Dynamic Core Community Detection and Information Diffusion Processes on Networks

by

Wei Bao

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
2018

Doctoral Committee:

Professor George Michailidis, Co-Chair
Assistant Professor Emanuel Gull, Co-Chair
Assistant Professor Hui Jiang
Assistant Professor Xiaoming Mao
Professor Mark Newman

Wei Bao

wbao@umich.edu

ORCID iD: 0000-0001-9883-9848

For my family.

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my advisor Professor George Michailidis for his thorough guidance through out the years of my PhD study. Without him, non of the materials in this thesis would be possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Interest in network science has been increasingly shared among various research communities due to its broad range of applications. Many real world systems can be abstracted as networks, a group of nodes connected by pairwise edges, and examples include friendship networks, metabolic networks, and world wide web among others. Two of the main research areas in network science that have received a lot of focus are community detection and information diffusion. As for community detection, many well developed algorithms are available for such purposes in static networks, for example, spectral partitioning and modularity function based optimization algorithms. As real world data becomes richer, community detection in temporal networks becomes more and more desirable and algorithms such as tensor decomposition and generalized modularity function optimization are developed. One scenario not well investigated is when the core community structure persists over long periods of time with possible noisy perturbations and changes only over periods of small time intervals. The contribution of this thesis in this area is to propose a new algorithm based on low rank component recovery of adjacency matrices so as to identify the phase transition time points and improve the accuracy of core community structure recovery. As for information diffusion, traditionally it was studied using either threshold models or independent interaction models as an epidemic process. But information diffusion mechanism is different from epidemic process such as disease transmission because of the *reluctance to tell stale news* and to address this issue other models such as DK model was proposed taking into consideration of the reluctance of spreaders to diffuse

the information as time goes by. However, this does not capture some cases such as the losing interest of information receivers as in viral marketing. The contribution of this thesis in this area is we proposed two new models coined susceptible-informed-immunized (SIM) model and exponentially time decaying susceptible-informed (SIT) model to successfully capture the *intrinsic time value* of information from both the spreader and receiver points of view. Rigorous analysis of the dynamics of the two models were performed based mainly on mean field theory. The third contribution of this thesis is on the information diffusion optimization. Controlling information diffusion has been widely studied because of its important applications in areas such as social census, disease control and marketing. Traditionally the problem is formulated as identifying the set of k seed nodes, informed initially, so as to maximize the diffusion size. Heuristic algorithms have been developed to find approximate solutions for this NP-hard problem, and measures such as k-shell, node degree and centrality have been used to facilitate the searching for optimal solutions. The contribution of this thesis in this field is to design a more realistic objective function and apply binary particle swarm optimization algorithm for this combinatorial optimization problem. Instead of fixating the seed nodes size and maximize the diffusion size, we maximize the profit defined as the revenue, which is simply the diffusion size, minus the cost of setting those seed nodes, which is designed as a function of degrees of the seed nodes or a measure that is similar to the centrality of nodes. Because of the powerful algorithm, we were able to study complex scenarios such as information diffusion optimization on multilayer networks.

CHAPTER I

Introduction

Network (graph) theory dates back to 1736 when Euler formulated the Konigsberg's bridge problem as a graph theoretic one. To do so, he developed basic definitions and concepts of network theory, including the notions of a vertex set V, the corresponding edge set $E \subseteq V \times V$ together with the first tools for their analysis. Since then, research on networks has experienced an explosive growth, especially in the last 10-15 years both from a theoretical standpoint and also in terms of application areas, including technical work in mathematics and physics, and applications in the social, health and information sciences. Examples include protein-protein interaction networks, metabolic networks and food webs in biology, friendship networks (Facebook, Twitter) in the social sciences, communication networks and the World Wide Web in computer science and various networks in many other fields such as politics, economics and finance [23, 51, 68, 7, 41]. Further, a number of social dynamics mechanisms such as disease transmission, rumor spreading, information diffusion, new product advertisement, and political propaganda among others can be mapped to corresponding stochastic processes on networks [21, 11, 33, 105]. In the following sections, we review some background fundamentals of network theory that the work on subsequent chapters makes extensive use of.

1.1 Network Basics

1.1.1 Mathematical formulations

A very popular way to represent networks is by its adjacency matrix A, where

$$A_{ij} = \begin{cases} 1, & \text{if node } i \text{ and node } j \text{ are connected} \\ 0, & \text{otherwise} \end{cases} \tag{1.1}$$

This representation applies to undirected binary networks, for which the adjacency matrix is symmetric. However in general, networks can have both directed and weighted connections, in which case we can set the element A_{ij} as the weight of the connection from node j to i. If the network is sparse (few connections among its nodes), using the adjacency matrix as the representation method is very expensive in terms of storage and memory in calculations; therefore, two alternative ways have been proposed, the adjacency list and the edge list representations, respectively. In the former representation, we store for each node its neighbors in list form, while for the edge list representation, we store only the edge information (which nodes are connected).

Further, in many applications the actual network can be partitioned to many network structures, thus giving rise to a *multilayer* network structure [11]. The family of multilayer networks can be formally defined as a pair $(\mathbf{G}, \mathbf{C})$, where $\mathbf{G}$ is a family of single layered networks $\mathbf{G} = \{G_\alpha, \alpha \in \{1, \cdots, M\}\}$ with $G_\alpha = (V_\alpha, E_\alpha)$ representing the layer α, and assuming there are M layers in total. On the other hand,

$$\mathbf{C} = \{E_{\alpha\beta} \subseteq V_\alpha \times V_\beta; \alpha, \beta \in \{1, \cdots, M\}, \alpha \neq \beta\} \tag{1.2}$$

is the set of cross layer connections. If we look at the multilayer networks from the

viewpoint of single layer, then the projection network $(\mathbf{V}_M, \mathbf{E}_M)$is

$$\begin{aligned} \mathbf{V}_M &= \bigcup_{\alpha=1}^{M} V_\alpha \\ \mathbf{E}_M &= \Big(\bigcup_{\alpha=1}^{M} E_\alpha\Big) \bigcup \Big(\bigcup_{\alpha,\beta=1,\alpha\neq\beta}^{M} E_{\alpha\beta}\Big) \end{aligned} \tag{1.3}$$

A multiplex network is a special type of a multilayer one and admits its mathematical representation from the multilayer network, but with $V_\alpha = V_\beta, \alpha, \beta \in \{1, \cdots, M\}, \alpha \neq \beta$ and $E_{\alpha,\beta} = \{(v, v); v \in V\}$ [118, 11]. In other words, the set of nodes are exactly the same across different layers and the cross layer connections exist only between each node and its counterparts from different layers. If we look at a multiplex network, we might even think of it as an ordinary single network with multi-edges between nodes, with each edge representing a channel from a certain layer.

A temporal network, or time series networks, can also be represented as a multilayer network with $\mathbf{G} = \{G_t, t \in \{1, \cdots, T\}\}$ and G_t represents the slice of network at time t. The cross layer connections become $E_{\alpha\beta} = \emptyset$ if $\beta \neq \alpha + 1$ and $E_{t,t+1} = \{(v, v), v \in V_t \bigcap V_{t+1}\}$. This formulation is fairly general, since both the nodes and edges across different layers over time are allowed to evolve.

1.1.2 Random networks generation models

There has been a lot of work on defining simple network generation models, whose behavior mimics those observed in the real world. Some of the most popular random network generation models include the Erdos-Renyi (ER) model, the Gilbert model, the onfiguration model, and various preferential attachment models, of which the Barabasi-Albert (BA) model is one of the most extensively studied and frequently used [45, 37, 8]. On the other hand, if we are interested in generating synthetic networks with an embedded community structure, the stochastic block model is the most popular one. The latter was generalized to capture the heterogeneity of the

degree distribution observed in real life networks [59].

In the **ER model**, we fix the number of nodes N and the total number of connections M and assume that all networks with these two parameter values are realized with the same probability. In practice, we place the M edges between nodes randomly and each pair gets the edge with the same probability. In the **Gilbert model**, each pair of nodes are connected with a fixed probability p, and the degree distribution of nodes follow a Binomial distribution. When we have $N \to \infty$, $p \to 0$ and Np converges to a constant as $\frac{2M}{N}$, the degree distribution of both the ER and Gilbert models follows a Poisson distribution with mean degree $\langle K \rangle = \frac{2M}{N}$ for the ER model and $\langle K \rangle = Np$ for the Gilbert model.

The **configuration model** is similar to the ER model, in the sense that it is also a statistical ensemble. On the other hand, the configuration model is much more flexible than the ER model, since it assumes that we are given a sequence of degrees (number of conenctions) for the nodes instead of treating each node equivalently. In practice, we can start from a given degree distribution, say, a power law (Pareto distribution) one given by $P(k) \sim k^{-\gamma}$, then:(1) Generate a list of degrees for the list of nodes and each degree will be attached a stub (half edge); (2) Randomly choose pair of stubs, with each pair being chosen with the same probability, and connect them into edges.

In the **preferential attachment model**, scale free networks with degree following power law distributions can be generated from a linear preference function [34, 67, 8]. To be specific, in practice, we start with m_0 nodes with a certain configuration, and then at each subsequent time step, we add one node and connect to $m \leq m_0$ of the existing nodes with each existing node i been connected with probability as a linear function of its degree: $p(K_i) \propto C + K_i$. Then, the degree distribution of the network obtained when $N \to \infty$ is approximately $P(K) \sim K^{-\gamma}$ with $\gamma = 3 + \frac{C}{m}$, and in particular when $C = 0$ we have $\gamma = 3$, which is exactly the **BA model**.

Naturally, random networks generated by the ER, the Gilbert and the configuration model are all uncorrelated in terms of the degree distribution of neighbors. Interestingly, the BA model is frequently treated as *almost uncorrelated*, due to the anomalously weak degree correlation between first neighbors, defined as the direct neighbors of a node. Under this condition, the joint distribution of the degrees of two neighbors can be factored into the product of the distribution of degree for each neighbor separately.

1.1.3 Giant component and percolation process

The giant component (GC) in a network is defined as the subnetwork of size $N_G \sim O(N)$ that scales up with the size of the network N as $N \to \infty$, in which each node can reach another nodes through edges belonging to the component; further, there can be only one GC. To see this, suppose we have two GCs of size N_1 and N_2 respectively, and for illustrative purposes suppose we operate under the Gilbert model. Then, the probability of a node i from GC1 not connecting to any node in GC2 is $(1-p)^{N_2}$ and therefore, the probability of no connection between the two GCs would be $\prod_{i=1}^{N_1}(1-p)^{N_2} = (1-p)^{N_1 N_2} \overset{N\to 0}{=} 0$. Further, the size of the GC in a random network can be obtained in the following way.

Following an arbitrary edge, the probability of reaching an edge of degree k is

$$\frac{kN_k}{\sum_{k=1}^{N-1} kN_k} = \frac{NP(k)k}{\sum_{k=1}^{N-1} P(k)kN} = \frac{P(k)k}{\langle k \rangle} \tag{1.4}$$

where $P(k)$ represents the degree distribution of the network. Therefore, the distribution of excess degree, defined as the number of edges of the node other than the one that led to it, is $q(k) = P(k+1)(k+1)/\langle k \rangle$. In addition, the average excess

degree would be

$$\sum_{k=0}^{N-2} q(k)k = \sum_{k=0}^{N-2} \frac{P(k+1)(k+1)k}{\langle k \rangle} = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \tag{1.5}$$

which is also known as the mean branching coefficient. The average number of first neighbors is $\langle k \rangle$, the average number of second nearest neighbors is thus $\langle k \rangle \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} = \langle k^2 \rangle - \langle k \rangle$ and similarly, the average number of l_{th} nearest neighbors is $\langle k \rangle \big(\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}\big)^{l-1}$ [99]. From this we can see that in order to have a GC in the network, we need $\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} > 1$

Suppose u is the probability of not reaching a first neighbor belonging to the GC following an arbitrary edge, then this should be equal to the probability that this first neighbor is not connected to any of its first neighbor belonging to the GC. In other words, this is the probability that none of the second neighbors belong to the GC following this arbitrary edge, and therefore, we have

$$u = \sum_{k=0}^{\infty} q(k)u^k \tag{1.6}$$

On the other hand, suppose the average size of the GC is $S = Ns$, then s is given by,

$$s = 1 - \sum_{k=0}^{\infty} P(k)u^k \tag{1.7}$$

In the case that the degree follows Poisson distribution,

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!} \Rightarrow q(k) = \frac{\lambda^{k+1} e^{-\lambda}}{(k+1)!} \frac{k+1}{\lambda} = \frac{\lambda^k e^{-\lambda}}{k!} = P(k) \tag{1.8}$$

So,

$$u = \sum_{k=0}^{\infty} q(k)u^k = \sum_{k=0}^{\infty} p(k)u^k \overset{N\to\infty}{=} \sum_{k=0}^{N-1} \binom{N-1}{k} p^k (1-p)^{N-1-k} u^k$$
$$= (1 - p + pu)^{N-1} \tag{1.9}$$

Hence, in the case of the Gilbert model, the probability for a node i not connecting to the GC via node j is either they are not connected or they are connected, but the neighbor j itself does not belong to the GC. We can apply this idea and make the connection to the percolation process. In the case of bond percolation, each edge in the network will be removed with the same probability $1 - T$; then, we can find the GC_p on this remainder network after bond percolation in the following way. Suppose the probability of not reaching the GC_p via an arbitrary edge on the original network before percolation is u, then either this edge is removed with probability $1 - T$, or the edge remains intact but none of the second neighbors following the edge are in GC_p. Then,

$$u = 1 - T + T \sum_{k=0}^{\infty} q(k)u^k \tag{1.10}$$

and similarly, the average size of the GC is $S = Ns$, with s given by

$$s = 1 - \sum_{k=0}^{\infty} P(k)u^k \tag{1.11}$$

On the other hand, in the case of site percolation, when we remove the nodes uniformly from the network with probability $1 - T$, the proportion of GC size in the network is

$$s = T\big[1 - \sum_{k=0}^{\infty} P(k)u^k\big] \tag{1.12}$$

with the function for u unchanged. In fact, there is a condition for the existence of a

GC in the remaining network after it went through the percolation process. Specifically, following the same argument above regarding the mean branching coefficient, if we define the reproduction number R_0 as the expected number of second neighbors on the remaining network following an arbitrary edge, then [95] shows that

$$R_0 = T\sum_{k=0}^{\infty} kq_k = T\sum_{k=0}^{\infty} k\frac{(k+1)p_{k+1}}{\langle k\rangle} = T\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}. \tag{1.13}$$

Hence, to have a GC, we need to ensure $R_0 > 0 \Rightarrow T > T_C = \frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}$. For a more rigorous and detailed treatment of this topic, we refer the readers to [100].

1.2 Information diffusion process on networks

The topic of information diffusion on networks has attracted attention from a wide range of research communities due to the wide range of applications in diverse fields. A lot of social spreading phenomena can be treated as an information diffusion process, such as rumor spreading, news services, technological innovation adoption, new product promotion via viral marketing and political propaganda among others [21, 61, 88, 49, 71]. Mathematical frameworks formulated to tackle this problem include two main directions: threshold models and independent interaction models [31, 21]. One representative threshold model is the linear threshold one that can be simply described as follows [92]. At time $t = 0$, a proportion ρ_0 of nodes are set in the infected (informed) state, or in other words, selected as seed nodes; then, at each subsequent time step, all nodes on the whole network update according to the following rule,

$$s_i(t+1) = \begin{cases} 1, & \text{if } \sum_{j\in N(i)} s_j(t) > \theta k_i \\ 0, & \text{otherwise} \end{cases} \tag{1.14}$$

where $s_i(t) \in \{0, 1\}$, with 1 standing for informed (infected) and 0 otherwise, θ is the threshold parameter, k_i is the degree of node i and $N(i)$ is the set of first neighbors of node i. This evolution is deterministic and there is only one direction for the nodes to evolve: from not active to active, and the system reaches equilibrium once the number of active nodes stops changing.

On the other hand, due to its similarity with epidemic processes such as disease transmission, the information diffusion process has been widely treated as an epidemic process and a lot of independent interaction models describing the dynamics are employed. For example, the susceptible-infected-recovered (SIR) model and the susceptible-infected-susceptible (SIS) model are popular ones, whose origins come from disease transmission in an epidemiological setting [21, 33, 105]. In a standard SIR model, nodes in the network are split into three exclusive, but exhaustive groups: susceptible (S), infected (I), and recovered (R). After some seed nodes are activated at time $t = 0$, each susceptible neighbor of nodes in the I state will get infected with a constant rate β and each of the nodes in the I state will get removed with constant rate γ. As the name implies, the evolution process is stochastic, instead of deterministic as in the threshold model, and successive trials of infection (diffusion) between nodes are independent with each other with the same probability. Similarly, different trials on nodes in the I state for the removal process are also independent with each other. The system will reach an equilibrium when all the nodes in the I state disappear. For the SIS model, instead of getting removed, those nodes in the I state will be reverted back to the S state with a rate γ and the system reaches equilibrium when at each time step the number of nodes converted to the S state is the same as the number of nodes converted to the I state.

Interestingly, the SIR model can be mapped to a bond percolation process, with the GC representing the size of diffusion, since the probability of transmission between

a node in the S state and a node in the I state is

$$T = \int_0^\infty e^{-\beta t} \beta e^{-\gamma t} dt = \frac{\beta}{\beta + \gamma} \tag{1.15}$$

which is a constant and can be mapped to be 1 minus the probability of bond removal in the bond percolation process if we start the diffusion process from a single node in the GC. Similarly, the condition for having an epidemic is

$$\frac{\beta}{\beta + \gamma} > \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

and the disease will reach size N_E that scales up with $N_E \sim O(N)$ as $N \to \infty$ before dying out.

On the other hand, generally speaking, the information diffusion process can be approximated using results from mean field theory [92, 105]. For example, in the SIR model, when we ignore the degree distribution of the network and assume a homogeneous degree across the network we get that

$$\begin{aligned} \dot{s}(t) &= -\beta \langle k \rangle \rho(t) s(t) \\ \dot{\rho}(t) &= \beta \langle k \rangle \rho(t) s(t) - \gamma \rho(t) \\ \dot{r}(t) &= \gamma \rho(t) \end{aligned} \tag{1.16}$$

where $s(t)$ is the proportion of nodes in the susceptible state, $\rho(t)$ is the proportion in the infected state and $r(t)$ is the proportion removed, β is the infection rate, γ is the removal rate, and $\langle k \rangle$ is the average degree of nodes in the network. However, it is rather unrealistic to assume a homogeneous degree, especially in real life networks where a power law distribution is very often observed. To address this issues, a

heterogeneous version of the mean field theory can be used,

$$\begin{aligned}
\dot{s}_k(t) &= -\beta k s_k(t)\Theta_k(t) \\
\dot{\rho}_k(t) &= \beta k s_k(t)\Theta_k(t) - \gamma\rho_k(t) \\
\dot{r}_k(t) &= \gamma\rho_k(t)
\end{aligned} \tag{1.17}$$

where $\Theta_k(t) = \frac{1}{\langle k\rangle}\sum_{k'}(k'-1)P(k')\rho_{k'}(t)$ is the probability of reaching an infected node following an arbitrary edge originated from a susceptible node with degree k.

It has been noted long ago that an information diffusion process is quite different from an epidemic process due to the *reluctance to tell stale news*. To address this issue, many other models were proposed. For example, in 1965 Daley and Kendall proposed to modify the removal dynamics to $\dot{r}(t) = \gamma\rho(t)(\rho(t)+r(t))$ to take into consideration the case where the news spreaders encounter people who are already aware of the information and become reluctant to spread it further. However this model only considers the spreader's point of view, which is quite limited especially in terms of application in marketing when it is not the spreaders (e.g. advertisement broadcasting platforms or the company itself), but rather the receivers who get bored with a news item as time goes by. Besides, accompanying the *reluctance to tell stale news* is also the *reluctance to talk about topics of stale news* from the point of view of susceptible nodes. Our contribution in this thesis to this area is to propose two different models, coined susceptible-informed-immunized (SIM) model and exponentially time decaying susceptible-informed (SIT) model to capture these missing elements from existing models. We employ mainly mean field theory approaches to rigorously study the evolution dynamics of these two models and also explain why percolation theory is not applicable, but rather new techniques need to be introduced for their analysis.

1.3 Information diffusion optimization

Information diffusion control has attracted a lot of attention in the scientific research community, due to a wide range of applications [122, 21]. For example, we might want to minimize the disease transmission by disseminating information about the disease, or we might be interested in minimizing the diffusion size of a specific piece of information or rumor or idea, as in the case of social census and political unrest. On the other hand, we might be interested in maximizing the diffusion size as in marketing, news services, and political propaganda. A lot of research efforts have been devoted on these topics. For example, for the first objective, a multilayer network framework representing the disease diffusion and information diffusion layers are proposed, and disease control can be obtained by considering the interaction between the two layers [129]. As for the second objective, in threshold models a lot of algorithmic approaches have been developed to approximately solve the computationally NP hard problem in which a set of seed nodes of size k would produce the maximal diffusion size [61]. On the other hand, for independent interaction models, a lot of measures such as k-shell, degree and betweenness centrality among others are used to identify the single (or alternatively a set of size k) seed node(s) in order to maximize the diffusion size [65].

However, in many applications the key consideration is not the diffusion size per se, but rather a profit function (revenue minus cost) associated with it. For example in marketing, one might be willing to invest more if the generated revenue is more promising compared to the other way around. Besides, different people or platforms will charge differently depending on some measure of the node, for example, its degree and centrality, and therefore it is not realistic to assume simply a budget of k seed nodes.

The contribution of this thesis is to address this issue and develop a more realistic model in terms of the objective function associated with this problem. The revenue

can be simply defined as the total diffusion size, while we provide two versions for the cost. The first is a function of the degree of the seed nodes and another is a function of a value that takes into consideration the topology of the network, a measure similar to centrality. The objective function we want to maximize is profit defined as revenue minus cost. To achieve the optimization objective, we employ a binary particle swarm optimization algorithm, which is a powerful combinatorial optimization algorithm developed by Kennedy and Eberhart in 1997 [63].

1.4 Community detection

Communities in networks can be intuitively defined as clusters of nodes that are densely connected internally and sparsely connected externally [42, 46]. One way to quantify the community structure is by a quality function named modularity [98],

$$Q = \frac{1}{2m} \sum_{ij} (\mathbf{A}_{ij} - \frac{k_i k_j}{2m}) \delta(g_i, g_j) \tag{1.18}$$

where $\mathbf{A}$ is the adjacency matrix, g_i stands for the community to which node i belongs, $k_i = \sum_j \mathbf{A}_{ij}$, $2m = \sum_{ij} \mathbf{A}_{ij}$, and $\delta(g_i, g_j)$ is 1 if $g_i = g_j$ and 0 otherwise. A higher value of Q indicates that the network is better partitioned.

Many real world networks possess intrinsic community structures and detecting them has been a widely studied topic in network science. One application is building up a strategic recommendation system where similar products can be recommended to groups of people of similar tastes based on their historical shopping similarities. Another important application would be information diffusion control or epidemic control if such processes evolve on networks of community structures since it is well studied that the community structures of networks play an important role in such processes [42, 92, 136, 109]. For static networks, many well developed algorithms are available for such purposes, for example, spectral partitioning and modularity

function based optimization algorithms [115, 83, 97, 94].

In terms of real world applications, it becomes more and more desirable to detect community structures in temporal networks because most of the real world networks evolve over time[89, 137, 104]. In recent decades, many methodologies have been proposed to tackle this problem, for example, piecing together community structure at different times, automatic detection based on minimum description principle, tensor decompositions, low rank membership matrix recovery, and optimization based on generalized modularity function [104, 120, 137, 89]. For example, in [89], the generalized modularity function is defined as,

$$\begin{aligned} Q &= \frac{1}{2\mu}\sum_{ijrs}\Big[\big(A_{ijs}-\gamma_s\frac{k_{is}k_{js}}{2m_s}\big)\delta_{sr}+C_{jrs}\delta_{ij}\Big]\delta(g_{is},g_{jr}) \\ &= \frac{1}{2\mu}\sum_{ijrs}\Big[\big(A_{ijs}-\gamma_s\frac{k_{is}k_{js}}{2m_s}\big)\delta_{sr}\delta(g_{is},g_{jr})+C_{jrs}\delta_{ij}\delta(g_{is},g_{jr})\Big] \end{aligned} \tag{1.19}$$

where A_{ijs} is the adjacency matrix element (edge) between nodes i and j at time s, C_{jrs} is the interslice coupling of node j with itself across time from time r to s, γ_s is a resolution parameter and μ is a normalization constant independent on the community structure. Further, γ_s and C_{jrs} are two parameters that can be tuned to control community structure consistency across time. Intuitively, the first part in the generalized modularity function represents the total modularity value of all the networks treated independently across the whole time range, while the second part represents the total coupling strength of nodes that belong to the same community across time. In other words, the first part represent the fit of communities in each individual time slice and the second part represents the consistency of community structures across time.

For application purposes, one scenario that is very likely to happen in real life applications, but is not well investigated in the available literature is the following [137, 111]. The core community structure does not change over a long period of time,

but possible small perturbations on isolated edges occur. Hence, starting at a certain time point t_1 and over a small period of time the core community structure evolves: communities might grow, shrink, disappear, or emerge. Then, the evolution repeats itself over the remaining time period. The contribution of this thesis in this area is to identify the phase transition time points where the core community structure changes and improve the community structure detection accuracy by proposing a new algorithm based on low rank component recovery from the available time series networks. Applications on both synthetic networks and real world network data are carried out.

1.5 Outline of chapters

In Chapter II, we develop a systematic methodology to recover the core community structures of temporally evolving networks, while at the same time detecting and identifying the time epochs where the core community structure changes significantly. The methodology is based on low rank component extraction from adjacency matrices of the time series networks and the efficiency of the methodology is demonstrated through applications on various networks, such as synthetic networks generated by standard random graph models including the stochastic block model (SBM) and its variants, networks representing the synchronization pattern of the Kuramoto model, and real life networks capturing voting similarities between Senators based on US Senate Roll Call data. In Chapters III and IV, we propose two different models, coined Susceptible-Informed-Immunized (SIM) model and Susceptible-Informed model with time decay diffusion rate (SIT) for modeling the mechanisms of an information diffusion process. The SIM model captures the characteristic of *slowing down* with time due to the *intrinsic time value* of *information* for general information diffusion processes from the perspectives of both the spreaders and the "ignorant" players indirectly by letting them go through an immunization process. While the

SIT model achieves the same objective directly by letting the diffusion rate decay exponentially over time. Both the SIT and SIM models are rigorously analyzed first by employing Mean Field (MF) approximation theory and then by constructing a Maximum Weight Tree (MWT). Specifically, in the MF approximation, the time dependent diffusion size can be calculated by solving a set of ordinary differential equations (ODEs), while closed form solutions at early stages can be obtained, based on which *phase transition* time points and *epidemic threshold* can be calculated. All these results are compared with those obtained from Monte Carlo (MC) simulation and good agreement is achieved. On the other hand, the expected total diffusion size can be approximated by the MWT, which assumes we start with a single seed node, defined as the node informed at the beginning of time, and diffuse the information on a static network. Good agreement between MWT and MC simulation estimates can be achieved for sparse networks, which is expected since in this case the probability of multi-channel diffusion is negligible and a tree like mechanism in approximation is satisfactory. In Chapter V, we employed a binary version of the particle swarm optimization (PSO) algorithm to approximately solve the NP-hard problem of identifying the optimal set of seed nodes so as to maximize the diffusion profit, defined as the expected total diffusion size minus the cost of selecting those seed nodes. To investigate the impact of different cost functions on the seed nodes allocation strategy, we carried out experiments on two sets of different cost functions: one depends on the degree, while the other depends on the MWT value. Further, to investigate the role played by the network topology in selecting seed nodes, we performed experiments on the following three settings: a single layer network with embedded community structure, a single layer scale free network with no community structure, and a two layer network obtained by coupling the previous two networks together. A possible application of our methodology is in marketing or advertising, where the marketeers need to identify an appropriate set of platforms on which to

promote the new products.

CHAPTER II

Core community structure recovery and phase transition detection in temporally evolving networks

2.1 Introduction

There has been a lot of work across different scientific communities including computer science, applied physics, statistics and the social sciences in developing methods for the analysis of network data[66]. The impetus for these developments has been the availability of new data in biology (e.g. protein-protein interactions, product-substrate relationships amongst compounds, or ecological communities of commensal, symbiotic and pathogenic microorganisms), friendship relationships in social media platforms such as Facebook, Instagram and Twitter, transactional data between consumers or business organizations, just to name a few [23, 7, 51, 68, 41]. Such data capture interactions between a set of entities (e.g. biomolecules, physical persons, companies) giving rise to a network structure.

A wide range of topics can be studied on networks, including constructing representations, effective visualization of their structure, descriptive analysis of their characteristics, study of network formation models and their behavior and study of dynamically evolving phenomena (like epidemics or information diffusion) on networks

[100, 66]. A topic that has recently attracted a lot of interest is that of identifying community structure in observed networks, as well as developing network formation models that exhibit such structure. A community on a network is heuristically defined as a set of nodes exhibiting high degree of interconnectivity in comparison to other nodes in the network. Its importance stems from the fact that it represents a defining characteristic of real world networks; for example, sets of close friends give rise to such communities in social networks, or sets of closely interacting biomolecules (functional pathways) in biological networks. A large number of fast and efficient algorithms have been proposed in the literature to identify such communities in networks, including minimum cut based graph partitioning, hierarchical clustering, k-means based clustering and spectral clustering (for a comprehensive review, see [42] and references therein). Another class of algorithms is based on maximization of a quality function known as "modularity" [1] over possible partitions of the network; given the computational complexity of this optimization problem various greedy variants, as well as algorithms based on simulated annealing and spectral optimization [97, 96, 94, 52] have been developed. Finally, on the network formation models front exhibiting community structure, the stochastic block model (SBM) and its variants (e.g. degree corrected SBM) have been the objects of intensive study [59, 30, 1].

However, most of the focus to date has been on static networks, where a single snapshot is available. In many applications, one has access to a sequence of network snapshots evolving over time. It is then of great interest to identify communities in such dynamically evolving networks and also investigate whether their structure remains fixed or exhibits changes over time. There has been some recent work on this

[1] It is a function defined as [98]:

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(g_i, g_j) \tag{2.1}$$

where A is the adjacency matrix, g_i stands for the community to which node i belongs, $k_i = \sum_j A_{ij}$, $2m = \sum_{ij} A_{ij}$ and $\delta(g_i, g_j) = 1$ if $g_i = g_j$ and 0 otherwise. A higher modularity value indicates a better partition for the network under consideration.

subject. For example [89] developed a generalized modularity quality function that reflects the temporal dynamics of a sequence of networks, while [137] introduced the concept of common communities and proposed a method to detect them in a sequence of networks by optimizing an objective function based on a node-wise membership matrix. Further, [10] developed a robust community detection algorithm for this problem, by optimizing a quality function (e.g. modularity) based on null statistical models assumed to generate the network data; examples of such models include the Newman-Girvan one and correlation/similarity ones [98, 101, 10].

However, an issue not adequately explored in the literature is the identification of time epochs where significant changes (phase transitions) occur in the network structure and also the community structure between them. Some previous work includes [137] and [111], the latter using an Ising model and assuming the existence of a single change in the network structure over time. Nevertheless, such changes are common in many applications. A recent example comes from the change in the connectivity patterns in networks of asset returns before, during and after the financial crisis of 2008 (for details see [73]), while another one relates to changes in brain connectivity before and after epileptic seizures [112]. A third example stems from changing patterns of political polarization amongst US legislators, which is examined later on in this study.

Our proposed modeling framework for the problem at hand assumes that given a sequence of T network snapshots, the community structure exhibits significant changes (phase transition) at time periods $\{[\tau_m^-, \tau_m^+]\}_{m=1}^M$, while it remains invariant within time segments (τ_m^+, τ_{m+1}^-), where τ_m^- and τ_m^+ represents the start and end time points of the m-th phase transition epoch, respectively. Possible changes during phase transitions include merging or division of existing communities, growth by adding members to them or their extinction altogether, while during stable periods the community can exhibit perturbations in its structure; either *dense* ones that in-

volve a number of nodes in the community, but small in magnitude in the sense that the strength of the links between nodes changes by a small amount, or *sparse* ones that involve isolated nodes, but the strength of the corresponding links to selected other nodes can be large in magnitude. Technically, we assume that the community structure can be captured by a *low-rank* weighted adjacency matrix, while the perturbations correspond either to the addition of small magnitude dense components or large magnitude sparse components.

2.2 Results

Model formulation and Optimization Strategy

Consider a sequence of T weighted adjacency matrices $\{A(t)\}_{t=1}^{T}$ that encaptulate the structure of a network comprising of n nodes and their corresponding edges. It is a symmetric matrix and the edge magnitude $A_{i,j}(t), i, j = 1, \cdots, n$ capture the strength of association between nodes i and j. In the simplest case, $A_{ij}(t) \in \{0, 1\}$ indicate whether nodes i and j are connected or not.

It is assumed that $A(t)$ can be decomposed as follows: $A(t) = L(t) + S(t) + E(t)$, where $L(t)$ is a low-rank matrix, $S(t)$ is a sparse one with most of its elements being zero and $E(t)$ a dense matrix with $||E(t)||_F < \epsilon$ for some small $\epsilon > 0$ and where $|| \cdot ||_F$ denotes the Frobenius norm. This model captures the presence of community structure in networks through the low-rank component, as well as possible small sparse and/or dense perturbations as explained in the previous section. It is also compatible with the popular network formation model that gives rise to community structure, namely the Stochastic Block Model (SBM) [30]. Specifically, the SBM assumes an undirected network on n nodes and that the nodes are partitioned into K blocks. Then, edges are formed according to the following stochastic mechanism. Node i is connected to node j with an edge, whose probability of occurring only depends on

the blocks (communities) to which i and j belong to. It is commonly assumed that the probabilities for edges between i and j in the same community are significantly higher than those for nodes i and j in different communities. The SBM has been the object of intense study in recent years [59, 30, 1]. It can be seen that this mechanism gives rise to a low-rank structure L, corrupted with noise E, thus captured by the posited model. In fact, the proposed model also allows for "spiky" noise in the form of the sparse matrix S and as already mentioned can accommodate weighted adjacency matrices as well.

In what follows, we use the time point τ_m to represent the m-th phase transition time period $[\tau_m^-, \tau_m^+]$. For a sequence of adjacency matrices, we make the additional assumption that the low rank structure is invariant between phase transition time points τ_m, while the perturbations are allowed to vary freely. This is consistent with the intuition that community structure can be *slowly evolving* over time, while individual edges (connections) between nodes can exhibit a higher degree of variability in their patterns. Hence, our modeling framework assumes that:

$$L(t) = L_m I(t \in (\tau_m, \tau_{m+1})) + S(t) + E(t), \;\; t = 1, \cdots, T, \;\; 0 = \tau_0 < \tau_1 < \cdots < \tau_m < \cdots < \tau_M = T.$$

Note that the problem of decomposing a matrix into low-rank and sparse/dense components has been investigated in the literature, due to its relevance in matrix completion problems [38, 20, 22, 121, 108, 18] that emerged from recommender systems, compressed sensing, system identification, anomaly detection and related applications [108, 77, 39, 40, 103]. Specifically, the low rank recovery problem can be formulated as a rank minimization problem subject to certain constraints. To make the problem computationally tractable (convex), a nuclear norm [2] was introduced in [38]. Subsequent work [20, 22] examined the following variant $\min_{L,S} ||L||_* + \gamma ||S||_1$, subject to

[2]The nuclear norm of a matrix $X \in \mathbf{R}^{M \times N}$ of rank r is defined as the summation of all the singular values $\|X\|_* := \sum_{i=1}^{r} \sigma_i$, where σ_i's are the singular values of X and are equal to the square roots of the eigenvalues of XX^T.

exact recovery $A = L + S$, while a further extension was studied in [121] where the objective function remained the same, but the constraints allowed low-rank recovery subject to sparse and also dense noise E; i.e. $A = L+S+E$ with $||E||_F \leq \epsilon, \epsilon > 0$. In the above, $||\cdot||_1$ denotes the ℓ_1 norm for the vectorized form of its matrix argument. Further, in theoretical work, [20, 121] showed that the problem is feasible and admits a correct solution if an *incoherence condition* between L, S and E is satisfied, which intuitively requires that the matrices E and S can not exhibit low-rank structure.

In the presence of dense noise, the corresponding optimization problem can be formulated as

$$\begin{aligned} &\min \ \|L\|_* + \gamma\|S\|_1 + \alpha\|E\|_F^2 \\ &\text{s.t. } L + S + E = A \end{aligned}$$

with the two tuning parameters γ and α controlling the trade-off amongst the low rank, sparse and dense noise components.

Turning our attention to the problem at hand, we can analogously formulate the problem, while we need to incorporate an additional constraint that would force recovery of the *same* low-rank (community) component between phase transition epochs. The latter task can be accomplished by adding a total variation penalty encouraging similarity between successive time estimates of $L(t)$, given by

$$\begin{aligned} &\min \ \sum_{t=1}^{T} \Big[\|L(t)\|_* + \gamma(t)\|S(t)\|_1 + \alpha(t)\|E(t)\|_F^2\Big] + \sum_{t=2}^{T} \lambda(t)\Big[\|L(t) - L(t-1)\|_F^2\Big] \\ &\text{s.t. } A(t) = L(t) + S(t) + E(t), \ t = 1 \cdots T \end{aligned} \tag{2.2}$$

with the tuning parameters $\lambda(t)$ controlling the degree of discrepancy between consecutive low rank components $L(t)$. It is worth mentioning that the squared Frobenius norm for the total variation penalty can be replaced by other norms such as the ℓ_1 to

achieve the same objective, and the proposed strategy below would go through with minor modifications.

Brief Overview of Change Point Analysis Methods: The problem has been extensively studied in the statistics literature for *single time series* data. Further, an extensive body of theoretical results -convergence rates of various estimators, as well as asymptotic distributions for the change point (phase transition epoch)- have been established when a *single* change point is assumed (for a comprehensive review see [55, 54]). More recently, the focus has shifted to developing fast procedures for identifying *multiple* change points in a single time series and also providing probabilistic guarantees for identifying the right number of them, as well as their locations [44]. In parallel, methods for identifying a single change point in multiple time series data emerged [5] and more recently extensions to high-dimensional settings appeared [127] together with their extension to the case of multiple change points. However, a key assumption has been that the time series under consideration are *independent*, which implies that a simple least squares criterion can be used to identify at least a single change point. On the other hand, the network setting considered in our study obviously violates the latter condition and hence a more complex criterion needs to be employed. Further, as previously mentioned, in the presence of network data streams the objective is not simply to identify transition epochs, but also the nodes in the network that gave rise to them.

Turning our attention to the problem formulation in 2.2, it can be seen that our interest is in finding the single core common community structure for each of the M time epochs, while at the same time detecting and identifying the phase transition time points τ_j. Therefore, to achieve this objective, the tuning parameter $\lambda(t)$ should be set as large as possible between phase transition time points while at phase transition time points we would like to set $\lambda(t)$ as small as possible. Naturally, the second requirement can be achieved by directly removing the total variation penalty, which

is equivalent to setting $\lambda(t) = 0$. While for the first requirement, since we assume a fixed low rank component L_m for each stable community time period, we can approximately recover the single low rank component for each period by simply taking the average of the individual low rank components $L(t)$ recovered without the total variation penalty. Assuming that the length of the time intervals $|\tau_{m+1} - \tau_m| \sim cT, c < 1$ scales linearly with time, we would expect that the variations in the estimates of $L(t), t \in (\tau_m, \tau_{m+1})$ cancel out and the average low-rank component converges to the true one that generated the weighted network adjacency matrix. Therefore, by considering the following relaxed version of the optimization problem above

$$\begin{aligned} &\min \; |||L(t)||_* + \gamma||S(t)||_1 + \alpha||E(t)||_F^2 \\ &\text{s.t. } A(t) = L(t) + S(t) + E(t) \end{aligned} \tag{2.3}$$

and then averaging the low-rank component estimates between phase transition epochs. Then, the question becomes of how to identify those τ_m epochs accurately. We propose to calculate and monitor over time the *thresholded rank* over time windows of a certain length. This rank is defined as the number of singular values exceeding a carefully selected threshold, which represents the effective number of communities detected. Finally, the core community structure L_m can be estimated by using any standard clustering technique for networks applied to the average $\overline{L}_m(t)I(t \in (\tau_m, \tau_{m+1}))$. A recommended clustering technique for this task is spectral clustering [124, 134] and the overall strategy is illustrated in the next sub-section.

To solve the problem in equation 3.2.1.8, we adopt an alternating splitting augmented Lagrangian method (ASALM) that was proposed in [121] and is a variant of the widely used alternating direction method of multipliers (ADMM) [125]. The advantage of this method is that its computational complexity of each iteration of the algorithm is dominated by one singular value decomposition (SVD), whose computa-

tional efficiency can be further improved by employing a partial SVD decomposition [74, 121] since only leading singular values and corresponding vectors need to be calculated. The augmented Lagrangian function of 3.2.1.8 is given by:

$$L = \|L\|_* + \gamma\|S\|_1 + \alpha\|E\|_F^2 + \left\langle \Lambda, A - L - S - E \right\rangle + \frac{\beta}{2}\|L + S + E - A\|_F^2 \quad (2.4)$$

where $\beta > 0$ is used to penalize violation of the constraint $A = L+S+E$ and $\left\langle \right\rangle$ is the trace inner product which is defined as $\left\langle X, Y \right\rangle := \sum_{ij} X_{ij}Y_{ij}$. Based on ASALM, we can minimize the Lagrangian function by splitting it into separate parts and minimize consecutively as follows:

$$\begin{cases} E^{k+1} = \text{argmin } \alpha\|E\|_F^2 + \frac{\beta}{2}\|E + L^k + S^k - \frac{1}{\beta}\Lambda^k - A\|_F^2 \\ S^{k+1} = \text{argmin } \gamma\|S\|_1 + \frac{\beta}{2}\|S + L^k + E^{k+1} - \frac{1}{\beta}\Lambda^k - A\|_F^2 \\ L^{k+1} = \text{argmin } \|L\|_* + \frac{\beta}{2}\|L + S^{k+1} + E^{k+1} - \frac{1}{\beta}\Lambda^k - A\|_F^2 \\ \Lambda^{k+1} = \Lambda^k + \beta(A^{k+1} - L^{k+1} - S^{k+1} - E) \end{cases}$$

The Lipschitz constant β controls the number of leading singular values we need to compute at each iteration in the general soft singular value thresholding algorithm for low rank matrix recovery/completion, and a lot of related work has been devoted in the area to investigate the strategy of choosing an appropriate value to ensure convergence to the correct results, including dynamically updated ones and fixed ones[74, 17, 121]. Following [121], for our purpose we choose a fixed $\beta = 0.15N^2/||A||_1$, where N is the size of the networks.

Therefore, we have the following iterative updating algorithm for iteration q

Start with $E, S, L, \Lambda = \mathbf{0} \in \mathbf{R^{N \times N}}$
Update E: $E^{q+1} = \frac{\beta}{2\alpha+\beta}(\frac{1}{\beta}\Lambda^q + A - L^q - S^q)$
Update S[14]: $S^{q+1} = \mathcal{S}_{\gamma/\beta}(\frac{1}{\beta}\Lambda^q + A - L^q - E^{q+1})$
where $(\mathcal{S}_\mu(T))_{ij} := \max\{abs(T)_{ij} - \mu, 0\} \cdot sign(T_{ij})$
Update L[17]: $L^{q+1} = \mathcal{D}_{1/\beta}(\frac{1}{\beta}\Lambda^q + A - E^{q+1} - S^{q+1})$
where $\mathcal{D}_\mu(Y) := U\mathcal{S}_\mu(\Sigma)V^\top$, and $U\Sigma V^\top$ is the SVD of Y
Update Λ: $\Lambda^{q+1} = \Lambda^k + \beta(A - L^{q+1} - S^{q+1} - E^{q+1})$

Note that the update for the dense noise component E involves a closed form formula, that for the sparse component a soft-thresholding step, while that for the low-rank component a SVD. Finally, the update of the Lagrange multipliers Λ is given by another closed form formula. The stopping condition for the algorithm is given by: $\frac{\|L^{q+1}-L^q\|_F^2+\|S^{q+1}-S^q\|_F^2}{\|L^q\|_F^2+\|S^q\|_F^2+1} \leq \epsilon$.

To employ the above algorithm, the tuning parameters α and γ need to be specified. Further, once the network adjacency matrices have been decomposed, the task becomes on how to identify the phase transition epochs and determine the number of communities and their membership. These issues are discussed next and illustrated on synthetic data generated according to the following mechanism.

Data Generation for Illustrative Example: We employ a statistical factor model to generate the sequence of network adjacency matrices. For each of the M stable time periods, a common low rank component $L(t) = L_m = U_m U_m^T$ is generated for all $t \in (\tau_m, \tau_{m+1})$, where each column $u_k(m)$ of U_m satisfies that for nodes belonging to community k, $u_k(m)_i \overset{i.i.d.}{\sim} Unif(r_{in}, 1)$ and for those nodes not in community k, $u_k(m)_i \overset{i.i.d.}{\sim} Unif(0, r_{out})$. It can be seen that by selecting parameters so that $r_{in} > r_{out}$, a node i belongs to community k if $u_{ik} \geq r_{in}$. Finally, the sparse and dense noise components, at each time point t are generated according to:

- $S(t) = S_1(t) \bullet S_2(t)$, where $S_1(t) \overset{i.i.d.}{\sim} Bernoulli(p_s)$, $S_2(t) \overset{i.i.d.}{\sim} Unif(-1, 1)$.
- $E(t) \overset{i.i.d.}{\sim} Unif(-r, +r)$, $r > 0$.

- $A(t) = \mathcal{P}_\Omega(L(t) + S(t) + E(t))$, Euclidean projection onto the space $\Omega := \left\{X \in \mathbf{R}^{N\times N} |\ 0 \leq X_{i,j} \leq 1, X_{ij} = X_{ji}\right\}$

where $X \bullet Y = [X_{ij}Y_{ij}]$ for $X, Y \in \mathbf{R^{M\times N}}$ denotes element wise matrix multiplication, p_s is the density of the sparse noise, and r is the upper bound for the magnitude of the dense noise component. For our illustrative example, we set the following values for the control parameters for the model: $r_{out} = 0.4, r_{in} = 0.6, p_s = 0.1$ and $r = 0.1$, and we generate $T = 200$ network snapshots, divided into $M = 4$ stable periods, with $11, 11, 7,$ and 12 densely connected communities respectively, and phase transition epochs occurring at times $\tau = 30 - 31, 60 - 65, 130 - 135$. The time dependent community membership is depicted in panel (a) of Figure 2.1 with community *-1* representing the missing nodes and the detail of the network sequence is as follows: At $t = 0$, we set the network size as $N = 900$ and there are 11 densely connected communities numbered from 0 to 10 of size $100, 100, 90, 90, 80, 80, 70, 70, 60, 60, 50$, respectively, with the remaining 50 nodes not belonging to any community and hence expected to be captured by the noise component. For pure notation purposes, we call these 50 outlying nodes, the "11th community." At time 31, an extra 100 nodes not exhibiting any community structure join the network and hence are assigned to the "11th community." From time 61-65 the 100 nodes that joined the network at time 31 leave, while community 10 joins 0, 9 joins 1, and 7 and 8 join 6. From times, 131-135, the network structure reverts to the one being present between time periods 31-60, with the addition that community 8 disintegrates and joins the "11th community", and communities 1 and 2 split into two separate communities each. This structure persists till the the end. It can be seen that a number of intricate changes occur to the network structure designed to showcase the power of the proposed methodology.

2.2.0.1 Selection of tuning parameters and decomposition

There are two parameters required in the optimization problem above; following suggestions in literature, we recommend selecting $\gamma = 1/\sqrt{N}$ for general purpose[19]. For α, extensive numerical work shows that searching over the range $(0.5/\sqrt{N+\sqrt{8N}}, 10/\sqrt{N+\sqrt{8N}})$ provides highly satisfactory results.

For our problem we sample 20 evenly spaced time points from the total $T = 200$ and plot the rank of the recovered low rank components for these time points across different values of α. To facilitate identifying the optimal α we also include the plot of inconsistency/difference of the $L(t)$ across consecutive values of α, defined as $\|L(t)_{\alpha_i} - L(t)_{\alpha_{i+1}}\|_F / \|L(t)_{\alpha_i} + 1\|_F$ with $\alpha_{i+1} - \alpha_i = 0.015$ in our case. The results are shown in Figure 2.1(b), and in order to recover robust $L(t)$ components, we suggest choosing a value of α that would result in low inconsistency from the wide range of α values that would recover consistent rank across time points. For example, in our case, the wide range is $\alpha \in [0.045, 0.105]$ and we choose $\alpha = 0.1$ to minimize the inconsistency.

With the above optimal set of parameters, the rank of the recovered $L(t)$ across time is shown in Figure 2.1(c) and simply by inspecting the evolution of rank across time, we can easily identify the two groups of phase transition time points around $t = 60$ and $t = 130$. However, this is not a very reliable method because on the one hand we did not capture the phase transition time points around $t = 30$ and on the other hand in terms of real life applications the rank of $L(t)$ might not be so clean. Therefore, we develop the following strategy based on the *thresholded rank* of averaged $L(t)$ to accurately and robustly identify phase transition time points.

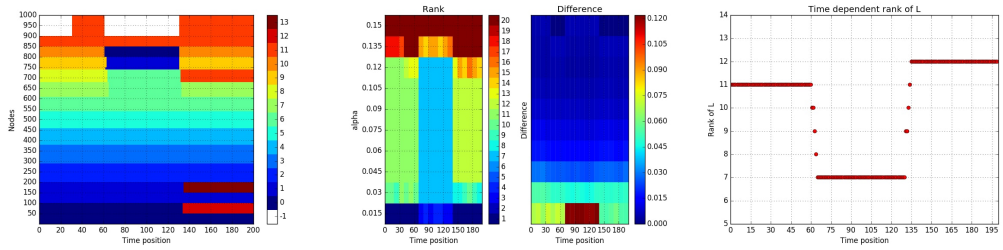


Figure 2.1: (a): Left panel, design of the time dependent community structure. (b): Middle panel, illustration of rank and inconsistency across time based on a range of values of α. (c): Right panel, rank of $L(t)$ recovered based on $\alpha = 0.1$

2.2.0.2 Phase transition identification

The strategy of identifying phase transition epochs $\{\tau_m\}$ is as follows. First, split the T available time points into $\sqrt{T}$ non-overlapping windows of equal length $\sqrt{T}$. Then, scan over these windows and within each window compute for a range of values of thresholds the *thresholded rank* of the averaged low rank component, defined as the number of singular values exceeding the threshold. If a window is within a certain stable period, following the argument in the section of model formulation, we would expect an enhanced modularity and therefore the *thresholded rank* would be consistent with the number of communities in the network across a wide range of threshold values. Therefore, if on one hand, the length of stable periods exceeds the window size, then we would expect a consistent *thresholded rank* across windows. On the other hand, if the window covers time points from two different stable periods having different community structures, the *thresholded rank* will exhibit volatility until it belongs to the new stable time period, wherein the *thresholded rank* starts to exhibit a stable behavior. Thus, to find the phase transition epochs, we simply choose an appropriate threshold value from a wide range of such values that would result in consistent *thresholded rank* across windows and identify the windows where the thresholded *rank* first exhibits volatility. For illustration purposes, we choose a window of length 7 and the wide range of threshold values is approximately $h \in$

$[1, 6]$ and we choose $h = 1.6$ for our purpose as is shown in (a) and (b) of Figure 2.2. The choice of the window length is supported by some theoretical work [76] and by extensive numerical experimentation. Of course, the accuracy level of the identification is limited by the window length $\sim \sqrt{T}$ and to improve accuracy, around the phase transition time points obtained from above, we can zoom in to windows of smaller length, say $\sim \sqrt{\sqrt{T}}$, and repeat the same procedure. Naturally, the window length can be further reduced to obtain a desired level of resolution for identifying the phase transition epochs. The selection of a $\sqrt{T}$ window length is dictated by the need for scalability of the procedure, in the presence of data sets involving a very large number of time points T.

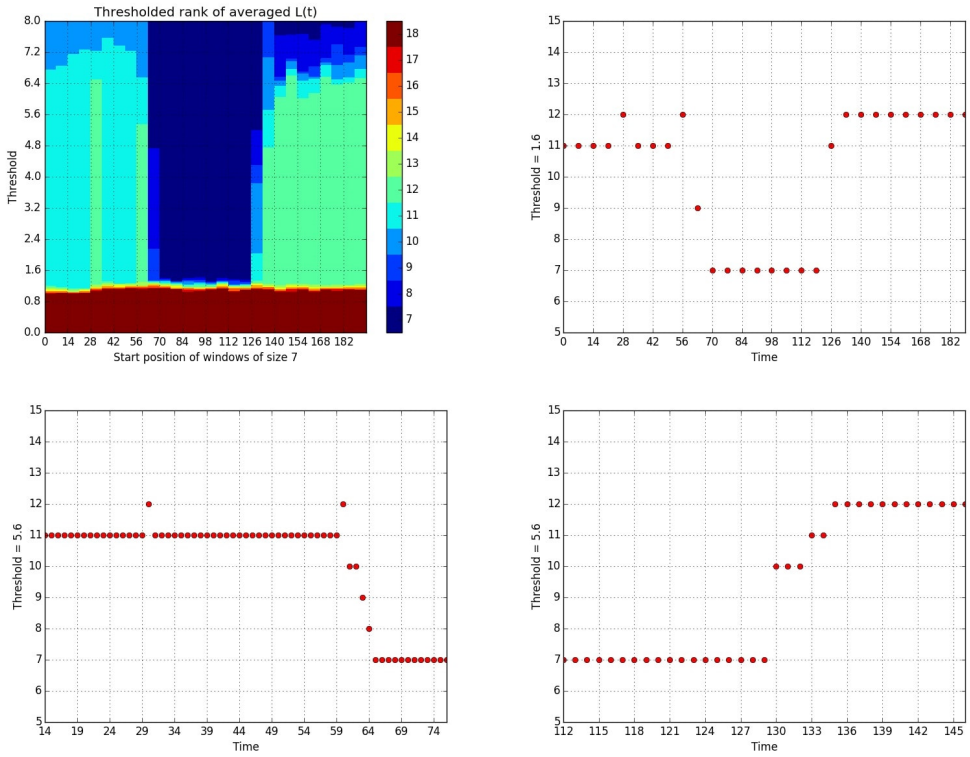


Figure 2.2: (a): Upper left panel, *thresholded rank* of windows of length 7 across time, based on different values of the *threshold*. (b): Upper right panel, *thresholded rank* of windows of length 7. (c): Lower left panel, *thresholded rank* by zooming-in using windows of length 2 around the first two phase transition epochs. (d): Lower right panel, *thresholded rank* of zoom-in windows of length 2 around the last phase transition epochs.

To illustrate the idea of the window scanning strategy, we consider the results of our illustrative example as shown in Figure 2.2. From the first level of windows of length 7 we identify the phase transition positions around $t = 28, 56 \sim 63, 126$. Then we zoom in around these time points with window length 2 and identify the phase transition time points $t = 31, 61 \sim 65, 131 \sim 135$ as shown in (c) and (d) of Figure 2.2. Here in order to scan windows over the whole time series, we cast the adjacency matrices of all networks with smaller size to matrices of size N_{max}, the maximum size of networks in the whole time range, and set entries 0 for those missing nodes. Note here we identified the phase transition time position $t = 30$ when new nodes join the network, while if we simply look at the time dependent rank of $L(t)$ in figure 2.1(c) we would fail to detect this since the two regimes share the same rank but different community structures.

2.2.0.3 Core community structure detection

The final step involves identifying the invariant core community structures between phase transition epochs based on averaged $\overline{L}_m(t)$. In our example, the average misclassification rate is 0.24% when clustering based on each individual $L(t)$ and 0% based on averaged $\overline{L}_m(t)$. For comparison, if we cluster simply based on the original time series adjacency matrices $A(t)$ corrupted with both dense and sparse noises, the average misclassification rate is as high as 0.89%.

2.2.1 Application to Synthetic Network Data

We employ the following three mechanisms to induce community structure in the sequences of networks snapshots observed. (i) a factor model, (ii) a stochastic block model (SBM) and (iii) a weighted stochastic block model (WSBM) [2, 30]. For simplicity, networks generated by the three models share the same time patterns on when the community structure changes. There are $T = 30$ network snapshots in total,

divided in $M = 3$ stable time periods and all networks have size $N = 1000$ nodes. At $t = 0$, there are 11 densely connected communities numbered from $0, \cdots, 10$ of size $120, 100, 100, 100, 80, 80, 70, 70, 70, 60, 60,$ while the remaining 90 nodes do not form any community. At time $t = 10$ community 10 joins 0 and 9 joins 1 and this structure is fixed till $t = 19$. At time $t = 20$, community 7 and 8 join together with 6 and similarly this structure is fixed till the end.

The specifics of the network formation mechanisms are described next:

(i) Factor model: We fix $r_{out} = 0.4, r_{in} = 0.6$ for each $L(t)$ and control the density p_s of the sparse noise and the magnitude r of the dense noise, which are generated the same way as described in the illustrative example.

(ii) SBM: Only sparse noise is introduced in this case since the entries of the adjacency matrices are binary. We fix the density of noise to be 0.1 and control the density of connections p_c within communities for each $L(t)$.

- Generate $L_{in} \overset{i.i.d.}{\sim} Bernoulli(p_c)$, and $L_{out} = 0$, where L_{in} stands for connections within communities and L_{out} for otherwise.
- Generate $S \overset{i.i.d.}{\sim} Categorical(\pi)$, such that $P(S_{ij} = 1) = 0.05, P(S_{ij} = -1) = 0.05$ and $P(S_{ij} = 0) = 0.9$.
- Project to adjacency matrices: $A(t) = \mathcal{P}_\Omega[L(t) + S(t)]$.

(iii) WSBM: Similar to the factor model, we fix $L(t)$ by generating them in the following way: $L_{in} \overset{i.i.d.}{\sim} Unif(0.4, 1)$ and $L_{out} \overset{i.i.d.}{\sim} Unif(0, 0.6)$. Then, the strength of the sparse and dense noise is generated in the same way as before. Note if we have an adequate number of network snapshots, the connection strength of $\overline{L}_m(t) \to \frac{1+0.4}{2} = 0.7$ for edges belonging to communities and is larger than $\overline{L}_m(t) \to \frac{0+0.6}{2} = 0.3$ for edges connecting different communities, although for any individual low-rank component $L(t)$ the connection strength of some edges might not be concordant with this ranking.

For both the factor and WSBM models, we design dense and sparse noise components with various levels of strength controlled by r or p_s as shown in Tables 2.1 and 2.2, while for the SBM model we control p_c to study the performance having different connection strength within communities, as given in Table 2.3. Applying the proposed methodology, we first successfully identify the phase transition time points $t = 9, 19$ for all the cases with optimal values for parameters γ, α given in the respective Tables. Subsequently, we compare the performance of community detection based on individual low-rank components $L(t)$, averaged ones $\overline{L}_m$ over the stable periods and the original adjacency matrices $A(t)$. For the factor model, the misclassification rate based on the recovered $L(t)$ from $A(t)$ are consistently lower than $A(t)$, while for the SBM and WSBM models, this is not necessarily the case. However, for all the scenarios examined, the misclassification rate based on the averaged $\overline{L}_m$ goes to zero except for some extreme cases. Note that the *recovery rate* is defined as the proportion of densely connected communities recovered, while the *error rate* is the averaged misclassification rate across time for those recovered communities.

Table 2.1: Dense noise case for factor model and WSBM

Models	r	γ	α ($\times 10^{-2}$)			Recovery Rate			Error Rate ($\times 10^{-3}$)								
									t1			t2			t3		
			t1	t2	t3	t1	t2	t3	$A(t)$	$L(t)$	$\overline{L}$	$A(t)$	$L(t)$	$\overline{L}$	$A(t)$	$L(t)$	$\overline{L}$
Factor	.3	.03	4.5	4.5	4.5	1	1	1	48	11	0	48	6	0	37	3	0
	.4	.03	3.2	2.8	2.8	1	1	1	61	24	0	41	12	0	33	6	0
	.5	.02	3.9	3.5	3.5	1	1	1	103	192	3	63	78	0	42	36	0
WSBM	.5	.02	3.5	3	3	1	1	1	0	1	0	0	0	0	0	0	0
	.7	.02	3	2.8	2.8	1	1	1	10	41	0	4	12	0	2	2	0
	.9	.02	2.1	2.1	2.1	4/11	4/9	5/7	76	223	0	78	130	0	46	57	0

Table 2.2: Sparse noise case for factor model and WSBM

Models	p_s	γ	α ($\times 10^{-2}$)			Recovery Rate			Error Rate ($\times 10^{-3}$)								
									t1			t2			t3		
			t1	t2	t3	t1	t2	t3	$A(t)$	$L(t)$	$\overline{L}$	$A(t)$	$L(t)$	$\overline{L}$	$A(t)$	$L(t)$	$\overline{L}$
Factor	.5	.02	15	4.5	4.5	1	1	1	183	3	0	129	18	0	81	11	0
	.6	.02	6	4.3	4.3	1	1	1	344	40	0	171	35	0	104	22	0
	.7	.02	4.5	4	4	1	1	1	515	169	36	273	109	0	150	47	0
WSBM	.5	.02	3.5	3	3	1	1	1	1	1	0	1	1	0	0	0	0
	.7	.02	3.4	3.2	3	1	1	1	19	32	0	8	10	0	2	2	0
	.9	.02	2.1	2.1	2.1	4/11	4/9	5/7	68	174	0	65	102	0	76	55	0

Table 2.3: Sparse noise case for SBM

Models	p_c	α	γ ($\times 10^{-2}$)			Recovery Rate			Error Rate ($\times 10^{-3}$)								
									t1			t2			t3		
			t1	t2	t3	t1	t2	t3	$A(t)$	$L(t)$	$\overline{L}$	$A(t)$	$L(t)$	$\overline{L}$	$A(t)$	$L(t)$	$\overline{L}$
SBM	.3	.3	5.5	5	4.5	1	1	1	5	9	0	3	6	0	2	3	0
	.2	.2	6.1	5.7	5.3	1	1	1	51	59	0	43	57	0	14	18	0
	.1	.3	20	20	20	1	1	1	556	466	7	366	459	58	281	377	10

To further demonstrate the effectiveness of the proposed methodology in both identifying phase transition epochs and in accurately recovering core community structures, we design the following time evolving networks of size $N = 500$ based on the WSBM model. There are $T = 40$ network snapshots in total, with $t = 1, \cdots, 20$ sharing the same community structure with respective sizes 200, 150, 100, and 50. At time $t = 21$ community 1 of size 200 splits into 2 of size 125 and 75 respectively, and this structure persists until $T = 40$. For obtaining the adjacency matrices, we set $L_{out} \overset{i.i.d.}{\sim} \text{Unif}(0, ub)$ and $L_{in} \overset{i.i.d.}{\sim} \text{Unif}(1 - ub, 1)$ and control the *signal strength* by varying ub. Further, we incorporate various levels of dense and sparse noise by controlling r and p_s. As is shown in Figure 2.3, when the noise level increases or equivalently the signal strength decreases, small communities tend to be captured by either the sparse or dense noise components, and only communities of larger size can be recovered. However, as long as a community is detected, its members can be re-

covered with very high accuracy across various levels of noise. As for phase transition epochs detection and identification, since we monitor the *thresholded rank* of averaged low rank matrices, as long as the change is driven by changes in the communities recovered by the proposed procedure, they can be also successfully identified. Indeed, this is the case for all the scenarios tested, except when $ub \geq 0.9$; the latter case is depicted in Figure 2.3(g) where no core community can be detected when $ub \geq 0.9$ and thus the proposed phase transition detection procedure fails. This demonstrates the overall robustness of the proposed procedure, even in the presence of fairly high noise in the data.

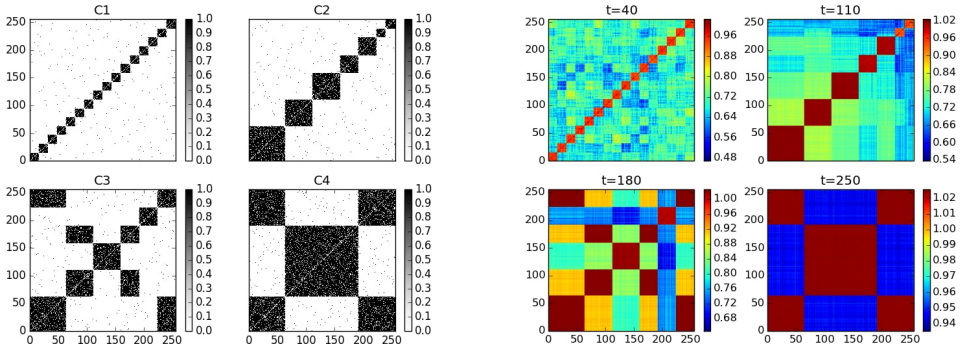


Figure 2.3: Upper row (a-c): presence of only dense noise, with $ub = 0.7$. Middle row (d-f): presence of both dense noise ($r = 0.5$) and varying degree of density of sparse noise with $ub = 0.7$. Lower row (g-i): fixed dense and sparse noise ($r = 0.5$, $p_s = 0.5$), and varying signal strength (ub). (a): Upper left panel, number of communities detected versus magnitude of dense noise r. (b): Upper middle panel, accuracy level of recovery for each community during the first stable period, with $C11$ representing the community of size 200, $C12$ that with 150 nodes, $C13$ with 100 and $C14$ with 50 nodes. (c): Upper right panel, accuracy level of recovery for each community during the second stable period, with $C21$ representing the community of size 150, $C22$ of 125, $C23$ of 100, $C24$ of 75 and $C25$ of 50 nodes, respectively. (d): Middle left panel, number of communities detected versus density of sparse noise p_s. (e): Middle middle panel, accuracy level of recovery for each community during the first stable period. (f): Middle right panel, accuracy level of recovery for each community during the second stable period. (g): Lower left panel, number of communities detected versus signal strength ub. (h): Lower middle panel, accuracy level of recovery for each community during the first stable period. (i): Lower right panel, accuracy level of recovery for each community during the second stable period.

In summary, across three network formation models and a wide range of scenarios examined, the proposed methodology is capable of correctly identifying phase transition epochs, accurately estimate the number of communities, as well as their membership.

2.2.2 Time-Evolving Synchronization Patterns in the Kuramoto Model

Next, we investigate synchronization patterns on networks over time by the dynamical system of Kuramoto oscillators. This model has been extensively studied in

the literature from various angles. In this work, we construct and study the behavior of resulting networks in the following two settings, also examined in [10, 4]: time evloving community structures and two level hierarchical community structures. In both settings, there are $N = 256$ coupled phase oscillators in total and the phase $\theta_i(t)$ of the i-th oscillator evolves in time according to

$$\frac{d\theta_i}{dt} = \omega_i + \sum_j \kappa C_{ij} \sin(\theta_j - \theta_i) \; i = 1, \cdots, N \tag{2.5}$$

where $\omega_i \overset{i.i.d.}{\sim} Normal(0, 1)$ denotes the natural (initial) phase of the oscillators, $\kappa = 0.25$ is the coupling strength, and C is the support coupling matrix, such that oscillators i and j are coupled, if and only if $C_{ij} = 1$. Mapping each oscillator to a node of a network, then the evolution of the model resorts to a time evolving network with the strength of connections represented by the similarities

$$A_{ij}(t) = \left\langle \left| \cos\left[\theta_i(t) - \theta_j(t)\right] \right| \right\rangle \tag{2.6}$$

where the angular brackets stands for the average over 40 different initial random phases.

In the first experimental setting, the total number of network snapshots is $T = 280$, with phase transition epochs occuring at times $\tau_1 = 70, \tau_2 = 140$ and $\tau_3 = 210$. To illustrate the structure of the resulting networks, the support coupling matrices and the corresponding adjacency matrices at times $40, 110, 180$ and 250 are depicted in (a) and (b) of Figure 2.4. Further, the structure of the support matrices where the communities correspond to the dark-colored nodes is generated as follows: for each node in community k of size N_k, there are exactly $14N_k/16$ connections with nodes inside the community, and 1 with nodes outside of it. In stable period 1 (i.e. for all $t \leq \tau_1$), there are 16 equal size communities comprising of $N_k = 16$ nodes each,

indexed by $0, \cdots, 15$; in stable period 2 ($t \in (\tau_1, \tau_2)$) communities 1, 2 and 3 merge with community 0, 5 and 6 with 4, 8 and 9 with 7, 11 with 10, and 13 with 12; in stable period 3 ($t \in (\tau_2, \tau_3)$), communities 14 and 15 merge with the enlarged community 0, community10 with 4 and finally in the last stable period ($t \geq \tau_3$) community 12 merges with 0, and 7 with 4. Time steps for synchronization for the 4 periods are $\tau = 0.1, 0.08, 0.05$ and 0.02 respectively.

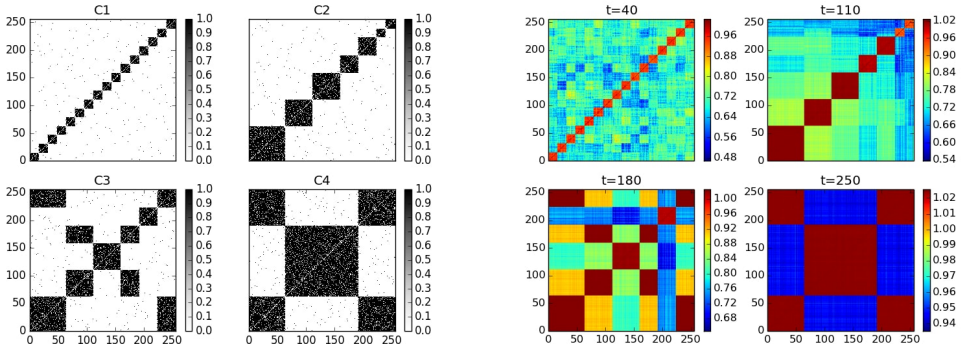


Figure 2.4: (a): Left panel, support coupling matrices for the 4 stable periods. (b): Right panel, adjacency matrices at times $40, 110, 180$ and 250.

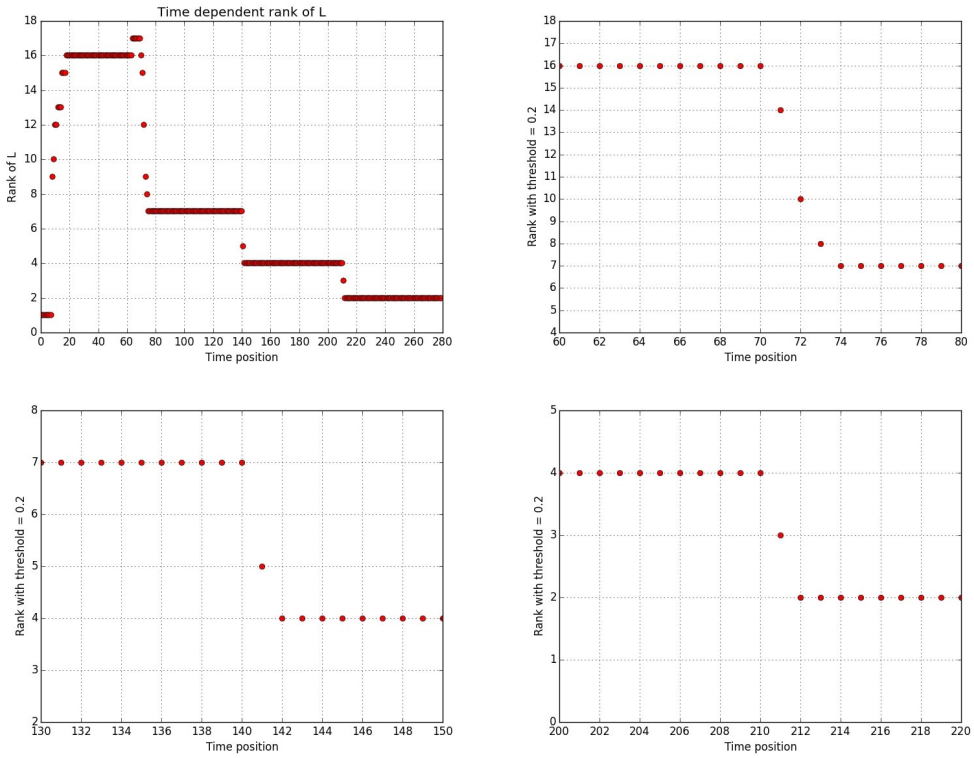


Figure 2.5: (a): Upper left panel, time dependent rank of $L(t)$ recovered. (b): Upper right panel, *thresholded rank* of windows of length 2 around the first phase transition epoch. (c): Lower left panel, *thresholded rank* of windows of length 2 around the second phase transition epoch. (d): Lower right panel, *thresholded rank* of windows of length 2 around the third phase transition epoch.

Following the strategy previously outlined for the selection of tuning parameters in the posited optimization problem, we identify the optimal $\alpha = 0.72$, based on which the rank of recovered $L(t)$ is depicted in Figure 2.5(a). With the final zoomed-in windows of length 2 shown in Figure 2.5(b-d) we estimate the 3 phase transition epochs as follows: $\tau_1 \in (71, 73), \tau_2 = 141$ and $\tau_3 = 211$. Note that in the presence of several small communities merging from stable period 1 to 2, there is some variability in the estimate of the epoch, while the other two epochs are identified precisely. Finally, we calculate the average of the low-rank matrices $L(t)$ in these four stable periods and cluster them to extract their community structure. The community membership extracted from all 280 snapshots (by clustering the $L(t)$ matrices) and

by the average ones over stable periods are shown in (a) and (b) of Figure 2.6. It can be seen that the average ones provide a much clearer identification of the members in each community.

To further investigate the cohesion of communities at different points in time, we also introduce a new metric coined *relative polarization* for a certain community k

$$P_k(t) = \frac{\sum_{i,j=1}^{N_k(t)} A_{ij}^{(k)}(t)}{N_k^2(t) \times P_{out}(t)}, \text{ with } P_{out}(t) = \frac{\sum_{i,j=1}^{N} A_{ij}(t) - \sum_{i,j=1}^{N_k(t)} A_{ij}^{(k)}(t)}{N^2 - \sum_{k=0}^{K-1} N_k^2(t)} \tag{2.7}$$

where $N_k(t)$ is the size of community $k \in \{0, \cdots, K-1\}$ at time t and $A_{ij}^{(k)}(t)$ denotes the edge weight between nodes i and j in community k at time t. Intuitively speaking, the relative polarization of community k is the ratio between the average connection strength within community k and the average strength between all communities.

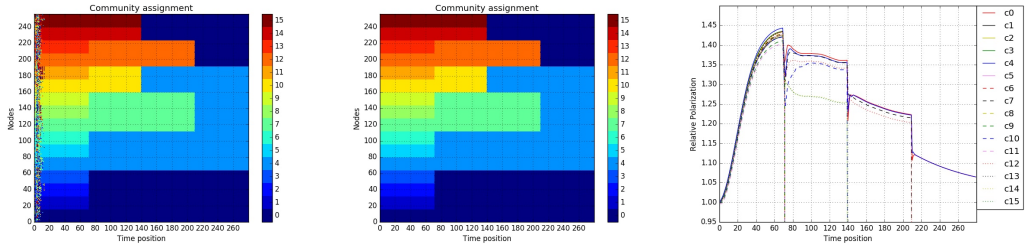


Figure 2.6: (a): Left panel, community membership based on each individual $L(t)$. (b): Middle panel, community membership based on averaged $\overline{L}_m$ of the 4 stable periods. (c): Right panel, relative polarization of communities based on $\overline{L}_m$ across time.

The relative polarization of each community consistently increased during the first stable time period, while in the following stable time periods this pattern reversed, which indicates that the intracommunity synchronization process within each community is dominant in early stages, while in the later stable periods the inter-community synchronization process becomes more important.

In the second experimental setting, we investigate the synchronization process of the Kuramoto model generated network, exhibiting hierarchical structure. Specifi-

cally, each network comprises of 256 nodes (oscillators) evolving across $T = 100$ time points. The coupling strength is $\kappa = 0.25$, time step for synchronization over the whole time range is $\tau = 0.1$, and the adjacency matrices are obtained based on an average over 40 different initial random phases. In the whole range, the community structure is fixed, and the support coupling matrices are designed as follows: there are 16 equally sized first level communities, and 4 equally sized second level communities. For each node, there are exactly 15 connections with nodes in the first level community, 3 in second level (outside the first level) and 1 with nodes outside both levels. This hierarchical structure corresponds to 16 tightly coupled communities that are also organized in 4 more loosely coupled ones. Hence, the presence of this hierarchical structure makes the identification of phase transition epochs, as well as the community structures a challenging problem.

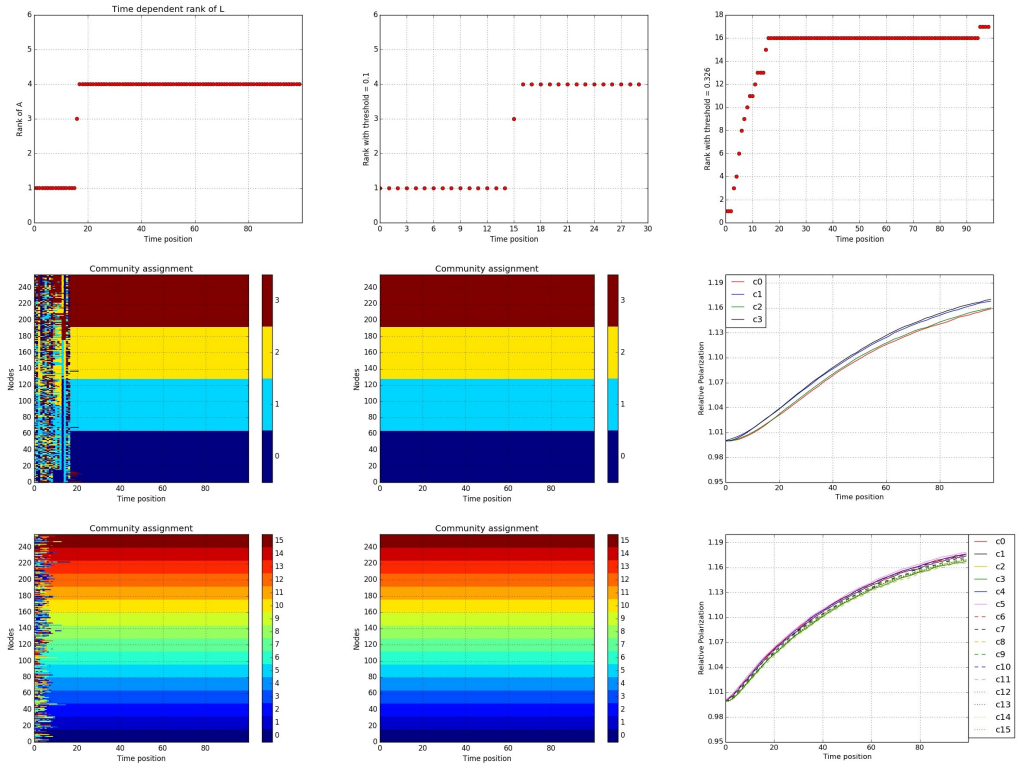


Figure 2.7: (a): Upper left panel, rank of recovered $L(t)$ across time. (b): Upper middle panel, *thresholded rank* of windows of length 2 around the first possible *phase transition epoch* for $L(t)$. (c): Upper right panel, *thresholded rank* of windows of size 2 over the whole time range for $E(t)$. (d): Middle left panel, community membership based on individual $L(t)$. (e): Middle middle panel, community membership based on averaged $\overline{L}$ for $t \geq 16$. (f): Middle right panel, relative polarization of communities based on $\overline{L}$ over time. (g): Lower left panel, community membership based on individual $E(t)$. (h): Lower middle panel, community membership based on averaged $\overline{E}$ for $t \geq 16$. (i): Lower right panel, relative polarization of communities based on $\overline{E}$ over time.

First, we find the optimal $\alpha = 0.35$, based on which the the rank of recovered $L(t)$ is depicted in Figiure 2.7(a). What is interesting is that the dense noise component $E(t)$ actually captures the first level community structure, as can be seen from the results base on window scanning of length 2 in Figure 2.7(c). A first sight on 2.7(b), the *thresholded rank* of averaged $L(t)$ of window length 2 might indicate a phase transition around $t = 16$, but a closer look will show that there is no community structure in the first *stable* period since the *thresholded rank* is 1. Therefore, we try

to recover community structure based on the *thresholded rank* at time $t \geq 16$ for both $L(t)$ and $E(t)$ in the whole time range. Not surprisingly, based on (d) and (g) of Figure 2.7 neither the first level nor the second level community structure forms concretely until approximately time $t = 20$ when the community memberships based on $L(t)$ and $E(t)$ start to match those based on $\overline{L}$ and $\overline{E}$.

2.2.3 Application to U.S. Senate Roll Call Voting Data

A topic of great interest in the political science literature is that of *polarization*, as well as understanding formation of coalitions (groups) over time of members of legislative bodies. Selected prior work includes the work in [106] that analyzed roll call data of the US Congress from 1879 to 1987 (both for the House of Representatives and the Senate), defined a distance measure between the two political parties and calculated its evolution over the corresponding 90-year period. More recently, [89, 87] examined roll call data from the US Senate for the period 1979 to 2012 and used a community modularity quality function to study the issue of polarization. Their key finding is that modularity exhibits a sharp change around early 1995, with members of the two parties drifting apart in their voting pattern. In an attempt to go beyond exploratory analysis, [111] developed a formal estimation framework for the presence of a single change point (phase transition) based on probabilistic graphical models and confirmed the main finding in [89, 87]. Other related findings regarding the time evolving community structures of legislative bodies can be found in [26] based on US Senate roll call data and using synchronization methods, and in [69] that examined legislation bill co-sponsorship networks of the Peruvian Congress and successfully captured the power shifts during the 2006-2011 period using time dependent community detection based on multilayer modularity maximization.

Before discussing the results, we briefly present the necessary data preprocessing steps undertaken. We examine all roll call votes of the US Senate from the 96th to the

113th Congress, covering the period 1979-2014. The nodes of the network represent a specific Senate seat for each state and we mapped the voting records of individual Senators over time to the corresponding seats, ensuring continuity of the voting record for each seat. For each Congress, the adjacency matrix is constructed element-wise as follows: $A_{ij} = \frac{1}{S_{ij}} \sum_{k=1}^{K} c_{ij}^{(k)}$, where S_{ij} is the total number of votes that both senators i and j participated in, K is the total number of votes that particular Congress undertook and $c_{ij}^{(k)} = 1$ if they voted the same way and 0, otherwise. The reason we consider the time-evolving roll call voting record network at the Congress time-scale is that election times that occur every two years wherein 1/3 of the Senate members are up for election represent the main external events that can induce changes in the network composition, including its community structure.

Based on the strategy discussed for selection of tuning parameters, we identify the optimal $\alpha = 0.4$ ($\gamma = \frac{1}{\sqrt{N}} = 0.1$), based on which, we obtain the time dependent rank of $L(t)$ and *thresholded rank* of the averaged $L(t)$ over windows of length 2, as shown in panels (a) and (b) of Figure 2.8. Based on the above results, we conclude that there is a phase transition that occurs in late 1992, when Congress 102 ended its tenure and 103 started its. Further, the identified phase transition occurs a bit earlier (end of 1992) than the one mentioned in [87, 111] (end of 1994). Note that the time dependent rank indicates three communities before the phase transition in late 1992, that correspond to the core Democrat Senate members, the core Republican ones, while the third one includes Senators exhibiting a higher degree of bipartisanship. For the stable period after 1993, the network structure coalesces to two core communities corresponding to the two parties. On the other hand, the averaged across Congresses community structure shows four communities before 1992, that can be categorized as the core Democrats, core Republicans, voting Democrat (most of the time) and voting Republican (most of the time), while the latter two coalesce to a weakly connected community that exhibits some degree of bipartisanship after 1993.

To obtain further insights from this analysis, we also compute for each node i the *sample standard deviation* of its connections with the other nodes for all votes undertaken during the tenure of a Congress and call it *variation*, which represents the polarity of the node:

$$V_i(t) := \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left[A_{ij}(t) - \frac{1}{N}\sum_{k=1}^{N}A_{ik}(t)\right]^2} \tag{2.8}$$

Hence, higher values of V_i indicate stronger agreement with the party vote, while lower values indicate a more bipartisan attitude for the node, since on certain votes they follow the party line and on other ones cross the aisle and follow with their political opponents. Hence, the community structure across all Congresses (before and after the phase transition point identified) are 3 communities. From panels (c) and (d) in Figure 2.8, we can see clearly that the recovered time dependent community structure accurately coincides with the time dependent variation structure in terms of outliers.

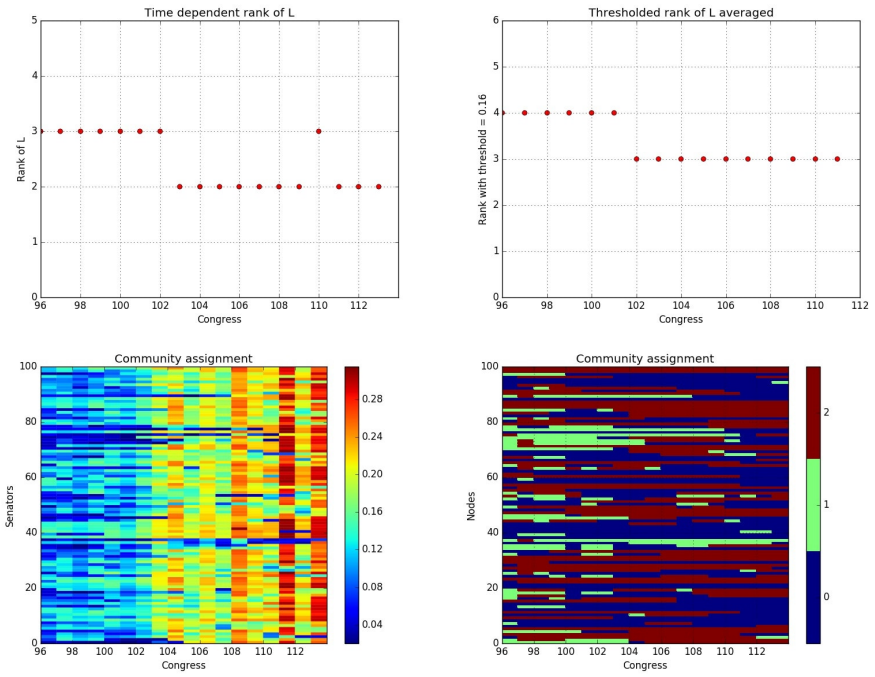


Figure 2.8: (a): Upper left panel, rank of recovered $L(t)$ across time. (b): Upper right panel, *thresholded rank* of windows of size 2 around the phase transition epoch. (c): Lower left panel, *variation* of each node across Congresses. (d): Lower right panel, community membership based on individual $L(t)$ across Congresses.

Further, we plot in Figure 2.9 the time dependent size and *relative polarization* (introduced in equation 2.7) for the community structure obtained based on individual low-rank matrices $L(t)$, as well as that obtained from the two averaged $\overline{L}$ for the before 1992 and after 1993 periods. For community structure based on individual $L(t)$, it can be seen that the size of the non-core community (the one exhibiting a bipartisan voting record) diminishes over time, while the relative polarization of all the three communities becomes higher. It is also worth noting that the relative polarization of the third community significantly increased and approached that of the other two over time. However, this result should not be over-interpreted, since the size of the latter community shrunk over time (note the variation of the respective nodes is still small (Figure 2.8(c)) because of the small community size). On the other hand, the polarization of the third community recovered based on $\overline{L}$ is consistently low. For the

first stable period, this indicates the success in capturing the core communities by the first two identified in the low-rank component, while for the second stable period, it is because the third community includes members from both the *polarized* D and R parties in each network snapshot, which induces the same amount of inter-community links as the intra-community ones, thus leading to low polarization. This is consistent with the results shown in Figure 2.4, where $C3$ for the second stable period captures mainly the party flipping seats as further elaborated on in the ensuing discussion.

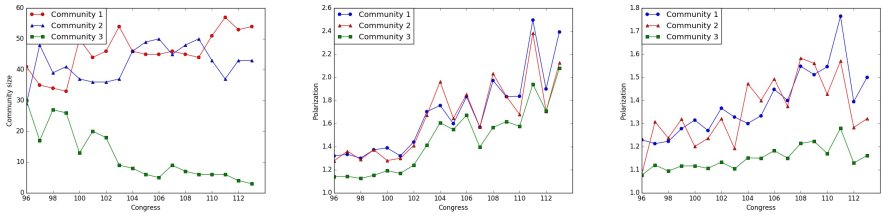


Figure 2.9: (a): Left panel, size of the three communities based on individual $L(t)$ across Congress. (b): Middle panel, relative polarization of the three communities based on individual $L(t)$ across Congress. (c): Right panel, relative polarization of the three communities based on the averaged $\overline{L}_m$ for the two stable periods.

To further understand the community structure, in each stable period we manually assign party affiliations to each node by counting the ratio of each party affiliation by the following criteria. Denoting Republican Senators by R and Democrats by D, we assign nodes to these labels if over 75% of the time the corresponding party affiliation is R (D), otherwise we assign to a so-called mixed M *party*. The mixed *party* indicates that the corresponding Senate seat flips between parties over time, while the core D/R seats are stably held by one party over time. Table 2.4 provides a correspondence between party affiliations (D, R, and M) and the identified communities ($C1$, $C2$, and $C3$) based on averaged $\overline{L}$ for the two stable periods that cover Congress 96-102 and Congress 103-113, respectively. For each of the two stable periods, C1 captures the core "Democratic Party" seats, C2 captures the core "Republican Party" seats, and C3 captures the "Outliers", in terms of voting similarities.

Table 2.4: Misclassification Table

Party	Congress 96-102			Congress 103-113		
	C1	C2	C3	C1	C2	C3
D	29	0	10	33	0	1
R	0	29	10	0	29	3
M	3	0	19	4	4	26

It is expected that nodes in the M *party* should be clustered in the third community ($C3$), which is consistent with the results in Table 2.4. Further, for both time ranges $C1$ basically captures the core members of D while $C2$ captures those of R. An interesting pattern that the analysis identifies is that almost all members of $C3$ for Congress 103-113 come from M, i.e. party flipping, while for Congress 96-102, more than half come from the two parties. This result agrees with what is shown in Figure 2.9(a): namely, there are fewer outliers in Congress 103-113 compared to the previous time range. Finally, to gain further insights of where the third non-core community is more predominant, as shown in Figure 2.10, we translate the results and findings of our analysis by laying them out on the US map, using the following notation for party identification purposes, which is in agreement with recent historical trends in the composition of Congress: $Strong\ D$ for $C1$, $Weak\ D$ for $D \cap C3$, $Strong\ R$ for $C2$, $Weak\ R$ for $R \cap C3$ and $Party\ Flip$ for $M \cap C3$. Intuitively speaking, in each stable period, $Strong\ D(R)$ represents those seats that consistently exhibit strong alignment with the core party members in $D(R)$ in their voting behavior throughout the stable time period, even though some of the seats may not be stably held by the corresponding party (e.g. $M \cap C1(C2)$). Further, $Weak\ D(R)$ indicates seats that consistently display bipartisan trends in voting behavior throughout the stable period although they are stably held by the $D(R)$ party, and finally $Party\ Flip$ represents those seats that, averaged over the stable time period, display bipartisan attitude, but mainly due to the fact that their holders lose the seat and are not re-elected (hence, the party flipping label).

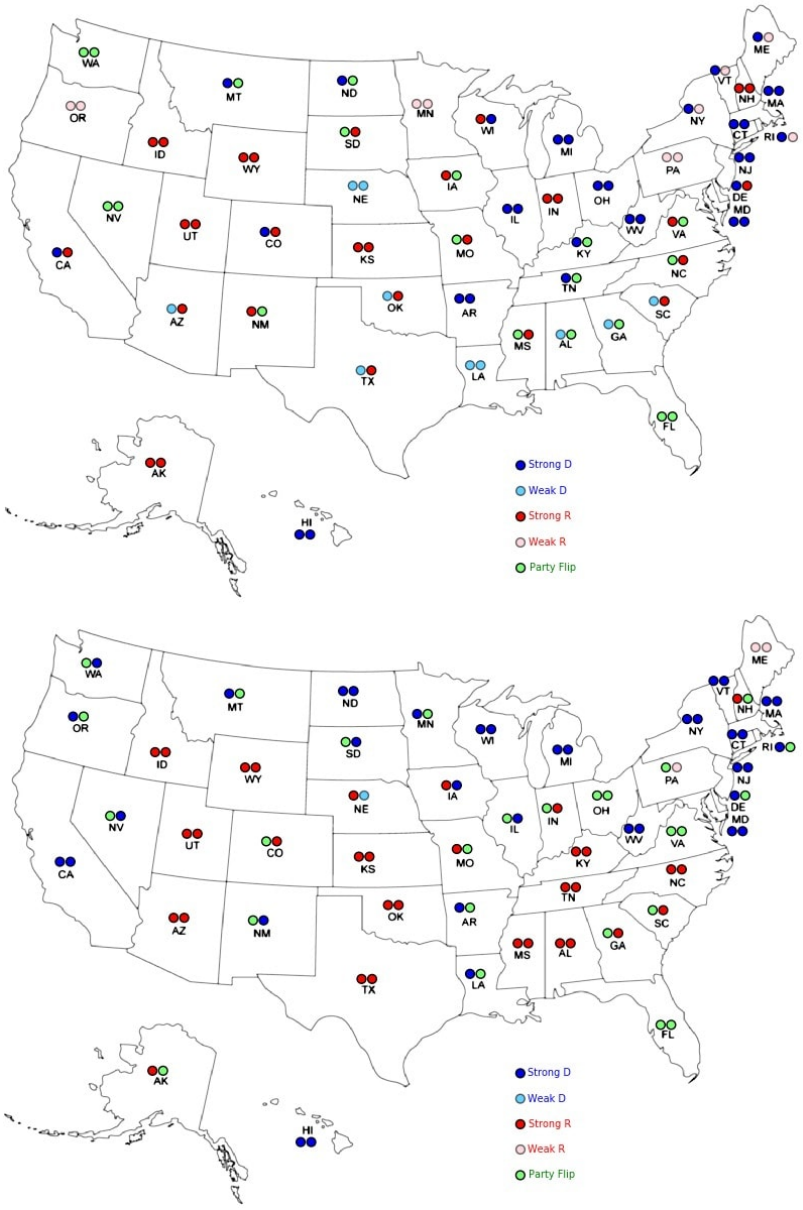


Figure 2.10: Clustering results based on averaged $\overline{L}_m$ over the two stable periods. (a): Upper panel, Congress 96-102. (b): Lower panel, Congress 103-113.

Concluding Remarks: The proposed methodology encompasses a number of network formation models that give rise to network community structure. As illustrated on a number of synthetic and real data examples, it is highly capable of identifying phase transition epochs, estimate accurately the number of communities and their membership. It is further computationally scalable, since obtaining the

decomposition of the network adjacency matrices at each time point is a highly parallelizable step. Hence, the main computational bottleneck stems from the Singular Value Decomposition in obtaining the low-rank components $L(t)$.

Further, note that the model can be further generalized and does not require a fixed low-rank component between phase transition epochs. The key requirement is that the community membership remains fixed as demonstrated through fixed rank, while the strength of their connections can fluctuate, as long as their average strengths over the length of the stable interval converges quickly. Hence, the proposed methodology would still be able to identify the transition epochs and extract the stable community structure provided that the length of the corresponding stable time intervals is adequate.

Finally, it would be of interest to couple the proposed strategy of identifying transition epochs, with more formal methods in change point analysis that come with statistical guarantees, as briefly outlined earlier on. One possibility is to extract the top eigenvectors from the adjacency matrices and then consider them as a multivariate time series. A possible challenge that requires careful study comes from the impact that the sparse and dense noise components considered may have on the extracted eigenvectors.

CHAPTER III

A Susceptible-Informed-Immunized (SIM) model with applications to information diffusion on networks

3.1 Introduction

The modeling and analysis of information diffusion mechanisms on networks, which encompasses a wide class of social spreading phenomena such as rumor and news propagation, adoption of technological innovations, and product advertisement by word of mouth, has attracted attention from various research communities, including physics, computer science and the social sciences [21, 129, 28, 132, 6, 90, 50, 61], due to its important applications in many fields, such as marketing, disease control and economics [86, 129, 88, 65, 49, 61]. A number of modeling formalisms have been used, appropriately modified, for the task at hand; two of the most popular ones are threshold models and independent interaction models [21, 31]. Threshold models that encompass the cascade, the linear threshold and popular vote models, as well as game theoretic approaches, assume that the diffusion occurs to a new node on the network when a certain proportion of neighbors that has already adopted the information exceeds a prespecified threshold [61, 88, 49, 132]. Another popular mechanism adopts the formalism of disease transmission processes and hence information diffu-

sion is treated as an epidemic process. Popular models for infectious disease spreading include the Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Removed (SIR) ones [33, 105, 129, 123, 86, 65]. Both these models belong to the class of independent interaction models that assume successive trials of infection/diffusion between two nodes are independent with the same probability p of success. There are a number of approaches to analyze these models and address questions such as the percentage of nodes being infected by a certain time point, including mean field theory and under certain conditions the SIR model can be mapped to bond percolation theory whose roots come from statistical physics [9, 113, 129, 47, 12, 16, 93, 95, 58].

Next, we elaborate on the mechanics of the SIR model. In this model, nodes in a network of size N are classified into three states: S for Susceptible, I for Infected/Infectious, and R for Removed/Recovered. The simplest version of the model assumes that nodes in state I would infect any of their neighbors in state S with probability β per unit time and will be removed or recover from state I to R with probability γ per unit time. Hence, a homogeneous rate of disease transmission between infectious and susceptible nodes is considered, which holds for unweighted networks [33, 105, 93]. By defining the expected transmissibility as the total probability of a susceptible node getting infected from a infectious neighbor and allowing the system to evolve over an infinite time horizon, then the former is given in the discrete time case,

$$T = \sum_{t=1}^{\infty} (1-\beta)^{t-1} \beta (1-\gamma)^{t-1} = \frac{\beta}{\beta + \gamma - \beta\gamma} \tag{3.1}$$

where it is assumed that the infection process precedes the removal/recovery process. Analogously, in the continuous time case, we obtain

$$T = \int_0^{\infty} e^{-\beta t} \beta e^{-\gamma t} dt = \frac{\beta}{\beta + \gamma} \tag{3.2}$$

which is a constant independent of the node from which the infection originates and leverages results in bond percolation theory. Specifically, in percolation theory, one is normally interested in two characteristics of the infection process: the size of the epidemic s, defined as the expected proportion of the network nodes being informed/infected at large times (as time grows to infinity), and the epidemic threshold T_C, defined as the minimum transmissibility T that can generate an epidemic; i.e., at very large times, the number of informed/infected nodes sN reaches the order $O(N)$ before dying out as the size of the network $N \to \infty$. The threshold T_C can be obtained from the perspective of the reproduction number R_0 in the following way: if we denote q_k as the *excess degree* (total degree -1) distribution of a neighbor reached by a randomly chosen edge, we have [95]:

$$R_0 = T\sum_{k=0}^{\infty} kq_k = T\sum_{k=0}^{\infty} k\frac{(k+1)p_{k+1}}{\langle k\rangle} = T\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle} \tag{3.3}$$

In order to have an epidemic, we need to ensure $R_0 > 1$, and therefore we have the transmissibility threshold $T_C = \frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}$, where $\langle k\rangle$ and $\langle k^2\rangle$ are the first and second moments of k.

To find the epidemic size s, we define u as the average probability of a node being not connected to the giant component (GC, cluster of connected nodes of size $O(N)$ as $N \to \infty$) via a randomly chosen edge. There are two ways for this event to ocur, either the edge is removed or the edge is intact but the neighbor following that edge is not connected to the GC via any of its excess edges. Denote s as the expected proportion of nodes in the network captured by the GC, one has[95]:

$$u = 1 - T + T\sum_{k=0}^{\infty} q_k u^k \tag{3.4}$$

$$s = 1 - \sum_{k=0}^{\infty} p_k u^k \tag{3.5}$$

Once we have the solution of u for Eq. 3.4 we can plug in to Eq. 3.5 and get the expected infection size s.

However, mechanisms of information diffusion can be quite different from the epidemic process since as pointed out in [27] the information diffusion process usually slows down over time because of the *reluctance to tell stale news* from the spreaders' point of view. Intuitively speaking, a piece of news/information from some time ago would be very unlikely mentioned today, while a disease outburst years ago would tend to infect a susceptible person equally easily today. Therefore, we can see that in general information diffusion process, the *information* often possesses *intrinsic time value* and thus it is not very appropriate to directly apply models such as SIR in epidemic process to the diffusion process. In order to address this problem, a lot of proposals in the literature have appeared, including the DK model proposed by Daley and Kendall in the year 1965 and the Maki-Thompson model proposed by Maki and Thompson in the year 1973, of which the former has gained significant popularity and has been very well studied [27, 119, 91, 78]. In the DK model, similar to the three states classification in SIR, the nodes are classified into three exclusive, but exhaustive classes: ignorants, spreaders and stiflers corresponding to S, I and R states in the SIR model and similarly we use s, i, r to represent the proportion of nodes in the three states. Instead of being proportional to $i(t)$, the DK model assumes that the removal/recovery rate of $i(t)$ is proportional to $i(t)[i(t) + r(t)]$, which takes into consideration the effect of "the losing value of news" when reaching a node already aware of it ($i(t)$ or $r(t)$) from the spreaders' perspective. In this way, we are effectively setting the recovery rate to be time dependent $\gamma(t) \sim \gamma[i(t) + r(t)]$.

One limitation of this model is that it does not capture the behavior of the "ignorant" actors/nodes. For example, in online news propagation, due to the competition for attention from new story lines, a *news cycle* arises because both spreaders and "ignorant" actors in the population tend to focus on more recent and dominant news

stories/topics [72]. Similarly, in rumor spreading, *reluctance to tell stale news* is co-existent with *reluctance to listen to stale news* indicating that the "ignorant" actors are not interested in talking about topics involving old news. Another example is in viral marketing where it is the customers who lose interest in adopting product as time goes by instead of the the recommenders/advertisements losing interest[71]. To address this issue, we introduce an alternative mechanism coined Susceptible-Informed-Immunized (SIM) model to capture the *intrinsic time value* of *information* from the perspectives of both the ignorant nodes and spreaders and provide a rigorous analysis of their behavior.

In the SIM model, we assume a constant infection/diffusion rate β, but let both the spreaders and ignorants go through an immunization process with rate γ. In this manner, the "losing interest" process is represented by the process of immunization that excludes those nodes from any further diffusion and thus the standard SI process would evolve on the residual network, defined as the network containing all nodes not immunized up to time t and their corresponding connections. This model successfully captures the time value property of information as follows: a certain node informed at an early time $t = \tau_1$ exhibits better capacity of diffusing further the information to the rest of the network compared to becoming informed at a later time $t = \tau_2 > \tau_1$, because some of the susceptible neighbors (expected proportion $1 - (1 - \gamma)^{(\tau_2 - \tau_1)} \sim 1 - e^{-\gamma(\tau_2 - \tau 1)}$) have already being immunized.

The analysis of this model exhibits a number of technical challenges. Standard arguments from percolation theory are no longer applicable -since the homogeneous expected transmissibility T mechanism for each edge does not hold- and thus other techniques are required. We employ a mean field (MF) approximation based on collective macro effects, and also construct a Maximum Weight Tree (MWT) based approximation that leverages individual micro dynamics to achieve our objective. The remainder of the paper is organized as follows: in Section 3.2, we consider both

homogeneous and heterogeneous versions of MF and compare the results of these approximations to those obtained from Monte Carlo (MC) simulation on a number of experimental settings. In section 3.3.1, we derive an analytical formula for the total probability of getting informed for each node in the network depending on its path to the "seed node", defined as the node informed at the initial stage. Subsequently, we construct a MWT to approximate the saturated total ratio of nodes getting informed at large times. The agreement between the MWT approximation and MC simulation results increases markedly, for sparser network topologies.

3.2 Mean Field Approximation

3.2.1 Homogeneous Mean Field Approximation

To facilitate the analysis for the general case, when the degree distribution is taken into consideration, we fist consider the homogeneous mean field (HMF) approximation where we assume all nodes are statistically equivalent with the same degree $\langle k \rangle$ [12, 9]. For the purpose of simplifying the statement of the main results, in the following we restrict ourselves on the residual network, except when otherwise stated. At time t, the expected number of susceptible nodes with degree $m(\leq \langle k \rangle)$ is

$$S_m(t) = \binom{\langle k \rangle}{m} \alpha_t^m (1 - \alpha_t)^{\langle k \rangle - m} S(t) \tag{3.2.1.1}$$

which represents the case when exactly m out of $\langle k \rangle$ neighbors survived the immunization process. Here $S(t)$ is the total number of susceptible nodes, and $\alpha_t = e^{-\gamma t}$ is the survival probability of the immunization process. If we denote $I_m(t)$ as the

number of informed nodes with degree m at time t, we have,

$$\begin{aligned} dI_m(t) =& S_m(t)\big[1-(1-\beta dt)^{m\Theta_m(t)}\big] - \gamma I_m(t)dt \\ \overset{dt\to 0}{=}& S_m(t)\beta m\Theta_m(t)dt - \gamma I_m(t)dt \end{aligned} \tag{3.2.1.2}$$

where $\Theta_m(t)$ is the probability of reaching an informed neighbor starting from an arbitrary edge of a susceptible node of degree m at time t. The expected total number of stubs (half edges) in the residual network is $\alpha_t N\alpha_t\langle k\rangle$, where N is the number of nodes in the original network, because the expected number of nodes in the residual network is $\alpha_t N$ and the expected degree of nodes is $\alpha_t\langle k\rangle$. Similarly, the expected number of stubs from seed nodes is $\alpha_t I(0)\alpha_t\langle k\rangle$, where $I(0)$ is the number of seed nodes. The expected number of newly informed nodes is $I(t)-\alpha_t I(0)$, where $I(t)$ is the total number of informed nodes in the residual network. If we examine these newly informed nodes on the original network, then each one of them has one edge connecting to another informed node from which it received the information. Then on the residual network, the survival probability of a node that transmitted the information to a newly informed node with degree m is approximately $\phi_m(t)\approx m/\langle k\rangle$, if we assume that the diffusion event occurs early enough such that its probability of survival is approximately the same as that of the other neighbors. Therefore, the expected number of stubs available for connection from the newly informed nodes of degree m is $[I(t)-\alpha_t I(0)]\binom{\langle k\rangle}{m}\alpha_t^m(1-\alpha_t)^{\langle k\rangle-m}[m-\phi_m(t)]$. Thus,

$$\begin{aligned} \Theta_m(t) &= \frac{\sum_{m=1}^{\langle k\rangle}\big[I(t)-\alpha_t I(0)\big]\binom{\langle k\rangle}{m}\alpha_t^m(1-\alpha_t)^{\langle k\rangle-m}[m-\phi_m(t)] + \alpha_t I(0)\alpha_t\langle k\rangle}{\alpha_t N\alpha_t\langle k\rangle} \\ &\approx \frac{\big[I(t)-\alpha_t I(0)\big](\alpha_t\langle k\rangle-\alpha_t)+\alpha_t I(0)\alpha_t\langle k\rangle}{\alpha_t N\alpha_t\langle k\rangle} = \frac{I(t)(\langle k\rangle-\delta_t)}{\alpha_t N\langle k\rangle} \end{aligned} \tag{3.2.1.3}$$

where $\delta_t = 1-\alpha_t I(0)/I(t)$ is represented as a function of time instead of simply been set as 0 or 1 as is normally done in the literatures[129, 12, 9]. This correction

significantly improves the agreement between MF and the MC results, as shown later on. For the total number of informed nodes, we then get

$$\begin{aligned} dI(t) &= \sum\nolimits_{m=0}^{\langle k\rangle} dI_m(t) = \sum\nolimits_{m=0}^{\langle k\rangle} S_m(t)\beta m\Theta_m(t)dt - \sum\nolimits_{m=0}^{\langle k\rangle} \gamma I_m(t)dt \\ &= \sum\nolimits_{m=0}^{\langle k\rangle} \binom{\langle k\rangle}{m} \alpha_t^m (1-\alpha_t)^{\langle k\rangle - m} S(t)\beta m\Theta_m(t)dt - \gamma I(t)dt \\ &= S(t)\beta\frac{I(t)}{N}(\langle k\rangle - \delta_t)dt - \gamma I(t)dt \end{aligned} \tag{3.2.1.4}$$

If we set $\rho(t) = I(t)/N$, $s(t) = S(t)/N$ and $r(t) = R(t)/N = 1 - \rho(t) - s(t)$, where $R(t)$ is the total number of nodes immunized by time t, we would have the following ordinary differential equations (ODE) governing the evolution of the system,

$$\begin{aligned} \dot{\rho}(t) &= s(t)\rho(t)\beta(\langle k\rangle - \delta_t) - \gamma\rho(t) \\ \dot{s}(t) &= -s(t)\rho(t)\beta(\langle k\rangle - \delta_t) - \gamma s(t) \\ \dot{r}(t) &= \gamma\big[\rho(t) + s(t)\big] = \gamma\big[1 - r(t)\big] \end{aligned} \tag{3.2.1.5}$$

After solving for $r(t) = 1 - e^{-\gamma t}$ directly, we obtain,

$$\dot{\rho}(t) = \big[e^{-\gamma t} - \rho(t)\big]\rho(t)\beta(\langle k\rangle - \delta_t) - \gamma\rho(t) \tag{3.2.1.6}$$

Note that $\rho(t)$ represents the proportion of nodes informed and currently active in the residual network, and to find the total proportion of nodes $q(t)$ ever informed up to time t, we observe that within an infinitesimal time interval $[t, t + dt]$, the change of $\rho(t)$ comprises two components. The first are the newly informed nodes, while the other are the immunized ones. For $q(t)$ only the first component contributes and therefore,

$$dq(t) = d\rho(t) + \gamma\rho(t)dt \Rightarrow \dot{q}(t) = \big[e^{-\gamma t} - \rho(t)\big]\rho(t)\beta(\langle k\rangle - \delta_t) \tag{3.2.1.7}$$

Equations 3.2.1.6 and 3.2.1.7 can be combined and solved numerically.

To study the early stage evolution, we start with a small $\rho(0)$, such that we can safely ignore the terms of order $O(\rho^2(t))$ when $t \to 0$, and for simplifying the calculation, we set $\delta_t = 1$ as commonly done in the literature. Then,

$$\dot{\rho}(t) = e^{-\gamma t}\rho(t)\beta(\langle k\rangle - 1) - \gamma\rho(t) \tag{3.2.1.8}$$

$$\Rightarrow \rho(t) = \rho(0)exp\big[\frac{\beta}{\gamma}(\langle k\rangle - 1)(1 - e^{-\gamma t}) \ - \gamma t\big] \approx \rho(0)exp\big\{\big[\beta(\langle k\rangle - 1) - \gamma\big]t\big\} \tag{3.2.1.9}$$

where we have used the approximation $e^{-\gamma t} \approx 1 - \gamma t$. To reach an epidemic, we need to ensure a positive exponent, and thus $(\beta/\gamma)_\rho > \frac{1}{\langle k\rangle - 1}$. As for the *phase transition time point* for $\rho(t)$, defined as the time that the number of informed nodes active in the residual network first starts to decrease, following Eq. 3.2.1.8 we obtain

$$\dot{\rho}(\tau) = 0 \Rightarrow \tau = \frac{1}{\gamma}log\big[\frac{\beta}{\gamma}(\langle k\rangle - 1)\big] \tag{3.2.1.10}$$

Similarly, the *phase transition time point* for $q(t)$, defined as the time that the growth in the total number of nodes ever informed first starts to decrease, following equations 3.2.1.7 and 3.2.1.8 and ignoring $O(\rho^2(t))$ terms,

$$\ddot{q}(t) = e^{-\gamma t}\big[\dot{\rho}(t) - \gamma\rho(t)\big]\beta(\langle k\rangle - 1) = e^{-\gamma t}\big[e^{-\gamma t}\beta(\langle k\rangle - 1) - 2\gamma\big]\rho(t)\beta(\langle k\rangle - 1) \tag{3.2.1.11}$$

Therefore, the phase transition time point for $q(t)$ is

$$\tau_q = \frac{1}{\gamma}log\big[\frac{\beta}{2\gamma}(\langle k\rangle - 1)\big] \tag{3.2.1.12}$$

while the corresponding *epidemic threshold* is $(\beta/\gamma)_q > \frac{2}{\langle k\rangle - 1}$.

A possible application of the phase transition time point in marketing is that it

provides of marketeers when to stop or decrease spending on advertisement since the marginal profit (e.g. the marginal diffusion size minus the expenditures on advertising per unit time) starts to decrease. This is consistent with results from the literature in the field of dynamic optimal advertising control in marketing, where expenditures on advertising are expected to be time dependent and shall be terminated after certain conditions are met in some models[114, 56].

As stated before, for simplicity we set $\delta_t = 1$ as is normally done in the literature, which results in inaccurate estimations. To improve the approximation, we next set $\delta_t = \delta_0 = 0$ and following equation 3.2.1.9 we obtain $\rho(t) \approx \rho(0)exp\big\{\big[\beta\langle k\rangle - \gamma\big]t\big\} \Rightarrow$ $\delta_t \approx 1 - e^{-\beta\langle k\rangle t}$. Then, plugging back into equation 3.2.1.6, and ignoring the order $O(\rho^2(t))$ terms, and solve to get

$$\begin{aligned}
&\dot{\rho}(t) = \big[e^{-\gamma t} - \rho(0)e^{(\beta\langle k\rangle-\gamma)t}\big]\rho(t)\beta\big[\langle k\rangle - 1 + e^{-\beta\langle k\rangle t}\big] - \gamma\rho(t) \Rightarrow \rho(t) = \rho(0)e^{f_1(t)-f_2(t)} \\
&f_1(t) = \frac{\beta}{\gamma}(\langle k\rangle - 1)(1 - e^{-\gamma t}) \ - \gamma t + \frac{\beta}{\gamma + \beta\langle k\rangle}\big[1 - e^{-(\gamma+\beta\langle k\rangle)t}\big] \\
&f_2(t) = \rho(0)\beta\Big[\frac{\langle k\rangle - 1}{\beta\langle k\rangle - \gamma}\big[e^{(\beta\langle k\rangle-\gamma)t} - 1\big] + \frac{1 - e^{-\gamma t}}{\gamma}\Big]
\end{aligned} \tag{3.2.1.13}$$

where we have assumed $\beta/\gamma > 1/\langle k\rangle$. Then, the phase transition time point τ is given by the solution of the equation,

$$\dot{\rho}(t) = 0 \Rightarrow \big[e^{-\gamma t} - \rho(0)e^{(\beta\langle k\rangle-\gamma)t}\big]\beta\big[\langle k\rangle - 1 + e^{-\beta\langle k\rangle t}\big] - \gamma = 0 \tag{3.2.1.14}$$

while the epidemic threshold is calculated from $t = 0$: $\big[1 - \rho(0)\big]\beta/\gamma \sim (\beta/\gamma)_\rho > 1/\langle k\rangle$. Similarly, τ_q can be obtained from the root of the equation

$$\ddot{q}(t) = 0 \Rightarrow \big[e^{-\gamma t} - \rho(0)e^{(\beta\langle k\rangle-\gamma)t}\big]\beta\big[\langle k\rangle - 1 + e^{-\beta\langle k\rangle t}\big] - 2\gamma = 0 \tag{3.2.1.15}$$

It follows the epidemic threshold for $q(t)$ is $(\beta/\gamma)_q > 2/\langle k\rangle$. Equations 3.2.1.14 and 3.2.1.15 can be solved numerically by standard root finding algorithms such as

Newton's method or the bisection method. Following bond percolation theory in the analysis of the SIR epidemic threshold, to reach an epidemic, the reproduction number originating from a newly informed node should be greater than 1. Therefore, to find the epidemic threshold we need to set $\delta_t = 1$ and thus $(\beta/\gamma)_\rho > \frac{1}{\langle k \rangle - 1}$ and $(\beta/\gamma)_q > \frac{2}{\langle k \rangle - 1}$. As will be shown in the following experiments, the superiority of incorporating $\delta_t = 1 - \alpha_t \rho(0)/\rho(t)$ in the analysis instead of simply setting $\delta_t = 1$ is that it results in better agreements between the MF with the MC results.

Performance Evaluation of the Homogeneous MF Approximation: We compare the MC simulation results with the theoretical ones obtained from the derived MF approximation. The networks are generated such that their size is $N = 5000$ and the degree of each node follows a Poisson distribution of rate λ: $k \overset{iid}{\sim} Pois(\lambda)$. Results in each of the settings are averaged over 10^3 realizations. As is well known, the accuracy of the MF approximation depends on the saturated infection/informed ratio [47], so we control the parameters β, γ for each case to ensure that the saturated ratios are all approximately $q(\infty) \sim 0.5$ for being able to make comparisons across experimental settings. Figure 3.1 depicts the results for $\lambda = 5, 10, 25$ representing sparse, medium and dense networks. In the settings considered, the agreement between the MC simulation estimates and the numerical solution to the ODEs for both $\rho(t)$ and $q(t)$ is very good for denser networks, that correspond to higher average excess degrees [47].

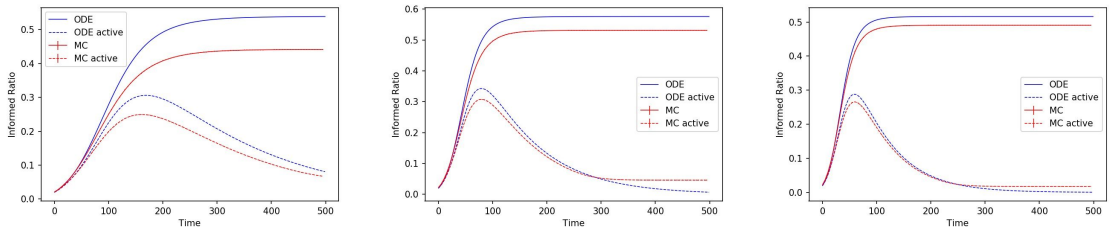


Figure 3.1: From left to right: (a-c), $\lambda = 5, 10, 25$. Time dependent diffusion ratio for both $\rho(t)$ and $q(t)$ based on MC simulation and the numerical solution of ODE equations for networks of different densities. *ODE* stands for the numerical solution of equations 3.2.1.6 and 3.2.1.7 and *MC* stands for MC simulation. $\rho(0) = 2\%$ for all cases.

Next, we investigate the early stage evolution by comparing the MC simulation with the analytical solutions 3.2.1.9 and 3.2.1.13, as is shown in Figure 3.2. The latter, after incorporating δ_t, shows a higher degree of concordance with the MC simulated results in the whole range of network densities, while for the former the agreement with MC simulation improves for denser networks, which is expected, since higher $\langle k \rangle$ would screen out the correction effect of the factor δ_t.

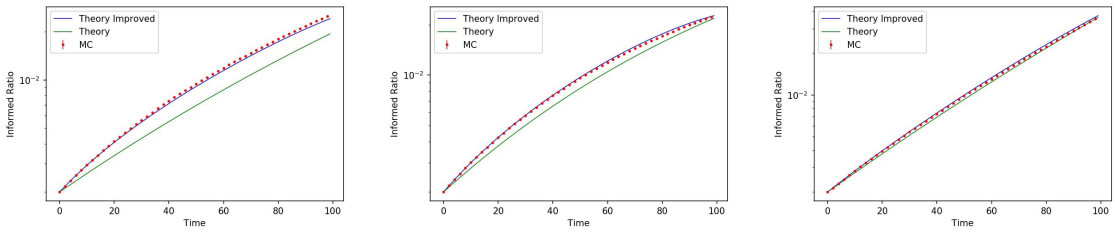


Figure 3.2: From left to right: (a-c), $\lambda = 3, 5, 10$. Early stage evolution of the diffusion ratio with comparison between MC simulation and analytical results for networks of different densities. *Theory Improved* stands for equation 3.2.1.13, and *Theory* for equation 3.2.1.9. For all the cases, $\rho(0) = 0.2\%$

Finally, we investigate the phase transition time points τ under different combinations of β and γ, and, as is shown in Figure 3.3, we see much better agreement with MC after incorporating δ_t. For illustration purposes, we also include the diffusion

ratio $q(\infty)$ under the same settings.

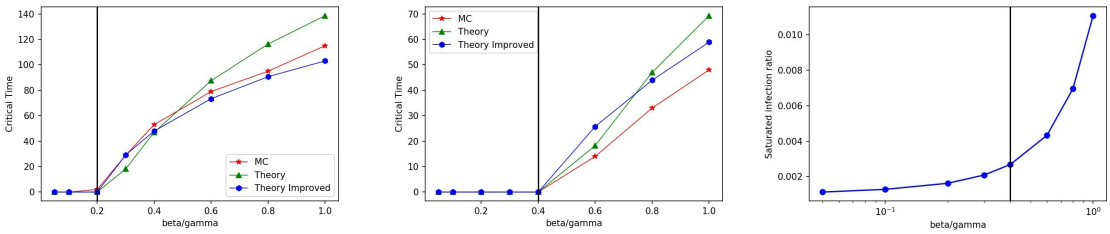


Figure 3.3: From left to right: (a), τ for $\rho(t)$; (b) τ for $q(t)$; (c) Total diffusion ratio $q(\infty)$. Phase transition time points for both $\rho(t)$ and $q(t)$ with random network parameter $\lambda = 5$ and the initial ratio of seed nodes $\rho(0) = 0.1\%$. Also included is the diffusion ratio for different values of β/γ. *Theory improved* stands for equations 3.2.1.14 and 3.2.1.15, and *Theory* stands for equations 3.2.1.10 and 3.2.1.12. For visual guidance, we include a black vertical line located at $1/\langle k \rangle$ in (a) and $2/\langle k \rangle$ in (b-c)

3.2.2 Heterogeneous Mean Field Approximation

For most applications of information diffusion, the assumption of homogeneous degree distribution of nodes is rather restrictive. Therefore, in the following, we take into consideration of the degree distribution and employ a heterogeneous MF approximation. We classify nodes into groups based on degrees in the original network and we assume that nodes within the same group would behave equivalently in the statistical sense. First, we constrain ourselves within the group of nodes with degree k on the original network; then, at time t the expected number of susceptible nodes with degree $m(\leq k)$ from this group in the residual network is,

$$S_m^{(k)}(t) = \binom{k}{m} \alpha_t^m (1-\alpha_t)^{k-m} S_k(t) \tag{3.2.2.1}$$

where $S_k(t)$ is the total number of susceptible nodes in the residual network coming from the group k and $\alpha_t = e^{-\gamma t}$ is the survival probability of the immunization process. Similarly, if we denote $I_m^{(k)}(t)$ as the total number of informed nodes with

degree m and $I_k(t)$ as the total number of informed nodes on the residual network that comes from the group k, we get,

$$\begin{aligned} dI_m^{(k)}(t) &= S_m^{(k)}(t)\big[1-(1-\beta dt)^{m\Theta_m^{(k)}(t)}\big] - \gamma I_m^{(k)}(t)dt \\ &\overset{dt\to 0}{=} S_m^{(k)}(t)\beta m\Theta_m^{(k)}(t)dt - \gamma I_m^{(k)}(t)dt \end{aligned} \tag{3.2.2.2}$$

where $\Theta_m^{(k)}(t)$ is the probability of reaching an informed neighbor following an arbitrary edge of a susceptible node with degree m on the residual network that comes from the group k. To simplify calculations, we assume that the degrees across groups are not correlated, which is the case for networks generated by some popular random graph models [33]. Then, $\Theta_m^{(k)}(t)$ can be calculated independent of k and this constraint of no degree degree correlation can be easily relaxed[12, 113]. Following the argument in homogeneous MF case, and considering the contribution from nodes of all possible degrees we have,

$$\begin{aligned} \Theta(t) &= \frac{\sum_{k'=1}^{N-1}\Big\{\sum_{m=1}^{k'}\big[I_{k'}(t)-\alpha_t I_{k'}(0)\big]\binom{k'}{m}\alpha_t^m(1-\alpha_t)^{k'-m}(m-\phi_m^{(k')}(t)) + \alpha_t I_{k'}(0)\alpha_t k'\Big\}}{\sum_{k'=1}^{N-1}\alpha_t N_{k'}\alpha_t k'} \\ &\approx \frac{\sum_{k'=1}^{N-1}\Big\{\big[I_{k'}(t)-\alpha_t I_{k'}(0)\big](\alpha_t k'-\alpha_t) + \alpha_t I_{k'}(0)\alpha_t k'\Big\}}{\sum_{k'=1}^{N-1}\alpha_t N_{k'}\alpha_t k'} \\ &= \frac{\sum_{k'=1}^{N-1} I_{k'}(t)[k'-\delta_{k'}(t)]}{\sum_{k'=1}^{N-1}\alpha_t N_{k'}k'} = \frac{1}{\alpha_t\langle k\rangle}\sum_{k'=1}^{N-1}[k'-\delta_{k'}(t)]\rho_{k'}(t)P(k') \end{aligned} \tag{3.2.2.3}$$

where $P(k)$ is the probability of a node with degree k on the original network, $N_k = NP(k)$, $\langle k\rangle = \sum_{k=1}^{N-1}P(k)k$ is the mean degree, $\rho_{k'}(t) = I_{k'}(t)/N_{k'}$, $\delta_{k'}(t) = 1 - \alpha_t\rho_{k'}(0)/\rho_{k'}(t)$, and similar to homogeneous MF case, we assumed that approximately

$\phi_m^{(k')}(t) \approx m/k'$. Thus,

$$\begin{aligned} dI_k(t) &= \sum_{m=0}^{k} dI_m^{(k)}(t) = \sum_{m=0}^{k} \left[S_m^{(k)}(t)\beta m \Theta_m^{(k)}(t)dt - \gamma I_m^{(k)}(t)dt \right] \\ &= \sum_{m=0}^{k} \binom{k}{m} \alpha_t^m (1-\alpha_t)^{k-m} S_k(t)\beta m \Theta_m^{(k)}(t)dt - \sum_{m=0}^{k} \gamma I_m^{(k)}(t)dt \\ &= S_k(t)\beta\alpha_t k\Theta(t)dt - \gamma I_k(t)dt \end{aligned} \tag{3.2.2.4}$$

In the following, we denote by $s_k(t) = S_k(t)/N_k$, $r_k(t) = R_k(t)/N_k$, and $q_k(t) = Q_k(t)/N_k$, where $R_k(t)$ is the total number of nodes immunized and $Q_k(t)$ is the total number of nodes ever informed that come from group k in the original network up to time t . We then obtain the following set of ODE equations,

$$\begin{aligned} \dot{\rho}_k(t) &= s_k(t)\beta\alpha_t k\Theta(t) - \gamma\rho_k(t) \\ \dot{s}_k(t) &= -s_k(t)\beta\alpha_t k\Theta(t) - \gamma s_k(t) \\ \dot{r}_k(t) &= \gamma[\rho_k(t) + s_k(t)] = \gamma[1 - r_k(t)] \\ \dot{q}_k(t) &= \dot{\rho}_k(t) + \gamma\rho_k(t) = s_k(t)\beta\alpha_t k\Theta(t) \end{aligned} \tag{3.2.2.5}$$

Plugging in equation 3.2.2.3 and the solution $r_k(t) = 1 - \alpha_t$, we get,

$$\dot{\rho}_k(t) = \frac{1}{\langle k \rangle}[\alpha_t - \rho_k(t)]\beta k \sum_{k'=1}^{N-1} [k' - \delta_{k'}(t)]\rho_{k'}(t)P(k') - \gamma\rho_k(t) \tag{3.2.2.6}$$

Similarly, equations 3.2.2.5 and 3.2.2.6 can be combined and solved numerically for $\rho(t)$ and $q(t)$. To investigate the epidemic property of the system, we assume a small $\rho(0)$ such that it is reasonable to ignore the terms of order $O(\rho^2(t))$ as $t \to 0$. Following the development in the homogeneous MF, we first set $\delta_k(t) = 1$ to obtain

$$\dot{\rho}_k(t) = \frac{\beta e^{-\gamma t}}{\langle k \rangle} \sum_{k'=1}^{N-1} k(k'-1)\rho_{k'}(t)P(k') - \gamma\rho_k(t) \tag{3.2.2.7}$$

Denote by $\vec{\rho}(t) = \big(\rho_1(t), \rho_2(t), \cdots, \rho_{N-1}(t)\big)^T$ and $A_t = \big[\frac{\beta e^{-\gamma t}}{\langle k \rangle} i(j-1)P(j) - \gamma\delta_{ij}\big] \in \mathbf{R^{(N-1)\times(N-1)}}$, where $\delta_{ij} = 1$ if $i = j$ and zero otherwise, then following equation 3.2.2.7, we get

$$\dot{\vec{\rho}}(t) = A_t\vec{\rho}(t) \tag{3.2.2.8}$$

If $A_\tau = A$ for $\tau \in [0, t]$ is a constant matrix, we have for equation 3.2.2.8 at time t a general solution,

$$\vec{\rho}(t) = \sum\nolimits_{i=1}^{N-1} C_i e^{\lambda_i t}\vec{v}_i \tag{3.2.2.9}$$

where λ_i and $\vec{v}_i$ are the i_{th} eigenvalue and the corresponding eigenvector of matrix A, and C_i are constants determined by the initial condition of the system $\vec{\rho}(0)$ and the system will reach an epidemic if at least one of the eigenvalues is positive, which can be achieved by ensuring that the maximum eigenvalue of the matrix A is positive[113]. Since A_τ is time dependent only through the factor $e^{-\gamma\tau}$, we can take advantage of the above general solution by replacing $e^{-\gamma\tau}$ with a constant ξ_t, say $\xi_t \approx (1 + e^{-\gamma t})/2$, to approximately capture its contribution. Denote by $A_\tau = \frac{\beta\xi_t}{\langle k \rangle} B - \gamma I$ for $\tau \in [0, t]$, where $B = \big[i(j-1)P(j)\big] \in \mathbf{R^{(N-1)\times(N-1)}}$ and I is an identity matrix of size $N-1$. It turns out that B has a unique eigenvalue $\lambda_B = \langle k^2 \rangle - \langle k \rangle$ and the corresponding eigenvector is $\vec{v}_B = (1, 2, \cdots, N-1)^T$. This is because on the one hand, all the rows of B are linearly dependent, indicating a unique eigenvalue for B, while on the other hand, if plug in the above results, we can easily see $B\vec{v}_B = \lambda_B\vec{v}_B$. Correspondingly, for A we have $\lambda_A(t) = \frac{\beta\xi_t}{\langle k \rangle}\big(\langle k^2 \rangle - \langle k \rangle\big) - \gamma$ and $\vec{v}_A = \vec{v}_B = (1, 2, \cdots, N-1)^T$. Plugging into equation 3.2.2.9, we get $\vec{\rho}(t) = Ce^{\lambda_A(t)t}\vec{v}_A$ if the initial condition $\rho_k(0) = Ck =$

$\rho(0)k/\langle k\rangle$ is satisfied, as shown bellow by setting $t = 0$,

$$\rho(0)N = \sum_{k=1}^{N-1} \rho_k(0)N_k = C\sum_{k=1}^{N-1} kN_k \Rightarrow C = \frac{\rho(0)}{\langle k\rangle}$$

Therefore, the solution for the whole system in early stages evolves according to,

$$\rho(t) = \sum_{k=1}^{N-1} P(k)\rho_k(t) \approx \rho(0)exp\big[\big(\frac{\beta\xi_t}{\langle k\rangle}\big(\langle k^2\rangle - \langle k\rangle\big) - \gamma\big)t\big] \tag{3.2.2.10}$$

It turns out we can remove the uncertainty in ξ_t in the following way. First, we notice with proper initial conditions, $\rho_k(t) \approx \rho_k(0)e^{\lambda_A(t)t} = \rho(0)k/\langle k\rangle e^{\lambda_A(t)t}$ and $\rho(t) \approx \rho(0)e^{\lambda_A(t)t}$. Then together with equation 3.2.2.7 we obtain,

$$\begin{aligned}
\dot{\rho}(t) &= \sum_{k=1}^{N-1} \dot{\rho}_k(t)P(k) \\
&= \sum_{k=1}^{N-1} \Big\{\frac{\beta e^{-\gamma t}}{\langle k\rangle}\sum_{k'=1}^{N-1} k(k'-1)\rho(0)k'/\langle k\rangle e^{\lambda_A(t)t}P(k') - \gamma\rho_k(t)\Big\}P(k) \\
&= \frac{\beta e^{-\gamma t}}{\langle k\rangle}\sum_{k=1}^{N-1} P(k)k\sum_{k'=1}^{N-1}(k'-1)\rho(0)k'/\langle k\rangle e^{\lambda_A(t)t}P(k') - \gamma\rho(t) \\
&= \frac{\beta e^{-\gamma t}}{\langle k\rangle}\langle k\rangle\rho(t)\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle} - \gamma\rho(t) \qquad (3.2.2.11) \\
\Rightarrow \rho(t) &= \rho(0)exp\Big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}(1 - e^{-\gamma t}) - \gamma t\Big] \qquad (3.2.2.12)
\end{aligned}$$

So, as long as we treat A_τ as a constant matrix for $\tau \in [0, t]$ by replacing the factor $e^{-\gamma\tau}$ with a constant ξ_t, and if we are given the initial condition $\rho_k(0) = \rho(0)k/\langle k\rangle$, then at the early stage, $\rho(t)$ is approximately given by equation 3.2.2.12, independent of how ξ_t is chosen. Similar to the homogeneous MF, we find for $\rho(t)$ the *phase transition time point* $\tau_\rho = \frac{1}{\gamma}log\big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\big]$ and *epidemic threshold* $(\beta/\gamma)_\rho > \frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}$. Similarly, for $q(t)$ the *phase transition time point* is $\tau_q = \frac{1}{\gamma}log\big[\frac{\beta}{2\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\big]$ and *epidemic*

threshold is $(\beta/\gamma)_q > \frac{2\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}$ since

$$\dot{q}(t) = \dot{\rho}(t) + \gamma\rho(t) = \frac{\beta(\langle k^2\rangle - \langle k\rangle)}{\langle k\rangle} e^{-\gamma t}\rho(t)$$
$$\Rightarrow \ddot{q}(t) = \frac{\beta(\langle k^2\rangle - \langle k\rangle)}{\langle k\rangle} e^{-\gamma t}\rho(t)\Big[\beta e^{-\gamma t}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle} - 2\gamma\Big] \tag{3.2.2.13}$$

For comparison purposes, the *phase transition time point* τ_ρ^* based on equation 3.2.2.10 and assuming $\xi_t = (1 + e^{-\gamma t})/2$ is given by

$$\frac{\beta\xi_{\tau_\rho^*}}{\langle k\rangle}\big(\langle k^2\rangle - \langle k\rangle\big) - \gamma = 0 \Rightarrow \tau_\rho^* = \frac{1}{\gamma} log\Big[\frac{\beta\big(\langle k^2\rangle - \langle k\rangle\big)}{2\gamma\langle k\rangle - \beta\big(\langle k^2\rangle - \langle k\rangle\big)}\Big] \tag{3.2.2.14}$$

Following the homogeneous MF, to incorporate $\delta_k(t)$ we first set $\delta_k(t) = \delta_k(0) = 0$, then from equation 3.2.2.12 we have $\rho_k(t) = \rho_k(0)e^{(\beta\langle k^2\rangle/\langle k\rangle - \gamma)t} \Rightarrow \delta_k(t) = 1 - \alpha_t\rho_k(0)/\rho_k(t) = 1 - e^{-\beta\langle k^2\rangle t/\langle k\rangle}$, and plugging into equation 3.2.2.11 we get

$$\dot{\rho}(t) = \beta e^{-\gamma t}\frac{\langle k^2\rangle - (1 - e^{-\beta\langle k^2\rangle t/\langle k\rangle})\langle k\rangle}{\langle k\rangle}\rho(t) - \gamma\rho(t)$$
$$\rho(t) = \rho(0)exp\Big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\big(1 - e^{-\gamma t}\big) - \gamma t + \frac{\beta}{\gamma + \beta\langle k^2\rangle/\langle k\rangle}\big(1 - e^{-(\gamma + \beta\langle k^2\rangle t/\langle k\rangle)}\big)\Big] \tag{3.2.2.15}$$

Similarly, the epidemic threshold for both ρ and q should not change, but the phase transition time points τ'_ρ for ρ and τ'_q for q are the solutions of the following equations,

$$\frac{\beta e^{-\gamma\tau'_\rho}}{\langle k\rangle}\Big[\langle k^2\rangle - (1 - e^{-\beta\langle k^2\rangle\tau'_\rho/\langle k\rangle})\langle k\rangle\Big] - \gamma = 0$$
$$\frac{\beta e^{-\gamma\tau'_q}}{\langle k\rangle}\Big[\langle k^2\rangle - (1 - e^{-\beta\langle k^2\rangle\tau'_q/\langle k\rangle})\langle k\rangle\Big] - 2\gamma = 0 \tag{3.2.2.16}$$

Performance Evaluation of the Heterogeneous MF Approximation: Next, we compare our theoretical results with those obtained from MC simulation. The networks are generated as having scale free topologies of size $N = 10^4$, while their node degree dis-

tribution follows a power law distribution $P(k) \sim k^{-\lambda}, k \in \{K_{min}, \cdots, N-1\}$ with $\lambda = 3.5$. Further, we control the network heterogeneity, defined as $\langle k^2 \rangle / \langle k \rangle$, by controlling the minimum degree K_{min}[9, 8]. In this case, $\langle k \rangle \approx 5K_{min}/3$, $\langle k^2 \rangle \approx 5K_{min}^2$ and therefore $\langle k^2 \rangle / \langle k \rangle \approx 3K_{min}$. Results in each of the settings are averaged over 10^3 realizations. First, we compare the saturated diffusion ratio from MC simulations, with the numerical solution of equations 3.2.2.5 and 3.2.2.6 and following the heterogeneous MF approximation, we choose appropriate β and γ so that $q(\infty) \to 0.5$. The results are shown in Figure 3.4. For comparison purposes, we include the results for both when $\rho_k(0) = \rho(0)k/\langle k \rangle$ and $\rho_k(0) = \rho(0)$. Similar to the homogeneous MF, we observe better agreement for both cases as K_{min} increases, corresponding to higher average excess degree.

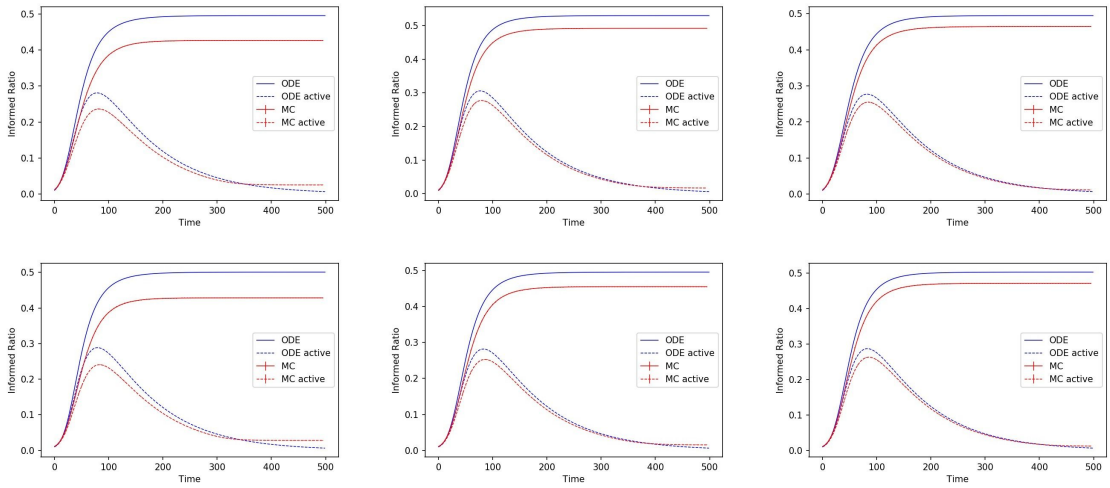


Figure 3.4: Left to right upper row: (a-c), $K_{min} = 5, 10, 15$; Left to right lower row: (d-f), $K_{min} = 5, 10, 15$. Time dependent diffusion ratio for both $\rho(t)$ and $q(t)$ based on MC simulation and numerical solution of ODE equations for networks of different densities. For comparison, we include the results for both when $\rho_k(0) = \rho(0)k/\langle k \rangle$ in (a-c) and when $\rho_k(0) = \rho(0)$ in (d-f) with $\rho(0) = 1\%$. *ODE* represents the numerical solution of equations 3.2.2.5 and 3.2.2.6

Next, we compare the analytical solutions at early stages ($t \to 0$) for the case $\rho_k(0) = \rho(0)k/\langle k \rangle$ with those obtained from MC simulations as shown in Figure 3.5.

We can see equations 3.2.2.10, 3.2.2.12 and 3.2.2.15 are very close to each other and good agreement with MC simulation is achieved.

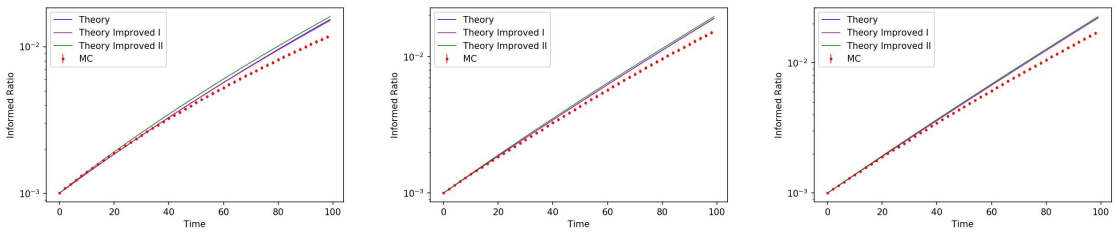


Figure 3.5: Left to right: (a-c), $K_{min} = 5, 10, 15$. Early stage evolution of the diffusion ratio and comparisons between the MC simulation and theoretical results for networks of different densities with $\rho(0) = 0.1\%$. *Theory* stands for equation 3.2.2.10, *Theory Improved I* for equation 3.2.2.12 and *Theory Improved II* for equation 3.2.2.15

Finally, we compare the phase transition time points, with results shown in Figure 3.6. We notice that the phase transition time points obtained based on equation 3.2.2.14 is so inaccurate such that $\tau_\rho^* \to \infty$ when $\beta/\gamma \geq 0.3$ while the other two agree well with the MC results.

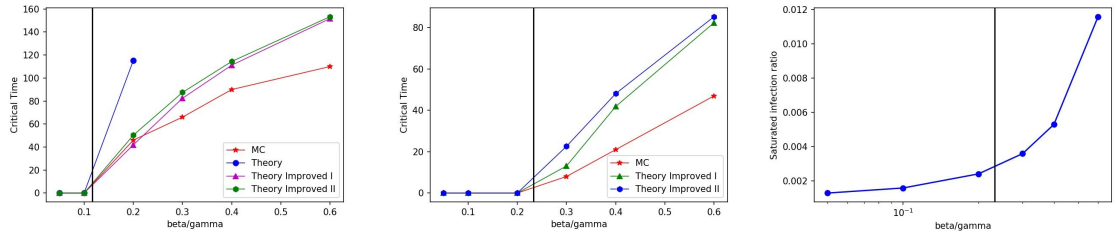


Figure 3.6: Left to right: (a), τ for $\rho(t)$; (b), τ for $q(t)$; (c), Total diffusion ratio $q(\infty)$. Phase transition time points for both $\rho(t)$ and $q(t)$ based on network parameter $K_{min} = 3$ and initial ratio of seed nodes $\rho(0) = 0.1\%$. Also included is the total diffusion ratio $\rho(\infty)$ for different values of β/γ. *Theory* stands for τ_ρ^* in equation 3.2.2.14, *Theory Improved I* stands for $\tau_\rho = \frac{1}{\gamma} log\left[\frac{\beta}{\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\right]$ and $\tau_q = \frac{1}{\gamma} log\left[\frac{\beta}{2\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\right]$, and *Theory Improved II* stands for equation 3.2.2.16. For visual guidance, we include black vertical lines at location $\beta/\gamma = \langle k\rangle/\langle k^2\rangle$ for $\rho(t)$ in (a) and $\beta/\gamma = 2\langle k\rangle/\langle k^2\rangle$ for $q(t)$ in (b-c).

Remark: In the above derivation process, we made the assumption that we can ignore all the terms of order $O(\rho^2(t))$ at early stages. However, this may not hold for some terms under our initial conditions, since for $k \to N$ we have $\rho_k(t) \sim \rho(0)k/\langle k \rangle \sim O(1)$, which seems to make our analytical solution 3.2.2.10 invalid. Nevertheless, we can show that this is not necessarily the case for some family of networks; a prominent example is scale free networks, where the degree distribution follows $P(k) = Ck^{-\lambda}$ for $k \in \{K_{min}, \cdots, N-1\}$. For simplicity, in the following we assume that $I(0)$ does not scale up with N and therefore, $\rho(0) \sim O(1/N)$.

Firstly, if $\lambda = 2 - \epsilon, \epsilon \in (0,1)$, then $C \approx (1-\epsilon)K_{min}^{1-\epsilon}$ and $\langle k \rangle \sim O(N^{\epsilon})$. Under this scenario, as $k \to N$, $\rho_k(t) \sim \rho(0)k/\langle k \rangle \sim O(\frac{1}{N}N\frac{1}{N^{\epsilon}}) \sim O(N^{-\epsilon})$ and it is safe to ignore $O(\rho^2(t))$ terms. On the other hand, if $\lambda = 2 + \epsilon, \epsilon > 0$, then $C \approx (1+\epsilon)K_{min}^{1+\epsilon}$ and $\langle k \rangle \approx \frac{1+\epsilon}{\epsilon}K_{min}$. In this case, for $k \sim O(N)$, $\rho_k(t) = \rho(0)k/\langle k \rangle \sim O(\frac{1}{N}N) \sim O(1)$, and it is not safe to ignore all $O(\rho_k^2(t))$ terms. The total relative error contributed to $\Theta(t)$ in equation 3.2.2.3 from $O(N)$ such terms is of order $N[k - \delta_k(t)]\rho_k^{err}(t)P(k) \sim O(NN\frac{1}{N^{2+\epsilon}}) \sim O(\frac{1}{N^{\epsilon}})$, and therefore the estimation for $\rho_k(t)$ for $k \sim O(1)$ is accurate up to order $O(\frac{1}{N^{\epsilon}})$. Therefore the total relative error contributed to $\rho(t)$ is $\sim NP(k)\rho_k^{err}(t) \sim N\frac{1}{N^{2+\epsilon}} \sim O(\frac{1}{N^{1+\epsilon}})$ for $k \sim O(N)$ plus $P(k)\rho_k^{err}(t) \sim O(\frac{1}{N^{\epsilon}})$ for $k \sim O(1)$ and of order $O(\frac{1}{N^{\epsilon}})$, making equation 3.2.2.10 valid.

3.3 Maximum Weight Tree (MWT) Approximation

3.3.1 Path dependent diffusion edge probabilities

In section 3.2, we considered the SIM model from a macroscopic perspective, when a proportion $\rho(0)$ of randomly selected nodes were set in the I state initially. Next, we investigate the problem from a microscopic perspective by fixing the network structure and explicitly compute the probability of each node getting informed depending on its actual path to the single seed node where the information diffusion originates. This

perspective is also instructive of why percolation theory is not applicable to SIM. For illustration purposes we include the results for both the discretized time step case useful for comparisons with computer simulations and for the continuous time step case. Further, it will be shown that the two cases converge to each other in the limit for small infection rate β and immunization rate γ .

For the *discretized time step* case we can simply assume $\Delta t = 1$ as in a computer simulation setting. The evolution of the system is as follows. At time step $t = 0$, first every neighbor j of the seed node i gets informed with probability $\beta_{ij} = 1 - (1 - \beta)^{W_{ij}}$, where W_{ij} is the weight of connection, and then all nodes in the network get immunized with a probability γ. The process then repeats itself on the residual network in the following time steps until all the nodes are either informed or immunized. The flow of information can be represented as in figure 3.7: At $t = 0$, node 0 is in state I and the remaining ones are in state S, with the diffusion probability per unit time between nodes i and j being β_j for ease of presentation.

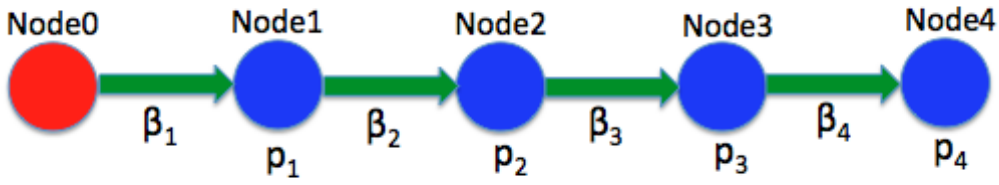


Figure 3.7: Information diffusion process starting from seed node 0.

The probability of node 1 to get informed is given by,

$$p_1^{(d)} = \sum_{t=0}^{\infty} (1-\gamma)^{2t}(1-\beta_1)^t\beta_1 = \frac{\beta_1}{1-(1-\gamma)^2(1-\beta_1)} \tag{3.3.1}$$

Similarly, node 2 gets informed in two steps. First node 1 gets informed from node

0, while at the same time none of the three nodes 0, 1 or 2 get immunized. Then, node 2 gets informed from node 1 in the same way as node 1 got informed from node 0 and the probability is given by

$$\begin{aligned} p_2^{(d)} &= \sum\nolimits_{t_1=0}^{\infty}(1-\gamma)^{3t_1}(1-\beta_1)^{t_1}\beta_1 \sum\nolimits_{t_2=0}^{\infty}(1-\gamma)^{2(t_2+1)}(1-\beta_2)^{t_2}\beta_2 \\ &= \frac{\beta_1}{1-(1-\gamma)^3(1-\beta_1)}\frac{\beta_2(1-\gamma)^2}{1-(1-\gamma)^2(1-\beta_2)} \end{aligned} \tag{3.3.2}$$

The extra one unit of time in the exponent of $1-\gamma$ for t_2 is because after node 1 gets informed, both nodes 1 and 2 need to survive one time step of the immunization process before they start the diffusion process.

Following the same procedure, the probability for node $k \geq 2$ at an arbitrary stage to get informed is:

$$\begin{aligned} p_k^{(d)} &= \sum\nolimits_{t_1=0}^{\infty}(1-\gamma)^{(k+1)t_1}(1-\beta_1)^{t_1}\beta_1 \cdots \sum\nolimits_{t_k=0}^{\infty}(1-\gamma)^{2(t_k+1)}(1-\beta_k)^{t_k}\beta_k \\ &= \frac{\beta_1}{1-(1-\gamma)^{k+1}(1-\beta_1)}\prod_{i=2}^{k}\frac{\beta_i(1-\gamma)^{k+2-i}}{1-(1-\gamma)^{k+2-i}(1-\beta_i)} \end{aligned} \tag{3.3.3}$$

In the above, we have assumed that the information diffusion process takes place before the immunization process, and if the case is the other way around, we simply need to make the system survive one step of the immunization process at the beginning of time $t = 0$ while the remainder of the evolution mechanism would be exactly the same as above; Therefore, in this case, the probability of getting informed for a node at an arbitrary stage k is given by

$$p_k^{(m)} = (1-\gamma)^{k+1}p_k^{(d)} = \prod_{i=1}^{k}\frac{\beta_i(1-\gamma_i)^{k+2-i}}{1-(1-\gamma_i)^{k+2-i}(1-\beta_i)} \tag{3.3.4}$$

Further, no special treatment for $k = 1$ is needed.

For the *continuous time step* case, in the limit $dt \to 0$, the probability of diffusion

in each time step is βdt and the survival probability of the immunization process for a time length of t is $(1-\gamma d_t)^{t/dt} = exp\big[\frac{t}{dt} log(1-\gamma dt)\big] \overset{dt\to 0}{=} e^{-\gamma t}$. Therefore, the probability for a node at an arbitrary stage to get informed is,

$$\begin{aligned} p_k &= \int_0^\infty e^{-(k+1)\gamma t_1} e^{-\beta_1 t_1} \beta_1 dt_1 \cdots \int_0^\infty e^{-(k+2-i)\gamma t_i} e^{-\beta_i t_i} \beta_i dt_i \cdots \int_0^\infty e^{-2\gamma t_k} e^{-\beta_k t_k} \beta_k dt_k \\ &= \prod_{i=1}^{k} \frac{\beta_i}{(k+2-i)\gamma + \beta_i} \end{aligned} \tag{3.3.5}$$

We can easily see that both equations 3.3.3 and 3.3.4 converge to equation 3.3.5 in the limit as $\beta \to 0$ and $\gamma \to 0$, which is expected since this is equivalent to the limit of continuous time step $dt \to 0$.

Next, we discuss how the above results can be extended to the multi-edge case. Suppose we have multiple edges between node i and j, which occurs in the case of coupled multilayer networks when considering them as a single layer with connections from different layers constituting multi-edges. Then the equivalent diffusion rate is given by

$$\begin{aligned} \beta_{ij} &= 1 - \textstyle\prod_{k=1}^{K} (1-\beta_k) = 1 - \prod_{k=1}^{K} (1-\beta_0)^{W_{ij}^{(k)}} \\ &= 1 - (1-\beta_0)^{\sum_{k=1}^{K} W_{ij}^{(k)}} \end{aligned} \tag{3.3.6}$$

where $W_{ij}^{(k)}$ is the weight of the k_{th} edge between node i and j.

For comparison purposes, we apply our analysis to the SIR model in *continuous time step* case, and the probability of a node in arbitrary stage k to get infected is,

$$p_k = \int_0^\infty e^{-\beta t_1} \beta e^{-\gamma t_1} dt_1 \cdots \int_0^\infty e^{-\beta t_k} \beta e^{-\gamma t_k} dt_k = \big(\frac{\beta}{\beta+\gamma}\big)^k \tag{3.3.7}$$

We can see that $p_{k+1}/p_k = \frac{\beta}{\beta+\gamma}$, which is the transmissibility of the edge connecting nodes of distance k and $k+1$ to the seed node, and is a constant, independent of

stage k. On the contrary, for the SIM model if we assume constant infection rate β, from equation 3.3.5, we get

$$\frac{p_{k+1}}{p_k} = \frac{\beta}{(k+2)\gamma + \beta} \tag{3.3.8}$$

which decreases with k and hence does not satisfy this key assumption in percolation theory.

3.3.2 MWT construction

In the following, we develop an alternative way of approximating the saturated ratio of information diffusion on a static network assuming that we start from a single seed node. The first assumption is reasonable since real world networks are not totally random with only certain parameters fixed as when we control λ in generating homogeneous random networks with degree following a Poisson distribution, or control λ and K_{min} in generating scale free networks. On the contrary, within a reasonably long period of time, it is reasonable to assume a fixed network topology in many applications, inlcuding Protein networks, Twitter networks and Facebook networks[81, 13, 3, 28]. The second assumption is also reasonable, namely that there exists a single seed node, for example, in marketing and advertising there is usually a highly influential node [13, 3]. We approximate the total saturated diffusion ratio by growing a MWT starting from seed node s to all the remaining nodes in the network. The diffusion process on the network is estimated by the tree and the weight for each node approximates the probability that the node will ever get informed. The accuracy of the approximation improves for sparser networks since in this case the probability of multi-channels in the diffusion would be lower and the tree based approximation for the diffusion process is satisfactory. To achieve the task of growing a MWT, we slightly modify the Dijkstra's algorithm in finding the shortest distance from seed

node s to the remaining nodes in the network. In each step, instead of identifying the node with the minimum distance to the seed node in the queue, we identify the node with the maximum weight. The detailed algorithm is presented in pseudo code Algorithm 1.

Algorithm 1 Find the Maximum Weight Tree over the network G starting from seed node s

Input: Network $G(V, E)$, seed node s, diffusion rate β, immunization rate γ and formula $F(B, \gamma)$ to calculate the diffusion probability/weight of any node v from the list B of β depending on its optimal path to the seed node s so as to achieve the highest weight.

```
for v ∈ V do
    weight[v] ← 0
    Ancestors[v] ← []
    B[v] ← []
end for
weight[s] ← 1
Q ← V
while Q ≠ ∅ do
    u ← ArgMax(Q, weight)
    Q ← Q − u
    for v ∈ neighbors[u] do
        B_temp ← B[u].append(β_uv)
        if F(B_temp, γ) > weight[v] then
            weight[v] ← F(B_temp, γ)
            Ancestors[v] ← Ancestors[u].append(u)
            B[v] ← B_temp
        end if
    end for
end while
return weight
```

Numerical Illustration of the MWT algorithm: To showcase the effectiveness of the MWT algorithm in estimating the diffusion ratio, we fix the immunization rate $\gamma = 0.01$ and carry out experiments with different diffusion rate β on synthetic scale free networks with degree distribution following $P(k) \sim k^{-\lambda}$. The network topology is controlled by changing the parameter λ. The network size $N = 10^4$ and the ratio of MWT over MC simulation results for each setting is obtained by averaging over 9000

realizations: namely, 30 different random network topologies, with 300 realizations of the diffusion process for each one. For each realization, we simply set the hub node with the highest degree as the seed node. The results are shown in Figure 3.8 and as expected for sparser networks (higher λ parameter) the agreement with the simulation results becomes markedly better.

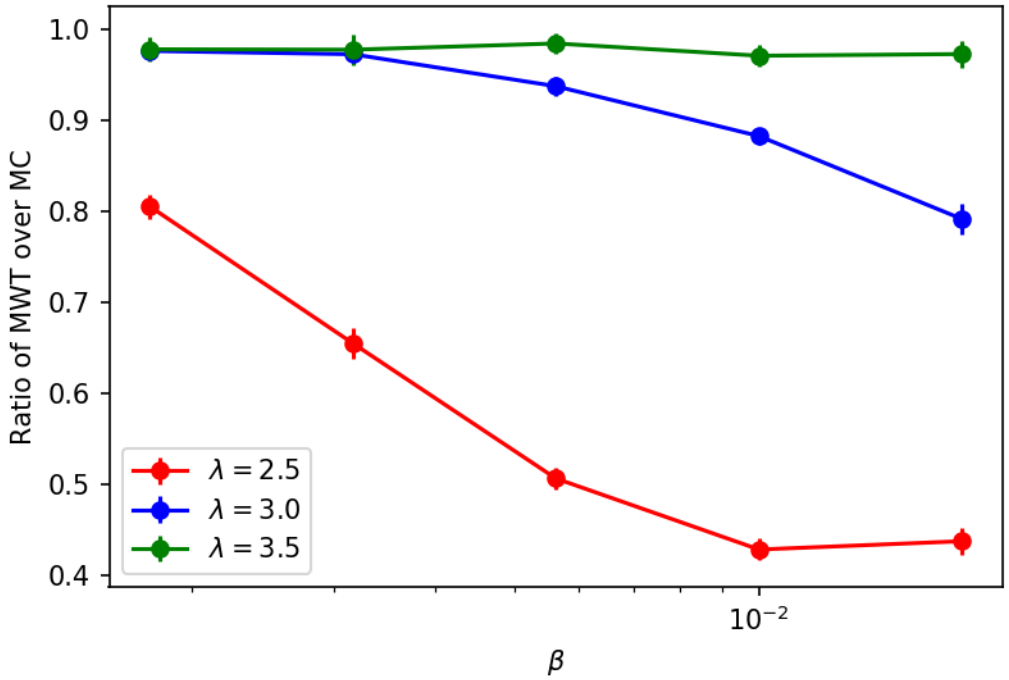


Figure 3.8: Ratio between diffusion ratio of MWT and MC simulations for different β with a fixed $\gamma = 0.01$ on networks of different densities realized by controlling the parameter λ.

CHAPTER IV

Exponentially time decaying Susceptible-Informed (SIT) model for information diffusion process on networks

4.1 Introduction

4.1.1 Epidemic models for information diffusion

A rich literature on information diffusion models has emerged, due to the timeliness of the topic in the era of massive social media platforms and their use for understanding news propagation, rumor spreading, product advertisement by word of mouth and technological innovation adoption [61, 49, 71, 29, 132, 50]. A popular class of models regards the information diffusion process akin to an epidemic and assumes that successive trials of infection/diffusion between two nodes are independent with the same probability p of success [33, 105, 129, 123, 86, 65]. One popular family of theoretical approaches for general purposes in studying these models include the mean field (MF) theory [9, 113, 129, 47, 12, 16], due to its simplicity and accuracy. The simplest one of the epidemic models is the Susceptible-Infected (SI) model, where the nodes in the network are grouped into two compartments, i.e., the infected (I), which are infected with the disease and can infect the nearest neighbors

through links, and the susceptible (S), which are susceptible and can be infected by an infected nearest neighbor. In the context of homogeneous mean field (MF) theory, where it is assumed that all nodes are statistically equivalent with the same degree of connectivity $\langle k \rangle$, and the population is fully mixed, if we denote by N as the total number of nodes in the network, and by $S(t)$ and $I(t)$ as the total number of nodes in the state S and I respectively at time t, and set $s(t) = S(t)/N$ and $\rho(t) = I(t)/N$, then,

$$\dot{\rho}(t) = \beta \langle k \rangle \rho(t)[1 - \rho(t)] \tag{4.1.1}$$

approximates the evolution process, were $\dot{\rho}(t) = d\rho(t)/dt$, and β is the infection rate. Obviously, in this simple model, the whole population will get infected at the end, which is not a good description of the disease transmission process in the real life. To overcome this shortcoming, another two popular models are proposed, namely, the susceptible-infected-susceptible (SIS) model and the susceptible-infected-recovered (SIR) model. In SIS model, instead of staying in the I state permanently once infected as in the SI model, the infected nodes can recover to the state S with a rate γ, and utilizing the same notations as above, we have for homogeneous MF approximation,

$$\dot{\rho}(t) = \beta \langle k \rangle \rho(t)[1 - \rho(t)] - \gamma \rho(t). \tag{4.1.2}$$

In this context, the disease will persist to a non-zero proportion of the population in equilibrium under the condition that the factor β/γ is larger than the *epidemic threshold*, $1/\langle k \rangle$, which can be obtained easily by setting $\dot{\rho}(t \to \infty) = 0$. A more general framework that can capture the degree distribution of the network is the heterogeneous MF approach, and if applied to the SIS model, the epidemic threshold is $\langle k \rangle / \langle k^2 \rangle$. Another popular epidemic model is the SIR model, where instead of

recovering back to state S and become susceptible again, the nodes in state I will recover with rate γ to an permanently removed/immunized state R, an additional compartment. Similarly, if we denote $R(t)$ as the number of nodes in the state R at time t, and set $r(t) = R(t)/N$, the evolution of the disease in this model under the context of homogeneous MF theory can be approximated as,

$$\begin{aligned} \dot{\rho}(t) &= \beta\langle k\rangle\rho(t)[1-\rho(t)] - \gamma\rho(t) \\ \dot{r}(t) &= \gamma\rho(t). \end{aligned} \tag{4.1.3}$$

In order to ensure that the infection size ($R(t \to \infty)$) scales up with the network size N, that is, a finite non-zero proportion of the population gets infected, similarly, the factor β/γ should be greater than the epidemic threshold $1/\langle k\rangle$. In the case of heterogeneous MF approach for a general network with arbitrary degree distribution, the corresponding epidemic threshold is $\langle k\rangle/(\langle k^2\rangle - \langle k\rangle)$. For further information about the critical behaviors, we refer to the review papers [33, 105]

To study the SIS and SIR models analytically, many other theoretical approaches have been employed other than the above mentioned homogeneous and heterogeneous MF approximations. For example, instead of only incorporating the degree distribution nodes in the network, the quenched mean field (QMF) approach captures the whole topology of the network by incorporating the adjacency matrix of the network directly in the ordinary differential equations (ODEs) governing the evolution of the disease [82]. To address the "echo chamber" issue suffered by the QMF approach, while retaining its strength of capturing the network topology, the dynamic message passing (DMP) approach was proposed by introducing the "cavity" state, such that the nodes in this state can be infected but cannot transmit the disease [57]. This approach is exact for tree-like network, but does not capture the dynamic correlations between neighbors. And this issue can be addressed in in the pairwise

approximation (PA) approach, which instead of studying the evolution of individual nodes, the disease transmission process is studied by considering the pair nodes states [35]. Specifically for the SIR model, due to the irreversibility of the evolution from the state I to state R, several other approaches can be employed. For example, since the probability of a susceptible node getting infected from an infectious neighbor is a constant, the SIR model can be mapped to bond percolation theory in statistical physics by treating the infection probability as the transmissibility of each edge [93, 95, 58]. Other approaches include the message passing (MP) approach and the Edge-based compartmental (EBC) approach that take into account issues such as degree-degree correlation and dynamical correlation between neighbors [60, 85]. For a more thorough review on this topic, we refer to [128].

Other than the standard SI, SIS and SIR models, a lot of efforts have been made in the literature in proposing improvement based on the three models in capturing various aspects of the disease transmission. For example, in [25], a modified SIS model was proposed where the infection rate of each infected node decreases exponentially with the number of times this node has ever been infected, with a maximum number L of such decays. At equilibrium, this effectively results in a standard SIS model with infection rate $\beta_0\epsilon^{-L}$, where β_0 is the initial infection rate and ϵ is the decreasing factor. A power law relationship between the $s(t \to \infty)$ and ϵ at equilibrium was obtained, which agrees well with the MC simulation results both on a fully mixed population and on d-dimensional lattices when $d > 6$, which effectively mimics a densely connected network, consistent with the well known knowlege in the literature that MF approximation improves with the density of networks. As for the SIR model, for example, in [75], to achieve more flexibility, new processes such as rebirth, death, pulse vaccination, and pulse treatment of nodes are included, and parameters such as rebirth rate, death rate, infection rate, and recovery rate are allowed to be time dependent, and the general nonlinear incidence rates are assumed for the infection

process. On the other hand, in [48], the SIR model is modified to include both birth and death processes and the recovery rate can be time dependent such that the time spent on state I follows Pearson distribution or state dependent such that the survival probability follows a power law distribution. Other examples include modifications made specifically for the Ebola Virus Disease (EVD) diffusion in west Africa. For example, in [107], a vaccination process was included in the standard SIR model such that the susceptible nodes have a certain rate of getting vaccinated and an optimal control strategy was obtained by minimizing an objective function consisting of the number infected individuals and the cost of vaccination. While in [15], a migration network based on the effective distances from a radiation migration model between administrative divisions to capture the topological characteristic of the EVD diffusion, and a Susceptible-Decreasingly Infections-Recovered (SDIR) model was proposed by letting the infection rate decrease exponentially with time, which can represent the public health intervention.

4.1.2 "Intrinsic time value" of information

However, on many occasions, especially on social media platforms, the process of information diffusion is quite different than that of an epidemic. An example comes from news propagation and/or rumor spreading, where the *reluctance to tell stale news* would slow down the diffusion process over time [27]. A related example is online news dissemination, where it is observed that in any *news cycle* the attention of actors is focused mainly on more recent and dominant topics [72]. Intuitively, a piece of information/news from further into the past tends to be mentioned less often at present, which is not the case for the outbreak of a disease, where recently infected actors have the same potential of infecting new ones, as those infected earlier on. Finally, in viral marketing and hashtag adoption on Twitter, it is observed that the marginal probability of adoption (new products, hashtags) would typically

decrease with the number of recommendations/exposures [71, 110]. If we treat this as a temporal event, we will notice that it coincides with the news spreading process in that as time goes by the diffusion process slows down. Therefore, on many instances the *information* often possesses *intrinsic time value*, which in turn renders the basic assumption of the SIR and related epidemic models rather inapplicable.

In order to address this issue, attempts have been made in the literature. For example, in [138], a special SIR model is studied for rumor spreading where it is assumed that a node in state I will randomly select one of its neighbors at each time step and transmit the information if the neighbor is in state I and lose interest or recover if the neighbor is in state I or R. This essentially captures the "losing interest effect" of the nodes in state I by letting the recovery process accelerates with the sizes $I(t)$ and $R(t)$. Analytical solutions for the final diffusion size based on both the homogeneous and heterogeneous MF are obtained and agree well with with MC simulation results, and the 80% final infection size in the literature for the homogenous MF case can be recovered as a special case of large $\langle k \rangle$, the average degree of connectivity of the network. Besides, models such as the Daley-Kendall (DK) and the Maki-Thompson ones were proposed [27, 119, 91, 78], the former becoming rather popular and being extensively studied and used. In the DK model, similar to the three states classification in SIR, the nodes are also classified into three classes: ignorants ($s(t)$), spreaders ($i(t)$) and stiflers ($r(t)$) corresponding to S, I and R states in the SIR model and following the convention in SIR we use s, i, r to represent the three states. Further, the removal/recovery rate of $i(t)$ in the DK model is proportional to $i(t)\big[i(t)+r(t)\big]$, with the effect of the *losing news value* when reaching a actor already aware of the information ($i(t)$ or $r(t)$) included from the perspective of spreaders. One limitation of the DK model is that it does not capture the role played by the "ignorant" actors. For example, in viral marketing and hashtag adoption on Twitter, it is the "ignorant" players who lose interest in adopting the products/hashtags.

On the other hand, in news/rumor spreading, the *reluctance to tell stale news* is co-existent with the *reluctance to listen to stale news*, i.e., the "ignorant" players are not interested in talking about topics involving old news. With a closer look at the *intrinsic time value* property of news/information, we can address this issue by developing naturally an alternative mechanism coined Susceptible-Informed model with exponentially time decaying diffusion rate, abbreviated as SIT, with T capturing the time decaying flavor.

In the SIT model, just as in the SI model, we assume the nodes in the network can be classified into two groups: susceptible, denoted by S, and informed/infected, denoted by I. There are two parameters in this model, initial diffusion rate β and diffusion decay rate γ, such that the diffusion rate at time t becomes $\beta_t = \beta e^{-\gamma t}$, which is time dependent. Within a unit time step at time t, all nodes in state I will diffuse the information to each of their neighbors in sate S with probability β_t, and the system will reach an asymptotic state of *equilibrium* either when all nodes in the network are informed or when β_t becomes negligibly small. In the context of standard homogeneous MF approach, following the treatment in equation 4.1.1 the evolution of the information diffusion process can be approximated as

$$\dot{\rho}(t) = \beta e^{-\gamma t} \langle k \rangle \rho(t)[1 - \rho(t)]. \tag{4.1.1}$$

Simple as it is, this model captures the information's characteristic of *losing value* with time from both the spreaders and "ignorant" actors' perspectives through an exponentially decreasing diffusion rate with time. To be specific, a node informed at an early time $t = \tau_1$ exhibits a better capacity of diffusing further the information to the rest of the network compared to getting informed at a later time $t = \tau_2 > \tau_1$ because the diffusion rate has decreased by a factor of $e^{-\gamma(\tau_2 - \tau_1)}$.

The analysis of the SIT model exhibits a number of technical challenges. Because

the expected transmissibility of an edge, i.e., the total probability of diffusing the information from a node in state I to another in state S through the edge, is dependent on its path to the seed node (the node in state I at the beginning of time), as will be shown in Section 4.3, the standard arguments from percolation theory are no longer applicable, and thus other techniques are required. In this work, we employ two theoretical approaches for the analysis. The first one is the MF approximation based on collective macro effects, where a factor δ_t is carefully and accurately handled, capturing the fact that at least one neighbor of any newly informed node v is in the state I transmitting to v the information in the first place, and significantly improved the accuracy of the MF approximation in terms of the agreement to the MC simulation results. To the best of our knowledge δ_t has always been set to 0 or 1 in the literature and this is the first time it is accurately incorporated. This analysis can be easily transferred to to the analysis of other epidemic models such the SIR and SIS models. Then we propose a novel approach in estimating the final diffusion size by constructing a Maximum Weight Tree (MWT) based approximation that leverages individual micro dynamics. Similar to the DMP approach, this novel approach will give an exact result for tree-like network, and are specifically designed for sparse networks, defined as the kind of network where the average degree $\langle k \rangle$ of the nodes in the network does not scale up with the network size N as N increases to a large number, or, equivalently, the total number of links in the network is in the order of the number of nodes N in the network.

The remainder of the paper is organized as follows: In Section 4.2 we analyzed the SIT model based on MF approximation. Specifically, for homogeneous case, we obtain a close form solution for the time dependent diffusion size in the whole time range. For heterogeneous case, on the other hand, we obtain a set of ordinary differential equations (ODE) based on which we can calculate time dependent diffusion size by numerically solving those ODEs. Besides, we obtain a close form solution for the

time dependent diffusion size at early stages of the evolution, based on which a *phase transition time* point, where the expected number of newly informed nodes in each unit time step reaches the maximum, and *epidemic threshold*, the critical value that β/γ needs to be greater than to ensure a positive acceleration in the information diffusion process at the beginning, can be derived. Two different initial conditions in terms of the seed nodes allocation are considered, and interestingly, we find that the *phase transition time* points and *epidemic threshold* of the two cases are identical, indicating the robustness of the results derived for the two in terms of initial conditions. To evaluate the performance of MF approximations, we compare the theoretical results derived with Monte Carlo (MC) simulation estimates and very good agreements are achieved. In Section 4.3, we first calculate explicitly the total probability of getting informed for each node depending on its actual path to the seed node, the single node initially in state I, then construct a MWT with the *weight* of each node corresponding to the probability. Very good agreement between the results of MWT and MC simulation can be achieved, especially for sparse networks, where the multi-channel diffusion is limited and tree approximation for the diffusion process is satisfactory. It should be noted that throughout the whole paper we assume $\gamma \neq 0$.

4.2 Mean Field Approximation

4.2.1 Homogeneous Mean Field Approximation

In this section we study the evolution of the system based on a homogeneous MF approximation, assuming all nodes are statistically equivalent of the same degree $\langle k \rangle$. An application of this scenario is for networks generated by some random graph models such as the ER model or the Gilbert models with an appropriate choice of the model parameters [37, 45]. At time t on a network of size N, if we denote by $S(t)$ the total number of susceptible nodes and by $I(t)$ the total number of informed/infected

nodes, then

$$dI(t) = S(t)[1 - (1 - \beta e^{-\gamma t} dt)^{\langle k \rangle \frac{I(t)(\langle k \rangle - \delta_t)}{N\langle k \rangle}}] \overset{dt \to 0}{=} S(t)\beta e^{-\gamma t} \frac{I(t)}{N}(\langle k \rangle - \delta_t)dt, \quad (4.2.1.1)$$

where δ_t takes into account the fact that at least one of the neighbors of each newly informed node i is in state I, from which node i received the information. Further, if we denote by $\rho(t) = I(t)/N$, $s(t) = S(t)/N$ and considering $\rho(t) + s(t) = 1$, we have:

$$d\rho(t) = [1 - \rho(t)]\beta e^{-\gamma t}\rho(t)(\langle k \rangle - \delta_t)dt. \quad (4.2.1.2)$$

By introducing this time dependent factor δ_t instead of directly setting $\delta_t = 1$ or 0 as is commonly done in the literatures [129, 12, 9], a much better agreement between the theoretical results derived based on MF and that of MC simulations can be achieved, as will be shown later. Denote by $I(0)$ the number of seed nodes, defined as the group of nodes informed at $t = 0$, then the factor $\delta_t = 1 - I(0)/I(t) = 1 - \rho(0)/\rho(t)$ can be obtained as follows

$$[I(t) - I(0)](\langle k \rangle - 1) + I(0)\langle k \rangle = I(t)(\langle k \rangle - 1) + I(0) = I(t)\big[\langle k \rangle - (1 - I(0)/I(t))\big]. \quad (4.2.1.3)$$

To obtain an analytical solution, we want to remove the dependence of δ_t on $\rho(t)$ and to derive an approximate formula, we first assume $\delta_t = 0$ which is true when $t = 0$. Then,

$$\begin{aligned} \frac{d\rho(t)}{\rho(t)[1 - \rho(t)]} = \beta e^{-\gamma t}\langle k \rangle dt \Rightarrow \log \frac{\rho(t)}{\rho(0)} - \log \frac{1 - \rho(t)}{1 - \rho(0)} = \frac{\beta\langle k \rangle}{\gamma}(1 - e^{-\gamma t}) \\ \Rightarrow \rho(t) = \big\{1 + \frac{1 - \rho(0)}{\rho(0)} \exp\big[-\frac{\beta\langle k \rangle}{\gamma}(1 - e^{-\gamma t})\big]\big\}^{-1} \\ \overset{t \to 0}{\approx} \big\{1 + (\frac{1}{\rho(0)} - 1)e^{-\beta\langle k \rangle t}\big\}^{-1} \approx \rho(0)e^{\beta\langle k \rangle t}, \end{aligned} \quad (4.2.1.4)$$

where we have used the approximation $1 - e^{-\gamma t} \overset{t\to 0}{\approx} 1 - (1 - \gamma t) = \gamma t$. Plugging in the above result, we get $\delta_t = 1 - \rho(0)/\rho(t) \approx 1 - e^{-\beta\langle k\rangle t}$, and plugging it back into Equation 4.2.1.2, gives

$$d\rho(t) = [1 - \rho(t)]\beta e^{-\gamma t}\rho(t)[\langle k\rangle - 1 + e^{-\beta\langle k\rangle t}]dt. \tag{4.2.1.5}$$

The later equation can be solved directly to get

$$\begin{aligned}
&\log\frac{\rho(t)}{\rho(0)} - \log\frac{1-\rho(t)}{1-\rho(0)} = \frac{\beta(\langle k\rangle - 1)}{\gamma}\big[1 - e^{-\gamma t}\big] + \frac{\beta}{\gamma + \beta\langle k\rangle}\big[1 - e^{-(\gamma+\beta\langle k\rangle)t)}\big] \\
\Rightarrow &\rho(t) = \Big\{1 + \frac{1-\rho(0)}{\rho(0)}\exp\big\{-\frac{\beta(\langle k\rangle - 1)}{\gamma}\big[1 - e^{-\gamma t}\big] - \frac{\beta}{\gamma + \beta\langle k\rangle}\big[1 - e^{-(\gamma+\beta\langle k\rangle)t)}\big]\big\}\Big\}^{-1}
\end{aligned} \tag{4.2.1.6}$$

and the saturated fraction of diffusion can be obtained by letting $t \to \infty$, so that

$$\begin{aligned}
\rho(\infty) &= \big\{1 + \frac{1-\rho(0)}{\rho(0)}\exp[-\frac{\beta}{\gamma}(\langle k\rangle - 1) - \frac{\beta}{\gamma + \beta\langle k\rangle}]\big\}^{-1} \\
&= \big\{1 + \frac{1-\rho(0)}{\rho(0)}\exp[-\theta(\langle k\rangle - 1) - \frac{\theta}{1 + \theta\langle k\rangle}]\big\}^{-1},
\end{aligned} \tag{4.2.1.7}$$

where $\theta = \beta/\gamma$. Hence the saturated fraction of diffusion $\rho(\infty)$ depends on β and γ only through the ratio $\theta = \beta/\gamma$.

If we simply set $\delta_t = 1$ or 0 as is normally done in the literature, instead of incorporating the information of $\rho(t)$ into the factor δ_t, we would get by setting $\delta_t = 1$,

$$\rho(t) = \big\{1 + \frac{1-\rho(0)}{\rho(0)}\exp\big[-\frac{\beta(\langle k\rangle - 1)}{\gamma}(1 - e^{-\gamma t})\big]\big\}^{-1}, \tag{4.2.1.8}$$

and correspondingly, $\rho(\infty) = \big\{1 + \frac{1-\rho(0)}{\rho(0)}\exp[-\theta(\langle k\rangle - 1)\big\}^{-1}$.

4.2.1.1 Numerical Experiments

Next, we compare the results from MC simulation with those derived based on the homogeneous MF approximation. The network of size $N = 5000$ is generated by the Gilbert model [45] while the degree of each node follows a Poisson distribution $k \overset{iid}{\sim} Poisson(\lambda)$, with the density of the network controlled by the parameter λ. For each setting, the results are obtained by averaging over 10^3 realizations. Since the accuracy of the MF approximation depends on the saturated fraction of infection/diffusion [47], we control the factor $\theta = \beta/\gamma$ for each case to ensure that the saturated fraction of diffusion in all settings is approximately $\rho(\infty) \sim 0.5$ for making the results comparable.

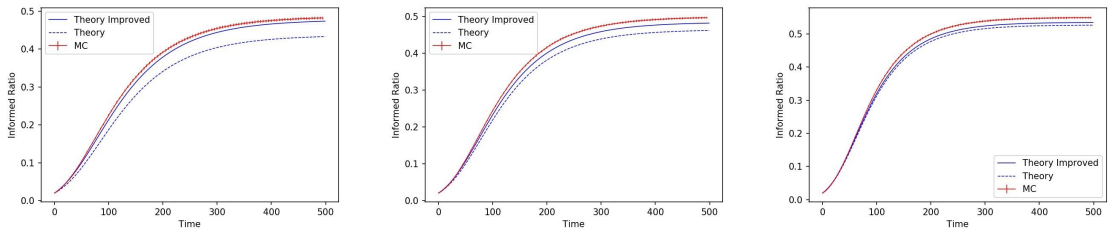


Figure 4.1: Comparison of results from the MC simulation estimates and theoretical results derived based on the homogeneous MF approximation for time dependent diffusion size in the whole time range. (a): Left panel, $\lambda = 5$; (b): Middle panel, $\lambda = 10$; (c): Right panel, $\lambda = 15$. *Theory improved* stands for Equation 4.2.1.6, *Theory* stands for Equation 4.2.1.8, and *MC* stands for MC simulation. For all the cases $\rho(0) = 2\%$

The results are depicted in Figure 4.1, and we can see that the results from the MC simulation agree well with the theoretical results in the whole range of network densities considered. Significant improvement over Equation 4.2.1.8 is achieved after incorporating the information of approximated $\rho(t)$ in the factor δ_t in Equation 4.2.1.6, especially for sparser networks since for dense networks $\langle k \rangle$ is large enough to shield the influence of changing from $\langle k \rangle - 1$ to $\langle k \rangle - \delta_t$.

4.2.2 Heterogeneous Mean Field Approximation

In terms of real life applications for information diffusion, it becomes rather restrictive to assume a homogeneous degree distribution for nodes since many real world networks are characterized by degree distributions exhibiting high variance. Therefore, in this section we investigate the heterogeneous version of the MF approximation which takes into consideration the degree distribution of nodes in the network. Specifically, nodes of the same degree can be treated equivalently and collectively, while nodes of different degrees need to be treated separately. At time t on a network of size N, if we denote by $S_k(t)$ the total number of susceptible nodes and by $I_k(t)$ the total number of informed/infected nodes of degree k respectively, then,

$$dI_k(t) = S_k(t)\big[1 - (1 - \beta e^{-\gamma t} dt)^{k\Theta_k(t)}\big] \stackrel{dt\to 0}{=} S_k(t)\beta e^{-\gamma t} k\Theta_k(t)dt, \tag{4.2.2.1}$$

where $\Theta_k(t)$ is the probability of reaching an informed neighbor at time t starting from an arbitrary edge of a susceptible node of degree k. For simplifying the calculation, we only consider the case where there is no degree correlation and therefore $\Theta_k(t) = \Theta(t)$ is independent of k, which is true for networks generated by some popular random graph models and their real life counter parties [33]. This assumption can be easily relaxed to incorporate the degree correlation structures as discussed in [12, 113]. Therefore, we have

$$\begin{aligned}\Theta(t) &= \frac{\sum_{k'=1}^{N-1}(k' - \delta_{k'}(t))I_{k'}(t)}{\sum_{k'=1}^{N-1} N_{k'}k'} = \frac{\sum_{k'=1}^{N-1}(k' - \delta_{k'}(t))\frac{I_{k'}(t)}{N_{k'}}N_{k'}}{N\langle k\rangle} \\ &= \sum_{k'=1}^{N-1}(k' - \delta_{k'}(t))\rho_{k'}(t)P(k')/\langle k\rangle,\end{aligned} \tag{4.2.2.2}$$

where $\rho_{k'}(t) = I_{k'}(t)/N_{k'}$, $\delta_{k'}(t) = 1 - I_{k'}(0)/I_{k'}(t) = 1 - \rho_{k'}(0)/\rho_{k'}(t)$, $N_{k'}$ is the expected number of nodes with degree k' in the network, and $P(k') = N_{k'}/N$. Then,

Equation 4.2.2.1 becomes

$$d\rho_k(t) = (1-\rho_k(t))\beta e^{-\gamma t} k\Theta(t)dt. \tag{4.2.2.3}$$

Together with Equation 4.2.2.2, the above ordinary differential equation (ODE) can be solved numerically given proper initial conditions.

In the following, we investigate the early stage evolution of the system when $t \to 0$ in order to study the outburst dynamics. Assume that we start with a small $\rho(0)$, such that we can safely ignore the terms of order $O(\rho^2(t))$ at early stages and approximately we set $\delta_k(t) = 1 - \rho_k(0)/\rho_k(t) \approx 1$ as is normally done in the literature. Take the time derivative of Equation 4.2.2.2 and incorporate Equation 4.2.2.3 to obtain

$$\begin{aligned}
\dot{\Theta}(t) &= \sum\nolimits_{k'=1}^{N-1}(k'-1)\dot{\rho}_{k'}(t)P(k')/\langle k\rangle = \sum\nolimits_{k'=1}^{N-1}(k'-1)\beta e^{-\gamma t}k'\Theta(t)P(k')/\langle k\rangle \\
&= \beta e^{-\gamma t}(\langle k^2\rangle - \langle k\rangle)\Theta(t)/\langle k\rangle \\
\Rightarrow \Theta(t) &= \Theta(0)\exp\big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}(1-e^{-\gamma t})\big].
\end{aligned} \tag{4.2.2.4}$$

If we assume a uniform density of initial seed nodes across groups of degrees $\rho_k(0) = \rho(0)$, from Equation 4.2.2.2 we then get

$$\Theta(0) = \sum\nolimits_{k'=1}^{N-1}(k' - \delta_{k'}(0))\rho(0)P(k')/\langle k\rangle = \rho(0),$$

where we have used the fact that $\delta_k(0) = 1 - \rho_k(0)/\rho_k(0) = 0$. Plugging the result of Equation 4.2.2.4 back into Equation 4.2.2.3 together with the initial condition $\Theta(0) = \rho(0)$, and ignoring terms of order $O(\rho^2(t))$ and solve, we obtain

$$\begin{aligned}
\rho_k(t) &= \rho_k(0) + \beta k\int_0^t e^{-\gamma t}\Theta(0)\exp\Big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\big(1-e^{-\gamma t}\big)\Big]dt \\
&= \rho(0) + k\rho(0)\frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}\Big\{\exp\Big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\big(1-e^{-\gamma t}\big)\Big] - 1\Big\}.
\end{aligned} \tag{4.2.2.5}$$

For two groups of degree k_1 and k_2 we can see that

$$\frac{\rho_{k_1}(t)-\rho_{k_1}(0)}{\rho_{k_1}(t)-\rho_{k2}(0)}=\frac{\rho_{k_1}(t)-\rho_{(}0)}{\rho_{k_1}(t)-\rho_{(}0)}=\frac{k_1}{k_2}; \tag{4.2.2.6}$$

Thus, the growth in the diffusion fraction (fraction of newly informed nodes) for each group is proportional to its degree k at early stages. This is expected since the diffusion process is localized on groups of high degrees at early stages and then gradually diffuse to low degree groups, as is discussed in [9]. On the other hand, $\sum_{k=1}^{N-1} N_k\rho_k(t)=N\rho(t)\Rightarrow\rho(t)=\sum_{k=1}^{N-1}P(k)\rho_k(t)$, therefore,

$$\begin{aligned}\rho(t)&=\sum_k P(k)\rho_k(t)\\&=\rho(0)\Big\{1+\frac{\langle k\rangle^2}{\langle k^2\rangle-\langle k\rangle}\Big\{\exp\Big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle-\langle k\rangle}{\langle k\rangle}\big(1-e^{-\gamma t}\big)\Big]-1\Big\}\Big\}.\end{aligned} \tag{4.2.2.7}$$

In order to find the *phase transition time point*, when the growth of the number of nodes informed in the network first starts to decrease, we take the first and second time derivatives of $\rho(t)$,

$$\begin{aligned}\dot\rho(t)&=\rho(0)\beta\langle k\rangle\exp\Big[-\gamma t+\frac{\beta}{\gamma}\frac{\langle k^2\rangle-\langle k\rangle}{\langle k\rangle}\big(1-e^{-\gamma t}\big)\Big]\\\ddot\rho(t)&=\rho(0)\beta\langle k\rangle\exp\Big[-\gamma t+\frac{\beta}{\gamma}\frac{\langle k^2\rangle-\langle k\rangle}{\langle k\rangle}\big(1-e^{-\gamma t}\big)\Big]\Big[-\gamma+\beta\frac{\langle k^2\rangle-\langle k\rangle}{\langle k\rangle}e^{-\gamma t}\Big].\end{aligned} \tag{4.2.2.8}$$

The *phase transition time point* τ can be calculated by simply setting $\ddot\rho(\tau)=0$ to get

$$-\gamma+\beta\frac{\langle k^2\rangle-\langle k\rangle}{\langle k\rangle}e^{-\gamma\tau}=0\Rightarrow\tau=\frac{1}{\gamma}\log\Big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle-\langle k\rangle}{\langle k\rangle}\Big]. \tag{4.2.2.9}$$

A possible application of the phase transition time point is in marketing. It provides marketeers an approximate time to stop or decrease spending on advertisement since it represents the time when the marginal profit, which can be defined as the

marginal diffusion size minus the expenditures on advertising per unit time, starts to decrease. This is consistent with results from the literature in the field of dynamic optimal advertising control in marketing, where expenditures on advertising are expected to be time dependent and shall be terminated after certain conditions are met as suggested by some theoretical models [114, 56].

To identify the *epidemic threshold* condition such that the growth of diffusion will not decrease from the beginning when $t = 0$, we need to ensure $\tau > 0$, and thus the epidemic threshold condition is $\beta/\gamma > \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$. Comparing it with the epidemic threshold found by employing percolation theory, where it is assumed that the reproduction number originating from newly informed nodes needs to be greater than 1 [93, 95], the epidemic threshold derived here by setting $\delta_t = 1$ is consistent with the former since it captures the contribution from the newly informed nodes.

Impact of seed nodes allocation: In the above, we assumed a uniform allocation of seed nodes across groups of different degrees. One interesting variation is to assume simply $\rho_k(0) = \rho(0)k/\langle k \rangle$, i.e., the fraction of seed nodes in each group is proportional to its degree k. As will be shown later, with this initial condition, we can incorporate the time dependent information of $\delta_k(t)$ in deriving $\rho(t)$ as was done in Section 4.2.1, instead of simply setting $\delta_k(t) = 1$ or 0 as is done in the previous derivation. In the following, we will first set $\delta_k(t) = 0$ at early stages and find an approximate analytical solution for $\rho_k(t)$, based on which we can find an analytical solution for $\delta_k(t)$ and then obtain an approximate solution of the early stage diffusion. Plugging $\rho_k(0) = \rho(0)k/\langle k \rangle$ and $\delta_k(t) = 0$ into Equation 4.2.2.2, we have

$$\Theta(0) = \sum\nolimits_{k=1}^{N-1} k \frac{\rho(0)k}{\langle k \rangle} P(k)/\langle k \rangle = \frac{\langle k^2 \rangle}{\langle k \rangle^2} \rho(0). \tag{4.2.2.10}$$

Plugging the result of Equation 4.2.2.4 back into Equation 4.2.2.3 together with the initial condition $\Theta(0) = \frac{\langle k^2 \rangle}{\langle k \rangle^2} \rho(0)$, and ignoring terms of order $O(\rho^2(t))$ and solve, we

obtain

$$\begin{aligned}\rho_k(t) &= \rho_k(0) + \beta k \int_0^t e^{-\gamma t}\Theta(0)\exp\Big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle}{\langle k\rangle}\big(1-e^{-\gamma t}\big)\Big]dt \\ &= \rho_k(0) + \frac{k\rho(0)}{\langle k\rangle}\Big\{\exp\Big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle}{\langle k\rangle}\big(1-e^{-\gamma t}\big)\Big]-1\Big\}. \\ &= \frac{k\rho(0)}{\langle k\rangle}\exp\Big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle}{\langle k\rangle}\big(1-e^{-\gamma t}\big)\Big] \overset{t\to 0}{\approx} \rho_k(0)e^{\beta\frac{\langle k^2\rangle}{\langle k\rangle}t}, \end{aligned} \tag{4.2.2.11}$$

where we have again used the approximation $1-e^{-\gamma t}\approx\gamma t$ and changed to $\delta_t = 0$ from 1 for Equation 4.2.2.4. Thus, $\delta_k(t) = \delta_t = 1-\rho_k(0)/\rho_k(t) \approx 1-e^{-\beta\frac{\langle k^2\rangle}{\langle k\rangle}t}$, independent of k. Intuitively, this is because we assign the seed node to each group of nodes with fraction proportional to the degree, while at the same time, the speed of information diffusion in each group is also proportional to the degree, therefore, the ratio $\rho_k(0)/\rho_k(t)$ is constant and the same across different groups of nodes of different degrees, at least in the early stage evolution. For two groups of degree k_1 and k_2, we can see that in this case

$$\frac{\rho_{k_1}(t)}{\rho_{k_2}(t)} \approx \frac{\rho_{k_1}(0)}{\rho_{k_2}(0)} = \frac{k_1}{k_2}, \tag{4.2.2.12}$$

the total fraction of diffusion in each group is also proportional to its degree k. Comparing this result to that in Equation 4.2.2.6, we find that after allocating seed nodes in a way such that the fraction of seed nodes in each group is proportional to its degree k, we essentially *moved* the system directly to its *equilibrium* state such that the total fraction of nodes informed in each group will always be proportional to its degree k instead of gradually converging to this ratio as in the case based on Equation 4.2.2.6.

As for $\rho(t)$, by taking into consideration the contribution from all groups, we can

obtain

$$\rho(t) = \sum_{k=1}^{N-1} \rho_k(t)P(k) = \rho(0)\exp\Big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle}{\langle k\rangle}\big(1-e^{-\gamma t}\big)\Big] \approx \rho(0)e^{\beta\frac{\langle k^2\rangle}{\langle k\rangle}t}. \quad (4.2.2.13)$$

Next, we use the approximate analytical solution of δ_t in deriving $\rho(t)$. Setting $\delta_k(t) = \delta_t$, using the result of Equation 4.2.2.11 in Equation 4.2.2.3 and combining with Equation 4.2.2.2, we get

$$\begin{aligned}
\dot{\rho}_k(t) &= (1-\rho_k(t))\beta e^{-\gamma t}k\sum_{k'=1}^{N-1}(k'-\delta_t)\rho_{k'}(t)P(k')/\langle k\rangle \\
&\approx (1-\rho_k(t))\beta e^{-\gamma t}k\sum_{k'=1}^{N-1}(k'-\delta_t)\rho_{k'}(0)e^{\beta\frac{\langle k^2\rangle}{\langle k\rangle}t}P(k')/\langle k\rangle \\
&= (1-\rho_k(t))\beta e^{-\gamma t}k\sum_{k'=1}^{N-1}(k'-\delta_t)\rho(0)k'/\langle k\rangle e^{\beta\frac{\langle k^2\rangle}{\langle k\rangle}t}P(k')/\langle k\rangle \\
&= (1-\rho_k(t))\beta e^{-\gamma t}\rho_k(t)\frac{\langle k^2\rangle - \delta_t\langle k\rangle}{\langle k\rangle}. \quad (4.2.2.14)
\end{aligned}$$

Taking into consideration the contribution from all groups and ignoring terms of order $O(\rho^2(t))$ at early stages, we have

$$\dot{\rho}(t) = \sum_{k=1}^{N-1} P(k)\dot{\rho}_k(t) \approx \beta e^{-\gamma t}\rho(t)\frac{\langle k^2\rangle - \delta_t\langle k\rangle}{\langle k\rangle}. \quad (4.2.2.15)$$

Finally, plugging in $\delta_t = 1 - e^{-\beta\frac{\langle k^2\rangle}{\langle k\rangle}t}$ and solve, we obtain

$$\rho(t) = \rho(0)\exp\Big\{\frac{\beta}{\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\big(1-e^{-\gamma t}\big) + \frac{\beta}{\gamma+\beta\langle k^2\rangle/\langle k\rangle}\big(1-e^{-(\gamma+\beta\langle k^2\rangle/\langle k\rangle)t}\big)\Big\}. \quad (4.2.2.16)$$

If we have simply set $\delta_t = 1$, then the above becomes

$$\rho(t) = \rho(0)\exp\big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\big(1-e^{-\gamma t}\big)\big]. \quad (4.2.2.17)$$

To find the *phase transition* time point τ, similarly, we take the second time derivative

of $\rho(t)$. To be consistent with the case when we assume uniform allocation of seed nodes across groups, here we set $\delta_t = 1$ and from Equation 4.2.2.15 we have

$$\begin{aligned}\ddot{\rho}(t) &= \beta\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\dot{\rho}(t)e^{-\gamma t} - \beta\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\gamma\rho(t)e^{-\gamma t} \\ &= \big(\beta\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}e^{-\gamma t} - \gamma\big)\beta\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\rho(t)e^{-\gamma t}. \end{aligned} \tag{4.2.2.18}$$

Thus the *phase transition* time point is

$$\tau = \frac{1}{\gamma}\log\Big[\frac{\beta}{\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\Big], \tag{4.2.2.19}$$

obtained by setting $\ddot{\rho}(\tau) = 0$. As for the *epidemic threshold*, following the same line of reason as before, we set $\delta_t = 1$ and ensure $\tau > 0$, thus, the condition is $\beta/\gamma > \frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}$.

We can see that after modifying the allocation of seed nodes from $\rho_k(0) = \rho(0)$ to $\rho_k(0) = \rho(0)k/\langle k\rangle$, neither the *phase transition* time point nor the *epidemic threshold* changes, indicating that the results obtained for the two are pretty robust in terms of the initial conditions.

4.2.2.1 Numerical Experiments

Next, we compare the results of MC simulation with those derived based on the heterogeneous MF approximation. The network of size $N = 10^4$ is generated as scale free by a configuration model, while the node degree distribution follows a power law one $P(k) \sim k^{-\lambda}$ with $\lambda = 3.5$. We control the network heterogeneity defined as $\langle k^2\rangle/\langle k\rangle$ by setting the minimum degree K_{min} [9, 8]. In this case, $\langle k\rangle \approx 5K_{min}/3$, $\langle k^2\rangle \approx 5K_{min}^2$ and therefore $\langle k^2\rangle/\langle k\rangle \approx 3K_{min}$. The results in each setting are obtained by averaging over 10^3 realizations.

First, we study the case of uniform seed nodes allocation across groups; namely, $\rho_k(0) \approx \rho(0)$.

We compare the MC simulation results to the numerical solution of the ODE equations for the time dependent diffusion fraction across the whole time range. Similarly, to facilitate comparison across different settings, we control β/γ for each setting to ensure that the saturated ratio is approximately $\rho(\infty) \sim 0.5$. The results are depicted in Figure 4.2.

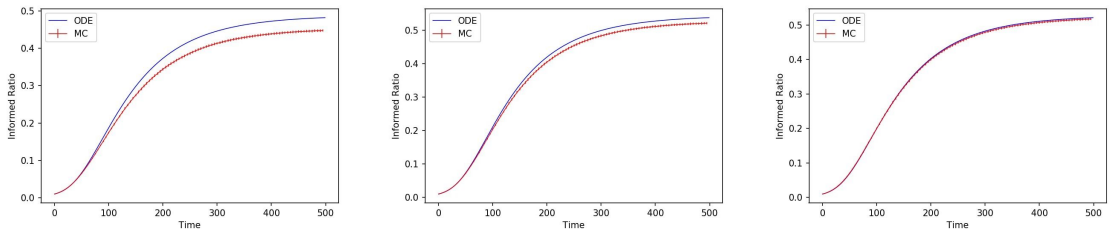


Figure 4.2: Comparison between MC simulations and numerical solutions of ODEs for time dependent diffusion size in the whole time range. (a): Left panel, $K_{min} = 5$; (b): Middle panel, $K_{min} = 10$; (c): Right panel, $K_{min} = 15$. *ODE* stands for numerical solution of Equations 4.2.2.3 and 4.2.2.2, and *MC* stands for MC simulation. For all the cases, $\rho(0) = 1\%$

A good agreement between the results of MC simulation and ODE solution is achieved for all three settings, and the agreement improves for higher K_{min} corresponding to denser networks. This is expected since higher accuracy of the MF approximation is obtained for diffusion on networks with higher average degree of neighbors [47] which is calculated as $\langle k^2 \rangle / \langle k \rangle \sim K_{min}$.

Next, we compare the MC simulation with the analytical solution of Equation 4.2.2.7 for the diffusion process at early stages. The results are depicted in Figure 4.3.

Similarly, a very good agreement between the theoretical results and the MC simulation is obtained for all three settings, and the agreement improves with higher K_{min}, corresponding to networks of higher density. Clearly, based on the semi-log plots for both the results of MC simulation and the close form solution based on MF analysis in Figure 4.4, the diffusion size increases exponentially with time at the early

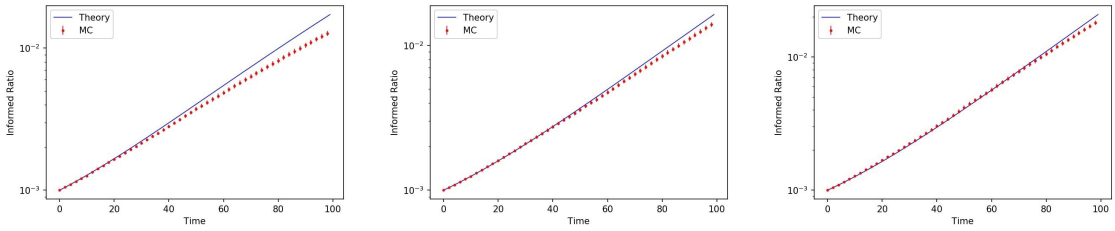


Figure 4.3: Comparison between MC simulation and theoretical results for the early stage evolution of the system. (a): Left panel, $K_{min} = 5$; (b): Middle panel, $K_{min} = 10$; (c): Right panel, $K_{min} = 15$. *Theory* stands for Equation 4.2.2.7 and *MC* stands for MC simulation. For all the cases, $\rho(0) = 0.1\%$

stage of the evolution of the information diffusion process.

Then, we compare the phase transition time point derived from Equation 4.2.2.9 and from the numerical solution of Equations 4.2.2.3 and 4.2.2.2 with that of MC simulation for different combinations of β and γ. Besides, we also include the saturated diffusion size $\rho(\infty)$ for all the settings to investigate the *epidemic threshold*. The results are depicted in Figure 4.4. Again, a good agreement for τ between the three

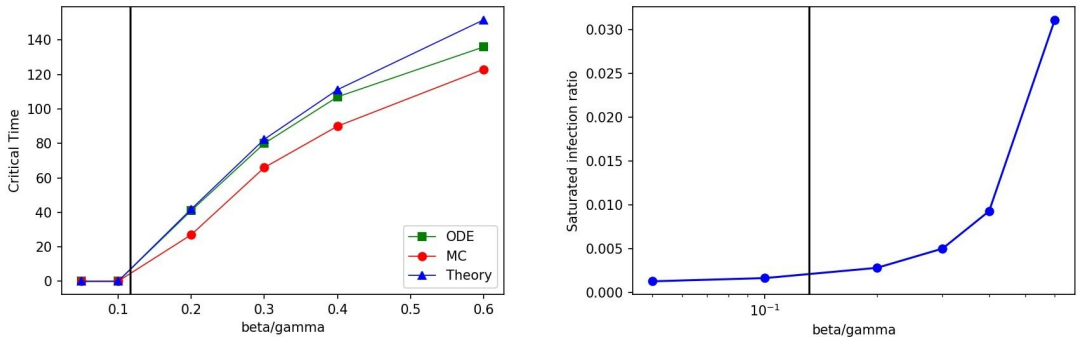


Figure 4.4: (a): Left panel, comparison of *phase transition* time points. *ODE* stands for numerical solution of Equations 4.2.2.3 and 4.2.2.2, *MC* stands for MC simulation and *Theory* stands for Equation 4.2.2.9. (b): Right panel, total fraction of diffusion $\rho(\infty)$ for different β/γ. For visual guidance, we also include a black vertical line located at $\beta/\gamma = \langle k\rangle/\langle k^2\rangle$ for (a) and $\beta/\gamma = \frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}$ for (b). Here $K_{min} = 3.0$ and $\rho(0) = 0.1\%$

methods is achieved. For visual guidance, we included a black vertical line located at $\beta/\gamma = \langle k\rangle/\langle k^2\rangle$ in Figure 4.4(a), different from the *epidemic threshold*, which is located at $\langle k\rangle/(\langle k^2\rangle - \langle k\rangle)$ in Figure 4.4(b). This is because the *epidemic threshold* is derived based on the basic reproduction number from the newly informed nodes, i.e., the size of the newly informed nodes from the existing newly informed nodes, which equivalently sets $\delta_t = 1$. On the other hand, the condition that the *phase transition time* equals 0 is equivalent to the case that the reproduction number from the seed nodes is smaller than 1, which equivalently sets $\delta_t = 0$, and thus the denominator becomes $\langle k^2\rangle$ in stead of $\langle k^2\rangle - \langle k\rangle$. As can be noted, when $\beta/\gamma \leq 0.1$, which is smaller than $\langle k\rangle/\langle k^2\rangle$, we should have $\tau = 0$, representing decaying evolution from the beginning, and this is captured by all three methods.

Next, we study the case where the seed nodes are allocated across groups such that the fraction in each group is proportional to the degree k: $\rho_k(0) = \rho(0)k/\langle k\rangle$.

Since the only differences after using this alternative initial condition are the early stage diffusion, *phase transition* time point, and *epidemic threshold*, in the following, we compare the derived expression with the results from MC simulation. The network settings are the same as above, and the results are depicted in Figure 4.5. Similarly, for early stage evolution, a good agreement between MC simulation and the theoretical results is obtained, and the agreement improves as the networks become denser, corresponding to higher K_{min}. As for the *phase transition* time points, the results obtained based on the two methods also agree well with each other.

In summary, the theoretical analysis shows very good agreement with experimental results, a finding consistent across many other settings considered (not shown for space consideration).

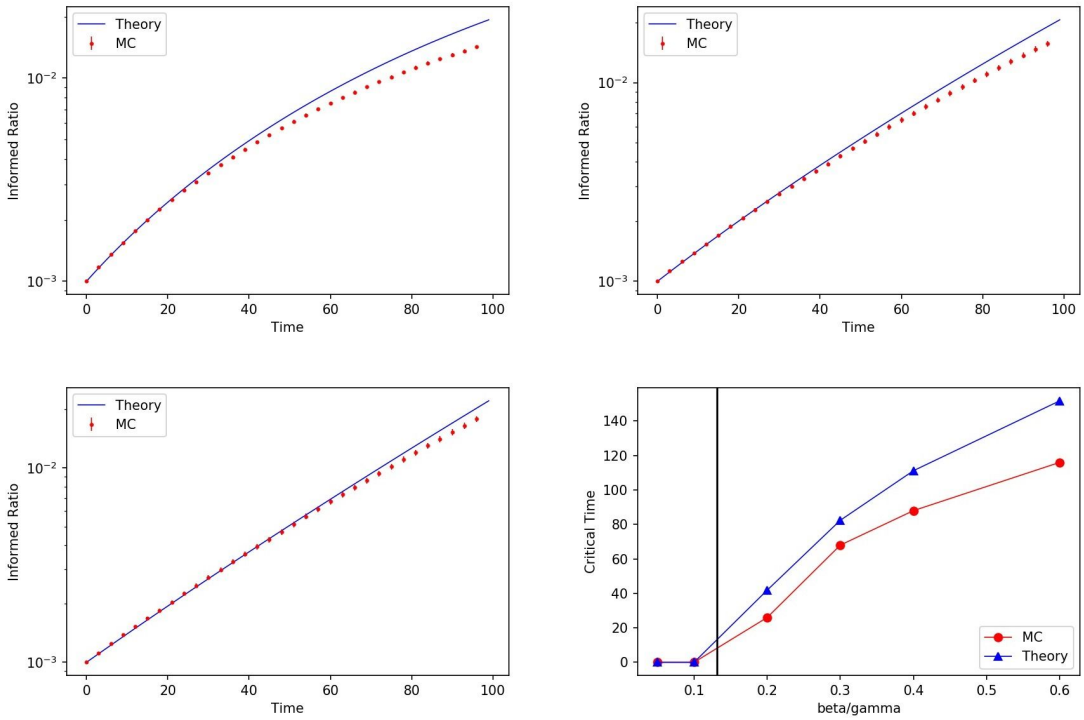


Figure 4.5: Comparison between MC simulation and theoretical analysis for both early stage diffusion size and *phase transition* time. (a): Upper left panel, $K_{min} = 3$; (b): Upper right panel: $K_{min} = 5$; (c): Lower left panel, $K_{min} = 10$; (d): Lower right panel, τ for $\rho(t)$. For (a-c), *Theory* stands for Equation 4.2.2.16 and $\rho(0) = 0.1\%$. For (d), *Theory* stands for Equation 4.2.2.19 and for visual guidance, we include a black vertical line located at $\beta/\gamma = \langle k \rangle / \langle k^2 \rangle$. $\rho(0) = 0.1\%$ and $K_{min} = 3$.

4.3 Maximum Weight Tree (MWT) approximation

4.3.1 Path dependent diffusion edge probabilities

In Section 4.2, we considered the SIT model from a macroscopic perspective when a randomly selected proportion $\rho(0)$ of nodes was set in the I state at the beginning of time. In this section, we investigate the problem from a microscopic perspective by fixing the network and explicitly compute the probability of getting informed for each node depending on its actual path to the single seed node. This analysis is also instructive on why the percolation theory is not applicable for the SIT model. For

keeping the presentation simple, we say a node is at stage l if it can reach the seed node by a minimum of l hops/edges.

To simplify the calculations, we assume a constant initial diffusion rate β, independent of edges, which decays with time $\beta_t = \beta e^{-\gamma t}$, and a continuous time evolution. Therefore, the survival probability of the diffusion process up to time t is $\prod_{i=0}^{M-1}(1-\beta e^{-\gamma t_i}dt) = \exp\big[\sum_{i=0}^{M-1}\log(1-\beta e^{-\gamma t_i}dt)\big] \approx \exp\big[\sum_{i=0}^{M-1} -\beta e^{-\gamma t_i}dt\big] \overset{dt\to 0}{=} e^{-\int_{\tau=0}^{t}\beta e^{-\gamma\tau}d\tau}$, where $t_i = idt$ and $t = Mdt$, where M represents the total number of time steps when t is discretized. Thus, the probability for a node at stage 1 to get informed is,

$$\begin{aligned} p_1 &= \int_{t=0}^{\infty} e^{-\int_{\tau=0}^{t}\beta e^{-\gamma\tau}d\tau}\beta e^{-\gamma t}dt = \int_{t=0}^{\infty} e^{-\frac{\beta}{\gamma}(1-e^{-\gamma t})}\beta e^{-\gamma t}dt = e^{-\frac{\beta}{\gamma}}\int_{t=0}^{\infty} e^{\frac{\beta}{\gamma}e^{-\gamma t}}\beta e^{-\gamma t}dt \\ &= e^{-\frac{\beta}{\gamma}}\big(-e^{\frac{\beta}{\gamma}e^{-\gamma t}}\big)\Big|_0^{\infty} = e^{-\frac{\beta}{\gamma}}(e^{\frac{\beta}{\gamma}}-1). \end{aligned} \tag{4.3.1}$$

In order for a node at stage 2 to get informed, first node 1 needs to get informed from node 0 by a certain time t_1 and then node 2 gets informed from node 1 at a later time $t_2 > t_1$. The survival probability of the first event is $e^{-\int_{\tau=0}^{t_1}\beta e^{-\gamma\tau}d\tau}$ and for the second event is $e^{-\int_{\tau=t_1}^{t_2}\beta e^{-\gamma\tau}d\tau}$. Therefore, the total probability for node in stage 2 to get informed is,

$$\begin{aligned} p_2 &= \int_{t_1=0}^{\infty} e^{-\int_{\tau=0}^{t_1}\beta e^{-\gamma\tau}d\tau}\beta e^{-\gamma t_1}dt_1 \int_{t_2=t_1}^{\infty} e^{-\int_{\tau=t_1}^{t_2}\beta e^{-\gamma\tau}d\tau}\beta e^{-\gamma t_2}dt_2 \\ &= e^{-\frac{\beta}{\gamma}}\int_{t_1=0}^{\infty} e^{\frac{\beta}{\gamma}e^{-\gamma t_1}}\beta e^{-\gamma t_1}dt_1 \int_{t_2=t_1}^{\infty} e^{\frac{\beta}{\gamma}\left(e^{-\gamma t_2}-e^{-\gamma t_1}\right)}\beta e^{-\gamma t_2}dt_2 \\ &= e^{-\frac{\beta}{\gamma}}\int_{t_1=0}^{\infty}\beta e^{-\gamma t_1}dt_1\Big[e^{\frac{\beta}{\gamma}e^{-\gamma t_1}}-1\Big] \\ &= p_1 - e^{-\frac{\beta}{\gamma}}\int_{t_1=0}^{\infty}\beta e^{-\gamma t_1}dt_1 = p_1 - e^{-\frac{\beta}{\gamma}}\frac{\beta}{\gamma}. \end{aligned} \tag{4.3.2}$$

Similarly,

$$
\begin{aligned}
p_3 &= \int_{t_1=0}^{\infty} e^{-\int_{\tau=0}^{t_1}\beta e^{-\gamma\tau}d\tau}\beta e^{-\gamma t_1}dt_1 \int_{t_2=t_1}^{\infty} e^{-\int_{\tau=t_1}^{t_2}\beta e^{-\gamma\tau}d\tau}\beta e^{-\gamma t_2}dt_2 \int_{t_3=t_2}^{\infty} e^{-\int_{\tau=t_2}^{t_3}\beta e^{-\gamma\tau}d\tau}\beta e^{-\gamma t_3}dt_3 \\
&= e^{-\frac{\beta}{\gamma}} \int_{t_1=0}^{\infty} e^{\frac{\beta}{\gamma}e^{-\gamma t_1}}\beta e^{-\gamma t_1}dt_1 \int_{t_2=t_1}^{\infty} e^{\frac{\beta}{\gamma}\left(e^{-\gamma t_2}-e^{-\gamma t_1}\right)}\beta e^{-\gamma t_2}dt_2 \int_{t_3=t_2}^{\infty} e^{\frac{\beta}{\gamma}\left(e^{-\gamma t_3}-e^{-\gamma t_2}\right)}\beta e^{-\gamma t_3}dt_3 \\
&= e^{-\frac{\beta}{\gamma}} \int_{t_1=0}^{\infty} \beta e^{-\gamma t_1}dt_1 \int_{t_2=t_1}^{\infty} \beta e^{-\gamma t_2}dt_2 \int_{t_3=t_2}^{\infty} e^{\frac{\beta}{\gamma}e^{-\gamma t_3}}\beta e^{-\gamma t_3}dt_3 \\
&= e^{-\frac{\beta}{\gamma}} \int_{t_1=0}^{\infty} \beta e^{-\gamma t_1}dt_1 \int_{t_2=t_1}^{\infty} \beta e^{-\gamma t_2}dt_2 \Big[e^{\frac{\beta}{\gamma}e^{-\gamma t_2}} - 1\Big] \\
&= p_2 - e^{-\frac{\beta}{\gamma}}\frac{1}{2}\frac{\beta^2}{\gamma^2}. \qquad (4.3.3)
\end{aligned}
$$

Following the same procedure, we can see that in general, for $l \geq 1$,

$$p_{l+1} = p_l - e^{-\frac{\beta}{\gamma}}\frac{1}{l!}\frac{\beta^l}{\gamma^l}. \qquad (4.3.4)$$

Pluging in $p_1 = 1 - e^{-\frac{\beta}{\gamma}}$, we get

$$p_{l+1} = 1 - \sum\nolimits_{i=0}^{l} \big(\frac{\beta}{\gamma}\big)^i / i! e^{-\frac{\beta}{\gamma}}. \qquad (4.3.5)$$

As $l \to \infty$, since $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$, we have,

$$p_l \overset{l\to\infty}{=} 1 - e^{\frac{\beta}{\gamma}}e^{-\frac{\beta}{\gamma}} = 0. \qquad (4.3.6)$$

SIR Model and Percolation Theory: For comparison purposes, we also study the probability of each node getting infected depending on the path to the seed node for the SIR model. Suppose as before, the infection rate is β and the node removal rate is γ, both of which are constant. Then,

$$p_l = \int_0^{\infty} e^{-\beta t_1}\beta e^{-\gamma t_1}dt_1 \cdots \int_0^{\infty} e^{-\beta t_k}\beta e^{-\gamma t_l}dt_l = \big(\frac{\beta}{\beta+\gamma}\big)^l. \qquad (4.3.7)$$

We can see here $p_{l+1}/p_l = \frac{\beta}{\beta+\gamma}$, representing the transmissibility of the edge linking nodes of stages l and $l+1$ is a constant, independent of stage l. On the contrary, for the SIT model from Equation 4.3.4 we get

$$\frac{p_{l+1}}{p_l} = 1 - \frac{1}{p_l} e^{-\frac{\beta}{\gamma}} \frac{1}{l!} \frac{\beta^l}{\gamma^l}; \tag{4.3.8}$$

the factor $e^{-\frac{\beta}{\gamma}} \frac{1}{l!} \frac{\beta^l}{\gamma^l}$ increases or decreases with l depending on the relationship between $\frac{\beta}{\gamma}$ and l, while p_l monotonically decreases with l, and thus the ratio p_{l+1}/p_l is dependent on l. So the key assumption in percolation theory of constant transmissibility is not satisfied.

4.3.2 MWT Approximation

Next, we make use of the results derived above and approximate the expected total fraction of diffusion of the network by growing a MWT starting from a seed node s, with the weight for each node approximates the probability that the node will ever get informed. Higher accuracy in approximation can be achieved for sparser networks, which is expected, since in this case the probability of multi-channel diffusion from seed node to the remaining nodes in the network is lower such that the tree approximation for the diffusion process is satisfactory. To achieve the task of growing a MWT, since we assume a constant initial diffusion rate β independent of edge, which is true for unweighted networks, we can directly apply the Dijkstra's algorithm in finding the shortest distance from seed node s to the remaining nodes in the network. For each node, suppose its shortest distance to node s is $l(\geq 1)$, then the weight would be $W_l = p_l = 1 - \sum_{i=0}^{l-1} \left(\frac{\beta}{\gamma}\right)^i / i! e^{-\frac{\beta}{\gamma}}$ from Equation 4.3.5. Sum up the weight of all nodes in the network would give us the approximated total fraction of diffusion by MWT.

Numerical Experiments To study the effectiveness of the MWT algorithm in estimat-

ing the expected total size of diffusion, we compare it with that of MC simulation estimates. The network of size $N = 10^4$ is generated with node degree following a power law distribution $P(k) \sim k^{-\lambda}$ and we consider three settings: $\lambda = 2.5, 3.0$ and 3.5. As for the parameters for the SIT model, we fix $\gamma = 0.01$ and carry out for each network setting experiments with different diffusion rate β. In each experiment, the ratio of the result based on MWT approximation over that of MC simulation is obtained by averaging over 2000 realizations: namely, 20 different random network topologies, with 100 realizations of the diffusion process for each one. As for the seed node selection, for each realization, we simply choose the hub node with the highest degree. The results are depicted in Figure 4.6 and as expected much better agreement between the MWT approximation and the MC simulation can be achieved for sparser networks.

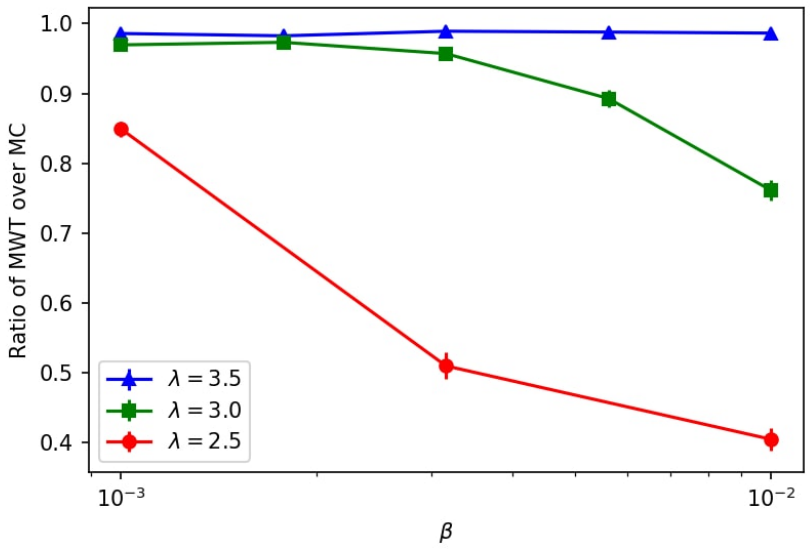


Figure 4.6: Ratio between the total diffusion fractions based on MWT and on MC simulations for different diffusion rate β with a fixed decay rate $\gamma = 0.01$. Different network densities are considered by controlling the parameter $\lambda = 2.5, 3.0, 3.5$.

4.4 Conclusion

To summarize, in this paper, the SIT model is proposed to model the information diffusion mechanism. Comparing to other popular and widely used models for epidemic processes such as SIR and SIS, the SIT model captures explicitly the "intrinsic time value", a key feature of information such as news, rumors, etc. Due to the path dependent diffusion probability along edges between susceptible and informed neighbors, traditional theoretical tools such as bond percolation approach is not applicable, which induces technical difficulties for analyzing the model from a theoretical perspective. Here we refer to the MF approach, covering both the homogeneous and heterogeneous cases. In particular, by incorporating the factor δ_t, capturing the fact that at least one neighbor of any newly informed neighbor is also informed, we are able to obtain much more accurate theoretical results from the MF analysis in terms of the agreement to the MC simulation results. Specifically, for homogeneous MF case, an analytical solution for the time dependent diffusion size at arbitrary time point is obtained, while for heterogeneous MF case, a close form solution of time dependent diffusion size at early stage of the evolution is obtained, which in turn presents us the *phase transition time points* as well as *epidemic threshold*. To the best of our knowledge, the factor δ_t has always been set to either 0 or 1 in the literature, and this is the first time it is accurately incorporated. On the other hand, we proposed a novel approach to estimate the final diffusion size based on constructing a MWT approximation which provides an exact approximation for tree-like networks and a highly accurate one for sparse networks. The latter is also fast to compute and provides another general tool for the analyst to obtain accurate approximations of the "epidemic's" size.

CHAPTER V

Information diffusion maximization on multilayer networks with community structures via Particle Swarm Optimization

5.1 Introduction

Information diffusion process has been extensively studied in the recent decades across various research disciplines due to its applications in various areas such as marketing, disease control, news propagation and political propaganda in terms of diffusion control [21, 71, 129, 90, 13, 3, 122]. On the one hand, we might be interested in minimizing some epidemic process such as disease transmission on the networks by disseminating information about the disease on the network, or minimizing the spread of certain news/information/rumor itself as in social census [129, 135, 130]. On the other hand, we are interested in maximizing the diffusion size as in marketing, news services and political propaganda. For example, in marketing application, viral marketing strategy is applied to promote a new product over a certain population based on the word of mouth diffusion of recommendations between individuals via connections of friendship or any other kind with the goal of maximizing the number of products purchasing [71]. Similarly in news propagation or political propaganda, we want to maximize the total size of population informed of the news or to adopt

some ideas. Naively speaking, if we do not have any budget constraint, we could well inform every one at time $t = 0$, in other words, we could set all the nodes as seed nodes, defined as the set of nodes aware of the information at the beginning. Of course, this is an over simplified problem and we would not expect to see this situation in real life given a large enough population.

Alternatively, we could set the budget constraint as the total number of seed nodes allowed to start with and the problem becomes finding k out of N nodes to target so as to maximize the diffusion size, which is a NP hard problem and different strategies are developed under different mathematical frameworks. In threshold models such as cascade model, linear threshold model and game theoretic approaches, for example, various algorithmic approaches were proposed to find approximate solutions [32, 61]. On the other hand, in independent interaction models such as the SIS and SIR models where the information diffusion process is treated as an epidemic process, various strategies based on measures such as k-shell, node degree and betweenness centrality are used to guide seed node selection process [65, 105, 43].

However, if we look closely at the problem in terms of application, the assumptions are still not very realistic as follows. (1) The cost of setting as seed node might be different across nodes for some applications and it is not very reasonable to assume a budget of k seed nodes allowance. For example, in marketing, the cost of a super star would be much higher compared to a random person you encounter on the street if you want him or her to advertise the product for you. And therefore, setting a hub node may quickly eat up your budget in a much higher rate than $1/k$ if we assume equivalent cost across nodes. (2) Once the budget is fixed, we did not take into consideration the cost any more in the optimization problem above, which is not reasonable since in some applications what really matters is the profit, which is equal to the revenue minus the cost. For example, in marketing, what we really want o maximize is the profit from the viral marketing activity, which for simplicity

purpose can be defined as the revenue from the products sold minus the cost invested in advertisement in the form of seed nodes allocation. Under this situation, we might be very flexible about the cost as long as the profit is large. For example, if the cost of advertisement is generally very cheap, we might ask more people and more platforms to advertise for us, while on the other hand, if the cost is generally very high, we might become pretty parsimonious.

In this work, we address this issue by the following. First of all, we use SIM model for our diffusion process and for simplicity reason define revenue as the total diffusion size in saturation, i.e., the total number of nodes ever get informed, or purchased the product in context of marketing, if we allow the system to evolve to infinite time. As for cost, we design the cost function for each node based on two characteristics. One plausible way is to define the personal cost of each node in terms of the degree of connectivity K, which in marketing context represents how popular/influential a person is in terms of spreading the recommendation for product purchase. However, this does not take into consideration of *personal perception*: some people might think of themselves very influential and thus would charge more because they know some people much more popular/influential. This means we need to take into consideration of the topology of the network starting from the seed node at hand, and one measure very suitable for this task is the personal MWT of the node, which is the expected diffusion size starting from the seed node assuming a tree like diffusion pattern.

Therefore, in the following experiments, we design the cost function as following. In terms of degree K, personal cost for each node can be proportional to its degree, and the total cost is thus $cost_{tot} = \sum_{i=1}^{M} K_i * W$ if we assume there are M nodes selected in the seen nodes set and W is the weight to control the cost. One possible variation is $cost_{tot} = \sum_{i=1}^{M} K_i^Q$, where the exponent Q is used to control the cost. On the other hand, in terms of MWT, we also can have two version of cost function, either $cost_{tot} = \sum_{i=1}^{M} MWT_i * W$ or $cost_{tot} = \sum_{i=1}^{M} MWT_i^Q$. And the total profit is defined

as $PT = F_{tot} - cost_{tot}$, where F_{tot} is simply the total size of diffusion in saturation if we start with the M seed nodes selected. And our objective is to optimize the total profit under different cost functions over a range of W and Q.

In fact, we can generalize the framework by assuming that the information diffuses on multilayer networks and for simplicity purpose, we include two layers, the primary, which is generated by standard stochastic block model with embedded community structure, and the secondary, which is generated as a scale free network. Information diffusion process on networks with community structure and on multilayer networks has been extensively studied, but not in our context in terms of profit optimization objective. The community structure in the primary layer captures the following feature: The network of influence or friendship of people tends to form communities, with dense connections within each community and sparse connections between communities[42]. And the diffusion process can evolve itself on this primary network and people exchange information or recommend new products by word of mouth. On the other hand, the scale free secondary layer network captures the channel of advertisements via mass media. For example, putting an advertisement spoken for by some super stars on platforms such as TV show, or on popular websites will behave in this similar way. As pointed out by some literature, people are becoming more and more resistant to traditional platforms of advertisements, and the hub nodes in the secondary layer are not as effective as their low degree counter parts in the primary layer[71]. To capture this feature, we set the diffusion rate on secondary layer to be $10\%\beta$, with β the diffusion rate between nodes from the same community on the primary layer. On the other hand, due to the lower influence across communities, we also set the diffusion rate between nodes from different communities to be $10\%\beta$. As for solving this optimization problem, we refer to the hybrid version of binary particle swarm optimization (PSO) algorithm, which is a powerful algorithm for general combinatorial optimization purpose[62, 63, 84].

The organization of the rest of the work is as following. In Section 5.2 we introduce the binary PSO algorithm and verify the feasibility of the algorithm by its successful performance on a set of test functions in numerical experiments. And in Section 5.3 we apply the algorithm on our profit optimization problem. We learned that when the cost function is very heterogeneous across candidate seed nodes as in the case when cost is a function of degree K, the optimal set of nodes selected will be mainly a set of low degree nodes, and only in the case of extremely small W or Q would high degree nodes be included. On the other hand, if the cost function is pretty homogeneous across nodes as in the case when cost is a function of MWT, the optimal seed nodes set will build up from high degree nodes to low degree nodes. For comparison, we also include the result based on a naive strategy: maximum personal profit (MPP), which is to identify the candidate seed node that will produce the highest profit if we set it as the single seed node. Specifically, for primary layer we find one seed node for each community and for secondary layer scale free network we find one for the whole network. When W or Q is not too small, good agreement between the optimal seed node allocation result based on PSO and that based on MPP is achieved. When we couple the two layer networks together, the community structure effect would be quenched in terms of the seed nodes allocation. And in our settings, utilizing two layers of networks for diffusion purpose outperforms single layer in terms of the total profit.

5.2 Particle Swarm Optimization

Particle swarm optimization algorithm was firstly introduced by Kennedy and Eberhart in 1995 for the optimization of continuous nonlinear functions and training neural networks [62]. The method was discovered through simulation of artificial social models such as bird flocking and fish schooling, and is a evolutionary computational technique, extremely useful in solving problems of optimizing an objective

function in the multidimensional space of parameters [80]. In the original setting, each particle flies in the space of parameters with its position representing a possible candidate solution and the velocity representing the speed of changing its position. The position of the i_{th} particle at time t is a vector $\mathbf{x}_i(t) = [x_{i1}(t), \cdots, x_{iD}(t)]$ with $i = 1, \cdots, N$, where N is the population size of the particles and D is the dimension of the problem at hand. The particle moves according to the following update equations,

$$\mathbf{v}_i(t+1) = \mathbf{v}_i(t) + c_1(\mathbf{p}_i(t) - \mathbf{x}_i(t))\mathbf{R}_1 + c_2(\mathbf{g}_i(t) - \mathbf{x}_i(t))\mathbf{R}_2 \tag{5.2.1}$$

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1) \tag{5.2.2}$$

Here c_1 and c_2 are two constants, usually in the range $c_1, c_2 \in (0, 4)$, $\mathbf{R}_1, \mathbf{R}_2$ are two diagonal matrices with $\mathbf{R}_1[d,d], \mathbf{R}_2[d,d] \sim Uniform(0,1)$, $\mathbf{p}_i(t)$ is the personal best position so far up to time t for particle i in terms of the objective function value, and $\mathbf{g}_i(t)$ is the global best position so far for all the particles up to time t. Intuitively, a particle's position at the next time step is a compromise amoung its previous position, previous velocity, best personal position and best global position, with c_1 and c_2 to control the weight of each contribution. In the following years after the introduction of the technique, a lot of attempts were made to improve the performance. In 1997 Shi and Eberhart introduced a new factor w named inertia weights for the previous velocity and soon after many variants appeared, including random inertia weight and linearly decreasing inertia weight among others[116, 36, 131]. And the update function for $\mathbf{v}_i(t)$ becomes

$$\mathbf{v}_i(t+1) = w(t)\mathbf{v}_i(t) + c_1(\mathbf{p}_i(t) - \mathbf{x}_i(t))\mathbf{R}_1 + c_2(\mathbf{g}_i(t) - \mathbf{x}_i(t))\mathbf{R}_2 \tag{5.2.3}$$

Then in 2002, Clerc and Kennedy introduced the constriction factor to ensure convergence of the algorithm and can be simplified as the following[24],

$$\begin{aligned} \mathbf{v}_i(t+1) &= K\left[\mathbf{v}_i(t) + c_1(\mathbf{p}_i(t) - \mathbf{x}_i(t))\mathbf{R}_1 + c_2(\mathbf{g}_i(t) - \mathbf{x}_i(t))\mathbf{R}_2\right] \\ K &= \frac{2}{\left|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}\right|}, \text{ where } \varphi = c_1 + c_2 > 4 \end{aligned} \tag{5.2.4}$$

On the other hand, a lot of attempts were also made on controlling the topology of the particles population, especially for the factor $\mathbf{g}_i(t)$ in order to achieve better performance. First of all, instead of as the global best position, $\mathbf{g}_i(t)$ was set as local best position, defined as the optimal one of the best personal positions of the k neighbors of the particle i. Some examples include random networks topology, 1D lattice and Von Neumann lattice[64].

The PSO algorithm was originally designed to solve continuous real valued problems, but fortunately, variants were proposed to solve discrete problems, of which the most representative and most popular one is the discrete binary version of the PSO algorithm proposed by Kennedy and Eberhart in 1997[63]. Since then, it has been widely studied and applied in various fields as a combinatorial optimization algorithm, including variable selection, signal processing, traveling salesman problem, and knapsack problems among others[80, 53, 79, 102, 126].

In the discrete binary version of the PSO algorithm proposed by Kennedy and Eberhart, the particles fly in the space restricted to binary 0 and 1 in each dimension, but the velocity update equation (2.1) is kept unchanged. Here the positions $\mathbf{x}_i(t) = [x_{i1}(t), \cdots, x_{iD}]$, $x_{id} \in \{0, 1\}, d = 1, \cdots, D$ still represent possible candidate solutions. However, just as in logistic regression, each dimension of the velocity $\mathbf{v}_{id}(t)$ here does not represent the real change in position anymore, instead, it represents the log odds of the d_{th} dimension of the particle i taking value 1 compared to 0. In this way, the position of particles in the space is ephemeral and probabilistic since the position

does not exist until realized by the following,

$$x_{id}(t+1) = \begin{cases} 1, & \text{if } S(v_{id}(t+1)) > u \\ 0, & \text{otherwise} \end{cases} \tag{5.2.5}$$

Here $S(v_{id}(t+1)) = \frac{e^{v_{id}(t+1)}}{1+e^{v_{id}(t+1)}} = \frac{1}{1+e^{-v_{id}(t+1)}}$ is sigmoid function, or, in terms of logistic regression, the probability form, and $u \sim Uniform(0,1)$ is a random number following uniform distribution in the range $[0,1]$. On the other hand, of course, those strategies developed for improving performance for the continuous version of PSO algorithm such as local neighbors, constriction factors and inertia weights still applies since the velocity update function remains. Since the introduction of this binary version of the PSO algorithms, due to its powerful applications in various fields, a lot of modifications/improvements were developed tailored for the tasks at hand, including charged binary PSO, genotype-phenotype binary PSO and PSO with mutation and cross over processes [117, 70, 84]. For example, in analogy to the genotype-phenotype mechanism in biology, Lee, etc, proposed viewing the velocity as genotype $x_{id}^{(g)}(t)$, hidden, and position as phenotype $v_{id}(t) = x_{id}^{(p)}(t)$, and the particles fly in the genotype space with possible mutation,

$$\begin{aligned} v_{id}(t+1) &= K\big[v_{id}(t) + c_1(p_{id}(t) - x_{id}(t))R_1 + c2(g_{id}(t) - x_{id}(t))R_2\big] \\ x_{id}^{(g)}(t+1) &= x_{id}^{(g)}(t+1) + v_{id}(t+1) \\ x_{id}^{(g)}(t+1) &= \begin{cases} -x_{id}^{(g)}(t+1), & \text{if } p_m > u \\ x_{id}^{(g)}(t+1), & \text{otherwise} \end{cases} \end{aligned} \tag{5.2.6}$$

Here as before $u \sim Uniform(0,1)$ and p_m is the probability of velocity mutation chosen for best performance. As a matter of fact, the only difference between this formulation and the original proposed by the Kennedy is the conceptual framework,

since we can easily combine the two update equations above and it becomes the version with inertia weight $w = 1 + K$.

$$v_{id}(t+1) = v_{id}(t) + K\big[v_{id}(t) + c_1(p_{id}(t) - x_{id}(t))R_1 + c_2(g_{id}(t) - x_{id}(t))R_2\big] \quad (5.2.7)$$

From empirical experience, we figured that the best version for our purpose is the a slightly modified version based on the hybrid binary PSO algorithm introduced in [84] because of its fast convergence speed and capability of reaching global optima. Both mutation from Genetic algorithm and cross over process back to previous personal best position are included, and the detailed procedure is shown in the pseudo code.

Algorithm 2 Binary Particle Swarm Optimization with Mutation and Crossover

Input: Particles population size N, dimension of the problem D, maximum allowed velocity V_{max}, velocity upper bound v_{max} for initialization, mutation probability p_m, cross over probability p_c, inertia weight w and weights c_1, c_2, formula $F(\mathbf{x}_i)$ that returns the value of the objective function based on the position $\mathbf{x}_i$, total number of iterations T and the stopping criteria ϵ.

Initialize $x_{id}(0) \overset{iid}{\sim} Bernoulli(p_i), i = 1, \cdots, N, d = 1, \cdots, D$, where $p_i = i/N$
$\mathbf{P} \leftarrow \mathbf{X}$, $\mathbf{g} \leftarrow ArgMax(\mathbf{X}, F)$, where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]$, and $\mathbf{P} = [\mathbf{p}_1, \cdots, \mathbf{p}_N]$
Initialize $v_{id}(0) \overset{iid}{\sim} Uniform(-v_{max}, v_{max}), i = 1, \cdots, N, d = 1, \cdots, D$
$t \leftarrow 0$
while $t < T$ and $F(\mathbf{g}) < \epsilon$ **do**
    for $i \leftarrow 1, \cdots, N$ **do**
        $\mathbf{R}_1[d,d] \overset{iid}{\sim} Uniform(0,1), \mathbf{R}_2[d,d] \overset{iid}{\sim} Uniform(0,1), d = 1, \cdots, D$
        $\mathbf{v}_i(t+1) \leftarrow w(t)\mathbf{v}_i(t) + c_1(\mathbf{p}_i(t) - \mathbf{x}_i(t))\mathbf{R}_1 + c_2(\mathbf{g}_i(t) - \mathbf{x}_i(t))\mathbf{R}_2$
        for $d \leftarrow 1, \cdots D$ **do**
            $v_{id}(t+1) \leftarrow min(V_{max}, max(-V_{max}, v_{id}(t+1)))$
            $u \leftarrow Uniform(0,1)$, $x_{id}(t+1) = bool\big(\frac{1}{1+e^{-v_{id}(t+1)}} > u\big)$
            $u_m \leftarrow Uniform(0,1)$, and $x_{id}(t+1) = \big|x_{id}(t+1) - 1\big|\big)$ if $u_m > p_m$
            $u_c \leftarrow Uniform(0,1)$, and $x_{id}(t+1) = p_{id}(t)\big)$ if $u_c > p_c$
        end for
        if $F(\mathbf{x}_i(t+1)) > F(\mathbf{p}_i(t))$ **then**
            $\mathbf{p}_i(t+1) \leftarrow \mathbf{x}_i(t+1))$
        else
            $\mathbf{p}_i(t+1) \leftarrow \mathbf{p}_i(t)$
        end if
    end for
    $\mathbf{g} \leftarrow ArgMax(\mathbf{P}, F)$
    $t \leftarrow t + 1$
end while
return $\mathbf{g}, F(\mathbf{g})$

To investigate the effectiveness of the algorithm in solving general combinatorial optimization problems, we apply the algorithm on the following set of test functions

following [70, 63] ,

$$
\begin{aligned}
&f_1 = \sum_{i=1}^{3} x_i^2,\ x_i \in [-5.12, 5.12],\ x_i^{(opt)} = 0 \\
&f_2 = 100(x_2 - x_1^2)^2 + 100(x_3 - x_2^2)^2 + (1 - x_1)^2 + (1 - x_2)^2,\ x_i \in [-5.12, 5.12],\ x_i^{(opt)} = 1 \\
&f_3 = 6 \times 3 + \sum_{i=1}^{3} \lfloor x_i \rfloor,\ x_i \in [-5.12, 5.12],\ x_i^{(opt)} \in [-5.12, -5) \\
&f_4 = \sum_{i=1}^{3} \big(x_i^2 - 10cos\big(2\pi x_i\big) + 10\big)\ x_i \in [-5.12, 5.12],\ x_i^{(opt)} = 0 \\
&f_5 = 0.5 + \frac{sin^2\big(\sqrt{x_1^2 + x_2^2 + x_3^2}\big) - 0.5}{\big[1 + 0.001(x_1^2 + x_2^2 + x_3^2)\big]^2}\ x_i \in [-51.2, 51.2],\ x_i^{(opt)} = 0 \\
&f_6 = \frac{1}{4000} \sum_{i=1}^{3} (x_i - 100)^2 - \prod_{i=1}^{3} cos\big(\frac{x_i - 100}{\sqrt{i}}\big) + 1,\ x_i \in [-512, 512],\ x_i^{(opt)} = 100
\end{aligned}
\tag{5.2.8}
$$

The first part is the function itself, second part is the search range we used, third part is the value of x_i in each dimension to achieve optimal value of objective function, which is $f_{min} = 0$ for all cases. These test functions were originally in continuous form and designed for testing continuous optimization algorithms. But we can convert floating point numbers into binary and still make use of these test functions for discrete case. For example, $x \in (-5.12, 5.12)$ can be converted into a 10 bits binary number $x^b = [x_1^b, \cdots, x_{10}^b]$ such that $x = \big(\sum_{i=1}^{10} 2^{x_i^b(10-i)} - 512\big)/100$. For the optimization process, we used the following set of parameters: $N = 50, D = 30, V_{max} = 6, v_{max} = 0.6, p_m = 0.02, p_c = 0.1, T = 100, w = 1, c_1 = c_2 = 2$. Then we run 100 realizations of the algorithm on all of the 6 functions and record the percentage number of realizations that could reach different levels of accuracy (defined as when the value of the objective function for global optimal is no larger than the error level) versus iteration. We can see that after 50 iterations, the true global optimal when objective function value is zero can be achieved for a small percentage of the realizations for most of the functions. Since for our purpose, we only need to find the optimal configuration as well as the corresponding value of the objective function once, we can

run the optimization algorithm several times, say M times, and each time set T not large, and choose the best from these several realizations. In practice for our usage, we set $M = 20, T = 50$.

For illustrative purpose, we plot one dimensional case for the three of the test functions f_4, f_5, f_6 as following,

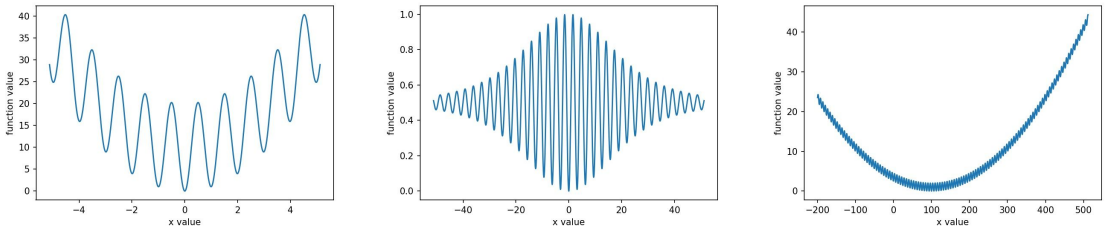


Figure 5.1: Left to right: (a), f_4, Rastrigin; (b), f_5, Schaffer's F6; (c), f_6, Griewank; 1 Dimensional plot of three representative test functions

We can see that the test functions are so complicated such that it is very easy to get trapped in local optima, but the optimization results are actually pretty good as shown bellow. Therefore, we are confident that we can apply the algorithm for our purpose following the above strategy.

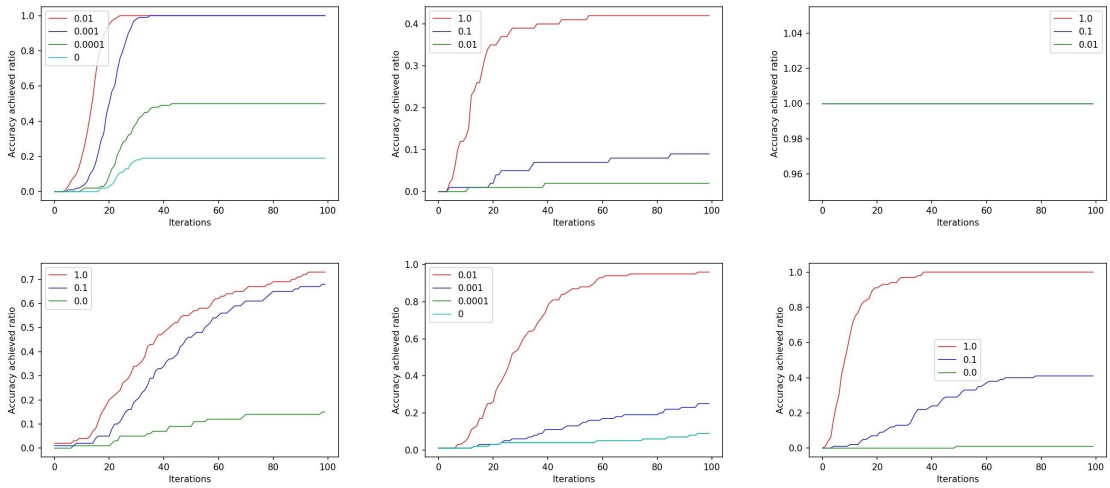


Figure 5.2: Upper row from left to right: (a), f_1, Spherical; (b), f_2, Rosenbrock; (c), f_3, Step. Lower row from left to right: (d), f_4, Rastrigin; (e), f_5, Schaffer's F6; (f), f_6, Griewank. Percentage of realizations reaching different level of accuracy versus iterations

5.3 Application of PSO on information diffusion

In this section, we apply the above hybrid binary PSO algorithm on seed nodes selection for information diffusion maximization purpose.

5.3.1 Experimental settings

The information diffusion process is carried out on three network settings, single layered network A with community structure generated by standard stochastic block model[30, 59, 42], single layered network B generated as a degree corrected scale free network[129, 133], and multilayered networks realized by coupling network A and B.

Network A is generated as of size $N = 1000$ with 5 communities, with size 400, 200, 200, 100 and 100. The probability of connections between nodes from the same community is $p_{in} = 0.02$ and between nodes from different communities is $p_{out} = 0.0002$ such that the degree distribution of community 1 is $P(k) \sim Poisson(8)$, for community 2 and 3 is $P(k) \sim Poisson(4)$ and for community 4 and 5 is $P(k) \sim$

$Poisson(2)$, total number of connections between community 1 and 2(3) is distributed as $P(M_{12}) \sim Poisson(16)$, between 1 and 4(5) is $P(M_{14}) \sim Poisson(8)$, between 2 and 3 is $P(M_{23}) \sim Poisson(8)$, between 2 and 4(5) is $P(M_{24}) \sim Poisson(4)$ and between 4 and 5 is $P(M_{45}) \sim Poisson(2)$. Network B is also of size $N = 1000$, with degree distribution $P(k) \sim k^{-\lambda}$, where $\lambda = 2.9$ and the average degree is set to be $\langle k \rangle = 5$.

To model the information diffusion process we use the susceptible-informed-immunized(SIM) model introduced in Chapter 3 with diffusion rate β_0 and immunization rate γ. Assume we start the diffusion process before immunization in a discrete time case, the probability of a node $k > 1$ hops from the seed node to finally get informed is

$$p_k = \frac{\beta_1}{1-(1-\gamma)^{k+1}(1-\beta_1)} \prod_{i=2}^{k} \frac{\beta_i(1-\gamma)^{k+2-i}}{1-(1-\gamma)^{k+2-i}(1-\beta_i)} \tag{5.3.1}$$

And for $k = 1$, we simply have:

$$p_1 = \frac{\beta_1}{1-(1-\gamma)^2(1-\beta_1)} \tag{5.3.2}$$

where in general $\beta_{ij} = 1-(1-\beta_0)^{\sum_{k=1}^{K} W_{ij}^{(k)}}$ with $W_{ij}^{(k)}$ being the weight of the k_{th} edge between node i and j, and for simplicity in notation we denote β_{ij} by β_i. As discussed before, we set the diffusion rate $\beta_0 = 0.05$ for within communities, and $\beta_0 = 0.005$ for diffusion both between communities on primary layer and on secondary layer.

To facilitate identifying a set of candidate seed nodes, we first study the structure of both MWT and saturated diffusion size F based on degrees of nodes for both layer A and B. In layer A, nodes split into 3 groups in terms of MWT and F such that community 1 forms one group, community 2 and 3 forms second group and community 4 and 5 forms the third group. Similarly, in layer B nodes naturally split into three clusters in terms of MWT vs degree, and it turns out that these three

clusters represent three groups of nodes with distinct distance from the hub node of degree $K = 364$, such that in cluster 1 all nodes are one hop from the hub, in cluster 2 all nodes are two hops from the hub and in cluster 3 all nodes are three or more hops from the hub.

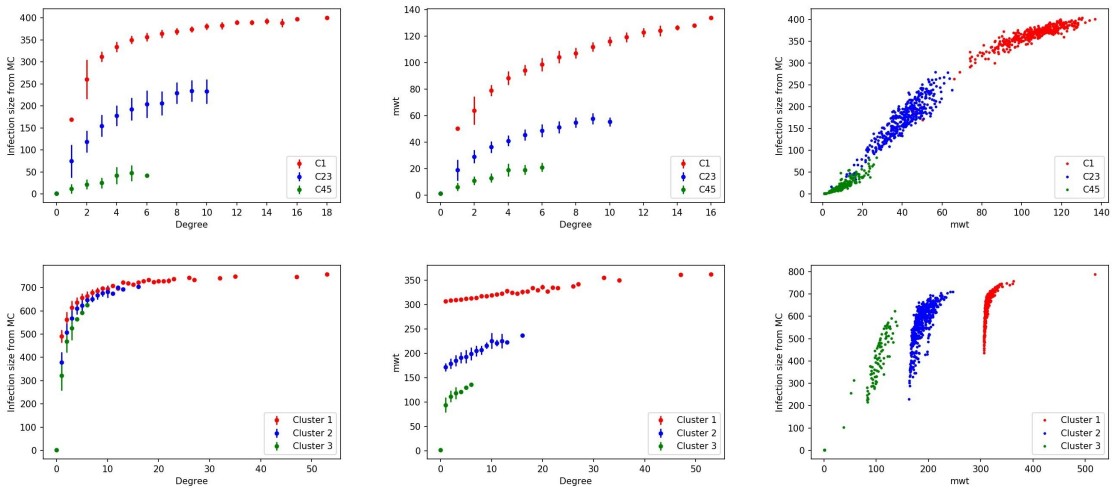


Figure 5.3: Upper row: (a-c), Layer A. Lower row: (d-f), Layer B. Structure of MWT and F vs degree for both network layer A and B

As can be seen in layer A both MWT and F are well separated for the three groups with larger group members producing both much higher MWT and F and the degree does not contribute very much. On the other hand, for layer B only MWT for the three clusters are well separated for the three groups and similarly degree does not have much contribution. On the contrary, for F the three groups are not so separated and F increases dramatically with degree especially when degrees are small.

In particle swarm optimization, as the dimension of the problem increases accuracy tends to decrease[70, 80], so we do not want to include all nodes as possible candidates, in which case the dimension is $D = 1000$. Instead, we choose a small subset $D = 30$ representative nodes from the whole network and search from these 30 nodes the optimal set of seed nodes under different circumstances to achieve maximum profit. To ensure that the 30 nodes are as representative as possible, in both layer A and B

we distribute the 30 nodes evenly in the three clusters with number of candidate seed nodes in each group proportional to the size of that group. Besides, within each group we select nodes over the whole range of degrees so as to cover as many degrees as possible, and for each degree set, we want to choose the one node with modest MWT and F. The candidate seed nodes selection process is illustrated by the community 1 in layer A as shown in the following, with green stars representing the nodes being selected as the candidate seed nodes.

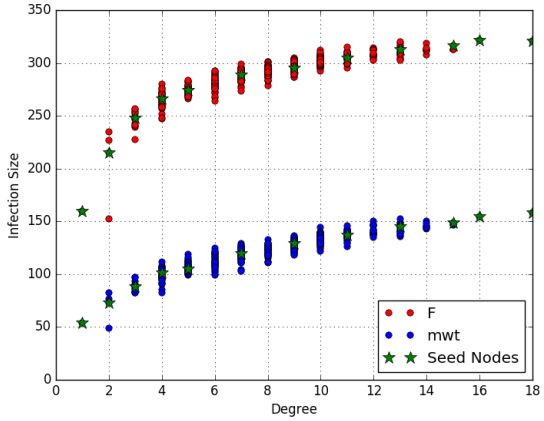


Figure 5.4: Seed nodes selection process for community 1 in layer A

5.3.2 Experiments I: Primary layer A only

In this section, we study the information diffusion maximization process on primary layer A only. In the tables shown in Figure 5.5, the column Cm represents the community the candidate seed nodes belong to, the column K represents the degree of connectivity of the candidate seed nodes in layer A, column MWT represents the MWT value and column F represents the saturated diffusion size if we start from the single one candidate seed node. The row PT represents the maximum total profit under the experimental setting obtained by PSO, and Pow represents the exponent Q in the definition of cost function.

In each experiment, for the finally identified optimal set of seed nodes by PSO,

1 indicates being included and 0 indicates otherwise, and for visual guidance, we colored 1s in black. The red and pink cells in each community for each experiment is achieved by MPP. Intuitively, if we want to maximize total profit, we would like to select at least one seed node from each community and the the red and pink cells represents this strategy when we choose one single seed node from each community and if the optimal node is also selected by PSO, it will be colored in red, otherwise, it will be pink.

Comparing the experimental results in table (a) and (c) in Figure 5.5 we have the following 3 main findings. (1) More seed nodes are included in the identified set by PSO in (a) than in (c) in the whole range values of W, and they form upper left triangle in (a) and V shape in (c). (2) Agreement on seed nodes selection by PSO and by MPP method for large W in (a), and agreement in (c) as long as W is not too small. (3) At least one seed node from each community is selected except when W is too large. To understand these findings, we can roughly replicate the structures by thinking of the searching process for the optimal set of seed nodes in each community in the first step as a competition in terms of total profit $PT = F_{tot} - cost_{tot}$ between the single one best seed node obtained based on MPP mechanism and a group of low degree nodes available. And then extra seed nodes will be added to the winner set step by step based on the same competition mechanism in terms of marginal profit $\Delta PT = \Delta F_{tot} - \Delta(cost)$ until $\Delta PT \leq 0$, where ΔF_{tot} is marginal revenue, defined as the extra diffusion size achieved by including the extra set of seed nodes, and $\Delta(cost)$ is the marginal cost, defined as the extra cost of including the extra set of seed nodes.

For illustrative purpose, in the following we use community 1 as an example and we will notice that the distinctive structure of seed nodes location for (a) and (c) comes from the higher homogeneity of values of MWT compared to degree K across nodes. In community 1, for example, the range of MWT is $(53.9, 158.2)$ with the maximum about 3 times of the minimum while for K the range is $(1, 18)$ and the

maximum is 18 times of the minimum. Suppose at step 1, we take two nodes of low degree $K = 2, 3$ as a candidate for original set of low degree seed nodes and compete with the node with degree $K = 18$ chosen by MPP mechanism when $W \leq 2.0$. If we assume that the diffusion processes starting from the two seed nodes are independent, when we start with the two low degree nodes the final diffusion size is approximately $F_{tot}^{(2,3)} \sim 400\big[1-(1-226.5/400)(1-254.6/400)\big] = 336.9$, which is close to the diffusion size when we start from the high degree node $F^{(18)} = 320.6$.

In (a) the objective function we want to maximize is profit $PT = F_{tot} - K_{tot} * W = F_{tot}\big(1 - \frac{K_{tot}}{F_{tot}}W\big)$, where F_{tot} is the saturated total diffusion size and K_{tot} is the sum of degrees of all seed nodes. For step 1, substitute the above two candidates into the objective function under the same W and assume an equal F_{tot}, then for the first one $K_{tot}^{(2,3)} = 5 \Rightarrow PT^{(2,3)} = F_{tot}\big(1 - \frac{5}{F_{tot}}W\big)$ and for the second one $K^{(18)} = 18 \Rightarrow PT^{(18)} = F_{tot}\big(1 - \frac{18}{F_{tot}}W\big)$. Obviously, the first one wins and we choose the set with two low degree seed nodes as the original optimal set. Of course, there might be other combination of low degree seed nodes that would produce higher PT, but still, this kind of candidates win. Then we add extra seed nodes step by step based on the same competition mechanism but in terms of the marginal profit $\Delta PT = \Delta F_{tot} - \Delta K_{tot} * W$. Following the same analysis, we would tend to add low degree nodes before high degree nodes. To investigate if the high degree node selected based on MPP will be included, we look at the difference in marginal profit between two seed nodes of different degrees $K_1 < K_2$

$$\Delta PT_2 - \Delta PT_1 = \big(\Delta F_{tot}^{(2)} - \Delta F_{tot}^{(1)}\big) - \big(\Delta K_{tot}^{(2)} - \Delta K_{tot}^{(1)}\big) * W = \big(\Delta F_{tot}^{(2)} - \Delta F_{tot}^{(1)}\big) - \big(K_2 - K_1\big) * W$$

With the same set of seed nodes, the difference in marginal revenue $\Delta F_{tot} = \Delta F_{tot}^{(2)} - \Delta F_{tot}^{(1)}$ between adding an extra node 2 versus 1 is constant, but the difference in marginal cost $\big(K_2 - K_1\big) * W$ will be negligible in the range $K \in [1, 18]$ if we have

extremely small W and therefore the high degree seed node will be included in sacrifice of some low degree seed nodes. On the other hand, as W increases, differences in marginal cost might exceed differences in marginal revenue such that high degree seed nodes are never included. However as W increases further the optimal one single seed node selected based on MPP might become so low in degree such that it overlaps with low degree seed nodes selected in step 1.

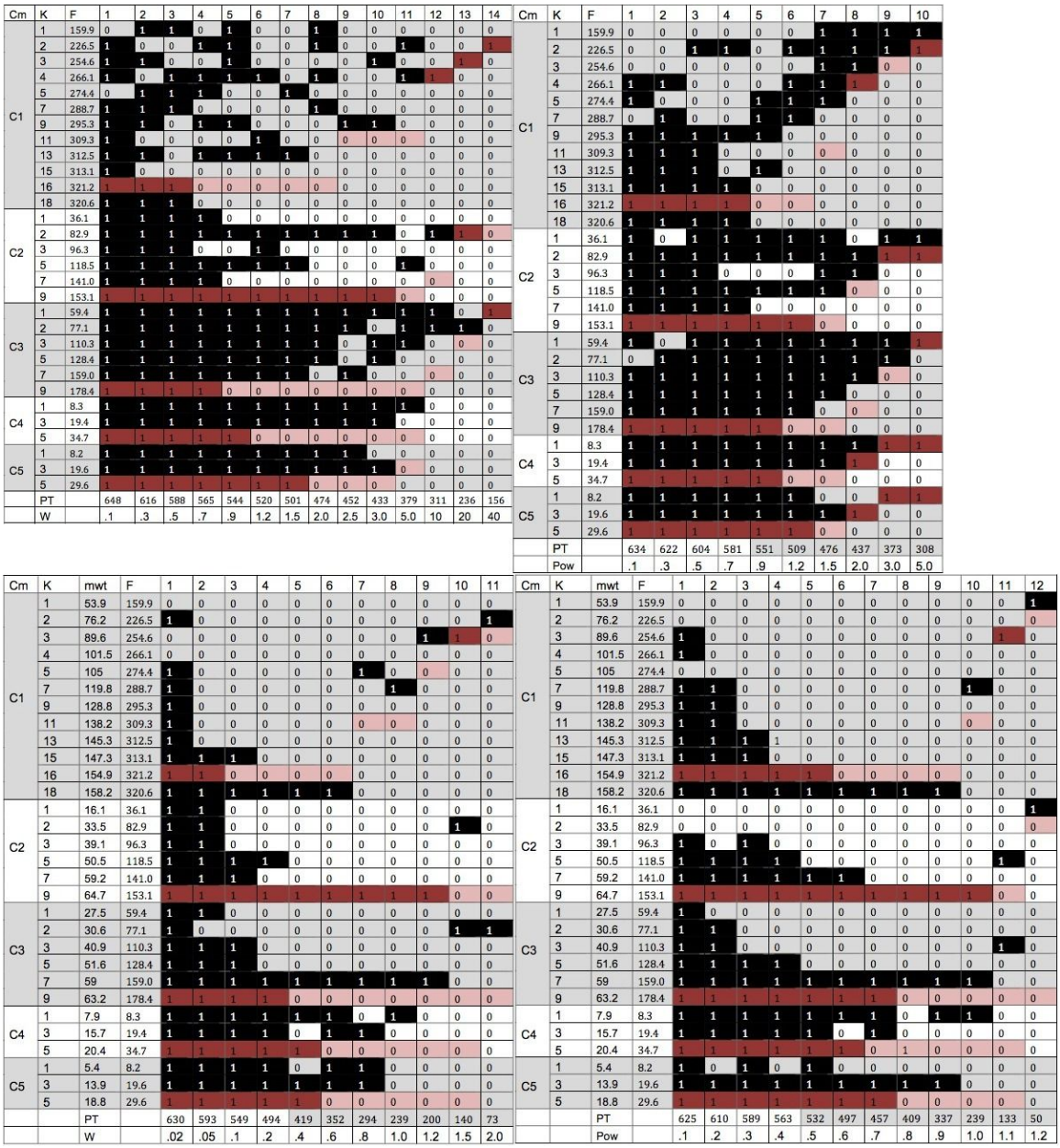
Similarly the objective function we want to maximize in (c) is $F_{tot} - MWT_{tot} * W = F_{tot}\big(1 - \frac{MWT_{tot}}{F_{tot}}W\big)$, following the same analysis, $MWT_{tot}^{(2,3)} = 76.2 + 89.6 = 165.8 \Rightarrow PT^{(2,3)} = F_{tot}\big(1 - \frac{165.8}{F_{tot}}W\big)$ while $MWT^{(18)} = 158.2 \Rightarrow PT^{(18)} = F_{tot}\big(1 - \frac{158.2}{F_{tot}}W\big)$. So the second one wins and we choose the one high degree node selected based on personal profit as the original optimal seed node. Then we add extra seed nodes based on the same competition mechanism but in terms of the marginal profit $\Delta PT = \Delta F_{tot} - \Delta MWT_{tot} * W$. Following the same analysis, we would tend to add high degree seed nodes before low degree nodes as long as $\Delta PT > 0$. In the region when W is extremely small, the marginal cost $\Delta MWT_{tot} * W$ would be negligible compared to ΔF_{tot} resulting in a positive ΔPT and therefore, the seed nodes would build up from high to low degrees, while when W becomes larger, it becomes unprofitable even to add one extra seed node and thus only the one single seed node selected based on personal profit is placed. This explains why the seed nodes number selected tends to be smaller in (c) compared to (a) in most values of W and better agreement between the set of seed nodes returned by PSO and by MPP is observed in (c).

As for finding (3), because of the extremely limited diffusion between communities, including one extra group of nodes from each community would result in a positive marginal profit ΔPT for most of the cases, except when $F_i - cost_i < 0$ for all seed nodes in the community in which case even the step 1 would return no seed nodes as any combination would produce negative profit. Here F_i is diffusion size if we set node i as the single one seed node, and $cost_i$ is the cost of doing so. This case is

represented when there is no red or pink cells in the community and we can observe pretty good agreement between the values of W or Q where pink/red disappears and where entry 1 disappears. For example in (a), pink cells disappear after $W = 5.0$ in community 5 and the entry 1 disappears only one step earlier and for community 4 both disappear after $W = 5.0$.

To sum up, in (a) we build up seed nodes in each community from low degree seed nodes up to high degree nodes and whether the high degree nodes will be included or not depends on the weight W. On the other hand, in (c) we build up the seed nodes in each community from high degree to low degree and good agreement between the set of seed nodes returned by PSO and by MPP is achieved if W is not too small.

For (b) and (d) we would expect similar phenomena, except in (b): (1) when Q is small, say $Q \leq 0.5$, we would expect fewer low degree seed nodes and (2) when Q is large, say $Q \geq 2.0$, we would expect more low degree nodes to be included. Both the two phenomena can be explained by the property of power function. Suppose we have two seed nodes and one's degree is 4 times the other, then under condition (1) $K_2/K_1 = 4 \Rightarrow (K_2/K_1)^{0.5} = 2$ cost only becomes 2 times and one extreme case is $K = 1 \Rightarrow cost = K^{0.5} = 1$ not negligible anymore. Under condition (2), similarly, $K_2/K_1 = 4 \Rightarrow (K_2/K_1)^2 = 16$ the cost becomes 16 times and one extreme case is $K = 1 \Rightarrow cost = K^4 = 1$. ΔPT would be positive in this case if we include the low degree nodes since fewer seed nodes are already included and including one extra seed node result in a higher ΔF_{tot} compared to when there are already many seed nodes selected as under condition (1).



Cm	K	F	1	2	3	4	5	6	7	8	9	10	11	12	13	14
C1	1	159.9	0	1	1	0	1	0	0	1	0	0	0	0	0	0
	2	226.5	1	0	0	1	1	0	0	1	0	0	1	0	0	1
	3	254.6	1	1	0	0	1	0	0	0	0	1	0	0	1	0
	4	266.1	1	0	1	1	1	1	0	1	0	0	1	1	0	0
	5	274.4	0	1	1	1	0	0	1	0	0	0	0	0	0	0
	7	288.7	1	1	1	0	0	0	0	1	0	0	0	0	0	0
	9	295.3	1	1	0	1	1	0	0	0	1	1	0	0	0	0
	11	309.3	1	0	0	0	0	1	0	0	0	0	0	0	0	0
	13	312.5	1	1	0	1	1	1	1	0	0	0	0	0	0	0
	15	313.1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	16	321.2	1	1	1	0	0	0	0	0	0	0	0	0	0	0
	18	320.6	1	1	1	0	0	0	0	0	0	0	0	0	0	0
C2	1	36.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
	2	82.9	1	1	1	1	1	1	1	1	1	0	1	1	0	
	3	96.3	1	1	1	0	0	1	0	0	0	0	0	0	0	
	5	118.5	1	1	1	1	1	1	1	0	0	0	1	0	0	
	7	141.0	1	1	1	1	0	0	0	0	0	0	0	0	0	
	9	153.1	1	1	1	1	1	1	1	1	1	1	0	0	0	
C3	1	59.4	1	1	1	1	1	1	1	1	1	1	1	1	0	1
	2	77.1	1	1	1	1	1	1	1	1	1	0	1	1	1	0
	3	110.3	1	1	1	1	1	1	1	1	0	1	1	0	0	0
	5	128.4	1	1	1	1	1	1	1	1	0	1	0	0	0	0
	7	159.0	1	1	1	1	1	1	1	1	0	0	0	0	0	0
	9	178.4	1	1	1	1	0	0	0	0	0	0	0	0	0	0
C4	1	8.3	1	1	1	1	1	1	1	1	1	1	1	0	0	0
	3	19.4	1	1	1	1	1	1	1	1	1	1	0	0	0	0
	5	34.7	1	1	1	1	1	0	0	0	0	0	0	0	0	0
C5	1	8.2	1	1	1	1	1	1	1	1	1	1	0	0	0	0
	3	19.6	1	1	1	1	1	1	1	1	1	1	0	0	0	0
	5	29.6	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	PT		648	616	588	565	544	520	501	474	452	433	379	311	236	156
	W		.1	.3	.5	.7	.9	1.2	1.5	2.0	2.5	3.0	5.0	10	20	40

Cm	K	F	1	2	3	4	5	6	7	8	9	10
C1	1	159.9	0	0	0	0	0	0	1	1	1	1
	2	226.5	0	0	1	1	0	1	1	1	1	1
	3	254.6	0	0	0	0	0	0	1	1	0	0
	4	266.1	1	1	0	0	0	1	1	1	0	0
	5	274.4	1	0	0	0	1	1	1	0	0	0
	7	288.7	0	1	0	0	1	1	0	0	0	0
	9	295.3	1	1	1	1	1	0	0	0	0	0
	11	309.3	1	1	1	0	0	0	0	0	0	0
	13	312.5	1	1	1	0	1	0	0	0	0	0
	15	313.1	1	1	1	1	0	0	0	0	0	0
	16	321.2	1	1	1	1	0	0	0	0	0	0
	18	320.6	1	1	1	1	0	0	0	0	0	0
C2	1	36.1	1	0	1	1	1	1	1	0	1	1
	2	82.9	1	1	1	1	1	1	1	1	1	1
	3	96.3	1	1	1	0	0	0	1	1	0	0
	5	118.5	1	1	1	1	1	1	1	0	0	0
	7	141.0	1	1	1	1	0	0	0	0	0	0
	9	153.1	1	1	1	1	1	1	0	0	0	0
C3	1	59.4	1	0	1	1	1	1	1	1	1	1
	2	77.1	0	1	1	1	1	1	1	1	1	0
	3	110.3	1	1	1	1	1	1	1	1	0	0
	5	128.4	1	1	1	1	1	1	1	0	0	0
	7	159.0	1	1	1	1	1	1	0	0	0	0
	9	178.4	1	1	1	1	1	0	0	0	0	0
C4	1	8.3	1	1	1	1	1	1	1	1	1	1
	3	19.4	1	1	1	1	1	1	1	1	1	0
	5	34.7	1	1	1	1	1	1	0	0	0	0
C5	1	8.2	1	1	1	1	1	1	0	0	1	1
	3	19.6	1	1	1	1	1	1	1	1	0	0
	5	29.6	1	1	1	1	1	1	0	0	0	0
	PT		634	622	604	581	551	509	476	437	373	308
	Pow		.1	.3	.5	.7	.9	1.2	1.5	2.0	3.0	5.0

Cm	K	mwt	F	1	2	3	4	5	6	7	8	9	10	11
C1	1	53.9	159.9	0	0	0	0	0	0	0	0	0	0	0
	2	76.2	226.5	1	0	0	0	0	0	0	0	0	0	1
	3	89.6	254.6	0	0	0	0	0	0	0	0	1	1	0
	4	101.5	266.1	0	0	0	0	0	0	0	0	0	0	0
	5	105	274.4	1	0	0	0	0	0	1	0	0	0	0
	7	119.8	288.7	1	0	0	0	0	0	0	1	0	0	0
	9	128.8	295.3	1	0	0	0	0	0	0	0	0	0	0
	11	138.2	309.3	1	0	0	0	0	0	0	0	0	0	0
	13	145.3	312.5	1	0	0	0	0	0	0	0	0	0	0
	15	147.3	313.1	1	1	1	0	0	0	0	0	0	0	0
	16	154.9	321.2	1	1	0	0	0	0	0	0	0	0	0
	18	158.2	320.6	1	1	1	1	1	1	0	0	0	0	0
C2	1	16.1	36.1	1	1	0	0	0	0	0	0	0	0	0
	2	33.5	82.9	1	1	0	0	0	0	0	0	0	1	0
	3	39.1	96.3	1	1	0	0	0	0	0	0	0	0	0
	5	50.5	118.5	1	1	1	1	0	0	0	0	0	0	0
	7	59.2	141.0	1	1	1	0	0	0	0	0	0	0	0
	9	64.7	153.1	1	1	1	1	1	1	1	1	1	0	0
C3	1	27.5	59.4	1	1	0	0	0	0	0	0	0	0	0
	2	30.6	77.1	1	0	0	0	0	0	0	0	0	1	1
	3	40.9	110.3	1	1	1	0	0	0	0	0	0	0	0
	5	51.6	128.4	1	1	1	0	0	0	0	0	0	0	0
	7	59	159.0	1	1	1	1	1	1	1	1	1	0	0
	9	63.2	178.4	1	1	1	1	0	0	0	0	0	0	0
C4	1	7.9	8.3	1	1	1	1	1	1	0	1	0	0	0
	3	15.7	19.4	1	1	1	1	0	1	1	0	0	0	0
	5	20.4	34.7	1	1	1	1	1	0	0	0	0	0	0
C5	1	5.4	8.2	1	1	1	1	0	1	1	0	0	0	0
	3	13.9	19.6	1	1	1	1	1	1	1	0	0	0	0
	5	18.8	29.6	1	1	1	1	1	0	0	0	0	0	0
		PT		630	593	549	494	419	352	294	239	200	140	73
		W		.02	.05	.1	.2	.4	.6	.8	1.0	1.2	1.5	2.0

Cm	K	mwt	F	1	2	3	4	5	6	7	8	9	10	11	12
C1	1	53.9	159.9	0	0	0	0	0	0	0	0	0	0	0	1
	2	76.2	226.5	0	0	0	0	0	0	0	0	0	0	0	0
	3	89.6	254.6	1	0	0	0	0	0	0	0	0	0	1	0
	4	101.5	266.1	1	0	0	0	0	0	0	0	0	0	0	0
	5	105	274.4	0	0	0	0	0	0	0	0	0	0	0	0
	7	119.8	288.7	1	1	0	0	0	0	0	0	0	1	0	0
	9	128.8	295.3	1	1	0	0	0	0	0	0	0	0	0	0
	11	138.2	309.3	1	1	0	0	0	0	0	0	0	0	0	0
	13	145.3	312.5	1	1	1	1	0	0	0	0	0	0	0	0
	15	147.3	313.1	1	1	1	0	0	0	0	0	0	0	0	0
	16	154.9	321.2	1	1	1	1	1	0	0	0	0	0	0	0
	18	158.2	320.6	1	1	1	1	1	1	1	1	1	0	0	0
C2	1	16.1	36.1	0	0	0	0	0	0	0	0	0	0	0	1
	2	33.5	82.9	0	0	0	0	0	0	0	0	0	0	0	0
	3	39.1	96.3	1	0	1	0	0	0	0	0	0	0	0	0
	5	50.5	118.5	1	1	1	1	0	0	0	0	0	0	1	0
	7	59.2	141.0	1	1	1	1	1	1	0	0	0	0	0	0
	9	64.7	153.1	1	1	1	1	1	1	1	1	1	1	0	0
C3	1	27.5	59.4	1	0	0	0	0	0	0	0	0	0	0	0
	2	30.6	77.1	1	1	0	0	0	0	0	0	0	0	0	0
	3	40.9	110.3	1	1	0	0	0	0	0	0	0	0	1	0
	5	51.6	128.4	1	1	1	1	0	0	0	0	0	0	0	0
	7	59	159.0	1	1	1	1	1	1	1	1	1	0	0	
	9	63.2	178.4	1	1	1	1	1	1	1	0	0	0	0	0
C4	1	7.9	8.3	1	1	1	1	1	1	1	0	1	1	0	0
	3	15.7	19.4	1	1	1	1	1	0	1	0	0	0	0	0
	5	20.4	34.7	1	1	1	1	1	1	0	1	0	0	0	0
C5	1	5.4	8.2	1	0	1	0	1	0	0	0	0	0	0	0
	3	13.9	19.6	1	1	1	1	1	1	1	1	1	0	0	0
	5	18.8	29.6	1	1	1	1	1	1	1	0	0	0	0	0
		PT		625	610	589	563	532	497	457	409	337	239	133	50
		Pow		.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0	1.1	1.2

Figure 5.5: Upper left: (a), cost = $K * W$; Upper right: (b), cost = K^Q; Lower left: (c), cost = $MWT * W$; Lower right: (d), cost = MWT^Q. Cost based on degree K and MWT for layer A

5.3.3 Experiments II: Secondary layer B only

In this section we investigate the diffusion maximization process on secondary layer B with all settings the same as above except stated otherwise. Similar to in

layer A we put candidate seed nodes in all three clusters and in the tables shown in Figure 5.6 red/pink cells represent the optimal seed node among candidate seed nodes within each cluster based on MPP while the blue cells represent the global optimal seed node among all the candidate seed nodes across the three clusters by MPP. As before, red and dark blue indicate that the optimal node is also included in the optimal set of seed nodes identified by PSO while pink and light blue indicate otherwise.

Similarly, comparing (a) and (c) we have 3 main findings. (1) More seed nodes are included in the identified set by PSO in (a) than in (c) in the whole range values of W, and they form upper left triangle in (a) and V shape in (c). (2) Agreement on seed nodes selection by PSO and by MPP in (c) as long as W is not too small. (3) For most of the cases, there is no necessity of placing at least one node in each cluster and we can basically treat all nodes the same way as if they are all in one community.

For illustrative purpose, we first focus on cluster 1. Following the same analysis as above we can roughly replicate the configurations by treating the searching process as a competition in terms of profit PT between one optimal single seed node based on personal profit and a group of low degree nodes available for the first step and then step by step adding extra seed nodes based on competition in terms of marginal profit ΔPT. It turns out that the values of $MWT \in (356.6, 605.4)$, where maximum is less than 2 times of minimum, are much more homogeneous than the degrees $K \in (1, 364)$, where maximum is 364 times of minimum, across nodes compared to layer A, which is reasonable since layer A was generated by standard SBM model while layer B was generated as a scale free network.

Following the same procedure, take the two seed nodes of degree $K = 7, 9$ as an example and compete with the optimal one seed node of degree $K = 364$ for most values of W. Assume that the information diffusion process from the two seed nodes are independent, then the total revenue when we start from both of them

is approximately $F_{tot}^{(7,9)} \sim 1000(1-(1-610.6/1000)(617.2/1000)) \sim 851$ which is comparable with the personal revenue of the one hub node $F^{(364)} = 780.2$. Then, we assume the same revenue F_{tot} for the two, and under the same W. In (a) $PT^{(7,9)} = F_{tot}^{(7,9)} - K_{tot}^{(7,9)} * W = F_{tot}^{(7,9)}\big(1 - \frac{K_{tot}^{(7,9)}}{F_{tot}^{(7,9)}} * W\big) \sim F_{tot}\big(1 - \frac{16}{F_{tot}} * W\big)$ and $PT^{(364)} \sim F_{tot}\big(1 - \frac{364}{F_{tot}} * W\big)$. Obviously the first candidate wins out and following the same analysis, we know that for (a) the seed nodes set build up from low degree ones and the hub node will be included only when W is extremely small when $\Delta PT_{hub} - \Delta PT_{K_{low}}$ is positive for some seed node of low degree K_{low}. The cluster does not make much contribution in this case which is reasonable since if we look at Figure 5.6 (d) we will notice that the $F \sim K$ plot is almost independent of the clusters. It is intuitively correct since we do not have any community structure in layer B and we would expect nodes of the same degree K to be statistically equivalent.

Similarly, in (c) $PT^{(7,9)} = F_{tot}^{(7,9)} - MWT_{tot}^{(7,9)} * W \sim F_{tot}\big(1 - \frac{365.7+369.8}{F_{tot}} * W\big)$ while $PT^{(364)} \sim F_{tot}\big(1 - \frac{605.4}{F_{tot}} * W\big)$. Obviously the second candidate wins out and we choose the optimal one based on MPP and then we add extra seed nodes based on marginal profit. Similarly we see that the set of seed nodes build up from high degree to low degree and when W is not too small perfect agreement between the optimal node returned by PSO and based on MPP is achieved. Note that when W is small some cluster dependent structure appears, that is at least one node from each cluster are included. For example when $W = 0.005$ if we treat all nodes as if they belong to the same community we would not expect nodes of degree smaller than $K = 14$ to be included from cluster 1 but one node of degree $K = 5$ is included from cluster 3. Another example is when $W = 0.01$ from cluster 1 we would not expect any node of degree $K < 22$ to be included but one node of degree $K = 16$ from cluster 2 is included. These findings are actually consistent with the community structured property of MWT values and none for F as shown in Figure 5.3 (d-e). Since MWT

of clusters 2 and 3 are much smaller compared to cluster 1

$$\begin{aligned}\Delta PT_{2,3} - \Delta PT_1 &= \left(\Delta F_{2,3} - \Delta F_1\right) - \left(MWT_{2,3} - MWT_1\right) * W \\ &= (MWT_1 - MWT_{2,3}) * W - \left(\Delta F_1 - \Delta F_{2,3}\right) \end{aligned} \tag{5.3.1}$$

When W is small enough such that the last node to be added following the above seed node building up procedure in cluster 1 is with small enough degree such that ΔF_1 and $\Delta F_{2,3}$ are so close that the difference in MWT values can compensate it even multiplied by W.

As for (b) and (d), (d) is similar to (c) while (b) is a little bit different just as in layer A when we expected less low degree nodes when Q is small and more low degree nodes when Q is large. As Q becomes small, difference in cost for two nodes of different degrees K_1, K_2 is much smaller and the cost becomes very homogeneous across nodes, which is very similar to the case MWT. And therefore, for low Q we would expect more similar structure to (c), and it turns out that it is the case in (b) where very steep V shape is visible. On the contrary when Q is large, cost from low degree nodes do not increase as fast as high degree nodes but ΔF_{tot} becomes more significant due to smaller number of seed nodes included and therefore we would expect to see more low degree nodes included, which is the case in (b).

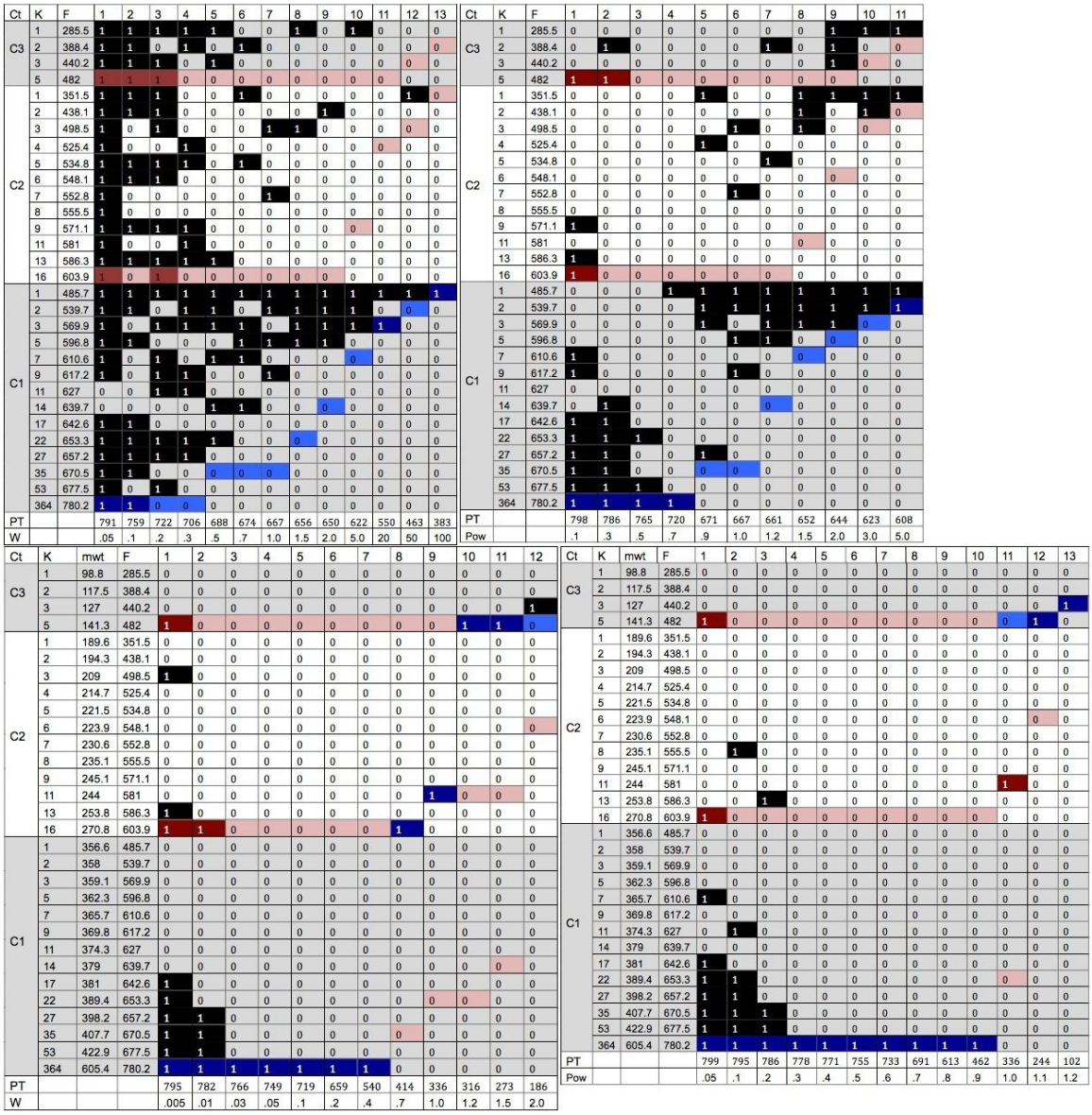


Figure 5.6: Upper left: (a), cost $= K * W$; Upper right: (b), cost $= K^Q$; Lower left: (c), cost $= mwt * W$; Lower right: (d), cost $= mwt^Q$. Cost based on degree K and mwt for layer B

5.3.4 Experiments III: Multilayer A and B

In this section, we couple layer A and layer B together and let the information diffuses on the two layered networks. We keep the candidate nodes position unchanged in layer A but modify the corresponding position of the seed nodes in layer B. For

tables (a-b) shown in Figure 5.7 we put all the candidate seed nodes in cluster 3 in layer B and for tables (c-d) we put all the candidate seed nodes in cluster 1 in layer B. For the two layered networks case, the cost of selecting a certain node i as a seed node is $\big(K_A(i)+K_B(i)\big)*W$ in terms of degree K and $\big(MWT_A(i)+MWT_B(i)\big)*W$ in terms of MWT, where $K_A(i)\big(K_B(i)\big)$ is the degree of connectivity of node i in layer $A(B)$, and $MWT_A(i)\big(MWT_B(i)\big)$ is the value of MWT of node i in layer $A(B)$. Similarly, the total cost is the summation of cost for each seed node.

From tables (a) and (c) in Figure 5.7 we have the following 2 findings. (1) Instead of forming upper left triangle the seed nodes form a more of a V shape. (2) Community structure effect is not as important as in single layer A, that is as W increases not so important to have at least one node from each community.

Phenomenon (1) can be explained from two perspectives. On the one hand, after coupling with the secondary layer B high degree nodes becomes much more effective in diffusing information while on the other hand, after including the secondary layer degree in the cost function the distribution of cost across nodes become more homogeneous and we would expect similar behavior as when we use MWT as cost. As we discussed before, more heterogeneous F combined with more homogeneous cost across nodes would result in V shape, and homogeneous F combined with heterogeneous cost would result in upper left triangle, thus the phenomenon. Phenomenon (2) is because adding one seed node from a new community does not result in an as high marginal profit as before in single layer A. On the one hand, in $\Delta PT = \Delta F_{tot} - \Delta K * W = \Delta F_{tot}\big(1 - \frac{\Delta K}{\Delta F_{tot}} * W\big)$. $\Delta K = \Delta K_A + \Delta K_B$ becomes larger while at the same time ΔF_{tot} is smaller because the contribution from secondary layer would diminish the marginal contribution from including an extra seed node. It is not profitable anymore to include any seed node from the community once $\Delta PT < 0$ and this happens earlier here compared to single layer A case. Note that for single layer A case, $\Delta F_{tot} \sim F$, the marginal revenue is approximately the personal revenue

if it is the only seed node in the community because of the extremely limited diffusion between communities, but here $\Delta F_{tot} < F$ because of the contribution from the secondary layer.

From tables (b) and (d) in Figure 5.7, similarly we have two findings. (1) When W is low, the behavior is pretty similar to the V shape we observed for single layer A. (2) As W increases we do not observe the right wing of V shape in terms of each community and community structure effect disappears.

Phenomenon (1) can be similarly explained by the homogeneous cost function and heterogeneous revenue function F across nodes. Phenomenon (2) can be explained by borrowing the same analysis as above. Adding one seed node from a new community does not result in an as high marginal profit as before in single layer A since in $\Delta PT = \Delta F_{tot} - \Delta MWT * W = \Delta F_{tot}\big(1 - \frac{\Delta MWT}{\Delta F_{tot}} * W\big)$ the factor $\Delta MWT = \Delta MWT_A + \Delta MWT_B$ becomes larger while at the same time ΔF_{tot} becomes smaller because the contribution from secondary layer would diminish the marginal contribution from including one extra seed node. As for the right wing of V shape, if we combine all the communities together we will be able to observe it since as W increases seed nodes disappears from big communities with high degree but remains or reappears in smaller communities with not as high degree.

Cm	KA	KB													
C1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0
	2	1	1	1	0	0	0	0	0	0	0	0	0	0	1
	3	2	1	1	0	0	0	0	0	0	0	0	0	0	0
	4	2	1	1	0	0	0	0	0	0	0	0	0	0	0
	5	1	1	0	0	0	0	0	1	0	0	0	0	1	0
	7	1	1	1	1	0	0	0	0	0	0	1	0	0	0
	9	1	1	1	1	0	0	1	0	0	1	0	1	0	0
	11	2	1	0	0	0	0	0	0	0	0	0	0	0	0
	13	2	1	1	0	1	1	0	0	0	0	0	0	0	0
	15	1	1	1	1	1	0	0	0	0	0	0	0	0	0
	16	3	1	1	0	0	0	0	0	0	0	0	0	0	0
	18	1	1	1	1	1	1	1	1	1	0	0	0	0	0
C2	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0
	2	3	1	0	1	0	1	0	0	0	0	0	0	0	0
	3	1	1	1	1	1	0	1	0	0	0	1	0	0	0
	5	2	1	1	1	1	1	0	0	0	0	0	0	0	0
	7	4	1	1	0	0	0	0	0	0	0	0	0	0	0
	9	1	1	1	1	1	1	1	1	1	1	0	0	0	0
C3	1	3	1	0	1	0	0	0	0	0	0	0	0	0	0
	2	1	1	1	0	0	1	0	0	1	0	0	0	0	0
	3	2	1	1	1	1	1	0	0	0	0	0	0	0	0
	5	1	1	1	1	1	1	1	1	0	1	1	0	0	0
	7	1	1	1	1	1	1	1	1	1	0	0	0	0	0
	9	2	1	1	1	0	0	0	0	0	0	0	0	0	0
C4	1	3	1	1	1	0	0	0	0	0	0	0	0	0	0
	3	1	1	1	1	1	1	1	1	1	0	0	0	0	0
	5	2	1	0	0	0	0	0	0	0	0	0	0	0	0
C5	1	2	1	1	1	0	0	0	0	0	0	0	0	0	0
	3	1	1	1	1	1	1	1	0	1	0	0	0	0	0
	5	2	1	1	1	0	0	0	0	0	0	0	0	0	0
PT			720	676	642	617	599	577	559	537	503	461	404	345	262
W			.1	.3	.5	.7	.9	1.2	1.5	2.0	3.0	5.0	10	20	40

Cm	mwtA	mwtB									
C1	53.9	1.72	0	0	0	0	0	0	0	0	0
	76.2	1.46	0	0	0	0	0	0	0	0	0
	89.6	1.93	0	0	0	0	0	0	0	0	0
	101.5	1.77	0	0	0	0	0	0	0	0	0
	105	1.59	0	0	0	0	0	0	0	0	0
	119.8	1.71	0	0	0	0	0	0	0	0	0
	128.8	1.57	0	0	0	0	0	0	0	0	0
	138.2	2.04	0	0	0	0	0	0	0	0	0
	145.3	2.5	0	0	0	0	0	0	0	0	0
	147.3	1.52	1	0	0	0	0	0	0	0	0
	154.9	2.37	1	1	0	0	0	1	0	0	0
	158.2	1.55	1	1	1	1	1	0	0	0	0
C2	16.1	2.32	1	0	0	0	0	0	0	0	0
	33.5	2.41	1	0	0	0	0	0	0	0	0
	39.1	1.52	1	0	0	0	0	0	0	0	0
	50.5	2.5	1	1	0	0	0	0	0	0	0
	59.2	2.68	1	0	0	0	0	0	0	0	0
	64.7	1.39	1	1	1	1	0	0	1	1	1
C3	27.5	2.35	1	0	0	0	0	0	0	0	0
	30.6	1.58	0	0	0	0	0	0	0	0	0
	40.9	1.72	1	1	0	0	0	0	0	0	0
	51.6	1.71	1	1	0	0	0	0	0	0	0
	59	1.83	1	1	1	0	0	0	0	0	0
	63.2	1.93	1	1	0	0	0	0	0	0	0
C4	7.9	2.39	1	1	0	0	0	0	1	0	1
	15.7	1.5	1	0	1	1	0	0	0	0	0
	20.4	2.2	1	1	0	0	0	0	0	0	0
C5	5.4	2.03	1	1	1	0	0	0	0	0	0
	13.9	1.8	1	1	0	0	1	0	1	1	0
	18.8	1.79	1	1	1	0	0	0	0	0	0
PT			664	618	530	472	414	382	361	334	298
W			.05	.1	.3	.5	.8	1.0	1.2	1.5	2.0

Cm	KA	KB													
C1	1	19	1	0	0	0	0	0	0	0	0	0	0	0	0
	2	7	1	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	1	0	0	0	0	0	0	0	0	0	0	0	1
	4	7	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	2	1	1	1	1	1	1	0	0	0	0	0	1	0
	7	2	1	0	0	1	1	1	1	0	0	1	1	0	0
	9	6	1	1	0	0	0	0	0	0	0	0	0	0	0
	11	6	1	0	0	0	0	0	0	0	0	0	0	0	0
	13	3	1	1	1	1	0	0	0	0	1	0	0	0	0
	15	2	1	1	1	1	0	0	0	0	0	0	0	0	0
	16	2	1	1	1	0	1	1	1	1	0	0	0	0	0
	18	4	1	1	0	0	0	0	0	0	0	0	0	0	0
C2	1	6	1	0	0	0	0	0	0	0	0	0	0	0	0
	2	5	1	1	0	0	0	0	0	0	0	0	0	0	0
	3	3	1	1	0	0	0	0	0	0	0	0	0	0	0
	5	3	1	1	1	0	0	0	0	1	1	0	0	0	0
	7	3	1	1	1	1	0	1	0	0	0	0	0	0	0
	9	6	1	1	1	1	1	0	1	0	0	0	0	0	0
C3	1	3	1	1	1	0	1	0	0	0	0	0	0	0	0
	2	2	1	1	0	0	0	1	0	1	1	0	0	0	0
	3	4	1	1	1	0	0	0	0	0	0	0	0	0	0
	5	21	1	0	0	0	0	0	0	0	0	0	0	0	0
	7	6	1	1	1	1	1	0	0	0	0	0	0	0	0
	9	3	1	1	1	1	1	1	1	1	0	0	0	0	0
C4	1	7	1	0	0	0	0	0	0	0	0	0	0	0	0
	3	6	1	1	0	0	0	0	0	0	0	0	0	0	0
	5	3	1	1	1	0	0	0	0	0	0	0	0	0	0
C5	1	4	1	1	0	0	0	0	0	0	0	0	0	0	0
	3	3	1	1	1	1	0	0	0	0	0	0	0	0	0
	5	5	1	1	1	0	0	0	0	0	0	0	0	0	0
PT			722	672	638	611	591	570	552	531	501	464	413	342	201
W			.1	.3	.5	.7	.9	1.2	1.5	2.0	3.0	5.0	10	20	40

Cm	mwtA	mwtB										
C1	53.9	18.55	0	0	0	0	0	0	0	0	0	0
	76.2	14.49	0	0	0	0	0	0	0	0	0	0
	89.6	13.61	0	0	0	0	0	0	0	0	0	0
	101.5	14.68	0	0	0	0	0	0	0	0	0	0
	105	13.31	0	0	0	0	0	0	0	0	0	0
	119.8	13.29	0	0	0	0	0	0	0	0	0	0
	128.8	14.78	0	0	0	0	0	0	0	0	0	0
	138.2	14.45	0	0	0	0	0	0	0	0	0	0
	145.3	13.63	0	0	0	0	0	0	0	0	0	0
	147.3	13.36	1	0	0	0	0	0	0	0	0	0
	154.9	13.28	1	0	0	0	0	0	0	0	0	0
	158.2	13.88	1	1	1	1	0	0	0	0	0	0
C2	16.1	14.4	0	0	0	0	0	0	0	0	0	0
	33.5	14.08	1	0	0	0	0	0	0	0	0	0
	39.1	13.65	0	0	0	0	0	0	0	0	0	0
	50.5	13.82	1	1	0	0	0	0	0	0	0	0
	59.2	13.55	1	0	0	0	0	0	0	0	0	0
	64.7	14.56	1	1	1	0	1	0	0	0	0	0
C3	27.5	13.58	0	0	0	0	0	0	0	0	0	0
	30.6	13.47	1	0	0	0	0	0	0	0	0	0
	40.9	14.1	0	0	0	0	0	0	0	0	0	0
	51.6	18.44	1	1	1	0	1	1	1	1	1	0
	59	14.34	1	1	0	0	0	0	0	0	0	0
	63.2	13.69	1	1	0	0	0	0	0	0	0	0
C4	7.9	14.49	1	1	0	0	0	0	0	0	0	0
	15.7	15.05	1	0	0	1	0	1	0	0	0	1
	20.4	13.71	1	1	0	0	0	0	0	0	0	0
C5	5.4	14.28	0	0	0	0	0	0	1	0	0	0
	13.9	13.5	1	1	0	0	0	0	0	0	0	0
	18.8	14.04	1	1	0	0	0	0	0	0	0	0
PT			664	617	528	473	423	399	386	351	316	219
W			.05	.1	.3	.5	.8	1.0	1.2	1.5	2.0	4.0

Figure 5.7: Upper left: (a), $(K_A + K_B) * W$ for $C3$; Upper right: (b), $(MWT_A + MWT_B) * W$ for $C3$; Lower left: (c), $(K_A + K_B) * W$ for $C1$; Lower right: (d), $(MWT_A + MWT_B) * W$ for $C1$. Layer A coupled with layer B

In order to investigate the effect on total profit of different seeding strategies in terms of the hub node placement in secondary layer B when coupled with layer A,

we perform a series of experiments in the following as shown in Figure 5.8. In (a-b) we keep everything the same as Figure 5.7 (a-b) except that we place one seed node in layer A to the hub node for each experiment. For example, in the experiment numbered 0 we place the node with degree $K_A = 1$ in community 1 on layer A in place of hub node on layer B and in the experiment numbered 17 we place the node with degree $K_A = 9$ in community 2 on layer A in the place of hub node on layer B. In (c-d) we keep everything the same as Figure 5.7 (c-d) except placing hub node following the same settings in (a-b) here. The degree of the hub node is $K_B = 364$ and $MWT_B = 94.75$

To compare the profit fairly under different experimental settings we need to convert the profit into comparable values. The last row of the four tables represent the *normalized final profit*, defined as the profit under the configuration given in the table but changing the weight W to the minimum in the series of experiment. For example, in table (b) we need to convert all the weight to the minimum $W_{min} = 1.0$ so for the experiment numbered 0 the normalized profit becomes $PT + (MWT^A_{tot} + MWT^B_{tot}) * (W - W_{min}) = 332 + (94.75 + 53.9) * (2.0 - 1.0) = 480.7$.

From Figure 5.7 we have the following 3 main findings. (1) For (a) and (c) only under extremely small value of W would the hub corresponding seed node be selected while for (b) and (d) it will always be selected except for very high W. (2) Almost all nodes are included in (a, c) while for (b, d) only one single optimal is included. (3) It would be more effective in terms of the normalized final profit to place hub in smaller community and in place of lower degree K_A in (b, d) but there is not significant difference in (a, c).

Phenomena (1) and (2) can be explained by the analysis we did in the previous sections. The cost function based on K in (a, c) is much more heterogeneous than the cost function based on MWT in (b, d) after including the hub node in layer B and therefore we expect the structure to be upper left triangle for (a, c) and V shape for

the (b, d). From previous analysis, in left upper triangle we build up the seed nodes from low degree to high degree and therefore only in the case of extremely small W would we expect to see the hub and when the hub first disappears the rest low degree nodes are still expected to remain. On the other hand, for V shape, the structure would agree with one single optimal seed node based on MPP, and we would expect only one hub node when W is moderate and the hub will disappear only when it is not the optimal one based on MPP.

Phenomenon (3) can be analyzed through investigating the forming components of total profit. For example, the difference in total profit between experiment numbered 0 and 6 is

$$PT_0 - PT_6 = F_0 - F_6 - \left[cost_0 - cost_6\right]$$

Here $cost_0 = \left(K^A_{tot} + K^B_{tot} + K^B_{hub} - K^B_{rep0} - (K^B_{rm0} + K^A_{rm0})\right) * W$ and similarly $cost_6 = \left(K^A_{tot} + K^B_{tot} + K^B_{hub} - K^B_{rep6} - (K^B_{rm6} + K^A_{rm6})\right) * W$ where K^A_{tot} is the total degree of all the candidate seed nodes in layer A, $K^B_{hub} = 364$ is the degree of the hub node on layer B, K^B_{rep0} is the original degree of the node on layer B replaced by the hub in experiment 0 and K^A_{rm0} is the total degree of all candidate seed nodes not included in the optimal set in layer A in experiment 0. Similar definition follows for others. For example in (a),

$$\begin{aligned} cost_0 - cost_6 &= \left[\left(K^B_{rep6} + (K^B_{rm6} + K^A_{rm6})\right) - \left(K^B_{rep0} + (K^B_{rm0} + K^A_{rm0})\right)\right] * W \\ &= [(1 + (1 + 2 + 1 + 1)) - (1 + (0 + 0))] * 0.05 = 0.25 \end{aligned}$$

Similarly, for other cases in (a) and (c), the differences in the values of cost functions are very small such that differences in contribution to the final total profit for different experiments can be negligible. Since the final profit are almost identical among placing hub node in different communities and degrees, we can conclude that difference in revenue is also negligible. So, in the case of designing the cost function in

terms of the degree K the placement of the hub node does not make a big difference.

As for (b) and (d), take (b) as an example,

$$\begin{aligned} cost_0 - cost_6 &= \left[(MWT^B_{hub} + MWT^A_0) - (MWT^B_{hub} + MWT^A_6)\right] * W \\ &= \left(MWT^A_0 - MWT^A_6\right) * W \\ &= (53.9 - 128.8) * 1.0 = -74.9 \end{aligned}$$

And therefore,

$$PT_0 - PT_6 = F_0 - F_6 - \left[cost_0 - cost_6\right] = F_0 - F_6 + 74.9$$

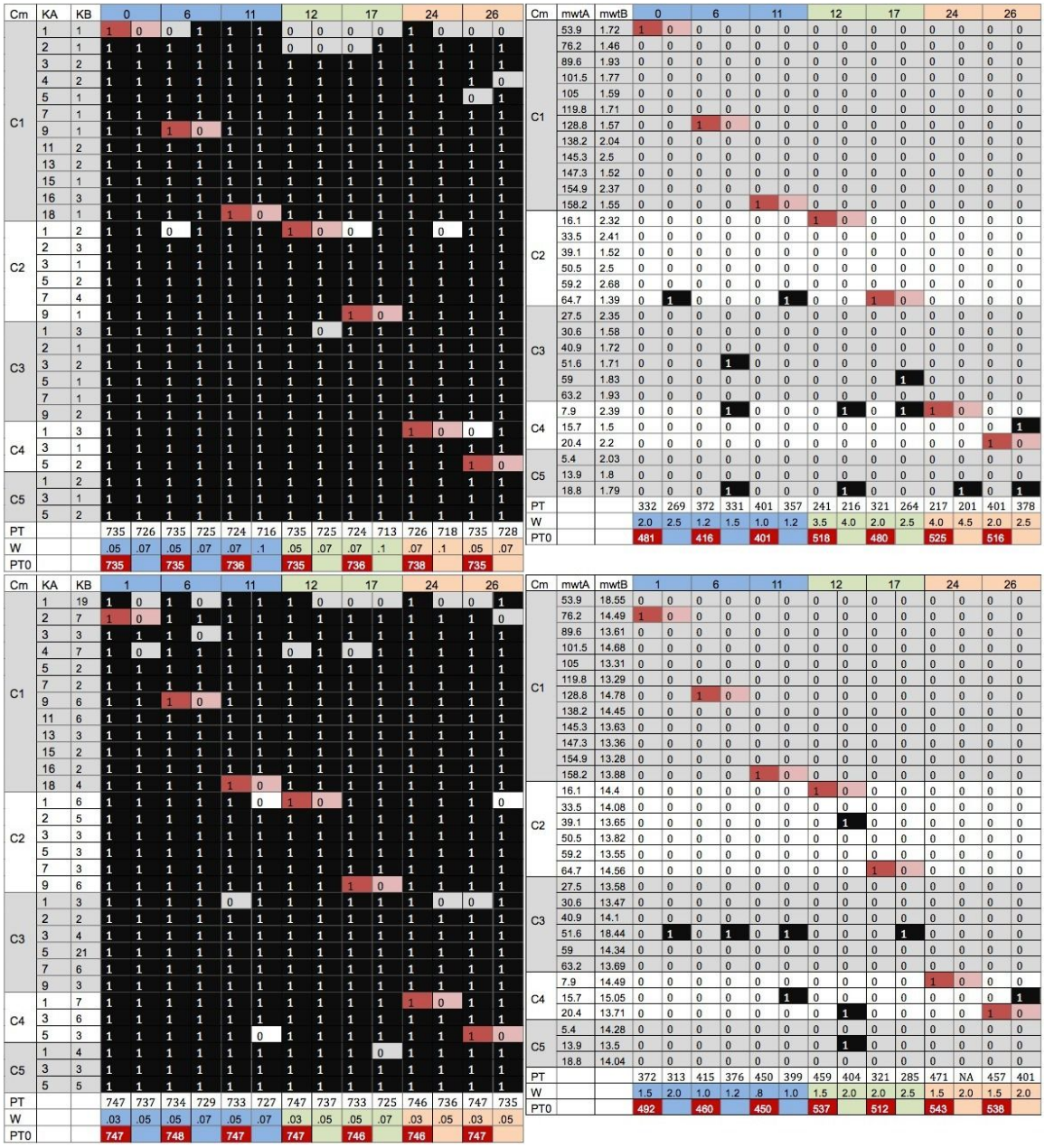
On the other hand, from the last row, we know $PT_0 - PT_6 = 481 - 416 = 65$ and therefore, $F_0 - F_6 \sim -10$. Therefore, we can conclude that the higher total profit achieved by placing the hub node in low degree mainly comes from lowered cost because the change in revenue is expected to be negative, and is more significant compared to the case (a, c) since in there many seed nodes are already included and changing the position of the hub does not make much difference.

Similarly, comparing experiment 12 and 0 in Figure 5.8 (b) for the effect of community size,

$$\begin{aligned} cost_{12} - cost_0 &= \left[(MWT^B_{hub} + MWT^A_{12}) - (MWT^B_{hub} + MWT^A_0)\right] * W \\ &= \left(MWT^A_{12} - MWT^A_0\right) * W \\ &= (16.1 - 53.9) * 1.0 = -37.8 \end{aligned}$$

On the other hand, from the last row, $PT_{12} - PT_0 = 518 - 481 = 37$ and therefore, $F_{12} - F_0 \sim -1$ and similarly, $F_{24} - F_{12} \sim -1$. So placing the hub node in smaller communities of the same degree does not make much difference in terms of revenue and the high profit comes from the lower cost.

To sum up, the difference in the total profit for different locations of hub node in terms of community size and degree in layer A in (b, d) mainly comes from the difference in cost value. As a matter of fact, the revenue would make contribution in the opposite direction.



Cm	KA	KB	0		6		11		12		17		24		26	
C1	1	1	1	0	0	1	1	1	0	0	0	0	1	0	0	0
	2	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1
	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	4	2	1	1	1	1	1	1	1	1	1	1	1	1	1	0
	5	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	9	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
	11	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	13	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	16	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	18	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
C2	1	2	1	1	0	1	1	1	1	0	0	1	1	0	1	1
	2	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	7	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	9	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
C3	1	3	1	1	1	1	1	1	1	0	1	1	1	1	1	1
	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	9	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C4	1	3	1	1	1	1	1	1	1	1	1	1	1	0	0	1
	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	2	1	1	1	1	1	1	1	1	1	1	1	1	1	0
C5	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PT			735	726	735	725	724	716	735	725	724	713	726	718	735	728
W			.05	.07	.05	.07	.07	.1	.05	.07	.07	.1	.07	.1	.05	.07
PT0			735		735		736		735		736		738		735	

Cm	mwtA	mwtB	0		6		11		12		17		24		26	
C1	53.9	1.72	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	76.2	1.46	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	89.6	1.93	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	101.5	1.77	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	105	1.59	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	119.8	1.71	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	128.8	1.57	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	138.2	2.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	145.3	2.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	147.3	1.52	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	154.9	2.37	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	158.2	1.55	0	0	0	0	1	0	0	0	0	0	0	0	0	0
C2	16.1	2.32	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	33.5	2.41	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	39.1	1.52	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	50.5	2.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	59.2	2.68	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	64.7	1.39	0	1	0	0	0	1	0	0	1	0	0	0	0	0
C3	27.5	2.35	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	30.6	1.58	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	40.9	1.72	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	51.6	1.71	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	59	1.83	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	63.2	1.93	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C4	7.9	2.39	0	0	0	1	0	0	0	1	0	1	1	0	0	0
	15.7	1.5	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	20.4	2.2	0	0	0	0	0	0	0	0	0	0	0	0	1	0
C5	5.4	2.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	13.9	1.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	18.8	1.79	0	0	0	1	0	0	0	1	0	0	0	1	0	1
PT			332	269	372	331	401	357	241	216	321	264	217	201	401	378
W			2.0	2.5	1.2	1.5	1.0	1.2	3.5	4.0	2.0	2.5	4.0	4.5	2.0	2.5
PT0			481		416		401		518		480		525		516	

Cm	KA	KB	1		6		11		12		17		24		26	
C1	1	19	1	0	1	0	1	1	1	0	0	0	1	0	0	1
	2	7	1	0	1	1	1	1	1	1	1	1	1	1	1	0
	3	3	1	1	1	0	1	1	1	1	1	1	1	1	1	1
	4	7	1	0	1	1	1	1	0	1	0	1	1	1	1	1
	5	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	7	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	9	6	1	1	1	0	1	1	1	1	1	1	1	1	1	1
	11	6	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	13	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	15	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	16	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	18	4	1	1	1	1	1	0	1	1	1	1	1	1	1	1
C2	1	6	1	1	1	1	1	0	1	0	1	1	1	1	1	0
	2	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	7	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	9	6	1	1	1	1	1	1	1	1	1	0	1	1	1	1
C3	1	3	1	1	1	1	0	1	1	1	1	1	0	0	1	1
	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	3	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	21	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	7	6	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	9	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C4	1	7	1	1	1	1	1	1	1	1	1	1	1	0	1	1
	3	6	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	3	1	1	1	1	1	0	1	1	1	1	1	1	1	0
C5	1	4	1	1	1	1	1	1	1	1	1	0	1	1	1	1
	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PT			747	737	734	729	733	727	747	737	733	725	746	736	747	735
W			.03	.05	.05	.07	.05	.07	.03	.05	.05	.07	.03	.05	.03	.05
PT0			747		748		747		747		746		746		747	

Cm	mwtA	mwtB	1		6		11		12		17		24		26	
C1	53.9	18.55	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	76.2	14.49	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	89.6	13.61	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	101.5	14.68	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	105	13.31	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	119.8	13.29	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	128.8	14.78	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	138.2	14.45	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	145.3	13.63	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	147.3	13.36	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	154.9	13.28	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	158.2	13.88	0	0	0	0	1	0	0	0	0	0	0	0	0	0
C2	16.1	14.4	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	33.5	14.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	39.1	13.65	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	50.5	13.82	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	59.2	13.55	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	64.7	14.56	0	0	0	0	0	0	0	0	1	0	0	0	0	0
C3	27.5	13.58	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	30.6	13.47	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	40.9	14.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	51.6	18.44	0	1	0	1	0	1	0	0	0	1	0	0	0	0
	59	14.34	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	63.2	13.69	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C4	7.9	14.49	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	15.7	15.05	0	0	0	0	0	1	0	0	0	0	0	0	0	1
	20.4	13.71	0	0	0	0	0	0	0	1	0	0	0	0	1	0
C5	5.4	14.28	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	13.9	13.5	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	18.8	14.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PT			372	313	415	376	450	399	459	404	321	285	471	NA	457	401
W			1.5	2.0	1.0	1.2	.8	1.0	1.5	2.0	2.0	2.5	1.5	2.0	1.5	2.0
PT0			492		460		450		537		512		543		538	

Figure 5.8: Upper left: (a), $(K_A + K_B) * W$ with hub for $C3$; Upper right: (b), $(MWT_A + MWT_B) * W$ with hub for $C3$; Lower left: (c), $(K_A + K_B) * W$ with hub for $C1$; Lower right: (d), $(MWT_A + MWT_B) * W$ with hub for $C1$. Layer A coupled with layer B with hub node

Finally we include the plot shown in Figure 5.9 for PT vs W for both when $cost = K * W$ and $cost = MWT * W$ including both the case with single layered A and the case with coupled A and B layers.

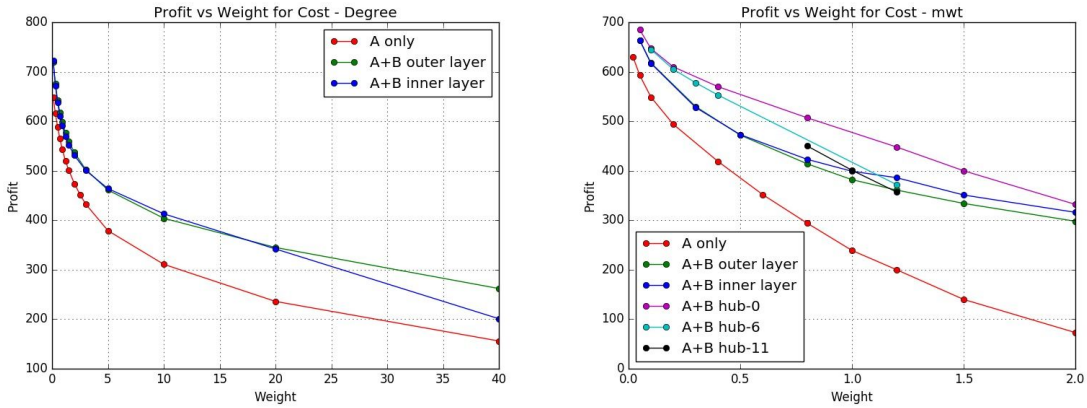


Figure 5.9: Left: (a), PT vs W for $cost = K * W$; Right: (b), PT vs W for $cost = MWT * W$. PT vs W for the two cost functions

From the plots we can see that for our settings, including extra layer of diffusion channel would result in a higher profit in the whole range of W values. In addition, for the case $cost = K * W$ as shown in (a) the profit would decrease very quickly with W when W is small, while on the other hand, the profit is pretty resilient with W increases further. This is consistent with the fact that under this scenario the system can always outperform the one optimal seed node configuration based on personal profit by choosing a group of low degree nodes, and the smaller W the stronger this effect. For the case $cost = MWT * W$ there is no early fast decreasing stage which is also consistent with the fact that under this scenario the system build up seed nodes from high to low degree and only under extremely small value of W would the system include a group of low degree nodes so as to outperform the single optimal node. Therefore, we would expect similar pattern for larger W range in (a) and basically the whole range of W in (c), which is the case.

5.4 Conclusion

In summary, we employed a hybrid version of the binary particle swarm optimization algorithm with mutation and cross over process from genetic algorithm in

optimal seed nodes selection for the profit maximization objective in information diffusion process on both single layered networks and coupled multi layered networks. For comparison we also developed a seed node selection strategy based on MPP. Roughly we can replicate the seed nodes location structure by treating the selection process as multiple steps. In the first step it is the competition between a group of low degree nodes and the single one optimal node based on MPP in terms of total profit. Then step by step new seed nodes are added into the optimal set identified in step 1 based on the same competition mechanism in terms marginal profit. When the cost function is heterogeneous, e.g., as a function of the degree K, we would expect the seed nodes to build up from low to high degree ones and form an upper left triangle shape if we table them in a way with increasing W or Q from left to right and increasing degree from top to bottom. On the other hand, cost function value would become homogeneous across nodes if we take the network topology into consideration, as in the case when cost is a function of MWT. Under this scenario, the seed nodes would build up from high degree nodes and V shape in the seed nodes location structure is expected. Besides, As W or Q increases, good agreement on seed nodes selection is achieved between the results by PSO algorithm and by the strategy of MPP. When we couple the primary layer of network containing community structure with the secondary layer scale free network, the community effect is quenched, just as expected, because the secondary layer would significantly facilitate the information diffusion process between communities.

CHAPTER VI

Conclusion

The focus of this thesis was on the following three topics: (i) new methods for identifying phase transition epochs in dynamically evolving networks exhibiting community structure, (ii) introduction of new models motivated by information diffusion problems along with the mathematical analysis of their evolution and (iii) development of computational methodology for selecting seed nodes that optimize a performance measure (e.g. a profit function) in information diffusion networks. The work provided analytical tools for a number of timely problems in dynamic network analysis.

Specifically, the low rank plus sparse noise model for time evolving networks provides a robust method for detecting phase transition epochs and at the same time diagnosing the changes in community structure that contribute to the regime changes. The methodology was assessed on a number of synthetic data sets, including the Kuramoto model of coupled oscillators and showed its effectiveness. Further, the analysis of the US Senate voting record over the last 35 years also provided insights into the onset of political polarization, as well which factors contributed to it.

On the second thrust, the thesis proposed the SIM and SIT models and rigorously analyzed their dynamics using mean field theory approximations for both the homogeneous and heterogeneous degree distribution case. Analytical solutions were obtained

for early stage evolution when $t \to 0$ and a very good agreement was achieved with Mote Carlo (MC) simulation based results. Regarding the expected saturated diffusion fraction, in most cases the result is obtained through the numerical solution of the corresponding ODE. However, for the homogeneous network SIT model, a close form solution for the whole time range is derived, and good agreement between the MC simulation results was also achieved, particularly for denser networks. Further, both *epidemic thresholds* and *phase transition time points* were obtained based on early stage evolution analysis. Finally, we also obtained an analytical solution for the probability of each node getting informed for both the SIT and SIM models depending on the path to the seed node, and approximated the expected saturated diffusion size by growing a MWT. Good agreement with numerical results can be achieved when the network has a tree-like topology or is rather sparse.

Finally, the thesis examined the topic of maximizing functionals related to the information diffusion process. By developing an appropriate optimal function and optimizing it leveraging ideas from binary particle swarm optimization algorithm, we were able to select the optimal set of seed nodes from a group of candidate seed nodes. The analysis showed that for networks exhibiting community structure, it is always a good strategy to allocate at least one seed node in each of the available communities, while if there is a channel through a secondary network (multi-layer network structure) having free scale structure, this is not the case anymore. We further noticed that if the cost component of the profit function is very heterogeneous across nodes -e.g. a function of degree of seed nodes- the optimal set of seed nodes corresponds to a group of low degree nodes, with the high degree nodes included only in the case of extremely small W or Q. On the other hand, if the cost is pretty homogeneous across nodes -e.g. a function of the value of its MWT- the optimal set comprises of high degree nodes, and good agreement between the optimal set returned by PSO and by MPP is achieved in this case. The analysis provides new insights on the interplay between

the topology structure of the underlying network and selecting a set of seed nodes to optimize information diffusion.

For future work, theoretical analysis of information diffusion processes based on the proposed SIM and SIT models, for networks with community structure and on multi-layer networks represents a promising, but also technically challenging direction. The theoretical study can provide deeper insights into the diffusion process mechnaisms on networks with complex topological structures, and can be used to guide the optimal seed nodes selection process so that extensive simulation work can be reduced to a minimum.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] E. Abbe. Community detection and stochastic block models: recent developments. *ArXiv e-prints*, March 2017.

[2] Christopher Aicher, Abigail Jacobs, and Aaron Clauset. Adapting the stochastic block model to edge-weighted networks. *arXiv*, 05 2013.

[3] Sinan Aral and Dylan Walker. Identifying influential and susceptible members of social networks. *Science*, 337(6092):337–341, 2012.

[4] Alex Arenas, Albert Díaz-Guilera, and Conrad J. Pérez-Vicente. Synchronization reveals topological scales in complex networks. *Phys. Rev. Lett.*, 96:114102, Mar 2006.

[5] Jushan Bai. Common breaks in means and variances for panel data. *Journal of Econometrics*, 157(1):78 – 92, 2010. Nonlinear and Nonparametric Methods in Econometrics.

[6] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 519–528, New York, NY, USA, 2012. ACM.

[7] Brian Ball and M.E.J. Newman. Friendship networks and social status. *Network Science*, 1(1):16–30, 2013.

[8] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[9] Marc Barthélemy, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.*, 92:178701, Apr 2004.

[10] Danielle S. Bassett, Mason A. Porter, Nicholas F. Wymbs, Scott T. Grafton, Jean M. Carlson, and Peter J. Mucha. Robust detection of dynamic community structure in networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(1):013142, 2013.

[11] S. Boccaletti, G. Bianconi, R. Criado, C.I. del Genio, J. Gmez-Gardees, M. Romance, I. Sendia-Nadal, Z. Wang, and M. Zanin. The structure and dynamics

of multilayer networks. *Physics Reports*, 544(1):1 – 122, 2014. The structure and dynamics of multilayer networks.

[12] Marián Boguñá and Romualdo Pastor-Satorras. Epidemic spreading in correlated complex networks. *Phys. Rev. E*, 66:047104, Oct 2002.

[13] Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. Locating privileged spreaders on an online social network. *Phys. Rev. E*, 85:066123, Jun 2012.

[14] Alfred M. Bruckstein, David L. Donoho, and Michael Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.

[15] Keith Burghardt, Christopher Verzijl, Junming Huang, Matthew Ingram, Binyang Song, and Marie-Pierre Hasne. Testing modeling assumptions in the west africa ebola outbreak. *Scientific Reports*, 6:34598 EP –, Oct 2016. Article.

[16] Chao-Ran Cai, Zhi-Xi Wu, Michael Z. Q. Chen, Petter Holme, and Jian-Yue Guan. Solving the dynamic correlation problem of the susceptible-infected-susceptible model on networks. *Phys. Rev. Lett.*, 116:258301, Jun 2016.

[17] Jian-Feng Cai, Emmanuel J. Cands, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[18] T. Tony Cai and Xiaodong Li. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Ann. Statist.*, 43(3):1027–1059, 06 2015.

[19] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *J. ACM*, 58(3):11:1–11:37, June 2011.

[20] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717, Apr 2009.

[21] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591–646, May 2009.

[22] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

[23] Jingchun Chen and Bo Yuan. Detecting functional modules in the yeast protein protein interaction network. *Bioinformatics*, 22(18):2283–2290, 2006.

[24] M. Clerc and J. Kennedy. The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6(1):58–73, Feb 2002.

[25] Nuno Crokidakis and Marcio Argollo de Menezes. Critical behavior of the sis epidemic model with time-dependent infection rate. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(05):P05012, 2012.

[26] Mihai Cucuringu. Synchronization over z2 and community detection in signed multiplex networks with constraints. *Journal of Complex Networks*, 3(3):469–506, 2015.

[27] D. J. Daley and D. G. Kendall. Epidemics and rumours. *Nature*, 204:1118 EP –, Dec 1964.

[28] M. De Domenico, A. Lima, P. Mougel, and M. Musolesi. The anatomy of a scientific rumor. *Scientific Reports*, 3:2980 EP –, Oct 2013. Article.

[29] M. De Domenico, A. Lima, P. Mougel, and M. Musolesi. The anatomy of a scientific rumor. *Scientific Reports*, 3:2980 EP –, Oct 2013. Article.

[30] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84:066106, Dec 2011.

[31] Peter Sheridan Dodds and Duncan J. Watts. Universal behavior in a generalized model of contagion. *Phys. Rev. Lett.*, 92:218701, May 2004.

[32] Pedro Domingos and Matt Richardson. *Mining the network value of customers*, pages 57–66. 2001.

[33] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Critical phenomena in complex networks. *Rev. Mod. Phys.*, 80:1275–1335, Oct 2008.

[34] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Phys. Rev. Lett.*, 85:4633–4636, Nov 2000.

[35] Ken T. D. Eames and Matt J. Keeling. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proceedings of the National Academy of Sciences*, 99(20):13330–13335, 2002.

[36] R. C. Eberhart and Yuhui Shi. Tracking and optimizing dynamic systems with particle swarms. In *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, volume 1, pages 94–100 vol. 1, 2001.

[37] Paul Erdős and Alfréd Rényi. On random graphs i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959 1959.

[38] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University, 2002.

[39] M. Fazel, E. Candes, B. Recht, and P. Parrilo. Compressed sensing and robust recovery of low rank matrices. In *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pages 1043–1047, Oct 2008.

[40] M. Fazel, H. Hindi, and S. P. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *Proceedings of the 2003 American Control Conference, 2003.*, volume 3, pages 2156–2162 vol.3, June 2003.

[41] G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee. Self-organization and identification of web communities. *Computer*, 35(3):66–70, Mar 2002.

[42] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75 – 174, 2010.

[43] Noah E. Friedkin. Theoretical foundations for centrality measures. *American Journal of Sociology*, 96(6):1478–1504, 1991.

[44] Piotr Fryzlewicz. Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, 42(6):2243–2281, 12 2014.

[45] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30(4):1141–1144, 12 1959.

[46] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[47] James P. Gleeson, Sergey Melnik, Jonathan A. Ward, Mason A. Porter, and Peter J. Mucha. Accuracy of mean-field theory for dynamics on real-world networks. *Phys. Rev. E*, 85:026106, Feb 2012.

[48] Scott Greenhalgh and Troy Day. Time-varying and state-dependent recovery rates in epidemiological models. *Infectious Disease Modelling*, 2(4):419 – 430, 2017.

[49] Daniel Gruhl, R. Guha, David Liben-Nowell, and Andrew Tomkins. Information diffusion through blogspace. In *Proceedings of the 13th International Conference on World Wide Web*, WWW '04, pages 491–501, New York, NY, USA, 2004. ACM.

[50] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A. Zighed. Information diffusion in online social networks: A survey. *SIGMOD Rec.*, 42(2):17–28, July 2013.

[51] Roger Guimerà and Luís A. Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433:895, Feb 2005.

[52] Roger Guimerà, Marta Sales-Pardo, and Luís A. Nunes Amaral. Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E*, 70:025101, Aug 2004.

[53] X.C. Guo, J.H. Yang, C.G. Wu, C.Y. Wang, and Y.C. Liang. A novel ls-svms hyper-parameter selection based on particle swarm optimization. *Neurocomputing*, 71(16):3211 – 3215, 2008. Advances in Neural Information Processing (ICONIP 2006) / Brazilian Symposium on Neural Networks (SBRN 2006).

[54] Lajos Horvath and Miklos Csorgo. *Limit Theorems in Change-Point Analysis*. Wiley, New York, NY, USA, 1997.

[55] Lajos Horvath and Gregory Rice. Extensions of some classical methods in change point analysis. *TEST*, 23(2):219–255, Jun 2014.

[56] Shlomo Kalish. A new product adoption model with price, advertising, and uncertainty. *Management Science*, 31(12):1569–1585, 1985.

[57] Brian Karrer and M. E. J. Newman. Message passing approach for general epidemic models. *Phys. Rev. E*, 82:016101, Jul 2010.

[58] Brian Karrer and M. E. J. Newman. Competing epidemics on complex networks. *Phys. Rev. E*, 84:036106, Sep 2011.

[59] Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83:016107, Jan 2011.

[60] Brian Karrer, M. E. J. Newman, and Lenka Zdeborová. Percolation on sparse networks. *Phys. Rev. Lett.*, 113:208702, Nov 2014.

[61] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 137–146, New York, NY, USA, 2003. ACM.

[62] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, volume 4, pages 1942–1948 vol.4, Nov 1995.

[63] J. Kennedy and R. C. Eberhart. A discrete binary version of the particle swarm algorithm. In *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, volume 5, pages 4104–4108 vol.5, Oct 1997.

[64] J. Kennedy and R. Mendes. Population structure and particle swarm performance. In *Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on*, volume 2, pages 1671–1676, 2002.

[65] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6:888 EP –, Aug 2010.

[66] Eric D. Kolaczyk. *Topics at the Frontier of Statistics and Network Analysis: (Re)Visiting the Foundations*. SemStat Elements. Cambridge University Press, 2017.

[67] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Phys. Rev. Lett.*, 85:4629–4632, Nov 2000.

[68] Ann E. Krause, Kenneth A. Frank, Doran M. Mason, Robert E. Ulanowicz, and William W. Taylor. Compartments revealed in food-web structure. *Nature*, 426:282, Nov 2003.

[69] Sang Hoon Lee, Jos Manuel Magallanes, and Mason A. Porter. Time-dependent community structure in legislation cosponsorship networks in the congress of the republic of peru. *Journal of Complex Networks*, 5(1):127–144, 2017.

[70] Sangwook Lee, Sangmoon Soak, Sanghoun Oh, Witold Pedrycz, and Moongu Jeon. Modified binary particle swarm optimization. *Progress in Natural Science*, 18(9):1161 – 1166, 2008.

[71] Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1(1), May 2007.

[72] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 497–506, New York, NY, USA, 2009. ACM.

[73] J. Lin and G. Michailidis. Regularized Estimation and Testing for High-Dimensional Multi-Block Vector-Autoregressive Models. *ArXiv e-prints*, August 2017.

[74] Z. Lin, M. Chen, and Y. Ma. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. *ArXiv e-prints*, September 2010.

[75] Xinzhi Liu and Peter Stechlinski. Infectious disease models with time-varying parameters and general nonlinear incidence rate. *Applied Mathematical Modelling*, 36(5):1974–1994, 2012.

[76] Z. Lu, M. Banerjee, and G. Michailidis. Intelligent sampling for multiple change-points in exceedingly long time series with rate guarantees. *ArXiv e-prints*, October 2017.

[77] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. Recommender systems with social regularization. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 287–296, New York, NY, USA, 2011. ACM.

[78] Daniel P Maki and (joint author.) Thompson, Maynard. *Mathematical models and applications : with emphasis on the social, life, and management sciences*. Englewood Cliffs, N.J. : Prentice-Hall, 1973. Includes bibliographies.

[79] Federico Marini and Beata Walczak. Finding relevant clustering directions in high-dimensional data using particle swarm optimization. *Journal of Chemometrics*, 25(7):366–374, 2011.

[80] Federico Marini and Beata Walczak. Particle swarm optimization (pso). a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 149(Part B):153 – 165, 2015.

[81] Sergei Maslov and Kim Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913, 2002.

[82] Angelica S. Mata and Silvio C. Ferreira. Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks. *EPL (Europhysics Letters)*, 103(4):48003, 2013.

[83] Marina Meila and Jianbo Shi. Learning segmentation by random walks. In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems 13*, pages 873–879. MIT Press, 2001.

[84] Muhammd Ilyas Menhas, MinRui Fei, Ling Wang, and Xiping Fu. *A Novel Hybrid Binary PSO Algorithm*, pages 93–100. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[85] Joel C. Miller, Anja C. Slim, and Erik M. Volz. Edge-based compartmental modelling for infectious disease spread. *Journal of The Royal Society Interface*, 9(70):890–906, 2012.

[86] Byungjoon Min, Sang-Hwan Gwak, Nanoom Lee, and K.-I. Goh. Layer-switching cost and optimality in information spreading on multiplex networks. *Scientific Reports*, 6:21392 EP –, Feb 2016. Article.

[87] James Moody and Peter J. Mucha. Portrait of political party polarization. *Network Science*, 1(1):119–121, 2013.

[88] Stephen Morris. Contagion. *The Review of Economic Studies*, 67(1):57–78, 2000.

[89] Peter J. Mucha, Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.

[90] Seth A. Myers, Chenguang Zhu, and Jure Leskovec. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 33–41, New York, NY, USA, 2012. ACM.

[91] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili. Theory of rumour spreading in complex social networks. *Physica A: Statistical Mechanics and its Applications*, 374(1):457 – 470, 2007.

[92] Azadeh Nematzadeh, Emilio Ferrara, Alessandro Flammini, and Yong-Yeol Ahn. Optimal network modularity for information diffusion. *Phys. Rev. Lett.*, 113:088701, Aug 2014.

[93] M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, Jul 2002.

[94] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, 69:066133, Jun 2004.

[95] M. E. J. Newman. Threshold effects for two pathogens spreading on a network. *Phys. Rev. Lett.*, 95:108701, Sep 2005.

[96] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74:036104, Sep 2006.

[97] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[98] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, Feb 2004.

[99] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118, Jul 2001.

[100] Mark Newman. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA, 2010.

[101] Jukka-Pekka Onnela, Daniel J. Fenn, Stephen Reid, Mason A. Porter, Peter J. Mucha, Mark D. Fricker, and Nick S. Jones. Taxonomies of networks from community structure. *Phys. Rev. E*, 86:036104, Sep 2012.

[102] L. Ouyang and Dongyun Wang. New particle swarm optimization algorithm for knapsack problem. In *2012 8th International Conference on Natural Computation*, pages 786–788, May 2012.

[103] R. Paffenroth, K. Kay, and L. Servi. Robust PCA for Anomaly Detection in Cyber Networks. *ArXiv e-prints*, January 2018.

[104] Gergely Palla, Albert-László Barabási, and Tamás Vicsek. Quantifying social group evolution. *Nature*, 446:664 EP –, Apr 2007.

[105] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87:925–979, Aug 2015.

[106] Keith T. Poole and Howard Rosenthal. *Congress : a political-economic history of roll call voting*. Oxford University Press, New York, 1997.

[107] Amira Rachah and Delfim F. M. Torres. Mathematical modelling, simulation, and optimal control of the 2014 ebola outbreak in west africa. *Discrete Dynamics in Nature and Society*, 2015(842792), 2015.

[108] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[109] Guangming Ren and Xingyuan Wang. Epidemic spreading in time-varying community networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24(2):023116, 2014.

[110] Daniel M. Romero, Brendan Meeder, and Jon Kleinberg. Differences in the mechanics of information diffusion across topics. *Proceedings of the 20th international conference on World wide web - WWW '11*, page 695, 2011.

[111] Sandipan Roy, Yves Atchade, and George Michailidis. Change point estimation in high dimensional markov random-field models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1187–1206, 2017.

[112] A. Safikhani and A. Shojaie. Joint Structural Break Detection and Parameter Estimation in High-Dimensional Non-Stationary VAR Models. *ArXiv e-prints*, November 2017.

[113] Anna Saumell-Mendiola, M. Ángeles Serrano, and Marián Boguñá. Epidemic spreading on interconnected networks. *Phys. Rev. E*, 86:026106, Aug 2012.

[114] Suresh P. Sethi. Dynamic optimal control models in advertising: a survey. *SIAM Review*, 19(4):685–725, 1977.

[115] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, August 2000.

[116] Yuhui Shi and Russell C. Eberhart. *Parameter selection in particle swarm optimization*, pages 591–600. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[117] Alexey N. Skvortsov. Estimation of rotation ambiguity in multivariate curve resolution with charged particle swarm optimization (cpso-mcr). *Journal of Chemometrics*, 28(10):727–739, 2014. CEM-14-0047.R1.

[118] Luis Sol, Miguel Romance, Regino Criado, Julio Flores, Alejandro Garca del Amo, and Stefano Boccaletti. Eigenvector centrality of nodes in multiplex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(3):033131, 2013.

[119] Aidan Sudbury. The proportion of the population never hearing a rumour. *Journal of Applied Probability*, 22(2):443–446, 1985.

[120] Jimeng Sun, Christos Faloutsos, Spiros Papadimitriou, and Philip S. Yu. Graphscope: Parameter-free mining of large time-evolving graphs. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '07, pages 687–696, New York, NY, USA, 2007. ACM.

[121] Min Tao and Xiaoming Yuan. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM Journal on Optimization*, 21(1):57–81, 2011.

[122] Thomas W. Valente. Network interventions. *Science*, 337(6090):49–53, 2012.

[123] Alessandro Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8:32 EP –, Dec 2011. Review Article.

[124] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, Dec 2007.

[125] Bo Wahlberg, Stephen Boyd, Mariette Annergren, and Yang Wang. An admm algorithm for a class of total variation regularized estimation problems*. *IFAC Proceedings Volumes*, 45(16):83 – 88, 2012. 16th IFAC Symposium on System Identification.

[126] Kang-Ping Wang, Lan Huang, Chun-Guang Zhou, and Wei Pang. Particle swarm optimization for traveling salesman problem. In *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*, volume 3, pages 1583–1585 Vol.3, Nov 2003.

[127] Tengyao Wang and Richard J. Samworth. High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):57–83.

[128] Wei Wang, Ming Tang, H Eugene Stanley, and Lidia A Braunstein. Unification of theoretical approaches for epidemic spreading on complex networks. *Reports on Progress in Physics*, 80(3):036603, 2017.

[129] Wei Wang, Ming Tang, Hui Yang, Younghae Do, Ying-Cheng Lai, and Gyu-Won Lee. Asymmetrically interacting spreading dynamics on complex layered networks. *Scientific Reports*, 4:5097 EP –, May 2014. Article.

[130] J. Xie, S. Sreenivasan, G. Korniss, W. Zhang, C. Lim, and B. K. Szymanski. Social consensus through the influence of committed minorities. *Phys. Rev. E*, 84:011130, Jul 2011.

[131] J. Xin, G. Chen, and Y. Hai. A particle swarm optimizer with multi-stage linearly-decreasing inertia weight. In *2009 International Joint Conference on Computational Sciences and Optimization*, volume 1, pages 505–508, April 2009.

[132] J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *2010 IEEE International Conference on Data Mining*, pages 599–608, Dec 2010.

[133] Zimo Yang, Ai-Xiang Cui, and Tao Zhou. Impact of heterogeneous human activities on epidemic spreading. *Physica A: Statistical Mechanics and its Applications*, 390(23):4543 – 4548, 2011.

[134] S. X. Yu and J. Shi. Multiclass spectral clustering. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 313–319 vol.1, Oct 2003.

[135] Hai-Feng Zhang, Jia-Rong Xie, Ming Tang, and Ying-Cheng Lai. Suppression of epidemic spreading in complex networks by local information based behavioral responses. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24(4):043106, 2014.

[136] Huiling Zhang, Zhi-Hong Guan, Tao Li, Xian-He Zhang, and Ding-Xue Zhang. A stochastic sir epidemic on scale-free network with community structure. *Physica A: Statistical Mechanics and its Applications*, 392(4):974 – 981, 2013.

[137] Shihua Zhang, Junfei Zhao, and Xiang-Sun Zhang. Common community structure in time-varying networks. *Phys. Rev. E*, 85:056110, May 2012.

[138] Jie Zhou, Zonghua Liu, and Baowen Li. Influence of network structure on rumor propagation. *Physics Letters A*, 368(6):458 – 463, 2007.