Spring 2015

Proclivity or Popularity? Exploring Agent Heterogeneity in Network Formation

Xiaotian Wang
Old Dominion University

PROCLIVITY OR POPULARITY?

EXPLORING AGENT HETEROGENEITY IN NETWORK FORMATION

by

Xiaotian Wang
B.S. September 2006, Shanxi University
M.S. June 2009, Shanxi University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

MODELING AND SIMULATION

OLD DOMINION UNIVERSITY
May 2015

Approved by:

Andrew Collins (Director)

Dean Chatfield (Member)

David Earnest (Member)

Roland Mielke (Member)

ABSTRACT

Proclivity or Popularity? Exploring Agent Heterogeneity in Network Formation

Xiaotian Wang
Old Dominion University, 2015
Director: Andrew Collins

The Barabasi-Albert model (BA model) is the standard algorithm used to describe the emergent mechanism of a scale-free network. This dissertation argues that the BA model, and its variants, rarely take agent heterogeneity into account in the analysis of network formation. In social networks, however, people's decisions to connect are strongly affected by the extent of similarity. In this dissertation, the author applies an agent-based modeling (ABM) approach to reassess the Barabasi-Albert model. This study proposes that, in forming social networks, agents are constantly balancing between instrumental and intrinsic preferences. After systematic simulation and subsequent analysis, this study finds that agents' preference of popularity and proclivity strongly shapes various attributes of simulated social networks. Moreover, this analysis of simulated networks investigates potential ways to detect this balance within real-world networks. Particularly, the scale parameter of the power-distribution is found sensitive solely to agents' preference popularity. Finally, this study employs the social media data (i.e., diffusion of different emotions) for Sina Weibo—a Chinese version Tweet—to valid the findings, and results suggest that diffusion of anger is more popularity-driven.

This thesis is dedicated to my family, especially to my husband.

ACKNOWLEDGMENTS

There are many people who have contributed to the successful completion of this dissertation. I extend many, many thanks to my committee members for their patience and hours of guidance on my research and editing of this manuscript. The extraordinary efforts of my major advisor Dr. Andrew Collins deserve special recognition. I also want to take my deeply thanks to my family, especially to my husband. Without their support, I could not have successfully completed my school career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Page

CHAPTER 1

INTRODUCTION

Social network analysis (SNA) has had an especially long tradition in various disciplines of social science [1]-[22]. Since its emergence in the 1930s, early SNA has focused primarily on the characteristics of individuals, assuming agent heterogeneity as the key factor in shaping the formation of different types of social networks [4], [7], [23]-[26]. In recent decades, due to the proliferation of the information and communications technologies (ICTs, e.g. the Internet and the mobile phone) [27]-[29], the scholarly interest in SNA has been reenergized. It is argued that the "digital revolution" has fundamentally reshaped the landscape of human society by empowering individuals in social networking, mass communication, and political mobilization [30]-[34]. A burgeoning literature, therefore, has been devoted to exploring large-scale complex networks [11], [18], [20], [35]-[43][1].

However, this dramatically increased visibility of SNA is owed mainly to statistical physicists [8], [21], [36], [37], [43]. Instead of emphasizing agent heterogeneity, statistical physicists focus more on aggregate properties and highlight the systematic regularities in spite of agent heterogeneity. In part, this changed focus has a lot to do with the rapid development of the deductive approaches to SNA, which propose that the structural and algorithmic aspects of abstract networks are the key to our understanding of network formation [8], [11], [28], [43]-[47].

[1] IEEE Transactions Journals style is used in the thesis for formatting figures, tables, and references.

Given these fundamental differences in how to approach SNA [8], [48]-[49], a question naturally arises: which approach should be adopted to explore social network formation?

1.1 Theoretical Formulations

This dissertation argues that a *generative* approach and its associated agent-based modeling (ABM) are particularly useful in bridging the seemingly incompatible inductive and deductive approaches, which helps reveal the complex nature of social network formation. First, on epistemological grounds, ABM implies that social networks can be best understood as an emergent phenomenon [49], [50]-[59]. Simple inductive or deductive conceptualization of SNA has been problematic largely because network formation is a function of complex interactions between agent characteristics and network structures. An emergent understanding of network formation, on the other hand, suggests that networks are spatially distributed systems of heterogeneous autonomous actors with bounded information and computing capacity who interact locally and, hence, is particularly helpful in SNA [11], [48]-[49].

Second, and from a theoretical perspective, ABM provides an avenue to integrate substantive theories in various disciplines in social science with abstract network models [49]. An important consequence of the cleavage between the inductive and deductive SNA is the limit of actual applications for deductive models such as BA model in social science [11], [60]-[62]. Part of this issue is that the network dynamics suggested by social theories are usually far more complex than those modeled in deductive SNA [48]. For instance, the BA model assumes uniformity of behavior of individuals, which is constantly violated in a real social network. Not surprisingly, in many social networks, considerable deviations from scale-free behaviors have been reported [11], [60]. ABM, on the other hand, is flexible enough to handle the heterogeneity, complexity, adaptability, and versatility of real social networks [48], [63].

Third, ABM is computationally advantageous in processing and simulating social networks at different scales. Both deductive and inductive SNA are plagued by analytical shortcomings. On one hand, most of the properties that are mathematically inferred in deductive SNA are present at the thermodynamic limit [48], [60], but the number of active participants in real social networks is normally much smaller, ranging only from 10^2 to 10^3 [64]-[66]. On the other hand, due to the complex dependency of the social network, inductive statistical models tend to behave poorly when social networks become large and complex. For instance, it is found that exponential random graph models (ERGMs) are "not a reasonable representation for most empirically observed social networks when they have more than 30 nodes, average degrees more than 2, and a transitivity index more than 2" [5, pp. 142].

For these reasons, this dissertation argues that ABM is a very promising tool in SNA, making the insights of different approaches complements each other [38], [48], [49]. Unfortunately, only limited studies in SNA have adopted an ABM approach, and even fewer have explicitly integrate substantive social theories with abstract networks in their ABM models.

1.2 Purpose

This dissertation proposes a simulation-oriented study of social network analysis. Specifically, this study intends to apply agent-based modeling approach and construct an ABM network model based key theories regarding social network. This ABM study allows for exploring the different patterns of network emergence when individuals' motivations vary. Drawing on arguments of individual psychologies [67]-[76], it is argued that people's motivation of social networking is a result of their two fundamental psychological needs: the *need for cognition* and the *need for affect*. Individuals who possess a high need for cognition are motivated to connect with people possessing a large column of information; that is, the people who are well connected. Individuals who possess a high need for affects are motivated to feel strong emotions and, therefore, tend to develop links with individuals with similar views. In other words, people are constantly weighting between popularity and proclivity in forming their social connections. Moreover, it is argued that how these two needs matter is contingent on costs of constructing and maintaining social ties.

This study therefore helps advance scholarly understanding of network formation by integrating substantive social theories with abstract network models by using ABM. First, this study considers the formation of complex social networks is driven primarily by the human's basic psychological needs. Hence, mechanisms of the network formation should have roots in the utility function of individuals.

$$Psychological\ needs \Rightarrow Networking\ heuristics \Rightarrow Dynamic\ formation \qquad (1)$$

Second, different psychological needs lead to varying strategies of emerging social networks. People with different needs tend to employ different network heuristics in searching their friends. Finally, individuals' strategies of networking serve as the momentum of the growth and evolution of social networks. Hence, finding out the driven mechanism for making decisions of people not only is significantly important for clarifying the network formation simulation, but also helps understanding the fundamental roots of social networking.

1.2.1 Need for Cognition and Principle of Popularity

Two motivations that are of particular relevance to social networking are: the need for cognition and the need for affect [71], [73], [75]. The need for cognition is fundamentally an instrumental concern and a stable disposition that explains individual differences in the tendency to use network to acquire such resources as information. Individuals who possess a high need for cognition have a strong motivation to connect with resourceful and well-informed people in a given social network. Consequently, individuals who are high in need for cognition are more likely to behave in line with the rational model: they should operate as if they possess a running tally about everyone's relative position in a network, which would in turn help make their decisions for networking.

Yet how can these individuals' running tally of network structure translate into abstract network models? Many works by sociologists have examined how individuals' instrumental concern can be realized in a network setting. One of the most influential works is Granovetter's "strength of weak ties" (SWT) theory [77]-[78], in which weak, bridging ties are argued to be beneficial to individuals because of their potential to introduce novel information. Burt [1] later refines the argument by emphasizing the relationship between "structural holes" and "information brokers." This, in turn, leads to Burt's conclusion that network brokers tend to enjoy various advantages because they can bridge otherwise isolated clusters. More recently, Podolny [79] argues that network structure matters not only because it serves as a "pipe" for resources, but also because it acts like a "prism," revealing important information about the inherent qualities of vertices (e.g., credibility). An individual's status (i.e., popularity) in a given network provides others a heuristic shortcut in assessing his/her credibility. In sum, these studies suggest that individuals motivated by need for cognition tend to be connected with people of high degrees in a network.

If individuals high in need for cognition are popularity-driven, what is its implication on network formation? Studies by statistical physicians on scale-free networks provide some clues. Yet, it should be noted that their concerns are mainly at the system level. As for large-scale complex networks, empirical results demonstrate that most of them are scale-free; their degree distribution follows a power law distribution [80]-[82]. The BA model [35], [37] is then introduced to describe this scale-free emergent mechanism. The BA model suggests that the growth of network size and preferential attachment are the necessary conditions for the emergence of scale-free networks. In other words, a social network that is solely composed of individuals who are high in need for cognition tends to follow a power law distribution.

In many social networks, however, significant deviations from scale-free behavior have been reported. Although numerous variants of the BA model have been developed to reproduce the growth process of social networks, most of them still share the instrumental assumption that a vertex's probability to be connected is determined primarily by its position (i.e., "popularity") in a given network [60], [62]. Instrumental concern or need for cognition, as elaborated above, is not the only motivation that governs individuals' social networking. For many, the need for affect is another important motivation.

1.2.2 The Need for Affect and Principle of Proclivity

The need for affect, in sharp contrast to need for cognition, is fundamentally intrinsic. It is a separate motivational construct that captures the degree to which people enjoy experiencing strong emotions [70], [72], [75], [83]. Individuals who are high in need for affect are more likely to view emotions as useful when making various decisions. Because these individuals tend to enjoy experiencing strong emotions, their attitudes tend to possess a stronger affective basis. This is not to say that the attitudes of individuals who are high in need for cognition are unaffected by emotion. Almost all attitudes carry some affective component. Rather, the argument here is that individuals high in need for affect possess attitudes that carry a more intense affective charge, and while affect may induce "biased reasoning" in most people, individuals high in need for affect should be especially prone to biased processing [70]-[72].

What are the implications of this "biased reasoning" for social networking? Pujol et al. [84] have pointed out that the assumptions of BA models usually lack sociological grounding. Wong et al. [84] argued that many network models have not taken the advantages of sociological and psychological insights of how social networks may be formed. It is also problematic since it assumes all the nodes possess the same preference (instrumental preferential attachment) and overlooks the potential impacts of agent heterogeneity on network formation (intrinsic preferential attachment). When joining a real social network, people are not only driven by instrumental calculation of connecting with the popular, but are also motivated by intrinsic affection of joining the like. In other words, people are constantly weighing between popularity and proclivity in forming their social connections. The impact of this mixed preferential attachment is particularly consequential on such social networks as political communication. More importantly, this assumption is supported from the social theory: homophily.

McPherson et al. [26] argue that homophily is the principle that a contact between similar people occurs at a higher rate than among dissimilar people, and the similarity could be regarding many types of personal characteristic positions, including gender, religion, social class, education and other intra-personal or behavioral characteristics. In fact, there are some models taking homophily into consideration, somehow not using the specific term but essentially the similar meaning. Robins et al. [86] presented network models for social selection process. Although where characteristic positions affecting the social relationship formation are concerned, it is broken between the local behavior and the global pattern. In other words, there is no analysis for the properties of large social networks. Newman and Girvan [87] conducted a network model discussing the mechanism of assortative mixing, which is, the nodes with a similar degree level like to link with each other. However, it actually is a special case of preferential attachment, albeit the similarity of nodes is concerned.

This study proposes an integrative model of preferential attachment encompassing both instrumental calculation and intrinsic similarity, which is a term transformed from homophily. Particularly, it emphasizes the ways in which both agent-heterogeneity and network position affect social network formation. Agent-based modeling is chosen as the paradigm to conduct this study. This integrative approach can strongly advance our understanding about the formation of social networks.

1.2.3 Individual Capacity and Mode of Mass Media

Up to this point, it is clear that both the principle of popularity and the principle of proclivity are key strategies adopted by individuals in social networking. Yet, it should be noted that the actual impacts of these two principles are contingent on the costs of acquiring information and establishing connections; that is, the constraints of communication modes.

Scholars have long found that the individual acquisition costs of information vary widely under different media settings. A key theme across many disciplines in social science is the question to what extent lower communications costs have made people more connected and less isolated [34], [88]. Recent years have seen an explosion of research on the topic. To some degree, the diffusion of ICT is at the heart of the latest surge of studies on social network. The development of the Internet has greatly increased the means by which people communicate with each other. In many respects, ICTs appears closer to traditional media than to mass media in the sense that they are used as personal communication platforms.

A review of relevant literature suggests that competing views on the effects of technological innovation and its resulting impact on ordinary people provide the basis for a rich debate on the relationship between social isolation and the fragmentation of society. Whereas few doubt that technology has greatly expanded our capacity to connect with others, the impact of technology on the perception of being connected is more controversial. For instance, many argue that ICT reduces or counteracts the impact of geography on structuring opportunities for social interactions, and that they take time away from other face-to-face activities [89].

Rather than joining the existing debate on the exact impacts of ICT, this study intends to explore to what extent the network formation is constrained by different modes of communication. An ABM approach to this problem helps elucidate the structural differences of social networks under different media setting.

In summary, the above three factors strongly affect the patterns of social networking among individuals with different inherent characteristics. People can be driven purely by one of the two mechanisms or jointly by a different combination of instrumental and intrinsic mechanisms. It should be noted that in either case the formation of social network is simultaneously affected by the context of the mass media; that is, the capacity of their own to construct their social networks with or without conscious. Accordingly, these different networking strategies lead to the emergence of variant social network patterns. Hence, the purpose of this dissertation is to reveal the different social network patterns outputted from the network system consisted and functioned of these three factors, and consequently to explain the characteristics owned by their relative network topologies.

Previous research in social science, actually, has investigated how the need for cognition (i.e., "popularity") and the need for affect (i.e., "proclivity") influence opinion formation, but this is the first study to the researcher's knowledge that investigates whether these motivations shape the formation of social networks. This inquiry makes no claims about the role that cognition or affects play in the evaluation of information. Rather, the purpose here is to investigate how individual differences in psychological motivations shape how people form their social circles by integrating ABM approach and abstract network models.

1.3 Problem

Motived by the research questions addressed above, this dissertation explicitly emphasizes an agent-based thinking and intends to explore three key factors in shaping the emergence of social networks.

To do so, two critical problems need to be systematically explored. The first problem is the agent heterogeneity in the emergence of social networks. Scholars have long noticed that the network, modeled usually as a "structure", strongly shapes the outcome on lots of social topics, such as, opinion formation, information diffusion [5], [11], [25]. However, the specific mechanisms that lead to the observed networks are severely underexplored. The actual social networks are a product of human behavior (if only located in the range of human communication) and are generated by interactive social agents are accordingly unpredictable and ever changing.

In this process, a key issue is how agent heterogeneity should be modeled. Until now, scholars have focused mostly on tracing the formation process of social networks or on inferring different social network topologies and their underlying mathematical properties. Unfortunately, the complication inherent in the human social world, particularly those derived from agent heterogeneity, is far more complex than expected. Knowledge about social networks is still insufficient to support an understanding of the existing social problems that were believed to have lots to do with social structures. The continuing exploration on social networking mechanisms requires new methods to incorporate agent heterogeneity into the modeling of social network formation.

The problem associated with agent heterogeneity naturally leads to the second problem: the social networking strategies of agents (people). The BA model, as described above, provides a new path toward the understanding the network formation, though Barabasi and Albert arbitrarily raise a simple assumption and generate a stylized network topology that explains one among many actual networks. This dissertation argues that a key contribution of the BA model is that it highlights one important driven mechanism: connecting with the popular. In light of their study, many would raise the question: in the real world networking are social agents governed only by one strategy? Obviously there are more. This dissertation proposes to flavor the agent with heterogeneous attributes. Besides the choice of connecting with the popular, another universal communicating rule would be connecting with the like. Relying on the thick research archive focusing on this topic, this dissertation introduces the "homophily" theory and raises another connecting strategy for social agents: connecting with the like. People living in the real social world usually are not performing to the extreme.

In this dissertation, thanks to advancements in the agent-based simulation technique, agents are installed with the tolerance of making choices in between the two rules; that is, agents are allowed to make choice like real people: an agent may consider connecting with people like itself, but it is also possible that the agent is more interested in constructing a relationship with the popular.

1.3.1 Uniqueness of Social Agents

Scholars are uncertain about how to obtain more accurate and realistic models of the social networks in the real world. Although many sophisticated models have been proposed (such as the BA model), this study suggests that the hidden micro mechanisms of their models are still underdeveloped. First, network topology in social networks is a graph analogy of the interactions of people [3], [8], [43]. When trying to make a model of a social network, it is important to note that what happens in the real world is intentional and rational (in most cases) communication, and thus connections are made between people with different personal characteristics. Hence, agent heterogeneity is the key factor that needs to be considered in the model process. However, most studies ignore the "humanity" of agents.

Second, rather than passively following the "iron law" of "connecting with the popular", which is proposed by Barabási and Albert [35] as one hypothesis, people usually intentionally and consciously build up connections with their own preferred "targets", and these connections can be strong ties or weak ties based on the purposes of people who are involved in a concrete event [8]. In some cases, people are building relationships not out of a "popularity" consideration but from sharing the same opinion. A convenient example is the tweeting and retweeting activities on the Internet. As a matter of fact, most activities of social media follow the logic of intrinsic preferential attachment (e.g., party orientation, [27]). The question at this point is that do the subjective opinions held by people affect the networks topologies? If so, how do they affect the networks topologies?

Third, social scientists from different disciplines have long emphasized the conditioning impacts of network structures on various sociopolitical outcomes, ranging from economic transaction to political participation [11], [67], [76], [90]-[93]. An emphasis on the consequences of social network, though necessary in revealing why and how social network matters, tends to distract us from a key fact: a social network per se is interactively constructed by intentional and heterogeneous individuals. Therefore, without a thorough examination of how various social networks are formed, our understanding about impacts of social networks is likely to be biased.

In summary, social networks, comparing to the well-explored existing networks, such as transportation networks or WWW networks, are significantly different in these facets [94]:

1. Social networks are emerged from the complex communication among a variety of individuals. Consequently, the existence and the structures are greatly driven and influenced by people's psychological needs and the social activities. Different motivations and intentions lead to different network structures. Whether the different network structures accordingly lead to the different network topologies needs further measurements, which is one of the key questions that this study intends to explore.
2. Social networks are constantly changing. People's social activities and the concerns involved in the process of the social activities cause the changes on the social networks.

3. Social networks are large. With the advancement of ICTs, the size of the social networks turns out to be extremely large.
4. Traditional computing methodologies are not good for the social network research. The relationships between the data are hardly assumed as linear. Social network studies ask for the new and effective methodology to be explored.

1.3.2 Network Heuristics of Agents

As discussed above, in recent decades, the scholarly interest in SNA is reenergized by the proliferation of the information and communications technologies (ICTs) like the Internet and the mobile phone network.

Particularly, Barabási and Albert [35] argue that the vertex connectivities in large networks tend to follow a scale-free power law distribution. The key underlying mechanism is that a vertex's probability to be connected is determined only by its relative position (i.e., connectiveness or "popularity") in the existing network. The formation of large networks is governed by this robust self-organizing mechanism that goes beyond the particulars of agents or individual systems. However, in many social networks considerable deviations from scale-free behaviors have been reported (Shirazi, Namaki, Roohi, and Jafari 2013). Numerous variants of the BA model, accordingly, have been developed to reproduce the growth process of social networks, and most of them still share the very key instrumental assumption that a vertex's probability to be connected is determined primarily by its "popularity" in a given network.

However, it should be noted that when joining a real social network, people are not only driven by instrumental calculation of connecting with the popular, but also motivated by the intrinsic affection of joining the like. The impact of this mixed network formation is particularly consequential to social networks such as political communication networks. For instance, when people appear in a new community and start to build their network, the two endogenous driven mechanisms would lead people to build up their social networks. Under extreme conditions, by following the preferential attachment only, people would only be interested in linking with popular people. By following homophily only, people would focus on connecting with the other people that have similar intrinsic properties with them, thus creating a more comfortable social ambiance. Certainly, the latter is a human behavior factor, which is labeled as an "intrinsic" intention to construct a network in this study. It is realized that in the real world, most people make decisions based on both mechanisms in different levels instead of in the extreme situations. Regarding the specifics of the heterogeneous attachment, more details will be introduced in Chapter 3 and 4.

1.4 Method and Procedure

In this dissertation, agent-based modeling (ABM) approach is used to reassess the Barabási-Albert (BA) model, the classical algorithm used to describe the emergent mechanism of scale-free networks. This approach allows for the incorporation of agent heterogeneity, which is rarely considered in the BA model and its extended models. The simulation will be constructed according to the decomposition of the decision process of heterogeneous agents. Hence, a statistical analysis will be conducted to export the information out of the data output from the simulation.

1.4.1 Simulation and Agent-Based Modeling

For at least three reasons, this dissertation proposes that agent-based thinking is the best way to formalize and conceptualize the network formation process. First, argued by Miller and Page [58], social networks are complex system and local interactions produce the global pattern. The global network topologies are grown from the local interactions between people—agents in the model.

ABM provides us a way to harvest on the global pattern by seeding the local agent activities. Analogously, the choice people made for linking with each other could generate and influence the structures of social networks. At this point, ABM is, so far, the only approach for linking the local behavior and the global pattern.

Second, agent-based modeling is the appropriate tool for learning complexity [95]. It has been widely accepted that there are many problems that cannot be calculated or represented by traditional mathematical methods; simulation provides a rational alternative to address these problems.

Here, ABM supplies an approach of simulating the human behavior. The traditional computational method does not contain this part of the function. It is emphasized that the method in this study consists of two parts: simulation then analysis. Simulation is used to construct a simulated social networking world based on rules that rely on the literature studies and the mathematical understanding of social activity. Without the simulating process, the study would be another piece of empirical computation, rather than an exploration to the unknown social communicating mechanism.

Third, validation is important for all the models, especially for agent-based modeling. In this proposed work, simulation experiments will be compared to real world network data to validate the findings. Validation is always in the first order when models used for conducting simulations, this study is not an exception. Real-world network data will be utilized for testing the results of the study.

For these reasons, this dissertation argues that ABM is a very promising tool in SNA, making the insights of different approaches complements each other [38], [48], [49], [96]. Unfortunately, only limited studies in SNA have adopted an ABM approach, and even fewer have explicitly integrate substantive social theories with abstract networks in their ABM models.

Several key assumptions of ABM involved in this dissertation need to be briefly explained here. Agents in ABM have their assumptions and they are conventionally and well acceptably described as having the following characteristics [57]:

- *Perception*. Agents perceive their environment. They may respond to changes in the environment and the presence of other agents and accordingly adjust their activity.
- *Performance*. Agents are installed with a set of behaviors that they are capable of performing. For instance, they are able to communicate with each other or send and receive messages.
- *Memory*. Advantaged by the ICTs, agents are able to carry memories. They are capable of recalling information of their past states and actions.
- *Policy*. Agents are performing according to the rule of the simulated world maker.

Behavioral Rule

In this study, all agents' abilities listed above are performed for accomplishing the research goal. Regarding these four points, agents in this study are capable of:

Perception. Agents in the model need to identify that the one they want to link with in each tick. They accomplish this goal through calculating the probability based on identifying the agent with more resources or with the similar characteristics to make their decision.

Performance. Agents observe the other agents (calculate) and make decisions on choosing the one to which they are going to link.

Memories. Agents need to memorize the characteristics (color) and identify the ones they linked with. That information is reported out of the system. This information is analyzed in the final stage.

Policy. Policy is fairly important in this study. Policy supports the whole simulated world. They are the hypothesis of this study and will be fully introduced in the following chapters.

Simulation Space

Combined with the tool NetLogo used in this dissertation, the simulation runs non-considering the physical location of agents. The structure of the networks is the focus of this study. Consequently, the "patch" setting for building up the environment of agents in NetLogo is not that important in the study. Agents are assumed to only consider the messages exchanged with each other, rather than taking care about their location.

Data Export

Data will be exported from the simulation and the analysis based the data will be introduced in detail in Chapter 4.

1.4.2 Statistical Analysis of the Network

Researchers of SNA have long employed a variety of measurement to describe structural characteristics of a network. In this dissertation, many vertex- and local-level measurements will be used to examine various properties of simulated networks. These network metrics provide a convenient way to evaluate the impacts of heterogeneous attachment on network formation. For the definitions and details of these measurements, please see Chapter 2. These measurements serve valuable purposes in describing and understanding network features that might bear on particular research questions. However, it should be noted that a single network metric is inherently limited in revealing the complex nature of a particular network. Beyond simple descriptive network statistics, many analytical methods have been introduced to explore the complex emergence of networks. The reasons are multiple.

First, social behavior is complex, and stochastic models allow us to capture both the regularities in the processes giving rise to network ties while at the same time recognizing that there is variability that is unlikely to be modeled in detail. Second, statistical models also allow inferences about whether certain network substructures—often represented in the model by one or a small number of parameters—are more commonly observed in the network than might be expected by chance. Third, sometimes, different social processes may make similar qualitative predictions about network structures and it is only through careful quantitative modeling that the differences in predictions can be evaluated. Therefore, it is important, if not necessary, to go beyond simple network statistics and search for a well-fitting analytical model of a simulated or observed network and in a particular statistical model.

1.4.3 Validation and Verification against an Actual Social Net

After simulation and statistical analysis, a remaining question is the extent to which the model proposed in this dissertation helps reveal micro-dynamics in actual social networks. From the perspective of agent-based modeling, this is about the issue of validation. Broadly speaking, validation in agent-based modeling concerns whether the simulation is a good model of the target. A model that can be relied on to reflect the behavior of the target is "valid." Gilbert and Troitzsch [97] suggest that validity can be ascertained by comparing the output of the simulation with data collected from the target. However, there are several caveats that must be borne in mind.

In this study, one enormous challenge is to conduct a micro-level validation of the agent-based model, which requires laboratory experiments testing under what circumstances individuals are popularity- or proclivity-oriented. Therefore, this dissertation turns to the approach of the macro-level validation. Specifically, relying on data collected from social media in China [64], this dissertation is able to validate the proposed model. After applying findings emerged in Chapter 4, this dissertation is able to clarify the micro-foundation of the diffusion patterns of different emotions.

1.5 Significance and Relevance

In three ways, this study contributes to current studies of network formation. First, by exploring the impacts agent heterogeneity, this study highlights an important yet less examined mechanism in network formation; that is, intrinsic preferential attachment. This mechanism becomes particularly important in the age of new media, in which individuals' capacity in homophilous sorting has been strongly boosted by information and communications technologies [98]-[101]. Therefore, an investigation of the impacts of intrinsic preferential attachment can significantly enrich our understanding about large-scale social networks.

Second, by emphasizing both micro-mechanisms in governing dyad formation and macro mathematical properties of large-scale networks, this dissertation concurs with Barabási [37] that "the structure and the evolution of networks are inseparable."

Third, joining many recent works [49], [58], [60], this study demonstrates that ABM, given its explicit emphasis on complexity and emergence, provides a promising perspective and a useful method to explore the dynamic evolution of large-scale social networks.

1.6 Organization of The Dissertation

Following this chapter, Chapter 2 intends to review some background knowledge about the existing network theory, theoretically and technically. Specifically, Chapter 2 focuses on both the statistical analysis and the mathematical analysis of network formation. To do so, Chapter 2 first aims to present a core set of methods and models for the analysis of measurements that are either of or from a system conceptualized as a network. These methods and models are of critical importance in characterizing the observed structures of given networks. Then, Chapter 2 turns to the key mathematical models that describe the emergence of various networks.

Chapter 3 is devoted to a review of extent studies on agent-based modeling and social networks, a discussion of the advantages associated with agent-based modeling in simulating network formation, and, finally, an explanation of operationalization of network formation based heterogeneous attachment.

In Chapter 4, this dissertation examines various properties of simulated social networks generated based upon different combination of popularity-proclivity parameter, λ, and individual capacity, m. For this examination, a large number of simulated social networks was generated based on individuals' heterogeneous attachment. These network properties encompass vertex-level, local-level, and global-level network statistics. Through examination of the relationships of these network statistics to the heterogeneous attachment, through λ, and individual capacity, m, the following questions will be addressed: To what extent does agent heterogeneity affect actual formation of social networks? More specifically, when people are popularity- and proclivity-orientated, will the resultant social network be better connected, equally connected, or efficiently connected?

Finally, Chapter 5 summarizes major findings that emerged from Chapter 4, and illuminate the key theoretical and practical implications that can be learned from these findings. The findings from this study significantly supplement our understanding of the emergence of social networks, particularly among intrinsically heterogeneous agents. In addition, the chapter discusses the limitations and drawbacks of this studies and possible directions for future studies.

CHAPTER 2

BACKGROUND OF THE STUDY: METHODS AND MODELS

Social network analysis—or in its general sense, network analysis—has a very long tradition. As noted by Brandes and Erlebach [8] as well as Kolaczyk [102], the seeds of network-based analysis in the sciences, particularly its mathematical foundation of graph theory, can be traced back to the Euler's solution to the well-known Konigsberg Bridge problem in 1735. In the study, Euler proved that it was impossible to traverse each of the seven bridges of that city each only once. Since then, particularly after the mid-1800's, network analysis has developed rapidly in a wide range of disciplines. Starting from the 1930s, network analysis has experienced another wave of rapid expansion thanks to the systematic collection and analysis of data on networks of one form or another in disciplines like sociology [5], [25], [43].

Provided with these intellectual accumulations, during the mid-1950s many sociologists started to take a particularly quantitative view towards the topic of social structure, they began developing the use of networks to characterize interactions within social groups [6], [7], [10], [86]. At the same time, as with the fields of operations research and computer science, networks were incorporated into solving problems involving transportation, allocation, and the like, producing network-based approaches to modeling and analysis of various complex systems [8], [102].

In recent decades, the proliferation of information and communications technologies (ICTs) has reenergized scholarly interest in network analysis. Particularly, statistical physicists and computer scientists contribute significantly to studies on dynamics on these large-scale complex networks [8], [21], [36], [37], [43]. The development of network analysis in such a wide range of disciplines, as emphasized by Brandes et al. [6], naturally leads to a question: Given it permeates a wide range of traditional disciplines (e.g., from the physical and mathematical sciences to the social sciences and humanities), is there a science of networks?

It should be noted that even when the disciplinary boundaries still tend to create academic "islands" of different interests, methods, and goals, there is much overlap in disciplinary network research. This dissertation explicitly emphasizes an agent-based modeling (ABM) approach to exploring the emergence of network formation. Despite its ABM, this dissertation also relies heavily on developments of network studies in other disciplines. Therefore, it is of critical importance to examine the must-known knowledge about the existing network theory, theoretically and technically. To do so, this chapter focuses on (1) the statistical analysis of *network data* and (2) the mathematical analysis of *network formation*. More specifically, this chapter first aims to present a core set of methods and models for the analysis of measurements that are either of or from a system conceptualized as a network. These methods and models are of critical importance in characterizing the observed structures of given networks. Then, this chapter turns to the key mathematical models that describe the emergence of various networks.

2.1 Literature Review: Network Analysis

As discussed above, social network analysis (SNA) has evolved over the years within its application to the social sciences [4], [7], [23]-[26]. It was not until 1930s that researchers started to employ the perspective of a "social network" for drawing the features and shape of social structures [4]. By the 1970s, a growing number of scholars applied SNA and combined the different domains to study various social relations. It has been primarily focused on the characteristics of individuals, assuming agent heterogeneity as the key factor in shaping the formation of different types of social networks [23], [26], [86], [103]-[107]. For example, Stanley Milgram, a professor and a creative practitioner of experimental psychology in Harvard social relations department, rediscovered "six degrees of separation" thesis which originally raised by Karinthy in 1929, and now, this theory is often cited to indicate a feature of the social networks: short path length [4].

Thanks to the rapid development of computational and modeling technology, the most recent 20 years have witnessed a dramatic growth in the innovative social network analysis. In this section, three items of basic commonly used network knowledge need to be introduced. They are: network properties, levels of analysis, and network models. Network properties are used for measuring and differentiating the features of networks. Levels of analysis help informing the objectives of networks in the research from different levels. Network models summarize the content about three main lines of well-studied network topologies.

2.1.1 Network Descriptions

For a relational system, in which agents are constantly interacting with one another, a network graph representation is possible. Such a network representation can be constructed from an appropriate set of measurements [8], [102], [108]-[117]. People exchange information, goods, or services with one another so frequently that social exchange has been viewed as one basic form of social behavior. An objective account of any social exchange would render concordance: both actors involved in the exchange would agree about the nature of their relationship.

Given a network graph representation of a social system, a natural question is how to explore and describe the characteristics and structural properties of the network. These tasks range from the calculation of simple metrics summarizing topological structure, both local and global, to the unsupervised extraction of complex relational patterns. In this chapter, this dissertation presents a core set of tools and techniques for examining various structural properties of a given network.

Network properties refer to several important definitions about quantities of measuring the features of network topologies. Some properties of networks can be quantified and thus used for describing characters and structures of the network. For example, various types of basic social dynamics can be represented by triplets of vertices with a particular pattern of ties among them (i.e., triads). Questions involving the movement of information or commodities usually can be posed in terms of paths on the network graph and flow along those paths. Certain notions of the "importance" of individual system elements may be captured by measures of how "central" the corresponding vertex is in the network. The search for "communities" and analogous types of unspecified "groups" within a system frequently may be addressed as a graph partitioning problem.

Accordingly, scholars often use these properties for measuring network characteristics and differentiating between them [4], [7], [8], [102]. The introduction of these definitions is originally rooted in graph theory, and they are well accepted and used by scholars from different research domains. The structural analysis of network graphs has traditionally been treated primarily as a descriptive task, as opposed to an inferential task, and the tools commonly used for such purposes derive largely from areas outside of 'mainstream' statistics. For example, overwhelming proportions of these tools is naturally graph-theoretic in nature, and thus have their origins in mathematics and computer science [8], [102]. Similarly, the field of social network analysis has been another key source, contributing tools usually aimed at capturing basic aspects of social structure and dynamics. More recently, the field of physics has also been an important contributor, with the proposed tools often motivated by analogues in statistical mechanics.

2.1.1.1 Basic Definitions

This chapter follows a common notation system in network analysis [4], [7], [8], [102]. Formally, a graph $G=(V,E)$ is a mathematical structure consisting of a set V of vertices and a set E of edges, where elements of E are unordered pairs $\{i,j\}$ of distinct vertices $\{i,j\}\in V$. Moreover, the number of vertices $N_v=|V|$ is commonly referred as the order of the graph G, and the number of edges $N_e=|E|$ is also known as the size of the graph G, respectively. Often, and without loss of generality, the vertices can be simply labeled with the integers $1,2,\ldots,N_v$. Similar notations can be used for the edges. A graph $S=(V_S,E_S)$ is a subgraph of another graph $G=(V_G,E_G)$, if $V_S\subseteq V_G$ and $E_S\subseteq E_G$. An induced subgraph of G' is a subgraph $G'=(V',E')$, where $V_S\subseteq V_G$ is a pre-specified subset of vertices and $E_S\subseteq E_G$ is the collection of edges to be found in G among that subset of vertices.

As defined, a graph has no edges for which both ends connect to a single vertex (called loops) and no pairs of vertices with more than one edge between them (called multi-edges) [4], [7], [8], [102]. A graph with either of these properties is called a multi-graph. For simplicity, and reflecting the bulk of common practice, the presentation in this dissertation will concentrate primarily on graphs, not multi-graphs, though reference to the latter will be made where appropriate. When it is necessary to indicate explicitly that a graph G is not a multi-graph, it will be referred to it as a simple graph, and its edges, as proper edges.

A graph G for which each edge in E has an ordering to its vertices (i.e., so that $\{i, j\}$ is distinct from $\{j,i\}$, for $\{j,i\}$) is called a directed graph or digraph. Such edges are called directed edges or arcs, with the direction of an arc $\{i, j\}$ reading from left to right, from the tail u to the head v. It should be noted that there is a natural extension of digraphs to multi-digraphs, where multiple arcs (i.e., multi-arcs) share the same head and tail. Moreover, digraphs may have two arcs between a pair of vertices without there being multi-arcs if the vertices play opposite roles of head and tail for the respective arcs. In this case, the two arcs are said to be mutual. However, in this study for the purpose of simplicity, only simple and undirected graphs are discussed.

Beyond this simple formal setup, it is necessary to describe the connectivity of a graph. One of the most basic notions of connectivity is that of adjacency. Two vertices $i, j \in V$ are said to be adjacent if joined by an edge in E. Similarly, two edges e $e_1, e_2 \in E$ are adjacent if joined by a common endpoint in V. A vertex $v \in V$ is incident on an edge $e \in E$ if v is an endpoint of e.

2.1.1.2 Degree, Degree Distribution, and Degree Correlation

Networks are usually described by the network graphs and network graphs consist of vertices and edges. The degree of a vertex provides a quantification of the edges connected with the vertex within a network graph. Provided with the above setup, the notion of the vertex degree of v, say d_v, can be defined as the number of edges incident on v. The degree sequence of a graph G is the sequence formed by arranging the vertex degrees d_v in non-decreasing order. The sum of the elements of the degree sequence is equal to twice the number of edges in the graph (i.e., twice the size of the graph). Note that for digraphs, the vertex degree is replaced by the in-degree (i.e., d_v^{in}) and the out-degree (i.e., d_v^{out}), which count the number of edges pointing in towards and out from a vertex, respectively. Hence, digraphs have both an in-degree sequence and an out-degree sequence. However, in this dissertation only d_v is considered given that only undirected graphs are examined.

One way to gauge the overall degree structure in the graph G is to calculate k_{ave}. Average degree (k_{ave}) refers to the number of links connected to a node, and average degree indicates the density of a network structure. The formula below provides a general meaning, and it can vary as in accordance with network structure,

$$k_{ave} = \frac{2E}{N} \tag{2}$$

where, E is the number of the edges in the network graph, and N is the size of the network.

Degree distribution is another important measure when the vertex degrees are considered in aggregate for indicating the characteristics of the network graph [102]. Given a network graph G, f_d is defined as fraction of vertices $v \in V$ with degree $d_v = d$. The collection $\{f_d\}_{d \geq 0}$ is called the degree distribution of G. For directed graphs, degree distributions may be defined analogously for in- and out-degrees. The degree distribution provides a natural summary of the connectivity in the graph.

Given an observed degree distribution, it is intuitive to summarize this distribution. Reporting basic summary statistics in forms such as moments and quantiles is common. It is also common in some fields to report the fit of some simple parametric families of distributions to the observed degree distribution (e.g., exponential distribution).

The degree distribution is useful as a composite summary of how the degree varies across vertices in the graph, but it does not provide any information on precisely which vertices are connected to which others. To capture information of this sort, it is helpful to establish summaries that describe the patterns of association among vertices of given degrees. In fact, such summaries can be quite important, as two graphs may have identical degree sequences and yet otherwise differ noticeably in the way their vertices are paired. A natural starting point is to define a two-dimensional analogue of the degree distribution, capturing the relative frequency with which the two vertices at the end of an arbitrarily selected edge in the graph have a given pair of degrees, and this is commonly captured by the concept of degree correlation.

2.1.1.3 Centrality Measurements

When trying to figure out the "importance" of a vertex in a network graph, measures of centrality are used to quantify such "importance." Centrality indices are to quantify an intuitive feeling that in most networks some vertices or edges are more central than others. As documented by Brandes and Erlebach [8], many vertex centrality indices were introduced for the first time in the 1950s: e.g., the Bavelas index [118], degree centrality, or a first feedback centrality, introduced by Seeley [119]. Given many different measurements of centrality, however, not every centrality index was suitable to every application. In this section, rather than conducting as a comprehensive review and comparison of all centrality measurement, this dissertation focuses primarily on three key measurements.

The first is closeness centrality. It measures the notion that a vertex be "close" to the other vertices [8], [102]. It is measured using the following formula, which calculates the inverse of the measure of the total distance of a vertex from all others:

$$c_{cl}(v) = \frac{1}{\sum_{u \in V} dist(v,u)} \tag{3}$$

where $\sigma(s,t \mid v)$ is the geodesic distance between the vertices. The shortest path or distance between all pairs of vertices in G (network graph) is needed for calculating the centrality. This measure assumes the graph G is connected, while all vertices in principle will have centrality $c_{cl}(v) = 0$, being of infinite distance from at least one other vertex.

It is evident that the focus of closeness centrality lies, for example, on measuring the closeness of a person to all other people in the network. People with a small total distance are considered as more important than those with a high total distance. Many variants of closeness-based measures have been developed. The measure defined in Eq. (3) is the most commonly employed.

Another popular class of centralities is based on the perspective that "importance" relates to where a vertex is located with respect to the paths in the network graph and betweenness centrality measures the extent to which a vertex is located "between" other pairs of vertices [8], [120]. Therefore, it is commonly referred as shortest-path betweenness centrality. Betweenness centrality score is defined as the following formula,

$$c_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s,t \mid v)}{\sigma(s,t)} \tag{4}$$

where $\sigma(s,t \mid v)$ is the total number of shortest paths between s and t that pass through v, and $\sigma(s,t) = \sum_{v}(s,t \mid v)$. There are two steps when calculating the collection of all betweenness centralities $c_B(v)$: (1) calculating the lengths of shortest paths among all pairs of vertices, and (2) computing the summation for each vertex in the formula above. The betweenness centrality index was introduced in [121] and is found in a wide field of applications.

There are many other indexes for indicating centrality of a vertex in a network graph. One of these measures indicates the notion of "status" or "rank." This index embodies the cognition that the more central the neighbors of a vertex are, the more central the vertex [8].

All centrality measures defined above have had many variations and extensions. One example is the variation to the betweenness centrality, where instead of counting all shortest paths passing through a vertex, modifying the underlying path set is suggested, such as, the shortest paths of length bounded by some k or shortest paths that are disjointed in their vertices.

2.1.1.4 Density and Clustering Coefficient

Density (D) is an indicator for showing the intensive level of edges in a network structure. It is defined as the ratio of the number of edges (E) to the possible potential number of edges for a network model. This can be calculated by the formula,

$$D = \frac{2E}{N(N-1)} \tag{5}$$

where, E is the number of the edges in the network graph, and N is the size of the network.

The concept of density defined in Eq. (5) can be used at a local level and help define a measure of local density. This local density can characterizes the extent to which subsets of vertices are dense. Such measures are commonly based on ratios of the number of edges among a subset of vertices to the total number of possible edges. For example, in a graph G with no self-loops and no multiple edges, the local density of a subgraph $S = (V_S, E_S)$ can be defined as,

$$D_S = \frac{2E_S}{N_S(N_S-1)} \tag{6}$$

where, E_S is the number of the edges in the subgraph S, and N_S is the size of the subgraph.

From Eq. (6), it is apparent that the interpretation of D_S depends heavily on the choice of subgraph S. If $S = G$, D_S then denotes the overall density of the graph G. Yet, if $S = S_v$ is the set of neighbors of a vertex $v \in V$ and the edges between them, D_S produces a measure of density in the immediate neighborhood of v. This particular measurement is also known as the clustering coefficient [7], [8], [102].

Given the definition $S = S_v$, the clustering coefficient can be used to estimate the situation of "all my friends know each other." It has become a standard quantity used in the analysis of network structure [8]. In fact, these values are tested quite often in real-world networks. It is reported that local clustering coefficient increase when the vertex degrees decrease, and vice versa. For example, Watts and Strogatz [41] proposed $D(S_v)$ as a summary of the extent to which there is "clustering" of edges local to v and propose that the average of $D(S_v)$ over all vertices be used as a clustering coefficient for the overall graph [120][2].

2.1.1.5 Connectivity and Cuts

Considering the connectivity of a graph G, a basic and intuitive question is whether a given graph can be separated into distinct subgraphs. If it cannot, it is necessary to quantify how close to being able to do so it is [120]. Intimately related to these issues are questions associated with the flow of "information" in the graph.

---

[2] In fact, these measures of clustering can be expressed alternatively in terms of the density of triangles among connected triples. A triangle is a complete subgraph of order three. A connected triple is a subgraph of three vertices connected by two. Intuitively, a measure of the frequency with which connected triples "close" to form triangles will provide some indication of the extent to which edges are 'clustered' in the graph.

If a graph G is connected, every vertex is reachable from every other (i.e., if for any two vertices, there exists a walk between the two); a connected component of a graph is a maximally connected subgraph. Often it is the case that one of the connected components in a graph G dominates the others in magnitude, in that it contains the vast majority of the vertices in G. Such a component is called, borrowing terminology from random graph theory, the giant component.

An intuitive way to gauge such a giant component or a given graph is diameter. A common notion of distance between vertices on a graph is defined as the length of the shortest path(s) between the vertices (which is set equal to infinity if no such path exists). This distance is often referred to as geodesic distance, with "geodesic" being another name for the shortest path(s). The value of the longest distance in a graph is called the diameter of the graph.

What is closely related to the measurement of diameter is the average shortest path length (L). Average shortest path length indicates the steps from one node in the network to another. It is defined as the average number of edges that must be traversed in the shortest path between any two pairs of vertices in the network [41].

2.1.1.6 Graph Partitioning

Partitioning refers to the segmentation grouped naturally. The concept is of critical importance where the phenomena like "cohesion" or "solidarity" are concerned [103], [122]. Formally, a partition $L = \{C_1, \ldots, C_K\}$ of a finite set S is a decomposition of S into K disjoint, nonempty subsets C_k such that $U_{k=1}^{K} C_k = S$ [7], [8], [102]. Partitioning takes care about finding the groups of vertices that naturally generated and vertices in each group demonstrate a "cohesiveness" regarding to some internal patterns. Rather than randomly grouped together, the "cohesive" subsets of vertices indicates that these vertices within a group are well connected among each other and are well separated from the vertices of the other groups. The partitioning problem is often referred to as "community detection" in complex networks research as well.

Compared to the definition of graph partitioning, hierarchical clustering is referred to as a more general concept and the former one is usually considered to be the variations of the latter one in data analysis.

Hierarchical clustering refers to many different methods used to explore the space of all possible partitions L in a defined network graph. Some of these methods are classified as agglomerative, based on merging processes for the successive partitions [123]-[125]. Some are classified as divisive, based on the refinement of partitions through the process of splitting [120]. In both of these methods, the candidate partition is modified in a way that minimizes the cost for measuring and there is lots of costs measuring methods have been proposed. However, in the analysis part of this study, the detailed measuring of hierarchical clustering is not going to be tested. No formula is/was provided for any methods consequently.

2.1.1.6 Assortativity

Assortativity, indicating a phenomenon wherein some vertices have priority to link with each other, based on a certain characteristics, usually refers to the term of assortative mixing in some of the social network research [87], [126], [127]. Relatively, assortativity coefficients refer to measuring the extent of assortativity mixing in a given network, which is described as the variations on the concept of correlation coefficients [120]. The vertex characteristics can be categorical, ordinal, or continuous. When taking the categorical case as the consideration, suppose that each vertex in graph G is labeled according to one of M categories, the assortativity coefficient is defined to be,

$$r_a = \frac{\sum_i f_{ii} - \sum_i f_{i+} f_{+i}}{1 - \sum_i f_{i+} f_{+i}} \tag{7}$$

where, f_{ij} is the fraction of edges in G that join a vertex in the ith category with a vertex in the jth category, and f_{i+} and f_{+i} indicate the ith marginal row and column sums of the resulting matrix.

The value r_a lies between -1 and 1. It is a value of zero if only randomness is involved in the linking among the vertices. It is a value of 1 if there is a perfect assortative mixing among the linkings. It is a value of -1 if there is a perfect disassortative mixing and every edge connects vertices of two different categories in the graph.

2.1.2 Levels of Analysis

An axiom of SNA is that instead of properties of the units themselves, social phenomena should be primarily conceived of and investigated through the properties of relations between and within the units [4]. Newman et al. [20] point out that the social network is self-organizing, emergent and complex, such that a globally coherent pattern appears from the local interaction of the elements that make up the system. The question is how scientists locate the scale of the networks. In practice, a different network scale leads to different computation quantity, computation speed, and a different level of consideration about the components of complexity. In light of this, social networks are analyzed at the scale in relation to the research questions. Generally, there are three levels of social network analysis scale and they are listed in the following section [8].

2.1.2.1 Micro-level

As noted by Brandes and Erlebach [8], research at this level particularly starts with the individual, or a small group of individuals under a specific social environment set. It can be categorized as follows.

1. Actor level. This analysis unit is often about an individual (e.g. actor and ego-centered network). As noted by Wasserman and Faust [7], "ego-centered network consists of a focal actor, termed ego, as set of alters who have ties to ego, and measurements on ties among these alters." Respectively, data of this kind of networks are obviously relational but limited, as links from each other are measured only to some alters. The measurements of these networks include size, density, centrality, strong or weak tie, bridge, etc.
2. Dyadic level. Dyad, originating in sociology, is a group of two people. Research at this level refers to one-mode dyadic networks about two single actors from one set of actors or two-mode dyadic networks about two actors from different set of actors [7]. The measurement of these networks is mainly about studying the structure of relationship, tendencies toward mutuality, etc.

3. Triadic level. Triadic relationship consists of one more individual than dyads. The measurements of such networks include balance, transitivity, and tendencies to reciprocity or mutuality.

4. Subset level is mainly about relation among a small subset of the network. The measurements of research are distance, reachability, cliques, and so on.

2.1.2.2 Meso-level

The population size of meso-level research ranges between micro-level and macro-level. Interestingly, meso-level research usually focuses on revealing the connections between micro- and macro- levels. It can be categorized as follows.

1. Organizations. Research on organizations focuses mainly on analyzing either intra-organizational or inter-organizational ties among actors of a social group sharing a collective goal.

2. Randomly distributed networks. Random networks have been studied since the 1950s, when Erdős and Rényi [128] and Gilbert [129] independently defined such networks. In 1980s, exponential random graph models of social networks became state-of-the-art methods in SNA [105]-[107]. The models are generally used for representing social structures commonly observed in many human social networks. More details are introduced when the specific model is described.

3. Scale-free networks. Scale-free network model refers to networks whose degree distribution follows a power law. This model framework is explained in detail in the following section.

2.1.2.3 Macro-level

Rather than tracing interpersonal interactions, macro-level analyses generally trace the outcomes of interactions, such as economic or other resource transfer interactions over a large population [11], [37]. Research at this level can be categorized as follows:

1. Large-scale networks. Generally, these networks are used for exploring the research problems in social and behavioral sciences and economics.
2. Complex networks. Complex networks provide an alternative prospective of SNA by emphasizing social complexity. This is different from the linearization relationships scholars used to exploring, which refers to the complex connections between elements that are neither purely regular nor purely random.

2.1.3 Network Topologies

Graph theory has its origin in the work of Euler in the eighteenth century [4]. The early work is mainly about solving problems of "small graphs with a high degree of regularity" [36].

Graphs, as defined by graph theory, are commonly used to represent individuals and other social actors by points and to represent their social relations by lines. Graphs represent the essential topological properties of a network by treating the network as a collection of nodes and edges [4]. A graph is defined as a pair of sets $G = \{P, E\}$, where P is a set of N nodes (or vertices and points) and E is a set of edges connecting two elements of P [36]. Defined more formally by Kolaczyk [102], a model for a network graph refers to one of the collection:

$$\left\{P_\theta(G), G \in g : \theta \in \Theta\right\} \tag{8}$$

Here, g is a collection of possible graphs, P_θ is a probability distribution on g, θ indicates a vertex of parameters in the range of Θ The richness of network graph greatly derives from how to specify $P(\cdot)$.

In the 1960s, two mathematicians, Erdős and Rényi [128] introduced the random graphs theory after Erdős discovered that applying probabilistic approach is useful in solving problems in graph theory. The study about random graphs has led to ideas very similar to those of statistical physics, and scholars believed that random networks could serve as the representations of many real world complex networks.

In practice, network models have been used for dealing with different problems, such as, exploring the mechanisms for generating commonly observed properties in the real world networks or testing the importance of certain target characteristics given a network graph.

In the following part of this section, three well-known and well-accepted network models are introduced: Erdos-Renyi model, Watts Model and Barabasi-Albert model.

2.1.3.1 The Erdős-Rényi Model (ER model)

The Erdős-Rényi network model is one of the classical random network graph models. From the mathematical perspective, random graph models are arguably taken as the most developed class of network graph models.

In Erdős and Rényi's classic article on random graphs, they defined the random graph as N labeled nodes connected by a number of edges, *m*. *m* is a random number chosen from $N(N-1)/2$ possible edges [128]. The ER model constructs a family of random graphs: indicated by $G_{N,p}$, namely the size of the network *N*, and the probability *p* that each pair of nodes is connected. In a summary, let $G(N,m)$ indicate a graph with N nodes and m links connected with the nodes, and it appears with a probability as follows,

$$P(G(N,m)) = p^{m}(1-p)^{M-m}, \text{ where } M = \frac{N(N-1)}{2} \tag{9}$$

The graphs of $G_{N,p}$ are also called Poisson random graphs, since in the limit of large N the binomial degree distribution converges to a Poisson distribution [11], [130], [131]. It is possible to create random graphs that have different degree distribution as summarized by Pacheco and Evans [132] and called the configuration model. The configuration model is a popular method for building up such kind of networks; it consists of: a) choosing a degree distribution; b) drawing a sequence of degrees from that distribution; and c) randomly connecting pairs of nodes.

2.1.3.2 The Watts-Strogatz Model (WS model)

The Small-World model generated by Watts and Strogatz is arguably one of the important innovations in modern network graph modeling [41], [42]. Explicitly, such types of models are designed for fulfilling the properties observed from the real world networks, often through the incorporation of the simple mechanism. The explorations of these models are usually involved with the relevant topic of communication, in the term of general and broad sense. Averagely, the structure of the small world networks brings the benefit of transmitting information quickly based on the 'neighbor to neighbor' exchanging feature. Small world network models have been used as the context of the spread of news, gossip, rumors, and even infectious disease.

Small world phenomena observed from real world refers to the networks featured by a small world characteristic that the networks (random graph) have short average path length. As demonstrated by many empirical studies, the real world networks own a small world characteristic, meanwhile, they have unusually large clustering coefficient [36], [131]. Low dimensional small world networks have been learned and tested for a long time by scholars, and many features of networks are understood, such as, clustering coefficient is independent of network size. However, all the network models share the combination of high clustering coefficient and large path length. The first successful attempt to generate the network models with high clustering coefficient and small path length which was close to the features of real world networks was those derived by Watts and Strogatz [41].

As emphasized by Watts [43], "their identification of a universal class of networks; that is, a family of networks that share certain aggregate properties regardless many of their individual details." Watts and Strogatz [41] significantly contribute to the literature of network analysis by making several related but distinct points of small world networks, Fig. 1(a) illustrates the rewiring process of WS model when p locate in [0, 1], and (b) indicate the changing process regarding to different p values of these two characteristics: path length ($L(p)$) and clustering coefficient ($C(p)$) for WS model.

1. Real world networks are neither completely ordered nor completely random, but display the distinct properties of both of them. "Order" refers to a uniform one-dimensional lattice. Each node in the lattice is connected with its k nearest neighbors. "Randomness" here refers to the variable parameter p (rewiring probability) that specifies the probability of randomly rewired links.

   Watts and Strogatz [41] quantified these properties with "simple statistics". The average shortest path length L defined to measure the average number of edges that must be traversed in the shortest path between any two pairs of vertices in the networks [43].

2. Clustering coefficient C defined to measure local density, which is the cliquishness of a typical neighborhood [41].

3. Quantified by Watts and Strogatz [41], when $p = 0$, completely ordered, the network is "large" and "highly clustered". When $p = 1$, completely random, the network is "small" and "poorly clustered". These statistics indicate "path lengths are short only when clustering is low" [43]. Interestingly, the model outputs two correlation relationships that the property of clustering is high relative to the randomness, conversely to it, average shortest path length significantly decrease when randomness is increased.

4. Since the requirements to the candidates belonging to "small world networks" for any networks are "relatively weak", Watts and Strogatz [41] predict that there are many real world networks that can be regarded as "small world networks". They claim that this network structure could have "dramatic implication to the collective dynamics of a system."

Arguably, based on his observation and analyses of the small world experiments by Travers and Milgram [133], Kleinberg [134] pointed out that social networks not only have the property of "small", but also have the "searchable" property. When links are uniformly random rewired, the Watts and Strogatz model, proved by Kleinberg [134], displays a short global path length, but does not display "searchability." He thus proposed "a class of generalized small world networks comprising an underlying *d*-dimensional lattice, and random links superposed on lattice, where the probability p of being connected randomly is $p \propto r^{-\gamma}$ [134], where r indicates the lattice distance from one node to the target node. He proved that only when $\gamma = d$ (the dimensions of lattice), the network would be "small" and "searchable". The arguments of Kleinberg suggest that the network structure is important both from the local view, and from the global view in the sense of "searchability" of the information or resources of individuals.

By far, the properties of small world networks have been deeply researched. Respectively, there are formulas for measuring and testing the properties, such as degree distribution, average path length, etc. More details are not given here and it is left to the reader to investigate them further.

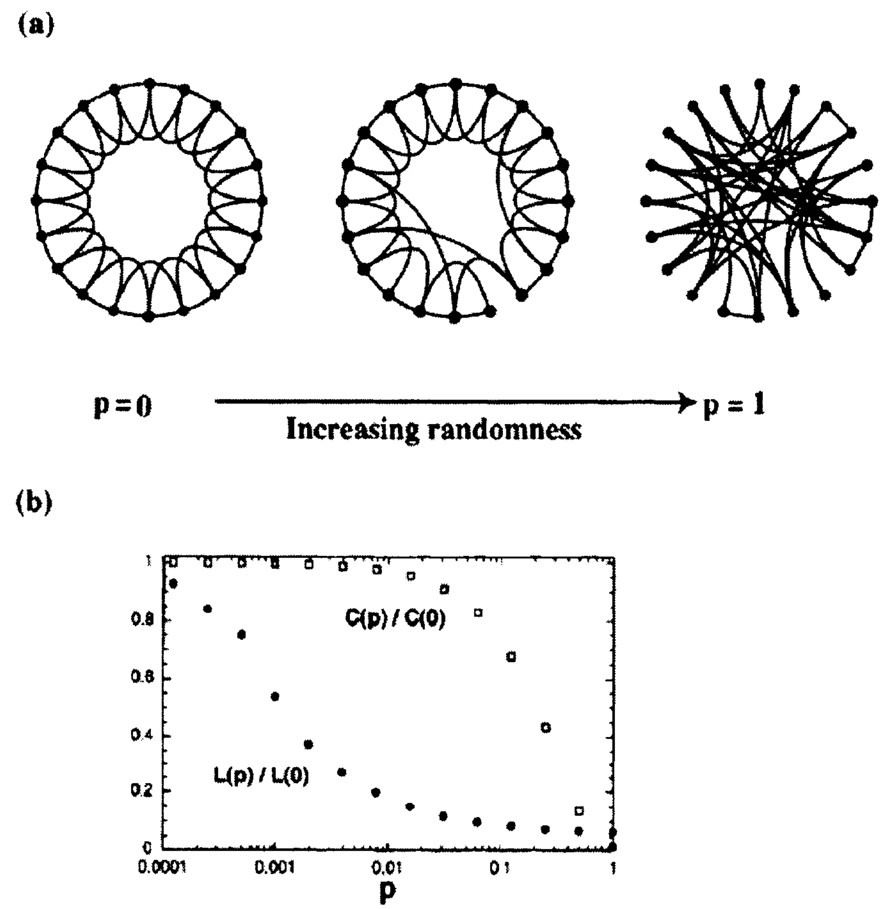


Fig. 1. The random rewiring procedure of Watts-Strogatz model. $N = 20$, $k = 4$. As p increases the network becomes randomly disordered. (b) Characteristic path length ($L(p)$) and clustering coefficient ($C(p)$) for WS model. Two statistics are normalized respectively by $L(0)$ and $C(0)$for a regular lattice [41].

Besides the well-known Watts small world model, there are other small world models as well. As noted by Kolaczyk and Csardi [120], some of them are the variations of the Watts model and have the amenable analytical calculation. Another popular one is the model where rewired edges are removed and a small number of new edges are added to randomly selected pairs of vertices [135]. In terms of simulation, it is argued that these models could be generated straightforwardly. There are some other studies that have generated the small world properties using different algorithms as well. For example, Kasturirangan [120] proposes a model, which starts from a one-dimensional lattice, then a small number of vertices are added to the center and each of them is connected with a large number of vertices of the original lattice; Kleinberg [134] proposed a model, which starts with a two-dimensional lattice, then adds short-cuts between vertices with a probability varying inversely proportional to the distance between them.

2.1.3.3 The Barabasi-Albert Model (BA model)

Another type of network model is network growth model. Real-world networks keep growing in time. For instance, the World Wide Web, scientific citation networks and Internet links keep expanding the size of themselves. Commonly in these studies, a mechanism is particularly studied for how the network changes at a given time point and the focus could be vertex preference, copying, age and so on. The goal of this series of studies is to seek what properties emerge from the networks given a limited number of nodes and certain time scales. It could be a reasonable explanation to some specific topic if such properties match those features observed in the real world networks. Among these many algorithms broadly discussed by the researchers, a preferential attachment model is one of the hottest discussing topics.

Preferential attachment refers as a mechanism that embodies a principle of the observed fact 'the rich get richer'. The driving motivation behind the introduction of this mechanism, argued by Kolaczyk [102], was a desire to generate the heavy-tail degree distribution that is usually observed in the real world networks, which is broadly mentioned as scale-free networks.

The empirical results demonstrate that many large networks are scale-free; that is, their degree distribution follows a power law for large k. In network analysis, there is a separate development that focuses studying on this degree distribution [43]. Different from the considering angle that is the modeling network topology of small world networks, scale-free networks stress a research angle of modeling the network assembly and evolution. Argued by Albert and Barabási [36], the goal of former models was to construct a graph with correct topological features, the scale-free networks model put the emphasis on capturing the network dynamics. Scholars believe that if the process that assembled the networks is captured correctly, the topology will be obtained correctly as well.

Barabási and Albert [35] first and successfully addressed the power law distribution of degree observed in the networks. The BA model embodied two mechanisms: population growth and preferential attachment. The former mechanism straightforwardly indicate that network expand itself as new members join in it. The latter mechanism illustrates that newly arriving nodes *a priori* choose to connect with hubs that are well connected rather than poorly connected nodes. The algorithm of the BA model is the following:

1. Growth: the network starts with a small number (m_0) of nodes, at each time step, a new node with m ($1 \leq m \leq m_0$) edges that link the new node to m different nodes already present in the system.

2. Preferential attachment: when choosing nodes to which the new node connects, the probability Π that a new node will be connected to node i depends on the degree k_i of node i, such that,

$$\Pi_{k_i} = \frac{k_i}{\Sigma_{j=1}^{N} k_j} \tag{10}$$

The network keep expanding itself follow these rules, after t time steps this procedure results in a network with $N = t + m_0$ nodes and m^*t edges. An important conclusion of this network model is that the degree of most nodes is below the average. Only few of them "have many times better connected than average" (Watts 2004). Barabási and Albert [35] employ "Power law" to describe the degree distribution. Power law has the asymptotic form: $p(k) \sim k^{-\alpha}$. Explained by Watts, the function means "the probability of a randomly chosen node having degree k decays like a power of k, where the exponentα, typically measured in the range (2, 3), determines the rate of decay." Summarized by Newman [131], although the BA model is elegant and simple when indicating some features of real world networks, it still lacks a number of features that are present in the real world networks, i.e., World Wide Web, such as:

1. The model is an undirected network model, where the real web is directed.
2. All vertices in the model belong to a single connected component. In the real Web, there are many separate components.

2.2 Limitations and Variants of Network Emergent Models

Emergent models like the ER model and the BA model have attracted an exceptional amount of attention in the literature. In addition to the analytic numerical studies of the model, many researchers have proposed the extensions or modifications of the model that alter its behavior or make it a more realistic presentation of processes taking place in the real-world networks. The relevant extension to the BA model will be discussed in the second part of this section.

This dissertation intends to expand the scale-free network, focusing particularly on the extension and modification models on the BA model. In this section, more details about scale-free networks and their formation mechanisms will be discussed.

There have been many extensions and variations on the classic BA model since its generation, including the undirected network model as the BA model and the directed network model. Introduced by Kolaczyk and Csardi [120], these concerns have been mainly discussed:

1. Do these models generate the power law degree distribution?
2. What is the algorithm if the power law degree distributed model is generated?
3. What are the parameters affect the power law exponent?
4. Many variations on changing format of the edges, addition and removal.

Since this study is originated from the BA model as well, more specific introductions to these extensions and variations will be discussed in the following section. Besides these mechanisms, another broadly discussed mechanism is the copying mechanism, which generates power law degree distribution as well. The distinct feature of copying mechanism is that it is often referred in the biological networks. Copying, here, refers to a normal observed tendency of reusing biological information when genomes of live organisms are evolving.

As discussed above, Barabási and Albert [35] argued that the scale-free nature of real networks is rooted in two generic mechanisms share by many real networks. They found that earlier network models all assume that the network models started with a limited number of vertices and then randomly linked or rewired. However, a majority of real world networks are open systems whose size is growing over time. A convenient example is the network of World Wide Web, which grows exponentially because of the addition of new web pages.

Second, Barabási and Albert [35] found that earlier models assume that the probability that two nodes are connected is independent of the nodes' degree. In contrast, Barabási and Albert [35] argue that most of the real world networks exhibit preferential attachment; that is, the connections between nodes are strongly correlated to the edges of nodes. For example, they illustrated that a web page will more likely include hyperlinks to popular documents with already high degrees [36]. This is because such highly connected documents are more likely to be searched and thus making the website well known. The above-discussed two mechanisms—summarized as growth and preferential attachment mechanisms—inspired the introduction of the BA model. This was the first type of network model that had the power law degree distribution. At this point, it is obvious that growth and preferential attachment play the key role in the network development. However, are both of them necessary for the emergence of power law scaling?

Barabási and Albert [35] investigated two models that contain one of two mechanisms to address the question above.

The first model keeps the growing character of the network but without the preferential attachment. The network model starts with a small number of nodes m_0, a new node with m ($m \le m_0$). Without the assumption of preferential attachment, the authors assume that the new node connects with equal probability to the nodes in the system (i.e., random attachment); that is,

$$\Pi(k_i) = \frac{1}{m_0 + t - 1} \tag{11}$$

This probability is independent of k_i. The continuum theory predicts that that k_i follows logarithmic time dependence, and for time $t \to \infty$, the degree distribution decays exponentially. The exponential character of the distribution indicates that the absence of preferential attachment eliminates the scale-free character of such network. The second model starts with N nodes and no edges. A node is selected randomly and connected with the probability $\Pi(k_i) = k_i / \Sigma_j k_j$ to a node i in the network. The model eliminates the growth process, and the number of nodes is constant during the network evolution. Their simulation results indicate that the model exhibits power law scaling at the early time, but $p(k)$ is not stationary. At the final stage, the system reaches a state in which all nodes in this network connected. Hence, they conclude that both growth and preferential attachment are the necessary conditions for formation of scale-free scaling. In the following literature, scholars extend the scale-free networks specifically by focusing on one of the formation mechanisms or both. In the following part of this section, this dissertation will review some recent works based upon the BA model. The discussion has been split into four themes relating to (2.2.1) preferential attachment, (2.2.2) dynamic growth, (2.2.3) edges and (2.2.4) growth constraints.

2.2.1 Preferential Attachment ($\Pi(k_i)$)

The BA model assumes that the probability $\Pi(k_i)$ that a node attaches to node i is proportional to the degree k of node i. The assumption involves two hypotheses: (1) $\Pi(k_i)$ depends on k; (2) the functional form of $\Pi(k_i)$ is linear in k. Barabási and Albert [35] stress that the studies have demonstrated that the degree distribution strongly depends on $\Pi(k_i)$ Hence, the precision of model of $\Pi(k_i)$ is fairly important.

Some researchers like Jeong et al. [136], Newman [81], Pastor-Satorras et al. [137] estimate the functional form of $\Pi(k_i)$ through measuring the real network data, such as co-authorship network, the citation network of articles, the actor collaboration network, and the Internet at the domain level. They consider that functional form of $\Pi(k_i)$ can be determined for networks for which the time each node joined the network is known. The process of capturing $\Pi(k_i)$ can be explained as follows:

1. Record the number of "old" nodes present in the network and their degree.
2. Measure the increase in the degree of the "old" nodes overtime interval ΔT, which is much shorter comparing to the age of the network.
3. According to the equation, $\Pi(k_i) = k_i / \Sigma_j k_j$, plotting the relative increase $\Delta k_i / \Delta k$ as a function of the earlier degree k_i for every node gives the $\Pi(k)$ function, where Δk is the number of edges added to the network in the time ΔT.

Finally, the obtained $\Pi(k_i)$ supports the existence of preferential attachment. Meanwhile, it appears that in each case $\Pi(k_i)$ follows a power law: $\Pi(k) \sim k^{-\alpha}$. In some cases, such as the Internet [136], the epidemic [137], the citation network [136], Medline, and the Los Alamos archive [138], We have $\alpha \cong 1$, i.e., $\Pi(k)$ that depends linearly on k as assumed in the BA model. For other networks the dependence is sub-linear, with $\alpha = 0.8 \pm 0.1$ for the neuroscience co-authorship and the actor collaboration networks [136].

Krapivsky et al. [139] proposed the effect of a nonlinear $\Pi(k_i)$ on network dynamics and topology. Instead of applying the equation $\Pi(k_i) = k_i / \Sigma_j k_j$ for adding new nodes, they calculate the average number of nodes $N_k(t)$ with $(k-1)$ incoming edges at time t. The time evolution of $N_k(t)$ follows (*rate equation approach*):

$$\frac{dN_k}{dt} = \frac{1}{M_\alpha}\left\{(k-1)^\alpha N_{k-1} - k^\alpha N_k\right\} + \delta_{k1} \tag{12}$$

where $M_\alpha(t) = \Sigma k^\alpha N_k(t)$ is the αth moment of $N_k(t)$. The first term accounts for the process in which a site with $(k-1)$ links is connected to the new site, thus increasing their degree to k. This happens with probability $(k-1)^\alpha / M_\alpha$. $k^\alpha N_k$ describes new nodes connecting to nodes with k edges, turning them into nodes with $(k+1)$ edges and hence decreasing the number of nodes with k edges. δ_{k1} indicates that the continuous introduction of new nodes with a single outgoing edge. According to the distinct α value, they identified the two different cases: the sub-linear case and the super-linear preferential attachment and conclude that the scale-free nature of the network is destroyed for nonlinear preferential attachment.

Dorogovtsev et al. [62] proposed an "attractiveness" model. At each time step a new site appears. Simultaneously, m new directed links coming out from no specified sites are introduced.

Let the connectivity q_s be the number of incoming links to a site s, i.e., to a site added at time s. The new links are distributed between sites according to the following rule: The probability that a new link points to a given site s is proportional to the following characteristic of the site:

$$A_s = A + q_s \tag{13}$$

Thereafter it may be called "attractiveness". All sites are born with some initial attractiveness $A \geq 0$, but afterwards it increases because of the q_s term. Parameter A, the initial attractiveness, determines the probability for "young" sites to get new links. They stress that they do not specify site from which the new links come out. The links may come out from the new site, from old sites, or even from outside of the network. The results have no relationship with the outgoing links setting. Hence, the model that they consider is equivalent to the BA model in particular case of the initial attractiveness (A) equal to m. In a summary, the model they defined is describing the likelihood that an isolated node will be discovered, such as a new article being cited the first time. The calculations indicate that the degree distribution follows $P(k) \sim k^{-\gamma}$ with $\gamma = 2 + A/m$.

Consequently initial attractiveness does not destroy scale-free nature of the degree distribution; it only changes the degree exponent.

2.2.2 Dynamic Growth

The BA model assumes that the number of nodes and edges increases linearly in time, and consequently the average degree of the network is constant. Scholars investigated the effect of nonlinear growth rates on the network dynamics and topology as well.

There are many real world networks growth patterns are measured by researchers. For instance, the average degree of the internet in November of 1997 was 3.42, but it increased to 3.96 by December of 1998 [140]. Similarly, the World Wide Web has increased its average degree from 7.22 to 7.86 in the five months between the measurements of Broder et al. [141]. The average degree of the co-authorship network of scientists has been found to continuously increase over an eight-year period [142]. Finally, comparison of metabolic networks of organisms of different sizes indicates that the average degree of the substrates increases approximately linearly with the number of substrates involved in the metabolism [136]. The increase of the average degree indicates that in many real systems the number of edges increases faster than the number of nodes, supporting the presence of a phenomenon called *accelerated growth.*

Dorogovtsev and Mendes [130] studied the effects of accelerated growth on the degree distribution. In their model, at every step a new node is added to the network, which receives n incoming edges from random nodes in the system. Additionally the amount of new edges is distributed, each of them being directed from a randomly selected node to a node with high incoming degree, with an asymptotically linear preferential attachment $\Pi(k_{in}) \propto A + k_{in}$ The authors show that accelerated growth, controlled by the exponent α, does not change the scale-free nature of the degree distribution, but it modifies the degree exponent.

2.2.3 Internal Edges, Rewiring and Edge Removal

One argument against BA modes is that the BA model only incorporates one mechanism for network growth: new nodes are kept connected with the aging nodes in the system. However, in the real world, a series of microscopic events shape network evolution, including the addition to or rewiring of new edges or the removal of nodes or edges. Several models have been proposed to investigate the effect of selected processes on the scale-free nature of the degree distribution, offering a more realistic description of various real networks. Any local change in the network topology can be obtained through a combination of four elementary processes: addition or removal of a node and addition or removal of an edge. But in reality these events come jointly.

An extended model by Albert and Barabasi [143] incorporates new edges between existing nodes and rewiring of edges. The model algorithm is starting withm_0 isolated nodes, each time step the model perform one of the following three operations:

1. With probability p, m $(m \leq m_0)$ new links are added:

   For this a node is randomly selected as the starting point of the new link, describing, for example, that a web developer decides to add a new hyperlink to a page. The other end of the link is selected with probability,

$$\Pi_{k_i} = \frac{k_i + 1}{\Sigma_j (k_j + 1)} \tag{14}$$

incorporating the fact that new links preferentially point to popular nodes, with a high number of connections. This process is repeated *m* times.

2. With probability q, m links are rewired:

   For node i and the link l_{ij} connected to it are randomly selected. Next, this link is removed and replaced with a new link $l_{ij'}$ that connects i with node j' chosen with probability $\Pi(k_j')$ given by equation above. This process is repeated m times.

3. With probability $(1-p-q)$, a new node is added:

   The new node has m new links that with probability Π_{k_i} are connected to nodes i already present in the system.

They conclude through the results of this model that in critical phenomena power law scaling is typically associated with universality, implying that the exponents are independent of the microscopic details of the model. The model demonstrated that no such universality exists for scale-free networks, the scaling exponents depending continuously on the network's parameters. On the other hand, results indicate the existence of a different criterion for universality based on the functional form of $P(k)$. Their model predicts the existence of two regimes, the scale-free and the exponential regime. Some of the large networks investigated so far, such as the www or the actor networks, are described by scale-free networks. However, a number of fundamental network models lead to a $P(k)$ that decays exponentially, indicating the usefulness of the exponential regime as well.

2.2.4 Growth Constraints

Scholars argue that nodes have a limited lifetime or a finite edge capacity (internet routers), thus they hsbr addressed some discussions on the topic.

Amaral et al. [144] suggested that while several networks do show deviations from the power law behavior, they are far from being random networks. For example, the degree distribution of the electric power grid of southern California is more consistent with a single-scale exponential distribution. Other networks, like the extended actor collaboration network, in which TV films and series are included, have a degree distribution in which power law scaling is followed by an exponential cutoff for a large k. In all these examples, there are constraints limiting the addition of new edges; for example, the actors have a finite active period during which they are able to collect new edges, while for the electrical power grid or neural networks there are constraints on the total number of edges a particular node can have, driven by economic, physical, or evolutionary reasons.

Amaral et al. [144] propose that in order to explain these deviations from a pure power law one must incorporate aging and cost or capacity constraints. The model studied evolves following growth and preferential attachment, but when a node reaches a certain age (aging) or has more than a critical number of edges (capacity constraints), new edges cannot connect to it. In both cases numerical simulations indicate that while for small k the degree distribution still follows a power law, for large k an exponential cutoff develops.

Dorogovtsev and Mendes [130] propose that in some systems the probability that a new node connects to a node *i* is not only proportional to the degree k_i of node *i*, but it also depends on its age. Papers or actors gradually lose their ability to attract more edges, the model assumes that this phase-out follows a power law. The calculations predict that the degree distribution depends on the exponent: power law scaling is present only for $\gamma < 1$, and the degree exponent depends on γ. When $\gamma > 1$, power law scaling completely disappears, the degree distribution approaching an exponential.

2.3 Conclusion and Discussion

After reviewing key network descriptive statistics and network formation models, this chapter helps lay out the methodological and technical foundations for this dissertation. Moreover, a review of the relevant literature suggests that the current network formation models are inherently limited in revealing the impacts of agent heterogeneity in shaping network formation. Particularly, the BA model assumes that nodes with more links (i.e., "popular nodes") are more likely to be connected when new nodes enter a system. More interesting, the BA model found that, as mentioned before, preferential attachment in a growing network leads to a power law degree distribution.

However, this line of research is problematic since it assumes all the nodes possess the same preference (instrumental preferential attachment) and overlooks the potential impacts of agent heterogeneity on network formation (intrinsic preferential attachment). When joining a real social network, people are not only driven by the instrumental calculation of connecting with the popular, but also motivated by the intrinsic affection of joining like-minded individuals. In other words, people are constantly weighing between popularity and proclivity in forming their social connections. The impact of this mixed preferential attachment is particularly consequential on such social networks as political communication.

In the following chapter, this dissertation proposes an integrative agent-based model of preferential attachment encompassing both instrumental calculation and intrinsic similarity. Particularly, it emphasizes the ways in which agent-heterogeneity affects social network formation. This integrative approach can strongly advance our understanding about the formation of various networks.

CHAPTER 3

METHODOLOGY: AGENT-BASED MODELING

From the previous chapters, it is evident that a key difficulty in modeling network formation is that fairly simple specifications at the micro-level, even with minimal heterogeneity, can give rise to complex networks at the macro-level. Although there are many insights that one can derive analytically or induce empirically, there are still many network dynamics that cannot be seen so directly. By examining the behavior of large computer-simulated societies is one possible way to try to observe this unseen behavior. Agent-based modeling provides one approach to generating this simulated society [148].

However, as with any form of analysis, there are important considerations in terms of how social networks should be formalized in agent-based modeling. This chapter, therefore, is dedicated to a review of extent studies on agent-based modeling and social network, a discussion of the advantages associated with agent-based modeling in simulating network formation, and, finally, an explanation of operationalization of network formation based heterogeneous attachment.

3.1 Literature Review: An Agent-based Approach to Social Network Analysis

This section intends to address why ABM approach is suitable to SNA. Specifically it addresses the question from two angles: the theoretical and the analytical. Theoretically speaking, ABM is desirable because of its emphasis on complexity and emergence, which are the regarding features of social networks as well. Analytically speaking, ABM is preferably because it is computationally advantageous in dealing with complex systems.

The importance of social networks analysis rests on three underlying assumptions about patterned relations and their effects [4], [7]. First, structural relations are more important for understanding observed behaviors than attributes of individuals. Second, social networks affect perceptions, beliefs, and actions through a variety of structural mechanisms that are socially constructed by relations among entities. The third assumption is that structural relations should be viewed as dynamic processes. Guided by these three assumptions, scholars have focused on different aspects of SNA. In recent decades, as the development of computational technology, there has been an explosion of such network analysis research [5], [25].

Up to this point, it has been found that although scholars have made important progress in SNA, the dominant guiding paradigm is unfortunately structure-oriented; that is, researchers focus mainly on systematic attributes of "regular" networks [4], [7], [8]. In this line of research, large-scale social networks are commonly regarded as complicated networks displaying certain variant attributes from the regular ones. Moreover, these studies have paid only limited attention to the emergent nature of network formation; that is, while isolated connection building between individual agents (or nodes, vertexes) by no means can exert any significant impact on the system, massive paralleling such interactions will lead to some fundamental systematic changes [37]. This study suggests the introduction of an agent-based approach can strongly enhance our understanding of SNA, particularly through its two anchoring concepts: complexity and emergence.

3.1.1 Complexity

A complex system, differing from complication common in a linear system, is composed of numbers of dependent elements [58]. These elements interact with each other through connections. The connections among them can be relatively simple and stable, or complicated and ever changing. Systems are formed by these elements along with their connections. In such systems, removing an element may destroy the overall system behavior, as well as the complexity [58]. The remarkable thing about the complexity in the social worlds is that the social agents must predict and react to the actions of other agents (adaptation). Varying connections among the adaptive agents exacerbate these activities when adaptive agents become close to one another. The outcome of such a system is that the relationships between social adaptive agents are highly nonlinear [58], so that the system is hard to decompose, and the complexity ensues.

3.1.2 Emergence

A key focus of ABM is that interacting automatous agents in the system lead to emergent phenomenon. Emergence can be regarded as a global phenomenon growing from localized, individual behavior. However, emergence, in most cases, disconnect with its details of local behavior. Consequently, emergent patterns still keep stable even if there are rational variations occurring in individual behavior [56]-[58].

Emergent behavior, in some cases, is considered as simply a reflection of scientific ignorance rather than some deeper underlying phenomenon. However, ignorance can drive the quest for understanding. Miller and Page [58], for instance, argue that two kinds of complexities can be well described by emergence: disorganized complexity and organized complexity. The feature of the former is that as the number of agents increases, individual variations begin to cancel one another out, and "system-wide" predictions become possible. It should be noted that, in this case, it is not observable how communication among agents and localized behavior aggregate into global phenomena that survive and last with the characteristics completely different than its components. Consequently, "explorations of complexity have begun to identify the emergent properties of interacting agents for want of a better term, organized complexity" [58]. In the case of organized complexity, agent variations no longer cancel one another out but, rather, become reinforcing.

A review of ABM research on networks reveals two main directions of research. While many researchers interested in seeking the process of network formation, many others are working on exploring the diffusion under different types of networks, i.e., the transmission process in the context of varying network topologies. For example, Hamill and Gilbert [49], in exploring network formation, argue that currently there is no such network models fit well with sociological observations of real social networks, and they provided an ABM model based upon the social circle theory. This is different agents with unequal social reach that create a wide variety of artificial social worlds labeled with properties of real world observed large scale networks. Similarly, Mitrović and Tadić [145] conducted an analysis of the empirical data and the ABM of the emotional behavior of users on Web portals where user interaction is mediated by posted comments. ABM here is used to simulate the dynamics and to capture the emergence of the emotional behaviors and communities. Gaston and des Jardins [146] provide the past findings of the structure of the artificial social network governing the agent interactions is strongly correlated with organizational performance. By the context of dynamic team formation, they proposed two strategies for agent-organized networks and evaluate their effectiveness for increasing organizational performance.

Up to this point, it is evident that social networks are complex, and information diffusion is an emergent phenomenon, which in turn suggests that an ABM approach is an appropriate tool for us to analyze the relationship between social network and information. Moreover, ABM is flexible, process-oriented, and adaptive. All of these entail important implications to future studies of social network [57].

3.2 Agent-Based Model of Heterogeneous Attachment

In retrospect, it is evident that traditional tools have significantly constrained the theorization of various social systems. Existing models are more about static, homogeneous system models composed of either several or infinite amount of agents, regulated in a world in which time and space matter little. In sharp contrast, computational techniques, a powerful tool arising from complex systems research, allow for a flexible number of heterogeneous agents to interact in a dynamic processing system subject to the limits of time and space. For at least two methodological reasons, an ABM approach is suitable for SNA; these reasons will be discussed in turn.

First, while some scholars argue that the computation methods are the outcome of the confusion between theory and the tools used to develop theories, it is reasonable to argue that theories need to be judged by how well they improve our understanding about the world we are interact with, rather than the tools used to develop them [58]. As long as the facts of the world are discovered or partially uncovered, theories are constructed withstand testing and tracing. In reference to SNA, ABM is usable for uncovering the network formation motivating factors intrinsic necessity of agents. Second, while many argue that computational models can only provide inductive proof, it is clear that axiomatic proof guaranteeing its outcome comes at the cost of being willing to sufficiently narrow the problem domain. The actual problem, on the other hand, is under what conditions the outcome is guaranteed. While the conditions of guarantee are exceedingly elaborate, it is possible that social scientists have the willingness to accept some inexactness in predictions in return for more favorable circumstances.

Finally, caution should be taken when differentiating between the computation in theory and computation as theory. Computation in theory is mainly accomplished through abstracting the behavior of the individual agents in the system into simplified agents. The collections of these agent-based objects, then, will be "solved" by allowing the objects to interact with each other using computation [58]. In contrast, computation as theory is a bottom-up simulation for providing a constructive existence proof of some propositions or for experiments to see if a set of rules can imply "lifelike" behavior, and top-down simulations that use the computer as a way to understand the implications of this set of abstraction-based objects. Obviously, in reference to SNA employing ABM, we will go through a bottom-up simulation process.

The review of the SNA literature in this dissertation is inevitably far from being exhaustive and leaves many recent developments in network analysis unexamined. However, several important conclusions stand out. First and foremost, it is evident from the literature that social networks are complex, and information diffusion is an emergent phenomenon, which in turn suggests that an ABM approach is an appropriate, if not the best, tool for us to analyze the relationship between social network and information.

Moreover, as listed in the earlier section, ABM is flexible, process-oriented, and adaptive. All of these entail important implications to future studies of social networks. In fact, an increasing number of scholars has started to use ABM methods to simulate the formation of different networks. For instance, Hamill and Gilbert [49] argue that none of the standard network models fit well with sociological observations of real social networks, so they presented a simple structure for use in agent-based models of large social networks. Mitrović and Tadić [145] applied ABM to investigate the dynamical network of bloggers' communities. Lodhi et al. [147] proposed an agent-based network formation model for the Internet at the autonomous system level.

However, in general, ABM approach to network analysis only occupies a marginal place in the literature. Most scholars still rely on statistical methods to examine variant network structures, ranging from centrality of a given network or connectivity between vertexes (i.e., nodes). The emphasis on the static attributes of already existent networks, although helpful in providing a clearer description of given networks, inevitably downplays the relationship between network formation and network structure. If social networking is a complex adaptive system, its emergent process, rather than its ending status, can tell us more about its fundamental attributes. As argued by Epstein [148], "if you didn't grow it, you didn't explain its emergence." For these reasons, it is surprising to note that only a very limited number of scholars are applying ABM to network analysis.

Given these, an ABM approach to more complex social networks can strongly advance our understanding of network analysis. The research goal of this dissertation therefore is to expand one of the most influential network models in SNA; that is, the scale-free network. Myriad studies have been devoted to exploring the nature and origin of social networks [2], [4], [7], [8]. Despite the nuances and complexities documented in the literature, there are essentially two overarching theoretical models regarding motivations of individuals: the instrumental model and the intrinsic model. Individuals are motivated to select friends that are either instrumental or intrinsic. This translates into an individual utility function of social networking as follows:

$$u_i(j) = K_j + G_{ij}, \tag{15}$$

where $u_i(j)$ represents individual i's utility of linking with individual j, K_j denotes the instrumental gains delivered by j's popularity in the given social network, G_{ij} represents the intrinsic benefits from the similarity between i and j. With Eq. (15), we can find a static model of an individuals' total utility in a given network, say, a network with n agents, $N = \{1,2,\ldots,\mathrm{n}\}$. We then can let g^N represent the complete graph, where every player is connected to every other player, and let $\{g \mid g \subseteq g^N\}$ represent the set of all-possible such graphs. If agent i and j are linked in graph g, we write $ij \in g$. Therefore, for a unique static graph g, each agent $i \in \{1,2,\ldots,\mathrm{n}\}$ receives a payoff, $u_i(g)$,

$$u_i(g) = \sum_{j:ij \in g} K_i + \sum_{i \neq j} G_{ij} - m, \tag{16}$$

where, m represents the cost constraint in a particular mode of social networking and is assumed uniform across agents in this study. Of course, this simplification of agents' utility in a networked setting is hardly the complete story. How can this static formal statement help integrate substantive theories and abstract network model? And how can this static statement help construct a dynamic ABM model?

3.2.1 *K*, Network Position, and the Principle of Popularity

Two motivations that are of particular relevance to social networking are the need for cognition and the need for affect [71], [73], [75]. Need for cognition is fundamentally an instrumental concern and a stable disposition that explains individual differences in the tendency to use network to acquire such resources as information. Individuals who possess a high need for cognition have a strong motivation to connect with resourceful and well-informed people in a given social network. Consequently, individuals who are high in need for cognition are more likely to behave in line with the rational model. They should operate as if they possess a running tally of everyone's relative position in a network, which in turn helps make their decision of networking.

Yet how can individuals' running tally of network structure translate into abstract network models? Many works by sociologists have examined how individuals' instrumental concern can be realized in a network setting. One of the most influential such works is Granovetter's "Strength of Weak Ties" (SWT) theory [77]-[78], in which weak, bridging ties are argued to be beneficial to individuals because of their potential in introducing novel information. Burt [1] later refines the argument by emphasizing the relationship between "structural holes" and "information brokers." This, in turn, leads to Burt's conclusion that network brokers tend to enjoy various advantages because they can bridge otherwise isolated clusters. More recently, Podolny [79] argues that network structure matters not only because it serves as "pipes" of resources, but also because it acts like "prisms," revealing important information about the inherent qualities of vertices (e.g., credibility). An individual's status (i.e., popularity) in a given network provides others a heuristic shortcut in assessing his/her credibility. In sum, these studies suggest that individuals motivated by need for cognition are tending to be connected with people of high degrees in a network.

If individuals high in need for cognition are popularity-driven, what is its implication on network formation? Studies by statistical physicians on scale-free networks provide some clues. Yet, it should be noted that their concerns are mainly at the system level. As for large-scale complex networks, empirical results demonstrate that most of them are scale-free; that is, their degree distribution follows a power law distribution [80]-[82]. The BA model [35], [37] then is introduced to describe this scale-free emergent mechanism. The BA model suggests that the growth of network size and preferential attachment are the necessary conditions for the emergence of scale-free networks. In other words, a social network that is solely composed of individuals high in need for cognition tends to follow a power law distribution.

In many social networks, however, significant deviations from scale-free behavior have been reported. Although numerous variants of the BA model have been developed to reproduce the growth process of social networks, most of them still share the very key instrumental assumption that a vertex's probability to be connected is determined primarily by its position (i.e., "popularity") in a given network [60], [62]. The instrumental concern or the need for cognition, as elaborated above, is not the only motivation that governs individuals' social networking. The need for affect is another important motivation of individual behavior.

3.2.2 *G*, Agent Heterogeneity, and the Principle of Proclivity

The need for affect, in sharp contrast to the need for cognition, is fundamentally intrinsic. It is a separate motivational construct that captures the degree to which people enjoy experiencing strong emotions [70], [72], [75], [83]. Individuals who are high in need for affect are more likely to view emotions as useful when making various decisions. Because these individuals tend to enjoy experiencing strong emotions, their attitudes tend to possess a stronger affective basis. This is not to say that the attitudes of individuals who are high in need for cognition are unaffected by emotion. Almost all attitudes carry some affective component. Rather, the argument here is that individuals high in need for affect possess attitudes that carry a more intense affective charge, and while affect may induce "biased reasoning" in most people, individuals high in need for affect should be especially prone to biased processing [70]-[72].

What are the implications of this "biased reasoning" for social networking? Pujol et al. [84] have pointed out that the assumptions of BA models usually lack sociological grounding. Wong et al. [84] argued that many network models have not taken the advantages of sociological and psychological insights of how social networks may be formed. It is also problematic since it assumes all the nodes possess the same preference (instrumental preferential attachment) and overlooks the potential impacts of agent heterogeneity on network formation (intrinsic preferential attachment). When joining a real social network, people are not only driven by instrumental calculation of connecting with the popular, but also motivated by intrinsic affection of joining the like. In other words, people are constantly weighing between popularity and proclivity in forming their social connections. The impact of this mixed preferential attachment is particularly consequential on such social networks as political communication. More importantly, the support for this assumption comes from the social theory: homophily.

McPherson et al. [26] argue that homophily is the principle that a contact between similar people occurs at a higher rate than among dissimilar people, and the similarity could be regarding many types of personal characteristic positions, including gender, religion, social class, education, and other intra-personal or behavioral characteristics. In fact, there are some models taking homophily into consideration, somehow not using the specific term but essentially the similar meaning. Robins et al. [86] presented network models for social selection process. Although characteristic positions affecting the social relationship formation are concerned, it is broken between the local behavior and the global pattern. In other words, there is no analysis for the properties of large social networks. Newman and Girvan [87] conducted a network model and discuss the mechanism of assortative mixing, which is, the nodes with similar degree level like to link with each other. However, it actually is a special case of preferential attachment, where the similarity of nodes is concerned.

This study proposes an integrative model of preferential attachment encompassing both instrumental calculation and intrinsic similarity, which is a term transformed from homophily. Particularly, it emphasizes the ways in which both agent-heterogeneity and network position affects social network formation. Agent-based modeling is chosen as the paradigm to conduct this study. This integrative approach can strongly advance our understanding about the formation of social networks.

3.2.3 *m*, Mass Media, and Individual Capacity

Up to this point, it is clear that both the principle of popularity and the principle of proclivity are key strategies adopted by individuals in social networking. Yet, it should be noted that the actual impact of these two principles is contingent on the costs of acquiring information and establishing connections; that is, the constraints of communication modes.

Scholars have long found that the individual acquisition costs of information vary widely under different media settings. A key theme across many disciplines in social science is the question to what extent lower communications costs have make people more connected and less isolated [34], [88]. Recent years have seen an explosion of research on the topic. To some degree, the diffusion of ICT is at the heart of the latest surge of studies on social network. The development of the Internet has greatly increased the means by which people communicate with each other. In many respects, ICT appears closer to traditional media than to mass media in the sense that they are used as personal communication platforms.

A review of relevant literature suggests that competing views about the effects of technological innovation and its resulting impact on ordinary people provide the basis for a rich debate on the relationship between social isolation and the fragmentation of society. Whereas few doubt that technology has greatly expanded our capacity to connect with others, the impact of technology on the perception of being connected is more controversial. For instance, many argue that ICT reduces or counteracts the impact of geography on structuring opportunities for social interactions, and second, they take time away from other face-to-face activities [89].

Rather than joining the existing debate on the exact impacts of ICT, this study intends to explore to what extent the network formation is constrained by different modes of communication. An ABM approach to this problem helps elucidates the structural differences of social networks under different media setting.

3.3 Model Hypothesis and Assumptions

As stated in earlier sections, there are two critical assumptions underlying the BA model, i.e., the scale-free network. First, the network continuously expands with the addition of new vertices that are connected to the vertices already present in the system. Second, the probability that two vertices are connected is neither random nor uniform; instead, there is a higher probability that it will be linked to a vertex that already has a large number of connections.

At the agent-level, the scale-free network suggests that agents (i.e. vertexes) are instrumentally motivated to be connected to agents that are "popular." To incorporate this type of preferential attachment, we can assume that the probability Π that a new vertex $j+1$ will be connected to vertex i depends on the connectivity k_i of that vertex, so that

$$\Pi(k_i) = \frac{k_i}{\Sigma_j k_j}. \tag{17}$$

The BA model assumes that all the agents are homogenous, sharing the same instrumental preference of connecting with the popular. Although this homogenous instrumental preference attachment is useful in describing micro dynamics of many social networks, it is often compromised by agents' inherent proclivity.

To capture the impact of agent heterogeneity on network formation, this dissertation proposes two more assumptions. First, vertexes are intrinsically different from each other on certain aspects. A convenient example is party affiliation [4]. With respect to US political party orientation, people can be labeled as Democrats and Republicans. Second, the probability that two vertices are connected is not solely determined by the connectivity of the existing vertexes. It is also determined by the instinct similarity between vertexes. Specifically, there is a higher probability that a new vertex will be linked to a vertex that shares similar characteristics. In other words, a new Democrat, when entering a social network, is prone to forming a connection with another Democrat but not a Republican. At the agent level, this theory suggests that agents are intrinsically rational, which implies that they are motivated to select parties that are characteristically proximate. This in term can be translated into a distance between vertexes.

$$\mathrm{P}(C_{N+1}, C_i) = \frac{1}{N-1}\left[1 - \frac{(C_i - C_{N+1})^2}{\Sigma_j (C_j - C_{N+1})}\right], \tag{18}$$

where C_i represents the characteristic position of vertex i; C_{N+1} represents the characteristic position of a new vertex $N + 1$ that is about to join the existing network. A smaller Euclidean distance translates into more utility and hence contributes to the likelihood that the new vertex will be attached to vertex i. It should be noted,

$$\sum_i \mathrm{P}\left(C_{j,}C_i\right) = 1. \tag{19}$$

Therefore it is possible to characterize this as intrinsic (or heterogeneous) preferential attachment in contrast to Barabasi and Albert's instrumental preferential attachment [35].

This dissertation assumes that homophily is another key driven mechanism, comparing to the mechanism of preferential attachment, for leading the social agents to make decisions for forming the different structures of the social networks. When joining a real social network, people are not only driven by instrumental calculation of connecting with the popular, but also motivated by intrinsic affection of joining the like. In other words, people are constantly weighting between popularity and proclivity when forming their social connections.

In light of this, the probability U can be updated as a new vertex $N + 1$ will be connected to vertex i depends on the connectivity k_i of that vertex as follows,

$$U\left(\Pi, \mathrm{P}\right) = f\left(k_i, C_i, C_{N+1}\right). \tag{20}$$

In this way, this heterogeneous attachment model can be used to describe and guide the emergence of real world social networks. In this model, it is imperative to introduce a parameter of the relative weight of instrumental preference, *U* is a weighted product of Π, the instrumental preferential probability, and P, that intrinsic preferential probability. In light of this, the BA model is only a special case of this integrative model.

The impact of this mixed network formation is particularly consequential on such social networks as political communication. For instance, when people appear in a new community and start to build their network, the two endogenous driven mechanisms would lead people to build up their social networks. Under the extremely conditions, by following the preferential attachment, people are only interested in linking with the popular people, so that the people have bigger potential to expand their networks. By following homophily, people would be more like to connect with the other people that have similar intrinsic properties with them, since they might be looking for a more comfortable social ambience or they even could gain more confidence from people owning the similar characteristic. Certainly, the latter is a human behavior factor, which is labeled as "intrinsic" intention to construct network in this study. In the real world, most people make decisions based on both mechanisms in different levels instead of in the extreme situations. The following chapter discusses the way of weighting for balancing these two mechanisms and the method of modeling homophily.

3.2.1 Model Assumptions

The heterogeneous attachment model proposed in this study therefore is rested on the following three key assumptions:

1. **Heterogeneity**: Vertices (i.e., agents) are intrinsically different from each other on certain aspects. All of the relevant characteristics of vertices are captured by a finite set of $c \geq 1$ types: $\{1, 2, \ldots, c\}$. Based on this finite set of relevant characteristics, it is possible to construct C_i, representing the characteristic position of node i.

2. **Dynamic Growth**: The network continuously expands by the addition of new vertices. The network starts with a small number (n_0) of nodes, at each time step t, a new node with m edges that link the new node to m different nodes already present in the system.

3. **Heterogeneous Attachment**: The probability that two vertices are connected is jointly determined by the connectivity of the existing vertices and the intrinsic similarity between vertices. The joint probability that a new vertex at time step $t + 1$ will be connected to vertex i depends on,

$$U = f\left(k_i, C_i, C_{t+1}\right) = \frac{\lambda k_{it}}{\Sigma_j k_{jt}} + (1 - \lambda) \cdot \delta\left(C_i, C_{t+1}\right), \tag{21}$$

where λ is a weighted product of the instrumental preferential probability and intrinsic preferential probability. Connectivity of node i at time step t thus is k_{it}. The probability of a new node and a random existing node are connected for intrinsic purpose at time $t + 1$ can be captured by $\delta(.)$, in which $\delta\left(C_i, C_{t+1}\right)$ decreases as $C_{diff}\left(C_i, C_j\right)$ increases for $i \neq j$. C_{t+1} is the characteristic position of a new node entering the network at time step $t + 1$, and C_i is that of node i already in the network, $i \in \{1,2,\ldots,t\}$.

4. **Media mode**: The global constraint is captured by agents' overall capacity in connecting, m.

Though Eq. (22) represents a generic means of determining the probability of attachment based on distance between two agents, a simplified version is used in this model because only the two types case is considered. In the original setting of the simulation in this dissertation, let $C_i \in \{\text{"Blue"},\text{"Red"}\}$, and,

$$\delta(C_{t+1},C_i)=\begin{cases}\dfrac{1}{\mu N_d+N_s} & \text{for } C_{t+1}=C_i\\ \dfrac{\mu}{\mu N_d+N_s} & \text{for } C_{t+1}\neq C_i\end{cases} \tag{22}$$

where N_s is the number of nodes $C_{t+1}=C_i$, N_d is the number of nodes $C_{t+1}\neq C_i$. Given there is a binary difference on the agents' characteristic position, and the above function provides a better performance than the Euclidean distance illustrated in Eq. (22).

Together, the above four assumptions suggest that each new node makes m new edges to remain in the network. Rather than solely attracted to the popular nodes, new nodes also weigh the extent to which other nodes are similar to themselves. The relative weights of this mixed preference are captured by λ. It is apparent that, $\lambda = 1$ will give rise to a purely rationality-driven social network (which are classical scale-free networks), and $\lambda = 0$ will result in a value-driven social network. Also note that replacement is assumed for the case when incoming nodes make more than one new edge. This assumption is made for simplicity. In the simulation, if a node is attached twice the second attachment is ignored and a new attachment is made.

3.4 Operationalization in NetLogo

In order to illustrate and model heterogeneous attachment, this study uses different colors to denote the different attributes of agents as stated in Hypothesis 1. Specifically, for the purpose of simplicity there are two types of agents in the system: blue agents and red agents. Assumption 2 is modeled by allowing the size of agents to increase at each tick, a default time unit in NetLogo. It should be noted that this study focuses primarily on the topology of social networks obtained on the final stage. Therefore, we assume simple dynamic growth of agents in this model. The network will stop growing when there are 10,000 agents in the network. As implied in Eq. (21) of Assumption 3, when $\lambda = 1$ a purely rationality-driven social network (i.e., a classical scale-free network) is expected to emerge, and when $\lambda = 0$, a value-driven social network is expected to be generated. Agents in the model would take the same color agents into their homophily consideration. The same process will repeat for 30 times for each λ, $\lambda \in \{0, 0.25, 0.5, 0.75, 1\}$.

More specifically, the simulation process can be described as the following steps:

1. Start with m + 1 connected nodes with random values (red/blue color).
2. A new node with a random value (red/blue color) wants to join the networks.
3. The nodes existent in network (for first run, the original m nodes) are in the choosing queue.

4. The incoming node weighs between popularity and proclivity. It chooses m node(s) from the queue to connect by calculating the probability given in Eq. (21). The node with higher probability in the existing network is more likely to be connected.
5. Loop from step 3.
6. Each run of simulation stops when there are 10,000 nodes in the network.
7. The simulation stops when the model runs 30 times for each $\lambda \in \{0, 0.25, 0.5, 0.75, 1\}$, and $m \in \{1, 2, 3\}$.

The following schematic example is for translating the model into real world instance. A new person living in a community has to building up a network relationship with other ones in the same community. At the point of determining which person he or she wants to link with, he or she is concerned with a balance between two features of other people: "Is he or she popular?" and "Is he or she a person more like me?" Relying on the result of his or her balancing, he or she makes a choice and links himself or herself in this community network. After joining, he or she will be evaluated by others if there are new persons joining to this community. The model outcome proves that super hubs are generated if people only care about "I want to link myself to a more popular people, in other words, people with more resources". That is, the the BA model is generated for the special case for $\lambda = 1$.

The main goal of this piece of work is to study how the intention from people affecting the structure of scale-free networks. Theoretically, network formation processes are affected by adding more heterogeneity to automatous agents. Based on the modified model introduced in the section, the network model is simulated respective to different values of people's intention rate "λ", where "λ" indicates the objective choice from people between linking to the person with more links or the one "more like me". The λ variable has five different values and they are listed and explained in Table 1.

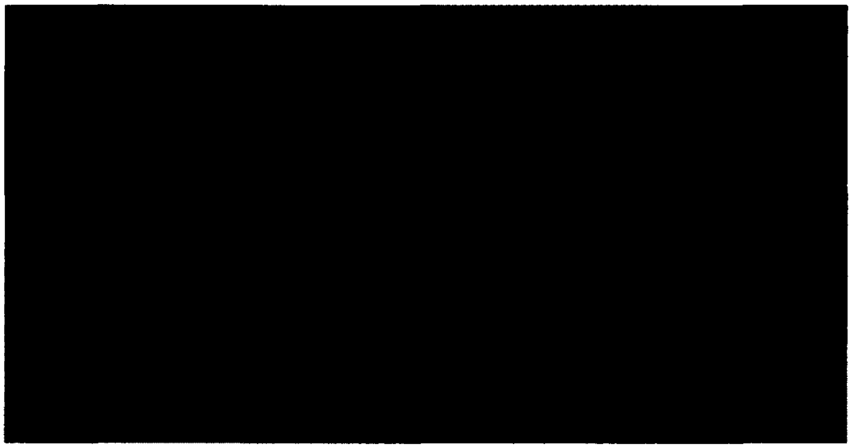
Table 1

THE SUBSTANTIAL MEANINGS OF λ

λ values	Categories
0	People are only concerned about "Are you more like me?" or "Are we in the same party?"
0.25	People are concerned more about "Are you on my side", but also care about "Do you have more links?" a little bit.
0.5	People are concerned about these two parameters equally.
0.75	People are more concerned about "Do you have more links?", but also care about "Are you on my side?" a little bit.
1	People only care about "How many links do you have?"

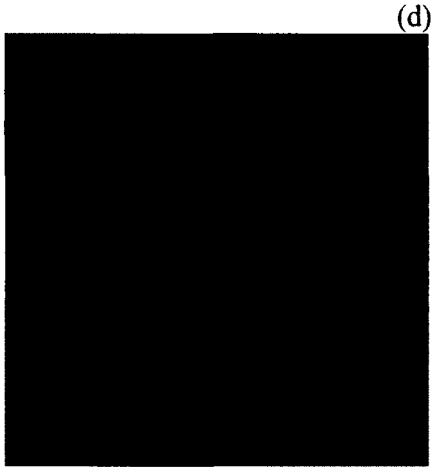
In summary, nodes in Netlogo are colored in red or blue to denote the difference or similarity between each other. The initial condition of the model starts with m nodes and these $m + 1$ nodes are connected with each other arbitrarily, without considering the probability calculating. After the initial stage, one node with a random color is generated in each time step and it decides to link with m existing nodes based on the proclivity and the popularity probability calculating. There are no separated nodes or networks generated accordingly. The simulation will be stopped when there are 10,000 nodes in the network and it would be run 30 times under each value setting of m and λ. The simulation will be run for 450 times in total. The network degree data would be outputted and used for data analysis.

The network structures are generated according to different intention rate values based on the integrative model. Illustrated in Fig. 2, as we can observe from the network structures of each graph, the network structures demonstrate that they are getting more clustering as the value of λ increases, and the size of the hubs are getting bigger. Explaining it in another way, the power (or resource) is distributed unevenly. For clarity reason, only 502 nodes are shown in these graphs.



(a) $\lambda = 0$ (b) $\lambda = 0.25$

(c) $\lambda = 0.5$ (d) $\lambda = 0.75$



(e) $\lambda = 1$

Fig. 2. Visualization of simulated networks, m = 1 and different λ values.

Visually and roughly, we can observe a pattern across the five graphs as the value of λ increases. When the value of λ increases, there are more super nodes (nodes with a large number of links) in the system. When λ value decreases, nodes are connected more evenly and there are almost no observable super nodes. In other words, we could explain this correlation as a result of people making decisions according to preferential attachment. There would be many more monopolies that owned lots of social resources in the emergent social networks. When people make decisions upon the heterogeneous attachment, the social resources may be evenly distributed. However, this is only an observation and a rough inference based on the model visualization we have so far. As suggested Barabási and Albert [35], degree distribution of preferential attachment follows a power law distribution. Degree distribution refers to the number of linker of each node. Instinctively, degree distribution of heterogeneous attachment should follow exponential distribution. In the next section, a statistical analysis on the degree distribution outputted from our agent-based model is provided.

In this dissertation, an integrative framework is proposed to understand the impacts of preferential attachment on the emergence of social networks. Particularly, this dissertation emphasizes that when joining a real social network, people are not driven simply by instrumental calculation of connecting with the popular, as stated in the BA model. They are also motivated by intrinsic affection of joining the like. In other words, people are constantly weighting between popularity and proclivity in forming their social connections.

3.5 After Simulation and Beyond Network Statistics

Researchers of SNA have long employed a variety of measurement to describe structural characteristics of a network [4], [7], [8]. In this dissertation, many vertex- and local-level measurements will be used to examine various properties of simulated networks. These network metrics provide a convenient way to evaluate the impacts of heterogeneous attachment on network formation. For the definitions and details of these measurements, please see Chapter 2. These measurements serve valuable purposes in describing and understanding network features that might bear on particular research questions. However, it should be noted that a single network metric is inherently limited in revealing the complex nature of a particular network. Beyond simple descriptive network statistics, many analytical methods have been introduced to explore the complex emergence of networks. The reasons are multiple.

First, social behavior is complex, and stochastic models allow us to capture both the regularities in the processes giving rise to network ties while at the same time recognizing that there is variability that we are unlikely to be able to model in detail. Second, statistical models also allow inferences about whether certain network substructures—often represented in the model by one or a small number of parameters—are more commonly observed in the network than might be expected by chance. Third, sometimes, different social processes may make similar qualitative predictions about network structures and it is only through careful quantitative modeling that the differences in predictions can be evaluated. Therefore, it is important, if not necessary, to go beyond simple network statistics and search for a well-fitting analytical model of a simulated or observed network and in a particular statistical model.

In this section, the degree distribution analytical method proposed by Clauset et al. [149] is discussed. With respect to the post-simulation analysis in this dissertation, this method is employed for the final statistical analysis as well.

3.5.1 Degree Distribution

Degree distribution provides a way to fit a given social network into a statistical model. Following statistical analysis techniques on the power law distribution provided by Clauset, Shalizi and Newman [149], it is possible conduct an analysis on the degree of distribution of the data generated from the simulation. The main goal of this statistical analysis is to learn to what degree distribution changes according to the different λ values, and furthermore, to test the hypotheses that when $\lambda = 1$, it is reasonable to conclude that the degree distribution follows a power law distribution and when $\lambda = 0$, the exponential or Poisson distribution is better to describe the degree distribution.

The first group of graphs puts 30 runs of data points in one figure with each different λ value. The Y-axis indicates the degree and the X-axis indicates the regarding numbers with the degree. Fig. 3 shows that there are big differences with different values. When λ is close to 1, there are some nodes with a high degree, but most of the nodes have lesser degree. Roughly, the graph with degree distribution of $\lambda = 1$ shows a feature of big tail. However, when λ is close to 0, the feature of big tail dispersed.

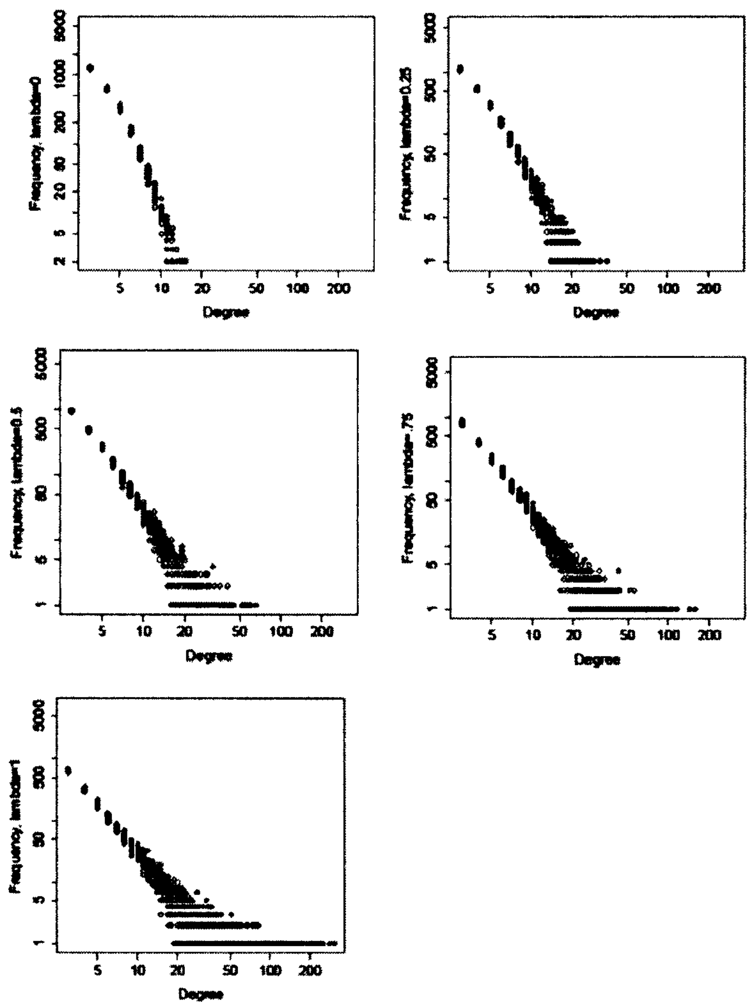


Fig. 3. Degree distribution of simulated network shown on log scales, *m*.

Mathematically, a quantity x obeys a power law if it is drawn from a probability distribution:

$$p(x) \propto x^{-\alpha}. \tag{23}$$

As noted by Watts (2004), the probability of a randomly chosen node having degree x decays like a power of x, where the exponent a, typically measured in the range of $2 < \alpha < 3$, determines the rate of decay (smaller a implies slower decay, hence a more skewed distribution). A distinguishing feature of power law distributions is that when plotted on a double logarithmic scale, a power law appears as a straight line with negative slope a. Argued by Clauset et. al [149], few empirical phenomena obey power laws for all values of x. Usually, the power law applies only for the values greater than some minimum $x_{\min}$. Basically, power law distribution has two different settings: continuous distributions with the continuous real number and discrete distributions with discrete set of positive integers. Since the data of this model are positive integers, the probability distribution should follow the form of:

$$p(x) = \Pr(X = x) = Cx^{-\alpha}. \tag{24}$$

This density function diverges at $x = 0$, implying a lower bound $x_{\min} > 0$ to the power law behavior. The normalizing constant can be calculated as in the following equations:

$$p(x)=\frac{Cx^{-\alpha}}{\zeta(\alpha,x_{\min})}. \tag{25}$$

Based on the power law distribution, provided by Clauset, Shalizi and Newman [149], the approach is used for analyzing the data generated from the simulation to test how the value of λ affects the degree distribution, and the system behavior of the network formation process. Specifically, this dissertation will conduct a two-step analysis:

1. Estimate the parameters of power law distribution: $x_{\min}$ and scaling parameter λ.
2. Compare the power law with alternative hypothesis. Here, a likelihood ratio test method is approached. For each alternative, if the calculated likelihood ratio is significantly different from zero, then its sign indicates whether or not the alternative is favored over the power law distribution.

3.5.2 Estimating Parameters

There are two key parameters that need to be estimated in a power law model: x_{min} and scaling parameter alpha. The fundamental idea in Clauset, Shalizi and Newman [149] is the lower bound $x_{\min}$ should make the probability distributions of the measured data which are above this lower bound and the best-fit power law model as similar as possible. The Kolmogorov-Smirnov or KS statistic is commonly used as a measurement of quantifying the distance between the two probability distributions, which indicates the maximum distance between the cumulative distribution functions (CDFs) of the measured data and the fitted model:

$$D = \max_{x \geq x_{min}} \left| S(x) - P(x) \right| \tag{26}$$

where $S(x)$ is the CDF of the simulated data for the observations with the value greater than x_{min}, and the $P(x)$is the CDF for the power law model that best fits the data in the region $x \geq x_{min}$. Hence the $\hat{x}_{min}$ is selected to be the x_{min} that minimizes the value of D.

Referring to the estimating, the method of maximum likelihood provably gives the accurate parameters estimates with the limit of large sample size. Assuming the data are drawn from a distribution following a power law for $x \geq x_{min}$, the scaling parameter estimation under discrete case can be derived through maximum likelihood estimators (MLEs) as:

$$\hat{\alpha} \simeq 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min} - \frac{1}{2}} \right]^{-1} \tag{27}$$

where x_i, $i = 1,\ldots,n$, are the observed values of x such that $x_i\ x_{min}$.

Following this procedure, we can estimate the parameters for the data from the simulation. According to the experiments of our simulation, with 5 different λ values, 30 repeated runs based on 10,000 data points would be estimated in the following chapter.

3.5.3 Comparing Alternative Distributions

While above estimated parameters indicate whether power law is an appropriate distribution in describing simulated networks, they provide no information about relative fitness of power law distribution against alternative distributions. This section provides a brief discussion about the statistical test if the simulated data are possibly drawn from a power law distribution. Is it still possible that another distribution, such as an exponential or a Poisson distribution, might give a fit as good or better? Given the method mentioned in the above two sections, we only need to run through the whole process again for different distribution candidates. Relying on the p-values, we make the judgment of accepting or rejecting the hypothesis.

$$p(x) = \Pr(X = x) = C_{\exp} e^{-\lambda x} \tag{28}$$

where,

$$C_{\exp} = \lambda e^{\lambda x_{\min}} \tag{29}$$

$$p(x) = \Pr(X = x) = C_{\text{log-normal}} \frac{1}{x} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right] \tag{30}$$

where,

$$C_{\text{log-normal}} = \sqrt{\frac{2}{\pi\sigma^2}} \left[\operatorname{erfc}\left(\frac{\ln x_{\min} - \mu}{\sqrt{2}\sigma} \right) \right]^{-1} \tag{31}$$

While Eq. (28) and Eq. (29) defines an exponential model to fit the degree distribution of the simulated network, Eq. (30) and Eq. (31) provides a way to fit a log-normal distribution. After the simulated network data are fitted against these two alternative distributions, we can conduct a test examining which distribution model fits the data best.

3.6 Conclusions and Discussion

In summary, three important methodological issues to this study are discussed in this chapter. First, this chapter elaborates why the agent-based modeling (ABM) approach is the one of the best solutions to simulating the dynamic formation of social networks. The flexibility of ABM allows for an exploration of heterogeneous agents in complex networks. The theoretical structure for this research then is provided; that is, besides the instrumental connection originated from the need of cognition, the intrinsic connection originated from the need of affect is another key driving mechanism to structure the social networks.

The main technical challenge in this simulation experiment is how to model agent heterogeneity. The agent heterogeneous attachment hypothesis and operative automata in simulation are introduced. In order to differentiate the agents, binary colors are used for agents to identify the similarity between each pair. The potential impact of different communication capacity is also discussed. Finally, the degree distribution analysis method is explained. This method helps to investigate the aggregate properties observed from the global view parameters.

CHAPTER 4

RESULTS AND APPLICATIONS

The discussion in the previous chapters has described the theoretical backgrounds and modeling procedures of agent-based simulations required for heterogeneous network formation. From the previous theoretical discussions, it is still not clear what the actual impacts were for the nature of heterogeneous attachment and the individual capacity on various properties of the resultant social networks. The results presented in this chapter address this crucial question by examining various properties of simulated social networks generated based upon different combination of popularity-proclivity parameter, λ, and individual capacity, m.

For this examination, a large number of simulated social networks were generated based on individuals' heterogeneous attachments. These network properties encompass vertex-level, local-level, and global-level network statistics. Through examination of the relationships of these network statistics to the heterogeneous attachment, λ, and individual capacity, m, the following questions will be addressed: To what extent does agent heterogeneity affect actual formation of social networks? More specifically, when people are popularity- and proclivity-orientated, will the resultant social network be better connected, equally connected, or efficiently connected?

Previous studies on the origins and formations of social networks have focused primarily on examining impacts of various generative mechanisms. While such studies provide important insights about the abstract mechanisms of network formation, very few of them juxtapose their generative models with empirical social networks (e.g., Hamill

Gilbert [49]). Thus there is only limited knowledge about the empirical applications and implications of these abstract models that has been obtained.

Recognizing these issues, findings emerged from the agent-based simulation results and analyses have been applied to exploring data of real world social networks. Specifically, when examining the relationships of heterogeneous attachment, λ, and individual capacity, m, to various network statistics of a simulated social network, reliable measures for differentiating different levels of λ are identified. Employing these measurements, the extent to which people are popularity- or proclivity-oriented in real social networks can be explored. To do so, the diffusion of different kinds of emotions in *Sina Weibo* (a Chinese Twitter-like social media) was examined. A comparison of the network patterns can help reveal different micro-foundations of diffusion of different emotions.

4.1 Network Descriptives

To what extent does the above-discussed heterogeneous attachment affect various properties of social network? The answer to this question is not only of critical importance to the evaluation of the impacts of heterogeneous attachment on network formation but, more importantly, such analyses can help identify the reliable and valid aggregate measures for individuals' heterogeneous attachment at the micro level. After all, in most cases it is impractical to survey how individuals balance between "popularity" and "proclivity." Particularly, when multiple networks are under consideration, researchers require measures that are effective in differentiating the extent of heterogeneous attachment and the level to which they are comparable over a broad variety of networks. Therefore, the goal in this section is two folded: (1) examine the effects of heterogeneous attachments on various properties of the simulated networks, and (2) identify reliable measures for heterogeneous attachment by evaluating and comparing different network properties.

Specifically, this section focuses on two classes of network properties. First, the relationship of heterogeneous attachment to various popular network statistics such as centrality measures, average path length, and diameter [7], [8], [21], [43], [102]. Then this section explores to what extent heterogeneous attachment affect degree distribution of simulated networks. Finally, a comparison of different measure is delivered.

4.1.1 Vertex-level Statistics

Centrality measures, in gauging the relative importance of nodes in a given network, perhaps are the most intuitive network statistics. This study here presents the impacts of heterogeneous attachment on three different centrality measures: degree, closeness, and betweenness centrality [3], [4], [7], [8], [102].

The relationship of average degree score to the nature of heterogeneous attachment λ and individual capacity m must be examined first. For a given network graph $G=(V,E)$, the degree d_v of a vertex v, as discussed in Chapter 2, calculates the total number of edges in E incident upon the vertex v. The standard deviation of each of the simulated networks, therefore, provides one of most intuitive measures about the connectivity of a particular network. In this dissertation, since the simulated networks are assumed to be undirected, no differentiation is made between in-degrees and out-degrees.

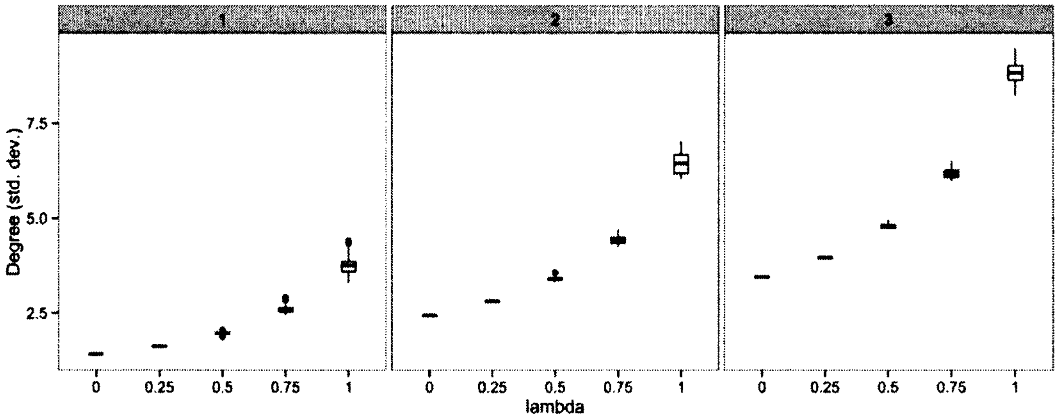


Fig. 4. The relationship of the standard deviation in degree score to λ and m.

To examine the impacts of λ and m on standard deviation in degree of the simulated networks respectively, the degree over all the vertices in the thirty simulated networks for each combination of λ and m values must be calculated. The results are presented in Fig.4, and three findings stand out. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, strongly affects the standard deviation in degree of simulated networks. Specifically, when controlling for m, the partial correlation between λ and the standard deviation in node degree is 0.902 and is statistically significant. With larger values of λ, the network-wise variation in degree centrality increases as well. In other words, as individual agents become more popularity-oriented, on average they tend to be less equal in possessing social links in a given social network. Therefore it is suggested that popularity orientation tends to make a social network less equally connected.

Second, after comparing the three subgraphs, significant and positive impacts of individual capacity was found. The partial correlation between m and the standard deviation in node degree is 0.897 and is statistically significant when λ is controlled. That is, with better networking media and stronger individual capacity, the resultant social network tends to be less equal in connectivity and higher in the standard deviation of degree score. Finally, it is evident that the standard deviation in the degree scores are sensitive to both the nature of heterogeneous attachment λ and individual capacity, *m*. Considering actual social networks vary simultaneously in these two dimensions, it is suggested that the standard deviation in degree score cannot serve as a reliable measure for heterogeneous attachment.

The degree score, though intuitive, only provides limited information about the connectivity of a network graph. It is therefore necessary to further examine the relationship of the average closeness centrality score to the nature of heterogeneous attachment λ and individual capacity *m*. For a given network graph $G=(V,E)$, the closeness score $c_{cl}(v)$ of a vertex v, as discussed in Eq. (3) in Chapter 2, captures the notion to what extent a vertex v is "close" to the other vertices in a given graph. Average closeness score, therefore, provides another way to gauge the connectivity of a particular network. It should be noted that, that this measure assumes the graph G is connected, as otherwise all vertices in principle will have closeness centrality $c_{cl}(v)=0$, which is a result of infinite distance from at least one other vertex.

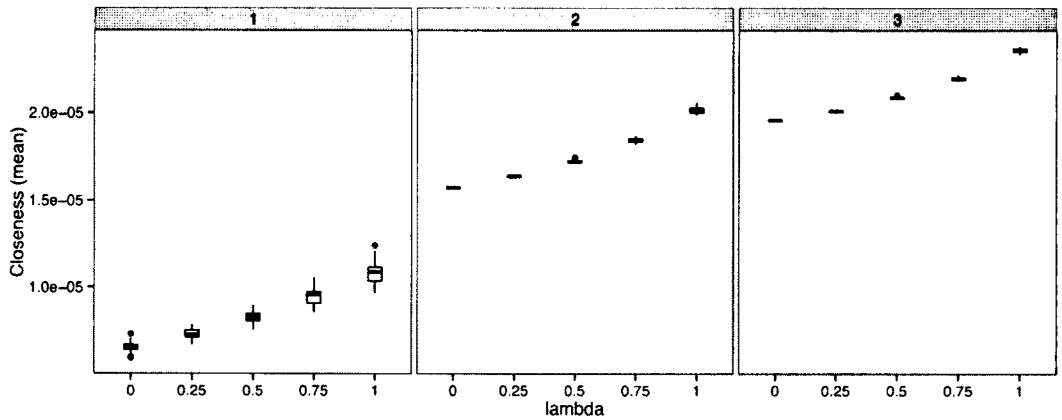


Fig. 5 The relationship of the average closeness score to λ and m.

To examine the impacts of λ and m on average closeness centrality score of the simulated networks respectively, its specific values were calculated over all the vertices in the thirty simulated networks for each combination of λ and m values. The results are presented in Fig. 5, and three findings stand out. First, after examining each of the three subgraphs, it is evident that the nature of individuals' heterogeneous attachment, λ, appears to strongly affect the average closeness centrality score of the simulated networks. The partial correlation between λ and the standard deviation in node degree is 0.739 and is statistically significant when m is controlled. With larger values of λ, the network-wise closeness centrality score increases as well. In other words, as individual agents become more popularity-oriented, on average they tend to be much closer to each other in a given social network. Therefore it is suggested that popularity orientation tends to make a social network more closely connected.

Second, after comparing the three subgraphs, significant and positive impacts of individual capacity, m was found. By controlling the value of λ, the partial correlation between m and the standard deviation in node degree is 0.968. That is, with better networking media and a stronger individual capacity, the resultant social network tends to be more closely connected and higher on average degree score. Finally, it is evident that the average closeness centrality scores are sensitive to the nature of heterogeneous attachment λ and individual capacity m, and thus the average closeness centrality score cannot serve as a reliable measure for heterogeneous attachment.

It should be noted that closeness score provides only the central tendency about the extent to what individual agents are close to each other. There is limited knowledge about to what extent such "closeness" varies across different agents. Yet the variations in the measure of closeness centrality score are of critical importance to our understanding about the relative "inequality" in connectedness. To a certain extent, the variations in closeness centrality scores can reveal the extent to which the simulated networks are evenly clustered.

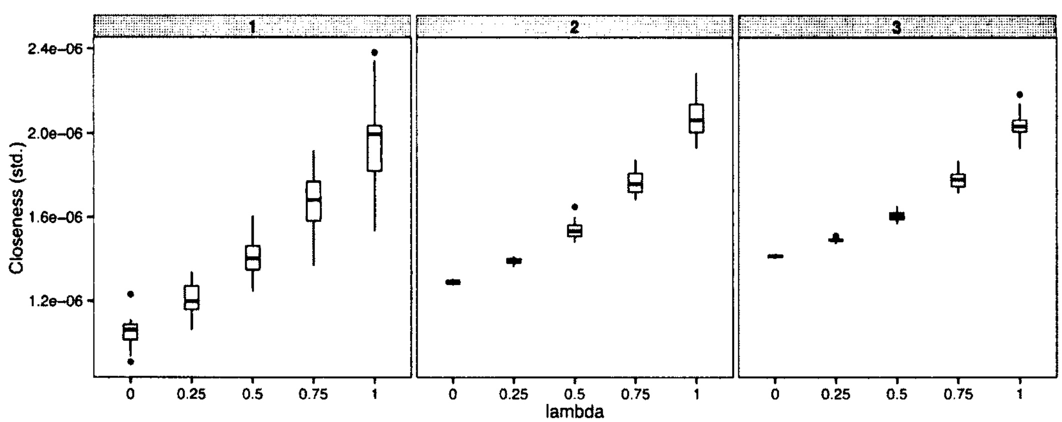


Fig. 6. The relationship of variations in closeness score to λ and m.

Specifically, the variations in the closeness centrality scores over all the vertices in the thirty simulated networks for each combination of λ and m values were calculated.

The results are presented in Fig. 6, and three conclusions are drawn. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, strongly affects the variations in the closeness centrality scores of the simulated networks. Specifically, when controlling for m, the partial correlation between λ and the standard deviation in closeness centrality score is 0.931 and is statistically significant. As the values of λ increase, the variations in the closeness centrality scores enlarge as well. In other words, when individual agents become more popularity-oriented, they tend to form a less equal network in which some agents are much closer to each other than others. Therefore it is suggested that popularity orientation, though making a social network better connected, tends to introduce unequal access for agents. Some agents are in more advantageous positions than others.

Second, after comparing the three subgraphs, significant and positive impacts of individual capacity, m was found. After a partial correlation analysis that controlled for λ, the result is 0.623 and is statistically significant. That is, with better networking media and stronger individual capacity, the resultant social network tends to be higher in variations in the closeness centrality score. In practical terms, the introduction of new communication tools and advancements in individual capacities fail to solve the inequality in mass communication. On the contrary, it tends to enlarge the existing gap in a given social network. Finally, it is evident that the variations in the closeness centrality scores are a function of both the nature of heterogeneous attachment λ and individual capacity m. Considering the fact actual social networks vary simultaneously in these two dimensions, it is concluded that the variations in the closeness centrality scores cannot serve as a reliable measure for heterogeneous attachment.

As addressed in Chapter 2, the closeness centrality score is one of many vertex-level measures for a given social network $G = (V, E)$. Another important vertex-level centrality measure is the betweenness centrality score. This centrality score intends to capture the perspective that "importance" relates to where an agent is located regarding the paths in the social network and this particular measure gauges the extent to which an agent is located "between" other pairs of agents. In light of this, the average of the betweenness centrality score provides another intuitive measure about the sparseness of a particular network. Therefore, a calculation of the betweenness centrality scores of all the vertices in all thirty simulated networks for each combination of λ and m values was made.

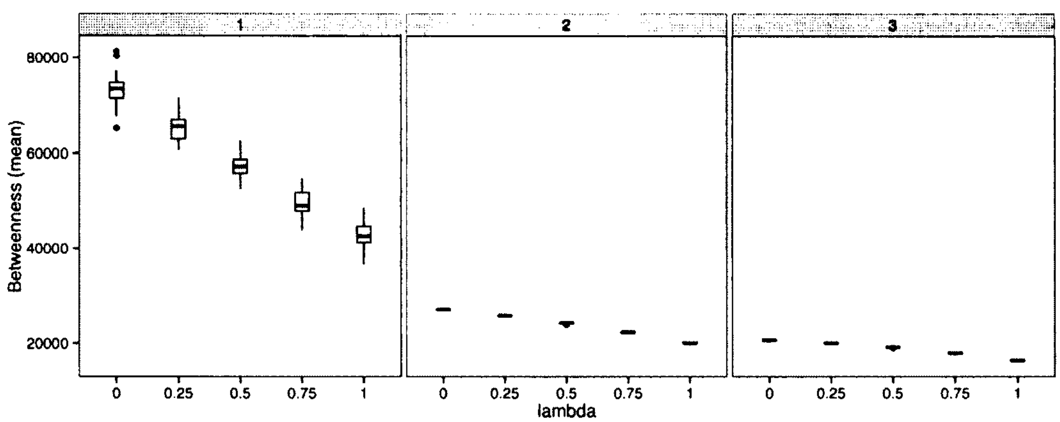


Fig. 7. The relationship of the average betweenness score to λ and m.

Firstly, an examination of the impacts of λ and m on the average betweenness centrality scores of the simulated networks respectively is conducted, and the results are presented in Fig. 7. After examining and comparing the three subgraphs, three conclusions are drawn are drawn. Firstly, examinations of each of the three subgraphs suggest that the nature of individuals' heterogeneous attachment, λ, strongly affects the average betweenness centrality score of the simulated networks. The partial correlation between the two is -0.517 and is statistically significantly. That is, with larger values of λ, the network-wise betweenness—the average number of shortest paths from all vertices to all others that pass through each node in a network—centrality score decreases accordingly. In other words, as individual agents become more popularity-oriented, on average they tend to become less "important" to each other in the sense that they are less likely "between" other pairs of agents. Therefore it is concluded that popularity orientation tends to make a social network less interconnected in the sense that there will be a smaller number of agents located in the paths of other pairs of agents.

Second, a comparison the three subgraphs indicated significant and negative impacts of individual capacity, *m*, and the partial correlation between m and the average betweenness centrality score is -0.890. With better networking media and stronger individual capacity, the resultant social network tends to be less interconnected and lower in the average betweenness centrality scores. Finally, it is evident that the average betweenness centrality scores are sensitive to the nature of heterogeneous attachment and individual capacity *m*. Considering actual social networks vary simultaneously in these two dimensions, it is argued that the average betweenness score cannot serve as a reliable measure for heterogeneous attachment.

Similar to the previous discussion about the closeness centrality score, it is evident that that average betweenness centrality score provides only the central tendency about the extent to what individual agents are close to each other, which in turn leaves us uninformed about to what extent such "betweenness" varies across different agents. Yet the variations in the measure of betweenness centrality score are of critical importance to our understanding about the relative "inequality" in the interconnectedness of a given network. To a certain extent, the variations in betweenness centrality scores can reveal the extent to which the simulated networks are evenly interconnected.

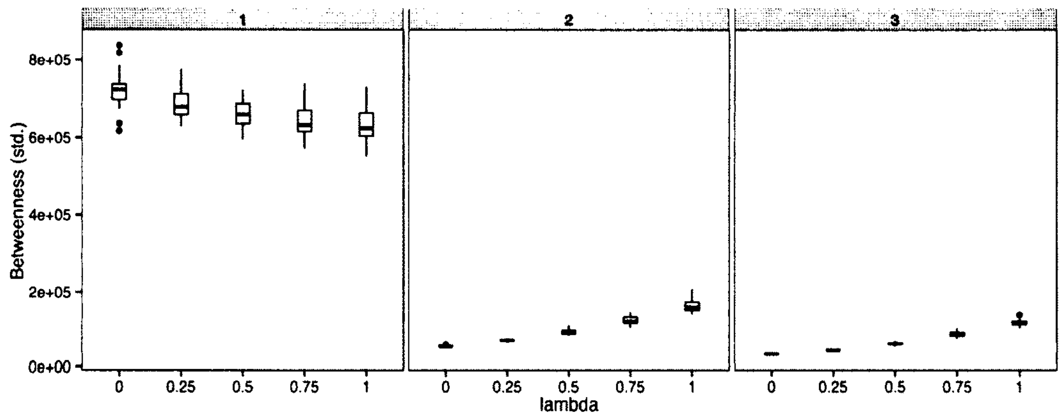


Fig. 8. The relationship of variations in betweenness score to λ and m.

To do so, the variations in the betweenness centrality score for all thirty simulated networks for each combination of λ and m values are calculated, and the results are presented in Fig. 8. Three conclusions are drawn. Firstly, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, strongly affects the average degree of simulated network. Particularly, when $m = 2$ and 3 with larger values of λ, the network-wise variations in the betweenness centrality score increases as well. The overall partial correlation when m is controlled is 0.096. In other words, when individual agents become more popularity-oriented, they tend to form a less equal network in which some agents are more likely located in paths of other paired agents. Therefore it is suggested that popularity orientation not only makes a social network less interconnected but also tends to introduce unequal "betweenness" for agents. Some agents are more interconnected than others.

Second, after comparing the three subgraphs, significant and negative impacts of individual capacity, *m*, were found, and the partial correlation score is -0.879. That is, with better networking media and stronger individual capacity, the resultant social network tends to higher in variations in the betweenness centrality score. In other words, the introduction of new communication tools and advancement in individual capacities fail to make social networks more interconnected. On the contrary, it tends to enlarge the existing gap in the betweenness for a given social network. Finally, it is evident that the variations in the betweenness centrality scores are a function of both the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is apparent that the variations in the betweenness centrality scores cannot serve as a reliable measure for heterogeneous attachment.

To take stock of the above, in this section, the impacts are examined for the nature of heterogeneous attachment λ and individual capacity m on three different vertex-level measures of simulated network; that is, degree, closeness centrality, and betweenness centrality score. In general, there are three important findings. Firstly, for all three vertex-level measures, significant impacts of the nature of individuals' heterogeneous attachment, λ, were found. Specifically, as the values of λ increase, the resultant social networks tend to be higher in the average degree and the closeness centrality score, but lower in the average betweenness score. However, it should be noted that impacts of λ

tend to be quite uneven. When individual agents become more popularity-oriented, they tend to less equal across all three measures, meaning variance increases. Second, significant impacts of individual capacity, m, were found. While m is positively associated with the degree score and closeness centrality score, it bears a negative correlation with the betweenness centrality score. Third, it is evident that none of these vertex-level measures can serve a reliable indicator for heterogeneous attachment, λ.

4.1.2 Local Network Statistics

Vertex-level measures, though intuitive, provide only limited information about the properties of a given social network. Recognizing this, many local, or meso-level, network measurements have been introduced by scholars [4], [7], [8], [102]. These measurements help reveal important information about connectivity of a given network. In this subsection, two local network statistics are focused on; that is, transitivity and diameter.

As addressed in Chapter 2, transitivity of the graph has long been a standard network statistics in the literature of SNA. As reveal in the name of the term, it is simply referred to the "fraction of transitive triples." Transitivity, therefore, helps highlight the extent to which triangular relationship is common in a particular social network. In other words, transitivity tells us where the friend of your friend is also a friend of yours.

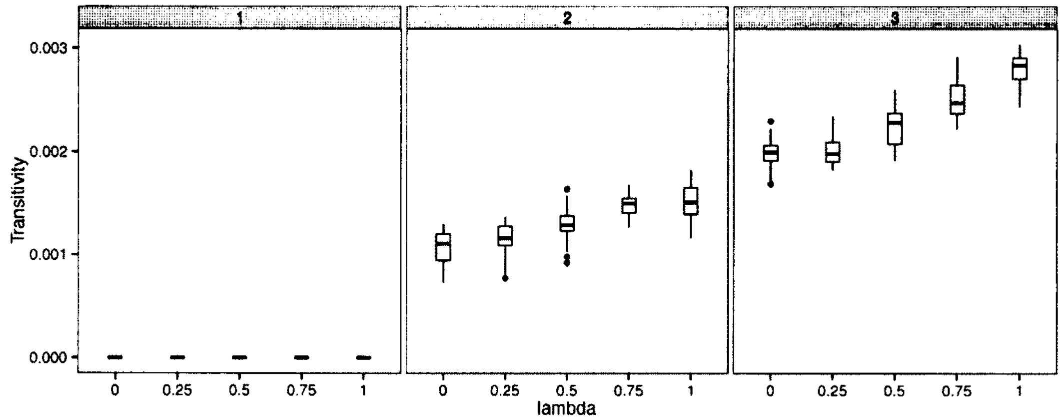


Fig. 9. The relationship of the transitivity score to λ and *m*.

To explore the impacts of λ and *m* on the extent to which triangular relationship is common in the simulated network, the transitivity score was calculated in the thirty simulated networks for each combination of λ and *m* values.[3] The results are presented in Fig. 9, and three findings stand out. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, strongly and positively affects the transitivity scores of simulated network. Specifically, when controlling for *m*, the partial correlation between the two is 0.628. That is, with larger values of λ, the network-wise transitivity score increases as well. In other words, as individual agents become more popularity-oriented, they tend to form more triangular connections among each other.

Second, after comparing the three subgraphs, significant and positive impacts of individual capacity, *m*, were found, and the partial correlation score is 0.980. That is, with better networking media and stronger individual capacity, the resultant social networks

[3] It should be noted that when m=1, the transitivity score remains zero for any given simulated network.

tend to be higher in the transitivity scores and encompass more triangular relationships. Finally, it is evident that the transitivity scores are sensitive to the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is evident that the transitivity score cannot serve as a reliable measure for heterogeneous attachment.

The relationship of the diameter score to the nature of heterogeneous attachment λ and individual capacity m was further examined. As elaborated in Chapter 2, diameter is introduced to describe the common notion of "distance" in a given graph; that is, the length of the shortest path(s) between the agents. It should be noted it is usually set equal to infinity when no such path exists. As for diameter, the measure is retrieved by calculating the longest distance in a particular graph (i.e., the geodesic distance, for more details see Chapter 2).

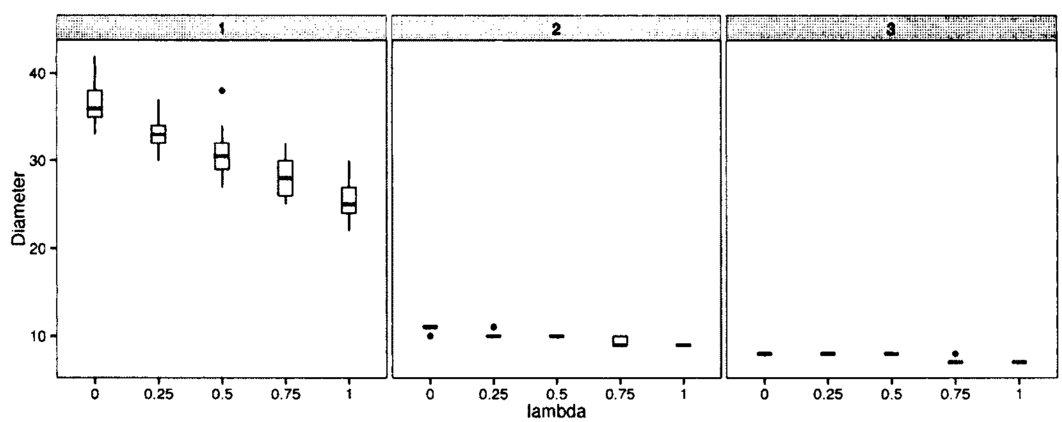


Fig. 10. The relationship of the diameter to λ and m.

To empirically examine the impacts of λ and m on diameter, the diameters of the thirty simulated networks for each combination of λ and m values were calculated. The results are presented in Fig. 10, and three findings stand out. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, moderately and negatively affects the average diameter of simulated network. Specifically, the corresponding partial correlation between the two is -0.318. That is, with larger values of λ, the network-wise diameter decreases substantively as well. In other words, as individual agents become more popularity-oriented, the social network that they form tends to be tightly connected. Therefore it is suggested that popularity orientation tends to make a social network more tightly connected.

Second, after comparing the three subgraphs, significant and negative impacts of individual capacity, m, were found, and the partial correlation score is -0.889. That is, with better networking media and stronger individual capacity, the resultant social network tends to be more tightly connected and lower in overall diameter. Finally, it is evident that the diameter scores are a function of both the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, the diameter of a social network cannot serve as a reliable measure for heterogeneous attachment.

4.1.3 Global Network Statistics

Besides the vertex-level and meso-level network statistics, some system-level, i.e. global, network statistics were also examined. Unlike the previously discussed measures, these measures focus more on systematic properties of a network, providing some important insights on the overall nature of a particular social network. Specifically, three global network statistics were examined: the average path length, the articulation points (i.e., the number of cuts), and finally the assortivity degree.

The average shortest path length is a common measure of the global network structure in the literature of SNA [41]-[43]. Generally, it intends to gauge the notion of the "steps" from one vertex in the network to another. More specifically, the average shortest path length is calculated as the average number of edges and links that must be traversed in the shortest path between any two pairs of nodes in a particular network. The average shortest path length is a global measure simply because determining the shortest path length between any two vertices requires information about the entire graph. For this particular reason, the average shortest path length is employed in many SNA such as the small world phenomena [41]-[43].

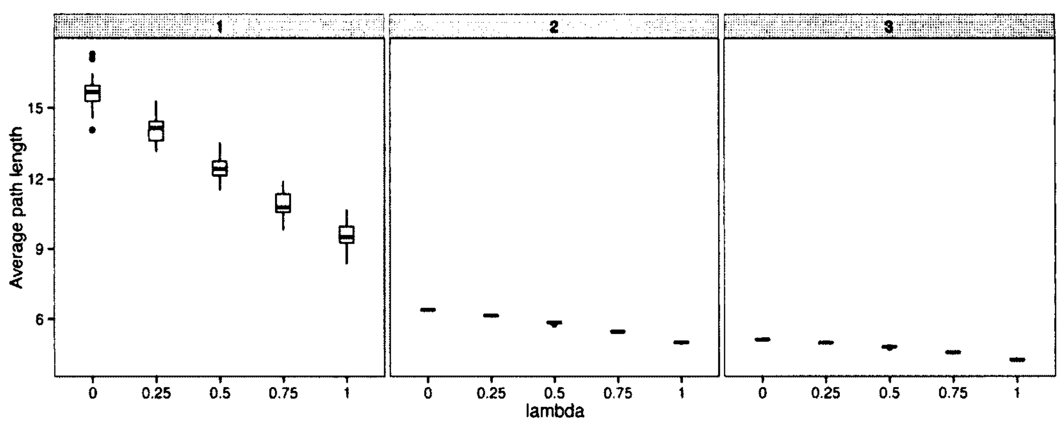


Fig. 11. The relationship of the average shortest path length to λ and m.

To examine the impacts of λ and m on the average shortest path length respectively, the average path length was calculated. The results are presented in Fig. 11, and three findings stand out. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, strongly affects the average shortest path length of simulated network, and the corresponding partial correlation between the two is -0.517. With larger values of λ, the average shortest path length decreases substantively. In other words, as individual agents become more popularity-oriented, on average they tend toward social networks in which the average number of edges and links that must be traversed between any two pairs of nodes in a particular network is low. Therefore it can be concluded that popularity orientation tends to make a social network more efficiently linked.

Second, after comparing the three subgraphs, significant and negative impacts of individual capacity, m, were found, and the corresponding partial correlation is -0.890. That is, with better networking media and stronger individual capacity, the resultant social network tends to be more efficiently connected and higher in the average path length. Finally, it is evident that the average path length is sensitive to the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is suggested that the average path length cannot serve as a reliable measure for heterogeneous attachment.

As addressed in Chapter 2, a basic question in SNA is whether a given social network can be separated into distinct sub-components or sub graphs. If there were not natural subcomponents, then there is a challenge in how to quantify the extent to which a given network can be cut into distinct subgraphs. The question is of particular importance when the flow of "information" in a particular social network is concerned. When a network contains a large number of "cut" points, it is vulnerable in the information flow.

One way to capture this notion, as addressed in Chapter 2, to calculate the number of articulation points, also known as the *vertex-cuts*. Specifically, the articulation points refer to a particular set of vertices in a particular network, whose removal will disconnect the graph, and a single vertex that disconnects the graph is called a cut vertex. Identification of such vertices can provide a sense of where a network is vulnerable (e.g., in the sense of a hacker attack, where disconnecting produces undesired consequences, such as a website hub in a computer network).

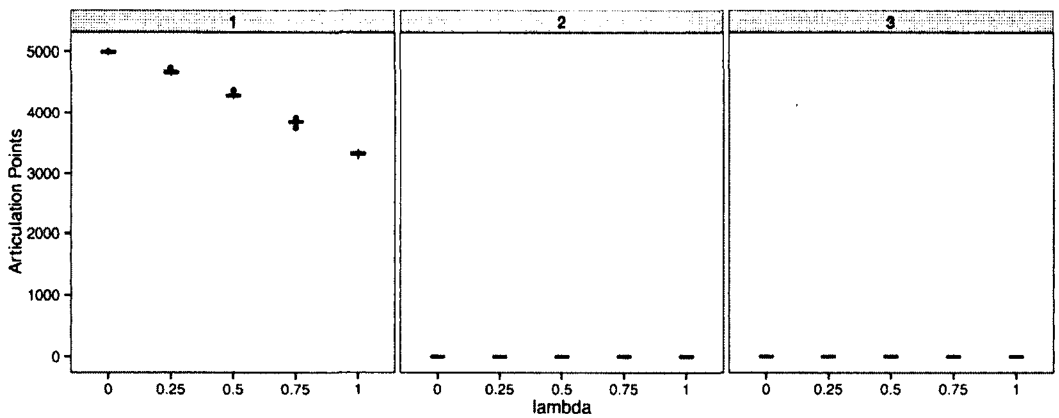


Fig. 12. The relationship of the number of articulation points to λ and m.

To examine the impacts of λ and m on the number of articulation points in the simulated network respectively, all the articulation points were identified and counted in the thirty simulated networks for each combination of λ and m values. The results are presented in Fig. 12, and three findings stand out. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, moderately affects the number of articulation points in the simulated networks. Specifically, the calculated partial correlation between the two when controlling for m is

-0.186; that is, with larger values of λ (especially when $m = 1$), the network-wise number of articulation points drops significantly. In other words, as individual agents become more popularity-oriented, on average, the social network they form tends to be less vulnerable to removal of individual vertices. Therefore it is suggested that popularity orientation tends to make a social network more viable and less vulnerable.

Second, after comparing the three subgraphs, significant and negative impacts of individual capacity, m, were found, and the corresponding partial correlation is -0.858. That is, with better networking media and stronger individual capacity, the resultant social network tends to be less vulnerable and possess a smaller set of articulation points. Moreover, the impacts of individual capacity, m appears to be more evident than those of heterogeneous attachment, λ. Finally, it is evident that the number of articulation points is a function of both the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is evident that the number of articulation points cannot serve as a reliable measure for heterogeneous attachment.

Another global measure of network structure that was examined is assortativity. Unlike the earlier two global measures probing the connectivity of a particular network, assortativity intends to explore the extent to which some vertices have priority to link with each other, based on a certain characteristics. As elaborated in Chapter 2, in some SNA research assortativity is also referred to the term of assortative mixing. For practical purposes, scholars rely on the assortativity coefficients to gauge the variations on the concept of correlation coefficients as specified in Eq. (10). The value of the assortativity coefficient in a particular graph changes between -1 and 1. While the assortativity coefficient of 0 indicates only randomness are involved in linking among all the vertices, the assortativity coefficient of 1 suggests there is a perfect assortative mixing.

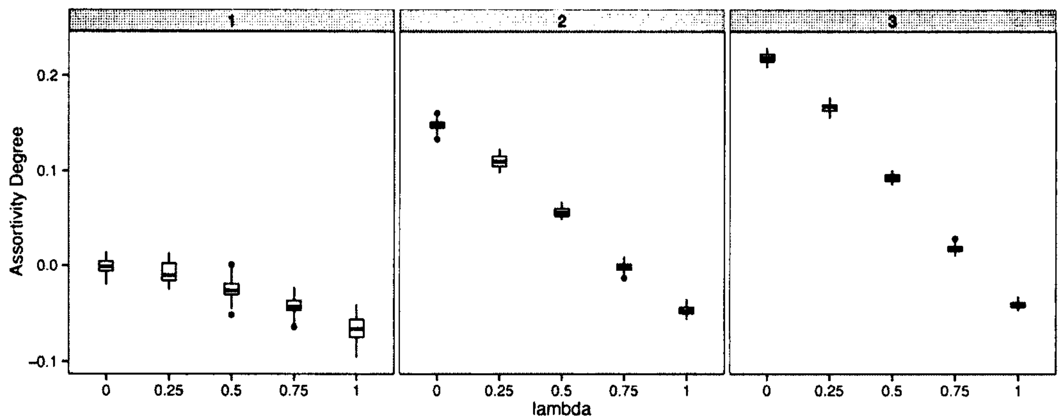


Fig. 13. The relationship of the assortativity degree score to λ and m.

So as to examine the impacts of λ and m on the assortativity degree score of the simulated network respectively, the assortativity mixing coefficients was calculated for the thirty simulated networks for each combination of λ and m values. The results are presented in Fig. 13, and three conclusions are drawn. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, strongly affects the assortativity coefficients of the simulated networks. Specifically, when controlling for m the partial correlation between the two is -0.887 and is statistically significant. With larger values of λ, the assortativity coefficients drop accordingly. In other words, as individual agents become more popularity-oriented, the networks they have are less likely to link with each other based on a certain characteristic (e.g., a node's degree). Therefore it is suggested that popularity orientation tends to make a social network less assortatively mixed.

Second, after comparing the three subgraphs, significant impacts of individual capacity, m, can be observed. Rather than a simple positive or negative impact, individual capacity, m, seems to amplify the impacts of heterogeneous attachments, λ. In other words, there is a significant interaction between individual capacity, λ, and heterogeneous attachment, m. The partial correlation is 0.832 and further confirms this. From a practical viewpoint, Fig. 13. shows that that with better networking media and stronger individual capacity, the resultant social network tends to have larger variations in the assortative coefficients. Finally, it is evident that the assortative coefficients are contingent upon both the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is suggested that the assortative coefficients cannot serve as a reliable measure for heterogeneous attachment.

4.1.4 Statistics of Graph Partitioning

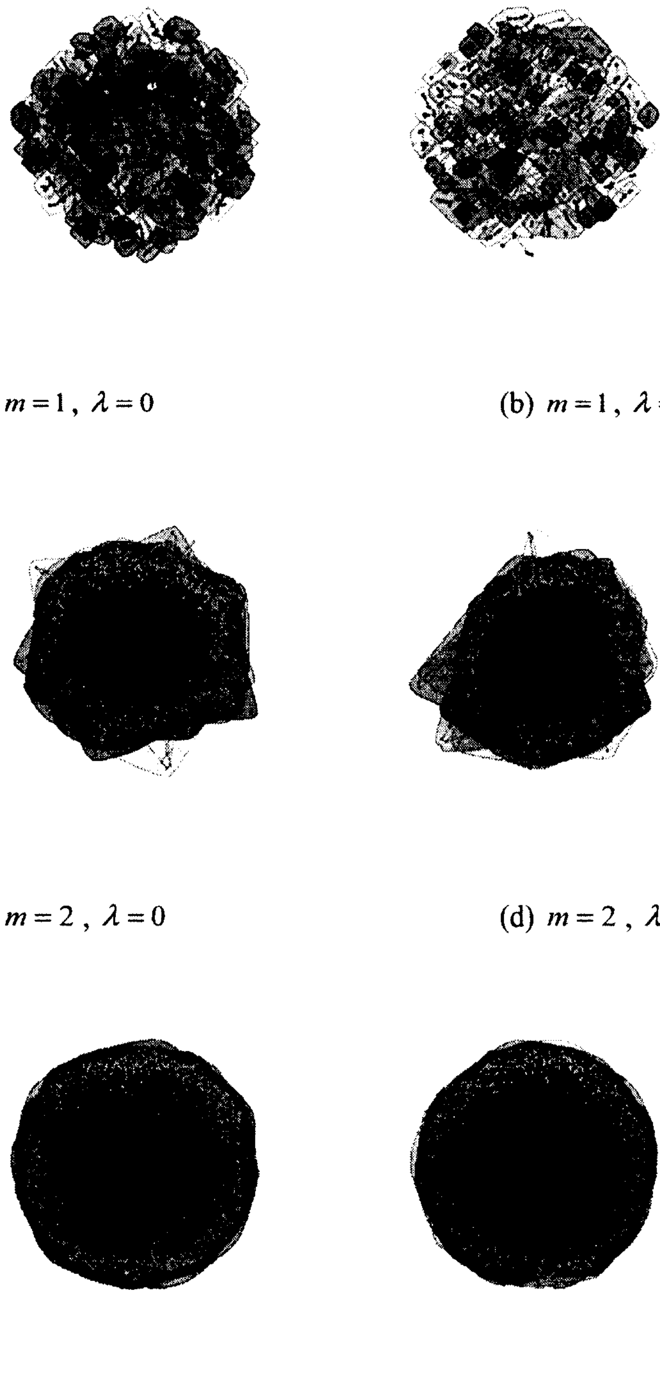
A further look at the extent to which the simulated networks vary in terms of different measures of graph partitioning is conducted next. Generally speaking, graph partitioning refers to the segmentation of a graph's objects into its natural subsets or sub-clusters; that is, the gathering of vertices into groups such that there is a higher density of edges within groups than between them. Many different disciplines require researchers to conduct graph partitioning [4], [7], [8], [102], and the aim is to locate and identify certain subsets of vertices that share and exhibit some forms of "cohesiveness" with respect to certain underlying relational structures. In biology, for instance, graph partitioning can be used to find possible protein complexes from existing protein interaction structures [109].

In studies of social networks, graph partitioning seems even more important, since it helps detect community structures in a given network.

It should be noted that there is no consensus on how to describe and capture the "cohesiveness" among vertices in a particularly network. A "cohesive" cluster of vertices is referred loosely to a subset of vertices that are closely connected within the cluster yet relatively shielded from vertices outside the cluster. In other words, graph partitioning intends to divide a large network into some small solidary groups. Given the large and ever increasing number of graph partitioning methods, this dissertation focuses mainly on two commonly cited methods of graph partitioning; that is, the method of "fast greedy" [150] and "walk trap" [151]. Although far from being exhaustive, an examination on these two measures can provide some basic insights on the impacts of heterogeneous attachment on the "cohesiveness" among vertices in a given graph.

Early approaches such as the Kernighan-Lin algorithm, spectral partitioning, or hierarchical clustering work well for specific types of problems (particularly graph bisection or problems with well-defined vertex similarity measures), but perform poorly in more general cases [102], [120]. Modularity is a property of a network and a specific proposed division of that network into communities. It measures when the division is a good one, in the sense that there are many edges within communities and only a few between them. Non-zero values represent deviations from randomness, and in practice it is found that a value above about 0.3 is a good indicator of significant community structure in a network.

The calculation of the global maximum modularity over all possible divisions is computationally costly in general. First, a discussion is given on the implementation of the fast greedy modularity optimization algorithm for finding community structure. The algorithm of a greedy optimization starts with each vertex being the sole member of a community of one, and then the algorithm repeatedly join together the two communities whose amalgamation produces the largest increase in the modularity.



(a) $m=1$, $\lambda=0$ (b) $m=1$, $\lambda=1$

(c) $m=2$, $\lambda=0$ (d) $m=2$, $\lambda=1$

(e) $m=3$, $\lambda=0$ (f) $m=3$, $\lambda=1$

Fig. 14. Network communities for different values of λ and m.

Fig. 14 provides an intuitive demonstration about graph partitioning using the optimization method of "fast greedy." Specifically, communities identified based upon the method of "fast greedy" is depicted as different colored areas, and vertices in the same colored areas then belong to the same community. An examination of Fig. 14 suggests that both the heterogeneous attachment, λ, and individual capacity, m, significantly affect the graph partitioning based on the method of "fast greedy." To examine the relationship thoroughly, this dissertation examines the impacts of the heterogeneous attachment, λ , and individual capacity, m, on (1) the number of communities, (2) the average size of network communities, and (3) the variations in the sizes of network communities.

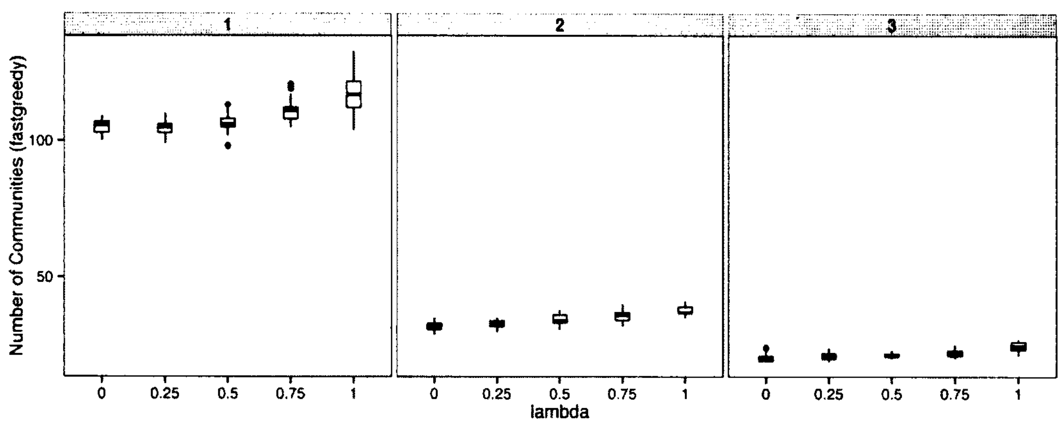


Fig. 15. The relationship of the number of network communities to λ and m , based on the community detection method of "fast greedy".

As for the impacts of λ and m on the number of network communities, the total number of "fast greedy" communities is counted in the thirty simulated networks for each combination of λ and m values, and the results are presented in Fig. 15. There are three interesting findings. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, shapes the total number of "fast greedy" communities in the simulated networks. The partial correlation score of 0.169 suggests that the association between the two is statistically significant yet substantially moderate. With larger values of λ, the network-wise number of "fast greedy" communities increases accordingly. In other words, as individual agents become more popularity-oriented, on average they tend to form more "fast greedy" communities in a given social network. Therefore it is suggested that popularity orientation helps create relatively isolated solidary groups in a social network.

Second and more important, after comparing the three subgraphs, significant yet negative impacts of individual capacity, m, were found, and the partial correlation between the two is -0.922 and is statistically significant. That is, with better networking media and stronger individual capacity, the resultant social network tends to be less internally segregated and lower in the number of "fast greedy" communities. Finally, it is evident that the number of "fast greedy" communities is sensitive to both the nature of heterogeneous attachment λ and individual capacity m. Therefore, it is suggested that the number of "fast greedy" communities cannot serve as a reliable measure for heterogeneous attachment.

Besides the total number of "fast greedy" communities in a social network, their sizes are of interest; that is, the number of vertices in each "fast greedy" community respectively. One way to do so is to calculate and compare the average size of "fast greedy" communities in the simulated networks.

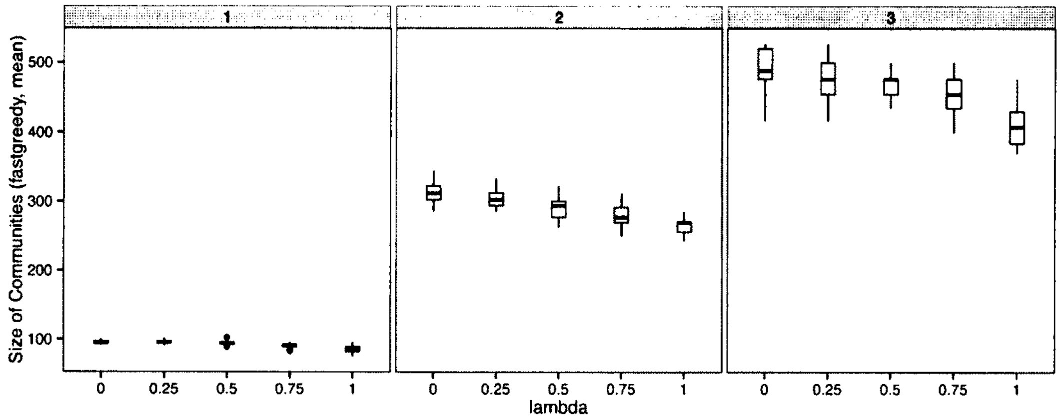


Fig. 16. The relationship of the average size of network communities to λ and m, based on the community detection method of "fast greedy".

In order to explore the impacts of λ and m on the average size of "fast greedy" communities of the simulated networks respectively, their average size is calculated over for all the thirty simulated networks with each combination of λ and m values. The results are presented in Fig. 16, and three findings stand out. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, strongly affects the average size of "fast greedy" communities in the simulated networks. Specifically, the partial correlation between λ and the average size of fast greedy communities is -0.572 and is statistically significant. With larger values of λ, the network-wise the average size of network communities drops as well. In other words, as individual agents become more popularity-oriented, on average they tend to form relatively smaller solidary groups in a given social network. Therefore it is suggested that popularity orientation tends to make a social network more sparsely isolated.

Second, after comparing the three subgraphs, significant and positive impacts of individual capacity are found, m, were found, and its partial correlation score with m is

0.990 and is statistically significant. That is, with better networking media and stronger individual capacity, the resultant social network tends to have much larger "fast greedy" communities and higher in the average size of network communities. Finally, it is evident that the average size of network communities are sensitive to the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is apparent that the average size of network communities cannot serve as a reliable measure for heterogeneous attachment.

It should be noted that the average size of "fast greedy" communities provides only the central tendency about the extent to what individual agents are able to form relatively segregated solidary groups. There is limited knowledge about to what extent the relative size of these communities varies. Yet the variations in the measure of the average size of "fast greedy" communities are of critical importance to our understanding about the relative "disparity" in the clustering. To a certain extent, the variations in the average size of "fast greedy" communities can reveal the extent to which the simulated networks are evenly clustered.

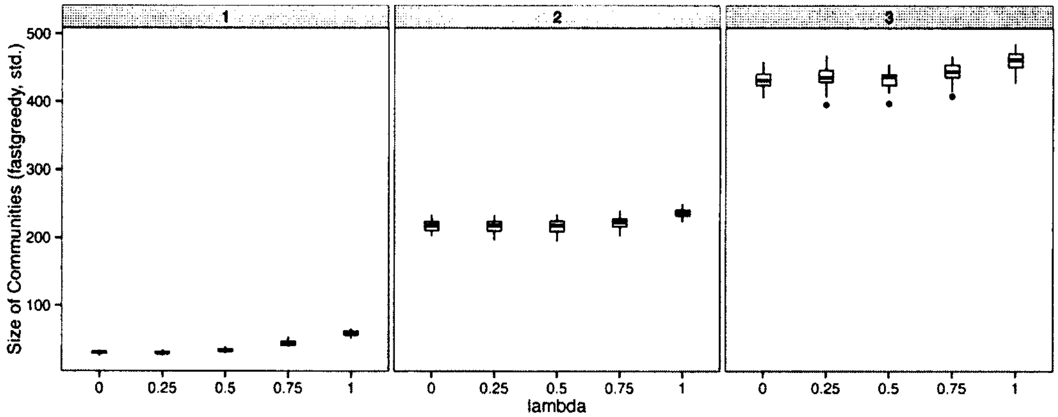


Fig. 17. The relationship of the variations in the sizes of network communities to λ and m, based on the community detection method of "fast greedy".

To examine the impacts of λ and m on the variations in the sizes of "fast greedy" communities respectively, the standard deviation is calculated, and the results are presented in Fig. 17. Three conclusions are drawn. First, after examining each of the three subgraphs, it is found that the nature of individuals' heterogeneous attachment, λ, strongly affects the standard deviations of the sizes of "fast greedy" communities in the simulated networks, its partial correlation score is as high as 0.521 and is statistically significant. Specifically, with larger values of λ, the network-wise variations in the size of "fast greedy" communities increase as well. In other words, as individual agents become more popularity-oriented, the "fast greedy" communities they formed tend to vary more significantly in a given social network. Therefore it is suggested that popularity orientation tends to make a social network less evenly clustered.

Second, after comparing the three subgraphs, significant and positive impacts of individual capacity are found, the partial correlation between the two is 0.996 and is statistically significant. That is, with better networking media and stronger individual capacity, the resultant social network tends to more uneven with regard to the sizes of the "fast greedy" communities. Finally, it is evident that the standard deviations of the sizes of the "fast greedy" communities are sensitive to both the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is suggested that the variations in the sizes of the "fast greedy" communities cannot serve as a reliable measure for heterogeneous attachment.

As discussed above, the "fast greedy" optimization is only one of many ways to calculate the modularity in a given social network. This dissertation thus explores another popular method of the modularity calculation proposed by Pons and Latapy [151]. Pons and Latapy [151] propose a measure of similarities between vertices based on random walks. Their approach is based on the following intuition: random walks on a graph tend to get "trapped" into densely connected parts corresponding to communities. They therefore begin with some properties of random walks on graphs. Using them, Pons and Latapy [151] further define a measurement of the structural similarity between vertices and between communities, thus defining a distance. This method "walk trap," as noted by Pons and Latapy [151], has several important advantages. Not only does the method of "walk trap" capture well the community structure in a network, it can also be computed efficiently used in an agglomerative algorithm to compute the community structure of a network. This approach, by emphasizing the "distance," can be regarded an expansion of existing spectral approaches of the problem.

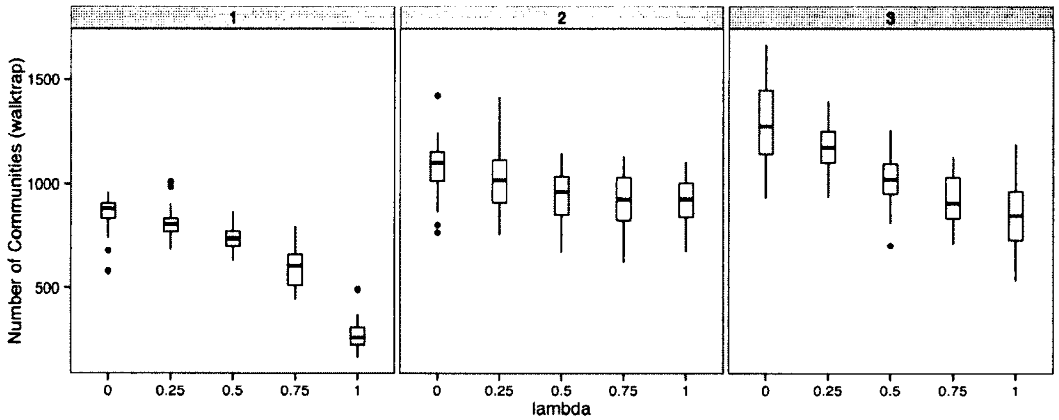


Fig. 18. The relationship of the number of network communities to λ and m, based on the community detection method of "walk trap".

Examining the impacts of λ and m on the total number of the "walk trap" communities in the simulated network respectively is tried first. All the "walk trap" communities are identified, and the results are presented in Fig. 18. After examining the results, three conclusions can be drawn. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, strongly affects the total number of the "walk trap" communities in the simulated networks, and the corresponding partial correlation score is -0.661 and is statistically significant. Specifically, with larger values of λ, the network-wise total number of the "walk trap" communities drops accordingly. In other words, as individual agents become more popularity-oriented, on average they tend to former fewer "walk trap" communities. Therefore it is suggested that popularity orientation tends to make a social network better interconnected.

Second, after comparing the three subgraphs, significant and positive impacts of individual capacity were found, m, and its partial correlation with the average size of walktrap community is 0.718 and is statistically significant. That is, with better networking media and a stronger individual capacity, the resultant social network tends to have more "walk trap" communities and thus is more densely interconnected. Finally, it is evident that the total number of the "walk trap" communities is a function of both the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is evident that the total number of the "walk trap" communities cannot serve as a reliable measure for heterogeneous attachment.

Besides the total number of the "walk trap" communities in a social network, their relative size is also of interest; that is, the number of vertices and agents in each "walk trap" community respectively. One way to do so is to calculate and compare the average size of the "walk trap" communities in the simulated networks.

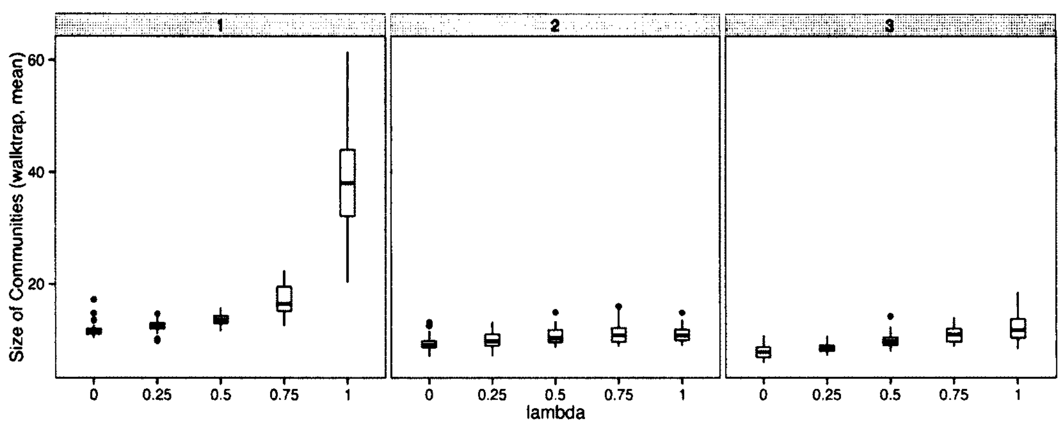


Fig. 19. The relationship of the average size of network communities to λ and m, based on the community detection method of "walk trap".

In order to explore the impacts of λ and m on the average size of "walk trap" communities of the simulated networks respectively, their average size is calculated. The results are presented in Fig. 19, and three findings stand out. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, appears to affect the average size of "fast greedy" communities in the simulated networks. Specifically, when m is controlled the partial correlation between the two is 0.504 and is statistically significant. With larger values of λ, the network-wise the average size of the "walk trap" communities increases as well. In other words, as individual agents become more popularity-oriented, on average they tend to form larger yet relatively more segregated solidary groups in a given social network. Therefore it is suggested that popularity orientation tends to make a social network more sparsely isolated.

Second, after comparing the three subgraphs, significant and negative impacts of individual capacity, m, were found, and the corresponding correlation is -0.516 and is statistically significantly. That is, with better networking media and stronger individual capacity, the average size of the "walk trap" communities decreases accordingly. Finally, it is evident that the average size of network communities are sensitive to the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is suggested that the average size of "walk trap" communities cannot serve as a reliable measure for heterogeneous attachment.

It should be noted that the average size of "walk trap" communities provides only the central tendency about the extent to what individual agents are able to form relatively segregated solidary groups. There is limited knowledge about to what extent the relative size of these communities varies. Yet the variations in the measure of the average size of "walk trap" communities are of critical importance to our understanding about the relative "disparity" in the clustering. To a certain extent, the variations in the average size of "walk trap" communities can reveal the extent to which the simulated networks are evenly clustered.

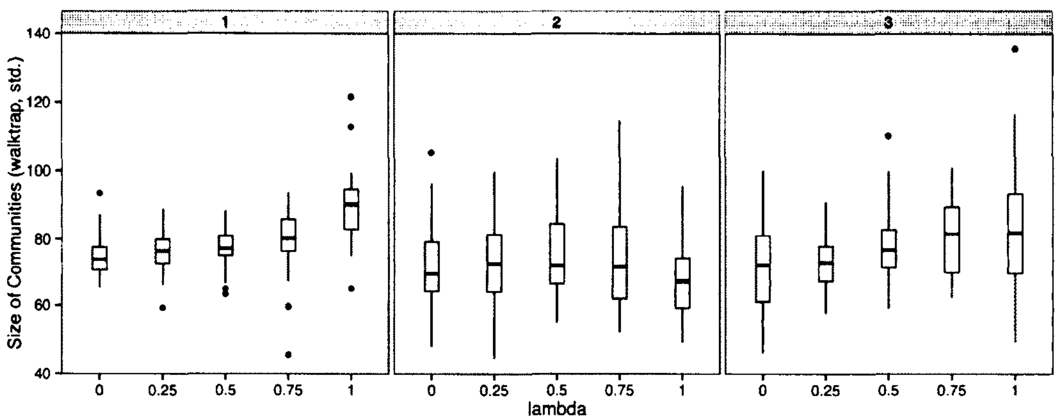


Fig. 20. The relationship of the variations in the sizes of network communities to λ and m, based on the community detection method of "walk trap".

To examine the impacts of λ and m on the variations in the sizes of the "walk trap" communities respectively, the standard deviation was calculated, and the results are presented in Fig. 20. Three conclusions were drawn. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, does not exert a strong impact on the standard deviations of the sizes of the "walk trap" communities in the simulated networks. Specifically, the partial correlation is 0.197 and substantially moderate, though statistically significantly. With larger values of λ, the network-wise variations in the sizes of "walk trap" community exhibits limited changes.

Second, after comparing the three subgraphs, insignificant and individual of individual capacity, m, was found. That is, with better networking media and stronger individual capacity, the resultant social network tends to display no meaningful changes with regard to the variations in sizes of the "walk trap" communities. Finally, it is evident that the standard deviations of the sizes of the "walk trap" communities are sensitive to neither the nature of heterogeneous attachment λ nor individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is suggested that the variations in the sizes of the "walk trap" communities cannot serve as a reliable measure for heterogeneous attachment.

To this point, in this section, the impacts of the nature of heterogeneous attachment λ and individual capacity m were examined on various network statistics of the simulated network at different levels; that is, vertex-, local-, and global level. After examining these associations, three general conclusions were drawn. First, for most network statistics, significant impacts of the nature of individuals' heterogeneous attachment, λ were found. Second, in parallel to the impacts of λ , significant impacts of individual capacity, m, were found. Third, for almost every single network statistics, a joint impact of both λ and m was found. In light of this, none of these vertex-level measures can serve a reliable indicator for heterogeneous attachment, λ.

4.2 Network Topologies: Degree Distribution

Different from the above discussed network statistics (netmetrics), another important way to characterize a network is the degree distribution, which is simply the histogram formed from the degree sequence, with bins of size one, centered on the non-negative integers. Mathematicians and physical statisticians particularly emphasize this approach, in that it provides a natural and reliable summary of the connectivity and other important properties in a graph network. This section therefore explores the extent to which the heterogeneous attachment, λ, affects degree distribution of simulated networks. Since the actual degree distributions of the simulated networks can be modeled against different candidate distributions, a comparison of these different distributions is delivered.

When examining the degree distributions of actual social network, scholars commonly plot them on a logarithmic scale for both axes (i.e., on a log-log scale). The log-log plot provides an intuitive view about the skewness in these distributions, which is of particular interests to researchers.

4.2.1 Fitting Power Law Distribution

When plotting the simulated network in log-log plots, for instance $m = 1$ as demonstrated in Fig.21, interesting patterns with respect to the skewness. Specifically, while the majority of agents are of very low degree, a nevertheless nontrivial number of agents are of much larger degree. When $\lambda =$ 0.75 and 1, a roughly linear decay of the points can be found in each plot and over almost all the range of each distribution. This, in turn, suggests the presence of a power law component to these distributions. In other words, Eq. (32),

$$p(x) \propto x^{-\alpha}, \tag{32}$$

holds approximately true [149]. However, when λ is equaled to smaller values like 0.5, 0.25, and 0, the decay patterns increasingly deviate from a linear one. Therefore, Fig.21 indicates that with decreasing values of λ, the degree distributions of simulated networks increasingly deviate from the power-low distributions.

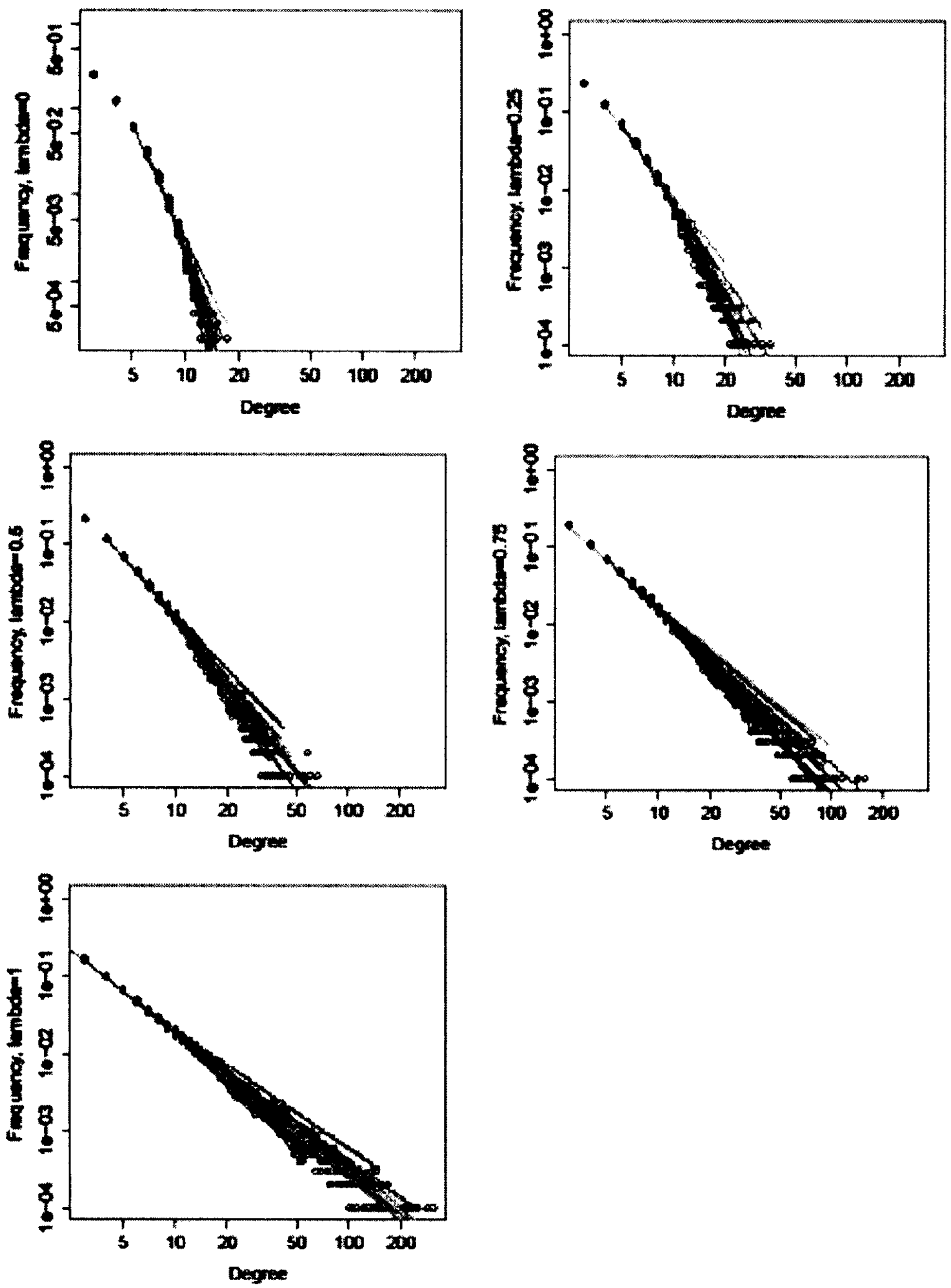


Fig. 21. Degree distribution of a simulated network (log/log), m = 1.

The method proposed by Clauset, Shalizi and Newman [149] was used to fit the degree distributions of the simulated network against the power law distribution. In Fig. 21, the colored lines represent the estimated power law distribution based on the method proposed in Clauset, Shalizi, and Newman [149]. Fig.21 is illustrative in revealing the associations. A comparison of the fitted lines further confirms the relationship between the value of λ and the relative fitness of the power law distribution. The approach proposed by Clauset, Shalizi, and Newman [149] is used in this dissertation to analyze the network data generated from the simulation to test how the values of λ affects the degree distribution, so as well as the system behavior of the network formation process. The below steps were followed:

1. Estimate the parameters of power law distribution: $x_{\min}$ and scaling parameter α.
2. Compare the power law with alternative hypothesis. Here, a likelihood ratio test method is approached. For each alternative, if the calculated likelihood ratio is significantly different from zero, then its sign indicates whether or not the alternative is favored over the power law distribution.

As explained in Chapter 3, the first key statistics involved in estimating a power law distribution are the Kolmogorov-Smirnov statistics. In order to estimate the key parameters of $x_{\min}$ and α, the Kolmogorov-Smirnov or the KS statistic is commonly used as a measurement of quantifying the distance between the two probability distributions, which indicates the maximum distance between the cumulative distribution functions (CDFs) of the measured data and the fitted model. Specifically, the

Kolmogorov-Smirnov statistics are calculated based on Eq. (32). The estimated parameter of $x_{\min}$ then is the one that minimizes the value of the Kolmogorov-Smirnov.

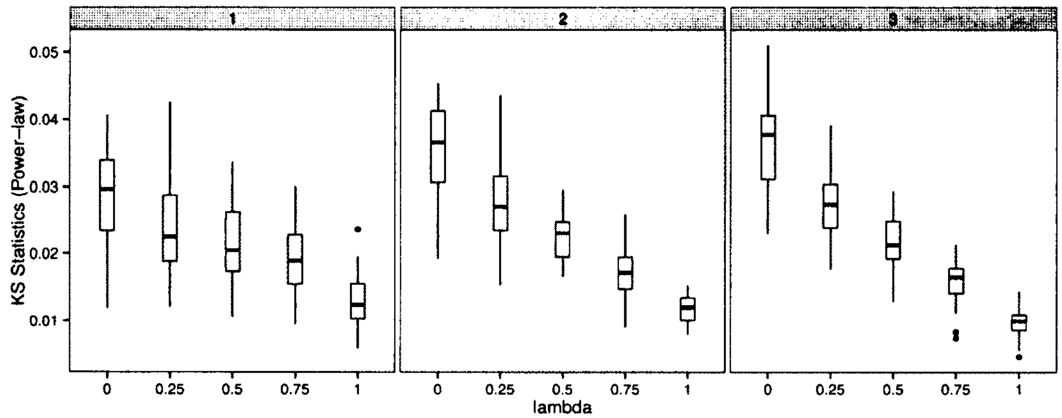


Fig. 22. The relationship of the Kolmogorov-Smirnov statistics to λ and m.

The impacts of λ and m on the Kolmogorov-Smirnov statistics were first explored. After calculating all the KS statistics of the simulated networks with each combination of λ and m values, they were plotted in Fig. 22. Three findings stand out. Firstly, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, appears to affect the value of the Kolmogorov-Smirnov statistics. Specifically, with larger values of λ, the Kolmogorov-Smirnov statistics for power-low distributions decreases accordingly. In other words, as individual agents become more popularity-oriented, the degree distribution of the network they have form is more likely to fit closely to a power law distribution, which is to be expected.

Second, after comparing the three subgraphs, significant impacts of individual capacity, m, were also found. That is, with better networking media and stronger individual capacity, the more the negative impacts of heterogeneous attachment, λ , have been amplified. In other words, with larger values of m, it was observed as having a much stronger impact of heterogeneous attachment, λ. Finally, it is evident that the Kolmogorov-Smirnov statistics are sensitive to the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is suggested that the Kolmogorov-Smirnov statistics cannot serve as a reliable measure for heterogeneous attachment.

After calculating the Kolmogorov-Smirnov statistics, it was then possible to identify the parameter x_{min} in fitting the power law distribution. It was possible to then examine the relationship of λ and m to the values of x_{min}. The results are presented in Fig. 23.

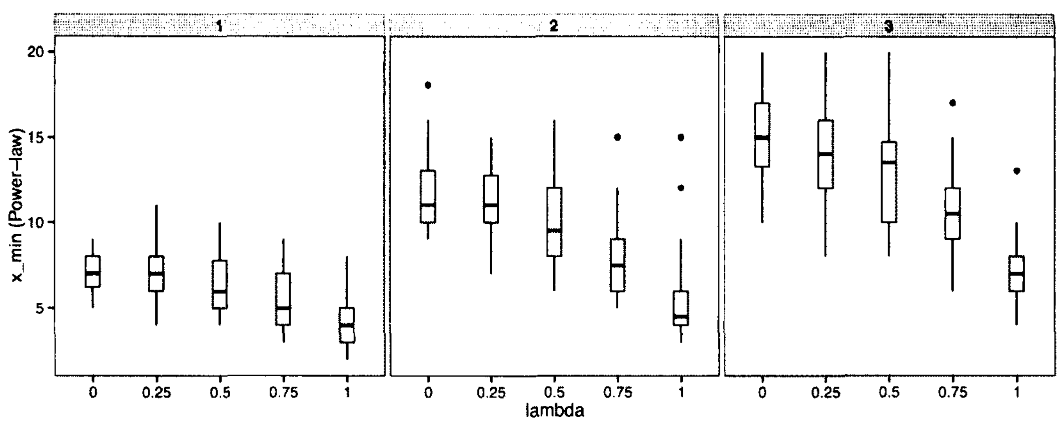


Fig. 23. The relationship of the parameter x_{min} to λ and m.

An examination of Fig. 23 points to three general findings. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ appears to affect the values x_{min}. Specifically, their corresponding partial correlation is -0.616 and is statistically significant. With larger values of λ, $x_{\min}$ for power-low distributions decreases accordingly. In other words, as individual agents become more popularity-oriented, the larger portion of the degree distribution of the network they have formed fits into a power law distribution.

Second, after comparing the three subgraphs, significant positive impacts of individual capacity, m, was found, and the partial correlation between the two is 0.692 and is statistically significant. That is, with better networking media and stronger individual capacity, a smaller portion of the degree distribution of the network they have formed fits into a power law distribution. In other words, with larger values of m, it is observed the degree distribution of resultant simulated networks deviate further from a power law distribution. Finally, it is evident that the values of $x_{\min}$ are a function of the nature of heterogeneous attachment λ and individual capacity m. Considering actual social networks vary simultaneously in these two dimensions, it is suggested that the values of $x_{\min}$ cannot serve as a reliable measure for heterogeneous attachment.

Finally, the skewness or the "shape" of the fitted power law distribution was determined by the parameter α. Clauset, Shalizi and Newman [149] found that method of maximum likelihood provably gives the accurate parameters estimates with the limit of large sample size. Assuming the data are drawn from a distribution following a power law for $x \geq x_{\min}$, the estimation of scaling parameter α under discrete case can be derived through maximum likelihood estimators (MLEs) as describe in Eq. (27) of Chapter 3.

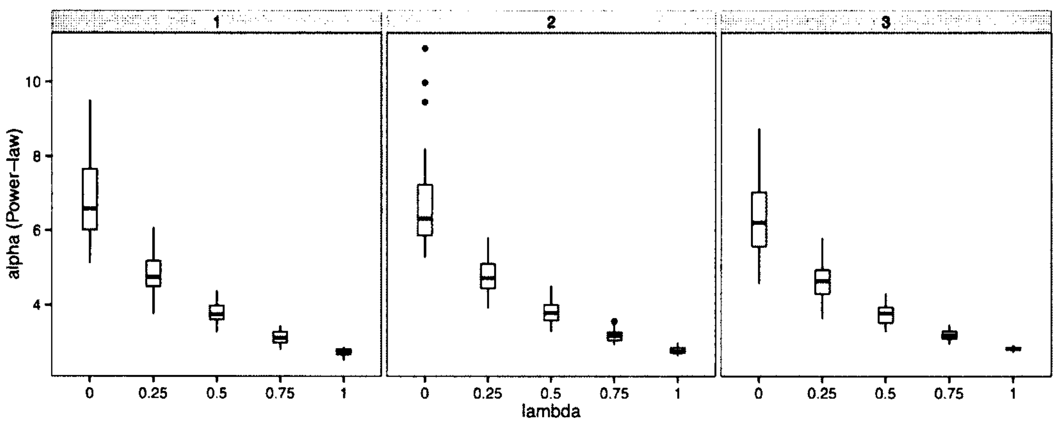


Fig. 24 The relationship of the parameter α to λ and m.

In order to examine the relationship between heterogeneous attachment, λ, and the scalar parameter of power law distribution, α, all the simulated network data were fit against the power law distribution, and the results are presented in Fig. 24. An examination of Fig. 24 points to three general findings. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, appears to strongly and significantly affect the values of the scalar parameter α. Specifically, the partial correlation between the two is -0.873 and is statistically significant. With larger values of λ, the scalar parameter α for power-low distributions decreases accordingly. In other words, as individual agents become more popularity-oriented, the linear decay of the points in the log-log plots appears to be flatter.

Second, after comparing the three subgraphs, no impacts of individual capacity, *m*, is found. That is, with better networking media and stronger individual capacity, then there is no change to the patterns of the linear decay of the points in the log-log plots at all. In other words, the scalar parameter α is found to be invariant to the changes in the individual capacities *m*. Finally, it is evident that the values of the scalar parameter α is sensitive solely to the nature of heterogeneous attachment λ. Considering actual social networks vary simultaneously in these two dimensions, it is suggested that the values of α can serve as a reliable measure for heterogeneous attachment.

4.2.2 Comparing Distributions

While above estimated parameters indicate whether power law is an appropriate distribution in describing simulated networks, they provide no information about relative fitness of the power law distribution against alternative distributions. This section provides a test to see if the simulated data are possibly drawn from a power law distribution. Is it still possible that another distribution, such as a log-normal or a Poisson distribution, might give a fit as good or better? The whole process of methods illustrated in the previous sections was used for the different distribution candidates. After comparing the relative fitness of these distributions, a judgment can be made on whether to accept or reject the hypothesis.

To answer these questions, the simulated network data based on the power-distribution were first compared to that based on the log-normal distribution. Following the specification described in Eq. (30) and Eq. (31), the degree distributions of the simulated networks is fitted against the log-normal distribution. The procedures are similar to those in fitting power law distribution. Rather than the distribution parameters, in this section the relative "fitness" of different distributions is of more interest. Therefore, one-sided tests for the power law distributions and the log-normal distributions respectively were conducted (for details about the test, see Chapter 3).The results are presented in Fig. 25.

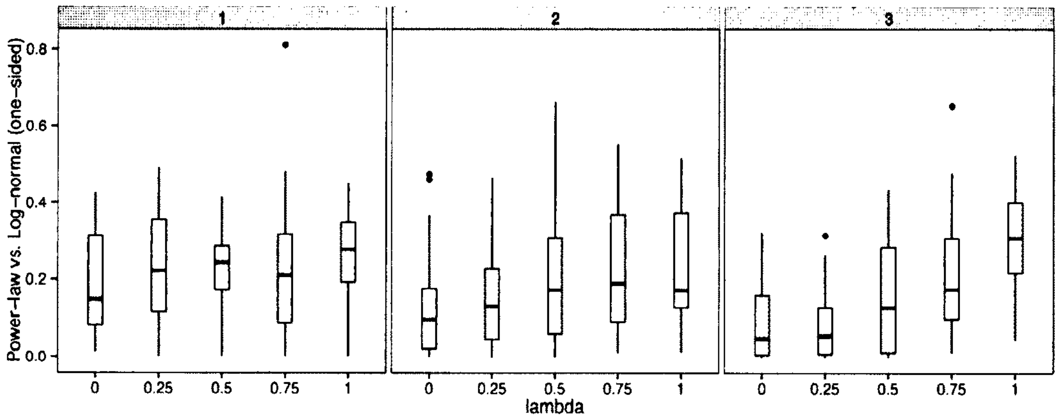


Fig. 25. One-sided test of the power law distribution against the log-normal distribution.

An examination of Fig. 25 leads to two general findings. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, appears to exert no meaningful impact on whether the power law distribution fits the simulated network data better. In other words, with changing values of λ, there is no significant difference in using the power law distribution or the log-normal distribution to fit the simulated network data.

Second, after comparing the three subgraphs, no impacts of individual capacity, *m*, were found. That is, with better networking media and stronger individual capacity, there is no consistent and meaningful change in the one-sided test between the power law distribution and the log-normal distribution in fitting the simulated network data. That is, similar to the nature of heterogeneous attachment, the individual capacities *m*. exert no influences on the question whether the power law distribution or the log-normal distribution is a better fit.

In light of this, it is evident that the test between the power law distribution and the log-normal distribution cannot serve as a reliable measure for heterogeneous attachment.

Besides the log-normal distribution, the Poisson distribution is another important distribution to describe the degree distribution of a given social network, particularly when the links or edges are formed completely randomly. Following the methods proposed by Clauset, Shalizi and Newman [149], the degree distributions of the simulated networks are fitted against the Poisson distribution. The procedures are similar to those in fitting power law distribution. One-sided tests for the power law distributions and the Poisson distributions, respectively, were also conducted. Results are presented in Fig. 26.

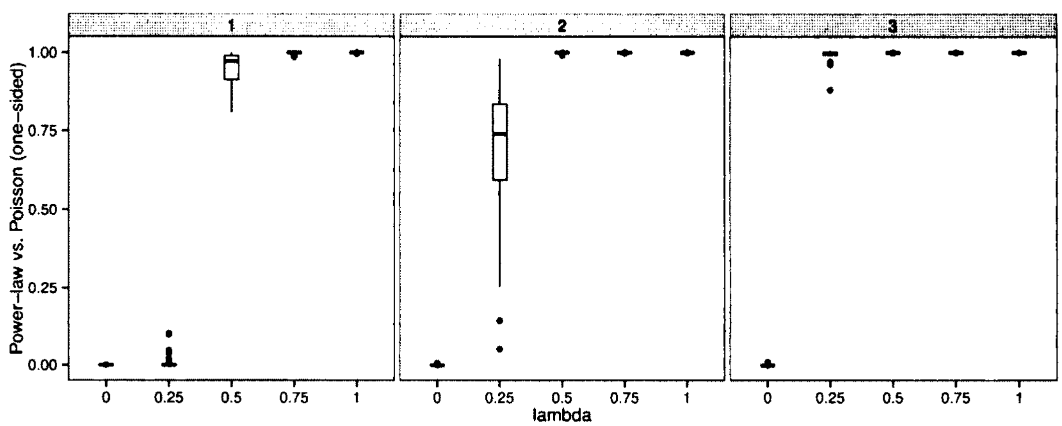


Fig. 26. One-sided test of the power law distribution against the Poisson distribution.

An examination of Fig. 26 leads to three general findings. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, strongly shapes whether the power law distribution outperforms the Poisson distribution in fitting the simulated data. Specifically, with larger values of λ, particularly when $\lambda \geq 0.5$, it was found that the power law distribution is increasingly outperforming the Poisson distribution in fitting the simulated network data.

Second, after comparing the three subgraphs, significant impacts of individual capacity, *m*, were found on the relative fitness of the power law distribution and the Poisson distribution. With better networking media and stronger individual capacity, the impacts of the heterogeneous attachment have been amplified. In other words, with regard to the one-sided test between the power law distribution and the Poisson distribution in fitting the simulated network data, there are significant interactive effects of λ and *m*.

Finally, it is evident that the one-sided test between the power law distribution and the Poisson distribution is sensitive solely to the nature of heterogeneous attachment, λ. Considering actual social networks vary simultaneously in these two dimensions, it is suggested that the one-sided test between the power law distribution and the Poisson distribution cannot serve as a reliable measure for heterogeneous attachment.

Further comparisons were conducted to fitting the simulated network data based on the power-distribution with that based on the exponential distribution. Following the specification described in Eq. (28) and Eq. (29), the degree distributions of the simulated networks is fitted against the exponential distribution. The procedures are similar to those in the fitting power law distribution. It was also conducted one-sided tests for the power law distributions and the exponential distribution respectively, and results are presented in Fig. 27.

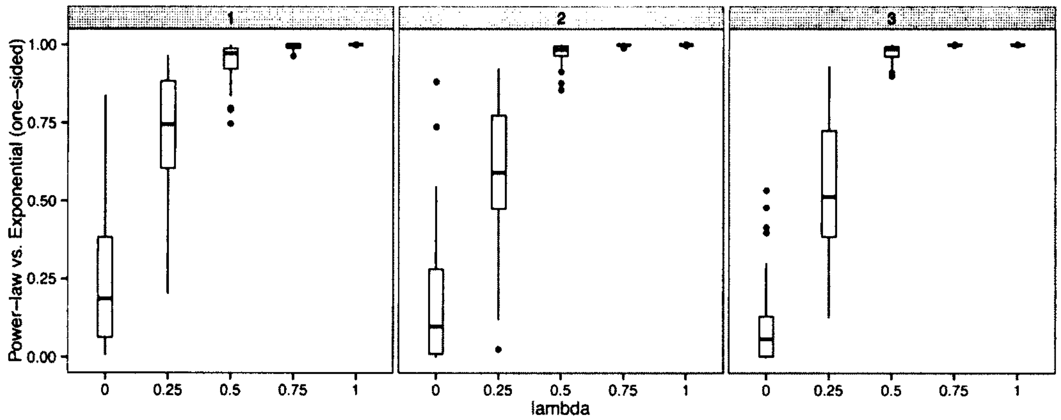


Fig. 27. One-sided test of the power law distribution against the exponential distribution.

An examination of Fig. 27 points to three general findings. First, after examining each of the three subgraphs, it was found that the nature of individuals' heterogeneous attachment, λ, appears to significantly affect the one-sided test between the power law distribution and the exponential distribution. Specifically, with larger values of λ, the p values of the one-sided test between the power law distribution and the exponential distribution approach the value of 1 accordingly. In other words, as individual agents become more popularity-oriented, the power law distribution significantly outperforms the exponential distribution.

Second, after comparing the three subgraphs, no impacts of individual capacity, m, are found. That is, with better networking media and stronger individual capacity, there is no change to the results of the one-sided test between the power law distribution and the exponential distribution. In other words, it was found that the one-sided test between the power law distribution and the exponential distribution is invariant to the changes in the individual capacities m. Finally, it is evident that the one-sided test between the power law distribution and the exponential distribution is sensitive solely to the nature of heterogeneous attachment λ. Considering actual social networks vary simultaneously in these two dimensions, it is suggested that the one-sided test between the power law distribution and the exponential distribution can serve as a reliable measure for heterogeneous attachment.

4.3 Validation and Application

To this point, the relationship of heterogeneous attachment to various popular network statistics such as centrality measures, average path length, and diameter have been examined. Moreover, the extent heterogeneous attachment affect degree distribution of simulated networks was explored. It is evident that the nature of heterogeneous attachment strongly affects the formation of social network, and, more importantly, the scalar parameter of the power law distribution, α can serve as a reliable measure of the heterogeneous attachment.

A remaining question is the extent to which the model proposed in this dissertation helps reveal micro-dynamics in actual social networks. From the perspective of the agent-based modeling, this is about the issue of validation. Broadly speaking, validation in agent-based modeling concerns whether the simulation is a good model of the target. A model which can be relied on to reflect the behavior of the target is "valid." Gilbert and Troitzsch [97] suggest that validity can be ascertained by comparing the output of the simulation with data collected from the target. However, there are several caveats that must be borne in mind.

First, both the model and the target processes are likely to be stochastic (that is, based partly on random factors). Exact correspondence would therefore not be expected on every occasion. Whether the difference between simulation and data from the target is so large as to cast doubt on the model depends partly on the expected statistical distribution of the output measures. Unfortunately, with simulations, these distributions are rarely known and not easy to estimate.

Second, many simulations are path-dependent: the outcomes depend on the precise initial conditions chosen because these affect the 'history' of the simulation. In other words, the outcomes may be very sensitive to the precise values of some of the assumptions in the model.

Third, even if the results obtained from the simulation match those from the target, there may be some aspects of the target that the model cannot reproduce. Fourth, one must not forget the possibility that the model is correct, but the data about the target are incorrect, or, more often, are themselves a result of making assumptions and estimates. Another kind of difficulty arises when the model is intentionally highly abstract. It may be hard to relate the conclusions drawn from the model to any particular data from the target.

Given the ultimate purpose of validation is to explore the extent to which the model does reflect the behaviors of interest, it is suggested that validation of ABMs can be done at both the micro and macro levels, so their falsifiability is really of two separate kinds. Using the Kalick and Hamilton [142] model as a simple example, Gilbert and Troitzsch [97] argue that researchers can validate an agent-based model by answering the following two kinds of questions

(a) Does their assumption about individual agent preferences match what is known about human mate preferences? and

(b) Does their model's generated outcome match what is observed in human populations?

Moss and Edmonds' [152] survey of the literature suggests that virtually all of the ABMs can be validated or compared to data at both of these levels. Of course, a match at both levels increases confidence in the validity of the model. Validation of the micro rules describing individual agent behavior is a task that is especially well suited for social psychology's most familiar and powerful research technique, lab-based experimental studies.

However, it should be noted that the specific methods of validation are contingent highly on the specific research design and the accessibility of the data regarding the questions of interests. The tightness or looseness of the model-data comparisons involved in validation (at either the micro or macro level) is a very difficult issue. A model may be asked to match what Epstein [148, pp. 46] called "stylized facts" or qualitative, generic empirical regularities, such as that residential segregation exists in Schelling [153] or that partner attractiveness correlates Kalick and Hamilton [142]. These are the kinds of broad empirical generalizations that might be the chief results of a meta-analysis of a research area—general summaries of what is empirically known rather than detailed results of a single, specific study.

It is argued that in many cases, this level of empirical validation is sufficient for the main purposes of ABM: the attaining of basic insights such as those offered by the models just mentioned (or many other examples in this dissertation). But in other cases, a much tighter and more precise match to data is demanded. Epstein [56] cited several examples of economic ABMs that have been developed to explain highly specific patterns in data, such as the distribution of firm sizes in the economy. Whether one seeks to validate relatively general, qualitative patterns or to match data in exact quantitative detail depends on the overall goals of a model and on the availability of suitable data sets.

More specifically, with regard to the basic criteria of validations, Casti [77] gives the following axes along which the validity of simulation models can be evaluated:

- Empirical: Does the model agree with observed data that are relevant to the problem under consideration?
- Theoretical: Does the model contradict any established theories?
- Consistency: Does the model contain any logical contradictions?
- Faith: Do specialists in the area being modeled agree that the model produces believable results?
- Testing: Can the model be tested in the real world?

In this study, one enormous challenge is to conduct a micro-level validation of the agent-based model, which requires laboratory experiments testing under what circumstances individuals are popularity- or proclivity-oriented. Therefore, this study turns to the approach of the macro-level validation. Specifically, this study relies on data collected from social media in China to validate the proposed model. After applying findings emerged in previous section, a demonstration of the micro-foundation of the diffusion patterns of different emotions is given.

4.3.1 Emotional Diffusion and Sina Weibo Data

The content in online social media like Twitter or *Sina Weibo* (微博) is mainly recorded in the form of text. Many approaches have been presented to mine sentiments from these texts in recent years. One of them is the lexicon-based method, in which the sentiment of a tweet is determined by counting the number of sentimental words, i.e., positive terms and negative terms. For example, Dodds and Danforth measured the happiness of songs, blogs and presidents [132]. They also employed Amazon Mechanical Turk to score over 10,000 unique English words on an integer scale from 1 to 9, where 1 represents sad and 9 represents happiness. Golder and Macy [133] collected 509 million English tweets from 2.4 million users in Twitter, then measured the positive and negative affects using Linguistic Inquiry and Word Count (LIWC)[4].

While another one is the machine learning based solution, in which different features are considered to perform the task of classification, including terms, smileys, emoticons and etc. The first step is taken by Pang et al. in [77], they treat the sentiment classification of movie reviews simply as a text categorization problem and investigate several typical classification algorithms. According to the experimental results, machine learning based classifiers outperform the baseline method based on human words list. Different from most work which just categorized the emotion into negative and positive, Fan et al. [64] divided the sentiment into four classes, then presented a framework based on emoticons without manually-labeled training tweets and achieved a convincing precision.

This dissertation relies on data provided by Fan et al. [64] to validate the proposed model. Specifically, Fan et al. [64] argue that the following relationship in Twitter-like social networks does not stand for the social interaction, while if two users reply, retweet or mention each other in their tweets for certain times, the online social tie between them is sufficient to present an alternative means of deriving a conventional social network.

[4] For more details, see http://www.liwc.net.

Therefore, Fan et al. [64] constructed an interaction network from the tweets crawled from *Weibo* during April 2010 to September 2010, where interaction means the number that two users retweet or mention each other is larger than a threshold T. From around 70 million tweets and 200,000 users, Fan et al. [64] crawled, an undirected but weighted graph $G(V, E, T)$ was constructed, in which V is the set of users, E represents the set of interactive links among V, and T is the minimum number of interactions on each link. For each link in E, the weight is the sum of retweet or mention times between its two ends in the specified time period.

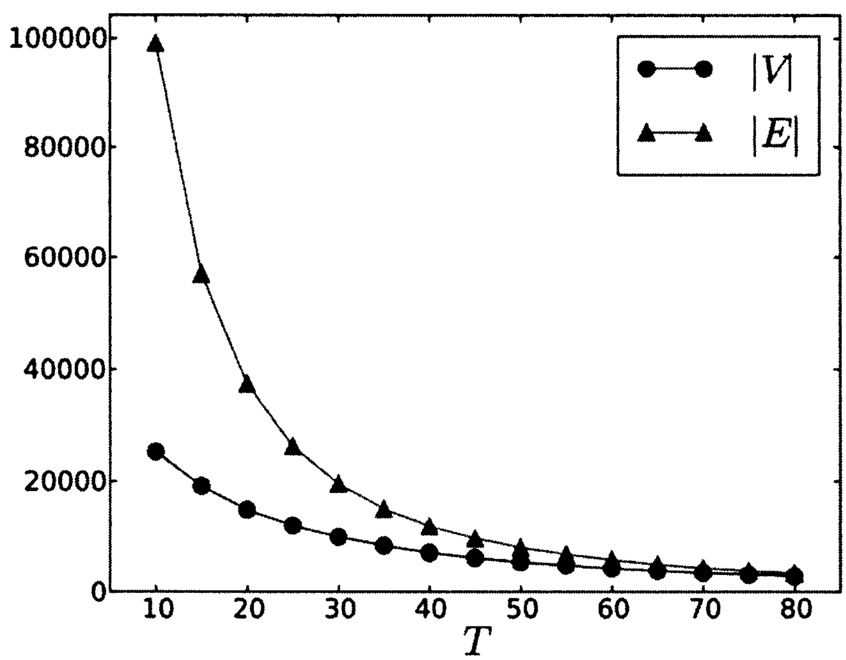
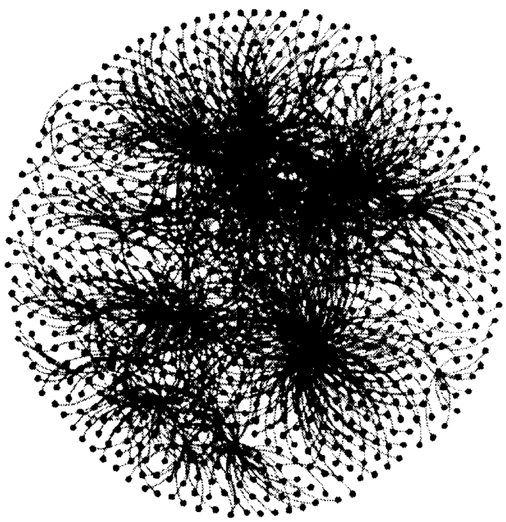
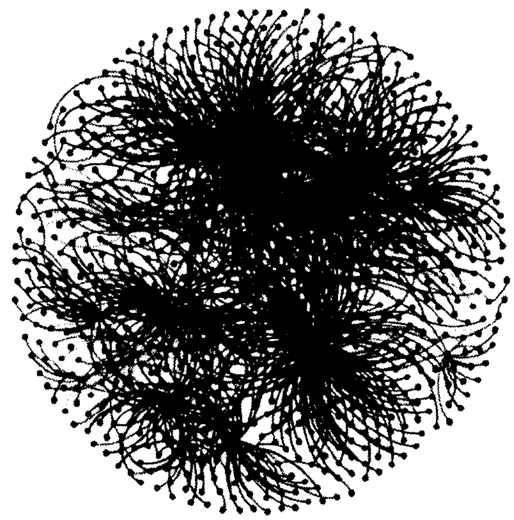


Fig. 28. Tweeters and their Tweets in *Sina Weibo*.

In order to exclude occasional users that are not truly involved in the *Weibo* social network, Fan et al. [64] reserved only those active users in the interaction network that posted more than one tweet every two days, on average, over six months. And to guarantee the validity of users' social interaction, if the number of retweets or mentions between two users was less than T, Fan et al. [64] would omit the connection between them. As shown in Fig. 29, by tuning T Fan et al. [64] were able to obtain networks of different scales. The number of nodes or edges varies for different interaction threshold T. In the following part of the present work, $T = 30$ is set to extract a large enough network with convincing interaction strength. Finally, Fan et al. [64] have set $T = 30$ and then the interaction network G contains 9868 nodes and 19517 links. The resultant dataset is publicly available.



(a) An interaction network.



(b) Node colored by emotions.

Fig. 29. The giant connected cluster of a network sample with $T = 30$ from Fan et al. [64]. (a) is the network structure, in which each node stands for a user and the link between two users represents the interaction between them. Based on this topology, Fan et al. [64] color each node by its emotion, i.e., the sentiment with the maximum tweets published by this node in the sampling period. In (b), the red stands for anger and the green represents joy, the blue stands for sadness and the black represents disgust. The regions of same color indicate that closely connected nodes share the same sentiment.

After locating the active subset of the *Weibo* social network, Fan et al. [64] try to classify and identify the diffusion of different emotions. The emotion is divided into four classes, including anger, sadness, joy, and disgust. Fan et al. [64] then employ the Bayesian classifier developed in previous work. Relying on the method in [77], Fan et al.

[64] use emoticons, which are pervasively used in *Weibo*, to label the sentiment of the tweets. At the first stage, 95 frequently used emoticons are manually labeled by different sentiments and then if a tweet only contains the emoticons of a certain sentiment, it would be labeled with this sentiment. From around 70 million tweets, 3.5 million tweets with valid emoticons were extracted and labeled. Using this data set as a training corpus, a simple but fast Bayesian classifier was built in the second stage to mine the sentiment of the tweets without emoticons, which are about 95% in *Weibo*.

The averaged precision of this classifier is convincing and particularly the large amount of tweets in the experiment can guarantee its accuracy further. Based on this framework, a sampled snapshot of interaction network with $T = 30$ is presented in Fig. 29. As shown in Fig. 29(b), each user is colored by its emotion. Roughly, Fan et al. [64] find that closely connected nodes generally share the same color, indicating emotion correlations in *Weibo* network. Besides, different colors show different clusterings.

4.3.2 Fitting the Power Law Distribution against *Sina Weibo* Data

As revealed in the previous section, after fitting all the simulated network data against the power-distribution, this dissertation finds that while the nature of individuals' heterogeneous attachment, λ, appears to significantly affect the values of the scalar parameter α, it is invariant to the changes in the individual capacities *m*. This, in turn, suggests that the values of the scalar parameter α in the fitted power law distribution can serve as a reliable measure for heterogeneous attachment. Therefore, this dissertation relies on the scalar parameter α to validate the proposed model. To do so, the diffusion of four different emotions against the power law distribution is fitted using the procedures described in Section 4.2.1, and the results are presented in Fig. 30.

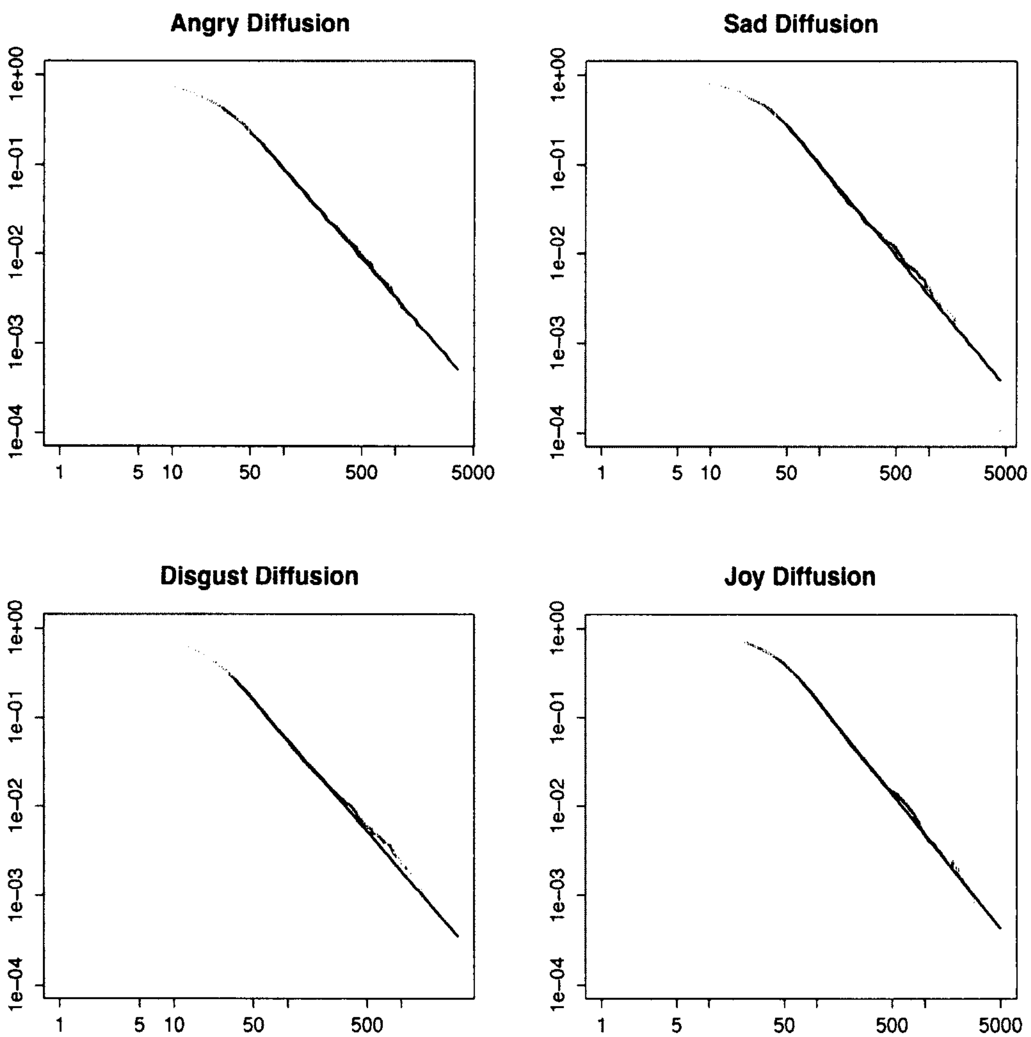


Fig. 30. Fitting power law distribution against the diffusion of four types of emotions: anger, sadness, disgust, and joy.

As revealed in Fig. 30, the grey spots represent the actual degree distribution of the diffusion patterns of the four different emotions, and the red lines highlight the fitted power law distributions. The different slopes across different emotions therefore are their respective "decaying" patterns; that is, the scalar parameter.

While Fig. 30 is intuitive in revealing the relative fitness of the power law distribution, the exact relationship between the diffusion patterns of different emotions is not known. After comparison, it seems that $1.5 < \alpha_{anger} < \alpha_{sadness} < \alpha_{disgust} < \alpha_{joy} < 2.2$. Comparing these values of α with Figure 4.21, they correspond to the values λ of as follows,

$$0.8 > \lambda_{anger} > \lambda_{sadness} > \lambda_{disgust} > \lambda_{joy} > 0.6.$$

In light of this, for diffusion of different emotions, the micro-foundations vary. While for contagion of anger, individuals tend to be more popularity-oriented, in diffusion of joy people are more proclivity-oriented. Although there is no laboratory-based research testing specifically the micro-foundation of the four different emotions, some psychological studies do suggest that there is a stronger homophily effect in joy diffusion than that of negative emotions. For instance, Schaefer, Kornienko, and Fox [154] investigate friend selection mechanisms responsible for similarity in depression among friends, and they find that people tend to withdraw friendship in reaction to such negative emotion like depression.

Mainly there are two important parameters tested in this study for exploring the process of the social network formation: λ and *m*. λ indicates people's intention of social networking strategies and *m* indicates the average people's social capacity. In this study, λ was found affected only by the alpha values when the social network data are fitted into the power law distribution, where the alpha value is the scale parameter of power law density function. Hence, estimating the alpha value provides a possible way of inferring the λ value. Accordingly, different levels of popularity and proclivity could be used for labeling the variant network topology.

Mainly there are two important parameters tested in this study for exploring the process of the social network formation: λ and *m*. λ indicates people's intention of social networking strategies and *m* indicates the average people's social capacity. In this study, λ is found affected only by the alpha values when the social network data are fitted into the power law distribution, where the alpha value is the scale parameter of power law density function. Hence, estimating the alpha value provides a possible way of inferring the λ value. Accordingly, different levels of popularity and proclivity could be used for labeling the variant network topology.

m was found to be variant for all measurements in this study, accordingly, it is not with evidence for picking any index for measuring the *m* parameter. In other words, after calculating the network data affected by *m* and λ, there is no index found that could be used for indicating the *m* value in the empirical network data. For proposing a reasonable way of validating the *m* value, an empirical study could be conducted, however, no empirical data were available for this validation. The study only limited the value of *m* to a maximum of three. Though it is expected that the current shown patterns would continue when *m* is greater than three this would need to be empirically tested. This further comparison has been left for future studies.

4.4 Conclusion and Discussion

To what extent does agent heterogeneity affect actual formation of social networks? When people are popularity- and proclivity-orientated, will the resultant social network better connected, equally connected, or efficiently connected? Is there any reliable measure to detect the nature of heterogeneous attachment at the individual level? These are the critical questions this chapter has intended to answer. So in order to answer these questions, this chapter examined the relationship of heterogeneous attachment to various popular network statistics such as centrality measures, average path length, and diameter. Then this chapter also explored to what extent heterogeneous attachment affects degree distribution of simulated networks. Finally, a comparison of different measures was delivered.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

The emergence of relational interactions is an important phenomenon in our society from economic transactions to social networking. Unfortunately, in contemporary social network analysis, the studies of network formation are skewed toward either an inductive statistical approach—empirical investigations of various factors in shaping actual formation of social networks, or a deductive approach—abstract inference of mathematical models in formalizing the social dynamics. As a consequence, our knowledge of network formation is regrettably divided and inadequately integrated.

By introducing an emergent conceptualization of network formation, this study intends to bridge the competing inductive and deductive approaches to network formation, and attempts to systematically and empirically explore how different networking heuristics of agents lead to varying patterns of network formation. Particularly, it is argued that one way to advance agent-based modeling (ABM) of network formation is to "bring agent heterogeneity back in." By explicitly modeling agent heterogeneity in our social networks, this study systematically evaluates to what extent individuals' concern over proclivity-popularity affects various attributes of network.

Specifically, this study tries to answer the following critical questions: (1) How can social networks emerge with heterogeneous agents (i.e., agents with varying popularity and proclivity orientation)? (2) As agents become increasingly popularity- or proclivity oriented, how will various network attributes at vertex-, dyad-, local-, and system-level change? (3) How can findings that emerged from an ABM analysis help advance our understanding about real-world social networks?

This concluding chapter first summarizes the empirical findings discussed in the previous chapters. Subsequently, key empirical and theoretical implications of the findings will be addressed. Finally, the limitations of this study and possible directions for future studies will be discussed.

5.1 Summary of Major Findings

The exploration of network formation in this dissertation starts with the psychological needs of agents. Most deductive approaches to network formation appear to be problematic since individual psychologies are mainly overlooked. In order to model the emergence of social network in reliable and valid way, this study first clarifies the major psychological foundations of social networking. By specifying these psychological foundations, this study was able to determine their corresponding networking heuristics in the process of network formation. Specifically, there are two individual psychologies that are particularly relevant in network formation.

A first important psychological need for individuals' social networking is the need for cognition. Fundamentally, the need for cognition is instrumental and describes how individuals' tendency to use their personal ties to acquire new information and resources. In light of this, people high in need for cognition are strongly motivated to connect with the "experts" or the "rich" in a given social network.

Second, the need for affect is another critical motivation for individuals' social networking. Specifically, the need for affect points to the emotional construct and captures the degree to which people enjoy experiencing strong emotions in social networking. In contrast to need for cognition, people high in need for emotion tend to possess a "biased reasoning" in their social networking.

After clarifying the key psychological motivations, this study explores these motivations' corresponding network heuristics. An examination of the competing psychological needs suggests that people will rely on different strategies in networking. Varying strategies can be used to satisfy different psychological needs. Specifically, this dissertation suggests that while people high in need for cognition tend to adopt a strategy of preferential attachment, people high in need for affect are more likely to use a strategy of homophily attachment. Together, individuals will use a combination of popularity- and proclivity-oriented strategies in their social networking.

However, a review of the existing literature on network formation suggests that recent studies, inspired by Barabási and Albert [35], focus primarily on the preferential attachment (i.e., popularity-oriented strategy) and overlook the potential impacts of agent heterogeneity in network formation. Recognizing the limitation, this dissertation formulates a heterogeneous attachment model of network formation. This model starts an assumption of heterogeneous agents; that is, agents are inherently different from each other on certain aspects. This assumption of agent heterogeneity in turn indicates that when making decisions about social networking, people are driven by both the instrumental calculation of connecting with the popular and the intrinsic affection of joining the like.

Based on this theory of heterogeneous attachment, this study constructed an agent-based model to simulate the dynamic emergence of social networks. Throughout the simulation, the degree of heterogeneous attachment is varied. Specifically, ranges between 0 and 1, and larger *l* indicate a stronger popularity orientation. Moreover, to explore the potential impacts of costs associated with networking, this study also varies ABM simulation on the average number of agent connections (i.e., *m*). With a different configuration of *l* and *m*, this study was able to simulate a repertoire of social networks with differing local dynamics.

In order to explore the relationship of heterogeneous attachment to network formation, this study analyzed the statistical association between *l* and *m*, on one hand, and various vertex-level, dyadic, local, and global attributes on the other hand. The results which emerged from partial correlation analysis indicate that all these network statistics vary significantly in accordance with different values of both *l* and *m*. The findings suggest that heterogeneous attachment significantly shapes the dynamic emergence of social network.

Yet, since all the key network statistics correlate with *both l* and *m*, it is hard, if not impossible, to evaluate the extent to which a social network is proclivity- or popularity-driven according to these statistics. To solve this problem, this study further examined the relationship between *l* and *m*, on one hand, and the degree distribution of simulated networks on the other hand. Particularly, the simulated networks are fitted into different distributions (i.e., power law, log-normal, and exponential distributions), and the key distributions are calculated and compared. The results of partial correlation analysis suggest that the key parameter of power law distribution (i.e., *l*) can serve as a convenient indicator of heterogeneous attachment. Specifically, not only is *a* insensitive to different values of *m*, *m* bears a significant and negative association with *a*.

Finally, as a preliminary effort of verification and application, this study examined the above relationship with real world data of social networking. Particularly, using data from *Sina Weibo* (a Chinese counterpart of Twitter), this study explores the diffusion networks of four types of emotions (i.e., anger, sad, disgust, and joy). After applying the findings emerged from simulated studies, this study reveals that the diffusion of anger tends to be the most popularity-driven, while joy is the least.

In summary, this dissertation has presented an ABM approach to exploration of dynamic emergence of social network, developed an integrated model of heterogeneous model, and provided a way to evaluate the degree heterogeneous attachment empirically. What does this study imply for future studies on social network analysis as well as for the studies on agent-based modeling? The following part of this chapter will address this important question.

5.2 Implications for Studies on Social Networks

The findings presented in this dissertation have important implications with respect to both social network analysis and agent-based modeling (ABM). Although most scholars focus either on an inductive empirical approach or a deductive mathematical approach, this dissertation points to a promising way to integrate these two competing paradigms; that is, the ABM approach.

On one hand, the ABM approach introduces more reliable computational simulation into empirical analysis of social networks and renders a rigorous way to test various findings emerged from empirical studies. On the other hand, the ABM approach can help develop more flexible frameworks, in which more complicated social theories can be modeled in a computational way.

5.3 Future Studies and Recommendations

It should be noted that this study assumes that the formation of social networks is unidirectional. Moreover, proclivity and popularity are assumed to be changing along a single continuum. That is, as proclivity increases, popularity decreases, and vice versa. However, relaxations on these two key assumptions could greatly expand the current study. First, if the formation of social networks is directional, the social dynamics could become quite complicated. In this case, in-degree and out-degree network statistics should be distinguished. Moreover, an introduction of directional assumption complicates the comparison and calculation of popularity and proclivity. For instance, in-degree popularity and out-degree popularity would have to be differentiated. Similarly, a distinction between in-degree proclivity and out-degree proclivity ought to be made. In combination, four sets of popularity-proclivity shape the dynamic formation of social network. Second, popularity and proclivity can change in the same direction. That is, people can be strongly motivated by both instrumental and intrinsic considerations. This, in turn, also leads to four different combinations of popularity-proclivity. Future studies are thus called for to explore these interesting dynamics.

As revealed in this study, the key parameter of power law distribution α is invariant under different values of m. It is also possible to construct invariant parameters using ratios of existing parameters. Specifically, the variation coefficient is defined as the ratio of the standard deviation to the mean. As for this particular study, the variation coefficient can be quite useful since the standard deviations of various network statistics and distribution parameters can be somehow correlated to their respective means. When the means are widely different, the variation coefficient is preferred. However, in this study, the key question is how λ and m jointly affect the dynamic emergence of social networks. Thus analyses based on the variation coefficients are not employed due to the static nature of the means in the generated networks. However, for the analysis of real world data, whose means of network statistics can vary considerably, the variation coefficient is recommended.

Future studies might be devoted to the ways in which network heuristics other than popularity- and proclivity-orientation might affect various vertex-, dyadic-, local-, and global-level network attributes. This dissertation argues that driven by the need for cognition and the need for affect, individuals can follow two simple strategies—connecting with people with similar interests and connecting with the rich—to form their social networks. Yet, additional local strategies like "a friend of my friend is a friend" can also serve as important heuristics to form social networks. Future studies, therefore, can benefit more from including more local strategies.

Second, future studies can benefit from modeling adaptive behaviors of agents in a networked setting. In this study, only network formation was explored. Network formation, though important, consists of only one aspect of how individuals interact with each other. After a network has been formed, individuals can choose to maintain or disconnect social ties. The dynamics associated with network evaluation require more systematic exploration. As revealed in this study, ABM approach is particularly useful in investigating these adaptivity problems.

Third, additional empirical studies are called for to validate and verify the model developed in this study. In this study, only second-hand data of social network were used to validate and verify the agent-based model. The current study can be strongly strengthened by more comprehensive data.

REFERENCES

[1] R. S. Burt, *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press, 1992.

[2] R. S. Burt, *et al.*, "Social Network Analysis: Foundations and Frontiers on Advantage," *Annual Review of Psychology,* vol. 64, pp. 527-547, 2013.

[3] P. J. Carrington, *et al.*, Eds., *Models and Methods in Social Network Analysis*. Cambridge and New York: Cambridge University Press, 2005.

[4] J. Scott, *Social Network Analysis: A Handbook*, 2nd ed. London: Sage, 2011.

[5] T. A. B. Snijders, "Statistical Models for Social Networks," *Annual Review of Sociology,* vol. 37, pp. 131-153, 2011.

[6] U. Brandes, *et al.*, "What is Network Science?," *Network Science,* vol. 1, pp. 1-15, 2013.

[7] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge and New York: Cambridge University Press, 1994.

[8] U. Brandes and T. Erlebach, Eds., *Network Analysis: Methodological Foundations*. New York: Springer, 2005.

[9] D. Knoke, *Political Networks: The Structural Perspective*. Cambridge and New York: Cambridge University Press, 1990.

[10] D. A. Kenny, *et al.*, *Dyadic Data Analysis*. New York, NY: The Guilford Press, 2006.

[11] M. O. Jackson, *Social and Economic Networks*. Princeton, NJ: Princeton University Press, 2010.

[12] M. O. Jackson and Y. Zenou, "Games on Networks," in *Handbook of Game Theory with Economic Applications*. vol. 4, H. P. Young and S. Zamir, Eds., 2015, pp. 95-163.

[13] Z. Maoz, *Networks of Nations: The Evolution, Structure, and Impact of International Networks, 1816-2001*. Cambridge and New York: Cambridge University Press, 2010.

[14] S. Bornholdt and H. G. Schuster, Eds., *Handbook of Graphs and Networks: From the Genome to the Internet*. Wiley-VCH, 2003.

[15] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function.* Cambridge and New York: Cambridge University Press, 2010.

[16] A. Barrat, *et al.*, *Dynamical Processes on Complex Networks.* Cambridge and New York: Cambridge University Press, 2008.

[17] A. K. Naimzada, *et al.*, Eds., *Networks, Topology and Dynamics: Theory and Applications to Economics and Social Systems.* Berlin: Springer, 2009.

[18] G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology.* Oxford and New York: Oxford University Press, 2007.

[19] T. Gross and H. Sayama, Eds., *Adaptive Networks: Theory, Models and Applications.* London and New York: Springer, 2009.

[20] M. Newman, *et al.*, Eds., *The Structure and Dynamics of Networks.* Princeton, NJ: Princeton University Press, 2006.

[21] A. Abraham, *et al.*, Eds., *Computational Social Network Analysis: Trends, Tools and Research Advances.* London: Springer, 2010.

[22] D. Easley and J. M. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* Cambridge and New York: Cambridge University Press, 2010.

[23] P. F. Lazarsfeld and R. K. Merton, "Friendship as A Social Process: A Substantive and Methodological Analysis," in *Freedom and Control in Modern Society*, M. Berger and T. Abel, Eds., New York: Van Nostrand, 1954, pp. 18-66.

[24] T. L. Huston and G. Levinger, "Interpersonal Attraction and Relationships," *Annual Review of Psychology,* vol. 29, pp. 115-156, 1978.

[25] R. S. Burt, "Models of Network Structure," *Annual Review of Sociology,* vol. 6, pp. 79-141, 1980.

[26] M. McPherson, *et al.*, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology,* vol. 27, pp. 415-444, 2001.

[27] G. Miller, "Social Scientists Wade Into the Tweet Stream," *Science,* vol. 333, pp. 1814-1815, 2011.

[28] S. P. Borgatti, *et al.*, "Network Analysis in the Social Sciences," *Science,* vol. 323, pp. 892-895, 2009.

[29] S. A. Golder and M. W. Macy, "Digital Footprints: Opportunities and Challenges for Online Social Research," *Annual Review of Sociology,* vol. 40, pp. 129-152, 2014.

[30] R. Ling, *New Tech, New Ties: How Mobile Communication is Reshaping Social Cohesion.* Cambridge, MA: The MIT Press, 2008.

[31] P. Parigi and L. Sartori, "The Political Party as a Network of Cleavages: Disclosing the Inner Structure of Italian Political Parties in the Seventies," *Social Networks,* vol. 36, pp. 54-65, 2014.

[32] S. D. McClurg and D. Lazer, "Political Networks," *Social Networks,* vol. 36, pp. 1-4, 2014.

[33] S. D. McClurg, "Social Networks and Political Participation: The Role of Social Interaction in Explaining Political Participation," *Political Research Quarterly,* vol. 56, pp. 449-464, 2003.

[34] M. McPherson, *et al.*, "Social Isolation in America: Changes in Core Discussion Networks over Two Decades," *American Sociological Review,* vol. 71, pp. 353-375, 2006.

[35] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science,* vol. 286, pp. 509-512, 1999.

[36] R. Albert and A.-L. Barabási, "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics,* vol. 74, pp. 47-97, 2002.

[37] A.-L. Barabási, "Scale-Free Networks: A Decade and Beyond," *Science,* vol. 325, pp. 412-413, 2009.

[38] D. Lazer, *et al.*, "Computational Social Science," *Science,* vol. 323, pp. 721-723, 2009.

[39] Y.-Y. Liu, *et al.*, "Controllability of Complex Networks," *Nature,* vol. 473, pp. 167-173, 05/12/print 2011.

[40] K. Heyman, "Making Connections," *Science,* vol. 313, pp. 604-606, 2006.

[41] D. J. Watts and S. H. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature,* vol. 393, pp. 440-442, 1998.

[42] D. J. Watts, "Networks, Dynamics, and the Small-World Phenomenon," *American Journal of Sociology,* vol. 105, pp. 493-527, 1999.

[43] D. J. Watts, "The 'New' Science of Networks," *Annual Review of Sociology,* vol. 30, pp. 243-270, 2004.

[44] M. O. Jackson and A. Watts, "The Evolution of Social and Economic Networks," *Journal of Economic Theory,* vol. 106, pp. 265-295, 2002.

[45] Y. Bramoullé, *et al.*, "Homophily and Long-run Integration in Social Networks," *Journal of Economic Theory,* vol. 147, pp. 1754-1786, 2012.

[46] F. Bloch and M. O. Jackson, "The Formation of Networks with Transfers among Players," *Journal of Economic Theory,* vol. 133, pp. 83-110, 2007.

[47] A. Watts, "A Dynamic Model of Network Formation," *Games and Economic Behavior,* vol. 34, pp. 331-341, 2001.

[48] F. Bergenti, *et al.*, "Agent-Based Interpretations of Classic Network Models," *Computational and Mathematical Organization Theory,* vol. 19, pp. 105-127, 2013.

[49] L. Hamill and N. Gilbert, "Social Circles: A Simple Structure for Agent-Based Social Network Models," *Journal of Artificial Societies and Social Simulation,* vol. 12, p. 3, 2009.

[50] R. Axelrod, *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton, NJ: Princeton University Press, 1997.

[51] R. Axelrod, "Advancing the Art of Simulation in the Social Sciences," *Complexity,* vol. 3, pp. 16-22, 1997.

[52] K. M. Carley, "Computational Approaches to Sociological Theorizing," in *Handbook of Sociological Theory*, J. H. Turner, Ed., ed: Springer US, 2001, pp. 69-83.

[53] L. E. Cederman, "Computational Models of Social Forms: Advancing Generative Process Theory," *American Journal of Sociology,* vol. 110, pp. 864-893, 2005.

[54] S. de Marchi and S. E. Page, "Agent-based Modeling," in *The Oxford Handbook of Political Methodology*, J. M. Box-Steffensmeier, *et al.*, Eds., Oxford: Oxford University Press, 2008, pp. 71-94.

[55] S. N. Durlauf, "Growing Artificial Societies," *Complexity,* vol. 2, pp. 47-49, 1997.

[56] J. M. Epstein, *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton and Oxford: Princeton University Press, 2006.

[57] N. Gilbert, *Agent-Based Models*. Los Angeles, CA: Sage, 2008.

[58] J. H. Miller and S. E. Page, *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, Oxford: Princeton University Press, 2007.

[59] R. K. Sawyer, "Emergence in Sociology: Contemporary Philosophy of Mind and Some Implications for Sociological Theory," *American Journal of Sociology*, vol. 107, pp. 551-585, 2001.

[60] A. H. Shirazi, *et al.*, "Transparency Effect in the Emergence of Monopolies in Social Networks," *Journal of Artificial Societies and Social Simulation*, vol. 16, pp. 1-9, 2013.

[61] A. Namaki, *et al.*, "Network Analysis of a Financial Market based on Genuine Correlation and Threshold Method," *Physica A: Statistical Mechanics and its Applications*, vol. 390, pp. 3835-3841, 2011.

[62] S. N. Dorogovtsev, *et al.*, "Structure of Growing Networks with Preferential Linking," *Physical Review Letters*, vol. 85, pp. 4633-4636, 2000.

[63] P. L. Krapivsky and S. Redner, "Network Growth by Copying," *Physical Review E*, vol. 71, p. 36118, 2005.

[64] R. Fan, *et al.*, "Anger is More Influential than Joy: Sentiment Correlation in Weibo," 2013.

[65] F. Xiong and Y. Liu, "Opinion Formation on Social Media: An Empirical Approach," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 24, pp.013130, 2014.

[66] M. A. Smith, *et al.*, "Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters," *Pew Researcher Center Report*, 2014.

[67] T. K. Ahn, *et al.*, "Expertise and Bias in Political Communication Networks," *American Journal of Political Science*, vol. 57, pp. 357-373, 2013.

[68] T. K. Ahn, *et al.*, "Communication, Influence, and Informational Asymmetries among Voters," *Political Psychology*, vol. 31, pp. 763-787, 2010.

[69] S. Aral, *et al.*, "Engineering Social Contagions: Optimal Network Seeding in the Presence of Homophily," *Network Science*, vol. 1, pp. 125-153, 2013.

[70] K. Arceneaux, *et al.*, "Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment," *Political Analysis*, vol. 14, pp. 37-62, 2006.

[71] K. Arceneaux and R. J. Vander Wielen, "The Effects of Need for Cognition and Need for Affect on Partisan Evaluations," *Political Psychology,* vol. 34, pp. 23-42, 2013.

[72] D. Enemark, *et al.*, "Knowledge and Networks: An Experimental Test of How Network Knowledge Affects Coordination," *Social Networks,* vol. 36, pp. 122-133, 2014.

[73] R. Huckfeldt and J. Sprague, "Networks in Context: The Social Flow of Political Information," *American Political Science Review,* vol. 81, pp. 1197-1216, 1987.

[74] R. Huckfeldt, *et al.*, "Political Environments, Cohesive Social Groups, and the Communication of Public Opinion," *American Journal of Political Science,* vol. 39, pp. 1025-1054, 1995.

[75] R. Huckfeldt, *et al.*, "Noise, Bias, and Expertise in Political Communication Networks," *Social Networks,* vol. 36, pp. 110-121, 2014.

[76] C. Panagopoulos, "I've Got My Eyes on You: Implicit Social-Pressure Cues and Prosocial Behavior," *Political Psychology,* vol. 35, pp. 23-33, 2014.

[77] M. S. Granovetter, "The Strength of Weak Ties," *American Journal of Sociology,* vol. 78, pp. 1360-1380, 1973.

[78] M. S. Granovetter, "The Strength of Weak Ties: A Network Theory Revisited," *Sociological Theory,* vol. 1, pp. 201-233, 1983.

[79] Joel M. Podolny, "Networks as the Pipes and Prisms of the Market," *American Journal of Sociology,* vol. 107, pp. 33-60, 2001.

[80] S. Redner, "How Popular is Your Paper? An Empirical Study of the Citation Distribution," *The European Physical Journal B: Condensed Matter and Complex Systems,* vol. 4, pp. 131-134, 1998.

[81] M. E. J. Newman, "Clustering and Preferential Attachment in Growing Networks," *Physical Review E,* vol. 64, p. 25102, 2001.

[82] S. H. Yook, *et al.*, "Weighted Evolving Networks," *Physical Review Letters,* vol. 86, pp. 5835-5838, 2001.

[83] R. Huckfeldt, *et al.*, *Political Disagreement: The Survival of Diverse Opinions within Communication Networks*. Cambridge and New York: Cambridge University Press, 2004.

[84] J. M. Pujol, *et al.*, "How Can Social Networks Ever Become Complex? Modelling the Emergence of Complex Networks from Local Social Exchanges," *Journal of Artificial Societies and Social Simulation,* vol. 8, p. 12, 2005.

[85] L. H. Wong, *et al.*, "A Spatial Model for Social Networks," *Physica A: Statistical Mechanics and its Applications,* vol. 360, pp. 99-120, 2006.

[86] G. Robins, *et al.*, "Network Models for Social Selection Processes," *Social Networks,* vol. 23, pp. 1-30, 2001.

[87] M. E. J. Newman and M. Girvan, "Mixing Patterns and Community Structure in Networks," in *Statistical Mechanics of Complex Networks.* vol. 625, R. Pastor-Satorras, *et al.*, Eds., Berlin: Springer, 2003, pp. 66-87.

[88] P. Parigi and W. Henson II, "Social Isolation in America," *Annual Review of Sociology,* vol. 40, pp. 153-171, 2014.

[89] M. A. Xenos and K. A. Foot, "Politics as Usual, or Politics Unusual? Position Taking and Dialogue on Campaign Websites in the 2002 U.S. Elections," *Journal of Communication,* vol. 55, pp. 169-185, 2005.

[90] J. Brundidge, *et al.*, "The 'Deliberative Digital Divide:' Opinion Leadership and Integrative Complexity in the U.S. Political Blogosphere," *Political Psychology,* vol. 35, pp. 741-755, 2014.

[91] K. Lee, *et al.*, "An Agent-Based Competitive Product Diffusion Model for the Estimation and Sensitivity Analysis of Social Network Structure and Purchase Time Distribution," *Journal of Artificial Societies and Social Simulation,* vol. 16, p. 3, 2013.

[92] R. N. Lupton, *et al.*, "The Moderating Impact of Social Networks on the Relationships among Core Values, Partisanship, and Candidate Evaluations," *Political Psychology*, forthcoming.

[93] D. A. Siegel, "Social Networks and the Mass Media," *American Political Science Review,* vol. 107, pp. 786-805, 2013.

[94] M. E. J. Newman and J. Park, "Why Social Networks are Different from Other Types of Networks," *Physical Review E,* vol. 68, p. 36122, 2003.

[95] C. M. Macal and M. J. North, "Tutorial on Agent-Based Modelling and Simulation," *Journal of Simulation,* vol. 4, pp. 151-162, 2010.

[96] D. F. Jung and D. A. Lake, "Markets, Hierarchies, and Networks: An Agent-Based Organizational Ecology," *American Journal of Political Science,* vol. 55, 2011.

[97] N. Gilbert and K. G. Troitzsch, *Simulation for the Social Scientist*. Philadelphia: Open University Press, 1999.

[98] Y. Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press, 2006.

[99] P. DiMaggio, *et al.*, "Social Implications of the Internet," *Annual Review of Sociology*, vol. 27, pp. 307-336, 2001.

[100] K. Lewis, *et al.*, "Social Selection and Peer influence in an Online Social Network," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 68-72, 2012.

[101] S. González-Bailón, *et al.*, "The Dynamics of Protest Recruitment through an Online Network," *Scientific Reports*, vol. 1, p. 197, 2011.

[102] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. New York: Springer, 2009.

[103] J. Moody and D. R. White, "Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups," *American Sociological Review*, vol. 68, pp. 103-127, 2003.

[104] J. Yap and N. Harrigan, "Why Does Everybody Hate Me? Balance, Status, and Homophily: The Triumvirate of Signed Tie Formation," *Social Networks*, vol. 40, pp. 103-122, 2015.

[105] S. M. Goodreau, "Advances in Exponential Random Graph (p*) Models Applied to a Large Social Network," *Social Networks*, vol. 29, pp. 231-248, 2007.

[106] G. Robins, *et al.*, "An Introduction to Exponential Random Graph (p*) Models for Social Networks," *Social Networks*, vol. 29, pp. 173-191, 2007.

[107] G. Robins, *et al.*, "Recent Developments in Exponential Random Graph (p*) Models for Social Networks," *Social Networks*, vol. 29, pp. 192-215, 2007.

[108] W. An and S. Schramski, "Analysis of Contested Reports in Exchange Networks Based on Actors' Credibility," *Social Networks*, vol. 40, pp. 25-33, 2015.

[109] A. K. Rider, *et al.*, "Networks' Characteristics Are Important for Systems Biology," *Network Science*, vol. 2, pp. 139-161, 2014.

[110] M. P. McAssey and F. Bijma, "A Clustering Coefficient for Complete Weighted Networks," *Network Science*, forthcoming.

[111] P. Cunningham, *et al.*, "Characterizing Ego-Networks using Motifs," *Network Science,* vol. 1, pp. 170-190, 2013.

[112] J. R. Bell, "Subgroup Centrality Measures," *Network Science,* vol. 2, pp. 277-297, 2014.

[113] J. M. Badham, "Commentary: Measuring the Shape of Degree Distributions," *Network Science,* vol. 1, pp. 213-225, 2013.

[114] P. P. Zubcsek, *et al.*, "Information Communities: The Network Structure of Communication," *Social Networks,* vol. 38, pp. 50-62, 2014.

[115] D. Savage, *et al.*, "Anomaly Detection in Online Social Networks," *Social Networks,* vol. 39, pp. 62-70, 2014.

[116] B. A. Desmarais, *et al.*, "Measuring Legislative Collaboration: The Senate Press Events Network," *Social Networks,* vol. 40, pp. 43-54, 2015.

[117] B. S. Anderson, *et al.*, "The Interaction of Size and Density with Graph-level Indices," *Social Networks,* vol. 21, pp. 239-267, 1999.

[118] A. Bavelas, "A Mathematical Model for Group Structure," *Human Organizations,* vol. 7, pp. 16-30, 1948.

[119] J. R. Seeley, "The Net of Reciprocal Influence: A Problem in Treating Sociometric Data," *Canadian Journal of Psychology,* pp. 234-240, 1949.

[120] E. D. Kolaczyk and G. Csardi, *Statistical Analysis of Network Data with R*. New York: Springer, 2014.

[121] L. C. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry,* vol. 40, pp. 35-41, 1977.

[122] D. Mani and J. Moody, "Moving beyond Stylized Economic Network Models: The Hybrid World of the Indian Firm Ownership Network," *American Journal of Sociology,* vol. 119, pp. 1629-1669, 2014.

[123] P. D. Hoff, *et al.*, "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association,* vol. 97, pp. 1090-1098, 2002.

[124] N. Krzysztof and T. A. B. Snijders, "Estimation and Prediction for Stochastic Blockstructures," *Journal of the American Statistical Association,* vol. 96, pp. 1077-1087, 2001.

[125] A. Žiberna, "Generalized Blockmodeling of Valued Networks," *Social Networks,* vol. 29, pp. 105-126, 2007.

[126] M. E. J. Newman, "Mixing Patterns in Networks," *Physical Review E,* vol. 67, p. 26126, 2003.

[127] M. E. J. Newman, "Assortative Mixing in Networks," *Physical Review Letters,* vol. 89, p. 208701, 2002.

[128] P. Erdős and A. Rényi, "On the Strength of Connectedness of a Random Graph," *Acta Mathematica Hungarica,* vol. 12, pp. 261-267, 1961.

[129] E. N. Gilbert, "Random Graphs," *The Annals of Mathematical Statistics,* vol. 30, pp. 1141-1144, 1959.

[130] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of Networks," *Advances in Physics,* vol. 51, pp. 1079-1187, 2002.

[131] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review,* vol. 45, pp. 167-256, 2003.

[132] A. T. Pacheco and T. Evans, "Social Dynamics Models on Complex Networks," Master's thesis, Imperial College, London, Department of Physics (January 2013), 2013.

[133] J. Travers and S. Milgram, "An Experimental Study of the Small World Problem," *Sociometry,* vol. 32, pp. 425-443, 1969.

[134] J. Kleinberg, "The Small-World Phenomenon: An Algorithmic Perspective," presented at the Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, Portland, Oregon, USA, 2000.

[135] R. Monasson, "Diffusion, Localization and Dispersion Relations on 'small-world' lattices," *The European Physical Journal B - Condensed Matter and Complex Systems,* vol. 12, pp. 555-567, 1999.

[136] H. Jeong, *et al.*, "Measuring Preferential Attachment in Evolving Networks," *Europhysics Letters,* vol. 61, p. 567, 2003.

[137] R. Pastor-Satorras, *et al.*, "Epidemic processes in complex networks," *arXiv preprint arXiv:1408.2701,* 2014.

[138] M. E. J. Newman, "Scientific Collaboration Networks. I. Network Construction and Fundamental Results," *Physical Review E,* vol. 64, p. 16131, 2001.

[139] P. L. Krapivsky, *et al.*, "Connectivity of Growing Random Networks," *Physical Review Letters,* vol. 85, pp. 4629-4632, 2000.

[140] M. Faloutsos, *et al.*, "On Power-Law Relationships of the Internet Topology," *SIGCOMM Computer Communication Review,* vol. 29, pp. 251-262, 1999.

[141] A. Broder, *et al.*, "Graph Structure in the Web," *Computer Networks,* vol. 33, pp. 309-320, 2000.

[142] A. L. Barabási, *et al.*, "Evolution of the Social Network of Scientific Collaborations," *Physica A: Statistical Mechanics and its Applications,* vol. 311, pp. 590-614, 2002.

[143] R. Albert and A.-L. Barabási, "Topology of Evolving Networks: Local Events and Universality," *Physical Review Letters,* vol. 85, pp. 5234-5237, 2000.

[144] L. A. N. Amaral, *et al.*, "Classes of Small-World Networks," *Proceedings of the National Academy of Sciences,* vol. 97, pp. 11149-11152, 2000.

[145] M. Mitrović and B. Tadić, "Patterns of Emotional Blogging and Emergence of Communities: Agent-based Model on Bipartite Networks," *arXiv preprint,* 2012.

[146] M. E. Gaston and M. des Jardins, "Agent-Organized Networks for Dynamic Team Formation," in *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, 2005, pp. 230-237.

[147] A. Lodhi, *et al.*, "Analysis of Peering Strategy Adoption by Transit Providers in the Internet," in *INFOCOM Workshops*, 2012, p. 177.

[148] J. M. Epstein, "Agent-Based Computational Models and Generative Social Science," *Complexity,* vol. 4, pp. 41-60, 1999.

[149] A. Clauset, *et al.*, "Power-Law Distributions in Empirical Data," *SIAM review,* vol. 51, pp. 661-703, 2009.

[150] A. Clauset, *et al.*, "Finding Community Structure in Very Large Networks," *Physical Review E,* vol. 70, p. 066111, 2004.

[151] P. Pons and M. Latapy, "Computing Communities in Large Networks Using Random Walks," *Journal of Graph Algorithms and Applications,* vol. 10, pp. 191-218, 2006.

[152] S. Moss and B. Edmonds, "Sociology and Simulation: Statistical and Qualitative Cross-Validation," *American Journal of Sociology,* vol. 110, pp. 1095-1131, 2005.

[153] T. C. Schelling, "Dynamic Models of Segregation," *Journal of Mathematical Sociology,* vol. 1, pp. 143-186, 1971.

[154] D. R. Schaefer, *et al.*, "Misery Does Not Love Company: Network Selection Mechanisms and Depression Homophily," *American Sociological Review*, vol. 76, pp. 764-785, 2011.

APPENDICES

Appendix: The Partial Correlation between Various Network Properties and m and λ

Network properties	m	λ	Significance
degree (*std.*)	0.897^{***}	0.902^{***}	both
closeness (*mean*)	0.968^{***}	0.739^{***}	both
closeness (*std.*)	0.623^{***}	0.931^{***}	both
betweenness	-0.890^{***}	-0.517^{***}	both
betweenness	-0.879^{***}	0.096^{*}	both
transitivity	0.980^{***}	0.628^{***}	both
diameter	-0.889^{***}	-0.318^{***}	both
average path length	-0.890^{***}	-0.517^{***}	both
articulation points	-0.858^{***}	-0.186^{***}	both
assortativity degree	0.832^{***}	-0.887^{***}	both
fast greedy community (no.)	-0.922^{***}	0.169^{***}	both
size of fast greedy community (*mean*)	0.990^{***}	-0.572^{***}	both
size of fast greedy community (*std.*)	0.996^{***}	0.521^{***}	both
walktrap community (no.)	0.718^{***}	-0.661^{***}	both
size of walktrap community (*mean*)	-0.516^{***}	0.504^{***}	both
size of walktrap community (*std.*)	-0.053	0.197^{***}	λ
power law distribution, x_{min}	0.692^{***}	-0.616^{***}	both
power law distribution, α	-0.065	-0.873^{***}	λ
exponential distribution, x_{min}	0.125^{***}	0.062	m
exponential distribution, α	-0.926^{***}	-0.069	m
log-normal distribution, x_{min}	0.677^{***}	-0.605^{***}	both
log-normal distribution, μ	-0.088	-0.755^{***}	λ
log-normal distribution, σ	0.253^{***}	0.900^{***}	both
Poisson distribution, x_{min}	0.369^{***}	0.236^{***}	both
Poisson distribution, μ	0.378^{***}	0.295^{***}	both
power law vs. exponential, one-sided p	-0.120^{*}	0.817^{***}	both
power law vs. log-normal, one-sided p	-0.165^{***}	0.298^{***}	both
power law vs. Poisson, one-sided p	0.315^{***}	0.805^{***}	both

VITA

Xiaotian Wang
Department of Modeling, Visualization, and Simulation Engineering
1300 Engineering and Computational Sciences Bldg.
Old Dominion University
Norfolk, VA 23529

Xiaotian comes from the People's Republic of China. In 2006, she was awarded a B.S. in Environmental Science at Shanxi University of China, Taiyuan, China. In 2009, she was awarded a M.S. in Environmental Engineering at Shanxi University of China, Taiyuan, China. Since January 2006, she has been a graduate student in Department of Modeling, Visualization, and Simulation Engineering at Old Dominion University, Norfolk. Her field of concentration has been modeling and simulation as well as social network analysis, with a particular interest in agent-based modeling.