamics, including ordinary differential equations, Markov chains, and martingales. With these concepts, we introduce our toolbox for characterizing those dynamics, with concentration of stochastic processes and linear algebra. In Section 2.4 we give a brief overview of dynamical systems. Finally, we define several common models for social networks and some related properties.

2.1 Dynamics on Social Networks

Dynamics on social networks model the temporal properties of a large number of individual agents interacting with each other. For example in contagions, each node can be represented as a binary state, infected or not infected, and each node gets infected based on the interaction between its neighbors. In social contagions, note that nodes cannot be uninfected and the process is monotone, and we are mostly interested in the span of contagions and the rate of contagions. On the other hand, for opinion formation, each node has a (possibly binary) opinion, and it updates its opinion based on its neighbors' choices. Here we are interested in does the process converges, and if the process reaches consensus. We will first define two concrete dynamics and generalize to mathematical modeling.

2.1.1 Social Contagions

Given a social network modeled as a graph $G = (V, E)$, in a cascade, a subset of nodes $S \subseteq V$ is chosen as the seed set; these seeds, being infected, then spread their influence across the graph according to some specified model. The general threshold model [104, 172] is a fairly general model to capture such intuition. Each node v has a

monotone function $g_v : \{0,1\}^{|\Gamma(v)|} \to [0,1]$, where $\Gamma(v)$ is the set of v's neighbors in a social networks. The functions g_v represents how much influence(via knowledge social pressure, ets) any set of neighbors has upon node v. Each node also has threshold R_v drawn uniformly and independently from the interval $[0,1]$. After an initial seed set is infected, a node v becomes infected at time t if $g_v(S_t) \geq R_v$ where S_t is the set of infected neighbors of v at time t. We will define two models for social contagions first and review other two later.

Definition 2.1 (r-complex contagions [104])**.** Given a constant $k \in \mathbb{N}$, a directed graph $G(V,E)$, and a set of initial infected nodes (seed) $X \subseteq V$, a *k-complex contagion* $\mathrm{CC}(G,r,X)$ is a contagion that initially infects vertices of X and spreads over graph G. The contagion proceeds in rounds. At each round, each vertex with at least r infected neighbors becomes infected.

We use $\sigma_{r,G}(S)$ to denote the total number of infected vertices at the end of the cascade, and $\sigma_{r,\mathcal{G}}(S) = \mathbb{E}_{G\sim\mathcal{G}}[\sigma_{k,G}(S)]$ if the graph G is sampled from some distribution $\mathcal{G}$. Notice that the function $\sigma_{r,G}(\cdot)$ is deterministic once the graph G and r are fixed.

Definition 2.2 (General threshold models)**.** Give a directed graph $G = (V,E)$, a set of initially infected nodes $X \subseteq V$, and a distribution $\mathcal{R}$ over non negative integer $\mathbb{R}_0$, a *General Threshold Contagion* $GTC(G,\mathcal{R},X)$ is a process $(X_t)_{t\geq 0}$ over subset of nodes. Initially, each node v samples a threshold R_v from the distribution $\mathcal{R}$ independently, and $X_0 = X$. The contagion proceeds in rounds. At each round $t+1$, each node v with at least threshold number of, R_v, infected out-neighbors/neighbors becomes infected,

$$X_{t+1} = X_t \cup \{v \in V : v \text{ has } R_v \text{ infected out neighbor}\}.$$

Other than the r-complex contagion, most cascade models are stochastic: the total number of infected vertices is nondeterministic but rather a *random variable*. $\sigma_G(S)$ usually refers to the *expected* number of infected vertices given the seed set S. A cascade model is *submodular* if, given any graph and $S \subseteq T \subseteq V$ and any vertex $v \in V \setminus T$, we have

$$\sigma_G(S \cup \{v\}) - \sigma_G(S) \geq \sigma_G(T \cup \{v\}) - \sigma_G(T),$$

and it is *nonsubmodular* otherwise. Typical submodular cascade models include *the linear threshold model* and *the independent cascade model* [128], which are studied

in an enormous past literature. The r-complex contagion, on the other hand, is a paradigmatic nonsubmodular model.

2.1.2 Opinion formations

A main difference between contagions and opinion formation is monotone property: in opinion formation, node can change back to its previous choice. This make the analysis become much more complicated. For example, in social contagions the process converges in linear number of iteration, but even showing the process converges is nontrival. Additionally, the relaveant time span for opinion formation dynamics to converge or consensus may be much longer which is challenging to analysis.

Suppose the opinion is binary. We can reuse the frame work of threshold model to capture opinion formation: Each node v has a monotone function $g_v : \{0,1\}^{|\Gamma(v)|} \to [0,1]$, where $\Gamma(v)$ is the set of v's neighbors in a social networks. The functions g_v represents how much influence(via knowledge social pressure, ets) any set of neighbors has upon node v. At each round, each node samples a threshold $R_{v,t}$ uniformly and independently from the interval $[0,1]$, and the node v picks the opinion 1 if $g_v(S_t) \geq R_{v,t}$ where S_t is the set of neighbors of v with opinion 1 at time t. Here are several examples:

Example 2.3. Given a graph (V,E), each agent $v \in V$ has an binary opinion $\chi(v) \in \{0,1\}$. We call $\chi = \{\chi(v) : v \in V\}$ the *opinion vector*, and $\sigma \in \{0,1\}$ we define the set of nodes with opinion σ as $\mathbf{x}(\sigma) = \{v \in V : \chi(v) = \sigma\}$. Given a set of function $\{g_v : \{0,1\}^{|\Gamma(v)|} \to [0,1] : v \in V\}$, and initial opinion vector χ_0, we introduce several well studies opinion formation models $(\chi_t)_{t\geq 0}$ which are Markov chains on opinion vectors:

Voter Model:[65, 114, 150, 151, 32, 47, 46, 212, 221, 222, 64] In the voter model, at each round, a random node chooses a random neighbor and updates to its opinion. Formally, let $g_v(\chi) = \frac{|\Gamma(v)\cap x(1)|}{|\Gamma(v)|}$. At round $t+1$, a random node $v \in V$ update to opinion 1 with probability $g_v(\chi_t)$.

Iterative majority:[136, 34, 125, 176, 224] In the iterative majority dynamics, in each round, a randomly chosen node updates to the opinion of the majority of its neighbors: At round $t+1$, a random node $v \in V$ update to opinion 1 with

probability $g_v(\chi_t)$ where

$$g_v(\chi) = \begin{cases} 1 & \text{if } x > 1/2; \\ 1/2 & \text{if } x = 1/2; \\ 0 & \text{if } x < 1/2. \end{cases} \quad \text{where } x = \frac{|\Gamma(v) \cap x(1)|}{|\Gamma(v)|}.$$

Iterative k-majority:[75, 66, 63, 31, 1] In this dynamics, in each round, a randomly chosen node collects the opinion of k randomly chosen (with replacement) neighbors and updates to the opinion of the majority of those k opinions: At round $t+1$, a random node $v \in V$ update to opinion 1 with probability $g_v(\chi_t)$ where

$$g_v(\chi) = \sum_{\ell=\lceil k/2 \rceil}^{k} \binom{k}{\ell} x^{\ell}(1-x)^{n-\ell} \text{ where } x = \frac{|\Gamma(v) \cap x(1)|}{|\Gamma(v)|}.$$

Iterative ρ-noisy majority model: [85, 100] In this dynamics, in each round, a randomly chosen node updates the majority opinion of its neighbors with probability $1-\rho$ and uniformly at random with probability ρ: At round $t+1$, a random node $v \in V$ update to opinion 1 with probability $g_v(\chi_t)$ where

$$g_v(\chi) = \begin{cases} 1-\rho/2 & \text{if } x > 1/2; \\ 1/2 & \text{if } x = 1/2; \\ \rho/2 & \text{if } x < 1/2. \end{cases} \quad \text{where } x = \frac{|\Gamma(v) \cap x(1)|}{|\Gamma(v)|}.$$

2.1.3 Decentralized routing algorithm

In the *decentralized routing algorithm*, a message is passed to one of its (local or long-range) contacts using only local information. Given the source s and destination t in the graph, we denote the routing process/algorithm $\mathbb{A}$: a sequence of nodes on the graph $(x_i)_{i\geq 0}$ where $x_0 = s$. The *delivery time* from s to t of algorithm $\mathbb{A}$ is defined as $\min\{i \geq 0 : x_i = t\}$ which is a random variable with σ-space generated by $HetK_{p,q,\mathcal{D}}(n)$ and the myopic routing algorithm. The *expected delivery time* of a decentralized algorithm $\mathbb{A}$ is the expected delivery time for uniformly chosen sources s and destinations t. The *myopic greedy algorithm* routes the message from the current location to be as close as possible to the destination vertex (according to the grid distance) using only one hop from the current node.

2.2 Models for Dynamics

Though the models in the above section have have high dimension spaces, e.g. $\{0,1\}^V$, we can often project those process into a smaller phase space such as the number of infected nodes in each round or the fraction of infection in the first k hightest degree nodes. Here we present three types of models capture the behavior of a large population of agents in a *phase space*, $\mathcal{X}$—a compact manifold space—that update in accord to some function $f : \mathcal{X} \to \mathcal{X}$. We will always use $\mathcal{X} = \mathbb{R}^d$, which, technically, must be compactified by adding infinity. We will say $f \in \mathcal{C}^r$ if the r-th derivative of f is continuous.

1. ***Vector fields*** or *ordinary differential equation with* f solve for:

$$\frac{d}{dt}\boldsymbol{x} = f(\boldsymbol{x}). \tag{2.1}$$

   The result is the continuous function $\varphi : \mathcal{X} \times \mathbb{R} \to \mathcal{X}$ such that $\varphi(x, 0) = x$ and $\frac{d}{dt}\varphi(x,t) = f(\varphi(x,t))$ for all $t \in \mathbb{R}, x \in \mathcal{X}$.

2. ***Maps*** or *difference equations with* f are discrete time processes. Given the step size at each round by a sequence of positive number (γ_k), a corresponding discrete process can be defined as follows:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \gamma_k f(\boldsymbol{x}_k), \tag{2.2}$$

   and the range of change of each update, $\|f(\boldsymbol{x}_k)\|$, is bounded when the process is in some compact set $B \subset \mathbb{R}^d$.

3. ***Reinforced random walks with*** f consider the evolution of a process subject to an unbiased stochastic perturbation. Let $(\boldsymbol{X}_k, \mathcal{F}_k)$ be a random process in $\mathcal{X}$ with filtration $\mathcal{F}$ which can be composed of a predictable part $f(\boldsymbol{X}_k)$ and noise part U_{k+1}:

$$\boldsymbol{X}_{k+1} = \boldsymbol{X}_k + \gamma_k \left(f(\boldsymbol{X}_k) + \boldsymbol{U}_{k+1}\right) \tag{2.3}$$

   such that for all $\boldsymbol{x} \in \mathbb{R}^d$, $\mathbb{E}[U_{k+1} \mid \mathcal{F}_k] = 0$.

Here are two examples for (2.3):

Example 2.4. A *stochastic approximation algorithm* $\{X_i\}$ is a stochastic process taking values in $\mathbb{R}$, adapted to the filtration $\mathcal{F}_i$, that can be represented as

$$X_{i+1} - X_i \mid \mathcal{F}_i = \gamma_{i+1}\left(f(X_i) + U_{i+1}\right).$$

A formal definition is in Definition 2.5. On the other hand, we can also set the step size γ_k be a constant $1/n$ as follows:

$$X_{i+1} - X_i \mid \mathcal{F}_i = \frac{1}{n} \left(f(X_i) + U_{i+1} \right).$$

The following two sections give a more detailed characterizations of above examples which can be skip in the first reading. Intuitively, there are two lines of work which both show the behavior of Equation (2.3) is close to the behavior of Equation (2.1). One line of work show the process converges to the fixed points of f when i is large enough but does not provide convergence rate. Another line of work show in constant step size $\gamma_i = 1/n$ the temporal behavior if close to (2.1) for all small enough $i = O(n)$. However, some of our problem need both rate of convergence for large i.

2.2.1 Stochastic approximation algorithm

Definition 2.5. A *stochastic approximation algorithm* $\{X_i\}$ is a stochastic process taking values in $\mathbb{R}$, adapted to the filtration $\mathcal{F}_i$, that can be represented as

$$X_{i+1} - X_i \mid \mathcal{F}_i = \gamma_{i+1} \left(g(X_i) + U_{i+1} \right),$$

and the following conditions hold almost surely: There exists positive constants c_l, c_h, K_u, K_g, K_e such that for all $i \geq 1$

1. $\frac{c_l}{i} \leq \gamma_i \leq \frac{c_h}{i}$,
2. $|U_i| \leq K_u$,
3. $|g(X_i)| \leq K_g$,
4. $|E[\gamma_{i+1} U_{i+1} \mid \mathcal{F}_i]| \leq K_e \gamma_i^2$,

Theorem 2.6 ([188, 189])**.** *If Y_n is a stochastic approximation algorithm with continuous feedback function g then:*

1. $\lim_{n\to\infty} Y_n$ *exists almost surely and is in* $Q_g = \{x : g(x) = 0\}$.
2. *Suppose there is an unstable fixed point p and $d > 0$ such that $\forall x : 0 < |x-p| < d$ and $K_l \leq E[U_{n+1}^2 \mid \mathcal{F}_n] \leq K_g$ holds for some $K_l, K_g > 0$, whenever $0 < |Y_n - p| < d$. Then $P[Y_n \to p] = 0$.*

3. *Suppose $p \in Q_g$ is a stable fixed point then $P[X_n \to p] > 0$.*

4. *If $p \in T_g$ and f is differentiable, $P[X_n \to p] > 0$.*

2.2.2 Wormald's system

There is extensive literature about stochastic processes and its mean field approximation e.g. [86]. Given a non-homogeneous random walk $X(t)$ in $\mathbb{Z}^\ell$ we can associate the behavior of it with the corresponding differential equation in $\mathbb{R}^\ell$. Formally, let $X_n(t)$ be a discrete time Markov chain on $\mathbb{Z}^\ell$ with parameter n which is time-homogeneous and the increments of the walk are uniformly bounded by β. As a result, random vectors $X_n(t+1) - X_n(t)$ have well defined moments, which depend on $X(t)$ and n. In particular, an important quantity is the one-step *mean drift vector* $F_n : \mathbb{R}^\ell \to \mathbb{R}^\ell$ defined to be

$$F_n(X) = \mathbb{E}[X_n(t+1) - X_n(t) \mid X_n(t) = X]. \tag{2.4}$$

In particular if there exists a function f independent of n such that $F_n(X) = f(\frac{X}{n})$, then there is a close relationship between X and the x which we define as a solution of the following autonomous differential system

$$x' = f(x) \tag{2.5}$$

with initial condition $x(0) = X(0)/n$.

The following theorem shows that the differential equation approximates the original random walk $X(t)$ such that $X(t) \approx n\hat{x}(\frac{t}{n})$ under proper conditions.

Theorem 2.7 (Wormald's method [239])**.** *For $1 \le \ell \le a$ where a is fixed, let $y_\ell : S^{(n)+} \to \mathbb{R}$ and $f_\ell : \mathbb{R}^{a+1} \to \mathbb{R}$ such that for some constant C_0 and all ℓ, $|y_\ell(h_t)| < C_0 n$ for all $h_t \in S^{(n)+}$ and n. Let $Y_\ell(t)$ denote the random counterpart of $y_\ell(h_t)$. Assume the following three conditions hold:*

1. *(Boundedness) For some functions $\beta = \beta(n) \ge 1$ and $\gamma = \gamma(n)$, the probability that*

$$\max_\ell |Y_\ell(t+1) - Y_\ell(t)| \le \beta$$

conditional upon H_t, is at least $1 - \gamma$ for $t < T_D$.

2. *(Trend) For some function $\lambda_1 = \lambda_1(n) = o(1)$, for all $\ell \le a$*

$$|\mathbb{E}[Y_\ell(t+1) - Y_\ell(t) \mid H_t] - f_\ell\left(\frac{t}{n}, \frac{Y_1(t)}{n}, \ldots, \frac{Y_a(t)}{n}\right)| \le \lambda_1$$

for $t \leq T_D$.

3. *(Lipschitz) Each function f_ℓ is continuous, and satisfies a Lipschitz condition, on*

$$D \cap \{(t, z_1, ..., z_a) : t \geq 0\},$$

with the same Lipschitz constant for each ℓ.

Then the following are true.

1. *For $(0, \hat{z}_1, ..., \hat{z}_a) \in D$ the system of differential equations*

$$\frac{dz_\ell}{dx} - f_\ell(x, z_1, .., z_a), \ell = 1, ..., z \tag{2.6}$$

have a unique solution in D for $z_\ell : \mathbb{R} \to \mathbb{R}$ passing through $z_\ell(0) = \hat{z}_\ell$ for $1 \leq \ell \leq a$, which extends to points arbitrarily close to the boundary of D;

2. *Let $\lambda > \lambda_1 + C_0 n\gamma$ with $\lambda = o(1)$. For a sufficiently large constant C with probability $1 - O\left(n\gamma + \frac{\beta}{\lambda} \exp\left(\frac{-n\lambda^3}{\beta^3}\right)\right)$,*

$$Y_\ell(t) = nz_\ell\left(\frac{t}{n}\right) + O(\lambda n) \tag{2.7}$$

uniformly for $0 \leq t \leq \sigma n$ and for each ℓ where $z_\ell(x)$ is the solution in Equation (2.6) with $\hat{z}_\ell = \frac{Y_\ell(t)}{n}$, and $\sigma = \sigma(n)$ is the supremum of those x to which the solution can be extended before reaching within ℓ^∞-distance $C\lambda$ of the boundary of D.

2.3 Mathematical tools

2.3.1 Markov chain and Martingale

Markov chain and hitting time Let $\mathcal{M} = (X_t, P)$ be a discrete time-homogeneous Markov chain with finite state space Ω and transition matrix P. For $x, z \in \Omega$, we define $\tau_a(x)$ to be the ***hitting time*** for a with initial state x:

$$\tau_a(x) \triangleq \min\{t \geq 0 : X_t = a, X_0 = x\},$$

and $\tau_A(x)$ to be the hitting time to a set of state $A \subseteq \Omega$:

$$\tau_A(x) \triangleq \min\{t \geq 0 : X_t \in A, X_0 = x\}.$$

By the Markov property, the expected hitting time can be written as linear equation.

$$\mathbb{E}_{\mathcal{M}}[\tau_A(x)] = \begin{cases} 1 + \sum_{y\in\Omega} P_{x,y}\mathbb{E}_{\mathcal{M}}[\tau_A(y)] & \text{if } x \notin A, \\ 0 & \text{if } x \in A \end{cases}$$

Due to the memory-less property of Markov chain, sometimes it is useful to analyze its first step. Let's consider a general measurable function $w : \Omega \mapsto \mathbb{R}$. If the Markov chain starts at state $X = x$, the next state is the random variable X', then the average change of $w(X)$ in one transition step is given by

$$(\mathcal{L}w)(x) \triangleq \mathbb{E}_{\mathcal{M}}[w(X') - w(X)|X = x] = \sum_{y\in\Omega} P_{x,y}w(y) - w(x)$$

To reduce the notation we will use $\mathbb{E}_{\mathcal{M}}[w(X')|X]$ to denote the expectation of the measurable function $w(X')$ given the previous state at X.

Definition 2.8. Given Markov chain $\mathcal{M}$ with state space Ω, $D \subsetneq \Omega$, and two real-valued functions ψ, ϕ with domain Ω, we define the ***Poisson equation*** as the problem of solving the function $w : \Omega \mapsto \mathbb{R}$ such that

$$\begin{aligned} \mathcal{L}w(x) &= -\phi(x) \text{ where } x \in D, \\ w(x) &= \psi(x) \text{ where } x \in \partial D. \end{aligned}$$

where the $\partial D \triangleq \cup_{x\in D} \text{supp } p(x, \cdot) \setminus D$ is the exterior boundary of D w.r.t the Markov chain.

Note that solving the expected hitting time of set A is a special case of the above problem by taking $D = \Omega \setminus A$, $\phi(x) = 1$ and $\psi(x) = 0$. The next fundamental theorem shows that super solutions to an associated boundary value problem provide upper bounds for the Poisson equation in Definition 2.8.

Theorem 2.9 (Maximum principle [81])**.** *Given Markov Chain $\mathcal{M}$ with state space Ω, $D \subsetneq \Omega$, and two real-valued functions ψ, ϕ with domain Ω, suppose $s : \Omega \mapsto \mathbb{R}$ is a non-negative function satisfying*

$$\begin{aligned} \mathcal{L}s(x) &\leq -\phi(x) \text{ where } x \in D, \\ s(x) &\geq \psi(x) \text{ where } x \in \partial D. \end{aligned}$$

Then $s(x) \geq w(x)$ for all $x \in D$.

Corollary 2.10 (Super solution for hitting time)**.** *Given Markov Chain $\mathcal{M}$ with state space Ω and a set of states $A \subsetneq \Omega$, suppose $s_A : \Omega \mapsto \mathbb{R}$ is a non-negative function satisfying*

$$\begin{aligned} \mathcal{L}s_A(x) &\leq -1 \text{ where } x \notin A, \\ s_A(x) &\geq 0 \text{ where } x \in A. \end{aligned} \tag{2.8}$$

Then $s_A(x) \geq \mathbb{E}_{\mathcal{M}}[\tau_A(x)]$ for all $x \notin A$. Moreover we call s_A a ***potential function*** *for short.*

Martingale and Stoppig time In this section we will define martingales and some of its properties. Let $\mathcal{F} = (\mathcal{F}_k)_k$ be a filtration, that is an increasing sequence of σ-field. A sequence X_k is said to be adapted to $\mathcal{F}_k$ if $X_k \in \mathcal{F}_k$ for all k. If X_k is sequence with 1) $\mathbb{E}|X_k| < \infty$, 2) X_k is adapted to $\mathcal{F}_k$, and 3) $\mathbb{E}[X_{k+1} \mid \mathcal{F}_k] = X_k$ for all k, X is saied to be a *martingale* with respect to $\mathcal{F}_k$.

T is called a *stopping time* for $\mathcal{F}$ if and only if $\{T = k\} \in \mathcal{F}_k, \forall k$. Intuitively, this condition means that the "decision" of whether to stop at time k must be based only on the information present at time k, not on any future information.

Theorem 2.11 (Martingale Stopping theorem)**.** *If $(W_k)_{0 \leq k \leq n}$ is a martingale with respect to $(\mathcal{F}_k)_{0 \leq k \leq n}$ and if T is a stopping time for $(\mathcal{F}_k)_{0 \leq k \leq n}$ such that W_k is bounded, T is bounded, $\mathbb{E}[T] < \infty$, and $\mathbb{E}[|W_{k+1} - W_k| \mid \mathcal{F}_k]$ is uniformly bounded, then*

$$\mathbb{E}[W_T] = \mathbb{E}[W_0].$$

2.3.2 Linear Algebra

In this section, we state some basic results of linear algebra. Given symmetric matrices $A, A' \in \mathbb{R}^{d \times d}$, $A \prec A'$ denotes $A' - A$ is positive definite.

Definition 2.12 (Majorize [116])**.** Given two real-valued sequences $x, y \in \mathbb{R}^d$, we say that x *majorizes* y if for all $k \leq n$ and for all length k-sub-sequence $i_1 < i_2 < \ldots < i_k$,

$$\sum_j^k x_{i_j} \geq \sum_j^k y_{i_j}$$

with equality for $k = d$.

The following characterization of the majorization relationship tells us that the

eigenvalues of the Hermitian part of a matrix A majorize the Hermitian parts of the eigenvalues of A.

Theorem 2.13. *Let $x \in \mathbb{R}^d$ and $z \in \mathbb{C}^d$. Then x majorizes $\Re(z)$ if and only if there is an $A \in \mathbb{R}^{d\times d}$ such that z is the vector of eigenvalues of A and x is the vector of eigenvalues of $H(A) = \frac{1}{2}(A + A^*)$*

Corollary 2.14 (Quadratic form)**.** *Let $A \in \mathbb{R}^{d\times d}$ with eigenvalues $\rho(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_d\}$ with $\lambda_{\min} \triangleq \min \Re(\lambda_i)$ and $\lambda_{\max} \triangleq \max \Re(\lambda_i)$. For all $v \in \mathbb{R}^d$,*

$$\lambda_{\min}\|v\|^2 \le v^\top Av \le \lambda_{\max}\|v\|^2.$$

Note that the process in (2.2) is exactly the explicit Euler method for (2.1). The following lemma is useful to show these two processes are close to each other.

Lemma 2.15 (Discrete Gronwall lemma)**.** *Let $a_{k+1} \le (1 + \frac{1}{n}L)a_k + b$ with $n > 0$, $L > 0$, $b > 0$ and $a_0 = 0$. Then*

$$a_k \le \frac{nb}{L}\left(\exp\left(\frac{k}{n}\right) - 1\right).$$

Theorem 2.16 (Convergence of Eular forward method [16])**.** *Let $f : D \to \mathbb{R}^d \in \mathcal{C}^1$ such that the derivative f' exists and is continuous with $\|f(x)\| \le M$, and $\|f(x,t) - f(z,t)\| \le L\|x - z\|$. Then in Equation (2.5), for all $t > 0$ the $Z^{(t)}$ differs from the true solution $\boldsymbol{z}$ by at most*

$$\|Z^{(nt)} - \boldsymbol{z}(t)\| \le \frac{M}{n}(e^{Lt} - 1).$$

2.3.3 Concentration of Stochastic Process

We say a sequence of events $\{A_n\}_{n\ge1}$ happens *with high probability* if $\lim_{n\to\infty} \Pr[A_n] = 1$ that is $\Pr[A_n] = 1 - o(1)$.

Theorem 2.17 (A Chernoff bound [78])**.** *Let $X \triangleq \sum_{i=1}^n X_i$ where X_i for $i \in [n]$ are independently distributed in $[0, 1]$. Then for $0 < \epsilon < 1$*

$$\Pr[X > (1 + \epsilon)\mathbb{E}X] \le \exp\left(-\frac{\epsilon^2}{3}\mathbb{E}X\right)$$
$$\Pr[X < (1 - \epsilon)\mathbb{E}X] \le \exp\left(-\frac{\epsilon^2}{2}\mathbb{E}X\right)$$

Theorem 2.18 (Azuma Inequality). *Let $(W_k)_{0\le k\le n}$ be a martingale with c_k such that $|W_{k+1} - W_k| \le c_k$. Then,*

$$\Pr[W_n \ge W_0 + t] \le \exp\left(-\frac{t^2}{2\sum c_k^2}\right).$$

The following theorem shows this concentration property is dimension free.

Theorem 2.19 (Vector-valued martingale [124, 110]). *Let g be a vector-valued function of n random variables $X = (X_1, \ldots, X_n)$. If $\sup_{x,x'} \|g(x) - g(x')\| \le c_i$ where x and x' only differ by one variable, $x = x_1, \ldots, x_i, \ldots, x_n, x' = x_1, \ldots, x_i', \ldots, x_n$. Then,*

$$\Pr\left[\|g(X) - \mathbb{E}[g(X)]\| \ge t\right] \le 20\exp\left(-\frac{t^2}{2\sum_i c_i^2}\right).$$

The following exponential inequality for maximum of martingales can save an extra union bound.

Theorem 2.20 (Maximum tail [93, 90]). *Let $W_0, W_1, \ldots$ be a martingale with c_k and D such that $|W_{k+1} - W_k| \le c_k$ and $\sup_k |W_{k+1} - W_k| \le D$. Then, for any $t \ge 0$*

$$\Pr\left[\max_{k\le n} W_k \ge W_0 + t\right] \le \exp\left(-\frac{t^2}{2\sum c_k^2 + Dt}\right).$$

If a bounded function g on a probability space (X, P) which is Lipschitz for most of the measure in X, then the following theorem prove a concentration property of g by using union bound and Azuma inequality.

Theorem 2.21 (Bad events [78]). *Let $(W_k)_{0\le k\le n}$ be a martingale which is bounded, $m \le W_n \le M$. Let $\mathcal{B}$ be a (bad) event such that there is a sequence c_k such that $|\mathbb{E}[W_T \mid \mathcal{F}_{k-1}, W_k, \neg\mathcal{B}] - \mathbb{E}[W_T \mid \mathcal{F}_{k-1}, W_k', \neg\mathcal{B}]| \le c_k$. Then,*

$$\Pr[W_n \ge W_0 + t + (M - m)\Pr[\mathcal{B}]] \le \exp\left(-\frac{2t^2}{\sum c_k^2}\right) + \Pr[\mathcal{B}].$$

The following statement is powerful when the random variables are not mutually infependent:

Theorem 2.22 ([25]). *Let $\mathcal{X} = (x_1, \ldots, x_N)$ be a finite set of N real numbers, that $X_1, \ldots, X_n$ denote a random sample without replacement from $\mathcal{X}$ and that $Y_1, \ldots, Y_n$ denote a random sample with replacement from $\mathcal{X}$. If $f : \mathbb{R} \mapsto \mathbb{R}$ is continuous and*

convex, then

$$\mathbb{E}f\left(\sum_{i=1}^{n} X_i\right) \leq \mathbb{E}f\left(\sum_{i=1}^{n} Y_i\right)$$

We state some concentration properties of random perfect matchings. The following proposition, which follows from a straightforward application of Azuma's inequality, appears as Theorem 2.19 in Wormald [240].

Proposition 2.23 (Concentration on uniform matchings)**.** *Let $m \in \mathbb{N}$ and $[2m]$ be a finite set with even cardinality and H be a real function on matchings of $[2m]$ such that*

$$|H(\sigma') - H(\sigma)| \leq c$$

if σ, σ' differ by at most a switch. Then, if σ is a uniformly drawn matching of $[2m]$, and $\lambda > 0$

$$\begin{matrix} \Pr[H(\sigma) \geq \mathbb{E}H(\sigma) + \lambda] \\ \Pr[H(\sigma) \leq \mathbb{E}H(\sigma) - \lambda] \end{matrix} \leq \exp\left(\frac{-\lambda^2}{2mc^2}\right).$$

We will consider functions H which count the number of edges between two sets of stubs S, T.

Given a matching σ on $[2m]$ and two disjoint subsets S and T of $[2m]$, let $C(\sigma; S, T)$ denote the size of matching between S, T. We will use random variable $C(S, T)$ to denote $C(\sigma; S, T)$ when σ is a uniformly drawn matching of $[2m]$.

Lemma 2.24 (Multiplicative error)**.** *Given two disjoint subsets S and T of $[2m]$ with cardinality s and t respectively,*

$$\Pr\left[(1-\delta)\frac{s\,t}{2m-1} < C(S,T) < (1+\delta)\frac{s\,t}{2m-1}\right] \geq 1 - 2\exp\left(-\frac{\delta^2 s^2 t^2}{2m(2m-1)^2}\right)$$

The following lemma is stronger when the size of S, T is relatively small.

Lemma 2.25 (Additive error)**.** *Given two disjoint subsets S and T of $[2m]$ with cardinality s and t respectively,*

$$\Pr[C(S,T) < k] \leq \exp\left(-\frac{st}{2m}\right)$$

when k is a constant and $s, t = \omega(1)$.

Proof. We reveal the matching in S one by one in a fixed but arbitrary order. For each step we pick a stub $\tau \in S$ and uniformly match it. Let X_τ denote the indicator

function that stub τ is connected to T. We have $\Pr[X_t = 0 | \sum_{\tau=0}^{t-1} X_\tau < k] \geq 1 - \frac{t-k}{2m-1} \geq 1 - \frac{t}{m}$ and there are at least $s/2$ such τ. Therefore

$$\Pr[C(S,T) < k] \leq \Pr\left[\sum_{\tau \in S} X_\tau < k\right] \leq \left(1 - \frac{t}{m}\right)^{s/2} \leq \exp(-\Theta(\frac{st}{m})).$$

□

2.4 Primer of dynamical systems

First, let us define some basic notions which are mostly from Robinson [197]. Let $\mathcal{X}$ be $\mathbb{R}^d$. A $\mathcal{C}^r$-flow φ is defined to be a $\mathcal{C}^r$-function $\varphi : \mathcal{X} \times \mathbb{R} \to \mathcal{X}$ with the property that $\forall x_0 \in \mathcal{X}, t_1, t_2 \in \mathbb{R}$,

$$\varphi(x_0, 0) = x_0; \qquad \varphi(x_0, t_1 + t_2) = \varphi(\varphi(x_0, t_1), t_2).$$

Given function $f \in \mathcal{C}^r$, initial condition $\boldsymbol{x} \in \mathcal{X}$, and time $t \in \mathbb{R}$, the solution of (2.1) forms a $\mathcal{C}^r$-flow $\varphi(\boldsymbol{x}, t; f)$ called ϕ the *flow with* f. We call a set $B \subseteq \mathcal{X}$ *positive invariant* if and only if for all $x \in B$ and $t \geq 0$, $\varphi(x,t) \in B$, *negative invariant* if and only if it is true for all $t \leq 0$, and *invariant* if and only if it is true for all $t \in \mathbb{R}$.

The *trajectory* or *orbit* of a point $\boldsymbol{x} \in \mathcal{X}$ is the set $\mathcal{O}_x = \{\varphi(\boldsymbol{x}, t; f) : t \in \mathbb{R}\}$. A point $\boldsymbol{x} \in \mathcal{X}$ is a ***fixed point*** if $\mathcal{O}_x = \{x\}$ that is $f(\boldsymbol{x}) = 0$, and we use Fix_f to denote the set of fixed points. The ω-limit set of $\boldsymbol{x}$ is the set of "limit points" such that $\omega(\boldsymbol{x}) = \{\boldsymbol{y} : \exists \{t_l\}_l \to +\infty, \lim_{l\to\infty} d(\varphi(\boldsymbol{x}, t_l), \boldsymbol{y}; f) = 0\}$ and α-limit is defined similarly with $t \to -\infty$.

2.4.1 Fundamental theorem of dynamical system

Now, we define different motion of "returning" which is important for understanding limit behavior of the system. For simplicity, we only define the notion for flow.

Fixed point A point $\boldsymbol{x} \in \mathcal{X}$ is a *fixed point* if $\mathcal{O}_x = \{x\}$ that is $f(\boldsymbol{x}) = 0$, and we use Fix_f to denote the set of fixed points.

Periodic point A point $\boldsymbol{x} \in \mathcal{X}$ is a *periodic point* of f if $\exists T \geq 0$ such that $\varphi(\boldsymbol{x}, T; f) = \boldsymbol{x}$, and we use Per_f to denote the set of periodic points.

ω-recurrent For other non-periodic points $\boldsymbol{x} \in$, the long term behavior can be characterized as *ω-limit set* of $\boldsymbol{x}$: $\omega(\boldsymbol{x}) = \{\boldsymbol{y} : \exists t_l \to +\infty, \lim_{l\to\infty} d(\varphi(\boldsymbol{x}, t_l), \boldsymbol{y}) = 0\}$,

and we call $\boldsymbol{x}$ *ω-recurrent* if $\boldsymbol{x} \in \omega(\boldsymbol{x})$ If we change $+\infty$ to $-\infty$ in above definition, it called *α-limit set* $\alpha(\boldsymbol{x})$ of $\boldsymbol{x}$. We call $L_f \triangleq \overline{\cup_{\boldsymbol{x}\in\mathcal{X}}\omega(\boldsymbol{x})} \cup \overline{\cup_{\boldsymbol{x}\in\mathcal{X}}\alpha(\boldsymbol{x})}$ the *limit set of* f.

Chain recurrent An ϵ-chain of length T from a point $\boldsymbol{x}$ to $\boldsymbol{y}$ is a sequence of points $(\boldsymbol{x}_\ell)_{0\le\ell\le n}$ and a sequence of time $(t_\ell)_{1\le\ell\le n}$ such that $\boldsymbol{x}_0 = \boldsymbol{x}$, $\boldsymbol{x}_n = \boldsymbol{y}$, and $d(\varphi(\boldsymbol{x}_{i-1}, t_i), \boldsymbol{x}_i) < \epsilon$ for $1 \le \ell \le n$ with $t_\ell \ge 1$ and $\sum_\ell t_\ell = T$. We define a relation $\sim_{\mathcal{CR}}$ on $\mathcal{CR}_f$. Similar to ω-limit we define $\Omega^+(\boldsymbol{x}) = \cap_{\epsilon>0,T>0}\{\boldsymbol{y} : \exists$an ϵ, T chain from $\boldsymbol{x}$ to $\boldsymbol{y}\}$, and a point $\boldsymbol{x}$ is said to be *chain recurrent* for the flow f if $\boldsymbol{x} \in \Omega^+(x)$. The set of chain recurrent points of f is called the *chain recurrent set* of f denoted as $\mathcal{CR}_f$. We say $\boldsymbol{x} \sim_{\mathcal{CR}} \boldsymbol{y}$ if and only if $\boldsymbol{x} \in \Omega^+(\boldsymbol{y})$ and $\boldsymbol{y} \in \Omega^+(\boldsymbol{x})$.

It is not hard to show

$$\mathrm{Fix}_f \subseteq \mathrm{Per}_f \subseteq L_f \subseteq \mathcal{CR}_f \subseteq \mathcal{X}$$

.

An opposite concept to "recurrence" is transit. How we show all the non recurrent points are transit? An ideal method is to find a "potential function", $\Psi : \mathcal{X} \to \mathbb{R}$ of the system such that Ψ decrease along the trajectory of the system.

Definition 2.26 (Complete Lyapunov function)**.** Let $\varphi(\cdot,\cdot; f)$ be a flow with f on a metric space $\mathcal{X}$. A *complete Lyapunov function* for $f \in \mathcal{C}^0$ is a continuous function $\Psi : \mathcal{X} \to \mathbb{R}$ such that

1. For all $s < t$ and $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{CR}_f$, $\Psi(\varphi(\boldsymbol{x}, s; f)) > \Psi(\varphi(\boldsymbol{x}, t; f))$,
2. for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{CR}_f$, $\boldsymbol{x} \sim_{\mathcal{CR}} \boldsymbol{y}$ if and only if $\Psi(\boldsymbol{x}) = \Psi(\boldsymbol{y})$.
3. $\Psi(\mathcal{CR}_f)$ is nowhere dense subset of $\mathbb{R}$.

By constructing a complete Lyapunov function, Conley [62] shows:

Theorem 2.27 (Fundamental theorem of dynamical system [62])**.** *Every flow on a compact metric space has a complete Lyapunov function $V : \mathcal{X} \to \mathbb{R}_+$.*

One interpretation of this theorem is that the space of the dynamics can be decomposed into two parts: points exhibiting a particular type of recurrence, and points proceed in a gradient-like fashion.

2.4.2 Local recurrence

a point $\bar{x} \in \mathbb{R}^\ell$ is called an *equilibrium point* of system (2.1) if $f(\bar{x}) = 0$. Moreover the equilibrium $\bar{x}$ is *asymptotically stable* if $\forall \epsilon > 0, \exists \delta > 0$ such that $||x(0) - \bar{x}|| \leq \delta \Rightarrow ||x(t) - \bar{x}|| \leq \epsilon, \forall t$ and $\exists \delta > 0$ such that $\lim_{t\to\infty} ||x(t) - \bar{x}|| = 0$. The stability of the system can be determined by the linearization of the system which is stated below.

Theorem 2.28 (Lyapunov's indirect method [220])**.** *Let x^* be an equilibrium point for $x' = f(x)$ where $f : \mathcal{D} \to \mathbb{R}^d$ is continuously differentiable and $\mathcal{D}$ is a neighborhood of x^*. Let $A = \frac{\partial f}{\partial x}|_{x=x^*}$ then x^* is asymptotically stable if A is* Hurwitz*, that is $Re(\lambda_i) < 0$ for all eigenvalues of A.*

Moreover, there exists an close set $U \subseteq \mathcal{D}$ and $x^* \in U$ and a potential function $V : U \to \mathbb{R}$ such that $V(x^*) = 0$, and $V(x) > 0, \frac{d}{dt}(V(x)) < 0$ for $x \in U \setminus x^*$.

This section has three parts. We first introduce linear flows and linear mappings, then talk about gradient-like flows which contain gradient flows as a special case. Finally we state some results of planar flow.

2.4.3 Linear dynamics

Here we introduce some important properties of linear flow (and mapping) in $\mathbb{R}^d$. Given a matrix $A \in \mathbb{R}^{d\times d}$,

$$\frac{d}{dt}x(t) = Ax(t) \qquad (\text{and } x_{k+1} = Ax_k)$$

which has a closed form solution $\varphi(x_0, t; A) = \exp(At)x_0$ and $\varphi(x_0, k; A) = A^k x_0$ respectively.

The long term behavior (e.g., converges to 0, diverges to infinite, or rotating) of the above systems both depend on the real part of eigenvalues of A. For linear flow, we denote the set of eigenvalues for the (real) matrix A by

$$\rho(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_s, \lambda_{s+1}, \ldots, \lambda_{s+u}, \lambda_{s+u+1}, \ldots, \lambda_{s+u+c}\},$$

where $\Re(\lambda_i) < 0$ for all $1 \leq i \leq s$, $\Re(\lambda_{s+i}) > 0$ for all $1 \leq i \leq u$, and $\Re(\lambda_{s+u+i}) = 0$

for all $1 \leq i \leq c$. We define the *stable/unstable/center eigenspace* of A,

$$E^s = \{v : v \text{ is a generalized eigenvector for an eigenvalue } \lambda_i, \Re(\lambda_i) < 0\};$$
$$E^u = \{v : v \text{ is a generalized eigenvector for an eigenvalue } \lambda_{s+i}, \Re(\lambda_{s+i}) > 0\};$$
$$E^c = \{v : v \text{ is a generalized eigenvector for an eigenvalue } \lambda_{s+u+i}, \Re(\lambda_{s+u+i}) = 0\}.$$

Moreover the subspaces E^s, E^u and E^c are invariant under the flow and map with A.

Definition 2.29. We say $A \in \mathbb{R}^{d\times d}$ is ***hyperbolic*** if $E^c = \emptyset$, i.e., for all $\lambda \in \rho(A)$

$$\Re(\lambda) \neq 0.$$

A hyperbolic A is called *attracting* (or *repelling*) if for all $\lambda \in \rho(A), \Re(\lambda) < 0$, (or $\Re(\lambda) > 0$) respectively. Finally, if A is neither attracting nor repelling, we call it *saddle*.

Therefore, there is a *hyperbolic splitting* of $\mathbb{R}^d$ such that $\mathbb{R}^d = E^s \oplus E^u$, and two positive real numbers λ_s and λ_u such that

$$\max_{1\leq i\leq s} \Re(\lambda_i) = -\lambda_s < 0 \text{ and } \min_{1\leq i\leq u} \Re(\lambda_{s+i}) = \lambda_u > 0.$$

2.4.4 Nonlinear dynamics: Morse-Smale, gradient-like, and gradient flow

For nonlinear dynamics, we first characterize some local properties of fixed points.

Definition 2.30 (Attracting, repelling, and saddle points)**.** Given a fixed point $\boldsymbol{x} \in \mathcal{X}$, with the linear approximation matrix $A = \nabla f|_{\boldsymbol{x}}$. A fixed point $\boldsymbol{x}$ is called ***hyperbolic*** if A is hyperbolic (Definition 2.29). Similarly, a fixed point $\boldsymbol{x}$ is respectively an ***attracting, repelling or saddle fixed point*** if A is attracting, repelling or saddle.

We use $\text{Attract}_f, \text{Repel}_f$, and Saddle_f to denote the sets of attracting, repelling and saddle fixed points respectively. If all the fixed points are hyperbolic, $\text{Fix}_f = \text{Attract}_f \cup \text{Repel}_f \cup \text{Saddle}_f$.

Moreover, there is a more general family of dynamic system called ***Morse-Smale flows*** which allows the ω- and α-limit to be cycles which is introduced in Definition 2.31. Gradient-like systems share several properties with gradient flows: no

complicated recurrent motion and the existence of "potential function" that is decreasing along trajectories. Furthermore, by the *Fundamental theorem of dynamical system* (Theorem 2.27). Before introducing Morse-Smale, we first define several notions.

Given a hyperbolic fixed point $\boldsymbol{x}$ for a C^r function f, and a neighborhood U of $\boldsymbol{x}$, the *local stable set/manifold* for $\boldsymbol{x}$ in the neighbor U is defined as:

$$W^s_{loc}(\boldsymbol{x}, U, f) \triangleq \{\boldsymbol{y} \in U : \varphi(\boldsymbol{y}, t; f) \in U, \forall t > 0 \text{ and } d(\varphi(\boldsymbol{y}, t; f), \boldsymbol{x}) \to 0 \text{ as } t \to \infty\}$$
$$W^u_{loc}(\boldsymbol{x}, U, f) \triangleq \{\boldsymbol{y} \in U : \varphi(\boldsymbol{y}, t; f) \in U, \forall t < 0 \text{ and } d(\varphi(\boldsymbol{y}, t; f), \boldsymbol{x}) \to 0 \text{ as } t \to -\infty\}$$

Opposite to the notion of tangency, *transversality* is a geometric notion of the intersection of manifolds. Let $\boldsymbol{x} \in \mathcal{X}$ M and N are C^r manifolds in $\mathcal{X}$. M, N are said to be *transversal* at $\boldsymbol{x}$ if $\boldsymbol{x} \notin M \cap N$; or if $\boldsymbol{x} \in M \cap N$, $T_xM + T_xN = \mathbb{R}^d$ where T_xM and T_xN denote the tangent space of M and N respectively at point x. M and N are said to be *transversal* if they are transversal at every point $\boldsymbol{x} \in \mathcal{X}$.

Definition 2.31 (Morse-Smale flow)**.** Let $\varphi(\cdot, \cdot; f)$ be a flow on $\mathcal{X} = \mathbb{R}^d$. φ is called *Morse-Smale flow* if there are a constant collection of periodic orbits $P_1, \ldots, P_l$ such that

1. P_i is hyperbolic $i = 1, \ldots, l$
2. $\mathcal{CR}_f = \mathrm{Per}_f$
3. $W^U(P_i)$ and $W^S(P_j)$ are transversal for all $1 \le i, j \le l$.

Furthermore, if the Morse-Smale system does not have cycle, it is further called *gradient-like*.

Note that the gradient flow is a special case of gradient-like flow

Definition 2.32 (Gradient flow)**.** A flow $\varphi(\cdot, \cdot)$ on $\mathbb{R}^d$ is call *gradient flow* if there is a real valued function $V : \mathbb{R}^d \to \mathbb{R}$ such that

$$\frac{d}{dt}\varphi(x, t) = -\nabla V(x).$$

Proposition 2.33. *Let $V : \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{C}^2$ function such that each critical point is nondegenerate, i.e., at each point β where $\nabla V(x) = 0$, the matrix of second partial derivatives $\nabla^2 V(\beta)$ has nonzero determinant. Then the gradient flow with V has all the fixed points are hyperbolic and the chain recurrent set for the flow equals the set of fixed points.*

The above proposition shows the (non-degenerate) gradient flows are Morse-Smale system if and only if the stable and unstable manifolds are transverse.

Let $\{\beta_1, \ldots, \beta_m\} = \mathrm{Fix}_f$ be the set of fixed point of f, and W_i^s and W_i^u be the stable and unstable manifold associated to β_i. The Morse-Smale system has the following property.

Lemma 2.34. *Let f be a Morse-Smale system on $\mathcal{X}$. Let $\beta_i \succ \beta_j$ mean there is a trajectory not equal to β_i or β_j whose α-limit set is β_i and whose ω-limit set is β_j. Then $\succ$ satisfies:*

anti-reflexive *It is never true that $\beta_i \succ \beta_i$*

partial order *if $\beta_i \succ \beta_j$ and $\beta_j \succ \beta_k$ then $\beta_i \succ \beta_k$*

transversal *If $\beta_i \succ \beta_j$ then $\dim W_i^u \geq W_j^u$*

Morse-Smale systems share several properties with gradient fields: no complicated recurrent motion and existence of "potential function"— Morse function— that is decreasing along trajectories. Furthermore, by the Fundamental theorem of dynamical system 2.27 we have[1]

Corollary 2.35 (Theorem 12 in Akin [5])**.** *If $f \in \mathcal{C}^2$ is a Morse-Smale system then there exists a complete Lyapunov function $V : \mathcal{X} \to \mathbb{R}$ such that*

1. *$V \in \mathcal{C}^2$ is smooth.*
2. *$\frac{d}{dt} V(\varphi(x_0, t))|_{t=0} < 0$ for all non fixed points of f.*

We use Lie derivative to simply the notion, $\mathcal{L}_f V(x_0) \triangleq \frac{d}{dt} V(\varphi(x_0, t))|_{t=0}$.

2.4.5 One and two-dimensional flows

Given $a < b \in \mathbb{R}$, we use $[a, b]$ to represent the close set and (a, b) as the open set. We also use $D(c, d)$ to represent the open set $(c - d, c + d)$. We can characterized the stability of (2.1) relatively easily if the phase space is a compact space in $\mathbb{R}$,

Definition 2.36 (Fixed points)**.** Given a function $f : \mathbb{R} \to \mathbb{R}$, $y^* \in \mathbb{R}$ is a *fixed point* of f if and only if $f(y^*) = y^*$. Moreover, a fixed point y^* is

[1] For Morse Smale system, we have a stronger notion of potential function ξ-function [163]. However, it often requires the flow to be smooth

- a *attracting point* if and only if there exists $\epsilon > 0$ such that $f(x) < x$ if $x \in (y^*, y^* + \epsilon]$ and $f(x) > x$ if $x \in [y^* - \epsilon, y^*)$. Let S_f be the set of all attracting points.

- a *repelling point* if and only if there exists $\epsilon > 0$ such that $f(x) > x$ if $x \in (y^*, y^* + \epsilon]$ and $f(x) < x$ if $x \in [y^* - \epsilon, y^*)$. U_f is defined as the set of all repelling points.

- a *touch point* if and only if there exists $\epsilon > 0$ for all x such that $0 < |x - y^*| < \epsilon, f(x) > x$ or for all x s.t. $0 < |x - y^*| < d, f(x) < x$. Let T_f be the set of touch points.

We define Q_f be the set of fixed points $\{x : f(x) = x\}$.

If the system (2.1) is planar, the recurrent behavior is much simpler: it only have ω-limit. To state the theorem we need to introduce more terminology. A set is bounded if it is contained in some cycle $\{x \in \mathbb{R}^2 | |x - \alpha < C\}$ for some $\alpha \in \mathbb{R}^2$ and $C > 0$. A point $p \in \mathbb{R}^2$ is called an ω-limit point of the trajectory $\Gamma_{z_0} = \{z(t) | t \geq 0, z(0) = z_0\}$ of the system (2.5) if there is a sequence $t_n \to \infty$ such that $\lim_{n\to\infty} x(t_n) = p$.

Theorem 2.37 (Poincare-Bendixon Theorem [220]). *Let $z' = H(z)$ be a system of differential equations defined on E an open subset in $\mathbb{R}^2$ where H is differentiable. Suppose a forward orbit with initial condition z_0 $\Gamma_{z_0} = \{z(t) | t \geq 0, z(0) = z_0\}$ is bounded. Then either*

- *$\omega(z_0)$ contains a fixed point*

- *$\omega(z_0)$ is a periodic orbit*

The following theorem gives us a sufficient condition for nonexistence of periodic orbit

Theorem 2.38 (Bendixson's Criteria [220]). *Let H be differentiable in E where E is a simply connected region in $\mathbb{R}^2$. If the divergence of the vector field H is not identically zero and does not change sign in E then $z'H(x)$ has no closed periodic orbit lying entirely in E.*

Note that the theorem only holds for two dimensions system and fails in general.

A flow with f is called a ***gradient flow*** if there exists a real value function $V : \mathbb{R}^d \to \mathbb{R}$ such that $f = -\nabla V$. However, there is a more general family of dynamics called ***gradient-like flow*** that contains the gradient flows with a mild restriction which is discussed at Proposition 2.33. Here we gives a sufficient condition for gradient-like flow on two dimensional manifolds.

Proposition 2.39. *Let $\mathcal{X} = \mathbb{R}^2$. A vector field with $f \in \mathcal{C}^r(\mathbb{R}^2, \mathbb{R}^2)$ is a gradient-like flow if:*

1. *f has a finite number of fixed points which are all hyperbolic;*
2. *there are no saddle-connections that is an orbit whose α- and ω -limits are saddle points; and*
3. *each orbit has a unique fixed point as its α-limit and has a unique fixed point as its ω-limit.*

We further call the function f gradient-like.

2.5 Models of Social Networks

2.5.1 Erdös-Rényi Random Graphs

Here we present the definition of Erdös-Rényi random graphs and show several properties of them that we need.

Definition 2.40 (Erdös-Rényi Random Graph)**.** $G_{n,p}$ is a random undirected graph on node set $V = [n]^2$ where each pair of nodes is independently connected with a fixed probability p. We further use $\mathcal{G}$ to denote this random object.

Let A_G be the *adjacency matrix of* G, so $(A_G)_{i,j} = 1$ if $v_i \sim v_j$ and 0 otherwise, and $\bar{A} = \mathbb{E}_{\mathcal{G}}[A_G]$, so $\bar{A}_{i,j} = p$ if $i \neq j$ and 0 otherwise. Let $\deg(v)$ be the degree of node v.

Definition 2.41. The *weighted adjacency matrix of undirected graph* G is defined by

$$M_G(i,j) = \begin{cases} \frac{1}{\sqrt{deg(v_i)deg(v_j)}} & \text{if } (A_G)_{i,j} = 1; \\ 0 & \text{otherwise .} \end{cases}$$

Definition 2.42 (Expansiveness [56])**.** For $\lambda \in [0,1]$, we say that a undirected graph G is a λ-expander if if $\lambda_k(M_G) \leq \lambda$ for all $k > 1$ where $\lambda_k(M_G)$ is the k-th largest eigenvalue.

[2] $[n] = \{1, 2, \ldots, n\}$

Theorem 2.43 (Spectral profile of $G_{n,p}$ [55]). *For $G_{n,p}$, we denote I as identity matrix and J as the matrix that has ones. If $G_{n,p}$ has $p = \omega(\frac{\log n}{n})$, then with probability at least $1 - 1/n$, for all k*

$$|\lambda_k(M_G) - \lambda_k(\bar{M})| = O\left(\sqrt{\log n/(np)}\right)$$

where $(\bar{M})_{i,j} = \frac{1}{n-1}$ if $i \neq j$ and $(\bar{M})_{i,i} = 0$.

Because the spectrum of $\bar{M}$ is $\{1, -1/(n-1)\}$ where $-1/(n-1)$ has multiplicity $n-1$, we can have the following corollary

Corollary 2.44. *If $p = \omega(\frac{\log n}{n})$, the $G \sim \mathcal{G}$ is $O\left(\sqrt{\frac{\log n}{np}}\right)$-expander with probability $1 - O(1/n)$,*

Let $e(S,T)$ denote the number of edges between S and T (double counting edges from $S \cap T$ to itself), and let $vol(S)$ count the number of edges adjacent to S. The following lemma relates the number of edges between two sets of nodes in an expander to their expected number in a random graph.

Lemma 2.45 (Irregular mixing lemma [54]). *If G is a λ-expander, then for any two subsets $S, T \subseteq V$:*

$$\left| e(S,T) - \frac{vol(S)vol(T)}{vol(G)} \right| \leq \lambda \sqrt{vol(S)vol(T)}$$

Finally, let $E(\delta_d; v)$ denote the event that the degree of some fixed node v is between $(1-\delta_d)np$ and $(1+\delta_d)np$ and let $E(\delta_d) = \cap_{v \in V} E(\delta_d; v)$ a *nearly uniform degree event*. By applying Theorem 2.17 we yields the following lemma.

Lemma 2.46 (Uniform degree). *For any $v \in V$, if $G \sim \mathcal{G}$*

$$\Pr[E(\delta_d; v)] \leq 2\exp\left(-\delta_d^2 np/3\right) \tag{2.9}$$

Furthermore, by union bound

$$\Pr[E(\delta_d)] \leq 2n\exp\left(-\delta_d^2 np/3\right) \tag{2.10}$$

2.5.2 Configuration Models with Power-law Degree Distributions

We use the *configuration model* introduced by Bollobás and McKay [41] to define a distributions over multigraphs.

Definition 2.47 (Configuration Model)**.** Let $\mathbf{d} = (d_1, \ldots, d_n)$ be a decreasing degree sequence where the sum of the terms is even. Define $V = [n]$ (Here we use integers $\{1, 2, \ldots, n\}$ to denote the vertices, and call nodes with lower indexes "earlier". Because the degrees decrease, earlier nodes have higher degrees). Let m be such that $2m = \sum_i d_i$. To create the m (multi-)edges, we first assign each node i d_i *stubs*. Next we choose a perfect matching of the stubs uniformly at random and for each pair of matched stubs construct an edge connecting the corresponding nodes.

We use CM($\mathbf{d}$) to denote the *Configuration Model* with respect to the degree distribution $\mathbf{d}$.

For any decreasing degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$ where the sum of the terms is even, we define the empirical distribution function of the degree distribution,

$$F_{\mathbf{d}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[d_i \leq x] \; \forall x \in [1, \infty)$$

which is the fraction of nodes that have degree less than x. Moreover, we define $N_{\mathbf{d}}(x) = n(1 - F_{\mathbf{d}}(x))$ be the number of nodes with degree at least x, $S_{\mathbf{d}}(x)$ be the number of stubs from nodes with degree at least x, and $s_{\mathbf{d}}(x)$ be the number of stubs from nodes with index less than x. We will omit the index $\mathbf{d}$ when there is no ambiguity.

Definition 2.48 (Power-law degree distributions)**.** Adopting the notation of Van Der Hofstad [230], we say a series $\mathbf{d}$ has power-law distribution with exponent α if there exists $0 < C_1 < C_2$ and $0 < x_0$ such that 1) $F_{\mathbf{d}}(x) = 0$ for $x < x_0$; 2) $F_{\mathbf{d}} = 1$ for $x > d_1 = n^{2/(\alpha+1)}$, and 3) for all $x_0 \leq x \leq d_1$ then

$$C_1 x^{-\alpha+1} \leq 1 - F_{\mathbf{d}}(x) \leq C_2 x^{-\alpha+1}$$

Let $\mathbf{d}$ have power-law distribution of *power law with exponent* α then it is easy to check that:

Lemma 2.49. *If* $\mathbf{d} = (d_1, \ldots, d_n)$ *is a power-law distribution with exponent* α *then*

1. $N(x) = \Theta(nx^{-\alpha+1})$
2. $S(x) = \Theta(nx^{-\alpha+2})$.
3. $d(i) = \Theta\left(\left(\frac{n}{i}\right)^{1/(\alpha-1)}\right)$
4. $s(i) = \Theta(n^{1/(\alpha-1)} i^{\frac{\alpha-2}{\alpha-1}})$

2.5.3 Small-World Models

Definition 2.50 (Kleinberg's small world model [133])**.** Given $p, q, n \in \mathbb{N}$ and positive real number $\gamma > 1$, *Kleinberg's small world model* with parameter (p, q, γ, n) defined as follows: Nodes are on a $n \times n$ planar grid.[3] Each node u connects to nodes within grid Manhattan distance $\lceil q \rceil$, and these edges are referred to as *strong ties*. In addition, each node generates p random outgoing edges (without replacement), termed *weak ties*. The probability that node u connects to node v via a random edge is $1/\lambda_\gamma d(u,v)^\gamma$, in which $d(u,v)$ is the Manhattan distance of u, v and $\lambda_\gamma = \sum_v d(u,v)^{-\gamma}$ is a normalization factor.

2.5.4 Preferential attachment models

Definition 2.51 (preferential attachment models)**.** Given positive integers k and n with $k < n$, the *Preferential Attachment Model*, $\mathrm{PA}_k(n)$ is a distribution over directed graphs with n nodes. It generates $G_n = (V_n, E_n)$ as follows:

1. Set $V_n = [n] = \{1, 2, \ldots, n\}$ and $E = \emptyset$, and we call node s is earlier than node t if $s < t$.
2. Let E_{k+1} consists of the edge between the first $k+1$ nodes, $\{1, \ldots, k+1\}$, which forms a $(k+1)$-clique. Set $G_{k+1} := ([k+1], E_{k+1})$.
3. At each subsequent step $t+1 > k+1$, given $G_t = ([t], E_t)$ the node $t+1$, chooses $w_1, w_2, \ldots, w_k$ vertices independently according to preferential attachment rule.[4] Formally, for all l between 1 and k, the node $t+1$ sets $w_l = s$ with probability,

$$\Pr[w_l = s] = \frac{\deg(s;t)}{2|E_t|}$$

   where $\deg(s;t)$ is the in-degree and out-degree of node s and $|E_t|$ is the total number of edges before $t+1$ adds edges. Then node $t+1$ adds k new directed edges $(t+1, w_1), \ldots, (t+1, w_k)$. We call the resulting graph $G_{t+1} = ([t+1], E_{t+1})$.

We use $\bar{G}_n$ to denote the undirected graph of G_n by changing all the directed edges into undirected.

[3] In order to eliminate the boundary effect, we wrap up the grid into a torus – i.e., the top boundary is identified with the bottom boundary and the left boundary is identified with the right boundary.

[4] This may cause multiple edges.

Note that by the definition the marginal distribution of G_t has the same distribution as $\mathrm{PA}_k(t)$ which is an important property for our theoretical analysis.

Here we prove a lower bound of degree for preferential attachment models. First note that we can approximate the expected degree $\mathbb{E}[\deg(s;t)]$ easily: Because $\mathbb{E}[\deg(s;t+1) \mid \deg(s;t)] = \deg(s;t) + \frac{k \deg(s;t)}{2kt}$, we can use the tower property of expectation and have

$$\mathbb{E}[\deg(s;n)] = \deg(s;s) \prod_{t>s}^{n} \left(1 + \frac{1}{2t}\right) \approx \sqrt{\frac{n}{s}} k.$$

With this estimation, we can use Markov inequality to have a decent upper bound for the probability that the degree of node s being small.

We further show the following stronger lemma, and Lemma 5.8 is a corollary. Note that if s and d are constants and n is large, the following lemma show the probability $\deg(s)$ is smaller than d is $\tilde{O}\left(\frac{1}{n^{k/2}}\right)$ which is better then the bounded by Markov inequality.

Lemma 2.52. *Given positive integers k, s, d and n such that $d \geq k$ and n large enough $k+1 < \frac{1}{e^2}(n+1)$,*

$$\Pr[\deg(s) \leq d \mid \mathrm{PA}_k(s)] \leq \binom{d-1}{k-1} \left(\frac{s}{n}\right)^{k/2} \left(\ln \frac{kn+1}{ks+1}\right)^d.$$

Note that the randomness comes from $\mathrm{PA}_m(n)$ conditioned on $\mathrm{PA}_m(k)$.

The idea of this proof is based on an estimation of $\Pr[\deg(k) \leq d \mid \mathrm{PA}_m(k)]$.

Proof. **A special case when $k = 1$:** Firstly we consider the special case when $k = 1$ (i.e., one edge is added for each node). Let $P_s(d)$ be the probability that node s has d neighbors. Let $S_d = \{N_d : N_d \subset (s, n], \text{ and } |N_d| = d-1\}$ is the set of $d-1$ nodes after s. Given a $d-1$ set $N_d = (w_2, \ldots, w_d)$ where $s < w_2 < \ldots < w_d$, we call $P_s(N_d)$ be the probability that node s has degree d, and $d-1$ of them are in N_d and the other is before s. Then we have

$$\Pr[\deg(s) \leq d \mid \mathrm{PA}_1(s)] = \sum_{i \leq d} P_s(i), \text{ and } P_s(d) = \sum_{N_d \in S_d} P_s(N_d). \tag{2.11}$$

Thus, it is sufficient to upper bound $P_s(N_d)$. Because node s has degree at least 1 for all node t after s, the probability that $t \notin N_d$ is not a neighbor of s is upper bounded

by $1 - \frac{1}{2(t-1)}$. Hence,

$$P_s(N_d) \leq \prod_{s<t\leq n: t\notin N_d} \left(1 - \frac{1}{2(t-1)}\right) \cdot \prod_{j=2}^{d} \frac{j-1}{2(w_j-1)}.$$

By some basic computation, we have a more concrete upper bound.

$$\begin{aligned}
P_s(N_d) &\leq \prod_{s<t\leq n} \left(1 - \frac{1}{2(t-1)}\right) \cdot \prod_{j=2}^{d} \frac{j-1}{2(w_j-1)} \left(1 - \frac{1}{2(w_j-1)}\right)^{-1} \\
&\leq \prod_{s<t\leq n} \left(1 - \frac{1}{2(t-1)}\right) \cdot (d-1)! \prod_{j=2}^{d} \frac{1}{w_j - 1}. \quad \text{(because } 2\left(1 - \tfrac{1}{2(w_j-1)}\right) \geq 1\text{)} \\
&\leq \exp\left(\sum_{t>s}^{n} \frac{-1}{2(t-1)}\right) \cdot (d-1)! \prod_{j=2}^{d} \frac{1}{w_i - 1} \qquad (1 + x \leq e^x) \\
&\leq \sqrt{\frac{s}{n}} \cdot (d-1)! \prod_{j=2}^{d} \frac{1}{w_j - 1}.
\end{aligned}$$

Apply this upper bound to Equation (2.11) we have,

$$P_s(d) \leq \sum_{N_d \in S_d} \sqrt{\frac{k}{n}} \cdot (d-1)! \prod_{j=2}^{d} \frac{1}{w_j - 1} = \sqrt{\frac{s}{n}} \cdot (d-1)! \sum_{N_d \in S_d} \prod_{j=2}^{d} \frac{1}{w_j - 1}. \tag{2.12}$$

For the second term, we can exchange the order of produce and summation by relaxing the condition that $w_1, \ldots, w_d$ need to be distinct.

$$(d-1)! \sum_{N_d \in S_d} \prod_{i=2}^{d} \frac{1}{w_i - 1} \leq \prod_{i=2}^{d} \sum_{w_i > s}^{n} \frac{1}{w_i - 1} = \left(\sum_{w_i > s}^{n} \frac{1}{w_i - 1}\right)^{d-1} \leq \left(\ln \frac{n+1}{s+1}\right)^{d-1}.$$

Combining this inequality to (2.12), we have

$$P_s(d) \leq \sqrt{\frac{s}{n}} \left(\ln \frac{n+1}{s+1}\right)^{d-1}.$$

Therefore we finish our bound by (2.11) and the above inequality:

$$\Pr[\deg(s) \leq d \mid \mathrm{PA}_1(s)] \leq \sum_{i=1}^{d} \sqrt{\frac{s}{n}} \left(\ln \frac{n+1}{s+1}\right)^{i-1} \leq \sqrt{\frac{s}{n}} \frac{\left(\ln \frac{n+1}{s+1}\right)^{d} - 1}{\ln \frac{n+1}{s+1} - 1} \leq \sqrt{\frac{s}{n}} \left(\ln \frac{n+1}{s+1}\right)^{d}.$$

The last inequality holds because $\frac{n+1}{s+1} > e^2$.

General case k of $\mathrm{PA}_k(n)$: We now handle the more general case of m. By Definition 2.51, each new vertex with index t forms k edges independently to nodes with indices smaller than t. By labeling those k edges with indices $k(t-1)+1, k(t-1)+2, \ldots, kt$ respectively, we can label every edge with a unique index. Therefore edge e is incident to node $\lceil \frac{e}{k} \rceil$. Let $P_s(d)$ be the probability that node s has d neighbors, and $S_d = \{N_d : N_d \subset (ks, kn], \text{ and } |N_d| = d-k\}$ is the set of $d-k$ edges after s. Given a $d-k$ set $N_d = (e_{k+1}, \ldots, e_d)$ where $ks < e_{k+1} < \ldots < e_d$, we call $P_s(N_d)$ be the probability that node s has degree d, and $d-k$ of them are in N_d and the others are before s.

We can reuse (2.11), and construct an upper bound for $P_s(N_d)$

$$\begin{aligned} P_s(N_d) \leq & \prod_{t>ks}^{kn} \left(1 - \frac{k}{2k(\lceil \frac{t}{k} \rceil - 1)}\right) \cdot \prod_{j>k}^{d} \frac{j-1}{2k(\lceil \frac{w_j}{k} \rceil - 1)} \left(1 - \frac{k}{2k(\lceil \frac{w_j}{k} \rceil - 1)}\right)^{-1} \\ \leq & \prod_{\ell>s}^{n} \left(1 - \frac{1}{2(\ell-1)}\right)^{k} \cdot \frac{(d-1)!}{(k-1)!} \prod_{j>k}^{d} \frac{1}{w_j - 1} \\ \leq & \left(\frac{s}{n}\right)^{k/2} \cdot \frac{(d-1)!}{(k-1)!} \prod_{j>k}^{d} \frac{1}{w_j - 1} \end{aligned}$$

Apply this upper bound to Equation (2.11), and we have,

$$P_s(d) \leq \sum_{N_d \in S_d} \left(\frac{s}{n}\right)^{k/2} \cdot \frac{(d-1)!}{(k-1)!} \prod_{j>k}^{d} \frac{1}{w_j - 1} = \left(\frac{s}{n}\right)^{k/2} \sum_{N_d \in S_d} \frac{(d-1)!}{(k-1)!} \prod_{j>k}^{d} \frac{1}{w_j - 1}.$$

Changing the order of summation and product in the second term, we yield:

$$\frac{(d-1)!}{(k-1)!} \sum_{N_d \in S_d} \prod_{j>k}^{d} \frac{1}{w_j - 1} \leq \binom{d-1}{k-1} \prod_{j>k}^{d} \sum_{w_j > ks}^{kn} \frac{1}{w_j - 1} \leq \binom{d-1}{k-1} \left(\ln \frac{kn+1}{ks+1}\right)^{d-k},$$

and we have

$$P_s(d) \leq \left(\frac{s}{n}\right)^{k/2} \binom{d-1}{k-1} \left(\ln \frac{kn+1}{ks+1}\right)^{d}$$

Therefore we can bound the probability that the degree of node k is smaller than d, $\Pr[\deg(s) \leq d \mid \mathrm{PA}_k(s)]$:

$$\left(\frac{s}{n}\right)^{k/2} \sum_{i=k}^{d} \binom{i-1}{k-1} \left(\ln \frac{kn+1}{ks+1}\right)^{i} \leq \left(\frac{s}{n}\right)^{k/2} \binom{d-1}{k-1} \left(\ln \frac{kn+1}{ks+1}\right)^{d}.$$

□

2.5.5 Stochastic Block Models

Definition 2.53 (Stochastic Block Models [113])**.** Fixing n vertices in V, let $\mathcal{B} = (B_1, \ldots, B_\ell)$ be a partition of the set V and $L = [\ell]$ be the set of labels. Let $W_n : L \times L \mapsto [0, 1]$ be a bounded non-negative function with $W_n(i, j) = W_n(j, i)$ for all $i, j \in L$. Given $v \in V$, we call i_v the label of node v such that $v \in B_{i_v}$.

A **stochastic block model** $\mathcal{G}(n, \mathcal{B}, W_n)$ is a distribution over undirected graph $G = (V, E)$ over nodes in V defined as follow: For each pair $u, v \in V$, independently add the edge (u, v) to E with probability $W_n(i_u, i_v)$.

Intuitively, the each node has a label/attribute in L, and the community structure is represented as W_n where $W_n(i, j)$ encodes the similarity between labels i and j. Notice that we allow the weight function W_n depends on the size of the graph n to allow asymptotically different density of connection.

CHAPTER 3

General Kleinberg's Small World Models

Kleinberg's small world model [133] simulates social networks with both strong and weak ties. In his original paper, Kleinberg shows how the distribution of weak-ties, parameterized by γ, influences the efficacy of myopic routing on the network. Recent work on social influence by k-complex contagion models shows that the distribution of weak-ties also impacts the spreading rate in a crucial manner on Kleinberg's small world model [99]. In both cases the parameter of $\gamma = 2$ proves special: when γ is anything but 2 the properties no longer hold.

In this chapter, we propose a natural generalization of Kleinberg's small world model to allow node heterogeneity: instead of a single global parameter γ, each node has a personalized parameter γ chosen independently from a distribution $\mathcal{D}$. In contrast to the original model, we show that this model enables myopic routing and k-complex contagions on a large range of the parameter space, improving the robustness of the model. Moreover, we show that our generalization is supported by real-world data. Analysis of four different social networks shows that the nodes do show heterogeneity in terms of the variance of the lengths of edges incident to the same node.

3.1 Introduction

In Milgram's "Small World" experiments [164, 227], he gave envelops to random residents of Wichita, Kansas and Omaha, Nebraska, and asked them to forward the envelopes to a personal contact so that they might eventually reach a specific banker in Massachusetts. The success of this experiment (which has since been observed in numerous other contexts – see related work) motivated Kleinberg's small work model which studies why such local decisions work [133]. This ingenious model shows not only that short paths between arbitrary nodes exist (this so-called "small world"

phenomena was already embedded into several fundamental models [234, 179, 40]), but also that these short paths can be easily discovered by myopic routing (i.e., using purely local knowledge).

Kleinberg's small world model considers an underlying metric space capturing the diversity of the population in various social attributes. Social ties are classified into two categories: strong ties that connect an individual to those similar in the social attribute space, and weak ties that may connect individuals far away. Kleinberg's model considers one parameter γ in determining how the weak ties are placed. Each node p takes a weak tie edge to a node q with probability proportional to $1/|pq|^\gamma$ where $|pq|$ denotes the distance between p and q in the social space. Thus at $\gamma = 0$ the weak ties are uniformly randomly distributed, and as γ increases shorter connections are increasingly favored.

However, in this model when the nodes are placed in a 2-dimensional grid the navigability only holds for a particular parameter choice: $\gamma = 2$. At this "sweet spot," a message can be delivered to the destination in $O(\log^2 n)$ hops, by hopping to the neighbor closest to the destination in the Euclidean metric. For *any* constant $\gamma \neq 2$, myopic routing, or, in general, any deterministic routing algorithm using only local information, provably fails to quickly deliver the message. Intuitively why $\gamma = 2$ is crucial, because at this sweet spot each weak tie edge uniformly at random lands in one of the annuli with inner radius 2^i and outer radius 2^{i+1}, for all i. Therefore, no matter where the destination is, there is a neighbor with probability roughly $1/\log n$ such that taking this neighbor reduces the Euclidean distance to the destination by half. If $\gamma < 2$, it turns out that the weak tie edges are too random and myopic routing loses its sense of direction. If $\gamma > 2$, the weak ties are simply too short and any path to the destination discoverable from local information necessarily takes many hops.

Other good properties also hold at special ranges of the parameter γ. In recent work on understanding complex social influence, it was shown how the distribution of weak-ties impacts the spreading behavior of k-complex contagions, in which a node becomes infected if at least k neighbors are infected [99, 83]. Again it was shown that when $\gamma = 2$, for any constant k, the k-complex contagion spreads in a polylogarithmic number of rounds to the entire network while when $\gamma \neq 2$ complex contagions necessarily require a polynomial number of rounds. The analysis here connects to the intuition presented earlier for myopic routing. The sweet spot $\gamma = 2$ substantially speeds up the spreading of the contagions.

While the existence of the sweet spot is both insightful and elegant, it has raised new questions for modeling practical networks. The model feels fragile if the good

properties only hold at a single parameter value and stop holding even with slight deviation. As put by Jackson [118]: "It is unlikely that societies just happen to hit the right balance. More likely there is something missing from the models, and it is clear the the network-formation process underlying many social networks is much more complex than in these models." If Jackson is correct, then a theoretical model that more robustly justifies the empirical observations of Milgram and those who followed is needed.

Our Results. In this work, we generalize Kleinberg's small world model by considering a personalized, possibly heterogeneous γ_u for each node u in the network. In particular, each node u chooses its parameter $\gamma_u \in [0, \infty)$ i.i.d from a distribution $\mathcal{D}$. The weak tie edges issued by u will be placed on node v with probability proportional to $1/|uv|^{\gamma_u}$, where $|uv|$ denotes the distance between u and v in some underlying metric.

This model is motivated by both intuition and observations in real world data sets. It is natural to believe that some people have weak ties that are more/less dispersed (geographically or otherwise) that others. We also provide empirical evidence for node heterogeneity using real world social network data. Given a network, we can embed it in Euclidean space using spectral methods and examine the length of the edges attached to each node. We find that the empirical variance of the lengths of edges incident on the same vertex is substantially less than when the edge lengths are randomly permuted — suggesting that lengths of edges incident on the same vertex are indeed more correlated. See Section 3.8 for details.

In this paper the main technical results we report is that both myopic routing and k-complex contagions operate quickly in the new model as long as the distribution $\mathcal{D}$ for the personalized γ has *non-negligible mass* around 2. Thus our model provides a robust justification for the observed properties of both myopic routing and k-complex contagions. Moreover it does this by only slightly tweaking Kleinberg's original model.

In particular, we can show that even if there is just $\Omega(\epsilon^\alpha)$ mass in the interval $[2-\epsilon, 2+\epsilon]$ of the distribution $\mathcal{D}$, where $\alpha > 0$ is *any* constant, then myopic routing and k-complex contagions (for any k) still only take polylogarithmic time! For example, it is enough that $\mathcal{D}$ be uniform on the interval $[a, b]$ for any $0 \leq a \leq 2 \leq b$. Note that in such a case, no particular γ_u will be exactly 2 (with probability 1). However, it turns out that enough of the γ_u are close enough to 2, which still enables these social processes.

We also show lower bounds. For myopic routing we show that if for some ϵ,

there is no mass in $[2-\epsilon, 2+\epsilon]$, then the typical myopic routing time is polynomial. This is not obvious, as there can be a distribution $\mathcal{D}$ that allows weak ties that are short — connecting nodes nearby, and weak tie that are long – connecting nodes far away. Recall that in the original Kleinberg proof it was shown that *short ties only*, or *long ties only*, are not enough to enable myopic routing but it did not exclude the possibility when *both long and short ties* exist simultaneously. We show that in fact the combination of these weak tie edges are still not enough for enabling efficient myopic routing. In particular, there is a range of distances when none of the two types of ties are helpful, which forces the greedy routing to take a long time.

For complex contagions, our first lower bound shows that if for some $\epsilon > 0$, there is no mass in $[2-\epsilon, 2+\epsilon]$, then there is some k such that k-complex contagions require a polynomial time to spread. Again we must show that the synergy between short and long weak ties cannot enable complex contagions to quickly spread.

The above results for complex contagion apply for *any* k. We also study what happens for a particular k. Here we show that for each k there is an interval $[2, \beta_k)$ where $\beta_k = \frac{2(k+1)}{k}$ such that when $\mathcal{D}$ has constant support on $[2, \beta_k)$, k-complex contagions spread in polynomial time, but when, for any $\epsilon > 0$, $\mathcal{D}$ has no support on $[2-\epsilon, \beta_k+\epsilon]$, then k-complex contagions requires polynomial time to spread with high probability.

3.2 Related Work

Small World Graphs The small world property—that there exists short paths between two random members of a network–appears in many real world complex networks in vastly different contexts ranging from film collaboration networks and neural networks [234] to email networks [74], food webs [238] and protein interaction networks [120].

It has been discovered in a number of settings that random edges introduced to a graph can dramatically reduce the network diameter, creating a small world graph. This observation was made in the Watts-Strogatz model [234] (when edges are rewired to a random destination) as well as for regular random graphs [40] (a graph in which all nodes have the same constant degree and edges are uniformly randomly placed). Kleinberg's small world model can be considered as an extension to such models. In particular, the Newmann-Watts model [179] (a variant of the Watts-Strogatz model in which random edges are added in *addition* to existing edges) is a special case of Kleinberg's model for choosing $\gamma = 0$ — i.e., the weak ties are uniformly randomly

added.

Navigability Milgram's "Small World" experiments [164, 227] illustrated not only the small world property—that short paths exist—but, in fact, showed a stronger property—that such paths can be efficiently found using only local information—called navigability. A short path was discovered through a *local* algorithm with the participants forwarding to a friend who *they believed* to be more likely to know the target. Although forwarding decision-making was not systematically recorded, geographical proximity was found to be an important forwarding criterion in some cases. Other criteria such as profession and popularity may have been used as well. A later study using email-chains [74] confirms this as well, finding that at least half of the choices were due to either geographical proximity of the acquaintance to the target or occupational similarity.

Besides the Kleinberg's small world model, several other models also considered using metric distances in modeling social ties. For example, Kumar et al. [140] extended the Kleinberg's model to include the underlying metrics with low-doubling dimension. This model also requires a specific distribution of the weak ties.

Another line of work diverges from distance function defined over some low-dimensional space, but instead defines a distance function based on some hierarchical structure. For example, Watts et al. [235] considered a hierarchical professional organization of individuals and a homophilous network with ties added between two nodes closer in the hierarchy with a higher probability. If each node has a fixed probability of dropping the message, they show a greedy routing algorithm sending packages to the neighbor most similar to the target (called homophily-based routing) successfully delivers a fraction of the messages before they are dropped. Kleinberg [134] also confirmed similar results on a hierarchical network, in which the nodes are represented as leaf nodes of a hierarchical organization structure and random edges are added to the leaves with probability dependent on their tree distance. When each node has polylogarithmic out-degree, greedy routing based on the tree distance arrives at the destination in $O(\log n)$ hops. While the aforementioned models also successfully create a more robust network model for myopic routing, in doing so they abandoned the spatial structure of Kleinberg's small world model. While certain structures can be modeled well as a hierarchy, others are much more natural as a continuum, as in Kleinberg's model—e.g. distances, wealth, political ideology, and education.

Boguna et al. [38] proposed a model that assumes a social metric space and the power law degree distribution. They considered nodes on a ring and assigned target

degrees from a power law distribution. An edge is then placed between two nodes with a probability positively dependent on their distance on the ring and negatively dependent on their degrees. They investigated greedy routing with the distances on the ring as a means of navigating in the network. Papadopoulos et al. [186] considered using a hyperbolic plane as the hidden social space. Nodes are uniformly distributed in a radius R disk in a hyperbolic plane with edges placed in pairs with distance smaller than r. They show that such a graph has power law degree distribution and that greedy routing with hyperbolic distance has a high success rate.

3.3 Definition of General Kleinberg's Small World Model

Similar to Kleinberg's small world model defined in Definition 2.50, for *Heterogeneous Kleinberg's small world* $\mathrm{HetK}_{p,q,\mathcal{D}}(n)$, we define p, q, n as in the original model, but, instead of one global γ, each node u independently chooses its personalized parameter γ_u from the distribution $\mathcal{D}$ on $[0, \infty)$ with probability density function $f_{\mathcal{D}}$ and cumulative distribution function $F_{\mathcal{D}}$. Let $M_{\mathcal{D}}(\epsilon) = F_{\mathcal{D}}(2+\epsilon) - F_{\mathcal{D}}(2-\epsilon)$ measure the "mass" of $\mathcal{D}$ around 2.

Lemma 3.1 (The Normalization Factor)**.** *For $n \geq 4$, the normalization factor $\lambda_\gamma = \sum_v d(u,v)^{-\gamma}$ can be bounded above as follows:*

$$\begin{aligned}
&\lambda_\gamma \leq 4\left(1 + \tfrac{1}{\gamma-2}\right) && \text{if } \gamma > 3 \\
&\lambda_\gamma \leq \tfrac{8}{\gamma-2} && \text{if } 2 + \tfrac{1}{\ln n} < \gamma \leq 3 \\
&\lambda_\gamma \leq 8 \ln n && \text{if } 2 - \tfrac{1}{\ln n} \leq \gamma \leq 2 + \tfrac{1}{\ln n} \\
&\lambda_\gamma \leq \tfrac{8}{2-\gamma} n^{2-\gamma} && \text{if } 0 \leq \gamma < 2 - \tfrac{1}{\ln n}
\end{aligned}$$

For the lower bound,

$$\begin{aligned}
\tfrac{2}{\gamma-2} &\leq \lambda_\gamma && \text{if } 2 + \tfrac{1}{\ln(n/2)} < \gamma \\
2\ln(n/2) &\leq \lambda_\gamma && \text{if } 2 - \tfrac{1}{\ln(n/2)} \leq \gamma \leq 2 + \tfrac{1}{\ln(n/2)} \\
\tfrac{2}{2-\gamma}(n/2)^{2-\gamma} &\leq \lambda_\gamma && \text{if } 0 \leq \gamma < 2 - \tfrac{1}{\ln(n/2)}
\end{aligned}$$

3.4 Myopic Routing Upper Bounds

In this section, we prove the following theorem about myopic routing (Section 2.1.3).

Theorem 3.2 (Myopic Routing Upper Bounds)**.** *Given a* $\text{HetK}_{p,q,\mathcal{D}}(n)$ *with constant* $p, q \geq 1$ *and distribution* $\mathcal{D}$*. If there exists some constants* $\epsilon_0 > 0$*,* $\alpha \geq 1$ *and* $K > 0$ *such that* $\forall \epsilon < \epsilon_0$*,* $M_{\mathcal{D}}(\epsilon) \geq K\epsilon^{\alpha}$*, the expected delivery time of the myopic greedy algorithm is at most* $O(\log^{2+\alpha} n)$*.*

The above theorem proves fast myopic routing over a large class of Heterogeneous Kleinberg's Small world models. The only distributions that this theorem fails to apply to are distributions with negligible mass near 2. In particular, if $\mathcal{D}$ is uniform over *any* finite interval containing 2, then myopic routing will take time at most $O(\log^3 n)$, and as long as the mass near 2 is non-trivial (i.e., lower bounded by the inverse of some fixed polynomial), then delivery only takes poly-log time.

Remark 3.3. Note that if the random variable associated with $\mathcal{D}$ is a constant random variable that takes a constant value 2, the $\text{HetK}_{p,q,\mathcal{D}}(n)$ degenerates to the original Kleinberg's model with $\gamma = 2$, and the Theorem 3.2 is tight which yields the same $O(\log^2 n)$ upper bound on delivery time on myopic greedy routing algorithm.

The proof of Theorem 3.2 follows the general outline of the proof in Kleinberg's original paper: measure the progress of process $\mathbb{A} = (x_i)_{i \geq 0}$ in terms of phases which will be defined later and show the following: (1) monotone property of the process, (2) upper bound the total number of phase, (3) lower bound the probability of finishing each phase.

3.5 Myopic Routing Lower Bounds

In this section we prove a lower bound for any decentralized algorithms on the Heterogeneous Kleinberg Small World $\text{HetK}_{p,q,\mathcal{D}}(n)$ in the following theorem:

Theorem 3.4. *Given a Heterogeneous Kleinberg's Small World network* $\text{HetK}_{p,q,\mathcal{D}}(n)$ *with constant parameters* p, q *and probabilistic density function* $f_{\mathcal{D}}$ *for the distribution* $\mathcal{D}$ *on the personalized* γ_u *for each node* u*, if there exists a constant* $\epsilon_0 > 0$ *such that* $F(2+\epsilon_0) - F(2-\epsilon_0) = 0$*, where* F *is the cumulative density function of* $\mathcal{D}$*, then the expected routing time for all decentralized algorithms is* $\Omega(n^{\xi})$ *where* $\xi = \frac{\epsilon_0}{3(3+\epsilon_0)}$*.*

In the original Kleinberg's model [133], all nodes use the same γ parameter. When γ is greater than 2 the weak ties are too short in expectation such that it would need a polynomial number of hops to reach a far away destination. When γ is smaller than 2 the edges are too random to be useful for nearby destinations. But in a heterogeneous model, the nodes may have different γ values. The nodes with $\gamma_u > 2$

have *concentrated edges* while those with $\gamma_u < 2$ have *diffuse edges*. A network with only concentrated edges or only diffuse edges cannot support polylogarithmic myopic routing. But it is unclear whether the combination of them, as in the heterogeneous model, can lead to polylogarithmic delivery time.Theorem 3.4 states that this is not true. We show this by considering a scope where neither type of edges is helpful.

Proof. Fix a decentralized algorithm $\mathbb{A}$ and consider the source s and destination t chosen uniformly at random in the network. With probability at least $1/2$, they are separated by a lattice distance of at least $n/4$. The decentralized algorithm $\mathbb{A}$ has the knowledge of the grid coordinates of the current node, all neighbors, and the destination.

We define B_j to be the disk with center t and radius 2^j, i.e., $B_j = \{x | d(x,t) \leq 2^j\}$ and we denote by $\neg B_j$ to be the nodes outside B_j. The goal of routing is to move from larger disks (with high values of j) to smaller disks (with smaller values of j). It turns out that diffuse weak ($\gamma < 2$) ties are mainly useful when $j > \frac{1}{2}\log n$ and concentrated weak ties ($\gamma > 2$) are mainly useful when $j < \frac{1}{2}\log n$. Near $j = \frac{1}{2}\log n$, neither will be very useful. We will show that $\mathbb{A}$ spends at least $\Omega(n^\xi)$ steps, from when it enters B_{j^*} with $j^* = \frac{1+\delta}{2}\log n$ to when it reaches $B_{\frac{1}{2}(\log n)}$ for some constant δ defined later.

We define Fast to be the event that the routing time T is less than n^ξ. Now we only need to show that this event is unlikely, i.e., $\Pr[\text{Fast}] = o(1)$. Basically,

$$\mathbb{E}[T] = \mathbb{E}[T|\text{Fast}] \cdot \Pr[\text{Fast}] + \mathbb{E}[T|\neg\text{Fast}] \Pr[\neg\text{Fast}] = \Omega(n^\xi),$$

because even if we allow the first term to be 0, in the second term $\mathbb{E}[T|\neg\text{Fast}] = \Omega(n^\xi)$ and $\Pr[\neg\text{Fast}] = 1 - o(1)$.

In order to show that Fast is unlikely, we define another event Jump, which occurs if there exists a jump from $\neg B_{j^*+1}$ to B_{j^*} during the routing process. Then we have that:

$$\begin{aligned} \Pr[\text{Fast}] &= \Pr[\text{Fast} \wedge \text{Jump}] + \Pr[\text{Fast} \wedge \neg\text{Jump}] \\ &\leq \Pr[\text{Jump}|\text{Fast}] + \Pr[\text{Fast}|\neg\text{Jump}] \end{aligned} \tag{3.1}$$

Now we are going to upper bound the two probabilities separately.

We denote by S_i the set of nodes that have been explored by time i, termed the *explored nodes*. Additionally, we denote by x_i, the node that has the message at time i, and $d_i = d(x_i, t)$ the grid distance from x_i to the target. Since each step the

algorithm may visit at most a new node, the size of the explored nodes at time i is at most i. And the total number of weak edges from the explored nodes S_i is bounded by $q|S_i| \leq qi$.

If Jump is true. Here we show that

$$\Pr[\text{Fast} \wedge \text{Jump}] < \Pr[\text{Jump}|\text{Fast}] = O\left(\frac{\epsilon_0}{n^{\frac{\epsilon_0(3-\epsilon_0)}{2(3+\epsilon_0)}}}\right).$$

Intuitively, if Fast is true, then the routing process touches T nodes at most. The total number of ties that could be used by $\mathbb{A}$, the ties of the nodes touched by $\mathbb{A}$, is not enough to get a good probability for Jump to happen, since Jump for any particular node is an event with small probability. Now we carry out the calculation.

Let $\delta = \frac{2\epsilon_0}{3+\epsilon_0}$. We now define the event Jump_i to be that a jump occurs from node x_i. We can lower bound $\Pr[\text{Jump}_i]$ by considering $\Pr[\text{Jump}_i|\gamma_i]$ and maximizing over all possible γ_i.

$$\Pr[\text{Jump}_i|\gamma_i = 2+\epsilon] = O\left(q\frac{n^{1+\delta}}{\lambda_{2+\epsilon} n^{\frac{1+\delta}{2}\cdot(2+\epsilon)}}\right) = \begin{cases} O\left(\frac{\epsilon}{n^{(1+\delta)\epsilon/2}}\right) & \text{if } \epsilon > 0 \\ O\left(\frac{|\epsilon|}{n^{(1-\delta)|\epsilon|/2}}\right) & \text{if } \epsilon < 0 \end{cases},$$

and $\Pr[\text{Jump}_i \mid \gamma_i = 2+\epsilon] = O\left(\frac{\epsilon_0}{n^{(1-\delta)\epsilon_0/2}}\right)$. Conditioning on the event Fast, and applying a union bound over all the explored nodes in $x_i \in S_T \setminus B_{j^*+1}$, we have

$$\Pr[\text{Jump}|\text{Fast}] = O\left(n^{\xi} \cdot \frac{\epsilon_0}{n^{(1-\delta)\epsilon_0/2}}\right) = O\left(\frac{\epsilon_0}{n^{(1-\delta)\epsilon_0/2-\xi}}\right). \tag{3.2}$$

If Jump is not true. Here we show that

$$\Pr[\text{Fast} \wedge \neg\text{Jump}] \leq \Pr[\text{Fast}|\neg\text{Jump}] \leq O\left(\frac{\epsilon_0}{n^{\frac{\epsilon_0(3-\epsilon_0)}{6(3+\epsilon_0)}}}\right).$$

In this case, since Jump is not true, we do not have 'long' ties. Then we show that a message not using any long ties cannot travel quickly. Thus Fast is unlikely to be true.

Given a process of $\mathbb{A}$, we can define $1 \leq \tau < \sigma \leq T$ such that x_τ is the last node outside of B_{j^*+1} and σ is the first node inside $B_{\frac{1}{2}\log n}$:

$$\tau = \arg\max\{x_t \notin B_{j^*+1}\} \text{ and } \sigma = \arg\min\{x_t \in B_{\frac{1}{2}\log n}\} \tag{3.3}$$

Now suppose Jump is not true. Since x_τ is the last node outside of B_{j^*+1}, the next

hop $x_{\tau+1}$ stays inside B_{j^*+1} but cannot reach B_{j^*}, i.e., $x_{\tau+1} \in B_{j^*+1} \setminus B_{j^*}$. Therefore, $\forall i$ such that $\tau < i < \sigma, n^{\frac{1}{2}} \leq d_i \leq n^{\frac{1}{2}(1+\delta)}$. We will consider the nodes x_i for i inside this range (τ, σ).

We consider events $Hop_i(\rho)$ to be that the message moves a $\rho = \frac{1}{n^\xi}$ fraction of the distance closer using a tie from x_i, that is, $d_{i+1} < (1-\rho)d_i$. Now,

$$\Pr\left[Hop_i(\rho) | \gamma_{x_i} = 2+\epsilon\right] = \begin{cases} O\left(\frac{\epsilon}{\rho^{2+\epsilon} d^\epsilon}\right) = O\left(\frac{\epsilon_0}{\rho^{2+\epsilon_0} n^{\epsilon_0/2}}\right) & \text{if } \epsilon > 0 \\ O\left(\frac{|\epsilon| d^{|\epsilon|}}{\rho^{2-|\epsilon|} n^{|\epsilon|}}\right) = O\left(\frac{\epsilon_0}{\rho^{2-\epsilon_0} n^{\epsilon_0(1-\delta)/2}}\right) & \text{if } \epsilon < 0 \end{cases}$$

The last equation is due to the fact that $n^{\frac{1}{2}} \leq d_i \leq n^{\frac{1}{2}(1+\delta)}$, for all i in the range of $\tau < i \leq \sigma$.

We can then partition our event Fast$|\neg$Jump with respect to $Hop = \bigcup_{\tau \leq i \leq \sigma} Hop_i(\rho)$ as follows

$$\begin{aligned} \Pr \text{Fast} | \neg \text{Jump} &= \Pr\left[\text{Fast}, Hop | \neg \text{Jump}\right] + \Pr\left[\text{Fast}, \neg Hop | \neg \text{Jump}\right] \\ &\leq \Pr\left[Hop | \neg \text{Jump}, \text{Fast}\right] + \Pr\left[\text{Fast}, \neg Hop | \neg \text{Jump}\right] \end{aligned}$$

and we upper bound these two terms. The first term can be bounded above by taking a union bound over all i between σ and τ, which is bounded by $T < n^\xi$. The probability that there exists such a hop between τ and σ is

$$\Pr\left[\bigcup_{\tau \leq i \leq \sigma} Hop_i(\rho)\right] \leq O\left(\frac{\epsilon_0}{n^{\epsilon_0/2-(3+\epsilon_0)\xi}}\right) + O\left(\frac{\epsilon_0}{n^{\epsilon_0(1-\delta)/2-(3-\epsilon_0)\xi}}\right) = O\left(\frac{\epsilon_0}{n^{\frac{\epsilon_0(3-\epsilon_0)}{6(3+\epsilon_0)}}}\right).$$

For the second term, if none of the Hop_i events is true, $d_{i+1} \geq (1-\rho)d_i$. Thus,

$$d_\sigma \geq (1-\rho)^{\sigma-\tau-1} d_{\tau+1} \geq \left(1 - \frac{1}{n^\xi}\right)^{n^\xi} 2^{j^*+1} \geq \frac{1}{e} n^{\frac{1}{2}(1+\delta)} \geq n^{1/2}$$

This contradicts with the fact that $x_\sigma \in B_{\frac{1}{2}\log n}$. Combining these two we have,

$$\Pr\left[\text{Fast} | \neg \text{Jump}\right] = O\left(\frac{\epsilon_0}{n^{\frac{\epsilon_0(3-\epsilon_0)}{6(3+\epsilon_0)}}}\right) + 0 \tag{3.4}$$

Apply Equation (3.2) and (3.4) to Equation (3.1) we know $\Pr\left[\text{Fast}\right] = o(1)$ and thus the theorem is true.

□

3.6 Complex Contagion Upper Bounds

The spreading of k-complex contagion (Definition 2.1) on the original Kleinberg's model has been fully characterized in [99] and [83]. If a k-seed cluster is infected initially, the contagion spreads to the entire network in $O(\text{polylog}(n))$ rounds if $\gamma \in [2, \beta_k)$, where $\beta_k = \frac{2(k+1)}{k}$, and in $\Omega(\text{poly}(n))$ rounds otherwise.

3.6.1 Non-negligible Mass Near 2

In the heterogeneous Kleinberg model, we first show a result that is analogous to our results for myopic routing: as long as the distribution $\mathcal{D}$ for γ_u has a non-negligible amount of mass near 2, then for any k, k-complex contagions spread in polylog time—but the exponent of $\log n$ depends on k and $\mathcal{D}$.

Theorem 3.5. *Fix a distribution $\mathcal{D}$, an integer $k > 0$ and $\eta > 0$. If there exist constants $\epsilon_0 > 0$ and $\alpha \geq 0$ where $M_{\mathcal{D}}(\epsilon) \geq K\epsilon^{\alpha}$ for all $\epsilon \leq \epsilon_0$, and $p, q \geq k$, there exists $\kappa = k\alpha + \frac{k(k+1)}{2}$, such that a k-complex contagion $\text{CC}(\text{HetK}_{p,q,\mathcal{D}}(n), k, I)$ starting from a k-seed cluster $\mathcal{I}$ takes at most $O(\log^{(3+\kappa)/2} n)$ rounds[1] to spread to the whole network with probability at least $1 - n^{-\eta}$ over the randomness of $\text{HetK}_{p,q,\mathcal{D}}(n)$.*

The theorem is based on the observation that the infected region doubles its size in a polylogarithmic number of steps. In this way the general proof framework is similar to that in [99].

Choose some node $t \in I$, which, recall, is of constant size. Define *ball* B_j as the set of nodes with Manhattan distance less than 2^j from t and *annulus* A_j as the set of nodes with Manhattan distance greater or equal to 2^j and less than 2^{j+1} from t. We say that the contagion is in *phase* j when the all the nodes in B_j are infected but not all nodes in B_{j+1} are infected. Note that the largest value of j is bounded by $O(\log n)$ and the value of phase j is non-decreasing, because nodes do not become uninfected.

By definition, $(u_1, \ldots, u_k)$ is a k-seed cluster only if for all $1 \leq i < j \leq k$, $d(u_i, u_j) \leq k \leq q$ there are $s - 1$ strong ties from u_s to $u_1 \ldots, u_{s-1}$. A k-seed cluster $(u_1, \ldots, u_k)$ in annulus A_j will be infected in k round after phase j if u_s has $k + 1 - s$ weak ties to nodes in B_{j-1}, and we call the k-seed cluster $(u_1, \ldots, u_k)$ to be *good*. We first bound the probability of a good seed cluster.

[1] the scalar depends on the the constants k, η, α, K

Lemma 3.6. *There exists some constant $C > 0$ depending on $\mathcal{D}$ and k such that for all $0 \le j \le \log n$*

$$\Pr\left[\textit{a } k\textit{-seed cluster in annulus } j \textit{ is good}\right] \ge \frac{C}{\log^{\kappa} n} \tag{3.5}$$

where $\kappa = k\alpha + \frac{k(k+1)}{2}$, and these events for all nonoverlap k-seed clusters are mutually independent.

The proof is fairly technical and is delayed to the appendix.

Now we are ready to prove Theorem 3.5.

Proof of Theorem 3.5. Firstly we partition the nodes in A_j into $\Theta\left(\frac{C|A_j|}{k(\eta+2)\log^{1+\kappa} n}\right)$ balls such that each ball has size $\Theta\left(\frac{k(\eta+2)}{C}\log^{1+\kappa} n\right)$ and contains $\frac{\eta+2}{C}\log^{1+\kappa} n$ disjoint k-seed clusters. If each ball in A_j has at least one good k-seed cluster, this good k-seed will be infected in k rounds after phase j. Then the time for all noded in this ball to be infected is bounded above by the diameter of the ball, $\Theta\left(\log^{(1+\kappa)/2} n\right)$. Therefore, it takes $O\left(\log^{(1+\kappa)/2} n\right)$ rounds to enter the next phase. There are at most $O(\log n)$ phases. So the total time for all nodes to be infected is $O\left(\log^{3/2+\kappa/2} n\right)$. Moreover, by Equation (3.5) and the union bound, the probability that all such balls have a good k-seed cluster is at least $1 - n^2\left(1 - \frac{C}{\log^{\kappa} n}\right)^{\frac{\eta+2}{C}\log^{1+\kappa} n} \ge 1 - \frac{1}{n^{\eta}}$. □

3.6.2 Fixed k

For a specific k, we can show that as long as the distribution $\mathcal{D}$ has constant mass in the interval $[2, \beta_k)$ (recall for the beginning of the section that $\beta_k = \frac{2(k+1)}{k}$), then the k-complex contagion will spread to the entire network in a polylogarithmic number of rounds. Recall that the results in Theorem 3.5 only require non-negligible mass near 2. Here we require constant mass, but the mass need not be asymptotically close to 2 as long as it is in the interval $(2, \beta_k)$.

Theorem 3.7. *Fix a distribution $\mathcal{D}$, an integer $k > 0$ and $\eta > 0$. If $\Pr_{\gamma\sim\mathcal{D}}\gamma \in [\mathbf{2}, \boldsymbol{\beta}_{\mathbf{k}}) > \mathbf{0}$ where $\beta_k = \frac{2(k+1)}{k}$, and $p, q \ge k$. There exists $\xi > 0$ depending on $\mathcal{D}$ and k such that, the speed of a k-complex contagion $\mathrm{CC}(\mathrm{HetK}_{p,q,\mathcal{D}}(n), k, I)$ starting from a k-seed cluster I is at most $O\left(\log^{\xi} n\right)$ with probability at least $1 - n^{-\eta}$.*

The proof of Theorem 3.7 uses the same divide and conquer strategy as in [83]. We first state Definition 3.8 from [83] and a technical lemma, Lemma 3.9, that says Definition 3.8 is sufficient for a k-complex contagion to spread fast.

Definition 3.8 ([83])**.** Fix constants δ, c, η, k, let λ_γ be the normalization factor such that $\lambda_\gamma = \sum_{v \neq u} d(u,v)^{-\gamma}$ and let constant $r = O((\frac{6\eta}{\lambda_\gamma})^c)$. We say that a $\mathrm{HetK}_{p,q,\mathcal{D}}(n)$ model is (δ, c, η, k)-*recursively spreading* if whenever there exists some constant $\gamma > 2$ such that

1. S is an ℓ-sized square ($\sqrt{\ell} \times \sqrt{\ell}$) of vertices in $\mathrm{HetK}_{p,q,\mathcal{D}}(n)$ where $\ell > (r \log^c(n))^{\frac{1}{1-\delta}}$;
2. A and B are any two disjoint $\ell^{1-\delta}$-sized subsquares of S; and
3. A is fully infected,

then with probability at least $1 - \ell^{2(1-\delta)}/n^\eta$, there is a new k-seed cluster in B that is infected in at most k rounds. The probability is over the coin flips of the $\mathrm{HetK}_{p,q,\mathcal{D}}(n)$ model.

Lemma 3.9 ([99])**.** *Fix constants δ, c, η, k. If a $\mathrm{HetK}_{p,q,\mathcal{D}}(n)$ model is (δ, c, η, k)-recursively spreading , then if we start a k-complex contagion from a k-seed cluster, it takes at most $O\left(\log^\xi n\right)$ rounds for the contagion to spread to the whole network with probability at least $1 - n^{-\eta}$, where $\xi = \frac{c}{2} + \log_{\frac{1}{1-\delta}} 2$, and the probability is over the coin flips of the $\mathrm{HetK}_{p,q,\mathcal{D}}(n)$ model.*

By the above Lemma, we only need to prove that $\mathrm{HetK}_{p,q,\mathcal{D}}(n)$ in Theorem 3.7 is (δ, c, η, k)-recursively spreading.

Lemma 3.10. *Suppose $\Pr[\gamma \in (2, \beta_k)] > 0$, and let*

$$\tilde{\gamma} = \inf\left\{2 \leq \gamma < \beta_k : \Pr[\gamma \in [2, \gamma]] \geq \frac{1}{2}\Pr[\gamma \in [2, \beta_k)]\right\},^2$$

$0 < \delta < 1 - \tilde{\gamma}/\beta_k$, $\frac{1-\delta}{(k+1)(1-\delta) - k\tilde{\gamma}/2} \leq c$, and $0 < \eta$, then $\mathrm{HetK}_{p,q,\mathcal{D}}(n)$ is (δ, c, η, k)-recursively spreading.

The proof is fairly technical and is delayed to the appendix.

3.7 Complex Contagion Lower Bounds

In this section, we describe a polynomial time lower bound for the spreading rate of k-complex contagion on the Heterogeneous Kleinberg Small World $\mathrm{HetK}_{p,q,\mathcal{D}}(n)$, when the distribution $\mathcal{D}$ on the personal parameter γ_u has zero weight around two. Here we first state the theorem for a fixed k, and the result near two is a natural corollary.

[2] $\tilde{\gamma}$ is the median of γ between 2 and β_k which measure how nice the distribution $\mathcal{D}$ behaves in interval $[2, \beta_k)$ on average.

Theorem 3.11 (Lower bound for fixed k)**.** *Given distribution $\mathcal{D}$, constant integers $k, p, q > 0$, and $\epsilon_0 > 0$ such that $M_{\mathcal{D}}(\beta_k + \epsilon_0) - F_{\mathcal{D}}(2 - \epsilon_0) = 0$, then there exist constants $\xi, \eta > 0$ depending on $\mathcal{D}$ and k, such that the time it takes a k-contagion starting at seed-cluster $\mathcal{I}$, $\mathrm{CC}(\mathrm{HetK}_{p,q,\mathcal{D}}(n), k, \mathcal{I})$, to infect all nodes is at least $\Omega(n^{\xi})$ with probability at least $1 - O(n^{-\eta})$ over the randomness of $\mathrm{HetK}_{p,q,\mathcal{D}}(n)$.*

If $\mathcal{D}$ satisfies the condition in Theorem 3.11, we can partition the support into two disjoint sets $Supp\{\mathcal{D}\} = D_1 \cup D_2$ such that $\gamma_1 = 2 - \epsilon_1 = \sup\{\gamma \in D_1\} < 2 - \epsilon_0$, and $\gamma_2 = 2 + \epsilon_2 = \inf\{\gamma \in D_2\} > 2(1 + 1/k) + \epsilon_0$.

Ebrahimi et al. [83] proved for the original Kleinberg model if $\gamma > \frac{2(k+1)}{k}$ the weak ties will be too short to create remote k-seeds; on the other hand, if $\gamma < 2$ the weak ties will be too random to form k-seeds at all. Similar to proving the lower bound for myopic routing, the challenge in proving this theorem is the synergy between concentrated and diffuse edges which can possibly be exploited by k-complex contagions in the heterogeneous Kleinberg model. We resolve this by considering a scale where neither type of edges is helpful.

Before proving Theorem 3.11 we state a corollary concerning a lower bound when there is no mass around 2.

Corollary 3.12 (Lower bound for no mass around 2)**.** *Given distribution $\mathcal{D}$, constant integers $p, q > 0$, and $\epsilon_0 > 0$ such that $F_{\mathcal{D}}(2 + \epsilon_0) - F_{\mathcal{D}}(2 - \epsilon_0) = 0$, there exist a constant integer $k > 0$ and $\xi, \eta > 0$ such that the time it takes a k-contagion starting at seed-cluster $\mathcal{I}$, $\mathrm{CC}(\mathrm{HetK}_{p,q,\mathcal{D}}(n), k, \mathcal{I})$, to infect all nodes is at least n^{ξ} with probability at least $1 - O(n^{-\eta})$ over the randomness of $\mathrm{HetK}_{p,q,\mathcal{D}}(n)$.*

The corollary follows directly from Theorem 3.11 by taking a sufficiently large k.

Proof of Theorem 3.11. Given a k-complex contagion $\mathrm{CC}(\mathrm{HetK}_{p,q,\mathcal{D}}(n), k, I)$, let $I_t \subseteq V$ be the set of infected nodes at time t. We let $r_t = \max_{u \in I_t} d(u, s)$ be the *radius of infected nodes at time t* by fixing some arbitrary node $s \in I$. Since the k-complex contagion begins with $r_0 = O(1)$ and does not infect the whole graph until time T with $r_T = \Theta(n)$, to lower bound the time of contagions it suffices to upper bound the radius r_t for the t-th round. To achieve that, we first make the following definition.

Definition 3.13. Denote $\rho_1 = n^{\delta_1}, \rho_2 = n^{\delta_2}$, where

$$\delta_1 < \min\left(\frac{k\epsilon_1}{k\epsilon_1 + 2}, \frac{k-1}{k}\right) \text{ and } \frac{2}{k\epsilon_2}\delta_1 < \delta_2 < \delta_1 \text{ are constants.}$$

Let the event $\mathcal{J}$ be as follows (see Figure 3.1):

$$\mathcal{J} = \{\exists\, v \text{ with } k \text{ edges } (v,u_1),\ldots,(v,u_k) \text{ s.t. } 1 \le i \le k, d(u_i,s) \le \rho_1 \text{ and } d(v,u_i) \ge \rho_2\}$$

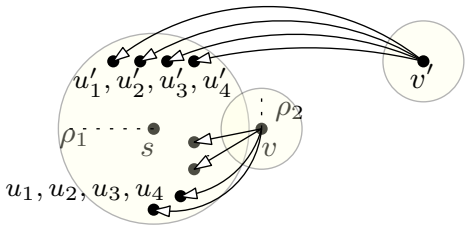


Figure 3.1: Event $\mathcal{J}$.

We will use the following technical lemma, whose proof is deferred.

Lemma 3.14. *$\mathcal{J}$ occurs with probability at most $o(n^{-\eta})$ where $\eta > 0$ and depends on $\delta_1, \delta_1, \epsilon_1, \epsilon_2$ and k.*

Now, let's suppose $r_\sigma = \rho_2$ and $r_\tau = \rho_1$ and even $\mathcal{J}$ does not happen. Each round $\sigma \le t \le \tau$ the increment of radius r_t is bounded by δ_2, $\tau - \sigma = \Omega\left(\frac{n^{\delta_1}}{n^{\delta_2}}\right) = \Omega(n^{\delta_1-\delta_2})$. Thus the statement is true by taking

$$\begin{aligned}
&\xi \le \delta_1 - \delta_2;\\
&\eta < \min\left(\frac{k\epsilon_1}{k\epsilon_1+2} - \delta_1, \frac{k-1}{k} - \delta_1, (\epsilon_2 - \frac{2}{k})\delta_1, \delta_2 - \frac{2}{k\epsilon_2}\delta_1\right).
\end{aligned}$$

□

3.8 Experiments

In this section, we report the analysis of several real world social networks that motivated our model. In Kleinberg's model all nodes are homogeneous and follow the same parameter γ in selecting their edges. We denote this as the *assumption of homogeneity*. The opposite of this assumption is that nodes are heterogeneous. The length of the weak ties (defined in some underlying social metric space) are not independent. A node with one long edge is likely to have other long edges and vice versa.

To test this assumption we examine a number of real world social networks. Firstly, we sort the edges according to *Jaccard similarity* [148] (or termed the neighborhood overlap) of their endpoints which is defined in section 3.8.2, and partition edges into strong and weak ties. Secondly, we find an embedding of the network using

the strong ties in some metric space (Σ, d) and calculate the length of each weak tie using this embedding. We sort the weak ties by length and produce a ranking of the edges with increasing lengths. Finally we evaluate the *variance of the ranks of the weak ties incident to the same node* u, for every node u in the network. In comparison, we randomly rewire those weak ties and re-compute the rank of these permuted weak ties on the same embedding and the average variance of the rank.

If the nodes in the network are homogeneous, i.e., follow the same model in choosing social ties, then the averaged variance of the ranking of edges incident to the same node should not change much before and after the permutation. If we observe a difference in the variance, say, the variance before permutation is smaller, then it must be that some of the nodes have a biased preference to long ties (or short ties), which implies that the nodes are not homogeneous. Formally the Null Hypothesis to test here states that the nodes are homogeneous and thus the variance before and after the permutation test should stay similar to each other. We run the permutation test to reject this hypothesis.

Notice that the above test is only for the homogeneity of the nodes and does not make any additional assumptions on how the social ties are selected. Further, we use the variance of the ranking of the edges. This is more robust than the variance of the edge lengths, which depends more on the specific embedding used.

3.8.1 Dataset Description

We use 4 social network data sets on the Stanford Network Analysis Project: Facebook [157] and Twitter [157] as well as the Wikipedia voting network [143] and the Epinion network [196]. The number of nodes in these networks vary between 4000 and 100,000 nodes.

The Facebook data set consists of friend ties collected from participants using a certain Facebook app. The Twitter data has similar attributes to the Facebook one and it was crawled from public sources. The Wikipedia data collected 2794 admin elections with 103,663 total votes and 7066 users participating in the elections. The Epinion data set is a who-trust-whom online social network of a general consumer review site. These networks vary in size and degree distributions. We summarize the basic statistics of the network data sets we used in Table 3.1.

Social network	Facebook	Wiki-vote	Twitter	Epinion
Nodes	4039	7115	81,306	75,879
Edges	88,234	103,689	1,768,149	508,837
Average degree	21.85	14.57	21.75	6.70
Nodes in 2-core	3964	4786	76,311	37,300
Edges in 2-core	88,159	98,456	1,337,315	367,162

Table 3.1: Data set statistics

3.8.2 Implementation Details

The *Jaccard similarity* of $e = (u, v)$ measures the tie strength and is defined as $J(u, v) = \frac{|N_u \cap N_v|}{|N_u \cup N_v|}$, where N_v is the set of vertices adjacent to node v. If the Jaccard similarity is high, the strength of the edge is 'strong', and vice versa. We take the 10% edges with the smallest Jaccard similarity as weak ties E_W, and the other as strong ties E_S where $E = E_S \cup E_W$. Computing the embedding on the strong ties and the average variance of rank over weak ties helps to avoid the interference between embedding and our test.

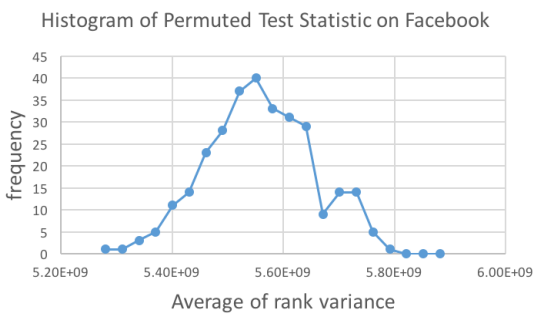
The most popular method to embed a graph into metric space is to use spectral techniques [108], which, intuitively tries to minimize the distance between neighboring vertices in the embedding. Formally, we create $x : V \to \mathbb{R}^d$ such that $x = (x_1, ..., x_d)$ where $x_i : V \to \mathbb{R}$. x_i then can be seen as vector where $x_0 = 1^n$ and for $i > 0$ we obtain x_i by solving

$$x_i = \underset{x^\top x_k = 0, \forall 0 \leq k < i}{\arg\min} \sum_{(u,v) \in E_S} \|x(u) - x(v)\|^2.$$

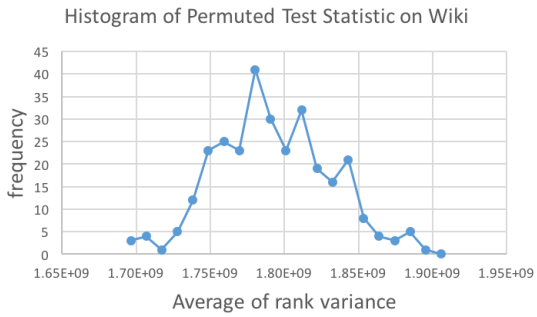
This can be solved efficiently and x_i will be the eigenvector corresponding to the ith largest eigenvalue of the Laplacian matrix of the graph. In this experiment, we take $d = 2$ for embedding. The Figure 3.3 (b) is our embedding on Kleinberg's small world after removing edges with low Jaccard similarity, E_W, and provide evidence that this embedding captures the hidden metric structure of this graph.

A rewired graph $G' = (V', E')$ of $G = (V, E_W)$ is constructed by carrying out a series of switching steps ($10|E_W|$ times), in which a pair of edges are selected uniformly at random (among the weak ties) and the endpoints are exchanged.

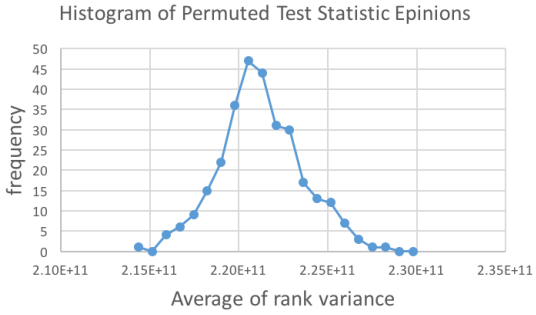
A commonly known issue with spectral embedding is that it can be greatly influenced by nodes that are barely connected to the rest of the graph. For this reason, we first take the k-core of the network. A k-core is a maximal connected subgraph of the original graph in which all vertices have degree at least k. This can be obtained by



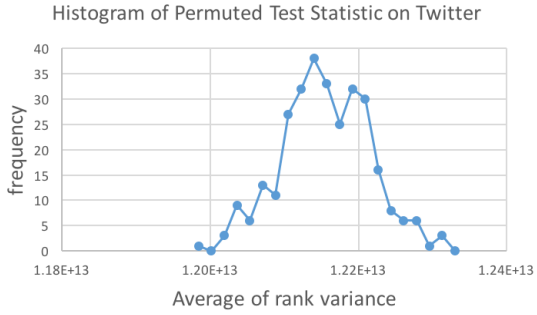
Facebook $(4.5 \cdot 10^9)$



Wiki $(9.2 \cdot 10^8)$



Epinions $(1.6 \cdot 10^{11})$



Twitter $(8.1 \cdot 10^{12})$

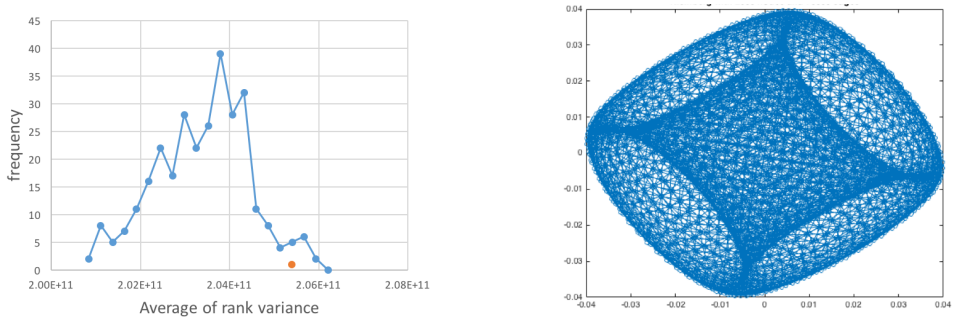
Figure 3.2: Histogram of permuted test statistics for each real social network. Since the observed test statistic is much smaller than the permuted ones, we only plot the histogram of 300 permuted test statistics, and the p-value are all below 0.0033

iteratively removing nodes of degree less than k. In our experiments we take $k = 2$. In the networks we study, the node degrees are reasonably high so we did not lose many edges by doing this.

3.8.3 Experimental Results

Figure 3.2 reports the averaged rank variance of our permutation test with 300 trials. In all four networks, the averaged rank variance before random permutation is much smaller than the value obtained in all the 300 permutation trials. Using the empirical distribution we obtain 0.33% as the p-value on the statistical hypothesis that the nodes are homogeneous in choosing their social ties.

As a control example, we consider the standard Kleinberg's small world model in Figure 3.3 (a). The result shows that more than 5% of permuted test statistics (16 out of 300) are larger than the original test statistics. Therefore the null hypothesis is not rejected on the standard Kleinberg's small world.



(a) Histogram of permuted test statistics (b) Spectral embedding

Figure 3.3: In the control group, we take the 20% edges with the smallest Jaccard similarity as weak ties and conduct the same permutation test. In (a), we consider Kleinberg's small world of $10,000$ nodes with $p = 2, q = 3$ the observed test statistic is represented as the red dot on the same scale with the histogram of 300 permuted test statistics. In (b), we plot the spectral embedding of Kleinberg's small world of $2,500$ nodes with $p = 2, q = 3$ after removing the weak ties.

CHAPTER 4

Configuration Models with Power-law Degree distributions

In this chapter we continue analysis of k-complex contagions (sometimes called boot-strap percolation). Our main result show if the graph is configuration model with power-law distribution with exponent $\alpha \in (2,3)$ (defined in Section 2.5.2), then with high probability, the single seed of the highest degree node will infect a constant fraction of the graph within time $O\left(\log^{\frac{\alpha-2}{3-\alpha}}(n)\right)$. This complements the prior work which shows that for $\alpha > 3$ boot strap percolation does not spread to a constant fraction of the graph unless a constant fraction of nodes are initially infected. This also establishes a threshold at $\alpha = 3$.

The case where $\alpha \in (2,3)$ is especially interesting because it captures the exponent parameters often observed in social networks (with approximate power-law degree distribution). Thus, such networks will spread complex contagions even lacking any other structures.

We additionally show that our theorem implies that $\omega\left(n^{\frac{\alpha-2}{\alpha-1}}\right)$ random seeds will infect a constant fraction of the graph within time $O\left(\log^{\frac{\alpha-2}{3-\alpha}}(n)\right)$ with high probability. This complements prior work which shows that $o\left(n^{\frac{\alpha-2}{\alpha-1}}\right)$ random seeds will have no effect with high probability, and this also establishes a threshold at $n^{\frac{\alpha-2}{\alpha-1}}$.

4.1 Introduction

Janson et al. [119] show that k-complex contagions do not spread on sparse $G(n,p)$ random graphs. Such cascades require $\Omega(n)$ seeds to infect a constant fraction of vertices. Balogh and Pittel [22] extended these results to configuration model graphs with regular degree distributions.

However, many networks do not have regular degree distributions. In a graph with power law degree distribution, the number of nodes having degree d is proportional to $1/d^{\alpha}$, for a positive constant α. In 1965, Price [193] showed that the number of citations to papers follows a power law distribution. Later, studies of the World Wide Web reported that the network of webpages also has a power law degree distribution [24, 43]. Observations of many different types of social networks also found power law degree distributions, as well as biological, economic and semantic networks [218, 7, 177].

Additional work by Amini [12] studies the configuration model with power-law degree distribution for $\alpha > 3$ and showed and shows theorem which implies (see Section 4.6) that, with high probability, infecting a constant fraction of the nodes requires an initial seed that comprises a constant fraction of the graph.

Amini and Fountoulakis [13] also have examined the Chung-Lu model with power-law exponent $2 < \alpha < 3$. They show that there exists a function $a(n) = o(n)$ such that if the number of initial seeds is $\ll a(n)$, the process does not evolve w.h.p.; and if the number of initial seeds is $\gg a(n)$, then a constant fraction of the graph is infected with high probability. However, this function is still super-constant—$n^{\Omega(1)}$.

The question remained open, can non-submodular cascades spread and spread quickly from a constant-sized seed set on sparse graphs with no other structure imposed besides a skewed degree distribution.

4.1.1 Our Contributions

Our main result is that for a configuration model graph with power-law exponent $\alpha \in (2, 3)$, with high probability, the single seed of the highest degree node will infect a constant fraction of the graph within time $O(\log^{\frac{\alpha-2}{3-\alpha}}(n))$. This complements the prior work which showed that for $\alpha > 3$ boot strap percolation does not spread to a constant fraction of the graph unless a constant fraction of nodes are initially infected. This also establishes a threshold at $\alpha = 3$.

The case where $\alpha \in (2, 3)$ is especially interesting because it captures the exponent parameters often observed in social networks (with approximate power-law degree distribution). Thus, such networks will spread complex contagions even lacking any other structure.

We additionally show that our main theorem implies that $\omega(n^{\frac{\alpha-2}{\alpha-1}})$ random seeds will infect a constant fraction of the graph within time $O(\log^{\frac{\alpha-2}{3-\alpha}}(n))$. This complements the prior work which shows that $o(n^{\frac{\alpha-2}{\alpha-1}})$ random seeds will have no effect with

high probability. This also establishes a threshold at $n^{\frac{\alpha-2}{\alpha-1}}$.

To prove these results, we provide new analysis that circumvents previous difficulties. While our results are similar to those of Ebrahimi et al. [82] (they study the preferential attachment model, while we study the configuration model), the techniques required are completely different. For example, it is an easy observation that k-complex contagions spread on the configuration model (if k is greater than the minimum degree), but much more difficult to show it spreads quickly.

The previous analyses on the configuration model required that the graph was locally tree-like, an assumption that fails in our case, and then were able to approximate the process using differential equations and obtain rigorous results by applying Wormald's Theorem [239]. However, their analysis fails when the degree distribution is power-law with exponent between 2 and 3.

4.2 Main Theorem: power-law exponent $\alpha \in (2, 3)$

In this section, we state and prove our main theorem: in a configuration model graph with the power-law exponent $\alpha \in (2, 3)$, with high probability, the single seed of the highest degree node will infect a constant fraction of the graph within time $O(\log^{\frac{\alpha-2}{3-\alpha}}(n))$.

We use the *configuration model* introduced by Bollobás and McKay [41] to define a distributions over multigraphs. The definition is in 2.5.2.

Theorem 4.1. *Given a power law distribution $\boldsymbol{d} = (d_1, ..., d_n)$ with exponent $\alpha \in (2, 3)$ and $d_1 > n^{\frac{3-\alpha}{\alpha+1}}$, with probability $1 - O\big(\frac{\log^{\frac{\alpha-1}{3-\alpha}} n}{n}\big)$, the k-complex contagion on configuration model $CM(\boldsymbol{d})$ with constant k and initial infection being the highest degree node $I = \{1\}$, $CC(CM(\boldsymbol{d}), k, I)$, infects $\Omega(n)$ vertices within time $O(\log^{\frac{\alpha-2}{3-\alpha}} n)$.*

4.2.1 Proof Setup

We consider a restricted form of contagion where nodes can only be infected by those proceeding them in the ordering. Formally, recall the nodes $\{d_i\}$ are ordered in terms of their degree. Node d_i will only be infected if $|\{j : j < i \text{ and } d_j \text{ is infected }\}| \geq k$ neighbors are infected. Hence, the total number of infected nodes in this process will be fewer than the number of infected nodes in original complex contagions, and it is sufficient to prove that a constant fraction of nodes become infected in this restricted contagion with high probability.

Buckets We first partition the nodes $V = [n]$ into buckets. We design the buckets to have at least (and about the same number of) stubs $b = \Theta(\frac{n}{\log^{\frac{\alpha-2}{3-\alpha}} n})$. We can define N_ℓ as follows

$$N_1 = \frac{n}{\log^{\frac{\alpha-1}{3-\alpha}} n} \text{ , and } N_{\ell+1} = \arg\min_{i>N_\ell}\{s(i) - s(N_\ell) \geq b\}$$

Since $d(N_1) = \Theta(\log^{1/(3-\alpha)} n) = o(b)$ and $\forall i > N_1, d(i) \leq d(N_1)$,

$$b < s(N_{\ell+1}) - s(N_\ell) \leq b + o(b) < 2b.$$

Therefore, we have $\ell b \leq s(N_\ell) \leq 2\ell b$ and $N_\ell = \Theta\big(\frac{n}{\log^{\frac{\alpha-1}{3-\alpha}} n} \ell^{\frac{\alpha-1}{\alpha-2}}\big)$ by (4), and so the total number of buckets is $L \leq \frac{s(n)}{b} = O(\log^{\frac{\alpha-2}{3-\alpha}} n)$.

We define our buckets to be $B_1 = \{1, .., N_1\}, B_2 = \{N_1 + 1, .., N_2\}, .., B_{\ell+1} = \{N_\ell + 1, .., N_{\ell+1}\}, ..., B_L = \{N_{L-1} + 1, .., N_L\}$.

Filtration We now state our filtration.

$\mathcal{F}_0$**:** The node i starts with d_i stubs of edges without revealing any edges.

$\mathcal{F}_1$**:** In the first stage we reveal all edges within the first bucket B_1,

$\mathcal{F}_\ell$**,** $1 \leq \ell \leq L$**:** In the stage $\ell > 1$, we reveal/match all the edges from B_ℓ to early nodes in $B_{<\ell}$.

4.2.2 Proof Summary

There are two parts of the proof.

1. All of the nodes in the first bucket would be infected with high probability.
2. For some constant $\rho > 0$, in the first $L' = \rho L$ buckets $B_1, .., B_{L'}$ a constant fraction ϵ of nodes will be infected. Because $N_{L'} = \Omega(n)$ nodes, the total number of infection also constant fraction.

In the first part of the proof is capture by the following lem:powerlaw

Lemma 4.2 (Base)**.** *Given at $\mathcal{F}_0$ $d_1 > n^{\frac{3-\alpha}{\alpha+1}}$, at $\mathcal{F}_1$ all the nodes in B_1 will be infected within $O(\log\log(n))$ steps with probability greater than $1 - O(\frac{1}{n})$.*

To prove this lemma we further decompose the first bucket into $O(\log\log(n))$ finer intervals, which we call bins. We first argue that every node in the first bin will have at least k multi-edges to the first node, and we inductively show the nodes in following bin will have at least k edges to the previous bins. The analysis is by straight-forward probabilistic techniques.

The time for the first bucket's infection is at most the number of the bins because inclusion of each bin only costs 1 step.

We need some additional notation to state the lemma which will imply the second part. Let X_ℓ be the number of stubs from buckets $B_{<\ell}$ to $B_{\geq\ell}$. Let Y_ℓ be the number of uninfect stubs from $B_{<\ell}$ to $B_{\geq\ell}$ before stage ℓ, of which $Y_\ell^{(1)}$ issue from $B_{<\ell-1}$ and the remaining $Y_\ell^{(2)}$ issue from $B_{\ell-1}$. We use $\mathbb{I}_i$ as the indicator variable that node $i \in B_\ell$ is not infected after stage ℓ. Let $\epsilon > 0$ be some constant we define later. Let $\delta_n = \Theta(\frac{1}{\log^{\frac{\alpha-2}{3-\alpha}} n})$.

Now we can formally define $\mathcal{A}_\ell$ as the intersection of the following three events:

1. connection: $(1-\delta_n)\mathbb{E}[X_\ell] \leq X_\ell \leq (1+\delta_n)\mathbb{E}[X_\ell]$;
2. number of uninfected nodes: $\sum_{i\in B_{\ell-1}} \mathbb{I}_i \leq 2\mu_H$ where $\mu_H = K\,\frac{|B_\ell|\ell^{\frac{3-\alpha}{\alpha-2}}}{\log n}$ for some constant K independent of ℓ and n;
3. number of uninfected stubs: $Y_\ell \leq \epsilon X_\ell$

Lemma 4.3 (Induction)**.** *Fix sufficiently small $\epsilon > 0$, $\rho > 0$. Let $\ell < \rho L$, and suppose $\Pr[\mathcal{A}_\ell] > 0.5$, then we have*

$$\Pr[\mathcal{A}_{\ell+1}|\mathcal{A}_\ell] = 1 - O(1)\frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}}$$

This lemma will be proved by showing that each of three events happens with high probability conditioned on A_ℓ. The most technically challenging of these is the second event, where we need to apply Chebychev's Inequality twice. One challenging is that the edges from $B_{<\ell}$ to B_ℓ are not independent. Another challenge is that if the buckets are to small, we fail to have concentration properties, but if they are too large, then the fraction of infected nodes at each stage will drop too quickly.

4.2.3 Proof of Theorem 4.1

Proof. If $\bigcap_{\ell=1}^{L'} \mathcal{A}_\ell$ happens, then the total fraction of infected nodes is $\Omega(n)$.

Using Lemma 4.2 as the base case and Lemma 4.3 as the induction steps we see that

$$\Pr\left[\bigcap_{\ell=1}^{L'}\mathcal{A}_\ell\right] \geq 1 - \sum_{\ell=1}^{L'} O(1)\frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}} - O\left(\frac{1}{n}\right) = 1 - O\left(\frac{\log^{\frac{\alpha-1}{3-\alpha}} n}{n}\right)$$

which is arbitrarily close to 1.

Moreover, the total time spent is the time in first bucket plus the number of buckets (because the infection spreads from bucket to bucket in only 1 step). Therefore the total time spent is

$$O(\log\log n) + O(\log^{\frac{\alpha-2}{3-\alpha}} n) = O(\log^{\frac{\alpha-2}{3-\alpha}} n)$$

which completes our proof. □

4.3 Proof of Lemma 4.2: Contagion in the First Bucket

In this section, we will show that with high probability, the contagion process infects all nodes within the first bucket. Recall that $N_1 = \frac{n}{\log^{\frac{\alpha-1}{3-\alpha}} n}$ and the number of stubs within the first bucket is $S(N_1) = b$.

We partition the first bucket into finer bins such that $B_1 = \bigcup_{t=1}^{T} V_t$ and $V_t = \{v_{t-1}+1, ..., v_t\}, t = 1, ..., T$ with ascending order and $v_0 = 1$. The v_t will be specified in Lemma 4.5. We define the event that every nodes in bin V_t is infected as E_t, then the event that all the nodes in B_1 are infected is equal to $\bigcap_{t=1}^{T} E_t$

We recall lemma 4.2:

Lemma 4.2 (Base)**.** *Given at $\mathcal{F}_0$ $d_1 > n^{\frac{3-\alpha}{\alpha+1}}$, at $\mathcal{F}_1$ all the nodes in B_1 will be infected within $O(\log\log(n))$ steps with probability greater than $1 - O(\frac{1}{n})$.*

We will use two Lemmas in the proof of Lemma 4.2, which will be a proof by induction. The first lemma will form the base case of the induction. It states the high degree nodes will all be infected by the first node by showing any high degree node forms k multi-edges to the first node.

Lemma 4.4. *Given $d_1 > n^{\frac{3-\alpha}{\alpha+1}}$ we define node $v_1 = \max\{v : d(v) \geq n^{\frac{3-\alpha}{\alpha+1}}\}$. (Recall nodes are ordered by degree.) Then all the nodes in $V_1 = \{1, ..., v_1\}$ will be infected in*

one step with probability

$$\Pr[E_1] = 1 - n^{\frac{3-\alpha}{\alpha+1}} \exp\big(-\Theta(1)n^{\frac{3-\alpha}{\alpha+1}}\big)$$

Proof. Consider $u \in V_1, u \neq 1$. By Equation (3) $d_1, d_u \geq \Theta(1)n^{2/(\alpha+1)}$. Thus by Lemma 2.25:

$$\Pr[C(1,u) \leq k] \leq \exp\big(-\Theta(1)n^{\frac{3-\alpha}{\alpha+1}}\big).$$

Applying a union bound on all $u \in V_1$ we have

$$\Pr[\bigcap_{u \in V_1} (C(1,u) \leq k)] \geq 1 - |V_1| \exp\big(-\Theta(1)n^{\frac{3-\alpha}{\alpha+1}}\big)$$

□

The second Lemma will form the inductive step in the proof of Lemma 4.2. It can be proved by induction itself.

Lemma 4.5. *Let* $v \in V_t = \{v_{t-1}+1, ..., v_t\}$ *and* $v_t = max\{v : d(v_t) \geq \frac{n}{\log^{\frac{\alpha-1}{3-\alpha}} n^{(\alpha-2)^t}}\}$, *then*

$$\Pr\left[u \text{ is not infected} \mid \bigcap_{s=1}^{t-1} E_s\right] \leq \frac{1}{n^2}.$$

Moreover, $T = O(\log\log n)$.

Proof. For any $u \in V_t$ condition on all the nodes in V_{t-1} being infected. Then u might not be infected if it has fewer than k neighbors in V_{t-1}.

Applying Lemma 2.25 and using Properties 3 and 4 of power-law distribution we have,

$$\Pr[C(V_{t-1}, u) < k] \leq \exp\left(-\Theta\left(\frac{n^{1/(\alpha-1)} v_{t-1}^{\frac{\alpha-2}{\alpha-1}} \cdot (n/v_t)^{1/(\alpha-1)}}{n}\right)\right)$$

On the other hand, by taking $v_t = \frac{Cn}{\log^{\frac{\alpha-1}{3-\alpha}}(n^{(\alpha-2)^t})}$, $v_t \leq C\frac{n^{3-\alpha}}{\log^{\alpha-1} n} \cdot v_{t-1}^{\alpha-2}$, so

$$2\log n \leq 2C\frac{n^{1/(\alpha-1)} v_{t-1}^{\frac{\alpha-2}{\alpha-1}} \cdot (n/v_t)^{1/(\alpha-1)}}{n}.$$

Combining these two inequalities, for large enough constant C we have

$$\Pr[C(V_{t-1}, u) < k] \leq \frac{1}{n^2}$$

Moreover the number of bins is at most $T = O(\log\log n)$ because $N_1 = \frac{n}{\log^{\frac{\alpha-1}{3-\alpha}} n}$. □

Lemma 4.2. The proof is by induction. For the base case, Lemma 4.4 ensures every node in the first bin will be infected. Suppose all nodes before v_{t-1} are infected. We can use a union bound to show every node in V_t will be also infected. Moreover, in each bin the contagion only takes one time step which implies that the infection time for the first bucket is at most $O(\log\log n)$.

For the probability that all these events hold, we apply a union bound

$$\begin{aligned}
&\Pr[\text{all the nodes in } B_1 \text{ are infected}] \\
=&\Pr\left[\bigcap_{t=1}^{T} E_t\right] \\
\geq&1 - \Pr[\neg E_1] - \sum_{t=2}^{T} \Pr\left[\neg E_t \mid \bigcap_{s=1}^{t-1} E_s\right] && \text{(union bound)} \\
\geq&1 - n^{\frac{3-\alpha}{\alpha+1}} \exp\big(-\Theta(1) n^{\frac{3-\alpha}{\alpha+1}}\big) - \frac{1}{n^2}|B_1| && \text{(by Lemma 4.4 and 4.5)}
\end{aligned}$$

□

4.4 Proof of Lemma 4.3: Contagion from Buckets to Bucket

In this section we prove Lemma 4.3.

Lemma 4.3 (Induction)**.** *Fix sufficiently small $\epsilon > 0$, $\rho > 0$. Let $\ell < \rho L$, and suppose $\Pr[\mathcal{A}_\ell] > 0.5$, then we have*

$$\Pr[\mathcal{A}_{\ell+1} | \mathcal{A}_\ell] = 1 - O(1)\frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}}$$

Recall that $\mathcal{A}_\ell$ is the intersection of the three events, we will show that at stage ℓ if these three events happen, then the requirements in Lemma 4.3 will be met, and those events would be proven in Lemma 4.6, 4.7, 4.9 respectively.

4.4.1 First Event: Connection

We first prove that the first event holds with high probability, which follows almost immediately from Lemma 2.24 in Section 2.5.2.

Lemma 4.6. *Let* $\delta_n = \Theta\left(\frac{1}{\log^{\frac{\alpha-2}{3-\alpha}} n}\right)$, *if* $\Pr[\mathcal{A}_\ell] \geq 0.5$

$$\Pr\left[|X_{\ell+1} - \mathbb{E}[X_{\ell+1}]| \leq \delta_n \mathbb{E}[X_{\ell+1}] | \mathcal{A}_\ell\right] \geq 1 - 4\exp\left(-\Theta(\frac{n}{\log^{6\cdot\frac{\alpha-2}{3-\alpha}} n})\right).$$

Here the constant only depends on the product of δ_n *and* L.

Proof. Because $\delta_n \mathbb{E}[X_{\ell+1}] = \frac{\ell(L-\ell)b^2}{Lb-1} \geq \frac{\delta_n}{L} b = \Theta(\frac{b}{L^2})$, apply Lemma 2.24 we have

$$\begin{aligned}
&\Pr\left[(1-\delta_n)\mathbb{E}[X_{\ell+1}] \leq X_{\ell+1} \leq (1+\delta_n)\mathbb{E}[X_{\ell+1}]\right] \\
\geq& 1 - 2\exp(\frac{-(\delta_n \mathbb{E}[X_{\ell+1}])^2}{2m}) \\
\geq& 1 - 2\exp(-\frac{b^2}{m\, L^4}) \\
\geq& 1 - 2\exp(-\Theta(\frac{n}{\log^{6\cdot\frac{\alpha-2}{3-\alpha}} n})
\end{aligned}$$

For arbitrary events A, B where $\Pr[A] \geq 0.5$: $\Pr[B|A] \geq 1 - \Pr[\neg B]/\Pr[A] \geq 1 - 2\Pr[\neg B]$, and by taking $B = (1-\delta_n)\mathbb{E}[X_{\ell+1}] \leq X_{\ell+1} \leq (1+\delta_n)\mathbb{E}[X_{\ell+1}]$, and $A = \mathcal{A}_\ell$, we have

$$\Pr[(1-\delta_n)\mathbb{E}[X_{\ell+1}] \leq X_{\ell+1} \leq (1+\delta_n)\mathbb{E}[X_{\ell+1}] | \mathcal{A}_\ell] \geq 1 - 4\exp(-\Theta(\frac{n}{\log^{6\cdot\frac{\alpha-2}{3-\alpha}} n}))$$

□

4.4.2 Second Event: Number of infected nodes

Now we will prove the second events holds with high probability.

Lemma 4.7 (Number of uninfected nodes in a single bucket)**.** *For sufficiently small* $\epsilon > 0$, *conditioned on* $\mathcal{A}_\ell$

$$\Pr\left[\sum_{i \in B_\ell} \mathbb{I}_i \geq 2\mu_H \mid \mathcal{A}_\ell\right] \leq O(1)\frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}}$$

where $\mu_H = K\, \frac{|B_\ell| \ell^{\frac{3-\alpha}{\alpha-2}}}{\log n}$ *and* K *is independent of* ℓ *and* n.

The proof relies on an application of Chebyshev's inequality and the following Lemma, which is in turn proved using Chebyshev's inequality.

Lemma 4.8 (Infection of a single node)**.** *If $\mathcal{F}_\ell \subseteq \mathcal{A}_\ell$ for some constant $0 < \epsilon < 1/2$ and $\delta_n = \Theta(\frac{1}{\log^{\frac{\alpha-2}{3-\alpha}} n}) < 1/2$, then the probability any node $i \in B_\ell$ is not infected is*

$$\Pr[\mathbb{I}_i | \mathcal{A}_\ell] \leq O(1) \frac{\ell^{\frac{3-\alpha}{\alpha-2}}}{\log n}$$

where the constant $O(1)$ only depends on α, k, ρ if δ_n, ϵ is small enough, and $\rho \leq 0.3\frac{\alpha-1}{\alpha-2}k^{\frac{\alpha-2}{3-\alpha}}$.

The proof uses Chebyshev's inequality to show each node in bucket ℓ has a high chance of being infected when the fraction of uninfected stubs from $B_{<\ell}$, ϵ, is small.

Proof. Conditioning on $\mathcal{F}_\ell \subseteq \mathcal{A}_\ell$, so X_ℓ is known. For a node i in B_ℓ with degree d_i and s an infected stub from $B_{<\ell}$, we let $I_{i,s}$ be the indicator function for if the infected stub s connects to node i. $\Pr[I_{i,s} = 1 | \mathcal{F}_\ell] = \frac{d_i}{(L-\ell+1)b}$. Because the number of infected stubs s is at least $(1-\epsilon)X_\ell$, the expected number of infected neighbors from $B_{<\ell}$ of node i is

$$\mathbb{E}[\#\text{ neighbor of } i \text{ in } B_{<\ell} | \mathcal{A}_\ell] = \mathbb{E}[\sum_{s \in X_\ell} I_{i,s}] = (1-\epsilon_\ell)\frac{d_i X_\ell}{(L-\ell+1)b}$$

On the other hand, we have to argue the variance of $\sum_{s \in X_\ell} I_{i,s}$ is small. Because for all $s \neq s', \Pr[I_{i,s} = 1 | I_{i,s'} = 1] = \frac{d_i - 1}{(L-\ell+1)b-1} \leq \frac{d_i}{(L-\ell+1)b} = \Pr[I_{i,s} = 1]$, $\operatorname{Cov}(I_{i,s}, I_{i,s'}) < 0$, and $\operatorname{Var}(\sum_{s \in X_\ell} I_{i,s}) = \sum \operatorname{Var} I_{i,s} + \sum \operatorname{Cov}(I_{i,s}, I_{i,s'})$, we have

$$\operatorname{Var}(\sum_{s \in X_\ell} I_{i,s}) \leq \sum \operatorname{Var} I_{i,s} \leq \mathbb{E} \sum I_{i,s}.$$

Therefore by Chebyshev's inequality

$$\Pr[\sum_{s \in X_\ell} I_{i,s} < 1/2 \mathbb{E}[\sum_{s \in X_\ell} I_{i,s}] | \mathcal{A}_\ell] \leq \frac{4 \operatorname{Var} \sum_{s \in X_\ell} I_{i,s}}{\mathbb{E}[\sum_{s \in X_\ell} I_{i,s}]^2} \leq \frac{4}{\mathbb{E} \sum_{s \in X_\ell} I_{i,s}} \tag{4.1}$$

Because $\mathcal{F}_\ell \subseteq \mathcal{A}_\ell$, we have

$$\mathbb{E}[\sum_{s \in X_\ell} I_{i,s} | \mathcal{A}_\ell] \geq (1-\epsilon)(1-\delta_n)\frac{d_i(\ell-1)b}{Lb-1} \geq (1-\epsilon)(1-\delta_n)\frac{d_i \ell}{2L}$$

Because node i is in bucket ℓ, its degree d_i is higher than node N_ℓ where $d_{N_\ell} =$

$\big(\ell\frac{\alpha-2}{\alpha-1}\big)^{-1/(\alpha-2)}\log^{1/(3-\alpha)} n$, and the total number of bucket is $L=\frac{\alpha-1}{\alpha-2}\log^{\frac{\alpha-2}{3-\alpha}} n$

$$\mathbb{E}[\sum_{s\in X_\ell} I_{i,s}|\mathcal{A}_\ell] \geq \frac{(1-\epsilon)(1-\delta_n)}{2}\big(\frac{\alpha-1}{\ell(\alpha-2)}\big)^{\frac{3-\alpha}{\alpha-2}}\log n \tag{4.2}$$

Combining (4.1) and (4.2), if $k<0.5(1-\epsilon)(1-\delta_n)\big(\frac{\alpha-1}{\ell(\alpha-2)}\big)^{\frac{3-\alpha}{\alpha-2}}\log n$ for $\ell=O(L)$ we have the probability that i is not infected,equivalently the number of infected edges $\sum_{s\in X_\ell} I_{i,s}$ is smaller than k as follows

$$\begin{aligned}
&\Pr[\text{node } i \text{ is not infected}|\mathcal{A}_\ell] = \Pr[\sum_{s\in X_\ell} I_{i,s} < k] \\
&\leq \Pr[\sum_{s\in X_\ell} I_{i,s} < 1/2\mathbb{E}[\sum_{s\in X_\ell} I_{i,s}]] \\
&\leq \frac{8}{(1-\delta_n)(1-\epsilon)}\big(\frac{\ell(\alpha-2)}{\alpha-1}\big)^{\frac{3-\alpha}{\alpha-2}}\cdot\frac{1}{\log n} \\
&\leq O(1)\ \frac{\ell^{\frac{3-\alpha}{\alpha-2}}}{\log n}
\end{aligned}$$

Note that the constant $O(1)$ only depends on α if $\delta_n,\epsilon<1/2$. □

We can now proof Lemma 4.7 which states the total number of infection in bucket ℓ is high.

Proof of Lemma 4.7. Recall that $\mathbb{I}_i$ denote the node i not being infected, and $\mathbb{I}_i=1$ if and only if the number of infected edges is smaller than k, $\sum_{s\in X_\ell} I_{i,s}<k$ Applying Lemma 4.8, we have

$$\mathbb{E}[\sum_{i\in B_\ell}\mathbb{I}_i|\mathcal{A}_\ell] \leq O(1)\ \frac{|B_\ell|\ell^{\frac{3-\alpha}{\alpha-2}}}{\log n}$$

We will use μ_H to denote $O(1)\ \frac{|B_\ell|\ell^{\frac{3-\alpha}{\alpha-2}}}{\log n}$

As in the proof of Lemma 4.8, in order to use the second moment method, it is sufficient to show the variance of $\sum\mathbb{I}_i$ is small. For all $j\neq i, \Pr[\mathbb{I}_j=1|\mathbb{I}_i=1]\leq \Pr[\mathbb{I}_j=1|\mathbb{I}_i=0]$, because whether we condition on $\mathbb{I}_i=1$ or $\mathbb{I}_i=0$, the revealing of the d_j edges of node j will be among the same number of random matching except in the later case the number of infected free stubs from is strictly less than the former. Therefore,

$$\Pr[\sum_{i\in B_\ell}\mathbb{I}_i \geq 2\mu_H] \leq \frac{1}{\mu_H}$$

There is a subtle issue that we need to use an upper bound of expectation, μ_H, in

second moment method instead of the original expectation, $\mathbb{E}\sum_{i\in B_\ell}\mathbb{I}_i$ because the expectation itself is too hard to compute. To bound the probability of this bad event $\sum_{i\in B_\ell}\mathbb{I}_i \geq 2\mathbb{E}\sum_{i\in B_\ell}\mathbb{I}_i$, we need to estimate μ_H which needs an upper bound for $|B_\ell|$

$$N_\ell - N_{\ell-1} = O(1)\frac{n}{(\log n)^{\frac{\alpha-1}{3-\alpha}}}(\ell^{\frac{\alpha-3}{\alpha-2}} - (\ell-1)^{\frac{\alpha-3}{\alpha-2}}) \leq O(1)\frac{n\ell^{1/(\alpha-2)}}{(\log n)^{\frac{\alpha-1}{3-\alpha}}}$$

The last inequality comes from the convexity.

Finally we have

$$\Pr[\sum_{i\in B_\ell}\mathbb{I}_i \geq 2\mu_H] \leq O(1)\frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}}$$

□

4.4.3 Third Event: Number of uninfected stubs

Lemma 4.9. *Suppose A_ℓ, the first event, $(1-\delta_n)\mathbb{E}[X_{\ell+1}] \leq X_{\ell+1} \leq (1+\delta_n)\mathbb{E}[X_{\ell+1}]$ and the second event, $\sum_{i\in B_\ell}\mathbb{I}_i \leq 2\mu_H$ is true (this is the conclusion of Lemma 4.7), then*

$$\Pr\left[Y_{\ell+1} \leq \epsilon X_{\ell+1}\middle| |X_{\ell+1} - \mathbb{E}[X_{\ell+1}]| \leq \delta_n\mathbb{E}[X_{\ell+1}] \wedge \sum_{i\in B_\ell}\mathbb{I}_i \leq 2\mu_H \wedge \mathcal{A}_\ell\right]$$

is greater than $1-\exp\left(-\Theta(\frac{n}{\log^{5\cdot\frac{\alpha-2}{3-\alpha}}})\right)$ when $\rho > 0$ is small enough and $\delta_n > 0$ is smaller than some constant.

For the third event, in Lemma 4.9 we want to argue the *fraction* of uninfected stubs is smaller then ϵ after stage ℓ. That requires both that $X_{\ell+1}$ is large and that $Y_{\ell+1}$—which is the summation of $Y_{\ell+1}^{(1)}$ and $Y_{\ell+1}^{(2)}$—is small. Upper bounds on $Y_{\ell+1}^{(1)}$ and $Y_{\ell+1}^{(2)}$ will be proven by Lemma 4.10 and 4.11 respectively.

Lemma 4.10. *Let $Y_\ell^{(1)}$ be the number of free uninfected stubs from $B_{<\ell}$ to $B_{>\ell}$ over the probability space $\mathcal{F}_{\ell+1}|\mathcal{F}_\ell$, then*

$$\Pr\left[Y_{\ell+1}^{(1)} \geq (1+\delta_n)\epsilon X_\ell|\mathcal{A}_\ell\right] \leq \exp\left(-\Theta(\frac{n}{\log^{5\cdot\frac{\alpha-2}{3-\alpha}}})\right)$$

Here the constant only depends on $\delta_n \cdot L$, ϵ and ρ

Proof. Given $\mathcal{F}_\ell \subseteq \mathcal{A}_\ell$, the uninfected free stubs from $B_{<\ell}$ to $B_{>\ell}$ will still around the

same portion after stage ℓ revealing the connection from B_ℓ to previous nodes in $B_{<\ell}$, and the proof is similar to Lemma 2.24 by using concentration of Doob martingale.

The expectation of $Y_{\ell+1}^{(1)}$is

$$\mathbb{E}[Y_{\ell+1}^{(1)}|\mathcal{F}_\ell] = \frac{L-\ell}{L-\ell+1}\epsilon_\ell X_\ell \leq \epsilon(1+\delta_n)\frac{(\ell-1)(L-\ell)}{L}b$$

Therefore applying Azuma's inequality 2.18 we have

$$\begin{aligned}\Pr[Y_{\ell+1}^{(1)} \geq (1+\delta_n)\epsilon_\ell X_\ell|\mathcal{F}_\ell] &\leq \exp(-\frac{1}{2b}(\frac{\delta_n(1+\delta_n)\epsilon(\ell-1)(L-\ell)}{L}b)^2) \\ &= \exp(-\Theta(\frac{\delta_n^2 n}{\log^{3\cdot\frac{\alpha-2}{3-\alpha}}})) = \exp(-\Theta(\frac{n}{\log^{5\cdot\frac{\alpha-2}{3-\alpha}}}))\end{aligned}$$

□

Lemma 4.11. *Suppose A_ℓ and the $\sum_{i\in B_\ell}\mathbb{I}_i \leq 2\mu_H$ is true (this is the conclusion of Lemma 4.7), then $Y_{\ell+1}^{(2)}$, the total number of uninfected stubs from B_ℓ to $B_{>\ell}$ is*

$$Y_{\ell+1}^{(2)} = O(1)\frac{(\log n)^{\frac{\alpha}{3-\alpha}}}{n\ell^{2/(\alpha-2)}}$$

Proof. Given Lemma 4.7, the number of uninfected nodes in B_ℓ is small

$$O(1)\frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}},$$

and the degree of each node in B_ℓ is smaller than

$$d_{N_{\ell-1}} = O(1)\frac{(\log n)^{1/(3-\alpha)}}{\ell^{1/(\alpha-2)}}$$

Therefore the total number of uninfected stubs from B_ℓ to $B_{>\ell}$ is smaller than

$$O(1)\frac{(\log n)^{\frac{\alpha}{3-\alpha}}}{n\ell^{2/(\alpha-2)}}$$

□

Proof of Lemma 4.9. By Lemma 4.10 and 4.11, we have

$$Y_{\ell+1} = Y_{\ell+1}^{(1)} + Y_{\ell+1}^{(2)} \leq \epsilon(1+\delta_n)^2\frac{(\ell-1)(L-\ell)}{L}b + O(1)\frac{(\log n)^{\frac{\alpha}{3-\alpha}}}{n\ell^{2/(\alpha-2)}} \tag{4.3}$$

Suppose $L' = \rho L$ is small enough such that $(1+\delta_n)^2(\ell-1) < (1-\delta_n)\ell$ by taking $\rho < \frac{1}{4L\delta_n} = \Theta(1)$. We have $1 - \frac{1}{\ell} \leq 1 - \frac{1}{\rho L} \leq 1 - 4\delta_n$ and

$$1 - \delta_n - (1+\delta_n)^2(1 - \frac{1}{\ell}) = \delta_n + O(\delta_n^2) \tag{4.4}$$

By the assumption the first event we have

$$X_{\ell+1} \geq (1-\delta_n)\mathbb{E}X_{\ell+1} \geq \frac{\ell(L-\ell)}{L} b$$

Let $\epsilon X_{\ell+1}$ subtracted by $Y_{\ell+1}$ and apply (4.3) and (4.4) we have

$$\epsilon X_{\ell+1} - Y_{\ell+1} \geq \frac{\epsilon(L-\ell)\ell}{L} b\big(1 - \delta_n - (1+\delta_n)^2(1-\frac{1}{\ell})\big) - O(1)\frac{\log^{\frac{\alpha}{3-\alpha}} n}{n\ell^{\frac{2}{\alpha-2}}} \geq 0$$

Therefore we have $Y_{\ell+1} \leq \epsilon X_{\ell+1}$, and the probability is

$$\begin{aligned}
&\Pr\left[Y_{\ell+1} \leq \epsilon X_{\ell+1} \mid (1-\delta_n)\mathbb{E}[X_{\ell+1}] \leq X_{\ell+1} \leq (1+\delta_n)\mathbb{E}[X_{\ell+1}] \wedge \sum_{i \in B_\ell} \mathbb{I}_i \leq 2\mu_H\right] \\
\geq &1 - \exp\left(-\Theta\left(\frac{n}{\log^{5\cdot\frac{\alpha-2}{3-\alpha}}}\right)\right)
\end{aligned}$$

□

4.4.4 Proof of Lemma 4.3

Proof. Recall the the event $\mathcal{A}_{\ell+1}$ is the intersection of the three events, so

$$\Pr[\mathcal{A}_{\ell+1}|\mathcal{A}_\ell] \geq 1 - \Pr[\neg(|X_\ell - \mathbb{E}[X_\ell]| \leq \delta_n \mathbb{E}[X_\ell])|\mathcal{A}_\ell] \tag{4.5}$$

$$- \Pr\left[\sum_{i \in B_{\ell-1}} \mathbb{I}_i \geq 2\mu_H | \mathcal{A}_\ell\right] \tag{4.6}$$

$$- \Pr\left[Y_\ell \leq \epsilon X_\ell | (|X_\ell - \mathbb{E}[X_\ell]| \leq \delta_n \mathbb{E}[X_\ell]) \wedge \sum_{i \in B_{\ell-1}} \mathbb{I}_i \leq 2\mu_H \wedge \mathcal{A}_\ell\right] \tag{4.7}$$

Applying Lemma 4.6 to Equation (4.5), Lemma 4.7 to Equation (4.6), and Lemma 4.9 to Equation (4.7), and we have

$$\Pr[\mathcal{A}_{\ell+1}|\mathcal{A}_\ell] \geq 1 - 4\exp\left(-\Theta(\frac{n}{\log^{6\cdot\frac{\alpha-2}{3-\alpha}} n})\right) - O(1)\frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}} - \exp\left(-\Theta(\frac{n}{\log^{5\cdot\frac{\alpha-2}{3-\alpha}}})\right)$$

Therefore

$$\Pr[\mathcal{A}_{\ell+1}|\mathcal{A}_\ell] \geq 1 - O(1)\frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}}.$$

□

4.5 Infection with Random Seeds

Theorem 4.1 together with prior results in Ebrahimi et al. [82] immediately implies the following corollary:

Corollary 4.12. *For a configuration model graph with power-law exponent α, if $\Omega(n^{\frac{\alpha-2}{\alpha-1}})$ initially random seeds are chosen, then with probability $1 - o(1)$ k-complex contagion infects a constant fraction of nodes.*

We first restate two results from Ebrahimi et al. [82].

Proposition 4.13 ([82])**.** *For any graph, let u be a node with degree d. If $\Omega(d/n)$ initial random seeds are chosen, then with probability $1 - o(1)$ u is infected after one round.*

Proof. The initial node has $\Theta(n^{\frac{1}{\alpha-1}})$ neighbors. If there are $\Omega(n^{\frac{\alpha-2}{\alpha-1}})$ initial seeds then by Proposition 4.13 the first seed is infected with probability $1 - o(1)$. However, then by Theorem 4.1 a constant fraction of the remaining nodes are infected in $\log^{O(\alpha)}(1)$ rounds.

□

This is tight as in Ebrahimi et al. [82] the following was proven:

Proposition 4.14 ([82])**.** *For any graph, with power law distribution α, if $o(n^{\frac{\alpha-2}{\alpha-1}})$ initially random seeds are chosen, then with probability $1 - o(1)$, no additional nodes are infected.*

4.6 Power-law exponent $\alpha > 3$

For the case of power-law degree distribution with $\alpha > 3$, Amini [12] shows how to analyze k-complex contagions using a differential equation method [239]. This approach heavily depends on the variance of the degree distribution and fails when $\alpha < 3$. For the case where the seed set contains all nodes with degree greater than $\rho > 0$ we can state their theorem as follows:

Theorem 4.15 ([12])**.** *Given a power law distribution* $\mathbf{d}$ *with exponent* $\alpha > 3$ *and* $d_1 < n^{1/\alpha - 1}$*, the* k*-complex contagion on configuration* $CM(\mathbf{d})$ *with constant* k *and seed set* $I_\rho = \{i | d_i \geq \rho\}$ *where* $0 \leq \rho \leq n$*. Then with high probability*

$$|CC(CM(\mathbf{d}), k, I_\rho)| = n \left(1 - \sum_{\substack{1 \leq d < \rho, \\ 0 \leq j < k}} p_{\mathbf{d}}(d) \binom{d}{j} (y^*)^{d-j} (1 - y^*)^j + o(1) \right) \tag{4.8}$$

where $p_{\mathbf{d}}(d) = (F_{\mathbf{d}}(d+1) - F_{\mathbf{d}}(d))$ *and* $0 < y^* \leq 1$ *is the largest root such that* $f(y) = 0$ *and*

$$f(y) = y^2 \left(\sum_{1 \leq d} d\, p_{\mathbf{d}}(d) \right) - y \left(\sum_{\substack{1 \leq d < \rho, \\ 0 \leq j < k}} d\, p_{\mathbf{d}}(d) \binom{d-1}{j} y^{d-1-j} (1-y)^j \right) \tag{4.9}$$

Before stating our corollary, we wish to give a brief idea of the proof of Theorem 4.15. They consider a Markov chain which results in the same number infected nodes as a k-complex contagion, but proceeds using the randomness of the configuration model. The Markov chain starts with the initially infected nodes and at each step the process reveals one of the unmatched edges from the set of infected nodes. This process needs only track: the number of unmatched edges, and the number of d-degree uninfected nodes with j infected neighbors, for each $j < k$. The Markov chain stops when all the agent are infected, or there are no unmatched edges from already infected nodes. As an example, if the revealed edge connects to an uninfected node v with $k-1$ infected neighbors, the node v will become infected and the number of unmatched edges will increase by $d_v - k$. It turns out, that if $\alpha > 3$, the process is smooth and we can use the corresponding differential equations to approximate this Markov chain and derive the fraction of infections.

With their results we can prove that to infect a constant fraction of nodes requires the initial seed need to also be a constant fraction of nodes. Note that if our initial

seed set infects the highest degree nodes, but does not infect a constant fraction of the nodes, then the greatest degree node not in the initially infected set has degree $\omega(1)$.

Corollary 4.16. *Given a power law distribution* $\mathbf{d}$ *with exponent* $\alpha > 3$ *and* $d_1 < n^{1/\alpha-1}$*, the* k*-complex contagion on configuration* $CM(\mathbf{d})$ *with constant* k *and seed set* $I_\rho = \{i | d_i \geq \rho\}$ *where* $\rho = \omega(1)$*, the* $|CC(CM(\mathbf{d}), k, I_\rho)| = o(n)$ *with high probability.*

Proof. Note that $\sum_{0\leq j<k} \binom{d}{j}(y^*)^{d-j}(1-y^*)^j$ is the probability of a binomial random variable $Bin(d, \epsilon)$ smaller than k where $\epsilon = 1 - y^*$. As ϵ increases, the number of infection increases. As a result, the proof has two parts: $\epsilon = 1 - y^* = o(1)$ and $\epsilon = o(1) \Rightarrow |CC(CM(\mathbf{d}), k, I_\rho)| = o(n)$.

$1 - y^* = o(1) \Rightarrow |CC(CM(\mathbf{d}), k, I_\rho)| = o(n)$ Rewrite the contagion fraction (4.8)

$$\begin{aligned}
|CC(CM(\mathbf{d}), k, I_\rho)| &= n\left(1 - \sum_{\substack{1\leq d<\rho,\\ 0\leq j<k}} p_{\mathbf{d}}(d)\binom{d}{j}(y^*)^{d-j}(1-y^*)^j + o(1)\right) \\
&= n\left(\sum_{1\leq d} p_{\mathbf{d}}(d) - \sum_{1\leq d<\rho} p_{\mathbf{d}}(d)\Pr[Bin(d,\epsilon) < k] + o(1)\right) \\
&= n\left(\sum_{\rho\leq d} p_{\mathbf{d}}(d) + \sum_{1\leq d<\rho} p_{\mathbf{d}}(d)\Pr[Bin(d,\epsilon) \geq k] + o(1)\right) \\
&= |I_\rho| + n\left(\sum_{1\leq d<\rho} p_{\mathbf{d}}(d)\Pr[Bin(d,\epsilon) \geq k] + o(1)\right)
\end{aligned}$$

Therefore besides the initial seed $|I_\rho| = o(n)$ the number of additional infected node is

$$\begin{aligned}
& n\left(\sum_{1\leq d<\rho} p_{\mathbf{d}}(d)\Pr[Bin(d,\epsilon) \geq k] + o(1)\right) \\
\leq & n\sum_{1\leq d<\rho} p_{\mathbf{d}}(d)\frac{d\epsilon}{k} + o(n) && \text{(Markov inequality)} \\
\leq & O\left(\frac{n\epsilon}{k}\sum_{1\leq d<\rho}\frac{1}{d^{\alpha-1}}\right) + o(n) \\
\leq & O(\epsilon n) = o(n)
\end{aligned}$$

Therefore the second part is proven.

$\epsilon = 1 - y^* = o(1)$ By definition $y^* \leq 1$ is the largest root for (4.9), because f is continuous and $f(1) > 0$, if there exists $\xi = o(1)$ such that $f(1-\xi) < 0$, we have $1 - y^* = o(1)$ by intermediate value theorem.

Let $\xi = K\frac{1}{\rho^{\alpha-2}}$ for some constant $K > 0$ which will be defined later.

$$
\begin{aligned}
& f(1-\xi) \\
&= (1-\xi)^2 \left(\sum_{1 \leq d} d\, p_{\mathbf{d}}(d) \right) - (1-\xi) \left(\sum_{\substack{1 \leq d < \rho, \\ 0 \leq j < k}} d\, p_{\mathbf{d}}(d) \binom{d-1}{j} (1-\xi)^{d-1-j} (\xi)^j \right) \\
&= (1-\xi) \left((1-\xi) \sum_{1 \leq d} d\, p_{\mathbf{d}}(d) - \sum_{1 \leq d < \rho} d\, p_{\mathbf{d}}(d) \Pr[Bin(d-1, \xi) < k] \right) \\
&= (1-\xi) \left(\sum_{\rho \leq d} d\, p_{\mathbf{d}}(d) + \sum_{1 \leq d < \rho} d\, p_{\mathbf{d}}(d) \Pr[Bin(d-1, \xi) \geq k] - \xi \sum_{1 \leq d} d\, p_{\mathbf{d}}(d) \right)
\end{aligned}
$$

To have a better estimate of $\Pr[Bin(d-1, \xi) \geq k]$ we can use multiplicative form of Chernoff bound

$$
\begin{aligned}
& f(1-\xi) \\
&\leq (1-\xi) \left(\frac{S_{\mathbf{d}}(\rho)}{n} + \sum_{1 \leq d < \rho} d\, p_{\mathbf{d}}(d) \exp(-O(\frac{k^2}{d\xi})) - \xi \frac{S_{\mathbf{d}}(1)}{n} \right) && \text{(chernoff bound)} \\
&\leq (1-\xi) \left(\frac{S_{\mathbf{d}}(\rho)}{n} + \sum_{1 \leq d < \rho} d\, p_{\mathbf{d}}(d) O((\frac{d\xi}{k^2})^2) - \xi \frac{S_{\mathbf{d}}(1)}{n} \right) \\
&\leq (1-\xi) \left(\frac{C_1}{\rho^{\alpha-2}} + \frac{C_2 \rho \xi^2}{k^4} - C_3 \xi \right) && \text{(where } C_1, C_2, C_3 \text{ are positive constants)} \\
&\leq (1-\xi) \left(\frac{C_1}{\rho^{\alpha-2}} - C_3' \xi \right) && (\rho\xi = o(1)) \\
&\leq (1-\xi) \left(\frac{C_1 - K C_3'}{\rho^{\alpha-2}} \right) < 0 && \text{(by taking } K > C1/C_3')
\end{aligned}
$$

□

CHAPTER 5

General Preferential Attachment Models

In this chapter we study a more general contagions model—the *general threshold model* which is parameterized by a distribution over the natural numbers, the collective influence from infected neighbors of an individual u, once beyond the threshold, will trigger the infection of u. By varying the choice of the distribution, the general threshold model can model cascades with or without the submodular property.

We provide both analytical and experimental results for how cascades from a general threshold model spread in a general growing network model which contains preferential attachment models as special cases. We show that if we choose the initial seeds as the early arriving nodes, the contagion can spread to a good fraction of the network and this fraction crucially depends on the fixed points of a function derived only from the specified distribution. We also show that, using a coauthorship network derived from DBLP databases and the Stanford web network, our theoretical results can be used to predict the infection rate up to a decent degree of accuracy, while the configuration models does the job poorly.

5.1 Introduction

Social Contagion Models. The *general threshold model* [104, 172] is a fairly general model to capture such intuition. Each node v has a monotone function $g_v : \{0,1\}^{|\Gamma(v)|} \to [0,1]$, where $\Gamma(v)$ indicates the set of v's neighbors in a social network. The function g_v represents how much influence (via knowledge, social pressure, etc) any set of neighbors has upon node v. In the general threshold model, each node also has threshold R_v drawn uniformly and independently from the interval $[0,1]$. After an initial seed set is infected, a node v becomes infected if $g_v(S) \geq R_v$ where S is the set of infected neighbors of v.

The general threshold model captures many other models as special cases. For

example, a special case is the *linear threshold model*, in which each edge (u, v) has an influence weight $w(u, v)$, and the function g_v is then the sum of the influence from all infected neighbors of v. [1] Another example of a class of cascades captured by the general threshold model is the independent cascade model [102]. In this model, there is some fixed parameter ρ, and each infected node has one chance to infect each uninfected neighbor node with probability ρ (iid).

We call contagions *simple* when the influence g_v is submodular—that is $g_v(S' \cup \{x\}) - g_v(S') \leq g_v(S \cup \{x\}) - g_v(S)$, if $S \subseteq S'$—and call contagions *complex* when this fails to hold (e.g., contagions that require infection from multiple neighbors). In a simple contagion, the effect of an additionally infected neighbor is marginally decreasing. In a complex contagion, there could be an initial barrier such that no activation is possible until the barrier is crossed. There can be synergy between neighbors such that the total influence from them is not just a simple sum. If we define $f(S)$ as the expected number of infected nodes when the vertices in S are chosen as the initial seeds, then if g_v is submodular for all nodes, then f is submodular as well [172].

The monotonicity and submodularity have greatly helped with the analysis of the diffusion behavior with respect to the choice of seeds. In particular, one can apply the greedy set cover algorithm to choose the set of k best seeds to maximize the final scope of the contagion. This will give a $1 - 1/e$ approximation to the maximum scope obtained by any k seeds. In contrast, for the general threshold model, this is a very hard question and not much is known in the literature other than that is NP-hard to even approximate [128]. The two special cases, the linear threshold model and the independent cascade model, have received a lot of attention because they both have the submodular property [128].

While this result has been well recognized and celebrated, a natural question one may ask is whether the submodularity assumption holds in reality and whether the result can be generalized. Sociologists observe that in the case of the adoption of pricey technology innovations, the change of social behaviors, and the decision to participate in a migration, etc [61, 156], an additional confirmation is crucial, suggesting the model of complex contagion. In practice, threshold distributions are usually computed from data of contagions by using the empirical fraction of agents who adopt directly after r ties adopt, given that they had not previously. The distributions found depend on which cascades are analyzed, however, this conditional probability

[1] Often an additional restriction is imposed that for all nodes v: $\sum_{u \in \gamma(v)} w(u, v) \leq 1$ to ensure that g_v is always in $[0, 1]$.

typically increases with r until some small constant of at least 2, and then tapers off.Examples include LiveJournal [18], DBLP [18], Twitter [199], and Facebook [229]. Some of these data sets indeed show diminishing return of the influence function, but others do not. They find that the second affected neighbor often has more marginal effect than the first. Additionally, the study in the Facebook data set shows that the number of *connected components* in the infected neighbors is a much better predictor on the probability of joining Facebook, compared to the number of infected neighbors.

Work done on complex contagions is much more limited and so far focused on a simplistic single threshold model called r-complex contagions. In r-complex contagions, all nodes have the same threshold r. A node becomes infected if and only if at least r of its neighbors have been infected. It has been shown that a r-complex contagion is generally slower and more delicate than simple contagion $r = 1$ [99, 83, 82]. One of the limitations of this r-complex contagion model is the dependency on the fixed threshold r for all nodes in the network. In practice there are people who like to try out new things and are more risk driven while others are risk averse. Therefore the threshold function is not necessarily uniform.

In this paper we consider one step of generalizing the r-complex contagion model by considering the threshold coming from a distribution $\mathbb{R}$ on positive intergers. The initial adoption barrier can still exist which makes the adoption function to be non-submodular. We provide analysis on the spreading behaviors on a general family of networks that grow over time.

Stochastic Attachment Network Model. In addition to a model of cascade, the model of network is also important. A lot of mathematical models have been developed to capture some of the attributes of real world social networks. A celebrated set of results are the family of small world graphs [179, 234, 133, 235, 134] and the family of graphs that produce power law degree distribution [24, 135, 138, 139].

In this work we examine a growing network in which newcomers connect stochastically to nodes already in the network. This family of networks, which we call the *stochastic attachment network* model, has the preferential attachment network model as a special case. In the preferential attachment models [24], nodes arrive in a sequential order. Each node chooses k edges from the nodes that arrive earlier. When an edge is added, the neighbor is selected with probability proportional to its current degree. This model generates graphs with a power law degree distribution and has been used to explain the observations in web graphs and social networks. We examine a more general model in which new edges are not necessarily preferentially attached to existing nodes and each newcomer may have a varying number of edges. The key

feature that is used in our analysis is that the network is formed over time, when new nodes arrive sequentially and attach to existing nodes.

We study contagions on both directed and undirected version of the stochastic attachment network. In the first case, we consider each edge issued by a newcomer u as directional, pointing to an earlier node v. This edge can be interpreted as u following edge v. A social contagion spreads in the *reverse* direction of an edge. This models information spreading in Twitter-type social networks, in which messages or information only travels along the direction of the edges. A node u will be influenced only by the neighbors u follows and not the neighbors that follow u. In the second case, all edges are treated as undirected, allowing contagions to spread in both directions. For example, consider a co-authorship network in which a new researcher choose to work with senior researchers/advisors, but here information or social influence is bidirectional.

An additional consideration is where the initially infected nodes in the contagions reside within the network structure. In this paper we consider the scenario when some entity is trying to initiate a cascade. The entity is allowed to choose where the nodes go. We model this case by letting the seed equal the first nodes (in arrival order) or a subset of these nodes.

In our earlier work we show that due to the evolutionary nature r-complex contagions spread to the entire network in preferential attachment models and the contagion spreads very fast [82], when $r < k$ and the first few nodes in the arriving order are selected as the initial seeds in both the directed and undirected cases. This paper provides significant generalizations in both models of contagions and models of networks. The proof ideas are also completely new.

Our Results. In this paper we study the behavior of a contagion following a general threshold model on both directed and undirected stochastic attachment graphs. We provide the most detailed analysis in the case of preferential attachment and later generalize to other scenarios.

We show that the number of infected nodes on preferential attachment model depends critically on the threshold distribution $\mathbb{R}$ and the number of edges a new node connect to existing nodes k. In the directed case of preferential attachment model, we derive a function $f_{k,\mathcal{R}} : [0,1] \to [0,1]$ describing the probability of the i-th arriving node being infected, which depends only on a single number summarizing the status of the nodes with earlier arriving order, i.e., their threshold and whether they are infected or not. This function $f_{k,\mathcal{R}}$ has fixed points, which may be either stable or unstable. The ratio of the infected nodes in the network converges to one of these

stable fixed points with high probability. When there are multiple fixed stable points, the contagion may converge to any one of them with at least constant probability.

In the undirected case of preferential attachment model, we note that the number of infected nodes will be no fewer than the directed case, since the edges can possibly spread social influence both ways. However, we show something much stronger than this, that, with high probability, the total number of infected nodes will always be a constant fraction higher than the highest stable fixed point of function $f_{k,\mathcal{R}}$, when non-zero stable fixed points exist.

We performed both simulations and experiments with real world data sets. On various stochastic attachment graphs we observe the same behaviors as predicted in theory. We also tested real world networks. We used two datasets, the coauthorship network derived from DBLP database which is an undirected graph and the Stanford web graph (which is naturally directed). On both datasets we infer the arriving order by using k-core decomposition – i.e., removing nodes with degree k for k starting from 1 recursively. We show that using the stochastic attachment model one can get fairly accurate prediction of the contagion rate. On the other hand, if we use the same degree distribution and generate a graph using the configuration model[2], the contagion behaviors differ significantly from that of the real netwrok. These experiments confirm the validity and utility of our model and analysis in helping to understand and predict contagions on real world graphs.

In this paper we consider the edges are considered directed or undirected. In the directed case, each edge is issued by a node s and points towards a node t earlier in the arriving order. We consider this as s following t. Thus the contagion propagates in the reversed direction of edge (s, t). A node s is infected if the number of infected nodes that t follows is greater than its threshold R_s. In the undirected case, the edges are undirected and infection can happen in both directions.

5.1.1 Preliminary

Definition 5.1 (contagions model)**.** Give a directed/undirected graph $G = (V, E)$, a set of initially infected nodes $X \subseteq V$, and a distribution over non negative integer $\mathbb{R}$, a ***General Threshold Contagion*** $GTC(G, \mathbb{R}, X)$ is a process $(X_t)_{t\geq 0}$ over subset of nodes. Initially, each node v samples a threshold R_v from the distribution $\mathbb{R}$ independently, and $X_0 = X$. The contagion proceeds in rounds. At each round $t + 1$, each node v with at least threshold number of, R_v, infected out-neighbors/neighbors

[2]In a configuration model we fix the degree distribution first and then match the half edges at the nodes randomly.

becomes infected,

$$X_{t+1} = X_t \cup \{v \in V : v \text{ has } R_v \text{ infected out neighbor}\}.$$

Note that nodes cannot be uninfected and the process is monotone the process reaches its limit in $|V|$ rounds. In this work, we are interested in the span of infection, i.e. what is the fraction of infected nodes at the end, which can be written as $|X_\infty|/|V|$ where $X_\infty \subseteq V$ is the set of nodes eventually infected.

In this paper we mostly consider *preferential attachment model* (Definition 2.51), and the edges are considered directed or undirected. In the directed case, each edge is issued by a node s and points towards a node t earlier in the arriving order. We consider this as s following t. Thus the contagion propagates in the reversed direction of edge (s, t). A node s is infected if the number of infected nodes that t follows is greater than its threshold R_s. In the undirected case, the edges are undirected and infection can happen in both directions.

5.2 Directed preferential attachment models

In this section, we analyze the general threshold contagion models $GTC(G, \mathbb{R}, X)$ on a directed preferential attachment graph, $G \sim \mathrm{PA}_k(n)$. In this paper the initial seeds are chosen as a fraction (or all) of the first few nodes.

Instead of considering the sampling of graph and contagions separately, we couple these two processes together: Each node is only evaluated once. The first k nodes are the initial seeds X. We start at node of index $k+1$ and process each of the following nodes in their order of arrival in the graph. When a node is being processed we reveal both its threshold and its outgoing edges, and based on its threshold and the status (being infected or not) of its outgoing edges, it is determined if the current node will be infected or not. To evaluate this probability we give some definitions.

Assume that node t is the t-th node in the arrival order in G. Let V_{t-1} be the set of first $t-1$ nodes in G and X_{t-1} be the set of infected nodes in V_{t-1}. If t's threshold is $R_t = r$, t is infected if and only if among the k edges t issues, at least r of them land in nodes in X_{t-1}. Now give a specific edge of t, we define Y_t as the probability that this edge lands in an infected node (e.g., in X_{t-1}). Under preferential rule, Y_t is the *ratio of the infected degree* $Y_t = \sum_{s \in X_{t-1}} \deg(s) / \sum_{u \in V_{t-1}} \deg(u)$. Recall that $\deg(s) = \deg(s; t)$ is the total degree of each node s (counting both incoming and outgoing edges).

Next we can compute the probability of node t being infected when its threshold is $R_t = r$. For that to happen, among the k (out) edges of t, at least r of them need to land on a node in X_{t-1}.

$$\Pr[t \text{ is infected} \mid R_t = r] = \sum_{\ell=r}^{k} \binom{k}{\ell} Y_t^{\ell} (1 - Y_t)^{(k-\ell)} \tag{5.1}$$

Now, the probability of infection of node t can be described by a function $f_{k,\mathbb{R}}$:

$$f_{k,\mathcal{R}}(y) = \sum_{r} \Pr_{\mathbb{R}}(r) \sum_{\ell=r}^{k} \binom{k}{\ell} y^{\ell} (1 - y)^{(k-\ell)}, \tag{5.2}$$

such that $\Pr[t \text{ is infected}] = f_{k,\mathcal{R}}(Y_t)$. Thus, the random process (Y_t) for $t = k+1, ..., n$ is a Markov chain. Moreover $f_{k,\mathcal{R}}$ is a non-decreasing polynomial function from $[0, 1]$ to $[0, 1]$. (c.f. appendix B.1). By Brower's fixed point theorem $f_{k,\mathcal{R}}$ has fixed points in $[0, 1]$. We will show that the behavior of the contagion depends crucially on the *fixed points* of this function $f_{k,\mathcal{R}}$.

5.2.1 Main Results for Directed Preferential Attachment

Now we state the main theorem that characterizes the behavior of general threshold contagion on *preferential attachment* graphs $\mathrm{PA}_k(n)$.

Theorem 5.2. *Given natural numbers n and k, a distribution $\mathbb{R}$ of threshold, and a set of initially infected nodes $X \subset [k]$, the ratio of infected degree Y_n of the general threshold contagion $GTC(\mathrm{PA}_k(n), \mathbb{R}, X)$ depends on the values of the fixed points of function $f_{k,\mathcal{R}}$ in Equation (5.2) as follows:*

1. *If $f_{k,\mathcal{R}}(x)$ has a unique fixed point y^* which is stable, Y_n converges to y^* as n large enough.*

   (a) *For all $\delta > 0$ and $\epsilon > 0$, there exists $\tau_0 = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, such that for all $n \geq \tau_0$*

   $$\Pr[|Y_n - y^*| > \epsilon] \leq \delta$$

   (b) *If $f_{k,\mathcal{R}}{}'(y^*) < 1$, then for all γ, $0 < \gamma < (1 - f_{k,\mathcal{R}}{}'(y^*))/2$, we have*

   $$|Y_n - y^*| = O_P(n^{-\gamma}).$$

That is for any $\delta > 0$ there exists $M > 0$ and τ_0 such that for all $n \geq \tau_0$, we have $\Pr[|Y_n - y^*| > Mn^{-\gamma}] < \delta$.

2. *If $f_{k,\mathcal{R}}(x)$ has a finite number of fixed points, then $\lim_{n\to\infty} Y_n$ exists almost surely, and*
$$\Pr\left[\lim_{n\to\infty} Y_n \in S_{f_{k,\mathcal{R}}} \cup T_{f_{k,\mathcal{R}}}\right] = 1.$$

3. *If $f_{k,\mathcal{R}}(x)$ has an infinite number of fixed points, the process $\{Y_t\}$ is a martingale and converges almost surely to some random variable Y.*

5.2.2 Proof outline of Theorem 5.2

Let's first understand the fixed points of function $f_{k,\mathcal{R}}$ and the recursive structure for Y_t, i.e., the probability for a specific edge from the t-th arriving node landing on an infected node.

Assume that t nodes have arrived and picked their edges. Given information $\mathcal{F}_t$ at time t which consists of the subgraph $\mathrm{PA}_k(t)$ and all the threshold of nodes with index smaller than t, we want to compute the value of Y_{t+1} when the $t+1$-th node is added. For this there are three components that contribute to Y_{t+1}:

- First from previous steps we have $2ktY_t$ infected degrees.
- If the new added node $t+1$ is infected, then the k degrees of the edges that $t+1$ issues are infected. Thus, $t+1$ will contribute $kf_{k,\mathcal{R}}(Y_t)$ infected degree in expectation, where $f_{k,\mathcal{R}}(Y_t)$ is the probability of $t+1$ being infected.
- When $t+1$ is added, it issues k edges to previous t nodes. Some of these neighbors are already infected, so the new edges will contribute kY_t degrees in expectation.

Let $\mathrm{Bin}(n,p)$ be the random variable following binomial distribution, i.e., the total number of successful events out of a total of n events when each event succeeds with probability p independent of the others. Hence we get the following recurrence: for all $k < t \leq n$

$$(2k(t+1))Y_{t+1} \mid \mathcal{F}_t = 2ktY_t + \mathrm{Bin}(k, Y_t) + k \cdot \mathrm{Bin}(1, f_{k,\mathcal{R}}(Y_t)).$$

Note that $|Y_{t+1} - Y_t| \leq 1/(t+1)$, and it can be decomposed as predictable part g and noise part U:

$$Y_{t+1} - Y_t \mid \mathcal{F}_t = \frac{1}{t+1}(g(Y_t) + U_{t+1}) \tag{5.3}$$

where

$$g(y) = \frac{1}{2}(f_{k,\mathcal{R}}(y) - y), \text{ and} \tag{5.4}$$

$$U_{t+1} = \frac{1}{2}\left(\frac{\text{Bin}(k, Y_t)}{k} + \text{Bin}(1, f_{k,\mathcal{R}}(Y_t)) - Y_t - f_{k,\mathcal{R}}(Y_t)\right) \tag{5.5}$$

Define $W_s = \sum_{t=k+1}^{s} U_t/t$. Because $\mathbb{E}[U_{t+1}|\mathcal{F}_t] = 0$ and $|U_{t+1}| \leq 1$, $\{W_t : k < t \leq n\}$ is a martingale and we can rewrite the process as:

$$Y_t = Y_k + \sum_{s=k+1}^{t} \frac{1}{s} g(Y_{s-1}) + W_t, \text{ for } t \geq k \tag{5.6}$$

Now we present an outline of our proof.

- For Theorem 5.2 1a and 1b, we consider the function $f_{k,\mathcal{R}}$ only has a single fixed point y^*. The main observation of the proof is the noise term W_t is Cauchy-like(Lemma 5.3). That is given $\epsilon > 0$ there exists a large enough τ such that for all $\tau < s < t$ the different between $|W_t - W_s|$ is smaller than ϵ. Therefore, the behavior of Y_t in Equation (5.6) can be approximated by ignoring W_t.
  - With this observation, for Theorem 5.2 1a, given $\epsilon, \delta > 0$ we show the predictable part $g(Y_t)$ along will push the process Y_t toward the stable fixed point y^*, and afterward it will stay in the neighborhood $(y^* - \epsilon, y^* + \epsilon)$ forever with large probability. We present the proof in Section 5.2.3
  - For Theorem 5.2 1b, we consider the fixed point y^* is not only a stable fixed point, but also hyperbolic, ${f_{k,\mathcal{R}}}'(y^*) < 1$. In this case, instead of just staying in some fixed neighborhood $(y^* - \epsilon, y^* + \epsilon)$, we show process Y_t converges to y^* asymptotically, and the convergence rate depends on the slope of $f_{k,\mathcal{R}}$ at y^*. We present the proof in Section 5.2.4.
- For multiple fixed point, we show the limit behavior, $\lim_{t\to\infty} Y_t$. In Theorem 5.2 2, we show if $f_{k,\mathcal{R}}$ has finite fixed points, the process Y_t converges to stable and touch points in probability. In Theorem 5.2 3, we also show if $f_{k,\mathcal{R}}$ has infinite fixed points, the process Y_t is a martingale and converges to some random variable. We defer the proofs in appendix (Section B.2).

5.2.3 Proof of Theorem 5.2 1a

We first analyze the case when $f_{k,\mathcal{R}}$ has a unique stable fixed point (Theorem 5.2 1a).

Given an interval $D(y^*, \epsilon)$ of length 2ϵ centered at the fixed point y^* of the function $f_{k,\mathcal{R}}$, we will show that the process will stay in the interval with probability $1-\delta$) after some time $\tau = O(\epsilon^{-2} \log 1/\delta)$.

Our proof has two parts. First, Lemma 5.3 shows that the noise part, W_t in Equation (5.6), is Cauchy-like. This says that after a sufficiently large time τ_0 the distance of two noise terms, $|W_s - W_t|$, for $s > t > \tau_0$, would be small. Second, in Lemma 5.4, given an interval $D(y^*, \epsilon)$, if at certain time τ_0 the noise part is smaller than the width of the interval, then after $O(\tau_0)$ time the process will stay within $D(y^*, \epsilon)$ forever with large probability. The proofs of the two Lemmas can be found in Appendix B.2.

Lemma 5.3. *Given $\delta_0, \epsilon_0 > 0$, there exists $\tau_0 = \frac{2}{\epsilon_0^2} \ln(1/(2\delta_0))$ such that for all $s > t > \tau_0$, $\Pr[|W_s - W_t| > \epsilon_1] < \delta_1$.*

Now we are ready to prove the Theorem 5.2 1a. Here we give a stronger result.

Lemma 5.4. *Let $y_L^* = \min\{y \in Q_{f_{k,\mathcal{R}}}\}$ and $y_H^* = \max\{y \in Q_{f_{k,\mathcal{R}}}\}$. For all $\delta > 0$ and $\epsilon > 0$, there exists $\tau = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$*

$$\Pr[\forall n \geq \tau, y_L^* - \epsilon \leq Y_n \leq y_H^* + \epsilon] \leq \delta.$$

It's easy to see Theorem 5.2 1a is a special case of the above lemma since Q_f only contains a fixed point, so $y^* = y_L^* = y_H^*$.

5.2.4 Proof of Theorem 5.2 1b

Theorem 5.2 1b is a stronger result than Theorem 5.2 1a. It says that when the fixed point y^* is hyperbolic, ${f_{k,\mathcal{R}}}'(y^*) < 1$, the process $\{Y_t\}$ will converge to the fixed point y^* with polynomial convergence rate with probability arbitrarily close to 1.

We decompose the process into two phases: In the first phase with large probability the process would enter and stay in the good interval $D(y^*, \epsilon)$ which will be defined later; in the second phase the process would approach y^* fast.

Now we define the good interval $D(y^*, \epsilon)$. By the definition of γ we have $0 < \gamma < \frac{1-{f_{k,\mathcal{R}}}'(y^*)}{2}$, and take $\rho > 0$ small enough such that $1 \geq (1-\gamma\rho)(1+\rho)^\gamma$. Furthermore, we take γ_1, γ_2 such that $\gamma < \gamma_2 < \gamma_1 < |g'(y^*)| \leq 1/2$. We define $\epsilon > 0$ such that

$$x \in (y^*, y^* + \epsilon], g(x) < -\gamma_1(x - y^*), \text{ and } x \in [y^* - \epsilon, y^*), g(x) > \gamma_1(y^* - x). \quad (5.7)$$

By Lemma 5.4, it's not hard to see the process will enter and stay in the good interval $D(y^*, \epsilon)$. For the second phase, we define a random sequence (e_t) and two reference sequences (α_t) and (β_t). Given constants A and B specified later,

$$\alpha_t = \frac{A}{t^\gamma}, \beta_t = \frac{B}{t^\gamma}, \text{ and } e_t = |y^* - Y_t|. \tag{5.8}$$

We will prove after time τ_0 the distance between Y_t and the fixed point y^* can be bounded by β_t with large probability which completes the proof.

Proof. For the first phase, given ϵ defined in (5.7), for any $\epsilon' < \epsilon$, there exists a constant τ_0 such that $Y_{\tau_0} \in D(y^*, \epsilon')$ with probability $1 - \delta/2$ by Lemma 5.4. For the second phase, we focus on the process Y_t for time between σ and $(1+\rho)\sigma$. We first pick A such that $\epsilon' = \alpha_{\tau_0}$, so $e_{\tau_0} \le \alpha_{\tau_0}$. If $e_t \le \alpha_t$ for all $t \ge \tau_0$, we are done, since $\alpha_t = O(1/t^\gamma)$. On the other hand, if there exists a bad transition $\sigma > \tau_0$ such that $e_{\sigma-1} \le \alpha_{\sigma-1}$ and $e_\sigma > \alpha_\sigma$, we want to show with probability $1 - \exp(-C'\sigma^{1-2\gamma})$ (1) there exists τ between σ and $(1+\rho)\sigma$ such that the deviation e_τ is smaller than α_τ and (2) the deviation e_t between σ and $(1+\rho)\tau$ are smaller than β_t. Formally,

$$\Pr\left[\exists \tau, \forall t, \sigma \le t < \tau \le (1+\rho)\sigma, e_\tau \le \alpha_\tau \wedge e_t \le \beta_t \mid e_\sigma > \alpha_\sigma\right] = 1 - \exp\left(-C'\sigma^{1-2\gamma}\right) \tag{5.9}$$

The arguments for the above two claims are similar to Lemma 5.4. The first one is more complicated, so we first present this one.

Part 1 Suppose

$$\forall \sigma < s \le (1+\rho)\sigma, e_s > \alpha_{(1+\rho)\sigma} \tag{5.10}$$

We want to reach a contradiction. By the property of τ_0, we have $Y_{\tau_0} \in D(y^*, \epsilon')$. By taking ρ and d' small enough we have $Y_s \in D(y^*, \epsilon)$ for all $\sigma < s < (1+\rho)\sigma$. Moreover by (5.7) and definition of ρ we have $\alpha_{(1+\rho)\sigma} \ge (1 - \gamma\rho)\alpha_\sigma$, and

$$g(Y_s) \ge \gamma_1 \alpha_{(1+\rho)\sigma} \ge \gamma_1(1-\gamma\rho)\alpha_\sigma. \tag{5.11}$$

Because the diameter of interval $D(\alpha_{(1+\rho)\sigma})$ is greater than $1/t$ for all $\sigma \le t$, we can assume the process Y_t is either greater than $y^* + \alpha_{(1+\rho)\sigma}$ or below $y^* - \alpha_{(1+\rho)\sigma}$, and remove the absolute value. Suppose all Y_t are above $y^* + \alpha_{(1+\rho)\sigma}$. By Equation (5.6)

$$e_{(1+\rho)\sigma} \le Y_\sigma - y^* + Y_{(1+\rho)\sigma} - Y_\sigma \le \alpha_{\sigma-1} + \frac{1}{\sigma} + \sum_{t=\sigma+1}^{(1+\rho)\sigma} \frac{1}{t} g(Y_{t-1}) + (W_{(1+\rho)\sigma} - W_\sigma). \tag{5.12}$$

The last inequality holds because after the bad transition σ e_σ is still close to $\alpha_{\sigma-1}$. Combining (5.10) (5.11) and (5.12) we have

$$(1-\gamma\rho)\gamma_1\alpha_\sigma\sum_{t=\sigma+1}^{(1+\rho)\sigma}\frac{1}{t}+\alpha_{(1+\rho)\sigma}-\alpha_{\sigma-1}-\frac{1}{\sigma}<W_{(1+\rho)\sigma}-W_\sigma$$

which gives an lower bound on the noise $W_{(1+\rho)\sigma}-W_\sigma$, and it's sufficient to give a lower bound on the left hand side.

$$\begin{aligned}
&(1-\gamma\rho)\gamma_1\alpha_\sigma\sum_{t=\sigma+1}^{(1+\rho)\sigma}\frac{1}{t}+\alpha_{(1+\rho)\sigma}-\alpha_{\sigma-1}-\frac{1}{\sigma}\\
\geq&(1-\gamma\rho)\gamma_1\alpha_\sigma\ln\left(\frac{(1+\rho)\sigma}{\sigma+1}\right)+\alpha_{(1+\rho)\sigma}-\alpha_{\sigma-1}-\frac{1}{\sigma} && \text{(integral test)}\\
\geq&(1-\gamma\rho)\gamma_1\alpha_\sigma\left(\rho+o(\rho)+O\left(\frac{1}{\sigma}\right)\right)+\alpha_{(1+\rho)\sigma}-\alpha_{\sigma-1}-\frac{1}{\sigma} && \text{(small } \rho \text{ and large } \sigma)\\
\geq&\left[(1-\gamma\rho)\gamma_1\left(\rho+o(\rho)+O\left(\frac{1}{\sigma}\right)\right)+(1-\gamma\rho)-1\right]\alpha_\sigma-O\left(\frac{1}{\sigma}\right)\\
\geq&\left[(\gamma_1-\gamma)\rho+o(\rho)+O\left(\frac{1}{\sigma}\right)\right]\alpha_\sigma-O\left(\frac{1}{\sigma}\right)\geq(\gamma_2-\gamma)\rho\alpha_\sigma
\end{aligned}$$

With small enough ρ, we have

$$W_\sigma-W_{(1+\rho)\sigma}\geq\big((\gamma_2-\gamma)\rho\big)\alpha_\sigma$$

With Azuma-Hoeffding inequality in Theorem 2.18, there exist constants C and C' such that

$$\Pr[W_\sigma-W_{(1+\rho)\sigma}\geq\big((\gamma_2-\gamma)\rho\big)\alpha_\sigma]\leq\exp\left(-\frac{(\gamma_2-\gamma)^2\rho^2\alpha_\sigma^2}{\frac{C}{\sigma}\frac{\rho}{1+\rho}}\right)\leq\exp\left(-C'\sigma^{1-2\gamma}\right).$$

The last inequality is true because $\gamma<1/2$ and $\sigma\geq\tau_0=\Theta\left(\ln n\right)$. Therefore we prove the first claim,

$$\Pr\left[\exists\tau,\sigma<\tau\leq(1+\rho)\sigma,e_\tau\leq\alpha_\tau\mid e_\sigma>\alpha_\sigma\right]=1-\exp\left(-C'\sigma^{1-2\gamma}\right). \tag{5.13}$$

Part 2 For the second claim, we wan to show the deviation e_t from σ to $(1+\rho)\sigma$ are smaller than β_t. It is identical to the above proof. Given ρ and A we can take B large enough such that $\beta_{(1+\rho)\sigma}\geq3\alpha_\sigma$ which is independent to σ. Suppose there

exists t where $\sigma < t \leq \tau$ such that $e_t \geq \beta_t$. Using similar argument in (5.12) we have

$$3\alpha_\sigma \leq e_t \leq \alpha_\sigma + \frac{1}{\sigma} + W_t - W_\sigma.$$

By Azuma-Hoeffding inequality (Theorem 2.18), $\Pr[W_t - W_\sigma \geq \alpha_\sigma] = \exp\left(-C'\sigma^{1-2\gamma}\right)$, so

$$\Pr\left[\exists \forall t, \sigma \leq t \leq (1+\rho)\sigma, e_t \leq \beta_t \mid e_\sigma > \alpha_\sigma\right] = 1 - \exp\left(-C'\sigma^{1-2\gamma}\right) \tag{5.14}$$

which proves (5.9) by taking union bound of (5.13) and (5.14).

By union bound on (5.9) on all possible bad transition σ, we have

$$\Pr[\forall t \geq \tau_0, Y_t \leq \beta_t] \geq 1 - \frac{\delta}{2} - \sum_{s \geq \tau_0}^{n} \exp\left(-C's^{1-2\gamma}\right) \geq 1 - \delta$$

by taking constant τ_0 large enough. □

5.3 Undirected Preferential Attachment Graphs

In this section, we analyze the fraction of infected degree $\bar{Y}_n$ when the underlying network is an undirected preferential attachment graph $\bar{\mathrm{PA}}_k(n)$. Here we categorize the behavior of $\bar{Y}_n$ based on the values of the stable fixed points of the *directed* version of Y_n.

Theorem 5.5. *Let $GTC(\bar{\mathrm{PA}}_k(n), \mathbb{R}, X)$ be the general threshold contagions on an undirected preferential attachment graph $\bar{\mathrm{PA}}_k(n)$, $\bar{Y}_n$ be the infected ratio, and $f_{k,\mathcal{R}}(y)$ defined on a directed $\mathrm{PA_k(n)}$ in Equation (5.2). We have:*

1. *If $1 \in Q_{f_{k,\mathcal{R}}}$, $0 \notin Q_{f_{k,\mathcal{R}}}$, and the initial infected nodes $X \neq \emptyset$, then the whole network will be infected, i.e., $\Pr[\bar{Y}_n = 1] = 1 - o(1)$.*

2. *If $0, 1 \notin Q_{f_{k,\mathcal{R}}}$ and $y^*_H = \max(Q_{f_{k,\mathcal{R}}})$ which is a stable fixed point, there exists a constant $\Delta > 0$ such that $\bar{Y}_n$ converges to a value greater than y^*_H with high probability,*

$$\Pr[\bar{Y}_n > y^*_H + \Delta] = 1 - o(1).$$

5.3.1 Proof outline

The theorem shows the fraction of infected degree on undirected preferential attachment graph is constantly larger than the one on directed cases. To show this lower

bound, in our proof, we will restrict how the contagion can proceed, which will serve to establish a lower bound for the infection ratio of the undirected contagion.

First we only consider the contagion passing from low indexed nodes to high indexed nodes – just as in the directed case; then we only consider the contagion passing from high indexed nodes to low indexed nodes. We call these two processes the *forward* and *backward* processes respectively. We will repeat these processes twice. Each time, we only need to reveal the edges that can help spread an infection (i.e., the edges that point to an infected node), and the remaining edges are revealed later. We will use this to carefully manage (in)dependence so that we may employ concentration bounds.

The intuition in the analysis is the following. The first forward process is essentially the same as the directed contagion case. If there are non-zero stable fixed points then the contagion will infect a constant fraction of nodes. In fact, since the stochastic process in the directed case converges fast, among the nodes of high indices there is a good fraction of infected nodes and these infected nodes are roughly uniformly distributed. Therefore in the first backward process, these nodes will infect the nodes with small indices with high probability, which will continue to boost the propagation in the next forward round. The following analysis will make this rigorous.

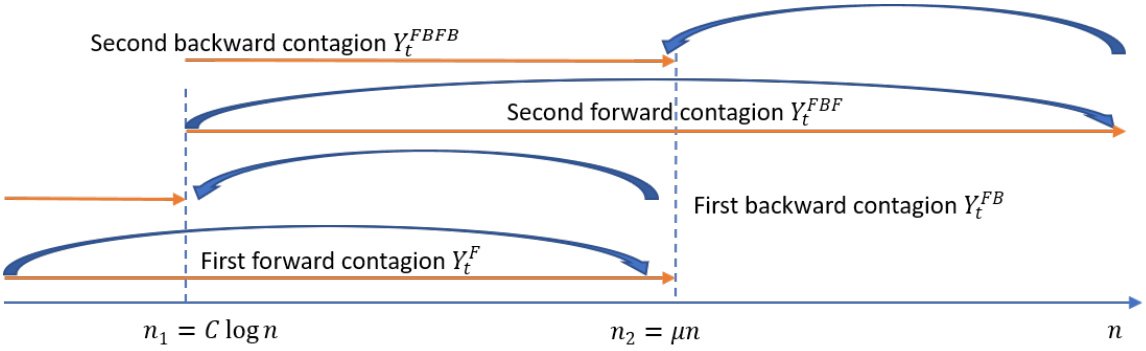


Figure 5.1: Let $\tau_1 = C_1 \log(n)$ and $\tau_2 = C_2 n$ where constants C_1 and C_2 will be specified later. In our first forward/backward process, we will actually only process nodes from 1 to τ_2 and back to 1, but in the second round, we will process all nodes with index from 1 to n and back down to 1.

For the sake of the proof, we divide these processing steps up into three phases and present them in the following subsections. The goal is to show that some specific properties happen at the end of each phase:

1. *First forward and backward contagion:* Run the infection in the forward direction from node 1 to the node with index τ_2. For each node $t \leq \tau_2$, we set I_t^F

as the indicator variable of the node being infected in the first forward process. We also use Y_t^F be the fraction of infected degree in the first forward infection of all nodes with index between 1 and t. Then we run the backward contagion from τ_2 back to 1. Define I_t^{FB} and Y_t^{FB} accordingly. Lemmas 5.7 and 5.9 show that all the first τ_1 node will be infected with high probability, i.e., $\Pr[Y_{\tau_1}^{FB} = 1] = 1 - o(1)$ is high.

2. *Converge to highest fixed point:* Conditioned on $Y_{\tau_1}^{FB} = 1$, run the second forward infection to node τ_2 again. We show that the infection ratio after the first τ_2 nodes, denoted by $Y_{\tau_2}^{FBF}$, is around the highest stable fixed point y_H^* (Lemma 5.10).

3. *Constant separation:* Conditioned on the infection ratio $Y_{\tau_2}^{FBF}$ being around the highest stable fixed point, y_H^*, run the infection in the forward direction from τ_2 to n and backward from n to 1 to show that the infection fraction, Y_n^{FBFB} , is incremented by a constant in the second backward round (Lemma 5.11 and 5.12).

Let $M_y^F(s,t)$ be the event that all nodes with index within $(s,t]$ are infected with probability greater than y after the first forward process. Similarly define $M_y^{FB}(s,t)$ and $M_y^{FBF}(s,t)$ accordingly after the first backward process and the next forward process respectively. Recall that $\deg(s;t)$ be the degree of node s at time t. For example $\deg(s;s) = k$, $\deg(s;n) = \deg(s)$, and $\deg(s;n) - \deg(s;t)$ is the number of neighbor in $(t,n]$.

5.3.2 First forward and backward contagion

After the first forward phase, the fraction of infected degrees is a positive constant, around one of the (non-zero) stable fixed point. The crucial part is to examine what happens in the first backward phase. We use two facts: fixing a node t, $t \leq \tau_1$, all neighbors of node $s > \tau_1$ will still have a large probability to get infected, in the first forward phase (proven in Lemma 5.7); furthermore, with a fact that with high probability, early nodes have large degree (proven in Lemma 5.8). Finally, we use a union bound to prove that all nodes t where $t \leq \tau_1$ will get infected with high probability. Some of the technical proofs are in the appendix.

Before jumping into the proof, note that by Lemma 5.4 we have $M_y^F(t,n)$ happens with very high probability. Formally,

Corollary 5.6. *Given δ, ϵ and y, if $\delta, \epsilon > 0$, $y < \min Q_{f_{k,\mathcal{R}}} - \epsilon$, then there exists $\tau = O\left(\frac{\log 1/\delta}{\epsilon^2}\right)$ such that*

$$\Pr\left[M_y^F(\tau, n)\right] \geq 1 - \delta.$$

Moreover, by the definition, for all y and $s \leq s' < t' \leq t \leq n$, $M_y^F(s', t') \subseteq M_y^F(s, t)$.

Lemma 5.7 (Uniform infection)**.** *Given an arbitrary node $s \leq \tau_1$, for all t such that $\tau_1 < t \leq \tau_2$. Let $0 < y < y_L^* = \min Q_{f_{k,\mathcal{R}}}$ and $N(t, s)$ be the event that node t connects to s. Then there exists a constant $q^* > 0$ such that for all $\mathcal{F}_{t-1}$ we have*

$$\Pr\left[I_t^F = 1 \mid \mathcal{F}_{t-1}, M_y^F(\tau_1, \tau_2), N(t, s)\right] \geq q^*.$$

Informally, this lemma ensures t has constant probability to be infected given t being a neighbor of s, $N(t, s)$. The following lemma show the number of neighbor between τ_1 and τ_2 is large. Note that the quantities $\deg(s; \tau_2) - \tau(s; \tau_1)$ greater than $\deg(\tau_1; \tau_2) - k$, and the degree of τ_1 can be computed directly.

Lemma 5.8 (Degree lower bound)**.** *If $s \leq \tau_1$,*

$$\Pr\left[\deg(s; \tau_2) - \deg(s; \tau_1) < \log\log(n)\right] = o\left(\frac{1}{\log^2(n)}\right).$$

Applying Lemma 5.7 and Lemma 5.8, we will finish the first phase by proving that $Y_{\tau_1}^{FB} = 1$ with high probability.

Lemma 5.9. *Given the network generated by the preferential attachment model with only the top τ_2 nodes, $\bar{\mathrm{PA}}_k(\tau_2)$,*

$$\Pr\left[Y_{\tau_1}^{FB} = 1\right] > 1 - O(1/\log \tau_2)$$

over the randomness of $\mathrm{PA}_k(\tau_2)$ and thresholds of the first τ_2 nodes.

5.3.3 Converge to the highest fixed point

Now we enter the second phase to show that the ratio, $Y_{\tau_2}^{FBF}$ will be around the highest stable fixed point y_H^*. We consider the following two events:

1. Event $Y_{\tau_1}^{FB} = 1$;
2. Event that $Y_{\tau_2}^{FBF}$ would be around the highest stable fixed point y_H^* conditioned on $Y_{\tau_1}^{FB} = 1$.

The intersection of these two events is what we want to prove. Since the first event happens with probability at least $1 - o(1)$ by Lemma 5.9. The second event can be shown through the identical argument in Lemma 5.4.

Lemma 5.10. *Let y_H^* be the highest fixed point of $f_{k,\mathcal{R}}$ which is stable. Conditioned on $Y_{\tau_1}^{FB} = 1$, for all $\epsilon > 0$,*

$$\Pr\left[|Y_{\tau_2}^{FBF} - y_H^*| \leq \epsilon\right] = 1 - o\left(\frac{1}{n}\right).$$

5.3.4 Constant separation phase

Finally, for the third phase, we reveal the edges from node τ_2 to n, and show that the infected ratio Y_n^{FBFB} after the second backward contagion will have a constant improvement, i.e., $Y_n^{FBFB} > y_H^* + \Delta$ where $\Delta > 0$ is independent of n.

Let X^{FBF} be the set of infected nodes after the second forward infection, and $Z^{FBF} = \{s \leq \tau_2 : s \notin X^{FBF}\}$. Given a node $s \in Z^{FBF}$, let I_s^{FBFB} be the indicator function that node t is not infected during the second forward infection but getting infected in the second backward infection. First note that the extra infected degree is lower bounded by $\sum_{s \leq \tau_2} \deg(s; \tau_2) I_s^{FBFB}$,

$$Y_n^{FBFB} \geq Y_n^{FBF} + \frac{1}{2kn} \sum_{s \leq \tau_2} \deg(s; \tau_2) I_s^{FBFB}.$$

so it is sufficient to show the random variable $L = \sum_{s \in Z^{FBF}} \deg(s; \tau_2) I_s^{FBFB}$ is $\Omega(n)$ with high probability. We use the second moment method to show this claim in Lemma 5.13 with the following two components:

1. $\mathbb{E}[L]$ is large. We first show for all $s \in Z^{FBF}$ the probability of getting infected in the second backward infection is large by argument similar to the first backward contagion. Second, by Lemma 5.10, with high probability $Y_{\tau_2}^{FBF}$ is strictly smaller than 1 so the number of uninfected degree is large, $\sum_{s \in Z^{FBF}} \deg(s; \tau_2) = \Omega(n)$.

2. $\mathrm{Var}[L]$ is small. In Lemma 5.12, we show the indicator functions I_s^{FBFB} are negative correlated, so the variance of L is smaller than the sum of variance of I_s^{FBFB}.

The proof of these lemmas can be found in the Appendix B.3.

Lemma 5.11 (Expectation)**.** *If $\tau_2 = C_2 n$, then there exists $y < y_H^*$ and $C_2 > 0$ such that for all $s \le \tau_2$ that is uninfected before the second backward infection*

$$\mathbb{E}\left[I_s^{FBFB} \mid \mathcal{F}_{\tau_2}, M_y^{FBF}(\tau_2, n)\right] \ge \frac{1}{4}.$$

Moreover, for all d, $\mathbb{E}\left[I_s^{FBFB} \mid \mathcal{F}_{\tau_2}, M_y^{FBF}(\tau_2, n), \deg(s; \tau_2) \ge d\right] \ge 1 - O\left(\frac{1}{\sqrt{d}}\right)$.

The moreover part ensures the high degree nodes also have high probability get infected. Combining this observation and the negative correlation in Lemma 5.12, we can show the variance of L is small.

Lemma 5.12 (Negative correlated)**.** *For all s and t smaller than τ_2,*

$$\mathbb{E}\left[I_s^{FBFB} I_t^{FBFB} \mid \mathcal{F}_{\tau_2}\right] \le \mathbb{E}\left[I_s^{FBFB} \mid \mathcal{F}_{\tau_2}\right] \mathbb{E}\left[I_t^{FBFB} \mid \mathcal{F}_{\tau_2}\right].$$

Lemma 5.13 (Second moment method)**.** *Given $\mathcal{F}_{\tau_2}$ such that $Y_{\tau_2}^{FBF} < \frac{1}{2}(1 + y_H^*)$, then*

$$\Pr\left[\sum_{s \le \tau_2} \deg(s; \tau_2) I_s^{FBFB} \ge \Delta n \mid \mathcal{F}_{\tau_2}, M_y^{FBF}(\tau_2, n)\right] \ge 1 - o(1).$$

Now we can prove Theorem 5.5:

Proof of Theorem 5.5. The first part of the proof is derived from Lemma 5.9 since $Y_{\tau_1} = 1$ with high probability and $f(1) = 1$ is a fixed point, then all the nodes after τ_1 will get infected and $\Pr[Y_n = 1] = \Pr[Y_{\tau_1} = 1] = 1 - o(1)$ by argument similar to Lemma 5.10.

In second part, by Lemma 5.9 and 5.10 $Y_{\tau_2}^{FBF} < \frac{1}{2}(1 + y_H^*)$ happens with high probability, and we finish the proof by applying Lemma 5.13. By union bound, the event fails with probability less than $1 - o(1)$. □

5.4 General threshold cascade on stochastic attachment graph

In this section, we show how to extend our result (Theorem 5.2) to a general family of time evolving graph, stochastic attachment graph.

5.4.1 Stochastic attachment graphs

We can generalize preferential attachment models $\mathrm{PA}_k(n)$ in two axes: the outgoing degree of each node can be different, and the attachment rule.

Definition 5.14 (attachment rule)**.** An *attachment rule* $\mathbb{A}$ is a random function mapping from a graph $G(V, E)$ to a node in the graph. For example,

uniform $\Pr[\mathbb{A}_u(V, E) = v] = 1/|V|$ for all $v \in V$.

preferential $\Pr[\mathbb{A}_p(V, E) = v] = \deg(v)/(2|E|)$ for all $v \in V$.

α-mixture Given $\alpha \in [0, 1]$, the attchment rule is the mixture between uniform and preferential attachment $\Pr[\mathbb{A}_\alpha(V, E) = v] = \alpha \deg(v)/(2|E|) + (1 - \alpha)1/|V|$ for all $v \in V$.

Definition 5.15 (Stochastic attachment model)**.** Given a distribution $\mathcal{K}$ over integers with support $\mathrm{supp}(\mathcal{K})$ between 0 and K_H, $n > K_H$, and an attachment rule $\mathbb{A}$, the ***Stochastic Attachment Model***, $\mathrm{SA}_{\mathcal{K},\mathbb{A}}(n)$ is a distribution over directed graphs with n nodes. $\mathrm{SA}_{\mathcal{K},\mathbb{A}}(n)$ generate $G_n = (V_n, E_n)$ as follows:

1. Set $V_n = [n] = \{1, 2, \ldots, n\}$ and $E = \emptyset$, and we call node s is earlier than node t if $s < t$.

2. Let E_{K_H+1} consists of the edge between the first K_H+1 nodes, $\{1, \ldots, K_H+1\}$, which forms a $(K_H + 1)$-clique. We set $G_{K_H+1} := ([K_H + 1], E_{K_H+1})$.

3. At each subsequent step $t + 1 > K_H + 1$, given $G_t = ([t], E_t)$ the node $t + 1$ samples a number k from the distribution $\mathcal{K}$, chooses $w_1, w_2, \ldots, w_k$ vertices according to the attachment rule $\mathbb{A}(G_t)$ independently,[3] and adding k new edges $(t + 1, w_1), \ldots, (t + 1, w_k)$. We call the resulting graph $G_{t+1} = ([t + 1], E_{t+1})$.

Note that the preferential attachment model is a special case, $\mathrm{PA}_k(n) = \mathrm{SA}_{k,\mathbb{A}_p}(n)$ where $\mathbb{A}_p$ is the preferential attachment rule, $\mathcal{K}$ is a constant k.

5.4.2 General threshold contagions on directed stochastic attachment graph

The analysis we did before is for the case of preferential attachment graph. Here we give the analysis for the more general case, when 1) the number of edges of the

[3]This may cause multiple edges.

newcomer to previous nodes is sampled from a bounded distribution $\mathcal{K}$ with range between 1 and K_H and $E[K] = \mu_K$; 2) when the attachment rule can be either preferential or uniformly at random.

Preferential attachment rule Similar to the analysis in Section 5.2, we first look at the case of preferential attachment when each newcomer may choose different number of edges from a common distribution $\mathcal{K}$. Now we consider $\{(Z_t, Y_t) : K_H+1 \leq t \leq n\}$ where Z_t is the total degree of graph G_t normalized by μ_K, and Y_t is the fraction of infected degree before node t arrives. Similarly, we can compute the probability of the t-th node being infected when the threshold is $R_t = r$ and $k_t = k$ edges go to the previous nodes V_{t-1},

$$\Pr[t \text{ is infected} \mid R_t = r, k_t = k] = \sum_{\ell=r}^{k} \binom{k}{\ell} Y_t^{\ell}(1-Y_t)^{(k-\ell)}$$

Using the same argument as in Section 5.2, we have

$$Y_{t+1} - Y_t = \frac{1}{\mu_K Z_{t+1}} \left(\mathrm{Bin}(k_t, Y_t) + k_t \,\mathrm{Bin}\left(1, \sum_{\ell=R_t}^{k_t} \binom{k_t}{\ell} Y_t^{\ell}(1-Y_t)^{(k_t-\ell)}\right) - 2k_t Y_t \right). \tag{5.15}$$

In analogous to Theorem 5.2 2, we define a deterministic function $f^p_{\mathcal{K},\mathbb{R}}(y)$ as the expectation of $\frac{1}{\mu_K}\left(\mathrm{Bin}(k_t, Y_t) + k_t \sum_{\ell=R_t}^{k_t} \binom{k}{\ell} Y_t^{\ell}(1-Y_t)^{(k-\ell)} - 2k_t Y_t\right)$ given $Y_t = y$. With some rearrangement, we have

$$f^p_{\mathcal{K},\mathbb{R}}(y) = \frac{1}{\mu_K} \mathbb{E}_{\mathcal{K},\mathbb{R}} \left[k \sum_{\ell=r}^{k} \binom{k}{\ell} y^{\ell}(1-y)^{(k-\ell)} \right] - y.^4 \tag{5.16}$$

Intuitively, we can expect the process Y_t converges to the zeros of $f^p_{\mathcal{K},\mathbb{R}}$. Although the total degree Z_t is a stochastic process, the asymptotic behavior is rather simple, $Z_t = \Theta(t)$ with probability 1. We can achieve an analogy to Theorem 5.2 2.

Uniform attachment rule Let Y_t be the *fraction of infected nodes*. Using the same argument, we have

$$Y_{t+1} - Y_t = \frac{1}{t+1} \left(\mathrm{Bin}\left(1, \sum_{\ell=R_t}^{k_t} \binom{k_t}{\ell} Y_t^{\ell}(1-Y_t)^{(k_t-\ell)}\right) - Y_t \right), \tag{5.17}$$

[4]Note that if $\mathcal{K}$ is constant k $f^p_{\mathcal{K},\mathbb{R}}(y)$ reduces to $f_{k,\mathcal{R}}$ defined in (5.2)

and we can also define a deterministic function $f^u_{\mathcal{K},\mathbb{R}}(y)$ to be

$$f^u_{\mathcal{K},\mathbb{R}}(y) = \mathbb{E}_{\mathcal{K},\mathbb{R}}\left[\sum_{\ell=r}^{k}\binom{k}{\ell}y^\ell(1-y)^{(k-\ell)}\right] - y, \tag{5.18}$$

such that $\mathbb{E}[Y_{t+1} - Y_t \mid Y_t = y] = \frac{1}{t+1}f^u_{\mathcal{K},\mathbb{R}}(y)$.

Now we are ready to state the theorem, and the proofs are deferred to appendix (Section B.4).

Theorem 5.16. *Given a natural number n, a distribution $\mathcal{K}$ of out degree, a distribution $\mathbb{R}$ of threshold, and a set of initially infected nodes $X \subset [K_H]$, the infection of general threshold contagion on stochastic attachment models depends on the fixed points of certain functions. Specifically,*

1. *For preferential attachment, the fraction of infected degree of the general threshold contagion $GTC(\mathrm{SA}_{\mathcal{K},\mathbb{A}_p}(n), \mathbb{R}, X)$, Y_n defined in (5.15), converge almost surely to the stable fix point of $f^p_{\mathcal{K},\mathbb{R}}$ defined in (5.16).*
2. *For uniform random attachment, the fraction of infected node of the general threshold contagion $GTC(\mathrm{SA}_{\mathcal{K},\mathbb{A}_u}(n), \mathbb{R}, X)$, Y_n defined in (5.17) will converge almost surely to the stable fix point of $f^p_{\mathcal{K},\mathbb{R}}$ defined in (5.18).*

5.5 Simulations

We ran simulations on model networks and real world data sets to understand the behavior of a general threshold contagion and its dependency on threshold distribution D, the network structure, and the selection of initial seeds.

Model networks We generate graphs using the stochastic attachment model and run a contagion in both the directed and undirected version. We use two threshold distributions $\mathcal{R}_1$ and $\mathcal{R}_2$. In $\mathcal{R}_1$, the probability of taking a threshold of $1, 2, 7$ is $0.22, 0.39, 0.49$ respectively; in $\mathcal{R}_2$, the probability of taking a threshold of $1, 2, 5, 7$ is $0.1, 0.4, 0.45, 0.05$ respectively. Using definition of function f in Equation 2, with $m = 5$ and $\mathcal{R}_1$, f has one fixed point equal to 0.558.With $m = 6$ and $\mathcal{R}_2$, f has two fixed points 0.875 and 0.521. In each run of the simulation, we vary I to be a fraction $\beta \in [0, 1]$ of a constant number of the first 6 nodes for $\mathcal{R}_1$ and first 7 nodes for $\mathcal{R}_2$ of the network.

Directed network. We create a network G_1 based on Definition 2.51, in which each newcomer choose m edges that are preferentially attached to earlier nodes. G_1

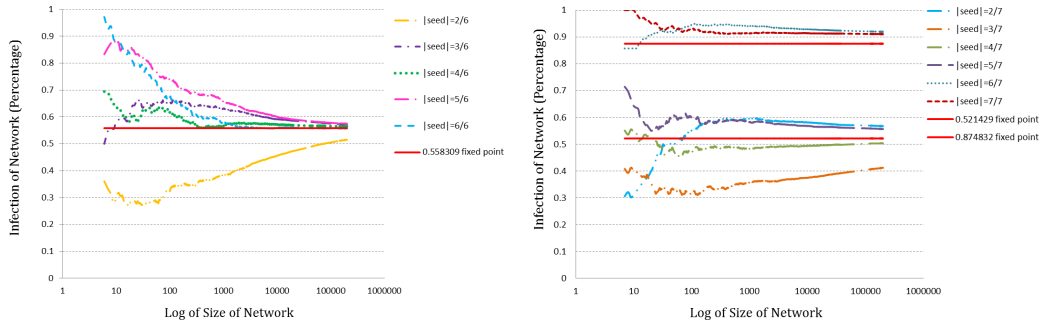


Figure 5.2: Contagion using threshold distribution $\mathcal{R}_1$ in (Left) and $\mathcal{R}_2$ in (Right) with different initial seeds on the directed preferential attachment graph.

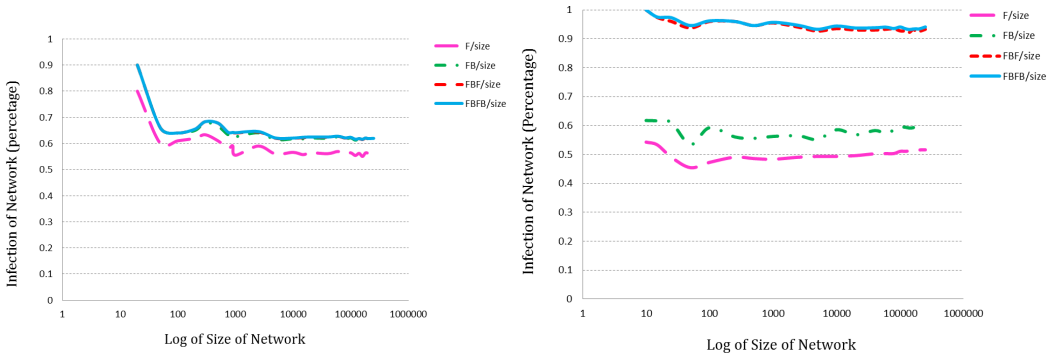


Figure 5.3: Contagion using threshold distribution $\mathcal{R}_1$ in (Left) and $\mathcal{R}_2$ in (Right) with different initial seeds on undirected preferential attachment graph.

is directed, each edge pointing from a high indexed node to a low indexed node. Figure 5.2 show the results of running a contagion over G_1 using $\mathcal{R}_1$ and $\mathcal{R}_2$ with different sets of seeds. For different runs, the ratio of infected nodes converges to one of the stable fixed points. When f has multiple fixed points (as in the case of $\mathcal{R}_2$), the way that the first few nodes are infected typically determine the infection rate of the entire network.

Undirected network. We take G_1, make all edges undirected, call it G_2. Then we run contagion in alternating forward and backward steps. See Figure 5.3. The first forward step behaves the same way as contagion on the directed network. The first backward step uniformly infects more nodes everywhere. In the case of $\mathcal{R}_2$, the next foward phase infects a large number of nodes. Additional steps do not change the infection state much.

DBLP and Web graphs We use two real world networks: the Stanford web graph (a directed network) and the DBLP co-authorship network (an undirected network).

1. Stanford web graph: Each node represents a page from Stanford University (stanford.edu) and there is a directed edge from u to v if u has a hyperlink to v. The network contains 281,903 nodes and 2,312,497 edges.

2. DBLP co-authorship network: The nodes are authors and there is an undirected between two nodes if they have published at least one paper together. This data set has 317,080 nodes and 1,049,866 edges.

To understand contagion on real networks, we first try to fit our stochastic attachment graph model. For that, we generate an arriving order from the real world graphs. There can be multiple ways to do so. Here we iteratively remove the lowest degree node, with ties broken arbitrarily. Then we take the reversed order and use it as the arriving order of the nodes. If the network is directed, we iteratively remove the node of lowest in-degree. Next, each node v has a degree d_v referring to the number of edges to the lower indexed nodes. We collect all such degrees d_v, $\forall v$, and use it for the outgoing degree distribution M. Then we generate a network G' using the stochastic attachment model with outgoing degree distribution M. Here, we set the number of nodes of the network to be $300,000$, which is almost the same as the number of nodes in both Stanford and DBLP data sets. We create a complete graph of m nodes, where m is the expectation of the outgoing degree distribution M, which is 6 for the Stanford data set and 3 for the DBLP data set. For the attachment rule we introduce a parameter $\alpha \in [0, 1]$ as the probability that an edge is attached using the preferential rule. If $\alpha = 0$ all edges are attached uniformly at random; if $\alpha = 1$ all edges are attached preferentially. In experiments, we use $\alpha = 0, 0.25, 0.5, 0.75, 1$.

For contagion model, we take two approaches. First, we take $\mathcal{R}$, the threshold distribution to be the Poisson distribution with parameter λ. We start each of the experiments from $\lambda = 1$ and increase its value until the total infection rate of the network drops below 1%. Second, we run a k-complex contagion model, in which all nodes have threshold k. We take seeds as the 25 lowest indexed nodes.

We run these two contagions over both real networks and their corresponding generated model networks. For comparison, we also generate a network using the configuration model following the same degree distribution of the real world network.

Figure 5.4 shows the results where the threshold distribution is a Poisson one. It can be observed that the behavior of contagion on the generated stochastic attachment graph (especially the one with $\alpha = .75$) matches the behavior of the real world graph fairly well, while the configuration model (though having the same degree distribution) does so poorly.

Figure 5.5 shows the results for k-complex contagion. Our models, though with infection rate shifted away from the behavior of the real world graph, is still much better than the behavior of configuration model (for which the infection rate is zero for any k complex contagions, $k \geq 2$). In particular, we believe this is partly due to

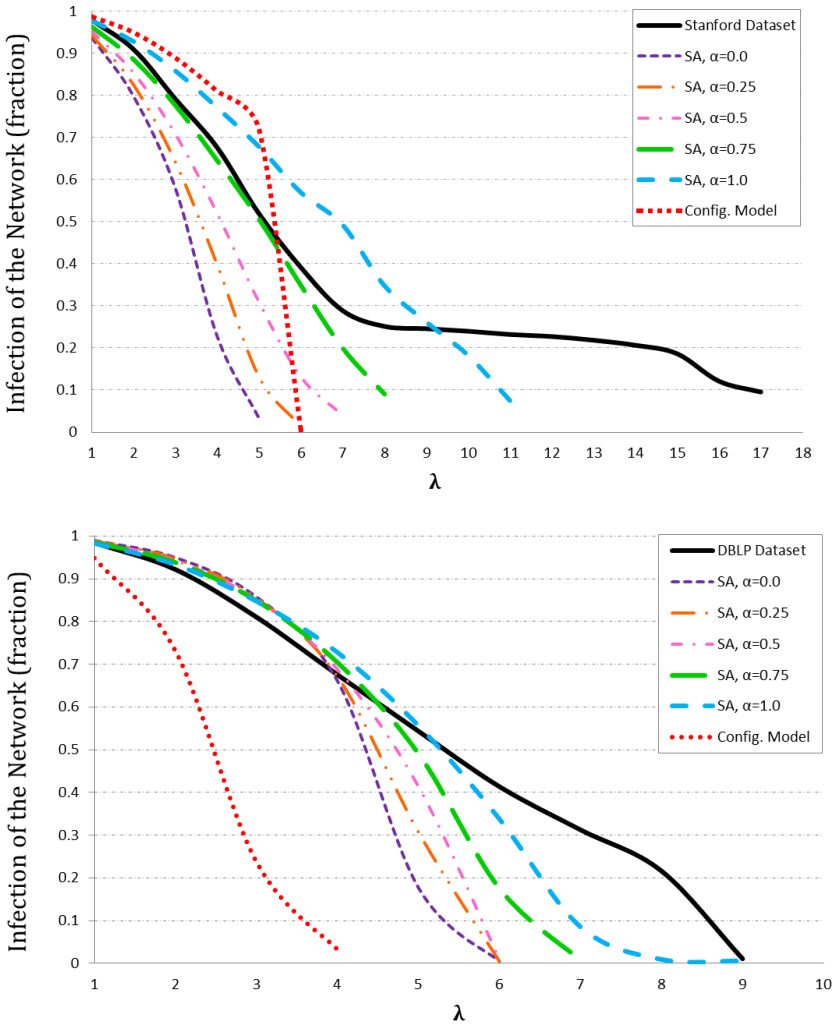


Figure 5.4: Contagion on (Top) Stanford web graph and (Bottom) the DBLP coauthorship graph, stochastic attachment models and configuration models.

the lack of community structures in the configuration model.

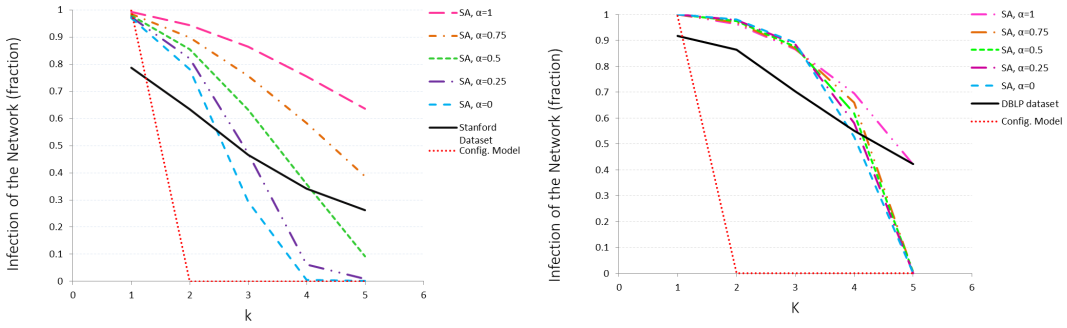


Figure 5.5: Contagion on Stanford web graph (left) and the DBLP coauthorship graph (right) under stochastic attachment models and configuration models.

5.6 Conclusion

This paper initiates the study of complex contagion with general thresholds. One takeaway is that stochastic attachment graph model can be used to estimate the behavior of contagion on real data sets better than configuration models.

CHAPTER 6

Stochastic Block Models

This is the last chapter about contagions. We study an important algorithmic problem of the r-complex contagion—influence maximization problem where one chooses a fixed number of initial seeds in a social network to maximize the spread of their influence. We focus on a random graph model named the *stochastic hierarchical block model*, which is a special case of the well-studied *stochastic block model*. We prove that, under certain mild assumptions, the optimal seeding strategy is to put all the seeds in a single community, when the graph is not exceptionally sparse, in particular, when each edge appears with probability $\omega(n^{-(1+1/r)})$. This matches the intuition that it is beneficial to put seeds near each other to maximize their synergy in a nonsubmodular cascade model, which is in sharp contrast to the seeding strategy in submodular cascade models (e.g., the independent cascade model and the linear threshold model).

Finally, we show that this observation yields a polynomial time dynamic programming algorithm which outputs optimal seeds if each edge appears with a probability either in $\omega(n^{-(1+1/r)})$ or in $o(n^{-2})$.

6.1 Introduction

Influence maximization is extensively studied when the contagion process is submodular (a node's marginal probability of becoming infected after a new neighbor is infected decreases when the number of previously infected neighbors increases [128]). However, many examples of nonsubmodular contagions have been reported, including pricey technology innovations, the change of social behaviors, the decision to participate in a migration, etc [61, 156, 199, 18, 141]. In this case, a node's marginal influence may increase in the presence of other nodes—creating a kind of synergy.

Network structure and seed placement We address this lack of understanding for nonsubmodular influence maximization by characterizing the optimal seed positions for certain settings which we will remark on shortly. In these settings, the optimal seeding strategy is to put all the seeds in a single community. This is arrestingly different than in the submodular setting, where the seeds can only erode each others' efficacy. Thus, in the submodular case, the optimal solutions tends to spread out the seeds, lest they erode each others' influence. We demonstrate this in Sect. 6.4 by presenting an example of submodular influence maximization where the optimal seeding strategy is to spread out the seeds.

This formally captures the intuition, as proposed by Angell and Schoenebeck [14], that it is better to target one market to saturate first (act locally) and then to allow the success in this initial market to drive broader success (think globally) rather than initially attempt to a scatter-shot approach (act globally). It also cautions that we must understand the particular nature of the contagion before blindly applying influence maximization tools.

We consider a well-known nonsubmodular cascade model which is also the most extreme one, the r-complex contagion [104, 49, 50, 88] (a node is infected if and only if at least r of its neighbors are infected, also known as *the bootstrap percolation*) when $r \geq 2$.

We consider *the stochastic hierarchical block model* [204] networks which are a special case of the stochastic block model [73, 113, 237] equipped with a hierarchical structure. Vertices are partitioned into m blocks. The blocks are arranged in a hierarchical structure which represents blocks merging to form larger and larger blocks (communities). The probability of an edge between two vertices is based solely on smallest block to which both the vertices belong. This model captures the intuitive hierarchical structure which is also observed in many real-world networks [101, 57]. The stochastic hierarchical block model is rather general and captures other well-studied models (e.g. the Erdos-Renyi random graphs, and the planted community model) as special cases.

Result 1: We first prove that, for the influence maximization problem on the stochastic hierarchical block model with the r-complex contagion, under certain mild technical assumptions, the optimal seeding strategy is to put all the seeds in a single community, if the probability that each edge appears satisfies $p_{uv} = \omega(n^{-(1+1/r)})$. Notice that this assumption captures many real life social networks. In fact, it is well-known that an Erdős-Rényi graph $\mathcal{G}(n, p)$ with $p = o(1/n)$ is globally disconnected: with probability $1 - o(1)$, the graph consists of a union of tiny connected

components, each of which has size $O(\log n)$.

The technical heart of this results is a novel coupling argument in Proposition 6.16. We simultaneously couple four cascade processes to compare two probabilities: 1) the probability of infection spreading throughout an Erdős-Rényi graph after the $(k+1)$-st seed, conditioned on not already being entirely infected after k seeds; 2) the probability of infection spreading throughout the same graph after the $(k+2)$-nd seed, conditioned on not already being entirely infected after $k+1$ seeds. This shows that the marginal rate of infection always goes up, revealing the supermodular nature of the r-complex contagion.

Our result is in sharp contrast to Balkanski et al.'s observation. Balkanski et al. [21] studies the stochastic block model with a well-studied submodular cascade model, *the independent cascade model*, and remarks that "when an influential node from a certain community is selected to initiate a cascade, the marginal contribution of adding another node from that same community is small, since the nodes in that community were likely already influenced."

Algorithmic Aspects For influence maximization in submodular cascades, a greedy algorithm efficiently finds a seeding set with influence at least a $(1-1/e)$ fraction of the optimal [128], and much of the work following Kempe et al. [128], which proposed the greedy algorithm, has attempted to make greedy approaches efficient and scalable [51, 52, 154, 58, 228].

Greedy approaches, unfortunately, can perform poorly in the nonsubmodular setting [14]. Moreover, in contrast to the submodular case which has efficient constant approximation algorithms, for general nonsubmodular cascades, it is NP-hard even to approximation influence maximization to within an $O(n^{1-\epsilon})$ factor of optimal [127], and the inapproximability results have been extended to several more restrictive nonsubmodular models [53, 147, 204]. The intrinsic reason why nonsubmodular influence maximization is hard is that one needs to take into account the potential synergy of multiple seeds. This is in sharp contrast to submodular influence maximization, where the submodularity enables a seed-picker to consider placing seeds one at a time in a myopic way, as it is in the greedy algorithm.

Can the $\Omega(n^{1-\epsilon})$ inapproximability results of Kempe et al. [127] be circumvented if we further assume the stochastic hierarchical block model? On the one hand, the stochastic hierarchical structure seems optimized for a dynamic programming approach: performing dynamic programming from the bottom to the root in the tree-like community structure. On the other hand, Schoenebeck and Tao [204] show

that the $\Omega(n^{1-\epsilon})$ inapproximability results extend to the setting where the networks are stochastic hierarchical block models.

Result 2: However, our observation that, when the network is reasonably dense, putting all the seeds in a single community is optimal can naturally be extended to a dynamic programming algorithm. We show this algorithm is optimal if the probability p_{uv} that each edge appears does not fall into a narrow regime. Interestingly, a heuristic based on dynamic programming works fairly well in practice [14]. Our second result theoretically justifies the success of this approach, at least in the setting of r-complex contagions.

Our Results We first prove that, for the influence maximization problem on the stochastic hierarchical block model with the r-complex contagion, under certain mild technical assumptions, the optimal seeding strategy is to put all the seeds in a single community, if the probability that each edge appears satisfies $p_{uv} = \omega(n^{-(1+1/r)})$.

Secondly, we extend this observation to a polynomial time algorithm which outputs optimal seeds for influence maximization on r-complex contagions on hierarchical stochastic block models if each p_{uv} satisfies either $p_{uv} = \omega(n^{-(1+1/r)})$ or $p_{uv} = o(n^{-2})$, in which case the graph can be viewed as many "almost isolated" dense subgraphs such that $p_{uv} = \omega(n^{-(1+1/r)})$ for u, v in the same subgraph and $p_{uv} = o(n^{-2})$ for u, v in different subgraphs. Our algorithm uses dynamic programming to decide the optimal number of seeds allocated to each subgraph; within a single subgraph and with the number of seeds allocated for this subgraph decided, our first result implies that the optimal seeding strategy is to put all these seeds in a same block.

6.2 Model

We study complex contagions (Definition 2.1) on social networks with community structure. This section defines the notions of our model for social network with community structure.

6.2.1 Stochastic Hierarchical block models

We study the *stochastic hierarchical block model* first introduced in [204]. The stochastic hierarchical block model is a special case of the *stochastic block model* [113]. Specifically and intuitively, the stochastic block model is a stochastic graph model dealing with networks with community structures, and the stochastic hierarchical block model

further assumes that the communities form a hierarchical structure. Our definition in this section follows closely to (the full version of) [204].

Definition 6.1. A *stochastic hierarchical block model* is a distribution $\mathcal{G} = (V, T)$ of unweighted undirected graphs sharing the same vertex set V, and $T = (V_T, E_T, w, v)$ is a weighted tree T called a *hierarchy tree*. The third parameter is the weight function $w : V_T \mapsto [0, 1]$ satisfying $w(t_1) < w(t_2)$ for any $t_1, t_2 \in V_T$ such that t_1 is an ancestor of t_2. Let $L_T \subseteq V_T$ be the set of leaves in T. Each leaf node $t \in L_T$ corresponds to a subset of nodes $V(t) \subseteq V$, and the $V(t)$ sets partition the nodes in V. In general, if $t \notin L_T$, we denote $V(t) = \bigcup_{t' \in L_T : t' \text{ is an offspring of } t} V(t')$. Let the function $v : V_T \mapsto 2^V$ in the forth parameter denote such correspondence.

The graph $G = (V, E)$ is sampled from $\mathcal{G}$ in the following way. The vertex set V is deterministic. For $u, v \in V$, the edge (u, v) appears in G with probability equal to the weight of the least common ancestor of u and v in T. That is $\Pr((u, v) \in E) = \max_{t : u, v \in V(t)} w(t)$.

In the rest of this paper, we use the words "node" and "vertex" to refer to the vertices in V_T and V respectively. In Definition 6.1, $V(t) \subseteq V$, corresponded by the node $t \in V_T$, represents a community in the social network, and $V(t_1), V(t_2), \ldots \subseteq V$, corresponded by the children $t_1, t_2, \ldots$ of t, partition $V(t)$ into sub-communities. Naturally, the relation between two nodes is stronger if they are in a same sub-community in a lower level. This justifies our assumption $w(t_1) < w(t_2)$ for any $t_1, t_2 \in V_T$ such that t_1 is an ancestor of t_2.

To consider the algorithmic aspect of influence maximization problem prior setting, because the graph G is not revealed, nodes with same label are symmetric, we can measure the complexity with respect to the complexity of stochastic hierarchical block (how complicate the tree structure T is), instead of the size of the graph (how large n is). As a result, we consider the following variant of the stochastic hierarchical block model.

Definition 6.2. A *succinct stochastic hierarchical block model* is a distribution $\mathcal{G}(n, T)$ of unweighted undirected graphs sharing the same vertex set V with $|V| = n$, where n is an integer which is assumed to be extremely large. The hierarchy tree $T = (V_T, E_T, w, v)$ is the same as it is in Definition 6.1, except for the followings.

1. Instead of mapping a node t to a weight in $[0, 1]$, the weight function w is a function $w : V_T \mapsto \mathcal{F}$, where $\mathcal{F} = \{f \mid f : \mathbb{Z}^+ \mapsto [0, 1]\}$ is the space of functions mapping an integer to a weight in $[0, 1]$. The weight of t is then defined by

$(w(t))(n)$. We assume $\mathcal{F}$ is the space of all functions that can be succinctly encoded.

2. For each node $t \in V_T$, we do not record the subset of vertices $V(t) \subseteq V$ that t corresponds to. Instead, we record a real number $v(t) \in (0, 1]$ which denotes the fraction $|V(t)|/n$. Naturally, we have $\sum_{t \in L_T} v(t) = 1$ and $\sum_{t':t' \text{ is a child of } t} v(t') = v(t)$.

We assume throughout that $\mathcal{G}(n, T)$ has the following properties.

Large communities For all node $t \in V_T$, we only consider the value of $v(t)$ that is independent of n. As a result, for each $t \in V_T$, $|V(t)| = v(t)n = \Theta(n)$ tends to infinity as $n \to \infty$.

Proper separation $w(t_1) = o\left(w(t_2)\right)$ for any $t_1, t_2 \in V_T$ such that t_1 is an ancestor of t_2. That is the connection between sub-community t_2 is asymptotically (with respect to n) denser than its super-community t_1.

The reason we define w and v in this seemly strange way is that we want to define $\mathcal{G}(n, T)$ in a way such that a same hierarchy tree $T = (V_T, E_T, w, v)$ is compatible with varying n. Thus, we need to encode each $w(t)$ and $v(t)$ in a way that is independent of V and n. The way we define w makes sure $w(t)$ is independent, and the large communities assumption above makes sure $v(t)$ is also independent. As we will see in the next subsection, we will take $n \to \infty$ when considering InfMax, and we would not like n or $|V|$ to be one of the inputs to the InfMax problem. As a result, T, being one of the inputs to the InfMax problem, should be independent of n. Notice that such property of the definition of InfMax enables us to consider graphs having exponentially many vertices.

6.2.2 The InfMax Problem

We study the r-complex contagion on the succinct stochastic hierarchical block model. Roughly speaking, given hierarchy tree T and an integer K, we want to choose K seeds which maximize the expected total number of infected vertices, where the expectation is taken over the graph sampling $G \sim \mathcal{G}(n, T)$ as $n \to \infty$.

Definition 6.3. The *influence maximization problem* InfMax is an optimization problem which takes as inputs an integer r, a hierarchy tree $T = (V_T, E_T, w, v)$ in

Definition 6.2, and an integer K, and outputs $\boldsymbol{k} \in \mathbb{N}_{\geq 0}^{|L_T|}$—an allocation of K seeds into the leaves L_T with $\sum_{t \in L_T} k_t = K$ that maximizes

$$\Sigma_{r,T}(\boldsymbol{k}) := \lim_{n \to \infty} \frac{\mathbb{E}_{G \sim \mathcal{G}(n,T)} \left[\sigma_{r,G}(S_{\boldsymbol{k}})\right]}{n},^{1}$$

the expected fraction of infected vertices in $\mathcal{G}(n,T)$ with the seeding strategy defined by $\boldsymbol{k}$, where $S_{\boldsymbol{k}}$ denotes the seed set in G generated according to $\boldsymbol{k}$.

Before we move on, the following remark is very important throughout the paper.

Remark 6.4. In Definition 6.3, n is not the input of the InfMax instance. Instead, the tree T is given as an input to the instance, and we take $n \to \infty$ to compute $\Sigma_{r,T}(\boldsymbol{k})$ *after* the seed allocation is determined. Therefore, asymptotically, all the input parameters to the instance, including K, r and the encoding size of T, are *constants* with respect to n. Thus, there are two different asymptotic scopes in this paper: *the asymptotic scope with respect to the input size* and *the asymptotic scope with respect to* n. Naturally, when we are analyzing the running time of an InfMax algorithm, we should use the asymptotic scope with respect to the input size. For example, an algorithm runs in polynomial time would mean the running time is a polynomial of the input size, not of n. On the other hand, when we are analyzing the number of infected vertices after the cascade, we should use the asymptotic scope with respect to n.

In this paper, we use $O_I(\cdot), \Omega_I(\cdot), \Theta_I(\cdot), o_I(\cdot), \omega_I(\cdot)$ to refer to the asymptotic scope with respect to the input size, and we use $O(\cdot), \Omega(\cdot), \Theta(\cdot), o(\cdot), \omega(\cdot)$ to refer to the asymptotic scope with respect to n. For example, $K = \Theta_I(|V_T|^2)$ means the number of seeds is asymptotically the square of the tree size, while with respect to n we always have $r = \Theta(1)$, $K = \Theta(1)$ and $|V_T| = \Theta(1)$. As another example, the two asymptotic notions, $|V(t)| = v(t)n = \Theta(n)$ and $w(t_1) = o(w(t_2))$, appeared in the large communities and the proper separation assumptions are all in the scope of n, so we have not put the subscripts I under Θ and o.

Lastly, we have assumed that $r \geq 2$, so that the contagion is nonsubmodular. When $r = 1$, the cascade model becomes a special case of the *independent cascade model* [128], which is a submodular cascade model. For submodular InfMax, a simple greedy algorithm is known to achieve a $(1-1/e)$-approximation to the optimal influence [128, 127, 173].

[1] The purpose we divided the expected number of infected vertices by n is to avoid the infinite limit. However, as a result of this, our analysis naturally ignores lower order terms.

6.2.3 Complex Contagion on Erdős-Rényi Graphs

In this section, we consider the r-complex contagion on the Erdős-Rényi random graph $\mathcal{G}(n,p)$. We review some results from [119] which are used in our paper.

Definition 6.5. The *Erdős-Rényi random graph* $\mathcal{G}(n,p)$ is a distribution of graphs with the same vertex set V with $|V| = n$ and we include an edge $(u,v) \in E$ with probability p for each pair of vertices u, v.

The InfMax problem in Definition 6.3 on $\mathcal{G}(n,p)$ is trivial, as there is only one possible allocation of those K seeds: allocate all the seeds to the single leaf node of T, which is the root. Therefore, $\sigma_{r,T}(\cdot)$ in Definition 6.3 depends only on the *number* of seeds $K = |\boldsymbol{k}|$, not on the seed allocation $\boldsymbol{k}$ itself. In this section, we slightly misuse the notation σ for a function mapping an *integer* to $\mathbb{R}_{\geq 0}$ (rather than mapping *an allocation of K seeds* to $\mathbb{R}_{\geq 0}$ as it is in Definition 6.3), and let $\sigma_{r,\mathcal{G}(n,p)}(k)$ be the expected number of infected vertices after the cascade given k seeds. Correspondingly, let $\sigma_{r,G}(k)$ be the actual number of infected vertices after the graph G is sampled from $\mathcal{G}(n,p)$.

Theorem 6.6 (A special case of Theorem 3.1 in [119])**.** *Suppose $r \geq 2$, $p = o(n^{-1/r})$ and $p = \omega(n^{-1})$. We have*

1. *if k is a constant, then $\sigma_{r,\mathcal{G}(n,p)}(k) \leq 2k$ with probability $1 - o(1)$;*
2. *if $k = \omega\left((1/np^r)^{1/(r-1)}\right)$, then $\sigma_{r,\mathcal{G}(n,p)}(k) = n - o(n)$ with probability $1 - o(1)$.*

Theorem 6.7 (Theorem 5.8 in [119])**.** *If $r \geq 2$, $p = \omega(n^{-1/r})$ and $k \geq r$, then $\Pr_{G\sim\mathcal{G}(n,p)}\left[\sigma_{r,G}(k) = n\right] = 1 - o(1)$.*

When $p = \Theta(n^{-1/r})$, the probability that k seeds infect all the n vertices is positive, but bounded away from 1. We use $\mathrm{Po}(\lambda)$ to denote the Poisson distribution with mean λ.

Theorem 6.8 (Theorem 5.6 and Remark 5.7 in [119])**.** *If $r \geq 2$, $p = cn^{-1/r} + o(n^{-1/r})$ for some constant $c > 0$, and $k \geq r$ is a constant, then*

$$\lim_{n\to\infty} \Pr\left(\sigma_{r,\mathcal{G}(n,p)}(k) = n\right) = \zeta(k,c),$$

for some $\zeta(k,c) \in (0,1)$. Furthermore, there exist numbers $\zeta(k,c,\ell) > 0$ for $\ell \geq k$ such that

$$\lim_{n\to\infty} \Pr\left(\sigma_{r,\mathcal{G}(n,p)}(k) = \ell\right) = \zeta(k,c,\ell)$$

for each $\ell \geq k$, *and* $\zeta(k,c) + \sum_{\ell=k}^{\infty} \zeta(k,c,\ell) = 1$.

Moreover, the numbers $\zeta(k,c,\ell)$*'s and* $\zeta(k,c)$ *can be expressed as the hitting probabilities of the following inhomogeneous random walk. Let* $\xi_\ell \in \mathrm{Po}\left(\binom{\ell-1}{r-1}c^r\right)$, $\ell \geq 1$ *be independent, and let* $\tilde{S}_\ell := \sum_{j=1}^{\ell}(\xi_j - 1)$ *and* $\tilde{T} := \min\{\ell : k + \tilde{S}_\ell = 0\} \in \mathbb{N} \cup \{\infty\}$. *Then*

$$\zeta(k,c) = \Pr\left(\tilde{T} = \infty\right) = \Pr\left(k + \tilde{S}_\ell \geq 1 \text{ for all } \ell \geq 1\right) \tag{6.1}$$

and $\zeta(k,c,\ell) = \Pr(\tilde{T} = \ell)$.

We have the following corollary for Theorem 6.8, saying that when $p = \Theta(n^{-1/r})$, if not all vertices are infected, then the number of infected vertices is constant. As a consequence, if the cascade spreads to more than constantly many vertices, then all vertices will be infected.

Corollary 6.9 (Lemma 11.4 in [119])**.** *If* $r \geq 2$, $p = cn^{-1/r} + o(n^{-1/r})$ *for some constant* $c > 0$, *and* $k \geq r$, *then*

$$\lim_{n\to\infty} \Pr\left(\phi(n) \leq \sigma_{r,\mathcal{G}(n,p)}(k) < n\right) = 0$$

for any function $\phi : \mathbb{Z}^+ \mapsto \mathbb{R}^+$ *such that* $\lim_{n\to\infty} \phi(n) = \infty$.

6.3 Our main result

Before presenting our main result, we need the following definition which defines the *density* of a leaf.

Definition 6.10. Given a hierarchy tree $T = (V_T, E_T, w, v)$ and a leaf $t \in L_T$, the *density* of the leaf $\rho(t) = w(t) \cdot (v(t)n)^{1/r}$.

Our main result is the following theorem, which states that the optimal seeding strategy is to put all the seeds in a community with the highest density, when the root has a weight in $\omega(1/n^{1+1/r})$.

Theorem 6.11. *Consider the* InfMax *problem with* $r \geq 2$, $T = (V_T, E_T, w, v)$, $K > 0$ *and the weight of the root node satisfying* $w(root) = \omega(1/n^{1+1/r})$. *Let* $t^* \in \arg\max_{t\in L_T} \rho(t)$ *and* $\boldsymbol{k}^*$ *be the seeding strategy that puts all the* K *seeds on* t^*. *Then* $\boldsymbol{k}^* \in \arg\max_{\boldsymbol{k}} \Sigma_{r,T}(\boldsymbol{k})$.

Notice that the assumption $w(\text{root}) = \omega(1/n^{1+1/r})$ captures many real life social networks. In fact, it is well-known that an Erdős-Rényi graph $\mathcal{G}(n,p)$ with $p = o(1/n)$

is globally disconnected: with probability $1 - o(1)$, the graph consists of a union of tiny connected components, each of which has size $O(\log n)$. Therefore, if we were to use the stochastic hierarchical block model to model the social networks in our daily life, we should expect that the root node has weight in $\Omega(1/n)$.

The remaining part of this section is dedicated to proving Theorem 6.11. We assume $w(\text{root}) = \omega(1/n^{1+1/r})$ in this section from now on. It is worthy noticing that, in many parts of this proof, and also in the proof of Theorem 6.18, we have used the fact that an infection of $o(n)$ vertices contributes 0 to the objective $\Sigma_{r,T}(\boldsymbol{k})$, as we have taken the limit $n \to \infty$ and divided the expected number of infections by n in Definition 6.3.

Definition 6.12. Given $T = (V_T, E_T, w, v)$, a node $t \in V_T$ is *supercritical* if $w(t) = \omega(1/n^{1/r})$, is *critical* if $w(t) = \Theta(1/n^{1/r})$, and is *subcritical* if $w(t) = o(1/n^{1/r})$.

From the results in the last section, if we allocate $k \geq r$ seeds on a supercritical leaf $t \in L_T$, then with probability $1 - o(1)$ all vertices in $V(t)$ will be infected; if we allocate k seeds on a subcritical leaf $t \in L_T$, at most a negligible amount of vertices, $2k = \Theta(1)$, will be infected; if we allocate $k \geq r$ seeds on a critical leaf $t \in L_T$, the number of infected vertices in $V(t)$ follows Theorem 6.8.

We say a node $t \in V_T$ is *activated* in a cascade process if the number of infected vertices in $V(t)$ is $v(t)n - o(n)$, i.e., almost all vertices in $V(t)$ are infected. Given a seeding strategy $\boldsymbol{k}$, let $P_{\boldsymbol{k}}$ be the probability that at least one node is activated when $n \to \infty$. Notice that this is equivalent to at least one leaf being activated. The proof of Theorem 6.11 consists of two parts. We will first show that, $P_{\boldsymbol{k}}$ completely determines $\Sigma_{r,T}(\boldsymbol{k})$ (Lemma 6.13). Secondly, to maximize $P_{\boldsymbol{k}}$, an optimal seeding strategy is to put all seeds on a single leaf with the maximum density (Lemma 6.14).

Lemma 6.13. *Given any two seeding strategies $\boldsymbol{k}_1, \boldsymbol{k}_2$, if $P_{\boldsymbol{k}_1} \leq P_{\boldsymbol{k}_2}$, then $\Sigma_{r,T}(\boldsymbol{k}_1) \leq \Sigma_{r,T}(\boldsymbol{k}_2)$.*

Lemma 6.14. *Let $\boldsymbol{k}$ be the seeding strategy that allocates all the K seeds on a leaf $t^* \in \arg\max_{t \in L_T}(\rho(t))$. Then $\boldsymbol{k}$ maximizes $P_{\boldsymbol{k}}$.*

Lemma 6.13 and Lemma 6.14 imply Theorem 6.11.

6.3.1 Proof Sketch of Lemma 6.13

We sketch the proof here, and the full proof is in the appendix.

Proof (sketch). Let E be the event that at least one leaf (or node) is activated at the end of the cascade. Theorem 6.7 and Corollary 6.9 imply that the number of infected vertices in a critical or supercritical leaf t can only be either a constant or $v(t)n$, and Theorem 6.6 indicates that a subcritical leaf will not have $\omega(1)$ infected vertices by a constant number of seeds (as it is in our setting $K = \Theta(1)$). Therefore, if E does not happen, we only have $o(n)$ infected vertices in V, regardless of the seeding strategy.

If E happens, we can show that the expected total number of infected vertices does not vary significantly by different seeding strategies. If a leaf t_1 is activated, the probability that a vertex $v \in V(t_2)$ is infected due to the influence of $V(t_1)$ is at least

$$\binom{v(t_1)n}{r} w(t)^r (1 - w(t))^{v(t_1)n-r} = \omega\left(n^r \left(\frac{1}{n^{1+\frac{1}{r}}}\right)^r \cdot 1\right) = \omega\left(\frac{1}{n}\right),$$

where in the above equality we have further assumed $w(t) = o(1/n)$ (which can only further reduce the probability if $w(t)$ were in $\Omega(1/n)$) so that $(1 - w(t))^{v(t_1)n-r} = \omega((1 - 1/n)^{v(t_1)n}) = \omega(1)$. Thus, there are $\omega(1/n) \cdot \Theta(n) = \omega(1)$ infected vertices in $V(t_2)$. Theorem 6.7 and Corollary 6.9 show that t_2 will be activated if t_2 is critical or supercritical. Therefore, when E happens, all the critical and supercritical will be activated. As for subcritical leaves, the number of infected vertices may vary, but Theorem 6.6 intuitively suggests that adding a constant number of seeds is insignificant (we will handle this rigorously in the full proof). Therefore, the expected total number of infections equals to the number of vertices in all critical and supercritical leaves, plus the expected number of infected vertices in subcritical leaves which does not significantly depend on the seeding strategy $\boldsymbol{k}$.

In conclusion, the number of infected vertices only significantly depends on whether or not E happens. In particular, we have a fixed fraction of infected vertices whose size does not depend on $\boldsymbol{k}$ if E happens, and a negligible number of infected vertices if E does not happen. Therefore, $P_{\boldsymbol{k}}$ characterizes $\Sigma_{r,T}(\boldsymbol{k})$, and a larger $P_{\boldsymbol{k}}$ implies a larger $\Sigma_{r,T}(\boldsymbol{k})$. □

6.3.2 Proof of Lemma 6.14

We first handle some corner cases. If $K < r$, then the cascade will not even start, and any seeding strategy is considered optimal. If T contains a supercritical leaf, the leaf with the highest density is also supercritical. Putting all the $K \geq r$ seeds in this leaf, by Theorem 6.7, will activate the leaf with probability $1 - o(1)$. Therefore, this strategy makes $P_{\boldsymbol{k}} = 1$, which is clearly optimal. In the remaining part of this

subsection, we shall only consider $K \geq r$ and all the leaves are either critical or subcritical. Notice that, by the proper separation assumption, all internal nodes of T are subcritical.

We split the cascade process into two phases. In Phase I, we restrict the cascade within the leaf blocks ($V(t)$ where $t \in L_T$), and temporary assume there are no edges between two different leaf blocks (similar to if $w(t) = 0$ for all $t \notin L_T$). After Phase I, Phase II consists of the remaining cascade process.

Proposition 6.15 shows that maximizing $P_{\boldsymbol{k}}$ is equivalent to maximizing the probability that a leaf is activated in Phase I. Therefore, we can treat T such that all the leaves, each of which corresponds to a $\mathcal{G}(n,p)$ random graph, are isolated.

Proposition 6.15. *If no leaf is activated after Phase I, then with probability* $1-o(1)$ *no vertex will be infected in Phase II, i.e., the cascade will end after Phase I.*

We sketch the proof here, and the full proof is omitted due to the space limit.

Proof (sketch). Consider any critical leaf t and an arbitrary vertex $v \in V(t)$ that is not infected after Phase I. Let K_{in} be the number of infected vertices in $V(t)$ after Phase I, and K_{out} be the number of infected vertices in $V \setminus V(t)$. If no leaf is activated after Phase I, Theorem 6.6 and Corollary 6.9 suggest that $K_{in} = O(1)$ and $K_{out} = O(1)$. The probability that v is connected to any of the K_{in} infected vertices in $V(t)$ can only be less than $w(t) = \Theta(n^{-1/r})$ conditioning on the cascade inside $V(t)$ does not carry to v, so the probability that v has a infected neighbors in $V(t)$ is $O(n^{-a/r})$. On the other hand, the probability that v has $r-a$ neighbors among the K_{out} outside infected vertices is $o(n^{-(r-a)/r})$. Therefore, the probability that v is infected in the next iteration is $\sum_{a=0}^{r-1} O(n^{-a/r}) \cdot o(n^{-(r-a)/r}) = o(1/n)$, and the expected total number of vertices infected in the next iteration after Phase I is $o(1)$. The proposition follows from the Markov's inequality. □

Since Theorem 6.6 suggests that any constant number of seeds will not activate a subcritical leaf, we should only consider putting seeds in critical leaves. In Proposition 6.16, we show that in a critical leaf t, the probability that the $(i+1)$-th seed will activate t conditioning on the first i seeds failing to do so is increasing as i increases. Intuitively, Proposition 6.16 reveals a super-modular nature of the r-complex contagion on a critical leaf, making it beneficial to put all seeds together so that the synergy effect is maximized, which intuitively implies Lemma 6.14.

Proposition 6.16. *Consider an Erdős-Rényi random graph* $\mathcal{G}(n,p)$ *with* $p = cn^{-1/r} + o(n^{-1/r})$*, and assume an arbitrary order on the* n *vertices. Let* E_k *be the event that*

seeding the first k vertices does not make all the n vertices infected. We have $\Pr(E_{k+2} \mid E_{k+1}) < \Pr(E_{k+1} \mid E_k)$ *for any* $k \geq r-1$.

Proof. Since the event E_{k+1} implies E_k, we have

$$\Pr(E_{k+1} \mid E_k) = \frac{\Pr(E_{k+1} \wedge E_k)}{\Pr(E_k)} = \frac{\Pr(E_{k+1})}{\Pr(E_k)},$$

and the inequality we are proving is equivalent to

$$\Pr(E_{k+2}) \Pr(E_k) < \Pr(E_{k+1}) \Pr(E_{k+1}). \tag{6.2}$$

We adopt the inhomogeneous random walk interpretation in Theorem 6.8, and view E_k as the following process: the random walk starts at $x = k$; in the i-th iteration, x moves to the right by 1 unit, and moves to the left by $\alpha_i \sim \mathrm{Po}\left(\binom{i-1}{r-1} c^r\right)$ units; $\Pr(E_k)$ is then the probability that the random walk reaches $x = 0$. In this proof, we let $\lambda_i = \mathrm{Po}\left(\binom{i-1}{r-1} c^r\right)$, and in particular, $\lambda_0 = \lambda_1 = \cdots = \lambda_{r-1} = 0$.

To prove (6.2), we consider two random walks in $\mathbb{Z}^2$, starting at $(k+2, k)$ and $(k+1, k+1)$ respectively. In each iteration i, we move from (x, y) to $(x-1+\alpha_i, y-1+\beta_i)$ where α_i and β_i are sampled from $\mathrm{Po}(\lambda_i)$ independently. If we hit the x-axis after a certain iteration t, then we stick to the x-axis, i.e., for any $i > t$, the update in the i-th iteration is from $(x, 0)$ to $(x - 1 + \alpha_i, 0)$; similarly, we stick to the y-axis and update $(0, y - 1 + \beta_i)$ if we hit the y-axis after certain iteration. Then, $\Pr(E_{k+2}) \Pr(E_k)$ (or $\Pr(E_{k+1}) \Pr(E_{k+1})$) is the probability that the random walk starting from $(k+2, k)$ (or $(k+1, k+1)$) reaches $(0, 0)$. Let A be the random walk starting from $(k+2, k)$, and B be the random walk starting from $(k+1, k+1)$. To prove (6.2), we define a coupling between the two random walks such that, 1) whenever A reaches $(0, 0)$, B also reaches $(0, 0)$, and 2) with positive probability, B reaches $(0, 0)$ but A fails to reach $(0, 0)$.

Before we define the coupling, we reinterpret the random walk a little bit by breaking down each *iteration* i into T *steps*:

- at step 0 of iteration i, we update (x, y) to $(x-1, y-1)$ (or $(x-1, y)$ if $y = 0$, $(x, y-1)$ if $x = 0$);
- at each step j for $j = 1, \ldots, T$, sample two Bernoulli random valuables independently, $\alpha_{ij} \sim \mathrm{Be}(\lambda_i/T)$, $\beta_{ij} \sim \mathrm{Be}(\lambda_i/T)$, and update from (x, y) to $(x + \alpha_{ij}, y + \beta_{ij})$ (notice that in each step we have at most 1 unit movement in each direction, and we can only move further away from both the x-axis and

y-axis);

- at the end of iteration i, if $x = 0$ (or $y = 0$), the random walk is stuck to the y-axis (or the x-axis) forever (notice that we only do this at the end of an iteration, i.e., after step T; in particular, if a random walk hits one of the x-axis or y-axis after step 0 in an iteration i, say, the x-axis, it will only stick to the x-axis if $\beta_{ij} = 0$ for all $j = 1, \ldots, T$).

Standard results from Poisson process indicate that, when $T \to \infty$, the effect of the T steps from 1 to T is equivalent as sampling $\alpha_i \sim \text{Po}(\lambda_i)$ and $\beta_i \sim \text{Po}(\lambda_i)$ (see, for example, Definition 8.4 and Theorem 8.7 in [166]).

Now we are ready to describe the coupling. Figure illustrations are available in Fig. 6.1 and Fig. 6.2. Let (x_{ij}^A, y_{ij}^A), and (x_{ij}^B, y_{ij}^B) be the coordinates for A and B respectively after iteration i step j. Similarly, let α_{ij}^A, and α_{ij}^B denote the x-direction movements of both walks, and β_{ij}^A, and β_{ij}^B denote the y-direction movements. The coupling consists of two phases.

Phase I A and B move in exactly the same way, i.e., $\alpha_{ij}^A = \alpha_{ij}^B$ and $\beta_{ij}^A = \beta_{ij}^B$, until one of the following two events happens.

Event (a) The current position of A and B are symmetric with respect to the line $y = x$, i.e., $x_{ij}^A - x_{ij}^B = y_{ij}^B - y_{ij}^A$ and $x_{ij}^A + x_{ij}^B = y_{ij}^A + y_{ij}^B$. Notice that (a) may happen in some middle step j of an iteration i. When (a) happens, we move on to Phase II(a).

Event (b) A hits the x-axis *at the end of an iteration*. Notice that this means A is then stuck to the x-axis forever. When (b) happens, we move on to Phase II(b)

It is important to notice that A is always below the line $y = x$ before (a) happens, so A will never hit the y-axis in Phase I. To see this, A can only have four types of movements in each step: lower-left $(x, y) \mapsto (x - 1, y - 1)$, up $(x, y) \mapsto (x, y + 1)$, right $(x, y) \mapsto (x + 1, y)$, and upper-right $(x, y) \mapsto (x + 1, y + 1)$. It is easy to see that, 1) A will never step across $y = x$ in one step, and 2) if A ever reaches the $y = x$ at (w, w), then A must be at $(w, w - 1)$ in the previous step. However, when A is at $(w, w - 1)$, B should be at $(w - 1, w)$ according to the relative position of A, B. In this case event (a) already happens.

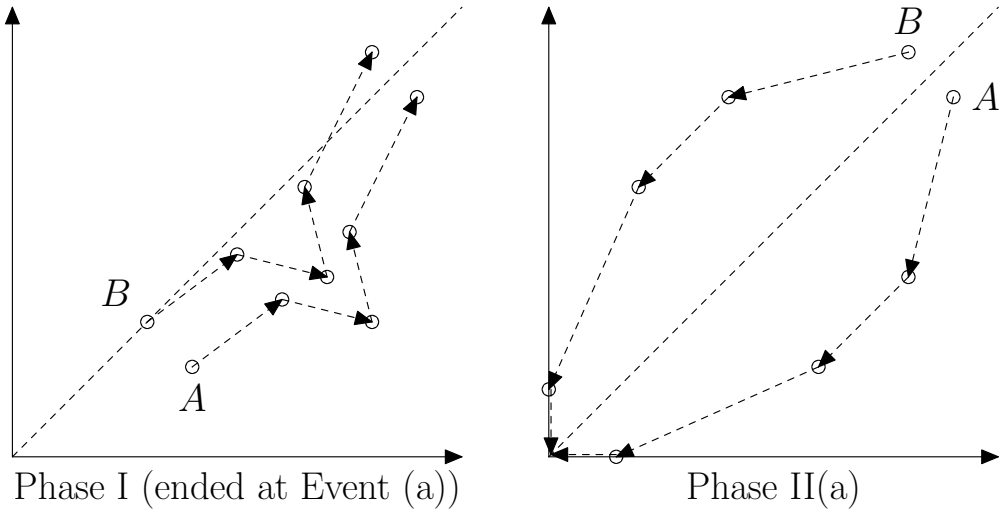


Figure 6.1: The coupling with Phase I ended at Event (a)

Phase II(a) A and B move in a way that is symmetric to the line $y = x$: $\alpha_{ij}^A = \beta_{ij}^B$ and $\beta_{ij}^A = \alpha_{ij}^B$. In this phase, by symmetry, A hits the x-axis if and only if B hits the y-axis, and A hits the y-axis if and only if B hits the x-axis. The coupling carries on when one of these happens. For example, if A hits the x-axis and B hits the y-axis, then we always have $\alpha_{ij}^A = \beta_{ij}^B$ and $\beta_{ij}^A = \alpha_{ij}^B = 0$. Notice that, in Phase II(a), A may cross $y = x$, after which A is above $y = x$ while B is below.

Phase II(b) We temporary freeze B's movement in y-direction, and let the x-direction movements of A and B be the same. To be specific, suppose Phase II(b) starts at iteration i_0. In each iteration $i \geq i_0$ and step j, we couple $\alpha_{ij}^A = \alpha_{ij}^B$, and we do not sample $\beta_{ij}^B \sim \text{Be}(\lambda_i/T)$ (β_{ij}^A is always 0 now, as A is stick to the x-axis). Till now, the relative position of A and B is always the same: $x_{ij}^A = x_{ij}^B + 1$ and $y_{ij}^A = y_{ij}^B - 1$. If B hits the y-axis at the end of an iteration i_1, we terminate the coupling. Notice that the current positions for A and B can only be that $(x^A, y^A) = (1, 0)$ and $(x^B, y^B) = (0, 1)$.

To prove Eqn. (6.2), it suffices to show that

1. if the coupling moves to Phase II(a), A reaches $(0, 0)$ if and only if B reaches $(0, 0)$;

2. with positive probability, the coupling moves to Phase II(b) and Phase II(b) terminates;

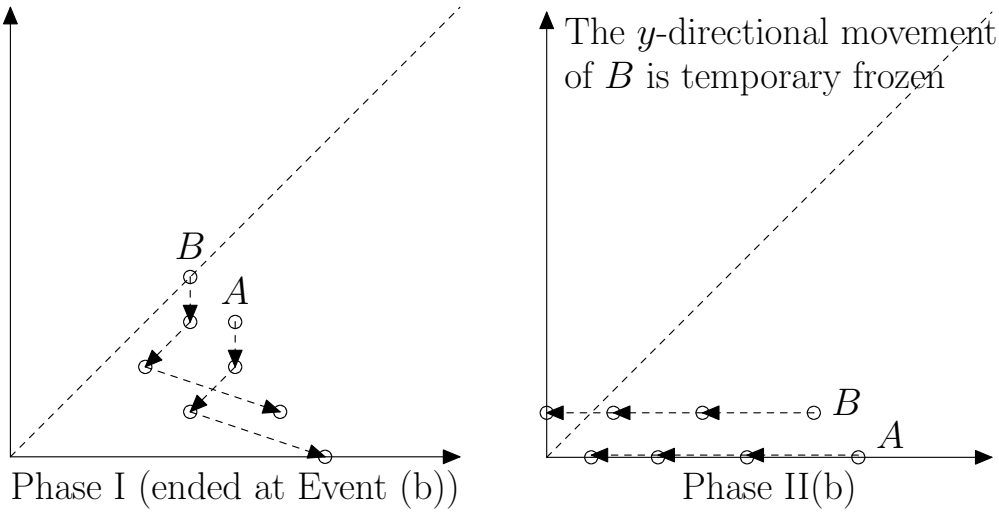


Figure 6.2: The coupling with Phase I ended at Event (b)

3. if the coupling moves to Phase II(b) and Phase II(b) terminates, then B has strictly higher chance to reach $(0,0)$ than A.

By symmetry, 1 is trivial. With positive probability $\beta_{ij}^A = 0$ for all $i = 1, \ldots, k$ and all $j = 1, \ldots, T$ (this is the probability that we independently sample k Poisson random variables with means $\lambda_1, \ldots, \lambda_k$ respectively, and obtain 0 for all of them), in which case event (b) happens, so 2 is also trivial. To see 3, by the time Phase II(b) terminates, A is stuck to the x-axis and one step away from $(0,0)$, while B is stuck to the y-axis and also one step away from $(0,0)$. Thus, we only need to consider a one-dimensional random walk for both A and B. In each future iteration $i_1 + t$ (for $t = 1, 2, \ldots$), A moves one step closer to $(0,0)$ and moves further by a distance sampled from $\mathrm{Po}(\lambda_{i_1+t})$, and meanwhile, taking into account that we have frozen B's y-direction movements between iterations i_0 and i_1, B also moves one step closer to $(0,0)$ and moves further by a distance sampled from $\mathrm{Po}(\lambda_{i_0+t})$. Since we have assumed $k \geq r-1$, it takes at least $r-1$ steps for A to reach the x-axis, so $i_0 \geq r-1$. Since λ_i is strictly increasing for $i \geq r-1$, we have $\lambda_{t_0+t} < \lambda_{t_1+t}$ for all $t \geq 0$. Therefore, in each future iteration after Phase II(b) terminates, the distance that B moves away from the destination is sampled from a Poisson distribution with mean strictly less than the mean of the Poisson distribution from which the distance that A moves away is sampled. This implies that B is strictly more likely to reach $(0,0)$ than A, which implies 3.

From 1, 2 and 3, we learn that the probability that B reaches $(0,0)$ is strictly larger than that of A, which implies (6.2) and concludes the proof. □

Equipped with Proposition 6.16, to show Lemma 6.14, we show that the seeding strategy that allocates $K_1 > 0$ seeds on a critical leaf t_1 and $K_2 > 0$ seeds on a critical leaf t_2 cannot be optimal. Firstly, it is obvious that both K_1 and K_2 should be at least r, for otherwise those K_1 (K_2) seeds on t_1 (t_2) are simply wasted.

Let E_k be the event that the first k seeds on t_1 fail to activate t_1 and F_k be the event that the first k seeds on t_2 fail to activate t_2. By Proposition 6.16, we have

$$\Pr(E_{K_1+1} \mid E_{K_1}) < \Pr(E_{K_1} \mid E_{K_1-1})$$

and

$$\Pr(F_{K_2+1} \mid F_{K_2}) < \Pr(F_{K_2} \mid F_{K_2-1}),$$

which implies

$$\begin{aligned}&\frac{\Pr(E_{K_1+1})\Pr(F_{K_2-1})}{\Pr(E_{K_1})\Pr(F_{K_2})} \cdot \frac{\Pr(E_{K_1-1})\Pr(F_{K_2+1})}{\Pr(E_{K_1})\Pr(F_{K_2})} \\ &= \frac{\Pr(E_{K_1+1} \mid E_{K_1})\Pr(F_{K_2+1} \mid F_{K_2})}{\Pr(E_{K_1} \mid E_{K_1-1})\Pr(F_{K_2} \mid F_{K_2-1})} < 1.\end{aligned}$$

Therefore, we have either $\frac{\Pr(E_{K_1+1})\Pr(F_{K_2-1})}{\Pr(E_{K_1})\Pr(F_{K_2})}$ or $\frac{\Pr(E_{K_1-1})\Pr(F_{K_2+1})}{\Pr(E_{K_1})\Pr(F_{K_2})}$ is less than 1. This means either the strategy putting $K_1 + 1$ seeds on t_1 and $K_2 - 1$ seeds on t_2 makes it less likely that none of t_1, t_2 is activated, or the strategy putting $K_1 - 1$ seeds on t_1 and $K_2 + 1$ seeds on t_2 makes it less likely that none of t_1, t_2 is activated, which implies that the strategy putting K_1 and K_2 seeds on t_1, t_2 respectively cannot be optimal. Therefore, in an optimal strategy, we should not allocate seeds on more than one leaf.

Finally, a critical leaf t with $v(t)n$ vertices and weight $w(t)$ can be viewed as an Erdős-Rényi random graph $\mathcal{G}(m, p)$ with $m = v(t)n$ and $p = w(t) = \rho(t) \cdot (v(t)n)^{-1/r} = \rho(t)m^{-1/r}$, where $\rho(t) = \Theta(1)$ when t is critical. Taking $c = \rho(t)$ in Theorem 6.8, we can see that ξ_ℓ has a larger Poisson mean if c is larger, making it more likely that the $\mathcal{G}(m, p)$ is fully infected (to see this more naturally, larger c means larger p if we fix m). Thus, given that we should put all the K seeds in a single leaf, we should put them on a leaf with the highest density. This concludes Lemma 6.14.

6.4 Optimal Seeds in Submodular InfMax

We have seen that putting all the K seeds in a single leaf is optimal for r-complex contagion, when the root node has weight $\omega(1/n^{1+1/r})$. To demonstrate the sharp difference between r-complex contagion and a submodular cascade model, we present

a submodular InfMax example where the optimal seeding strategy is to put no more than one seed in each leaf. The hierarchy tree T in our example meets all the assumptions we have made in the previous sections, including large communities, proper separation, and $w(\text{root}) = \omega(1/n^{1+1/r})$, where r is now an arbitrarily fixed integer with $r \geq 2$.

We consider a well-known submodular cascade model, *the independent cascade model* [128], where, after seeds are placed, each edge (u, v) in the graph appears with probability p_{uv} and vertices in all the connected components of the resultant graph that contain seeds are infected. In our example, the probability p_{uv} is the same for all edges, and it is $p = 1/n^{1-\frac{1}{4r}}$. The hierarchy tree T contains only two levels: a root and K leaves. The root has weight $1/n^{1+\frac{1}{2r}}$, and each leaf has weight 1. After $G \sim \mathcal{G}(n, T)$ is sampled and each edge in G is sampled with probability p, the probability that an edge appears between two vertices from different leaves is $(1/n^{1-\frac{1}{4r}}) \cdot (1/n^{1+\frac{1}{2r}}) = o(1/n^2)$, and the probability that an edge appears between two vertices from a same leaf is $1 \cdot (1/n^{1-\frac{1}{4r}}) = \omega(\log n/n)$. Therefore, with probability $1 - o(1)$, the resultant graph is a union of K connected components, each of which corresponds to a leaf of T. It is then straightforward to see that the optimal seeding strategy is to put a single seed in each leaf.

6.5 A Dynamic Programming Algorithm

In this section, we present an algorithm which finds an optimal seeding strategy when $w(t)$ satisfies either $w(t) = \omega(1/n^{1+1/r})$ or $w(t) = o(1/n^2)$ for each $t \in V_T$, and we will assume this for $w(t)$ throughout this section. Since a parent node always has less weight than its children (see Definition 6.1), we can decompose T into *the upper part* and *the lower part*, where the lower part consists of many subtrees whose roots have weights in $\omega(1/n^{1+1/r})$, and the upper part is a single tree containing only nodes with weight in $o(1/n^2)$ and whose leaves are the parents of those roots of the subtrees in the lower part. We call each subtree in the lower part a *maximal dense subtree* defined formally below.

Definition 6.17. Given a hierarchy tree $T = (V_T, E_T, w, v)$, a subtree rooted at $t \in V_T$ is a *maximal dense subtree* if $w(t) = \omega(1/n^{1+1/r})$, and either t is the root, or $w(t') = O(1/n^{1+1/r})$ where t' is the parent of t.

Since we have assumed either $w(t) = \omega(1/n^{1+1/r})$ or $w(t) = o(1/n^2)$, $w(t') = O(1/n^{1+1/r})$ in the definition above implies $w(t') = o(1/n^2)$.

The idea of our algorithm is the following: firstly, after the decomposition of T into the upper and lower parts, we will show that the weights of nodes in the upper part, falling into $w(t) = o(1/n^2)$, are negligible so that we can treat the whole tree T as a forest with only those maximal dense subtrees in the lower part (that is, we can remove the entire upper part from T); secondly, Theorem 6.11 shows that when we have decide the number of seeds to be allocated for each maximal dense subtree, the optimal seeding strategy is to put all the seeds together in a single leaf that has the highest density, where the density of a leaf $t \in L_T$ is defined in Definition 6.10; finally, the only problem remaining is how to allocate the K seeds among those maximal dense subtrees, and we decide this allocation by a dynamic programming approach.

Now, we are ready to describe our algorithm, presented in Algorithm 6.1.

Input: $r \in \mathbb{Z}$ with $r \geq 2$, $T = (V_T, E_T, w, v)$, and $K \in \mathbb{Z}^+$

1: Find all maximal dense subtrees $T_1, \ldots, T_m$, and let $r_1, \ldots, r_m$ be their roots.

2: For each T_i and each $k = 0, 1, \ldots, K$, let $\boldsymbol{s}_i^*(k)$ be the seeding strategy that puts k seeds in the leaf $t \in L_{T_i}$ with the highest density, and let

$$h(T_i, k) = \lim_{n\to\infty} \frac{\mathbb{E}_{G\sim\mathcal{G}(v(r_i)\cdot n, T_i)}[\sigma_{r,G}(\boldsymbol{s}_i^*(k))]}{n}$$

be the expected number of infected vertices in the subgraph defined by T_i, normalized by the total number of vertices in the whole graph.

3: Let $S[i, k]$ store a seeding strategy that allocates k seeds in the first i subtrees $T_1, \ldots, T_i$, and let $H[i, k]$ be the expected total number of infected vertices corresponding to $S[i, k]$, divided by n.

4: **for** $k = 0, 1, \ldots, K$ **do**

5: set $S[1, k] = \boldsymbol{s}_i^*(k)$ and $H[1, k] = h(T_1, k)$.

6: **end for**

7: **for** each $i = 2, \ldots, m$ **do**

8: **for** $k = 0, 1, \ldots, K$ **do**

9: $k_i = \underset{k_i\in\{0,1,\ldots,k\}}{\arg\max}\ H[i-1, k-k_i] + h(T_i, k_i)$;

10: set $S[i, k]$ be the strategy that allocates $k - k_i$ seeds among $T_1, \ldots, T_{i-1}$ according to $S[i-1, k-k_i]$ and puts the remaining k_i seeds in the leaf of T_i with the highest density;

11: set $H[i, k] = H[i-1, k-k_i] + h(T_i, k_i)$;

12: **end for**

13: **end for**

14: **Output:** the seeding strategy $S[m, K]$.

Program 6.1: The InfMax algorithm

The correctness of Algorithm 6.1 follows immediately from Theorem 6.18 (below) and Theorem 6.11. Recall Theorem 6.18 shows that we can ignore the upper part of

T and treat T as the forest consisting of all the maximal dense subtrees of T when considering the InfMax problem. Theorem 6.11 shows that for each subtree T_i and given the number of seeds, the optimal seeding strategy is to put all the seeds on the leaf with the highest density.

Theorem 6.18. *Given $T = (V_T, E_T, w, v)$, let $\{T_1, \ldots, T_m\}$ be the set of all T's maximal dense subtrees and let T^- be the forest consists of $T_1, \ldots, T_m$. For any seeding strategy $\boldsymbol{k}$ and any $r \geq 2$, we have $\Sigma_{r,T}(\boldsymbol{k}) = \Sigma_{r,T^-}(\boldsymbol{k})$.*

Proof. Let $V(T_i)$ be the set of vertices corresponding to the subtree T_i. Since the total number of possible edges between those $V(T_i)$'s is upper bounded by n^2 and each edge appears with probability $o(1/n^2)$, the expected number of edges is $o(1)$. By Markov's inequality the probability there exists edges between those $V(T_i)$'s is $o(1)$. Therefore, we have

$$\frac{1}{n}\mathbb{E}_{G\sim\mathcal{G}(n,T)}\left[\sigma_{r,G}(\boldsymbol{k})\right] = \frac{1}{n}\left(o(1)O(n) + (1-o(1))\mathbb{E}_{G\sim\mathcal{G}(n,T')}\left[\sigma_{r,G}(\boldsymbol{k})\right]\right).$$

Taking $n \to \infty$ concludes the proof. □

Finally, it is straightforward to see the time complexity of Algorithm 6.1, in terms of the number of evaluations of $\Sigma_{r,\mathcal{G}(n,T)}(\cdot)$.

Theorem 6.19. *Algorithm 6.1 requires $O_I(|V_T|K^2)$ computations of $\Sigma_{r,\mathcal{G}(n,T)}(\cdot)$.*

CHAPTER 7

Sybil Detection Using Strong Ties and Latent Network Structure

Sybil attacks, in which an adversary creates a large number of identities, present a formidable problem for the robustness of recommendation systems. One promising method of sybil detection is to use data from social network ties to imerlicitly infer trust.

Previous work along this dimension typically 1) Assumes that it is difficult/costly for an adversary to create edges to honest nodes in the network. 2) Limit the amount of damage done for each such edge, using conductance-based methods. However, these methods fail to detect a simple class of sybil attacks which have been identified in online systems. Indeed, conductance-based methods seem inherently unable to do so, as they are based on the assumption that creating many edges to honest nodes is difficult, which seems to fail in real-world settings.

We create a sybil defense system that accounts for the adversary's ability to launch such attacks yet provably withstands them by:

1. Not assuming any restriction on the number of edges an adversary can form, but instead making a much weaker assumption that creating edges from the sybil nodes to most of the honest nodes is difficult, but allowing that the remaining nodes can be freely connected to.

2. Relaxing the goal from classifying all nodes as honest or sybil to the goal of classifying the "core" nodes of the network as honest, and avoiding classifying any sybil nodes as honest.

3. Exploiting a property of social networks, that nodes can be embedded in low-dimensional spaces, that has not previously been used in Sybil detection.

7.1 Introduction

The creation of multiple false identities, so-called sybil attacks [76], can enable actors undo influence in recommendation systems or other algorithms that harness user-generated data [167]. Controlling even just a small portion of the allege user-base can enable nefarious actors to hide their ill-gotten influence over recommendation systems [247]. Such recommendation systems might be used to classify spam, recommend products, or filter user-generated content (e.g. on an online-social networking site). Due to society's increasing reliance on the results of harnessing user-generated content/feedback (e.g. "big data"), guarding the veracity of the results will become increasingly important. Manipulation can have economically important (such as product recommendation) and politically important (as a public show of support) outcomes which provides rational actors incentives to manipulate outcomes to match their desires.

This has been recognized as a problem and addressed in the literature via a variety of methods (see Section 7.1.2). This paper focuses on a particularly promising method of using network ties to (implicitly) infer trust.

The models of prior work tend to restrict the adversary by making an ***edge-limiting*** assumption: the number of ties that the adversary can forge between sybils and honest nodes is restricted [245, 246, 226, 68, 236].

Armed with the edge-limiting assumption and additionally assuming that the honest nodes of a network are "well-connected," these works show that one of two outcomes occurs: A) The adversary does not create many sybils; B) The adversary creates many sybils, but there is a detectable "sparse cut" in the graph. This sparse cut is caused by the assumption that there are few edges between the many sybil nodes and the honest nodes. Moreover, it is unique due to the assumption that the honest nodes are well-connected.

Thus, even if a powerful adversary can create many sybils, and moreover, endow them with high degree by connecting them with each other, the adversary cannot well integrate the sybils back into the rest of the network due to the limited number of ties that the adversary can forge between sybils and honest nodes.

While this defence does indeed (provable) protect against certain types of sybil attacks, the edge-limiting assumption seems to be too strong in practice [10]. Indeed Yang et al [242] recently showed evidence that in the RenRen social network, sybil attacks did not look like those that the prior work was anticipating, but instead were characterized by isolated sybils connected by many edges to honest nodes. We call

these ***periphery attacks*** for reasons that will be made clear shortly. In periphery attacks, the number of sybils is only a fraction of the number of edges, yet Yang et al [242] found many sybil nodes in such an attack pattern. As such, these attacks violate the edge-limiting assumption; so the guarantees of the conductance-based sybil defences appear not to apply. Indeed Alvisi et al [10] showed via simulation on a real network, that the conductance-based defences do a poor job defending against such attacks.

Such attacks seem difficult to attenuate, in particular because often time the majority of nodes in a social network have a similar appearance. For example Leskovec et al [142] showed that networks have a "core/periphery" structure, with many nodes on the periphery poorly connected to the core of the network, which was difficult to partition. Additionally, Yardi et al [243] showed that in Twitter, the majority of nodes in twitter only had a few friends, and that the spammers looked like-wise. Alvisi et al [10] looked into a collection of network topology properties and showed that the only one that was useful to sybil detection is conductance, which failed in thwarting periphery attacks.

7.1.1 Our Contribution

We create a framework that accounts for the adversary's ability to launch periphery attacks. Additionally, we create a network topology based sybil defense system that both accounts for and provably withstands periphery attacks. Our work builds upon and advances prior work in three main ways:

1. We replace the edge-limiting assumption with a new assumption: A random fraction of the honest nodes are ***compromisable*** and can easily be tricked into connecting with sybil nodes; but the remainder of the honest nodes are ***trustworthy*** and will refuse connections from sybils. With such an assumption, periphery attacks are easy for an adversary to launch. The adversary can test which nodes are gullible, and then connect to them at will with his sybil network.

2. We relax the goal from classifying all nodes as honest or sybil to the goal of classifying the "core" nodes of the network as honest; and classifying no sybil nodes as honest. Our model acknowledges the difficult of differentiating between the "periphery" nodes of the honest network and nodes that are part of a coordinated periphery sybil attack. Indeed this seems impossible to do with only information about network topology.

3. We identify a new network property namely, that they can be embedded in low-dimensional spaces as useful to detecting sybils. For a sybil to "blend in" with the core of the topology structure of a network it is not enough that he has many ties; rather the sybil needs a large number of ties amongst other nodes that are "close" in the network. A sybil that connects to random nodes, will not have a "location" in the network the way an honest node might.

Starting with Watts and Strogatz [234] works from the mathematical, computer science, sociology, and statistics communities have also mathematically modeled social networks as coming from a low-dimension latent space and use the guiding principal that nodes which are "closer" in the latent space are more likely to be attached [2, 57, 91, 109, 112, 130, 133, 137, 194, 203, 202]. Furthermore, there is evidence of the accuracy of such models [3, 19, 44, 149, 159, 168]. Thus, while new to sybil detection literature, our model is well grounded in the social network literature.

For many applications, like learning algorithms, or implicit community voting algorithms, having white-listed nodes is enough [10]. The system needs a representative sample of nodes. If the nodes on the periphery are not counted, then, as long as the nodes in "core" are sufficiently numerous, the system can succeed. Also, the zero false positives is important because even a few sybils can distort recommendations [247]. For other applications (e.g. spam), such a classification might not to enough. There other tools must be used (e.g. user feed-back on spam; setting participation limitations for new nodes, etc).

7.1.2 Related Work

Well-mixed networks A growing number of works look to using a network topology to aid in sybil detection.

Yu et al create SybilGuard [245] and SybilLimit [246], which use a random walk technique to bound the number of sybils that an adversary can produce for each edge that they can produce to honest nodes. This bound is $O(\sqrt{n}\log(n))$ for SybilGuard and was improved to $O(\log(n))$ in SybilLimit. However, in our setting where we do not restrict the number of edges that sybils can make to honest but gullible nodes, these guarantees are empty.

These works are typically called "conductance-based" and require an assumption that the network of honest nodes is well-mixing (and thus has high conductance). The intuition is that if there are many sybil nodes, but not many edges between the sybils

and the honest nodes, then these algorithms will find a sparse cut. The well-mixing assumption is required to ensure that this sparse cut is unique.

Since these original works, several others works have made improvements along certain dimensions. Danezis et al [68] create SybilInfer which, instead of classifying nodes as safe or unsafe, using Bayesian reasoning, outputs confidence. Unlike aforementioned conductance-based work SybilInfer is a centralized algorithm. They point out that the run times of prior, distributed work are very slow because they detect one sybil at a time and show that SybilInfer scales better. Likewise, Wei [236] et al propose SybilDefender which uses random walks, but is centralized and has improved scaling properties. They also suggest looking at tie strength as a method for improving results. Tran et al [226] propose Gatekeeper which achieves the same worst-case bound as SybilLimit, but improves upon it when the number of honest-sybil edges is very small.

Clustered Honest Networks All of theses works must assume that the network among honest nodes is well-mixing. The SybilLimit [246] paper provides some empirical evidence for this, but the claim is generally disputed. For example, Viswanath et al [231] analyze the state of current network-based Sybil defenses, showing that they rely on local community structure, and have trouble when their are cuts in the honest networks because they have difficulty distinguishing between the natural partitions in network of the honest nodes, and the sparse cuts between the sybils and honest nodes. They propose borrowing techniques from the community detection literature.

Alvisis et al [10] also believe that the network will be too fragmented to employ the previous techniques, and show rigorous theoretical bounds to substantiate this claim. Without the "well-mixing" assumption, they fear the problem may be intractable as distinguishing between honest and sybil communities seems impossible. For example, consider the extreme case where all communities, both sybil and honest, are small and disjoint. Instead of sybil detection, they suggested "personalized white-lists". They point out that there is no need to distinguish between sybil and honest communities as long as you use the recommendation of each community for the nodes in it. A draw-back of this is that if some communities are small, there may not be enough data to provide optimal recommendations. Like Alvisis et al, this work provides a white-listing strategy. However, we provide a global (not local) white list, and the honest nodes our model cannot classify are nodes on the periphery that belong to no community.

Cai and Jermaine [45] also address the problem of potential community structure

within the honest nodes. There algorithm first partitions the network into disjoint communities, and then tries to ferret out the honest communities from the sybil communities by embedding them into a low dimensional space. They argue that the sybil communities will be on the periphery of this latent community graph. To get this result, their model assumes 1) the network of honest and sybil nodes partitions into well-structured and detectable communities, 2) that honest nodes connect to nodes in other communities according to a latent network of communities, and 3) that some communities are easy for sybils to attach to, while other communities are difficult for sybils to attach to. Our work differs in several ways. Most fundamentally, their algorithm does not guard against periphery attacks. In fact, their model does not allow periphery attacks because they make a necessary (in their setting) edge-limiting assumption. Moreover, they use machine learning techniques and thus do not obtain rigorous security results. Finally, our network models differ: our model of latent structure applies to the nodes and not communities; and in our model which nodes are vulnerable is decided at the node level rather than the community level.

Other strategies There are also an increasing sequence of work looking at information beyond the social graph such as users click-stream data [233]; entry and exit times [181], number of rejected friend requests [10], etc [242]. It is clear that they currently provide large practical benefits [242]. Moreover, they can be usefully combined with network topology based techniques [10]. Thus it seems like this is a useful orthogonal direction to prosue in ensuring the validity of recommendations. However, a key disadvantage of many of these techniques is that the rely on an uninformed adversary, that does not understand the behavior of honest nodes well enough to mimic them. Thus, there usefulness may wane as they are increasing deployed and understood.

Another approach is to integrate sybil detection together with opinion aggregation (e.g. SumUp[225]). A key advantage here is that the sybil nodes do not have to be completely eliminated; but instead can be "down-weighted". However, a disadvantage of such approaches is that if they depend too sharply on the specific aggregation method, they loose some generality.

Another, someone disjoint, line of inquiry is for setting where a central authority can restrict the entry of sybils through some verification or payment (e.g. Captchas [232] or Netflix). And defense in sensor networks [155, 244] where, the solutions concept offered is light-weight cryptography (so that it can be efficiently executed).

7.2 Preliminaries

A ***metric space*** is an ordered pair $M = (V, d)$ where V is a set and d is a metric on V mapping $V \times V$ to $\mathbb{R}^+$ such that for any $u, v, w \in V$, the following holds: $d(u, v) \geq 0$; $d(u, u) = 0$; $d(u, v) = d(v, u)$; and $d(u, v) \leq d(u, w) + d(w, v)$. We say that $M' = (V', d')$ is a ***metric subspace*** of $M = (V, d)$ if $V' \subseteq V$ and $d' = d|_{V' \times V'}$. We only consider finite metric spaces, i.e. $|V| \in \mathbb{N}$.

A ***metric graph*** $G = (V, E, d)$ is an undirected graph with distances defined between all pairs in V such that (V, d) is a metric space.

We define $B_M(u, r) = \{x \in V : d(u, x) < r\}$ as a ball with radius r centered at u in metric space M. We will often drop the subscript when it is clear from context, and denote $B(u, 1)$ by $B(u)$.

To capture the idea of low dimension in such a metric space, we use the notion of doubling dimension defined as follows: the ***doubling dimension*** $dim(M)$ of a metric space $M = (V, d)$ is the minimum k such that every ball of radius r is covered by 2^k balls of radius $R/2$; i.e. $\forall c \in V, r > 0, B(c, r) \subseteq V$, there exists $c_1, c_2, ..., c_m$ where $m \leq 2^k$ such that $B(c, r) \subseteq \bigcup_i B(c_i, r/2)$.

The doubling dimension is a very general definition of dimension. When it is applied to Euclidean vector spaces, it recovers the usual definition of dimension, but it also can apply to arbitrary metric spaces. Additionally, note that all finite metric spaces have finite doubling dimensions.

We define the ***neighbors*** of u in metric graph (V, E, d) to be $N(u) = \{v : (u, v) \in E\}$, and the ***core neighbors*** of u to be $CN(u) = B(u) \cap N(u)$, i.e the neighbors of u at distance at most 1.

7.3 Sybil Detection Framework

7.3.1 Metric Space Properties

We first define some properties of a metric space $M = (V, d)$ which we will make use of throughout.

Definition 7.1. The ***density*** of a metric space is $den(M) = \min_{u \in V} |B_u|$ which is the minimum cardinality of a unit ball.

Definition 7.2. We say that U is an r***-code*** of a metric space $M = (V, d)$ if $U \subseteq V$ and $\forall u, v \in U, d(u, v) > r$ and $V \subseteq \bigcup_{u \in U} B(u, r)$. That is U is a maximal set of points of distance strictly more than r from each other.

Definition 7.3. We define the ***volume*** of a metric space $M = (V, d)$ to be $vol(M) = \max\{|U| : U \text{ is a 2-code of } M\}$.

We show a natural relation between the density, the volume, and the cardinality of a metric space.

Lemma 7.4. *Let $M = (V, d)$ be a metric space with density $den(M)$ and volume $vol(M)$. Then*

$$den(M) \cdot vol(M) \leq |V|.$$

Proof. Let Y be a 2-code of M such that $|Y| = vol(M)$. On the one hand we have that

$$den(M) \cdot vol(M) \leq \sum_{y \in Y} |B(y)|$$

because for any $v \in V$, $den(M) \leq |B(y)|$ (by Definition 7.1) and $vol(M) = |Y|$ (by Definition 7.3).

On the other hand, we have that

$$\sum_{y \in Y} |B(y)| = |\bigcup_{y \in Y} B(y)| \leq |V|$$

because the $B(y)$ are disjoint—recall that for all $x, y \in Y$ we have $d(x, y) > 2$—and $\bigcup_{y \in Y} B(y) \subseteq V$. □

Here we provide an efficient algorithm to compute an approximation of the largest 2-code.

Lemma 7.5. *Let $M = (V, d)$ be a metric space and $dim(M) = k$, then there exists a polynomial algorithm f, such that $f(M)$ is a 2-code and $\frac{vol(M)}{4^k} \leq |f(M)| \leq vol(M)$.*

Proof. Let Y be the maximum 2-code of M, then by definition $|Y| = vol(M)$. The algorithm f iteratively inserts a node x into X, and removes all the nodes in $B(x, 2)$. Therefore each pair in X has distance more than 2, and $|X| \leq |Y| = vol(M)$ by definition.

On the other hand, consider a 1-code Z of metric space M, because $\forall u, v \in Y, d(u, v) > 2$, every unit ball of Z contains at most one $y \in Y$. Thus

$$|Y| \leq |Z|.$$

Moreover because $dim(M) = k$, $B(x, 2)$ can be covered by 4^k $\frac{1}{2}$ balls, and each $\frac{1}{2}$ ball can contains at most 1 element in Z. Thus

$$|Z| \leq 4^k |X|.$$

Putting this together $|X| \leq |Y| = vol(M) \leq Z \leq 4^k |X|$ which yields the lemma. □

Definition 7.6. Given a metric space $M = (V, d)$, we define a graph $H_r(M) = (V, E)$ where $(u, v) \in E$ if $d(u, v) \leq r$.

Definition 7.7. If $H_1(M)$ is connected, we say a metric space M is ***hyper-connected***.

This characterizes the metric space as "well connected" so that for all pairs of nodes there exists a sequence of points such that the distance between each pair of consecutive nodes is less than 1.

Definition 7.8. We say that $\hat{M} = (\hat{V}, \hat{d})$ is a ***core space*** with density Δ of a metric space $M = (V, d)$ if $\hat{M}$ is a submetric of M; density $\Delta = \min_{v \in \hat{V}} |B_M(v)|$; and $H_1(\hat{M})$ is connected.

This idea of a core space is important, because we only hope to classify nodes in the "core" of the network, not those in the periphery. This is a somewhat connected region with density above some threshold.

7.3.2 Network of Honest Nodes

In this section we both highlight exactly what we require of honest networks and provide motivation for this model.

We will consider metric graphs that are generated on top of a metric $M = (V, d)$ on n points. We would like that these points a) have doubling dimension bounded by some parameter k; and b) have a "large" core space $\hat{M}$ with density Δ where Δ is again a parameter.

The edges of the graph are generated by including each possible edge (u, v) where $d(u, v) \leq 3$, with probability ρ. Any additional edges may then be added to the graph after the outcomes of these random edges are realized.

Recapping, the important parameters are n, the number of nodes; k the doubling dimension; Δ the density of the core; and ρ, the minimum probability that edges appear between nodes close in the metric.

We think that this is a rather general model that is well-justified. First, the assumption that nodes are embedded in a low-dimensional space where nearby nodes are connected is implicit in many well-regarded network models. For example, in the Watts-Strogatz model [234] nodes are arranged on ring (which is just a one-dimensional lattice) and any two nodes within some distance d on the ring, are connected via an edge with some probability that is a parameter of the model. Similarly, Kleinberg's Small World Model [133] has the nodes embedded into a low dimensional lattice structure where nodes are connected to neighbors. Additionally, Kumar et al. [140] allows an arbitrary metric space with low doubling dimension and requires an additional property which is similar to our core space requirement. Though the latent space Abraham et al. [2] considers is not necessary a metric space, our method can be easily applied to their model, because once having the distance function of all categories, we can removed individuals which fail to have enough common neighbors in all categories.

A host of other works from the mathematical, computer science, sociology, and statistics communities have also mathematically modeled social networks as coming from a low-dimension latent space and use the guiding principal that nodes which are "closer" in the latent space are more likely to be attached [2, 57, 91, 109, 112, 130, 137, 194, 203, 202].

The intuition behind these models is that the location of a node in a metric space encodes some key properties of the individual, e.g. geographic location, income, political beliefs on a spectrum, education level, etc; and that these attributes are sufficient so that when individuals are "close" in this space, they are likely (with probability ρ) to be friends. Notice that in most of the aforementioned models, nodes are *always* neighbors with the nearby nodes in the metric; where as we only require that nearby nodes are neighbors with some constant, non-zero probability.

Furthermore, there is evidence of the accuracy of such models [3, 19, 44, 149, 159, 168]. In Section 7.6, we provide our own experimental result which confirms that, for the networks we look at, they can be fruitfully embedded in a low-dimensional latent space. An additional feature of our model is that additional edges may be added to the graph in any, even *adversarial*, manner.

Second, our model additionally requires that the nodes be sufficiently dense in the metric. Notice that most of the aforementioned models have the nodes spread out uniformly, so their are no sparse regions of the network. We additionally relax this assumption and only require that there is a "large" dense region. To a first approximation, this dense region is the area we will be able to white-list; while nodes

in sparse regions may not be included in the white list. The necessity of dealing with sparse regions is empirically motivated by aforementioned findings of Leskovec et al. [142], Alvisi et al. [10], and Yardi et al. [243] which all identify nodes on the periphery with low-degree and/or that can be disconnected from the network by only removing a few edges.

7.3.3 Detection Game

In this section we propose a formal model for sybil detection as a game with two agents: the *adversary* and the *distinguisher*.

The adversary will be given a metric graph G. We say that the nodes of G are the ***honest*** nodes. This set of honest nodes is partitioned into a set of ***compromisable*** nodes C that the adversary can attach to and a set of ***trustworthy*** nodes T that the adversary cannot attach to. The adversary must output a new metric graph G' which is the same as G except that the adversary can add up to Σ ***sybil*** nodes and any edges that it likes except those between trustworthy nodes and sybil nodes.

The distinguisher will then be given the adversary's output graph (as well as some parameters), and must create a white-list of as many nodes as possible without including any sybil nodes.

Definition 7.9. Let $A : (G, C, p, \rho, \Sigma) \to G'$ be a (possibly random) function where $G = (V, E, d)$ and $G' = (V', E', d')$ are metric graphs, $C \subseteq V$ is a set of "compromisable" nodes, p, ρ are real values between 0 and 1, and $\Sigma > 0$. We say that A is an ***adversary*** if for every input G, C, p, ρ, Σ:

1. $|S| < \Sigma$ where $S = V' \setminus V$.
2. The distance function d' is a metric that extends d to $V \cup S$.
3. $E \subseteq E'$ but E' contains no edges from $V \setminus C$ to S. However E' may contain additional edges between V and itself, between S and itself, and between C and S.

Our definition limits the adversary in two keys ways: first, he can only introduce so many sybil nodes. Such a condition is necessary because otherwise the adversary could just create a completely new graph on a disjoint set of vertices which is identical to the original graph; no detection algorithm could distinguish the ordinal graph from the identical facsimile. Second, the adversary can only connect sybils to the original network via compromisable nodes. The intuition is that some set of nodes

can by tricked or bribed into connecting with the sybils. The remaining vertices are more trustworthy, concerned, aware, and/or vigilant and are thus immune from the adversaries attempts to connect. This aligns with the observations of Yang et al. [242] that software toolkits which facilitate the creating of sybil nodes for the Renren cite were available and would attempt to identify network nodes that would likely accept a sybil's tie request (e.g. nodes with extremely large degree).

Note especially that the adversary can also add ties between honest nodes. This is not meant to model that the adversary could or would actually compel honest nodes to add a tie (though it does capture this as well). Rather it is meant to model that, apart from the ties in the network that we assume to exist from the low-dimensional embedding (that are included in G and cannot be removed), the rest of the graph is adversarial bad. In actuality, we think that the graph on the honest nodes would come from nature. However, we do not wish to prescribe anything more about the honest graph other than that nodes which are "close" in the low-dimensional latent space are often connected; and may be connected in a way that is not helpful to the "distinguisher."

Definition 7.10. A ***distinguisher*** D is a (possibly random) function which maps $(G', p, den(M), vol(M))$ to W where $G' = (V', E', d')$ is a metric graph, p, $den(M)$, $vol(M)$ are real valued parameters, and $W \subseteq V'$.

Now we formally define a *detection game* on a metric space $M = (V, d)$.

Definition 7.11. We define a ***detection game*** Γ with input $(M, p, \rho, \Sigma, A, D)$ where M is a metric space, p, ρ are real values between 0 and 1, $\Sigma > 0$, A is an adversary, and D is a defender as follows,

1. Based on $M = (V, d)$, a metric graph $G = (V, E, d)$ is instantiated where E is created by independently including each edges (u, v) with probability p if $d(u, v) < 3$, and otherwise with probability 0. [Note that in Step 3, the adversary can add *any* additional ties it likes between honest nodes in an attempt to thwart the distinguisher. At that point the adversary knows which nodes are trustworthy and compromisable, so the additional edges can depend on those labels.]

2. We randomly partition V into two sets T (for trusthworthy) and C (for compromisable). Each agent $v \in V$ will, independently, be included in set C with probability ρ and in set T otherwise.

3. The adversary A creates a new metric graph $G' = A(G, C, p, \rho, \Sigma)$.

4. The distinguisher D outputs a list of nodes W with input $(G', p, den(M), vol(M))$

5. If $W \subseteq V$ we say that the distinguisher ***succeeds with score*** $|W|$; otherwise, if $W \cap S \neq \emptyset$ we way that the distinguisher *fails*.

We note that we give the distinguisher help via the parameters $p, den(M), vol(M)$. In general, we do not feel this assumption is overly restrictive, as distinguisher could likely learn these over time.

We also node that the detection game maps onto our definition of honest networks in Section 7.3.2. In particular, this gives the adversely (perhaps unrealistic) power to manipulate the graph of honest nodes by adding additional edges between any pair of vertices even after the random edges have been realized and the compromisable nodes have been determined. However, this only makes our results stronger.

7.4 Sybil Detection Algorithm

Theorem 7.12. *Fix* $0 < \epsilon < \frac{1}{\sqrt{2}}$ *and let* $\Gamma(M, p, \rho, \Sigma, A, D)$ *be a detection game where* p, ρ *are probabilities such that* $\frac{1+\epsilon}{1-\epsilon}\rho < p$, $0 \leq \Sigma$, *and* M *is a metric space that has* n *nodes and doubling dimension* k *with core-space* $\hat{M} = (\hat{V}, \hat{d})$ *with density* Δ *with* $m = |\hat{V}|$. *Then if*

$$\Sigma < (1-\epsilon)\frac{p}{2 \cdot 128^k} den(\hat{M}) \cdot vol(\hat{M}) - (1+\epsilon)\rho n$$

there exists a detection algorithm D *such that for any adversary* A *the detection algorithm* D *will succeed with score at least* m *with probability*

$$1 - n^2 \exp(-\frac{\epsilon^2}{2} p\Delta) - n \exp(-\frac{\epsilon^2}{3}\rho\Delta) - \exp(-\frac{\epsilon^2}{3}\rho n).$$

Note that the size of the white-list is at least as large as the dense core of M. The parameters of the theorem can cover a variety of settings. For example, if $\Delta = \omega(\frac{\log(n)}{p\epsilon^2})$ and $\epsilon^2\rho = o(n/\log n)$, then the probability of error is negligible (less than the inverse of any polynomial).

To the end of proving Theorem 7.12, we propose the *detection algorithm* which is specified in Algorithm 7.1.

Before we dig into the proof we sketch the intuition behind the detection algorithm. Verification goes as follow: the algorithm pretends that there is no sybil node in the

Input: $G' = (V', E', d')$, p, and $\Delta, vol(\hat{M})$
Output: W, denoting the white-listed nodes.

Find a 2-code Y of $H_2(V', d')$ by the algorithm in Lemma 7.5.
Obtain (V'', E'', d'') from G' by iteratively finding nodes $u \in V'$ where $|N_{G'}(u) \cap B_{G'}(u, 2)| < (1 - \epsilon)p \cdot \Delta$ and removing these nodes and all incident edges.
for $y \in Y$ **do**
  $G_y(V_y, E_y, d_y) \leftarrow (V'', E'', d'')$, $W_y \leftarrow \emptyset$ and $U_y \leftarrow \emptyset$
  while $U_y = \emptyset$ and $|B_{G_y}(y)| \geq \Delta$ or $\exists v \in U_y$ such that $\exists u \in B_{G_y}(v) \setminus U_y$ where $|B_{G_y}(u)| \geq \Delta$ **do**
    if $U_y = \emptyset$ **then**
      $u \leftarrow y$
    else
      Set u to be some u from Step 5
    end if
    $U_y \leftarrow U_y \cup \{u\}$
    for $v \in B_{G_y}(u, 2)$, and $v \notin W_y$ **do**
      if $|N_{G_y}(v) \cap B_{G_y}(u)| > (1 - \epsilon)p|B_{G_y}(u)|$ **then**
        $W_y = W_y \cup \{v\}$
      else
        Remove v and all its edges from G_y.
      end if
    end for
  end while
end for

Program 7.1: Detection algorithm

starting region $B(y)$ for some y from Step 3 and attempts to certify nodes $v \in B(y, 2)$ by checking whether they have many neighbors in $B(y)$. Then the algorithm moves to a different ***center*** u in Step 11 and verifies the region $B(u, 2)$. Doing this, it will iteratively remove the sybils on the boundary; allowing it to grow a white-listed region in the graph to cover the entire core.

The remaining difficulty is to find a good starting point y. In Step 3, we say $y \in V$ is a ***good starting point*** if $B(y) \cap S = \emptyset$ and $|B(y)| \geq \Delta$, and say $y \in V$ is a ***bad starting point*** if $B(y) \cap S \neq \emptyset$ and $|B(y)| \geq \Delta$. The main idea is that the adversary cannot corrupt every region of the graph with many nodes. Thus after Step 2 there will be many regions of the graph with no sybils. In Step 1, we get a maximal independent set corresponding to a 2-code of (V', d') which ensures that we are exploring many diverse regions of the network.

The proof can be separated into two parts:

1. (completeness/soundness) If y from Step 3 is a good starting point, then with high probability, this algorithm will white-list every honest node in the core space and no sybil nodes will be white-listed;

2. (majority) There are many $y \in Y$ that are good starting points, and not too many bad starting points.

We first prove three lemmas about structural properties of the network that occur with high probability. The first of these lemmas shows that if node v is near a node u with many nodes within unit distance, then node v has large degree. The second says that if node v has many nodes within unit distance, then v does not (fractionally) have too many compromisable nodes within unit distance. The third lemma bounds the total number of compromisable nodes.

We will then show that if these properties hold, then our detection algorithm succeeds.

Lemma 7.13. *Let $\Gamma(M, p, \rho, \Sigma, A, D)$ be a detection game, let $n = |M|$, and let $\Delta \in \mathbf{R}_{\geq 0}$. Then with probability $1 - n^2 \exp(-\frac{\epsilon^2}{2} p\Delta)$ for every $u, v \in M$ with $d(u, v) \leq 2$ and $|B_M(u)| \geq \Delta$, it is the case that $|N_G(v) \cap B_M(u)| \geq (1 - \epsilon)p|B_M(u)|$.*

Proof. By Definition 7.11 in Step 1, the edges will independently form between v and nodes in $B(v, 3)$. Because the adversary cannot remove any edges, he can only increase $|N_{G'}(v) \cap B_M(u)|$.

We known $\forall w \in B_M(u), d(v, w) \leq d(v, u) + d(u, w) \leq 2 + 1 = 3$. Thus $Pr[w \in N_{G'}(v)] \geq p$ and $E[|N_{G'}(v) \cap B_M(u)|] = \sum_{w \in B_M(u)} Pr[w \in N_{G'}(v)] \geq p|B_M(u)|$.

Moreover because the realization of edges are independent, we can apply a Chernoff bound and get

$$Pr[|N_{G'}(v) \cap B_M(u)| \leq (1-\epsilon)p|B_M(u)|] \tag{7.1}$$

$$\leq \exp(-\frac{\epsilon^2}{2}p|B_M(u)|) \leq \exp(-\frac{\epsilon^2}{2}p\Delta) \tag{7.2}$$

The lemma follows from a union bound. □

Lemma 7.14. *Let $\Gamma(M, p, \rho, \Sigma, A, D)$ be a detection game, let $n = |M|$, and let $\Delta \in \mathbf{R}_{\geq 0}$. Then with probability $1-n\exp(-\frac{\epsilon^2}{3}\rho\Delta)$ for every $u \in M$ with $|B_M(u)| \geq \Delta$, it is the case that $|B_M(u) \cap C| \leq (1+\epsilon)\rho|B_M(u)|$.*

Proof. In Definition 7.11, $\forall u \in B_M(u)$ whether v is in C are independently random indicator variables and $|B_M(u) \cap C|$ is the sum of these indicator variables with expectation $\rho|B_M(u)|$. Applying a Chernoff bound,

$$Pr[|B_M(u) \cap C| > (1+\epsilon)\rho|B_M(u)|] \tag{7.3}$$

$$\leq \exp(-\frac{\epsilon^2}{3}\rho|B_M(u)|) \tag{7.4}$$

$$\leq \exp(-\frac{\epsilon^2}{3}\rho\Delta) \tag{7.5}$$

Applying union bound and taking the complement, the lemma follows. □

Lemma 7.15. *Let $\Gamma(M, p, \rho, \Sigma, A, D)$ be a detection game and let $n = |M|$, then with probability $1 - \exp(-\frac{\epsilon^2}{3}\rho n)$, $|C| < (1+\epsilon)\rho n$.*

Proof. Because the expectation of $|C|$ is ρn and it is the sum of independent indicator variables, we can apply a Chernoff bound and get:

$$Pr[|C| \geq (1+\epsilon)\rho n] < \exp(-\frac{\epsilon^2}{3}\rho n)$$

□

Now notice that by a union bound, the statements of Lemmas 7.13, 7.14, and 7.15 holds with probability $1 - n^2\exp(-\frac{\epsilon^2}{2}p\Delta) - n\exp(-\frac{\epsilon^2}{3}\rho\Delta) - \exp(-\frac{\epsilon^2}{3}\rho n)$.

We now assume that all these statements hold, and show that when this is the case, our detection algorithm works. The next lemma shows that no honest node within unit distance of a node with high density is removed in Step 2.

Lemma 7.16. *Let $\Gamma(M, p, \rho, \Sigma, A, D)$ be a detection game where D is our detection algorithm with inputs G', p, Δ, vol. Let $v \in V$ with $|B_G(v)| \geq \Delta$ then, assuming statement of Lemma 7.13 holds, after Step 2, $B_{G''}(u) \cap V = B_G(u) \cap V$.*

Proof. Let $W = \bigcup_{v \in V : |B_G(v)| \geq \Delta} B_G(v)$. We claim that there can be no "first" $w \in W$ that is removed from V'. For each $w \in W$ let v_w be a node such that $w \in B_G(v_w)$ and $|B_G(v_w)| \geq \Delta$. By the definition of w, we know that such a node v_w exists, and that $v_w \in W$.

For the sake of contradiction say that w is the first $w \in W$ removed. Then

$$\begin{aligned} &N_G(w) \cap B_G(v_w, 1) \\ &\subseteq N_{G'}(w) \cap B_{G'}(v_w, 1) \\ &\subseteq N_{G'}(w) \cap B_{G'}(w, 2) \end{aligned}$$

where the first $\subseteq$ is because $E \subseteq E'$, and the second $\subseteq$ is because $B_{G'}(v_w, 1) \subseteq B_{G'}(w, 2)$.

However, by Lemma 7.14, we have that

$$(1 - \epsilon)p\Delta \leq (1 - \epsilon)p|B_G(v_w, 1)| \leq |N_G(w) \cap B_G(v_w, 1)|.$$

This is a contradiction because in this case $|N_{G'}(w) \cap B_{G'}(w, 2)| \geq (1 - \epsilon)p\Delta$ and so w would not have been removed in Step 2. □

Lemma 7.17. *Let M be a metric space and let $\hat{M}$ be a core space with density Δ. Let $\Gamma(M, p, \rho, \Sigma, A, D)$ be a detection game where D is our detection algorithm with inputs $G', p, \Delta, vol(\hat{M})$. Assume that the conditions in Lemmas 7.13 and Lemma 7.14 are true, and let y be a good starting point. Then the Detection algorithm will output $W_y \subseteq V$. Moreover, if $y \in \hat{M}$ then $\hat{V} \subseteq W_y \subseteq V$*

Proof. We assume the statements of Lemma 7.13 and Lemma 7.14 and that y is a good starting point and then we will show that the following always hold:

1. $V_y \cap V = V'' \cap V$,
2. $W_y \cap S = \emptyset$,
3. For all $u \in U_y$ and $u' \in B_{G_y}(u)$ where either $|B_{G_y}(u')| \geq \Delta$ or $|B_G(u')| \geq \Delta$, we have $B_{G_y}(u') = B_G(u') \subseteq W_y$.

If we prove this, then, by the second statement, we know that $W_y \subseteq V$. We must also show that if $y \in \hat{M}$ then $\hat{M} \subseteq W_y$. We show something stronger: each node in $\hat{M}$ is eventually included in U_y. This is a stronger statement because, by Statement 3, if $u \in U_y$, then $B_{G_y}(u) = B_G(u) \subseteq W_y$. Say that some node $w \in \hat{M}$ is never added to U_y. By the hyper-connection property of $\hat{M}$ we can create a spanning tree on the nodes of $H_1(\hat{M})$ rooted at y, and let w be a "closest" node to y (in the tree) that is not included and let v be its parent.

However, from the third statement above, we know $B_G(w) = B_{G_y}(w)$ because $d(w, v) \leq 1$, $v \in U_y$ and $|B_G(w)| \geq \Delta$. Thus w will also be processed as a center, and this is a contradiction.

We now show that the three properties always hold via induction on $|U_y|$. For $|U_y| = 0$, the first statement holds because at that point $V_y = V''$; while the second statement holds because $W_y = \emptyset$ and the third statement holds because $U_y = \emptyset$.

We now show the inductive step, that if the three statements hold when $|U_y| = k$, they will also hold when $|U_y| = k + 1$.

Lets say that u is the $k + 1$st node chosen for a center in Step 11. We know that $B_{G_y}(u) = B_G(u)$ either because $u = y$ and then it follows from the fact that y is a good starting point and Lemma 7.16, or because there must exist $w \in U_y$ such that $d(w, u) \leq 1$, and then it follows from the third assumption (note that $|B_{G_y}(u)| \geq \Delta$ because u was chosen to be a center).

Before processing center u, a node $v \in V_y(u) \cap V$ has two cases:

1) $v \notin B_{G_y}(u, 2)$ then v will certainly be in V_y after the process;

2) if $v \in B_{G_y}(u, 2)$, since Lemma 7.13 holds, we have $|N_G(v) \cap B_G(u)| \geq (1 - \epsilon)p|B_G(u)|$. Because $B_{G_y}(u) = B_G(u)$, we have also have $|N_{G_y}(v) \cap B_{G_y}(u)| \geq (1 - \epsilon)p|B_{G_y}(u)|$. Thus $v \in V_y$ holds after the process, and that proves $V_y \cap V = V'' \cap V$, and $B_{G_y}(u, 2) \subseteq W_y$.

On the other hand, let $s \in B_{G_y}(u, 2)$ be a sybil node. Then s can only connect to the compromised nodes in $B_{G_y}(u)$ because, by assumption, $B_{G_y}(u) = B_G(u)$, which contains no sybil nodes. Formally, we see:

$$\begin{aligned} |N_{G_y}(s) \cap B_{G_y}(u)| &\leq |C \cap B_{G_y}(u)| = |C \cap B_G(u)| \\ &< (1 - \epsilon)\rho|B_G(u)| = (1 - \epsilon)\rho|B_{G_y}(u)| \\ &\leq (1 + \epsilon)p|B_{G_y}(u)|. \end{aligned}$$

The first equality is from the assumption that $B_{G_y}(u) = B_G(u)$, the second inequality is from the assumption of Lemma 7.14, and the final inequality is because

$\frac{1+\epsilon}{1-\epsilon}\rho < p$. This proves $W_y \cap S = \emptyset$.

It remains to show that part 3) holds. We break the analysis into two cases by partitioning U_y into $U_y \setminus \{u\}$ and $\{u\}$.

First, let $v \in U_y \setminus \{u\}$ and let $u' \in B_{G_y}(v)$ where $|B_{G_y}(u')| \geq \Delta$ or $|B_G(u')| \geq \Delta$. Then, by the inductive hypothesis, after the time the v was processed, we had that $B_{G_y}(u') = B_G(u') \subseteq W_y$. No node in W_y is ever removed, so this still must be the case.

Second, let $u' \in B_{G_y}(u)$ with $|B_{G_y}(u')| \geq \Delta$ or $|B_G(u')| \geq \Delta$ and so that u' was not considered above. Then, we must show $B_{G_y}(u') = B_G(u') \subseteq W_y$.

Note that combining the facts that $B_{G_y}(u') \subseteq B_{G_y}(u, 2)$ and $B_{G_y}(u, 2) \subseteq W_y$ (argued above) we see, that $B_{G_y}(u') \subseteq W_y$. Using that $B_{G_y}(u') \subseteq W_y$ and $W_y \cap S = \emptyset$ we see that $B_{G_y}(u') \cap S = \emptyset$, which means that $B_{G_y}(u')$ has no sybils and so $B_{G_y}(u') \subseteq B_G(u')$. This additionally implies that $|B_G(u')| \geq \Delta$.

It remains to show that $B_G(u') \subseteq B_{G_y}(u')$. Intuitively, the one problem we could encounter is that some nodes of $B_G(u')$ might have been removed in Step 2. However, this does not happen. Rather $B_G(u') = B_{G''}(u') \cap V$ because $|B_G(u')| \geq \Delta$ and so by Lemma 7.16 $B_{G''}(u') \cap V = B_G(u')$. We use this to get:

$$\begin{aligned} B_G(u') =& B_{G''}(u') \cap V = B_{G''}(u') \cap V'' \cap V \\ =& B_{G''}(u') \cap V_y \cap V = B_{G_y}(u') \cap V \subseteq B_{G_y}(u') \end{aligned}$$

The third equality is because $V'' \cap V = V_y \cap V$, as proved above.

Putting everything together we have $B_G(u') = B_{G_y}(u') \subseteq W_y$ and this concludes the proof of the lemma. □

Lemma 7.18. *(Majority) Let $\Gamma(M, p, \rho, \Sigma, A, D)$ be a detection game and assume that the condition in Lemma 7.15 is true, and let Y be the 2-code D gets after step 1, then at most $\frac{|Y|}{2 \cdot 4^k} \leq \frac{vol(\hat{M})}{2 \cdot 4^k}$ points in Y are bad starting points.*

Proof. Suppose the lemma is false. Then we consider the subset $Y' \subseteq Y$ such that every $y' \in Y'$ is a bad starting point and

$$|Y'| \geq \frac{|Y|}{2 \cdot 4^k}.$$

We consider some $X \subseteq Y'$ such that X is a 8-code for Y'. For each $x \in X$, $|B(x, 8) \cap Y'| \leq 8^k$ because $B(x, 8)$ can be covered by less than 8^k unit balls by

definition of doubling dimension and each unit ball contains at most 1 element of Y'. The cardinality of 2-code Y is greater than $\frac{vol(\hat{M})}{4^k}$ by Lemma 7.5. Thus

$$|X| \geq \frac{|Y'|}{8^k} \geq \frac{|Y|}{2 \cdot 32^k} \geq \frac{vol(\hat{M})}{2 \cdot 128^k} \tag{7.6}$$

By the assumption of Y' every $x \in X \subseteq Y$ is a bad starting point which means $\forall x \in X, \exists s_x \in B(x)$ which is a sybil node, and since s_x survives after step 2, $|B(s_x, 2)| > (1-\epsilon)p \cdot \Delta$. Moreover, for all $x, z \in X$ and $x \neq z$, $d(s_x, s_z) \geq d(x,z) - d(x,s_x) - d(z,s_z) > 8 - 4 = 4$, $B(s_x, 2), B(s_z, 2)$ are disjoint. On one hand,

$$|C \cup S| \geq |\bigcup_x B(s_x, 2)| \geq |X|(1-\epsilon)p \cdot \Delta$$

Using (7.6) and the condition on S, we get

$$|C| \geq -|S| + (1-\epsilon)p\frac{vol(\hat{M})\Delta}{2 \cdot 128^k} > (1+\epsilon)\rho n$$

On the other hand, by Lemma 7.15, $|C| \leq (1+\epsilon)\rho n$, and so we get a contradiction. □

Now we can prove the Theorem 7.12

Proof. First, we note that the statements of Lemmas 7.13, 7.14, and 7.15 hold with probability

$$1 - n^2 \exp(-\frac{\epsilon^2}{2}p\Delta) - n\exp(-\frac{\epsilon^2}{3}\rho\Delta) - \exp(-\frac{\epsilon^2}{3}\rho n) \tag{7.7}$$

In the case that y is a good starting point, we never add a sybil nodes to W_y by Lemma 7.17. By Lemma 7.18, there are only $\frac{|Y|}{2 \cdot 4^k}$ bad starting points in Y. Thus no sybil meets the threshold in Step 5 to be included in W.

However, by Lemma 7.5, for any node v in the $\hat{M}$-core, there are $\frac{vol(\hat{M})}{4^k}$ start nodes in Y. Moreover, less than $\frac{vol(\hat{M})}{2 \cdot 4^k}$ of them can be bad. Thus at least $\frac{vol(\hat{M})}{2 \cdot 4^k}$ of them are good. By Lemma 7.17 for these y, $\hat{M} \subseteq W_y$ and thus, $\hat{M}$ will be included in W. □

7.5 Sybil Detection with a Trustworthy Seed

In this section, we will make the additional assumption that the distinguisher is given one trustworthy node as advice. The advantage is that in this setting, we do not have to limit the number of sybil nodes (nor the doubling dimension). However,

to take advantage of this advice, we will lose a fair bit in the trade-off between the parameters of ρ and p—the fraction of nodes that are compromisable and the fraction of edges present between honest nodes which are close in the underlying latent space, respectively. Before we required that $\frac{1+\epsilon}{1-\epsilon}\rho < p$, but in this section, we will require $\frac{1+\epsilon}{1-\epsilon}\rho < p^3$.

To formally define this new setting we modify Definition 7.9 (Adversary) to additionally require the adversary to choose a single trustworthy node t which will be used as advice from by the distinguisher. The idea is that the adversary can choose a "worst-case" trustworthy seed.

Definition 7.19. Let $A : (G, C, p, \rho) \to (G', t)$ be a (possibly random) function where $G = (V, E, d)$ and $G' = (V', E', d')$ are metric graphs, $C \subseteq V$ is a set of "compromisable" nodes, p, ρ are real values between 0 and 1, and $t \in V'$ is a node in the output graph. We say that A is an ***seeding-adversary*** if for every input G, C, p, ρ:

1) The distance function d' is a metric that extends d to $V \cup S$.

2) $E \subseteq E'$ but contain no edges from $V \setminus C$ to S. However E' may contain edges between V and V, between C and S, and between S and S.

3) $t \in V \setminus C$.

Next we must modify Definition 7.10 (Distinguisher) to take as input a single node t (which it will assume is trustworthy) in lieu of $vol(M)$.

Definition 7.20. A ***seeded-distinguisher*** D is a (possibly random) function $D : (G', p, den(M), vol(M), t) \to W$ where $G' = (V', E', d')$ is a metric graph, $p, den(M), vol(M)$ are real valued parameters, and $W \subseteq V'$.

Finally, we accordingly modify Definition 7.11 (Detection Game).

Definition 7.21. We define a ***seeded-detection game*** identically to the *detection game* (see Definition 7.11) except that in Step 3 and Step 4 the adversary and the distingisher are replaced by a seeding-adversary and a seeded-distinguisher, respectively.

In our Theorem, will only require that the distinguisher succeed if t is in the core of the network.

Theorem 7.22. *Fix $0 < \epsilon < \frac{1}{\sqrt{2}}$ and let $\Gamma(M, p, \rho, A, D)$ be a seeded-detection game where p, ρ are probabilities such that $\frac{1+\epsilon}{1-\epsilon}\rho < p^3$, and M is a metric space that has n nodes with core-space $\hat{M} = (\hat{V}, \hat{d})$ with density Δ with $m = |\hat{V}|$. Then there exists a*

detection algorithm D such that against adversary A that outputs a trustworthy node $t \in \hat{M}$, D will succeed with score at least m with probability

$$1 - 2n^2 \exp(-\frac{\epsilon^2}{2} p\Delta) - n \exp(-\frac{\epsilon^2}{3} \rho\Delta).$$

The idea is rather straightforward, given the prior results. From Lemma 7.17 we know that we can succeed if we can find a good starting point. The idea is then to make t into a good starting point.

Proof. (Sketch) Our distinguisher will simply run a slightly modified version of the Detection Algorithm. On input $(G', p, den(W), t)$, it will

1) Obtain V'' by removing all nodes in $B_{V'}(t)$ where the number of neighbors in $B_{V'}(t) \cap N_{V'}(t)$ is fewer than $(1-\epsilon)p^2\Pi(t)$ where $\Pi(t) = |B_{V'}(t) \cap N_{V'}(t)|$.

Then it will set y to t in Step 3 and return the W_t it obtains.

We first show that with probability $1 - n^2 \exp(-\frac{\epsilon^2}{2} p^2\Delta) - n \exp(-\frac{\epsilon^2}{3} \rho\Delta)$ that for all trustworthy nodes t in core space, if t is output by the seeding-adversary then:

1) all the sybils in $B_{V'}(t)$ are removed; and

2) all the honest nodes in $B_{V'}(t)$ remain.

If node $v \in H$ and $v \in B_{V'}(t)$, in expectation $|N_{V'}(v) \cap N_{V'}(t) \cap B_{V'}(t)| \geq |N_V(v) \cap N_V(t) \cap B_V(t)| \geq p^2|B_V(t)| \geq p^2\Pi(t)$ where the second inequality holds because adversary cannot remove edges between honest nodes. By a Chernoff bound, with probability at most $\exp(-\frac{\epsilon^2}{2} p^2\Delta)$, $|N_{V'}(v) \cap N_{V'}(t) \cap B_{V'}(t)| \leq (1-\epsilon)p^2\Pi(t)$ and v would be removed.

Taking a union bound over all possible t and v we get error with probability at most $n^2 \exp(-\frac{\epsilon^2}{2} p^2\Delta)$.

On the other hand, by Lemma 7.14 we have that for all t: $|C \cap B_V(t)| \leq (1+\epsilon)\rho\Pi(t)$ with probability at least $1 - n \exp(-\frac{\epsilon^2}{3} \rho\Delta)$. Note that

$$(1+\epsilon)\rho|B_V(t)| = (1+\epsilon)\frac{\rho}{p}|B_V(t) \cap N_V(t)| \leq (1+\epsilon)\frac{\rho}{p}\Pi(t) < (1-\epsilon)p^2\Pi(t).$$

Because the common neighbors between any sybil and trustworthy node t are compromisable nodes, if Lemma 7.14 is satisfied, then any sybil in $B_{V'}(t)$ will be removed.

At this point, if $t \in \hat{M}$ then t is a good starting point because $S \cap B_{V''}(t) = \emptyset$ and $|B_{V''}(t)| \geq \Delta$.

While the situation is slightly changed, and we cannot directly apply Lemma 7.17, it is straightforward to check that the exact same proof applies. The only difference come from not having run Step 2 in the Detection Algorithm. Note that Lemma

7.17 requires that both Lemma 7.13 and Lemma 7.14 are satisfied. We have already accounted for the small probability that Lemma 7.14 is not satisfied. Noting that Lemma 7.13 is satisfied with probability $1 - n^2 \exp(-\frac{\epsilon^2}{2} p\Delta)$ we get the theorem. □

We also consider the additional assumption that the distinguisher is given one trustworthy node as advice. In the full version, we show that we can obtain similar results as Theorem 7.12 but with no limit on the number of sybil nodes (nor the doubling dimension). However, to take advantage of this advice, we will lose a fair bit in the trade-off between the parameters of ρ and p—the fraction of nodes that are compromisable and the fraction of edges present between honest nodes which are close in the underlying latent space, respectively. Before we required that $\frac{1+\epsilon}{1-\epsilon}\rho < p$, but in this case, we will require $\frac{1+\epsilon}{1-\epsilon}\rho < p^3$.

7.6 Experiments

We previously noted that our assumptions hold in many generative models, and general versions of these assumptions are typically assumed to be true. We conducted several experiments to evaluate and further study our specific assumptions in different online communities and social networks.

In our main theorems, we assume the social network G can be embedded into a low doubling dimension space $M(G, d)$; that a large fraction of nodes forms a *core space* with density Δ, that every node in the core space has at least Δ nodes whose distance is smaller than 3, and the edges between the node and nearby nodes form independently with probability p. Collectively, we refer to this set of assumptions our *low-dimension assumption*.

Note that because our low dimensional assumptions are stochastic in nature, we cannot exactly test them empirically. Instead we will ensure that a node connects to some p fraction of neighbors within distance 3 (rather than a random set of neighbors).

7.6.1 Dataset Description

Our experiments use all 4 social network data sets on Stanford Network Analysis Project with between $4,000$ and $100,000$ nodes: this includes networks collected from Facebook [157] and Twitter [157] as well as the Wikipedia voting network [143] and the Epinion network [196]. The Facebook data set consists of 'circles' (or 'friends lists') from Facebook, and was collected from survey participants using a certain Facebook app. The Twitter data has similar attributes to Facebooks and it was crawled from

public sources. Wikipedia is a free encyclopedia written collaboratively by volunteers around the world. The promotion to adminship of Wikipedia community is done through public discussion or voting and an edge represents one volunteer voting for another. This data set collects $2,794$ elections with $103,663$ total votes and $7,066$ users participating in the elections. Epinion data set is a who-trusts-whom online social network of a general consumer review site. Members of the site can decide whether to "trust" each other. These networks vary in size and degree distributions and provide insight on the effect network parameters have on the low dimension latent space assumption.

We summarize the basic statistics of the network data sets we used in Table 7.1.

Social network	Facebook	Wiki-vote	Twitter	Epinion
Nodes	4039	7115	81,306	75,879
Edges	88,234	103,689	1,768,149	508,837
Average degree	21.85	14.57	21.75	6.70
Nodes in 6-core	3478	3343	58,787	13,911
Edges in 6-core	86,492	94,179	1,279,919	303,324

Table 7.1: Data set statistics

We found:

1. Under proper parameters, there are large fraction of nodes in social networks in the core space.

2. Moreover, we determined that this property is a result of the structure of the network. We compared the results to an equivalent networks where the edges were randomly "rewired" so that the network had the same number of nodes, same number of edges, and same degree distribution, but the edges were distributed randomly. Such networks did not exhibit large core regions.

7.6.2 Implementation Details

To test the low-degree assumption on each network we used spectral embedding techniques to embed the 6-core of the graph into $\mathbb{R}^d$, and then measured the ***core-fraction*** of the resulting metric graph as follows: for given parameters r, p, and Δ, we first removed all nodes in the 6-core that either a) did not have Δ nodes within distance r, or b) were not neighbors with a p fraction of the nodes within distance r. We then created a graph of the remaining nodes by connecting those within distance r. We output the size of the largest component divided by the size of the 6-core, the core-fraction.

We additionally, randomly "rewired" the 6-core of each graph and again embedded that into $\mathbb{R}^d$, and then measured properties of the resulting metric graphs.

We now describe each step of the procedure in detail.

- **Finding the k-core of the original social network.** A k-core is a maximal connected subgraph of the original graph in which all vertices have degree at least k. This can be obtained by iteratively removing nodes of degree $< k$. We fix $k = 6$.

- **Rewire graph** A rewired graph $G' = (V', E')$ of $G(V, E)$ is constructed starting out with G, and then by letting $V' = V$ and carrying out a series of switching steps ($10|E|$ times) whereby a pair of edges are selected uniformly at random and the ends are exchanged[165]. A *rewired k-core* takes two steps 1) take k-core of the original graph 2) rewire the k-core.

- **Embedding a graphs into $\mathbb{R}^d$ by spectral embedding [108].** Formally, we create $x : V \to \mathbb{R}^d$ such that $x = (x_1, ..., x_d)$ where $x_i : V \to \mathbb{R}$. x_i then can be seen as vector where $x_0 = 1^n$ and for $i > 0$ we obtain x_i by solving

$$x_i = \underset{x^\top x_k = 0, \forall 0 \leq k < i}{\arg\min} \sum_{(u,v) \in E} ||x(u) - x(v)||^2.$$

  This can be solved efficiently and x_i will be the eigenvector corresponding to the i-th largest eigenvalue of the Laplacian matrix of the graph.

- **Computing core-faction** We fixed $\Delta = 10$. For values of p between 0 and 0.2 in .01 increments we tested different values of r and used the r which maximized the core-fraction at $p = 0.2$.

Explanation of Procedure Though no canonical way of embedding a graph into metric space, one conventional way is to use spectral techniques [108], which, intuitively tries to embed the graph in such a way as to minimize the distance between two connected vertices in the embedding. A known issue with spectral embedding is that it can be greatly influence by nodes that are disconnected from (or barely to) the rest of the graph. For this reason, we first took the k-core. Additionally, we expected that much of the core of the graph would be in the k-core, so that we would not lose too much by doing this.

Recall that in our model, nodes in the core were required to have a sufficient density of nearby nodes, and were randomly connected to a p fraction of these nodes.

Additionally, it was required that the core was connected. We first determined what "close" (that is r) should mean in terms of the embedding. Note that if r is too small, then no nodes will have sufficient density, and thus will not be included in the core. If r is too large, then nodes will not be connected to large fraction of the nodes within distance r. We chose the r that worked best when $p = 0.2$, which corresponds to choosing the "scale" of the embedding. We optimized r separately for the original and rewired networks. The rewiring of the graph essential destroys any "structure" of the graph, and serves as a sort of null hypothesis. This mostly serves as a sanity check that the parameters of our test do not hold for all graphs, but that there is something specific in real-world networks that support our assumptions.

Note that our goal here was not to find the "best" embedding, but to find one that would effectively illustrate that our low-dimensional assumptions hold. If our assumptions hold in this embedding, we know that the graph contains low-dimensional structure; however, if our assumptions fail to hold in this embedding, we only know that this particular embedding is a poor choice.

7.6.3 Experimental Results

The results with $\Delta = 10$, $0 < p < 0.2$, and $d = 2, 3, 4$, and 5 (recall d is the dimension of the embedding) are shown in Figure 7.1. We generally found that there is a large fraction of nodes in core space, with the Twitter and Facebook networks embedding more effectively than E-pinions or Wikipedia. This is promising because these data sets are the closest to traditional social networks.

Note that the charts only show the fraction of the 6-core in the core. Table 7.1 additionally shows the fraction of the nodes in the 6-core, which averages around 50% but varies greatly between datasets. For example, even when we required that a core node be connected to 20% of the close nodes in the Facebook data set, about 22% of the 6-core nodes remained in the core. Because in this dataset over 86% of the nodes are in the 6-core, this means that about 19% of the nodes are in the core. If we only require that a core node be connected to 10% of the close nodes, then the overall fraction of core nodes jumps to 34%.

The exception was the E-pinions network. In this network, even when we only required a core node be connected to 5% of the close nodes only 12% of the 6-core remained in the core. Also, unique to this network is that the 6-core only represented about 18% of the nodes. So at this point only 2% of the nodes are in the core. While we cannot definitively say, we postulate that one reason for this failure is the low

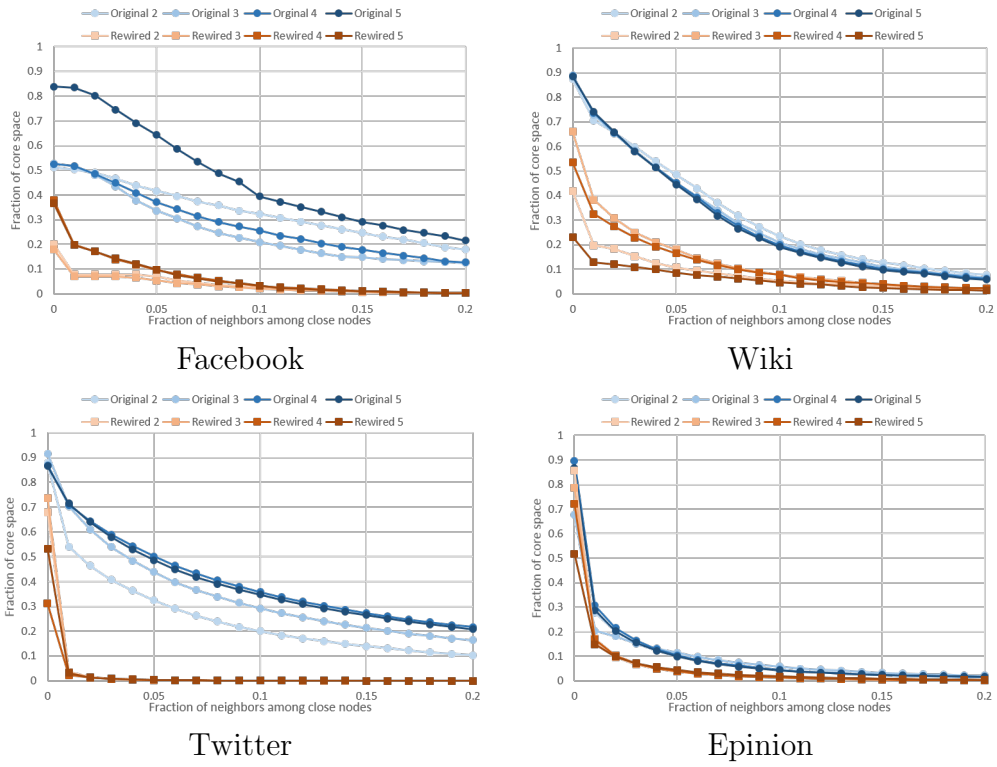


Figure 7.1: The relation between fraction of core space to graph under required fraction of neighbors among close nodes and in different dimensions.

average degree of the E-pinion network, which is less than half of any other network. Additionally, we note that even though the numbers are small, the faction of nodes in the core of the E-pinions network is still a factor of 10 greater than in the rewired E-pinions network.

The dimension for which we embedded a network seemed not to make a systematic difference, though it seemed like slightly larger dimensions were more effective in Twitter and Facebook.

In the rewiring setting, the experiments show that the embeddings of the rewired networks do not do as well placing neighbors close by. This indicates that the link structure in original networks contains features that the rewired networks do not. In particular, the core-fraction of the rewired networks when $p = 0.2$ was about 0.46% in Facebook, 2.2% in Wikipedia, 0.01% in Twitter, and 0.39% in E-pinion.

7.7 Conclusion

Sybil attacks, in which an adversary creates a large number of identities, present a formidable problem for the robustness of recommendation systems. We create a sybil

defense system that provably defends against sybil attacks, even in the setting where an adversary can create an arbitrary number of links, though only to a constant fraction of the network. This defends against periphery attacks, which were found in practice, but not accounted for by previous sybil defense works. We did this by exploiting a new, for sybil detection, social network property, namely, that nodes can be embedded in low-dimensional spaces, which we empirically verified.

CHAPTER 8

The Volatility of Weak Ties: Co-evolution of Selection and Influence in Social Networks

In this chapter, we look at opinion formation and the effects of two phenomena both of which promote consensus between agents connected by ties: *influence*, agents changing their opinions to match their neighbors; and *selection*, agents re-wiring to connect to new agents when the existing neighbor has a different opinion. In our agent-based model, we assume that only weak ties can be rewired and strong ties do not change. The network structure as well as the opinion landscape thus co-evolve with two important parameters: the probability of influence versus selection; and the fraction of strong ties versus weak ties. Using empirical and theoretical methodologies we show that on a two-dimensional spatial network:

- With no/low selection, the presence of weak ties enables fast consensus. This conforms with the classical theory that weak ties are helpful for quickly mixing and spreading information, and strong ties alone act much more slowly.
- With high selection, too many weak ties inhibit any consensus at all—the graph partitions. The weak ties reinforce the differing opinions rather than mixing them. However, sufficiently many strong ties promote convergence, though at a slower pace.

We additionally test the aforementioned results using a real network. Our study relates two theoretical ideas: the strength of weak ties—that weak ties are useful for spreading information; and the idea of echo chambers or filter bubbles, that people are typically bombarded by the opinions of like-minded individuals. The difference is in how (much) selection operates.

8.1 Introduction

Social ties are not static, they evolve over time and the evolution is driven by two processes. One is *selection* where an individual may seek out and form new ties; often with others that have similar attributes [217]. The other social process is *influence* in which two individuals already connected by a social tie may influence one another and converge on their personal attributes (interest, tastes, etc) [201, 219]. Both of them result in neighboring nodes being more similar than two random nodes.

The sociology literature has, for a long time, acknowledged and studied the difference of social ties [105, 106, 145]. Strong ties refer to the ties that people regularly spend effort to maintain, such as family members, close friends, and colleagues. Weak ties, on the other hand, are relatively effortless to keep and typically are much more numerous than strong ties. The difference in the type of ties is also reflected structurally. Strong ties tend to be clustered with a high clustering coefficient, while weak ties are important bridges that connect remote communities. In the seminal paper "The Strength of the Weak Ties" Granovetter [105] showed how information spreads through weak ties. While strong ties connect people who are more similar to each other (due to homophily), weak ties tend to bring fresh information to a social group, which can be extremely valuable, for example, in the case of looking for new jobs.

One of the interesting aspects of this paper is to examine the evolution of strong ties and weak ties, with selection and influence considered. By definition, strong ties and weak ties also differ in their stability or fragility. The physical constraints that form a strong tie are often stable in time and are hard to change. Many of the strong ties are not formed by selection. We are born with family ties and they stay with us for a lifetime except in extreme cases. Neighbors and colleagues are also relatively hard to change without some serious effort or cost. But weak ties, especially those discovered on a social platform, are a lot easier to form or break, making it convenient to block opinions that one does not like and stay in a comfortable "echo chamber" [20, 223].

The political science literature has confirmed the observation of geographical segregation and partisan alignment [160, 94] and of 'ideology sorting', that people tend to "segregate themselves into their own political worlds, blocking out discordant voices and surrounding themselves with reassuring news and companions" [35]. In the online setting, the sorting process can possibly happen at a much faster rate and a larger scale [122, 152, 20, 37, 162]. Online forums allow people to seek out like-minded individuals, including those holding unpopular views that have been shunned elsewhere [69]. Moreover, social media research clearly shows that unfriending on

Facebook [210] and unfollowing on Twitter [241, 132] disproportionately affect weak ties as compared to strong ties. Between 16% and 26% of American SNS users have disconnected a tie for reasons to do with politics [191, 122, 210]. While such selection processes indeed limits the information input to certain users, it was also observed that the disconnections helped to sustain user participation in the social network [152].

Our Approach In this work we develop a model of opinion formation and changes with two competing opinions/behaviors. Examples include political views (liberal v.s. conservative) or behaviors (smoking/non-smoking, drug use/no drug use). The opinions are influenced by one's friends which could be connected by strong ties or weak ties. Generally speaking, one's opinion is going to move toward the majority opinion in his/her friend circle over time. Meanwhile, selection may also happen such that a node re-wires ties when he/she has different opinion from his/her friends. In our model, we assume that only the weak ties can be rewired and strong ties do not change. The network structure as well as the opinion landscape thus co-evolve with two important parameters: p_{select}, the probability of a selection as the next action as opposed to influence; and q_{strong}, the fraction of strong ties in the network. The objective of this paper is to answer the following question: does the opinion distribution converge and if so how fast does it converge with respect to the two parameters?

Related Work There has been work on co-evolution of social ties and opinions without separating strong/weak ties. Holme and Newman [115]

show a phase transition from a segregated network to a homogeneous network, controlled by p_{select}. Durrett et al. [80] built on top of the Holme-Newman model and considered two models of selection: rewire-to-random, and rewire-to-same. Cohen et al. [59] study a problem of opinion formation with continuous values with influence and selection. Kempe et al. [129] considered agents with multiple dimensions/attribute types and only agents who are similar in many dimensions can influence each other. They characterized the equilibrium outcome and proved convergence.

An expansive literature attempted to validate selection and influence models using real-world data, although some of them are limited as they assume independent observations and no external factors [217]. Lewis et al. [146] considered Facebook data and discovered that there could be a large variation of whether selection or influence is more prominent, depending on the studied attributes. Further, selection and influ-

ence can be heavily entangled. For example, in a static network (when selection does not exist), both cooperative and selfish behaviors are contagious. But in a dynamic network, selfish behavior is still contagious, but cooperative behavior is not [123]. Thus selection and influence in network co-evolution definitely deserve further study in different social settings.

8.2 Model

8.2.1 Model of Agent Network

To encode the interaction among people, we use a directed graph $G = (V, E_S, E_W)$ with V as the set of nodes and two types of edges— *strong ties*, E_S, and *weak ties*, E_W. For $v \in V$ let $d_S(v)$ be the *out-degree* of strong ties of node v and the i-th strong out-neighbor of node v is denoted by $\delta_S(v)_i$. We define $d_W(v)$ and $\delta_W(v)_i$ analogously. We allow multi-edges and self loops in both E_S and E_W.

8.2.2 Dynamics of Influence and Selection

Each agent $v \in V$ has an opinion $\chi(v) \in \{0, 1\}$. We call $\chi = \{\chi(v) : v \in V\}$ the *opinion vector*. For $\sigma \in \{0, 1\}$, let $\boldsymbol{x}(\sigma) \triangleq \{v \in V : \chi(v) = \sigma\} \subseteq V$ denote the set of nodes with opinion σ. Let $R^\chi_S(v) = \frac{|\{i:\chi(\delta_S(v)_i)=1\}|}{d_S(v)}$ be the fraction of strong ties which have an endpoint with opinion 1, and similarly define $R^\chi_W(v)$.

The process Sel-Inf$(G^{(0)}, f_{\text{inf}}, p_{\text{select}}, q_{\text{strong}})$ is a discrete time Markov chain over state space $\{(\chi, G)\}$ where $G^{(0)}$ is the initial network of agents, $f_{\text{inf}} : [0, 1] \mapsto [0, 1]$ is an influence function, parameter $p_{\text{select}} \in [0, 1)$ denotes the amount of selection (versus influence), and $q_{\text{strong}} \in [0, 1]$ denotes the influence of the strong ties (versus weak ties). To this end we define $R^\chi(v) = q_{\text{strong}} R^\chi_S(v) + (1 - q_{\text{strong}}) R^\chi_W(v)$ to be the weighted fraction of v's neighbors that are 1.

The dynamics Sel-Inf$(G^{(0)}, f_{\text{inf}}, p_{\text{select}}, q_{\text{strong}})$ start with the graph $G^{(0)}$ and initial opinions that are uniformly and independently randomly selected. Given state $Y^{(t)} = (\chi^{(t)}, G^{(t)})$ at time t, the dynamics updates to $Y^{(t+1)}$ as follows: initially set $Y^{(t+1)} = Y^{(t)}$, choose an agent v uniformly at random and update $Y^{(t+1)}$ with one of the following two operations:

Selection. With probability p_{select}, agent v randomly chooses a weak tie and rewires if they disagree: select a random number k between $1, \ldots, d_W(v)$, and let

$u = \delta_W^{(t)}(v)_k$. Then

$$\delta_W^{(t+1)}(v)_k = \begin{cases} u & \text{, if } \chi^{(t)}(v) = \chi^{(t)}(u) \\ \text{a random node in } V & \text{, otherwise.} \end{cases} \tag{8.1}$$

Influence.[206] Otherwise (with probability $1 - p_{\text{select}}$), agent u updates its opinion,

$$\chi_v^{(t+1)} = \begin{cases} 1 & \text{with probability } f_{\text{inf}}\left(R^{\chi^{(t)}}(v)\right) \\ 0 & \text{, otherwise.} \end{cases} \tag{8.2}$$

recall that $R^{\chi^{(t)}}(v)$ is the q_{strong} weighted fraction of v's neighbors with opinion 1 at time t.

We say the process reaches *consensus* if all agents have the same opinion, and we use the number of influence steps as the consensus time.

Remark 8.1. Our model is similar to the Holme-Newman model [115]. In the selection phase of our model, the chosen node picks a random edge, and when the endpoint has a different opinion rewires the edge to a random node (rewired when disagreeing). In their model, a random edge is rewired to a random node with the same opinion (rewired to the same). For the influence phase, their model uses the voter model to update opinions.

Remark 8.2. We will describe our simulation results using ρ_{select} instead of p_{select} where $p_{\text{select}} = \frac{d\rho_{\text{select}}}{1+(d-1)\rho_{\text{select}}}$ and d is the average degree of the graph. Here ρ_{select} just rescales p_{select} to correctly normalize for the degree. This way, if v is a node of degree d, the rate that the opinion of $\delta_W(v)_i$ is updated via selection versus influence is ρ_{select} versus $1 - \rho_{\text{select}}$ and does not depend on d.

8.2.3 Choices of Influence dynamics

We consider k-majority dynamics (choose k neighbors according to their edge weights independently with replacement and change the opinion to the majority opinion of these k neighbors),

$$f_{\text{inf}}(x) = \sum_{\ell=\lceil k/2 \rceil}^{k} \binom{k}{\ell} x^{\ell}(1-x)^{k-\ell}. \tag{8.3}$$

This generalizes several previously studied models:

- Voter Model ($k = 1$): agent u chooses a neighbor v with probability proportional to the weight and updates to v's opinion, $f_{\text{inf}}(x) = x$ [114].

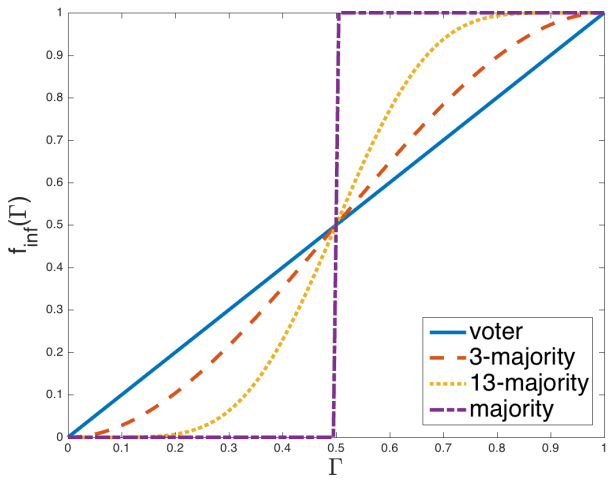


Figure 8.1: The function f_{inf} for different influence dynamics. The k-majority model, with an increasing k, changes from the voter model to the majority model.

- Majority ($k \to \infty$): agent u updates to the opinion with maximum weight, when there is a tie, the opinion is chosen at random [176].
- 3-majority dynamics ($k = 3$): agent u polls the opinion from three random neighbors and takes the majority as the new opinion [31].

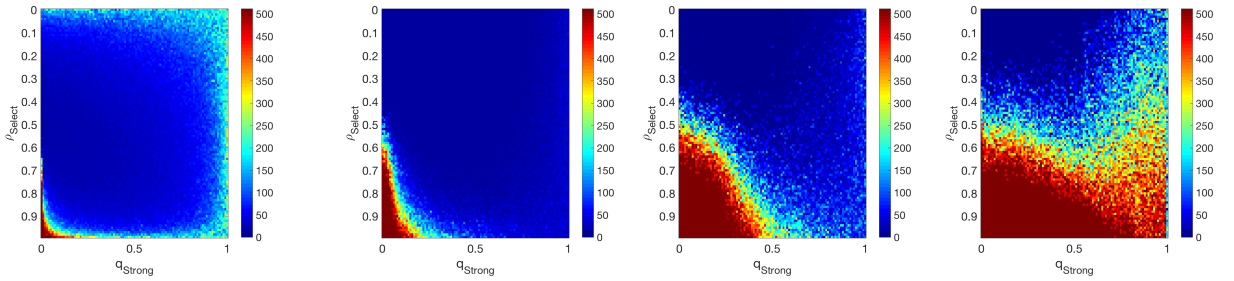
For $k > 1$ this family of influence dynamics can be seen as the smooth version of majority dynamic with "the rich get richer property"— if $R_u > 1/2$, more than half of u's neighbors are 1 then the probability that agent u updates to 1 is greater than R_u, the fraction of u's 1 neighbors; moreover on a complete graph if the number of agents with opinion 1 is greater than the number of agents with opinion 0 there is a "drift" for opinion 1 such that the number of agents with opinion 1 tends to increase. We are primarily interested in the case where $k > 1$, but include the $k = 1$ case for contrast.

8.2.4 Our Problem

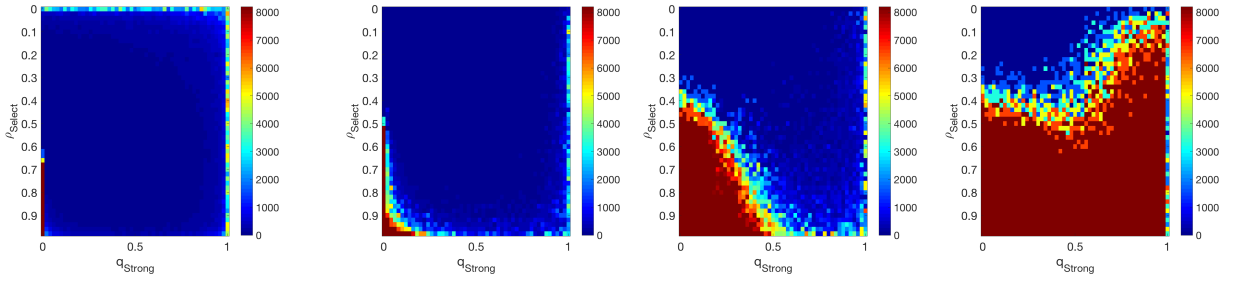
In this paper we try to understand the role of weak ties in promoting consensus with two main parameters: ρ_{select}, the probability of selection as the next action as opposed to influence; and q_{strong}, the fractional influence of the strong ties in the network. We consider the entire parameter space: $\rho_{\text{select}} \in [0, 1)$ and $q_{\text{strong}} \in [0, 1]$. For shorthand, we refer to this as Sel-Inf$(f_{\text{inf}}, p_{\text{select}}, q_{\text{strong}})$, when the graph is clear.

In this paper we consider a number of graph topologies, networks generated by the Newman-Watts model and a real-world ego-network from Facebook [158].

8.3 Spatial Networks



(a) Voter Model (k=1) (b) 3-Majority (c) 13-Majority (d) Majority(16×16)



(a) Voter Model (k=1) (b) 3-Majority (c) 13-Majority (d) Majority(64×64)

Figure 8.2: Consensus time on spatial network. The color at each point $(\rho_{\text{select}}, q_{\text{strong}})$ in this bit map represents the average number of influence steps before consensus (or timeout). The size of graph in the top row is 256 and the bottom row is 4096.

8.3.1 Simulation setting

In this section, the initial graph we study is based on the Newman-Watts model [180]. The nodes form a two dimensional lattice wrapped into a torus. Each node has 12 strong ties connecting it to nodes with Hamming distance less than 2, and 10 weak ties to random nodes drawn uniformly and independently with replacement.

We run simulations on networks of size ranging from 16×16 to 64×64 (256 to 4096 nodes). A representative figure on the number of influence steps until consensus is shown in Figure 8.2. The color at each point $(\rho_{\text{select}}, q_{\text{strong}})$ represents the number of influence steps before consensus (or timeout) normalized by the the size of the graph and averaged among the trials of the dynamics Sel-Inf$(f_{\text{inf}}, p_{\text{select}}, q_{\text{strong}})$. We stop the dynamics if the total number of influence steps is more than twice the square of the size of the graph. In the larger graph, this corresponds to 33,554,432 influence steps and, for some parameter settings, over 10 billion total steps. For the 256 node graph, we run 10 trials for each of 100×101 parameter settings. For the 4096 node graph, we run 5 trials for each of 50×51 parameter settings.

8.3.2 Simulation Results Overview

To better understand Figure 8.2, we first consider what happens with different selection rates. When $\rho_{\text{select}} < 0.5$, which is the upper part of the plots, the majority-like processes (3-majority, 13-majority, and majority) reach consensus faster if the weight of weak ties is larger (q_{strong} being smaller). This is natural because the graph topology is more stable when ρ_{select} is small. Once the number of nodes with different opinions become imbalanced *the weak ties act like sampling a complete graph and help the opinions to mix, strengthening the imbalance.* If q_{strong} is close to 1, the network has mostly only the strong ties that connect local neighbors. Even though there may exist a global imbalance of opinions, it still takes a long time to spread this imbalance through strong ties.

However, when selection rate is high ($\rho_{\text{select}} > 0.5$, the lower part of the plots), the majority-like processes (3-majority, 13-majority, and majority) reach consensus slower or even get stuck if there are a large fraction of weak ties (when q_{strong} is small). In contrast to the low selection setting, here the weak tie weights are frequently updated and form stronger connections among the agents with the same opinion. Informally, *the weak ties form community structures which hinder the agents from communicating between different opinions and prevent the opinions from mixing.* As a result, the higher the selection rate is, the harder for the agents to reach consensus.

We hypothesize that there are three distinct theoretical cases:

Fast Consensus Consensus takes a logarithmic number of steps (per node).

Slow Consensus Consensus is reached in polynomial time.

No Consensus Consensus is either never reached or takes exponential time.

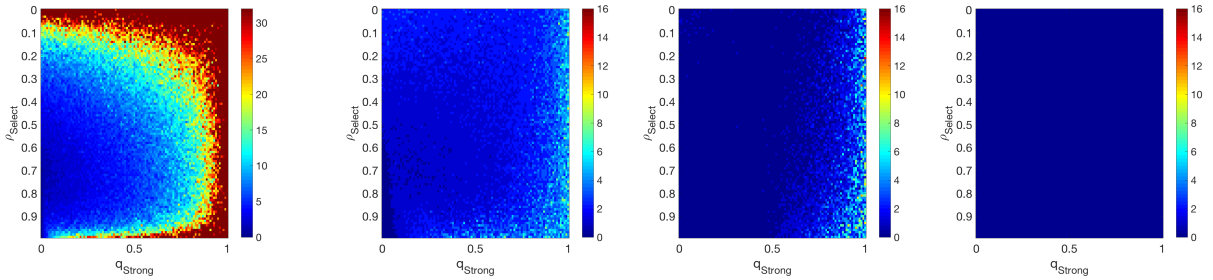
Roughly speaking: we expect fast consensus is represented by the deep blue region; no consensus by the deep red region; and slow consensus by the other colors. Notice that when there are no strong ties ($q_{\text{strong}} = 0$) the transition from fast consensus to no consensus is rapid. We hypothesize that the there is a threshold here. Moreover, that there is a "triple point" incident on each of these three regions.

In the remainder of our analysis we focus on the three "edges": either $q_{\text{strong}} = 0$ or $\rho_{\text{select}} \in \{0, 1\}$, and we change the other parameters. Note that when $q_{\text{strong}} = 1$ selection cannot operate and the value of $\rho_{\text{select}} \in \{0, 1\}$ is immaterial. So this case is omitted.

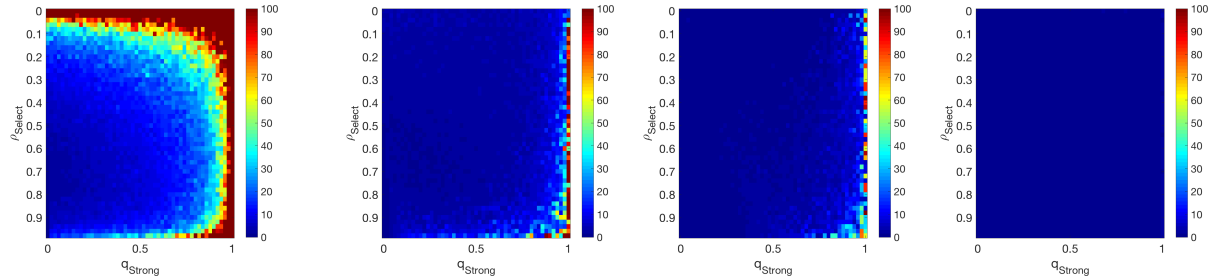
8.3.3 Weak Ties Only ($q_{\text{strong}} = 0$)

In this section we study the effects of the relative frequency between selection and influence (ρ_{select}) on the consensus time of Sel-Inf($f_{\text{inf}}, p_{\text{select}}, q_{\text{strong}}$) when the strong ties are absent, $q_{\text{strong}} = 0$. This corresponds to the left edge of the plots in Figure 8.2.

We can see that if $\rho_{\text{select}} = 0$, then the dynamics quickly converge in all but the voter model, where it slowly converges. On the other hand if $\rho_{\text{select}} \to 1$, then it nearly always times out before converging. We hypothesize that in this case there is no consensus. One way we can see this is in Figure 8.3, which plots the number of times nodes switch opinions, normalized by the size of the graph, before the processes reach consensus. A switch is an influence step when the chosen agent changes its opinion. The total number of switches is quite small in this region. This indicates that no real progress is being made.



(a) Voter Model (b) 3-Majority (c) 13-Majority (d) Majority



(a) Voter Model (k=1) (b) 3-Majority (c) 13-Majority (d) Majority

Figure 8.3: Switches on Spacial Network. The color at each point ($\rho_{\text{select}}, q_{\text{strong}}$) in this a bit map represents the total number of switches (before consensus or timing out) normalized by the size of the network for Sel-Inf($f_{\text{inf}}, p_{\text{select}}, q_{\text{strong}}$). The size of graph in the top row is 256 and the bottom row is 4096.

k > 1 First we consider $k > 1$—recall f_{inf} is k-majority. We see that on the left side of the plots in Figure 8.2 the time quickly transitions from fast to very slow. Again the data in Figure 8.3 backs up the story that the process transitions from making quick progress (with few switches) to making no progress (with a lot of switches).

In the following section we use theoretical analysis to show that in the mean field approximation the k-majority dynamics (for odd k) converges to *segregation* if the relative frequency of selection is high enough. We present theoretical results on the mean field approximation of this setting in Section 8.4.

$\mathbf{k = 1}$ Turning toward the case $k = 1$, we notice a large difference. Here the dynamics appear to converge slowly at $\rho_{\text{select}} = 0$. The time to consensus is intermediate (Figure 8.2), and requires many switches (Figure 8.3). However, as ρ_{select} increases, the process transitions to fast consensus (fast time and few switches). Finally, as ρ_{select} continues to increase we transition to increasingly timing out (slow time and few switches). The slow consensus at $\rho_{\text{select}} = 0$ is expected, because the voter model has no drift. However, the fast consensus time for intermediate values of ρ_{select} is surprising. We hypothesize that it is due to the details of the selection process which induces a rich-get-richer drift. When updating, if a node is in the minority, then its selections acts slower (because the updates are additive, but the total mass of its weak ties is smaller). This means that minority nodes are more likely to be connected to majority nodes than vice versa.

8.3.4 No Selection, Only Influence ($\rho_{\text{select}} = 0$)

In this section, we consider the setting when there is no selection. Therefore the process boils down to influence in a static network with strong and weak ties. The results are at the top edge of the plots in Figure 8.2.

For k-majority models for $k > 1$, we hypothesize that any non-zero fraction of weak ties leads to fast consensus, which is supported in the simulation results. The reason is that as soon as an opinion is a global leader, the weak ties introduce a global drift. Since there is no selection, each node connects uniformly to all nodes via weak ties. The strong ties can make local imbalances, but these cancel each other out as the size of the "boundary" for each opinion is necessarily the same. In Figure 8.3, the number of switches increases when there are more strong ties (with q_{strong} increasing). When q_{strong} is small, on average each node switches fewer than 4 times before consensus is reached — weak ties help to spread the imbalance of opinion quickly and in most of the influence steps the chosen agent updates to the global majority correctly.

However, with just strong ties ($q_{\text{strong}} = 1$, the top right corner), the process predominantly changes only at the boundary of regions of different opinions. Since

the boundary of each opinion is the same, the process takes an unbiased walk (without drift) and converges slowly.

For $k = 1$, we have the voter model, which has no drift regardless of q_{strong}. However, as there are more weak ties, the graph mixes better and convergence speeds increase slightly. Indeed, as the fraction of strong ties increases, the number of switches in Figure 8.3 increases. However, compared to majority-like dynamics the voter model has a much larger number of switches regardless of the value of q_{strong}.

8.3.5 Lots of Selection ($\rho_{\text{select}} \to 1$)

In this section, we want to understand when ρ_{select} is nearly 1, which is near the bottom edge of the plots in Figure 8.2. When $\rho_{\text{select}} = 1$, i.e., no influence, the opinions do not change. Thus the network does not reach consensus.

When q_{strong} and ρ_{select} is nearly 1 (near the right bottom corner), there are no weak ties. Although almost all actions are selections, there are simply no weak ties to work on, and so the selection steps do not affect. (Note that Figure 8.2 only counts influence steps.) Thus, as discussed in the earlier section it converges but slowly.

When $\rho_{\text{select}} \to 1$ and q_{strong} is increasing, the strong ties increasingly help with consensus, but the weak ties are almost surely connecting nodes of the same opinion. Conversely, as the number of weak ties increases, they increasingly promote segregation.

For the majority model, it is abruptly not stuck when $q_{\text{strong}} = 1$. Here it is, in theory, possible that the dynamics get stuck (for example if an 8×16 region of nodes in the torus have opinion 0 and the other 8×16 region have opinion 1. All agents will have three neighbors of their type. However, in our empirical results, these trials never do become stuck. Since there are only strong ties, we hypothesize, that in the case the dynamics do converge it cannot be done quickly (in logarithmic time per node) but must take a polynomial time per node to converge.

8.4 Theoretical Results

In this section, we analyze the process Sel-Inf when the d-regular random graph which only has weak ties, and we show the mean field approximation process converges to segregation when the selection rate is higher than a certain threshold which depends on the influence function f_{inf} and the degree d.

Formally, we consider Sel-Inf$(G^{(0)}, f_{\text{inf}}, p_{\text{select}}, q_{\text{strong}})$ where the initial weak graph

$E_W^{(0)}$ is a directed d-regular random graph (i.e., each node has d out neighbors selected at random), $q_{\text{strong}} = 0$, and f_{inf} is the k-majority influence dynamics with $k \geq 3$. We note that the nodes with the same initial state will have the same *expected* behavior. Specifically we can partition the nodes by their initial opinions into $U_0 \triangleq \boldsymbol{x}^{(0)}(0)$ and $U_1 \triangleq \boldsymbol{x}^{(0)}(1)$ and can assume $|U_0| = |U_1| = n/2$.

For $\sigma \in \{0, 1\}$ we call $v \in \boldsymbol{x}^{(0)}(\sigma)$ a *type* σ node, and similarly define type $\tau \in \{0, 1\}$ nodes. We set $X_\sigma(t)$ to be the average probability of type σ nodes having opinion 1 at time t, and $C_{\sigma,\tau}(t)$, the expected cut of the weak ties between a type σ node and a type τ node at time t. Formally,

$$\begin{cases} X_\sigma(t) \triangleq \frac{1}{|U_\sigma|} \sum_{v \in U_\sigma} \mathbb{E}\left[x_v^{(t)}\right] \\ C_{\sigma,\tau}(t) \triangleq \frac{1}{|U_\sigma|} \sum_{v \in U_\sigma} \frac{1}{d_W(v)} \mathbb{E}\left[\left|\{i : \delta_W^{(t)}(v)_i \in U_\tau\}\right|\right] \end{cases} \tag{8.4}$$

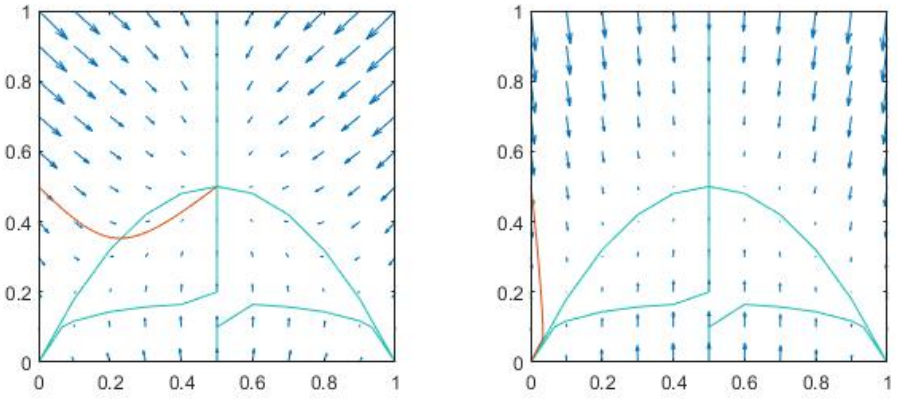
Theorem 8.3. *Given constants $k > 1$ odd and d, let $G^{(0)}$ be a directed d-regular random graph with n nodes, and $q_{\text{strong}} = 0$, there exists ${p_{\text{select}}}^* \in (0, 1)$ such that for all $p_{\text{select}} > {p_{\text{select}}}^*$ for sufficiently large n, the mean field approximation of* Sel-Inf *with parameters $(G^{(0)}, f_{\text{inf}}, p_{\text{select}}, q_{\text{strong}})$ defined in Equation (8.4), the system converges to segregation:*

$$\lim_{t \to \infty} X_0(t) = 0, \lim_{t \to \infty} X_1(t) = 1 \tag{8.5}$$

$$\lim_{t \to \infty} C_{0,1}(t) = \lim_{t \to \infty} C_{1,0}(t) = 0. \tag{8.6}$$

Intuitively, this theorem shows in the mean field approximation, the cut between two sets $\boldsymbol{x}^{(0)}(0)$ and $\boldsymbol{x}^{(0)}(1)$ converges to zero, the agents in $\boldsymbol{x}^{(0)}(0)$ converge to opinion 0, and the agents in $\boldsymbol{x}^{(0)}(1)$ converge to opinion 1.

Now we give some intuitions of the proof. We first show that as n increases the recurrence relation can be (rigorously) quantitatively approximated by a system of ordinary differential equation (ODE) (c.f. Figure 8.4). We analyze the corresponding system of ODE using tools from dynamical systems theory. One major challenge of Theorem 8.3 is to argue the limits of system (8.4) converges to $(0, 0)$ without knowing their analytic solutions. We achieve this by using tools in the qualitative analysis of dynamical systems which is of independent interest.



$p_{\text{select}} = 0.5$ $p_{\text{select}} = 0.9$

Figure 8.4: The vector field for dynamical system of (8.4) for 3-majority under different p_{select} where the initial condition $(X_0(t), C_{0,1}(0))$ is $(0, 0.5)$. The green lines represent the zeros of the system of differential equations, and the red path is the numerical solution of the dynamical system. On the left-hand side (small p_{select}), the dynamical system mixes and the probability of having opinion 1 and the connection between two types of nodes converges to $(0.5, 0.5)$. On the right-hand side (large p_{select}), the system segregates— the connection/cut between two types of nodes converges from 0.5 to 0 which is characterized in the Theorem 8.3.

8.4.1 Symmetry in Equation (8.4)

Note that by the definition $C_{0,0}(t) + C_{0,1}(t) = C_{1,0}(t) + C_{1,1}(t) = 1$. For all $\sigma \in \{0, 1\}$, denote the difference of a sequence (a_t) as $\Delta(a_t) \triangleq a_{t+1} - a_t$

$$\begin{aligned}\Delta(X_\sigma(t)) =& \frac{1 - p_{\text{select}}}{2|U_\sigma|} \left(f_{\text{inf}}(R_\sigma(t)) - X_\sigma\right) \\ \Delta(C_{\sigma,\sigma'}(t)) =& \frac{p_{\text{select}}}{4d|U_\sigma|} \left[C_{\sigma,\sigma}(2X_\sigma(1 - X_\sigma)) - C_{\sigma,\sigma'}(X_\sigma + X_{\sigma'} - 2X_\sigma X_{\sigma'})\right]\end{aligned}$$

where $R_\sigma(t) \triangleq C_{\sigma,\sigma}(t)X_\sigma(t) + C_{\sigma,\sigma'}(t)X_{\sigma'}(t)$ and σ' is the complement of σ such that $\sigma, \sigma' \in \{0, 1\}$ and $\sigma' \neq \sigma$.

For the initial conditions, by definition, $X_0(0) = 0, X_1(0) = 1$, and the initial weak graph $E_W^{(0)}$ is a directed d-regular random graph, so $C_{00}(0) = C_{01}(0) = C_{10}(0) = C_{11}(0) = 0.5$. Thus, for all $t \geq 0$, $X_0(t) = 1 - X_1(t)$, $C_{0,0}(t) = C_{1,1}(t)$, and $C_{0,1}(t) = C_{1,0}(t)$.

With these symmetries, we further define $\boldsymbol{Z}^{(t)} = \left(Z_1^{(t)}, Z_2^{(t)}\right)$ where $Z_1^{(t)} \triangleq X_0(t)$

and $Z_2^{(t)} \triangleq C_{0,1}(t)$. We can reduce the number of parameters from 6 to 2 and have

$$\begin{cases} Z_1^{(t+1)} - Z_1^{(t)} = \frac{1}{n}(1 - p_{\text{select}}) f_1\left(\boldsymbol{Z}^{(t)}\right) \\ Z_2^{(t+1)} - Z_2^{(t)} = \frac{1}{n}\frac{p_{\text{select}}}{2d} f_2\left(\boldsymbol{Z}^{(t)}\right) \end{cases} \tag{8.7}$$

where

$$\begin{cases} f_1(\boldsymbol{Z}) = \left(f_{\text{inf}}\left(Z_1 + Z_2(1 - 2Z_1)\right) - Z_1\right) \\ f_2(\boldsymbol{Z}) = \left(-Z_2 + 2Z_1(1 - Z_1)\right) \end{cases}. \tag{8.8}$$

Observe that as n increases, the above process can be approximated by the following ODE by Theorem 2.16:

$$\begin{cases} \frac{d}{dt} z_1 = (1 - p_{\text{select}}) f_1(\boldsymbol{z}) \\ \frac{d}{dt} z_2 = \frac{p_{\text{select}}}{2d} f_2(\boldsymbol{z}) \end{cases} \tag{8.9}$$

8.4.2 Proof of Theorem 8.3

The main idea of the proof has three parts:

1. There exists a ${p_{\text{select}}}^*$ such that for all $p_{\text{select}} > {p_{\text{select}}}^*$, $Z^{(t)}$ converges to $(0, 0)$ if there is t_0 such that $Z^{(t_0)}$ is close to $(0, 0)$.

2. Given $p_{\text{select}} > {p_{\text{select}}}^*$ there exists t_0 large enough such that $\boldsymbol{z}$ hits an asymptotically stable region for $(0, 0)$ at time t_0.

3. Given t_0, there exists a n large enough such that $Z^{(nt_0)}$ and $\boldsymbol{z}(t_0)$ are close.

We formalize these three statements in Lemmas 8.4, 8.5 and 8.7. The proof of Theorem 8.3 is deferred to the full version.

Lemma 8.4. *For all p_{select}, there exist $\delta_{p_{\text{select}}} > 0$ and large enough n such that if there is $t_0 \geq 0, \left\|Z^{(t_0)} - 0\right\| \leq \delta_{p_{\text{select}}}$, then*

$$\lim_{t \to \infty} \left\|Z^{(t)} - 0\right\| = 0.$$

The detailed proof is deferred to the appendix. To prove Lemma 8.4, there are two parts: by Theorem 2.28, we can show 0 is asymptotically stable for (8.9) and there is a potential function V. Then we can show the $\mathbf{Z}^{(t)}$ in (8.7) converges to 0 when $\mathbf{Z}^{(0)}$ is close to 0 by showing $V\left(\mathbf{Z}^{(t)}\right)$ is decreasing as t increases when n sufficiently large.

Lemma 8.5. *There exists ${p_{\text{select}}}^* < 1$ large enough such that for all $p_{\text{select}} > {p_{\text{select}}}^*$ and $\delta > 0$, there is t_0, $\|\boldsymbol{z}(t_0) - 0\| \leq \delta/3$.*

The statement says starting from the initial condition $(0, 0.5)$, $\boldsymbol{z}$ converges to 0 when p_{select} is large enough. The proof of Lemma 8.5 is more complicated, and we need the following lemma to prove it.

Lemma 8.6 (stability)**.** *There exists ${p_{\text{select}}}^* < 1$, a region $R_A \subset \mathbb{R}^2$ containing $(0, 0)$, and $t_0 > 0$. If $p_{\text{select}} \geq {p_{\text{select}}}^*$ and $\boldsymbol{z}(0) = (0, 0.5)$, $\boldsymbol{z}(t_0) \in R_A$, and $\boldsymbol{z}(t) \in R_A$ for all $t \geq t_0$.*

The detailed proof of Lemma 8.6 is in the appendix. Informally, to prove the second part of Lemma 8.6, we first define our stable region $R_A = \{(x_1, x_2) : 0 \leq x_1 \leq x_1^*, 0 \leq x_2 \leq x_2^*\}$.[1] where (x_1^*, x_2^*) is the fixed point of Equation (8.8) with smallest positive x_1^*. We must show at each boundary the drift is inward such that if the $\boldsymbol{z}(t)$ is at the boundary the $\boldsymbol{z}(t+\epsilon)$ will go back to the stable region. For the first part, we show $\boldsymbol{z}$ hits the stable region R_A fast by taking ${p_{\text{select}}}^*$ large enough. With Lemma 8.6 the rest of the proof of Lemma 8.5 goes as follows:

Proof of Lemma 8.5. Our system is two dimensional, so the solution $\boldsymbol{z}$ is a Jordan curve, and it is bounded in R_A if $\boldsymbol{z} \in R_A$ for $t > \tau_0$ by Lemma 8.6. Therefore by Theorem 2.37 $\boldsymbol{z}$ converges to either a fixed point or a limit cycle.

We first show no limit cycle. By Theorem 2.38, it is sufficient to show the divergence of f is not identically zero and does not change sign in R_A

$$\nabla f = (1 - p_{\text{select}}) \left(-1 + f'_{\text{inf}} \left(Z_1 + Z_2(1 - 2Z_1)\right)\right) - \frac{p_{\text{select}}}{2d}.$$

Because a k-majority function defined in (8.3) is Lipschitz such that there exists $L_k > 0$ for all $x \in [0, 1], |f'_{\text{inf}}(x)| \leq L_k$, we can take ${p_{\text{select}}}^*$ large enough such that for all x and $p_{\text{select}} \geq {p_{\text{select}}}^*$, $\nabla f(\boldsymbol{x}) \leq (1 - p_{\text{select}}) (-1 + L_k) - \frac{p_{\text{select}}}{2d} \leq (1 - p_{\text{select}})(L_k - 1 + 1/2d) - 1/2d < 0$. Since 0 is the only fixed point in R_A and there is no limit cycle, $\lim_{t \to 0} \boldsymbol{z}(t) = 0$ □

Lemma 8.7. *Given constants $t_0 \geq 0$, $\delta > 0$, and p_{select} there exists n large enough such that $\|Z^{(n\, t_0)} - \boldsymbol{z}(t_0)\| \leq \delta/3$.*

Since a k-majority function (8.3) is smooth, Lemma 8.7 is a corollary of Theorem 2.16.

[1] Technically, we need our regions to avoid the fixed point, so $R_A = [0, y_1^*] \times [0, y_2^*]$ where $y_1^* < x_1^*$ and $y_2^* < x_2^*$. By the continuity of the system and because the fixed point (x_1^*, x_2^*) is a saddle point, the stability argument still holds.

Proof of Theorem 8.3. By Lemma 8.5 and 8.7, we have $\exists\, p_{\text{select}}{}^*$, $\forall\, p_{\text{select}} > p_{\text{select}}{}^*$, $\forall\, \delta > 0, \exists\, t_0 > 0$ such that for large enough n

$$\|Z^{(n\, t_0)} - 0\| \leq \|Z^{(n\, t_0)} - \boldsymbol{z}(t_0)\| + \|\boldsymbol{z}(t_0) - 0\| < \delta.$$

The proof is completed by taking $\delta = \delta_{p_{\text{select}}}$ and applying Lemma 8.4. □

8.5 Real Social Network

8.5.1 Simulation Setting

We use a dataset consisting of social circles (egocentric networks) collected from Facebook [158]. The graph has 4039 nodes and 88,234 edges. In this section, we only consider the 10-core[2] of Facebook graph as our base network, which contains 2987 nodes and 83,181 edges. We take V as the set of vertices of the 10-core of Facebook graph. Then we use Jaccard similarity[3] to measure tie strength and take the top 80% edges with the highest Jaccard similarity as strong ties edges, and rest as the initial weak ties.

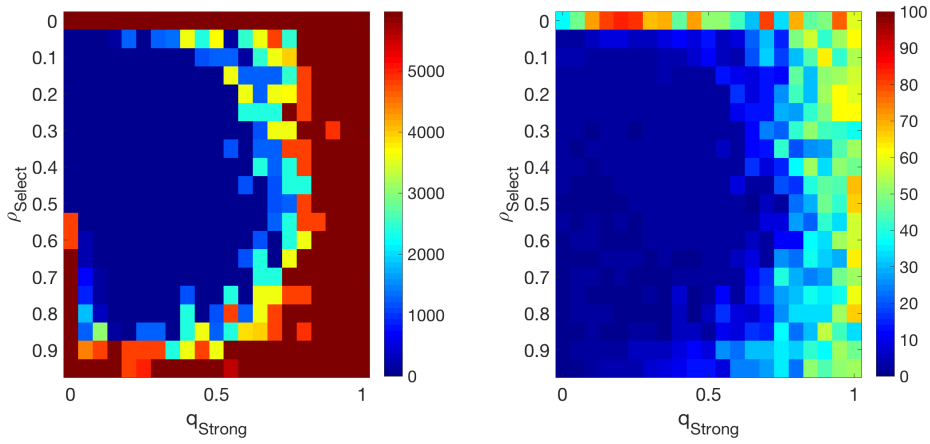
8.5.2 Results

We run the influence-selection dynamics with the 3-majority influence model on the initial graph defined in Section 8.5.1, and show the number of influence steps until consensus in Figure 8.5. We stop a trial if the total number of influence step is more than the two times the square of the size of the graph which is 17,844,338. The setting of bit map is similar to Figure 8.2, but there are 20 parameters ρ_{select} ranging from 0 to 0.95 with even space, and 21 parameters of q_{strong} ranges from 0 to 1 with even space.

Small $\mathbf{q}_{\text{strong}}$ We first consider the case where q_{strong} is small (the left part of the plots). When $\rho_{\text{select}} = 0$ the dynamics almost always time out and the number of switches is high which indicates influence may be not enough for the system to consensus when the graph has a rich structure. Interestingly, when $0 < \rho_{\text{select}} < 0.5$ (upper-left quadrant except for the top boundary), the processes reach consensus

[2]Nodes with fewer than 10 neighbors are iteratively removed.

[3]The Jaccard similarity between u, v defined as $J(u,v) = \frac{|N_u \cap N_v|}{|N_u \cup N_v|}$, where N_v is the set of vertices adjacent to node v. The Jaccard coefficient is commonly used to measure the strength of an edges [185].



Consensus time number of switches

Figure 8.5: Consensus time in Facebook and number of switches before consensus. The color at each point $(\rho_{\text{select}}, q_{\text{strong}})$ in this 21×20 bitmap represents the average number of influence steps before consensus (or timeout) of 5 trials of the dynamics Sel-Inf$(f_{\text{inf}}, p_{\text{select}}, q_{\text{strong}})$ with 3-majority measured in influence steps.

quickly, as the weak ties help the opinions to mix. This result shows moderate selection encourages agents to form (random) connections and helps the system mix. However, when selection is dominantly taken, $\rho_{\text{select}} > 0.5$ (lower-left quadrant), the processes often time out, as the selection process creates local community structures by the weak ties that hinder communication between agents of different opinions, preventing the opinions from mixing.

Large $\mathbf{q}_{\text{strong}}$ In the right part of the plots with large q_{strong}, the processes often reach timeout. This may due to the community structures in strong ties of the real graph.

Interestingly, in the region of a medium-high selection rate (center height of the plots), the processes times out if the graph mostly consists of either weak ties or strong ties when q_{strong} is near 0 or 1, because of structures in strong ties and weak ties. However, if q_{strong} is near $1/2$, the graph has a mixture of strong and weak ties. The community structures within the strong and weak ties seem to override each other, and so the processes reach consensus fast. This suggests multiple independent community structures help the processes reach consensus, even if individually, the community structures would stifle agreement.

The results of the simulation on the real-world graph and the synthetic one are similar when q_{strong} is small and p_{select} is large. This is not surprising because the initial condition does not matter in the above condition. When q_{strong} is large or

p_{select} is small the initial graph matters a lot. Our real-world social network has 10-20 rather distinct communities, but our spatial networks, Newman Watt's model, are more uniform. Because of this, the processes on real-world network become stuck substantially more often.

8.6 Conclusion

As discovered by [105], the strength of weak ties is to get new information and fresh ideas into the comfort zone created by strong ties. However, in a time-evolving spatial network, especially one where selection happens at a substantially higher rate than influence, the role of strong ties and weak ties, in terms of spreading fresh ideas, are swapped. The weak ties are too fragile, and the power of spreading information diminishes. The selection causes the forming of weak ties that only repeat and reinforce the same opinion that the person already holds, which ironically, does not bring any new thoughts. It is nevertheless the strong ties that hold the network together, prevent it from being fully divided, and motivate the participants to compromise.

CHAPTER 9

Erdös-Rényi Graphs

Interacting Particle Systems—exemplified by the voter model, iterative majority, and iterative $k-$majority processes—have found use in many disciplines including distributed systems, statistical physics, social networks, and Markov chain theory. In these processes, nodes update their "opinion" according to the frequency of opinions amongst their neighbors.

We propose a family of models parameterized by an update function that we call Node Dynamics: every node initially has a binary opinion. At each round a node is uniformly chosen and randomly updates its opinion with the probability distribution specified by the value of the update function applied to the frequencies of its neighbors' opinions.

In this work, we prove that the Node Dynamics converges to consensus in time $\Theta(n \log n)$ in complete graphs and dense Erdös-Rényi random graphs when the update function is from a large family of "majority-like" functions. Our technical contribution is a general framework that upper bounds the consensus time. In contrast to previous work that relies on handcrafted potential functions, our framework systematically constructs a potential function based on the state space structure.

9.1 Introduction

We propose the following stochastic process—that we call ***Node Dynamics***—on a given network of n agents parameterized by an update function $f : [0, 1] \to [0, 1]$. In the beginning, each agent holds a binary "opinion", either red or blue. Then, in each round, an agent is uniformly chosen and updates its opinion to be red with probability $f(p)$ and blue with probability $1 - f(p)$ where p is the fraction of its neighbors with the red opinion.

Node dynamics generalizes processes of interest in many different disciplines including distributed systems, statistical physics, social networks, and even biology.

Voter Model: In the voter model, at each round, a random node chooses a random neighbor and updates to its opinion. This corresponds to the Node Dynamics with

$$f(x) = x.$$

This models has been extensively studied in mathematics [65, 114, 150, 151], physics [32, 47], and even in social networks [46, 212, 221, 222, 64]. A key question studied is how long it takes the dynamics to reach consensus on different network typologies.

Iterative majority: In the iterative majority dynamics, in each round, a randomly chosen node updates to the opinion of the majority of its neighbors. This corresponds to the Node Dynamics where

$$f(x) = \begin{cases} 1 & \text{if } x > 1/2; \\ 1/2 & \text{if } x = 1/2; \\ 0 & \text{if } x < 1/2. \end{cases}$$

Typically works about Majority Dynamics study 1) when the dynamics converge, how long it takes the dynamics to converge, and whether they converge to the original majority opinion—that is, does majority dynamics successfully aggregate the original opinion [136, 34, 125, 176, 224].

Iterative k-majority: In this dynamics, in each round, a randomly chosen node collects the opinion of k randomly chosen (with replacement) neighbors and updates to the opinion of the majority of those k opinions. This corresponds to the Node Dynamics where

$$f(x) = \sum_{\ell=\lceil k/2 \rceil}^{k} \binom{k}{\ell} x^{\ell}(1-x)^{n-\ell}.$$

A synchronized variant of this dynamics is proposed as a protocol for stabilizing consensus: collection of n agents initially hold a private opinion and interact with the goal of agreeing on one of the choices, in the presence of $O(\sqrt{n})$-*dynamic adversaries* which can adaptively change the opinions of up to $O(\sqrt{n})$ nodes at every round. In the synchronized variant of this dynamics, Doerr et al.

[75] prove 3-majority reaches "stabilizing almost" consensus on the complete graph in the presence of $O(\sqrt{n})$-dynamic adversaries. Many works extend this result beyond binary opinions [66, 63, 31, 1].

Iterative ρ-noisy majority model: [85, 100] In this dynamics, in each round, a randomly chosen node updates the majority opinion of its neighbors with probability $1-\rho$ and uniformly at random with probability ρ.

$$f(x) = \begin{cases} 1-\rho/2 & \text{if } x > 1/2; \\ 1/2 & \text{if } x = 1/2; \\ \rho/2 & \text{if } x < 1/2. \end{cases}$$

Genetic Evolution Model: In biological systems, the chance of survival of an animal can depend on the frequencies of its kin and foes in the network [15, 170]. Moreover, this frequency depending dynamics is also known to model the dynamics for maintaining the genetic diversities of a population [131, 198].

Our Contribution We focus on a large set of update functions f that are symmetric, smooth, and satisfy a property well call "majority-like", intuitively meaning that agents update to the majority opinion strictly more often than the fraction of neighbors holding the majority opinion. We obtain tight bounds for the consensus time—the time that it takes the system to reach a state where each node has an identical opinion—on Erdös-Rényi random graphs.

Our main technical tool is a novel framework for upper bounding the hitting time for a general discrete-time homogeneous Markov chain $(\mathcal{X}, P)$, including non-reversible and even reducible Markov chains. This framework decomposes the problem so that we only need to upper bound two sets of parameters for all $x \in \mathcal{X}$—the reciprocal of the probability of decreasing the distance to target $1/p^+(x)$ and the ratio of the probability of decreasing the distance to the target and the probability of increasing the distance to the target: $p^-(x)/p^+(x)$. Our technique can give much stronger bounds than simply lower bounding $p^-(x)$ and upper bounding $p^+(x)$.

Once we apply this decomposition to our consensus time problem, the problem becomes very manageable. We show the versatility of our approach by extending the results to a variant of the stabilizing consensus problem, where we show that all majority-like dynamics convergence quickly to the "stabilizing almost" consensus on the complete graph in the presence of adversaries.

A large volume of literature is devoted to bounding the hitting time of different

Markov process and achieving fast convergence. The techniques typically employed are (1) showing the Markov chain has fast mixing time [171], (2) reducing the dimension of the process into small set of parameters (e.g. the frequency of each opinion) and using a mean field approximation and concentration property to control the behavior of the process [31], or (3) using handcrafted potential functions [176].

Our results fill in a large gap that these results do not adequately cover. Mixing time is not well-defined in non-reversible or reducible Markov chains, and so does not apply to Markov chains with multiple absorption states, like in the consensus time question we study. Reducing the dimension and using a mean field approximation fails for two reasons. First, summarizing with a small set of parameters is not possible when the process of interest has small imperfections (like in a fixed Erdös-Rényi graph). Second, the mean-field of our dynamics has unstable fixed points; in such cases the mean field does not serve as a useful proxy for the Markov process. Handcrafting potential functions also runs into several problems: the first is that because we consider dynamics on random graphs, the dynamic is not a priori well specified; so there is no specific dynamic to handcraft a potential function for. Secondly, we wish to solve the problem for a large class of update functions f, and so cannot individually hand-craft a potential function for each one. Typically, the potential function is closely tailored to the details of the process.

Additional Related Work Our model is similar to that of Schweitzer and Behera [208] who study a variety of update functions in the homogeneous setting (complete graph) using simulations and heuristic arguments. However, they leave a rigorous study to future work.

9.2 Preliminaries

9.2.1 Node Dynamics

Given an undirected graph $G = (V, E)$ let $\Gamma(v)$ be the neighbors of node v and $\deg(v) = |\Gamma(v)|$.

We define a ***configuration*** $x^{(G)} : V \to \{0, 1\}$ to assign the "color" of each node $v \in G$ to be $x^{(G)}(v)$ so that $x^{(G)} \in \{0, 1\}^n$. We will usually suppress the superscript when it is clear. We will use uppercase (e.g., $X^{(G)}$) when the configuration is a random variable. Moreover we say v is ***red*** if $x(v) = 1$ and is ***blue*** if $x(v) = 0$. We then write the set of red vertices as $x^{-1}(1)$. We say that a configuration x is ***in consensus*** if

$x(\cdot)$ is the constant function (so all nodes are red or all nodes are blue). Given a node v in configuration x we define $r_x(v) = \frac{|\Gamma(v) \cap X^{-1}(1)|}{\deg(v)}$ to be its fraction of red neighbors.

Definition 9.1. An ***update function*** is a mapping $f : [0,1] \mapsto [0,1]$ with the following properties:

Monotone $\forall x, y \in [0,1]$, if $x < y$, then $f(x) \leq f(y)$.

Symmetric $\forall t \in [0, 1/2]$, $f(1/2 + t) = 1 - f(1/2 - t)$.

Absorbing $f(0) = 0$ and $f(1) = 1$.

We define node dynamics as follows:

Definition 9.2. A ***node dynamics*** $\mathrm{ND}(G, f, X_0)$ with an undirected graph $G = (V, E)$, update function f and initial configuration X_0 is a stochastic process over configurations at time t, $\{X_t\}_{t \geq 0}$ where X_0 is the initial configuration. The dynamics proceeds in rounds. At round t, a node v is picked uniformly at random, and we update

$$X_t(v) = \begin{cases} 1 & \text{with probability } f(r_{X_{t-1}}(v)) \\ 0 & \text{otherwise} \end{cases}$$

This formulation is general enough to contain many well known dynamics such as the aforementioned voter model, iterated majority model, and 3-majority dynamics.

Note that in some of the original definitions the nodes syncronously update; whereas, to make our presentation more cohesive, we only consider asynchronous updates.

In this paper, we will focus on the interaction between the update function f and geometric structure of G. More specifically, we are interested in the consensus time defined as following.

Definition 9.3. The ***consensus time*** of node dynamics $\mathrm{ND}(G, f, X_0)$ is a random variable $T(G, f, X_0)$ denoting the first time step that ND is in a consensus configuration. The ***(maximum) expected consensus time*** $\mathrm{ME}(G, f)$ is the maximum expected consensus time over any initial configuration, $\mathrm{ME}(G, f) = \max_{X_0} \mathbb{E}[T(G, f, X_0)]$.

Now we define some properties of functions.

Definition 9.4. Given positive M_1, M_2, a function $f : I \subseteq \Re \mapsto \Re$ is called M_1-***Lipschitz*** in $I \subseteq \Re$ if for all $x, y \in I$,

$$|f(x) - f(y)| \leq M_1 |x - y|.$$

Moreover, f is M_2-***smooth*** in $I \subseteq \Re$ if for all $x, y \in I$,

$$|f'(x) - f'(y)| \leq M_2|x - y|.$$

9.3 Warm-up: Majority-liked Update Function on Complete Graph

In this section we consider majority-like node dynamics on the complete graph K_n with n nodes in which every pair of nodes has an edge (no self-loops). We use this as a toy example to give intuition for dense Erdös-Rényi graphs even though we will obtain better bounds later.

Theorem 9.5. *Let $\mathcal{M} = \mathrm{ND}(K_n, f, X_0)$ be a node dynamic over the complete graph K_n with n nodes. If the update function f satisfies $\forall x : 1/2 < x < 1$ then $x \leq f(x)$, then the maximum expected consensus time of a node dynamic over K_n is*

$$\mathrm{ME}(K_n, f) = O\left(n^2\right).$$

A standard method of proving fast convergence is to guess a potential function of each state and prove the expectation decreases by 1 after every step—this is just an application of corollary 2.10.

As a warm-up, we will prove theorem 9.5 by guessing a potential function and applying corollary 2.10.

Proof of Theorem 9.5. Given a configuration x, define $Pos(x) \triangleq |x^{-1}(1)|$ then for all red nodes v where $x^{(K_n)}(v) = 1$ have $r_x(v) = \frac{Pos(x)-1}{n-1}$; otherwise $r_x(v) = \frac{Pos(x)}{n-1}$.

Because the node dynamics $\mathcal{M}$ is on the complete graph, $\mathcal{M}$ is *lumpable* with respect to partition $\{\Sigma_l\}_{0 \leq l \leq n}$ where $\Sigma_l = \{x \in \Omega : Pos(x) = l\}$ such that for any subsets Σ_i and Σ_j in the partition, and for any states x, y in subset Σ_i,

$$\sum_{z \in \Sigma_j} P(x, z) = \sum_{z \in \Sigma_j} P(y, z)$$

Furthermore, inspired by an analysis of Voter Model [9] we consider $\psi : [n] \mapsto \Re$ as

$$\begin{aligned}\psi(k) = (n-1)\big[&k(H(n-1) - H(k-1)) \\ &+ (n-k)(H(n-1) - H(n-k-1))\big]\end{aligned}$$

where $H(k) \triangleq \sum_{\ell=1}^{k} \frac{1}{\ell}$, and define the potential function as

$$\phi(x) = \psi(Pos(x)) \tag{9.1}$$

The proof of the following claim is deferred to the full version, and here we just give some intuition as to why this potential function for the voter model works. The sequence $(Pos(X_t))_{t\geq 0}$ can be seen as a random walk on $0, 1, \ldots, n$ with *drift*[1]. Moreover the drift depends on $f(Pos(x)) - Pos(x)$. For voter model $f(x) = x$, there is no drift. For majority-like function because there is a positive drift toward n when $Pos(x) > n/2$; and a negative drift toward 0 when when $Pos(x) < n/2$. Informally the drift is always *helping* and thus the potential function for voter models works.

Claim 9.6. *Our definition of ϕ satisfies the inequalities* (2.8)*: Given Markov Chain $\mathcal{M} = \mathrm{ND}(K_n, f, X_0)$ in theorem 9.5, ϕ defined in* (9.1) *are non-negative and satisfy*

$$\mathcal{L}\phi(x) \leq -1 \text{ where } x \neq 0^n, 1^n,$$
$$\phi(x) \geq 0 \text{ where } x = 0^n, 1^n.$$

Combining claim 9.6 and corollary 2.10, we have

$$\mathbb{E}_{\mathcal{M}}[T(K_n, f, x)] \leq \phi(x).$$

By direct computation, if $0 \leq k < n$, $\psi(k+1) - \psi(k) = (n-1)(H(n-k-1) - H(k))$. Therefore, the maximum $\psi(k)$ happens at $k = \lfloor n/2 \rfloor$,

$$\mathrm{ME}(K_n, f) \leq \psi(\lfloor n/2 \rfloor) \leq (\ln 2)n^2,$$

and completes our proof. □

9.4 Smooth Majority-like Update Function on Dense G_{np}

In this section, we consider the smooth Majority-liked update function defined as follows:

Definition 9.7. We call an update function f a *smooth majority-like update function* if it satisfies $\forall x : 1/2 < x < 1$ then $x < f(x)$ and the following technical conditions

[1]The formal definition of drift is in Equation (9.5).

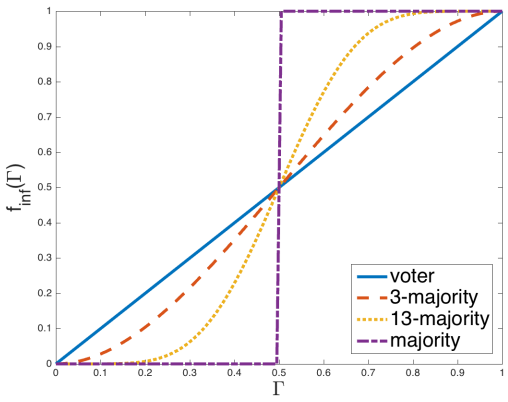


Figure 9.1: An example of smooth majority-like update function.

hold:

Lipschitz There exists M_1 such that f is M_1-Lipschitz in $[0, 1]$.

Condition at $1/2$ There exists an open interval $I_{1/2}$ containing $1/2$ and constants $1 < \check{M}_1$, $0 < \check{M}_2$ such that f is $\check{M}_2$-smooth in $I_{1/2}$ and $1 < \check{M}_1 \leq f'(1/2)$.

Condition at 0 **and** 1 There exists intervals $I_0 \ni 0$, $I_1 \ni 1$ and a constant $\hat{M}_1 < 1$ such that $\forall x \in I_0$, $f(x) \leq \hat{M}_1 x$ and $\forall x \in I_1$, $1 - f(x) \leq \hat{M}_1(1 - x)$.

Intuitively, the majority-like update function should be "smooth" and not tangent with $y = x$. The following figure shows an example of smooth majority-like update function. Now we are ready to state our main theorem.

Theorem 9.8. *Let $\mathcal{M} = \mathrm{ND}(G, f, X_0)$ be a node dynamic over $G \sim G_{np}$ with $p = \Omega(1)$, and let f be a smooth majority-like function. Then the expected consensus time of a node dynamic over G is*

$$\mathrm{ME}(G, f) = O\left(n \log n\right)$$

with high probability.

This theorem shows the fast convergence rate of this process. Note that there is some chance of getting a disconnected graph $G \sim \mathcal{G}$ which results in a reducible Markov chain $\mathcal{M}$ which cannot converge from some initial configurations. Therefore, we can only ask for the fast convergence result *with high probability*.

We note that, the technical conditions exclude interactive majority updates, which we leave for future work.

9.4.1 Proof Overview

Here we will first outline the structure of the proof. In section 9.4.2 we propose a paradigm for proving an upper bound for the hitting time when the state space has special structure. In section 9.4.3, we use the result in section 9.4.2 to prove theorem 9.8.

where large literature have devote to different process. Most of them achieve fast convergence result by using handcraft potential function or showing the Markov chain has fast mixing time. However it is not easy to find clever potential function for any process, and the fast mixing time is not well defined in reducible Markov chain. Recall that the expected consensus time is

$$\tau(x) \triangleq \mathbb{E}_{\mathcal{M}}[T(G, f, x)]$$

which is exactly the hitting time of states 0^n and 1^n. However in contrast to section 9.3, finding a clever potential function is much harder here. We prove theorem 9.8 using that *the expected hitting can be formulated as a system of linear equations* (2.8) and by explicitly estimating an upper bound of this system of linear equations. Moreover, following the intuition in section 9.3, the Markov chain $\mathcal{M}$ can be *nearly* characterized by one parameter $Pos(x)$ when the node dynamics is on a graph that is close to the complete graph. We exploit this structure of our Markov chain and *construct* a potential function for Equations (2.8).

9.4.2 A Framework for Upper Bounding the Hitting Time

We want to upper bound the hitting time from arbitrary state x to $\{0^n, 1^n\}$ denoted as $\tau(x)$ of a given time-homogeneous Markov chain $\mathcal{M} = (\Omega, P)$ with finite state space $\Omega = \{0, 1\}^n$ where $P(x, y) > 0$ only if the states x, y only differ by one digit, $|x - y| \leq 1$.

We let $Pos(x)$ be the *position* of state $x \in \Omega$:

$$Pos(x) \triangleq |x^{-1}(1)|, \text{ and } pos(x) \triangleq Pos(x)/n \tag{9.2}$$

and the *bias* of x as

$$Bias(x) \triangleq |n/2 - Pos(x)|, \text{ and } bias(x) \triangleq Bias(x)/n \tag{9.3}$$

Note that the $Bias(x) = n/2$ if and only if $x = 0^n, 1^n$.

Suppose that $\mathcal{M}$ can be "almost" characterized by one parameter $Bias(x)$. Informally, we want the transitions at states x and y to be similar if $Bias(x) = Bias(y)$. Therefore with the notion of first step analysis we define $\{(p_G^+(x), p_G^-(x))\}_{x\in\Omega}$ where

$$\begin{aligned} p_G^+(x) &= \Pr_{\mathcal{M}}[Bias(X') = Bias(X) + 1 | X = x], \\ p_G^-(x) &= \Pr_{\mathcal{M}}[Bias(X') = Bias(X) - 1 | X = x]. \end{aligned} \tag{9.4}$$

Moreover, we call $p_G^+(x)$ the ***exertion*** and define the ***drift*** of state x as follows

$$D(x) \triangleq \mathbb{E}_{\mathcal{M}}[Bias(X') - Bias(X) | X = x]. \tag{9.5}$$

It is easy to see $D(x) = p^+(x) - p^-(x)$.

Since $\mathcal{M}$ can be almost characterized by one parameter, $Bias(x)$, $\mathcal{M}$ is almost lumpable with respect to the partition induced by $Bias(\cdot)$. The following lemma gives us a scheme for constructing an upper bound for the hitting time:

Lemma 9.9 (Pseudo-lumpability lemma)**.** *Let $\mathcal{M} = (\Omega, P)$ have finite state space $\Omega = \{0,1\}^n$ with even n[2] and $P(x,y) > 0$ only if the states x and y differ in at most one coordinate and*

$$\begin{aligned} d_0 &= \max_{x:Bias(x)=0} \frac{1}{p^+(x)} \\ d_l &= \max_{x:Bias(x)=l} \frac{1}{p^+(x)} + \max_{x:Bias(x)=l} \left(\frac{p^-(x)}{p^+(x)}\right) d_{l-1} \end{aligned} \tag{9.6}$$

where $0 < l < n/2$, and $\{(p^+(x), p^-(x)\}_{x\in\Omega}$ are as defined in (9.4). Then the maximum expected hitting time from state x to $\{0^n, 1^n\}$ can be bounded as follows:

$$\max_{x\in\Omega} \mathbb{E}_{\mathcal{M}}[\tau(x)] \leq \sum_{0\leq\ell<n/2} d_\ell$$

where $\tau(x)$ denotes the hitting time from state x to $\{0^n, 1^n\}$.

Remark 9.10. At first glance it appears this lemma "couples" the process $\mathcal{M}$ with a birth-and-death chain [144], but is actually stronger as the following example illustrates. We define an unbiased random walk where the self transition probability of nodes differs. For all $x \in \{0,1\}^n \setminus \{0^n, 1^n\}$ let $p^+(x) = p^-(x) = \frac{1}{2+x_1}$, and 0^n and 1^n are absorbing states. This lemma yield a polynomial time upper bound because $1/p^+(x) = 3$ and $p^-(x)/p^+(x) = 1$. On the other hand, consider a birth-and-

[2]To avoid cumbersome notion of parity we only consider n to be even here.

death chain on $\{0, 1, \ldots, n/2\}$, such that $P(k, k+1) = \min_{x\in\Omega:Bias(x)=k} p^+(x)$ and $P(k, k-1) = \max_{x\in\Omega:Bias(x)=k} p^-(x)$. Because $P(k, k+1) = 1/3$ and $P(k, k-1) = 1/2$ for all $0 < k < n/2$, the corresponding birth-and-death chain takes exponential time to reach $n/2$.

lemma 9.9 can be derived from corollary 2.10 and is proven in appendix E.1. Intuitively, to get a potential function $s(x)$ for hitting time $\tau(x)$, we order the states in terms of the value of $Bias(\cdot)$, and take the process as a non-uniform random walk on $[n]$. Then we recursively estimate $s(x)$ for each x in increasing order of $Bias(x)$.

To use lemma 9.9, to upper bound $\tau(x)$ we need to prove for *every* configurations $x \in \Omega$

1. An upper bound for $1/p^+(x)$.

2. An upper bound for $p^-(x)/p^+(x)$.

In theorem 9.11 we give a framework that uses the upper bounds for $1/p^+(x)$ and $p^-(x)/p^+(x)$ to obtain upper bounds for expected hitting time. To have some intuition about the statement of the theorem, observe that if the drift $D(x)$ is bounded below by some positive constant both $1/p^+(x)$ and $p^-(x)/p^+(x)$ have nice upper bounds. However, this bound fails when the drift is near zero or even negative. Taking our node dynamics on dense G_{np} as an example, when the states have either very small or very large $bias(x)$ the drift $D(x)$ can be very close to zero or even negative. The drift near $1/2$ is close to 0 because the effects of red and blue largely cancel each other. The drift near the extreme point is small because there are very few nodes outside the majority.

As a result, we partition the states into subsets, and take addition care on the sets of states with small drift. In theorem 9.11 we partition the states into $\Sigma^{\mathrm{s}}, \Sigma^{\mathrm{m}}, \Sigma^{\mathrm{l}}$ according to the bias as follows:

$$\begin{aligned} \Sigma^{\mathrm{s}} &= \{x \in \Omega : bias(x) < \check{\epsilon}\} \\ \Sigma^{\mathrm{m}} &= \{x \in \Omega : \check{\epsilon} \le bias(x) \le 1/2 - \hat{\epsilon}\} \\ \Sigma^{\mathrm{l}} &= \{x \in \Omega : 1/2 - \hat{\epsilon} < bias(x)\} \end{aligned} \tag{9.7}$$

The small constants $\hat{\epsilon}$ and $\check{\epsilon}$ depend on the process.

Theorem 9.11. *Given $\mathcal{M} = (\Omega, P)$ defined in lemma 9.9, if there exist constants $\hat{\epsilon}$ and $\check{\epsilon}$ defining the partition $\Sigma^{\mathrm{s}}, \Sigma^{\mathrm{m}}$, and Σ^{l} and some constants $p^+, A_1, A_2, A_3, B_1 > 0$,*

and $0 < r, A_2, A_3 < 1$ *such that*

$$p^+ < p^+(x) \leq 1 \text{ if } x \in \Sigma^{\mathrm{s}}, \Sigma^{\mathrm{m}} \tag{9.8}$$

$$r < \frac{p^+(x)}{(1/2 - bias(x))} \leq 1 \text{ if } x \in \Sigma^{\mathrm{l}} \tag{9.9}$$

and

$$\frac{p^-(x)}{p^+(x)} \leq 1 + A_1 \left(\frac{B_1}{\sqrt{n}} - bias(x) \right) \text{ if } x \in \Sigma^{\mathrm{s}} \tag{9.10}$$

$$\frac{p^-(x)}{p^+(x)} \leq 1 - A_2 \text{ if } x \in \Sigma^{\mathrm{m}} \tag{9.11}$$

$$\frac{p^-(x)}{p^+(x)} \leq 1 - A_3 \text{ if } x \in \Sigma^{\mathrm{l}}, \tag{9.12}$$

the maximum expected hitting time is

$$\max_{x \in \Omega} \mathbb{E}_{\mathcal{M}}[\tau(x)] = O(n \log n)$$

where $\tau(x)$ *is the hitting time from state* x *to* $\{0^n, 1^n\}$.

The proof of theorem 9.11, it is rather straightforward using lemma 9.9, and carefully constructing the potential function from the recursive Equation (9.6).

9.4.3 Proof of Theorem 9.8

In this section, we will use theorem 9.11, to prove an $O(n \log n)$ time bound by exploiting properties of our process. Specifically, let our node dynamic $\mathcal{M} = \mathrm{ND}(G, f, X_0)$ be a node dynamic over G sampled from $\mathcal{G}$, it is sufficient to prove an upper bound for $1/p_G^+(x)$ and $p_G^-(x)/p_G^+(x)$. Note that we use subscripts to emphasize the dependency of the graph G.

To apply theorem 9.11, we partition the states into three groups Σ^s, Σ^m, and Σ^l defined in (9.7). The constants $\check{\epsilon}$ and $\hat{\epsilon}$ depend on the update function f and the probability of an edge p and will be specified later. Figure 9.4.3 illustrates the partitions of the states.

The following lemma upper bounds $1/p_G^+(x)$:

Lemma 9.12 (lower bound for $p_G^+(x)$)**.** *Given node process* $\mathcal{M}$ *on* G*, if* G λ*-expander with nearly uniform degree* $E(\delta_d)$*,* $\delta_d < 1$ *and* $\lambda^2 < \frac{1-\delta_d}{1+\delta_d} \cdot \min\{\frac{\hat{\epsilon}}{18}, \frac{(1/2-\hat{\epsilon})^2}{2}\}$*, then for*

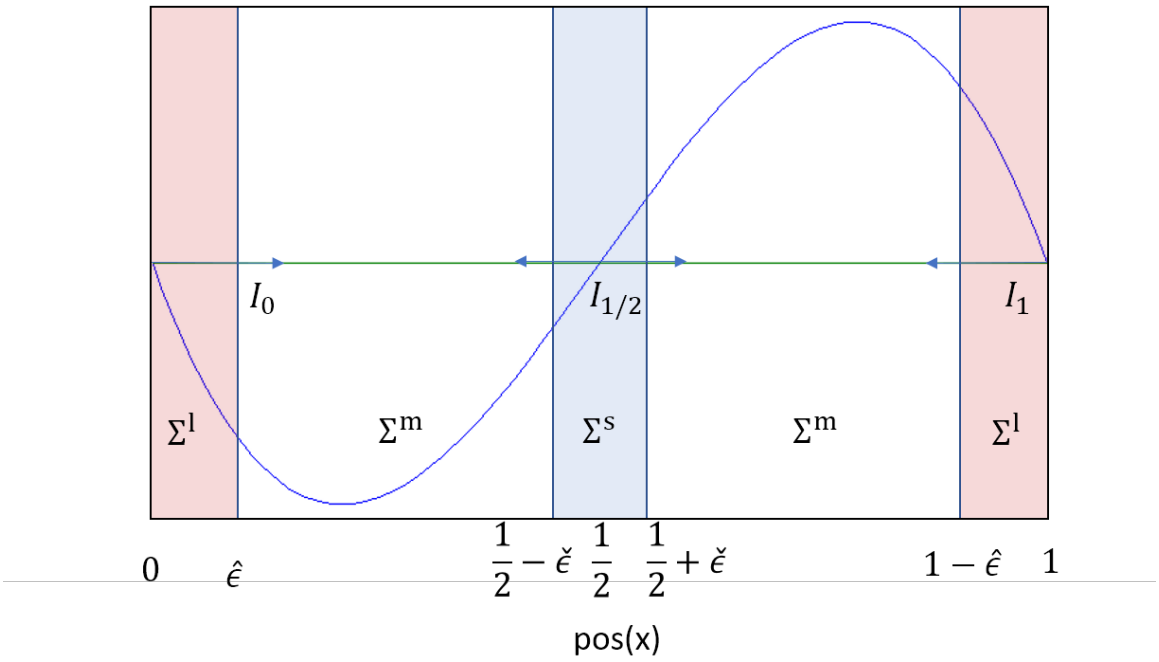


Figure 9.2: An illustration of partition in section 9.4.3.

$p^+ \triangleq \frac{\hat{\epsilon}}{2} f\left(\frac{\hat{\epsilon}}{2}\right)$,

$$p^+ < p_G^+(x) \leq 1 \text{ if } x \in \Sigma^s \cup \Sigma^m \tag{9.13}$$

$$\frac{1}{4} < \frac{p_G^+(x)}{(1/2 - bias(x))} \leq 1 \text{ if } x \in \Sigma^l. \tag{9.14}$$

This lemma is proved by apply mixing lemma 2.45 to show that the probability of increasing bias is (1) larger than some constant for $x \in \Sigma^s \cup \Sigma^m$ lemma E.1 and (2) proportional to the size of minority in Σ^l in lemma E.2. The proof details are in appendix E.2.

The second part follows from the following lemma:

Lemma 9.13 (upper bound for $p_G^-(x)/p_G^+(x)$)**.** *Given node process $\mathcal{M}$ on G, if $G \sim \mathcal{G}$, then there exist positive constant A_1, A_2, A_3, B_1, and $0 < A_2, A_3 < 1$ such that, with high probability,*

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 + A_1 \left(\frac{B_1}{\sqrt{n}} - bias(x) \right) \text{ if } x \in \Sigma^s \tag{9.15}$$

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 - A_2 \text{ if } x \in \Sigma^m \tag{9.16}$$

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 - A_3 \text{ if } x \in \Sigma^l \tag{9.17}$$

Instead of bounding $p_G^-(x)/p_G^+(x)$ directly, the drift $D_G(x) \triangleq p_G^+(x) - p_G^-(x)$ is more natural to work with. Taking the complete graph as example, $D_G(x) = f(pos(x)) - pos(x)$. Therefore instead of proving an upper bound of $p_G^-(x)/p_G^+(x)$ directly, we prove an lower bound for the drift in Appendix E.2 (lemma E.3, E.4, E.5, and E.9). Combining with lemma 9.12, these gives us an desired upper bound for $p_G^-(x)/p_G^+(x)$.

Proof of theorem 9.8. By corollary 2.44 $G \sim \mathcal{G}$ is a $O\left(\sqrt{\frac{\log n}{np}}\right)$-expander with high probability. Thus, we can apply lemma 9.12 and 9.13 to theorem 9.11, which finishes the proof. □

9.5 The Stabilizing Consensus Problem

The consensus problem in the presence of an adversary (known as Byzantine agreement) is a fundamental primitive in the design of distributed algorithms.

For the stabilizing-consensus problem—a variant of the consensus problem, Doerr et al. [75] proves synchronized 3-majority converges fast to an almost stable consensus on a complete graph in the presence of $O(\sqrt{n})$-dynamic adversaries which, at every round, can adaptively change the opinions of up to $O(\sqrt{n})$ nodes.

Here we consider an asynchronous protocol for this problem:

Definition 9.14. Given a complete network of n anonymous nodes with update function f, and $F \in \mathbb{N}$. In the beginning configuration, each node holds a binary opinion specified by $x_0(\cdot)$. In each round:

1. An adaptive dynamic adversary can arbitrarily corrupt up to F agents, and change the reports of their opinions in this run (the true opinion of these nodes is restored and will be reported once the adversary stops corrupting them).

2. A randomly chosen node updates its opinion according to node dynamics. (If the chosen node is corrupted by adversary in that run, the adversary can arbitrarily update the opinion of the chosen node.)

Definition 9.15 (n^γ-almost consensus)**.** We say a complete network of n anonymous nodes reaches an n^γ-almost consensus if all but $O(n^\gamma)$ of the nodes support the same opinion.

Our analysis in section 9.4 can be naturally extended to the stabilizing consensus problem and proves all majority-liked update functions (definition 9.8) are stabilizing almost consensus protocols and have the same convergence rate.

Theorem 9.16. *Given n nodes, fixed $\gamma > 1/2$, $F = O(\sqrt{n})$, and initial configuration $X_0 \in \{0,1\}^n$, the node dynamic $\mathrm{ND}(K_n, f, x_0)$ on a complete graph with update function f reaches an n^γ-almost consensus in the presence of any F-corrupt adversary within $O(n \log n)$ rounds with high probability.*

Remark 9.17. The goal of this section is not to promote majority-liked node dynamics as a state-of-art protocol for the stabilizing consensus problem, but to show the versatile power of the our framework of proving convergence time in section 9.2.1. Additionally we modify the formulation of the problem here to make our presentation more cohesive.

Let the random process with the presence of some fixed F-dynamic adversary $\mathcal{A}_F$ defined in theorem 9.16 be denoted $\mathcal{X}(\mathcal{A}_F) = (X_t)_{t\geq 0}$. Observe that our framework in section 9.4.2 only works for Markov chain, but with the presence of adaptive adversary the process is no longer a Markov chain. As a result, we "couple" this process with a nice Markov chain $\mathcal{Y}(F) = (Y_t)_{t\geq 0}$, and use the Markov chain as a proxy to understand the original process.

The proof has two parts: we first define the proxy Markov chain $\mathcal{Y}(F)$ and prove an upper bound of almost consensus time by using the tools in section 9.4.2. Secondly, we construct a monotone coupling between $\mathcal{Y}(F)$ and $\mathcal{X}(\mathcal{A}_F)$ to prove $\mathcal{X}(\mathcal{A}_F)$ also converges to almost consensus fast.

9.5.1 Upper Bounding the Expected Almost Consensus Time for $\mathcal{Y}(F)$.

With the notation defined in section 9.4, we define $\mathcal{Y}(F)$. Informally, we construct $\mathcal{Y}(F)$ as a pessimistic version of $\mathrm{ND}(K_n, f, X_0)$ with the presence of an adversary: at every round the adversary tries to push the state toward the unbiased configuration, and it always corrupts F nodes with the minority opinion.

Initially, $Y_0 = X_0$. At time t if we set $y = Y_t$, Y_{t+1} is uniformly sampled from

$$\begin{aligned} &\{y' \in \Omega : \exists i \in [n], \forall j \neq i, y'_j = y_j\} \\ &\cap \{y' \in \Omega : Bias(y') = Bias(y) + 1\} \end{aligned} \tag{9.18}$$

with probability $\max\{f\left(\frac{1}{2} + bias(y)\right)\left(\frac{1}{2} - bias(y)\right) - \frac{(M_1+1)F}{n}, 0\}$, or uniformly sampled from

$$\begin{aligned} &\{y' \in \Omega : \exists i \in [n], \forall j \neq i, y'_j = y_j\} \\ &\cap \{y' \in \Omega : Bias(y') = Bias(y) - 1\} \end{aligned} \tag{9.19}$$

with probability $\min\{f\left(\frac{1}{2} - bias(y)\right)\left(\frac{1}{2} + bias(y)\right) + \frac{(M_1+1)F}{n}, 1\}$; otherwise Y_{t+1} stays the same: $Y_{t+1} = y$.

Recall that the time to reach an n^γ-almost consensus is the hitting time to the set of states

$$A_\gamma \triangleq \{y \in \Omega : bias(y) > 1/2 - n^{-(1-\gamma)}\},$$

and we use $T_\gamma(z)$ to denote the hitting time to a set of state A_γ.

Lemma 9.18. *The expected n^γ-almost consensus time of the Markov chain $\mathcal{Y}$ is* $\max_y \mathbb{E}_{\mathcal{Y}(F)}[T_\gamma(y)] = O(n \log n)$.

This lemma is very similar to theorem 9.8 and we defer the proof to the full version.

9.5.2 Monotone Coupling Between $\mathcal{Y}(F)$ And $\mathcal{X}(\mathcal{A}_F)$.

To transfer the upper bound of $\mathcal{Y}(F)$ to $\mathcal{X}(\mathcal{A}_F)$, we need to build a "nice" coupling between them which is characterized as follow:

Definition 9.19 (Monotone Coupling)**.** Let X, Y be two random variables on some partially ordered set $(\Sigma, \geq)$. Then a *monotone coupling between X and Y* is a measure $(\tilde{X}, \tilde{Y})$ on $\Sigma \times \Sigma$ such that

- The marginal distributions $\tilde{X}$ and X have the same distribution;
- The marginal distributions $\tilde{Y}$ and Y have the same distribution;
- $\Pr_{(\tilde{X},\tilde{Y})}[\tilde{X} \geq \tilde{Y}] = 1$.

Note that the function $bias(\cdot)$ induces a natural total order $\leq_{bias}$ of our state space $\Omega = \{0,1\}^n$ such that for $x, y \in \Omega$, $x \leq_{bias} y$ if and only if $bias(x) \leq bias(y)$. We can also define a partial order over sequences of states: given two sequences $(X_t)_{t\geq 0}$, $(Y_t)_{t\geq 0}$ we call $(X_t)_{t\geq 0} \leq_{bias} (Y_t)_{t\geq 0}$ if $\forall t \geq 0$ $X_t \leq_{bias} Y_t$. We use calligraphic font to represent the whole random sequence, e.g. $\mathcal{Z} = (Z_t)_{t\geq 0}$.

Lemma 9.20. *There exists a monotone coupling $(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}})$ between $\mathcal{X}(\mathcal{A}_F)$ and $\mathcal{Y}(F)$ under the partial order $\leq_{bias}$*

The proof of this lemma is straightforward, and we defer the proof to the full version.

9.5.3 Proof of Theorem 9.16

Proof of theorem 9.16. We call an event A *increasing* if $x \in A$ implies that any $y \geq x$ is also in A. Observe that $A_\gamma := \{y \in \Omega : bias(y) > 1/2 - n^{-(1-\gamma)}\}$ is increasing with respect to $\leq_{bias}$. Therefore given a random sequence $\mathcal{Z} = (Z_t)_{t\geq 0}$

$$\Pr_{\mathcal{Z}}[T_\gamma(z) > \tau] = \Pr_{\mathcal{Z}}\left[\max_{t\leq\tau} bias(Z_t) \leq 1/2 - n^{-(1-\gamma)}\right]$$

By lemma 9.20, for fixed $\tau > 0$ and initial configuration $z \in \Omega$:

$$\begin{aligned}
&\Pr_{\mathcal{X}(\mathcal{A}_F)}[T_\gamma(z) > \tau] \\
&= \Pr_{\mathcal{X}}\left[\max_{t\leq\tau} bias(X_t) \leq 1/2 - n^{-(1-\gamma)}\right] \\
&= \Pr_{(\tilde{\mathcal{X}},\tilde{\mathcal{Y}})}\left[\max_{t\leq\tau} bias(\tilde{X}_t) \leq 1/2 - n^{-(1-\gamma)}\right] \\
&= \Pr_{(\tilde{\mathcal{X}},\tilde{\mathcal{Y}})}\left[\max_{t\leq\tau} bias(\tilde{X}_t) \leq 1/2 - n^{-(1-\gamma)}, \tilde{\mathcal{X}} \geq_{bias} \tilde{\mathcal{Y}}\right] \\
&\leq \Pr_{(\tilde{\mathcal{X}},\tilde{\mathcal{Y}})}\left[\max_{t\leq\tau} bias(\tilde{Y}_t) \leq 1/2 - n^{-(1-\gamma)}\right] \\
&= \Pr_{\mathcal{Y}(F)}[T_\gamma(z) > \tau].
\end{aligned}$$

On the other hand, applying Markov's inequality

$$\Pr_{\mathcal{Y}(F)}[T_\gamma(z) > \tau] \leq \frac{E_{\mathcal{Y}(F)}[T_\gamma(z)]}{\tau},$$

and by lemma 9.18, $\Pr_{\mathcal{Y}(F)}[T_\gamma(z) > \tau]$ can be arbitrary small by taking $\tau = O(n \log n)$ which finishes the proof. □

CHAPTER 10

Planted Community Networks

Continuing the study in Chapter 9, we study the Node Dynamics on networks with two communities. Each node has one of two opinions and updates its opinion as a "majority-like" function of the frequency of opinions among its neighbors. The networks we consider are weighted graphs each of which is comprised of two equally sized communities where intracommunity edges have weight p, and inter-community edges have weight q. Thus, q and p parameterize the connectivity between the two communities.

We prove a dichotomy theorem about the interaction of the two parameters: 1) the "majority-like" update function, and 2) the level of inter-community connectivity. For each setting of parameters, we show that either the system quickly converges to consensus with high probability in time $\Theta(n \log(n))$, or the system can get "stuck" and take time $2^{\Theta(n)}$ to reach consensus. We note that $O(n \log(n))$ is optimal because it takes this long for each node to even update its opinion. Technically, we achieve this fast convergence result by exploiting the connection between a family of reinforced random walks and dynamical systems literature. Our main result shows if the systems is a reinforced random walk with a gradient-like function, it converges to an arbitrary neighborhood of a local attracting point in $O(n \log n)$ time with high probability. This result adds to the recent literature on saddle-point analysis and shows a large family of stochastic gradient descent algorithm converges to a local minimal in $O(n \log n)$ time when the step size is $O(1/n)$.

Our opinion dynamics model captures a broad range of systems, sometimes called interacting particle systems, exemplified by the voter model, the iterative majority, and iterative k-majority. The interacting particle systems have found use in many disciplines including distributed systems, statistical physics, social networks, and Markov chain theory.

10.1 Introduction

Opinion dynamics on networks study how a set of opinions evolve over a network. In this case, we study how two mutually exclusive competing opinions evolve. This general model captures important settings in sociology (competing beliefs or practices), economics (competing technologies/brands), statistical physics (e.g., the Ising Model), distributed computer systems (consensus), and evolutionary biology (genetic inheritances).

We study the maximum expected consensus time on a general set of stochastic process called ***Node dynamics*** [206, 208]. Node dynamics are parameterized by an update function $f : [0, 1] \to [0, 1]$. In the beginning, each agent holds a binary "opinion", either red or blue. Then, in each round, an agent is uniformly chosen and updates its opinion to red with probability $f(r)$ and blue with probability $1 - f(r)$ where r is the fraction of its neighbors with the red opinion.

By changing f, one can capture many previously studied dynamics including:

Voter Model: Update a node's opinion to that of a randomly chosen neighbor.

Iterative majority: Update a node's opinion to the majority opinion its neighbors.

Iterative k-majority: Update a node's opinion to the majority opinion of k randomly chosen (with replacement) neighbors.

Iterative ρ-noisy majority model: [85, 100] Update a node's opinion to majority opinion its neighbors with probability $1 - \rho$ and uniformly at random with probability ρ.

We focus on a specific regime of f that we call "majority-like" (definition 10.2). In particular f should be monotone, symmetric, twice continuously differentiable, strictly convex in $[0, .5]$, and have $f(0) = 0$. Though node dynamics with majority-like function do not contain the iterative majority (non-smooth) nor voter model (not strictly convex), this still captures rich get richer property and yields a very general class of functions including k-majority. Such functions promote consensus within a single homogeneous community [206]. This result is intuitive because once symmetry is broken, the majority should retain its advantage. Here we study whether such dynamics also quickly reach consensus in models with community structure from arbitrary initial states.

While previous work on this general model only considered Erdös-Rényi graphs, we study what happens when community structure is present. We model this with a

planted community model where n nodes on a complete weighted graph are divided into two equal sets which we call communities. Edges within each community have weight p while edges spanning both communities have weight q. This can also be thought of a block-model which has a long history in the sociology literature [209].

Our Contributions We prove a dichotomy theorem about *the interaction of the update function and the level of inter-community connectivity*. For each "majority-like" function we show a threshold such that if the communities (the difference between p and q) are more connected than this threshold value, they will converge to consensus in time $\Theta(n \log(n))$ from arbitrary initial states. However, if they are more isolated than this threshold value, an attracting, non-convergence fixed point will emerge which can delay consensus to $2^{\Omega(n)}$ steps. For technical reasons, there may exist a single point in each region (above and below the threshold) that we cannot classify.

There are two challenges for showing fast consensus from worst case initial state: *breaking symmetry* and *measuring progress*. For breaking symmetry, if initially precisely half of the agents have the red opinion in both each community, our process does not move toward consensus in expectation. Thus, we show the randomness in the process can break this symmetry and help escaping such singular states. On the other hand, we also need to measure the global progress of the dynamics to prevent it making cycle or having other complicated recurrent behavior.

Our analysis is mostly decoupled from the particular problem we are solving, but instead relies on two properties: the mean-field dynamics is a gradient-like flow (which implies the existence of a potential function for us to measure the progress), and the dynamics are reinforcing random walks (which enables us to break symmetry). Thus we believe there will be other applications of it in the future.

To break symmetry, our main technical lemma shows that a dynamics with the two properties mentioned above can quickly escape from non-attracting fixed points. This result adds to the recent literature on saddle-point analysis. In particular, the process studied in Theorem 10.6 greatly generalize Ge et al. [98], Jin et al. [121], and we prove the convergence time is $O(\ell \log \ell)$ as long as the noise is well-behaved, and the objective function has a continuous third derivative.

Finally, our work has applications to ideological polarization [36]. The threshold behavior implies that even when the dynamics are very polarized, a small change in the network or processes can lead to large-scale consensus. Conversely, if the dynamics are far from the threshold, small measures may yield no effect at all.

10.1.1 Related Work

The most closely related work is the rigorous treatment of node dynamics on dense Erdös-Rényi graphs [206] which shows that it converges in time $O(n \log(n))$. The technical difficulties dealt with by these works are largely orthogonal. The difficulty with Erdös-Rényi graphs is that the exponentially many configurations are all slightly different. Thus the potential function must be custom designed for each function f but done in an automated fashion. However, the current treatment assumes a complete graph with weighted edges, and so the state can be succinctly represented. Here the difficulty in constructing a nice potential function comes from the fact that there may be more than one non-attracting fixed point.

As mentioned, our model extends several previously studied dynamics including the voter model, iterative majority, iterative k-majority. The voter model has been extensively studied in mathematics [65, 114, 150, 151], physics [32, 47], and even in social networks [46, 212, 221, 222, 64]. A major theme of this work is how long it takes the dynamics to reach consensus on different network topologies. Works about iterative majority dynamics [136, 34, 125, 176, 224, 248] often study when the dynamics converge and how long it takes them to do so. Another interest question, orthogonal to those explored here, is whether the dynamics converge to the original majority opinion—that is, successfully aggregate the original opinion. Doerr et al. [75] prove 3-majority reaches "stabilizing almost" consensus on the complete graph in the presence of $O(\sqrt{n})$-dynamic adversaries. Many works extend this result beyond binary opinions [66, 63, 31, 1].

Another line of related literature is about designing and analyzing algorithms for consensus on social networks. When dealing with binary opinions, these works typically study more elaborate dynamics which, in particular, include nodes having memory beyond their opinion [126, 190, 33, 174]. Another line of work deals with agents selecting an opinion from among a large (or infinite) set of options [27, 96]. There are also myriad models where the opinions space is continuous instead of discrete. Typically agents either average their neighbors' opinions [72], or a subset of their neighbors' opinions which are sufficiently aligned [111, 71]. Finally, models involving the coevolution of the opinions and the network [115, 79] have been studied using simulations and heuristic arguments.

A large volume of literature is devoted to bounding the hitting time of different Markov process and achieving fast convergence. The techniques typically employed are (1) showing the Markov chain has fast mixing time [171], (2) reducing the dimension of the process into small set of parameters (e.g., the frequency of each opinion)

and using a mean field approximation and concentration property to control the behavior of the process [31], or (3) using handcrafted potential functions [176]. Our results extend the second approach. We map our high dimensional process into a process on a low dimensional space ($\mathbb{R}^2$). This new process is a reinforced random walk with small step size which is closely related to the solution of an ordinary differential system which can be seen as a mean-field approximation of our random walk. However, the mean-field of our dynamics has unstable fixed points and does necessarily not have a nice potential function. We circumvent these challenges by exploiting the literature of dynamical systems and showing the existence of a potential function by analyzing the phase portrait of the flow. Additionally, we show the process leaves unstable fixed points by using the stochastic nature of our process.

Recently, there is a long line of research of stochastic gradient descent on non-convex functions, see [98, 121] and the reference therein. Searching for the minimum value of a non-convex function is in general unfeasible, and those work focus on finding local minimal efficiently which is achieved by showing that stochastic gradient decent leaves non minimal singular points (repelling and saddle fixed points) efficiently.

10.2 Graph with community structure and node dynamics

In this work, we consider blockmodels with two communities:

Definition 10.1 (bi-blockmodel [73, 237])**.** Given $p > q > 0$, and the set of n vertices V which can be decomposed into two equal size communities V_1 and V_2, we define a weighted complete graph $K(n, p, q) = (V, w)$ a *Graph* where

$$w(u, v) = \begin{cases} p \text{ if } u, v \text{ are in the same community;} \\ q \text{ otherwise.} \end{cases} \tag{10.1}$$

A *configuration* $\sigma^{(G)} : V \to \{0, 1\}$ assigns the "color" of each node $v \in G$ to be $\sigma^{(G)}(v)$, equivalently $\sigma^{(G)} \in \{0, 1\}^n$. We will usually suppress the superscript when it is clear. Moreover in a configuration σ we say v is *red* if $\sigma(v) = 1$ and is *blue* if $\sigma(v) = 0$. We then write the set of red vertices as $\sigma^{-1}(1)$. We say that a configuration σ is *in consensus* if $\sigma(\cdot)$ is the constant function (so all nodes are red or all nodes are blue), and call these two states *consensus states*.

Given a node v in configuration σ we define

$$r_\sigma(v) = \frac{\|w(v, \sigma^{-1}(1))\|}{\|w(v, V)\|} = \frac{\sum_{u \in V:\sigma(u)=1} w(v,u)}{\sum_{u \in V} w(v,u)} \tag{10.2}$$

to be its fractional weight of the red neighbors $\sigma^{-1}(1)$.

Definition 10.2. An *update function* is a $\mathcal{C}^2$ function $f_{\text{ND}} : [0,1] \to [0,1]$ with the following properties:

Monotone $\forall x, y \in [0,1]$, if $x < y$, then $f_{\text{ND}}(x) \leq f_{\text{ND}}(y)$.

Symmetric $\forall t \in [0,1]$, $f_{\text{ND}}(1/2 + t) = 1 - f_{\text{ND}}(1/2 - t)$.

Absorption $f_{\text{ND}}(0) = 0$ and $f_{\text{ND}}(1) = 1$.

In this work, we further assume the update function has an "S" shape— f is strictly convex in $[0, 0.5]$, and strictly concave in $[0.5, 1]$, and called such function ***smooth majority-like update function***

We define node dynamics as follows:

Definition 10.3. Given a undirected edge-weighted graph $G = (V, w)$, an update function f_{ND} and an initial configuration σ_0, a ***node dynamic*** $\text{ND}(G, f_{\text{ND}}, \sigma_0)$ is a stochastic process over configurations, $\left\{S_k^{\text{ND}}\right\}_{k \geq 0}$ where $S_0^{\text{ND}} = \sigma_0$ is the initial configuration. The dynamics proceeds in rounds. At round $k+1$, a node v is picked uniformly random, v updates its opinion

$$S_{k+1}^{\text{ND}}(v) = \begin{cases} 1 & \text{with probability } f_{\text{ND}}(r_{S_k^{\text{ND}}}(v)), \\ 0 & \text{otherwise} \end{cases}$$

where $r_{S_k^{\text{ND}}}(v)$ is the fractional weight of the red neighbors with configuration S_k^{ND} defined in Equation (10.2), and we further define $S_k = S_k^{\text{ND}}$ and $r_k \triangleq r_{S_k^{\text{ND}}}(v)$ in the later discussion.

In this paper, we will use consensus time to study on the interaction between update function f in Definition 10.2 and community structure of G in definition 10.1. Note that we can assume $K(n, p, q)$ with $p + q = 1$ which does not change the node dynamics.

Definition 10.4. The *consensus time* of a node dynamic $\text{ND}(G, f_{\text{ND}}, \sigma_0)$ is a stopping time $T(G, f_{\text{ND}}, \sigma_0)$ denoting the first time step that ND is in a consensus configuration. The *maximum consensus time* $\text{ME}(G, f_{\text{ND}})$ is the maximum consensus time over any initial configuration, $\text{ME}(G, f_{\text{ND}}) = \max_{\sigma_0} \mathbb{E}[T(G, f_{\text{ND}}, \sigma_0)]$.

10.3 Main results

Theorem 10.5. *Given a smooth majority-like function f_{ND} in Definition 10.2, let $(S_k^{\mathrm{ND}})_{k\geq 0} = \mathrm{ND}(G, f_{\mathrm{ND}}, \sigma_0)$ be a node dynamic over $K(n,p,q)$ where $p > q > 0$ and $p + q = 1$. There are three constants δ', δ^* and δ'' such that $0 < \delta' < \delta^* \leq \delta'' < 1$*

1. *If $p - q \in (0, \delta^*) \setminus \{\delta'\}$, the maximum expected consensus time*

$$\mathrm{ME}(K(n,p,q), f_{\mathrm{ND}}) = O(n \log n).$$

2. *If $p - q \in (\delta^*, 1) \setminus \{\delta''\}$, the maximum expected consensus time*

$$\mathrm{ME}(K(n,p,q), f_{\mathrm{ND}}) = \exp\left(\Omega(n)\right).$$

We prove the first part of the Theorem 10.5, fast convergence result, in three parts:

1. We first construct a function ϕ and show both the process $\phi(S_k^{\mathrm{ND}})$ is a reinforced random walk with a gradient-like function and only the images of consensus states are the attracting fixed points of the gradient-like flow (Theorem 10.7).

2. We next show a general theorem that a family of reinforced random walks with a gradient-like function reaches an arbitrary neighborhood of some attracting fixed point in $O(n \log n)$ with high probability under mild conditions on the perturbation (Theorem 10.6). Combining these we can show our process S_k^{ND} gets close to the consensus states in $O(n \log n)$ with high probability. In section F.3 we additionally show the process indeed hits the consensus states after arriving the neighborhoods of consensus states.

The second part is relatively straightforward, and proved in Section F.3.

10.3.1 Fast convergence result of reinforced random walk

Informally, if the Markov chain S_k of interest can be mapped into $X_k \triangleq \phi(S_k)$ such that X_k is a reinforced random walk in $\mathbb{R}^d$ with a gradient-like function f, then Theorem 10.6 shows that the behavior of the reinforced random walk with f is closely related to its mean field—the flow with f. By the definition of the gradient-like flow with f, the flow (mean field) converges to the (repelling, attracting, and saddle) fixed points of f. The theorem, on the other hand, shows the process X_k converges

to an arbitrary neighborhood of an attracting fixed point fast as long as the noise around repelling and saddle points is sufficiently large. Intuitively, this noise allows the process X_k to quickly escape from any non-attracting fixed point, this is unlike analogous the flow (mean field).

Theorem 10.6 (Hitting time of reinforced random walk)**.** *Let S_k be a time homogeneous Markov chain on state space Ω. If there exist constants $d \in \mathbb{N}$, $D, d_1, d_2 \in \mathbb{R}_+$, a function $\phi : \Omega \to \mathbb{R}^d$, a compact set $B \subset \mathbb{R}^d$, and $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^d)$ with the set of fixed points $\mathrm{Fix}_f = \{\beta_1, \ldots, \beta_m\}$ for any $\{N_i\}_{1 \le i \le m}$ which is a set of open sets in $\mathbb{R}^d$ with $\beta_i \in N_i$, such that*

1. *the flow with f in (2.1) is a gradient-like system and B is positive invariant under the flow with f,*
2. *$\{\boldsymbol{X}_k\}_{k \in \mathbb{N}} = \{\phi(S_k)\}_{k \in \mathbb{N}}$ is a function of the Markov chain in $\mathbb{R}^d$, and*

$$\boldsymbol{X}_{k+1} = \boldsymbol{X}_k + \frac{1}{n}\left(f(\boldsymbol{X}_k) + \boldsymbol{U}(S_k)\right)$$

such that $X_k \in B$ with probability 1 and the noise is well-behaved: let $\boldsymbol{U}_{k+1} \triangleq \boldsymbol{U}(S_k)$.

   (a) For all $X_k \in B$, $\mathbb{E}[\boldsymbol{U}_{k+1} \mid S_k] = 0$,

   (b) For all $X_k \in B$, $\|\boldsymbol{U}_{k+1}\| \le D$

   (c) For all $X_k \in \cup_{i:\beta_i \in \mathrm{Repel}_f \cup \mathrm{Saddle}_f} N_i$,

$$d_1 \mathbb{I}_d \prec \mathrm{Cov}[\boldsymbol{U}_{k+1} \mid S_k] \prec d_2 \mathbb{I}_d$$

there exists $\tau = O(n \log n)$ for all $\boldsymbol{X}_0 \in B$, $\boldsymbol{X}_T \in \cup_{i:\beta_i \in \mathrm{Attract}_f} N_i$ for some $T \le \tau$ with high probability.

The proof is in Section 10.4. Note that when the reinforced function is gradient-like, by Corollary 2.35, there is a complete Lyapunov function for the flow, and we can use it to control the global movement of the reinforced random walk. However, the technical difficulty is how to control the behavior of the reinforced random walk around fixed points, and this is carefully analyzed in Section 10.4.

In Theorem 10.18, we show the connection between stochastic gradient descent on non-convex functions and Theorem 10.6. Informally, Theorem 10.6 ensures that the reinforced random walk with the gradient of a non-convex function converges to a local minimal (attracting fixed point) in $O(n \log n)$.

10.3.2 Phase portrait

To apply Theorem 10.6, we need to show our node dynamics on bi-blockmodel are time homogeneous Markov chains which can be mapped to $\mathbb{R}^2$ as a reinforced random walk with gradient-like function. In this section, we define such a mapping ϕ, and show $(\phi(S_k^{\mathrm{ND}}))$ itself is a Markov chain which is stronger than reinforced random walk defined in Equation (2.3).

Because of the symmetry of S_k^{ND} and bi-blockmodel defined in Definition 10.1, we define a mapping from $\sigma = \{0,1\}^d$ to the numbers of red nodes in those two communities, $Pos : \{0,1\}^n \to [n/2] \times [n/2]$ and $\phi = pos \triangleq 2Pos/n$ such that

$$Pos(\sigma) = \left(\|V_1 \cap \sigma^{-1}(1)\|, \|V_2 \cap \sigma^{-1}(1)\|\right) = \left(\sum_{v \in V_1} \sigma(v), \sum_{u \in V_2} \sigma(u)\right). \tag{10.3}$$

Then we define the process: for all $k \in \mathbb{N}_0$

$$X_k^{\mathrm{ND}} \triangleq \phi(\sigma_k), \tag{10.4}$$

which is a function of Markov chain S_k^{ND}, and it is a two-dimensional reinforced random walk, as defined in (2.3), with $F_{\mathrm{ND}} : [0,1]^2 \to \mathbb{R}^2$ defined as follows:

$$F_{\mathrm{ND}}(x_1, x_2) \triangleq \left(f_{\mathrm{ND}}\left(p\, x_1 + q\, x_2\right) - x_1,\ f_{\mathrm{ND}}\left(p\, x_2 + q\, x_1\right) - x_2\right). \tag{10.5}$$

Moreover, X_k^{ND} itself is also a Markov chain and the consensus states 0^n and 1^n is mapped o $(0,0)$ and $(1,1)$ respectively. As a result, we will use X^{ND} and S^{ND} interchangeably as our node dynamics.

As mentioned in section 2.2, we summarize three closely related dynamics here:

1. $\left\{S_k^{\mathrm{ND}}\right\}_{k \in \mathbb{N}_0}$ the stochastic sequence on state space $\Omega = \{0,1\}^n$, the dynamics in theorem 10.5, $\mathrm{ND}(G, f_{\mathrm{ND}}, \sigma_0)$ with $G = K(n,p,q)$,

2. $\left\{X_k^{\mathrm{ND}}\right\}_{k \in \mathbb{N}_0}$ the stochastic sequence on state space $\{0, 2/n, 4/n, \ldots, 1\}^2 \subset \mathbb{R}^2$ defined in Equation (10.4) which is a function of Markov chain and a reinforced random walk with F_{ND}.

3. $\left\{x^{\mathrm{ND}}(t)\right\}_{t \in \mathbb{R}}$ the deterministic flow in $\mathbb{R}^2$ associated with F_{ND}.[1]

[1] We need to extend the domain of F_{ND} into $\mathbb{R}^2$ which is defined in Section 10.5 and call it as $\bar{F}_{\mathrm{ND}}$.

Theorem 10.7 (Phase portrait)**.** *Given f_{ND} and p, q in the Node Dynamics defined in Theorem 10.5, there exist three constants $0 < \delta' < \delta^* \leq \delta'' < 1$ such that the flow with $\bar{F}_{\mathrm{ND}}$ defined in* (10.5) *has three cases:*

1. *When $p - q \in (0, \delta^*) \setminus \{\delta'\}$, the flow is a gradient-like system, and the consensus states $(0,0), (1,1)$ are the only attracting fixed point.*
2. *When $p - q \in (\delta^*, 1) \setminus \{\delta''\}$, F_{ND} has an attracting fixed point $\beta_a \neq (0,0), (1,1)$.*

A more detailed characterization of δ', δ^* and δ'' is in Theorem 10.14.

10.3.3 From neighborhood of fixed points to the fixed points

In Section F.3, we complete the proof of Theorem 10.5. For the first part of Theorem 10.5, we want to show the process indeed reach a consensus state 0^n or 1^n fast. Theorem 10.6 ensures our process reach a *neighborhood* of those consensus states which is not enough. In section F.3, we show after reaching arbitrary neighborhoods of consensus states the process can reach a consensus state in $O(n \log n)$ steps with constant probability. We achieve this by constructing a coupling between our process and a birth-and-death process in Lemma F.4, and an upper bound for expected hitting time of birth-and-death in Lemma F.6 and F.7.

Finally, for the second part, Theorem 10.7 shows the existence of an attracting fixed point β_a other than consensus states. By a standard argument (Lemma F.5) if the process starts at β_a, the probability of leaving Q' a neighborhood of β_a in n steps is exponentially small. Therefore the expected time of reach consensus states is $\exp(\Omega(n))$.

10.4 Reinforced random walks of gradient-like function and the proof of Theorem 10.6

This section is concerned with proving Theorem 10.6. To show the process reach a neighborhood of an attracting fixed point fast, we need to show two parts: locally, the process does not stuck at any small neighborhood; globally, the process to progress without making circle or having complicated recurrent behavior.

For global characterization, because the flow is gradient-like, by Corollary 2.35, there exists a smooth complete Lyapunov function V for the flow. With this real-value function V, we can control the behavior of the reinforced random walk X_k.

Locally, for each fixed point $\beta_i \in \mathrm{Fix}(f)$, we define a small neighborhood N_i around it containing no additional fixed points, and we set the index such that $\beta_i \in N_i$. There are two cases: either $x \in \mathcal{X} \setminus (\cup_i N_i)$, and we say x is a *regular point*. In this case the complete Lyapunov function V has large (linear) decrements. Otherwise, $x \in N_i$ for some i, we say that x is a *neighborhood point* and V decrements increasingly slowly as it approaches the fixed point β_i.

The first lemma deals with the regular points, and shows that from them the trajectory will quickly reach a non-regular point. The proof is in appendix.

Lemma 10.8 (regular points)**.** *If $X_0 \notin \cup N_i$, there exists i and $T = O(n)$ such that $X_T \in N_i$ and $V(\beta_i) < V(X_0)$ with probability $1 - o(1)$.*

The next lemma says that as long as β_i is not an attracting fixed point, then from any point in its neighborhood, the process will quickly leave the neighborhood in a manner that decreases the potential function.

Lemma 10.9 (non attracting fixed points)**.** *If $X_0 \in N_i$ and β_i is not an attracting point, there exists $\delta > 0$ such that $\tau = O(n \log n)$, $X_T \notin N_i$, and $V(\beta_i) > \delta + V(X_0)$ for some $T \leq \tau$ with high probability.*

This is proved in the appendix. The proof relies heavily on our main technical lemma, Lemma 10.10, which shows that the processes leaves saddle points (or unstable fixed points). Lemma 10.10 is proved in Section 10.4.1.

Proof of Theorem 10.6. Combining the above two characterizations, we can study the process in two alternating stages.

1. Given an initial condition $x_0 \in B$ where B is compact and positive invariant, if $x_0 \notin \cup_i N_i$, it converges to some N_i in $O(n)$ with high probability by Lemma 10.8.

2. If β_i is not an attracting point by Lemma 10.9, the process leaves the region N_i and $V(x) < V(\beta_i) - \delta$ in $O(n \log n)$ time with high probability.

3. After leaving N_i, by Lemma 10.8, the process converges to N_j a neighborhood of another fixed point β_j where $V(\beta_j) < V(\beta_i)$ in $O(n)$ steps with high probability.

4. We can repeat these arguments until the process reaches some attracting point. The processes can never return to the neighborhood of the same fixed point twice because $V(\beta(i))$ is always decreasing. Moreover since the number fixed points are constant (and independent to the step size), the alternation between the above stages stops in constant rounds.

□

10.4.1 Escaping local saddle points

In this section we state and prove our main technical lemma, which shows that our process will quickly leave the neighborhood of a saddle point or unstable fixed point.

Lemma 10.10 (Leaving non-attracting fixed point)**.** *Given the setup in Theorem 10.6, there are a sufficiently small constant $r > 0$ and $\tau_i = O(n \log n)$, such that if the process starting in N_i for some non-attracting fixed points, after $T_i \leq \tau_i$,*

$$\Pr[X_{T_i} \in B(\beta_i, 3r/4) \setminus B(\beta_i, r/2)] = 1 - o(1).$$

Roughly, around the saddle point β_i the dynamics can be approximated by a linear flow with $A = \nabla f(\beta_i)$ defined in section 2.4.3. That is the process is expanding in the subspace E^u, and contracting in subspace E^s with respect to A. However, because of the non-linearity of the process, there is a quadratic error term $O(\|Z_k\|^2)$. To handle this, we partition the process into $O(\log n)$ phases illustrated in Figure 10.1 such that as long as the difference between $\|Z_k^u\|$ and $\|Z_k^s\|$ is not too large, the errors are comparable (and small).

The proof has three parts. Intuitively, Lemma 10.11 shows the magnitude in contracting subspace decrease rapidly. Lemma 10.12 shows if the process is very close to or at β_i, the noise of the process can ensure the unstable part of the process can be $\Omega((\log n)^{1/3}/\sqrt{n})$ far away from β_i in $O(n \log n)$ times. Finally, Lemma 10.13 shows if the unstable part of the process is $\Omega((\log n)^{1/3}/\sqrt{n})$ away from β_i, the unstable part double in $O(n)$ time with probability $1 - \exp\left(-\Omega\left(\sqrt{\log n}\right)\right) = 1 - o(1/\log n)$.

Proof. Because the fixed points of the Gradient-like system are hyperbolic we can rewrite the process around $B(\beta_i, r)$ as,

$$Z_{k+1} - Z_k = \frac{1}{n}(AZ_k + \overrightarrow{O(\|Z_k\|^2)} + \text{noise}) \tag{10.6}$$

where $A = \nabla f(\beta_i)$ is hyperbolic and $Z_k = X_k - \beta_i$. Note that here we use $\overrightarrow{O(\|Z_k\|^2)}$ to denote an error *vector* such that each coordinate of this vector is $O(\|Z_k\|^2)$.

Furthermore, given the matrix A, we can decompose the tangent space $\mathbb{R}^d$ into the stable and the unstable subspaces E^s and E^u with respect to A (Section 2.4.3). Let P^u and P^s be the projection operators for E^s and E^u respectively. Without loss of generality, we consider β_i to be a saddle point.

We can consider the following two (correlated) processes which are updated by the original process decomposed into operating on the (not necessarily orthogonal)

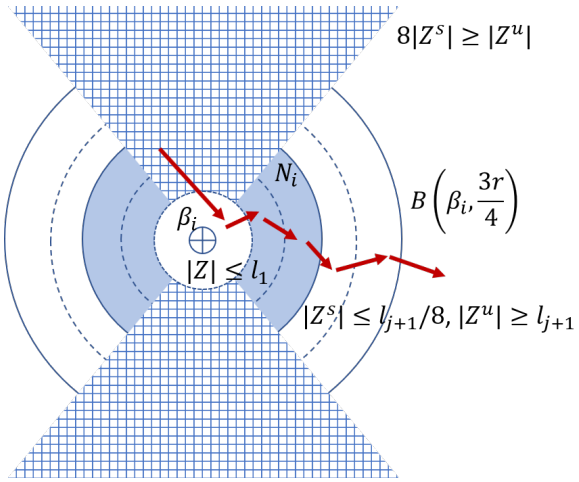


Figure 10.1: The solid blue circle represents N_i, dash circles are different phases of the process, non-grid region are those $\|Z^u\| \geq 8\|Z^s\|$, and the other solid circle is $B(\beta_i, 3r/4)$. For Lemma 10.10, we partition the process around saddle point β_i into $O(\log n)$ phases: In phase 0 the process hits N_i(the solid blue circle), and Lemma 10.11 shows the process hits $\|Z\| \leq l_1$(the smallest dash circle), in $O(n \log n)$ with probability $1-o(1)$. In the phase 1, by Lemma 10.12, after hitting $\|Z\| \leq l_1$, the process enter $\|Z^u\| \geq l_1$ (the non-grid region of the bigger dash circle) in $O(n \log n)$. The Lemma 10.13 shows if $\|Z_0^u\| \geq l_j$ the process will enter the $\|Z^u\| \geq l_{j+1}$ (the non-grid region of the next bigger dash circle) in additional $O(n)$ times. Finally, the process leaves the region $B(\beta_i, 3r/4)$.

spaces E^s and E^u:

$$Z_{k+1}^u - Z_k^u = \frac{1}{n}(AZ_k^u + \overrightarrow{O(\|Z_k\|^2)} + \text{noise}^u) \in E^u \tag{10.7}$$

$$Z_{k+1}^s - Z_k^s = \frac{1}{n}(AZ_k^s + \overrightarrow{O(\|Z_k\|^2)} + \text{noise}^s) \in E^s \tag{10.8}$$

where $\text{noise}^u \triangleq P^u \boldsymbol{U}_{k+1} \in E^u$ and $\text{noise}^s \triangleq P^s \boldsymbol{U}_{k+1} \in E^s$. We call Z_k^u and Z_k^s *unstable component* and *stable component* of the process Z_k respectively

If we can show after $\tau_i = O(n \log n)$ steps both in the stable manifold (10.8) we have $\|Z_{\tau_i}^s\| \le r/4$ and in unstable manifold (10.7) we have $\|Z_{\tau_i}^u\| \ge 3r/4$, the $\|Z_{\tau_i}\| \ge \|Z_{\tau_i}^u\| - \|Z_{\tau_i}^s\| = r/2$ which completes the proof.[2]

Let $\lambda_u = \min\{\Re(\lambda_i)\} > 0$ which is minimum real part of eigenvalue of A in E^u. We define a length $J = O(\log n)$ sequence

$$l_1 = \frac{(\log n)^{1/3}}{\sqrt{n}}, l_{j+1} = 2l_j \text{ for } j = 1, 2, \ldots, J-1, \text{ and } l_J = 3r/4. \tag{10.9}$$

With the sequence (l_j), we partition the processes in $B(\beta_i, 3r/4)$ into $O(\log n)$ phases, and say the process Z_k^u is in phases j if and only if $l_{j-1} \le \|Z_k^u\| < l_j$ and $\|Z_k^s\| \le \|Z_k^u\|/8$.

First in Lemma 10.11, we show either the stable component $\|Z^s\|$ is smaller than the unstable component $\|Z^u\|$ or enters the phase 0, $\|Z\| \le l_1$ in $O(n \log n)$ time with high probability.

Secondly, by Lemma 10.12, suppose the process is at phase 0, $\|Z_0\| \le l_1$, the process reach phase 1 within $O(n \log n)$ steps with probability $1 - o(1)$.

Finally, by Lemma 10.13, starting at phase j, the process reach phase $j+1$ with in $O(n)$ steps with probability $1 - \exp\left(-\Omega\left(\sqrt{\log n}\right)\right) = 1 - o(1/\log n)$. Thus the proof completes by taking union bound on these $J = O(\log n)$ phases. □

Due to the space constrain, we put all of the proofs of the following lemmas in to the appendix.

10.4.1.1 Phase 0: decreasing the stable component

Lemma 10.11 (Phase 0)**.** *If $X_0 \in N_i$, in time $\tau_0 = O(n \log n)$, there exists $T_0 \le \tau_0$ such that $\|Z_{T_0}^u\| \ge 8\|Z_{T_0}^s\|$ or $\|Z_{T_0}\| \le l_1$ with probability $1 - o(1)$.*

[2]Although the process Z_k may even leave $B(\beta, r)$ before τ_i such that Equation (10.6) does not hold anymore, we can define another process by Equation (10.6) and couple it with the original process when the process is in $B(\beta, 3r/4)$. We analyze the new process instead and show it leaves $B(\beta, 3r/4)$ with high probability. Therefore the original process also leave it with high probability.

10.4.1.2 Phase 1: leaving the fixed point

For Lemma 10.12, because the drift of the process is too small, we use the anti-concentration of noise (Lemma F.1) to show in expectation it can reach $l_1 = \Omega((\log n)^{1/3}/\sqrt{n})$ after $O(n(\log n)^{2/3})$ steps. By Markov inequality, we show it will happen in $O(n \log n)$ with probability $1 - o(1)$.

Lemma 10.12 (Phase 1)**.** *If $\|Z_0\| \le l_1$, there are $\tau_1 = O(n \log n)$ and $T_1 \le \tau_1$ such that $\|Z^u_{T_1}\| \ge 2l_1$ and $\|Z^s_{T_1}\| = o(l_1)$ with probability at least $1 - o(1)$.*

10.4.1.3 Phase j: amplifying the unstable component

To the end we want to show $\|Z^u_k\|$ in (10.7) increases rapidly which depends on three things: the linear part AZ^u_k is large, the nonlinear term $O(\|Z_k\|^2)$ is small and the noise, noise^u, is small. However, $O(\|Z_k\|^2)$ depends both on Z^u_k and Z^s_k, so we need to upper bound the value of $\|Z^s_k$ as well. Therefore in contrast to Lemma 10.11, to prove the $\|Z^u_k\|$ reach large value fast, we use induction because to control the process multiple quantities $\|Z^s_k\|/\|Z^u_k\|$, $\|Z_k\|$, and $\|Z^u_k\|$, and it requires more delicate argument than optional stopping time theorem.

For non-linearity because $f \in \mathcal{C}^2$ is smooth we can upper bound the *quadratic* values $\|Z_k\|^2$ by $\|Z_0\| = o(l_j)$ for all $0 \le k \le T$ with high probability. However, the standard Chernoff bound and union bound are not enough, so use a more advanced tail bound for the *maximum deviation* (Theorem 2.20). For the noise part, condition on $\|Z_k\|^2$ being small we use linear approximation of f to study two aspect to the Doob martingale $Y_k = \mathbb{E}[Z_T|Z_0, \dots, Z_k]$: 1) the effect variance $\sum c_i^2$ is small and 2) the expectation $Y_0 = \mathbb{E}[Z_T]$ is nice.

Lemma 10.13 (Phase $j > 1$)**.** *If $\|Z^s_0\| \le \frac{1}{8}\|Z^u_0\|$ and $l_j \le \|Z^u_0\| \le l_{j+1}$, $\tau_j = O(n)$ such that $\|Z^s_{\tau_j}\| \le \frac{1}{8} l_{j+1}$ and $\|Z^u_{\tau_j}\| > l_{j+1}$ with probability $1 - \exp\left(-\Omega\left(\sqrt{\log n}\right)\right)$.*

Note that in contrast to Lemmas 10.11 and 10.12 which show upper bounds for hitting times, this lemma characterizes the behavior of Z at time τ_j.

10.5 Phase portrait

In this section, we prove Theorem 10.7 (which will follow immediately from theorem 10.14), by analyzing the fixed points of the function F_{ND} defined in (10.5). We can classify the fixed points into three types: symmetric, anti-symmetric and eccentric. Lemma 10.15 characterizes the property of symmetric fixed points; Lemma 10.16,

anti-symmetric fixed points; and Lemma 10.17, eccentric fixed points. The following section introduces the symmetry property of the flow on F_{ND} and Theorem 10.14 is proved in the next one.

10.5.1 Setup and examples

The fixed points of the system x^{ND} are the zeroes of F_{ND} which can be parameterized by $\delta \triangleq p - q$:

$$\begin{aligned} 0 &= f_{\mathrm{ND}}\left(p\,x_1 + q\,x_2\right) - x_1, \\ 0 &= f_{\mathrm{ND}}\left(p\,x_2 + q\,x_1\right) - x_2. \end{aligned} \tag{10.10}$$

Denote the solutions of equation (10.10) as

$$\begin{aligned} \gamma_1 &= \left\{(x_1^{(1)}, x_2^{(1)}) \in [0,1]^2 : x_1^{(1)} = f_{\mathrm{ND}}\left(p\,x_1^{(1)} + q\,x_2^{(1)}\right)\right\} \\ \gamma_2 &= \left\{(x_1^{(2)}, x_2^{(2)}) \in [0,1]^2 : x_2^{(2)} = f_{\mathrm{ND}}\left(p\,x_2^{(2)} + q\,x_1^{(2)}\right)\right\}. \end{aligned} \tag{10.11}$$

Note that the system of Equation 10.10 is symmetric with respect to two axes $x_1 = x_2$ and $x_1 + x_2 = 1$, so we define four disjoint regions of $[0,1]^2$:

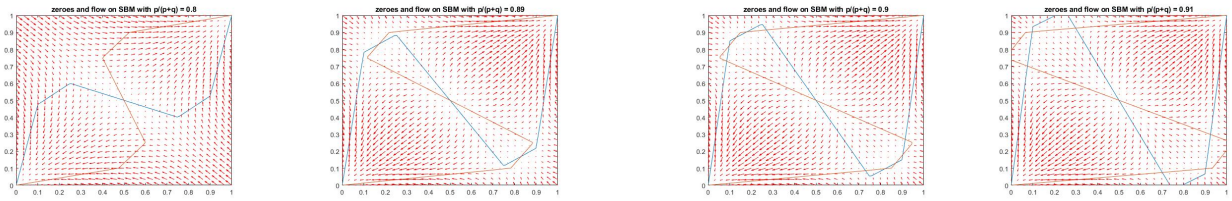
$$\begin{aligned} R_1 &= \{(x_1, x_2) \in [0,1]^2 : x_1 < x_2 \text{ and } x_1 + x_2 < 1\}, \\ R_2 &= \{(x_1, x_2) \in [0,1]^2 : x_1 < x_2 \text{ and } x_1 + x_2 > 1\}, \\ R_3 &= \{(x_1, x_2) \in [0,1]^2 : x_1 > x_2 \text{ and } x_1 + x_2 < 1\}, \text{ and} \\ R_4 &= \{(x_1, x_2) \in [0,1]^2 : x_1 > x_2 \text{ and } x_1 + x_2 < 1\}. \end{aligned}$$

With this symmetric property, we classify the fixed points of (10.10) into three types:

- symmetric fixed points: $(x_1^{(s)}, x_2^{(s)})$ such $x_1^{(s)} = x_2^{(s)}$,
- anti-symmetric fixed points:$(x_1^{(a)}, x_2^{(a)})$ such $x_1^{(a)} + x_2^{(a)} = 1$,
- eccentric fixed points:$(x_1^{(e)}, x_2^{(e)})$ such $x_1^{(e)} + x_2^{(e)} > 1$ and $x_1^{(e)} < x_2^{(e)}$.

Figure 10.2 shows some examples of a dynamic with different p, q.

To consider the dynamic $x^{\mathrm{ND}}(t)$ as a flow, there is a caveat: the function F_{ND} only has domain in $[0,1]^2$ instead of $\mathbb{R}^2$, and the set $[0,1]^2$ is not invariant since the $x^{\mathrm{ND}}(t)$ leaves $[0,1]$ if we reverse the time t. Fortunately, it's not hard to extend the domain of F_{ND} without changing the structure: let $m_1 = \lim_{x\to 1^-} f'_{\mathrm{ND}}(x)$ and $m_0 = \lim_{x\to 0^+} f'_{\mathrm{ND}}(x)$



(a) $p - q < \delta_{\text{symm}}$ (b) $\delta_{\text{symm}} < p - q < \delta_{\text{ecce}}$ (c) $\delta_{\text{ecce}} < p - q < \delta_{\text{anti}}$ (d) $\delta_{\text{anti}} < p - q$

Figure 10.2: In Theorem 10.14 there are three critical values $\delta_{\text{symm}}, \delta_{\text{ecce}}$ and δ_{anti}. In the case (a), the difference $p-q$ is smaller than $\delta_{\text{symm}} = 1/f'_{\text{ND}}(1/2)$, and there are only three fixed points characterized in Lemma 10.16. In case (b), the $p-q$ is bigger such that there are two extra saddle anti-symmetric fixed points. For some specific update function f_{ND} there is case (c) such that there are two extra eccentric fixed points but the antisymmetric fixed points are saddle which is discussed in Lemma 10.17. Finally in case (d), the $p-q$ is big enough such that the antisymmetric fixed points become attracting which is characterized in Lemma 10.16.

$$\bar{f}_{\text{ND}}(x) = \begin{cases} m_1 x \text{ if } x < 0 \\ f_{\text{ND}}(x) \text{ if } x \in [0,1] \\ m_1(x-1)+1 \text{ if } x > 1 \end{cases} .$$

We can have $\bar{F}_{\text{ND}}$ by using $\bar{f}_{\text{ND}}$ in (10.5) instead of f_{ND}.[3]

10.5.2 Proof of Theorem 10.7

The following theorem is a detailed characterization of the flow x^{ND} with F_{ND}, and Theorem 10.7 is an corollary of it. In the first case, we take $(\delta', \delta^*, \delta'') = (\delta_{\text{symm}}, \delta_{\text{ecce}}, \delta_{\text{anti}})$ and $(\delta_{\text{symm}}, \delta_{\text{anti}}, \delta_{\text{anti}})$ in the second case.

Theorem 10.14 (Phase portrait)**.** *Fix the flow x^{ND} with p, q and $\bar{F}_{\text{ND}}$ defined in* (10.5)*, depending on the property of f_{ND} there are two situations*

1. *If there exists δ_e such that equation* (10.10) *with $p_e = (1+\delta_e)/2$ has an eccentric fixed point $(x_1^{(e)}, x_2^{(e)})$ where $x_1^{(e)} + x_2^{(e)} > 1$ and $x_1^{(e)} < x_2^{(e)}$ there are three constants $\delta_{\text{symm}} < \delta_{\text{ecce}} < \delta_{\text{anti}}$ where $\delta_{\text{anti}} = 1/f'_{\text{ND}}(1/2)$ is defined in Lemma 10.15 and δ_{anti} is defined in Lemma 10.16 and δ_{ecce} defined in Lemma 10.17 such that there are three cases:*

[3]To make $\bar{f}_{\text{ND}} \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$, we can consider $\epsilon > 0$ and set $f''(x) = 0$ if $x < -\epsilon$ and set the intermediate value in $[-\epsilon, 0]$ smoothly. Then we have an $\mathcal{C}^2$ function moreover it can be arbitrary close to the above definition if we take ϵ small enough.

(a) *When* $p-q < \delta_{\text{symm}}$*, there are only three fixed points* $(0,0),(0.5,0.5),(1,1)$*. The system is a gradient-like system, and the consensus states* $(0,0),(1,1)$ *are the only attracting fixed point.*

(b) *When* $\delta_{\text{anti}} < p-q < \delta_{\text{ecce}}$*, there are five fixed points,* $(0,0),(0.5,0.5),(1,1)$ *and two anti-symmetric saddle points. The system is a gradient-like system and the consensus states* $(0,0),(1,1)$ *are the only attracting fixed point.*

(c) *When* $\delta_{\text{ecce}} < p-q < \delta_{\text{anti}}$ *or* $\delta_{\text{anti}} < p-q$*, there exists an attracting fixed point* $\beta \neq (0,0),(1,1)$*.*

2. *Otherwise, there are two constants* $\delta_{\text{symm}} < \delta_{\text{anti}}$ *where* $\delta_{\text{symm}} = 1/f'_{\text{ND}}(1/2)$ *is defined in Lemma 10.15 and* δ_{anti} *is defined in Lemma 10.16 such that the following three cases:*

   (a) *When* $p-q < \delta_{\text{symm}}$*, there are only three fixed points* $(0,0),(0.5,0.5),(1,1)$*. The system is a gradient-like system, and the consensus states* $(0,0),(1,1)$ *are the only attracting fixed point.*

   (b) *When* $\delta_{\text{symm}} < p-q < \delta_{\text{anti}}$*, there are five fixed points,* $(0,0),(0.5,0.5),(1,1)$ *and two anti-symmetric saddle points. The system is a gradient-like system and the consensus states* $(0,0),(1,1)$ *are the only attracting fixed point.*

   (c) *When* $\delta_{\text{anti}} < p-q$*, there exists an attracting fixed point* $\beta \neq (0,0),(1,1)$*.*

We will use two lemmas to proof Theorem 10.14.

Lemma 10.15 (symmetric fixed points)**.** *Given* F_{ND} *with* p,q *and* f_{ND}*, let* $0 < \delta_{\text{symm}} \triangleq 1/f'_{\text{ND}}(1/2)$*. There are three symmetric fixed points:* $(0,0),(1,1)$ *are attracting points, and* $(0.5,0.5)$ *which is a saddle point if* $(p-q) < \delta_{\text{symm}}$ *and a repelling point when* $(p-q) > \delta_{\text{symm}}$*. Moreover, when* $(p-q) < \delta_{\text{symm}}$*, the system in* (10.10) *only has the above three fixed points.*

Lemma 10.16 (anti-symmetric fixed points)**.** *Given* F_{ND} *with* p,q *and* f_{ND} *and* δ_{symm} *in Lemma 10.15, there exists* $\delta_{\text{anti}} > \delta_{\text{symm}}$ *such that there are two cases for the anti-symmetric fixed points in Equation* (10.10) *depending on the value of* $p-q$*:*

saddle *If* $\delta_{\text{symm}} < p-q < \delta_{\text{anti}}$*, there are anti-symmetric fixed points which are saddle.*

attracting *If* $\delta_{\text{anti}} < p-q$*, there are anti-symmetric fixed points which are stable.*

With Lemma 10.16, one might guess the systems only have consensus as stable fixed points when $p - q < \delta_{\text{anti}}$, and have two extra stable fixed points when $p - q > \delta_{\text{anti}}$. However, as $p - q$ increases there is some f_{ND} such that the system has extra stable eccentric fixed points before the anti-symmetric fixed points become stable, e.g. Figure 10.2. Though we can use simulation to estimate the phase space, the following lemma shows: Given f_{ND} suppose there exists $\delta_e < \delta_{\text{anti}}$ such that the system with $\delta_e = p_e - q_e$ in Equation (10.10) has an eccentric fixed point. Then there exists $\delta_{\text{ecce}} < \delta_{\text{anti}}$ such that for all p'_e such that $\delta_{\text{ecce}} < p'_e - q'_e < \delta_{\text{anti}}$ the system (10.10) has attracting eccentric stable fixed points fixed points. By symmetry, we only state the result in R_2.

Lemma 10.17 (eccentric fixed points)**.** *Given F_{ND} with p, q, f_{ND}, δ_{symm} and δ_{anti} in Lemma 10.15, 10.16, if there exists $\delta_e < \delta_{\text{anti}}$ such that equation* (10.10) *with $p_e = (1+\delta_e)/2$ has an eccentric fixed point $(x_1^{(e)}, x_2^{(e)}) \in R_2$, then for all $\delta_e < \delta'_e < \delta_{\text{anti}}$ the system in* (10.10) *with p'_e has an eccentric fixed point $(x_1^{(e)'}, x_2^{(e)'}) \in R_2$ which is a stable fixed point.*

We call $\delta_{\text{ecce}} = \min\{\delta_e\}$ which is the smallest δ_e such that the there exists a eccentric fixed point and anti-symmetric saddle points.

Now we are ready to prove Theorem 10.14.

Proof of Theorem 10.14. The main statement of theorem is proved by Lemma 10.16 and 10.17. Now we prove the case 1 and 2 are indeed gradient-like. Because it's only a two dimensional system, by Proposition 2.39, we only need to show 1) the system only have constant hyperbolic fixed points, 2) there is no saddle connections 3) there is no cycle.

For the first case, by Lemma 10.16, the system have constant hyperbolic fixed points and no saddle connections. By symmetric and positive invariant property of $[0, 1]^2$, suppose there is cycle in the system, it should contained in one of the triangles, R_1, R_2, R_3 or R_4. However, it is impossible, since there is no fixed point within those four region.

For the second case, by Lemma 10.16 and 10.17, the system only have 5 fixed points. Secondly, the saddle point have stable manifold in $\{(x_1, x_2) : x_1 + x_2 = 1\}$, so there is no saddle connection. No limit cycle argument is similar to the first case. □

10.6 Stochastic gradient descent and Theorem 10.6

Several machine learning and signal processing applications induce optimization problems with non-convex objective functions. The global optimization of a non-convex objective is an NP-hard problem in general. As a result, a much sought-after goal in applications with non-convex objectives is to find a local minimum of the objective function. One main hurdle in achieving local optimality is the presence of saddle points which can mislead local search method by stalling their process.

Our analysis in Section 10.4 can be applied to these problems. Formally, given an objective function $F : \mathbb{R}^d \to \mathbb{R}$, an popular heuristic to minimize F is by gradient descent method:

$$x_{t+1} = x_t - \eta \nabla F(x_t), \tag{10.12}$$

The gradient descent is well-studied when the objective function is convex: for any constant ϵ, $|F(x_t) - \min_{x \in \mathbb{R}^d} F(x)| \leq \epsilon$ in time $O(1/\eta)$. In this section, we want to study the convergence property when F is non-convex. In particular, we are interested in the time complexity with respect to the step size η.

10.6.1 Bounded stochastic gradient descent algorithm

We now state a general stochastic gradient descent with bounded martingale difference perturbation, and show such processes converge to a local minimal.

Input: An objective function $F : \mathbb{R}^d \to \mathbb{R}$, the step length η, the running time T, and the initial point x_1
Output: A point $x \in \mathbb{R}^d$
for t = 1,2..., T **do**
Sample a perturbation U_{t+1} with properties defined in Theorem 10.18
$x_{t+1} = x_t - \eta \left(\nabla F(x_t) + U_{t+1} \right)$
end for

Program 10.1: Bounded Stochastic Gradient Descent Algorithm

Using the same argument for Theorem 10.6 and Proposition 2.33, we have:

Theorem 10.18 (Bounded Stochastic Gradient Descent Algorithm)**.** *Given a constant d, an objective function $F \in \mathcal{C}^3(\mathbb{R}^d, \mathbb{R}^d)$, a compact set $B \subset \mathbb{R}^d$ which is positive invariant for Equation* (10.12)*, constants $D, d_1, d_2 > 0$, such that*

1. *The objective function F has*

(a) a continuous third derivative.

(b) a constant number of fixed points in B, $\{\beta \in B : \nabla F(\beta) = 0\}$. Moreover, these fixed points are non-degenerate: $\nabla^2 F(\beta)$ is invertible.

2. The perturbation of the process in Algorithm 10.1 satisfies

(a) $X_t \in B$ with probability 1.

(b) $\mathbb{E}[U_{t+1} \mid x_t] = 0$ for all $x_t \in B$,

(c) $\|U_{t+1}\| \leq D$,

(d) For all $x_t \in \cup_{\beta:\nabla F(\beta)=0} B(\beta, \epsilon)$, $d_1 \mathbb{I}_d \prec \mathrm{Cov}[\boldsymbol{U}_{k+1} \mid X_k] \prec d_2 \mathbb{I}_d$

where $B(\beta, r)$ is a ball in $\mathbb{R}^d$, $\{y : \|y - \beta\| \leq r\}$.

Then for any $\epsilon > 0$, there exist $\eta > 0$ and $T = O\left(\frac{\log 1/\eta}{\eta}\right)$ such that for all initial points $x_1 \in B$ $\|x_t - x^*\| \leq \epsilon$ for some local minimal x^* and $t \leq T$ with high probability.

Note that though this theorem only shows the hitting time is $O\left((\log 1/\eta)/\eta\right)$ with high probability, with Lemma F.5 it is not hard to show the process stays in an arbitrary neighborhood of a local minimal after hitting a neighborhood of the local minimal.

10.6.2 Related work

For the time complexity with respect to the step size η, this framework contains several previous results as special cases, and provides a tighter convergence time upper bound. For example, Ge et al. [98] propose the following algorithm:voterbm

Input: An objective function $F : \mathbb{R}^d \to \mathbb{R}$, the step length η, the running time T, and the initial point x_1

Output: A point $x \in \mathbb{R}^d$

for t = 1,2..., T **do**

Sample a perturbation $U_{t+1} \sim S^{d-1}$(Random point on unit sphere)

$x_{t+1} = x_t - \eta\left(\nabla F(x_t) + U_{t+1}\right)$

end for

Program 10.2: Noisy Gradient Descent

They show the convergent time to constant neighborhood of some local minima is $O(1/\eta^2)$ which is weaker than Theorem 10.18 when the objective function satisfies our condition.

Similarly, Jin et al. [121] proposes a perturbed gradient descent algorithm:voterbm

Input: An objective function $F : \mathbb{R}^d \to \mathbb{R}$, the step length η, the running time T, and the initial point x_1

Output: A point $x \in \mathbb{R}^d$

for t = 1,2... **do**
  if $\|\nabla F(x_t)\|$ is small **then**
    $U_{t+1} \sim S^{d-1}$
  else
    $U_{t+1} = 0$
  end if
  $x_{t+1} = x_t - \eta\left(\nabla F(x_t) + U_{t+1}\right)$
end for

Program 10.3: Perturbed Gradient Descent

They show the convergent time to a constant neighborhood of some local minimal is $O((\log 1/\eta)^4/\eta)$ which is weaker than Theorem 10.18 when the objective function satisfies our condition.

Remark 10.19. Here we put some comparison between Theorem 10.18 and previous work.

1. The running time is optimal with respect to step size $1/n$, $O(n \log n)$.

2. This result applies to a larger family of stochastic gradient descent algorithm. Instead of requiring the perturbation to be a uniform point in the unit sphere, our result only requires the noise is bounded and the covariance matrix is positive definite (Theorem 10.18).

3. In gradient flow, the stable and unstable manifold are orthogonal at the saddle point (the Hessian of the function is symmetric), but it is not true for hyperbolic saddle points of non-gradient flow. Our result extends to reinforced random walks with non-gradient flows.

On the other hand, our result doesn't handle some aspects in Ge et al. [98], Jin et al. [121]:

1. We consider the step size η is small enough, but do not provide a closed-form upper bound.

2. We do not work out the dependency of running time on the dimension (and several parameters), although we believe our analysis is dimension free.

3. The number of fixed points in our work is constant, and they avoid this condition by assuming a uniform lower bound of positive eigenvalue of all saddle points which ensures a universal constant improvement after escape from any saddle points.

CHAPTER 11

Engineering Agreement: The Naming Game with Asymmetric and Heterogeneous Agents

A popular topic in language evolution, cognitive science, and culture dynamics, the Naming Game has been widely used to analyze how agents reach global consensus via communications in multi-agent systems. Most prior work considers networks that are symmetric and homogeneous (e.g., vertex transitive). In this paper we consider asymmetric or heterogeneous settings that complement the current literature: 1) we show that increasing asymmetry in network topology can improve convergence rates. Empirically, the star graph converges faster than all previously studied graphs; 2) we consider graph topologies that are particularly challenging for Naming Game such as disjoint cliques and multi-level trees. For these cases, we and ask how much extra homogeneity (random edges) is required to allow convergence or fast convergence.

We provide theoretical analysis which was confirmed by simulations; 3) we analyze how consensus can be manipulated when stubborn nodes are introduced at different points in the process. Early introduction of stubborn nodes can easily influence the outcome in certain families of networks, while stubborn nodes introduced later have much less power.

11.1 Introduction

The analysis of shared conventions in multi-agent systems and complex decentralized social networks has been the focus of study in several diverse fields, such as linguistics, sociology, cognitive science, and computer science. The problem of how such conventions can be established, from among countless options, without a central coordinator has been addressed by several disciplines [182, 42]. Among them, the multi-agent

models and mathematical approaches gain the most attention by accounting for both the network topology and opinion change over time [215, 184, 30, 192, 92]. It has been shown that the emergence of new political, social, economic behaviors, and culture transmission are highly dependent on such convention dynamics [18, 117, 183].

In order to analyze the social dynamics in multi-agent systems in depth, we focus on one stylized model, the Naming Game, in which agents negotiate conventions through local pairwise interactions [214, 26]. The Naming Game captures the generic and essential features of an agreement process in networked agent-based systems. Briefly speaking, when two agents wish to communicate, one agent, the speaker, randomly selects one convention from her list of current conventions and uses this convention to initiate communication with the listener. If the listener recognizes that convention, both the speaker and listener purge their lists of current conventions to only include that "successful" convention. If the listener does not recognize that convention, she adds it to her list of known conventions.

This simple model is able to account for the emergence of shared conventions in a homogeneous population of agents. Both simulations and experiments have been conducted on various network topologies.

However many key questions, especially those related to asymmetric and heterogeneous agents, remain open. For example: what network topologies enable the fastest convergence? Does community structure help or harm convergence? Does homogeneity or heterogeneity help or harm convergence? How robust are the dynamics to possible manipulations by a small number of agents? Moreover, rigorous theoretical analysis is almost entirely absent in previous work on the Naming Game. In this paper we aim to fill in the literature in the following aspects:

1. We discovered that the star graph empirically converges faster than all previously considered graphs for the Naming Game. This network differs from previously analyzed topologies in that it is not symmetric (vertex transitive). In some sense, it is not too surprising that the star graph, an asymmetric graph, works so well to reach consensus, which is a symmetry breaking problem. Though, from first principles, this is far from obvious, and other asymmetric graphs, for example a multi-level tree, perform extremely poorly.

2. To understand network topologies that inhibit fast convergence of the Naming Game, we study two networks with community structures: agents divided into two disconnected communities; and a multi-level tree. For the first network, it is clear that it cannot converge to consensus (it is disconnected). We investi-

gate how much inter community communication needs to be added in order to facilitate convergence. Empirically we observe a sharp threshold on the level of inter community communication: above this threshold, fast convergence is guaranteed, and below it the dynamics fail to converge before time out. We give theoretical justifications for this threshold by showing that convergence takes exponentially long if inter community communication is insufficient (below the threshold). For the second network, the multi-level tree, we observe via simulations that it converges exceedingly slowly—we conjecture that it takes exponential time. For this network, we perform the same simulation tests for adding homogeneity and obtain similar results.

We show that with added communication, the community divisions that thwart consensus can be overcome. Perhaps surprisingly, the amount of intercommunity communication required after disagreement is entrenched, is not substantially more than the amount of communication needed to avoid such division in the first place.

3. Finally, we analyze a third way of introducing asymmetry and heterogeneity: including "stubborn" nodes that do not follow the standard Naming Game protocol. Our experimental results suggest the following hypothesis: in some graphs (e.g. cliques) even a small constant (e.g. 5) number of stubborn nodes can assure convergence to a particular name. However, in others networks (e.g. star graphs, grid graphs, Kleinberg's small world models), the number of nodes required seems to grow with the size of the graph. Additionally, we prove that in a complete graph, manipulation after convergence is much harder than before: there exists a value p such that if an adversary controls more than a p fraction of the nodes, consensus results can be easily manipulated; otherwise it takes exponential time to manipulate the consensus.

   The results on stubborn nodes have implications for the use of the Naming Game in distributed systems. In Steels and McIntyre [216] it was assumed that the protocol would be robust to manipulation. We confirmed this claim if the stubborn nodes appear after the system has converged. But in certain networks these protocols are immensely vulnerable to rogue agents that appear from the start.

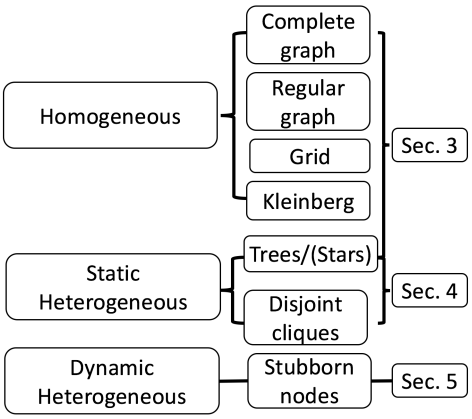


Figure 11.1: Overview of considered graph structures.

11.1.1 Related Work

Baronchelli et al. [28] proposed the Naming Game as a simple multi-agent framework that accounts for the emergence of shared conventions in a structured population. One of the most important problems for Naming Game is to understand how fast the global consensus can be reached and what factors affect it. Some research has been conducted to analyze the effect of network topology on the Naming Game dynamics [67]. Lu et al. [153] show via simulations on real-world graphs that communities show speedy convergence of the dynamics. Centola and Baronchelli [48], using human-subject study, empirically demonstrate the spontaneous creation of universally adopted social conventions and show simple changes in a population's network structure can greatly change the dynamics of norm formation. Baronchelli et al. [29] show that finite connectivity, combined with the small-world property, ensures superior performance in terms of memory usage and convergence rate to that of the grid or complete network. Additionally, a dynamically evolving topology of co-evolution of language and social structure has been studied by Gong et al. [103], for a more complex language game.

One common way to influence the social dynamics and facilitate the converging process towards the consensus is to break the symmetry. Lu et al. [153]Lu et al. have made use of a special kind of agents called "committed" nodes, who will stick to a preferred opinion without deviating, and show that such agents often reduce the time needed to reach consensus. However, in their work they did not evaluate how these nodes might influence which name was converged upon. Additionally, they did not study how the network topology interacted with stubborn nodes or how robust the communication protocol is.

11.2 Preliminary

We present here the version of the *Naming Game* introduced in Baronchelli et al. [26] in which agents negotiate conventions (names), i.e. associations between forms and meaning. The process stops when all agents reach consensus on a single 'name.' The Naming Game is played by agents on a (weighted) graph $G = (V, E, w)$ and proceeds in steps. At each step t, each agent v, is characterized by its inventory (list of names) $A_t(v) \subseteq \mathbb{S}$. At time 0 each agent has an initial inventory $A_0(\cdot)$ which is possibly empty. At each time step $s = 1, 2...$

- An edge is randomly chosen with probability proportional to its weight; and with equal chance one vertex incident to the edge is considered as the speaker and the other as the listener.
- The speaker v selects a word c uniformly at random from its inventory $A_t(v)$ and sends c to the listener u. If the speaker's inventory is empty, the speaker invents a new word c (one that is not in the list of any other agent).
- If the word is in the listener's inventory, $c \in A_s(u)$, the interaction is a "success", and both the speaker and listener remove all words besides c from their inventories.
- If the word is not in the listener's inventory, $c \notin A_s(u)$, the interaction is a "failure" and the listener adds c to its inventory.

The process stops when all the inventories are a singleton of the same name, and we say the process has reached consensus. Notice that the only time a node can have an empty inventory is if it starts that way and has yet to engage in any interaction.

The way in which agents may interact with each other is determined by the topology of the underlying contact network. Here we will introduce the models considered in this paper.

1. *Complete graphs*: all agents are mutual connects.
2. *Regular random graph* $G_{n,k}$ (see Bollobás [39]): every node has degree $k = 8$ and the connection is randomly sample under this constrain.
3. *Kleinberg's small world model* [133]: in standard Kleinberg's model the nodes are on two dimensional grid. Each node u connects to every other node within Manhattan distance p as strong ties, and there are q weak ties which connects

to other nodes v proportional to $d(u,v)^{\alpha}$. In our simulation, the each nodes has 4 strong tie which is $p = 1$, and 4 weak ties with $\alpha = 2$.

4. *Watts-Strogatz's small world model* [234]: the nodes are on one-dimensional ring, and connect to 8 nearest nodes with respect to Manhattan distance, then we rewire the edges of independently with probability 0.5.

5. *Complete bipartite graph* is a bipartite graph such that every pair of graph vertices in the two sets are adjacent. If there are p and q graph vertices in the two sets, the complete bipartite graph is denoted $K_(p,q)$.

6. The *trees* in this paper refer to perfect k-ary trees with height h—that is, a rooted tree with h levels where each node except leaf nodes has exactly k children and the leaf nodes are all at the level h. Note that a *star graph* with n leaves is the complete bipartite graph $K_{1,n}$. Alternatively, a star graph can also be defined as rooted tree of branching factor $n-1$ with depth 1.

11.3 Networks with Fast and Slow Convergence

In this section we study the convergence rate of various graphs. Here we show that a family of asymmetric graphs, the star graphs, empirically converge faster than previously proposed graphs. Next, we point out, perhaps surprisingly, that the convergence time of a multi-level tree is extremely slow. We will engineer and analyze fast converge versions of trees by adding random edges in Section 11.4.1.

We first examine the convergence time for different graphs on a large scale. Here we calculate the time in terms of the number of communication steps denoted as "s". We look at complete graphs, random regular graphs ($G_{n,k}$ graphs), Kleinberg's small world graphs, Watts-Strogatz graphs, as well as star and tree graphs. Unless mentioned otherwise, we will use the same setting defined above in Section 11.2. From Figure 11.2 we can see that the star graph converges the fastest. The tree graph is in fact the slowest. If the tree has two levels with 5000 nodes, after 10^7 steps the nodes still cannot reach consensus. Therefore we did not present the consensus time of the tree in the figure. Among the rest of the graphs, the Kleinberg's small world model is the second slowest, while the other graphs have convergence rate roughly a constant factor of each other.

The network topology's impact on the Naming Game's consensus time is fairly intriguing. To better understand the results, let us consider the best and worst

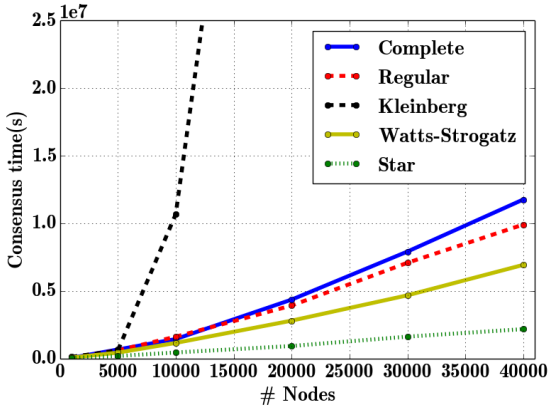


Figure 11.2: Evaluation of the consensus time for different graphs with size growing until 40000.

topology scenarios for multiple agents to reach consensus. The best (quickest) way to reach consensus is to have a specific node to inform all the other nodes of the name. In other words, it is represented by a star graph and the center node is always the speaker. In the naming game framework, even when the speaker/listener role assignment is uniformly random, the star graph is still the fastest in reaching global consensus. This is partly attributed by the asymmetry inherent in the star graph topology.

To analyze the effect of asymmetry, we simulate the graph morphing from a balanced complete bipartite graph to a star by increasing the number of vertices in the larger side of a complete bipartite graph. Figure 11.3 shows the converge time for various complete bipartite graphs. Moving to the right in the figure, the graph becomes more asymmetric and we see that the convergence time decreases. Note that at $m = n$ ($m/n = 1$), this is a balanced bipartite network, and at $m = 2n-1$ ($m/n \approx 2$) this is a star graph. This finding is also aligned with the idea that breaking symmetry can improve consensus efficiency for naming game via "stubborn" agents [153] (and see Section 11.5). On the other hand, the worst graph topology for reaching global consensus is the multi-level tree graph. We hypothesize that this is due to the "community structure" embedded in the tree that converge fast by themselves. In a two-level tree, the subtrees of the main tree are themselves star graphs. Such community structure enables fast "local" convergence of the dynamics within the communities, but face challenges in reaching global convergence — the communities are trying to influence each other but each community has more internal influence than external influence. This phenomena is the topic of the next section, where we give both empirical and rigorous theoretical analysis.

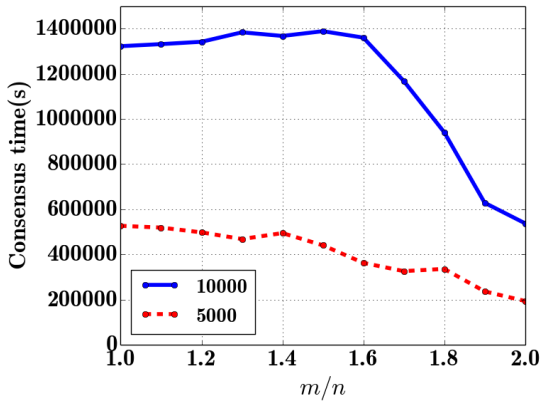


Figure 11.3: Evaluation for converging time for various complete bipartite graphs $K_{m,2n-m}$ where m is the cardinality of the larger partition of vertices.

11.4 Effects of Community Structure

In this section we study the effects of community structure using two network models, one of them is a dense graph and the other one a sparse graph. The first is a graph of heterogeneous agents divided into two disconnected communities. The simplicity of this model permits theoretical analysis of precisely how and when community structure can exhibit convergence. The second is a multi-level tree introduced in the previous section.

Given a weighted graph G where the sum of the weights is W we construct $\mathrm{Hom}(G, p)$ by adding $\frac{p}{1-p}\frac{W}{\binom{n}{2}}$ mass to each edge (creating a new edge if it does not exist). This effectively samples the complete graph with probability p and the graph G with probability $1 - p$.

For each network, we first examine the convergence rate of $\mathrm{Hom}(\cdot, p)$ using simulations. We show that adding a sufficient amount of homogeneity overcomes the heterogeneity. For the first network, we will provide a theoretical analysis which predicts, supports, and explains the empirical results.

11.4.1 Disjoint Cliques

Naturally, a graph G of $2n$ heterogeneous agents divided into two equally sized disconnected communities will not converge to consensus. As p increases from 0 toward 1 $\mathrm{Hom}(G, p)$ becomes a network of increasingly interconnected communities.

Additionally, the behavior of the Naming Game depends on the initial states, i.e., the collection of names at these nodes at the beginning. We consider two situa-

tions for the initial states. 1) "Empty" start, where all nodes start with empty lists $\forall v \in V, A_0(v) = \phi$. 2) "Segregated" start, in which the two groups have different initial opinions, $\forall v \in V_1, A_0(v) = \{0\}$ and $\forall v \in V_2, A_0(v) = \{1\}$. Clearly it is more challenging for the Naming Game to reach global convergence under the segregated initial state.

Simulation Results. Figure 11.4 (row 1 (a)) shows the convergence time for different values of p under different initial scenarios on graphs of size n. For each setting we run the simulation multiple times and plot the time to reach consensus for each run as a dot in the figure. In certain situations it is hard to reach consensus even after a long time. Therefore we set 10^7 as the time-out criteria – i.e., if no consensus is reached after 10^7 rounds and we stop the simulation. From Figure 11.4 (row 1 (a)) we can see that when p is smaller it is harder to reach consensus for all situations. When p is sufficiently small all situations may hit the timeout condition before consensus is reached. In addition, the threshold of p which allows this happen is larger for the "segregated" initial setup compared to the empty initial setup. Similarly, for graphs of larger size it is easier to hit the time out condition. When $p > 0.2$ the time to reach consensus for all situations is small so we chose not to plot it.

To further analyze the naming game behavior when p is in between $[0, 0.25]$, we show in Figure 11.4 (row 2 (a)) the fraction of trials failing to reach consensus (before timing out) with different values of p. It is clear that for the empty start initial condition, the game will time out at about $p = 0.24$, while for the segregated start case, the game will time out when p is around 0.26. This threshold value changes with the size of the local community.

Curiously, for the "empty" start, graphs with smaller sizes are more likely to encounter timeouts than their larger counterparts. This may be because the smaller size of each community results in a greater chance of quickly reaching local consensus, which resembles the segregated start scenario. Therefore, it takes longer for graphs with smaller sizes to break the local consensus and escape the so called "stuck" situation.

However, for the segregated start, it immediately starts with the "worst" case setting where the two communities have diverging opinions, so overall it takes longer to leave "stuck" situation compared with graphs of the same size in the "empty" start scenario. Additionally, graphs with larger sizes in the segregated setting more easily encounter a timeout. This may be because larger graphs occasionally time out even if they are not really "stuck" because they take longer to reach consensus in any event.

Theoretical Analysis. Next we will analyze the consensus time for the naming game

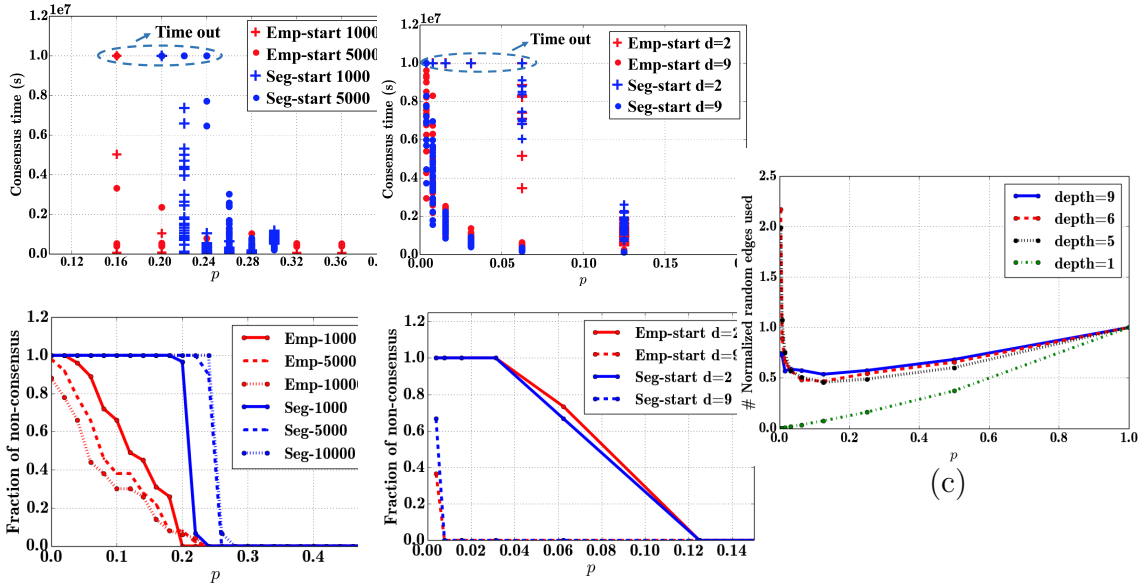


Figure 11.4: Evaluations of reaching global consensus for different initialization scenarios and sizes of graphs. Row 1: Consensus time for (a) disjoint cliques (b) tree structure; row 2: fraction of nodes failing to reach consensus, based on different probability of random edges p for (a) disjoint cliques (b) tree structure; (c) normalized number of random edges used for communication as a function of the probability of random edges p.

on $Hom(G, p)$ where G has $2n$ agents divided into two equally sized disconnected communities with segregated start.

Theorem 11.1. *Let G be the disjoint union of two n cliques, each of size n. Then for the segregated start naming game on $Hom(G, p)$, there exists a constant $p_0 \approx 0.110$ such that if $0 \leq p < p_0$ the expected consensus time is $\exp(\Omega(n))$.*

Here we sketch a proof of theorem. A full proof appears in the appendix.

To prove this theorem, we formulate the naming game as a nonhomogenous random walk on $\mathbb{Z}^4$ and relate this nonhomogenous random walk to a corresponding autonomous system in $\mathbb{R}^4$.

In the segregated start scenario, every node has an initial opinion, therefore no new name will be generated, and nodes inventory will be either $\{0\}, \{1\}$, or $\{0, 1\}$. Due to the symmetry among nodes, at each step t we only need to keep track of the number of nodes in the two groups whose inventory corresponds to the three categories above. Moreover, because the total number of two communities are n, we can use four variables to discribes this random process: fraction of $\{0\}, \{1\}$ nodes in two communities.

As the size of community increase, the above process is closely related to its mean field which can be seen as a autonomous system in $\mathbb{R}^4$. We show that this system has a stable fixed point as long as $0 \leq p < p_0$. To proof Theorem 11.1 we show two things with the help in autonomous system:

1. Global behaviour: the random process $X(t)$ will initially "converge" to a point corresponding to the stable fixed point of the autonomous system.

2. Local behaviour: the random process $X(t)$ takes exponential time to leave the regions corresponding to the regions around stable fixed point of the autonomous system.

11.4.2 Tree Structure

In this section, we systematically study the Naming Game on trees and examine how the naming game converges when applying $\mathrm{Hom}(\cdot, p)$ to the tree structure. We show that convergence is substantially sped up for random edges added with small probabilities. In Figure 11.4 (row 1 (b)) we evaluate the time to reach consensus based on the probability p of choosing random edges. It is clear that for trees with smaller depth (d) and more branches, the time to reach consensus is larger. Compared with

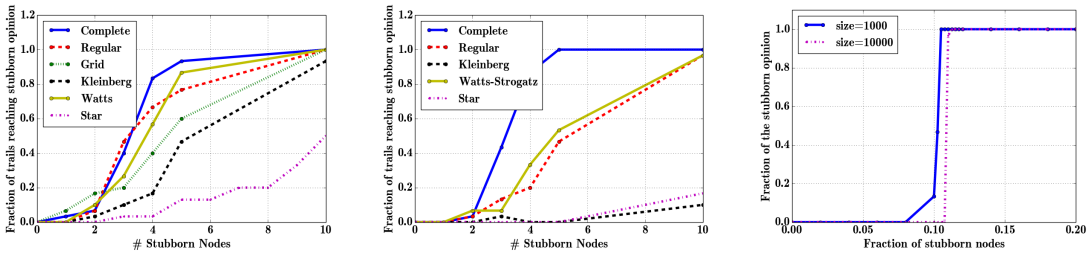


Figure 11.5: Evaluation for early stage coordinated stubborn nodes within different graphs. (a) fraction of trials converging to the stubborn nodes' opinion, as a function of the number of stubborn nodes with size 1000; (b) fraction of trials converging to the stubborn nodes' opinion, as a function of the number of stubborn nodes with size 10000; (c) fraction of nodes converging to the stubborn nodes' opinion in the late stage situation for complete graph of size 1000 and 10000.

Figure 11.4 (row 1 (a)) we see that by adding random edges, the tree graph is much less likely to encounter a timeout than the densely connected community graph. Figure 11.4 (row 2 (b)) show the fraction of agents failing to reach consensus as a function of p. Additionally, though the additional pairs can break up the sparse community structure and help to accelerate the converging process, redundant communications may be introduced at the same time. Therefore in Figure 11.4 (c) for various p, we present the total number of time the dynamics choose a homogeneous edge before consensus is reached, normalized by time it takes homogeneous graph (clique) to reach consensus. We can see that there is actually an tipping point where the homogeneous edges are used the least, which implies the edges of the original tree actually help towards consensus. Above this point, the homogeneous edges provide unnecessary communication redundancy.

11.5 Stubborn Nodes

In this section we introduce another aspect of asymmetry and heterogeneity. We introduce special agents called "stubborn" nodes, which never change their own opinions and aim to influence the whole network. The topic is also related to the robustness of the naming game in the real world setting, in which a small number of nodes can be malicious and not follow the protocol. The primary question we want to ask is: how and when can such nodes affect the opinion/name to which the dynamics converge? There are two important factors to consider here – the network topology and the

time when the stubborn nodes are activated. Here we consider two situations: 1) the stubborn nodes join at the beginning of the game; 2) the nodes become "stubborn" after the graph has converged to one global opinion. Figure 11.5 (a) (b) show the fraction of trials converging to the stubborn nodes' preference based on the number of stubborn nodes in situation 1) for graphs of size 1000 and 10000, respectively. From Figure 11.5 (a), it is clear that in some graphs (e.g. the clique) even a small constant (e.g. 5) number of stubborn nodes can guarantee convergence to a particular name. Note that as the number of nodes increase, the curve barely changes, and if anything, becomes a sharper threshold. However, in others networks (e.g. star, grid, Kleinberg model), the number of required nodes seems to grow with the size of the graph.

This shows that in certain networks these protocols are not robust to rogue/stubborn agents. By comparing Figure 11.5 (a) and (b)), we see that the complete graph is not affected much by its size in terms of the influence efficiency of the "stubborn" nodes. However, in the Kleinberg and star graphs the number of stubborn nodes needed greatly depends on the size of the network. Note that here we choose the same number of stubborn nodes because complete, regular and Watts-Strogatz graphs actually perform similarly with size 1000 on these number of stubborn nodes.

Additionally, we show that in the complete graph, manipulating the name after convergence is much harder than before: there exists a value $p_0 \in [0, 1]$, such that if an adversary controls more than p_0 fraction of the node, consensus can be easily manipulated and otherwise it will take exponential time to manipulate the consensus. In Figure 11.5 (c) we verify this empirically by showing the fraction of trials converging to the stubborn nodes' preference (before timing out) based on the fraction of "stubborn" nodes within the network. It shows that at least 10% such stubborn nodes are needed to manipulate the opinion of the original graph empirically.

We provide theoretical analysis on the lower bound for the number of "stubborn" nodes required to manipulate the global consensus to align with the "stubborn" nodes' in a complete graph. In completed graph if the naming game converges to opinion 1, we want to answer the following question: what fraction of stubborn nodes with opinion 0 are required in order to convert the graph's consensus to 0 in polynomial time?

Theorem 11.2. *Given the naming game with p fraction of stubborn nodes defined above, there exists a constant $p_0 \approx 0.108$ such that for all $0 \leq p < p_0$ the expected consensus time is $\exp(\Omega(n))$. Additionally, if $p_0 < p \leq 1$ for all $\epsilon > 0$ the fraction of original opinion is smaller than ε after $O(n)$ steps.*

The proof appears in the appendix.

BIBLIOGRAPHY

[1] M. A. Abdullah and M. Draief. Global majority consensus by local majority polling on graphs of a given degree sequence. *Discrete Applied Mathematics*, 180:1–10, 2015.

[2] I. Abraham, S. Chechik, D. Kempe, and A. Slivkins. Low-distortion inference of latent similarities from a multiplex social network. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '13, pages 1853–1883. SIAM, 2013. ISBN 978-1-611972-51-1. URL http://dl.acm.org/citation.cfm?id=2627817.2627949.

[3] L. Adamic and E. Adar. How to search a social network. *Social Networks*, 27: 187–203, 2005.

[4] L. A. Adamic and N. Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM, 2005.

[5] E. Akin. *The general topology of dynamical systems*, volume 1. American Mathematical Soc., 2010.

[6] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[7] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, Jan 2002. doi: 10.1103/RevModPhys.74.47. URL http://link.aps.org/doi/10.1103/RevModPhys.74.47.

[8] R. Albert, H. Jeong, and A.-L. Barabási. Internet: Diameter of the world-wide web. *nature*, 401(6749):130, 1999.

[9] D. Aldous et al. Interacting particle systems as stochastic social dynamics. *Bernoulli*, 19(4):1122–1149, 2013.

[10] L. Alvisi, A. Clement, A. Epasto, S. Lattanzi, and A. Panconesi. Sok: The evolution of sybil defense via social networks. In *Security and Privacy (SP), 2013 IEEE Symposium on*, pages 382–396. IEEE, 2013.

[11] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the national academy of sciences*, 97(21): 11149–11152, 2000.

[12] H. Amini. Bootstrap percolation and diffusion in random graphs with given vertex degrees. *Electronic Journal of Combinatorics*, 17(1):1–20, 2010.

[13] H. Amini and N. Fountoulakis. What I tell you three times is true: bootstrap percolation in small worlds. In *Proceedings of the 8th international conference on Internet and Network Economics*, WINE'12, pages 462–474, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-35310-9. doi: 10.1007/978-3-642-35311-6_34. URL http://dx.doi.org/10.1007/978-3-642-35311-6_34.

[14] R. Angell and G. Schoenebeck. Don't be greedy: leveraging community structure to find high quality seed sets for influence maximization. In *International Conference on Web and Internet Economics*, pages 16–29. Springer, 2017.

[15] J. Antonovics and P. Kareiva. Frequency-dependent selection and competition: empirical approaches. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 319(1196):601–613, 1988.

[16] H. Antosiewicz. Ordinary differential equations (g. birkhoff and gc rota). *SIAM Review*, 5(2):160, 1963.

[17] R. Arratia, L. Goldstein, and L. Gordon. Poisson approximation and the chenstein method. *Statistical Science*, pages 403–424, 1990.

[18] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: Membership, growth, and evolution. In *Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 44–54, 2006.

[19] L. Backstrom, E. Sun, and C. Marlow. Find me if you can: Improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 61–70, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-799-8. doi: 10.1145/1772690.1772698. URL http://doi.acm.org/10.1145/1772690.1772698.

[20] E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132, 2015.

[21] E. Balkanski, N. Immorlica, and Y. Singer. The importance of communities for learning to influence. In *Advances in Neural Information Processing Systems*, pages 5862–5871, 2017.

[22] J. Balogh and B. Pittel. Bootstrap percolation on the random regular graph. *Random Struct. Algorithms*, 30:257–286, 2007.

[23] A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson. The diffusion of microfinance. *Science*, 341(6144):1236498, 2013.

[24] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[25] R. Bardenet, O.-A. Maillard, et al. Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385, 2015.

[26] A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, and L. Steels. Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(06):P06014, 2006.

[27] A. Baronchelli, V. Loreto, L. Dall'Asta, and A. Barrat. Strategy, topology, and all that. In *The Evolution of Language: Proceedings of the 6th International Conference (EVOLANG6), Rome, Italy, 12-15 April 2006*, page 11. World Scientific, 2006.

[28] A. Baronchelli, V. Loreto, L. DallAsta, and A. Barrat. Bootstrapping communication in language games: Strategy, topology and all that. In *Proceedings of the 6th International Conference on the Evolution of Language*, volume 2006, pages 11–18. World Scientific Press, 2006.

[29] A. Baronchelli, L. Dall'Asta, A. Barrat, and V. Loreto. The role of topology on the dynamics of the naming game. *The European Physical Journal Special Topics*, 143(1):233–235, 2007.

[30] A. Baronchelli, V. Loreto, and L. Steels. In-depth analysis of the naming game dynamics: the homogeneous mixing case. *International Journal of Modern Physics C*, 19(05):785–812, 2008.

[31] L. Becchetti, A. Clementi, E. Natale, F. Pasquale, and L. Trevisan. Stabilizing consensus with many opinions. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 620–635. Society for Industrial and Applied Mathematics, 2016.

[32] E. Ben-Naim, L. Frachebourg, and P. L. Krapivsky. Coarsening and persistence in the voter model. *Physical Review E*, 53(4):3078, 1996.

[33] F. Bénézit, P. Thiran, and M. Vetterli. Interval consensus: from quantized gossip to voting. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 3661–3664. IEEE, 2009.

[34] I. Benjamini, S.-O. Chan, R. ODonnell, O. Tamuz, and L.-Y. Tan. Convergence, unanimity and disagreement in majority dynamics on unimodular graphs and random graphs. *Stochastic Processes and their Applications*, 126(9):2719–2733, 2016.

[35] B. Bishop. *The Big Sort: Why the Clustering of Like-minded America is Tearing Us Apart*. Houghton Mifflin Harcourt, 2009.

[36] B. Bishop. *The big sort: Why the clustering of like-minded America is tearing us apart*. Houghton Mifflin Harcourt, 2009.

[37] L. Bode. Pruning the news feed: Unfriending and unfollowing political content on social media. *Research & Politics*, 3(3), 2016.

[38] M. Boguna, D. Krioukov, and K. C. Claffy. Navigability of complex networks. *Nature Physics*, 5(1):74, 2009.

[39] B. Bollobás. Random graphs. In *Modern Graph Theory*, pages 215–252. Springer, 1998.

[40] B. Bollobás and F. R. K. Chung. The diameter of a cycle plus a random matching. *SIAM J. Discrete Math.*, 1(3):328–333, 1988. doi: 10.1137/0401033. URL https://doi.org/10.1137/0401033.

[41] B. Bollobás and B. D. McKay. The number of matchings in random regular graphs and bipartite graphs. *Journal of Combinatorial Theory, Series B*, 41(1): 80–91, 1986.

[42] H. Brighton and S. Kirby. The survival of the smallest: Stability conditions for the cultural evolution of compositional language. In *European Conference on Artificial Life*, pages 592–601. Springer, 2001.

[43] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer networks*, 33(1-6):309–320, 2000.

[44] C. T. Butts. *Predictability of large-scale spatially embedded networks*. na, 2003.

[45] Z. Cai and C. Jermaine. The latent community model for detecting sybil attacks in social networks. In *VLDB*, 2011.

[46] C. Castellano, D. Vilone, and A. Vespignani. Incomplete ordering of the voter model on small-world networks. *EPL (Europhysics Letters)*, 63(1):153, 2003.

[47] C. Castellano, V. Loreto, A. Barrat, F. Cecconi, and D. Parisi. Comparison of voter and glauber ordering dynamics on networks. *Physical review E*, 71(6): 066107, 2005.

[48] D. Centola and A. Baronchelli. The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences*, 112(7):1989–1994, 2015.

[49] D. Centola and M. Macy. Complex contagions and the weakness of long ties. *American journal of Sociology*, 113(3):702–734, 2007.

[50] J. Chalupa, P. L. Leath, and G. R. Reich. Bootstrap percolation on a bethe lattice. *Journal of Physics C: Solid State Physics*, 12(1):L31, 1979.

[51] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *ACM SIGKDD*, pages 199–208. ACM, 2009.

[52] W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 88–97. IEEE, 2010.

[53] W. Chen, T. Lin, Z. Tan, M. Zhao, and X. Zhou. Robust influence maximization. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 795–804. ACM, 2016.

[54] F. Chung and R. Graham. Quasi-random graphs with given degree sequences. *Random Structures & Algorithms*, 32(1):1–19, 2008.

[55] F. Chung and M. Radcliffe. On the spectra of general random graphs. *the electronic journal of combinatorics*, 18(1):P215, 2011.

[56] F. R. Chung and F. C. Graham. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.

[57] A. Clauset, C. Moore, and M. E. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98, 2008.

[58] E. Cohen, D. Delling, T. Pajor, and R. F. Werneck. Sketch-based influence maximization and computation: Scaling up with guarantees. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 629–638. ACM, 2014.

[59] R. Cohen, A. Tsang, K. Vaidyanathan, and H. Zhang. Analyzing opinion dynamics in online social networks. *Big Data & Information Analytics*, 1(4): 279–298, 2016.

[60] J. Coleman, E. Katz, and H. Menzel. The diffusion of an innovation among physicians. *Sociometry*, 20:253–270, 1957.

[61] J. S. Coleman, E. Katz, and H. Menzel. *Medical Innovation: A Diffusion Study*. Bobbs-Merrill Co, 1966.

[62] C. C. Conley. *Isolated invariant sets and the Morse index*. Number 38. American Mathematical Soc., 1978.

[63] C. Cooper, R. Elsässer, and T. Radzik. The power of two choices in distributed voting. In *International Colloquium on Automata, Languages, and Programming*, pages 435–446. Springer, 2014.

[64] C. Cooper, T. Radzik, N. Rivera, and T. Shiraga. Fast plurality consensus in regular expanders. *arXiv preprint arXiv:1605.08403*, 2016.

[65] J. T. Cox and D. Griffeath. Diffusive clustering in the two dimensional voter model. *The Annals of Probability*, pages 347–370, 1986.

[66] J. Cruise and A. Ganesh. Probabilistic consensus via polling and majority rules. *Queueing Systems*, 78(2):99–120, 2014.

[67] L. Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto. Agreement dynamics on small-world networks. *EPL (Europhysics Letters)*, 73(6):969, 2006.

[68] G. Danezis and P. Mittal. Sybilinfer: Detecting sybil nodes using social networks. In *NDSS*. San Diego, CA, 2009.

[69] W. De Koster and D. Houtman. 'stormfront is like a second home to me' on virtual community formation by right-wing extremists. *Information, Communication & Society*, 11(8):1155–1176, 2008.

[70] I. de Sola Pool and M. Kochen. Contacts and influence. *Social networks*, 1(1): 5–51, 1978.

[71] G. Deffuant, F. Amblard, G. Weisbuch, and T. Faure. How can extremism prevail? a study based on the relative agreement interaction model. *Journal of artificial societies and social simulation*, 5(4), 2002.

[72] M. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, pages 118–121, 1974.

[73] P. DiMaggio. Structural analysis of organizational fields: A blockmodel approach. *Research in organizational behavior*, 1986.

[74] P. S. Dodds, R. Muhamad, and D. J. Watts. An experimental study of search in global social networks. *science*, 301(5634):827–829, 2003.

[75] B. Doerr, L. A. Goldberg, L. Minder, T. Sauerwald, and C. Scheideler. Stabilizing consensus with the power of two choices. In *Proceedings of the twenty-third annual ACM symposium on Parallelism in algorithms and architectures*, pages 149–158. ACM, 2011.

[76] J. R. Douceur. The sybil attack. In *Peer-to-peer Systems*, pages 251–260. Springer, 2002.

[77] R. M. D'souza, C. Borgs, J. T. Chayes, N. Berger, and R. D. Kleinberg. Emergence of tempered preferential attachment from optimization. *Proceedings of the National Academy of Sciences*, 104(15):6112–6117, 2007.

[78] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[79] R. Durrett, J. P. Gleeson, A. L. Lloyd, P. J. Mucha, F. Shi, D. Sivakoff, J. E. Socolar, and C. Varghese. Graph fission in an evolving voter model. *Proceedings of the National Academy of Sciences*, 2012.

[80] R. Durrett, J. P. Gleeson, A. L. Lloyd, P. J. Mucha, F. Shi, D. Sivakoff, J. E. S. Socolar, and C. Varghese. Graph fission in an evolving voter model. *PNAS*, 109(10):3682–3687, 6 Mar. 2012.

[81] A. Eberle. Markov processes. *Lecture Notes at University of Bonn*, 2009.

[82] R. Ebrahimi, J. Gao, G. Ghasemiesfeh, and G. Schoenebeck. How complex contagions spread quickly in the preferential attachment model and other time-evolving networks. *arXiv preprint arXiv:1404.2668*, 2014.

[83] R. Ebrahimi, J. Gao, G. Ghasemiesfeh, and G. Schoenebeck. Complex contagions in kleinberg's small world model. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*, pages 63–72, 2015. doi: 10.1145/2688073.2688110. URL https://doi.org/10.1145/2688073.2688110.

[84] R. Ebrahimi, J. Gao, G. Ghasemiesfeh, and G. Schoenbeck. How complex contagions spread quickly in preferential attachment models and other time-evolving networks. *IEEE Transactions on Network Science and Engineering*, 4 (4):201–214, 2017.

[85] G. Ellison. Learning, local interaction, and coordination. *Econometrica: Journal of the Econometric Society*, pages 1047–1071, 1993.

[86] G. Ellison. Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *The Review of Economic Studies*, 67(1):17–45, 2000.

[87] P. Erdös and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61, 1960.

[88] J. W. Essam. Percolation theory. *Reports on Progress in Physics*, 43(7):833, 1980.

[89] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM computer communication review*, volume 29, pages 251–262. ACM, 1999.

[90] X. Fan, I. Grama, and Q. Liu. Hoeffdings inequality for supermartingales. *Stochastic Processes and their Applications*, 122(10):3545–3559, 2012.

[91] P. Fraigniaud, E. Lebhar, and Z. Lotker. Recovering the long-range links in augmented graphs. *Theor. Comput. Sci.*, 411(14-15):1613–1625, Mar. 2010. ISSN 0304-3975. doi: 10.1016/j.tcs.2010.01.009. URL http://dx.doi.org/10.1016/j.tcs.2010.01.009.

[92] H. Franks, N. Griffiths, and A. Jhumka. Manipulating convention emergence using influencer agents. *Autonomous Agents and Multi-Agent Systems*, 26(3): 315–353, 2013.

[93] D. A. Freedman. On tail probabilities for martingales. *the Annals of Probability*, pages 100–118, 1975.

[94] J. F. Gainsborough. Voters in context: Cities, suburbs, and presidential vote. *American Politics Research*, 33(3):435–461, 2005.

[95] J. Gao, G. Ghasemiesfeh, G. Schoenebeck, and F. Yu. General threshold model for social cascades: Analysis and simulations. In *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, pages 617–634, 2016. doi: 10.1145/2940716.2940778. URL https://doi.org/10.1145/2940716.2940778.

[96] J. Gao, B. Li, G. Schoenebeck, and F. Yu. Engineering agreement: The naming game with asymmetric and heterogeneous agents. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 537–543, 2017. URL http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14986.

[97] J. Gao, G. Schoenebeck, and F.-Y. Yu. Cascades and myopic routing in nonhomogeneous kleinbergs small world model. In *International Conference on Web and Internet Economics*, pages 383–394. Springer, 2017.

[98] R. Ge, F. Huang, C. Jin, and Y. Yuan. Escaping from saddle pointsonline stochastic gradient for tensor decomposition. In *Conference on Learning Theory*, pages 797–842, 2015.

[99] G. Ghasemiesfeh, R. Ebrahimi, and J. Gao. Complex contagion and the weakness of long ties in social networks: revisited. In *Proceedings of the fourteenth ACM conference on Electronic Commerce*, pages 507–524. ACM, 2013.

[100] R. Gheissari and A. B. Hamou. Aimpl: Markov chain mixing times, available at http://aimpl.org/markovmixing.

[101] M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[102] J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*, 9(3):1–18, 2001.

[103] T. Gong, J. Ke, J. W. Minett, and W. S. Wang. A computational framework to simulate the coevolution of language and social structure. In *Artificial Life IX: Proceedings of the 9th International Conference on the Simulation and Synthesis of Living Systems*, pages 158–64, 2004.

[104] M. Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.

[105] M. S. Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.

[106] M. S. Granovetter. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1(1983):201–233, 1983. URL http://www.jstor.org/stable/202051?origin=crossref.

[107] G. Grimmett and D. Stirzaker. *Probability and random processes*. Oxford university press, 2001.

[108] K. M. Hall. An r-dimensional quadratic placement algorithm. *Management science*, 17(3):219–229, 1970.

[109] M. S. Handcock, A. E. Raftery, and J. M. Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354, 2007.

[110] T. P. Hayes. A large-deviation inequality for vector-valued martingales. *Combinatorics, Probability and Computing*, 2005.

[111] R. Hegselmann, U. Krause, et al. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of artificial societies and social simulation*, 5(3), 2002.

[112] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97 (460):1090–1098, 2002.

[113] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[114] R. A. Holley and T. M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, pages 643–663, 1975.

[115] P. Holme and M. E. J. Newman. Nonequilibrium phase transition in the coevolution of networks and opinions. *Phys. Rev. E*, 74(5 Pt 2), 2006.

[116] R. A. Horn, R. A. Horn, and C. R. Johnson. *Matrix analysis*. Cambridge university press, 1990.

[117] J. R. Hurford. Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*, 77(2):187–222, 1989.

[118] M. O. Jackson. *Social and economic networks*. Princeton university press, 2010.

[119] S. Janson, T. Łuczak, T. Turova, T. Vallier, et al. Bootstrap percolation on the random graph $g_{-}\{n, p\}$. *The Annals of Applied Probability*, 22(5):1989–2047, 2012.

[120] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41, 2001.

[121] C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan. How to escape saddle points efficiently. *arXiv preprint arXiv:1703.00887*, 2017.

[122] N. A. John and S. Dvir-Gvirsman. "I don't like you any more": Facebook unfriending by Israelis during the Israel–Gaza conflict of 2014. *J. Commun.*, 65 (6):953–974, 1 Dec. 2015.

[123] J. J. Jordan, D. G. Rand, S. Arbesman, J. H. Fowler, and N. A. Christakis. Contagion of cooperation in static and fluid social networks. *PLoS One*, 8(6), 2013.

[124] O. Kallenberg and R. Sztencel. Some dimension-free features of vector-valued martingales. *Probability Theory and Related Fields*, 88(2):215–247, 1991.

[125] Y. Kanoria, A. Montanari, et al. Majority dynamics on trees and the dynamic cavity method. *The Annals of Applied Probability*, 21(5):1694–1748, 2011.

[126] M. Kearns and J. Tan. Biased voting and the democratic primary problem. In *International Workshop on Internet and Network Economics*, pages 639–652. Springer, 2008.

[127] D. Kempe, J. Kleinberg, and E. Tardos. Influential nodes in a diffusion model for social networks. In *Proceedings of the 32nd international conference on Automata, Languages and Programming*, pages 1127–1138, 2005.

[128] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. *Theory of Computing*, 11:105–147, 2015. doi: 10.4086/toc.2015.v011a004. URL https://doi.org/10.4086/toc.2015.v011a004.

[129] D. Kempe, J. Kleinberg, S. Oren, and A. Slivkins. Selection and influence in cultural dynamics. *Network Science*, 4(1):1–27, 2016.

[130] A.-M. Kermarrec, V. Leroy, and G. Trédan. Distributed social graph embedding. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1209–1214. ACM, 2011.

[131] M. Kimura and G. H. Weiss. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49(4):561, 1964.

[132] F. Kivran-Swaine, P. Govindan, and M. Naaman. The impact of network structure on breaking ties in online social networks: unfollowing on twitter. In *SIGCHI*, 2011.

[133] J. Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 163–170. ACM, 2000.

[134] J. M. Kleinberg. Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001, December 3-8, 2001, Vancouver, British Columbia, Canada]*, pages 431–438, 2001. URL http://papers.nips.cc/paper/2061-small-world-phenomena-and-the-dynamics-of-information.

[135] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins. The web as a graph: measurements, models, and methods. In *Computing and combinatorics*, pages 1–17. Springer, 1999.

[136] P. L. Krapivsky and S. Redner. Dynamics of majority rule in two-state interacting spin systems. *Physical Review Letters*, 90(23):238701, 2003.

[137] P. N. Krivitsky, M. S. Handcock, A. E. Raftery, and P. D. Hoff. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social networks*, 31(3):204–213, 2009.

[138] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Extracting large-scale knowledge bases from the web. In *VLDB*, volume 99, pages 639–650. Citeseer, 1999.

[139] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 57–, 2000. ISBN 0-7695-0850-2.

[140] R. Kumar, D. Liben-Nowell, and A. Tomkins. Navigating low-dimensional and hierarchical population networks. In *Algorithms - ESA 2006, 14th Annual European Symposium, Zurich, Switzerland, September 11-13, 2006, Proceedings*, pages 480–491, 2006. doi: 10.1007/11841036_44. URL https://doi.org/10.1007/11841036_44.

[141] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1(1), 2007.

[142] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, 2008. doi: 10.1080/15427951.2009.10129177. URL https://doi.org/10.1080/15427951.2009.10129177.

[143] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg. Signed networks in social media. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Atlanta, Georgia, USA, April 10-15, 2010*, pages 1361–1370, 2010. doi: 10.1145/1753326.1753532. URL https://doi.org/10.1145/1753326.1753532.

[144] D. A. Levin and Y. Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[145] D. Z. Levin and R. Cross. The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management science*, 50 (11):1477–1490, 2004.

[146] K. Lewis, M. Gonzalez, and J. Kaufman. Social selection and peer influence in an online social network. *Proc. Natl. Acad. Sci. U. S. A.*, 109(1):68–72, 3 Jan. 2012.

[147] Q. Li, W. Chen, X. Sun, and J. Zhang. Influence maximization with ε-almost submodular threshold functions. In *NIPS*, pages 3804–3814, 2017.

[148] D. Liben-Nowell and J. M. Kleinberg. The link-prediction problem for social networks. *JASIST*, 58(7):1019–1031, 2007. doi: 10.1002/asi.20591. URL https://doi.org/10.1002/asi.20591.

[149] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33):11623–11628, 2005.

[150] T. M. Liggett. Coexistence in threshold voter models. *The Annals of Probability*, pages 764–802, 1994.

[151] T. M. Liggett et al. Stochastic models of interacting systems. *The Annals of Probability*, 25(1):1–29, 1997.

[152] B. Light and E. Cassidy. Strategies for the suspension and prevention of connection: Rendering disconnection as socioeconomic lubricant with Facebook. *New Media & Society*, 16(7):1169–1184, 24 July 2014.

[153] Q. Lu, G. Korniss, and B. K. Szymanski. The naming game in social networks: community formation and consensus engineering. *Journal of Economic Interaction and Coordination*, 4(2):221–235, 2009.

[154] B. Lucier, J. Oren, and Y. Singer. Influence at scale: Distributed computation of complex contagion in networks. In *ACM SIGKDD*, pages 735–744. ACM, 2015.

[155] S. Lv, X. Wang, X. Zhao, and X. Zhou. Detecting the sybil attack cooperatively in wireless sensor networks. In *Computational Intelligence and Security, 2008. CIS'08. International Conference on*, volume 1, pages 442–446. IEEE, 2008.

[156] J. S. MacDonald and L. D. MacDonald. Chain migration ethnic neighborhood formation and social networks. *The Milbank Memorial Fund Quarterly*, 42(1): 82–97, 1964.

[157] J. J. McAuley and J. Leskovec. Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States.*, pages 548–556, 2012. URL http://papers.nips.cc/paper/4532-learning-to-discover-social-circles-in-ego-networks.

[158] J. J. McAuley and J. Leskovec. Learning to discover social circles in ego networks. In *NIPS*, volume 2012, pages 548–56, 2012.

[159] D. D. McFarland and D. J. Brown. Social distance as a metric: a systematic introduction to smallest space analysis. *EO Laumann. Bonds of Pluralism: The Form and Substance of Urban Social Networks. New York: John Wiley*, pages 213–252, 1973.

[160] S. C. Mckee. Rural voters in presidential elections. In *Forum-A Journal of Applied Research in Contemporary Politics*, volume 5, 2007.

[161] R. Mermelstein, S. Cohen, E. Lichtenstein, J. S. Baer, and T. Kamarck. Social support and smoking cessation and maintenance. *Journal of consulting and clinical psychology*, 54(4):447, 1986.

[162] S. Messing and S. J. Westwood. Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communic. Res.*, 41(8):1042–1063, 31 Dec. 2012.

[163] K. R. Meyer. Energy functions for morse smale systems. *American Journal of Mathematics*, 90(4):1031–1040, 1968.

[164] S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[165] R. Milo, N. Kashtan, S. Itzkovitz, M. E. Newman, and U. Alon. On the uniform generation of random graphs with prescribed degree sequences. *arXiv preprint cond-mat/0312028*, 2003.

[166] M. Mitzenmacher and E. Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge university press, 2005.

[167] B. Mobasher, R. Burke, and J. J. Sandvig. Model-based collaborative filtering as a defense against profile injection attacks. In *AAAI*, volume 6, page 1388, 2006.

[168] D. Mok, B. Wellman, et al. Did distance matter before the internet?: Interpersonal contact and support in the 1970s. *Social networks*, 29(3):430–461, 2007.

[169] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180, 1995.

[170] J. Molofsky, R. Durrett, J. Dushoff, D. Griffeath, and S. Levin. Local frequency dependence and global coexistence. *Theoretical population biology*, 55(3):270–282, 1999.

[171] A. Montanari and A. Saberi. Convergence to equilibrium in local interaction games and ising models. *arXiv preprint arXiv:0812.0198*, 2008.

[172] E. Mossel and S. Roch. On the submodularity of influence in social networks. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 128–134. ACM, 2007.

[173] E. Mossel and S. Roch. Submodularity of influence in social networks: From local to global. *SIAM J. Comput.*, 39(6):2176–2188, 2010.

[174] E. Mossel and G. Schoenebeck. Arriving at consensus in social networks. In *The First Symposium on Innovations in Computer Science (ICS 2010)*, January 2010.

[175] E. Mossel and O. Tamuz. Opinion exchange dynamics. *Probability Surveys*, 14: 155–204, 2017.

[176] E. Mossel, J. Neeman, and O. Tamuz. Majority dynamics and aggregation of information in social networks. *Autonomous Agents and Multi-Agent Systems*, 28(3):408–429, 2014.

[177] M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[178] M. E. Newman and D. J. Watts. Scaling and percolation in the small-world network model. *Physical review E*, 60(6):7332, 1999.

[179] M. E. Newman, C. Moore, and D. J. Watts. Mean-field solution of the small-world network model. *Physical Review Letters*, 84(14):3201, 2000.

[180] M. E. J. Newman and D. J. Watts. Renormalization group analysis of the small-world network model. *Phys. Lett. A*, 263(4):341–346, 6 Dec. 1999.

[181] G. Noh, H. Oh, Y.-m. Kang, and C.-k. Kim. Psd: Practical sybil detection schemes using stickiness and persistence in online recommender systems. *Information Sciences*, 281:66–84, 2014.

[182] M. A. Nowak and D. C. Krakauer. The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14):8028–8033, 1999.

[183] M. A. Nowak, J. B. Plotkin, and D. C. Krakauer. The evolutionary language game. *Journal of Theoretical Biology*, 200(2):147–162, 1999.

[184] M. A. Nowak, J. B. Plotkin, and V. A. Jansen. The evolution of syntactic communication. *Nature*, 404(6777):495–498, 2000.

[185] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the national academy of sciences*, 104(18):7332–7336, 2007.

[186] F. Papadopoulos, D. Krioukov, M. Boguñá, and A. Vahdat. Greedy forwarding in dynamic scale-free networks embedded in hyperbolic metric spaces. In *2010 Proceedings IEEE INFOCOM*, pages 1–9. IEEE, 2010.

[187] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguná, and D. Krioukov. Popularity versus similarity in growing networks. *Nature*, 489(7417):537, 2012.

[188] R. Pemantle. When are touchpoints limits for generalized pólya urns? *Proceedings of the American Mathematical Society*, 113(1):235–243, 1991.

[189] R. Pemantle et al. A survey of random processes with reinforcement. *Probab. Surv*, 4(0):1–79, 2007.

[190] E. Perron, D. Vasudevan, and M. Vojnovic. Using three states for binary consensus on complete graphs. In *INFOCOM 2009, IEEE*, pages 2527–2535. IEEE, 2009.

[191] Pew. Political polarization & media habits, 2014.

[192] W. Pickering and C. Lim. Solution of the multistate voter model and application to strong neutrals in the naming game. *Physical Review E*, 93(3):032318, 2016.

[193] D. J. D. S. Price. Networks of scientific papers. *Science*, pages 510–515, 1965.

[194] A. E. Raftery, X. Niu, P. D. Hoff, and K. Y. Yeung. Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics*, 21(4):901–919, 2012.

[195] A. Rapoport. Spread of information through a population with socio-structural bias: I. assumption of transitivity. *The bulletin of mathematical biophysics*, 15 (4):523–533, 1953.

[196] M. Richardson, R. Agrawal, and P. M. Domingos. Trust management for the semantic web. In *The Semantic Web - ISWC 2003, Second International Semantic Web Conference, Sanibel Island, FL, USA, October 20-23, 2003, Proceedings*, pages 351–368, 2003. doi: 10.1007/978-3-540-39718-2_23. URL https://doi.org/10.1007/978-3-540-39718-2_23.

[197] C. Robinson. *Dynamical systems: stability, symbolic dynamics, and chaos*. CRC press, 1998.

[198] F. J. Rohlf and G. D. Schnell. An investigation of the isolation-by-distance model. *The American Naturalist*, 105(944):295–324, 1971.

[199] D. M. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics : Idioms , political hashtags , and complex contagion on twitter. In *Proceedings of the 20th international World Wide Web Conference*, pages 695–704. ACM, 2011. URL http://dl.acm.org/citation.cfm?id=1963503.

[200] N. Ross et al. Fundamentals of stein's method. *Probability Surveys*, 8:210–293, 2011.

[201] B. Ryan and N. C. Gross. The diffusion of hybrid seed corn in two iowa communities. *Rural sociology*, 8(1):15, 1943.

[202] P. Sarkar and A. W. Moore. Dynamic social network analysis using latent space models. *ACM SIGKDD Explorations Newsletter*, 7(2):31–40, 2005.

[203] P. Sarkar, D. Chakrabarti, and A. W. Moore. Theoretical justification of popular link prediction heuristics. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, page 2722, 2011.

[204] G. Schoenebeck and B. Tao. Beyond worst-case (in)approximability of non-submodular influence maximization. In *International Conference on Web and Internet Economics*, pages 368–382. Springer, 2017.

[205] G. Schoenebeck and F.-Y. Yu. Complex contagions on configuration model graphs with a power-law degree distribution. In *International Conference on Web and Internet Economics*, pages 459–472. Springer, 2016.

[206] G. Schoenebeck and F.-Y. Yu. Consensus of interacting particle systems on erdös-rényi graphs. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1945–1964. SIAM, 2018.

[207] G. Schoenebeck, A. Snook, and F. Yu. Sybil detection using latent network structure. In *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, pages 739–756, 2016. doi: 10.1145/2940716.2940747. URL https://doi.org/10.1145/2940716.2940747.

[208] F. Schweitzer and L. Behera. Nonlinear voter models: the transition from invasion to coexistence. *The European Physical Journal B-Condensed Matter and Complex Systems*, 67(3):301–318, 2009.

[209] J. Scott. *Social network analysis*. Sage, 1988.

[210] C. Sibona. Facebook fallout: Future contact avoidance after being unfriended on facebook. In *2013 46th Hawaii International Conference on System Sciences*, pages 3272–3281. IEEE, 2013.

[211] R. Solomonoff and A. Rapoport. Connectivity of random nets. *The bulletin of mathematical biophysics*, 13(2):107–117, 1951.

[212] V. Sood and S. Redner. Voter model on heterogeneous graphs. *Physical review letters*, 94(17):178701, 2005.

[213] C. Sower. Medical innovation: A diffusion study., 1967.

[214] L. Steels. A self-organizing spatial vocabulary. *Artificial life*, 2(3):319–332, 1995.

[215] L. Steels. The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. *Connection science*, 17(3-4):213–230, 2005.

[216] L. Steels and A. McIntyre. Spatially distributed naming games. *Advances in complex systems*, 1(04):301–323, 1998.

[217] C. Steglich, T. A. B. Snijders, and M. Pearson. 8. dynamic networks and behavior: Separating selection from influence. *Sociol. Methodol.*, 40(1):329–393, 15 June 2010.

[218] M. Steyvers and J. B. Tenenbaum. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, 29(1):41–78, 2005.

[219] D. Strang and S. A. Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual review of sociology*, 24(1):265–290, 1998.

[220] S. H. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press, 2014.

[221] K. Suchecki, V. M. Eguiluz, and M. San Miguel. Conservation laws for the voter model in complex networks. *EPL (Europhysics Letters)*, 69(2):228, 2005.

[222] K. Suchecki, V. M. Eguíluz, and M. San Miguel. Voter model dynamics in complex networks: Role of dimensionality, disorder, and degree distribution. *Physical Review E*, 72(3):036132, 2005.

[223] C. R. Sunstein. *Republic.com 2.0*. Princeton University Press, 17 Aug. 2009.

[224] O. Tamuz and R. J. Tessler. Majority dynamics and the retention of information. *Israel Journal of Mathematics*, 206(1):483–507, 2015.

[225] D. N. Tran, B. Min, J. Li, and L. Subramanian. Sybil-resilient online content voting. In *NSDI*, volume 9, pages 15–28, 2009.

[226] N. Tran, J. Li, L. Subramanian, and S. S. Chow. Optimal sybil-resilient node admission control. In *INFOCOM, 2011 Proceedings IEEE*, pages 3218–3226. IEEE, 2011.

[227] J. Travers and S. Milgram. An experimental study of the small world problem. In *Social Networks*, pages 179–197. Elsevier, 1977.

[228] G. Tullock. Toward a theory of the rent-seeking society, chapter efficient rent seeking,(pp. 112), 1980.

[229] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg. Structural diversity in social contagion. *Proc. National Academy of Sciences*, 109(16):5962–5966, April 2012.

[230] R. Van Der Hofstad. Random graphs and complex networks. *Available on http://www. win. tue. nl/rhofstad/NotesRGCN. pdf*, page 11, 2009.

[231] B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove. An analysis of social network-based sybil defenses. *ACM SIGCOMM Computer Communication Review*, 41(4):363–374, 2011.

[232] L. Von Ahn, M. Blum, N. J. Hopper, and J. Langford. Captcha: Using hard ai problems for security. In *Advances in CryptologyEUROCRYPT 2003*, pages 294–311. Springer, 2003.

[233] G. Wang, T. Konolige, C. Wilson, X. Wang, H. Zheng, and B. Y. Zhao. You are how you click: Clickstream analysis for sybil detection. In *Usenix Security*, pages 241–256, 2013.

[234] D. J. Watts and S. H. Strogatz. Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442, 1998.

[235] D. J. Watts, P. S. Dodds, and M. E. Newman. Identity and search in social networks. *science*, 296(5571):1302–1305, 2002.

[236] W. Wei, F. Xu, C. C. Tan, and Q. Li. Sybildefender: Defend against sybil attacks in large social networks. In *INFOCOM, 2012 Proceedings IEEE*, pages 1951–1959. IEEE, 2012.

[237] H. C. White, S. A. Boorman, and R. L. Breiger. Social structure from multiple networks. i. blockmodels of roles and positions. *American Journal of Sociology*, 81(4):730–780, 1976. URL http://www.jstor.org/stable/2777596.

[238] R. J. Williams, E. L. Berlow, J. A. Dunne, A.-L. Barabási, and N. D. Martinez. Two degrees of separation in complex food webs. *Proceedings of the National Academy of Sciences*, 99(20):12913–12916, 2002.

[239] N. C. Wormald. Differential equations for random processes and random graphs. *The annals of applied probability*, pages 1217–1235, 1995.

[240] N. C. Wormald. Models of random regular graphs. *London Mathematical Society Lecture Note Series*, pages 239–298, 1999.

[241] B. Xu, Y. Huang, H. Kwak, and N. S. Contractor. Structures of broken ties: exploring unfollow behavior on twitter. In *Computer Supported Cooperative Work, CSCW 2013, San Antonio, TX, USA, February 23-27, 2013*, pages 871–876, 2013. doi: 10.1145/2441776.2441875. URL https://doi.org/10.1145/2441776.2441875.

[242] Z. Yang, C. Wilson, X. Wang, T. Gao, B. Y. Zhao, and Y. Dai. Uncovering social network sybils in the wild. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 8(1):2, 2014.

[243] S. Yardi, D. Romero, G. Schoenebeck, et al. Detecting spam in a twitter network. *First Monday*, 15(1), 2009.

[244] J. Yin and S. K. Madria. Sybil attack detection in a hierarchical sensor network. In *Security and Privacy in Communications Networks and the Workshops, 2007. SecureComm 2007. Third International Conference on*, pages 494–503. IEEE, 2007.

[245] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. Sybilguard: defending against sybil attacks via social networks. *ACM SIGCOMM Computer Communication Review*, 36(4):267–278, 2006.

[246] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao. Sybillimit: A near-optimal social network defense against sybil attacks. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pages 3–17. IEEE, 2008.

[247] H. Yu, C. Shi, M. Kaminsky, P. B. Gibbons, and F. Xiao. Dsybil: Optimal sybil-resistance for recommendation systems. In *Security and Privacy, 2009 30th IEEE Symposium on*, pages 283–298. IEEE, 2009.

[248] A. N. Zehmakan. Opinion forming in binomial random graph and expanders. *CoRR*, abs/1805.12172, 2018. URL http://arxiv.org/abs/1805.12172.

APPENDIX A

Proofs for General Kleinberg's Small World Models

A.1 Proof in the preliminary

In contrast to the original Kleinberg model, when the parameter γ is allow to scale with the size of the graph n the normalization factor λ_γ requires more careful treatment than the O-notation.

Proof of Lemma 3.1. The proof is easy, and is include for completeness. Recall that to eliminate the boundary effect, we wrap up the grid into a torus, and by definition of λ_γ, we have

$$4\sum_{k=1}^{\lfloor n/2\rfloor}\frac{k}{k^\gamma} \le \lambda_\gamma = \sum_v d(u,v)^{-\gamma} \le 4\sum_{k=1}^{n}\frac{k}{k^\gamma} \tag{A.1}$$

Now we use integration $\int x^{1-\gamma}\,dx$ to approximate the summation $\sum k^{1-\gamma}$. There are two cases: if $\gamma < 1$, $x^{1-\gamma}$ is increasing, and if $\gamma \le 1$ is decreasing. As a result,

$$\int_1^{m+1} x^{1-\gamma}\,dx \le \sum_{k=1}^{m} k^{1-\gamma} \le 1 + \int_1^m x^{1-\gamma}\,dx \qquad \text{if } \gamma > 1 \tag{A.2}$$

$$\int_0^{m} x^{1-\gamma}\,dx \le \sum_{k=1}^{m} k^{1-\gamma} \le \int_1^{m+1} x^{1-\gamma}\,dx \qquad \text{if } \gamma \le 1 \tag{A.3}$$

Now we are ready the prove the above inequality.

Case: $\gamma \ge 3$**.** Apply Equation (A.2) into (A.1), and we have

$$\lambda_\gamma \le 4\sum_{k=1}^{n}\frac{k}{k^\gamma} \le 4\left(1+\int_1^n x^{1-\gamma}\,dx\right) \le 4\left(1+\frac{1}{\gamma-2}(1-n^{-(\gamma-2)})\right) \le 4\left(1+\frac{1}{\gamma-2}\right)$$

For the lower bound, we have

$$\lambda_\gamma \geq 4\sum_{k=1}^{\lfloor n/2\rfloor}\frac{k}{k^\gamma} \geq 4\left(\int_1^{n/2} x^{1-\gamma}\,dx\right) \geq 4\left(\frac{1}{\gamma-2}(1-(n/2)^{-(\gamma-2)})\right)$$
$$\geq 2\left(\frac{1}{\gamma-2}\right) \qquad \text{(Because } \gamma-2\geq 1\text{, it holds when } n\geq 4.\text{)}$$

Case: $2<\gamma<3$**.** For the upper bound, by similar computation we have

$$\lambda_\gamma \leq 4\sum_{k=1}^{n}\frac{k}{k^\gamma} \leq 4\left(1+\int_1^n x^{1-\gamma}\,dx\right) \leq 4+\frac{4}{\gamma-2}\left(1-n^{-(\gamma-2)}\right)$$

Here we need to be more careful. Let $r_1 \triangleq (\gamma-2)\ln n$, and we have $\lambda_\gamma \leq 4+\frac{4}{\gamma-2}\left(1-\exp\left(-r_1\right)\right)$. If $r_1\geq 1$, because $\gamma-2<1$

$$\lambda_\gamma \leq 4+\frac{4}{\gamma-2} \leq 8\left(\frac{1}{\gamma-2}\right).$$

Otherwise, if $r_1<1$, since $\exp(-r_1)\geq 1-r_1$ and

$$\lambda_\gamma \leq 4+\frac{4\ln n}{r_1}r_1 \leq 8\ln n.$$

For the lower bound, after similar computation we have

$$\lambda_\gamma \geq \frac{4}{\gamma-2}\left(1-(n/2)^{-(\gamma-2)}\right) = \frac{4}{\gamma-2}\left(1-\exp\left(-(\gamma-2)\ln\frac{n}{2}\right)\right)$$

Let $r_2 \triangleq (\gamma-2)\ln\frac{n}{2}$, and we have $\lambda_\gamma \geq \frac{4}{\gamma-2}\left(1-\exp\left(-r_2\right)\right)$. If $r_2\geq 1$, $\exp(-r_2)\leq e^{-1}$ and

$$\lambda_\gamma \geq \frac{4}{\gamma-2}\left(1-e^{-1}\right) \geq \frac{2}{\gamma-2}.$$

Otherwise, if $r_2<1$, by convexity of exp we have $\exp(-r_2)\leq 1-(1-e^{-1})r_2$ and

$$\lambda_\gamma \geq \frac{4\ln\frac{n}{2}}{r_2}\left((1-e^{-1})r_2\right) \geq 2\ln\frac{n}{2}.$$

Case: $\gamma=2$**.** We just use Equation (A.2) and if $n\geq e$ we have

$$\lambda_\gamma \leq 4\sum_{k=1}^{n}\frac{k}{k^\gamma} \leq 4\left(1+\int_1^n x^{1-\gamma}\,dx\right) \leq 4\left(1+\ln n\right) \leq 8\ln n$$

For the lower bound, if we have

$$\lambda_\gamma \geq 4 \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{k}{k^\gamma} \geq 4 \left(\int_1^{n/2} x^{1-\gamma}\, dx \right) \geq 4 \ln \frac{n}{2}$$

Case: $1 \leq \gamma < 2$**.** We just use Equation (A.2) and have

$$\lambda_\gamma \leq 4 \sum_{k=1}^{n} \frac{k}{k^\gamma} \leq 4 \left(1 + \int_1^{n} x^{1-\gamma}\, dx \right) = 4 \left(1 + \frac{1}{2-\gamma} \left(n^{2-\gamma} - 1 \right) \right)$$

Here we need to be more careful. Let $r_3 \triangleq (2-\gamma) \ln n$, and we have $\lambda_\gamma \leq 4 + \frac{4}{2-\gamma} (\exp(r_3) - 1)$. If $r_3 \geq 1$, because $2 - \gamma \leq n^{2-\gamma}$ when $n \geq e$

$$\lambda_\gamma \leq 4 + \frac{4}{\gamma - 2} \leq \frac{8}{2-\gamma} n^{2-\gamma}.$$

Otherwise, if $r_3 < 1$, since $\exp(r) \leq 1 + (e-1)r$ and

$$\lambda_\gamma \leq 4 + \frac{4(e-1) \ln n}{r_3} r_3 \leq 8 \ln n.$$

For the lower bound, we have

$$\lambda_\gamma \geq \frac{4}{2-\gamma} \left((n/2)^{2-\gamma} - 1 \right) = \frac{4}{2-\gamma} \left(\exp\left((2-\gamma) \ln \frac{n}{2} \right) - 1 \right)$$

Let $r_4 \triangleq (2-\gamma) \ln \frac{n}{2}$, and we have $\lambda_\gamma \geq \frac{4}{2-\gamma} (\exp(r_4) - 1)$. If $r_4 \geq 1$, $\exp(r_4) - 1 \geq (1/2) \exp(r_4)$ and

$$\lambda_\gamma \geq \frac{4}{2-\gamma} (e-1) \geq \frac{2}{2-\gamma} \left(\frac{n}{2} \right)^{2-\gamma}.$$

Otherwise, if $r_4 < 1$, we have $\exp(r_4) \geq 1 + r_4$ and

$$\lambda_\gamma \geq \frac{4 \ln \frac{n}{2}}{r_4} (r_4) \geq 2 \ln \frac{n}{2}.$$

Case: $\gamma < 1$**.** Apply Equation (A.3) into (A.1), and if $n \geq 4$ we have

$$\lambda_\gamma \leq 4 \sum_{k=1}^{n} \frac{k}{k^\gamma} \leq 4 \left(\int_1^{n+1} x^{1-\gamma}\, dx \right) \leq \frac{4}{2-\gamma} \left((n+1)^{2-\gamma} - 1 \right) \leq \frac{8}{2-\gamma} n^{2-\gamma}$$

For the lower bound, if $n \geq 7$ we have

$$\lambda_\gamma \geq 4 \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{k}{k^\gamma} \geq 4 \left(\int_0^{n/2-1} x^{1-\gamma}\, dx \right) \geq \frac{4}{\gamma - 2} (n/2 - 1)^{2-\gamma} \geq \frac{2}{\gamma - 2} (n/2)^{2-\gamma}$$

□

A.2 Proofs for Myopic Routing

Proof of Theorem 3.2. As defined in section 3.3, we use $\mathbb{A} = (x_i)_{i\geq 0}$ to denote the routing process of the myopic greedy algorithm towards the destination t. Recall the definition of the myopic greedy algorithm which greedily routes the message from the "current location" to a neighbor that is as close as possible to the destination vertex. This process is memory-less. Because one of the grid neighbors will always have a smaller distance to t, the distance to the destination is strictly decreasing, and the sequence $(x_i)_{i\geq 0}$ is a simple path. Therefore for each node u that $\mathbb{A}$ encounters, the value of γ_u and the formation of node u's long-range contacts are independent of other nodes, and the process $(x_i)_{i\geq 0}$ is a Markov chain, and we can study its behavior independent of its history.

To measure the progress of process $\mathbb{A} = (x_i)_{i\geq 0}$, we say that the i-th run of $\mathbb{A}$ is in *phase* j when the lattice distance from the current node to t is between 2^j and 2^{j+1}— $2^j < d(x_i, t) \leq 2^{j+1}$. Note that the phase of myopic greedy algorithm $\mathbb{A}$ is always non-increasing because the distance to destination is strictly decreasing, and the initial value of j is at most $\log n$ which prove the first two parts.

Now we are ready to address the last part of this proof: *what is the probability P_j that phase j will end after this step?* Note that this event is well defined because the the phases is non-increasing. Suppose there exists constant $L > 0$ depending on α and K such that

$$P_j \geq L \cdot \frac{1}{\log^{\alpha+1} n} \tag{A.4}$$

Because this event is independent for each node, the expected number of steps spent in phase j is $(\log^{\alpha+1} n)/L$. Therefore, the expected routing time from phase $\log n$ is $O\left(\log^{\alpha+2} n\right)$, and the expected delivery time is $O\left(\log^{\alpha+2} n\right)$ which finishes the proof.

To prove Equation (A.4), note that if node u has a neighbor to a node v such that $d(v, t) < 2^j$, the phase j will end. We fix the γ_u, and compute the probability of ending phase j—P_{j,γ_u}. Then we use the property of $\mathcal{D}$ and can compute $P_j = \mathbb{E}_{\gamma_u\sim\mathcal{D}}[P_{j,\gamma_u}]$. With Lemma 3.1 if $|\epsilon| < 1/\log n$, a direct computation:

$$\begin{aligned} P_{j,2+\epsilon} &\triangleq \Pr\left[\mathbb{A} \text{ exits phase } j | \gamma_u = 2+\epsilon\right] = \frac{2 \cdot 2^{2j}}{\lambda_{2+\epsilon}(2^{j+2})^{2+\epsilon}} \\ &\geq \begin{cases} \frac{1}{256} \cdot \frac{1}{(\log n) 2^{j\epsilon}} & \text{if } 0 \leq \epsilon \leq 1/\log n \\ \frac{1}{256} \cdot \frac{1}{(\log n) 2^{(\log n - j)|\epsilon|}} & \text{if } -1/\log n \leq \epsilon < 0 \end{cases} \end{aligned}$$

The first equality is because there are $2(2^j)^2$ nodes at distance at most 2^j from t, and each is distance at most 2^{j+2} away from the current location. The last equality follows from the computation of λ_γ given in Lemma 3.1. Because $0 \leq j/\log n \leq 1$, we have

$$P_{j,2+\epsilon} \triangleq \Pr\left[\mathbb{A} \text{ exists phase } j | \gamma_u = 2+\epsilon\right] \geq \frac{1}{512} \cdot \frac{1}{\log n} \tag{A.5}$$

By applying the condition of $\mathcal{D}$ and taking n large enough such that $1/\log n < \epsilon_0$,

$M_{\mathcal{D}}\left(\frac{1}{\log n}\right) \geq \frac{K}{\log^{\alpha} n}$. Then we can use a point-wise comparison of $P_{j,2+\epsilon}$ and derive

$$\begin{aligned} P_j = \mathbb{E}_{\gamma\sim\mathcal{D}}[P_{j,\gamma}] &\geq \int_{-1/\log n}^{1/\log n} P_{j,2+\epsilon}\, f_{\mathcal{D}}(2+\epsilon)d\epsilon \\ &\geq \frac{1}{512}\cdot\frac{1}{\log n} M_{\mathcal{D}}\left(\frac{1}{\log n}\right) \geq \frac{K}{512}\cdot\frac{1}{\log^{\alpha+1} n} \end{aligned}$$

which proves Equation A.4 with $L = \frac{K}{512}$.

□

A.3 Proofs for Complex Contagions

Proof of Lemma 3.6. By Equation (3.1) we have

$$\Pr[\text{A fixed weak tie of } u \text{ connects to } B_j|2+\epsilon] \geq \begin{cases} \frac{1}{256}\cdot\frac{1}{\log n 2^{j\epsilon}} & \text{if } 0 \leq \epsilon < 1/\log n \\ \frac{1}{256}\cdot\frac{1}{\log n 2^{(\log n - j)|\epsilon|}} & \text{if } -1/\log n < \epsilon \leq 0 \end{cases}$$

Because fixing a node u in A_j the connection of its p weak ties are mutually independent given γ_u, for any $\ell \leq k$

$$\Pr[\ell \text{ of } u_1\text{'s } p \text{ weak ties connect to } B_j] \geq \int \Pr[\text{A fixed weak tie of } u \text{ connects to } B_j|\gamma_u = x]^{\ell} f_{\mathcal{D}}(x)$$

Similar arguments in the proof of Theorem 3.2 yields

$$\Pr[\ell \text{ of } u_1\text{'s } p \text{ weak ties connect to } B_j] \geq \frac{K}{2^{9\ell}}\cdot\frac{1}{(\log n)^{\ell+\alpha}}$$

Since weak ties from different nodes are mutually independent by definition, the probability that a k-seed cluster is good is at least $\frac{C}{\log^{k\alpha+k(k+1)/2} n}$ with some constant $C > 0$ that depends on $\mathcal{D}$ and k. This finishes the proof. □

Proof of Lemma 3.10. Essentially, we want to lower bound the probability of forming a long connection which yields the statement of this lemma by careful arguments.

Fix $\ell > (r\log^c n)^{(\frac{1}{1-\delta})}$ where constant r will be defined later, an ℓ-sized square S and two disjoint $\ell^{(1-\delta)}$-sized subsquares A and B, and partition B into $\ell^{(1-\delta)}/k$ disjoint seed clusters $(u_1, \ldots, u_k)$.

Let $\mu_2 = \Pr[\gamma \in (2, \tilde{\gamma})]$.

- $Q_1(\ell) = \Pr[u_1 \text{ is connected to } A, \text{ via } k \text{ weak ties}]$
- $Q'_s(\ell) = \Pr[u_s \text{ has } k \text{ weak ties to } u_1]$
- $P_1(\ell) = \Pr[(u_1, \ldots u_k) \text{ form a } \textit{new seed} \text{ in } B]$
- $P_2(\ell) = \Pr[\text{a } \textit{new seed} \text{ forms in } B]$

For Q_1 and Q'_s we have

$$\begin{aligned}
Q_1(\ell) &= \int_2^{\beta_k} \left(\Pr\left[u \text{ has a weak tie to } A, \text{ via a particular edge} | \gamma_u = \gamma\right]\right)^k f_{\mathcal{D}}(\gamma) d\gamma \\
&\geq \int_2^{\beta_k} \left(\frac{|A|\lambda_\gamma}{d(u,a)^\gamma}\right)^k f_{\mathcal{D}}(\gamma) d\gamma \geq \int_2^{\beta_k} \left(\lambda_\gamma \ell^{1-\delta-\gamma/2}\right)^k f_{\mathcal{D}}(\gamma) d\gamma \\
&\geq \int_2^{\tilde{\gamma}} \left(\lambda_\gamma \ell^{1-\delta-\gamma/2}\right)^k f_{\mathcal{D}}(\gamma) d\gamma \\
&= \mu_2 \left(\lambda_{\hat{\gamma}} \ell^{1-\delta-\hat{\gamma}/2}\right)^k \text{ where } \hat{\gamma} \in (2, \tilde{\gamma}) && \text{(by Mean value theorems)} \\
&= \Omega\left(\ell^{1-\delta-\tilde{\gamma}/2}\right) && \text{(since } \hat{\gamma} < \tilde{\gamma} \text{ and the constant depends on } \mathcal{D} \text{ and } k)
\end{aligned}$$

and

$$\begin{aligned}
Q'_s(\ell) &= \int_2^{\beta_k} \left(\Pr\left[u \text{ has a weak tie to } u_1, \text{ via a particular edge} | \gamma_u = \gamma\right]\right)^k f_{\mathcal{D}}(\gamma) d\gamma \\
&\geq \int_2^{\beta_k} \left(\frac{\lambda_\gamma}{k^\gamma}\right)^k f_{\mathcal{D}}(\gamma) d\gamma && \text{(decreasing with respect to distance)} \\
&\geq \int_2^{\tilde{\gamma}} \left(\frac{\lambda_\gamma}{k^\gamma}\right)^k f_{\mathcal{D}}(\gamma) d\gamma \\
&= \mu_2 \left(\frac{\lambda_{\bar{\gamma}}}{k^{\bar{\gamma}}}\right)^k \text{ where } \bar{\gamma} \in (2, \tilde{\gamma}) && \text{(by Mean value theorems)} \\
&= \Omega(1) && \text{(the constant depends on } \mathcal{D} \text{ and } k)
\end{aligned}$$

For $P_1(\ell)$, the event $(u_1, \ldots, u_k)$ becoming a *new seed*, we have

$$P_1(\ell) = Q_1(\ell) \times Q'_2(\ell) \times \ldots \times Q'_k(\ell) = \Omega\left(\ell^{k(1-\delta-\tilde{\gamma}/2)}\right)$$

and all disjoint seed clusters in B are mutually independent, so

$$P_2(\ell) = 1 - (1 - P_1(\ell))^{|B|/k} \geq 1 - \exp\left(P_1(\ell)|B|/k\right) \geq 1 - \exp\left(-\Omega(\ell^{(k+1)(1-\delta)-\frac{k\tilde{\gamma}}{2}})\right)$$

Note that P_2 is increasing in ℓ. Therefore, the smallest probability happens when the ℓ is the smallest; that is $\ell = (r \log^c n)^{\frac{1}{1-\delta}}$, and

$$P_2(\ell) \geq 1 - \exp\left(-\Omega\left(r' \log^{\frac{c}{1-\delta}\left((k+1)(1-\delta)-\frac{k\tilde{\gamma}}{2}\right)} n\right)\right)$$

where constant $r' = r^{(k+1)(1-\delta)-\frac{k\tilde{\gamma}}{2}}$ By the definition of c, $c \geq \frac{1-\delta}{(k+1)(1-\delta)-k\tilde{\gamma}/2}$, and taking r large enough, we have

$$P_2(\ell) \geq 1 - n^{-\eta}.$$

□

Proof of Lemma 3.14. Let's define $\mathcal{J}_v$ to be the event that v satisfies the conditions in event $\mathcal{J}$. Clearly $\mathcal{J} = \bigcup_{v\in V} \mathcal{J}_v$. Now we consider $\Pr[\mathcal{J}]$ when the nodes' personal γ parameters have been revealed and partition the events $\{\mathcal{J}_v\}$ by two conditions: 1) if $d(v,s)$ is less than $2\rho_1$ or greater than $2\rho_1$; 2) if γ_v is less than γ_1 or greater than γ_2.

$$\begin{aligned}
&\Pr[\mathcal{J}] = \Pr\left[\bigcup_{v\in V} \mathcal{J}_v\right] \le \sum_{v\in V} \Pr[\mathcal{J}_v] \\
=&\sum_{\substack{v\in V:\\ d(v,s)\le 2\rho_1,\\ \gamma_v<\gamma_1}} \Pr[\mathcal{J}_v] + \sum_{\substack{v\in V:\\ d(v,s)\le 2\rho_1,\\ \gamma_v>\gamma_2}} \Pr[\mathcal{J}_v] + \sum_{\substack{v\in V:\\ d(v,s)>2\rho_1,\\ \gamma_v<\gamma_1}} \Pr[\mathcal{J}_v] + \sum_{\substack{v\in V:\\ d(v,s)>2\rho_1,\\ \gamma_v>\gamma_2}} \Pr[\mathcal{J}_v]
\end{aligned}$$

There are four terms on the right hand side, if all of them are $O(n^{-\eta})$, then we finish the proof. Now we estimate them term by term.

To make the above four terms work, the choice of ρ_1 and ρ_2 are critical: ρ_1 should be small to prevent term 1 nodes from forming new seeds, and also so we can apply a union bound over the term 2 nodes. But ρ_1 should also be large enough to prevent term 3 and 4 nodes from forming new seeds. Additionally, ρ_2 should be large enough to prevent nodes in term 2 from forming new seeds. But ρ_2 also needs to be much smaller than ρ_1 to derived a polynomial lower bound on the contagions time.

Type 1: $d(v,s) \le 2\rho_1$ **and** $\gamma_v < \gamma_1$ In this case, v is nearby and has a small γ_v, i.e., a set of diffuse edges. Note that by triangle inequality the ball $B_{\rho_1}(s) = \{u : d(u,s) \le \rho_1\}$ is contained in the ball $B_{3\rho_1}(v) = \{u : d(u,v) \le 3\rho_1\}$. Thus the probability that one weak tie issued by v falls inside $B_{\rho_1}(s)$ is no greater than the probability of this weak tie connecting to nodes in $B_{3\rho_1}(v)$. The latter can be bounded from above:

$$\begin{aligned}
&\Pr[\text{A given edge of } v \text{ connects to some node in } B_{3\rho_1}(v) | \gamma_v \le \gamma_1] \\
\le&\int_{3\rho_1}^{1} \frac{\lambda_\gamma}{y^\gamma}\Theta(y)dy = O(\lambda_\gamma(3\rho_1)^{2-\gamma}) = O\left(\frac{n^{(2-\gamma)\delta_1}}{n^{2-\gamma}}\right) \\
=&O\left(\frac{1}{n^{\epsilon_1(1-\delta_1)}}\right) && \text{(by the definition of } \epsilon_1)
\end{aligned}$$

Recall that each edge issued by v is independent of the others,

$$\begin{aligned}
&\Pr[J_v | \gamma_v \le \gamma_1] \\
\le&\Pr[\text{ At least } k \text{ weak ties of } v \text{ connect to some node in } B_{3\rho_1}(v)] \\
\le&\binom{p}{k} \cdot \Pr[\text{A given edge of } v \text{ connects to some node in } B_{3\rho_1}(v)]^k && \text{(union bound)} \\
=&O\left(\frac{1}{n^{k(\epsilon_1(1-\delta_1))}}\right) && (p \text{ is constant})
\end{aligned}$$

Therefore

$$\sum_{\substack{v\in V:\\ d(v,s)\le 2\rho_1,\\ \gamma_v<\gamma_1}} \Pr[\mathcal{J}_v] = \sum_{\substack{v\in V:\\ d(v,s)\le 2\rho_1,\\ \gamma_v<\gamma_1}} O\left(\frac{1}{n^{k(\epsilon_1(1-\delta_1))}}\right)$$
$$= O\left(\frac{\rho_1^2}{n^{k(\epsilon_1(1-\delta_1))}}\right) = O(1/n^{(k\epsilon_1+2)\left(\frac{k\epsilon_1}{k\epsilon_1+2}-\delta_1\right)}) = O(n^{-\eta})$$

Type 2: $d(v,s) \le 2\rho_1$ **and** $\gamma_v > \gamma_2$ v is nearby but has a large γ_v, i.e., a set of concentrated edges. The probability that one weak tie issued by v has length more than ρ_2 can be bounded from above:

$$\begin{aligned}
&\Pr[\text{A given edge of } v \text{ connects to a node } u \text{ in } B_{\rho_1}(s) \text{ and } d(u,v) > \rho_2]\\
\le &\Pr[\text{A given edge of } v \text{ connects to a node } u \text{ such that } d(u,v) > \rho_2]\\
= &\int_{\rho_2}^{n} \frac{\lambda_\gamma}{y^\gamma}\Theta(y)dy = O(\rho_2^{2-\gamma}) = O\left(\frac{1}{n^{\epsilon_2\delta_2}}\right)
\end{aligned}$$

Again by applying the union bound, $\Pr[J_v] \le O(1/n^{k\epsilon_2\delta_2})$.

Therefore we can upper bound the second type as follows

$$\sum_{\substack{v\in V:\\ d(v,s)\le 2\rho_1,\\ \gamma_v>\gamma_2}} \Pr[\mathcal{J}_v] = \sum_{\substack{v\in V:\\ d(v,s)\le 2\rho_1,\\ \gamma_v>\gamma_2}} O\left(\frac{1}{n^{k\epsilon_2\delta_2}}\right) = O\left(\frac{\rho_1^2}{n^{k\epsilon_2\delta_2}}\right) = O(1/n^{k\epsilon_2\left(\delta_2-\frac{2}{k\epsilon_2}\delta_1\right)}) = O(n^{-\eta})$$

Type 3: $d(v,s) > 2\rho_1$ **and** $\gamma_v < \gamma_1$ Since $d(v,s) > 2\rho_1$, $\forall u \in B_{\rho_1}(s)$ we have $d(v,u) > d(v,s) - \rho_1 > d(v,s)/2$ by triangle inequality, and we have the following upper bound

$$\begin{aligned}
&\Pr[\text{A given edge of } v \text{ connects to some node in } B_{\rho_1}(s)]\\
&= O\left(|B_{\rho_1}(s)|\frac{\lambda_\gamma}{d(v,s)^\gamma}\right) = O\left(\frac{\lambda_\gamma\rho_1^2}{d(v,s)^\gamma}\right)
\end{aligned}$$

Since the edges issued by v are independent, $\Pr[\mathcal{J}_v] = O\left(\binom{m}{k}\left(\frac{\lambda_\gamma \rho_1^2}{d(v,s)^\gamma}\right)^k\right)$ and

$$\begin{aligned}
&\sum_{\substack{v\in V:\\ d(v,s)>2\rho_1,\\ \gamma_v<\gamma_1}} \Pr[\mathcal{J}_v] = \sum_{\substack{v\in V:\\ d(v,s)>2\rho_1,\\ \gamma_v<\gamma_1}} O\left(\left(\frac{\lambda_\gamma \rho_1^2}{d(v,s)^\gamma}\right)^k\right)\\
=&O\left(\int_{\rho_1}^{n}\left(\frac{\lambda_\gamma \rho_1^2}{y^\gamma}\right)^k y dy\right) = O\left(\lambda_\gamma^k \rho_1^{2k}\int_{\rho_1}^{n} y^{1-k\gamma} dy\right)\\
=&\begin{cases}\left(\frac{\rho_1^2}{n^{2-\gamma}}\right)^k n^{2-k\gamma}, 0\le\gamma<2/k,\\ \left(\frac{\rho_1^2}{n^{2-\gamma}}\right)^k \log n, \gamma = 2/k\\ \left(\frac{\rho_1^2}{n^{2-\gamma}}\right)^k \rho_1^{2-k\gamma}, 2/k<\gamma<\gamma_1\end{cases} = \begin{cases}O\left(1/n^{2k\cdot(\frac{k-1}{k}-\delta_1)}\right), 0\le\gamma<2/k,\\ O\left(\log n/n^{2k(\frac{k-1}{k}-\delta_1)}\right), \gamma = 2/k\\ O\left(1/n^{(k\epsilon_1+2)(\frac{k\epsilon_1}{k\epsilon_1+2}-\delta_1)}\right), 2/k<\gamma<\gamma_1\end{cases}\\
=&O(n^{-\eta})
\end{aligned}$$

Type 4: $d(v,s) > 2\rho_1$ **and** $\gamma_v > \gamma_2$ Similar to type 3, we have $\Pr[\mathcal{J}_v] = O\left(\left(\frac{\rho_1^2}{d(v,s)^\gamma}\right)^k\right)$, and

$$\begin{aligned}
\sum_{\substack{v\in V:\\ d(v,s)>2\rho_1,\\ \gamma_v>\gamma_2}} \Pr[\mathcal{J}_v] = &\sum_{\substack{v\in V:\\ d(v,s)>2\rho_1,\\ \gamma_v<\gamma_2}} O\left(\left(\frac{\rho_1^2}{d(v,s)^\gamma}\right)^k\right) = O\left(\int_{\rho_1}^{n}\left(\frac{\rho_1^2}{y^\gamma}\right)^k y dy\right)\\
=&O\left(\rho_1^{2k}\int_{\rho_1}^{n} y^{1-k\gamma} dy\right) = O\left(\rho_1^{2k}\rho_1^{2-k\gamma}\right) = O\left(1/n^{\delta_1(k\epsilon_2-2)}\right) = O(n^{-\eta})
\end{aligned}$$

□

APPENDIX B

Proofs for Preferential attachment models

B.1 Properties of the function $f_{k,\mathcal{R}}$

In this section, we will explore some relationships between the feedback function $f_{k,\mathcal{R}}$ and threshold distribution $\mathbb{R}$, monotonicity and degeneracy, which help us to state the result in Theorem 5.2.

Lemma B.1. *Given a distribution $\mathbb{R}$ and a constant $k \in \mathbb{N}$, the function $f_{k,\mathcal{R}}$ defined in (5.2) is a non-decreasing polynomial function from $[0,1]$ to $[0,1]$.*

Moreover, $f_{k,\mathcal{R}}(x) = x$ for all $x \in [0,1]$ if and only if $\Pr_{\mathbb{R}}(r) = \frac{1}{k}$ for $r = 1, \ldots, k$.

Proof. By Equation (5.2), $f_{k,\mathcal{R}}$ is a convex combination of a finite number of Bernstein polynomials $b_{\ell,k}(x) = \binom{k}{\ell} x^{\ell}(1-x)^{(k-\ell)}$, so $f_{k,\mathcal{R}}$ is a polynomial of degree at most k Because $\sum_{\ell=r}^{k} b_{\ell,k}(x)$ is non-decreasing in interval $[0,1]$, as a convex combination of $\{\sum_{\ell=k}^{m} b_{\ell,m}(x)\}_{k=0\ldots m}$, $f_{k,\mathcal{R}}$ is also non-decreasing.

For the second part, we can rewrite $f_{k,\mathcal{R}}$ as

$$f_{k,\mathcal{R}}(x) = \sum_{\ell=0}^{k} \Pr_{\mathbb{R}}[r \leq \ell] b_{\ell,k}(x).$$

Because Bernstein polynomials $\{b_{\ell,m} : 0 \leq \ell \leq m\}$ are linearly independent[1] and

$$\sum_{\ell=0}^{k} \frac{\ell}{k} b_{\ell,k}(x) = x$$

Therefore, $f_{k,\mathcal{R}}(x) = x$ if and only if $\Pr_{\mathbb{R}}[r \leq \ell] = \frac{\ell}{k}$ which finishes the proof. □

B.2 Proof of Directed case

B.2.1 Proof of Theorem 5.2 1a

Proof of Lemma 5.3. Recall that in Equation (5.5) the random variable U_{i+1} is bounded by 1 and the expectation is $E[U_{i+1} \mid \mathcal{F}_i] = 0$. Hence W_t defined in Equation (5.6) is

[1] That is given $a_\ell \in \mathbb{R}, \ell = 0, \ldots, m \forall x \in [0,1], \sum_{a_\ell} a_\ell b_{\ell,m}(x) = 0 \iff \forall \ell, a_\ell = 0$

a martingale. Applying Azuma-Hoeffding inequality in Theorem 2.18,

$$\Pr[|W_s - W_t| < \epsilon_1] \leq 2\exp\left(\frac{-\epsilon_1^2}{2\sum_{i=s}^{t} |\frac{1}{i}U_i|^2}\right)$$

However, $2\sum_{i=t+1}^{s} |\frac{1}{i}U_i|^2 \leq 2\sum_{i=t+1}^{s} \frac{1}{i^2} \leq \frac{2}{t} - \frac{2}{s} \leq \frac{2}{\tau_0}$ where the second inequality is true by an integration test. Therefore,

$$\Pr[|W_s - W_t| < \epsilon_1] \leq 2\exp(-\tau_0\epsilon_1^2/2))$$

which is smaller than δ_1 if $\tau_0 = \frac{2\ln(1/(2\delta_1))}{\epsilon_1^2}$. □

Proof of Lemma 5.4. By symmetry, we only need to show one side of the statement, $\Pr[Y_n \geq y^* - \epsilon] \leq \delta$. If $y^* - \epsilon \leq 0$, the statement is trivial, because $Y_n \geq 0$. If $y^* - \epsilon > 0$, 0 is not a fixed point and $g(y) > 0$ for all $y \in [0, y_L^*)$ because $f_{k,\mathcal{R}}(y) \in [0, 1]$. Hence, let $g_L = \min\{g(y) : y \in [0, y_L^* - \epsilon/2]\} > 0$ which is positive.

For the first part we want to show Y_t hits $(y_L^* - \epsilon/2, y_H^* + \epsilon/2)$ with small t. We set $\tau_0 = \frac{2\cdot 4}{\epsilon^2}\ln(1/(\delta))$, and $\tau_1 = \exp(2/g_L)\tau_0$. By Lemma 5.3, we have for all $s, t \geq \tau_0$, $|W_s - W_t| > \frac{\epsilon}{2}$ with probability less than $\delta/2$. Suppose $Y_{\tau_0} \leq y_L^* - \epsilon/2$, and $Y_t \notin (y_L^* - \epsilon/2, y_L^* + \epsilon/2)$ for all t between τ_0 and τ_1. Because the step size at time t is bounded by $1/t$ by (5.3), $\forall \tau_0 \leq t \leq \tau_1, Y_t \leq y_L^* - \epsilon/2$. By rearranging Equation (5.6),

$$\begin{aligned} W_{\tau_1} - W_{\tau_0} &= \sum_{k=\tau_0+1}^{\tau_1} \frac{1}{k} g(X_{k-1}) + Y_{\tau_0} - Y_t \geq \min_{\tau_0 < k \leq \tau_1} (g(X_{k-1}) \left(\sum_{k=\tau_0+1}^{\tau_1} \frac{1}{k}\right) - 1 \\ &\geq g_L \ln\left(\frac{\tau_1}{\tau_0}\right) - 1 = 1 > \frac{\epsilon}{2} \end{aligned}$$

Therefore,

$$\Pr\left[\exists \tau_0 \leq t \leq \tau_1, Y_t \in \left(y_L^* - \frac{\epsilon}{2}, y_H^* + \frac{\epsilon}{2}\right)\right] \geq 1 - \frac{\delta}{2}. \tag{B.1}$$

For the second part, if $Y_{\sigma_0} \in (y_L^* - \epsilon/2, y_H^* + \epsilon/2)$, at time $\tau_0 \leq \sigma_0 \leq \tau_1$, we want to show it will stay in the interval $(y_L^* - \epsilon, y_H^* + \epsilon)$. Suppose there exists $\tau_3 > \sigma_1$ such that, $Y_{\tau_3} < y_L^* - \epsilon$. Then there exist σ_1 and σ_2 such that $\tau_0 \leq \sigma_0 \leq \sigma_1 \leq \sigma_2 \leq \tau_3$,

$$Y_{\sigma_1} > y_L^* - \frac{\epsilon}{2}, \text{ and } Y_{\sigma_2} < y_L^* - \epsilon \tag{B.2}$$

$$\forall \sigma_1 \leq t < \sigma_2, y_L^* - \epsilon \leq Y_t \leq y_H^* - \frac{\epsilon}{2}. \tag{B.3}$$

Using similar argument,

$$W_{\sigma_1} - W_{\sigma_2} = Y_{\sigma_1} - Y_{\sigma_2} + \sum_{k=\sigma_1}^{\sigma_2} \frac{1}{k} g(X_{k-1}) \geq Y_{\sigma_1} - Y_{\sigma_2} \geq \frac{\epsilon}{2}$$

Because (B.3) $g(Y_t) \geq g_L > 0$, the first inequality holds. The second inequality is

true by (B.2). Therefore by union bound, there exists $\tau = O\left(\frac{1}{\epsilon^2}\ln(1/\delta)\right)$ such that

$$\Pr\left[\forall t \geq \tau, Y_t \in [y_L^* - \epsilon, y_H^* + \epsilon]\right] \geq 1 - \delta.$$

□

B.2.2 Proof of Theorem 5.2 2

In this case, $f_{k,\mathcal{R}}$ has multiple stable fixed points, and the process (5.3) is indeed a special case of stochastic approximation algorithms defined in Definition 2.5 such that

1. $\gamma_{i+1} = 1/(i+1)$,
2. $U_{i+1} = \frac{1}{2}(\mathrm{Bin}(k, Y_i)/k - Y_i + \mathrm{Bin}(1, f_{k,\mathcal{R}}(Y_i)) - f_{k,\mathcal{R}}(Y_i))$ is a martingale with $K_u = 4k$ and $E[\gamma_{i+1}U_{i+1} \mid \mathcal{F}_i] = 0$,
3. $g(Y_i) = E[(\mathrm{Bin}(k, Y_i) + k \cdot \mathrm{Bin}(1, f_{k,\mathcal{R}}(Y_i)) - 2kY_i) \mid Y_i]/2k = (f_{k,\mathcal{R}}(Y_i) - Y_i)/2$ is bounded by $K_g = 1$,
4. $|E[\gamma_{n+1}U_{n+1} \mid \mathcal{F}_n]| = |E[\frac{U_{i+1}}{i+1}]| = 0 \leq K_e\gamma_n^2$, where $K_e = 1$.

To prove this convergence property, we apply Theorem 2.6.

Proof. The first statement is a result of Theorem 2.6 (1) because g is a polynomial. The second is a result of Theorem 2.6 (1) and 2.6 (2). However to apply Theorem 2.6 (2) we have to prove $E[U_{i+1}^2]$ is bounded below by constant K_L which in our case is sufficient to prove the variance of $\mathrm{Bin}(k, Y_i) - kY_i + k\,\mathrm{Bin}(1, g(Y_i) - kg(Y_i)$ is nonzero when $0 < |Y_i - p| < d$. Formally,

$$\begin{aligned}
&\mathrm{Var}(\mathrm{Bin}(k, Y_i) - kY_i + k\,\mathrm{Bin}(1, g(Y_i) - kg(Y_i)) \\
=&\,\mathrm{Var}(\mathrm{Bin}(k, Y_i)) + \mathrm{Var}(k\,\mathrm{Bin}(1, g(Y_i)) + 2\,\mathrm{Cov}(\mathrm{Bin}(k, Y_i), k\,\mathrm{Bin}(1, g(Y_i)) \\
\geq&\, kY_i(1 - Y_i) + k^2 g(Y_i)(1 - g(Y_i)) > 0
\end{aligned}$$

The last inequality comes from $\mathrm{Cov}(\mathrm{Bin}(k, Y_i), k\,\mathrm{Bin}(1, g(Y_i)) \geq 0$ by FKG inequality. Finally, Theorem 2.6 (3) and 2.6 (4) show that Y_i will converge to an arbitrary stable or touch point with positive probability. □

B.2.3 Proof of Theorem 5.2 3

In the special case when $f_{k,\mathcal{R}}$ has an infinite number of fixed points, because $f_{k,\mathcal{R}}$ is a polynomial with degree at most m, we have $f_{k,\mathcal{R}}(x) = x$ by Fundamental Theory of Algebra. As a result, the predictable part $g(x) = 0$ in (5.4) and U_i is a martingale difference such that $\mathbb{E}[U_{i+1} \mid \mathcal{F}_i] = 0$ and $|U_{i+1} \mid \mathcal{F}_i| \leq 1$. Therefore our random process $\{Y_i\}$ is the martingale $Y_i = \sum_{\ell=m+1}^{i} \frac{1}{\ell}\{U_\ell\}$. To prove the convergence of martingale $\{Y_i\}$ we can use standard martingale convergence theorem (c.f. Theorem 1 in Chapter 7.8 in [107]) to prove convergence. Because $\mathbb{E}[Y_i^2] = \sum_{\ell=m+1}^{i} \frac{1}{\ell^2}|U_\ell|^2 \leq$

$\sum_{\ell=m+1}^{i} \frac{1}{\ell^2} < \infty$ for all i, there exists a random variable Y such that Y_i converges to Y almost surely.

B.3 Proofs of Undirected Case

B.3.1 Phase 1: first forward and backward contagion

Proof. The intuition is that if $M_y^F(\tau_1, t-1)$ happens, Y_{t-1} is large and the node t is likely to be infected. Our goal is to argue that even when conditioned on $\mathcal{F}_{t-1}, M_y^F(t-1, \tau_2), N_k^t$, the probability node t being infected is still large.

First, we use the Markov property to argue that $\mathcal{F}_{t-1}$ does not affect the probability after conditioning on Y_{t-1}. Second we use positive correlation to remove condition $M_y^F(t-1, \tau_2)$. Finally, we use independence among edges of node t to show that conditioning on $N(t,s)$ will not affect the probability too much.

Note that I_t^F only depends on Y_t^F, we have

$$
\begin{aligned}
&\Pr\left[I_t^F \mid \mathcal{F}_{t-1}, M_y^F(\tau_1, \tau_2), N(t,s)\right] \\
\geq &\Pr[I_t^F \mid Y_{t-1} > y, M_y^F(t-1, \tau_2), N(t,s), \neg I_s^F] \\
\geq &\Pr[I_t^F \mid (Y_{t-1} > y), N_k^t, \neg I_s^F] \\
&\qquad \text{(positively correlated between } I_t^F \text{ and } M_y^F(t-1,\tau_2)) \\
= &\sum_r \Pr[R_t = r \mid (Y_{t-1} > y), N(t,s), \neg I_s^F] \Pr[I_t^F \mid (Y_{t-1} > y), N_k^t, \neg I_s^F, R_t = r].
\end{aligned}
$$

Note that for all r $\Pr[R_t = r | (Y_{t-1} > y), N_k^t, I_k] = \Pr[R_t = r | (Y_{t-1} > y)]$ and it is greater than $\Pr[R_t = r] - \Pr[\neg(Y_{t-1} > y)]$. Moreover, the second term is very small.

Now we want to have a lower bound for $\Pr[I_t^F \mid (Y_{t-1} > y), N_k^t, \neg I_s^F, R_t = r]$. Because all k edges of node t are added independently, there are three different outcomes of each edge:

- Connected to an infected node with probability Y_{t-1};
- Connected to node s with probability $(1 - Y_{t-1})q$;
- Connected to an uninfected node except node s with probability $(1-Y_{t-1})(1-q)$,

where $q > 0$ is the probability that an edge goes to node s conditioned on the edge not connected to infected nodes.

Therefore the distribution is a multinomial distribution. Let X_1 be the number of edges that land on infected nodes, X_2 be the number of edges that land on node s, and X_3 be the number of edges that land on uninfected nodes other than s. With

these notation,

$$\begin{aligned}
&\Pr[I_t^F|(Y_{t-1}>y), N(t,s), \neg I_s^F, R_t=r] = \Pr[X_1 \geq r | X_2 > 0] \\
=&\frac{\sum_{\ell \geq r} \binom{k}{\ell} Y_{t-1}^{\ell}(1-Y_{t-1})^{k-\ell}\left(1-(1-q)^{k-\ell}\right)}{1-(1-(1-Y_{t-1})q)^k} \qquad (\ell \text{ of } k \text{ edges to an infected node}) \\
\geq &\min_{\ell<k} \frac{1-(1-q)^{k-\ell}}{1-(1-(1-Y_{t-1})q)^k} \cdot \sum_{\ell \geq s} \binom{k}{\ell} Y_{t-1}^{\ell}(1-Y_{t-1})^{k-\ell}
\end{aligned}$$

Because $1-(1-(1-Y_{t-1})q)^k \leq 1-(1-k(1-Y_{t-1})q) = k(1-Y_{t-1})q$, and $1-(1-q)^{k-\ell} \geq q$. The first term can be lower bounded as following:

$$\frac{1-(1-q)^{m-\ell}}{1-(1-(1-Y_{t-1})q)^m} \geq \frac{q}{m(1-Y_{t-1})q} = \frac{1}{m(1-Y_{t-1})}. \tag{B.4}$$

Moreover, for the second term $\sum_{\ell=s}^{m-1} \binom{m}{\ell} Y_{t-1}^{\ell}(1-Y_{t-1})^{m-\ell}$ is non-decreasing with respect to Y_{t-1} when it is between $[0,1]$, $Y_{t-1} > y$, so

$$\sum_{\ell=s}^{m-1} \binom{m}{\ell} Y_{t-1}^{\ell}(1-Y_{t-1})^{m-\ell} \geq \sum_{\ell=s}^{m-1} \binom{m}{\ell} y^{\ell}(1-y)^{m-\ell} \tag{B.5}$$

Combining Equations (B.4) and (B.5) we get

$$\Pr[I_t|\mathcal{F}_{t-1}, M_y^F(\tau_1,\tau_2), N(t,s), R_t=r] \geq \frac{1}{m(1-y)} \cdot \sum_{\ell=s}^{m-1} \binom{m}{\ell} y^{\ell}(1-y)^{m-\ell},$$

and have $\Pr[I_t^F|(Y_{t-1}>y), N(t,s), \neg I_s^F, R_t=r]$ is greater than

$$\frac{1}{m(1-y)} \sum_{r \geq 0} \Pr[R_t=r] \sum_{\ell=r}^{k-1} \binom{k}{\ell} y^{\ell}(1-y)^{(k-\ell)} - k \Pr[\neg M_y^F(\tau_1,\tau_2)]$$

Let $q^* = \frac{1}{2k(1-y_L^*)} \sum_{r \geq 0} \Pr[R_t=r] \sum_{\ell=r}^{k-1} \binom{k}{\ell} (y_L^*)^{\ell}(1-y_L^*)^{(k-\ell)}$. By Corollary 5.6, we can take $\epsilon_0 = y_L^* - y$ and $\delta_0 = q^*/(100k)$, and there exists some constant $\tau_0 = O\left(\frac{\log 1/\delta_0}{\epsilon_0^2}\right)$ such that $\tau_1 > \tau_0$ and $\Pr[M_y^F(\tau_1,\tau_2)] \leq \Pr[M_y^F(\tau_1,n)] \leq q^*/(100k)$ and complete the proof. □

Proof of Lemma 5.9. We first prove that a node $s < \tau_1$ will get infected with high probability and then use the union bound. Given nodes s and t such that $s < \tau_1$ and $s < t$, let $N(t,s)$ be the indicator that node t is a neighbor of s, and $R_H = \max\{r : \Pr[R_u = r] \neq 0\}$, which is the maximum of the support of threshold distribution. Then for node s, the random variable of the number of infected neighbor at time τ_2 can be written as $\sum_{t>s}^{\tau_2} I_t^F N(t,s)$. If this value is greater than the maximum possible threshold K, s will be infected in the first backward infection.

First we show that with additional condition $M_y^F(\tau_1,\tau_2)$ the probability $\Pr[I_s^{FB}]$

is not affected too much

$$\begin{aligned}
\Pr[I_s^{FB}=1] &\geq \Pr\left[\sum_{t>s}^{\tau_2} I_s^F \cdot N(t,s) \geq R_H\right] \geq \Pr\left[\sum_{t>\tau_1}^{\tau_2} I_t^F \cdot N(t,s) \geq R_H\right] \quad (k \leq \tau_1) \\
&\geq \Pr\left[\sum_{t>\tau_1}^{\tau_2} I_t^F N(t,s) \geq R_H \mid M_y^F(\tau_1,\tau_2)\right] - \Pr[\neg M_y^F(\tau_1,\tau_2)] \\
&\geq \Pr\left[\sum_{t>\tau_1}^{\tau_2} I_t^F \cdot N(t,s) \geq R_H \mid M_y^F(\tau_1,\tau_2)\right] - \frac{1}{n} \qquad \text{(by Corollary 5.6)}
\end{aligned}$$

With this condition $M_y^F(\tau_1,\tau_2)$ and Lemma 5.7, the following proof will show that at time τ_2 the node s will, with high probability, both have 1) degree $\deg(s)$ higher than $d = 2\log\log n$ and 2) have more than R_H infected neighbors.

$$\begin{aligned}
&\Pr[I_s^{FB}] \geq \Pr\left[\sum_{t>\tau_1}^{\tau_2} I_t^F \cdot N(t,s) \geq R_H \mid M_y^F(\tau_1,\tau_2)\right] - \frac{1}{n} \\
\geq &\Pr\left[\deg(s;\tau_2) - \deg(s;\tau_1) > d \wedge \sum_{t>\tau_1}^{\tau_2} I_t^F \cdot N(t,s) \geq R_H \mid M_y^F(\tau_1,\tau_2)\right] - \frac{1}{n} \\
\geq &1 - \Pr\left[\sum_{t>\tau_1}^{\tau_2} I_t^F \cdot N(t,s) < R_H \mid \deg(s;\tau_2) - \deg(s;\tau_1) > d, M_y^F(\tau_1,\tau_2)\right] \\
&- \Pr[\deg(s;\tau_2) - \deg(s;\tau_1) \leq d] - \frac{2}{n}
\end{aligned}$$

If the degree of node $s < \tau_1$ at time τ_2 is at least d, $\deg(s;\tau_2) - \deg(s;\tau_1) \geq d$, there are at least $d/k - 1$ neighbors at time t, because each neighbor of node s, t, contributes the degree of node s at most k. On the other hand, by Lemma 5.7 the probability of any neighbor of node k gets infected is lower bounded by a constant p^* which is independent of the history. Thus we can couple each neighbor of node s getting infected or not with a Bernoulli trial which succeeds with probability p^* which is independent of history. Therefore the probability that the node s with $d/k - 1$ neighbors has more than R_H infected neighbors is bounded below by $\Pr[\mathrm{Bin}(d/k - 1, p^*) < R_H]$. Moreover,

$$\begin{aligned}
&\Pr[I_t^{FB} \mid M_y^F(\tau_1,\tau_2), \deg(s;\tau_2) - \deg(s;\tau_1) > d] \\
\geq &1 - \Pr[\mathrm{Bin}(d/k-1,p^*) < R_H] - \Pr[\deg(s;\tau_2) - \deg(s;\tau_1) \leq d] - \frac{1}{n} \\
\geq &1 - \frac{1}{(\log n)^2} \qquad \text{(by Lemma 5.8)}
\end{aligned}$$

We apply union bound over all nodes with index less than $\tau_1 = C_1 \log n$:

$$\Pr[Y^{FB}_{\tau_1} = 1] > 1 - \frac{C_1}{\log n} > 1 - \epsilon.$$

□

B.3.2 Phase 3: Constant separation

Proof of Lemma 5.11. Because for nodes $t > \tau_2$ nothing is revealed until the last phase, I^{FBF}_t, I^F_t have the same distribution. Therefore given that node $s < \tau_2$ is not infected at time τ_2, we can use coupling and the same argument in Lemma 5.9,

$$\begin{aligned}\mathbb{E}[I^{FBFB}_s] \geq& \Pr\left[\sum_{t>\tau_2}^{n} I^{FBF}_t \cdot N(t,s) > R_H \mid M^{FBF}_y(\tau_2, n)\right] - o(1) \quad \text{(for all } y < y^*_H\text{)} \\ =& \Pr\left[\sum_{t>\tau_2}^{n} I^{F}_t \cdot N(t,s) > R_H \mid M^{FBF}_y(\tau_2, n)\right] - o(1).\end{aligned}$$

To have event $\sum_{t>\tau_2}^{n} I^F_t \cdot N(t,s) > R_H$, it is sufficient that 1) the total degree of node s is greater than $d = \frac{7mR_H}{p^*}$ and 2) there are R_H infected neighbor among the first d neighbor of s. Let $\deg(s;t)$ is the degree of s at time $t > s$. By Lemma 2.52, for all $s \leq \tau_2 \leq e^{-2}n$ and $d \geq m$, the number of neighbor in τ_2 to n is $\deg(s;n) - \deg(s;\tau_2) \geq \deg(\tau_2;n) = \deg(\tau_2)$. That is

$$\Pr[\deg(s;n) - \deg(s;\tau_2) < d] \leq \binom{d-1}{k-1} \left(\frac{\tau_2}{n}\right)^{k/2} \ln\left(\frac{kn+1}{k\tau_2+1}\right)^d \leq \frac{1}{4} \tag{B.6}$$

by taking constant r small enough.

By argument similar to Lemma 5.7 each t will be infected with probability at least p^* defined the statement, so

$$\begin{aligned}&\Pr[I^{FBFB}_s \mid M^{FBF}_y(\tau_2, n), \deg(s;n) - \deg(s;\tau_2) > d] \\ \geq& 1 - \Pr[\mathrm{Bin}(d/k - 1, p^*) < R_H] - \Pr[\deg(s;n) - \deg(s;\tau_2) \leq d] - o(1) \\ \geq& \frac{3}{4} - \Pr[\mathrm{Bin}(6R_H/p^*, p^*) < R_H] - \Pr[\deg(s;n) - \deg(s;\tau_2) \leq d] \qquad \text{(by (B.6))} \\ \geq& \frac{3}{4} - \frac{1}{6} \qquad \text{(by Markov inequality)}\end{aligned}$$

Hence we get $\Pr[I^{FBFB}_s] > 1/4$.

For the moreover part, if $\deg(s;\tau_2) = d$, the expected number of neighbors from τ_2 to n is proportional to d, by Markov inequality we have $\Pr[\deg(s;n) - \deg(s;\tau_2) \leq \sqrt{d}] \leq O(1/\sqrt{d})$ and $\Pr[I^{FBFB}_s \mid M^{FBF}_y(\tau_2, n), \deg(s;n) - \deg(s;\tau_2) \geq \sqrt{d}] \geq 1 - 1/\sqrt{d}$ which completes the proof. □

Proof. Here we are going to show that I^{FBFB}_s and I^{FBFB}_t are negatively dependent,

i.e.,

$$\Pr[I_s^{FBFB} = 1 \mid I_t^{FBFB} = 1] \leq \Pr[I_s^{FBFB} = 1].$$

Suppose nodes s, t are not infected in the forward infection, if we reveal all the edges that do not go to either s or t as condition $\mathcal{H}_{s,t}$, then the result of forward infection is revealed since the edges that go to nodes s, t do not affect the infection. We will show if $I_t^{FBFB} = 1$ which t is infected, s has fewer edges than s has without conditioning on $I_t^{FBFB} = 1$.

Formally, conditioned on $\mathcal{H}_{s,t}$ given $s < t$ the set of edges go to either s or t is fixed. We set $N = \{e_1, \ldots, e_l\}$ to be the set of edges, and define an order on the edges in N based on their arriving order.[2] Under this order, given the ℓ-th edge in N e_ℓ we consider the (random) sets $N(\ell)$,$N_1(s;\ell)$,$N_2(s;\ell)$, $N_1(t;\ell)$, and $N_2(t;\ell)$. $N(\ell)$ is the set of the edges before edge $e_{\ell+1}$ that go to node s or t. $N_1(s;\ell)$ ($N_1(s;\ell)$) is the set of edges goes to s (t) given $\mathcal{H}_{s,t}$ and $I_t^{FBFB} = 1$, and $N_2(s;\ell)$ ($N_2(t;\ell)$) is the set given $\mathcal{H}_{s,t}$. Note that given $\mathcal{H}_{s,t}$, the set of edges go to either s or t is fixed, that is

$$N(\ell) = N_1(s;\ell) \cup N_1(t;\ell) = N_2(s;\ell) \cup N_2(t;\ell).$$

With these notions, our goal is to prove for any u the cardinality of $N_2(s;\ell)$ stochastic dominates the cardinality of $N_1(s;\ell)$. We will show this by constructing a coupling $(\hat{N}_1(s;\ell), \hat{N}_2(s;\ell))$ between $N_1(s;\ell)$ and $N_2(s;\ell)$ such that $\hat{N}_1(s;\ell) \subseteq \hat{N}_2(s;\ell)$ and the marginal distribution $\hat{N}_1(s;\ell)$ ($\hat{N}_2(s;\ell)$) is equal to $N_1(s;\ell)$ ($N_2(s;\ell)$).

Now we define the coupling by induction. Given $e_\ell \in N(u)$, we define the event $E(s,\ell)$ such that the edge e_ℓ goes to s, and $\neg E(s,\ell)$ would be the event when the edge e_ℓ goes to t.

For the base case, when edge e_1 reveals its connection, the degree of s and t are the same, so we have $\Pr[E(s,1) \mid I_j^{FBFB} = 1, \mathcal{H}_{s,t}] \leq \Pr[E(s,1) \mid \mathcal{H}_{s,t}]$. Therefore, we can construct a coupling such that

$$\hat{N}_1(s;1) \subseteq \hat{N}_2(s;1).$$

For the inductive steps, suppose $\hat{N}_1(s;\ell) \subseteq \hat{N}_2(s;\ell)$. For the edge $e_{\ell+1}$, because of the positive feedback property of preferential attachment graph,

$$\Pr[E(s;\ell+1) \mid \hat{N}_1(s;\ell), \mathcal{H}_{s,t}] \leq \Pr[E(s;\ell+1) \mid \hat{N}_2(s;\ell), \mathcal{H}_{s,t}], \tag{B.7}$$

On the other hand, the more edges j has, the higher chance $I_j^{FBFB} = 1$:

$$\Pr[I_t^{FBFB} = 1 \mid E(s;\ell+1), \hat{N}_1(s;\ell), \mathcal{H}_{s,t}] \leq \Pr[I_t^{FBFB} = 1 \mid \neg E(s;\ell+1), \hat{N}_1(s;\ell), \mathcal{H}_{s,t}].$$

Therefore

$$\Pr[E(s;\ell+1) \mid I_t^{FBFB} = 1, \hat{N}_1(s;\ell), \mathcal{H}_{s,t}] \leq \Pr[E(s;\ell+1) \mid \hat{N}_1(s;\ell), \mathcal{H}_{s,t}], \tag{B.8}$$

[2]To compare two edges, we first compare their oldest endpoints, and break tie arbitrarily.

Combining inequalities (B.7) and (B.8) we have,

$$\Pr[E(s;\ell+1) \mid I_t^{FBFB}=1, \hat{N}_1(s;\ell), \mathcal{H}_{s,t}] \le \Pr[E(s;\ell+1) \mid \hat{N}_2(s;\ell), \mathcal{H}_{s,t}],$$

so we can have

$$\hat{N}_1(s;\ell+1) \subseteq \hat{N}_2(s;\ell+1).$$

Therefore we complete the proof. □

Proof of Lemma 5.13. Firstly, by Lemma 5.11, the expectation is large. Formally, if $Y_{\tau_2}^{FBF} < \frac{1}{2}(1+y_H^*)$, we know there are constant fraction of degree at time τ_2 being uninfected after the second forward infection,

$$\sum_{s\in U^{FBF}: s\le\tau_2} \deg(s;\tau_2) = 2k\tau_2(1-Y_{\tau_2}^{FBF}) \ge k\tau_2(1-y_H^*). \tag{B.9}$$

By Lemma 5.11 and linearity of expectation we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{s\le\tau_2} \deg(s;\tau_2) I_s^{FBFB} \mid M_y^{FBF}(\tau_2,n)\right] \\
=&\sum_{s\in U^{FBF}: s\le\tau_2} \deg(s;\tau_2)\mathbb{E}\left[I_s^{FBFB} \mid M_y^{FBF}(\tau_2,n)\right] && (\deg(s;\tau_2) \text{ are deterministic}) \\
\ge&\frac{1}{4}\sum_{s\in U^{FBF}: s\le\tau_2} \deg(s;\tau_2) && (\text{Lemma 5.11}) \\
\ge&\frac{1}{4}k\tau_2(1-y_H^*) = \frac{kC_2(1-y_H^*)}{4}n && (\text{by (B.9)})
\end{aligned}$$

Second, we need to upper bound the variance. By Lemma 5.12, the indicator functions I_s^{FBFB} are negative correlated, and the variance can be upper bounded by sum of individual variance. That is

$$\begin{aligned}
&\operatorname{Var}\left[\sum_{s\le\tau_2} \deg(s;\tau_2) I_s^{FBFB} \mid M_y^{FBF}(\tau_2,n)\right] \\
\le&\sum_{s\in U^{FBF}: s\le\tau_2} \deg(s;\tau_2)^2 \operatorname{Var}\left[I_s^{FBFB} \mid M_y^{FBF}(\tau_2,n)\right] \\
\le&\sum_{s\in U^{FBF}: s\le\tau_2} \deg(s;\tau_2)^2\mathbb{E}\left[I_s^{FBFB} \mid M_y^{FBF}(\tau_2,n)\right] && (\text{indicator function}) \\
=&O\left(\sum_{s\in U^{FBF}: s\le\tau_2} \deg(s;\tau_2)^2\frac{1}{\sqrt{\deg(s;\tau_2)}}\right) && (\text{Lemma 5.11}) \\
=&O\left(\sum_{s\in U^{FBF}: s\le\tau_2} \deg(s;\tau_2)\sqrt{\deg(s;\tau_2)}\right) = O(n\sqrt{n})
\end{aligned}$$

Therefore applying the second moment method for all $\Delta < \frac{kC_2(1-y_H^*)}{4}$ we have

$$\Pr\left[\sum_{s\le\tau_2}\deg(s;\tau_2)I_s^{FBFB} \ge \Delta n \mid M_y^{FBF}(\tau_2,n)\right] \ge 1 - O\left(\frac{n\sqrt{n}}{(\frac{1}{4}(kC_2(1-y_H^*)) - \Delta)^2 n^2}\right)$$

which is greater than $1 - O(1/\sqrt{n})$. □

B.4 Proof for Directed Stochastic Attachment Model

Proof of the Theorem 5.16. We first show Part (1).

Let $D(Y_i, m_{i+1}) = (-2m_{i+1}Y_i + \mathrm{Bin}(m_{i+1}, Y_i) + m_i \,\mathrm{Bin}(1, f_0^{m_{i+1}}(Y_i)))$. We have

$$\begin{aligned}
&Y_{i+1} - Y_i \mid \mathcal{F}_i = \frac{1}{d_{i+1}}\left(g_0(Y_i) + U_{i+1}\right)\\
&g_0(y) = \mathbb{E}[D(y,m) \mid y] = -2\mu_m y + \mu_m y + \mathbb{E}[m f_0^m(y)]\\
&U_{i+1} = D(Y_i, m_{i+1}) - g_0(Y_i) = D(Y_i, m_{i+1}) - \mathbb{E}[D(Y_i, m_{i+1})|Y_i]
\end{aligned}$$

Now we check the conditions in Definition 2.5

1. Because $i \le d_i \le c_h i$, $\frac{1}{c_h i} \le \gamma_i = 1/d_i \le \frac{1}{i}$.
2. $|U_i| \le 4m_{i+1} \le 4c_h$.
3. $|g(Y_i)| \le 4c_h$
4. $|\mathbb{E}[\frac{1}{d_{i+1}}U_{i+1}|\mathcal{F}_i]| \le K_e/d_i^2$

Here we prove Property (4).

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{d_{i+1}}U_{i+1}|\mathcal{F}_i\right] = \mathbb{E}\left[\frac{1}{d_{i+1}}(D(Y_i,m_{i+1})\right] - \mathbb{E}\left[D(Y_i,m_{i+1})|Y_i]\right)|\mathcal{F}_i]\\
=&\mathbb{E}\left[\frac{1}{d_{i+1}}D(Y_i,m_{i+1})|\mathcal{F}_i\right] - \mathbb{E}\left[\frac{1}{d_{i+1}}|\mathcal{F}_i\right]\mathbb{E}[D(Y_i,m_{i+1})|\mathcal{F}_i] &\text{(B.10)}\\
=&\,\mathrm{Cov}\left[\frac{1}{d_{i+1}}, D(Y_i,m_{i+1})|\mathcal{F}_i\right] &\text{(B.11)}\\
\le&\sqrt{\mathrm{Var}(\frac{1}{d_{i+1}}|\mathcal{F}_i)\cdot\mathrm{Var}\left[D(Y_i,m_{i+1})|\mathcal{F}_i\right]} &\text{(B.12)}
\end{aligned}$$

Equation B.10 is true by linearity of expectation and the tower property for conditional expectations. By the definition of covariance we have Equation (B.11), and Equation (B.12) comes from the Cauchy inequality.

Consider a random variable X restricted to $[a,b]$, then $\mathbb{E}[X]$ is also in $[a,b]$ and $|X - \mathbb{E}[X]| \le (a-b)$. Hence $\mathrm{Var}[X] = \mathbb{E}[(X-\mu_X)^2] \le (a-b)^2$. By definition $|Y_i| \le 1$

and $1 \le |m_{i+1}| \le c_h$. Because $0 \le D(Y_i, m_{i+1}) \le |-2m_{i+1}Y_i + \mathrm{Bin}(m_{i+1}, Y_i) + m_i \,\mathrm{Bin}(1, f_0^{m_{i+1}}(Y_i))| \le 4|m_{i+1}| \le 4c_h$ is a bounded random variable:

$$\sqrt{\mathrm{Var}(D(Y_i, m_{i+1}))} \le 4c_h$$

On the other hand, we have to analyze $\mathrm{Var}(\frac{1}{d_{i+1}}|\mathcal{F}_i)$ carefully. Because $\frac{1}{d_{i+1}}|\mathcal{F}_i$ is restricted to $[\frac{1}{d_i+2c_h}, \frac{1}{d_i+2}]$,

$$\sqrt{\mathrm{Var}\left(\frac{1}{d_{i+1}}|\mathcal{F}_i\right)} \le \left(\frac{1}{d_i+2c_h} - \frac{1}{d_i+2}\right) \le \frac{2c_h}{d_i^2}$$

Property (4) is true by taking large enough constant $K_e \ge 8c_h^2$,

$$\left|\mathbb{E}\left[\frac{1}{d_{i+1}}U_{i+1}|\mathcal{F}_i\right]\right| \le 8c_h^2/d_i^2 \le K_e/d_i^2$$

Therefore Y_i converges by applying Theorem 2.6.

For the random uniform attachment, part (2), the proof is similar. □

APPENDIX C

Proofs for Blockmodel

C.1 Full Proofs

C.1.1 Proof of Lemma 6.13

Proposition C.1. *Suppose the root of T has weight $\omega(1/n^{1+1/r})$ and consider a leaf t. If there are $\Theta(n)$ infected vertices in $V \setminus V(t)$, then these infected vertices outside $V(t)$ will infect $\omega(1)$ vertices in $V(t)$ with probability $1 - o(1)$.*

Proof. Let $X = \Theta(n)$ be the number of infected vertices in $V \setminus V(t)$. For each $u \in V(t)$ and $v \in V \setminus V(t)$, we assume that the probability p_{uv} that the edge (u, v) appears satisfies $p_{uv} = \omega(1/n^{1+1/r})$ and $p_{uv} = o(1/n)$, where $p_{uv} = \omega(1/n^{1+1/r})$ holds since the root of T has weight $\omega(1/n^{1+1/r})$, and assuming $p_{uv} = o(1/n)$ may only decrease the number of infected vertices in $V(t)$ if the least common ancestor of the two leaves containing u and v has weight $\Omega(1/n)$. Let p be the minimum probability among those p_{uv}'s, and we further assume that each edge (u, v) appears with probability p, which again may only reduce the number of infected vertices in $V(t)$.

For each vertex $u \in V(t)$, by only accounting the probability that it has exactly r neighbors among those X outside infected vertices, the probability that u is infected is at least

$$\rho := \binom{X}{r} p^r (1-p)^{X-r} = \omega\left(n^r \cdot \left(\frac{1}{n^{1+1/r}}\right)^r \left(1 - \frac{1}{n}\right)^n\right) = \omega\left(\frac{1}{n}\right),$$

and the expected number of infected vertices in $V(t)$ is $v(t)n \cdot \rho = \omega(1)$.

Let Y be the number of vertices in $V(t)$ that are infected due to the influence of $V \setminus V(t)$, so we have $\mathbb{E}[Y] = v(t)n\rho$. Applying Chebyshev's inequality,

$$\Pr\left(Y \le \frac{1}{2}v(t)n\rho\right) \le \Pr\left(|Y - \mathbb{E}[Y]| \ge \frac{1}{2}v(t)n\rho\right)$$

$$\le \frac{\mathrm{Var}(Y)}{(\frac{1}{2}v(t)n\rho)^2} = \frac{v(t)n\rho(1-\rho)}{\frac{1}{4}v(t)^2n^2\rho^2} = o(1),$$

where we have used the fact that $n\rho = \omega(1)$ and the variance of the Binomial random variable with parameter n, p is $np(1-p)$. Therefore, with probability $1 - o(1)$, the

number of infected vertices in $V(t)$ is at least $\frac{1}{2}v(t)n\rho = \omega(1)$. □

Let E be the event that at least one leaf (or node) is activated at the end of the cascade. By our definition, $P_{\boldsymbol{k}} = \lim_{n\to\infty} \Pr(E)$. Given a seeding strategy $\boldsymbol{k}$, let $\sigma(\boldsymbol{k}) := \mathbb{E}_{G\sim\mathcal{G}(n,T)}[\sigma_{r,G}(\boldsymbol{k})]$ be the expected number of infected vertices, $\sigma(\boldsymbol{k} \mid E) := \mathbb{E}_{G\sim\mathcal{G}(n,T)}[\sigma_{r,G}(\boldsymbol{k}) \mid E]$ be the expected number of infected vertices conditioning on event E happens, and $\sigma(\boldsymbol{k} \mid \neg E) := \mathbb{E}_{G\sim\mathcal{G}(n,T)}[\sigma_{r,G}(\boldsymbol{k}) \mid \neg E]$ be the expected number of infected vertices conditioning on E does not happen. We have

$$\sigma(\boldsymbol{k}) = \Pr(E) \cdot \sigma(\boldsymbol{k} \mid E) + (1 - \Pr(E)) \cdot \sigma(\boldsymbol{k} \mid \neg E),$$

and

$$\begin{aligned}\Sigma_{r,T}(\boldsymbol{k}) &= \lim_{n\to\infty} \frac{\sigma(\boldsymbol{k})}{n} \\ &= P_{\boldsymbol{k}} \cdot \lim_{n\to\infty} \frac{\sigma(\boldsymbol{k} \mid E)}{n} + (1 - P_{\boldsymbol{k}}) \cdot \lim_{n\to\infty} \frac{\sigma(\boldsymbol{k} \mid \neg E)}{n}. \end{aligned} \tag{C.1}$$

First, we show that $\sigma(\boldsymbol{k} \mid \neg E) = o(n)$, so the second term in (C.1) is always 0. If there is no critical or supercritical leaf in T, given that the total number of seeds $K = \Theta(1)$ is a constant, Theorem 6.6 suggests that there can be at most $2K = \Theta(1)$ infected vertices. To be specific, we can take the maximum weight $w^*(t)$ over all the leaves, and assume the entire graph is the Erdős-Rényi graph $\mathcal{G}(n, w^*(t))$. This makes the graph denser, so the expected number of infected vertices increases. However, even in this case, Theorem 6.6 implies that the total number of infected vertices is less than $2K$. If there is at least one critical or supercritical leaf t, for the sake of contradiction we assume the total number of infected vertices is $\Theta(n)$ and E does not happen. Since the number of leaves is a constant, there exists $t' \in L_T$ such that the number of infected vertices in $V(t')$ is $\Theta(n)$. Theorem 6.7 and Corollary 6.9 indicate that, with probability $1 - o(1)$, the number of infected vertices in $V(t)$ is either a constant or $v(t)n$. Therefore, if $t' = t$, with probability $o(1)$, those $\Theta(n)$ infected vertices in $V(t)$ will not activate t, and this probability becomes 0 after taking the limit $n \to \infty$, which makes no contribution to the second term in (C.1). If $t' \neq t$, let $X = \Theta(n)$ be such that with probability $1 - o(1)$ the number of infected vertices in $V(t')$ is more than X, then the total number of vertices in $V(t)$ that are infected by those X vertices in $V(t')$ is $\omega(1)$ according to Proposition C.1. Theorem 6.7 and Corollary 6.9 suggest that those $\omega(1)$ infected vertices in $V(t)$ will further spread and activate t, which again contradicts to that E does not happen.

Secondly, to conclude the proof, it remains to show that the first term in (C.1) only depends on $P_{\boldsymbol{k}}$, or $\sigma(\boldsymbol{k} \mid E) = cn + o(n)$ for some constant c which does not depend on $\boldsymbol{k}$. As an intuitive argument, Proposition C.1, Theorem 6.7 and Corollary 6.9 suggest that, when E happens, a single activated leaf will activate all the critical and supercritical leaves, and the number of vertices corresponding to all the critical and supercritical leaves is fixed and independent of $\boldsymbol{k}$; based on the tree structure and the number of infected outside vertices, the number of infected vertices in a subcritical leaf may vary; however, we will see that the seeding strategy $\boldsymbol{k}$, adding only a constant number of infections, is too weak to affect the number of infected

vertices in a subcritical leaf.

To break it down, we first show that all critical and supercritical leaves will be activated if E happens. This is straightforward: Proposition C.1 shows that an activated leaf can cause $\omega(1)$ infected vertices in every other leaf, and Theorem 6.7 and Corollary 6.9 indicate that those critical and supercritical leaves will be activated by those $\omega(1)$ infected vertices.

Lastly, assuming all critical and supercritical leaves are activated, we show that the number of infected vertices in any subcritical leaf does not depend on $\boldsymbol{k}$. We do not need to worry about those seeds that are put in the critical or supercritical leaves, as all vertices in those leaves will be infected later. As a result, we only need to show that a constant number of seeds in subcritical leaves has negligible effect to the cascade.

We say a subcritical leaf t is *vulnerable* if there exists a criticial or supercritical leaf t' such that the least common ancestor of t and t' has weight $\Omega(1/n)$, and we say t is *not-very-vulnerable* otherwise. It is easy to see that a vulnerable leaf t will always be activated, even if no seed is put into it. Since each $v \in V(t)$ is connected to one of the $v(t')n$ vertices in $V(t')$ with probability $\Omega(1/n)$, the number of infected neighbors of v follows a Binomial distribution with parameter $(v(t')n, p)$ where $p = \Omega(1/n)$. We only consider $p = \Theta(1/n)$, as there can only be more infected vertices if $p = \omega(1/n)$. If $p = \Theta(1/n)$, the Binomial distribution becomes a Poisson distribution with a constant mean λ for $n \to \infty$. In this case, with constant probability $e^{-\lambda}\frac{\lambda^r}{r!}$, v has r infected neighbors. Therefore, v will be infected with constant probability, and $V(t)$ has $\Theta(n)$ vertices that are infected by $V(t')$ outside. The second part of Theorem 6.6 suggests that, these $\Theta(n)$ infected vertices will further spread and activate t. Therefore, the seeds on those vulnerable subcritical leaves have no effect, since vulnerable subcritical leaves will be activated regardless the seeding strategy.

Let $t_1, \ldots, t_M$ be all the not-very-vulnerable subcritical leaves. Suppose we are at the stage of the cascade process where all those critical, supercritical and vulnerable subcritical leaves have already been activated (as they will with probability $1 - o(1)$ since we assumed that E has happened) and we are revealing the edges between $V \setminus \bigcup_{m=1}^{M} V(t_m)$ and $\bigcup_{m=1}^{M} V(t_m)$ to consider the cascade process in $\bigcup_{m=1}^{M} V(t_m)$. For each $i = 0, 1, \ldots, r-1$ and each $m = 1, \ldots, M$, let χ_i^m be the number of vertices in $V(t_m)$ that have *exactly* i infected neighbors among $V \setminus \bigcup_{m=1}^{M} V(t_m)$, which can be viewed as a random variable. For each $m = 1, \ldots, M$, let χ_r^m be the number of vertices in $V(t_m)$ that have *at least* r infected neighbors. If there are K_m seeds in $V(t_m)$, we increase the value of χ_r^m by K_m. Let $\boldsymbol{\chi}^m = (\chi_0^m, \chi_1^m, \ldots, \chi_r^m)$. Since $(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ completely characterize the expected number of infected vertices in the subcritical leaves, we let $\sigma(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ be the total number of infected vertices in the subcritical leaves, given $(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$. We aim to show that adding $K_1, \ldots, K_M$ seeds in $V(t_1), \ldots, V(t_M)$ only changes the number of infected vertices by $o(n)$. Let $(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ correspond to the case where no seed is added, and $(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)$ correspond to the case where K_m seeds are added to t_m for each $m = 1, \ldots, M$. The outline of the proof is that, we first show that the total variation distance of the two distributions $(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ and $(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)$ is $o(1)$; then we show that $\sigma(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ and $\sigma(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)$ can only differ by $o(n)$ in expectation.

To show the first claim, noticing that M is a constant and $\boldsymbol{\chi}^{m_1}$ is independent of $\boldsymbol{\chi}^{m_2}$ for any m_1, m_2 (the appearances of edges between $V(t_{m_1})$ and $V \setminus \bigcup_{m=1}^{M} V(t_m)$ are independent of the appearances of edges between $V(t_{m_2})$ and $V \setminus \bigcup_{m=1}^{M} V(t_m)$), it is sufficient to show that the total variation distance between $\boldsymbol{\chi}^m$ and $\bar{\boldsymbol{\chi}}^m$ is $o(1)$. For each vertex $v \in V(t_m)$, it is connected to an arbitrary vertex in a critical or supercritical leaf with probability between $\omega(1/n^{1+1/r})$ (since the root has weight $\omega(1/n^{1+1/r})$) and $o(1/n)$ (otherwise t_m is vulnerable). Since the number of infected vertices in $V \setminus \bigcup_{m=1}^{M} V(t_m)$ is $\Theta(n)$, the number of v's infected neighbors follows a Binomial distribution, $\mathrm{Bin}(n, \theta)$, with mean $n\theta$ between $\omega(1/n^{1/r})$ and $o(1)$, we can use Poisson distribution $\mathrm{Po}(n\theta)$ to approximate it. Formally, the total variation distance is $d_{TV}(\mathrm{Bin}(n, \theta), \mathrm{Po}(n\theta)) \leq n\theta^2 = o(1/n)$. Thus, this approximation only changes the total variation distance of $\boldsymbol{\chi}^m$ by $o(1)$.

Observing these, the proposition below shows the total variation distance between $\boldsymbol{\chi}^m$ and $\bar{\boldsymbol{\chi}}^m$ is $o(1)$.

Proposition C.2. *Let λ be such that $\lambda = \omega(1/n^{1/r})$ and $\lambda = o(1)$. Let $Y_1, \ldots, Y_n \in \mathbb{Z}$ be n independently and identically distributed random variables where each Y_i is sampled from a Poisson distribution with mean λ, Let $Z_1, \ldots, Z_n \in \mathbb{Z}$ be n random variables where $Z_1 = \cdots = Z_K = r$ with probability 1, and $Z_{K+1}, \ldots, Z_n$ are independently sampled from a Poisson distribution with mean λ. For $i = 0, 1, \ldots, r-1$, let χ_i be the numbers of random variables in $\{Y_1, \ldots, Y_n\}$ that have value i, and $\bar{\chi}_i$ be the numbers of random variables in $\{Z_1, \ldots, Z_n\}$ that have value i. Let χ_r be the numbers of random variables in $\{Y_1, \ldots, Y_n\}$ that have values at least r, and $\bar{\chi}_r$ be the numbers of random variables in $\{Z_1, \ldots, Z_n\}$ that have values at least r. The total variation distance between $\boldsymbol{\chi} = (\chi_0, \chi_1, \ldots, \chi_r)$ and $\bar{\boldsymbol{\chi}} = (\bar{\chi}_0, \bar{\chi}_1, \ldots, \bar{\chi}_r)$ is $d_{TV}(\boldsymbol{\chi}, \bar{\boldsymbol{\chi}}) = o(1)$ if $K = \Theta(1)$.*

To show this for random vectors χ and $\bar{\chi}$ that have small total variation distance, we use some straightforward computations and the Poisson approximation [17, 200]. We first decouple the correlation between $\boldsymbol{\chi} = (\chi_0, \chi_1, \ldots, \chi_r)$, and consider $r+1$ coordinate-wise independent Poisson $\boldsymbol{\zeta} = (\zeta_0, \zeta_1, \ldots, \zeta_r)$ with the same expectation $\mathbb{E}[\boldsymbol{\chi}] = \mathbb{E}[\boldsymbol{\zeta}]$. Then we define $\bar{\boldsymbol{\zeta}}$ similarly. Finally the total variational distance between two coordinate-wise independent Poisson vectors is well studied in the literature. The full proof of this proposition is omitted due to the space limit.

To show the second claim, notice that the range of the function $\sigma(\cdot)$ falls into the interval $[0, n]$. The total variation distance of $(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ and $(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)$ being $o(1)$ implies that

$$\left|\mathbb{E}_{(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)}[\sigma(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)] - \mathbb{E}_{(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)}[\sigma(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)]\right| = o(n),$$

by a standard property of total variation distance (see, for example, Proposition 4.5 in [144]).

This concludes that the seeds on subcritical leaves can only affect $o(n)$ infections.

Adding together, $\sigma(\boldsymbol{k} \mid E)$ equals to the number of vertices in all critical and supercritical leaves which is independent of the seeding strategy, plus the expected number of infected vertices in those subcritical leaves for which different seeding

strategies only affect a value in $o(n)$. This implies that $\lim_{n\to\infty} \frac{\sigma(\boldsymbol{k}|E)}{n}$ in the first term of (C.1) does not depends on $\boldsymbol{k}$. Therefore, Eqn. (C.1) reveals that $\Sigma_{r,T}(\boldsymbol{k})$ is proportional to $P_{\boldsymbol{k}}$, which implies the lemma.

APPENDIX D

Proofs for Selection and Influence

D.0.1 Proof of Lemma 8.4

The proof has two parts. With Theorem 2.28, we can show $(0, 0)$ is asymptotic stable in the system (8.9) by considering the first order approximation of the system at $(0, 0)$. Therefore, we derive a potential function for system (8.9). Then since the linear approximation of f at 0 in (8.9) is Hurwitz, there is a smooth potential function V defined in Theorem 2.28, so we can argue V is also a valid potential function for the discrete process defined in (8.7).

Proof. Because

$$\left.\frac{\partial f}{\partial \boldsymbol{z}}\right|_0 = \begin{pmatrix} \frac{1-p_{\text{select}}}{2}\left(f'_{\text{inf}}(0) - 1\right) & \frac{1-p_{\text{select}}}{2} f'_{\text{inf}}(0) \\ 2\frac{p_{\text{select}}(1-\gamma)}{2} & -\frac{p_{\text{select}}(1-\gamma)}{2} \end{pmatrix}$$

with the influence function of k-majority defined in Equation (8.3),

$$f_{\text{inf}}(x) = \sum_{\ell=\lceil k/2 \rceil}^{k} \binom{k}{\ell} x^{\ell}(1-x)^{n-\ell},$$

$\left.\frac{\partial f}{\partial \boldsymbol{z}}\right|_0$ is Hurwitz. By Theorem 2.28, there exists $\delta_{p_{\text{select}}} > 0$, $B(\delta_{p_{\text{select}}}) = \{x \in \mathbb{R}^2 : \|x - 0\| \leq \delta_{p_{\text{select}}}\}$ and a potential function $V : B(\delta_{p_{\text{select}}}) \mapsto \mathbb{R}$ such that $V(0) = 0$, and $V(x) > 0$, $\frac{d}{dt}(V(x)) < 0$ for $x \in B(\delta_{p_{\text{select}}})$, $x \neq 0$.

We want to show V is also a potential function for (8.7) as n large enough, the series $V(Z^{(i)})$ is decreasing. Because V is smooth there exists a function, we have

$$V(Z^{(i+1)}) - V(Z^{(i)}) = \nabla V(Z^{(i)}) \cdot (Z^{(i+1)} - Z^{(i)}) + O\left((Z^{(i+1)} - Z^{(i)})^2\right)$$

Because $\nabla V(Z^{(i)}) \cdot (Z^{(i+1)} - Z^{(i)}) \leq \frac{1}{n} \frac{d}{dt} V(\boldsymbol{z})\big|_{\boldsymbol{z}=Z^{(i)}}$ and $\|Z^{(i+1)} - Z^{(i)}\| \leq C/n$ for some constant C, we have

$$V(Z^{(i+1)}) - V(Z^{(i)}) = \frac{1}{n}\left(\frac{d}{dt} V(\boldsymbol{z}) + O(1/n)\right)\Bigg|_{\boldsymbol{z}=Z^{(i)} \in N} < 0$$

Therefore, $\lim_{t\to\infty} \|Z^{(t)} - 0\| = 0$. □

D.0.2 Proof of Lemma 8.6

The first part of Lemma 8.6 is quite straightforward. We first define our stable region R_A and show at each boundary the drift is inward such that if the $\boldsymbol{z}(t)$ is at the boundary the $\boldsymbol{z}(t+\epsilon)$ will go back to the stable region. For the second part, we show $\boldsymbol{z}$ hits the stable region R_A fast by taking ${p_{\text{select}}}^*$ large enough.

Proof of Lemma 8.6. Because f_{inf} is k majority which is strictly increasing which is invertible, the zeros of Equation (8.8) can be denoted as two continues curves $\left(x_1, x_2^{(1)}(x_1)\right)$ and $\left(x_1, x_2^{(2)}(x_1)\right)$ where

$$x_2^{(1)}(t) = \frac{f_{\text{inf}}^{-1}(t) - t}{1-2t}, \text{ and } x_2^{(2)}(t) = 2t(1-t) \tag{D.1}$$

, see Figure 8.4 for an example. Now we want to show these two curves have an intersection (x_1^*, x_2^*), and we pick the one with the smallest positive x_1^* and define

$$R_A = \{(x_1, x_2) : 0 \le x_1 \le x_1^*, 0 \le x_2 \le x_2^*\}.$$

To show these two curves have an intersection, since $f'_{\text{inf}}(0) = 0$, by Equation (D.1), there exists $\tilde{x}_1 > 0$ such that $x_2^{(1)}(\tilde{x}_1) \ge x_2^{(2)}(\tilde{x}_1)$. On the other hand, because

$$\lim_{t \to 1/2} x_2^{(1)}(t) = \frac{1}{2}\left(1 - \frac{1}{f'_{\text{inf}}(0.5)}\right) < \frac{1}{2} = x_2^{(2)}(0.5)$$

, there exists $\bar{x}_1$ such that $x_2^{(1)}(\bar{x}_1) \le x_2^{(2)}(\bar{x}_1)$. By intermediate value theorem, we have x_1^* and x_2^* such that $x_2^* = x_2^{(1)}(x_1^*) = x_2^{(2)}(x_1^*)$.

Now we want to prove this is a stable region for $\boldsymbol{z}$ in (8.9). To prove boundedness, we only need to prove $\boldsymbol{z}$ cannot leave the upper and right boundaries. Suppose $\boldsymbol{z}(t_0) \in R_A$, and given arbitrary $t_1 > t_0$. At the right boundary $x \in \{(x_1, x_2) : 0 \le x_1 \le x_1^*, x_2 = x_2^*\}$ by (8.8) we have $f_1(\boldsymbol{x}, t) \le 0$ for all $t > 0$ so the process cannot increase and $z_1(t_1) \le x_1^*$. On the upper boundary $\{(x_1, x_2) : x_1 = x_1^*, 0 \le x_2^*\}$ by (8.8) we have $f_2(\boldsymbol{x}, t) \le 0$ so the process cannot increase and $z_2(t_1) \le x_2^{(2)}(x_1^*)$. Therefore we prove R_A is a stable region for Equation (8.9).

Now we want to prove the second part: If $\boldsymbol{z}(0) = (0, 0.5)$, there exists ${p_{\text{select}}}^* < 1$, and τ_0 such that $\boldsymbol{z}(\tau_0) \in R_A$. We define $\delta = \min\{x_1^*, x_2^*\}$ which is a positive constant independent to p_{select}. Note that in (8.9) we have $z_1(t) \le \frac{(1-p_{\text{select}})}{2} t$ Therefore if $t \le (1-\gamma)\delta/4(1-p_{\text{select}})$,

$$z_1(t) \le (1-\gamma)\delta/8 \tag{D.2}$$

Additionally, by (8.9) and (D.2) we have $z_2' \le -p_{\text{select}}(1-\gamma)/2\, z_2 + (1-\gamma)\delta/8$ and

$$z_2(t) \le \left(\frac{1}{2} - \frac{\delta}{4p_{\text{select}}}\right) \exp\left(-\frac{p_{\text{select}}(1-\gamma)}{2} t\right) + \frac{\delta}{4p_{\text{select}}}$$

If we take $\tau_0 = (1-\gamma)\delta/4(1-p_{\text{select}})$, we have

$$z_1(\tau_0) \leq \frac{(1-\gamma)\delta}{8} \text{ and } z_2(\tau_0) \leq \left(\frac{1}{2} - \frac{\delta}{4p_{\text{select}}}\right) \exp\left(-\frac{p_{\text{select}}}{1-p_{\text{select}}} \frac{\delta(1-\gamma)^2}{8}\right) + \frac{\delta}{4p_{\text{select}}}$$

which are smaller than δ if we take $1-p_{\text{select}}$ small enough. Therefore $\boldsymbol{z}(\tau_0) \in R_A$. $\square$

APPENDIX E

Proofs for Gnp

E.1 Proofs in Section 9.4.2

Proof of lemma 9.9. Define $s : \Omega \mapsto \mathbb{R}$ as follows

$$s(x) = \sum_{\ell=Bias(x)}^{n/2-1} d_\ell, \text{ where } Bias(x) < n/2$$
$$s(x) = 0, \text{ where } Bias(x) = n/2$$

Note that the value of s only depends on the bias of each state and for $x, y \in \Omega$ with $Bias(x) = Bias(y)$ we have $s(x) = s(y)$, so we can abuse the notion and consider potential function with integral domain $s : [0, n/2 - 1] \mapsto \mathbb{R}$ such that $s(l) \triangleq s(x)$ for some x such that $Bias(x) = l$.

To prove s is a valid super solution of τ, by corollary 2.10 it is sufficient for us to show that

$$\mathcal{L}s(x) \leq -1 \text{ where } Bias(x) < n/2, \tag{E.1}$$
$$s(x) \geq 0 \text{ where } Bias(x) = n/2. \tag{E.2}$$

For the Equation (E.1), if $Bias(x) = \ell$ and $0 < \ell < n/2$,

$$\begin{aligned}\mathcal{L}s(x) &= \sum_{y\in\Omega} P_{x,y}s(y) - s(x) \\ &= \sum_{y\in\Omega} P_{x,y}\big(s(Bias(y)) - s(Bias(x))\big)\end{aligned}$$

By the definition of $\mathcal{M}$, $P(x, y) > 0$ only if the states x, y differ by at most one digit, we only need to consider the states y such that $|Bias(y) - \ell| \leq 1$, by the defintion of

s,

$$\begin{aligned}
\mathcal{L}s(x) &= \sum_{y:Bias(y)=\ell+1} P_{x,y}\big(s(\ell+1) - s(\ell)\big) \\
&+ \sum_{y:Bias(y)=\ell-1} P_{x,y}\big(s(\ell-1) - s(\ell)\big) \\
&= -\left(\sum_{y:Bias(y)=\ell+1} P_{x,y}\right) d_\ell + \left(\sum_{y:Bias(y)=\ell-1} P_{x,y}\right) d_{\ell-1} \\
&= -\Pr_{\mathcal{M}}[Bias(X') = \ell+1 | X = x] d_\ell \\
&+ \Pr_{\mathcal{M}}[Bias(X') = \ell-1 | X = x] d_{\ell-1}
\end{aligned}$$

by the definition of $p^+(x)$ and $p^-(x)$

$$\begin{aligned}
\mathcal{L}s(x) &= -p^+(x)d_\ell + p^-(x)d_{\ell-1} \\
&\leq -p^+(x)\left(\frac{1}{p^+(x)} + \frac{p^-(x)}{p^+(x)} d_{\ell-1}\right) + p^-(x)d_{\ell-1} = -1
\end{aligned}$$

where the last equality comes from the definition of d_ℓ On the other hand, if $Bias(x) = 0$

$$\begin{aligned}
\mathcal{L}s(x) &= \sum_{y:Bias(y)=1} P_{x,y}\big(s(1) - s(0)\big) \\
&= -\Pr_{\mathcal{M}}[Bias(X') = 1 | X = x] d_0 = -p^+(x)d_0 \leq -1.
\end{aligned}$$

Equation (E.2) automatically holds by the definition of s.

Therefore, applying corollary 2.10 we have $\max_{x\in\Omega} \tau(x) \leq \max_{x\in\Omega} s(x) = \sum_{\ell=0}^{n/2-1} d_\ell$. □

The proof of theorem 9.11, which is rather straightforward but tedious, uses lemma 9.9 and a careful estimation of the potential function from the recursive equation (9.6).

is rather straightforward but tedious by use of lemma 9.9, and estimation of the potential function from the recursive Equation (9.6) carefully.

Proof of theorem 9.11. With the help of lemma 9.9, we only need to give an upper bound the recursive equations (9.6). With the condition in the statements, suppose we prove the following equations: There exists some positive constant C_1, C_2, C_3, C_4, D_1

such that

$$d_\ell \le C_1\sqrt{n} \text{ where } \ell < D_1\left\lceil\sqrt{n}\right\rceil, \tag{E.3}$$

$$d_\ell \le C_2\frac{n}{\ell} \text{ where } D_1\left\lceil\sqrt{n}\right\rceil \le \ell \le \check{\epsilon}n, \tag{E.4}$$

$$d_\ell \le C_3 \text{ where } \check{\epsilon}n < \ell \le (1/2-\hat{\epsilon})n, \tag{E.5}$$

$$d_\ell \le C_4\frac{n}{n/2-\ell} \text{ where } (1/2-\hat{\epsilon})n < \ell < n/2 \tag{E.6}$$

Supposing the above inequalities are true, by lemma 9.9 we can complete the proof as follows:

$$\begin{aligned}
&\max_{x\in\Omega}\mathbb{E}_{\mathcal{M}}[T(G,f,x)] \le \sum_{\ell=0}^{n/2-1} d_\ell \\
\le &\sum_{\ell=0}^{D_1\left\lceil\sqrt{n}\right\rceil-1} C_1\sqrt{n} + \sum_{\ell=D_1\left\lceil\sqrt{n}\right\rceil}^{\check{\epsilon}n} C_2\frac{n}{\ell} + \sum_{\ell=\check{\epsilon}n+1}^{(1/2-\hat{\epsilon})n} C_3 + \sum_{\ell=(1/2-\hat{\epsilon})n+1}^{n/2-1} C_4\frac{n}{n/2-\ell} \\
\le &D_1\left\lceil\sqrt{n}\right\rceil\cdot C_1\sqrt{n} + C_2 n\sum_{\ell=D_1\left\lceil\sqrt{n}\right\rceil}^{\check{\epsilon}n}\frac{1}{\ell} + C_3(1/2-\hat{\epsilon}-\check{\epsilon})n + C_4 n\sum_{\ell=1}^{\hat{\epsilon}n-1}\frac{1}{\ell} \\
= &O(n\ln n)
\end{aligned}$$

Now we are going to use induction to prove Equations (E.3), (E.4), (E.5), and (E.6).

Equation (E.3)**:** We first use induction to prove the following inequality: If $A(n) = \frac{1}{p^+} + \frac{\sqrt{n}}{p^+A_1B_1}$ and $B(n) = \frac{\sqrt{n}}{p^+A_1B_1}$, for all $\ell, 0 \le \ell \le D_1\left\lceil\sqrt{n}\right\rceil$

$$d_\ell \le A(n)\left(1+\frac{A_1B_1}{\sqrt{n}}\right)^\ell - B(n) \tag{E.7}$$

Because for all constant D_1 there exists some constant $C_1 > 0$ such that

$$A(n)\left(1+\frac{A_1B_1}{\sqrt{n}}\right)^\ell - B(n) \le C_1\sqrt{n}$$

for all $\ell \le D_1\left\lceil\sqrt{n}\right\rceil$, the Equation (E.3) is proven once the Equation (E.7) is true. Now, let's prove (E.7).

For $\ell = 0$, applying Equation (9.8) to Equation (9.6), we have

$$\begin{aligned} d_0 &= \max_{x\in\Omega:Bias(x)=0} \frac{1}{p^+(x)} \\ &\le \max_{x\in\Sigma^s\cup\Sigma^m} \frac{1}{p^+(x)} && \text{(because } \{x\in\Omega : Bias(x)=0\} \subset \Sigma^s\cup\Sigma^m) \\ &\le 1/p^+ && \text{(by Equation (9.8))} \\ &= A - B && \text{(by the definition of } A \text{ and } B) \end{aligned}$$

Suppose $d_{\ell-1} \le A\left(1+\frac{A_1B_1}{\sqrt{n}}\right)^{\ell-1} - B$ for some $1 < \ell < D_1\lceil\sqrt{n}\rceil$. Since $\ell < D_1\lceil\sqrt{n}\rceil - 1 < \check{\epsilon}n$, $\{x\in\Omega : Bias(x)=\ell\} \subset \Sigma^s$ and we can apply equation (9.10) and (9.8) to equation (9.6) and have

$$\begin{aligned} d_\ell &\le \frac{1}{p^+} + \left(1 + A_1\left(\frac{B_1}{\sqrt{n}} - \ell\right)\right) d_{\ell-1} \\ &\le \frac{1}{p^+} + \left(1+\frac{A_1B_1}{\sqrt{n}}\right) d_{\ell-1} \end{aligned} \tag{E.8}$$

By induction hypothesis, and definition of B we have

$$\begin{aligned} &\le \frac{1}{p^+} + \left(1+\frac{A_1B_1}{\sqrt{n}}\right)\left(A\left(1+\frac{A_1B_1}{\sqrt{n}}\right)^{\ell-1} - B\right) \\ &\le A\left(1+\frac{A_1B_1}{\sqrt{n}}\right)^{\ell} - B - \left(\frac{A_1B_1}{\sqrt{n}}B - \frac{1}{p^+}\right) \\ &\le A\left(1+\frac{A_1B_1}{\sqrt{n}}\right)^{\ell} - B \end{aligned}$$

Equation (E.4)**:** We use induction again to prove Equation (E.4) holds for $D_1\lceil\sqrt{n}\rceil \le \ell \le \check{\epsilon}n$.

For $\ell = D_1\lceil\sqrt{n}\rceil$, we already have $d_\ell \le C_1\sqrt{n} \le C_2\frac{n}{D_1\lceil\sqrt{n}\rceil}$ so if we take $C_2 \ge C_1D_1$

$$d_\ell \le C_1\sqrt{n} \le C_2\frac{n}{D_1\lceil\sqrt{n}\rceil} = C_2\frac{n}{\ell}.$$

Suppose $d_{\ell-1} \le C_2\frac{n}{\ell-1}$ for some $D_1\lceil\sqrt{n}\rceil < \ell < \check{\epsilon}n$. Because $\{x\in\Omega : Bias(x) = \ell\} \subset \Sigma^s$, by equation (E.8) and induction hypothesis we have

$$\begin{aligned} d_\ell &\le \frac{1}{p^+} + \left(1 + A_1\left(\frac{B_1}{\sqrt{n}} - \ell\right)\right) d_{\ell-1} \\ &= \frac{1}{p^+} + \left(1 - \left(\frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)\right) d_{\ell-1} \end{aligned}$$

and

$$d_\ell \le \frac{1}{p^+} + \left(1 - \left(\frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)\right) C_2 \frac{n}{\ell - 1}$$
$$= \frac{C_2 n}{\ell} + \left(\frac{1}{p^+} + \left(1 - \frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)\frac{C_2 n}{\ell - 1} - \frac{C_2 n}{\ell}\right).$$

Therefore equation (E.4) is proven if $\frac{1}{C_2 p^+} + \left(1 - \frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)\frac{n}{\ell-1} - \frac{n}{\ell} \le 0$. By taking $C_2 \ge \frac{2}{p^+ A_2}$ and $D_1 \ge 4B_1$ and $D_1^2 \ge 4/A_1$ we have

$$\begin{aligned}
\frac{A_1}{2} &\le A_1 - \frac{A_1B_1}{D_1} - \frac{1}{D_1^2} \\
&\le \frac{n}{\ell}\left(\frac{A_1\ell - A_1B_1\sqrt{n}}{n} - \frac{1}{\ell}\right) && \text{(because } \ell > D_1\lceil\sqrt{n}\rceil\text{)} \\
&\le \frac{n}{\ell - 1}\left(\frac{A_1\ell - A_1B_1\sqrt{n}}{n} - \frac{1}{\ell}\right) \\
&= \frac{n}{\ell} - \frac{n}{\ell - 1}\left(1 - \frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)
\end{aligned}$$

Because $C_2 \ge \frac{2}{p^+ A_2}$, we have $\frac{1}{C_2 p^+} \le \frac{A_1}{2}$ and using the above inequality we get

$$\frac{1}{C_2 p^+} \le \frac{n}{\ell} - \frac{n}{\ell - 1}\left(1 - \frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)$$

which completes proving Equation (E.4). Finally, by the Equation (E.4)

$$d_{\check{\epsilon}n} \le C_2 \frac{n}{\check{\epsilon}n} = C_2/\check{\epsilon}. \tag{E.9}$$

Equation (E.5)**:** We use induction to prove d_ℓ is bounded above by some constant C_3 for all ℓ such that $\check{\epsilon}n < \ell \le (1/2 - \hat{\epsilon})n$.

For $\ell = \check{\epsilon}n + 1$, because $\{x \in \Omega : Bias(x) = \check{\epsilon}n + 1\} \subset \Sigma^m$, we can apply (9.16) and (9.13) into Equation (9.6) and have

$$\begin{aligned}
d_\ell &\le \frac{1}{p^+} + (1 - A_2)\, d_{\ell-1} && \text{(E.10)} \\
&\le \frac{1}{p^+} + (1 - A_2)\, C_2/\check{\epsilon} && \text{(by Equation (E.9))} \\
&\le A_2 \frac{1}{p^+ A_2} + (1 - A_2)\, C_2/\check{\epsilon}
\end{aligned}$$

Because $0 \le A_2 < 1$ if we take $C_3 = \max\{\frac{1}{p^+ A_2}, C_2/\check{\epsilon}\}$ the base case of (E.3) is true. Suppose $d_{\ell-1} \le C_3$ for some $\check{\epsilon}n < \ell < (1/2 - \hat{\epsilon})n$, because $\{x \in \Omega : Bias(x) = \ell\} \subset \Sigma^m$

we can use (E.10) and

$$\begin{aligned} d_\ell &= \frac{1}{p^+} + (1 - A_2)\, d_{\ell-1} \le \frac{1}{p^+} + (1 - A_2)\, C_3 \\ &\le C_3 - \left(A_2 C_3 - \frac{1}{p^+} \right) \le C_3, \end{aligned}$$

because $C_3 = \max\{\frac{1}{p^+ A_2} C_2/\check{\epsilon}\}$

This finishes the proof of Equation (E.5).

Equation (E.6)**:** Because $\{x \in \Omega : Bias(x) = \ell\} \subset \Sigma^l$ for all $(1/2 - \hat{\epsilon})n < \ell < n/2$ we can apply (9.17) and (9.14) into Equation (E.9), and get

$$d_\ell \le \frac{4}{1/2 - \ell/n} + (1 - A_3) d_{\ell-1} = \frac{4n}{n/2 - \ell} + (1 - A_3) d_{\ell-1}$$

Recursively applying this relation, d_ℓ is upper bounded by

$$\sum_{j=(1/2-\hat{\epsilon})n+1}^{l} \frac{4n(1 - A_3)^{\ell-j}}{n/2 - j} + (1 - A_3)^{\ell-(1/2-\check{\epsilon})n} d_{(1/2-\hat{\epsilon})n}$$

because of Equation (E.5) this is at most

$$\begin{aligned} &\le \sum_{j=(1/2-\hat{\epsilon})n+1}^{l} \frac{4n(1 - A_3)^{\ell-j}}{n/2 - j} + C_3 \\ &\le 4n \sum_{i=0}^{\ell-(1/2-\hat{\epsilon})n-1} \frac{(1 - A_3)^i}{n/2 - \ell + i} + C_3 \qquad \text{(taking } i = \ell - j\text{)} \\ &\le \frac{4n}{n/2 - \ell} \sum_{i=0}^{\ell-(1/2-\hat{\epsilon})n-1} (1 - A_3)^i \frac{n/2 - \ell}{n/2 - \ell + i} + C_3 \end{aligned}$$

Because $(1 - A_3)^i \frac{n/2-\ell}{n/2-\ell+i} \le (1 - A_3)^i$ and taking $C_4 \ge 2/A_3 + \hat{\epsilon} C_3$, d_ℓ is bounded above by

$$\frac{4n}{n/2 - \ell} \sum_{i=0}^{\infty} (1 - A_3)^i + C_3 = \frac{4}{A_3} \frac{n}{n/2 - \ell} + C_3 \le \frac{C_4 n}{n/2 - \ell}$$

□

E.2 Exertion and Drift: Proof of Lemma 9.12 and 9.13

In this section, we want to control the exertion $p_G^+(x)$ and drift $p_G^+(x) - p_G^-(x)$ of the process $\mathcal{M}$ on graph $G \sim \mathcal{G}$. To achieve these upper bounds, we prove several

properties of dense Erdös-Rényi graphs which might seem ad-hoc, but there is a common thread under these lemmas: concentration phenomena in dense Erdös-Rényi graph. Our main tools are the spectral property of random graph and several variants of Chernoff bounds.

E.2.1 Exertion and Lemma 9.12

We partition the lemma 9.12 into lemma E.1 and E.2, and use the mixing lemma 2.45 to show all configurations have $p_G^+(x)$ close to that of the complete graph if G is a good expander.

Lemma E.1 (Exertion of Σ^s, Σ^m)**.** *If G λ-expander with nearly uniform degree $E(\delta_d)$, $\delta_d < 1$ and $\lambda^2 < \frac{1-\delta_d}{1+\delta_d} \cdot \frac{\hat{\epsilon}}{18}$, for all x with $Bias(x) < 1/2 - \hat{\epsilon}$,*

$$\frac{\hat{\epsilon}}{2} f\left(\frac{\hat{\epsilon}}{2}\right) < p_G^+(x) \leq 1.$$

Proof. Let's consider a fixed configuration x where $\hat{\epsilon} \leq pos(x) < 1/2$ and where the number of red nodes is less than the number of blue nodes'. We can partition the V into three sets of vertices $S_x, T_x, U_x \subset V$ such that

$$S_x = \{v \in V : x(v) = 0, r_x(v) < \frac{\hat{\epsilon}}{2}\}, \tag{E.11}$$

$$T_x = \{v \in V : x(v) = 0, r_x(v) \geq \frac{\hat{\epsilon}}{2}\}, \text{ and} \tag{E.12}$$

$$U_x = \{v \in V : x(v) = 1\}. \tag{E.13}$$

Observe that U_x is the set of red nodes in configuration x, and $S_x \cup T_x$ is the set of blue nodes so $|S_x \cup T_x| = Pos(x) \geq \hat{\epsilon}n$. Moreover by the definition of $\mathcal{M}$ with update function f, the definition in (E.12), and the monotone property of f, the probability a node $v \in T_x$ becomes red in the next step, given v is chosen and the current configuration is x, is greater than $f(\frac{\hat{\epsilon}}{2})$. As a result, every node in T_x has a constant probability to change if chosen, and

$$p_G^+(x) \geq \frac{|T_x|}{n} \cdot f\left(\frac{\hat{\epsilon}}{2}\right)$$

Therefore, if we prove the following inequality

$$|S_x| < \frac{\hat{\epsilon}}{2}|V| \tag{E.14}$$

then the size of set T_x is greater than $\frac{\hat{\epsilon}}{2}|V|$, and we have $p_G^+(x) \geq \frac{\hat{\epsilon}}{2} f\left(\frac{\hat{\epsilon}}{2}\right)$ which finishes the proof.

Now it is sufficient for us to prove equation (E.14). By the definition in (E.11) we can upper bound the number of edges between S_x and U_x, $e(S_x, U_x)$, and use mixing lemma 2.45 to upper bound the size of S_x.

First, since the degree of nodes are nearly uniform, the volume of S_x, and U_x can be bounded

$$(1-\delta_d)np|S_x| \le vol(S_x) \le (1+\delta_d)np|S_x| \tag{E.15}$$
$$(1-\delta_d)np|U_x| \le vol(U_x) \le (1+\delta_d)np|U_x|, \tag{E.16}$$

and by the definition of S_x in (E.11) the number of edges between S_x and U_x can be bounded as follows:

$$e(S_x, U_x) \le \frac{\hat{\epsilon}}{2} \cdot vol(S_x) \le \frac{\hat{\epsilon}}{2} \cdot (1+\delta_d)np|S_x| \tag{E.17}$$

Applying mixing lemma 2.45 on sets S_x and U_x, and we have

$$\begin{aligned}
&\left| e(S_x, U_x) - \frac{vol(S_x)vol(U_x)}{vol(G)} \right| \le \lambda\sqrt{vol(S_x)vol(U_x)} \\
&\frac{vol(S_x)vol(U_x)}{vol(G)} - e(S_x, U_x) \le \lambda\sqrt{vol(S_x)vol(U_x)} \\
&\frac{vol(S_x)vol(U_x)}{vol(G)} - \frac{\hat{\epsilon}}{2}vol(S_x) \le \lambda\sqrt{vol(S_x)vol(U_x)} \qquad \text{(by equation (E.17))} \\
&\left(\frac{vol(U_x)}{vol(G)} - \frac{\hat{\epsilon}}{2}\right)\sqrt{vol(S_x)} \le \lambda\sqrt{vol(U_x)} \qquad \text{(E.18)}
\end{aligned}$$

For the left hand side, because the degree of G is near uniform, we can approximate the ratio of $\frac{vol(U_x)}{vol(G)}$ by the ratio of $\frac{|U_x|}{|G|}$ as follows

$$\begin{aligned}
&\left(\frac{vol(U_x)}{vol(G)} - \frac{\hat{\epsilon}}{2}\right)\sqrt{vol(S_x)} \\
&\ge \left(\frac{(1-\delta_d)|U_x|}{(1+\delta_d)|V|} - \frac{\hat{\epsilon}}{2}\right)\sqrt{vol(S_x)}
\end{aligned}$$

Because $pos(x) < 1/2$, this is

$$\begin{aligned}
&\ge \left(\frac{(1-\delta_d)/2}{(1+\delta_d)} - \frac{\hat{\epsilon}}{2}\right)\sqrt{vol(S_x)} \\
&\ge \frac{1}{2}\left(\frac{1-\delta_d}{1+\delta_d} - \hat{\epsilon}\right)\sqrt{vol(S_x)} \\
&\ge \frac{1}{3}\sqrt{vol(S_x)}. \qquad \text{(E.19)}
\end{aligned}$$

For the right hand side, we can upper bound the volume of U_x by

$$vol(U_x) \le vol(V) \le (1+\delta_d)n^2p. \tag{E.20}$$

Applying equations (E.19) and (E.20) into equation (E.18) yields

$$\begin{aligned}\frac{1}{9}vol(S_x) &\le \lambda^2 vol(U_x)\\ vol(S_x) &\le 9\lambda^2(1+\delta_d)n^2p\\ (1-\delta_d)np|S_x| &\le 9\lambda^2(1+\delta_d)n^2p\\ |S_x| &\le 9\lambda^2\frac{1+\delta_d}{1-\delta_d}n = o(n)\end{aligned}$$

which is smaller than $\frac{\hat{\epsilon}}{2}n$ because $\lambda^2 < \frac{1-\delta_d}{1+\delta_d}\cdot\frac{\hat{\epsilon}}{18}$. □

Lemma E.2 (Exertion of Σ^l)**.** *If G is a λ-expander with nearly uniform degree $E(\delta_d)$, $\delta_d < 1$ and $\lambda^2 < \frac{1-\delta_d}{1+\delta_d}\cdot\frac{(1/2-\hat{\epsilon})^2}{2}$, for all x with $bias(x) > 1/2-\hat{\epsilon}$,*

$$\frac{1}{4}(1/2 - bias(x)) < p_G^+(x) \le (1/2 - bias(x)).$$

Proof. Without lose of generality, we consider the configuration x where $pos(x) < \hat{\epsilon}$.

The proof of the upper bound is straightforward. Suppose $H_v = \{v$ changes from blue to red in the step given v is chosen and the configuration is $x\}$

$$\begin{aligned}p_G^+(x) &= \Pr_{\mathcal{M}}[Bias(X_1) = Bias(x)+1 | X_0 = x]\\ &= \frac{1}{n}\sum_{v\in V}\Pr_{\mathcal{M}}[H_v]\\ &\le \frac{1}{n}\sum_{v\in V}\mathbb{I}[v \text{ is blue}] = pos(x) = (1/2 - bias(x)).\end{aligned}$$

For the lower bound, similar to lemma E.1, given a configuration, we partition the set of nodes V into three sets S'_x, T'_x, U'_x

$$S'_x = \{v \in V : x(v) = 1, r_x(v) \ge \frac{1}{2}\}, \tag{E.21}$$

$$T'_x = \{v \in V : x(v) = 1, r_x(v) < \frac{1}{2}\}, \tag{E.22}$$

$$U'_x = \{v \in V : x(v) = 1\} = S'_x \cup T'_x \tag{E.23}$$

To show a lower bound for $p_G^+(x)$, it is sufficient to show that the fraction of red nodes $T'_x \subset V$ is large and has constant probability to change to blue if selected to update. Because the probability that node $v \in T'_x$ becomes blue in the next step given v is chosen with configuration x is $f(1-r_x(v))$, by the definition in (E.22) and by the monotone property of f

$$f(1-r_x(v)) \ge f\left(\frac{1}{2}\right) \ge 1/2.$$

Suppose

$$|S'_x| < \frac{1}{2} Pos(x) \tag{E.24}$$

then the size of set T'_x is greater than $\frac{1}{2}Pos(x)$, and we have a lower bound for $p^+_G(x)$: $\frac{1}{2}\frac{|T'_x|}{n} \geq \frac{1}{4}pos(x)$ which finishes the proof

Now it is sufficient for us to prove equation (E.24). By the definition in (E.21) we can upper bound the number of edges between S'_x and U'_x, $e(S'_x, U'_x)$, and use mixing lemma 2.45 to upper bound the size of S'_x.

First, since the degree of nodes are nearly uniform, the volume of S_x, and U_x can be bounded

$$(1-\delta_d)np|S'_x| \leq vol(S'_x) \leq (1+\delta_d)np|S'_x| \tag{E.25}$$

$$(1-\delta_d)np|U'_x| \leq vol(U'_x) \leq (1+\delta_d)np|U'_x|, \tag{E.26}$$

and by the definition of S'_x in (E.11) the number of edges between S'_x and U'_x can be bounded as follows

$$e(S'_x, U'_x) \geq \frac{1}{2} \cdot vol(S'_x) \tag{E.27}$$

Applying mixing lemma 2.45 on sets S'_x and U'_x, we have

$$\left| e(S'_x, U'_x) - \frac{vol(S'_x)vol(U'_x)}{vol(G)} \right| \leq \lambda \sqrt{vol(S'_x)vol(U'_x)}$$

$$e(S'_x, U'_x) \leq \frac{vol(S'_x)vol(U'_x)}{vol(G)} + \lambda \sqrt{vol(S'_x)vol(U'_x)}$$

By equation (E.27)

$$\frac{1}{2} \cdot vol(S'_x) \leq \frac{vol('S_x)vol(U'_x)}{vol(G)} + \lambda \sqrt{vol(S'_x)vol(U'_x)}$$

Reorganizing the last inequality we have,

$$vol(S'_x) \leq \left(\frac{\lambda}{\frac{1}{2} - \frac{vol(U'_x)}{vol(G)}} \right)^2 vol(U'_x)$$

Because $\frac{1}{2} - \frac{vol(U'_x)}{vol(G)} = \frac{1}{2} - pos(x) > 1/2 - \hat{\epsilon}$,

$$vol(S'_x) \leq \frac{\lambda^2}{(1/2-\hat{\epsilon})^2} vol(U'_x).$$

Finally by equations (E.25) and (E.26) and taking δ_d small enough

$$|S'_x| \leq \frac{\lambda^2}{(1/2-\hat{\epsilon})^2} \cdot \frac{1+\delta_d}{1-\delta_d} vol(U'_x) < \frac{1}{2} vol(U'_x)$$

The last inequality holds because $\lambda^2 < \frac{1-\delta_d}{1+\delta_d} \cdot \frac{(1/2-\hat{\epsilon})^2}{2}$. □

E.2.2 Drift and Lemma 9.13

In this section, we want to prove lemma 9.13 As discussed in section 9.4.3, we will prove lower bounds for drift $D_G(x)$ in Σ^s, Σ^m and Σ^l separately, and use the lower bound for $p_G^+(x)$ in lemma 9.12 to prove lemma 9.13.

E.2.2.1 Drift in Σ^s and Σ^m

The high level idea is to use a serial of triangle inequalities: Given a configuration $x \in \Omega$:

1. The drift $D_G(x)$ is close to its expectation $\mathbb{E}_{\mathcal{G}}[D_G(x)]$;
2. The expectation $\mathbb{E}_{\mathcal{G}}[D_G(x)]$ is close to the drift on complete graphs $D_{K_n}(x)$; and
3. The drift on the complete graph $D_{K_n}(x)$ is lower bounded by its $bias(x)$.

The third part is easy because when $pos(x) > 1/2$ the drift $D_{K_n}(x)$ is

$$D_{K_n}(x) = p^+_{K_n}(x) - p^-_{K_n}(x) = f(pos(x)) - pos(x) \tag{E.28}$$

and equations (9.15), (9.16), and (9.17) can be obtained by the definition of f.

Therefore, our strategy for states in Σ^s, Σ^m is to argue the value of $\{D_G(x)\}_{x\in\Omega}$ is close to $\{D_{K_n}(x)\}_{x\in\Omega}$ with high probability. The first part is proved in lemmas E.3 and E.4, and the second part is fulfilled in lemma E.5. Informally lemma E.3 shows $\mathbb{E}_{\mathcal{G}}[D_G(x)] - D_{K_n}(x) = O(1/n)$, and lemma E.4 shows $\mathbb{E}_{\mathcal{G}}[D_G(x)] - D_{K_n}(x) = O(\sqrt{(\log n)/n})$.

Before digging into the lemmas, let's rewrite $D_G(x)$. Without lost of generality if $pos(x) > 1/2$,

$$\begin{aligned} D_G(x) &= \mathbb{E}_{\mathcal{M}}[Pos(X')|X = x] - Pos(x) \\ &= \frac{1}{n}\sum_{v\in V} \Pr_{\mathcal{M}}[X'(v) = 1 | v \text{ is chosen}, X = x] - pos(x) \\ &= \frac{1}{n}\sum_{v\in V} f(r_x(v)) - pos(x), \end{aligned}$$

and by symmetry of $\mathcal{G}$ we can fixed arbitrary node $v \in V$ and have

$$\mathbb{E}_{\mathcal{G}}[D_G(x)] = \mathbb{E}_{\mathcal{G}}[f(r_x(v))] - pos(x). \tag{E.29}$$

Lemma E.3 (Expected Drift in Σ^s)**.** *If $x \in \Sigma^s$ where $bias(x) < \check{\epsilon}$ then there exists constant $K_1 > 0$ such that for large enough n,*

$$\mathbb{E}_{\mathcal{G}}[D_G(x)] \geq f\left(\frac{1}{2} + bias(x)\right) - \left(\frac{1}{2} + bias(x)\right) - \frac{K_1}{n}.$$

Lemma E.4 (Expected Drift in Σ^m)**.** *If $x \in \Sigma^m$ where $\check{\epsilon} \leq bias(x) \leq 1/2 - \hat{\epsilon}$ then there exists constant $K_2 > 0$ such that for large enough n, $\mathbb{E}_{\mathcal{G}}[D_G(x)]$ is greater than*

$$f\left(\frac{1}{2} + bias(x)\right) - \left(\frac{1}{2} + bias(x)\right) - K_2\sqrt{\frac{\log n}{n}}.$$

Lemma E.5 (Small noise in Σ^s and Σ^m)**.** *For all $x \in \Sigma^s \cup \Sigma^m$, $D_G(x)$ there exists a constant $L > 0$ such that when n large enough*

$$D_G(x) - \mathbb{E}_{\mathcal{G}}[D_G(x)] > -\frac{L}{\sqrt{n}}$$

happens with high probability over the randomness of $\mathcal{G}$.

Properties of $r_x(v)$. Due to equation (E.29), to prove lemma E.3 and E.4, it is sufficient for us to analyze $\{\mathbb{E}_{\mathcal{G}}[f(r_x(v))]\}_{x\in\Omega}$ is close to $\{f(pos(x))\}_{x\in\Omega}$ for some fixed node $v \in V$. We use the principal of deferred decisions—*We reveal the randomness of the graph $G \sim \mathcal{G}$ after fixing node v and configuration x*, and apply a union bound over all configurations $x \in \Omega$.

Fixing a configuration and node v, let's consider a bin with $Pos(x)$ red balls and $n - Pos(x)$ blue balls, if we sample k balls without replacement, the expected number of red balls among those k ball is $pos(x) \cdot k$, and this random number has the same distribution as the random variable $r_x(v) \cdot k$ if $G \sim \mathcal{G}$ is conditioned on the degree of v being k.

We define $E_x(\delta_r; v)$ to be the event

$$E_x(\delta_r; v) \triangleq \{G : |r_x(v) - pos(x)| \leq \delta_r pos(x)\}. \tag{E.30}$$

Since $r_x(v) \cdot k$ can be seen as a sample without replacement, a standard argument combining theorems 2.17 and 2.22 upper bounds the probability of it deviating from expectation by the one of sampling with replacement.

$$\Pr_{\mathcal{G}}[E_x(\delta_r; v)|deg(v) = k] \leq 2\exp\left(-\frac{\delta_r^2 k\, pos(x)}{3}\right) \tag{E.31}$$

Proof of the Lemmas As discussed below equation (E.29), we want to prove the difference between $\mathbb{E}_{\mathcal{G}}[f(r_x(v))]$ and $f(pos(x))$ is of order $O(1/n)$. However, in contrast to the $O(\sqrt{(\log n)/n})$ error in lemma E.4 we need a smoothness property of f around $1/2$ to derived this stronger result. The following two lemmas prove some basic results about smooth functions and conditional variance.

Lemma E.6. *Given $I \subseteq \mathbb{R}$, and X is a random variable with support in I and expectation $\mathbb{E}X$, if $g : \mathbb{R} \mapsto \mathbb{R}$ is M_2-smooth in I, then*

$$\left|\mathbb{E}[g(X)] - g(\mathbb{E}X)\right| \leq \frac{M_2}{2}\left(\mathbb{E}X^2 - (\mathbb{E}X)^2\right).$$

Lemma E.7. *Given a real-valued random variable X and $\epsilon > 0$ such that $\Pr[\mathbb{E}X - \epsilon \leq X \leq \mathbb{E}X + \epsilon)] > 0$, we have*

$$\mathrm{Var}[X|\mathbb{E}X - \epsilon \leq X \leq \mathbb{E}X + \epsilon] \leq \mathrm{Var}[X]$$

Proof of lemma E.6. Let $h(t) \triangleq g(\mathbb{E}X + t(X - \mathbb{E}X))$. Because g is smooth, we use the fundamental theorem of Calculus, and have

$$\begin{aligned}
&\mathbb{E}[g(X)] - g(\mathbb{E}X) = \mathbb{E}_X[g(X) - g(\mathbb{E}X)] = \mathbb{E}\left[\int_0^1 h'(t)dt\right] \\
&= \mathbb{E}\left[\int_0^1 g'\big(\mathbb{E}X + t(X - \mathbb{E}X)\big)(X - \mathbb{E}X)dt\right]
\end{aligned}$$

Because g is M_2-smooth, we have for all $a, a + b \in I$ $g'(a) - M_2|b| \leq g'(a + b) \leq g'(a) + M_2|b|$ and by taking $a = \mathbb{E}X$ and $b = t(X - \mathbb{E}X)$

$$\begin{aligned}
&\mathbb{E}_X[g(X)] - g(\mathbb{E}X) \\
&\leq \mathbb{E}_X\left[\int_0^1 \left(g'(\mathbb{E}X) + M_2 t(X - \mathbb{E}X)\right)(X - \mathbb{E}X)dt\right] \\
&= \mathbb{E}_X\left[g'(\mathbb{E}X)(X - \mathbb{E}X) + \frac{M_2}{2}(X - \mathbb{E}X)^2\right] \\
&= \frac{M_2}{2}\mathbb{E}_X\left[(X - \mathbb{E}X)^2\right]
\end{aligned}$$

The lower bound $-\frac{M_2}{2}\mathbb{E}_X\left[(X - \mathbb{E}X)^2\right] \leq \mathbb{E}_X[g(X)] - g(\mathbb{E}X)$ is can be derived similarity. □

Proof of lemma E.7. Let A the event that X is in the interval $[\mathbb{E}X - \epsilon, \mathbb{E}X + \epsilon]$

$$\begin{aligned}
&\mathrm{Var}[X|(1 - \delta_r)\mathbb{E}X \leq X \leq (1 + \delta_r)\mathbb{E}X] && \text{(E.32)} \\
&= \mathrm{Var}[X|A] \\
&= \mathbb{E}\left[\big(X - \mathbb{E}[X|A]\big)^2|A\right] \\
&\leq \mathbb{E}\left[\big(X - \mathbb{E}[X]\big)^2|A\right] && \text{(E.33)}
\end{aligned}$$

The last inequality is true, because for all z, $\mathbb{E}\left[(Z - z)^2\right] \geq \mathbb{E}\left[(Z - \mathbb{E}Z)^2\right]$.

On the other hand, $\mathrm{Var}[X]$ is equal to

$$\mathbb{E}\left[(X - \mathbb{E}[X])^2\middle|\, A\right]\Pr[A] + \mathbb{E}\left[(X - \mathbb{E}[X])^2\middle|\, \neg A\right](1 - \Pr[A])$$

Because $|X - \mathbb{E}[X]| \geq \epsilon$ conditioned on $\neg A$ and $\epsilon \geq |X - \mathbb{E}[X]|$ if A happens,

$$\mathbb{E}\left[(X - \mathbb{E}[X])^2\middle|\, A\right] \leq \mathrm{Var}[X] \qquad \text{(E.34)}$$

The proof is completed by combining (E.33), and (E.34). □

Proof of lemma E.3. Following equation (E.29), our goal is to derive a better approximation of

$$|\mathbb{E}_{\mathcal{G}}[f(r_x(v))] - f(pos(x))|$$

We take $\check{\epsilon}$ small enough so that $[1/2 - 2\check{\epsilon}, 1/2 + 2\check{\epsilon}] \subseteq I_{1/2}$, and so by the definition the update function f is $\check{M}_2$-smooth function in $[1/2 - 2\check{\epsilon}, 1/2 + 2\check{\epsilon}]$. Moreover we take constants δ_r, δ_d to that $\delta_r \leq \check{\epsilon}$ and $\delta_d < 1$.

Let $\mathcal{E}$ be the event $E_x(\delta_r; v) \wedge E(\delta_d; v)$ defined in equation (E.30) and lemma 2.46 respectively. Informally, if $\mathcal{E}$ happens that means the value of $r_x(v)$ is close to its expectation $pos(x)$ and the degree is nearly uniform. Therefore we can decompose $\mathbb{E}_{\mathcal{G}}[f(r_x(v))] - f(pos(x))$ as follows

$$\begin{aligned}
&|\mathbb{E}_{\mathcal{G}}[f(r_x(v))] - f(pos(x))| \\
&\leq |\mathbb{E}_{\mathcal{G}}[f(r_x(v))|\mathcal{E}] - f(pos(x))| + \Pr_{\mathcal{G}}[\neg\mathcal{E}] \\
&\leq \left|\mathbb{E}_{\mathcal{G}}\left[f(r_x(v))|\mathcal{E}\right] - f\left(\mathbb{E}_{\mathcal{G}}[r_x(v)|\mathcal{E}]\right)\right| \\
&+ \left|f\left(\mathbb{E}_{\mathcal{G}}[r_x(v)|\mathcal{E}]\right) - f(pos(x))\right| + \Pr_{\mathcal{G}}[\neg\mathcal{E}]
\end{aligned}$$

Now we want to give upper bounds for these three terms.

For the first term, $|\mathbb{E}_{\mathcal{G}}\left[f(r_x(v))|\mathcal{E}\right] - f\left(\mathbb{E}_{\mathcal{G}}[r_x(v)|\mathcal{E}]\right)|$ only depends on the random variable $r_x(v)|\mathcal{E}$. By the definition of $\mathcal{E}$ the random variable $r_x(v)|\mathcal{E}$ has support in $[(1-\delta_r)pos(x), (1+\delta_r)pos(x)]$ and $pos(x) \in [1/2 - \check{\epsilon}, 1/2 + \check{\epsilon}]$. Therefore the support of $r_x(v)|\mathcal{E}$ is in $I_{1/2}$ and

$$\begin{aligned}
&\left|\mathbb{E}_{\mathcal{G}}[f(r_x(v))|E] - f\left(\mathbb{E}_{\mathcal{G}}[r_x(v)|\mathcal{E}]\right)\right| \\
&\leq \frac{\check{M}_2}{2} \operatorname{Var}_{\mathcal{G}}[r_x(v)|\mathcal{E}] \\
&\leq \frac{\check{M}_2}{2} \operatorname{Var}_{\mathcal{G}}[r_x(v)|E(\delta_d)].
\end{aligned}$$

The first inequality is true because f is smooth in $I_{1/2}$ and lemma E.6. The second comes from lemma E.7. Now we want to upper bound $\operatorname{Var}[r_x(v)|E(\delta_d)]$. Recall that we observed that the random variable $k \cdot r_x(v)|deg(v) = k$ can be seen as sampling balls from bin with a $pos(x)$ fraction of red balls without replacement. Because the variance is a convex function by theorem 2.22, the value of $\operatorname{Var}[r_x(v)|deg(v) = k]$ is upper bounded by the variance of sampling k balls from the same bin with replacement, $\frac{pos(x)(1-pos(x))}{k}$. As a result,

$$\operatorname{Var}_{\mathcal{G}}[r_x(v)|E(\delta_d)] \leq \frac{pos(x)(1-pos(x))}{(1-\delta_d)np}.$$

Because $x \in \Sigma^s$, $1/2 - \check{\epsilon} < pos(x) < 1/2 + \check{\epsilon}$, and δ_d is some constant independent of n

$$|\mathbb{E}_{\mathcal{G}}\left[f(r_x(v))|\mathcal{E}\right] - f\left(\mathbb{E}_{\mathcal{G}}[r_x(v)|\mathcal{E}]\right)| = \frac{1/4 - \check{\epsilon}^2}{(1-\delta_d)p} \cdot \frac{1}{n}. \tag{E.35}$$

For the second term, because the update function f is Lipschitz, it is sufficient to prove an upper bound for $|\mathbb{E}_{\mathcal{G}}[r_x(v)|E] - pos(x)|$. Note that in the properties of $r_x(v)$ we show that $\mathbb{E}_{\mathcal{G}}[r_x(v)] = pos(x)$. By the law of total probability we have, $|\mathbb{E}_{\mathcal{G}}[r_x(v)] - \mathbb{E}_{\mathcal{G}}[r_x(v)|E]| \leq |\mathbb{E}_{\mathcal{G}}[r_x(v)|\neg E] - \mathbb{E}_{\mathcal{G}}[r_x(v)|\mathcal{E}]| \Pr_{\mathcal{G}}[\neg\mathcal{E}]$ which is less than $\leq 2\Pr_{\mathcal{G}}[\neg\mathcal{E}]$ because $0 \leq r_x(v) \leq 1$. Therefore we have

$$|\mathbb{E}_{\mathcal{G}}[r_x(v)|\mathcal{E}] - pos(x)| \leq 2 \Pr_{\mathcal{G}}[\neg\mathcal{E}]. \tag{E.36}$$

For the last term, $\Pr_{\mathcal{G}}[\neg\mathcal{E}]$ we just use a union bound:

$$\begin{aligned}
&\Pr_{\mathcal{G}}[\neg\mathcal{E}] = \Pr_{\mathcal{G}}[\neg E_x(\delta_r; v) \cup \neg\mathcal{E}(\delta_d; v)] \\
&\leq \Pr_{\mathcal{G}}[\neg E_x(\delta_r)|E(\delta_d; v)] + \Pr_{\mathcal{G}}[\neg E(\delta_d; v)] \\
&\leq 2\exp\left(-\frac{1}{3}\delta_r^2(1-\delta_d)np \cdot pos(x)\right) + 2\exp\left(\frac{-\delta_d^2 np}{3}\right).
\end{aligned} \tag{E.37}$$

Equation (E.37) is derived from equation 2.9 and equation (E.31). Because $p > 0$ and $pos(x) \geq 1/2 - \check{\epsilon}$ are constants when $x \in \Sigma^s$ for large enough n we have

$$\Pr_{\mathcal{G}}[\neg\mathcal{E}] \leq \frac{1}{n}. \tag{E.38}$$

Recall that δ_d, δ_r are constants independent of n. Combining equations (E.35), (E.36), and (E.38), we finish the proof with $K_1 = \frac{1/4-\check{\epsilon}^2}{(1-\delta_d)p} + 3$. □

For lemma E.4, we want to prove the difference between $\mathbb{E}_{\mathcal{G}}[f(r_x(v))]$ and $f(pos(x))$ is of order $O(\sqrt{\log n/n})$ which is much weaker than lemma E.3, and we only need to use the Lipschitz properties of update function f, and concentration phenomenon for $r_x(v)$ shown in equation (E.31).

Proof of lemma E.4. Let $\mathcal{E}$ be the event of $E_x(\delta_r; v) \wedge E(\delta_d; v)$ defined in equation (E.30) and lemma 2.46 respectively. Informally, if $\mathcal{E}$ happens that means the value of $r_x(v)$ is close to expectation $pos(x)$ and the degree is nearly uniform. Therefore we can decompose $\mathbb{E}_{\mathcal{G}}[f(r_x(v))] - f(pos(x))$ as follows

$$\begin{aligned}
&|\mathbb{E}_{\mathcal{G}}[f(r_x(v))] - f(pos(x))| \\
&\leq |\mathbb{E}_{\mathcal{G}}[f(r_x(v))|\mathcal{E}] - f(pos(x))| + \Pr_{\mathcal{G}}[\neg\mathcal{E}]
\end{aligned} \tag{E.39}$$

The first term $|\mathbb{E}_{\mathcal{G}}[f(r_x(v))|\mathcal{E}] - f(pos(x))|$ Since the update function f is Lipschitz with Lipschitz constant M_1, if the event $\mathcal{E}$ happens $|r_x(v) - pos(x)| \leq \delta_r pos(x)$ and,

$$\begin{aligned}
&|f(r_x(v)) - f(pos(x))| \leq M_1 \cdot |r_x(v) - pos(x)| \\
&\leq M_1 \cdot \delta_r pos(x) \leq M_1 \delta_r
\end{aligned}$$

By taking $\delta_r = A\sqrt{\log n/n}$ for some constant A which will be specified later we have

$$|\mathbb{E}_{\mathcal{G}}[f(r_x(v))|\mathcal{E}] - f(pos(x))| = M_1 A\sqrt{\frac{\log n}{n}} \tag{E.40}$$

For the second term by equation (E.37), $\Pr_{\mathcal{G}}[\neg\mathcal{E}]$ is smaller than

$$2\exp\left(-\frac{\delta_r^2(1-\delta_d)np\cdot pos(x)}{3}\right) + 2\exp\left(\frac{-\delta_d^2 np}{3}\right)$$

because $pos(x) \geq \hat{\epsilon} = \Omega(1)$ when $x \in \Sigma^m$. If δ_d is some small constant and $\delta_r = A\sqrt{\log n/n}$, then by taking A large enough $\Pr_{\mathcal{G}}[\neg\mathcal{E}]$ is smaller than $2\exp\left(-\frac{A^2(1-\delta_d)p\cdot pos(x)}{3}\log n\right) + 2\exp\left(\frac{-\delta_d^2 np}{3}\right)$ Therefore

$$\Pr_{\mathcal{G}}[\neg\mathcal{E}] \leq \sqrt{\frac{\log n}{n}} \tag{E.41}$$

Combining equation (E.40) and (E.41) into equation (E.39), and have

$$|\mathbb{E}_{\mathcal{G}}[f(r_x(v))] - f(pos(x))| \leq (M_1 A + 1)\sqrt{\frac{\log n}{n}}$$

and the proof is completed by taking $K_2 = (M_1 A + 1)$. □

Proof of lemma E.5. Given a fixed configuration $x \in A$, random variable $D_G(x)$ has expectation $D(x)$ with randomness over $\mathcal{G}$. Assuming the following claim which we will later prove:

Claim E.8. *If δ_d is some fixed constant, there exists some constant $K > 0$ such that for all $t > 1/\sqrt{n}$, then*

$$\Pr_{\mathcal{G}}[D_G(x) - \mathbb{E}_{\mathcal{G}} D_G(x) < -Kt|E(\delta_d)] \leq \exp\left(-n^2t^2\right)$$

By taking $t = \sqrt{2\ln 2/n}$

$$\begin{aligned}&\Pr_{\mathcal{G}}\left[D_G(x) - \mathbb{E}_{\mathcal{G}} D_G(x) < -\frac{K\sqrt{2\ln 2}}{\sqrt{n}}\Big|E(\delta_d)\right] \\ &\geq 1 - \exp\left(-2n\ln 2\right) = 1 - \frac{1}{4^n}.\end{aligned} \tag{E.42}$$

Apply a union bound over all configurations $x \in \Omega = \{0,1\}^n$ we will derived a high

probability result with $L = K\sqrt{2\ln 2}$

$$\begin{aligned}
&\Pr_{\mathcal{G}}\left[\forall x, D_G(x) - \mathbb{E}_{\mathcal{G}} D_G(x) \geq -\frac{K\sqrt{2\ln 2}}{\sqrt{n}}\right] \\
&\geq \Pr_{\mathcal{G}}\left[\forall x, D_G(x) - \mathbb{E}_{\mathcal{G}} D_G(x) \geq -\frac{K\sqrt{2\ln 2}}{\sqrt{n}} \Big| E(\delta_d)\right] \\
&- \Pr_{\mathcal{G}}[\neg E(\delta_d)].
\end{aligned}$$

By union bound, it is greater than

$$\begin{aligned}
&1 - 2^n \Pr_{\mathcal{G}}\left[D_G(x) - \mathbb{E}_{\mathcal{G}} D_G(x) \geq -\frac{K\sqrt{2\ln 2}}{\sqrt{n}} \Big| E(\delta_d)\right] \\
&- \Pr_{\mathcal{G}}[\neg E(\delta_d)].
\end{aligned}$$

By equation (E.42), this is lower bounded by

$$1 - 2^n \cdot 4^{-n} - \Pr_{\mathcal{G}}[\neg E(\delta_d)] = 1 - o(1).$$

Therefore, it is sufficient for us to prove claim E.8.

Following the analysis in equation (E.29), if $pos(x) > 1/2$,

$$D_G(x) = \frac{1}{n}\sum_{v \in V} f(r_x(v)) - pos(x)$$

Now we think of $\{f(r_{G,x}(v))\}_{x\in\Omega, v\in V}$ as a set real-valued functions with input G indexed by x and v. Similarly we think $\{D_G(x)\}_{x\in\Omega}$ as a set of real-valued functions with input G. We will apply Theorem 2.19 with event $E(\delta_d)$ to prove claim E.8 which consists of two parts: showing the maximum effect/Lipschitz constant c_i is small, and showing the event $E(\delta_d)$ happens with high probability so that it does not change the expectation too much.

For the first part, recall that the update function f is M_1-Lipschitz. Because given x, v if the degree of the node v is k then adding/removing a single edge in G changes the value of $r_{G,x}(v)$ by at most $1/k$, $r_{G,x}(v)$ is $1/k$-Lipschitz. Therefore the Lipschitz constants of $\{f(r_{G,x}(v))\}_{x\in\Omega, v\in V}$ are uniformly bounded by $O(M_1/k) = O(1/k)$. Moreover fixing x if every node have degree at least k, adding/removing a single edge in G only affects two endpoints, and changes the value of $\frac{1}{n}\sum_{v\in V} f(r_x(v))$ by at most $O(\frac{1}{nk})$.

As a result, if $E(\delta_d)$ happens, every node has nearly uniform degree with constant δ_d. For all G, G' in $E(\delta_d)$ that differ in just the presence of a single edge e, we can take $c_e = \max_{G,G'} |D_G(x) - D_{G'}(x)|$ and

$$c_e = O\left(\frac{1}{n \min_{v\in V} deg(v)}\right) = O\left(\frac{1}{n^2}\right) \tag{E.43}$$

Therefore, there exists some constant $\xi > 0$ such that $\sum_e c_e^2 = \xi/n^2$ and $0 \le D_G(x) \le 1$, so we can apply theorem 2.19 and

$$\Pr_{\mathcal{G}}\left[D_G(x) - \mathbb{E}_{\mathcal{G}} D_G(x) < -t' - \Pr_{\mathcal{G}}[\neg E(\delta_d)] | E(\delta_d)\right] \le \exp\left(-\frac{2}{\xi} n^2 t'^2\right)$$

Note that by equation (2.10) when δ_d is some fixed constant and n is large enough $\Pr_{\mathcal{G}}[\neg E(\delta_d)] \le 1/\sqrt{n}$, and we finish the proof of equation (E.8) by taking $K \ge \sqrt{\xi/2} + 1$ and $t \ge 1/\sqrt{n}$. □

E.2.2.2 Drift in Σ^l

Here we consider the phase of the process when the fraction of red nodes is almost 1. The laziness $1/p_G^+(x)$ should be roughly the inverse of the fraction of blue nodes and increases as the bias increases. As a result to prove equation (9.17) we need to give a better lower bound for the drift $D_G(x)$.

Lemma E.9. *There exists small enough constants $\delta_d > 0$, $\hat{\epsilon} > 0$, and $K_3 > 0$. If G has nearly uniform degree, $E(\delta_d)$, such that $D_G(x) \ge K_3(1/2 - bias(x))$ for all $x \in \Sigma^l$.*

The following proof is basically a counting argument: when $x \in \Sigma^l$ the number of red nodes is so small for any node to ahve a majority of red neighbors.

Proof. Without lose of generality, we only consider configurations x where $pos(x) < \hat{\epsilon}$ and $pos(x) = 1/2 - bias(x)$. Given p, δ_d we can take $\hat{\epsilon}$ small enough such that $\frac{\hat{\epsilon}}{(1-\delta_d)p} \in I_0$. Because there are at most $\hat{\epsilon} n$ red nodes and for all $v \in V$ $deg(v) \ge (1-\delta_d)np$, we have $r_x(v) \in I_0$ and by the property of update function

$$f(r_x(v)) \le \hat{M}_1 \cdot r_x(v) < r_x(v) \tag{E.44}$$

If we define $R_x = \{u \in V : x(u) = 1\}$ to be the set of red nodes, by similarly to equation (E.29)[1] we have

$$p_G^+(x) - p_G^-(x) = pos(x) - \frac{1}{n}\sum_{v \in V} f(r_x(v)).$$

[1] In contrast to equation (E.29) where $pos(x) > 1/2$, here $pos(x) < 1/2$

By the equation (E.44), this is greater than

$$\begin{aligned}
&\geq pos(x) - \frac{1}{n}\sum_{v\in V}\hat{M}_1 r_x(v) \\
&= pos(x) - \frac{\hat{M}_1}{n}\sum_{v\in V}\frac{e(S_x, v)}{deg(v)} \\
&\geq pos(x) - \frac{\hat{M}_1}{n}\frac{e(S_x, V)}{\min_{v\in V} deg(v)}.
\end{aligned}$$

The last is true because $deg(v) \geq \min_{v\in V} deg(v)$ and $\sum_v e(S_x, v) = e(S_x, V)$. Because $\sum_v e(S_x, v) \leq |S_x| \max_{u\in S_x} deg(u)$, and $\frac{|S_x|}{n} = pos(x)$,

$$\begin{aligned}
p_G^+(x) - p_G^-(x) &\geq \left(1 - \frac{\hat{M}_1 \max_{u\in S_x} deg(u)}{\min_{v\in V} deg(v)}\right) pos(x) \\
&\geq \left(1 - \frac{1+\delta_d}{1-\delta_d}\hat{M}_1\right) pos(x) \\
&> K_3 pos(x) = K_3(1/2 - bias(x))
\end{aligned}$$

The last inequality is true by taking δ_d small enough and $0 < K_3 \leq 1 - \frac{1+\delta_d}{1-\delta_d}\hat{M}_1$. □

E.2.3 Proof of Lemma 9.13

Proof. We prove each equation in turn.

Equation (9.15). First, for drift $D_G(x) = p_G^+(x) - p_G^-(x)$ we apply the idea illustrated at the beginning of section E.2.2.1. For the first and second steps, we have

$$\begin{aligned}
&p_G^+(x) - p_G^-(x) - \left(f\left(\frac{1}{2} + bias(x)\right) - \left(\frac{1}{2} + bias(x)\right)\right) \\
&= p_G^+(x) - p_G^-(x) - (p^+(x) - p^-(x)) + (p^+(x) - p^-(x)) \\
&- \left(f\left(\frac{1}{2} + bias(x)\right) - \left(\frac{1}{2} + bias(x)\right)\right)
\end{aligned}$$

By lemma E.3 and E.5, with high probability, this is greater than

$$\geq -\frac{K_1}{n} - \frac{L}{\sqrt{n}} \geq -\frac{K_1 + L}{\sqrt{n}}$$

For the last step, because the update function f satisfies $f'(1/2) \geq \check{M}_1 > 1$, we can take $\check{\epsilon}$ small enough such that for all h such that $0 \leq h < \check{\epsilon}$,

$$f\left(\frac{1}{2} + h\right) - f\left(\frac{1}{2}\right) \geq \frac{\check{M}_1 + 1}{2}h$$

As a result, with high probability we have for all $x \in \Sigma^s$ where $bias(x) < \check{\epsilon}$

$$p_G^-(x) - p_G^+(x) \leq -\frac{\check{M}_1 - 1}{2} bias(x) + \frac{K_1 + L}{\sqrt{n}} \tag{E.45}$$

On the other hand, by equation (9.13), we have

$$1 \leq \frac{1}{p_G^+(x)} < \frac{1}{\frac{\hat{\epsilon}}{2} f\left(\frac{\hat{\epsilon}}{2}\right)} \tag{E.46}$$

Multiplying equation (E.45) by equation (E.46) we have,

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 + -\frac{\check{M}_1 - 1}{\hat{\epsilon} f\left(\frac{\hat{\epsilon}}{2}\right)} bias(x) + \frac{K_1 + L}{\frac{\hat{\epsilon}}{2} f\left(\frac{\hat{\epsilon}}{2}\right)} \frac{1}{\sqrt{n}}$$

which finishes the proof of equation (9.15) by taking $A_1 = \frac{\check{M}_1 - 1}{\hat{\epsilon} f\left(\frac{\hat{\epsilon}}{2}\right)}$, and $B_1 = \frac{2(K_1+L)}{\check{M}_1 - 1}$ which are positive constants.

For equation (9.16). First, for drift $D_G(x) = p_G^+(x) - p_G^-(x)$ using argument similar to the proof of (9.15), we have with high probability by lemmas E.4 and E.5, for all $x \in \Sigma^m$, $p_G^+(x) - p_G^-(x)$ is greater than

$$f\left(\frac{1}{2} + bias(x)\right) - \left(\frac{1}{2} + bias(x)\right) - \frac{K_2}{\sqrt{n}} - \frac{L}{n}. \tag{E.47}$$

Recalled that the update function f is Lipschitz and $\forall 0 < h < 1/2, f(1/2 + h) > 1/2 + h$, we can define its minimum over a compact set $[1/2 + \check{\epsilon}, 1 - \hat{\epsilon}]$

$$0 < \delta_f \triangleq \min_{\check{\epsilon} \leq h \leq 1/2 - \hat{\epsilon}} f\left(\frac{1}{2} + h\right) - \left(\frac{1}{2} + h\right) \tag{E.48}$$

Combining equations (E.47), and (E.48), for large enough n we have with high probability for all $x \in \Sigma^m$

$$p_G^-(x) - p_G^+(x) \leq -\delta_f + \frac{K_2}{\sqrt{n}} + \frac{L}{n} \leq -\frac{\delta_f}{2} \tag{E.49}$$

Multiplying equation (E.49) by equation (E.46) we have,

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 - \frac{\delta_f}{\hat{\epsilon} f\left(\frac{\hat{\epsilon}}{2}\right)}$$

which finishes the proof of equation (9.16) by taking $A_2 = \frac{\delta_f}{\hat{\epsilon} f\left(\frac{\hat{\epsilon}}{2}\right)}$ and $0 < A_2 < 1$.

For equation (9.17), by lemmas E.9 and 9.12, we have $p_G^-(x) - p_G^+(x) \leq -K_3(1/2 -$

$bias(x)$), and $\frac{1}{4}(1/2 - bias(x)) \leq p_G^+(x)$. Therefore

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 - 4K_3$$

This finishes the proof by taking $A_3 = 4K_3$. □

APPENDIX F

Proofs for Blockmodel

F.1 Missing proofs for Section 10.4

Proof of Lemma 10.8. Since (2.1) is a Morse-Smale system and V is a complete Lyapunov function, starting at x_0 we know there exists a fixed point $\beta_i \in \mathrm{Fix}_f$ such that $\liminf_{t\to\infty} d(\varphi(x_0,t),\beta_i) = 0$ and $V(\beta_i) \le V(x_0)$. Therefore, given $r > 0$ a neighborhood of β_i, $B(\beta_i, r)$, there is a constant t such that $\varphi(x_0,t) \in B(\beta_i,r)$.

Moreover by Wormald's method (Theorem 2.7), the random component X_k converges to $B(\beta_i, 2r)$ in $O(n)$ steps with high probability. Fixing N_i we can take r small enough and finish the proof. □

F.1.1 Non attracting fixed points—Lemma 10.9

The rest of this section is devoted to prove the Lemma 10.9. Lemma 10.9 is very similar to the literature of leaving saddle points, and we discuss more details in section 10.6.

The proof of Lemma 10.9 has two parts: we first show the process is constant away from the fixed point β_i within time $T_1 = O(n\log n)$ with high probability in Lemma 10.10, and we use the property of complete Lyapunov function, and show the value of $V(X_{T_1})$ is not much bigger than $V(\beta_i)$. In the second part, we run the process for extra $T_2 = O(n)$ steps. Because the process is far from fixed point, the decrease rate of V is large and $V(X_{T_1+T_2})$ is constantly smaller than $V(\beta_i)$.

To define this two parts formally, We first define several neighborhood of β_i: $N_i \subset B(\beta_i, r/2) \subset B(\beta_i, 3r/4) \subset B(\beta_i, r)$ where $B(\beta_i, r)$ is the open ball with radius β_i and center at β_i. Lemma 10.9 keeps track of the process when it enter the region N_i and stop after leaving $B(\beta_i, r)$. Taking r small enough such that $\overline{B(\beta_i, r)}$ only has a single fixed point β_i. Because the complete Lyapunov function $V \in \mathcal{C}^2$ and $\mathcal{L}_f V(\boldsymbol{x}) < 0$ for all $x \in \overline{B(\beta_i, r)} \setminus B(\beta_i, r/2)$ which is a compact set, there exists $\kappa > 0$ such that

$$\forall x \in \overline{B(\beta_i, r)} \setminus B(\beta_i, r/2),\ \mathcal{L}_f V(\boldsymbol{x}) < -\kappa. \tag{F.1}$$

Fixing r with κ, because f is smooth, there exists D' such that $D' = \max \|f(x)\| + D$ for all $x \in \overline{B(\beta_i, r)}$ which is an upper bound for the movement of the process in

$\overline{B(\beta_i, r)}$. Finally we can take N_i small enough such that

$$\forall \boldsymbol{x} \in N_i,\ \|V(\boldsymbol{x}) - V(\beta_i)\| \leq \frac{\kappa r}{32D'}. \tag{F.2}$$

Proof of Lemma 10.9. Suppose the process starting at $X_0 \in N_i$. Let $V(k) \triangleq V(X_k)$, by Equation (F.2),

$$V(0) \leq V(\beta_i) + \frac{\kappa r}{32D'}$$

We show $T_1 = O(n \log n/\rho)$ time the process starting at N_i leaves β_i: $X_{T_1} \in B(\beta_i, 3r/4) \backslash B(\beta_i, r/2)$ with probability $1 - \rho$ in Lemma 10.10.

Because by direct computation the value of complete Lyapunov function V is a almost a supermartingale , $\mathbb{E}[V(X_{k+1})] \leq V(X_k) + O\left(\frac{1}{n^2}\right)$, by Azuma's inequality(Theorem 2.18), with high probability,

$$V(T_1) \leq V(0) + \frac{\kappa r}{32D'} \leq V(\beta_i) + \frac{\kappa r}{16D'}.$$

By Equation (F.1), $\mathcal{L}_f V(x) \leq -\kappa$ for all $x \in \overline{B(\beta_i, r)} \setminus B(\beta_i, r/2)$, we run the process for additional $T_2 = \frac{rn}{4D'}$ steps then

$$\begin{aligned}
V(T_1 + T_2) &= V(T_1) + \sum_{k=T_1}^{T_1+T_2} V(k+1) - V(k) \\
&= V(T_1) + \sum_{k=T_1}^{T_1+T_2} \left(\frac{d}{dt} V(X_k) + O(\frac{1}{n^2})\right) \frac{1}{n} \\
&\leq V(T_1) + \sum_{k=T_1}^{T_1+T_2} \left(-\kappa + O(\frac{1}{n^2})\right) \frac{1}{n} \\
&\leq V(T_1) - \frac{\kappa r}{4D'} + O\left(\frac{1}{n^2}\right) \\
&\leq V(\beta_i) - \frac{\kappa r}{8D'}
\end{aligned}$$

which shows the process leaves the neighborhood N_i in $O(n \log n/\rho)$ time with probability $1 - \rho$. □

F.1.2 Proofs for Sect. 10.4.1

Proof of Lemma 10.11. This is proved by optional stopping time theorem. Given $Z_0 + \beta_i = X_0 \in N_i$, Suppose T_0 is the stopping time such that $\|Z^u_{T_0}\| \geq 8\|Z^s_{T_0}\|$ or $\|Z_{T_0}\| \leq l_1$. We consider the following random variables $W^s_k \triangleq \left(1 - \frac{\lambda_s}{2n}\right)^{-k} \|Z^s_k\|_2^2$. Suppose W^s_k is a super martingale and r small enough, and by optional stopping time theorem

$$\mathbb{E}[W^s_{T_0}] \leq W^s_0 \leq r^2 \leq 1. \tag{F.3}$$

On the other hand, let $p = \Pr[T_0 \le n \log n]$

$$\begin{aligned}
\mathbb{E}[W^s_{T_0}] &= \mathbb{E}\left[\left(1-\frac{\lambda_s}{2n}\right)^{-T_0} \|Z^s_{T_0}\|_2^2\right] \\
&\ge \mathbb{E}\left[\left(1-\frac{\lambda_s}{2n}\right)^{-T_0} l_1^2/2\right] && \text{(stop before } \|Z_{T_0}\|^2 < l_1^2/2) \\
&\ge p l_1^2/2 + (1-p)\left(1-\frac{\lambda_s}{2n}\right)^{-n\log n} l_1^2/2 \\
&\ge p l_1^2/2 + (1-p) n^{\lambda_s/2} l_1^2/2.
\end{aligned}$$

Therefore combining the equation (F.3) and the above, we have $\Pr[T \le n\log n] = p \ge 1 - \frac{l_1^2-2}{(n^{\lambda_s/2}-1)l_1^2} = 1 - o(1)$.

Now, let's use induction show W^s_T is a supermartingale before stopping time T_0:

$$\begin{aligned}
&\left(1-\frac{\lambda_s}{2n}\right)^{k+1} \mathbb{E}[W^s_{k+1} \mid \mathcal{F}_k] \\
=&\mathbb{E}\left[\|Z^s_{k+1}\|^2 \mid \mathcal{F}_k\right] \\
=&\mathbb{E}\left[\left\|Z^s_k + \frac{1}{n}AZ^s_k + \frac{1}{n}(O(\|Z_k\|^2) + \text{noise}^s)\right\|^2 \mid \mathcal{F}_k\right] \\
=&\|Z^s_k\|^2 + \mathbb{E}\left[\left\langle Z^s_k, \frac{1}{n}AZ^s_k + \frac{1}{n}(O(\|Z_k\|^2) + \text{noise}^s)\right\rangle \mid \mathcal{F}_k\right] + O\left(\frac{1}{n^2}\right) \quad \text{(by (10.6))} \\
=&\|Z^s_k\|^2 + \frac{1}{n}(Z^s_k)^\top AZ^s_k + \frac{1}{n}O(\|Z_k\|^3) + O\left(\frac{1}{n^2}\right) \\
\le&\left(1-\frac{\lambda_s}{n}\right)\|Z^s_k\|^2 + \frac{1}{n}O(\|Z_k\|^3) + O\left(\frac{1}{n^2}\right) \quad \text{(by Corollary 2.14)}
\end{aligned}$$

If we take N_i small enough and n large enough,

$$\left(1-\frac{\lambda_s}{2n}\right)^{k+1} \mathbb{E}[W^s_{k+1} \mid \mathcal{F}_k] \le \left(1-\frac{\lambda_s}{2n}\right)\|Z^s_k\|^2 = \left(1-\frac{\lambda_s}{2n}\right)^{k+1} W^s_k.$$

This completes the proof. □

Proof of Lemma 10.12. Let T_1 be the stopping time that $\|Z_{T_1}\| \ge Cl_1$ for some constant C. We first show the expectation of T_1 is much smaller than τ_1. Then we show the stable component $\|Z^s_k\|$ is small for all $k \le \tau_1$. By union bound on these two event, we show with high probability $\|Z_{T_1}\|$ is large and $\|Z^s\|$ is small before τ_1.

For the first part, because we are in a Euclidean space, the principle angle between E^u and E^s is bounded that is

$$\theta_{us} = \min\{\arccos(|\langle v_s, v_u\rangle|) : v_s \in E^s, \|v_s\| = 1, v_u \in E^u, \|v_u\| = 1\} > 0. \tag{F.4}$$

As a result for all $Z = Z^u + Z^s$, we can lower bound

$$\|Z\| \geq \|Z^u\| / \sin\theta_{us}, \tag{F.5}$$

so it is sufficient to lower bound the magnitude of unstable component, $\|Z^u\|$. Let $a^u_{\text{noise}} \triangleq d_1 \operatorname{Tr}((P^u)^\top P^u) > 0$ and $W_k \triangleq \|Z^u_k\|_2^2 - \frac{a^u_{\text{noise}}}{n^2} k$. If W_k is a submartingale, by optional stopping theorem (Theorem 2.11) $\mathbb{E}[W_{T_1}|\mathcal{F}_0] \geq \mathbb{E}[W_0] \geq 0$ and

$$\mathbb{E}[\|Z^u_{T_1}\|_2^2] \geq a^u_{\text{noise}} \frac{\mathbb{E}[T_1]}{n^2}. \tag{F.6}$$

Therefore by (F.5) and (F.6),

$$\mathbb{E}[T_1] \leq \frac{n^2}{a^u_{\text{noise}}} \mathbb{E}[\|Z^u_{T_1}\|_2^2] \leq \frac{(n\sin\theta_{us})^2}{a^u_{\text{noise}}} \mathbb{E}[\|Z_{T_1}\|_2^2] \leq \frac{(Cn\sin\theta_{us} l_1)^2}{a^u_{\text{noise}}} = O(n\log^{2/3} n).$$

By Markov inequality there exists $\tau_1 = O(n\log n)$ such that $\|Z_k\|$ is greater than Cl_1 for some $k \leq \tau_1$ with probability $1 - 1/(\log n)^{1/3} = 1 - o(1)$.

Now, let's show W_k is a submartingale with respect to $\mathcal{F}_k$ before stopping time T_1. Let $Z^u_{k+1} = Z^u_k + D_k$ where $D_k = \frac{1}{n}(AZ^u_k + O(\|Z_k\|^2) + \text{noise}^u)$:

$$\begin{aligned}
\mathbb{E}[W_{k+1}|\mathcal{F}_k] &= \mathbb{E}\left[\|Z^u_{k+1}\|_2^2 - \frac{a^u_{\text{noise}}}{n^2}(k+1)|\mathcal{F}_k\right] \\
&= \mathbb{E}\left[\langle Z^u_k + D_k, Z^u_k + D_k\rangle|\mathcal{F}_k\right] - \frac{a^u_{\text{noise}}}{n^2}(k+1) \\
&= W_k + 2\mathbb{E}\left[\langle Z^u_k, D_k\rangle|\mathcal{F}_k\right] + \mathbb{E}\left[\langle D_k, D_k\rangle|\mathcal{F}_k\right] - \frac{a^u_{\text{noise}}}{n^2}.
\end{aligned}$$

To prove $\mathbb{E}[W_{k+1}|\mathcal{F}_k] \geq W_k$, it is sufficient to show the following two claims:

$$2\mathbb{E}\left[\langle Z^u_k, D_k\rangle|\mathcal{F}_k\right] - o(1/n^2) \geq 0 \tag{F.7}$$

$$\mathbb{E}\left[\langle D_k, D_k\rangle|\mathcal{F}_k\right] \geq \frac{a^u_{\text{noise}}}{n^2} \tag{F.8}$$

For (F.7), we need to use the fact that A is expanding is subspace of E^u before stopping time,

$$\begin{aligned}
2\mathbb{E}\left[\langle Z^u_k, D_k\rangle|\mathcal{F}_k\right] &= 2\langle Z^u_k, \mathbb{E}\left[D_k|\mathcal{F}_k\right]\rangle \\
&= \frac{2}{n}\langle Z^u_k, AZ^u_k + O(\|Z_k\|^2)\rangle && (\mathbb{E}[\text{noise}] = 0) \\
&= \frac{2}{n}\left((Z^u_k)^\top AZ^u_k + O(\|Z_k\|^3)\right) \\
&\geq \frac{2}{n}\left(\lambda_u\|Z^u_k\|^2 + O(\|Z_k\|^3)\right) && (\text{by Corollary 2.14}) \\
&> \frac{1}{n}O(\|Z_k\|^3) = o(1/n^2). && (\ \|Z^s_k\| = O(\|Z_k\|) = O(l_1))
\end{aligned}$$

For (F.7), we use the variance of noise is bounded below by some constant

$$\begin{aligned}\mathbb{E}\left[\langle D_k, D_k\rangle|\mathcal{F}_k\right] &= \frac{1}{n^2}\|AZ_k^u + O(\|Z_k\|^2)\|^2 + \frac{1}{n^2}\mathbb{E}\left[\langle AZ_k^u + O(\|Z_k\|^2), \text{noise}^u\rangle\right] + \frac{1}{n^2}\mathbb{E}\left[\|\text{noise}^u\|^2\right] \\ &\geq \frac{1}{n^2}\mathbb{E}\left[\|\text{noise}^u\|^2\right] \\ &\geq \frac{1}{n^2}d_1 \operatorname{Tr}((P^u)^\top P^u) = \frac{1}{n^2}a_{\text{noise}}^u \\ &\qquad\qquad \text{(by Lemma F.1 and definition of } a_{\text{noise}}^u)\end{aligned}$$

For the second part, $\|Z_k^s\| = o(l_1)$ for all $k \leq \tau_1$, we can use similar argument in Lemma 10.11 to show it's true with high probability.

Finally because with high probability $T_1 < \tau_1$ such that $\|Z_{T_1}\| \geq Cl_1$ and $\|Z_k^s\| = o(l_1)$ for all $k \leq \tau_1$, we have $\|Z_{T_1}^u\| \geq 2l_1$ which completes the proof. □

Lemma F.1 (projected noise)**.** *Given a d-dimensional random vector $X \in \mathbb{R}^d$, matrices $P, S \in \mathbb{R}^{d\times d}$, and $0 < d_1 < d_2$ where $\mathbb{E}[X] = 0$, $\operatorname{Cov}[X] = S$, P is not the zero matrix, and S is positive definite matrix with $d_1\mathbb{I}_d \prec S \prec d_2\mathbb{I}_d$, then*

$$0 < d_1 \operatorname{Tr}(P^\top P) < \mathbb{E}\left[\|PX\|^2\right] < d_2 \operatorname{Tr}(P^\top P).$$

Proof.

$$\begin{aligned}\mathbb{E}\left[\|PX\|^2\right] =& \mathbb{E}\left[\operatorname{Tr}\left(X^\top P^\top PX\right)\right] \\ =& \mathbb{E}\left[\operatorname{Tr}\left(P^\top PXX^\top\right)\right] \\ =& \operatorname{Tr}\left(P^\top P\mathbb{E}\left[XX^\top\right]\right) && \text{(linearity of trace)} \\ =& \operatorname{Tr}\left(P^\top PS\right) > 0\end{aligned}$$

Because S is positive definite and $P^\top P$ is positive semi-definite and not the zero matrix.

Finally, since $d_1\mathbb{I}_d \prec S$, $S - d_1\mathbb{I}_d$ is positive definite, and

$$\mathbb{E}\left[\|PX\|^2\right] - d_1 \operatorname{Tr}(P^\top P) = \operatorname{Tr}\left(P^\top PS\right) - d_1 \operatorname{Tr}(P^\top P) = \operatorname{Tr}\left(P^\top P\left(S - d_1\mathbb{I}_d\right)\right) > 0.$$

□

Proof of Lemma 10.13. Let $\tau_j = Cn$, T_j be the stopping time, $T_j = \arg\min\{X_t \notin B(\beta_i, \sqrt{r})\}$ given $X_0 = Z_0 + \beta_i$ defined in the statement of Lemma 10.13, and r small enough such that (10.6) holds. Here we abuse the notation and define Z_k as a new process by Equation (10.6) and couple it with the original process until T_j. Therefore, the lemma can be proved with the following are three equations:

1. With very high probability the stopping time T_j is greater than τ_j ,

$$\Pr[T_j > \tau_j] = 1 - o(1/\log n); \tag{F.9}$$

2. The expectation at time τ_j, $\mathbb{E}[Z_{\tau_j}]$, is nice,

$$l_{j+1} \geq 8\|\mathbb{E}[Z^s_{\tau_j}]\| \text{ and } \|\mathbb{E}[Z^u_{\tau_j}]\| > l_{j+1}, \textit{and} \tag{F.10}$$

3. Z_{τ_j} is concentrated

$$\Pr\left[l_{j+1} \geq 8\|Z^s_{\tau_j}\| \text{ and } \|Z^u_{\tau_j}\| > l_{j+1}\right] = 1 - o(1/\log n). \tag{F.11}$$

Before proving these, let's do some computation to gain some intuition. To compute the $\mathbb{E}[Z_{\tau_j}]$ suppose $T_j > \tau_j$ we can use the linear function Ax to approximate $f(x)$ and tower property of expectation:

$$\begin{aligned}
\mathbb{E}[Z_{k+1}] &= \mathbb{E}[\mathbb{E}[Z_{k+1} \mid \mathcal{F}_k]] \\
&= \mathbb{E}\left[\mathbb{E}\left[Z_k + \frac{1}{n}(AZ_k + O(\|Z_k\|^2) + \text{noise}) \mid \mathcal{F}_k\right]\right] \quad \text{(by Equation (10.6))} \\
&= \left(1 + \frac{1}{n}A\right)\mathbb{E}\left[Z_k\right] + \frac{1}{n}\overrightarrow{\mathbb{E}\left[O(\|Z_k\|^2)\right]}.
\end{aligned}$$

Apply this recursive relation repeatedly and we have,

$$\mathbb{E}[Z_{\tau_j}] = \left(1 + \frac{1}{n}A\right)^{\tau_j}\mathbb{E}\left[Z_0\right] + \frac{1}{n}\sum_{k<\tau_j}\left(1 + \frac{1}{n}A\right)^{\tau_j - k}\mathbb{E}\left[\overrightarrow{O(\|Z_k\|^2)}\right]. \tag{F.12}$$

Therefore, suppose the norm $\|Z_k\|^2$ for all $0 \leq k < \tau_j$ are small, the value $\mathbb{E}[Z_{\tau_j}]$ can be approximated by the first term, $\left(1 + \frac{1}{n}A\right)^{\tau_j}\mathbb{E}\left[Z_0\right]$. Formally, it is sufficient to show for all constant $\epsilon > 0$,

$$\|Z_k\|^2 \leq \epsilon\|Z_0\| \text{ for all } 0 \leq k < \tau_j. \tag{F.13}$$

Equation (F.9)**:** We define $W_k \triangleq \left(1 + \frac{2\lambda_{\max}}{n}\right)^{-k}\|Z_k\|_2^2$ where $\lambda_{\max} = \max_{\lambda\in\rho(A)}\Re(\lambda)$ is the maximum real part of eigenvalues of A. By Corollary 2.14 and similar argument in Lemma 10.11, W_k is a supermartingale such that $\mathbb{E}[W_{k+1} \mid \mathcal{F}_k] \leq W_k$.

Let's apply Theorem 2.20 on (W_k). Because for all $k \leq Cn$ $|\|Z_{k+1}\|_2^2 - \|Z_k\|_2^2| = O\left(\frac{1}{n}\right)$ uniformly, $D = O(1/n)$, $c_k = |W_{k+1} - W_k| = O\left(\left(1 + \frac{2\lambda_{\max}}{n}\right)^{-k}\frac{1}{n}\right)$, and $\sum c_i^2 = O\left(\frac{1}{n}\right)$. Let $\delta = \frac{(\log n)^{1/4}}{\sqrt{n}}$. By Theorem 2.20,

$$\Pr\left[\max_{k\leq\tau_j} W_k \geq W_0 + \delta\right] \leq \exp\left(-\frac{\delta^2}{2\sum_{k\leq\tau_j} c_k^2 + D\delta}\right) = \exp\left(-\Omega\left(\sqrt{\log n}\right)\right).$$

Let $\mathcal{E}$ be the good event that $\max_{k\leq\tau_j} W_k < W_0 + \delta$. Note that condition on $\mathcal{E}$, with

probability $\Pr[\mathcal{E}] = 1 - \exp\left(-\Omega\left(\sqrt{\log n}\right)\right)$ we have Equation (F.13) for all $0 \le k \le \tau_j$

$$\|Z_k\|^2 \le \left(1 + \frac{2\lambda_{\max}}{n}\right)^k \left(\|Z_0\|^2 + \delta\right) \le 2r \exp\left(2\lambda_{\max} C\right) \cdot \|Z_0\|. \tag{F.14}$$

Given $\epsilon, C, \lambda_{\max} > 0$, we can take r small enough such that $2r \exp\left(2\lambda_{\max} C\right) \le \epsilon \le 1$. Moreover, this also proves Equation (F.9), because $\|Z_k\| \le \sqrt{\epsilon r} \le \sqrt{r}$.

Equation (F.10)**:** Now we are ready to prove the first part. By Equation (F.12) and (F.14), let $\mathcal{E}$ be the event defined in (F.13) we have

$$\begin{aligned}
&\mathbb{E}[Z_{\tau_j}] - \left(1 + \frac{1}{n}A\right)^{\tau_j} \mathbb{E}\left[Z_0\right] \\
=&\frac{1}{n} \sum_{k<\tau_j} \left(1 + \frac{1}{n}A\right)^{\tau_j - k} \overrightarrow{\mathbb{E}\left[O(\|Z_k\|^2)\right]} && \text{(by (F.12))} \\
=&\frac{1}{n} \sum_{k<\tau_j} \left(1 + \frac{1}{n}A\right)^{\tau_j - k} \left(\mathbb{E}\left[\overrightarrow{O(\|Z_k\|^2)} \mid \mathcal{E}\right] + \mathbb{E}\left[\overrightarrow{O(\|Z_k\|^2)} \mid \neg\mathcal{E}\right]\right) \\
=&\frac{1}{n} \sum_{k<\tau_j} \left(1 + \frac{1}{n}A\right)^{\tau_j - k} \left(\mathbb{E}\left[\overrightarrow{O(\|Z_k\|^2)} \mid \mathcal{E}\right] + \overrightarrow{O\left(\Pr[\neg\mathcal{E}]\right)}\right) \\
& && \text{(by (F.14) and } Z_{\tau_j} \in B) \\
\le& r\overrightarrow{O(\|Z_0\|)} + \overrightarrow{\exp\left(-\Omega\left(\sqrt{\log n}\right)\right)} = \overrightarrow{\epsilon l_j}. && \text{(by } \mathcal{E})
\end{aligned}$$

Therefore, we have if $\|Z_0\| \le l_j$,

$$\mathbb{E}[Z_{\tau_j}] = \left(1 + \frac{1}{n}A\right)^{\tau_j} Z_0 + r\overrightarrow{O(l_j)}, \tag{F.15}$$

and for the unstable component and stable component we have

$$\|\mathbb{E}[Z^u_{\tau_j}]\| \ge \left(1 + \frac{\lambda_u}{n}\right)^{\tau_j} \|Z^u_0\| \ge \exp\left(\lambda_u C\right) l_j - rO(l_j). \tag{F.16}$$

$$\|\mathbb{E}[Z^s_{\tau_j}]\| \le \frac{\exp(-\lambda_s C)}{8} l_j + rO(l_j). \tag{F.17}$$

and the constant of $O(l_j)$ depends on C, dimension d, and the principle angle θ_{us} defined in (F.4). This proves Equation (F.10) by taking C large enough and r small enough.

Equation (F.11)**:** We define a vector-valued Doob martingale,

$$Y_k(Z_0, \ldots, Z_k) = \mathbb{E}[Z_{\tau_j} | Z_0, \ldots, Z_k] \in \mathbb{R}^d. \tag{F.18}$$

and prove Equation (F.11) by using concentration property of vector-valued martingale Y_k (Theorem 2.19 and 2.21). With good event $\mathcal{E}$, we want to bound $\{c_k\}_{0 \le k \le \tau_j}$

the "variability" of each variable $Z_0, \ldots, Z_{\tau_j}$ on the martingale Y_k condition on this good event defined in (F.18),

$$c_k = \sup \left\{ \left\| \mathbb{E}[Z_{\tau_j} | Z_0, \ldots Z_{k-1}, Z_k = z_k, \mathcal{E}] - \mathbb{E}[Z_{\tau_j} | Z_0, \ldots Z_{k-1}, Z_k = z_k', \mathcal{E}] \right\| \right\}.$$

Equivalently, c_k is the 2-norm error with initial difference $\|z_k - z_k'\| = O(1/n)$. Formally by (F.15) and $\mathcal{E}$, we have $c_k = O(1/n)$ for all $k \le \tau_j$ and $\sum_{k=0}^{\tau_j} c_k^2 = O(1/n)$. By concentration property of vector-valued martingale Y_k (Theorem 2.19), for any constant $D' > 0$

$$\Pr\left[\|Z_{\tau_j} - \mathbb{E}[Z_{\tau_j}]\| \ge \frac{l_j}{16D'} \right] \le O\left(\exp\left(-\Omega(n l_j^2)\right)\right) + \Pr[\neg\mathcal{E}] = \exp\left(-\Omega\left(\sqrt{\log n}\right)\right) \tag{F.19}$$

Therefore, by Equations (F.16), (F.17), and (F.19), with probability $1 - \exp\left(-\Omega\left(\sqrt{\log n}\right)\right) = 1 - o(1/\log n)$ we have,

$$\|Z_{\tau_j}^u\| \ge \|\mathbb{E}[Z_{\tau_j}^u]\| - \frac{l_j}{16D'} \ge \left(\exp(\lambda_u C) - O(r) - \frac{1}{16D'}\right) l_j \ge 2l_j = l_{j+1}.$$

This the last inequality can be true by first take D' large, C large, and r small enough. The stable component can be upper bounded as follows

$$\|Z_{\tau_j}^s\| \le \|\mathbb{E}[Z_{\tau_j}^s]\| + \frac{l_j}{16D'} \le \left(\exp(-\lambda_s C) + O(r) + \frac{1}{2D'}\right)\frac{l_j}{8} \le \frac{1}{8} l_j \le \frac{1}{8} l_{j+1}.$$

which proves Equation (F.11). □

F.2 Missing proofs for Sect. 10.5

Proof of Lemma 10.15. We first show there is no fixed point outside $[0,1]^2$, that is the curve γ_1 and γ_2 do not have intersection outside.

Let $(x_1, x_2) \in \gamma_1 \cap \gamma_2$. When $m_0 = f_{k,\mathcal{R}}{}'(0) = 0$, if $p\,x_1 + q\,x_2 \le 0$ or $p\,x_2 + q\,x_1 \le$ by the definition of $\bar{f}_{\text{ND}}$ and γ_1, $(x_1, x_2) = (0,0)$. On the other hand, when $m_0 = f_{k,\mathcal{R}}{}'(0) > 0$, $\bar{f}_{\text{ND}}$ is monotone, the above solution curve can be rewritten with respect to

$$g(z) \triangleq \frac{1}{q}\left(f_{k,\mathcal{R}}{}^{-1}(z) - pz\right) \tag{F.20}$$

$$\begin{aligned} \gamma_1 &= \left\{(x_1, x_2) \in [0,1]^2 : x_2 = g(x_1)\right\} \\ \gamma_2 &= \left\{(x_1, x_2) \in [0,1]^2 : x_1 = g(x_2)\right\} \end{aligned} \tag{F.21}$$

For $x_1 < 0$, because $(x_1, x_2) \in \gamma_1$, $x_2 < x_1$, and because $(x_1, x_2) \in \gamma_2$, $x_2 > x_1$. Therefore there is no fixed point out side $[0,1]^2$.

If $\delta_{\text{symm}} = 1/f_{k,\mathcal{R}}{}'(1/2)$, we want to show $(0,0), (1,1)$ and $(0.5, 0.5)$ are the only intersections between γ_1 and γ_2 in $[0,1]^2$ which by symmetry is enough to show the curve γ_1 is in $R_1 \cup R_3 \cup \{(0,0), (1,1), (0.5, 0.5)\}$. By Definition 9.1, $f_{k,\mathcal{R}}(0) = 0$, $f_{k,\mathcal{R}}(1/2) = 1/2$, and $f_{k,\mathcal{R}}$ is strictly convex in [0.0.5], $g(0) = 0$, $g(0.5) = 0.5$, and g is

strictly concave in $[0, 0.5]$, so for all $x_1 \in (0, 0.5)$,

$$g(x_1) = g\left((1-2x_1)\cdot 0 + 2x_1 \cdot \frac{1}{2}\right) > \left((1-2x_1)\cdot g(0) + 2x_1 \cdot g\left(\frac{1}{2}\right)\right) = x_1, \tag{F.22}$$

and we show γ_1 is above $x_1 = x_2$.

On the other hand, since g is strictly concave and $\mathcal{C}^2$ in $[0, 0.5]$, $g'(x_1) > g'(0.5)$, and $g'(0.5) = \frac{1}{q}\left(\frac{1}{f'(0.5)} - p\right) > -1$, since $p - q < \delta_{\text{symm}} = 1/f'_{\text{ND}}(0.5)$. Thus we have

$$g(x_1) = g(0.5) + \int_{0.5}^{x_1} g'(s)\, ds = 0.5 - \int_{x_1}^{0.5} g'(s)\, ds < 0.5 + (0.5 - x_1),$$

and show

$$g(x_1) + x_1 < 1. \tag{F.23}$$

Combining equations (F.22) and (F.23) we prove the number of fixed points is exactly 3.

For the property of these three fixed points for all p and q. By Definition 2.30 it is sufficient to study the linear approximation of the dynamics at these points:

$$\nabla \bar{F}_{\text{ND}}|_{(x_1,x_2)} = \begin{bmatrix} -1 + p\bar{f}'_{\text{ND}}(p\,x_1 + q\,x_2) & q\bar{f}'_{\text{ND}}(p\,x_1 + q\,x_2) \\ q\bar{f}'_{\text{ND}}(q\,x_1 + p\,x_2) & -1 + p\bar{f}'_{\text{ND}}(q\,x_1 + p\,x_2) \end{bmatrix} \tag{F.24}$$

When $(x_1, x_2) = (0, 0)$, $\nabla \bar{F}_{\text{ND}}|_{(0,0)} = \begin{bmatrix} -1 + p\bar{f}'_{\text{ND}}(0) & q\bar{f}'_{\text{ND}}(0) \\ q\bar{f}'_{\text{ND}}(0) & -1 + p\bar{f}'_{\text{ND}}(0) \end{bmatrix}$ has trace $2(p\bar{f}'_{\text{ND}}(0) - 1)$ and determinant $((p-q)\bar{f}'_{\text{ND}}(0) - 1)(\bar{f}'_{\text{ND}}(0) - 1)$. Thus $\nabla \bar{F}_{\text{ND}}|_{(0,0)}$ has two negative real eigenvalues since ${f_{k,\mathcal{R}}}'(0) < 1$.

Similarly there are two cases for the fixed point $(0.5, 0.5)$: if $1 < {f_{k,\mathcal{R}}}'(0.5) < 1/(p-q)$, the determinant is negative $((p-q)\bar{f}'_{\text{ND}}(0) - 1)(\bar{f}'_{\text{ND}}(0) - 1) < 0$, so $(0.5, 0.5)$ is a saddle point. On the other hand if ${f_{k,\mathcal{R}}}'(0.5) > 1/(p-q)$, $(0.5, 0.5)$ is a repelling point. □

Proof of Lemma 10.16. We first show the number of anti-symmetric fixed points is two, than analyze the property of those fixed points.

Because $p - q > \delta_{\text{symm}}$, we have $g'(0.5) = \frac{1}{q}\left(\frac{1}{f'(0.5)} - p\right) < -1$, so the curve γ_1 overlaps with R_2. Therefore there exists a non-symmetric intersection between γ_1 and the line $x_1 + x_2 = 1$, $(x_1^{(a)}, x_2^{(a)})$ with $x_1^{(a)} \neq x_2^{(a)}$ which is also in the intersection of γ_1 and γ_2 due to the symmetry.

$$\begin{cases} x_1^{(a)} & = \bar{f}_{\text{ND}}\left(p\,x_1^{(a)} + q\,x_2^{(a)}\right) \\ x_2^{(a)} & = \bar{f}_{\text{ND}}\left(p\,x_2^{(a)} + q\,x_1^{(a)}\right) \\ 1 & = x_1^{(a)} + x_2^{(a)} \text{ and } x_1^{(a)} < x_2^{(a)} \end{cases} \tag{F.25}$$

Because f is convex in $[0, 0.5]$, the system only has two anti-symmetric fixed points $(x_1^{(a)}, x_2^{(a)})$ and $(1 - x_1^{(a)}, 1 - x_2^{(a)})$.

Now we want to show the property of these fixed points. Let $\delta = p - q$ and $s^{(a)} = p\, x_1^{(a)} + q\, x_2^{(a)}$ and $t^{(a)} = p\, x_2^{(a)} + q\, x_1^{(a)}$. Rearrange the above equations we have,

$$1 = f_{k,\mathcal{R}}(s^{(a)}) + f_{k,\mathcal{R}}(t^{(a)}) \tag{F.26}$$

$$\frac{p+q}{p-q} = \frac{f_{k,\mathcal{R}}(s^{(a)}) - f_{k,\mathcal{R}}(t^{(a)})}{s^{(a)} - t^{(a)}} \tag{F.27}$$

$$1 = s^{(a)} + t^{(a)} \text{ and } s^{(a)} > t^{(a)} \tag{F.28}$$

Because $1 = x_1^{(a)} + x_2^{(a)}$ and the symmetry of $f_{k,\mathcal{R}}$, we have $\bar{f}'_{\mathrm{ND}}\left(s^{(a)}\right) = \bar{f}'_{\mathrm{ND}}\left(t^{(a)}\right)$ and call it $m^{(a)}(\delta)$. By Equation (F.27) and the convexity of $f_{k,\mathcal{R}}$, as δ increases, the derivative at $s^{(a)}$, $m^{(a)}(\delta)$, decreases. By the monotone property, there exists $\delta_{\mathrm{anti}} > \delta_{\mathrm{symm}}$ such that $m^{(a)}(\delta) < 1$ for all $\delta = p - q < \delta_{\mathrm{anti}}$, and $m^{(a)}(\delta) > 1$ for all $\delta < \delta_{\mathrm{anti}}$.

Using Equation (F.24) the matrix $\nabla \bar{F}_{\mathrm{ND}}|_{(x_1^{(a)}, x_2^{(a)})}$ has the trace $2(pm^{(a)}(\delta) - 1)$ and the determinant $((p-q)m^{(a)}(\delta) - 1)(m^{(a)}(\delta) - 1)$, so

attracting Both eigenvalues are negative, when $m^{(a)}(\delta) < 1$.

saddle One positive and negative eigenvalues, when $\frac{1}{p-q} < m^{(a)}(\delta) < 1$.

Note it is impossible that $\frac{1}{p-q} > m^{(a)}(\delta)$; otherwise, $g'(x_1^{(a)}) < -1$ and implies there are more than two anti-symmetric fixed points contradicting the property of $f_{k,\mathcal{R}}$. □

Proof of Lemma 10.17. Let $(x_1^{(a)}, x_2^{(a)})$ be the anti-symmetric fixed point defined in (F.25). Given p_e, q_e and $\delta_e < \delta_{\mathrm{anti}}$, let $(x_1^{(e)}, x_2^{(e)}) \in R_2$ be the eccentric fixed point such that $x_1^{(e)}$ is the smallest value that greater than $x_1^{(a)}$.

We first characterize the local behavior of $(x_1^{(e)}, x_2^{(e)})$. Because $f_{k,\mathcal{R}}$ is a $\mathcal{C}^2$ function by implicit function theorem, we can parametrize curves (10.11) as $(x_1^{(1)}, x_2^{(1)})$ and $(x_1^{(2)}, x_2^{(2)})$ of γ_1, and γ_2 respectively. Given $\delta_e < \delta_{\mathrm{anti}}$, by Lemma 10.16 $(x_1^{(a)}, x_2^{(a)})$ is a saddle point,

$$m^{(a)}(\delta_e) = \left.\frac{dx_2^{(1)}}{dx_1^{(1)}}\right|_{(x_1^{(a)}, x_2^{(a)})} < 1 < \left.\frac{dx_2^{(2)}}{dx_1^{(2)}}\right|_{(x_1^{(a)}, x_2^{(a)})} = \frac{1}{m^{(a)}(\delta_e)}.$$

By convexity of $f_{k,\mathcal{R}}$ and definition of $(x_1^{(e)}, x_2^{(e)})$ we have

$$\left.\frac{dx_2^{(2)}}{dx_1^{(2)}}\right|_{(x_1^{(e)}, x_2^{(e)})} \leq \left.\frac{dx_2^{(1)}}{dx_1^{(1)}}\right|_{(x_1^{(e)}, x_2^{(e)})} < m^{(a)}(\delta_e) < 1 \tag{F.29}$$

Let $I \subseteq (\delta_e, \delta_{\mathrm{anti}})$ be the set of δ such that the system (F.21) has eccentric fixed points. We want to show the system has an eccentric fixed point when δ is between δ_e and δ_{anti}— $I = (\delta_e, \delta_{\mathrm{anti}})$. Since $(\delta_e, \delta_{\mathrm{anti}})$ is connected, it is sufficient to show the set I is relative open and closed. By the continuity of system (F.21), we know the set I is closed. To show I is open, without loss of generality, we show there is a neighborhood

of δ_e contained in I. Given $(x_1^{(e)}, x_2^{(e)})$ with δ_e, fixing $x_1 = x_1^{(e)}$, let's consider and the movement of $x_2^{(1)}(\delta)$ and $x_2^{(2)}(\delta)$ as δ changes around δ_e where $x_2^{(1)}(\delta)$ (and $x_2^{(2)}(\delta)$) is the highest intersection between $x_1 = x_1^{(e)}$ and γ_1 (γ_2 respectively).

$$\frac{d}{d\delta}\left(x_2^{(1)} - x_2^{(2)}\right) > 0. \tag{F.30}$$

Informally, by Equation (F.21), as δ changes, the curve γ_1 is stretched vertically (x_2 direction) and the movement is proportional to the change rate of δ. On the other hand, γ_2 is stretched horizontally (x_1 direction), and by Equation (F.29) the slope is smaller than 1, so the vertically increment rate is smaller than the rate of δ. Therefore the $x_2^{(1)}(\delta)$ should increase faster than $x_2^{(2)}(\delta)$ in x_2. Now let give a formal argument. Through direct computation on Equation (F.21),

$$\frac{dx_2^{(1)}}{d\delta} = \frac{1}{2(1-\delta)}(x_2^{(1)} - x_1^{(1)}) = \frac{1}{2(1-\delta)}(x_2^{(e)} - x_1^{(e)}).$$

Similarly,

$$\left(1 + \frac{1}{1-\delta}\left(\frac{1}{{f_{k,\mathcal{R}}}'({f_{k,\mathcal{R}}}^{-1}(x_2^{(e)}))} - 1\right)\right)\frac{dx_2^{(2)}}{d\delta} = \frac{1}{2(1-\delta)}(x_2^{(e)} - x_1^{(e)})$$

Therefore, to prove Equation (F.30), it is sufficient to show

$$\left(1 + \frac{1}{1-\delta}\left(\frac{1}{{f_{k,\mathcal{R}}}'({f_{k,\mathcal{R}}}^{-1}(x_2^{(e)}))} - 1\right)\right) > 1. \tag{F.31}$$

This can be proved by taking derivative at Equation (F.21) with respect to $x_1^{(2)}$ and applying Equation (F.29),

$$1 = \left(1 + \frac{1}{1-\delta}\left(\frac{1}{{f_{k,\mathcal{R}}}'({f_{k,\mathcal{R}}}^{-1}(x_2^{(e)}))} - 1\right)\right)\frac{dx_2^{(2)}}{dx_1^{(2)}} < \left(1 + \frac{1}{1-\delta}\left(\frac{1}{{f_{k,\mathcal{R}}}'({f_{k,\mathcal{R}}}^{-1}(x_2^{(e)}))} - 1\right)\right).$$

Now, let's prove the eccentric fixed point is stable. Note that by (F.29) and (F.30), for all $\delta > \delta_e$,

$$0 < \left.\frac{dx_2^{(2)}}{dx_1^{(2)}}\right|_{(x_1^{(e)}, x_2^{(e)})} < \left.\frac{dx_2^{(1)}}{dx_1^{(1)}}\right|_{(x_1^{(e)}, x_2^{(e)})} < 1. \tag{F.32}$$

Rewrite the above inequality in terms of $f_{k,\mathcal{R}}$ we have,

$$1 > \frac{1}{1-\delta}\left(\frac{1}{{f_{k,\mathcal{R}}}'({f_{k,\mathcal{R}}}^{-1}(x_1^{(e)}))} - \delta\right) > \left[\frac{1}{1-\delta}\left(\frac{1}{{f_{k,\mathcal{R}}}'({f_{k,\mathcal{R}}}^{-1}(x_2^{(e)}))} - \delta\right)\right]^{-1} > 0.$$

By Equation (F.24), the matrix $\nabla \bar{F}_{\mathrm{ND}}|_{(x_1^{(e)}, x_2^{(e)})}$ is

$$\begin{bmatrix} -1 + p\bar{f}'_{\mathrm{ND}}(f_{k,\mathcal{R}}{}^{-1}(x_1^{(e)})) & q\bar{f}'_{\mathrm{ND}}(f_{k,\mathcal{R}}{}^{-1}(x_1^{(e)})) \\ q\bar{f}'_{\mathrm{ND}}(f_{k,\mathcal{R}}{}^{-1}(x_2^{(e)})) & -1 + p\bar{f}'_{\mathrm{ND}}(f_{k,\mathcal{R}}{}^{-1}(x_2^{(e)})) \end{bmatrix}.$$

The trace is negative, because $f_{k,\mathcal{R}}{}'(f_{k,\mathcal{R}}{}^{-1}(x_1^{(e)})) < 1$ and $f_{k,\mathcal{R}}{}'(f_{k,\mathcal{R}}{}^{-1}(x_2^{(e)})) < 1/\delta$. The determinant is positive, because $\left(\frac{1}{f_{k,\mathcal{R}}{}'(f_{k,\mathcal{R}}{}^{-1}(x_1^{(e)}))} - \delta\right) \cdot \left(\frac{1}{f_{k,\mathcal{R}}{}'(f_{k,\mathcal{R}}{}^{-1}(x_2^{(e)}))} - \delta\right) > (1-\delta)^2$. Therefore, the $(x_1^{(e)}, x_2^{(e)})$ is a stable fixed point.

□

F.3 Node dynamics on bi-blockmodel

To prove the first part, our proof has two steps: given an arbitrary neighborhood of consensus states Q, the Markov chain X^{ND} reaches Q in $O(n \log n)$ with high probability, and it hits the consensus states in $O(n \log n)$ with constant probability if $X_0^{\mathrm{ND}} \in Q$ when Q small enough. The first one is proved in Lemma F.2 and the second part is proved in Lemma F.4.

Lemma F.2 (Reaching neighborhood Q)**.** *In case* 1 *of Theorem 10.5, given arbitrary neighborhoods $M_0, M_1 \subset [0,1]^2$ such that $(0,0) \in M_0$ and $(1,1) \in M_1$, the hitting time of X^{ND} to set $Q \triangleq M_0 \cup M_1$ is*

$$\Pr[\forall \sigma_0 \in \{0,1\}^n, T_Q(\sigma_0) = O(n \log n)] = 1 - o(1),$$

where $T_Q(\sigma_0)$ denotes the stopping time of S^{ND} such that $pos(S^{\mathrm{ND}}_{T_Q}) \in Q$ from the initial state σ_0.

Proof of Lemma F.2. With Theorem 10.7, and 10.6, X^{ND} reaches a fixed neighborhood of consensus states $(0,0), (1,1)$, Q in $O(n \log n)$ with high probability if the noise is well-behaved:

$$\exists d_1, d_2 > 0, \forall x \in \Omega_X \setminus Q, d_1 \mathbb{I}_d \prec \mathrm{Cov}[\boldsymbol{U}(x)] \prec d_2 \mathbb{I}_d. \tag{F.33}$$

which is proved in Lemma F.3. □

Lemma F.3 (Well-behaved noise)**.** *Given X^{ND} defined in* (10.4)*, there exist $d_1, d_2 > 0$, for all $x \in \Omega_X \setminus Q$,*

$$d_1 \mathbb{I}_d \prec \mathrm{Cov}[\boldsymbol{U}(x)] \prec d_2 \mathbb{I}_d$$

where $U(x) \triangleq n\,(\boldsymbol{X}' - \mathbb{E}[\boldsymbol{X}'])$ condition on $\boldsymbol{X} = x$.

Lemma F.4 (Reaching consensus)**.** *In the first case of Theorem 10.5, there exist $T = O(n \log n)$, neighborhoods M_0, M_1, and Q in $[0,1]^2$ where $(0,0) \in M_0$, $(1,1) \in M_1$, and $Q \triangleq M_0 \cup M_1$, such that for all $pos(\sigma_0) \in Q$*

$$\Pr[T(\sigma_0) \le T] \ge 1/6$$

where $T(\sigma_0)$ denotes the hitting time of S^{ND} to consensus states 0^n or 1^n with initial state σ_0.

Lemma F.5 (Potential wall)**.** *Given a time homogeneous Markov chain S_k with state space Ω, if there exist constants $d \in \mathbb{N}$, $D \in \mathbb{R}_+$ compact sets $Q \subset Q' \subset \mathbb{R}^d$, functions $\phi : \Omega \to \mathbb{R}^d$, and $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^d)$ with an attracting fixed point $\beta_a \in Q$, if $\{\boldsymbol{X}_k\}_{k\in\mathbb{N}} = \{\phi(S_k)\}_{k\in\mathbb{N}}$ is a function of Markov chain in $\mathbb{R}^d$, and*

1. *$\boldsymbol{X}_{k+1} = \boldsymbol{X}_k + \frac{1}{n}\left(f(\boldsymbol{X}_k) + \boldsymbol{U}(S_k)\right)$, and set $\boldsymbol{U}_{k+1} \triangleq \boldsymbol{U}(S_k)$.*
2. *For all $X_k \in \Omega$, $\mathbb{E}[\boldsymbol{U}_{k+1} \mid S_k] = 0$,*
3. *For all $X_k \in \Omega$, $\|\boldsymbol{U}_{k+1}\| \le D$*

For all s_0 with $\phi(s_0) \in Q$ and $T \ge 1$

$$\Pr[\forall k < T, X_k \in Q'] \ge 1 - T\exp(-\Omega(n)).$$

With above four lemmas, we are ready to prove the Theorem 10.5

Proof of Theorem 10.5. For the first part, by Lemma F.2 X^{ND} reaches a fixed neighborhood of consensus states $(0,0), (1,1)$, Q in $O(n\log n)$ with high probability. By Lemma F.4, if Q is small enough, the process hits consensus states in $T_c = O(n\log n)$ with probability at least $1/7$. Therefore

$$\Pr[\forall \sigma_0 \in \{0,1\}^n, T(\sigma_0) \le T_c] \ge 1/7. \tag{F.34}$$

Because the X^{ND} is a Markov chain which is bounded in $B \subseteq [0,1]^2$ and the upper bound in (F.34) is independent of initial states, we can show upper bound the expected hitting time to consensus states. Let E_c be the event of hitting Q in time $T_c = O(n\log n)$, and by Equation (F.34) we have $\Pr[E_c] \ge 1/7$. Because the process X^{ND} is a Markov chain and bounded, we can partition the time into intervals with duration T_c, the process X^{ND} reaches consensus with probability at most $\Pr[E_c]$ and this bounds are independent of different time intervals so the expected number of intervals for E_c to happens is bounded by the expectation of a geometric random variable with success probability $\Pr[E_c]$ which is constant. Therefore, $\mathrm{ME}(K(n,p,q), f_{k,\mathcal{R}}) = O(n\log n)$.

For the second part, by Theorem 10.7 there is an extra attracting fixed point β_a of $\bar{F}_{\mathrm{ND}}$. By Lemma F.5, there exists neighborhoods of β_a, Q and Q' such that for any σ_0 with $\phi(\sigma_0) \in Q$ and $T \ge 1$ $\Pr[X_T \in Q'] \ge 1 - T\exp(-\Omega(n))$. Therefore, with initial state σ_0

$$\Pr[T(G, f_{k,\mathcal{R}}, \sigma_0) \ge k] \ge \Pr[X_k \in Q'] \ge 1 - k\exp(-\Omega(n))$$

Because the hitting time is a non-negative random variable

$$\mathbb{E}[T(G, f_{k,\mathcal{R}}, s_0)] = \sum_k \Pr[T(G, f_{k,\mathcal{R}}, s_0) \ge k] \ge \sum_k 1 - k\exp(-\Omega(n)) = \exp(\Omega(n)).$$

□

Proof of Lemma F.5. Because β_a is a attracting fixed point, all the eigenvalue of $A \triangleq \nabla f|_{\beta_a}$ has negative real part which is called stable matrix (or sometimes Hurwitz matrix), and by Lyapunov theorem there exists a positive definite matrix P such that $PA + A^\top P = -\mathbb{I}_d$. We define $V(x) \triangleq x^\top P x$. Therefore, with Taylor expansion on V and the property of A, we have

$$\begin{aligned}
\mathbb{E}[V(\boldsymbol{X}_{k+1})|S_k] &= \mathbb{E}\left[V(\boldsymbol{X}_k) + \frac{1}{n}\nabla V(\boldsymbol{X}_k)\cdot(f(\boldsymbol{X}_k) + \boldsymbol{U}_{k+1}) + O\left(\frac{1}{n^2}\right) \mid S_k\right] \\
&\le V(\boldsymbol{X}_k) + \frac{2}{n}\boldsymbol{X}_k^\top P \cdot (f(\boldsymbol{X}_k)) + O\left(\frac{1}{n^2}\right) \\
&\le V(\boldsymbol{X}_k) + \frac{2}{n}\boldsymbol{X}_k^\top PA\boldsymbol{X}_k + \frac{L}{n}\|\boldsymbol{X}_k\|^3 \\
&\qquad\qquad (L \text{ bounded because } f \in \mathcal{C}^2 \text{ in } B) \\
&\le V(\boldsymbol{X}_k) + \frac{1}{n}\boldsymbol{X}_k^\top (PA + A^\top P)\boldsymbol{X}_k + \frac{L}{n}\|\boldsymbol{X}_k\|^3 \\
&\le V(\boldsymbol{X}_k) - \frac{1}{n}\|\boldsymbol{X}_k\|^2 + \frac{L}{n}\|\boldsymbol{X}_k\|^3 \qquad (PA + A^\top P = -\mathbb{I}_d)
\end{aligned}$$

Therefore the value $V(X_k)$ is a super martingale and there exists $r > 0$ such that $\mathbb{E}[V(\boldsymbol{X}_{k+1})|S_k] - V(\boldsymbol{X}_k) \le -r$ for all $\boldsymbol{X}_k \in Q'$ when Q' are small enough. Furthermore, because P is positive definite we can take $Q \subset Q'$ small enough such that the potential value has constant separation: $\max_{x\in Q} V(x) < \min_{x\notin Q'} V(x)$.

Suppose there exists $0 \le l \le T$ such that $X_l \notin Q'$. Because $X_0 = \phi(s_0) \in Q$, there exists an interval of time from k to l such that $X_k \in Q$, $X_l \notin Q'$ and $X_\ell \in Q' \setminus Q$ for all $k < \ell < l$, we define this event as E_l. Because each step the process X_k can only increase by $1/n$ and the potential value in Q and outside Q' has constant separation, the time interval is $l - k \ge cn$ for some constant $c > 0$ However, such event E_l happen with probability

$$\Pr[X_l \notin Q'] \le \Pr[E_l] \le \exp(-\Omega(n))$$

by Azuma's inequality. The proof is finished by taking union bound on l. □

Proof of Lemma F.3. By the definition of X^{ND}, given $\boldsymbol{X} = x = (x_1, x_2) \in \Omega_X \setminus Q$, define the difference to be $\boldsymbol{Y} \triangleq n(\boldsymbol{X}' - \boldsymbol{X})$ where $\boldsymbol{Y} = (Y_1, Y_2) \in \{(0,0),(1,0),(-1,0),(0,1),(0,-1)\}$ only have these five possible outcomes, and we can compute these directly:

$$\begin{aligned}
p_1^+(x) &\triangleq \Pr[\boldsymbol{Y} = (1,0) \mid \boldsymbol{X} = x] = \frac{1 - x_1}{2} f_{\mathrm{ND}}(px_1 + qx_2), \\
p_1^-(x) &\triangleq \Pr[\boldsymbol{Y} = (-1,0) \mid \boldsymbol{X} = x] = \frac{x_1}{2}\left(1 - f_{\mathrm{ND}}(px_1 + qx_2)\right), \\
p_2^+(x) &\triangleq \Pr[\boldsymbol{Y} = (0,1) \mid \boldsymbol{X} = x] = \frac{1 - x_2}{2}\left(f_{\mathrm{ND}}(qx_1 + px_2)\right), \\
p_2^-(x) &\triangleq \Pr[\boldsymbol{Y} = (0,-1) \mid \boldsymbol{X} = x] = \frac{x_1}{2}\left(1 - f_{\mathrm{ND}}(qx_1 + px_2)\right).
\end{aligned}$$

We omit x when it is clear. Then by the definition of $\boldsymbol{U}(x)$ and Y,

$$\begin{aligned}
\mathrm{Cov}[U(x)] &= \mathrm{Cov}[n\,(\boldsymbol{X}' - \mathbb{E}[\boldsymbol{X}']) \mid \boldsymbol{X} = x] \\
&= \mathrm{Cov}[n\,(\boldsymbol{X}' - x) \mid \boldsymbol{X} = x] \\
&= \mathrm{Cov}[\boldsymbol{Y} \mid \boldsymbol{X} = x] \\
&= \begin{bmatrix} \mathrm{Var}[Y_1] & \mathbb{E}[Y_1Y_2] - \mathbb{E}[Y_1]\mathbb{E}[Y_2] \\ \mathbb{E}[Y_1Y_2] - \mathbb{E}[Y_1]\mathbb{E}[Y_2] & \mathrm{Var}[Y_2] \end{bmatrix} \\
&= \begin{bmatrix} p_1^+ + p_1^- - (p_1^+ - p_1^-)^2 & -(p_1^+ - p_1^-)(p_2^+ - p_2^-) \\ -(p_1^+ - p_1^-)(p_2^+ - p_2^-) & p_2^+ + p_2^- - (p_2^+ - p_2^-)^2 \end{bmatrix}.
\end{aligned}$$

Let $S_1 = p_1^+ + p_1^-$, $S_2 = p_2^+ + p_2^-$, $D_1 = p_1^+ - p_1^-$, and $D_2 = p_2^+ - p_2^-$, and $\mathrm{Cov}[U(x)]$ can be simplified as,

$$\mathrm{Cov}[U(x)] = \begin{bmatrix} S_1 - D_1^2 & -D_1D_2 \\ -D_1D_2 & S_2 - D_2^2 \end{bmatrix}. \tag{F.35}$$

Because $\mathrm{Cov}[U(x)]$ is symmetric, the eigenvalues are real. By Gershgorin circle theorem and (F.35), the eigenvalues are upper bounded by

$$\max\left\{S_1 - D_1^2 + |D_1D_2|, S_2 - D_2^2 + |D_1D_2|\right\} \le 1,$$

and lower bounded by

$$\min\left\{S_1 - D_1^2 - |D_1D_2|, S_2 - D_2^2 - |D_1D_2|\right\}, \tag{F.36}$$

so to find d_1 it is sufficient to lower bound Equation (F.36).

Because $Q = M_0 \cup M_1$, there exists constant $\alpha > 0$ such that 1-norm balls $\{x \in [0,1]^2 : \|x\|_1 \le \alpha\}$ and $\{x \in [0,1]^2 : \|x - (1,1)\|_1 \le \alpha\}$ are insides Q. Thus, if $(x_1, x_2) \in \Omega_X \setminus Q$, $px_1 + qx_2$, $qx_1 + \bar{b}x_2$ are in $[q\alpha, p(1-\alpha)]$, so

$$0 < f_{\mathrm{ND}}(q\alpha) \le f_{\mathrm{ND}}(px_1 + qx_2) \text{ and } f_{\mathrm{ND}}(qx_1 + px_2) \le f_{\mathrm{ND}}(p(1-\alpha)) < 1 \tag{F.37}$$

As a result, p_1^+, p_1^-, p_2^+ and p_2^- are smaller or equal to $\frac{1}{2}f_{\mathrm{ND}}(p(1-\alpha))$, and $|D_1|, |D_2| \le \frac{1}{2}f_{\mathrm{ND}}(p(1-\alpha))$. Moreover,

$$\begin{aligned}
\text{(F.35)} &\ge \min\left\{S_1 - f_{\mathrm{ND}}(p(1-\alpha))|D_1|, S_2 - f_{\mathrm{ND}}(p(1-\alpha))|D_2|\right\} \\
&\ge (1 - f_{\mathrm{ND}}(p(1-\alpha)))\min\left\{S_1, S_2\right\}.
\end{aligned}$$

Because $S_1 = p_1^+ + p_1^-$ is a convex combination of $f_{\mathrm{ND}}(px_1 + qx_2)/2$ and $(1 - f_{\mathrm{ND}}(px_1 + qx_2))/2$, and $S_2 = p_2^+ + p_2^-$ is a convex combination of $f_{\mathrm{ND}}(qx_1 + px_2)/2$ and $(1 - f_{\mathrm{ND}}(qx_1 + px_2))/2$, by (F.37), $\min\{S_1, S_2\} \ge \frac{1}{2}\min\{f_{\mathrm{ND}}(q\alpha), 1 - f_{\mathrm{ND}}(p(1-\alpha))\}$,

$$\text{(F.35)} \ge (1 - f_{\mathrm{ND}}(p(1-\alpha))) \cdot \frac{1}{2}\min\{f_{\mathrm{ND}}(q\alpha), 1 - f_{\mathrm{ND}}(p(1-\alpha))\} > 0$$

Therefore, we can take $0 < d_1 < \frac{1}{2}(1 - f_{\mathrm{ND}}(p(1-\alpha))) \cdot \min\{f_{\mathrm{ND}}(q\alpha), 1 - f_{\mathrm{ND}}(p(1-\alpha))\}$

and $d_2 = 2$ which completes the proof. □

Proof of Lemma F.6. Let $\psi(k) = \sum_{1 \le \ell \le k} d_\ell$ and $\psi(0) = 0$. By direct computation, for all $0 < k < m$

$$\begin{aligned}\mathcal{L}\psi(k) =& p^+(k)\left(\psi(k+1) - \psi(k)\right) - p^-(k)\left(\psi(k) - \psi(k-1)\right) \\ =& p^+(k)d_{k+1} - p^-(k)d_k && \text{(definition of } \psi) \\ \le& -1 && \text{(definition of } d_k)\end{aligned}$$

Finally, $\mathcal{L}\psi(m) = -p^-(k)\left(\psi(k) - \psi(k-1)\right) - p^-(k)d_k \le -1$. Therefore $\psi(m)$ is a upper bound for the maximum expected hitting time by Corollary 2.10. □

F.3.1 From neighborhood of attracting fixed points to fixed points

In this section, we want to prove Lemma F.4: once the process X^{ND} hits the set Q defined in Lemma F.2 process reaches consensus states with constant probability within $O(n \log n)$ time. We achieve this by coupling the process with a birth-and-death chain. In Lemma F.6, we give a simple upper bound for hitting time of birth-and-death chain. In Lemma F.7, a uniform bound for (F.38) is given for our process.

Lemma F.6 (Hitting time of birth-and-death chains)**.** *Let discrete time Markov chain W_k be a birth-and-death chain on space $\Omega = \{0, 1, \ldots, m\}$ such that in each transition the state can increase or decrease by at most 1 where*

$$\begin{aligned}\Pr[W' = W + 1 \mid W = \ell] &= p^+(\ell) \\ \Pr[W' = W \mid W = \ell] &= 1 - p^+(\ell) - p^-(\ell) \\ \Pr[W' = W - 1 \mid W = \ell] &= p^-(\ell)\end{aligned}$$

Let $d_1, \ldots, d_m$ be a positive sequence such that

$$d_m \ge \frac{1}{p^-(m)} \text{ and } d_{l-1} \ge \frac{1}{p^-(\ell-1)} + \left(\frac{p^+(\ell+1)}{p^-(\ell-1)}\right) d_l \tag{F.38}$$

Then the maximum expected hitting time from state ℓ to 0 can be bounded as follows:

$$\max_{\ell \in \Omega} \mathbb{E}[T_0(x)] \le \sum_{0 < \ell \le m} d_\ell$$

where $T_0(x)$ denotes the hitting time from state x to state 0.

Lemma F.7. *Let $h(\sigma) \triangleq \|Pos(\sigma)\|_1$. There exist positive constants α, γ and ϵ, such that for all S_k^{ND} with $h(S_k^{\mathrm{ND}}) \le \epsilon n$,*

$$\Pr\left[h(S_{k+1}^{\mathrm{ND}}) = h(S_k^{\mathrm{ND}}) - 1 \mid S_k^{\mathrm{ND}} = \sigma_0\right] \ge \gamma h(\sigma_0)/n, \tag{F.39}$$

and

$$\frac{\Pr\left[h(S_{k+1}^{\mathrm{ND}}) = h(S_k^{\mathrm{ND}}) + 1\right]}{\Pr\left[h(S_{k+1}^{\mathrm{ND}}) = h(S_k^{\mathrm{ND}}) - 1\right]} \leq 1 - \alpha. \tag{F.40}$$

Proof of Lemma F.4. Without loss of generality, suppose $pos(\sigma_0) \in M_0$. Consider a function $h : \Omega \to \mathbb{N}_0$ where $h(\sigma) \triangleq \|Pos(\sigma)\|_1$. Let $V_k = h(S_k)$ is a stochastic process on $\mathbb{N}_0$ and the process S_k reaches 0^n if and only $h(S_k^{\mathrm{ND}}) = 0$. With M_0 and h we define $m_0 = \max\{h(\sigma) : pos(\sigma) \in M_0\} = \Theta(n)$.

To show the process hits 0^n in $O(n \log n)$ with probability $1/6$, the proof has two steps: we first upper bound the expected optional stopping time, $T = \min\{k : V_k = 0 \vee V_k \geq 2m_0\}$,

$$\mathbb{E}[T] = \tau' = O(n \log n) \tag{F.41}$$

Then show

$$\Pr[V_T = 0] \geq \Pr[V_T \geq 2m_0] \tag{F.42}$$

With the above two equations, we have

$$\begin{aligned}
\Pr[T \leq 3\tau'] &\geq \Pr[T \leq 3\tau' \wedge V_T = 0] \\
&\geq 1 - \Pr[V_T \neq 0] - \Pr[T \geq 3\tau'] && \text{(union bound)} \\
&\geq 1/2 - 1/3 = 1/6 && \text{(by Markov inequality and (F.42))}
\end{aligned}$$

Now let's prove the Equation (F.41) and (F.42). For Equation (F.41) we couple the process V_k with a birth-and-death chain W_k as follows: W_k is a Markov chain on space $\{0, 1, \ldots, 2m_0\}$, one step the state can increase or decrease by at most 1 such that for all $0 < \ell < 2m_0$

$$\begin{array}{ll}
\Pr[W' = W + 1 \mid W = \ell] & = \max_{\sigma : h(\sigma) = \ell} \Pr[V' = V + 1 \mid V = h(\sigma)] \\
\Pr[W' = W - 1 \mid W = \ell] & = \min_{\sigma : h(\sigma) = \ell} \Pr[V' = V - 1 \mid V = h(\sigma)]
\end{array} \tag{F.43}$$

recalled that we use W' to denote state of single transition of a discrete time Markov chain starting at W. For the boundary states 0 and $2m_0$, we set $\Pr[W' = W + 1 \mid W = 2m_0] = 0$ and $\Pr[W' = W - 1 \mid W = 0] = 0$.

By Lemma F.7 and F.6, the expected hitting time of W_k to state 0 is upper bounded by $\sum_{\ell \leq 2m_0} d_\ell$ where d_ℓ is defined in Lemma F.6. By Lemma F.7, we can set $d_{2m_0} = \frac{n}{\gamma 2m_0} = O(1)$, for all $1 \leq \ell < 2m_0$, $d_\ell = \frac{1}{\gamma \ell} + (1 - \alpha) d_{\ell+1}$. By induction there exists C such that $d_\ell \leq \frac{Cn}{\ell}$ for all $1 \leq \ell \leq 2m_0$. Therefore

$$\mathbb{E}[\min\{k : W_k = 0\}] \leq \sum d_\ell = O(n \log n).$$

By the definition of W_k, we can couple these two process V_k and W_k before the process hits the boundary such that $W_k \geq V_k$ for all $k \leq \tau$. Therefore, we can upper bound $\mathbb{E}[\tau] \leq \mathbb{E}[\min\{k : W_k = 0\}] = O(n \log n)$.

Finally Equation (F.42) is true, because V_k is a supermartingale, $\mathbb{E}[V_{k+1} \mid S_k^{\mathrm{ND}}] \leq V_k$ by Lemma F.7. □

Proof of Lemma F.7. This Lemma shows if the fraction of opinion 1 in V_1 and V_2 is

smaller than α, the number of 1 opinion decrease fast. Given configuration S_k, let a_k, b_k be the number of 1 opinion in V_1, V_2 at time k. Note that the update function $f_{k,\mathcal{R}}$ is smooth and strictly concave in $[0.5, 1]$ and $f_{k,\mathcal{R}}(1) = 1, f_{k,\mathcal{R}}(0.5) = 0.5$, there exists m_1 such that ${f_{k,\mathcal{R}}}'(1) < m_1 < 1$ and for all $0 < 1 - x < \epsilon$

$$f_{k,\mathcal{R}}(x) \leq 1 + m_1(x - 1). \tag{F.44}$$

Similarly there exists m_0 such that ${f_{k,\mathcal{R}}}'(0) < m_0 < 1$ and for all $0 < x < \epsilon$

$$f_{k,\mathcal{R}}(x) \geq m_0 x. \tag{F.45}$$

Let's first prove (F.39). The event that $h(S_{k+1}) = h(S_k) - 1$ is equal at time $k+1$ a node with opinion 1 is chosen and updates its opinion to 0,

$$\begin{aligned}
&\Pr[h(S_{k+1}) = h(S_k) - 1 \mid S_k] \\
=&\frac{a_k}{n}\Pr[v_1 \in V_1 \text{ updates to0}] + \frac{b_k}{n}\Pr[v_2 \in V_2 \text{ updates to0}] \\
=&\frac{a_k}{n}\left(1 - f_{k,\mathcal{R}}\left(p\frac{2a_k}{n} + q\frac{2b_k}{n}\right)\right) + \frac{b_k}{n}\left(1 - f_{k,\mathcal{R}}\left(q\frac{2a_k}{n} + p\frac{2b_k}{n}\right)\right) \\
\geq&\frac{a_k}{n}m_1\left(1 - p\frac{2a_k}{n} - q\frac{2b_k}{n}\right) + \frac{b_k}{n}m_1\left(1 - q\frac{2a_k}{n} - p\frac{2b_k}{n}\right) && \text{(by (F.44))} \\
\geq&\frac{a_k + b_k}{n}m_0\left(1 - 2\epsilon\right) \\
\geq&\frac{m_1}{2n}(a_k + b_k) = \frac{m_1}{2}h(S_k)/n && \text{(if } \epsilon \text{ smaller than } 1/4)
\end{aligned}$$

Therefore this proves (F.39) by taking $0 < \gamma < \frac{m_1}{2}$

For the (F.40), with (F.39), it is sufficient to show there exists δ such that $\Pr\left[h(S_{k+1}) = h(S_k) - 1\right] - \Pr\left[h(S_{k+1}) = h(S_k) + 1\right]$ is greater than $\delta h(S_k)/n$. This can be done by computation

$$\begin{aligned}
&\Pr\left[h(S_{k+1}) = h(S_k) - 1\right] - \Pr\left[h(S_{k+1}) = h(S_k) + 1\right] \\
=&\mathbb{E}[h(S_{k+1})] - h(S_k) \\
=&\mathbb{E}[a_{k+1} + b_{k+1}] - a_k - b_k \\
=&f_{k,\mathcal{R}}(pa_k/n + qb_k/n) + f_{k,\mathcal{R}}(qa_k/n + pb_k/n) \\
\geq&m_0(pa_k/n + qb_k/n) + m_0(qa_k/n + pb_k/n) && \text{(by (F.45))} \\
\geq&m_0(a_k/n + b_k/n) = m_0 h(S_k)/n
\end{aligned}$$

, and these complete the proof for (F.40). □

APPENDIX G

Proofs for Naming game

G.1 Proof of Theorem 11.1

Recall that we want to formulate the naming game as nonhomogenous random walk on $\mathbb{Z}^4$ and relate this nonhomogenous random walk to a correpsonding autonomous system in $\mathbb{R}^4$ to study consensus time. Note that we can use four variables to describe this random process: fraction of $\{0\}, \{1\}$ nodes in two communities by following notations.

A_t	community1	community2
$\{0\}$	$R_1(t)$	$R_2(t)$
$\{1\}$	$B_1(t)$	$B_2(t)$
$\{0,1\}$	$M_1(t)$	$M_2(t)$

Since $n = R_1(t) + B_1(t) + M_1(t) = R_2(t) + B_2(t) + M_2(t)$ for all t, it's sufficient to consider $X(t) = (R_1(t), B_1(t), R_2(t), B_2(t))$ in $\mathbb{Z}^4$ with initial state $X(0) = (n, 0, 0, n)$ and the naming game reaches consensus at T when $X(T) = (n, 0, n, 0)$ or $(0, n, 0, n)$.

We can now define $F(\cdot)$ as the mean field of this system (as in Equation (2.4)):

$$F(X(t)) = \mathbb{E}[X(t+1) - X(t)|X(t)]. \tag{G.1}$$

Our approach to understand the behavior of X is mainly inspired by the stability property of nonlinear autonomous systems. We define $f(\cdot)$ such that $F_n(X) = f(\frac{X}{n})$ and then we can relate the nonhomogeneous random walk X to the solution of $x' = f(x)$ as in (2.5).

Intuitively we will prove that there exists p such that the system has an "undesirable" asymptotically stable points x^* (which will be defined mathematically in appendix)

$$x^* = (r^*, b^*, b^*, r^*)$$

where $r^* = \frac{e^2+\sqrt{-4e+6e^2-e^4}}{2e}$, $b^* = \frac{e^2-\sqrt{-4e+6e^2-e^4}}{2e}$ and $p = \frac{2}{3}(1-e)$ such that the random process $X(t)$ in Equation (G.1) will

1. Reach some region of nx^*.
2. Given $X(T_0)$ is in some region of nx^* the expected consensus time of the corresponding naming game is exponential in the size of each group $\exp(\Omega(n))$.

These two conclusions can be proved by the following two lemmas, respectively and the proof of Theorem 11.1 follows directly from the above two Lemmas.

Lemma G.1. *Given the naming game defined above, if* $0 \leq p < \frac{4-2\sqrt{3}}{3} \approx 0.178$ *given arbitrary constant* $r_a > 0$ *the random walk* $X(t)$ *will converge to* x^**. That is there exist* $T_0 = O(n)$ *such that* $||X(T_0)/n - x^*|| \leq r_a$ *with probability* $1 - O(\frac{\log n}{\exp(\frac{-n}{\log^3 n})})$

Lemma G.2. *Given the naming game defined above, there exists a constant* $p_0 \approx 0.110$ *such that for all* $0 \leq p < p_0$ *there exists some constant* $r_a > 0$ *such that if* $||X(T_0)/n - x^*|| \leq r_a$ *then the consensus time is* $\exp(\Omega(n))$

Now we need to quantify the evolution of this process. Recalled that our naming game defined in (G.1)

$$\begin{aligned}
\mathbb{E}[R_1(t+1) - R_1(t)|X(t)] &= \frac{1}{2}\big\{(1 - \frac{R_1}{n} - 2\frac{B_1}{n} + (\frac{B_1}{n})^2) \\
&+ \frac{p}{2}(\frac{-R_1}{2n} + \frac{B_1}{n} + \frac{R_2}{2n} - \frac{B_2}{n} - (\frac{B_1}{n})^2 - \frac{3R_1B_2}{2n^2} - \frac{B_1R_2}{2n^2})\big\} \\
\mathbb{E}[B_1(t+1) - B_1(t)|X(t)] &= \frac{1}{2}\big\{(1 - \frac{B_1}{n} - 2\frac{B_1}{n} + (\frac{R_1}{n})^2) \\
&+ \frac{p}{2}(\frac{-B_1}{2n} + \frac{R_1}{n} + \frac{B_2}{2n} - \frac{R_2}{n} - (\frac{R_1}{n})^2 - \frac{3B_1R_2}{2n^2} - \frac{R_1B_2}{2n^2})\big\} \\
\mathbb{E}[R_2(t+1) - R_2(t)|X(t)] &= \frac{1}{2}\big\{(1 - \frac{R_2}{n} - 2\frac{B_2}{n} + (\frac{B_2}{n})^2) \\
&+ \frac{p}{2}(\frac{-R_2}{2n} + \frac{B_2}{n} + \frac{R_1}{2n} - \frac{B_1}{n} - (\frac{B_2}{n})^2 - \frac{3R_2B_1}{2n^2} - \frac{B_2R_1}{2n^2})\big\} \\
\mathbb{E}[B_2(t+1) - B_2(t)|X(t)] &= \frac{1}{2}\big\{(1 - \frac{B_2}{n} - 2\frac{B_2}{n} + (\frac{R_2}{n})^2) \\
&+ \frac{p}{2}(\frac{-B_2}{2n} + \frac{R_2}{n} + \frac{B_1}{2n} - \frac{R_1}{n} - (\frac{R_2}{n})^2 - \frac{3B_2R_1}{2n^2} - \frac{R_2B_1}{2n^2})\big\} \\
R_1(0) = n, B_1(0) &= 0, R_2(0) = 0, B_2(0) = n
\end{aligned}$$

has corresponding autonomous differential system as follow.

$$
\begin{aligned}
r_1' &= \frac{1}{2}\big\{(1 - r_1 - 2b_1 + b_1^2 \\
&\quad + \frac{p}{2}(\frac{-1}{2}r_1 + b_1 + \frac{1}{2}r_2 - b_2 - b_1^2 - \frac{3}{2}r_1 b_2 - \frac{1}{2}b_1 r_2)\big\} \\
b_1' &= \frac{1}{2}\big\{(1 - b_1 - 2r_1 + r_1^2 \\
&\quad + \frac{p}{2}(\frac{-1}{2}b_1 + r_1 + \frac{1}{2}b_2 - r_2 - r_1^2 - \frac{3}{2}b_1 r_2 - \frac{1}{2}r_1 b_2)\big\} \\
r_2' &= \frac{1}{2}\big\{(1 - r_2 - 2b_2 + b_2^2 \\
&\quad + \frac{p}{2}(\frac{-1}{2}r_2 + b_2 + \frac{1}{2}r_1 - b_1 - b_2^2 - \frac{3}{2}r_2 b_1 - \frac{1}{2}b_2 r_1)\big\} \\
b_2' &= \frac{1}{2}\big\{(1 - b_2 - 2r_2 + r_2^2 \\
&\quad + \frac{p}{2}(\frac{-1}{2}b_2 + r_2 + \frac{1}{2}b_1 - r_1 - r_2^2 - \frac{3}{2}b_2 r_1 - \frac{1}{2}r_2 b_1)\big\} \\
&r_1(0) = 1, b_1(0) = 0, r_2(0) = 0, b_2(0) = 1
\end{aligned} \tag{G.2}
$$

G.1.1 Proof of Lemma G.2

With Theorem G.3, to prove Lemma G.2, it is sufficient to prove x^* is a stable fixed point.

Proof of Lemma G.2. With Theorem 2.28, it is sufficient to show all the eigenvalues of $A = \frac{\partial f}{\partial x}|_{x=x^*}$ are negative. By elementary computation, the eigenvalues of A are

$$
\begin{aligned}
&\frac{-e-5}{6} - D_1, \frac{-e-5}{6} + D_1 \\
&\frac{e^2-3}{2} - D_2, \frac{e^2-3}{2} + D_2
\end{aligned}
$$

where $p = \frac{2}{3}(1-e)$ and

$$
\begin{aligned}
D_1 &= \frac{1}{6}\sqrt{\frac{(1-e)(-8e^4 - 36e^3 + 7e^2 + 153e + 64)}{e}} \\
D_2 &= \frac{1}{2}\sqrt{(1-e)(-e^3 - 5e^2 + e + 25)}
\end{aligned}
$$

Therefore A is Hurwitz and x^* is asymptotically stable if $e > 0.835$ and $0 \leq p < 0.110$ □

G.1.2 Proof of Lemma G.1

To proof Lemma G.1 we prove two claims:

1. The solution x to the differential equation in (G.2) converges to x^*;

2. the limit behavior of random process in (G.1) can be approximated by x in (G.2), that is $\lim_{n\to\infty} X(nt)/n \approx x(t)$.

With these two claims we can conclude given any $r_a > 0$ there exists t_0 such that $||X(t)/n - x^*|| < r_a$ for all $t > t_0$ with high probability. For the first claim we use Poincare-Bendixon Theorem 2.37 and use Wormald's differential equation method 2.7 to prove the second.

Proof of Lemma G.1. First, by the symmetry of the system and initial conditions $\hat{r}_1 = \hat{b}_2 = 1$ and $\hat{b}_1 = \hat{r}_2 = 0$. we can assume that $\hat{r}_1(t) = \hat{b}_2(t)$ and $\hat{b}_1(t) = \hat{r}_2$ for all $t \geq 0$, and the system of differential equations is equivalent to the following

$$\begin{aligned} r' &= (1 - r - 2b + b^2) + \frac{1-e}{2}(b - r - b^2 - r^2) \\ b' &= (1 - b - 2r + r^2) + \frac{1-e}{2}(r - b - r^2 - b^2) \\ &\text{where } r(0) = 1, \text{ and } b(0) = 0 \end{aligned}$$

where $r(t) = \hat{r}_1(t) = \hat{b}_2(t)$ $b(t) = \hat{r}_1(t) = \hat{b}_2(t)$ and $p = \frac{1-e}{3}$, and the system will have stable fixed point $r^* = \frac{e^2+\sqrt{-4e+6e^2-e^4}}{2e}$ and $b^* = \frac{e^2-\sqrt{-4e+6e^2-e^4}}{2e}$, and we take

$$x^* = (r^*, b^*, b^*, r^*)$$

Note that such x^* exists if $-4 + 6e - e^3 \geq 0$, i.e. $0 \leq p \leq \frac{4-2\sqrt{3}}{3} \approx 0.178$.

To apply Theorem 2.37 in (G.3), we need to show the orbit of (r, b) is bounded and there is no periodic cycle. It is easy to see that $r(t), b(t)$ are bounded in interval $[0, 1]$. Moreover because if $r(t) = b(t)$ for some t then $r(t') = b(t')$ for all $t' \geq t$, we have $r(t) \geq b(t)$. Combining these two observations we have (r, b) is bounded in $\Omega = \{(r, b)|r \geq b, 0 \leq r, b \leq 1\}$. On the other hand, because $\nabla \cdot H = -2 + \frac{1-e}{2}(-2 - 2r - 2b) < 0 \forall (r, b) \in \Omega$ which, by Theorem 2.38, proves there is no closed orbit. Therefore we have proven the first claim: $\lim_{t\to\infty}(r(t), b(r)) = (r^*, b^*)$ by Theorem 2.37. Furthermore in (G.2) we have

$$||x(t) - x^*|| < 0.5 r_a \forall t > t_0. \tag{G.3}$$

For the second claim, we want to show the original process in (G.1) can be approximated by (G.2). It is not hard to show that the process is bounded by $\beta = 1$ and $\gamma = 0$, and by taking $\lambda = O(\frac{1}{\log(n)})$ we have with probability $1 - O(\log n \exp(-\frac{n}{\log^3 n})$

$$X(nt)/n = x(t) + O(\frac{1}{\log n}) \tag{G.4}$$

in terms of each component.

Combining (G.3) and (G.4) we have with probability $1 - O(\log n \exp(-\frac{n}{\log^3 n}))$

$$||X(nt)/n - x^*|| \leq r_a, \forall t > t_0$$

when n is large enough. □

G.2 Main Results

The main idea used to prove both Theorem 11.1 and Theorem 11.2 is to show the existence of a stable fixed point x^* of the solution to differential system (2.5) and then to relate this stable fixed point to the nonhomogeneous random walk (2.4) by showing:

1. Global behaviour: the random process $X(t)$ will initially "converge" to a point corresponding to the stabile fixed point of the autonomous system.

2. Local behaviour: random process $X(t)$ takes exponential time to leave the region corresponding to a regions around stabile fixed point of the autonomous system.

Here we prove a auxiliary theorem for the second part.

Theorem G.3. *If x^* is an asymptotically stable equilibrium of (2.5), given a closed set U containing x^* there exists $r_a > 0$ such that in system (2.4) if $||X(t_0)/n - x^*|| \leq r_a$ then*

$$\mathbb{E}[\arg\min_{\tau > t_0}\{X(\tau) \notin U\} \big| ||X(t_0)/n - x^*|| \leq r_a] = \exp(\Omega(n)).$$

To prove Lemma G.3, we use the second part of Lyapunov's indirect method Theorem 2.28, which shows the existence of a potential function $\hat{V}(x)$ at some region around the asymptotically stable fixed point in system (2.5) such that the value of potential function is strictly decrease along the trajectory. On the other hand, the counterpart of that potential function in (2.4) will be a supermartingale $V(X(t))$ and we use the optional stopping time to show that it takes an exponential time for the supermartingale to increase by constant.

Proof of Lemma G.3. By Theorem 2.28, we know that there exists a potential function V and an open region $U \subseteq \mathcal{D}$ such that $V(x^*) = 0$, and $V(x) > 0$, $\frac{d}{dt}(V(x)) < 0$ for $x \in U \setminus x^*$. Now we consider a random process

$$W(i) = V\left(\frac{X(i)}{n}\right)$$

and the conditional expectation is

$$\begin{aligned}
&\mathbb{E}[W(i+1) - W(i)|X(i)] \\
=&\mathbb{E}[V(\frac{X(i+1)}{n}) - V(\frac{X(i)}{n})|X(t)] \\
=&\nabla V(\frac{X(i)}{n}) \cdot (\frac{\mathbb{E}[X(i+1) - X(i)|X(i)]}{n}) + O(\frac{1}{n^2}) \\
=&\nabla V(\frac{X(i)}{n}) \cdot \frac{f(\frac{X(i)}{n})}{n} + O(\frac{1}{n^2}) \\
=&\frac{1}{n}\frac{d}{dt}V(x)\Big|_{x=\frac{X(i)}{n}} + O(\frac{1}{n^2})
\end{aligned} \tag{G.5}$$

Therefore $W(i)$ is a supermartingale such that $\mathbb{E}[W(i+1) - W(i)|X(i)] < 0$ when $\frac{X(t)}{n} \in U \setminus \{x^*\}$ and n is large enough.

The idea is to use the optional stopping theorem by proving the process $X(t)$ is not likely to pass through the annulus $B_{r_b} \setminus B_{r_a}$ for some properly choosen r_a, r_b. Here we need to use the properties of the potential function V from Theorem 2.28. Note that U is open, there exists $r_b > 0$ such that a open set $B_{r_b} = \{||x - x^*|| < r_b\} \subseteq U$. Because the boundary $U \setminus B_{r_b}$ is compact and V is continuous, there exists $\min_{x \in B_{r_b}} V(x)$ which is denoted as l_b. On the other hand, because $V(x^*) = 0$ and V is continuous, there exists a close set $\bar{B_{r_a}}$ where $0 < r_a < r_b$ such that $l_a = \max_{x \in \bar{B_{r_a}}} V(x) \leq 0.3 l_b$.

Given such r_a, r_b if $X(t_0)/n \in B_{r_a}$ at some time t_0 and the system leaves the stable region U at time $t_1 > t_0$ there exists σ,τ when n is large enough such that

$$\tau = \arg\min_{t_0 < t < t_1} \{X(t)/n \in U \setminus B_{r_b}\}$$

$$\sigma = \arg\max_{t_0 < s < \tau} \{X(s)/n \in B_{r_a}\}$$

which gives us

$$W(\sigma) < 0.5 l_a, \text{and } W(\tau) \geq l_b$$

Moreover by the definition of σ, τ, for all $\sigma \leq t < \tau$ the random process $X(t)$ would stay in the annulus $B_{r_b} \setminus B_{r_a}$. Therefore for all t such that $\sigma \leq t < \tau$, we have $W(t)$ is a strict supermartingale

$$W(t) = \frac{1}{n}\frac{d}{dt}V(x)\Big|_{x=\frac{X(i)}{n}} + O(\frac{1}{n^2}) = \frac{-h}{2n} < 0$$

where constant $-h = \max_{x \in B_{r_b} \setminus B_{r_a}} \frac{d}{dt}V(x)\Big|_x < 0$ since the annulus is compact.

Therefore by standard optional stopping time theorem with initial state $W(\sigma+1)$ where $l_a < W(\sigma+1) < l_b$ the average time for $W(t)$ to hit $W(t) \geq l_b$ is $\exp(\Omega(hn)) = \exp(\Omega(n))$. □

G.3 Proof of Theorem 11.2

We define stubborn node which has different behavior in naming game. A node s is stubborn if its inventory will not change the process $A_t(s) = A_0(s)$ even when it is the speaker or listener, and we call node s is *stubborn node* with $A_0(s)$, and we call other node as *ordinary nodes*. Here we consider that on completed graph if the naming game is already consensus on opinion 1. The Theorem 11.2 gives a way to understand the following question: how many nodes stubborn with opinion 0 do we make in order to change the graph consensus on opinion 0 in polynomial time?

Theorem G.4 (Restate theorem 11.2)**.** *Given the naming game with p fraction of stubborn nodes defined above, there exists a constant $p_0 \approx 0.108$ such that for all $0 \leq p < p_0$ the expected consensus time is $\exp(\Omega(n))$. Additionally, if $p_0 < p \leq 1$ for all $\epsilon > 0$ the fraction of original opinion is smaller than ε after $O(n)$ steps.*

Similar to the proof of theorem 11.1, we formulate this process as nonhomogenous random walk on $\mathbb{Z}^2$ and relate this nonhomogenous random walk to a correpsonding autonomous system in $\mathbb{R}^2$ to study consensus time.

G.3.1 Model Description

Given a completed graph G which has n nodes and the weight of every pair of node is uniform, if every nodes consensus on 1, we want to make p fraction of nodes stubborn on 0, and all the set of stubborn nodes S such that $|S| = pn$. That is $\forall s \in S, A_0(s) = \{0\}$ and for all ordinary node $v \in V(G) \setminus S, A_0(v) = \{0\}$.

Because the symmetry of the completed graph, only the number of stubborn nodes matters, and we apply the same method in theorem 11.1 to simplify the notations. At time t, we define $X(t) = (R(t), B(t))$ as our state of Markov chain where $R(t)$ the number of ordinary node with inventory $\{0\}$, $B(t)$ the number of ordinary node with inventory $\{1\}$ and $M(t)$ be the number of ordinary node with inventory $\{0, 1\}$. Moreover we use $\tilde{n}$ to denote the number of ordinary nodes, $\tilde{n} = |V(G) \setminus S| = (1-p)n$. Here we have

$$\begin{aligned}
&\mathbb{E}[R(t+1) - R(t)|X(t)] \\
=&(1-p)^2(\frac{R}{\tilde{n}}\frac{M}{\tilde{n}} + (\frac{M}{\tilde{n}})^2 - \frac{R}{\tilde{n}}\frac{B}{\tilde{n}}) + p(1-p)\frac{3}{2}\frac{M}{\tilde{n}} \\
&\mathbb{E}[B(t+1) - B(t)|X(t)] \\
=&(1-p)^2(\frac{B}{\tilde{n}}\frac{M}{\tilde{n}} + (\frac{M}{\tilde{n}})^2 - \frac{R}{\tilde{n}}\frac{B}{\tilde{n}}) - p(1-p)\frac{B}{\tilde{n}}
\end{aligned}$$

and the corresponding autonomous differential system is

$$\begin{aligned} r' =& (1-p)^2(rm + m^2 - rb) + p(1-p)\frac{3}{2}m \\ b' =& (1-p)^2(bm + m^2 - rb) - p(1-p)b \end{aligned}$$

G.3.2 Proofs

Similar to theorem 11.1, when $p < 0.108$ it is striaghtforward to show there exists a stable fixed point $x^* \neq (1, 0)$ and derived the following two lemmas to prove the first part of Theorem 11.2.

Lemma G.5. *Given the naming game defined above, there exists* $p_0 \approx 0.108$ *such that for all constant* $0 \leq p < p_0$ *there exists some constant* $r_a > 0$ *such that if* $||X(T_0)/n - x^*|| \leq r_a$ *then the consensus time is* $\exp(\Omega(n))$

Lemma G.6. *Given the naming game defined above, if constant* $0 \leq p < 0.108$ *given arbitrary constant* $r_a > 0$ *the random walk* $X(t)$ *will converge to* x^**. That is there exists* $T_0 = O(n)$ *such that* $||X(T_0)/n - x^*|| \leq r_a$ *with probability* $1 - O(\frac{\log n}{\exp(\frac{-n}{\log^3 n})})$

For the second part of Theorem 11.2, since if $p > p_0$ the consensus point,$c^* = (1, 0)$ is the only fixed point of the system, we can use similar technique in Lemma G.1 and Theorem G.3 to prove given arbitrary small constant $\epsilon > 0$, $b(t) \leq \epsilon$ for $t = O(n)$.