A Study of Phase Transition in New Random Graph Families

by

Mehrdad Moharrami

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2020

Doctoral Committee:

Professor Mingyan Liu, Co-Chair
Associate Professor Vijay Gautam Subramanian, Co-Chair
Professor Jinho Baik
Researcher Charles Bordenave, Institut de Mathématiques de Marseille
Professor Bruce Hajek, University of Illinois at Urbana-Champaign
Professor Cristopher Moore, Santa Fe Institute
Professor Rajesh Sundaresan, Indian Institute of Science

Mehrdad Moharrami
moharami@umich.edu

ORCID iD: 0000-0003-3907-8406

Dedicated to my parents
Abbas Moharamigargari and Roghieh Afsari

ACKNOWLEDGEMENTS

I cannot begin to express my thanks to my advisors, Mingyan Liu and Vijay Subramanian, who guided me and encouraged me to carry on through these years. I'm deeply indebted to them for their patience and understanding of the hardships of the first few years. I have been very fortunate to work with advisors who gave me the freedom to explore the area of my interest, while being deeply involved in my research. I would also like to extend my deepest gratitude to other committee members: Jinho Biak, Charles Bordenave, Bruce Hajek, Cristopher Moore, and Rajesh Sundaresan. Your high standards on the quality of research, as well as ethical conduct, are aspects I hope to emulate in my future career. I hope I will get a chance to collaborate and to learn from you in the years to come. I am most grateful to Cristopher Moore, a dear friend of mine, and the people of Santa Fe Institute for their hospitality.

I must thank the entire faculty of the Department of Electrical Engineering and the Department of Mathematics for creating a great learning atmosphere. I would like to express my special gratitude to Demos Teneketzis and Achilleas Anastasopoulos for being excellent teachers and great mentors. I would also like to thank Jinho Biak and Mark Rudelson for contributing to almost all of my mathematical knowledge.

Thanks should also go to to my friends and my colleagues who contributed to wonderful memories in Ann Arbor and made the journey possible. I feel very lucky to have been surrounded by amazing friends in Ann Arbor. I cannot list all the names here, but you are always on my mind.

None of this would have been possible without the endless love and support of my parents, Abbas and Roghieh. I am in my parents debt for showing faith in me and supporting my decisions throughout my life. I salute you all for the selfless love, care, pain and sacrifice you did to shape my life. I would never be able to pay back the love and affection showered upon by my parents. Also I express my thanks to my siblings for their support and love.

I owe thanks to a very special person, my wife, Niloufar for her continued and unfailing love, support and understanding during my pursuit of Ph.D degree that made the completion of thesis possible. I greatly value her contribution and deeply appreciate her belief in me. I am the luckiest for her unconditional love and support.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

ABSTRACT

Random graphs are mathematical models for understanding real-world networks. Important properties can be captured, processes studied, and rigorous predictions made. Phase transitions (sudden changes in structural properties caused by varying an underlying parameter) are commonly observed in random graphs. Our work focuses on phase transitions in three models. We study emergence of cascades and impact of community structure on phase transition in threshold-based contagion models using modular random graphs generated by configuration model and differential equation method. Using local weak analysis, we study a new graph model generated by bilateral agreement of individuals and analyze when a giant component emerges. Using the objective method and motivated by particle tracking in physics and object tracking in videos, we study detectability threshold of a hidden planted matching in a complete bipartite randomly weighted graph.

CHAPTER I

Introduction

1.1 Motivation

Networks are everywhere. The World Wide Web is a network of Web pages connected by hyperlinks to each other. Social networks such as Facebook are networks of user accounts with the connections governed by friendship status. Food webs are also networks where the linkage illustrate feeding pathways. They also arise in many other biological systems: e.g., the brain is a network of nerve cells connected by axons and even interactions between proteins in the cells can be modeled using networks.

The study of networks has attracted increased attention during the last few decades. Scientists from different fields are trying to understand the structure and the properties of real-world networks. One of the main tools that have been used for the study of networks is graph theory. The first use of graph theory dates back to the celebrated work of Euler on the bridges of Königsberg in 1736. Ever since, graph theory has been developed into a rich branch of discrete mathematics.

One of the central questions in network science is the following: given some examples of real-world networks, how might you reason that observed phenomena are "typical" to such networks? Often, there is only limited information available about the network. This may arise due to lack of data, or the difficulty in analyzing a large data set. Despite this, we need to understand and forecast the behavior of a process on a typical example of the network. Running a process over the given network can be impossible or too costly, and is prone to error because of a limited number of samples. The scientific method suggests using mathematical models to make predictions, and to compare them with the observed real-world phenomena. Based on these concerns, a reasonable and fruitful procedure is to generate graphs randomly, using the structural properties of real-world networks.

The introduction of probabilistic methods to graph theory inspired a new branch of mathematics called random graph theory. The term "random" usually refers to the ran-

dom construction of the graph; generally speaking, a random graph model is a probability distribution over graphs. The work of Erdös and Rényi in [2] and Gilbert in [3] were the first formal introduction of random graphs. However, the first discussion related to random graphs appeared in [4], where Erdös provided an easy and yet elegant lower bound for the Ramsey number $R(k, k)$.

The rich structure of random graphs captures some important properties of real-world networks, and meanwhile, makes it possible to mathematically analyze processes over the network and also to make rigorous predictions. These models have many applications in different fields including, but not limited to, network sciences, social sciences, life sciences, biological science, health studies, macroeconomics, etc. They have also been studied in a large body of research over many years, but despite this many mathematical challenges and open questions remain.

Among different phenomena associated with real-world networks, phase transitions particularly stand out. A phase transition refers to a sudden change in global properties of the network, when an underlying parameter changes from one side of a critical value to the other. Some examples of this phenomenon are: emergence of a giant connected network, belief propagation in social networks, spread of epidemic disease on networks, detectability of a hidden structure in networks, etc. Most random graphs experience a phase transition as well. We say a random graph model experience a phase transition if the behavior of a typical realization of this model is radically different when some parameter changes around its critical value. The focus of this thesis is to understand phase transitions in some recently introduced random graph models.

In short, my work rigorously studies the impact of macroscopic structural properties of random graphs and microscopic foundational level schemes for constructing them on the emergence of global-level phenomena. In this thesis, we study three specific problems related to this. First, we study the impact of community structure on the emergence of cascades, a phase transition where a large portion of the population adopts the new behavior, in threshold-based contagion models using modular random graphs generated by the configuration model and the differential equation method for the analysis. Next, using local weak analysis, we study a new graph model generated by bilateral agreement of individuals and analyze when a giant component emerges. We extensively study the associated branching process, using a point process perspective. Finally, using the objective method and motivated by particle tracking in physics and object tracking in videos, we study the detectability threshold of a hidden planted matching in a complete bipartite randomly weighted graph.

The rest of the introduction is organized as follows. In Section 1.2, we review some well-known random graph models. In Section 1.3, we discuss the main techniques that are used

in this thesis. Finally, in Section 1.4, we provide an overview of the thesis and summarize our contributions.

1.2 Some Well-Known Random Graph Models

There are many random graph models, each of which generates a random graph in a unique manner and defines a probability measure over the space of graphs. In general, a random graph model is defined in one of the following ways: either by specifying a sample space together with a probability measure, or by defining a random process that generates a graph. In this section, we discuss some well-studied random graph models from each of these two cases.

Erdös-Rényi Random Graph: The earliest random graph models are due to Erdös and Rényi in [2] and Gilbert in [3]. They introduced two closely related models: $G(n, m)$ and $G(n, p)$. The random graph model $G(n, m)$ is the uniform probability distribution over the set of all graphs with n nodes and m edges. The random graph model $G(n, p)$ is the distribution of random graphs of n nodes where each edge exists with probability p. The later is more common, probably because of the inherent independence structure, but these two models are closely related when $m = \binom{n}{2}p$.

Fixed Degree Sequence Random Graph: One natural modification of $G(n, m)$ is to fix the degree sequence rather than the total number of edges. The random graph model $G(n, \boldsymbol{d}_n)$ is the uniform probability distribution over the space of all graphs with n nodes and a fixed degree sequence $\boldsymbol{d}_n = (d_1, d_2, \cdots, d_n)$. Indeed, we need to assume $\boldsymbol{d}_n$ is graphic, that is to say, the space of graphs with degree sequence $\boldsymbol{d}_n$ is non-empty. One important question though, is how to generate such a random graph. The earliest and the most well-known generative model is due to Bollobás [5], which is referred to as the configuration model in the literature. This generative model realizes a graph by randomly pairing half-edges to form edge, where d_i half-edges are assigned to node i. Under some regularity conditions, the resulted graph is simple with positive probability. When these conditions hold, then conditioned on the graph begin simple, it is distributed as $G(n, \boldsymbol{d}_n)$.

Preferential Attachment Model: Some real-world networks, such as World Wide Web, are the outcome of some random process after it has been executed for a period of time. The preferential attachment model tries to capture this behavior. The construction begins with a fixed graph G_0. At time step $t + 1$, the random graph G_{t+1} is generated by adding a node v with degree m to G_t. The probability that the node w is connected to the node v in G_{t+1} is proportional to the degree of w in G_t. Hence, node v picks m neighbors from G_t according

to a multinomial distribution with parameters m and $p_1, p_2, \cdots, p_{N_t}$, where $p_i \propto \deg(i, G_t)$ and N_t is the total number of nodes in G_t. This random graph model was first considered by Price [6, 7] in the study of the citation network. Krapivsky, Redner and Leyvraz studied various properties of a related model, called the growing random networks, in [8] and [9]. The name "preferential attachment" is due to Barabási and Albert [10], who proposed the application of this model to the growth of the World Wide Web.

The above three models are the most well-known and heavily studied random graph models in the literature. There are many other random graph models such as random intersection graphs, random geometric graphs, small world model, etc. For an application based survey on different types of random graph models see [11], and for a more rigorous study see [12], or more recently [13] and [14].

1.3 Mathematical Tools and Methods

There are various tools to study a random graph model. The first and probably the most well-known is the first and the second moment method. However, especially due to the increased use and development of martingale methods in the recent decades, there are many more rigorous techniques to understand and predict the behavior of random graphs, e.g., concentration inequalities, differential equations method, entropy based analysis, etc. In this section, we provide an overview of the methods that we use in this thesis.

The first and the second moment methods: As their name suggests, these two methods are concerned with the first and the second moments of a random variable. Consider a non-negative integer-valued random variable X. The first moment method is an easy application of Markov inequality to bound the probability of the event $\{X > 0\}$: $P(X > 0) \leq \mathbb{E}[X]$. The second moment method involved a more careful analysis and typically uses the Paley-Zygmund inequality that can be proved directly using the Cauchy-Schwarz inequality: $P(X > 0) \geq (\mathbb{E}[X])^2/\mathbb{E}[X^2]$. The main applications of these two methods is to bound the probability of of the event $\{X = 0\}$. The first moment method is used to prove $X = 0$ with high probability, while the second moment method is used to prove $X > 0$ with high probability.

One application of these methods is the disappearance of isolated nodes in $G(n, p)$ for $p = c \log n/n$. Let $X = I_1 + I_2 + \cdots + I_n$ denote the total number of isolated nodes, where I_i is an indicator random variable indicating whether node i is isolated or not. Using the first moment method, it is easy to see that for $c > 1$ we have $P(X > 0) \to 0$. On the other hand,

using the second moment method, for $c \leq 1$ we have $P(X > 0) \to 1$. Hence, there is a phase transition at the critical value $c = 1$ for $G(n, c \log n/n)$ random graph model.

While the simplicity of these methods has resulted in their immense popularity and success [15, 16], they are not always applicable. The reason usually is the difficulty of finding the first and/or the second moments of the random variable of interest. In chapter III, we will see that even finding the first moment can be very challenging when the structure of the random graph is interdependent. We face the same problem in chapter IV to find the second moment.

Differential Equation Method: This method provides a deterministic approximation for the dynamics of a random process. It was originally introduced by Kurtz [17] to study a continuous time random process. The earliest application of the differential equation method to random graphs is due to the work of Karp and Sipser [18]. They used this method in the analysis of a random greedy matching algorithm on $G(n, \lambda/(n-1))$, where $\lambda > 0$ is a constant.

The basic idea is simple: calculate the expected drift of the random variables of interest, treat these expected changes as continuous functions, and write down the corresponding differential equations suggested by these functions. The analysis uses large deviations results and concentration inequalities to show the concentration of the random variables around the solution of the differential equations. We use this method in chapter II to approximate belief propagation in random graphs with community structure, generated by configuration model. For an extensive survey and detailed discussion see [19].

The Objective Method: This method is applicable to a variety of combinatorial optimization problems on random structures. The basic idea is as follows: identify an infinite probabilistic object whose local properties are the same as the local limiting properties of a class of finite problems, solve the optimization problem on this infinite object, and then use this solution to construct a near-optimal solution to the class of finite problems as the size increases without bound. Intuitively speaking, this method works if there is a correlation decay; that is to say, the structure of the optimal solution are mostly given by the local properties of the problem. The first rigorous application of this method is due Aldous [20, 21]. In a series of papers, he proved the expected cost of the minimum matching on a complete bipartite graph with uniformly distributed edge weights converges to $\pi^2/6$; this established one of Mézard and Parisi's conjectures [22].

In chapter III, we use this method to identify the corresponding infinite object of the sparse random graph introduced in [23]. In chapter IV, we closely follow the work of Aldous [20, 21]

to find a hidden planted matching in a complete bipartite graph with random edge weights. See the survey by Aldous and Steele [24] and the lecture notes of Bordenave [25] for a detailed discussion of this method.

Branching Processes: The idea of branching processes was first introduced by Bienaymé [26] and more formally by Galton and Watson [27] to study the extinction of family names. A branching process is a mathematical model for population dynamics. In the case of the Galton-Watson branching process, there is only one individual in generation 0, and each individual in generation n produces an *i.i.d.* random number of individuals in generation $n+1$ (with 0, i.e., death, also allowed). This model arises naturally in the study of locally tree-like random graphs [28, 29].

In chapter III, we extensively study the branching process associated with the sparse random graph introduced in [23]. Because of the interdependent structure of the associated branching process, the common techniques such as the first and the second moment method fail to provide any meaningful insight. Instead, we use the point process perspective introduced by Harris [30, Chapter 3] to study this branching process. An in-depth exposition of branching processes can be found in the books by Harris [30] and Athreya and Ney [31].

1.4 Summary of Thesis

Using the techniques discussed in Section 1.3, we study phase transitions in three new random graph families. We will discuss each of these techniques in greater detail in the related chapters. For now, we provide the summary of the main technical chapters of the thesis.

1.4.1 Chapter II: Impact of Community Structure on Cascades

In this chapter, we study the threshold model of cascades on random graphs with community structure. The threshold model is widely used to study the propagation of opinions, behaviors and technologies in social networks, see e.g., [32]. In this model, individuals drop the old behavior and adopt the new behavior based on how many neighbors have already chosen it. A small portion of the population are early adopters, who introduce the new behavior to the network. A cascade is said to happen if the number of individuals adopting the new behavior is substantially greater than the number of early adopters.

Specifically, we consider the permanent adoption model where individuals that have adopted the new behavior cannot change their state. We study cascades under the threshold

model on sparse random graphs with community structure to see whether the existence of communities affects the number of individuals who finally adopt the new behavior.

This problem has been rigorously analyzed when the underlying network is given by a single community, see e.g., [33–36]. In the physics literature, however, this problem has been studied in the case of multiple communities using heuristic methods, see e.g., [37,38]. In this work, we close the gap by providing a rigorous analysis of this problem on random graphs with community structure. These random graphs are given by a concatenation of multiple sparse random graphs and bipartite random graphs.

A heuristically driven cavity-method like approximation developed in the physics literature [39,40] suggests that the number of individuals who eventually adopt the new behavior is related to a fixed point of a particular set of mean-field equations. This approximation is based on the fact that sparse random graphs are locally tree-like. In the case of two communities, this mean-field approximation suggests the final proportion of adopters is $\boldsymbol{\phi} = \boldsymbol{\Phi}(\boldsymbol{\mu})$ where $\boldsymbol{\mu} = (\mu^{(1,1)}, \mu^{(1,2)}, \mu^{(2,1)}, \mu^{(2,2)})$ is a fixed point of the mean-field equations $\boldsymbol{\mu} = \boldsymbol{F}(\boldsymbol{\mu})$. The components of functions $\boldsymbol{\Phi}(\cdot)$ and $\boldsymbol{F}(\cdot)$ are defined as follows:

$$F_{(j,j)}(\boldsymbol{\mu}) = \mathbb{E}\bigg[\frac{D_j}{\mathbb{E}(D_j)}\,(1-\alpha_j(D_j,D_m))\,\mathbf{1}\Big\{\sum_{i=1}^{D_j-1} Y_i^{(j,j)} + \sum_{i=1}^{D_m} Y_i^{(j,-j)} \le K_j(D_j,D_m)\Big\}\bigg],$$

$$F_{(-j,j)}(\boldsymbol{\mu}) = \mathbb{E}\bigg[\frac{D_m}{\mathbb{E}(D_m)}\,(1-\alpha_j(D_j,D_m))\,\mathbf{1}\Big\{\sum_{i=1}^{D_m-1} Y_i^{(j,-j)} + \sum_{i=1}^{D_j} Y_i^{(j,j)} \le K_j(D_j,D_m)\Big\}\bigg],$$

$$\Phi_j(\boldsymbol{\mu}) = \mathbb{E}\bigg[(1-\alpha_j(D_j,D_m))\,\mathbf{1}\Big\{\sum_{i=1}^{D_j} Y_i^{(j,j)} + \sum_{i=1}^{D_m} Y_i^{(j,-j)} \le K_j(D_j,D_m)\Big\}\bigg],$$

where for every $j', j \in \{1,2\}$, $Y_i^{(j,j')}$ are *i.i.d.* Bernoulli random variables with parameter $1-\mu^{(j,j')}$, D_j is a random variable with the community j degree distribution, D_m has the inter-community degree distribution, $K_j(\cdot,\cdot)$ is the threshold function for nodes in community j, and $\alpha_j(D_j,D_m)$ is the probability of a node in community j to be an early adopter, given her neighbors in community j and $-j$ to be d_j and d_m respectively. Notationally, we use $\{-j\}$ to denote $\{1,2\} \setminus \{j\}$.

As the first step to prove the validity of the mean-field equations, we construct a Markov chain that couples the evolution of the adoption process with the process of generating the random graph. We then calculate the expected drift of the random variables associated with this Markov process, treating them as continuous functions. However, there is a caveat here, as the resulted continuous functions are not Lipschitz. This makes the analysis very challenging, since to apply the differential equation method we need to have Lipschitz continuous functions.

One way to fix this is to consider a truncated version of the original process, in which we assume the state of nodes with high degrees does not change during the process. We also have to stop the process before it terminates, to guarantee the Lipschitz property. We can then use the differential equation approach to approximate the sample-path of the truncated version of the Markov process of adoption, using a finite set of ordinary differential equations (ODEs) given by these Lipschitz continuous functions. Next, we prove that the dimension of this set of ODEs can be reduced significantly (in the case of 2 communities, the new dimension is 4). This "surprising" dimension reduction has also been observed in the case of one community [33,34]. We sketch a plausible reason behind this simplification by providing a probabilistic solution to this set of ODEs.

We then show that in the case of two communities, the trajectory of the simplified ODEs is the same as the trajectory of the following ODEs:

$$\frac{d\boldsymbol{\mu}}{dt} = \boldsymbol{F}(\boldsymbol{\mu}) - \boldsymbol{\mu} \qquad \boldsymbol{\mu}(\mathbf{0}) = \mathbf{1}, \qquad \boldsymbol{\mu} \in [0,1]^4. \tag{1.1}$$

By analyzing this set of finite dimensional ODEs, we prove the validity of the mean-field equations. We also show that the absorbing point of this ODEs is given by $\boldsymbol{\mu}_* = \lim_{s\to\infty} \boldsymbol{F}^s(\mathbf{1})$, where $\mathbf{1} = (1,1,1,1)$ and $\boldsymbol{F}^s$ is the sth iteration of the function $\boldsymbol{F}$. This absorbing point is a fixed point of the mean-field equations, and we show that this fixed point is the correct fixed point to study for understanding the properties of contagion when $\boldsymbol{\mu}_*$ is a stable equilibrium of the ODEs in (1.1).

In addition, for the case of linear thresholds (as a function of the total degree) we characterize both necessary and sufficient conditions for contagion to happen no matter how small the set of initial adopters is. Specifically, we prove that a contagion happens if and only if $\rho\big(\boldsymbol{J}_{\boldsymbol{F}}(\mathbf{0},\mathbf{1})\big)$ is greater than 1, where $\rho(\cdot)$ is the Perron-Frobenius eigenvalue of a non-negative matrix and $\boldsymbol{J}_{\boldsymbol{F}}(\mathbf{0},\mathbf{1})$ is the Jacobian matrix of $\boldsymbol{F}$ at point $\mathbf{1}$ when $\alpha_j(\cdot,\cdot)$ is set to be zero for all $j \in \{1,2\}$.

We also investigate the problem of optimum seeding given a budget constraint, and propose a gradient-based heuristic seeding strategy, that numerically, dispels commonly held beliefs in the literature that suggest the best seeding strategy is to seed over the nodes with the highest number of neighbors.

Our preliminary results were presented at EC'2016 [41] with a more detailed version on arXiv [42]. This work is in collaboration with Vijay Subramanian, Mingyan Liu, and Marc Lelarge.

1.4.2 Chapter III: Erlang Weighted Tree

In this chapter, we propose a new branching process which we named Erlang Weighted Tree(EWT). This branching process arises naturally in the study of a locally tree-like random graph model proposed in [23] by La and Kabkab. This model tries to capture the self-optimizing behavior of individuals in which links are made based on the cost/benefit of the connection. Cooper and Frieze [43] analyzed a related model, in which a one-sided interest is enough for a connection to be established. We analyze the asymptotic behavior of this graph model [23] in the sparse setting as the number of nodes increases without bound. We prove that the model locally weakly converges [24, 44] to the rooted tree associated with the EWT. We then study the main properties of the EWT such as the probability of extinction, the emergence of phase transition and growth rate.

The graph construction starts with a complete graph $K_n = ([n], E_n)$, a fixed sequence of positive integers $\boldsymbol{d}_n = (d_1(n), d_2(n), \cdots, d_n(n))$, and a random function C_n that assigns independent and exponentially distributed random variables with mean n to the edges of K_n. The empirical distribution of $\boldsymbol{d}_n$ converges to some probability distribution $P(\cdot)$ as $n \to \infty$. The value of d_i indicates the number of neighbors that node i wants to connect to, and the value assigned to each edge by C_n represents the cost of the edge. Thereafter, each node i selects the d_i lowest cost incident edges and declares them to be preferred edges. The random graph $G_n = ([n], \widetilde{E}_n)$ is constructed by keeping the edges of E_n that are preferred by both end nodes, where $\widetilde{E}_n$ is the set of edges that are kept.

We study the local weak limit of $\mathbb{E}(U_{G_n})$, where U_{G_n} is the random measure associated with the connected component of a uniformly selected node in G_n over the space of rooted graphs. Note that U_{G_n} inherits the randomness of C_n, and thus is a random measure. Moreover, the uniform choice of the root in $\mathbb{E}(U_{G_n})$ implies some modest symmetry which is preserved under the weak limit. This property is called unimodularity which is equivalently described via a Mass-Transport Principle [44, 45]. Intuitively speaking, unimodularity means the expected mass that the root receives is equal to the expected mass it sends out. This notion is also equivalent to involution invariance: a measure over the space of rooted graphs is said to be involution invariant if it is invariant with respect to the root swapping operation which swaps the root node with one of its uniformly selected neighbors [24, 25, 44]. We prove $\mathbb{E}(U_{G_n})$ converges weakly to $Er(P)$, the unimodular probability measure associated with the EWT. Then we analyze the main properties of the EWT.

The branching process EWT is defined as follows. Let $\mathbb{N}^f$ denote the set of all sequences of positive integers of length k together with $\emptyset$. For each $\boldsymbol{i} \in \mathbb{N}^f$, define the following set of random variables: 1) $n_{\boldsymbol{i}}$ which is a non-negative integer distributed as $P(\cdot)$ if $\boldsymbol{i} = \emptyset$ and $\widehat{P}(\cdot)$ otherwise, where $\widehat{P}(\cdot)$ is the shifted version of $P(\cdot)$, i.e., $\widehat{P}(k) = P(k+1)$ for all $k \geq 0$;

2) $v_{\boldsymbol{i}}$ which is an Erlang random variable[1] with parameters $n_{\boldsymbol{i}}+1$ and 1; and 3) $\{\zeta_{(\boldsymbol{i},j)}\}_{j=1}^{n_{\boldsymbol{i}}}$ which are *i.i.d.* uniformly distributed random variables on $[0, v_i]$ conditioned on v_i. Insert an edge between $\boldsymbol{i}$ and $\boldsymbol{i}j$ if $j \leq n_{\boldsymbol{i}}$ and $v_{\boldsymbol{i}j} \geq \zeta_{\boldsymbol{i}j}$, where $\boldsymbol{i}j \in \mathbb{N}^f$ is the concatenation of the sequence $\boldsymbol{i} \in \mathbb{N}^f$ and the positive integer $j \in \mathbb{N}$. The Erlang Weighted Tree is the connected component of $\emptyset$, and $Er(p)$ is the measure generated by EWT on the space of rooted graphs.

Unlike the Galton-Watson branching process, the degree distribution of nodes in EWT depends on its generation (number of edges between the root and the node). This is closely related to the bilateral agreement required for a connection to survive. As a result, it is not possible to find a closed-form expression for the degree distribution of any node other than the root. The interdependent structure also causes the first and second moment methods to fail. Hence, to study this branching process we take a different approach, namely, a point process perspective. This viewpoint was developed by Harris in [30] to study general branching processes. He proved that, under some mild conditions, the growth/extinction operator scales with its largest eigenvalue; establishing existence and positivity, and determining the multiplicity of the eigenvalue is an integral part of the results.

The EWT does not satisfy the conditions required in [30], hence, it needs a separate treatment. Using the point process perspective, we first simplify the growth operator and show that it is sufficient to analyze a compact and bounded liner operator mapping a Hilbert space of L^2 functions from $\mathbb{R}_+$ to itself. Next, we show the existence of an eigenvalue with non-negative eigenfunction for the simplified operator. To prove that this eigenvalue is the unique Krein-Rutman eigenvalue [46, 47], we define a continuous state Markov process and study the convergence of this Markov process to its stationary distribution using Baxandale's theorem [48]; this Markov process is reversible, which is closely related to the fact that EWT is unimodular. Therefore, we derive the growth/extinction rate (denoted as β_0) of the EWT by proving the existence and the uniqueness of the Krein-Rutman eigenvalue β_0 and characterizing the corresponding eigenfunction f_0. Based on this analysis, we show that the expected value of Z_l, the number of nodes that are l hub away from the root, scales as follows:

$$\frac{E[Z_l]}{{\beta_0}^l} \xrightarrow{l\to\infty} \left(\sum_{m=1}^{\infty} P(m) \int_{x=0}^{\infty} \frac{e^{-x}x^m}{m!} \times \frac{m}{x} f_0(x)\, dx\right) \left(\int_0^{\infty} \sum_{k=1}^{\infty} P(k) \frac{e^{-z}z^{k-1}}{(k-1)!} f_0(z)\, dz\right).$$

Moreover, we prove that in the case of $\beta_0 > 1$, the random variable $Z_l/{\beta_0}^l$ converges almost

[1]The probability density function of an Erlang random variable with parameter i is given by

$$f(x) = \frac{e^{-x}x^{i-1}}{(i-1)!}.$$

surely and in L^2 to a random variable W, and in particular $Z_l \sim {\beta_0}^l W$.

We also study the probability of extinction. We prove that the probability of extinction is given by a fixed point of another operator. However, this analysis does not relate β_0 with the probability of extinction. To establish the classical connection between β_0 and the probability of extinction, we again combine the analysis of this operator with the point process perspective. We then prove this operator has a unique non-trivial fixed point. Finally, we relate the probability of extinction with β_0: $\beta_0 > 1$ if and only if $\Pr(\text{extinction}) < 1$. We leave the connection between the probability of extinction and the finite graph model for future work as the following conjecture: if $\Pr(\text{extinction}) < 1$, then the size of the giant component in the finite graph is proportional to $1 - \Pr(\text{extinction})$.

A preliminary version of this work was presented as an invited paper at the 2018 Allerton conference [49], with a more detailed version on arXiv [50]. This is a joint work with Vijay Subramanian, Mingyan Liu, and Rajesh Sundaresan.

1.4.3 Chapter IV: The Planted Matching Problem

In this chapter, we study the problem of recovering a planted matching in randomly weighted complete bipartite graphs $K_{n,n}$. For some unknown perfect matching M^*, the weight of an edge is drawn from one distribution P if $e \in M^*$ and another distribution Q if $e \notin M^*$. The question then is whether it is possible to find the planted matching or not – more specifically, to determine the conditions when it is possible to almost exactly recover the planted matching (i.e., find a matching which coincides with the planted matching on almost all the edges except $o(n)$ ones). This model is motivated by the problem of tracking moving objects in a video, such as flocks of birds, motile cells, or particles in a fluid [1].

Automated tracking of mobile objects/particles is of broad interest in the natural sciences as well as in robotics and security or surveillance activities. The goal is to extract the identities of the objects/particles from successive images of the system, obtained at an appropriate rate, and then to track them. Ideally, the rate at which the images are sampled is high, and the density of particles is low, so it is possible to easily track the particles. But when the density of particles increases or if the sampling is not fast enough, this task becomes much more difficult and in some cases even impossible. Chertkov et al. [1] studied this problem and proposed a canonical mathematical model. The model consists of a complete bipartite graph with random edge weights–nodes at each side of the bipartite graph model the particles, and edge weights model the distance between them in two successive images. There is an exact matching M^* which matches each particle on the left side with the corresponding one on the right side of the bipartite graph.

In this chapter we study this canonical model rigorously using local weak analysis and the

objective method [24,44]. For the general model, we first identify the asymptotic object to be studied as a rooted infinite tree together with a planted matching imposed on it. We call this object the planted Poisson Weighted Infinite Tree (planted PWIT), because of its natural connection to the Poisson Weighted Infinite Tree [24]. We relate the minimum-weighted matching on the complete bipartite graph, $M_{\min}$, to a class of matchings on the planted PWIT using an approach similar to that pioneered in [20, 21, 24]. This class of matchings needs to satisfy a modest symmetry, namely involution invariance; again, this means that the matching is not affected by swapping the root of the infinite tree "uniformly" with one of its neighbors.

We study the minimum-weighted involution invariance matching, $\mathscr{M}_{\infty,\text{opt}}$, on the planted PWIT. This matching is related to the following recursive distributional equation:

$$\begin{aligned} X &\stackrel{d}{=} \min(\{\zeta_i - Y_i\}_{i=0}^{\infty}), \\ Y &\stackrel{d}{=} \min(\eta - X, \{\zeta_i - Y_i\}_{i=0}^{\infty}) \end{aligned} \tag{1.2}$$

where X is independent of everything else, $\{Y_i\}_{i=1}^{\infty}$ and Y are *i.i.d.*, $\{\zeta_i\}_{i=1}^{\infty}$ are the arrivals of a Poisson process with rate 1, and η has the same distribution as the planted edges. We prove this matching is unique as long as the system of recursive distributional equations (1.2) has a non-trivial solution. We then relate the $\mathscr{M}_{\infty,\text{opt}}$ on the planted PWIT to the $M_{\min}$ on the complete bipartite graph in two steps. First, using a simple compactness argument, we prove that any subsequence of the minimum-weighted matchings on the complete bipartite graph converges to an involution invariant matching on the planted PWIT in the local weak sense. Next, we construct a matching on the finite graph with cost converging to the $\mathscr{M}_{\infty,\text{opt}}$ on the planted PWIT. Given the uniqueness property of the $\mathscr{M}_{\infty,\text{opt}}$, we deduce that the minimum-weighted matching on the complete bipartite graph converges to the minimum-weighted involution invariant matching on the planted PWIT in the local weak sense.

In the remainder of this chapter, we specialize our results to the case when the edge weights of $K_{n,n}$ are independent exponentially distributed random variable, where the rate parameter of the distribution is given by $\lambda > 0$ for planted edges and $1/n$ for all the other edges; that is to say, $P = Exp(\lambda)$ and $Q = Exp(1/n)$. In this case the maximum-likelihood estimator of M^* is the minimum-weighted matching. Chertkov et al. [1] used the cavity method to analyze this model when P is a folded Gaussian and Q is the uniform distribution over $[0, n]$. They observed there is a phase transition after which the $M_{\min}$ cannot almost exactly recover the planted matching.

Our analysis reveals that almost exact recovery using an algorithm that find the minimum weighted matching is not possible if the recursive distributional equation has a finite solution.

In particular, if the edge weights of the planted matching are distributed as $Exp(\lambda)$, the minimum weighted matching algorithm fails to almost exactly recover M^* for $\lambda < 4$. On the other hand, using the first moment method, we prove that the $M_{\min}$ almost exactly recovers M^* for $\lambda \geq 4$. We also calculate the expected overlap for $\lambda < 4$:

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}|M_{\min} \cap M^*| = 1 - 2\int_0^\infty (1 - F(x))\,(1 - G(x))\,V(x)W(x)\,dx < 1\,,$$

where (F, G, V, W) is the unique solution to the following coupled system of ODEs

$$\frac{\mathrm{d}F}{\mathrm{d}x} = (1 - F(x))(1 - G(x))V(x), \qquad \frac{\mathrm{d}V}{\mathrm{d}x} = \lambda(V(x) - F(x)),$$
$$\frac{\mathrm{d}G}{\mathrm{d}x} = -(1 - F(x))(1 - G(x))W(x), \qquad \frac{\mathrm{d}W}{\mathrm{d}x} = \lambda(G(x) - W(x)),$$

with the boundary conditions

$$F(-\infty) = V(-\infty) = G(+\infty) = W(+\infty) = 0,$$
$$F(+\infty) = V(+\infty) = G(-\infty) = W(-\infty) = 1.$$

This work was presented as an invited talk at Allerton conference 2019, with a more detailed version on arXiv [51]. This is a joint work with Cristopher Moore and Jiaming Xu.

1.5 Summary of Contribution

In summary, the major contributions of this thesis, ordered in terms of the chapters, are as follows:

• Using the differential equation method, we provide a tight approximation for the evolution of the stochastic process adoption under the threshold model. Our analysis also proves the validity of the cavity method for finding the final proportion of adopters. Using this method, we are able to analyze the impact of advertising by means of seeding of the nodes with the new technology or opinion.

• Using a point process perspective, we extensively analyze a novel branching process. We derive the main properties of this branching process such as the probability of extinction, emergence of phase transition, growth/extinction rate, etc. We also prove that this branching process is related to the random graph model introduced in [23].

• Using the machinery of local weak convergence, we study the detectability of a planted matching in a complete randomly weighted bipartite graph. We explicitly calculate the expected overlap between the minimum matching and the planted matching when the edge

weights of the planted matching are distributed as $Exp(\lambda)$. We also show that there is a phase transition at $\lambda = 4$, before which the minimum matching cannot almost perfectly recover the planted matching.

Notation of the Thesis: Notation varies across chapters. However, each chapter is self-contained and the notation of each chapter is defined in that chapter. Appendices follow the notation of the corresponding chapter.

CHAPTER II

Impact of Community Structure on Cascades

Notation: Random variables are denoted by capital letters (sometimes using a bold typeset too); realizations or deterministic quantities are in small letters. Vectors are denoted by using a bold typeset and individual components without it. Adhering to game theoretic notation, a node's community is denoted by $j \in \{1, 2\}$ and the other community by $-j = \{1, 2\} \setminus \{j\}$. Multigraphs are denoted by an asterisk and simple graphs without one. The words *community* and *side* are used interchangeably.

2.1 Introduction

In this chapter, we investigate a type of cascade problem on graphs that has been used to study the spread of new technology or opinions in social networks, see e.g., [32, 52–56]. This spread is also referred to as a contagion in networks. The underlying model typically consists of a few (selected) initial adopters (nodes in the network) or "seeds" and a particular adoption model that determines the condition under which a node will choose to adopt given the states of its neighbors. A commonly studied model is the threshold model [57, 58], whereby individuals adopt the new technology based on how many neighbors have already chosen it.

Prior work in this area has generally focused on analyzing what happens when the underlying network is given by a single community modeled as a sparse random graph, either heuristically, see e.g., [58, 59], or more rigorously, see e.g., [33–36]. In this work, we instead consider graphs with a type of community structure (also known as modular networks), whereby multiple sparse random graphs are weakly interconnected. This could model for instance segments of the population (e.g., different age or ethnic groups), where members of a single segment are more strongly connected (with a relatively high node degree) and cross-segment connections are weak, i.e., fewer members are connected to those from a different segment. This would be a more realistic and interesting model for many practical

scenarios and serves as a natural next step to the studies done with a single community. We are particularly interested in whether the existence of communities affects the number of individuals who eventually adopt the new technology. Also of interest is the question whether seeding in all communities is a better strategy in terms of maximizing the number of eventual adopters than exclusively in one community or, in particular, the optimum seeding strategy given budget constraints. While earlier works have looked at this problem using heuristic methods, see e.g., [37, 38, 59–61], we present a mathematically rigorous analysis of this problem.

Specifically, we consider the permanent adoption model where nodes that have adopted the new technology cannot change their state. Our analysis presents a differential-equation-based tight approximation to the stochastic process of adoption under the threshold contagion model. While this is a similar approach to the original analysis of contagions in a single community [34], the additional community structure requires significant technical development to establish the validity of this approach in the new scenario. The analysis of the differential equation leads to a correctness proof of a mean-field equation for the contagion in a large network, as well as an algorithm to calculate the properties of the contagion. Using this analysis, we are able to analyze the impact of advertising by means of seeding of the nodes with the new technology or opinion. The differential equation also leads to a characterization of the sample-path of the adoption process as well as a sharp characterization of the contagion threshold for the linear threshold model.

Our main contributions can be summarized as follows.

1. We prove the validity of a mean-field analysis of the contagion process over infinite trees. This analysis yields a fixed point equation whose solution can be used to exactly determine the final fraction of the population that are eventual adopters (the size of the cascade). Furthermore, when the fixed point equation has multiple solutions, we identify the correct solution among these and provide an algorithmic means to calculate it.

2. We provide a tight differential equation approximation to the sample-path of the contagion process. This allows us to track the evolution of the cascade, in particular, how it moves from one community to the other. We also provide a probabilistic approach to solve the differential equation, which also explains the "surprising" dimension reduction observed in [33, 34]. Specifically, we show that the solution of the differential equation can be obtained by solving a much simpler k^2-dimensional differential equation, where k is the number of communities. This dimension reduction is crucial to developing a comprehensive understanding of the contagion process.

3. For general thresholds we provide a sharp characterization of the contagion threshold—the condition on the thresholds for which a contagion occurs with a finite set of seed nodes—in terms of the Perron-Frobenius eigenvalue of an associated matrix.

4. Specializing to Poisson degree distributions and linear thresholds we prove that the community structure does not matter for global properties like the contagion threshold for the linear threshold model of [57,58]. In particular, when seeding a small number of agents with the new technology, we find that the community structure has little effect on the final proportion of adopters.

5. We numerically study the impact of the community structure on the viral seeding of nodes. We find that seeding a fraction of population with the new technology has a significant impact on the cascade, with the optimal seeding strategy depending on how strongly the communities are connected.

6. We propose a gradient-based heuristic seeding strategy to maximize the size of the cascade given budget constraints. Empirically, the algorithm dispels all commonly held beliefs in the literature that suggest the best seeding strategy is to seed over the nodes with the highest number of neighbors. Most notably, we can demonstrate many cases wherein our seeding algorithm achieves a global cascade, while seeding over the nodes with highest degree fails to spread much further from the seeds.

The remainder of this chapter is organized as follows. We present our model in Section 2.2 and a literature review in Section 2.3. In Section 2.4 we present a mean-field approximation of the adoption process, whose validity is then established in Sections 2.5 through 2.9: in Section 2.5, we construct a Markov process coupling the evolution of the adoption process with the process generating the random graph; this Markov process is then approximated using a set of Ordinary Differential Equations (ODEs) in Section 2.6; we then provide a probabilistic approach to solve this set of ODEs in Section 2.7; the analysis of the ODEs is presented in Sections 2.8 and 2.9. We discuss the results on the contagion threshold for general thresholds in Section 2.10. Most of the results are specialized to the case of Poisson degree distributions in Section 2.11. We present numerical results and discuss the optimal seeding strategy in Section 2.12.

2.2 System Model

Consider a set $[n] = \{1, \ldots, n\}$ of agents that are organized into two communities, community 1 $\{1, 2, \ldots, n_1\}$ and community 2 $\{n_1 + 1, \ldots, n\}$ with $n_2 := n - n_1$ individuals. Assume that we are given three sequences of non-negative integers: $\mathbf{d}_1 = (d^n_{1,i})_1^{n_1}$,

$\mathbf{d}_2 = (d^n_{2,i})^n_{n_1+1}$, and $\mathbf{d}_m = (d^n_{m,i})^n_1$, which satisfy the following conditions: 1) $\sum_{i=1}^{n_1} d^n_{1,i}$ is even; 2) $\sum_{i=n_1+1}^{n} d^n_{2,i}$ is even; and 3) $\sum_{i=1}^{n_1} d^n_{m,i} = \sum_{i=n_1+1}^{n} d^n_{m,i}$. The sequence $\mathbf{d}_j$ is the degree sequence of the sub-graph for community j for $j \in \{1, 2\}$ and $\mathbf{d}_m$ is the degree sequence of the bipartite graph connecting the two communities.

Construct a two-community random multigraph (allowing for self-loops and multiple links) with given degree sequences $\mathbf{d}_1$, $\mathbf{d}_2$ and $\mathbf{d}_m$ generated by the configuration model [12] as the concatenation of $G^*(n_1, \mathbf{d}_1)$, $G^*(n_2, \mathbf{d}_2)$ (both generated via the configuration model) and a random bipartite multigraph $G^*(n_1, n_2, \mathbf{d}_m)$: generate half-edges for each node corresponding to the different degree sequences and combine the half-edges into edges by a uniform random matching of the set of half-edges of each sequence. Conditioned on the random multigraphs and the random bipartite graph being simple graphs, we obtain uniformly distributed random graphs $G(n_1, \mathbf{d}_1)$, $G(n_2, \mathbf{d}_2)$, and $G(n_1, n_2, \mathbf{d}_m)$ with the given degree sequences. The concatenation of these produces a simple two-community graph $G(n, \mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_m)$ with the desired distributions. In Definition II.4 we impose standard regularity assumptions [62] on the degree sequences so that the resulted multigraphs are simple with positive probability. We assume that $\lim_{n\to\infty} n_1/n = \beta$ (equivalently $\lim_{n\to\infty} n_2/n = 1 - \beta$). The stochastic block model [63] is a prototypical example of a two-community graph.

Following Lelarge [35], we analyze the threshold model of Morris [57] and Watts [58] on the two-community random graph model described above. In this model, nodes have the choice between two types of opinions/technologies, A and B; we sometimes also use "inactive" to denote type A and "active" to denote type B. All nodes initially start in type A, i.e., are inactive. Each node has a threshold that is a function of its community and degrees (in the same community and across to the other community); the value of the threshold is fixed and allowed to be any non-negative real number. If a node finds that the number of its neighbors (across both communities) who have chosen type B is greater than its threshold, then it will permanently choose to switch to type B. Again following [35] we initially seed nodes with type B using a Bernoulli random variable (1 implying that a node gets seeded with type B) that is independently chosen with the mean depending on the node's parameters, namely, community and degrees. Note that a degree and/or community-unaware seeding strategy would imply an appropriate uniformity in the means of the seeding random variables. After the seeding process is completed, the remaining nodes then react to the seed nodes and decide whether to adopt type B. This process continues until a final state of the nodes is reached. A cascade is said to happen if the number of nodes adopting type B is substantially greater than the seed set.

2.3 Literature Review

The threshold model [32, 52–54, 64] is a well accepted model for explaining the adoption of a new technology, opinion or behavior in a population that interacts via a social network. The linear threshold model, where the threshold is a function of the degree, was analyzed for the contagion threshold for specific graphs in [57], and using heuristically derived formulae for single community random graphs in [58, 59]. The results on the single community random graphs were rigorously proved using branching processes in [35], where the importance of pivotal players (those whose degree is low enough that one neighbor will make them adopt the new behavior) was identified and studied. Similar results were derived using the differential equation method in [33, 34], and in [36] for the non-permanent adoption model.

The threshold model has been studied for networks with communities, but using heuristically derived mean-field approximations and approximate differential equations [37, 38, 59–61]. In these studies, it was numerically shown in [37, 38] for the linear threshold model that the community structure leads to a different dynamic in terms of the evolution of the cascade itself. It is important to note that the authors in these works postulate both the mean-field equation and the differential equations in an *ad hoc* manner without a formal proof. This is particularly the case for the multi-community work in [37, 38] where the authors combine the adoption processes in the different communities without a proper mathematical justification.

The problem of maximizing influence propagation in networks, by targeting certain influential nodes that have the potential to influence many others, has been an important follow-up problem [32, 52–54] once the impact of a social network on behavior adoption was discovered. While this problem is known to be NP hard for many influence models, several approximate methods have been designed, see e.g., [65, 66]. A contrasting strategy to identifying and targeting influential nodes is to use viral marketing [52, 53, 55]. A randomized version of viral marketing, also referred to as seeding or advertising in the chapter, was studied in [34, 35] where the resulting cascade was precisely identified. The results in [35] also suggested that targeting higher degree nodes is a better seeding strategy over degree-unaware random seeding. With community structure, [38, 60, 61] showed that the seeding strategies could be dramatically different from the one-community optimal strategies. Typically asymmetric seeding strategies, wherein the seeding is principally carried out in one community over another, were shown to perform better than more uniform (over the communities) seeding strategies.

2.4 Mean-Field Approximation

We start by presenting a mean-field approximation of the process of adoption of type B, i.e., becoming active, in a typical simple graph generated through the configuration model described in Section 2.2. The graphs that we consider are locally tree-like [67] so that the structure up to any finite depth when viewed from a uniformly selected node of the graph is a tree with high probability. Therefore, the local structure of a simple graph produced by configuration model converges [24, 25, 44] to a rooted unimodular Galton-Watson Multi-type Tree (GWMT_*). In a rooted unimodular Galton-Watson Tree, the degree distribution of any non-root vertex is the size-biased/sampling-biased distribution of the root: for a random variable $D \in \mathbb{Z}_+$ with distribution $\mathbb{P}(\cdot)$ and finite mean $\mathbb{E}[D]$, the size-biased/sampling-biased distribution $\mathbb{P}_*(\cdot)$ is given by $\mathbb{P}_*(d) = d\mathbb{P}(d)/\mathbb{E}[D]$ for all $d \in \mathbb{Z}_+$. In the case of GWMT_*, the degree distribution of each child depends on the community of its parent. The joint degree distribution of the root node is $\mathbb{P}_{j,m}$ if the community of the root is $j \in \{1, 2\}$. The degree distributions of each child is then given by the size-biased/sampling-biased distribution for the community of the parent and the regular distribution for the other community. In particular, if the parent is in community $j \in \{1, 2\}$ and the child node is in community j too, then the joint degree distribution is the size-biased distribution $\mathbb{P}_{j*,m}$ given by $\mathbb{P}_{j*,m}(d_j, d_{-j}) = d_j\mathbb{P}_{j,m}(d_j, d_{-j})/\sum_k k\mathbb{P}_{j,m}(k, d_{-j})$ for all $d_j, d_{-j} \in \mathbb{Z}_+$; on the other hand, if the parent is in community $j \in \{1, 2\}$ and the child node is in community $-j \in \{1, 2\} \setminus \{j\}$, then the joint degree distribution of the child is the size-biased distribution $\mathbb{P}_{-j,m*}$ given by $\mathbb{P}_{-j,m*}(d_{-j}, d_j) = d_j\mathbb{P}_{j,m}(d_{-j}, d_j)/\sum_k k\mathbb{P}_{j,m}(d_{-j}, k)$ for all $d_j, d_{-j} \in \mathbb{Z}_+$. We denote a random variable with the size-biased distribution by $D^* + 1$ where D^* takes values in $\mathbb{Z}_+$. For a Poisson random variable with parameter $\lambda > 0$, i.e., $D \sim \text{Poi}(\lambda)$, we have $D^* \sim \text{Poi}(\lambda)$, so that the size-biased/sampling-biased distribution is a shifted Poisson distribution. This is the only distribution with this property.

An example of the limiting rooted GWMT_* is shown in Figure 2.1 where the root node is in community 1.

Assume that we have a rooted GWMT_* (with root node ψ) denoted by T_ψ. For a node $l \neq \psi$ let l_p be its parent, indicated by $l \longrightarrow l_p$, and $T_{l \longrightarrow l_p}$ be the sub-tree rooted at l when the link (l_p, l) is excised. Then assuming that l_p is inactive, state of node l only depends on the state of her children in sub-tree $T_{l \longrightarrow l_p}$. Next, we define a few random variables that will aid in describing the mean-field approximation.

$X_\psi^{(j)}$: Bernoulli r.v; $= 1$ if root node ψ of the rooted GWMT_* is on side j and inactive.

$Y_l^{(j,j)}$: Bernoulli r.v; $= 1$ if node $l(\neq \psi)$ and its parent l_p are both on side j and node l is inactive on $T_{l \to l_p}$.

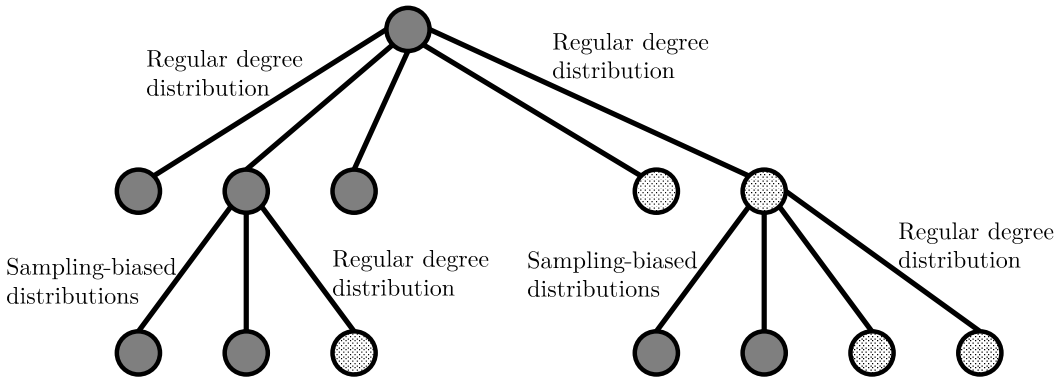


Figure 2.1: Illustration of the limiting rooted unimodular Galton-Watson Multi-type Tree. Solid circles denote nodes in community 1, and dotted circles denote nodes in community 2.

$Y_l^{(j,-j)}$: Bernoulli r.v; $= 1$ if node $l(\neq \psi)$ is on side $-j$ and its parent l_p is on side j and node l is inactive on $T_{l\to l_p}$.
$\boldsymbol{\alpha}_l^{(j)}$: Bernoulli r.v; $= 1$ if node l on side j is a seed node.
$K_l^{(j)}$: Threshold of node l on side j that is determined by number of its neighbors in either community, i.e., by the degrees of the node on side j and $-j$.
Then we can write down the following equations:
(i) A non-root node $l \neq \psi$ remains inactive on $T_{l\to l_p}$, if it is not seeded initially and the number of her children who are active does not exceed her threshold, i.e.,

$$Y_l^{(j,j)} = \left(1-\boldsymbol{\alpha}_l^{(j)}\right)\mathbf{1}\Big\{\sum_{i\longrightarrow l}\left(1-Y_i^{(j,j)}\right)+\sum_{i\longrightarrow l}\left(1-Y_i^{(j,-j)}\right)\leq K_l^{(j)}\Big\}, \tag{2.1}$$

$$Y_l^{(j,-j)} = \left(1-\boldsymbol{\alpha}_l^{(-j)}\right)\mathbf{1}\Big\{\sum_{i\longrightarrow l}\left(1-Y_i^{(-j,-j)}\right)+\sum_{i\longrightarrow l}\left(1-Y_i^{(-j,j)}\right)\leq K_l^{(-j)}\Big\}, \tag{2.2}$$

where $\mathbf{1}\{O\}$ is the indicator function of set O.
(ii) Root node ψ (on side j) remains inactive if it is not seeded initially and the number of her active children falls below her threshold, i.e.,

$$X_\psi^{(j)} = \left(1-\boldsymbol{\alpha}_\psi^{(j)}\right)\mathbf{1}\Big\{\sum_{i\longrightarrow \psi}\left(1-Y_i^{(j,j)}\right)+\sum_{i\longrightarrow \psi}\left(1-Y_i^{(j,-j)}\right)\leq K_\psi^{(j)}\Big\}. \tag{2.3}$$

For the mean-field approximation it is assumed that the random variables $Y_l^{(1,1)}$, $Y_l^{(1,2)}$, $Y_l^{(2,1)}$, and $Y_l^{(2,2)}$ for $l \neq \psi$ are, respectively, identically distributed when considering l as the variable and keeping (j,j) or $(j,-j)$ fixed. Moreover, it is assumed that all these

random variables are mutually independent. These random variables are then related via the following Recursive Distributional Equations (RDEs), where equality below should be interpreted in terms of distribution.

$$\begin{aligned}\widetilde{Y}^{(j,j)} \stackrel{d}{=} \left(1-\bar{\boldsymbol{\alpha}}^{(j)}(D_j^*+1, D_m)\right)\mathbf{1}\Big\{\sum_{i=1}^{D_j^*}\left(1-\widetilde{Y}_i^{(j,j)}\right) \\ +\sum_{i=1}^{D_m}\left(1-\widetilde{Y}_i^{(j,-j)}\right) \leq K^{(j)}(D_j^*+1, D_m)\Big\},\end{aligned} \tag{2.4}$$

$$\begin{aligned}\widetilde{Y}^{(j,-j)} \stackrel{d}{=} \left(1-\bar{\boldsymbol{\alpha}}^{(-j)}(D_{-j}, D_m^*+1)\right)\mathbf{1}\Big\{\sum_{i=1}^{D_m^*}\left(1-\widetilde{Y}_i^{(-j,j)}\right) \\ +\sum_{i=1}^{D_{-j}}\left(1-\widetilde{Y}_i^{(-j,-j)}\right) \leq K^{(-j)}(D_{-j}, D_m^*+1)\Big\},\end{aligned} \tag{2.5}$$

where for every $j \in \{1,2\}$, $\widetilde{Y}^{(j,j)}$ and $\widetilde{Y}_i^{(j,j)}$ as well as $\widetilde{Y}^{(j,-j)}$ and $\widetilde{Y}_i^{(j,-j)}$ are *i.i.d.* copies (Bernoulli random variables with unknown parameters). We also have a set of random variables: D_j is a random variable with the community j degree distribution, D_j^*+1 is a random variable with the size-biased distribution of D_j, D_m has inter-community degree distribution, and D_m^*+1 is a random variable with the size-biased distribution of D_m; the joint distribution of (D_j^*+1, D_m) is given by $\mathbb{P}_{j*,m}$, and the joint distribution of (D_j, D_m^*+1) is given by $\mathbb{P}_{j,m*}$ (for all $d_j, d_m \in \mathbb{Z}_+$, we have $\mathbb{P}_{j*,m}(d_j,d_m) = d_j\mathbb{P}_{j,m}(d_j,d_m)/\mathbb{E}[D_j]$ and $\mathbb{P}_{j,m*}(d_j,d_m) = d_m\mathbb{P}_{j,m}(d_j,d_m)/\mathbb{E}[D_m]$). We have also assumed, without loss of generality, that the seeding Bernoulli random variables have means that depend on the community and the degrees of the nodes, namely, $\alpha_j(d_j, d_{-j})$ for $j \in \{1,2\}$ and $d_j, d_{-j} \in \mathbb{Z}_+$. We also assume that threshold random variables are deterministic functions of the community and degrees of the nodes, namely, $K_j(d_j, d_{-j})$ for $j \in \{1,2\}$ and $d_j, d_{-j} \in \mathbb{Z}_+$. These are then used to construct the random variables $\bar{\boldsymbol{\alpha}}^{(j)}(D_j^*+1, D_m)$, $\bar{\boldsymbol{\alpha}}^{(-j)}(D_{-j}, D_m^*+1)$, $K^{(j)}(D_j^*+1, D_m)$ and $K^{(-j)}(D_{-j}, D_m^*+1)$.

Since we have RDEs with Bernoulli random variables, we can equivalently obtain the solutions by taking expectations and solving for the means of the underlying random variables. We set $\mathbb{E}[X_\psi^{(j)}] = \phi_j$, $\mathbb{E}[\widetilde{Y}^{(j,j)}] = \mu^{(j,j)}$ and $\mathbb{E}[\widetilde{Y}^{(j,-j)}] = \mu^{(j,-j)}$, taking expectation in (2.4)-(2.5) and then (2.3) yields

$$\begin{aligned}\mu^{(j,j)} = \sum_{u_j+u_{-j}\leq K_j(d_j,d_{-j})} \mathbb{P}_{j*,m}(d_j,d_{-j})(1-\alpha_j(d_j,d_{-j})) \\ \times Bi(u_j; d_j-1, 1-\mu^{(j,j)})Bi(u_{-j}; d_{-j}, 1-\mu^{(j,-j)}),\end{aligned} \tag{2.6}$$

$$\mu^{(j,-j)} = \sum_{u_j+u_{-j}\le K_{-j}(d_{-j},d_j)} \mathbb{P}_{-j,m*}(d_{-j},d_j)(1-\alpha_{-j}(d_{-j},d_j)) \times Bi(u_j;d_j-1,1-\mu^{(-j,j)})Bi(u_{-j};d_{-j},1-\mu^{(-j,-j)}), \tag{2.7}$$

$$\phi_j = \sum_{u_j+u_{-j}\le K_j(d_j,d_{-j})} \mathbb{P}_{j,m}(d_j,d_{-j})(1-\alpha_j(d_j,d_{-j})) \times Bi(u_j;d_j,1-\mu^{(j,j)})Bi(u_{-j};d_{-j},1-\mu^{(j,-j)}), \tag{2.8}$$

where $Bi(k;n,p) := \binom{n}{k}p^k(1-p)^{n-k}$ is the probability mass function of the binomial distribution.

To find the probability of a node in community $j \in \{1,2\}$ remaining inactive, i.e. $\phi_j = 1$, one needs to first solve the fixed point equations (2.6)-(2.7), and then substitute the result into (2.8). For ease of understanding we write equations (2.6)-(2.8) as follows:

$$\boldsymbol{\mu} = \boldsymbol{F}(\boldsymbol{\mu}), \text{ and } \boldsymbol{\phi} = \boldsymbol{\Phi}(\boldsymbol{\mu}), \tag{2.9}$$

for functions $\mathbf{F}(\cdot)$ and $\boldsymbol{\Phi}(\cdot)$ defined component-wise via (2.6)-(2.7), and (2.8), respectively.

A basic question at this point is whether one can rigorously justify (2.9), particularly given the various independence and uniformity assumptions for the derivation. A few other questions also arise: i) Does a solution to (2.9) exist; ii) Are there multiple solutions to (2.9)? Numerically, we observed that there are many cases where (2.9) has multiple solutions; and iii) Which solution should one pick if there are multiple solutions? Note that for every $\boldsymbol{\mu} \in [0,1]^4$ and $j \in \{1,2\}$, we have

$$\begin{aligned}\phi_j &= \sum_{d_j,d_{-j}} \mathbb{P}_{j,m}(d_j,d_{-j})(1-\alpha_j(d_j,d_{-j}))\times \\ &\qquad \sum_{u_j+u_{-j}\le K_j(d_j,d_{-j})} Bi(u_j;d_j-1,1-\mu^{(j,j)})Bi(u_{-j};d_{-j},1-\mu^{(j,-j)}) \\ &\le \sum_{d_j,d_{-j}} \mathbb{P}_{j,m}(d_j,d_{-j})(1-\alpha_j(d_j,d_{-j})) = \mathbb{P}(\boldsymbol{\alpha}_\psi^{(j)}=0),\end{aligned} \tag{2.10}$$

so that the seeding distribution gets automatically accounted in any solution of (2.9), and the final population of active nodes includes at least the seed nodes.

Before proceeding, we should again point out that equations of a similar form were heuristically postulated in the literature [37, 38, 59–61]. An important contribution of our work is thus to rigorously prove the validity of (2.9), and to identify the correct solution to choose. As discussed in [35], the existence of multiple solutions and a lack of "monotonicity" makes it extremely challenging to use the techniques developed in [24, 44] to prove the needed

results.

2.5 Markov Process of Adoption

As the first step to prove the validity of the mean-field equations, we construct a Markov process that couples the evolution of the adoption process with the process of generating the random graph using the configuration model.

The typical way to create a random graph with a given degree sequence $(d_i^n)_1^n$ using the configuration model is as follows: first label nodes of the graph $1, 2, \ldots, n$ such that node i has d_i^n half-edges sticking out of it; next iterate through all the unpaired half-edges so that at each step, two half-edges are paired randomly; and declare the final graph as the desired random graph. In our setting, when there are two communities, the basic idea of generating the random graph using the configuration model is just the same. However, to analyze the adoption process we work a little differently. We start by realizing the early adopter nodes using the seeding random variables. We set the early adopters to be active and make all their half-edges active. Any other node and its half-edges will initially be counted as inactive. We then run the adoption process and draw the random graph simultaneously by iterating through the active half-edges (if any). At each iteration, we pick an active half-edge, i.e., an half-edge connected to an active node, and connect it to some other half-edge that belongs to the appropriate community. Then we remove both half-edges from the graph. Moreover, if the second half-edge belongs to an inactive node, we reduce its threshold by one. If the threshold of the inactive node becomes minus one after this change, we activate this node and also all the half-edges that are still connected to this node. Note that this process stops when all active half-edges have been omitted, and the remainder of the graph (containing only inactive half-edges) is not realized (or can be realized but will not influence the contagion process). This process is described in Algorithm 1. We keep

Algorithm 1: Process of jointly generating the random graph and running the adoption process.

Data: Degree sequences and early adopters
Result: Sub-graph of the final random graph that contains all Active nodes
initialization;
while *There is an active half-edge* **do**
    Randomly choose an active half-edge;
    Randomly choose another half-edge belongs to proper community;
    Omit two selected half-edges from the set of half-edges;
    Update the state of the inactive node (if any);
end

track of active half-edges, inactive nodes and number of times that the process described in Algorithm 1 picks half-edges from each community. The random variables associated with these quantities are given as follows:
$A_j(k)$: Number of active half-edges belonging entirely to community j at time k.
$A_m^{(j)}(k)$: Number of active half-edges between the two communities belonging to nodes in community j at time k.
$T_j(k)$: Number of times the algorithm visits community j up to time k where a visit means removing two half-edges within the same community.
$I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)$: Number of inactive nodes in community j with d_j initially assigned half-edges corresponding to community j where u_j of them have been removed by k, and similarly, d_{-j} initially assigned half-edges corresponding to community $-j$ where u_{-j} of them have been removed by k. Note that $u_j + u_{-j} \leq K_j(d_j, d_{-j})$, $0 \leq u_j \leq d_j$, and $0 \leq u_{-j} \leq d_{-j}$.
It is easily verified that $\{X^n(k)\}_{k\in\mathbb{Z}_+}$ is a discrete-time Markov chain, where

$$X^n(k) := (A_j(k), A_m^{(j)}(k), T_j(k), I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k), \dots),$$

and $j \in \{1, 2\}$. For ease of explanation we denote the number of edges entirely in community j with $m_j(n)$ and the number of edges between the two communities by $m_m(n)$; these can be determined once the degrees have been realized.

Remark II.1. We use the following notational convention throughout the chapter: whenever the term $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}$ appears as a member of some sequence, it represents all $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}$ for $j \in \{1, 2\}$, $(d_j, u_j) \in \mathbb{Z}_+^2$, $(d_{-j}, u_{-j}) \in \mathbb{Z}_+^2$, $u_j \leq d_j$, $u_{-j} \leq d_{-j}$, and $u_j + u_{-j} \leq K_j(d_j, d_{-j})$. We use the same notational convention for $A_j(k)$, $A_m^{(j)}(k)$, and $T_j(k)$.

The mean-field analysis [19, 34, 68, 69] proceeds by scaling both space and time by n and considering the one-step drift of the scaled process. We will now present the one-step drift analysis of our Markov chain (for the unscaled variables). At each iteration, one of the following events will happen:

1.) Two active half-edges will be omitted. This event results in the half-edges being "wasted", in a manner of speaking. Here two sub-cases are possible:
i.) Both half-edges belongs to community j: This event happens with probability

$$\frac{A_j(k)\,(A_j(k)-1)}{\Big(A_1(k)+A_2(k)+A_m^{(1)}(k)+A_m^{(2)}(k)\Big)\,(2m_j(n)-2T_j(k)-1)}.$$

In this case, we should update the corresponding variables as follows:

$$A_j(k+1) = A_j(k) - 2,\ T_j(k+1) = T_j(k) + 1.$$

ii.) Half-edges belongs to different sides: This event happens with probability

$$\frac{2A_m^{(1)}(k)A_m^{(2)}(k)}{\left(A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)\right)\left(m_m(n) - (k - T_1(k) - T_2(k))\right)}.$$

In this case, we should update the variables as follows:

$$A_m^{(2)}(k+1) = A_m^{(2)}(k) - 1,\ A_m^{(1)}(k+1) = A_m^{(1)}(k) - 1.$$

2.) One active half-edge and one inactive half-edge will be omitted, while the inactive half-edge belongs to inactive nodes in community j. Four sub-cases arise here:
i.) The inactive node belongs to $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)$ and the active half-edges belongs to community j, while $K_j(d_j, d_{-j}) \geq u_j + u_{-j} + 1$. This event results in the threshold of an inactive node in community j being lowered by 1 owing to a node within its own community. This occurs with probability

$$\frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{(d_j - u_j)\, I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)}{2m_j(n) - 2T_j(k) - 1}$$

In this case, we should update variables as follows:

$$\begin{aligned} &A_j(k+1) = A_j(k) - 1,\ I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k+1) = I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k) - 1, \\ &T_j(k+1) = T_j(k) + 1,\ I^{(j)}_{d_j,d_{-j},u_j+1,u_{-j}}(k+1) = I^{(j)}_{d_j,d_{-j},u_j+1,u_{-j}}(k) + 1. \end{aligned}$$

ii.) The inactive node belongs to $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)$ and the active half-edge belongs to community j, while $K_j(d_j, d_{-j}) = u_j + u_{-j}$. Note that an inactive node becomes active during this event and all remaining half-edges also become active. This is an important growth event for our process. This occurs with probability

$$\frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{(d_j - u_j)\, I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)}{2m_j(n) - 2T_j(k) - 1}$$

Here we update the variables as follows:

$$A_j(k+1) = A_j(k) - 1 + d_j - u_j - 1,\ T_j(k+1) = T_j(k) + 1,$$
$$A_m^{(j)}(k+1) = A_m^{(j)}(k) + d_{-j} - u_{-j},\ I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k+1) = I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k) - 1.$$

iii.) The inactive node belongs to $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)$ and the active half-edge comes from the other community, while $K_j(d_j, d_{-j}) \geq u_j + u_{-j} + 1$. Here the threshold of an inactive node is being reduced by a node from the other community. This occurs with probability

$$\frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{(d_{-j} - u_{-j})\, I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)}{m_m(n) - (k - T_1(k) - T_2(k))}$$

Here we update the variables as follows:

$$A_m^{(-j)}(k+1) = A_m^{(-j)}(k) - 1,\ I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k+1) = I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k) - 1,$$
$$I^{(j)}_{d_j,d_{-j},u_j,u_{-j}+1}(k+1) = I^{(j)}_{d_j,d_{-j},u_j,u_{-j}+1}(k) + 1.$$

iv.) The inactive node belongs to $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)$ and the active half-edge comes from the other community, while $K_j(d_j, d_{-j}) = u_j + u_{-j}$. This is another important growth event for our process wherein an inactive node becomes active owing to a node from the other community. This occurs with probability

$$\frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{(d_{-j} - u_{-j})\, I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)}{m_m(n) - (k - T_1(k) - T_2(k))}$$

Here we update the variables as follows:

$$A_m^{(-j)}(k+1) = A_m^{(-j)}(k) - 1,\ A_j(k+1) = A_j(k) + d_j - u_j,$$
$$A_m^{(j)}(k+1) = A_m^{(j)}(k) + d_{-j} - u_{-j} - 1,\ I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k+1) = I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k) - 1.$$

Finally, note that these random variables satisfy the balance equations given by the realization of degrees. For $j \in \{1, 2\}$ we have

$$A_j(k) + \sum_{u_j+u_{-j}\leq K_j(d_j,d_{-j})} (d_j - u_j) I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k) = 2m_j(n) - 2T_j(k),$$
$$A_m^{(j)}(k) + \sum_{u_j+u_{-j}\leq K_j(d_j,d_{-j})} (d_{-j} - u_{-j}) I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k) = m_m(n) - (k - T_j(k) - T_{-j}(k)), \tag{2.11}$$

where the summations above are understood to be over both the degrees $(d_j, d_{-j}) \in \mathbb{Z}_+^2$ and the used half-edges $(u_j, u_{-j}) \in \mathbb{Z}_+^2$ meeting the constraint listed underneath.

Remark II.2. The the coordinates of the discrete-time Markov chain $\{X^n(k)\}_{k\in\mathbb{Z}_+}$ are dependent because of the balance equations. Hence, we only need to keep track of T_j and $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}$ for $j \in \{1,2\}$ and different values of d_j, d_{-j}, u_j, and u_{-j}.

The one-step drift for the unscaled random variables is obtained by summing over all possible events, given the current state of the Markov chain. The details can be found in Appendix A.2.

We now conclude this section by stating the regularity conditions on the degree sequences and some consequences of these conditions. Recall that $\mathbf{d}_1 = (d^n_{1,i})_1^{n_1}$ and $\mathbf{d}_2 = (d^n_{2,i})_{n_1+1}^{n}$ are the corresponding degree sequence of the sub-graph for community 1 and 2, respectively, and $\mathbf{d}_m = (d^n_{m,i})_1^n$ is the degree sequence of the bipartite graph connecting the two communities; these sequences satisfy the following basic conditions: 1) $\sum_{i=1}^{n_1} d^n_{1,i}$ is even; 2) $\sum_{i=n_1+1}^{n} d^n_{2,i}$ is even; and 3) $\sum_{i=1}^{n_1} d^n_{m,i} = \sum_{i=n_1+1}^{n} d^n_{m,i}$. The size of the community 1 is n_1 and the size of the community 2 is $n_2 = n - n_1$. We assume that $\lim_{n\to\infty} n_1/n = \beta$, and $\lim_{n\to\infty} n_2/n = 1-\beta$. We denote these two quantities by variables $\beta_1 = \beta$ and $\beta_2 = 1-\beta$. We also assume that as $n \to \infty$, both $m_j(n)$ for $j \in \{1,2\}$ and $m_m(n)$ goes to infinity, where $m_j(n)$ denotes the number of edges in community j and $m_m(n)$ denotes the number of edges between the two communities.

Definition II.3. We say the degree sequences $\mathbf{d}_1$, $\mathbf{d}_2$, and $\mathbf{d}_m$ satisfy the **degree regularity conditions** if the followings hold:

1. The empirical degree distribution of nodes in community $j \in \{1,2\}$ converges to some joint probability distribution:

$$|\{i : d^n_{j,i} = r, d^n_{m,i} = s \text{ and } i \in \text{Community } j\}|/n_j \to \mathbb{P}_{j,m}(r,s) \text{ for every } r,s \geq 0.$$

2. The average degree of nodes, and the number of nodes concentrates:

$$|2m_j(n)/n_j - \lambda_{j,j}| = O(n_j^{-\eta}), \quad |m_m(n)/n_j - \lambda_{j,m}| = O(n_j^{-\eta}), \quad |n_j/n - \beta_j| = O(n^{-\eta}),$$

where $\eta > \frac{1}{3}$, $\lambda_{j,j} > 0$, and $\lambda_{j,m} > 0$ for $j \in \{1,2\}$ are some constants. Given assumption 1, the following equalities hold:

$$\lambda_{j,j} = \sum_{r\geq 0} r\, \mathbb{P}_{j,m}(r, \mathbb{Z}_+) \in (0,\infty), \qquad \lambda_{j,m} = \sum_{s\geq 0} s\, \mathbb{P}_{j,m}(\mathbb{Z}_+, s) \in (0,\infty),$$

where $\mathbb{P}_{j,m}(\mathbb{Z}_+, s) \coloneqq \sum_r \mathbb{P}_{j,m}(r,s)$, and $\mathbb{P}_{j,m}(r,\mathbb{Z}_+) \coloneqq \sum_s \mathbb{P}_{j,m}(r,s)$ for $j \in \{1,2\}$.

Definition II.4. We say the degree sequences $\mathbf{d}_1$, $\mathbf{d}_2$, and $\mathbf{d}_m$ satisfy the **graph regularity conditions** if they satisfy the degree regularity conditions and the followings hold:

1. The random multigraphs $G^*(n_1, \mathbf{d}_1)$ and $G^*(n_2, \mathbf{d}_2)$ are asymptotically simple random graphs with positive probability. Mathematically, the condition is

$$\sum_i (d^n_{j,i})^2 = O(m_j(n)) \text{ for } j \in \{1, 2\},$$

which implies $\liminf_{n\to\infty} \mathbb{P}(G^*(n_j, \mathbf{d}_j) \text{ is simple}) > 0$ for $j \in \{1, 2\}$ [62, Theorem 1.1].

2. The random bipartite multigraph $G^*(n_1, n_2, \mathbf{d}_m)$ is asymptotically simple random bipartite graph with positive probability. Mathematically, the condition is

$$(i) \sum_{i=1}^{n_1} \sum_{i'=n_1+1}^{n} d^n_{m,i}(d^n_{m,i} - 1) d^n_{m,i'}(d^n_{m,i'} - 1) = O((m_m(n))^2),$$

(ii) for any $M \geq 1$,

$$\sum_{i=\min(d_2^{\max}, M)}^{n_1} d^n_{m,(i)} = \Omega(m_m(n)) \text{ and } \sum_{i=n_1+\min(d_1^{\max}, M)}^{n} d^n_{m,(i)} = \Omega(m_m(n)),$$

where $(d^n_{m,(i)})_{i=1}^{n_1}$ is the descending-sorted version of $(d^n_{m,i})_{i=1}^{n_1}$, $(d^n_{m,(i)})_{i=n_1+1}^{n}$ is the descending-sorted version of $(d^n_{m,i})_{i=n_1+1}^{n}$, $d_1^{\max} = d^n_{m,(1)}$, and $d_2^{\max} = d^n_{m,(n_1+1)}$. This implies that [62, Theorem 6.1]

$$\liminf_{n\to\infty} \mathbb{P}(G^*(n_1, n_2, \mathbf{d}_m) \text{ is simple}) > 0.$$

Recall that $x = \Omega(N(n))$ means $\liminf_{n\to\infty} x/N(n) > 0$.

Janson in [70] proved that the probability of the event "the random multigraph generated by the configuration model is simple" is strictly positive if and only if assumption 1 holds. Various sufficient conditions for this property were given by the authors of [5, 12, 71, 72]. The final result in [70] provides a necessary and sufficient condition. Blanchet and Stauffer in [73] proved similar result for bipartite random graphs: "the random bipartite multigraph generated by the configuration model is simple" with strictly positive probability if and only if assumption 2 holds. Both results are presented in [62].

Remark II.5. Assumption 2 part (i) corresponds to assumption 1. Also, note that part (ii) of assumption 2 follows from part (i) if $d_1^{\max} = o(m_m(n))$ and $d_2^{\max} = o(m_m(n))$ [62, Remark 6.1].

Remark II.6. Almost all the results presented in this chapter only need the degree regularity conditions. The additional assumptions in the graph regularity conditions extend this results

to uniformly sampled simple graphs. Note that in [34], assumption 2 appears only as a convergence rather than a concentration. However, this assumption is necessary for the analysis of [34] as well.

Remark II.7. Suppose the degree regularity conditions hold. Since $\sum_{i=1}^{n_1} d_{m,i}^n = \sum_{i=n_1+1}^{n} d_{m,i}^n$, we have $\lambda_{1,m}\beta_1 = \lambda_{2,m}\beta_2$. We denote this quantity by $\lambda_m \coloneqq \lambda_{m,1}\beta_1 = \lambda_{m,2}\beta_2$. Similarly, we define $\lambda_1 \coloneqq \lambda_{1,1}\beta_1$ and $\lambda_2 \coloneqq \lambda_{2,2}\beta_2$. Note that by assumption 2:

$$\lambda_1 = \lim_{n\to\infty} 2m_1(n)/n, \qquad \lambda_m = \lim_{n\to\infty} m_m(n)/n, \qquad \lambda_2 = \lim_{n\to\infty} 2m_2(n)/n.$$

2.6 Convergence to ODE

In this section, we use techniques developed for the mean-field analysis [19, 34, 68, 69] of the resulting population density-dependent Markov processes to approximate the process by a system of ordinary differential equations (ODEs). We show that a scaled-version of the Markov process of adoption from Section 2.5 converges (in probability) to a set of continuous functions obtained from the solution of a set of ODEs. We start by highlighting why the analysis is non-trivial:

Point 1. The first point concerns some of the terms that appear in the one-step drift. Notice that we have many terms like

$$\frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)}.$$

In terms of the scaled variables, these terms are not Lipschitz unless there is a lower bound on the value of the (scaled) denominator. Owing to this, in our ODE approximation we will have to stop the Markov process of adoption just before the sum of these scaled variables hits zero (corresponding to the denominator above), i.e., before all the active half-edges have been omitted; it is important that this be the sum and not the individual components. For the exact same reason, we have to stop the process before we ran out of half-edges in any of the two communities or between the communities.

Point 2. The second point is regarding the one-step drift of variables like $A_j(k)$, i.e., the number of active half-edges based on the community structure. The one-step drift can be unbounded as the increase can equal the number of nodes (in the appropriate community) minus one. However, owing to the balance equations as we pointed out in Remark II.2, there is no need to keep track of the random variables associated with the number of active half-edges.

There is, however, another technical issue with the one-step drift of other quantities as they depend on all terms $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)$ through an infinite sum associated with $A_1(k) + A_2(k) + A^{(1)}_m(k) + A^{(2)}_m(k)$. For any finite n, we only need to account for a finite number of terms but in the limit we have a countable number of terms leading to a similar property for the functions associated with these variables. Since the coefficient of these variables are increasing without bound, the associated functions are not Lipschitz continuous. This precludes the direct application of the results of [19, 68]. On the other hand, given degree regularity conditions, this should be a superficial problem as the total number of half-edges associated with inactive nodes with high degrees is small.

To address the specific scenario outlined above we bound the original Markov chain, from above and below, using two truncated versions. We denote these Markov chains with $X^n_{U,\delta}$ and $X^n_{L,\delta}$ respectively. Set $d_\delta > 0$ large enough so that the following inequalities hold for all n:

$$\sum_{i=1}^{n_1} (d^n_{1,i} + d^n_{m,i}) \mathbf{1}\{d^n_{1,i} + d^n_{m,i} > d_\delta\} \leq \delta n_1,$$
$$\sum_{i=n_1+1}^{n} (d^n_{2,i} + d^n_{m,i}) \mathbf{1}\{d^n_{2,i} + d^n_{m,i} > d_\delta\} \leq \delta n_2.$$

$X^n_{U,\delta}$ is defined by activating all nodes with total degree larger than d_δ, i.e., by setting $\alpha_j(d_j, d_{-j}) = 1$ for all $d_j + d_{-j} > d_\delta$ and $j \in \{1, 2\}$. X^n_L is defined by assuming $K_j(d_j, d_{-j}) = d_j + d_{-j}$ for all $j \in \{1, 2\}$ and all $d_j + d_{-j} > d_\delta$. Since none of the inactive nodes with degree higher than d_δ can be activated, instead of tracking the random variables $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}$ for these nodes, we track the total number of half-edges associated with these random variables.

Intuitively speaking, the total number of inactive nodes with degree less than or equal to d_δ at the natural stopping time of X^n is bounded between the same quantities for $X^n_{L,\delta}$ and $X^n_{U,\delta}$; it is important to note that this bound only works for the stopping time of the processes and not the whole trajectory. We use the techniques developed by Wormald [19,68] to approximate $X^n_{L,\delta}$ and $X^n_{U,\delta}$. In Section 2.9, we then show that the difference between these two approximations can be made arbitrary small by tuning δ.

2.6.1 Convergence to ODE for Truncated Processes

As per say above, we focus on the case where inactive nodes with degree higher than some constant $d_{\max}$ cannot be activated. Let us define a new set of random variables, to keep track of half-edges associated with these inactive nodes:
$W_j(k)$: Number of remaining half-edges belonging to inactive nodes with degree higher than

$d_{\max}$ in community j at time k.
$W_m^{(j)}(k)$: Number of remaining half-edges between the two communities belonging to inactive nodes with degree higher than $d_{\max}$ in community j at time k.
At each step of the Markov process of adoption, the value of $W_j(k)$ or $W_m^{(j)}(k)$ can reduce at most by one.

Remark II.1. The one step drifts of these new set of random variables are given as follows:

$$\begin{aligned}
&\mathbb{E}[W_j(k+1) - W_j(k)|X^n(k)] = \\
&\qquad \frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{-W_j(k)}{2m_j(n) - 2T_j(n) - 1}, \\
&\mathbb{E}[W_m^{(j)}(k+1) - W_m^{(j)}(k)|X^n(k)] = \\
&\qquad \frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{-W_m^{(j)}(k)}{m_m(n) - (k - T_1(k) - T_2(k))}.
\end{aligned}$$

Note that these random variables do not change the dynamic of the Markov process of adoption; hence, the one step drift of all other random variables remain the same. Also note that the sum in the balance equations (2.11) has now a finite number of summands (less than $(d_{\max}+1)^4$) as the other terms are replaced with either $W_j(k)$ or $W_m^{(j)}(k)$:

$$\begin{aligned}
A_j(k) + \sum_{\substack{u_j+u_{-j}\le K_j(d_j,d_{-j}) \\ d_j+d_{-j}\le d_{\max}}} (d_j - u_j) I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k) + W_j(k) = 2m_j(n) - 2T_j(k), \\
A_m^{(j)}(k) + \sum_{\substack{u_j+u_{-j}\le K_j(d_j,d_{-j}) \\ d_j+d_{-j}\le d_{\max}}} (d_{-j} - u_{-j}) I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k) + W_m^{(j)}(k) = m_m(n) - T_m(k),
\end{aligned} \tag{2.12}$$

where $T_m(k) \coloneqq k - T_j(k) - T_{-j}(k)$.

Recall that $m_j(n)$ denotes the total number of edges on side $j \in \{1,2\}$, and $m_m(n)$ denotes the total number of edges between the two communities. Recall also that by Remark II.7: $\lim_{n\to\infty} 2m_1(n)/n = \lambda_1$, $\lim_{n\to\infty} m_m(n)/n = \lambda_m$, and $\lim_{n\to\infty} 2m_2(n)/n = \lambda_2$. Then the ODEs follow by defining the real functions $\tau_j(t)$, $i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t)$, $w_j(t)$, and $w_m^{(j)}(t)$ to model the behavior of their discrete counterpart, i.e., intuitively speaking:

$$\begin{aligned}
\tau_j(t) &= \lim_{n\to\infty} \frac{1}{n} T_j(tn), \qquad i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t) = \lim_{n\to\infty} \frac{1}{n} I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(tn), \\
w_j(t) &= \lim_{n\to\infty} \frac{1}{n} W_j(tn), \quad w_m^{(j)}(t) = \lim_{n\to\infty} \frac{1}{n} W_m^{(j)}(tn),
\end{aligned} \tag{2.13}$$

where all the limits are in probability and sample-path-wise. We can then use the one-step drift from Appendix A.2 and Remark II.1 to derive the ODEs corresponding to the modified

process. The details are in Appendix A.3.

As we mentioned in Point 1, we have to stop the process just before we ran out of half-edges in each community or between the two communities and before we ran out of active half-edges. Using the same technique as in [19, Section 5], we can approximate the (truncated) discrete-time Markov chain as long as the associated functions in the ODEs are bounded away from zero.

Theorem II.2. *Fix $\epsilon > 0$ small enough, and consider the Markov process of adoption. Assume there is a constant $d_{\max} > 0$, independent of n, such that $K_j(d_j, d_{-j}) = d_j + d_{-j}$ for all $d_j + d_{-j} > d_{\max}$ and $j \in \{1, 2\}$. Assume the degree regularity conditions given in Definition II.3 hold. Assume n is large enough such that the total number of half-edges within community $j \in \{1, 2\}$ is bounded by $4\lambda_j n$ and between the two communities is bounded by $4\lambda_m n$, where λ_1, λ_2 and λ_m are given as in Remark II.7. Condition on the event that the total number of active half-edges is greater than $2\epsilon n$. Let $\theta = O(n^{-\gamma})$ for some $\gamma \in (1/3, \eta)$. Then, with probability $1 - O\left(\theta^{-1} \exp(-n\theta^3)\right)$*

$$\begin{aligned}
&\left| I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t) - n i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t/n) \right| = O(n\theta),\\
&|T_j(t) - n\tau_j(t/n)| = O(n\theta),\\
&|W_j(t) - n w_j(t/n)| = O(n\theta),\\
&\left| W^{(j)}_m(t) - n w^{(j)}_m(t/n) \right| = O(n\theta),
\end{aligned}$$

uniformly for $0 \le t \le \sigma_\epsilon n$, where w_j, $w^{(j)}_m$, $i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}$ and τ_j are the solution of the ODEs given in Appendix A.3 with the corresponding initial conditions, and $\sigma_\epsilon = \sigma_\epsilon(n)$ is the supremum of those x to which the solution of the ODEs can be extended before reaching within l^∞-distance $C\theta$ of the boundary of $\mathcal{D}_\epsilon$, for a sufficiently large constant C. The open connected set $\mathcal{D}_\epsilon$ is defined as follows:

$$\begin{aligned}
\mathcal{D}_\epsilon \coloneqq \Big\{ &(t, \tau_1, \tau_2, w_1, w_2, w^{(1)}_m, w^{(2)}_m, i^{(1)}_{d_1,d_2,u_1,u_2}, i^{(2)}_{d_2,d_1,u_2,u_1}) \in R^K :\\
&-\epsilon < t - \tau_1 - \tau_2 < \lambda_m - \epsilon,\\
&\text{for } j \in \{1,2\}\colon\ -\frac{\epsilon}{2} < \tau_j < \frac{\lambda_j - \epsilon}{2},\ -\epsilon < w_j < 4\lambda_j,\ -\epsilon < w^{(j)}_m < 2\lambda_m,\\
&\text{for } j \in \{1,2\},\ u_j \le d_j,\ u_{-j} \le d_{-j},\ \text{and } d_j + d_{-j} \le d_{\max}\colon\ -\epsilon < i^{(j)}_{d_j,d_{-j},u_j,u_{-j}} < 2,\\
&\epsilon < a_1 + a_2 + a^{(1)}_m + a^{(2)}_m < 4(\lambda_1 + \lambda_2 + \lambda_m) \Big\}
\end{aligned}$$

where $K \leq 6 + (d_{\max} + 1)^4$ *is a constant,*

$$a_j := - \sum_{\substack{u_j + u_{-j} \leq K_j(d_j, d_{-j}) \\ d_j + d_{-j} \leq d_{\max}}} (d_j - u_j) i^{(j)}_{d_j, d_{-j}, u_j, u_{-j}} + \lambda_j - 2\tau_j - w_j,$$

$$a^{(j)}_m := - \sum_{\substack{u_j + u_{-j} \leq K_j(d_j, d_{-j}) \\ d_j + d_{-j} \leq d_{\max}}} (d_{-j} - u_{-j}) i^{(j)}_{d_j, d_{-j}, u_j, u_{-j}} + \lambda_m - \tau_m - w^{(j)}_m,$$

and $\tau_m := t - \tau_1 - \tau_2$.

Proof. See Appendix A.4.1. □

2.7 A Probabilistic Method to Solve the ODE

To get a better understanding of the ODEs, we provide a probabilistic sketch to find the form of the solution. Note that the purpose of this section is to provide a heuristic argument on how to solve the ODEs given in Appendix A.3. Similar ideas were used in [19, Subsection 3.3.3]. Our sketch also provides an answer to the "surprising simplification" that has been observed in the solution of the ODEs in [33, 34].

Each step of our Markov process of adoption has two phases: first, we pick an active half-edge and then, we pair it with a randomly chosen edge. Consider a fixed half-edge e in community j that belongs to an inactive node v at the beginning of the Markov process of adoption. We want to estimate the probability that e is not paired with any other half-edge upto time k.

Condition on the event that v is inactive at time k. Then e is not paired with any half-edge, if it has not been picked in the second phase of any step of the process up to time k. Any additional dependencies introduced by the conditioning should fade away as $n \to \infty$, and we will proceed by ignoring them. Since half-edges are chosen uniformly at random in the second phase of each step, we have

$$\mathbb{P}(\{e \text{ is not paired with any other half-edge}\} | \{v \text{ is inactive}\}) \approx \left(1 - \frac{1}{2m_j(n) - 1}\right) \times \left(1 - \frac{1}{2m_j(n) - 2 - 1}\right) \times \cdots \times \left(1 - \frac{1}{2m_j(n) - 2T_j(k) - 1}\right).$$

Using the simple approximation $1 - x \approx \mathrm{e}^{-x}$ for small values of x, we get

$$\begin{aligned}
&\mathbb{P}(\{e \text{ is not paired with any other half-edge}\}|\{v \text{ is inactive}\}) \\
&\quad \approx \exp\left(-\sum_{i=1}^{T_j(k)} \frac{1}{2m_j(n) - 2i}\right) \approx \exp\left(-\frac{1}{2}\int_{m_j(n)-T_j(k)}^{m_j(n)} \frac{1}{z}\, dz\right) = \left(1 - \frac{T_j(k)}{m_j(n)}\right)^{\frac{1}{2}}.
\end{aligned}$$

Next, we use the same argument for the half-edges that are supposed to connect to the nodes in the other community. However, there is an important distinction here as two phases happen in different communities: if in the first phase we pick an active half-edge from community j, in the second phase we pick a random half-edge from community $-j$. This makes the direct use of the above argument almost impossible. To fix it, we track two half-edges, dangling from two nodes in communities 1 and 2. Consider two fixed half-edges e_1 and e_2 between the communities such that that e_j belongs to an inactive node v_j in community j for $j \in \{1, 2\}$. Now, we can use the exact same argument as above:

$$\begin{aligned}
&\mathbb{P}(\{e_1 \text{ and } e_2 \text{ are not paired with any other half-edges}\}|\{v_1 \text{ and } v_2 \text{ are inactive}\}) \\
&\quad \approx \left(1 - \frac{1}{m_m(n)}\right) \times \left(1 - \frac{1}{m_m(n) - 1 - 1}\right) \times \cdots \times \left(1 - \frac{1}{m_m(n) - T_m(k) - 1}\right) \\
&\quad \approx \left(1 - \frac{T_m(k)}{m_m(n)}\right).
\end{aligned}$$

where $T_m(k)$ denotes the number of times the algorithm removes one half-edge from each community. Note that $T_m(k) = k - T_1(k) - T_2(k)$. For large values of n, intuitively speaking, the events $\{e_j \text{ is not paired with any other half-edges}\}$ for $j \in \{1, 2\}$ are independent, and we can write:

$$\begin{aligned}
&\mathbb{P}(\{e_1 \text{ and } e_2 \text{ are not paired with any other half-edges}\}|\{v_1 \text{ and } v_2 \text{ are inactive}\}) \\
&\quad \approx \mathbb{P}(\{e_1 \text{ is not paired with any other half-edge}\}|\{v_1 \text{ is inactive}\}) \\
&\quad\quad \times \mathbb{P}(\{e_2 \text{ is not paired with any other half-edge}\}|\{v_2 \text{ is inactive}\})
\end{aligned}$$

If we denote the two quantities on the RHS as $Z_1(k)$ and $Z_2(k)$ respectively, we expect to have

$$Z_1(k)Z_2(k) \approx \left(1 - \frac{T_m(k)}{m_m(n)}\right).$$

Next, consider a vertex v in community j with d_j half-edges in community j and d_{-j} half-edges between the two communities. The above sketch together with some independence

assumptions (which can be justified as $n \to \infty$) suggests the following proximate equality at time k:

$$\mathbb{P}\Big(\Big\{\begin{matrix}u_j \text{ out of } d_j \text{ half-edges and } u_{-j} \text{ out of } d_{-j} \text{ half-edges of} \\ v \text{ has been removed where } u_j + u_{-j} \leq K_j(d_j, d_{-j})\end{matrix}\Big\}\Big) \approx$$
$$Bi\left(u_j, d_j; 1 - \left(1 - \frac{T_j(k)}{m_j(n)}\right)^{\frac{1}{2}}\right) \times Bi\left(u_{-j}, d_{-j}; 1 - Z_j(k)\right),$$

which further implies

$$\mathbb{E}\big[I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)\big|I^{(j)}_{d_j,d_{-j},0,0}(0)\big] \approx$$
$$I^{(j)}_{d_j,d_{-j},0,0}(0) \times Bi\left(u_j, d_j; 1 - \left(1 - \frac{T_j(k)}{m_j(n)}\right)^{\frac{1}{2}}\right) \times Bi\left(u_{-j}, d_{-j}; 1 - Z_j(k)\right). \tag{2.14}$$

Now, we are ready to simplify the differential equations given in Appendix C. Let $\mu^{(j,j)}$ and $\mu^{(j,-j)}$ model the limiting value of the probabilities defined above; intuitively speaking

$$\mu^{(j,j)}(t) = \lim_{n\to\infty}\left(1 - \frac{T_j(tn)}{m_j(n)}\right)^{\frac{1}{2}}, \text{ and } \mu^{(j,-j)}(t) = \lim_{n\to\infty} Z_j(tn). \tag{2.15}$$

Then the heuristic equality (2.14) suggests:

$$i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t) = i^{(j)}_{d_j,d_{-j},0,0}(0)\, Bi(u_j; d_j, 1 - \mu^{(j,j)}(t))\, Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}(t)). \tag{2.16}$$

Also, by the definition of scaled variables

$$\tau_j(t) = \frac{\lambda_j}{2}\left(1 - \mu^{(j,j)}(t)^2\right), \tag{2.17}$$
$$\tau_m(t) = \lambda_m\left(1 - \mu^{(1,2)}(t)\mu^{(2,1)}(t)\right), \tag{2.18}$$

where τ_m is the continuous counterpart of T_m. Moreover, we expect the following equality to hold:

$$\frac{\lambda_1}{2}\mu^{(1,1)}(t)^2 + \frac{\lambda_2}{2}\mu^{(2,2)}(t)^2 + \lambda_m\mu^{(1,2)}(t)\mu^{(2,1)}(t) = \lambda_m + \frac{\lambda_1}{2} + \frac{\lambda_2}{2} - t, \tag{2.19}$$

as it is equivalent to the equality $T_1(k) + T_2(k) + T_m(k) = k$. The importance of the above equality is that given the vector $\big(\mu^{(1,1)}(t), \mu^{(1,2)}(t), \mu^{(2,1)}(t), \mu^{(2,2)}(t)\big)$, the value of t is uniquely determined.

2.8 Analysis of the ODE

Following the heuristic argument provided in Section 2.7, we start with presenting the solution of the ODEs in Appendix A.3. The following lemma characterizes the solution of the differential equations that (with high probability) approximate the adoption process.

Lemma II.1. *The solution of differential equations* (A.1)–(A.4) *with initial condition* (A.5) *in* $\mathcal{D}_\epsilon$*, is given by equations* (2.16), (2.17)*, and the following:*

$$w_j(t) = w_j(0)\mu^{(j,j)}(t), \qquad w_m^{(j)}(t) = w_m^{(j)}(0)\mu^{(j,-j)}(t), \tag{2.20}$$

where $j \in \{1,2\}$ *and* $\mu^{(j,j)}$ *and* $\mu^{(j,-j)}$ *are the unique solution of the following four dimensional differential equation,*

$$\frac{-a_j(t)}{a_1(t)+a_2(t)+a_m^{(1)}(t)+a_m^{(2)}(t)} = \lambda_j \frac{d\mu^{(j,j)}}{dt}\left(\mu^{(j,j)}(t)\right) \tag{2.21}$$

$$\frac{-a_m^{(-j)}(t)}{a_1(t)+a_2(t)+a_m^{(1)}(t)+a_m^{(2)}(t)} = \lambda_m \frac{d\mu^{(j,-j)}}{dt}\left(\mu^{(-j,j)}(t)\right) \tag{2.22}$$

with the initial condition,

$$\left(\mu^{(1,1)}(0), \mu^{(1,2)}(0), \mu^{(2,1)}(0), \mu^{(2,2)}(0)\right) = (1,1,1,1) \tag{2.23}$$

and $\left(\mu^{(1,1)}(t), \mu^{(1,2)}(t), \mu^{(2,1)}(t), \mu^{(2,2)}(t)\right) \in \widetilde{\mathcal{D}}_\epsilon$*. The set* $\widetilde{\mathcal{D}}_\epsilon$ *is defined as follows:*

$$\begin{aligned}\widetilde{\mathcal{D}}_\epsilon \coloneqq \Bigg\{ & (\mu^{(1,1)}, \mu^{(1,2)}, \mu^{(2,1)}, \mu^{(2,2)}) \in [0,1]^4 : \\ & \text{for all } j \in \{1,2\}\text{: } \sqrt{\frac{\epsilon}{\lambda_j}} < \mu^{(j,j)},\ \sqrt{\frac{\epsilon}{\lambda_m}} < \mu^{(j,-j)} \\ & \epsilon < a_1 + a_2 + a_m^{(1)} + a_m^{(2)} \Bigg\}.\end{aligned}$$

The functions $a_j(t)$ *and* $a_m^{(j)}(t)$ *are given as follows:*

$$a_j(t) = -\sum_{\substack{u_j+u_{-j}\le K_j(d_j,d_{-j}) \\ d_j+d_{-j}\le d_{\max}}} (d_j - u_j) i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t) + \lambda_j - 2\tau_j(t) - w_j(t),$$

$$a_m^{(j)}(t) = -\sum_{\substack{u_j+u_{-j}\le K_j(d_j,d_{-j}) \\ d_j+d_{-j}\le d_{\max}}} (d_{-j} - u_{-j}) i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t) + \lambda_m - \tau_m(t) - w_m^{(j)}(t),$$

and $\tau_m(t)$ *is given by* (2.18)*. Also, the solution of the four dimensional differential equation satisfies the equality* (2.19)*.*

Proof. See Appendix A.4.2. □

The significance of this result is in demonstrating that the set of ODEs from Section 2.6 can be reduced to a set of four dimensional ODEs (which has a unique solution). Note that this dimension reduction applies to the sample-path of the adoption process and not just the final population of active nodes as suggested by the mean-field approximation of Section 2.4.

The denominator of all four equations given by (2.21) and (2.22) are the same. Since in $\widetilde{\mathcal{D}}_\epsilon$ this quantity is bounded away from zero by $\epsilon > 0$, it is safe to remove this term from the denominator of the differential equations for equilibrium analysis. More specifically, if we consider a particle at $(1, 1, 1, 1)$ whose movement is governed by (2.21) and (2.22), removing the denominator will not change the trajectory of the particle, but will affect its speed. Hence, after some simple algebra we find that the trajectory of (2.21)-(2.22) is the same as the trajectory of the following system of differential equation:

$$\frac{d\mu^{(j,j)}}{dt} = F_{(j,j)}(\mu^{(j,j)}, \mu^{(j,-j)}) - \mu^{(j,j)}, \tag{2.24}$$

$$\frac{d\mu^{(j,-j)}}{dt} = F_{(j,-j)}(\mu^{(-j,-j)}, \mu^{(-j,j)}) - \mu^{(j,-j)}, \tag{2.25}$$

for $j \in \{1, 2\}$ with the same initial conditions (2.23), where the functions $F_{(j,j)}$ and $F_{(j,-j)}$ are given as follows:

$$\begin{aligned} F_{(j,j)}(\mu^{(j,j)}, \mu^{(j,-j)}) \coloneqq & \sum_{\substack{u_j+u_{-j}\le K_j(d_j,d_{-j}) \\ d_j+d_{-j}\le d_{\max}}} \frac{d_j}{\lambda_j}\, i^{(j)}_{d_j,d_{-j},0,0}(0) \\ & Bi(u_j; d_j - 1, 1 - \mu^{(j,j)}) Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}) + \frac{w_j(0)}{\lambda_j}, \end{aligned} \tag{2.26}$$

$$\begin{aligned} F_{(j,-j)}(\mu^{(-j,-j)}, \mu^{(-j,j)}) \coloneqq & \sum_{\substack{u_j+u_{-j}\le K_j(d_j,d_{-j}) \\ d_j+d_{-j}\le d_{\max}}} \frac{d_j}{\lambda_m}\, i^{(-j)}_{d_{-j},d_j,0,0}(0) \\ & Bi(u_j; d_j - 1, 1 - \mu^{(-j,j)}) Bi(u_{-j}; d_{-j}, 1 - \mu^{(-j,-j)}) + \frac{w^{(-j)}_m(0)}{\lambda_m}. \end{aligned} \tag{2.27}$$

Remark II.2. In derivation of (2.24) and (2.25), we have used the following equalities, which

are true as long as $\boldsymbol{\mu} = \left(\mu^{(1,1)}, \mu^{(1,2)}, \mu^{(2,1)}, \mu^{(2,2)}\right) > \mathbf{0}$ component-wise or $\boldsymbol{\mu} = \mathbf{0}$:

$$a_j = \lambda_j \mu^{(j,j)}(\mu^{(j,j)} - F_{(j,j)}(\mu^{(j,j)}, \mu^{(j,-j)})) \tag{2.28}$$
$$a_m^{(-j)} = \lambda_m \mu^{(-j,j)}(\mu^{(j,-j)} - F_{(j,-j)}(\mu^{(-j,-j)}, \mu^{(-j,j)})) \tag{2.29}$$

These equalities are algebraic and straightforward.

Remark II.3. There is a clear connection between the function $\boldsymbol{F}$ given by (2.9) and the function $\boldsymbol{F} \coloneqq (F_{(1,1)}, F_{(1,2)}, F_{(2,1)}, F_{(2,2)})$. Note that as $n \to \infty$:

$$\begin{aligned}
&i^{(j)}_{d_j,d_{-j},0,0}(0)\, d_j/\lambda_j \to \mathbb{P}_{j*,m}(d_j, d_{-j})(1 - \alpha_j(d_j, d_{-j})),\\
&i^{(-j)}_{d_{-j},d_j,0,0}(0)\, d_j/\lambda_m \to \mathbb{P}_{-j,m*}(d_{-j}, d_j)(1 - \alpha_{-j}(d_{-j}, d_j)),\\
&w_j(0)/\lambda_j \to \sum_{d_j + d_{-j} > d_{\max}} \mathbb{P}_{j*,m}(d_j, d_{-j})(1 - \alpha_j(d_j, d_{-j})),\\
&w_m^{(-j)}(0)/\lambda_m \to \sum_{d_j + d_{-j} > d_{\max}} \mathbb{P}_{-j,m*}(d_{-j}, d_j)(1 - \alpha_{-j}(d_{-j}, d_j)).
\end{aligned}$$

The same terms appear in (2.6) and (2.7) if we assume $K_j(d_j, d_{-j}) = d_j + d_{-j}$ for all $d_j + d_{-j} > d_{\max}$ and $j \in \{1, 2\}$. Almost all properties of these two functions are the same. Abusing notation, we use the same symbol for both, the distinction is clear from the context.

Note that the function $\boldsymbol{F}$ is well-defined on $[0, 1]^4$, hence, we do not need any restriction on its domain. We continue with some basic properties of the function $\boldsymbol{F}$. As the first step, we prove that $\boldsymbol{F}$ is increasing in each of its components.

Lemma II.4. *If $\boldsymbol{\mu} \geq \boldsymbol{\mu}'$ component-wise with $\boldsymbol{\mu} \neq \boldsymbol{\mu}'$, then $\boldsymbol{F}(\boldsymbol{\mu}) \geq \boldsymbol{F}(\boldsymbol{\mu}')$ component-wise, and $\boldsymbol{F}(\boldsymbol{\mu}) \neq \boldsymbol{F}(\boldsymbol{\mu}')$.*

Proof. See Appendix A.4.3. □

To analyze the equilibrium of (2.24)-(2.25), we use the LaSalle Invariance Principle [74]. The following Lemma characterizes the most important properties of $\boldsymbol{F}$ which then enables us to invoke this principle.

Lemma II.5. *Let $\mathcal{U} \subseteq [0, 1]^4$ be the largest connected set containing $\mathbf{1} \coloneqq (1, 1, 1, 1)$ such that $\forall \boldsymbol{\mu} \in \mathcal{U}$, $\boldsymbol{\mu} \geq \boldsymbol{F}(\boldsymbol{\mu})$. Then we have the followings:*

(i) $\boldsymbol{F}(\mathcal{U}) \subseteq \mathcal{U}$.

(ii) $\mathcal{U}$ *is closed and compact.*

(iii) $\forall \boldsymbol{u} \in \mathcal{U}$, $\lim_{n\to\infty} \boldsymbol{F}^n(\boldsymbol{u})$ *converges to some point* $\boldsymbol{u}_* \in \mathcal{U}$, *which is a fixed point of* $\boldsymbol{F}$.

(iv) If $\boldsymbol{u}_* \in [0,1]^4$ is a fixed point of $\boldsymbol{F}$, then for any $\boldsymbol{u} \geq \boldsymbol{u}_*$ such that $\boldsymbol{u}$ and $\boldsymbol{u}_*$ are equal in at least one component, we have $\boldsymbol{u} \notin \mathcal{U}$.

Proof. See Appendix A.4.4. □

An immediate and important corollary of the above Lemma is the following.

Corollary II.6. *Let $\boldsymbol{\mu}_*$ denote the closest fixed point of $\boldsymbol{F}$ to $\mathbf{1}$ in sup norm, i.e.,*

$$\boldsymbol{\mu}_* \coloneqq \operatorname*{arg\,min}_{\boldsymbol{u}:\boldsymbol{F}(\boldsymbol{u})=\boldsymbol{u}} \|\boldsymbol{u} - \mathbf{1}\|_\infty,$$

where $\|\boldsymbol{x} - \boldsymbol{y}\|_\infty \coloneqq \max_i |x_i - y_i|$. Then, we have $\boldsymbol{\mu}_* \in \mathcal{U}$, and for all $\boldsymbol{u} \in \mathcal{U} \cap \{\boldsymbol{x} : \mathbf{1} \geq \boldsymbol{x} \geq \boldsymbol{\mu}_*\}$:

$$\boldsymbol{\mu}_* = \lim_{n\to\infty} \boldsymbol{F}^n(\boldsymbol{u}).$$

Proof. The proof follows by parts (iii) and (iv) of Lemma II.5, and the fact that $\mathbf{1} \in \mathcal{U}$. □

Finally, we characterize the equilibrium point at which the ODE (2.24)-(2.25) settles starting from $(1,1,1,1)$ and also provide an iterative method to find it.

Theorem II.7. *Consider the following ODE:*

$$\frac{d\boldsymbol{\mu}}{dt} = \boldsymbol{F}(\boldsymbol{\mu}) - \boldsymbol{\mu} \qquad \boldsymbol{\mu}(0) = \mathbf{1}, \qquad \boldsymbol{\mu} \in [0,1]^4. \tag{2.30}$$

The solution to the ODE settles at $\boldsymbol{\mu}_*$, given by Corollary II.6. Equivalently, it converges to $\boldsymbol{F}^\infty(\mathbf{1})$, where $\boldsymbol{F}^\infty(\boldsymbol{u}) := \lim_{n\to\infty} \boldsymbol{F}^n(\boldsymbol{u})$.

Proof. See Appendix A.4.5. □

Let us for the moment pretend that the initial condition of the ODE in Appendix A.3 is not random. This assumption is not far from being correct since as $n \to \infty$ the initial values given by (A.5) concentrate around their mean (as we also pointed out in Remark II.3). Now, solving the ODE (2.30), we obtain the trajectory of the ODE (2.21)-(2.22) as long as $\boldsymbol{\mu}(t) \in \widetilde{\mathcal{D}}_\epsilon$.

By part (iv) of Lemma II.5 and Corollary II.6, the trajectory of ODE (2.30) hits the set

$$\mathcal{A} \coloneqq \{\boldsymbol{x} : \mathbf{1} \geq \boldsymbol{x} \geq \boldsymbol{\mu}_* \text{ and } \exists i,j \in \{1,2\} : \boldsymbol{\mu}_*^{(i,j)} = \boldsymbol{x}^{(i,j)}\} \tag{2.31}$$

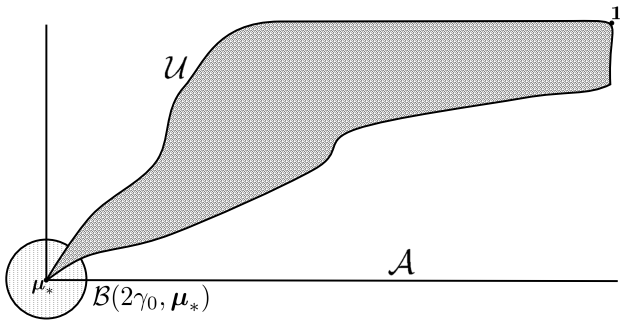


Figure 2.2: Schematic of $\mathcal{A}$, $\mathcal{U}$, and the choice of $\gamma_0 > 0$ in (2.32) .

at $\boldsymbol{\mu}_*$. More specifically, $\mathcal{A} \cap \mathcal{U} = \boldsymbol{\mu}_*$. Now, given the fact that both $\mathcal{A}$ and $\mathcal{U}$ are closed and compact, we can pick $\gamma_0 > 0$ small enough such that

$$\{\boldsymbol{x} \in \mathcal{U} : \exists y \in \mathcal{A} \text{ such that } \|\boldsymbol{x} - \boldsymbol{y}\|_\infty < \gamma_0\} \subset \mathcal{B}(2\gamma_0, \boldsymbol{\mu}_*) \tag{2.32}$$

where $\mathcal{B}(2\gamma_0, \boldsymbol{\mu}_*)$ is a ball of radius $2\gamma_0$ centered at $\boldsymbol{\mu}_*$. Figure 2.2 provides a 2-dimensional schematic for the choice of $\gamma_0 > 0$.

It is easy to see that the equalities (2.28)-(2.29) holds for all $\boldsymbol{\mu} \in \mathcal{U}$. Also, note that $\boldsymbol{\mu}_*$ is the only point in $\mathcal{U}$ such that $a_1 + a_2 + a_m^{(1)} + a_m^{(2)} = 0$. This justifies our claim earlier in this section that removing the denominator of the ODE (2.21)-(2.22) will not affects its trajectory. Since the function $\boldsymbol{F}$ is continuous, it is easy to see that there is an $\epsilon_0 > 0$ such that for all $\epsilon < \epsilon_0$,

$$\mathcal{U} \cap \{\boldsymbol{x} : \mathbf{1} \geq \boldsymbol{x} \geq \boldsymbol{\mu}_*\} \setminus \mathcal{B}(2\gamma_0, \boldsymbol{\mu}_*) \subset \widetilde{\mathcal{D}}_\epsilon \subset \mathcal{U} \cap \{\boldsymbol{x} : \mathbf{1} \geq \boldsymbol{x} \geq \boldsymbol{\mu}_*\}. \tag{2.33}$$

Combining the above argument with Lemma II.1 and Theorem II.2, we can track the Markov process of adoption (the truncated version) upto any ϵ ($< \max(\epsilon_0, \gamma_0)$) neighborhood of $\boldsymbol{\mu}_*$. Intuitively speaking, if $\boldsymbol{\mu}_*$ is an stable equilibrium point of ODE (2.30), then $\boldsymbol{\mu}_*$ should correspond to the natural stopping point of the Markov process of adoption. We can also use equation (2.19) to estimate the natural stopping time. Define t_* as follows:

$$t_* \coloneqq \lambda_m + \frac{\lambda_1}{2} + \frac{\lambda_2}{2} - \frac{\lambda_1}{2}\left(\mu_*^{(1,1)}\right)^2 - \frac{\lambda_2}{2}\left(\mu_*^{(2,2)}\right)^2 - \lambda_m \mu_*^{(1,2)} \mu_*^{(2,1)}.$$

Then, conjecturally, the natural stopping point of the Markov process of adoption is $\sim t_* n$. However, both of these arguments are far from being rigorous.

2.9 Asymptotic Behavior of the Markov Process of Adoption

As we mentioned in Point 2 of Section 2.6, we have assumed that inactive nodes of degree higher than $d_{\max}$ cannot be activated. In this section, we will study the behavior of the original Markov process of adoption at its natural stopping time for large values of n. In what follows, we focus on the truncated version; however, this assumption is relaxed in the statement of the main theorem.

Recall the definition of ϵ_0 and γ_0 from the previous section and let $\epsilon < \max(\epsilon_0, \gamma_0)/16$. We can track the truncated version of the Markov process of adoption upto $\epsilon/2$ neighborhood of $\boldsymbol{\mu}_*$. However, the question is whether the process stops at $\boldsymbol{\mu}_*$. If $\boldsymbol{\mu}_* = \mathbf{0}$, then the answer is clear as we have removed almost all the half-edges. However, the same cannot be said if $\boldsymbol{\mu}_* \neq \mathbf{0}$.

The basic idea is to introduce a perturbation to the process, right after the total number of active half-edges gets close enough to ϵn. Let t_ϵ denote the first time that $a_1(t_\epsilon) + a_2(t_\epsilon) + a_m^{(1)}(t_\epsilon) + a_m^{(2)}(t_\epsilon) \leq \epsilon$, where $a_j(t)$ and $a_m^{(j)}(t)$ are as in Lemma II.1. By Theorem II.2, at time step $t_\epsilon n$, the total number of active half-edges is around ϵn. At this point, we add $2\epsilon n$ active half-edges to each community, ϵn of them for the in-community connections and the rest for the connections between the two communities (in total, we add $4\epsilon n$ active half-edges). Now, the question is how this type of perturbation changes $\boldsymbol{\mu}_*$. Intuitively speaking, if $\boldsymbol{\mu}_*$ is a stable equilibrium point of ODE (2.30), then for small enough $\epsilon > 0$ the above perturbation should not change $\boldsymbol{\mu}_*$ much.

As a side note, recall that $\boldsymbol{\mu}_*$ depends on the initial condition, which is random in our problem setting. However, as $n \to \infty$, the initial conditions (A.5) in Appendix A.3 converges in probability to its mean. More specifically,

$$
\begin{aligned}
& i^{(j)}_{d_j,d_{-j},0,0}(0) \xrightarrow{P} \beta_j \, \mathbb{P}_{j,m}(d_j, d_{-j})(1 - \alpha_j(d_j, d_{-j})), \\
& w_j(0) \xrightarrow{P} \beta_j \sum_{d_j + d_{-j} > d_{\max}} d_j \, \mathbb{P}_{j,m}(d_j, d_{-j})(1 - \alpha_j(d_j, d_{-j})), \\
& w_m^{(j)}(0) \xrightarrow{P} \beta_j \sum_{d_j + d_{-j} > d_{\max}} d_{-j} \, \mathbb{P}_{j,m}(d_j, d_{-j})(1 - \alpha_j(d_j, d_{-j})), \\
& a_j(0) \xrightarrow{P} \beta_j \sum_{d_j, d_{-j}} d_j \, \mathbb{P}_{j,m}(d_j, d_{-j}) \alpha_j(d_j, d_{-j}), \\
& a_m^{(j)}(0) \xrightarrow{P} \beta_j \sum_{d_j, d_{-j}} d_{-j} \, \mathbb{P}_{j,m}(d_j, d_{-j}) \alpha_j(d_j, d_{-j}).
\end{aligned}
\tag{2.34}
$$

where $\beta_j \coloneqq \lim_{n\to\infty} n_j/n$. These initial values appear as coefficients in the definition of function $\boldsymbol{F}$. Hence, as long as the initial values are in a small neighborhood of their mean,

the trajectory of the ODE (2.30) and the properties of the absorbing point $\boldsymbol{\mu}_*$ are almost the same.

The exact same reasoning can be used to justify the fact that $X^n_{U,\delta}$ and $X^n_{L,\delta}$ (defined in in Point 2 of Section 2.6) are close to each other. Let $\boldsymbol{F}_{U,\delta}$ and $\boldsymbol{F}_{L,\delta}$ denote the functions given by (2.26)-(2.27) for $X^n_{U,\delta}$ and $X^n_{L,\delta}$ respectively, where $d_{\max} \coloneqq d_\delta$. Clearly, for all $\boldsymbol{\mu} \in [0,1]^4$, we have $\boldsymbol{F}_{U,\delta}(\boldsymbol{\mu}) \leq \boldsymbol{F}_{L,\delta}(\boldsymbol{\mu})$ as the only difference between these two functions is the value of $(w_1(0), w_m^{(1)}(0), w_m^{(2)}(0), w_2(0))$. Moreover, $\boldsymbol{F}_{L,\delta}(\boldsymbol{\mu}) - \boldsymbol{F}_{U,\delta}(\boldsymbol{\mu}) \leq c\mathbf{1}$, where $c > 0$ is a constant independent of $\boldsymbol{\mu}$, and $c \to 0$ as $\delta \to 0$. Hence, the trajectory of $\boldsymbol{\mu}_{U,\delta}$ and $\boldsymbol{\mu}_{L,\delta}$ given by ODE (2.30) are close to each other. Moreover, using Lemma II.4, we have $\boldsymbol{\mu}_{U,\delta}(t) \leq \boldsymbol{\mu}_{L,\delta}(t)$ for all t. Note that the later inequality may not hold for ODE (2.21)-(2.22) as it has a different time scale.

Theorem II.1. *Consider the Markov process of adoption. Assume there is a constant $0 < d_{\max} \leq \infty$, independent of n, such that $K_j(d_j, d_{-j}) = d_j + d_{-j}$ for all $d_j + d_{-j} > d_{\max}$ and $j \in \{1,2\}$. Suppose the degree regularity conditions given in Definition II.3 hold. Let $\boldsymbol{\mu}(t)$ denote the solution of ODE (2.30) where the function $\boldsymbol{F} = (F_{(1,1)}, F_{(1,2)}, F_{(2,1)}, F_{(2,2)})$ is given by the right-hand side of (2.6)-(2.7) (Mean-Filed equations). Let $\boldsymbol{\mu}_* = \lim_{s\to\infty} \boldsymbol{F}^s(\mathbf{1})$ to be the closest fixed point of $\boldsymbol{F}$ to $\mathbf{1}$ in sup norm. Define the function $T : [0,1]^4 \to \mathbb{R}_+$ using equation (2.19) as follows:*

$$T(\boldsymbol{\mu}) \coloneqq \frac{\lambda_1}{2}\left(1 - \left(\mu^{(1,1)}\right)^2\right) + \frac{\lambda_2}{2}\left(1 - \left(\mu^{(2,2)}\right)^2\right) + \lambda_m\left(1 - \mu^{(1,2)}\mu^{(2,1)}\right), \tag{2.35}$$

and let $t_* \coloneqq T(\boldsymbol{\mu}_*)$. For $t \leq t_*$, define

$$i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t) = \mathbb{P}_{j,m}(d_j, d_{-j})\, Bi(u_j; d_j, 1-\mu_t^{(j,j)})\, Bi(u_{-j}; d_{-j}, 1-\mu_t^{(j,-j)}), \tag{2.36}$$

$$\tau_j(t) = \left(1 - \left(\mu_t^{(j,j)}\right)^2\right) \tag{2.37}$$

where $\boldsymbol{\mu}_t \coloneqq T^{-1}(t)$, and $T^{-1} : [0, t_*] \to \{\boldsymbol{\mu}(x) : \boldsymbol{\mu}$ is the solution of ODE$\}$ is the inverse of the function T restricted to the trajectory of ODE.

1. *Assume $d_{\max} < \infty$ is finite. Then, for any $t < t_*$, we have*

$$\frac{I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(tn)}{n} \xrightarrow{P} i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t),$$
$$\frac{T_j(tn)}{n} \xrightarrow{P} \tau_j(t).$$

2. *Assume $d_{\max} \leq \infty$, where $d_{\max} = \infty$ is interpreted as no constraint on the threshold*

function $K_j(d_j, d_{-j})$. Then, at the natural stopping time of the Markov process of adoption K_n, we have:

$$\frac{I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(K_n)}{n} \xrightarrow{P} i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t_*), \qquad \frac{T_j(K_n)}{n} \xrightarrow{P} \tau_j(t_*),$$

given $\boldsymbol{\mu}_*$ is a stable equilibrium, or equivalently given that the absolute value of all eigenvalues of $\boldsymbol{J}_{\boldsymbol{F}}(\boldsymbol{\mu}_*)$ are smaller than 1, where $\boldsymbol{J}_{\boldsymbol{F}}(\boldsymbol{\mu}_*)$ is the Jacobian matrix of $\boldsymbol{F}$ at $\boldsymbol{\mu}_*$.

Proof. See Appendix A.4.6. □

Remark II.2. As we commented in Point 2 of Section 2.6, the natural stopping time of the original Markov process of adoption X^n is bounded between the same quantities for the truncated versions $X^n_{L,\delta}$ and $X^n_{U,\delta}$. However, this bound does not apply to the whole trajectory and the proof of part 1 of Theorem II.1 is restricted to the case $d_{\max} < \infty$. Nonetheless, we conjecture that part 1 holds for $d_{\max} = \infty$ as well.

Tallying all the (scaled) inactive nodes we can determine the total (scaled) number of inactive nodes in community j. This is an immediate corollary of Theorem II.1.

Corollary II.3. *Let $I_n(k)$ denote the total number of inactive nodes at time step k of the Markov process of adoption X^n.*

1. *Assume $d_{\max} < \infty$ is finite. Then for all $t < t_*$, we have*

$$\frac{I_n(tn)}{n} \xrightarrow{P} \sum_{u_j+u_{-j}\le K_j(d_j,d_{-j})} i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t) = \beta_1\Phi_1(\mu_t^{(1,1)}, \mu_t^{(1,2)}) + \beta_2\Phi_2(\mu_t^{(2,2)}, \mu_t^{(2,1)}).$$

   where $\boldsymbol{\Phi} = (\Phi_1, \Phi_2)$ is given by (2.9).

2. *Assume $d_{\max} \le \infty$ and suppose that $\boldsymbol{\mu}_*$ is a stable equilibrium of ODE (2.30). Then, we have*

$$\frac{I_n(K_n)}{n} \xrightarrow{P} \sum_{u_j+u_{-j}\le K_j(d_j,d_{-j})} i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t_*) = \beta_1\Phi_1(\mu_*^{(1,1)}, \mu_*^{(1,2)}) + \beta_2\Phi_2(\mu_*^{(2,2)}, \mu_*^{(2,1)}),$$

   where $\boldsymbol{\Phi} = (\Phi_1, \Phi_2)$ is given by (2.9).

Similar statements are true for $\frac{A_j(tn)}{n}$ and $\frac{A^{(j)}_m(tn)}{n}$.

Proof. The proof follows from Theorem II.1, and the fact that for large enough $d_{\max}$, the total number of nodes with degree higher than $d_{\max}$ is arbitrary small. □

This machinery can easily be generalized to any finite number of communities. We conclude this section by presenting an obvious generalization to k communities. The degree conditions need not to be revised for this setting. Note that the only assumptions we used in the proof of Theorem II.1 are the degree regularity conditions given in Definition II.3. The other two assumptions of the regularity conditions are necessary to get a uniform simple random graph using configuration model, with positive probability.

Theorem II.4. *Assume there are k communities, and size of communities are given by n_1, n_2, $\cdots$, n_k such that $\sum_i n_i = n$. Assume $\lim_{n\to\infty} n_i/n = \beta_i$ for all $i \in \{1, 2, \cdots, k\}$. Suppose the regularity conditions hold, and define $\lambda_r := \lambda_{r,r}\beta_r$ and $\lambda_m^{(r,s)} := \lambda_{r,s}\beta_r$ (note that $\lambda_{r,s}\beta_r = \lambda_{s,r}\beta_s$), for all $r, s \in \{1, 2, \cdots, k\}$. Assume there is a constant $0 < d_{\max} \leq \infty$, independent of n, such that the inactive nodes with degree higher than $d_{\max}$ cannot be activated. Let X^n denote the Markov process of adoption. Let $\boldsymbol{\mu}(t) = (\mu^{(i,j)}(t))_{i,j\in\{1,2,\cdots,k\}}$ denote the solution of k^2-dimensional ODE*

$$\frac{d\boldsymbol{\mu}}{dt} = \boldsymbol{F}(\boldsymbol{\mu}) - \boldsymbol{\mu} \qquad \boldsymbol{\mu}(0) = \mathbf{1}, \qquad \boldsymbol{\mu} \in [0,1]^{k^2}. \tag{2.38}$$

where the function $\boldsymbol{F}$ is given by Mean-Field equations. Let $\boldsymbol{\mu}_* = \lim_{s\to\infty} \boldsymbol{F}^s(\mathbf{1})$ to be the closest fixed point of $\boldsymbol{F}$ to $\mathbf{1}$ in sup norm. Define $T(\boldsymbol{\mu})$ as follows:

$$T(\boldsymbol{\mu}) := \sum_{r=1}^{k} \frac{\lambda_r}{2}\left(1 - \left(\mu^{(r,r)}\right)^2\right) + \sum_{\substack{r,s=1\\ s\neq r}}^{k} \frac{\lambda_m^{(r,s)}}{2}\left(1 - \mu^{(r,s)}\mu^{(s,r)}\right) \tag{2.39}$$

and let $t_* := T(\boldsymbol{\mu}_*)$. Now, the result of Theorem II.1 and Corollary II.3 holds by using the following functions:

$$\begin{aligned}
i^{(j)}_{d_1,d_2,\cdots,d_r,u_1,u_2,\cdots,u_r} &= \mathbb{P}_{j,m}(d_1, d_2, \cdots, d_k) \prod_{r=1}^{k} Bi(u_r; d_r, 1 - \mu_t^{(j,r)}) \\
\tau_r(t) &= \frac{\lambda_r}{2}\left(1 - \left(\mu_t^{(r,r)}\right)^2\right) \\
\tau_m^{(r,s)} &= \lambda_m^{(r,s)}\left(1 - \mu_t^{(r,s)}\mu_t^{(s,r)}\right)
\end{aligned}$$

where $\boldsymbol{\mu}_t := T^{-1}(t)$. Note that our notation is slightly different from the case of two communities, as we use $I^{(j)}_{d_1,d_2,\cdots,d_r,u_1,u_2,\cdots,u_r}$ to denote the number of nodes in community j with d_r half-edges in community r, such that u_r of them have been already removed.

Proof. The proof of the generalized k follows by recycling the proof of $k = 2$. □

2.10 Contagion Threshold

Recall that by definition, node i in community j with $d^n_{j,i}$ neighbors in community j and $d^n_{m,i}$ neighbors in the other community is an early adopter with probability $\alpha_j(d^n_{j,i}, d^n_{m,i})$. If we assume $K_j(d_j, d_{-j}) \equiv \theta(d_j + d_{-j})$ for some $\theta \in (0,1)$, then the largest value of θ that results in a cascade (i.e., $O(n)$ nodes becoming active) when a small number of nodes ($o(n)$, often taken to be a constant number) are initially seeded is called the contagion threshold; denote it by θ_*. Morris [57] showed that $\theta_* \leq 0.5$ and the upper-bound is loose for many graphs. It's argued that the contagion threshold of the graph family can be calculated by choosing $\alpha_j(d_j, d_{-j}) \equiv \alpha$, letting $\alpha \to 0$, and varying θ. In this section we will formalize this intuition and characterize the contagion condition for more general threshold functions.

Let $\boldsymbol{\alpha} = \{\alpha_j(d_j, d_{-j})\}_{j,d_j,d_{-j}}$ represent the seeding strategy. Let's rewrite the function $\boldsymbol{F}(\boldsymbol{\mu})$ as $\boldsymbol{F}(\boldsymbol{\alpha}, \boldsymbol{\mu})$ to emphasis on the dependency of function $\boldsymbol{F}$ over the seeding strategy. The question of interest is the final proportion of adopters, if the seeding effects only a finite number of population, i.e., the proportion of early adopters goes to 0 as $n \to \infty$. We provide an answer this question in the following theorem.

Theorem II.1. *Consider an arbitrary sequence $\{\boldsymbol{\alpha}_s\}_{s=1}^{\infty}$ that represent a sequence of non-zero seeding strategies that converges to zero in sup norm, i.e., $\|\boldsymbol{\alpha}_s\|_\infty \to 0$. Let $\mathcal{U}(\mathbf{0}) \subseteq [0,1]^4$ to be the largest connected set containing $\mathbf{1}$ such that $\forall \boldsymbol{\mu} \in \mathcal{U}(\mathbf{0}), \boldsymbol{\mu} \geq \boldsymbol{F}(\mathbf{0}, \boldsymbol{\mu})$. If $\mathcal{U}(\mathbf{0})$ is singleton, i.e., $\mathcal{U}(\mathbf{0}) = \{\mathbf{1}\}$, then the final proportion of adopters converges to 0 as $\|\boldsymbol{\alpha}_s\|_\infty \to 0$. Otherwise, the final proportion of adopters is strictly positive, and we have*

$$\lim_{s\to\infty} \boldsymbol{\mu}_*(\boldsymbol{\alpha}_s) = \lim_{r\to\infty} \boldsymbol{F}^r(\mathbf{0}, \boldsymbol{u}) \qquad \forall \boldsymbol{u} \in \mathcal{U}(\mathbf{0}) \cap \{\boldsymbol{x} : \boldsymbol{\mu}_*(\mathbf{0}) \leq \boldsymbol{x} \leq \mathbf{1}\} \setminus \{\mathbf{1}\},$$

where $\boldsymbol{\mu}_*(\boldsymbol{\alpha}_s) \coloneqq \lim_{r\to\infty} \boldsymbol{F}^r(\boldsymbol{\alpha}_s, \mathbf{1})$, and $\boldsymbol{\mu}_*(\mathbf{0})$ is the closest fixed point of $\boldsymbol{F}(\mathbf{0}, \mathbf{1})$ to $\mathbf{1}$ other than $\mathbf{1}$ itself.

Proof. See Appendix A.4.7. □

As in the proof of Lemma A.2, it is easy to see that $\mathcal{U}(\mathbf{0})$ is singleton if and only if $\mathbf{0}$ is an stable equilibrium of $\boldsymbol{F}(\mathbf{0}, \cdot)$. This gives the following easy criteria to check whether contagion happens or not: contagion does not happen if and only if all eigenvalues of $\boldsymbol{J}_{\boldsymbol{F}(\mathbf{0},\cdot)}(\mathbf{1})$ are smaller than 1, where $\boldsymbol{J}_{\boldsymbol{F}(\mathbf{0},\cdot)}(\mathbf{1})$ is the Jacobian matrix of $\boldsymbol{F}(\mathbf{0}, \cdot)$ at $\mathbf{1}$. From the proof of Lemma II.4, all elements of $\boldsymbol{J}_{\boldsymbol{F}(\mathbf{0},\cdot)}(\mathbf{1})$ are non-negative. Now, using Perron-Frobenius theorem we get the following equivalent result.

Theorem II.2. *Contagion happens if and only if $\rho\big(\boldsymbol{J}_{\boldsymbol{F}(\mathbf{0},\cdot)}(\mathbf{1})\big)$ is greater than 1, where $\rho(\cdot)$ is the Perron-Frobenius eigenvalue of a non-negative matrix, and $\boldsymbol{J}_{\boldsymbol{F}(\mathbf{0},\cdot)}(\mathbf{1})$ is the Jacobian matrix of $\boldsymbol{F}(\mathbf{0}, \cdot)$ at $\mathbf{1}$.*

The discussion on contagion can also be generalized to k communities with exactly the same statement as in Theorem II.2.

2.11 Poisson Degree Distributions

We will now specialize our results to Poisson degree distributions. An Erdos-Renyi random graph is an example of a graph family that asymptotically yields a Poisson degree distribution. The two community stochastic block model is then the appropriate generalization of the Erdos-Renyi random graph that will asymptotically produce Poisson degree distributions within the community and across the communities. We will show in the following results that under some symmetry assumptions for the threshold and the advertising strategy, the solution of ODE (2.30) simplifies considerably. In the case of Poisson degree distribution, we assume

$$\mathbb{P}_{j,m}(d_j, d_{-j}) = \mathrm{e}^{-\lambda_{j,j}} \frac{(\lambda_{j,j})^{d_j}}{d_j!} \times \mathrm{e}^{-\lambda_{j,m}} \frac{(\lambda_{j,m})^{d_{-j}}}{d_{-j}!}$$

Note that $\mathbb{P}_{j*,m}(d_j, d_{-j}) = \mathbb{P}_{j,m}(d_j - 1, d_{-j})$ and $\mathbb{P}_{j,m*}(d_j, d_{-j}) = \mathbb{P}_{j,m}(d_j, d_{-j} - 1)$.

Theorem II.1. *Assume the threshold of each node depends on its community and the total number of its neighbors, i.e.* $K_j(d_j, d_{-j}) = K_j(d_j + d_{-j})$. *Moreover, assume the advertisement strategy is based on the community affiliation and the total number of neighbors, i.e.* $\alpha_j(d_j, d_{-j}) = \alpha_j(d_j + d_{-j})$. *Now, if the asymptotic degree distributions are Poisson with parameters* $\lambda_{1,1}$, $\lambda_{1,m}$, $\lambda_{2,m}$, *and* $\lambda_{2,2}$, *then the solution of ODE* (2.30) *with the function* $\boldsymbol{F}$ *given by the right hand side of* (2.6)-(2.7) *simplifies as follows:* $\mu^{(1,1)}(x) = \mu^{(2,2)}(x)$ *and* $\mu^{(2,1)}(x) = \mu^{(1,2)}(x)$ *for all* $x \geq 0$*; that is to say, the dimension of the differential equations reduces to 2.*

Proof. See Appendix A.4.8. □

The next theorem concerns general distributions.

Theorem II.2. *Assume both the advertisement strategy and the threshold function are symmetric in the following sense:* $\alpha_j(d_j, d_{-j}) = \alpha_{-j}(d_{-j}, d_j)$ *and* $K_j(d_j, d_{-j}) = K_{-j}(d_{-j}, d_j)$ *for all* $d_j, d_{-j} \geq 0$ *and* $j \in \{1, 2\}$. *Also assume that the asymptotic degree distribution in both communities are the same, i.e.* $\mathbb{P}_{1,m} = \mathbb{P}_{2,m}$. *Then,* $\mu^{(1,1)}(x) = \mu^{(2,1)}(x)$ *and* $\mu^{(2,2)}(x) = \mu^{(1,2)}(x)$ *for all* $x \geq 0$*; that is to say, the dimension of the differential equations reduces to 2.*

Proof. See Appendix A.4.9. □

Similar generalization holds for the case of k communities. Given similar assumptions to Theorem II.1 or Theorem II.2, in the case of k communities, the dimension of ODE reduces to k. An immediate corollary is the following which assert that if both the assumptions hold, then the dimension reduces to 1. This is also true for general k.

Corollary II.3. *Assume the assumptions of Theorems II.1 and Theorems II.2 hold, then the dimension of ODE* (2.30) *reduces to one. In this case, the formulation is the same as if there was only one community with asymptotic degree distribution given by Poisson*$(\lambda_{1,1} + \lambda_{1,m})$.

It is interesting to note that given assumptions of Theorems II.1 and Theorems II.2, the contagion threshold is the same as if there was only one community. The derivation of contagion threshold matches with the ones presented in [34, 35] for case of one community.

2.12 Numerical Investigation

We present some numerical results using the analysis presented above. The main point is to show how the community structure impacts seeding strategies. A natural question to ask is "what is the best seeding strategy given a budget constrain?". In this section, we also formalize this question and provide a partial answer to it using a gradient based heuristic algorithm.

By Theorem II.1, the solution of the ODE (2.30) tracks the Markov process of adoption upto points arbitrary close to $\boldsymbol{\mu}_*$. Moreover, if $\boldsymbol{\mu}_*$ is an stable equilibrium, the process stops at $\boldsymbol{\mu}_*$ with high probability. Now, by Corollary II.3, our strategy is to pick $\boldsymbol{\alpha}_*$ that minimizes $\beta_1\Phi_1(\boldsymbol{\mu}_*) + \beta_2\Phi_2(\boldsymbol{\mu}_*)$.

Specifically, consider the following budget constraint which constraints the expected number of early adopters:

$$\sum_{j,d_j,d_{-j}} \beta_j \, \mathbb{P}_{j,m}(d_j, d_{-j}) \, \alpha_j(d_j, d_{-j}) = \mathscr{B},$$

where $\mathscr{B} > 0$ is the total available budget. We now formulate the "optimum seeding strategy"

as follows:

$$\begin{aligned}
\inf_{\boldsymbol{\alpha}} \quad & \beta_1 \Phi_1(\boldsymbol{\alpha}, \boldsymbol{\mu}_*) + \beta_2 \Phi_2(\boldsymbol{\alpha}, \boldsymbol{\mu}_*) \\
\text{subject to} \quad & (i) \text{ Budget constraint: } \sum_{j, d_j, d_{-j}} \beta_j \, \mathbb{P}_{j,m}(d_j, d_{-j}) \, \alpha_j(d_j, d_{-j}) = \mathscr{B}, \\
& (ii) \text{ Definition of } \boldsymbol{\mu}_*\text{: } \boldsymbol{\mu}_* = \operatorname*{arg\,min}_{\boldsymbol{u}: \boldsymbol{F}(\boldsymbol{\alpha}, \boldsymbol{u}) = \boldsymbol{u}} \|\boldsymbol{u} - \mathbf{1}\|_\infty, \\
& (ii)' \text{ Definition of } \boldsymbol{\mu}_*\text{: } \boldsymbol{\mu}_* = \lim_{s \to \infty} \boldsymbol{F}^s(\boldsymbol{\alpha}, \mathbf{1}).
\end{aligned}$$

Note that by Corollary II.6, the conditions (ii) and $(ii)'$ are equivalent. As the constraints are highly nonlinear, we take a heuristic approach to provide a partial answer to this optimization problem. We comment that if $\boldsymbol{\mu}_*$ is not a stable equilibrium point, then the process may not stop at $\boldsymbol{\mu}_*$ as we don't know what happens beyond this point. However, the above formulation is heuristically our best bet to maximize the contagion.

For sake of simplicity, let us assume the degrees are uniformly bounded. This assumption is justified by noting that the total number of half-edges associated with high degree nodes is small. The fixed point $\boldsymbol{\mu}_*$ is a function of $\boldsymbol{\alpha}$. Abusing notation, we denote this by $\boldsymbol{\mu}_*(\boldsymbol{\alpha})$. Note that $\boldsymbol{\mu}_*(\boldsymbol{\alpha}) = \boldsymbol{F}(\boldsymbol{\alpha}, \boldsymbol{\mu}_*(\boldsymbol{\alpha}))$. Using the chain rule we have

$$\boldsymbol{J}_{\boldsymbol{\mu}_*}(\boldsymbol{\alpha}) = \boldsymbol{J}_{\boldsymbol{F}(\cdot, \boldsymbol{\mu}_*(\boldsymbol{\alpha}))}(\boldsymbol{\alpha}) + \boldsymbol{J}_{\boldsymbol{F}(\boldsymbol{\alpha}, \cdot)}(\boldsymbol{\mu}_*(\boldsymbol{\alpha})) \, \boldsymbol{J}_{\boldsymbol{\mu}_*}(\boldsymbol{\alpha})$$

where $\boldsymbol{J}_{\boldsymbol{\mu}_*}(\boldsymbol{\alpha})$ is the Jacobian matrix of $\boldsymbol{\mu}_*(\cdot)$ at $\boldsymbol{\alpha}$, $\boldsymbol{J}_{\boldsymbol{F}(\cdot, \boldsymbol{\mu})}(\boldsymbol{\alpha})$ is the Jacobian matrix of $\boldsymbol{F}(\cdot, \boldsymbol{\mu})$ at $\boldsymbol{\alpha}$, and $\boldsymbol{J}_{\boldsymbol{F}(\boldsymbol{\alpha}, \cdot)}(\boldsymbol{\mu})$ is the Jacobian matrix of $\boldsymbol{F}(\boldsymbol{\alpha}, \cdot)$ at $\boldsymbol{\mu}$. Hence,

$$\boldsymbol{J}_{\boldsymbol{\mu}_*}(\boldsymbol{\alpha}) = \big(\boldsymbol{I} - \boldsymbol{J}_{\boldsymbol{F}(\boldsymbol{\alpha}, \cdot)}(\boldsymbol{\mu}_*(\boldsymbol{\alpha}))\big)^{-1} \boldsymbol{J}_{\boldsymbol{F}(\cdot, \boldsymbol{\mu}_*(\boldsymbol{\alpha}))}(\boldsymbol{\alpha}), \tag{2.40}$$

where $\boldsymbol{I}$ is the identity matrix.

The heuristic seeding algorithm is an iterative algorithm that has two stages. The logic behind the algorithm is simple: at *Stage 1*, the algorithm tries to find the best direction for updating the seeding strategy while keeping the budged constraint, and at *Stage 2*, the algorithm validates the choice of the direction. Fix some $\xi_0 > 0$, and pick $\xi_0 < \xi < 1$ arbitrary. Also pick $\boldsymbol{\alpha}$ arbitrary such that conditions (i), the budget constraint, holds. The heuristic algorithm is given as follows:

Stage 1. Numerically solve the following linear optimization problem:

$$\begin{aligned}
\inf_{\Delta\boldsymbol{\alpha}=[\Delta\boldsymbol{\alpha}_j(d_j,d_{-j})]_{j,d_j,d_{-j}}} \quad & \left\langle \Delta\boldsymbol{\alpha}\,,\, \sum_j \boldsymbol{J}_{\Phi_j(\cdot,\boldsymbol{\mu}_*(\boldsymbol{\alpha}))}(\boldsymbol{\alpha}) + \boldsymbol{J}_{\Phi_j(\boldsymbol{\alpha},\cdot)}(\boldsymbol{\mu}_*(\boldsymbol{\alpha}))\, \boldsymbol{J}_{\boldsymbol{\mu}_*}(\boldsymbol{\alpha}) \right\rangle \\
\text{subject to} \quad & \sum_{j,d_j,d_{-j}} \beta_j\, \mathbb{P}_{j,m}(d_j,d_{-j})\, (\alpha_j(d_j,d_{-j}) + \xi\, \Delta\boldsymbol{\alpha}_j(d_j,d_{-j})) = \mathscr{B}, \\
& \alpha_j(d_j,d_{-j}) + \xi\, \Delta\boldsymbol{\alpha}_j(d_j,d_{-j}) \in [0,1] \text{ for all } j,d_j,d_{-j} \\
& \|\Delta\boldsymbol{\alpha}\|_F = 1
\end{aligned}$$

where $\langle A, B\rangle \coloneqq \sum a_{i,j} b_{i,j}$ is the Frobenius inner product, $\boldsymbol{J}_{\Phi_j(\cdot,\boldsymbol{\mu})}(\boldsymbol{\alpha})$ is the Jacobian matrix of $\Phi_j(\cdot,\boldsymbol{\mu})$ at $\boldsymbol{\alpha}$, $\boldsymbol{J}_{\Phi_j(\boldsymbol{\alpha},\cdot)}(\boldsymbol{\mu})$ is the Jacobian matrix of $\Phi_j(\boldsymbol{\alpha},\cdot)$ at $\boldsymbol{\mu}$, and $\|\cdot\|_F$ is the Frobenius norm.

Stage 2. Let $\boldsymbol{\alpha}_{\text{new}} = [\alpha_j(d_j,d_{-j}) + \xi\, \Delta\boldsymbol{\alpha}_j(d_j,d_{-j})]_{j,d_j,d_{-j}}$. If

$$\beta_1\Phi_1(\boldsymbol{\alpha}, \boldsymbol{\mu}_*(\boldsymbol{\alpha})) + \beta_2\Phi_2(\boldsymbol{\alpha}, \boldsymbol{\mu}_*(\boldsymbol{\alpha})) > \beta_1\Phi_1(\boldsymbol{\alpha}_{\text{new}}, \boldsymbol{\mu}_*(\boldsymbol{\alpha}_{\text{new}})) + \beta_2\Phi_2(\boldsymbol{\alpha}_{\text{new}}, \boldsymbol{\mu}_*(\boldsymbol{\alpha}_{\text{new}})),$$

then update $\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha}_{\text{new}}$ and go to *Stage 1*. Otherwise, update $\xi \leftarrow \xi/2$. If $\xi < \xi_0$ terminate the algorithm, otherwise go to *Stage 1*.

Next, we compare different seeding strategies using Theorem II.1, Corollary II.3, and Corollary II.6. Most of our results will be for Poisson degree distributions, owing to analytical simplifications and the fact there are only three parameters to tune. Moreover, for simplicity we assume that the threshold function is given by $k_j(d_j,d_{-j}) = \theta \times (d_j + d_{-j}) - 1$ where $\theta = 0.25$. Henceforth, we assume $\beta_1 = \beta_2$.

The nodes that are seeded by the advertisers are early adopters. A few strategies that we will consider are: (1) Random seeding: First, we assume the advertiser does not even know about the existence of two communities. This scenario is named as *global seeding*. Second, we assume the advertiser knows the community structure, and decides to seed just asymmetrically in the two communities. This advertisement strategy is denoted by *local seeding*. (2) Degree-targeted seeding: the advertiser knows the degree distribution of the network and the identity of the nodes that possess a certain degree, but does not know the underlying connectivity structure.

In Figure 2.3, we assume the both in-community degree distribution and out-community degree distribution are Poisson distributions with parameters $\lambda_{1,1} = \lambda_{2,2} = \lambda_{\text{in}}$ and $\lambda_{1,2} = \lambda_{2,1} = \lambda_{\text{out}}$. The figure suggests that if the communities are symmetric, and if they are well-connected ($\lambda_{\text{out}} = 1$), then the best strategy is to put the whole budget in one community. In Figure 2.4 we consider the general case where distributions can have different parameters

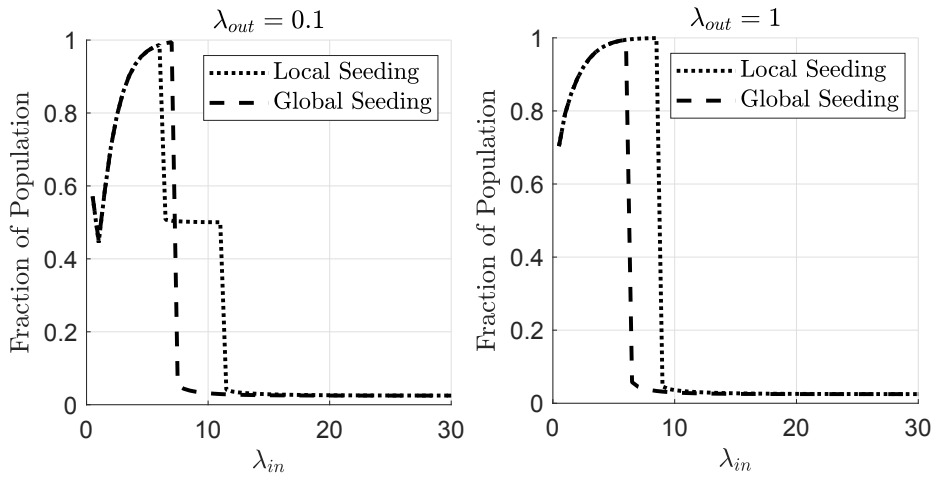


Figure 2.3: Random seeding strategy on symmetric communities. The ratio of of early adopters is 2.5% of the population. λ_{in} denotes the in-community connectivity, and λ_{out} denotes the out-community connectivity.

in the two communities, i.e., $\lambda_{1,1}$ and $\lambda_{2,2}$ need not to be equal. We also assume $\lambda_{1,2} = \lambda_{2,1} = \lambda_{\text{out}} = 1$. In this case, the community structure dramatically changes the cascade potential: there are scenarios where global seeding can cause a cascade while local seeding won't, and *vice-versa*.

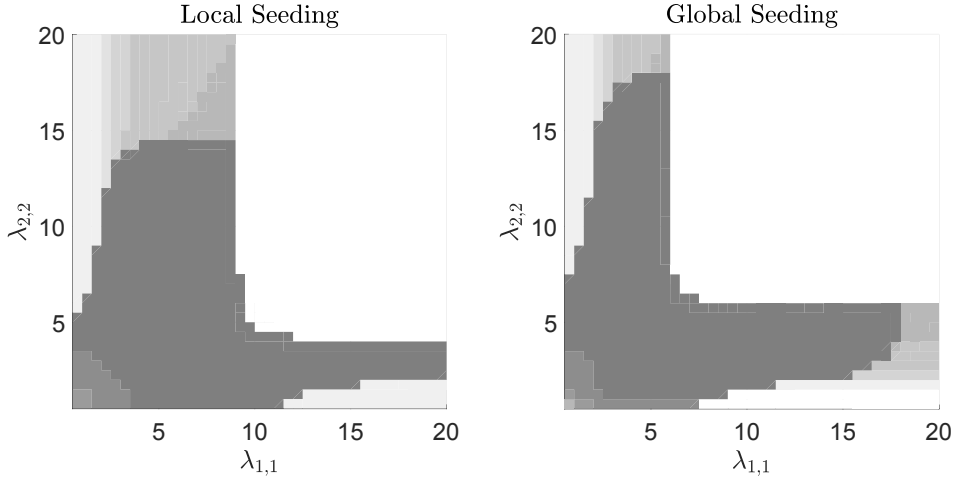


Figure 2.4: Random seeding strategy on asymmetric communities. The ratio of of early adopters is 2.5% of the population. $\lambda_{1,1}$ and $\lambda_{2,2}$ are parameters of the in-community distributions. $\lambda_{1,2} = \lambda_{2,1} = \lambda_{\text{out}} = 1$ in both cases. Intensity of grayscale indicates the final proportion of adopters.

Next, we consider degree-targeted seeding in Figures 2.5 and 2.6. In general, high-degree nodes can potentially stop a cascade if they are not early adopters; hence it might make sense to seed these nodes in each community. We will consider the following case: the budget is

spread equally in both communities, denoted by $(0.5, 0.5)$; the budget is concentrated in community 1, denoted by $(1, 0)$; and that 25% of budget is in community 1, denoted by $(0.25, 0.75)$. The outer connectivity is given by $\lambda_{1,2} = \lambda_{2,1} = \lambda_{\text{out}} = 1$. The main observation is the dramatic difference in the proportion of final adopters based on how asymmetric the targeting is. Additionally, a higher inter-community connectivity leads to a bigger cascade. Also note that seeding nodes with the highest degree gives better result than random seeding. In Figure 2.6, we compare the highest degree strategy with the heuristic algorithm proposed in the beginning of the section, using the same parameters as in Figure 2.5. The results are dramatically different. We also illustrate the resulted seeding distribution, $\alpha_1(d_1, d_2)$ and $\alpha_2(d_2, d_1)$, for $\lambda_{1,1} = 18$ and $\lambda_{2,2} = 10.5$ in Figure 2.7. These two figures highlights the importance of community structure on the optimum seeding strategy.

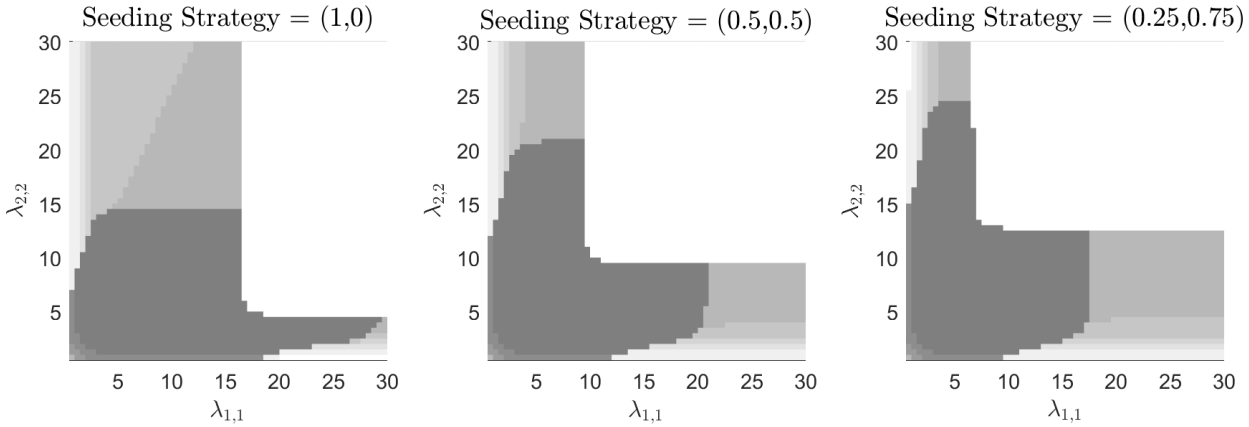


Figure 2.5: Highest degree seeding strategy on asymmetric communities.The ratio of of early adopters is 2.5% of the population. $\lambda_{1,1}$ and $\lambda_{2,2}$ are parameters of the in-community distributions. $\lambda_{1,2} = \lambda_{2,1} = \lambda_{\text{out}} = 1$ in all three cases. Intensity of grayscale indicates the final proportion of adopters.

Next, we discuss the evolution of cascade using Theorem II.1 and Corollary II.3. Figure 2.8 illustrates the evolution of active half-edges and inactive nodes in the second community for $\lambda_{1,1} = 7$, $\lambda_{2,2} = 12$, and $\lambda_{1,2} = \lambda_{2,1} = 1$ when the seeding strategy is to put the whole budget in the first community. Figure 2.4 suggests that global seeding strategy will not result in any cascade. On the other hand, a global cascade emerges following local seeding strategy: it develops in the first community and then moves to the next community; this happens when the inactive nodes in community 2 with $d_1 > \theta \times (d_1 + d_2) - 1$ become active, causing a cascade in the second community. In this figure, we also present the total number of active half-edges in the second community for the Markov process of adoption, for $n = 20000$, to numerically validate the result of Theorem II.1.

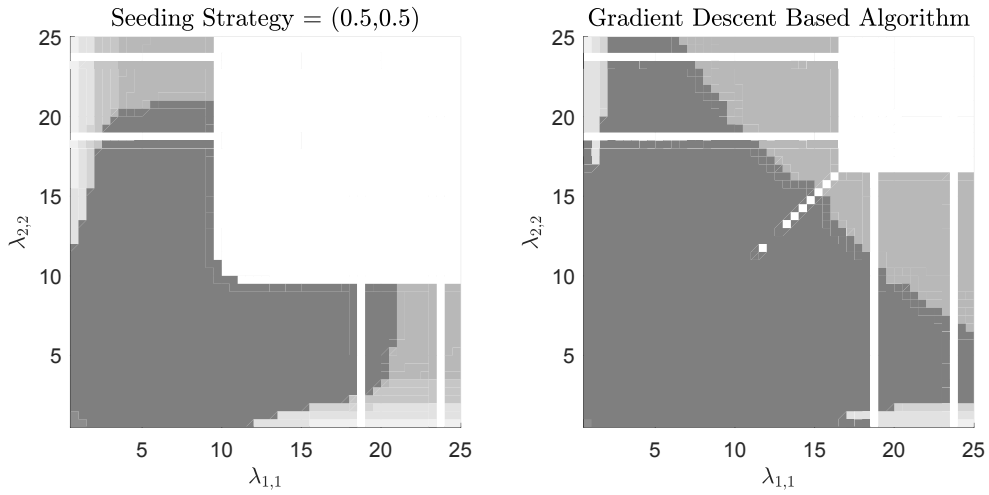


Figure 2.6: Highest degree seeding strategy vs proposed heuristic seeding strategy on asymmetric communities.The ratio of of early adopters is 2.5% of the population. $\lambda_{1,1}$ and $\lambda_{2,2}$ are parameters of the in-community distributions. $\lambda_{1,2} = \lambda_{2,1} = \lambda_{\text{out}} = 1$ in all three cases. Intensity of grayscale indicates the final proportion of adopters.

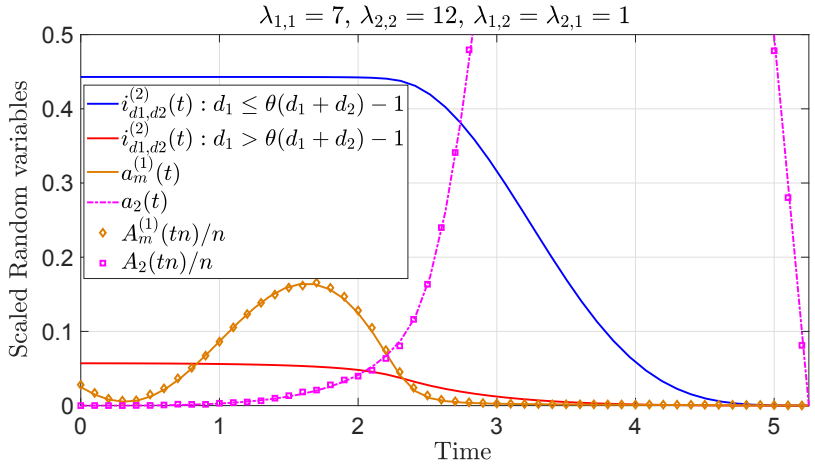


Figure 2.8: Evolution of cascade for $\lambda_{1,1} = 7$, $\lambda_{2,2} = 12$, and $\lambda_{1,2} = \lambda_{2,1} = 1$ using local seeding strategy. The total proportion of early adopters is 2.5%.

Finally, there are scenarios where neither global nor local seeding strategy can cause cascade. Figure 2.9 illustrate the evolution of active half-edges, using $(0.25, 0.75)$ seeding strategy. As can be seen from Figure 2.5, the only seeding strategy (among the ones in the figure) that can cause a global cascade for $\lambda_{1,1} = 17$, $\lambda_{2,2} = 12$ and $\lambda_{1,2} = \lambda_{2,1} = 1$ is $(0.25, 0.75)$. Active half-edges in both communities get close to zero, but nevertheless a cascade happens in the second community. This cascade then moves to the first community, and almost all nodes adopt the new technology. This example illustrates the importance of active half-edges $a^{(2)}_m(t)$ in triggering a cascade in the first community. We also include

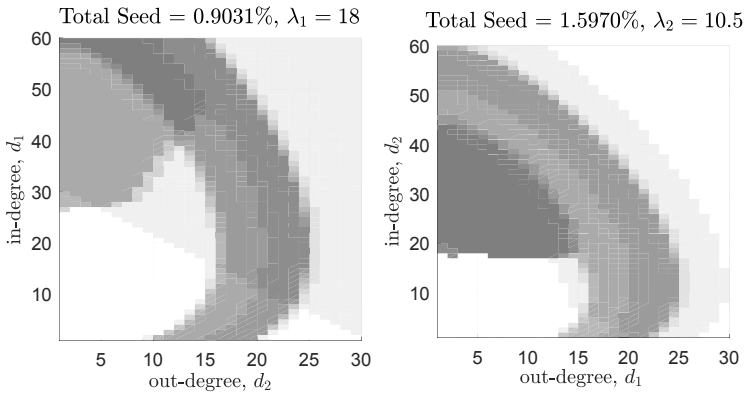


Figure 2.7: Distribution of $\alpha_1(d_1, d_2)$ (left) and $\alpha_2(d_2, d_1)$ (right) in the proposed heuristic seeding algorithm for $\lambda_{1,1} = 18$, $\lambda_{2,2} = 10.5$ and $\lambda_{\text{out}} = 1$. The ratio of of early adopters is 2.5% of the population. Intensity of grayscale indicates the value of $\alpha_j(d_j, d_{-j})$.

the total number of active half-edges for the Markov process of adoption, for $n = 20000$, to numerically validate the result of Theorem II.1.

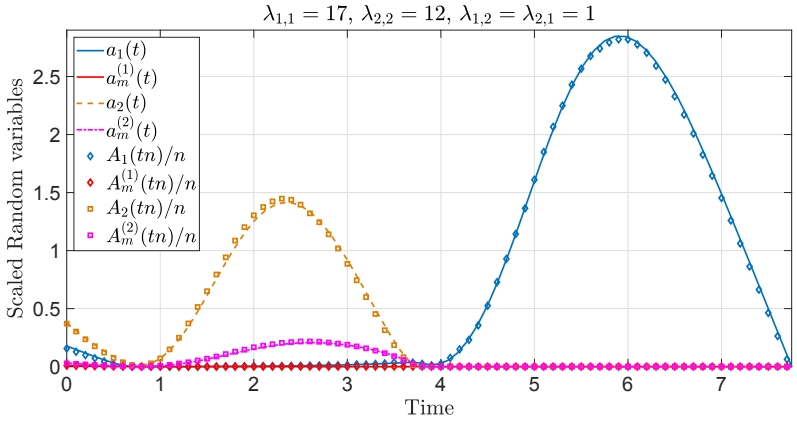


Figure 2.9: Evolution of cascade, $\lambda_{\text{out}} = 1$, $\lambda_{in,1} = 17$ and $\lambda_{in,2} = 12$; proportion of early adopters is 2.5%, x axis is time and y axis is the quantity of corresponding scaled variables. The seeding strategy is given by $(0.25, 0.75)$.

CHAPTER III

Erlang Weighted Tree

Notation

Bold symbols are used for sequences while random variables are denoted by capital letters and their realization by small letters. $\mathbb{R}_+$ denotes the set of non-negative real numbers. Similarly, $\mathbb{Z}_+$ denotes the set of non-negative integers. The set of natural numbers is denoted by $\mathbb{N}$. The set of all finite sequences of $\mathbb{N}$ is denoted by $\mathbb{N}^f = \cup_{i=0}^{\infty}\mathbb{N}^i$ with the convention $\mathbb{N}^0 = \{\emptyset\}$. The set of positive integers less than or equal to n is denoted by $[n]$, i.e., $[n] = \{1, 2, \ldots, n\}$. Let $L(\mathbb{R}_+; [0,1])$ be the set of Lebesgue measurable functions from $\mathbb{R}_+$ to $[0,1]$. Let $C^1(\mathbb{R}_+; [0,1])$ be the set of continuously differentiable functions from $\mathbb{R}_+$ to $[0,1]$. The Erlang distribution with parameters $k \in \mathbb{N}$ and $\lambda > 0$ is denoted by $Erlang(\cdot\,; k, \lambda)$, and the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$ is denoted by $Bi(\cdot\,; n, p)$. The Poisson distribution with parameter λ is denoted by $Poiss(\lambda)$, and the geometric distribution with parameter p and support on $\mathbb{N}$ is denoted by $geo(p)$. A random variable X is said to have a moment generating function at $t \in \mathbb{R}$, if $\mathbb{E}[\mathrm{e}^{tX}] < \infty$. For a set S, $\mathcal{P}(S)$ is the set of all Borel probability measures defined on S.

3.1 Introduction

This chapter studies a random tree object called the Erlang Weighted Tree (EWT). The construction of the EWT begins with the construction of the "backbone tree". The backbone tree has more edges, some of which are then pruned to obtain the EWT.

Let $\mathbb{N}^f$ denote the labels of vertices of an infinite tree. Each $\boldsymbol{i} \in \mathbb{N}^f$ is associated with three types of random variables: 1) $n_{\boldsymbol{i}}$ which is the potential number of descendants of the vertex $\boldsymbol{i}$, 2) $v_{\boldsymbol{i}}$ which is the value associated with the vertex $\boldsymbol{i}$, and, 3) $\{\zeta_{(\boldsymbol{i},j)}\}_{j=1}^{n_{\boldsymbol{i}}}$ which represents the cost of the potential edges $\{\boldsymbol{i}, (\boldsymbol{i}, j)\}$ for $j \in \{1, 2, \ldots, n_{\boldsymbol{i}}\}$. The probability distribution of $n_\emptyset$ is given by $P \in \mathcal{P}(\mathbb{N})$ which is assumed to have a finite mean and $P(1) < 1$. The

probability distribution of $n_{\boldsymbol{i}}$ for $\boldsymbol{i} \in \mathbb{N}^f \setminus \mathbb{N}^0$ is given by the shifted distribution $\widehat{P} \in \mathcal{P}(\mathbb{Z}_+)$, i.e., $\widehat{P}(k) = P(k+1)$ for all $k \geq 0$. Conditioned on $n_{\boldsymbol{i}}$, $v_{\boldsymbol{i}}$ is distributed as $Erlang(\cdot\,; n_{\boldsymbol{i}}+1, \lambda)$ for a positive and fixed real value λ. Conditioned on $n_{\boldsymbol{i}}$ and $v_{\boldsymbol{i}}$, $\{\zeta_{(\boldsymbol{i},j)}\}_{j=1}^{n_{\boldsymbol{i}}}$ are $n_{\boldsymbol{i}}$ independent and uniformly distributed random variables over the interval $[0, v_{\boldsymbol{i}}]$. When $n_{\boldsymbol{i}} = 0$, there are no potential edges emanating from vertex $\boldsymbol{i}$. The backbone tree is the connected component of $\emptyset$ with the potential edges as its edge set.

The edges of the backbone tree are pruned to obtain the EWT. Define a rooted tree $\mathbb{T}_\circ = (V, E, \emptyset, w_v, w_e)$, rooted at $\emptyset$, by preserving the edge between the vertices $\boldsymbol{i}$ and $(\boldsymbol{i}, j)$ if and only if $\zeta_{(\boldsymbol{i},j)} < v_{(\boldsymbol{i},j)}$. The mark functions are defined as follows,

$$w_v : V \to \mathbb{N} \times \mathbb{R}_+, \qquad w_v(\boldsymbol{i}) = \begin{cases} (n_\emptyset, v_\emptyset) & \boldsymbol{i} = \emptyset \\ (n_{\boldsymbol{i}} + 1, v_{\boldsymbol{i}}) & \text{otherwise} \end{cases}$$

$$w_e : E \to \mathbb{R}_+, \qquad w_e(\{\boldsymbol{i}, (\boldsymbol{i}, j)\}) = \zeta_{(\boldsymbol{i},j)}.$$

The random rooted tree $\mathbb{T}_\circ$ is called an Erlang Weighted Tree with distribution for the potential degree given by P. Henceforth, we call P the *potential degree distribution*. Let $[\mathbb{T}_\circ]$ denote the equivalence class of $\mathbb{T}_\circ$ up to isomorphisms (over vertex relabelings that preserve the root). Denote by $Er(P, \lambda)$ the probability distribution of $[\mathbb{T}_\circ]$ in G_*, which denotes the set of rooted marked graphs up to isomorphisms. For a formal definition of G_* and related background material, see Section 3.2.1.

Remark III.1. The parameter λ in the definition of $Er(P, \lambda)$ appears only as a scaling factor. Usually, this value is set to be 1, and for ease of notation, $Er(P)$ is used instead of $Er(P, 1)$.

Remark III.2. Throughout this chapter, a non-root vertex $\boldsymbol{i}$ with the mark $(n_{\boldsymbol{i}} + 1, v_{\boldsymbol{i}})$ will be referred to as a vertex of *type* $(n_{\boldsymbol{i}}, v_{\boldsymbol{i}})$.

We will show that EWT appears as the local weak limit of a random graph model introduced by La and Kabkab in [23]. The graph construction starts with a complete graph $K_n = ([n], E_n)$, a sequence of positive integers $\boldsymbol{d}_n = (d_1(n), d_2(n), \cdots, d_n(n))$ and a random cost function C_n that assigns non-negative real values to the edges of K_n, independently. The value of $d_i(n)$ indicates the number of neighbors that vertex i wants to connect to. The value assigned to each edge by C_n is an independent exponentially distributed random variable with parameter $1/n$ that represents the cost of the edge. Each vertex i then selects the $d_i(n)$ lowest cost incident edges and declares them to be its preferred edges. The random graph $G_n = ([n], \widetilde{E}_n)$ is constructed by keeping only those edges of E_n that are preferred by both end vertices. This model is closely related to the k-th nearest neighbor graphs presented by Cooper and Frieze in [43], in which a connection survives as long as at least one individual

involved in the connection is interested in it. The bilateral agreement required in the above random graph model makes the analysis much more challenging.

Main Results

In this work, we derive the following properties of the EWT:

(i) EWT is unimodular. We shall shortly define unimodularity.

(ii) Let $\bar{F}_k(\cdot)$ denote the complementary cumulative distribution function of $Erlang(\cdot\,; k, \lambda)$. The degree distribution of the root is given by,

$$\mathbb{P}(D_\emptyset = d) = \sum_{m=1}^{\infty} P(m) \int_0^\infty \frac{\mathrm{e}^{-x} x^m}{m!} Bi\left(d; m, \int_0^x \frac{1}{x} \sum_{k=1}^{\infty} P(k) \bar{F}_k(y)\, dy\right) dx$$
$$\mathbb{E}[D_\emptyset] = \sum_{m=1}^{\infty} \sum_{k=1}^{\infty} P(m)P(k) \int_0^\infty \bar{F}_k(y)\bar{F}_m(y)\, dy.$$

Note that this is the asymptotic degree distribution of the random graph family in [23]. Unlike the canonical branching processes, the degree distribution of a vertex at depth $l \geq 0$ depends on l.

(iii) The probability of extinction is given by,

$$\mathbb{P}(\{\text{extinction}\}) = \sum_{m=1}^{\infty} P(m) \int_{x=0}^{\infty} \frac{\mathrm{e}^{-x} x^m}{m!} (q(x))^m\, dx,$$

where $q(\cdot) \in C^1(\mathbb{R}_+; [0,1])$ is the smallest fixed point (point-wise smaller than all the other fixed points) of the operator $T : L(\mathbb{R}_+; [0,1]) \to C^1(\mathbb{R}_+; [0,1])$ defined as,

$$T(f)(x) := \begin{cases} \frac{1}{x} \sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} \left(\int_{z=0}^{y} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz + \int_{z=y}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f(z)^{k-1}\, dz \right) dy, & \text{x>0} \\ \sum_{k=1}^{\infty} P(k) \int_{z=0}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f(z)^{k-1}\, dz, & \text{x=0} \end{cases}.$$

This fixed point is also the pointwise limit of $T^l(\mathbf{0})(\cdot)$ as l goes to infinity where $\mathbf{0}(\cdot)$ is the zero function. If the probability of extinction equals 1, then the function $q(x) \equiv 1$ for all $x \geq 0$ is the unique fixed point of T (upto sets of measure 0). If the probability of extinction is smaller than 1, then assuming that the moment generating function of $n_\emptyset$ exists for some $\theta > 0$, the operator T has exactly two fixed points: $q(\cdot)$ and $\mathbf{1}(\cdot)$, where $\mathbf{1}(\cdot)$ is the all 1 function.

(iv) Assume that the moment generating function of $n_\emptyset$ exists for some $\theta > 0$. Define the function $L(\beta, x)$ as follows,

$$L(\beta, x) = \sum_{i=0}^{\infty} G_i(x) \left(\frac{-1}{\beta}\right)^i,$$

where $g_2(x) = \mathrm{e}^{-x} \sum_{k=2}^{\infty} P(k) \frac{x^{k-2}}{(k-2)!}$ and the function $G_i(x)$ is defined recursively via

$$\begin{aligned} G_0(x) &= 1, \\ G_i(x) &= \int_x^{\infty} \int_{z=y}^{\infty} g_2(z) G_{i-1}(z)\, dz\, dy \qquad \forall i > 0. \end{aligned}$$

Let Z_l denote the number of vertices at generation l. We have,

$$\frac{\mathbb{E}[Z_l]}{{\beta_0}^l} \xrightarrow{l\to\infty} \left(\int_0^{\infty} \sum_{k=1}^{\infty} P(k) \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f_0(z)\, dz \right)^2$$

where β_0 is the smallest zero of the function $L(\beta, 0)$, $f_0(x) = L(\beta_0, x)\sqrt{C_N}$ for all $x \in \mathbb{R}_+$, and $C_N = (\int_0^{\infty} g_2(y) L(\beta_0, y)^2\, dy)^{-1}$ is the normalization factor so that

$$\int_0^{\infty} g_2(y) f_0(y)^2 dy = 1.$$

(v) Let the assumption of part (iv) hold and let $\beta_0 > 1$. Then there is a random variable W such that $Z_l/{\beta_0}^l$ converges to W almost surely and in L^2. Moreover, $Z_l \sim {\beta_0}^l W$, i.e., β_0 is the growth rate of Z_l, and the proportion of various types converges to a non-random limit.

(vi) Let the assumption of part (iv) hold. If $\beta_0 > 1$, then the probability of extinction is less than 1, otherwise it equals 1. Moreover, if $\beta_0 > 1$, then the number of vertices at generation n as $n \to \infty$, goes to either 0 or ∞.

The organization of the rest of the chapter is as follows: In Section 3.2, we provide the necessary background: a short background on random graphs and local weak convergence, a short note on the point process perspective of a branching process, and a short description of spectral theory for compact self-adjoint bounded linear operators. In Section 3.3, we describe the finite graph model and discuss the local weak convergence of the finite graph model to the EWT. In Section 3.4, we begin with the unimodularity of EWT. Then we derive the degree distribution of the root vertex, expected number of vertices at generation l and the

probability of extinction. Finally, we discuss the point process perspective and derive the growth rate of the branching process and the phase transition. In Section 3.5, we present some numerical illustrations of our results. Some proofs are presented in the Appendix for ease of presentation.

3.2 Background Material

In this section, we present the necessary background for the rest of the chapter. The essential background on "random graphs and local weak convergence" is mostly based on lecture notes by Bordenave [25] and the work of Aldous and Lyons [44]. The background on the "point process perspective of a branching process" is based on chapter 3 of Harris's book [30]. We use this background in Section 3.4.5 which proves the most significant result of our work. The background on the "spectral theorem for compact self-adjoint bounded linear operators" is based on a classical text book in functional analysis by Lax [75] and the work of Toland [47]. The related topics from this subject are used in Section 3.4.5; however, we will rederive the main theorems presented in this section using a probabilistic approach.

3.2.1 Random Graphs and Local Weak Convergence

We start with a few graph terminologies that are used in the chapter. Let $G = (V, E)$ denote an undirected graph, where V is the set of vertices (finite or countably infinite), and E is the set of edges. A rooted graph $G_\circ = (V, E, \emptyset)$ is a graph with a distinguished vertex $\emptyset \in V$. Vertices $v_1, v_2 \in V$ are said to be neighbors, if $\{v_1, v_2\} \in E$. The degree of a vertex $v \in V$, denoted by d_v, is the number of its neighbors. A graph G is said to be locally-finite if the degree of each vertex is finite. A path p of length $n-1$ is an ordered sequence of vertices $(v_1, v_2, \ldots, v_n)$ where $\{v_i, v_{i+1}\} \in E,\ \forall i < n$. A graph G is said to be connected if there is a path between every pair of vertices.

Two graphs $G = (V, E)$ and $G' = (V', E')$ are said to be isomorphic if there is a bijection σ from V to V' such that $\{v_1, v_2\} \in E$ if and only if $\sigma(\{v_1, v_2\}) \coloneqq \{\sigma(v_1), \sigma(v_2)\} \in E'$. The function σ is called an isomorphism from G to G'. A rooted-isomorphism between two rooted graphs is an isomorphism that maps the root vertices to each other.

A network $N = (V, E, w_v, w_e)$ is a graph (V, E) with mark functions $w_v : V \to \Omega_1$ and $w_e : E \to \Omega_2$, where Ω_1 and Ω_2 are the mark spaces. A rooted network is a network with a distinguished vertex as the root vertex. In this chapter, the mark spaces are assumed to be $\Omega_1 = \mathbb{N} \times \mathbb{R}_+$ and $\Omega_2 = \mathbb{R}_+$, which are complete separable metric spaces equipped with the

following metrics,

$$d_{\Omega_1}\left((m,x),(n,y)\right) = \sqrt{(m-n)^2+(x-y)^2} \qquad \forall m,n \in \mathbb{N},\ \forall x,y \in \mathbb{R}_+$$
$$d_{\Omega_2}\left(x,y\right) = |x-y| \qquad \forall x,y \in \mathbb{R}_+$$

Two networks N and N' are said to be isomorphic if there is a bijection map from V to V' that preserve the edges as well as the marks. A rooted-isomorphism between two rooted networks $N_\circ$ and $N'_\circ$ is an isomorphism that maps the root of one network to the other. For a rooted network $N_\circ = (V, E, \emptyset, w_v, w_e)$, $[N_\circ]$ denotes the class of rooted networks that are isomorphic to $N_\circ$. Let $G_*(\Omega_1, \Omega_2)$ denote the set of all isomorphism classes $[N_\circ]$, where $N_\circ$ ranges over all connected locally-finite rooted networks with mark spaces Ω_1 and Ω_2. For notational simplicity, we use G_* instead of $G_*(\Omega_1, \Omega_2)$.

There is a natural way to define a metric on G_*. Consider a connected rooted network $N_\circ = (V, E, \emptyset, w_v, w_e)$[1] and the corresponding rooted graph $G_\circ = (V, E, \emptyset)$. The depth of a vertex $v \in V$ is defined to be the infimum length of the paths from v to the root vertex. Let $(G_\circ)_t = (V_t, E_t, \emptyset)$ denote the subgraph of $G_\circ$ where V_t is the set of vertices in V at depth less than or equal to t from ϕ, and E_t is the set of edges in E between the vertices in V_t. For any $[N_\circ], [N'_\circ] \in G_*$, a natural way to define a distance is given by

$$d_{G_*}([N_\circ],[N'_\circ]) = \frac{1}{R+1},$$

where

$$R = \sup\left\{ t \geq 0 : \begin{array}{c} \text{there exists a rooted-isomorphism } \sigma \text{ from } (G_\circ)_t \text{ to} \\ (G'_\circ)_t \text{ such that } \forall \boldsymbol{v} \in V_t \text{ and } \forall \boldsymbol{e} \in E_t,\ d_{\Omega_1}(w_v(\boldsymbol{v}), \\ w'_v(\sigma(\boldsymbol{v}))) < t^{-1} \text{ and } d_{\Omega_2}(w_e(\boldsymbol{e}), w'_e(\sigma(\boldsymbol{e}))) < t^{-1} \end{array} \right\}.$$

Note that in the definition of R, the isomorphism is between the rooted graphs and not the corresponding rooted networks. The space G_* equipped with d_{G_*} is a complete separable metric (Polish) space [25]. Define $\mathcal{P}(G_*)$ as the set of all probability measures on G_* and endow this space with the topology of weak convergence. Since G_* is a Polish space, the space $\mathcal{P}(G_*)$ is a Polish space as well [25] with the Lévy-Prokhorov metric.

The members of G_* are unlabeled connected locally-finite rooted networks; however, there is a way to generalize the framework to unrooted, not necessarily connected, finite networks. Consider a finite network $N = (V, E, w_v, w_e)$. For every vertex $v \in V$, define $N(v)$ to be the connected component of the vertex v in the network N. Let $N_\circ(v)$ denote the rooted version

[1] Strictly speaking, $N_\circ$ is a member of the equivalence class $[N_\circ]$.

of $N(v)$, rooted at v, and define $\delta_{[N_\circ(v)]} \in \mathcal{P}(G_*)$ to be the Dirac measure that assigns 1 to $[N_\circ(v)]$ and 0 to any other member of G_*. Define $U(N) \in \mathcal{P}(G_*)$ as follows,

$$U(N) = \frac{1}{|V|} \sum_{v \in V} \delta_{[N_\circ(v)]}. \tag{3.1}$$

The probability measure $U(N)$ is the law of $[N_\circ(\emptyset)]$, where $\emptyset \in V$ is picked uniformly at random. This probability measure captures the local structure of N as viewed from a randomly chosen vertex. The notion of local weak convergence studies the weak limit of $\{U(N_n)\}_{n\geq 0}$ for a sequence of finite networks $\{N_n\}_{n\geq 0}$.

Definition III.1. (**Local Weak Limit**) A sequence of finite networks $\{N_n\}_{n\geq 1}$ has a *local weak limit* $\rho \in \mathcal{P}(G_*)$ if $U(N_n) \xrightarrow{w} \rho$.

A necessary condition for a probability measure $\rho \in \mathcal{P}(G_*)$ to be a local weak limit is *unimodularity* [44] which is defined next. Let $G_{**}(\Omega_1, \Omega_2)$, or more simply G_{**} denote the set of isomorphism classes of connected locally-finite networks with an ordered pair of distinct vertices. Let $N_{\circ\circ}(\emptyset, v)$ denote a network in G_{**}. Equip G_{**} with the natural metric $d_{G_{**}}$ which is defined in the same way as d_{G_*}.

Definition III.2. (**Unimodularity**) A measure $\rho \in \mathcal{P}(G_*)$ is said to be *unimodular* if for all Borel functions $f : G_{**} \to \mathbb{R}_+$,

$$\int \sum_{v \in V} f([N_{\circ\circ}(\emptyset, v)])\, d\rho([N_\circ(\emptyset)]) = \int \sum_{v \in V} f([N_{\circ\circ}(v, \emptyset)])\, d\rho([N_\circ(\emptyset)]). \tag{3.2}$$

The function f in the definition of unimodularity ranges over all Borel functions from G_{**} to $\mathbb{R}_+$; however, it is sufficient to consider Borel functions $f : G_{**} \to \mathbb{R}$ that assign a non-zero value to a doubly rooted network only if the roots are adjacent. This property is known as involution invariance [44].

Lemma III.3. *(***Involution Invariance***) A measure $\rho \in \mathcal{P}(G_*)$ is unimodular if and only if the equality* (3.2) *holds for all Borel functions $f : G_{**} \to \mathbb{R}_+$ such that $f([N_{\circ\circ}(\emptyset, v)]) = 0$ unless $\{\emptyset, v\} \in E$.*

It is easy to show that the class of local weak limits are unimodular. The question of whether the class of unimodular measures and local weak limits coincide or not, is still an open problem.

3.2.2 Point Process Perspective of a Branching Process

Let $\Omega = \mathbb{Z}_+ \times \mathbb{R}$ denote the type space. A point distribution $\omega = ((m_1, x_1), a_1; (m_2, x_2), a_2; \ldots; (m_k, x_k), a_k)$ on type space Ω is a finite set of vertices that consists of a_j vertices of

type (m_j, x_j) for $k \in \mathbb{Z}_+ \setminus \{0\}$, and $k = 0$ corresponds to null point distribution. Let $\mathscr{P}_\Omega$ denote the set of all point distributions. A point distribution $\omega \in \mathscr{P}_\Omega$ defines a natural set function $\widetilde{\omega}(\cdot)$ over all subsets of Ω,

$$\widetilde{\omega}(\mathcal{A}) \coloneqq \sum_{(m_j, x_j) \in \mathcal{A}} a_j, \qquad \forall \mathcal{A} \subset \Omega.$$

It is easy to see that there is a one-to-one correspondence between point distributions and set functions satisfying the following conditions:

(a) for any $\mathcal{A} \subset \Omega$, $\widetilde{\omega}(\mathcal{A})$ is a non-negative integer.

(b) if $\mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_k$ are disjoint subsets of Ω, then $\widetilde{\omega}\left(\cup_j \mathcal{A}_j\right) = \sum_j \widetilde{\omega}(\mathcal{A}_j)$.

(c) if $\mathcal{A}_1 \supset \mathcal{A}_2 \supset \ldots$ are subsets of Ω and $\cap_j \mathcal{A}_j = \emptyset$, then $\widetilde{\omega}(\mathcal{A}_j) = 0$ for all sufficiently large j.

Abusing notation, we write $\omega(\cdot)$ as the set function generated by the point distribution $\omega \in \mathscr{P}_\Omega$. We now define a σ-algebra on $\mathscr{P}_\Omega$.

A rational interval is a subset of Ω with elements of the form (m, x) such that $\underline{q}_1 \leq m < \overline{q}_1$ and $\underline{q}_2 \leq x < \overline{q}_2$, where $\underline{q}_1$ and $\overline{q}_1$ are non-negative integers, $\underline{q}_2$ and $\overline{q}_2$ are non-negative rational numbers, and $\overline{q}_1$ and $\overline{q}_2$ are allowed to be ∞. A basic set is a finite union of rational intervals or the empty set. Given a collection of basic sets $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k$ and a set of non-negative integers $r_1, r_2, \cdots, r_k$, a cylinder set in $\mathscr{P}_\Omega$ is defined as follows:

$$\mathcal{C}(\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k; r_1, r_2, \cdots, r_k) = \{\omega \in \mathscr{P}_\Omega : \omega(\mathcal{A}_j) = r_j, \ \forall j \in [k]\}.$$

Let $\mathscr{A}$ denote the σ-algebra generated by the cylinder sets. The following theorem defines a probability measure on $(\mathscr{P}_\Omega, \mathscr{A})$ using a set of probability distributions defined over basic sets. The proof is based on the Kolmogorov extension theorem [30].

Theorem III.4. *Let functions $p(\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k; r_1, r_2, \cdots, r_k)$ be given, defined for any collection of basic sets and non-negative integers, satisfying the following.*

(a) $p(\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k; r_1, r_2, \cdots, r_k)$ is a probability distribution on k-tuples of non-negative integers $r_1, r_2, \cdots, r_k$.

(b) $p(\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k; r_1, r_2, \cdots, r_k)$ is permutation invariant, that is to say $\forall \sigma \in S_k$

$$p(\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k; r_1, r_2, \cdots, r_k) = p(\mathcal{A}_{\sigma(1)}, \mathcal{A}_{\sigma(2)}, \cdots, \mathcal{A}_{\sigma(k)}; r_{\sigma(1)}, r_{\sigma(2)}, \cdots, r_{\sigma(k)}).$$

(c) The functions p are consistent,

$$p(\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k; r_1, r_2, \cdots, r_k) = \sum_{r_{k+1}=0}^{\infty} p(\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k, \mathcal{A}_{k+1}; r_1, r_2, \cdots, r_k, r_{k+1}).$$

(d) If $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k$ are disjoint sets and $\mathcal{A} = \cup_{j=1}^{k} \mathcal{A}_j$, then $p(\mathcal{A}, \mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k; r, r_1, r_2, \cdots, r_k) = 0$ unless $r = \sum_{j=1}^{k} r_j$.

(e) If $\mathcal{A}_1 \supset \mathcal{A}_2 \supset \cdots$ and $\cap_{j=1}^{\infty} \mathcal{A}_j = \emptyset$, then $\lim_{j\to\infty} p(\mathcal{A}_j; 0) = 1$.

Then there exists a unique probability measure P on $\mathscr{A}$ that coincides with the functions p whenever $\mathcal{A}_j$'s are basic sets,

$$P(\omega(\mathcal{A}_1) = r_1, \omega(\mathcal{A}_2) = r_2, \cdots, \omega(\mathcal{A}_k) = r_k) = p(\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k; r_1, r_2, \cdots, r_k).$$

For a point distribution $\omega = ((m_1, x_1), a_1; (m_2, x_2), a_2; \ldots; (m_k, x_k), a_k) \in \mathscr{P}_\Omega$ and a function $h : \Omega \to \mathbb{R}$, the random integral $\int h d\omega$ is defined as $\sum_{j=1}^{k} a_j \times h(m_j, x_j)$. The term "random" refers to the randomness of ω. Given a probability distribution P on $(\Omega, \mathscr{P}_\Omega)$, the Moment Generating Functional (MGF) of P is defined as follows:

$$\Phi(s) = \mathbb{E}\mathrm{e}^{-\int s\, d\omega} = \int_{\mathscr{P}_\Omega} \mathrm{e}^{-\int s\, d\omega}\, dP(\omega),$$

where $s : \Omega \to \mathbb{R}_+$ is a non-negative function. Similarly, given some conditions on a functional Φ defined over non-negative functions $s : \Omega \to \mathbb{R}_+$, there exists a unique probability measure P on $(\Omega, \mathscr{P}_\Omega)$ with MGF Φ [30]. This correspondence implies the following theorem:

Theorem III.5. *Let $\Phi_1, \Phi_2, \cdots, \Phi_k$ be MGF's on $(\Omega, \mathscr{P}_\Omega)$. Then the functional $\Phi(s) = \Phi_1(s)\Phi_2(s)\cdots\Phi_k(s)$ defines an MGF on $(\Omega, \mathscr{P}_\Omega)$.*

Now, we revisit the EWT from point processes perspective. For any collection of basic sets $\{\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k\}$ and non-negative integers $\{r_1, r_2, \cdots, r_k\}$ define $p_{(m,x)}(\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k; r_1, r_2, \cdots, r_k)$ to be the probability that a vertex of type (m, x) has r_j children of type $\mathcal{A}_j$ for $j \in [k]$. Then, the functions $p_{(m,x)}$ determines a unique probability measure $P^{(1)}_{(m,x)}$ on $(\mathscr{P}_\Omega, \mathscr{A})$ (Theorem III.4). The probability measure $P^{(1)}_{(m,x)}$ determines, in turn, an MGF $\Phi^{(1)}_{(m,x)}$. Note that $p_{(m,x)}$, for any fixed set of arguments $\mathcal{A}_i$s and r_is , is a Borel-measurable function of $(m, x) \in \Omega$ where Ω is equipped with the same metric as Ω_1. Using the Theorem III.5, for any point distribution $\omega = ((m_1, x_1), a_1; (m_2, x_2), a_2; \ldots; (m_k, x_k), a_k) \in \mathscr{P}_\Omega$ the functional

$\Phi_\omega^{(1)}$

$$\Phi_\omega^{(1)}(s) = [\Phi_{(m_1,x_1)}^{(1)}(s)]^{a_1}[\Phi_{(m_2,x_2)}^{(1)}(s)]^{a_2}\cdots[\Phi_{(m_1,x_1)}^{(1)}(s)]^{a_k},$$

is an MGF and induces a probability measure $P_\omega^{(1)}$ on $(\mathscr{R}_\Omega, \mathscr{A})$. The probability measure $P_\omega^{(1)}$ is the transition probability function of a generalized Markov chain defined by the branching process,

$$P_\omega^{(1)}(\mathcal{A}) = \mathbb{P}(Z_{l+1} \in \mathcal{A} | Z_l = \omega) \qquad \forall \mathcal{A} \in \mathscr{A},$$

where Z_l is the point distribution of vertices at depth l (abusing the notation). As in regular Markov chains, the $m+n$-step transition probability function satisfies the following Chapman–Kolmogorov recurrence relation,

$$P_\omega^{(m+n)}(\mathcal{A}) = \int_{\mathscr{R}_\Omega} P_{\omega'}^{(n)}(\mathcal{A})\, dP_\omega^{(m)}(\omega') \qquad \forall \mathcal{A} \in \mathscr{A}.$$

The MGF of $P_\omega^{(n)}$ is denoted by $\Phi_\Omega^{(n)}$ which satisfies the following recurrence relation,

$$\Phi_\Omega^{(m+n)} = \Phi_\Omega^{(n)}(-\log \Phi_\cdot^{(m)}).$$

3.2.3 Spectral Theorem for Compact Self-adjoint Bounded Linear Operators

A linear space $\mathcal{X}$ equipped with a norm $\|\cdot\|_\mathcal{X}$ is called normed linear space. A complete normed linear space is called Banach space. Every Banach space is a metric space. A metric space $(\mathcal{X}, d)$ is called separable if it has a countable dense subset, i.e, a set $\{x_1, x_2, x_3, \cdots\}$ with the property that for all $\epsilon > 0$ there exists x_n such that $d(x_n, x) < \epsilon$. A linear space equipped with an inner-product is called an inner-product space. We say $S = \{e_\alpha\}_{\alpha\in I}$ is an orthonormal basis of an inner-product space $\mathcal{X}$, if $\forall x \in \mathcal{X}$ we have $x = \sum_{\alpha\in I}\langle x, e_\alpha\rangle$ and $\langle e_\alpha, e_\beta\rangle = 0$ when $\alpha \neq \beta$ and $\langle e_\alpha, e_\alpha\rangle = 1$. A Banach space with a norm induced by an inner-product is called Hilbert space. It is easy to prove that a Hilbert space is separable if and only if it has a countable orthonormal basis.

Let $\mathcal{X}$ and $\mathcal{U}$ be normed linear spaces over $\mathbb{C}$ with norms $\|\cdot\|_\mathcal{X}$ and $\|\cdot\|_\mathcal{U}$, respectively. A map $M : \mathcal{X} \to \mathcal{U}$ is called a bounded linear map if it is linear and there exists $b > 0$ such that $\forall x \in \mathcal{X},\ \|Tx\|_\mathcal{U} \leq b\|x\|_\mathcal{X}$. Let $\mathcal{L}(\mathcal{X},\mathcal{U})$ denote the set of all bounded linear maps from $\mathcal{X}$ to $\mathcal{U}$ and equip this space with the natural norm $\|M\|_\mathcal{L} = \sup_{x\in\mathcal{X},\|x\|_\mathcal{X}=1}\|Mx\|_\mathcal{U}$. Then $(\mathcal{L}(\mathcal{X},\mathcal{U}), \|\cdot\|_\mathcal{L})$ is a normed linear space. It is easy to check that if $\mathcal{U}$ is a Banach space then $\mathcal{L}(\mathcal{X},\mathcal{U})$ is also a Banach space.

Consider $\mathcal{L}(\mathcal{X}, \mathcal{X})$ together with its natural binary map, i.e., if $N, M \in \mathcal{L}(\mathcal{X}, \mathcal{X})$ then $N \cdot M(x) \coloneqq N(M(x))$ for all $x \in \mathcal{X}$. This forms an algebra over $\mathbb{C}$ which is called a normed algebra. A complete normed algebra is called Banach algebra. A operator M in a Banach algebra is called invertible if $\exists N \in \mathcal{L}(\mathcal{X}, \mathcal{X})$ such that $N \cdot M = M \cdot N = I$, where $I \in \mathcal{L}(\mathcal{X}, \mathcal{X})$ is the identity map.

Let $\mathcal{L}(\mathcal{X}, \mathcal{X})$ be a Banach algebra over $\mathbb{C}$ and let $M \in \mathcal{L}(\mathcal{X}, \mathcal{X})$. The resolvent set of M is given by

$$\rho(M) = \{\lambda \in \mathbb{C} : \lambda I - M \text{ is invertible } \}.$$

The set $\sigma(M) = \mathbb{C} \setminus \rho(M)$ is called the spectrum of M. If $\lambda \in \sigma(M)$ then, 1) if $\lambda I - M$ is not one-to-one then λ is called an eigenvalue of M, 2) if $\lambda I - M$ is one-to-one, but $\overline{R(\lambda I - M)} \neq \mathcal{X}$, where $R(N)$ is the range of N, then λ is called a residual of $\sigma(M)$, and 3) if λ is neither an eigenvalue nor a residual, then it is said to be in the continuous spectrum of M. The eigenvalues of M are denoted by $\sigma_p(M)$, the residual spectrum of M is denoted by $\sigma_r(M)$, and the continuous spectrum of M is denoted by $\sigma_c(M)$. The spectrum of M is nonempty, bounded and closed in $\mathbb{C}$. The spectral radius of an operator M is defined as $|\sigma(M)| \coloneqq \max_{\lambda \in \sigma(M)} |\lambda|$.

Theorem III.6. *We have* $|\sigma(M)| = \lim_{n \to \infty} \left(\|M^n\|_{\mathcal{L}}\right)^{\frac{1}{n}}$

Let $\mathcal{X}$ and $\mathcal{U}$ be Banach spaces. A set $S \subset \mathcal{X}$ is called precompact if $\overline{S}$ is compact. A map $M \in \mathcal{L}(\mathcal{X},\mathcal{U})$ is a compact operator if $M(B)$, where B is the ball of radius 1 in $\mathcal{X}$, is precompact in $\mathcal{U}$. The following theorem is the Riesz-Schauder Theorem which is a spectral theorem for compact operators.

Theorem III.7. *Let $\mathcal{X}$ be a Banach space and let $M \in \mathcal{L}(\mathcal{X}, \mathcal{X})$ be a compact operator. Then the spectrum of M satisfies the following:*

(i) *0 is in the spectrum of M unless the dimension of $\mathcal{X}$ is finite.*

(ii) *All non-zero elements of $\sigma(M)$ are in $\sigma_p(M)$.*

(iii) *If λ is a non-zero eigenvalue of M, then λ has finite multiplicity, i.e., the dimension of the null space of $\lambda I - M$ is finite.*

(iv) *If λ_0 is an accumulation point of $\sigma(M)$ then $\lambda_0 = 0$.*

Let $\mathcal{H}$ denote a Hilbert space and let $A \in \mathcal{L}(H, H)$. The adjoint of A, written as A^*, is defined by $\langle x, A^* y\rangle_{\mathcal{H}} \coloneqq \langle Ax, y\rangle_{\mathcal{H}}$ for all $x, y \in \mathcal{H}$. If $A^* = A$ or equivalently $\langle Ax, y\rangle = \langle x, Ay\rangle$, $\forall x, y \in \mathcal{H}$, we say A is symmetric or self-adjoint. The spectral theorem for compact symmetric operators on a Hilbert space $\mathcal{H}$ is given as follows.

Theorem III.8. *Let $\mathcal{H}$ be a Hilbert space and let $A \in \mathcal{L}(\mathcal{H}, \mathcal{H})$ be a compact symmetric operator. Then the spectrum of A satisfies the following properties.*

(i) The spectrum of A is a subset of $\mathbb{R}$.

(ii) If $\lambda, \lambda' \in \sigma_p(A)$ and $\lambda \neq \lambda'$ then the null space of $\lambda I - A$ is orthogonal to the null space of $\lambda' I - A$.

(iii) There exists $x_0 \in \mathcal{H}$ with $\|x_0\|_{\mathcal{H}} = 1$ such that $|\langle Ax_0, x_0\rangle_{\mathcal{H}}| = \sup_{\|x\|_{\mathcal{H}}=1} |\langle Ax, x\rangle_{\mathcal{H}}| = \|A\|_{\mathcal{L}}$, and moreover, x_0 is an eigenvector of A, i.e., $Ax_0 = \lambda x_0$ for some $\lambda \in \mathbb{R}$. The corresponding eigenvalue λ is the largest eigenvalue of A in magnitude.

(iv) (Hilbert-Schmidt) There exists an orthonormal basis of $\mathcal{H}$ consisting of the eigenvectors of A.

Let $\mathcal{H}$ be a Hilbert space. A cone $\mathcal{K} \subset \mathcal{H}$ is a closed convex subset of $\mathcal{H}$ such that for all $\lambda \geq 0$, $\lambda\mathcal{K} \subset \mathcal{K}$ and $\mathcal{K} \cap (-\mathcal{K}) = \{\mathbf{0}\}$ where $(-1)\mathcal{K}$ is denoted as $-\mathcal{K}$. A closed subset $\mathcal{S}$ of $\mathcal{H}$ is said to be invariant under $A \in \mathcal{L}(\mathcal{H}, \mathcal{H})$ if $A\mathcal{S} \subseteq \mathcal{S}$. The following theorem by Toland [47] is a version of the Krein-Rutman Theorem [46] for compact self-adjoint operators.

Theorem III.9. *Suppose $\mathcal{K} \subseteq \mathcal{H}$ is a closed cone such that $\mathcal{K}^{\perp} \coloneqq \{x \in \mathcal{H} : \langle x, y\rangle = 0, \forall y \in \mathcal{K}\} = \{\mathbf{0}\}$. Let $A \in \mathcal{L}(H, H)$ be a compact self-adjoint operator such that $A : \mathcal{K} \to \mathcal{K}$. Define $\mathscr{X}(A) \coloneqq \sup\{\langle Aw, w\rangle_{\mathcal{H}} : \|w\|_{\mathcal{H}} = 1, w \in \mathcal{K}\}$. We have the following.*

(i) $\mathscr{X}(A) > 0$ is the largest eigenvalue of A in magnitude and $\mathscr{X}(A)$ has an eigenvector in $\mathcal{K}$.

(ii) $\mathscr{X}(A) > 0$ is a simple eigenvalue of A.

3.3 Finite Graph Model

Let $K_n = ([n], E_n)$ denote a complete graph, i.e., $E_n = \{\{i, j\} : i, j \in [n], i \neq j\}$. Consider some probability mass function $P(\cdot)$ defined over $\mathbb{N}$. Let $\boldsymbol{d}_n = (d_1(n), d_2(n), \ldots, d_n(n)) \in \mathbb{N}^n$ denote the sequence of potential degrees such that $d_i(n) \leq n-2$ and, as $n \to \infty$, its empirical distribution converges to $P(\cdot)$, i.e.,

$$P_{\boldsymbol{d}_n}(k) = \frac{1}{n}\sum_{i=1}^{n} \delta_{d_i(n)}(k) \xrightarrow{n\to\infty} P(k) \qquad \forall k \in \mathbb{N}.$$

Often, we just write d_i for $d_i(n)$. Let $C_n : E_n \to \mathbb{R}_+$ denote a random function that assigns *i.i.d.* random variables distributed as $Exp(1/n)$ to the edges of K_n. The value of an edge corresponds to the cost of the edge.

Remark III.1. Without loss of generality, we assume the cost of all the edges in K_n are different.

For each vertex i, let $\mathcal{T}_i$ and $\mathcal{P}_i$ denote the threshold and the set of potential neighbors of the vertex i,

$$\mathcal{T}_i = d_i + 1^{st} \text{ smallest value in } \{C_n(\{i,j\}) : j \in [n] \setminus \{i\}\} \tag{3.3}$$

$$\mathcal{P}_i = \{j \in [n] \setminus \{i\} : C_n(\{i,j\}) < \mathcal{T}_i\}. \tag{3.4}$$

Vertices of the graph have the following self-optimizing behavior: they are willing to form an edge only if the cost of the edge is less than each of their thresholds in (3.3) and an edge is formed only if both endpoint vertices are willing. Call the resulting random graph the random graph $G_n = ([n], \widetilde{E}_n)$ with

$$\widetilde{E}_n = \{\{i,j\} \in E_n : i \in \mathcal{P}_j \text{ and } j \in \mathcal{P}_i\}.$$

The bilateral agreement required for establishing an edge causes an interdependence structure; more precisely, inclusion of an edge into $\widetilde{E}_n$ depends on the preference of both ends, which is in turn dictated by the values of all the incident edges. This makes the analysis of the finite graph intricate; however, it is possible to study the model, using the framework of local weak convergence.

Consider the random network $N_n = ([n], \widetilde{E}_n, \widetilde{W}_{v,n}, \widetilde{W}_{e,n})$, where the mark functions are defined as follows:

$$\begin{aligned} &\widetilde{W}_{v,n} : [n] \to \mathbb{N} \times \mathbb{R}, \qquad \widetilde{W}_{v,n}(i) = (d_i, \mathcal{T}_i) && \forall i \in [n], \\ &\widetilde{W}_{e,n} : \widetilde{E}_n \to \mathbb{R}, \qquad \widetilde{W}_{e,n}(\{i,j\}) = C_n(\{i,j\}) && \forall \{i,j\} \in \widetilde{E}_n. \end{aligned}$$

Let $\mathcal{N}(n, \boldsymbol{d}_n)$ denote the law of the random network N_n. Define the random probability measure $U(N_n)$ over G_* as follows,

$$U(N_n) = \frac{1}{n} \sum_{i \in [n]} \delta_{[N_{n,\circ}(i)]},$$

where $N_n \sim \mathcal{N}(n, \boldsymbol{d}_n)$ and $N_{n,\circ}(i)$ is the connected component of i in N_n rooted at i. Taking

expectation with respect to the randomness of the network, for every event $A \in G_*$,

$$\begin{aligned}\mathbb{E}U(N_n)(A) \coloneqq \mathbb{E}\left[U(N_n)(A)\right] &= \frac{1}{n}\sum_{i\in[n]} \mathbb{E}\left[\delta_{[N_{n,\circ}(i)]}(A)\right] \\ &= \frac{1}{n}\sum_{i\in[n]} \mathbb{P}([N_{n,\circ}(i)] \in A).\end{aligned}$$

Hence, $\mathbb{E}U(N_n)$ is the law of $[N_{n,\circ}(\emptyset)]$ where $\emptyset \in [n]$ is a random vertex chosen uniformly from $[n]$. Then the primary motivation of our work is the claim that the sequence of random networks N_n converges locally weakly to the EWT, i.e., $\mathbb{E}U(N_n) \xrightarrow{w} Er(P)$.

As is suggested by Aldous and Steele in [24], the first step to establish the local weak convergence is to guess the object that the finite graph model converges to. Next, we provide an argument to justify the EWT guess.

Aldous [21] proved that the complete graph K_n with *i.i.d.* edge weights distributed as $Exp(1/n)$ is locally tree-like, and it converges to the Poisson Weighted Infinite Tree(PWIT). The idea is to modify the structure of PWIT to capture the behavior of the finite graph model while preserving unimodularity of the asymptotic object. In our graph family the root vertex $\emptyset$ is potentially connected to $n_\emptyset$ other vertices; hence, the $n_\emptyset + 1^{st}$ edge weight in the PWIT is considered as the threshold of the vertex $n_\emptyset$. On the other hand, any non-root vertex with label $\boldsymbol{i}$, needs to know the edge weight of its $n_{\boldsymbol{i}}^{th}$ descendant to decide whether to connect to its "parent" or not. Hence, the edge weight of the $n_{\boldsymbol{i}}^{th}$ descendant in the PWIT is taken to be its real-valued threshold mark if $\boldsymbol{i}$ belongs to the connected component of $\emptyset$. Moreover, a pruning process is added to include the fact that the survival of an edge is based on the marks at both endpoint vertices. Finally, the labels of the descendants of each vertex are permuted to remove the order. This is an essential step to make the object unimodular.

However, there are quite a few technical issues to resolve to make this intuition work. For example, there is interdependence beyond just pairs. The fact that this interdependence can be ignored as was done in the intuitive reasoning that led to the pruned PWIT needs a rigorous proof. It is worth mentioning that the mark space of PWIT and EWT are different and the local weak convergence viewpoint is not the same for these two objects.

Theorem III.2. *Let $N_n \sim \mathcal{N}(n, \boldsymbol{d}_n)$, where $P_{\boldsymbol{d}_n}$ converges weakly to $P(\cdot)$. Then*

$$\mathbb{E}U(N_n) \xrightarrow{w} Er(P).$$

Sketch of the proof. The main body of the proof consists of four steps:

1. Recall that $\mathbb{E}U(N_n)$ is the law of $[N_{n,\circ}(r)]$ for a uniformly chosen $r \in [n]$. The first step is to redefine the construction of the random network N_n, as viewed from r.

2. The random network N_n has an interdependence structure; however as n grows, the dependency weakens. The second step is to exploit this weak dependence and to prove that as n goes to infinity, the connected component of the vertex r becomes locally tree-like.

3. As the dependency weakens, the local structure of $[N_{n,\circ}(r)]$ gets close to the local structure of a rooted tree generated by $Er(P)$. The third step is to prove that for every finite rooted network $\mathbb{T}_\circ \in G_*$ with depth t, the measure assigned to $A_{\mathbb{T}_\circ} = \{[N_\circ] \in G_* : d_{G_*}([N_\circ], \mathbb{T}_\circ) < (1+t)^{-1}\}$ by $\mathbb{E}U(N_n)$ converges to the measure assigned to $A_{\mathbb{T}_\circ}$ by $Er(P)$.

4. Finally, since G_* is a Polish space, the Portmanteau Theorem is applied to show the desired convergence.

The formal proof of the theorem is given in Appendix B.1. □

3.4 Properties of Erlang Weighted Tree

3.4.1 Unimodularity of EWT

From general results on local weak convergence, Theorem III.2 implies that $Er(P, \lambda)$ is unimodular; however, unimodularity of $Er(P, \lambda)$ can be proved directly too. The proof provides more insight into the structure of the EWT.

Theorem III.1. *If $P \in \mathcal{P}(\mathbb{N})$ has a positive and finite mean and $\lambda \in (0, \infty)$, then $Er(P, \lambda)$ is a unimodular measure in $\mathcal{P}(G_*)$*

Proof. Using the involution invariance property, we need to prove for all Borel measurable non-negative functions $f : G_{**} \to \mathbb{R}_+$,

$$\mathbb{E}\left(\sum_{v\sim\emptyset} f(G, \emptyset, v)\right) = \mathbb{E}\left(\sum_{v\sim\emptyset} f(G, v, \emptyset)\right), \tag{3.5}$$

where the expectation is with respect to $Er(P, \lambda)$. Let us expand the left-hand side of (3.5) by conditioning on the potential degree of the root vertex. By linearity of expectation, we have,

$$\mathbb{E}\left(\sum_{v\sim\emptyset} f(G, \emptyset, v)\right) = \sum_{m=1}^{\infty} P(m)\mathbb{E}\left(\sum_{i\sim\emptyset} f(G, \emptyset, i)\middle| n_\emptyset = m\right)$$

$$
\begin{aligned}
&= \sum_{m=1}^{\infty} P(m)\mathbb{E}\left(\sum_{i=1}^{m} f(G,\emptyset,i)\mathbb{1}_{i\sim\emptyset}\middle| n_\emptyset = m\right)\\
&= \sum_{m=1}^{\infty} P(m)\sum_{i=1}^{m}\mathbb{E}\left(f(G,\emptyset,i)\mathbb{1}_{i\sim\emptyset}|n_\emptyset = m\right)\\
&= \sum_{m=1}^{\infty} mP(m)\mathbb{E}\left(f(G,\emptyset,1)\mathbb{1}_{1\sim\emptyset}|n_\emptyset = m\right).
\end{aligned}
$$

where the last equality is based on the symmetric and conditionally independent structure of $\{\zeta_j\}_{j=1}^{n_\emptyset}$ and $\{(n_j, v_j)\}_{j=1}^{n_\emptyset}$ conditioned on $n_\emptyset$. We now expand the term $\mathbb{E}(f(G,\emptyset,1)\mathbb{1}_{1\sim\emptyset}|n_\emptyset = m)$ by realizing the values of $v_\emptyset$, ζ_1, n_1, and v_1:

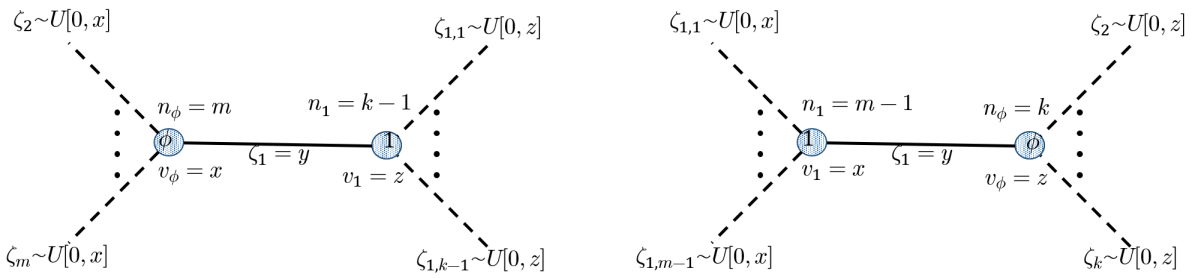
$$
\begin{aligned}
&\mathbb{E}(f(G,\emptyset,1)\mathbb{1}_{1\sim\emptyset}|n_\emptyset = m)\\
&\quad= \sum_{k=1}^{\infty}\widehat{P}(k-1)\int_{x=0}^{\infty}\frac{\lambda \mathrm{e}^{-\lambda x}(\lambda x)^m}{m!}\int_{y=0}^{x}\frac{1}{x}\int_{z=y}^{\infty}\frac{\lambda \mathrm{e}^{-\lambda z}(\lambda z)^{k-1}}{(k-1)!}\times\\
&\qquad\qquad \mathbb{E}(f(G,\emptyset,1)|n_\emptyset = m, v_\emptyset = x, \zeta_1 = y, n_1 = k-1, v_1 = z)\,dz\,dy\,dx\\
&\quad= \sum_{k=1}^{\infty}P(k)\int_{x=0}^{\infty}\int_{z=0}^{\infty}\int_{y=0}^{\min(x,z)}\frac{\lambda^3 \mathrm{e}^{-\lambda(x+z)}(\lambda x)^{m-1}(\lambda z)^{k-1}}{m!(k-1)!}\times\\
&\qquad\qquad \mathbb{E}(f(G,\emptyset,1)|n_\emptyset = m, v_\emptyset = x, \zeta_1 = y, n_1 = k-1, v_1 = z)\,dy\,dz\,dx,
\end{aligned}
$$

where the last equality is obtained by changing the order of the integration and replacing $\widehat{P}(k-1)$ by $P(k)$. Putting it all together, we have

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{v\sim\emptyset} f(G,\emptyset,v)\right) =\\
&\quad\sum_{m=1}^{\infty}\sum_{k=1}^{\infty}\frac{P(m)P(k)\lambda^{m+k}}{(m-1)!(k-1)!}\int_{x=0}^{\infty}\int_{z=0}^{\infty}\int_{y=0}^{\min(x,z)}\lambda \mathrm{e}^{-\lambda(x+z)}x^{m-1}z^{k-1}\times\\
&\qquad\mathbb{E}(f(G,\emptyset,1)|n_\emptyset = m, v_\emptyset = x, \zeta_1 = y, n_1 = k-1, v_1 = z)\,dy\,dz\,dx. \qquad (3.6)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{v\sim\emptyset} f(G,v,\emptyset)\right) =\\
&\quad\sum_{m=1}^{\infty}\sum_{k=1}^{\infty}\frac{P(m)P(k)\lambda^{m+k}}{(m-1)!(k-1)!}\int_{x=0}^{\infty}\int_{z=0}^{\infty}\int_{y=0}^{\min(x,z)}\lambda \mathrm{e}^{-\lambda(x+z)}x^{m-1}z^{k-1}\times\\
&\qquad\mathbb{E}(f(G,1,\emptyset)|n_\emptyset = m, v_\emptyset = x, \zeta_1 = y, n_1 = k-1, v_1 = z)\,dy\,dz\,dx. \qquad (3.7)
\end{aligned}
$$



(a) Structure of $N_{\circ\circ}(\emptyset, 1)$ (b) Structure of $N_{\circ\circ}(1, \emptyset)$

Figure 3.1: Structure of doubly rooted graphs $(G, 1, \emptyset)$ and $(G, \emptyset, 1)$ conditioned on a realization of $n_\emptyset$, $v_\emptyset$, ζ_1, n_1 and v_1 such that $\zeta_1 < v_1$, where $(G, \emptyset)$ is distributed as $Er(P, \lambda)$

In order to complete the proof, the following observation is crucial. Let $(G, \emptyset)$ be a realization of $Er(P, \lambda)$; conditioned on $n_\emptyset = m$, $v_\emptyset = x$, $\zeta_1 = y$, $n_1 = k - 1$ and $v_1 = z$ such that $\min(x, z) > y$, the structure and distribution of the doubly rooted graph $(G, \emptyset, 1)$ is the same as the structure and distribution of the doubly rooted graph $(G, 1, \emptyset)$ conditioned on $n_\emptyset = k$, $v_\emptyset = z$, $\zeta_1 = y$, $n_1 = m - 1$ and $v_1 = x$. This symmetry property is evident from Figure 3.1. Based on the above discussion, we have

$$\begin{aligned}&\mathbb{E}(f(G, \emptyset, 1)|n_\emptyset = m, v_\emptyset = x, \zeta_1 = y, n_1 = k - 1, v_1 = z)\\ &\quad = \mathbb{E}(f(G, 1, \emptyset)|n_\emptyset = k, v_\emptyset = z, \zeta_1 = y, n_1 = m - 1, v_1 = x).\end{aligned}$$

which implies (3.6) and (3.7) are equal. This completes the proof. □

3.4.2 Degree Distribution

Next, we characterize the degree distribution of EWT. The conditional degree distribution of a vertex conditioned on its type and the degree distribution of the root vertex is given as follows.

Theorem III.2. *Let $D_{\boldsymbol{i}}$ denote the degree of the vertex $\boldsymbol{i} \in T \sim Er(P)$. The conditional distribution of $D_{\boldsymbol{i}}$, conditioned on the type of the vertex $\boldsymbol{i}$ is given as follows:*

$$\mathbb{P}\left(D_{\boldsymbol{i}} = d|n_{\boldsymbol{i}} = m, v_{\boldsymbol{i}} = x\right) = Bi\left(d; m, \int_0^x \frac{1}{x}\sum_{k=1}^{\infty} P(k)\bar{F}_k(y)\, dy\right),$$

where $\bar{F}_k(\cdot)$ is the complementary cumulative distribution function of $Erlang(k)$ and $Bi(d; m, \eta) = C(m, d)\eta^d(1-\eta)^{m-d}$, with $C(m, d) = m!/(d!(m-d)!)$. Consequently, the degree distri-

bution of the root vertex and its mean are given as follows:

$$\mathbb{P}(D_{\emptyset} = d) = \sum_{m=1}^{\infty} P(m) \int_0^{\infty} \frac{\mathrm{e}^{-x} x^m}{m!} Bi\left(d; m, \int_0^x \frac{1}{x} \sum_{k=1}^{\infty} P(k) \bar{F}_k(y)\, dy\right) dx,$$

$$\mathbb{E}[D_{\emptyset}] = \sum_{m=1}^{\infty} \sum_{k=1}^{\infty} P(m) P(k) \int_0^{\infty} \bar{F}_k(y) \bar{F}_m(y)\, dy.$$

Proof. The proof is presented in Appendix B.2. □

It is easy to derive in closed-form the degree distribution of the root vertex. However, the degree distribution of a vertex at depth l is rather complex. To see why, let us focus on the vertices at the first generation, i.e., the neighbors of the root vertex $\emptyset$. For a unimodular measure ρ with support on rooted trees, the following equality holds,

$$\mathbb{E}_{\rho}\left[\sum_{v \sim \emptyset} \mathbf{1}_{deg(\emptyset)=k}\right] = \mathbb{E}_{\rho}\left[\sum_{v \sim \emptyset} \mathbf{1}_{deg(v)=k}\right]. \tag{3.8}$$

The above relation is obtained by using the following function in the definition of the unimodularity,

$$f_k([N_{\circ\circ}(\emptyset, v)]) = \begin{cases} 1 & \text{if } deg(\emptyset) = k \text{ and } v \sim \emptyset \\ 0 & \text{otherwise} \end{cases}.$$

It is easy to check that the function f_k is a Borel function from G_{**} to $\mathbb{R}$. Let D_1 and $D_{\emptyset}$ denote the degree of a vertex at the first generation and the degree of the root vertex, respectively. Simplifying (3.8), we have

$$k\mathbb{P}(D_{\emptyset} = k) = \mathbb{E}_{\rho}\left[D_{\emptyset} \mathbb{P}(D_1 = k | D_{\emptyset})\right]. \tag{3.9}$$

Now if D_1 and $D_{\emptyset}$ are independent, then D_1 has the size-biased distribution corresponding to $D_{\emptyset}$. This is the case for the unimodular Galton-Watson Tree [25]. However, D_1 and $D_{\emptyset}$ are not independent in our setting. Another interesting observation is that the degree distribution of different generations are not the same since the probability of the events $n_{\boldsymbol{i}} = m$ and $v_{\boldsymbol{i}} = x$ depends on the depth of the vertex $\boldsymbol{i}$. Owing to this interdependence structure, at present we do not have a characterization of the degree distribution at any level other than the root. We also believe that D_l, the degree distribution of a typical vertex at generation l, converges in distribution to the size-biased distribution of $D_{\emptyset}$ as $l \to \infty$ but establishing this is also an open problem. In Section 3.5, we investigate these claims via

numerical simulation.

3.4.3 Probability of Extinction

The next natural quantity to study is the probability that the component containing the root is finite, i.e., the probability of extinction. This is an important quantity associated with the EWT which should be related to the size of the giant component in the finite graph model, as in the unimodular Galton-Watson tree. Let us start with the definition of the probability of extinction.

Definition III.3. Let Z_l denote the number of vertices at depth l. The probability of extinction is defined as:

$$\mathbb{P}(\{\text{extinction}\}) \coloneqq \mathbb{P}\left(\bigcup_{l=1}^{\infty}\{Z_l = 0\}\right).$$

Observe that the event $\{Z_i = 0\}$ is a subset of the event $\{Z_j = 0\}$ for every $j < i$; hence, the continuity of probability measures implies that

$$\mathbb{P}(\{\text{extinction}\}) = \lim_{l\to\infty} \mathbb{P}(\{Z_l = 0\}).$$

Using this, we can characterize the probability of extinction.

Theorem III.4. *Consider the operator* $T : L(\mathbb{R}_+; [0,1]) \to C^1(\mathbb{R}_+; [0,1])$ *defined as*

$$T(f)(x) \coloneqq \begin{cases} \dfrac{1}{x}\displaystyle\sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} \left(\int_{z=0}^{y} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz + \int_{z=y}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f(z)^{k-1} dz \right) dy, & x > 0 \\ \displaystyle\sum_{k=1}^{\infty} P(k) \int_{z=0}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f(z)^{k-1}\, dz, & x = 0 \end{cases} \tag{3.10}$$

with the convention $0^0 = 1$. *Then the probability of extinction is*

$$\mathbb{P}(\{\textit{extinction}\}) = \sum_{m=1}^{\infty} P(m) \int_{x=0}^{\infty} \frac{\mathrm{e}^{-x} x^m}{m!} (q(x))^m\, dx,$$

where the function $q(\cdot)$ *is the smallest fixed point of the operator* T, *i.e., for any other fixed point of* $T(\cdot)$ *say* $f(\cdot) \in C^1(\mathbb{R}_+; [0,1])$, $f(x) \geq q(x)$ *for all* $x \in \mathbb{R}_+$. *Equivalently, the function* $q(\cdot)$ *is the point of convergence of* $T^l(\mathbf{0})(\cdot)$ *as* l *goes to infinity, where* $\mathbf{0}(\cdot)$ *is the null function, i.e.,* $\mathbf{0}(x) \equiv 0\ \forall x$.

Sketch of the proof. The main idea is to find the probability of the event $\{Z_l = 0\}$ and then, let l increases to infinity. This can be done through the following steps.

1. Observe that conditioned on the type of the root vertex to be (m, x), there are m potential branches and the probability that the depth of each branch is less than or equal to $l-1$ depends only on the value of x.

2. Starting from the first generation, all the vertices have the same behavior, i.e., for any non-root vertex $\boldsymbol{i}$, the distribution of $n_{\boldsymbol{i}}$ is given by $\widehat{P}$. Hence, it is possible to write the probability that the depth of a branch is less than or equal to $l-1$ via a recursion.

3. Taking the limit and using monotonicity, the result follows.

Proof. We now fill in the details. The theorem claims that the range of T is $C^1(\mathbb{R}_+; [0,1])$ and that there exists a fixed point $q(\cdot)$ such that for any other fixed point $f(\cdot)$ of T,

$$q(x) = T(q)(x) \leq T(f)(x) = f(x) \qquad \forall x \in \mathbb{R}_+,$$

i.e., it is the smallest fixed point of the operator T. The theorem also claims that

$$q(\cdot) = \lim_{l \to \infty} T^l(\mathbf{0}).$$

Let us start with these important properties of the operator T.

Lemma III.5. *Let $\mathbf{1}(\cdot)$ be the constant function with value 1 everywhere. The following hold:*

(i) For every $f(\cdot) \in L(\mathbb{R}_+; [0,1])$, the function $T(f)(\cdot)$ is non-decreasing and it belongs to $C^1(\mathbb{R}_+; [0,1])$. Moreover, $T(f)(\cdot) \equiv 1$ if and only if $f(x) = 1$ for almost every $x \in \mathbb{R}_+$.

(ii) The largest fixed point of the operator T is the constant function $\mathbf{1}(\cdot)$. Moreover, if $f(\cdot) \neq \mathbf{1}(\cdot)$ is a fixed point of T, then $f(\cdot)$ is strictly increasing.

(iii) For every pair of functions $f_1(\cdot), f_2(\cdot) \in C^1(\mathbb{R}_+, [0,1])$ with the property that $f_1(x) < f_2(x)$ $\forall x \in \mathbb{R}_+$, we have

$$T(f_1)(x) < T(f_2)(x) \qquad \forall x \in \mathbb{R}_+.$$

(iv) The function $T^l(\mathbf{0})$ converges point-wise to some function $q(\cdot) \in C[0,1]$ as l goes to infinity, which is the smallest fixed point of the operator T.

Proof of Lemma III.5. The proof is algebraic and does not use the connection between the operator $T(\cdot)$ and the probability of extinction.

(i) As the first step, we want to show the range of $T(f)(\cdot)$ is $[0,1]$. The inequality $T(f)(x) \geq 0$ is trivial. For the other side of the inequality, note that $f(x) \leq 1$ for all $x \in \mathbb{R}_+$; hence,

$$\begin{aligned} T(f)(x) &\leq \frac{1}{x}\sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} \left(\int_{z=0}^{y} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz + \int_{z=y}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz \right) dy \qquad (3.11) \\ &= \frac{1}{x}\sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} dy = 1. \end{aligned}$$

The equality holds if and only if $f(x) = 1$ for almost every $x \in \mathbb{R}_+$. To see $T(f)(\cdot)$ is non-decreasing, we show that it has a continuous non-negative derivative. Let $x > 0$. We then have

$$\begin{aligned} \frac{\mathrm{d}T(f)(x)}{\mathrm{d}x} &= -\frac{1}{x^2}\sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} \left(\int_{z=0}^{y} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz + \int_{z=y}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f(z)^{k-1}\, dz \right) dy \\ &\quad + \frac{1}{x}\sum_{k=1}^{\infty} P(k) \left(\int_{z=0}^{x} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz + \int_{z=x}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f(z)^{k-1}\, dz \right) \\ &= -\frac{1}{x^2}\sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} \left(\int_{z=0}^{x} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz + \int_{z=x}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f(z)^{k-1}\, dz \right) dy \\ &\quad - \frac{1}{x^2}\sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} \left(-\int_{z=y}^{x} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz + \int_{z=y}^{x} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f(z)^{k-1}\, dz \right) dy \\ &\quad + \frac{1}{x}\sum_{k=1}^{\infty} P(k) \left(\int_{z=0}^{x} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz + \int_{z=x}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f(z)^{k-1}\, dz \right) \\ &= \frac{1}{x^2}\sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} \int_{z=y}^{x} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} (1 - f(z)^{k-1})\, dz\, dy \\ &= \frac{1}{x^2}\sum_{k=1}^{\infty} P(k) \int_{z=0}^{x} \frac{\mathrm{e}^{-z} z^{k}}{(k-1)!} (1 - f(z)^{k-1})\, dz \geq 0. \end{aligned}$$

Observe that the derivative exists and is continuous for all $x > 0$. Taking the limit $x \downarrow 0$, we have

$$\lim_{x \downarrow 0} \frac{\mathrm{d}T(f)(x)}{\mathrm{d}x} = \lim_{x \downarrow 0} \frac{1}{2x}\sum_{k=1}^{\infty} P(k) \frac{\mathrm{e}^{-x} x^{k}}{(k-1)!} (1 - f(x)^{k-1}) = 0.$$

Since $f(0) \coloneqq \lim_{x \to 0} f(x)$, the have right-hand derivative of $T(f)(\cdot)$ at $x = 0$ is zero.

Hence, $T(f) \in C^1(\mathbb{R}_+; [0,1])$ is non-decreasing which completes the proof of (i).

(ii) It is easy to see that $\mathbf{1}(\cdot)$ is the largest fixed point of T. Moreover, for any other fixed point of $T(\cdot)$ say $f(\cdot) \in C^1(\mathbb{R}_+; [0,1])$, from (3.11) the function $T(f)(\cdot)$ is strictly less than 1; hence, $f(x) < 1 \ \forall x \in \mathbb{R}_+$. Using the proof of part (i), it is easy to see that the derivative of $T(f)$ is strictly positive; hence, the fixed point $f(\cdot)$ is strictly increasing.

(iii) The proof is straightforward.

(iv) Using part (iii), since $\mathbf{0}(x) < T(\mathbf{0})(x) < 1$ for all $x > 0$,

$$0 \leq T^l(\mathbf{0})(x) < T^{l+1}(\mathbf{0})(x) < 1 \qquad \forall x \in \mathbb{R}_+, l \in \mathbb{N}.$$

Let $f_l(x) = T^l(\mathbf{0})(x)$. Since, for every fixed value of x, the sequence $\{f_l(x)\}_{l=0}^{\infty}$ is strictly increasing, it converges. Define $q(x) = \lim_{l\to\infty} f_l(x) \ \ \forall x \in \mathbb{R}_+$. We then have

$$\begin{aligned} q(x) &= \lim_{l\to\infty} \frac{1}{x} \sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} \left(\int_{z=0}^{y} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz + \int_{z=y}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f_l(z)^{k-1}\, dz \right) dy \\ &= \frac{1}{x} \sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} \left(\int_{z=0}^{y} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz + \int_{z=y}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \lim_{l\to\infty} f_l(z)^{k-1}\, dz \right) dy \\ &= T(q)(x). \end{aligned}$$

The second equality follows from monotone convergence theorem, which allows interchanging the order of the summation, the integration, and the limit.

To show that $q(\cdot)$ is the smallest fixed point of T, consider any other fixed pint of T, $\widetilde{q} = T(\widetilde{q})$. Since $\mathbf{0}(x) < \widetilde{q}(x)$ for all $x \in \mathbb{R}_+$, the inequality $\widetilde{q}(x) = T(\widetilde{q})(x) > f_l(x)$ holds for all values of $l \in \mathbb{N}$ and $x \in \mathbb{R}_+$; hence, passing to the limit as $l \to \infty$, we get $q(x) \leq \widetilde{q}(x)$.

□

We now get back to the proof of the main theorem. As we mentioned, the main idea is to characterize the probability of the event $\{Z_l = 0\}$. Define $Z_{l,i}$ to be the number of children at depth l in the i^{th} subtree connected to the root. Fix an $l > 1$. $Z_l = 0$ if for all $i \in [n_\emptyset]$ either (i) $v_i < \zeta_i$, i.e., the i^{th} edge does not form, or (ii) the i^{th} edge forms but there are no children at its l^{th} level, i.e., $Z_{l,i} = 0$. Recall that for $i \in [n_\emptyset]$, ζ_i is the cost of the potential edge $\{\emptyset, i\}$. Hence, for $l \geq 2$ we have

$$\mathbb{P}\left(\{Z_l = 0\} | n_\emptyset = m, v_\emptyset = x\right)$$

$$
\begin{aligned}
&= \prod_{i=1}^{m} \mathbb{P}\Big(\big\{ v_i < \zeta_i \big\} \cup \big\{ \{ v_i \geq \zeta_i \} \cap \{ Z_{l,i} = 0 \} \big\} | n_\emptyset = m, v_\emptyset = x \Big) \\
&= \Big(\mathbb{P}\big(\{ v_1 < \zeta_1 \} | n_\emptyset = m, v_\emptyset = x \big) + \mathbb{P}\big(\{ v_1 \geq \zeta_1 \} \cap \{ Z_{l,1} = 0 \} | n_\emptyset = m, v_\emptyset = x \big) \Big)^m \\
&= \Bigg(\sum_{k=1}^{\infty} \widehat{P}(k-1) \int_{y=0}^{x} \frac{1}{x} \int_{z=0}^{y} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \, dz \, dy + \\
&\qquad \sum_{k=1}^{\infty} \widehat{P}(k-1) \int_{y=0}^{x} \frac{1}{x} \int_{z=y}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \mathbb{P}\left(\{ Z_{l,1} = 0 \} | n_1 = k-1, v_1 = z \right) \, dz \, dy \Bigg)^m .
\end{aligned} \tag{3.12}
$$

Conditioning on the type of the vertex 1, the probability distribution of $Z_{l,1}$ for $l > 1$ is exactly the same as the probability distribution of Z_{l-1} conditioned on the corresponding type of the root vertex; in particular,

$$
\mathbb{P}\left(\{ Z_{l,1} = 0 \} | n_1 = k-1, v_1 = z \right) = \mathbb{P}\left(\{ Z_{l-1} = 0 \} | n_\emptyset = k-1, v_\emptyset = z \right). \tag{3.13}
$$

A crucial observation is that $\mathbb{P}\left(\{ Z_l = 0 \} | n_\emptyset = m, v_\emptyset = x \right)$ depends on m only through the exponent. Define the function $f_l(\cdot)$ without the m^{th}-power as follows,

$$
\begin{aligned}
f_l(x) \coloneqq &\sum_{k=1}^{\infty} \widehat{P}(k-1) \int_{y=0}^{x} \frac{1}{x} \int_{z=0}^{y} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \, dz \, dy + \\
&\sum_{k=1}^{\infty} \widehat{P}(k-1) \int_{y=0}^{x} \frac{1}{x} \int_{z=y}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \mathbb{P}\left(\{ Z_{l,1} = 0 \} | n_1 = k-1, v_1 = z \right) \, dz \, dy.
\end{aligned}
$$

The function $f_l(\cdot)$ does not depend on the value of m and further,

$$
\mathbb{P}\left(\{ Z_l = 0 \} | n_\emptyset = m, v_\emptyset = x \right) = f_l(x)^m .
$$

Using (3.13) and the definition of the function $f_l(\cdot)$, for every $l > 0$, we have

$$
\begin{aligned}
f_l(x) &= \sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} \frac{1}{x} \left(\int_{z=0}^{y} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \, dz + \int_{z=y}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f_{l-1}(z)^{k-1} \, dz \right) dy \\
&= T(f_{l-1})(x),
\end{aligned}
$$

where $f_1(\cdot)$ should be taken to be $T(\mathbf{0})(\cdot)$ for consistency with (3.12) at $l = 2$. Lemma III.5 implies that $f_l(\cdot) = T^l(\mathbf{0})(\cdot)$ converges to $q(\cdot)$, the smallest fixed point of T, point-wise.

Hence,

$$\begin{aligned}\mathbb{P}\left(\{\text{extinction}\}|n_\emptyset = m, v_\emptyset = x\right) &= \lim_{l\to\infty} \mathbb{P}\left(\{Z_l = 0\}|n_\emptyset = m, v_\emptyset = x\right) \\ &= \lim_{l\to\infty} \left(T^l(\mathbf{0})(x)\right)^m \\ &= q(x)^m.\end{aligned}$$

Taking expectation with respect to $n_\emptyset$ and $v_\emptyset$ and using monotone convergence theorem, we have

$$\mathbb{P}(\{\text{extinction}\}) = \sum_{m=1}^{\infty} P(m) \int_{x=0}^{\infty} \frac{\mathrm{e}^{-x} x^m}{m!} (q(x))^m \, dx.$$

□

The above theorem suggests that for all $f(\cdot) \in L(\mathbb{R}_+; [0,1])$, the function $T^l(f)(\cdot)$ converges point-wise to a fixed point of T, as l goes to infinity; however, it is not clear how many fixed points the operator T has and, if there is more than one fixed point, to which one does $T^l(f)(\cdot)$ converge. An immediate corollary is the following.

Corollary III.6. $\mathbb{P}(\{\textit{extinction}\}) = 1$ *if and only if* $\mathbf{1}(\cdot)$ *is the unique fixed point of the operator* T*.*

A sufficient condition to check $\mathbb{P}(\{\text{extinction}\}) < 1$ is to find a test function $f(\cdot) \in L(\mathbb{R}_+; [0,1])$ such that $T(f)(x) \leq f(x)$ for all $x \in \mathbb{R}_+$ and $f(\cdot) \neq \mathbf{1}(\cdot)$. One natural choice is

$$f_{x_0,\epsilon}(x) \coloneqq \begin{cases} 1-\epsilon, & \text{if } x \leq x_0 \\ 1, & \text{otherwise.} \end{cases}$$

Choosing $\epsilon > 0$ to be small enough, we get the following corollary.

Corollary III.7. *Assume that there is an* $x_0 > 0$ *such that for all* $x \in [0, x_0]$*,*

$$\int_{z=0}^{x_0} z \frac{\min(x,z)}{x} g_2(z) \, dz > 1,$$

where $g_2(z) = \sum_{k=2}^{\infty} P(k) \frac{\mathrm{e}^{-z} z^{k-2}}{(k-2)!}$*, and* $\min(x,z)/x$ *for* $x = 0$ *is interpreted as* 1*. Then* $\mathbb{P}(\textit{extinction}) < 1$*.*

Proof. Note that for $\epsilon \in (0,1)$ we have $(1-\epsilon)^{k-1} \leq (1+(k-1)\epsilon)^{-1}$. Using this, for all

$x \leq x_0$ we have

$$f_{x_0,\epsilon}(x) - T(f_{x_0,\epsilon})(x) = \frac{1}{x}\sum_{k=1}^{\infty} P(k) \int_{z=0}^{x_0} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \min(x,z)(1-(1-\epsilon)^{k-1})\, dz - \epsilon$$
$$\geq \epsilon \left(\sum_{k=1}^{\infty} \frac{(k-1)}{1+(k-1)\epsilon} \times P(k) \int_{z=0}^{x_0} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \frac{\min(x,z)}{x}\, dz - 1 \right)$$

We want to show that the given condition in the Corollary implies that $f_{x_0,\epsilon}(x) - T(f_{x_0,\epsilon})(x) \geq 0$ for all x. It is sufficient to prove that the the right hand side of the above inequality is non-negative for all $x \leq x_0$ when ϵ is small enough. Equivalently, we want to show the following inequality holds

$$\lim_{\epsilon \downarrow 0} \sum_{k=1}^{\infty} \frac{(k-1)}{1+(k-1)\epsilon} \times P(k) \int_{z=0}^{x_0} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \frac{\min(x,z)}{x}\, dz > 1$$

Using the monotone convergence theorem, the result follows by changing the order of summation and the limit. □

The assumption of the corollary is not tight, i.e., there are examples where $\mathbb{P}(\text{extinction}) < 1$, but the assumption of the above corollary fails. Two natural follow-up questions are: 1) Is there a general test function $f(\cdot)$ such that $\mathbb{P}(\{\text{extinction}\}) < 1$ if and only if $f \geq T(f)$? 2) If the answer is yes, what is the closed-form of f?

The idea of using test functions, as simple as it seems, combined with point process perspective turns out to be a powerful tool for analyzing the branching process. We revisit this idea in Section 3.4.8.

3.4.4 Expected Number of Vertices at Depth l

Let Z_l and W_l denote the number of vertices and the number of potential vertices, respectively, at depth l. The expected value of Z_l and W_l are related to the growth rate of the EWT. These are also closely related to the probability of extinction via the following:

$$\mathbb{E}[Z_l] < \text{Const for all } l \text{ if and only if } \mathbb{P}(\{\text{extinction}\}) = 1. \tag{3.14}$$

The proof of (3.14) is based on a classical property of branching processes that Z_n goes to either 0 or ∞. We will revisit this property later on. For now, we state the following.

Theorem III.8. *We have*

$$\mathbb{E}[W_l] = \mathbb{E}[n_\emptyset] \times (\mathbb{E}[(n_\emptyset - 1)])^{l-1}$$

$$\mathbb{E}[Z_l] = \sum_{m=1}^{\infty} \lambda P(m) \sum_{k_1=2}^{\infty} \lambda P(k_1) \cdots \sum_{k_{l-1}=2}^{\infty} \lambda P(k_{l-1}) \sum_{k_l=1}^{\infty} P(k_l)$$
$$\int_{y_l=0}^{\infty} \int_{y_{l-1}=0}^{\infty} \cdots \int_{y_1=0}^{\infty} \bar{F}_m(y_1) \bar{F}_{k_1-1}(\max(y_1, y_2)) \ldots$$
$$\bar{F}_{k_{l-1}-1}(\max(y_{l-1}, y_l)) \bar{F}_{k_l}(y_l)\, dy_1\, dy_2 \ldots\, dy_l.$$

where, as before, $\bar{F}_k(\cdot)$ is the complementary cumulative distribution function of the $Erlang(k)$ distribution.

Proof. The proof is presented in Appendix B.3. □

A necessary but not a sufficient condition for $\mathbb{P}(\{\text{extinction}\})$ to be non-zero, is stated in the following corollary.

Corollary III.9. *If the expected number of the potential neighbors of the root vertex, i.e., $\mathbb{E}[n_\emptyset]$, is smaller than 2, then the population will eventually go extinct. See Section 3.4.8.*

Proof. If $\mathbb{E}[n_\emptyset] < 2$, then $\mathbb{E}[Z_l] \leq \mathbb{E}[W_l] = \mathbb{E}[n_\emptyset] \times (\mathbb{E}[(n_\emptyset - 1)])^{l-1} \xrightarrow{l\to\infty} 0$. □

Theorem III.8 does not provide an easy way to check whether $\mathbb{E}[Z_l]$ goes to zero or not. There is no recursive representation for the quantities provided by the theorem either; however, using the point process perspective leads to a full characterization of the growth rate and provides a necessary and sufficient condition for the probability of extinction to be less than 1.

3.4.5 Krein-Rutman Eigenvalue and the Corresponding Eigenfunctions

To obtain the growth rate of EWT more work needs to be done. We follow the discussion of Chapter 3 of Harris [30]. Harris analyzes general branching processes from a point process perspective. Although we use the same idea, our assumptions are different and the results from Harris's book [30] do not apply to our setting and require a generalization.

Abusing notation, let $Z_l(k-1, A)$ denote the number of vertices at depth l of type $(k-1, z)$ where $k \in \mathbb{N}$ and $z \in A$ with $A \subset \mathbb{R}_+$ being a Borel set. Let $M_l(m, x; k-1, A)$ denote the expected value of $Z_l(k-1, A)$, conditioned on $n_\emptyset = m$ and $v_\emptyset = x$, i.e.,

$$M_l(m, x; k-1, A) \coloneqq \mathbb{E}[Z_l(k-1, A) | n_\emptyset = m, v_\emptyset = x].$$

Let $m_l(m, x; k-1, z)$ denote the density of $M_l(m, x; k-1, A)$ at (m, x):

$$M_l(m, x; k-1, A) = \int_{z\in A} m_l(m, x; k-1, z)\, dz.$$

We show that $\beta^{-l} M_l(m, x; \mathbb{Z}_+, \mathbb{R}_+)$ converges to some fixed function independent of l, for a suitable β. Moreover, we show that $\beta^{-l} m_l(m, x; k-1, z)$ converges to $\mu(m, x)\nu(k-1, z)$. The quantity β is the largest eigenvalue of M_1, and the functions $\mu(\cdot\,, \cdot)$ and $\nu(\cdot\,, \cdot)$ are the unique right and the left eigenfunctions corresponding to the eigenvalue β, respectively.

Definition III.10. Let m_1 denote the density of M_1. If there exists a non-zero function $\mu(\cdot\,, \cdot)$ and a $\beta \in \mathbb{R}$ such that

$$\beta\mu(m, x) = \int_{z=0}^{\infty} \sum_{k=1}^{\infty} m_1(m, x; k-1, z)\mu(k-1, z)\, dz, \tag{3.15}$$

then $\mu(\cdot\,, \cdot)$ is called the right eigenfunction of M_1 corresponding to the eigenvalue β. Similarly, the left eigenfunction corresponding to the eigenvalue β is defined as follows,

$$\beta\nu(k-1, z) = \int_{x=0}^{\infty} \sum_{m=0}^{\infty} m_1(m, x; k-1, z)\nu(m, x)\, dx. \tag{3.16}$$

The main goal of this section is to prove a generalization of the Perron–Frobenius theorem. We show that a version of Krein-Rutman Theorem by Toland [47] applies to our setting. However, it does not provide an easy way to find the spectral radius. The specific structure of the EWT makes it possible to directly prove the convergence of $\beta^{-l} m_l(m, x; k-1, z)$ to $\mu(m, x)\nu(k-1, z)$ and to show that $\beta^{-l} M_l(m, x; \mathbb{R}_+, \mathbb{Z}_+)$ converges to some function that only depends on x and m. Before presenting the main theorems and their proofs, let us simplify the operator of interest,

$$\begin{aligned} M_1(m, x; k-1, A) &= m \int_{y=0}^{x} \frac{1}{x} \int_{z \geq y, z \in A} \widehat{P}(k-1) f_k(z)\, dz\, dy \\ &= \frac{m}{x} \int_{z \in A} \min(x, z) P(k) \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz, \end{aligned}$$

where $f_k(\cdot)$ is the probability density function of $Erlang(k)$. Hence,

$$m_1(m, x; k-1, z) = \frac{m}{x} \min(x, z) P(k) \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}. \tag{3.17}$$

Let β be an arbitrary eigenvalue of M_1. By (3.15) the right eigenfunction of β then satisfies the following equation:

$$\beta\mu(m, x) = \int_{z=0}^{\infty} \sum_{k=1}^{\infty} \frac{m}{x} \min(x, z) P(k) \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \mu(k-1, z)\, dz.$$

Dividing both sides by m, the right-hand side is independent of m (note that $\mu(0,x)=0$); hence, $\mu(m,x)$ is linear in m and we can write

$$x\mu(m,x)/m =: \widetilde{\mu}(x), \tag{3.18}$$

where $\widetilde{\mu}(\cdot)$ is a solution to the following equation

$$\beta\widetilde{\mu}(x) = \int_{z=0}^{\infty} g_2(z)\min(x,z)\widetilde{\mu}(z)\,dz, \tag{3.19}$$

and $g_2(x) = \mathrm{e}^{-x}\sum_{k=2}^{\infty} P(k)\frac{x^{k-2}}{(k-2)!}$. Note that if $\widetilde{\mu}(\cdot)$ satisfies the above relation, then a right eigenfunction of M_1 corresponding to the eigenvalue β is given by $\mu(x,m) \coloneqq m\widetilde{\mu}(x)/x$. Similarly, for the left eigenfunction, we have

$$\begin{aligned}
\beta\nu(k-1,z) &= \int_{x=0}^{\infty}\sum_{m=0}^{\infty} m_1(m,x;k-1,z)\nu(m,x)\,dx \\
&= \int_{x=0}^{\infty}\sum_{m=0}^{\infty}\frac{m}{x}\min(x,z)P(k)\frac{\mathrm{e}^{-z}z^{k-1}}{(k-1)!}\nu(m,x)\,dx \\
&= P(k)\frac{\mathrm{e}^{-z}z^{k-1}}{(k-1)!}\int_{x=0}^{\infty}\sum_{m=0}^{\infty}\frac{m}{x}\min(x,z)\nu(m,x)\,dx \\
&= P(k)\frac{\mathrm{e}^{-z}z^{k-1}}{(k-1)!}\int_{y=0}^{z}\int_{x=y}^{\infty}\sum_{m=0}^{\infty}\frac{m}{x}\nu(m,x)\,dx\,dy.
\end{aligned}$$

Note that the dependence of $\nu(k-1,z)$ in k is through the term $P(k)\frac{\mathrm{e}^{-z}z^{k-1}}{(k-1)!}$. Hence, we can write

$$\nu(k-1,z) = \widetilde{\nu}(z)P(k)\frac{\mathrm{e}^{-z}z^{k-1}}{(k-1)!}, \tag{3.20}$$

for a suitable $\widetilde{\nu}(\cdot)$ that is a solution to the following equation,

$$\begin{aligned}
\beta\widetilde{\nu}(z) &= \int_{y=0}^{z}\int_{x=y}^{\infty}\sum_{m=0}^{\infty}\frac{m}{x}P(m+1)\frac{\mathrm{e}^{-x}x^{m}}{m!}\widetilde{\nu}(x)\,dx\,dy \\
&= \int_{y=0}^{z}\int_{x=y}^{\infty}\sum_{m=2}^{\infty}P(m)\frac{\mathrm{e}^{-x}x^{m-2}}{(m-2)!}\widetilde{\nu}(x)\,dx\,dy \\
&= \int_{y=0}^{z}\int_{x=y}^{\infty} g_2(x)\widetilde{\nu}(x)\,dx\,dy \\
&= \int_{x=0}^{\infty}\min(x,z)g_2(x)\widetilde{\nu}(x)\,dx.
\end{aligned}$$

Observe that $\widetilde{\nu}(\cdot)$ satisfies the same equation as $\widetilde{\mu}(\cdot)$ does. To study this equation, we define a new operator and rely on the background materials discussed in Section 3.2.3. Since H_1 is a compact self-adjoint operator, classical results from operator theory says H_1 indeed

Let $\mathcal{H} = L^2(\mathbb{R}_+, \upsilon)$ denote the set of real-valued square integrable functions with respect to a measure υ. It is easy to prove that $L^2(\mathbb{R}_+, \upsilon)$ together with the inner product $\langle f, g\rangle = \int_0^\infty f(x)g(x)d\upsilon(x)$ is a real Hilbert space. Let $H_1 \in \mathcal{L}(\mathcal{H}, \mathcal{H})$ be an integral operator with integrand $\min(\cdot\,,\cdot) \in L^2(\mathbb{R}_+ \times \mathbb{R}_+, \upsilon \times \upsilon)$, i.e.,

$$H_1 f(x) = \int_0^\infty \min(x, y) f(y) d\upsilon(y),$$

where $\upsilon(\cdot)$ is a finite measure with Radon-Nikodym derivative $g_2(\cdot)$ with respect to Lebesgue measure. The integral operator H_1 is self-adjoint since its integrand is symmetric. Moreover, H_1 is compact since $\mathcal{H}$ is separable (the proof follows by the fact that $\mathcal{H}$ has a countable orthonormal basis). Putting these together, we see that H_1 is a compact self-adjoint operator.

Let $\mathcal{K}$ denote the set of all non-negative functions in $\mathcal{H}$. The set $\mathcal{K}$ is closed and convex. Moreover, $\lambda\mathcal{K} \subset \mathcal{K}$ and $\mathcal{K} \cap (-\mathcal{K}) = \{\mathbf{0}\}$; hence, $\mathcal{K}$ is a cone. Actually, it is a total cone, i.e., $\mathcal{H} = \mathcal{K} - \mathcal{K}$. The following theorem is a direct implication of Theorem III.7-III.9.

Theorem III.11. *The largest eigenvalue of H_1 in magnitude is,*

$$\mathscr{X}(H_1) = \max_{\substack{f(\cdot)\in\mathcal{H}, \|f\|_{\mathcal{H}}=1,\\ f(\cdot) \text{ is non-negative}}} \int_0^\infty \int_0^\infty \min(x, y) f(x) f(y) d\upsilon(y) d\upsilon(x).$$

$\mathscr{X}(H_1) > 0$ is a simple eigenvalue and corresponds to a non-negative eigenfunction. Moreover, all the eigenvalues of H_1 are real and if $\zeta(\cdot)$ is an eigenfunction of H_1 with some eigenvalue $\mu \neq \mathscr{X}(H_1)$, we have

$$\int_0^\infty f(y)\zeta(y)d\nu(y) = 0.$$

The following simplification will help in finding the Perron-Frobenius eigenvalue of H_1 and the corresponding eigenfunction. Changing the order of integration, the operator H_1 can be written as follows,

$$H_1 f(x) = \int_{y=0}^{x} \int_{z=y}^{\infty} f(z) d\upsilon(z)\, dy. \tag{3.21}$$

Define the operator $\widetilde{H_1}$ as follows,

$$\widetilde{H_1} f(x) = \int_{y=x}^{\infty} \int_{z=y}^{\infty} f(z) d\upsilon(z)\, dy.$$

Using (3.21), we have

$$\begin{aligned} H_1 f(x) + \widetilde{H_1} f(x) &= \int_0^{\infty} \int_y^{\infty} f(z) d\upsilon(z)\, dy \\ &= \int_0^{\infty} z f(z) d\upsilon(z) = \langle f, \mathbb{I} \rangle_{\mathcal{H}}, \end{aligned}$$

where $\mathbb{I}(\cdot)$ is the identity function, i.e., $\mathbb{I}(x) = x$ for all $x \in \mathbb{R}_+$. The Perron-Frobenius eigenvalue of H_1 and the corresponding eigenfunction are related to the operator $\widetilde{H_1}$.

Theorem III.12. *Consider the function $L(\beta, x)$ for $x \in \mathbb{R}_+$ and $\beta \in \mathbb{C}$ defined as follows,*

$$L(\beta, x) \coloneqq \sum_{i=0}^{\infty} G_i(x) \left(\frac{-1}{\beta} \right)^i,$$

where the function $G_i(x)$ is defined recursively via

$$\begin{aligned} &G_0(x) \coloneqq 1, \\ &G_i(x) \coloneqq \int_{y=x}^{\infty} \int_{z=y}^{\infty} g_2(z) G_{i-1}(z)\, dz\, dy = \widetilde{H}_1 G_{i-1}(x) \qquad \forall i > 0, \end{aligned}$$

where

$$g_2(x) \coloneqq \mathrm{e}^{-x} \sum_{k=2}^{\infty} P(k) \frac{x^{k-2}}{(k-2)!}.$$

Assuming the moment generating function of $n_{\text{ø}}$ exists for some $\theta > 0$, the function $L(\beta, x)$ satisfies the following properties,

(i) The function $L(\beta, x)$ is well-defined for all $\beta \in \mathbb{C}$ and $x \in \mathbb{R}_+$, i.e., the series converges in the absolute sense.

(ii) The second partial derivative of $L(\beta, x)$ with respect to x, satisfies the following equality,

$$\beta \frac{\partial^2 L(\beta, x)}{\partial x^2} = -g_2(x) L(\beta, x).$$

(iii) For every fixed $x \in \mathbb{R}_+$, all the zeros of the function $L(\beta, x)$ are real-valued.

(iv) *There exists a real value* $\beta_0 \in \Big(\max_x \left(x\upsilon([x,\infty))\right), \mathbb{E}[n_\emptyset] - 1\Big)$ *such that for every real* $\beta > \beta_0$*, the function* $L(\beta, x)$ *is uniformly positive, i.e.,* $\exists \varepsilon_\beta > 0$ *such that* $L(\beta, x) > \varepsilon_\beta$, $\forall x \in \mathbb{R}_+$*. Moreover, the function* $L(\beta_0, x)$ *is non-negative for all* $x \in \mathbb{R}_+$ *and* $L(\beta_0, 0) = 0$*. Finally, for all* $\beta \geq \beta_0$*,*

$$\left.\frac{\partial L(\beta, x)}{\partial x}\right|_{x=x_0} > 0 \qquad \forall x_0 \in \mathbb{R}_+.$$

(v) *The function* $\frac{x}{L(\beta_0,x)}$ *is well-defined, is strictly positive for all* $x \in \mathbb{R}_+$*, and is strictly increasing.*

Proof. In the course of the proof, it will become apparent that $L(\beta, x)$ and the Bessel function of the first kind of zeroth order share similar properties.

(i) Using the Chernoff bound,

$$\mathbb{P}(n_\emptyset \geq k) \leq \frac{\mathbb{E}[\mathrm{e}^{\theta n_\emptyset}]}{\mathrm{e}^{\theta k}} < \infty.$$

We then have

$$\begin{aligned} g_i(z) &\coloneqq \sum_{k=i}^{\infty} P(k) \frac{\mathrm{e}^{-z} z^{k-i}}{(k-i)!} \\ &\leq \sum_{k=i}^{\infty} \mathbb{P}(n_\emptyset \geq k) \frac{\mathrm{e}^{-z} z^{k-i}}{(k-i)!} \\ &\leq \sum_{k=i}^{\infty} \frac{\mathbb{E}[\mathrm{e}^{\theta n_\emptyset}]}{\mathrm{e}^{\theta k}} \frac{\mathrm{e}^{-z} z^{k-i}}{(k-i)!} \\ &= \frac{\mathbb{E}[\mathrm{e}^{\theta n_\emptyset}]}{\mathrm{e}^{\theta i}} \sum_{k=i}^{\infty} \frac{\mathrm{e}^{-z} z^{k-i}}{\mathrm{e}^{\theta(k-i)}(k-i)!} \\ &= \frac{\mathbb{E}[\mathrm{e}^{\theta n_\emptyset}]}{\mathrm{e}^{\theta i}} \exp\big(-z(1-\mathrm{e}^{-\theta})\big). \end{aligned} \tag{3.22}$$

Let $C = \mathbb{E}[\mathrm{e}^{\theta n_\emptyset}]/\mathrm{e}^{2\theta}$ and $\Upsilon = 1 - \mathrm{e}^{-\theta}$. It is easy to prove that $G_i(x)$ is upper bounded by $C^i \,\mathrm{e}^{-i\Upsilon x}/(\Upsilon^{2i} i! i!)$, by induction. Indeed,

$$\begin{aligned} G_0(x) &= 1 \\ G_{i+1}(x) &= \int_{y=x}^{\infty} \int_{z=y}^{\infty} G_i(z) g_2(z)\, dz\, dy \\ &\leq \int_{y=x}^{\infty} \int_{z=y}^{\infty} \frac{C^i}{\Upsilon^{2i}} \frac{\mathrm{e}^{-i\Upsilon z}}{i!i!} \times C\mathrm{e}^{-z\Upsilon}\, dz\, dy \end{aligned}$$

$$= \frac{C^{i+1}}{\Upsilon^{2(i+1)}} \frac{\mathrm{e}^{-(i+1)\Upsilon x}}{(i+1)!(i+1)!},$$

which proves the upper bound by induction. Collectively we then have

$$\sum_{i=0}^{\infty} G_i(x) \left(\frac{1}{|\beta|}\right)^i \leq \sum_{i=0}^{\infty} \frac{C^i}{\Upsilon^{2i}} \frac{\mathrm{e}^{-i\Upsilon x}}{i!i!} \left(\frac{1}{|\beta|}\right)^i = J_0\left(\sqrt{\frac{-4C\mathrm{e}^{-\Upsilon x}}{\Upsilon^2|\beta|}}\right)$$
$$= I_0\left(\sqrt{\frac{4C\mathrm{e}^{-\Upsilon x}}{\Upsilon^2|\beta|}}\right) \in (0, \infty),$$

where $J_0(\cdot)$ is the Bessel function of the first kind of order 0 and $I_0(\cdot)$ is the modified Bessel function of the first kind of order 0. This establishes that the series converges absolutely.

(ii) Using the definition of $L(\beta, x)$ and part (i),

$$\begin{aligned}\beta \frac{\partial^2 L(\beta, x)}{\partial x^2} &= \beta \sum_{i=0}^{\infty} \frac{\mathrm{d}^2 G_i(x)}{\mathrm{d}x^2} \left(\frac{-1}{\beta}\right)^i \\ &= \sum_{i=1}^{\infty} -G_{i-1}(x) g_2(x) \left(\frac{-1}{\beta}\right)^{i-1} \\ &= -g_2(x) L(\beta, x).\end{aligned}$$

(iii) Fix some $x \in \mathbb{R}_+$. Consider the function $H_1(\beta, x)$ defined as follows,

$$H_1(\beta, x) \coloneqq L(\beta, x) \frac{\partial L(\bar{\beta}, x)}{\partial x} - L(\bar{\beta}, x) \frac{\partial L(\beta, x)}{\partial x},$$

where $\bar{\beta}$ is the complex conjugate of β. The partial derivative of $H_1(\beta, x)$ with respect to x, using the part (ii), is given as follows,

$$\begin{aligned}\frac{\partial H_1(\beta, x)}{\partial x} &= \frac{\partial L(\beta, x)}{\partial x} \frac{\partial L(\bar{\beta}, x)}{\partial x} - \bar{\beta}^{-1} L(\beta, x) L(\bar{\beta}, x) g_2(x) \\ &\quad - \frac{\partial L(\bar{\beta}, x)}{\partial x} \frac{\partial L(\beta, x)}{\partial x} + \beta^{-1} L(\bar{\beta}, x) L(\beta, x) g_2(x) \\ &= (\beta^{-1} - \bar{\beta}^{-1}) |L(\beta, x)|^2 g_2(x),\end{aligned}$$

where the last equality is obtained by the fact that $\overline{L(\beta, x)} = L(\bar{\beta}, x)$. Note that,

$$\lim_{x \to \infty} L(\beta, x) = 1 \text{ and } \lim_{x \to \infty} \frac{\partial L(\beta, x)}{\partial x} = 0,$$

since $\lim_{x\to\infty} G_i(x) = \lim_{x\to\infty} \frac{\mathrm{d}G_i(x)}{\mathrm{d}x} = 0$ for all $i > 0$ and $L(\beta, x)$ is absolutely summable. Hence, $\lim_{x\to\infty} H_1(\beta, x) = 0$. Therefore,

$$\int_x^\infty (\beta^{-1} - \bar{\beta}^{-1})|L(\beta, y)|^2 g_2(y)\, dy = -H_1(\beta, x).$$

Since for every fixed $x \in \mathbb{R}_+$ the coefficients of $L(\beta, x)$ are real-valued, $L(\beta, x) = 0$ implies $L(\bar{\beta}, x) = 0$. Moreover, if $L(\beta, x) = 0$ for some $x \in \mathbb{R}_+$ and $\beta \in \mathbb{C}$, then $H_1(\beta, x) = 0$; hence,

$$\int_x^\infty (\beta^{-1} - \bar{\beta}^{-1})|L(\beta, y)|^2 g_2(y)\, dy = 0,$$

from which we conclude that $\beta = \bar{\beta}$, i.e., $\beta \in \mathbb{R}$.

(iv) Pick any real-valued $\beta \geq \mathbb{E}[n_\emptyset] - 1$. For all $i \geq 1$, we have,

$$\begin{aligned} &G_i(x)\left(\frac{1}{\beta}\right)^i - G_{i+1}(x)\left(\frac{1}{\beta}\right)^{i+1} \\ &\quad = \int_{y=x}^\infty \int_{z=y}^\infty g_2(z)\frac{1}{\beta}\left(G_{i-1}(z)\left(\frac{1}{\beta}\right)^{i-1} - G_i(z)\left(\frac{1}{\beta}\right)^i\right) dz\, dy, \end{aligned}$$

and for $i = 0$,

$$1 - G_1(x)\frac{1}{\beta} = 1 - \int_{y=x}^\infty \int_{z=y}^\infty g_2(z)\frac{1}{\beta}\, dz\, dy.$$

For each $i \in \mathbb{N}$, the function $G_i(x)$ is decreasing; hence, the function $1 - \beta^{-1}G_1(x)$ is increasing and it achieves its minimum at $x = 0$, so

$$\begin{aligned} 1 - G_1(0)\frac{1}{\beta} &= 1 - \int_{y=0}^\infty \int_{z=y}^\infty g_2(z)\frac{1}{\beta}\, dz\, dy \\ &= 1 - \int_{z=0}^\infty z g_2(z)\frac{1}{\beta}\, dz \\ &= 1 - \int_{z=0}^\infty z \sum_{k=2}^\infty P(k)\frac{\mathrm{e}^{-z}z^{k-2}}{(k-2)!}\frac{1}{\beta}\, dz \\ &= 1 - \frac{1}{\beta}\sum_{k=2}^\infty (k-1)P(k)\int_{z=0}^\infty \frac{\mathrm{e}^{-z}z^{k-1}}{(k-1)!}\, dz \\ &= 1 - \frac{\mathbb{E}[n_\emptyset] - 1}{\beta} \geq 0. \end{aligned}$$

By induction,

$$G_i(x)\left(\frac{1}{\beta}\right)^i - G_{i+1}(x)\left(\frac{1}{\beta}\right)^{i+1} > 0 \qquad \forall x \geq 0,\ i > 0.$$

Hence, for every real value $\beta \geq \mathbb{E}[n_\emptyset] - 1$, by rewriting $L(\beta, x)$, we get

$$L(\beta, x) = \sum_{i=0}^{\infty}\left(G_{2i}(x)\left(\frac{1}{\beta}\right)^{2i} - G_{2i+1}(x)\left(\frac{1}{\beta}\right)^{2i+1}\right) > 0 \qquad \forall x \in \mathbb{R}_+.$$

Moreover, if for a fixed real value $\beta > 0$ and for all $x \in \mathbb{R}_+$ the function $L(\beta, x)$ is non-negative, then the function $L(\beta, x)$ is strictly increasing:

$$\begin{aligned}
\frac{\partial L(\beta, x)}{\partial x} &= \sum_{i=1}^{\infty}\frac{\mathrm{d}G_i(x)}{\mathrm{d}x}\left(\frac{-1}{\beta}\right)^i \\
&= \sum_{i=1}^{\infty}\int_{y=x}^{\infty} -g_2(y)G_{i-1}(y)\,dy\left(\frac{-1}{\beta}\right)^i \\
&= \frac{1}{\beta}\int_{y=x}^{\infty} g_2(y)\sum_{i=1}^{\infty}G_{i-1}(y)\left(\frac{-1}{\beta}\right)^{i-1}dy \\
&= \frac{1}{\beta}\int_{y=x}^{\infty} g_2(y)L(\beta, y)\,dy > 0. \qquad (3.23)
\end{aligned}$$

Next, we prove that for some $\beta \in \mathbb{R}_+$ and $x \in \mathbb{R}_+$, the function $L(\beta, x)$ is negative. Let us rewrite the function $L(\beta, x)$,

$$\begin{aligned}
L(\beta, x) &= 1 + \left(\frac{-1}{\beta}\right)\sum_{i=1}^{\infty}G_i(x)\left(\frac{-1}{\beta}\right)^{i-1} \\
&= 1 - \frac{1}{\beta}\int_{y=x}^{\infty}\int_{z=y}^{\infty}L(\beta, z)g_2(z)\,dz\,dy \qquad (3.24)
\end{aligned}$$

where the last equality is based on the recursive relation between $G_i(x)$ and $G_{i-1}(x)$. Using the above equality we have,

$$\begin{aligned}
L(\beta, x) - L(\beta, 0) &= \frac{1}{\beta}\left(\int_{y=0}^{\infty}\int_{z=y}^{\infty}L(\beta, z)g_2(z)\,dz\,dy - \int_{y=x}^{\infty}\int_{z=y}^{\infty}L(\beta, z)g_2(z)\,dz\,dy\right) \\
&= \frac{1}{\beta}\int_{y=0}^{x}\int_{z=y}^{\infty}L(\beta, z)g_2(z)\,dz\,dy \\
&= \frac{1}{\beta}\int_{0}^{\infty}\min(x, z)L(\beta, z)g_2(z)\,dz
\end{aligned}$$

$$= \frac{1}{\beta}\int_0^x zL(\beta,z)g_2(z)\,dz + \frac{1}{\beta}\int_x^\infty xL(\beta,z)g_2(z)\,dz, \tag{3.25}$$

where the third equality follows by changing the order of integration.

Suppose that for all $\beta \in \mathbb{R}_+$ and all $x \in \mathbb{R}_+$, the function $L(\beta,x)$ is non-negative. Hence, for any fixed $\beta \in \mathbb{R}_+$, the function $L(\beta,x)$ is strictly increasing and,

$$-L(\beta,0) - \frac{1}{\beta}\int_0^x zL(\beta,z)g_2(z)\,dz \geq L(\beta,x)\left(\frac{x}{\beta}\int_{y=x}^\infty g_2(y)\,dy - 1\right) \qquad \forall x \in \mathbb{R}_+.$$

However, the left-hand side of the above equation is negative for all $\beta \in \mathbb{R}_+$ but the right-hand side, for small enough β, is positive which is a contradiction. The above argument shows that if there exist some $\widehat{x} > 0$ such that $\beta \leq \widehat{x}\upsilon([\widehat{x},\infty])$, then the function $L(\beta,\cdot)$ takes negative values. Moreover, for every $\beta \geq \mathbb{E}[n_\emptyset] - 1$ the function $L(\beta,x)$ is strictly positive. Combining these together and considering the fact that $L(\beta,x)$ is a continuous function of $x \in \mathbb{R}_+$ and $\beta \in \mathbb{R}_+$, we conclude that there exists a largest $\beta_0 > 0$ such that the function $L(\beta_0,x)$ is non-negative, and $L(\beta_0,x_0) = 0$ for some $x_0 \in \mathbb{R}_+$. The already established strictly increasing property of $L(\beta_0,x)$ implies that $x_0 = 0$, and the proof is complete.

(v) Using 3.25, $L(\beta_0,x) > 0$ for all $x > 0$. Moreover, using the L'Hospital rule,

$$\lim_{z\to 0}\frac{z}{L(\beta_0,z)} = \frac{1}{\frac{\partial L(\beta_0,x)}{\partial x}\big|_{x=0}},$$

which is well-defined since $\frac{\partial L(\beta_0,x)}{\partial x}\big|_{x=0}$ is strictly positive. Next, taking the derivative of $\frac{x}{L(\beta_0,x)}$, we get

$$\frac{\partial\left(x/L(\beta_0,x)\right)}{\partial x} = \frac{L(\beta_0,x) - x\frac{\partial L(\beta_0,x)}{\partial x}}{(L(\beta_0,x))^2}, \tag{3.26}$$

Note that $L(\beta_0,0) = 0$ and $L(\beta_0,x)$ is a strictly concave function due to part (ii) and (iv); therefore,

$$L(\beta_0,0) < L(\beta_0,x) + \frac{\partial L(\beta_0,x)}{\partial x} \qquad \forall x > 0.$$

Hence the expression (3.26) is strictly positive for every $x > 0$, and we have established that the function $\frac{x}{L(\beta_0,x)}$ is strictly increasing.

□

The following immediate corollary guarantees the existence of an eigenfunction $f(\cdot)$ and an eigenvalue β of the operator H_1.

Corollary III.13. *Let β_0 be the largest zero of $L(\cdot\,,0)$. Let $f_0(x) = L(\beta_0, x)$, then the constant β_0 and the function $f_0(\cdot)$ satisfy the following fixed point equation,*

$$\beta_0 f_0(x) = \int_{y=0}^{x} \int_{z=y}^{\infty} g_2(z) f_0(z)\, dz\, dy. \tag{3.27}$$

Proof. Substituting the function $L(\beta_0, x)$ in the above equation, we get,

$$\begin{aligned}
&\int_{y=0}^{x} \int_{z=y}^{\infty} g_2(z) L(\beta_0, z)\, dz\, dy \\
&\quad = \sum_{i=0}^{\infty} \int_{y=0}^{x} \int_{z=y}^{\infty} g_2(z) G_i(z)\, dz\, dy \left(\frac{-1}{\beta_0}\right)^i \\
&\quad = \sum_{i=0}^{\infty} \left(G_{i+1}(0) - G_{i+1}(x)\right) \left(\frac{-1}{\beta_0}\right)^i \\
&\quad = -\beta_0 (L(\beta_0, 0) - L(\beta_0, x)) = \beta_0 L(\beta_0, x),
\end{aligned}$$

where the last equality follows from part (iv) of Theorem III.12, since $L(\beta_0, 0) = 0$. □

Using the Corollary III.13 and the equations (3.18) and (3.20), a left and a right eigenfunction of M_1 for the eigenvalue β_0 are obtained.

Observe that, from (3.17), $m_l(m, x; k-1, z)$ satisfies the following recursive equation:

$$\begin{aligned}
m_l(m, x; k-1, z) &= \int_{z'=0}^{\infty} \sum_{k'=1}^{\infty} m_{l-1}(m, x; k'-1, z') m_1(k'-1, z'; k-1, z)\, dz' \\
&= \int_{z'=0}^{\infty} \sum_{k'=2}^{\infty} m_{l-1}(m, x; k'-1, z') \frac{k'-1}{z'} \min(z', z) P(k) \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\, dz' \\
&= P(k) \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \int_{z'=0}^{\infty} \sum_{k'=2}^{\infty} m_{l-1}(m, x; k'-1, z') \frac{k'-1}{z'} \min(z', z)\, dz'.
\end{aligned}$$

The terms related to the values of k and m can be factored out. However, to avoid dividing by zero, we consider the function $h_l(\cdot\,,\cdot)$ defined recursively as follows:

$$\begin{aligned}
h_l(x, z) &= \int_{z'=0}^{\infty} h_{l-1}(x, z') g_2(z') h_1(z', z)\, dz' \quad l \geq 2, \\
h_1(x, z) &= \min(x, z).
\end{aligned} \tag{3.28}$$

It is easy to see that the function m_l is related to the function h_l via the following equation;

indeed, the relation holds between m_1 and h_1, which is just (3.17), and for a general l the proof holds via induction:

$$m_l(m, x; k-1, z) = h_l(x, z) \times \frac{P(k)\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \frac{m}{x}. \tag{3.29}$$

Recall that the kernel of the operator H_1 is symmetric, hence, any right eigenfunction is also a left eigenfunction. Moreover, Corollary III.13 implies that $f_0(\cdot)$ is an eigenfunction of H_1 with eigenvalue β_0, i.e.,

$$\beta_0 f_0(x) = \int_{z=0}^{\infty} \min(x, y) f_0(y) d\upsilon(y). \tag{3.30}$$

Hence, the question of whether or not β_0 is the Krein-Rutman eigenvalue of M_1 with right eigenfunctions $\mu(\cdot\,,\cdot)$ and left eigenfunction $\nu(\cdot\,,\cdot)$, boils down to the same question for H_1 with right and left eigenfunctions f_0.

To show that β_0 is the Perron-Frobenius eigenvalue of H_1, we define a continuous state Markov chain and prove uniform geometric ergodicity for the chain. Consider a continuous state Markov chain, with the following transition probability kernel:

$$p(x, y) \coloneqq \frac{h_1(x, y) g_2(y) f_0(y)}{\beta_0 f_0(x)} \qquad \forall x, y \in \mathbb{R}_+, \tag{3.31}$$

where the transition probability at $x = 0$ is defined by taking the limit of $p(x, \cdot)$ as x goes to 0, i.e.,

$$p(0, y) \coloneqq \lim_{x \to 0} \frac{h_1(x, y) g_2(y) f_0(y)}{\beta_0 f_0(x)} = \frac{g_2(y) f_0(y)}{\beta_0 f_0{}'(0)}.$$

By Theorem III.12 part (iv), the term $f_0{}'(0)$ is strictly positive; hence, the function $p(\cdot\,,\cdot)$ is well-defined. Moreover, the function $p(\cdot\,,\cdot)$ is indeed a valid transition probability kernel since

$$\begin{aligned}
\int_{z=0}^{\infty} p(x, z)\, dz &= \int_{z=0}^{\infty} \frac{\min(x, z) g_2(z) f_0(z)}{\beta_0 f_0(x)}\, dz = \int_{y=0}^{x} \int_{z=y}^{\infty} \frac{g_2(z) f_0(z)}{\beta_0 f_0(x)}\, dz = 1 \\
\int_{z=0}^{\infty} p(0, z)\, dz &= \int_{z=0}^{\infty} \frac{g_2(z) f_0(z)}{\beta_0 f_0{}'(0)}\, dz \overset{(*)}{=} \frac{1}{\beta_0 f_0'(0)} \left. \frac{\partial \left(\beta_0 f_0(x)\right)}{\partial x} \right|_{x=0} = 1,
\end{aligned}$$

where $(*)$ follows from (3.23). By induction, it is easy to observe from (3.28) that the l step

transition probability kernel is related to the function $h_l(\cdot\,,\cdot)$ via the following equation,

$$p^{(l)}(x,y) = \int_{z=0}^{\infty} p^{(l-1)}(x,z)p(z,y)\,dz = \frac{h_l(x,y)g_2(y)f_0(y)}{{\beta_0}^l f_0(x)}. \tag{3.32}$$

The stationary density of the Markov chain can now be verified to be $\pi(y) = C_N g_2(y)(f_0(y))^2$, where C_N is the normalization factor. Indeed, from (3.30) and (3.31), we have

$$\int_{x=0}^{\infty} \pi(x)p(x,y)\,dx = C_N \int_{x=0}^{\infty} g_2(x)f_0(x)\frac{\min(x,y)g_2(y)f_0(y)}{\beta_0}dx = \pi(y).$$

Observe that the stationary distribution equals the product of the left and the right eigenfunctions of H_1 upto a normalization factor. Recall that $g_2(\cdot)$ is the Radon-Nikodym derivative of $\upsilon(\cdot)$. Moreover, the Markov chain is reversible with respect to the stationary distribution $\pi(\cdot)$, i.e., $\pi(x)p(x,y) = \pi(y)p(y,x)$.

It is natural to expect $p^{(l)}(x,y)$ converges point-wise to $\pi(y)$ as l goes to infinity. To prove this, we invoke the following result by Baxendale [48],

Theorem III.14 (Baxendale 2005)**.** *Let $\{X_n : n > 0\}$ be a time homogeneous Markov chain on a state space $(\mathcal{S}, \mathcal{S})$. Let $P(x, A)$, $x \in S$, $A \in \mathcal{S}$ denote the transition probability and by abusing notation let P denote the corresponding operator on measurable functions $\mathcal{S} \to \mathbb{R}$. Assume that the following assumptions hold:*

(A1) Minorization condition: There exists $C \in \mathcal{S}$, $\widetilde{\beta} > 0$ and a probability measure ν on $(\mathcal{S}, \mathcal{S})$ such that,

$$P(x, A) \geq \widetilde{\beta}\nu(A),$$

for all $x \in C$ and $A \in \mathcal{S}$.

(A2) Drift condition: There exist a measurable function $V : \mathcal{S} \to [1, \infty)$ and constants $\lambda < 1$ and $K < \infty$ satisfying,

$$PV(x) \leq \begin{cases} \lambda V(x), & \text{if } x \notin C \\ K, & \text{if } x \in C \end{cases}.$$

(A3) Strong aperiodicity condition: There exists $\widehat{\beta} > 0$ such that $\widetilde{\beta}\nu(C) \geq \widehat{\beta}$.

Then $\{X_n : n > 0\}$ has a unique stationary probability measure π, say, and $\int V\,d\pi < \infty$. Moreover, there exists $\rho < 1$ depending only (and explicitly) on $\widehat{\beta}$, $\widetilde{\beta}$, λ and K such that

whenever $\rho < \gamma < 1$ there exists $M < \infty$ depending only (and explicitly) on γ, $\widehat{\beta}$, $\widetilde{\beta}$, λ and K such that,

$$\sup_{|g| \le V} \left| (P^n g)(x) - \int g \, d\pi \right| \le M V(x) \gamma^n,$$

for all $x \in \mathcal{S}$ and $n \ge 0$, where the supremum is taken over all possible measurable functions $g : \mathcal{S} \to \mathbb{R}$ satisfying $|g(x)| \le V(x)$ for all $x \in \mathcal{S}$. In particular, $P^n g(x)$ and $\int g \, d\pi$ are both well-defined whenever

$$\|g\|_V \equiv \sup\{|g(x)|/V(x) : x \in \mathcal{S}\} < \infty.$$

Baxendale [48] provides explicit values for ρ and M and improves the constants if the corresponding Markov chain is reversible, which holds in our case. In the following lemma, we prove that the Markov chain with transition probability $p(x, y)$ from (3.31) satisfies the assumptions $(A1) - (A3)$.

Lemma III.15. *Assume the moment generating function of $n_\varnothing$ exists for some $\theta > 0$ and is finite. Then, the Markov chain defined by the transition probability kernel $p(x, y)$ on state space $(\mathbb{R}_+, \mathscr{B})$ satisfies the assumptions $(A1) - (A3)$ of Theorem (III.14) where the set C, the constants $\widetilde{\beta}$, λ, K, $\widehat{\beta}$, the function $V : \mathbb{R} \to [1, \infty)$ and the probability measure $\nu(x)$ are given as follows:*

$$\begin{gathered}
C \coloneqq [0, c], \quad \widetilde{\beta} \coloneqq \int_0^\infty W(y) \, dy, \quad \lambda \coloneqq \frac{1}{2}, \\
K \coloneqq f_0'(0) \frac{c}{f_0(c)} \frac{\mathbb{E}[\mathrm{e}^{\theta n_\varnothing}]}{\beta_0 \mathrm{e}^{2\theta}} \frac{1}{(1 - \mathrm{e}^{-\theta} - \eta)^2}, \quad \widehat{\beta} \coloneqq \min\left(\frac{1}{\beta_0 f'(0)}, \frac{c/2}{\beta_0 f_0(c)} \right) \int_{c/2}^{c} g_2(y) f_0(y) \, dy, \\
V(x) \coloneqq f_0'(0) \mathrm{e}^{\eta x} \frac{x}{f_0(x)}, \\
\nu(A) \coloneqq \frac{1}{\widetilde{\beta}} \int_A W(y) \, dy,
\end{gathered} \tag{3.33}$$

where the constants η and c, and the function $W(y)$ are defined as follows,

$$\begin{gathered}
\eta \coloneqq \frac{1 - \mathrm{e}^{-\theta}}{2}, \quad c \coloneqq \max\left(\frac{1}{\eta} \ln\left(\frac{\mathbb{E}[\mathrm{e}^{\theta n_\varnothing}]}{\beta_0 \mathrm{e}^{2\theta}} \frac{2}{(1 - \mathrm{e}^{-\theta} - \eta)^2} \right), 1 \right), \\
W(y) \coloneqq \begin{cases} \frac{1}{\beta_0 f_0'(0)} f_0(y) g_2(y) & \text{if } y \notin [0, c] \\ \min(\frac{1}{\beta_0 f_0'(0)}, \frac{y}{\beta_0 f_0(c)}) f_0(y) g_2(y) & \text{if } y \in [0, c] \end{cases}.
\end{gathered} \tag{3.34}$$

Proof. First, we prove that the assumption $(A2)$ holds and derive the constants c, λ and K

as well as the function $V(\cdot)$. Next, we show that the assumption $(A1)$ holds and derive the probability measure ν and the constant $\widetilde{\beta} > 0$. Finally, we illustrate that the assumption $(A3)$ holds and derive the constant $\widehat{\beta}$.

Assumption $(A2)$*:* Define the operator P by its action on non-negative measurable functions as follows:

$$\begin{aligned} PV(x) &\coloneqq \int_0^\infty V(y) \frac{\min(x,y) g_2(y) f_0(y)}{\beta_0 f_0(x)}\, dy \\ &\leq \frac{x}{\beta_0 f_0(x)} \int_0^\infty V(y) g_2(y) f_0(y)\, dy. \end{aligned}$$

Assuming the moment generating function of $n_\emptyset$ exists for some $\theta > 0$ and using inequality (3.22), we have

$$PV(x) \leq \frac{x}{\beta_0 f_0(x)} \int_0^\infty V(y) \frac{\mathbb{E}[\mathrm{e}^{\theta n_\phi}]}{\mathrm{e}^{2\theta}} \exp\big(-y(1-\mathrm{e}^{-\theta})\big) f_0(y)\, dy. \tag{3.35}$$

Let $V(x) = f_0{}'(0)\mathrm{e}^{\eta x} \frac{x}{f_0(x)}$ where the constant $\eta > 0$ is small enough such that $1 - \mathrm{e}^{-\theta} - \eta > 0$. Part (v) of Theorem III.12 states that the function $\frac{x}{f_0(x)}$ is strictly increasing. Hence, $V(\cdot)$ is a strictly increasing function and its range is $[1, \infty)$. Substituting the function $V(\cdot)$ into (3.35), we get

$$\begin{aligned} PV(x) &\leq f_0{}'(0) \frac{x}{f_0(x)} \frac{\mathbb{E}[\mathrm{e}^{\theta n_\phi}]}{\beta_0 \mathrm{e}^{2\theta}} \int_0^\infty y \exp\big(-y(1-\mathrm{e}^{-\theta}-\eta)\big)\, dy \\ &= f_0{}'(0) \frac{x}{f_0(x)} \frac{\mathbb{E}[\mathrm{e}^{\theta n_\phi}]}{\beta_0 \mathrm{e}^{2\theta}} \frac{1}{(1-\mathrm{e}^{-\theta}-\eta)^2}. \end{aligned} \tag{3.36}$$

Consider the constants c, K and λ as in the statement of the Theorem. For every $x \leq c$, the right-hand side of the (3.36) is bounded by K. Moreover, for every $x > c$, the following inequality holds

$$\frac{x}{f_0(x)} \frac{\mathbb{E}[\mathrm{e}^{\theta n_\phi}]}{\beta_0 \mathrm{e}^{2\theta}} \frac{1}{(1-\mathrm{e}^{-\theta}-\eta)^2} \leq \frac{x}{f_0(x)} \frac{1}{2} \mathrm{e}^{\eta c} \leq \frac{1}{2} \frac{x}{f_0(x)} \mathrm{e}^{\eta x}.$$

Hence, the assumption $(A2)$ is satisfied.

Assumption $(A1)$*:* Recall that $P(x, A)$ is defined as follows,

$$P(x, A) = \int_{y \in A} p(x,y)\, dy = \frac{1}{\beta_0 f_0(x)} \int_{y \in A} f_0(y) g_2(y) \min(x,y)\, dy.$$

Define the set $A_x = A \cap [0, x]$ and $A_{\bar{x}} = A \cap (x, \infty)$. Using $A = A_x \cup A_{\bar{x}}$, we have,

$$P(x, A) = \frac{1}{\beta_0 f_0(x)} \int_{y \in A_x} f_0(y) g_2(y) y \, dy + \frac{x}{\beta_0 f_0(x)} \int_{y \in A_{\bar{x}}} f_0(y) g_2(y) dy.$$

Consider the function $W(y) = \min_{x \in [0,c]} p(x, y)$. Using the fact that $\frac{x}{f_0(x)}$ and $f_0(x)$ are increasing functions, the function $W(\cdot)$ is given as in (3.34). Note that $W(\cdot)$ is integrable since it is upper bounded by the integrable function $\beta_0 {f_0}'(0)^{-1} g_2(y) f_0(y)$. Define the probability measure ν as follows,

$$\nu(A) = \frac{1}{\widetilde{\beta}} \int_{y \in A} W(y) \, dy,$$

where $\widetilde{\beta}$ is the normalization factor. The inequality $P(x, A) \geq \widetilde{\beta} \nu(A)$ for all $x \in [0, c]$ holds because of the following inequalities:

$$\begin{aligned}
\frac{1}{\beta_0 f_0(x)} \int_{y \in A_x} f_0(y) g_2(y) y \, dy &\geq \int_{y \in A_x} \min\left(\frac{1}{\beta_0 {f_0}'(0)}, \frac{y}{\beta_0 f_0(c)}\right) f_0(y) g_2(y) dy \\
&= \int_{y \in A_x} W(y) \, dy, \\
\frac{x}{\beta_0 f_0(x)} \int_{y \in A_{\bar{x}}} f_0(y) g_2(y) \, dy &\geq \int_{y \in A_{\bar{x}}} \frac{1}{\beta_0 {f_0}'(0)} f_0(y) g_2(y) \, dy \geq \int_{y \in A_{\bar{x}}} W(y) \, dy.
\end{aligned}$$

From here, the assumption $(A1)$ immediately follows.

Assumption $(A3)$*:* Using the definition of the probability measure ν, we have,

$$\widetilde{\beta} \nu([0, c]) = \int_0^c W(y) \, dy \geq \min\left(\frac{1}{\beta_0 {f_0}'(0)}, \frac{c/2}{\beta_0 f_0(c)}\right) \int_{c/2}^c g_2(y) f_0(y) \, dy = \widehat{\beta} > 0.$$

□

Remark III.16. The function $V(\cdot)$ in (3.33) provides us with more freedom, i.e., it is possible to choose a function $g : \mathbb{R}_+ \to \mathbb{R}$ that goes to infinity.

Lemma III.15 implies that the Theorem III.14 holds for the continuous state Markov chain with transition probability $p(x, y)$. The first implication is that the stationary distribution $\pi(x) = C_N g_2(x) f_0(x)^2$ is unique. Moreover, there exists $M < \infty$ and $0 < \gamma < 1$ such that all the measurable functions $g : \mathbb{R}_+ \to \mathbb{R}$ with the property that $|g(x)| \leq V(x)$ for all $x \in \mathbb{R}_+$,

satisfy

$$\left|(P^n g)(x) - \int g\, d\pi\right| \leq MV(x)\gamma^n.$$

Since $V(0) = 1$, and $V(x)$ is increasing as can be gleaned from (3.33) and Theorem III.12 part (v), geometric ergodicity follows by restricting the function $g(\cdot)$ to satisfy $|g(x)| \leq 1$, for all $x \geq 0$, that is

$$\|P^n(x,\cdot) - \pi\|_{TV} \leq MV(x)\gamma^n.$$

However, it is possible to prove uniform ergodicity by an appropriate choice of function $V(\cdot)$.

Lemma III.17. *Let* $V(x) = 1 + a \times \mathbb{1}_{x>x_0}$. *Let*

$$\lambda \coloneqq \frac{3}{4} \qquad K \coloneqq a + 1 \qquad c = x_0,$$

where the constant a *is defined as follows,*

$$a \coloneqq \frac{8}{\beta_0}\frac{\mathbb{E}[\mathrm{e}^{\theta n_\phi}]}{\mathrm{e}^{2\theta}} \times \frac{1}{(1-\mathrm{e}^{-\theta})^2},$$

and the constant x_0 *is large enough such that* $f_0(x_0) \geq 0.5$ *and the following inequality is satisfied for all* $x > x_0$*:*

$$\frac{2}{\beta_0}\frac{\mathbb{E}[\mathrm{e}^{\theta n_\phi}]}{\mathrm{e}^{2\theta}} \times \frac{(x+1)\mathrm{e}^{-x(1-\mathrm{e}^{-\theta})}}{(1-\mathrm{e}^{-\theta})^2} < \frac{1}{4}.$$

Then, for a suitable $\widetilde{M} > 0$ *and* $\widetilde{\gamma} < 1$, *we have*

$$\|P^n(x,\cdot) - \pi\|_{TV} \leq \widetilde{M}(a+1)\widetilde{\gamma}^n, \qquad \forall x \geq 0.$$

Proof. Again, we apply Baxendale's Theorem (Theorem III.14), but this time the $V(\cdot)$ is bounded. The only assumption affected by choice of the function $V(\cdot)$ is the assumption $(A2)$. Recall that the transition probability is given by,

$$p(x,y) = \frac{\min(x,y)g_2(y)f_0(y)}{\beta_0 f_0(x)},$$

Hence, the operator P performs on the measurable function $V(\cdot)$ as,

$$\begin{aligned} PV(x) &= \int_0^\infty V(y)\frac{\min(x,y)g_2(y)f_0(y)}{\beta_0 f_0(x)}\,dy \\ &= \frac{1}{\beta_0 f_0(x)}\int_0^x V(y)yg_2(y)f_0(y)\,dy + \frac{x}{\beta_0 f_0(x)}\int_x^\infty V(y)g_2(y)f_0(y)\,dy. \end{aligned}$$

Recall that the function $f_0(x)$ is an increasing function, $f_0(0) = 0$ and $\lim_{x\to\infty} f_0(x) = 1$, see for e.g., (3.24). Consider the function $V(x) = 1 + a \times \mathbb{1}_{\{x>x_0\}}$, where a and x_0 are constants to be specified later. Substituting the choice of function $V(\cdot)$, we get

$$\begin{aligned} PV(x) &= \begin{cases} \dfrac{1}{\beta_0 f_0(x)}\displaystyle\int_0^x yg_2(y)f_0(y)\,dy + \dfrac{x}{\beta_0 f_0(x)}\int_x^{x_0} g_2(y)f_0(y)dy \\ \qquad + \dfrac{(a+1)x}{\beta_0 f_0(x)}\displaystyle\int_{x_0}^\infty g_2(y)f_0(y)\,dy & \text{if } x \le x_0 \\ \dfrac{1}{\beta_0 f_0(x)}\displaystyle\int_0^{x_0} yg_2(y)f_0(y)\,dy + \dfrac{a+1}{\beta_0 f_0(x)}\int_{x_0}^x yg_2(y)f_0(y)\,dy \\ \qquad + \dfrac{(a+1)x}{\beta_0 f_0(x)}\displaystyle\int_x^\infty g_2(y)f_0(y)\,dy & \text{if } x > x_0 \end{cases} \\ &\le \begin{cases} \dfrac{a+1}{\beta_0 f_0(x)}\left(\dfrac{1}{a+1}\displaystyle\int_0^x yg_2(y)\,dy + x\int_x^{x_0} g_2(y)\,dy + x\int_{x_0}^\infty g_2(y)\,dy\right) & \text{if } x \le x_0 \\ \dfrac{a+1}{\beta_0 f_0(x)}\left(\dfrac{1}{a+1}\displaystyle\int_0^{x_0} yg_2(y)\,dy + \int_{x_0}^x yg_2(y)\,dy + x\int_x^\infty g_2(y)\,dy\right) & \text{if } x > x_0 \end{cases}. \end{aligned}$$

Assume $x > x_0$. Using the inequality (3.22), we have,

$$\begin{aligned} PV(x) &\le \frac{(a+1)\mathbb{E}[\mathrm{e}^{\theta n_\phi}]}{\mathrm{e}^{2\theta}\beta_0 f_0(x)}\left(\frac{1}{a+1}\int_0^{x_0} y\mathrm{e}^{-y(1-\mathrm{e}^{-\theta})}\,dy + \int_{x_0}^x y\mathrm{e}^{-y(1-\mathrm{e}^{-\theta})}\,dy + x\int_x^\infty \mathrm{e}^{-y(1-\mathrm{e}^{-\theta})}\,dy\right) \\ &\le \frac{a+1}{\beta_0 f_0(x)}\frac{\mathbb{E}[\mathrm{e}^{\theta n_\phi}]}{\mathrm{e}^{2\theta}}\left(\frac{1}{a+1}\frac{1}{(1-\mathrm{e}^{-\theta})^2} + \frac{(x_0(1-\mathrm{e}^{-\theta})+1)\mathrm{e}^{-x_0(1-\mathrm{e}^{-\theta})}}{(1-\mathrm{e}^{-\theta})^2} + x\frac{\mathrm{e}^{-x(1-\mathrm{e}^{-\theta})}}{1-\mathrm{e}^{-\theta}}\right). \end{aligned} \tag{3.37}$$

The last inequality follows from evaluating the integrals and removing the negative terms. The constants a and x_0 are chosen such that t $f_0(x_0) \ge 0.5$ and all the following inequalities are satisfied:

$$\frac{2}{\beta_0}\frac{\mathbb{E}[\mathrm{e}^{\theta n_\phi}]}{\mathrm{e}^{2\theta}} \times \frac{1}{a+1}\frac{1}{(1-\mathrm{e}^{-\theta})^2} < \frac{1}{4} \tag{3.38}$$

$$\frac{2}{\beta_0}\frac{\mathbb{E}[\mathrm{e}^{\theta n_\phi}]}{\mathrm{e}^{2\theta}} \times \frac{(z+1)\mathrm{e}^{-z(1-\mathrm{e}^{-\theta})}}{(1-\mathrm{e}^{-\theta})^2} < \frac{1}{4} \qquad \forall z > x_0. \tag{3.39}$$

Note that the left-hand side of (3.38) is decreasing in a and the left-hand side of (3.39) can be made arbitrary small by setting x_0 to be large enough. The second of these upper bounds the last two terms in (3.37), and we have $PV(x) \leq (3/4)(a+1)$ for $x > x_0$. For $x \leq x_0$, $PV(x) \leq 1 + a$. Given the above choice of constants a and x_0, for $\lambda = \frac{3}{4}$ and $K = a + 1$, taking $C = \{x : x \leq x_0\}$, the assumption $(A2)$ is satisfied; i.e.,

$$PV(x) \leq \begin{cases} \frac{3}{4}(a+1) & \text{if } x > x_0 \\ a+1 & \text{if } x \leq x_0. \end{cases}$$

An application of Baxendale's Theorem then completes the proof. □

An immediate consequence of uniform ergodicity is the following.

Corollary III.18. *For any $x, y \in \mathbb{R}_+$ and $l > 1$, we have*

$$\left|p^{(l)}(x,y) - \pi(y)\right| < 2\widetilde{M}(a+1)\widetilde{\gamma}^{l-1}. \tag{3.40}$$

Proof. The idea is same as in Doob [76, pages 216-217]. Note that $\pi(\cdot)$ is the unique stationary distribution. Hence, for any

$$\begin{aligned} \left|p^{(l)}(x,y) - \pi(y)\right| &= \left|\int_{z=0}^{\infty} p^{(1)}(z,y)\left(p^{(l-1)}(x,z) - \pi(z)\right)\, dz\right| \\ &\leq \left|\int_{p^{(l-1)}(x,z)>\pi(z)} \left(p^{(l-1)}(x,z) - \pi(z)\right)\, dz\right| \\ &\qquad + \left|\int_{p^{(l-1)}(x,z)<\pi(z)} \left(p^{(l-1)}(x,z) - \pi(z)\right)\, dz\right| \\ &\leq 2\widetilde{M}(a+1)\widetilde{\gamma}^{l-1} \end{aligned}$$

□

To get rid of the constant factor C_N, from now on, we assume the function f_0 is normalized such that,

$$\int_0^{\infty} g_2(y) f_0(y)^2\, dy = 1.$$

That is $f_0(y) = L(\beta_0, y)\sqrt{C_N}$, where $C_N = (\int_0^{\infty} g_2(y) L(\beta_0, y)^2\, dy)^{-1}$.

Then inequality (3.40) implies that for every $x \in \mathbb{R}_+$ and $y > 0$,

$$h_l(x,y) = {\beta_0}^l f_0(x) f_0(y) \left(1 + \frac{2\widetilde{M}(a+1)O(\widetilde{\gamma}^{l-1})}{g_2(y) {f_0(y)}^2}\right) \qquad l \geq 2. \tag{3.41}$$

Harris [77] assumes the density of the M_1 is uniformly positive and bounded, and deduces that the corresponding eigenfunction is uniformly positive as well. However, in our setting $f_0(0) = 0$ and $g(y) \to 0$ as $y \to \infty$. As a result, the error term for $h_l(x,y)/\beta_0^l$ explodes as y goes to 0 or ∞. On the other hand, induction using (3.28) implies $h_l(x,0) = h_l(0,y) = 0$. Hence, we should expect a uniform bound. The idea is to use the function $V(\cdot)$ in (3.33) and apply (3.28).

Lemma III.19. *For some constant $\widehat{M} > 0$, we have*

$$h_l(x,y) = {\beta_0}^l f_0(y) f_0(x) \left(1 + \widehat{M} O(\gamma^{l-2}) \frac{x}{{\beta_0}^2 f_0(y) f_0(x)}\right) \qquad l \geq 2. \tag{3.42}$$

Proof. Fix $z \in \mathbb{R}_+$ and define the function $g(\cdot)$ as follows,

$$g(x) = \begin{cases} \dfrac{h_1(x,z)}{f_0(x)} \times {f_0}'(0) & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}.$$

The function $g(\cdot)$ is a well-defined continuous function by Theorem III.12 part (v). Moreover, $|g(x)| \leq V(x)$ for all $x \in \mathbb{R}_+$ where $V(\cdot)$ is given by (3.33). Now using Lemma III.15 and Theorem III.14 (Baxendale's Theorem), we have

$$\left| \int_0^\infty \frac{h_l(x,y) g_2(y) f_0(y)}{{\beta_0}^l f_0(x)} \times \frac{h_1(y,z)}{f_0(y)} {f_0}'(0)\, dy - \int_0^\infty g_2(y) {f_0(y)}^2 \times \frac{h_1(y,z)}{f_0(y)} {f_0}'(0)\, dy \right| \leq M \gamma^l V(x).$$

Using (3.28) and (3.30), we get

$$\left| \frac{h_{l+1}(x,z)}{{\beta_0}^l f_0(x)} - \beta_0 f_0(z) \right| \leq M \gamma^l \frac{x \mathrm{e}^{\eta x}}{f_0(x)},$$

hence,

$$h_{l+1}(x,y) = {\beta_0}^{l+1} f_0(y) f_0(x) \left(1 + MO(\gamma^l) \frac{x \mathrm{e}^{\eta x}}{\beta_0 f_0(y) f_0(x)}\right).$$

Now using (3.28) again, we have

$$\begin{aligned} h_{l+2}(x,y) &= \int_0^\infty h_1(x,z)h_{l+1}(z,y)g_2(z)\,dz \\ &= \int_0^\infty \min(x,z){\beta_0}^{l+1}f_0(y)f_0(z)\left(1+MO(\gamma^l)\frac{z\mathrm{e}^{\eta z}}{\beta_0 f_0(y)f_0(z)}\right)dz \\ &= \int_0^\infty \min(x,z){\beta_0}^{l+1}f_0(y)f_0(z)g_2(z)\,dz + MO(\gamma^l)\int_0^\infty \min(x,z){\beta_0}^l z\mathrm{e}^{\eta z}g_2(z)\,dz. \end{aligned}$$

Applying inequality (3.22), we get

$$\left|h_{l+2}(x,y)-\beta_0^{l+2}f_0(x)f_0(y)\right| \le M\gamma^l \times \beta_0^l \frac{\mathbb{E}[\mathrm{e}^{\theta n_\emptyset}]}{\mathrm{e}^{2\theta}}\int_0^\infty \min(x,z)z\mathrm{e}^{\eta z}\exp\bigl(-z(1-\mathrm{e}^{-\theta})\bigr)\,dz.$$

Now the result follows by the fact that $\min(x,z)\le x$, and the fact that $\eta < 1-\mathrm{e}^{-\theta}$. Note that

$$\widehat{M} = M \times \frac{\mathbb{E}[\mathrm{e}^{\theta n_\emptyset}]}{\mathrm{e}^{2\theta}}\int_0^\infty z\mathrm{e}^{\eta z}\exp\bigl(-z(1-\mathrm{e}^{-\theta})\bigr)\,dz.$$

□

Remark III.20. In the proof of Lemma III.19, we bound $\min(x,z)$ by x instead of z. This gives us a uniform error bound for m_l. Specifically, as $x\to 0$ the error term in (3.43) stays bounded.

Combining (3.29) and (3.42), we get a similar bound for $m_l(m,x;k-1,z)$: for every $x\in\mathbb{R}_+$ and $z>0$,

$$m_l(m,x;k-1,z) = \frac{P(k)\mathrm{e}^{-z}z^{k-1}}{(k-1)!}\frac{m}{x}\times {\beta_0}^l f_0(x)f_0(z)\left(1+\widehat{M}O(\gamma^{l-2})\frac{x}{{\beta_0}^2 f_0(z)f_0(x)}\right)\ l\ge 2. \tag{3.43}$$

Note that the error term is uniformly bounded for all $x,z\in\mathbb{R}_+$ and $k\in\mathbb{N}$ (naturally, it is not uniform in m). Next we prove that β_0 is the Krein-Rutman eigenvalue of H_1 with the eigenfunction $f_0(x)$.

Theorem III.21. *Assume that the moment generating function of $n_\emptyset$ exists for some $\theta>0$. Then $\beta_0 \in \left(\max_x\left(x\upsilon([x,\infty))\right),\mathbb{E}[n_\emptyset]-1\right)$ is an eigenvalue of H_1 larger in magnitude than any other eigenvalue of H_1. The corresponding eigenfunction is $f_0(\cdot)$. Moreover, this is the only non-negative eigenfunction of H_1 up to a normalization factor.*

Proof. Assume there exists a real-valued function $\zeta(\cdot)$ and $\beta' \neq 0$ such that,

$$\beta'\zeta(x) = \int_{z=0}^{\infty} h_1(x,z)g_2(z)\zeta(z)\,dz.$$

Clearly, $\zeta(x)$ satisfies the following inequality,

$$\begin{aligned}|\zeta(x)| &\leq \frac{1}{|\beta'|}\int_{z=0}^{\infty} h_1(x,z)g_2(z)|\zeta(z)|\,dz \\ &\leq \frac{x}{|\beta'|}\int_{z=0}^{\infty} g_2(z)|\zeta(z)|\,dz = \texttt{Const} \times x.\end{aligned}$$

Moreover, $\zeta(0) = 0$ since $h_1(0,z) = 0$; hence the function $g(x) = \zeta(x)/f_0(x)$ for all $x \in \mathbb{R}_+$ is well-defined. Letting $V(x) = {f_0}'(0)\mathrm{e}^{\eta x}\frac{x}{f_0(x)} \times \max(\frac{\texttt{Const}}{{f_0}'(0)}, 1)$ in Lemma III.15, we have $|g(x)| \leq V(x)$ for all $x \in \mathbb{R}_+$. Using Baxendale's Theorem (Theorem III.14), we have

$$\left|\int_0^{\infty} \frac{h_l(x,y)g_2(y)f_0(y)}{{\beta_0}^l f_0(x)}\frac{\zeta(y)}{f_0(y)}\,dy - \int_0^{\infty} g_2(y).{f_0(y)}^2\frac{\zeta(y)}{f_0(y)}\,dy\right| < M\gamma^l V(x)$$

Hence,

$$\left|\frac{{\beta'}^l\zeta(x)}{{\beta_0}^l f_0(x)} - \int_0^{\infty} g_2(y)f_0(y)\zeta(y)\,dy\right| < M\gamma^l V(x).$$

As l goes to infinity, the right-hand side of the above inequality goes to zero. If $|\beta'| > \beta_0$, then the left-hand side explodes. If $|\beta'| = \beta_0$, then the left-hand side does not go to zero for all x. Hence, $|\beta'| < \beta_0$ and $\zeta(\cdot)$ and $f_0(\cdot)$ are orthogonal to each other, i.e.,

$$\int_0^{\infty} f_0(y)\zeta(y)d\upsilon(y) = 0.$$

The above equality also proves that $f_0(\cdot)$ is the only non-negative eigenfunction. □

We summarize the key conclusions in the following theorem.

Theorem III.22. *Assume the moment generating function of n_ϕ exists and is finite for some $\theta > 0$. Let β_0 and $f_0(\cdot)$ to be as in Theorem III.21. Then $\beta_0 \in \left(\max_x\left(x\upsilon([x,\infty))\right), \mathbb{E}[n_\phi] - 1\right)$ is the largest eigenvalue of M_1 in magnitude. The corresponding eigenfunctions are given as follows*

$$\textit{Right eigenfunction: } \mu(m,x) = \frac{m}{x}f_0(x),$$

$$\textit{Left eigenfunction: } \nu(k-1,z) = P(k)\frac{\mathrm{e}^{-z}z^{k-1}}{(k-1)!}f_0(z).$$

These eigenfunctions are the unique non-negative right and left eigenfunctions, respectively. Moreover, there exists $0<\gamma<1$ *and a constant* $\widehat{M}>0$ *independent of* x, m, z *and* k *such that for all* $x\in\mathbb{R}_+$, $y>0$, $k\geq 1$ *and* $m\geq 0$,

$$m_l(m,x;k-1,z) = \frac{P(k)\mathrm{e}^{-z}z^{k-1}}{(k-1)!}\frac{m}{x}\times {\beta_0}^l f_0(x)f_0(z)\left(1+\widehat{M}O(\gamma^{l-2})\frac{x}{{\beta_0}^2 f_0(z)f_0(x)}\right) \quad l\geq 2. \tag{3.44}$$

Finally, $m_l(m,x;k-1,z)$ *is related to the function* $h_l(x,y)$ *via the following equation,*

$$m_l(m,x;k-1,z) = h_l(x,z)\times\frac{P(k)\mathrm{e}^{-z}z^{k-1}}{(k-1)!}\frac{m}{x},$$

and for all functions $g:\mathbb{R}_+\to\mathbb{R}$ *satisfying* $|g(x)|\leq V(x)$ *for all* $x\in\mathbb{R}_+$,

$$\left|\int_0^\infty \frac{h_l(x,y)g_2(y)f_0(y)g(y)}{{\beta_0}^l f_0(x)}\,dy - \int_0^\infty g_2(y){f_0(y)}^2 g(y)\,dy\right| \leq M\gamma^l V(x) \qquad l\geq 2.$$

where $V(x) = {f_0}'(0)\exp(\eta x)\frac{x}{f_0(x)}$ *and* $\eta = (1-\mathrm{e}^{-\theta})/2$. *The constants* M *and* $0<\gamma<1$ *are independent of* x *and* l.

Using the above theorem, we get similar bounds for M_l which is useful for large l.

Corollary III.23. *The growth rate of* $M_l(m,x;\mathbb{R}_+,\mathbb{Z}_+)$ *equals* β_0 *which is given by Theorem III.22, i.e.,*

$$\left|\frac{M_l(m,x;\mathbb{R}_+,\mathbb{Z}_+)}{{\beta_0}^l} - \frac{m}{x}f_0(x)\int_0^\infty\sum_{k=1}^\infty P(k)\frac{\mathrm{e}^{-z}z^{k-1}}{(k-1)!}f_0(z)\,dz\right| = \frac{m}{{\beta_0}^2}\widehat{M}O(\gamma^{l-2}), \qquad l\geq 2$$

where the constant $0<\gamma<1$ *is independent of* x, m *and* l.

Proof. By Theorem III.22, we have

$$\begin{aligned} M_l(m,x;\mathbb{R}_+,\mathbb{Z}_+)/{\beta_0}^l &= \int_0^\infty\sum_{k=1}^\infty m_l(m,x;k-1,z)\,dz/{\beta_0}^l \\ &= \int_0^\infty\sum_{k=1}^\infty \frac{P(k)\mathrm{e}^{-z}z^{k-1}}{(k-1)!}\frac{m}{x}\times \\ &\qquad f_0(x)f_0(z)\left(1+\widehat{M}O(\gamma^{l-2})\frac{x}{{\beta_0}^2 f_0(z)f_0(x)}\right)\,dz \end{aligned}$$

$$= \frac{m}{x} f_0(x) \int_0^\infty \sum_{k=1}^\infty P(k) \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f_0(z)\, dz + \frac{m}{{\beta_0}^2} \widehat{M} O(\gamma^{l-2}) \int_0^\infty g_1(z) dz.$$

□

Recall that Z_l denotes the number of vertices at generation l. As an immediate Corollary, the growth/extinction rate of $\mathbb{E}[Z_l]$ is β_0 as well.

Corollary III.24. *We have*

$$\mathbb{E}[Z_l]/{\beta_0}^l \xrightarrow{l\to\infty} \left(\sum_{m=1}^\infty P(m) \int_{x=0}^\infty \frac{\mathrm{e}^{-x} x^m}{m!} \times \frac{m}{x} f_0(x)\, dx \right) \left(\int_0^\infty \sum_{k=1}^\infty P(k) \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f_0(z)\, dz \right).$$

If $\beta_0 > 1$, the expected number of vertices at generation l explodes as l goes to infinity. If $\beta_0 = 1$, the expected number of vertices at generation l stays bounded. If $\beta_0 < 1$, the expected number of vertices at generation l goes to zero.

3.4.6 Analysis of the Second Moments and Asymptotic Results for $\beta_0 > 1$

A follow-up question is the limit of the random variable Z_l/β_0^l: 1) If $\beta_0 < 1$, it is clear that $Z_l \to 0$ almost surely as $l \to \infty$ since the population will become extinct; however, conditioned on $Z_l > 0$, the distribution of the total number of vertices might be of interest. We leave this problem for future work. 2) If $\beta_0 > 1$, one way to study the limit is to analyze the second moment. This methodology was introduced by Harris in [78] and was generalized to finite type branching processes in [77]. In [79], Harris pointed out that a similar generalization is possible for general branching processes and discussed this further in [30, Chapter 3]. We follow his argument closely in this section. 3) The case $\beta = 1$ is tricky and is discussed in Section 3.4.7. We will prove that $Z_l \to 0$ almost surely as $l \to \infty$, however, a similar question as in 1) is left for future work.

Let $Z_l(\mathcal{A})$ denote the number of vertices at depth l of type $(k-1, \zeta) \in \mathcal{A}, \mathcal{A} \subset \Omega$. Recall that by the discussion of Section 3.2.2, $Z_l(\cdot)$ is a set function. For Borel sets $\mathcal{A}_1, \mathcal{A}_2 \subset \Omega$, define,

$$M_l^{(2)}(m, x; \mathcal{A}_1; \mathcal{A}_2) \coloneqq \mathbb{E}[Z_l(\mathcal{A}_1) Z_l(\mathcal{A}_2) | n_\emptyset = m, v_\emptyset = x], \qquad \forall l = 0, 1, \cdots, \tag{3.45}$$

$$v(m, x; \mathcal{A}_1; \mathcal{A}_2) \coloneqq M_1^{(2)}(m, x; \mathcal{A}_1; \mathcal{A}_2) - M_1(m, x; \mathcal{A}_1) M_1(m, x; \mathcal{A}_2). \tag{3.46}$$

The conditionally independent structure of EWT implies

$$M_1^{(2)}(m,x;\mathcal{A}_1;\mathcal{A}_2)=\begin{cases}\dfrac{m(m-1)}{x}\left(\sum\limits_{(k-1,\zeta)\in\mathcal{A}_1}\int P(k)\min(x,\zeta)\dfrac{\mathrm{e}^{-\zeta}\zeta^{k-1}}{(k-1)!}\,d\zeta\right)\times & \\ \qquad\left(\sum\limits_{(k-1,\zeta)\in\mathcal{A}_2}\int P(k)\min(x,\zeta)\dfrac{\mathrm{e}^{-\zeta}\zeta^{k-1}}{(k-1)!}\,d\zeta\right) & \text{if } x>0\\ \quad+\dfrac{m}{x}\left(\sum\limits_{(k-1,\zeta)\in\mathcal{A}_1\cap\mathcal{A}_2}\int P(k)\min(x,\zeta)\dfrac{\mathrm{e}^{-\zeta}\zeta^{k-1}}{(k-1)!}\,d\zeta\right), & \\ m(m-1)\left(\sum\limits_{(k-1,\zeta)\in\mathcal{A}_1}\int P(k)\dfrac{\mathrm{e}^{-\zeta}\zeta^{k-1}}{(k-1)!}\,d\zeta\right) & \\ \qquad\left(\sum\limits_{(k-1,\zeta)\in\mathcal{A}_2}\int P(k)\dfrac{\mathrm{e}^{-\zeta}\zeta^{k-1}}{(k-1)!}\,d\zeta\right) & \text{if } x=0\\ \quad+m\left(\sum\limits_{(k-1,\zeta)\in\mathcal{A}_1\cap\mathcal{A}_2}\int P(k)\dfrac{\mathrm{e}^{-\zeta}\zeta^{k-1}}{(k-1)!}\,d\zeta\right), & \end{cases}$$

where by $\sum\limits_{(k-1,\zeta)\in\mathcal{A}}\int d\zeta$ we mean $\sum_{k=1}^{\infty}\int_{\zeta:(k-1,\zeta)\in\mathcal{A}}d\zeta$. To get the above equality, note that

$$\begin{aligned}M_1^{(2)}(m,x;\mathcal{A}_1;\mathcal{A}_2)&=\mathbb{E}[Z_l(\mathcal{A}_1)Z_l(\mathcal{A}_2)-Z_l(\mathcal{A}_1\cap\mathcal{A}_2)|n_\emptyset=m,v_\emptyset=x]\\&\quad+\mathbb{E}[Z_l(\mathcal{A}_1\cap\mathcal{A}_2)|n_\emptyset=m,v_\emptyset=x],\end{aligned}$$

and also note that $Z_l(\mathcal{A}_1)Z_l(\mathcal{A}_2)-Z_l(\mathcal{A}_1\cap\mathcal{A}_2)$ equals the number of ways to select two different descendants of the root successively: first, a descendant of a type belongs to $\mathcal{A}_1$ and then a descendant of a type belongs to $\mathcal{A}_2$.

For any fixed (m,x), we can interpret $M_l^{(2)}(m,x;\mathcal{A}_1;\mathcal{A}_2)$ as the measure of the "rectangular" $\mathcal{A}_1\times\mathcal{A}_2$, i.e., the measure of points $(k_1-1,\zeta_1;k_2-2,\zeta_2)$ such that $(k_1-1,\zeta_1)\in\mathcal{A}_1$, $(k_2-1,\zeta_2)\in\mathcal{A}_2$, and $(k_i-1,\zeta_i)\in Z_l$, where Z_l (abusing notation) is the point distribution of vertices in generation l. To make the matters rigorous, we need to define bivariate measures and random double integrals.

Definition III.25. A function $F(\mathcal{A},\mathcal{B})$, where $\mathcal{A}$ and $\mathcal{B}$ are subsets of Ω, is called a bivariate measure if it satisfies the following conditions:

(a) it is finite and non-negative;

(b) if $\mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_k$ are disjoint subsets of Ω, then $F\left(\cup_j \mathcal{A}_j, \mathcal{B}\right) = \sum_j F(\mathcal{A}_j, \mathcal{B})$;

(c) if $\mathcal{B}_1, \mathcal{B}_2, \ldots \mathcal{B}_k$ are disjoint subsets of Ω, then $F\left(\mathcal{A}, \cup_j \mathcal{B}_j\right) = \sum_j F(\mathcal{A}, \mathcal{B}_j)$;

F is called a signed bivariate measure if $F = F_1 - F_2$, where F_1 and F_2 are bivariate measures.

Definition III.26. For a function $f(k_1 - 1, \zeta_1; k_2 - 1, \zeta_2)$ defined over $\Omega \times \Omega$, the random double integral is defined as follows:

$$\sum \!\!\!\!\!\!\int_{(k_2-1,\zeta_2)\in\Omega} \sum \!\!\!\!\!\!\int_{(k_1-1,\zeta_1)\in\Omega} f(k_1 - 1, \zeta_1; k_2 - 1, \zeta_2)\, d\omega(\zeta_1, k_1)\, d\omega(\zeta_2, k_2) = \sum_{i,j} a_i a_j f(m_i, x_i; m_j, x_j)$$

where ω is the point distribution $((m_1, x_1), a_1; (m_2, x_2), a_2; \ldots; (m_k, x_k), a_k)$.

By definition, $M_l^{(2)}(m, x; \mathcal{A}_1; \mathcal{A}_2)$ and $M_1(m, x; \mathcal{A}_1) M_2(m, x; \mathcal{A}_2)$ are bivariate measures, and $v(m, x; \mathcal{A}_1; \mathcal{A}_2)$ is a signed bivariate measure. Define a map $\mathcal{T}$ from the set of signed bivariate measures to itself as follows,

$$\mathcal{T}F(\mathcal{A}_1; \mathcal{A}_2) = \sum \!\!\!\!\!\!\int_{(k_2-1,\zeta_2)\in\Omega} \sum \!\!\!\!\!\!\int_{(k_1-1,\zeta_1)\in\Omega} M_1(k_1 - 1, \zeta_1; \mathcal{A}_1) M_1(k_2 - 1, \zeta_2; \mathcal{A}_2)\, dF(k_1 - 1, \zeta_1; k_2 - 1, \zeta_2).$$

To derive a recurrence relation between $M_l^{(2)}$ and $M_{l+1}^{(2)}$, write

$$M_{l+1}^{(2)}(m, x; \mathcal{A}_1; \mathcal{A}_2) = \mathbb{E}[\, \mathbb{E}[Z_{l+1}(\mathcal{A}_1) Z_{l+1}(\mathcal{A}_2) | Z_l = \omega]\, | n_\emptyset = m, v_\emptyset = x].$$

Conditioned on $Z_l = \omega \in \mathscr{P}$, the expected value of $Z_{l+1}(\mathcal{A}_1) Z_{l+1}(\mathcal{A}_2)$ is given by the following random integrals,

$$\begin{aligned}
&\sum \!\!\!\!\!\!\int_{(k_2-1,\zeta_2)\in\Omega} \sum \!\!\!\!\!\!\int_{(k_1-1,\zeta_1)\in\Omega} \mathbb{E}_{k_1-1,\zeta_1}[\widetilde{Z}_1(\mathcal{A}_1)]\, \mathbb{E}_{k_2-1,\zeta_2}[\widetilde{Z}_1(\mathcal{A}_2)]\, dZ_l(k_1 - 1, \zeta_1)\, dZ_l(k_2 - 1, \zeta_2) \\
&\qquad - \sum \!\!\!\!\!\!\int_{(k-1,\zeta)\in\Omega} \mathbb{E}_{k-1,\zeta}[\widetilde{Z}_1(\mathcal{A}_1)]\, \mathbb{E}_{k-1,\zeta}[\widetilde{Z}_1(\mathcal{A}_2)]\, dZ_l(k - 1, \zeta) \\
&\qquad + \sum \!\!\!\!\!\!\int_{(k-1,\zeta)\in\Omega} \mathbb{E}_{k-1,\zeta}[\widetilde{Z}_1(\mathcal{A}_1) \widetilde{Z}_1(\mathcal{A}_2)]\, dZ_l(k - 1, \zeta)
\end{aligned}$$

where $\widetilde{Z}_1$ is an *i.i.d.* copy of the point distribution Z_1 and $\mathbb{E}_{m_i,x_i}$ is the expected value conditioned on the type of the root to be (m_i, x_i). Now, taking expectation of the above

random integrals with respect to the point distribution Z_l, we derive the following recurrence relation,

$$M_{l+1}^{(2)}(m,x;\mathcal{A}_1;\mathcal{A}_2) = \mathcal{T}M_l^{(2)}(m,x;\mathcal{A}_1;\mathcal{A}_2) + \sum\int_{(k-1,\zeta)\in\Omega} v(k-1,\zeta;\mathcal{A}_1;\mathcal{A}_2)\, dM_l(m,x;k-1,\zeta). \tag{3.47}$$

Repeatedly using (3.47) and then applying (3.46), we get the following relation

$$M_{l+1}^{(2)}(m,x;\mathcal{A}_1;\mathcal{A}_2) = \mathcal{T}^l M_1(m,x;\mathcal{A}_1)M_1(m,x;\mathcal{A}_2) + \sum_{\widehat{l}=0}^{l} \mathcal{T}^{l-\widehat{l}} \left(\sum\int_{(k-1,\zeta)\in\Omega} v(k-1,\zeta;\mathcal{A}_1;\mathcal{A}_2)\, dM_{\widehat{l}}(m,x;k-1,\zeta) \right), \tag{3.48}$$

where $\mathcal{T}^0$ is the identity map. Finally, observe that

$$\mathcal{T}^l F(\mathcal{A}_1;\mathcal{A}_2) = \sum\int_{(k_1-1,\zeta_1)\in\Omega} \sum\int_{(k_2-1,\zeta_2)\in\Omega} M_l(k_1-1,\zeta_1;\mathcal{A}_1)M_l(k_2-1,\zeta_2;\mathcal{A}_2)\, dF(k_1-1,\zeta_1;k_2-1,\zeta_2), \tag{3.49}$$

which can be proved by induction and the following equality:

$$d\mathcal{T}F(k-1,\zeta;\widetilde{k}-1,\widetilde{\zeta}) = \sum\int_{(k_2-1,\zeta_2)\in\Omega} \sum\int_{(k_1-1,\zeta_1)\in\Omega} m_1(k_1-1,\zeta_1;k-1,\zeta)\, d\zeta\, m_1(k_2-1,\zeta_2;\widetilde{k}-1,\widetilde{\zeta})\, d\widetilde{\zeta}\, dF(k_1-1,\zeta_1;k_2-1,\zeta_2).$$

Now, we can use the analysis of the previous section to approximate $M_l^{(2)}(m,x;\mathcal{A}_1;\mathcal{A}_2)$ for large values of l. This is basically the same result as in [30, page 72, eqn. (13.5)].

Theorem III.27. *With β_0 and $\mu(m,x)$ as specified in Theorem III.22, the growth rate of $M_l^{(2)}(m,x;\mathcal{A}_1;\mathcal{A}_2)$ equals ${\beta_0}^2$, i.e.,*

$$M_l^{(2)}(m,x;\mathcal{A}_1,\mathcal{A}_2)/{\beta_0}^{2l} = U(m,x) \left(\sum\int_{(k-1,z)\in\mathcal{A}_1} \nu(k-1,z)\, dz \right) \left(\sum\int_{(k-1,z)\in\mathcal{A}_2} \nu(k-1,z) dz \right) \tag{3.50}$$

$$+ m^2\overline{M}O(\gamma^{l-2}), \qquad l \geq 2,$$

where the constants $\overline{M} > 0$ *and* $0 < \gamma < 1$ *are independent of* x, l, $\mathcal{A}_1$, *and* $\mathcal{A}_2$. *The function* $U(m,x)$ *is defined as follows,*

$$\begin{aligned} U(m,x) \coloneqq \\ \mu(m,x)^2 + \sum_{\widehat{l}=1}^{\infty} {\beta_0}^{-2\widehat{l}} \sum \!\!\!\!\!\!\int_{(k-1,z)\in\Omega} \Bigg(\sum \!\!\!\!\!\!\int_{(k_1-1,z_1)\in\Omega} \sum \!\!\!\!\!\!\int_{(k_2-1,z_2)\in\Omega} \mu(k_1-1,z_1)\mu(k_2-1,z_2) \\ dv(k-1,z;k_1-1,z_1;k_2-1,z_2) \Bigg) dM_{\widehat{l}-1}(m,x;k-1,z). \end{aligned} \tag{3.51}$$

Remark III.28. Note that using Theorem III.22 the summand in the definition of $U(m,x)$ is $O(\beta_0^{-\widehat{l}-1})$; hence, the sum is finite and $U(m,x)$ is well-defined.

Proof. As we pointed out in the proof of Corollary III.23, using Theorem III.22 for any $\mathcal{A} \in \Sigma$ we have,

$$M_l(m,x;\mathcal{A})/{\beta_0}^l = \sum \!\!\!\!\!\!\int_{(k-1,z)\in\mathcal{A}} \mu(m,x)\nu(k-1,z)\,dz + \frac{m}{{\beta_0}^2}\widehat{M}O(\gamma^{l-2}).$$

Substituting the above equality in (3.49), after some simple algebra we have

$$\begin{aligned} \mathcal{T}^l F(\mathcal{A}_1;\mathcal{A}_2)/{\beta_0}^{2l} = \\ \Bigg(\sum \!\!\!\!\!\!\int_{(k_1-1,z_1)\in\Omega} \sum \!\!\!\!\!\!\int_{(k_2-1,z_2)\in\Omega} \mu(k_1-1,z_1)\mu(k_2-1,z_2)\,dF(k_1-1,z_1;k_2-1,z_2) \Bigg) \times \\ \Bigg(\sum \!\!\!\!\!\!\int_{(k-1,z)\in\mathcal{A}_1} \nu(k-1,z)\,dz \Bigg) \Bigg(\sum \!\!\!\!\!\!\int_{(k-1,z)\in\mathcal{A}_2} \nu(k-1,z)\,dz \Bigg) + C_F\,\widehat{M}O(\gamma^{l-2}), \end{aligned}$$

Now the result follows by combining (3.47), the above equality, and the following relation

$$\mathcal{T}^{l-1}M_1(m,x;\mathcal{A}_1)M_1(m,x;\mathcal{A}_2) = M_l(m,x;\mathcal{A}_1)M_l(m,x;\mathcal{A}_2).$$

which can be proved using induction similar to (3.49). The constant C_F depends on the choice of the function F. It is easy to check that for $v(m,x;\mathcal{A}_1,\mathcal{A}_2)$, we can replace $C_F\widehat{M}$ with $m^2\overline{M}$ for some $\overline{M} > 0$ independent of x, l, $\mathcal{A}_1$, and $\mathcal{A}_2$ (note that $\min(x,z)/x \leq 1$). □

Remark III.29. Fix the value of $\widetilde{l} > 0$ and consider $\mathbb{E}[Z_l(\mathcal{A}_1)Z_{l+\widetilde{l}}(\mathcal{A}_2)|n_\emptyset = m, v_\emptyset = x]$. Using the same argument as above, the conditional expectation converges to the same value as in (3.50) with the error bounded by $m^2\overline{M}_{\widetilde{l}}O(\gamma^{l-2})$.

Now, combining the above theorem and remark, we get a similar result as in [30, Theorem 14.1, page 72].

Theorem III.30. *Let $\mathcal{A} \subset \Omega$ and set $W_l(\mathcal{A}) = Z_l(\mathcal{A})/{\beta_0}^l$. Then, conditioned on $n_\emptyset = m$ and $v_\emptyset = x$, where $x \in \mathbb{R}_+$ and $m \in \mathbb{N}$, there is a random variable $W(\mathcal{A})$ such that $W_l(\mathcal{A})$ converges to $W(\mathcal{A})$ in L^2 and almost surely. The first and the second moments of $W(\mathcal{A})$ are,*

$$\mathbb{E}\left[W(\mathcal{A})|n_\emptyset = m, v_\emptyset = x\right] = \frac{m}{x}f_0(x)\left(\sum\int_{(k-1,z)\in\mathcal{A}} P(k)\frac{\mathrm{e}^{-z}z^{k-1}}{(k-1)!}f_0(z)\,dz\right),$$

$$\mathbb{E}\left[W(\mathcal{A})^2\middle|n_\emptyset = m, v_\emptyset = x\right] = U(m,x)\left(\sum\int_{(k-1,z)\in\mathcal{A}} \nu(k-1,z)\,dz\right)^2,$$

where the function $U(x,m)$ is given by (3.51). *Furthermore, if $\mathcal{A}$ and $\mathcal{B}$ are subsets of Ω such that $\sum\int_{(k-1,z)\in\mathcal{A}} \nu(k-1,z)\,dz > 0$, then*

$$W(\mathcal{B}) = \frac{\sum\int_{(k-1,z)\in\mathcal{B}} \nu(k-1,z)\,dz}{\sum\int_{(k-1,z)\in\mathcal{A}} \nu(k-1,z)\,dz}W(\mathcal{A}) \qquad a.s.$$

Proof. Remark III.29 and equation (3.50) imply that $\mathbb{E}[(W_l(\mathcal{A})-W_{l+\widehat{l}}(\mathcal{A}))^2] = m^2\overline{M}_{\widehat{l}}O(\gamma^{l-2})$. Hence, $\{W_l(\mathcal{A})\}_l$ satisfies the Cauchy criteria and converges to $W(\mathcal{A})$ in L^2. Since for any $\widehat{l} > 0$

$$\sum_{l=1}^{\infty}\mathbb{E}[(W_l(\mathcal{A}) - W_{l+\widehat{l}}(\mathcal{A}))^2] < \infty,$$

$\{W_l(\mathcal{A})\}_l$ converges to $W(\mathcal{A})$ almost surely as well. Finally, the relation between $W(\mathcal{B})$ and $W(\mathcal{A})$ follows by the following relation between $W_l(\mathcal{A})$ and $W_l(\mathcal{B})$:

$$\mathbb{E}\left[\left(W_l(\mathcal{B}) - \frac{\sum\int_{(k-1,z)\in\mathcal{B}} \nu(k-1,z)\,dz}{\sum\int_{(k-1,z)\in\mathcal{A}} \nu(k-1,z)\,dz}W_l(\mathcal{A})\right)^2\right] = m^2\overline{M}\,O(\gamma^{l-2}).$$

□

An immediate corollary of the above theorem and Corollary III.24 is the following, which connects the growth rate and the probability of extinction.

Corollary III.31. *If $\beta_0 > 1$ then the probability of extinction is less than 1. If $\beta_0 < 1$ then the probability of extinction equals 1.*

Proof. By Theorem III.30 if $\beta_0 > 1$, then $W(\mathcal{A})$ is positive with non-zero probability. Hence, the probability of extinction is less than 1. The second part follows by Markov inequality and Corollary III.24. □

3.4.7 Transience of Z_l

To analyze the case of $\beta_0 = 1$ and to show that $Z_l \sim {\beta_0}^l W$ we need to show transience of Z_n, i.e., Z_n either goes to zero or infinity. Consider the generalized Markov Chain introduced in Section 3.2.2. Recall that $Z_l(\mathcal{A})$ is the number of vertices $(k-1, \zeta) \in \mathcal{A}$, and $Z_l(\Omega)$ is the total number of vertices at generation l. Note that for any $\kappa \in \mathbb{N}$, the probability of extinction after κ steps conditioned on $n_\emptyset = m$, $v_\emptyset = x$ can be arbitrary small when m is large. As a result, the same proof technique as in [30, Theorem 11.2, page 69] does not work in our problem setting. To show the transience of Z_l, more work needs to be done. The following lemma establishes the transience of Z_l. We follow the notation introduced in Section 3.2.2.

Lemma III.32. *For all $k \geq 1$ and for all $\omega \in \mathscr{P}_\Omega$ we have,*

$$\mathbb{P}(\{0 < Z_l(\Omega) \leq k, \ \textit{infinitely often}\}) = 0.$$

Proof. Define $\mathscr{P}_{\Omega_0}$ to be the set of non-null point distributions with at most k vertices,

$$\mathscr{P}_{\Omega_0} \coloneqq \{\omega \in \mathscr{P}_\Omega \,|\, 0 < \omega(\Omega) \leq k\}.$$

Let $\mathscr{P}_{\Omega_0,m}$ be the set of point distributions $\omega = ((m_1, x_1), a_1; (m_2, x_2), a_2; \ldots; (m_{\widetilde{k}}, x_{\widetilde{k}}), a_{\widetilde{k}}) \in \mathscr{P}_{\Omega_0}$ such that $m_i \leq m$ for all i. Recall that $\emptyset$ denote the null point distribution.

Step 1: Using the same argument as in [30, Theorem 11.2, page 69], we show that $\mathbb{P}(Z_l \in \mathscr{P}_{\Omega_0,m}) = 0$. Define $R_m(\omega)$ for $\omega \in \mathscr{P}_\Omega$ as follows:

$$R_m(\omega) = \mathbb{P}(\{Z_l \in \mathscr{P}_{\Omega_0,m}, \ \text{infinitely often}\} \,|\, Z_0 = \omega).$$

For $\mathscr{P} \subset \mathscr{P}_{\Omega_0,m}$ let $Q_{m,2}(\omega, \mathscr{P})$ be the conditional probability that, conditioned on $Z_0 = \omega$, at least one of the point distributions Z_2, Z_3, $\cdots$ are in $\mathscr{P}_{\Omega_0,m}$ and if Z_l is the first one, then $Z_l \in \mathscr{P}$. Then

$$R_m(\omega) = \int_{\mathscr{P}_{\Omega_0,m}} R_m(\omega') dQ_{m,2}(\omega, \omega').$$

Let $\overline{R}_m := \sup_{\omega \in \mathscr{P}_{\Omega_0,m}} R_m(\omega)$. We have

$$R_m(\omega) \leq \overline{R}_m \int_{\mathscr{P}_{\Omega_0,m}} dQ_{m,2}(\omega, \omega') = \overline{R}_m \, Q_{m,2}(\omega, \mathscr{P}_{\Omega_0,m}). \tag{3.52}$$

In the proof of Theorem III.4, we show that, if $Z_0 = (m_i, x_i)$, the probability of extinction after 2 generations is given by $(T^2(\mathbf{0})(x_i))^{m_i}$. Recall that $T^2(\mathbf{0})(\cdot)$ is a decreasing and strictly positive function. Hence,

$$Q_{m,2}(\omega, \mathscr{P}_{\Omega_0,m}) \leq 1 - \prod_{i=1}^{\widetilde{k}} \left(T^2(\mathbf{0})(x_i)\right)^{a_i m_i} \leq 1 - \left(T^2(\mathbf{0})(0)\right)^{mk} < 1 - \epsilon. \tag{3.53}$$

where $\omega = ((x_1, m_1), a_1; (x_2, m_2), a_2; \cdots; (x_{\widetilde{k}}, m_{\widetilde{k}}), a_{\widetilde{k}})$ and $\epsilon > 0$ is a constant which depends only on m and k. Contradiction follows by taking supremum from both sides of (3.52).

Remark III.33. In *Step 1*, we proved that the probability of the event $\{Z_l \in \mathscr{P}_{\Omega_0,m}, \text{ infinitely often}\}$ is zero. However (as $\mathscr{P}_{\Omega_0,m} = \cup_m \mathscr{P}_{\Omega_0,m}$) this approach cannot rule out the possibility of the event $\{Z_l \in \mathscr{P}_{\Omega_0}, \text{ infinitely often}\}$. As an example, there might be a sequence $\{l_i\}_{i=1}^{\infty}$ such that $Z_{l_i} \in \mathscr{P}_{\Omega_0,m_{l_i}}$ where $m_{l_1} < m_{l_2} < m_{l_3} < \cdots$. In *Step 2*, we will prove that such sequences are unlikely.

Step 2: For the sake of notational simplicity, we prove the result for $k = 1$ and then discuss the general case. Fix the value of m. Note that by the first step, the probability of hitting $\mathscr{P}_{\Omega_0,m}$ infinitely often is zero. Hence, we need to show that the probability of hitting $\widetilde{\mathscr{P}}_{\Omega_0,m} := \mathscr{P}_{\Omega_0} \setminus \mathscr{P}_{\Omega_0,m}$ infinitely often is zero to complete the proof.

Remark III.34. By conditional independence, the transition kernel of the generalized Markov chain from the point distribution $\omega = ((m_1, x_1), a_1; (m_2, x_2), a_2; \ldots; (m_{\widetilde{k}}, x_{\widetilde{k}}), a_{\widetilde{k}}) \in \mathscr{P}_{\Omega_0}$ is exactly the same as the transition kernel from the point distribution $\omega = ((1, x_1), m_1 \times a_1; (1, x_2), m_2 \times a_2; \ldots; (1, x_{\widetilde{k}}), m_{\widetilde{k}} \times a_{\widetilde{k}})$ which may or may not be in $\mathscr{P}_{\Omega_0}$.

Assume $k = 1$ and let $\kappa \in \mathbb{N}$. Define $\widetilde{Q}_{m,\kappa}(\omega, \mathscr{P})$ and $\widetilde{R}_m(\omega)$ similar to $Q_{m,2}(\omega, \mathscr{P})$ and $R_m(\omega)$ by considering the set $\widetilde{\mathscr{P}}_{\Omega_0,m}$ instead of $\mathscr{P}_{\Omega_0,m}$. Specifically, let $\widetilde{Q}_{m,\kappa}(\omega, \mathscr{P})$ be the conditional probability that, conditioned on $Z_0 = \omega$, at least one of the point distributions

$Z_\kappa, Z_{\kappa+1}, Z_{\kappa+2}, \cdots$ are in $\widetilde{\mathscr{P}}_{\Omega_0,m}$ and if Z_l is the first one, then $Z_l \in \mathscr{P}$. Similarly, define $\widetilde{R}_m(\omega)$ for $\omega \in \widetilde{\mathscr{P}}_\Omega$ as follows:

$$\widetilde{R}_m(\omega) = \mathbb{P}(\{Z_l \in \widetilde{\mathscr{P}}_{\Omega_0,m},\ \text{infinitely often}\} \,|\, Z_0 = \omega).$$

Assume $Z_0 = \omega = (m_1, x_1)$, where $m_1 \geq m$. Note that the first time $Z_l \in \widetilde{\mathscr{P}}_{\Omega_0,m}$ for some $l > 0$, $m_1 - 1$ out of m_1 branches of Z_0 go extinct. Hence, by Remark III.34

$$\widetilde{R}_m(\omega) = m_1 q(x_1)^{m_1-1} \widetilde{R}_m(\widehat{\omega}) \tag{3.54}$$

where $q(\cdot)$ is the smallest fixed point of the operator T defined in Theorem III.4, and $\widehat{\omega} = (1, x_1)$ is a point distribution with only one point of type $(1, x_1)$. Next, using the same argument as in step 1, we have

$$\widetilde{R}_m(\widehat{\omega}) \leq \overline{\widetilde{R}}_m\, \widetilde{Q}_{m,\kappa}(\widehat{\omega}, \widetilde{\mathscr{P}}_{\Omega_0,m}) \leq \overline{\widetilde{R}}_m(1 - T^\kappa(\mathbf{0})(x_1)).$$

Note that $\widetilde{R}_m(\widehat{\omega})$ does note depend on κ. Now, if we take κ to infinity, we have

$$\widetilde{R}_m(\widehat{\omega}) \leq \overline{\widetilde{R}}_m\,(1 - q(x_1)). \tag{3.55}$$

Combining (3.54) and (3.55), and taking supremum with respect to ω, the result follows by the fact that $m_1 q(x_1)^{m_1-1}(1 - q(x_1)) = Bi(1\,; m_1, 1 - q(x_1)) \leq 0.5$ for $m_1 \geq m > 1$.

Now consider the case $k = 2$ and pick $\omega = ((m_1, x_1), a_1; (m_2, x_2), a_2)$, where $a_1 + a_2 \leq 2$. Assume $\widetilde{R}_m(\omega) > 0$. Note that the point distribution ω has $m_1 a_1 + m_2 a_2$ potential branches dangling from it. Conditioned on $\{Z_l \in \widetilde{\mathscr{P}}_{\Omega_0,m},\ \text{infinitely often}\} \cap \{Z_0 = \omega\}$, if two of these potential branches survive, do not go extinct at all, then by Remark III.34 we have

$$\mathbb{P}(\{Z_l(\Omega) = 1,\ \text{infinitely often}\} \,|\, Z_0 = (1, x_i)) > 0,$$

for some i which is a contradiction. That is to say, conditioned on $\{Z_l \in \widetilde{\mathscr{P}}_{\Omega_0,m}, \text{infinitely often}\}$ $\cap \{Z_0 = \omega\}$, if two out of $m_1 a_1 + m_2 a_2$ potential branches survives with positive probability then each one will hit point distributions with only one vertex infinitely often.

Hence, only one of these branches can survive. Following the similar logic as before, we have

$$\widetilde{R}_m(\omega) = m_1 q(x_1)^{a_1 m_1 - 1} q(x_2)^{a_2 m_2} \widetilde{R}_m(\widehat{\omega}_1) + m_2 q(x_1)^{a_1 m_1} q(x_2)^{a_2 m_2 - 1} \widetilde{R}_m(\widehat{\omega}_2) \tag{3.56}$$

where $\widehat{\omega}_1 = (1, x_2)$ and $\widehat{\omega}_2 = (1, x_2)$. The result follows by same argument using (3.55). The

exact same argument holds for any $k > 2$ as well, and we get similar relation as (3.56). $\square$

The above lemma together with Corollary III.31 and Corollary III.24 have an important implication which completes the connection between probability of extinction and the growth rate.

Corollary III.35. *If $\beta_0 > 1$ then the probability of extinction is less than 1. If $\beta_0 \leq 1$ then the probability of extinction equals 1.*

3.4.8 Probability of Extinction Revisited

To show that growth rate of Z_n is β_0 when $\beta_0 > 1$, i.e., $Z_n \sim {\beta_0}^n W$, we need to show that $\mathbb{P}(W = 0 \,|\, Z_n \to \infty) = 0$. As Harris points out in [30, Remark 1, page 28], if there is a positive probability that $Z_n \to \infty$ at a rate less than β_0, then $\mathbb{P}(W = 0 \,|\, Z_n \to \infty) > 0$. To rule out such a scenario, we need to show that $\mathbb{P}(W = 0 \,|\, n_\emptyset = m, v_\emptyset = x) = \mathbb{P}(\{\text{extinction}\} \,|\, n_\emptyset = m, v_\emptyset = x) = q(x)^m$, where $q(\cdot)$ is given by Theorem III.4. In fact, it is easy to see that $\mathbb{P}(W = 0 \,|\, n_\emptyset = 1, v_\emptyset = x)$ is a fixed point of the operator T. However, to complete the argument we need to show that $T(\cdot)$ does not have any fixed point other than $q(\cdot)$ and $\mathbf{1}(\cdot)$.

Using the point process perspective, we can rewrite the operator T as follows:

$$T(f)(x) = \int\limits_{\omega=(m_1,x_1)\in\mathscr{P}_\Omega \text{ or } \omega=\emptyset} (f(x_1))^{m_1} dP^{(1)}_{\omega_0}(\omega)$$

where $\omega_0 = (1, x)$ is the type of the root vertex and $P^{(1)}_{\omega_0}$ is the one step transition probability defined in Section 3.2.2. For ease of representation, we define

$$\int\limits_{\omega=\emptyset} (f(x_1))^{m_1} dP^{(1)}_{\omega_0}(\omega) \coloneqq \int\limits_{\omega=\emptyset} dP^{(1)}_{\omega_0}(\omega) = \mathbb{P}(Z_1(\Omega) = 0 | Z_0 = (1, x)). \tag{3.57}$$

Inductively, using the same argument as in Remark III.34 we have

$$T^l(f)(x) = \int\limits_{\omega=((m_1,x_1),a_1;\cdots;(m_k,x_k),a_k)\in\mathscr{P}_\Omega,\, k\geq 0} (f(x_1))^{m_1 a_1} \cdots (f(x_k))^{m_k a_k} dP^{(l)}_{\omega_0}(\omega), \tag{3.58}$$

where by $k = 0$ we mean $\omega = \emptyset$ which follows the same definition as in (3.57). The above equality combined with an appropriate test function becomes a powerful tool to study properties of the operator T and the branching process in general. Recall that $q(\cdot)$ is the smallest fixed point of the operator T (Theorem III.4).

Lemma III.36. *If $\beta > 1$, then the operator T has two fixed points, one of which is : $q(\cdot)$ and $\mathbf{1}(\cdot)$. Moreover, for any function $f \in L(\mathbb{R}_+; [0,1])$ such that the Lebesgue measure of the set $\{x \in \mathbb{R}_+ : f(x) < 1\}$ is positive, $T^l(f)(x) \to q(x)$ for all $x \in \mathbb{R}_+$.*

Proof. Consider the function $\bar{f}_{x_0,\epsilon}(\cdot)$ defined as follows

$$\bar{f}_{x_0,\epsilon}(x) \coloneqq \begin{cases} \epsilon & \text{if } x \le x_0 \\ 1 & \text{otherwise} \end{cases}.$$

The goal is to show that for every large enough x_0, there is an $\epsilon > 0$ such that $T(\bar{f}_{x_0,\epsilon})(x) \le q(x)$ for all $x \in \mathbb{R}_+$, where $q(\cdot)$ is the smallest fixed point of the operator T. One important implication of this inequality is:

$$\lim_{l\to\infty} T^l(\bar{f}_{x_0,\epsilon})(x) = q(x), \qquad \forall x \in \mathbb{R}_+. \tag{3.59}$$

Note that for $x > 0$,

$$\begin{aligned} q(x) - T(\bar{f}_{x_0,\epsilon})(x) &= \frac{1}{x}\Bigg(\sum_{k=1}^{\infty} P(k) \int_{z=0}^{x_0} \frac{e^{-z} z^{k-1}}{(k-1)!} \min(x,z)(q(z)^{k-1} - \epsilon^{k-1})\, dz \\ &\qquad - \sum_{k=1}^{\infty} P(k) \int_{z=x_0}^{\infty} \frac{e^{-z} z^{k-1}}{(k-1)!} \min(x,z)(1 - q(z)^{k-1})\, dz \Bigg) \\ &\ge \frac{1}{x}\Bigg(\sum_{k=1}^{\infty} P(k)(q(0)^{k-1} - \epsilon^{k-1}) \int_{z=0}^{x_0} \frac{e^{-z} z^{k-1}}{(k-1)!} \min(x,z)\, dz \\ &\qquad - \sum_{k=1}^{\infty} P(k) \int_{z=x_0}^{\infty} \frac{e^{-z} z^{k-1}}{(k-1)!} \min(x,z)\, dz \Bigg) \end{aligned}$$

By choosing x_0 to be large enough, we can make the second term in the parenthesis to be arbitrary small. Fixing x_0, we can choose $\epsilon > 0$ to be small enough such that $q(x) - T(\bar{f}_{x_0,\epsilon})(x) > 0$ for all $x \in \mathbb{R}_+$. Note that $q(0) > 0$ and $q(x)$ is a strictly increasing function.

Now that we have proved (3.59), we use the alternative representation of $T^l(\bar{f}_{x_0,\epsilon})$ as in (3.58) to prove the lemma. Define the set $\mathscr{P}_M$ as

$$\mathscr{P}_M \coloneqq \left\{ w = ((m_1,x_1),a_1;(m_2,x_2),a_2;\cdots;(m_k,x_k),a_k),\ k \ge 0 \;\middle|\; \sum_{i \text{ s.t. } x_i \le x_0} m_i a_i < M \right\}.$$

Now, we have

$$T^l(\bar{f}_{x_0,\epsilon})(x) = \mathbb{P}(Z_l = 0 \,|\, Z_0 = (1,x)) + \int\limits_{\omega \in \mathscr{P}_M} \epsilon^{\sum m_i a_i} dP^{(l)}_{\omega_0}(\omega_1) + \int\limits_{\omega \in \mathscr{R}_\Omega \setminus \mathscr{P}_M} (f(x_1))^{m_1 a_1} (f(x_2))^{m_2 a_2} \cdots (f(x_k))^{m_k a_k} dP^{(l)}_{\omega_0}(\omega_1).$$

Note that,

$$\int\limits_{\omega \in \mathscr{P}_M} \epsilon^{\sum m_i a_i} dP^{(l)}_{\omega_0}(\omega_1) \geq \epsilon^M \mathbb{P}(Z_l \in \mathscr{P}_M | Z_0 = (1,x))$$

However, by (3.59), the left hand-side of the above inequality goes to 0 as l goes to infinity. Hence,

$$\mathbb{P}(Z_l \in \mathscr{P}_M | Z_0 = (1,x)) \to 0 \text{ as } l \to \infty.$$

For sake of contradiction, assume that T has another fixed point $\widetilde{q}(\cdot)$. By Lemma III.5, we already know that $q(x) < \widetilde{q}(x) < 1$ for all $x \in \mathbb{R}_+$ and that $\widetilde{q}(\cdot)$ is strictly increasing. Note that,

$$\widetilde{q}(x) = T^l(\widetilde{q})(x) = \mathbb{P}(Z_l = 0 \,|\, Z_0 = (1,x)) + \int\limits_{\omega \in \mathscr{P}_M} (\widetilde{q}(x_1))^{m_1 a_1} (\widetilde{q}(x_2))^{m_2 a_2} \cdots (\widetilde{q}(x_k))^{m_k a_k} dP^{(l)}_{\omega_0}(\omega_1) + \int\limits_{\omega \in \mathscr{R}_\Omega \setminus \mathscr{P}_M} (\widetilde{q}(x_1))^{m_1 a_1} (\widetilde{q}(x_2))^{m_2 a_2} \cdots (\widetilde{q}(x_k))^{m_k a_k} dP^{(l)}_{\omega_0}(\omega_1).$$

As l goes to infinity, the first term converges to $q(x)$. Using the analysis of $\bar{f}_{x_0,\epsilon}$, the second term goes to 0 since

$$\int\limits_{\omega \in \mathscr{P}_M} (\widetilde{q}(x_1))^{m_1 a_1} (\widetilde{q}(x_2))^{m_2 a_2} \cdots (\widetilde{q}(x_k))^{m_k a_k} dP^{(l)}_{\omega_0}(\omega_1) \leq \mathbb{P}(Z_l \in \mathscr{P}_M | Z_0 = (1,x)) \xrightarrow{l \to \infty} 0.$$

Finally, we can bound the third term as follows,

$$\int\limits_{\omega\in\mathscr{R}_\Omega\setminus\mathscr{P}_M} (\widetilde{q}(x_1))^{m_1a_1}(\widetilde{q}(x_2))^{m_2a_2}\cdots(\widetilde{q}(x_k))^{m_ka_k}dP^{(l)}_{\omega_0}(\omega_1)$$
$$\leq \widetilde{q}(x_0)^M\mathbb{P}(Z_l\in\mathscr{P}_M|Z_0=(1,x))\leq\widetilde{q}(x_0)^M$$

since $\widetilde{q}(\cdot)$ is non-decreasing (Lemma III.5). Combining these inequalities, we have,

$$\widetilde{q}(x)=\lim_{l\to\infty}T^l(\widetilde{q})(x)\leq q(x)+\widetilde{q}(x_0)^M$$

The result follows by letting M to infinity.

Finally, if $f\in L(\mathbb{R}_+;[0,1])$ such that the Lebesgue measure of the set $\{x\in\mathbb{R}_+ : f(x)<1\}$ is positive, then by same analysis and the fact that $T(f)(x)<1$ for all $x\in\mathbb{R}_+$, we have

$$\lim_{l\to\infty}T^l(f)(x)=q(x)\qquad\forall x\in\mathbb{R}_+.$$

□

As we pointed out in Section 3.4.6, one implication of the above lemma is $Z_n\sim{\beta_0}^nW$.

Theorem III.37. *If $\beta_0>1$, then the growth rate of Z_n is β_0, i.e., $\mathbb{P}(W=0\,|\,Z_n\to\infty)=0$. Moreover, conditioned on $Z_n\to\infty$, the proportions of different types converges to a constant.*

Proof. Let $f(x)=\mathbb{P}(W=0|Z_0=(1,x))$. Note that,

$$\mathbb{P}(W=0|Z_0=(m,x))=(f(x))^m,$$

and

$$\mathbb{P}(W=0|Z_0=(m,x))=\int\limits_{\omega=(m_1,x_1)\in\mathscr{R}_\Omega}(f(x_1))^{m_1}dP^{(1)}_{\omega_0}(\omega).$$

Hence, $f(x)$ is a fixed point of the operator T. On the other hand, by Theorem III.30, $\mathbb{P}(W=0|Z_0=(1,x))<1$ for all $x\in\mathbb{R}_+$. Hence, by Lemma III.36, $f(\cdot)=q(\cdot)$. Now, the result follows by law of total probability. The second part is just a corollary of the first part and Theorem III.30. □

3.5 Numerical Simulation

In this section, we present some numerical results for the case when P, the distribution of $n_\emptyset$, is the geometric distribution. We start by explicitly determining the the degree distribution of the root vertex, and Krein-Rutman eigenvalue and the corresponding eigenfunctions of M_1. Then we investigate various properties of the resulted EWT and compare its structural properties with unimodular Galton-Watson Trees [25].

Proposition III.1. *Assume P is the geometric distribution with parameter p, i.e., $P(k) = (1-p)^{k-1}p$ for all $k \in \mathbb{N}$. The following hold:*

(i) The probability distribution of the root vertex is given as follows:

$$\mathbb{P}(D_\emptyset = d) = \frac{p}{(1-p)^2}\left(1 - \sum_{m=0}^{d} \frac{\left(\frac{1-p}{p}\right)^m \mathrm{e}^{-\frac{1-p}{p}}}{m!}\right) - \frac{p}{1-p}\mathbb{1}\{d=0\} \qquad \forall d \geq 0$$

(ii) The extinction operator T is given as follows:

$$T(f)(x) \coloneqq \begin{cases} \dfrac{px - 1 + \mathrm{e}^{-px}}{px} + \dfrac{p}{x}\displaystyle\int_{z=0}^{\infty} \min(x,z) \exp\left(-z\left(1-(1-p)f(z)\right)\right) dz, & x > 0 \\ p\displaystyle\int_{z=0}^{\infty} \exp\left(-z\left(1-(1-p)f(z)\right)\right) dz, & x = 0 \end{cases}$$

(iii) The Krein-Rutman eigenvalue and the corresponding eigenfunctions of M_1 are given as follows:

$$\textit{Eigenvalue: } \beta_0 = \frac{4(1-p)}{{r_0}^2 p},$$

$$\textit{Right eigenfunction: } \mu(m,x) = \frac{m}{x} \frac{J_0\left(r_0 \mathrm{e}^{-\frac{p}{2}x}\right)}{\sqrt{\int_0^\infty p(1-p)\mathrm{e}^{-py} J_0\left(r_0 \mathrm{e}^{-\frac{p}{2}y}\right)^2 dy}},$$

$$\textit{Left eigenfunction: } \nu(k-1,z) = P(k)\frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \frac{J_0\left(r_0 \mathrm{e}^{-\frac{p}{2}z}\right)}{\sqrt{\int_0^\infty p(1-p)\mathrm{e}^{-py} J_0\left(r_0 \mathrm{e}^{-\frac{p}{2}y}\right)^2 dy}},$$

where $J_0(\cdot)$ is the zeroth-order Bessel function of first kind $J_0(x) = \sum_{i=0}^{\infty} \frac{(-1)^i}{i!\, i!}\left(\frac{x}{2}\right)^{2i}$, and $C_N = \left(\int_0^\infty g_2(y) L(\beta_0, y)^2 dy\right)^{-1}$ is a normalization factor.

Proof. The proofs of part (i) and (ii) are elementary and are presented in the Appendix B.4.

(iii) Recall the definition of g_2 and G_i as in Theorem III.12. We have

$$g_2(x) = \mathrm{e}^{-x} \sum_{k=2}^{\infty} (1-p)^{k-1} p \frac{x^{k-2}}{(k-2)!} = p(1-p)\mathrm{e}^{-px}.$$

Using the above equality together with a simple induction, we get

$$G_i(x) = \left(\frac{1-p}{p}\right)^i \frac{\mathrm{e}^{-ipx}}{i!\, i!}$$

Substituting the above equality into the definition of $L(\beta, x)$, we have

$$L(\beta, x) = \sum_{i=0}^{\infty} \left(\frac{4(1-p)\mathrm{e}^{-px}}{2p\beta}\right)^i \frac{(-1)^i}{i!\, i!} = J_0\left(\sqrt{\frac{4(1-p)\mathrm{e}^{-px}}{p\beta}}\right)$$

Note that $J_0\left(\sqrt{\frac{4(1-p)\mathrm{e}^{-px}}{p\beta}}\right)$ is the solution of the following differential equation:

$$\beta \frac{\mathrm{d}^2 q}{\mathrm{d}x^2} + p(1-p)\mathrm{e}^{-px}\, q(x) = 0$$

as we mentioned in Theorem III.12 part (ii) for the function $L(\beta, x)$.

Now, Let $r_0 \approx 2.4048$ denote the smallest zero of $J_0(\cdot)$. Recall that β_0 is the smallest root of $L(\cdot, 0)$, and the eigenfunction $f_0(\cdot)$ is given by $L(\beta_0, \cdot)\sqrt{C_N}$ where $C_N = (\int_0^\infty g_2(y) L(\beta_0, y)^2\, dy)^{-1}$. Then, by simple algebra

$$\beta_0 = \frac{4(1-p)}{{r_0}^2\, p}, \qquad \text{and} \qquad f_0(x) = \frac{J_0\left(r_0 \mathrm{e}^{-\frac{p}{2}z}\right)}{\sqrt{\int_0^\infty p(1-p)\mathrm{e}^{-py}\, J_0\left(r_0 \mathrm{e}^{-\frac{p}{2}y}\right)^2 dy}}.$$

□

The simple form of the geometric distribution makes it easier to study the associated Erlang Weighted Tree. Next, we numerically compare the degree distribution of EWT with unimodular Galton-Watson Trees (GWT_*). A GWT_* with degree distribution $Q \in \mathcal{P}(\boldsymbol{N})$ is a rooted tree, rooted at $\emptyset$, such that the number of descendants of the root is distributed as Q, and for all the other vertices, the offspring distribution is given by the size-biased distribution Q_*:

$$Q_*(k-1) = \frac{k\, Q(k)}{\sum_r r\, Q(r)}. \tag{3.60}$$

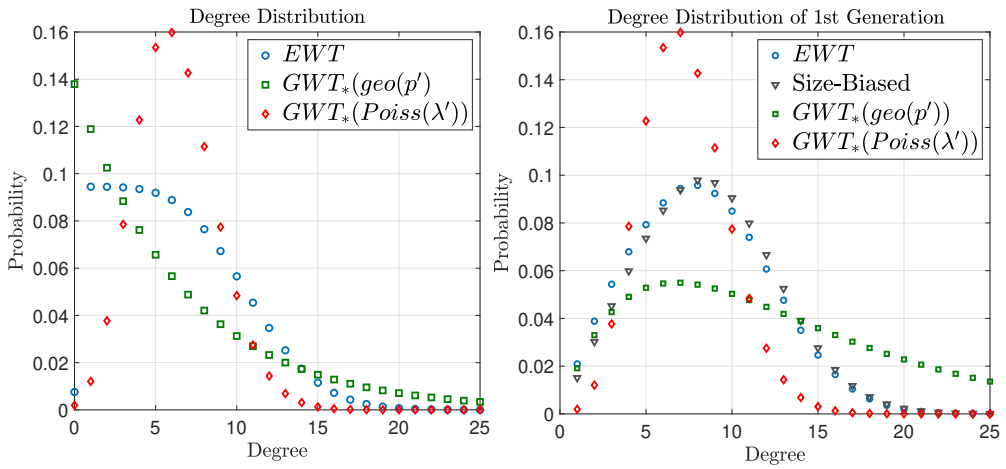


Figure 3.2: The degree distribution of the root vertex (zeroth generation) and the first generation of Erlang Weighted Tree (with potential degree distribution $geo(p)$), unimodular Galton-Watson Trees (with degree distribution $Poiss(\lambda')$ and $geo(p')$), and the size-biased degree distribution of the root of EWT. $p = 0.08$ and the parameters p' and λ' are chosen so that the expected degree of the root vertex is the same as in EWT.

In Figure 3.2, we compare the degree distribution of the zeroth and the first generation of EWT with GWT_*. We consider a GWT_* that has a Poisson degree distribution with parameter λ', and a GWT_* that has a geometric degree distribution with parameter p'. Both p' and λ' are chosen so that the expected degree of the root vertex is the same as in EWT. We also consider the size-biased distribution of the root vertex of EWT, using (3.60) and Theorem III.2. In this figure, the potential degree distribution of EWT is the geometric distribution with parameter 0.08. The degree distribution of EWT has different behavior compared with GWT_*. Most notably, the degree distribution of the first generation is not the size-biased distribution of the root vertex, as we also mentioned in Section 3.4.2. Since there is no closed-form for the degree distribution of the first generation in EWT, we numerically derive this distribution by averaging over 10^6 rooted trees.

Next, we compare the degree distribution of different generations of EWT. In Figure 3.3, we illustrate the degree distribution of the root, the first generation and the second generation of EWT with potential degree distribution $geo(0.08)$. Since EWT is the random weak limit of the finite graph model, we numerically derive the degree distribution of the first two generations of EWT by averaging over 1000 graphs with 10000 vertices. The error bars are also included in Figure 3.3. Note that the degree distribution of the first generation in Figures 3.2 and 3.3 are the same. Given the interdependence structure of EWT, the digree distribution of different generations are not the same. Note that the size-biased distribution of the root node is close to the degree distribution of the second generation. Intuitively

speaking, this means that the dependency between the degree distribution of generation l and the root node fades away as $l \to \infty$. We conjecture that the degree distribution of the lth generation converges to the size-biased degree distribution of the root vertex. This also suggest that the growth/extinction rate of EWT should be close to the growth/extinction rate of GWT_* with probability distribution given by the degree distribution of the root vertex in EWT.

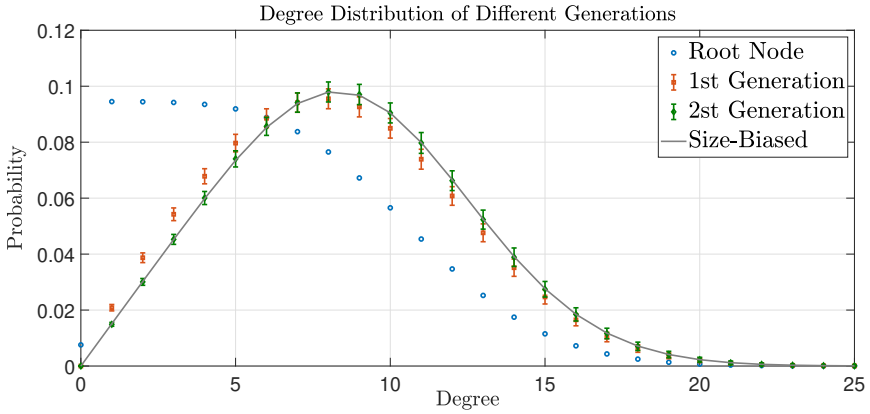


Figure 3.3: The degree distribution of the root vertex (zeroth generation), the first generation, and the second generation of a Erlang Weighted Tree with potential degree distribution $geo(0.08)$.

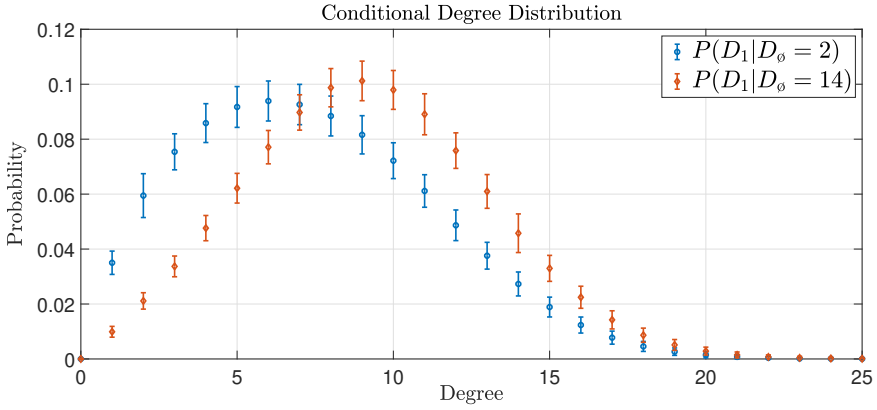


Figure 3.4: The conditional degree distribution of the first generation, conditioned on the degree of the root vertex of a Erlang Weighted Tree with potential degree distribution $geo(0.08)$.

Next, in Figure 3.4 we compare the conditional degree distribution of the first generation, conditioned on the degree of the root node. Similar to Figure 3.3, we use the finite graph model to estimate the conditional degree distribution. As we pointed out earlier, the degree distribution of the first generation, D_1, depends on the degree distribution of the root vertex,

$D_\emptyset$. In particular, higher value of $D_\emptyset$ increases the probability of observing larger values of D_1.

In Figure 3.5 we compare growth/extinction rate of the EWT with GWT$_*$. We consider a GWT$_*$ that has a Poisson degree distribution with parameter λ', a GWT$_*$ that has a geometric degree distribution with parameter p', and a GWT$_*$ with degree distribution given by the degree distribution of the root vertex of EWT. As we mentioned, the growth/extinction rate of EWT is close to the growth/extinction rate of GWT$_*$, however, they are not the same.

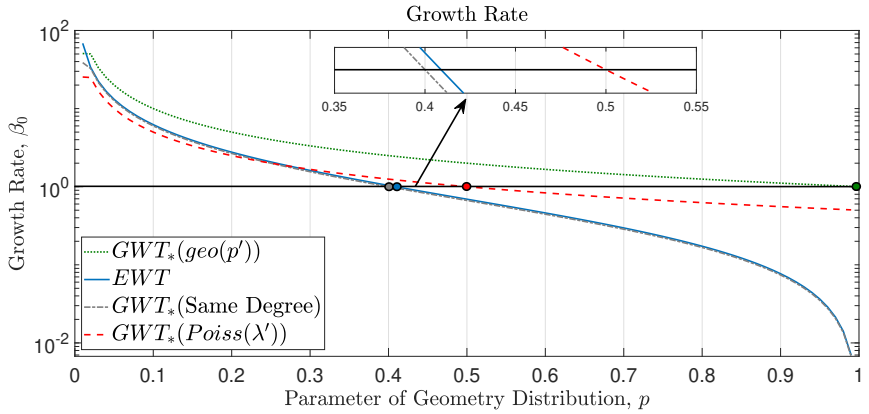


Figure 3.5: The growth rate of Erlang Weighted Tree (with potential degree distribution $geo(p)$) and unimodular Galton-Watson Trees (with degree distribution $Poiss(\lambda')$, $geo(p')$, and the degree distribution of the root vertex in EWT). $p = 0.08$ and the parameters p' and λ' are chosen so that the expected degree of the root vertex is the same as in EWT.

Finally, in Figure 3.6 we compare the probability of extinction of EWT with GWT$_*$. We consider the same set of unimodular Galton-Watson Trees as before. We also compare the ratio of vertices in the giant component of the finite graph model (with potential degree distribution $geo(p)$), with the random graphs generated by the configuration model (using the same degree distribution as in the associated GWT$_*$) and the Erdös-Renyi random graph (with parameter λ'/n, where n is the number of vertices), in Figure 3.7. We derive the size of the giant component of the finite graph model by averaging over 1000 graphs with 10000 vertices. The error bars are also included. The configuration model generates a random graph by uniformly pairing the half-edges assigned to vertices of the graph, where the number of half-edges assigned to a vertex is given by a fixed degree distribution. The Erdös-Renyi random graph with parameter λ'/n is given by connecting pairs of nodes to each other with probability λ'/n. For the configuration model and the Erdös-Renyi random graph model, this ratio equals $1 - \mathbb{P}(\{\text{extiction}\})$, where $\mathbb{P}(\{\text{extiction}\})$is the probability of extinction of the associated GWT$_*$ [25]. Figures 3.6 and 3.7 suggests that this is also true for EWT.

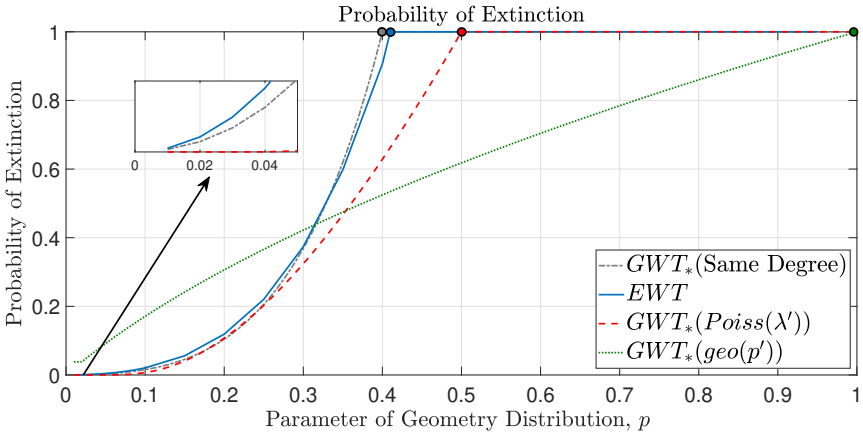
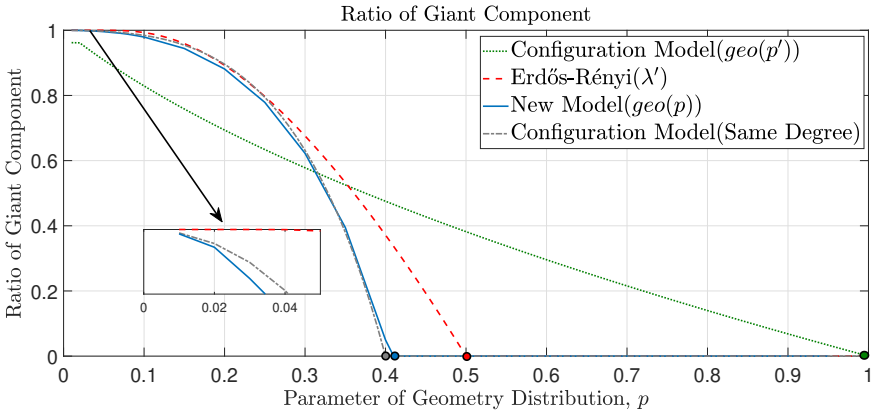


Figure 3.6: The probability of extinction of Erlang Weighted Tree (with potential degree distribution $geo(p)$) and unimodular Galton-Watson Trees (with degree distribution $Poiss(\lambda')$, $geo(p')$, and the degree distribution of the root vertex in EWT). The parameters p' and λ' are chosen so that the expected degree of the root vertex is the same as in EWT.



H

Figure 3.7: The ratio of the giant component of the finite graph model (with potential degree distribution $geo(p)$), random graphs generated by the configuration model (with degree distributions $geo(p')$, and the degree distribution of the root vertex in EWT), and the Erdös-Renyi random graph (with parameter λ'/n). The parameters p' and λ' are chosen so that the degree distribution are the same.

CHAPTER IV

Planted Matching Problem

Notation: The exponential distribution with rate ν is denoted by $Exp(\nu)$. Its cumulative distribution function is $\mathbb{P}[x > t] = \mathrm{e}^{-\nu t}$ and its mean is $1/\nu$. For a Borel space $(S, \mathcal{S})$ consisting of a set S and a σ-algebra $\mathcal{S}$, $\mathcal{P}(S)$ is the set of all Borel probability measures defined on S. We use $\mathbb{Z}$ for the integers, $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and $\mathbb{N}_+ = \{1, 2, \ldots\}$ for the natural numbers, and $\mathbb{R}_+$ for the set of non-negative real numbers. The number of elements of a set A is denoted by $|A|$. Random variables are denoted by capital letters; when we need to refer to a specific realization we sometimes use small letters.

4.1 Introduction

Consider a weighted complete bipartite graph $K_{n,n}$ with an unknown perfect matching M^*, where for each edge e the weight w_e is independently distributed according to P when $e \in M^*$ and Q when $e \notin M^*$. The goal is to recover the "hidden" or "planted" matching M^* from the edge weights.

This problem is inspired by the long history of planted problems in computer science, where an instance of an optimization or constraint satisfaction problem is built around a planted solution in some random way. As we vary the parameters used to generate these instances, such as the size of a hidden clique or the density of communities in the stochastic block model of social networks, we encounter phase transitions in our ability to find this planted solution, exactly or approximately. In an inference problem, the instance corresponds to some noisy observation, such as a data set produced by a generative model, and the planted solution corresponds to the ground truth—the underlying structure we are trying to discover.

More concretely, we are motivated by the problem of tracking moving objects in a video, such as flocks of birds, motile cells, or particles in a fluid. Figure 4.1, taken from [1], shows two frames of such a video, where each particle has moved from its original position by some

amount. Our goal is then to find the most-likely matching between the two frames, assuming some probability distribution of these displacements.

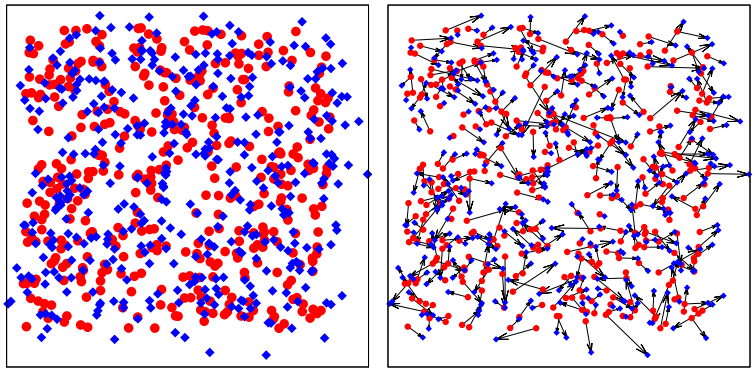


Figure 4.1: On the left, the positions of particles in two frames of a video, with one frame red and the other blue. On the right, an inferred matching, hypothesizing how each particle has moved from one frame to the next. Taken from [1].

For many planted problems such as Hidden Clique (e.g. [80]) or community detection in the stochastic block model (e.g. [81, 82]), there are two types of thresholds: information-theoretic and computational. When these are distinct, the region in between them has the interesting property that finding the planted solution, or at least approximating it better than chance, is information-theoretically possible but (conjecturally) computationally hard. These regions are also known as statistical-computational gaps.

In the planted matching problem, one obvious estimator to try is the minimum weight matching (a.k.a. the linear assignment problem) which can be found in polynomial time. The natural question is then, as a function of the distributions P and Q on the planted and un-planted edges, how much the minimum matching $M_{\min}$ has in common with the planted matching M^*. In general, we define the *overlap* of an estimator M' with M^* as (assuming that $|M'| = n$)

$$\text{overlap}(M^*, M') = 1 - \frac{1}{2n}|M^* \triangle M'| = \frac{1}{n}|M^* \cap M'|\,. \tag{4.1}$$

We say that M' achieves *almost perfect recovery* if $\mathbb{E}[\text{overlap}(M^*, M')] = 1 - o(1)$, or equivalently if $\text{overlap}(M^*, M') = 1 - o(1)$ with high probability. We say that M achieves *partial recovery* if $\mathbb{E}[\text{overlap}(M^*, M')] > 0$ as $n \to \infty$.

Chertkov et al. [1] studied the case where $P = |\mathcal{N}(0, \kappa)|$ is a folded Gaussian and Q is the uniform distribution over $[0, n]$. When $\kappa = O(1)$, the planted edges are competitive with the lightest un-planted edges at each vertex, which have expected weight 1. This suggests a phase transition in this regime, and indeed they predicted a transition from almost perfect

recovery to partial recovery at $\kappa \approx 0.17$ using the cavity method of statistical physics.

We focus on exponential weight distributions, $P = Exp(\lambda)$ and $Q = Exp(1/n)$, so that the planted and un-planted weights have expectation $1/\lambda$ and n respectively. For this family of distributions we obtain exact results, proving a transition from almost perfect recovery to partial recovery at $\lambda = 4$, and determining the expected overlap between M^* and $M_{\min}$ for $\lambda < 4$.

Many of our results apply more generally for any distribution of un-planted edge weights with density $Q'(0) = 1/n$, such as when Q is uniform in the interval $[0, n]$. However, our assumption that the planted weights P are exponentially distributed is important for two reasons. First, it makes possible to exactly analyze a message-passing algorithm, and obtain precise results for the expected overlap. Secondly, it has the pleasing consequence of making $M_{\min}$ the maximum-likelihood estimator for M^*. To see this, note that all $n!$ matchings are equally likely a priori. Let G denote the observed complete bipartite graph with edge weights W. The posterior probability for a given matching M', i.e., $\mathbb{P}[M^* = M' \mid G]$, is proportional to the density

$$\begin{aligned} \mathbb{P}[G \mid M'] = \prod_{e \in M'} P(w_e) \prod_{e \notin M'} Q(w_e) &\propto \prod_{e \in M'} \exp(-(\lambda - 1/n) w_e) \\ &= \exp\left(-(\lambda - 1/n) \sum_{e \in M'} w_e \right) . \end{aligned} \tag{4.2}$$

Thus maximizing the likelihood is equivalent to minimizing the total weight of M'.

Our main results are as follows.

- In Theorem IV.1, we show that the minimum matching $M_{\min}$ achieves almost perfect recovery with high probability whenever $\lambda \geq 4$. This proof is a simple first-moment argument using the expected number of augmenting cycles of each length.

- In Theorem IV.1, we compute the expected overlap between M^* and $M_{\min}$ for $\lambda < 4$, showing that it is an explicit function $\alpha(\lambda)$ given by a system of differential equations.

The proof of Theorem IV.1 takes up most of the chapter. Our proof is inspired by Aldous' analysis of the minimum matching in the un-planted case where all edges have the same weight distribution with $Q'(0) = 1/n$. Using the machinery of local weak convergence [20, 21, 24] Aldous gave a rigorous justification for the cavity method of statistical physics [22], modeling $K_{n,n}$ as a Poisson-weighted infinite tree (PWIT). The cost of matching a vertex with one of its children then follows a probability distribution which is the fixed point of a recursive distributional equation (RDE) which can then be transformed into an ordinary

differential equation (ODE). Solving this ODE proves the conjecture of Mézard and Parisi [22] that the expected cost per vertex is $\zeta(2) = \pi^2/6$.

Generalizing Aldous' analysis to the planted case presents several challenges. We now have an infinite weighted tree we call the *planted PWIT* with two types of edges and two types of vertices, since the partner of a vertex in M^* can be its parent or one of its children. The cost of matching a vertex with a child follows a pair of probability distributions fixed by a system of RDEs, which (when P is exponential) we can transform into a system of four coupled ODEs. We use techniques from dynamical systems to show that this system has a unique solution consistent with its boundary conditions, and express the expected overlap $\alpha(\lambda)$ as an integral involving this solution.

While we focus on the case where P is exponential, we claim that a qualitatively similar picture to Theorems IV.1 and Theorem IV.1 holds for other distributions of planted weights. Indeed, much of our proof applies to any distribution P, including the general framework of a message-passing algorithm on the planted PWIT, and the resulting system of RDEs. Thus while the location of the threshold and the overlap would change, in any one-parameter family of distributions P we expect there to be a phase transition from almost-perfect to partial recovery when P's expectation crosses some critical value.

The organization of the rest of the chapter is as follows: In Section 4.2, we use the first moment method to prove that almost exact recovery is possible for $\lambda \geq 4$. In Section 4.3, we calculate the exact value of $\text{overlap}(M^*, M_{\min})$ for $\lambda < 4$, as number of nodes grows without bound, using local weak framework. The remaining of the chapter focuses on the details of the local weak convergence: in Section 4.4, we discuss the machinery of local weak convergence; in Section 4.5, we define the planted PWIT, and we show the convergence of the planted model to the planted PWIT in local weak sense; in Section 4.6, we characterize the minimum matching on planted PWIT, that satisfies some modest symmetry condition called involution invariance; Finally in Section 4.7, we prove that the minimum matching on bipartite graphs given by the planted model converges to the minimum involution invariant matching on planted PWIT. Some proofs are presented in the Appendix.

4.2 Almost Perfect Recovery for $\lambda \geq 4$

We start by proving that the minimum matching achieves almost perfect recovery whenever $\lambda \geq 4$.

Theorem IV.1. *For any $\lambda \geq 4$, we have $\mathbb{E}[\text{overlap}(M^*, M_{min})] = 1 - o(1)$. In particular, $\mathbb{E}[|M^* \triangle M_{min}|]$ is $O(1)$ for $\lambda > 4$ and $O(\sqrt{n})$ for $\lambda = 4$.*

To prove Theorem IV.1, we use the following Chernoff-like bound on the probability that one Erlang random variable exceeds another. The proof is elementary and appears in Appendix C.1.

Lemma IV.2. *Suppose X_1 is the sum of t independent exponential random variables with rate λ_1, and X_2 is the sum of t independent exponential random variables with rate λ_2 (and independent of X_1) where $\lambda_1 > \lambda_2$. Then*

$$\mathbb{P}[X_1 > X_2] \le \left(\frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2}\right)^t \le \left(\frac{4\lambda_2}{\lambda_1}\right)^t.$$

Proof of Theorem IV.1. An *alternating cycle* is a cycle in $K_{n,n}$ that alternates between planted and un-planted edges, and an *augmenting cycle* is an alternating cycle C where the total weight of its planted edges $C \cap M^*$ exceeds that of its un-planted edges $C \setminus M^*$.

Now recall that the symmetric difference $M^* \triangle M_{\min}$ is a disjoint union of augmenting cycles. The number of cyclic permutations of t things is $(t-1)!$. Thus the number of alternating cycles of length $2t$, i.e., containing t planted edges and t un-planted edges, is at most

$$\binom{n}{t}(t-1)! = \frac{1}{t}\, n^t \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{t-1}{n}\right) \le \frac{1}{t}\, n^t\, \mathrm{e}^{-t(t-1)/(2n)}. \tag{4.3}$$

Applying Lemma IV.2 with $\lambda_1 = \lambda$ and $\lambda_2 = 1/n$, the probability that a given alternating cycle of length $2t$ is augmenting is at most $(4/(\lambda n))^t$.

Now the size of the symmetric difference $|M^* \triangle M_{\min}|$ is at most the total length of all augmenting cycles. By the linearity of expectation, its expectation is bounded by

$$\mathbb{E}[|M^* \triangle M_{\min}|] \le \sum_{t=1}^{n} 2t \left(\frac{4}{\lambda n}\right)^t \frac{1}{t}\, n^t\, \mathrm{e}^{-t(t-1)/(2n)} \le 2\mathrm{e}^{1/2} \sum_{t=1}^{\infty} \left(\frac{4}{\lambda}\right)^t \mathrm{e}^{-t^2/(2n)}.$$

When $\lambda > 4$ the geometric sum $\sum_{t=1}^{\infty}(4/\lambda)^t$ converges, giving $\mathbb{E}[|M^* \triangle M_{\min}|] = O(1)$. When $\lambda = 4$, we have $\sum_{t=1}^{\infty} \mathrm{e}^{-t^2/(2n)} \le \int_0^{\infty} \mathrm{e}^{-t^2/(2n)} \mathrm{d}t = \sqrt{\pi n/2}$, so $\mathbb{E}[|M^* \triangle M_{\min}|] = O(\sqrt{n})$.

To complete the proof, let $\omega(1)$ be any function of n that tends to infinity. By Markov's inequality, with high probability $|M^* \triangle M_{\min}|$ is less than $\omega(1)$ times its expectation, and (4.1) gives w.h.p. $\mathrm{overlap}(M^*, M_{\min}) = 1 - o(1)$. □

We note that when $\lambda > 4$ is sufficiently large we have $\mathbb{E}[|M^* \triangle M_{\min}|] < 1$, implying that $M_{\min}$ achieves perfect recovery, i.e., $M_{\min} = M^*$, with positive probability. We also note that a similar argument shows that, for $\lambda < 4$, the overlap is w.h.p. at least $1 - 2\log\frac{4}{\lambda}$. But this bound is far from tight, and below we give much more precise results.

4.3 Exact Results for the Expected Overlap when $\lambda < 4$

In this section we provide a characterization of the asymptotic overlap of $M_{\min}$, showing exactly how well $M_{\min}$ achieves partial recovery when $\lambda < 4$.

Theorem IV.1. *Suppose $0 < \lambda < 4$ is a fixed constant. Then the expected overlap between the minimum matching and the planted one is*

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}[|M_{min} \cap M^*|] = \alpha(\lambda)\,,$$

where

$$\alpha(\lambda) = 1 - 2\int_0^\infty (1 - F(x))\,(1 - G(x))\,V(x)W(x)\,dx < 1\,, \tag{4.4}$$

and where (F, G, V, W) is the unique solution to the coupled system of ordinary differential equations (4.11)–(4.14) *given below with boundary conditions* (4.15)–(4.16).

Proof. We start by relating the planted model $(\ell_n, K_{n,n})$ where ℓ_n denotes the random edge weights, to a type of weighted infinite tree (ℓ_∞, T_∞) as Aldous did for the un-planted model [20, 21]. This tree corresponds to the neighborhood of a uniformly random vertex, where "local" is defined in terms of shortest path length (sum of edge weights). While $K_{n,n}$ has plenty of short loops, this neighborhood is locally treelike since it is unlikely to have any short loops consisting entirely of low-weight edges.

Starting at a root vertex $\emptyset$, we define the tree T_∞ shown in Figure 4.2. The root has a planted child, i.e., a child connected to it by a planted edge (bold in red), and a series of un-planted children (solid blue). We label these vertices with strings of integers as follows: the root is labeled with the empty string $\emptyset$. Appending 0 to a label indicates the planted child of that parent, if it has one—that is, if its partner in the planted matching is a child rather than its parent. We indicate the un-planted children by appending i for $i \in \{1, 2, 3, \ldots\}$.

We sort the un-planted children of each vertex so that the one labeled with i is the ith lightest, i.e., has the ith lightest edge. Since the distribution of un-planted weights has density $Q'(0) = 1/n$ at 0, these weights are asymptotically described by the arrivals of a Poisson process with rate 1, while the weight of the planted edges are distributed as $Exp(\lambda)$. We call the resulting structure the planted Poisson weighted infinite tree, or planted PWIT, and use ℓ_∞ to denote its edge weights. We define all this formally in Section 4.4 and Section 4.5, and prove that the finite planted model $(\ell_n, K_{n,n})$ weakly converges to (ℓ_∞, T_∞).

Following Aldous [21], in Section 4.6 we then construct a matching $\mathscr{M}_{\infty,\text{opt}}$ on the planted PWIT. Crucially, it has a symmetry property called *involution invariance*, which roughly speaking means that it treats the root just like any other vertex in the tree. We prove that it is the unique involution invariant matching that minimizes the expected cost at the root.

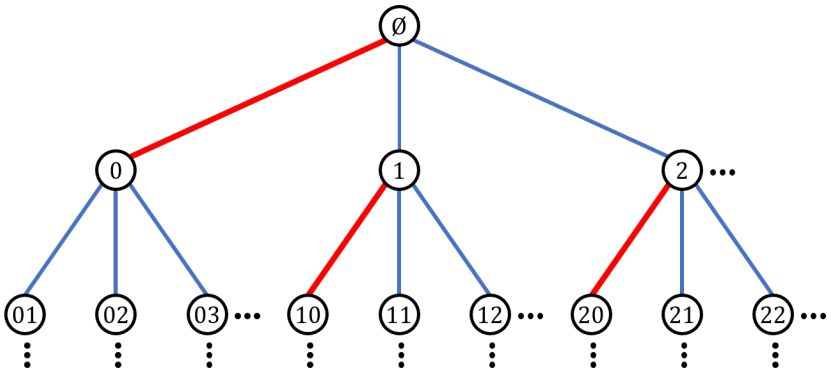


Figure 4.2: The planted Poisson weighted infinite tree (planted PWIT) (ℓ_∞, T_∞) with the labeling scheme described in the text. The bold red edges are planted edges and the solid blue edges are un-planted. The root is the empty string $\emptyset$. Appending 0 to the label of a vertex indicates its planted child, if any, while appending $i \geq 1$ indicates its un-planted child with the ith lightest edge.

We define $\mathscr{M}_{\infty,\text{opt}}$ in terms of the fixed point of a message-passing algorithm that computes, for each vertex v, the cost of matching v with its best possible child. This cost is the minimum over v's children w of the weight of the edge between them, minus the analogous cost for w:

$$X_v = \min_{\text{children } w \text{ of } v} (\ell_\infty(v, w) - X_w)\,.$$

Now suppose that the X_w's are independent, and our goal is to compute the distribution of X_v. Unlike the un-planted model, the two types of children will have their X_w drawn from two different distributions. In the first case, w is v's planted child, and w's children are all un-planted. In the second case, w is an un-planted child of v, and has a planted child of its own. Let X and Y denote the distributions of X_w in these two cases. Then assuming that X_v obeys the appropriate distribution gives the following system of recursive distributional equations (RDE)s:

$$X \stackrel{d}{=} \min\{\zeta_i - Y_i\}_{i=1}^{\infty} \tag{4.5}$$

$$Y \stackrel{d}{=} \min(\eta - X, \{\zeta_i - Y_i\}_{i=1}^{\infty}) \stackrel{d}{=} \min(\eta - X, X')\,, \tag{4.6}$$

where the Y_i's are *assumption*, X and X' are *assumption*, $\eta \sim Exp(\lambda)$, and the ζ_i for $i = 1, 2, \ldots$ are jointly distributed as the arrivals of a Poisson process of rate 1.

In general, analyzing recursive distributional equations (RDEs) is very challenging, since they act on the infinite-dimensional space of probability distributions over the reals. How-

ever, it is sometimes possible to "collapse" them into a finite-dimensional system of ordinary differential equations. For the un-planted case of the random matching problem, Aldous [21] derived a single differential equation whose solution is the logistic distribution. For the planted case, we use a similar approach, but arrive at a more complicated system of four coupled ODEs.

Lemma IV.2. *Let $f_X, f_Y, F_X(x) = \mathbb{P}[X < x]$ and $F_Y(y) = \mathbb{P}[Y < y]$ denote the probability density functions and cumulative distribution functions (CDFs) of X and Y, and let $\bar{F}_X = 1 - F_X$ and $\bar{F}_Y = 1 - F_Y$. If (4.5)–(4.6) have a solution, then*

$$\frac{\mathrm{d}F_X(x)}{\mathrm{d}x} = \bar{F}_X(x)\bar{F}_X(-x)\,\mathbb{E}[F_X(\eta + x)]. \tag{4.7}$$

Proof. First note that (4.6) gives

$$\bar{F}_Y(y) = \bar{F}_X(y)\,\mathbb{E}[F_X(\eta - y)]\,. \tag{4.8}$$

Now the pairs $\{(\zeta_i, Y_i)\}$ in (4.5) form a two-dimensional Poisson point process $\{(z, y)\}$ on $\mathbb{R}_+ \times \mathbb{R}$ with density $f_Y(y)\,\mathrm{d}z\,\mathrm{d}y$. We have $X > x$ if and only if none of these points have $z - y < x$, so

$$\bar{F}_X(x) = \exp\biggl(-\iint_{z-y<x} f_Y(y)\,\mathrm{d}z\,\mathrm{d}y\biggr) = \exp\biggl(-\int_{z=-x}^{\infty} \bar{F}_Y(z)\,\mathrm{d}z\biggr)\,. \tag{4.9}$$

Taking derivatives of both sides of this equation with respect to x and using (4.8) gives

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} = -\frac{\mathrm{d}\bar{F}_X(x)}{\mathrm{d}x} = \bar{F}_X(x)\bar{F}_Y(-x) = \bar{F}_X(x)\bar{F}_X(-x)\,\mathbb{E}[F_X(\eta + x)]\,.$$

□

For the sake of simplicity, we omit the subscript X in $F_X(\cdot)$ in the sequel. Define

$$G(x) = F(-x), \quad V(x) = \mathbb{E}[F(\eta + x)], \quad W(x) = V(-x). \tag{4.10}$$

Lemma IV.3. *When $\eta \sim Exp(\lambda)$, F is a solution to (4.7) if and only if (F, G, V, W) is a solution to the following four-dimensional system of ordinary differential equations (ODEs):*

$$\frac{\mathrm{d}F}{\mathrm{d}x} = (1 - F(x))(1 - G(x))V(x) \tag{4.11}$$

$$\frac{\mathrm{d}V}{\mathrm{d}x} = \lambda(V(x) - F(x)) \tag{4.12}$$

$$\frac{\mathrm{d}G}{\mathrm{d}x} = -(1 - F(x))(1 - G(x))W(x) \tag{4.13}$$
$$\frac{\mathrm{d}W}{\mathrm{d}x} = \lambda(G(x) - W(x)) \tag{4.14}$$

with the boundary conditions

$$\begin{aligned} F(-\infty) = V(-\infty) = G(+\infty) = W(+\infty) = 0 \\ F(+\infty) = V(+\infty) = G(-\infty) = W(-\infty) = 1. \end{aligned} \tag{4.15}$$

and

$$0 \le F, G \le 1, \quad 0 < V, W \le 1\,. \tag{4.16}$$

Proof. For one direction, suppose F is a solution to (4.7). Then (4.11) and (4.13) directly follow from (4.7) by plugging in the definition of (F, G, V, W); thus they hold for any distribution of η. In contrast, (4.12) and (4.14) are derived via integration by parts under the assumption that $\eta \sim Exp(\lambda)$. The conditions (4.15) and (4.16) hold because F must be a valid CDF. Note that $V(x), W(x) > 0$ for any finite x by definition, as η is larger than any fixed threshold with a positive probability.

For the other direction, suppose F is a solution to the system of ODEs (4.11)–(4.14) with conditions (4.15)–(4.16). Clearly F satisfies (4.7). We only need to verify that F is a valid CDF, which is equivalent to checking (1) F is non-decreasing; (2) $F(+\infty) = 1$ and $F(-\infty) = 0$; and (3) F is right continuous. All these properties are satisfied automatically. □

We comment that RDEs can be solved exactly for some other problems with random vertex or edge weights in the case of the exponential distribution, such as maximum weight independence sets and maximum weight matching in sparse random graphs [83–85]. In some cases this is simply because the minimum of a set of exponential random variables is itself an exponential random variable. To our knowledge our situation involving integration by parts is more unusual.

An interesting consequence of (4.11)–(4.16) is the following conservation law:

$$F(x)W(x) + G(x)V(x) - V(x)W(x) = 0, \tag{4.17}$$

which further implies that $V(0) = 2F(0)$.

Surprisingly, we find that the system (4.11)–(4.14) exhibits a sharp phase transition at $\lambda = 4$. On the one hand, when $\lambda \ge 4$, they have no solution consistent with (4.15)–(4.16), corresponding to Theorem IV.1 that we have almost perfect recovery in that case. To see

this, assume that $V(x) \neq 0$ and introduce a new function $U(x)$ as

$$U(x) = \frac{F(x)}{V(x)}\,.$$

Then $U(x)$ is differentiable and satisfies:

$$\frac{\mathrm{d}U}{\mathrm{d}x} = -\lambda U(1-U) + (1-F)(1-G). \tag{4.18}$$

Lemma IV.4. *If $\lambda \geq 4$, then the system of ODEs* (4.11)–(4.14) *with conditions* (4.15)–(4.16) *has no solution.*

Proof. We prove by contradiction. Suppose the system of ODEs (4.11)–(4.14) has a solution satisfying the conditions (4.15)–(4.16). Then $U(x) \to 1$ as $x \to +\infty$ and by the conservation law $U(0) = F(0)/V(0) = 1/2$. Moreover, since $F, G \geq 0$, it follows from (4.18) that

$$\frac{\mathrm{d}U}{\mathrm{d}x} \leq -\lambda U(1-U) + 1.$$

We claim that $U(x) \leq 1/2$ for all $x \geq 0$. Suppose not. By the differentiability of $U(x)$ and $U(0) = 1/2$, there must exist $x_0 > 0$ s.t. $U(x_0) = 1/2$ and $U'(x_0) > 0$, which contradicts $U'(x) \leq 0$ whenever $U(x) = 1/2$ and $\lambda \geq 4$. Thus $U(x) \leq 1/2$ for all x, which contradicts the fact that $U(x) \to 1$ as $x \to \infty$. □

On the other hand, Theorem C.1 in Appendix C.2 proves that for all $\lambda < 4$, there is a unique solution to (4.11)–(4.14) consistent with the conditions (4.15)–(4.16), and hence giving the CDFs of X and Y. The idea hinges on a dynamical fact, namely that the $(U, V, W) = (1, 1, 0)$ is a saddle point, and there is a unique initial condition that approaches it as $x \to \infty$ along its unstable manifold.

Along with Lemma IV.3, this unique solution to the ODEs gives the unique solution to the RDEs (4.5) and (4.6). Moreover Theorem IV.1 in Section 4.7 tells us that the expected overlap of $M_{\min}$ converges to that of $\mathscr{M}_{\infty,\mathrm{opt}}$, which in turn is the probability that the edge weight of a planted edge is less than the cost of matching its endpoints to other vertices:

$$\lim_{n\to\infty} \frac{1}{n}\,\mathbb{E}[|M_{\min} \cap M^*|] = \alpha(\lambda) = \mathbb{P}[\eta < X + X']\,,$$

where X and X' are *assumption* with CDF given by F and $\eta \sim Exp(\lambda)$ is independent.

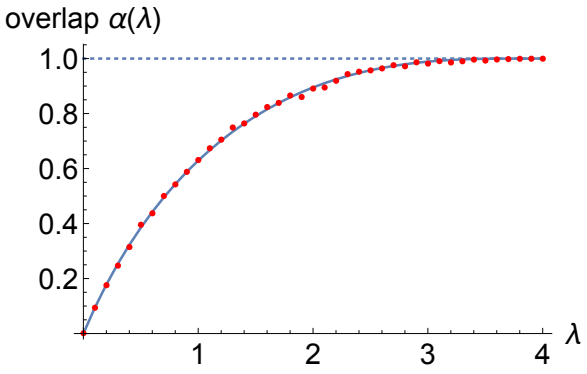


Figure 4.3: The solid blue line is $\alpha(\lambda)$ computed by numerically solving the system of ODEs (4.11)–(4.14). The red dots are computed by finding the minimum matching on bipartite graphs generated by the planted model with $n = 1000$. Each dot is the average of 10 independent trials.

Finally, we compute $\mathbb{P}[\eta < X + \widehat{X}]$ as follows:

$$\begin{aligned}
\mathbb{P}[\eta < X + \widehat{X}] &= 1 - \mathbb{E}_\eta \left[\int_{-\infty}^{+\infty} f(x) F(\eta - x) \, \mathrm{d}x \right] \\
&= 1 - \int_{-\infty}^{+\infty} \frac{\mathrm{d}F(x)}{\mathrm{d}x} \, \mathbb{E}_\eta[F(\eta - x)] \, \mathrm{d}x \\
&= 1 - \int_{-\infty}^{+\infty} (1 - F(x))(1 - G(x)) V(x) W(x) \, \mathrm{d}x \\
&= 1 - 2 \int_{0}^{+\infty} (1 - F(x))(1 - G(x)) V(x) W(x) \, \mathrm{d}x \, .
\end{aligned}$$

where in the last line we used the fact that the integrand is an even function of x.

This completes the proof of Theorem IV.1. To illustrate our results, we plot the function $\alpha(\lambda)$ in Figure 4.3, and compare with experimental results from finite graphs with $n = 1000$. □

We comment that the connection between the finite planted model and the planted PWIT is an integral part of the above argument. We explore this connection in detail in Sections 4.4–4.7. The results presented in these sections are true for any distribution of un-planted edge weights with density $Q'(0) = 1/n$, and any distribution of planted edge weights P.

4.4 Planted Networks and Local Weak Convergence

In this section and the succeeding ones we define the planted Poisson Weighted Infinite Tree (planted PWIT), define a matching $\mathscr{M}_{\infty,\mathrm{opt}}$ on it, prove that it is optimal and unique,

and prove that the minimum weight matching $M_{n,\min}$ on $K_{n,n}$ converges to it in the local weak sense. We follow the strategy of Aldous' celebrated proof of the $\pi^2/6$ conjecture in the un-planted model [20, 21], and in a few places the review article of Aldous and Steele [24]. There are some places where we can simply re-use the steps of that proof, and others where the planted model requires a nontrivial generalization or modification, but throughout we try to keep our proof as self-contained as possible.

In this section we lay out our notation, and formally define local weak convergence. We apologize to the reader in advance for the notational complications they are about to endure: there are far too many superscripts, subscripts, diacritical marks, and general doodads on these symbols. But some level of this seems to be unavoidable if we want to carefully define the various objects and spaces we need to work with.

Our graphs will be simple and undirected unless otherwise specified. Given an undirected graph $G = (V, E)$, a *(perfect) matching* $M \subset E$ is a set of edges where every vertex $v \in V$ is incident to exactly one edge in M. For each v, we refer to the unique v' such that $\{v, v'\} \in M$ as the *partner* to v, and will sometimes denote it as $M(v)$; then $M(M(v)) = v$. In a bipartite graph, a matching defines a one-to-one correspondence between the vertices on the left and those on the right. In a forgivable abuse of notation, will often write $M(v, v') = 1$ if $\{v, v'\} \in M$ and 0 otherwise.

A *rooted graph* $G_\circ = (V, E, \emptyset)$ is a graph $G = (V, E)$ with a distinguished vertex $\emptyset \in V$. The *height* of a vertex $v \in V$ in a rooted graph $G_\circ = (V, E, \emptyset)$ is the shortest-path distance from $\emptyset$ to v, i.e., the minimum number of edges among all paths from $\emptyset$ to v.

A *planted graph* $G = (V, E, M^*)$ is a graph (V, E) together with a *planted matching* $M^* \subset E$. Similarly a *rooted planted graph* $G_\circ = (V, E, M^*, \emptyset)$ is a planted graph with a distinguished vertex $\emptyset$. We refer to the edges in M^* and $E \setminus M^*$ as the planted edges and un-planted edges respectively.

Two planted graphs $G = (V, E, M^*)$ and $G' = (V', E', M^{*\prime})$ are said to be isomorphic if there exists a bijection $\gamma : V \to V'$ such that $\{v_1, v_2\} \in E$ if and only if $\gamma(\{v_1, v_2\}) \coloneqq \{\gamma(v_1), \gamma(v_2)\} \in E'$, and $\{v_1, v_2\} \in M^*$ if and only if $\gamma(\{v_1, v_2\}) = \{\gamma(v_1), \gamma(v_2)\} \in M^{*\prime}$. Thus the isomorphism γ preserves the planted and un-planted edges. A rooted isomorphism from $G_\circ = (V, E, M^*, \emptyset)$ to $G'_\circ = (V', E', M^{*\prime}, \emptyset')$ is an isomorphism between $G = (V, E, M^*)$ and $G' = (V', E', M^{*\prime})$ such that $\gamma(\emptyset) = \emptyset'$.

Next we endow a planted graph with a weight function. A *planted network* $N = (G, \ell)$ is a planted graph $G = (V, E, M^*)$ together with a function $\ell : E \to \mathbb{R}_+$ that assigns weights to the edges. For the sake of brevity, we write $\ell(v, w)$ instead of $\ell(\{v, w\})$.

Now let $K_{n,n} = (V_n, E_n, M_n^*)$ denote a complete bipartite graph together with a planted matching. We use $[n]$ to denote the set of integers $\{1, 2, \ldots, n\}$. We label the vertices on the

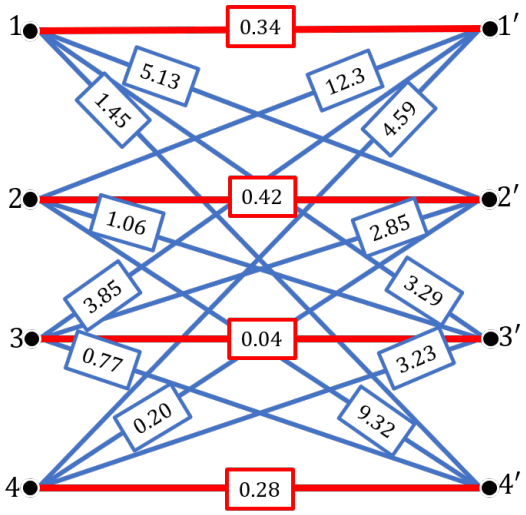


Figure 4.4: A realization of $(K_{4,4}, \ell_4)$ for $n = 4$ and $\lambda = 1$. Red bold edges are in M_n^* (planted edges), and solid blue edges are in $E_n \setminus M_n^*$.

left-hand side of $K_{n,n}$ as $\{1, 2, \ldots, n\}$, and the vertices on the right-hand side as $\{1', 2', \ldots, n'\}$. In a slight abuse of notation, we denote these sets of labels $[n]$ and $[n']$ respectively. Thus $V_n = [n] \cup [n']$, $E_n = \{\{i, j'\} : i \in [n] \text{ and } j' \in [n']\}$, and $M_n^* = \{\{i, i'\} : i \in [n] \text{ and } i' \in [n']\}$.

Let ℓ_n denote a random function that assigns weights to the edges of $K_{n,n}$ as follows: if $e \in M^*$, then $\ell_n(e) \sim Exp(\lambda)$, and if $e \notin M_n^*$ then $\ell_n(e) \sim Exp(1/n)$. We denote the resulting planted network as $(K_{n,n}, \ell_n)$. We denote the minimum matching on $(K_{n,n}, \ell_n)$ as $M_{n,\min}$. Figure 4.4 illustrates a realization of the planted model.

We want to define a metric on planted networks, or rather on their isomorphism classes. Two planted networks $N = (G, \ell)$ and $N' = (G', \ell')$ are isomorphic if there is an isomorphism γ between G and G' that preserves the length of the edges, i.e., if $\ell(v_1, v_2) = \ell'(\gamma(v_1), \gamma(v_2))$. A rooted planted network $N_\circ = (G_\circ, \ell)$ is a rooted planted graph $G_\circ$ together with a weight function ℓ, and we define rooted isomorphism as before. Let $[N_\circ]$ denote the class of rooted planted networks that are isomorphic to $N_\circ$. Henceforth, we use $N_\circ$ to denote a typical member of $[N_\circ]$.

Next, we define a distance function $d_\ell(v, v')$ as the shortest-path weighted distance between vertices but treating planted edges as if they have zero weight. That is,

$$d_\ell(v, v') := \inf_{\text{paths } p \text{ from } v \text{ to } v'} \sum_{e \in p \setminus M^*} \ell(e)\,. \tag{4.19}$$

For any vertex $v \in V$ and any $\rho \in \mathbb{R}_+$, we can consider the neighborhood $N_\rho(v) = \{v' : d_\ell(v, v') \le \rho\}$. A network is *locally finite* if $|N_\rho(v)|$ is finite for all v and all ρ.

Now let $\mathcal{G}_*$ denote the set of all isomorphism classes $[N_\circ]$, where $N_\circ$ ranges over all connected locally finite rooted planted networks. There is a natural way to equip $\mathcal{G}_*$ with a metric. Consider a connected locally finite rooted planted network $N_\circ = (G_\circ, \ell)$. Now, for $\rho \in \mathbb{R}_+$, we can turn the neighborhood $N_\rho(\emptyset)$ into a rooted subgraph $(G_\circ)_\rho$. To be precise, $(G_\circ)_\rho = (V_\rho, E_\rho, M^*_\rho, \emptyset)$ is given as follows:

1. *Vertex set*: $V_\rho = N_\rho(\emptyset) = \{v : d_\ell(v, \emptyset) \le \rho\}$.
2. *Edge set*: $e \in E_\rho$ if $e \in p$ for some path p starting from $\emptyset$ such that $\sum_{e \in p \setminus M^*} \ell(e) \le \rho$.
3. *Planted matching*: $M^*_\rho = M^* \cap E_\rho$.

Given this definition, for any $[N_\circ], [N'_\circ] \in \mathcal{G}_*$ a natural way to define a distance is

$$d([N_\circ], [N'_\circ]) = \frac{1}{R+1},$$

where R is the largest ρ at which the corresponding rooted subnetworks $((G_\circ)_\rho, \ell)$ and $((G'_\circ)_\rho, \ell')$ cease to be approximately isomorphic in the following sense:

$$R = \sup \left\{ \rho \ge 0 : \begin{array}{l} \text{there exists a rooted isomorphism } \gamma_\rho : (G_\circ)_\rho \to \\ (G'_\circ)_\rho \text{ such that } \forall e \in E_\rho,\ |\ell(e) - \ell'(\gamma_\rho(e))| < 1/\rho \end{array} \right\}. \tag{4.20}$$

(Note that this isomorphism is between the rooted subgraphs $(G_\circ)_\rho$ and $(G'_\circ)_\rho$, not the corresponding rooted networks, so it is not required to preserve the weights exactly.) In other words, $N_\circ$ and $N'_\circ$ are close whenever there is a large neighborhood around $\emptyset$ where the edge weights are approximately the same, up to isomorphism. In particular, a continuous function is one that we can approximate arbitrarily well by looking at larger and larger neighborhoods of the root.

Equipped with this distance, we say that a sequence $([N_{n,\circ} = (G_{n,\circ}, \ell_n)])_{n=1}^\infty$ *converges locally to* $[N_{\infty,\circ}]$, and write $[N_{n,\circ}] \xrightarrow{\text{loc}} [N_{\infty,\circ}]$, if the following holds: for all $\rho \in \mathbb{R}_+$ such that $N_{\infty,\circ}$ does not have a vertex at a distance exactly ρ from the root $\emptyset$, there is an $n_\rho \in \mathbb{N}_0$ such that for all $n > n_\rho$ there is a rooted isomorphism $\gamma_{n,\rho} : (G_{n,\circ})_\rho \to (G_{\infty,\circ})_\rho$ such that $\ell_n(\gamma_{n,\rho}^{-1}(e)) \to \ell(e)$ for all $e \in E_\rho$ where E_ρ is defined from $G_{\infty,\circ}$ as above. That is, as n increases, $N_{n,\circ}$ becomes arbitrarily close to $N_{\infty,\circ}$ on arbitrarily large neighborhoods.

It is easy to check that d defines a metric on $\mathcal{G}_*$. Moreover, $\mathcal{G}_*$ equipped with this metric is a Polish space: a complete metric space which is *separable*, i.e., it has a countable dense subset. Hence, we can use the usual tools in the theory of weak convergence to study sequence of probability measures on $\mathcal{G}_*$. More precisely, define $\mathcal{P}(\mathcal{G}_*)$ as the set of all probability measures on $\mathcal{G}_*$ and endow this space with the topology of weak convergence: a

sequence $\mu_n \in \mathcal{P}(\mathcal{G}_*)$ converges weakly to μ_∞, denoted by $\mu_n \xrightarrow{w} \mu$, if for any continuous bounded function $f : \mathcal{G}_* \to \mathbb{R}$,

$$\int_{\mathcal{G}_*} f \, d\mu_n \to \int_{\mathcal{G}_*} f \, d\mu_\infty.$$

Since $\mathcal{G}_*$ is a Polish space, $\mathcal{P}(\mathcal{G}_*)$ is a Polish space as well with the Lévy-Prokhorov metric [86, pp. 394–395, Thm. 11.3.1 and Thm. 11.3.3]. Also, Skorokhod's theorem [87, p. 79, Thm. 4.30] implies that μ_n converges weakly to μ_∞ if an only if there are random variables $[N_{n,\circ}]$ and $[N_{\infty,\circ}]$ defined over $\mathcal{G}_*$ such that $[N_{n,\circ}] \sim \mu_n$, $[N_{\infty,\circ}] \sim \mu_\infty$, and $[N_{n,\circ}] \xrightarrow{\text{loc}} [N_{\infty,\circ}]$ almost surely.

This notion of convergence in $\mathcal{G}_*$ was first discussed by Aldous and Steele in [24]. It is called *local weak convergence* to emphasize the fact that this notion of convergence only informs us about the local properties of measure around the root. We are going to use this framework to study the asymptotics of a sequence of finite planted networks. This methodology is known as *the objective method* [24] and has been used to analyze combinatorial optimization problems in a variety of random structures (e.g. [24, 83–85, 88, 89]).

In order to apply this machinery to random finite planted networks, consider a finite planted network $N = (G, \ell)$. For a vertex $v \in V$, let $N_\circ(v)$ denote the planted network rooted at v consisting of v's connected component. Then we can define a measure $U(N) \in \mathcal{P}(\mathcal{G}_*)$ as follows,

$$U(N) = \frac{1}{|V|} \sum_{v \in V} \delta_{[N_\circ(v)]}, \tag{4.21}$$

where $\delta_{[N_\circ(v)]} \in \mathcal{P}(\mathcal{G}_*)$ is the Dirac measure that assigns 1 to $[N_\circ(v)] \in \mathcal{G}_*$ and 0 to to any other member of $\mathcal{G}_*$. In other words, $U(N)$ is the law of $[N_\circ(\emptyset)]$ where $\emptyset$ is picked uniformly from V. Now, to study the local behavior of a sequence of finite networks $(N_n)_n$, the objective method suggests studying the weak limit of the sequence of measures $(U(N_n))_n$.

Definition IV.1. (**Random Weak Limit**) A sequence of finite planted networks $(N_n)_{n=1}^\infty$ has a random weak limit $\mu \in \mathcal{P}(\mathcal{G}_*)$ if $U(N_n) \xrightarrow{w} \mu$.

If N_n is a random planted network, we replace $U(N_n)$ in the above definition with $\mathbb{E}U(N_n)$, where

$$\mathbb{E}U(N)(A) \coloneqq \mathbb{E}[U(N)(A)] \qquad \text{for all Borel sets } A \subseteq \mathcal{G}_*, \tag{4.22}$$

and the expectation is taken with respect to the randomness of N. For us, in both $(K_{n,n}, \ell_n)$ and the weighted infinite tree $(N_{\infty,\circ}, \ell_\infty)$ we define below, the only source of randomness is the edge weights. It is easy to see that if N is vertex transitive, so that every vertex has the same distribution of neighborhoods, then $\mathbb{E}U(N)$ is the law of $[N(\emptyset)]$ (or of $[N(v)]$ for any

vertex v). In many settings, e.g. sparse Erdős-Rényi graphs, $U(N)$ converges in distribution to $N(\emptyset)$, since averaging over all possible root vertices effectively averages over N as well. But taking the expectation over N as we do here avoids having to prove this.

Not all probability measures $\mu \in \mathcal{P}(\mathcal{G}_*)$ can be random weak limits. The uniform rooting in the measure associated with finite networks implies a modest symmetry property on the asymptotic measure. One necessary condition for a probability measure to be a random weak limit is called *unimodularity* [44].

To define unimodularity, let $\mathcal{G}_{**}$ denote the set of all isomorphism classes $[N_{\circ\circ}]$, where $N_{\circ\circ}$ ranges over all connected locally finite *doubly-rooted* planted networks—that is, networks with an ordered pair of distinguished vertices. We define $\mathcal{G}_{**}$ as the set of equivalence classes under isomorphisms that preserve both roots, and equip it with a metric analogous to (4.20) to make it complete and separable. A continuous function $f([N_{\circ\circ}(\emptyset, v)])$ is then one which we can approximate arbitrarily well by looking at neighborhoods of increasing size that contain both $\emptyset$ and v.

Then we can define unimodularity as follows:

Definition IV.2. (**Unimodularity**) A probability measure $\mu \in \mathcal{P}(\mathcal{G}_*)$ is *unimodular* if for all Borel functions $f : \mathcal{G}_{**} \to \mathbb{R}_+$,

$$\int_{\mathcal{G}_*} \sum_{v \in V} f([N_{\circ\circ}(\emptyset, v)])\, d\mu([N_\circ(\emptyset)]) = \int_{\mathcal{G}_*} \sum_{v \in V} f([N_{\circ\circ}(v, \emptyset)])\, d\mu([N_\circ(\emptyset)])\,. \tag{4.23}$$

In other words, the expectation over μ of the sum (either finite or $+\infty$) over all v of $f([N_{\circ\circ}(\emptyset, v)])$ remains the same if we swap $\emptyset$ and v. Since in a connected graph we can swap any vertex v with $\emptyset$ by a sequence of swaps between $\emptyset$ and its neighbors, each of which moves v closer to the root, this definition is equivalent to one where we restrict f to Borel functions with support on $\{[N_{\circ\circ}(\emptyset, v)] \mid \emptyset \text{ and } v \text{ are neighbors}\}$. With this restriction, unimodularity is known as *involution invariance* [44, Prop. 2.2]:

Lemma IV.3. ***(Involution Invariance)*** *A probability measure $\mu \in \mathcal{P}(\mathcal{G}_*)$ is unimodular if and only if* (4.23) *holds for all Borel functions $f : \mathcal{G}_{**} \to \mathbb{R}_+$ such that $f([N_{\circ\circ}(\emptyset, v)]) = 0$ unless $\{\emptyset, v\} \in E$.*

Aldous in [24] uses another characterization of involution invariance. Given a probability measure $\mu \in \mathcal{P}(\mathcal{G}_*)$, define a measure $\widetilde{\mu}$ on $\mathcal{G}_{**}$ as the product measure of μ and the counting measure on the neighbors of the root, i.e.,

$$\widetilde{\mu}(\cdot) \coloneqq \int_{\mathcal{G}_*} \sum_{v : \{\emptyset, v\} \in E} \mathbf{1}([N_{\circ\circ}(\emptyset, v)] \in \cdot)\, d\mu([N_\circ(\emptyset)]), \tag{4.24}$$

where $\mathbf{1}$ is the indicator function. Like μ, $\widetilde{\mu}$ is a σ-finite measure. Throughout the following sections, we use the $\widetilde{\text{tilde}}$ to distinguish a measure associated with doubly-rooted planted networks from the corresponding measure associated with singly-rooted ones.

Then Aldous' definition of involution invariance in [24] is as follows.

Definition IV.4. (**Involution Invariance, again**) A probability measure $\mu \in \mathcal{P}(\mathcal{G}_*)$ is said to be involution invariant if the induced measure $\widetilde{\mu}$ on $\mathcal{G}_{**}$ is invariant under the involution map $\iota : \mathcal{G}_{**} \to \mathcal{G}_{**}$, i.e.,

$$\widetilde{\mu}(A) = \widetilde{\mu}(\iota^{-1}(A)) \qquad \text{for all Borel sets } A \subseteq \mathcal{G}_{**},$$

where $\iota([N_{\circ\circ}(\emptyset, v)]) = [N_{\circ\circ}(v, \emptyset)]$.

Crucially, unimodularity and involution invariance are preserved under local weak convergence. Any random weak limit satisfies unimodularity and is involution invariant [24, 44] (although the converse is an open problem).

The theory of local weak convergence is a powerful tool for studying random combinatorial problems. In the succeeding sections we will prove a series of propositions analogous to [20,21] showing local weak convergence between our planted model of randomly weighted graphs $K_{n,n}$ and a kind of infinite tree $N_{\infty,\circ}$. These propositions make a rigorous connection between the minimum matching on $K_{n,n}$ and the minimum involution invariant matching $\mathscr{M}_{\infty,\text{opt}}$ on $N_{\infty,\circ}$. Finally, we analyze $\mathscr{M}_{\infty,\text{opt}}$ using the RDEs that we solved with differential equations above.

4.5 PWIT and Planted PWIT

In this section we define the planted Poisson Weighted Infinite Tree, and show that it is the weak limit of the planted model $(K_{n,n}, \ell_n)$.

Let us ignore the planted matching for the moment and assume that $\ell_n(e) \sim Exp(1/n)$ for all $e \in E_n$. The problem of finding the minimum matching on this un-planted network is known as the random assignment problem. Kurtzberg [90] introduced this problem with *i.i.d.* uniform edge lengths on $[0, n]$, and Walkup [91] proved that the expected cost of the minimum matching is bounded and is independent of n. In the succeeding years, many researchers tightened the bound for $\mathbb{E}[X_n]$ (e.g. [92–94]). Using powerful but non-rigorous methods from statistical physics, Meźard and Parisi [22] conjectured that $\mathbb{E}[X_n]$ has the limiting value $\zeta(2) = \pi^2/6$ as $n \to \infty$. Aldous first proved [20] that $\mathbb{E}[X_n]$ indeed has a limit, and then [21] proved the $\pi^2/6$ conjecture, using the local weak convergence approach we follow here.

Other methods have been introduced to study this problem [95–97], including the marvelous fact that for finite n, the expected cost of the minimum matching is the sum of the first n terms of the Riemann series for $\zeta(2)$, namely $1 + 1/4 + 1/9 + \cdots + 1/n^2$. But these methods rely heavily on the specifics of the matching problem, and we will not discuss them here.

As the first step in applying local weak convergence to the planted problem, we are going to identify the weak limit of the planted model according to Definition IV.1: that is, the kind of infinite randomly weighted tree that corresponds to $K_{n,n}$ with weights drawn from our model. To be more precise, we are interested in a probability measure $\mu_\infty \in \mathcal{P}(\mathcal{G}_*)$ that $\mu_n = \mathbb{E}U(N_n) \in \mathcal{P}(\mathcal{G}_*)$ converges to in the local weak sense, where $N_n = (K_{n,n}, \ell_n)$ is the planted model, $U(N_n)$ is the random measure defined in (4.21) by rooting N_n at a uniformly random vertex, and $\mathbb{E}U(N_n)$ is the measure defined in (4.22). Since every neighborhood has the same distribution of neighborhoods in the planted model, the root might as well be at vertex 1, so μ_n is simply the distribution of $[N_{n,\circ}(1)]$. Thus

$$\begin{aligned}\mu_n(A) &= \mathbb{E}U(N_n)(A) \\ &= \frac{1}{2n}\sum_{v\in V_n} \mathbb{E}[\delta_{[N_n(v)]}(A)] = \mathbb{P}[[N_{n,\circ}(1)] \in A], \qquad \text{for all Borel sets } A \subseteq \mathcal{G}_*.\end{aligned} \tag{4.25}$$

In the un-planted model studied by Aldous and others, the weak limit of the random matching problem is the Poisson Weighted Infinite Tree (PWIT). The planted case is similar but more elaborate: the weights of the un-planted edges are Poisson arrivals, but the weights of the planted edges have to be treated separately. We call this the *planted PWIT*, and define it as follows.

We label the vertices V_∞ of the planted PWIT with sequences over $\mathbb{N}_0$, which we denote with bold letters. The root is labeled by the empty sequence $\emptyset$. The children of a vertex $\boldsymbol{i} = (i_1, i_2, \cdots, i_t)$, are $\boldsymbol{i}j \coloneqq (i_1, i_2, \cdots, i_t, j)$ for some $j \in \mathbb{N}_0$, and if $t > 0$ its parent is $\text{parent}(\boldsymbol{i}) \coloneqq (i_1, i_2, \cdots, i_{t-1})$. We say that $\boldsymbol{i}$ belongs to the tth generation of the tree, and write $\text{gen}(\boldsymbol{i}) = t$.

Appending $j \in \{1, 2, \ldots\}$ to $\boldsymbol{i}$ gives the jth non-planted child, i.e., the child with the jth smallest edge weight among the non-planted edges descending from the parent $\boldsymbol{i}$. However, appending $j = 0$ indicates $\boldsymbol{i}$'s planted child if any, i.e., $\boldsymbol{i}$'s partner in the planted matching if its partner is one of its children instead of its parent. Since the planted partner of a planted child is its parent, these sequences never have two consecutive zeroes. (Note that the root has a planted child, so the first entry in the sequence is allowed to be 0.) We denote the set of such sequences of length t as Σ^t, and the set of all finite such sequences as $\Sigma = \bigcup_{t\in\mathbb{N}_0} \Sigma^t$. Thus the edge set is $E_\infty = \{\{\boldsymbol{i}, \boldsymbol{i}j\} \mid j \in \mathbb{N}_0 \text{ and } \boldsymbol{i}, \boldsymbol{i}j \in \Sigma\}$, and the planted matching

$M_\infty^* \subset E_\infty$ consists of the edges $\{\{\boldsymbol{i}, \boldsymbol{i}0\} \mid \boldsymbol{i}, \boldsymbol{i}0 \in \Sigma\}$. Let $T_\infty = (V_\infty = \Sigma, E_\infty, M_\infty^*)$ be the resulting planted tree.

Next we define the random edge weights $\ell_\infty : E_\infty \to \mathbb{R}_+$. The weights of the un-planted edges are distributed just as in the PWIT: that is, for each vertex $\boldsymbol{i} \in V_\infty$, the sequence $\big(\ell_\infty(\boldsymbol{i}, \boldsymbol{i}j)\big)_{j=1,2,\dots}$ is distributed jointly as the arrivals $\zeta_1, \zeta_2, \dots$ of a Poisson process with rate 1. Then we have the planted edges: if $\boldsymbol{i}0 \in \Sigma$, then $\ell_\infty(\boldsymbol{i}, \boldsymbol{i}0) \sim Exp(\lambda)$ independent of everything else. Note that these random weights are independent for different parents $\boldsymbol{i} \in V_\infty$.

Finally, let $N_\infty = (T_\infty, \ell_\infty)$ denote the random planted tree and let $N_{\infty,\circ}$ denote the version of N_∞ rooted at $\emptyset$. We call $N_{\infty,\circ}$ the *planted Poisson Weighted Infinite Tree* or the planted PWIT for short. Its structure is shown in Figure 4.2.

As in Section 4.4, let $[N_{\infty,\circ}]$ denote the equivalence class of $N_{\infty,\circ}$ up to rooted isomorphisms, and denote by $\mu_\infty \in \mathcal{P}(\mathcal{G}_*)$ the probability distribution of $[N_{\infty,\circ}]$ in $\mathcal{G}_*$. The following theorem shows that μ_n converges weakly to μ_∞.

Theorem IV.1. *The planted PWIT is the random weak limit of the planted model on $K_{n,n}$, i.e, $\mu_n \xrightarrow{w} \mu_\infty$.*

Sketch of the proof. Similar to the un-planted case [20, Lemmas 10 and 11] the proof follows from the following steps:

1. Recall that μ_n is the distribution of $[N_{n,\circ}(1)]$. We define an exploration process that explores the vertices of $N_{n,\circ}(1)$ starting from the root vertex 1 in a series of stages. At stage m, this process reveals a tree of depth $m+1$ and maximum arity $m+1$, where the children of each vertex are its m lightest un-planted neighbors (among the remaining vertices) and possibly its planted partner (if its planted partner is not its parent).

2. In the limit $n \to \infty$, the tree explored at each stage is asymptotically the same as a truncated version of the planted PWIT, i.e., the analogous stage-m neighborhood of the root $\emptyset$.

3. For large enough m (independent of n), the ρ-neighborhood $(G_{n,\circ}(1))_\rho$ of vertex 1 in $N_{n,\circ}(1)$ is a subgraph of the explored tree at stage m of the process with high probability. This is due to the fact that, while $K_{n,n}$ has plenty of cycles that are topologically short, it is very unlikely that any short cycle containing vertex 1 consists entirely of low-weight edges.

4. Finally, the result follows by using the Portmanteau Theorem, which enables us to extend the convergence of distributions on local neighborhoods in total variation distance to the desired local weak convergence.

The complete proof is presented in Appendix C.3. □

Since the planted model on $K_{n,n}$ converges to the planted PWIT, we have every reason to believe that—just as Aldous showed for the un-planted problem—the minimum matching on the planted model converges locally weakly to the minimum involution invariant matching on the planted PWIT. We make this statement rigorous in the following sections, following and generalizing arguments in [20, 21, 24].

4.6 Optimal Involution Invariant Matching on Planted PWIT

In this section we define the optimal involution invariant random matching $\mathscr{M}_{\infty,\text{opt}}$ on the planted PWIT—or more precisely, the joint distribution $(\ell_\infty, \mathscr{M}_{\infty,\text{opt}})$. We define it in terms of fixed points of a message-passing algorithm, construct it rigorously on the infinite tree, and prove that it is optimal and unique.

Since the planted PWIT is an infinite tree, the total weight of any matching is infinite. This makes it unclear whether there is a well-defined notion of a minimum-weight matching. But since we are ultimately interested in the cost per vertex of the minimum matching on $K_{n,n}$, we call a random matching $(\ell_\infty, \mathscr{M}_\infty)$ on the planted PWIT optimal if it minimizes the expected cost of the edge incident to the root, $\mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}_\infty(\emptyset))]$.

However, since μ_n is involution invariant and involution invariance is preserved under weak limit, we need to restrict our search for minimum matching to involution invariant matchings. This restriction is crucial. For instance, if we simply want to minimize the expected cost at the root, we could construct a matching as follows, akin to a greedy algorithm: first match the root to its lightest child, i.e., the one with the lowest edge weight. Then match each of its other children with their lightest child, and so on. For this matching, $\mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}_\infty(\emptyset))] = \mathbb{E}[\min(\eta, \zeta)] = 1/(1+\lambda)$ where $\eta \sim Exp(\lambda)$ is the weight of the root's planted edge and $\zeta \sim Exp(1)$ is the weight of its lightest un-planted edge.

However, as pointed out by Aldous for the un-planted model [21, Section 5.1], this matching is not involution invariant. For instance, suppose 1 is $\emptyset$'s lightest child, but that 1 has a descending edge whose weight is even less. In this case, if we swap $\emptyset$ and 1, we won't include the edge $\{1, \emptyset\}$ in the resulting matching. Indeed, in the un-planted case the optimal involution invariant matching has expected weight $\pi^2/6$ per vertex, while this greedy matching has expected weight 1. The lesson here is that the only matchings on the PWIT (or the planted PWIT) that correspond to genuine matchings on $K_{n,n}$ are those that are involution invariant.

Before we proceed, we make a small increment to our formalism. For a network N we define $\mathcal{M}[N]$ as the set of all matchings on N. Now, a *random matching* $(\ell, \mathscr{M})$ on N is a

joint distribution of edge weights and matchings, i.e., a probability measure on $\mathbb{R}_+^E \times \mathcal{M}[N]$ with marginal ℓ on $\mathbb{R}_+^E$. Intuitively, the reader would probably interpret the phrase "random matching" as a measurable function from $\mathbb{R}_+^E \to \mathcal{M}[N]$, assigning a distribution of matchings to each realization of the edge weights ℓ. However, here we follow Aldous by using it to mean a distribution over both ℓ and $\mathscr{M}$. Note that $\mathscr{M}$ may have additional randomness even after conditioning on ℓ; we will eventually learn, however, that $\mathscr{M}_{\infty,\mathrm{opt}}$ does not.

4.6.1 Message-Passing Algorithm

We start by describing a message-passing algorithm on the planted PWIT that we will use to define $\mathscr{M}_{\infty,\mathrm{opt}}$. We have already discussed this, but we do it here in our notation for the infinite tree.

If $(\ell_\infty, \mathscr{M})$ is involution invariant, $\mathbb{E}[\ell_\infty(v, \mathscr{M}(v))]$ is independent of the choice of $v \in V_\infty$. Let us pretend for now that the total weight of the minimum involution invariant matching $\mathscr{M}_{\infty,\mathrm{opt}}$ is finite, and minimize it with a kind of message-passing algorithm.

For a vertex $v \in V_\infty$, let $T_\infty(v)$ denote the subtree consisting of v and its descendants, rooted at v (in particular, $T_\infty(\emptyset) = T_\infty$). Let $\ell_\infty(T_\infty(v))$ and $\ell_\infty(T_\infty(v) \setminus \{v\})$ denote the total weight of the minimum involution invariant matching on $T_\infty(v)$ and $T_\infty(v) \setminus \{v\}$ respectively. The difference between these, which we denote

$$X_v = \ell_\infty(T_\infty(v)) - \ell_\infty(T_\infty(v) \setminus \{v\}), \tag{4.26}$$

is the cost of matching v with one of its children, as opposed to leaving it unmatched (or rather matching it with its parent, without including the cost of that edge). This is the difference between two infinite quantities, but as Aldous and Steele say [24] we should "continue in the brave tradition of physical scientists" and see where it leads. While we have already seen the resulting RDEs in the proof of Theorem IV.1, it will be helpful to restate them here in this more precise notation.

Suppose that in a realization of $\mathscr{M}_{\infty,\mathrm{opt}}$, $\emptyset$ is matched with its child i. Then we have

$$\begin{aligned}\ell_\infty(T_\infty(\emptyset)) &= \ell_\infty(\emptyset, i) + \ell_\infty(T_\infty(i) \setminus \{i\}) + \sum_{j \neq i} \ell_\infty(T_\infty(j)) \\ &= \ell_\infty(\emptyset, i) + \ell_\infty(T_\infty(i) \setminus \{i\}) + \ell_\infty(T_\infty(\emptyset) \setminus \{\emptyset\}) - \ell_\infty(T_\infty(i)).\end{aligned}$$

Rearranging and using (4.26), we have

$$X_\emptyset = \ell_\infty(\emptyset, i) - X_i .$$

We can read this as follows: by matching ø with its child i, we pay the weight $\ell_\infty(\text{ø}, i)$ of the edge between them, but avoid the cost X_i of having i matching with one of its own children. But of course we want to match ø with whichever child minimizes this cost, giving

$$X_\text{ø} = \min_{j \geq 0} \left(\ell_\infty(\text{ø}, j) - X_j \right) . \tag{4.27}$$

Using the same argument, this relation holds for any vertex $v \in V_\infty$. Recalling that the children of v are labeled vj (i.e., v's label sequence with j appended) for $j \geq 0$, we have

$$X_v = \min_{j \geq 0, vj \in V_\infty} \left(\ell_\infty(v, vj) - X_{vj} \right) . \tag{4.28}$$

Now recall that v's planted partner is either its parent or its 0th child. If the former, then this minimization ranges over v's un-planted children vj for $j \geq 1$. If the latter, then it also includes v's planted child $v0$. Let us assume that X_v is drawn from one of two distributions over $\mathbb{R}$, and denote this random variable X in the first case and Y in the second case. We expect these distributions to be fixed if we draw X_{vj} independently for each j, and obtain X_v by applying (4.28). Since v's un-planted children have planted children, but v's planted child (if any) only has un-planted children, we get the following recursive distributional equations (RDEs):

$$X \stackrel{d}{=} \min(\{\zeta_i - Y_i\}_{i=0}^\infty), \tag{4.29}$$

$$Y \stackrel{d}{=} \min(\eta - X, \{\zeta_i - Y_i\}_{i=0}^\infty) \tag{4.30}$$

where X is independent of everything else, $\{Y_i\}_{i=1}^\infty$ and Y are *i.i.d.* and $\{\zeta_i\}_{i=1}^\infty$ are the arrivals of a Poisson process with rate 1, and $\eta \sim Exp(\lambda)$ is the weight $\ell(v, v0)$ of the planted edge—these are the edge weights of the planted PWIT described in Section 4.5.

As we saw in Section 4.3, the distributional equations (4.29)–(4.30) have a unique fixed point supported on $\mathbb{R}$ whenever $\lambda < 4$. Our next task is to turn this heuristic derivation into a rigorous construction of random variables on the planted PWIT, and use them to construct the minimum involution invariant random matching $\mathscr{M}_{\infty,\text{opt}}$.

4.6.2 A Rigorous Construction of $\mathscr{M}_{\infty,\text{opt}}$

The construction is similar to the one in the un-planted model (see [21, Section 4.3] and [24, Section 5.6]). We draw random variables X from a fixed point of the system of recursive distributional equations (4.29)–(4.30). Then we show that these random variables generate an involution invariant random matching, by constructing it (randomly) on finite

neighborhoods, and then extending it to the infinite tree. In the next subsection, we analyze this matching and show that it is optimal.

Define the set of directed edges $\overleftrightarrow{E}_\infty = \overrightarrow{E}_\infty \cup \overleftarrow{E}_\infty$ of T_∞ by assigning two directions to each edge $e \in E_\infty$: for an edge $e = \{\boldsymbol{i}, \boldsymbol{i}j\}$ let $\overrightarrow{e} = (\boldsymbol{i}, \boldsymbol{i}j) \in \overrightarrow{E}_\infty$ denote the edge directed downward, i.e., away from the root, and let $\overleftarrow{e} = (\boldsymbol{i}j, \boldsymbol{i}) \in \overleftarrow{E}_\infty$ denote the edge directed upward toward the root. We use $\overleftrightarrow{e}$ to denote a typical member of $\overleftrightarrow{E}_\infty$. We extend the edge weights to $\overleftrightarrow{E}_\infty$, as $\ell_\infty(\overleftarrow{e}) = \ell_\infty(\overrightarrow{e}) = \ell_\infty(e)$.

The following lemma shows how to define "costs" or "messages" on $\overleftrightarrow{E}_\infty$. It is essentially identical to [21, Lemma 14] and [24, Lemma 5.8], except that we have different distributions of messages on the planted and un-planted edges.

Lemma IV.1. *Let (X_0, Y_0) be a solution of the system of recursive distributional equations (4.29)–(4.30). Jointly with the edge weights ℓ_∞, we can construct a random function $X : \overleftrightarrow{E}_\infty \to \mathbb{R}$ such that the following holds:*

(i) For every edge $(v, w) \in \overleftrightarrow{E}_\infty$ we have

$$X(u, v) = \min_{(v,w) \in E_\infty, w \neq u} (\ell_\infty(v, w) - X(v, w)) . \tag{4.31}$$

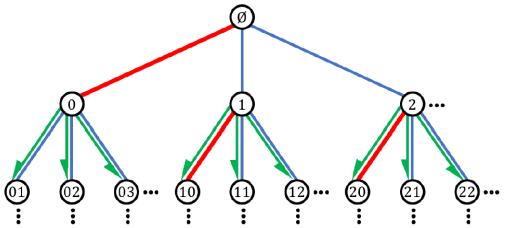
(ii) For every planted edge $e \in M^*_\infty$, $X(\overrightarrow{e})$ and $X(\overleftarrow{e})$ each have the same distribution as X_0.

(iii) For every un-planted edge $e \notin M^*_\infty$, $X(\overrightarrow{e})$ and $X(\overleftarrow{e})$ each have the same distribution as Y_0.

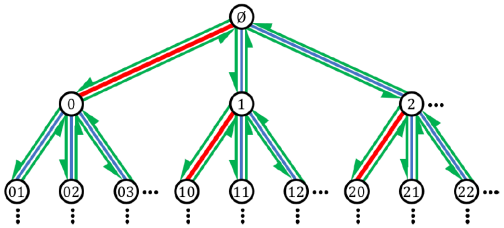
(iv) For every edge $e \in E_\infty$, $X(\overrightarrow{e})$ and $X(\overleftarrow{e})$ are independent.

Proof. The idea is to construct these random variables on the subtree consisting of all edges up to a given depth h. We do this by initially "seeding" them on the downward-pointing edges at that depth, drawing their X independently from the appropriate fixed-point distribution. We then use the message-passing algorithm given by (4.31) to propagate them through this subtree. As with belief propagation on a tree, this propagation consists of one sweep upward to the root, and then one sweep back downward toward the leaves. Finally, we use the Kolmogorov consistency theorem [87, p. 115, Theorem 6.16] to take the limit $h \to \infty$, extending the distribution on these finite-depth subtrees to T_∞.

Formally, let $h \in \mathbb{N}_+$. Let $\overrightarrow{E}_\infty(h)$ and $\overleftarrow{E}_\infty(h)$ respectively denote the set of downward- and upward-directed edges at depth $h - 1$, and let $\overleftrightarrow{E}_\infty(\leq h)$ denote the set of all directed



Green arrows are the members of $\overrightarrow{E}_\infty(2)$.



Green arrows are the members of $\overleftrightarrow{E}_\infty(\leq 2)$.

Figure 4.5: The sets $\overrightarrow{E}_\infty(2)$ and $\overleftrightarrow{E}_\infty(\leq 2)$. Bold red edges are in M^*_∞ and solid blue edges are in $E_\infty \setminus M^*_\infty$.

edges up to depth h:

$$\begin{aligned}
\overrightarrow{E}_\infty(h) &= \{\overrightarrow{e} = (v, vj) : \mathrm{gen}(v) = h-1,\ \{v, vj\} \in E_\infty\},\\
\overleftarrow{E}_\infty(h) &= \{\overleftarrow{e} = (vj, v) : \mathrm{gen}(v) = h-1,\ \{v, vj\} \in E_\infty\},\\
\overleftrightarrow{E}_\infty(\leq h) &= \{\overleftrightarrow{e} = (v, w) : \mathrm{gen}(v), \mathrm{gen}(w) \leq h,\ \{v, w\} \in E_\infty\}.
\end{aligned}$$

In particular, $\overrightarrow{E}_\infty(h)$ is the set of downward-pointing edges at the leaves of the subtree of depth h, and $\overleftrightarrow{E}_\infty(\leq h)$ is the set of all edges, pointing in both directions, within that subtree (see Figure 4.5). Our goal is to define X on $\overleftrightarrow{E}_\infty(\leq h)$.

To initialize the process, for each $\overrightarrow{e} \in \overrightarrow{E}_\infty(h)$ we assign the random variable $X(\overrightarrow{e})$ by drawing independently from X_0 if $e \in M^*_\infty$ and from Y_0 if $e \notin M^*_\infty$. We then use (4.31) recursively to define $\{X(\overrightarrow{e}) : \overrightarrow{e} \in \overrightarrow{E}_\infty(k)\}$ for $k \in \{h-1, h-2, \ldots, 1\}$. Once we have $X(\overrightarrow{e})$ for all edges incident to the root, we use (4.31) to obtain $X(\overleftarrow{e})$ for these edges, i.e., for $\overleftarrow{E}_\infty(1)$. We then move back down the tree, using (4.31) at each level to define $\{X(\overleftarrow{e}) : \overleftarrow{e} \in \overleftarrow{E}_\infty(k)\}$ for $k \in \{1, 2, \ldots, h\}$.

Parts (ii) and (iii) of the lemma follow from the fact that (X_0, Y_0) are fixed points of (4.29)–(4.30). Part (iv) follows from the fact that, for all $e \in E_\infty(\leq h)$, $X(\overrightarrow{e})$ and $X(\overleftarrow{e})$ are determined by disjoint subsets of $\{X(\overrightarrow{e}) : \overrightarrow{e} \in \overrightarrow{E}_\infty(h)\}$ and hence are independent.

Finally, we extend these random variables to the entire planted PWIT. For each finite depth h, the above construction gives a collection of random variables

$$\mathcal{X}_h = \{(\ell_\infty(\overleftrightarrow{e}), X(\overleftrightarrow{e})) : \overleftrightarrow{e} \in \overleftrightarrow{E}_\infty(\leq h)\},$$

that satisfies (i), (ii), (iii), and (iv). Moreover, the marginal distribution of $\mathcal{X}_{h+1}$ restricted to depth h is the same as the distribution of $\mathcal{X}_h$. Now, by the Kolmogorov consistency theorem, there exists a collection of random variables $\mathcal{X}_\infty$ that satisfies (i), (ii), (iii), and (iv), such that the marginal distribution of $\mathcal{X}_\infty$ restricted to depth h is the same as the distribution of

$\mathcal{X}_h$. □

One important implication of Lemma IV.1 is the following corollary.

Corollary IV.2. *Consider the collection of random variables $\mathcal{X}_\infty$ given by Lemma IV.1.*

(i) Let $e = \{\varnothing, 0\}$ denote the planted edge incident to the root. Then $X(\overrightarrow{e})$ and $X(\overleftarrow{e})$ are independent and identically distributed as X_0, and are independent of $\ell_\infty(e)$.

(ii) Suppose we condition on the existence of an un-planted edge $e = \{\varnothing, i\}$ incident to $\varnothing$ with $\ell_\infty(e) = \zeta$. Then $X(\overrightarrow{e})$ and $X(\overleftarrow{e})$ are independent and identically distributed as Y_0.

Proof. Part (i) follows immediately from the construction in Lemma IV.1. Part (ii) follows from the fact that if we condition on the existence of a Poisson arrival at time ζ, the other arrivals are jointly distributed according to the same Poisson process. There is a subtlety here in that it is important to condition on ζ but not on i, since knowing where ζ is in the sorted order of the un-planted weights affects their distribution. On the other hand, if we fix an edge e before doing this sorting, then $X(\overrightarrow{e})$ and $X(\overleftarrow{e})$ are independent of $\ell_\infty(e)$ for both planted and un-planted edges, and we will use this fact below. □

Our next task is to transform the above construction into a random matching (ℓ_∞, $\mathscr{M}_{\infty,\mathrm{opt}}$). There are two ways we might do this. One would be to define a function on V_∞ that yields a proposed partner w for each vertex v. As in (4.31), matching v with w would cost the weight of the edge between them, but remove the cost of having w pair with one of its other neighbors. Minimizing this total cost over all neighbors w (rather than over all but one as in the message-passing algorithm) gives

$$\mathscr{M}_{\infty,\mathrm{opt}}(v) = \operatorname*{arg\,min}_{w:\{v,w\}\in E_\infty} \left(\ell_\infty(v,w) - X(v,w)\right) . \tag{4.32}$$

Since each edge weight $\ell_\infty(v,w)$ is drawn from a continuous distribution, and Corollary IV.2 implies that it is independent of $X(v,w)$, with probability 1 the elements of the set we are minimizing over are distinct and this arg min is well-defined.

Alternately, we could define a mark function on E_∞ as described above, namely the indicator function for the event that an edge e is in the matching. Including e in the matching makes sense if $\ell_\infty(e)$ is less than the cost of matching each of its endpoints to one of their other neighbors. So (abusing notation) this suggests

$$\mathscr{M}_{\infty,\mathrm{opt}}(e) = \begin{cases} 1 & \text{if } \ell_\infty(e) < X(\overrightarrow{e}) + X(\overleftarrow{e}) \\ 0 & \text{otherwise.} \end{cases} \tag{4.33}$$

A priori, there is no guarantee that either of these functions is a matching, or that they agree with each other. The following lemma (which is a reformulation of [24, Lemma 5.9]) gives the good news that they are, and they do.

Lemma IV.3. *The following are equivalent:*

1. $u = \arg\min_{w:\{v,w\}\in E_\infty} (\ell_\infty(v,w) - X(v,w))$
2. $v = \arg\min_{w:\{u,w\}\in E_\infty} (\ell_\infty(u,w) - X(u,w))$
3. $\ell_\infty(u,v) < X(u,v) + X(v,u)$.

Therefore, $u = \mathscr{M}_{\infty,opt}(v)$ if and only if $v = \mathscr{M}_{\infty,opt}(u)$ (with $\mathscr{M}_{\infty,opt}$ defined as in (4.32)), and these are equivalent to $\mathscr{M}_{\infty,opt}(u,v) = 1$ (with $\mathscr{M}_{\infty,opt}$ defined as in (4.33)).

Proof. By (4.31), condition (1) holds if and only if

$$\ell_\infty(v,u) - X(v,u) < \operatorname*{arg\,min}_{w:\{v,w\}\in E_\infty, w\neq u} (\ell_\infty(v,w) - X(v,w)) = X(u,v)\,.$$

Rearranging gives (3), so (1) and (3) are equivalent. Since (3) is symmetric with respect to swapping u and v, (2) and (3) are also equivalent. □

Finally, given the symmetric dependency of $\mathscr{M}_{\infty,\mathrm{opt}}(e)$ on the values of $X(\overrightarrow{e})$ and $X(\overleftarrow{e})$, it is intuitive that the random matching $(\ell_\infty, \mathscr{M}_{\infty,\mathrm{opt}})$ is involution invariant. The following lemma corresponds to [21, Lemma 24] in the un-planted case, but defining the involutions in a way that preserves the (un)planted edges takes a little more work. We give the proof in Appendix C.4.

Proposition IV.4. *The random matching $(\ell_\infty, \mathscr{M}_{\infty,opt})$ is involution invariant.*

4.6.3 Optimality of $\mathscr{M}_{\infty,\mathbf{opt}}$

Now that we have constructed $(\ell_\infty, \mathscr{M}_{\infty,\mathrm{opt}})$, it is time to prove that $(\ell_\infty, \mathscr{M}_{\infty,\mathrm{opt}})$ is *the* minimum involution invariant random matching. The steps we take to prove this claim are mostly the same as in [21, Sections 4.4 and 4.5], but a few details differ in the planted model, so for the sake of completeness and consistency with our notation we give a self-contained proof.

As the first step, we are going to prove that $(\ell_\infty, \mathscr{M}_{\infty,\mathrm{opt}})$ is a minimum involution invariant matching: that is, it achieves the minimum expected length at the root. We follow the discussion at the beginning of Section 4.5 in [21].

Proposition IV.5. *Let $(\ell_\infty, \mathscr{M}'_\infty)$ be an involution invariant random matching on the planted PWIT. Then $\mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}'_\infty(\emptyset))] \geq \mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}_{\infty,opt}(\emptyset))]$.*

Proof. Note that in addition to depending on the edge weights ℓ_∞, $\mathscr{M}'_\infty$ might also have additional randomness. However, we can always couple $(\ell_\infty, \mathscr{M}'_\infty)$ and $(\ell_\infty, \mathscr{M}_{\infty,\mathrm{opt}})$ so that if we condition on ℓ_∞ then $\mathscr{M}'_\infty$ and $\mathscr{M}_{\infty,\mathrm{opt}}$ are independent. Let A be the event that $\mathscr{M}'_\infty(\emptyset) \neq \mathscr{M}_{\infty,\mathrm{opt}}(\emptyset)$, and assume without loss of generality that $\mathbb{P}[A] > 0$.

Conditioned on A, there is a doubly-infinite alternating path that passes through the root $\emptyset$, alternating between edges in $\mathscr{M}'_\infty$ and $\mathscr{M}_{\infty,\mathrm{opt}}$. That is to say, there is a doubly-infinite sequence of distinct vertices $\cdots, v_{-2}, v_{-1}, v_0, v_1, v_2, \cdots$ where $v_0 = \emptyset$, $v_1 = \mathscr{M}_{\infty,\mathrm{opt}}(\emptyset)$, and $v_{-1} = \mathscr{M}'_\infty(\emptyset)$, and where for all even integers m we have $\mathscr{M}_{\infty,\mathrm{opt}}(v_m) = v_{m+1}$ and $\mathscr{M}'_\infty(v_m) = v_{m-1}$.

By the construction of $\mathscr{M}_{\infty,\mathrm{opt}}$, we know that v_1 achieves the minimum in Equation (4.32):

$$\ell_\infty(v_0, v_1) - X(v_0, v_1) = \min_{w:\{v_0,w\}\in E_\infty} (\ell_\infty(v_0, w) - X(v_0, w)) \,. \tag{4.34}$$

We also have the message-passing equation (4.31) for $X(v_{-1}, v_0)$,

$$X(v_{-1}, v_0) = \min_{w:\{v_0,w\}\in E_\infty, w\neq v_{-1}} (\ell_\infty(v_0, w) - X(v_0, w)) \,. \tag{4.35}$$

The right-hand sides of (4.34) and (4.35) are the same except that v_{-1} is excluded in (4.35). But since the minimum is achieved by v_1, excluding v_{-1} makes no difference, and the right-hand sides are equal. Rearranging gives

$$\ell_\infty(v_0, v_1) = X(v_0, v_1) + X(v_{-1}, v_0) \,. \tag{4.36}$$

On the other hand, (4.31) also implies $X(u, v) \leq \ell_\infty(v, w) - X(v, w)$ for any u, v, w where u and w are distinct neighbors of v, and in particular

$$X(v_1, v_0) \leq \ell_\infty(v_0, v_{-1}) - X(v_0, v_{-1}) \,. \tag{4.37}$$

Now, using (4.36), the expected difference in the length at the root is

$$\begin{aligned} \mathbb{E}[(\ell_\infty(\emptyset, \mathscr{M}'_\infty(\emptyset)) - \ell_\infty(\emptyset, \mathscr{M}_{\infty,\mathrm{opt}}(\emptyset)))] &= \mathbb{E}[(\ell_\infty(v_0, v_{-1}) - \ell_\infty(v_0, v_1))\,\mathbf{1}_A] \\ &= \mathbb{E}[(\ell_\infty(v_0, v_{-1}) - X(v_0, v_1) - X(v_{-1}, v_0))\,\mathbf{1}_A] \,. \end{aligned} \tag{4.38}$$

Now we use the fact that $\mathscr{M}_{\infty,\mathrm{opt}}$ and $\mathscr{M}'_\infty$ are both involution invariant. There is a subtlety

here in that conditioning on A breaks involution invariance, since it requires $\mathscr{M}_{\infty,\mathrm{opt}}$ and $\mathscr{M}'_\infty$ to differ at the root specifically. However, the involutions that swap v_0 with v_1 or with v_{-1} maintain this conditioning, since $\mathscr{M}_{\infty,\mathrm{opt}}$ and $\mathscr{M}'_\infty$ differ at these vertices as well. It follows that $X(v_0, v_1)$ and $X(v_1, v_0)$ have the same conditional distribution and hence the same conditional expectation, and similarly for $X(v_0, v_{-1})$ and $X(v_{-1}, v_0)$. Then (4.38) becomes

$$\mathbb{E}[(\ell_\infty(\emptyset, \mathscr{M}'_\infty(\emptyset)) - \ell_\infty(\emptyset, \mathscr{M}_{\infty,\mathrm{opt}}(\emptyset)))] = \mathbb{E}[(\ell_\infty(v_0, v_{-1}) - X(v_1, v_0) - X(v_0, v_{-1}))\,\mathbf{1}_A]\,, \tag{4.39}$$

which is greater than or equal to zero by (4.37). □

Even given Proposition IV.5, it is still possible a priori that there might be a random involution invariant matching $(\ell_\infty, \mathscr{M}'_\infty)$ with the same expected length at the root as $(\ell_\infty, \mathscr{M}_{\infty,\mathrm{opt}})$. If we were simply trying to calculate the expected length of the minimum matching, this would not be an issue. But our object is the overlap, not the length. If there are two minimal matchings with the same length but different overlap, it would not be clear which is the weak limit of the minimum matching on $K_{n,n}$.

Happily, we can follow a path similar to [21, Section 4.4 and 4.5] to show that $\mathscr{M}_{\infty,\mathrm{opt}}$ is unique, making the inequality in Proposition IV.5 strict. The following is essentially Proposition 18 of [21].

Proposition IV.6. *Let $(\ell_\infty, \mathscr{M}'_\infty)$ be an involution invariant random matching on the planted PWIT. If $\mathbb{P}[\mathscr{M}'_\infty(\emptyset) \neq \mathscr{M}_{\infty,opt}(\emptyset)] > 0$ then $\mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}'_\infty(\emptyset))] > \mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}_{\infty,opt}(\emptyset))]$.*

Proof. For sake of contradiction, assume there is an involution invariant random matching $(\ell_\infty, \mathscr{M}'_\infty)$ such that $\mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}'_\infty(\emptyset))] = \mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}_{\infty,\mathrm{opt}}(\emptyset))]$. By the proof of Proposition IV.5, we have $\mathbb{E}[D\mathbf{1}_A] = 0$ where

$$D = \ell_\infty(v_0, v_{-1}) - X(v_1, v_0) - X(v_0, v_{-1}) \geq 0\,,$$

and where A is again the event $\{\mathscr{M}'_\infty(\emptyset) \neq \mathscr{M}_{\infty,\mathrm{opt}}(\emptyset)\}$, and where the inequality $D \geq 0$ is given by Equation (4.37). Therefore, conditioned on A, almost surely

$$X(v_1, v_0) = \ell_\infty(v_0, v_{-1}) - X(v_0, v_{-1})\,. \tag{4.40}$$

Now recall that v_1 achieves the minimum, over all w in v_0's neighborhood, of $\ell_\infty(v_0, w) - X(v_0, w)$. By Equation (4.31), $X(v_1, v_0)$ is the minimum of this same quantity over all $w \neq v_1$. But this is the second minimum, i.e., the second-smallest value, and (4.40) implies

$$v_{-1} = \arg\min_i{}^{[2]}(\ell_\infty(\emptyset, i) - X(\emptyset, i))\,, \tag{4.41}$$

where $\min^{[2]}$ denotes the second minimum. Thus the following holds almost surely: either $\mathscr{M}'_\infty$ agrees with $\mathscr{M}_{\infty,\mathrm{opt}}$ at the root, or it matches the root with the second minimum of $\ell_\infty(\emptyset, i) - X(\emptyset, i)$ rather than the minimum. That is, without conditioning on A,

$$\mathbb{P}\left(\mathscr{M}'_\infty(\emptyset) \in \left\{\arg\min_i(\ell_\infty(\emptyset, i) - X(\emptyset, i)) \text{ or } \arg\min_i{}^{[2]}(\ell_\infty(\emptyset, i) - X(\emptyset, i))\right\}\right) = 1\,.$$

Since $(\ell_\infty, \mathscr{M}'_\infty)$ is involution invariant, the same relation holds for each vertex $v \in V_\infty$, i.e.,

$$\mathbb{P}\left(\mathscr{M}'_\infty(v) \in \left\{\arg\min_{w:\{w,v\}\in E_\infty}(\ell_\infty(v, w) - X(v, w)) \text{ or } \arg\min_{w:\{w,v\}\in E_\infty}{}^{[2]}(\ell_\infty(v, w) - X(v, w))\right\}\right) = 1\,. \tag{4.42}$$

Thus any matching with the same expected length as $\mathscr{M}_{\infty,\mathrm{opt}}$ must, almost surely at almost all vertices v, match v with its best or second-best partner according to $\ell_\infty(v, w) - X(v, w)$.

Surprisingly, no involution-invariant matching can choose the second-best partner with nonzero probability. The following proposition shows that (4.42) cannot hold unless $\mathscr{M}'_\infty = \mathscr{M}_{\infty,\mathrm{opt}}$ almost surely.

Proposition IV.7 (Proposition 20 of [21])**.** *The only involution invariant random matching that satisfies* (4.42) *is* $\mathscr{M}_{\infty,opt}$*.*

Proof. The reader might be wondering why we can't simply assign everyone to their second-best partner. But recall the key fact from Lemma IV.3 that if

$$\mathscr{M}_{\infty,\mathrm{opt}}(v) = \arg\min_{w:\{v,w\}\in E_\infty}(\ell_\infty(v, w) - X(v, w)),$$

then $\mathscr{M}_{\infty,\mathrm{opt}}(\mathscr{M}_{\infty,\mathrm{opt}}(v)) = v$ and $\{\{v, \mathscr{M}_{\infty,\mathrm{opt}}(v)\} : v \in V_\infty\}$ is indeed a matching. The problem is that this fact does not generally hold if we replace arg min with $\arg\min^{[2]}$.

If $\mathscr{M}'_\infty$ and $\mathscr{M}_{\infty,\mathrm{opt}}$ differ anywhere with positive probability, then by involution invariance they differ at the root with positive probability. In that case, as before, there is a doubly-infinite alternating path from the root to infinity. Thus once $\mathscr{M}'_\infty$ matches the root with its second-best partner, it must keep doing this forever on that path. But in order for $\mathscr{M}'_\infty$ to be involution invariant, it must make the same choices if we follow the path in reverse, and so each vertex on this path must be the second-best partner of its second-best partner. We will see that the probability that this is true on every step of the path, all the way to infinity, is zero.

Let $\cdots, v_{-2}, v_{-1}, v_0, v_1, v_2, \cdots$ be the alternating path defined as follows. First let $v_0 = \emptyset$. To define v_t for $t > 0$, we extend the path by alternately apply the best and second-best

rules,

$$v_{t+1} = \begin{cases} \arg\min_{u:\{u,v_t\}\in E_\infty} (\ell_\infty(v_t,u) - X(v_t,u)) & \text{if } t \text{ is even} \\ \arg\min^{[2]}_{u:\{u,v_t\}\in E_\infty} (\ell_\infty(v_t,u) - X(v_t,u)) & \text{if } t \text{ is odd} \end{cases}$$

Similarly, for $t < 0$ we extend the path backwards,

$$v_{t-1} = \begin{cases} \arg\min^{[2]}_{u:\{u,v_t\}\in E_\infty} (\ell_\infty(v_t,u) - X(v_t,u)) & \text{if } t \text{ is even} \\ \arg\min_{u:\{u,v_t\}\in E_\infty} (\ell_\infty(v_t,u) - X(v_t,u)) & \text{if } t \text{ is odd} \end{cases}$$

In particular, $v_1 = \mathscr{M}_{\infty,\mathrm{opt}}(\emptyset)$ and (if A holds) $v_{-1} = \mathscr{M}'_\infty(\emptyset)$.

Now for each odd integer t, define the event B_t that v_t and v_{t+1} are the second-best partners of each other. For odd $t > 0$ we can write

$$B_t = \left\{ v_t = \operatorname*{arg\,min}^{[2]}_{u:\{u,v_{t+1}\}\in E_\infty} (\ell_\infty(v_{t+1},u) - X(v_{t+1},u)) \right\},$$

and for odd $t < 0$,

$$B_t = \left\{ v_{t+1} = \operatorname*{arg\,min}^{[2]}_{u:\{u,v_t\}\in E_\infty} (\ell_\infty(v_t,u) - X(v_t,u)) \right\}.$$

As discussed above, since $\mathscr{M}'_\infty$ is involution invariant A implies B_t, in particular, for all $t = 1, 3, 5, \ldots$. Thus

$$A \subset \overline{B}_\infty \coloneqq \bigcap_{t=1,3,5,\ldots}^{\infty} B_t \,.$$

Writing $\overline{B}_t = \bigcap_{t'=1,3,5,\ldots}^{t} B_{t'}$, this implies

$$\mathbb{P}[A] \le \mathbb{P}[\overline{B}_\infty] = \prod_{t=1,3,5,\ldots} \mathbb{P}[B_{t+2} \mid \overline{B}_t] = \prod_{t=1,3,5,\ldots} \frac{\mathbb{P}[\overline{B}_{t+2}]}{\mathbb{P}[\overline{B}_t]} \,.$$

and so

$$\text{if } \mathbb{P}[A] > 0 \text{ then } \lim_{t\to\infty,\, t \text{ odd}} \frac{\mathbb{P}[\overline{B}_{t+2}]}{\mathbb{P}[\overline{B}_t]} = 1. \tag{4.43}$$

Now we use involution invariance again. If we root the planted PWIT at v_2 instead of v_0, sliding the alternating path two steps to the left, the event $\overline{B}_{t+2}$ becomes the event $B_{-1} \cap \overline{B}_t$ (and A still holds). By involution invariance the probability of these two events is the same, so

$$\frac{\mathbb{P}[\overline{B}_{t+2}]}{\mathbb{P}[\overline{B}_t]} = \frac{\mathbb{P}[B_{-1} \cap \overline{B}_t]}{\mathbb{P}[\overline{B}_t]} = \mathbb{P}[B_{-1} \mid \overline{B}_t] \,.$$

By continuity of probability measure, if $\mathbb{P}[\overline{B}_\infty] > 0$ — which holds if $\mathbb{P}[A] > 0$ — we also have

$$\lim_{t\to\infty} \mathbb{P}[B_{-1} \cap \overline{B}_t] = \mathbb{P}[B_{-1} \cap \overline{B}_\infty] \quad \text{and} \quad \lim_{t\to\infty} \mathbb{P}[\overline{B}_t] = \mathbb{P}[\overline{B}_\infty]\,,$$

in which case

$$\lim_{t\to\infty} \mathbb{P}[B_{-1} \mid \overline{B}_t] = \mathbb{P}[B_{-1} \mid \overline{B}_\infty]\,.$$

Thus (4.43) demands that this conditional probability is 1. But the following lemma, which generalizes Lemma 22 of [21] to the planted case, shows that this is not so.

Lemma IV.8. *If* $\mathbb{P}[\overline{B}_\infty] > 0$ *then* $\mathbb{P}[B_{-1} \mid \overline{B}_\infty] < 1$.

Proof. As in [21, Remark on p. 402], the idea is that B_{-1} only depends on what happens on the "leftward" branch of the alternating path, $v_0, v_{-1}, v_{-2}, \ldots$, while $\overline{B}_\infty$ depends only on the "rightward" branch $v_0, v_1, v_2, \ldots$ For the details, see Appendix C.5. □

Lemma IV.8 implies that $P(A) = 0$, and by the discussion above that $\mathscr{M}'_\infty = \mathscr{M}_{\infty,\mathrm{opt}}$ almost surely. This completes the proof of Proposition IV.7... □

...which completes the proof of Proposition IV.6. □

An immediate corollary of Proposition IV.6 is the following.

Corollary IV.9. *In the minimum involution invariant random matching* $(\ell_\infty, \mathscr{M}_{\infty,opt})$, $\mathscr{M}_{\infty,opt}$ *is a function of the edge lengths* ℓ_∞. *That is to say, given a realization of* $(\ell_\infty(e), e \in E_\infty)$, $\mathscr{M}_{\infty,opt}$ *is a fixed matching on the planted PWIT.*

Proof. Consider a coupling $(\ell_\infty, \mathscr{M}_{\infty,\mathrm{opt}}, \mathscr{M}'_{\infty,\mathrm{opt}})$ such that conditioned on $(\ell_\infty(e), e \in E_\infty)$, $\mathscr{M}_{\infty,\mathrm{opt}}$ and $\mathscr{M}'_{\infty,\mathrm{opt}}$ are *i.i.d.*. Then, by Proposition IV.6 we have $\mathscr{M}_{\infty,\mathrm{opt}} = \mathscr{M}'_{\infty,\mathrm{opt}}$ almost surely. □

In other words, $\mathscr{M}_{\infty,\mathrm{opt}}$ does not have any additional randomness besides its dependence on ℓ_∞. This was left as an open question for the un-planted case in [21, Remark (d)], although we claim that that paper in fact resolved it! As later stated in [24], this implies that if we use the construction of Section 4.6.2 to define random variables X on neighborhoods of depth h, then (conditioning on ℓ_∞) the random matching defined by these variables becomes concentrated around a single matching as $h \to \infty$.

This does not quite imply that the messages X on the directed edges of the planted PWIT are determined by ℓ_∞. This was shown for the un-planted case by Bandyopadhyay using the concept of endogeny [98]. We believe endogeny holds for the planted case, but we leave this as an open question. In any case, as long as the system of recursive distributional

equations (4.29)–(4.30) has a solution supported on $\mathbb{R}$, whether it is unique or not, the minimum involution invariant random matching $\mathscr{M}_{\infty,\mathrm{opt}}$ is uniquely defined. Therefore, whenever we focus on a realization of ℓ_∞, there is no need to call $\mathscr{M}_{\infty,\mathrm{opt}}$ a random matching.

4.6.4 Uniqueness of The Solution of RDEs

Recall from Section 4.6.2 that $\mathscr{M}_{\infty,\mathrm{opt}}$ is defined by drawing messages at the boundary of neighborhoods of increasing size from a fixed point of the RDEs (4.29)–(4.30), propagating these messages throughout the neighborhood, and then including edges (u, v) whose weights $\ell_\infty(u, v)$ are less than the sum of their messages $X(u, v) + X(v, u)$.

However, Corollary IV.9 shows that $\mathscr{M}_{\infty,\mathrm{opt}}$ is a function of the weights ℓ_∞. As we commented there, this doesn't quite imply that the messages X are also functions of ℓ_∞. However, Corollary IV.9 imposes strong conditions on the possible solutions of the RDEs. Specifically, if the RDEs have more than one solution, then each one must somehow result in the same matching $\mathscr{M}_{\infty,\mathrm{opt}}$ given the edge weights. In this section, we show that this implies that the fixed point is indeed unique. This provides an interesting counterpart to the dynamical proof of uniqueness given in Theorem C.1.

First we show that any solution has a well-defined moment generating function in a neighborhood of the origin.

Lemma IV.10. *Let (X, Y) be a solution of the system of recursive distributional equations (4.29)–(4.30) supported on $\mathbb{R}$. Then the random variable X has a finite moment generating function $\mathbb{E}[\mathrm{e}^{\mu X}]$ for μ in an open neighborhood of 0.*

Proof. Recall that F_X and F_Y denote the cumulative distribution functions of X and Y respectively, and $\bar{F}_X$ and $\bar{F}_Y$ denote their complements. On the one hand, by (4.9), for all $x > 0$ we have

$$\bar{F}_X(x) = \exp\biggl(-\int_{z=-x}^{\infty} \bar{F}_Y(z)\,\mathrm{d}z\biggr) \le \exp\bigl(-x\bar{F}_Y(0)\bigr)\,.$$

On the other hand, for every $x_0 > 0$, Lemma IV.2 gives

$$f_X(x_0) = \bar{F}_X(x_0)\bar{F}_X(-x_0)\,\mathbb{E}[F_X(\eta + x_0)] \ge \bar{F}_X(-x_0)\bar{F}_X(x_0)\,\mathbb{E}[F_X(\eta - x_0)] = f_X(-x_0)\,,$$

where the inequality follows by the fact that $F_X(\eta - x_0) \le F_X(\eta + x_0)$ for all η. Hence,

$$\mathbb{P}[X < -x] \le \mathbb{P}[X > x] \le \exp\bigl(-x\bar{F}_Y(0)\bigr)\,,$$

and $\bar{F}_X(0) \geq 1/2$. Then (4.8) implies

$$\bar{F}_Y(0) = \bar{F}_X(0)\,\mathbb{E}[F_X(\eta)] \geq \frac{1}{2}\,\mathbb{E}[F_X(\eta)] > 0\,.$$

where the last inequality holds because η can be arbitrarily large and X is supported on $\mathbb{R}$. The result now follows by simple algebra. If $0 \leq \mu < \bar{F}_Y(0)$ we have

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{\mu X}] &= \int_0^\infty \mathbb{P}[\mathrm{e}^{\mu X} > s]\,\mathrm{d}s \leq 1 + \int_1^\infty \mathbb{P}[\mathrm{e}^{\mu X} > s]\,\mathrm{d}s \\ &= 1 + \int_1^\infty \mathbb{P}\left[X > \frac{\ln s}{\mu}\right]\,\mathrm{d}s \leq 1 + \int_1^\infty s^{-\bar{F}_Y(0)/\mu}\,\mathrm{d}s < \infty\,,\end{aligned}$$

and the proof for $-\bar{F}_Y(0) < \mu \leq 0$ is similar. Hence $\mathbb{E}[\mathrm{e}^{\mu X}] < \infty$ for $\mu \in (-\bar{F}_Y(0), \bar{F}_Y(0))$. □

Now recall that by Lemmas IV.1 and Lemma IV.3, given $\ell(u,v) = x$, the probability that $(u,v) \in \mathscr{M}_{\infty,\mathrm{opt}}$ equals $\mathbb{P}[X + X' > x]$ where $X = X(u,v)$ and $X' = X(v,u)$ are *i.i.d.* copies of the random variable X. If the RDEs have two distinct solutions (X_1, Y_1) and (X_2, Y_2), Corollary IV.9 implies that $\mathbb{P}[X_1 + X_1' > x] = \mathbb{P}[X_2 + X_2' > x]$ for all x, so that $X_1 + X_1'$ and $X_2 + X_2'$ have the same distribution. But since

$$\mathbb{E}[\mathrm{e}^{\mu(X_1 + X_1')}] = \left(\mathbb{E}[\mathrm{e}^{\mu X_1}]\right)^2\,,$$

and similarly for X_2, this implies that X_1 and X_2 have the same moment generating function, which by Lemm IV.10 is well-defined in a neighborhood of the origin. It follows that X_1 and X_2 have the same distribution [99, Theorem 1]. Using (4.6), Y_1 and Y_2 are equidistributed as well, and we have proved the following theorem:

Theorem IV.11. *Assume the system of recursive distributional equations* (4.29)–(4.30) *has a solution supported on* $\mathbb{R}$. *Then such solution is unique.*

4.7 Convergence of the Minimum Matching on $(K_{n,n}, \ell_n)$ to $\mathscr{M}_{\infty,\mathrm{opt}}$

At this point we have constructed $(\ell_\infty, \mathscr{M}_{\infty,\mathrm{opt}})$ and shown that it is the unique involution invariant matching on the planted PWIT that minimizes the weight at the root. It is finally time to show that the minimum matching $(\ell_n, M_{n,\min})$ on our original planted model on $K_{n,n}$ converges to $(\ell_\infty, \mathscr{M}_{\infty,\mathrm{opt}})$ in the local weak sense. This implies that these two objects have the same joint distribution of edge weights, and which edges they include in the matching, on neighborhoods of any finite radius. In particular, they have the same expected overlap—which is the overlap we computed in Section 4.3. Thus we finally complete the proof of

Theorem IV.1.

To use the framework of local weak convergence to study minimum matchings, we append $\{0,1\}$ to the edges of planted networks in $\mathcal{G}_*$. In a slight abuse of terminology, we add a $\widehat{\text{hat}}$ and also call $\widehat{N} = (G, \ell, \mathscr{M})$ a planted network where ℓ is the weight function and $\mathscr{M} : E \to \{0,1\}$ is a mark function (which may or may not be a matching). This lets us discuss the joint distribution of edge weights, the planted matching, and the minimum matching of vertex neighborhoods in either model.

In particular, let $\widehat{\mu}_n$ be this distribution in the vicinity of a uniformly random vertex in the finite model,

$$\widehat{\mu}_n = \mathbb{E}U(\widehat{N}_n) \quad \text{is the law of} \quad [\widehat{N}_{n,\circ}(1)] \quad \text{where} \quad \widehat{N}_n = (K_{n,n}, \ell_n, M_{n,\min})\,,$$

and let $\widehat{\mu}_\infty$ be the analogous distribution at the root of the planted PWIT,

$$\widehat{\mu}_\infty \quad \text{is the law of} \quad [\widehat{N}_{\infty,\circ}] \quad \text{where} \quad \widehat{N}_\infty = (T_\infty, \ell_\infty, \mathscr{M}_{\infty,\text{opt}})\,.$$

We will show that $\widehat{\mu}_\infty$ is the weak limit of $\widehat{\mu}_n$. Thus the two models have all the same local statistical properties, including their expected weight and overlap.

The proof consists of two main steps, namely, the easier half and the harder half. In the easier half, using a simple compactness argument we prove that any subsequence of probability measures $\widehat{\mu}_n$ has a subsequence that converges to an involution invariant random matching on planted PWIT. Using Skorokhod's theorem this shows that the weight of the minimum matching on $K_{n,n}$ is at least that of $\mathscr{M}_{\infty,\text{opt}}$:

$$\liminf_{n\to\infty} \mathbb{E}[\ell_n(1, M_{n,\min}(1))] \geq \mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}_{\infty,\text{opt}}(\emptyset))]\,. \tag{4.44}$$

Informally, this follows by contradiction. If $\liminf_n \ell_n(1, M_{n,\min}(1))$ were smaller than this, then the subsequence of sizes n on which it converges to that smaller value would itself have a subsequence that convergences to an involution invariant matching on the planted PWIT with that weight... but this would contradict the optimality of $\mathscr{M}_{\infty,\text{opt}}$.

In the harder half, using $(\ell_\infty, \mathscr{M}_{\infty,\text{opt}})$ we follow the strategy of [20]. First we construct an almost-perfect matching on $(K_{n,n}, \ell_n)$ with weight close to $\mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}_{\infty,\text{opt}}(\emptyset))]$. Then, we fix this almost-perfect matching matching to make a perfect matching on $(K_{n,n}, \ell_n)$ without changing the weight too much. This proves that

$$\limsup_{n\to\infty} \mathbb{E}[\ell_n(1, M_{n,\min}(1))] \leq \mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}_{\infty,\text{opt}}(\emptyset))]\,. \tag{4.45}$$

Combining (4.44) and (4.45), we have

$$\lim_{n\to\infty} \mathbb{E}[\ell_n(1, M_{n,\min}(1))] = \mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}_{\infty,\mathrm{opt}}(\emptyset))]\,. \tag{4.46}$$

As in Aldous' proof of the $\zeta(2)$ conjecture for the un-planted model, this establishes the expected weight of the minimum matching in the planted model. But much more is true. Since $\mathscr{M}_{\infty,\mathrm{opt}}$ is unique, we get the following theorem.

Theorem IV.1. *The random weak limit of* $(K_{n,n}, \ell_n, M_{n,\min})$ *is* $(T_\infty, \ell_\infty, \mathscr{M}_{\infty,opt})$*, i.e.,* $\widehat{\mu}_n \xrightarrow{w} \widehat{\mu}_\infty$*. In particular, their expected overlap is equal to*

$$\begin{aligned}\alpha(\lambda) &\coloneqq \lim_{n\to\infty} \frac{1}{n}\,\mathbb{E}[|M_{n,\min} \cap M_n^*|] \\ &= \lim_{n\to\infty} \mathbb{P}(\{1,1'\} \in M_n^*) = \mathbb{P}(\{\emptyset, 0\} \in \mathscr{M}_{\infty,opt})\,.\end{aligned}$$

Proof. In the easy-half proof, we show that every subsequence of $\widehat{\mu}_n$ has a further subsequence that converges to an involution random matching on the planted PWIT (see next subsection). Now, by (4.46) and Proposition IV.6, every subsequence of $\widehat{\mu}_n$ converges to $\widehat{\mu}_\infty$; so does the whole sequence. Hence, the random weak limit of $(K_{n,n}, \ell_n, M_{n,\min})$ is $(T_\infty, \ell_\infty, \mathscr{M}_{\infty,\mathrm{opt}})$. □

Finally, using Corollary IV.2 we have for the expected overlap

$$\mathbb{P}(\{\emptyset, 0\} \in \mathscr{M}_{\infty,\mathrm{opt}}) = \mathbb{P}(X + \widehat{X} > \ell_\infty(\emptyset, 0))\,, \tag{4.47}$$

where X and $\widehat{X}$ are independent copies of X_0 and (X_0, Y_0) is the unique solution of the system of recursive distributional equations (4.29)–(4.30). But this computation is exactly what we have done in Section 4.3 by transforming these distributional equations into a system of ordinary differential equations. This completes the proof of Theorem IV.1.

4.7.1 The Easy Half: A Simple Compactness Argument

As the first step toward the proof of the local weak convergence of $\widehat{\mu}_n$ to $\widehat{\mu}_\infty$, we show that for any sequence of n that tends to infinity, there is a subsequence that converges weakly to some involution invariant random matching $\widehat{\mu}'_\infty$ on the planted PWIT. Saying this again in symbols, for any sequence (n_k) there is an involution invariant $\widehat{\mu}'_\infty$ and a subsequence $(n_j) \subseteq (n_k)$ such that $\widehat{\mu}_{n_j} \xrightarrow{w} \widehat{\mu}'_\infty$. Our argument is somewhat simplified from [24, Section 5.8, pp. 53-54].

By Theorem IV.1, we already know $\mu_{n_k} \xrightarrow{w} \mu_\infty$: that is, the two models agree on their local distributions of weighted neighborhoods. Since $\mathcal{G}_*$ is a Polish space, the Prokhorov

theorem [87, p. 309, Thm. 16.3] implies that the sequence μ_{n_k} is tight, i.e., for every $\epsilon > 0$ there is a compact set $\mathcal{K} \subset \mathcal{G}_*$ such that $P([N_{n_k,\circ}(1)] \in \mathcal{K}) > 1 - \epsilon$ (recall that μ_{n_k} is the law of $[N_{n_k,\circ}(1)]$).

Define $\widehat{\mathcal{K}}$ by appending $\{0,1\}$ to the edges of each member of $\mathcal{K}$. Since $\mathcal{K}$ is compact, so is $\widehat{\mathcal{K}}$. Moreover, since $P([N_{n_k,\circ}(1)] \in \mathcal{K}) > 1 - \epsilon$, so is $P([\widehat{N}_{n_k,\circ}(1)] \in \widehat{\mathcal{K}}) > 1 - \epsilon$. Hence, the sequence $\widehat{\mu}_{n_k}$ is also tight, and by the Prokhorov theorem there is a further subsequence n_j such that $\widehat{\mu}_{n_j} \xrightarrow{w} \widehat{\mu}'_\infty$ where $\widehat{\mu}'_\infty$ is some random matching on the planted PWIT. Since involution invariance passes through limit, $\widehat{\mu}'_\infty$ is involution invariant.

By Skorokhod's theorem we can assume that $[\widehat{N}_{n_j,\circ}(1)] \xrightarrow{\text{loc}} [\widehat{N}'_{\infty,\circ}]$ almost surely, where $[\widehat{N}'_{\infty,\circ}] \sim \widehat{\mu}'_\infty$, and $\widehat{N}'_\infty = (T_\infty, \ell_\infty, \mathscr{M}'_\infty)$. By the definition of local convergence,

$$\ell_{n_j}(1, M_{n_j,\min}(1)) \to \ell_\infty(\emptyset, \mathscr{M}'_\infty(\emptyset)) \text{ as } n \to \infty, \text{ almost surely.}$$

Using Fatou's Lemma, we have

$$\liminf_{n_j \to \infty} \mathbb{E}[\ell_{n_j}(1, M_{n_j,\min}(1))] \geq \mathbb{E}[\ell_\infty(\emptyset, M'_\infty(\emptyset))].$$

The lower bound (4.44) follows by assuming n_k is a subsequence of n that achieves $\liminf_{n\to\infty} \mathbb{E}[\ell_n(1, M_{n,\min}(1))]$.

4.7.2 The Harder Half

In the "easy half" above we proved the inequality (4.44), namely that the average weight of the minimum matching on $(K_{n,n}, \ell_n)$ is bounded below by $\mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}_{\infty,\text{opt}}(\emptyset))]$. Now, we are going to prove the inequality (4.45) in the opposite direction, and therefore that inequality can be replaced by equality within arbitrarily small ϵ. The key idea is to construct a low weight matching on $(K_{n,n}, \ell_n)$ using $(\ell_\infty, \mathscr{M}_{\infty,\text{opt}})$. In particular, we want the average weight of this matching to be arbitrarily close to $\mathbb{E}[\ell_\infty(\emptyset, \mathscr{M}_{\infty,\text{opt}}(\emptyset))]$ for large enough n.

Recall that the weight of planted and un-planted edges in $(K_{n,n}, \ell_n)$ are distributed as $Exp(\lambda)$ and $Exp(1/n)$ respectively, independent of everything else. Intuitively speaking, $K_{n,n}$ viewed from the planted directed edge $(1, 1')$ corresponds to the doubly-planted PWIT viewed from the root's planted edge $(\emptyset, 0)$, and $K_{n,n}$ viewed from an un-planted directed edge such as $(1, 2)$ corresponds to the doubly-rooted planted PWIT viewed from an un-planted edge incident to the root, namely $(\emptyset, i)$ where $i \in \mathbb{N}$ is arbitrary.

Now, using this "edge-centric" viewpoint, and following the approach of Aldous in [20,21], we will assign possibly fractional values to the edges of $(K_{n,n}, \ell_n)$ such that the value assigned to the edge $e = \{i, j'\} \in E_n$ corresponds to the probability that $\mathscr{M}_{\infty,\text{opt}}(e) = 1$, assuming its

local neighborhood in $(K_{n,n}, \ell_n)$ is a realization of the planted PWIT.

Consider the $n \times n$ matrix Q_n where $q_{i,j'}$ is the value assigned to the edge (i, j'). If these values corresponded exactly to probabilities, then Q_n would be doubly-stochastic. That is, Q_n would be a fractional matching, i.e., an element of the matching polytope, with weight

$$\ell_n(Q_n) = \sum_{i \in [n], j \in [n']} q_{i,j'} \ell_n(i, j') \,.$$

Since minimizing the weight is a linear programming problem, there is a vertex of this polytope, i.e., an honest matching $M(\pi)_{i,j'} = 1$ if $j' = \pi(i)$ and 0 otherwise for some permutation π, whose weight is less than or equal to $\ell_n(Q_n)$. Alternatively, we could use the Birkhoff theorem to write Q as a convex combination of permutation matrices,

$$Q = \sum_{\pi} c_\pi M(\pi) \,.$$

Then if we choose a random matching π with probability c_π, the expected weight would be $\ell_n(Q_n)$. Finally, if $(1/n)\ell_n(Q_n)$ also converges to the expected weight of $\mathscr{M}_{\infty,\mathrm{opt}}$, we would be done.

All this is almost true. As we will see, we will define Q_n by looking at a bi-infinite version of the planted PWIT, extending it in either direction from an edge [21, Section 5.2]. By looking at $(\ell_\infty, \mathscr{M}_{\infty,\mathrm{opt}})$ on large neighborhoods of this edge, we will obtain probabilities that almost, but not quite, sum to 1, since the true partner of a vertex in $\mathscr{M}_{\infty,\mathrm{opt}}$ might be outside this neighborhood. As a result, Q_n is almost doubly-stochastic in a certain sense. Following [20], we then build an almost-perfect matching with weight close to that of $\mathscr{M}_{\infty,\mathrm{opt}}$, and then—by swapping a small fraction of edges—convert this into a perfect matching within increasing the weight very much. At this point in the proof, we can use lemmas in [20] virtually unchanged.

Remark IV.2. For sake of notational simplicity, for the remainder of the section, we will drop the subscripts ∞ and $\infty\infty$. Thus objects without the subscript n live on the planted PWIT, while those with n live on the finite model.

4.7.2.1 The Bi-infinite Planted PWITs

As the first step toward the proof of the "harder half", we change the viewpoint from a vertex to an edge. There are two doubly-rooted infinite versions of the planted PWIT that we need to study. One is rooted at a planted edge $(\varnothing, 0)$, and the other is rooted at an un-planted edge $(\varnothing, i)$ for some $i \in \mathbb{N}$. We illustrate these in Figure 4.6.

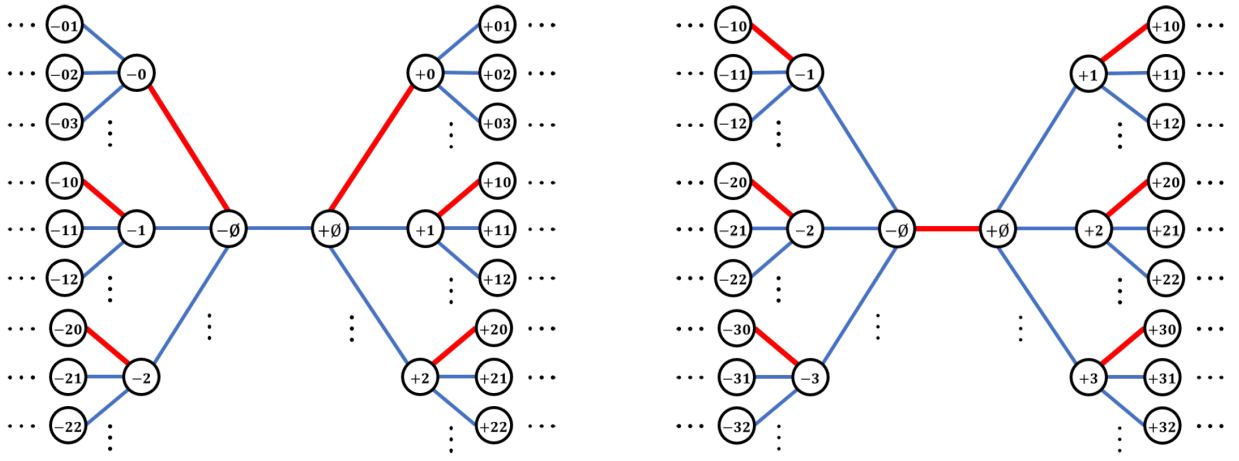


Figure 4.6: The structure of (a) $T_u^{\leftrightarrow}$ and (b) $T_p^{\leftrightarrow}$. The bold red edges are planted edges and the solid blue edges are un-planted.

The measure on $i \in \mathbb{N}$ will be the uniform counting measure—that is, every position in the order of ø's un-planted edges has the same measure. Although this measure is not normalizable, we will speak informally of i as "uniformly chosen." One intuition for this uniformity comes from the finite model $K_{n,n}$. If we choose uniformly from the $n(n-1)$ un-planted edges, or (by symmetry) from the $n-1$ un-planted edges of the root vertex 1, then its order in the sorted list of edge weights is uniform on the set $\{1, \ldots, n-1\}$.

The other intuition is as follows. The edge weight of an un-planted edge is distributed as $Exp(1/n)$, which for weights of constant size is asymptotically $1/n$ times the Lebesgue measure on $\mathbb{R}_+$. If the weight is x, the probability that the Poisson process of weight 1 generates $i-1$ arrivals in the interval $[0, x]$ is $\mathrm{e}^{-x}x^{i-1}/(i-1)!$. The total Lebesgue measure of this event is then

$$\int_0^\infty \frac{\mathrm{e}^{-x}x^{i-1}}{(i-1)!}\mathrm{d}x = \frac{\Gamma(i)}{(i-1)!} = 1 \quad \text{for all } i\,, \tag{4.48}$$

so every i has equal measure.

We make this intuition rigorous below, showing how the appropriate measure on bi-infinite PWITs around both types of edge is related to the planted PWIT by extending the strategy of [21, Section 5.2] to the planted case.

In the case of the planted edge $\{ø, 0\}$, the corresponding bi-infinite tree is just a relabeling of the vertices: relabel ø as $-ø$ and relabel 0 as $+ø$, and then relabel all the other vertices "below" these two roots as we did before (see Figure 4.6). However, for a planted edge $\{ø, i\}$ for a "uniformly" selected $i \in \mathbb{N}$, things are a bit more complicated. Since $i \in \mathbb{N}$ is "uniformly" selected, the cost of the edge $\{ø, i\}$ is "uniformly distributed" over $\mathbb{R}_+$, i.e.,

Lebesgue measure on $\mathbb{R}_+$. On the other hand, as we pointed out in the proof of Corollary IV.2, the other un-planted children of $\emptyset$ are still arrivals of a Poisson process. Specifically, if we remove $\{\emptyset, i\}$, the remaining connected component of $\emptyset$ is still a planted PWIT. Hence, the corresponding bi-infinite tree is obtained by gluing two independent copies of the planted PWIT, using an edge with edge weight "distributed" as Lebesgue measure, and then relabeling the vertices: $\emptyset$ as $-\emptyset$, i as $+\emptyset$, and the others accordingly (See Figure 4.6). In the rest of the section, we are going to give detailed construction of the bi-infinite trees, and then we show that they are equivalent to the corresponding doubly-rooted PWIT.

Remark IV.3. To distinguish the edge-centric viewpoint from the singly-rooted vertex viewpoint, we use the superscript $\leftrightarrow$.

Let T_u denote the planted PWIT rooted at $\emptyset$, and let T_p denote the subtree, rooted at 0, obtained by removing the edge $\{\emptyset, 0\}$. Relabel the root of T_p to be $\emptyset$, and then relabel all the vertices of T_p using the same rule as in the planted PWIT. Let T_u^- and T_u^+ be two independent copies of T_u. Similarly, let T_p^- and T_p^+ be two independent copies of T_p. Relabel the vertices of T_u^+ and T_p^+ by adding "$+$" sign to the original labels, and relabel the vertices of T_u^- and T_p^- by adding "$-$" sign to the original labels.

Now, let $T_u^{\leftrightarrow}$ denote a bi-infinite tree, rooted at $(-\emptyset, +\emptyset)$, obtained by joining the roots of T_u^- and T_u^+. Let $V_u^{\leftrightarrow}$ and $E_u^{\leftrightarrow}$ denote the vertices and the edges of $T_u^{\leftrightarrow}$ respectively. Let $M_u^{\leftrightarrow}$ denote the set of planted edges of $T_u^{\leftrightarrow}$, which is the union of the planted edges in T_u^- and T_u^+. Let $\ell_u^{\leftrightarrow}$ denote the function that assigns weight to the edges of $T_u^{\leftrightarrow}$ by using the weight of the edges in T_u^- and T_u^+, and specifying the weight of $\{-\emptyset, +\emptyset\}$ to be uniformly "distributed" on $[0, \infty)$, i.e., Lebesgue measure on $\mathbb{R}_+$, independent of everything else. Write $T_u^{\leftrightarrow} = (G_u^{\leftrightarrow}, \ell_u^{\leftrightarrow})$, where $G_u^{\leftrightarrow} = (V_u^{\leftrightarrow}, E_u^{\leftrightarrow}, M_u^{\leftrightarrow}, (-\emptyset, +\emptyset))$, and let $\mu_u^{\leftrightarrow}$ denote the σ-finite measure associated with $T_u^{\leftrightarrow}$.

Similarly, define $T_p^{\leftrightarrow}$ by joining the roots of T_p^- and T_p^+. However, this time include $\{-\emptyset, +\emptyset\}$ as a planted edge in $M_p^{\leftrightarrow}$, and specify the weight of $\{-\emptyset, +\emptyset\}$ to be an exponentially distributed random variable with parameter λ, independent of everything else. Write $T_p^{\leftrightarrow} = (G_p^{\leftrightarrow}, \ell_p^{\leftrightarrow})$, where $G_p^{\leftrightarrow} = (V_p^{\leftrightarrow}, E_p^{\leftrightarrow}, M_p^{\leftrightarrow}, (-\emptyset, +\emptyset))$, and let $\mu_p^{\leftrightarrow}$ denote the probability distribution of $T_p^{\leftrightarrow}$. Figure 4.6 illustrates a realization of $T_u^{\leftrightarrow}$ and $T_p^{\leftrightarrow}$.

Recall that the doubly rooted PWIT is the product measure $\mu \times \texttt{count}$ on $\{0, 1, 2, 3, \cdots\}$, where `count` is the counting measure. We can think of it as a product measure on $[0, \infty)^E \times \{0, 1, 2, \cdots\}$. Similarly, we can think of $\mu_u^{\leftrightarrow}$ and $\mu_p^{\leftrightarrow}$ as a σ-finite measure on $[0, \infty)^{E_u^{\leftrightarrow}}$ and a probability measure on $[0, \infty)^{E_p^{\leftrightarrow}}$, respectively. Now, depending on whether the second root is 0 or not, there is a natural map from the doubly rooted PWIT to $T_p^{\leftrightarrow}$ or $T_u^{\leftrightarrow}$: (1) If the second root is 0, relabel the vertices as

i. relabel 0 as $+\emptyset$ and $\emptyset$ as $-\emptyset$,

ii. relabel any sequence $0i_1i_2\cdots i_l$ for $l \geq 0$ as $+i_1i_2\cdots i_l$,

iii. relabel any sequence $i_1i_2\cdots i_l$ for $l \geq 1$ such that $i_1 \neq 0$ as $-i_1i_2\cdots i_l$.

This relabeling induces a bijection $\psi\big|_p : [0,\infty)^E \times \{0\} \to [0,\infty)^{E^{\leftrightarrow}_p}$; (2) If the second root is $k \in \{1,2,3,\cdots\}$, relabel the vertices as

i. relabel k as $+\emptyset$ and $\emptyset$ as $-\emptyset$,

ii. relabel any sequence $ki_1i_2\cdots i_l$ for $l \geq 0$ as $+i_1i_2\cdots i_l$,

iii. relabel any sequence $i_1i_2\cdots i_l$ for $l \geq 1$ such that $i_1 > k$ as $-(i_1-1)i_2\cdots i_l$,

iv. relabel any sequence $i_1i_2\cdots i_l$ for $l \geq 1$ such that $i_1 < k$ as $-i_1i_2\cdots i_l$.

This relabeling induces a bijection $\psi\big|_u : [0,\infty)^E \times \{1,2,3,\cdots\} \to [0,\infty)^{E^{\leftrightarrow}_u}$. Clearly, $\psi\big|_p$ maps $\mu \times \delta_0$ to $\mu^{\leftrightarrow}_p$, where δ_0 is the delta measure on the second root. It is also easy to see that $\psi\big|_u$ maps $\mu \times \texttt{count}$ on $\{1,2,3,\cdots\}$ to $\mu^{\leftrightarrow}_u$ using the following lemma. Finally, note that

$$\mu \times \texttt{count} \text{ on } \{0,1,2,3,\cdots\} = \mu \times \delta_0 + \mu \times \texttt{count} \text{ on } \{1,2,3,\cdots\}.$$

Lemma IV.4. *[21, Lemma 25] Write $\Delta := \{(x_i) : 0 < x_1 < x_2 < \cdots, x_i \to \infty\}$. Write $\texttt{Pois}$ for the probability measure on Δ which is the distribution of the Poisson process of rate 1. Consider the map $\mathcal{X} : \Delta \times \{1,2,3,\cdots\} \to \Delta \times [0,\infty)$ which takes $((x_i),k)$ to $((x_i, i \neq k), x_k)$. Then $\mathcal{X}$ maps $\texttt{Pois} \times \texttt{count}$ to $\texttt{Pois} \times \texttt{Leb}$, where $\texttt{Leb}$ is the Lebesgue measure on $[0,\infty)$.*

Remark IV.5. For sake of brevity, we use $E^{\leftrightarrow}_{\cdot}$ instead of $E^{\leftrightarrow}_u$ or $E^{\leftrightarrow}_p$ whenever the subscript does not affect the discussion. Similar changes applies to other symbols. More specifically, all the following arguments are true if we replace all the "$_{\cdot}$" with "$_u$" or "$_p$".

We introduce two additional pieces of notation that we will use in the following subsections. Recall that $G^{\leftrightarrow}_{\cdot} = (V^{\leftrightarrow}_{\cdot}, E^{\leftrightarrow}_{\cdot}, M^{\leftrightarrow}_{\cdot}, (-\emptyset, +\emptyset))$ is a bi-infinite tree (without the edge weights). We define $G^{\leftrightarrow}_{\cdot,B}$ to be a subtree of $G^{\leftrightarrow}_{\cdot}$ induced by $V^{\leftrightarrow}_{\cdot,B} := \{\pm\emptyset\} \cup \{\pm i_1i_2\cdots i_l \in V^{\leftrightarrow}_{\cdot} : i_s \in \{0,1,\cdots,B\}\}$, i.e., the subtree obtained by restricting the number of un-planted children of each vertex to B. We also define $G^{\leftrightarrow}_{\cdot,B,H}$ to be a subtree of $G^{\leftrightarrow}_{\cdot,B}$ induced by $V^{\leftrightarrow}_{\cdot,B,H} := \{\pm\emptyset\} \cup \{\pm i_1i_2\cdots i_l \in V^{\leftrightarrow}_{\cdot,B} : l \leq H+1 \text{ and } l = H+1 \text{ iff } i_l = 0\}$, i.e., the subtree obtained by restricting the depth of vertices to H or $H+1$, depending on whether the vertex is a planted pair of its parent or not. Define $\ell^{\leftrightarrow}_{\cdot,B}$ to be the restriction of $\ell^{\leftrightarrow}_{\cdot}$ to $\ell^{\leftrightarrow}_{\cdot,B}$, and $\ell^{\leftrightarrow}_{\cdot,B,H}$

to be the restriction of $\ell^{\leftrightarrow}_{\cdot,B}$ to $\ell^{\leftrightarrow}_{\cdot,B,H}$. Now, define $T^{\leftrightarrow}_{\cdot,B} \coloneqq (G^{\leftrightarrow}_{\cdot,B}, \ell^{\leftrightarrow}_{\cdot,B})$ and $T^{\leftrightarrow}_{\cdot,B,H} \coloneqq (G^{\leftrightarrow}_{\cdot,B,H}, \ell^{\leftrightarrow}_{\cdot,B,H})$, and let $\mu^{\leftrightarrow}_{\cdot,B}$ and $\mu^{\leftrightarrow}_{\cdot,B,H}$ to be the associated measures. Note that there is a natural restriction $\rho_{\cdot,H}$ that maps $T^{\leftrightarrow}_{\cdot,B}$ to $T^{\leftrightarrow}_{\cdot,B,H}$.

Thus far, we show that the doubly rooted planted PWITs and bi-infinite trees are equivalent. Next, we are going establish the connection between the planted model $(K_{n,n}, \ell_n)$ and the bi-infinite trees.

4.7.2.2 The Unfolding Map

Now that we have discussed how to view the planted PWIT from an edge (a planted edge or a "uniformly" selected un-planted edge), we are going to discuss the similar viewpoint in $K_{n,n} = (V_n, E_n, M_n^*)$. This is done via an unfolding map that unfolds $(K_{n,n}, \ell_n)$ viewed from a planted or un-planted edge. This unfolding map is similar to the one discussed in [20, Section 3.2] with two additional properties: the number of un-planted children of every vertex is the same (for a fixed B when n is large enough), and the set of un-planted edges M_n^* is a matching on the unfolded graph. Eventually, we want to show that the local neighborhood of an edge in $K_{n,n}$ is the same as the local neighborhood of the edge in the bi-infinite trees. This should not come as a surprise, given Theorem IV.1.

Next, we are going to give detailed construction of the unfolding map. Fix $B \in \mathbb{N}_+$, and fix some edge $\{i, j'\}$ from $K_{n,n}$. We are going to unfold $(K_{n,n}, \ell_n)$ from the viewpoint of the edge $\{i, j'\}$ and construct the doubly rooted tree $U^{\leftrightarrow}_{B,n}(i, j')$, rooted at (i, j'), with maximum $B + 1$ arity. This unfolding map resembles the exploration process discussed in the proof of Theorem IV.1, with one key difference: this map unfolds $(K_{n,n}, \ell_n)$ from an edge viewpoint.

Include the edge $\{i, j'\}$ in $U^{\leftrightarrow}_{B,n}(i, j')$. The unfolding map proceeds as follows. In this process, all vertices will be "live", "dead", or "neutral". The live vertices will be contained in a queue. Initially, i and j' are live and the queue consists of only i, j' in order and all the other vertices are neutral. At each time step, a live vertex v is popped from the head of the queue. If v's planted neighbor $M_n^*(v)$ is neutral, then include the edge $\{v, M_n^*(v)\}$ in the doubly rooted tree $U^{\leftrightarrow}_{B,n}(i, j')$, add $M_n^*(v)$ to the end of the queue and let $M_n^*(v)$ live. Let v_1, $v_2, \ldots, v_B$ denote the B closest un-planted, neutral neighbors of vertex v in a non-decreasing order of distance to v. Include all the edges $\{v, v_k\}$ in the doubly rooted tree $U^{\leftrightarrow}_{B,n}(i, j')$ for $1 \le k \le B$. Also, include all the edges $\{v_k, M_n^*(v_k)\}$ in $U^{\leftrightarrow}_{B,n}(i, j')$ for $1 \le k \le B$. The popped vertex v is dead. Add those neutral vertices $v_1, v_2, \ldots, v_B$ to the end of the queue in order and they are live. Also, add those neutral vertices $M_n^*(v_1), M_n^*(v_2), \ldots, M_n^*(v_B)$ to the end of the queue in order and they are live. The process ends when the queue is empty. According to this rule, the order of the vertices that are selected after j' is the planted neighbor i_0 of i (if $\{i, j'\}$ is not planted), i_1, i_2, $\cdots$, i_B, $M_n^*(i_1)$, $M_n^*(i_2)$, $\cdots$, $M_n^*(i_B)$, the planted neighbor j_0' of

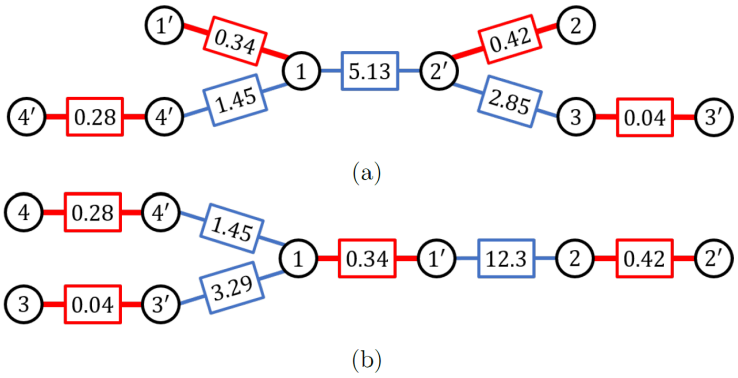


Figure 4.7: The doubly rooted trees (a) $U^{\leftrightarrow}_{1,4}(1,2')$ and (b) $U^{\leftrightarrow}_{1,4}(1,1')$ obtained by unfolding the planted network in Figure 4.4.

j' (if $\{i,j'\}$ is not planted), j'_1, j'_2, $\cdots$, j'_B, $M^*_n(j'_1)$, $M^*_n(j'_2)$, $\cdots$, $M^*_n(j'_B)$, etc. This unfolding process stops when all vertices of $K_{n,n}$ are included in $U^{\leftrightarrow}_{B,n}(i,j')$. Figure 4.7 illustrates the $U^{\leftrightarrow}_{1,4}(1,2')$ and $U^{\leftrightarrow}_{1,4}(1,1')$ for the planted network given in Figure 4.4. Note that in both cases, M^*_n is a matching.

Let us define a relabeling bijection $\phi^{i,j'}_{B,n}$ (not to be confused with $\phi\big|_p$ and $\phi\big|_u$) from V_n to a subset of $+\Sigma \cup -\Sigma$, where $+\Sigma := \{+\boldsymbol{i} : \boldsymbol{i} \in \Sigma\}$ and $-\Sigma := \{-\boldsymbol{i} : \boldsymbol{i} \in \Sigma\}$ denote the set of vertex labelings of the bi-infinite planted PWIT. Define $\phi^{i,j'}_{B,n}(i) = -\emptyset$ and $\phi^{i,j'}_{B,n}(j') = +\emptyset$. At each step of the unfolding process, when we pick vertex v and add $\{v, v_k\}$ and $\{v_k, M^*_n(v_k)\}$ to $U^{\leftrightarrow}_{B,n}(i,j')$, set $\phi^{i,j'}_{B,n}(v_k) = \phi^{i,j'}_{B,n}(v)k$ and $\phi^{i,j'}_{B,n}(M^*_n(v_k)) = \phi^{i,j'}_{B,n}(v_k)0$.

Now, using the bijection $\phi^{i,j'}_{B,n}$ we can define the restriction map $\rho_{\cdot,H}$ on $U^{\leftrightarrow}_{B,n}(i,j')$. It is easy to see that the graph structure of $\rho_{u,H}\left(U^{\leftrightarrow}_{B,n}(i,j')\right)$ for $j' \neq i'$ and all sufficiently large n is isomorphic to $G^{\leftrightarrow}_{u,B,H}$, and the graph structure of $\rho_{p,H}\left(U^{\leftrightarrow}_{B,n}(i,i')\right)$ for all sufficiently large n is isomorphic to $G^{\leftrightarrow}_{p,B,H}$. Let $\mu^{\leftrightarrow}_{n,u,B,H}$ denote the probability measure associated with $\rho_{u,H}\left(U^{\leftrightarrow}_{B,n}(i,j')\right)$ for $j' \neq i'$, and let $\mu^{\leftrightarrow}_{n,p,B,H}$ denote the probability measure associated with $\rho_{u,H}\left(U^{\leftrightarrow}_{B,n}(i,i')\right)$. Note that if $D_{u,x} = \{T^{\leftrightarrow}_{u,B,H} : \ell^{\leftrightarrow}_{u,B,H}(\{-\emptyset,\emptyset\}) < x\}$, then $\mu^{\leftrightarrow}_{n,u,B,H}(D_{u,x}) = 1 - \exp(-x/n) \approx x/n$ as $n \to \infty$. Also, note that $\mu^{\leftrightarrow}_{u,B,H}(D_{u,x}) = x$. We now generalize Lemma 10 in [20] to the planted case.

Lemma IV.6. *For any fixed $B, H \in \mathbb{N}_+$, and $x > 0$ we have*

$$n\mu^{\leftrightarrow}_{n,u,B,H}(D_{u,x} \cap \cdot\,) \xrightarrow{TV} \mu^{\leftrightarrow}_{u,B,H}(D_{u,x} \cap \cdot\,)$$
$$\mu^{\leftrightarrow}_{n,p,B,H} \xrightarrow{TV} \mu^{\leftrightarrow}_{p,B,H},$$

where the total variation convergence of positive measures is defined as:

$$\mu_n \xrightarrow{TV} \mu \text{ iff } \sup_A |\mu_n(A) - \mu_(A)| \to 0.$$

Proof. The proof is almost identical to the proof of Lemma C.1, and has been omitted. The only subtle difference is the restriction to $D_{u,x}$, which causes no problem since $n \times \frac{1}{n} \mathrm{e}^{-x/n} \to 1$ as $n \to \infty$. □

Next, we are going to use the involution invariant random matching $(\ell, \mathscr{M}_{\text{opt}})$ on the planted PWIT to assign values to the edges of $(K_{n,n}, \ell_n)$. Ideally, the value assigned to an edge $e = \{i, j'\}$ corresponds to the probability of e being in the matching that we want to construct.

4.7.2.3 Assigning Values to the Edges of $(K_{n,n}, \ell_n)$

Now it's time to assign fractional values to the edges of $(K_{n,n}, \ell_n)$. This is done by pretending that the local neighborhood of an edge e in $(K_{n,n}, \ell_n)$, is a realization of the local neighborhood of the corresponding bi-infinite tree. So, as the first step toward assigning values to the edges of $(K_{n,n}, \ell_n)$, we need to know how to assign value to $\{-\emptyset, +\emptyset\}$ in the bi-infinite tree using the minimum matching on the planted PWIT. The idea is to use the inverse image of $\psi\big|_{\cdot}$ and map the edge $\{-\emptyset, +\emptyset\}$ to the corresponding edge on the planted PWIT. This gives us a function $g_{\cdot} : [0, \infty)^{E_{\cdot}^{\leftrightarrow}} \to [0, 1]$, which we then use to assign fractional values to the edges of $(K_{n,n}, \ell_n)$, by conditioning on its neighborhood. We are going to discuss these steps in detail. We follow the discussion in [21, Section 5.5] and then [20, Introduction].

Recall that $\widetilde{\mu} = \mu \times \texttt{count}$ on $\{0, 1, 2, 3, \cdots\}$ is the measure associated with the doubly rooted planted PWIT. Also, recall that since μ is involution invariant, the product measure $\widetilde{\mu}$ is invariant under the involution map ι that swaps the roots. Following the discussion of [21, Section 5.5], we define the function $\gamma_{\text{opt}} : [0, \infty)^E \times \{0, 1, 2, \cdots\} \to [0, 1]$ by

$$\gamma_{\text{opt}}(\boldsymbol{w}, i) = \mathbb{P}[\mathscr{M}_{\text{opt}}(\emptyset) = i | \ell(e) = w(e) \, \forall e \in E]$$

As Aldous points out, the function γ_{opt} satisfies certain consistency properties:

(i) $\sum_{i=0}^{\infty} \gamma_{\text{opt}}(\boldsymbol{w}, i) = 1$ since $\mathscr{M}_{\text{opt}}$ is a matching.

(ii) $\gamma_{\text{opt}}(\iota(\boldsymbol{w}, i)) = \gamma_{\text{opt}}(\boldsymbol{w}, i)$ since the random matching $(\ell, \mathscr{M}_{\text{opt}})$ is involution invariance, where $\iota(\boldsymbol{w}, i)$ swaps i and $\emptyset$ given $\{\ell(e) = w(e) \, \forall e \in E\}$.

Also, we have (iii) $\mathbb{E}[\ell(\mathscr{M}_{\text{opt}}(\emptyset),\emptyset)] = \int_{\boldsymbol{w}} \sum_{i=0}^{\infty} \gamma_{\text{opt}}(\boldsymbol{w},i) w(\emptyset,i)\, \mu(d\boldsymbol{w})$. Now, define $g_{\cdot} : [0,\infty)^{E_{\cdot}^{\leftrightarrow}} \to [0,1]$ as follows:

$$g_{\cdot}(\boldsymbol{w}_{\cdot}^{\leftrightarrow}) = \gamma_{\text{opt}}(\psi\big|_{\cdot}^{-1}(\boldsymbol{w}_{\cdot}^{\leftrightarrow})),$$

where $\psi\big|_{\cdot}$ is the bijection map that we defined earlier in Section 4.7.2.1, and $\psi\big|_{\cdot}^{-1}$ is its inverse. We can think of the function $g_{\cdot}$ as the probability that the edge $\{-\emptyset,+\emptyset\}$ is in the matching. The function $g_{\cdot}$ satisfies similar consistency properties corresponding to (i) and (ii):

(i′) The function $g_{\cdot}$ assigns honest probabilities to the neighbors of $-\emptyset$ as well as $+\emptyset$. To be more specific, let $\iota_{p,i}^{+}, \iota_{p,i}^{-} : [0,\infty)^{E_p^{\leftrightarrow}} \to [0,\infty)^{E_u^{\leftrightarrow}}$ denote the root swapping maps, that change the root from $(-\emptyset,+\emptyset)$ to $(-\emptyset,-i)$, and from $(-\emptyset,+\emptyset)$ to $(+i,+\emptyset)$ respectively (and then relabel all the vertices). Similarly, define $\iota_{u,i}^{+}$ and $\iota_{u,i}^{-}$. We have:

(a) In $T_p^{\leftrightarrow}$, the values assigned to the neighbors of $-\emptyset$ as well as $+\emptyset$ sums to one:

$$g_p(\boldsymbol{w}_p^{\leftrightarrow}) + \sum_{i=1}^{\infty} g_u(\iota_{p,i}^{\pm}(\boldsymbol{w}_p^{\leftrightarrow})) = 1.$$

(b) In $T_u^{\leftrightarrow}$, the values assigned to the neighbors of $-\emptyset$ as well as $+\emptyset$ sums to one:

$$g_u(\boldsymbol{w}_u^{\leftrightarrow}) + g_p(\iota_{u,0}^{\pm}(\boldsymbol{w}_u^{\leftrightarrow})) + \sum_{i=1}^{\infty} g_u(\iota_{u,i}^{\pm}(\boldsymbol{w}_u^{\leftrightarrow})) = 1.$$

(ii′) Let $\iota_{\cdot}^{\leftrightarrow}$ denote the root swamping map, that swaps the root $(-\emptyset,+\emptyset)$ to $(+\emptyset,-\emptyset)$ (and then relabels all the vertices). We have $g_{\cdot}(\boldsymbol{w}_{\cdot}^{\leftrightarrow}) = g_{\cdot}(\iota_{\cdot}^{\leftrightarrow}(\boldsymbol{w}_{\cdot}^{\leftrightarrow}))$.

Also note that the function $g_{\cdot}$ is measurable with respect to the product σ-algebra on $[0,\infty)^{E_{\cdot}^{\leftrightarrow}}$. Hence, (iii) becomes (iii′)

$$\begin{aligned}\mathbb{E}[\ell(\mathscr{M}_{\text{opt}}(\emptyset),\emptyset)] = &\int_{\boldsymbol{w}_p^{\leftrightarrow}} \boldsymbol{w}_p^{\leftrightarrow}(-\emptyset,+\emptyset)\, g_p(\boldsymbol{w}_p^{\leftrightarrow})\, \mu_p^{\leftrightarrow}(d\boldsymbol{w}_p^{\leftrightarrow}) \\ &+ \int_{\boldsymbol{w}_u^{\leftrightarrow}} \boldsymbol{w}_u^{\leftrightarrow}(-\emptyset,+\emptyset)\, g_u(\boldsymbol{w}_u^{\leftrightarrow})\, \mu_u^{\leftrightarrow}(d\boldsymbol{w}_u^{\leftrightarrow}).\end{aligned}$$

Now, we use the function $g_{\cdot}$ to assign fractional values to the edges of the planted model. We store the values assigned to the edges of $(K_{n,n},\ell_n)$ in an $n\times n$ matrix $Q_n = [q_{i,j'}]_{i,j'}$, where $q_{i,j'}$ is the value assigned to the edge $\{i,j'\}$. Let $g_{\cdot,B,H}$ to be the conditional expectation of $g_{\cdot}$ given $\sigma(\ell_{\cdot}^{\leftrightarrow}(e), e \in E_{\cdot,B,H}^{\leftrightarrow})$, with respect to the measure $\mu_{\cdot}^{\leftrightarrow}$. Also, let $g_{u,x,B,H} = g_{u,B,H}\mathbf{1}_{D_{u,x}}$.

Now, $q_{i,j'}$ is defined as follows:

$$q_{i,j'} \coloneqq \begin{cases} g_{u,x,B,H}\left(\rho_{u,H}\left(U^{\leftrightarrow}_{B,n}(i,j')\right)\right) & \text{if } j' \neq i' \\ g_{p,B,H}\left(\rho_{p,H}\left(U^{\leftrightarrow}_{B,n}(i,i')\right)\right) & \text{if } j' = i' \end{cases} \tag{4.49}$$

As we mentioned before, if Q_n were doubly-stochastic, we could have used it to construct a matching on $K_{n,n}$. However, it is not hard to see that $\mathbb{P}[Q_n \text{ is doubly-stochastic}] = 0$. Nevertheless, we expect Q_n to become almost doubly-stochastic, as a larger neighborhood is revealed.

Let us define a discrimination factor, as Aldous does in [20, Introduction]:

$$\mathcal{X}(Q_n) = \frac{1}{n}\sum_{i=1}^{n}\left|1 - \sum_{j'=1'}^{n'} q_{i,j'}\right| + \frac{1}{n}\sum_{j'=1'}^{n'}\left|1 - \sum_{i=1}^{n} q_{i,j'}\right|$$

Note that if Q_n is doubly-stochastic, then $\mathcal{X}(Q_n) = 0$. Naturally, we should expect that $\mathbb{E}[\mathcal{X}(Q_n)] \approx 0$, for large values x, B, H, and n. We should also expect the average expected cost of Q_n to be close to the expected cost of $\mathscr{M}_{\text{opt}}$, i.e.,

$$\frac{1}{n}\mathbb{E}[\sum_{i,j'} q_{i,j'}\ell_n(i,j')] \approx \mathbb{E}[\ell(\mathscr{M}_{\text{opt}}(\emptyset),\emptyset)],$$

for large enough values of x, B, H, and n. Using the same set of inequalities as in [20, Section 3.4], it follows that the both intuitions are correct.

Lemma IV.7. *(i) For any $\epsilon > 0$, there is an x_0, B_0, H_0 and n_0 such that for all $x > x_0$, $B > B_0$, $H > H_0$, and $n > n_0$ we have $\mathbb{E}[\mathcal{X}(Q_n)] < \epsilon$. In the other words,*

$$\lim_{x\to\infty}\limsup_{B\to\infty}\limsup_{H\to\infty}\limsup_{n\to\infty}\mathbb{E}[\mathcal{X}(Q_n)] = 0.$$

(ii) For any $\delta > 0$, there is an x_0 and n_0 such that for all $x > x_0$ and $n > n_0$, the average expected cost of Q_n is in the δ neighborhood of the cost of $\mathscr{M}_{opt}$ on the planted PWIT, for all B and H. In the other words,

$$\lim_{x\to\infty}\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[\sum_{i,j'} q_{i,j'}\ell_n(i,j')] = \mathbb{E}[\ell(\mathscr{M}_{opt}(\emptyset),\emptyset)] \qquad \text{for all } B \text{ and } H.$$

Proof. The proof is almost identical to the proof presented in [20, Section 3.4]. The proof of part (ii) is a direct consequence of Lemma IV.6 and linearity of expectation. The proof of part (i) needs more work, but the inequalities are the same as the one presented in [20, Section

3.4]. The key factor is the consistency properties of functions g_u and g_p, Lemma IV.6, and the fact that there is no short cycle containing the root consists entirely of low-weight edges (as mentioned in the proof of Theorem IV.1). □

Now that we know Q_n eventually becomes a doubly-stochastic matrix with weight close to that of $\mathscr{M}_{\text{opt}}$, we will construct a perfect matching on $(K_{n,n}, \ell_n)$ with (within ϵ) the same weight. First, by invoking [20, Proposition 7], we construct a partial matching with the cost close to the cost of Q_n. Next, using [20, Proposition 9], we construct a perfect matching by swapping operation while keeping the cost almost the same. The changes required to extend this analysis to the planted case are very minor, but we present the strategy in the next section for completeness.

4.7.2.4 Construction of the Matching

Finally, we are going to give the precise construction of the low cost matching on $(K_{n,n}, \ell_n)$ and prove (4.45). By Lemma IV.7, we know that Q_n becomes arbitrary close to a doubly-stochastic matrix, i.e., $\mathcal{X}(Q_n)$ becomes arbitrary close to 0 with high probability. Now, using Q_n for sufficiently large n, we construct a low cost partial matching which matches most of the vertices and leaves a small fraction of vertices isolated. This is done by invoking Proposition 7 in [20]. Then we use Proposition 9 in [20] and swap some of the edges to obtain a perfect matching on $(K_{n,n}, \ell_n)$ while keeping the cost almost the same. On the rare occasions that we fail to either construct the partial matching or swap the edges, we use M_n^*. Since these events are rare, the use of planted matching won't affect the total cost. We follow the discussion in [20, Section 2].

Let us begin with describing how to construct a partial matching using an almost doubly-stochastic matrix. We say Q_n is an almost doubly-stochastic matrix if its discrimination factor is close enough to 0, or more precisely, if $\mathcal{X}(Q_n) < 1/200$. We say $\nu_n : E_n \to \{0, 1, \emptyset\}$ is a $(1-\theta)$ partial matching if vertices in $U(\nu_n) = \{i \in [n] : \nu_n(i, j') = 1 \text{ for some } j' \in [n']\}$ are matched to different vertices in $[n']$ and $|U(\nu_n)| \geq (1-\theta)n$.

The first step is to convert Q_n to a doubly-stochastic matrix. Define an $n \times n$ matrix $A_n = [a_{i,j'}]$ as follows:

$$a_{i,j'} \coloneqq \frac{q_{i,j'}}{\max(1, q_{i,:}) \max(1, q_{:,j'})},$$

where $q_{i,:} \coloneqq \sum_{j' \in [n']} q_{i,j'}$ and $q_{:,j'} \coloneqq \sum_{i \in [n]} q_{i,j'}$. Similarly, define $a_{i,:}$ and $a_{:,j'}$ for all $i \in [n]$

and $j' \in [n']$. Note that $a_{i,:} \leq 1$ and $a_{:,j'} \leq 1$. Define an $n \times n$ matrix $B_n = [b_{i,j'}]$ as follows:

$$b_{i,j'} = \frac{(1 - a_{i,:})(1 - a_{:,j'})}{n - \sum_{i \in [n]} \sum_{j' \in [n']} a_{i,j'}}.$$

It is easy to check that $A_n + B_n$ is doubly-stochastic. Hence, by Birkhoff–von Neumann theorem $A_n + B_n$ can be written as a convex combination of permutations. Hence, there exists a random matching $\mathscr{M}_n$ on (K_n, ℓ_n) such that $\forall i \in [n]$, $j' \in [n'] : P(\mathscr{M}_n(i) = j') = a_{i,j'} + b_{i,j'}$. Note that for all $i \in [n]$ and $j' \in [n']$ we have $a_{i,j'} \leq q_{i,j'}$, however, there is no such a bound for $b_{i,j'}$. As a result, we may end up assigning high probabilities to undesired edges which can affect the expected cost of the matching $\mathscr{M}_n$. Now, the idea is to use some part of the matching $\mathscr{M}_n$ that is behaving well enough.

Proposition IV.8. *[20, Proposition 7] Let $Q_n = [q_{i,j'}]$ and $\ell_n = [\ell_n(i, j')]$ be given non-random $n \times n$ matrices. Suppose $200\mathcal{X}(Q_n) \leq \theta < 1$. Consider the random matching $\mathscr{M}_n$ given as above and define the random set $D(\mathscr{M}_n) \coloneqq \{i \in [n] : b_{i,\mathscr{M}_n(i)} \leq \eta a_{i,\mathscr{M}_n(i)}\}$ where $\eta = \sqrt{3\mathcal{X}(Q_n)/\theta}$. Then*

$$\mathbb{E}[\sum_{i \in D(\mathscr{M}_n)} q_{i,\mathscr{M}_n(i)} \ell_n(i, \mathscr{M}_n(i))] \leq (1 + \eta) \sum_{i \in [n]} \sum_{j' \in [n']} q_{i,j'} \ell_n(i, j'),$$
$$\mathbb{P}[|D(\mathscr{M}_n)| \geq (1 - \theta)n] \geq 1 - \frac{3(1 + \eta^{-1})\mathcal{X}(Q_n)}{\theta},$$

that is, the random matching $\mathscr{M}_n$ restricted to the random set $D(\mathscr{M}_n)$ is a $(1 - \theta)$ matching with high probability with cost close to the cost of Q_n. Specifically, there is a $(1 - \theta)$ partial matching ν_n(non-random) such that

$$\sum_{i \in U(\nu_n)} q_{i,\nu_n(i)} \ell_n(i, \nu_n(i)) \leq (1 + 4\sqrt{\mathcal{X}(Q_n)/\theta}) \sum_{i \in [n]} \sum_{j' \in [n']} q_{i,j'} \ell_n(i, j').$$

Next, we patch the partial matching ν_n given by the Proposition IV.8 (which exists for almost all realization of the edge weights and Q_n), to construct a perfect matching without distorting the expected cost and the given partial matching too much. Aldous suggests using the greedy algorithm to do that. In the planted setting, the idea is to simply remove all the planted edges and then use the greedy algorithm.

Proposition IV.9. *[20, Proposition 9] Fix $0 < \theta < 1/10$, and let $k = \lfloor \theta n \rfloor$. Let $\ell_n = [\ell_n(i, j')]$ denote the matrix of the edge weights. Let ν_n denote a $1 - \theta$ partial matching. Let $N_0 = [n] \setminus U(\nu_n)$ and $N_0' = [n'] \setminus \nu_n(U(\nu_n))$ denote the set of unpaired vertices in either side. (Here, N_0, N_0' and ν_n may depend on ℓ_n in an arbitrary way.) Then there exists a random*

subset $S \subset [n] \setminus N_0$ *of size* k*, random bijections* $\nu_1 : S \to N_0'$*,* $\nu_2 : N_0 \to \nu_n(S)$*, and events* Ω_n *with* $P(\Omega_n) \to 0$ *such that*

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\left[1_{\Omega_n^c}\left(\sum_{i\in N_0} \ell_n(i,\nu_2(i)) + \sum_{i\in S} \ell_n(i,\nu_1(i))\right)\right] \le 24\,\theta^{1/2}.$$

It remains to combine Proposition IV.8 and Proposition IV.9 to construct a matching on $K_{n,n}$. The key idea is to rewrite the edge weights of $K_{n,n}$ as the minimum of two independent exponential random variables.

Lemma IV.10. *If* $X \sim Exp(\mu_1)$ *and* $Y \sim Exp(\mu_2)$ *are independent, then* $\min(X,Y) \sim Exp(\mu_1+\mu_2)$*.*

Now, for any fixed $0 < \alpha < 1$, we can write

$$\ell_n(i,j') = \min\left(\frac{\ell_n^1(i,j')}{1-\alpha}, \frac{\ell_n^2(i,j')}{\alpha}\right), \tag{4.50}$$

where $\ell_n^1(i,j')$ and $\ell_n^2(i,j')$ are independent copies of $\ell_n(i,j')$. We use $\ell_n^1(i,j')$ to construct the partial matching, and then $\ell_n^2(i,j')$ to patch the partial matching and construct a complete matching. On the event Ω_n^* that the construction is not possible, that is either Proposition IV.8 or Proposition IV.9 failed, we can always use the planted matching to construct a matching on $K_{n,n}$. Since the probability of this failure goes to zero, this does not affect the cost of the matching that much. Specifically, we need to show that $L_n \coloneqq \frac{1}{n}\sum_{i\in[n]} \ell_n(i,i')$ is uniformly integrable.

Lemma IV.11. *There exists a function* $\delta(\cdot)$ *with* $\delta(x) \to 0$ *as* $x \to 0$ *such that for arbitrary events* Ω_n^**,*

$$\limsup_{n\to\infty} \mathbb{E}[L_n \mathbf{1}_{\Omega_n^*}] \le \delta(\epsilon), \tag{4.51}$$

where $\epsilon = \limsup_{n\to\infty} \mathbb{P}[\Omega_n^*]$*.*

Proof. Since $\mathbb{E}[(L_n)^2] < \infty$, L_n is uniformly integrable. In particular, by Cauchy-Schwarz inequality, we have $\mathbb{E}[L_n \mathbf{1}_{\Omega_n^*}] \le \sqrt{\mathbb{E}[(L_n)^2]\mathbb{P}[\Omega_n^*]} = \sqrt{\mathbb{E}[(L_n)^2]\epsilon}$. □

Now, we are ready to present the formal proof of (4.45), which closely follows the proof of Proposition 2 in [20]. Fix $0 < \theta < 1$ and let $\epsilon = \theta^3/200$. Construct $Q_n = [q_{i,j'}]_{i,j'}$ as per (4.49) using the edge cost ℓ_n^1. It follows from Lemma IV.7 that for all sufficiently large

x, B, H, n, $\mathbb{E}[\mathcal{X}(Q_n)] < \epsilon$ and

$$\frac{1}{n}\mathbb{E}[\sum_{i,j'} q_{i,j'}\ell_n^1(i,j')] \le \mathbb{E}[\ell(\mathscr{M}_{\text{opt}}(\emptyset,\emptyset)] + \epsilon. \tag{4.52}$$

Define event $\Omega_n^1 = \{200\mathcal{X}(Q_n) > \theta^2\}$. Then by Markov's inequality,

$$\mathbb{P}[\Omega_n^1] = \mathbb{P}[200\mathcal{X}(Q_n) > \theta^2] \le \frac{200\mathbb{E}[\mathcal{X}(Q_n)]}{\theta^2} \le \frac{200\epsilon}{\theta^2} = \theta.$$

Outside event Ω_n^1, we have $200\mathcal{X}(Q_n) \le \theta^2 < \theta$. It follows from Proposition IV.8 that outside event Ω_n^1, there exists a $(1-\theta)$ partial matching ν_n such that

$$\sum_{i\in U(\nu_n)} q_{i,\nu_n(i)}\ell_n^1(i,\nu_n(i)) \le (1+\theta)\sum_{i\in[n]}\sum_{j'\in[n']} q_{i,j'}\ell_n^1(i,j'), \tag{4.53}$$

where $U(\nu_n)$ is the set of vertices i matched under ν_n. Now, condition on a realization (outside of Ω_n^1) of ℓ_n^1, and apply Proposition IV.9 to ℓ_n^2. Since ℓ_n^1 and ℓ_n^2 are independent, it follows from Proposition IV.9 that there exist an event Ω_n^2 and bijections ν_1 and ν_2 such that for all sufficiently large n, $\mathbb{P}[\Omega_n^2] \le \theta$ and

$$\frac{1}{n}\mathbb{E}\left[1_{(\Omega_n^1)^c}\left(\sum_{i\in[n]\setminus U(\nu_n)}\ell_n^2(i,\nu_2(i)) + \sum_{i\in\nu_n^{-1}(\nu_2([n]\setminus U(\nu_n)))}\ell_n^2(i,\nu_1(i))\right)\right] \le 24\,\theta^{1/2}. \tag{4.54}$$

Outside event $\Omega_n^* = \Omega_n^1 \cup \Omega_n^2$, we can construct a complete matching $\pi_n : [n] \to [n']$ such that $\pi_n(i) = \nu_2(i)$ if $i \in [n]\setminus U(\nu_n)$; and $\pi_n(i) = \nu_1(i)$ if $i \in \nu_n^{-1}(\nu_2([n]\setminus U(\nu_n))$; and $\pi_n(i) = \nu_n(i)$ otherwise. On event Ω_n^*, we just let π_n to be the planted matching. Combining (4.50), (4.52), (4.53), (4.54), and (4.51) yields that

$$\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[\sum_{i,j'}\pi_n(i,j')\ell_n(i,j')] \le \frac{(1+\theta)\left(\mathbb{E}[\ell(\mathscr{M}_{\text{opt}}(\emptyset,\emptyset)] + \epsilon\right)}{1-\alpha} + \frac{24\theta^{1/2}}{\alpha} + \delta(2\theta).$$

Letting θ (and hence ϵ) $\to 0$, then letting $\alpha \to 0$, we establish (4.45).

CHAPTER V

Open Problems and High-Level Takeaways

In this thesis, we discussed three examples of phase transitions in new random graph families. We close the thesis by discussing some open problems related to each chapter, and then providing the high-level takeaways of the thesis.

5.1 Open Problems

5.1.1 Chapter II: Impact of Community Structure on Cascades

1. What if $\boldsymbol{\mu}_* = \lim_{s\to\infty} \boldsymbol{F}^s(\mathbf{1})$ is not an stable equilibrium point of the ODE (2.30)? Intuitively speaking, the process stops at $\boldsymbol{\mu}_*$ with positive probability, or skips to the next fixed point with positive probability. This problem has not been studied in the literature due to the technicality of the problem. Another issue here is to develop the process-level approximation beyond $\boldsymbol{\mu}_*$.

2. How many fixed points does $\boldsymbol{F}(\cdot)$ have? In a related work, Balogh and Pittel [33] showed that the answer is two for regular random graphs (given some additional conditions).

3. What is the trajectory of the contagion process? The evolution of the Markov process of adoption is not the same as the evolution of cascades on the network. Since the random graph model converges locally weakly to GWMT_* defined in Section 2.4, the trajectory of the contagion process on the random graph is related to the trajectory of the contagion process on GWMT_*. Moreover, it is easy to see that the evolution of cascades on GWMT_* is given by the iterations of function $\boldsymbol{F}(\cdot)$ starting from $\mathbf{1}$ (See [36] for a related discussion). This intuitive argument also justifies the connection between $\lim_{s\to\infty} \boldsymbol{F}^s(\mathbf{1})$ and the equilibrium of the ODE (2.30).

4. Finally, what is the optimum seeding strategy? We proposed a heuristic seeding strategy in Section 2.12 that performs well in different scenarios (Figure 2.6). However, we

have no proof that this seeding strategy is either optimum or its gap to optimality if not. This question is also related to the question #2 above.

5.1.2 Chapter III: Erlang Weighted Tree

1. Conditioned on $Z_l > 0$, for $\beta_0 \leq 1$, what is the asymptotic distribution of Z_l as l grows without bound? This problem has been studied for general multi-type branching processes, e.g., [100–102].

2. What is the connection between the reversibility of the continuous state Markov process and the unimodularity of the branching process? Exploring this connection can provide a general framework to study an important class of branching processes.

3. What is the connection between the probability of extinction and the ratio of the giant component in the finite graph model? For other random graph models (e.g. configuration model, Erdos Renyi random graph, etc.) the ratio of the giant component converges to $1 - \mathbb{P}[\{Extinction\}]$, where $\mathbb{P}[\{Extinction\}]$ is the probability the associated branching process goes extinct eventually. We have observed the same relation via numerical simulation in Figures 3.6 and 3.7 between the finite graph model and the EWT.

4. What is the local weak limit if vertices in the finite graph model iterate to use all their budget, given by their potential degree? Naturally, one can imagine a scenario in which after the realization of G_n, all vertices i with degree less than $d_i(n)$ have a second chance to find more neighbors by announcing an updated set of potential neighbors. Of particular interest is the case when vertices can iterate as many times as possible until they achieve $d_i(n)$ or have checked all other $n - 1$ nodes.

5. How general is the methodology we developed in this thesis for finding the Krein-Rutman eigenvalue and the corresponding eigenfunction?

6. Finally, what is the connectivity threshold of the random graph model, when the potential degree of all the vertices are same, and is equal to $k(n) = c \log(n)$? In [23, Conjecture 1], it is conjectured that the phase transition happens at $c = 1$, i.e., the probability of the event {the graph is connected} goes to one if $c > 1$ and goes to zero if $c < 1$.

5.1.3 Chapter IV: Planted Matching Problem

1. We have focused here on the maximum-likelihood estimator, which for the exponential distribution is also the minimum-weight matching. In physical terms, we have considered this problem at zero temperature. In contrast, the posterior distribution $\mathbb{P}[M^* = M' \mid G]$ given in

(4.2) is a Gibbs distribution at nonzero temperature. The estimator with the largest expected overlap would then be the maximum marginal estimator, i.e., the set of edges e for which $\mathbb{P}[e \in M^* \mid G] \geq \mathbb{P}[e \notin M^* \mid G]$. This estimator is not generally a matching or even of size n; nevertheless, one can restrict to estimators which are perfect matchings while increasing the expected misclassification rate $|M' \triangle M^*|/(2n)$ at most by a factor of two. We leave for future work the problem of computing the expected overlap of this estimator. It is possible that it achieves almost-perfect recovery for some $\lambda < 4$, i.e., that the information-theoretic threshold for almost-perfect recovery is different from the threshold we have computed here, but we conjecture this is not the case.

2. In physics, a phase transition is called *continuous* if the order parameter (in this case, the overlap) is continuous at the threshold, and as *pth order* if its $(p-2)$th derivative is continuous. Although we have not proved this, $\alpha(\lambda)$ in Figure 4.3 appears to have zero derivative at $\lambda = 4$. This suggests that the transition in the optimal overlap is of third or higher order, unlike other well-known problems in random graphs such as the emergence of the giant component (second order) [12], the stochastic block model with two groups (second order) or with four or more groups (first order) [82], or the appearance of the k-core for $k \geq 3$ (first order) [103].

3. A related question is how the minimum matching changes when the graph undergoes a small perturbation. Aldous and Percus [104] introduced this problem formally and classified combinatorial optimization problems based on how the cost of the optimal solution scales with the size of the perturbation. Using a cavity–based analysis and Monto Carlo simulation, they suggested that the minimum cost among all perfect matchings that differ from the minimum matching by at least δn edges is larger than the cost of the minimum matching by $\Theta(\delta^3)$. This framework has been studied rigorously in [105] and [106] for different combinatorial optimization problems. It would be interesting to explore this same kind of perturbation in the planted model, where we study the minimum cost among all matchings that differ from the planted matching by at least δn edges.

4. Does the overlap of $M_{\min}$ with M^* concentrate around its expectation? This would follow from correlation decay of messages in the planted PWIT, but we have not proved this.

5. Can Lemma IV.4 be turned into a proof of almost perfect recovery? More generally, when the RDEs (4.5) and (4.6) lack a solution supported on $\mathbb{R}$ (i.e., excluding weights in $\{\pm\infty\}$) does this imply almost perfect recovery?

6. We have given two proofs that that the RDEs have a unique solution if $\lambda < 4$. Theorem C.1 uses the dynamics of the ODEs, while Theorem IV.11 uses the uniqueness of $\mathscr{M}_{\infty,\text{opt}}$

on the planted PWIT. These two types of reasoning seem completely orthogonal, but they must be connected. When do the properties of the optimum involution invariant object on an appropriate type of infinite tree imply the dynamical fact that a system of RDEs has a unique fixed point?

7. What can we say about distributions $P(w)$ of planted weights other than exponential? For what distributions is it possible to collapse the RDEs into a finite-dimensional system of ODEs? As stated above, Chertkov et al. [1] studied the folded Gaussian distribution $P = |\mathcal{N}(0, \kappa)|$, but we have been unable to reduce the RDEs to ODEs in this case. Nevertheless, in the spirit of universality classes in physics, we expect any reasonable family of distributions P to undergo a phase transition similar to what we have shown here for the exponential distribution, namely from almost-perfect to partial recovery at some critical value of P's expectation (where this critical value may depend on the shape of the distribution P). Moreover, with respect to question #2 above, we expect the order of this phase transition, and other scaling properties in its vicinity, to be robust as long as $P'(0) > 0$.

8. Finally, what about planted models with spatial structure, as in the original problem of particle tracking from [1]?

5.2 High-Level Takeaways

In this thesis, we studied phase transitions in three new random graph families. We have used different techniques in each case to characterize the phase transition and to understand the behavior of the random graph before and after the critical value. We close the thesis by discussing the main ideas developed in each technical chapter.

In chapter II, we used the differential equation method to provide a tight-approximation for the trajectory of the Markov process of adoption. The differential equation method is specifically useful in the study of discrete processes when the expected drift of random variables of interest results in a simple ODE. However, the expected changes in the Markov process of adoption is highly non-linear, and hence, the dimension of the resulted ODE is high too. Nevertheless, we solve this ODE using a probabilistic method, which shows that the dimension of the ODE is superficially high. The form of the solution is also related to the mean-field equations. The differential equation method can be used in a variety of problems to prove the correctness of mean-field type approximations.

In chapter III, we used a point process perspective to analyze the Erlang Weighted Tree. The first and the second moment methods are the most common tools in the study of branching processes. However, the interdependent structure of the EWT causes these

methods to fail. Using a point process perspective, we found the Krein-Rutman eigenvalue and the corresponding eigenfunction of the growth operator of the EWT. We also explored the connection between this eigenvalue and the probability of the extinction of the branching process and hence, random graphs. This method is quite general in the study of branching processes. An important property that made the analysis possible is the unimodularity of the branching process. We believe the techniques developed in this chapter are applicable to other unimodular processes.

Finally, in chapter IV, we used the machinery of local weak convergence to analyze the performance of the minimum matching algorithm to find the planted matching in $K_{n,n}$. This framework has been used for different combinatorial optimization problems. This technique is applicable as long as there is a correlation decay; that is to say, the solution of the optimization problem is mostly given by the local structure of the random graph. Note that, in general, the optimization problem may depend on the whole graph. However, the structural properties of the random graph model may dictate the optimum solution in a local manner. This makes it possible to solve the optimization problem on the asymptotic object and then to use this solution to provide a near-optimum answer for the finite random graph.

APPENDICES

APPENDIX A

Appendix of Chapter II

A.1 When Greedy Maximization Is Bad

We now give an example showing how the results in [65, 66] break if θ_v is assumed to be fixed. We build a network as follows: start from a $3n \times 3n$ torus, i.e. node (i, j) with $1 \leq i, j \leq 3n$ has four neighbors: $(i+1, j), (i-1, j), (i, j+1), (i, j-1)$ where operation are done modulo $3n$. Now for each $1 \leq j \leq 3n$, and $0 \leq k \leq n-1$, we add a node $v(j, k)$ connected to the nodes of the torus $(3k+1, j)$, $(3k+2, j)$ and $(3k+3, j)$. Finally each of these nodes $v(j, k)$ are part of a cycle of size $K \geq 3$ with no other common point with the rest of the graph except through $v(j, k)$. In summary, we have $9n^2$ nodes on the torus, and $3n^2$ disjoint cycles of size K which are connected to the torus only through the nodes $v(j, k)$. There is a total of $9n^2 + 3n^2K$ nodes. Note that the degree of a node on the torus is 5 (4 neighbors on the torus and 1 on a cycle) as well as for the nodes $v(j, k)$. We take $\theta = 2/5$ so that a node of degree d becomes active as soon as θd of its neighbors are active. In particular a node on the torus or a $v(j, k)$ needs only 2 active neighbors to become active. Moreover, activating a node $v(j, k)$ will activate all the K nodes on the cycle. Because of this, it is easy to see that any greedy algorithm with budget $b \leq 3n^2$ will only activate the nodes $v(j, k)$. Note however that by acting the set of nodes on the torus: $(1, 1), (1, 2), \ldots, (1, 3n)$ and $(2, 1)$ will result in a global activation of the network. Hence for any $3n + 1 \leq b \leq 3n^2$, we can find a set activating the $9n^2 + 3n^2K$ nodes of the networks, whereas the greedy algorithm only activate Kb nodes which is far from the optimum solution.

A.2 One-Step Drift

Recall that $m_j(n)$ denotes the total number of edges on side $j \in \{1,2\}$, and $m_m(n)$ denotes the total number of edges between the two communities. The one-step drift of the random variables associated with the Markov process of adoption are given as follows:

- One-step drift of $A_j(\cdot)$ for $j \in \{1,2\}$:

$$\begin{aligned}
\mathbb{E}[A_j(k+1) - A_j(k) | X^n(k)] = \\
&- \frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \\
&- \frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{A_j(k) - 1}{2m_j(n) - 2T_j(k) - 1} \\
&+ \frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \\
&\qquad \sum_{u_j + u_{-j} = K_j(d_j, d_{-j})} (d_j - u_j - 1) \times \frac{(d_j - u_j) I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k)}{2m_j(n) - 2T_j(k) - 1} \\
&+ \frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \\
&\qquad \sum_{u_j + u_{-j} = K_j(d_j, d_{-j})} (d_j - u_j) \times \frac{(d_{-j} - u_{-j}) I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k)}{m_m(n) - (k - T_1(k) - T_2(k))}
\end{aligned}$$

- One-step drift of $T_j(\cdot)$ for $j \in \{1,2\}$:

$$\mathbb{E}[T_j(k+1) - T_j(k) | X^n(k)] = + \frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)}$$

- One-step drift of $A_m^{(j)}(\cdot)$ for $j \in \{1,2\}$:

$$\begin{aligned}
\mathbb{E}[A_m^{(j)}(k+1) - A_m^{(j)}(k) | X^n(k)] = \\
&- \frac{A_m^{(j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \\
&- \frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{A_m^{(j)}(k)}{m_m(n) - (k - T_1(k) - T_2(k))} \\
&+ \frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times
\end{aligned}$$

$$\sum_{u_j+u_{-j}=K_j(d_j,d_{-j})} (d_{-j}-u_{-j}-1)\times\frac{(d_{-j}-u_{-j})I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)}{m_m(n)-(k-T_1(k)-T_2(k))}$$
$$+\frac{A_j(k)}{A_1(k)+A_2(k)+A^{(1)}_m(k)+A^{(2)}_m(k)}\times$$
$$\sum_{u_j+u_{-j}=K_j(d_j,d_{-j})} (d_{-j}-u_{-j})\times\frac{(d_j-u_j)I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)}{2m_j(n)-2T_j(k)-1}$$

- One-step drift of $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(\cdot)$ for $j\in\{1,2\}$:

$$\mathbb{E}[I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k+1)-I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)|X^n(k)]=$$
$$-\frac{A_j(k)}{A_1(k)+A_2(k)+A^{(1)}_m(k)+A^{(2)}_m(k)}\times\frac{(d_j-u_j)I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)}{2m_j(n)-2T_j(k)-1}$$
$$-\frac{A^{(-j)}_m(k)}{A_1(k)+A_2(k)+A^{(1)}_m(k)+A^{(2)}_m(k)}\times\frac{(d_{-j}-u_{-j})I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k)}{m_m(n)-(k-T_1(k)-T_2(k))}$$
$$+\frac{A_j(k)}{A_1(k)+A_2(k)+A^{(1)}_m(k)+A^{(2)}_m(k)}\times\frac{(d_j-u_j+1)I^{(j)}_{d_j,d_{-j},u_j-1,u_{-j}}(k)}{2m_j(n)-2T_j(k)-1}$$
$$+\frac{A^{(-j)}_m(k)}{A_1(k)+A_2(k)+A^{(1)}_m(k)+A^{(-j)}_m(k)}\times\frac{(d_{-j}-u_{-j}+1)I^{(j)}_{d_j,d_{-j},u_j,u_{-j}-1}(k)}{m_m(n)-(k-T_1(k)-T_2(k))}$$

Although we've presented the one-step drift of A_j and $A^{(j)}_m$ for $j\in\{1,2\}$, we are not going to use them. As is pointed out in Remark II.2, we only need to keep track of T_j and $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}$. In particular, using the balance equations, we replace all the terms A_j and $A^{(j)}_m$ by an affine function of T_j and $I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}$.

A.3 Derivation of ODEs

The scaled variables are supposed to model the behavior of their discrete counterpart, as we mentioned in (2.13). Using assumption 2 of Definition II.3, Remark II.7, Point 2 in

Section 2.6, Remark II.1, and one-step drifts in Appendix A.2, the ODEs are given as follows:

$$\begin{aligned}\frac{di^{(j)}_{d_j,d_{-j},u_j,u_{-j}}}{dx} &= f_{j,d_j,d_{-j},u_j,u_{-j}}(x,\tau_1,\tau_2,w_1,w_2,w_m^{(1)},w_m^{(2)},i^{(1)}_{d_1,d_2,u_1,u_2},i^{(2)}_{d_2,d_1,u_2,u_1}) \coloneqq \\ &- \frac{a_j(x)}{a_1(x)+a_2(x)+a_m^{(1)}(x)+a_m^{(2)}(x)} \times \frac{(d_j-u_j)i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(x)}{\lambda_j-2\tau_j(x)} \\ &- \frac{a_m^{(-j)}(x)}{a_1(x)+a_2(x)+a_m^{(1)}(x)+a_m^{(2)}(x)} \times \frac{(d_{-j}-u_{-j})i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(x)}{\lambda_m-\tau_m(x)} \\ &+ \frac{a_j(x)}{a_1(x)+a_2(x)+a_m^{(1)}(x)+a_m^{(2)}(x)} \times \frac{(d_j-u_j+1)i^{(j)}_{d_j,d_{-j},u_j-1,u_{-j}}(x)}{\lambda_j-2\tau_j(x)} \\ &+ \frac{a_m^{(-j)}(x)}{a_1(x)+a_2(x)+a_m^{(1)}(x)+a_m^{(2)}(x)} \times \frac{(d_{-j}-u_{-j}+1)i^{(j)}_{d_j,d_{-j},u_j,u_{-j}-1}(x)}{\lambda_m-\tau_m(x)},\end{aligned} \tag{A.1}$$

$$\begin{aligned}\frac{d\tau_j}{dx} &= f_j(x,\tau_1,\tau_2,w_1,w_2,w_m^{(1)},w_m^{(2)},i^{(1)}_{d_1,d_2,u_1,u_2},i^{(2)}_{d_2,d_1,u_2,u_1}) \coloneqq \\ &\frac{a_j(x)}{a_1(x)+a_2(x)+a_m^{(1)}(x)+a_m^{(2)}(x)},\end{aligned} \tag{A.2}$$

$$\begin{aligned}\frac{dw_j}{dx} &= f_{j+2}(x,\tau_1,\tau_2,w_1,w_2,w_m^{(1)},w_m^{(2)},i^{(1)}_{d_1,d_2,u_1,u_2},i^{(2)}_{d_2,d_1,u_2,u_1}) \coloneqq \\ &\frac{a_j(x)}{a_1(x)+a_2(x)+a_m^{(1)}(x)+a_m^{(2)}(x)} \times \frac{-w_j(x)}{\lambda_j-2\tau_j(x)},\end{aligned} \tag{A.3}$$

$$\begin{aligned}\frac{dw_m^{(j)}}{dx} &= f_{j+4}(x,\tau_1,\tau_2,w_1,w_2,w_m^{(1)},w_m^{(2)},i^{(1)}_{d_1,d_2,u_1,u_2},i^{(2)}_{d_2,d_1,u_2,u_1}) \coloneqq \\ &\frac{a_m^{(-j)}(x)}{a_1(x)+a_2(x)+a_m^{(1)}(x)+a_m^{(2)}(x)} \times \frac{-w_m^{(j)}(k)}{\lambda_m-\tau_m(x)},\end{aligned} \tag{A.4}$$

where $\tau_m(x) \coloneqq x-\tau_1(x)-\tau_2(x)$,

$$a_j(x) \coloneqq -\sum_{\substack{u_j+u_{-j}\le K_j(d_j,d_{-j})\\ d_j+d_{-j}\le d_{\max}}} (d_j-u_j)i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(x)+\lambda_j-2\tau_j(x)-w_j(x), \text{ and}$$

$$a_m^{(j)}(x) \coloneqq -\sum_{\substack{u_j+u_{-j}\le K_j(d_j,d_{-j})\\ d_j+d_{-j}\le d_{\max}}} (d_{-j}-u_{-j})i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(x)+\lambda_m-\tau_m(x)-w_m^{(j)}(x).$$

The initial conditions are given as follows for $j \in \{1, 2\}$:

$$\begin{aligned}
&\tau_j(0) = 0, \\
&w_j(0) = \sum_{d_j + d_{-j} > d_{\max}} d_j \frac{I^{(j)}_{d_j, d_{-j}, 0, 0}(0)}{n}, \\
&w^{(j)}_m(0) = \sum_{d_j + d_{-j} > d_{\max}} d_{-j} \frac{I^{(j)}_{d_j, d_{-j}, 0, 0}(0)}{n}, \\
&i^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(0) = \begin{cases} \frac{1}{n} I^{(j)}_{d_j, d_{-j}, 0, 0}(0) & \text{if } u_j = u_{-j} = 0 \text{ and } d_j + d_{-j} \le d_{\max} \\ 0 & \text{otherwise} \end{cases}.
\end{aligned} \tag{A.5}$$

Note that if $\alpha_j(d_j, d_{-j}) = 1$ for all $d_j + d_{-j} > d_{\max}$, then $w_j(x) = w^{(j)}_m(x) = 0$, $\forall x$.

A.4 Proofs of Theorems and Lemmas

A.4.1 Proof of Theorem II.2

By the assumptions of Theorem II.2, we have

$$\Big(\frac{0}{n}, \frac{T_1(0)}{n}, \frac{T_2(0)}{n}, \frac{W_1(0)}{n}, \frac{W_2(0)}{n}, \frac{W^{(1)}_m(0)}{n}, \frac{W^{(2)}_m(0)}{n}, \frac{I^{(1)}_{d_1,d_2,u_1,u_2}(0)}{n}, \frac{I^{(2)}_{d_2,d_1,u_2,u_1}(0)}{n}\Big) \in \mathcal{D}_\epsilon.$$

Moreover, it is easy to see that the functions given in Appendix A.3 satisfies a Lipschitz condition on

$$\mathcal{D}_\epsilon \cap \{(t, \tau_1, \tau_2, w_1, w_2, w^{(1)}_m, w^{(2)}_m, i^{(1)}_{d_1,d_2,u_1,u_2}, i^{(2)}_{d_2,d_1,u_2,u_1}) : t \ge 0\}$$

with the same Lipschitz constant ("*Lipschitz hypothesis*"). Also, for $\theta_1 = O(n^{-\eta})$

$$\begin{aligned}
&\Big|\mathbb{E}(T_j(k+1) - T_j(k) \mid X^n(k)) - \\
&\qquad f_j\Big(\frac{k}{n}, \frac{T_1(k)}{n}, \frac{T_2(k)}{n}, \cdots, \frac{I^{(1)}_{d_1,d_2,u_1,u_2}(k)}{n}, \frac{I^{(2)}_{d_2,d_1,u_2,u_1}(k)}{n}\Big)\Big| \le \theta_1, \\
&\Big|\mathbb{E}(W_j(k+1) - W_j(k) \mid X^n(k)) - \\
&\qquad f_{j+2}\Big(\frac{k}{n}, \frac{T_1(k)}{n}, \frac{T_2(k)}{n}, \cdots, \frac{I^{(1)}_{d_1,d_2,u_1,u_2}(k)}{n}, \frac{I^{(2)}_{d_2,d_1,u_2,u_1}(k)}{n}\Big)\Big| \le \theta_1,
\end{aligned}$$

$$\Big|\mathbb{E}(W_m^{(j)}(k+1) - W_m^{(j)}(k) \mid X^n(k)) - f_{j+4}\Big(\frac{k}{n}, \frac{T_1(k)}{n}, \frac{T_2(k)}{n}, \cdots, \frac{I^{(1)}_{d_1,d_2,u_1,u_2}(k)}{n}, \frac{I^{(2)}_{d_2,d_1,u_2,u_1}(k)}{n}\Big)\Big| \leq \theta_1,$$

and,

$$\Big|\mathbb{E}(I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k+1) - I^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(k) \mid X^n(k)) - f_{j,d_j,d_{-j},u_j,u_{-j}}\Big(\frac{k}{n}, \frac{T_1(k)}{n}, \frac{T_2(k)}{n}, \cdots, \frac{I^{(1)}_{d_1,d_2,u_1,u_2}(k)}{n}, \frac{I^{(2)}_{d_2,d_1,u_2,u_1}(k)}{n}\Big)\Big| \leq \theta_1,$$

for all $k < T_{\mathcal{D}_\epsilon}$, where $T_{\mathcal{D}_\epsilon}$ is the minimum $k > 0$ such that

$$\Big(\frac{k}{n}, \frac{T_1(k)}{n}, \frac{T_2(k)}{n}, \frac{W_1(k)}{n}, \frac{W_2(k)}{n}, \frac{W_m^{(1)}(k)}{n}, \frac{W_m^{(2)}(k)}{n}, \frac{I^{(1)}_{d_1,d_2,u_1,u_2}(k)}{n}, \frac{I^{(2)}_{d_2,d_1,u_2,u_1}(k)}{n}\Big) \notin \mathcal{D}_\epsilon$$

("*Trend hypothesis*"). Finally, the changes for each random variable in the successive steps of the Markov process of adoption is bounded by 1 ("*Bounded hypothesis*"). Now, Theorem II.2 follows by the direct application of Wormald's Theorem [19, Theorem 5.1].

A.4.2 Proof of Lemma II.1

The proof follows by plugging in the form of the solution and check the validity of the corresponding differential equations. Before plugging in the form of the solution, let us prove the equality (2.19). Summing up (2.21) and (2.22) for $j \in \{1, 2\}$, we have

$$\lambda_1 \frac{d\mu^{(1,1)}}{dt}\left(\mu^{(1,1)}(t)\right) + \lambda_2 \frac{d\mu^{(2,2)}}{dt}\left(\mu^{(2,2)}(t)\right) + \lambda_m \frac{d\mu^{(1,2)}}{dt}\left(\mu^{(2,1)}(t)\right) + \lambda_m \frac{d\mu^{(2,1)}}{dt}\left(\mu^{(1,2)}(t)\right) = -1.$$

Equality (2.19) follows by integrating the both sides of the above equation from 0 to t, where the constant of integration is determined by the initial condition (2.23). Note that (2.19) suggests $\tau_m(t) = t - \tau_1(t) - \tau_2(t)$, where $\tau_m(t)$ is defined by (2.18). Also, by (2.17), (2.21), and (2.22) we have

$$\begin{aligned}\frac{d\mu^{(j,j)}}{dt}\left(\mu^{(j,j)}(t)\right)^{-1} &= \lambda_j \frac{d\mu^{(j,j)}}{dt}\left(\mu^{(j,j)}(t)\right) \times \left(\lambda_j \mu^{(j,j)}(t)^2\right)^{-1} \\ &= \frac{-a_j(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} \times \frac{1}{\lambda_j - 2\tau_j(t)}, \end{aligned} \tag{A.6}$$

and

$$\frac{d\mu^{(j,-j)}}{dt}\left(\mu^{(j,-j)}(t)\right)^{-1} = \lambda_m \frac{d\mu^{(j,-j)}}{dt}\mu^{(-j,j)}(t)\left(\lambda_m \mu^{(-j,j)}(t)\mu^{(j,-j)}(t)\right)^{-1}$$
$$= \frac{-a_m^{(-j)}(t)}{a_1(t)+a_2(t)+a_m^{(1)}(t)+a_m^{(2)}(t)} \times \frac{1}{\lambda_m - (t-\tau_1-\tau_2)}. \qquad \text{(A.7)}$$

Consider the form of $i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t)$ given by (2.16). We have

$$\begin{aligned}
\frac{di^{(j)}_{d_j,d_{-j},u_j,u_{-j}}}{dt} = i^{(j)}_{d_j,d_{-j},0,0}(0)\times \\
\Bigg\{ &+ (d_j-u_j)\frac{d\mu^{(j,j)}}{dt}\binom{d_j}{u_j}\left(1-\mu^{(j,j)}(t)\right)^{u_j}\left(\mu^{(j,j)}(t)\right)^{d_j-u_j-1} \\
&\qquad \times Bi\left(u_{-j};d_{-j},1-\mu^{(j,-j)}(t)\right) \\
&+ (d_{-j}-u_{-j})\frac{d\mu^{(j,-j)}}{dt}\binom{d_{-j}}{u_{-j}}\left(1-\mu^{(j,-j)}(t)\right)^{u_{-j}} \times \left(\mu^{(j,-j)}(t)\right)^{d_{-j}-u_{-j}-1} \\
&\qquad \times Bi\left(u_j;d_j,1-\mu^{(j,j)}(t)\right) \\
&- u_j\frac{d\mu^{(j,j)}}{dt}\binom{d_j}{u_j}\left(1-\mu^{(j,j)}(t)\right)^{u_j-1}\left(\mu^{(j,j)}(t)\right)^{d_j-u_j} \\
&\qquad \times Bi\left(u_{-j};d_{-j},1-\mu^{(j,-j)}(t)\right) \\
&- u_{-j}\frac{d\mu^{(j,-j)}}{dt}\binom{d_{-j}}{u_{-j}}\left(1-\mu^{(j,-j)}(t)\right)^{u_{-j}-1}\left(\mu^{(j,-j)}(t)\right)^{d_{-j}-u_{-j}} \\
&\qquad \times Bi\left(u_j;d_j,1-\mu^{(j,j)}(t)\right) \Bigg\}.
\end{aligned}$$

Using (2.16), we have

$$\begin{aligned}
\frac{di^{(j)}_{d_j,d_{-j},u_j,u_{-j}}}{dt} = \; & (d_j-u_j)\times\frac{d\mu^{(j,j)}}{dt}\left(\mu^{(j,j)}(t)\right)^{-1}\times i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t) \\
&+ (d_{-j}-u_{-j})\times\frac{d\mu^{(j,-j)}}{dt}\left(\mu^{(j,-j)}(t)\right)^{-1}\times i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t) \\
&- (d_j-u_j+1)\times\frac{d\mu^{(j,j)}}{dt}\left(\mu^{(j,j)}(t)\right)^{-1}\times i^{(j)}_{d_j,d_{-j},u_j-1,u_{-j}}(t) \\
&- (d_{-j}-u_{-j}+1)\times\frac{d\mu^{(j,-j)}}{dt}\left(\mu^{(j,-j)}(t)\right)^{-1}\times i^{(j)}_{d_j,d_{-j},u_j,u_{-j}-1}(t).
\end{aligned}$$

Now (A.1) follows by plugging in (A.6) and (A.7) into the above equality. Next, consider

the function $\tau_j(t)$ given by (2.17). It is easy to see that,

$$\frac{d\tau_j}{dt} = -\lambda_j \mu^{(j,j)} \frac{d\mu^{(j,j)}}{dt} = \frac{a_j(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)}$$

Finally, for the functions $w_j(t)$ and $w_m^{(j)}(t)$ given by (2.20), we have

$$\frac{dw_j}{dt} = w_j(0)\frac{d\mu^{(j,j)}}{dt} = w_j(t)\frac{d\mu^{(j,j)}}{dt}\left(\mu^{(j,j)}(t)\right)^{-1}$$
$$\frac{dw_m^{(j)}}{dt} = w_m^{(j)}(0)\frac{d\mu^{(j,-j)}}{dt} = w_m^{(j)}(t)\frac{d\mu^{(j,-j)}}{dt}\left(\mu^{(j,-j)}(t)\right)^{-1}$$

Now, using (A.6) and (A.7), the equations (A.3) and (A.4) follows.

A.4.3 Proof of Lemma II.4

Let $Y(x) \coloneqq Bi(u, d, 1-x)$. Then we have:

$$\begin{aligned}\frac{\mathrm{d}Y}{\mathrm{d}x} &= \binom{d}{u}\left((d-u)\times x^{d-u-1}(1-x)^u - u \times x^{d-u}(1-x)^{u-1}\right)\\ &= d \times \left(Bi(u, d-1, 1-x) - Bi(u-1, d-1, 1-x)\right)\end{aligned}$$

Now, the proof follows by straightforward algebraic simplification.

$$\begin{aligned}\frac{\partial F_{(j,j)}}{\partial \mu^{(j,j)}} &= \sum_{\substack{u_j+u_{-j}\le K_j(d_j,d_{-j})\\ d_j+d_{-j}\le d_{\max}}} \frac{d_j}{\lambda_j}\, i^{(j)}_{d_j,d_{-j},0,0}(0)\, Bi(u_{-j}; d_{-j}, 1-\mu^{(j,-j)}) \times (d_j - 1)\\ &\qquad \left(Bi(u_j; d_j-2, 1-\mu^{(j,j)}) - Bi(u_j-1; d_j-2, 1-\mu^{(j,j)})\right)\\ &= \sum_{\substack{u_{-j}\le K_j(d_j,d_{-j})\\ d_j+d_{-j}\le d_{\max}}} \frac{d_j}{\lambda_j}\, i^{(j)}_{d_j,d_{-j},0,0}(0)\, Bi(u_{-j}; d_{-j}, 1-\mu^{(j,-j)}) \times (d_j - 1)\\ &\qquad Bi(K_j(d_j,d_{-j}) - u_{-j}; d_j-2, 1-\mu^{(j,j)}) > 0\end{aligned}$$

$$\begin{aligned}\frac{\partial F_{(j,j)}}{\partial \mu^{(j,-j)}} &= \sum_{\substack{u_j+u_{-j}\le K_j(d_j,d_{-j})\\ d_j+d_{-j}\le d_{\max}}} \frac{d_j}{\lambda_j}\, i^{(j)}_{d_j,d_{-j},0,0}(0)\, Bi(u_j; d_j-1, 1-\mu^{(j,j)}) \times d_{-j}\\ &\qquad \left(Bi(u_{-j}; d_{-j}-1, 1-\mu^{(j,-j)}) - Bi(u_{-j}-1; d_{-j}-1, 1-\mu^{(j,-j)})\right)\\ &= \sum_{\substack{u_j\le K_j(d_j,d_{-j})\\ d_j+d_{-j}\le d_{\max}}} \frac{d_j}{\lambda_j}\, i^{(j)}_{d_j,d_{-j},0,0}(0)\, Bi(u_j; d_j-1, 1-\mu^{(j,j)}) \times d_{-j}\\ &\qquad Bi(K_j(d_j,d_{-j}) - u_j; d_{-j}-1, 1-\mu^{(j,-j)}) > 0\end{aligned}$$

Similar inequalities holds for other cases.

A.4.4 Proof of Lemma II.5

(i) Fix $\boldsymbol{\mu} \in \mathcal{U}$. Consider the closed, convex, and compact set

$$\mathcal{S} = \left\{ \boldsymbol{x} \in [0,1]^4 : \boldsymbol{F}(\boldsymbol{\mu}) \leq \boldsymbol{x} \leq \boldsymbol{\mu} \right\},$$

where the inequalities are interpreted component-wise. Note that $\mathcal{S}$ is a hyperrectangle. By Lemma II.4, $\boldsymbol{F}(\boldsymbol{s}) \leq \boldsymbol{F}(\boldsymbol{\mu}) \leq \boldsymbol{s}$ for all $\boldsymbol{s} \in \mathcal{S}$ since $\boldsymbol{s} \leq \boldsymbol{\mu}$. Hence, $\mathcal{S}$ is a subset of $\mathcal{U}$.

(ii) Define $\boldsymbol{G}(\boldsymbol{x}) \coloneqq \boldsymbol{x} - \boldsymbol{F}(\boldsymbol{x})$ for $\boldsymbol{x} \in [0,1]$. Clearly, the set $\mathcal{A} = \{\boldsymbol{x} : \boldsymbol{G}(\boldsymbol{x}) \geq \boldsymbol{0}\}$ is a closed set as $\boldsymbol{G}(\cdot)$ is a continuous function. Since $\mathcal{U}$ is the largest connected component of $\mathcal{A}$ that contains $\boldsymbol{1} = (1,1,1,1)$, $\mathcal{U}$ is closed as well. Now, compactness follows from the fact that $\mathcal{U}$ is bounded.

(iii) Consider the sequence $\left\{\boldsymbol{F}^k(\boldsymbol{u})\right\}_{k=1}^{\infty}$ for some $\boldsymbol{u} \in \mathcal{U}$. Since $\boldsymbol{F}(\mathcal{U}) \subset \mathcal{U}$, we have $\boldsymbol{F}^k(\boldsymbol{u}) \in \mathcal{U}$ for all k. By compactness of $\mathcal{U}$, this sequence has a subsequence that converges to a point $\boldsymbol{u}_* \in \mathcal{U}$. Now the result follows by the fact that $\boldsymbol{F}^{k+1}(\boldsymbol{u}) \leq \boldsymbol{F}^k(\boldsymbol{u})$ for all $k \geq 0$, where $\boldsymbol{F}^0(\boldsymbol{u}) \coloneqq \boldsymbol{u}$.

(iv) Recall that $\boldsymbol{F} \coloneqq (F_{(1,1)}, F_{(1,2)}, F_{(2,1)}, F_{(2,2)})$, and for any $\boldsymbol{\mu} \in [0,1]^4$ we use the notation $\boldsymbol{\mu} = \left(\mu^{(1,1)}, \mu^{(1,2)}, \mu^{(2,1)}, \mu^{(2,2)}\right)$. Also, recall that $F_{(j,j)}$ is a function of $\mu^{(j,j)}$ and $\mu^{(j,-j)}$, and $F_{(j,-j)}$ is a function of $\mu^{(-j,-j)}$ and $\mu^{(-j,j)}$, for $j \in \{1,2\}$. Now, by symmetry there are two cases that we need to consider:

(a) $u^{(j,j)} = u_*^{(j,j)}$: If $u^{(j,-j)} > u_*^{(j,-j)}$, then by Lemma II.4 and equality $F_{(j,j)}(u_*^{(j,j)}, u_*^{(j,-j)}) = u_*^{(j,j)}$, we have $u^{(j,j)} < F_{(j,j)}(u^{(j,j)}, u^{(j,-j)})$. If $u^{(j,-j)} = u_*^{(j,-j)}$, then either $u^{(-j,j)} > u_*^{(-j,j)}$ or $u^{(-j,-j)} > u_*^{(-j,-j)}$. Now, if $u^{(-j,-j)} > u_*^{(-j,-j)}$ then by the same argument we have $u^{(j,-j)} < F_{(j,-j)}(u^{(-j,-j)}, u^{(j,-j)})$. Otherwise, we have $u^{(-j,j)} > u_*^{(-j,j)}$ and by the same argument $u^{(-j,-j)} < F_{(-j,-j)}(u^{(-j,-j)}, u^{(-j,j)})$.

(b) $u^{(j,-j)} = u_*^{(j,-j)}$: The argument is exactly the same as the previous case, and we avoid repetition.

A.4.5 Proof of Theorem II.7

By Corollary II.6, $\boldsymbol{\mu}_* = \lim_{n\to\infty} \boldsymbol{F}^n(\boldsymbol{1})$ is a fixed point of $\boldsymbol{F}$. Let

$$\mathcal{N} \coloneqq \mathcal{U} \cap \left\{ \boldsymbol{x} \in [0,1]^4 : \boldsymbol{\mu}_* \leq \boldsymbol{x} \leq \boldsymbol{1} \right\}.$$

For any arbitrary point $\boldsymbol{u} \in \mathcal{N}$, define $\mathcal{S}_{\boldsymbol{u}} \coloneqq \{x : \boldsymbol{F}(\boldsymbol{u}) \leq x \leq \boldsymbol{u}\}$. By the proof of Lemma II.5 part (i), $\mathcal{S}_{\boldsymbol{u}}$ is a subset of $\mathcal{U}$. Moreover, if $\boldsymbol{u} \neq \boldsymbol{\mu}_*$, then $\boldsymbol{F}(\boldsymbol{u}) \geq \boldsymbol{F}(\boldsymbol{\mu}_*)$ and hence $\mathcal{S}_{\boldsymbol{u}}$ is a subset of $\mathcal{N}$. So we have $\boldsymbol{F}(\mathcal{N}) \subset \mathcal{N}$. It is also easy to see that $\mathcal{N}$ is closed and compact, and $\boldsymbol{\mu}_*$ is the unique fixed point of $\boldsymbol{F}$ in $\mathcal{N}$.

Now, consider the ODE (2.30). Note that the initial conditions lies in $\mathcal{N}$. Moreover, if $\boldsymbol{\mu}(t) \in \mathcal{N}$, then $\dot{\boldsymbol{\mu}}(t) = \boldsymbol{F}(\boldsymbol{\mu}(t)) - \boldsymbol{\mu}(t)$ is directing toward $\mathcal{N}$ as $\boldsymbol{\mu}(t) + \delta(\boldsymbol{F}(\boldsymbol{\mu}(t)) - \boldsymbol{\mu}(t)) \in \mathcal{S}_{\boldsymbol{\mu}(t)} \subset \mathcal{N}$ for all $\delta \in [0, 1]$. Hence, $\mathcal{N}$ is a positive invariant set.

Consider the function $V(\boldsymbol{\mu}) \coloneqq (\boldsymbol{\mu} - \boldsymbol{\mu}_*)(\boldsymbol{\mu} - \boldsymbol{\mu}_*)^T$. Note that $\forall \boldsymbol{\mu} \in \mathcal{N} \setminus \{\boldsymbol{\mu}_*\}$, we have

$$\begin{aligned}
\frac{\dot{V}(\boldsymbol{\mu})}{2} &= \big(\boldsymbol{\mu} - \boldsymbol{\mu}_*\big)(\boldsymbol{F}(\boldsymbol{\mu}) - \boldsymbol{\mu})^T \\
&= \big(\boldsymbol{\mu} - \boldsymbol{F}(\boldsymbol{\mu}) + \boldsymbol{F}(\boldsymbol{\mu}) - \boldsymbol{\mu}_*\big)(\boldsymbol{F}(\boldsymbol{\mu}) - \boldsymbol{\mu})^T \\
&= -(\boldsymbol{\mu} - \boldsymbol{F}(\boldsymbol{\mu}))(\boldsymbol{\mu} - \boldsymbol{F}(\boldsymbol{\mu}))^T + \big(\boldsymbol{F}(\boldsymbol{\mu}) - \boldsymbol{\mu}_*\big)(\boldsymbol{F}(\boldsymbol{\mu}) - \boldsymbol{\mu})^T < 0,
\end{aligned}$$

where the last inequality follows by the fact that $\boldsymbol{\mu} \geq \boldsymbol{\mu}_*$ implies $\boldsymbol{F}(\boldsymbol{\mu}) \geq \boldsymbol{F}(\boldsymbol{\mu}_*) = \boldsymbol{\mu}_*$. Also, note that $\dot{V}(\boldsymbol{\mu}_*) = V(\boldsymbol{\mu}_*) = 0$. Now, the proof of Theorem II.7 follows by the LaSalle Invariance Principle. Specifically, all trajectories with initial value in $\mathcal{N}$ converges to $\boldsymbol{\mu}_*$.

A.4.6 Proof of Theorem II.1

Throughout the proof, we use the subscript n to denote the case of finite n (total number of nodes), and we use no subscript to denote the behavior as $n \to \infty$, i.e., when we use the right-hand side of (2.34) as initial conditions rather than the actual initial conditions (A.5) given in Appendix A.3 (which is random). The proof is done via two steps: first, we analyze the case when $d_{\max} < \infty$; next, we relax this restriction for part 2 of Theorem II.1. Note that we use the function $\boldsymbol{F} = (F_{(1,1)}, F_{(1,2)}, F_{(2,1)}, F_{(2,2)})$ given by (2.26)-(2.27) which is the same as the function $\boldsymbol{F}$ given by the right-hand side of (2.6)-(2.7), if we use the right-hand side of (2.34) as the initial conditions.

Before proceeding with the discussion of these two steps, we present four preliminary lemmas that is used in the proof. The first lemma concerns with fixed points on the axis of origin. The other 3 lemmas relates the trajectory of $\boldsymbol{\mu}_n(t)$ to the trajectory of $\boldsymbol{\mu}(t)$, where $\boldsymbol{\mu}_n(t)$ and $\boldsymbol{\mu}(t)$ are the solutions of the ODE (2.30) with functions $\boldsymbol{F}_n$ and $\boldsymbol{F}$ respectively.

Lemma A.1. *Let $\boldsymbol{\mu}_* \coloneqq \lim_{s\to\infty} \boldsymbol{F}^s(\mathbf{1})$ to be the closest fixed point of $\boldsymbol{F}$ to $\mathbf{1}$ in sup norm. Assume at least one of the components of $\boldsymbol{\mu}_*$ is zero. Then, we have $\boldsymbol{\mu}_* = \mathbf{0}$.*

proof of Lemma. The proof is similar to the proof of Lemma II.5 part (iv). Without loos of generality, assume $\mu_*^{(1,1)} = 0$. Since $\boldsymbol{F}$ is a non-negative function, using Lemma II.4, we have $\mu_*^{(1,2)} = 0$ as $\mathbf{0} \leq \boldsymbol{\mu}_*$. Using the exact same argument, we have $\mu_*^{(2,2)} = \mu_*^{(2,1)} = 0$. □

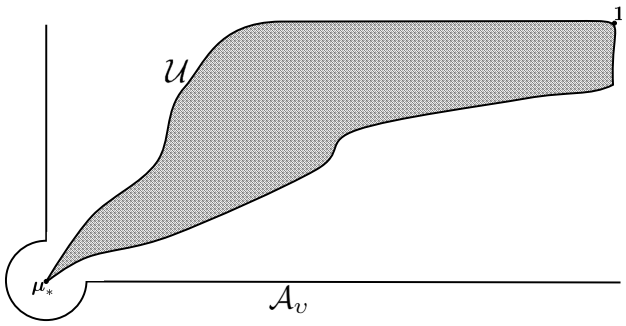


Figure A.1: Schematic of $\mathcal{A}_\upsilon$.

Lemma A.2. *Let $\boldsymbol{\mu}_* \coloneqq \lim_{s\to\infty} \boldsymbol{F}^s(\mathbf{1})$ to be the closest fixed point of $\boldsymbol{F}$ to $\mathbf{1}$ in sup norm. Assume $\boldsymbol{\mu}_*$ is an stable equilibrium of ODE (2.30), and assume $\boldsymbol{\mu}_* > \mathbf{0}$ component-wise. For any $\upsilon > 0$, define the set $\mathcal{A}_\upsilon$ as follows:*

$$\mathcal{A}_\upsilon \coloneqq (\mathcal{B}(\upsilon, \boldsymbol{\mu}_*) \cup \mathcal{A}) \setminus \mathit{int}(\mathcal{B}(\upsilon, \boldsymbol{\mu}_*))$$

where $\mathcal{A}$ is given by (2.31), $\mathcal{B}(\upsilon, \boldsymbol{x})$ is a ball of radius υ centered at $\boldsymbol{x}$, and $\mathit{int}(\mathcal{C})$ is the interior of the set $\mathcal{C}$. The 2-dimensional schematic of $\mathcal{A}_\upsilon$ is given in Figure A.1. Then, there is a small enough $\upsilon_0 > 0$ such that $\mathcal{A}_{\upsilon_0} \subset (0,1]^4$ and $\boldsymbol{F}(\boldsymbol{u}) \nleq \boldsymbol{u}$ component-wise, $\forall \boldsymbol{u} \in \mathcal{A}_{\upsilon_0}$. Moreover, we have $\boldsymbol{F}_n(\boldsymbol{u}) \nleq \boldsymbol{u}$ with high probability; that is to say, for any $p > 0$ there is a large enough $n_p > 0$ such that for any $n > n_p$, the event $\{\boldsymbol{F}_n(\boldsymbol{u}) \nleq \boldsymbol{u},\ \forall \boldsymbol{u} \in \mathcal{A}_{\upsilon_0}\}$ holds with probability greater than $1 - p$.

proof of Lemma. By part (iv) of Lemma II.5, for all $\boldsymbol{u} \in \mathcal{A}$ we have $\boldsymbol{F}(\boldsymbol{u}) \nleq \boldsymbol{u}$. Since $\boldsymbol{\mu}_*$ is an stable equilibrium point, $\exists \upsilon_1 > 0$ such that starting from any point $\boldsymbol{u} \in \mathcal{B}(\upsilon_1, \boldsymbol{\mu}_*) \subset (0,1]^4$, the ODE (2.30) converges to $\boldsymbol{\mu}_*$. Next, we show that if $\boldsymbol{F}(\boldsymbol{u}) \leq \boldsymbol{u}$ for some $\boldsymbol{u} \in \mathcal{B}(\upsilon_1, \boldsymbol{\mu}_*)$, then $\boldsymbol{u} \geq \boldsymbol{\mu}_*$ component-wise.

By the same argument as in the proof of Theorem II.7, if $\boldsymbol{F}(\boldsymbol{u}) \leq \boldsymbol{u}$ then the ODE (2.30) starting from $\boldsymbol{u}$ converges to the fixed point $\lim_{s\to\infty} \boldsymbol{F}^s(\boldsymbol{u})$ which is smaller than $\boldsymbol{u}$ component-wise. Hence, $\boldsymbol{u} \geq \boldsymbol{\mu}_*$ so that $\boldsymbol{\mu}_* = \lim_{s\to\infty} \boldsymbol{F}^s(\boldsymbol{u})$.

Let $\upsilon_0 = \upsilon_1/2$. By the above argument, we have $\boldsymbol{F}(\boldsymbol{u}) \nleq \boldsymbol{u}$ component-wise, $\forall \boldsymbol{u} \in \mathcal{A}_{\upsilon_0}$. Pick $\varepsilon_0 > 0$ small enough such that $\boldsymbol{F}(\boldsymbol{u}) - \varepsilon_0 \mathbf{1} \nleq \boldsymbol{u}$ for all $\boldsymbol{u} \in \mathcal{A}_{\upsilon_0}$. Now, using (2.34) and the fact that the dependency of $\boldsymbol{F}$ to the initial conditions is through coefficients, for any $p > 0$ we can pick $n_p > 0$ large enough such that for any fixed $n > n_p$ the event $\{\forall \boldsymbol{u} \in \mathcal{A}_{\upsilon_0} : \|\boldsymbol{F}(\boldsymbol{u}) - \boldsymbol{F}_n(\boldsymbol{u})\|_\infty < \varepsilon_0/2\}$ holds with probability greater than $1 - p$. □

Lemma A.3. *Let $\boldsymbol{\mu}_* \coloneqq \lim_{s\to\infty} \boldsymbol{F}^s(\mathbf{1})$ and $\boldsymbol{\mu}_{*,n} \coloneqq \lim_{s\to\infty} \boldsymbol{F}_n^s(\mathbf{1})$ to be the closest fixed point of $\boldsymbol{F}$ and $\boldsymbol{F}_n$ to $\mathbf{1}$ in sup norm, receptively. Note that $\boldsymbol{\mu}_{*,n}$ is a random variable as it*

depends on the initialization. Assume $\boldsymbol{\mu}_*$ is an stable equilibrium of ODE (2.30)*. Then, for any $\zeta > 0$, $\boldsymbol{\mu}_{*,n} \in \mathcal{B}(2\zeta, \boldsymbol{\mu}_*)$, with high probability.*

proof of Lemma. Pick k large enough such that $\boldsymbol{F}^k(\mathbf{1}) \in \mathcal{B}(\zeta, \boldsymbol{\mu}_*)$. Now, using (2.34), we have $\boldsymbol{F}_n^k(\mathbf{1}) \in \mathcal{B}(2\zeta, \boldsymbol{\mu}_*)$ with high probability. If $\boldsymbol{\mu}_* = \mathbf{0}$, the proof is complete. Otherwise, by Lemma A.1, we have $\boldsymbol{\mu}_* > 0$ component-wise. In the later case, using Lemma A.2, for small enough $\upsilon_0 < 2\zeta$, we have $\mathcal{A}_{\upsilon_0} \cap \mathcal{U}_n = \emptyset$ with high probability (Recall that $\mathcal{U}_n$ is the largest connected set containing $\mathbf{1} \coloneqq (1,1,1,1)$ such that $\forall \boldsymbol{u} \in \mathcal{U}$, $\boldsymbol{u} \geq \boldsymbol{F}_n(\boldsymbol{u})$.). Hence, $\lim_{s\to\infty} \boldsymbol{F}_n^s(\mathbf{1}) \in \mathcal{B}(2\zeta, \boldsymbol{\mu}_*)$ with high probability. □

Step 1: Assume there is a constant $\infty > d_{\max} > 0$ such that for all $d_j + d_{-j} > d_{\max}$, we have $K_j(d_j, d_{-j}) = d_j + d_{-j}$; that is to say, the inactive nodes with degree higher than $d_{\max}$ cannot be activated.

Now, the proof of part 1 of Theorem II.1 follows immediately from Theorem II.2, Lemma II.1, and the fact that the trajectory of $\boldsymbol{\mu}_n(x)$ and $\boldsymbol{\mu}(x)$ are uniformly close to each other with high probability. The last statement follows by the exact same argument as in the proof of Lemma A.2: for any $\varepsilon_0 > 0$, we can pick $n_p > 0$ large enough such that the probability of the event $\{\forall \boldsymbol{u} \geq \boldsymbol{\mu}_t : \|\boldsymbol{F}(\boldsymbol{u}) - \boldsymbol{F}_n(\boldsymbol{u})\|_\infty < \varepsilon_0\}$ is less than $1-p$ for all $n > n_p$.

For the proof of part 2 of Theorem II.1, we use the idea that we introduced in Section 2.9. We skip some minor details, as we have already used similar arguments in other parts of the chapter. If $\boldsymbol{\mu}_* = \mathbf{0}$ there is nothing to prove as all the half-edges has been used and the process has to stop. Note that by Lemma A.3, $\boldsymbol{\mu}_{*,n}$ is in any fixed ball around $\boldsymbol{\mu}_*$ with high probability.

Now, assume $\boldsymbol{\mu}_* \neq \mathbf{0}$. By Lemma A.1, we know $\boldsymbol{\mu}_* > \mathbf{0}$ component-wise. By Lemma A.2, we can pick $\upsilon_0 > 0$ small enough such that $\mathcal{A}_{\upsilon_0} \cap \mathcal{U} = \emptyset$ and for all $\boldsymbol{u} \in \mathcal{A}_{\upsilon_0}$ we have $\boldsymbol{u} > 0$ component-wise. Following the proof of Lemma A.2, let $\varepsilon_0 > 0$ to be small enough such that $\boldsymbol{F}(\boldsymbol{u}) - \varepsilon_0 \mathbf{1} \not\leq \boldsymbol{u}$ for all $\boldsymbol{u} \in \mathcal{A}_{\upsilon_0}$. Note that the event $\{\forall \boldsymbol{u} \in \mathcal{A}_{\upsilon_0} : \|\boldsymbol{F}(\boldsymbol{u}) - \boldsymbol{F}_n(\boldsymbol{u})\|_\infty < \varepsilon_0/2\}$ holds with high probability. Define $\boldsymbol{F}_{n,\kappa}(\boldsymbol{u})$ as in (2.26)-(2.27) by replacing the the terms λ_1, λ_2, and λ_m with $\lambda_1 + \kappa$, $\lambda_2 + \kappa$, and $\lambda_m + \kappa$. Let $\kappa_0 > 0$ to be small enough such that $\{\forall \boldsymbol{u} \in \mathcal{A}_{\upsilon_0} : \|\boldsymbol{F}_{n,\kappa_0}(\boldsymbol{u}) - \boldsymbol{F}_n(\boldsymbol{u})\|_\infty < \varepsilon_0/4\}$ holds with high probability. Also, note that by Lemma A.3, we have $\boldsymbol{\mu}_{*,n} \in \mathcal{B}(\upsilon_0/4, \boldsymbol{\mu}_*)$ with high probability.

To summarize the crucial results so far, we have $\{\forall \boldsymbol{u} \in \mathcal{A}_{\upsilon_0} : \|\boldsymbol{F}_{n,\kappa_0}(\boldsymbol{u}) - \boldsymbol{F}(\boldsymbol{u})\|_\infty < 3\varepsilon_0/4\}$ and $\boldsymbol{\mu}_{*,n} \in \mathcal{B}(\upsilon_0/4, \boldsymbol{\mu}_*)$ with high probability, and $\boldsymbol{F}(\boldsymbol{u}) - \varepsilon_0 \mathbf{1} \not\leq \boldsymbol{u}$ for all $\boldsymbol{u} \in \mathcal{A}_{\upsilon_0}$. Recall the choice of $\gamma_0 > 0$ as in (2.32) (see Figure 2.2). Consider a point $\boldsymbol{\mu}_{0,n} = \boldsymbol{\mu}_n(x_0) \in \mathcal{B}(\gamma_0, \boldsymbol{\mu}_*)$ on the trajectory of $\boldsymbol{\mu}_n$ such that $a_{1,n}(x_0) + a_{2,n}(x_0) + a_{m,n}^{(1)}(x_0) + a_{m,n}^{(2)}(x_0) = \kappa_0/2$. Note that such a point exists with high probability. At the corresponding time step of the process, the total number of active half-edges is of order of $\kappa_0 n$ with high probability, and

we expect the Markov process of adoption stop shortly thereafter.

At this point, we modify the process by adding one active node to each community, each of which has $\kappa_0 n$ active half-edges for in community connections and $\kappa_0 n$ half-edges for out community connections. In terms of one step drifts, we can recycle the one step drifts in Appendix A.2 by replacing the terms $m_1(n)$, $m_2(n)$, and $m_m(n)$ with $m_1(n)+\kappa_0 n$, $m_2(n)+\kappa_0 n$, and $m_m(n)+\kappa_0 n$. Similarly, this modification reflects in ODE by replacing $\boldsymbol{F}_n$ with $\boldsymbol{F}_{n,\kappa_0}$. Let $\boldsymbol{\mu}^{\kappa_0}_{*,n}$ denote the equilibrium point after modification. Trivially, $\boldsymbol{\mu}^{\kappa_0}_{*,n} \leq \boldsymbol{\mu}_{*,n}$ as $\mathcal{U}_n \subset \mathcal{U}^{\kappa_0}_n$. Moreover, $\mathcal{A}_{\upsilon_0} \cap \mathcal{U}^{\kappa_0}_n = \emptyset$ with high probability as $\{\forall \boldsymbol{u} \in \mathcal{A}_{\upsilon_0} : \|\boldsymbol{F}_{n,\kappa_0}(\boldsymbol{u}) - \boldsymbol{F}(\boldsymbol{u})\|_\infty < 3\varepsilon_0/4\}$ with high probability. Finally, since $\boldsymbol{\mu}_{*,n} \in \mathcal{B}(\upsilon_0/4, \boldsymbol{\mu}_*)$ with high probability, we have $\boldsymbol{\mu}^{\kappa_0}_{*,n} \in \text{int}(\mathcal{B}(\upsilon_0, \boldsymbol{\mu}_*))$ with high probability as well.

Using the exact same techniques as in Theorem II.2 and Lemma II.1, we can track the modified process using the modified ODE (with the new set of initial conditions given by the final values of the original ODE at the point of modification) up to any neighborhood of $\boldsymbol{\mu}^{\kappa_0}_{*,n}$. In particular, we can track the modified process upto when $a_{1,n}(x) + a_{2,n}(x) + a^{(1)}_{m,n}(x) + a^{(2)}_{m,n}(x) = \kappa_0/2$ for the modified ODE. At the corresponding time step of the modified process, we have already used all the original active half-edges from the original process, and most of the active half-edges from the modified one. Hence, if we ignore the modification, the original process runs out of active half-edges. Note that since $\boldsymbol{\mu}_* > \boldsymbol{0}$, we can pick υ_0 small enough such that most of the newly added active half-edges are used to be connected to inactive nodes, i.e., they are not wasted. Now, the result follows by adjusting the constant υ_0 to be arbitrary small.

Step 2: Consider the truncated processes $X^n_{L,\delta}$ and $X^n_{U,\delta}$ defined in Point 2 of Section 2.6. Let $\boldsymbol{F}_{L,\delta}$ and $\boldsymbol{F}_{U,\delta}$ denote the functions correspond to these processes, with the coefficients given by the right-hand side of (2.34). Let $\boldsymbol{\mu}_{*,L,\delta}$ and $\boldsymbol{\mu}_{*,U,\delta}$ denote the closets fixed points of $\boldsymbol{F}_{L,\delta}$ and $\boldsymbol{F}_{U,\delta}$ to $\mathbf{1}$ in sup norm respectively.

First off, note that for any $\delta > 0$, we have $\boldsymbol{F}_{U,\delta}(\boldsymbol{u}) < \boldsymbol{F}(\boldsymbol{u}) < \boldsymbol{F}_{L,\delta}(\boldsymbol{u})$ for all $\boldsymbol{u} \in [0,1]^4$. Moreover, $\boldsymbol{F}_{L,\delta}(\boldsymbol{u}) - \boldsymbol{F}_{U,\delta}(\boldsymbol{u}) < 2\delta$. Now, if we pick $\delta > 0$ small enough, the trajectory of the corresponding ODEs (given by the functions $\boldsymbol{F}_{L,\delta}$, $\boldsymbol{F}_{U,\delta}$, and $\boldsymbol{F}$) are uniformly close to each other. It is also easy to see that the corresponding equilibrium points are arbitrary close to each other by small choice of δ, whenever $\boldsymbol{\mu}_*$ is stable. Moreover, if $\boldsymbol{\mu}_*$ is stable, so are $\boldsymbol{\mu}_{*,L,\delta}$ and $\boldsymbol{\mu}_{*,U,\delta}$ for all small enough $\delta > 0$.

Now, the proof of part 2 of Theorem II.1 for the case $d_{\max} = \infty$ follows by the fact that for small $\delta > 0$, both truncated processes $X^n_{U,\delta}$ and $X^n_{L,\delta}$ stops at points arbitrary close to $\boldsymbol{\mu}_*$ (as $\boldsymbol{\mu}_{*,U,\delta}$ and $\boldsymbol{\mu}_{*,U,\delta}$ are arbitrary close to $\boldsymbol{\mu}_*$).

A.4.7 Proof of Theorem II.1

We begin with the simple observation that given $\|\boldsymbol{\alpha}_s\|_\infty \to 0$, we have $\boldsymbol{F}(\boldsymbol{\alpha}_s, \boldsymbol{u}) \to \boldsymbol{F}(\mathbf{0}, \boldsymbol{u})$ for all $\boldsymbol{u} \in [0,1]^4$.

Lemma A.4. *Assume $\{\boldsymbol{\alpha}_s\}_{s=1}^\infty$ converges to zero in sup norm. Then, we have*

$$\|\boldsymbol{F}(\boldsymbol{\alpha}_s, \boldsymbol{u}) - \boldsymbol{F}(\mathbf{0}, \boldsymbol{u})\|_\infty \to 0$$

uniformly over $\boldsymbol{u} \in [0,1]^4$.

Proof. Fix $\delta > 0$. Pick S_δ large enough such that $\|\boldsymbol{\alpha}_s\|_\infty < \delta$ for all $s > S_\delta$. It is easy to see that $\mathbf{F}(\mathbf{0}, \boldsymbol{u}) \geq \boldsymbol{F}(\boldsymbol{\alpha}_s, \boldsymbol{u}) > \boldsymbol{F}(\delta \mathbf{1}, \boldsymbol{u})$ component-wise for all $\boldsymbol{u} \in [0,1]^4$. Now, for any $\boldsymbol{u} \in [0,1]^4$ we have

$$\begin{aligned} \|\mathbf{F}(\mathbf{0}, \boldsymbol{u}) - \mathbf{F}(\boldsymbol{\alpha}_s, \boldsymbol{u})\|_\infty &\leq \|\mathbf{F}(\mathbf{0}, \boldsymbol{u}) - \mathbf{F}(\delta \mathbf{1}, \boldsymbol{u})\|_\infty \\ &= \delta \|\mathbf{F}(\mathbf{0}, \boldsymbol{u})\|_\infty \end{aligned} \tag{A.8}$$

The proof follows by small choice of δ. □

Define the set $\mathcal{U}(\boldsymbol{\alpha})$ to be largest connected set containing $\mathbf{1}$ such that for every $\boldsymbol{u} \in \mathcal{U}(\boldsymbol{\alpha})$, we have $\boldsymbol{u} \geq \boldsymbol{F}(\boldsymbol{\alpha}, \boldsymbol{u})$. It is easy to see that for any $\boldsymbol{\alpha} \neq \mathbf{0}$, we have $\boldsymbol{F}(\boldsymbol{\alpha}, \boldsymbol{u}) < \boldsymbol{F}(\mathbf{0}, \boldsymbol{u})$, and hence, $\mathcal{U}(\mathbf{0}) \subset \mathcal{U}(\boldsymbol{\alpha})$.

Using the same idea as in the proof of Lemma A.4, for any $\delta > 0$, we have $\boldsymbol{F}(\mathbf{0}, \boldsymbol{u}) > \boldsymbol{F}(\boldsymbol{\alpha}_s, \boldsymbol{u}) > \boldsymbol{F}(\delta \mathbf{1}, \boldsymbol{u})$ for all $s > S_\delta$. Hence, $\mathcal{U}(\mathbf{0}) \subset \mathcal{U}(\boldsymbol{\alpha}_s) \subset \mathcal{U}(\delta \mathbf{1})$ for all $s > S_\delta$. Now, using the fact that $\cap_{k=1}^\infty \mathcal{U}(1/k \times \mathbf{1}) = \mathcal{U}(\mathbf{0})$, we have $\lim_{s\to\infty} \mathcal{U}(\mathbf{0}) \cap \mathcal{U}(\boldsymbol{\alpha}_s) = \mathcal{U}(\mathbf{0})$.

By Corollary II.6, $\boldsymbol{\mu}_*(\boldsymbol{\alpha}_s) = \lim_{r\to\infty} F^r(\boldsymbol{\alpha}_s, \mathbf{1})$ is the closest fixed point of $F^r(\boldsymbol{\alpha}_s, \cdot)$ to $\mathbf{1}$ in sup norm. Define $\boldsymbol{\mu}_*(\mathbf{0}) \in \mathcal{U}(\mathbf{0})$ as follows: if $\mathcal{U}(\mathbf{0})$ is singleton, define $\boldsymbol{\mu}_*(\mathbf{0}) \coloneqq \{\mathbf{1}\}$, otherwise, define $\boldsymbol{\mu}_*(\mathbf{0})$ to be the closest fixed point of $\boldsymbol{F}(\mathbf{0}, \cdot)$ to $\mathbf{1}$ other than $\mathbf{1}$ itself. Note that for all $\boldsymbol{u} \in \mathcal{U}(\mathbf{0}) \cap \{\boldsymbol{x} : \boldsymbol{\mu}_*(\mathbf{0}) \leq \boldsymbol{x} \leq \mathbf{1}\} \setminus \{\mathbf{1}\}$, using the same argument as in the proof of Lemma II.4, we have $\lim_{r\to\infty} \boldsymbol{F}^r(\mathbf{0}, \boldsymbol{u}) = \boldsymbol{\mu}_*(\mathbf{0})$. Now, the sequence $\{\boldsymbol{\mu}_*(\boldsymbol{\alpha}_s)\}_{s=0}^\infty$ is sandwiched between the sequence $\{\boldsymbol{\mu}_*(1/k \times \mathbf{1})\}_{s=0}^\infty$ and $\boldsymbol{\mu}_*(\mathbf{0})$. Hence, if $\mathcal{U}(\mathbf{0})$ is singleton, then the final proportion of adopters converges to 0 as $\boldsymbol{\alpha_s} \to \mathbf{0}$. Otherwise, the final proportion of adopters is strictly positive, and we have

$$\lim_{s\to\infty} \boldsymbol{\mu}_*(\boldsymbol{\alpha}_s) = \lim_{r\to\infty} F^r(\mathbf{0}, \boldsymbol{u}) \qquad \forall \mu \in \mathcal{U}(\mathbf{0}) \cap \{\boldsymbol{x} : \boldsymbol{\mu}_*(\mathbf{0}) \leq \boldsymbol{x} \leq \mathbf{1}\} \setminus \{\mathbf{1}\}.$$

A.4.8 Proof of Theorem II.1

By definition of $\boldsymbol{F}$ given in (2.6)-(2.7), and the fact that $\mathbb{P}_{j*,m}(d_j, d_{-j}) = \mathbb{P}_{j,m}(d_j - 1, d_{-j})$ and $\mathbb{P}_{j,m*}(d_j, d_{-j}) = \mathbb{P}_{j,m}(d_j, d_{-j} - 1)$, we have

$$\begin{aligned}
F_{(j,j)}(\mu^{(j,j)}, \mu^{(j,-j)}) &= \sum_{u_j + u_{-j} \le K_j(d_j + d_{-j})} \mathbb{P}_{j,m}(d_j - 1, d_{-j})(1 - \alpha_j(d_j + d_{-j}))\times \\
&\qquad Bi(u_j; d_j - 1, 1 - \mu^{(j,j)}) Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}) \\
&= \sum_{u_j + u_{-j} \le K_j(d_j + d_{-j} + 1)} \mathbb{P}_{j,m}(d_j, d_{-j})(1 - \alpha_j(d_j + d_{-j} + 1))\times \\
&\qquad Bi(u_j; d_j, 1 - \mu^{(j,j)}) Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}) \\
F_{(-j,j)}(\mu^{(j,j)}, \mu^{(j,-j)}) &= \sum_{u_j + u_{-j} \le K_j(d_j + d_{-j})} \mathbb{P}_{j,m}(d_j, d_{-j} - 1)(1 - \alpha_j(d_j + d_{-j}))\times \\
&\qquad Bi(u_{-j}; d_{-j} - 1, 1 - \mu^{(j,-j)}) Bi(u_j; d_j, 1 - \mu^{(j,j)}) \\
&= \sum_{u_j + u_{-j} \le K_j(d_j + d_{-j} + 1)} \mathbb{P}_{j,m}(d_j, d_{-j})(1 - \alpha_j(d_j + d_{-j} + 1))\times \\
&\qquad Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}) Bi(u_j; d_j, 1 - \mu^{(j,j)})
\end{aligned}$$

Hence, $F_{(1,1)}(\mu^{(1,1)}, \mu^{(1,2)}) = F_{(2,1)}(\mu^{(1,1)}, \mu^{(1,2)})$ and $F_{(2,2)}(\mu^{(2,2)}, \mu^{(2,1)}) = F_{(1,2)}(\mu^{(2,2)}, \mu^{(2,1)})$. Now, if we also have $\mu^{(1,1)} = \mu^{(2,1)}$ and $\mu^{(2,2)} = \mu^{(1,2)}$, then we have

$$\begin{aligned}
F_{(1,1)}(\mu^{(1,1)}, \mu^{(1,2)}) - \mu^{(1,1)} &= F_{(2,1)}(\mu^{(1,1)}, \mu^{(1,2)}) - \mu^{(1,1)} \\
F_{(2,2)}(\mu^{(2,2)}, \mu^{(2,1)}) - \mu^{(2,2)} &= F_{(1,2)}(\mu^{(2,2)}, \mu^{(2,1)}) - \mu^{(1,2)}
\end{aligned}$$

Since these equalities hold at the initial stage of ODE (2.30) with the function $\boldsymbol{F}$ given in (2.6)-(2.7), it holds on the whole trajectory; that is to say, $\mu^{(1,1)}(x) = \mu^{(2,1)}(x)$ and $\mu^{(2,2)}(x) = \mu^{(1,2)}(x)$ for all $x \ge 0$ where $\boldsymbol{\mu}(x)$ is the solution of the ODE.

A.4.9 Proof of Theorem II.2

If $\mu^{(1,1)} = \mu^{(2,2)}$ and $\mu^{(2,1)} = \mu^{(1,2)}$ then we have $F_{(1,1)}(\mu^{(1,1)}, \mu^{(1,2)}) = F_{(2,2)}(\mu^{(2,2)}, \mu^{(2,1)})$ and $F_{(2,1)}(\mu^{(1,1)}, \mu^{(1,2)}) = F_{(1,2)}(\mu^{(2,2)}, \mu^{(2,1)})$. Using the same argument as in the proof of Theorem II.1, we have $\mu^{(1,1)}(x) = \mu^{(2,2)}(x)$ and $\mu^{(2,1)}(x) = \mu^{(1,2)}(x)$ for all $x \ge 0$.

APPENDIX B

Appendix of Chapter III

B.1 Proof of Theorem III.2

Before presenting the proof, we revisit some basic properties of the order statistics of n independent and identically distributed random variables.

Lemma B.1. *Let $\{X_i\}_{i=1}^m$ denote a set of* i.i.d. *random variables. Let $F(\cdot)$ and $f(\cdot)$ represent the cumulative distribution function and probability density function of X_1 respectively. Consider the order statistics of $\{X_i\}_{i=1}^m$ and denote it by $\{X^{(i)}\}_{i=1}^m$. For every $x_1 \leq x_2 \leq \cdots \leq x_n$ and $l \leq m$, we have*

$$f_{X^{(1)},X^{(2)},\ldots,X^{(l)}}(x_1, x_2, \ldots, x_l) = l!\binom{m}{l} \times \prod_{i=1}^{l} f(x_i) \times (1 - F(x_l))^{m-l},$$

$$f_{X^{(l)},X^{(l+1)},\ldots,X^{(m)}}(x_l, x_{l+1}, \ldots, x_m) = (m-l+1)!\binom{m}{m-l+1} \times \prod_{i=l}^{m} f(x_i) \times F(x_l)^{l-1},$$

$$f_{X^{(l)}}(x_l) = l\binom{m}{l} \times f(x_l) \times F(x_l)^{l-1} \times (1 - F(x_l))^{m-l},$$

$$f_{X^{(1)},X^{(2)},\ldots,X^{(l-1)}|X^{(l)}}(x_1, x_2, \ldots, x_{l-1}|x_l) = (l-1)!\frac{\prod_{i=1}^{l-1} f(x_i)}{F(x_l)^{l-1}},$$

$$f_{X^{(l+1)},X^{(l+2)},\ldots,X^{(m)}|X^{(l)}}(x_{l+1}, x_{l+2}, \ldots, x_m|x_l) = (m-l)!\frac{\prod_{i=l+1}^{m} f(x_i)}{(1 - F(x_l))^{m-l}}.$$

Corollary B.2. *Let $\{Y_i\}_{i=1}^{l-1}$ denote a random permutation of $\{X^{(i)}\}_{i=1}^{l-1}$, i.e., pick a permutation $\sigma \in S_{l-1}$ uniformly at random and set $Y_i = X^{(\sigma(i))}$ for all $1 \leq i \leq l-1$. Similarly, let*

$\{Z_i\}_{i=l+1}^m$ denote a random permutation of $\{X^{(i)}\}_{i=l+1}^m$. Then we have

$$f_{Y_1,Y_2,\ldots,Y_{l-1}|X^{(l)}}(y_1,y_2,\ldots,y_{l-1}|x_l) = \frac{\prod_{i=1}^{l-1} f(y_i)}{F(x_l)^{l-1}},$$
$$f_{Z_{l+1},Z_{l+2},\ldots,Z_m|X^{(l)}}(z_{l+1},z_{l+2},\ldots,z_m|x_l) = \frac{\prod_{i=l+1}^{m} f(z_i)}{(1-F(x_l))^{m-l}}.$$

Moreover, $\{Y_i\}_{i=1}^{l-1}$ are identically distributed and conditioned on $X^{(l)}$, they are independent. Same holds for $\{Z_i\}_{i=l+1}^m$:

$$f_{Y_i|X^{(l)}}(y_i|x_l) = \frac{f(y_i)}{F(x_l)} \qquad \forall i \le l-1$$
$$f_{Z_i|X^{(l)}}(y_i|x_l) = \frac{f(z_i)}{1-F(x_l)} \qquad \forall i \ge l+1$$

Corollary B.3. *Let $\{X_i\}_{i=1}^m$ to be independent exponentially distributed random variables with mean n. Consider the random variables $\{Y_i\}_{i<l}$ and $\{Z_i\}_{i>l}$ as are defined in Corollary B.2. Then, the conditional distribution of these random variables are given as follows:*

$$f_{X^{(i)}}(x_i) = i\binom{n}{i} \times (1-\mathrm{e}^{-\frac{x_i}{n}})^{i-1} \times \frac{1}{n}\mathrm{e}^{-(n-i+1)\frac{x_i}{n}} \xrightarrow{n\to\infty} \frac{\mathrm{e}^{-x_i}{x_i}^{i-1}}{(i-1)!} \qquad \forall i \in [n]$$
$$f_{Y_i|X^{(l)}}(y_i|x_l) = \frac{\frac{1}{n}\mathrm{e}^{-y_i/n}}{1-\mathrm{e}^{-x_l/n}} \xrightarrow{n\to\infty} \frac{1}{x_l} \qquad \forall i \le l-1$$
$$f_{Z_i|X^{(l)}}(z_i|x_l) = \frac{\frac{1}{n}\mathrm{e}^{-z_i/n}}{\mathrm{e}^{-x_l/n}} = \frac{1}{n}\mathrm{e}^{-(z_i-x_l)/n} \qquad \forall i \ge l+1$$

Most notably, the conditional distribution of Y_i for $i < l$ conditioned on $X^{(l)} = x_l$ converges to the uniform distribution over $[0, x_l]$, as n goes to infinity. Moreover, the distribution of $X^{(i)}$ converges to Erlang(i).

As we mentioned, $\mathbb{E}U(N_n)$ is the law of $[N_{n,\circ}(r)]$ for a uniformly chosen $r \in [n]$. The idea is to first define an exploration process over K_n that realizes the connected component of the vertex r in N_n. Then, we show that the connected component is locally tree-like and the distribution of the connected component up to any finite time step of the exploration process converges to $Er(P)$. Finally, using the Portmanteau Theorem, we prove $\mathbb{E}U(N_n) \xrightarrow{w} Er(P)$.

Step 1: Exploration Process

The first step is to define a process that explores K_n and realizes the connected component of a randomly selected vertex $r \in [n]$ in N_n. Let $E_n = \{\{i,j\} : i \neq j \in [n]\}$ denote the set of all edges in K_n. In order to track the process, we also construct a map ϕ from $\mathbb{N}^f$ to the connected component of r and a fictitious vertex. In particular, ϕ maps $\mathbb{N}^f \setminus \{\phi^{-1}(v) : v \text{ in connected component of } r\}$ to the fictitious vertex. The exploration is on E_n and the cost of edges in E_n; at each step of the exploration process, E_n is partitioned into five sets, defined as follows:

$\mathcal{A}_t = \{(\{i,j\}, C_n(\{i,j\})) : \{i,j\} \text{ is active at time } t\}$
$\mathcal{C}_t = \{(\{i,j\}, C_n(\{i,j\})) : \{i,j\} \text{ belongs to the connected component at time } t\}$
$\mathcal{D}_t = \{(\{i,j\}, C_n(\{i,j\})) : \{i,j\} \text{ does not belong to the connected component at time } t\}$
$\mathcal{R}_t = \{(\{i,j\}, C_n(\{i,j\})) : \text{the cost of the non-active edge } \{i,j\} \text{ has been realized by time } t\}$
$\mathcal{U}_t = \{\{i,j\} : \text{the cost of the edge } \{i,j\} \text{ has not been realized by time } t\}$

Remark B.4. During the proof, we may abuse the notation by saying $\{i,j\} \in \mathcal{A}_t$ without including the cost. Even though $\mathcal{A}_t$ is a set of edges and their costs, we say $i \in \mathcal{A}_t$ (notationally), if there is a vertex $j \in [n]$ such that $\{i,j\} \in \mathcal{A}_t$. Finally, we say a vertex $i \in [n]$ has been explored by time step t, if both the threshold of i, $\mathcal{T}_i$ given by (3.3), and the set of potential neighbors of i, $\mathcal{P}_i$ given by (3.4), have been realized.

Remark B.5. The partition of E_n at time t satisfies the following properties:

1. $\mathcal{A}_t$: The set of active edges, $\mathcal{A}_t$, consists of all the edges $\{v,z\}$ such that: *i*) The cost of $\{v,z\}$ has been realized; *ii*) Exactly one of v or z (but not both) belongs to the connected component at time t; and *iii*) If $\mathcal{P}_v$ has been realized, then $z \in \mathcal{P}_v$. If $\mathcal{P}_z$ has been realized, then $v \in \mathcal{P}_z$.

2. $\mathcal{C}_t$: The set of voted-in edges, $\mathcal{C}_t$, consists of all the edges $\{v,z\}$ such that: *i*) The cost of $\{v,z\}$ has been realized; *ii*) The vertices v and z belong to the connected component at time t; and *iii*) Each vertex is a potential neighbor of the other, i.e., $z \in \mathcal{P}_v$ and $v \in \mathcal{P}_z$.

3. $\mathcal{D}_t$: The set of erased edges, $\mathcal{D}_t$, consists of all the edges $\{v,z\}$ such that: *i*) The cost of $\{v,z\}$ has been realized; and *ii*) If only $\mathcal{P}_v$ ($\mathcal{P}_z$) has been realized, then $z \notin \mathcal{P}_v$ ($v \notin \mathcal{P}_z$); if $\mathcal{P}_v$ and $\mathcal{P}_z$ have been realized, then either $z \notin \mathcal{P}_v$ or $v \notin \mathcal{P}_z$ (or both).

4. $\mathcal{R}_t$: The set of realized edges, $\mathcal{R}_t$, consists of all the edges $\{v,z\}$ such that: *i*) The cost of $\{v,z\}$ has been realized; *ii*) Neither v nor z belongs to the connected component

at time t; and *iii*) If $\mathcal{P}_v$ has been realized, then $z \in \mathcal{P}_v$; if $\mathcal{P}_z$ has been realized, then $v \in \mathcal{P}_z$.

5. $\mathcal{U}_t$: The set of unrealized edges, $\mathcal{U}_t$, consists of all the edges $\{v, z\}$ such that the cost of $\{v, z\}$ has not been realized.

Remark B.6. At each step of the exploration process, we may add at most one vertex to the connected component of r. Moreover, if the vertex v is added to the connected component at time $t+1$, i.e., $v \in \mathcal{C}_{t+1}$, then v is active at time t, i.e., $v \in \mathcal{A}_t$ and the exploration process explores an edge $\{j, v\}$ such that $j \in \mathcal{C}_t$.

Exploration process details and an alternative viewpoint: The exploration process starts by realizing the sets for $t = 0$. Set $\phi(r) = \emptyset$ and define $v_0 \coloneqq r$ and $k \coloneqq d_r(n)$. Let $\mathcal{T}_0$ and $\mathcal{P}_0$ denote the threshold and the set of potential neighbors of v_0, respectively. By definition, $\mathcal{T}_0$ and $\mathcal{P}_0$ are given by

$$\begin{aligned}\mathcal{T}_0 &= k+1^{th} \text{ smallest value in } \{C_n(\{v_0, j\}) : j \in [n] \setminus \{v_0\}\},\\ \mathcal{P}_0 &= \{j \in [n] \setminus \{v_0\} : C_n(\{v_0, j\}) < \mathcal{T}_0\}.\end{aligned}$$

Next, we present an alternative way to realize $\mathcal{T}_0$ and $\mathcal{P}_0$ without realizing the cost of $\{v_0, j\}$ for all $j \in [n] \setminus \{v_0\}$. This alternative construction of the finite graph is an essential part of the proof of the weak convergence result, and is used at all time steps $t \geq 0$ as well.
Pick a vertex $z_0 \in [n] \setminus \{v_0\}$ uniformly at random and assume the threshold of the vertex v_0 is equal to the cost of the edge $\{v_0, z_0\}$, i.e., $\mathcal{T}_0 = C_n(\{v_0, z_0\})$. Realize the value of $C_n(\{v_0, z_0\})$; according to Lemma B.1, the density function of $C_n(\{v_0, z_0\})$ is given by

$$f_{C_n(\{v_0,z_0\})}(w) = (k+1)\binom{n-1}{k+1} \times \frac{1}{n}\mathrm{e}^{-w(n-k-1)/n} \times (1-\mathrm{e}^{-w/n})^k.$$

Next, pick $\mathcal{I}_0 = \{z_1, z_2, \ldots, z_k\}$, a subset of size k, from $[n]\setminus(\{z_0\}\cup\{v_0\})$ uniformly at random and assume $\mathcal{I}_0$ is the set of potential neighbors of v_0, i.e., $\mathcal{P}_0 = \mathcal{I}_0$. Pick a permutation $\mathfrak{m}_0$ over $[|\mathcal{I}_0|]$ uniformly at random and define $\phi(z_i) = \mathfrak{m}_0(i)$ for all $i \in [k]$. Realize the values of $\{C_n(\{v_0, z_i\})\}_{i=1}^k$; by Corollary B.2, the conditional joint density function of these random variables is given by

$$f_{C_n(\{v_0,z_1\}),C_n(\{v_0,z_2\}),\ldots,C_n(\{v_0,z_k\})|\mathcal{T}_0}(w_1, w_2, \ldots, w_k|w_0) = \frac{\prod_{i=1}^k \frac{1}{n}\mathrm{e}^{-w_i/n}}{(1-\mathrm{e}^{-w_0/n})^k}.$$

Start the exploration process with

$$\mathcal{A}_0 = \{(\{v_0, j\}, c_n(\{v_0, j\})) : j \in \mathcal{P}_0\}, \tag{B.1a}$$

$$\mathcal{C}_0 = \{\}, \tag{B.1b}$$

$$\mathcal{D}_0 = \{(\{v_0, z_0\}, c_n(\{v_0, z_0\}))\}, \tag{B.1c}$$

$$\mathcal{R}_0 = \{\}, \tag{B.1d}$$

$$\mathcal{U}_0 = E_n \setminus \{\{v_0, j\} : j \in I_0 \cup \{z_0\}\}. \tag{B.1e}$$

The description of the above equations is as follows:

1. Equation (B.1a): The vertex v_0 is the root of the connected component. All the potential neighbors of v_0 are included in $\mathcal{A}_0$.

2. Equation (B.1b): Although the vertex v_0 is the root of the connected component, there is no edge in the connected component yet; hence, the set $\mathcal{C}_0$ is set to be empty at the initial stage.

3. Equation (B.1c): The connection $\{v_0, z_0\}$ determines the threshold of the vertex v_0; hence, the vertex $z_0 \notin \mathcal{P}_0$ and the edge $\{v_0, z_0\}$ does not survive.

4. Equation (B.1d): The vertex v_0 is the root of the connected component; hence, none of the edges of form $\{v_0, z\}$ belongs to $\mathcal{R}_0$. The set $\mathcal{R}_0$ is set to be empty at the initial stage.

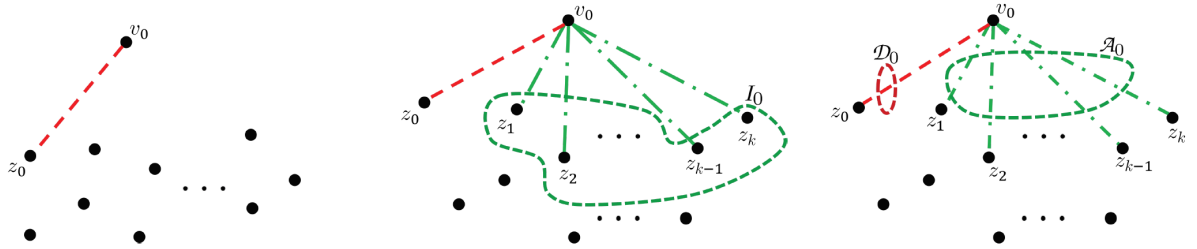
5. Equation (B.1e): All the edges $\{v_0, j\}$ such that $C_n\{v_0, j\}$ has been realized are removed from E_n to construct $\mathcal{U}_0$.

Figure B.1 depicts the preparation step for the exploration process. Define $\widehat{\mathcal{T}}_0$ to be equal to $\mathcal{T}_0$. These two values might be different for $t > 0$. The definition and the role of $\widehat{\mathcal{T}}_t$ will become clear later on.

Before proceeding with the exploration process, we need to define an order on $\mathbb{N}^f$: for two sequence $\boldsymbol{i} = (i_1, i_2, \ldots, i_l)$ and $\boldsymbol{j} = (j_1, j_2, \ldots, j_{l'})$, we say $\boldsymbol{i} \prec \boldsymbol{j}$ if $l < l'$ or $l = l'$ and there exist some $g \in \mathbb{Z}_+$ such that $(i_1, i_2, \ldots, i_{g-1}) = (j_1, j_2, \ldots, j_{g-1})$ and $i_g < j_g$.

Remark B.7. For the sake of notational simplicity, we denote the set of potential neighbors and the threshold of the vertex v_t by $\mathcal{P}_t$ and $\mathcal{T}_t$ instead of $\mathcal{P}_{v_t}$ and $\mathcal{T}_{v_t}$. We may also use $\mathcal{P}_j$ as the set of potential neighbors of the vertex j. The distinction is clear from the context.

The exploration process for $t \geq 0$ is as follows; let $e_{t+1} = \{\phi^{-1}(\boldsymbol{i}), \phi^{-1}(\boldsymbol{j})\} \in \mathcal{A}_t$ such that $\boldsymbol{i}$ is minimal among $\{\phi(v) : v \in \mathcal{A}_t\}$ and $\boldsymbol{j}$ is minimal among $\{\phi(z) : \{\phi^{-1}(\boldsymbol{i}), z\} \in \mathcal{A}_t\}$. The



(a) Set $\phi(v_0) = \emptyset$, pick z_0 uniformly at random, realize cost of $\{v_0, z_0\}$ such that $\mathcal{T}_0 = C_n(\{v_0, z_0\})$

(b) Pick $I_0 = \{z_1, z_2, \ldots, z_k\}$ uniformly at random. Set $\phi(z_i) = m_0(i)$ for all $i \in [k]$ and realize the cost of $\{v_0, z_i\}$ such that $\mathcal{P}_0 = I_0$

(c) Initialize the sets $\mathcal{A}_0$ and $\mathcal{D}_0$. Set $\mathcal{U}_0 = E_n \setminus (\mathcal{A}_0 \cup \mathcal{D}_0)$ and define $\mathcal{C}_0 = \mathcal{R}_0 = \{\}$

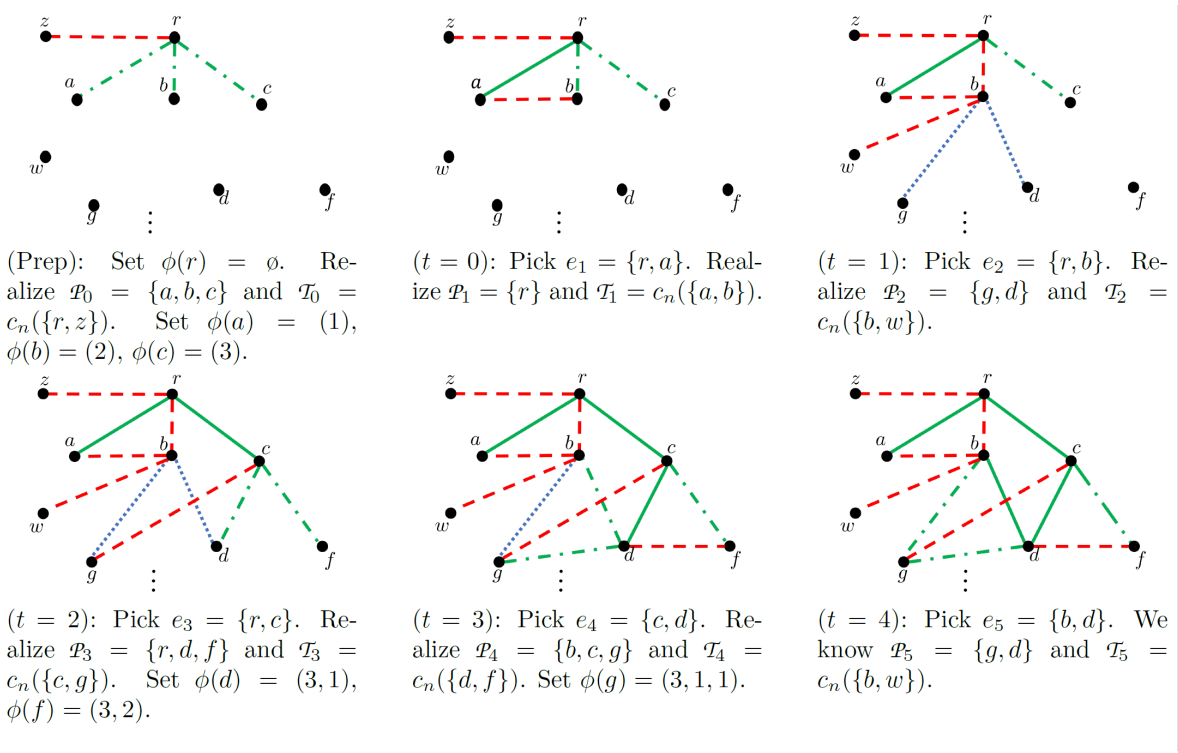
Figure B.1: Preparation step for the exploration process

choice of e_{t+1} corresponds to the breadth-first search algorithm. As an example for $t = 0$, the set $\{\phi(v) : v \in \mathcal{A}_0\}$ equals to $\{\emptyset, 1, 2, \ldots, k\}$; hence $\boldsymbol{i} = \emptyset$ and $\phi^{-1}(\emptyset) = r$. Moreover, the set $\{\phi(z) : \{\phi^{-1}(\emptyset), z\} \in \mathcal{A}_t\}$ equals $\{1, 2, \ldots, k\}$; hence $\boldsymbol{j} = 1$ and $\phi^{-1}(1) = z_{m_0^{-1}(1)}$. Hence, $e_1 = \{\phi^{-1}(\emptyset), \phi^{-1}(1)\} = \{r, z_{m_0^{-1}(1)}\}$.

Remark B.8. Let $\phi(v) = (i_1, i_2, \ldots, i_g)$ and define $par(v) \coloneqq \phi^{-1}(i_1, i_2, \ldots, i_{g-1})$. The exploration process ensures that $par(v)$ belongs to the connected component of r; moreover, $par(v)$ is the first vertex in the connected component such that v belongs to the set of the potential neighbors of $par(v)$, i.e., for every z in the connected component if $v \in \mathcal{P}_z$ then $par(v)$ is attached to the connected component before z. However, it is possible to have $\{par(v), v\} \in \mathcal{D}_t$ for some $t > 0$, which is the case if $par(v) \notin \mathcal{P}_v$ and the vertex v has been explored by time step t. Still, v may connect to the connected component through some other vertex v', i.e., $\{v', v\} \in \mathcal{C}_{t'}$ for some $t' > t$. Figure B.2 illustrates such a situation, where $par(b) = r$, $\{par(b), b\} \in \mathcal{D}_2$ and $\{d, b\} \in \mathcal{C}_5$. Note that the labeling is based on being a "potential neighbor" rather than being an actual neighbor.

Remark B.9. A vertex $v \neq r$ belongs to the connected component of r by time step t if and only if $v \in \mathcal{C}_t$. A vertex $v \in [n] \setminus \{r\}$ has been explored by time step t if and only if v belongs to the connected component, or there is a vertex $v' \in \mathcal{C}_t$ such that $\{v', v\} \in \mathcal{D}_t \cup \mathcal{C}_t$ and v belongs to the set of potential neighbors of v', i.e., $v \in \mathcal{P}_{v'}$. Note that in the later case, the vertex v' may not be the vertex $par(v)$; As an example, in Figure B.2 the vertex g is explored at time step $t = 5$ (since $e_6 = \{b, g\}$), but $\{par(g), g\} = \{d, g\} \in \mathcal{A}_6$.

Remark B.10. An important observation is that for every $\{v, z\} \in \mathcal{A}_t$ exactly one of v or z (but not both) belongs to the connected component of the vertex r at time t. Moreover, at least one of the vertices v or z has been explored; hence, at each time step we may explore at most one vertex.



(Prep): Set $\phi(r) = \emptyset$. Realize $\mathcal{P}_0 = \{a, b, c\}$ and $\mathcal{T}_0 = c_n(\{r, z\})$. Set $\phi(a) = (1)$, $\phi(b) = (2)$, $\phi(c) = (3)$.

($t = 0$): Pick $e_1 = \{r, a\}$. Realize $\mathcal{P}_1 = \{r\}$ and $\mathcal{T}_1 = c_n(\{a, b\})$.

($t = 1$): Pick $e_2 = \{r, b\}$. Realize $\mathcal{P}_2 = \{g, d\}$ and $\mathcal{T}_2 = c_n(\{b, w\})$.

($t = 2$): Pick $e_3 = \{r, c\}$. Realize $\mathcal{P}_3 = \{r, d, f\}$ and $\mathcal{T}_3 = c_n(\{c, g\})$. Set $\phi(d) = (3, 1)$, $\phi(f) = (3, 2)$.

($t = 3$): Pick $e_4 = \{c, d\}$. Realize $\mathcal{P}_4 = \{b, c, g\}$ and $\mathcal{T}_4 = c_n(\{d, f\})$. Set $\phi(g) = (3, 1, 1)$.

($t = 4$): Pick $e_5 = \{b, d\}$. We know $\mathcal{P}_5 = \{g, d\}$ and $\mathcal{T}_5 = c_n(\{b, w\})$.

Figure B.2: A realization of the exploration process up to $t = 5$. Cost of the edges, threshold of the vertices, potential degree of the vertices, the permutation π_t, and the sets $\mathcal{A}_t$, $\mathcal{U}_t$, $\mathcal{C}_t$, and $\mathcal{D}_t$ are not mentioned. Solid green edges belong to $\mathcal{C}_t$, dashed red edges belong to $\mathcal{D}_t$, dashed dotted green edges belong to $\mathcal{A}_t$, and dotted blue edges belong to $\mathcal{R}_t$. Note that $par(b)$ is defined to be r although b is connected to the root via d at time $t = 5$. Moreover, $par(g)$ is d since the vertex d is the first vertex in the connected component such that $g \in \mathcal{P}_d$; although $g \in \mathcal{P}_b$, the vertex b is connected to the connected component after the vertex d. Based on the exploration process, $\{b, g\} \in \mathcal{A}_5$ and $e_6 = \{b, g\}$.

Based on the exploration strategy the vertex $\phi^{-1}(\boldsymbol{i})$ has been explored, but it may not belong to the connected component. More explicitly, $par(\phi^{-1}(\boldsymbol{i}))$ belongs to the connected component (Remark B.8), and $\phi(par(\phi^{-1}(\boldsymbol{i}))) \prec \boldsymbol{j}$; hence, the edge $\{par(\phi^{-1}(\boldsymbol{i})), \phi^{-1}(\boldsymbol{i})\} \in \mathcal{C}_t \cup \mathcal{D}_t$ or equivalently, $\phi^{-1}(\boldsymbol{i})$ has been explored by time t (Remark B.9). However, the vertex $\phi^{-1}(\boldsymbol{j})$ has two different possibilities,

- ***Subcase I***, where $\phi^{-1}(\boldsymbol{j})$ has not been explored: in this case, the vertex $\phi^{-1}(\boldsymbol{i})$ belongs to the connected component. Let $v_{t+1} = \phi^{-1}(\boldsymbol{j})$. Let $m \leq t + 1$ denote the number of explored vertices by time step t. Note that at time $t = 0$, the root vertex has already been explored and for each $t > 0$, we may explore at most one vertex at each time step (Remark B.10). Define $k' := \min(n - m - 2, d_{v_{t+1}}(n))$. If $n - m - 2 < 0$, which may happen if the graph is fully connected and the process is reaching to its end,

then let $k' = 0$. In order to explore v_{t+1}, the first step is to choose $\mathcal{B}_{t+1} = \{z_1, z_2, \ldots, z_{k'}\}$, a subset of size k', uniformly at random from the set of unexplored vertices (there are $n - m - 1$ unexplored vertices other than v_{t+1}). Next, pick a vertex z_0 out of remaining unexplored vertices uniformly at random (there are $n - m - 1 - k'$ option for z_0). Assume that the cost of the edges $\{v_{t+1}, z_i\}_{i=1}^{k'}$ are the least k' values in $\{C_n(\{v_{t+1}, z\}) : z \text{ is not explored}\}$ and the cost of $\{v_{t+1}, z_0\}$ is exactly the $k' + 1^{th}$ smallest one. As in $t = 0$, we do not realize the cost of $\{v_{t+1}, z\}$ for all unexplored vertices $z \in [n]$. Using Lemma B.1 and Corollary B.3, the joint density function of $\{C_n(\{v_{t+1}, z\})\}_{i=0}^{k'}$ is given by,

$$f_{C_n(\{v_{t+1},z_0\}),C_n(\{v_{t+1},z_1\}),\ldots,C_n(\{v_{t+1},z_{k'}\})}(w_0, w_1, \ldots, w_{k'}) = (k'+1)\binom{n-m-1}{k'+1} \times \prod_{i=0}^{k'} \frac{1}{n} \mathrm{e}^{-w_i/n} \times \mathrm{e}^{-w_0(n-m-1-(k'+1))/n}$$

where $w_i \leq w_0$ for all $i \in [k']$. Notice that for every vertex $v \notin \mathcal{B}_{t+1} \cup \{z_0\}$ such that v has not been explored and the cost of $\{v_{t+1}, v\}$ has not been realized, the value of $C_n(\{v_{t+1}, v\})$ is greater than $c_n(\{v_{t+1}, z_0\})$. Define $\widehat{\mathcal{T}}_{t+1}$ to be $c_n(\{v_{t+1}, z_0\})$,

$$\widehat{\mathcal{T}}_{t+1} \coloneqq c_n(\{v_{t+1}, z_0\})$$

Remark B.11. If, after realizing $\mathcal{B}_{t+1} \cup \{z_0\}$, the set of unexplored vertices v such that $\{v_{t+1}, v\}$ has not been realized is non-empty, then $d_{v_{t+1}}(n) < n-m-2$ and $\mathcal{T}_{t+1} \leq \widehat{\mathcal{T}}_{t+1}$.

The second step to explore v_{t+1} is to realize the cost of all the edges between v_{t+1} and the explored vertices; by Corollary B.3, for every explored vertex v such that $\{v_{t+1}, v\} \in \mathcal{U}_t$, the density of $C_n(\{v_{t+1}, v\})$ conditioned on $\widehat{\mathcal{T}}_v = w_v$ is given by

$$f_{C_n(\{v_{t+1},v\})|\widehat{\mathcal{T}}_v}(w|w_v) = \frac{1}{n}\mathrm{e}^{-(w-w_v)/n}$$

Remark B.12. Assume the vertex v has been explored but the value of $C_n(\{v_{t+1}, v\})$ has not been realized. Since v has been explored, we already know that $v_{t+1} \notin \mathcal{P}_v$ and $C_n(\{v_{t+1}, v\}) > \mathcal{T}_v$. However, by the first step of the exploration process for the vertex v we have $C_n(\{v_{t+1}, v\}) > \widehat{\mathcal{T}}_v$. Moreover, Remark B.11 suggests $\widehat{\mathcal{T}}_v \geq \mathcal{T}_v$ since $\{v_{t+1}, v\} \in \mathcal{U}_t$.

Note that the potential neighbors of v_{t+1} are either explored or belongs to $\mathcal{B}_{t+1} \cup \{z_0\}$.

Define $k \coloneqq d_{v_{t+1}}(n)$ and set the threshold and the set of potential neighbors of v_{t+1},

$\mathcal{T}_{t+1} = k + 1^{th}$ smallest value in $\{c_n(\{v_{t+1}, j\}) : j \in [n]$ is explored or $j \in \mathcal{B}_{t+1} \cup \{z_0\}\}$
$\mathcal{P}_{t+1} = \{j \in [n] : c_n(\{v_{t+1}, j\}) < \mathcal{T}_{t+1}$ and $j \in [n]$ is explored or $j \in \mathcal{B}_{t+1} \cup \{z_0\}\}$

Remark B.13. The value of k' is less than or equal to k. As the process reaches to its end or if $d_{v_{t+1}}(n) > n - m - 2$, we have $k' < k$; hence, it is possible to have $z_0 \in \mathcal{P}_{t+1}$.

Sub-subcase I.1: If $c_n(e_{t+1}) \geq \mathcal{T}_{t+1}$, then the connection e_{t+1} does not survive; however, all the potential neighbors of v_{t+1} has been realized and the vertex v_{t+1} has been explored. In this case, update the sets as follows:

$$\mathcal{A}_{t+1} = \mathcal{A}_t \setminus \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : j \notin \mathcal{P}_{t+1} \text{ and } \{v_{t+1}, j\} \in \mathcal{A}_t\} \tag{B.2a}$$

$$\mathcal{C}_{t+1} = \mathcal{C}_t \tag{B.2b}$$

$$\begin{aligned}\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : j \notin \mathcal{P}_{t+1} \text{ and } C_n(\{v_{t+1}, j\}) \text{ is realized}\} \\ \cup \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : \ j \text{ has been explored and } v_{t+1} \notin \mathcal{P}_j\}\end{aligned} \tag{B.2c}$$

$$\begin{aligned}\mathcal{R}_{t+1} = (\mathcal{R}_t \cup \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : j \in \mathcal{P}_{t+1} \text{ and } j \text{ has not been explored}\}) \\ \setminus \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : j \notin \mathcal{P}_{t+1} \text{ and } \{v_{t+1}, j\} \in \mathcal{R}_t\}\end{aligned} \tag{B.2d}$$

$$\mathcal{U}_{t+1} = \mathcal{U}_t \setminus \{\{v_{t+1}, j\} : C_n(\{v_{t+1}, j\}) \text{ is realized}\} \tag{B.2e}$$

The description of the above equations is as follows:

1. Equation (B.2a): All the active edges $\{v_{t+1}, j\}$ in $\mathcal{A}_t$ such that $j \notin \mathcal{P}_{t+1}$ are removed, including e_{t+1}. Note that if $\{v_{t+1}, j\} \in \mathcal{A}_t$, then $v_{t+1} \in \mathcal{P}_j$ (Remark B.10); however, after exploring the vertex v_{t+1}, it is clear whether j is a potential neighbor of v_{t+1} or not. If $j \notin \mathcal{P}_{t+1}$ then the edge $\{v_{t+1}, j\}$ is moved to $\mathcal{D}_{t+1}$. On the other hand, if $j \in \mathcal{P}_{t+1}$, then $\{v_{t+1}, j\}$ survives; however, this edge needs to be revisited at a later time in order to add new members to the set of active edges.
2. Equation (B.2b): The vertex v_{t+1} is not connected to the connected component through the edge e_{t+1}. Note that there might be some other vertex j such that $\{v_{t+1}, j\} \in \mathcal{A}_t$ and $j \in \mathcal{P}_{t+1}$, i.e., $\{v_{t+1}, j\}$ survives (Remark B.10); however, the exploration of the edge $\{v_{t+1}, j\}$ is postponed to some $t' > t$.
3. Equation (B.2c): All the edges $\{v_{t+1}, j\}$ such that $C_n\{v_{t+1}, j\}$ has been realized and $j \notin \mathcal{P}_{t+1}$ do not survive. Moreover, for all explored vertices j such that

$v_{t+1} \notin \mathcal{P}_j$, the edge $\{v_{t+1}, j\}$ does not survive as well.

4. Equation (B.2d): For all $j \in \mathcal{P}_{t+1}$ such that the vertex j has not been explored, $\{v_{t+1}, j\}$ is added to $\mathcal{R}_{t+1}$. Note that the cost of $\{v_{t+1}, j\}$ has been realized and neither v_{t+1} nor j belong to the connected component. Moreover, for each explored vertex j, if $\{v_{t+1}, j\} \notin \mathcal{R}_t$ then either $v_{t+1} \notin \mathcal{P}_j$ or j belongs to the connected component; hence, $\{v_{t+1}, j\}$ need not be included in $\mathcal{R}_{t+1}$. Finally, for all edges $\{v_{t+1}, j\} \in \mathcal{R}_t$, the vertex v_{t+1} is a potential neighbor of the vertex j; however, if $j \notin \mathcal{P}_{t+1}$ then $\{v_{t+1}, j\}$ does not survive.
5. Equation (B.2e): All the edges $\{v_{t+1}, j\}$ such that $C_n\{v_{t+1}, j\}$ has been realized are removed from $\mathcal{U}_{t+1}$.

Remark B.14. Consider an edge $e = \{v_{t+1}, j\}$ such that the cost of e has been realized. If the vertex $j \notin \mathcal{P}_{t+1}$, then the edge e does not survive and it belongs to $\mathcal{D}_{t+1}$. Now assume $j \in \mathcal{P}_{t+1}$. If the vertex j has not been explored, then e belongs to $\mathcal{R}_{t+1}$. If the vertex j has been explored and $v_{t+1} \notin \mathcal{P}_j$, then the edge e does not survive and it belongs to $\mathcal{D}_{t+1}$. Assume j has been explored and $v_{t+1} \in \mathcal{P}_j$. If j belongs to the connected component, then $e \in \mathcal{A}_t$. If j does not belong to the connected component, then $e \in \mathcal{R}_t$. In either case, e needs no update, and it is included in the corresponding set at time step $t + 1$.

Sub-subcase I.2: If $c_n(e_{t+1}) < \mathcal{T}_{t+1}$, then the connection e_{t+1} survives and v_{t+1} belongs to the connected component. Define $\mathcal{I}_{t+1} = \{p \in \mathcal{P}_{t+1} : \phi(p) \text{ is not defined}\}$. Let $I_{t+1} = \{p_1, p_2, \ldots, p_{|\mathcal{I}_{t+1}|}\}$. Pick a permutation $\boldsymbol{m}_{t+1}$ over $[|\mathcal{I}_{t+1}|]$ uniformly at random and set $\phi(p_l) = (\boldsymbol{j}, \boldsymbol{m}_{t+1}(l))$ for all $l \in [|\mathcal{I}_{t+1}|]$, where $\boldsymbol{j} = \phi(v_{t+1})$. Update the sets as follows,

$$\begin{aligned}\mathcal{A}_{t+1} = \big(&\mathcal{A}_t \cup \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : j \in \mathcal{P}_{t+1} \text{ and } j \text{ has not been explored}\} \\ &\cup \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : j \text{ has been explored and } j \in \mathcal{P}_{t+1}, v_{t+1} \in \mathcal{P}_j\}\big) \\ &\setminus \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : j \text{ belongs to the connected component}\}\end{aligned} \tag{B.3a}$$

$$\mathcal{C}_{t+1} = \mathcal{C}_t \cup \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : j \in \mathcal{P}_{t+1} \text{ and } \{v_{t+1}, j\} \in \mathcal{A}_t\} \tag{B.3b}$$

$$\begin{aligned}\mathcal{D}_{t+1} = \mathcal{D}_t &\cup \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : j \notin \mathcal{P}_{t+1} \text{ and } C_n(\{v_{t+1}, j\}) \text{ is realized}\} \\ &\cup \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : j \text{ has been explored and } v_{t+1} \notin \mathcal{P}_j\}\end{aligned} \tag{B.3c}$$

$$\mathcal{R}_{t+1} = \mathcal{R}_t \setminus \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : \{v_{t+1}, j\} \in \mathcal{R}_t\} \tag{B.3d}$$

$$\mathcal{U}_{t+1} = \mathcal{U}_t \setminus \{\{v_{t+1}, j\} : C_n(\{v_{t+1}, j\}) \text{ is realized}\} \tag{B.3e}$$

The description of the above equations is as follows:

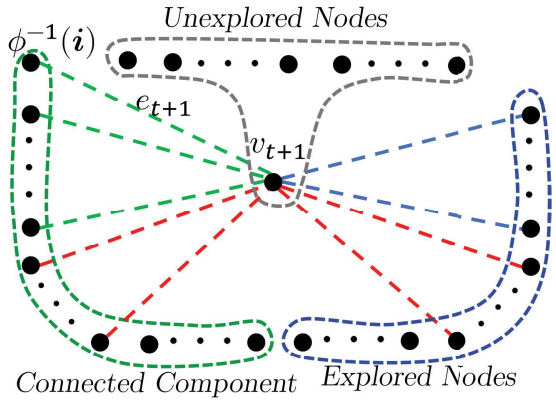
1. Equation (B.3a): All the edges $\{v_{t+1}, j\}$ such that $j \in \mathcal{P}_{t+1}$ and j has not been explored are added to $\mathcal{A}_t$. Moreover, all the edge $\{v_{t+1}, j\}$ such that j has been explored, j do not belongs to the connected component, $j \in \mathcal{P}_{t+1}$ and $v_{t+1} \in \mathcal{P}_j$ are also included in $\mathcal{A}_{t+1}$.
2. Equation (B.3b): The vertex v_{t+1} is connected to the connected component through the edge e_{t+1}; however, all the edges $\{v_{t+1}, j\} \in \mathcal{A}_t$ such that $j \in \mathcal{P}_{t+1}$ are also included in $\mathcal{C}_{t+1}$; since for each edge $\{v_{t+1}, j\} \in \mathcal{A}_t$ the vertex j belongs to the connected component and $v_{t+1} \in \mathcal{P}_j$.
3. Equation (B.3c): All the edges $\{v_{t+1}, j\}$ such that $C_n\{v_{t+1}, j\}$ has been realized and $j \notin \mathcal{P}_{t+1}$ do not survive. Moreover, for all explored vertex j such that $v_{t+1} \notin \mathcal{P}_j$, the edge $\{v_{t+1}, j\}$ does not survive as well.
4. Equation (B.3d): Since v_{t+1} is connected to the connected component, no edge needs to be added to $\mathcal{R}_t$; however, all the edges $\{v_{t+1}, j\} \in \mathcal{R}_t$ is removed from $\mathcal{R}_t$, since one end of such an edge belongs to the connected component.
5. Equation (B.3e): All the edges $\{v_{t+1}, j\}$ such that $C_n\{v_{t+1}, j\}$ has been realized is removed from $\mathcal{U}_{t+1}$.

Remark B.15. Consider an edges $e = \{v_{t+1}, j\}$ such that the cost of e has been realized. If the vertex $j \notin \mathcal{P}_{t+1}$, then the edge e does not survive and it belongs to $\mathcal{D}_{t+1}$. Assume $j \in \mathcal{P}_{t+1}$. If the vertex j has not been explored, then e belongs to $\mathcal{A}_{t+1}$. If the vertex j has been explored and $v_{t+1} \notin \mathcal{P}_j$, then the edge e does not survive, and it belongs to $\mathcal{D}_{t+1}$. Assume j has been explored and $v_{t+1} \in \mathcal{P}_j$. If j belongs to the connected component, then $e \in \mathcal{A}_t$ and e is moved to $\mathcal{C}_{t+1}$. If j does not belong to the connected component, then $e \in \mathcal{R}_t$ and e is moved to $\mathcal{A}_{t+1}$.

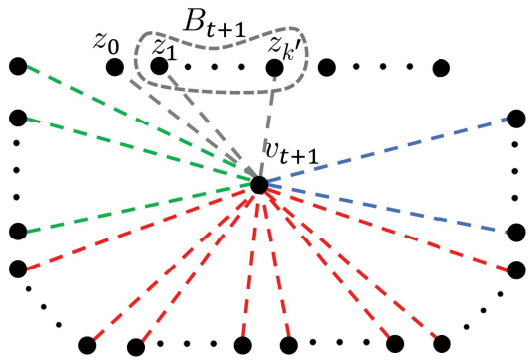
Figure B.3 illustrates the update process for the case where only $\phi^{-1}(\boldsymbol{i})$ has been explored.

- ***Subcase II***, $\phi^{-1}(\boldsymbol{j})$ has been explored: Let v_{t+1} denote the one, amongst $\phi^{-1}(\boldsymbol{j})$ and $\phi^{-1}(\boldsymbol{i})$, which is not connected to the connected component. Since v_{t+1} has already been explored, all the potential neighbors of the vertex v_{t+1} has been realized.

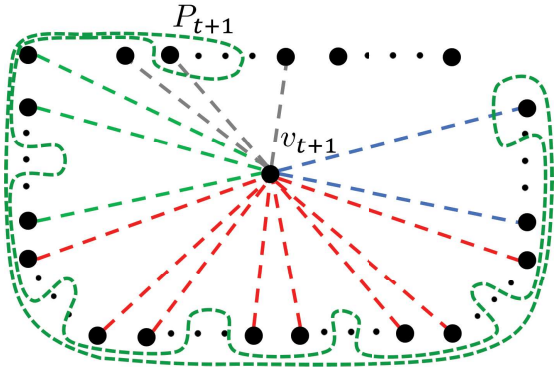
  Remark B.16. Since the vertex v_{t+1} has been explored and it does not belong to the connected component by time t, there is a vertex $v \in [n]$, which belongs to the connected component of r by time t and $v_{t+1} \in \mathcal{P}_v$ and $\{v, v_{t+1}\} \in \mathcal{D}_t$. Note that v may or may not be $par(v_{t+1})$. To clarify the reason, consider the following cases,



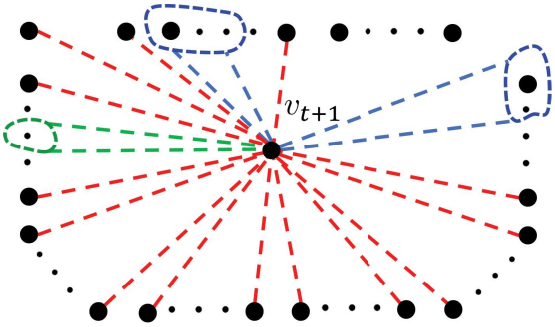
(a) Let $e_{t+1} = \{\phi^{-1}(i), \phi^{-1}(j)\}$ and $v_{t+1} = \phi^{-1}(j)$. Red dashed lines belong to D_t, green dashed lines belong to A_t and blue dashed lines belong to R_t.



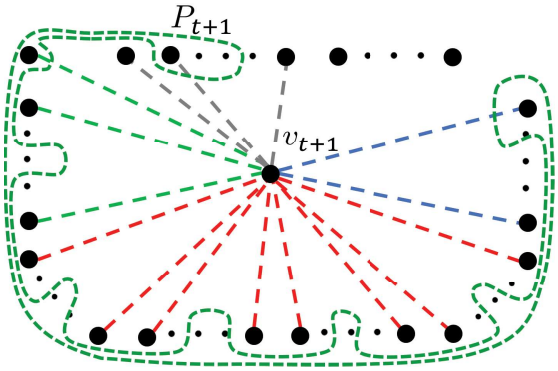
(b) Realize B_{t+1}, z_0 and the corresponding edge costs. Realize the cost of all edges $\{v_{t+1}, j\}$ for explored vertices j as well. For each explored vertex j such that $\{v_{t+1}, j\} \in U_t$, we have $\{v_{t+1}, j\} \in D_{t+1}$.



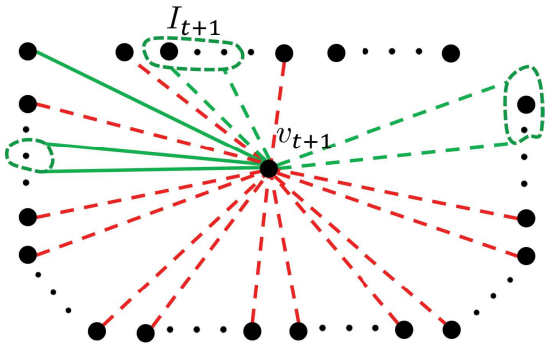
(c1) Realize T_{t+1} and P_{t+1}. Consider the case where e_{t+1} does not survive, i.e., $c_n(e_{t+1}) \geq T_{t+1}$.



(d1) Update the sets for time step $t+1$.Red dashed lines belong to D_{t+1}, green dashed lines belong to A_{t+1} and blue dashed lines belong to R_{t+1}.



(c2) Realize T_{t+1} and P_{t+1}. Consider the case where e_{t+1} survives, i.e., $c_n(e_{t+1}) < T_{t+1}$. Define $I_{t+1} = \{p_1, p_2, \ldots, p_{|I_{t+1}|}\}$ such that $\phi(p)$ is not defined for all $p \in I_{t+1}$. Set $\phi(p_l) = (j, l)$ for all $l \in [|I_{t+1}|]$.



(d2) Update the sets for time step $t+1$. Red dashed lines belong to D_{t+1}, green dashed lines belong to A_{t+1}, and solid green lines belong to C_{t+1}.

Figure B.3: The exploration process at time step t, when the vertex $\phi^{-1}(j)$ has not been explored. (c1) and (d1) illustrate the case when $c_n(e_{t+1}) \geq \mathcal{T}_{t+1}$ while (c2) and (d2) illustrate the case when $c_n(e_{t+1}) < \mathcal{T}_{t+1}$.

1. Consider the case where $\phi^{-1}(\boldsymbol{j})$ belongs to the connected component. As is mentioned in Remark B.8, the vertex $par(\phi^{-1}(\boldsymbol{i}))$ has been explored; hence, $\{par(\phi^{-1}(\boldsymbol{i})), \phi^{-1}(\boldsymbol{i})\} \in \mathcal{D}_t$. In Figure B.2, at $t = 4$, we have $\boldsymbol{i} = (2)$ and $\phi^{-1}(2) = b$, and $\boldsymbol{j} = (3, 1)$ and $\phi^{-1}(\boldsymbol{j}) = d$; however, d belongs to the connected component and b does not and the edge $\{par(b), b\} = \{r, b\} \in \mathcal{D}_4$.
2. Consider the case where $\phi^{-1}(\boldsymbol{i})$ belongs to the connected component. In this case, the edge $\{par(\phi^{-1}(\boldsymbol{j})), \phi^{-1}(\boldsymbol{j})\}$ may belong to $\mathcal{A}_t$. In Figure B.2, at $t = 5$, we have $\boldsymbol{i} = (2)$ and $\phi^{-1}(2) = b$, and $\boldsymbol{j} = (3, 1, 1)$ and $\phi^{-1}(\boldsymbol{j}) = g$; assuming $b \notin \mathcal{P}_g$ and $d \in \mathcal{P}_g$, the connection e_6 does not survive but the vertex g is explored and $\{par(g), g\} = \{d, g\} \in \mathcal{A}_6$.

Without loss of generality, assume $\phi^{-1}(\boldsymbol{i})$ belongs to the connected component; hence, $v_{t+1} = \phi^{-1}(\boldsymbol{j})$. Define $k \coloneqq d_{v_{t+1}}(n)$ and set the threshold and the set of potential neighbors of v_{t+1},

$$\mathcal{T}_{t+1} = k + 1^{th} \text{ smallest value in } \{c_n(\{v_{t+1}, j\}) : j \in [n] \text{ and } \{v_{t+1}, j\} \in \mathcal{R}_t \cup \mathcal{A}_t \cup \mathcal{D}_t\}$$
$$\mathcal{P}_{t+1} = \{j \in [n] : \{v_{t+1}, j\} \in \mathcal{R}_t \cup \mathcal{A}_t \cup \mathcal{D}_t \text{ and } c_n(\{v_{t+1}, j\}) < \mathcal{T}_{t+1}\}$$

Remark B.17. Given that both $\phi^{-1}(\boldsymbol{i})$ and $\phi^{-1}(\boldsymbol{j})$ have been explored and one of them does not belong to the connected component, the survival of $\{\phi^{-1}(\boldsymbol{i}), \phi^{-1}(\boldsymbol{j})\}$ should have been determined, i.e., it survives. The edge $\{\phi^{-1}(\boldsymbol{i}), \phi^{-1}(\boldsymbol{j})\}$ has been added to the set of active edges to revisit the vertex v_{t+1} and add new potential edges to $\mathcal{A}_t$.

As is mentioned in Remark B.17, the connection e_{t+1} survives and v_{t+1} belongs to the connected component. Define $I_{t+1} = \{z \in \mathcal{P}_{t+1} : \phi(z) \text{ is not defined}\}$. Let $I_{t+1} = \{z_1, z_2, \ldots, z_{|I_{t+1}|}\}$. Pick a permutation π_{t+1} over $[|I_{t+1}|]$ uniformly at random and set $\phi(z_l) = (\boldsymbol{j}, \pi_{t+1}(l))$ for all $l \in [|I_{t+1}|]$, where $\boldsymbol{j} = \phi(v_{t+1})$. Update the sets as follows,

$$\begin{aligned}\mathcal{A}_{t+1} = (\mathcal{A}_t \cup \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : \{v_{t+1}, j\} \in \mathcal{R}_t\}) \\ \setminus \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : \{v_{t+1}, j\} \in \mathcal{A}_t\}\end{aligned} \tag{B.4a}$$
$$\mathcal{C}_{t+1} = \mathcal{C}_t \cup \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : \{v_{t+1}, j\} \in \mathcal{A}_t\} \tag{B.4b}$$
$$\mathcal{D}_{t+1} = \mathcal{D}_t \tag{B.4c}$$
$$\mathcal{R}_{t+1} = \mathcal{R}_t \setminus \{(\{v_{t+1}, j\}, c_n(\{v_{t+1}, j\})) : \{v_{t+1}, j\} \in \mathcal{R}_t\} \tag{B.4d}$$
$$\mathcal{U}_{t+1} = \mathcal{U}_t \tag{B.4e}$$

The description of the above equations is as follows:

1. Equation (B.4a): All the edges $\{v_{t+1}, j\} \in \mathcal{R}_t$ is added to $\mathcal{A}_t$; since, for every $\{v_{t+1}, j\} \in \mathcal{R}_t$, the vertex j is a potential neighbor of v_{t+1} and if j has been explored, then $v_{t+1} \in \mathcal{P}_j$ as well. In addition, all the edges $\{v_{t+1}, j\} \in \mathcal{A}_t$ are removed from $\mathcal{A}_t$; since, j belongs to the connected component at time t (Remark B.10), the edge $\{v_{t+1}, j\}$ survives (Remark B.17) and we do not need to revisit the vertex v_{t+1} at a later time.
2. Equation (B.4b): All the edges $\{v_{t+1}, j\} \in \mathcal{A}_t$ are moved to $\mathcal{C}_{t+1}$; since, if $\{v_{t+1}, j\} \in \mathcal{A}_t$ then $j \in \mathcal{P}_{t+1}$, $v_{t+1} \in \mathcal{P}_j$ and the vertex j belongs to the connected component (Remark B.10 and Remark B.17).
3. Equation (B.4c): Note that both $\phi^{-1}(\boldsymbol{i})$ and $\phi^{-1}(\boldsymbol{j})$ have been explored; hence, the cost of none of the edges in $\mathcal{U}_t$ is realized and the set $\mathcal{D}_t$ needs no update.
4. Equation (B.4d): All the edges $\{v_{t+1}, j\} \in \mathcal{R}_t$ are removed from $\mathcal{R}_t$, since exactly one end of such an edge belongs to the connected component. All of these edges are moved to $\mathcal{A}_{t+1}$.
5. Equation (B.4e): The cost of none of the edges in $\mathcal{U}_t$ is realized; hence, $\mathcal{U}_t$ needs no update.

Remark B.18. Consider an edges $e = \{v_{t+1}, j\}$ with realized cost. If $e \in \mathcal{A}_t$, then j belongs to the connected component, $v_{t+1} \in \mathcal{P}_j$ (Remark B.10) and $j \in \mathcal{P}_{t+1}$ (vertex v_{t+1} has been explored); hence, e is moved to $\mathcal{C}_{t+1}$. If the edge $e \in \mathcal{D}_t$, then e needs no update. If the edge $e \in \mathcal{R}_t$, then e is moved to $\mathcal{A}_{t+1}$ since v_{t+1} belongs to the connected component. Finally, e does not belong to $\mathcal{U}_t$ nor $\mathcal{C}_t$.

Remark B.19. Recall that for any $\{v, z\} \in \mathcal{R}_t$, if v has been explored then $z \in \mathcal{P}_v$. Moreover, neither z nor v belongs to the connected component of r by time t.

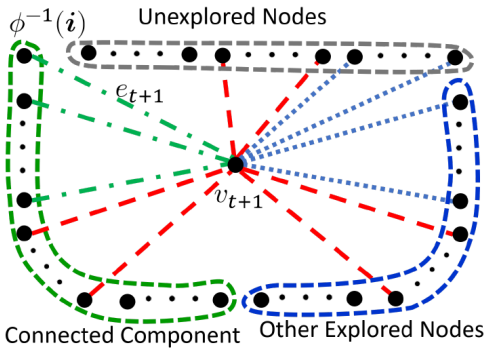
Figure B.4 illustrates the updating process for the case where both $\phi^{-1}(\boldsymbol{i})$ and $\phi^{-1}(\boldsymbol{j})$ have been explored.

Exploration phase: The exploration terminates when $\mathcal{A}_t = \emptyset$. Consider the following filtration,

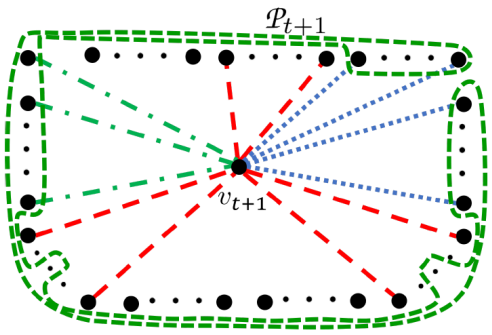
$$\mathcal{F}_t = \sigma((\mathcal{A}_0, \mathcal{C}_0, \mathcal{D}_0, \mathcal{R}_0, \mathcal{U}_0), \ldots, (\mathcal{A}_t, \mathcal{C}_t, \mathcal{D}_t, \mathcal{R}_t, \mathcal{U}_t))$$

Let τ denote the time that the algorithm terminates. Indeed, τ is a stopping time of the filtration where $\tau = \inf\{t \geq 1 : \mathcal{A}_t = \emptyset\}$.

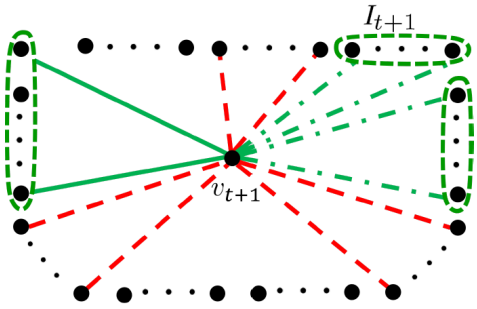
Step 2: Locally tree-like property



(a) Let $e_{t+1} = \{\phi^{-1}(i), \phi^{-1}(j)\}$ and $v_{t+1} = \phi^{-1}(j)$. Dashed red edges belong to $\mathcal{D}_t$, dashed dotted green edges belong to $\mathcal{A}_t$, and dotted blue edges belong to $\mathcal{R}_t$.



(b) Determine the set of potential neighbors of v_{t+1}, i.e., $\mathcal{P}_{t+1}$. Define $I_{t+1} = \{z_1, z_2, \ldots, z_{|I_{t+1}|}\}$ such that $\phi(z)$ is not defined for all $z \in I_{t+1}$. Set $\phi(z_l) = (j, m_{t+1}(l))$ for all $l \in [|I_{t+1}|]$.



(c) Update the sets for time step $t+1$. Solid green edges belong to $\mathcal{C}_{t+1}$, dashed red edges belong to $\mathcal{D}_{t+1}$, and dashed dotted green edges belong to $\mathcal{A}_{t+1}$.

Figure B.4: The exploration process at time step t, when both the vertices $\phi^{-1}(j)$ and $\phi^{-1}(i)$ have been explored.

In the second step, the goal is to show that the rooted graph induced by $\mathcal{C}_{t\wedge\tau}$ for any fixed t becomes a tree as the number of vertices, n, goes to infinity. This implies that the graph G_n, induced by the network N_n after removing the marks, is asymptotically locally tree-like. In fact, a stronger property holds: for every fixed $t > 0$, the probability that the vertex v_l, for all $l \in [t \wedge \tau]$, has been *touched* twice during the exploration process prior to time step l goes to zero as $n \to \infty$. The term "touching" is defined as follows,

Definition B.20. A vertex v is said to be touched at time $t' \leq \tau$ if the cost of $\{v_{t'}, v\}$ is realized at time t', *i.e.*, $\{v_{t'}, v\} \in \mathcal{U}_{t'-1} \setminus \mathcal{U}_{t'}$. The vertex $v_{t'}$ is chosen according to the exploration process. Note that the vertex v may have or may not have been explored.

If for every $l \in [t \wedge \tau]$, the vertex v_l has been touched only once before the time step l, then $e_l = \{par(v_l), v_l\}$; moreover, for every $l' < l$ such that $v_{l'} \neq par(v_l)$, the vertex v_l is not the potential neighbor of the vertex $v_{l'}$. This implies that the rooted graph induced by $\mathcal{C}_{t\wedge\tau}$

is a tree. A stronger condition is proved in the following lemma: with high probability, for all $l \in [t \wedge \tau]$ the potential neighbors of the vertex v_l are touched for the first time, except maybe $par(v_l)$.

Lemma B.21. Locally tree-like property *For $t' > 0$, let $J_{t'}$ denote the set of vertices j such that $C_n(\{v_{t'}, j\}) \leq \mathcal{T}_{t'}$ and j has been touched at least twice during the exploration process up to time t', once at time step t' and at least once at some time step $\widetilde{t} < t'$, i.e.,*

$$J_{t'} = \{j \in [n] : C_n(\{v_{t'}, j\}) \leq \mathcal{T}_{t'}, \{v_{t'}, j\} \in \mathcal{U}_{t'-1} \setminus \mathcal{U}_{t'}, \text{ and } \exists \widetilde{v} \neq v_{t'} \text{ such that } \{\widetilde{v}, j\} \notin \mathcal{U}_{t'-1}\}$$

Consider a fixed value $t > 0$, then we have,

$$\lim_{n \to \infty} \mathbb{P}(\exists l \in [t \wedge \tau] \text{ such that } |J_l| \neq 0) = 0. \tag{B.5}$$

Remark B.22. Consider the event $\{J_l = \emptyset\}$ for all $l \in [t \wedge \tau]$. This implies that for every vertex j such that $C_n(\{v_l, j\}) \leq \mathcal{T}_l$, either j is touched for the first time at time step l or the value of $C_n(\{v_l, j\})$ has been realized by time step $l-1$. However, if $j \neq par(v_l)$, then the later case is impossible; otherwise, the vertex v_l should have been touched at least twice during the exploration process up to time $l-1$: once when we realized $C_n(\{v_l, j\})$ and once when we realized $C_n(\{par(v_l), v_l\})$.

Remark B.23. Even if the rooted graph induced by $\mathcal{C}_{t \wedge \tau}$ is a tree, it does not mean that the exploration process satisfies the property which is mentioned in Lemma B.21. In Figure B.2, vertex b has been touched twice during the exploration process up to time step $t = 1$: at time steps $t = 0$ (by the vertex a) and $t = 1$; however, $\mathcal{C}_1$ is a tree.

Proof. Observe that $J_0 = \emptyset$. Fix $t > 0$. An obvious upper-bound for the left-hand side of (B.5) is given by applying the union bound:

$$\begin{aligned}
\mathbb{P}(\exists l \in [t \wedge \tau] \text{ such that } |J_l| \neq 0) &= \mathbb{P}\left(\bigcup_{l=1}^{t \wedge \tau} \{|J_l| \neq 0\}\right) \\
&= \mathbb{P}\left(\bigcup_{l=1}^{t} (\{l \leq \tau\} \cap \{|J_l| \neq 0\})\right) \\
&\leq \sum_{l=1}^{t} \mathbb{P}(\{l \leq \tau\} \cap \{|J_l| \neq 0\}) \\
&= \sum_{l=1}^{t} \mathbb{E}\left[\mathbb{1}\{l \leq \tau\} \mathbb{P}(\{|J_l| \neq 0\} \,|\mathcal{F}_{l-1})\right].
\end{aligned} \tag{B.6}$$

We provide an upper-bound for each term on the right-hand side. If the vertex v_l has been explored by time step $l-1$, then we do not need to touch any vertex at time l and $J_l = \emptyset$. In Figure B.2, the vertex b has already been explored at time step $t = 2$ and $J_5 = \emptyset$. Hence, we only need to consider sample paths where v_l has not been explored. Thus,

$$\mathbb{P}(|J_l| = 0|\mathcal{F}_{l-1}) = \mathbb{1}\{v_l \text{ has been explored}\} + \mathbb{1}\{v_l \text{ has not been explored}\}\mathbb{P}(|J_l| = 0|\mathcal{F}_{l-1}).$$

Consider the sets δ_l, ε_l, $\epsilon_l \in \mathcal{F}_{l-1}$ defined as follows,

1. δ_l: set of vertices $j \neq v_l$ with at least one incident edge such that the cost of the edge has been realized prior to time l. Equivalently, δ_l is the set of all vertices except v_l that have been touched prior to time l,

$$\delta_l = \{j \in [n] \setminus \{v_l\} : \exists i \neq v_l \text{ such that, } \{i, j\} \notin \mathcal{U}_{l-1}\}.$$

2. ε_l: set of all vertices except v_l that have been explored prior to time l,

$$\varepsilon_l = \{j \in [n] \setminus \{v_l\} : j \text{ has been explored by } l-1\}.$$

3. ϵ_l: set of vertices $j \neq v_l$ such that the cost of $\{v_l, j\}$ has been realized prior to time l,

$$\epsilon_l = \{j \in [n] \setminus \{v_l\} : \{v_l, j\} \notin \mathcal{U}_{l-1}\}.$$

Observe that $|\epsilon_l| \geq 1$ since $v_l \in \mathcal{A}_l$. Moreover, at each step of the time we may explore at most one vertex (there might be cases in which we revisit an explored vertex); hence, $|\varepsilon_l| \leq l$. Furthermore, for all sample paths in $\mathcal{F}_{l-1}$ in which v_l has not been explored, $\epsilon_l \subseteq \varepsilon_l$ since if $\{v_l, j\}$ has been realized and v_l has not been explored, then j has been explored. Finally, at each time step l', we may touch at most $d_{v_{l'}}(n) + 1$ new vertices; hence, $|\delta_l| \leq 1 + \sum_{i=0}^{l-1} \left(d_{v_i(n)} + 1\right)$

Let $k \coloneqq d_{v_l}(n)$ denote the potential degree of the vertex v_l. Let $\widetilde{k} \coloneqq \min(k, n - |\epsilon_l| - 2)$, where $n - |\epsilon_l| - 1$ equals to the number of vertices j such that $\{v_l, j\} \in \mathcal{U}_{l-1}$. Note that $n - |\epsilon_l| - 1 > 0$ if v_l has not been explored and $n > l$. Define $\widetilde{\mathcal{T}}_l$ and $\widetilde{\mathcal{P}}_l$ to be modified versions of $\mathcal{T}_l$ and $\mathcal{P}_l$, i.e.,

$$\widetilde{\mathcal{T}}_l = \widetilde{k} + 1^{th} \text{ smallest value in } \{C_n(\{v_l, j\}) : j \in [n] \text{ and } \{v_l, j\} \in \mathcal{U}_{l-1}\}$$
$$\widetilde{\mathcal{P}}_l = \{j \in [n] : \{v_l, j\} \in \mathcal{U}_{l-1} \text{ and } C_n(\{v_l, j\}) < \widetilde{\mathcal{T}}_l\}$$

Recall that $\mathcal{T}_l$ and $\mathcal{P}_l$ is defined as follows,

$$\mathcal{T}_l = k + 1^{th} \text{ smallest value in } \{C_n(\{v_l, j\}) : j \in [n]\}$$
$$\mathcal{P}_l = \{j \in [n] : C_n(\{v_l, j\}) < \mathcal{T}_l\}$$

In the definition of $\mathcal{T}_l$, all possible vertices are considered; however, the definition of $\widetilde{\mathcal{T}}_l$ skips all the vertices j, such that $\{v_l, j\}$ has been realized prior to time step l. Hence, if $\widetilde{k} = k$, then $\mathcal{T}_l \leq \widetilde{\mathcal{T}}_l$. Moreover, for every vertex $j \in \mathcal{P}_l$ such that the cost of $\{v_l, j\}$ is realized at time l, i.e., $\{v_l, j\} \in \mathcal{U}_{l-1}$, we have $j \in \widetilde{\mathcal{P}}_l$. To see this, consider the two cases: 1)If $\widetilde{k} = k$, then $j \in \mathcal{P}_l$ implies $C_n(\{v_l, j\}) < \mathcal{T}_l \leq \widetilde{\mathcal{T}}_l$. 2)If $\widetilde{k} < k$, then $\widetilde{\mathcal{P}}_l$ contains all the vertices j such that $\{v_l, j\} \in \mathcal{U}_{l-1}$.

To realize $\widetilde{\mathcal{T}}_l$ and $\widetilde{\mathcal{P}}_l$, we need to pick the $\widetilde{k}+1$ closest vertices to v_l, based on the cost of the connection. For an unexplored vertex j, the cost of $\{v_l, j\}$ is an exponentially distributed random variable with parameter $\frac{1}{n}$. For an explored vertex j such that $\{v_l.j\} \in \mathcal{U}_{l-1}$, the cost of $\{v_l, j\}$ conditioned on $\widehat{\mathcal{T}}_j$ is a shifted exponentially distributed random variable with parameter $\frac{1}{n}$ (Corollary B.3); that is $C_n(\{v_l, j\}) \equiv \widehat{\mathcal{T}}_j + \exp\big(\frac{1}{n}\big)$, where $\widehat{\mathcal{T}}_j$ is defined before Remark B.11. Hence, we need to pick the $\widetilde{k}+1$ smallest value in $H_1 \cup H_2$ where,

$$\begin{aligned} H_1 &= \{C_n(\{v_l, j\}) : j \in [n], j \text{ has not been explored and } \{v_l, j\} \in \mathcal{U}_{l-1}\} \\ &\equiv \left\{Y_1, Y_2, \ldots, Y_{n-|\varepsilon_l \cup \epsilon_l|-1} : Y_i \overset{i.i.d.}{\sim} \exp\left(\frac{1}{n}\right)\right\} \end{aligned}$$

and,

$$\begin{aligned} H_2 &= \{C_n(\{v_l, j\}) : j \in [n], j \text{ has been explored and} \{v_l, j\} \in \mathcal{U}_{l-1}\} \\ &\equiv \left\{\exp\left(\frac{1}{n}\right) + \widehat{\mathcal{T}}_j : j \in [n], j \text{ has been explored and} \{v_l, j\} \in \mathcal{U}_{l-1}\right\} \end{aligned}$$

Instead of H_2 we consider $\widehat{H}_2$, defined as follows,

$$\begin{aligned} \widehat{H}_2 &= \left\{C_n(\{v_l, j\}) - \widehat{\mathcal{T}}_j : j \in [n], j \text{ has been explored and} \{v_l, j\} \in \mathcal{U}_{l-1}\right\} \\ &\equiv \left\{Y'_1, Y'_2, \ldots, Y'_{|\varepsilon_l \setminus \epsilon_l|} : Y'_i \overset{i.i.d.}{\sim} \exp\left(\frac{1}{n}\right)\right\} \end{aligned}$$

In fact, $\widehat{H}_2$ is obtained by replacing $C_n(\{v_l, j\})$ with $C_n(\{v_l, j\}) - \widehat{\mathcal{T}}_j$ for all explored vertices j such that $\{v_l, j\} \in \mathcal{U}_{l-1}$. Note that if $\widetilde{k}+1$ smallest values in $H_1 \cup H_2$ correspond to $\{u_0, u_1, \ldots, u_{\widetilde{k}}\}$, then the $\widetilde{k}+1$ smallest values in $H_1 \cup \widehat{H}_2$ correspond to $\{\widehat{u}_0, \widehat{u}_1, \ldots, \widehat{u}_{\widetilde{k}}\}$ where $\widehat{u}_i$ is either u_i or $u_i - \mathcal{T}_j$ for some explored vertex $j \in [n]$. Note that if a member of H_2

is amongst $\widetilde{k}+1$ smallest values in $H_1 \cup H_2$, then the corresponding element is also amongst $\widetilde{k}+1$ smallest values in $H_1 \cup \widehat{H}_2$. Collecting everything together, we have

$$\begin{aligned}
\mathbb{P}(|J_l|=0|\mathcal{F}_{l-1}) &= \mathbb{P}\left(\left\{\begin{array}{l}\forall\{v_l,u\}\in\mathcal{U}_{l-1} \text{ such that } C_n(\{v_l,u\})\leq\mathcal{T}_l, \text{ the}\\ \text{vertex } u \text{ is touched for the first time at time}\\ \text{step } l\end{array}\right\}\middle|\mathcal{F}_{l-1}\right)\\
&\geq \mathbb{P}\left(\left\{\begin{array}{l}\forall\{v_l,u\}\in\mathcal{U}_{l-1} \text{ such that } C_n(\{v_l,u\})\leq\widetilde{\mathcal{T}}_l, \text{ the}\\ \text{vertex } u \text{ is touched for the first time at time}\\ \text{step } l\end{array}\right\}\middle|\mathcal{F}_{l-1}\right)\\
&= \mathbb{P}\left(\left\{\begin{array}{l}\text{The } \widetilde{k}+1 \text{ smallest values in } H_1\cup H_2 \text{ correspond}\\ \text{to the vertices that are touched for the first}\\ \text{time at time step } l\end{array}\right\}\middle|\mathcal{F}_{l-1}\right)\\
&\geq \mathbb{P}\left(\left\{\begin{array}{l}\text{The } \widetilde{k}+1 \text{ smallest values in } H_1\cup \widehat{H}_2 \text{ correspond}\\ \text{to the vertices that are touched for the first}\\ \text{time at time step } l\end{array}\right\}\middle|\mathcal{F}_{l-1}\right),
\end{aligned}$$

where the last inequality follows from the fact that members of H_2 correspond to the vertices that have been touched before time step l (note that some members of H_1 may also correspond to the vertices that have already been touched.). However, all the values in $H_1 \cup \widehat{H}_2$ are independent and exponentially distributed with parameter $\frac{1}{n}$. There are $n-|\epsilon_l|-1$ vertices $j \neq v_l$ such that $\{v_l,j\}\in\mathcal{U}_{l-1}$ and the number of the vertices j that has not been touched prior to time step l is $n-|\delta_l|-1$; hence,

$$\begin{aligned}
&\mathbb{P}\left(\left\{\begin{array}{l}\text{The } \widetilde{k}+1 \text{ smallest values in } H_1\cup \widehat{H}_2 \text{ correspond}\\ \text{to the vertices that are touched for the first}\\ \text{time at time step } l\end{array}\right\}\middle|\mathcal{F}_{l-1}\right)\\
&\qquad = \frac{\binom{n-|\delta_l|-1}{\widetilde{k}+1}}{\binom{n-|\epsilon_l|-1}{\widetilde{k}+1}} \geq \frac{\binom{\max\left(0,n-\left(\sum_{i=0}^{l-1}\left(d_{v_i}(n)+1\right)\right)-2\right)}{\widetilde{k}+1}}{\binom{n-1}{\widetilde{k}+1}}\\
&\qquad \geq \left(\max\left(0,\frac{n-l-\sum_{i=0}^{l}d_{v_i}(n)-2}{n}\right)\right)^{d_{v_l}(n)+1}.
\end{aligned}$$

Recall that $\widetilde{k}\leq k = d_{v_l}(n)$. Finally, $\sum_{i=0}^{l} d_{v_i}(n) < M$ with arbitrary high probability for a large enough constant M since the unique elements of the sequence $(d_{v_i}(n))_{i=0}^{l}$ are chosen uniformly at random (without replacement) from $\boldsymbol{d_n}$ and empirical distribution of $\boldsymbol{d_n}$ converges to P; hence,

$$\mathbb{E}\left[\mathbb{1}\{l\leq\tau\}\mathbb{P}\left(\{|J_l|\neq 0\}\,|\mathcal{F}_{l-1}\right)\right]$$

$$= \mathbb{E}\left[\mathbb{1}\{l \le \tau\}\left(1 - \mathbb{P}\left(\{|J_l| = 0\}\,|\mathcal{F}_{l-1}\right)\right)\right]$$
$$\le 1 - \mathbb{E}\left[\left(\max\left(0, 1 - \frac{l + \sum_{i=0}^{l} d_{v_i}(n) + 2}{n}\right)\right)^{d_{v_l}(n)+1}\right] \xrightarrow{n\to\infty} 0,$$

using the law of total probability. Now, the result follows from the fact that the summation in (B.6) has only t summands, each of which converges to zero as n goes to ∞. □

Step 3: Convergence of the Exploration

In the third step, we study the local structure of the rooted graph induced by $\mathcal{C}_{t\wedge\tau}$ for any fixed t. The goal is to analyze the joint distribution of the sequence $(X_0^{(n)}, X_1^{(n)}, X_2^{(n)}, \ldots, X_{t\wedge\tau}^{(n)})$ as n goes to infinity, where

$$X_0^{(n)} := \left(d_{v_0}(n), \mathcal{T}_0, C_n(\{v_0, j_1\}), C_n(\{v_0, j_2\}), C_n(\{v_0, j_3\}), \ldots, C_n(\{v_0, j_{d_{v_0}(n)}\})\right)$$

such that $C_n(\{v_0, j_s\}) < \mathcal{T}_0$ for all $s \in [d_{v_0}(n)]$ and $\phi(j_1) \prec \phi(j_2) \prec \cdots \phi(j_{d_{v_0}(n)})$, and for all $l \in [t \wedge \tau]$ the random vector $X_l^{(n)}$ is given by

$$X_l^{(n)} := \left(d_{v_l}(n), \overline{\mathcal{T}}_l, C_n(\{v_l, j_1\}), C_n(\{v_l, j_2\}), C_n(\{v_l, j_3\}), \ldots, C_n(\{v_l, j_{d_{v_l}(n)-1}\})\right)$$

such that $C_n(\{v_l, j_s\}) < \overline{\mathcal{T}}_l$ for all $s \in [d_{v_l}(n) - 1]$ and $\phi(j_1) \prec \phi(j_2) \prec \cdots \phi(j_{d_{v_l}(n)-1})$; the term $\overline{\mathcal{T}}_l$ is the $d_{v_l}(n)^{th}$ smallest value in the set $\{C_n(\{v_l, j\}) : j \in [n] \text{ and } \{v_l, j\} \neq e_l\}$. Note that the second component of $X_0^{(n)}$ equals the threshold of the vertex v_0 and the remaining components correspond to the cost of connections between v_0 and its potential neighbors. Recall that $d_{v_l}(n)$ is the potential degree of vertex v_l, and that the edge e_l is picked according to the exploration process.

An important observation is that for each $l \in [t \wedge \tau]$, $\mathcal{T}_l = \overline{\mathcal{T}}_l$ if $C_n(e_l) < \overline{\mathcal{T}}_l$; moreover, if $C_n(e_l) > \overline{\mathcal{T}}_l$, then the edge e_l does not survive (note that by Remark III.1, we have $C_n(e_l) \neq \overline{\mathcal{T}}_l$.). Hence, the first two components of $X_l^{(n)}$ together are the type of the vertex v_l if and only if the edge e_l survives. Note that the value of $X_l^{(n)}$ depends on the number of vertices.

Let us extend the sequence to $(X_0^{(n)}, X_1^{(n)}, X_2^{(n)}, \ldots, X_t^{(n)})$: for each $l > t \wedge \tau$, the first component of $X_l^{(n)}$ is defined to be $d_v(n)$ where the vertex v is chosen uniformly at random such that $v \notin \{v_0, v_1, \ldots, v_{l-1}\}$, the second component is set to be $d_v(n)^{th}$ smallest value in $\mathcal{S}_l^{(n)} = \{s_1, s_2, \ldots, s_n : s_i \overset{i.i.d.}{\sim} \exp\left(\frac{1}{n}\right)\}$, and, the remaining components are defined to be $(s_{l_1}, s_{l_2}, \ldots, s_{l_{d_v(n)-1}})$ such that $l_1 < l_2 < \cdots < l_{d_v(n)-1}$ and s_{l_i} is among the $d_v(n) - 1$ smallest values in $\mathcal{S}_l^{(n)}$.

The following Lemma states that the sequence $(X_0^{(n)}, X_1^{(n)}, X_2^{(n)}, \ldots, X_t^{(n)})$ has the same distribution as the corresponding sequence $(X_0, X_1, X_2, \ldots, X_t)$ generated by $Er(P)$ and extended up to time t. The proof is given by using a coupling argument.

Lemma B.24. Convergence of the Exploration Process *The sequence* $(X_0^{(n)}, X_1^{(n)}, \ldots, X_t^{(n)})$ *converges to the sequence* $(X_0, X_1, \ldots, X_t)$*,*

$$\begin{aligned} X_0 &\coloneqq (\mathrm{D}_0, \mathrm{T}_0, \mathrm{C}_1^{(0)}, \mathrm{C}_2^{(0)}, \ldots, \mathrm{C}_{\mathrm{D}_l}^{(0)}) \\ X_l &\coloneqq (\mathrm{D}_l, \mathrm{T}_l, \mathrm{C}_1^{(l)}, \mathrm{C}_2^{(l)}, \ldots, \mathrm{C}_{\mathrm{D}_l-1}^{(l)}) \qquad \forall l > 0, \end{aligned}$$

in distribution where D_l *is distributed as* $P(\cdot)$ *for all* $l \geq 0$*,* T_l *is distributed as* $Erlang(\mathrm{D}_l)$ *for all* $l \in [t]$ *and* T_0 *is distributed as* $Erlang(\mathrm{D}_0 + 1)$*,* $\{\mathrm{C}_i^{(l)}\}_i$ *are* i.i.d. *random variables uniformly distributed on* $[0, \mathrm{T}_l]$ *for all* $l \geq 0$*, and* X_i*s are independent.*

Proof. Fix the value of n. Let $l > 0$ and consider the random vector

$$\widetilde{X}_l^{(n)} = (\widetilde{\mathrm{d}}_{(l)}(n), \widetilde{\mathrm{T}}_l, \widetilde{\mathrm{C}}_1, \widetilde{\mathrm{C}}_2, \ldots, \widetilde{\mathrm{C}}_{\widetilde{\mathrm{d}}_{(l)}(n)-1})$$

where $\big(\widetilde{\mathrm{d}}_{(i)}(n)\big)_{i=0}^{n-1}$ is a random reordering of $\big(d_i(n)\big)_{i=1}^{n}$, $\widetilde{\mathrm{T}}_l$ is the $\widetilde{\mathrm{d}}_{(l)}(n)^{th}$ smallest value in $\mathcal{S}_l^{(n)} = \{s_1, s_2, \ldots, s_{n-2} : s_i \overset{i.i.d.}{\sim} \exp\big(\frac{1}{n}\big)\}$ and $\widetilde{\mathrm{C}}_i$ equals to s_{l_i} where $l_1 < l_2 < \cdots < l_{\widetilde{\mathrm{d}}_{(l)}(n)-1}$ and $s_{l_i} < \widetilde{\mathrm{T}}_l$. Using Corollary B.3, it is easy to see that for any fixed $l > 0$, $\widetilde{X}_l^{(n)}$ converges in distribution to $(\mathrm{D}_l, \mathrm{T}_l, \mathrm{C}_1^{(l)}, \mathrm{C}_2^{(l)}, \ldots, \mathrm{C}_{\mathrm{D}_l-1}^{(l)})$ as n goes to infinity. Similarly, for a proper definition of $\widetilde{X}_0^{(n)}$, the same property holds. Note that the distribution of $\widetilde{X}_l^{(n)}$ depends on n.

The idea of the proof is to first construct a coupling between $(X_l^{(n)})_{l=0}^t$ and $(Y_l^{(n)})_{l=0}^t$ where conditioned on $\bigcap_{l \in [t \wedge \tau]} \{J_l = \emptyset\}$, $(Y_l^{(n)})_{l=0}^t$ has the same distribution as $(\widetilde{X}_l^{(n)})_{l=0}^t$, and then show that

$$\lim_{n \to \infty} \mathbb{P}\left((X_0^{(n)}, X_1^{(n)}, X_2^{(n)}, \ldots, X_t^{(n)}) \neq (Y_0^{(n)}, Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_t^{(n)})\right) = 0.$$

For all $l > t \wedge \tau$, let $Y_l^{(n)} = X_l^{(n)}$. Moreover, let $Y_0^{(n)} = X_0^{(n)}$. For all $l \in [t \wedge \tau]$, let the first component of $Y_l^{(n)}$ to be equal to the first component of $X_l^{(n)}$. Conditioned on $\mathcal{F}_{l-1}$, construct the set $\mathcal{S}_l^{(n)}$ as follows,

- For each vertex j such that the vertex j has not been explored and the value of $C_n(\{v_l, j\})$ has not been realized by time step $l-1$, include $C_n(\{v_l, j\})$ in $\mathcal{S}_l^{(n)}$.
- For each vertex j such that the vertex j has been explored, but the value of $C_n(\{v_l, j\})$

has not been realized by time step $l-1$, include $C_n(\{v_l, j\}) - \widehat{\mathcal{T}}_j$ in $\mathcal{S}_l^{(n)}$, where $\widehat{\mathcal{T}}_j$ is defined before Remark B.11.

- For each vertex j such that the value of $C_n(\{v_l, j\})$ has been realized by time step $l-1$ and $\{v_l, j\} \neq e_l$, add an exponentially distributed random variable with mean n to $\mathcal{S}_l^{(n)}$.

Now define the second component of $Y_l^{(n)}$ be the $Y_l^{(n)}(1)^{th}$ smallest value in $\mathcal{S}_l^{(n)}$ and let the remaining $Y_l^{(n)}(1) - 1$ components of $Y_l^{(n)}$ to be the $Y_l^{(n)}(1) - 1$ smallest values in $\mathcal{S}_l^{(n)}$ (randomly ordered). Clearly conditioned on $\bigcap_{l \in [t \wedge \tau]} \{J_l = \emptyset\}$, $(Y_l^{(n)})_{l=0}^t$ and $(\widetilde{X}_l^{(n)})_{l=0}^t$ are equidistributed.

The event $X_l^{(n)} \neq Y_l^{(n)}$ for some $l \in [t \wedge \tau]$ may happen if 1)the vertex v_l has been touched twice during the exploration process up to time step $l-1$ or if 2)the value of $C_n(\{v_l, j\}) - \widehat{\mathcal{T}}_j$ for an explored vertex j is smaller than $\overline{\mathcal{T}}_l$. Recall that in the proof of the Lemma B.21, we replaced the set H_2 with the set $\widehat{H}_2$ where each value in $\widehat{H}_2$ corresponds to $C_n(\{v_l, j\}) - \widehat{\mathcal{T}}_j$ for an explored vertex j such that $\{v_l, j\} \in \mathcal{U}_{l-1}$. We also proved the following inequality:

$$\mathbb{P}\left(\left\{ \begin{array}{l} \text{The } \widetilde{k}+1 \text{ smallest values in } H_1 \cup \widehat{H}_2 \text{ corre-} \\ \text{sponds to the vertices that are touched for the} \\ \text{first time at time step } l \end{array} \right\} \middle| \mathcal{F}_{l-1} \right) \geq \left(\max\left(0, \frac{n - l - \sum_{i=0}^{l} d_{v_i}(n) - 2}{n}\right)\right)^{d_{v_l}(n)+1}$$

Hence, using the above inequality and the union bound, for all $l \in [t \wedge \tau]$ we have

$$\begin{aligned} &\mathbb{P}(X_l^{(n)} \neq Y_l^{(n)} | \mathcal{F}_{l-1}) \\ &\quad \leq \mathbb{1}\{v_l \text{ has been touched at least twice}\} + \\ &\qquad 1 - \left(\max\left(0, \frac{n - l - \sum_{i=0}^{l} d_{v_i}(n) - 2}{n}\right)\right)^{d_{v_l}(n)+1} \\ &\quad \leq \mathbb{1}\left\{ \bigcup_{i=1}^{l-1} \{|J_i| \neq 0\} \right\} + 1 - \left(\max\left(0, \frac{n - l - \sum_{i=0}^{l} d_{v_i}(n) - 2}{n}\right)\right)^{d_{v_l}(n)+1} \end{aligned}$$

Using Lemma $B.21$ and the same reasoning as in its proof, we get

$$\mathbb{P}((X_0^{(n)}, X_1^{(n)}, \ldots, X_t^{(n)}) \neq (Y_0^{(n)}, Y_1^{(n)}, \ldots, Y_t^{(n)})) \leq \sum_{l=1}^{t} \mathbb{E}\left[\mathbb{1}\{l \leq \tau\} \mathbb{P}\left(\{X_l^{(n)} \neq Y_l^{(n)}\} | \mathcal{F}_{l-1} \right) \right] \xrightarrow{n \to \infty} 0.$$

□

Step 4: Portmanteau Theorem

The final step is to prove the weak convergence of $\mathbb{E}U(N_n)$ to $Er(P)$ by using the Portmanteau theorem. Let $\overline{\rho}_n = \mathbb{E}U(N_n)$ and $\rho = Er(P)$. The goal is to prove $\overline{\rho}_n \xrightarrow{w} \rho$. For a finite rooted tree $[\mathrm{T}_\circ] \in G_*$ of depth R, define the set $A_{\mathrm{T}_\circ}$ as follows,

$$A_{\mathrm{T}_\circ} = \left\{ [N_\circ] \in G_* : d_{G_*}([N_\circ], [\mathrm{T}_\circ]) < \frac{1}{1+R} \right\}$$

Note that if $[N_\circ] \in A_{\mathrm{T}_\circ}$, then the rooted subgraph $(G_\circ)_R$ obtained by removing the marks as well as all the vertices of depth more than R from $N_\circ$ is homeomorphic to the graph structure of $\mathrm{T}_\circ$. Moreover, the first component of the mark of each vertex in $N_\circ$ up to depth R is equal to the one in $\mathrm{T}_\circ$. Recall that this value for each vertex corresponds to the number of potential neighbors while the first component of the type of each vertex equals the number of potential descendants (see Remark III.2).

The first step is to prove that the measure assigned to $A_{\mathrm{T}_\circ}$ by $\overline{\rho}_n$ converges to the measure assigned by ρ. Let $l < \infty$ denote the sum of the first component of the type of the vertices in $\mathrm{T}_\circ$. To see whether the rooted network generated by $\overline{\rho}_n$ is in $A_{\mathrm{T}_\circ}$ or not, we need to look at the first l steps of the exploration process; however, by Lemma B.24 the sequence corresponds to the first l steps of the exploration process converges to the one generated by ρ in distribution. Therefore we have

$$\begin{aligned} &|\overline{\rho}_n(A_{\mathrm{T}_\circ}) - \rho(A_{\mathrm{T}_\circ})| = \\ &\quad \left| \mathbb{P}\left((X_0^{(n)}, X_1^{(n)}, X_2^{(n)}, \ldots, X_l^{(n)}) \in \mathcal{K} \right) - \mathbb{P}\left((X_0, X_1, X_2, \ldots, X_l) \in \mathcal{K} \right) \right| \xrightarrow{n\to\infty} 0, \end{aligned}$$

where $\mathcal{K}$ is defined such that $(X_0^{(n)}, X_1^{(n)}, X_2^{(n)}, \ldots, X_l^{(n)}) \in \mathcal{K}$ if and only if the rooted network induced by $\mathcal{C}_{l \wedge \tau}$ belongs to the set $A_{\mathrm{T}_\circ}$.

The second step is to prove that for any bounded uniformly continuous function f,

$$\left| \int f d\overline{\rho}_n - \int f d\rho \right| \xrightarrow{n\to\infty} 0.$$

Fix the value of $\varepsilon > 0$. Since f is continuous, there exists a $\delta > 0$ such that for every $N_\circ$ and $N'_\circ$ in G_*, $d_{G^*}(N_\circ, N'_\circ) < \delta$ implies $|f(N_\circ) - f(N'_\circ)| < \varepsilon$. Let t to be large enough such that $(t+1)^{-1} < \delta$.

Note that the space G_* is separable; hence the restriction of G_* to the rooted trees is also separable. Moreover, ρ assigns zero measure to the set of rooted networks in G_* that

are not rooted trees. Hence, there exists a finite set $\mathcal{S}$ of rooted trees of depth less than or equal to t in G_* such that,

$$\sum_{\mathrm{T}_\circ \in \mathcal{S}} \rho(A_{\mathrm{T}_\circ}) > 1 - \varepsilon$$

Moreover, since $\overline{\rho}_n(A_{\mathrm{T}_\circ})$ converges to $\rho(A_{\mathrm{T}_\circ})$, for large enough n we have $\sum_{\mathrm{T}_\circ \in \mathcal{S}} \overline{\rho}_n(A_{\mathrm{T}_\circ}) > 1 - 2\varepsilon$. Using all of these points, we have

$$\left| \int f d\overline{\rho}_n - \int f d\rho \right| \leq 3\varepsilon \|f\|_\infty + \sum_{\mathrm{T}_\circ \in \mathcal{S}} f(\mathrm{T}_\circ) \left| \overline{\rho}_n(A_{\mathrm{T}_\circ}) - \rho(A_{\mathrm{T}_\circ}) \right| + 2\varepsilon$$

Finally, let n go to infinity and then ε to zero, and the apply the Portmanteau Theorem to complete the proof.

B.2 Proof of Theorem III.2

Conditioned on $n_{\boldsymbol{i}} = m$ and $v_{\boldsymbol{i}} = x$, the probability of the event $\left\{\zeta_{(\boldsymbol{i},j)} < v_{(\boldsymbol{i},j)}\right\}$ is given as follows,

$$\begin{aligned}
\mathbb{P}\left(\left\{\zeta_{(\boldsymbol{i},j)} < v_{(\boldsymbol{i},j)}\right\} | n_{\boldsymbol{i}} = m, v_{\boldsymbol{i}} = x\right) &= \int_{y=0}^{x} \frac{1}{x} \left(\sum_{k=1}^{\infty} \widehat{P}(k-1) \int_{y}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} \, dz \right) dy \\
&= \int_{y=0}^{x} \frac{1}{x} \sum_{k=1}^{\infty} P(k) \bar{F}_k(y) \, dy.
\end{aligned}$$

The symmetric and conditionally independent structure of EWT implies that the random variable $D_{\boldsymbol{i}}$ conditioned on $n_{\boldsymbol{i}} = m$ and $v_{\boldsymbol{i}} = x$ has the binomial distribution. Hence,

$$\begin{aligned}
\mathbb{P}\left(D_{\boldsymbol{i}} = d | n_{\boldsymbol{i}} = m, v_{\boldsymbol{i}} = x\right) &= \mathbb{P}\left(\sum_{j=1}^{n_{\boldsymbol{i}}} \mathbb{1}\left\{\zeta_{(\boldsymbol{i},j)} < v_{(\boldsymbol{i},j)}\right\} = d \middle| n_{\boldsymbol{i}} = m, v_{\boldsymbol{i}} = x \right) \\
&= Bi\left(d; m, \int_0^x \frac{1}{x} \sum_{k=1}^{\infty} P(k) \bar{F}_k(y) \, dy \right).
\end{aligned}$$

The degree distribution of the root follows immediately by integrating/summing over all possible values of $v_\emptyset$ and $n_\emptyset$. The mean of $D_\emptyset$ is obtained as follows:

$$\mathbb{E}[D_\emptyset] = \sum_{d=1}^{\infty} d \times \mathbb{P}(D_\emptyset = d)$$

$$
\begin{aligned}
&= \sum_{m=1}^{\infty} P(m) \int_0^{\infty} \frac{\mathrm{e}^{-x} x^m}{m!} \sum_{d=1}^{m} d \times Bi\left(d; m, \int_0^x \frac{1}{x} \sum_{k=1}^{\infty} P(k) \bar{F}_k(y)\, dy\right) dx \\
&= \sum_{m=1}^{\infty} P(m) \int_0^{\infty} \frac{\mathrm{e}^{-x} x^m}{m!} \times m \int_0^x \frac{1}{x} \sum_{k=1}^{\infty} P(k) \bar{F}_k(y)\, dy\, dx \\
&= \sum_{m=1}^{\infty} P(m) \sum_{k=1}^{\infty} P(k) \int_0^{\infty} \bar{F}_k(y) \int_y^{\infty} \frac{\mathrm{e}^{-x} x^{m-1}}{(m-1)!}\, dx\, dy \\
&= \sum_{m=1}^{\infty} \sum_{k=1}^{\infty} P(m) P(k) \int_0^{\infty} \bar{F}_k(y) \bar{F}_m(y)\, dy.
\end{aligned}
$$

B.3 Proof of Theorem III.8

Let $W_{l,\boldsymbol{i}}$ denote the number of potential vertices at depth l on the backbone tree, all of whose paths to the root vertex pass through the potential vertex $\boldsymbol{i} \in \mathbb{N}^f$. In the following, we write $\boldsymbol{i} = (i_1, i_2, \cdots, i_k)$ where $k \geq 0$. We have

$$
\begin{aligned}
\mathbb{E}[W_l] &= \sum_{m=1}^{\infty} P(m) \mathbb{E}[W_l | n_\emptyset = m] \\
&= \sum_{m=1}^{\infty} P(m) \mathbb{E}\left[\sum_{j=1}^{m} W_{l,(j)} \middle| n_\emptyset = m\right] \\
&= \sum_{m=1}^{\infty} m P(m) \mathbb{E}\left[W_{l,(1)}\right] \\
&= \sum_{m=1}^{\infty} m P(m) \sum_{k_1=0}^{\infty} \widehat{P}(k_1) \mathbb{E}\left[W_{l,(1)} \middle| n_1 = k_1\right] \\
&= \sum_{m=1}^{\infty} m P(m) \sum_{k_1=0}^{\infty} P(k_1+1) \mathbb{E}\left[\sum_{j=1}^{k_1} W_{l,(1,j)} \middle| n_1 = k_1\right] \\
&= \sum_{m=1}^{\infty} m P(m) \sum_{k_1=1}^{\infty} k_1 P(k_1+1) \mathbb{E}\left[W_{l,(1,1)}\right] \\
&\vdots \\
&= \sum_{m=1}^{\infty} m P(m) \sum_{k_1=2}^{\infty} (k_1-1) P(k_1) \cdots \sum_{k_{l-1}=2}^{\infty} (k_{l-1}-1) P(k_{l-1}) \times 1 \\
&= \mathbb{E}[n_\emptyset] \times \left(\mathbb{E}[(n_\emptyset - 1)]\right)^{l-1}.
\end{aligned}
$$

For the expected number of vertices at depth l, rewrite Z_l as the sum of indicator functions of survival over the potential vertices at depth l. A vertex at depth l survives if and only if all the potential edges on its path to the root survive. Writing $\boldsymbol{t}^j = (t_1, t_2, \ldots, t_j)$ and $\boldsymbol{t}^0 = \emptyset$ by convention, we then have

$$\begin{aligned}
\mathbb{E}[Z_l] &= \mathbb{E}\left[\sum_{\substack{(t_1,t_2,\ldots,t_l) \\ s.t.\ t_j \in [n_{\boldsymbol{t}^{j-1}}]}} \mathbb{1}\left(\bigcap_{j=1}^{l} \{\zeta_{\boldsymbol{t}^j} < v_{\boldsymbol{t}^j}\}\right)\right] \\
&= \sum_{m=1}^{\infty} P(m) \times \mathbb{E}\left[\sum_{\substack{(t_1,t_2,\ldots,t_l) \\ s.t.\ t_j \in [n_{\boldsymbol{t}^{j-1}}]}} \mathbb{1}\left(\bigcap_{j=1}^{l} \{\zeta_{\boldsymbol{t}^j} < v_{\boldsymbol{t}^j}\}\right) \middle| n_\emptyset = m\right]
\end{aligned}$$

Using the symmetric structure of the EWT, we have

$$\begin{aligned}
\mathbb{E}[Z_l] &= \sum_{m=1}^{\infty} mP(m) \times \mathbb{E}\left[\sum_{\substack{(t_1=1,t_2,\ldots,t_l) \\ s.t.\ t_j \in [n_{\boldsymbol{t}^{j-1}}]}} \mathbb{1}\left(\bigcap_{j=1}^{l} \{\zeta_{\boldsymbol{t}^j} < v_{\boldsymbol{t}^j}\}\right) \middle| n_\emptyset = m\right] \\
&\vdots \\
&= \sum_{m=1}^{\infty} mP(m) \sum_{k_1=2}^{\infty} (k_1 - 1)P(k_1) \cdots \sum_{k_{l-1}=2}^{\infty} (k_{l-1} - 1)P(k_{l-1}) \sum_{k_l=1}^{\infty} P(k_l) \times \\
&\quad \mathbb{E}\left[\mathbb{1}\left(\bigcap_{j=1}^{l} \{\zeta_{\mathbf{1}^j} < v_{\mathbf{1}^j}\}\right) \middle| n_\emptyset = m, \bigcap_{j=1}^{l} \{n_{\mathbf{1}^j} = k_j - 1\}\right] \\
&= \sum_{m=1}^{\infty} mP(m) \sum_{k_1=2}^{\infty} (k_1 - 1)P(k_1) \cdots \sum_{k_{l-1}=2}^{\infty} (k_{l-1} - 1)P(k_{l-1}) \sum_{k_l=1}^{\infty} P(k_l) \times \\
&\quad \int_{x=0}^{\infty} f_{m+1}(x) \int_{y_1=0}^{x} \frac{1}{x} \int_{z_1=y_1}^{\infty} f_{k_1}(z_1) \int_{y_2=0}^{z_1} \frac{1}{z_1} \int_{z_2=y_2}^{\infty} f_{k_2}(z_2) \int_{y_3=0}^{z_2} \frac{1}{z_2} \\
&\quad \cdots \int_{y_l=0}^{z_{l-1}} \frac{1}{z_{l-1}} \int_{z_l=y_l}^{\infty} f_{k_l}(z_l)\, dz_l\, dy_l \ldots\, dz_1\, dy_1\, dx
\end{aligned}$$

where $f_l(\cdot)$ is the probability density function of $Erlang(l)$ and $\mathbf{1}^j \in \mathbb{N}^f$ is a sequence of all 1 of length j. Using the equality $f_k(x) \times (k-1)/x = f_{k-1}(x)$, interchanging order of integration in pairs, e.g., z_l and y_{l-1}, and using the complementary cumulative distribution

functions to simplify the integrals involving the z's, we have,

$$\mathbb{E}[Z_l] = \sum_{m=1}^{\infty} P(m) \sum_{k_1=2}^{\infty} P(k_1) \cdots \sum_{k_{l-1}=2}^{\infty} P(k_{l-1}) \sum_{k_l=1}^{\infty} P(k_l)$$
$$\int_{y_l=0}^{\infty} \int_{y_{l-1}=0}^{\infty} \cdots \int_{y_1=0}^{\infty} \bar{F}_m(y_1) \bar{F}_{k_1-1}(\max(y_1, y_2)) \ldots$$
$$\bar{F}_{k_{l-1}-1}(\max(y_{l-1}, y_l)) \bar{F}_{k_l}(y_l)\, dy_1\, dy_2 \ldots\, dy_l$$

B.4 Proof of Proposition III.1 part (i) and part (ii)

(i) Let $\{X_i\}_{i=1}^{\infty}$ denote a set of independent and exponentially distributed random variables with mean 1. Let $N \sim geo(p)$ be independent of $\{X_i\}_{i=1}^{n}$. Recall that $\bar{F}_k(\cdot)$ is the complementary cumulative distribution function of $Erlang(\cdot\,; k, 1)$. It is easy to see that

$$\mathbb{P}\left(\sum_{i=1}^{N} X_i > y\right) = \sum_{k=1}^{\infty} P(k) \bar{F}_k(y), \tag{B.7}$$

since $Erlang(\cdot\,; k, 1)$ is the distribution of a sum of k independent exponential variables with mean 1. On the other hand,

$$\mathbb{E}\left[\mathrm{e}^{t\sum_{i=1}^{N} X_i}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathrm{e}^{t\sum_{i=1}^{N} X_i} | N\right]\right] = \mathbb{E}\left[\left(\frac{1}{1-t}\right)^N\right] = \frac{p}{p-t}$$

which is the moment generating function of an exponentially distributed random variable with rate parameter p. Hence,

$$\sum_{k=1}^{\infty} P(k) \bar{F}_k(y) = \mathbb{P}\left(\sum_{i=1}^{N} X_i > y\right) = \mathrm{e}^{-py}$$

We treat the case $d \geq 1$ and $d = 0$ separately. Assume $d \geq 1$. Using Theorem III.2, we have

$$\mathbb{P}(D_\emptyset = d) = \sum_{m=1}^{\infty} P(m) \int_0^{\infty} \frac{\mathrm{e}^{-x} x^m}{m!} Bi\left(d; m, \int_0^x \frac{1}{x} \sum_{k=1}^{\infty} P(k) \bar{F}_k(y)\, dy\right) dx$$
$$= \sum_{m=1}^{\infty} p(1-p)^{m-1} \int_0^{\infty} \frac{\mathrm{e}^{-x} x^m}{m!} Bi\left(d; m, \frac{1-\mathrm{e}^{-px}}{px}\right) dx$$

$$
\begin{aligned}
&= \int_0^\infty \sum_{m=d}^{\infty} p(1-p)^{m-1} \frac{\mathrm{e}^{-x}}{d!\,(m-d)!} \frac{(1-\mathrm{e}^{-px})^d\ (px-1+\mathrm{e}^{-px})^{m-d}}{p^m}\,dx \\
&= \int_0^\infty \frac{\mathrm{e}^{-x}}{d!} \left(\frac{1-p}{p}\right)^{d-1} \left(1-\mathrm{e}^{-px}\right)^d \sum_{m=d}^{\infty} \left(\frac{1-p}{p}\right)^{m-d} \frac{(px-1+\mathrm{e}^{-px})^{m-d}}{(m-d)!}\,dx \\
&= \int_0^\infty \frac{\mathrm{e}^{-px}}{d!} \left(\frac{1-p}{p}\right)^{d-1} \left(1-\mathrm{e}^{-px}\right)^d \exp\left(-\frac{1-p}{p}\left(1-\mathrm{e}^{-px}\right)\right)\,dx \\
&= \frac{p}{(1-p)^2} \int_0^1 \frac{\left(\frac{1-p}{p}\right)^{d+1} z^d \exp\left(-\frac{1-p}{p}z\right)}{d!}\,dz \\
&= \frac{p}{(1-p)^2} \left(1 - \sum_{m=0}^{d} \frac{\left(\frac{1-p}{p}\right)^m \mathrm{e}^{-\frac{1-p}{p}}}{m!}\right),
\end{aligned}
$$

where the penultimate equality follows by a change of variable, and the last equality follows by the fact that the integrand is the probability density function of Erlang distribution with parameters $d+1 \in \mathbb{N}$ and $\frac{1-p}{p} > 0$. Note that the third equality does not hold for the case $d = 0$.

Next, consider the case $d = 0$. Using Theorem III.2 and similar to above, we have

$$
\begin{aligned}
\mathbb{P}(D_\emptyset = 0) &= \sum_{m=1}^{\infty} P(m) \int_0^\infty \frac{\mathrm{e}^{-x} x^m}{m!} Bi\left(d; m, \int_0^x \frac{1}{x} \sum_{k=1}^{\infty} P(k) \bar{F}_k(y)\,dy\right) dx \\
&= \int_0^\infty \sum_{m=1}^{\infty} p(1-p)^{m-1} \frac{\mathrm{e}^{-x}\ (px-1+\mathrm{e}^{-px})^m}{p^m\, m!}\,dx \\
&= \frac{p}{1-p} \int_0^\infty \mathrm{e}^{-x} \left(\exp\left(\frac{1-p}{p}\left(px-1+\mathrm{e}^{-px}\right)\right) - 1\right) dx \\
&= \frac{p}{(1-p)^2}\left(1-\mathrm{e}^{-\frac{1-p}{p}}\right) - \frac{p}{1-p}.
\end{aligned}
$$

(ii) Let us consider the case $x > 0$. We have,

$$
\begin{aligned}
T(f)(x) &= \frac{1}{x} \sum_{k=1}^{\infty} P(k) \int_{y=0}^{x} \left(\int_{z=0}^{y} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!}\,dz + \int_{z=y}^{\infty} \frac{\mathrm{e}^{-z} z^{k-1}}{(k-1)!} f(z)^{k-1}\,dz\right) dy \\
&= \frac{p}{x} \int_{y=0}^{x} \left(\int_{z=0}^{y} \mathrm{e}^{-z} \sum_{k=1}^{\infty} \frac{(1-p)^{k-1} z^{k-1}}{(k-1)!}\,dz + \right. \\
&\qquad \left. \int_{z=y}^{\infty} \mathrm{e}^{-z} \sum_{k=1}^{\infty} \frac{(1-p)^{k-1} z^{k-1}}{(k-1)!} f(z)^{k-1}\,dz \right) dy
\end{aligned}
$$

$$= \frac{px - 1 + \mathrm{e}^{-px}}{px} + \frac{p}{x} \int_{z=0}^{\infty} \min(x, z) \exp\left(-z\left(1 - (1-p)f(z)\right)\right)\, dz$$

The derivation for $x = 0$ is similar and is omitted.

APPENDIX C

Appendix of Chapter IV

C.1 Proof of Lemma IV.2

The moment generating function for an exponential random variable Y with rate λ is

$$\mathbb{E}[e^{\mu Y}] = \frac{\lambda}{\lambda - \mu}.$$

Since X_1 and X_2 are independent, the exponential generating function for $X_1 - X_2$ is

$$\mathbb{E}[e^{\mu(X_1 - X_2)}] = \mathbb{E}[e^{\mu X_1}]\,\mathbb{E}[e^{-\mu X_2)}] = \left(\frac{\lambda_1\lambda_2}{(\lambda_1 - \mu)(\lambda_2 + \mu)}\right)^t, \tag{C.1}$$

By Markov's inequality

$$\mathbb{P}[X_1 > X_2] = \mathbb{P}[e^{\mu(X_1 - X_2)} > 1] \leq \mathbb{E}[e^{\mu(X_1 - X_2)}]$$

for any $\mu > 0$. The right-hand side of (C.1) is minimized when $\mu = (\lambda_1 - \lambda_2)/2$, giving the desired result.

C.2 Analysis of System of ODEs

In this section, we state and prove Theorem C.1.

Theorem C.1. *When $\lambda < 4$, the system of ODEs* (4.11)–(4.14) *has a unique solution* (F, G, V, W) *satisfying conditions* (4.15)–(4.16).

When $V(x) \neq 0$, recall that

$$U(x) = \frac{F(x)}{V(x)};$$

Thus, when $V(x) \neq 0$ and $W(x) \neq 0$, the conservation law $FW + GV - VW = 0$ is equivalent to

$$\frac{G(x)}{W(x)} = 1 - U(x).$$

Also, recall that from the conservation law we have $V(0) = W(0) = 2F(0) = 2G(0)$. Hence, the previous 4-dimensional system of ODEs (4.11)–(4.14) with conditions (4.15)–(4.16) reduces to the following 3-dimensional system of ODEs:

$$\begin{aligned} \frac{\mathrm{d}U}{\mathrm{d}x} &= -\lambda U(1-U) + (1-UV)\,(1-(1-U)W) \\ \frac{\mathrm{d}V}{\mathrm{d}x} &= \lambda V(1-U) \\ \frac{\mathrm{d}W}{\mathrm{d}x} &= -\lambda WU \end{aligned} \tag{C.2}$$

with initial condition

$$U(0) = \frac{1}{2}, \quad V(0) = W(0) = \epsilon, \quad \epsilon \in [0,1]. \tag{C.3}$$

Note that the partial derivatives of the right hand side of (C.2) with respect to (U, V, W) are continuous. Therefore, by the standard existence and uniqueness theorem for solutions of systems of ODEs (see e.g. [107, Theorem 2]) it follows that the system (C.2) with the initial condition (C.3) has a unique solution for a fixed $\epsilon \in [0,1]$. We write this unique solution as $U(x,\epsilon)$, $V(x,\epsilon)$, and $W(x,\epsilon)$, which we abbreviate as (U, V, W) whenever the context is clear.

Therefore, to prove Theorem C.1, it suffices to show that the system of ODEs (C.2) with the initial condition (C.3) has a solution $(U(x,\epsilon_0), V(x,\epsilon_0), W(x,\epsilon_0))$ for $x \in [0, +\infty)$ satisfying the boundary condition $U(+\infty,\epsilon_0) = V(+\infty,\epsilon_0) = 1$ and $W(+\infty,\epsilon_0) = 0$ for a unique $\epsilon_0 \in [0,1]$. Geometrically speaking, this is due to the fact that $(U = 1, V = 1, W = 0)$ is a *saddle point*, and there is a unique choice of ϵ_0 so that the trajectory $(U(x,\epsilon_0), V(x,\epsilon_0), W(x,\epsilon_0))$ falls into the stable manifold i.e., set of initial conditions $(U(0,\epsilon_0), V(0,\epsilon_0), W(0,\epsilon_0))$ such that $(U(x,\epsilon_0), V(x,\epsilon_0), W(x,\epsilon_0)) \to (1,1,0)$ as $x \to +\infty$. For any other choice of $\epsilon \neq \epsilon_0$, the trajectory $(U(x,\epsilon), V(x,\epsilon), W(x,\epsilon))$ veers away from $(1,1,0)$ to infinity.

The outline of the proof is as follows. We first prove some basic properties satisfied by the solution (U, V, W) in Appendix C.2.1. Then based on these properties, in Appendix C.2.2 we prove that the solution satisfies some monotonicity properties with respect to ϵ by

studying the sensitivity of the solution to the initial condition. Next, in Appendix C.2.3 we characterize the limiting behavior of the solution depending on whether it hits 1 or not. The monotonicity properties and the limiting behavior enable us to completely characterize the basins of attraction in Appendix C.2.4. In particular, we show that the basin of attraction for $(U = 1, V = 1, W = 0)$ is a singleton, i.e., there is a unique choice of $\epsilon_0 \in [0, 1]$ such that $(U(x, \epsilon_0), V(x, \epsilon_0), W(x, \epsilon_0) \to (1, 1, 0)$ as $x \to +\infty$. Finally, we connect the 3-dimensional system of ODEs (C.2) back to the 4-dimensional system of ODEs (4.11)–(4.14) and finish the proof of Theorem C.1 in Appendix C.2.5.

C.2.1 Basic Properties of the Solution

In the following two lemmas, we prove some basic properties of the solution.

Lemma C.2. *Fix any $\epsilon \in [0, 1]$. Then for any $x \in [0, +\infty)$ such that the unique solution (U, V, W) is well-defined (not equal to $\pm\infty$), it holds that*

$$V(x) = W(x)\,\mathrm{e}^{\lambda x}, \quad UV < 1, \quad (1 - U)W < 1, \quad U > 0.$$

Proof. It follows from (C.2) that

$$V(x) = \epsilon \exp\left(\lambda \int_0^x (1 - U(y))\,\mathrm{d}y\right), \tag{C.4}$$

$$W(x) = \epsilon \exp\left(-\lambda \int_0^x U(y)\mathrm{d}y\right). \tag{C.5}$$

Hence $V(x) = W(x)\,\mathrm{e}^{\lambda x}$. Thus the conservation law $FW + GV - VW = 0$ implies that

$$V = F + G\mathrm{e}^{\lambda x}, \qquad W = F\mathrm{e}^{-\lambda x} + G.$$

Recall that $F = UV$ and $G = (1 - U)W$. Then

$$\begin{aligned} \frac{\mathrm{d}F}{\mathrm{d}x} &= (1 - F)\,(1 - G)\,V = (1 - F)\,(1 - G)\left(F + G\mathrm{e}^{\lambda x}\right) \\ \frac{\mathrm{d}G}{\mathrm{d}x} &= -\,(1 - F)\,(1 - G)\,W = -\,(1 - F)\,(1 - G)\left(F\mathrm{e}^{-\lambda x} + G\right). \end{aligned} \tag{C.6}$$

For the sake of contradiction, suppose $\max\{F(x), G(x)\} \geq 1$ for some finite $x > 0$. Since $F(x)$ and $G(x)$ are continuous in x and $F(0) = G(0) < 1$, there is an $x_0 > 0$ such that $\max\{F(x_0), G(x_0)\} = 1$. Define $\widetilde{F}(x) \equiv F(x_0)$ and $\widetilde{G}(x) \equiv G(x_0)$. Then $(\widetilde{F}(x), \widetilde{G}(x))$ is a solution to ODE (C.6) in $x \in [0, x_0]$ running backward with its initial value at $x = x_0$ given by $(F(x_0), G(x_0))$. Note that $(F(x), G(x))$ is also a solution to ODE (C.6) in the

backward time $x \in [0, x_0]$ with its initial value at $x = x_0$ given by $(F(x_0), G(x_0))$. Also note that the right hand side of ODE (C.6) is continuous in x and the partial derivatives with respect to F and G are continuous. By existence and uniqueness [107, Theorem 2] it follows that $F(x) \equiv F(x_0)$ and $G(x) \equiv G(x_0)$ for $x \in [0, x_0]$. Hence, $\max\{F(0), G(0)\} = 1$, which contradicts that $F(0) = G(0) = \epsilon/2 < 1$. Thus, $F = UV < 1$ and $G = (1-U)W < 1$.

Next, we argue that $U > 0$. Suppose not, since $U(0) = 1/2$, by the differentiability of U in x, there exists a finite $x_0 > 0$ such that $U(x_0) = 0$ and $U'(x_0) \leq 0$. However,

$$\left.\frac{\mathrm{d}U}{\mathrm{d}x}\right|_{x=x_0} = [1 - U(x_0)V(x_0)]\,[1 - (1 - U(x_0))W(x_0)] = [1 - F(x_0)]\,[1 - G(x_0)] > 0,$$

which leads to a contradiction. □

The following lemma shows that V and W are positive when $U(x, \epsilon) < +\infty$.

Lemma C.3. *Fix $\epsilon \in (0, 1)$ and $\lambda < 4$. Suppose $U(x, \epsilon) < +\infty$ for $x \in [0, c]$ for a finite constant $c > 0$. Then for all $x \in [0, c]$, $V(x, \epsilon) > 0$ and $W(x, \epsilon) > 0$.*

Proof. Suppose not. Since $V(0) = \epsilon > 0$ and $V(x) = W(x)\,\mathrm{e}^{\lambda x}$, by the continuity of V in x, there exists a finite $x_0 \in (0, c]$ such that $V(x_0) = 0$ and $W(x_0) = 0$. Note that by assumption $U(x_0) < +\infty$.

Consider the following ODE backward in time $x \in [0, x_0]$ with the initial value $U(x_0)$:

$$\frac{\mathrm{d}U}{\mathrm{d}x} = -\lambda U(1 - U) + 1. \tag{C.7}$$

Since the derivative of the right hand side of the above ODE with respect to U is continuous, by existence and uniqueness [107, Theorem 2] there exists a unique solution of the above ODE with the initial value $U(x_0)$, denoted by $\widetilde{U}(x)$. Since $\lambda < 4$, it follows that $dU/dx > 0$ and thus $\widetilde{U}(x)$ is monotonically increasing. As a consequence, either $\widetilde{U}(x)$ is well defined over the entire interval $[0, x_0]$ or there exists $x_1 \in (0, x_0)$ such that $\widetilde{U}(x)$ is well defined over $(x_1, x_0]$ and $\widetilde{U}(x) \to -\infty$ as x approaches x_1 from the above. Let $\mathcal{I}$ denote the interval where $\widetilde{U}(x)$ is well defined.

Let $\widetilde{V}(x) \equiv V(x_0)$ and $\widetilde{W}(x) \equiv W(x_0)$. Then $(\widetilde{U}(x), \widetilde{V}(x), \widetilde{W}(x))$ is a solution to ODE (C.2) with initial values $(U(x_0), V(x_0), W(x_0))$ in the backward time $x \in \mathcal{I}$. Note that $(U(x), V(x), W(x))$ is also a solution to ODE (C.2) with the same initial values $(U(x_0), V(x_0), W(x_0))$ in the backward time $x \in [0, x_0]$. Also note that the partial derivatives of the right hand side of ODE (C.2) with respect to (U, V, W) are continuous. By the existence and uniqueness [107, Theorem 2], it follows that $V(x) \equiv V(x_0) = 0$ and $U(x) = \widetilde{U}(x)$ for $x \in \mathcal{I}$. When $\mathcal{I} = [0, x_0]$, then we get $V(0) = 0$, contradicting $V(0) = \epsilon > 0$. When $\mathcal{I} = (x_1, x_0]$, we

get $U(x) \to -\infty$ as x approaches x_1 from the above, contradicting $U(x) > 0$ for $x \in [0, +\infty)$ in view of Lemma C.2. □

C.2.2 Monotonicity to the Initial Condition

The key to our proof is to study how the solution of the system of ODEs (C.2) changes with respect to the initial condition (C.3).

Standard ODE theory (see [107, Theorem 15]) shows that $U(x, \epsilon)$ is differentiable in ϵ and the mixed partial derivatives satisfy

$$\frac{\partial^2 U(x,\epsilon)}{\partial x \partial \epsilon} = \frac{\partial^2 U(x,\epsilon)}{\partial \epsilon \partial x};$$

similarly for V and W. Moreover, the partial derivatives $(\partial U/\partial\epsilon, \partial V/\partial\epsilon, \partial W/\partial\epsilon)$ satisfy the following system of equations:

$$\begin{aligned}
\frac{\partial}{\partial x}\frac{\partial U}{\partial \epsilon} &= [-\lambda(1-2U) - V\,(1-(1-U)W) + (1-UV)W]\,\frac{\partial U}{\partial \epsilon} \\
&\qquad - U\,(1-(1-U)W)\,\frac{\partial V}{\partial \epsilon} - (1-UV)(1-U)\frac{\partial W}{\partial \epsilon} \\
\frac{\partial}{\partial x}\frac{\partial V}{\partial \epsilon} &= -\lambda V \frac{\partial U}{\partial \epsilon} + \lambda(1-U)\frac{\partial V}{\partial \epsilon} \\
\frac{\partial}{\partial x}\frac{\partial W}{\partial \epsilon} &= -\lambda W \frac{\partial U}{\partial \epsilon} - \lambda U \frac{\partial W}{\partial \epsilon},
\end{aligned} \tag{C.8}$$

with initial condition

$$\frac{\partial U(0,\epsilon)}{\partial \epsilon} = 0, \quad \frac{\partial V(0,\epsilon)}{\partial \epsilon} = 1, \quad \frac{\partial W(0,\epsilon)}{\partial \epsilon} = 1. \tag{C.9}$$

The system of equations (C.8) is known as the *system of variational equations* and can be derived by differentiating (C.2) with respect to ϵ and interchange ∂x and $\partial\epsilon$. The initial condition (C.9) can be derived by differentiating (C.3) with respect to ϵ.

The following key lemma shows that whenever $U(x, \epsilon) \le 1$, $U(x, \epsilon)$ is decreasing in ϵ, while $V(x, \epsilon)$ and $W(x, \epsilon)$ are increasing in ϵ.

Lemma C.4. *Fix $\epsilon \in (0, 1)$. Suppose $U(x, \epsilon) \le 1$ for $x \in (0, c]$ for a finite constant $c > 0$. Then for all $x \in (0, c]$,*

$$\frac{\partial U(x,\epsilon)}{\partial \epsilon} < 0, \quad \textit{and} \quad \frac{\partial W(x,\epsilon)}{\partial \epsilon} > 0. \tag{C.10}$$

Moreover, it follows that all $x \in (0, c]$,

$$\frac{\partial V(x,\epsilon)}{\partial \epsilon} \geq \exp\left(\lambda \int_0^x (1 - U(t,\epsilon))\, dt\right) = \frac{V(\epsilon)}{\epsilon} \geq 1. \tag{C.11}$$

Proof. We first show that (C.11) holds whenever $\partial U(x,\epsilon)/\partial \epsilon < 0$ for $x \in (0, c]$. Recall that in Lemma C.3 we have shown that $V > 0$. It follows from (C.8) that for all $x \in (0, c]$

$$\frac{\partial}{\partial x}\frac{\partial V}{\partial \epsilon} \geq \lambda(1 - U)\frac{\partial V}{\partial \epsilon}.$$

Thus for all $x \in (0, c]$

$$\frac{\partial V(x,\epsilon)}{\partial \epsilon} \geq \exp\left(\lambda \int_0^x (1 - U(s))\, ds\right) = \frac{V(x,\epsilon)}{\epsilon} \geq 1,$$

where the equality holds due to (C.4).

Next we show (C.10). For the sake of contradiction, suppose not, i.e., there exists a $x_0 \in (0, c]$ such that either $\frac{\partial U(x_0,\epsilon)}{\partial \epsilon} \geq 0$ or $\frac{\partial W(x_0,\epsilon)}{\partial \epsilon} \leq 0$.

Define

$$a = \inf\left\{x \in (0, c] : \frac{\partial U(x,\epsilon)}{\partial \epsilon} \geq 0\right\}$$

and

$$b = \inf\left\{x \in (0, c] : \frac{\partial W(x,\epsilon)}{\partial \epsilon} \leq 0\right\},$$

with the convention that the infimum of an empty set is $+\infty$. Then $\min\{a, b\} \leq x_0 \leq c$.

Case 1: Suppose $a \leq b$. Due to the initial condition (C.9) and the initial condition (C.3), we have that

$$\frac{\partial U(0,\epsilon)}{\partial \epsilon} = 0, \quad \frac{\partial}{\partial x}\frac{\partial U(0,\epsilon)}{\partial \epsilon} = -\left(1 - \frac{\epsilon}{2}\right) < 0.$$

Then we have $a > 0$. Moreover, by the differentiability of $\frac{\partial U(x,\epsilon)}{\partial \epsilon}$ in x and the definition of a, we have

$$\frac{\partial U(x,\epsilon)}{\partial \epsilon} < 0, \quad \forall x \in (0, a), \quad \frac{\partial U(a,\epsilon)}{\partial \epsilon} = 0, \quad \text{and} \quad \frac{\partial}{\partial x}\frac{\partial U(a,\epsilon)}{\partial \epsilon} \geq 0.$$

It follows from our previous argument for proving (C.11) that $\partial V(x,\epsilon)/\partial \epsilon \geq 1$ for all $x \in (0, a]$. Since $a \leq b$, we also have that

$$\frac{\partial W(a,\epsilon)}{\partial \epsilon} \geq 0.$$

Recall that in Lemma C.3 we have shown that $UV < 1$, $(1-U)W < 1$ and $U > 0$. Moreover, by assumption we have $U \leq 1$. Thus we get from ODE (C.8) that

$$\frac{\partial}{\partial x}\frac{\partial U(a,\epsilon)}{\partial \epsilon} = -U\left(1-(1-U)W\right)\frac{\partial V(a,\epsilon)}{\partial \epsilon} - (1-UV)(1-U)\frac{\partial W(a,\epsilon)}{\partial \epsilon} < 0,$$

which contradicts $\frac{\partial}{\partial x}\frac{\partial U(a,\epsilon)}{\partial \epsilon} \geq 0$.

Case 2: Suppose $a > b$. Due to the initial condition (C.8), we have that

$$\frac{\partial W(0,\epsilon)}{\partial \epsilon} = 1.$$

Thus $b > 0$. By the differentiability of $\frac{\partial W(x,\epsilon)}{\partial \epsilon}$ in x, we have that

$$\frac{\partial W(b,\epsilon)}{\partial \epsilon} = 0, \quad \text{and} \quad \frac{\partial}{\partial x}\frac{\partial W(b,\epsilon)}{\partial \epsilon} \leq 0, \quad \text{and} \quad \frac{\partial U(b,\epsilon)}{\partial \epsilon} < 0.$$

Recall that in Lemma C.3 we have shown that $W > 0$. It follows from ODE (C.8) that

$$\frac{\partial}{\partial x}\frac{\partial W(b,\epsilon)}{\partial \epsilon} = -\lambda W\frac{\partial U(b,\epsilon)}{\partial \epsilon} > 0,$$

which contradicts $\frac{\partial}{\partial x}\frac{\partial W(b,\epsilon)}{\partial \epsilon} \leq 0$. □

Based on Lemma C.4, we prove another "monotonicity" lemma, showing that if $U(x, \epsilon_0) < 1$ for all $x \geq 0$ and some $\epsilon_0 \in (0,1)$, then $U(x,\epsilon) < 1$ for all $x \geq 0$ and all $\epsilon \in (\epsilon_0, 1)$.

Lemma C.5. *Suppose $U(x,\epsilon_0) < 1$ for all $x \geq 0$ and some $\epsilon_0 \in (0,1)$. Then $U(x,\epsilon) < 1$ for all $\epsilon \in (\epsilon_0, 1)$ and all $x \geq 0$.*

Proof. Fix an arbitrary but finite $x_0 > 0$. We claim that $U(x_0,\epsilon) < 1$ for all $\epsilon \in (\epsilon_0, 1)$. Suppose not. Then define

$$\epsilon_1 \triangleq \inf\left\{\epsilon \in (\epsilon_0, 1) : 1 \leq U(x_0,\epsilon) < +\infty\right\}$$

Note that by assumption, $U(x_0,\epsilon_0) < 1$. By the definition of ϵ_1 and the differentiability of $U(x_0,\epsilon)$ in ϵ, we have

$$U(x_0,\epsilon_1) = 1, \quad \frac{\partial U(x_0,\epsilon_1)}{\partial \epsilon} \geq 0.$$

We claim that $U(x,\epsilon_1) < 1$ for all $x \in (0, x_0)$. If not, then there exists an $x_1 \in (0, x_0)$ such that $U(x_1,\epsilon_1) = 1$. Note that $\frac{\mathrm{d}U(x,\epsilon_1)}{\mathrm{d}x} > 0$ if $U(x) \geq 1$. Thus $U(x,\epsilon_1) > 1$ for all $x > x_1$, which contradicts the fact that $U(x_0,\epsilon_1) = 1$. Therefore, we can apply Lemma C.4 with

$c = x_0$ and get that

$$\frac{\partial U(x_0, \epsilon_1)}{\partial \epsilon} < 0,$$

which contradicts the fact that $\frac{\partial U(x_0,\epsilon_1)}{\partial \epsilon} \geq 0$. Since x_0 is arbitrarily chosen, we conclude that $U(x, \epsilon) < 1$ for all $\epsilon \in (\epsilon_0, 1)$ and all $x > 0$. □

C.2.3 Limiting Behavior of (U, V, W)

In this section, we characterize the limiting behavior of (U, V, W), depending on whether U or V hit 1.

First, we state a simple lemma, showing that if both U and V do not hit 1 in finite time, then they converge to 1 as $x \to \infty$.

Lemma C.6. *If $U(x, \epsilon_0) < 1$ and $V(x, \epsilon_0) < 1$ for all $x \geq 0$ and some $\epsilon_0 \in (0, 1)$, then $U(x, \epsilon_0) \to 1$, $V(x, \epsilon_0) \to 1$, and $W(x, \epsilon_0) \to 0$ as $x \to \infty$.*

Proof. By Lemma C.2, we have $W(x) = V(x)\,\mathrm{e}^{-\lambda x} \to 0$ as $x \to \infty$. Recall that according to (C.2),

$$\frac{\mathrm{d}V}{\mathrm{d}x} = \lambda V\,(1 - U) > 0\,.$$

Since $V(x) < 1$ for all $x \geq 0$, it follows that $\frac{\mathrm{d}V}{\mathrm{d}x} \to 0$ and hence $U(x) \to 1$ as $x \to +\infty$. Thus, as $x \to +\infty$,

$$\frac{\mathrm{d}U}{\mathrm{d}x} = -\lambda U(1 - U) + (1 - UV)\,(1 - (1 - U)W) \to 0\,,$$

which implies that $V(x) \to 1$ as $x \to +\infty$. □

The next lemma shows the behavior of U and V if they hit 1 for finite x.

Lemma C.7. *Let $x_0 > 0$ be finite.*

- *If $V(x_0) = 1$, then $V(x)$ monotonically increases to $+\infty$ and $U(x) \to 0$ for $x \geq x_0$.*
- *If $U(x_0) = 1$, then $U(x)$ monotonically increases to $+\infty$ and $V(x) \to 0$ for $x \geq x_0$.*

Proof. Suppose $V(x_0) = 1$. Recall that in Lemma C.2, we have shown that $UV < 1$. According to ODE (C.2), we get that $dV/dx > 0$ if $V \geq 1$ as $UV < 1$. Thus $V(x)$ monotonically increases to $+\infty$ for $x \geq x_0$. Moreover $U(x) \to 0$ for $x \geq x_0$.

Suppose $U(x_0) = 1$. Recall that in Lemma C.2, we have shown that $(1 - U)WV < 1$. According to ODE (C.2), we get that $dU/dx > 0$ if $U \geq 1$. Hence, $U(x)$ monotonically increases to $+\infty$ for $x \geq x_0$. As $UV < 1$, it further follows that $V(x) \to 0$ for $x \geq x_0$. □

C.2.4 Basins of Attraction

In view of Lemma C.6 and Lemma C.7, define the basin of attraction for ($U = 0$, $V = +\infty$) as

$$S_1 = \{\epsilon \in [0,1] : V(x,\epsilon) \geq 1 \text{ for some finite } x > 0 \};$$

the basin of attraction for ($U = +\infty, V = 0$) as

$$S_2 = \{\epsilon \in [0,1] : U(x,\epsilon) \geq 1 \text{ for some finite } x > 0 \};$$

and the basin of attraction for ($U = 1, V = 1$) as

$$S_0 = \{\epsilon \in [0,1] : U(x,\epsilon) < 1 \text{ and } V(x,\epsilon) < 1 \text{ for all finite } x > 0 \}.$$

When ϵ is either 0 or 1, we have the following simple characterizations of the solution.

Lemma C.8. *Suppose $\lambda < 4$.*

- *If $\epsilon = 0$, then $V(x) \equiv 0$, $W(X) \equiv 0$ and $U(x)$ monotonically increases to $+\infty$.*
- *If $\epsilon = 1$, then $V(x)$ monotonically increases to $+\infty$ and $U(x) \to 0$.*

Proof. First, consider the case $\epsilon = 0$. Then according to the system of ODEs (C.2), we immediately get that $V(x) \equiv 0$, $W(X) \equiv 0$. Thus

$$\frac{\mathrm{d}U}{\mathrm{d}x} = -\lambda U(1-U) + 1 > 0,$$

where the last inequality holds due to $\lambda < 4$. Hence, $U(x)$ monotonically increases to $+\infty$.

The conclusion in the case $\epsilon = 1$ simply follows from Lemma C.7. □

Now, we are ready to prove a lemma, which completely characterizes the basins of attraction S_0, S_1, and S_2.

Lemma C.9. *Suppose $\lambda < 4$. Then there exists a unique $\epsilon_0 \in (0,1)$ such that*

$$S_0 = \{\epsilon_0\}, \quad S_1 = (\epsilon_0, 1], \quad S_2 = [0, \epsilon_0). \tag{C.12}$$

Proof. Lemma C.8 implies that $1 \in S_1$ and $0 \in S_2$. Note that $UV < 1$ by Lemma C.2. Thus it follows from Lemma C.7 that S_1 and S_2 are disjoint.

We first prove that S_1 is left open. Fix any $\epsilon \in S_1$. Since $V(x,\epsilon) \geq 1$ for some finite x, it follows from Lemma C.7 that there exists an x_0 such that $V(x_0,\epsilon) > 1$. By the continuity

of $V(x_0, \epsilon)$ in ϵ, there exists a $\delta > 0$ such that for all $\epsilon' \in [\epsilon - \delta, \epsilon]$, $V(x_0, \epsilon') > 1$, and thus $V(x, \epsilon') \to +\infty$ and $U(x, \epsilon') \to 0$ as $x \to +\infty$. Hence, $[\epsilon - \delta, \epsilon] \subset S_1$. Thus S_1 is left open.

Analogously, we can prove that S_2 is right open. Note that $S_0 = [0, 1] \setminus (S_1 \cup S_2)$, and S_1 and S_2 are disjoint. It follows that S_0 is non-empty. Let ϵ_0 be any point in S_0. Next we prove (C.12).

We first fix any $\epsilon \in (\epsilon_0, 1)$. Since $\epsilon_0 \in S_0$, it follows that $U(x, \epsilon_0) < 1$ and $V(x, \epsilon_0) < 1$ for all $x \geq 0$. In view of Lemma C.5, we have that $U(x, \epsilon) < 1$ for all $\epsilon \in (\epsilon_0, 1)$ and all $x > 0$. It follows from Lemma C.4 that $\partial V(x, \epsilon)/\partial \epsilon \geq 1$ for all $x > 0$ and all $\epsilon \in (\epsilon_0, 1)$. Thus for all $x \geq 0$,

$$V(x, \epsilon) = V(x, \epsilon_0) + \int_{\epsilon_0}^{\epsilon} \frac{\partial V(x, \eta)}{\partial \eta} d\eta \geq V(x, \epsilon_0) + (\epsilon - \epsilon_0).$$

Since $V(x, \epsilon_0) \to 1$ as $x \to +\infty$, there exists an x_0 such that for all $x \geq x_0$,

$$V(x, \epsilon_0) \geq 1 - (\epsilon - \epsilon_0)/2.$$

Combining the last two displayed equation gives that for all $x \geq x_0$,

$$V(x, \epsilon) \geq 1 + (\epsilon - \epsilon_0)/2 > 1.$$

We conclude that $\epsilon \in S_1$ and thus $(\epsilon_0, 1] \subset S_1$.

Next we fix any $\epsilon \in (0, \epsilon_0)$ and show that $\epsilon \in S_2$. Suppose not. Then there exists an $\epsilon_1 \in (0, \epsilon_0)$ such that $U(x, \epsilon_1) < 1$ for all $x \geq 0$. By Lemma C.5, we have that $U(x, \epsilon) < 1$ for all $\epsilon \in (\epsilon_1, 1)$ and all $x > 0$. In view of Lemma C.4, it immediately follows that (C.11) holds for all $x > 0$ and all $\epsilon \in (\epsilon_1, 1)$. Thus,

$$V(x, \epsilon_0) = V(x, \epsilon_1) + \int_{\epsilon_1}^{\epsilon_0} \frac{\partial V(x, \epsilon)}{\partial \epsilon} d\epsilon \geq V(x, \epsilon_1) + (\epsilon_0 - \epsilon_1).$$

Note that since $\epsilon_0 \in S_0$, $V(x, \epsilon_0) < 1$ for all $x \geq 0$, it follows that for all $x \geq 0$,

$$V(x, \epsilon_1) < 1 - (\epsilon_0 - \epsilon_1),$$

which contradicts the conclusion of Lemma C.6. Thus we conclude that $\epsilon \in S_2$ and thus $[0, \epsilon_0) \subset S_2$.

Since S_0, S_1, and S_2 are all disjoint, the desired (C.12) readily follows. □

C.2.5 Proof of Theorem C.1

We are now ready to prove Theorem C.1. Let $S_0 = \{\epsilon_0\}$, and let $(U(x,\epsilon_0), V(x,\epsilon_0), W(x,\epsilon_0))$ be the unique solution to the system of ODEs (C.2) with the initial condition (C.3). For $x \in [0,+\infty)$, define

$$\begin{aligned}
&F(x) = U(x,\epsilon_0)V(x,\epsilon_0), \quad F(-x) = (1-U(x,\epsilon_0))\,W(x,\epsilon_0),\\
&V(x) = V(x,\epsilon_0), \quad V(-x) = W(x,\epsilon_0),\\
&G(x) = F(-x), \quad W(x) = V(-x).
\end{aligned}$$

We show that (F,G,V,W) is a solution to the system of ODEs (4.11)–(4.14) with conditions (4.15)–(4.16). First, by construction (F,G,V,W) satisfy the system of ODEs (4.11)–(4.14). Second, since $\epsilon_0 \in S_0$, by definition $U(x,\epsilon_0) < 1$ and $V(x,\epsilon_0) < 1$ for all $x \geq 0$. Thus it follows from Lemma C.6 that as $x \to +\infty$, $U(x,\epsilon_0) \to 1$, $V(x,\epsilon_0) \to 1$, and $W(x,\epsilon_0) \to 0$. Hence, (F,G,V,W) satisfy condition (4.15). Thirdly, in view of Lemma C.2 and Lemma C.3, we have that $U(x,\epsilon_0), V(x,\epsilon_0), W(x,\epsilon_0) > 0$, $W(x,\epsilon_0) < 1$, $U(x,\epsilon_0)V(x,\epsilon_0) < 1$, and $(1-U(x,\epsilon_0))\,W(x,\epsilon_0) < 1$. Therefore, $0 < V, W < 1$ and $0 < F, G < 1$, satisfying condition (4.16).

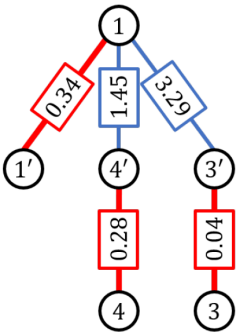
Next, we show that the solution (F,G,V,W) is unique. Let $(\widetilde{F},\widetilde{G},\widetilde{V},\widetilde{W})$ denote another solution to system of ODEs (4.11)–(4.14) with conditions (4.15)–(4.16). Let $\widetilde{U} = \widetilde{F}/\widetilde{V}$. Then $(\widetilde{U},\widetilde{V},\widetilde{W})$ is a solution to the system of ODEs (C.2), satisfying the initial condition (C.3) with $\epsilon = \widetilde{V}(0) = \widetilde{W}(0)$. Moreover, $\widetilde{U}(x) < 1$ and $\widetilde{V}(x) < 1$ for all $x \geq 0$, because otherwise by Lemma C.7, either $\widetilde{U}(x) \to +\infty$ or $\widetilde{V}(x) \to +\infty$, violating that $\widetilde{F}(x), \widetilde{V}(x) \to 1$. As a consequence, $\widetilde{V}(0) \in S_0$. It follows from Lemma C.9 that $\epsilon_0 = \widetilde{V}(0)$. By the uniqueness of the solution to system of ODEs (C.2) with the initial condition (C.3), we have $\widetilde{U}(x) \equiv U(x,\epsilon_0)$, $\widetilde{V}(x) \equiv V(x,\epsilon_0)$, $\widetilde{W}(x) \equiv V(x,\epsilon_0)$. Thus, $(\widetilde{F},\widetilde{G},\widetilde{V},\widetilde{W}) = (F,G,V,W)$.

C.3 Proof of Theorem IV.1

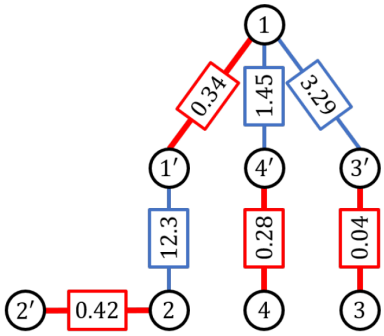
Step 1: The Exploration Process

As we pointed out in the sketch of the proof, the first step is to define an exploration process that explores vertices of $N_{n,\circ}(1)$ in a series of stages. The stage m of the exploration process reveals a rooted subtree of $(K_{n,n}, \ell_n)$, denoted by $N_{n,\circ}[m]$. The root of $N_{n,\circ}[m]$ is vertex 1, the number of un-planted children of every vertex is m, and the set of planted edges restricted to $N_{n,\circ}[m]$ is a matching. Next, we provide a formal construction of $N_{n,\circ}[m]$.

The construction begins with vertex 1. Include the edge $\{1, M_n^*(1) = 1'\}$ in $N_{n,\circ}[m]$. Let $\{v_1, v_2, \cdots, v_m\}$ denote the m closest un-planted neighbors of vertex 1. Add all the edges



Step 1 of the construction of $N_{n,\circ}[2]$.



Step 2 of the construction of $N_{n,\circ}[2]$.

Figure C.1: Stage 2 of the exploration process on $(K_{4,4}, \ell_4)$ given by Figure 4.4.

$\{1, v_k\}$ and then $\{v_k, M_n^*(v_k)\}$ to $N_{n,\circ}[m]$. Next, continue with the vertex $1'$. Let $\{w_1, w_2, \cdots, w_m\}$ denote the m closest un-planted neighbors of vertex $1'$, among all the vertices that has not been added to $N_{n,\circ}[m]$. Include all the edges $\{1', w_k\}$ and then $\{w_k, M_n^*(w_k)\}$ to $N_{n,\circ}[m]$. At each step of the construction, we follow two simple rules: 1) the next vertex to pick is the oldest one in $N_{n,\circ}[m]$; 2) when we add m closest un-planted children of this vertex and their planted pairs, we avoid all the vertices that has already been added to $N_{n,\circ}[m]$.

The construction continues until we are about to pick a vertex at depth m, at which point it stops. Note that the only vertices at depth $m+1$ are the planted partners of the vertices at depth m. Let $V_n^{(m)}$ and $E_n^{(m)}$ denote the set of the vertices and the edges of $N_{n,\circ}[m]$ respectively. Note that for all sufficiently large n, $|V_n^{(m)}|$ and $|E_n^{(m)}|$ are independent of n. Let $\mu_n^{(m)} \in \mathcal{P}(\mathcal{G}_*)$ denote the law of $[N_{n,\circ}[m]]$. Figure C.1 demonstrates the construction of $N_{n,\circ}[2]$ for the graph given by Figure 4.4.

Step 2: A Total Variation Convergence

The rooted planted tree $N_{n,\circ}[m]$ has the same graph structure as a truncated version of the planted PWIT: remove all vertices $\boldsymbol{i} = (i_1, i_2, \cdots i_l)$ such that either (1) $i_s > m$ for some $s \in \{1, 2, \cdots, l\}$, or (2) $l > m+1$, or (3) $l = m+1$ and $i_l \neq 0$. In particular, the number of un-planted children of every vertex in the truncated version is m, the depth of vertices are bounded by $m+1$, and the only vertices at depth $m+1$ are the planted pairs of the vertices at depth m. Let $N_\infty[m]$ denote the truncated planted PWIT, and let $\mu_\infty^{(m)}$ denote the law of $[N_\infty[m]]$. Now, using the same approach as in [20, Lemma 10], we show that $\mu_n^{(m)}$ converges to $\mu_\infty^{(m)}$ in total variation norm.

Lemma C.1. *For any fixed m, $\mu_n^{(m)} \xrightarrow{TV} \mu_\infty^{(m)}$ where the total variation convergence of*

positive measures is defined as follows:

$$\mu_n \xrightarrow{TV} \mu \text{ iff } \sup_A |\mu_n(A) - \mu(A)| \to 0$$

Proof. It is easy to see that $\mu_n^{(m)}$ is absolutely continuous with respect to $\mu_\infty^{(m)}$. Moreover, the Radon-Nikodym derivative of $\mu_n^{(m)}$ with respect to $\mu_\infty^{(m)}$ equals the likelihood ratio.

Consider similar steps on the planted PWIT to construct $N_\infty[m]$. Conditioned on the first $t-1$ steps of the construction of $N_\infty[m]$ and $N_{n,\circ}[m]$, we will calculate the ratio of the conditional densities for the next step of the construction. Since planted edges have the same $\text{Exp}(\lambda)$ distribution in both cases, we are only interested in the corresponding ratio of un-planted edges.

At tth step of the construction of $N_\infty[m]$, the conditional density of $(x_1, x_1+x_2, \cdots, x_1+\cdots+x_m)$ is $\exp(-(x_1+x_2+\cdots+x_m))$. At tth step of the construction of $N_{n,\circ}[m]$, using the memoryless property of exponential random variables, the conditional density of $(x_1, x_1+x_2, \cdots, x_1+\cdots+x_m)$ is

$$\prod_{i=0}^{m-1} \frac{|I_{t-1}|-i}{n} \exp\left(-\frac{x_{i+1}(|I_{k-1}|-i)}{n}\right).$$

where I_{t-1} is the set of vertices that has not been added to $N_\infty[m]$ yet up to the t-th step. Hence, the ratio of conditional densities is at least

$$\prod_{i=0}^{m-1} \frac{|I_{k-1}|-i}{n} \geq \left(\frac{n-\left|V^{(m)}\right|}{n}\right)^m,$$

and we have

$$\frac{d\mu_n^{(m)}}{d\mu_\infty^{(m)}} \geq \left(1-\frac{\left|V^{(m)}\right|}{n}\right)^{\left|E^{(m)}\right|-\left|V^{(m)}\right|/2},$$

where the exponent $\left|E^{(m)}\right| - \left|V^{(m)}\right|/2$ is the number of un-planted edges explored.

Note that $\left|V^{(m)}\right|$ and $\left|E^{(m)}\right|$ do not depend on n, for all sufficiently large n. Hence, as $n \to \infty$ the right-hand side of the above inequality goes to 1. Now, the result follows by the fact that $\mu_n^{(m)}$ and $\mu_\infty^{(m)}$ are probability measures. □

Step 3: Locally Tree-Like Property

Fix some $\rho > 0$. Recall that $(G_{n,\circ}(1))_\rho$ denotes the ρ-neighborhood of vertex 1 in $N_{n,\circ}(1)$ as is defined in Appendix 4.4. Similarly, $(G_{n,\circ}[m])_\rho$ denotes the neighborhood ρ of node

1 in $N_{n,\circ}[m]$. The question is, whether these two neighborhoods are the same. Note that $(G_{n,\circ}[m])_\rho$ is a tree but $(G_{n,\circ}(1))_\rho$ is not necessary. However, it becomes a tree with high probability.

Lemma C.2. *Fix $\epsilon > 0$ and $\rho > 0$. Then there exists large enough $m_0(\epsilon, \rho)$ such that for all fixed $m > m_0(\epsilon, \rho)$,*

$$\mathbb{P}[(G_{n,\circ}(1))_\rho \neq (G_{n,\circ}[m])_\rho] \leq \epsilon, \qquad \text{as } n \to \infty.$$

Proof. Let m_0 to be large enough such that

$$\mathbb{P}[\text{number of vertices in } (G_{n,\circ}(1))_\rho > m_0] < \epsilon/2.$$

Fix $m > m_0$. Consider the event $\{(G_{n,\circ}(1))_\rho \neq (G_{n,\circ}[m])_\rho\}$. This event may happens if either the number of vertices in $(G_{n,\circ}(1))_\rho$ is greater than m or there are two vertices v, $w \in V_n^{(m)}$ such that $\{v, w\} \notin E_n^{(m)}$ but $\ell_n(v, w) \leq \rho$. The probability of the first event is bounded by $\epsilon/2$. For the other event, note that if $v, w \in V_n^{(m)}$ and $\{v, w\} \in E_n \setminus E_n^{(m)}$, then $\ell_n(v, w)$ is dominated by an exponentially distributed random variable with mean n. (To see this, assume that v was revealed earlier than w. Then we know that the cost of the edge $\{v, w\}$ is larger than the cost c_m of the edge $\{v, v_m\}$, where v_m is the mth closest un-planted neighbor of v. Hence, the probability distribution of the weight of $\{v, w\}$, using the memoryless property, is $1/n\text{Exp}((x-c_m)/n)$ which is stochastically larger than a random variable distributed as $\text{Exp}(1/n)$. Hence,

$$\begin{aligned}\mathbb{P}[\exists v, w \in V_n^{(m)} \text{ such that } \{v, w\} \notin E_n^{(m)} \text{ and } \ell_n(v, w) \leq \rho] \\ \leq \binom{|V_n^{(m)}|}{2}(1 - \exp(-t/n)) \to 0 \text{ as } n \to \infty.\end{aligned}$$

As we mentioned before, $|V_n^{(m)}|$ is independent of n for all sufficiently large n. Now, the result follows by combining the last two displayed inequalities. □

Now, combining Lemma C.1 and Lemma C.2, we get the following corollary.

Corollary C.3. *Fix $\rho > 0$. Let $\mu_{n,\rho}$ denote the law of $[((G_{n,\circ}(1))_\rho, \ell_n)]$, and let $\mu_{\infty,\rho}$ denote the law of $[((G_\infty)_\rho, \ell_\infty)]$. Then $\mu_{n,\rho} \xrightarrow{TV} \mu_{\infty,\rho}$.*

Step 4: Portmanteau Theorem

For a fixed $R > 0$, since the condition $d([N_\circ], [T_\circ]) < (R+1)^{-1}$ is equivalent to $d([((G_\circ)_R, \ell)], [T_\circ]) < (R+1)^{-1}$, Corollary C.3 implies that, for all finite rooted planted trees $[T_\circ] \in \mathcal{G}_*$,

we have

$$|\mu_n(A_{T_\circ}) - \mu_\infty(A_{T_\circ})| \to 0 \text{ as } n \to \infty\,, \tag{C.13}$$

where $A_{T_\circ}$ is defined as

$$A_{T_\circ} \coloneqq \{[N_\circ] \in \mathcal{G}_* : d([N_\circ], [T_\circ]) < (R+1)^{-1}\}.$$

Note that the support of μ_∞ is rooted planted trees. Moreover, recall that $\mathcal{G}_*$ is separable, hence, the restriction of $\mathcal{G}_*$ to the rooted planted trees is also separable. Since μ_∞ is a probability measure, for any $R > 0$ and any $\epsilon > 0$, there exists a finite set $S(R, \epsilon)$ consisting of rooted planted trees $T_\circ = (G_\circ, \ell)$ with $(G_\circ)_R = G_\circ$ such that

$$\mu_\infty\left(\bigcup_{T_\circ \in S(R,\epsilon)} A_{T_\circ}\right) > 1 - \epsilon.$$

Using Corollary C.3, there exists $n_0(\epsilon) \in \mathbb{N}_+$ such that for all $n > n_0(\epsilon)$,

$$\mu_n\left(\bigcup_{T_\circ \in S(R,\epsilon)} A_{T_\circ}\right) > 1 - 2\epsilon.$$

Now, we are going to prove that $\mu_n \xrightarrow{w} \mu_\infty$. By definition $\mu_n \xrightarrow{w} \mu_\infty$, if for any continuous bounded function $f : \mathcal{G}_* \to \mathbb{R}$,

$$\int_{\mathcal{G}_*} f\, d\mu_n \to \int_{\mathcal{G}_*} f\, d\mu_\infty.$$

Using the Portmanteau Theorem, we can restrict our attention to the uniformly continuous bounded functions. Let $f : \mathcal{G}_* \to \mathbb{R}$ be a uniformly continuous bounded function. Now, for any $\epsilon > 0$, there is a $\delta > 0$ such that if $d([N_\circ], [N'_\circ]) < \delta$ then $|f([N_\circ]) - f([N'_\circ])| \le \epsilon$. Fix the value of ϵ and let R to be large enough such that $(R+1)^{-1} < \delta$. We have,

$$\left|\int_{\mathcal{G}_*} f d\mu_n - \int_{\mathcal{G}_*} f d\mu_\infty\right| \le 3\epsilon|f|_\infty + \sum_{T_\circ \in S(R,\epsilon)} f(T_\circ)\,|\mu_n(A_{T_\circ}) - \mu_\infty(A_{T_\circ})| + 2\epsilon.$$

where $|f|_\infty \coloneqq \sup_{N \in \mathcal{G}_*} |f(N)|$. The result follows by arbitrary choice of ϵ, the fact that $|S(R,\epsilon)| < \infty$, and (C.13).

C.4 Proof of Proposition IV.4

By Lemma IV.3, $\mathscr{M}_{\infty,\mathrm{opt}}$ is a deterministic function of the collection of random variables

$$\mathcal{C}_\infty = \{\ell_\infty(e), X(\overleftrightarrow{e}); e \in E_\infty, \text{ and } \overleftrightarrow{e} \text{ is directed}\}.$$

Also note that $X(\overleftrightarrow{e})$ satisfies (4.31), which does not depend on the relabeling of the vertices. Now, by construction, we only need to show that the distribution of $(X(\overrightarrow{e}); e \in \overrightarrow{E}_\infty(h))$ for $h \geq 1$ is invariant with respect to the involution map ι.

Abusing the notation, let $\widehat{\mu}_{\infty,\mathrm{opt}}$ denote the law of $\mathcal{C}_\infty$ and define $\widetilde{\mu}_{\infty,\mathrm{opt}}$ to be $\widehat{\mu}_{\infty,\mathrm{opt}} \times \texttt{count}$ on $\mathbb{N}_+$, similar to (4.24). Fix the second root $k \geq 0$ and let

$$B = \Big\{(X(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{E}_\infty(h)) \in \cdot\,,\ k \text{ is distinguished}\Big\} \tag{C.14}$$

denote a measurable subset on $\widehat{\mathcal{G}}_{**}$, where $\widehat{\mathcal{G}}_{**}$ is defined similar to $\mathcal{G}_{**}$ (the set all isomorphism classes of connected locally finite doubly-rooted planted networks) with an additional mark on the edges $\overleftrightarrow{e}$ representing $X(\overleftrightarrow{e})$. Let (X_0, Y_0) denote a solution of the system of recursive distributional equations (4.29)–(4.30) as in Lemma IV.3. Recall that if $\overrightarrow{e}$ is a planted edge then $X(\overrightarrow{e})$ and X_0 have the same distribution; otherwise, $X(\overrightarrow{e})$ and Y_0 have the same distribution. Note that the collection of random variables $(X(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{E}_\infty(h))$ are independent since they depend on messages received from disjoint subtrees.

We need to show that

$$\widetilde{\mu}_{\infty,\mathrm{opt}}(\iota^{-1}(B)) = \widetilde{\mu}_{\infty,\mathrm{opt}}(B).$$

We treat the cases $k = 0$ or $k > 0$ separately.

(i) If $k = 0$, then we have

$$\iota^{-1}(B) = \big\{\text{the double root is } (\emptyset, 0)\big\} \cap \big\{(X(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{E}_\infty(h, 0)) \in \cdot\,\big\},$$

where $\overrightarrow{E}_\infty(h,0) \coloneqq \{(v, vj) : \{v, vj\} \in E_\infty \text{ s.t. } v = i_1 i_2 i_3 \cdots i_{h-2} \text{ with } i_1 \neq 0 \text{ or } v = 0 i_2 i_3 \cdots i_h\}$. Now to complete the proof, it suffices to show that $(X(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{E}_\infty(h, 0))$ has the same distribution as $(X(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{E}_\infty(h))$. Clearly, the collection of random variables $(X(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{E}_\infty(h,0))$ are independent, and $X(\overrightarrow{e})$ has the same distribution as X_0 or Y_0 depending on whether $\overrightarrow{e}$ is planted or un-planted. Thus it remains to prove that there is a one-to-one map from $\overrightarrow{E}_\infty(h,0)$ to $\overrightarrow{E}_\infty(h)$ that maps (un)planted edges to (un)planted ones. Consider the relabeling function ϕ defined as

follows

$$\begin{aligned}
&\phi(0i_2i_3\cdots i_h) = i_2i_3\cdots i_h && \forall\, 0i_2i_3\cdots i_h \in V_\infty,\\
&\phi(i_1i_2\cdots i_{h-2}) = 0i_1i_2\cdots i_{h-2} && \forall\, i_1i_2\cdots i_{h-2} \text{ with } i_1 \neq 0,
\end{aligned}$$

and define $\gamma : \overrightarrow{E}_\infty(h,0) \to \overrightarrow{E}_\infty(h)$ by $\gamma(\{v, vj\}) \coloneqq \{\phi(v), \phi(v)j\}$.

(ii) If $k > 0$, then we have

$$\iota^{-1}(B) = \bigcup_{l>0} \Big\{ \big\{\text{the double root is } (\varnothing, l)\big\} \cap A_l \cap \big\{(X(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{E}_\infty(h,l)) \in \cdot\,\big\} \Big\},$$

where

$$A_l \coloneqq \big\{\ell_\infty(l, l(k-1)) < \ell_\infty(\varnothing, l) < \ell_\infty(l, lk)\big\},$$

and $\overrightarrow{E}_\infty(h,l)$ is defined similar to $\overrightarrow{E}_\infty(h,0)$. Note that the events $\{(X(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{E}_\infty(h,l)) \in \cdot\,\}$ and A_l are independent, and the distribution of $(X(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{E}_\infty(h, l))$ does not depend on l. Also, note that by (4.48), we have

$$\sum_{l>0} \mathbb{P}[A_l] = \int_0^\infty \mathbb{P}[\text{exactly } k-1 \text{ arrivals before } x]\,\mathrm{d}x = 1.$$

The result then follows using the same argument as in the previous case.

C.5 Proof of Lemma IV.8

There are two branches that we are interested in: the alternating path from ø through v_{-1}, and the alternating path from ø through v_1. It is more convenient to study these two branches on the doubly rooted planted PWIT, rooted at $(\varnothing, v_1)$. The proof uses the discussion of bi-infinite planted PWITs $T_u^{\leftrightarrow}$ and $T_p^{\leftrightarrow}$ in Appendix 4.7.2.1. We follow the same notation and simplification (Remark IV.2 in Appendix 4.7.2.1) here.

Using the relabeling maps $\psi\big|_p$ and $\psi\big|_u$, we already know that $\mu_\infty \times \delta_0$ is equivalent to $\mu_p^{\leftrightarrow}$ and $\mu_\infty \times \texttt{count}$ on $\{1, 2, 3, \cdots\}$ is equivalent to $\mu_u^{\leftrightarrow}$. We can use the relabeling map $\psi\big|_\cdot$ to define $\{X_\cdot(\overrightarrow{e}), \overrightarrow{e}$ is a directed edge in $E_\cdot^{\leftrightarrow}\}$ jointly with $\{\ell_\cdot^{\leftrightarrow}(e), e \in E_\cdot^{\leftrightarrow}\}$. Note that the joint distribution of $\{\ell_\cdot^{\leftrightarrow}(e), X_\cdot(\overrightarrow{e}); e \in E_\cdot^{\leftrightarrow}$ and $\overrightarrow{e}$ is directed$\}$ is exactly the same as if we

use the construction of Lemma IV.1 by redefining $\overrightarrow{E}(h)$ as

$$\overrightarrow{bE}_{\cdot}^{\leftrightarrow}(h) \coloneqq \{\overrightarrow{e} = (-v, -vj) : \text{gen}(v) = h-1\} \cup \{\overrightarrow{e} = (+v, +vj) : \text{gen}(v) = h-1\}. \tag{C.15}$$

Now, given $\{X_{\cdot}(\overrightarrow{e}), \overrightarrow{e}$ is a directed edge in $E_{\cdot}^{\leftrightarrow}\}$ we can define a minimum matching $\mathscr{M}_{\cdot,\text{opt}}^{\leftrightarrow}$ on $T_{\cdot}^{\leftrightarrow}$, same as in Lemma IV.3, i.e.,

$$\forall v \in V_{\cdot}^{\leftrightarrow} : \mathscr{M}_{\cdot,\text{opt}}^{\leftrightarrow}(v) = \underset{w:\{v,w\}\in E_{\cdot}^{\leftrightarrow}}{\arg\min} \left(\ell_{\cdot}^{\leftrightarrow}(v,w) - X_{\cdot}(v,w)\right) \tag{C.16}$$

$$\forall e \in E_{\cdot}^{\leftrightarrow} : \mathscr{M}_{\cdot,\text{opt}}^{\leftrightarrow}(e) = 1 \text{ if and only if } \ell_{\cdot}^{\leftrightarrow}(e) < X_{\cdot}(\overrightarrow{e}) + X_{\cdot}(\overleftarrow{e}). \tag{C.17}$$

Next, we are going to show that the bi-infinite tree $T_u^{\leftrightarrow}$ ($T_p^{\leftrightarrow}$) restricted to $\mathscr{M}_{u,\text{opt}}^{\leftrightarrow}(-\emptyset,+\emptyset) = 1$ ($\mathscr{M}_{p,\text{opt}}^{\leftrightarrow}(-\emptyset,+\emptyset) = 1$) is equivalent to the doubly rooted planted PWIT, rooted at $\{\emptyset, \mathscr{M}_{\text{opt}}(\emptyset)\}$, restricted to $\mathscr{M}_{\text{opt}}(\emptyset) \neq 0$ ($\mathscr{M}_{\text{opt}}(\emptyset) = 0$).

On the planted PWIT T_u (recall that $T_u = T(\emptyset)$ as defined in Appendix 4.7.2.1), define $X_u^{\downarrow} = \min_{i\geq 0}(\ell(\emptyset,i) - X(\emptyset,i))$. Let $\nu_u(x)$ denote the conditional distribution of the set

$$\{\ell(e), X(\overrightarrow{e}); e \in E_h \text{ and } \overrightarrow{e} \text{ is directed away from } \emptyset\}$$

given $X_u^{\downarrow} = x$. Similarly, on the subtree T_p (recall that T_p is a relabeling of $T(0)$ as defined in Appendix 4.7.2.1), define $X_p^{\downarrow} = \min_{i\geq 1}(\ell(\emptyset,i) - X(\emptyset,i))$. Let $\nu_p(x)$ denote the conditional distribution of the set

$$\{\ell(e), X(\overrightarrow{e}); e \in E_p \text{ and } \overrightarrow{e} \text{ is directed away from } \emptyset\}$$

given $X_p^{\downarrow} = x$. On the bi-infinite tree $T_{\cdot}^{\leftrightarrow}$, define $\mu_{\cdot}^1$ to be the measure obtained by restricting $\mu_{\cdot}^{\leftrightarrow}$ to the set $\{\ell_{\cdot}^{\leftrightarrow}(-\emptyset,+\emptyset) < X_{\cdot}(-\emptyset,+\emptyset) + X_{\cdot}(+\emptyset,-\emptyset)\}$, i.e., $\mathscr{M}_{\cdot,\text{opt}}^{\leftrightarrow}(-\emptyset,+\emptyset) = 1$. Let $E_{\cdot}^{\leftrightarrow+}$ and $E_{\cdot}^{\leftrightarrow-}$ denote all edges of form $\{+v,+vj\}$ and $\{-v,-vj\}$ respectively. Clearly, $E_{\cdot}^{\leftrightarrow} = E_{\cdot}^{\leftrightarrow+} \cup E_{\cdot}^{\leftrightarrow-} \cup \{-\emptyset,+\emptyset\}$. Let (X_0, Y_0) be a solution of the system of recursive distributional equations (4.29)–(4.30).

Lemma C.1. *The measures μ_u^1 and μ_p^1 are finite positive measures. The total mass of μ_u^1 equals $\mathbb{P}[\mathscr{M}_{opt}(\emptyset) \neq 0]$ and the total mass of μ_p^1 is $\mathbb{P}[\mathscr{M}_{opt}(\emptyset) = 0]$. Under $\mu_{\cdot}^1$ we have:*

(i) The joint density of $(\ell_u^{\leftrightarrow}(-\emptyset,+\emptyset), X_u(+\emptyset,-\emptyset), X_u(-\emptyset,+\emptyset))$ at point (l,x_1,x_2) is $f_u(x_1)f_u(x_2)\mathbf{1}_{(0<l<x_1+x_2)}$, where $f_u(\cdot)$ is the density of Y_0 and $\mathbf{1}_{(0<l<x_1+x_2)}$ is the indicator function; the joint density of $\left(\ell_p^{\leftrightarrow}(-\emptyset,+\emptyset), X_p(+\emptyset,-\emptyset), X_p(-\emptyset,+\emptyset)\right)$ at point (l,x_1,x_2) is $f_p(x_1)f_p(x_2)\lambda Exp(-\lambda l)$, where $f_p(\cdot)$ is the density of X_0.

(ii) Conditioned on $(\ell_\cdot^{\leftrightarrow}(-\emptyset,+\emptyset), X_\cdot(+\emptyset,-\emptyset), X_\cdot(-\emptyset,+\emptyset)) = (l, x_1, x_2)$ *with* $x_1 + x_2 > l$*, the distribution of the family*

$$\{\ell_\cdot^{\leftrightarrow}(e), X_\cdot(\overrightarrow{e}); e \in E_\cdot^{\leftrightarrow -} \text{ and } \overrightarrow{e} \text{ is directed away from } -\emptyset\}$$

is the image of $\nu_\cdot(x_1)$ *under the natural embedding* $T_\cdot \to T_\cdot^- \subset T_\cdot^{\leftrightarrow}$*; the distribution of the family*

$$\{\ell_\cdot^{\leftrightarrow}(e), X_\cdot(\overrightarrow{e}); e \in E_\cdot^{\leftrightarrow +} \text{ and } \overrightarrow{e} \text{ is directed away from } +\emptyset\}$$

is the image of $\nu_\cdot(x_2)$ *under the natural embedding* $T_\cdot \to T_\cdot^+ \subset T_\cdot^{\leftrightarrow}$ *and these two families are conditionally independent.*

Remark C.2. Conditioned on $(\ell_\cdot^{\leftrightarrow}(-\emptyset,+\emptyset), X_\cdot(+\emptyset,-\emptyset), X_\cdot(-\emptyset,+\emptyset)) = (l, x_1, x_2)$ with $x_1 + x_2 > l$, we have $\mathscr{M}_{\cdot,\mathrm{opt}}^{\leftrightarrow}(-\emptyset,+\emptyset) = 1$. Now, by (C.16) and the construction of $X_\cdot$ on $T_\cdot^{\leftrightarrow}$, we have $X_\cdot(-i,-\emptyset) = \ell_\cdot^{\leftrightarrow}(-\emptyset,+\emptyset) - X_\cdot(-\emptyset,+\emptyset)$ for all i. Similarly, $X_\cdot(+i,+\emptyset) = \ell_\cdot^{\leftrightarrow}(+\emptyset, -\emptyset) - X_\cdot(+\emptyset,-\emptyset)$ for all i. This combined with the families in part (ii) of Lemma C.1, specifies $X_\cdot$ on $T_\cdot^{\leftrightarrow}$ under $\mu_\cdot^1$.

Proof. By construction of $X_\cdot$ on $T_\cdot^{\leftrightarrow}$, we already know that $X_\cdot(+\emptyset,-\emptyset)$ and $X_\cdot(-\emptyset,+\emptyset)$ are independent with density $f_\cdot(\cdot)$. Moreover, $\ell_u^{\leftrightarrow}(-\emptyset,+\emptyset)$ has uniform "distribution" on $[0,\infty)$, and $\ell_p^{\leftrightarrow}(-\emptyset,+\emptyset)$ is an exponentially distributed random variable with parameter λ. Hence, the joint density has the form mentioned in (i). Moreover, the total mass of μ_u^1 is

$$\int_{x_1=-\infty}^{\infty}\int_{x_2=-\infty}^{\infty}(x_1+x_2)^+ f_u(x_1) f_u(x_2)\,dx_2\,dx_1, \tag{C.18}$$

and the total mass of μ_p^1 is

$$\int_{x_1=-\infty}^{\infty}\int_{x_2=-\infty}^{\infty}(1-\exp\bigl(-\lambda(x_1+x_2)^+\bigr)) f_p(x_1) f_p(x_2)\,dx_2\,dx_1, \tag{C.19}$$

where $(x_1+x_2)^+ = \max(x_1+x_2, 0)$. Now, using the joint density above, we can calculate the total mass of μ_u^1 and μ_p^1 as shown by the following lemma.

Lemma C.3. *The equation* (C.18) *equals* $\mathbb{P}[\mathscr{M}_{opt}(\emptyset) \neq 0]$*, and the equation* (C.19) *equals* $\mathbb{P}[\mathscr{M}_{opt}(\emptyset) = 0]$.

proof of Lemma C.3. Let X_1 and X_2 denote two independent copies of X_0, and Y_1 and Y_2 denote two independent copies of Y_0. Let η denote an exponentially distributed random variable with parameter λ.

Using Corollary IV.2, we have

$$\begin{aligned}
\text{Equation (C.18)} &= \mathbb{E}[(Y_1 + Y_2)^+] \\
&= \int_{x=0}^{\infty} \mathbb{P}[Y_1 + Y_2 > x]\, dx \\
&= \int_{\zeta=0}^{\infty} \mathbb{P}[X(\overrightarrow{e}) + X(\overleftarrow{e}) > \ell(e) \,|\, \exists e = \{\emptyset, i \geq 1\}, \ell(e) = \zeta]\, d\zeta \\
&= \mathbb{P}[\mathscr{M}_{\mathrm{opt}}(\emptyset) \neq 0], \\
\text{Equation (C.19)} &= \mathbb{P}[X_1 + X_2 > \eta] = \mathbb{P}[X(\emptyset, 0) + X(0, \emptyset) > \ell(\emptyset, 0)] = \mathbb{P}[\mathscr{M}_{\mathrm{opt}}(\emptyset) = 0].
\end{aligned}$$

□

Next, by construction based on $(X_\cdot(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{bE}_\cdot^{\leftrightarrow}(h))$ as in Lemma IV.1, under $\mu_\cdot^{\leftrightarrow}$ the families

$$\{\ell_\cdot^{\leftrightarrow}(e), X_\cdot(\overrightarrow{e}); e \in E_\cdot^{\leftrightarrow -} \text{ and } \overrightarrow{e} \text{ is directed away from } -\emptyset\} \cup X_\cdot(+\emptyset, -\emptyset),$$

and

$$\{\ell_\cdot^{\leftrightarrow}(e), X_\cdot(\overrightarrow{e}); e \in E_\cdot^{\leftrightarrow +} \text{ and } \overrightarrow{e} \text{ is directed away from } +\emptyset\} \cup X_\cdot(-\emptyset, +\emptyset),$$

are independent of each other and $\ell_\cdot^{\leftrightarrow}(-\emptyset, +\emptyset)$. Therefore, the desired conditional independence in part (ii) follows, when conditioned on $(\ell_\cdot^{\leftrightarrow}(-\emptyset, +\emptyset), X_\cdot(+\emptyset, -\emptyset), X_\cdot(-\emptyset, +\emptyset)) = (l, x_1, x_2)$.

Finally, note that each families under $\mu_\cdot^{\leftrightarrow}$ is distributed as the image of corresponding family on $T_\cdot$, where $X_\cdot(+\emptyset, -\emptyset)$ (or $X_\cdot(-\emptyset, +\emptyset)$) corresponds to $X_\cdot^{\downarrow}$. Now, the independence of these two families under $\mu_\cdot^{\leftrightarrow}$ implies that the conditional distribution of families under $\mu_\cdot^1$ depends only on the corresponding value of $X_\cdot^{\downarrow}$, i.e., x_1 for the first family and x_2 for the second one. □

Recall that $\psi\big|_p : [0,\infty)^E \times \{0\} \to [0,\infty)^{E_p^{\leftrightarrow}}$ maps $\mu \times \delta_0$ to $\mu_p^{\leftrightarrow}$, and $\psi\big|_u : [0,\infty)^E \times \{1, 2, 3, \cdots\} \to [0,\infty)^{E_u^{\leftrightarrow}}$ maps $\mu \times \texttt{count}$ on $\{1, 2, 3, \cdots\}$ to $\mu_u^{\leftrightarrow}$. Note that the inverse image of the event $\{\ell_\cdot^{\leftrightarrow}(-\emptyset,+\emptyset) < X_\cdot(+\emptyset, -\emptyset) + X_\cdot(-\emptyset, +\emptyset)\}$ under $\psi\big|_\cdot$ is the event

$$\{(\emptyset, \mathscr{M}_{\mathrm{opt}}(\emptyset)) \text{ is the double root}\}.$$

Hence, $\psi\big|_p^{-1}$ maps the measure μ_p^1 to $\mu \times \delta_0$ restricted to $\{\text{the second root is } \mathscr{M}_{\mathrm{opt}}(\emptyset) = 0\}$,

and $\psi\big|_u^{-1}$ maps the measure μ_u^1 to $\mu \times \texttt{count}$ on $\{1, 2, 3, \cdots\}$ restricted to

$$\{\text{the second root is } \mathscr{M}_{\text{opt}}(\emptyset) \text{ and } \mathscr{M}_{\text{opt}}(\emptyset) \neq 0\}.$$

Hence, to study the events $\overline{B}_\infty$ and B_{-1} on the planted PWIT, we can relabel the vertices by setting $\emptyset$ to be $-\emptyset$, $\mathscr{M}_{\text{opt}}(\emptyset)$ to be $+\emptyset$, map the doubly rooted PWIT, rooted at $(\emptyset, \mathscr{M}_{\text{opt}}(\emptyset))$, to the corresponding bi-infinite tree ($T_u^{\leftrightarrow}$ or $T_p^{\leftrightarrow}$ depending on whether $\mathscr{M}_{\text{opt}}(\emptyset) = 0$ or not), and study the image of these events under μ_p^1 and μ_u^1.

C.5.1 A Lemma on T_u and T_p

Before analyzing the image of B_{-1} on $T_u^{\leftrightarrow}$ (or $T_p^{\leftrightarrow}$ depending on whether $\mathscr{M}_{\text{opt}}(\emptyset) = 0$ or not), let us present a technical lemma which generalizes Lemma 23 in [20] to the planted case.

Lemma C.4. *On the planted PWIT, define*

$$\begin{aligned}
X_u^{\downarrow} &= \min_{i \geq 0}(\ell(\emptyset, i) - X(\emptyset, i)),\\
X_p^{\downarrow} &= \min_{i \geq 1}(\ell(\emptyset, i) - X(\emptyset, i)),\\
I_u &= \arg\min_{i \geq 0}(\ell(\emptyset, i) - X(\emptyset, i)),\\
I_p &= \arg\min_{i \geq 1}(\ell(\emptyset, i) - X(\emptyset, i)).
\end{aligned}$$

For $-\infty < b < a < \infty$ define

$$\begin{aligned}
g_u(a, b) &\coloneqq \mathbb{P}[\ell(\emptyset, I_u) - b > \min_{j:\{I_u, I_u j\} \in E}^{[2]}(\ell(I_u, I_u j) - X(I_u, I_u j)) \mid X_u^{\downarrow} = a],\\
g_p(a, b) &\coloneqq \mathbb{P}[\ell(\emptyset, I_p) - b > \min_{j:\{I_p, I_p j\} \in E}^{[2]}(\ell(I_p, I_p j) - X(I_p, I_p j)) \mid X_p^{\downarrow} = a].
\end{aligned}$$

Then $g_u(a, b), g_p(a, b) > 0$.

Proof. Let us begin with an observation, which is the continuous analogous of the splitting property of a Poisson process.

Observation. Let $\{X_i\}_{i=1}^{\infty}$ be independent real-valued continuous random variables with common distribution α_X. Let $\{\zeta_i\}_{i=1}^{\infty}$ denote the arrivals of a Poisson process with parameter 1. Then $\{(\zeta_i, X_i)\}_{i=1}^{\infty}$ forms a Poisson point process on $[0, \infty) \times (-\infty, +\infty)$ with mean intensity $\beta(z, x)dzdx = dz\alpha_X(dx)$.

Now, let $Y_i = \zeta_i - X_i$. The set of points $\{Y_i\}_{i=1}^{\infty}$ forms a certain inhomogeneous Poisson process on $(-\infty, \infty)$ with mean intensity $\gamma(y)dy = \alpha_X([-y, \infty))dy$. Finally, it is easy to see

that conditioned on the time of the first arrival to be y_0, the other points in $\{Y_i\}_{i=1}^{\infty}$ are the points of a certain inhomogeneous Poisson process on (y_0, ∞). Similar statement holds if we condition on no arrival before y_0.

Note that by (4.31), we have

$$X(\emptyset, I_\cdot) = \min_{j:\{I_\cdot, I_\cdot j\}\in E} (\ell(I_\cdot, I_\cdot j) - X(I_\cdot, I_\cdot j)).$$

Now, by the above observation, the set of points $\{\ell(I_\cdot, I_\cdot j) - X(I_\cdot, I_\cdot j), j \geq 1\}$ conditioned on $X(\emptyset, I_\cdot) = x$, are the points of a certain inhomogeneous Poisson process on (x, ∞) (note that the claim is true regardless of whether $\arg\min_{j:\{I_\cdot, I_\cdot j\}\in E}(\ell(I_\cdot, I_\cdot j) - X(I_\cdot, I_\cdot j)) = 0$ or not). Hence,

$$\mathbb{P}[\min_{j:\{I_\cdot, I_\cdot j\}\in E}^{[2]} (\ell(I_\cdot, I_\cdot j) - X(I_\cdot, I_\cdot j)) \in [y, y + dy] \,|\, X(\emptyset, I_\cdot) = x] \geq \widetilde{\beta}_x(y)dy,$$

where $\widetilde{\beta}_x(y) > 0$ for all $y > x$. Since the above term does not depend on the value of $\ell(\emptyset, I_\cdot)$, we have

$$\widetilde{g}_\cdot(a, b, x) \coloneqq \mathbb{P}[\ell(\emptyset, I_\cdot) - b > \min_{j:\{I_\cdot, I_\cdot j\}\in E}^{[2]} (\ell(I_\cdot, I_\cdot j) - X(I_\cdot, I_\cdot j)) \,|\, X(\emptyset, I_\cdot) = x, \ell(\emptyset, I_\cdot) = a + x] > 0,$$

for all $-\infty < b < a < \infty$, and $-\infty < x < \infty$. Now, the result follows by

$$g_\cdot(a, b) = \mathbb{E}[\widetilde{g}_\cdot(a, b, X(\emptyset, I_\cdot)) \,|\, X_\cdot^{\downarrow} = a],$$

since $X_\cdot^{\downarrow} = \ell(\emptyset, I_\cdot) - X(\emptyset, I_\cdot)$. □

C.5.2 Calculating with the Bi-infinite Tree

On the bi-infinite tree $T_\cdot^{\leftrightarrow}$, define the event $C_{\cdot,-1}$ as

$$C_{\cdot,-1} \coloneqq \left\{ -\emptyset = \operatorname*{arg\,min}_{y:\{y,-I_\cdot\}\in E_\cdot^{\leftrightarrow}}^{[2]} (\ell_\cdot^{\leftrightarrow}(-I_\cdot, y) - X_\cdot(-I_\cdot, y)) \right\},$$

where

$$-I_\cdot = \operatorname*{arg\,min}_{-i:\{-\emptyset,-i\}\in E_\cdot^{\leftrightarrow}} (\ell_\cdot^{\leftrightarrow}(-\emptyset, -i) - X_\cdot(-\emptyset, -i)).$$

The event $C_{\cdot,-1}$ under $\mu^1_\cdot$ corresponds to the event B_{-1} on the doubly rooted PWIT, rooted at $(\emptyset, v_1)$, where $v_1 = \mathscr{M}_{\text{opt}}(\emptyset)$. Define the following σ-algebras:

$$\begin{aligned}
\mathcal{F}^-_\cdot &= \sigma\left(X_\cdot(\overrightarrow{e}), \ell^{\leftrightarrow}_\cdot(e) : e \in E^{\leftrightarrow -}_\cdot \text{ and } \overrightarrow{e} \text{ is directed}\right), \\
\mathcal{F}^+_\cdot &= \sigma\left(X_\cdot(\overrightarrow{e}), \ell^{\leftrightarrow}_\cdot(e) : e \in E^{\leftrightarrow +}_\cdot \text{ and } \overrightarrow{e} \text{ is directed}\right), \\
\mathcal{F}^\emptyset_\cdot &= \sigma\left(\ell^{\leftrightarrow}_\cdot(-\emptyset, +\emptyset), X_\cdot(+\emptyset, -\emptyset), X_\cdot(-\emptyset, +\emptyset)\right).
\end{aligned}$$

Lemma C.5. $\mu^1_\cdot(C^c_{\cdot,-1} \big| \mathcal{F}^+_\cdot, \mathcal{F}^\emptyset_\cdot) = g_\cdot\left(X_\cdot(+\emptyset, -\emptyset), \ell^{\leftrightarrow}_\cdot(-\emptyset, +\emptyset) - X_\cdot(-\emptyset, +\emptyset)\right)$, *where* $C^c_{\cdot,-1}$ *is the complement of the event* $C_{\cdot,-1}$.

Proof. By Remark C.2 right after Lemma C.1,

$$\begin{aligned}
\mathcal{F}^-_\cdot \cap \mathcal{F}^\emptyset_\cdot &= \mathcal{F}^\emptyset_\cdot \cap \sigma\left(X_\cdot(\overrightarrow{e}), \ell^{\leftrightarrow}_\cdot(e) : e \in E^{\leftrightarrow -}_\cdot \text{ and } \overrightarrow{e} \text{ is directed away from } -\emptyset\right), \\
\mathcal{F}^+_\cdot \cap \mathcal{F}^\emptyset_\cdot &= \mathcal{F}^\emptyset_\cdot \cap \sigma\left(X_\cdot(\overrightarrow{e}), \ell^{\leftrightarrow}_\cdot(e) : e \in E^{\leftrightarrow +}_\cdot \text{ and } \overrightarrow{e} \text{ is directed away from } +\emptyset\right).
\end{aligned}$$

Now, since $C_{\cdot,-1}$ is $\mathcal{F}^-_\cdot$ measurable, by conditional independence of Lemma C.1 part (ii) we have

$$\mu^1_\cdot(C^c_{\cdot,-1} \big| \mathcal{F}^+_\cdot, \mathcal{F}^\emptyset_\cdot) = \mu^1_\cdot(C^c_{\cdot,-1} \big| \mathcal{F}^\emptyset_\cdot).$$

Hence, we need to show that for all (l, x_1, x_2),

$$\mu^1_\cdot\left\{C^c_{\cdot,-1} \big| (\ell^{\leftrightarrow}_\cdot(-\emptyset, +\emptyset), X_\cdot(+\emptyset, -\emptyset), X_\cdot(-\emptyset, +\emptyset)) = (l, x_1, x_2)\right\} = g_\cdot\left(x_1, l - x_2\right).$$

By Lemma C.1, conditioned on the event $(\ell^{\leftrightarrow}_\cdot(-\emptyset, +\emptyset), X_\cdot(+\emptyset, -\emptyset), X_\cdot(-\emptyset, +\emptyset)) = (l, x_1, x_2)$, the distribution of the family

$$\{\ell^{\leftrightarrow}_\cdot(e), X_\cdot(\overrightarrow{e}); e \in E^{\leftrightarrow -}_\cdot \text{ and } \overrightarrow{e} \text{ is directed away from } -\emptyset\},$$

is the image of $\nu_\cdot(x_1)$ under the natural embedding $T_\cdot \to T^-_\cdot \subset T^{\leftrightarrow}_\cdot$. Recall that $\nu_\cdot(x_1)$ is the distribution of $\{\ell(e), X(\overrightarrow{e}); e \in E_\cdot \text{ and } \overrightarrow{e} \text{ is directed away from } \emptyset\}$ on $T_\cdot$, given $X^\downarrow_\cdot = x_1$. Hence,

$$\begin{aligned}
g_\cdot\left(x_1, l - x_2\right) = \mu^1_\cdot\Bigg\{ & \ell^{\leftrightarrow}_\cdot(-\emptyset, -I_\cdot) - (l - x_2) > \\
& \min^{[2]}_{j:\{-I_\cdot, -I_\cdot j\} \in E^{\leftrightarrow -}_\cdot} (\ell^{\leftrightarrow}_\cdot(-I_\cdot, -I_\cdot j) - X_\cdot(-I_\cdot, -I_\cdot j)) \\
& \Big| (\ell^{\leftrightarrow}_\cdot(-\emptyset, +\emptyset), X_\cdot(+\emptyset, -\emptyset), X_\cdot(-\emptyset, +\emptyset)) = (l, x_1, x_2)\Bigg\}.
\end{aligned}$$

However, under this conditioning

$$\begin{aligned}\ell_{\cdot}^{\leftrightarrow}(-\emptyset,-I_{\cdot})-(l-x_2)&=\ell_{\cdot}^{\leftrightarrow}(-\emptyset,-I_{\cdot})-(\ell_{\cdot}^{\leftrightarrow}(-\emptyset,+\emptyset)-X_{\cdot}(-\emptyset,+\emptyset))\\&=\ell_{\cdot}^{\leftrightarrow}(-I_{\cdot},-\emptyset)-X_{\cdot}(-I_{\cdot},-\emptyset),\end{aligned}$$

where the last equality follows by (4.31) and the fact that under $\mu_{\cdot}^1$, $\mathscr{M}_{\cdot,\mathrm{opt}}^{\leftrightarrow}(-\emptyset,\emptyset)=1$ (see Remark C.2). Finally, note that under $\mu_{\cdot}^1$,

$$C_{\cdot,-1}^{c}=\left\{\ell_{\cdot}^{\leftrightarrow}(-I_{\cdot},-\emptyset)-X_{\cdot}(-I_{\cdot},-\emptyset)>\min_{j:\{-I_{\cdot},-I_{\cdot}j\}\in E_{\cdot}^{\leftrightarrow -}}{}^{[2]}(\ell_{\cdot}^{\leftrightarrow}(-I_{\cdot},-I_{\cdot}j)-X_{\cdot}(-I_{\cdot},-I_{\cdot}j))\right\}.$$

□

Now, we have all the machinery to finish the proof of Lemma IV.8. By using the relabeling bijections, $\mathbb{P}[B_{-1}\,|\,\overline{B}_{\infty},\{\mathscr{M}_{\mathrm{opt}}(\emptyset)=0\}]$ equals $\mu_p^1\left(C_{p,-1}\,|\,C_p\right)$ where C_p is a certain event which is measurable with respect to $\mathcal{F}_p^{+}\cap\mathcal{F}_p^{\emptyset}$ such that $\mu_p^1(C_p)=\mathbb{P}[\overline{B}_{\infty}\cap\{\mathscr{M}_{\mathrm{opt}}(\emptyset)=0\}]$. Similarly, $\mathbb{P}[B_{-1}\,|\,\overline{B}_{\infty},\{\mathscr{M}_{\mathrm{opt}}(\emptyset)\neq 0\}]$ equals $\mu_u^1\left(C_{u,-1}\,|\,C_u\right)$ for a certain event C_u that is defined similar to C_p. Now, by Lemma C.4 and Lemma C.5, if $\mu_{\cdot}^1(C_{\cdot})>0$, then we have

$$\begin{aligned}\mu_{\cdot}^1\left(C_{\cdot,-1}^{c}\cap C_{\cdot}\right)&=\mathbb{E}_{\mu_{\cdot}^1}\left[\mathbf{1}_{C_{\cdot}}\,\mu_{\cdot}^1(C_{\cdot,-1}^{c}\big|\mathcal{F}_{\cdot}^{+},\mathcal{F}_{\cdot}^{\emptyset})\right]\\&=\mathbb{E}_{\mu_{\cdot}^1}\left[\mathbf{1}_{C_{\cdot}}\,g_{\cdot}\left(X_{\cdot}(+\emptyset,-\emptyset),\ell_{\cdot}^{\leftrightarrow}(-\emptyset,+\emptyset)-X_{\cdot}(-\emptyset,+\emptyset)\right)\right]>0.\end{aligned}$$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] M. Chertkov, L. Kroc, F. Krzakala, M. Vergassola, and L. Zdeborová, "Inference in particle tracking experiments by passing messages between images," *PNAS*, vol. 107, no. 17, pp. 7663–7668, 2010.

[2] P. Erd́os and A. Rényi, "On random graphs i," *Publ. Math. Debrecen*, vol. 6, p. 290, 1959.

[3] E. N. Gilbert, "Random graphs," *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1141–1144, 12 1959. [Online]. Available: https://doi.org/10.1214/aoms/1177706098

[4] P. Erd́os, "Some remarks on the theory of graphs," *Bull. Amer. Math. Soc.*, vol. 53, no. 4, pp. 292–294, 04 1947. [Online]. Available: https://projecteuclid.org:443/euclid.bams/1183510596

[5] B. Bollobás, "A probabilistic proof of an asymptotic formula for the number of labelled regular graphs," *Eur. J. Comb.*, vol. 1, no. 4, pp. 311 – 316, 1980. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0195669880800308

[6] D. J. de Solla Price, "Networks of scientific papers," *Science*, vol. 149, no. 3683, pp. 510–515, 1965. [Online]. Available: https://science.sciencemag.org/content/149/3683/510

[7] D. D. S. Price, "A general theory of bibliometric and other cumulative advantage processes," *Journal of the American Society for Information Science*, vol. 27, no. 5, pp. 292–306, 1976. [Online]. Available: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.4630270505

[8] P. L. Krapivsky, S. Redner, and F. Leyvraz, "Connectivity of growing random networks," *Phys. Rev. Lett.*, vol. 85, pp. 4629–4632, Nov 2000. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.85.4629

[9] P. L. Krapivsky and S. Redner, "Organization of growing random networks," *Phys. Rev. E*, vol. 63, p. 066123, May 2001. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.63.066123

[10] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999. [Online]. Available: https://science.sciencemag.org/content/286/5439/509

[11] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi, "A survey of statistical network models," *Found. Trends Mach. Learn.*, vol. 2, no. 2, p. 129–233, Feb. 2010. [Online]. Available: https://doi.org/10.1561/2200000005

[12] B. Bollobás, *Random Graphs*, 2nd ed., ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001.

[13] A. Frieze and M. Karoński, *Introduction to Random Graphs*. Cambridge University Press, 2015.

[14] R. v. d. Hofstad, *Random Graphs and Complex Networks*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016, vol. 1.

[15] M. Molloy and B. Reed, *Graph Colouring and the Probabilistic Method*. Springer-Verlag Berlin Heidelberg, 2002.

[16] N. Alon and J. H. Spencer, *The Probabilistic Method*, 4th ed. Wiley-Interscience Series in Discrete Mathematics and Optimization, 2016.

[17] T. G. Kurtz, "Solutions of ordinary differential equations as limits of pure jump markov processes," *J. Appl. Probab.*, vol. 7, no. 1, pp. 49–58, 1970. [Online]. Available: http://www.jstor.org/stable/3212147

[18] R. M. Karp and M. Sipser, "Maximum matching in sparse random graphs," in *22nd Annual Symposium on Foundations of Computer Science (sfcs 1981)*, Oct 1981, pp. 364–375.

[19] N. Wormald, *The differential equation method for random graph processes and greedy algorithms*. Wydawnictwo Naukowe Pwn, 1999, pp. 73–155.

[20] D. Aldous, "Asymptotics in the random assignment problem," *Probability Theory and Related Fields*, vol. 93, no. 4, pp. 507–534, Dec 1992. [Online]. Available: https://doi.org/10.1007/BF01192719

[21] ——, "The $\zeta(2)$ limit in the random assignment problem," *Random Structures & Algorithms*, vol. 18, no. 4, pp. 381–418, 2001. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.1015

[22] Marc Mézard and Giorgio Parisi, "On the solution of the random link matching problems," *J. Phys. France*, vol. 48, no. 9, pp. 1451–1459, 1987. [Online]. Available: https://doi.org/10.1051/jphys:019870048090145100

[23] R. J. La and M. Kabkab, "A new random graph model with self-optimizing nodes: Connectivity and diameter," *Internet Math.*, vol. 11, no. 6, pp. 528–554, 2015. [Online]. Available: https://doi.org/10.1080/15427951.2015.1022626

[24] D. Aldous and M. J. Steele, *The Objective Method: Probabilistic Combinatorial Optimization and Local Weak Convergence*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 1–72. [Online]. Available: https://doi.org/10.1007/978-3-662-09444-0_1

[25] C. Bordenave, "Lecture notes on random graphs and probabilistic combinatorial optimization," April 2016. [Online]. Available: https://www.math.univ-toulouse.fr/~bordenave/coursRG.pdf

[26] I.-J. Bienaymé, "De la loi de multiplication et de la durée des familles," *Soc. Philomat. Paris Extraits*, vol. 13, p. 131–132, 1845.

[27] F. Galton and H. W. Watson, "On the probability of the extinction of families," *J. Anthropol. Inst. Great Britain and Ireland*, vol. 4, pp. 138–144, 1875.

[28] B. Bollobás and O. Riordan, *Random Graphs and Branching Processes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 15–115. [Online]. Available: https://doi.org/10.1007/978-3-540-69395-6_1

[29] R. Durrett, *Random Graph Dynamics*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2006.

[30] T. E. Harris, *The Theory of Branching Processes*. Springer-Verlag Berlin Heidelberg, 1963.

[31] K. B. Athreya and P. E. Ney, *Branching Processes*. Springer-Verlag Berlin Heidelberg, 1972.

[32] M. Granovetter, "Threshold models of collective behavior," *Am. J. Sociol.*, pp. 1420–1443, 1978.

[33] J. Balogh and B. G. Pittel, "Bootstrap percolation on the random regular graph," *Random Struct. Algor.*, vol. 30, no. 1-2, pp. 257–286, 2007. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20158

[34] H. Amini, "Bootstrap percolation and diffusion in random graphs with given vertex degrees," *Electronic Journal of Combinatorics*, vol. 17, p. R25, 2010.

[35] M. Lelarge, "Diffusion and cascading behavior in random networks," *Games and Economic Behavior*, vol. 75, no. 2, pp. 752–775, 2012.

[36] W. S. Rossi, G. Como, and F. Fagnani, "Threshold models of cascades in large-scale networks," *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 2, pp. 158–172, April 2019.

[37] A. Galstyan and P. Cohen, "Cascading dynamics in modular networks," *Physical Review E*, vol. 75, no. 3, p. 036109, 2007.

[38] J. P. Gleeson, "Cascades on correlated and modular random networks," *Physical Review E*, vol. 77, no. 4, p. 046117, 2008.

[39] P. Curie, *Propriétés magnétiques des corps a diverses températures*, 1895, vol. 5.

[40] P. Weiss, "L'hypothèse du champ moléculaire et la propriété ferromagnétique," *J. Phys. Theor. Appl.*, vol. 6, no. 1, pp. 661–690, 1907.

[41] M. Moharrami, V. Subramanian, M. Liu, and M. Lelarge, "Impact of community structure on cascades," in *Proceedings of the 2016 ACM Conference on Economics and Computation*, ser. EC '16. New York, NY, USA: ACM, 2016, pp. 635–636. [Online]. Available: http://doi.acm.org/10.1145/2940716.2953924

[42] ——, "Impact of community structure on cascades," `arXiv:1606.00858 [math.PR]`.

[43] C. Cooper and A. Frieze, "On the connectivity of random k-th nearest neighbour graphs," *Comb. Probab. Comput.*, vol. 4, no. 4, p. 343–362, 1995.

[44] D. Aldous and R. Lyons, "Processes on unimodular random networks," *Electron. J. Probab.*, vol. 12, pp. 1454–1508, 2007. [Online]. Available: https://doi.org/10.1214/EJP.v12-463

[45] I. Benjamini and O. Schramm, "Recurrence of distributional limits of finite planar graphs," *Electron. J. Probab.*, vol. 6, p. 13 pp., 2001. [Online]. Available: https://doi.org/10.1214/EJP.v6-96

[46] M. G. Krein and M. A. Rutman, "Linear operators leaving invariant a cone in a banach space," *Uspekhi Mat. Nauk*, vol. 3, pp. 3–95, 1948.

[47] J. F. Toland, "Self-adjoint operators and cones," *Journal of the London Mathematical Society*, vol. 53, no. 1, pp. 167–183, 1996. [Online]. Available: https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/jlms/53.1.167

[48] P. H. Baxendale, "Renewal theory and computable convergence rates for geometrically ergodic markov chains," *Ann. Appl. Probab.*, vol. 15, no. 1B, pp. 700–738, 02 2005. [Online]. Available: https://doi.org/10.1214/105051604000000710

[49] M. Moharrami, V. G. Subramanian, M. Liu, and R. Sundaresan, "Local weak convergence based analysis of a new graph model," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2018, pp. 502–503.

[50] ——, "The erlang weighted tree, a new branching process," `arXiv:2002.03993 [math.PR]`.

[51] M. Moharrami, C. Moore, and J. Xu, "The planted matching problem: Phase transitions and exact results," `arXiv:1912.08880v3 [cs.DS]`.

[52] T. C. Schelling, *Micromotives and Macrobehavior*. W. W. Norton & Company, October 1978.

[53] R. Naylor, "A social custom model of collective action," *European Journal of Political Economy*, vol. 6, no. 2, pp. 201–216, 1990.

[54] D. J. Watts and P. S. Dodds, "Influentials, networks, and public opinion formation," *Journal of consumer research*, vol. 34, no. 4, pp. 441–458, 2007.

[55] F. Vega-Redondo, *Complex social networks*. Cambridge University Press, 2007, no. 44.

[56] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[57] S. Morris, "Contagion," *The Review of Economic Studies*, vol. 67, no. 1, pp. 57–78, 2000.

[58] D. J. Watts, "A simple model of global cascades on random networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, pp. 5766–5771, 2002.

[59] D. López-Pintado, "Diffusion in complex social networks," *Games and Economic Behavior*, vol. 62, no. 2, pp. 573–590, 2008.

[60] J. P. Gleeson and D. J. Cahalane, "Seed size strongly affects cascades on random networks," *Physical Review E*, vol. 75, no. 5, p. 056103, 2007.

[61] A. Galstyan, V. Musoyan, and P. Cohen, "Maximizing influence propagation in networks with community structure," *Physical Review E*, vol. 79, no. 5, p. 056102, 2009.

[62] S. Janson, "The probability that a random multigraph is simple. ii," *J. Appl. Probab.*, vol. 51A, pp. 123–137, 12 2014. [Online]. Available: https://doi.org/10.1239/jap/1417528471

[63] E. Mossel, J. Neeman, and A. Sly, "Stochastic block models and reconstruction," *arXiv preprint arXiv:1202.1499*, 2012.

[64] M. Shrestha and C. Moore, "Message-passing approach for threshold models of behavior in networks," *Phys. Rev. E*, vol. 89, p. 022805, Feb 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.89.022805

[65] D. Kempe, J. Kleinberg, and E. 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*, ser. KDD '03. New York, NY, USA: ACM, 2003, pp. 137–146. [Online]. Available: http://doi.acm.org/10.1145/956750.956769

[66] E. Mossel and S. Roch, "Submodularity of influence in social networks: From local to global," *SIAM Journal on Computing*, vol. 39, no. 6, pp. 2176–2188, 2010. [Online]. Available: http://dx.doi.org/10.1137/080714452

[67] R. v. d. Hofstad, *Random Graphs and Complex Networks*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016, vol. 1.

[68] N. C. Wormald, "Differential equations for random processes and random graphs," *The annals of applied probability*, pp. 1217–1235, 1995.

[69] M. Molloy and B. Reed, "The size of the giant component of a random graph with a given degree sequence," *Combinatorics, probability and computing*, vol. 7, no. 03, pp. 295–305, 1998.

[70] S. JANSON, "The probability that a random multigraph is simple," *Combinatorics, Probability and Computing*, vol. 18, no. 1-2, p. 205–225, 2009.

[71] E. A. Bender and E. Canfield, "The asymptotic number of labeled graphs with given degree sequences," *Journal of Combinatorial Theory, Series A*, vol. 24, no. 3, pp. 296 – 307, 1978. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0097316578900596

[72] B. D. McKay and N. C. Wormald, "Asymptotic enumeration by degree sequence of graphs with degreeso(n1/2)," *Combinatorica*, vol. 11, no. 4, pp. 369–382, Dec 1991. [Online]. Available: https://doi.org/10.1007/BF01275671

[73] J. Blanchet and A. Stauffer, "Characterizing optimal sampling of binary contingency tables via the configuration model," *Random Structures & Algorithms*, vol. 42, no. 2, pp. 159–184, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20403

[74] J. LaSalle, "Some extensions of liapunov's second method," *IRE Transactions on Circuit Theory*, vol. 7, no. 4, pp. 520–527, 1960.

[75] P. D. Lax, *Functional analysis*, ser. Pure and applied mathematics. Wiley, 2002. [Online]. Available: https://books.google.com/books?id=-jbvAAAAMAAJ

[76] J. L. Doob, *Stochastic Processes*. Wiley, 1953.

[77] T. E. Harris, "Some mathematical models for branching processes," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, Calif.: University of California Press, 1951, pp. 305–328. [Online]. Available: https://projecteuclid.org/euclid.bsmsp/1200500236

[78] ——, "Branching processes," *Ann. Math. Statist.*, vol. 19, no. 4, pp. 474–494, 12 1948. [Online]. Available: https://doi.org/10.1214/aoms/1177730146

[79] ——, "A theorem on general branching processes," *AMS*, vol. 6, no. 36, p. 55, 2 1959.

[80] B. Barak, S. B. Hopkins, J. A. Kelner, P. Kothari, A. Moitra, and A. Potechin, "A nearly tight sum-of-squares lower bound for the planted clique problem," in *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS*, 2016, pp. 428–437.

[81] E. Abbe, "Community detection and stochastic block models: Recent developments," *J. Mach. Learn. Res.*, vol. 18, pp. 177:1–177:86, 2017.

[82] C. Moore, "The computer science and physics of community detection: Landscapes, phase transitions, and hardness," *Bulletin of the EATCS*, vol. 121, 2017.

[83] D. Gamarnik, "Linear phase transition in random linear constraint satisfaction problems," *Probability Theory and Related Fields*, vol. 129, no. 3, pp. 410–440, Jul 2004.

[84] D. Gamarnik, T. Nowicki, and G. Swirszcz, "Maximum weight independent sets and matchings in sparse random graphs. exact results using the local weak convergence method," *Random Structures & Algorithms*, vol. 28, no. 1, pp. 76–106, 2006. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20072

[85] ——, "Invariant probability measures and dynamics of exponential linear type maps," *Ergodic Theory and Dynamical Systems*, vol. 28, no. 5, pp. 1479–1495, 2008.

[86] R. M. Dudley, *Real Analysis and Probability*, 2nd ed., ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002.

[87] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed., ser. Probability and Its Applications. Springer-Verlag New York, 2002.

[88] J. Salez and D. Shah, "Belief propagation: An asymptotically optimal algorithm for the random assignment problem," *Mathematics of Operations Research*, vol. 34, no. 2, pp. 468–480, 2009. [Online]. Available: http://www.jstor.org/stable/40538393

[89] M. Khandwawala and R. Sundaresan, "Belief propagation for optimal edge cover in the random complete graph," *Ann. Appl. Probab.*, vol. 24, no. 6, pp. 2414–2454, 12 2014. [Online]. Available: https://doi.org/10.1214/13-AAP981

[90] J. M. Kurtzberg, "On approximation methods for the assignment problem," *J. ACM*, vol. 9, no. 4, pp. 419–439, oct 1962. [Online]. Available: http://doi.acm.org/10.1145/321138.321140

[91] D. W. Walkup, "On the expected value of a random assignment problem," *SIAM Journal on Computing*, vol. 8, no. 3, pp. 440–442, 1979. [Online]. Available: https://doi.org/10.1137/0208036

[92] R. M. Karp, "An upper bound on the expected cost of an optimal assignment," in *Discrete Algorithms and Complexity*, D. S. Johnson, T. Nishizeki, A. Nozaki, and H. S. Wilf, Eds. Academic Press, 1987, pp. 1 – 4. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B978012386870150006X

[93] A. J. Lazarus, "Certain expected values in the random assignment problem," *Operations Research Letters*, vol. 14, no. 4, pp. 207 – 214, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/016763779390071N

[94] M. X. Goemans and M. S. Kodialam, "A lower bound on the expected cost of an optimal assignment," *Mathematics of Operations Research*, vol. 18, no. 2, pp. 267–274, 1993. [Online]. Available: https://doi.org/10.1287/moor.18.2.267

[95] C. Nair, B. Prabhakar, and M. Sharma, "Proofs of the Parisi and Coppersmith-Sorkin conjectures for the finite random assignment problem," in *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, Oct 2003, pp. 168–178.

[96] S. Linusson and J. Wästlund, "A proof of Parisi's conjecture on the random assignment problem," *Probability Theory and Related Fields*, vol. 128, no. 3, pp. 419–440, Mar 2004. [Online]. Available: https://doi.org/10.1007/s00440-003-0308-9

[97] J. Wästlund, "An easy proof of the $\zeta(2)$ limit in the random assignment problem," *Electron. Commun. Probab.*, vol. 14, pp. 261–269, 2009. [Online]. Available: https://doi.org/10.1214/ECP.v14-1475

[98] A. Bandyopadhyay, "Bivariate uniqueness and endogeny for the logistic recursive distributional equation," *Technical Report #629, Department of Statistics, UC Berkeley, arXiv preprint arXiv:math/0401389*, 2002.

[99] J. H. Curtiss, "A note on the theory of moment generating functions," *Ann. Math. Statist.*, vol. 13, no. 4, pp. 430–433, 1942.

[100] M. I. Goldstein, "Critical age-dependent branching processes: Single and multitype," *Probability Theory and Related Fields*, vol. 17, no. 1, pp. 74–88, Mar 1971.

[101] A. Joffe and F. Spitzer, "On multitype branching processes with $\rho \leq 1$," *Journal of Mathematical Analysis and Applications*, vol. 19, no. 3, pp. 409 – 430, 1967.

[102] T. W. Mullikin, "Limiting distributions for critical multitype branching processes with discrete time," *Transactions of the American Mathematical Society*, vol. 106, no. 3, pp. 469–494, 1963.

[103] B. Pittel, J. Spencer, and N. Wormald, "Sudden emergence of a giant k-core in a random graph," *Journal of Combinatorial Theory, Series B*, vol. 67, no. 1, pp. 111–151, 1996.

[104] D. Aldous and A. G. Percus, "Scaling and universality in continuous length combinatorial optimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 20, pp. 11 211–11 215, 2003. [Online]. Available: https://www.pnas.org/content/100/20/11211

[105] D. J. Aldous, C. Bordenave, and M. Lelarge, "Near-minimal spanning trees: A scaling exponent in probability models," *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 44, no. 5, pp. 962–976, 10 2008. [Online]. Available: https://doi.org/10.1214/07-AIHP138

[106] ——, "Dynamic programming optimization over random data: The scaling exponent for near-optimal solutions," *SIAM J. Comput.*, vol. 38, no. 6, p. 2382–2410, mar 2009. [Online]. Available: https://doi.org/10.1137/070709037

[107] L. S. Pontryagin, *Ordinary Differential Equations*. Addison-Wesley, 1962, translated from the Russian by Leonas Kacinskas and Walter B. Counts.