Essays on Social Networks and Political Economy

Thesis by

Euncheol Shin

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2016
Defended May 13th, 2016

Euncheol Shin
ORCID: 0000-0002-6053-4050

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Leeat Yariv, my mentor and dissertation committee chair. It is not an exaggeration to say that she has provided dedicated and unique guidance for all the research papers that I have developed at the California Institute of Technology. I would like to offer equal thanks to Federico Echenique, my dissertation committee member. His guidance has played a fundamental role in the completion of my thesis. Specifically, the weekly meetings with him and other graduates working on economic theory have been a unique and fascinating learning experience during my Ph.D. studies.

In addition, many sincere thanks are due to two other members of my dissertation committee, Matthew Elliott and Matthew Shum. They were upfront in evaluating my research papers so that I was able to improve them in new and better ways. I am very thankful to two senior professors, John Ledyard and Tomas Palfrey, for sharing their wisdom and knowledge on economics and political science. I owe a very important debt to Ben Gillen, Jean-Laurent Rosenthal, Kota Saito, and Omer Tamuz for helpful guidance for job market preparation. I am also pleased to acknowledge a unique opportunity to interact with other professors in the Division of the Humanities and Social Sciences at Caltech. I especially want to thank Laurel Auchampaugh for various support from the beginning to the end of my Ph.D. studies.

Throughout my Ph.D. studies, I had the great fortune of meeting great research colleagues. I thank the previous Caltech students who started their Ph.D. studies with me: Kyle Carlson, Matthew Chao, Jonathan Chapman, Khai Chiong, Geoff Fisher, Chris Hagel, Jin Huang, Matthew Kovach, Samantha Myers, Allyson Pellissier, and Kirill Pogorelskiy. I also thank my officemates from my last year at Caltech: Liam Clegg, Sergio Fortes, Yifei Huang, Duk Gyoo Kim, Gerelt Tserenjigmid, and Qiaoxi Zhang. For my research, I have also benefited from conversations with scholars from different universities and departments: Hyoung Jun Ahn, Yeon-Koo Che, Jaypil Choi, Jong-Hee Hahn, Jinwook Hur, Jinwoo Kim, Youngse Kim, Jihong Lee, Kyu-Min Lee, SangMok Lee, Jaeok Park, Chris Shannon, Adam Wierman, and Seungil You.

And last but not least, I would like to express my deep gratitude to Mimi Jeon, my lovely wife, for her constant support. She has been of great help, both academically and personally. She is the person who has shown the utmost patience and support throughout my Ph.D. studies. My daughter, Irene Yerin Shin, has generated endless smiles and joy during my whole life in Pasadena, as predicted by Leeat Yariv.

ABSTRACT

This dissertation consists of two original studies in social networks and one original study in political economy. In the first two chapters, I study (i) how social networks form, and (ii) how economic agents optimize their behaviors for a given network structure. In the last chapter, I examine how election rules affect individual voting decisions and ultimate election outcomes.

In Chapter 1, "Social Network Formation and Strategic Interaction in Large Networks," 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.

In Chapter 2, "Monopoly Pricing and Diffusion of a (Social) Network Good," I present a model of dynamic pricing and diffusion of a network good sold by a monopolist. In the model, the network good is a *subscription social network good*. This means that in each period, each consumer has to pay a subscription price to use the good, and the utility derived from subscribing to the good increases as more of her neighboring consumers subscribe. Consumers myopically optimize their subscription decisions, and the monopolist chooses a sequence of subscription prices that maximizes his discounted sum of per-period profits. Three main results emerge. First, I characterize a unique steady state of the monopoly market where both the monopolist and consumers do not change their decisions. Second, I find that optimal sequences of subscription prices oscillate around the subscription price at the steady state as time passes. Third, I analyze how changes in the monopolist's discount factor and the density of the social network affect the subscription price, subscription rate, and deadweight loss at the steady state.

In Chapter 3, "A Model of Pre-Electoral Coalition Formation," I study how two different election rules, simple plurality (e.g., as in South Korea) and two-round runoff (e.g., as in France), affect political candidates' incentives to form

pre-electoral coalitions (PECs). In my model, three candidates compete for a single office, and two candidates can form a PEC. Since the candidates are both policy- and office-motivated, one candidate can incentivize the other candidate to withdraw his candidacy by choosing a joint policy platform. I find that PECs are more likely to form in plurality elections than in two-round runoff elections. I further examine how other electoral environments, such as ideological distance and pre-election polls, influence incentives to form PECs.

PUBLISHED CONTENTS

As the date of May 13th, 2016, I declare that (i) Chapter 2 in this dissertation, "Monopoly Pricing and Diffusion of (Social) Network Goods," has been revised and resubmitted to the *Games and Economic Behavior*, and (ii) Chapter 3 in this dissertation, "A Model of Pre-Electoral Coalition Formation," is under review by the *Journal of Public Economics*.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

SOCIAL NETWORK FORMATION AND STRATEGIC INTERACTION IN LARGE NETWORKS

1.1 Introduction

1.1.1 Overview

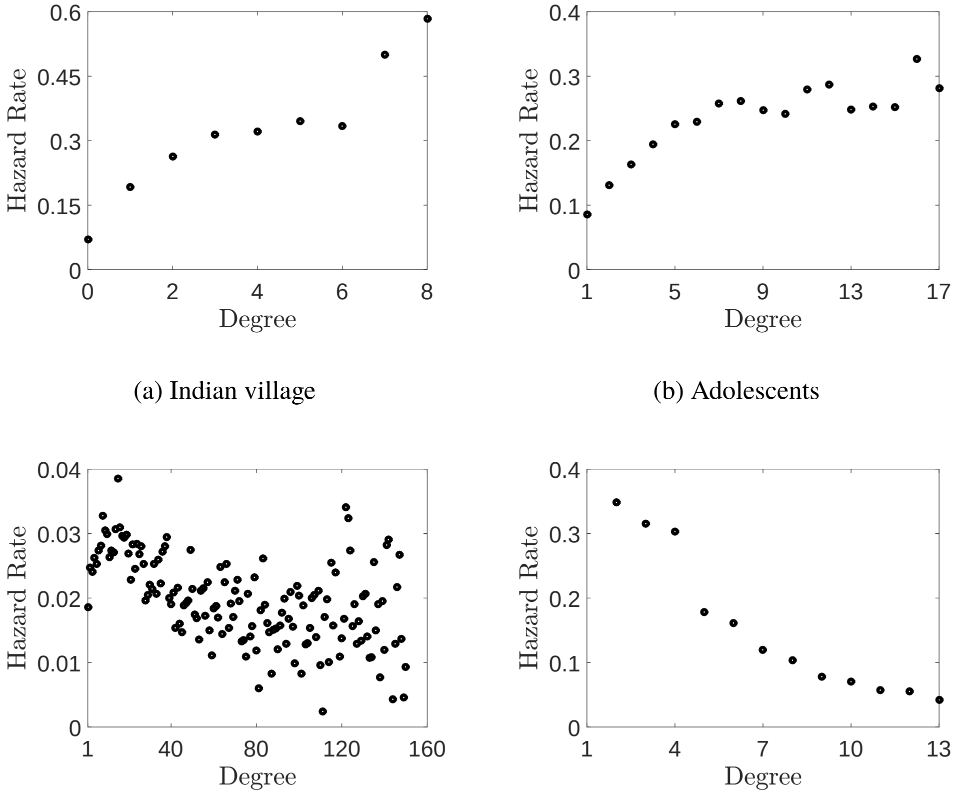
People are linked together through social relationships, and these relationships influence their economic decisions. The growing literature on *network games* analyzes various economic settings such as contagion behavior, criminal activity, political alliances, pricing of network goods, public good provision, 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* (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.

However, empirical degree distributions exhibit both increasing and decreasing

[1] For 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.



(a) Indian village

(b) Adolescents

(c) Facebook users

(d) Webpages

Figure 1.1: Different patterns of hazard rate functions

patterns of hazard rates. For example, Figure 1.1 plots the empirical hazard rate functions for four network datasets:[2] (a) one social network of a rural Indian village,[3] (b) a friendship network of adolescents in the United States, (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),

[2]In 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 1.1, I plot the hazard rate function only for degrees that account for the degrees of 95% of nodes.

[3]Source: 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 1.1 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.

[4]Sources: http://www.cpc.unc.edu/projects/addhealth/documentation is for the friendship network of adolescents in the United States; 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 webpage network at the Notre Dame University.

but decreasing patterns for (c) and (d).

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 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, 2016; Feri and Pin, 2015; Galeotti et al., 2010; Ghiglino, 2012; Jackson and Yariv, 2007). 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 distribution. The allocation rule of an optimal mechanism maximizes the *virtual value*, which is a multiplication of the usual virtual type and the *social value*. The social value represents the magnitude 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 sum of virtual types is not necessarily optimal. By increasing allocations to every customer, the seller can increase the social value. Since an increase of the social value raises the virtual value of buyers, 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 characterize the optimal mechanism in a restricted environment where the seller cannot price discriminate.

1.1.2 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.[5] 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.[6]

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 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

[5] A 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.

[6] Some 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.

condition for the IHRP is not satisfied.[7]

There have been many related papers on strategic interaction in networks that adopt the incomplete information setting introduced by Galeotti et al. (2010). Shin (2016a) is a 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 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).

[7] Since 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.

1.2 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.

1.2.1 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 $f(\cdot, \mathbf{G}) : \mathbb{N} \to [0, 1]$ 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 $F(d, \mathbf{G}) := \sum_{d' \le d} f(d', \mathbf{G})$. Last, $\overline{F}(\cdot, \mathbf{G})$ denotes the *complementary cumulative degree distribution* defined as $\overline{F}(d, \mathbf{G}) := \sum_{d' \ge d} f(d', \mathbf{G})$.

Dynamic network formation. I build a model of dynamic network formation process by recursively defining a random sequence of networks denoted by $(\mathbf{G}^t)_{t \ge 1}$. 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 $t \ge 2$ 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.[8] 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 $\frac{d_i(\mathbf{G}^t)}{\sum_{j=1}^{t} d_j(\mathbf{G}^t)}$.

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

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 $\Phi(\cdot) : \mathbb{N} \to (0,1]$ 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 with an existing node successfully.[9] 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

[9] By 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 $(\mathbf{X}_1, \ldots, \mathbf{X}_s)$, 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. Thus, the process ends in a finite length almost surely.

over the real numbers with full support. Assuming a decreasing marginal return of additional links is tantamount to $w(d)$ decreasing in d. Thus, 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.

1.2.2 The Rate Equation

Following a standard approach in the literature (e.g., Bollobás et al., 2001; Dorogovtsev et al., 2000; Ghiglino, 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 $\mathscr{G}^t$ for the probability space of undirected networks in which a random network $\mathbf{G}^t$ has its distribution. Let $(\mathscr{F}^t)_{t\geq 1}$ 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

$$\mathbf{N}(d,t) := \sum_{j=1}^{t} \mathbb{1}\{d_j(\mathbf{G}^t) = d\},$$

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

$\mathbf{N}(d,t)$ 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

$$\begin{aligned}&\mathbb{E}\left[\mathbf{N}(d,t+1) - \mathbf{N}(d,t)|\mathbf{G}^t\right] \\ &= \underbrace{\mathbb{1}\{d=1\}}_{\text{(i)}} - \underbrace{\frac{d\Phi(d)}{\mathbf{M}(t)}\mathbf{N}(d,t)}_{\text{(ii)}} + \underbrace{\frac{(d-1)\Phi(d-1)}{\mathbf{M}(t)}\mathbf{N}(d-1,t)\mathbb{1}\{d\geq 2\}}_{\text{(iii)}}. \end{aligned} \tag{1.2.1}$$

Each term in equation (1.2.1) represents the following:

(i) The degree of a new node is always 1. Thus, one additional node with degree 1 emerges.

(ii) 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

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

(iii) 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 (1.2.1) 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:[10]

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:

For $d = 1$:

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

For $d \geq 2$:

$$\mathbb{E}[\mathbf{N}(d,t+1)] = \left(1 - \frac{d\Phi(d)}{\mu t}\right)\mathbb{E}[\mathbf{N}(d,t)] + \frac{(d-1)\Phi(d-1)}{\mu t}\mathbb{E}[\mathbf{N}(d-1,t)] + \varepsilon(d,t), \tag{2.3}$$

where the correction term $\varepsilon(d,t)$ converges to zero as the network size t becomes large.[11] Moreover, as shown in the proof of Proposition 1, I can ignore the correction terms regardless of their convergence rates.

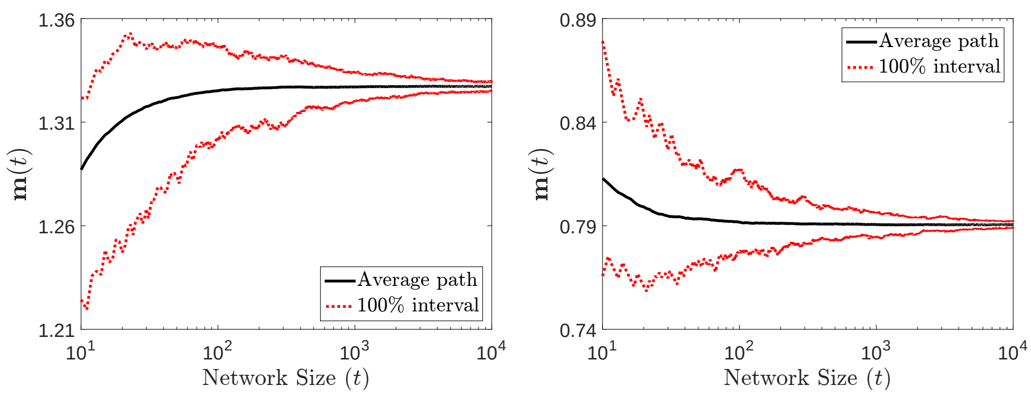
The previous dynamic network formation models using the rate equation approach make Assumption 1 implicitly (e.g., Bollobás et al., 2001; Dorogovtsev et al., 2000). For instance, in the PA model with $\Phi(d) = 1$, $\mathbf{m}(t)$ calculates the average degree in the network at the end of period t. Since only one addition link is added

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

[11]See Section A.2 for a proof.

in each period, $\mathbf{m}(t)$ converges to 2 as $t \to \infty$, and so Assumption 1 is trivially satisfied.

To evaluate the validity of Assumption 1, I present numerical simulation results in Figure 1.2. In each figure, the horizontal axis is in logarithmic scale. For $\Phi(d) = d^{-1/2}$, Figure 1.2(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 1.2(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 $\Phi(\cdot)$.[12]



(a) 1000 repetitions for $\Phi(d) = d^{-1/2}$

(b) 1000 repetitions for $\Phi(d) = d^{-3/2}$

Figure 1.2: Simulation results for parametric examples of $\Phi(\cdot)$

1.3 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.

1.3.1 Characterization of the Asymptotic Degree Distribution

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

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

[12] As 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 $\alpha = 1$.

$f(d,t)$ 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 $f(\cdot):\mathbb{N}\to[0,1]$ as the pointwise limit of $(f(\cdot,t))_{t\geq 1}$: for a fixed $\varepsilon>0$,

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

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 $f(\cdot):\mathbb{N}\to[0,1]$ 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.[13] 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*

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

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.[14] First, by using 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.

[13] See Section A.2 for a proof.

[14] Stationarity of the asymptotic degree distribution defined in Proposition 1 is obvious because it is a unique solution of the following stationarity equations:

$$\text{For } d=1: \qquad f(d)=1-\frac{d\Phi(d)}{\mu}f(d),$$
$$\text{For } d\geq 2: \qquad f(d)=-\frac{d\Phi(d)}{\mu}f(d)+\frac{(d-1)\Phi(d-1)}{\mu}f(d-1).$$

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

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

For large t, the expected number of nodes with degree 1 in period $t+1$ is approximated as[15]

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

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 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-Hoeffiding 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 $\varepsilon_d > 0$,

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

[15] I 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.

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 $f(d)$.

I finally note that there exists a unique choice of μ for Assumption 1. Since the degree distribution sequence converges, I have $\mu = \lim_{t\to\infty} \mathbb{E}\left[\sum_{d=1}^{\infty} d\Phi(d)f(d,t)\right]$. This observation in turn implies that

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

The last expression is continuous and strictly decreasing in μ.[16] Moreover, it diverges to infinity as $\mu \to 0$, but it converges to zero as $\mu \to \infty$. Therefore, the choice of μ satisfying the above equation is unique.

1.3.2 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 $h(d) := \frac{f(d)}{\overline{F}(d)}$, in which $\overline{F}(d)$ 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

[16] See Section A.2 for a proof of continuity.

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 $h(d) = 1 - \frac{\overline{F}(d+1)}{\overline{F}(d)}$, and it implies that

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

Since $f(d) = \overline{F}(d) - \overline{F}(d+1)$, 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.[17]

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

[17]The 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 $h(d) = \frac{\mu}{d\Phi(d)}$.

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. Thus, 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.

1.4 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.

1.4.1 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.[18]

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

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

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 $\delta\lambda_{\max}(\mathbf{G}) < 1$.

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, $\sum_{s=0}^{\infty} \left(\delta\lambda_{\max}(\mathbf{G})\mathbf{x}'\right)^s$, 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

[18] By 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.

of the productivity distribution. He assumes that the productivity of an idea (node) depends on its *parental* and *offspring* ideas.[19] 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 $\sigma_\theta \ll 1$.[20] Let $\mathbf{x}^t$ be the $t \times 1$ vectors with entries x_i^t. Then, $\mathbf{x}^t$ satisfies

$$\mathbf{x}^t = \left(I - \delta\theta A^t\right)^{-1} \theta \mathbb{1}_{t\times 1} = \sum_{s=0}^{\infty} \left(\delta\theta A^t\right)^s \theta \mathbb{1}_{t\times 1}. \tag{1.4.1}$$

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; 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 (1.4.1) 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 $\sigma_\theta \ll 1$.[21]

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.[22] 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

---

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

[20] I 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.

[21] König 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 $\frac{1}{\sqrt{2m(n-1)/n}}$.

[22] The 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 $n \to \infty$, assuming that $np(n)$ is a constant.

Poisson distribution of any parameter.[23] Thus, researchers have been interested in building dynamic network formation models that generate heavy-tailed degree distributions (e.g., Barabási and Albert, 1999; Ghiglino, 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 $\varepsilon > 0$,*

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

1.4.2 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.

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).[24] 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 $\frac{\Phi(1)}{2t}$:

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

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

[23]See Chapter 3 in Jackson (2010) for examples and discussions.

[24]A proof is provided in Section A.2.

second Borel-Cantelli lemma (Durrett, 2005).[25] 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 complementary degree distribution as

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

Suppose the IHRP, and so $\frac{d\Phi(d)}{\mu+d\Phi(d)}$ is decreasing in d. Then, since $\frac{\Phi(d)}{\mu+\Phi(d)} \le \frac{\Phi(1)}{\mu+\Phi(1)}$ for all d, the value of the complementary degree distribution $\overline{F}(d)$ 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*

[25] The 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 Section A.2.

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*

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

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.

1.5 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.

1.5.1 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, 2016; Feri and Pin, 2015; Galeotti et al., 2010; Ghiglino, 2012; Shin, 2016a). 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 $\widetilde{f}(\cdot) : \mathbb{N} \to [0, 1]$ the probability density function of a neighboring agent's degree, which is calculated as

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

where $\langle d \rangle = \sum_{d=1}^{\infty} df(d)$ is the average degree.[26] I call $\widetilde{f}(\cdot)$ the *conditional degree distribution*. The conditional degree distribution $\widetilde{f}(\cdot)$ captures the idea that a highly connected agent is more likely to be an agent's neighbor: $\widetilde{f}(d) > f(d)$ for all $d > \langle d \rangle$. $\widetilde{F}(\cdot)$ and $\widetilde{h}(\cdot)$ are the corresponding cumulative distribution function and the hazard rate function, respectively.

[26] To 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, $\widetilde{f}(\cdot)$ is the probability density function of a neighbor's degree.

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 $\widetilde{f}(\cdot)$.[27]

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

Lemma 1 *If $f(\cdot)$ satisfies the increasing hazard rate property, then $\widetilde{f}(\cdot)$ 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 $>_{\mathbf{LR}}$ and assume that $f_\theta(\cdot) >_{\mathbf{LR}} f_{\theta'}(\cdot)$ if $\theta > \theta'$.[28] For each degree distribution $f_\theta(\cdot)$, let $\widetilde{f}_\theta(\cdot)$ be the corresponding conditional degree distribution. Similarly, I denote by $h_\theta(\cdot)$ and $\widetilde{h}_\theta(\cdot)$ the corresponding hazard rate functions.

The likelihood ratio order has the following three useful properties:

(i) The likelihood ratio order between two degree distributions is preserved for the corresponding conditional degree distributions:[29]

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

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

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

[29] For all $d > d'$,

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

(ii) 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:[30]

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

(iii) 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:

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

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

When agent i with degree d_i chooses action $x_i \in \mathbb{R}_+$, her expected utility is

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

where $\widetilde{\mathbb{E}}[\sigma] := \sum_{d=1}^{\infty} \sigma(d)\widetilde{f}(d)$ 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,*

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

1.5.2 Properties of Equilibria

The expected utility of agent i with degree d_i is $U(x_i, \sigma, d_i) = x_i - \frac{1}{2}x_i^2 + \delta d_i \widetilde{\mathbb{E}}[\sigma]\, x_i$ 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:

$$x^B(\sigma, d_i) = 1 + \delta d_i \widetilde{\mathbb{E}}[\sigma]\,. \tag{1.5.1}$$

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

[30] The 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 $\widetilde{f}_\theta(\cdot) \not>_{\mathbf{FOSD}} \widetilde{f}_{\theta'}(\cdot)$ because $\widetilde{F}_\theta(1) = 0.25 > \widetilde{F}_{\theta'}(1) = 0.15$ but $\widetilde{F}_\theta(2) = 0.58 < \widetilde{F}_{\theta'}(1) = 0.62$.

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 (1.5.1), it must hold that

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

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

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

Thus, 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 $\widetilde{\mathbb{E}}[\sigma^1] = 1$. 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 $\widetilde{\mathbb{E}}[\sigma^2] = 1 + \delta\frac{\langle d^2\rangle}{\langle d\rangle}$. Thus, 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.[31] 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

[31] This 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, 2016a). 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.

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.[32] 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.*

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

[32] For 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.

of finding the second moment is $O(n^2)$.[33] 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.[34]

1.6 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 assuming its IHRP, I characterize an optimal mechanism.

1.6.1 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* (*mechanisms*): buyers directly report their types, and an allocation vector and a payment

[33] The 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.

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

vector are determined according to a pre-determined rule. 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 $\mathbf{x} = (x_1, \ldots, x_n)$ where $x_i(\cdot) : \mathcal{D}^n \to [0,1]$ is an allocation rule for buyer i. Similarly, the payment scheme is denoted by $\mathbf{p} = (p_1, \ldots, p_n)$ where $p_i(\cdot) : \mathcal{D}^n \to \mathbb{R}_+$ is a payment scheme for buyer i. Therefore, when the reported profile is $\mathbf{d} = (d_1, \ldots, d_n)$, buyer i obtains $x_i(\mathbf{d})$ 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 $\xi_i(\cdot) : \mathcal{D} \to [0,1]$ and the conditional expected payment function $\pi_i(\cdot) : \mathcal{D} \to \mathbb{R}_+$ as

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

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

$$V_i(d_i', d_i) = \xi_i(d_i') \sum_{j \in N_i} \widetilde{\mathbb{E}}\left[\xi_j(d_j)\right],$$

where $\widetilde{\mathbb{E}}\left[\xi_j(d_j)\right] := \sum_{d \in \mathcal{D}} \widetilde{f}(d_j)\xi_j(d_j)$ 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, 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 $\pi(d_i')$. Thus, for a buyer of type d, her expected value from reporting d' is

$$V(d', d) = \xi(d') d \widetilde{\xi},$$

where $\widetilde{\xi} = \sum_{d=1}^{d_{\max}} \xi(d) \widetilde{f}(d)$ is the expected allocation of a neighboring buyer.

A mechanism (ξ, π) is called *incentive compatible* if

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

In addition, a mechanism is called (interim) *individually rational* if for all $i \in N$,

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

Since the seller's per-capita expected revenue is $\sum_{d=1}^{d_{\max}} f(d)\pi(d)$, his mechanism design problem is formulated as

$$\begin{aligned}
\underset{(\xi,\pi)}{\text{maximize}} \quad & \sum_{d=1}^{d_{\max}} f(d)\pi(d) \\
\text{subject to} \quad & V(d,d) - \pi(d) \geq V(d',d) - \pi(d') \text{ for all } d', d \in \mathcal{D}, \\
& V(d,d) - \pi(d) \geq 0 \text{ for all } d \in \mathcal{D}, \\
& \xi(d) \in [0,1] \text{ for all } d \in \mathcal{D}.
\end{aligned} \tag{1.6.1}$$

1.6.2 Revenue-Maximizing Mechanism

For any mechanism generating a positive revenue to the seller, $V(\cdot,\cdot)$ 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:[35]

$$V(d,d) - \pi(d) \geq V(d+1,d) - \pi(d+1) \text{ for all } d = 0,\ldots,d_{\max}-1 \tag{1.6.2}$$

$$V(d,d) - \pi(d) \geq V(d-1,d) - \pi(d-1) \text{ for all } d = 1,\ldots,d_{\max}. \tag{1.6.3}$$

Needless to say, all the downward incentive compatibility constraints (1.6.3) 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

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

for all $d \geq 1$, where $\widetilde{\xi} = \sum_{d=1}^{d_{\max}} \xi(d)\widetilde{f}(d)$ 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 (1.6.2) are satisfied. Thus, the seller's problem becomes

$$\begin{aligned}
\underset{\xi:\mathcal{D}\to[0,1]}{\text{maximize}} \quad & \left\{ \sum_{d'=1}^{d_{\max}} \widetilde{f}(d')\xi(d') \right\} \left[\sum_{d=0}^{d_{\max}} f(d) \left(\xi(d) \Big(d - \frac{1-F(d)}{f(d)} \Big) \right) \right] \\
\text{subject to} \quad & 0 \leq \xi(0) \leq \xi(1) \leq \cdots \leq \xi(d_{\max}) \leq 1.
\end{aligned} \tag{1.6.4}$$

[35] See Chapter 6 in Vohra (2011) for a proof.

The seller's objective function can be rewritten as

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

where $\widetilde{\xi} = \sum_{d=1}^{d_{\max}} \xi(d)\widetilde{f}(d)$ is the expected allocation of a neighboring buyer. The term presents in the summation, $\widetilde{\xi}(d - \frac{1-F(d)}{f(d)})$, is the *virtual value* of a buyer of type d. Due to the allocative externalities between neighboring buyers, the virtual value consists of two components: $(d - \frac{1-F(d)}{f(d)})$ and $\widetilde{\xi}$. The first component that does not depend on other buyers' allocations is the virtual type (Myerson, 1981). The second component $\widetilde{\xi}$ 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 $\widetilde{\xi}$ 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:

$$\begin{aligned} &\underset{\xi:\mathcal{D}\to[0,1]}{\text{maximize}} && \sum_{d=1}^{d_{\max}} f(d) \left(\xi(d)\Big(d - \frac{1 - F(d)}{f(d)}\Big) \right) \\ &\text{subject to} && 0 \le \xi(0) \le \xi(1) \le \cdots \le \xi(d_{\max}) \le 1. \end{aligned} \tag{1.6.5}$$

The solution of the above problem maximizes $\xi(d)(d - \frac{1-F(d)}{f(d)})$ for each d (Myerson, 1981). Note that the virtual type $(d - \frac{1-F(d)}{f(d)})$ 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 (1.6.5) does not maximize the seller's true per-capita revenue in (1.6.4). 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 $\widetilde{f}_\theta(\cdot) >_{\mathbf{FOSD}} \widetilde{f}_{\theta'}(\cdot)$.

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, $\widetilde{f}_\theta(\cdot) >_{\mathbf{FOSD}} \widetilde{f}_{\theta'}(\cdot)$, implies that the social value strictly increases as

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

These two observations imply that for any given allocation rule, the seller's objective function in (1.6.4) 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.*

1.6.3 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. Thus, 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 $\pi = V(\underline{d}, \underline{d}) = \underline{d}(1 - \widetilde{F}(\underline{d} - 1))$, and the seller's revenue function for a given threshold degree d is

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

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 $\triangle\Psi(d) := \Psi(d+1) - \Psi(d)$. The marginal revenue satisfies (i) $\triangle\Psi(0) > 0$, and (ii) a single-crossing property that $\triangle\Psi(d) \leq 0$ implies $\triangle\Psi(d') < 0$ 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

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

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

$$\frac{1}{d^*+1} = h(d^*) + \widetilde{h}(d^*) - \widetilde{h}(d^*)h(d^*). \tag{1.6.6}$$

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 $\widetilde{\xi}$, 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 $d = \pi/\widetilde{\xi}$, 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 $\widetilde{\xi}d(1-F(d-1))$. The seller's revenue is uniquely maximized at d' satisfying $\frac{1}{d'+1} = h(d')$, and it is independent of the social value $\widetilde{\xi}$. Therefore, the first term in equation (1.6.6) 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 (1.6.6), the latter term $\left(\widetilde{h}(d) - \widetilde{h}(d)h(d)\right)$ captures this factor. Since $\widetilde{h}(d) - \widetilde{h}(d)h(d)$ is strictly positive, the threshold degree d^* solving equation (1.6.6) is strictly smaller than d'.

The optimal choice of threshold degree d^* as strictly smaller than d' has the following meaning. Let $\pi^* = d^*(1 - \widetilde{F}(d^* - 1))$ and $\pi' = d'(1 - \widetilde{F}(d' - 1))$ be the corresponding prices. Since the degree distribution $f(\cdot)$ satisfies the strictly IHRP, it follows that $\pi^* < \pi'$.[36] 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.*

1.7 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.

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

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

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 compared to what the previous strategic network formation models predict.

Chapter 2

MONOPOLY PRICING AND DIFFUSION OF (SOCIAL) NETWORK GOODS

2.1 Introduction

2.1.1 Overview

There are many goods for which a consumer's valuation is influenced by her existing social relationships. For instance, when a consumer is thinking of joining an online communication service (e.g., Skype or WhatsApp), she considers how many of her friends and co-workers are currently subscribing the same service. Because of this interdependency of valuations, demand for a good depends on social relationships between all consumers. Social relationships between all consumers can be represented by a *social network* that consists of consumers and a set of links between them. It then follows that a monopolistic seller would factor social network structures into his optimal marketing strategy.

In this paper, I analyze an optimal dynamic pricing strategy for a *subscription social network good* sold by a monopolist. In the model, the good is a social network good in the sense that positive network effects are generated only between consumers sharing a link within a social network. Subscription means that in each period, consumers need to pay a subscription price (e.g., monthly service fees of cell phone services) to use the good. I assume that in each period, there is only one subscription price that applies to all the consumers, and that each consumer myopically best responds to population behavior in the previous period (mean-field approximation). Under these assumptions, the monopolist maximizes the discounted sum of per-period profits by choosing a sequence of subscription prices, which I call a *pricing plan*.

The current monopoly market for a subscription social network good is different from other types of network good markets. First, in contrast with a network good market in which each consumer benefits from all other subscribers, a consumer benefits only from her neighboring consumers' subscription. I assume that consumers know only their own degree and are uncertain about the subscription decisions of other consumers. Thus, other things equal, a consumer's value from subscribing to the good depends on her *degree*, the number of her neighboring consumers. For this, in the model, consumer heterogeneity is summarized by the *degree distribution* of the social network. Thus, the monopolist takes the degree distribution into account when determining his optimal pricing plan.

Second, in contrast with a durable good market in which a consumer permanently leaves the market after her initial purchase, no consumer permanently leaves the market for a subscription good. As a consequence, the monopolist does not compete with his future selves in demand. Rather, as a function of the degree distribution, his dynamic problem is how to adjust the size of positive network effects by changing subscription prices over time.

In my model, the monopolist balances an intertemporal tradeoff in each period: maximizing the current profit versus increasing future profits by encouraging more consumers to subscribe to the good. In period t, if the monopolist lowers the subscription price below the price that maximizes profit in period t, more consumers subscribe to the good. This will lead consumers to expect that their neighboring consumers are more likely to subscribe to the good in period $t + 1$; the demand shifts up, and it provides a chance for the monopolist to obtain a higher profit in period $t + 1$. The monopolist balances the above tradeoff across periods in order to maximize his discounted sum of per-period profits.

I characterize optimal pricing plans and analyze their dynamic properties. Interestingly, any optimal pricing plan has the property that if at some period the subscription price is lower than its steady state level, then it will become higher than its steady state level within finite periods, and vice versa. This pattern originates from the fact that, for example, when the size of network effects is high due to a high subscription rate in the previous period, the monopolist has an incentive to set a high subscription price in order to obtain a higher profit in that period. Moreover, since an optimal pricing plan oscillates around its steady state level, the resulting subscription rate also exhibits oscillating patterns.

The rest of my analysis focuses on properties of the unique steady state of the monopoly market where both the monopolist and consumers do not change their decisions.[1] First, the subscription rate at the steady state of the market is at the highest level that is consistent with the steady state price level. Because of this, the steady state is robust to small perturbations to consumers' belief in the subscription rate. Second, I find a closed-form expression for the deadweight loss from the monopoly that consists of two parts: welfare loss from excluded consumers and welfare loss from the lost network effects. Third, I examine the effects of changes

[1] Since I consider an incomplete information setting, the unique steady state can be interpreted as a perfect Bayesian equilibrium (Jackson and Yariv, 2007).

in the monopolist's discount factor and the density of social networks in terms of the likelihood ratio property.

2.1.2 Related Literature

The current paper is related to the literature on diffusion in social networks. Bass (1969) introduces a non-strategic diffusion model of product adoptions, and expresses changes of the adoption rate by using ordinary differential equations. Granovetter (1978) proposes an alternative dynamic model where in each period, each agent best responds to population behavior in the previous period. Both Bass (1969) and Granovetter (1978) do not account for the impact of social network structures on diffusion outcomes. Several recent papers highlight the impact of particular social network structures (e.g., Jackson and Rogers, 2007b; Jackson and Yariv, 2007; López-Pintado, 2008; Young, 2009). Although these models can be used to analyze consumers' adoption behavior, they do not introduce a firm that might influence diffusion processes in order to maximize his profit. I take the framework by Jackson and Yariv (2007) and consider a monopolist who changes subscription incentives by changing subscription prices. This leads my model to predict a unique steady state.

The current paper is also related to the growing literature on optimal marketing strategy when consumers, connected to one another in a social network, influence one another (e.g., Fainmesser and Galeotti, 2016; Campbell, 2013).[2] Fainmesser and Galeotti (2016) consider a *static* price discrimination problem rather than the dynamic setting I consider. Specifically, they consider a monopolist who knows the in-degree and out-degree distributions of the underlying social network. They show that the average and variance of the two degree distributions are sufficient statistics to characterize the optimal price discrimination rule. In contrast, in my dynamic setting, the steady state price level depends on the hazard rate function.

Campbell (2013) models the diffusion process as a *percolation* process, and finds the value of informed consumers asymptotically. A monopolist can influence the percolation process by changing a purchasing price *only at the outset of the process*. However, in my model, the monopolist optimally sets a subscription price in every period, depending on the present subscription rate. Because of this, the

[2] In the computer science literature, several papers study the complexity of finding an optimal strategy in various settings. For example, Hartline et al. (2008) and Arthur et al. (2009) consider a setting in which a monopolist sequentially approaches consumers and offers a different price for each consumer. Since finding an optimal strategy is NP-hard, the authors of these papers propose an algorithm that generates a simple strategy that returns approximately the maximum expected revenue.

comparative static results on the steady state rely on different statistics of the degree distribution.

Conceptually, the current paper is related to the classical literature on dynamic pricing of *subscription* network goods sold by a monopolist.[3] Rohlfs (1974) is one of the first to inspect consequences of dynamic pricing and diffusion. Dhebar and Oren (1985) show that assuming a particular subscription pattern, an optimal pricing plan monotonically increases to the steady state price level. Similarly, Radner et al. (2014) prove that assuming a concave distribution of consumer heterogeneity, an optimal pricing plan monotonically decreases (increases) to the steady state price level if the initial subscription rate is above (below) the subscription rate at the steady state. In my model, consumer utilities derived from subscribing to the good depend on connectivity in the social network. Moreover, my model does not assume a particular subscription pattern, so that I obtain oscillating optimal pricing plans instead of monotone plans in the previous papers.

2.2 The Model

Social network. I consider a finite (but large) set of consumers $N = \{1, \ldots, i, \ldots, n\}$. Social relationships between the consumers are represented by a *social network* $\langle N, L \rangle$, where $L \subseteq N \times N$ is the set of *links* that represents pairwise social relationships among the consumers. The social network is *undirected*: $(i, j) \in L$ if and only if $(j, i) \in L$. For consumer i, $N^i := \{j \in N | (i, j) \in L\}$ is the set of *neighbors*, and its cardinality $x^i := |N^i|$ is called consumer i's *degree*. As will be explained later, a key property of the social network is its *degree distribution* $f : \mathbb{N}_0 \to \mathbb{R}^+$, where $f(x)$ represents the fraction of consumers with degree x.[4] To obtain clear results, I assume that the degree distribution is approximated by a continuous probability density function, which has full support $\mathbb{R}_0^+$ and is continuously differentiable.[5] I denote by F the corresponding *cumulative degree distribution*. I define a function $\widetilde{f} : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ as $\widetilde{f}(x) := \frac{x f(x)}{\langle x \rangle}$, where $\langle x \rangle$ is the expected degree under the degree

[3] There are several papers (e.g., Bensaid and Lesne, 1996; Cabral et al., 1999) that find conditions under which an optimal pricing of a *durable* network good is increasing, contrary to the Coase theorem (Coase, 1972). Bensaid and Lesne (1996) find that if the network effect on consumer utility is substantial, the monopolist can credibly increase prices as time passes. Cabral et al. (1999) propose a two-period model, which shows that if consumers have incomplete information about the demand size, there exists a unique perfect Bayesian equilibrium in which price in the second period is higher than price in the first period. See surveys by Katz and Shapiro (1994), Farrell and Klemperer (2007), and Shy (2011).

[4] I use the following notation: $\mathbb{R}^+ := (0, \infty)$ and $\mathbb{R}_0^+ := \{0\} \cup \mathbb{R}^+$.

[5] This approximation is reasonable when the population size n is large. For example, the degree distribution of an undirected Erdős-Rényi graph $G(n, p)$ is approximated by the normal distribution $N(np, np)$.

distribution f. I call $\widetilde{f}$ the *conditional degree distribution*, and $\widetilde{f}(x)$ represents the conditional probability that a consumer's degree is x, conditional on being at the end of a randomly and independently chosen link in the social network.

I take the following assumption of the *increasing hazard rate property* (IHRP):

Assumption 2 *For all $x \in \mathbb{R}^+$,*

$$\frac{d}{dx}\left(\frac{f(x)}{1-F(x)}\right) \geq 0.$$

The IHRP of f indicates that, among consumers having at least x neighbors, the conditional fraction of consumers with degree exactly x increases as x increases. Several degree distributions exhibit the IHRP. For example, an exponential degree distribution generated through a random attachment model satisfies the IHRP.[6] A degree distribution generated by the Watts-Strogatz model (Watts and Strogatz, 1998) also satisfies the IHRP. Moreover, several empirical degree distributions exhibit the IHRP (Shin, 2016b).[7]

The following lemma provides useful properties of the hazard rate functions of f and $\widetilde{f}$ that will play important roles throughout the paper.

Lemma 2 *If f satisfies the IHRP, then the following properties hold:*

(i) The strictly IHRP of $\widetilde{f}$: for all $x \in \mathbb{R}^+$,

$$\frac{d}{dx}\left(\frac{\widetilde{f}(x)}{1-\widetilde{F}(x)}\right) > 0. \tag{2.2.1}$$

(ii) The dominance relationship: for all $x \in \mathbb{R}^+_0$,

$$\frac{f(x)}{1-F(x)} > \frac{\widetilde{f}(x)}{1-\widetilde{F}(x)}. \tag{2.2.2}$$

(iii) The single crossing property: there exist $\overline{x}, \widetilde{x} \in \mathbb{R}^+$ such that $\overline{x} < \widetilde{x}$ and

$$\frac{1}{\overline{x}} = \frac{f(\overline{x})}{1-F(\overline{x})} \quad \textit{and} \quad \frac{1}{\widetilde{x}} = \frac{\widetilde{f}(\widetilde{x})}{1-\widetilde{F}(\widetilde{x})}. \tag{2.2.3}$$

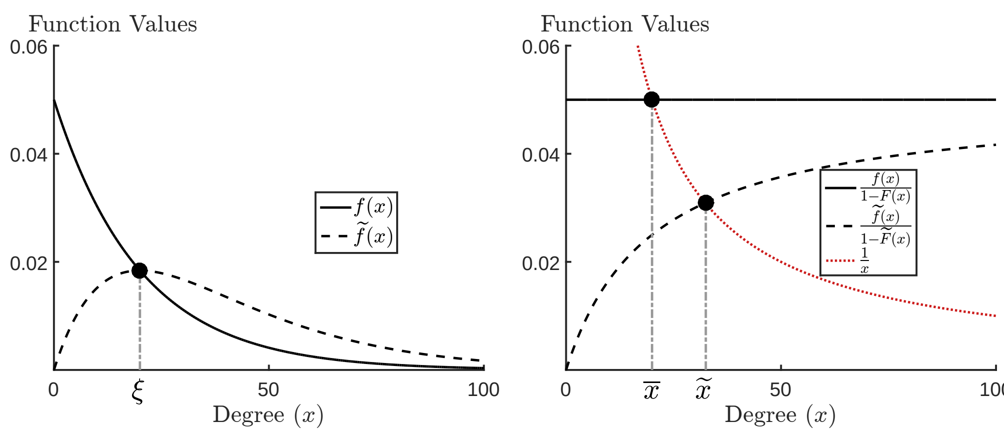
[6] A simple discrete-time random attachment model generates a variant of exponential degree distributions in expectation. In the model, nodes newly join to existing networks and form links to the existing nodes. Then, in expectation, the resulting degree distribution becomes an exponential distribution. See Jackson (2010) for details.

[7] Of course, there are some degree distributions not satisfying the IHRP. A scale-free distribution generated by the preferential attachment model (Barabási and Albert, 1999) is an example.

The strictly IHRP of $\widetilde{f}$ has a probabilistic interpretation that knowing that a consumer's neighbor has at least x neighbors, the conditional probability that the neighbor has degree exactly x increases as x increases. The hazard rate function of $\widetilde{f}$ is strictly smaller than that of f as stated in (2.2.2) because $\widetilde{f}$ is calculated based on additional information that a neighbor's degree is not zero. The single crossing property is a direct consequence of (2.2.1) and (2.2.2). The following example illustrates a case in which the degree distribution is an exponential degree distribution, and properties in Lemma 2 are easily illustrated.

Example (Exponential Degree Distribution) Consider the case of $f(x) = \frac{e^{-x/\xi}}{\xi}$. Then, $\widetilde{f}(x) = \frac{xe^{-x/\xi}}{\xi^2}$, and the corresponding hazard rate functions are calculated as

$$\frac{f(x)}{1-F(x)} = \frac{1}{\xi} \quad \text{and} \quad \frac{\widetilde{f}(x)}{1-\widetilde{F}(x)} = \frac{x}{\xi(\xi+x)}.$$



(a) $f(x)$ and $\widetilde{f}(x)$

(b) hazard rate functions of f and $\widetilde{f}$

Figure 2.1: Illustration of Lemma 2

Figure 2.1(a) illustrates $f(x)$ and $\widetilde{f}(x)$ when $\xi = 20$. Observe that $\widetilde{f}(x)$ (the dotted line) dominates $f(x)$ (the solid line) for all $x \geq \langle x \rangle = \xi$. This captures the idea that in a social network, highly connected consumers are involved in a larger fraction of the links.[8] As depicted in Figure 2.1(b), the hazard rate functions of f and $\widetilde{f}$ satisfy properties in Lemma 2: (i) the hazard rate function of $\widetilde{f}$ (the black

[8]One may interpret this result as a probabilistic version of *friendship paradox* in the social network literature. Roughly speaking, the friendship paradox states that on average, most people have fewer friends than their friends have. See Feld (1991) for introduction of the related literature and Hodas et al. (2013) for evidence in recent social network services.

dotted line) is strictly increasing in x, (ii) it is strictly smaller than the hazard rate function of f (the solid line), and (iii) there are two crossing points, $\overline{x}$ and $\widetilde{x}$.

Network good market. I consider a monopoly market for a *subscription social network good*. The monopolist produces the good at zero marginal production cost and sells it to the consumers for infinitely many periods $t = 1, 2, 3, \ldots$. The monopolist's discount factor is denoted by $\beta \in (0, 1)$. The monopolist chooses a *pricing plan*, a sequence of non-negative subscription prices $\{p_t\}_{t=1}^{\infty}$, to maximize his discounted sum of per-period profits. I will formulate the monopolist's dynamic optimization problem at the end of the current section.

The good is a *subscription* good, which indicates that a consumer needs to pay a subscription price p_t to use the good in each period t. Without loss of generality, I assume that a consumer obtains a utility of zero if she does not subscribe to the good. To qualify as a *social network good*, I assume that positive network effects are generated only between consumers sharing a link. Formally, if a consumer with degree x subscribes to the good, she gets a per-period utility of $u_x(m, p_t) = m - p_t$, where m ($\leq x$) is the number of consumers who share a link with her as well as subscribe to the good in period t. Without loss of generality, I assume that in each period, consumers maximize their per-period utility.[9]

Diffusion process and the monopolist's problem. For a given pricing plan $\{p_t\}_{t=1}^{\infty}$, I model consumers' subscription decisions as a *diffusion process* on the social network by employing the approach proposed by Jackson and Yariv (2007). In terms of consumers' information about the underlying social network, I assume that the consumers know their own degrees but not their neighbors' degrees, so that each consumer is uncertain about interactions of other consumers including her neighbors. Specifically, I consider consumers who believe that (i) there is no correlation between consumers' degrees, and that (ii) each neighbor's degree is an independent random draw from the conditional degree distribution across periods.[10] The former means that given that two consumers i and j share a link, consumer i believes that consumer j's degree is x with the probability $\widetilde{f}(x)$, and vice versa.[11]

[9]To see why this assumption can be made without loss of generality, suppose that a consumer maximizes her discounted sum of per-period utilities. Suppose also that her per-period utility is strictly greater than zero if she subscribes to the good in period t. In this case, cancelling her subscription in period t strictly decreases her discounted sum of utilities because by doing so, she obtains a utility of zero in period t, but her future utilities remain the same, independent of this cancellation. Thus, only the per-period utility matters for her decision in each period.

[10]These two assumptions rule out any possibility that consumers can learn their neighbors' degrees from realized utilities in previous periods.

[11]The current incomplete information setting is used to model strategic interactions in a large

With the above incomplete information structure, I consider the following *mean-field approximation* of the diffusion process in which the consumers myopically best respond to the population subscription rate in the previous period.[12] As an initial condition, suppose that any consumer with degree at least x_0 subscribes to the good; or equivalently, suppose that a strictly positive fraction $1 - \widetilde{F}(x_0)$ of the consumers is randomly chosen to subscribe to the good.[13] In period $t = 1$, x_0 and p_1 are known to all consumers. Then, consumers with degree at least $x_1 = \frac{p_1}{1-\widetilde{F}(x_0)}$ subscribe to the good because the expected utility of a consumer with degree x is

$$\left[\sum_{m=0}^{x} (1 - \widetilde{F}(x_0))^m \widetilde{F}(x_0)^{x-m}\right] - p_1 = x(1 - \widetilde{F}(x_0)) - p_1,$$

presuming that each neighboring consumer subscribes to the good with probability $1 - \widetilde{F}(x_0)$.

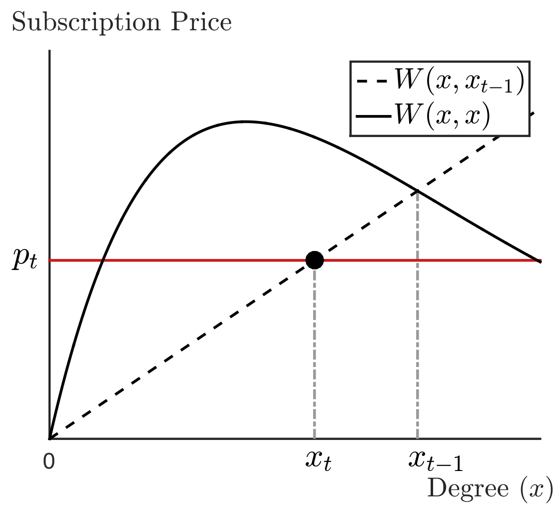


Figure 2.2: The dynamic demand structure

Generally, for given p_t and x_{t-1}, a new threshold degree x_t is uniquely determined and satisfies $p_t = W(x_t, x_{t-1})$, where $W(x, x') := x(1 - \widetilde{F}(x'))$ is the willingness-to-pay (or, the inverse demand) of consumers with degree x when each neighboring

social network. In fact, there is some evidence suggesting that people are not perfectly informed of their neighbors' connectivity even in a small world network (e.g., Kumbasar et al., 1994).

[12]The myopia considered in this paper is only about consumers' beliefs on their neighbors' decisions. As explained by Dhebar and Oren (1985), there are three factors affecting a consumers' subscription decision in each period when population size n is sufficiently large: her degree, the current subscription price, and her expectation about her neighbors' current subscription decisions. Among the three factors, the former two do not require a myopia assumption because a consumer's degree is time-invariant, and the current subscription price is observable.

[13]This can be done, for example, by distributing free units of the good to $1 - F(x_0)$ fraction of the consumers.

consumer subscribes to the good with probability $1 - \widetilde{F}(x')$.[14] As illustrated in Figure 2.2, this one-to-one correspondence between p_t and x_t for a given threshold degree x_{t-1} determines the demand size as $1-F(x_t)$ in period t, which is the fraction of consumers with degree at least x_t.[15]

Note that either a pricing plan or the corresponding sequence of threshold degrees determined by the diffusion process returns the same discounted sum of profits to the monopolist. For this reason, I formulate the monopolist's dynamic optimization problem as an optimal choice of a sequence of threshold degrees $\{x_t\}_{t=1}^{\infty}$, which I call a *diffusion plan*. Therefore, the monopolist's dynamic optimization problem is formulated as

$$\underset{\{x_t\}_{t=1}^{\infty}}{\text{maximize}} \quad \sum_{t=1}^{\infty} \beta^{t-1} \pi(x_{t-1}, x_t),$$

where $\pi(x, y) = y\big(1 - \widetilde{F}(x)\big)\big(1 - F(y)\big)$ is the per-period profit function.

2.3 Preliminary Results

In this section, I first provide characterizations of an optimal diffusion plan and its unique steady state level. Then, I analyze dynamic properties of optimal diffusion and pricing plans.

2.3.1 Optimal Diffusion Plan and Unique Steady State

I find that the monopolist never chooses $x_t > \overline{x}$ in any period t, so that any optimal diffusion plan must be in a compact interval $[0, \overline{x}]$ where $\overline{x}$ solves equation $\frac{1}{x} = \frac{f(x)}{1-F(x)}$. To see why, let π_i denote the partial derivative of π with respect to the i-th variable. By the single crossing property in Lemma 2, $\pi_2(x, y) \geq 0$ if and only if $y \in [0, \overline{x}]$. This observation implies that $\pi(x, y)$ is single-peaked at $y = \overline{x}$, so that the monopolist never chooses $x_t > \overline{x}$ in any period t.[16] This ensures that an optimal

[14] Recall that each neighbor's degree is an independent random draw from the conditional degree distribution across periods. In addition, there is no correlation between neighboring consumers' degrees. Thus, for a consumer with degree x, the number of neighbors subscribing to the good in period t follows the binomial distribution with parameters x and $1-\widetilde{F}(x_{t-1})$. Therefore, the expected number of subscribing neighbors is $x(1 - \widetilde{F}(x_{t-1}))$.

[15] This mean-field approximation of the diffusion process is different from that considered in Jackson and Rogers (2007b) and López-Pintado (2008). In these papers, the mean-field approximation does not require an incomplete information setting because agents simply myopically react to what they observe in the previous period. In contrast, the current mean-field approximation with an incomplete information structure corresponds to that in Jackson and Yariv (2007), which induces the property that consumers with the same degree take the same action.

[16] Let a diffusion plan $\{x_t\}_{t=1}^{\infty}$ with $x_{t'} > \overline{x}$ for some $t' \geq 1$ be given. Define an alternative diffusion plan $\{\widehat{x}_t\}_{t=1}^{\infty}$ as $\widehat{x}_t = \overline{x}$ if $t = t'$, and $\widehat{x}_t = x_t$ otherwise. Then, the alternative diffusion plan returns a strictly greater discounted sum of per-period profits than the given diffusion plan.

diffusion plan $\{x_t^*\}_{t=1}^{\infty}$ exists.

Moreover, an optimal diffusion plan is contained in the interior of $[0, \overline{x}]$ because the monopolist has an incentive to sacrifice current profit for a higher future profit by choosing x_t strictly smaller than $\overline{x}$. By doing so, the profit in period t is not maximized, but the profit in period $t+1$ increases because consumers are more willing to subscribe to the good due to increased subscription probability of neighboring consumers. In addition, the monopolist never chooses x_t equal to zero because it is not profitable to sell the good to the solitary consumers with degree zero. Therefore, x_t^* must be contained in $(0, \overline{x})$ for all $t \geq 1$.

Since any optimal diffusion plan is in $(0, \overline{x})$, I characterize an optimal diffusion plan by employing a Bellman approach as follows. A value function $V : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is defined as

$$V(x) := \max_{y \in [0, \overline{x}]} \{\pi(x, y) + \beta V(y)\},$$

where $V(x)$ represents the maximum discounted sum of profits that the monopolist can achieve when the consumers believe that consumers with degree at least x are expected to subscribe to the good. Assuming that V is differentiable, x_t^* solves a first-order condition,

$$\pi_2(x_{t-1}^*, x_t^*) + \beta V'(x_t^*) = 0,$$

for given x_{t-1}^*. Along with an envelope condition, $V'(x_t^*) = \pi_1(x_t^*, x_{t+1}^*)$, I characterize an optimal diffusion plan as follows.

Proposition 8 *If the value function is differentiable, then there exists an optimal diffusion plan* $\{x_t^*\}_{t=1}^{\infty}$ *such that*

$$\pi_2(x_{t-1}^*, x_t^*) + \beta \pi_1(x_t^*, x_{t+1}^*) = 0.$$

Since the per-period profit function is not necessarily concave on its whole domain, it is not obvious whether V is differentiable. Introducing a sufficient condition for the differentiability of V is delayed until the end of the current section because it depends on a feature of optimal diffusion plans discussed below.

The optimality condition in Proposition 8 represents how the monopolist balances intertemporal tradeoffs. For given x_{t-1}^* and x_{t+1}^*, there are two effects of a change in x_t on the monopolist's discounted sum of profits. First, an increase of x_t within the interval $(0, \overline{x})$ strictly increases the monopolist's profit in period t because $\pi(x_{t-1}^*, x_t)$ is single-peaked at $x_t = \overline{x}$. The magnitude of this effect is $\pi_2(x_{t-1}^*, x_t)$.

Second, an increase of x_t will strictly decrease the profit in period $t+1$ because consumers will be less willing to subscribe to the good. Before discounting, the magnitude of this effect is $\pi_1(x_t, x_{t+1}^*) = -\widetilde{f}(x_t)x_{t+1}^*(1-F(x_{t+1}^*))$. I normalize the magnitudes of these two effects by dividing them by $\widetilde{f}(x_t)$. After this normalization, I call the first effect a *profit effect*, which strictly decreases to zero as x_t increases to $\overline{x}$. I call the second effect a *friendship effect*, which is now independent of x_t. Therefore, an optimal choice of x_t^* is unique and balances the two effects:

$$\underbrace{\frac{\pi_2(x_{t-1}^*, x_t^*)}{\widetilde{f}(x_t^*)}}_{\textbf{profit effect}} = \beta \underbrace{x_{t+1}^*(1-F(x_{t+1}^*))}_{\textbf{friendship effect}}. \tag{2.3.1}$$

I call a fixed point of equation (2.3.1) a *steady state diffusion level*, which represents that when the consumers believe that consumers with degree at least x^* are expected to subscribe, the monopolist perfectly balances the profit effect and the friendship effect without any change in the diffusion plan. The corresponding *steady state price level* is $p^* = W(x^*, x^*)$ that makes a marginal consumer with degree x^* indifferent between subscription and non-subscription. The following proposition finds that a unique steady state diffusion level exists, and that this level is characterized by the hazard rate functions of f and $\widetilde{f}$.

Proposition 9 *If the value function is differentiable, then there exists a unique steady state diffusion level x^* such that*

$$\frac{1}{x^*} = \frac{f(x^*)}{1-F(x^*)} + \beta\frac{\widetilde{f}(x^*)}{1-\widetilde{F}(x^*)}. \tag{2.3.2}$$

The uniqueness directly follows from the IHRPs of f and $\widetilde{f}$. The second term in the right-hand side of the equation in Proposition 9 captures the magnitude of the discounted friendship effect that the monopolist takes into account in exchange for sacrificing the per-period profit maximization. Before discounting, this term solely depends on $\widetilde{f}$ because consumers' belief about their neighbors' subscription probability depends on $\widetilde{f}$.

I finally introduce a sufficient condition under which the value function is differentiable.[17] Although the per-period profit function is not concave on its whole domain, the current model satisfies all the conditions for Theorem 3 and Theorem 4

[17] A proof is provided in the proof for Proposition 8. In fact, Theorem 4 in Milgrom and Segal (2002) provides that V is absolutely continuous, which further implies that V is differentiable except on a set of Lebesgue measure zero.

in Milgrom and Segal (2002). Theorem 3 confirms that directional derivatives are well-defined at each x_t^*. Theorem 4 finds that the directional derivatives equal to each other (equivalently, V is differentiable) if and only if the *diffusion policy* that maps a current threshold degree to the threshold degree in the next period is single-valued.[18] The diffusion policy is single-valued if the per-period profit function is concave on a small region $[\underline{x}, \overline{x}]^2$, where $\underline{x}$ uniquely satisfies

$$\frac{\pi_2(\overline{x}, \underline{x})}{\widetilde{f}(\underline{x})} = \beta \underbrace{\overline{x}(1 - F(\overline{x}))}_{\substack{\text{the maximum} \\ \text{friendship effect}}}.$$

Lower bound $\underline{x}$ represents a minimum threshold degree that the monopolist may choose to maximize the friendship effect, even though only consumers with degree at least $\overline{x}$ are currently expected to subscribe to the good.[19]

This concavity assumption on a small region is not too restrictive. First, the uniqueness of steady state diffusion level is independent of this assumption. Second, for a given degree distribution, one can calculate the two bounds and check the concavity of the per-period profit function on $[\underline{x}, \overline{x}]^2$. For example, a direct calculation provides that when the underlying degree distribution is an exponential degree distribution, the corresponding per-period profit function is strictly concave on $[\underline{x}, \overline{x}]^2$. Last, if the degree distribution is decreasing, then a sufficient condition for the concavity can be written as $\frac{1}{x} < \frac{|f'(x)|}{f(x)}$ for $x \in [\underline{x}, \overline{x}]$.

2.3.2 Dynamics

I now analyze how an optimal diffusion plan and the corresponding pricing plan behave as time passes. For this, I first observe that when x_{t-1}^* and x_{t+1}^* are fixed, a change in x_t only influences the per-period profits in period t and $t+1$. Suppose that $x_{t-1}^* < x^*$, which means that the size of network effects in period t is larger than that at the steady state of the market where both the monopolist and consumers do not change their decisions. In this scenario, the only reason to set x_t smaller than x^* is to aim for a higher profit in period $t+1$ through the friendship effect, while sacrificing profit in period t. For this higher future profit, equation (2.3.1) requires x_{t+1}^* to be strictly greater than x^*. As a consequence, any optimal diffusion plan does not contain three consecutive threshold degrees strictly smaller than the steady state diffusion level. Similarly, it can be shown that any optimal diffusion plan does

[18] A correspondence $k : \mathbb{R}_0^+ \rightrightarrows \mathbb{R}_0^+$ is said to be a diffusion policy if $x_{t+1}^* \in k(x_t^*)$ for all t.

[19] $\underline{x}$ is the unique minimum threshold degree that the monopolist might choose because $\frac{\pi_2(x,y)}{\widetilde{f}(y)}$ is strictly decreasing in x, but $x(1 - F(x))$ is strictly increasing in x whenever $x \in (0, \overline{x})$.

not contain three consecutive threshold degrees strictly greater than the steady state diffusion level. Therefore, optimal diffusion plans not staying at the steady state alternate around the steady state diffusion level.

To obtain an alternating diffusion plan, the monopolist has to set a pricing plan that alternates around the steady state price level. For example, if $x^*_{t-1} > x^*$, it is required to set p^*_t strictly smaller than p^* to achieve $x^*_t < x^*$. The following proposition summarizes discussions.

Proposition 10 *Any optimal diffusion plan does not contain three consecutive threshold degrees either strictly smaller or strictly greater than the steady state diffusion level:*

$$x^*_t < x^* \text{ and } x^*_{t+1} < x^* \implies x^*_{t+2} > x^*,$$
$$x^*_t > x^* \text{ and } x^*_{t+1} > x^* \implies x^*_{t+2} < x^*.$$

Moreover, any optimal pricing plan does not contain four consecutive subscription prices either strictly smaller or strictly greater than the steady state price level:

$$p^*_t < p^*, p^*_{t+1} < p^*, \text{ and } p^*_{t+2} < p^* \implies p^*_{t+3} > p^*,$$
$$p^*_t > p^*, p^*_{t+1} > p^*, \text{ and } p^*_{t+2} > p^* \implies p^*_{t+3} < p^*.$$

Proposition 10 illustrates an important difference between subscription good markets and durable good markets. In a durable good market, a consumer permanently leaves the market after her purchase. As a result, if the monopolist lowers the purchasing price in a period, then he loses all the possible profits he can earn from consumers who buy the good in the period. In other words, the monopolist competes on consumer demands with his future selves. On the contrary, there is no such self competition in a subscription good market, which in turn implies that it is optimal for the monopolist to lower the subscription price under the steady state price level occasionally.

2.4 Properties of the Steady State

In this section, I first study efficiency and stability of the steady state diffusion level. Then, I find a closed-form expression for a per-period deadweight loss from monopoly at the steady state of the market.

2.4.1 Efficiency and Stability

Recall that $W(x, x) = x(1-\widetilde{F}(x))$ measures the willingness-to-pay of consumers with degree x when consumers with degree at least x are expected to subscribe to the good. By the single crossing property in Lemma 2, $W(x,x)$ is single-peaked at $x = \widetilde{x}$, where $\widetilde{x}$ solves equation $\frac{1}{x} = \frac{\widetilde{f}(x)}{1-\widetilde{F}(x)}$. Thus, with respect to the steady state price level, there are two distinctive degrees x^H and x^L satisfying $x^H < \widetilde{x} < x^L$ and $W(x^H, x^H) = W(x^L, x^L) = p^*$. Since the subscription rate at x^H is strictly higher than the subscription rate at x^L, there is a coordination problem between consumers because either expectation that consumers with degree at least x^H or x^L subscribe to the good is self-fulfilling.

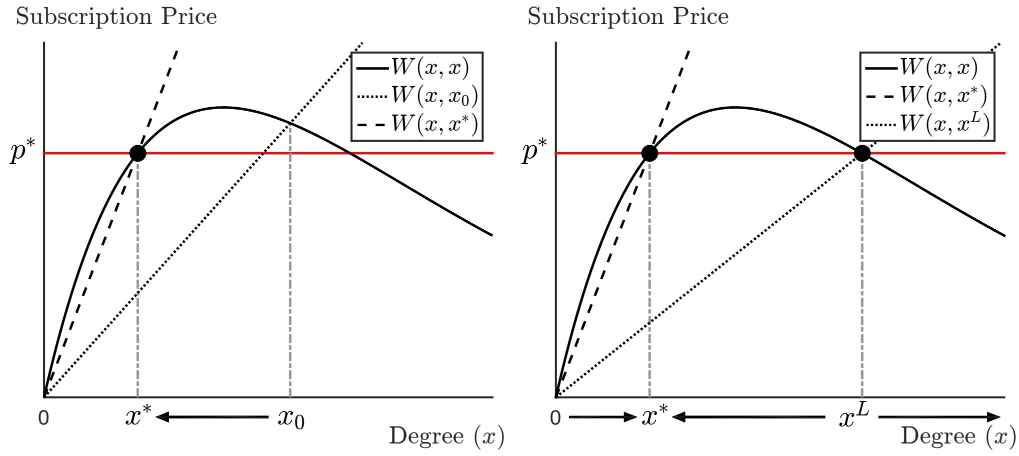
Note that the expectation that consumers with degree at least x^* subscribe to the good is also self-fulfilling with respect to subscription price p^*. For this, I call x^* is *diffusion-efficient* if $x^* = x^H$, because $x^* = x^L$ means that there is a coordination failure between consumers at the steady state of the market. The following proposition finds that the steady state diffusion level is diffusion-efficient.

Proposition 11 *The steady state diffusion level is diffusion-efficient.*

The proof directly follows from the dominance relationship in Lemma 2, which implies $x^* < \overline{x} < \widetilde{x}$. In fact, as in other models for a network good market, this diffusion-efficiency of x^* is equivalent to a dynamic stability: when the subscription price is fixed at p^*, the steady state diffusion level is robust to small perturbations to consumers' belief in the subscription rate $1 - \widetilde{F}(x^*)$. Perturbations can arise, for instance, due to incorrect reports of the subscription rate by the media, or because the monopolist tries to exaggerate some statistics about the subscription rate. I formally define this stability notion as the following:

Definition 4 *x^* is said to be **diffusion-stable** if there exists $\varepsilon > 0$ such that for all $x_0 \in (x^* - \varepsilon, x^* + \varepsilon)$, a diffusion process defined as $x_t := \frac{p^*}{1-\widetilde{F}(x_{t-1})}$ converges to x^* as $t \to \infty$.*

To see why x^* is called diffusion-stable, suppose that the subscription price is fixed as p^* and that consumers' belief is perturbed in a way that each neighboring consumer subscribes with probability $1 - \widetilde{F}(x_0)$ as illustrated in Figure 2.3(a). For this perturbed belief, the inverse demand shifts down to $W(x, x_0)$. In this shifted inverse demand curve, consumers with degree in $[x_1, x_0)$ where $x_1 = \frac{p^*}{1-\widetilde{F}(x_0)} < x_0$ are still willing to subscribe at p^*. Thus, in the next period, the inverse demand curve shifts up from $W(x, x_0)$ to $W(x, x_1)$. This positive feedback continues until the diffusion process reaches x^*, the steady state diffusion level. An analogous process



(a) x_t converges to x^*

(b) x^* is diffusion-stable iff $x^* < \widetilde{x}$

Figure 2.3: Illustration of the diffusion stability of x^*

with negative feedbacks takes place when consumers believe that more consumers are subscribing than the actual subscription rate.[20]

However, the above convergence does not arise if $x_0 > x^L$. For example, as illustrated in Figure 2.3(b), if consumers' belief is perturbed in a way that each neighboring consumer subscribes with probability less than $1 - \widetilde{F}(x^L)$, then the diffusion process diverges to infinity.

2.4.2 Deadweight Loss

Since the marginal production cost is zero, social surplus is maximized when every consumer subscribes to the good at the subscription price of zero. For this, I define the *steady state deadweight loss ratio* from monopoly as the ratio of the per-period deadweight loss at the steady state of the market to the maximum social surplus.

First note that the maximum per-period social surplus is the mean of the degree distribution. If every consumer subscribes to the good at zero subscription price, a consumer with degree x interacts with all of her neighbors, so that she obtains a utility of x. Thus, the total consumer surplus is $\int_0^\infty x f(x)\, dx$, the mean of f. Since the monopolist's per-period profit is zero in this case, the mean of the degree distribution is the maximum per-period social surplus. On the contrary, the per-

[20]When consumers believe that the subscription rate is higher than $1 - \widetilde{F}(x^*)$, some consumers immediately cancel their subscriptions because p^* is higher than their willingness-to-pay. This negative feedback continues until the diffusion process converges to x^*.

period social surplus at the steady state of the market is $\int_{x^*}^{\infty} x(1-\widetilde{F}(x^*))\,dx$ because (i) each neighboring consumer is expected to subscribe with probability $1-\widetilde{F}(x^*)$, and (ii) only consumers with degree at least x^* subscribe to the good.

Therefore, the steady state deadweight loss ratio from monopoly, denoted by DW, is

$$DW := \frac{\int_0^{\infty} xf(x)\,dx - \int_{x^*}^{\infty} x(1-\widetilde{F}(x^*))\,dx}{\int_0^{\infty} xf(x)\,dx}.$$

The following proposition finds an expression for DW that solely depends on the conditional degree distribution.

Proposition 12 *The steady state deadweight loss ratio is*

$$DW = \widetilde{F}(x^*)(1-\widetilde{F}(x^*)) + \widetilde{F}(x^*). \tag{2.4.1}$$

Figure 2.4 illustrates the decomposition of two terms for given steady state diffusion and price levels. The lower dark triangle depicts the first term in expression (2.4.1) before normalization, corresponding to the classical deadweight loss generated by restricting some consumers from taking the positive network effect. If consumers with degree in $[0, x^*)$ were to subscribe to the good, they would derive a positive network effect of $1-\widetilde{F}(x^*)$ from each neighboring consumer's subscription. The amount of this welfare loss is calculated by multiplying the positive network effect and the average degree of the consumers excluded from the market.

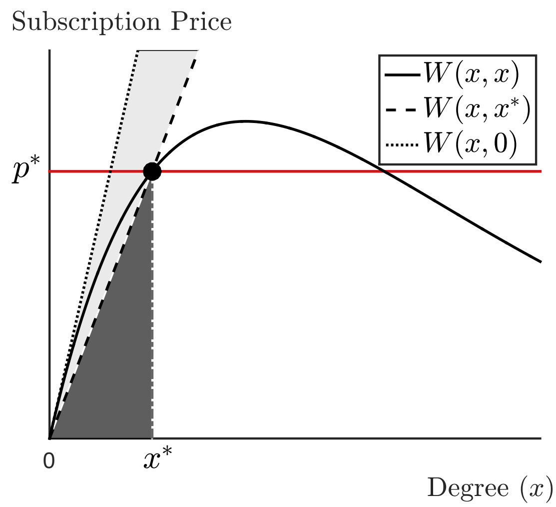


Figure 2.4: Illustration of two sources of welfare loss

The upper light triangle represents the other source of welfare loss. This inefficiency originates from the lost positive network effect at the steady state of the market, and it is measured by the second term in expression (2.4.1) after the normalization. If all consumers subscribe to the good, then the inverse demand shifts up from $W(\cdot, x^*)$ to $W(\cdot, 0)$ because every consumer can derive an additional positive network effect of $\widetilde{F}(x^*)$ from each neighboring consumer's subscription. The amount of this deadweight loss is calculated by multiplying the lost positive network effect to the average degree of the whole population.

2.5 Comparative Statics

In this section, I investigate how the steady state diffusion level, price level, and deadweight loss ratio from monopoly depend on two fundamentals of the monopoly market: the monopolist's discount factor and the social network density.

2.5.1 Changes in Discount Factor

As functions of β, I denote (i) by $x^*(\beta)$ the steady state diffusion level, (ii) by $p^*(\beta)$ the steady state price level, and (iii) by $DW(\beta)$ the steady state deadweight loss ratio. I find that these quantities are decreasing in β as follows.

Proposition 13 *The steady state diffusion level, price level, and deadweight loss ratio are decreasing in the monopolist's discount factor.*

As the discount factor increases, the monopolist has an incentive to aim for a higher friendship effect. This automatically implies that $x^*(\beta)$ is decreasing in β. Observe also that since the steady state diffusion level is diffusion-efficient, $(x^*(\beta), p^*(\beta))$ is located on the graph of W where $W(x, x)$ is increasing in x. This means that $p^*(\beta)$ is decreasing in β. In turn, that $x^*(\beta)$ and $p^*(\beta)$ are decreasing in β imply that $DW(\beta)$ is also decreasing in β.

This comparative static result has a policy implication. One possible interpretation of the monopolist's discount factor β is that it represents how confident he is that he will continue to hold exclusive rights in the market in future periods. Consider a developing country whose government tries to introduce a new monopoly market for a social network good, such as landline phone service in an area. The above result indicates that giving a strong monopoly power may be optimal for the government if its goal is to minimize the steady state deadweight loss ratio.

2.5.2 Changes in Network Structure

I consider an increase in density of social networks in terms of the *monotone likelihood ratio property* (MLRP). For this, I represent two social networks by their degree distributions $f(\cdot, \xi_0)$ and $f(\cdot, \xi_1)$ satisfying all the assumptions for the existence of the steady state diffusion level discussed in Section 2.3. I assume the MLRP of $f(\cdot, \xi_0)$ and $f(\cdot, \xi_1)$ as follows.

Assumption 3 *For all $x_1 > x_0$,*

$$\frac{f(x_1;\xi_1)}{f(x_1;\xi_0)} \geq \frac{f(x_0;\xi_1)}{f(x_0;\xi_0)}.$$

Intuitively, the assumption means that the social network represented by $f(\cdot;\xi_1)$ is *denser* than that represented by $f(\cdot;\xi_0)$. I denote by $\widetilde{f}(\cdot,\xi_0)$ and $\widetilde{f}(\cdot,\xi_1)$ the corresponding conditional degree distributions. Assumption 3 implies the following statistical orderings for the degree distributions and the conditional degree distributions.

(i) The MLRP of $\widetilde{f}(\cdot,\xi_0)$ and $\widetilde{f}(\cdot,\xi_1)$:[21] for all $x_1 \geq x_0$,

$$\frac{\widetilde{f}(x_1;\xi_1)}{\widetilde{f}(x_1;\xi_0)} \geq \frac{\widetilde{f}(x_0;\xi_1)}{\widetilde{f}(x_0;\xi_0)}. \tag{2.5.1}$$

(ii) The first-order stochastic dominance property: for all $x \in \mathbb{R}_0^+$,

$$1 - F(x;\xi_1) \geq 1 - F(x;\xi_0) \quad \text{and} \quad 1 - \widetilde{F}(x;\xi_1) \geq 1 - \widetilde{F}(x;\xi_0). \tag{2.5.2}$$

(iii) The monotone hazard rate property: for all $x \in \mathbb{R}_0^+$,

$$\frac{f(x;\xi_1)}{1-F(x;\xi_1)} \leq \frac{f(x;\xi_0)}{1-F(x;\xi_0)} \quad \text{and} \quad \frac{\widetilde{f}(x;\xi_1)}{1-\widetilde{F}(x;\xi_1)} \leq \frac{\widetilde{f}(x;\xi_0)}{1-\widetilde{F}(x;\xi_0)}. \tag{2.5.3}$$

I finally denote by $x^*(\xi_i)$ and $p^*(\xi_i)$ the steady state diffusion and price levels under $f(\cdot;\xi_i)$. The following proposition finds that $x^*(\xi_i)$ and $p^*(\xi_i)$ are increasing in the density of the social network.

Proposition 14 *The steady state diffusion and price levels increase as the density of the social network increases in the likelihood ratio: $x^*(\xi_1) > x^*(\xi_0)$ and $p^*(\xi_1) > p^*(\xi_0)$.*

[21] Let $\mu(\xi_i) = \int_0^\infty x f(x;\xi_i)dx$ for $i = 0, 1$. Then, for all $x_1 \geq x_0$,

$$\frac{\widetilde{f}(x_1;\xi_1)}{\widetilde{f}(x_1;\xi_0)} = \frac{x_1 f(x_1;\xi_1)/\mu(\xi_1)}{x_1 f(x_1;\xi_0)/\mu(\xi_0)} \geq \frac{x_0 f(x_0;\xi_1)/\mu(\xi_1)}{x_0 f(x_0;\xi_0)/\mu(\xi_0)} = \frac{\widetilde{f}(x_0;\xi_1)}{\widetilde{f}(x_0;\xi_0)}.$$

Because of the monotone hazard rate property, other things equal, the monopolist has an incentive to choose a larger x_t under $f(\cdot;\xi_1)$ than under $f(\cdot;\xi_0)$ in order to maximize the profit in period t. In addition, the friendship effect is higher under $f(\cdot;\xi_1)$ than under $f(\cdot;\xi_0)$ by the first-order stochastic dominance property. Hence, the monopolist's former incentive to choose a higher x_t does not conflict with his latter incentive to choose a lower x_t to aim for a higher friendship effect. Therefore, the steady state diffusion level increases as the density of the social network increases.

To see why the steady state price level also increases, note that $(x^*(\xi_i), p^*(\xi_i))$ is located on the upward slope of $W(x, x; \xi_i) := x(1 - \widetilde{F}(x; \xi_i))$ as illustrated in Figure 2.5. This indicates that in the denser social network represented by $f(\cdot;\xi_1)$, consumers believe that their neighbors are more likely to have a higher degree. Knowing this, the monopolist sets a higher subscription price for a higher profit at the steady state of the market. Thus, the steady state price level increases in the density of the social network as well.

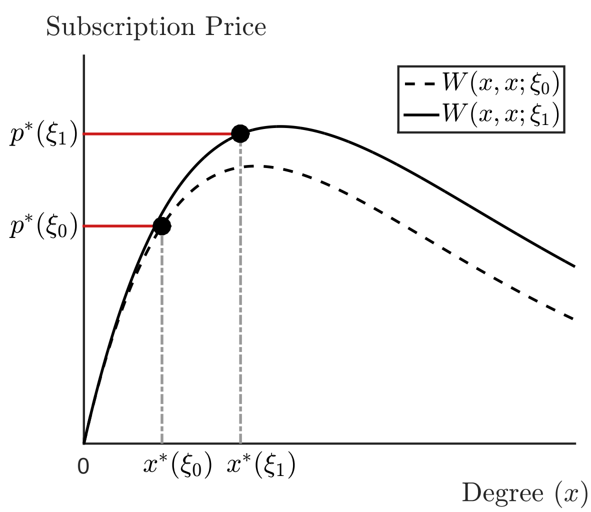


Figure 2.5: Monotonicity of the steady state subscription and price levels

By combining the above two monotonicity results, the following proposition finds that the monopolist can earn a higher per-period profit at the steady state of the market as the social network becomes denser.

Corollary 2 *The per-period profit at the steady state of the market increases as the density of the social network increases in likelihood ratio.*

This comparative static result is different from Jackson and Yariv (2007), in which monotonicity results regarding the stable steady state follows from the first-order stochastic dominance property of the degree distributions. On the contrary, due to the presence of a monopolist who optimally chooses a pricing plan, the first-order stochastic dominance property does not ensure a monotonicity of the steady state diffusion level in the current model.

In fact, considering the MLRP has three benefits. First, the MLRP only concerns the degree distributions, and it implies the same property for the conditional degree distributions.[22] Second, it provides monotonicity of the steady state price level as well as of the steady state diffusion level. Third, having these two monotonicity results leads to monotonicity of the steady state monopoly profit.

2.6 Concluding Remarks

This paper presents a novel approach for studying how a monopolist takes into account the diffusion of subscription decisions in a social network for its optimal dynamic pricing strategy. Depending on connections within the social network, consumers derive heterogeneous utilities from subscribing to a network good. Due to the presence of the monopolist who optimally balances his intertemporal tradeoffs, I obtain optimal pricing plans exhibit oscillating patterns around the steady state price level. I also provide monotone comparative static results concerning the monopolist's discount factor and the social network density in terms of monotone likelihood ratio order.

Though the current paper has focused on a monopoly market, it might also be worth analyzing a competitive market in which multiple firms simultaneously optimize their pricing policies through time. Another salient extension would be to consider consumers who have richer information about their neighboring consumers' connectivity and decisions. One drawback of the mean-field approximation considered in this paper is that each neighboring consumer's degree (identity) is assumed to be independently and identically redrawn from the degree distribution across all the periods. As a result, consumers ignore correlations between neighbors' decisions and identities. In reality, however, consumers continuously interact with one another, so that they may better understand neighbors' decisions based on their information on neighbors' identities in a social network. Analyzing how subscription decisions diffuse within a social network and how a monopolist's optimal pricing plan changes in such a richer environment will be an interesting future research direction.

[22] On the contrary, the first-order stochastic dominance property of two degree distributions does not imply the same property for the corresponding conditional degree distributions (Shin, 2016b).

Chapter 3

A MODEL OF PRE-ELECTORAL COALITION FORMATION

3.1 Introduction

3.1.1 Overview

In many elections with three or more candidates, coalitions can form between candidates in order to defeat a common opponent. These *pre-electoral coalitions* (*PECs*) often require that one candidate withdraw from the race and throw his support to another candidate in exchange for policy concessions from that candidate. These types of coalitions can frequently occur in single-office elections, such as presidential or gubernatorial elections. Bormann and Golder (2013) find that between 2001 and 2011, the average number of effective candidates in 147 worldwide presidential elections was 3.1.[1] This means that the possibility of PEC formation is quite high in presidential elections. Indeed, PECs have formed in presidential elections in many countries (e.g., France, Mexico, and South Korea). In France, a PEC carried François Hollande to victory in 2012. In Mexico, winners of two of the recent three presidential elections were also beneficiaries of PECs.[2] In South Korea, every presidential election since 1992 has included at least one PEC, and both Dae Jung Kim and Moo-Hyun Roh won elections in 1997 and 2002 in part due to PECs.[3]

Previous literature on PECs has focused primarily on parliamentary democracies. For instance, Golder (2006b) observes that in 23 parliamentary democracies between 1946 and 2002, 47.8% of the 364 legislative elections had at least one PEC.[4] Moreover, she finds that 19.2% of all the elections in her data produced a government that was based on a PEC agreement. Given their prevalence and impact, however, the literature on PECs is surprisingly thin (Powell, 2000). Since PECs can only exist in multi-candidate elections, the limitation might be a result of the literature's

---

[1] The median is 2.86. The number of effective candidates in an election is calculated by $1/\sum_i q_i^2$, where q_i is candidate i's proportion of all votes in the election.

[2] Following the definition proposed by Bormann and Golder (2013), I include only democratic presidential elections from 2000. Vicente Fox won the 2000 presidential election as the candidate of Alliance for Change, which was a coalition between the National Action Party and the Ecological Green Party of Mexico. In 2012, Enrique Peña Nieto won the election as the candidate of the Institutional Revolutionary Party and the Ecological Green Party of Mexico.

[3] I focus on presidential elections taking place since 1992 because the 1992 presidential election was the first free and fair election in which South Korea elected its first civilian president (Golder, 2006b).

[4] One can find more PEC examples and related empirical results in parliamentary democracies in Bandyopadhyay et al. (2011) and references therein.

focus on American politics, where two dominant parties competing is norm. The assumption of two party elections is also supported by Duverger's law, which asserts that elections under plurality rule tend to favor a two-party system (Duverger, 1954).

From a theoretical perspective, PECs in single-office elections involve a different set of incentives than other elections. In single-office elections, only one member of a PEC can occupy office; this is unlike in parliamentary systems, in which a PEC can lead to a sharing of seats. Hence, if candidates obtain utility from holding office, utilities are not fully transferable between coalition partners. For this reason, PEC formation incentives in presidential elections call for a model different from these used for parliamentary elections (e.g., Bandyopadhyay et al., 2011; Golder, 2006b). This paper investigates and compares PEC incentives in single-office elections, as a function of electoral environments such as election rules, ideological distance between candidates, and pre-election polls. To this end, I analyze a sequential game in which three candidates compete in large elections for one office.

In my model, there are three candidates, $L1$, $L2$, and R, who are motivated by both ideology and office value.[5] I assume that due to their ideological similarity, only $L1$ and $L2$ can form a PEC by nominating one representative candidate and choosing a joint policy platform. There are n voters, and each voter is one of three types: (i) a type t_{L1} voter who prefers $L1$ to $L2$, and $L2$ to R, (ii) a type t_{L2} voter who prefers $L2$ to $L1$, and $L1$ to R, and (iii) a type t_R voter who prefers R to $L1$ and $L2$, but evaluates $L1$ and $L2$ equally. Types of voters are independently and identically distributed according to a probability distribution, which is unknown to both voters and candidates. Voters and candidates observe a set of public signals drawn from the same distribution of voter types, which models a pre-election opinion poll result. After observing the poll result, candidates $L1$ and $L2$ publicly decide whether to form a PEC. Voters cast their ballots based on their private information and the PEC outcome. Finally, one candidate is elected according to the given election rule. I consider two election rules, *plurality* and the *two-round runoff*, which are the two most frequently used rules for presidential elections; for instance, Bormann and Golder (2013) find that about 87% of presidential elections between 2001 and 2011 utilized these rules.[6]

[5]There are three lines of theoretical models in terms of assumptions of political candidates' motivations. Following Hotelling (1929), one stream assumes that candidates are purely office-motivated. Another stream assumes that candidates are purely policy-motivated. Finally, a third stream makes assumptions between the previous two extremes, and the current paper belongs to this stream. See Callander (2008) and references therein for more detailed discussions.

[6]In their original dataset that covers worldwide presidential elections between 1946 to 2011, 76% of elections used these two rules. Among those elections, about 40% of elections were under

This model captures the main incentives that each candidate faces. When a candidate chooses to form a PEC in which he is not the representative candidate, he must evaluate the tradeoff between reducing his likelihood of holding office (to zero) and the increase in likelihood of a policy more to his liking. The latter is driven by the fact that the PEC increases his partner's chances of winning the election and implementing the joint policy platform of the PEC. In my model, if candidates are purely policy-motivated agents, then $L1$ and $L2$ always form a PEC; by doing so, they can strictly increase their probability of winning and split additional benefits by choosing a compromised joint policy platform. If candidates are purely office-motivated, however, no candidate has an incentive to form a PEC. Thus, *ceteris paribus*, there is a threshold office value such that $L1$ and $L2$ form a PEC whenever the office value is lower than that threshold. The likelihood of forming a PEC under a particular electoral environment is measured through this threshold office value.

This paper produces three main results: (i) PECs are more likely to form in plurality elections than in two-round runoff elections; (ii) in two-round runoff elections, PECs are more likely to form as the threshold for first-round victory decreases; and (iii) conditional on divided support, PECs are more likely to form as the ideological distance between PEC partners increases.[7]

3.1.2 Related Literature

There is a small but growing body of literature on PEC formation. Golder (2006b) provides a comprehensive survey of PECs around the world. In addition, she builds a simple bargaining model to derive a set of testable comparative static results. Carroll and Cox (2007) and Debus (2009) find empirical evidence that PECs significantly affect election outcomes and government formations. Bandyopadhyay et al. (2011) compare coalition outcomes in proportional representation and plurality voting systems under different bargaining protocols. They assume a continuum of voters, which in turn implies that voters are not strategic because no single voter can change election outcomes. Thus, election outcomes are purely a function of parties' coalition decisions. In contrast, I allow strategic voting behavior so that the election outcome depends on voters' equilibrium voting behavior as well as candidates' PEC decisions.

In terms of motivation, the current paper studies PEC formation incentives in

plurality rule.

[7]Prediction (i) implies that under the presence of PECs, plurality elections are more likely to include fewer candidates than two-round runoff elections. This is consistent with several empirical findings (e.g., Bormann and Golder, 2013; Fujiwara, 2011; Golder, 2006a).

presidential elections, which are not a primary concern of the above papers. In particular, in my model, office value is *non-transferable*, and candidates are not allowed to form a coalition after the election. This feature dramatically changes the PEC formation incentives. In terms of results, the current paper predicts that in presidential elections, PECs are more likely to form under plurality rule. However, Nagashima (2011) finds the opposite result in legislative elections in which party leaders might be able to share office value by allocating seats to other coalition partners.

My model is related to the literature on multi-candidate elections with strategic voters (e.g., Bouton, 2013; Cox, 1997; Hummel, 2014; Myatt, 2007). A substantial portion of this literature is aimed at rationalizing the Duvergerian equilibria where a Condorcet loser would be elected.[8] Cox (1997) provides a good summary of strategic voting behavior from both theoretical and empirical perspectives. Myatt (2007) employs techniques of global games (Morris and Shin, 2003). He considers the voters who use their private signals to infer the true relative popularity of their favorable candidates. He proves that in a plurality election, there exists a unique equilibrium with partial coordination. As a result, a Condorcet loser could be elected. Bouton (2013) offers a tractable framework to study strategic voting in two-round runoff elections by considering a Poisson voting game (Myerson, 1998; Myerson, 2000). He shows that two-round runoff elections with a threshold below 50% may elect the Condorcet loser. In the current paper, by adopting the model by Hummel (2014), I show that in both plurality and two-round runoff elections, the possibility of electing the Condorcet loser remains even with PEC formation.[9]

Several papers compare plurality and two-round runoff rules in terms of strategic voting behavior in multi-candidate elections (e.g., Bordignon et al., 2013; Niou, 2001). Hummel (2014) studies sequential voting in large elections, and he examines voters' incentives in responding to pre-election polls. I extend Hummel (2014) to account for important electoral details (e.g., election rules, ideological distance, and pre-election polls) that may affect strategic voting and PEC formation incentives. Furthermore, I also analyze different strategic voting behavior in plurality elections

[8]A candidate is called the Condorcet loser if a majority would vote against him in a two-candidate race between himself and any one of the other candidates (Myerson and Weber, 1993).

[9]Both Myatt (2007) and Hummel (2014) assume that there exists *aggregate uncertainty* about the true distribution from which voters' private signals are drawn. This assumption provides that the odds of pivotal events do not diverge as the number of voters becomes infinitely large, so that voters may vote strategically. Although this assumption is unusual in formal models considering strategic voters (Hummel, 2014), there is a growing literature adopting it (e.g., Chamberlain and Rothschild, 1981; Dewan and Myatt, 2007; Ekmekci, 2009; Good and Mayer, 1975; Hummel, 2012; Hummel, 2014; Myatt, 2007).

and two-round runoff elections.

The current paper is also broadly related to other literature on coalition formation and bargaining (e.g., Chatterjee et al., 1993; Bandyopadhyay et al., 2011; Eraslan and Merlo, 2002; Okada, 2011; Ray, 2008). Ray (2008) provides an excellent survey of this literature. These models are interested in how coalition formation and election outcomes depend on bargaining protocols, and are related to cooperative bargaining solutions such as the Nash bargaining solution. Although my model takes the Nash bargaining solution as a bargaining protocol, its cooperative feature is not a driving force of the results. For example, as I will briefly discuss in Section 3.3, a Rubinstein alternating-offers model with outside options, which is a non-cooperative bargaining protocol, generates the same bargaining outcome. The main focus of the current paper is the comparison of PEC formation incentives under different election rules, rather than the identification of outcomes of the PEC bargaining process.

The rest of the paper is organized as follows. Section 3.2 describes a sequential game of four periods. Section 3.3 analyzes the outcomes of plurality rule. Section 3.4 studies the outcomes of the two-round runoff rule and compares PEC formation incentives under these two election rules. Section 3.5 analyzes voter welfare. Section 3.6 presents comparative static results. Finally, Section 3.7 concludes. All proofs are gathered in Appendix C.

3.2 The Model

Candidates and voters. $C = \{L1, L2, R\}$ is a set of three candidates competing for one office. If a candidate wins the election alone, then he obtains a utility of $V + 1$ where V is the value of holding office, and 1 is the value of choosing his ideal policy. $L1$ and $L2$ are left candidates who are similar in their ideologies; if $L2$ wins, $L1$ obtains a utility of $1/2$ and vice versa.[10] On the other hand, R obtains a utility of 0 if either $L1$ or $L2$ wins. Similarly, if R wins, he obtains a utility of $V + 1$, but both $L1$ and $L2$ obtain a utility of 0.

$N = \{1, \ldots, n\}$ is a finite set of voters. $T = \{t_{L1}, t_{L2}, t_R\}$ is a set of voters' types. A voter is said to be of type t_j if her most preferred candidate is j. Voters are only interested in candidates' ideologies. Thus, a type t_{L1} voter obtains a utility of 1 if $L1$ wins, $1/2$ if $L2$ wins, and 0 if R wins; a type t_{L2} voter obtains a utility of 1 if $L2$ wins, $1/2$ if $L1$ wins, and 0 if R wins. Finally, a type t_R voter obtains a utility of $3/2$ if R wins, but 0 otherwise.[11]

[10] Section 3.6.2 presents a case where ideological distance between $L1$ and $L2$ is parameterized by $\theta \in [0, 1]$.

[11] I assume a utility of $3/2$ for a type t_R voter's utility to normalize the sum of welfare levels

The sequential game. I consider a sequential game of four periods. At the outset of the game, types of voters are determined as independently and identically distributed (i.i.d.) draws from a probability distribution p, where p takes on value t_{L1} with probability p_{L1}, t_{L2} with probability p_{L2}, and t_R with probability $p_R = 1 - p_{L1} - p_{L2}$; that is, a voter's type is t_j with probability p_j. I assume that the exact value of p is unknown, but it is *common knowledge* that p is distributed according to distribution f with full support over $\triangle^2 = \{(p_{L1}, p_{L2}) \in \mathbb{R}^2 | 0 \leq p_{L1}, p_{L2} \leq 1 \text{ and } 0 \leq p_{L1} + p_{L2} \leq 1\}$.[12] f is assumed to be symmetric and continuously differentiable.[13] Voters know their types but do not know any of the other voters' types.

In period 2, voters and candidates observe the result of a single pre-election opinion poll. I model the opinion poll as a number of public signals, each of which is an i.i.d. draw from p. Without loss of generality, $\sigma = (\sigma_{L1}/m, \sigma_{L2}/m)$ represents a result of the opinion poll, where σ_j denotes the number of j signals out of m.[14] $\Sigma = \{(\sigma_{L1}/m, \sigma_{L2}/m) \in \triangle^2 | 0 \leq \sigma_{L1}, \sigma_{L2} \leq m \text{ and } 0 \leq \sigma_{L1} + \sigma_{L2} \leq m\}$ is the set of all possible results of the opinion poll.

In period 3, bargaining between $L1$ and $L2$ takes place over whether and how to form a PEC. In the bargaining process, they decide (i) who becomes the nominated candidate (i.e., who withdraws from the race) and (ii) how to choose their joint policy platform. I denote a PEC by $\xi(k, \lambda)$, where $k \in \{L1, L2\}$ represents the nominated candidate, and $\lambda \in [0, 1]$ measures how much the coalition reflects $L1$'s ideal policy. I assume that $\xi(k, \lambda)$ is determined as a Nash bargaining solution (Nash, 1950).[15] Table 3.1 summarizes the utility structure of the game. Note that for a given $\xi(k, \lambda)$, only the nominated candidate k takes the office value V.

In period 4, voters cast a ballot for a single candidate. An election rule is denoted by $\mathcal{E}$, which is either plurality ($\mathcal{P}$) or the two-round runoff ($\mathcal{R}$). Under plurality rule, the candidate who obtains the relative majority in a single round wins. I assume that if there are some candidates that tie, then each of those candidates is elected with equal probability.

Under the two-round runoff rule, the candidate who is ranked first by taking

corresponding to all three types a constant. This normalization simplifies the welfare analysis in Section 3.5.

[12] There are two degrees of freedom for the support of f because p is a probability distribution such that $p_{L1} + p_{L2} + p_R = 1$.

[13] Symmetry of f is defined as $f(x, y) = f(y, x)$ for all $(x, y) \in \triangle^2$. All the results except the welfare calculations in Section 3.5 do not depend on this symmetricity assumption.

[14] Since $\sigma_{L1} + \sigma_{L2} + \sigma_R = m$, there are two degrees of freedom for σ.

[15] With symmetricity of the distribution function f, the symmetry property of the Nash bargaining solution simplifies the calculation of ex-ante payoff for the voters.

Winners	Voter Types			Candidates		
	t_{L1}	t_{L2}	t_R	$L1$	$L2$	R
$L1$	1	$1/2$	0	$V+1$	$1/2$	0
$L2$	$1/2$	1	0	$1/2$	$V+1$	0
R	0	0	$3/2$	0	0	$V+1$
$\xi(L1,\lambda)$	$(1+\lambda)/2$	$(2-\lambda)/2$	0	$V+(1+\lambda)/2$	$(2-\lambda)/2$	0
$\xi(L2,\lambda)$	$(1+\lambda)/2$	$(2-\lambda)/2$	0	$(1+\lambda)/2$	$V+(2-\lambda)/2$	0

Table 3.1: The utility structure

more than or equal to 50% of the vote share in the first round wins outright, and there is no second round.[16] If everyone receives less than 50% of the vote share in the first round, then the top two candidates who received the most votes in the first round advance to the second round, and the candidate who obtains the majority of the vote share in that round ultimately wins. I also assume that if there are some candidates that tie in any round, then each of those candidates is chosen to advance or win with equal probability.

I finally assume that each election rule nullifies any ballot for a withdrawn candidate whenever a PEC forms. That is, for instance, if $L1$ and $L2$ decide to form a PEC by nominating $L1$, but some voters vote for $L2$, then an election rule excludes their ballots for $L2$ in calculation of the vote share. Table 3.2 summarizes the timeline of the game.

Period 1	Types of voters are determined
Period 2	Poll result is realized
Period 3	$L1$ and $L2$ bargain over PEC formation
Period 4	Voters vote, and the election rule decides the winner

Table 3.2: The timeline of the game

Information. The election winner is determined by the exact value of p as the number of voters n increases to infinity. For instance, suppose that p satisfies $p_{L1} > p_{L2} > p_R$ and that voters select their most preferred candidates. Then, by the law of large numbers, the probability that $L1$ wins the election converges to one as n increases to infinity. Since voters and candidates infer the value of p from the poll result σ, their updated beliefs on f affect their decisions.

[16] Section 3.6.1 studies parameterized two-round runoff rules where a candidate wins outright if he obtains more than $\zeta \leq 1/2$ vote shares in the first round. This limited attention of thresholds is empirically relevant: to the best of my knowledge, the 1996 presidential election in Sierra Leone was the only presidential election in which the threshold was set higher than 55%.

There are two stages in which voters update their beliefs. The first is when voters learn their types (private information) in period 1, and the second is when they observe the poll result (public information) in period 2. I denote by $f_j(\cdot|\sigma)$ the type t_j voters' beliefs after the second update, as a function of the poll result σ. For example, a type t_{L1} voter obtains $1+m$ signals from realization of her own type and m public signals until period 4. Consequently, her updated belief in period 4 is

$$f_{L1}(x,y|\sigma) = \frac{x^{\sigma_{L1}+1} y^{\sigma_{L2}} (1-x-y)^{m-(\sigma_{L1}+\sigma_{L2})} f(x,y)}{\int_0^1 \int_0^{1-w} z^{\sigma_{L1}+1} w^{\sigma_{L2}} (1-z-w)^{m-(\sigma_{L1}+\sigma_{L2})} f(z,w)\, dz\, dw}. \quad (3.2.1)$$

For candidates, there is only one stage in which beliefs are updated after observing σ in period 2. I denote the candidates' updated belief by $f_C(\cdot|\sigma)$ defined as

$$f_C(x,y|\sigma) = \frac{x^{\sigma_{L1}} y^{\sigma_{L2}} (1-x-y)^{m-(\sigma_{L1}+\sigma_{L2})} f(x,y)}{\int_0^1 \int_0^{1-w} z^{\sigma_{L1}} w^{\sigma_{L2}} (1-z-w)^{m-(\sigma_{L1}+\sigma_{L2})} f(z,w)\, dz\, dw}. \quad (3.2.2)$$

Note that since realizations of types and results of the opinion poll are independent of election rules, $f_j(\cdot|\sigma)$ and $f_C(\cdot|\sigma)$ do not depend on the election rule $\mathcal{E}$.

Voting strategies. I focus on symmetric pure strategies such that voters of the same type behave in the same way. I make another weak assumption about voting strategies that voters only make voting decisions that are *weakly undominated*. This assumption requires that no voter votes for candidate j who is dominated in the sense that there exists another candidate j', such that voting for j' never yields a lower expected payoff for the voter than voting for j does, and that there exists at least one situation that voting for j' returns a strictly higher expected payoff for the voter than voting for j does. As will be explained soon, this assumption characterizes a unique equilibrium voting behavior as a function of the poll result σ under each election rule.

Since the set of voting strategies depends on election rules, I define voting strategies for each election rule separately. In plurality elections, a voting strategy s is a function $s : T \times \Sigma \to C$; that is, a voting strategy depends on types and results of the opinion poll. Define S as the set of all voting strategies.[17] I denote by $BR : T \times \Sigma \times S \to C$ the best response function. I impose a tie-breaking rule that if a voter of either type t_{L1} or type t_{L2} is indifferent to candidates $L1$ and $L2$ in terms of her expected utility, then she votes for her most preferred candidate, without loss of further generality.

[17]With this notation, the following strategies are weakly dominated for all $\sigma \in \Sigma$: $s(t_R, \sigma) = L1$, $s(t_R, \sigma) = L2$, $s(t_{L1}, \sigma) = R$, and $s(t_{L2}, \sigma) = R$. In addition, when a PEC $\xi(k, \lambda)$ forms, $s(t_{L1}, \sigma) = k'$ and $s(t_{L2}, \sigma) = k'$ are weakly dominated strategies for $k' \neq k$.

In two-round runoff elections, the voting strategy is a bit complicated because it depends not only on what voters do in the second round, but also on how voters form their beliefs about the other voters' decisions in the second round (if it exists). To avoid any such complexity, I make the following assumptions about the voting decisions in the second round. If one of the left candidates and candidate R compete, then it is clear that type t_{L1} and type t_{L2} voters vote for the advanced left candidate, and that type t_R voters vote for R. When $L1$ and $L2$ reach the second round, it is natural to assume that type t_{L1} voters vote for $L1$ and type t_{L2} voters vote for $L1$. For type t_R voters' decisions, I assume that when there are r number of type t_R voters who are indifferent to $L1$ and $L2$, $\lfloor r/2 \rfloor$ number of voters vote for $L1$ and the other $r - \lfloor r/2 \rfloor$ number of type t_R voters vote for $L2$.[18]

With the simplifying assumptions described above, it suffices to consider type $L1$ and type $L2$ voters' voting strategies in the first round. Likewise for plurality elections, a voting strategy is a function $s : T \times \Sigma \to C$. I denote by S the set of voting strategies, and $BR : T \times \Sigma \times S \to C$ the best response function. I assume the same tie-breaking rule for type t_{L1} and type $L2$ voters.

Equilibrium notion. I consider Bayesian equilibria (*voting equilibria*) that satisfy a certain criterion. Recall that I restrict my attention to symmetric strategies, in which voters of the same type vote in the same way. Voters know the possible types of other voters and observe the same poll results. Given the uncertainty over other voters' types, however, each voter can only form beliefs about how the other voters will vote. Thus, each voter optimizes her voting decision by taking into account this uncertainty, depending on her own type and the poll results. Therefore, a voting equilibrium is denoted by a voting strategy that voters play in it.

Before defining the criterion, I first define an *uninformative* voting strategy, in which every voter votes for their most preferred candidate independent of their private information. Formally, it is defined as a voting strategy $\overline{s} \in S$ such that $\overline{s}(t_j, \sigma) = j$ for all $t_j \in T$ and all $\sigma \in \Sigma$. I focus on a voting equilibrium in which each voter's decision not only best responds to the equilibrium voting strategy, but also to the uninformative voting strategy. I call this voting equilibrium *strong* and define it as follows.

[18] For any weakly undominated strategy for type t_R voters, they always vote for R in the first round regardless of the realization of the poll result σ. Similarly, for any weakly undominated strategy for type t_{L1} and t_{L2} voters, they always vote for one of the left candidates. Thus, by observing an election result of the first round, voters can figure out the exact number of type t_R voters. Hence, type t_R voters are able to make decisions according to the described behavior. Alternatively, one can simply assume that type t_R voters vote for $L1$ with probability $1/2$.

Definition 5 *A voting equilibrium s^* is called* ***strong*** *if for all $t_j \in T$ and all $\sigma \in \Sigma$,*

(i) *s^* is weakly undominated;*

(ii) *$s^* = BR(t_j, \sigma, s^*)$;*

(iii) *$s^* = BR(t_j, \sigma, \overline{s})$.*

The concept of strong voting equilibrium captures the idea of conservative voting behavior. In reality, people might have several reasons to vote for their most preferred candidates, even though voting for that candidate does not add any value in terms of election outcome. Perhaps, voters may vote for their most preferred candidates to express their identities rather than obtaining better election outcomes (Fiorina, 1976).[19] In addition, some voters might not fully follow the opinion poll results. These voters might be poorly informed about the population preferences and willing to vote for their most preferred candidate based on their limited information. In this regard, the strong voting equilibrium describes an equilibrium voting behavior in which voters best respond to all those behaviors.

As will be shown in the following section, the strongness criterion characterizes a unique equilibrium voting behavior as a function of the poll result when the population size n becomes infinitely large. For this reason, I analyze the limiting properties of strong voting equilibria for sufficiently large n. From a theoretical perspective, the uniqueness enriches the scope of my analysis. In particular, I can calculate the expected utility of the voters according to their types for each poll result, which in turn enables me to find ex-ante utility of voters under each election rule. Therefore, I compare the ex-ante utilities of voters of each type, and investigate which election rule is more beneficial to which type of voters under the presence of a PEC.

3.3 Plurality Elections

In this section, I analyze the game when the election is governed by plurality rule. In the spirit of backward induction, I first characterize a unique strong voting equilibrium in period 4. Then, I formulate and analyze the PEC formation as a Nash bargaining problem.

[19] This motivation is called *expressive motivation* in the voting literature. See Schuessler (2000) for more detailed discussions.

3.3.1 Strong Voting Equilibrium

To begin, I provide a lemma that states how type t_{L1} and type t_{L2} voters best respond to the uninformative voting strategy.[20]

Lemma 3 If $\mathcal{E} = \mathcal{P}$, then $BR(t_{L1}, \sigma, \overline{s}) = L1$ if and only if

$$\int_{1/3}^{1/2} f_{L1}(x, x|\sigma)\, dx + \int_{1/3}^{1/2} f_{L1}(x, 1-2x|\sigma)\, dx \geq \frac{1}{2}\int_{1/3}^{1/2} f_{L1}(1-2x, x|\sigma)\, dx, \tag{3.3.1}$$

and $BR(t_{L2}, \sigma, \overline{s}) = L2$ if and only if

$$\int_{1/3}^{1/2} f_{L2}(x, x|\sigma)\, dx + \int_{1/3}^{1/2} f_{L2}(1-2x, x|\sigma)\, dx \geq \frac{1}{2}\int_{1/3}^{1/2} f_{L2}(x, 1-2x|\sigma)\, dx. \tag{3.3.2}$$

To see the intuition behind Lemma 3, consider a type t_{L1} voter's decision. Given the other $n-1$ voters' decisions, there are three types of scenarios in which her decision is pivotal: (i) $L1$ and $L2$ obtain nearly the same vote shares, (ii) $L1$ and R obtain nearly the same vote shares, and (iii) $L2$ and R obtain nearly the same vote shares.

For scenario (i), there are three types of events: (i-a) $L1$ obtains exactly one more vote than $L2$, (i-b) $L1$ and $L2$ obtain exactly the same support, and (i-c) $L1$ obtains exactly one less support than $L2$. Now, if she (a type t_{L1} voter) votes for $L1$ rather than for $L2$, she receives additional utility of $1/4$ for event (i-a), $1/2$ for event (i-b), and $1/4$ for event (i-c).[21] When n is sufficiently large, the probability of each event is proportional to $\int_{1/3}^{1/2} f_{L1}(x, x|\sigma)\, dx$. Hence, the first term in the left-hand side of inequality (3.3.1) represents the marginal utility of voting for $L1$ instead of $L2$.

Similarly, when n is sufficiently large, the probabilities of scenarios (ii) and (iii) are proportional to $\int_{1/3}^{1/2} f_{L1}(x, 1-2x|\sigma)\, dx$ and $\int_{1/3}^{1/2} f_{L1}(1-2x, x|\sigma)\, dx$, respectively. The marginal utility of voting for $L1$ instead of $L2$ is one for scenario (ii), but $-1/2$ for scenario (iii). Therefore, inequality (3.3.1) represents the condition in which voting for $L1$ is optimal for a type t_{L1} voter.

Given Lemma 3, the following lemma describes how voters take the results of the pre-election opinion poll into account for their voting decisions.

[20] Since the lemma is essentially the same as Proposition 1 in Hummel (2012), I omit its proof.

[21] For instance, for event (i-a), a type t_{L1} voter obtains a utility of 1 for sure if she votes for $L1$, but her expected utility becomes $3/4$ if she votes for $L2$.

Lemma 4 *Let $\mathcal{E} = \mathcal{P}$. For each $j \in \{L1, L2\}$, there exists a function $\gamma_j : [0,1] \to [0,1]$ such that $BR(t_{j'}, \sigma, \overline{s}) = j$ for all $t_{j'} \in \{t_{L1}, t_{L2}\}$ if and only if $\sigma_j/m > \gamma_j(\sigma_R/m)$.*

The lemma states that voters who prefer the left candidates would concentrate their support for a common left candidate if one candidate is doing significantly better than the other left candidate. An opinion poll result σ provides information about the expected performance of $L1$ in the election. For a given σ_R/m, the larger σ_{L1}/m is the better expected performance of $L1$ in the election. Thus, if σ_{L1}/m is larger than a threshold $\gamma_{L1}(\sigma_R/m)$, then a type t_{L2} voter chooses $L1$ although she believes that other type t_{L2} voters choose $L2$.

On the other hand, any poll result where $L1$ and $L2$ are doing almost equally well in the poll does not encourage type t_{L1} or type t_{L2} voters to choose their second preferred candidate. Consider the case of $\sigma_{L1} = \sigma_{L2}$. Although the opinion poll equally evaluates the expected performance of candidates $L1$ and $L2$, type t_{L1} and type t_{L2} voters differently evaluate their pivotal probabilities, and they vote for their most preferred candidates. For instance, since a type t_{L1} voter has private information from realization of her own type, she believes that $L1$ and R, or $L1$ and $L2$ are more likely to compete than $L2$ and R. Thus, she votes for $L1$, her most preferred candidate.[22]

Lemma 3 and Lemma 4 clearly characterize the set of opinion poll results in which voters concentrate their support in the strong voting equilibrium. Let $\Sigma_{L1}^{\mathcal{P}}$ be the set of opinion poll results where type t_{L1} and type t_{L2} voters vote for $L1$ as the best responses to the uninformative voting strategy; that is, the set of opinion poll results such that inequality (3.3.1) holds, but inequality (3.3.2) does not. To qualify these behaviors as an equilibrium behavior, I still have to check whether voters are best-responding to other voters' behavior. In fact, since n is supposed to be sufficiently large and there are only two types of left voters, no voter has an incentive to change her mind.

Specifically, first note that type t_{L1} voters have no reason to switch their decisions because the probability that $L1$ wins the election is maximized. For type t_{L2} voters, if they believe that other type t_{L1} and type t_{L2} voter votes for $L1$, then any unilateral deviation to vote for $L2$ is weakly dominated: $L2$ never wins the election whenever there are at least three voters, but there is a possibility that R wins with one margin

[22] Therefore, there is no σ such that $\sigma_j/m > \gamma_j(\sigma_R/m)$ and $\sigma_{j'}/m > \gamma_{j'}(\sigma_R/m)$. Moreover, the impossibility of $\sigma_j/m > \gamma_j(\sigma_R/m)$ and $\sigma_{j'}/m > \gamma_{j'}(\sigma_R/m)$ holds for asymmetric f, as shown in Appendix C.

of victory. Hence, for any poll result $\sigma \in \Sigma^{\mathcal{P}}_{L1}$, voting for $L1$ is an equilibrium behavior for both type t_{L1} and type t_{L2} voters.

Likewise, I define $\Sigma^{\mathcal{P}}_{L2}$ to be the set of opinion poll results where type t_{L1} and type t_{L2} voters vote for $L2$, as the best responses to the uninformative voting strategy. For any poll result $\sigma \in \Sigma^{\mathcal{P}}_{L2}$, voting for $L2$ is an equilibrium behavior for both type t_{L1} and type t_{L2} voters as for $\Sigma^{\mathcal{P}}_{L1}$. For any poll result $\sigma \notin \Sigma^{\mathcal{P}}_{L1} \cup \Sigma^{\mathcal{P}}_{L2}$, each voter votes for her most preferred candidate, as the best response to the uninformative voting strategy. This is obviously an equilibrium behavior because each voter's belief about the other voters, voting for their most preferred candidates, is self-fulfilling.

The following proposition summarizes the strong voting equilibrium.

Proposition 15 *In plurality elections, there exists a unique strong voting equilibrium in which*

(i) *type t_{L1} and type t_{L2} voters vote for $L1$ if $\sigma \in \Sigma^{\mathcal{P}}_{L1}$;*

(ii) *type t_{L1} and type t_{L2} voters vote for $L2$ if $\sigma \in \Sigma^{\mathcal{P}}_{L2}$;*

(iii) *type t_{L1} voters vote for $L1$, but type t_{L2} voters vote for $L2$ if $\sigma \in \Sigma^{\mathcal{P}}_{S}$.*

Moreover, $\Sigma^{\mathcal{P}}_{L1}$, $\Sigma^{\mathcal{P}}_{L2}$, and $\Sigma^{\mathcal{P}}_{S}$ are mutually exclusive and exhaustive.

3.3.2 Pre-Electoral Coalition Bargaining

I formulate the PEC formation bargaining process as a Nash bargaining problem. To define a bargaining environment, I find the status quo utilities (or disagreement utilities) of the left candidates, and the set of feasible agreements. These two elements depend on the opinion poll result σ, which represents the information candidates have in period 3.

The status quo utilities for the left candidates are the expected utilities when they decline to form a PEC. Formally, for each $j \in \{L1, L2\}$, let $\Phi^{\mathcal{P}}_j(\sigma)$ be the probability that candidate j wins the plurality election if the left candidates run independently. Then, the status quo utility of a left candidate j is

$$\overline{u}^{\mathcal{P}}_j(\sigma) = (V+1)\Phi^{\mathcal{P}}_j(\sigma) + \frac{1}{2}\Phi^{\mathcal{P}}_{j'}(\sigma),$$

where j' denotes the other left candidate.

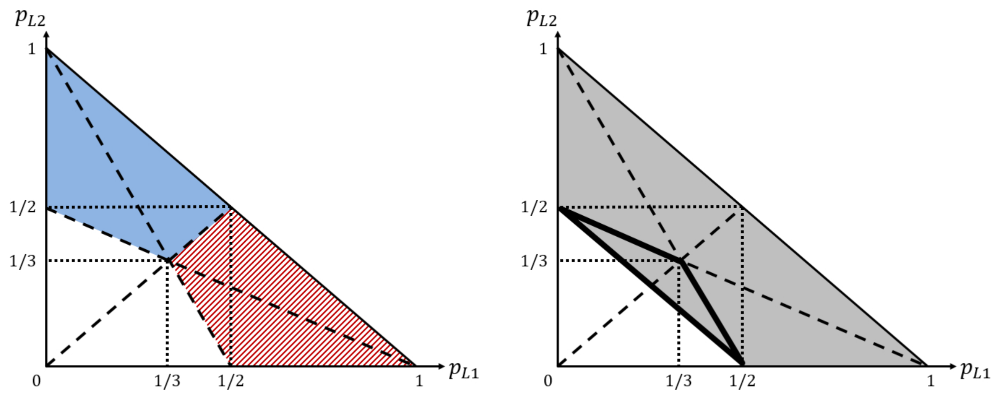
To define the set of feasible agreements, let $\Phi^{\mathcal{P}}_{\xi}(\sigma)$ be the probability that a PEC wins at σ. If $\xi(L1, \lambda)$ forms and wins, for example, then $L1$ obtains a utility of $V + (1+\lambda)/2$ and $L2$ obtains a utility of $(2-\lambda)/2$. The set of feasible agreements by nominating $L1$ is

$$G^{\mathcal{P}}_{L1}(\sigma) = \left\{(x, y) \in \mathbb{R}^2 | (x, y) = \Phi^{\mathcal{P}}_{\xi}(\sigma)(V + (1+\lambda)/2, (2-\lambda)/2), \lambda \in [0, 1]\right\}.$$

Similarly, the set of feasible agreements having $L2$ as the nominated candidate is

$$G_{L2}^{\mathcal{P}}(\sigma) = \left\{(x, y) \in \mathbb{R}^2 | (x, y) = \Phi_{\xi}^{\mathcal{P}}(\sigma)((1 + \lambda)/2, V + (2 - \lambda)/2), \lambda \in [0, 1]\right\}.$$

Therefore, the set of feasible agreements at σ is $G^{\mathcal{P}}(\sigma) = G_{L1}^{\mathcal{P}}(\sigma) \cup G_{L2}^{\mathcal{P}}(\sigma)$.[23]



(a) winning events without a PEC

(b) winning events with a PEC

Figure 3.1: Illustration of the additional winning event under plurality rule

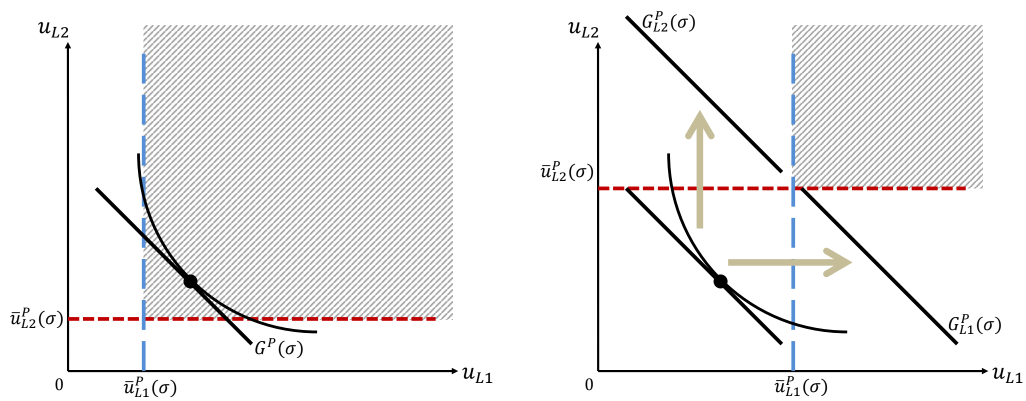
Figure 3.1 illustrates how to calculate the probabilities of winning for the left candidates with and without forming a PEC at $\sigma \in \Sigma_S^{\mathcal{P}}$. Under plurality rule with large n, $L1$ wins the election if $p_{L1} = \max\{p_{L1}, p_{L2}, p_R\}$, which corresponds to the case where p is located in the red stripped region in Figure 3.1(a). Similarly, $L2$ wins if p is located in the blue region, and R wins if p is located in the white region. On the other hand, if $\xi(k, \lambda)$ forms, then the coalition wins if $p_{L1} + p_{L2} \geq 1/2$, which corresponds to the grey region in Figure 3.1(b). This region is strictly larger than the union of the red stripped and blue regions in Figure 3.1(a); that is, if $L1$ and $L2$ form a coalition, then there exists an additional winning event, which corresponds to the bordered region in Figure 3.1(b). When p is located in this region, type t_{L1} voters vote for $L1$, and type t_{L2} voters vote for $L2$. Since $\max\{p_{L1}, p_{L2}\} < p_R < p_{L1} + p_{L2}$ in this region, R wins if and only if there is no PEC. Therefore, since f has full support, the left candidates strictly increase their probability of winning by forming a PEC, and they split the extra benefit by transferring utilities within the set of feasible agreements.

[23] $G^{\mathcal{P}}(\sigma)$ may not be convex as the standard Nash bargaining problem. Alternatively, one can include coin-tossing bargaining outcomes to make the set of feasible agreements convex. In this case, however, the same qualitative comparison of the election rules follows.

With the above expressions, I formulate the PEC formation problem as follows:

$$\begin{aligned} \text{maximize} \quad & \left(u_{L1} - \overline{u}^{\mathcal{P}}_{L1}(\sigma)\right)\left(u_{L2} - \overline{u}^{\mathcal{P}}_{L2}(\sigma)\right) \\ \text{subject to} \quad & (u_{L1}, u_{L2}) \geq (\overline{u}^{\mathcal{P}}_{L1}(\sigma), \overline{u}^{\mathcal{P}}_{L2}(\sigma)) \\ & (u_{L1}, u_{L2}) \in G^{\mathcal{P}}(\sigma). \end{aligned} \tag{3.3.3}$$

Figure 3.2(a) illustrates the bargaining problem when $V = 0$ and $\sigma \in \Sigma^{\mathcal{P}}_{S}$. The blue and red dotted lines represent the status quo utilities, and the solid line represents the set of feasible agreements. The solid curve is the objective function of the bargaining problem (3.3.3). Note that since the probability that a PEC wins is strictly greater than the sum of probabilities that each left candidate wins without a PEC, the left candidates can agree to form a PEC and obtain strictly higher payoffs than the status quo utilities. In the figure, this fact is captured by $G^{\mathcal{P}}(\sigma)$ intersecting the stripped region where $(u_{L1}, u_{L2}) \geq (\overline{u}^{\mathcal{P}}_{L1}(\sigma), \overline{u}^{\mathcal{P}}_{L2}(\sigma))$. Consequently, the left candidates form a PEC represented by the dot in the figure.



(a) $V = 0$: a PEC forms

(b) $V > \overline{V}^{\mathcal{P}}(\sigma)$: no PEC forms

Figure 3.2: Illustration of the Nash bargaining problem

On the other hand, the intersection of $G^{\mathcal{P}}(\sigma)$ with the stripped region shrinks and eventually disappears as V increases. Figure 3.2(b) illustrates the situation where $\Phi^{\mathcal{P}}_{L1}(\sigma) > \Phi^{\mathcal{P}}_{L2}(\sigma)$ and V is larger than the threshold $\overline{V}^{\mathcal{P}}(\sigma)$. Note that $G^{\mathcal{P}}_{L1}(\sigma)$ is located below the line of $u_{L2} = \overline{u}^{\mathcal{P}}_{L2}$. This condition corresponds to the situation where, having $L1$ as the nominated candidate, there is no way to persuade $L2$ to form a coalition even if the coalition chooses $L2$'s ideology. Of course, since $\Phi^{\mathcal{P}}_{L1}(\sigma) > \Phi^{\mathcal{P}}_{L2}(\sigma)$ implies $\overline{u}^{\mathcal{P}}_{L1}(\sigma) > \overline{u}^{\mathcal{P}}_{L2}(\sigma)$, there is no way to form a PEC by nominating $L2$. Therefore, the left candidates end up in disagreement. With this observation, I formalize the condition in which (3.3.3) has a solution as follows.

Proposition 16 *In plurality elections, the left candidates form a PEC after observing opinion poll result σ if and only if the office value V is less than or equal to a threshold $\overline{V}^{\mathcal{P}}(\sigma)$.*

A solution of the bargaining problem (3.3.3) is unique in terms of the expected utilities of the left candidates; that is, there exists a unique pair of utilities $(u^*_{L1}(\sigma), u^*_{L2}(\sigma))$ such that $(u^*_{L1}(\sigma), u^*_{L2}(\sigma))$ solves (3.3.3) for a given poll result σ. However, there might be two different ways to represent the bargaining solution as the PEC formation, say $\xi(L1, \lambda^*)$ and $\xi(L2, \lambda^{**})$, that return the same utilities for the left candidates. One pair represents a PEC having $L1$ as the nominated candidate, and the other pair represents a PEC having $L2$ as the nominated candidate.[24]

To see this, suppose $V > 0$ and there are two bargaining representations, $\xi(L2, \lambda^{**})$ and $\xi(L2, \lambda^{**})$. Since the office value is strictly greater than V, it must be that $\lambda^* < \lambda^{**}$; that is, by being nominated, $L1$ must transfer more utilities to $L2$ by choosing a joint policy platform close to $L2$'s ideal policy. This in turn implies that type t_{L1} voters obtain lower utility from the PEC of $\xi(L1, \lambda^{**})$ than from the PEC of $\xi(L2, \lambda^*)$. Therefore, when a PEC forms at $\sigma \in \Sigma^{\mathcal{P}}_S$, there are multiples ways to calculate type t_{L1} voters' expected utility, which is problematic for welfare analysis in Section 3.5.

To solve this multiplicity problem, I assume that the candidate who has a higher probability of winning becomes the nominated candidate of the coalition.[25] In particular, this assumption is consistent with Proposition 16 in that when $V = \overline{V}^{\mathcal{P}}(\sigma)$, the only way to form a PEC is by nominating the candidate with more support.

I further suppose that, if vote concentration is expected, then $L1$ and $L2$ form a trivial coalition by nominating the candidate who is expected to receive the support and choosing his ideal policy. For example, if $\sigma \in \Sigma^{\mathcal{P}}_{L1}$, then $L1$ and $L2$ form a PEC having $L1$ as a nominated candidate and $\lambda = 1$. This assumption is empirically supported by the fact that candidates often withdraw their candidacy when they are unlikely to receive a significant amount of the vote share in elections. Theoretically, the supported candidate can offer any arbitrarily small offer to the other candidate, and the unappealing candidate takes it. In Proposition 16, this assumption is captured by the fact that $\overline{V}(\sigma)$ is infinity at $\sigma \in \Sigma^{\mathcal{P}}_{L1} \cup \Sigma^{\mathcal{P}}_{L2}$.

[24]This multiplicity of PEC representation exists only when $G^{\mathcal{P}}_{L1}(\sigma) \cap G^{\mathcal{P}}_{L2}(\sigma)$.

[25]This assumption fits the PEC formation in the 2002 Korean presidential election. Two coalition partners, Chung and Roh, used opinion polls to decide a representative candidate. As a result, Roh obtained more support and was nominated.

Observe that a micro-foundation for the above Nash bargaining problem is found as follows.[26] Consider a Rubinstein alternating-offers model with outside options where there is an exogenous risk of breakdown where the left candidates will no longer be able to continue bargaining, so that they have to run a three-way race independently. In this case, the expected utilities from the three-way race become the status quo utilities of the left candidates.[27] As a result, the resulting (subgame perfect) equilibrium payoffs approximate the Nash bargaining solution payoffs with the same status quo utilities in the current section, when the risk is small and the candidates make offers frequently.

3.4 Runoff Elections

In this section, I analyze the game under the two-round runoff rule. I then compare the two election rules in terms of the threshold office value V to form a PEC.

3.4.1 Voting Equilibrium

I first show that regardless of opinion poll results, the best response to the uninformative voting strategy is to vote for the most preferred candidate in two-round runoff elections.

Lemma 5 *If $\mathcal{E} = \mathcal{R}$, then $BR(t_j, \sigma, \overline{s}) = j$ for all $t_j \in T$ and all $\sigma \in \Sigma$.*

To gain the intuition behind this lemma, recall that voters care only about pivotal probabilities in the first round due to the simplifying assumptions for the second round voting behavior. There are two types of events where a voter's decision can be pivotal in the first round. The first case is when one candidate obtains almost half of the vote share and one of the other candidates obtains another half of the votes. In this case, depending on what a voter decides, one candidate may win outright, or the two candidates can advance to the second round. The other case is when one top candidate obtains fewer than half of the votes, and the other two candidates obtain almost the same vote shares. In this case, depending on a voter's choice, only one of the trailing candidates advances to the second round.

When n is sufficiently large, the events of the first type are negligible relative to the events of the second type in terms of probability. To see this, first consider

[26] See Muthoo (1999) for a proof and more discussions about the relationships between Nash Bargaining problems and Rubinstein alternating-offers models.

[27] Each left candidate's status quo utility is the utility he obtains when the left candidates perpetually disagree in the bargaining process. If p is an exogenous probability that the bargaining process terminates, then the candidate j's status quo utility is calculated as $p\overline{u}_j^{\mathcal{P}}(\sigma) \sum_{s=0}^{\infty}(1-p)^s = \overline{u}_j^{\mathcal{P}}(\sigma)$.

an event of the first type in which $L1$ and $L2$ get fifty percent of the vote share approximately. Then, in large elections, the probability that both $L1$ and $L2$ receive the fraction of the vote share equal to $1/2$ is close to the probability that $p_{L1} = p_{L2} = 1/2$, which is in turn related to the density at $(p_{L1}, p_{L2}) = (1/2, 1/2)$. Consider another event of the second type in which $L1$ and $L2$ obtain almost the same vote share, and R takes first place. The probability of this event is close to the probability that $p_{L1} = p_{L2} \leq p_R \leq 1/2$, which is in turn related to the density at $(p_{L1}, p_{L2}) = (x, x)$ with $x \in [1/4, 1/3]$.

As n increases, the probability of the first event decreases to zero at a speed of the order of n^{-2}, but the probability of the second event decreases to zero at a speed of the order of n^{-1}. The same convergence rates hold for other pivotal events of the same types. Therefore, due to the different convergence rates, the events of the first type are negligible in terms of a type t_{L1} or t_{L2} voter's choice when n is sufficiently large.

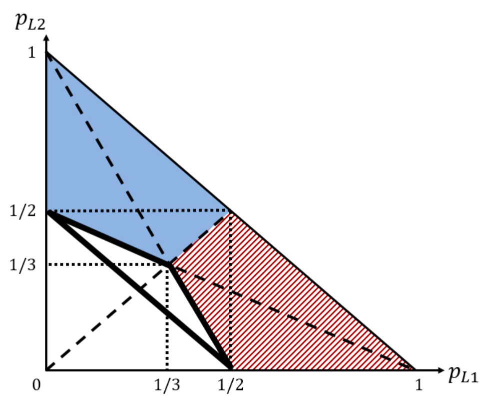
Among the events of the second type, only the event where R takes first place (but gains less than fifty percent of the vote share) and $L1$ and $L2$ get almost half of the remaining vote share affects decisions of type t_{L1} and type t_{L2} voters. To see this, first observe that when n is sufficiently large, a type t_{L1} or type t_{L2} voter does not consider the event where $L1$ or $L2$ takes first place. For instance, $L1$ becomes the top candidate in the first round if $p_{L1} > \max\{p_{L2}, p_R\}$. Then, in the second round, $L1$ gets more than half of the vote share regardless which candidates advance to the second round. Thus, only the event where R takes first place and $L1$ and $L2$ get almost the same vote share is relevant. Given this, it is obviously optimal for a type t_j voter to vote for candidate j. Therefore, characterization of the strong voting equilibrium under the two-round runoff rule is straightforward, and we have the following proposition.

Proposition 17 *In two-round runoff elections, there exists a unique strong voting equilibrium in which every voter votes for her most preferred candidate for all opinion poll results.*

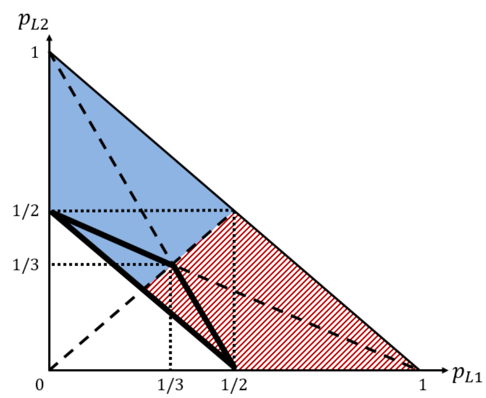
3.4.2 Pre-Electoral Coalition Bargaining

The probabilities of winning significantly change under the two-round runoff rule. Consider the case where $p_{L1} = 1/3 + \varepsilon$, $p_{L2} = 1/6$, and $p_R = 1/2 - \varepsilon$ for some positive small real number ε. In this case, R wins outright under plurality rule if no PEC forms. On the other hand, under the two-round runoff rule, R cannot obtain majority support in the first round, and he reaches the second round with $L1$. In the second round, both type t_{L1} and type t_{L2} voters choose $L1$; consequently, $L1$ ultimately wins.

Figure 3.3 illustrates the above differences. Assuming that there is no PEC and voters vote for their most preferred candidates, the red stripped regions represent the winning events for candidate $L1$ under different election rules; likewise, the blue regions represent the winning events for candidate $L2$ under different election rules.



(a) winning events under plurality rule



(b) winning events under the two-round runoff rule

Figure 3.3: Illustration of the winning events under different election rules

There are two facts to notice. First, the winning event for each left candidate is strictly larger under the two-round runoff rule than the one under plurality rule. This implies that each left candidate is more likely to win under the two-round rule without a PEC. In turn, if other things remain the same, $L1$ and $L2$ are more likely to win the election alone, and their status quo utilities are strictly greater under the two-round runoff rule than under plurality rule. Second, the union of the red stripped and blue regions in Figure 3.3(b) includes the bordered region in Figure 3.3(a). This feature means that if other things remain the same, the probability that a PEC wins the election is the same under the two election rules.

In sum, the left candidates can obtain higher utilities in the two-round runoff elections without PEC formation, and the chances to win the elections by forming a PEC remains the same. In terms of the bargaining environment, the two effects stated above diminish incentives to form a PEC, which in turn means that the threshold V decreases for each opinion poll result σ, compared to the threshold V under plurality rule.

The following proposition summarizes the characterization and comparison of PEC formation incentives under the two election rules.

Proposition 18 *In two-round runoff elections, the left candidates form a PEC after observing opinion poll result σ if and only if the office value V is less than or equal to a threshold $\overline{V}^{\mathcal{R}}(\sigma)$. In addition, the threshold under plurality rule is strictly larger than the threshold under the two-round runoff rule for any given opinion poll result σ: $\overline{V}^{\mathcal{P}}(\sigma) > \overline{V}^{\mathcal{R}}(\sigma)$ for all $\sigma \in \Sigma$.*

This result induces a clear testable hypothesis: PECs are more likely to form in plurality elections than in two-round runoff elections. This prediction is the result of two effects. First, strategic voters merge their support only under plurality rule, and whenever this behavior is expected, candidates form a trivial PEC. Second, even when voters are expected not to integrate their support, it is easier for the left candidates to form a PEC under plurality rule. These two forces have the same impact in terms of the ex-ante likelihood of PEC formation; thus, more PECs should be observed in presidential elections under plurality rule.

The above hypothesis is consistent with observations of French and Korean presidential elections. French presidential elections are under the two-round runoff rule, but Korean presidential elections are under plurality rule. Among the presidential elections in these two countries, I observe that 80% of Korean presidential elections had at least one PEC, and this fraction is significantly higher than 33% of elections with PECs in French presidential elections.[28] Thus, it is definitely worth working on more careful empirical analysis on the relationships between election rules and PEC formation with a richer dataset.

3.5 Welfare Analysis

In this section, I analyze which election rule is more favorable for which types of voters under the potential presence of PECs. I define the type t_j voter's *welfare* as her ex-ante utility before starting the game. Formally, I define a function $W_j^{\mathcal{E}} : \mathbb{R}_+ \to \mathbb{R}_+$ such that $W_j^{\mathcal{E}}(V)$ is the welfare of a type t_j voter under election rule $\mathcal{E}$ at office value V. Note that since f is symmetric and the Nash bargaining solution is symmetric, $W_{L1}^{\mathcal{E}}(V) = W_{L2}^{\mathcal{E}}(V)$ for all V. Moreover, it suffices to consider the welfare of a

[28]The calculations include nine French presidential elections in the period of 1965 - 2012 and five Korean presidential elections in the period of 1992–2012. I exclude Korean presidential elections before 1992; the 1992 presidential election was the first free and fair election in which South Korea elected its first civilian president (Golder, 2006b)

type t_{L1} voter because the sum of welfare levels corresponding to all three types is assumed to be a constant.

The following proposition asserts that the two-round runoff elections return a larger welfare to the type t_{L1} voters.

Proposition 19 *The welfare of type $L1$ voters is higher under plurality rule than under the two-round runoff rule: $W^{\mathcal{R}}_{L1}(V) \geq W^{\mathcal{P}}_{L1}(V)$ for all $V \in \mathbb{R}_+$.*

The proposition can be understood with Figure 3.4 that depicts $W^{\mathcal{R}}_{L1}(V)$ and $W^{\mathcal{P}}_{R}(V)$ as functions of V for $m = 1$ and the distribution of p is uniform. When $m = 1$, voters always choose their most preferred candidate under plurality rule for any realized opinion poll result σ because inequalities (3.3.1) and (3.3.2) are simultaneously satisfied. This means that $\Sigma^{\mathcal{P}}_{S} = \Sigma$, and the welfare of type t_{L1} voters is solely determined by whether the left candidates form a PEC.

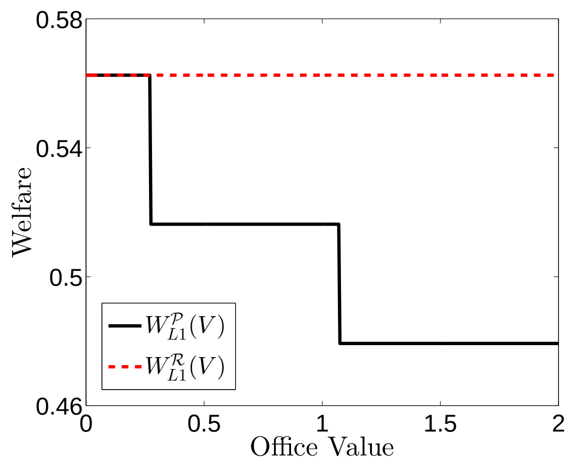


Figure 3.4: Illustration of the welfare of a type t_{L1} voter

Consider two different office values, $V = 0, 1/2$. When the candidates are purely policy-motivated ($V = 0$), the left candidates form a PEC regardless of the realization of the opinion poll σ as discussed in Section 3.3.2. However, when candidates obtain value by holding the office ($V = 1/2$), the left candidates do not form a PEC if the opinion poll realizes as $\sigma = (1, 0)$ because this poll result indicates that the probability that candidate $L1$ wins without any PEC is substantially high, and it is better for him to finish the race alone and try to take office value $1/2$. Likewise, if the opinion poll realizes as $\sigma = (0, 1)$, then no PEC forms because $L2$ wants to take the office value alone.

This selfish motivation when $V = 1/2$ indeed decreases the welfare of type t_{L1} voters because voters only care about the ideal policy of the candidates. Suppose that p is located in the bordered region in Figure 3.3(a). When $V = 0$, one of the left candidates wins for sure, but when $V = 1/2$, neither of them wins if the opinion poll result realizes as $\sigma = (1,0), (0,1)$. Hence, type t_{L1} voters obtain strictly less ex-ante utility if $V = 1/2$. In Figure 3.4, this decrease of type t_{L1} voter's welfare is captured by a drop of $W^{\mathcal{P}}_{L1}(V)$ at $V = 3/11$.[29]

However, under the two-round runoff rule, the welfare of a t_{L1} voter is constant. To see why, consider the situation where $V = 2$ and p is located in the region where the left candidates can win only by forming a PEC under plurality rule. Regardless of opinion poll results, no PEC forms because V is large.[30] However, one left candidate still wins the election under the two-round runoff rule because type t_{L1} and type t_{L2} voters are the majority of the population. Thus, no matter whether a PEC forms, a type t_{L1} voter obtains a strictly positive payoff whenever p is located in the union of the red and blue regions in Figure 3.3(b). Thus, the welfare of a type t_R is constant. Since f is symmetric, the ex-ante utilities of type t_{L1} and type t_{L2} voters' welfare are the same, which in turn implies that a type t_{L1} voter's welfare is also constant.

The key factor determining welfare of voters is whether not only voters, but also candidates correctly guess the true p and make the right decision to win. Intuitively, as their size increases, poll results convey more accurate information on the distribution of voters' preferences. Thus, voters and candidates can correctly infer the true value of p and merge their support when a PEC is necessary to win; in turn, the welfare of the voters who prefer the left candidates increases.

To formalize the above discussion, let $W^{\mathcal{P}}_{L1}(V,m)$ be the welfare of a type t_{L1} voter under plurality rule at (V,m). The following proposition states the welfare effect of the poll size.

Proposition 20 *In terms of the welfare of a type t_{L1} voter, plurality and the two-round runoff rules become indifferent as the poll size increases to infinity: for a given $V \in \mathbb{R}_+$, $W^{\mathcal{P}}_{L1}(V,m)$ converges to $W^{\mathcal{R}}_{L1}(V)$ as $m \to \infty$.*

Figure 3.5 illustrates the proposition. The convergence is driven by more PEC formations of the left candidates as well as more vote concentration when p is located in the region where the left candidates can win only by forming a PEC. As before, type t_{L1} and type t_{L2} voters never concentrate their support if $\sigma_{L1} \approx \sigma_{L2}$. On the other hand, for large m, the left candidates form a PEC whenever $\max\{\sigma_{L1}, \sigma_{L2}\} <$

[29] Another drop around $V = 1$ is due to no PEC formation at $\sigma = (0,0)$.

[30] This observation follows from Proposition 18 and the fact that no PEC forms at $V = 2$ for all $\sigma \in \Sigma$ under plurality rule.

$\sigma_R < \sigma_{L1} + \sigma_{L2}$; if any such opinion poll result is disclosed, the left candidates can correctly predict that they can defeat R only by forming a PEC. Therefore, as m increases, plurality rule and the two-round runoff rule become indifferent in terms of welfare.

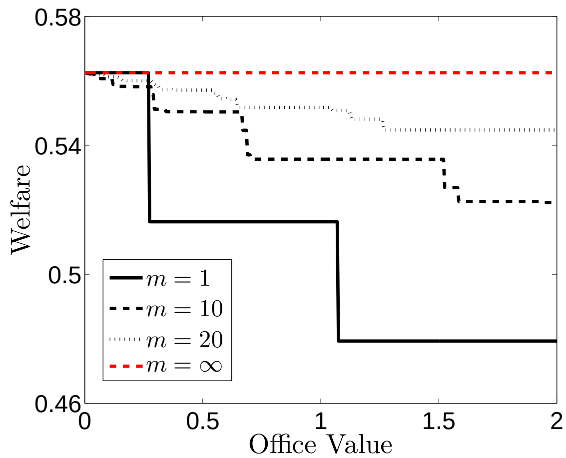


Figure 3.5: Convergence of the welfare of a type t_{L1} voter

3.6 Comparative Statics

In this section, I provide two comparative static results. I first parameterize two-round runoff elections and compare PEC incentives. Then, I study how PEC incentives change as a function of ideological distance between the two left candidates.

3.6.1 General Two-Round Runoff Elections

General two-round runoff rules can be represented by a single parameter $\zeta \in (1/3, 1/2)$ such that ζ is the required threshold for a candidate to win outright.[31] To begin, the arguments underlying Lemma 5 imply that voters always choose their most preferred candidates for all ζ. Thus, I focus on the threshold V to form a PEC and obtain the following comparative static result.

Proposition 21 *Let $\overline{V}^{\zeta}(\sigma)$ be the threshold office value for the left candidates to form a PEC when ζ is the threshold for the first-round victory, after observing opinion poll result σ. Then, the threshold office value is smaller than the threshold under plurality, but larger than the threshold under the two-round runoff rule:*

[31] I follow the approach of Bouton (2013). If $\zeta = 1/3$, the corresponding two-round runoff rule is equivalent to plurality rule. The two-round runoff rule studied in the previous sections is equivalent to a runoff rule with $\zeta = 1/2$.

$\overline{V}^{\mathcal{P}}(\sigma) > \overline{V}^{\zeta}(\sigma) > \overline{V}^{\mathcal{R}}(\sigma)$. *In addition, the threshold office value decreases as the threshold for first-round victory increases.*

Figure 3.6 illustrates the events where the left candidates can win in a two-round runoff election with threshold ζ. The bordered region represents the set of events such that the left candidates can win only by forming a PEC. If p is located in this region, R obtains not only majority support in the first round, but also the support higher than the threshold ζ. Thus, R wins outright if $L1$ and $L2$ do not form a PEC. On the other hand, the left candidates can defeat R by forming a PEC. For each threshold ζ, the additional event that guarantees victory of the left candidates is strictly contained in the added winning event under plurality rule. Thus, the left candidates are more likely to win without a PEC under any runoff rule other than under plurality rule, and so the threshold V under plurality rule is higher than any runoff rule.

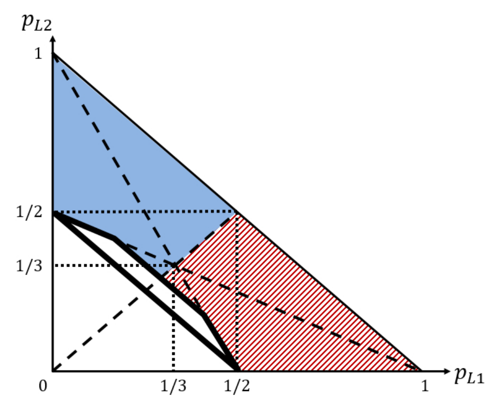


Figure 3.6: Illustration of the winning events when $\mathcal{E} = \zeta$

Note that the size of this additional event decreases in ζ because the higher ζ is in favor of the left candidates. This observation implies that the threshold V is larger than the threshold under the two-round runoff rule with threshold $1/2$. Therefore, in two-round runoff elections, PECs are more likely to form as the threshold for first-round victory decreases.

3.6.2 Ideological Distance

I now parameterize the ideological distance between the left candidates by $\theta \in [0, 1]$: if $L1$ wins the election alone, then $L2$ and type t_{L2} voters obtain a utility of $1 - \theta$, and vice versa. With this parametrization, as an extension of Lemma 3, I have the following lemma.

Lemma 6 *If $\mathcal{E} = \mathcal{P}$, then $BR(t_{L1}, \sigma, \overline{s}) = L1$ if and only if*

$$2\theta \int_{1/3}^{1/2} f_{L1}(x, x|\sigma)\, dx + \int_{1/3}^{1/2} f_{L1}(x, 1-2x|\sigma)\, dx \geq (1-\theta) \int_{1/3}^{1/2} f_{L1}(1-2x, x|\sigma)\, dx,$$

and $BR(t_{L2}, \sigma, \overline{s}) = L2$ if and only if

$$2\theta \int_{1/3}^{1/2} f_{L2}(x, x|\sigma)\, dx + \int_{1/3}^{1/2} f_{L2}(1-2x, x|\sigma)\, dx \geq (1-\theta) \int_{1/3}^{1/2} f_{L2}(x, 1-2x|\sigma)\, dx.$$

This lemma is quite intuitive because as ideological distance increases, the marginal benefit of voting for the second preferred candidate decreases. Although voters are less likely to merge their support, the next proposition states that the left candidates's incentives to merge their support increase as their ideological distance increases.

Proposition 22 *For any opinion poll result σ that predicts divided support, the threshold office value $\overline{V}^{\mathcal{E}}(\sigma, \theta)$ for the left candidates to form a PEC after observing σ increases as the ideological distance increases.*

The proposition states that an increase in ideological distance has a positive effect on the probability of winning through increasing PEC formation incentives. To see why, first note that ideological distance has no influence on the probability of winning. The maximum utility transfer by choosing the opponent's ideal policy is θ and the status quo utility is $1-\theta$. Hence, as θ increases, the left candidates' status quo utilities decrease, and the set of feasible utility transfers expands, which in turn implies that $L1$ and $L2$ are more willing to form a PEC. Therefore, conditional on divided support, PECs are more likely to form as the ideological distance between potential coalition partners increases.

3.7 Concluding Remarks

Although only one coalition member can occupy the office, PECs have frequently formed and influenced outcomes in single-office elections. In my model, candidates are both office- and policy-motivated. For this reason, forming a PEC could be beneficial to all coalition members including the one who withdraws his candidacy. By throwing his support behind the representative candidate, the coalition becomes more likely to win and implement a policy more to his liking. In this setting, the current paper provides a series of empirically testable predictions in the literature on PEC formation.

Adding one more candidate who can form a coalition with candidate R does not change the qualitative nature of PEC incentives. Of course, incentives to form a PEC between coalition partners will depend on their beliefs about what candidates of the other-side would do. For instance, if the right candidates are expected to merge their support by forming a PEC, then the left candidates are more willing to form a coalition. However, the comparative static results in this paper remain the same. As an example, *ceteris paribus*, the increase of probability of winning by forming a PEC under plurality rule is strictly larger than the increase under the two-round runoff rule. Thus, PECs are more likely to form in plurality elections than runoff elections.

With the generalization of the number of candidates, one salient extension would be to allow some candidate to choose their coalition partners. In reality, it is easier for ideologically moderate candidates to form a PEC with other candidates than ideological extremists, and they utilize this flexibility as a leverage in the PEC bargaining process. One thing to notice is that if their office motivation dominates policy motivation, such candidates would try to be nominated and occupy the office in exchange for implementing their ideal policy. Therefore, the resulting policy choice by a PEC including those candidates might be more extreme than without the possibility of PEC formation. Since this new asymmetric incentive structure affects voter welfare, it is worth considering this extension as a future research topic.

BIBLIOGRAPHY

Arthur, D. et al. (2009). "Pricing Strategies for Viral Marketing on Social Networks". *Unpublished Manuscript*.

Azuma, K. (1967). "Weighted Sums of Certain Dependent Random Variables". *Tohoku Mathematical Journal* 19.3, pp. 357–367.

Baetz, O. (2015). "Social Activity and Network Formation". *Theoretical Economics* 10, pp. 315–340.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006). "Who's Who in Networks. Wanted: The Key Player". *Econometrica* 74.5, pp. 1403–1417.

Bandyopadhyay, S., K. Chatterjee, and T. Sjostrom (2011). "Pre-Electoral Coalitions and Post-Electoral Bargaining". *Quarterly Journal of Political Science* 6.1, pp. 1–53.

Barabási, A.-L. and R. Albert (1999). "Emergence of Scaling in Random Networks". *Science* 286.5439, pp. 509–512.

Bass, F. M. (1969). "A New Product Growth for Model Consumer Durables". *Management Science* 15.5, pp. 215–227.

Belhaj, M., Y. Bramoullé, and F. Deroän (2014). "Network Games under Strategic Complementarities". *Games and Economic Behavior* 88, pp. 310–319.

Bender, E. A. and E. R. Canfield (1978). "The Asymptotic Number of Labeled Graphs with Given Degree Sequences". *Journal of Combinatorial Theory, Series A* 24.3, pp. 296–307.

Bensaid, B. and J.-P. Lesne (1996). "Dynamic Monopoly Pricing with Network Externalities". *International Journal of Industrial Organization* 14.6, pp. 837–855.

Billingsley, P. (2012). *Probability and Measure*. 4th. Wiley.

Bollobás, B. et al. (2001). "The Degree Sequence of A Scale-Free Random Graph Process". *Random Structures and Algorithms* 18.3, pp. 279–290. ISSN: 1042-9832.

Bordignon, M., T. Nannicini, and G. Tabellini (2013). "Moderating Political Extremism: Single Round vs Runoff Elections under Plurality Rule". *Unpublished Manuscript*.

Bormann, N.-C. and M. Golder (2013). "Democratic Electoral Systems Around the World, 1946 - 2011". *Electoral Studies* 32.2, pp. 360–369.

Bouton, L. (2013). "A Theory of Strategic Voting in Runoff Elections". *American Economic Review* 103.4, pp. 1248–1288.

Bramoullé, Y., R. Kranton, and M. D'amours (2014). "Strategic Interaction and Networks". *American Economic Review* 104.3, pp. 898–930.

Bubeck, S., E. Mossel, and M. Z. Rácz (2015). "On the Influence of the Seed Graph in the Preferential Attachment Model". *IEEE Transactions on Network Science and Engineering* 2.1, pp. 30–39.

Cabral, L. M. B., D. J. Salant, and G. A. Woroch (1999). "Monopoly Pricing with Network Externalities". *International Journal of Industrial Organization* 17.2, pp. 199–214.

Callander, S. (2008). "Political Motivations". *Review of Economic Studies* 75.3, pp. 671–697.

Campbell, A. (2013). "Word-of-Mouth Communication and Percolation in Social Networks". *American Economic Review* 103.6, pp. 2466–2498.

Carroll, R. and G. W. Cox (2007). "The Logic of Gamson's Law: Pre-election Coalitions and Portfolio Allocations". *American Journal of Political Science* 51.2, pp. 300–313.

Chamberlain, G. and M. Rothschild (1981). "A Note on the Probability of Casting a Decisive Vote". *Journal of Economic Theory* 25.1, pp. 152–162.

Chatterjee, K. et al. (1993). "A Noncooperative Theory of Coalitional Bargaining". *Review of Economic Studies,* 60.2, pp. 463–477.

Chung, F., L. Lu, and V. Vu (2003). "Spectra of Random Graph with Given Expected Degrees". *Proceedings of the National Academy of Sciences of the United States of America* 100.11, pp. 6313–6318.

Coase, R. H. (1972). "Durability and Monopoly". *Journal of Law and Economics* 15.1, pp. 143–149.

Cooper, C. and A. Frieze (2003). "A General Model of Web Graphs". *Random Structures and Algorithms* 22.3, pp. 311–335.

Coppersmith, D. and S. Winograd (1990). "Matrix Multiplication via Arithmetic Progressions". *Journal of Symbolic Computation* 9.3, pp. 251–280.

Cox, G. W. (1997). *Making Votes Count: Strategic Coordination in the World's Electoral Systems (Political Economy of Institutions and Decisions)*. Cambridge University Press.

Currarini, S., M. O. Jackson, and P. Pin (2009). "An Economic Model of Friendship: Homophily, Minorities, and Segregation". *Econometrica* 77.4, pp. 1003–1045.

Cvetković, D. and P. Rowlinson (1990). "The Largest Eigenvalue of a Graph: A Survey". *Linear and Multilinear Algebra* 28.1-2, pp. 3–33.

Debus, M. (2009). "Pre-Electoral Commitment and Government Formation". *Public Choice* 138, pp. 45–64.

Dewan, T. and D. P. Myatt (2007). "Leading the Party: Coordination, Direction, and Communication". *American Political Science Review* 101.4, pp. 827–845.

Dhebar, A. and S. S. Oren (1985). "Optimal Dynamic Pricing for Expanding Networks". *Marketing Science* 4.4, pp. 336–351.

Dorogovtsev, S. N., J. F. F. Mendes, and A. N. Samukhin (2000). "Structure of Growing Networks with Preferential Linking". *Physical Review Letters* 85.21, p. 4633.

Dorogovtsev, S. N. and J. F. F. Mendes (2001). "Scaling Properties of Scale-Free Evolving Networks: Continuous Time Approach". *Physical Review E* 63.5, p. 056125.

Durrett, R. (2005). *Probability: Theory and Examples*. 3rd. Duxbury Advanced Series. Thomson.

Durrett, R. (2010). *Random Graph Dynamics*. 1st. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Duverger, M. (1954). *Political Parties: Their Organization and Activity in the Modern State*. Wiley.

Ekmekci, M. (2009). "Manipulation Through Political Endorsements". *Journal of Economic Theory* 144.3, pp. 1227–1248.

Eraslan, H. and A. Merlo (2002). "Majority Rule in a Stochastic Model of Bargaining". *Journal of Economic Theory* 103.1, pp. 31–48.

Erdős, P. and A. Rényi (1959). "On Random Graphs". *Publicationes Mathematicae Debrecen* 6, pp. 290–297.

Fainmesser, I. P. and A. Galeotti (2016). "Pricing Network Effects". *Review of Economic Studies* 83.1, pp. 165–198.

Farrell, J. and P. Klemperer (2007). "Coordination and Lock-In: Competition with Switching Costs and Network Effects". Ed. by M. Armstrong and R. H. Porter. Vol. 3. Handbook of industrial organization. North Hollan. Chap. 31, pp. 1967–2072.

Feld, S. L. (1991). "Why Your Friends Have More Friends Than You Do". *American Journal of Sociology* 96.6, pp. 1464–1477.

Feri, F. and P. Pin (2015). "The Effect of Externalities Aggregation on Network Games Outcomes". *Unpublished Manuscript*.

Fiorina, M. P. (1976). "The Voting Decision: Instrumental and Expressive Aspects". *Journal of Politics* 38.2, pp. 390–413.

Flaxman, A., A. Frieze, and T. Fenner (2005). "High Degree Vertices and Eigenvalue in the Preferential Attachment Graph". *Internet Mathematics* 2.1, pp. 1–19.

Fujiwara, T. (2011). "A Regression Discontinuity Test of Strategic Voting and Duverger's Law". *Quarterly Journal of Political Science* 6.3-4, pp. 197–233.

Galeotti, A. et al. (2010). "Network Games". *Review of Economic Studies* 77.1, pp. 218–244.

Ghiglino, C. (2012). "Random Walk to Innovation: Why Productivity Follows a Power Law". *Journal of Economic Theory* 147.2, pp. 713–737.

Golder, M. (2006a). "Presidential Coattails and Legislative Fragmentation". *American Journal of Political Science* 50.1, pp. 34–48.

Golder, S. (2006b). *The Logic of Pre-Electoral Coalition Formation*. Ohio State University Press.

Good, I. J. and L. S. Mayer (1975). "Estimating the Efficacy of a Vote". *Behavioral Science* 20.1, pp. 25–33.

Granovetter, M. (1978). "Threshold Models of Collective Behavior". *American Journal of Sociology* 83.6, pp. 1420–1443.

Hartline, J., V. S. Mirrokni, and M. Sundararajan (2008). "Optimal Marketing Strategies over Social Networks". *Proceedings of the 17th international conference on World Wide Web*, pp. 189–198.

Hodas, N. O., F. Kooti, and K. Lerman (2013). "Friendship Paradox Redux: Your Friends Are More Interesting Than You". *Unpublished Manuscript*.

Hoeffding, W. (1963). "Probability Inequalities for Sums of Bounded Random Variables". *Journal of the American Statistical Association* 58.301, pp. 13–30.

Hotelling, H. (1929). “Stability in Competition”. *Economic Journal* 39.153, pp. 41–57.

Hummel, P. (2012). “Sequential Voting in Large Elections with Multiple Candidates”. *Journal of Public Economics* 96.3-4, pp. 341–348.

Hummel, P. (2014). “Pre-Eelection Polling and Third Party Candidates”. *Social Choice and Welfare* 42.1, pp. 77–98.

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O. and B. W. Rogers (2007a). “Meeting Strangers and Friends of Friends: How Random Are Social Networks?” *American Economic Review* 97.3, pp. 890–915.

Jackson, M. O. and B. W. Rogers (2007b). “Relating Network Structure to Diffusion Properties Through Stochastic Dominance”. *The B.E. Journal of Theoretical Economics (Advances)* 7.1, pp. 1–13.

Jackson, M. O. and L. Yariv (2007). “Diffusion of Behavior and Equilibrium Properties in Network Games”. *American Economic Review* 97.2, pp. 92–98.

Jehiel, P., B. Moldovanu, and E. Stacchetti (1996). “How (Not) to Sell Nuclear Weapons”. *American Economic Review* 86.4, pp. 814–829.

Karlin, S. and H. Rubin (1956). “The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio”. *Annals of Mathematical Statistics* 27.2, pp. 272–299.

Katz, M. L. and C. Shapiro (1994). “Systems Competition and Network Effects”. *Journal of Economic Perspectives* 8.2, pp. 93–115.

König, M. D., C. J. Tessone, and Y. Zenou (2014). “Nestedness in Networks: A Theoretical Model and Some Applications”. *Theoretical Economics* 9.3, pp. 695–752.

Krapivsky, P. L., S. Redner, and F. Leyvraz (2000). “Connectivity of Growing Random Networks”. *Physical Review Letters* 85.21, pp. 4629–4632.

Kumbasar, E., A. K. Rommey, and W. H. Batchelder (1994). “Systematic Biases in Social Perception”. *American Journal of Sociology* 100.2, pp. 477–505.

López-Pintado, D. (2008). “Diffusion in Complex Social Networks”. *Games and Economic Behavior* 62.2, pp. 573–590.

Milgrom, P. and I. Segal (2002). “Envelope Theorems for Arbitrary Choice Sets”. *Econometrica* 70.2, pp. 583–601.

Morris, S. and H. S. Shin (2003). “Advances in Economics and Econometrics, the Eighth World Congress”. Ed. by M. Dewatripont, L. Hansen, and S. Turnovsky. Cambridge: Cambridge University Press. Chap. Global Games: Theory and Applications, pp. 56–114.

Muthoo, A. (1999). *Bargaining Theory with Applications*. Cambridge University Press.

Myatt, D. P. (2007). “On the Theory of Strategic Voting”. *Review of Economic Studies* 74.1, pp. 255–281.

Myerson, R. B. (1981). “Optimal Auction Design”. *Mathematics of Operations Research* 6.1, pp. 58–73.

Myerson, R. B. (1998). “Population Uncertainty and Poisson Games”. *International Journal of Game Theory* 27.3, pp. 375–392.

Myerson, R. B. (2000). “Large Poisson Games”. *Journal of Economic Theory* 94.1, pp. 7–45.

Myerson, R. B. and R. J. Weber (1993). “A Theory of Voting Equilibria”. *American Political Science Review* 87.1, pp. 102–114.

Nagashima, D. (2011). “The Formation of Pre-Electoral Coalitions: Actor Strategy and Institutional Constraints”. *Unpublished Manuscript*.

Nash, J. F. (1950). “The Bargaining Problem”. *Econometrica* 18.2, pp. 155–162.

Newman, M. E. J. (2003). “The Structure and Function of Complex Networks”. *SIAM Review* 45.2, pp. 167–256.

Niou, E. M. S. (2001). “Strategic Voting Under Plurality and Runoff Rules”. *Journal of Theoretical Politics* 13.2, pp. 209–227.

Okada, A. (2011). “Coalitional Bargaining Game with Random Proposers: Theory and Application”. *Games and Economic Behavior* 73.1, pp. 227–235.

Powell, G. B. (2000). *Elections as Instruments of Democracy*. Yale University Press.

Radner, R., A. Radunskaya, and A. Sundararajan (2014). “Dynamic Pricing of Network Goods with Boundedly Rational Consumers”. *Proceedings of the National Academy of Sciences of the United States of America* 111.1, pp. 99–104.

Ray, D. (2008). *A Game-Theoretic Perspective on Coalition Formation*. Oxford University Press.

Rohlfs, J. (1974). “A Theory of Interdependent Demand for a Telecommunications Service”. *The Bell Journal of Economics and Management Science* 5.1, pp. 16–37.

Schuessler, A. A. (2000). *A Logic of Expressive Choice*. Princeton University Press.

Shin, E. (2016a). “Monopoly Pricing and Diffusion of (Social) Network Goods”. *Unpublished Manuscript*.

Shin, E. (2016b). “Social Network Formation and Strategic Interaction in Large Networks”. *Unpublished Manuscript*.

Shy, O. (2011). “A Short Survey of Network Economics”. *Review of Industrial Organization* 38.2, pp. 119–149.

Stokey, N. L. and R. E. Lucas (1989). *Recursive Methods in Economic Dynamics*. Havard University Press.

Vohra, R. V. (2011). *Mechanism Design: A Linear Programming Approach*. Econometric Society Monographs. Cambridge University Press.

Watts, D. J. and S. H. Strogatz (1998). “Collective Dynamics of ‘Small-World’ Networks”. *Nature* 393.6684, pp. 440–442.

Young, H. P. (2009). “Innovation Diffusion in Heterogeneous Populations: Contagion, Cosial Influence, and Social learning”. *American Economic Review* 99.5, pp. 1899–1924.

Appendix A

PROOFS OF CHAPTER 1: SOCIAL NETWORK FORMATION AND STRATEGIC INTERACTION IN LARGE NETWORKS

A.1 Proofs of Results

Proof of Proposition 1

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

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

Claim 1 *For each $d \in \mathbb{N}$, $\frac{\overline{\mathbf{N}}(d,t)}{t}$ converges to $f(d)$ as $t \to \infty$.*

Proof. I start from the following rate equations:

(i) For $d = 1$:

$$\overline{\mathbf{N}}(1,t+1) = 1 + \left(1 - \frac{\Phi(1)}{\mu t}\right)\overline{\mathbf{N}}(1,t) + \varepsilon(1,t). \tag{A.1.1}$$

(ii) For $d \geq 2$:

$$\overline{\mathbf{N}}(d,t+1) = \frac{(d-1)\Phi(d-1)}{\mu t}\overline{\mathbf{N}}(d-1,t) + \left(1 - \frac{d\Phi(d)}{\mu t}\right)\overline{\mathbf{N}}(d,t) + \varepsilon(d,t). \tag{A.1.2}$$

I solve the rate equations inductively. By letting $a(1) = \frac{\Phi(1)}{\mu}$, (A.1.1) becomes

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

$$+\underbrace{\sum_{s=1}^{t}\left[\prod_{r=s+1}^{t}\left(1-\frac{a(1)}{r}\right)\varepsilon(1,r-1)\right]}_{\text{(iii)}}.$$

For a large t, I have the following approximation:

$$\prod_{r=s+1}^{t}\left(1-\frac{a(1)}{r}\right)\approx e^{-\sum_{r=s+1}^{t}\frac{a(1)}{r}}\approx e^{-a(1)(\log t-\log s)}=\left(\frac{s}{t}\right)^{a(1)}. \tag{A.1.3}$$

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

(i) $\displaystyle\sum_{s=1}^{t}\prod_{r=s+1}^{t}\left(1-\frac{a(1)}{r}\right)\approx\frac{1}{t^{a(1)}}\int_0^t s^{a(1)}\,ds=\frac{1}{t^{a(1)}}\frac{1}{1+a(1)}t^{a(1)+1}=\frac{t}{1+a(1)},$

(ii) $\displaystyle\overline{\mathbf{N}}(1,1)\prod_{r=1}^{t}\left(1-\frac{a(1)}{r}\right)\approx\overline{\mathbf{N}}(1,1)\left(\frac{s}{t}\right)^{a(1)},$

(iii) $\displaystyle\sum_{s=1}^{t}\prod_{r=s+1}^{t}\left(1-\frac{a(1)}{r}\right)\varepsilon(1,r-1)\approx\varepsilon(1,t-1)\frac{t}{1+a(1)}.$

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

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

Let $f(d-1)$ be given. Define $a(d)$ and $b(d-1,t)$ as

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

I observe that

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

Then, by using the approximation technique for $d=1$, I rewrite equation (A.1.2) as

$$\begin{aligned}
&\overline{\mathbf{N}}(d,t+1)\\
&=\left(1-\frac{a(d)}{t}\right)\overline{\mathbf{N}}(d,t)+b(d-1,t)+\varepsilon(d,t)\\
&=b(d-1,t)+\left(1-\frac{a(d)}{t}\right)b(d,t-1)+\left(1-\frac{a(d)}{t}\right)\left(1-\frac{a(d)}{t-1}\right)\overline{\mathbf{N}}(d,t-1)\\
&\quad+\varepsilon(d,t)+\left(1-\frac{a(d)}{t}\right)\varepsilon(d,t-1)
\end{aligned}$$

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

By dividing t and taking the limit, the latter two terms become zero. In addition, since $\lim\limits_{t\to\infty} b(d-1,t) = \frac{(d-1)\Phi(d-1)}{\mu} f(d-1)$, it follows that

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

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

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

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

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

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

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

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

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

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

$$
\overline{\mathbf{N}}(d,t+1) = \overline{\mathbf{N}}(d,t) + \frac{z_{d-1}}{\mu t} \overline{\mathbf{N}}(d-1,t) - \frac{z_d}{\mu t} \overline{\mathbf{N}}(d,t) + \varepsilon(d,t),
$$

$$(t+1)f(d) = tf(d) + \frac{z_{d-1}}{\mu t} tf(d-1) - \frac{z_d}{\mu t} tf(d).$$

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

$$\begin{aligned}
&\left|\overline{\mathbf{N}}(d,t+1) - (t+1)f(d)\right| \\
&\leq \left|\left(\overline{\mathbf{N}}(d,t) - tf(d)\right)\left(1 - \frac{z_d}{\mu t}\right) + \frac{z_{d-1}}{\mu t}\left(\overline{\mathbf{N}}(d-1,t) - tf(d-1)\right)\right| + \varepsilon(d,t) \\
&\leq \left(1 - \frac{z_d}{\mu t}\right)\delta(d,t) + \frac{z_{d-1}}{\mu}\nu(d-1,t) + \varepsilon(d,t) \\
&\leq \left(1 - \frac{z_d}{\mu t}\right)\left(1 - \frac{z_d}{\mu(t-1)}\right)\delta(d,t-1) + \left(1 - \frac{z_d}{\mu t}\right)\frac{z_{d-1}}{\mu}\nu(d-1,t-1) \\
&\quad + \frac{z_{d-1}}{\mu}\nu(d-1,t) + \left(1 - \frac{z_d}{\mu t}\right)\varepsilon(d,t-1) + \varepsilon(d,t) \\
&\leq \underbrace{\delta(d,1)\prod_{k=1}^{t}\left(1 - \frac{z_d}{\mu k}\right)}_{\to 0 \text{ as } t \to 0} + \underbrace{\sum_{s=1}^{t}\left\{\left[\frac{z_{d-1}}{\mu}\nu(d-1,s) + \varepsilon(d,s)\right]\prod_{r=1}^{t}\left(1 - \frac{z_d}{\mu r}\right)\right\}}_{\leq \left(\frac{z_{d-1}}{\mu}\nu(d-1,s) + \varepsilon(d,s)\right)t} \\
&\leq \nu(d,t)t. \qquad \blacksquare
\end{aligned}$$

I now show concentration of the degrees as follows.

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

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

Proof. I first find that a sequence of random variables $(\mathbf{N}(d,s))_{s=1}^{t}$ defined as $\mathbf{N}(d,s) := \mathbb{E}\left[\mathbf{N}(d,t)|\mathcal{F}^s\right]$ is a *Doob Martingale*. To see why, first find that

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

Moreover, $\mathbf{N}(d,0) = \overline{\mathbf{N}}(d,t)$ and $\mathbf{N}(d,t) = \mathbf{N}(d,t)$ because

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

$(\mathbf{N}(d,s))_{s=1}^{t}$ further has a uniformly bounded difference property: For each $d \in \mathbb{N}$, there exists $M_d > 0$ such that for all $1 \leq s \leq t$, $|\mathbf{N}(d,s) - \mathbf{N}(d,s-1)| \leq M_d$. To see why, I first note that it suffices to show it only for $d \geq 2$ because one can easily repeat similar steps for $d = 1$. Now, for any given $d \geq 2$, I set $M_d = \frac{2}{z_1} \max\{1, z_d, z_{d-1}\}$ and verify the uniformly bounded difference property by using mathematical induction in the difference $k = t - s \geq 0$.

First, when $k = 0$, I have

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

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

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

Having the properties of $(\mathbf{N}(d,s))_{s=1}^t$ as the above, the Azuma-Hoeffiding inequality states that for any $\varepsilon_d > 0$,

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

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

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

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

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

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

Proof. By Claim 2, there exists T_1 such that $|\frac{\overline{\mathbf{N}}(d,t)}{t} - f(d)| < \varepsilon/3$ whenever $t \geq T_1$. Then, for all $t \geq T_1$,

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

By Claim 3, it follows that

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

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

Claim 5 *Let* $\Gamma(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ *be a function defined as*

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

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

Proof. Define a function $\gamma_d(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ 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 $\Gamma_n(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ 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. Therefore, $\Gamma(\cdot)$ is continuous. ■

Therefore, the theorem is proven. ■

Proof of Proposition 2

Proof. The proposition is fully proven in the main text. ■

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 $d_1(\mathbf{G}^t) \to \infty$ almost surely as $t \to \infty$.

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

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

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

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

Thus, $\mathbb{P}(\sum_{t=1}^{\infty} \mathbf{I}_t = \infty) = 1$, so that $d_1(\mathbf{G}^t) \to \infty$ almost surely as $t \to \infty$. ■

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

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

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 $z_d \to \infty$ as $d \to \infty$. For a fixed $\varepsilon > 0$, there exists $\delta > 0$ such that $\varepsilon = \log(1 + \delta)$. Then, I have

$$\lim_{d\to\infty} \log \frac{\overline{F}(d)}{e^{-\varepsilon d}} = \lim_{k\to\infty} \sum_{s=1}^{d-1} \log\left(\frac{z_s}{\mu + z_s}\right) + \varepsilon d$$

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

Therefore, the proposition is proven. ■

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

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

By its construction, I have

$$
\begin{aligned}
f(d) &= (1 - a(d+1)) \prod_{k=1}^{d} a(k), \\
h(d) &= 1 - a(d+1).
\end{aligned}
$$

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 $\widetilde{f}(\cdot)$ is

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

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

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

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

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 $\widetilde{f}(\cdot)$ is *strictly* decreasing in d. Thus, the proposition follows. ■

Proof of Proposition 5

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

Proof of Proposition 7

Proof. Fix a mechanism (ξ, π). The social value $\widetilde{\xi} = \widetilde{E}\left[\xi(d)\right]$ 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

$$\begin{aligned}
V(d', d_1) - V(d, d_1) &= \widetilde{\xi} d_1\left(\xi(d') - \xi(d)\right)\\
&> \widetilde{\xi} d_2\left(\xi(d') - \xi(d)\right)\\
&= V(d', d_2) - V(d, d_2).
\end{aligned}$$

Then, I have the following claim:

Claim 6 *Suppose that $V(\cdot,\cdot)$ is strictly supermodular. Then, the following hold:*

(i) *$\xi(\cdot)$ is incentive compatible if and only if $\xi(\cdot)$ is monotone.*

(ii) *$\xi(\cdot)$ is incentive compatible if the following inequalities hold:*

$$\begin{aligned}
&V(d,d) - \pi(d) \geq V(d-1, d) - \pi(d-1) \textit{ for all } d = 1, \ldots, d_{\max},\\
&V(d,d) - \pi(d) \geq V(d+1, d) - \pi(d+1) \textit{ for all } d = 0, \ldots, d_{\max} - 1.
\end{aligned}$$

Proof. See Chapter 6 in Vohra (2011) for a proof. ■

Claim 7 *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 $\pi(\cdot) : \mathcal{D} \to \mathbb{R}_+$ such that $\pi(0) := 0$ and

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

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

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

Therefore, the claim is proven.

The previous claims imply that to solve the seller's problem (1.6.1), 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 (1.6.1). Thus, given that $\xi(0) = 0$ and $\pi(0) = 0$ without loss of generality, I can fix the payment schedule defined as $\pi(d) := \widetilde{\xi} \sum_{k=1}^{d} (\xi(k)k - \xi(k-1)k)$. This implies that the seller's problem is

$$\underset{\xi:\mathcal{D}\to[0,1]}{\text{maximize}} \quad \left\{ \sum_{d=1}^{d_{\max}} \widetilde{f}(d)\xi(d) \right\} \left[\sum_{d=1}^{d_{\max}} f(d) \left(\xi(d) \Big(d - \frac{1-F(d)}{f(d)} \Big) \right) \right]$$
$$\text{subject to} \quad 0 = \xi(0) \leq \xi(1) \leq \cdots \leq \xi(d_{\max}) \leq 1.$$

Consider two degree distributions $f(\cdot)$ and $f'(\cdot)$ with $f(\cdot) >_{\mathbf{LR}} f'(\cdot)$. 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 $\widetilde{f}(\cdot) >_{\mathbf{FOSD}} \widetilde{f}'(\cdot)$, it also follows that $\sum_{d=1}^{d_{\max}} \widetilde{f}(d)\xi(d) \geq \sum_{d=1}^{d_{\max}} \widetilde{f}'(d)\xi(d)$ 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. $\blacksquare$

Proof of Proposition 7

Proof. The proposition follows by the discussions in the main text. $\blacksquare$

A.2 Additional Proofs

I here provide additional proofs and results in probability theory.

Approximation of Random Variables

Claim 8 *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* $t \to \infty$,

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

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 $t \geq T$. Hence, $t \geq T$ implies that

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

Therefore, the claim is proven. ■

Convergence of Random Variables in $\mathbb{N}^\infty$

Claim 9 *Let* $(\mathbf{X}_n)_{n\geq 1}$ *be a sequence of random variables. Let* $\mathbf{X}_\infty$ *be a random variable distributed over* $\mathbb{N}^\infty$. *Then,* $\mathbf{X}_n$ *converges in distribution to* $\mathbf{X}_\infty$ *as* $n \to \infty$ *if and only if for all* $x \in \mathbb{R}$,

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

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

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

This part is shown in Theorem 29.1 in Billingsley (2012). Therefore, for any $x \notin \mathbb{N}$, there exists (a, b) such that $x \in (a, b)$, $(a, b) \cap \mathbb{N} = \emptyset$, and $\mathbb{P}(\mathbf{X}_\infty = x) = 0 = \lim_{n\to\infty} \mathbb{P}(\mathbf{X}_n = x)$.

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

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

where the second and the last inequalities are by Theorem 29.1 in Billingsley (2012). Therefore, it follows that $\lim_{n\to\infty} \mathbb{P}(\mathbf{X}_n = x) = \mathbb{P}(\mathbf{X}_\infty = x)$.

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

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

Note that the above argument does not hold if the support of $\mathbf{X}_\infty$ is not $\mathbb{N}$.[1] ■

Bounds of the Largest Eigenvalue

Claim 10 *Let $\mathbf{G} = \langle N, A \rangle$ be a network with size n. Its largest eigenvalue satisfies*

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

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

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

for all $\mathbf{q} \in \mathbb{R}^n$. Without loss of generality, suppose that nodes $2, \dots, k$ have, respectively, a link to node 1, but nodes $k + 1, \dots, n$ do not. Choose $\mathbf{q}$ such that

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

[1] A simple counter example is a sequence of integer-valued random variables $(\mathbf{X}_n)_{n\ge1}$ such that $\mathbb{P}(\mathbf{X}_n = k) = 1/n$ if and only if $1 \le k \le n$. Then, $\lim_{n\to\infty} \mathbb{P}(\mathbf{X}_n = x) = 0$ 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 $\mathbf{X}_\infty$ as $n \to \infty$.

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

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

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 so

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

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

$$|\lambda_{\max}(\mathbf{G})||\mathbf{x}_1| \leq \sum_{j\in N_1} |\mathbf{x}_j| \leq d_{\max}(\mathbf{G})|\mathbf{x}_1|. \tag{A.2.1}$$

Therefore, the claim is proven. ■

The Second Borel-Cantelli Lemma

Claim 11 *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

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

Thus, $\mathbb{P}(\cup_{k=n}^{\infty} E_n) = 1$ for all n. Since $\cup_{k=n}^{\infty} E_n$ decreases monotonically to $\limsup_n E_n$ as $n \to \infty$, it follows that $\mathbb{P}(E_n \text{ i.o.}) = 1$. ■

Appendix B

PROOFS OF CHAPTER 2: MONOPOLY PRICING AND DIFFUSION OF (SOCIAL) NETWORK GOODS

Proof of Lemma 2

Proof. The strictly increasing hazard rate property of $\widetilde{f}$: Define a function $\lambda : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\lambda(x) := \frac{f(x)}{1-F(x)}$, which is differentiable and $\lambda'(x) \geq 0$ by the IHRP assumption. Then, I can rewrite f and F as $f(x) = \lambda(x)e^{-\int_0^x \lambda(s)\,ds}$ and $F(x) = 1 - e^{-\int_0^x \lambda(s)\,ds}$. With these new expressions, it follows that

$$\begin{aligned}
\frac{1-\widetilde{F}(x)}{\widetilde{f}(x)} &= \frac{\int_x^\infty sf(s)\,ds}{xf(x)} \\
&= \frac{\Big[s(-1+F(s))\Big]_x^\infty - \int_x^\infty \big(-1+F(s)\big)\,ds}{x\lambda(x)e^{-\int_0^x \lambda(s)\,ds}} \\
&= \frac{xe^{-\int_0^x g(s)\,ds} + \int_x^\infty e^{-\int_0^y \lambda(s)\,ds}\,ds}{x\lambda(x)e^{-\int_0^x \lambda(s)\,ds}} \\
&= \frac{1}{\lambda(x)} + \frac{\int_x^\infty e^{-\int_0^s \lambda(u)\,du}\,ds}{x\lambda(x)}.
\end{aligned}$$

Since $\lambda'(x) \geq 0$, $\big(\int_x^\infty e^{-\int_0^s \lambda(u)\,du}\,ds\big)' < 0$, and $(x\lambda(x))' > 0$, the inverse hazard rate function of $\widetilde{f}$ is strictly decreasing.

The dominance relationship: Observe that for all $x \in \mathbb{R}^+$,

$$\begin{aligned}
\frac{\widetilde{f}(x)}{1-\widetilde{F}(x)} &= \frac{xf(x)}{\int_x^\infty sf(s)\,ds} \\
&= \frac{f(x)}{\int_x^\infty f(s)\,ds + \int_x^\infty \frac{(s-x)}{x}f(s)\,ds} \\
&= \frac{f(x)}{1-F(x) + \int_x^\infty \frac{(s-x)}{x}f(s)\,ds} \\
&< \lambda(x) \quad \text{because } \frac{(s-x)}{x}f(s) > 0 \text{ for all } s > x.
\end{aligned}$$

The single crossing property: By the IHRP assumption, it suffices to show existence of crossing points. A proof for the existence is straightforward by the intermediate value theorem as follows. Define a function $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ as

$\varphi(x) = x(1-F(x))$, which is differentiable, $\varphi(0) = 0$, $\varphi'(0) > 0$, and $\lim_{x\to\infty} \varphi(x) = 0$. Hence, there exists $\overline{x} \in \mathbb{R}^+$ such that $\varphi'(\overline{x}) = 1 - F(\overline{x}) - \overline{x} f(\overline{x}) = 0$, which is the desired single crossing point of $\lambda(x)$ with $\frac{1}{x}$. The same exercise with a function $\widetilde{\varphi} : \mathbb{R}^+ \to \mathbb{R}^+$ defined as $\widetilde{\varphi}(x) = x(1 - \widetilde{F}(x))$ shows existence of single crossing point $\widetilde{x}$. ■

Proof of Proposition 8

Proof. I first show that an optimal diffusion policy exists. Then, assuming differentiability of V, I derive the characterization equation in Proposition 8. Finally, I provide proofs for the discussions about a sufficient condition for the differentiability of V.

Note that $\pi_1(x, y) = y(-\widetilde{f}(x))(1 - F(y)) < 0$ for all $x, y \in \mathbb{R}^+$. By single crossing property (2.2.3), it follows that for all $x \in \mathbb{R}^+$,

$$\pi_2(x, y) = (1 - \widetilde{F}(x))(1 - F(y)) + y(1 - \widetilde{F}(x))(-f(y)) \geq 0 \text{ if and only if } y \leq \overline{x}.$$

The equality holds only at $y = \overline{x}$. Thus, for given $x \in \mathbb{R}^+$, $\pi(x, y)$ is uniquely maximized at $y = \overline{x}$.

The following claim finds that an optimal plan must be contained a compact interval:

Claim 12 *For any $t \geq 1$, it is not optimal to choose $x_t \notin [0, \overline{x}]$.*

Proof. By stationarity of the monopolist's dynamic optimization problem, let $\{x_t\}_{t=1}^{\infty}$ be a diffusion policy such that $x_1 > \overline{x}$, without loss of generality. Then, since $\pi_1(x, y) < 0$ and $\pi_2(x, y) < 0$ for all $y > \overline{x}$ and $x \in \mathbb{R}^+$, I have

$$\begin{aligned}\sum_{t=1}^{\infty} \beta^{t-1}\pi(x_{t-1}, x_t) &= \pi(x_0, x_1) + \beta\pi(x_1, x_2) + \sum_{t=3}^{\infty} \beta^{t-1}\pi(x_{t-1}, x_t) \\ &< \pi(x_0, \overline{x}) + \beta\pi(\overline{x}, x_2) + \sum_{t=3}^{\infty} \beta^{t-1}\pi(x_{t-1}, x_t). \qquad ■\end{aligned}$$

I prove existence of an optimal diffusion policy. By Claim 12, without loss of generality, I define a feasible policy correspondence $\Gamma : [0, \overline{x}] \to [0, \overline{x}]$ such that $\Gamma(x) = [0, \overline{x}]$ for all $x \in [0, \overline{x}]$. Γ satisfies the following properties:

(i) $\Gamma(x)$ is nonempty for all $x \in [0, \overline{x}]$,

(ii) For any feasible policy $\{x_t\}_{t=1}^{\infty}$, $\lim_{n\to\infty} \sum_{t=1}^{n} \beta^{t-1}\pi(x_{t-1}, x_t)$ exists because π is bounded, and $\sum_{t=1}^{n} \beta^{t-1}\pi(x_{t-1}, x_t)$ is strictly increasing in n,

(iii) $[0, \overline{x}]$ is a convex subset of $\mathbb{R}_0^+$, and Γ is compact-valued and continuous, and

(iv) $\pi : [0, \overline{x}] \times [0, \overline{x}] \to [0, \overline{x}]$ is continuous, and $\beta \in (0, 1)$.

Under the above four properties, it can be shown that there exists a unique continuous function $V : [0, \overline{x}] \to \mathbb{R}$ such that $V(x) = \sup\{y \in \Gamma(x) | \pi(x, y) + \beta V(y)\}$ (Stokey and Lucas, 1989). The Maximum theorem provides that an optimal diffusion policy exists.

I now show that for any optimal diffusion policy $\{x_t^*\}_{t=1}^{\infty}$, it must be that $x_t^* \in (0, \overline{x})$. I prove it by using mathematical induction. For $t = 1$, I calculate

$$\frac{\partial}{\partial y}\left(\pi(x_0, y) + \beta\pi(y, x_2^*) + \sum_{t=3}^{\infty} \beta^{t-1}\pi(x_{t-1}^*, x_t^*)\right)$$
$$= (1 - \widetilde{F}(x_0))(1 - F(y)) + y(1 - \widetilde{F}(x_0))(-f(y)) + \beta x_2^*(-\widetilde{f}(y))(1 - F(x_2^*)).$$

The value of the last expression at $y = 0$ is $1 - \widetilde{F}(x_0) > 0$ because $\widetilde{f}(0) = 0$. Hence, $x_1^* > 0$. Suppose now that $x_s^* > 0$ for all $1 \le s \le t$. Then, at $y = 0$,

$$\frac{\partial}{\partial y}\left(\sum_{s=1}^{t} \beta^{s-1}\pi(x_{s-1}^*, x_s^*) + \beta^t\pi(x_t^*, y) + \beta^{t+1}\pi(y, x_{t+1}^*) + \sum_{s=t+3}^{\infty} \beta^{s-1}\pi(x_{s-1}^*, x_s^*)\right)$$
$$= \beta^t(1 - \widetilde{F}(x_t^*)) > 0.$$

This implies that $x_{t+1}^* > 0$. Therefore, $x_t^* > 0$ for all $t \ge 1$.

Similarly, at $y = \overline{x}$, I have that since $x_2^* > 0$,

$$\frac{\partial}{\partial y}\left(\pi(x_0, y) + \beta\pi(y, x_2^*) + \sum_{t=3}^{\infty} \beta^{t-1}\pi(x_{t-1}^*, x_t^*)\right)$$
$$= (1 - \widetilde{F}(x_0))(1 - F(\overline{x})) + y(1 - \widetilde{F}(x_0))(-f(\overline{x})) + \beta x_2^*(-\widetilde{f}(\overline{x}))(1 - F(x_2^*))$$
$$= \beta x_2^*(-\widetilde{f}(\overline{x}))(1 - F(x_2^*)) < 0.$$

Hence, $x_1^* < \overline{x}$. Suppose now that $x_s^* < \overline{x}$ for all $1 \le s \le t$. Then, at $y = \overline{x}$, I have

$$\frac{\partial}{\partial y}\left(\sum_{s=1}^{t} \beta^{s-1}\pi(x_{s-1}^*, x_s^*) + \beta^t\pi(x_t^*, y) + \beta^{t+1}\pi(y, x_{t+1}^*) + \sum_{s=t+3}^{\infty} \beta^{s-1}\pi(x_{s-1}^*, x_s^*)\right)$$
$$= \beta^t\Big((1 - \widetilde{F}(x_t^*))(1 - F(\overline{x})) - \overline{x}(1 - \widetilde{F}(x_t^*))f(\overline{x}) - \beta x_{t+2}^*\widetilde{f}(\overline{x})(1 - F(x_{t+2}^*))\Big)$$
$$= -\beta^{t+1}x_{t+2}^*\widetilde{f}(\overline{x})(1 - F(x_{t+2}^*)) < 0.$$

Therefore $x_t^* < \overline{x}$ for all t.

Suppose that V is differentiable. Then, it follows that $\pi_2(x^*_{t-1}, x^*_t) + \beta V'(x^*_t) = 0$ for all $t \geq 1$ because $x^* \in \mathbb{R}^+$. By plugging $V'(x^*_t) = \pi_1(x^*_t, x^*_{t+1})$, I obtain that for all $t \geq 1$,

$$\pi_2(x^*_{t-1}, x^*_t) + \beta\pi_1(x^*_t, x^*_{t+1}) = 0.$$

I prove discussions about the differentiability of V. I first show that directional derivatives are well-defined at x^*_t for all $t \geq 1$ by checking conditions in the following claim:

Claim 13 (Theorem 3 in Milgrom and Segal (2002)) *Suppose that the family of functions $\{\pi_1(\cdot, y)\}_{y\in\Gamma(x)}$ is equi-differentiable at $\widehat{x} \in [0, \overline{x}]$, that $\sup_{y\in\Gamma(x)} |\pi_1(\widehat{x}, y)| < \infty$, and that $k(x) = \operatorname{argmax}_{y\in\Gamma(x)} \pi(x, y) + \beta V(y)$ is not empty for all $x \in [0, \overline{x}]$. Then, V is left- and right-hand differentiable at $\widehat{x}$. For any selection $y^*(x) \in k(x)$, the directional derivatives are*

$$V'(\widehat{x}+) = \lim_{x\to\widehat{x}+} \pi_1(\widehat{x}, y^*(x)) \text{ for } \widehat{x} < \overline{x},$$
$$V'(\widehat{x}-) = \lim_{x\to\widehat{x}-} \pi_1(\widehat{x}, y^*(x)) \text{ for } \widehat{x} > 0.$$

V is differentiable at $\widehat{x} \in (0, \overline{x})$ if and only if $\pi_1(\widehat{x}, y^*(x))$ is continuous in x at $x = \widehat{x}$.

I check the following three conditions:

(i) $\{\pi_1(\cdot, y)\}_{y\in\Gamma(x)}$ is equi-differentiable at $\widehat{x} \in \Gamma(x)$: It suffices to show that $\{\pi_{11}(\cdot, y)\}_{y\in\Gamma(x)}$ is equi-continuous in x (Milgrom and Segal, 2002). Since $\pi_{11}(\cdot, y) = y(-\widetilde{f}'(\cdot))(1 - F(y))$ and $\widetilde{f}'$ is continuous, $\{\pi_{11}(\cdot, y)\}_{y\in\Gamma(x)}$ is equi-continuous.

(ii) $\sup_{y\in\Gamma(x)} |\pi_1(\widehat{x}, y)| < \infty$: By single crossing property (2.3), for all $y \in \Gamma(x)$,

$$\pi_1(\widehat{x}, y) = y(-\widetilde{f}(\widehat{x}))(1 - F(y)) \leq \overline{x}(-\widetilde{f}(\widehat{x}))(1 - F(\overline{x})) \leq 0.$$

(iii) $k(x) = \operatorname{argmax}\{y \in \Gamma(x) | \pi(x, y) + \beta V(y)\} \neq \emptyset$ for all $x \in [0, \overline{x}]$: This property is proven by the existence of an optimal diffusion policy.

In addition, it follows that $\pi_1(\widehat{x}, y^*(x))$ is continuous in x at $x = \widehat{x}$ if k is continuous. Moreover, k is single-valued if π is concave on $[\underline{x}, \overline{x}]$. This further provides that k is continuous (Theorem 4 in Milgrom and Segal (2002)). Therefore, if π is concave on $[\underline{x}, \overline{x}]$, then there exists a unique optimal diffusion policy $\{x^*_t\}^\infty_{t=1}$ such that $\pi_2(x^*_{t-1}, x^*_t) + \beta\pi_1(x^*_t, x^*_{t+1}) = 0$ for all $t \geq 1$. ■

Proof of Proposition 9

Proof. Since equation (2.3.2) has a unique solution, the proposition is proven. ■

Proof of Proposition 10

Proof. Define two functions $L : [0, \overline{x}] \times [0, \overline{x}] \to \mathbb{R} \cup \{\infty\}$ and $R : [0, \overline{x}] \to [0, \overline{x}]$ as

$$L(x, y) := \begin{cases} (1 - \widetilde{F}(x)) \left(\frac{1 - F(y) - y f(y)}{\widetilde{f}(y)} \right) & \text{if } x > 0, \\ \infty & \text{if } x = 0, \end{cases}$$

$$R(z) := \beta z (1 - F(z)).$$

Observe that $L(x, y)$ is strictly decreasing in x and y, and $R(z)$ is strictly increasing in z.

Let $\{x_t^*\}_{t=1}^{\infty}$ be an optimal diffusion policy. Then, $L(x_{t-1}^*, x_t^*) = R(x_{t+1}^*)$ for all $t \geq 1$. In addition, by the definition of x^*, $L(x^*, x^*) = R(x^*)$. Suppose that $x_{t-1}^* < x^*$ and $x_t^* < x^*$. Then, it follows that

$$R(x_{t+1}^*) = L(x_{t-1}^*, x_t^*) > L(x^*, x^*) = R(x^*).$$

Since $R(z)$ is strictly increasing in z, $x_{t+1}^* > x_t^*$. Similarly, $x_{t-1}^* > x^*$ and $x_t^* > x^*$ imply that $x_{t+1}^* < x_t^*$. Thus, $\{x_t^*\}_{t=1}^{\infty}$ alternates around x^* within length 3.

Let $\{p_t^*\}_{t=1}^{\infty}$ be an optimal pricing policy. In the following, I show that $p_t^* > p^*, p_{t+1}^* > p^*, p_{t+2}^* > p^*$ implies $p_{t+3}^* < p^*$. There are two cases to consider:

Case 1: Suppose $x_{t-1}^* < x^*$. There exists τ_1 with $t \leq \tau_1 \leq t+1$ such that $x_{\tau_1}^* > x^*$ because $\{x_t^*\}_{t=1}^{\infty}$ alternates around x^* within length 3. Suppose $\tau_1 = t+1$. Then, $p_{t+2}^* > p^*$ implies $x_{t+2}^* > x^*$. It follows that $x_{t+3}^* < x^*$ because $\{x_t^*\}_{t=1}^{\infty}$ alternates around x^* within length 3. Therefore, $p_{t+3}^* < p^*$. Suppose $\tau_1 = t$. Then, $p_{t+1}^* > p^*$ and $p_{t+2}^* > p^*$ imply that $x_t^* > x^*$, $x_{t+1}^* > x^*$, and $x_{t+2}^* > x^*$, which contradict that $\{x_t^*\}_{t=1}^{\infty}$ alternates around x^*.

Case 2: Suppose $x_{t-1}^* \geq x^*$. Since $p_t^* > p^*$ and $p_{t+1}^* > p^*$, it follows that $x_t^* > x^*$ and $x_{t+1}^* > x^*$. Now, since $\{x_t^*\}_{t=1}^{\infty}$ alternates around x^* within length 3, it must be that $x_{t+2}^* < x^*$, which implies that $p_{t+2}^* < p^*$.

A similar argument shows that $p_t^* < p^*$, $p_{t+1}^* < p^*$, and $p_{t+2}^* < p^*$ imply $p_{t+3}^* < p^*$. Therefore, $\{p_t^*\}_{t=1}^{\infty}$ alternates around p^* within length 4. ■

Proof of Proposition 11

Proof. Equation (2.3.2) and $\overline{x} < \widetilde{x}$ imply that $x^* < \widetilde{x}$.

To prove diffusion-stability, it suffices to show that $x^* < \widetilde{x}$ implies the diffusion-stability of x^*. Let $\varepsilon = \min\{|x^* - \widetilde{x}|, |x^*|\}/2$. It follows that $0 < g(x_0) < g(\widetilde{x})$ by the construction of ε for any $x_0 \in (x^* - \varepsilon, x^* + \varepsilon)$. Note that $x_1 < x_2$ is equivalent to $g(x_1) < g(x_2)$ for any $x_1, x_2 \in [0, \widetilde{x}]$.

Suppose that $x^* < x_0$. Then, the corresponding diffusion process $\{x_t\}_{t=1}^{\infty}$ is defined as $x_t = \frac{p^*}{1-\widetilde{F}(x_{t-1})}$. This diffusion process is monotone decreasing and bounded below by x^* as follows. First, $x^* < x_1 < x_0$ because

$$x^* < x_0 \Rightarrow x^* = \frac{p^*}{1 - \widetilde{F}(x^*)} < \frac{p^*}{1 - \widetilde{F}(x_0)} = x_1,$$

and

$$g(x^*) < g(x_0) \Rightarrow x_1 = \frac{p^*}{1 - \widetilde{F}(x_0)} = \frac{g(x^*)}{\left(x_0\left(\frac{1-\widetilde{F}(x_0)}{x_0}\right)\right)} = \frac{g(x^*)}{g(x_0)} x_0 < x_0.$$

Second, (ii) $x^* < x_t < x_{t-1}$ implies $x^* < x_{t+1} < x_t$ because

$$x^* < x_t \Rightarrow x^* = \frac{p^*}{1 - \widetilde{F}(x^*)} < \frac{p^*}{1 - \widetilde{F}(x_t)} = x_1,$$

and

$$g(x_t) < g(x_{t-1}) \Rightarrow x_{t+1} = \frac{p^*}{1 - \widetilde{F}(x_t)} = \frac{g(x^*)}{g(x_t)} x_t = \frac{g(x_{t-1})}{g(x_t)} \frac{g(x^*)}{g(x_{t-1})} x_t < x_t.$$

Hence, x_t decreases to x^* as $t \to \infty$. Similarly, for $x_0 < x^*$, it can be shown that the corresponding diffusion process $\{x_t\}_{t=1}^{\infty}$ increases to x^*. ■

Proof of Proposition 12

Proof. The maximum social welfare is $\langle x \rangle$, and the social welfare at the steady state is $\mathbb{E}_f[x(1 - \widetilde{F}(x^*))|x \geq x^*]$. Thus, the deadweight loss from monopoly is

$$\begin{aligned}
&\langle x \rangle - \mathbb{E}_f[x(1 - \widetilde{F}(x^*))|x \geq x^*] \\
&= \int_0^{\infty} x f(x)\, dx - (1 - \widetilde{F}(x^*)) \int_{x^*}^{\infty} x f(x)\, dx \\
&= \int_0^{x^*} x f(x)\, dx + \widetilde{F}(x^*) \int_{x^*}^{\infty} x f(x)\, dx \\
&= \langle x \rangle \int_0^{x^*} \widetilde{f}(x)\, dx + \langle x \rangle \widetilde{F}(x^*) \int_{x^*}^{\infty} \widetilde{f}(x)\, dx \\
&= \langle x \rangle \widetilde{F}(x^*) + \langle x \rangle \widetilde{F}(x^*)(1 - \widetilde{F}(x^*)).
\end{aligned}$$

By dividing the last expression by $\langle x \rangle$, I obtain the deadweight loss *ratio* as

$$\widetilde{F}(x^*) + \widetilde{F}(x^*)(1 - \widetilde{F}(x^*)).$$

Therefore, Proposition 12 is proven. ■

Proof of Proposition 13

Proof. $x^*(\beta)$ is decreasing in β by equation (2.3.2). Since $x^*(\beta) < \widetilde{x}$ and $x^*(\beta)$ is decreasing in β, it follows that $p^*(\beta)$ is decreasing in β.

To show that the $DW(\beta)$ is decreasing in β, note that $F(x)$ is strictly increasing in x. Thus, $F(x^*(\beta))$ is decreasing in β, so that $DW(\beta)$ is decreasing in β. ■

Proof of Proposition 14

Proof. Monotonicity of $x^*(\xi)$ directly follows from Proposition 9, Assumption 3, and equation (2.5.1). To show monotonicity of $p^*(\xi)$, observe that for all $\xi_1 \geq \xi_0$,

$$\begin{aligned}
&p^*(\xi_0) \\
&= x^*(\xi_0)(1 - \widetilde{F}(x^*(\xi_0);\xi_0)) \\
&\leq x^*(\xi_0)(1 - \widetilde{F}(x^*(\xi_0);\xi_1)) \quad \text{because of the first-order stochastic dominance} \\
&\leq x^*(\xi_1)(1 - \widetilde{F}(x^*(\xi_1);\xi_1)) \quad \text{because } g(\cdot;\xi) \text{ increases in } \xi \text{ and } x^*(\xi_1) \geq x^*(\xi_0) \\
&= p^*(\xi_1).
\end{aligned}$$

Therefore, Proposition 14 is proven. ■

Proof of Corollary 2

Proof. Let $\xi_0 \leq \xi_1$. Since $f(\cdot;\xi_1)$ has first-order stochastic dominance over $f(\cdot;\xi_0)$, the per-period profit by charging $\widehat{p} = x^*(\xi_0)(1 - \widetilde{F}(x^*(\xi_0);\xi_1))$ is higher than the per-period profit under $f(\cdot;\xi_0)$ by charging $p^*(\xi_0) = x^*(\xi_0)(1 - \widetilde{F}(x^*(\xi_0);\xi_0))$. Since $p^*(\xi_1) \geq \widehat{p}$, the equilibrium monopoly profit under $f(\cdot;\xi_1)$ is greater than that under $f(\cdot;\xi_0)$. ■

Appendix C

PROOFS OF CHAPTER 3: A MODEL OF PRE-ELECTORAL COALITION FORMATION

Proof of Lemma 4

Proof. By symmetry, I prove the lemma for $j = L1$. I first define $int(\Sigma)$ as

$$int(\Sigma) = \{(\sigma_{L1}/m, \sigma_{L2}/m) \in \Sigma | 0 < \sigma_{L1}, \sigma_{L2} < m \text{ and } \sigma_{L1} + \sigma_{L2} < m\}.$$

To construct γ_{L1}, I state and prove the following lemma:

Lemma 7 *Let $\mathcal{E} = \mathcal{P}$. If $BR(t_{L2}, \sigma, \overline{s}) = L1$ at $\sigma = (\sigma_{L1}/m, \sigma_{L2}/m) \in int(\Sigma)$, then $BR(t_{L2}, \sigma', \overline{s}) = L1$ where $\sigma' = ((\sigma_{L1} + 1)/m, (\sigma_{L2} - 1)/m)$.*

Proof. $BR(t_{L2}, \sigma, \overline{s}) = L1$ at $\sigma \in int(\Sigma)$ if and only if

$$\int_{1/3}^{1/2} f_{L2}(x, x|\sigma)\, dx + \int_{1/3}^{1/2} f_{L2}(1 - 2x, x|\sigma)\, dx < \frac{1}{2} \int_{1/3}^{1/2} f_{L2}(x, 1 - 2x|\sigma)\, dx.$$

Then, since $(1 - 2x)/x \leq 1$ for all $x \in [1/3, 1/2]$, I have

$$\begin{aligned}
&\int_{1/3}^{1/2} f_{L2}(x, x|\sigma')\, dx + \int_{1/3}^{1/2} f_{L2}(1 - 2x, x|\sigma')\, dx \\
&= \mu \int_{1/3}^{1/2} x^{\sigma_{L1}+\sigma_{L2}+1}(1 - 2x)^{\sigma_R} f(x, x)\, dx \\
&\quad + \mu \int_{1/3}^{1/2} (1 - 2x)^{\sigma_{L1}+1} x^{\sigma_{L2}+\sigma_R} f(1 - 2x, x)\, dx \\
&< \mu \int_{1/3}^{1/2} x^{\sigma_{L1}+\sigma_{L2}+1}(1 - 2x)^{\sigma_R} f(x, x)\, dx \\
&\quad + \mu \int_{1/3}^{1/2} (1 - 2x)^{\sigma_{L1}} x^{\sigma_{L2}+\sigma_R+1} f(1 - 2x, x)\, dx \\
&< \frac{\mu}{2} \int_{1/3}^{1/2} x^{\sigma_{L1}+\sigma_R}(1 - 2x)^{\sigma_{L2}+1} f(x, 1 - 2x)\, dx \quad \text{because } BR(t_{L2}, \sigma, \overline{s}) = L1 \\
&< \frac{\mu}{2} \int_{1/3}^{1/2} x^{\sigma_{L1}+\sigma_R+1}(1 - 2x)^{\sigma_{L2}} f(x, 1 - 2x)\, dx \\
&= \frac{1}{2} \int_{1/3}^{1/2} f_{L2}(x, 1 - 2x|\sigma')\, dx,
\end{aligned}$$

where

$$\mu^{-1} = \int_0^1 \int_0^{1-w} z^{\sigma_{L1}} w^{\sigma_{L2}+1}(1 - z - w)^{\sigma_R} f(z, w)\, dz\, dw.$$

Therefore, $BR(t_{L2}, \sigma', \overline{s}) = L1$, and the lemma is proven. ■

Define a function $\widehat{\gamma}_{L1} : \{0, 1/m, \ldots, 1\} \to \{0, 1/m, \ldots, 1\}$ such that

$$\widehat{\gamma}_{L1}\Big(\frac{\sigma_R}{m}\Big) = \sup\Big\{\frac{\sigma'_{L1}}{m}\Big|BR(t_{L2}, \sigma', \overline{s}) = L2, \sigma'_{L1} + \sigma'_{L2} = m - \sigma_R\Big\}.$$

$\widehat{\gamma}_{L1}$ is well-defined by the following lemma.

Lemma 8 *$\{\sigma'_{L1}/m | BR(t_{L2}, \sigma', \overline{s}) = L2, \sigma'_{L1} + \sigma'_{L2} = m - \sigma_R\} \neq \emptyset$ for all m.*

Proof. By direct calculations, one can easily show the statement for $m = 0, 1$. Thus, I prove the statement for $m \geq 2$ by proving that $BR(t_{L2}, \sigma', \overline{s}) = L2$ for all $\sigma = (\sigma_{L1}/m, \sigma_{L2}/m)$ with $\sigma_{L1} = \sigma_{L2}$. Suppose such a σ be given. Let $a = \int_{1/3}^{1/2} f_C(x, x|\sigma)\, dx$, $b = \int_{1/3}^{1/2} f_C(x, 1-2x|\sigma)\, dx$, and $c = \int_{1/3}^{1/2} f_C(1-2x, x|\sigma)\, dx$. Note that $b = c > 0$ because $\sigma_{L1} = \sigma_{L2}$.

Since t_{L1} and t_{L2} are symmetric, it suffices to show that $BR(t_{L1}, \sigma, \overline{s}) = L1$. Suppose, by a way of contradiction, that $BR(t_{L1}, \sigma, \overline{s}) = L2$. Then, it follows that

$$\begin{aligned}
&\frac{1}{3}(a+b) \\
&= \nu \int_{1/3}^{1/2} \frac{1}{3} x^{\sigma_{L1}+\sigma_{L2}} (1-2x)^{\sigma_R} f(x,x)\, dx \\
&\quad + \nu \int_{1/3}^{1/2} \frac{1}{3} x^{\sigma_{L1}+\sigma_R} (1-2x)^{\sigma_{L2}} f(x, 1-2x)\, dx \\
&< \nu \int_{1/3}^{1/2} x^{\sigma_{L1}+\sigma_{L2}+1} (1-2x)^{\sigma_R} f(x,x)\, dx \\
&\quad + \int_{1/3}^{1/2} x^{\sigma_{L1}+\sigma_R+1} (1-2x)^{\sigma_{L2}} f(x, 1-2x)\, dx \\
&< \frac{\nu}{2} \int_{1/3}^{1/2} (1-2x)^{\sigma_{L1}+1} x^{\sigma_{L2}+\sigma_R} f(1-2x, x)\, dx \quad \text{because } BR(t_{L1}, \sigma, \overline{s}) = L2 \\
&< \frac{\nu}{2} \int_{1/3}^{1/2} \frac{1}{3} (1-2x)^{\sigma_{L1}} x^{\sigma_{L2}+\sigma_R} f(1-2x, x)\, dx \\
&= \frac{1}{6} c,
\end{aligned}$$

where

$$\nu^{-1} = \int_0^1 \int_0^{1-w} z^{\sigma_{L1}+1} w^{\sigma_{L2}} (1-z-w)^{\sigma_R} f(z,w)\, dz\, dw.$$

Since $b = c$, $a + b < 2c$ implies $a < 0$, which is a contradiction. Thus, $BR(t_{L1}, \sigma, \overline{s}) = L1$. ■

I finally define $\gamma_{L1} : [0,1] \to [0,1]$ such that $\gamma_{L1}(z) = \widehat{\gamma}_{L1}(\lfloor mz \rfloor / m)$. By its construction, γ_{L1} satisfies the desired property. ■

Proof of Proposition 15

Proof. For each $j \in \{L1, L2\}$, I define $\Sigma_j^{\mathcal{P}} = \{\sigma \in \Sigma | \sigma_j / m > \gamma_j(\sigma_R / m)\}$ and $\Sigma_S^{\mathcal{P}} = \Sigma \setminus \left(\Sigma_{L1}^{\mathcal{P}} \cup \Sigma_{L2}^{\mathcal{P}}\right)$.

Mutual exclusiveness: For each $j \in \{L1, L2\}$, $\Sigma_j^{\mathcal{P}} \cap \Sigma_S^{\mathcal{P}} = \emptyset$ by the definition of $\Sigma_S^{\mathcal{P}}$. I now show that $\Sigma_{L1}^{\mathcal{P}} \cap \Sigma_{L2}^{\mathcal{P}} \neq \emptyset$. Suppose, by a way of contradiction, that there exists $\sigma \in \Sigma_{L1}^{\mathcal{P}} \cap \Sigma_{L2}^{\mathcal{P}}$. For each $j \in \{L1, L2\}$, let $a_j = \int_{1/3}^{1/2} f_j^2(x, x|\sigma)\, dx$, $b_j = \int_{1/3}^{1/2} f_j^2(x, 1-2x|\sigma)\, dx$, and $c_j = \int_{1/3}^{1/2} f_j^2(1-2x, x|\sigma)\, dx$. Since $x \geq 1 - 2x$ for all $x \in [1/3, 1/2]$, I have $b_{L1} - b_{L2} > 0$ and $c_{L2} - c_{L1} > 0$.

Note that $\sigma \in \Sigma_{L1}^{\mathcal{P}} \cap \Sigma_{L2}^{\mathcal{P}}$ implies that $a_{L1} + b_{L1} < c_{L1}/2$ and $a_{L2} + c_{L2} < b_{L2}/2$. These two inequalities imply that

$$(a_{L1} + a_{L2}) + \left(b_{L1} - \frac{b_{L2}}{2}\right) + \left(c_{L2} - \frac{c_{L1}}{2}\right) < 0,$$

which is impossible because $b_{L1} - b_{L2} > 0$ and $c_{L2} - c_{L1} > 0$. Thus, $\Sigma_{L1}^{\mathcal{P}} \cap \Sigma_{L2}^{\mathcal{P}} = \emptyset$.

Bayesian equilibrium: By the previous results, it follows that for all $j \in \{L1, L2\}$,

$$BR(t_j, \sigma, \overline{s}) = \begin{cases} j & \text{if } \sigma \in \Sigma_S^{\mathcal{P}}, \\ L1 & \text{if } \sigma \in \Sigma_{L1}^{\mathcal{P}}, \\ L2 & \text{if } \sigma \in \Sigma_{L2}^{\mathcal{P}}. \end{cases}$$

Define $s^* : T \times \Sigma \to C$ such that

$$s^*(t_j, \sigma) = \begin{cases} j & \text{if } \sigma \in \Sigma_S^{\mathcal{P}}, \\ L1 & \text{if } \sigma \in \Sigma_{L1}^{\mathcal{P}}, \\ L2 & \text{if } \sigma \in \Sigma_{L2}^{\mathcal{P}}. \end{cases}$$

Then, it suffices to show that $BR(t_j, \sigma, s^*) = BR(t_j, \sigma, \overline{s})$ all $j \in \{L1, L2\}$ and all $\sigma \in \Sigma$.

I first observe that by the definition of $\overline{s}$, $BR(t_j, \sigma, s^*) = BR(t_j, \sigma, \overline{s}) = j$ for all $\sigma \in \Sigma_S^{\mathcal{P}}$. To prove $BR(t_j, \sigma, s^*) = L1$ for $\sigma \in \Sigma_{L1}^{\mathcal{P}}$, among $n-1$ voters, let a be the number of votes for $L1$, b be the number of votes for $L2$, and c be the number of votes for R. Then, $\sigma \in \Sigma_{L1}^{\mathcal{P}}$ implies that $b = 0$ and

$$\lim_{n\to\infty} n \left\{ \left[\frac{\Pr[a = c+1|\sigma, n] + \Pr[a = c|\sigma, n]}{2} \right] - \left[\frac{\Pr[c = 0|\sigma, n]}{2} \right] \right\} > 0.$$

Thus, $BR(t_j, \sigma, s^*) = L1$ for all $j \in \{L1, L2\}$ and all $\sigma \in \Sigma_{L1}^{\mathcal{P}}$. Moreover, by symmetry, $BR(t_j, \sigma, s^*) = L2$ for all $j \in \{L1, L2\}$ and all $\sigma \in \Sigma_{L2}^{\mathcal{P}}$. ■

Proof of Proposition 16

Proof. By setting $\overline{V}^{\mathcal{P}}(\sigma) = \infty$ for all $\sigma \in \Sigma_{L1}^{\mathcal{P}} \cup \Sigma_{L2}^{\mathcal{P}}$, it suffices to set $\overline{V}^{\mathcal{P}}(\sigma)$ for $\sigma \in \Sigma_S^{\mathcal{P}}$.

For a given $\sigma \in \Sigma_S^{\mathcal{P}}$, a Nash bargaining solution exists if and only if $G^{\mathcal{P}}(\sigma) \cap Q^{\mathcal{P}}(\sigma) \neq \emptyset$ where $Q^{\mathcal{P}}(\sigma) = \{(x, y) \in \mathbb{R}^2 | (x, y) \geq (\overline{u}_{L1}^{\mathcal{P}}(\sigma), \overline{u}_{L2}^{\mathcal{P}}(\sigma))\}$.

If $\Phi_{L1}^{\mathcal{P}}(\sigma) \geq \Phi_{L2}^{\mathcal{P}}(\sigma)$, then $\overline{u}_{L1}^{\mathcal{P}}(\sigma) \geq \overline{u}_{L2}^{\mathcal{P}}(\sigma)$. Hence, $G^{\mathcal{P}}(\sigma) \cap Q^{\mathcal{P}}(\sigma) \neq \emptyset$ if and only if $\Phi_{\xi}^{\mathcal{P}}(\sigma) \geq \overline{u}_{L2}^{\mathcal{P}}(\sigma)$, which is equivalent to

$$V \leq \widetilde{V}_2^{\mathcal{P}}(\sigma) = \frac{\Phi_{\xi}^{\mathcal{P}}(\sigma) - \Phi_{L2}^{\mathcal{P}}(\sigma) - \Phi_{L1}^{\mathcal{P}}(\sigma)/2}{\Phi_{L2}^{\mathcal{P}}(\sigma)}.$$

$\widetilde{V}_2^{\mathcal{P}}(\sigma)$ is well-defined and strictly positive because both the numerator and the denominator are strictly positive. Similarly, if $\Phi_{L2}^{\mathcal{P}}(\sigma) \geq \Phi_{L1}^{\mathcal{P}}(\sigma)$, then $G^{\mathcal{P}}(\sigma) \cap Q^{\mathcal{P}}(\sigma) \neq \emptyset$ if and only if

$$V \leq \widetilde{V}_1^{\mathcal{P}}(\sigma) = \frac{\Phi_{\xi}^{\mathcal{P}}(\sigma) - \Phi_{L1}^{\mathcal{P}}(\sigma) - \Phi_{L2}^{\mathcal{P}}(\sigma)/2}{\Phi_{L1}^{\mathcal{P}}(\sigma)}.$$

Therefore, $G^{\mathcal{P}}(\sigma) \cap Q^{\mathcal{P}}(\sigma) \neq \emptyset$ if and only if $\overline{V}^{\mathcal{P}}(\sigma) = \min\{\widetilde{V}_1^{\mathcal{P}}(\sigma), \widehat{V}_2^{\mathcal{P}}(\sigma)\}$. ■

Proof of Lemma 5

Proof. The proof is based on several results in the supplementary appendix for Hummel (2012). I here assume that $n = 2k$, without loss of generality. I use the following notations:

- $\Pr[a, b, c | \sigma, k]$: the conditional probability that among $2k - 1$ voters, there are a number of type t_{L1} voters, b number of type t_{L2} voters, and c number of type t_R voters at σ;
- $\alpha = a/(2k - 1)$, $\beta = b/(2k - 1)$;
- $F = \sup_{p \in \triangle^2} f_j(p|\sigma)$;
- $D = \sup_{p \in \triangle^2} \left\{ \max \left\{ \frac{\partial f_j(p|\sigma)}{\partial p_{L1}}, \frac{\partial f_j(p|\sigma)}{\partial p_{L2}} \right\} \right\}$.

With the above notations, I state the following claim, which appears in Lemma 3 in the supplementary appendix for Hummel (2012).

Claim 14 *For any $\varepsilon > 0$,*

$$\frac{f_j(\alpha,\beta|\sigma) - 2D\varepsilon}{2k(2k+1)} - \frac{Fe^{-2(2k-1)\varepsilon^2}}{2} \leq \Pr[a,b,c|\sigma,k]$$
$$\leq \frac{f_j(\alpha,\beta|\sigma) - 2D\varepsilon}{2k(2k+1)} + \frac{Fe^{-2(2k-1)\varepsilon^2}}{2}.$$

With this claim, I have the following two lemmas.[1]

Lemma 9 $\lim_{k\to\infty} k \Pr[a = c = k, b = 0|\sigma,k] = \lim_{k\to\infty} k \Pr[a = k-1, b = 1, c = k|\sigma,k] = 0.$

Proof. By Claim 14, it follows that for any $\varepsilon > 0$,

$$\frac{f_j(\frac{1}{2},0|\sigma) - 2D\varepsilon}{2k(2k+1)} - \frac{Fe^{-2(2k-1)\varepsilon^2}}{2} \leq \Pr[a = c = k, b = 0|\sigma,k]$$
$$\leq \frac{f_j(\frac{1}{2},0|\sigma) - 2D\varepsilon}{2k(2k+1)} + \frac{Fe^{-2(2k-1)\varepsilon^2}}{2}.$$

Consider $\varepsilon(k) = k^{1/3}$. Then, $\lim_{k\to\infty} k\frac{D\varepsilon(k)}{2k(2k+1)} = \lim_{k\to\infty} k\frac{Fe^{-2(2k-1)\varepsilon(k)^2}}{2} = 0$. Thus, it follows that

$$\lim_{k\to\infty} \left| k\frac{f_j(\frac{1}{2},0|\sigma)}{2k(2k+1)} - k\Pr[a = c = k, b = 0|\sigma,k]\right| = 0.$$

Since $f_j(1/2,0|\sigma) < \infty$, $\lim_{k\to\infty} k\Pr[a = c = k, b = 0|\sigma,k] = 0$. I omit the proof for $\lim_{k\to\infty} k\Pr[a = k-1, b = 1, c = k|\sigma,k] = 0$ because it is basically identical to the current proof. ■

Lemma 10 $\lim_{k\to\infty} k\Pr[2k/3 \leq a = b < c \leq k|\sigma,k] = \frac{1}{2}\int_{1/4}^{1/3} f_j(x,x|\sigma)\,dx.$

Proof. Observe that $\lim_{k\to\infty} k\frac{D\varepsilon(k)}{2k(2k+1)} = 0$ and $\lim_{k\to\infty} k\frac{Fe^{-2(2k-1)\varepsilon(k)^2}}{2} = 0$. Thus, I have

$$\lim_{k\to\infty} k\Pr\left[\frac{2k}{3} \leq a = b < c \leq k\middle|\sigma,k\right] = \lim_{k\to\infty} k\sum_{l=\lfloor\frac{2k}{3}\rfloor}^{k} \Pr[a = b = l|\sigma,k]$$
$$= \lim_{k\to\infty} k\sum_{l=\lfloor\frac{2k}{3}\rfloor}^{k} \frac{f_j(\frac{l}{2k-1},\frac{l}{2k-1}|\sigma)}{2k(2k+1)} = \lim_{k\to\infty}\frac{1}{2}\sum_{l=\lfloor\frac{2k}{3}\rfloor}^{k}\frac{f_j(\frac{l}{2k-1},\frac{l}{2k-1}|\sigma)}{2k+1}$$
$$= \frac{1}{2}\int_{1/4}^{1/3} f_j(x,x|\sigma)\,dx.$$

[1]The proofs are based on the proof of Lemma 4 in the supplementary appendix for Hummel (2012).

Thus, the lemma is proven. ■

Lemma 9 and Lemma 10 imply that for each type t_j voter, the marginal utility of voting for j is strictly positive for sufficiently large n. Therefore, Lemma 5 follows. ■

Proof of Proposition 17

Proof. The proposition directly follows from Lemma 5 and the second part of the proof of Proposition 15. ■

Proof of Proposition 18

Proof. I omit the proof of the first part because it is essentially a replication of the proof of Proposition 16.

To prove the second part, without loss of generality, suppose that $\Phi^{\mathcal{R}}_{L1}(\sigma) \geq \Phi^{\mathcal{R}}_{L2}(\sigma)$. Then, since $\Phi^{\mathcal{R}}_{j}(\sigma) > \Phi^{\mathcal{P}}_{j}(\sigma)$ for all $j \in \{L1, L2\}$ and all $\sigma \in \Sigma^{\mathcal{P}}_{S}$, it follows that

$$\begin{aligned}\overline{V}^{\mathcal{P}}(\sigma) &= \frac{\Phi^{\mathcal{P}}_{\xi}(\sigma) - \left(\Phi^{\mathcal{P}}_{L2}(\sigma) + \Phi^{\mathcal{P}}_{L1}(\sigma)/2\right)}{\Phi^{\mathcal{P}}_{L2}(\sigma)} \\ &> \frac{\Phi^{\mathcal{R}}_{\xi}(\sigma) - \left(\Phi^{\mathcal{R}}_{L2}(\sigma) + \Phi^{\mathcal{R}}_{L1}(\sigma)/2\right)}{\Phi^{\mathcal{R}}_{L2}(\sigma)} = \overline{V}^{\mathcal{R}}(\sigma).\end{aligned}$$

Therefore, the proposition is proven. ■

Proof of Proposition 19

Proof. Since $W^{\mathcal{E}}_{L1}(V) = W^{\mathcal{E}}_{L2}(V)$ for all V, $W^{\mathcal{E}}_{R}(V) + 2W^{\mathcal{E}}_{L1}(V)$ is constant for all $\mathcal{E} = \mathcal{P}, \mathcal{R}$. Thus, it suffices to show that (i) $W^{\mathcal{P}}_{L1}(0) = W^{\mathcal{R}}_{L1}(0)$, (ii) $W^{\mathcal{R}}_{L1}(V)$ is constant, and (iii) $W^{\mathcal{P}}_{L1}(V)$ is decreasing in V.

To begin, I introduce several notations for proofs. Let $\Sigma^{\mathcal{E}}_{CO}(V) = \{\sigma \in \Sigma^{\mathcal{E}}_{S} | V \leq \overline{V}^{\mathcal{E}}(\sigma)\}$ and $\Sigma^{\mathcal{E}}_{NC}(V) = \Sigma^{\mathcal{E}}_{S} \setminus \Sigma^{\mathcal{E}}_{CO}(V)$. Note that $\Sigma^{\mathcal{E}}_{NC}(V)$ is increasing in V in the sense of set inclusion. Let $\xi(k(\sigma), \lambda(\sigma))$ be the Nash Bargaining solution at σ (if it exists). For each $j \in \{L1, L2\}$, $\Omega^{\mathcal{E}}_{j}$ is the subset of $\triangle^2$ where j wins by running the race alone. $\Omega^{\mathcal{E}}_{\xi} = \{(x, y) \in \triangle^2 | x + y \geq 1/2\}$ is the set of events where the left candidates can win the election by forming a PEC. Note that the ex-post utility of type t_{L1} voters is positive if and only if $p \in \Omega^{\mathcal{E}}_{\xi}$: conditional on the fact that $p \in \Omega^{\mathcal{E}}_{\xi}$, a type t_{L1} voter's utility is 1 if $\sigma \in \Sigma^{\mathcal{E}}_{L1}$; 1/2 if $\sigma \in \Sigma^{\mathcal{E}}_{L2}$; $(1 + \lambda(\sigma))/2$ if $\sigma \in \Sigma^{\mathcal{E}}_{CO}(V)$; and 0 if $\sigma \in \Sigma^{\mathcal{E}}_{NC}(V)$. Let $\Pr[\sigma | x, y]$ be the conditional probability that σ is observed conditional on that $p = (x, y)$:

$$\Pr[\sigma|x, y] = \binom{m}{\sigma_{L1}, \sigma_{L2}, \sigma_R} x^{\sigma_{L1}} y^{\sigma_{L2}} (1 - x - y)^{\sigma_R}.$$

Then, since $f(x, y) = f(y, x)$, it follows that

$$\int_{(x,y)\in\Omega_\xi^{\mathcal{E}}} \left(\sum_{\sigma\in\Sigma_{L1}^{\mathcal{E}}} \Pr[\sigma|x, y] \right) f(x, y)\, dA = \int_{(x,y)\in\Omega_\xi^{\mathcal{E}}} \left(\sum_{\sigma\in\Sigma_{L2}^{\mathcal{E}}} \Pr[\sigma|x, y] \right) f(x, y)\, dA,$$

and

$$\begin{aligned}
&\int_{(x,y)\in\Omega_\xi^{\mathcal{E}}} \left(\sum_{\sigma\in\Sigma_{CO}^{\mathcal{E}}} \frac{1 + \lambda(\sigma)}{2} \Pr[\sigma|x, y] \right) f(x, y)\, dA \\
&\quad = \int_{(x,y)\in\Omega_\xi^{\mathcal{E}}} \left(\sum_{\sigma\in\Sigma_{CO}^{\mathcal{E}}} \frac{3}{4} \Pr[\sigma|x, y] \right) f(x, y)\, dA.
\end{aligned}$$

To show (i), recall that $\Sigma_{CO}^{\mathcal{E}}(0) = \Sigma_S^{\mathcal{E}}$. Thus, I have

$$\begin{aligned}
&W_{L1}^{\mathcal{E}}(0) \\
&= \int_{(x,y)\in\Omega_\xi^{\mathcal{E}}} \left(\sum_{\sigma\in\Sigma_{L1}^{\mathcal{E}}} \Pr[\sigma|x, y] + \sum_{\sigma\in\Sigma_{L2}^{\mathcal{E}}} \frac{1}{2} \Pr[\sigma|x, y] \right) f(x, y)\, dA \\
&\quad + \int_{(x,y)\in\Omega_\xi^{\mathcal{E}}} \frac{1 + \lambda(\sigma)}{2} \Pr[\sigma|x, y] f(x, y)\, dA \\
&= \int_{(x,y)\in\Omega_\xi^{\mathcal{E}}} \left(\sum_{\sigma\in\Sigma_{L1}^{\mathcal{E}}\cup\Sigma_{L2}^{\mathcal{E}}} \frac{3}{4} \Pr[\sigma|x, y] + \sum_{\sigma\in\Sigma_S^{\mathcal{E}}} \frac{3}{4} \Pr[\sigma|x, y] \right) f(x, y)\, dA \\
&= \frac{3}{4} \int_{(x,y)\in\Omega_\xi^{\mathcal{E}}} f(x, y)\, dA.
\end{aligned}$$

Hence, it follows that $W_{L1}^{\mathcal{P}}(0) = W_{L1}^{\mathcal{R}}(0)$ because $\Omega_\xi^{\mathcal{P}} = \Omega_\xi^{\mathcal{R}}$.

To prove (ii), observe that

$$\begin{aligned}
W_{L1}^{\mathcal{R}}(V) &= \int_{(x,y)\in\Omega_\xi^{\mathcal{R}}} \left(\sum_{\sigma\in\Sigma_{L1}^{\mathcal{R}}\cup\Sigma_{L2}^{\mathcal{R}}} \frac{3}{4} \Pr[\sigma|x, y] + \sum_{\sigma\in\Sigma_{CO}^{\mathcal{R}}} \frac{3}{4} \Pr[\sigma|x, y] \right) f(x, y)\, dA \\
&+ \underbrace{\int_{(x,y)\in\Omega_{L1}^{\mathcal{R}}} \sum_{\sigma\in\Sigma_{NC}^{\mathcal{R}}} \Pr[\sigma|x, y] f(x, y)\, dA + \int_{(x,y)\in\Omega_{L2}^{\mathcal{R}}} \sum_{\sigma\in\Sigma_{NC}^{\mathcal{R}}} \frac{1}{2} \Pr[\sigma|x, y] f(x, y)\, dA}_{= \int_{(x,y)\in\Omega_\xi^{\mathcal{R}}} \sum_{\sigma\in\Sigma_{NC}^{\mathcal{R}}} \frac{3}{4} \Pr[\sigma|x,y] f(x,y)\, dA \text{ because } \Omega_\xi^{\mathcal{R}} = \Omega_{L1}^{\mathcal{R}} \cup \Omega_{L2}^{\mathcal{R}} \text{ and } f(x,y) = f(y,x)}
\end{aligned}$$

$$= \frac{3}{4}\int_{(x,y)\in\Omega_{\xi}^{\mathcal{R}}} f(x,y)\,dA.$$

Hence, $W_{L1}^{\mathcal{R}}(V)$ is constant.

To show (iii), note that

$$W_{L1}^{\mathcal{P}}(V) = \int_{(x,y)\in\Omega_{\xi}^{\mathcal{P}}}\left(\sum_{\sigma\in\Sigma_{L1}^{\mathcal{P}}\cup\Sigma_{L2}^{\mathcal{P}}}\frac{3}{4}\Pr[\sigma|x,y] + \sum_{\sigma\in\Sigma_{CO}^{\mathcal{P}}}\frac{3}{4}\Pr[\sigma|x,y]\right)f(x,y)\,dA$$
$$+ \int_{(x,y)\in\Omega_{L1}^{\mathcal{P}}}\sum_{\sigma\in\Sigma_{NC}^{\mathcal{P}}}\Pr[\sigma|x,y]f(x,y)\,dA + \int_{(x,y)\in\Omega_{L2}^{\mathcal{P}}}\sum_{\sigma\in\Sigma_{NC}^{\mathcal{P}}}\frac{1}{2}\Pr[\sigma|x,y]f(x,y)\,dA$$
$$= \int_{(x,y)\in\Omega_{\xi}^{\mathcal{P}}}\left(\sum_{\sigma\in\Sigma_{L1}^{\mathcal{P}}\cup\Sigma_{L2}^{\mathcal{P}}}\frac{3}{4}\Pr[\sigma|x,y] + \sum_{\sigma\in\Sigma_{CO}^{\mathcal{P}}}\frac{3}{4}\Pr[\sigma|x,y]\right)f(x,y)\,dA$$
$$+ \int_{(x,y)\in\Omega_{L1}^{\mathcal{P}}}\sum_{\sigma\in\Sigma_{NC}^{\mathcal{P}}}\frac{3}{4}\Pr[\sigma|x,y]f(x,y)\,dA$$
$$+ \int_{(x,y)\in\Omega_{L2}^{\mathcal{P}}}\sum_{\sigma\in\Sigma_{NC}^{\mathcal{P}}}\frac{3}{4}\Pr[\sigma|x,y]f(x,y)\,dA$$
$$= \frac{3}{4}\left(\int_{(x,y)\in\Omega_{\xi}^{\mathcal{P}}} f(x,y)\,dA - \int_{(x,y)\in\Omega_{\xi}^{\mathcal{P}}\setminus\left(\Omega_{L1}^{\mathcal{P}}\cup\Omega_{L2}^{\mathcal{P}}\right)}\left(\sum_{\sigma\in\Sigma_{NC}^{\mathcal{P}}}\Pr[\sigma|x,y]\right)f(x,y)\,dA\right).$$

Thus, $W_{L1}^{\mathcal{P}}(V)$ is decreasing in V because $\Sigma_{NC}^{\mathcal{P}}(V)$ is increasing in V. ■

Proof of Proposition 20

Proof. Let $\Sigma_{CO}^{\mathcal{P}}(m) = \{\sigma_m \in \Sigma_S^{\mathcal{P}}(m)|V \le \overline{V}^{\mathcal{P}}(\sigma, m)\}$ and $\Sigma_{NC}^{\mathcal{P}}(m) = \Sigma_S^{\mathcal{P}} \setminus \Sigma_{CO}^{\mathcal{P}}(m)$. To economize notation, let $\Omega^* = \Omega_{\xi}^{\mathcal{P}} \setminus \left(\Omega_{L1}^{\mathcal{P}} \cup \Omega_{L2}^{\mathcal{P}}\right)$. Then, it suffices to show that

$$\lim_{m\to\infty}\Pr\left[\left\{\sigma \in \Sigma_{NC}^{\mathcal{P}}(m)\right\}\middle|\left\{p \in \Omega^*\right\}\right] = 0,$$

which is equivalent to prove that

$$\lim_{m\to\infty}\Pr\left[\left\{\sigma_m \in \left\{\sigma'_m \in \Sigma(m)|V \le \overline{V}^{\mathcal{P}}(\sigma'_m)\right\}\right\}\middle|\left\{p \in \Omega^*\right\}\right] = 1, \tag{C.0.1}$$

where $\overline{V}^{\mathcal{P}}(\sigma'_m)$ is the minimum of

$$\frac{\Phi_{\xi}^{\mathcal{P}}(\sigma'_m) - \left(\Phi_{L1}^{\mathcal{P}}(\sigma'_m) + \Phi_{L2}^{\mathcal{P}}(\sigma'_m)/2\right)}{\Phi_{L1}^{\mathcal{P}}(\sigma'_m)}$$

and

$$\frac{\Phi_{\xi}^{\mathcal{P}}(\sigma'_m) - \left(\Phi_{L2}^{\mathcal{P}}(\sigma'_m) + \Phi_{L1}^{\mathcal{P}}(\sigma'_m)/2\right)}{\Phi_{L2}^{\mathcal{P}}(\sigma'_m)}.$$

To prove (C.0.1), I first observe that since f is symmetric,

$$\lim_{m\to\infty} \Pr\left[\left\{p \in \Omega^*\right\}\middle|\left\{\sigma_m \in \Omega^*\right\}\right] = \lim_{m\to\infty} \frac{\Pr\left[\left\{\sigma_m \in \Omega^*\right\}\middle|\left\{p \in \Omega^*\right\}\right] \cdot \Pr\left[\left\{p \in \Omega^*\right\}\right]}{\Pr\left[\left\{\sigma_m \in \Omega^*\right\}\right]}$$
$$= \lim_{m\to\infty} \Pr\left[\left\{\sigma_m \in \Omega^*\right\}\middle|\left\{p \in \Omega^*\right\}\right] = 1.$$

Thus, it follows that

$$\lim_{m\to\infty} \Pr\left[\left\{\sigma_m \in \left\{\sigma'_m \in \Sigma(m) | V \le \overline{V}^{\mathcal{P}}(\sigma'_m)\right\}\right\}\middle|\left\{\sigma_m \in \Omega^*\right\}\right] = 1.$$

Similarly, I also have that

(i) $\lim_{m\to\infty} \Pr\left[\left\{\sigma_m \in \left\{\sigma'_m \in \Sigma(m) | V \le \overline{V}^{\mathcal{P}}(\sigma'_m)\right\}\right\}\middle|\left\{\sigma_m \notin \Omega^*\right\}\right] = 0;$

(ii) $\lim_{m\to\infty} \Pr\left[\left\{\sigma_m \in \Omega^*\right\}\middle|\left\{p \in \Omega^*\right\}\right] = 1;$

(iii) $\lim_{m\to\infty} \Pr\left[\left\{\sigma_m \notin \Omega^*\right\}\middle|\left\{p \in \Omega^*\right\}\right] = 0.$

Thus, since the inside of the limit expression in (C.0.1) is decomposed as

$$\Pr\left[\left\{\sigma_m \in \left\{\sigma'_m \in \Sigma(m) | V \le \overline{V}^{\mathcal{P}}(\sigma'_m)\right\}\right\}\middle|\left\{p \in \Omega^*\right\}\right]$$
$$= \Pr\left[\left\{\sigma_m \in \left\{\sigma'_m \in \Sigma(m) | V \le \overline{V}^{\mathcal{P}}(\sigma'_m)\right\}\right\}\middle|\left\{\sigma_m \in \Omega^*\right\}\right] \Pr\left[\left\{\sigma_m \in \Omega^*\right\}\middle|\left\{p \in \Omega^*\right\}\right]$$
$$+ \Pr\left[\left\{\sigma_m \in \left\{\sigma'_m \in \Sigma(m) | V \le \overline{V}^{\mathcal{P}}(\sigma'_m)\right\}\right\}\middle|\left\{\sigma_m \notin \Omega^*\right\}\right] \Pr\left[\left\{\sigma_m \notin \Omega^*\right\}\middle|\left\{p \in \Omega^*\right\}\right],$$

it follows that

$$\lim_{m\to\infty} \Pr\left[\left\{\sigma_m \in \left\{\sigma'_m \in \Sigma(m) | V \le \overline{V}^{\mathcal{P}}(\sigma'_m)\right\}\right\}\middle|\left\{p \in \Omega^*\right\}\right] = 1.$$

Therefore, the proposition is proven. ■

Proof of Proposition 21

Proof. Let $\Phi_j^{\zeta}(\sigma)$ be the probability that candidate j wins the election alone at σ when the election rule is the two-round runoff rule with threshold ζ. Note that if $\zeta = 1/3$, $\Phi_j^{\zeta}(\sigma)$ represents the corresponding probability under plurality rule; if $\zeta = 1/2$, then it represents the corresponding probability under the two-round runoff rule. Therefore, it suffices to show that $\overline{V}^{\zeta}(\sigma)$ is strictly decreasing in ζ assuming that a type t_j voter votes for j for all $j \in \{L1, L2\}$ at all $\sigma \in \Sigma$.

Without loss of generality, suppose that $\Phi_{L1}^{\zeta}(\sigma) \ge \Phi_{L2}^{\zeta}(\sigma)$. Let $\zeta_1 > \zeta_0$. Then, since $\Phi_j^{\zeta_1}(\sigma) > \Phi_j^{\zeta_0}(\sigma)$, it follows that

$$\overline{V}^{\zeta_1}(\sigma) = \frac{\Phi_{\xi}^{\zeta_1}(\sigma) - \left(\Phi_{L2}^{\zeta_1}(\sigma) + \Phi_{L1}^{\zeta_1}(\sigma)/2\right)}{\Phi_{L2}^{\zeta_1}(\sigma)}$$

$$< \frac{\Phi^{\zeta_1}_{\xi}(\sigma) - \left(\Phi^{\zeta_0}_{L2}(\sigma) + \Phi^{\zeta_0}_{L1}(\sigma)/2\right)}{\Phi^{\zeta_1}_{L2}(\sigma)} = \overline{V}^{\zeta_0}(\sigma).$$

Therefore, the proposition is proven. ■

Proof of Proposition 22

Proof. Let $\Sigma^{\mathcal{P}}_{S}(\theta)$ be the set of poll results such that the inequalities in Lemma 6 are satisfied. Since $\Sigma^{\mathcal{P}}_{S}(\theta)$ is decreasing in θ, in the sense of set inclusion, it suffices to show that for each $\mathcal{E}$, $\overline{V}^{\mathcal{E}}(\sigma, \theta)$ is increasing in θ, assuming that voters vote for their most preferred candidates.

Without loss of generality, suppose that $\Phi^{\mathcal{E}}_{L1}(\sigma) \geq \Phi^{\mathcal{E}}_{L1}(\sigma)$. Then, $\theta_1 > \theta_0$ implies that

$$\begin{aligned}
\overline{V}^{\mathcal{E}}(\sigma, \theta_1) &= \frac{\Phi^{\mathcal{E}}_{\xi}(\sigma) - \left(\Phi^{\mathcal{E}}_{L2}(\sigma) + (1-\theta_1)\Phi^{\mathcal{E}}_{L1}(\sigma)\right)}{\Phi^{\mathcal{E}}_{L2}(\sigma)} \\
&> \frac{\Phi^{\mathcal{E}}_{\xi}(\sigma) - \left(\Phi^{\mathcal{E}}_{L2}(\sigma) + (1-\theta_0)\Phi^{\mathcal{E}}_{L1}(\sigma)\right)}{\Phi^{\mathcal{E}}_{L2}(\sigma)} = \overline{V}^{\mathcal{E}}(\sigma, \theta_0).
\end{aligned}$$

Therefore, the proposition follows. ■