論文 / 著書情報
Article / Book Information

題目(和文)	社会ネットワークにおける2つの動的発展現象の研究：オンラインソーシャルメディアにおける噂の拡散現象と企業間の倒産伝播現象
Title(English)	Data-driven Study on Two Dynamic Evolution Phenomena of Social Networks: Rumor Diffusion in Online Social Media and Bankruptcy Evolution Among Firms
著者(和文)	WangShihan
Author(English)	Shihan Wang
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10665号, 授与年月日:2017年9月20日, 学位の種別:課程博士, 審査員:寺野 隆雄,新田 克己,出口 弘,高安 美佐子,小野 功
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10665号, Conferred date:2017/9/20, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

DATA-DRIVEN STUDY ON TWO DYNAMIC EVOLUTION PHENOMENA OF SOCIAL NETWORKS: RUMOR DIFFUSION IN ONLINE SOCIAL MEDIA AND BANKRUPTCY EVOLUTION AMONG FIRMS

A Dissertation Presented to The Academic Faculty

By

Shihan Wang

Doctor of Engineering
in
Computational Intelligence and Systems Science

Supervisor: Prof. Takao Terano

Tokyo Institute of Technology

July 2017

ABSTRACT

The fast growth of computational technologies and unprecedented volume of data have revolutionized the way we understand our society. While the social network structure is commonly used to conceptualize and describe individuals and collectives in the highly connected world, social network analysis becomes an important means of exploring insights behind this social structure. The social networks usually keep evolving slowly over time, this evolution can become very dramatic when facing external influences, which raised new challenges for scholars in understanding the complex social phenomenon.

In the thesis, we concentrate on the dynamic evolution phenomena of social networks caused by external factors from both interpersonal and inter-organizational perspectives: rumor diffusion in online social media (i.e. interpersonal social network) and bankruptcy evolution among the firms (i.e. inter-organizational network). Driven by real big data resources, we applied various computational technologies to explore the behavioral patterns in dynamic social networks and provide implications for solving these social problems.

From the individual perspective, we explore the rumor diffusion phenomenon in online social media (i.e. Twitter in particular). With the extremely fast and wide spread of information, online trending rumors cause devastating socioeconomic damage before being effectively identified and corrected. To fix the gap in real-time situation, we propose an early detection mechanism to monitor and identify rumors in the online streaming social media as early as possible. The rumor-related patterns (combining features of users attitude and network structure in the information propagation) are first defined, as well as a pattern matching algorithm for tracking the patterns in streaming data. Then, we analyze the snapshots of data stream and alarm matched patterns automatically based on the sliding window mechanism. The experiments in two different real Twitter datasets show that our approach captures early signal patterns of trending rumors and have a good potential to be used in real-time rumor discovery.

From the organizational perspective, we understand the dynamic evolution phenomenon of inter-firm network emerging from bankruptcy. When the bankruptcy transfers as a chain among trade partners (i.e. firms), it causes serious socioeconomic concerns. Beyond previous studies in statistical analysis and propagation modeling, we focus on one underlying human-related factor, the social network among senior executives of firms, and investigate its effects on this social phenomenon. Based on empirical analysis of real Japanese firms data in ten years, an agent-based model is particularly proposed to understand the role of this human factor in two perspectives: the number of social partners and the local interaction mechanism among firms (i.e. triangle structure in inter-firm social network). Using both real and artificial datasets, the beneficial effects of a number of social partners are well examined and validated in various simulated scenarios from both micro and macro levels. Our results also indicate the influential strategies to keep firms resilience when facing the bankrupt emergency.

In this context, besides the contributions we made in each research field respectively, the study on two social phenomena enhances the understanding of dynamic independent and interactive behaviors in complex social networks, and provides a good perspective to seek solutions in other computational social problems.

ACKNOWLEDGEMENTS

First of all, I sincerely thank my supervisor, Prof. Takao Terano, for supporting me wholeheartedly during my doctoral study in the past three years. His trust and patience provides me an amazing environment to explore my research interests in data science. He inspires me all the time by his magic whimsy and guided me to start an interdisciplinary research journey in Computational Social Science. Helping me to open the door of becoming a good researcher, he is both my academic and spiritual mentor in the life.

I am also grateful to my thesis committees, Prof. Deguchi, Prof. Nitta, Prof. Takayasu and Prof. Ono, for their patience and critiques during my defense. I learned a lot from their precious advices and comments. I give my particular thanks to the Rotary Yoneyama Memorial Foundation, as well as professors and members involved in the Center for TDB Advanced Data Analysis and Modeling. Thank you so much for your supports in my doctoral study. Meanwhile, I would like to thank Prof. Dirk Helbing from ETH Zurich. I really appreciate the four months' exchange opportunity in your team. This experience motivated my research to become a better one.

Many thanks should also be given to the colleges and friends in Terano Lab, in Tokyo Tech and around the world. Thanks to Dr. Chang for answering my questions patiently in agent-based modeling. Thanks to Dr. Iza and Mohsen for the kind collaboration in two projects. Thanks to Dr. Dong for providing advices about my daily confusions. Thanks to Uchida San for helping me to adapt in our lab at the beginning. Thanks to Yamaguchi San for supporting my complex administration documents in the past three years. Thanks to my dear Dr. Jiang for being there in all the past years.

Last but not least, I would like to give my deep thanks to my parents and lovely husband Qi Ji. Your unconditional love and supports helped me to go through all the days and nights with laugh and tears. I may never be able to accomplish this without you. This thesis is dedicated to you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In this chapter, I introduce the basic definition of social networks and indicate the important role of social network structure on understanding complex social problems. With this motivation, the main contributions and overall structure of this thesis are also outlined.

1.1 Basic Definitions

The social network is a social structure made up of a set of social actors (such as individuals or organizations), sets of dyadic ties, and other social interactions between actors. With various social characters, researchers commonly study the social network to identify local and global patterns, locate influential entities, and examine network dynamics. Overall, the social network perspective provides a way to analyze the structure of whole social entities using different methods, as well as to explain the patterns observed in these structures based on a variety of theories.

There are two basic structural parts in a social network: the node (i.e. social actor) and the link (i.e. relations between actors). Figure 1.1 represents an example of the social network extracted from the friendship within a social group, where each node is a person, and links are social relations between them. In the figure, the color represents the corresponding modularity community of the nodes. Similar to modularity, other properties (such as size, density, degree distribution, average distance, clustering coefficient, connected components, etc.) are also commonly applied to measure the social network [1].

On the one side, each node in social networks can be defined based on three aspects, including its position, event and relation [2]. In the position-based case, only actors who are members of an organization or hold particular defined positions are considered as members of the network [3]. An event-based approach defines the boundaries of the network

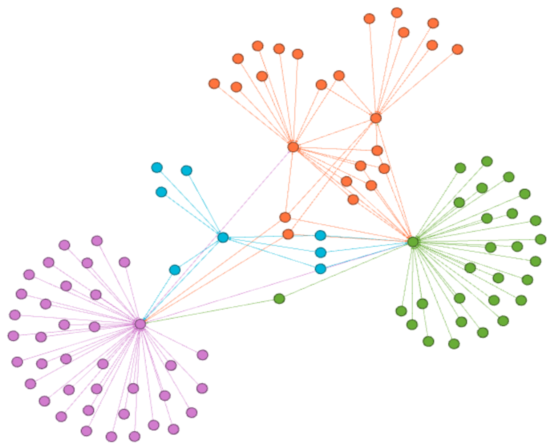


Figure 1.1: A simple example of the social network based on friendship

based on whether a social actor has participated in focal events [4]. From the final aspect, researchers extract a small population of people involved in the same interest group, then expand new actors that having particular types of relationship with existing ones into the network [5].

On the other side, the relations between nodes can be addressed from four categories: similarity, interactions, social relations and flows [2]. First, many approaches capture the similarities between actors, when they share typical kinds of variable-based attributes [6], such as demographic characteristics, locations or group memberships. Interactions are addressed as behavioral linkages among social actors [7] , such as calling, playing with or helping. Social relations refer to the interpersonal ties [8], including kinship (or other types of well-defined role relations like friendship) and affective ties (i.e. people's feeling for other members in the network). Flows are defined as exchanging or transferring relations between actors, for instance, retweet flow [9]. Meanwhile, flow-based relations may occur with other types of social relations and be used as proxies for each other. For instance, the links in a retweet social work could also include ties between friends in Twitter.

Compared to the individual-level exploration, the network-based study provides different fundamental explanations of various social questions, such as whether relations between individuals or organizations should matter, and which ones matter more. Thus, as an important carrier of various computational technologies, the social networks are commonly analyzed by researchers to understand the underlying insights of complex social problems. In this context, I particularly investigate two complex social phenomena through the use of social networks.

1.2 Summary of Contributions

In the thesis, I aim to explore two dynamic evolution phenomena of social networks: rumor diffusion in online social media (i.e. interpersonal social network) and bankruptcy evolution among the firms (i.e. inter-organizational network). Driven by real big data resources, I applied various computational technologies to explore behavioral patterns in dynamic social networks and provide implications for solving these social problems. The objectives of both research problems are stated as follows respectively:

In the dynamic phenomenon of rumor diffusion in online social media, I made contributions to identify rumor-related patterns at the early stage of trending rumor events in streaming social media data. First, rumor patterns are designed combining both structure and behavior features. Second, a pattern matching algorithm is developed to detect patterns in streaming data automatically. Third, a sliding window mechanism is particularly proposed to analyze the snapshots of data streams in the real-time.

In the dynamic phenomenon of bankruptcy evolution among firms, I studied the influential role of one inter-firm human relationship (raised by firms' senior executives) in the bankrupt evolution process of the inter-firm trade network. The empirical data of Japanese firms over ten years are analyzed firstly. Then, an agent-based model is proposed to simulate the evolutionary process of bankruptcy-related phenomenon. Finally, two factors relevant to this inter-firm human relationship are examined and measured using series of simulation scenarios.

Besides the contributions I made in each research field respectively, the study of two computational social phenomena in this thesis enhances the understanding of dynamic independent and interactive behaviors in complex social phenomena, and provides good implications to seek solutions to other computational social problems.

1.3 Outline of the Thesis

To track the research problems and objectives indicated above, the thesis is organized as follows: In **Chapter 2 Background**, I review the relevant literature on dynamic social network analysis from both technology and application perspectives and address its importance. Then, the researches of two dynamic evolution phenomena in social networks are presented in **Chapter 3 Rumor Diffusion in Online Social Media** and **Chapter 4 Bankruptcy Evolution among Firms** respectively. On the one side, an early rumor detection mechanism is proposed to capture the early signal patterns of trending rumor in streaming Twitter data. On the other side, the effects of the inter-firm human factor in bankruptcy evolution are well analyzed, modeled and examined using the agent-based simulation. The correlations of two studies are discussed in **Chapter 5 Discussion and Implication**, together with their implications for understanding other computational social problems. In the last chapter, I conclude this thesis and suggest the potential directions for future work. The overall structure is described in Figure 1.2.

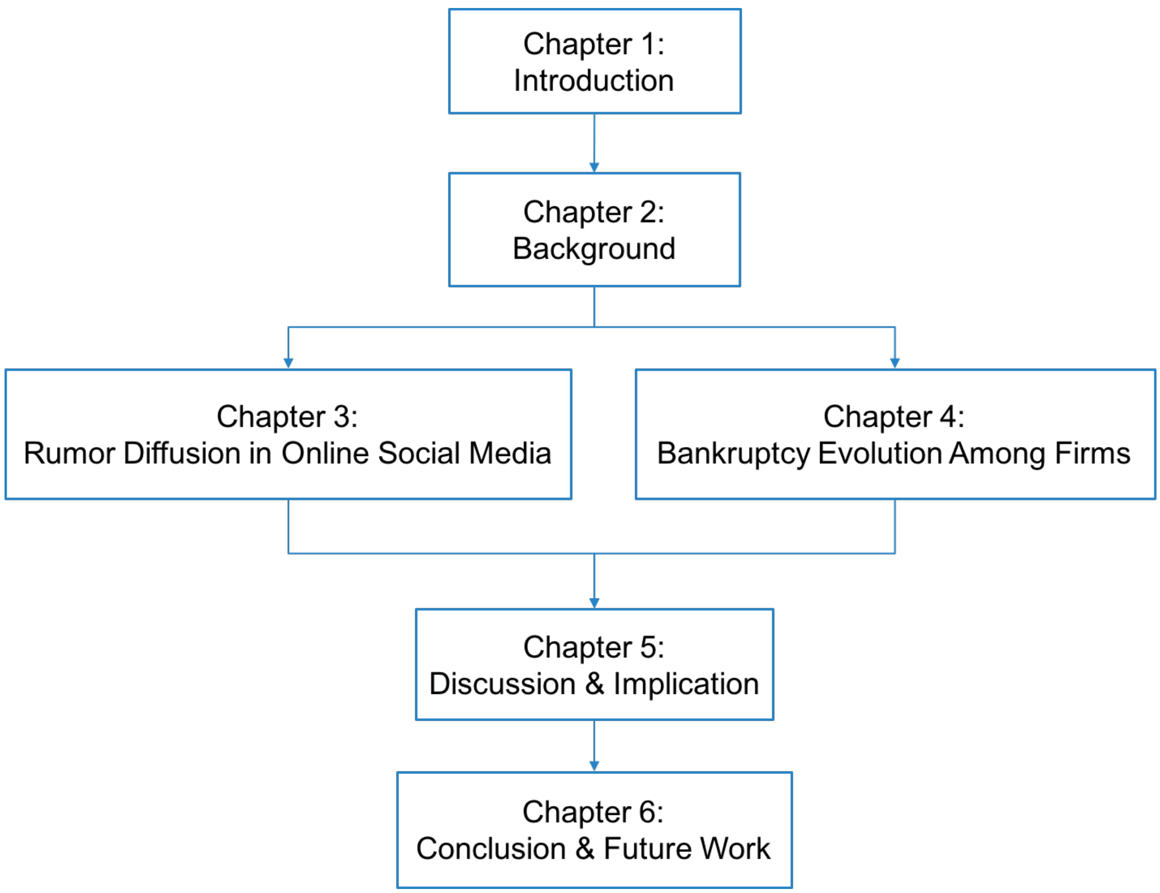


Figure 1.2: The overall structure of this thesis

CHAPTER 2

BACKGROUND

This chapter reviews the literature in the field of social network dynamic evolution. Starting from the social network evolution phenomena, I describe the related research work from both technology and application perspectives. Then, by discussing new challenges in various dynamic social phenomena, two social problems, rumor diffusion in online social media and bankruptcy diffusion among firms, and their importance are explained and argued.

2.1 Dynamic Evolution of Social Networks

In most real-life complex social scenarios, the involved actors usually join or leave from time to time, while their social relations may also be created or dissolved consequently. When applying the network structure to describe these scenarios, the corresponding social networks are evolving over time. Such phenomenon is known as dynamic evolution phenomena of social networks and computational research on this phenomenon is defined as evolutionary social networks analysis [10].

One real example is given to depict the evolving social network phenomenon over time in Figure 2.1. Since Apple released its first iPhone in 2007, the company has grown as one of biggest technology giants in the world. To study Apple's inventor innovation information, Andr Vermeij collected its internal patent collaboration data from 2007 to 2012 and reflected it in three visualized social network snapshots, where each node is an inventor and each link represents the patent collaboration [11].

The dynamic process of social networks raised much inspiration for scholars to explore interesting insights of the complex social phenomenon. Compared with the traditional social network analysis, its study requires the use of **time** in addition to the description of

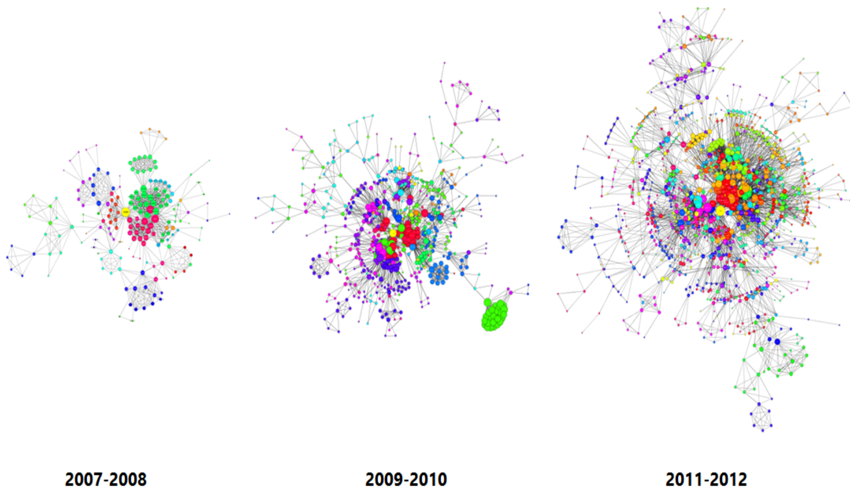


Figure 2.1: An real example: the dynamic evolution of apple's patent collaboration social network over 6 years

social network structures [12], which brings new challenges in the many perspectives. For example, key properties of social networks may change dramatically during the evolution period; standard methodologies for static social networks cannot be easily extended in the incremental (even streaming) environment; the adaptation of social behaviors and their effects in the network also be tricky and attractive for researchers, etc. Facing these challenges, researchers have made many efforts to study the dynamic evolution process of social networks and understand key concepts implied in the corresponding social phenomenon.

2.2 Literature Review

2.2.1 Overview of Research

Similar to the social network analysis (abbreviated as SNA), evolutionary social network analysis (abbreviated as ESNA) is neither a theory nor a methodology [2]. Instead, it provides a way of looking at a problem (i.e. social problem), then a way of explaining the problem, or even a potential way of solving the problem. In a nutshell, SNA and ESNA provides an environment to better understand various social phenomenons. Their studies involve various computational technologies and perform on the social networks. From this perspective, ESNA is considered as an extension of SNA, which concentrates on the dynamic (rather than static) network structure. Main research domains of the SNA and ESNA are presented in Figure 2.2.

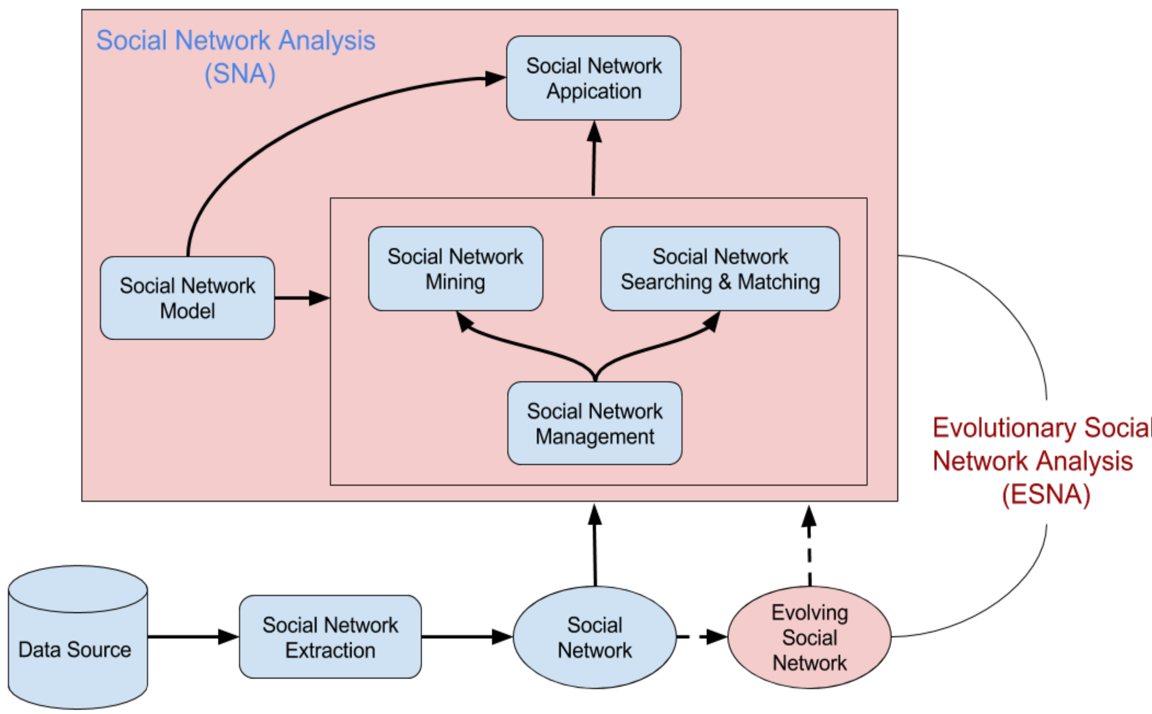


Figure 2.2: Main research domains of social network analysis (blue parts) and evolutionary social network analysis (red parts)

In the Figure 2.2, starting from raw data sources, the study in the SNA extracts a social network based on the social problem. Several research domains (including data management, data mining, pattern searching and matching) can be conducted on this social network environment. Then, various applications are studied by applying these social network techniques, while social network models are built to quantify the construction of different network structures (as well as key properties shown in). As this social network is evolving, all of those research domains can be extended to the series of evolving social networks, which are the main studies conducted in the ESNA. Also, evolution and transmission models of social networks are particularly involved in this new research field.

The literature in ESNA can mainly be distinguished into two categories, including maintenance methods and evolutionary analysis [10]:

On the one side, **maintenance methods** aim to maintain the results of analyzing process continuously and incrementally in the evolving social networks. Therefore, research works in this category generally extend the traditional methods (like classification, clustering, link prediction and pattern searching) or develop new algorithms to handle the challenging of dynamic networks.

On the other side, **evolutionary analysis** is desirable to quantify and understand the observed changes in the underlying social networks. Thus, researchers either model the structural changes to measure the evolution of networks or exam the mechanisms (or factors) engaged in the evolving process.

In this thesis, our work is mainly concentrated on the second parts (i.e. evolutionary analysis) of the ESNA. Nevertheless, the research from this perspective is reviewed from both methodology and application level in the following sections.

2.2.2 Methodology-level Review

In the methodological level, research in two classes of the ESNA has clear connections, since the maintenance methods can also be used to enhance the understanding of natures in the social network evolutionary. Therefore, instead of only distinguishing various methodologies, I organize the related work based on their purpose of using the methodology in this section. The categorizations of different scenarios for social network evolutionary analysis are summarized in Table 2.1.

Scholars in the first group made many efforts on the empirical data analysis from the statistical inference. They not only studied the change of structural properties in the series of social networks snapshots [3, 13] but also analyzed their temporal patterns occurred in the evolution process from both macroscopic [14] and microscopic [15] levels. Such data analysis can be characterized with a variety of traditional network measurements (such as density, distribution, centrality and community structure), as well as time-related measure-

Table 2.1: Key researches in the evolutionary analysis

Goal of Research	**Main Methodology**
1. Analyze empirical data to quantify the evolution over time	Statistical methods
2. Study how entire social networks form & evolve over time	Mathematical models
3. Study how finer aspects in social networks form & evolve over time	Mathematical models and machine learning methods
4. Understand social processes & issues performed in the dynamic evolving social networks	Machine learning and simulation methods

ments like arriving speed, frequency, temporal ratio and so on.

The literature in another aspect of evolutionary analysis modeled the laws of evolution in various social networks. In general, the evolving behavior of social networks is addressed as a *stochastic* dynamic process, which indicates how local mechanisms (i.e. addition or deletion of the nodes and links) shapes the global evolutionary of social networks. As the traditional Erdős-Rényi model [16] is not able to well present many properties of social networks, the preferential attachment models are proposed to observe the power-law degree distribution in real social networks [17] [15]. Later, the increasing density and decreasing diameter are taken into account in the Forest Fire model [18]. Also, Jackson and Watts considered the incentives of individuals for the definition of improving paths [19], while Tom A.B. Snijders et al. proposed to use stochastic actor-based models to flexibly represent the dynamic process of social networks based on observed longitudinal data [20]. While several typical mathematical models are discussed here, more related work can be referred in the surveys [21, 22, 23].

In addition, a parallel body of work concentrated on the measurements of finer aspects in evolving social networks. Statistical (i.e. mathematical) models are widely developed to capture various finer structural characteristics in the social networks, and to uncover their formation mechanisms. Social network formation models regarding triangle structure [24], star and component structure [14] are presented and discussed in the literature. Moreover, the formation and evolution analysis based on community structure particularly attracted academical attentions [25, 26], since the significant community structure is usually corresponding to a social group. Tracking emergence of new communities and death of old ones in dynamic social networks, researchers is desirable to understand the social inferences under identified structural changes.

The final technical scope studied the network-dependent social processes and issues. As a rather new research domain, it is becoming increasingly important and popular in recent years. Compared with the above three classes, this research category looks at the social network evolutionary more from the social perspective, instead of a physical aspect. It mainly concerns about how the outcomes observed in dynamic social networks can *shed light on* various practical and substantive topics in our society [27]. Thus, those research subjects are usually diverse and contain key concepts in the practical social phenomenon. For instance, Ghiassi et al. studied customer demands by analyzing sentiments of users in an online network towards various brands [28], and Sakaki et al. considered network-based users as social sensors to detect earthquakes in Japan on real-time [29]. In fact, both of our works presented in the thesis belong to this scope. Various machine learning and social simulation technologies are widely employed and developed to achieve the research goal, which allows us to not only investigate the social behaviors embedded in dynamic social network environment but also explain the network-based effects on the social phenomenon [5].

2.2.3 Application-level Review

In this section, I introduce the related work on evolutionary analysis from an application aspect, discussing the different kinds of evolving social networks.

Various social networks require different kinds of evolutionary analysis as they evolve at a totally different rate. According to Aggarwal and Subbian's review [10], two classes of evolving social networks are defined based on their evolution rate. First, **slowly evolving social networks** evolves naturally slow over time, where the edges are updated on the time scale of weeks or months. For example, in bibliographic networks, it may take coauthors months (or even years) to write an article. Thus, the offline analysis on a series of snapshots can be used effectively in this case. On the contrast, **fast evolving social networks** update links (or nodes) in faster time series by transient interactions (like people can make mobile calls every minute or second). In some scenarios, the evolution of social networks can even happen at the streaming rate, which typically requires real-time analysis methods. An overview about various kinds of evolving social networks in two classes are presented in Table 2.2. Among numerous applications in this field, I only indicate some typical kinds of social networks, and detailed applications can be found from the discussions in [30, 31, 32].

Table 2.2: An overview of applications (i.e. various kinds of evolving social networks) in the evolutionary analysis

Class of Evolving Social Networks	**Research Work**
Slowly evolving social networks	Scientific bibliographic network [5], actor network [33], business network [34, 35], blog network [13], etc.
Fast evolving social networks	Email network [18], mobile call network [36] social media network [29, 37], etc.

2.3 Two Dynamic Evolution Phenomena of Social Networks

2.3.1 Background

While not all social networks evolve equally fast, their individual evolution is not always at the same rate as well. In fact, most dramatic changes in both slowly and fast evolving social networks occurred in a short period of time, often caused by the external influences [10]. For example, the big business meeting usually benefits the creation of a great number of new commercial linkages, and lots of interesting communities in Facebook are formed by popular virtual events. Such key evolutionary scenarios of social networks are commonly defined as the **events** and lead to a number of significant social tasks. Identifying the events and investigating effects of them embedded in the evolution of social networks provides an important sketch about the underlying natures of those dynamic social processes.

2.3.2 Importance and Motivation

Driven by this background, I studied two dynamic evolution phenomena of social networks raised by external factors: rumor diffusion in online social media and bankruptcy evolution among firms.

First of all, the literature has well addressed that social networks in online social media plays influential role in our society [38, 39]. Since online social media drastically engaged in our daily life, people has changed their social behaviors and tend to communicate anywhere and anytime via online platforms (in particular microblogs like Twitter). Considering the importance of information sharing among users [40], the dynamic evolution of retweet social network is studied. When rumor (i.e. controversial and confusing information) appears in this online social network as an external factor, the network may grow at a fast speed based on the information spreading among tremendous audiences. In this work, I desire to discover such dynamic social phenomenon from the beginning and decrease the devastating socioeconomic damage it may cause.

At the same time, the evolution of a business network influenced by bankruptcy is analyzed in the second part. After a firm bankrupted in the inter-firm network, it affects the sustainment and dissolution of entities and links dynamically, and may even cause a chain of economic collapse. To prevent companies from this emergency, the investigation of such social phenomenon is very significant for understanding not only company's individual economic activities, but also the overall commercial environment [41]. Here, I particularly concentrate on the inter-firm social network regarding information of their senior executives (i.e. entrepreneurs), as they generally take charge of most business decisions [42] and play substantial role in the inter-organizational alliances [43].

In this thesis, I explore the dynamic evolution phenomena of two levels of social networks: interpersonal (social media user network) and inter-organizational (inter-firm network). Meanwhile, the social media user network evolves in real-time,while the evolution of inter-firm network perform slowly over weeks or months. The combination of two different studies allows us to comprehensively investigate the dyanmic social behaviors embedded in social networks and explain the network-based effects on the complex social phenomenon.

2.4 Concluding Remarks

In this chapter, the background of dynamic evolution in social networks is reviewed, including explanations of its phenomenon, challenges, research domains and a variety of related work. We also propose evolutionary analysis research of two social networks, which will be described in the following Chapter 3 and 4 respectively.

CHAPTER 3

RUMOR DIFFUSION IN ONLINE SOCIAL MEDIA

This chapter concentrates on the study of rumor diffusion phenomenon in online social media. With the extremely fast and wide spread of information, online rumor causes devastating socioeconomic damage before being effectively identified and corrected. To fix the gap in real-time situation, I propose a method for monitoring and detecting rumors in the online streaming social media (Twitter in particular) as early as possible. First, the rumor-related pattern combining features of users attitude and structure in the propagation social network is defined. A pattern matching algorithm tracking these patterns in streaming data is also proposed. Based on the sliding window mechanism, I overcome the streaming challenge by only analyzing the snapshots of data stream and detecting matched patterns automatically. The experiments in two different real datasets show that my approach captures early signal patterns of rumor (that trending in online social media), and have a good potential to be used in real-time rumor discovery. In a nutshell, this chapter presents the work from following parts: research statement, research review, pattern design, methodology, experiments on real data and concluding remarks of research contributions.

3.1 Research Statement

In the past decade, the proliferation of online social media has drastically changed our way of social communication. Thanks to its fast and wide reachability, an ever-increasing audience turn to use Microblog platforms, such as Twitter, for real-time information sharing. However, this powerful dissemination tool provides a prolific environment for the fast and easy spread of not only valid news but also a variety of rumors.

In psychology, rumors are known as pieces of controversial (i.e. cannot be verified as true or false) information or statements that spreading from person to person [44]. Among thousands of rumors in online social media, I particularly concentrate on those are emerging as popular topics and attracting tremendous social attention. I refer to them by *trending rumor events*. In my work, trending rumor event is defined as a group of online social media posts indicating one controversial emerging topic (i.e. statement or information), which is discussed frequently and transferred widely within a specific time period.

Understanding trending rumor events are essential as they fuel mistrust, cause panic and often prompt irrational behaviors in society. Beyond online rumors with limited attention, trending rumor events brings uncertainty into wider audience and raise up stronger social chaos. The implications are obvious in emergency and disaster situations before one trending rumor event is verified as fake information and effectively corrected. For example, medical researchers are raising concerns that public health misinformation could impede efforts to limit the spread of a virus. However, the implications go beyond this. Social media platforms and Twitter, in particular, are changing the way journalism is conducted nowadays, by providing real time information, reactions, and public opinions during breaking stories. This 24 hours news cycle provides an important news source for journalists, then reports information back to the public.

While many scholars made efforts to automatically determine the credibility of rumors, there is still a significant research gap of identifying trending rumor events in the real-time online social media. Therefore, I focus on identifying the candidates of trending rumor events (who have high likelihood to be false) automatically and efficiently from online social media. There are three important keywords devoted in my contributions: *early*, *signal* and *streaming*: 1) I track trending rumor events in online social media at their very *early* stage (i.e. as soon as possible or near real-time); 2) I capture the *signal* of trending rumor event candidates to reach a good recall accuracy; 3) I overcome challenges (like the limitation of data access and calculation) raised by the *streaming* data in real-time social media scenarios.

Following those important considerations, in this context, I analyze early signal features of trending rumor events, then propose an approach to detect the candidates of them in Twitter data stream based on the sliding window mechanism. First, the rumor-related patterns are defined and analyzed to distinguish the property of trending rumors in short term time-series. Second, a pattern matching algorithm is developed to automatically track significant patterns in streaming data. Finally, I analyze snapshots of the Twitter data stream and capture early significant patterns matched in the series of sliding windows. According to empirical analysis of patterns in two different Twitter datasets, I address stable results for the following questions: how early the signal patterns can be discovered in trending rumor events; and once detected, how reliable are they for distinguishing between false and true trending rumor events. Applying these early signal patterns as potential indicators, I expect to detect candidates of trending rumor events before they cause too much social and economic damages while spreading in online social media.

3.2 Research Review

Following the common usage of online social media in our daily life, a wide range of research domains are raised and investigated in recent years. In one scope, scholars addressed a variety of negative entities involved in online social media (like memes, misinformation, and rumors) and analyzed their significant properties [45, 46, 47]. Meanwhile, a great many works are done to study the social phenomena raised by the real-time spreading entities from different aspects. For instance, researchers defined strategies to collect and annotate data for rumor classification [48] developed systems for tracking misinformation or memes in real-time [49, 50, 51] and identified the certification of emerging news in social media streams [52].

Among various entities spreading in online social media, those relevant to the rumor are particularly concentrated on in this thesis. While rumors have been reported as the big social issue by psychologists for a long time [53], computer scientists started to concentrate on automatically detecting rumor-related entities in online social media from recent years.

In this year, Zubiaga et al. [54] initially summarizes the state-of-the-art approaches in rumor analysis and detection in online social media. According to scopes addressed there, my contributions are involved in the research field of detecting rumor in real-time online microblogs (in particular Twitter and Sina Weibo). Thus, the studies of analyzing rumor-related features (i.e. patterns) are firstly introduced in this section, followed by a discussion of current literature and remained challenges about the real-time rumor detection.

Rumor-related Patterns in Online Social Media Castillo et al. [55] initialized this research aspect by examining relevant features in four categories (text, author, propagation and topic properties), and particularly emphasized the importance of propagation features. Kwon et al. [56] extend their work by first understanding the temporal properties in rumor spreads, while Wu et al. [57] further analyzed propagation patterns with temporal behaviors in false rumors of Sina Weibo. Ma et al. [58] focused on a different angle of widely used features, that is the slopes of features between consecutive time intervals. Those works well analyzed the signal features in rumors of online social media, however, all their analysis are based on the entail historical data by the current time.

In addition, several pieces of literature concentrated on time-sensitive signals of online rumors, which are essential in real-time detection scenarios. Castillo et al. [59] first captured early features by using tweets before the first activity peak in static data. Later, Liu et al. [60] indicated the influence of belief features at early stage of false rumors dynamically but computed the feature vector based on an accumulative data collected over time, instead of a data stream. Kwon et al. [61] examined the effectiveness of various features in different time period of rumor diffusion over the long and incremental time windows (per days and weeks). The work that is the most similar to ours is by Zhao et.al. [62]. They focused on inquiry patterns appearing early in the disputed factual claims (i.e. rumors) and examined its significant role in the Twitter stream of Boston bombing events. Extending their definition of rumor cluster, I define the trending rumor event by combining the "trending" characters (defined in [63]). To overcome the potential limitation of text-based regular expressions, I propose to compute early signals in streaming environment from both context and information propagation perspectives and evaluate them in short-term sliding window (per hours), which has never been considered in rumor-related social media events.

Rumor Detection in Real-time Online Social Media The work flow of real-time rumor detection starts from identifying trending events within the data of online microblogs, which is addressed as one independent research area called event detection (or emerging topic detection) [64, 65]. Related researches in this field have well developed multiple methods to detect emerging events in Twitter data stream [66, 67, 68], and return them as a set of sub-stream.

Given each trending event as a post stream, I mainly review works for identifying rumor-related entities automatically, which is the next step in the overall flow. In this step, scholars consider rumors from two kinds of perspectives. First, on a post-level, each post is analyzed to explore whether it is relevant to the false rumor [69]. Second, Castillo et al. [55] proposed the definition of newsworthy topic (a group of tweets) and assess its credibility automatically. Regardless of literature focus on either kind of rumor, I summarize them from the methodological perspective.

First of all, most researchers collect and download all of the trending events over a period of time offline. Then, they are proceeded one by one to evaluate the credibility of its claim. Here, the rumor detection task is commonly considered as a binary classification problem. And supervised learning approach is utilized to automatically determine whether one trending topic that is spreading is true or false, based on various kinds of features. At the same time, some other research went beyond classification approach to achieve a similar goal. For instance, Ennals et al. [70] used pattern matching techniques to highlight disputed claims from the web. Their method automatically searched lexical patterns for claims, then filtered claims by a classifier and provided a corpus of disputed claims only. Chen et al. [71] viewed this problem as an anomaly detection task in the first time, then performed and returned anomalies as the possible false rumors. Experimental results of previous works report an accuracy of around 90% in static/ off-line datasets. However, processing all trending events in the off-line approach requires high computational costs. Moreover, the static data analysis restricts the detection time, as most features (even temporal ones) are only available after the rumor has widely spread. Therefore, it is not very practical to use such approaches in a real-time situation, while rumors are necessary to be detected and corrected before they cause serious socioeconomic damage.

To overcome these shortcomings, some scholars adopt a different approach in the recent works: they first filter trending events online, then analyze only those with higher probabilities of being false rumors in the offline environment. Zhao et al. [62] identified such controversial trending events based on early inquiry content patterns. Their work identified an earlier detection characteristic and obtained a precision of 50% when applied to a cleaned dataset capturing the Boston bombing events. However, the recall is ignored in their work. While recall accuracy is more significant than precision in rumor detection task, my methodology proposed in this thesis is designed to narrow down trending events with high probability of being rumor in streaming data.

Some other related works are also proposed to discover rumor-related events in real-time social media data. Qin et al. [72] developed an early detection system by consulting data sources additional than the social media messages, while Yang et al. [73] combined rumor identification within the process of Twitter 'bursty' detection. Moreover, the crowd sourcing techniques [74, 75] are applied to annotate and identify trending events with high likelihood of being false rumors, by involving the feedbacks from crowds.

3.3 Rumor-based Patterns

In this section, I introduce the definition of my rumor-related patterns, as well as the design principles.

To apply rumor-based patterns for identifying candidates of trending rumor events in streaming data, there are two main properties involved in the design: *complexity*, patterns should be easily and efficiently acquired in data streams; *accuracy*, patterns capture the important temporal features that can distinguish false and true trending rumor events. As studied in the previous empirical data, the more complex rumor patterns usually turn out a better accuracy for false trending rumor detection [55]. However, the streaming environment restricts the complexity of my patterns. Because of the one pass constraint in data streams, it is very difficult to do iterative or time-consuming calculations [76]. Therefore, I balance the above two aspects (complexity and accuracy) and define a set of rumor-related patterns.

Overall, designed patterns combine two most influential and efficient features: information propagation structure and users' attitude towards the focal trending rumor event. In a nutshell, my patterns are labeled and directed graphs, where directed network cascades are extracted to represent the information diffusion process, while various users' attitudes toward the topic are conceptualized as labels of nodes. In Figure 3.1, two examples are particularly given to clarify the patterns.

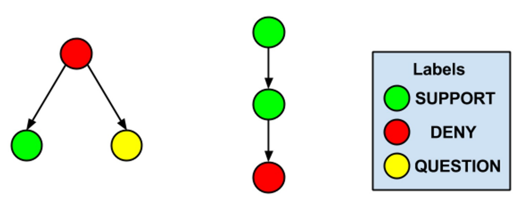


Figure 3.1: Two examples of the rumor-based patterns

According to Figure 3.1, two essential network cascades (star and path graph) are employed as the structural base in this context,while three different labels (SUPPORT, DENY, and QUESTION) are designed to capture a variety of user attitudes. In the following part, I detailedly explain the importance of both properties and the theoretical base of how they involve in the patterns.

The Structural Feature. The network structure of information propagation has been regarded as one of the most influential features for determining the information credibility

[55, 57]. As it can observe the whole process that a topic is transferring among people, its structural patterns generally imply the significant ”trending” properties.

Thus, I follow the previous studies and review the important topological features shown in the propagation of trending topics. Instead of macro-level topologies, I concentrate on micro cascade motifs that present representative characteristics in the network structure. Zhou et al. [77] and Fan et al. [78] analyzed the trace of information propagation in trending topics of two important Microblogs. And the top seven frequent non-trivial shapes obtained from Twitter and Sina Weibo data respectively are shown in Figure 3.2.

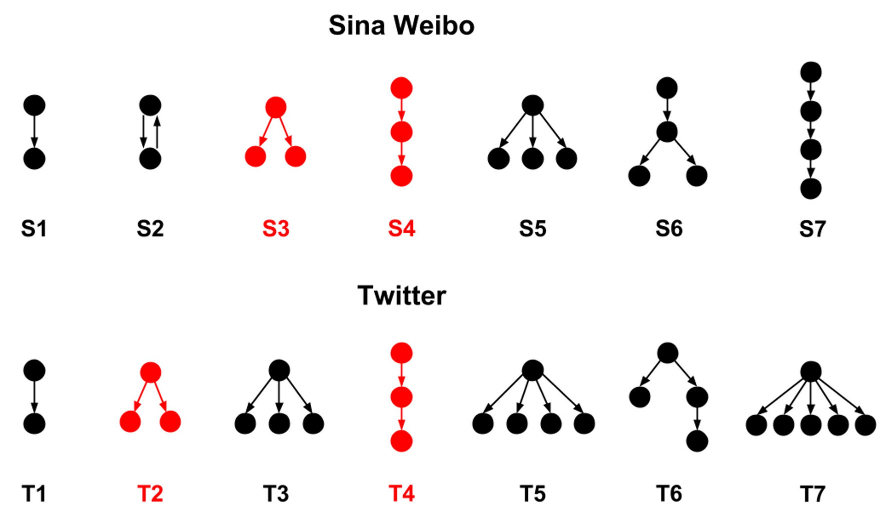


Figure 3.2: Frequent-ordered nontrivial cascades in the propagation of trending topics in Twitter [77] and Sina Weibo [78]

In Figure 3.2, besides the basic shapes with two nodes (S1, S2, and T1), cascade star (i.e. T2 and S3) and path (i.e. T4 and S4) are observed as the most frequent shapes among all the common structures. Also, all the other frequent cascades can also be decomposed into a set of them. These observations indicate that star and path play an essential role in the diffusion process of trending topics. Meanwhile, while social media data is coming as a stream in the real situation, the propagation network starts with the most basic structures, then keeps growing bigger and bigger. Thus, I combine those two simple propagation structures (star and path) in my patterns, as they have higher chances to be identified at the beginning stage of trending topic propagation.

The User Attitude Feature. At the same time, users attitude towards the target trending topic, as another significant feature is also combined with the rumor-based patterns. Mendoza et al. [79] firstly announced a promising report about the user attitude analysis in Twitter data after 2010 Chile earthquake. As exhibited in Table 3.1 exhibits, people tend to have obviously different opinions when spreading tweets of false and true topics.

Table 3.1: The percentage of tweets shown various user attitudes in the cases of 2010 Chile Earthquake [79]

	Confirmed true trending news	**Confirmed false trending rumors**
Support	95.9%	45%
Deny	0.4%	38%
Questioning	3.5%	17.3%

In general, more negative and doubted attitudes [62] are related to false trending rumor events, as positive users actively involved in the credible trending topics [55]. Therefore, I extract four types of user attitude (SUPPORT, DENY, QUESTION and NEUTRAL) to better distinguish two kinds of trending rumor events. Here, the NEUTRAL label is used when there is no clear attitude of the user.

In this context, considering all possible combinations of the four labels and two shapes (star and pattern), I totally construct 112 candidate patterns. For instance, two patterns shown in the Figure 3.2 are defined as: $\{SUPPORT \leftarrow DENY \rightarrow QUESTION\}$ and $\{SUPPORT \rightarrow SUPPORT \rightarrow DENY\}$ respectively. Also, since the two children nodes are symmetric with respect to the propagation structure in the ”star” shape, I consider the patterns $\{SUPPORT \leftarrow DENY \rightarrow QUESTION\}$ and $\{QUESTION \leftarrow DENY \rightarrow SUPPORT\}$ to be equivalent in the work.

3.4 The Real-time Framework and Approach

This section presents the multi-step real-time framework and my proposed approach that seeks to capture the defined patterns in streaming social media data. Overall, I propose a work flow of methodology (described in Figure 3.3), to process raw post streams of social media data into a series of detected rumor-related patterns.

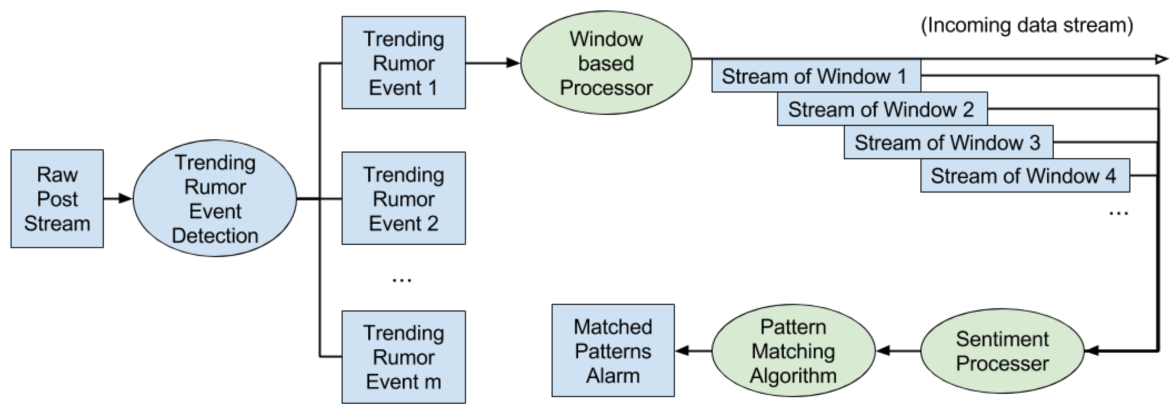


Figure 3.3: The overall framework of my methodology

According to Figure 3.3, the framework begins with extracting trending rumor events from the overall data stream. Then, with every newly received post allocated to the relevant trending rumor event, each event holds an incremental post stream. To keep all the processes in this real-time framework online, I decompose every raw post stream of trending rumor event into a sequential sub-streams, based on the sliding window principle. In this way, I further analyze the data stream in each individual window and detect rumor-related patterns using a pattern matching algorithm. During this process, sentiment analysis of post context is first applied to support necessary features for pattern matching. In the last step, after a pattern is matched, an alarm signal is reported with its current time step, carrying a potential indicator for the focal trending rumor event.

In the following subsections, I introduce three main technical parts in this work flow, which is colored as green in Figure 3.3).

3.4.1 The Sliding Window Mechanism

The sliding window mechanism is first applied as the underlying principle in my methodology. Insights provided by recent state-of-the-art literature have emphasized the usefulness of the sliding window mechanism for real-time research problems [80]. It decreases the computational complexity of mining data streams by limiting the amount of processed data. Falling into the category of real-time analysis, my approach focuses on the snapshots of most recent data, instead of the entire historical dataset. In this way, I not only overcome the challenges of streaming environment but also capture effective properties in the most recent period of time (i.e. as soon as possible).

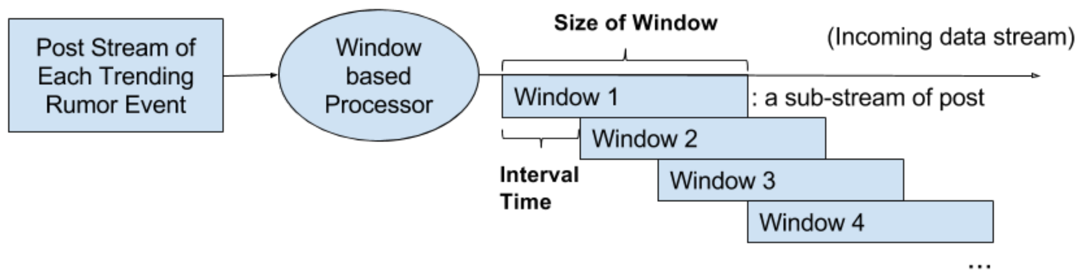


Figure 3.4: The detailed process of window processor in Figure 3.3

As shown in Figure 3.4, given a fixed window size and the interval time between windows, the window-based processor decomposes the data stream of each trending rumor event, then returns a sequence of *windows*. In my work, the overlapping sliding windows are considered to efficiently decrease the interval of detection. It allows us to balance the tradeoff between real-time data mining and sufficient data collection with interesting patterns. More details about setups of the sliding window mechanism are discussed in section 3.6.

3.4.2 The Process of Sentiment Analysis

Given the stream of posts in each window, in the next step, I extract user attitude of each post based on its context information. In total, four types of user attitude (including SUPPORT, DENY, QUESTION and NEUTRAL) are identified in my approach. The detailed processes are described in Figure 3.5.

In a nutshell, I employed sentiment analysis [81] techniques to determine user opin-

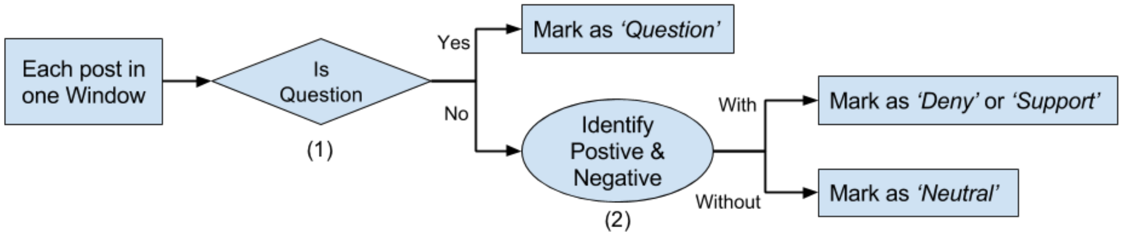


Figure 3.5: The detailed process of sentiment processor in Figure 3.3

ion from the tweet content. It is implemented based on the Natural Language Toolkit (`nltk`) [82], which is able to learn positive (SUPPORT) or negative (DENY) opinion from the text of each tweet (step 2 in Figure 3.5). Meanwhile, I identify posts with 'questioning' attitude based on several lexical expressions validated in previous researches [62, 83]. The list of used lexical patterns is summarized in Table 3.2. The top three patterns in Table 3.2 captured the inquiring sentiment about the target post. And 4th to 9th patterns match question asking cases that 5W1H question words (What, Why, Who, When, Where and How) appear at the beginning of one sentence, while the question mark ends this sentence. Besides three types of clear sentiments, the group of users who do not show any obvious attitude is labeled as 'NEUTRAL'.

Table 3.2: A list of lexical patterns used to filter strong and weak QUESTION attitude

Number	**Pattern Regular Expression**
1	$is(that \vert this \vert it)true$
2	$wh[a]*t[?!][?1]*$
3	$(real? \vert really? \vert unconfirmed)$
4	$(\backslash b[how \vert How][\wedge\backslash.!?]*[?])$
5	$(\backslash b[what \vert What][\wedge\backslash.!?]*[?])$
6	$(\backslash b[why \vert Why][\wedge\backslash.!?]*[?])$
7	$(\backslash b[who \vert Who][\wedge\backslash.!?]*[?])$
8	$(\backslash b[when \vert When][\wedge\backslash.!?]*[?])$
9	$(\backslash b[where \vert Where][\wedge\backslash.!?]*[?])$

Given the sentiment information of each post as a stream, I propose a pattern matching algorithm to further detect the rumor-related patterns in data streams.

3.4.3 The Pattern Matching Algorithm

In this subsection, I present an algorithm that can automatically discover matches of rumor-based patterns from data streams and return them with detected timestamps.

Relational Index Structure. To dynamically label the search of patterns, I firstly propose a data structure, called *Relational Index* (R-index). R-index is responsible for storing attitude (label) information related to each node. It contains label information of the current node, as well as that of all nodes link to this one. To save the storage space, total numbers of in-degree and out-degree for each kind of label are counted and collected, instead of every individual node ids. This information supports us adequate information to discover incremental patterns of each step as edges are updating in streaming. Considering four kinds of labels in my pattern graph (SUPPORT, DENY, QUESTION and NEUTRAL), an example of the basic structure of R-index is shown in Figure 3.6.

NodeId	Support_{in}	Support_{out}	Deny_{in}	Deny_{out}	Question_{in}	Question_{out}	Neutral_{in}	Neutral_{out}

Figure 3.6: The format of R-index structure

Graph-based pattern matching algorithm. Following the definition of rumor-related patterns, my approach is also developed based on a network structure. Before introducing the details of the algorithm, I firstly conceptualize raw posts holding diffusion relations (like retweet and reply) as a stream of edges, whose directions are determined by the information spreading direction. Thus, given a pattern, the proposed graph-based pattern matching algorithm tracks its shape and label information (i.e. learned by above sentiment analysis) to identify its matches. Here are some basic definitions I used in the algorithm.

Definition 1. Given a set of labeled nodes $N_T = \{n_1, n_2, n_3...\}$, each edge contains two nodes and time when it is shown, defined as $e =< n_{start}, n_{end}, time >$. **Edge Stream** is the continual sequence of edges, defined as $ES = \{e_1, e_2, e_3...\}$.

Definition 2. Since there are two kinds of pattern structure, **Pattern** is defined in the following two types: $p = \{'Star', n_{root}.label, n_{left}.label, n_{right}.label\}$ and $p = \{'Path', n_{root}.label, n_{up}.label, n_{down}.label\}$. A set of patterns is defined as $P_T = \{p_1, p_2, p_3 ...\}$. For example, two patterns in Figure 3.1 are defined as $\{'Star'$, $n_{root}.label = DENY$, $n_{left}.label = SUPPORT$, $n_{right}.label = QUESTION\}$ and $\{'Path'$, $n_{root}.label = SUPPORT$, $n_{up}.label = SUPPORT$, $n_{right}.label = DENY\}$ respectively.

Algorithm 1 matchGraphPattern(ES, P_T)

```
1:  graph G ← ∅
2:  for each e =< n_start, n_end, time >∈ ES do
3:      for all n_i ∈ {n_start, n_end} do
4:          createNodeIfNew(n_i) in G
5:          for all p_i ∈ P_T do
6:              if n_i.label matches p_i.n_root.label then
7:                  n_root ← n_i
8:                  if e is subgraph of p_i then
9:                      num ← getNumOfNewPattern(n_root, e, p_i)
10:                     updateResult(p_i, num, e.time)
11:                 end if
12:             end if
13:         end for
14:         updateIndex(n_i)
15:     end for
16: end for
```

The input to **matchGraphPattern** algorithm is an edge stream ES and a set of query patterns P_T. For every coming edge e, all of its nodes that are new for graph G are added into the graph at first (line 4). I iteratively go through every query pattern (p_i) to identify matches (line 5). Then, every node of e that shares the same label with a root node in the given pattern is selected and recorded as the root node of possible matches (line 6-7). Next, I utilize basic subgraph isomorphism to check whether this new edge is a subgraph of p_i, which is the necessary condition for further identification (line 8). As R-index maintains all previous label-related information of root node n_{root}, it is efficient to acquire the total amount of nodes that have been linked to n_{root} and matches another label of p_i (line 9). After that, the algorithm provides real-time updating matches with this new edge e in the format of $< p_i, num, e.time >$ (number of new matched query patterns and time stamp) (line 10). In the end, the R-index of both nodes is updated for future calculation(line 14).

An example is given to explain the main matching procedure. Given the star pattern in Figure 3.1, $p = \{'Star', n_{root}.label = DENY, n_{left}.label = SUPPORT, n_{right}.label = QUESTION\}$ and a new edge $e =< n_{start}, n_{end}, time >$ ($n_{start}.label = DENY, n_{end}.label = SUPPORT$), I firstly find that n_{start} is root node n_{root} and e is a subgraph of p. In the next step, I process into **getNumOfNewPattern**. As p is 'Star' type, I continue to find matches of another part in p, which is an edge with $n_{start}.label = DENY$ and $n_{end}.label = QUESTION$. Therefore, I check whether out-degree of label ***QUESTION*** ($Question_{-out}$) in R-index of n_{start} is zero. If not, it means I successfully discover new matched patterns of p that are contributed by this new coming edge. In this way, I capture the amount of new patterns and their discovered time ($e.time$).

3.5 Twitter Datasets and Preprocess

In this research, I totally utilized two real Twitter datasets to validate my methodology. On the one side, a clean multi-topic dataset is provided by a KAIST team [56], which is considered as my training data. On the other side, a raw dataset relevant to Zika virus is collected by ourselves through the Twitter Streaming API. It is not only used to test my findings but also considered as a promising case near to real-time streaming data situation. In the following section, I introduce the detailed information of both datasets, as well as how I collect and preprocess them for the further analysis.

3.5.1 The Clean Multi-topic Twitter Dataset

The first dataset collected 109 emerging topics in the Twitter from 2006 to 2009. Those trending events are related to diverse topics and contain various numbers of tweets (from 10 to 33401). And, each trending event includes a full set of tweets, as well as its confirmed credibility label (either true or false). The validation of false or true labels has been well annotated and evaluated in their work based on both investigation websites and human participants. More information about this dataset can be referred in the publication of [56].

Referring to the definition of trending rumor events, I select 10 false events containing over 500 tweets in the dataset, as they attracted a great many attentions (i.e. overall tweet amount). At the same time, 10 true trending events having similar tweet size with the false ones are selected for a better comparison between two groups. On average, each selected trending event has over 3000 tweets. Figure 3.7 illustrates the snapshots of two trending events in online media, which were selected in this dataset. In the left part of Figure 3.7, the report corrects a false claim (i.e. a false trending event I selected) that you may get swine flu by eating the pork, while the right news discusses the controversial truth about a pregnant man (i.e. a true trending event I selected).


COMMENTS

Despite the rumors, you can't get swine flu by eating pork

By Keith Magill
Executive Editor
Published: Monday, May 11, 2009 at 11:51 a.m.
Last Modified: Monday, May 11, 2009 at 11:51 a.m.

HOUMA — Your concern about swine flu might be justified, but any fears of catching the disease by eating pork is not.

"Eating properly handled and cooked pork and pork products is safe," said Beth Reames, a nutrition professor with the LSU AgCenter. "Pork should be cooked to an internal temperature of 160 degrees. And you must make sure you don't cross-contaminate raw and cooked pork as you prepare it."

The 'Pregnant Man' finalizes bitter three year divorce after agreeing to give his shark tooth collection to his ex-wife

- Thomas Beatie began to change sex before his marriage to Nancy Beatie in Hawaii in 2003
- He retained female reproductive organs and became pregnant three times when his wife couldn't have children
- The couple split in 2012 after Beatie alleged Nancy was an alcoholic who was prone to violent attacks on him
- In 2013 a judge denied a divorce request, ruling state ban on same-sex marriages meant union was not recognized as valid
- Decision was overturned last year by Arizona's Court of Appeals and the divorce has now been finalized
- But his ex-wife made one last demand - that Beatie hand over his beloved shark tooth collection

a) One used false trending rumor event

b) One used true trending rumor event

Figure 3.7: The real online snapshots of two selected trending events in the experiment

In the next step, by ranking tweets in each topic by their time stamp, I process every group of tweets into a stream of posts, which is consistent with the data in real-time. In order to briefly understand the temporal property of trending events in short-term, I count the tweet frequency per hour of each event and present 10 examples of them in Figure 3.8. In each image of Figure 3.8, one unit of x-axis is one hour, and the y-axis represents how many tweets are posted within this time period. I observed that true trending events generally show dramatic fluctuations, while false ones commonly have one sharp peak. It indicates that even in the short-term time series, different kinds of trending events can hold different properties, which makes it possible to distinguish them in the streams of sliding windows.

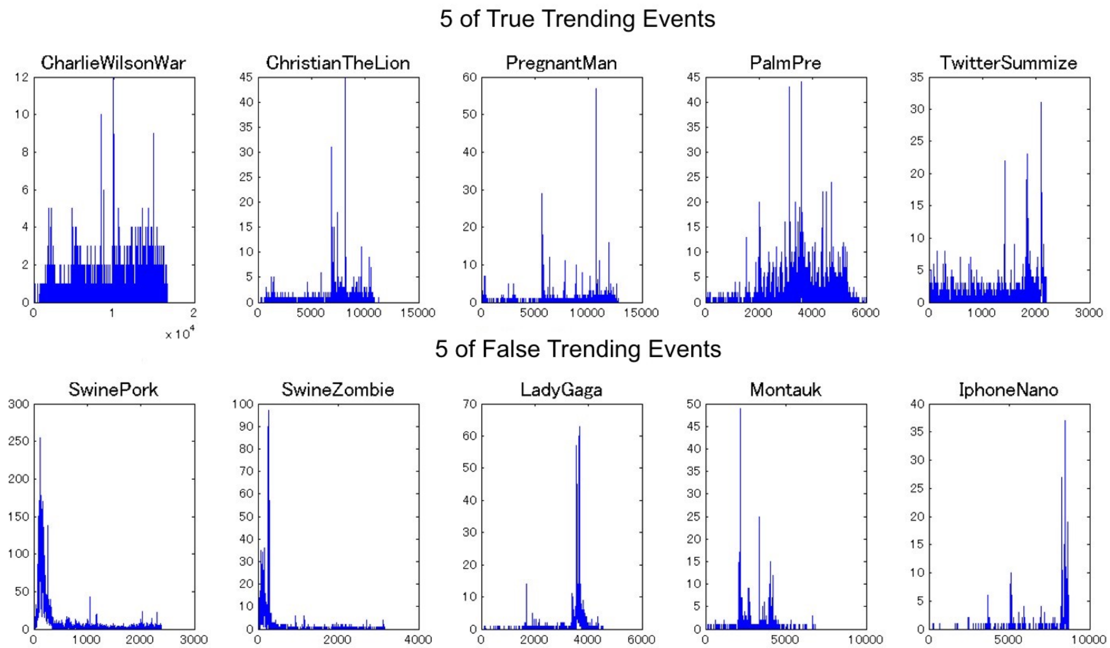


Figure 3.8: Tweet frequency of false and true trending events in short-term series

3.5.2 The Zika-related Twitter Dataset

I also acquire and process a real-time streamed Twitter dataset, which captures the conversation on an important and recent topic: the Zika virus. This dataset is referred as the *Zika* dataset, throughout the rest of thesis. In this subsection, I mainly provide an overview of the data acquisition and preprocessing steps for the Zika dataset.

Data Acquisition. The Zika dataset I consider in this paper was collected through the Twitter Streaming API[1], which grants public access to the 1% of public tweets. I define

[1]Source: https://dev.twitter.com/rest/public
and https://dev.twitter.com/streaming/overview

a custom filter using keywords from the official Twitter channels of the Center for Disease Control and Prevention (CDC) and the World Health Organization (WHO), keywords related to the zika virus: `zika,mosquito,aedes,microcephaly.`

In addition to the text of the tweet, the streaming mechanism also provides relevant information about the tweet, such as posting time and date, geo-tags, author information, retweet counts, and network indicators (number of followers and friends). These metadata fields are returned by the API in JavaScript Object Notation (JSON) format. The implementation of the streaming process was done based on Storm[2]. Table 3.3 provides a descriptive summary of the dataset.

Table 3.3: Information of the Zika Twitter dataset

Number of files	17716
Total size	256.1 GB
Number of tweets	59 930 468
Start date	13th April 2016
End date	31st August 2016

Data Preprocessing. In the following, I describe the multi-step pipeline that transforms raw tweets into trending rumor events that can be provided as input to complex analysis. Overall, as a proxy for identifying trending rumor events in raw data stream, I develop a method that proceeds in two steps: first, I rely on topic modeling and clustering, for identifying the Zika-related topics (i.e. data clusters of tweets) that received a lot of attention in the Twitter conversation; in a second phase, I investigated the credibility of Zika-related topics by consulting several credible websites. In this way, I collect a list of labeled trending rumor events in the Zika dataset.

Step 1: Extract Topics in Data: I first acquire the most discussed topics in the Zika dataset by topic modeling and clustering techniques.

The phase starts with extracting message field from each tweet and transforms the text into a "bag of words" representation. In the bag of words model, a text is represented as a

[2]Source: http://storm.apache.org/

multi-set of its words, disregarding grammar and even word order but keeping multiplicity. I implemented this first phase on top of Hadoop[3], by filtering out all non-latin characters and removing @mentions, hashtags, URLs, as well as stopwords (most common words used in the English language).

Then, I use a powerful technique widely used in data mining and information retrieval, topic modeling, to extract the distribution of hidden or *latent* topics from input tweet texts. The Mahout[4] implementation of Latent Dirichlet Allocation (LDA) [84] is applied in my work. The topics I extracted from the dataset, mostly reflect real-world events that created an important echo in the Twitter-sphere. Figure 3.9 illustrates an example of two topics, in the form of size-coded word-clouds: a larger font for a specific word indicates a higher probability of the word to appear in the topic definition.

(a) (b)

Figure 3.9: An illustration of topics discovered by running lda on the zika dataset.

In Figure 3.9, Topic (a) captures the event from sports world, that triggered a lot of echo in the news and social media: by mid July, shortly before the start of the Olympics, the Canadian tennis player Milos Raonic announced his official withdrawal from participation to the Rio 2016 Olympic Games. Several other athletes (tennis player Tomas Berdych and golfer Rory Mcllroy) expressed similar concerns amid fears and risks associated with the virus. Topic (b) highlights the first travel advisory issued by public health officials at the beginning of June. the Zika virus has prompted public health officials to warn pregnant

[3]Source: http://www.hadoop.org
[4]Source: http://mahout.apache.org/

women to avoid traveling to a part of the continental US, in response to a growing outbreak of the mosquito-borne disease in South Florida. 10 new cases have been reported in Miami, with strong evidence that mosquito control efforts were not working as well as officials hoped.

In the next step, I cluster the tweets into captured trending events based on their similarity. In a nutshell, each tweet is represented as a vector in the overall topic space (i.e. a distribution of topics). I perform the clustering algorithm on the overall tweet-topic matrix. The resulted clusters identify groups of tweets that are defined by similar topics. Finally, I sort the topics in descending order, based on the number of their tweets. By doing so, I am able to construct a proxy to discovering emerging topics in my dataset.

Step 2: Collect Credible Statements: I further enhance my dataset by collecting Zika-related statements corresponding to the trending topics, based on the topics discovered at the previous step.

I rely on websites of rumor tracking, popular news and official health organization channels (such as *snopes, emergent.info, Times, BBC, WHO and CDC*), for the collection process. By cross-matching the timestamps of these statements with the time range of my dataset, I compile a list of true and false Zika statements, with a prior confirmed veracity. Within the obtained latent topics which correspond to the verified statements, I further select that have the largest number of tweets and various user attitude (referring to my definition of trending rumor event). Thus, I am able to extract several groups of Zika-related tweets which are labeled as true or false. Table 3.4 summarizes a list of the 6 selected trending rumor events, with their associated labels. Among them, the second and fifth trending events are corresponding to the example (b) and (a) in Figure 3.9 respectively.

In the following experimental phase, I rely on those selected emerging topics in both Twitter datasets to further analyze trending rumor events (with a group of tweets and its false or true label).

Table 3.4: A list of selected trending rumor events in the Zika dataset

Trending Rumor Event	Veracity	Tweet Amount
Brazil's Rio Olympic will cause globe Zika spread.	False	41892
Miami outbreaks Zika, CDC issued travel warning.	True	7125
Congress rejected the Zika funding Obama proposed.	True	23412
First female-to-male Zika sexual transmission reported.	True	6331
Milos Raonic out of Olympic because of the fear for Zika.	True	3729
Zika vaccine go ahead for human trail.	True	2164

3.6 Experiments and Results

This section describes the empirical analysis on two real Twitter datasets and examines the efficiency and stability of defined rumor-related patterns and the proposed methodology.

3.6.1 Data Processing and Setup

According to section 3.4, the data stream of each trending rumor event is transformed into a sliding sequence of sub-streams. Thus, I first examine two important parameters about the sliding window, including the window size and interval time between two windows, by collecting tweet frequency in each window over a trending event. Our observation in data shows that the value of interval time between two windows is less influential than that of window time since the trend of tweet frequency in sliding windows remains relatively constant when the window length is constant. Based on several initial evaluations, I set values for the window size and interval length to 5 hours and 1 hour respectively in my work.

Next, the information of user attitude and propagation structure should be extracted from each raw post as well. When an individual post contains the diffusion relations, it is transformed into a directed edge. Here, direction of the edge follows the direction of information diffusion, while the label of two nodes in this edge is defined by user sentiment of its corresponding post.

On the one hand, the *retweet*, *mention* and *reply* information can be acquired from the metadata of each individual tweet through the official Twitter API. For example, given a tweet t_i, a set of its mentioned tweets can be acquired $T_4 = \{t_m, \|t_n, t_j, \|t_k\}$. Among them, I can identify that t_i retweets t_m and replies t_k. Then, I captured linkages within the propagating information. In this example, the *retweet* and *reply* implies that information is transferred from t_m to t_i and from t_i to t_k respectively. For the rest of mentions, the direction of transferring is from t_i to mentioned nodes (t_n and t_j).

On the other hand, I learn the sentiment information of each tweet relying on approaches developed in subsection 3.4.2. Table 3.5 summarizes the identification results for four types of attitude, in twenty trending rumor events of clean dataset. In the table, I calculate the average percentage of tweets in each kind of trending rumor events.

Table 3.5: Sentiment analysis results of 20 trending events in the clean dataset

User Attitude	**True Trending Rumor Events**	**False Trending Rumor Events**
SUPPORT	52.54%	36.55%
DENY	11.2%	17.05%
QUESTION	4.84%	13.3%

Overall, the results of user opinion are consistent with previous studies [79, 37] and reliable for further experiments. While more than half of tweets involved in true trending rumor events are positive, more users tend to deny and question the statement of a false trending rumor event.

3.6.2 Experiments and Results on Data

Following the three keywords (*early, signal and streaming*) considered in this research, I address two questions in the experimental phase: how early can I detect these defined rumor signals in streaming Twitter data; once detected, how useful are they for distinguishing false trending rumor events from true ones. In this subsection, I investigated those two questions firstly in the clean dataset, then in the Zika dataset, which provides an experience closer to the real-world online data.

1. How Early the Detected Patterns in the Clean Dataset. Starting from the clean dataset, I proceeded the data stream of every trending rumor event following the work flow in Figure 3.3. In this experiment, all possible patterns described in section 3.3 are considered. Given those patterns, the matched patterns in each window are detected and collected consequentially. According to this output, I counted the frequency of both tweet and matched patterns across the succession of sliding windows in each trending rumor events. The comparison between two values of frequency in ten false trending rumor events is presented in Figure 3.11, where each unit in x axis corresponds to one window and y axis is the count number.

Following the study of peak time in social media topics [59], I denoted a *peak window* as the window (i.e a period of time) with the highest tweet frequency for the considered event (this is also referred to as the full flourish of the trending rumor event). According to Figure 3.11, in all of the false trending rumor events, I were able to identify the defined patterns before the peak window (with the largest amount of tweets). In some cases (like event LadyGG and ObamaAT), the first patterns appeared in very early windows of the false trending rumor events, when the tweet frequency is still low (i.e. the events have not been noticed widely yet). The same trend can also be found in 10 true trending rumor events. It indicates that those rumor related patterns are capable to capture signals of trending rumor events at very early stage.

Moreover, I further analyzed time stamp of the first window where patterns were ever detected and found that it is much earlier than the peak time. In average, rumor related patterns are firstly identified around 5 months before the peak flourish of tweets in trending rumor events, and at least 38 hours earlier. This observation quantifies the early property of the defined rumor-related patterns.

2. How reliable the Detected Patterns to Distinguish Two Kinds of Trending Rumor Events in the Clean Dataset. Secondly, I further investigated the early signal patterns shown in trending rumor events of the clean dataset. Here, frequent pattern mining technique is applied to identify frequent patterns shown in individual windows, that not only appear early in trending rumor events but also can distinguish false and true trending rumor events.

In my approach, each window is considered as an entity with a list of matched patterns. By relying on the `pymining` Python library [85], a set of frequent patterns that appear in various windows earlier than the peak window are collected in particular. The illustration of frequent pattern mining process is given in Figure 3.10. In both kinds of trending rumor events, two different time periods are explored based on this method: 1) the first window in which I ever capture matched patterns; 2) the set of sliding windows within 20 hours before the peak window (referring to the time period closest to a fully flourish).

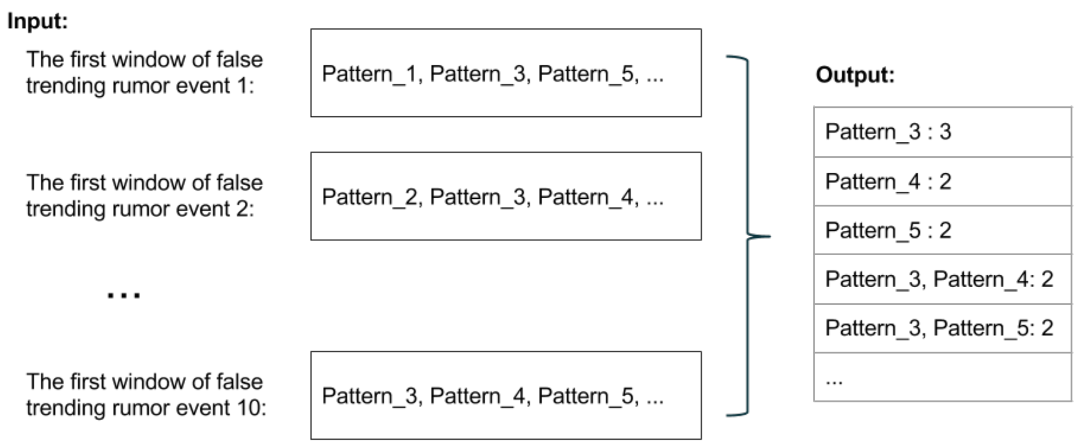


Figure 3.10: An illustration of identifying frequent matched patterns in the first window of false trending rumor events

The analysis results show that the trained trending rumor events generally hold more patterns in the earlier windows (the first window) than within 20 hours near the peak window. It indicates that sliding windows at the early stages already contain recognizable properties, which are captured by the designed patterns.

Meanwhile, there is a similar trend of frequent patterns across windows in two different periods: in the false trending rumor events, frequently detected patterns have QUESTION or DENY labels, while only a few valid events have QUESTION or DENY related patterns. Such similarity implies a potential use of those patterns to evaluate false trending rumor events in the very early sliding windows. Thus, considering either QUESTION or DENY related patterns, I calculated the number of different trending rumor events with and without QUESTION or DENY related patterns are summarized in Table 3.6.

Table 3.6: Number of events with QUESTION or DENY patterns in the clean dataset

	With	**Without**
False Trending Rumor Events	8	2
True Trending Rumor Events	4	6

As shown in 3.6, eight out of ten false trending rumor events in the clean dataset contain QUESTION or DENY related patterns, which contributes to a good recall of 80% to distinguish two kinds of trending rumor event, with reasonable precision (at 66.7%) and accuracy (at 70%). In this way, my work not only indicates the significant role of DENY labeled patterns, that has been overlooked in previous research but also contributes to a better recall accuracy for early trending rumor event detection.

3. How the Early Signal Patterns Perform in the Zika Dataset. Finally, I conducted the same experiments done in the clean dataset by using the Zika dataset. Figure 3.12 depicts tweet frequency and pattern frequency across the sliding windows in one false trending rumor event (left) and five true trending rumor events (right five). The results validated the above finding that the rumor related patterns can be detected much earlier than the peak time. In average, it is around 24 days (i.e. 378 hours) before the peak window in all trending rumor events, and particularly over 2 months in the false trending rumor event.

Another essential observation emphasizes the importance of DENY related patterns. Using the same frequent pattern mining approach, I only captured patterns relevant to DENY or QUESTION in 20% true trending rumor events (i.e. one out of five), during both time periods. In contrast, DENY related patterns are identified in both the first matched window and 20 hours before the peak of this false trending event. In fact, the most frequent patterns within 20 hours before peak are $\{DENY \leftarrow NEUTRAL \rightarrow NEUTRAL\}$ (9 occurrences) and $\{QUESTION \leftarrow QUESTION \rightarrow QUESTION\}$ (7 occurrences). This result is consistent with findings in the clean dataset and further indicates the importance of previously ignored DENY patterns, and contributes to a better recall of the early detection of trending rumor events.

3.7 Concluding Remarks

In this chapter, I address a real-time trending rumor event detection task in the phenomenon of rumor diffusion in online social media. Aiming to detect candidates of trending rumor events in online social media as early as possible, I analyze the signal patterns appearing early in Twitter data and propose a framework for monitoring and identifying these rumor related patterns in streaming data automatically. The experiment results of two Twitter datasets are also performed to validate the stability of my approach. Overall, I made three contributions in this research: First, I extend previous work and design rumor-based patterns combining properties of propagation structure and user sentiment information. Second, I proposed a pattern matching algorithm to track the matched patterns in raw post streams. Third, my framework captured signal patterns that appear at the early stage of trending rumor events' spread, by considering snapshots of data stream contained in the sliding windows. In a nutshell, the approach can identify the signal patterns relevant to false trending rumor events much earlier than the time of its fully flourish. Given these early signal patterns overlooked in previous work, my study indicates a good implication for detecting trending rumor events in the real-time online social media system.

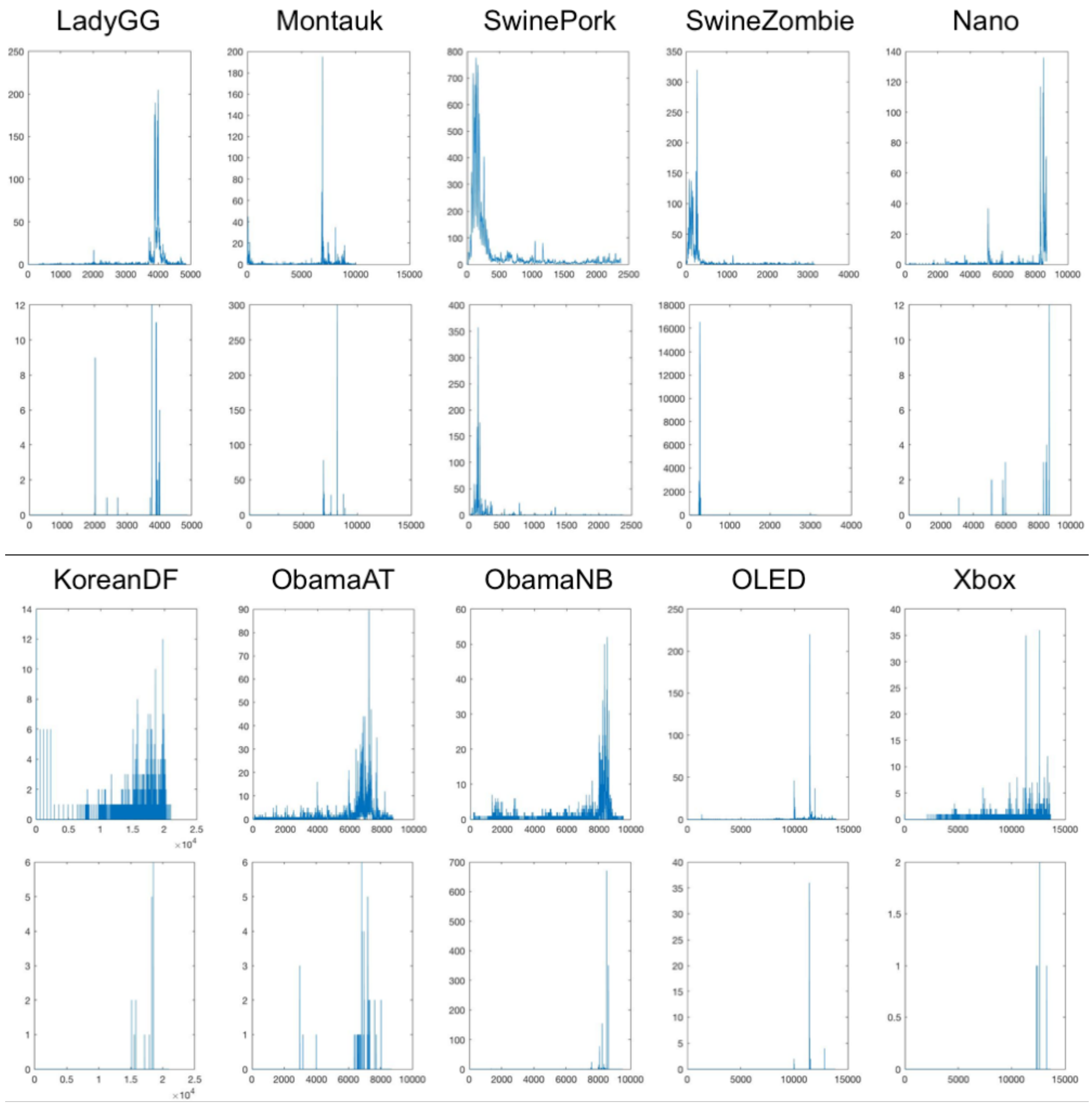


Figure 3.11: Sliding window-based tweet frequency (up side in each event) and matched pattern frequency (down side in each event) in 10 false trending rumor events of the clean dataset

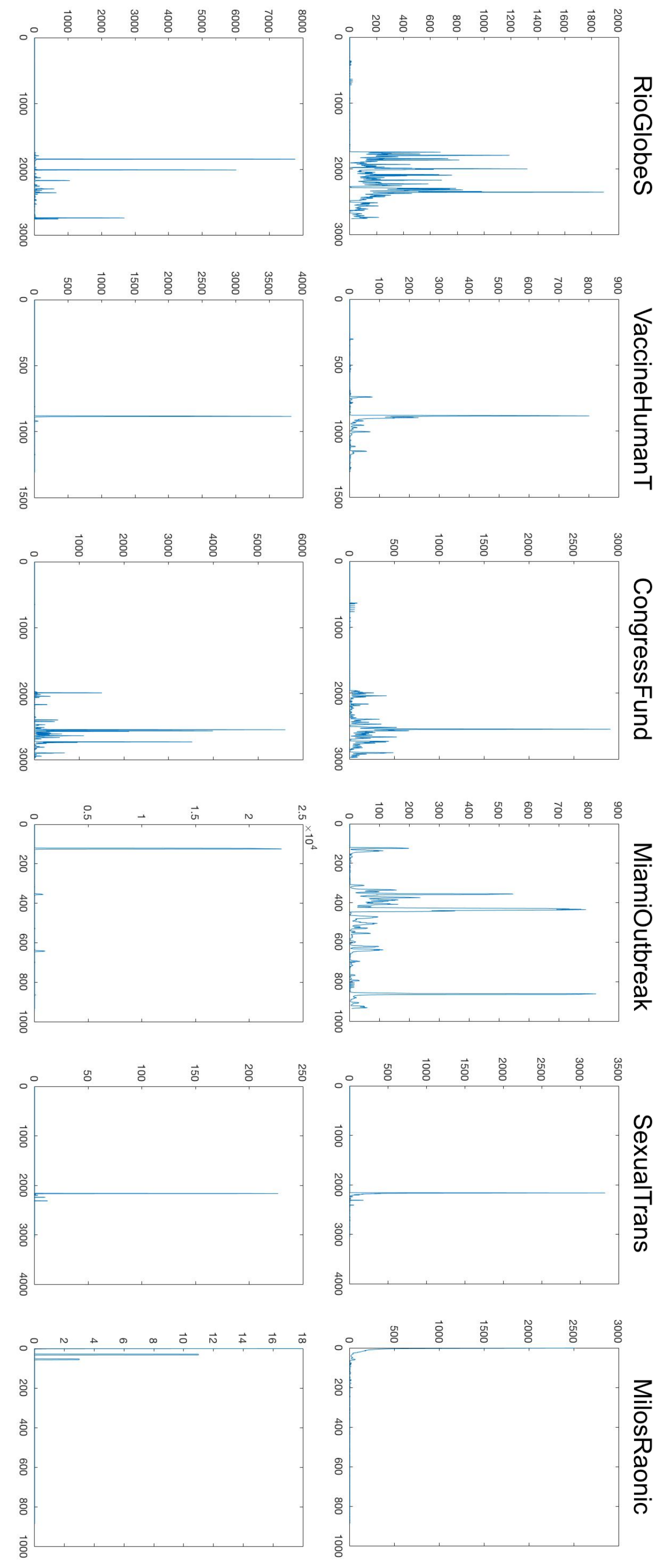


Figure 3.12: Tweet frequency (up) and matched pattern frequency (down) in sliding windows of zika trending events (event name corresponds with statements in Table 3.4)

CHAPTER 4
BANKRUPTCY EVOLUTION AMONG FIRMS

In this chapter, I explore the dynamic evolution phenomenon of inter-firm network emerging from bankruptcy. When the bankruptcy transfers as a chain among trade partners (i.e. firms), it causes serious socioeconomic concerns. Beyond previous studies in statistical analysis and propagation modeling, I focus on one underlying human-related factor, i.e. the senior executives of firms in this phenomenon, and investigate the effects of an inter-firm social network conducted by them. Based on empirical analysis of real Japanese firms data in ten years, an agent-based model is particularly proposed to understand the role of this human factor in two perspectives. Firstly, the beneficial effects of the number of social partners are well examined in various simulated scenarios from both micro and macro levels. Secondly, the local interaction mechanism among firms (i.e. triangle structure in inter-firm social network) is studied in a series of simulations using the real network of one industry. Both results indicate the important role of this inter-firm social network, which enhances my understanding of this inter-firm bankruptcy evolution phenomenon. Overall, this chapter presents the work from following parts: research question, research review, empirical data analysis, agent-based modeling, simulations, and results, as well as a summary of research contributions.

4.1 Research Question and Hypotheses

The inter-firm trade network generated from companies' trading interactions plays a significant role in understanding economic situation of both individual firms and the whole market [86]. While an inter-firm trade network dynamically evolves over time, the evolution can become dramatic within a short period when facing external influences like bankrupt [10]. Once a firm went bankrupted (i.e. inter-firm trade environment), its negative effects can grow as a snowball in the market and transfer to a wide range of firms. Such evolutionary of inter-firm trade network is also known as the bankrupt chain [87]. The evolution of bankruptcy among firms can cause serious socioeconomic concerns and even shake the foundation of our society. For instance, since several US Coal Giants (like Arch, Peabody) declared bankruptcy in 2016, a lot of firms in the US coal industry have been involved in economic trouble [88]. To avoid such collapse, the study of bankruptcy evolution among firms becomes essential and attracts the increasing attentions in recent years.

Many scholars have made efforts to explore the evolving inter-firm trade network raised by bankruptcy. They analyzed the statistical features from empirical data [89] and modeled its dynamic propagating process using simulation approaches [90]. However, some underlying factors of this evolution phenomenon have been ignored in the current literature. Among them, the role of senior executives in companies can be crucial, as most of business activities are decided and operated by them [91]. To fill this research gap, I concentrate on senior executives of firms in this work and investigate their effects on the dynamic evolution of inter-firm trade network emerging from bankruptcy.

Based on the information of senior executives, the social relations between firms are extracted and defined as an inter-firm social linkage in my work. In a nutshell, when an executive has ever belonged to more than one firms, there is an inter-firm human relation constructed between each pair of them. An illustration is shown in Figure 4.1. I refer this social link among firms as *inter-firm human relationship* in the rest of this thesis.

Overall, I measure the role of inter-firm human relationship by considering the network structure generated by it, which is referred as emphthe inter-firm human network. Then, two mechanisms relevant to this inter-firm human network are mainly explored, including the degree of inter-firm human link (i.e. the number of social partners) and a closed triangle structure within inter-firm human network (i.e. local interaction property). I expect to understand the various effects of those mechanisms in the evolutionary phenomenon of inter-firm trade network.

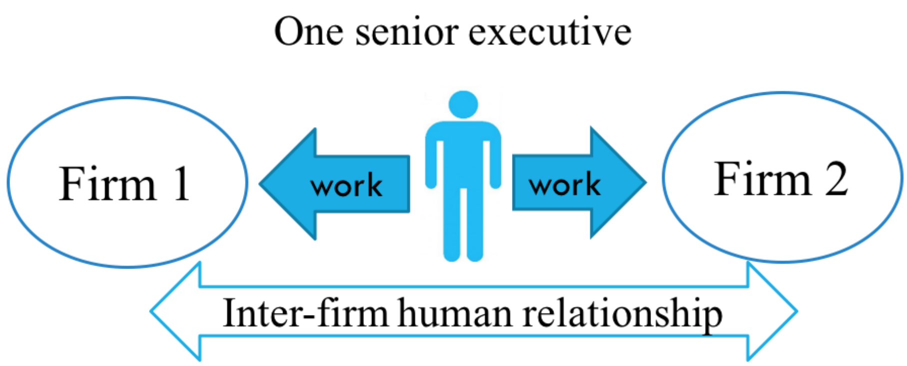


Figure 4.1: An illustration for the definition of inter-firm human relationship

Thus, I claim the following hypotheses and examine them relying on the workflow shown in Figure 4.2. According to the figure, I answer the research question by analyzing the empirical data firstly. Then, an agent-based model is particularly proposed to further investigate different human-related factors. Using both real and artificially generated data, I finally simulate and evaluate the dynamic evolution process of inter-firm trade network in a variety of scenarios.

- **Hypothesis 1:** The inter-firm human relationship influences the evolution of inter-firm trade network caused by bankruptcy.

- **Hypothesis 2:** The number of human partners in the inter-firm human network influences the evolution of inter-firm trade network caused by bankruptcy.

- **Hypothesis 3:** The triangle structure within the inter-firm human network influences the evolution of inter-firm trade network caused by bankruptcy.

Before describing each part of the workflow in Figure 4.2, I firstly review the theoretical base of my hypotheses in the next section.

4.2 Theoretical Review

In this section, I mainly review the theoretical supports for proposed hypotheses, by indicating the importance of human being (i.e. senior executives) in the inter-firm networking

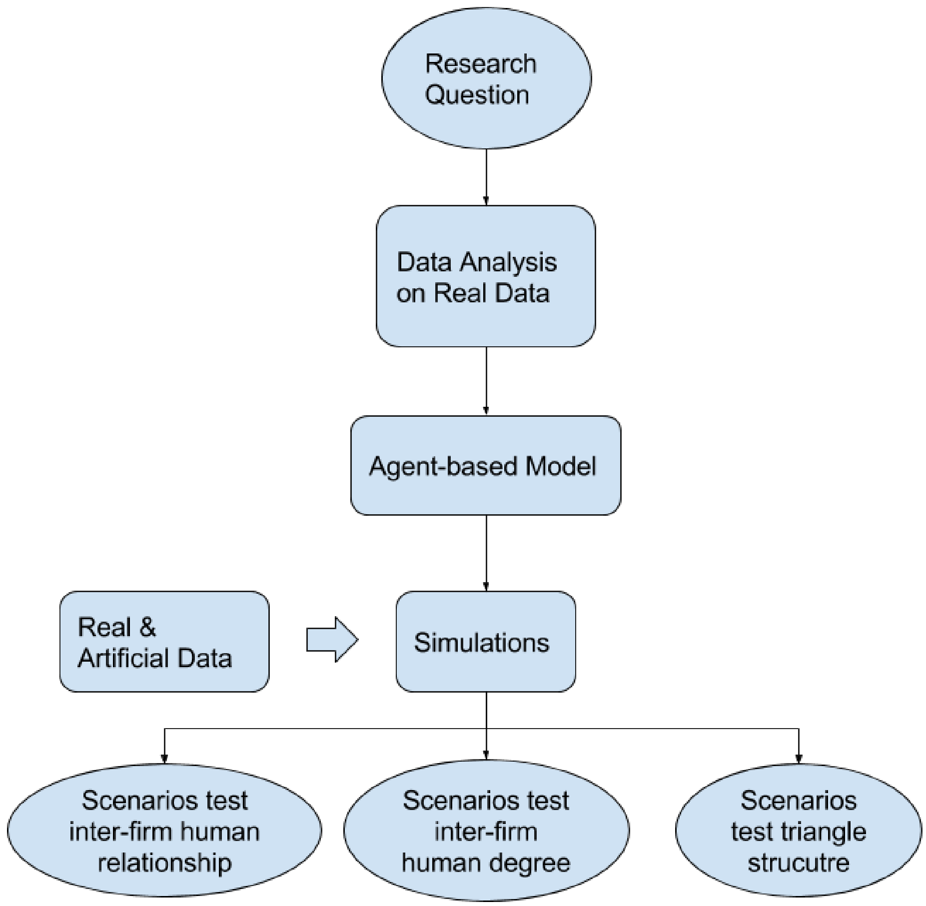


Figure 4.2: The overall flow of how I examine the hypotheses in this research

environment.

In the organization science, entrepreneurs generally play an essential role in both internal and external actions. Senior executives and founders are considered as *resource coordinators* for the firm [92], as they often bring their personal human relationship to firms as a potential source of competitive advantage [42]. The formation process of organizational alliance also usually begins with the negotiation among their senior executives [93]. Therefore, social relationship based on firms' executive is often utilized to support the business achieves [43]. Especially in the critical and emergent situations, the inter-organizational relationship tends to be driven by social level linkages among senior executives [92]. Based on these arguments, I propose the hypothesis about the importance of inter-firm human relationship when firms are facing bankrupts in a trade environment.

Meanwhile, the importance of inter-firm human relationship has been indicated in both facilitative and substantive levels [93]. In the facilitative perspective, this social relationship mainly enhances the collaboration chance [94]. However, it benefits more substantively by providing wide opportunities for the resource and information exchange among firms [92]. For example, executives communicate with each other and share their knowledge, which gives the partners access to extra information and resources [95]. Rather than marginal effect in facilitative level, I mainly consider the substantive benefits of inter-firm human relationship in my work.

In particular, I also discuss the theoretical motivation that I consider the triangle structure in the inter-firm network as a remarkable factor. Since 1908, scientists have reported the triangle structure (also known as triadic closure or closure) as a significant local interaction mechanism in multiple social phenomena. Simmel firstly pointed out that, compared with the dyad structure (i.e. two linked people), social triad (a social group with three people) increases the strength and stability between social partners [96]. While motifs with three nodes are emphasized as the cornerstone of building most networks [97], the triangle-related structures are particularly important in social networks [3], where the density of triangles is remarkably high [98]. Among several triangle cascades, I only take a simple structure (i.e. close and undirected triangle) into consideration.

In addition, triangle structure in the inter-firm network has also been widely studied in a variety of literature. In supply network, buyersuppliersupplier triads offer an interesting managerial decision implications [99]. Phelpss work indicates the beneficial role of closure structure in alliance network on the exploratory innovation of firms [100], while Kreiser suggests that network closure enhances the impact of entrepreneurial orientation on firms experimental learning [101]. In one most recent work, researchers measure the importance of triadic closure within the formation of inter-firm ties driven by shared directors using exponential random graph models [102]. Those implications have motivated us to address the impact of triangle structure in inter-firm human network on the phenomenon of bankruptcy evolution.

4.3 Japanese Firm Dataset and Empirical Data Analysis

In this section, the above research question is firstly explored from an empirical dataset about Japanese firms' information.

4.3.1 The Japanese Firm Dataset

The used Japanese firm dataset is provided by a joint project of TDB (TEIKOKU DATABANK, LTD.) and Tokyo Institute of Technology. It contains the financial information of Japanese companies over the past 10 years and has been well cleaned by engineers. In this work, I utilized data from January 2005 to December 2015, involving about 2 million firms and over 1 million senior executives belong to them. Also, Table 4.1 presents the databases applied for data analysis.

Table 4.1: Used data in the Japanese firm dataset

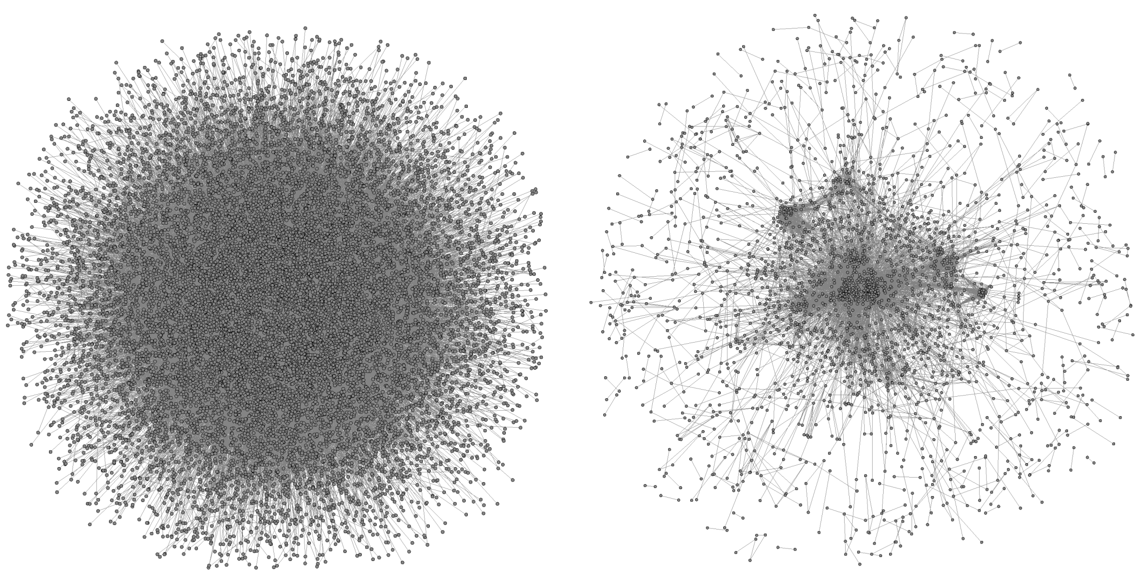
No.	Information	Schema	Number of Item
1	Firm data	$< ID_{firm}, ID_{industry}, foundTime >$	1.95 million
2	Trade recorder data	$< ID_{outFirm}, ID_{inFirm}, Time >$	1.1 million
3	Senior executive data	$< ID_{firm}, ID_{seniorExecutives}, Time >$	4.1 million
4	Bankrupted firm data	$< ID_{firm}, BankruptedTime >$	0.46 million

4.3.2 Empirical Data Analysis

Relying on real data in Table 4.1, I analyzed the importance of defined inter-firm human relationship and its relevant properties.

The Extraction and Observation of Two Inter-firm Networks. First of all, the inter-firm trade and human linkages are extracted from the real dataset respectively. On the one side, about 8 millions trade links are constructed among Japanese firms from their historical trade records (given in Table 4.1-2). On the other side, according to data in Table 4.1-3, over 20% senior executives are serving in more than one Japanese firms. Thus, I generated more than 2 million inter-firm human links from, by considering two firms sharing the same senior executive member, which is consistent with the definition of inter-firm human relationship in section 4.1. A large number of inter-firm human links (almost a quarter of the number of trade links) implies that this human relationship is a very important property among firms.

Next, I concentrated on the firms' data in one industry and constructed the inter-firm trade network with about 55,000 trade links, representing a real trade environment. Meanwhile, another inter-firm network is generated by inter-firm human links in this industry, containing 20041 edges and 2460 nodes (firms). These two extracted inter-firm networks are visualized in Figure 4.3.



(a) (b)

Figure 4.3: The topological snapshots of real inter-firm trade (a) and human (b) network extracted from one industry

A distinguished difference between the inter-firm human and trade network can be observed firstly. Although both inter-firm networks follow *a power law degree distribution*, their topological structure is quite different. While the trade-related network holds a typical scale-free property, the inter-firm human network has a *strong community-related feature*, where dense connections are held between firms in each community (with a modularity of 0.64). Moreover, firms in human-related network have a low and sparse connectivity by separating in more than 200 components, while all firms are connected in two big components by inter-firm trade linkages. Table 4.2 presents a set of detailed measurements of two networks. As shown in the table, only about 20% firms in the inter-firm trade network have inter-firm human links. To simplify the environment setup, only firms with human links are considered in the following simulation study.

Table 4.2: Detailed information of two real inter-firm networks in Figure 4.3

	Inter-firm Human Network	**Inter-firm Trade Network**
Node Number	2460	13531
Edge Number	20041	55261
Average degree	16.29	8.17
Modularity	0.64	0.26
Component Number	225	2

In addition, empirical analysis results indicate a very significant property in the inter-firm human network, which is *the triangle closure*. There are in total 320,608 triangles in this real human-related network, resulting in an average clustering coefficient at 0.467. The overview of triangle structure in the real inter-firm human network is presented in Figure 4.4, where only edges in close triangles remain and grayscale shows the local clustering coefficient value of each node.

Observation in Bankrupted Firms. On the other hand, I explored the information of bankrupted firms to understand their inter-firm human relationship. By observing inter-firm human links in the bankrupted firms from data in Table 4.1-3 and 4, I found that the number of inter-firm human links of bankrupted firms is quite diverse. Overall, it follows a power-law distribution degree. While most bankrupted companies do not have human partners in the inter-firm environment, few firms who went bankrupted involve a great many of inter-firm human links (over 60% of them have more than 10 human partners).

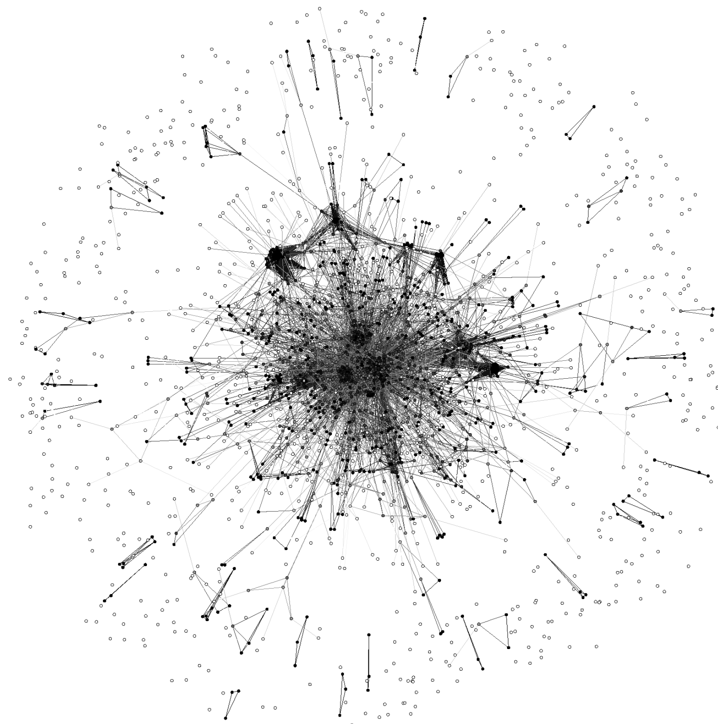


Figure 4.4: A snapshot of the real inter-firm human network about its triangle closure information (where only edges in close triangles are shown, and various levels of gray scale represent different local clustering coefficient)

This result implies that the inter-firm human links play a variety of effects to bankruptcy. It also motivated me to examine the internal impact of this human relationship in bankrupt diffusion using the agent-based simulation technique.

4.4 The Agent-based Model

This section presents an agent-based model following the ODD (Overview, Design concepts, Details) protocol [103], considering both internal adaptation and external interaction behaviors of firms.

4.4.1 The Purpose

I propose the agent-based model (abbreviate as ABM) to further investigate the effects of several mechanisms relevant to defined inter-human relationship in the evolution process of inter-firm trade network emerging from bankruptcy. The ABM technique is well used for explaining complex social phenomenon, by representing actions and interactions of autonomous agents (both individual or collective entities such as organizations or groups). It is particularly efficient to understand macro level phenomenon (bankrupt evolution) from micro level perspectives (the human-related factor). In this work, following the research hypotheses, three questions are mainly examined by applying this agent-based model: 1) whether the inter-firm human relationship affects this bankrupt evolution process; 2) whether the number of human partners in the inter-firm human network affects this bankrupt evolution process, and how; 3) whether the triangle structure within the inter-firm human network affects this bankrupt evolution process, and how. By exploring these questions, I aim to understand the internal mechanism of inter-firm bankrupt diffusion, then provide managerial guidelines about proper strategies (relevant to this human relationship) in defending bankrupt emergencies.

4.4.2 Entities, State Variables and Scales

In this model, three types of entities are presented from the real phenomenon. The inter-firm trade network is extracted from a market environment, and each firm is considered as an individual firm agent. Senior executives of the firm are conceptualized as manager entities of each firm to measure its human relationships. The conceptualization of firm and its senior executives are illustrated in Figure 4.5.

Table 4.3: Entities and state variables.

Entities	**Description**	**State variables**
Environment	Overall market environment	Firm agents ($agent_1 ... agent_n$), Human links, Trade links, Social aspiration level, Time stamp
Firm agent ($agent_i$)	Individual firm i	ID = i, Status, Performance, Internal fitness, External resource, Managers agents ($manager_1^i ... manager_m^i$), Historical aspiration level
Manager agent ($manager_j^i$)	Senior executive j of firm i	ID = j, Knowledge vector, Fitness

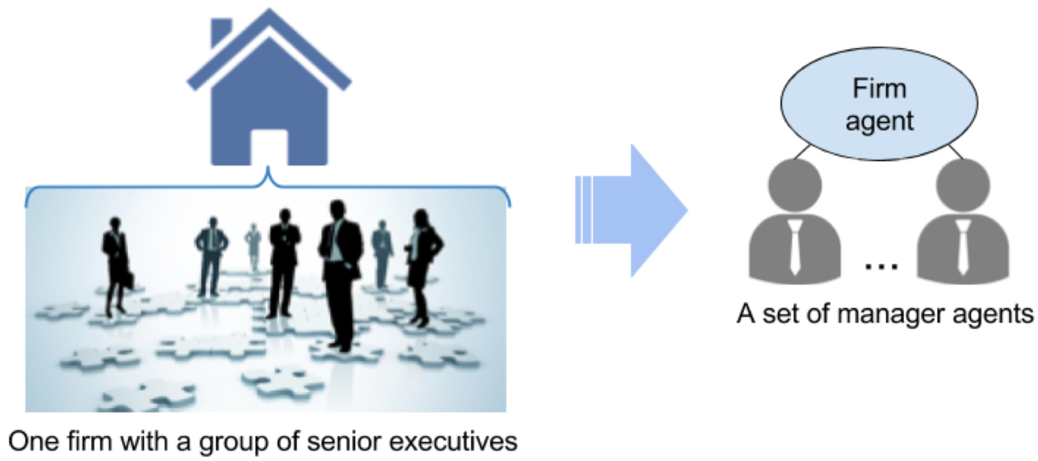


Figure 4.5: The illustration of one firm agent and its manager agents

Table 4.4: State variables in detail.

State variables	Description	Detail
Trade relationship	Supplier-customer tie between firms	R_t is directed
Human relationship	Executive-based tie among firms	R_h is undirected
Time stamp	Simulation time	$t \in [0, 20]$
Social aspiration level	Overall performance in the environment	$AS \in (0, 1)$
Status	Economy status of firm i	$S^i \in \{S_h, S_d, S_b\}$
Performance	Economy performance of firm i	$P^i \in (0, 1)$
Historical aspiration level	Performance history of firm i	$AH^i \in (0, 1)$
Internal fitness	Internal fitness of firm i	$IF^i \in (0, 1)$
External resource	External resource of firm i	$ER^i \in (0, 1)$
Knowledge vector	Knowledge vector of manager j in firm i	$V^{ji} = (0, 1, 1...)$
Fitness	Fitness of manager j in firm i	$F^{ji} \in (0, 1)$

While Table 4.3 shows a description of entities and their state variables, the detailed descriptions and scales of state variables are shown in Table 4.4. The environment entity contains both human (R_h) and trade relationship (R_t). Among them, trade links represent firm's trade records with directions, while a human link is the undirected inter-firm human relationship between firms (i.e. same as defined in 4.1). In the environment, simulation time is defined as a time stamp variable t, which varies from 0 to 20 (as one-time step represents half a year and simulations are run for 10 years), and the average performance of all firms is defined as the social aspiration level (AS) [104].

Each focal firm agent i holds a status (S^i) and performance (P^i) representing its current economy situation. Among them, the performance is modeled as a weighted summation of the internal fitness (IF^i) and external resource (ER^i), while its performance history is conceptualized as the historical aspiration level (AH^i) [104]. Meanwhile, I define three different statuses for firm agents: healthy, distressed and bankrupt [105], which has been taken in similar risk diffusion research [106]. Healthy status (S_h) implies the focal firm is successful and wants to persist its current trend. Later, it may turn to the distressed status (S_d) if the firm starts taking action to survive from near bankrupt when facing an economic emergency. In the worst case, the firm is completely failed and result in a bankrupt status (S_d). In my work, various states transfer from each other, whose detailed process is described in the following section 4.4.4.

Finally, I model knowledge evolution of manager j of agent i as a binary vector (V^j). This vector is being updated over simulation period, and consequently, that results in updates of manager's fitness (F^j).

4.4.3 The Process Overview and Scheduling

I present an overall workflow of this agent-based model in Figure 4.6, considering both agents' internal adaptation and interaction within the inter-firm environment.

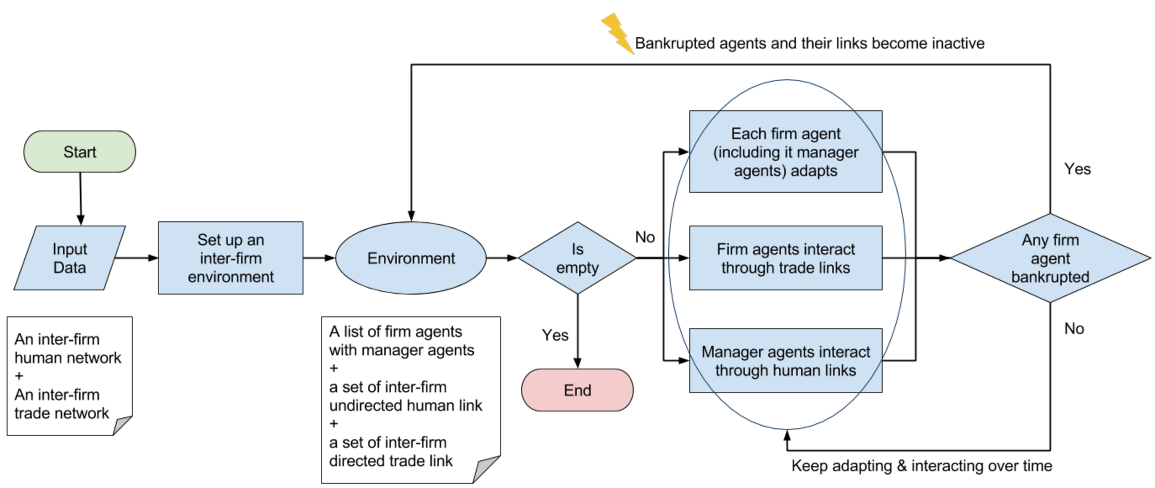


Figure 4.6: Overall work flow of the agent-based model

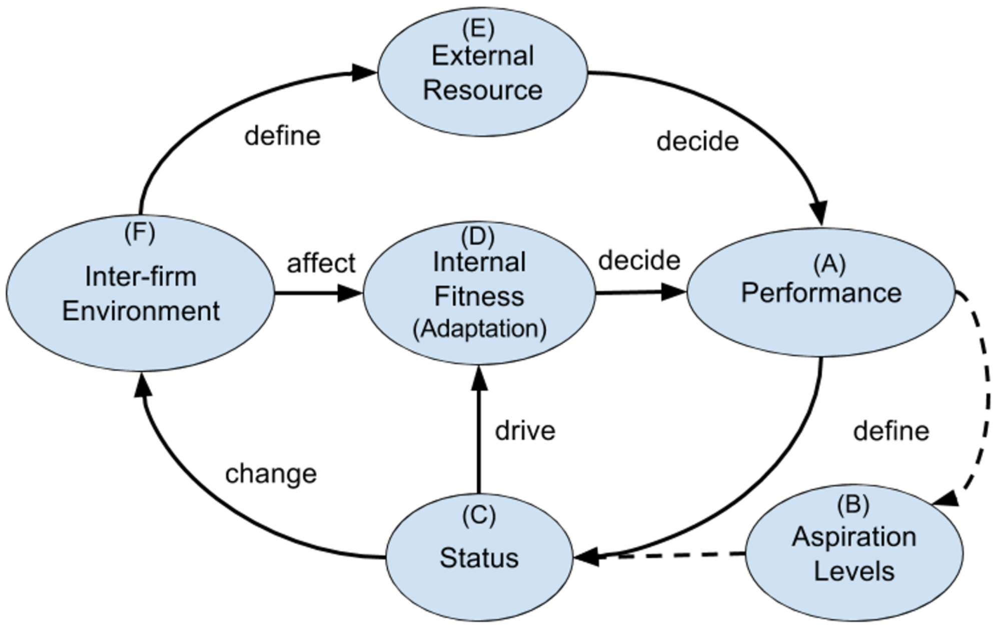
In Figure 4.6, the model firstly initialized the defined environment entity based on two given inter-firm networks, which contain the data of inter-firm links generated by trade and human relationship respectively. During the setting up process, firm agents and its corresponding manager agents are randomly generated. Over time, each active firm agent (including its manager agents) take actions adaptively in the environment. Especially, the interactive behaviors raised by trade links are done by the firm agents, while their manager agents get involved in the human-based interactions. Later, when some firm agents experiencing a severe performance failure and turn to the bankrupted status, they (together with their manager agents) become inactive and dissolve all their involved links in the environment. In this way, the environment keeps evolving until there are no more agents left.

Moreover, I further introduce the overview of simulation circle from the firm agent's perspective. According to Figure 4.7, each firm agent holds a performance (A) at the current time step t. Considering its performance with aspiration levels (B), the agent has an economy status (C). Different statuses drive different adaptive behaviors, and also change the inter-firm environment (F). The agent updates the internal fitness (D) for the next time step by internal adaptation, while the market environment forms its external resource (E). Combine the internal and external economic situation, the agent has a new performance, which drives the adaptation of simulation time $t + 1$. In this way, firm agents take actions to adjust and interact with the environment over time.

4.4.4 Design Concepts

Basic Principles. As each firm agent represents one firm, the organization behavior theories are mainly employed as the basic design concepts. Overall, there are two kinds of principles applied in my model, considering the perspectives of individual agent behavior and interaction with the inter-firm environment.

A common assumption in organization learning is that the firm tends to learn from its experience and take actions based on its performance [107]. To better understand current



(a)

Figure 4.7: The simulation circle of each firm agent in the agent-based model

situation, the aspiration level is often used as *the borderline between perceived success and failure and starting point of doubt* [108]. The model follows the well-studied assumption and integrates two aspiration levels (social and historical) [109]. The historical and social aspiration levels of focal agent i at time t are defined in equations (4.1) and (4.2) respectively [104]. In the equation (4.1), α is the weight factor of performance influence on historical aspiration (AH_t^i). Social aspiration level (AS_t) is the mean of performance of all firms in the inter-firm environment at time t.

$$AH_t^i = \alpha P_{t-1}^i + (1-\alpha) AH_{t-1}^i. \tag{4.1}$$

$$AS_t = \frac{1}{n}\sum_{i=1}^{n} P_{t-1}^i. \tag{4.2}$$

Therefore, the first principle of model behavior design is that a variety of agent statuses drive various agent behaviors, as the firm behaves differently based on whether the performance attains its goal [109]. In particular, various agent statuses are determined by agent's performance relative to the aspiration levels. The flow of status transition is shown in Figure 4.8 and three statuses of agent i at time t are defined in equation (4.3). If the performance (P_t^i) is larger than both social (AS_t) and historical (AH_t^i) aspiration levels, agent i is healthy (S_h). When the performance is between two of them, agent i is distressed (S_d). In the worst case, the firm agent goes bankrupt (S_b) and cannot recover from failure anymore.

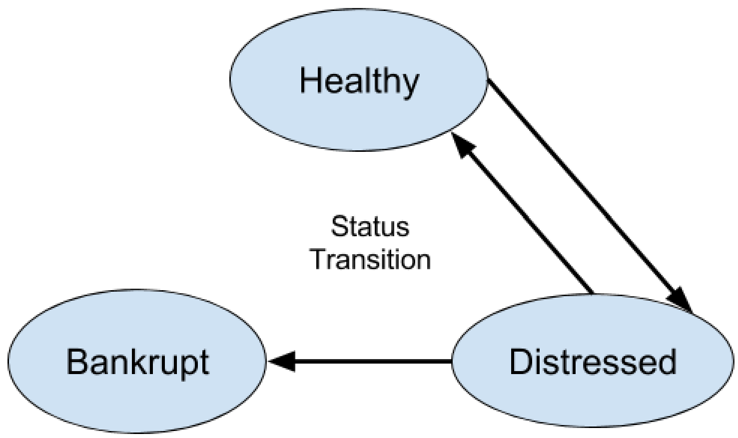


Figure 4.8: The status transition flow of each firm agent

$$S_t^i = \begin{cases} S_h, & \text{if } P_t^i > Max(AH_t^i, AS_t); \\ S_d, & \text{if } Max(AH_t^i, AS_t) > P_t^i > Min(AH_t^i, AS_t); \\ S_b, & \text{if } P_t^i < Min(AH_t^i, AS_t). \end{cases} \tag{4.3}$$

At the same time, firms continue interacting with each other in the inter-firm environment. In this regard, trade and human linkages constitute the environmental effects, and they consequently affect the linked firms in various ways, which is the second basic concept in the model design. In the organizational resource perspectives, inter-firm ties are forms of resource [110], which provide value for related firms [111]. Facing with emergency of trade partner's bankrupt, which is equivalent to alliance dissolution, the firm often takes action in response to the external resource challenge [95]. Thus, the trade link affects the external resource of relevant firms in my model. By contrast, inter-firm social ties play a more substantive role on partner firms [95], as the executives are important inter-firm knowledge transfer [112]. By carrying information across firms to overcome local search [113], the human relationship affects the internal adaptation of manager agents, then forwards the updates to the internal fitness of firm agents they belong to.

Input Data. There are two input data in this work, including the inter-firm trade network and inter-firm human network, which is constructed by the inter-firm trade and human relationship respectively. Figure 4.9 describes how both inter-firm links are conceptualized in the agent-based model.

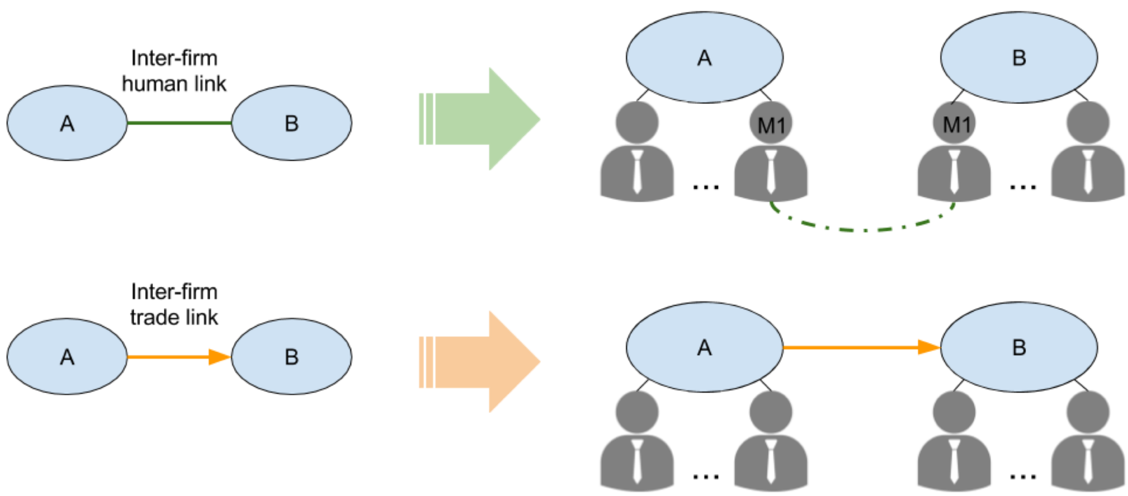


Figure 4.9: The way of inter-firm trade and human links involved in the agent-based model respectively (refer figure 4.5 for definition of entities)

As shown in Figure 4.9, the direct inter-firm trade tie links the relevant firm agents together, while the manager agents involved in the undirected inter-firm human link. Here, the linked two manager agents represent a same senior executive M1, who is working in both firm A and B, which is consistent with the defined inter-firm human relationship in section 4.1.

Objectives. One important objective in the model is performance, which measures the success of firm's economy and drives firm agent's behavior. The performance is defined in equation (4.4), where it combines both internal fitness (IF_t^i) and external resource (ER_t^i). Parameter ε is the adjustment factor that defines the weight of internal and external influences on performance.

$$P_t^i = \varepsilon IF_t^i + (1-\varepsilon)ER_t^i. \tag{4.4}$$

At the same time, as adaptive trait for each manager entity, the fitness of manager shows the overall operation situation at each time step. The internal fitness of a focal firm agent i at time t is conceptualized as the average fitness (F_t^j) of its managers in equation (4.5), where agent i has m managers in total.

$$IF_t^i = \frac{1}{m}\sum_{j=1}^{m} F_t^j \tag{4.5}$$

Adaptation. In the model, adaptation of a firm agent is executed by its manager agents. Each independent manager has a knowledge vector showing the current situation and obtains the fitness based on a function of knowledge vector in the NK model [114]. At each time stamp, the manager updates the knowledge vector to pursue a better fitness.

Figure 4.10 gives an example of the adaptation process. With the human relationship, a senior executive not only can search locally by self-experience, but also acquires information from his/her other knowledge vectors in different firms [115]. Such knowledge exchange behavior provides the manager agent j more information to act and update the fitness for the next simulation time. With many possible choices (local knowledge and that of all social partner firms), each manager agent generally has two kinds of behaviors: 1) filter the candidate situations with better fitness and select one based on the majority rule [116]; 2) choose the update at random.

In general, as the unhealthy status of firm agents restricts the energy of their managers to behave actively, the status of manager's agent determines how it updates. In particular, the

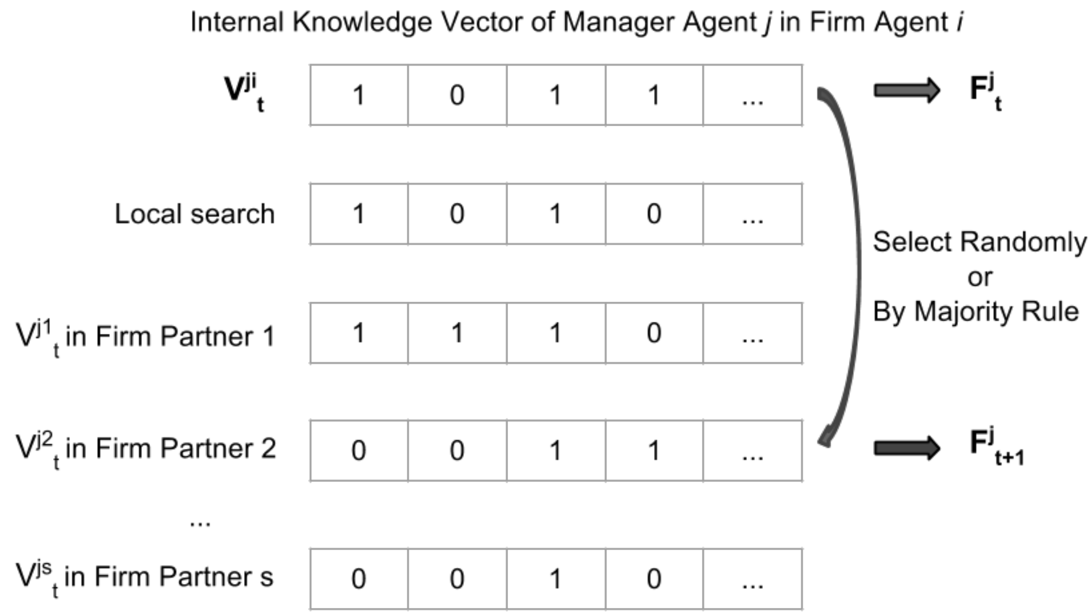


Figure 4.10: An example of internal adaptation

managers of agents in distressed select from candidate updates randomly, while the healthy managers follow a majority rule and only chooses the knowledge vector that contributes to a better fitness.

Sensing. Firm agents are sensing the inter-firm trade network environment to define their external resource. The trade links of agent i generally build an ego economic environment, which brings the agent valuable resources. While the trade linkage equally affects the trade partner on both sides, the amount of transaction brings different resources [94]. Thus, the weight of trade linkage between firm agent i and k is defined as W_k^i. Moreover, if the trade partners of agent i are in different statuses, the impact from them are also different. In particular, the effect factor of firm agent k (a trade partner of firm agent i) on agent

i is defined in equation (4.6), where S_t^k is the status of firm agent k at time t.

$$f_t^k = \begin{cases} 1, & \text{as } S_t^k = S_h; \\ 0.5, & \text{as } S_t^k = S_d; \\ 0, & \text{as } S_t^k = S_b. \end{cases} \tag{4.6}$$

Equation (4.7) presents the external resource of a focal firm agent i with total p trade ties at time t. Within the two multiplying parts in equation (4.7), the right part implies the overall economic situation around agent i, which is conceptualized as the average internal fitness of all trade partners. The higher average internal fitness provides firm agent i more valuable resource. Moreover, the left part of equation (4.7) shapes a healthy ratio of economic environment around agent i. It presents the percentages of outside valuable resources that the agent can receive. When all trade partners of agent i are healthy, the ratio reaches the highest at 1, in contrast, the bankruptcy of all partners brings 0 resources to agent i.

$$ER_t^i = \frac{\sum_{k=1}^{p} W_k^i f_{t-1}^k}{\sum_{k=1}^{p} W_k^i} \times \frac{\sum_{k=1}^{p} IF_{t-1}^k}{k} \tag{4.7}$$

Initialization. At the initial state, the model first constructs the firm agents and environment from the given inter-firm trade network. Then, managers of each agent are randomly generated and linked together based on the given inter-firm human network. Within state variables, knowledge vector of each manager is a set of random binary bits, which contributes to a value of manager fitness. In addition, the initial ratio of three agent statuses in the environment is controlled, while the status of each agent is set up at random. Other initial variables of agent and environment are calculated automatically by the model equations (See Table 4.4 and equations).

Emergency. Experiencing a severe performance failure, a firm agent goes bankrupt in the model, which not only makes itself and its manager agents inactive but also creates new risk to related firm agents by deleting its relevant human and trade links. Such dissolution of linkages changes the inter-firm environment consequently over time.

4.5 Simulations and Results

In this section, I introduce the configuration of a series of scenarios and evaluate their simulation results.

4.5.1 The Simulation Configuration

Overall, in each group of my simulation run, the diversity of scenarios are set up by applying different input data for the inter-firm human network. The overview of simulation configuration is presented in Figure 4.11. In the figure, relying on the same inter-firm trade network and a variety of inter-firm human networks, I set up a set of different simulation environments. Then, given each inter-firm environment, I follow the agent-based model work flow in Figure 4.6 and Figure 4.7. In this way, I continuously obverse important features within the network evolution, to measure effects of various influential factors inside this bankruptcy diffusion process.

In this work, I observed the evolving measurements of overall environment and two inter-firm networks from both macro and micro perspectives. Table 4.5 lists several important features mainly considered in my simulation process. To validate the results, I generally collected the data of those properties and average them over multiple runs for each group of simulation (with the same input).

Table 4.5: A list of significant observations in my simulation

Name of Feature	Measurement Level
Temporal number of surviving firm agents	Macro level
Temporal social aspiration level	Macro level
Temporal features of inter-firm human network (for instance density)	Macro level
Temporal features of inter-firm trade network (for instance density)	Macro level
Temporal performance of focal firm agents	Micro level

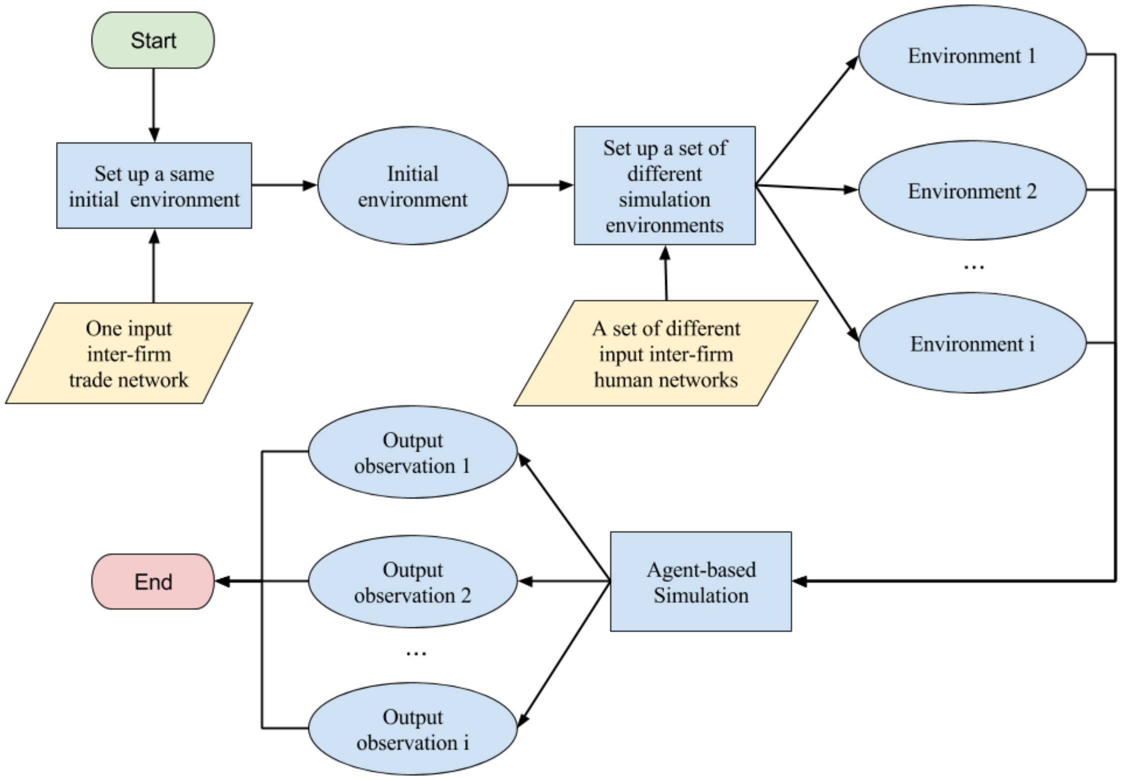


Figure 4.11: The overview of simulation configuration

4.5.2 The Simulation Setup

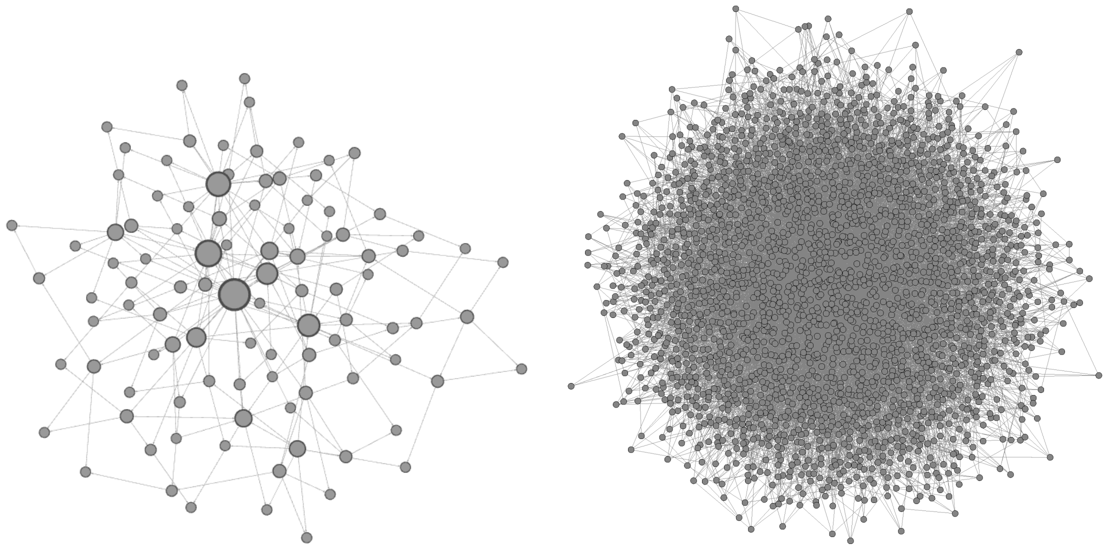
This section describes how to generate all inter-firm networks (i.e. input data) used in the simulations, as well as their parameter settings in details.

Input Data Generation. To keep simulating scenarios consistent with the practical inter-firm environment and validate it implicitly, I involved the real inter-firm networks (presented in section 4.3) in my simulations. Meanwhile, based on properties of real ones, several artificial inter-firm trade and human networks are also generated respectively to construct a variety of scenarios in the simulation.

On the one side, both my empirical analysis in section 4.3 and a recent statistical research reported that the Japanese inter-firm trade network holds a scare-free property [117]. Thus, I apply the well-known Barabsi-Albert model [118] in Networkx[5] to construct all of

[5]Source: https://networkx.github.io/

the used inter-firm trade networks. Two examples of the generated ones are particularly shown in Figure 4.12, with both small and large number of nodes. Among them, the node size in Figure 4.12-a represents the degree of each node, combining its in-degree and out-degree.

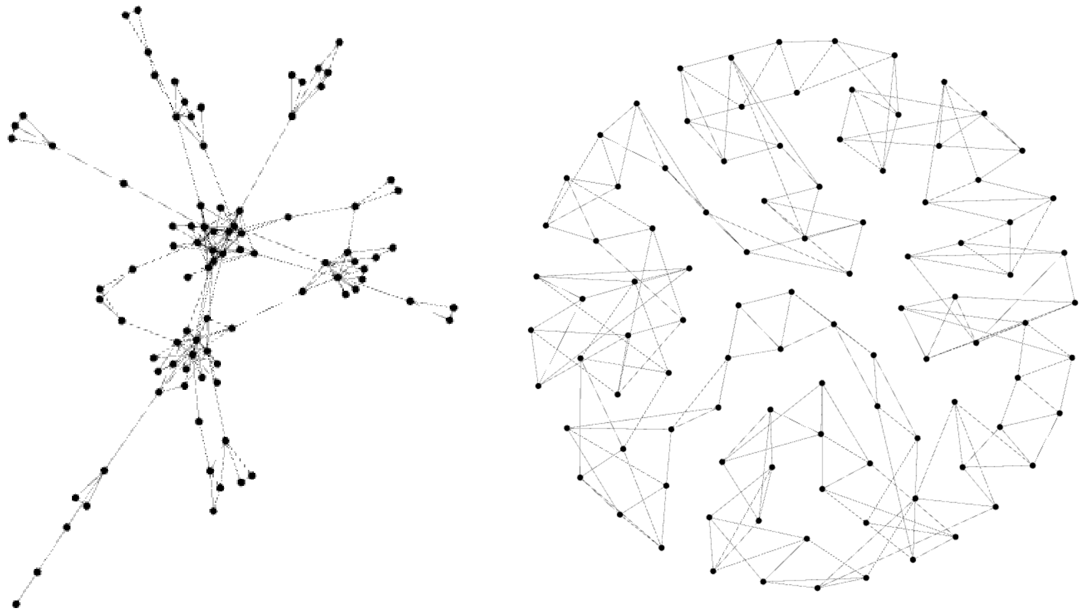


(a) (b)

Figure 4.12: Two examples of the generated inter-firm trade networks used in simulations (with 100 (a) and 2460 (b) nodes respectively)

On the other side, to examine various factors relevant to inter-firm human relationship, I generate the artificial inter-firm human networks in three different ways for a variety of purposes. First, copying a wide community structure and power law degree distribution of real inter-firm human network, Lancichinetti-Fortunato-Radicchi (LFR) benchmark [119] is utilized to generate and evaluate a set of resemble networks holding various number of nodes and the same average degree. Secondly, I generate several regular networks for examining the effects of different average degree, regardless of the influences caused by different number of nodes or topological structure. The examples of generated LFR and regular inter-firm human networks are both shown in Figure 4.13.

Third, the average clustering coefficient is particularly considered to measure the density of triangles in an inter-firm human network [120, 121]. In general, higher average clus-



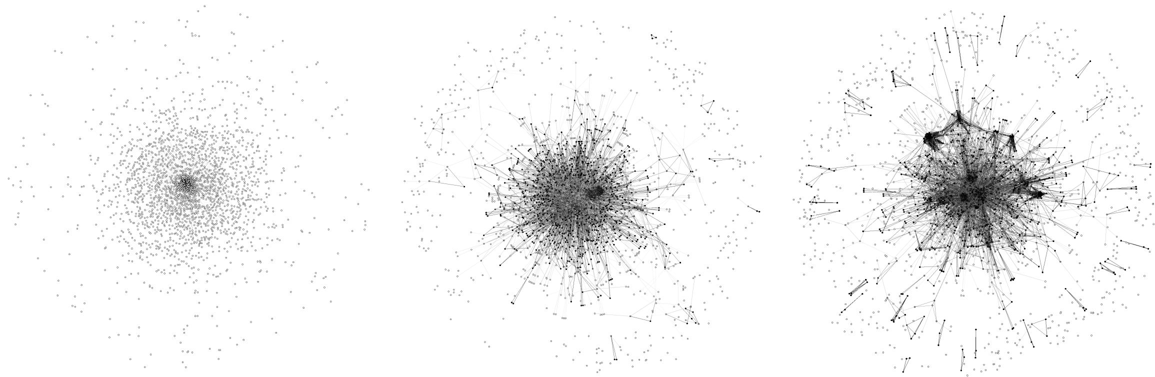
(a) (b)

Figure 4.13: The examples of used LFR (a) and regular (b) inter-firm human networks respectively

tering coefficient means that the network holds a heightened global triangle structure. The RandNetGen tool[6] is employed to generate two artificial networks based on main properties of the real inter-firm human network. This generator algorithm has been well validated [122] and used to study the effect of a certain topological feature in a network [123]. In my work, two generated inter-firm human networks holds the same number of nodes (2460 nodes), edges (20041 edges) and degree distribution with the real one, but adjusting their average clustering coefficient at 0.015 and 0.428 respectively. In this case, the effects of triangle structure in inter-firm human network can be explicitly considered. The observations on this set of inter-firm human networks are shown in Fig. 4.14, where only the triangle closure information remains and the grayscale represents local clustering coefficient value of each node.

Parameter Setting. In my model, there are five senior executives (i.e. manager agents) in each firm (i.e. firm agent), which is as same as the real data. The settings of other

[6]Code Source: https://github.com/polcolomer/RandNetGen



(a) (b) (c)

Figure 4.14: The triangle-related snapshots of three used inter-firm human networks holding average clustering coefficient at 0.015 (a), 0.428 (b) and 0.467 (c) respectively (where panel (c) is the real extracted network, and panel (a) and (b) are artificially generated).

parameters are presented in Table 4.6. I evaluate several groups of parameter settings and select the one with the most stable results. In NK model, the contribution and elements' relation table are both randomly generated at the beginning of each simulation. Initially, for each simulation run the statuses of all agents are randomly generated once, following a ratio of 1% bankrupt, 9% distressed and 90% healthy. To include varying trade amounts in the trade network, the weight on every trade link is also selected randomly in set W. Each simulation is running for 20 timestamps (i.e. 10 years), as each time step represents half a year. Then, this is repeated for either 50 or 100 times in different experiments.

Table 4.6: Parameter setting.

Parameter name	Setting
NK model	N = 10, K = 2
Initial agent status ratio	Bankrupt = 5%, Distressed = 25%, Healthy = 70%
Weight in equation (4.7)	$W_k^i \in W,$ where $W = \{1, 5, 10, 50\}$
Weight in equation (4.1)	$\alpha = 0.5$
Weight in equation (4.4)	$\varepsilon = 0.5$

4.5.3 Simulations and Results

In this subsection, I describe the observations in simulations and discuss their results. Overall, three groups of experiments are conducted to evaluate: 1) evolutionary effects of the human relationship; 2) evolutionary effects of different number of inter-firm human links; 3) evolutionary effects of different triangle structure in inter-firm human network.

1. Impact of the Human Relationship. To study the effects of human relationship, in the first set of experiments, I simulated the inter-firm trade network evolving scenarios with and without the inter-firm human network. When there are no human links, manager agents adapt their knowledge vectors only by local search in the agent-based model.

The simulation results of the both scenarios are shown in Figure 4.15, including temporal agent's bankrupt information (upper side) and social aspiration level (lower side) in three individual experiments. Simulations with 200, 300 and 500 firm agents are set up respectively based on the same parameter setting (given in Table 4.6) and same input networks (shown in Figure 4.12-a and 4.13-a for the inter-firm trade and human network).

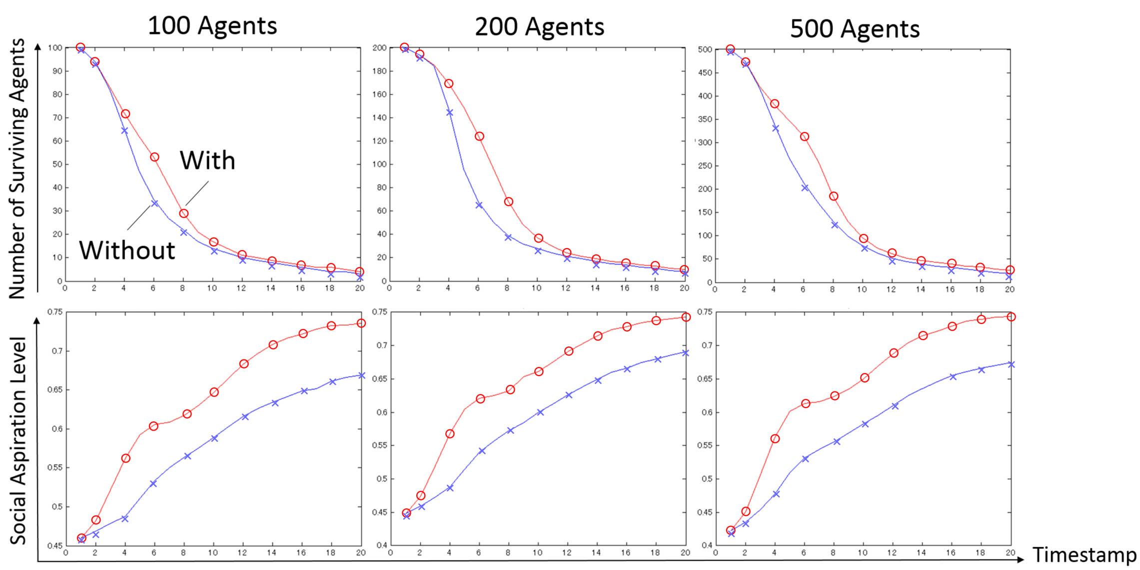


Figure 4.15: Temporal bankrupt and social aspiration level information with human relationship ($\circ$) and without human relationship ($\times$)

According to Figure 4.15, similar results are observed in three experiments with various number of agents. When bankrupt happens in the inter-firm trade network, the size of surviving agents follows a power-law distribution, which matches the empirical evidence of firm size distribution [124]. By comparing two scenarios, I noticed that bankrupt diffusion of the inter-firm trade network delays and the overall social aspiration level increases, when having inter-firm human relationship. This stable phenomenon indicates that the human relationship is influential on inter-firm bankrupt transferring and promotes the average environmental performance.

2. Impact of Different Number of Human Partners. Having observed the effect of defined human relationship in the above experiment, I further understand the impact of different number of human partners, which is conceptualized as degree in the inter-firm human network, from both macro and micro perspectives.

Macro-level Degree in the Inter-firm Human Network. In the macro level, I evaluated the effects of different number of human partners on the overall bankrupt evolution process, by examining the number of surviving firm agents and average performance over time. The simulation results in this set of experiments are shown in Figure 4.16. Given the same initial inter-firm trade network, I set up a group of inter-firm human networks holding different average degree (representing various number of overall human links in the network). Here, regular network (as shown in Figure 4.13-b) is applied to minimum the effects of inter-firm human network structure. Also, each group of simulation includes 100 agents with the same parameter setting (given in Table 4.6).

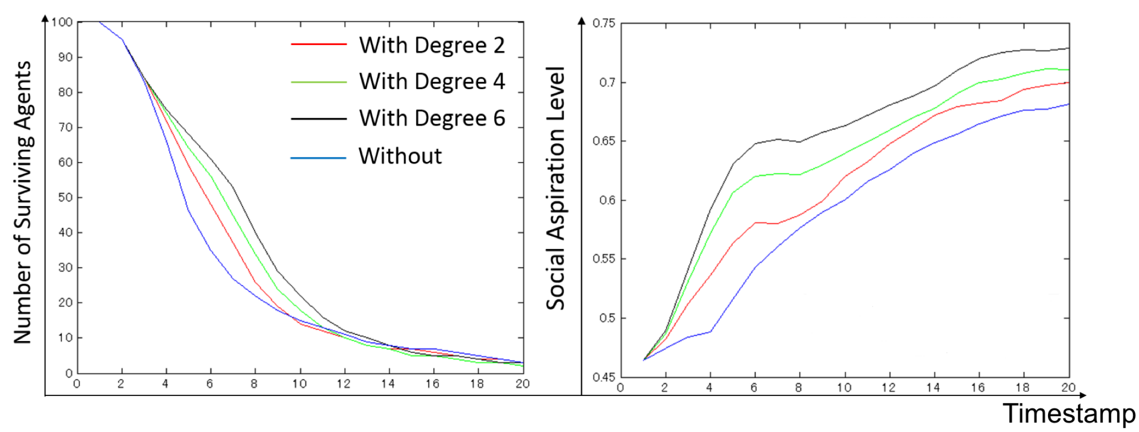


Figure 4.16: Temporal bankrupt and social aspiration level information for inter-firm human networks having different average degree

The results indicate that different number of inter-firm human links has various effects on the bankrupt evolutionary process at the macro level. In general, higher average degree (i.e. more inter-firm human links in the network) benefits more in the bankrupt diffusion process and promotes a better social aspiration level. It implies that, in a denser connected inter-firm social environment, firms can receive more beneficial help when facing bankrupt emergency.

Micro-level Degree in the Inter-firm Human Network. The local human degree of each firm agent is also studied to explore whether different number of human partners can affect the individual firm agent in the bankrupt risk.

When one trade partner of a focal firm agent bankrupts, it can be strongly affected and go bankrupt as well [125]. Thus, the following scenario is designed particularly in this experiment. One important trade agent b (for instance the biggest node in Figure 4.12-a) is initialized in the bankrupted status. While one of its trade partner agent f has different number of human partners (i.e. local human degree), its performance in the subsequent time periods is observed and analyzed. In this simulation, both trade and human networks have 100 firm agents and share the same network structure (Figure 4.12-a and 4.13-a). In particular, to make sure the bankrupt firm agent b brings the same influence to partner agent f in each simulation, I set up the weight of trade link W_b^f as 1, instead of a random value in W. Figure 4.17 presents the average performance of the focal agent f with different local human degree in 50 simulation runs.

The different effects of local human degree are observed in above Figure 4.17. In the early time after bankruptcy of the trade partner, higher local degree (i.e. more human partners) brings more benefits in firm agent's performance. In addition, I averaged the results of performance in 50 simulation runs, and it can be 0 once the firm is bankrupted. Thus, the shown performance also implies a probability of going bankrupt. According to the results, while the firm agent with the fewest human ties (blue line in Figure 4.17) are most likely to bankrupt, firm agents with more human links (red line in Figure 4.17) usually survive as bankrupt transferring. This result suggests that a great number of individual

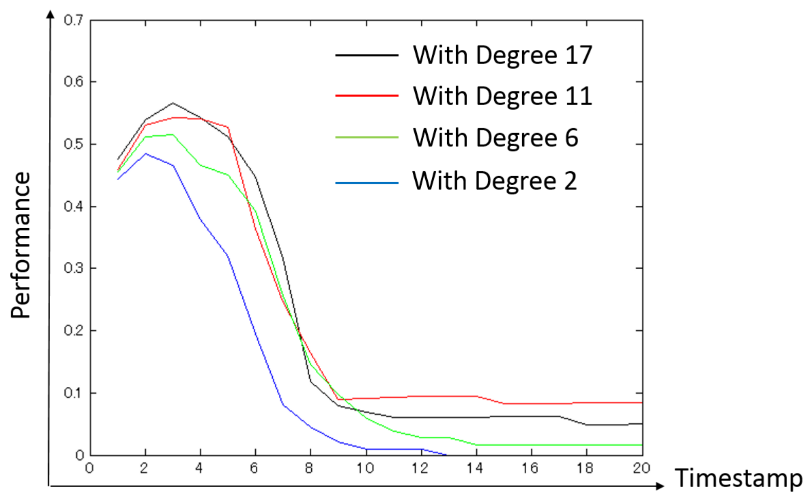


Figure 4.17: Temporal performance of focal firm agent f with various local human degree (averaged in 50 runs)

human partners not only is beneficial in bankrupt diffusion but also can help the firm to recover from an urgent bankrupt influence.

At the same time, an interesting phenomenon is shown that firm agents with higher human degree do not always result in a higher performance in the long-term. Such phenomenon matches the data analysis result that firms with high human degree may go bankrupt. It drives us to explore other important mechanisms related to this inter-firm human relationship, as human degree is not the only influential factor in this bankruptcy phenomenon.

3. Impact of Different Triangle Structure in Inter-firm Human Network. Next, the impact of triangle structure in inter-firm human network is discussed in various simulation scenarios. Overall, three sets of results are observed and analyzed to answer the following questions: in bankruptcy evolution phenomena, 1) whether the triangle structure affects; 2) whether the triangle structure affects who go bankrupt; 3) whether the triangle structure survives.

Whether Triangle Structure Affects. To study the impact of triangle structure, I first simulated the bankrupt evolution of a same inter-firm trade network (Figure 4.12-b) using inter-firm human networks with and without triangle structure (Figure 4.14-c and Figure 4.14-a respectively). The temporal bankruptcy information and social aspiration level are presented in Fig. 4.18.

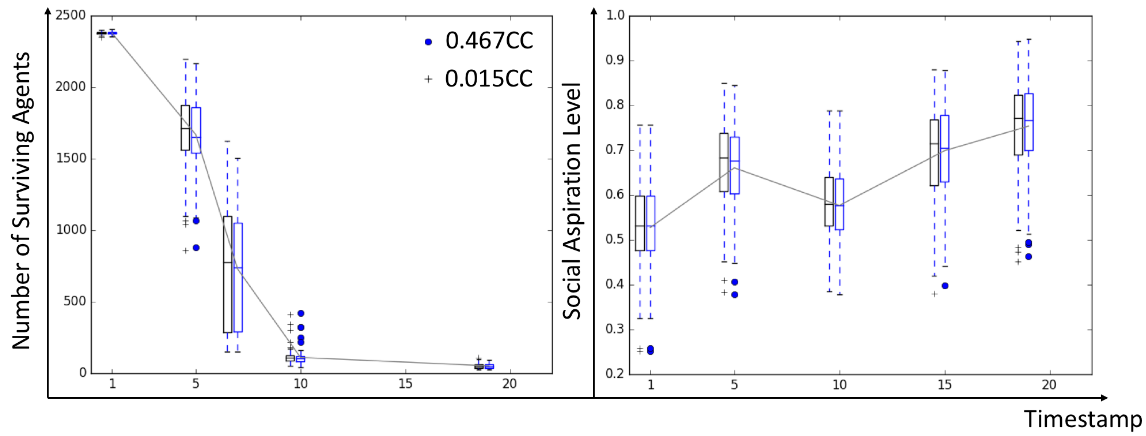


Figure 4.18: Temporal bankrupt and social aspiration level information for inter-firm human networks without (almost to zero) and with the triangle structure

In Fig. 4.18, blue circle (O) shows information of the real inter-firm human network with average clustering coefficient (abbreviated as CC) at 0.467, while black plus (+) represents the artificial network with 0.015 CC. It shows a similar trend between two scenarios, not only at the median value but also within the overall scale. This stable phenomenon indicates that presence or lack of triangle structure in inter-firm human network does not significantly affect the global bankrupt distribution and average performance in the inter-firm trade network.

Whether Triangle Structure Affects Who bankrupted. Three inter-firm human networks (given in Figure 4.14) with different average clustering coefficients are used with the same inter-firm trade network (Figure 4.12-b) in this set of simulations. To further study how different triangle structure in inter-firm human network affect the individual firms, I consequently analyzed some features of evolving human and trade networks during the bankruptcy diffusion process. Figure 4.19 shows the comparison among various scenarios, measuring the density of both trade and human network, as well as average clustering coefficient (abbreviated as CC) of the human networks.

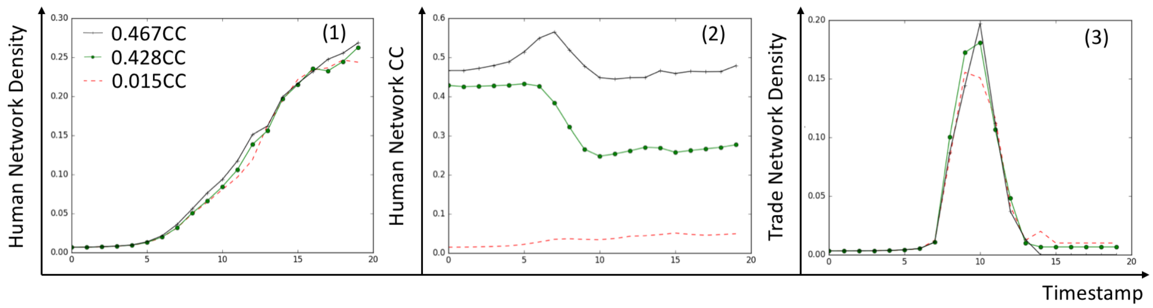


Figure 4.19: Three temporal measurements of evolving inter-firm networks (where black plus (+) represents the real inter-firm human network extracted in section 4.3; green circle (o) and red dotted line (- - -) are artificial networks)

In panel (2) of Figure 4.19, networks with higher CC keep their advantages on CC during the bankrupt propagation. In the real data case (i.e. black +), the global CC of human network is particularly rising at the early stage of bankruptcy, which implies that firm agents (i.e. firms) with lower local CC (i.e. within fewer triangles) have higher risk to bankrupt at the beginning of the bankruptcy emergency.

The different effects of triangle structure in inter-firm human network are also found in Figure 4.19-(1) and Figure 4.19-(3). In general, the higher CC drives a higher density in both inter-firm human and trade networks. It means the firm agents (i.e. firms) holding high trade and human degree are easier to survive during the bankrupt diffusion if they involve in a social environment with a great many of triangles (i.e. higher CC). This result indicates that the triangle structure in inter-firm human networks can help firms to obtain more benefits from partners via inter-firm links.

In addition, I observed a correlated influence between the trend of CC in inter-firm human network and density (equivalent to average degree) in inter-firm trade network. Considering Figure 4.19-(1) and Figure 4.19-(2), when triangle structure in human network starts to decrease quickly at simulation time stamp 7 (referring to wide broken of triangle structure), the density of trade network follows a dramatic increase. Afterwards, the density of inter-firm trade network returns back, while the CC of human network becomes stable again around simulation time 10. Such phenomenon suggests that the change of CC in the inter-firm human network affects which firms go bankrupt in the inter-firm trade network. Moreover, since the big increase of density in inter-firm trade network implies that most firm agents with low local trade degree are getting bankrupt during this time, while firms with more trade partners still can survive. This observation indicates that the dissolution of triangle structure in inter-firm human network can cause the collapse of firms with few local trade partners. Meanwhile, a stable triangle structure in the inter-firm human network (before time 7 or after time 10) can balance the chance of firms with different trade degrees to survive in a bankrupt emergency.

Whether Triangle Structure Survives. According to Figure 4.19-(2) and Figure 4.18, both average clustering coefficient of inter-firm human networks and bankruptcy diffusion is getting smooth again in the long-term. It drives us to further explore the role of triangle structure in the remaining real human network after this dramatic bankruptcy. Thus, I took a snapshot of topological structure in the real inter-firm human network at simulation time stamp 10, which is shown in Figure 4.20.

In total, the above network has 17 nodes, 21 links and 10 triangles (i.e. average clustering coefficient at 0.43). The survived firm agents (i.e. firms) holds a strong triangle structure in the left real inter-firm human network. This result indicates again the significant role of triangle structure in the inter-firm human network in this bankruptcy phenomenon. In general, a strong triangle structure helps the firms to survive after the bankrupt emergency.

4.6 Concluding Remarks

This chapter introduces the study on dynamic evolutionary phenomenon in inter-firm trade network emerging from bankruptcy. To understand the underlying mechanism within this process, I concentrate on one influential but overlooked human-related factor (i.e. the inter-firm human relationship raised by firms' senior executives), and explore its influence in this phenomenon based on an inter-firm human network structure. First, the real inter-firm hu-

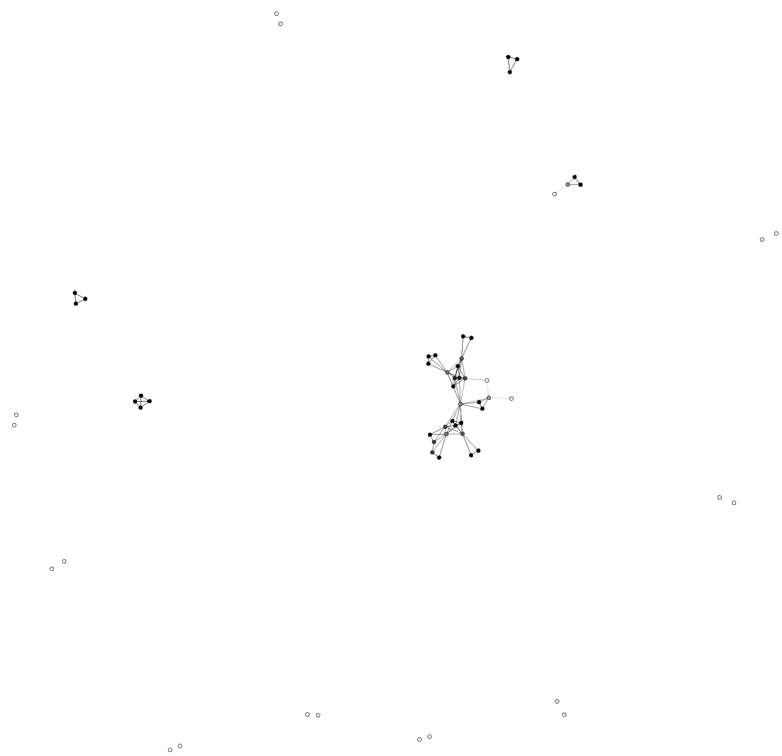


Figure 4.20: A snapshot of the remaining real inter-firm human network after the dramatic bankruptcy

man and trade networks are extracted and analyzed from empirical Japanese firm data over 10 years. Then, an agent-based model is particularly proposed by applying both empirical properties and theoretical supports. By setting up the same input inter-firm trade network with various inter-firm human networks, I simulated three sets of diverse scenarios and observed well-studied measurements of both inter-firm networks consequently. Overall, the validated observations indicate the influential role of defined inter-firm human relationship and its related features: it not only affects overall bankrupt diffusion process but also plays a significant role in determining which individual firms have higher risk of going bankrupt. These implications can be further explained from three perspectives: 1) the inter-firm human relationship generally delays bankrupt diffusion and promotes average social performance in the market; 2) more human partners benefits related firms to survive in a bankrupt emergency in both macro and micro aspects; 3) a stable triangle structure in the inter-firm human network provides individual firms with fewer trade partners the equal chance to survive in the bankrupt emergency. Given those implications, I expect to provide managerial guidelines about proper strategies (relevant to this human relationship) in defending bankrupt emergencies.

CHAPTER 5
DISCUSSION AND IMPLICATION

This chapter summarizes the correlations between two research problems and discusses the significance of studying those data-driven complex social phenomena. Moreover, I argue the good perspectives of expanding the study on other computational social problems.

5.1 Correlations of Two Researches

Two important social phenomenons are studied in the context, rumor spreading in online social media and bankruptcy propagation among firms. Besides that both pieces of research focus on the dynamic evolution phenomena of social networks caused by external factors (i.e. rumor or bankruptcy), there is another significant correlation between their methodology, which is shown in Figure 5.1.

According to the figure, although two research problems studied in this thesis share a similar technical procedure in understanding the corresponding social problem. Based on various social theories, both practical research problems are firstly identified and driven by social phenomena observed in empirical social data. From a network-dependent perspective, two social networks are processed afterward to describe involved social entities and their relations. Applying various computational approaches in these networks (the technical usage can be flexible because of diverse research requirements), the analysis and explanation of evolutionary patterns can help us to well understand the complex social phenomenon.

The methodological procedure in Figure 5.1 not only is a common base for two studied social phenomena but also have a broader applicability for other problems in computational social science, whose importance and implications are discussed in the next section.

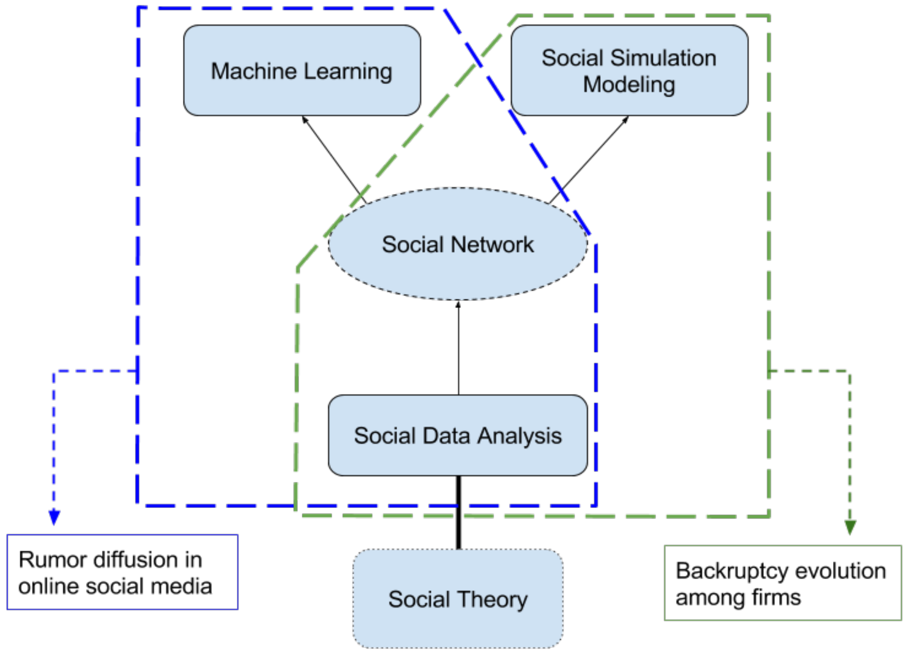


Figure 5.1: The correlation between two research tasks in methodology

5.2 Implications in Computational Social Science

In this section, I address the implications of our work in the emerging research area, computational social science.

Computational social science (CSS) is an interdisciplinary academic field that integrates a variety of innovative methodologies in both computational science and social science. In 2009, it was firstly identified as a new research area in a publication of *Science* [126]. Lazer et al. empathized the rise of computational capacity and indicated the important role of CSS in revolutionizing our understanding in lives, organizations, and societies. Since this remarkable starting point, CSS attracts more and more attention and experienced an extraordinary academical increase in the last years [127].

Defined as the ′***interdisciplinary investigation of the social universe on many scales, ranging from individual actors to large groupings, through the medium of computation***′ [128], CSS analyzes, models and simulates the social phenomenon by applying computational approaches. In particular, under the presence of data flood, it creates new opportuni-

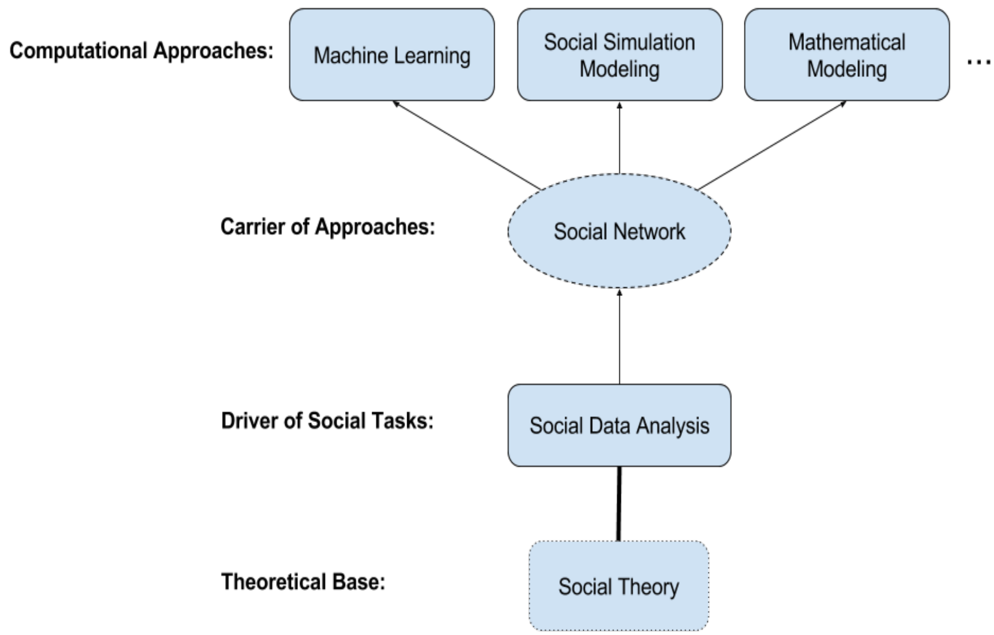


Figure 5.2: A methodological research paradigm of computational social science

ties for research to quantify the study on complex social systems [129].

The field of data-driven computational social science covers a wide range of innovative approaches and practical categories [130]. In fact, both of social tasks studied in this thesis belong to the scope of data-driven computational social science phenomena. From the CSS perspective, our work provides a paradigm for understanding data-driven complex social systems and their evolutionary. Figure 5.2 presents this technical research procedure in details.

Starting from the social theories, computational social phenomena can be studied based on empirical big data first, which addresses social tasks to be potentially explored further. Then, the social network structure is employed to describe complex social actors, relations, and patterns formed by them. As a carrier of methods, it provides a good technical environment, so various computational approaches (including but not limited to machine learning, social simulation, and mathematical modeling) can be developed smartly to understand social problems. In the nutshell, such methodological research paradigm can be easily applied or expended in other computational social problems to understand their complex systems and evolutions. This provides an important guidance for scientists in CSS on where to look for questions and how to seek solutions.

5.3 Concluding Remarks

In this chapter, I discuss one important methodological correlation between our two pieces of research, as well as its implications in computational social science research field. Moreover, I indicated the potential to explore other computational social phenomena using the proposed methodological research paradigm in future.

CHAPTER 6
CONCLUSION

In the thesis, I study two evolutionary phenomena of social networks: rumor diffusion in online social media and bankruptcy evolution among the firms. Driven by real big data analysis, I used a variety of computational technologies to understand the behavioral patterns in the dynamic social networks from both interpersonal and inter-organizational levels. The main contributions of my work are summarized in the following list:

1) In the dynamic phenomenon of rumor diffusion in online social media, a methodology is proposed to discover rumor-related patterns at the early stage of trending rumor events in streaming social media data. First, the rumor patterns are designed regarding both social network structural features and user behavioral features. Second, a pattern matching algorithm is particularly developed to track these patterns in streaming data automatically. Third, the sliding window mechanism is applied in a framework to detect and analyze the snapshots of the data stream in the real-time. Experiments in two different real Twitter datasets show that my approach reliably captures early signal patterns of trending rumors and very potentially useful in real-time rumor detection system.

2) In the dynamic phenomenon of bankruptcy evolution among firms, I addressed the influential role of social network between senior executives in the bankrupt propagation process among firms. The empirical data of millions of Japanese firms are first analyzed, including properties of the extracted inter-firm social network based on senior executives' information. Second, an agent-based model is proposed to examine various factors in the evolutionary process of this phenomenon using the simulation technique. According to the results from various simulating scenarios, the influential role of inter-firm social network is evaluated. In general, the resilience of firms facing bankrupt emergency can benefit from the large number of social partners and a stable local interaction mechanism among firms (i.e. triangle structure in the inter-firm social network).

3) Overall, the combination of two studies provided a good perspective to understand evolutionary individual and interactive behaviors, and their effects in the complex social phenomenon. Regarding the correlations between two research, I proposed a methodological procedure that is commonly applicable to explore other social phenomena. Given this implication, I expect to apply the same function to explore the underlying nature in other computational social problems.

This thesis provides a good start point for understanding complex dynamic social phenomena. Still, there are many potential directions that can be explored in two individual research, carrying out a variety of potential computational social tasks to be advanced in a long-term future.

1) For the research of rumor diffusion in online social media, I propose to integrate the work into a real-time trending rumor event discovery and analysis system (which is part of on-going work). Overall, the behavior of framework is modeled as a state-transition process including three possible states: normal (the initial state which mainly contains work in this context), alarm and detected. The state transition is triggered by two decision mechanisms, that process streams of data and raise certain flags based on the observed features and patterns. The first decision mechanism is implemented by the early signal alarm: if this initial step detects the presence of early signals in trending rumor events (which can be conducted by the work presented in section 3), it triggers a state transition from normal to alarm state. Moreover, while the framework is in alarm state, a second decision mechanism further investigates the flags risen by the early detection. During this decision step, each candidate might transit into two states: the normal one, if there was not sufficient data or patterns for a reliable decision making; or in the detected state, if the framework is able to validate (with high accuracy) a trending rumor event as false. Using this early detection framework of trending rumor events, I expect to take immediate actions to restrict and stop its spread by limiting the audience it can reach or emphasize the accurate information.

2) For the research of bankruptcy evolution among firms, I first propose to further explore the causal effects of defined human factors in this phenomenon, beyond their correlated impact studied in this work. Thus, multiple micro-level measurements of firms could be examined consequently using the simulation. Secondly, I would like to enhance empirical analysis of this work from various perspectives, such as involving real bankrupt chain evolutionary data for the validation, expanding other important behavioral features relevant to the defined inter-firm social relationship, or even exploring effects of other inter-firm social relations (like data of social relationship between firms' senior executives). In addition, the work is currently based on data in one particular industry of one country. As my agent-based model is conceptually designed from general organization theory, I conjecture that the model can be extended to other business environments by feeding extracted features in different data. A behavioral comparison between various environments could also be a promising direction.

3) The combination of two studies has good implications for understanding the evolutionary phenomena and its underlying factors in complex social problems. Thus, my proposed methodological research paradigm is particularly applicable to be used in a variety of other computational social problems. In real society, emerging events (such as traffic accident, big social events, and health disaster) always affect the social and business environment dramatically and cause serious concern in the crowd [131]. From a longer term, I expect to expand this work and make more contributions in computational social problems.

Appendices

APPENDIX A

PUBLICATION LIST

Paper 1: S. Wang, and T. Terano. ”Detecting Rumor Patterns in Streaming Social Media.” Proceedings of 2015 IEEE International Conference on Big Data (Big Data). IEEE, 2015. $DOI: 10.1109/BigData.2015.7364071T$

Paper 2: S. Wang, M.J. Songhori, S. Chang and T. Terano. ”The Impact of Human Relationship on Bankruptcy-related Evolution of Inter-firm Trade Network.” Proceedings of Winter Simulation Conference (WSC). IEEE, 2016. $DOI: 10.1109/WSC.2016.7822371$

Paper 3: S. Wang, M.J. Songhori, S. Chang and T. Terano. ”How Triangle Structure in Inter-firm Human Network Affects Bankruptcy Evolution: An Agent-based Simulation Study with Real and Artificial Data.” Advances in Human Factors in Simulation and Modeling. Springer International Publishing, 2017. $DOI: 10.1007/978-3-319-60591-3\_26$

Paper 4: S. Wang, I. Moise, D. Helbing and T. Terano. ”Early Signals of Trending Rumor Event in Streaming Social Media.” Proceedings of IEEE 41th Annual Computer Software and Applications Conference (COMPSAC). IEEE, 2017. $DOI: 10.1109/COMPSAC.2017.115$

BIBLIOGRAPHY

[1] Mark Newman. *Networks: an introduction*. Oxford university press, 2010.

[2] Alexandra Marin and Barry Wellman. "Social network analysis: An introduction". In: *The SAGE handbook of social network analysis* 11 (2011).

[3] Gueorgi Kossinets and Duncan J Watts. "Empirical analysis of an evolving social network". In: *science* 311.5757 (2006), pp. 88–90.

[4] Xingjie Liu et al. "Event-based social networks: linking the online and offline social worlds". In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2012, pp. 1032–1040.

[5] Albert-Laszlo Barabâsi et al. "Evolution of the social network of scientific collaborations". In: *Physica A: Statistical mechanics and its applications* 311.3 (2002), pp. 590–614.

[6] Eunjoon Cho, Seth A Myers, and Jure Leskovec. "Friendship and mobility: user movement in location-based social networks". In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2011, pp. 1082–1090.

[7] Nathan Eagle, Alex Sandy Pentland, and David Lazer. "Inferring friendship network structure by using mobile phone data". In: *Proceedings of the national academy of sciences* 106.36 (2009), pp. 15274–15278.

[8] Patti M Valkenburg, Jochen Peter, and Alexander P Schouten. "Friend networking sites and their relationship to adolescents' well-being and social self-esteem". In: *CyberPsychology & Behavior* 9.5 (2006), pp. 584–590.

[9] Shihan Wang et al. "Early Signals of Trending Rumor Event in Streaming Social Media". In: *Computer Software and Applications Conference (COMPSAC), 2017 IEEE 41th Annual*. Vol. 2. IEEE. 2017, pp. 654–659.

[10] Charu Aggarwal and Karthik Subbian. "Evolutionary network analysis: A survey". In: *ACM Computing Surveys (CSUR)* 47.1 (2014), p. 10.

[11] Andr Vermeij. *Apples Internal Innovation Network Unraveled Part 1 Evolving Networks*. https://www.kenedict.com/apples-internal-innovation-network-unraveled-part-1-evolving-networks/. 2013. (Visited on 07/30/2017).

[12] Patrick Doreian and Frans N Stokman. *Evolution of social networks*. Vol. 1. Psychology Press, 1997.

[13] Ravi Kumar et al. “On the bursty evolution of blogspace”. In: *World wide web* 8.2 (2005), pp. 159–178.

[14] Ravi Kumar, Jasmine Novak, and Andrew Tomkins. “Structure and evolution of online social networks”. In: *Link mining: models, algorithms, and applications*. Springer, 2010, pp. 337–357.

[15] Jure Leskovec et al. “Microscopic evolution of social networks”. In: *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2008, pp. 462–470.

[16] Paul Erdos and Alfréd Rényi. “On the evolution of random graphs”. In: *Publ. Math. Inst. Hung. Acad. Sci* 5.1 (1960), pp. 17–60.

[17] Réka Albert and Albert-László Barabási. “Statistical mechanics of complex networks”. In: *Reviews of modern physics* 74.1 (2002), p. 47.

[18] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. “Graph evolution: Densification and shrinking diameters”. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1.1 (2007), p. 2.

[19] Matthew O Jackson and Alison Watts. “The evolution of social and economic networks”. In: *Journal of Economic Theory* 106.2 (2002), pp. 265–295.

[20] Tom AB Snijders, Gerhard G Van de Bunt, and Christian EG Steglich. “Introduction to stochastic actor-based models for network dynamics”. In: *Social networks* 32.1 (2010), pp. 44–60.

[21] Deepayan Chakrabarti and Christos Faloutsos. “Graph mining: Laws, generators, and algorithms”. In: *ACM computing surveys (CSUR)* 38.1 (2006), p. 2.

[22] Béla Bollobás and Oliver M Riordan. “Mathematical results on scale-free random graphs”. In: *Handbook of graphs and networks: from the genome to the internet* (2003), pp. 1–34.

[23] Stefano Boccaletti et al. “Complex networks: Structure and dynamics”. In: *Physics reports* 424.4 (2006), pp. 175–308.

[24] Riitta Toivonen et al. “A comparative study of social network models: Network evolution models and nodal attribute models”. In: *Social Networks* 31.4 (2009), pp. 240–254.

[25] Lars Backstrom et al. “Group formation in large social networks: membership, growth, and evolution”. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2006, pp. 44–54.

[26] Myra Spiliopoulou. “Evolution in social networks: A survey”. In: *Social network data analytics*. Springer, 2011, pp. 149–175.

[27] Stephen P Borgatti et al. “Network analysis in the social sciences”. In: *science* 323.5916 (2009), pp. 892–895.

[28] Manoochehr Ghiassi, James Skinner, and David Zimbra. “Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network”. In: *Expert Systems with applications* 40.16 (2013), pp. 6266–6282.

[29] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. “Earthquake shakes Twitter users: real-time event detection by social sensors”. In: *Proceedings of the 19th international conference on World wide web*. ACM. 2010, pp. 851–860.

[30] Sergei N Dorogovtsev and José FF Mendes. *Evolution of networks: From biological nets to the Internet and WWW*. OUP Oxford, 2013.

[31] Luciano da Fontoura Costa et al. “Analyzing and modeling real-world phenomena with complex networks: a survey of applications”. In: *Advances in Physics* 60.3 (2011), pp. 329–412.

[32] David Burth Kurka, Alan Godoy, and Fernando J Von Zuben. “Online social network analysis: A survey of research applications in computer science”. In: *arXiv preprint arXiv:1504.05655* (2015).

[33] Duncan J Watts and Steven H Strogatz. “Collective dynamics of’small-world’networks”. In: *nature* 393.6684 (1998), p. 440.

[34] Ranjay Gulati. “Alliances and networks”. In: *Strategic management journal* 19.4 (1998), pp. 293–317.

[35] Shihan Wang et al. “The impact of human relationship on bankruptcy-related evolution of inter-firm trade network”. In: *Winter Simulation Conference (WSC), 2016*. IEEE. 2016, pp. 3405–3416.

[36] Derek Greene, Donal Doyle, and Padraig Cunningham. “Tracking the evolution of communities in dynamic social networks”. In: *Advances in social networks analysis and mining (ASONAM), 2010 international conference on*. IEEE. 2010, pp. 176–183.

[37] Shihan Wang and Takao Terano. “Detecting rumor patterns in streaming social media”. In: *Big Data (Big Data), 2015 IEEE International Conference on*. IEEE. 2015, pp. 2709–2715.

[38] Nicole B Ellison, Charles Steinfield, and Cliff Lampe. “The benefits of Facebook friends: Social capital and college students use of online social network sites”. In: *Journal of Computer-Mediated Communication* 12.4 (2007), pp. 1143–1168.

[39] Haewoon Kwak et al. “What is Twitter, a social network or a news media?” In: *Proceedings of the 19th international conference on World wide web*. ACM. 2010, pp. 591–600.

[40] Bongwon Suh et al. “Want to be retweeted? large scale analytics on factors impacting retweet in twitter network”. In: *Social computing (socialcom), 2010 ieee second international conference on*. IEEE. 2010, pp. 177–184.

[41] Anna Grandori and Giuseppe Soda. “Inter-firm networks: antecedents, mechanisms and forms”. In: *Organization studies* 16.2 (1995), pp. 183–214.

[42] Howard E Aldrich, Ben Rosen, and Bill Woodward. “The Impact of Social Networks on Business Foundings and Profit: A Longitudinal Study”. In: *Frontiers of Entrepreneurship Research* (1987), pp. 154–168.

[43] Anat BarNir and Ken A Smith. “Interfirm Alliances in the Small Business: The Role of Social Networks”. In: *Journal of Small Business Management* 40.3 (2002), pp. 219–232.

[44] Gordon W Allport and Leo Postman. *The psychology of rumor.* 1947.

[45] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. “Meme-tracking and the dynamics of the news cycle”. In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2009, pp. 497–506.

[46] Michela Del Vicario et al. “The spreading of misinformation online”. In: *Proceedings of the National Academy of Sciences* 113.3 (2016), pp. 554–559.

[47] Adrien Friggeri et al. “Rumor Cascades.” In: *ICWSM*. 2014.

[48] Arkaitz Zubiaga et al. “Towards Detecting Rumours in Social Media”. In: *arXiv preprint arXiv:1504.04712* (2015).

[49] Chengcheng Shao et al. “Hoaxy: A platform for tracking online misinformation”. In: *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee. 2016, pp. 745–750.

[50] Jacob Ratkiewicz et al. “Truthy: mapping the spread of astroturf in microblog streams”. In: *Proceedings of the 20th international conference companion on World wide web*. ACM. 2011, pp. 249–252.

[51] Amira Ghenai and Yelena Mejova. “Catching Zika Fever: Application of Crowdsourcing and Machine Learning for Tracking Health Misinformation on Twitter”. In: *arXiv preprint arXiv:1707.03778* (2017).

[52] Xing Zhou et al. “Real-Time News Certification System on Sina Weibo”. In: *Proceedings of the 24th International Conference on World Wide Web*. ACM. 2015, pp. 983–988.

[53] Robert H Knapp. “A psychology of rumor”. In: *Public Opinion Quarterly* 8.1 (1944), pp. 22–37.

[54] Arkaitz Zubiaga et al. “Detection and Resolution of Rumours in Social Media: A Survey”. In: *arXiv preprint arXiv:1704.00656* (2017).

[55] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. “Information credibility on twitter”. In: *Procs of the 20th Intl Conf on World wide web*. ACM. 2011, pp. 675–684.

[56] Sejeong Kwon et al. “Prominent features of rumor propagation in online social media”. In: *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE. 2013, pp. 1103–1108.

[57] Ke Wu, Song Yang, and Kenny Q Zhu. “False rumors detection on sina weibo by propagation structures”. In: *31st Intl Conf on Data Engineering*. IEEE. 2015.

[58] Jing Ma et al. “Detect rumors using time series of social context information on microblogging websites”. In: *Procs of the 24th ACM Intl Conf on Information and Knowledge Management*. ACM. 2015.

[59] Eni Mustafaraj Markus Strohmaier Harald Schoen Gayo-Avello Panagiotis Takis Metaxas et al. “Predicting information credibility in time-sensitive social media”. In: *Internet Research* 23.5 (2013), pp. 560–588.

[60] Xiaomo Liu et al. “Real-time rumor debunking on twitter”. In: *Procs of the 24th ACM Intl Conf on Information and Knowledge Management*. ACM. 2015, pp. 1867–1870.

[61] Sejeong Kwon, Meeyoung Cha, and Kyomin Jung. “Rumor detection over varying time windows”. In: *PloS one* 12.1 (2017), e0168344.

[62] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. "Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts". In: *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee. 2015, pp. 1395–1405.

[63] Michael Mathioudakis and Nick Koudas. "Twittermonitor: trend detection over the twitter stream". In: *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*. ACM. 2010, pp. 1155–1158.

[64] Jianshu Weng and Bu-Sung Lee. "Event detection in twitter." In: *ICWSM* 11 (2011), pp. 401–408.

[65] Mario Cataldi, Luigi Di Caro, and Claudio Schifanella. "Emerging topic detection on twitter based on temporal and social terms evaluation". In: *Proceedings of the tenth international workshop on multimedia data mining*. ACM. 2010, p. 4.

[66] Saša Petrović, Miles Osborne, and Victor Lavrenko. "Streaming first story detection with application to twitter". In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics. 2010, pp. 181–189.

[67] Xiangmin Zhou and Lei Chen. "Event detection over twitter social media streams". In: *The VLDB journal* 23.3 (2014), pp. 381–400.

[68] Qi Dang, Feng Gao, and Yadong Zhou. "Early detection method for emerging topics based on dynamic bayesian networks in micro-blogging networks". In: *Expert Systems with Applications* 57 (2016), pp. 285–295.

[69] Vahed Qazvinian et al. "Rumor has it: Identifying misinformation in microblogs". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. 2011, pp. 1589–1599.

[70] Rob Ennals et al. "What is Disputed on the Web?" In: *Proceedings of the 4th workshop on Information credibility*. ACM. 2010, pp. 67–74.

[71] Weiling Chen et al. "Behavior deviation: An anomaly detection view of rumor preemption". In: *Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016 IEEE 7th Annual*. IEEE. 2016, pp. 1–7.

[72] Yumeng Qin et al. "Spotting Rumors via Novelty Detection". In: *arXiv preprint arXiv:1611.06322* (2016).

[73] Zhifan Yang et al. "Emerging rumor identification for social media with hot topic detection". In: *Web Information System and Application Conference (WISA), 2015 12th*. IEEE. 2015, pp. 53–58.

[74] Arkaitz Zubiaga et al. “Crowdsourcing the annotation of rumourous conversations in social media”. In: *Proceedings of the 24th International Conference on World Wide Web*. ACM. 2015, pp. 347–353.

[75] Richard McCreadie, Craig Macdonald, and Iadh Ounis. “Crowdsourced rumour identification during emergencies”. In: *Proceedings of the 24th International Conference on World Wide Web*. ACM. 2015, pp. 965–970.

[76] Jian Zhang. “A survey on streaming algorithms for massive graphs”. In: *Managing and Mining Graph Data*. Springer, 2010, pp. 393–420.

[77] Zicong Zhou et al. “Information resonance on Twitter: watching Iran”. In: *Proceedings of the first workshop on social media analytics*. ACM. 2010, pp. 123–131.

[78] Pengyi Fan et al. “Measurement and analysis of topology and information propagation on sina-microblog”. In: *Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on*. IEEE. 2011, pp. 396–401.

[79] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. “Twitter Under Crisis: Can we trust what we RT?” In: *Proceedings of the first workshop on social media analytics*. ACM. 2010, pp. 71–79.

[80] Brian Babcock et al. “Models and issues in data stream systems”. In: *Procs of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM. 2002, pp. 1–16.

[81] Bo Pang and Lillian Lee. “Opinion mining and sentiment analysis”. In: *Foundations and trends in information retrieval* 2.1-2 (2008), pp. 1–135.

[82] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python*. " O’Reilly Media, Inc.", 2009.

[83] Baichuan Li et al. “Question identification on twitter”. In: *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM. 2011, pp. 2477–2480.

[84] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. “Latent Dirichlet Allocation”. In: *J. Mach. Learn. Res.* 3 (Mar. 2003), pp. 993–1022.

[85] Christian Borgelt. “Simple algorithms for frequent item set mining”. In: *Advances in machine learning II*. Springer, 2010, pp. 351–369.

[86] Takaaki Ohnishi, Hideki Takayasu, and Misako Takayasu. “Network Motifs in an Inter-firm Network”. In: *Journal of Economic Interaction and Coordination* 5.2 (2010), pp. 171–180.

[87] Yuichi Ikeda et al. “Agent Simulation of Chain Bankruptcy”. In: *arXiv preprint arXiv:0709.4355* (2007).

[88] Dana Varinsky. *Nearly half of US coal is produced by companies that have declared bankruptcy and Trump wont fix that*. http://www.businessinsider.com/us-coal-bankruptcy-trump-2016-12. 2016. (Visited on 07/30/2017).

[89] Yoshi Fujiwara. “Chain of Firm’s Bankruptcy: A Macroscopic Study of Link Effect in a Production Network”. In: *Advances in Complex Systems* 11.05 (2008), pp. 703–717.

[90] Stefano Battiston et al. “Credit Chains and Bankruptcy Propagation in Production Networks”. In: *Journal of Economic Dynamics and Control* 31.6 (2007), pp. 2061–2084.

[91] Garry Robins and Malcolm Alexander. “Small worlds among interlocking directors: Network structure and distance in bipartite graphs”. In: *Computational & Mathematical Organization Theory* 10.1 (2004), pp. 69–94.

[92] Amar Bhide. *The Origin and Evolution of New Businesses*. Oxford University Press on Demand, 2000.

[93] Gautam Ahuja. “The Duality of Collaboration: Inducements and Opportunities in the Formation of Interfirm Linkages”. In: *Strategic Management Journal* 21.3 (2000), pp. 317–343.

[94] Ranjay Gulati. “Network Location and Learning: The Influence of Network Resources and Firm Capabilities on Alliance Formation”. In: *Strategic Management Journal* 20.5 (1999), pp. 397–420.

[95] Julie M Hite and William S Hesterly. “The Evolution of Firm Networks: From Emergence to Early Growth of the Firm”. In: *Strategic Management Journal* 22.3 (2001), pp. 275–286.

[96] Georg Simmel. *The sociology of georg simmel*. Vol. 92892. Simon and Schuster, 1950.

[97] Ron Milo et al. “Network motifs: simple building blocks of complex networks”. In: *Science* 298.5594 (2002), pp. 824–827.

[98] Mark EJ Newman and Juyong Park. “Why social networks are different from other types of networks”. In: *Physical Review E* 68.3 (2003), p. 036122.

[99] Thomas Y Choi and Zhaohui Wu. “Triads in supply networks: theorizing buyer–supplier–supplier relationships”. In: *Journal of Supply Chain Management* 45.1 (2009), pp. 8–25.

[100] Corey C Phelps. “A longitudinal study of the influence of alliance network structure and composition on firm exploratory innovation”. In: *Academy of management journal* 53.4 (2010), pp. 890–913.

[101] Patrick M Kreiser. “Entrepreneurial orientation and organizational learning: The impact of network range and network closure”. In: *Entrepreneurship Theory and Practice* 35.5 (2011), pp. 1025–1050.

[102] Michael Howard, Emily Cox Pahnke, Warren Boeker, et al. “Understanding network formation in strategy research: Exponential random graph models”. In: *Strategic management journal* 37.1 (2016), pp. 22–44.

[103] Volker Grimm et al. “The ODD Protocol: A Review and First Update”. In: *Ecological Modelling* 221.23 (2010), pp. 2760–2768.

[104] Henrich R Greve. “Performance, Aspirations, and Risky Organizational Change”. In: *Administrative Science Quarterly* (1998), pp. 58–86.

[105] Patti Cybinski. “Description, Explanation, Prediction-the Evolution of Bankruptcy Studies?” In: *Managerial Finance* 27.4 (2001), pp. 29–44.

[106] Rahul C Basole and Marcus A Bellamy. “Supply network structure, visibility, and risk diffusion: A computational approach”. In: *Decision Sciences* 45.4 (2014), pp. 753–789.

[107] George P Huber. “Organizational Learning: The Contributing Processes and the Literatures”. In: *Organization Science* 2.1 (1991), pp. 88–115.

[108] Lola L Lopes. “Between Hope and Fear: The Psychology of Risk”. In: *Advances in Experimental Social Psychology* 20.3 (1987), pp. 255–295.

[109] Richard Michael Cyert and James G March. “A Behavioral Theory of the Firm”. In: *Englewood Cliffs, NJ* 2 (1963).

[110] Jeffrey Pfeffer and Gerald R Salancik. *The External Control of Organizations: A Resource Dependence Perspective*. Stanford University Press, 2003.

[111] Joseph T Mahoney and J Rajendran Pandian. “The Resource-based View within the Conversation of Strategic Management”. In: *Strategic Management Journal* 13.5 (1992), pp. 363–380.

[112] Edward J Malecki. “Technology and Economic Development: The Dynamics of Local, Regional, and National Change”. In: *University of Illinois at Urbana-Champaign’s Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship* (1997).

[113] Lori Rosenkopf and Paul Almeida. “Overcoming Local Search Through Alliances and Mobility”. In: *Management Science* 49.6 (2003), pp. 751–766.

[114] Stuart A Kauffman and Edward D Weinberger. “The NK Model of Rugged Fitness Landscapes and its Application to Maturation of the Immune Response”. In: *Journal of Theoretical Biology* 141.2 (1989), pp. 211–245.

[115] Yves L Doz. “The Evolution of Cooperation in Strategic Alliances: Initial Conditions or Learning Processes?” In: *Strategic Management Journal* 17.S1 (1996), pp. 55–83.

[116] Milton Harris and Artur Raviv. “Corporate Governance: Voting Rights and Majority Rules”. In: *Journal of Financial Economics* 20 (1988), pp. 203–235.

[117] Hayato Goto, Hideki Takayasu, and Misako Takayasu. “Empirical Analysis of Firm-Dynamics on Japanese Interfirm Trade Network”. In: *Proceedings of the International Conference on Social Modeling and Simulation, plus Econophysics Colloquium 2014*. Springer. 2015, pp. 195–204.

[118] Réka Albert and Albert-László Barabási. “Statistical Mechanics of Complex Networks”. In: *Reviews of Modern Physics* 74.1 (2002), p. 47.

[119] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. “Benchmark Graphs for Testing Community Detection Algorithms”. In: *Physical Review E* 78.4 (2008), p. 046110.

[120] Mark EJ Newman. “The structure and function of complex networks”. In: *SIAM review* 45.2 (2003), pp. 167–256.

[121] Thomas Schank and Dorothea Wagner. *Approximating clustering-coefficient and transitivity*. Universität Karlsruhe, Fakultät für Informatik, 2004.

[122] Chiara Orsini et al. “Quantifying randomness in real networks”. In: *Nature communications* 6 (2015), p. 8627.

[123] Pol Colomer-de Simón and Marián Boguñá. “Double percolation phase transition in clustered complex networks”. In: *Physical Review X* 4.4 (2014), p. 041020.

[124] Byoung Hee Hong, Kyoung Eun Lee, and Jae Woo Lee. “Power Law in Firms Bankruptcy”. In: *Physics Letters A* 361.1 (2007), pp. 6–8.

[125] Domenico Delli Gatti et al. "Business Fluctuations and Bankruptcy Avalanches in an Evolving Network Economy". In: *Journal of Economic Interaction and Coordination* 4.2 (2009), pp. 195–212.

[126] David Lazer et al. "Computational Social Science". In: *Science* 323.5915 (2009), pp. 721–723.

[127] Nicola Lettieri. "Computational Social Science, the Evolution of Policy Design and Rule Making in Smart Societies". In: *Future Internet* 8.2 (2016), p. 19.

[128] Claudio Cioffi-Revilla. "Introduction to computational social science". In: *Berlin/New York: Springer* 10 (2014), pp. 978–1.

[129] Ray M Chang, Robert J Kauffman, and YoungOk Kwon. "Understanding the paradigm shift to computational social science in the presence of big data". In: *Decision Support Systems* 63 (2014), pp. 67–80.

[130] Rosaria Conte et al. "Manifesto of computational social science". In: *European Physical Journal-Special Topics* 214 (2012), p–325.

[131] Dirk Helbing, Anders Johansson, and Habib Zein Al-Abideen. "Dynamics of crowd disasters: An empirical study". In: *Physical review E* 75.4 (2007), p. 046109.