

Social Network Formation and Strategic Interaction in Large Networks*

Euncheol Shin[†]

Job Market Paper

Recent Version: <http://people.hss.caltech.edu/~eshin/pdf/DSNF.pdf>

November 8, 2015

Abstract

I present a dynamic network formation model that aims to explain why some empirical degree distributions exhibit the *increasing hazard rate property* (IHRP). In my model, a sequentially arriving node forms a link with one existing node through a bilateral agreement. A newborn node prefers a highly linked node; however, the more links an existing node has, the more the marginal return from an additional link diminishes. I prove that the IHRP emerges if and only if the latter effect prevails over the former. I present two implications of the IHRP for strategic interactions in networks. First, when there is uncertainty about neighboring agents' connectivity, the IHRP guarantees that a unique Bayesian equilibrium exists in a network game with strategic complementarities. Second, the IHRP characterizes a monotone revenue-maximizing mechanism with allocative externalities.

Keywords: Degree distribution; Increasing hazard rate property; Mechanism design; Network game; Random network

*I am indebted to Federico Echenique, Matthew Elliott, Matthew Shum, and Leeat Yariv for their invaluable guidance and encouragement. I have also benefited from conversations with Hyoungh Jun Ahn, Marina Agranov, Yeon-Koo Che, Ben Gillen, Jong-Hee Hahn, Yinghua He, Jinwoo Kim, Fuhito Kojima, Matthew Kovach, Nicolas Lambert, John Ledyard, SangMok Lee, Brendan Lucier, Mallesh Pai, Thomas Palfrey, Kirill Pogorelskiy, Jean-Laurent Rosenthal, Larry Samuelson, Robert Sherman, Li Song, Omer Tamuz, Adam Wierman, and Seungil You.

[†]Division of the Humanities and Social Sciences (MC 228-77), California Institute of Technology, Pasadena, CA, 91125. <http://www.hss.caltech.edu/~eshin>. E-mail: eshin@caltech.edu.

1 Introduction

1.1 Overview

People are linked together through social relationships, and these relationships influence their economic decisions. The growing literature on *network games* analyzes various economic settings such as contagion behavior, criminal activity, political alliances, pricing of network goods, public good provision, and so forth. In many contexts, researchers analyze games in *large* networks, i.e., networks consisting of a large number of agents and their relationships. Since equilibrium outcomes depend on certain properties of the underlying network, it is important to identify key properties of large networks and understand how the social network formation process generates those properties.

A fundamental characteristic that represents connectivity of a large network is its *degree distribution*. The value of a degree distribution at integer d is the proportion of nodes having d links (notated as *degree d*). In this paper, I highlight one crucial property of the degree distribution of large networks that has been overlooked: whether the degree distribution satisfies the *increasing hazard rate property* (henceforth, IHRP). The value of the *hazard rate function* of a degree distribution at d is the conditional probability that a randomly selected node has *exactly* d links given that it has *at least* d links. The IHRP indicates that the hazard rate function is increasing in d .

The literature on dynamic network formation, in which newborn nodes form links with existing nodes, offers possible explanations for various properties observed in real large networks.¹ Most of the models in this literature tend to generate only the degree distributions that have decreasing hazard rates. For instance, the preferential attachment (PA) model by Barabási and Albert (1999) and the network-based search model by Jackson and Rogers (2007a) produce *strictly decreasing* hazard rate functions regardless of the model parameters.

¹For example, the small-world property with high clustering and short-average path lengths (Jackson and Rogers, 2007a), nestedness (König et al., 2014), and the scale-free property (Barabási and Albert, 1999) are supported by this literature.

However, empirical degree distributions exhibit both increasing and decreasing patterns of hazard rates. For example, Figure 1 plots the empirical hazard rate functions for four network datasets:² (a) one social network of a rural Indian village,³ (b) a collaboration network of jazz musicians, (c) an online friendship network of Facebook users, and (d) the network of the webpages at Notre Dame University.⁴ The hazard rate functions exhibit increasing patterns for (a) and (b), but decreasing patterns for (c) and (d).⁵

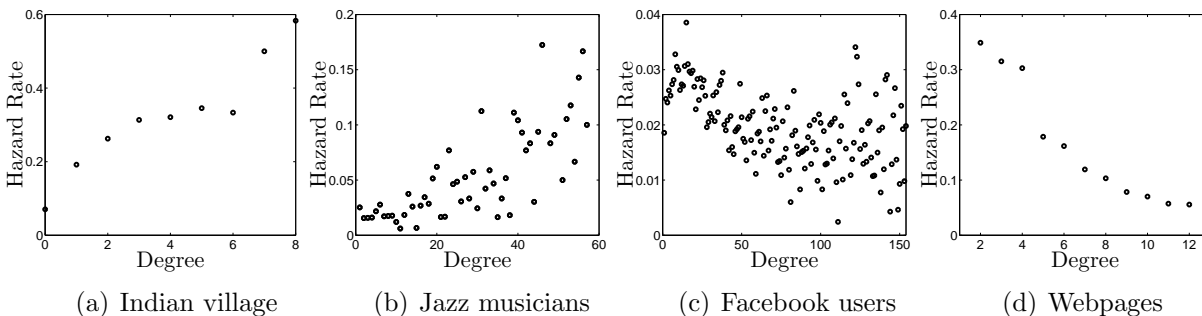


Figure 1: Different patterns of hazard rate functions

To understand the logic behind different patterns of the hazard rate function, I consider a dynamic network formation model formed by bilateral and costly link formations. Nodes arrive sequentially. Upon arrival, a new node randomly finds a single existing node with a probability that is proportional to the degree of the existing node. Once an existing node is

²In a network dataset, one can identify the hazard rate at integer d as the number of nodes with d links divided by the number of nodes having at least d links. By its definition, the degree hazard rate at the largest degree is always one for any finite network dataset. This constraint makes the hazard rate function tend to increase around the largest degree. Thus, in Figure 1, I plot the hazard rate function only for degrees that account for the degrees of 95% of nodes.

³Source: <http://web.stanford.edu/~jacksonm/Data.html>. The whole dataset consists of the social networks of 75 rural Indian villages, and the hazard function in Figure 1 corresponds to the 58th village as an illustrating example. For the 75 villages as a whole dataset, I observe that increasing hazard rates are observed at more than 75% of all points. The 75 social networks are generated as follows. For each village, to obtain relevant social relationships, the authors construct 12 individual-level adjacency matrices. The adjacency matrices are based on questions such as “Name 4 non-relatives to whom you speak to the most,” and “In your free time, whose house do you visit?” Then, they create a unified adjacency matrix by taking the union of the 12 adjacency matrices. That is, in the unified social network, nodes i and j are connected by a link if either i or j identified the other’s name at least once in the questions. Finally, the authors reduce the adjacency matrix to a household-level matrix and delete the self-loops. Therefore, the unified adjacency matrix is symmetric with each entry in $\{0, 1\}$.

⁴Sources: <http://deim.urv.cat/~alexandre.arenas/data/welcome.htm> is for the collaboration network of jazz musicians; <https://snap.stanford.edu/data/egonets-Facebook.html> is for the friendship network of Facebook users; and <https://www.aeaweb.org/articles.php?doi=10.1257/aer.97.3.890> is for the network of webpages. Detailed description about these datasets are available in the online appendix.

⁵More empirical implications of these datasets are provided in the online appendix.

identified, it decides whether to form a link with the new node. Since the marginal benefit from one additional link is decreasing but link formation is costly, the identified node is less likely to agree to form a link as its degree increases. As such, the probability that an existing node forms one additional link is determined by the combination of the new node's desire to form a link with a highly connected node, and the existing node's decreasing marginal utility from one additional link. When the diminishing marginal utility from additional links is substantial, a node will be less likely to form additional links as its degree increases.

I prove that the IHRP emerges if and only if a node is *less* likely to form additional links as its degree *increases* (Proposition 2). This characterization directly explains why previous models are not able to produce degree distributions having the IHRP. The previous models mostly consider *unilateral* link formations: existing nodes never reject any link formation offers by newly joining nodes. Since newborn nodes are more willing to form links with the existing nodes having more links, a node is more likely to form additional links as its degree increases. This is exactly the condition for the decreasing hazard rate property of the resulting degree distribution.

There are many theoretical implications of the IHRP for modeling network games. I consider an incomplete information setting in which agents are not aware of the exact structure of the underlying network, but know its degree distribution. I employ the *degree independence* assumption as a way to simplify uncertainty about neighboring agents' connectivity (e.g., Fainmesser and Galeotti, 2015; Feri and Pin, 2015; Galeotti et al., 2010; Ghiglino, 2012; Jackson and Yariv, 2007; Shin, 2015). Specifically, under this assumption, agents believe that their shared links are independently and randomly chosen from the underlying network. Because of independence, the only private information that remains for the agents is their degree. Therefore, the *type* distribution of the agents is the degree distribution.

I explore two particular theoretical implications of the IHRP. First, I consider a network game in which agents interact with neighboring agents. There are *strategic complementarities* between linked agents: an agent's incentive to perform an action increases in her neighboring

agents' actions. For example, the individual cost of engaging in criminal activity becomes lower as more criminal friends engage in the same criminal activity; or the value of using a computer software becomes higher as more acquaintances use the same software. I show that as long as the second moment of the degree distribution is finite, a Bayesian equilibrium exists even when the action space is unbounded (Proposition 5).

The IHRP guarantees that *all* moments of the degree distribution are finite (Proposition 4), thereby an equilibrium exists. However, degree distributions generated by prominent dynamic network formation models have an infinite second moment, and so no equilibrium exists. To see why, note first that taking a high action is always desirable for the agents who have an enormous number of links because of strategic complementarities. The IHRP implies that the probability that an agent is linked to such highly linked agents is very small. As such, although agents' actions feed back into one another, their best response dynamics converges even if the action space is unbounded. However, for many prominent degree distributions, the probability that an agent is linked to very highly linked agents will be substantial, and so agent's best response dynamics diverges.

Second, I study a revenue-maximizing Bayesian incentive compatible mechanism design problem. I consider an environment in which there is a single seller who produces divisible objects at zero production cost. There are *allocative externalities* between linked buyers: each buyer's valuation of her allocation depends on allocations of neighboring buyers. This environment is relevant for many settings such as a monopolistic telecommunications company that provides data plan services. The better data plans friends have, the higher valuation a customer obtains. Therefore, the company has to investigate how its sales to individual customers generate positive network externalities to their neighbors. For a given mechanism (a pair of allocation rule and price scheme), the induced network game with the IHRP provides a tractable framework where the seller can take into account the amount of network externalities generated by the equilibrium behavior of buyers.

I characterize a revenue-maximizing mechanism, assuming the IHRP of the degree distri-

bution (Proposition 6). The allocation rule of an optimal mechanism maximizes the *virtual value*, which is a multiplication of the usual (individual) virtual value and the *social value*. The social value calculates the amount of network externalities in the equilibrium of the game induced by the optimal mechanism. Thus, different from a canonical mechanism design problem (Myerson, 1981), the allocation rule that maximizes the individual value is not necessarily optimal. By increasing allocations to every customer, the seller can increase the social value. Since increase of the social value raises the virtual value of the consumers, the seller can charge a higher price to every buyer, and it ultimately returns a higher revenue to the seller. Although a closed-form solution of the optimal mechanism is not generally obtainable, I fully characterize the optimal mechanism in a restricted environment where the seller cannot price discriminate (Proposition 7).

1.2 Related Literature

There is a large and growing literature on dynamic network formation models. In these models, new nodes are born over time and form links to existing nodes. The seminal model is the PA model by Barabási and Albert (1999), which attempts to explain the *scale-free* property of degree distributions.⁶ There has been a variety of extensions of the PA model (e.g., Cooper and Frieze, 2003; Dorogovtsev and Mendes, 2001; Krapivsky et al., 2000). Jackson (2010) and references therein explain network properties that emerge from those models. To the best of my knowledge, my model is the first dynamic network formation model that identifies a condition that produces an IHRP for the degree distribution.⁷

In terms of the modeling approach, the current paper takes the *rate equation* approach introduced by Bollobás et al. (2001). They formalize the dynamic network formation process generated by the PA model, and prove that the resulting degree distribution sequence

⁶A degree distribution is said to have the scale-free property if it has a functional form of $f(d) = cd^{-\gamma}$ where c is a normalization factor.

⁷Some random network formation models can generate the IHRP. Examples are the Poisson random network model by Erdős and Rényi (1959) and the small-world model by Watts and Strogatz (1998). Although the resulting degree distributions by these two models always generate the IHRP, none of these models explain why this property can emerge.

converges. In the current paper, I prove the convergence of the degree distribution sequence in a more general setting. I find a closed-form expression of the limiting degree distribution that provides a condition under which the limiting degree distribution satisfies the IHRP.

Dynamic network formation models are largely mechanical in that there are few reasons why links are formed according to their descriptions. My paper provides a micro-foundation in which agents optimize their link formation decisions (e.g., Baetz, 2015; Currarini et al., 2009; Ghiglino, 2012; Jackson and Rogers, 2007a; König et al., 2014). Ghiglino (2012) and König et al. (2014) are two notable dynamic and strategic network formation models using linear utility functions. They assume that the largest eigenvalue of the relevant network is bounded regardless of the network size. However, my model finds that as the network size becomes large, the largest eigenvalue of an undirected network diverges almost surely to infinity if the condition for the IHRP is not satisfied.⁸

There have been many related papers on strategic interaction in networks that adopt the incomplete information setting introduced by Galeotti et al. (2010). Shin (2015) is a very closely related paper. In that paper, I study optimal dynamic pricing of a subscription network good sold by a monopolist. Each consumer's value of the good increases as more of her friends use the good, and consumers need to pay a subscription price in each period. By assuming the IHRP of the degree distribution, I characterize a unique equilibrium in which the monopolist does not change the subscription price. In the current paper, I examine a similar problem in a static setting where the monopolist can price discriminate consumers according to their number of friends.

The current paper is also related to the literature on network games with strategic complementarities. Galeotti et al. (2010) study a more general framework than my model in that they allow correlations in the degrees of agents' neighbors. In the current paper, by assuming degree independence, I obtain a clear characterization of a unique Bayesian equilibrium, and find its relation to the IHRP of the degree distribution. Belhaj et al. (2014) examine network

⁸Since the number of nodes is fixed in König et al. (2014), the authors identify an upper bound of the largest eigenvalue in a footnote. Ghiglino (2012) avoids this problem by considering a directed network.

games with strategic complementarities when agents have complete information about the underlying network. However, the current paper and Galeotti et al. (2010) analyze network games of incomplete information.

One important application of the IHRP is on the mechanism design theory. Myerson (1981) considers a problem where an auctioneer wants to sell a single object to one of many buyers. The types of buyers are their valuations of the object. Assuming the IHRP of the type distribution, he characterizes the seller's optimal mechanism. Jehiel et al. (1996) study a mechanism design problem with allocative externalities as in the current paper. In particular, they consider a two-dimensional type space: each buyer's type is a pair of her value of the object and the externalities that she generates to the other agents. I examine an environment in which buyers' types are their degrees, and a buyer's valuation of her allocation is endogenously determined by her neighboring buyers' allocations. Because of the endogenously determined externalities, my characterization of an optimal mechanism is different from a canonical solution in Myerson (1981).

2 Dynamic Network Formation

In this section, I introduce terminology and establish a model of dynamic network formation. Then, I derive the rate equations, which are essential to analyze the resulting degree distribution sequence in the next section.

2.1 Setup

Terminology. A *network* is represented by $\mathbf{G} = \langle N, A \rangle$, where $N = \{1, \dots, n\}$ is a set of *nodes*, and A is the *adjacency matrix*, an $n \times n$ symmetric matrix with each entry in $\{0, 1\}$. $A_{ij} = 1$ indicates that nodes i and j are connected by a *link*. For a given network \mathbf{G} , $N_i(\mathbf{G}) := \{j \in N | A_{ij} = 1\}$ is the set of *neighbors* of node i . $d_i(\mathbf{G}) := |N_i(\mathbf{G})|$ is called the *degree* of node i . The *degree distribution* is a function $f(\cdot, \mathbf{G}) : \mathbb{N} \rightarrow [0, 1]$ with $\sum_{d=0}^{\infty} f(d, \mathbf{G}) = 1$, in which $f(d, \mathbf{G})$ represents the fraction of nodes with degree d . $F(\cdot, \mathbf{G})$ is the corresponding

cumulative degree distribution defined as $F(d, \mathbf{G}) := \sum_{d' \leq d} f(d', \mathbf{G})$. Last, $\bar{F}(\cdot, \mathbf{G})$ denotes the *complementary cumulative degree distribution* defined as $\bar{F}(d, \mathbf{G}) := \sum_{d' \geq d} f(d', \mathbf{G})$.

Dynamic network formation. I build a model of dynamic network formation process by recursively defining a random sequence of networks denoted by $(\mathbf{G}^t)_{t \geq 1}$. Nodes arrive sequentially, and only one node joins the existing network in each period t . $N^t = \{1, \dots, t\}$ represents the set of nodes that have emerged by period t . As such, t also denotes the size of the network in period t .

To make the process well-defined, I focus on formation of random networks after $t \geq 2$ with the initial conditions

$$A^1 = \begin{bmatrix} 0 \end{bmatrix} \quad \text{and} \quad A^2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

where A^1 represents a network of one node without any link, and A^2 expresses a network of two nodes sharing one link.⁹ As will be explained in the following sections, results of the current paper are independent of these initial conditions.

For a given network \mathbf{G}^t , a network \mathbf{G}^{t+1} is randomly formed by adding one new node $t + 1$ together with one link between node $t + 1$ and node $i \in N^t$. Upon arrival, node $t + 1$ randomly identifies a single existing node with a probability that is proportional to the degree of the existing node. I call this step *preferential search*. Formally, node $t + 1$ finds node i with probability

$$\frac{d_i(\mathbf{G}^t)}{\sum_{j=1}^t d_j(\mathbf{G}^t)}.$$

Once node i is identified by the new node $t + 1$, node i probabilistically agrees to form a link with node $t + 1$. The probability of forming a link decreases as its degree increases. I call this step *constrained match*. Formally, node i agrees to form a link with probability $\Phi(d_i(\mathbf{G}^t))$ where $\Phi : \mathbb{N} \rightarrow (0, 1]$ is a decreasing function. If node i rejects node $t + 1$'s link formation offer, node $t + 1$ independently and randomly repeats the two steps until it forms one link

⁹The dynamic network formation process is well-defined with these initial conditions in the sense that for every period $t \geq 2$, each node has at least one link, so that every node has a positive probability of being found by new nodes after their arrival.

with an existing node successfully.¹⁰ Since trials are independent, the probability that node $t + 1$ forms a link with node i is

$$\frac{d_i(\mathbf{G}^t)\Phi(d_i(\mathbf{G}^t))}{\sum_{j=1}^t d_j(\mathbf{G}^t)\Phi(d_j(\mathbf{G}^t))}.$$

One interpretation of the above two-step process is as follows. Consider the evolution of a collaboration network in which nodes represent researchers, and links denote experiences of collaborations between them. Establishing a new collaborative relationship is clearly bilateral and costly. A researcher’s productivity increases as she has more collaborators because she exchanges new ideas, receives more comments about her ongoing projects, obtains other indirect benefits from her collaborators’ colleagues, etc. When a junior researcher tries to build a new collaborative relationship, he is more likely to find distinguished researchers who have many existing collaborators. Thus, more collaborations will make a researcher more likely to attract junior researchers. However, for a senior researcher, the marginal utility from having one additional relationship is decreasing due to constraints such as limited time and energy as she has more existing collaborators. Therefore, more collaborations will make a researcher reject collaboration offers more frequently.

A degree-dependent utility function provides a micro-foundation for the current model. Suppose that new nodes find existing nodes according to the preferential search step. Consider myopic link formation decisions in which existing nodes look at only the marginal utility from one additional link. Let the marginal utility of a node with degree d be

$$w(d) - c - \eta,$$

where $w(d)$ is the marginal value of forming one additional link, $c > 0$ is the marginal cost of forming one additional link, and η is a random factor distributed over the real numbers with full support. Assuming a decreasing marginal return of additional links is

¹⁰By repeating the two-step process, node $t + 1$ forms one link within a finite number of trials almost surely. To see this, one can consider node $t + 1$ ’s trials as a Bernoulli process $(\mathbf{X}_1, \dots, \mathbf{X}_s)$, where each entry represents a Bernoulli trial, and s represents the first time that a success is achieved. Since trials are independent and identical, the variables are independently and identically distributed with a strictly positive probability of success. Hence, the process ends in a finite length almost surely.

tantamount to $w(d)$ decreasing in d . Hence, the probability of accepting a link formation offer is $\Phi(d) = \mathbb{P}[\eta \leq w(d) - c]$, and it is clearly decreasing in d .

2.2 The Rate Equation

Following a standard approach in the literature (e.g., [Bollobás et al., 2001](#); [Dorogovtsev et al., 2000](#); [Ghigino, 2012](#)), I derive the *rate equation* for each degree d , which describes the dynamics of the expected number of nodes with degree d .

I write \mathcal{G}^t for the probability space of undirected networks in which a random network \mathbf{G}^t has its distribution. Let $(\mathcal{F}^t)_{t \geq 1}$ be the σ -field generated by the dynamic network formation process. For a given network \mathbf{G}^t , I define two random variables $\mathbf{N}(d, t)$ and $\mathbf{M}(t)$ as

$$\begin{aligned}\mathbf{N}(d, t) &:= \sum_{j=1}^t \mathbb{1}\{d_j(\mathbf{G}^t) = d\}, \\ \mathbf{M}(t) &:= \sum_{d=1}^t d\Phi(d)\mathbf{N}(d, t).\end{aligned}$$

$\mathbf{N}(d, t)$ is the number of nodes with degree d , and $\mathbf{M}(t)$ is a weighted sum of $(\mathbf{N}(d, t))_{d \geq 1}$.

With the above notation, for a given network \mathbf{G}^t , I express changes in the conditional expectations of $\mathbf{N}(d, t)$ from t to $t + 1$ by

$$\begin{aligned}\mathbb{E} [\mathbf{N}(d, t + 1) - \mathbf{N}(d, t) | \mathbf{G}^t] \\ = \underbrace{\mathbb{1}\{d = 1\}}_{(a)} - \underbrace{\frac{d\Phi(d)}{\mathbf{M}(t)}\mathbf{N}(d, t)}_{(b)} + \underbrace{\frac{(d-1)\Phi(d-1)}{\mathbf{M}(t)}\mathbf{N}(d-1, t)\mathbb{1}\{d \geq 2\}}_{(c)}.\end{aligned}\quad (2.1)$$

Each term in equation (2.1) represents the following:

- (a) The degree of a new node is always 1. Thus, one additional node with degree 1 emerges.
- (b) If the new node in period $t + 1$ attaches to a node with degree d , its degree becomes $d + 1$. Consequently, the number of nodes of degree d decreases by 1, but the number of nodes of degree $d + 1$ increases by 1. The probability of this event is

$$\frac{d\Phi(d)\mathbf{N}(d, t)}{\sum_{d=1}^t d\Phi(d)\mathbf{N}(d, t)} = \frac{d\Phi(d)}{\mathbf{M}(t)}\mathbf{N}(d, t).$$

(c) If the new node in period $t + 1$ attaches to a node with degree $d - 1$, its degree turns into d . Consequently, the number of nodes of degree d increases by 1, but the number of nodes of degree $d - 1$ decreases by 1. The probability of this event is

$$\frac{(d-1)\Phi(d-1)\mathbf{N}(d-1,t)}{\sum_{d=1}^t d\Phi(d)\mathbf{N}(d,t)} = \frac{(d-1)\Phi(d-1)}{\mathbf{M}(t)}\mathbf{N}(d-1,t).$$

Equation (2.1) is not linear with respect to $\mathbf{N}(d,t)$ and $\mathbf{N}(d-1,t)$ because $\mathbf{M}(t)$ appears in the denominators of the second and third terms. This is an obstacle for characterizing the asymptotic degree distribution by using the rate equation approach.

To make my analysis tractable, I introduce a technical assumption that enables me to consider *linear* rate equations. Before introducing the assumption, note that $\mathbf{M}(t)$ can be written as $\mathbf{m}(t)t$ by setting $\mathbf{m}(t) := \sum_{d=1}^t d\Phi(d)\frac{\mathbf{N}(d,t)}{t} \in [\frac{\Phi(1)}{2}, 2]$. I assume that $\mathbf{m}(t)$ converges in probability to a constant:¹¹

Assumption 1 $\mathbf{m}(t)$ converges in probability to $\mu \in [\frac{\Phi(1)}{2}, 2]$.

Assumption 1 enables me to consider linear rate equations with correction terms as

$$\text{For } d = 1: \quad \mathbb{E}[\mathbf{N}(1, t + 1)] = 1 + \left(1 - \frac{\Phi(1)}{\mu t}\right) \mathbb{E}[\mathbf{N}(1, t)] + \varepsilon(1, t), \quad (2.2)$$

$$\begin{aligned} \text{For } d \geq 2: \quad \mathbb{E}[\mathbf{N}(d, t + 1)] &= \left(1 - \frac{d\Phi(d)}{\mu t}\right) \mathbb{E}[\mathbf{N}(d, t)] + \frac{(d-1)\Phi(d-1)}{\mu t} \mathbb{E}[\mathbf{N}(d-1, t)] \\ &+ \varepsilon(d, t), \end{aligned} \quad (2.3)$$

where the correction term $\varepsilon(d, t)$ converges to zero as the network size t becomes large.¹² Moreover, as shown in the proof of Proposition 1, I can ignore the correction terms regardless of their convergence rates.

The previous dynamic network formation models using the rate equation approach make Assumption 1 implicitly (e.g., Bollobás et al., 2001; Dorogovtsev et al., 2000). For instance, in the PA model with $\Phi(d) = 1$, $\mathbf{M}(t)$ only counts the number of links in the network at the

¹¹When $d\Phi(d)$ is bounded, it suffices to assume that the expectation of $\mathbf{m}(t)$ converges to μ . A proof is available upon request.

¹²See Appendix B for a proof.

end of period t . $\mathbf{m}(t)$ is a deterministic sequence as $\mathbf{m}(t) = 2 - 1/t$, and [Assumption 1](#) is trivially satisfied.

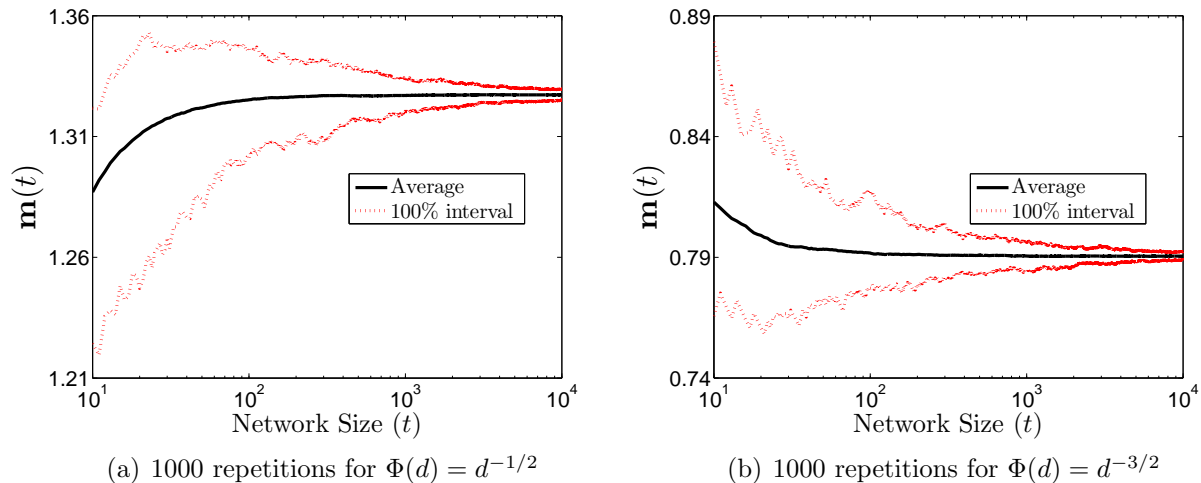


Figure 2: Simulation results for parametric examples of $\Phi(\cdot)$

To evaluate the validity of [Assumption 1](#), I present numerical simulation results in [Figure 2](#). In each figure, the horizontal axis is in logarithmic scale. For $\Phi(d) = d^{-1/2}$, [Figure 2\(a\)](#) illustrates that $\mathbf{m}(t)$ converges in probability to a constant. The average path of $\mathbf{m}(t)$ (the solid black line) is generated from 1000 repetitions, and it converges to a constant as the network size t increases. The size of the 100% interval (the dotted red lines) clearly shrinks to zero as the network size t increases. These observations indicate that $\mathbf{m}(t)$ converges in probability. In [Figure 2\(b\)](#), the same observations follow for another functional form $\Phi(d) = d^{-3/2}$. These simulations illustrate that [Assumption 1](#) reasonably holds for these parametric examples of $\Phi(\cdot)$.¹³

3 Results

In this section, I define the asymptotic degree distribution and find its closed-form expression. Then, I characterize a sufficient and necessary condition for the IHRP of the resulting asymptotic degree distribution.

¹³As will be shown in the next section, when $\Phi(d) = d^{-\alpha}$ with $\alpha \geq 0$, the hazard rate function is strictly increasing if $\alpha > 1$, strictly decreasing if $\alpha < 1$, and constant if $\alpha = 1$.

3.1 Characterization of the Asymptotic Degree Distribution

I define $f(\cdot, t) : \mathbb{N} \rightarrow [0, 1]$ the degree distribution at the end of period t by

$$f(\cdot, t) := \left(\frac{\mathbf{N}(1, t)}{t}, \dots, \frac{\mathbf{N}(d, t)}{t}, \dots \right).$$

$f(d, t)$ represents a probability that one randomly selected node at the end of period t has d links. $(f(\cdot, t))_{t \geq 1}$ is the sequence of degree distributions. I define the asymptotic degree distribution $f(\cdot) : \mathbb{N} \rightarrow [0, 1]$ as the pointwise limit of $(f(\cdot, t))_{t \geq 1}$: for a fixed $\varepsilon > 0$,

$$\lim_{t \rightarrow \infty} \mathbb{P}(|f(d, t) - f(d)| > \varepsilon) = 0 \quad \text{for all } d \in \mathbb{N}.$$

I define the asymptotic degree distribution as the pointwise limit because it ensures that the degree distribution sequence *converges in distribution* to the asymptotic degree distribution. To see this, I first clarify the notion of convergence in distribution in the current setup. Convergence in distribution means that $f(\cdot, t)$ and $f(\cdot)$ are approximately the same when the network size t is large. Since the degree of a randomly selected node is an integer, it is natural to consider a probability density function $f(\cdot) : \mathbb{N} \rightarrow [0, 1]$ as a limit of the degree distribution sequence. Thus, both $f(\cdot)$ and $(f(\cdot, t))_{t \geq 1}$ are defined over the set of integers, and it implies that as the network size t becomes infinitely large, the degree distribution sequence converges in distribution to $f(\cdot)$ if and only if $f(d, t)$ *converges in probability* to $f(d)$ for all d .¹⁴ Therefore, the asymptotic degree distribution is equivalently identified as a pointwise limit of the sequence of degree distributions.

The following presents a closed-form expression of the asymptotic degree distribution:

Proposition 1 *As the network size becomes infinitely large, the degree distribution sequence converges in distribution to $f(\cdot)$, which is defined recursively as*

$$f(1) = \frac{\mu}{\mu + \Phi(1)} \quad \text{and} \quad f(d) = \frac{(d-1)\Phi(d-1)}{\mu + d\Phi(d)} f(d-1) \quad \text{for } d \geq 2.$$

Since the rate equation approach is relatively new in the economics literature, I explain details of the proof of [Proposition 1](#). The proof consists of two parts.¹⁵ First, by using

¹⁴See [Appendix B](#) for a proof.

¹⁵Stationarity of the asymptotic degree distribution defined in [Proposition 1](#) is obvious because it is a

rate equations (2.2) and (2.3), I prove that the expected proportion of nodes with degree d converges to $f(d)$ as the network size increases to infinity. Second, for each d , I show that the difference between the random proportion of nodes with degree d and its expectation converges in probability to zero. These two observations will provide that the degree distribution sequence converges in probability to the asymptotic degree distribution. For expositional simplicity, I consider the linear rate equations, ignoring correction terms.

For $d = 1$, the iterations of rate equation (2.2) provides that

$$\begin{aligned}\mathbb{E}[\mathbf{N}(1, t+1)] &= 1 + \left(1 - \frac{\Phi(1)}{\mu t}\right) + \left(1 - \frac{\Phi(1)}{\mu t}\right) \left(1 - \frac{\Phi(1)}{\mu(t-1)}\right) \mathbb{E}[\mathbf{N}(1, t-1)] \\ &= \sum_{s=1}^t \prod_{r=s+1}^t \left(1 - \frac{\Phi(1)}{\mu r}\right) + \left\{ \prod_{r=1}^t \left(1 - \frac{\Phi(1)}{\mu r}\right) \right\} \mathbb{E}[\mathbf{N}(1, 1)].\end{aligned}$$

For large t , the expected number of nodes with degree 1 in period $t+1$ is approximated as¹⁶

$$\mathbb{E}[\mathbf{N}(1, t+1)] \approx \frac{1}{t^{\Phi(1)/\mu}} \int_0^t s^{\Phi(1)/\mu} ds = \frac{\mu t}{\mu + \Phi(1)}.$$

By dividing $\mathbb{E}[\mathbf{N}(1, t+1)]$ by $t+1$, I find the limit of the expected fraction of nodes with degree 1 as $f(1) = \frac{\mu}{\mu + \Phi(1)}$.

Second, for $d \geq 2$, the expected number of nodes with degree d in period $t+1$ relies on $\mathbb{E}[\mathbf{N}(d-1, t)]$ as well as $\mathbb{E}[\mathbf{N}(d, t)]$:

$$\mathbb{E}[\mathbf{N}(d, t+1)] = \left(1 - \frac{d\Phi(d)}{\mu t}\right) \mathbb{E}[\mathbf{N}(d, t)] + \frac{(d-1)\Phi(d-1)}{\mu t} \mathbb{E}[\mathbf{N}(d-1, t)].$$

[Assumption 1](#) enables me to replace $\mathbb{E}[\mathbf{N}(d-1, t)]$ by $tf(d-1)$ for sufficiently large t . Hence, by following a similar procedure for $d = 1$, I identify the limit of the expected fraction of

unique solution of the following stationarity equations:

$$\begin{aligned}\text{For } d = 1: \quad & f(d) = 1 - \frac{d\Phi(d)}{\mu} f(d), \\ \text{For } d \geq 2: \quad & f(d) = -\frac{d\Phi(d)}{\mu} f(d) + \frac{(d-1)\Phi(d-1)}{\mu} f(d-1).\end{aligned}$$

¹⁶I here use the following approximation:

$$\prod_{r=s+1}^t \left(1 - \frac{\Phi(1)}{\mu r}\right) \approx e^{-\frac{\Phi(1)}{\mu} \sum_{r=s+1}^t \frac{1}{r}} \approx \left(\frac{s}{t}\right)^{\frac{\Phi(1)}{\mu}}.$$

Since this product converges to zero as t becomes large, the second term in the previous equation is ignored.

nodes with degree d as it appears in [Proposition 1](#).

The remaining step is to show that as a random variable, the proportion of nodes with degree d is very close to its expectation when the network size is large. This step is proven by applying the *Azuma-Hoeffding* inequality ([Azuma, 1967](#); [Hoeffding, 1963](#)). The Azuma-Hoeffding inequality states that the number of nodes with degree d is located around its expectation within a bounded range. That is, for a fixed d , there exists a constant $M_d > 0$ such that for any $\varepsilon_d > 0$,

$$\mathbb{P}\left(\left|\mathbf{N}(d, t) - \mathbb{E}[\mathbf{N}(d, t)]\right| \geq \varepsilon_d\right) \leq 2e^{-\frac{\varepsilon_d^2}{2M_d^2 t}}.$$

By choosing $\varepsilon_d = 2M_d\sqrt{t \log t}$, it follows that the probability that the proportion of nodes with degree d is different from its expectation becomes arbitrarily small as the network size t becomes infinitely large:

$$\mathbb{P}\left(\left|\frac{\mathbf{N}(d, t)}{t} - \mathbb{E}\left[\frac{\mathbf{N}(d, t)}{t}\right]\right| \geq \frac{2M_d\sqrt{t \log t}}{t}\right) \leq o(1).$$

In order to finalize that the random proportion of nodes with degree d converges in probability to $f(d)$, I still need to show that $\mathbb{E}\left[\frac{\mathbf{N}(d, t)}{t}\right]$ is quite close to $f(d)$ for large t . In fact, [Assumption 1](#) provides that for a given $\varepsilon > 0$, $|\mathbb{E}\left[\frac{\mathbf{N}(d, t)}{t}\right] - f(d)| < \frac{\varepsilon}{3}$ whenever the network size t is larger than some constant T_ε . Thus, $t \geq T_\varepsilon$ implies that

$$\begin{aligned} \mathbb{P}\left(\left|\frac{\mathbf{N}(d, t)}{t} - f(d)\right| \geq \varepsilon\right) &\leq \mathbb{P}\left(\left|\frac{\mathbf{N}(d, t)}{t} - \mathbb{E}\left[\frac{\mathbf{N}(d, t)}{t}\right]\right| + \left|\mathbb{E}\left[\frac{\mathbf{N}(d, t)}{t}\right] - f(d)\right| \geq \varepsilon\right) \\ &\leq \mathbb{P}\left(\left|\frac{\mathbf{N}(d, t)}{t} - \mathbb{E}\left[\frac{\mathbf{N}(d, t)}{t}\right]\right| \geq \frac{2}{3}\varepsilon\right). \end{aligned}$$

Therefore, since the last term converges to zero as the network size becomes infinitely large, the random proportion of nodes with degree d converges in probability to $f(d)$.

Finally, I state that there exists a unique choice of μ for [Assumption 1](#). Due to the convergence of the degree distribution sequence, μ satisfies $\mu = \lim_{t \rightarrow \infty} \mathbb{E}[\sum_{d=1}^{\infty} d\Phi(d)f(d, t)]$.

This observation in turn implies that

$$1 = \frac{1}{\mu} \lim_{t \rightarrow \infty} \mathbb{E}\left[\sum_{d=1}^{\infty} d\Phi(d)f(d, t)\right] = \frac{1}{\mu} \mathbb{E}\left[\sum_{d=1}^{\infty} d\Phi(d)f(d)\right] = \sum_{d=1}^{\infty} \prod_{k=1}^d \left(1 + \frac{\mu}{k\Phi(k)}\right)^{-1}.$$

The last expression is continuous and strictly decreasing in μ .¹⁷ Moreover, it diverges to infinity as $\mu \rightarrow 0$, but it converges to zero as $\mu \rightarrow \infty$. Therefore, the choice of μ satisfying the above equation is unique.

3.2 The Hazard Rate Function

I characterize a condition under which the hazard rate function of the asymptotic degree distribution is increasing. Recall that the hazard rate function is defined as $h(d) := \frac{f(d)}{\bar{F}(d)}$, in which $\bar{F}(d)$ is the complementary cumulative degree distribution. This definition suggests an interpretation that the value of the hazard rate function at d is a conditional probability: $h(d)$ is the probability that a randomly selected node has exactly d links, given that it has at least d links. To characterize a condition for $h(d)$ to increase in d , I first relate the expression of the complementary cumulative degree distribution to the hazard rate function. The hazard rate at d can be written as $h(d) = 1 - \frac{\bar{F}(d+1)}{\bar{F}(d)}$, and it implies that

$$\bar{F}(d) = \frac{\bar{F}(d)}{\bar{F}(d-1)} \frac{\bar{F}(d-1)}{\bar{F}(d-2)} \dots \frac{\bar{F}(3)}{\bar{F}(2)} \frac{\bar{F}(2)}{\bar{F}(1)} = \prod_{k=1}^{d-1} (1 - h(k)).$$

Since $f(d) = \bar{F}(d) - \bar{F}(d+1)$, it follows that

$$f(d) = h(d) \prod_{k=1}^{d-1} (1 - h(k)).$$

Recall that the asymptotic degree distribution has the following recursive formula:

$$f(d) = \frac{\mu}{\mu + d\Phi(d)} \prod_{k=1}^{d-1} \frac{k\Phi(k)}{\mu + k\Phi(k)}.$$

Since the hazard rate function is uniquely defined for the asymptotic degree distribution, it directly follows that the hazard rate function is simply expressed as

$$h(d) = \frac{\mu}{\mu + d\Phi(d)}.$$

Therefore, I characterize the increase of the hazard rates from d to $d+1$ as follows.¹⁸

¹⁷See [Appendix B](#) for a proof of continuity.

¹⁸The value of the hazard rate function at d is sometimes defined as $h(d) := \frac{f(d)}{(1-F(d))}$ for a discrete probability distribution. The characterization of the IHRP by [Proposition 2](#) is still valid for this alternative definition as $h(d) = \frac{\mu}{d\Phi(d)}$.

Proposition 2 $h(d) \leq h(d+1)$ if and only if $d\Phi(d) \geq (d+1)\Phi(d+1)$.

Proposition 2 provides a natural interpretation of the IHRP in terms of the dynamic network formation process. Let $d_i(t)$ be the degree of node i at the end of period t . By Assumption 1, the logarithm of the probability that node i forms a link with the new node entering in period $t+1$ is approximated by

$$\log \left(d_i(\mathbf{G}^t)\Phi(d_i(\mathbf{G}^t))/\mathbf{M}(t) \right) \approx \log \left(d_i(\mathbf{G}^t)\Phi(d_i(\mathbf{G}^t)) \right) - \log \left(\mu t \right).$$

The first term on the right-hand side explains how the probability of forming one additional link depends on node i 's degree. Therefore, Proposition 2 provides a dynamic interpretation that the IHRP emerges if and only if a node is less likely to form additional links with newly entering nodes as its degree increases.

The IHRP is difficult to observe in network datasets where link formation decisions are unilateral. In many contexts of growing networks such as collaborations between scholars and links between webpages, new nodes are more willing to link to the more popular nodes. Hence, if the link formation decision is unilateral, more links will cause a node to form more new links. For example, in the PA model, a link formation decision is clearly unilateral because a webpage freely creates a link from itself to an existing webpage. As a result, nodes are always more likely to form additional links as their degree increases, and so the degree distribution generated by the PA model satisfies the *decreasing* hazard rate property.

In my model, however, link formation decisions are bilateral. This feature separates the new node's desire to form a link with a node having many links (preferential search) and the limitation of existing nodes to form additional links with newly entering nodes (constrained match). The constrained match step may cause links to make a node form fewer new links. The hazard rate function is increasing if this limitation is so strict that it nullifies the new nodes' desire in the preferential search step. Therefore, one can expect the IHRP in a network dataset where maintaining links is very costly and link formations are bilateral.

A parametric example considered in the previous section illustrates the above discussion.

Let $\Phi(d) = d^{-\alpha}$ be the probability that an existing node agrees to form a link when it is identified by a new node. $\alpha \geq 0$ is the parameter that measures the cost of forming links. The hazard rate function is strictly increasing in d if and only if $\alpha > 1$. The knife-edge case of this parametric example is $\alpha = 1$, which corresponds to the random attachment model in which link formation does not depend on the degree of nodes.

4 Relations to Other Properties of Large Networks

In this section, I compare the IHRP to other properties of a large network: (i) the size of its largest eigenvalue and (ii) heavy-tailedness of its degree distribution. I first introduce these characteristics and explain why they have received much attention in the literature. Then, I identify relations between these two properties and the IHRP under the assumption that the hazard rate function is monotonically either increasing or decreasing.

4.1 Definitions

The largest eigenvalue. The largest eigenvalue of network $\mathbf{G} = \langle N, A \rangle$ is defined as the largest eigenvalue of its adjacency matrix A , denoted by $\lambda_{\max}(\mathbf{G})$. Since I focus on undirected networks, and therefore on symmetric A , $\lambda_{\max}(\mathbf{G})$ is a positive real number.¹⁹

The largest eigenvalue of a network has many implications for strategic interactions in a network. In particular, equilibrium conditions of network games often depend on the size of the largest eigenvalue (e.g., [Ballester et al., 2006](#); [Bramoullé et al., 2014](#)). For example, consider the network game in [Ballester et al. \(2006\)](#), where each node i takes a positive action $x_i \in \mathbb{R}_+$ and obtains the payoff

$$u_i(x_i, \mathbf{x}_{-i}) = x_i - \frac{1}{2}x_i^2 + \delta \sum_{j=1}^n A_{ij}x_ix_j.$$

Suppose that δ is strictly positive, which means that actions are strategic complements. As [Ballester et al. \(2006\)](#) show, a Nash equilibrium exists if and only if $\delta\lambda_{\max}(\mathbf{G}) < 1$.

¹⁹By the Perron-Frobenius theorem, all eigenvalues of a symmetric adjacency matrix are real numbers. Since all the diagonal entries are zero, the trace of the adjacency matrix is zero. The trace equals the sum of all eigenvalues, and it follows that largest eigenvalue of the adjacency matrix is strictly positive.

To see why this condition is necessary, note that node i 's best response with respect to other nodes' action profile \mathbf{x}_{-i} is linear as

$$x_i^B(\mathbf{x}_{-i}) = 1 + \delta \sum_{j=1}^n A_{ij} x_j.$$

Let \mathbf{x}' be an eigenvector corresponding to the largest eigenvalue. The vector of nodes' myopic best responses to \mathbf{x}' is

$$\mathbf{x}^B(\mathbf{x}') = 1 + \delta A \mathbf{x}' = 1 + \delta \lambda_{\max}(\mathbf{G}) \mathbf{x}'.$$

The myopic best reply dynamics constructed by repeating the above steps converges if and only if the summation of externalities, $\sum_{s=0}^{\infty} (\delta \lambda_{\max}(\mathbf{G}) \mathbf{x}')^s$, converges (Ballester et al., 2006). Obviously, the summation converges if and only if $\delta \lambda_{\max}(\mathbf{G}) < 1$, which means that the maximum marginal influence of nodes' actions on other nodes is bounded.

The above restriction on the size of the largest eigenvalue is also required for some strategic dynamic network formation models (e.g., Ghiglino, 2012; König et al., 2014). For example, Ghiglino (2012) tries to explain the scale-free property of the productivity distribution. He assumes that the productivity of an idea (node) depends on its *parental* and *offspring* ideas.²⁰ Only one idea is newly created in each period. Since a new idea inherits its parental idea's productivity, it attempts to form a link to an old idea with many offspring ideas. Specifically, the productivity of idea i when used in knowledge creation in period t is

$$x_i^t = \theta + \delta \sum_{j \in N_i} \theta A_{ij}^t x_j^t,$$

where $\delta > 0$, and $\theta \sim N(1, \sigma_\theta)$ with $\sigma_\theta \ll 1$.²¹ Let \mathbf{x}^t be the $t \times 1$ vectors with entries x_i^t . Then, \mathbf{x}^t satisfies

$$\mathbf{x}^t = (I - \delta \theta A^t)^{-1} \theta \mathbf{1}_{t \times 1} = \sum_{s=0}^{\infty} (\delta \theta A^t)^s \theta \mathbf{1}_{t \times 1}. \quad (4.1)$$

A new idea strategically forms a link to an old idea with the highest productivity. This network formation process requires the productivity of ideas to be finite for all periods;

²⁰Node i is called a parental (offspring) node of node j if there is a link from j to i (from i to j).

²¹I here simplify the notation in Ghiglino (2012). In addition, note that the adjacency matrix A^t is not necessarily symmetric because the author considers a directed network.

otherwise, the productivity of an idea will be infinite after some period, and only this idea will have offsprings beyond that time. Therefore, \mathbf{x}^t in equation (4.1) has to be finite with probability one for all period t . This condition is satisfied if and only if $\delta\lambda_{\max}(\mathbf{G}^t) < 1$ with probability one for all $t \geq 1$ because $\sigma_\theta \ll 1$.²²

Heavy-tailed degree distribution. In the Poisson random network model by Erdős and Rényi (1959), a link between two nodes is formed independently of other pairs of nodes with a fixed probability. The resulting degree distribution is approximated by the Poisson distribution if the network size is infinitely large.²³ The Poisson distribution with parameter λ has the form of $f(d; \lambda) = \frac{\lambda^d e^{-\lambda}}{d!}$, and its tail decreases at an exponential rate.

One important observation in real large networks is that their degree distributions are *heavy-tailed*: there tend to be more nodes with very large degrees than the Poisson distribution of any parameter.²⁴ Thus, researchers have been interested in building dynamic network formation models that generate heavy-tailed degree distributions (e.g., Barabási and Albert, 1999; Ghigliano, 2012; Jackson and Rogers, 2007a; König et al., 2014). The following definition formalizes the heavy-tailedness of a degree distribution:

Definition 1 A degree distribution $f(\cdot)$ is said to be **heavy-tailed** if for all $\varepsilon > 0$,

$$\limsup_{d \rightarrow \infty} \frac{\bar{F}(d)}{e^{-\varepsilon d}} = \infty.$$

4.2 Relations and Implication

I find the relations between three properties of an infinitely large network: (i) finiteness of the largest eigenvalue, (ii) heavy-tailedness of its degree distribution, and (iii) the IHRP of its degree distribution.

²²König et al. (2014) consider a dynamic network formation model with a finite number of agents. For any network between n agents, its largest eigenvalue has an upper bound of $\sqrt{2m(n-1)/n}$ where $m = \binom{n}{2}$ (Cvetković and Rowlinson, 1990). Thus, they explicitly assume that the parameter representing the magnitude of positive externalities between linked agents is strictly smaller than $1/\sqrt{2m(n-1)/n}$.

²³The Poisson random network is represented by $\mathbf{G}(n, p(n))$ such that there are n nodes, and each pair of nodes forms a link independently at random with probability $p(n)$. The resulting degree distribution is a binomial distribution, and it converges to the Poisson distribution with parameter $np(n)$ as $n \rightarrow \infty$, assuming that $np(n)$ is a constant.

²⁴See Chapter 3 in Jackson (2010) for examples and discussions.

I first present lower and upper bounds of the largest eigenvalue of a network that will be useful for illustrating the relationships between the three properties under consideration. For any finite network \mathbf{G} , its largest eigenvalue $\lambda_{\max}(\mathbf{G})$ satisfies

$$\sqrt{d_{\max}(\mathbf{G})} \leq \lambda_{\max}(\mathbf{G}) \leq d_{\max}(\mathbf{G}),$$

where $d_{\max}(\mathbf{G})$ is the largest degree of the network (Cvetković and Rowlinson, 1990).²⁵ This observation suggests that the limiting behavior of the largest eigenvalue is closely related to the limiting behavior of the maximum degree.

I now find a relation between the finiteness of the largest eigenvalue and the IHRP. In the current model, the hazard rate function of the asymptotic degree distribution is decreasing if and only if a node is more likely to form additional links as its degree increases. Suppose this, and consider the evolution of node i 's degree. Given a network \mathbf{G}^t , the probability that node i forms one additional link with the new node entering in period $t + 1$ is at least $\frac{\Phi(1)}{2t}$:

$$\mathbb{P}\left(\left\{\text{node } i \text{ forms a link with node } t + 1\right\}\right) = \frac{d_i(\mathbf{G}^t)\Phi(d_i(\mathbf{G}^t))}{\sum_{j=1}^t d_j(\mathbf{G}^t)\Phi(d_j(\mathbf{G}^t))} \geq \frac{\Phi(1)}{2t}.$$

Thus, the growth of node i 's degree is faster than its growth when new nodes independently and randomly form a link with probability $\frac{\Phi(1)}{2t}$. When this is the case, the probability that node i forms infinitely many links becomes one by the second Borel-Cantelli lemma (Durrett, 2005).²⁶ Therefore, the degree of node i becomes infinitely large as the network size increases under the condition of decreasing hazard rates.

Proposition 3 *If the hazard rate function is decreasing, then the largest eigenvalue diverges almost surely to infinity as the network size becomes infinitely large.*

I now identify a relation between the heavy-tailedness of the asymptotic degree distribution and its IHRP. From the definition of the hazard rate function, I can express the

²⁵A proof is provided in [Appendix B](#).

²⁶The second Borel-Cantelli lemma states that for a given set of independent events, say $\{E_n\}_{n=1}^{\infty}$, if $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty$, then

$$\mathbb{P}(E_n \text{ i.o.}) = 1.$$

A proof is provided in [Appendix B](#).

complementary degree distribution as

$$\bar{F}(d) = \prod_{k=1}^{d-1} (1 - h(k)) = \prod_{k=1}^{d-1} \left(\frac{k\Phi(k)}{\mu + k\Phi(k)} \right).$$

Suppose the IHRP, and so $\frac{d\Phi(d)}{\mu + d\Phi(d)}$ is decreasing in d . Then, since $\frac{\Phi(d)}{\mu + \Phi(d)} \leq \frac{\Phi(1)}{\mu + \Phi(1)}$ for all d , the value of the complementary degree distribution $\bar{F}(d)$ decreases at least at a geometric rate of $\frac{\Phi(1)}{\mu + \Phi(1)} < 1$. Therefore, the asymptotic degree distribution is *not* heavy-tailed if the hazard rate function is increasing.

The asymptotic degree distribution can be heavy-tailed even when its hazard rate function strictly decreases. In particular, the asymptotic degree distribution is heavy-tailed only if its hazard rate function not only decreases, but also converges to zero. In the current model, the hazard rate function converges to zero whenever $d\Phi(d)$ diverges to infinity without any bound as d becomes infinitely large.

In many models, however, the strictly decreasing hazard rate property of the asymptotic degree distribution coincides with its heavy-tailedness. For example, consider the parametric example of $\Phi(d) = d^{-\alpha}$ in the current model. $d\Phi(d)$ becomes infinitely large as the network size increases whenever $\alpha < 1$, which is the condition for the strictly decreasing hazard rate function. Thus, the asymptotic degree distribution is heavy-tailed if and only if the hazard rate function is strictly decreasing. Indeed, in other network formation models such as Barabási and Albert (1999) and Jackson and Rogers (2007a), the resulting asymptotic degree distributions are heavy-tailed and satisfy the strictly decreasing hazard rate property simultaneously. The following proposition summarizes this point.

Proposition 4 *If the hazard rate function is increasing, then the asymptotic degree distribution is not heavy-tailed. If the hazard rate function is strictly decreasing and converges to zero as the degree becomes infinitely large, then the asymptotic degree distribution is heavy-tailed.*

Negative implication. By Proposition 3 and Proposition 4, it follows that if a dynamic network formation model generates a heavy-tailed degree distribution, then the largest eigenvalue of the network becomes infinitely large as the network size increases.

Corollary 1 *If the asymptotic degree distribution is heavy-tailed, then*

$$\lim_{t \rightarrow \infty} \mathbb{P} \left(\lambda_{\max}(\mathbf{G}^t) < \infty \right) = 0.$$

Corollary 1 proves that by using a standard utility function in the literature, it is *impossible* to build a bilateral dynamic network formation model that generates a heavy-tailed degree distribution. In a bilateral dynamic network formation model, the resulting asymptotic degree distribution is heavy-tailed only if a node is more likely to form new links as its degree increases. **Corollary 1** shows that this condition implies that largest eigenvalue becomes arbitrarily large as the network size becomes large. Thus, the value of forming a link to a particular node becomes arbitrarily large beyond a certain period. As such, if nodes can strategically choose a node to link, all nodes entering after that period will choose a particular node. Therefore, when it comes to bilateral link formations, the strategic link formation and a heavy-tailed degree distribution are incompatible.

5 Application I: Network Games

In this section, I present how the IHRP helps to characterize equilibria in network games. I adopt the incomplete information setting introduced by [Galeotti et al. \(2010\)](#), in which agents are not aware of the exact structure of the underlying network, but know its degree distribution. I characterize a unique Bayesian equilibrium, and explain how it is related to the IHRP.

5.1 Network Games with Incomplete Information

Network and utilities. There is a countable set of agents, $N = \{1, \dots, n\}$. Connections between agents are represented by a network $\mathbf{G} = \langle N, A \rangle$, in which A is a symmetric matrix of size n with each entry in $\{0, 1\}$. For notational simplicity, let N_i be the set of agent i 's neighbors. $f(\cdot)$ is the degree distribution of the underlying network, which is common knowledge amongst the agents.

Each agent i simultaneously takes an action $x_i \in \mathbb{R}_+$. I denote by $\mathbf{x} = (x_i, \mathbf{x}_{-i}) \in \mathbb{R}_+^n$ the

action profile of the agents, where \mathbf{x}_{-i} is the action profile of all agents except agent i . For an action profile \mathbf{x} , the utility of agent i with degree d_i is given by

$$u_i(x_i, \mathbf{x}_{-i}, d_i) = \underbrace{x_i - \frac{1}{2}x_i^2}_{\text{idiosyncratic utility}} + \underbrace{\delta x_i \sum_{j \in N_i} x_j}_{\text{network externality}},$$

where $\delta > 0$ represents positive network externalities between agents' actions. This utility function satisfies a property in which adding a link to an agent taking action 0 generates no additional value to agent i 's utility. Note that the utility function is independent of agent i 's identity in the network in the sense that agents i and j obtain the same utility if their degrees are identical and their neighbors' actions coincide. Thus, I represent agent i 's utility by $u(x_i, \mathbf{x}_{N_i}, d_i)$ where $\mathbf{x}_{N_i} \in \mathbb{R}_+^{d_i}$ is the action profile of agent i 's neighbors.

Information. Before deciding on her action, the information available to agent i is her degree d_i and the degree distribution $f(\cdot)$. Thus, each agent can update her beliefs about the degrees of her neighbors based on her private information. To simplify this belief updating process, I employ the assumption of *degree independence*, which is quite common in the literature (e.g., Fainmesser and Galeotti, 2015; Feri and Pin, 2015; Galeotti et al., 2010; Ghiglini, 2012; Shin, 2015). The degree independence assumption states that agent i believes that the link between herself and each of her neighboring agents is an *i.i.d.* draw from a given degree distribution. Under this assumption, I denote by $\tilde{f}(\cdot) : \mathbb{N} \rightarrow [0, 1]$ the probability density function of a neighboring agent's degree, which is calculated as

$$\tilde{f}(d) = \frac{df(d)}{\langle d \rangle},$$

where $\langle d \rangle = \sum_{d=1}^{\infty} df(d)$ is the average degree.²⁷ I call $\tilde{f}(\cdot)$ the *conditional degree distribution*. The conditional degree distribution $\tilde{f}(\cdot)$ captures the idea that a highly connected agent is

²⁷To understand this calculation, suppose that each agent's degree is either one or two. Let $f(1)$ and $f(2)$ be the fraction of agents with degree one and two, respectively. Divide the set of links into two categories: (i) set L_1 containing links attached to an agent with degree one, and (ii) set L_2 containing links attached to an agent with degree two. Then, the fraction of links in set L_1 is proportional to $f(1)$ and the fraction of links in set L_2 is proportional to $2f(2)$. Due to degree independence, the probability that the degree of a randomly selected neighbor is d is equal to the probability that a randomly selected link is chosen from set L_d . Thus, after normalization, $\tilde{f}(\cdot)$ is the probability density function of a neighbor's degree.

more likely to be an agent’s neighbor: $\tilde{f}(d) > f(d)$ for all $d > \langle d \rangle$. $\tilde{F}(\cdot)$ and $\tilde{h}(\cdot)$ are the corresponding cumulative distribution function and the hazard rate function, respectively.

The degree independence assumption is plausible for large networks because the degrees of two neighboring agents are approximately independently distributed. Indeed, the *configuration model* confirms that for each agent i , knowing only agent i ’s degree provides no additional information about the degrees of her neighbors, as the number of agents becomes large (Bender and Canfield, 1978). Therefore, a neighboring agent’s degree is considered as an *i.i.d.* draw from the degree distribution, and its probability density function is $\tilde{f}(\cdot)$.²⁸

The following lemma states that the conditional degree distribution satisfies the IHRP.

Lemma 1 *If $f(\cdot)$ satisfies the increasing hazard rate property, then $\tilde{f}(\cdot)$ satisfies the strictly increasing hazard rate property.*

Partial order on degree distributions. To compare equilibria where the underlying network changes in density, I consider a family of degree distributions $\{f_\theta(\cdot)\}_{\theta \in \Theta}$ indexed by an ordered set Θ in which all members have the common support \mathbb{N} . I use the *likelihood ratio order* (Karlin and Rubin, 1956) as a partial order on $\{f_\theta(\cdot)\}_{\theta \in \Theta}$.

Definition 2 Degree distribution $f_\theta(\cdot)$ is said to stochastically dominate $f_{\theta'}(\cdot)$ according to the *likelihood ratio order* if for all $d, d' \in \mathbb{N}$ with $d > d'$,

$$\frac{f_\theta(d)}{f_{\theta'}(d)} > \frac{f_\theta(d')}{f_{\theta'}(d')}.$$

I denote this stochastic dominance order by $>_{\text{LR}}$ and assume that $f_\theta(\cdot) >_{\text{LR}} f_{\theta'}(\cdot)$ if $\theta > \theta'$.²⁹ For each degree distribution $f_\theta(\cdot)$, let $\tilde{f}_\theta(\cdot)$ be the corresponding conditional degree distribution. Similarly, I denote by $h_\theta(\cdot)$ and $\tilde{h}_\theta(\cdot)$ the corresponding hazard rate functions.

The likelihood ratio order has the following three useful properties:

²⁸However, many dynamic network formation models generate correlations between neighboring nodes. In fact, real network datasets often exhibit positive or negative neighbor affiliations (Newman, 2003).

²⁹The likelihood ratio order is not a complete ordering of an arbitrary family of degree distributions.

- The likelihood ratio order between two degree distributions is preserved for the corresponding conditional degree distributions:³⁰

$$f_\theta(\cdot) >_{\mathbf{LR}} f_{\theta'}(\cdot) \text{ implies } \tilde{f}_\theta(\cdot) >_{\mathbf{LR}} \tilde{f}_{\theta'}(\cdot).$$

- The likelihood ratio order $>_{\mathbf{LR}}$ induces the *first-order stochastic dominance* order, denoted by $>_{\mathbf{FOSD}}$ for the degree distributions and the conditional degree distributions.³¹

$$f_\theta(\cdot) >_{\mathbf{LR}} f_{\theta'}(\cdot) \text{ implies } f_\theta(\cdot) >_{\mathbf{FOSD}} f_{\theta'}(\cdot) \text{ and } \tilde{f}_\theta(\cdot) >_{\mathbf{FOSD}} \tilde{f}_{\theta'}(\cdot).$$

- The likelihood ratio order $>_{\mathbf{LR}}$ provides the *monotone hazard rate* order for the hazard rate functions of the degree distributions and the conditional degree distributions:

$$f_\theta(\cdot) >_{\mathbf{LR}} f_{\theta'}(\cdot) \text{ implies } h_\theta(d) > h_{\theta'}(d) \text{ and } \tilde{h}_\theta(d) > \tilde{h}_{\theta'}(d) \text{ for all } d.$$

As will be presented in later sections, the above properties are useful to analyze comparative statics of network game outcomes.

Strategy and equilibria. A strategy for agent i is a map $\sigma_i : \mathbb{N} \rightarrow \Delta(\mathbb{R}_+)$ where $\Delta(\mathbb{R}_+)$ is the set of probability distributions over \mathbb{R}_+ . I consider symmetric *Bayesian equilibria* (henceforth, *equilibria*). Thus, an equilibrium is represented by a strategy $\sigma(\cdot)$, and so each agent's equilibrium strategy depends only on her degree.

Given agent i with degree d_i , let $\Psi(\mathbf{x}_{N_i}, \sigma, d_i)$ be the probability distribution over $\mathbb{R}_+^{d_i}$ induced by the conditional degree distribution $\tilde{f}(\cdot)$ and strategy $\sigma(\cdot)$. When she chooses action x_i , the expected utility of agent i with degree d_i is

$$U(x_i, \sigma, d_i) = \int_{\mathbf{x}_{N_i} \in \mathbb{R}_+^{d_i}} u(x_i, \mathbf{x}_{N_i}, d_i) d\Psi(\mathbf{x}_{N_i}, \sigma, d_i)$$

³⁰The proof is as follows: for all $d > d'$,

$$\frac{\tilde{f}_\theta(d)}{\tilde{f}_{\theta'}(d)} = \frac{df_\theta(d)/\langle d \rangle_\theta}{df_{\theta'}(d)/\langle d \rangle_{\theta'}} > \frac{d' f_\theta(d')/\langle d \rangle_\theta}{d' f_{\theta'}(d')/\langle d \rangle_{\theta'}} = \frac{\tilde{f}_\theta(d')}{\tilde{f}_{\theta'}(d')},$$

where $\langle d \rangle_\theta = \sum_{d=1}^{\infty} df_\theta(d)$ and $\langle d \rangle_{\theta'} = \sum_{d=1}^{\infty} df_{\theta'}(d)$.

³¹The first-order stochastic dominance order between the degree distributions does not induce the first-order stochastic dominance order between the conditional degree distributions. To see this, consider two degree distributions, $f_1(1) = 0.30$, $f_1(2) = 0.45$, and $f_1(3) = 0.25$; $f_2(1) = 0.45$, $f_2(2) = 0.30$ and $f_2(3) = 0.25$. Then, $f_\theta(\cdot) >_{\mathbf{FOSD}} f_{\theta'}(\cdot)$, but $\tilde{f}_\theta(\cdot) \not>_{\mathbf{FOSD}} \tilde{f}_{\theta'}(\cdot)$ because $\tilde{F}_\theta(1) = 0.25 > \tilde{F}_{\theta'}(1) = 0.15$ but $\tilde{F}_\theta(2) = 0.58 < \tilde{F}_{\theta'}(2) = 0.62$.

$$= x_i - \frac{1}{2}x_i^2 + \delta d_i \tilde{\mathbb{E}}[\sigma] x_i,$$

where $\tilde{\mathbb{E}}[\sigma] := \sum_{d=1}^{\infty} \sigma(d) \tilde{f}(d)$ is the expected action of a neighboring agent. A strategy $\sigma(\cdot)$ establishes an equilibrium if $\sigma(d_i)$ is a best response for any agent with degree d_i .

Definition 3 A strategy $\sigma(\cdot)$ is an *equilibrium* if for any agent i with degree d_i ,

$$U(x_i, \sigma, d_i) \geq U(x'_i, \sigma, d_i) \text{ for all } x'_i \in \mathbb{R}_+ \text{ and } x_i \in \text{supp}(\sigma(d_i)).$$

5.2 Properties of Equilibria

The expected utility of agent i with degree d_i is $U(x_i, \sigma, d_i) = x_i - \frac{1}{2}x_i^2 + \delta d_i \tilde{\mathbb{E}}[\sigma] x_i$ for given strategy $\sigma(\cdot)$. Maximizing the expected utility $U(x_i, \sigma, d_i)$ with respect to x_i yields a linear best reply function:

$$x^B(\sigma, d_i) = 1 + \delta d_i \tilde{\mathbb{E}}[\sigma]. \quad (5.1)$$

In any equilibrium, the corresponding equilibrium strategy $\sigma^*(\cdot)$ is a best reply with respect to $\sigma^*(\cdot)$: $x^B(\sigma^*, \cdot) = \sigma^*(\cdot)$. In any equilibrium, agents' beliefs must be consistent: by taking expectations of both sides in equation (5.1), it must hold that

$$\tilde{\mathbb{E}}[\sigma^*] = 1 + \delta \langle \tilde{d} \rangle \mathbb{E}[\sigma^*] = \frac{1}{1 - \delta \langle \tilde{d} \rangle},$$

where $\langle \tilde{d} \rangle = \sum_{d=1}^{\infty} d \tilde{f}(d)$ is the expectation of a neighboring agent's degree. Thus, an equilibrium exists if and only if $\delta \langle \tilde{d} \rangle < 1$. This equilibrium condition can be further simplified as the ratio of the first and second moments of degree distribution $f(\cdot)$ as:

$$\langle \tilde{d} \rangle = \sum_{d=1}^{\infty} d \tilde{f}(d) = \sum_{d=1}^{\infty} d \left(\frac{df(d)}{\langle d \rangle} \right) = \frac{\langle d^2 \rangle}{\langle d \rangle}.$$

Therefore, the sufficient and necessary condition for existence of equilibria is $\delta \frac{\langle d^2 \rangle}{\langle d \rangle} < 1$.

To understand the intuition behind the above equilibrium condition, consider a myopic best reply dynamics such that $\sigma^1(\cdot) := 1$, and recursively define $\sigma^n(\cdot) := x^B(\sigma^{n-1}, \cdot)$ for $n \geq 2$. The initial strategy $\sigma^1(\cdot)$ corresponds to a strategy in which individuals take the minimum action. When agent i myopically best responds to $\sigma^1(\cdot)$, she assumes that her neighboring agents take action $\tilde{\mathbb{E}}[\sigma^1] = 1$. Thus, when her degree is d_i , her best response is

$\sigma^2(d_i) = 1 + \delta d_i$. Now, for the next best reply, agents optimize their actions by assuming that the other agents play strategy $\sigma^2(\cdot)$. Each agent i finds the expectation of her neighboring agent's action as $\tilde{\mathbb{E}}[\sigma^2] = 1 + \delta \frac{\langle d^2 \rangle}{\langle d \rangle}$. Hence, when her degree is d_i , agent i 's myopic best reply is $\sigma^3(d_i) = 1 + \delta d_i + \delta^2 \frac{\langle d^2 \rangle}{\langle d \rangle} d_i$. This myopic best reply dynamics continues until it converges.³² To establish that this dynamics converges, it must hold that the influence of the neighboring agent's expected actions converges as $\delta \frac{\langle d^2 \rangle}{\langle d \rangle} < 1$.

In the above dynamics, the agents with very high degrees serve as conduits for accelerating actions of other agents with low degrees. When the degree distribution is heavy-tailed, there is a sufficient number of agents with enormously many neighbors. Since the action space is unbounded, the presence of such agents will significantly increase other agents' actions. Thus, the above myopic best reply dynamic will diverge. For example, consider a network having a scale-free degree distribution. Since a scale-free degree distribution has a functional form of $f(d) = cd^{-\gamma}$ where c is a normalization factor, its second moment is infinite if and only if $\gamma \leq 3$. The scale parameter is frequently estimated to take values within the (2, 3) interval.³³ Therefore, the empirical scale-free degree distributions predict that the myopic best reply dynamics diverges.

However, the degree distribution of a network is not heavy-tailed if it satisfies the IHRP. Specifically, the IHRP provides a finite second moment, and so the myopic best reply dynamics converges as long as δ is not too large. Moreover, since the best reply function is linear, the equilibrium is uniquely exists. The following proposition summarizes these points.

Proposition 5 *If the degree distribution satisfies the increasing hazard rate property, then there exists an equilibrium if and only if $\delta \frac{\langle d^2 \rangle}{\langle d \rangle} < 1$. The equilibrium is unique if it exists.*

³²This myopic best reply dynamics is called the mean-field dynamics and frequently used in diffusion models (e.g., Jackson and Rogers, 2007b; López-Pintado, 2008; Shin, 2015). These dynamic models implicitly assume that (i) in each period, agents consider a new strategic interaction in a network, and (ii) the stochastic dynamics is represented by a deterministic dynamics. These two assumptions remarkably simplify the models, and allow researchers to compare diffusion outcomes in terms of the network structure that underlies.

³³For example, Barabási and Albert (1999) measure the scale-free parameter for the social network between movie actors. Two actors share a link if they have appeared in at least one movie together. The authors identify the scale parameter 2.3 for the degree distribution. See Chapter 1 in Durrett (2010) for more examples of the estimated scale parameters.

I finally remark on two theoretical features of the equilibrium in the current incomplete information setting, comparing to the equilibria in a complete information setting. First, the current equilibrium condition is less restrictive in that it is independent of the network size that underlies. In the network games with complete information, Nash equilibria exist if and only if $\delta\lambda_{\max}(\mathbf{G}) < 1$. However, in prominent dynamic network formation models, the largest eigenvalue diverges to infinity as the size of the network increases. For example, in the PA model, it grows at the rate of \sqrt{n} where n is the size of the network (e.g., [Chung et al., 2003](#); [Flaxman et al., 2005](#)). Moreover, as shown in the previous section, the largest eigenvalue can be very large even when the IHRP is satisfied. Thus, the equilibrium condition under complete information is more restrictive.

Second, the equilibrium under incomplete information is easier to calculate. In the network game with complete information, the equilibrium action profile is a function of the eigenvalues of the adjacency matrix ([Ballester et al., 2006](#)), which are more expensive to calculate than finding the second moment of the degree distribution. Specifically, for a given $n \times n$ adjacency matrix, the time complexity of finding the second moment is $O(n^2)$.³⁴ On the other hand, the time complexity of an algorithm that finds all eigenvalues is roughly tied to the complexity of matrix multiplications, which require at least $O(n^{2.3})$ time.³⁵

6 Application II: Mechanism Design

I study a revenue-maximizing Bayesian incentive compatible mechanism. I consider a monopolistic seller who determines allocations to buyers. Buyers are connected to one another, and a buyer's valuation of her allocation depends on her neighbors' allocations as well. The buyers know their degree but have incomplete information about the degrees of their neighboring buyers. Thus, the degree distribution is a *type* distribution of the buyers. By

³⁴The computation of a degree distribution can be done in $O(n^2)$ time. For example, the naïve algorithm, which simply iterates through each element of the adjacency matrix to count the number of neighbors each node has, achieves the bound of $O(n^2)$ to find the degree distribution. Given a degree distribution, one can easily calculate the second moment in $O(n)$ time.

³⁵For instance, [Coppersmith and Winograd \(1990\)](#) suggest an algorithm that achieves the bound of $O(n^{2.376})$. There is no known algorithm that achieves $O(n^2)$.

assuming its IHRP, I characterize an optimal mechanism.

6.1 The Model

Consider the following mechanism design problem. There are n buyers indexed by i , and $N = \{1, \dots, n\}$ is the set of buyers. There is a single seller who owns an infinite number of an object. The object is divisible and has no value for the seller. I assume that each buyer has a unit demand for the object. The buyers and seller know the degree distribution of the underlying network between the buyers.

Agent i 's *type* is her degree $d_i \in \mathcal{D} = \{0, \dots, d_{\max}\}$, and \mathcal{D}^n is the set of all type profiles. By the assumption of degree independence, buyer i 's type is drawn from the degree distribution $f(\cdot)$, and is independent of other buyers' types. I assume that each buyer's degree is her private information. The joint type distribution, except buyer i 's type, is $f_{-i}(\mathbf{d}_{-i}) = \prod_{j \neq i} f(d_j)$ where $\mathbf{d}_{-i} \in \mathcal{D}^{n-1}$ is the type profile of all buyers except buyer i 's type. Although realizations of types are independent, buyer i 's utility has *allocative externalities*. Let $\mathbf{x} = (x_i, \mathbf{x}_{-i}) \in [0, 1]^n$ be an allocation vector of the buyers, where \mathbf{x}_{-i} represents the allocation vector except buyer i 's allocation. Buyer i 's value of allocation vector \mathbf{x} is $v_i(x_i, \mathbf{x}_{-i}) = x_i \sum_{j \in N_i} x_j$. I consider buyers with a quasi-linear utility function: when buyer i pays p_i to the seller, her utility is $v_i(x_i, \mathbf{x}_{-i}) - p_i$.

By the revelation principle, I focus on *direct revelation mechanisms* (henceforth, mechanisms): buyers directly report their types, and an allocation vector and a payment vector are determined according to a pre-determined rule.³⁶ Formally, let $\mathbf{X} = [0, 1]^n$ be the set of allocation vectors. The seller specifies a direct revelation mechanism (\mathbf{x}, \mathbf{p}) , where \mathbf{x} is an allocation rule, and \mathbf{p} is a payment scheme. The allocation rule is represented by $\mathbf{x} = (x_1, \dots, x_n)$ where $x_i(\cdot) : \mathcal{D}^n \rightarrow [0, 1]$ is an allocation rule for buyer i . Similarly, the payment scheme is denoted by $\mathbf{p} = (p_1, \dots, p_n)$ where $p_i(\cdot) : \mathcal{D}^n \rightarrow \mathbb{R}_+$ is a payment scheme for buyer i . Therefore, when the reported profile is $\mathbf{d} = (d_1, \dots, d_n)$, buyer i obtains $x_i(\mathbf{d})$

³⁶The revelation principle states that an equilibrium outcome of any mechanism can be replicated by a truthful equilibrium of a direct mechanism.

unit of the object and pays $p_i(\mathbf{d})$ to the seller.

Given a mechanism (\mathbf{x}, \mathbf{p}) , I define for each buyer i the conditional expected allocation function $\xi_i(\cdot) : \mathcal{D} \rightarrow [0, 1]$ and the conditional expected payment function $\pi_i(\cdot) : \mathcal{D} \rightarrow \mathbb{R}_+$ as

$$\begin{aligned}\xi_i(d_i) &:= \sum_{\mathbf{d}_{-i} \in \mathcal{D}^{n-1}} x_i(d_i, \mathbf{d}_{-i}) f_{-i}(\mathbf{d}_{-i}), \\ \pi_i(d_i) &:= \sum_{\mathbf{d}_{-i} \in \mathcal{D}^{n-1}} p_i(d_i, \mathbf{d}_{-i}) f_{-i}(\mathbf{d}_{-i}).\end{aligned}$$

Suppose buyer i believes that other buyers report their types truthfully. When her type is d_i , buyer i 's expected valuation by reporting type d'_i is

$$V_i(d'_i, d_i) = \xi_i(d'_i) \sum_{j \in N_i} \tilde{\mathbb{E}}[\xi_j(d_j)],$$

where $\tilde{\mathbb{E}}[\xi_j(d_j)] := \sum_{d \in \mathcal{D}} \tilde{f}(d_j) \xi_j(d_j)$ is the expectation of neighbor j 's allocation.

I restrict my attention to *anonymous* mechanisms: if $d_i = d_j$, then $\xi_i(d_i) = \xi_j(d_j)$ and $\pi_i(d_i) = \pi_j(d_j)$ for all $i, j \in N$. With this restriction, since the expected valuation is represented by the expected allocations and payments, a mechanism is simply expressed by a pair of two functions (ξ, π) : if buyer i reports type d_i , she receives $\xi(d_i)$ unit of the object and pays $\pi(d_i)$. The seller's problem is to maximize his expected revenue from one buyer.

Buyer i 's expected valuation does not depend on her identity by anonymity of the mechanism. Let $V(d'_i, d_i)$ be her expected value if she reports d'_i when her type is d_i :

$$V(d'_i, d_i) = \xi(d'_i) \sum_{j \in N_i} \tilde{\mathbb{E}}[\xi(d_j)].$$

Given this, a mechanism (ξ, π) is called *incentive compatible* if

$$V(d_i, d_i) - \pi(d_i) \geq V(d'_i, d_i) - \pi(d'_i) \text{ for all } d'_i, d_i \in \mathcal{D}.$$

It is called (interim) *individually rational* if for all $i \in N$,

$$V(d_i, d_i) - \pi(d_i) \geq 0 \text{ for all } d_i \in \mathcal{D}.$$

Since the seller's per-capita expected revenue is $\sum_{d \in \mathcal{D}} f(d) \pi(d)$, his mechanism design problem is formulated as

$$\begin{aligned}
& \underset{(\xi, \pi)}{\text{maximize}} && \sum_{d \in \mathcal{D}} f(d) \pi(d) \\
& \text{subject to} && V(d, d) - \pi(d) \geq V(d', d) - \pi(d') \text{ for all } d', d \in \mathcal{D}, \\
& && V(d, d) - \pi(d) \geq 0 \text{ for all } d \in \mathcal{D}, \\
& && \xi(d) \in [0, 1] \text{ for all } d \in \mathcal{D}.
\end{aligned} \tag{6.1}$$

One interpretation of the above mechanism design problem is as follows. There is a telecommunications company that provides data plan services at zero production cost. There are n consumers each with demand for the service of up to one unit. A consumer's valuation of her data plan depends on her friends' plans as well: the better data plans her friends have, the higher value she obtains. Given this environment, the company wants to construct a list of data plans that incentivizes consumers to truthfully report their number of friends and maximizes the company's revenue. Hence, when a consumer tries to join the service, a service manager will ask how many friends the consumer has. Depending on her answer, the manager recommends a specific data plan from the list of available data plans, and the consumer will take the recommended service because the mechanism is incentive compatible.

6.2 Revenue-Maximizing Mechanism

Any mechanism for a strictly positive revenue to the seller provides $\tilde{E}[\xi(d)] > 0$. Thus, for any mechanism with positive revenue, the expected valuation $V(\cdot, \cdot)$ is strictly supermodular: for all $d > d'$ and $k > k'$,

$$V(d, k) - V(d', k) > V(d, k') - V(d', k').$$

For this strict supermodularity, a mechanism (ξ, π) is incentive compatible if and only if the allocation rule $\xi(\cdot)$ is *monotone*: $\xi(d) \geq \xi(d')$ for all $d \geq d'$. The monotonicity simplifies the incentive compatibility constraints by the adjacent incentive compatibility constraints:

$$V(d, d) - \pi(d) \geq V(d + 1, d) - \pi(d + 1) \text{ for all } d = 0, \dots, d_{\max} - 1 \tag{6.2}$$

$$V(d, d) - \pi(d) \geq V(d - 1, d) - \pi(d - 1) \text{ for all } d = 1, \dots, d_{\max}. \tag{6.3}$$

Needless to say, all the downward incentive compatibility constraints (6.3) must be binding if a mechanism (ξ, π) maximizes the seller's revenue. Since any isolated buyer with zero degrees takes no value from his allocation, I set $\pi(0) = 0$ without loss of generality. Since the downward incentive compatibility constraints are binding, the payment scheme satisfies

$$\pi(d) = \pi(d-1) + (V(d, d) - V(d-1, d)) = \tilde{\xi} \sum_{k=1}^d (\xi(k)k - \xi(k-1)k)$$

for all $d \geq 1$, where $\tilde{\xi} = \sum_{d=1}^{d_{\max}} \xi(d)f(d)$ is the expected allocation of a neighboring buyer. It follows that if $\xi(\cdot)$ is monotone, the above payment scheme provides that all the upward incentive compatibility constraints (6.2) are satisfied. Thus, the seller's problem becomes

$$\begin{aligned} & \underset{\xi: \mathcal{D} \rightarrow [0,1]}{\text{maximize}} && \left\{ \sum_{d'=1}^{d_{\max}} \tilde{f}(d')\xi(d') \right\} \left[\sum_{d=1}^{d_{\max}} f(d) \left(\xi(d) \left(d - \frac{1-F(d)}{f(d)} \right) \right) \right] \\ & \text{subject to} && 0 \leq \xi(0) \leq \xi(1) \leq \dots \leq \xi(d_{\max}) \leq 1. \end{aligned} \quad (6.4)$$

The seller's objective function can be rewritten as

$$\sum_{d=1}^{d_{\max}} f(d) \left(\tilde{\xi}\xi(d) \left(d - \frac{1-F(d)}{f(d)} \right) \right),$$

where $\tilde{\xi} = \sum_{d=1}^{d_{\max}} \xi(d)f(d)$ is the expected allocation of a neighboring buyer. The term presents in the summation, $\tilde{\xi}\xi(d) \left(d - \frac{1-F(d)}{f(d)} \right)$, is the *virtual value* of a buyer of type d . Due to the allocative externalities between linked buyers, the virtual value has two components: $\xi(d) \left(d - \frac{1-F(d)}{f(d)} \right)$ and $\tilde{\xi}$. Since the first component does not depend on other buyers' allocations, I call it the *individual value*. Note that $\left(d - \frac{1-F(d)}{f(d)} \right)$ is the analogue of Myerson's virtual type (Myerson, 1981), which is ubiquitous in mechanism design problems when types are realized independently (e.g., Jehiel et al., 1996, 1999). The second component $\tilde{\xi}$ newly appears in the current model, and I call this the *social value*. Since the social value depends on allocation rule $\xi(\cdot)$, the seller takes it into account his revenue maximization problem.

If the seller fully knows the buyers' types, he can choose the efficient allocation rule where $\xi(d) = 1$ and the payment rule $\pi(d) = d$ for all d . Since every buyer knows that the other neighboring buyers obtain one unit of the object, this allocation rule maximizes the social value $\tilde{\xi}$ as one. However, when the seller has incomplete information, he has to incentivize

the buyers to truthfully report their types.

Note that if allocative externalities do not exist, and the IHRP is satisfied, the seller can incentivize the buyers by solving the following pointwise maximization problem of the individual values:

$$\begin{aligned} & \underset{\xi: \mathcal{D} \rightarrow [0,1]}{\text{maximize}} && \sum_{d=1}^{d_{\max}} f(d) \left(\xi(d) \left(d - \frac{1 - F(d)}{f(d)} \right) \right) \\ & \text{subject to} && 0 \leq \xi(0) \leq \xi(1) \leq \dots \leq \xi(d_{\max}) \leq 1. \end{aligned} \tag{6.5}$$

The solution of the above problem maximizes $\xi(d) \left(d - \frac{1 - F(d)}{f(d)} \right)$ for each d (Myerson, 1981). Note that the virtual type $\left(d - \frac{1 - F(d)}{f(d)} \right)$ undercuts the buyers' true types, and so the resulting social value is strictly less than one. Since the seller's true objective function contains the social value, the solution of alternative problem (6.5) may not maximize the seller's true per-capita revenue in (6.4). This suggests that the seller has to balance the maximization of the social value and the maximization of the individual value. By using the example in the following section, I will clearly illustrate the tension between these two maximizations.

Comparative statics. Although it is impossible to obtain a closed-form solution of the optimal mechanism, it follows that the seller's per-capita revenue monotone increases as the density of the underlying network increases. By using the notion of likelihood ratio order, let $f_{\theta}(\cdot)$ and $f_{\theta'}(\cdot)$ be the degree distributions of two networks with $f_{\theta}(\cdot) >_{\mathbf{LR}} f_{\theta'}(\cdot)$. As shown in the previous section, $\frac{f_{\theta}(d)}{1 - F_{\theta}(d)} < \frac{f_{\theta'}(d)}{1 - F_{\theta'}(d)}$ for all d , and $\tilde{f}_{\theta}(\cdot) >_{\mathbf{FOSD}} \tilde{f}_{\theta'}(\cdot)$.

Suppose that the seller chooses the same allocation rule $\xi(\cdot)$ for both networks. Then, as the degree distribution changes from $f_{\theta'}(\cdot)$ to $f_{\theta}(\cdot)$, the individual value strictly increases as

$$d - \frac{1 - F_{\theta}(d)}{f_{\theta}(d)} > d - \frac{1 - F_{\theta'}(d)}{f_{\theta'}(d)}$$

for all d . In addition, the first-order stochastic dominance, $\tilde{f}_{\theta}(\cdot) >_{\mathbf{FOSD}} \tilde{f}_{\theta'}(\cdot)$, implies that the social value strictly increases as

$$\sum_{d=1}^{d_{\max}} \xi(d) \tilde{f}_{\theta}(d) > \sum_{d=1}^{d_{\max}} \xi(d) \tilde{f}_{\theta'}(d).$$

These two observations imply that for any given allocation rule, the seller's objective function

in (6.4) is strictly increasing as the underlying network changes its density in terms of the likelihood ratio order. Therefore, the seller's revenue strictly increases. The following proposition summarizes this idea.

Proposition 6 *If the degree distribution satisfies the increasing hazard rate property, then there exists a monotone revenue-maximizing mechanism. The seller's revenue strictly increases as the degree distribution increases in terms of likelihood ratio order.*

6.3 Example: Uniform Pricing

I here analyze the seller's optimal mechanism problem when he cannot price discriminate the buyers. The seller chooses only one price $\pi \geq 0$. Since the valuation of the object is strictly increasing in degree, any buyer with a degree higher than a threshold degree will be willing to pay the posted price π and obtain one unit of the object. Hence, in any equilibrium of the game with incomplete information among the buyers, the buyers with the threshold degree must obtain zero utility.

The above observation provides that it suffices to consider a binary allocation rule $\xi(\cdot)$ characterized by a threshold degree \underline{d} as $\xi(d) = 1$ if $d \geq \underline{d}$, and $\xi(d) = 0$ otherwise. By the incentive compatibility constraint for \underline{d} , the expected utility of the buyers with the threshold degree \underline{d} must be zero. Thus, the price π satisfies $\pi = V(\underline{d}, \underline{d}) = \underline{d}(1 - \tilde{F}(\underline{d} - 1))$, and the seller's revenue function for a given threshold degree d is

$$\Psi(d) = \underbrace{d(1 - \tilde{F}(d - 1))}_{\text{posted price}} \underbrace{(1 - F(d - 1))}_{\text{demand}}.$$

I now characterize the seller's optimal choice of the threshold degree. I note that the seller's revenue function is single-peaked if the degree distribution satisfies the IHRP. To explain why, I define the seller's marginal revenue by increasing the threshold degree d to $d + 1$ as $\Delta\Psi(d) := \Psi(d + 1) - \Psi(d)$. The marginal revenue satisfies (i) $\Delta\Psi(0) > 0$, and (ii) a single-crossing property that $\Delta\Psi(d) \leq 0$ implies $\Delta\Psi(d') < 0$ for all $d' > d$. In words, (i) represents that the seller excludes the buyers who have no neighboring buyers: otherwise,

the seller has to make the price zero to sell the object to those buyers. The single-crossing property (ii) establishes that the seller's revenue function has a unique maximizer.

Indeed, the single-crossing property follows from the fact that both the degree distribution and the conditional degree distribution simultaneously satisfy the IHRP. To demonstrate, I write the marginal revenue as

$$\Delta\Psi(d) = \frac{1}{(1 - \tilde{F}(d-1))(1 - F(d-1))} \left\{ 1 + (d+1) \left(\tilde{h}(d)h(d) - \tilde{h}(d) - h(d) \right) \right\}.$$

Since $\tilde{h}(d)h(d) - (\tilde{h}(d) + h(d))$ is strictly decreasing by [Lemma 1](#), the marginal revenue satisfies the single-crossing property. The seller's revenue is maximized at $d^* = \inf\{d | \Delta\Psi(d) \leq 0\}$. By ignoring the integer-value problem, the optimal choice of threshold degree d^* is easily characterized by setting $\Delta\Phi(d^*) = 0$:

$$\frac{1}{d^* + 1} = h(d^*) + \tilde{h}(d^*) - \tilde{h}(d^*)h(d^*). \quad (6.6)$$

This characterization directly shows the existence and uniqueness of a threshold degree: the left-hand side strictly decreases in d , but the right-hand side strictly increases in d .

The above characterization equation has the following economic interpretation. Suppose that the social value is fixed as $\tilde{\xi}$, and that the seller proposes a take-it-or-leave-it offer to a buyer at price π . Since the buyer obtains a zero utility if she has degree $d = \pi/\tilde{\xi}$, the probability that this buyer accepts the offer is $1 - F(d-1)$. Thus, the seller's revenue function in terms of threshold degree d is $\tilde{\xi}d(1 - F(d-1))$. The seller's revenue is uniquely maximized at d' satisfying $\frac{1}{d'+1} = h(d')$, and it is independent of the social value $\tilde{\xi}$. Therefore, the first term in equation (6.6) explains the optimal choice of threshold degree when he does not take into account the change of the social value by his mechanism. However, the seller has to consider the social value to maximize his revenue as explained for the general model. In equation (6.6), the latter term $(\tilde{h}(d) - \tilde{h}(d)h(d))$ captures this factor. Since $\tilde{h}(d) - \tilde{h}(d)h(d)$ is strictly positive, the threshold degree d^* solving equation (6.6) is strictly smaller than d' .

The optimal choice of threshold degree d^* as strictly smaller than d' has the following meaning. Let $\pi^* = d^*(1 - \tilde{F}(d^* - 1))$ and $\pi' = d'(1 - \tilde{F}(d' - 1))$ be the corresponding prices.

Since the degree distribution $f(\cdot)$ satisfies the strictly IHRP, it follows that $\pi^* < \pi'$.³⁷ Therefore, by setting a lower price π^* , the seller increases the demand. In turn, the lower price provides a higher return to the seller than a price that only maximizes the individual value. The following proposition summarizes the observations.

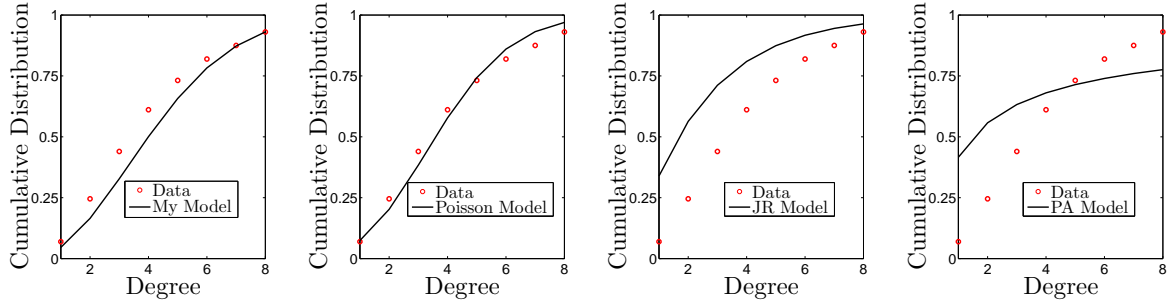
Proposition 7 *If the degree distribution satisfies the increasing hazard rate property, then there exists a unique threshold degree d^* that maximizes the seller's revenue.*

7 Concluding Remarks

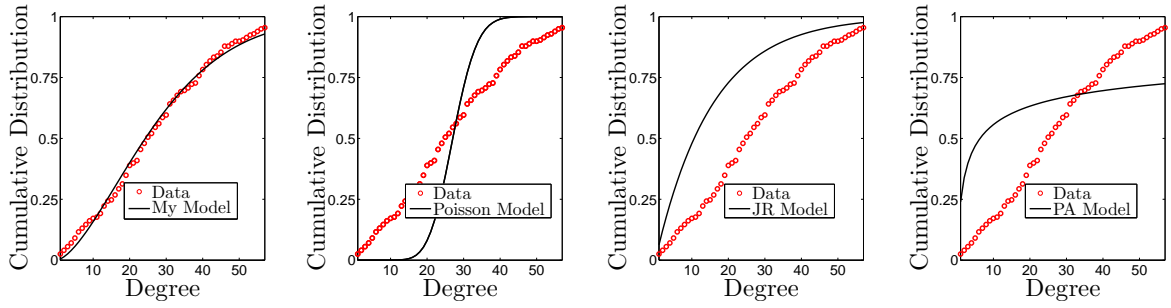
Researchers are interested in analyzing strategic interaction in large networks. In the modeling perspective, they often consider an incomplete information setting in which agents know their own connections, but have uncertainty about connectivity of their neighboring agents. In this setting, the IHRP of the degree distribution plays a key role in characterizing equilibrium outcomes. In addition to the theoretical implications of the IHRP, the current paper presents a dynamic network formation model that explains why empirical hazard rates exhibit different patterns. This network formation model with empirical observations justifies the use of the IHRP as an assumption of network games.

My model has many empirical implications, and I discuss these in more detail in the online appendix. My model outperforms other network formation models in terms of data-fitting performance. In the appendix, I compare my model to three prominent models: (i) the Poisson model by Erdős and Rényi (1959), (ii) the network-based search model by Jackson and Rogers (2007a) (JR model), and (iii) the PA model by Barabási and Albert (1999). In Figure 3, I present results only for three network datasets: (a) one social network of a rural Indian village, (b) a collaboration network between jazz musicians, (c) a friendship network of Facebook users. As illustrated in the figure, my model fits the empirical degree distributions better than the other models.

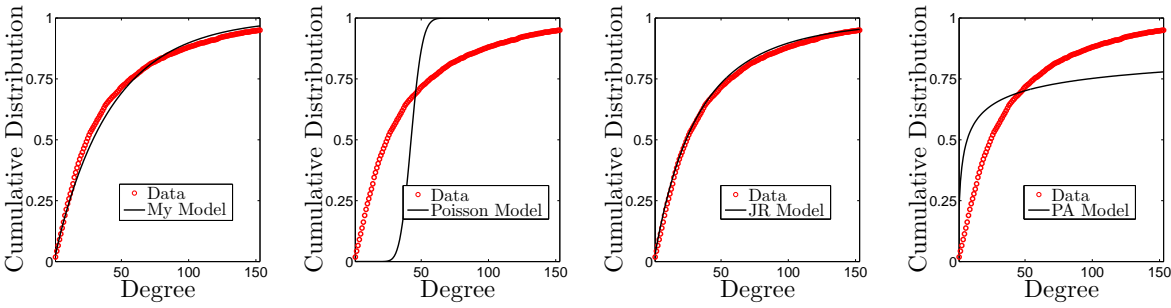
³⁷To see this, let d'' be the degree that maximizes $d(1 - \tilde{F}(d - 1))$. Since $\tilde{f}(\cdot)$ satisfies the strictly IHRP, such d'' is unique. It follows that $d' < d''$ by $h(d) > \tilde{h}(d)$ for all d . Note also that $d(1 - \tilde{F}(d - 1))$ is strictly increasing in d for $d \leq d''$. Therefore, $\pi^* < \pi'$ because $d^* < d' < d''$.



(a) Indian village



(b) Jazz musicians



(c) Facebook users

Figure 3: In each plot, the horizontal axis represents degrees, and the vertical axis represents the empirical (the red dots) and estimated (the black line) cumulative degree distributions.

One important factor should be taken into account in future research. It has been recently shown that networks having the same degree distribution may have very different network structures. Specifically, [Bubeck et al. \(2015\)](#) show that the initial network has a great impact on the limiting graph generated by the PA model. This is a surprising result because the degree distribution generated by the PA model converges almost surely to a scale-free distribution regardless of the initial network. Beyond just finding a limiting degree distribution, the dynamic network formation literature is evolving in a direction of identifying the limiting *distribution of networks*.

Therefore, in line with the literature on strategic network formation, it is definitely worth building a strategic dynamic network formation model that incorporates how agents form beliefs about the future networks for a given network, and how it affects their decision on link formation. I conjecture that this approach will generate a probability distribution over a set of multiple networks, and so it will enrich the limiting equilibrium network structure comparing to what the previous strategic network formation models predict.

A Proofs of Results

Proof of Proposition 1

Proof. The proof consists of two parts. In Part I, by defining $\bar{\mathbf{N}}(d, t) := \mathbb{E}[\mathbf{N}(d, t)]$ for each d , I show that $\bar{\mathbf{N}}(d, t)/t$ converges to $f(d)$. In Part II, I prove that for each d , $\mathbf{N}(d, t)/t$ converges in probability to $f(d)$.

Part I. I first show that the expected fraction of nodes with degree d converges.

Proposition A.1 *For each $d \in \mathbb{N}$, $\bar{\mathbf{N}}(d, t)/t$ converges to $f(d)$ as $t \rightarrow \infty$.*

Proof. I start from the following rate equations:

(i) For $d = 1$:

$$\bar{\mathbf{N}}(1, t+1) = 1 + \left(1 - \frac{\Phi(1)}{\mu t}\right) \bar{\mathbf{N}}(1, t) + \varepsilon(1, t). \quad (\text{A.1})$$

(ii) For $d \geq 2$:

$$\bar{\mathbf{N}}(d, t+1) = \frac{(d-1)\Phi(d-1)}{\mu t} \bar{\mathbf{N}}(d-1, t) + \left(1 - \frac{d\Phi(d)}{\mu t}\right) \bar{\mathbf{N}}(d, t) + \varepsilon(d, t). \quad (\text{A.2})$$

I solve the rate equations inductively. By letting $a(1) = \Phi(1)/\mu$, equation (A.1) becomes

$$\bar{\mathbf{N}}(1, t+1) = 1 + \left(1 - \frac{a(1)}{t}\right) \bar{\mathbf{N}}(1, t) + \varepsilon(1, t).$$

By iteration, I have

$$\begin{aligned} & \bar{\mathbf{N}}(1, t+1) \\ &= 1 + \left(1 - \frac{a(1)}{t}\right) + \left(1 - \frac{a(1)}{t}\right) \left(1 - \frac{a(1)}{t-1}\right) \bar{\mathbf{N}}(1, t-1) + \varepsilon(1, t) + \left(1 - \frac{a(1)}{t}\right) \varepsilon(1, t-1) \\ &= \underbrace{\sum_{s=1}^t \left[\prod_{r=s+1}^t \left(1 - \frac{a(1)}{r}\right) \right]}_{(i)} + \underbrace{\bar{\mathbf{N}}(1, 1) \prod_{r=1}^t \left(1 - \frac{a(1)}{r}\right)}_{(ii)} + \underbrace{\sum_{s=1}^t \left[\prod_{r=s+1}^t \left(1 - \frac{a(1)}{r}\right) \varepsilon(1, r-1) \right]}_{(iii)}. \end{aligned}$$

For a large t , the product is approximated by

$$\prod_{r=s+1}^t \left(1 - \frac{a(1)}{r}\right) \approx e^{-\sum_{r=s+1}^t a(1)/r} \approx e^{-a(1)(\log t - \log s)} = \left(\frac{s}{t}\right)^{a(1)}. \quad (\text{A.3})$$

By the above approximation, I find approximations for (i) - (iii) as

$$\begin{aligned}
(i) \quad & \sum_{s=1}^t \prod_{r=s+1}^t \left(1 - \frac{a(1)}{r}\right) \approx \frac{1}{t^{a(1)}} \int_0^t s^{a(1)} ds = \frac{1}{t^{a(1)}} \frac{1}{1+a(1)} t^{a(1)+1} = \frac{t}{1+a(1)}, \\
(ii) \quad & \overline{\mathbf{N}}(1, 1) \prod_{r=1}^t \left(1 - \frac{a(1)}{r}\right) \approx \overline{\mathbf{N}}(1, 1) \left(\frac{s}{t}\right)^{a(1)}, \\
(iii) \quad & \sum_{s=1}^t \prod_{r=s+1}^t \left(1 - \frac{a(1)}{r}\right) \varepsilon(1, r-1) \approx \varepsilon(1, t-1) \frac{t}{1+a(1)}.
\end{aligned}$$

By dividing by t and taking the limit, both (ii) and (iii) become zero. Hence, it follows that

$$\lim_{t \rightarrow \infty} \frac{\overline{\mathbf{N}}(1, t)}{t} = \frac{1}{1+a(1)} = \frac{\mu}{\mu + \Phi(1)} = f(1).$$

Suppose that $f(d-1)$ is given. Define $a(d)$ and $b(d-1, t)$ as

$$a(d) := \frac{d\Phi(d)}{\mu} \quad \text{and} \quad b(d-1, t) := \frac{(d-1)\Phi(d-1)\overline{\mathbf{N}}(d-1, t)}{\mu t}.$$

I observe that

$$\lim_{t \rightarrow \infty} b(d-1, t) = \frac{(d-1)\Phi(d-1)}{\mu} \lim_{t \rightarrow \infty} \frac{\overline{\mathbf{N}}(d-1, t)}{t} = \frac{(d-1)\Phi(d-1)}{\mu} f(d-1).$$

Then, by using the approximation technique for $d=1$, I rewrite equation (A.2) as

$$\begin{aligned}
\overline{\mathbf{N}}(d, t+1) &= \left(1 - \frac{a(d)}{t}\right) \overline{\mathbf{N}}(d, t) + b(d-1, t) + \varepsilon(d, t) \\
&= b(d-1, t) + \left(1 - \frac{a(d)}{t}\right) b(d, t-1) + \left(1 - \frac{a(d)}{t}\right) \left(1 - \frac{a(d)}{t-1}\right) \overline{\mathbf{N}}(d, t-1) \\
&\quad + \varepsilon(d, t) + \left(1 - \frac{a(d)}{t}\right) \varepsilon(d, t-1) \\
&\approx \sum_{s=1}^t b(d-1, s) \prod_{r=s+1}^t \left(1 - \frac{a(d)}{r}\right) + \overline{\mathbf{N}}(d, 1) \prod_{r=1}^t \left(1 - \frac{a(d)}{r}\right) \\
&\quad + \sum_{s=1}^t \prod_{r=s+1}^t \left(1 - \frac{a(d)}{r}\right) \varepsilon(d, r-1).
\end{aligned}$$

By dividing t and taking the limit, the latter two terms become zero. In addition, since

$\lim_{t \rightarrow \infty} b(d-1, t) = \frac{(d-1)\Phi(d-1)}{\mu} f(d-1)$, it follows that

$$\lim_{t \rightarrow \infty} \frac{\overline{\mathbf{N}}(d, t)}{t} = \frac{\lim_{t \rightarrow \infty} b(d-1, t)}{1+a(d)} = \frac{(d-1)\Phi(d-1)}{\mu + d\Phi(d)} f(d-1) = f(d). \quad \blacksquare$$

Part II. For notational simplicity, I let $z_d = d\Phi(d)$ for each d . I find the following result:

Proposition A.2 For each $d \in \mathbb{N}$, there exists $(\nu(d, t))_{t \geq 1}$ such that

$$|\bar{\mathbf{N}}(d, t) - f(d)t| \leq \nu(d, t)t,$$

and $\nu(d, t) \rightarrow 0$ as $t \rightarrow \infty$.

Proof. I prove the statement inductively. For $d = 1$, I observe that

$$\begin{aligned}\bar{\mathbf{N}}(1, t+1) &= 1 + \left(1 - \frac{z_1}{\mu t}\right) \bar{\mathbf{N}}(1, t) + \varepsilon(1, t), \\ (t+1)f(1) &= 1 + \left(1 - \frac{z_1}{\mu t}\right) tf(1).\end{aligned}$$

Let $\delta(1, t) := |\bar{\mathbf{N}}(1, t) - tf(1)|$. Then, by setting $\nu(1, t) = \varepsilon(1, t) + 1/t$, it follows that

$$\begin{aligned}\left| \bar{\mathbf{N}}(1, t+1) - (t+1)f(1) \right| &\leq \left| \left(\bar{\mathbf{N}}(1, t) - tf(1) \right) - \frac{z_1}{\mu t} \left(\bar{\mathbf{N}}(1, t) - tf(1) \right) \right| + \varepsilon(1, t) \\ &= \left(1 - \frac{z_1}{\mu t}\right) \delta(1, t) + \varepsilon(1, t) \\ &\leq \underbrace{\delta(1, 1) \prod_{k=1}^t \left(1 - \frac{z_1}{\mu k}\right)}_{\rightarrow 0 \text{ as } t \rightarrow \infty} + \underbrace{\sum_{s=1}^t \prod_{r=s+1}^t \left(1 - \frac{z_1}{\mu r}\right) \varepsilon(d, r)}_{\leq \varepsilon(1, t)t} \\ &\leq \nu(1, t)t.\end{aligned}$$

Suppose now that the statement holds for $d - 1$. I observe that

$$\begin{aligned}\bar{\mathbf{N}}(d, t+1) &= \bar{\mathbf{N}}(d, t) + \frac{z_{d-1}}{\mu t} \bar{\mathbf{N}}(d-1, t) - \frac{z_d}{\mu t} \bar{\mathbf{N}}(d, t) + \varepsilon(d, t), \\ (t+1)f(d) &= tf(d) + \frac{z_{d-1}}{\mu t} tf(d-1) - \frac{z_d}{\mu t} tf(d).\end{aligned}$$

Let $\delta(d, t) := |\bar{\mathbf{N}}(d, t) - tf(d)|$. By setting $\nu(d, t) = \frac{z_{d-1}}{\mu} \nu(d-1, t) + \varepsilon(d, t) + 1/t$, I have

$$\begin{aligned}\left| \bar{\mathbf{N}}(d, t+1) - (t+1)f(d) \right| &\leq \left| \left(\bar{\mathbf{N}}(d, t) - tf(d) \right) \left(1 - \frac{z_d}{\mu t}\right) + \frac{z_{d-1}}{\mu t} \left(\bar{\mathbf{N}}(d-1, t) - tf(d-1) \right) \right| + \varepsilon(d, t) \\ &\leq \left(1 - \frac{z_d}{\mu t}\right) \delta(d, t) + \frac{z_{d-1}}{\mu} \nu(d-1, t) + \varepsilon(d, t) \\ &\leq \left(1 - \frac{z_d}{\mu t}\right) \left(1 - \frac{z_d}{\mu(t-1)}\right) \delta(d, t-1) + \left(1 - \frac{z_d}{\mu t}\right) \frac{z_{d-1}}{\mu} \nu(d-1, t-1) + \frac{z_{d-1}}{\mu} \nu(d-1, t) \\ &\quad + \left(1 - \frac{z_d}{\mu t}\right) \varepsilon(d, t-1) + \varepsilon(d, t)\end{aligned}$$

$$\leq \underbrace{\delta(d, 1) \prod_{k=1}^t \left(1 - \frac{z_d}{\mu k}\right)}_{\rightarrow 0 \text{ as } t \rightarrow 0} + \underbrace{\sum_{s=1}^t \left\{ \left[\frac{z_{d-1}}{\mu} \nu(d-1, s) + \varepsilon(d, s) \right] \prod_{r=1}^t \left(1 - \frac{z_d}{\mu r}\right) \right\}}_{\leq \left(\frac{z_{d-1}}{\mu} \nu(d-1, s) + \varepsilon(d, s)\right) t}$$

$$\leq \nu(d, t)t. \quad \blacksquare$$

I now show concentration of the degrees as follows.

Proposition A.3 *Let $d \in \mathbb{N}$ be fixed. Then, there exists a constant $K_d > 0$ such that*

$$\mathbb{P} \left(\left| \mathbf{N}(d, t) - \bar{\mathbf{N}}(d, t) \right| \geq K_d \sqrt{t \log t} \right) \leq o(1).$$

Proof. I find that $(\mathbf{M}(d, s))_{s=1}^t$ where $\mathbf{M}(d, s) := \mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^s]$ is a *Doob Martingale*:

Lemma A.1 *For each $d \in \mathbb{N}$, $(\mathbf{M}(d, s))_{s=1}^t$ is a Doob Martingale with respect to $(\mathcal{F}^s)_{s=0}^t$. Moreover, $\mathbf{M}(d, 0) = \bar{\mathbf{N}}(d, t)$ and $\mathbf{M}(d, t) = \mathbf{N}(d, t)$.*

Proof. The first part follows as:

$$\begin{aligned} \mathbb{E} [|\mathbf{M}(d, s)|] &= \mathbb{E} [\mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^s]] = \mathbb{E} [\mathbf{N}(d, t)] \leq t < \infty, \\ \mathbb{E} [\mathbf{M}(d, s) | \mathcal{F}^{s-1}] &= \mathbb{E} [\mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^s] | \mathcal{F}^{s-1}] = \mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^{s-1}] = \mathbf{M}(d, s-1). \end{aligned}$$

The second part follows as:

$$\begin{aligned} \mathbf{M}(d, 0) &= \mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^0] = \mathbb{E} [\mathbf{N}(d, t)], \\ \mathbf{M}(d, t) &= \mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^t] = \mathbf{N}(d, t). \end{aligned} \quad \blacksquare$$

$(\mathbf{M}(d, s))_{s=1}^t$ has the following uniformly bounded difference property:

Lemma A.2 *For each $d \in \mathbb{N}$, there exists $M_d > 0$ such that for all $1 \leq s \leq t$,*

$$|\mathbf{M}(d, s) - \mathbf{M}(d, s-1)| \leq M_d.$$

Proof. I prove the lemma only for $d \geq 2$ because one can easily repeat similar steps for $d = 1$. For any given $d \geq 2$, I set $M_d = \frac{2}{z_1} \max\{1, z_d, z_{d-1}\}$, and show the statement by using mathematical induction in the difference $k = t - s \geq 0$. First, when $k = 0$, I have

$$|\mathbf{M}(d, t) - \mathbf{M}(d, t-1)|$$

$$\begin{aligned}
&= \left| \mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^t] - \mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^{t-1}] \right| \\
&= \left| \mathbf{N}(d, t) - \mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^{t-1}] \right| \\
&\leq \left| \mathbf{N}(d, t) - \mathbf{N}(d, t-1) \right| + \frac{z_d}{\frac{z_1}{2}(t-1)} \mathbf{N}(1, t-1) + \frac{z_{d-1}}{\frac{z_1}{2}(t-1)} \mathbf{N}(1, t-1) \\
&\leq 1 + \frac{\max\{z_{d-1}, z_d\}}{\frac{z_1}{2}(t-1)} \left(\mathbf{N}(d-1, t) + \mathbf{N}(d, t-1) \right) \\
&\leq M_d.
\end{aligned}$$

Second, suppose the statement holds for all $k' \leq k$. For $k \geq 1$, I have

$$\begin{aligned}
|\mathbf{M}(d, s) - \mathbf{M}(d, s-1)| &= \left| \mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^s] - \mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^{s-1}] \right| \\
&\leq 1 + \frac{\max\{z_{d-1}, z_d\}}{\frac{z_1}{2}(s-1)} \left(\mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^s] + \mathbb{E} [\mathbf{N}(d, t) | \mathcal{F}^{s-1}] \right) \\
&\leq M_d. \quad \blacksquare
\end{aligned}$$

With the above two lemmas, the Azuma-Hoeffding inequality states that for any $\varepsilon_d > 0$,

$$\mathbb{P} \left(|\mathbf{N}(d, t) - \bar{\mathbf{N}}(d, t)| \geq \varepsilon_d \right) \leq 2e^{-\frac{\varepsilon_d^2}{2M_d^2 t}}.$$

By choosing $\varepsilon_d = 2M_d \sqrt{t \log t}$, it follows that

$$\mathbb{P} \left(|\mathbf{N}(d, t) - \bar{\mathbf{N}}(d, t)| \geq 2M_d \sqrt{t \log t} \right) \leq o(1). \quad \blacksquare$$

I finally prove that for each d , $\frac{\mathbf{N}(d, t)}{t}$ converges in probability to $f(d)$ as follows:

Proposition A.4 *Let $d \in \mathbb{N}$ be fixed. For any given $\varepsilon > 0$,*

$$\lim_{t \rightarrow \infty} \mathbb{P} \left(\left| \frac{\mathbf{N}(d, t)}{t} - f(d) \right| \geq \varepsilon \right) = 0.$$

Proof. By [Proposition A.2](#), there exists T_1 such that $|\frac{\bar{\mathbf{N}}(d, t)}{t} - f(d)| < \varepsilon/3$ whenever $t \geq T_1$.

Then, for all $t \geq T_1$,

$$\begin{aligned}
\mathbb{P} \left(\left| \frac{\mathbf{N}(d, t)}{t} - f(d) \right| \geq \varepsilon \right) &\leq \mathbb{P} \left(\left| \frac{\mathbf{N}(d, t)}{t} - \frac{\bar{\mathbf{N}}(d, t)}{t} \right| + \left| \frac{\bar{\mathbf{N}}(d, t)}{t} - f(d) \right| \geq \varepsilon \right) \\
&\leq \mathbb{P} \left(\left| \frac{\mathbf{N}(d, t)}{t} - \frac{\bar{\mathbf{N}}(d, t)}{t} \right| \geq \frac{2}{3}\varepsilon \right).
\end{aligned}$$

By [Proposition A.3](#), it follows that

$$\lim_{t \rightarrow \infty} \mathbb{P} \left(\left| \frac{\mathbf{N}(d, t)}{t} - \frac{\bar{\mathbf{N}}(d, t)}{t} \right| \geq \frac{2}{3} \varepsilon \right) = 0. \quad \blacksquare$$

I finally prove the unique choice of μ . For this, I state and prove the following lemma.

Lemma A.3 *Let $\Gamma(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a function defined by*

$$\Gamma(\mu) := \sum_{d=1}^{\infty} d\Phi(d)f(d) = \sum_{d=1}^{\infty} d\Phi(d) \left[\frac{\mu}{d\Phi(d)} \prod_{k=1}^d \left(\frac{k\Phi(k)}{\mu + k\Phi(k)} \right) \right].$$

Then, $\Gamma(\cdot)$ is continuous in μ .

Proof. Define a function $\gamma_d(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ as

$$\gamma_d(\mu) = d\Phi(d)f(d) = d\Phi(d) \left[\frac{\mu}{d\Phi(d)} \prod_{k=1}^d \left(\frac{k\Phi(k)}{\mu + k\Phi(k)} \right) \right].$$

Note that $0 \leq \gamma_d(\mu) \leq df(d)$ and $\sum_{d=1}^{\infty} df(d) = 2$. Define a function $\Gamma_n(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ as $\Gamma_n(\mu) = \sum_{d=1}^n \gamma_d(\mu)$. Then, $\Gamma_n(\cdot)$ is continuous in μ , and it converges uniformly to $\Gamma(\cdot)$ by the Weierstrass M test ([Marsden and Hoffman, 1993](#)). Therefore, $\Gamma(\cdot)$ is continuous. \blacksquare

Therefore, the theorem is proven. \blacksquare

Proof of [Proposition 2](#)

Proof. The proposition is fully proven in the main text. \blacksquare

Proof of [Proposition 3](#)

Proof. I observe the following lower bound of $\lambda_{\max}(\mathbf{G}^t)$:

$$\sqrt{d_{\max}(\mathbf{G}^t)} \leq \lambda_{\max}(\mathbf{G}^t),$$

where $d_{\max}(\mathbf{G})$ is the maximum degree of network \mathbf{G}^t . Thus, it suffices to show that if $d\Phi(d)$ is increasing in d , then $d_1(\mathbf{G}^t) \rightarrow \infty$ almost surely as $t \rightarrow \infty$.

Define a sequence of independent Bernoulli random variables $(\mathbf{I}_t)_{t \geq 1}$ such that

$$\mathbb{P}(\mathbf{I}_t = 1) := \frac{\Phi(1)}{2t}.$$

If the hazard rate function is not increasing, then it follows that

$$\mathbb{P}(d_1(\mathbf{G}^t) - d_1(\mathbf{G}^{t-1}) = 1 | \mathbf{G}^{t-1}) = \frac{d_1(\mathbf{G}^{t-1})\Phi(d_1(\mathbf{G}^{t-1}))}{\sum_{s=1}^{t-1} d_s(\mathbf{G}^s)\Phi(d_s(\mathbf{G}^{t-1}))} \geq \frac{\Phi(1)}{2t} = \mathbb{P}(\mathbf{I}_t = 1).$$

Thus, $d_1(\mathbf{G}^t) \geq \sum_{s=1}^t \mathbf{I}_s$. Note that $\sum_{t=1}^{\infty} \mathbb{P}(\mathbf{I}_t = 1) = \infty$. Since $(\mathbf{I}_t)_{t \geq 1}$ is a sequence of independent random variables, the second Borel-Cantelli lemma shows that

$$\mathbb{P}\left(\mathbf{I}_t = 1 \text{ i.o.}\right) = 1.$$

Thus, $\mathbb{P}\left(\sum_{t=1}^{\infty} \mathbf{I}_t = \infty\right) = 1$, and it implies that $d_1(\mathbf{G}^t) \rightarrow \infty$ almost surely as $t \rightarrow \infty$. ■

Proof of Proposition 4

Proof. Suppose that the hazard rate function is increasing, which implies that $z_d \leq z_1$ for all $d \geq 1$. Choose $\varepsilon = \log\left(1 + \frac{\Phi(1)}{4z_1}\right)$. Then, it follows that

$$\begin{aligned} \lim_{d \rightarrow \infty} \log \frac{\bar{F}(d)}{e^{-\varepsilon d}} &= \lim_{k \rightarrow \infty} \sum_{s=1}^{d-1} \log \left(\frac{z_s}{\mu + z_s} \right) + \varepsilon d \\ &= \lim_{d \rightarrow \infty} \sum_{s=1}^{d-1} \left(\log \left(\frac{z_s}{\mu + z_s} \right) + \log \left(1 + \frac{\Phi(1)}{4z_1} \right) \right) + \varepsilon \\ &\leq \lim_{d \rightarrow \infty} \sum_{s=1}^{d-1} \left(\log \left(\frac{z_1}{\mu + z_1} \right) + \log \left(1 + \frac{\Phi(1)}{4z_1} \right) \right) + \varepsilon \\ &= \lim_{d \rightarrow \infty} \sum_{s=1}^{d-1} \log \left(\frac{\Phi(1)/4 + z_1}{\mu + z_1} \right) + \varepsilon \\ &< \varepsilon. \end{aligned}$$

Hence, the asymptotic degree distribution is not heavy-tailed.

I now show that if the hazard rate function monotonically decreases to zero, then the asymptotic degree distribution is heavy-tailed. By the assumption, it follows that $z_d \rightarrow \infty$ as $d \rightarrow \infty$. For a fixed $\varepsilon > 0$, there exists $\delta > 0$ such that $\varepsilon = \log(1 + \delta)$. Then, I have

$$\begin{aligned} \lim_{d \rightarrow \infty} \log \frac{\bar{F}(d)}{e^{-\varepsilon d}} &= \lim_{k \rightarrow \infty} \sum_{s=1}^{d-1} \log \left(\frac{z_s}{\mu + z_s} \right) + \varepsilon d \\ &= \lim_{d \rightarrow \infty} \sum_{s=1}^{d-1} \left(\log \left(\frac{z_s}{\mu + z_s} \right) + \log(1 + \delta) \right) + \varepsilon \end{aligned}$$

$$\begin{aligned}
&= \lim_{d \rightarrow \infty} \sum_{s=1}^{d-1} \log \left(\frac{\delta z_s + z_s}{\mu + z_s} \right) + \varepsilon \\
&= \infty.
\end{aligned}$$

■

Proof of Corollary 1

Proof. The statement directly follows by and Proposition 3 and Proposition 4. ■

Proof of Lemma 1

Proof. I define a sequence $(a(d))_{d \geq 1}$ as

$$\bar{F}(d) = 1 - \left(\sum_{k=1}^{d-1} f(k) \right) = \prod_{k=1}^d a(k).$$

By its construction, I have

$$f(d) = \left(1 - a(d+1) \right) \prod_{k=1}^d a(k),$$

$$h(d) = 1 - a(d+1).$$

Thus, $h(d)$ is increasing in d if and only if $a(d)$ is decreasing in d for all $d \geq 2$.

By the summation by parts, I find that

$$\sum_{k=d}^{\infty} k f(k) = d \prod_{s=1}^d a(s) + \sum_{k=d+1}^{\infty} \prod_{s'=1}^k a(s').$$

Hence, the inverse hazard rate function of $\tilde{f}(\cdot)$ is

$$\begin{aligned}
\frac{1}{\tilde{h}(d)} &= \frac{\sum_{k=d}^{\infty} k f(k)}{d f(d)} \\
&= \frac{d \prod_{s'=1}^d a(s')}{d(1 - a(d+1)) \prod_{r'=1}^d a(r')} + \frac{\sum_{k=d+1}^{\infty} \prod_{s=1}^k a(s)}{d(1 - a(d+1)) \prod_{r=1}^d a(r)} \\
&= \frac{1}{h(d)} + \frac{1}{dh(d)} \left(\sum_{k=d+1}^{\infty} \prod_{s=d+1}^k a(s) \right).
\end{aligned}$$

Thus, it suffices to show that the inverse hazard rate function is decreasing in d .

Since $a(d)$ is decreasing in d ,

$$\begin{aligned}
& \sum_{k=d+1}^{\infty} \prod_{s=d+1}^k a(s) - \sum_{k'=d+2}^{\infty} \prod_{s'=d+2}^{k'} a(s') \\
&= (a(d+1) + a(d+1)a(d+2) + \dots) - (a(d+2) + a(d+2)a(d+3) + \dots) \\
&= (a(d+1) - a(d+2)) + (a(d+1)a(d+2) - a(d+2)a(d+3)) + \dots \\
&= (a(d+1) - a(d+2)) + \sum_{k=d+2}^{\infty} \left(\prod_{s=d+1}^k a(s) - \prod_{s'=d+2}^k a(s') \right) \\
&> 0.
\end{aligned}$$

Note here that terms in the summations are rearrangeable because each summation is absolutely convergent. With this observation, it follows that the inverse hazard rate function of $\tilde{f}(\cdot)$ is *strictly* decreasing in d . Therefore, the proposition follows. ■

Proof of Proposition 5

Proof. The proof directly follows by the equilibrium characterization. ■

Proof of Proposition 6

Proof. Fix a mechanism (ξ, π) . The social value $\tilde{\xi} = \tilde{E}[\xi(d)]$ is strictly positive if a mechanism returns a strictly positive revenue to the seller. Thus, for any mechanism with a positive revenue, the expected valuation $V(\cdot, \cdot)$ is strictly supermodular as

$$V(d', d_1) - V(d, d_1) = \tilde{\xi} d_1 (\xi(d') - \xi(d)) > \tilde{\xi} d_2 (\xi(d') - \xi(d)) = V(d', d_2) - V(d, d_2).$$

Then, the following lemmas follow:

Lemma A.4 *If $V(\cdot, \cdot)$ is strictly supermodular, then the incentive compatibility has the following properties:*

- $\xi(\cdot)$ is incentive compatible if and only if $\xi(\cdot)$ is monotone.
- $\xi(\cdot)$ is incentive compatible if the following inequalities hold:

$$V(d, d) - \pi(d) \geq V(d-1, d) - \pi(d-1) \text{ for all } d = 1, \dots, d_{\max},$$

$$V(d, d) - \pi(d) \geq V(d+1, d) - \pi(d+1) \text{ for all } d = 0, \dots, d_{\max} - 1.$$

Proof. See Chapter 6 in [Vohra \(2011\)](#) for proofs. ■

Lemma A.5 *For any monotone allocation $\xi(\cdot)$, there exists an expected payment schedule $\pi(\cdot)$ such that all the incentive comparability constraints are satisfied.*

Proof. Since any buyer with degree zero obtains zero utility, I set $\xi(0) = 0$ without loss of generality. Define a payment schedule $\pi(\cdot) : \mathcal{D} \rightarrow \mathbb{R}_+$ such that $\pi(0) := 0$ and

$$\pi(d) := \sum_{k=1}^d \left(V(d, d) - V(d-1, d) \right)$$

for all $d \geq 1$. Then, by the previous lemma, it suffices to show that all downward and upward incentive compatibility constraints are satisfied as:

$$\pi(d) - \pi(d-1) = V(d, d) - V(d-1, d) \text{ for all } d \geq 1,$$

$$\pi(d+1) - \pi(d) > V(d+1, d) - V(d, d) \text{ for all } d \geq 0.$$

Therefore, the lemma is proven. ■

The previous lemmas imply that to solve the seller's problem (6.1), it suffices to consider monotone allocation rules and a payment schedule satisfying the adjacent incentive compatibility constraints. Moreover, for any monotone allocation rule, there is a payment schedule satisfying the adjacent constraints. In particular, the payment schedule satisfies all downward incentive compatibility constraints.

Note that for any optimal mechanism (ξ, π) , the downward incentive compatibility constraints must be binding in the seller's problem (6.1). Thus, given that $\xi(0) = 0$ and $\pi(0) = 0$ without loss of generality, I can fix the payment schedule defined as $\pi(d) := \tilde{\xi} \sum_{k=1}^d (\xi(k)k - \xi(k-1)k)$. This implies that the seller's problem is

$$\underset{\xi: \mathcal{D} \rightarrow [0,1]}{\text{maximize}} \quad \left\{ \sum_{d=1}^{d_{\max}} \tilde{f}(d) \xi(d) \right\} \left[\sum_{d=1}^{d_{\max}} f(d) \left(\xi(d) \left(d - \frac{1 - F(d)}{f(d)} \right) \right) \right]$$

$$\text{subject to } 0 = \xi(0) \leq \xi(1) \leq \dots \leq \xi(d_{\max}) \leq 1.$$

Consider two degree distributions $f(\cdot)$ and $f'(\cdot)$ with $f(\cdot) >_{\text{LR}} f'(\cdot)$. Assuming the increasing hazard rate property for both degree distributions, it follows that $\frac{1-F(d)}{f(d)} > \frac{1-F'(d)}{f'(d)}$.

Since $\tilde{f}(\cdot) >_{\text{FOSD}} f'(\cdot)$, it also follows that $\sum_{d=1}^{d_{\max}} \tilde{f}(d)\xi(d) \geq \sum_{d=1}^{d_{\max}} f'(d)\xi(d)$ for any allocation rule $\xi(\cdot)$. Therefore, the seller's revenue strictly increases as the degree distribution increases in terms of the likelihood ratio order. ■

Proof of Proposition 7

Proof. The proposition follows by the discussions in the main text. ■

B Additional Proofs

I here provide additional proofs and results in probability theory.

Approximation

Claim. Let $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ be random variables such that $\mathbf{X}(t) \in [0, K]$ and $\mathbf{Y}(t) \in [\delta, L]$ with $\delta > 0$. If $(\mathbf{Y}(t))_{t \geq 1}$ converges in probability to a constant $y \in [\delta, 1]$, then as $t \rightarrow \infty$,

$$\left| \mathbb{E} \left[\frac{\mathbf{X}(t)}{\mathbf{Y}(t)} \right] - \mathbb{E} \left[\frac{\mathbf{X}(t)}{y} \right] \right| \rightarrow 0.$$

Proof. For a given $\varepsilon > 0$, let $\Omega(t) := \{\omega : |\mathbf{Y}(t) - y| < \frac{\varepsilon \delta^2}{12K}\}$. By the assumption, there exists T such that $\mathbb{P}(\Omega(t)^c) < \frac{\varepsilon \delta^2}{12KL}$ whenever $t \geq T$. Therefore, $t \geq T$ implies that

$$\begin{aligned} \left| \mathbb{E} \left[\frac{\mathbf{X}(t)}{\mathbf{Y}(t)} \right] - \mathbb{E} \left[\frac{\mathbf{X}(t)}{y} \right] \right| &\leq \left| \int_{\Omega(t)} \frac{\mathbf{X}(t)y - \mathbf{X}(t)\mathbf{Y}(t)}{y\mathbf{Y}(t)} dF \right| + \left| \int_{\Omega(t)^c} \frac{\mathbf{X}(t)y - \mathbf{X}(t)\mathbf{Y}(t)}{y\mathbf{Y}(t)} dF \right| \\ &\leq \frac{1}{\delta^2} \left(\int_{\Omega(t)} |\mathbf{X}(t)y - \mathbf{X}(t)\mathbf{Y}(t)| dF + \int_{\Omega(t)^c} |\mathbf{X}(t)y - \mathbf{X}(t)\mathbf{Y}(t)| dF \right) \\ &< \frac{1}{\delta^2} \left(\frac{\varepsilon \delta^2}{12K} \int_{\Omega(t)} |\mathbf{X}(t)| dF + 2L \int_{\Omega(t)^c} |\mathbf{X}(t)| dF \right) \\ &< \frac{1}{\delta^2} \left(\frac{\varepsilon \delta^2}{12} \mathbb{P}(\Omega(t)) + (2KL) \mathbb{P}(\Omega(t)^c) \right) \\ &< \varepsilon. \end{aligned} \quad \blacksquare$$

Convergence of Random Variables in \mathbb{N}^∞

Claim. Let $(\mathbf{X}_n)_{n \geq 1}$ be a sequence of random variables. Let \mathbf{X}_∞ be a random variable distributed over \mathbb{N}^∞ . Then, \mathbf{X}_n converges in distribution to \mathbf{X}_∞ as $n \rightarrow \infty$ if and only if for all $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{X}_n = x) = \mathbb{P}(\mathbf{X}_\infty = x).$$

Proof. Suppose that \mathbf{X}_n converges to \mathbf{X}_∞ in distribution as $n \rightarrow \infty$. Let $(a, b) \subset \mathbb{R}$ such that $(a, b) \cap \mathbb{N} = \emptyset$. Then, it follows that

$$\mathbb{P}(\mathbf{X}_\infty \in (a, b)) \leq \liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in (a, b)) = 0.$$

This part is shown in Theorem 29.1 in Billingsley (1995). Therefore, for any $x \notin \mathbb{N}$, there

exists (a, b) such that $x \in (a, b)$, $(a, b) \cap \mathbb{N} = \emptyset$, and $\mathbb{P}(\mathbf{X}_\infty = x) = 0 = \lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{X}_n = x)$.

Now fix $x \in \mathbb{N}$. By choosing $\varepsilon = 1/2$, I find that

$$\begin{aligned} \mathbb{P}(\mathbf{X}_\infty = x) &= \mathbb{P}(\mathbf{X}_\infty \in (x - \varepsilon, x + \varepsilon)) \\ &\leq \liminf_{n \rightarrow \infty} \mathbb{P}(\mathbf{X}_n \in (x - \varepsilon, x + \varepsilon)) \\ &\leq \limsup_{n \rightarrow \infty} \mathbb{P}(\mathbf{X}_n \in (x - \varepsilon, x + \varepsilon)) \\ &\leq \limsup_{n \rightarrow \infty} \mathbb{P}(\mathbf{X}_\infty \in (x - \varepsilon, x + \varepsilon)) \\ &= \mathbb{P}(\mathbf{X}_\infty = x), \end{aligned}$$

where the second and the last inequalities are by Theorem 29.1 in Billingsley (1995). Therefore, it follows that $\lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{X}_n = x) = \mathbb{P}(\mathbf{X}_\infty = x)$.

To show the converse, fix $z \in \mathbb{R}$, and assume that $F_\infty(\cdot)$ is continuous at z . Since $F_\infty(z) = \mathbb{P}(\mathbf{X}_\infty \leq \lfloor z \rfloor)$, it follows that

$$F_\infty(z) = \sum_{k \leq \lfloor z \rfloor} \mathbb{P}(\mathbf{X}_\infty = k) = \sum_{k \leq \lfloor z \rfloor} \lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{X}_n = \lfloor z \rfloor) = \lim_{n \rightarrow \infty} F_n(\lfloor z \rfloor) = \lim_{n \rightarrow \infty} F_n(z).$$

Note that the above argument does not hold if the support of \mathbf{X}_∞ is not \mathbb{N} .³⁸ ■

Bounds of the Largest Eigenvalue

Claim. Let $\mathbf{G} = \langle N, A \rangle$ be a network with size n . Its largest eigenvalue satisfies

$$\sqrt{d_{\max}(\mathbf{G})} \leq \lambda_{\max}(\mathbf{G}) \leq d_{\max}(\mathbf{G}).$$

Proof. Throughout the proof, let node 1's degree be the maximum degree d_{\max} .

To prove the lower bound, note that $\lambda_{\max}(\mathbf{G})$ satisfies

$$\lambda_{\max}(\mathbf{G}) \geq \frac{\mathbf{q}'A\mathbf{q}}{\mathbf{q}'\mathbf{q}}$$

for all $\mathbf{q} \in \mathbb{R}^n$. Without loss of generality, suppose that nodes $2, \dots, k$ have, respectively, a

³⁸A simple counter example is a sequence of integer-valued random variables $(\mathbf{X}_n)_{n \geq 1}$ such that $\mathbb{P}(\mathbf{X}_n = k) = 1/n$ if and only if $1 \leq k \leq n$. Then, $\lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{X}_n = x) = 0$ for all $x \in \mathbb{R}$. Let $\mathbf{X}_\infty \sim N(0, 1)$. Then, $\mathbb{P}(\mathbf{X}_\infty = x) = 0$, but \mathbf{X}_n does not converges in distribution to \mathbf{X}_∞ as $n \rightarrow \infty$.

link to node 1, but nodes $k + 1, \dots, n$ do not. Choose a vector \mathbf{q} such that

$$\mathbf{q} = (\sqrt{d_{\max}(\mathbf{G})}, \mathbf{1}_{1 \times (k-1)}, \mathbf{0}_{1 \times (n-k-1)})',$$

where $\mathbf{1}_{1 \times k-1}$ is a $1 \times (k-1)$ vector with entries of one, and $\mathbf{0}_{1 \times (n-k-1)}$ is a $1 \times (n-k-1)$ vector with entries of zero. Then, $\mathbf{q}'\mathbf{q} = 2d_{\max}(\mathbf{G})$ and

$$\mathbf{q}'A\mathbf{q} = (d_{\max}(\mathbf{G}))^{3/2} + \sqrt{d_{\max}(\mathbf{G})}(d_2(\mathbf{G}) + \dots + d_k(\mathbf{G})) = 2(d_{\max}(\mathbf{G}))^{3/2}.$$

Thus, $\lambda_{\max}(\mathbf{G}) \geq \sqrt{d_{\max}(\mathbf{G})}$.

To prove the upper bound, let \mathbf{x}_{\max} be an eigenvector corresponding to the maximum eigenvalue $\lambda_{\max}(\mathbf{G})$. By its definition, $A\mathbf{x} = \lambda_{\max}(\mathbf{G})\mathbf{x}$, and it follows that

$$\lambda_{\max}(\mathbf{G})\mathbf{x}_1 = \sum_{j \in N_1} \mathbf{x}_j,$$

where \mathbf{x}_i is the vector that has only one non-zero entry at the i -th entry. Hence,

$$|\lambda_{\max}(\mathbf{G})||\mathbf{x}_1| \leq \sum_{j \in N_1} |\mathbf{x}_j| \leq d_{\max}(\mathbf{G})|\mathbf{x}_1|. \quad \blacksquare$$

The Second Borel-Cantelli Lemma

Claim. Let $(E_n)_{n=1}^{\infty}$ be a sequence of independent events. If $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty$, then

$$\mathbb{P}(E_n \text{ i.o.}) = 1.$$

Proof. Independence and $1 - x \leq e^{-x}$ imply that

$$\mathbb{P}\left(\bigcap_{k=n}^{\infty} E_k^c\right) = \prod_{k=n}^{\infty} (1 - \mathbb{P}(E_k)) \leq \prod_{k=n}^{\infty} \exp(-\mathbb{P}(E_k)) = \exp\left(-\sum_{k=n}^{\infty} \mathbb{P}(E_k)\right) = 0.$$

Thus, $\mathbb{P}(\bigcup_{k=n}^{\infty} E_k) = 1$ for all n . Since $\bigcup_{k=n}^{\infty} E_k$ decreases monotonically to $\limsup_n E_n$ as $n \rightarrow \infty$, it follows that

$$\mathbb{P}(E_n \text{ i.o.}) = 1.$$

References

- Azuma, K. (1967). Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367.
- Baetz, O. (2015). Social activity and network formation. *Theoretical Economics*, 10:315–340.
- Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who’s who in networks. wanted: the key player. *Econometrica*, 74(5):1403–1417.
- Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 285.
- Belhaj, M., Bramoullé, Y., and Deroän, F. (2014). Network games under strategic complementarities. *Games and Economic Behavior*, 88:310–319.
- Bender, E. A. and Canfield, E. R. (1978). The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24:296–307.
- Billingsley, P. (1995). *Probability and Measure*. Wiley-Interscience, 3rd edition.
- Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, 18(3):279–290.
- Bramoullé, Y., Kranton, R., and D’amours, M. (2014). Strategic interaction and networks. *American Economic Review*, 104(3):898–930.
- Bubeck, S., Mossel, E., and Rácz, M. Z. (2015). On the influence of the seed graph in the preferential attachment model. *Transactions on Network Science and Engineering (Forthcoming)*.
- Chung, F., Lu, L., and Vu, V. (2003). Spectra of random graph with given expected degrees. *Proceedings of the National Academy of Sciences of the United States of America*, 100(11):6313–6318.

- Cooper, C. and Frieze, A. (2003). A general model of web graphs. *Random Structures and Algorithms*, 22(3):311–335.
- Coppersmith, D. and Winograd, S. (1990). Matrix multiplication via arithmetic progressions. *Journal of Symbolic Computation*, 9(3):251–280.
- Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: homophily, minorities, and segregation. *Econometrica*.
- Cvetković, D. and Rowlinson, P. (1990). The largest eigenvalue of a graph: a survey. *Linear and Multilinear Algebra*, 28:3–33.
- Dorogovtsev, S. N. and Mendes, J. F. F. (2001). Scaling properties of scale-free evolving networks: continuous time approach. *Physical Review E*, 63(5):056125.
- Dorogovtsev, S. N., Mendes, J. F. F., and Samukhin, A. N. (2000). Structure of growing networks with preferential linking. *Physical Review Letters*, 85(21):4633.
- Durrett, R. (2005). *Probability: theory and examples*. Duxbury Advanced Series. Thomson, 3rd edition.
- Durrett, R. (2010). *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1st edition.
- Erdős and Rényi, A. (1959). On random graphs I. *Publications Mathematicae Debrecen*, 6:290–297.
- Fainmesser, I. P. and Galeotti, A. (2015). Pricing network effects. *Review of Economic Studies (Forthcoming)*.
- Feri, F. and Pin, P. (2015). The effect of externalities aggregation on network games outcomes. *Working Paper*.

- Flaxman, A., Frieze, A., and Fenner, T. (2005). High degree vertices and eigenvalue in the preferential attachment graph. *Internet Mathematics*, 2(1):1–19.
- Galeotti, A., Goyal, S., Jackson, M. O., Vega-Redondo, F., and Yariv, L. (2010). Network games. *Review of Economic Studies*, 77:218–244.
- Ghiglino, C. (2012). Random walk to innovation: Why productivity follows a power law. *Journal of Economic Theory*, 147(2):713–737.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.
- Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.
- Jackson, M. O. and Rogers, B. W. (2007a). Meeting strangers and friends of friends: how random are social networks? *American Economic Review*, 97(3):890–915.
- Jackson, M. O. and Rogers, B. W. (2007b). Relating network structure to diffusion properties through stochastic dominance. *The B.E. Journal of Theoretical Economics*, 7(1).
- Jackson, M. O. and Yariv, L. (2007). Diffusion of behavior and equilibrium properties in network games. *American Economic Review*, 97(2):92–98.
- Jehiel, P., Moldovanu, B., and Stacchetti, E. (1996). How (not) to sell nuclear weapons. *American Economic Review*, 86(4):814–829.
- Jehiel, P., Moldovanu, B., and Stacchetti, E. (1999). Multidimensional mechanism design for auction with externalities. *Journal of Economic Theory*, 85:258–293.
- Karlin, S. and Rubin, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. *Annals of Mathematical Statistics*, XXVII:272–299.
- König, M. D., Tessone, C. J., and Zenou, Y. (2014). Nestedness in networks: a theoretical model and some applications. *Theoretical Economics*, 9:695–752.

- Krapivsky, P. L., Redner, S., and Leyvraz, F. (2000). Connetivity of growing random networks. *Physical Review Letters*, 85(21).
- López-Pintado, D. (2008). Diffusion in complex social networks. *Games and Economic Behavior*, 62(2):573–590.
- Marsden, J. E. and Hoffman, M. J. (1993). *Elementary classical analysis*. W. H. Freeman, 2nd edition.
- Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73.
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.
- Shin, E. (2015). Monopoly pricing and diffusion of (social) network goods. *Working Paper*.
- Vohra, R. V. (2011). *Mechanism design: a linear Programming approach*. Econometric Society Monographs. Cambridge University Press.
- Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393:440–442.

Online Appendix to “Social Network Formation and Strategic Interaction in Large Networks”

Euncheol Shin

Recent Version: <http://people.hss.caltech.edu/~eshin/pdf/DSNF-OA.pdf>

October 13, 2015

Abstract

In this online appendix, I present a continuous time approach to my model with a parametric form $\Phi(d) = d^\alpha$. The resulting degree distribution is the Weibull distribution (Weibull, 1951). I fit my model to four empirical degree distributions by using the method of maximum likelihood estimation. I compare the goodness-of-fit of my model with other three prominent degree distributions generated by Erdős and Rényi (1959), Barabási et al. (1999), and Jackson and Rogers (2007). I use the Kolmogorov-Smirnov statistic as a way to find the best fitting distribution.

1 Models and Degree Distributions

1.1 Continuous Time Approach

To obtain a tractable form of the resulting degree distribution, I apply a continuous time approach to my model.¹ In this approach, a node’s degree changes deterministically at a rate proportional to its expected change. Specifically, the degree of node i changes according to the following ordinary differential equation:

$$\frac{dd_i(t)}{dt} = \frac{d_i(t)}{\mu t} \frac{1}{d_i(t)^\alpha}, \quad (1.1)$$

where $d_i(t)$ is the degree of node i at time t , $\mu > 0$ is the parameter that represents the rate of new nodes entering the network, and $\alpha > 0$ is the parameter that captures the cost

¹See Chapter 5 in Jackson (2010) and references therein for examples and discussions about this approach.

of forming new links. The first term shows that higher degree nodes have a greater chance of being identified by newborn nodes, and the second term shows that higher degree nodes are more selective with additional links. When the parameter α is one, the probability that node i forms additional links is independent of its degree. However, when parameter α is larger than one, node i is *less* likely to form additional links as its degree *increases*.

I solve the differential equation (1.1) and find that

$$d_i(t)^\alpha - d_i(i)^\alpha = \frac{\alpha}{\mu} \log\left(\frac{t}{i}\right).$$

To find an expression for the complementary cumulative degree distribution, I identify the proportion of nodes with degrees that *exceed* a given degree d at time t . For this, it suffices to find the node having *exactly* degree d at time t because all nodes born before time t have larger degrees. Let $i_t(d)$ be the node that has degree d at time t : $d_{i_t(d)}(t) = d$. In addition, let d_0 be node $i_t(d)$'s initial degree: $d_{i_t(d)}(i_t(d)) = d_0$. Then, I have

$$\frac{i_t(d)}{t} = e^{-((\lambda d)^\alpha - (\lambda d_0)^\alpha)},$$

where $\lambda = \left(\frac{\mu}{\alpha}\right)^{-\alpha}$. Finally, by setting $d_0 = 0$, the complementary cumulative degree distribution takes a form of a Weibull distribution as $\bar{F}(d; \lambda, \alpha) = e^{-(\lambda d)^\alpha}$. The hazard rate function of the Weibull distribution is increasing if and only if the parameter α is larger than one. The following proposition summarizes the results.

Proposition 1 *The resulting complementary cumulative degree distribution has a form of*

$$\bar{F}(d; \lambda, \alpha) = e^{-(\lambda d)^\alpha},$$

where $\lambda = \left(\frac{\mu}{\alpha}\right)^{-\alpha}$. *The hazard rate function is strictly increasing if and only if $\alpha > 1$.*

1.2 Other Models and Degree Distributions

Poisson model. The uniform random network formation model by Erdős and Rényi (1959) generates the Poisson distribution as the resulting degree distribution. In their model, a link between two nodes is formed independently of other pairs of nodes with a fixed probability. The resulting degree distribution is approximated by the Poisson distribution when the

network size is sufficiently large. Specifically, the Poisson distribution with parameter $\rho > 0$ has the form of

$$f(d; \rho) = \frac{e^{-\rho} \rho^d}{d!}.$$

The hazard rate of the Poisson distribution is strictly increasing for any value of ρ .

Preferential attachment model. The scale-free distribution is another prominent degree distribution generated by the preferential attachment (PA) model (Barabási et al., 1999). In the PA model, a newborn node forms links with existing nodes probabilistically, and the probability of selecting a particular existing node is proportional to its degree. The resulting degree distribution is the scale-free distribution, which has the form of

$$f(d; \gamma) = cd^{-\gamma},$$

where $\gamma > 0$ is the scale parameter, and $c > 0$ is the normalization factor that is calculated as $c = \sum_{d=d_{\min}}^{\infty} d^{-\gamma}$ when d_{\min} is the minimum degree of a network.² Assuming that the minimum degree is one, the complementary cumulative degree distribution has the form of

$$\bar{F}(d; \gamma) = \frac{\zeta(\gamma, d)}{\zeta(\gamma)},$$

where $\zeta(\gamma) = \sum_{k=1}^{\infty} k^{-\gamma}$ is the Riemann zeta function, and $\zeta(\gamma, d) = \sum_{k=0}^{\infty} (k + d)^{-\gamma}$ for $d \geq 1$. Note that the Riemann zeta function is well-defined if and only if $\gamma > 1$.

Network-based search model. The network-based search model by Jackson and Rogers (2007) (henceforth, the JR model) produces another prominent degree distribution. In the JR model, nodes form links in two ways: (i) uniformly and randomly, and (ii) by searching locally through the given structure of the network. There are two parameters $m > 0$ and $r > 0$. m is the expected number of links that a new node forms, and r is the ratio of the number of links that are formed uniformly at random comparing to network-based meetings.

²Its name originates from the fact that for any two degrees of a fixed ratio, their probability ratios are independent of the scale of degrees: $f(d)/f(d') = f(sd)/f(sd')$ for all $d, d' \in \mathbb{N}$ and $s \in \mathbb{R}_+$. A scale-free distribution can be defined either over continuous real numbers or discrete positive integers. Although a continuous distribution intuitively explains its various properties, an estimation by using it generates systematic errors (Clauset et al., 2009).

Notice that r can be infinity, which means that the network formation is purely random. When $r < \infty$, the resulting complementary cumulative degree distribution has the form of

$$\bar{F}(d; r, m) = 1 - \left(\frac{d_0 + rm}{d + rm} \right)^{1+r},$$

where d_0 is the parameter that calculates the number of links that each node forms upon its birth. Following [Jackson and Rogers \(2007\)](#), I set $d_0 = 0$ for the later parameter estimation.

2 Fitting Models to Datasets

2.1 Datasets

I consider the following four real network datasets: (a) the 75 social networks of rural Indian villages, (b) a collaboration network of jazz musicians, (c) an online friendship network of Facebook users, and (d) the network among webpages at Notre Dame University.³ [Figure 1](#) plots their empirical hazard rate functions. I present the hazard rate function of village #58 as an illustrating example. One can observe that the empirical hazard rates exhibit increasing patterns for datasets (a) and (b), but decreasing patterns for datasets (c) and (d).

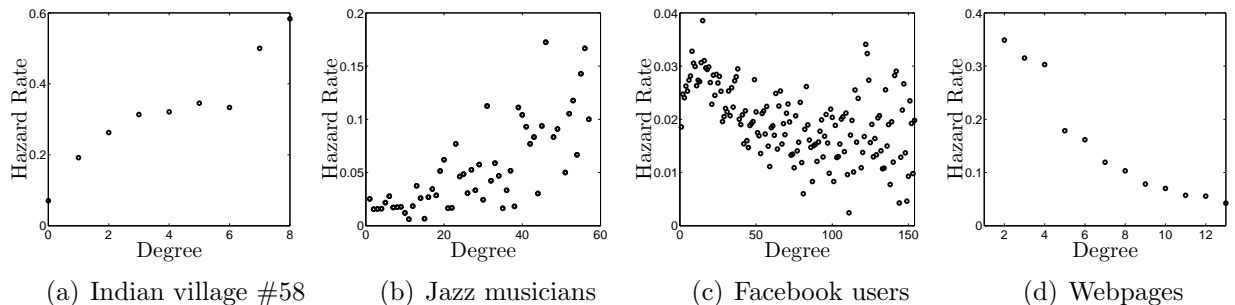


Figure 1: Different patterns of the hazard rate function across datasets

2.2 Estimation Strategy and Comparison

Estimation strategies. I fit the four theoretical degree distributions generated by network formation models to the empirical degree distributions of the above four network datasets. I

³I obtain dataset (a) from [Banerjee et al. \(2013\)](#), dataset (b) from [Gleiser and Danon \(2003\)](#), dataset (c) from [McAuley and Leskovec \(2012\)](#), and dataset (d) from [Jackson and Rogers \(2007\)](#). All datasets are publicly available from their corresponding author’s webpage.

explain how to estimate model parameters. First, I estimate parameters of the Weibull distribution by using the method of maximum likelihood estimation, which is well-documented in the statistics literature (e.g., [Rinne, 2008](#)). Specifically, for a given dataset $\{d_1, \dots, d_n\}$ where d_i is the degree of node $i \in \{1, \dots, n\}$, the maximum likelihood estimator of α is identified as a solution of

$$\frac{1}{\alpha} = \frac{\sum_{i=1}^n d_i^\alpha \log d_i}{\sum_{i=1}^n d_i^\alpha} - \frac{1}{n} \sum_{i=1}^n \log d_i.$$

One can find a solution of the above equation by using a standard iterative procedure (e.g., the Newton-Raphson method). The solution is unique because the left-hand side is strictly decreasing in α , but the right-hand side is strictly increasing in α .

Second, I estimate the parameter $\rho > 0$ for the Poisson distribution by using the method of maximum likelihood estimation. In particular, for a given dataset $\{d_1, \dots, d_n\}$, the maximum likelihood estimator $\hat{\rho}$ is the average degree from a dataset as

$$\hat{\rho} = \frac{1}{n} \sum_{i=1}^n d_i.$$

Third, I estimate the scale parameter γ by using the method of maximum likelihood estimation ([Clauset et al., 2009](#)), which solves the following equation:

$$\frac{\zeta'(\gamma)}{\zeta(\gamma)} = -\frac{1}{n} \sum_{i=1}^n \log d_i.$$

A solution of the above equation always exists and is unique because the left-hand side is strictly decreasing in γ , but the right-hand side is constant.

Finally, for the JR model, I estimate two parameters m and r according to the iteration method suggested by [Jackson and Rogers \(2007\)](#). To begin with, estimate m as the average degree from a dataset. To estimate r , note first that

$$\log \bar{F}(d; r, m) = (1 + r) \log (rm) - (1 + r) \log (d + rm).$$

I use a method of iterative least squares as follows. Start with an initial value of r , say r_0 , and plug in this value to obtain $\log (d + r_0 m)$. Estimate $-(1 + r)$ by using the method of least squares to obtain r_1 . Repeat the same procedure until a fixed point \hat{r} is achieved. If

the procedure converges, a fixed point is unique regardless of the initial value r_0 ; otherwise, set $\hat{r} = \infty$. When $\hat{r} = \infty$, use the exponential distribution as the limit degree distribution.

Goodness-of-fit. As a way to compare the models in terms of data fitting performance, I use the Kolmogorov-Smirnov (KS) statistic (Smirnov, 1944).⁴ The KS statistic is calculated as follows. Let $\{d_1, \dots, d_n\}$ be a given dataset. Then, I find the empirical cumulative degree distribution $\bar{F}_n(\cdot)$ as

$$F_n(d) = \frac{1}{n} \sum_{i=1}^n \mathbb{1} \{d_{\min} \leq d_i \leq d\}.$$

For each model, let $F(\cdot; \hat{\theta})$ be the estimated cumulative degree distribution from the estimated parameter $\hat{\theta}$. The KS statistic is defined as

$$\sup_d |F_n(d) - F(d; \hat{\theta})|.$$

By the Glivenko-Cantelli theorem, the KS statistic converges almost surely to zero as the dataset size n becomes large if the dataset is generated according to the given model (Mood et al., 1974). Thus, the smaller value of the KS statistic means the better data fitting performance of the model. Notice that this convergence is independent of the increasing hazard rate property of the true underlying degree distribution.

2.3 Results

I first introduce the key parameters of the models. In my model, $\alpha > 0$ is the key parameter that distinguishes my model from other models; the other parameter $\lambda > 0$ provides no additional information but the average degree of the network, which is available information from other three models. Recall that α indicates the behavior of the hazard rate function. In the Poisson model, there is only one parameter $\rho > 0$. There is also a single parameter $\gamma > 0$ in the PA model. In the JR model, the key parameter is $r \in (0, \infty]$ that represents the ratio of network-based meetings; the other parameter m calculates the average degree of the network. When the estimate for r is infinity, I report the KS statistic

⁴See Chapter 11 in Mood et al. (1974) for details of this test.

by using an exponential distribution as the limit distribution.⁵

In [Table 1](#), I report the estimates for the above four parameters and the KS statistics across datasets. The first and second lowest KS statistics are marked by $\diamond\diamond$ and \diamond , respectively.⁶ Since there are 75 networks in the Indian villages dataset, I present the minimum and the maximum values of the estimates and the KS statistics. [Figure 2](#) illustrates the goodness-of-fit of the models across the datasets.

Dataset	My model	Poisson model	PA model	JR model
	$\alpha > 0$ KS statistic	$\rho > 0$ KS statistic	$\gamma > 0$ KS statistic	$r > 0$ KS statistic
Indian villages	[1.41, 2.39] [0.051, 0.159] \diamond	[3.80, 7.40] [0.035, 0.151] $\diamond\diamond$	[1.49, 2.39] [0.300, 0.443]	[2.38, ∞] [0.098, 0.456]
Jazz musicians	1.57 0.042 $\diamond\diamond$	27.70 0.308 \diamond	1.28 0.391	∞ 0.380
Facebook users	0.96 0.053 \diamond	43.125 0.538	1.27 0.343	2.54 0.033 $\diamond\diamond$
Websites	0.74 0.248	4.20 0.530	1.98 0.016 $\diamond\diamond$	0.57 0.203 \diamond

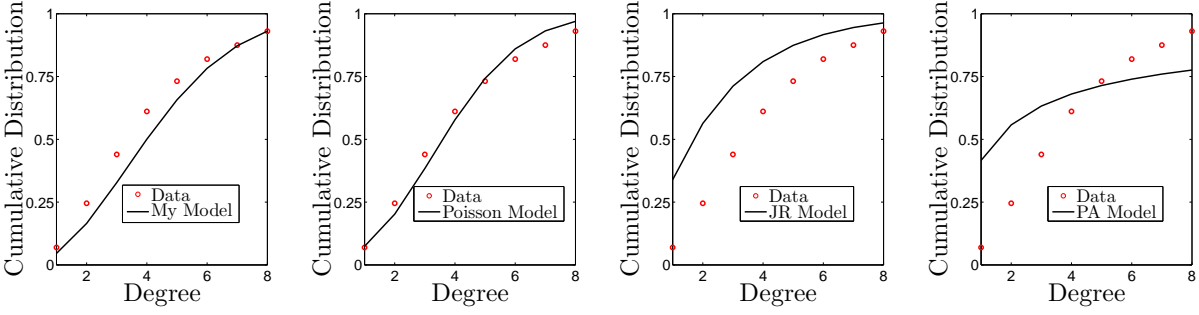
Table 1: Parameter estimates and the goodness-of-fit of the models across datasets

First of all, I find that all the estimates for the parameter α are consistent with simple eyeball tests by using [Figure 1](#). Specifically, estimates for the networks of Indian villages and jazz musicians are all strictly greater than one, which confirm that hazard rates are increasing. Since the KS statistics of my model are very low for these networks, the inference about the increasing hazard rates is significant. On the contrary, for the networks of Facebook users and websites, the estimates are all strictly smaller than one.

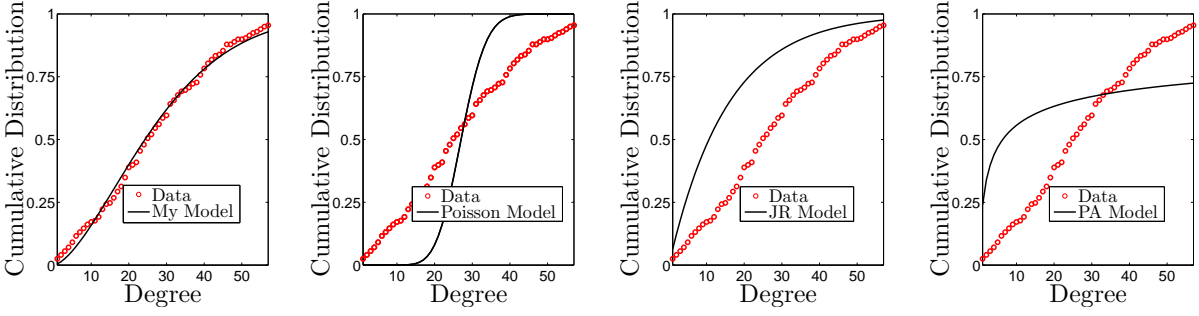
For the 75 social networks of Indian villages, my model and the Poisson model fit the datasets better. Although the lowest KS statistic of the JR model is 0.098, the second lowest KS statistic is 0.171, which is strictly greater than the largest KS statistic of my

⁵The exponential distribution is a special case of the Weibull distribution with $\lambda = m/2$ and $\alpha = 1$.

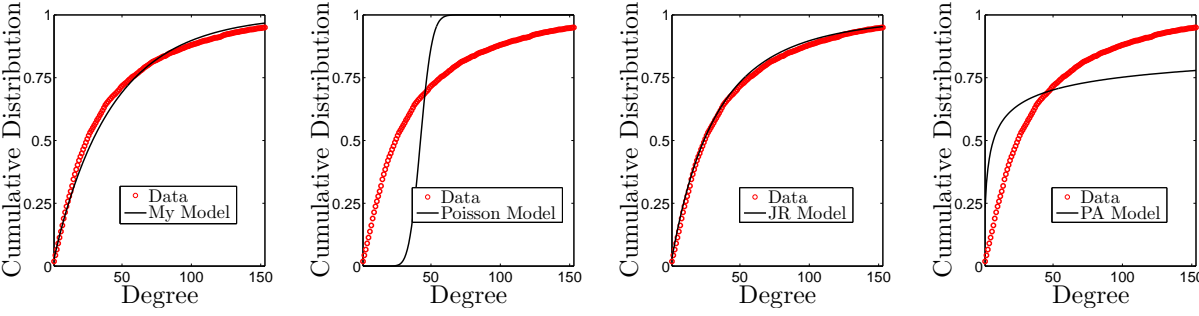
⁶The estimate of the network of websites using the PA model is different from the values reported by [Barabási and Albert \(1999\)](#). This result occurs because they use the usual least square estimates assuming a continuous scale-free distribution, but I use the method of maximum likelihood estimation assuming discrete scale-free distribution.



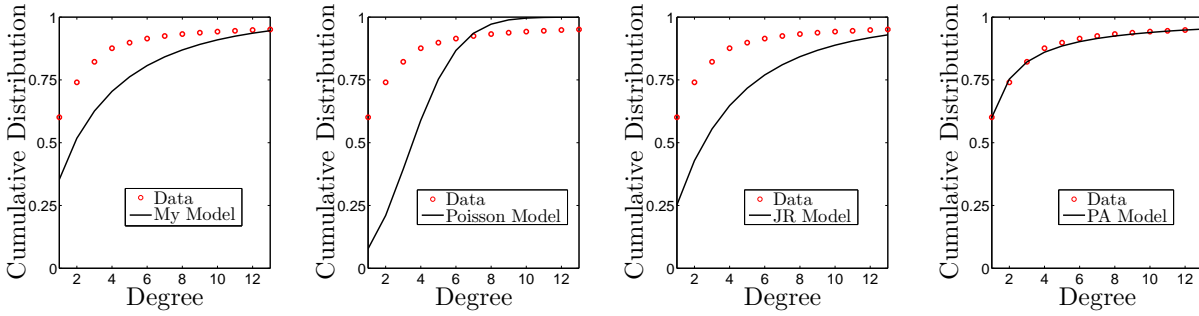
(a) Indian village #58



(b) Jazz musicians



(c) Facebook users



(d) Webpages

Figure 2: An illustration of the goodness-of-fit across datasets and models: In each plot, the horizontal axis represents degrees, and the vertical axis represents the empirical (the red circles) and estimated (the black line) cumulative degree distributions. For each model, the KS statistic measures the maximum difference between the two distributions.

model. Moreover, it turns out that for 71 social networks, the estimates for r are infinity.⁷ This implies that the JR model loses its explanatory power because $r = \infty$ corresponds to the random attachment model.

The difference between my model and the Poisson model becomes clearer when I compare their KS statistics for the networks of jazz musicians and Facebook users. My model still fits both datasets well (the KS statistics are about 0.05) but the Poisson model returns very large KS statistics that are greater than 0.3. This difference becomes clearer when observing the shape of the fitted cumulative distributions in Figure 2-(b) and -(c). The fitted distributions when using the Poisson model are *S-shaped*, but the empirical distributions show different patterns. In contrast, the shape of the fitted distributions by my model is flexible and performs well for the networks of jazz musicians and Facebook users.

The main force driving the flexibility of my model is that it contains two parameters, λ and α , that determine the size and shape of the distribution. For example, the size of the jazz musician network is large in that its average degree is about 27. In my model, the scale parameter λ explains the average degree, and the shape parameter α explains the shape of the empirical distribution. However, there is only one parameter ρ in the Poisson model, which indicates only the scale of the distribution. Thus, when the network is large, my model outperforms the Poisson distribution even when the empirical hazard rate function exhibits increasing patterns.

Finally, for the network of webpages, the PA model performs significantly better than any other models. The KS statistic is very small (0.016), which is the lowest value among all the KS statistics across datasets and models. In comparison to my model, this dominant performance of the PA model is not surprising: my model tries to explain networks with bilateral and costly link formation, but the network of webpages is based on unilateral and costless link formations. Therefore, I conclude that my model fits empirical degree distributions very well whenever link formations are bilateral and costly.

⁷Only villages #13, #19, #30, and #36 return finite estimates for r .

References

- Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144).
- Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 285.
- Barabási, A.-L., Albert, R., and Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A*, 272:173–187.
- Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.
- Erdős and Rényi, A. (1959). On random graphs i. *Publications Mathematicae Debrecen*, 6:290–297.
- Gleiser, P. M. and Danon, L. (2003). Community structure in jazz. *Advances in Complex Systems*, 6(565).
- Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.
- Jackson, M. O. and Rogers, B. W. (2007). Meeting strangers and friends of friends: how random are social networks? *American Economic Review*, 97(3):890–915.
- McAuley, J. and Leskovec, J. (2012). Learning to discover social circles in ego networks. *Advances in Neural Information Processing Systems 25*, pages 539–547.
- Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the theory of statistics*. McGraw-Hill series in probability and statistics. McGraw-Hill College.
- Rinne, H. (2008). *The Weibull distribution: A handbook*. Chapman and Hall/CRC, 1st edition.

Smirnov, N. V. (1944). Approximate laws of distributions of random variables from empirical data. *Uspekhi Mat. Nauk*, 10:179–206.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 18(3):293–297.