COMPREHENSIVE MODELLING OF INFLUENCE DIFFUSION IN COMPLEX SOCIAL NETWORKS, AN AGENT-BASED PERSPECTIVE

A THESIS SUBMITTED TO AUCKLAND UNIVERSITY OF TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Supervisors
Dr Quan Bai
Prof. Minjie Zhang

June 2018

By
Weihua Li
School of Engineering, Computer and Mathematical Sciences

Copyright

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

Signature of candidate

Acknowledgements

The process of pursuing the PhD is really very challenging. The completion of my PhD thesis turns out to be the outcome of a long journal of learning, exploring and researching new knowledge.

It is my great pleasure to acknowledge some of many people who have inspired, supported and encouraged me all along the way. First and foremost, I would like to express my sincere gratitude to my primary supervisor Dr Quan Bai who gave the opportunity to start my academic career. In particular, Quan provides me with continuous support, guidance and encouragement. I greatly appreciate his patience, enthusiasm and immense knowledge, which make him an excellent supervisor. I am also extremely grateful to my secondary supervisor, Professor Minjie Zhang, for her valuable advice, encouragement and sincere guidance. With the assistance of both of them, I obtained numerous opportunities to develop my research skills.

Second, I would like to give my great appreciation to my colleagues, both current and past, within the school of Engineering, Computer and Mathematical Sciences at the Auckland University of Technology, who have given me tremendous help during my PhD study, Doan Tung Nguyen, Ling Liang, Chang Jiang, Barry Dowdeswell, Da Zhang, Janet Tong, Jing Wang, Yingying Tao, Wenwang Pang, Saide Lo and Terry Brydon.

Finally, I would like to extend my love and gratitude to my family members for their understanding and support during my PhD research study.

Abstract

Influence diffusion modelling, analysis and applications draw tremendous attention to both researchers and practitioners since many organisations attempt to utilise its power to achieve business or political goals. A great many research works have been dedicated to the exploration of maximizing the spread of a particular influence in complex networks, e.g., social networks. However, influence appears to be a hybrid and complex effect caused by numerous factors, such as friendship affiliation, preferences, common communities, etc. Moreover, due to the sophisticated and dynamic environment where influences reside in, modelling influence diffusion in complex networks becomes a very challenging topic.

In this thesis, agent-based approaches and multi-agent systems have been employed to model the influence propagation in complex systems. In other words, the perspective of exploring the spread of influence diffusion stands at a microscopic level, where the dissemination of influences is driven by individuals' personalised traits and behaviours.

First, the thesis elaborates the hybrid effects of influence and systematically presents a generic architecture of modelling influences from an agent-based perspective. Second, based on the proposed framework, we further investigate agent-based approach with stigmergic interactions, to address the influence maximisation problem in a dynamic and complex environment. Third, driven by the business needs for long-term marketing, the generic agent-based model has been extended by incorporating the capabilities for maintaining long-lasting influences. Last but not the least, by considering the

coexistence of multiple influences, the agent-based model has been enhanced to handle the various relations of influences.

Publications

Li, W., Bai, Q., Zhang, M. & Nguyen T.D. Automated Influence Maintenance in Social Networks: an Agent-based Approach. In the IEEE Transactions on Knowledge and Data Engineering, Accepted in August 2018, DOI 10.1109/TKDE.2018.2867774

Li, W., Bai, Q. & Zhang, M. A Multi-Agent System for Modelling Preference-based Complex Influence Diffusion in Social Networks. In The Computer Journal, Accepted in July 2018, DOI 10.1093/comjnl/bxy078

Li, W., Bai, Q. & Zhang, M. SIMiner: A Stigmergy-based Model for Mining Influential Nodes in Dynamic Social Networks. In the IEEE Transactions on Big Data, Accepted in March 2018, DOI 10.1109/TBDATA.2018.2824826

Li, W., Bai, Q., Zhang, M. & Nguyen T.D. Modelling Multiple Influences Diffusion in On-line Social Networks. In Proceedings of International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), 2018

Li, W., Bai, Q., Nguyen T.D. & Zhang, M. Agent-based Influence Maintenance in Social Networks. In Proceedings of International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), 2017, Pages 1592-1594

Li, W., Bai, Q., Jiang, C. & Zhang, M. Stigmergy-based Influence Maximization in Social Networks. In Proceedings of Pacific Rim International Conferences on Artificial Intelligence (PRICAI), 2016, Pages 750-762

Li, W., Bai, Q. & Zhang, M. Comprehensive Influence Propagation Modelling for Hybrid Social Network. In Proceedings of the 29th Australasian Joint Conference on Artificial Intelligence, 2016, Pages 597-608

Li, W., Bai, Q. & Zhang, M. Agent-based Influence Propagation in Social Networks. In Proceedings of IEEE International Conference on Agents, 2016, Pages 51-56

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A complex network can be considered as a graph containing a large collection of nodes connected by links representing various complex interactions among the nodes (Ganguly, Deutsch & Mukherjee, 2009). Complex networks arise in many diverse contexts, including viral networks, computer networks, protein-protein interactions, social networks, etc. The modelling and analysis of these complex systems requires a broad effort spanning many disciplines (Bonato & Tian, 2012). A complex on-line social network, i.e., a sub-class of complex network, is an on-line platform for people to build social relations with others who share similar preferences, career activities or real-life connections (Boyd & Ellison, 2007). Moreover, a complex on-line social network is large-scale, having dynamic and evolving topological structures. The thesis is developed mainly based on this environment.

Within the context of complex on-line social networks, influence diffusion modelling has been extensively studied and widely applied in many research fields, such as maximization of product adoption (Kempe, Kleinberg & Tardos, 2003), culture dissemination analysis (Axelrod, 1997). These studies bring benefit to various domains, such as E-business, marketing and sociology. Particularly, in e-business and marketing field, one of the famous business strategies relying on the influence propagation is viral marketing (Kempe et al., 2003), where the spread of influence relies on one of the social phenomena, i.e., social influence, indicating that one's opinions or behaviours are affected by his or her contactable neighbours in the social network (Turner, 1991; Raven, 1964). With the adoption of such strategies, the influence diffuses among the individual's social circle through 'word-of-mouth' effect; this has been widely employed by advertising agencies to increase the brand awareness for their clients (Ahmed & Ezeife, 2013; East, Hammond & Lomax, 2008; W. Chen et al., 2011).

It is very challenging to model and analyse the influence diffusion process in complex networks. First, influence appears to be a hybrid effect, which can be decomposed into multiple components focusing on different activities and personalised traits of human beings (Anthony, 2009). Some common assumptions, e.g., "friendship-affiliation links represent influence-propagation channels" and "the strength of links is considered as the only factor affecting the influence propagation probability", cannot hold in general. Second, local information is merely available in most real situations. Diffusion models relying on a global view definitely lose the advantages. Third, networks appear to be highly dynamic. The nodes join and quit, while the links are forming and vanishing over time.

By considering these sophisticated features of influence diffusion, in this thesis, I present an agent-based stack to model the spread of influence in complex networks to address these challenging issues. Based on the proposed generic models, I further investigate the in-depth applications and the extended problems that can be addressed by utilising the models.

1.1 Influence Diffusion

This section aims to introduce the characteristics of state-of-the-art influence diffusion models and one of the classic applications, i.e., influence maximization.

1.1.1 Influence Diffusion Models

Most researchers have conducted social influence analysis and modelling based on two fundamental influence diffusion models, i.e., the Independent Cascade (IC) model and the Linear Threshold (LT) model (Kempe et al., 2003; W. Chen, Wang & Yang, 2009; W. Chen, Yuan & Zhang, 2010). In both models, each node has two states: the active state and the inactive state. At the beginning, a limited set of users, i.e., seed nodes, are supposed to be selected as the initial active nodes, which attempt to propagate influence and affect the inactive neighbours at a certain chance. If any neighbour is activated, the state of the node becomes active, and it starts to propagate influence to the neighbours. Moreover, both models have two key properties, i.e., propagation and attenuation. The influence initiates from the seed set, i.e., the selected activated nodes for propagating influence. These nodes transfer their influence through the correlation graph, whereas the power of the effect decreases when hopping further and further away from the activated nodes. The IC model and the LT model represent different deterministic strategies of influence though they share some common features.

There are also some differences between these two models. The IC model is a kind of non-deterministic diffusion model, where the receiver's state is not deterministically decided by itself but affected and influenced with a predefined probability by the senders (Y. Jiang & Jiang, 2015). In the IC model, when an active node v_i interacts with the adjacent inactive nodes $\mathit{\Gamma}(v_i)$, it has a single chance to activate each neighbour at a successful rate, whereas, the features of nodes are not taken into consideration (Y. Wang, Cong, Song & Xie, 2010). In contrast, the LT model is a deterministic diffusion model (Y. Jiang & Jiang, 2015). In this model, each node is assigned with a fixed threshold, where the threshold of node v_i can be represented as ϑ_i. Node v_i is influenced by the neighbours $\mathit{\Gamma}(v_i)$ when the sum of their weights exceeds ϑ_i. The threshold controls the opinion or state adoption for each node. Specifically, nodes with a high threshold have low probabilities to be influenced, while for those low-threshold nodes, they have a high tendency of becoming active.

Influence maximization is one of the typical applications based on these two fundamental influence diffusion models,

1.1.2 Influence Maximization

In on-line marketing, it is very important to investigate how to disseminate positive influence in on-line social networks with limited resources, so that the entire network evolves towards an expected beneficial direction (Domingos & Richardson, 2001). Motivated by this background, Kempe et al. (2003) formulate the influence maximization as a discrete optimization problem, which aims to select a finite set of influential users from the network, expecting that they can effectively spread the positive influence across the entire network. The selection process is name as *seed selection*, and the selected users are call *seed set*.

Influence maximization is a type of optimisation problem and is NP-hard. Many real-world applications are far more complex than influence maximization since sophisticated features, e.g., individual's personalised traits and behaviours, time-space dynamics, undiscovered global view, etc., are enabled.

Real-world influence diffusion applications are not just restricted to the influence maximization. In this thesis, I also investigate how to maintain a long-term influence, and how to effectively suppress an undesirable influence when multiple influences emerge.

1.2 Sophisticated Influence

The sophistication of modelling and analysing influence diffusion in complex networks is mainly reflected in two aspects: the complex nature of networks and characteristics of influence.

Many real-world systems in nature are complex networks, with a large collection of nodes connected by links representing various complex interactions among the nodes (Ganguly et al., 2009). These systems inherit three significant features, i.e., large scale, dynamics and complex structure (Newman, 2003; Boccaletti, Latora, Moreno, Chavez & Hwang, 2006; Ganguly et al., 2009). The paths of propagation vary all the time due to the dynamics of network topological structure. It is very complicated to track the influence-diffusion path in an evolving large-scale network. Furthermore, the influence can be triggered by various types of interactions among the nodes in the complex networks.

Apart from the complex context, the sophistication also stems from the characteristics of influence, summarised as follows:

- **Influence is homophily-driven.** Homogeneous influence suggests similarity lead users to interact, and interactions lead users to behave more similarly (McPherson, Smith-Lovin & Cook, 2001; Z. Li & Tang, 2012). Specifically, the higher common preference degree the users have, the greater chance they adopt the behaviours of their peers. The process of homophily and influence is capable of producing network convergence (Centola, Gonzalez-Avella, Eguiluz & San Miguel, 2007; Galam & Moscovici, 1991; DiMaggio, Evans & Bryson, 1996).

- **Individuals have different levels of influence acceptance.** Users in a social network are cognitive, and they possess different levels of prior commitment towards the same item, which is recognised as a critical factor associated with the preference, affecting individual's acceptance of the influence (Ahluwalia, Burnkrant & Unnava, 2000). More specifically, every user has an initial impression of either unknown or familiar item intrinsically, which is potentially more dominant than social influences.

- **People tend to be influenced by the neighbours in their social context.** Users in the same network are generally affected by others' thoughts, feelings and behaviours (Z. Li & Tang, 2012), and this conformity influence is also named as confluence (Tang, Wu & Sun, 2013). Users normally have different levels of individual conformity. In other words, some people are soft-headed and easy to be influenced by others. Whereas some people are stubborn and their actions do not necessarily rely on others' opinions.

- **Influence is bi-directional and weighted.** User's preference or prior knowledge are normally considered when ranking users by their influential capacities (Zhang, Wang, Han, Yang & Wang, 2015; D. Li et al., 2014). Two influencers normally exert influences of different degrees on each other, e.g., a computer expert could exert higher influence on the amateurs in selecting laptops than the others do, but the amateurs' opinions may not affect the computer experts a lot.

- **The users' state appear dynamic.** Driven by influences, users' attitude towards any innovation can be possibly revised. For example, Emily bought a smartphone and found a tiny piece of dust in the back-end camera. She shared the experience with her friend Leo. Leo was influenced and started to spread this incidence to his friends. However, Leo noticed a lot of positive news regarding this smartphone posted on the wall of his social network, so he bought one as well. Hence, individual's attitude is dynamic, which can be revised.

- **Influence is of two-sidedness.** Both positive and negative opinions towards the same entity normally coexist in the same environment, and people are getting influenced by both kind of messages (Fiske, 1980; Rozin & Royzman, 2001). For instance, 'word-of-mouth' effect may be positive to encourage the choice of a specific product, but it also can be negative to dampen the adoption (East et al., 2008).

Therefore, by considering both the real-world environment and complex nature of influence, it is incredibly challenging to model and analyse the real-world influence diffusion.

1.3 Research Motivations

From a broad point of view, there are many research motivations of exploring the modelling of influence propagation in complex systems. Influence diffusion modelling assists the sociological researchers with cultural groups' evolution and culture dissemination analysis (Centola et al., 2007; Axelrod, 1997). Furthermore, it also helps to investigate and predict the obnoxious things propagating through the networks (Kimura, Saito & Motoda, 2008). For instance, computer viruses are capable of spreading across the computer networks (Serazzi & Zanero, 2004), epidemics and infectious diseases diffuse via interacting individuals (Webb, 1981), malicious rumours spread in social networks among the individuals (Doerr, Fouz & Friedrich, 2012). In the context of the social network, influence-diffusion modelling helps to capture the patterns and trend of the network evolution, so that the appropriate business strategies can be adopted to reduce the negative impact and increase the positive influence.

In the contemporary research field, there is much literature describing the modelling of influence diffusion and its applications.The primary research gaps are identified as follows:

- Very few research works have been dedicated to the influence modelling in complex networks. The influence diffusion is oversimplified as a hopping and infecting process, but the individual's personalised characters and behaviours affecting the diffusion of influence are neglected.

- The assumed environment of the spread of influence is different from the complex networks in the real-world, which are normally large-scale, dynamic and with no global view. For example, some literature imitates the dynamics by capturing snapshots or storing all changes happening around, which seems unrealistic since this inevitably creates another set of big data. Furthermore, capturing snapshots or any alternations require a global view, which becomes another obstacle as a central component is required for monitoring the entire network in real-time.

- Most applications of influence diffusion modelling concentrate on the influence maximization problem. However, few research works are conducted to investigate long-term influence maintenance and situations when multiple influences emerge.

By considering the aforementioned challenging issues and the research gaps, the research motivations of this research are mainly explained in two folds from the following perspectives:

The Modelling Perspective:

- Model the influence diffusion in complex systems through defining the individuals' personalised characters and behaviours. This also enables the model to function without a global view.
- Model the influence diffusion in dynamic and large-scale networks.
- Model multiple influences and the corresponding relations.

The Business/Application Perspective:

Based on the proposed models, the following motivations are expanded:

- Achieve influence maximization via 'word-of-mouth' effect in large-scale and dynamic networks.
- Achieve long-term marketing by maintaining a particular influence.

- Suppress an undesirable influence effectively without modifying the topological structure of the networks.

1.4 Research Questions

According to the research motivations mentioned in the previous section, the objectives of this thesis can be reflected in the following three research questions:

Research Question 1: How to systematically model the influence diffusion for large-scale and dynamic networks even without discovering a global view?

- **Sub-Research Question 1.1:** How to generally model the individual's behaviours associated with influence diffusion?
- **Sub-Research Question 1.2:** How to extend the proposed generic model to handle the influence maximization in large-scale and dynamic social networks?

Research Question 2: How to maintain an influence message effectively in social networks?

- **Sub-Research Question 2.1:** How to develop a model for influence maintenance?
- **Sub-Research Question 2.2:** How to maintain an influence by utilising the proposed model?

Research Question 3: How to model multiple influences diffusion in social networks?

- **Sub-Research Question 3.1:** How to model multiple influences diffusion and their relationships?
- **Sub-Research Question 3.2:** How to utilise the proposed model to suppress the spread of an undesirable influence?

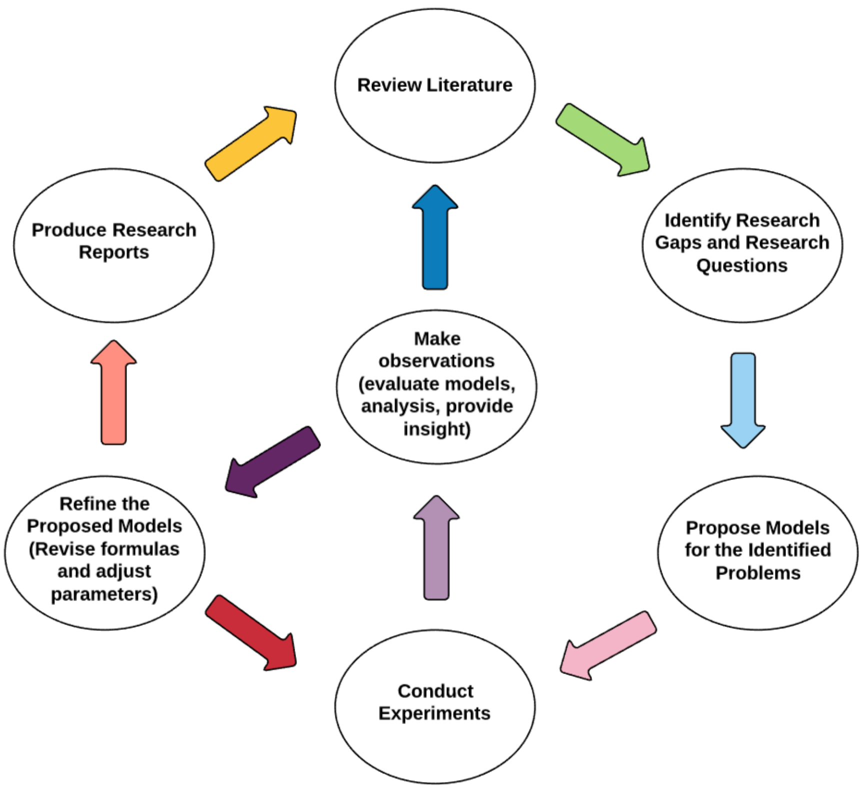


Figure 1.1: Research Methodology Leveraged in this Thesis

1.5 Design of Study

To answer the research questions, the research methodology utilised in this study is explained. Next, the proposed approach, i.e., agent-based modelling, leveraged in this research is introduced. Then, the evaluation methods of the proposed approach are explained.

1.5.1 Research Methodology

In my PhD study, the research presents an interactive process, including seven major steps. The research methodology leveraged in this thesis are demonstrated in Figure 1.1, adapted from the *Scientific Method as an Ongoing Process* (Garland, 2015).

The first step is reviewing literature of the relevant research area. Next, research gaps are supposed to be identified, including the issues that have been ignored by other researchers and performances to be improved. Subsequently, I raise the research questions based on the observed research gaps. After that, the models are proposed and developed to tackles the research issues. The next step is collecting valid public dataset on-line and conducting experiments by using the model. Then, make observation on the experimental results, explore the insights and evaluate the performance. If the outcome is satisfying, the model is finalised. Otherwise, the model is considered to be refined or parameters need to be adjusted, and then conduct experiments again. The final step is writing research reports and articles.

1.5.2 Agent-based Modelling

By researching and investigating complex problems in open and dynamic environments, it is universally acknowledged by the research community that the agent-based approach would be preferable and feasible in this scenario (Macal & North, 2009; Nguyen, 2017). Therefore, in this project, I attempt to leverage the advantages offered by the agent-based approach to model the influence diffusion in a complex environment.

Agent-Based Modelling (ABM), also named as individual-based modelling, has demonstrated many advantages in modelling complex systems (Bonabeau, 2002; Macal & North, 2009; van Maanen & van der Vecht, 2013; Holland, 1995). ABM is a specific individual-based computational model, where individuals are modelled the interactive autonomous agents. Compared with traditional centralised models, ABM is an appropriate approach to explore the macro world through defining the micro level of a social system (Bonabeau, 2002; Macal & North, 2009).

A simple and generic ABM is Cellular Automata (CA) (Shiffman, Fry & Marsh, 2012), where each agent keeps evolving by looking at the neighbours' states. Sophisticated ABM sometimes leverages artificial intelligence approaches, such as neural networks, evolutionary algorithms, etc., which enable the agents to learn and adapt their behaviours (Bonabeau, 2002; Chantemargue, Krone, Schumacher, Dagaeff & Hirsbrunner, 1998).

By utilising ABM to model the influence diffusion process in complex networks, nodes are modelled as proactive agents interacting with the neighbours by exchanging the influence messages; the network represents the global environment shared by all the agents; influence propagation process demonstrates a decentralised evolutionary pattern driven by the agents' actions. The entire network evolves through a number of discrete time steps according to a set of transitional or behavioural rules of each agent.

With ABM applied to influence diffusion modelling, I can concentrate on modelling individual's personalised traits and behaviours by neglecting the complexities coming from the global views.

1.5.3 Evaluation Methods

ABM is proposed to simulate the environments and influence-diffusion phenomenon. There are a few approaches to regulate the agents' decisions, including the heuristics, rules derived from social theories and parametrised utility functions adopted from game theory.

It is hard to evaluate an ABM itself based on traditional metrics, such as accuracy. This is because ABM produces and forecasts the trend of a certain phenomenon, rather than predicting particular values. However, the validation of the model can be obtained when it is applied to a particular problem, such as influence maximization. Two traditional evaluation metrics in the field of influence maximization are also considered.

- *Influence Activation Coverage (Effectiveness)*: the estimated number of active users, i.e., users who have been influenced, after the influence-propagation process.

- *Time to carry out the solution (Efficiency):* the time required to complete the seed-selection process.

The evaluation metrics in this thesis are not just restricted to these two traditional evaluation metrics. New evaluation metrics are defined according to the needs from the extended models and problems. The details are given in the experiment section of each chapter.

1.6 Contributions of the Thesis

In this thesis, I have proposed four major influence-diffusion models, focusing on different aspects of the propagation process. Based on these four models, the contributions of this research work are summarised as follows:

(1) I propose a generic agent-based framework to model the influence diffusion, i.e., Influence Diffusion Multi-Agent System (IDMAS), which serves as a basis for addressing the related issues in complex networks. The IDMAS does not require a global view and is capable of tracking a long-term evolutionary trend of social networks driven by influence, and handling the situations when group opinions revise according to the changing context. The models and related results are published in (W. Li, Bai & Zhang, 2016a).

(2) I propose a decentralised approach, called Stigmergy-based Influencers Miner (SIMiner), to address the influence maximization problem in a large-scale and dynamic environment efficiently and effectively. SIMiner is applicable in both static and dynamic complex environments and capable of adapting the solutions in an evolving context. The model and results are published in (W. Li, Bai, Jiang & Zhang, 2016) and in the journal of IEEE Transactions on Big Data (W. Li, Bai & Zhang, 2018b).

(3) I study long-term influence maintenance problem by extending the classic influence maximization problem. The Agent-based Timeliness Influence Diffusion

(ATID) model is proposed to accommodate the influence maintenance problem. ATID is capable of capturing two primary elements for maintaining long-lasting influence, i.e., the temporal feature of a social network and the status of particular influence. The model and results are published in (W. Li, Bai, Nguyen & Zhang, 2017).

(4) I model multiples influence diffusion, explore the pheromone when various influences coexist, and study the influence minimization problem based on a proposed model, called Agent-based Multiple Influences Diffusion (AMID) model. The related results have been published in (W. Li, Bai, Zhang & Nguyen, 2018b).

1.7 Thesis Structures

The remainder of the thesis is constructed as follows:

- **Chapter 2** reviews several state-of-the-art studies of influence diffusion and influence maximization in social networks, as well as agent-based modelling.
- **Chapter 3** introduces a hybrid model and a generic agent-based framework for modelling influence propagation, which serves as a basis for the following chapters to explore in-depth issues regarding influence diffusion.
- **Chapter 4** presents the Stigmergy-based Influencers Miner (SIMiner) model to tackle the influence maximization problem in large-scale and dynamic social networks, which aims to answer Research Question 1.
- **Chapter 5** presents an influence maintenance model and investigates how to maintain a particular influence, which tackles Research Question 2.
- **Chapter 6** models multiple influences diffusion and explores negative influence minimization by leveraging the proposed model. This chapter focuses on Research Question 3.

- **Chapter 7** concludes the thesis by summarising the advantages and limitations of the proposed approaches, as well as the outlines for the future work.

Chapter 2

Literature Review

The literature review of the thesis covers three major aspects, i.e., the influence diffusion modelling, influence maximization and agent-based modelling.

In the PhD study, I reviewed and compared different kinds of influence-diffusion models. Very few studies merely investigate the modelling of influence diffusion without applying the proposed model to particular problems. This is because the values are greatly reflected in the practical applications based on the diffusion models. Therefore, in this chapter, beside a review of influence-diffusion modelling, I also further explore the contemporary research works associated with the applications of the influence-diffusion modelling. Subsequently, I survey the influence maximization problem as one of the typical applications of the influence diffusion modelling, and then expand the scope to the extended diffusion models and problems. Next, studies of multi-agent systems and agent-based modelling in the field of influence diffusion have been reviewed. In the end, I summarise the findings and briefly explain the research gaps.

2.1 Influence Diffusion Modelling

There are quite a few existing influence-diffusion models, which are not strictly classified. In general, these models can be categorised as explanatory models and predictive models (Guille, Hacid, Favre & Zighed, 2013; M. Li, Wang, Gao & Zhang, 2017). Epidemics models are similar to the predictive models but sometimes are categorised to the explanatory models (M. Li et al., 2017).

2.1.1 Explanatory Models

The explanatory models are capable of retracing the influence-diffusion paths (Guille et al., 2013). Gomez et al. (2010) attempt to infer the structure of the influence spreading cascade by proposing NETINF. The model has been further extended as NETRATE, considering the pairwise transmission rates (Gomez-Rodriguez, Balduzzi, Schölkopf, Scheffer et al., 2011). As these two models cannot handle the dynamics of the social network, INFOPATH is proposed, as well as a time-varying inference algorithm (Gomez Rodriguez, Leskovec & Schölkopf, 2013).

The Independent Cascade Model with Pyramid Scheme (ICMPS) proposed in Section 4 is also a type of explanatory model, which tracks the contributions of activation of each user.

2.1.2 Epidemic Spread Models

The propagation process of both influence and epidemics share similar patterns. Epidemic spread models describe the transmission of communicable disease through individuals. In other words, the virus spreads from the infected individuals to the susceptible users via physical interactions. Similarly, the innovations also disseminate from senders to the reachable contacts in an analogous fashion. Some influence-diffusion models even stem from the epidemic models. Therefore, to model and analyse influence diffusion in complex networks, it is essential to review the diffusion models of virus contagions.

Fundamental Epidemic Models

The epidemic spread models rely on the physical networks, where each node presents an individual and each link denotes a physical connection between two nodes. There are four fundamental epidemic spread models, i.e., Susceptible-Infected (SI) model (Bailey et al., 1975; Pastor-Satorras & Vespignani, 2001), Susceptible-Infected-Susceptible (SIS) model (Pastor-Satorras & Vespignani, 2001), Susceptible-Infected-Removed (SIR) model (Xiong, Liu, Zhang, Zhu & Zhang, 2012) and Susceptible-Infected-Removed-Susceptible (SIRS) model (Jin, Wang & Xiao, 2007).

The differences among these four models are reflected in the rules of contagion and states. In both SI model and SIS model, only two discrete states exist in the context, “susceptible" (able to be infected) and “infected". Both models focus on the evolution of the proportions of nodes in each state. Different from SI model, SIS model allows the infected users to be cured after having been infected. As for SIR and SIRS model, another state, i.e., “recovered" (no longer able to infect or be infected), is added. SIR model also extends SI model, insisting that an individual can be immune from the contagion after getting cured. Specifically, at each time step, only the nodes infected in the last time step is able to infect the neighbours with a susceptible state at a certain probability. At the next step, the previously infected node switches to “recovered" and becomes immune from being infected. To extend the epidemic model further, SIRS believes that a cured user also can revert back to a susceptible individual with a certain probability.

Epidemic Models in Social Networks

The epidemic models have been extended and utilised in the social network, where the extensions mainly consider the contextual features of social networks, such as network topological structures, innovation content, social factors, etc. These models focus on either simulating the propagation process by defining different states and infecting rules or exploring the factors affecting the diffusion.

Some representative influence-diffusion models derived from the epidemic models include the Susceptible-Exposed-Infected-Removed (SEIR) model (M. Y. Li, Graef, Wang & Karsai, 1999), the infection recovery SIR (irSIR) model (Cannarella & Spechler, 2014), the Fractional SIR (FSIR) model (Feng et al., 2015) and the Emotion-based Spreader–Ignorant–Stifler (ESIS) model (Q. Wang, Lin, Jin, Cheng & Yang, 2015).

SEIR extends SIR by adding the “Expose" state (M. Y. Li et al., 1999). In the irSIR model, the infection recovery dynamics have been added to simulate the adoption and abandonment of user views (Cannarella & Spechler, 2014). FSIR model is developed based on the intuition that a user with more friends requires more repeated exposures to spread further the information (Feng et al., 2015). In this model, the infection probability is proportional to the number of infected neighbours. ESIS simulates the influence diffusion by considering the information weight with emotion, revealing that the information diffusion is associated with the propagation probability and transmission intensity (Q. Wang et al., 2015).

2.1.3 Predictive Models

Predictive models aim to estimate the future influence spread in the entire network according to certain factors or heuristics. In this subsection, I introduce three fundamental models, including Independent Cascade (IC) model, Linear Threshold (LT) model and General Threshold (GT) model. In these models, each node has two possible states, i.e., active and inactive. The set of initially active users are named as seed set, and the elements inside the seed set are called influencers or seeds.

The GT model is a broader framework of the IC model and the LT model. The IC model and the LT model are two seminal models which have been widely applied and extended in modelling influence diffusion in the contemporary research field (Kempe et al., 2003; W. Chen et al., 2009, 2010, 2011; Shakarian, Bhatnagar, Aleali, Shaabani & Guo, 2015). Therefore, only classic models are introduced in this subsection, and review the extended versions for tackling influence maximization problems in the following subsections.

Independent Cascade Model

The IC model generalises the SIR model introduced previously. The influences are initiated from seed set S. The active user $v_i, v_i \in S$ attempts to interact and activate the corresponding inactive neighbour $v_j, v_j \in \Gamma(v_i)$ with Probability p_{ij}. If v_j is activated successfully, the state is supposed to be converted to active, and v_j starts to exert influence on the inactive neighbours $\Gamma(v_j)$ at certain probabilities.

Each node has a single chance to activate any of the neighbours. To explain this further, let node v_i be one of the active seeds at time step t_m, and v_j be one of the inactive neighbours of v_i, $v_j \in \Gamma(v_i)$. If v_i fails to activate v_j at t_{m+1}, v_j is not supposed to be activated by v_i directly at any other time step $t_{m+n+1}, n \in \mathbb{N}$.

Linear Threshold Model

In the LT model, each node $v_i, v_i \in V$ selects a random threshold θ_i, $\theta_i \in [0,1]$, where θ_i affects the influence acceptance of v_i. A low threshold indicates the user can be easily affected, while a high threshold means it is difficult to influence this user. Having the same diffusion process as that of the IC model, influences transfer from the active nodes to the inactive neighbourhood. However, the activation rules are different. Suppose that node v_i is influenced by neighbour $v_j, v_j \in \Gamma(v_i)$, at a weight b_{v_i,v_j}, where

$$\sum_{v_j \in \Gamma(v_i)} b_{v_i,v_j} \leq 1. \tag{2.1}$$

Node v_i becomes active when:

$$\sum_{v_j \in \Gamma(v_i)} b_{v_i,v_j} \geq \theta_i \tag{2.2}$$

Both the IC model and the LT model inherit two key properties, i.e., propagation and attenuation. The influences are initiated from the seed set, and they transfer their influence via the network. Whereas, the power of this effect decreases when hopping further and further away from the active nodes.

Both models present different deterministic strategies of influence. The IC model is a kind of non-deterministic diffusion model, where the receiver's state is not deterministically decided by itself but affected and influenced with a predefined probability by the senders (Y. Jiang & Jiang, 2015). By contrast, the LT model is a deterministic diffusion model, where a pre-defined threshold controls the opinion adoption for each node. In other words, nodes with high threshold have lower chances to be influenced, while for those low-threshold nodes, they possess a higher tendency of becoming active.

General Threshold Model

The General Threshold (GT) model is a generalised framework of which the LT model and the IC model are special cases (Kempe et al., 2003; Ahmed & Ezeife, 2013). Suppose that an inactive node v_i and the set of its active neighbours $\Gamma(v_i)$ exist in the network. To measure the probability that v_i will become active, the joint influence probability of $\Gamma(v_i)$ on node v_i, i.e., $p_{v_i}(\Gamma(v_i))$, is supposed to be calculated first. $p_{v_i}(\Gamma(v_i))$ can be formulated in Equation 2.3.

$$p_{v_i}(\mathit{\Gamma}(v_i)) = 1 - \Pi_{v_j \in \mathit{\Gamma}(v_i)}(1 - p_{v_j,v_i}), \tag{2.3}$$

where $p_{j,i}$ refers to the influence probability of v_j on v_i. Node v_i turns active if $p_{v_i}(\mathit{\Gamma}(v_i)) \geq \theta_i$, where θ_i refers to the threshold of v_i.

These seminal approaches model the influence propagation as a simplified centralised diffusion process (W. Chen et al., 2009; Kempe et al., 2003). In particular, a central component is responsible for controlling the diffusion process. However, such models are not applicable without a global.

2.2 Influence Maximization

In this subsection, the influence maximization, its extended problems and the corresponding revised diffusion models are reviewed. It is almost impossible to separate the influence maximization and the extended problems from the influence-diffusion modelling, since most newly proposed seeding algorithms are developed on the basis of the existing or extended diffusion models.

Influence maximization is first studied by Domingos and Richardson as a probabilistic problem (Domingos & Richardson, 2001). Kempe et al. (2003) formulate the influence maximization as a discrete optimisation problem, which aims to select a finite set of k influencers, i.e., $S, |S| = k$, from the network, expecting that they can effectively spread the positive influence, and maximize the impact $f(S)$ across the entire network. The initially selected users S can be called seed set, and the seeding process is named as seed selection. The influence of seed set is defined as the number of active nodes at the end of the propagation process. This is usually referred to as influence spread or influence coverage. To find a solution S for the influence maximization problem is NP-hard.

Several classic seed selection approaches presented in the literature are listed as follows. They are normally regarded as the counterparts of the proposed algorithms.

- **Greedy Selection:** attempts to reach the maximum influence marginal gain in selecting each seed, coming with a $1 - 1/e$ approximation guarantee, since $f(.)$ satisfies normalization ($f(\varnothing) = 0$), monotonicity ($S \subseteq S', f(S) \leq f(S')$) and submodularity (diminishing returns property, i.e., $S', S'' \subseteq S, S' \subseteq S'', f(S' \cup \{s\}) - f(S') \geq f(S'' \cup \{s\}) - f(S'')$) conditions (Nemhauser, Wolsey & Fisher, 1978; Kempe et al., 2003).

  The greedy algorithm facilitated in the influence maximization problem is simulation-based, which selects each seed by running numerous times of Monte-Carlo simulations. For example, in seed selection, the first seed can be identified by estimating each individual's influence through the simulations. Next, having the first influencer involved, the second seed can be identified by estimating the influence coverage of all the possible combinations, and so on and so forth. Eventually, the seed set will be selected.

- **Random Selection:** selects each seed randomly, so that the seed set grows incrementally. This approach does not follow any heuristics.

- **Degree-based Selection:** ranks the users based on out-degree, i.e., the size of reachable users.

- **Weighted Degree-based Selection:** ranks the users based on the cumulative influence-diffusion probabilities to the neighbours.

- **Degree Discount Heuristics:** extends the degree-based selection, and is developed based on the intuition that many of the most central nodes may be clustered. Therefore, it is not necessary to target all of them (W. Chen et al., 2009). Specifically, whenever a neighbour of an inactive node becomes active, the degree of the inactive node decreases by one. Next, the node with the highest node degree in the network is supposed to be selected into seed set.

Greedy selection is one of the strongest baselines in the influence maximization problem, and it appears to be more effective than most of the existing algorithms. However, the greedy selection is not scalable for large-scale networks. Many studies aim to develop scalable algorithms without decreasing the effectiveness significantly. Some other research works in this field investigate the influence maximization by considering particular features affecting the information dissemination, such as users' preferences, communities and emergence of negative opinions. Furthermore, to relax the assumptions of the traditional influence maximization problem, complex environments are taken into consideration, including the dynamics of networks and temporal features.

Based on the discussions above, in the following subsections, the studies related to the influence maximization have been reviewed.

2.2.1 Scalable Seeding Algorithms

As mentioned previously, the greedy algorithm in the influence maximization problem can carry out relatively optimal solutions, which are normally better than the existing ones. However, the greedy selection is not applicable to the large-scale networks due to its scalability issue. Therefore, a great many studies have been dedicated to improving the efficiency and effectiveness of seed selection algorithms.

Chen et al. (2009) study the efficient influence maximization by improving the original greedy selection, and propose a novel seed selection approach, namely, Degree Discount Heuristics (DDH) for the uniform IC model.

Wang et al. (2012) handle the influence propagation within large-scale networks by proposing a new heuristic algorithm based on the IC model. The proposed algorithm is capable of reducing the operating time by limiting the computation related to local influence regions of nodes. Furthermore, a tunable parameter is defined to control the balance between effectiveness and operating time.

Leskovec et al. (2007) propose the CELF optimisation that significantly reduces the number of calls (Monte-Carlo Simulations) for the spread estimation procedure, which improve the running time of the classic greedy algorithm by up to 700 times. However, the CELF optimisation still turns out to be not efficient when being applied to super large networks since the first iteration is computationally expensive (W. Chen et al., 2010; A. Goyal, Bonchi & Lakshmanan, 2011). Therefore, Goyal et al. (2011a) improve CELF by proposing the CELF++, which further reduce the number of influence spread estimation calls. Based on the experimental report, the performance of CELF++ is approximately 33%-55% faster than that of CELF.

Goyal et al. (2011b) propose an efficient and effective algorithm based on the LT model, called SIMPATH, which includes three ways of optimising the computation and improving the quality of the seed selection. The quality of the seeds is on the basis of the fact that the larger its spread, the higher its quality.

On top of that, the community-detection approaches are also utilised in the influence maximization problem to improve the quality of the seed set, as well as addressing the efficiency issues.

Wang et al. (2010) introduce a community-based greedy algorithm to mine the top-k influential nodes on the basis of influence diffusion speed. The proposed algorithms detect communities in a social network by considering information diffusion and select communities to find the influential nodes.

Chen et al. (2014) leverage the community structure and develop two efficient algorithms, i.e., Community and Degree Heuristic with Kcut (CDHKcut) and Community and Degree Heuristic with SHRINK (CDH-SHRINK), which significantly decrease the number of candidate influential nodes. Both scalable algorithms significantly outperform state-of-the-art algorithms in efficiency.

However, very few studies handle the influence diffusion in both large-scale and dynamic environment. Furthermore, most proposed algorithms are not superior to the traditional greedy algorithm in terms of effectiveness though the efficiency has been greatly improved.

2.2.2 Considerations in Influence Maximization

As introduced in Chapter 1, in reality, influence possesses sophisticated characteristics which inevitably lead to difficulties in analysing the diffusion process. These characteristics become major considerations to tackle the real-world influence maximization problem.

It is important to understand the factors affecting the probabilities of influence. Goyal et al. (2010) research learning influence probabilities in social networks based on the users' past actions, and successfully predict the time by which a user may be expected to perform an action. Similarly, Saito et al. (2008) estimate influence probabilities by measuring pairwise influences among the individuals. Based on these studies, the probabilities are largely dependent on the individual's behaviours.

Preference-aware Influence

Preference turns out to be one of the most important factors in analysing the influence diffusion in social networks since it contributes to the prediction of influential link formation, as well as the estimation of the corresponding strength. According to the underlying social theories, social contagion describes a sort of phenomenon where user's preferences and actions are influenced by interpersonal contact, impacting the aggregate diffusion and spread of behaviours, new products and innovations (Watts & Dodds, 2007). The user preference can be explicitly reflected as the ratings to items in a heterogeneous network.

Arora and Allenby (1999) insist that users' purchasing behaviours are significantly impacted by the preference structures and the influence they have in a group. Jiang et al. (2015) design a Preference-based Trust IC (PTIC) model, which takes account not only the trust connectivity, but also user preference in constructing influence propagation network. The experimental results demonstrate a better outcome when considering both trust and preference in the influence maximization problem than that of considering trust relationship only. Wu et al. (2016) investigate modelling users' preference and social links by proposing a model with the explanatory ability and predictive power.

To further narrow down the scope of user's preferences, i.e., topical interest as a type of user preference, the investigation of topic-level influence has been studied. Tang et al. (2009) propose topical affinity propagation to model the topic-level social influence and measure the strength quantitatively. By considering user's preferences, Barbieri et al. (2012) extend the classic IC and LT models to be topic-aware, named as Topic-aware Independent Cascade (TIC) model and Topic-aware Linear Threshold (TLT) model, respectively. Chen et al. (2015) define and study topic-aware influence maximization problem, where the Maximum Influence Arborescence (MIA) model has been proposed to approximate the computation of influence spread.

Negative Influence

The real-world is filled with positive information, e.g., advertisement with positive opinions recommending the audience to purchase, while the negative, e.g., rumour, is usually rarer, making the adverse information more diagnostic (Fiske, 1980). In reality, the Word-of-Mouth (WoM) effect may be either positive or negative (East et al., 2008). Motivated by this background, many researchers extend the influence maximization problem and explore the approaches to minimise the adverse impact of an existing influence in a social network. To handle the emergence of negative opinions, the traditional influence diffusion models are supposed to be extended.

Liu et al. (2010) propose an Ising model to predict positive and negative opinion formation in a social network by considering the neighbourhood-based interactions. Wang et al. (2016) propose the IMIC-OC model to explain how users build opinions during the process of information spreading, where both positive and negative opinions are considered. Chen et al. (2011) model influence maximization in social networks when negative opinions may merge and propagate. The IC model with negative opinions, i.e., IC-N, has been proposed, where each individual possesses the same quality factor, referring to the probability of turning to negative.

A bulk of studies attempt to block an influence in a very straightforward way, i.e., altering the structure of a social network. For example, Kimura et al. (2008; 2009) claim to minimise the spread of influence contaminations by removing links. Similarly, Wang et al. (2013) suggest minimising the negative influence by blocking a limited number of nodes in social networks, and Yao et al. (2015) adopt the same solution from a topic modelling perspective.

However, these approaches can only be applied based on the assumption that the organisation is authorised to manage network topological structures. In reality, such modifications are generally not applicable.

On the other side, some researchers tend to achieve the negative influence minimisation by levering the power of competitive influence. In other words, the negative influence minimisation is developed based on the extended models of competitive influence. For example, He et al. (2012) address the influence blocking maximization problem by selecting seed nodes to inject the positive opinions to fight against the negative rumour.

Most studies relying on the competitive influence models intent to suppress an undesirable impact by introducing the opposite influence only. Whereas, the possibility of minimising the adverse impact by introducing other influences has not been experimented. More introductions and studies regarding competitive influence are provided in the following subsection.

Competitive Influence

In traditional influence maximization problem, the influencers are identified by considering a single influence. However, in real-world, various influences coexist in the same environment. Competitive influence diffusion and its corresponding influence maximization problem have been investigated by extending the classic IC model and the LT model.

Bharathi et al. (2007) extend the IC model and focus on the scenario when multiple innovations are competing within a social network. Based on the traditional IC model, Zhu et al. (2016) present the C-IC model to characterise how various influences are competing with others in social networks. Borodin et al. (2010) propose an extended version of the LT model to handle the competitive influence diffusion of two different technologies. Liu et al. (2016) extend the LT model to establish the diffusion-containment model, i.e., D-C model, by incorporating the realistic specialities of the containment of the competitive influence spread. He et al. (2012) attempt to tackle the influence blocking maximization problem and extend the LT model to incorporate competitive influence diffusion. Kostka et al. (2008) present the rumour game which models the dissemination of competing information in social networks. Similarly, Trpevski et al. (2010) model the competitive rumour spreading by extending a well-known epidemic SIS model. Goyal and Kearns (2014) study the product adoption competition between two firms by developing a game-theoretic framework.

However, in nearly all the research work mentioned above, only one type of influence is considered. In other words, these studies focus on the adoption of a particular product or opinion, while other influences in the same context have been ignored. With an exception, Tang et al. (2009) propose topical affinity propagation to model the topic-level social influence, which can identify the experts in different topics and measure the strength quantitatively. Nevertheless, Tang's work is developed based on the assumption that no dependencies are presented among the various topical influences. The impacts and relationships among the multiple influences are neglected.

2.2.3 Dynamics in Influence Maximization

Modelling emergent properties of social networks appears to be one of the pillars of social network science (Holme, 2015). Real-world social networks possess a highly dynamic nature and evolve rapidly over time (R. Kumar, Novak & Tomkins, 2010; Leskovec, Kleinberg & Faloutsos, 2007). More importantly, the network evolution is continuous. Some research works have been dedicated to modelling the influence propagation in dynamically temporal social networks (Holme, 2015; B. Wang, Chen, Fu, Song & Wang, 2017; Song, Li, Chen, He & Tang, 2017). Within such a dynamic environment, the influence maximization problem becomes more challenging. In this subsection, some representative studies of two categories are mainly reviewed, including the dynamic environment and the corresponding strategies.

Temporal Features and Dynamic Topological Structure

Chen et al. (2014) formulate the influence maximization problem by focusing on the temporal factors based on the heat diffusion model, a realistic model that simulates the social influence in accordance with a physical phenomenon, i.e., heat flow (Ma, Yang, Lyu & King, 2008) . Zhuang et al. (2013) study influence maximization in dynamic social networks and first introduce the concept of probing in dynamic networks. Similarly, Bao et al. (2016) study influence maximization in dynamic social networks by proposing an on-line randomised algorithm dealing with both unknown and non-stationary influence probabilities. In (Tong, Wu, Tang & Du, 2017) and (Gayraud, Pitoura & Tsaparas, 2015), traditional IC model and LT model have been extended to capture the dynamic aspect of real social networks. Karim and Holme (2013) attempt to handle the dynamics of users' interactions by proposing a threshold model of cascades. Gomez-Rodriguez et al. (2016) develop a flexible model, i.e., NetRate of the spatio-temporal structure underlying diffusion process.

Whereas, these approaches consider only a few dynamic features of a social network, such as dynamic propagation probabilities. The fully dynamic topological structure of real social networks is not handled when modelling the influence diffusion. With an exception, Naoto et al. (2016) propose the first real-time fully-dynamic index data structure designed for influence analysis on evolving networks, where five network operations, i.e., vertex additions, vertex deletions, edge additions, edge deletions, and propagation probability updates are included.

However, the state-of-the-art dynamic influence maximization solutions are only capable of processing hundreds of updates per second, which is still far from the updated rate in real-world (Ohsaka et al., 2016). In this sense, some research works have been dedicated to the dynamic social streams, which aim to investigate the possible solutions for real-time influence maximization in a dynamic environment.

Konstantin et al. (2013) present STRIP, the first streaming method computing influence probabilities. Subbian et al. (2016) propose an influence-query framework to mine influencers in a time-sensitive fashion from streaming social data. Wang et al. (2017) propose the Influential Checkpoints framework and a Sparse Influence Checkpoints framework to tackle the Stream Influence Maximization (SIM) querying processing.

In spite of the studies of social streaming, for large-scale networks, it is almost impossible to imitate dynamics by capturing snapshots or storing all changes happening around since this inevitably creates another set of big data. Furthermore, capturing snapshots or changes require a global view, which becomes another obstacle as a central component is required for monitoring the entire network in real-time.

Adaptive Influence Maximization

Adaptive influence maximization aims to investigate the adaptive budget allocations based on the changing environment (Golovin & Krause, 2011; Alon, Gamzu & Tennenholtz, 2012; Soma, Kakimura, Inaba & Kawarabayashi, 2014).

Hatano et al. (2016) address budget allocation for maximizing influence by considering adaptive strategies. Yang et al. (2016) model the continuous influence maximization problem and devise a coordinate descent framework. Similarly, Rodriguez and Schölkopf (2012) study influence maximization in continuous time diffusion networks by developing INFLUMAX model that accounts for the temporal dynamics underlying diffusion process. Song et al. (2017) address the influence maximization in dynamic social networks by leveraging an interchange greed approach. The elements in a seed set are constantly replaced with the evolution of the network, rather than reselecting all the influencers. An efficient algorithm based on UBI, i.e., UBI+ has been proposed, which can improve the computation of node replacement upper bound.

2.3 Multi-Agent Systems and Influence

Agent-Based Modelling (ABM) and Multi-Agent Systems (MAS) have demonstrated many advantages in modelling complex systems, simulating continuous variations and analysing the trend of a particular phenomenon (Bonabeau, 2002; Macal & North, 2009). Moreover, they are more suitable for exploring the macro world through defining a micro level of a social system (Bai, Zhang & Zhang, 2005; Ye, Zhang & Vasilakos, 2016).

MAS is comprised of numerous autonomous interactive agents within the same environment, which are supposed to be goal-driven and self-directed (Wooldridge, 2009; Gilbert, 2008). There are substantial and active literature of ABM and MAS, including methodologies and applications (Macal & North, 2009), and the agent-based approach has been widely applied to simulate a particular phenomenon at a microscopic level. Based on the existing research works (Sycara, 1998; Wooldridge, 2009; Su, Zhang & Bai, 2015; Bai et al., 2005; Ye et al., 2016), the major characteristics of MAS can be summarised as follows.

- **Autonomous:** agents access the environment and behave independently. They can make decisions based on the heuristics or rules.
- **Adaptive:** the behaviours of the agents are adapting based on the changing environment.
- **Local view:** each agent has its own local view, covering partial space/information of the entire environment, as well as limited understanding of other agents' states.
- **Limited capabilities:** the capabilities of each agent are limited. In particular, they are usually restricted by the design or energy.
- **Decentralised:** The MAS does not allow a central control. The information appears decentralised and is stored at the local storage of each individual agent.
- **Openness:** MAS is an open system, allowing the existing agents to discompose and new agents to join in the same environment.

On the other side, the agents can interact with each other directly or indirectly. More specifically, agents contact the reachable peers and deliver information directly. In contrast, the communications among the agents also can be mediated through leaving and accessing messages in the same environment. In this subsection, I will review the studies of both types of MAS being utilised in the influence spread analysis.

2.3.1 Simulations of Influence Diffusion

Many studies have been focusing on the agent-based simulations for recreating and predicting the appearance of a particular social phenomenon (Gustafsson & Sternad, 2010). Almost all the studies in this field leverage the direct communicational mechanism, where the agents exchange the information directly with each other.

The influence modelling using agent-based approaches defines three levels of social influence (Kiesling, Günther, Stummer & Wakolbinger, 2012).

- **Micro-level influence** travels through pairwise communication links. The WoM effect is one of the typical forms of micro-level influence, where the information is passing from person to person by oral communications.

- **Meso-level influence** stems from an agent's immediate social environment, such as neighbourhood or community. Meso-level social influence is associated with some typical concepts, incorporating group conformity, social comparison, herding behaviour, etc (Veblen, 2017).

- **Macro-level influence** is defined as global interactions at the level of the entire network as a whole. The network-level opinion aggregation pheromone demonstrates a macro-level influence (?, ?).

On top of that, social theories play a significant role in explaining various types of social phenomena and modelling influence diffusion. According to the social influence and homophily effect, homogeneous influence suggests that similarity leads users to interact and interaction leads users to behave more similarly (Z. Li & Tang, 2012; L. Wu et al., 2016). The demographic and preference similarities encourage the users to adopt the behaviours or accept the influence from the neighbours (Aral, Muchnik & Sundararajan, 2009). In other words, individuals tend to follow the behaviours and conform the ideas of their friends since they view them as a source of valid information;

adjacent users normally exhibit similar behaviours (Tang, Chang & Liu, 2014; Kelman, 1961). Many social influence simulations are developed based on such rules.

Centola et al. (2007) articulate that the influence propagation processes can produce global convergence. Axelrod (1997) also shows the homophily and influence are capable of acting of local convergence mechanisms, and both produce emergent social cleavages that lead to global polarisation. Tang et al. (2013) formally define the influence conformity and conduct contribution analysis of various social factors, and their studies show that the conformity plays a significant role in predicting influence acceptance. Van Maanen and Van der Vecht (2013) propose a multi-disciplinary agent-based approach for studying on-line social network influence. Li and Tang (2012) investigate the non-positive social impact on group polarization based on Hop-field network model by considering the dyad and triad influence and social pressure, so that social structure balance is achieved. Li and Tang (2015) investigate the microscopic social mechanisms through ABM and explore the non-positive social impact on group polarization, where the Voter Model (VM) and Hopfield Attractor Model (HAM) have been applied. Empirical data evidence proves the importance of non-positive influence in promoting voters' opinion polarisation. The social network demonstrates a convergent trend when considering three types of valence of the social identity tie. The proposed model considers the polarity relationship among the users, which has been measured by dyadic and triadic closure.

Whereas, the complex nature of influence is not considered in the most of existing research work. Moreover, one of the critical assumptions is that users' opinions are not supposed to be revised once formed. The dynamics of individual's complex behaviours and varying personalised features are ignored as well.

2.3.2 Stigmergic Interactions

The stigmergic interactions is a type of indirect communication utilised in the ant and stigmergy-based algorithms. The agents are not supposed to interact with each other directly, but allocating the messages to the same environment, which are accessible and can be referred by other agents. Stigmergy-based systems have shown that they can be applied to generate complicated and robust behaviours in the systems even if each ant has limited intelligence. Some researchers have applied stigmergy in the computer science field.

Dorigo et al. (2000) introduce how to solve the Travelling Salesman Problem (TSP) (K. Wang, Huang, Zhou, Pang et al., 2003) by leveraging ant and stigmergy-based algorithms, where the pheromones are allocated by considering the distances among the cities. Ahmed et al. (2014) propose a stigmergy-based approach for modelling dynamic interactions among Web service agents in decentralised environments. Takahashi et al. (2012) propose an anticipatory stigmergy model with allocation strategies for sharing near future traffic information related to traffic congestion management in a decentralised environment. Hadeli et al. (2004) introduce a novel design and prototype implementation for manufacturing control systems using stigmergy, which tends to handle the changes and disturbances. Lewis (2013) claims that the essential social networking behaviours of human beings are in fact forms of stigmergy, and attempts to explain a theory of group formation based on stigmergy.

The concept of stigmergic interactions is not new in the contemporary research field. However, this mechanism is not fully utilised in the influence maximization problem though it demonstrates its superior in handling optimisation problems in a distributed manner.

2.4 Summary

In this chapter, I have given a detailed review of the related works in the field of influence diffusion modelling, influence maximization and agent-based modelling.

By analysing the pros and cons of these studies, the gaps from the literature review are summarised as follows.

- Most studies consider the influence as a simple hopping and infecting process, but the complex nature of influence is ignored, as well as individuals' personal traits, behaviours, information intake capacity. These factors also contribute to the influence acceptance.
- It is difficult for the existing influence-diffusion models to function when the global view is not available.
- Very few studies focus on handling both dynamics and large-scale networks at the same time. In spite of this, the existing algorithms appear not superior to the classic greedy selection though the efficiency is greatly improved.
- Few studies explore the long-term evolutionary trend of social networks driven by influence. Furthermore, how to handle the situations when group opinions are revised according to the changing context is not systematically articulated.
- Influence maximization problem has been extensively studied. However, there is so far nearly no research work tracking the status of influence messages or investigating the maintenance of a particular influence.
- Almost no research work has been conducted to modelling multiple influences coexisting in the same environment. Such phenomenon is worthy to be investigated for suppressing the negative opinions especially without any control to the network topological structure.

The literature review has pointed out the above research gaps of influence-diffusion modelling and influence maximization. This also reveals the in-depth studies demanded in this field. The solutions presented in this thesis cover the aforementioned research gaps by proposing influence diffusion models from an agent-based perspective to address the influence maximization and the extended problems.

Specifically, in Chapter 3, the fundamental agent-based framework is proposed and articulated, serving as a basis for the further investigations. A novel decentralised model, i.e., SIMiner, introduced in Chapter 4 is capable of handling both dynamic and large-scale networks. In Chapters 5 and 6, studies of influence maintenance and multiple influences diffusion using agent-based approaches are conducted extensively, which attempt to address the real-world issues for the applications of influence diffusion. The detailed realisation of the agent-based models and solutions are presented in the following chapters.

Chapter 3

Agent-based Influence Diffusion Modelling

The challenges of modelling influence diffusion stem from its complexity and the evolving environment. Therefore, it is essential to develop a generic framework for handling both features before investigating further.

In this chapter, I systematically articulate the complex effects of social influence and model the influence-diffusion space as a Hybrid Social Network (HSN), which could be formed by merging a number of homogeneous or heterogeneous networks representing possible influence propagation channels. Furthermore, by considering the sophisticated influential relationships derived from the users' interactions, I proposed an agent-based framework to architect the influence diffusion in complex networks through modelling the individual's characteristics and behaviours that caused influence propagation.

In this thesis, the proposed generic hybrid and agent-based framework serve as the foundation for exploring in-depth issues of influence-diffusion in the following chapters. In Chapter 4, the proposed generic framework will be extended by incorporating the indirect communications among the agents. The extended influence-diffusion model can assist researchers in tackling the scalability issues and networked dynamics of the influence maximization problem. In Chapter 5, I further expand the generic agent-based framework by considering the status of influence messages and investigate long-term influence maintenance problem. Chapter 6 studies multiple influences on the basis of the proposed generic agent-based framework. The model inherits some of the key features from the influence-maintenance model introduced in Chapter 5.

3.1 Overview

Most research works investigate influence diffusion in social networks based on the existing network topological structure, where friendship-affiliation links represent influence propagation channels, and the strength of connections is considered as the only factor affecting the influence propagation probability. Therefore, an ordinary assumption is that friendship affiliation links are equivalent to influential links. However, this cannot hold in general, as friendship-affiliation links and influential links are naturally two different types of links coexisting in a social network.

Influence is a hybrid effect, which can be decomposed into multiple components focusing on different activities of human-beings (Anthony, 2009). Specifically, influence is presented as a mixed type of communications and interactions, such as perceiving information posted by the friends of on-line social networks, delivering messages or emails, conducting face-to-face discussions, reviewing the comments from web-pages, etc. Hence, any of these behaviours are capable of exerting influence and impacting individuals' decisions. However, most researchers ignore the multiple possible interactive diffusion channels. On the other side, in many situations, individuals are more likely getting influenced by the 'stimuli' left by others, especially in E-commerce domain. Feedbacks of a particular product, such as reviews, ratings, comments from previous buyers influence the purchasing decisions of others, even they are not adjacent neighbours and without any immediate interactions. Therefore, the influence is still capable of propagating through the users in the same context without explicit links, since they are affiliated implicitly through other features, such as similar preference and criteria for items. Apparently, this is an important feature to be considered in influence propagation modelling, but unfortunately, ignored in most existing works.

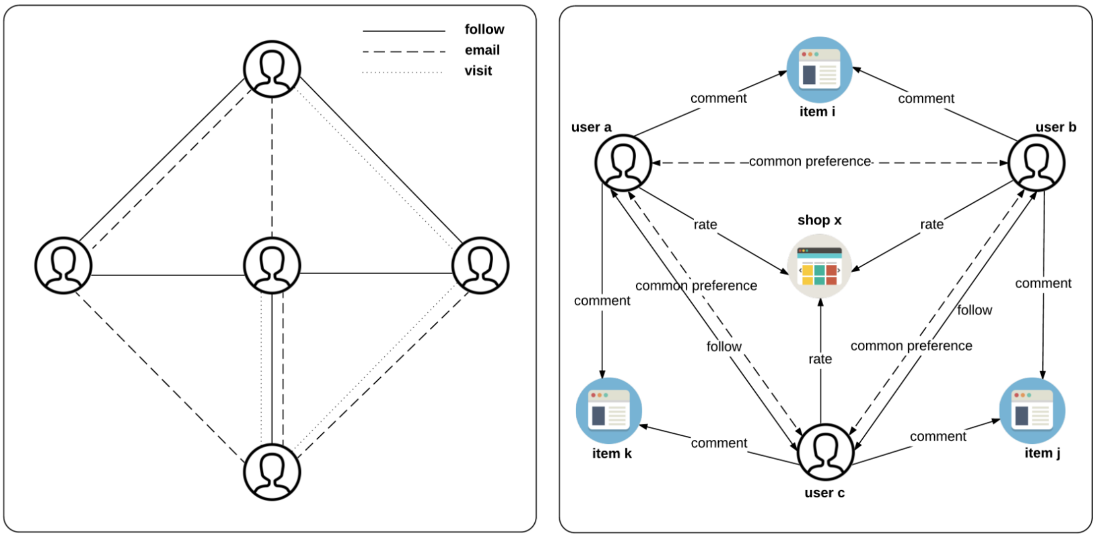
3.2 A Generic Hybrid Framework

3.2.1 Direct and Indirect Influence

Influence diffusion is a sort of communication which concerns the spread of messages perceived as new ideas or innovations (Y.-C. Chen et al., 2014). Thus, in this chapter, direct and indirect influence emphasises the types of communicational channels. More specifically, direct influence refers to the immediate interactions or message reciprocation among users with explicit connections, such as friendship affiliation.

The concept of indirect communicational influence stems from the *ant and stigmergy algorithms*, where ants interact each other and conduct group activities by leaving and sensing pheromones (Dorigo et al., 2000). By tailoring this idea, indirect influence describes a form of indirect communications among the users mediated by modifications of the environment. Specifically, some users are not connected explicitly, but diffuse influences by leaving the messages, such as ratings, comments, reviews, beliefs, etc. Meanwhile, individuals are getting influenced by reading the information produced by their counterparts, since they locate in the same environment. In other words, indirect influence implies something shared among the users, which can generally be regarded as common preferences from a broad view. Sometimes, the strength of indirect influence is even more prominent than that of direct influence, especially in sparse networks.

Figure 3.1 demonstrates two typical examples focusing on direct and indirect communicational influence. In Figure 3.1a, three possible direct influence-diffusion channels



(a) Direct Influence Propagation

(b) Indirect Influence Propagation

Figure 3.1: Influence Diffusion in Social Networks

exist. Whereas, other types of nodes, i.e., item and shop, are involved in Figure 3.1b, where users a and b potentially influence each other via the messages delivered to the item and shop, though they are not connected explicitly in this social context. As for the other two pairs of users, i.e., users a and c, users c and b, they share both implicit and explicit influential links.

3.2.2 Hybrid Social Network

Hybrid Social Network refers to an implicit heterogeneous user-centric network comprised of a number of social networks concerning possible direct and indirect influence propagation channels. It aims to model various influential relations among the individuals. Meanwhile, it implies the decomposition of influence effects, which gives high extensibility and flexibility. When other available influential factors are added or the existing elements are revised, the model can be adapted easily by updating a particular facet.

Figure 3.2 demonstrates a generic HSN model proposed in this thesis. Social

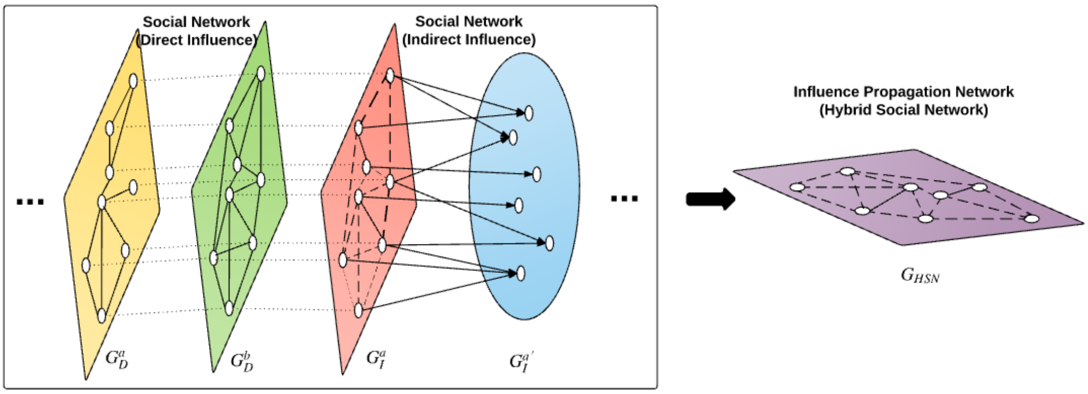


Figure 3.2: Hybrid Social Network Composition

networks inside the rectangle container represent different influence diffusion channels, and are supposed to be extracted from the original heterogeneous social network. There are two types of networks, i.e., the direct and indirect influence propagation network. The former is a homogeneous network, while the latter is a heterogeneous network, where the indirect influential relationships are established via their corresponding object layer. Specifically, in this figure, $G_I^{a\prime}$ is the object/item layer of network G_I^a. G_{HSN} is constructed by merging all the social networks.

HSN demonstrates a network of networks pattern driven by merging various social networks from a low level and reconstructing a hybrid influence-diffusion platform. As mentioned previously, a HSN is comprised of direct and indirect influence propagation networks. The former incorporates the explicit social networks with direct links, which show full lines in this figure. Whereas, the latter consists of a variety of implicit networks formed by examining the common features or behaviours of the users, where the implicit relations are denoted using dotted lines.

HSN can be characterised by the diverse entities and influential relations among them. Take Taobao[1] as an example, customers (users), shops, items, etc. coexist with various types of ties. Customers are not only capable of establishing affiliation

[1]http://world.taobao.com/

relationships by following others, but also linked with certain items by adding them to watching lists or rating them after transactions. In addition, users develop influential implicit relationships with others via other types of entities in the social context. Since users with common interests are eager to spend efforts on reading and writing the attributes of these entities. In other words, customers can interact indirectly by accessing the entities in the same context. Thus, the implicit influence channels can be established on the basis of common features of the customers, such as preferences.

The proposed HSN is a type of centralised model, which demonstrates a user-centric pattern in modelling influence diffusion in social networks. However, HSN requires attention to other types of entities, as well as the interactions among users and these entities. Moreover, HSN still needs the topological structure of various networks, which appears relatively difficult to be obtained. Therefore, inspired by the fundamental concepts derived from HSN, a generic multi-agent system is designed for modelling influence diffusion.

3.3 Influence Diffusion Multi-Agent System

To simplify the complicated influence-diffusion modelling process, I proposed a Multi-Agent System (MAS), namely, the Influence Diffusion Multi-Agent System (IDMAS), to model the propagation process as an evolutionary behavioural pattern of a social network.

Fig. 3.3 shows a generic framework of IDMAS. Users in a social network are represented as autonomous and self-directed agents (Macal & North, 2009). They locate in the centre of their local social environments, i.e., directed weighted influence ego-network. The influences among the agents are mediated by direct communications. Each agent has a *preference state* (see Definition 3.2) towards a hypothesis item, where the state is affected by two major factors, i.e., individual's valuation of the product

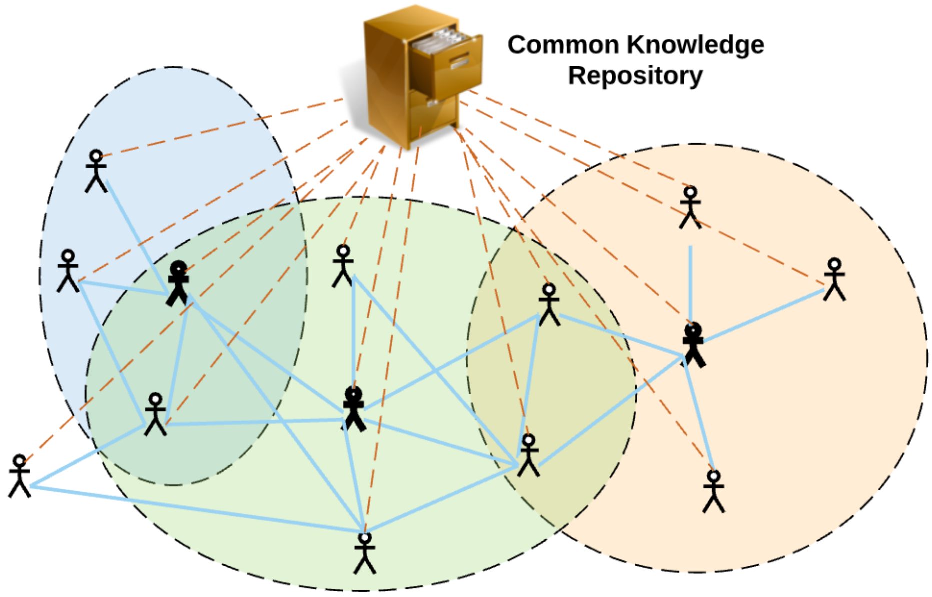


Figure 3.3: Framework of IDMAS

(Bhagat, Goyal & Lakshmanan, 2012) and social conformity (Tang et al., 2013). More specifically, an agent not only evaluates its intrinsic appreciation towards a particular item but also handles the social influence coming from the neighbours, especially those with diverse opinions. Hence, in IDMAS, agents are capable of adapting their preference states and behaviours based on the internal factor: predisposition towards an item or an event, as well as the external factor: social context. In this way, agents are attempting to reach a “comfortable" state over time. On the other side, limited global information, i.e., Common Knowledge Repository (CKR), is available for all the agents. CKR provides sufficient data for all the agents to evaluate personalised parameters by training themselves locally. Meanwhile, agents’ critical behaviours are also uploaded to the CKR.

Compared with the traditional influence diffusion models (Kempe et al., 2003), the proposed system demonstrates the advantages in three fields. **(1)** IDMAS is capable of capturing the complexity of influence diffusion due to the heterogeneous entities in a social environment since it concentrates on modelling individuals at a microscopic level.

IDMAS neglects influence propagation paths and attenuation degrees but focuses on the users' personalised characteristics, behaviours and the social influence undertaken from the local context at a particular time frame. Moreover, there can be numerous individualised parameters when analysing influence diffusion in a social network, such as behavioural, demographic, geographic, status-related, and preference similarities (Aral et al., 2009). By utilising ABM, individual agents can be heterogeneous and modelled by selecting the principle features associated with the influence capabilities in a specific context. **(2)** IDMAS can aid the analysis and predictions of networked evolution by considering time series. More specifically, the entire network evolves through a number of discrete time steps according to a set of transitional or behavioural rules of each agent. The network in time step t_i is regarded as generation g_i. The agents keep adapting and evolving, and the network becomes generation g_{i+1} in the next time step t_{i+1}. In this way, the network state at t_{i+n} can be predicted. Thus, IDMAS simulates the macroscopic influence trend through defining the micro-level agents. **(3)** Each user agent in IDMAS is capable of adapting its personalised parameters, such as the degree of stubbornness, i.e., a factor of a user reflecting how easily he or she can be influenced by others, based on the historical behavioural transactions and limited global information retrieved from CKR. This feature allows agents, even those with very few past transactions, to train themselves.

3.4 Modelling Agent-Based Influence Diffusion

In this section, I drill down into the details of the proposed model. Each individual agent's personalised features and actions with respect to influence are elaborated and formally defined.

Inspired by the idea presented in (Chantemargue et al., 1998), the architecture of individual agents is designed as the form shown in Figure 3.4. For each agent,

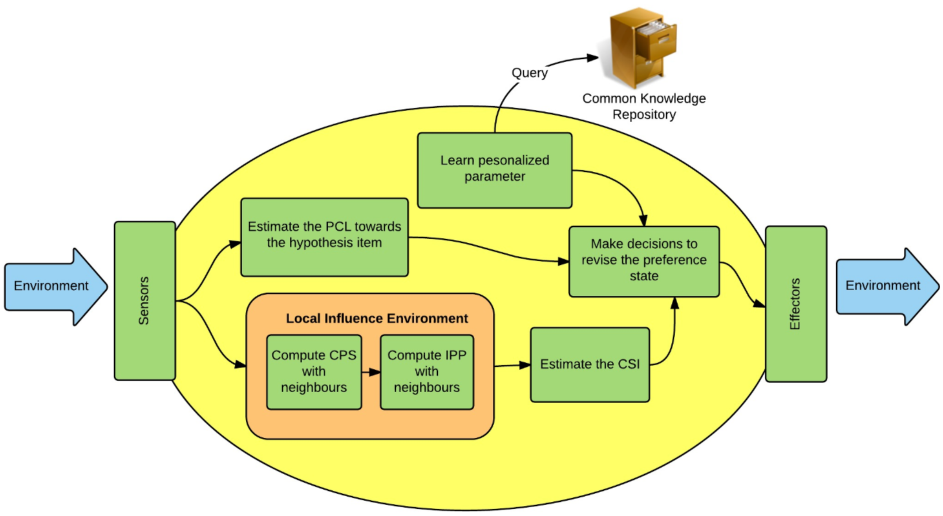


Figure 3.4: Individual Agent Architecture of IDMAS

three major actions are incorporated, i.e., constructing local influence environment, learning the personalised parameter and making decisions for preference state revision. Specifically, to establish a local environment, agents attempt to estimate the influential relationships based on the past transactions. Next, by examining both identical and adverse opinions, an agent measures the Comprehensive Social Influence (CSI) from its ego-network. Meanwhile, Prior Commitment Level (PCL) towards the hypothesis item is evaluated by the agent, indicating its predisposition of this item. Subsequently, the agent commits a preference state by considering both CSI and PCL (refer to Definitions 5 and 6). On the other hand, to balance the CSI/PCL trade-off, an agent queries the CKR and conducts training locally to learn and adjust the personalised parameter. The detailed modelling of agents and social behaviours are elaborated in the following subsections.

3.4.1 Individualised Agent

Definition 3.1: An Agent is defined as a vertex $v_i, v_i \in V$ in the directed weighed network $G = (V, E)$, where $V = \{v_1, ..., v_n\}$ denotes a set of agents and E represents a set of edges; $E = \{e_{ij} | 1 \leq i, j \leq n\}$, where $i, j \in \mathbb{N}, \{v_i, v_j\} \subseteq V$. The weight (strength) of edge e_{ij} denotes the influence propagation probability from v_i to v_j (refer to Definition 3.4). Agent v_i has a set of neighbours $\mathit{\Gamma}(v_i)$. If v_j is a neighbour of v_i, then $\{e_{ij}\} \cup \{e_{ji}\} \subseteq E$, where $v_j \in \mathit{\Gamma}(v_i)$. $|V|$ and $|E|$ denote the cardinality of agents and edges, respectively.

Items, another type of entities, also exist in the same context. I denotes the item set, i.e., $I = \{i_1, ..., i_k\}$. Agents interact with the item set by giving rating scores, and an agent v_j maintains its own rating list R_j locally. The preference of v_j towards item i_x derives from its ratings to items $\{r_{jx} | 1 \leq j, x \leq n\}, i, j \in \mathbb{N}, r_{jx} \in R_j$, where r_{jx} signifies the rating score of item i_x given by agent v_j. $\overline{r_j}$ signifies the average rating value given by v_j. Meanwhile, each agent has a preference state towards any item.

Definition 3.2: Preference State s_{jx} is defined as agent v_j's opinion towards item i_x. s_{jx}^t denotes the preference state in a particular time step t, and $\overline{s_{jx}}$ represents the initial opinion. There are three possible values of a preference state, i.e., $s_{jx}, \overline{s_{jx}} \in \{PA, NA, IA\}$:

- *Positively Activated* ($s_{jx} = PA$). Agent v_j shows its favour towards item i_x. v_j tends to diffuse positive influence to the neighbours $\mathit{\Gamma}(v_j)$, and enhances the agents with same opinion and tries to change those with opposite opinions.

- *Negatively Activated* ($s_{jx} = NA$). Agent v_j expresses disfavour towards item i_x, and $\mathit{\Gamma}(v_j)$ will be negatively influenced.

- *Inactivated* ($s_{jx} = IA$). Agent v_j holds a neutral opinion towards item i_x. v_j is not supposed to exert any influence on the neighbours, but tends to be influenced by any of $\mathit{\Gamma}(v_j)$ with non-neutral opinions.

3.4.2 Local Influence Environment

Each individual agent constructs and maintains its own influence environment, which is represented as an ego-network, incorporating all its adjacent neighbours and the influential relationships among them. The establishment of such an environment is based on one of the well-known social theories, i.e., homophily and influence (Centola et al., 2007). In the current setting, the influential strength, i.e., the weight of the links in each ego-network, can be measured by considering Common Preference Similarity (CPS).

Definition 3.3: Common Preference Similarity cps_{ij} is defined as the similarity degree of both v_i and v_j's valuations towards the items in a social network. The cps_{ij} can be measured based on the co-rating by using the Adjusted Cosine measure (ACOS) (Ahn, 2008; H. Liu, Hu, Mian, Tian & Zhu, 2014), which is formulated as follows.

$$cps_{ij} = \frac{\sum_{i_x \in I}(r_{ix} - \overline{r_i})(r_{jx} - \overline{r_j})}{\sqrt{\sum_{i_x \in I}(r_{ix} - \overline{r_i})^2}\sqrt{\sum_{i_x \in I}(r_{jx} - \overline{r_j})^2}} \tag{3.1}$$

Definition 3.4: Influence Propagation Probability (IPP) ipp_{ij} is defined as the probability that agent v_i propagates influence to v_j effectively, where v_i and v_j are influence source and target, respectively. As influence diffusion is directed, thus, $ipp_{ij} \neq ipp_{ji}$.

The IPP can be derived from the CPS by considering the rating counts. The more ratings an agent gives, the more powerful influence it propagates (Zhang et al., 2015). ipp_{ij} is formulated in Equation 3.2, where $|I_i|$ denotes the number of items rated by agent v_i.

$$ipp_{ij} = cps_{ij} \cdot \frac{|I_i|}{|I_i \cup I_j|} \tag{3.2}$$

3.4.3 Social Influence and Agent Behaviours

Social influence generally refers to the way that individuals' cognition, thoughts and behaviours are affected by others (Raven, 1964). In the current setting, social influence indicates an agent's capability of affecting the neighbours' sentiment or preference state towards any item. Each agent constantly receives the influence from its neighbours, which can be either positive or negative.

Given a hypothesis item, an agent with neutral opinions may turn positive, negative or retain the same preference state when getting influenced by the neighbours. Whereas, non-neutral agents can revise their views due to the varying social context. The revising preference state action normally depends on a set of transitional rules, which consider the agent's intrinsic behavioural trait, i.e., prior commitment level towards the hypothesis item, as well as the external factor, i.e., social influence from the peers around.

Definition 5: Prior Commitment Level (PCL) pcl_{jx} is defined as agent v_j's estimated prior preference state or predisposition towards a hypothesis or rated item i_x on the basis of past ratings or experiences. To be more specific, if an estimated or actual rating on item i_x almost reaches the highest rating score, v_j has a strong tendency of becoming PA, vice versa. Thus, pcl_{jx} of turning positive can be formulated in Equation 3.3 by using Min-Max Normalisation.

$$pcl_{jx} = \begin{cases} \frac{r_{jx} - min(R_j)}{max(R_j) - min(R_j)}, & max(R_j) \neq min(R_j) \\ 0.5, & max(R_j) = min(R_j) \end{cases} \tag{3.3}$$

In Equation 3.3, $max(R_j) - min(R_j)$ refers to the gap between the highest and the lowest rating values given by agent v_j. While $v_j's$ PCL of turning negative is represented as $1 - pcl_{jx}$; PCL of retaining neutral opinion on i_x is depicted as $1 - |pcl_{jx} - 0.5|$.

On the other side, PCL also reflects a user's possible initial preference state towards an item. User v_j's initial state towards i_x, i.e., $\overline{s_{jx}}$, can be evaluated based on the PCL by using a simple threshold model. Specifically, having two predefined thresholds θ_{PA} and θ_{NA} representing the PA threshold and NA threshold, respectively, then:

$$\overline{s_{jx}} = \begin{cases} PA, & if\ pcl_{jx} > \theta_{PA} \\ NA, & if\ pcl_{jx} < \theta_{NA} \\ IA, & otherwise \end{cases} \tag{3.4}$$

Definition 6: Comprehensive Social Influence (CSI) $csi_{jx}(s)$ is defined as a direct implicit influence on recipient v_j, who receives influence messages from the immediate neighbours $\Gamma(v_j)$, being encouraged to retain the current opinion towards item i_x, or to revise its opinion to the opposite preference state, where s denotes the targeting state, and $s \in \{PA, NA, IA\}$.

The comprehensive social influence degree can be measured by examining the opinions of reachable peers (Z. Li & Tang, 2012; Macy, Kitts, Flache & Benard, 2003). In other words, when v_j has more adverse-opinion neighbours with strong IPP, v_j has a higher tendency to update or revise its current opinion. Whereas, neighbours with the same preference states contribute supportive influence; such strengths are capable of alleviating the chances of changing the current preference state. Hence, the comprehensive social influence of a retaining/revising state to s effected on agent v_j can be estimated by using Equation 3.5.

$$csi_{jx}(s) = 1 - \prod\nolimits_{v_i \in \{v_m | v_m \in \Gamma(v_j) \wedge s_{mx} = s\}} (1 - ipp_{ij}) \tag{3.5}$$

In Equation 3.5, s represents a particular preference state of agent v_j towards item i_x. If $s_{jx} = s$, then $csi_{jx}(s)$ refers to the supportive strength to encourage v_j to retain its current preference state. Otherwise, $csi_{jx}(s)$ indicates the social pressure from the neighbours, suggesting v_j revise its state from s_{jx} to s.

As mentioned previously, agents can reconsider the preference state on the basis of both PCL and CSI. Hence, the probabilities of v_j turning PA, NA or retaining IA are formulated in Equations 3.6, 3.7 and 3.8, respectively.

$$p_{jx}(PA, s_{jx}) = \lambda_j \cdot pcl_{jx} + (1 - \lambda_j) \cdot \frac{csi_{jx}(PA)}{\sum_{s \in \{PA,NA,IA\}} csi_{jx}(s)} \tag{3.6}$$

$$p_{jx}(NA, s_{jx}) = \lambda_j \cdot (1 - pcl_{jx}) + (1 - \lambda_j) \cdot \frac{csi_{jx}(NA)}{\sum_{s \in \{PA,NA,IA\}} csi_{jx}(s)} \tag{3.7}$$

$$p_{jx}(IA, s_{jx}) = \lambda_j \cdot (1 - |pcl_{jx} - 0.5|) + (1 - \lambda_j) \cdot \frac{csi_{jx}(IA)}{\sum_{s \in \{PA,NA,IA\}} csi_{jx}(s)} \tag{3.8}$$

In Equation 3.6, $p_{jx}(PA, s_{jx})$ denotes the probability of agent v_j to revise the preference state towards i_x from any state to PA. Similarly, Equations 3.7 and 3.8 formulate the probability of revising or retaining the current preference state s_{jx} as NA and IA, respectively. The opinion reconsideration of an agent is triggered on a time step basis. Meanwhile, v_j's personalised parameter, i.e., λ_j, is a trade-off factor between the PCL and CSI of v_j, representing the stubbornness degree of the agent. The personalised parameter learning process is detailed in the following section.

3.4.4 Learning Personalised Parameter

As aforementioned, different agents have divergent personalised parameters, describing the characteristics of users in terms of influence acceptance. For example, $\lambda_j = 0$ means that user v_j's preference state reconsideration totally depends on the neighbours' opinions. Whereas, $\lambda_j = 1$ implies that v_j's behaviours rely on its own intrinsic opinion only. To measure λ_j for agent v_j, training is required to be conducted locally based

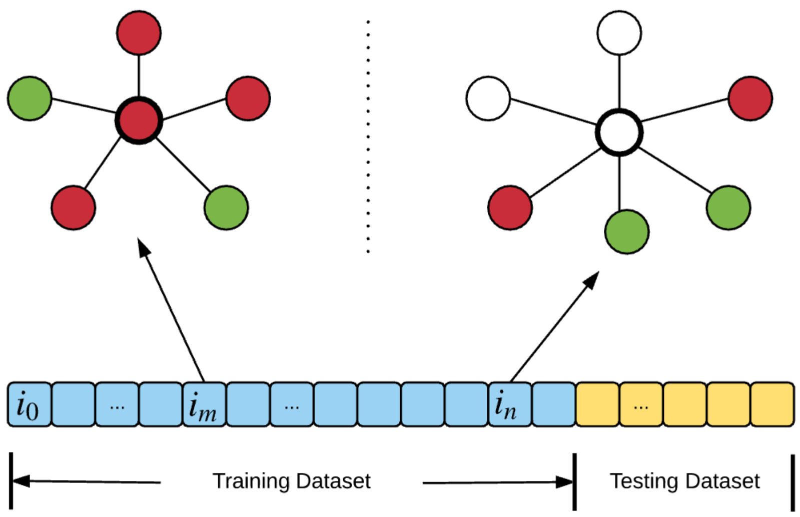


Figure 3.5: Personalised Parameter Learning

on the historical and predicted transactions, i.e., ratings to items. To be more specific, agent v_j's personalised parameter can be estimated by comparing all behaviours with the adjacent neighbours. If v_j's behaviours are similar to most of the neighbouring agents, it has a high tendency of following the neighbours' opinions. Otherwise, v_j appears not influenced by the social context easily.

Figure 3.5 demonstrates the idea of how agents train their personalised parameters. Based on the proposed individual agent architecture of IDMAS in Figure 3.4, each agent queries the CKR for rating information and establishes its local influence environment before initiating a training process. The rating set for each item is regarded as an element of a training dataset, reflecting the users' opinions. User agents derive the personalised parameter by comparing the item ratings among itself and corresponding neighbours.

The preliminary task is to obtain the user-item rating matrix, since it is impossible for each individual to rate every item, and the matrix is sparse with many missing values. Hence, collaborative filtering (Breese, Heckerman & Kadie, 1998; Linden, Smith & York, 2003) can be employed to tackle this issue. One of the popular approaches is the Probabilistic Matrix Factorisation (PMF), which performs well on a large, sparse, and imbalanced dataset (Mnih & Salakhutdinov, 2008). In the current setting, the input to PMF can be the user-item rating matrix with some missing values corresponding to individuals, who have not rated an item. PMF is capable of predicting the missing ratings and fill the matrix. After obtaining the dataset, it has been partitioned into training dataset and testing dataset. Agent v_j trains λ_j by minimising the objective function (Equation 3.9), where $pmf(r_{jx})$ denotes the predicted rating result produced by PMF.

$$\varphi(\lambda_j) = min \sum_{i_x \in I} \left| \frac{p_{jx}(PA, s_{jx})}{\sum_{s'_{jx} \in \{PA,NA,IA\}} p_{jx}(s'_{jx}, s_{jx})} - \frac{pmf(r_{jx})}{max(R_j)} \right| \tag{3.9}$$

3.5 Agent-based Influence Maximization

One typical application of the proposed model is to address the influence maximization problem (Kempe et al., 2003; Domingos & Richardson, 2001). The traditional influence maximization approaches aim to identify a limited set of influencers, expecting that they can propagate influence and maximize the positive impact across the entire social network. The selected users are called seeds, and the selection process is named as seed selection. *Activation/adoption coverage* (Bhagat et al., 2012) generally has been employed to evaluate the effectiveness of a seed selection algorithm, indicating the number of users getting influenced when the selection algorithm has been applied.

To extend the influence maximization problem a little bit, I consider time series by estimating the global network state in each time step, where the time series refer to a sequence taken at consecutive equally spaced points in time, representing a sequence of discrete-time data. Hence, I define agent-based influence maximization as follows. Given a fixed budget (seeds amount), i.e., $|A| = k$, the elements of seed set $A = \{a_1, a_2, ..., a_k\}$ from V, $A \subset V$, are selected to maximize the *Cumulative Influence*

Activation Coverage Difference (CIACD) over m time steps, i.e., $\eta(A, m)$, which is defined as follows.

$$\eta(A, m) = \sum_{t=1}^{m} f(S \cup A, t) - \sum_{t=1}^{m} f(S, t) \tag{3.10}$$

In Equation 3.10, S denotes the existing PA users in the social network, $S \subset V$. $f(.)$ is a function computing the weighted difference between PA users and NA users, which is formulated in Equation 3.11.

$$f(S, t) = \alpha |PA|_t - (1 - \alpha)|NA|_t, \tag{3.11}$$

where $|PA|_t$ denotes the cardinality of PA users in the social network at time t, likewise, $|NA|_t$ refers to the amount of NA ones at the same time step. A trade-off factor $\alpha \in (0, 1)$ has been introduced to present the different weights for PA and NA.

3.5.1 Seed Selections in IDMAS

In the studies of the influence maximization problem, there are two major types of seed-selection approaches, i.e., non-feedback model and full-feedback model (Kempe et al., 2003; Golovin & Krause, 2011). In the former, seeds are selected by ranking the node features in the social network, such as node degrees, betweenness or closeness. In other words, all the influencers are identified in advance. Whereas, in the latter, influencers are gradually identified wholly based on the feedback obtained from the social network. That means, in the seeding procedure, each seed is selected on the basis of previous observations. For example, greedy selection belongs to full-feedback models, which aims to obtain the maximum influence marginal gain in selecting each seed (Kempe et al., 2003; W. Chen et al., 2010).

The traditional seed selection approaches of both types are also applicable in IDMAS. However, I assume that the pre-existing negative users are not supposed to be selected as seeds, i.e., $\forall v_j \in A, \overline{s_{jx}} \neq NA$, but their preference states can be revised with the network evolutions. Furthermore, full-feedback models need to be adapted for obtaining the feedback from a decentralised environment. Hence, the traditional greedy selection can be tailored in Algorithm 1.

Algorithm 1 Adapted Greedy Selection Algorithm

```
Input: G = (V, E), k, i_x
Output: A
Initialise A := ∅
for i = 1 to k do
    for ∀v_j ∈ V \ A ∧ s̄_jx ≠ NA do
        Set seeds A ∪ {v_j}
        Network evolves for m time steps
        Calculate η(A, m) using Equations 3.10 and 3.11
    end for
    Select v_x which has the maximum η(A ∪ {v_x}, m)
    A := A ∪ {v_x}
end for
```

I assume that if any user is selected as an element of a seed set, e.g., the mouthpiece of a particular product, the corresponding status retains PA all along the entire networked evolutionary process. In other words, seeds are not affected by any other, but exert a positive influence of the hypothesis item on the adjacent neighbours.

3.5.2 Enhanced Evolution-Based Backward Seed Selection (2E2B)

In (W. Li, Bai & Zhang, 2016a), we proposed the Evolution-Based Backwards (EBB) selection algorithm, which was developed based on the intuition that minimising the negative impact is equivalent to improving the positive influence. The objective of EBB is to turn the "worst enemy" into the "best friend". EBB is a full-feedback model, since it does not rank the users on the basis of any node features, but identifies seeds by looking at the convergent status of the network. In other words, the negative user with the highest influential capabilities is selected as a seed, so that the most negative influential user is converted and starts to convey positive impacts to the neighbours.

However, as the previously proposed EBB model is probabilistic-based, the convergent state of a social network can be varied in different attempts. Multiple trials may reduce such uncertainty of social-network behaviours, but the individual's actions are not captured. Furthermore, EBB focuses on the final state of a social network only, while the evolutionary process is ignored. To overcome the issues mentioned above, the preliminary EBB algorithm has been revised by proposing 2E2B algorithm. 2E2B algorithm tracks the tendency of turning negative for all the agents during the entire influence-diffusion process, where v_j's tendency of turning negative, i.e., $\Psi(v_j)$, is formulated in Equation 3.12.

$$\Psi(v_j) = \frac{1}{m}\sum_{t=1}^{m}\psi(v_j,t)\cdot csi_{jx}(s_{jx}^t) \tag{3.12}$$

During the influence diffusion process, v_j's tendency of turning negative is increased if it becomes or retains NA in any time step. $\psi(v_j,t)$ denotes a characteristic function, which is formulated in Equation 3.13.

$$\psi(v_j,t) = \begin{cases} 1, & if\ s_{jx}^t = NA \\ 0, & otherwise \end{cases} \tag{3.13}$$

The detailed description of 2E2B algorithm by considering agents' tendency of turning negative is described by Algorithm 2. The inputs incorporate social network graph $G = (V,E)$, seed set size k and a hypothesis item i_x, while the output is the selected seed set A, $|A| = k$. Lines 2 - 10 comprise the main body of 2E2B; in each iteration, the user with maximum $\Psi(.)$ is selected. Meanwhile, $\Psi(.)$ of any user is obtained based on the observations of the previous networked evolutionary process.

Algorithm 2 2E2B Algorithm

Input: $G = (V, E), k, i_x$
Output: A

Initialise $A := \varnothing$
for $i = 1$ to k **do**
  Set seeds A into G
  Network evolves for m time steps
  for $\forall v_j \in V \setminus A \wedge \overline{s_{jx}} \neq NA$ **do**
    Calculate $\Psi(v_j)$ using Equations 3.12 and 3.13
  end for
  Select v_x, $\Psi(v_x)$ is maximum
  $A := A \cup \{v_x\}$
end for

3.6 Experiment and Analysis

I have conducted two experiments to evaluate IDMAS and 2E2B algorithm. In Experiment 1, I simulate influence diffusion in a social network by facilitating IDMAS. Through the simulation, I attempt to explore the of influence, as well as the long-term trend of the network status driven by the diffusion. Experiment 2 aims to evaluate the performance of 2E2B algorithm by comparing with some state-of-the-art influence maximization algorithms.

3.6.1 Experiment Setup

Dataset. The experiments are conducted by using the MovieLens dataset [2], which is a stable benchmark dataset incorporating approximate one million ratings (range from 1 to 5) for 3,900 movies given by 6,040 users (Harper & Konstan, 2015). There are no explicit links among the users, but implicit links can be generated based on the users' ratings of items. For simplification, 500 users have been selected randomly for the experiments.

[2]http://grouplens.org/datasets/movielens/

System Setup. A social context is simulated by creating user agents based on the public datasets. Each user agent manages its own local information, such as friendship affiliations, personalised features and behaviours. Meanwhile, a reporting agent is initialized for monitoring the entire multi-agent system and collecting global information.

There are three states in the process of IDMAS, which are listed below.

- **Initialisation:** User agents are initialised, including their neighbours, historical ratings, CKR and a hypothesis item. Agents initiate training to obtain the personalised parameter. The selected seeds are set into the environment.
- **Evolution:** All the user agents conduct social behaviours with a number of time steps, which drives the evolution of the entire system.
- **Termination:** The evolutionary process stops, and the outcome can be observed at this point.

By setting up the system, the parameters for the experiments are given in Table 3.1. One of the fundamental tasks is to obtain the user-rating matrix based on the existing ratings. As aforementioned, PMF is utilised for user-item rating matrix prediction and personalised parameter training; thus, the number of latent features has to be optimised to improve the quality of PMF prediction. Assume an interval of $[0, 25]$ for latent features count. By inspecting both *Mean Absolute Error (MAE)* and *Root Mean Squared Error (RMSE)* in Figure 3.6, the latent features count is assigned as 15, which gives the minimum MAE and RMSE. Based on the predicted user-rating matrix, training can be initiated by all the agents. I further inspect individuals' personalised parameter distribution of the current environment. In Figure 3.7, it can be observed that most agents' personalised parameters are lower than 0.4. The result implies that the individuals in this social network tend to be influenced by their social context much

Table 3.1: Experiment Parameters

Attribute / Parameter	**Value**
Nodes	500
Links	5838
Average degree	12.01
Graph density	0.025
Total time step m (Experiment 1)	1000
Total time step m (Experiment 2)	100
Evolution trials	100
α	0.7
Learning rate (PMF)	0.0001
Regularisation strength (PMF)	0.1
Latent Features Count (PMF)	15

more than that of retaining their own opinions. This is because the influence network is constructed based on the preference similarities among the individuals.

Next, a hypothesis item is selected. The distribution of users' PCL towards the hypothesis item has been demonstrated in Figure 3.8. It reveals that very few users show high appreciations towards the hypothesis item initially, whereas, a large fraction of the users may initially decline the adoptions of the hypothesis item. As explained in Definition 5, a user's initial preference state is assigned as PA if the corresponding PCL exceeds the *PA threshold*, i.e., θ_{PA}. Likewise, a user appears NA if its PCL falls below the *NA threshold*, i.e., θ_{NA}.

Experimental Scenarios. By varying the values of both θ_{PA} and θ_{NA}, different scenarios can be generated. I choose three typical experimental scenarios listed in Table 3.2. In Scenario 1, there are more initially existing PA users than that of NA users. Scenario 2 demonstrates an opposite situation, where the pre-existing NA users exceed the PA. Scenario 3 shows an extreme case that all the users have already possessed a clear attitude towards the hypothesis item, i.e., either PA or NA.

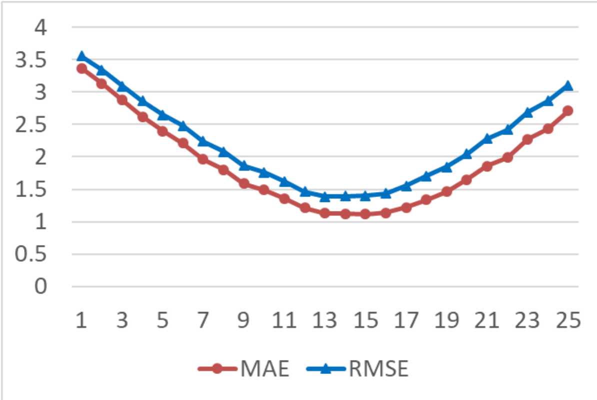


Figure 3.6: Number of Latent Features

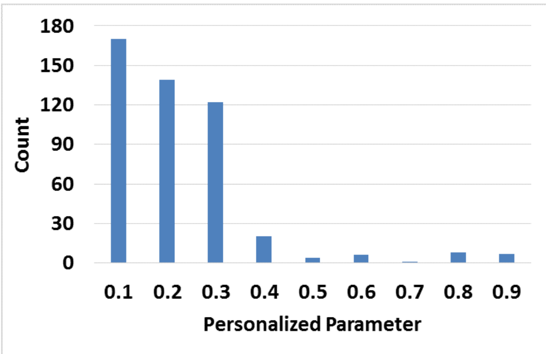


Figure 3.7: Personalised Parameter Distribution

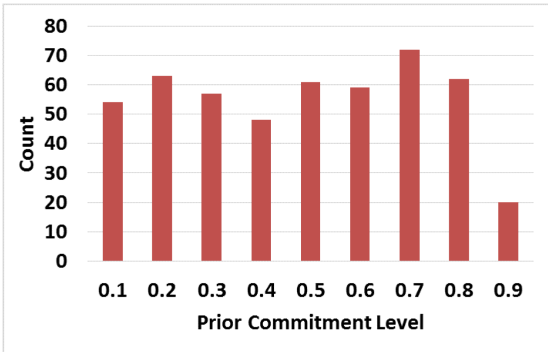


Figure 3.8: PCL Distribution

Table 3.2: Experimental Scenarios

Scenario	Parameter	Value
Scenario 1	θ_{PA}	0.7
	θ_{NA}	0.1
	Initial PA users	82
	Initial NA users	52
Scenario 2	θ_{PA}	0.8
	θ_{NA}	0.2
	Initial PA users	20
	Initial NA users	115
Scenario 3	θ_{PA}	0.5
	θ_{NA}	0.5
	Initial PA users	217
	Initial NA users	283

Performance Evaluation Metrics. To evaluate the performance of various seed-selection algorithms in IDMAS, CIACD formulated in Equation 3.10 is utilised as the evaluation metrics. Meanwhile, the activation coverage has been leveraged for exploring the details of a networked evolutionary process, where the activation coverage refers to the cardinality of the activated users.

3.6.2 Experiment 1

Experiment 1 aims to simulate influence diffusion in social networks by adopting IDMAS under Scenario 1 (refer to Table 3.2). The simulations are conducted without selecting any influential users (investment). There are two objectives in Experiment 1: **(1)** To explore the complex nature of influence based on the varying the factors. **(2)** To investigate the long-term networked trend driven by influences.

IDMAS focuses on the individual's personalised traits and behaviours, and the opinion formation/revision of each individual depends on both PCL and CSI. First, I assume that users are not supposed to revise their PCL over time. In other words, their decisions of opinion revision in any time step do not affect the intrinsic attitude towards the hypothesis item at all. Given a fixed time interval, i.e., 1000 time steps, as we can observe from Figure 3.9 that IA users have been converted rapidly, whereas, the activation coverage of both PA and NA oscillates significantly all along the way. In this case, the social network is not capable of reaching a convergent state.

Second, agents are enabled to update their PCL according to any of their decisions:

$$pcl'_{jx} = pcl_{jx} + \chi_j, \ subject\ to\ 1 - pcl_{jx} \geq \chi_j \geq -pcl_{jx}, \tag{3.14}$$

where pcl'_{jx} signifies the PCL of v_j towards the hypothesis item i_x in the coming time step, and χ_j denotes the PCL changing the rating of user v_j. If v_j holds a positive attitude towards the hypothesis item at any time step, thus $1-pcl_{jx} \geq \chi_j > 0$. On the contrary, if v_j

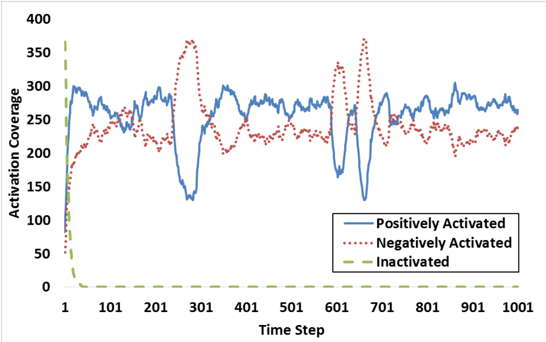


Figure 3.9: NO Updating PCL

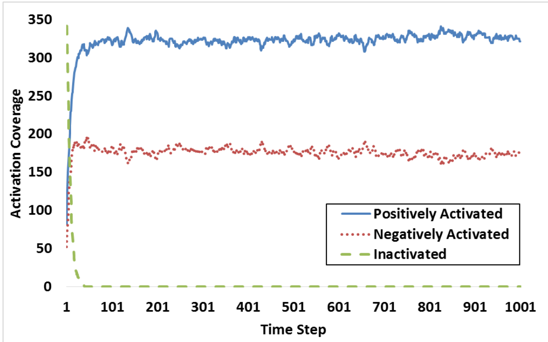


Figure 3.10: Updating PCL

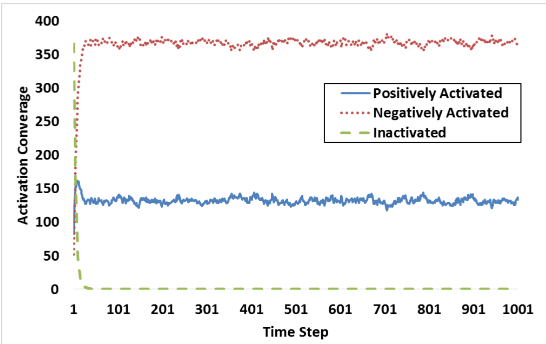


Figure 3.11: CSI Only

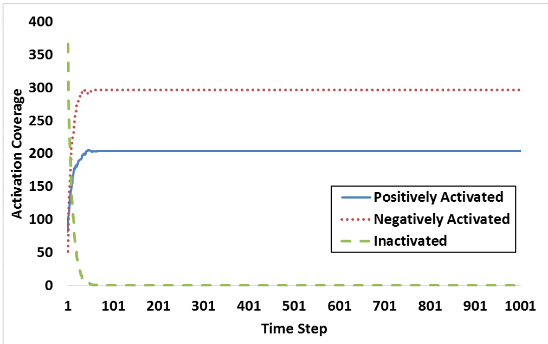


Figure 3.12: PCL Only

shows a negative opinion, then $0 > \chi_j \geq -pcl_{jx}$. Under such a setting, the convergence of both PA and NA can be observed in Figure 3.10, where the positive opinion dominates the social network. Based on the significant feature of influence that the diffusion process eventually leads to global convergence or polarisation (Centola et al., 2007; Sunstein, 2002; Z. Li & Tang, 2015; Axelrod, 1997), the situation demonstrated in Figure 3.10 appears more reasonable. Therefore, I can infer that the PCL tends to be an essential factor for influence propagation and users tend to revise PCL over time. Third, I assume agents ignore PCL when evaluating their activation tendency, and the trend is shown in Figure 3.11. The network converges rapidly but demonstrates an entirely different result from Figure 3.10 that the population of NA users far exceeds that of the PA. Furthermore, as illustrated in Figure 3.12, without the impact of CSI, most users in the social network appear negative.

To inspect the reason, I estimate both PA and NA strength from the initially activated users by considering the PCL only, using the formulas below:

$$PCL_{PA}^{i_x} = \sum\nolimits_{v_j \in V \wedge \overline{s_{jx}} = PA} |v_j| \cdot \lambda_j \cdot pcl_{jx} \tag{3.15}$$

$$PCL_{NA}^{i_x} = \sum\nolimits_{v_j \in V \wedge \overline{s_{jx}} = NA} |v_j| \cdot \lambda_j \cdot (1 - pcl_{jx}) \tag{3.16}$$

The results of the current setting can be obtained: $PCL_{PA}^{i_x} = 108.1$ and $PCL_{NA}^{i_x} = 52.7$. Therefore, the PCL of the pilot users plays a significant role in leading the evolutionary direction of a social network, e.g., if the positive strength of PCL exceeds the negative, the entire social network can be positive, even of most users have an NA tendency intrinsically.

Based on the aforementioned intriguing discoveries, IDMAS demonstrates its rationality of being facilitated to analyse the complex influence diffusion in social networks and the capability of capturing and predicting the long-term evolutionary trend.

3.6.3 Experiment 2

The second experiment aims to evaluate the performance of 2E2B algorithm in the influence maximization problem under IDMAS, where both efficiency (running time) and effectiveness (CIACD), formulated in Equation 3.10) are considered. Three different scenarios listed in Table 3.2 have been applied for the evaluation. Another five classic seed-selection algorithms of both non-feedback and full-feedback models as follows are involved as the counterparts.

- *Random Selection:* Select each seed randomly, so that the seed set grows incrementally.

- *Degree Ranking Selection:* Rank the users based on out-degree, i.e., the size of

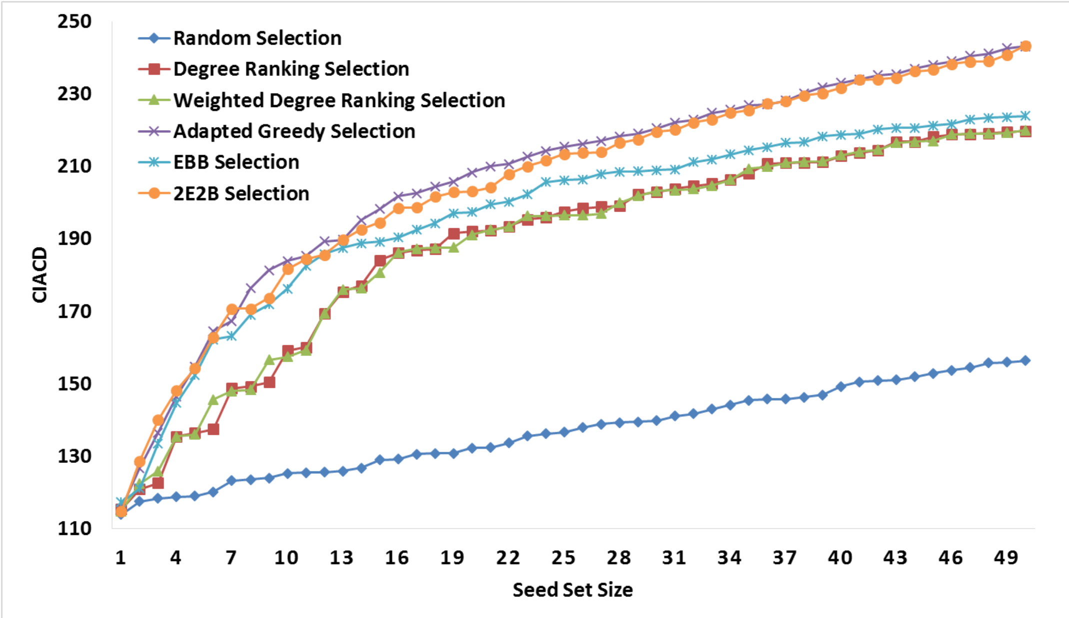


Figure 3.13: Performance Evaluation under Scenario 1

reachable users.

- *Weighted Degree Ranking Selection:* Rank the users based on the cumulative IPP to the neighbours.
- *Adapted Greedy Selection:* Obtain the maximum influence marginal gain in selecting each seed whose initial state is not NA.
- *EBB Selection:* Select the negative user whose influential capability is the highest in each networked evolutionary trial.
- *2E2B Selection:* Select the user with the highest negative tendency and influential capability in each networked evolutionary trial.

Effectiveness Comparison

In Experiment 2, I first compare the effectiveness of the above algorithms in the three different scenarios by using IDMAS. The comparison results are demonstrated in

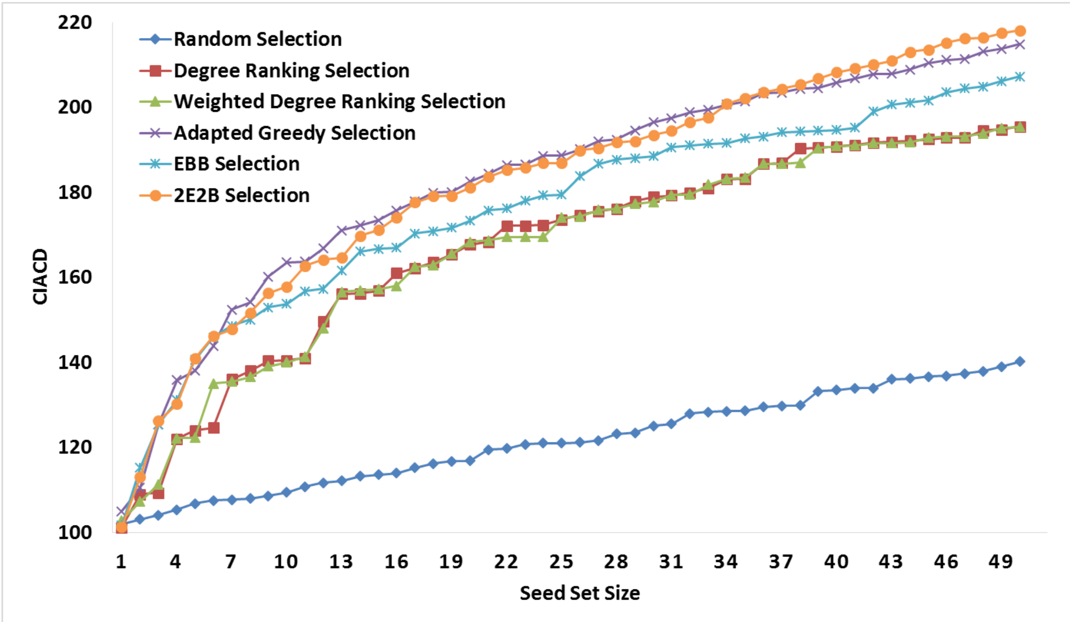


Figure 3.14: Performance Evaluation under Scenario 2

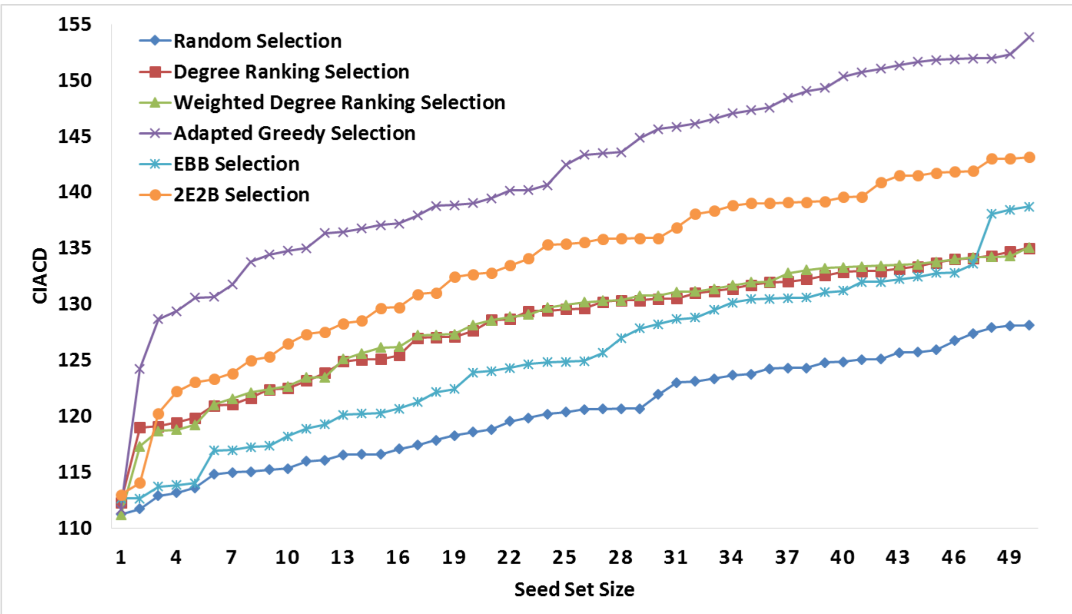


Figure 3.15: Performance Evaluation under Scenario 3

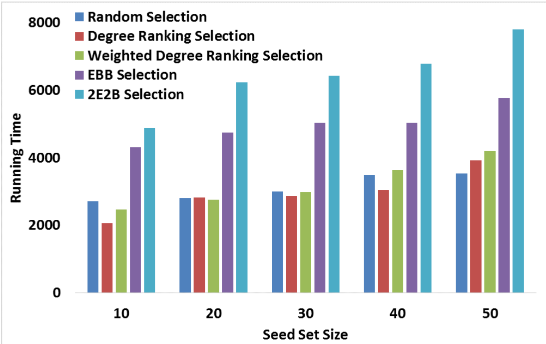


Figure 3.16: Efficiency Comparison 1

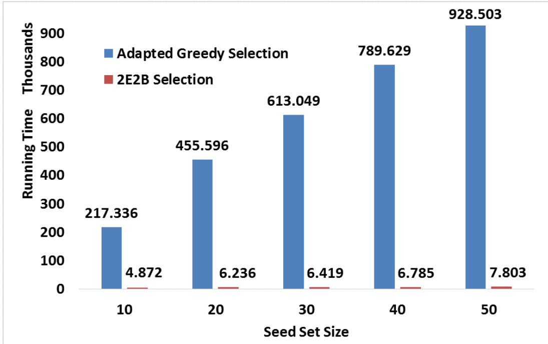


Figure 3.17: Efficiency Comparison 2

Figures 3.13, 3.14 and 3.15. In the figures, x-axis denotes the seed set size, and the y-axis represents the effectiveness, i.e., the value of CIACD. We could observe from these three figures that the full-feedback models (EBB, 2E2B and Adapted Greedy selection) generally outperforms the non-feedback models (random, degree ranking selection).

In Figures 3.13 and 3.14, it can be seen that 2E2B performs almost the same as that of the adapted greedy selection. 2E2B demonstrates its advantages especially when the seed set size enlarges over 30. This implies that with the increment of seed set, the negative impact has been suppressed and converted to the positive influence in the social network to a large extent; thus, CIACD presents significant upward trend. EBB also gives a relatively promising performance, though yields to 2E2B and adapted greedy selection. As for the two degree ranking selections, their performances are incredibly close to each other, but a bit far from that of 2E2B. As expected, the random selection gives the worst performance, since it is not based on any heuristics.

In Figure 3.15, an extreme scenario is presented, where all the users already possess a positive or negative attitude towards the hypothesis item. Under such a situation, 2E2B algorithm does not give a promising performance compared with the adapted greedy selection, but still outperforms the others. On the other side, the network almost reaches a convergent status before the simulation starts. It is very tough to change the situation, and minor increment of CIACD can be observed compared with the other two scenarios, i.e., the maximum rise of CIACD is only 45 when expanding the seed set size up to 50. From a business perspective, this phenomenon explicitly shows that the investment in a saturated market is not cost-effective, and strategies should be built before the marketing maturity.

Efficiency Comparison

Furthermore, the efficiency of 2E2B algorithm is evaluated by comparing the running time against other seed selection algorithms, which is demonstrated in Figures 3.16 and 3.17. By increasing the seed set size, we can observe that the running time required for 2E2B does not vary a lot. The running time of 2E2B is manageable, though it is less efficient compared with those none-feedback models (i.e., random selection, degree ranking selection and weighted degree ranking selection). In Figure 3.17, it is evident that the running time of the adapted greedy selection far exceeds that of 2E2B when selecting same amount of seeds, and increases dramatically with the enlarging of seed set size. Moreover, the adapted greedy selection is not scalable for large social networks, but 2E2B algorithm can be applied.

According to the results obtained from Experiment 2, it can be seen that 2E2B outperforms the other classic algorithms by considering both effectiveness and efficiency. As 2E2B algorithm is based on IDMAS, many individual features can be captured and utilised in seed selections. Namely, the advantages of 2E2B cannot be achieved without ABM. Hence, I claim that the proposed IDMAS is capable of producing a certain range of dynamical behaviours based on different parameter constellation and suitable for analysing and modelling the real-world complex influence diffusion and tracking the long-term trend of social networks.

3.7 Summary

In this chapter, a generic HSN is proposed to model the channels of influence propagation. I articulated the multi-faceted nature of influence, introduced the decomposition of influence effects and defined direct/indirect influence. Based on the key concepts derived from HSN, I further systematically modelled preference-based complex influence propagation by using multi-agent systems.

The information dissemination in social networks has been modelled as an evolutionary process driven by individuals' actions. The proposed agent-based model is capable of alleviating the issues raised by the complex nature of influence diffusion. Instead of obtaining the global view of the entire social network, I focus on user's personalised traits, preferences, behaviours and other factors, affecting the influence acceptance towards a hypothesis product. Moreover, a limited global view has been facilitated in the proposed model, i.e., CKR, which enables each agent to conduct training to learn its personalised parameter, representing the stubbornness degree. Another outstanding feature of the agent-based model is that it is capable of tracking a long-term evolutionary trend of social networks driven by influence, and handling the situations when group opinions revise according to the changing context. On the other side, a novel seed selection algorithm is developed based on the proposed model, namely, 2E2B selection. The experimental results prove that 2E2B outperforms the state-of-the-art approaches in a dynamic environment.

The agent-based framework proposed in this chapter serves as a basis for the remaining chapters, where I further extended the generic agent-based model of influence propagation to tackle in-depth issues in this field.

The related works of this chapter have been published in (W. Li, Bai & Zhang, 2016b), (W. Li, Bai & Zhang, 2016a) and (W. Li, Bai & Zhang, 2018a).

Chapter 4

SIMiner: A Stigmergy-based Model for Mining Influential Nodes

Based on the challenging issues discussed in Chapter 1, an agent-based approach is utilised to systematically model the influence diffusion for large-scale and dynamic networks without a global view. In this chapter, a collective intelligence model, i.e., stigmergy-based influencers miner, is proposed to explore influential nodes in a fully dynamic environment. The proposed model is capable of analysing influential relationships in a social network in decentralised manners and identifying the influencers more efficiently than traditional seed selection algorithms. Moreover, it is capable of adapting the solutions in complex dynamic environments without any interruptions or recalculations. Experimental results show that the proposed model achieves better performance than other traditional models in both static and dynamic social networks by considering both efficiency and effectiveness.

4.1 Overview

The major challenging issues of influence maximization problem stem from two intrinsic characteristics of the social networks, i.e., large scale and dynamics, and both often come up together in reality. Most social networks possess a vast number of nodes and links, as well as evolving topological structures. Nodes keep growing; links form and vanish; link strengths revise over time. To handle the dynamics of influence maximization, when the network topological structure changes, traditional seed selection approaches are supposed to recalibrate the solutions by running over again, which is computationally expensive. Moreover, using numerous static snapshots of a social network is a typical representation of dynamics but becomes unrealistic in a large-scale and continuously evolving environment. It is obvious that capturing snapshots in seconds in a big-data environment will inevitably create another set of big data, making the existing problem more complicated. On the other side, almost all the diffusion models require a global view of the social network for influence-diffusion simulations. However, in many applications, local information is available merely, e.g., the practitioners intervene in a population, where the observation of the social network is initially not discovered, and the data is required to be collected via a laborious process (Brautbar & Kearns, 2010; Borgs, Brautbar, Chayes, Khanna & Lucier, 2012). Therefore, by considering both features, it is extremely difficult to analyse and mine influential users by leveraging classic diffusion models and seeding algorithms.

Collective intelligence approach contributes to the shift of knowledge from individual to the collective, which appears more competent in handling the large-scale, dynamic and distributed environment. It disrupts the limitations and handles dynamics at the microscopic level using agents. Specifically, individuals revert the latest information by exploring the possible solution in a dynamic context through communications; thus, the solution can be adapted over time based on the evolutionary environment.

Furthermore, the decentralised model cuts the computational cost by sharing the workload and distributing the tasks to individuals, so as can be operated without a global view. Therefore, such approaches appear applicable in many applications domains with undiscovered global views, e.g., WeChat business (Lien & Cao, 2014). There are basically two major types of decentralised models regarding communications. One relies on direct communications among the individuals, such as cellular automata (Shiffman et al., 2012), where each cell in the grid adapts its state by looking at the adjacent neighbours based on a set of rules. The other focuses on indirect communications by reading or analysing the messages left by the peers. Stigmergy-based models (Dorigo et al., 2000) are a typical type of decentralised models applicable to the second category.

Stigmergy is defined as "stimulation of workers by the performance they have achieved" (Bonabeau, 1999), which is a particular indirect communication mechanism exhibited by tiny social insects, such as ants, to coordinate group activities. Their indirect communications are normally conducted through leaving a chemical substance, i.e., pheromones, on the trails, which evaporate over time. Inspired by the stigmergic interactions, the stigmergy and ant algorithms have been widely applied in many applications without global information, such as communication network routing, exploratory data analysis and diagram drawing, where the intelligent agents cooperate with each other by leaving and sensing the artificial pheromone, which indicates application-specific information (Mostafa et al., 2014).

In this chapter, I propose a collective intelligence model called Stigmergy-based Influencers Miner (SIMiner), which is able to analyse and mine influential nodes in dynamic social networks in a decentralised manner. The proposed model is applicable in both static and dynamic complex environments, and capable of adapting the solutions in an evolving context. SIMiner is a Multi-Agent System (MAS) (Ye et al., 2016), which incorporates two types of agents, i.e., user agents and ant agents. The former provide information actively for ant agents when requested. While the latter traverse the network based on the heuristics, seeking for the influencers in a social network. Ant agents' key behaviours, including tour formation and pheromone allocation, have been modelled for selecting appropriate influential nodes to achieve the maximum positive impact. The pheromone evaporation, another factor affecting the pheromone deposition, is also formulated in SIMiner. Specifically, tour formation refers to how ants walk and form a tour in the environment, and the latter aims to distribute pheromone to specific nodes based on the results of a local influence diffusion model. While pheromone evaporation is an exploration mechanism that delays faster convergence of the solutions. Empirical experiments have been conducted to analyse the convergence and evaluate the performance of SIMiner against other classic models under both static and dynamic social networks. The experimental results reveal that the proposed model can converge to an optimal solution gradually and function perfectly in a distributed environment without a global view. It outperforms state-of-the-art approaches by considering both influence efficiency and effectiveness. Moreover, SIMiner is able to mine influencers in dynamic environments without any interruption or recalculation, and the solution can be adapted quickly over time.

The remainder of this chapter is structured as follows. Section 4.2 reviews the literature related to this research work. Section 4.3 introduces the preliminaries and gives formal definitions. Section 4.4 systematically elaborates SIMiner model. Theoretical analysis of convergence is conducted in Section 4.5. In Section 4.6, experiments and experimental results are presented to evaluate the performance of SIMiner. The summary of this chapter is given in Section 4.7.

4.2 Related Work

4.2.1 Influence Maximization

In on-line marketing, it is very important to investigate how to propagate positive influence in a social network with limited resources. Motivated by this background, Kempe et al. (2003) formulate the influence maximization as a discrete optimization problem.

Recall that in Section 2.2, several popular seed selection approaches for the influence maximization are reviewed, such as greedy selection, degree ranking selection and random selection (Kempe et al., 2003). Many studies have been conducted to improve the efficiency and effectiveness of seed selection algorithms. However, most approaches can neither handle the dynamics of social networks nor function without a global view.

4.2.2 Dynamic Influence Maximization

How to handle influence propagation in dynamically temporal social networks also has drawn attention to some researchers (Holme, 2015; B. Wang et al., 2017; Song et al., 2017). In 2.2.3, I reviewed the contemporary studies of dynamic influence maximization. Different from most research works in this field, the collective intelligence approach proposed in this chapter only concentrates on modelling the features and behaviours of agents and can handle the topologically and temporally dynamic network automatically. Thus, the proposed model possesses excellent adaptation capability to explore solutions in a changing environment.

4.2.3 Ant and Stigmergy Algorithm

Ant and stigmergy-based algorithms do not necessarily require global network information, and the computation is decentralised. Stigmergy relies on the ant colony knowledge, as it demonstrates a particular mechanism exploited for indirect communication among ants to control and coordinate their tasks. In natural environments, stigmergy-based systems have shown that they can be utilised for generating complicated and robust behaviours in the systems even if each ant has limited intelligence.

Some researchers have applied stigmergic interactions in the computer science field. The relevant studies have been reviewed in Subsection 2.3.2. However, the stigmergy-based algorithm is not fully utilised in the influence maximization problem though it demonstrates its superiority in handling optimization problems in a distributed manner.

4.3 Preliminaries and Formal Definitions

4.3.1 Fully Dynamics of Social Networks

Modelling emergent properties of social networks appears to be one of the pillars of social network science (Holme, 2015). Real-world social networks possess a highly dynamic nature and evolve rapidly over time (R. Kumar et al., 2010; Leskovec, Kleinberg & Faloutsos, 2007). More importantly, the network evolution is continuous. Nearly all of the approaches in this research field utilise numerous static snapshots of consecutive discrete time steps to mimic the dynamics of social networks. For example, a fully dynamic social network can be represented in Figure 4.1. As shown in Figure 4.1, a social network evolves and updates over time, i.e., from time steps t_1 to t_n, incorporating the addition and deletion of nodes and links, as well as the weight updates.

However, for large-scale networks, it is almost impossible to imitate dynamics by capturing snapshots or storing all changes happening around since this inevitably creates another set of big data. Furthermore, capturing snapshots or changes require a global view, which becomes another obstacle as a central component is required for monitoring the entire network in real-time.

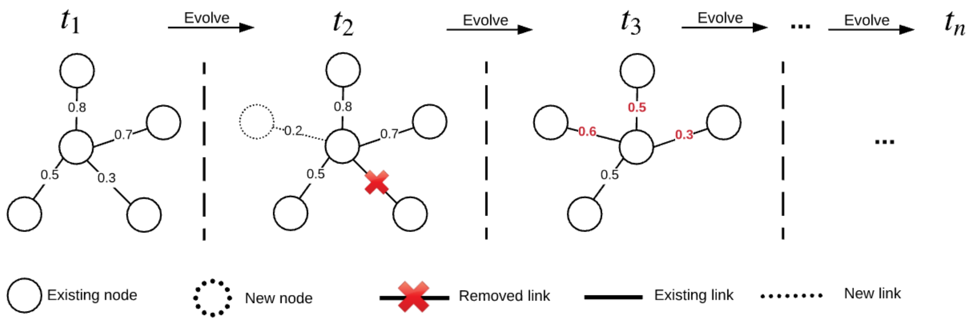


Figure 4.1: Fully Dynamic Social Network

To overcome the difficulties mentioned above, I argue that compared with other existing approaches, the stigmergy-based multi-agent system is more suitable for capturing the dynamic behaviours of social networks since individual agents have been deployed to explore the real-time changes in a local level from a microscopic point of view. This approach functions even without a global view.

4.3.2 Stigmergy and Multi-Agent Systems

In general, stigmergy refers to a series of behaviours initiated by simple animals, such as ants, termites and wasps, which is a collection of mechanisms that mediate the interactions among these animals (Lewis, 2013; Theraulaz & Bonabeau, 1999). The communications among the individuals are mediated by a sort of biological substance, i.e., pheromone. By borrowing the ideas from this biological phenomenon, many researchers model the ants as autonomous and self-directed agents (Dorigo et al., 2000; Mostafa et al., 2014; Takahashi et al., 2012). All of the agents work in the same context and form a MAS. The agents exchange messages indirectly. Specifically, they keep modifying the global environment by leaving certain amounts of pheromone trails based on their local experiences and discoveries. Meanwhile, the agents select walking paths according to the pheromone concentration and a partial network topological structure covered in the local view at a particular time frame. Therefore, the pheromone concentration and distribution represent the problem that the agents are currently working on (Hadeli et al., 2004). From a microscopic perspective, the MAS demonstrates an evolutionary pattern driven by the local behaviours of individual agents.

In this chapter, the social network is modelled as the dynamic context/environment of SIMiner, where two types of agents reside in the same context, i.e., *user agents* and *ant agents*. The former represent users or nodes in a social network, while the latter refer to the agents for analysing users' influential degrees. Each user agent possesses a local view covering the neighbourhood. Similarly, ant agents can move among user agents and sense the environment at a local level. Both types of agents are capable of interacting with each other, while ant agents can only communicate with their species indirectly. On the other side, the influence propagation process is simulated as the crawling behaviours of ant agents. The objective of the ant agents is to investigate the possible influential users from the social network. To explain this further, the ant agents keep moving through the network and allocate different amounts of pheromone on each node that they claw over as rewards. The pheromone rewarding strategy is on the basis of a local influence propagation model. The users with higher influential abilities will gain more pheromone and become prominent after a number of iterations.

In SIMiner, multiple ant agents keep crawling in a dynamic context simultaneously and iteratively. Figure 4.2 illustrates an overall picture of SIMiner, implying how SIMiner handles the dynamics from a microscopic level and in a distributed environment. Three ants are presented in this figure. Each ant captures the environment within the coverage of its local view when it reaches any node at a particular time step. Any environmental updates prior to the arriving of ants are not supposed to be discovered, but the modifications will be spotted when they are within the local view of any ants. From a microscopic viewpoint, ants do not care about the changes of network topological structures; instead, they concentrate on local time-spatial observations.

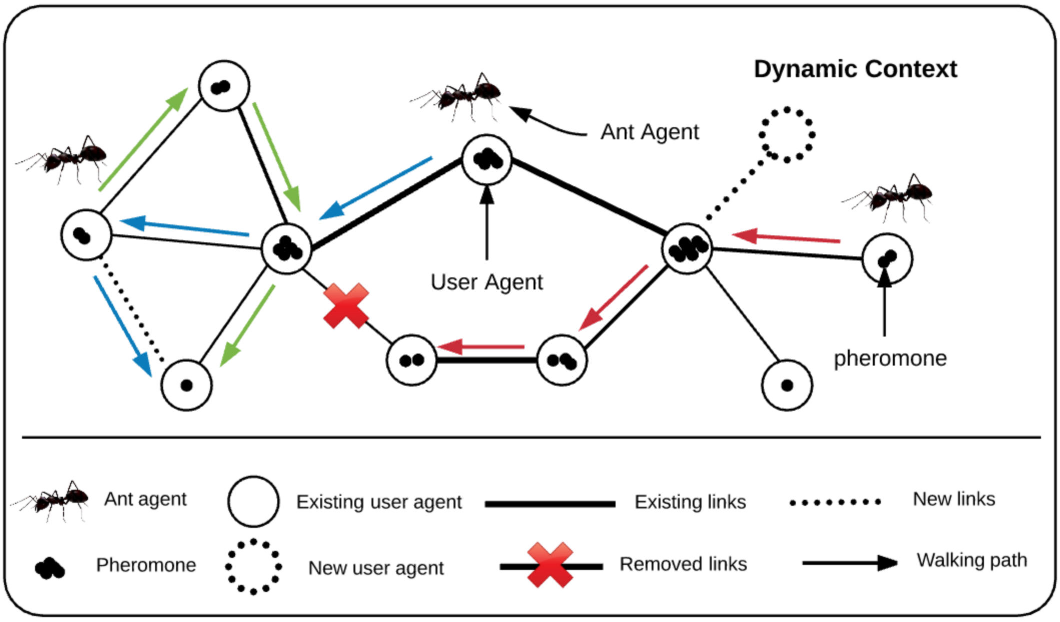


Figure 4.2: Stigmergy-based Influencers Miner Model

Therefore, SIMiner neither relies on a global view nor proactively captures global network snapshots but is capable of handling the dynamics locally, which explicitly simplifies the problem significantly.

4.3.3 Influence Diffusion Models

Both the IC model and the LT model can be facilitated to simulate the diffusion process when given a static network with a global topological structure. However, they cannot function only with local views; both require a substantial extension to handle the temporal features of a social network (Tong et al., 2017; Gayraud et al., 2015; Karimi & Holme, 2013). Moreover, the outcome of both traditional models merely demonstrates a global activation coverage and ignores the contribution of each individual.

In SIMiner, I propose and leverage an extended version of the IC model, i.e., Independent Cascade Model with Pyramid Scheme (ICMPS), to measure the individual's activation contribution or influential capabilities in a local social network, which provides

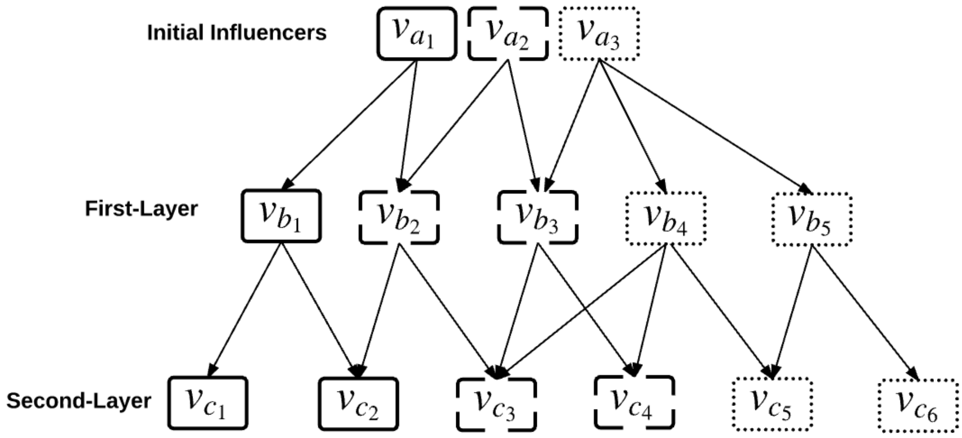


Figure 4.3: An Independent Cascade Model with Pyramid Scheme

the evidence for pheromone distribution. ICMPS inherits the key features from the IC model, where influence diffusion is demonstrated as a hopping and infecting process. Whereas ICMPS tends to capture the activation contribution of the influencers in a local environment, and the activation follows a pyramid pattern. Figure 4.3 illustrates a toy example of ICMPS in a static local network, where the three initially active users, i.e., v_{a_1}, v_{a_2} and v_{a_3}, initiate the influence effecting on the two-hop neighbourhood. As we can observe from the figure that v_{a_1} successfully activates its neighbour v_{b_1} only; both v_{b_2} and v_{b_3} are activated by v_{a_2}; v_{a_3} influences v_{b_4} and v_{b_5}. Next, the newly active nodes in the first layer attempt to exert influence on their direct neighbours, i.e., the nodes in the second layer. Therefore, activation contribution of v_{a_1}, v_{a_2} and v_{a_3} is 3 ,4 and 4, respectively. As ICMPS is a stochastic model, the results should be averaged over multiple trials.

4.3.4 Formal Definitions

In general, a social network at a particular time step k, i.e., $G(k) = (V(k), E(k))$, can be defined as a graph containing numerous of entities $V(k)$ with their connectivities $E(k)$, and it possesses an evolving and dynamic topological structure. In this research, the problem space is modelled as a distributed MAS. The environment at time step k, $Env(k) = (V(k), A)$, is considered as a shared working space of two types of agents, i.e., user agents $V(k) = \{v_1, v_2, ..., v_j\}$ and ant agents $A = \{a_1, a_2, ..., a_i\}$. In static networks, $\forall k \in \mathbb{N}$, $G(k) = G(k+1)$ and $Env(k) = Env(k+1)$. For simplification purpose, "(k)" can be omitted in a general context.

Definition 4.1: A user agent (node) v_i refers to an agent in the environment Env. $v_i.E$ denotes the friendship set/attribute of user v_i, where $v_i.E = \{e_{ij} | v_j \in V \land e_{ij} \in v_j.E, v_i \neq v_j\}$. A particular element in $v_i.E$ can be represented as a three-tuple, i.e., $e_{ij} = (v_i, v_j, w_{ij})$, where w_{ij} is the weight of e_{ij}, describing the affiliation strength, and it also can be denoted by using the notation $e_{ij}.w$. Meanwhile, any user agent v_i has a set of neighbours, i.e., $\Gamma(v_i) = \{v_j | e_{ij} \in v_i.E \cap v_j.E, v_i \neq v_j\}$. The friendship affiliation information is supposed to be maintained by each individual locally.

Definition 4.2: An ant agent a_m is defined as an autonomous agent working for a specific problem in the environment Env, which crawls across the nodes in the same environment based on users' relationships. The friendship affiliation signifies the available routes for ant agents.

There exist a number of ant agents in the environment, $A = \{a_1, a_2, ..., a_i\}$. Moreover, they are capable of communicating with user agents in order to discover and evaluate the amount of pheromone on the current user and the ones nearby, as well as examine the strength of the relationships.

Definition 4.3: A tour $T_m^n = (\overrightarrow{V_m^n}, \overrightarrow{\tau_m^n})$ is defined as the path that ant agent a_m walks through d nodes in Round n, where $\overrightarrow{V_m^n} =< v_1, v_2, ..., v_d >$ denotes a directed vector which contains the node sequence in the tour, while $\overrightarrow{\tau_m^n} =< t_1, t_2, ..., t_d >$ refers to the corresponding time when a_m passing each element in $\overrightarrow{V_m^n}$. For simplification purpose,

T is regarded as a tour in a general context.

Specifically, a_m randomly selects a starting point (a user agent). Next, it crawls from one node to currently existing adjacent neighbours at the next time step and eventually ceases when it reaches the *endpoint* v_d, where $\Gamma(v_d) \subset \{v_1, v_2, ..., v_d\}$.

Definition 4.4: N-layer Sub-network $G^n_m(N) = (V^n_m(N), E^n_m(N))$ is defined as a local static sub-graph generated by ant a_m after completing tour T^n_m. The edge set $E^n_m(N)$ includes all the links among $V^n_m(N)$, while $V^n_m(N)$ denotes the union set of the nodes in T^n_m and the corresponding N-hop neighbourhood. Thus, $V^n_m(N)$ can be formulated in Equation 4.1.

$$V^n_m(N) = \begin{cases} \{v_1, v_2, ..., v_d\}, & N = 0 \\ V^n_m(N-1) \cup \Gamma(V^n_m(N-1)), & N \neq 0 \end{cases} \tag{4.1}$$

Definition 4.5: Pheromone represents the information and heuristics passed by ant agents to their peers based on the local experiences. $p^n_m(v_i)$ represents the pheromone amount allocated to v_i in tour T^n_m. While $v_i.p(n)$ indicates the pheromone amount accumulated on user agent v_i in the n^{th} interactive walking around, and the value is constantly changing over time.

Pheromone also can be regarded as the reward granted to user agents. Specifically, each ant agent tends to allocate more pheromone on the nodes which exert high impact potentially. In other words, the more pheromone amount a user agent (node) possesses, the higher chance it becomes an influencer.

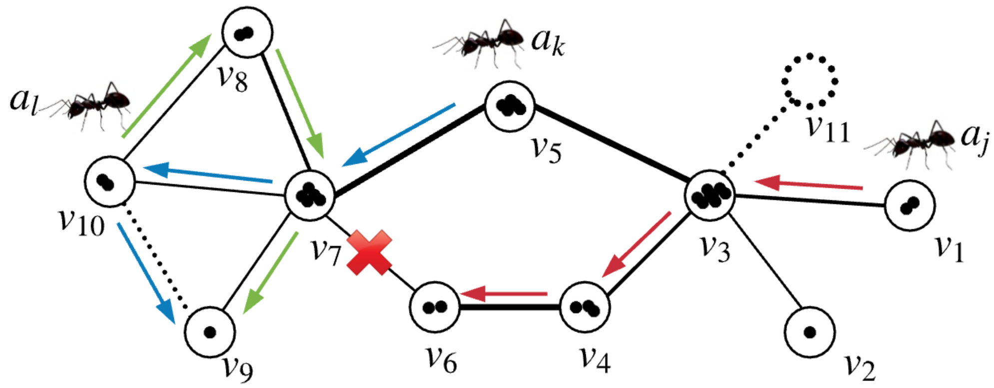


Figure 4.4: An Annotated Dynamic Social Network

4.4 Stigmergy-based Influencers Miner Modelling

SIMiner inherits several major features of the ant algorithm and agent-based modelling, which is capable of dynamically mining influencers by adapting the solutions over time. In SIMiner, numerous ant agents crawl simultaneously and update the shared environment by allocating pheromones on user agents. The influence propagation process is simulated as crawling behaviours of ant agents. The influencers can be selected based on the pheromone distribution and concentration in the network. The detailed modelling of SIMiner is elaborated in the following subsections.

4.4.1 Overall Process of SIMiner

In this subsection, the overall process of SIMiner is explained, including **(1)** how to handle the dynamic temporal and topological features of a social network, and **(2)** how to mine influencers using SIMiner.

First, the network is converted into a distributed environment shared by two types of agents, i.e., user agents and ant agents. By using the same sample dynamic network in Figure 4.2, all the agents are annotated and demonstrated in Figure 4.4. Assume that all the ant agents start tours at the same step t_1, and it takes one time step to crawl

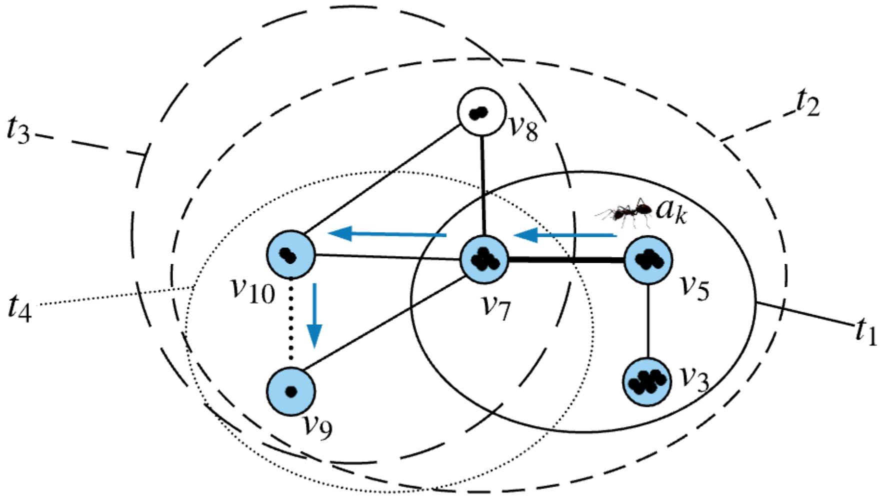


Figure 4.5: Dynamic Local View of Ant a_k based on Different Time Steps

from one node to another. As we can observe that all of the three ants complete the tour after three hops. Each ant agent possesses a dynamic local view which is driven by the changing structure of a social network, both temporally and spatially. Figure 4.5 shows the dynamic local view of ant a_k ranging from t_1 to t_4. It is obvious that each ant agent only focuses on the local context at a particular time step.

In Figure 4.4, I assume the following three events and corresponding time steps. Thus, the dynamic local view of ant agents at each time step is described in Table 4.1.

- Event 1: Link $e_{9,10}$ is formed at t_2. v_9 is not covered in ant a_l's local view at t_1, but included at t_3 and t_4.
- Event 2: Node v_{11} joins at t_3. v_8 is not spotted by a_j when it passes v_3 at t_2.
- Event 3: Link $e_{6,7}$ vanishes at t_3. Ant a_k can observe v_6 at t_2, but a_l cannot see v_6 when it reaches v_7 at t_3. Similarly, a_j arrives the end of its tour at t_4, due to the broken linkage.

Second, ant agents embark on the exploration of influencers concurrently, and they collaborate with each other by distributing pheromone on the nodes. The behaviours of

Table 4.1: Ant Agents' Local View Overtime

	t_1	t_2	t_3	t_4
a_j	v_1, v_3	v_3, v_5, v_4, v_1, v_2	v_4, v_6, v_3	v_6, v_4
a_k	v_5, v_7, v_3	$v_7, v_8, v_{10}, v_9, v_5$	v_{10}, v_8, v_7, v_9	v_9, v_{10}, v_7
a_l	v_{10}, v_8	v_8, v_{10}, v_7	$v_7, v_8, v_{10}, v_9, v_6, v_5$	v_9, v_{10}, v_7

an ant agent incorporate selecting paths, constructing local influence propagation sub-networks and distributing pheromones. Meanwhile, pheromone evaporation happens over time, which enables SIMiner to forget the poor choices in the past. The poor choices refer to the user agents that are seldom visited by the ant agents. When solutions start to converge, i.e., the pheromone ranking list has few changes, the influencers can be easily located purely based on the pheromone amount.

4.4.2 Tour Formation

Tour formation ascribes to a series of path selection and skipping behaviours from an ant agent.

Path Selection

Path selection is one of the ant's fundamental behaviours. An ant agent a_m needs to select the next node to crawl when having various choices $V_c = \{v_j | v_j \in \Gamma(v_i)\}$, where v_i denotes the node that a_m arrives at.

Basically, the path selection decision is based on two major aspects, incorporating the pheromone amount of v_j, i.e., $v_j.p$, and the weight of corresponding edge $e_{ij}.w$. The path selection behaviour is modelled as a probabilistic event by using Equation 4.2, where q_{ij} denotes the probability that an ant agent walks from v_i to v_j, α and β are two parameters balancing the weight of two factors.

$$q_{ij} = \begin{cases} \frac{(e_{ij}.w)^{\alpha} \cdot (v_j.p)^{\beta}}{\sum\limits_{v_x \in \Gamma(v_i) \setminus T} (e_{ix}.w)^{\alpha} \cdot (v_x.p)^{\beta}}, & e_{ij} \in v_i.E \\ 0, & e_{ij} \notin v_i.E \end{cases} \tag{4.2}$$

Furthermore, ant agents cannot choose the nodes they have walked through within the same tour, but can pass the nodes that other ant agents have walked before the current or previous round.

Skip

Walking from v_i to v_j, an ant agent can either select v_j or skip it by comparing the ratio of common neighbours between v_i and v_j. If the ratio is greater than a predefined threshold, v_j is supposed to be skipped. Subsequently, node v_j will be neither selected into the tour nor considered for pheromone allocation.

The rationale of defining "skip" behaviour can be reflected in two folds. **(1)** "Skip" is initiated based on the fact that many of the most central nodes in a social network may be clustered; thus, it is not necessary to target all of them (W. Chen et al., 2009; Boccaletti et al., 2006). For example, if two users, v_i and v_j, share many common neighbours, both are most likely influenced by each other since they own a broad range of communicational channels. The same concept has been applied in the traditional influence maximization problem. In other words, if one is selected as a seed, the other should be ignored, since selecting both cannot enlarge the global activation coverage to a large extent, and the impact generated by both may be pretty much close to choosing either of them. **(2)** "Skip" helps to tailor the ant algorithms to the influence maximization problem. Different from TSP, the identified influencers are not necessarily connected with each other or stay in the same path, but they can be scattered everywhere in the entire social network. "Skip" chops a continuous path selected by an ant agent into pieces, which satisfies the pattern of the solutions for influence maximization.

This behaviour significantly improves the quality of the selected influencers, which is explained in the Experiment section.

In SIMiner, ant agent a_m skips v_j if the common neighbour percentage comparing with previous node v_i, i.e., η_{ij}, exceeds a certain limit ω, where η_{ij} can be measured by using Equation 4.3.

$$\eta_{ij} = \frac{|\Gamma(v_i) \cap \Gamma(v_j)|}{|\Gamma(v_j)|} \tag{4.3}$$

Tour Formation Algorithm

As introduced in Definition 4.3, a tour can be represented as a sequential list of nodes and the corresponding time step of passing each node. In the shared environment, each ant agent keeps performing an iterative process: walking and selecting paths. Whereas, the actions stop when the ant agent reaches the endpoint v_d. In other words, the iterative process triggered by ant agent a_m in Round n produces a path vector, i.e., tour T_m^n.

Algorithm 3 describes the tour formation process. The inputs of this algorithm include ant agent a_m, round index n and the starting time step t_0, while the output is tour T_m^n. Line 3 shows the criteria of walking to the next node. Lines 5-10 demonstrate the targeting candidates selection, where σ is a predefined threshold to filter out those candidates with low probability. Lines 11-21 indicate the path selection process. Whereas, Lines 12-15 tend to judge whether an node should be included in the tour. The iterative walking process ends when all of the current node v_s's neighbours reside in tour T_m^n.

The complexity of Algorithm 3 is mainly determined by the loops in Line 3 and Line 5. The complexity is $O(ls)$, where l presents the length of the path and s denotes the cardinality of the neighbourhood.

Algorithm 3 Tour Formation Algorithm

Input: a_m, n, t_0
Output: $T_m^n = (\overrightarrow{V_m^n}, \overrightarrow{\tau_m^n})$

Initialise a_m and randomly select a starting point $v_s, v_s \in V$
Initialise a tour vector $\overrightarrow{V_m^n}$:=<>, a tour set $V_m^n := \varnothing$
while $\Gamma(v_s) \notin V_m^n$ **do**
    Initialise candidate list $V_c := \varnothing$
    for $\forall v_i \in \Gamma(v_s) \wedge v_i \notin V_m^n$ **do**
        Compute the probability q_{si} using Equation 4.2.
        if $q_{si} > \sigma$ **then**
            $V_c := V_c \cup \{v_i\}$
        end if
    end for
    $v_s := null$
    if $V_c \neq \varnothing$ **then**
        Choose the next node $v_n \in V_c$
        Compute η_{sn} using Equation 4.3
        if $\eta_{sn} \leq \omega$ **then**
            $\overrightarrow{V_m^n} :=< \overrightarrow{V_m^n}, v_n >$
            $\overrightarrow{\tau_m^n} :=< \overrightarrow{\tau_m^n}, t_0 >$
            $V_m^n := V_m^n \cup \{v_n\}$
        end if
        $v_s := v_n$, $t_0 := t_0 + 1$
    end if
end while

4.4.3 Pheromone Operations

Pheromone plays a significant role in the proposed model since it represents information or heuristics based on the ant agents' experiences which can be referred by their peers. In this section, pheromone operations, including evaporation, N-layer sub-network generation and allocation will be introduced in the following subsections.

Pheromone Evaporation

Pheromone evaporation is a common phenomenon, where the amount of allocated pheromone decreases over time. In stigmergy-based algorithms, this mechanism delays the faster convergence and avoids to converge to a locally optimal solution (P. Kumar & Raghavendra, 2011). In the current model, pheromone evaporation increases the randomness and encourages ant agents to select divergent walking paths. Furthermore, pheromone evaporation eliminates the poor choices in the past, so that the present influencers become more prominent.

Pheromone evaporates through each node within the scope of the whole network at the same time. At a justified time, all of the nodes in the network will evaporate a predefined unit of pheromone. The pheromone evaporation is quantified by using Equations 4.4 and 4.5, where ρ_n denotes evaporation rate, n represents the iterative time step, and a is a constant.

$$v_j.p(n+1) = v_j.p(n) \cdot \rho_n \tag{4.4}$$

$$\rho_n = 1 - \sqrt{\frac{n}{n+a}} \tag{4.5}$$

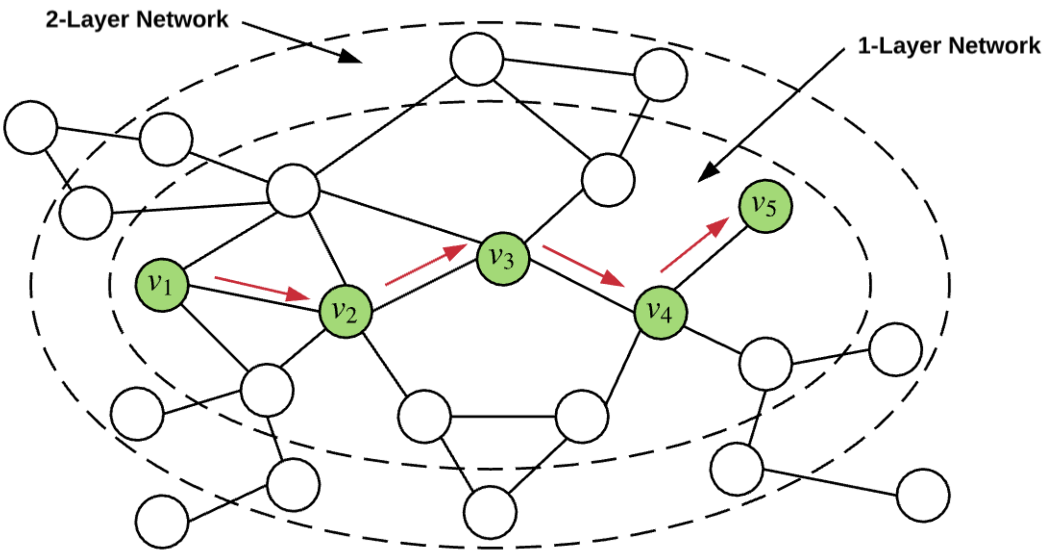


Figure 4.6: N-Layer Sub-Network

N-Layer Sub-network Generation

N-layer sub-network generation is the preliminary step of pheromone allocation, which constructs a local environment to estimate the influence contribution of each individual in a tour. "N" is natural number describing the depth of influence propagation.

Once ant agent a_m completes tour T_m^n, a corresponding local N-layer sub-network $G_m^n(N) = (V_m^n(N), E_m^n(N))$ is supposed to be generated based on the path that a_m walked through. $V_m^n(N)$ incorporates all the elements in tour T_m^n and the corresponding N-hop neighbourhood. For example, given $N = 2$, $V_m^n(2) = V_m^n(0) \cup \Gamma(V_m^n(0)) \cup \Gamma(\Gamma(V_m^n(0)))$. While the edge set $E_m^n(2)$ includes all the links among $V_m^n(2)$.

Figure 4.6 describes an example of a generated two-layer sub-network, i.e., $G_m^n(2)$. Ant a_m completes tour T_m^n in Round n, where $\overrightarrow{V_m^n} =< v_1, v_2, v_3, v_4, v_5 >$. The pheromone distribution initiated by each ant agent relies on the generated N-layer sub-network.

Pheromone Allocation

Pheromone allocation, in general, refers to how ant agents determine the amount of biological information left on the nodes that they have crawled over. From a broader point of view, the pheromone allocation unravels the rewarding strategies initiated by ant agents. Basically, ant agents tend to reward the potential influential users and update the environment by modifying the pheromone amount on each node. Thus, it demonstrates an evolutionary group work, and the solution is continuously being optimised.

The pheromone allocation is determined based on the generated N-layer sub-network; more pheromone will be allocated to the nodes in a sub-network with larger size and a path with a shorter length generally.

Next, the approach to disseminate pheromone to the environment is elaborated as follows. The pheromone amount Δq_m^n distributed by a_m to node $v_i, v_i \in T_m^n$ is proportional to the contribution of v_i in the tour. The individual's activation contribution of v_i in T_m^n, i.e., $c_m^n(v_i)$, refers to the number of nodes that have been activated by v_i in a sub-network, which can be calculated through the ICMPS model. The activation contribution of $v_i, v_i \in V_m^n(N)$, is denoted by using $c_m^n(v_i)$, and $p_m^n(v_i)$ can be formulated in Equation 4.6.

$$p_m^n(v_i) = g(v_i) \cdot \rho_n \tag{4.6}$$

In this equation, $g(v_i)$ refers to an adjustment function of v_i formulated in Equation 4.7, where D denotes a constant. $g(v_i)$ is proportional to $c_m^n(v_i)$.

$$g(v_i) = \frac{c_m^n(v_i)}{c_m^n(v_i) + D} \tag{4.7}$$

Therefore, by considering both evaporation and allocation, the pheromone update of v_i after tour T_m^n completed is described in Equation 4.8.

$$v_j.p(n+1) = \begin{cases} v_j.p(n) \cdot (1-\rho_n) + p_m^n(v_i) & , if\ v_j \in T_m^n \\ v_j.p(n) \cdot (1-\rho_n) & , otherwise \end{cases} \tag{4.8}$$

Algorithm 4 Pheromone Allocation Algorithm

Input: $T_m^n = (\overrightarrow{V_m^n}, \overrightarrow{\tau_m^n}), N$

Output: $\{v_m.p | v_m \in \overrightarrow{V_m^n}\}$

```
1: Initialise G_m^n(N) = (V_m^n(N), E_m^n(N))
2: Initialise V_m^n(N) := ∅, E_m^n(N) := ∅
3: Initialise temp variable κ = 0
4: for ∀v_i ∈ V_m^n(→) do
5:     Obtain the next element v_j from V_m^n(→)
6:     Calculate η_ij using Equation 4.3.
7:     if η_ij ≤ ω then
8:         V_m^n(N) := V_m^n(N) ∪ {v_i}
9:     end if
10: end for
11: while κ < N do
12:     for ∀v_i ∈ V_m^n(N) do
13:         E_m^n := E_m^n ∪ v_i.E
14:     end for
15:     V_m^n(N) := V_m^n(N) ∪ Γ(V_m^n(N))
16:     κ := κ + 1
17: end while
18: Initialise seed set S := V_m^n(→)
19: for ∀v_x ∈ S do
20:     v_x.activatedBy := v_x
21: end for
22: Given G_m^n(N) and S, estimate individual contribution using ICMPS
23: for ∀v_k ∈ S do
24:     for ∀v_x ∈ V_m^n(N) ∖ S do
25:         Calculate p_m^n(v_k) using Equation 4.6
26:         v_k.p := v_k.p + p_m^n(v_k)
27:     end for
28: end for
```

Algorithm 4 shows the pheromone allocation process initiated by ant a_m in tour T_m^n. The inputs of the algorithm include tour T_m^n and number of layers of sub-network N. While the output is a set of updated pheromone. Lines 1-3 initialise the environment and a temporal variable. Lines 4-10 identify the candidates for pheromone allocation by considering the threshold ω. The N-layer sub-network is constructed through Lines 11-17. While Lines 18-22 estimate the influence contribution of each node in the tour. In the end, Lines 23-28 aim to allocate the pheromone.

The complexity of Algorithm 4 is $O(n)$. More specifically, the complexity of Lines 9-15 is $O(n)$ since N is a constant. As the seed set size is limited, the complexity of Lines 21-26 is $O(n)$.

4.4.4 Mining Influential Users

Mining influential users is the last step of SIMiner, which happens when the convergence of the solution emerges. It aims to identify a limited set of influential users based on the environment modified by ants, having the pheromone with different intensity distributed. Nodes with high pheromone intensity demonstrate their important position and superior influential capability in the social network. Therefore, this group of users should be first targeted and selected as seeds.

The influential users mining algorithm applied in this context is a degree-based approach, which identifies the influential users by only considering the amount of pheromone on each node. Thus, the computational complexity is merely $O(n)$.

4.5 Theoretical Analysis

Given a random process, i.e., $X_n = \{P(n), E(n), T(n)\}, n \in N$, the status space S is defined over a finite set of discrete decision variables $X_i, i \in \mathbb{N}$. n represents the a particular time step; $P(n)$ refers to a collection of the pheromone distribution across the entire network, where $P(n) = \{v_j.p(n)|v_j \in V\}$. $E(n)$ indicates the edge set, and $T(n)$ denotes a set of tours generated by all the ants in the $Env(n)$, where $T(n) = \{T_m^n|a_m \in A\}$.

At time step n of a dynamic network, ant a_m's transfer probability $Q(n)$ merely depends on $P(n-1)$ and $E(n)$. Whereas, tour $T(n)$ relies on $T(n-1)$ and $Q(n)$. While $P(n)$ can be determined by Equation 4.8. Therefore, X_{n+1} is only affected by X_n, which explicitly shows a typical Markov process.

Let $B = \{(v_i, v_j)|v_i, v_j \in T(n), v_i \in \Gamma(v_j)\}$, and v_i is followed by v_j in $T(n)$. Then, $\forall a, b \in S$, the transfer probability $Q(n)$ satisfies:

$$Q(n) = Q_n(X_{n+1} = b|X_n = a) = \prod_{(v_i,v_j)\in B} q_{ij}(n) \tag{4.9}$$

Thus, $Q(n)$ depends on n only. The Markov process is a non-homogeneous Markov chain.

Theorem 1. *Given $N \in \mathbb{N}^+$, $\forall n \geq N$, if $\exists q_{min}(n) > 0$, having $v_j.p(n) \geq q_{min}(n) > 0$ and $\sum_{i=1}^{\infty} \rho_i = \infty$, then the non-homogeneous Markov process $X(n) = \{P(n), E(n), T(n)\}$ converges to the optimal solution with a probability of 1.0.*

Proof. $\forall e_{ij} \in v_i.E$, $\exists e_{min}, e_{max}$, having $0 < e_{min} \leq e_{ij} \leq e_{max} \leq 1$. Then:

$$\begin{aligned} q_{ij} &= \frac{(e_{ij}.w)^{\alpha} \cdot (v_j.p)^{\beta}}{\sum\limits_{v_x \in \Gamma(v_i)\setminus T} (e_{ix}.w)^{\alpha} \cdot (v_x.p)^{\beta}} \\ &\geq \frac{(e_{min})^{\alpha} \cdot (v_j.p)^{\beta}}{\sum\limits_{v_x \in \Gamma(v_i)\setminus T} (e_{max})^{\alpha} \cdot (v_x.p)^{\beta}} \\ &\geq \frac{(e_{min})^{\alpha} \cdot (v_j.p)^{\beta}}{\sum\limits_{v_x \in \Gamma(v_i)\setminus T} (v_x.p)^{\beta}} = q_{min} \end{aligned}$$

According to Equation 4.9, $q(n) \geq (q_{min})^{|B|} > 0$. Therefore, the lower bounds of probability that Markov chain X_n converges to the optimal solution is:

$$q^*(n) = 1 - (1 - \widehat{q})^n \geq 1 - (1 - (q_{min})^{|B|})^n$$

As $0 < (q_{min})^{|B|} < 1$, therefore $0 < 1-(q_{min})^{|B|} < 1$. When $n \to \infty$, $(1-(q_{min})^{|B|})^n = 0$, thus, $\lim_{n\to\infty} q^*(n) = 1$

The proof in Theorem 1 is a foundation of Theorem 3.

□

Lemma 2. *Given* $\rho_n = 1 - \sqrt{\frac{n}{n+a}}, a \in \mathbb{N}^+, \sum_{n=1}^{\infty} \rho_n = \infty$

Proof.

$$\begin{aligned}
\sum_{n=1}^{\infty} \rho_i &= \sum_{n=1}^{\infty}(1 - \sqrt{\frac{n}{n+a}}) \\
&= \sum_{n=1}^{\infty} \frac{\sqrt{n+a} - \sqrt{n}}{\sqrt{n+a}} \\
&= \sum_{n=1}^{\infty} \frac{a}{n+a+\sqrt{n(n+a)}} \\
&\geq a \cdot \sum_{n=1}^{\infty} \frac{1}{2(n+a)} \\
&= \frac{a}{2} \sum_{n=1}^{\infty} \frac{1}{n+a} = \infty
\end{aligned}$$

□

Theorem 3. *Let* ρ_n *as evaporation rate, non-homogeneous Markov chain* $X(n) = \{P(n), E(n), T(n)\}$ *converges to the optimal solution with a probability of 1.0.*

Proof. Suppose that v_i is not passed by any ant, the pheromone intensity of this node reaches the minimum at time step n.

$$
\begin{aligned}
v_j.p(n) &= \prod_{i=1}^{n}(1-\rho_i)\cdot v_j.p(0)\\
&= \prod_{i=1}^{n}\sqrt{\frac{i}{i+a}}\cdot v_j.p(0)\\
&= \prod_{i=1}^{a}\sqrt{\frac{i}{n+i}}\cdot v_j.p(0)
\end{aligned}
$$

Thus, $q_{min} = \prod_{i=1}^{a}\sqrt{\frac{i}{n+i}}\cdot v_j.p(0)$

□

In Lemma 2, I prove $\sum_{n=1}^{\infty}\rho_n = \infty$, so the Markov chain satisfies the conditions in Theorem 1.

4.6 Experiments and Analysis

Four experiments have been conducted to evaluate SIMiner. In Experiment 1, the convergence of SIMiner is explored. Experiment 2 intends to analyse the time to converge of SIMiner when giving the different size of ant agents. Experiment 3 evaluates the performance of SIMiner in static networks, and Experiment 4 analyses the behaviour of SIMiner in dynamic networks.

4.6.1 Experiment Setup

Datasets. The empirical analysis and evaluation have been conducted by using four public datasets as follows.

- **Ego-Facebook**[1] dataset, collected by McAuley et al. (2012) using a Facebook

[1]http://snap.stanford.edu/data/egonets-Facebook.html

application, which is archived in Stanford Large Network Dataset Collection (Leskovec & Krevl, 2014). It contains profile and network data from 10 ego-networks, consisting of 193 circles, 4,039 users and 88,234 edges.

- **Email-Enron**[2] dataset, which covers all the email communication. It has been posted to the web by the Federal Energy Regulatory Commission (Klimt & Yang, 2004; Leskovec, Lang, Dasgupta & Mahoney, 2009). The Enron email network has 36,692 nodes and 367,662 Edges. To diminish the computing time, a sub-graph with 10k nodes is captured for the experiment.

- **Wiki-Vote**[3] dataset, which incorporates administrator elections and votes history data from 3 January 2008. There are 2,794 elections with 103,663 total votes and 7,066 users participating in the elections. Nodes refer to Wikipedia users and edges represent votes from one user to another (Leskovec, Huttenlocher & Kleinberg, 2010a, 2010b).

- **Flixster**[4] is a movie social network which enables users to publish and share the reviews on movies by rating and commenting movies. The raw dataset has been collected by Jamali et al. (2010), which contains over 1 million users, 8,196,077 ratings and 7,058,819 undirected friendship affiliation links. The collected ratings range from December 2005 to November 2009. To process the raw dataset, duplicated links have been removed since the raw dataset has been crawled as a directed network. In the meanwhile, as for those users whose behaviours appear abnormal, their data have been removed. More specifically, the rating counts of such users are over 1,000 in a single day, and all the given rating scores are either 1 or 5; thus, they are most likely robots. On the other side, users with less than 100 ratings are also eliminated from the experiments.

[2] https://snap.stanford.edu/data/email-Enron.html
[3] https://snap.stanford.edu/data/wiki-Vote.html
[4] http://www.flixster.com/

Evaluation Metrics. To quantitatively evaluate the proposed model, two major performance metrics are utilised, i.e., the activation coverage and running time. Meanwhile, seed set variation rate is regarded as an indicator of convergence, and the corresponding indices are elaborated in Experiment 1.

- **Influence Activation Coverage** refers to the number of nodes that have been activated by the selected influential users, which represents the effectiveness of the algorithms. Specifically, by selecting the same amount of users from the same network using different selection algorithms, the higher the activation coverage, the better the performance. The IC model has been selected as the influence propagation model for evaluating the activation coverage of the selected seeds.
- **Time to Converge** describes the total time steps required by ant agents to carry out an optimal solution, at which point the convergence emerges. This means the ants already finished exploring the influencers from the environment, and it is ready for seeding process. Since SIMiner is a distributed approach, traditional estimation of running time may not be suitable.

Influence Probability and Estimation. Degree Weighted Activation (DWA) has been applied for influence probability estimation. DWA assigns the probability of each edge based on the degrees of both nodes. For example, the weight of edge is estimated as: $e_{ij}.w = 1/|\Gamma(v_j)|$.

The influence coverage of a seed set carried out by any algorithm is estimated by running numerous times of Monte-Carlo simulations.

Baselines. I compare SIMiner approach against the following algorithms in terms of influence activation coverage.

- **Greedy:** Attempt to the reach the maximum influence marginal gain in selecting each seed, coming with a $1 - 1/e$ approximation guarantee. Greedy selection

Table 4.2: Experiment Parameters

Parameter	**Default Value**
Seed set size	50
Ant agents size	10
Number of time steps	500
Sub-network layer N	2
α in Equation 4.2	0.8
β in Equation 4.2	0.4
Initial pheromone amount of v_i ($v_i.p(0)$)	0.5
Threshold of skip ω	0.4
Constant D in Equation 4.7	1000

is not scalable since it relies on a large number of Monte-Carlo simulations. However, in regards to the effectiveness, i.e., global influence activation coverage, it outperforms almost all the existing improved algorithms, such as IRIE (Jung, Heo & Chen, 2012).

- **Degree-based Selection:** Select the users with high degree, which is based on the intuition that users with larger friend circle can influence more users in the social network.
- **Degree Discount Heuristic:** If one node is nominated as a seed, all its adjacent neighbours' node degrees are discounted by one due to the presence of the node in the seed set (W. Chen et al., 2009).
- **Random Selection:** Select seeds randomly. It normally performs the worst since it is not based on any heuristics.

Experiment Parameters. The default parameters for running the experiments are listed in Table 4.2. The choice of the parameters is mainly based on the heuristics.

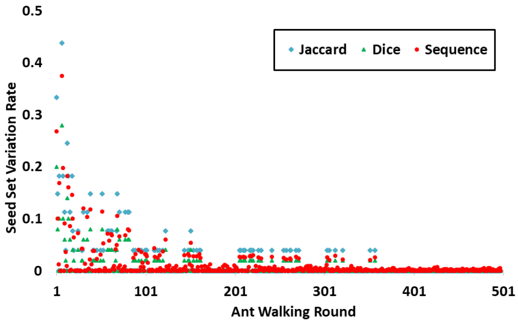


Figure 4.7: Seed Set Variations (Ego-Facebook)

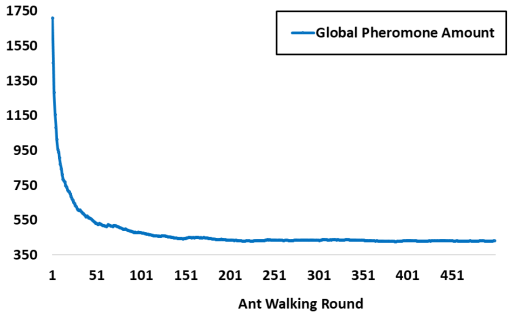


Figure 4.8: Pheromone Distribution (Ego-Facebook)

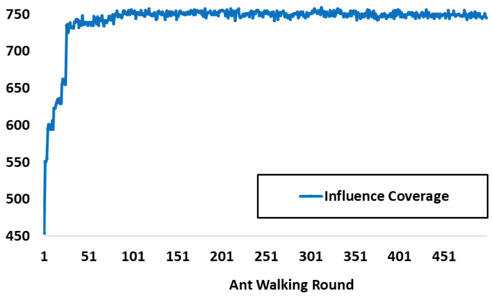


Figure 4.9: Activation Coverage (Ego-Facebook)

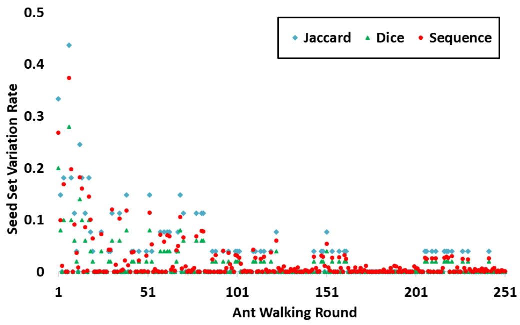


Figure 4.10: One Ant (Ego-Facebook)

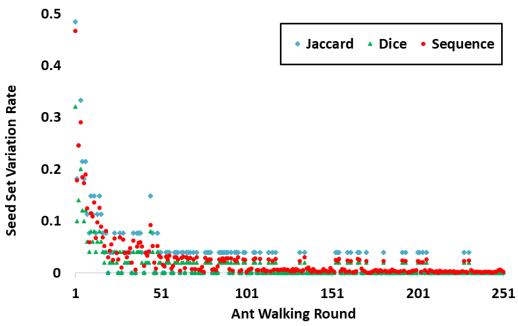


Figure 4.11: Five Ants (Ego-Facebook)

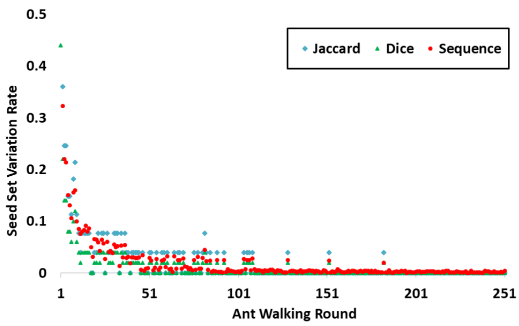


Figure 4.12: Ten Ants (Ego-Facebook)

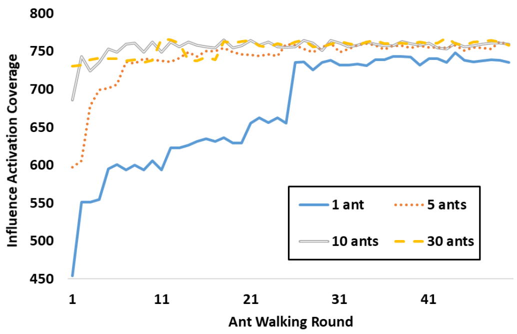


Figure 4.13: Different Ant Size (Ego-Facebook)

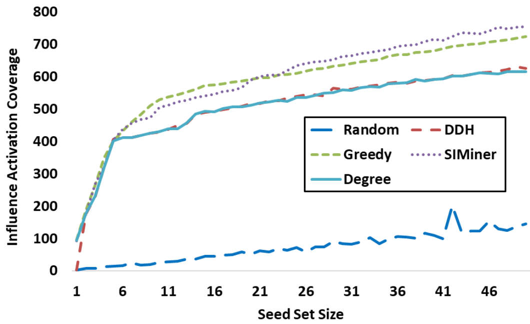


Figure 4.14: Ego-Facebook

4.6.2 Experiment 1 : Convergence Analysis

Experiment 1 aims to explore the convergence of SIMiner, showing the evolution of the solution, which is continuously optimised. In this experiment, I compute the seed set variation rate of the selected seeds over time and estimate the activation coverage of the seed set at each time step.

In SIMiner, a group of ant agents conducts global searching activities for exploring the influencers across the entire network. I regard the convergence emerges when the top k, i.e., the seed set size, users stay stable, without any significant seed-set variations in a number of consecutive time steps. In other words, low variation rate is a sign of convergence, and activation coverage of the seed set remains stable accordingly.

To quantify the variations between any two seed sets, three indices are adopted, i.e., Jaccard distance $d_{jcd}(S_1, S_2)$, Dice dissimilarity $d_{dic}(S_1, S_2)$ and sequential distance considering the index of the elements $d_{sqc}(S_1, S_2)$, which are formulated in Equations 4.10, 4.11 and 4.12, respectively. In these three equations, S_1 and S_2 denote two different seed sets, having the same cardinality, i.e., $S_1 \neq S_2, |S_1| = |S_2|$. $I(c|S_1)$ refers to the index of element c in set S_1.

$$d_{jcd}(S_1, S_2) = 1 - \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \tag{4.10}$$

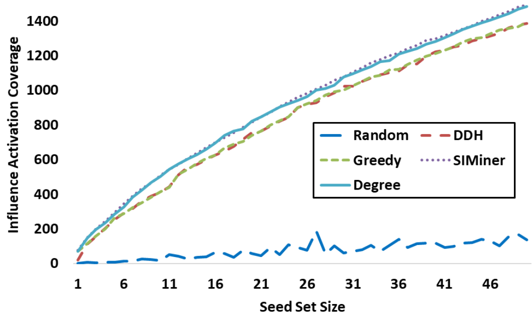


Figure 4.15: Wiki Vote

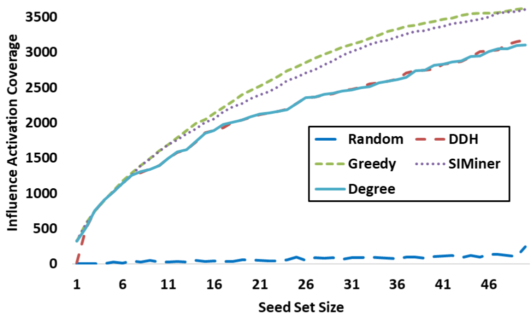


Figure 4.16: Email Enron

$$d_{dic}(S_1, S_2) = 1 - \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|} \tag{4.11}$$

$$d_{sqc}(S_1, S_2) = \frac{1}{|S_1|}\left(\sum_{c \in S_1 \cap S_2} \frac{|I(c|S_1) - I(c|S_2)|}{|S_1|} + |S_1 \setminus S_2|\right) \tag{4.12}$$

In this experiment, I explore the convergence phenomenon of SIMiner by using Ego-Facebook dataset only since the same trend can be observed by using other datasets. Figure 4.7 demonstrates the distribution of three indices within 500 walking rounds. The seed set variation rate declines slowly with the evolutionary effect from SIMiner. Meanwhile, Figure 4.8 shows the global pheromone distribution trend over time. As we can observe from Figure 4.9 that the influence activation coverage climbs up to around 750 rapidly within 50 walking rounds, and oscillates slightly. The experimental results explicitly reveal that the solutions carried out by SIMiner evolves and eventually start to converge from a certain point.

4.6.3 Experiment 2 : Time to Converge

To further explore SIMiner, Experiment 2 analyses the efficiency of SIMiner based on the previous experiment. As SIMiner is a distributed approach, the efficiency evaluation and improvement are very different from other centralised algorithms. Specifically, the traditional algorithm only can be improved by modifying the algorithm itself or leverage other statistical methods. By contrast, SIMiner models the individual agent, and the overall efficiency can be improved easily by adding more ant agents. In this experiment, the efficiency of SIMiner is analysed by exploring the time to converge.

By using the Ego-Facebook dataset, the number of the ant agents is increased to explore the trend of the convergence. The results of seed set variation rate when involving one ant, five ants and ten ants are demonstrated in Figures 4.10, 4.11 and 4.12, respectively. It is clear that in general the more ant agents are deployed, the faster convergence can be observed.

Figure 4.13 compares the activation coverage trend when involving the ants of the different size in SIMiner. It is evident that a larger number of ants have a higher starting point. The time to converge is longer when only one ant is working in the environment. 30 ants and 10 ants perform almost the same in terms of efficiency.

4.6.4 Experiment 3 : SIMiner in Static Networks

Experiment 3 aims to evaluate the performance of SIMiner in static networks. SIMiner is compared against some other state-of-the-art algorithms using three public datasets, i.e., Ego-Facebook, Email-Enron and Wiki-Vote. Greedy selection is one of the strongest baselines in the influence maximization problem, and it outperforms most of the existing algorithms but appears not scalable.

As we can see from Figures 4.14, 4.15 and 4.16 that SIMiner performs similarly or even better than Greedy algorithm.

Next, the parameters of SIMiner, including ω, α and β, are adjusted to investigate the performance. Figures 4.17 and 4.18 demonstrate the results of parameter choices using Ego-Facebook dataset. No significant changes but slight differences can be observed.

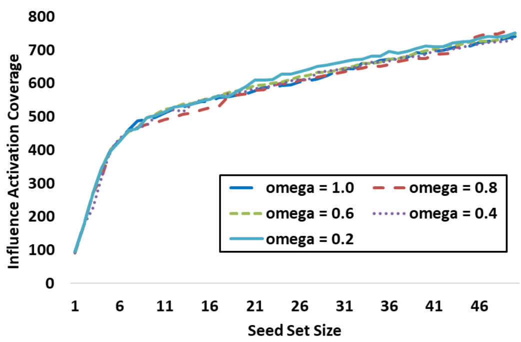


Figure 4.17: Parameter Choices ω

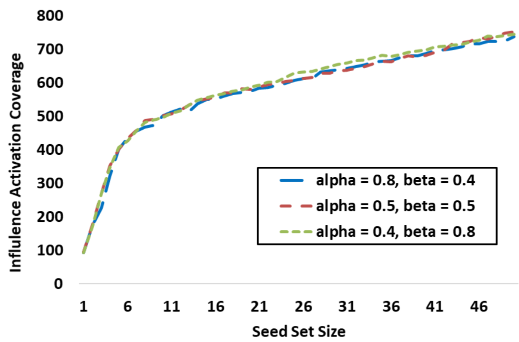


Figure 4.18: Parameter Choices α, β

4.6.5 Experiment 4 : SIMiner in Dynamic Networks

In Experiment 4, I analysed dynamic social networks and conducted influential nodes mining from a temporal perspective. This experiment describes SIMiner handles the dynamics by reusing the pheromone accumulated at previous time step in the environment.

The experiment encompasses two scenarios, i.e., the social network keeps expanding on a monthly basis, and the size of the network varies quarterly. In both scenarios, SIMiner keeps running over time, and no interruptions occur when changing the snapshots. However, the greedy selection algorithm identifies a set of influential nodes when given the first snapshot, and is executed again after a specific time interval, so that the influential user set is recalibrated. To simulate the dynamics and evolution of social networks, I have extracted a number of snapshots from Flixster dataset by following the assumptions below.

- A user is regarded as joining the social network when giving the first rating to any movie, while he or she quits the network after contributing the last rating.
- A user is considered as effective in a particular time frame if he or she has contributed any ratings during this period, i.e., a particular snapshot.

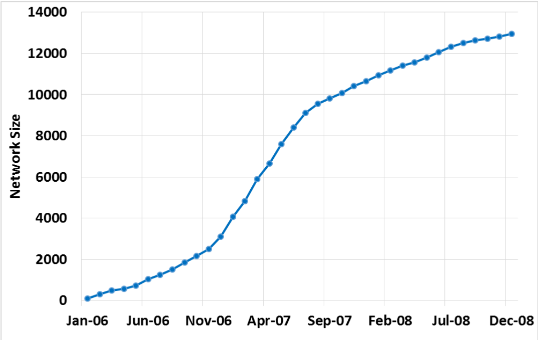


Figure 4.19: Monthly Flixster Network (Incremental)

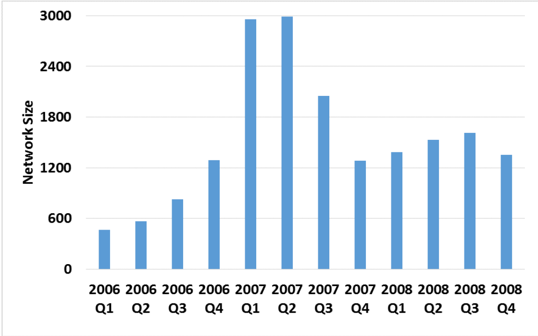


Figure 4.20: Quarterly Flixster Network (Dynamic)

- When users join the social network, the affiliation relationships are formed immediately.
- The influential nodes selected from a snapshot at k remain effective in the next snapshot at $k + 1$.

Figures 4.19 and 4.20 demonstrate the aforementioned two scenarios, i.e., 36-month and 12-quarter snapshots respectively, ranging from the year 2006 to 2008. Figures 4.21 and 4.22 reflect the dynamic features of social networks, where the users join and quit, and links are forming and vanishing over time. As can be observed from Figure 4.22 that the social network is actually 'changing the blood' over time, and the effective users' life cycle appears short. In Q1 2007, the difference between new joining users and quitting users is very high, while more users quit than that of joining in Q3 and Q4 2007.

The experimental results of the first scenario is presented in Figure 4.23, where the performance of both greedy selection and SIMiner have been compared in an incremental social network. SIMiner keeps running and selecting influential nodes. Whereas, the seed set identified by greedy algorithm is recalibrated every first month of the year. The seeds from greedy algorithm perform well for a few months. However, obvious performance degradation can be observed when the social network evolves.

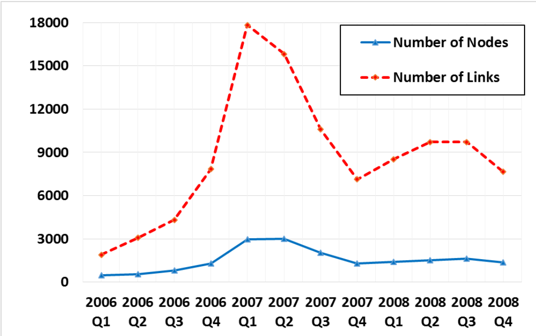


Figure 4.21: Quarterly Flixster Links Variation

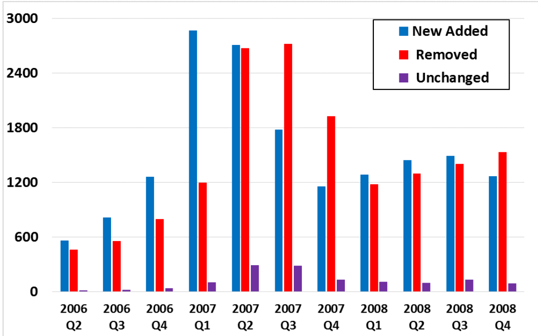


Figure 4.22: Quarterly Flixster Nodes Variation

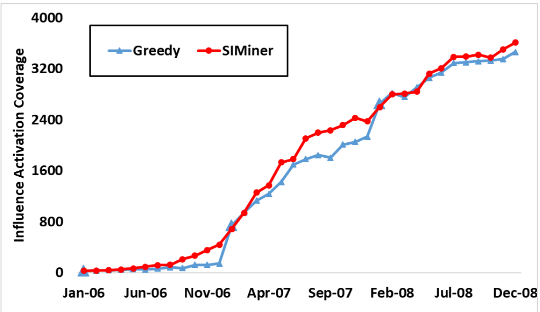


Figure 4.23: SIMiner in an Incremental Environment (Monthly)

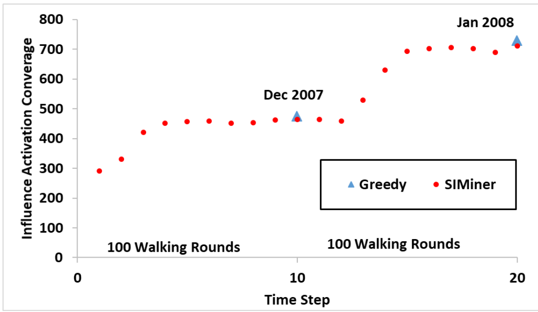


Figure 4.24: SIMiner Handling Incremental Environment

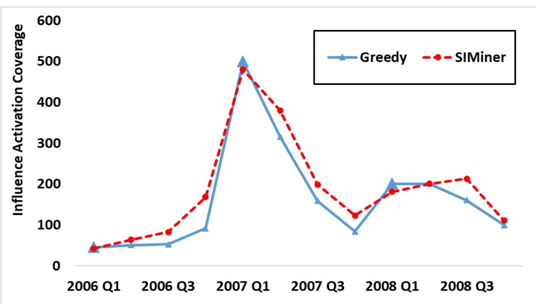


Figure 4.25: SIMiner in a Dynamic Environment (Quarterly)

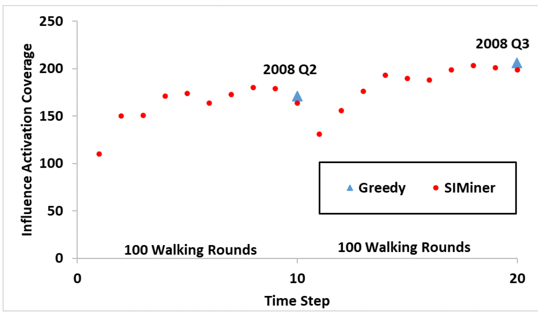


Figure 4.26: SIMiner Handling Dynamic Environment

To examine the detailed operations of SIMiner, two consecutive months are selected, i.e., December 2007 and January 2008. In Figure 4.24, it can be seen that when social network evolves, SIMiner is capable of adapting the solutions rapidly and carrying out an optimal solution.

Figure 4.25 demonstrates the results of the second scenario, where the effective users in the dynamic network vary on a quarterly basis. Greedy selection recalibrates the solution on the first quarter of each year. It is obvious that SIMiner is adaptive and performs well, whereas, without any calibration, the greedy selection loses its advantages when the social network evolves. In Figure 4.26, the performance of SIMiner degrades suddenly when the network snapshot changes, but increases immediately, approaching an optimal solution.

4.6.6 Discussion

Four experiments are conducted to explore the key features of SIMiner, as well as its performance in the influence maximization problem in both static and dynamic networks. SIMiner's advantages are demonstrated in mining influencers and handling dynamics from a microscopic level. The following insights can be uncovered from the experiments.

(1) In the influence maximization problem, SIMiner can always converge to an optimal solution within a number of iterations. Equipped with the same knowledge, the individual's behaviours lead to a global convergence. The identified seed set is continuously being revised and optimised, and eventually reaches a stable status. SIMiner is a global searching algorithm with evolutionary computing features, which are powered by stigmergic interactions. The influence effectiveness of the seed set from SIMiner is closer or even superior to that of the Greedy algorithm. Furthermore, SIMiner demonstrates a fast convergence, implying an excellent adaptability.

(2) SIMiner can handle large-scale networks when big data emerge. As SIMiner is a distributed approach, inheriting from ABM, the modelling merely focuses on individual's features and behaviours. To improve the efficiency in handling large-scale networks, more agents can be simply added to expedite the convergence speed. Therefore, SIMiner is not affected by the data size.

(3) SIMiner requires relatively complicated parameter choices, and the selection of these parameters are dataset-dependent. In SIMiner, parameters are supposed to be carefully determined to achieve a good performance. The selection still relies on the attempts of different combinations but I would like to set the investigation of parameter choices for SIMiner as one of the future work.

(4) The ant size appears to be a critical factor affecting the performance. Given a finite walking rounds within a limited time frame, employing one ant to explore a large network causes performance degradation. Because the solution may not fully converge before the walking finished. In this case, SIMiner more likely carries out a premature solution.

(5) SIMiner handles dynamics using agents in a local level. The environment changes are captured from a microscopic perspective. Given an undiscovered network, SIMiner requires a few time steps to "warm up", allocating pheromone to the environment, which can be regarded as an initialization. Any topological structure update is not supposed to affect the initial pheromone intensity. In other words, the exploration can be continued based on the existing pheromone and without any interruption.

4.7 Summary

This chapter presented a novel decentralised approach, i.e., SIMiner, to mine influential nodes in social networks. The outstanding merit of the proposed model is its adaptation ability. SIMiner is capable of adapting the solutions in complex dynamic environments.

Specifically, a set of artificial ant agents have been deployed, which keep crawling in the network and updating the environment via stigmergic interactions, thus, the solutions keep revising with the network evolution over time. Four experiments have been conducted to analyse and evaluate the performance of SIMiner in both static and dynamic environments. Experimental results reveal that SIMiner can fast converge to an optimal solution, whose effectiveness is closer or even better than Greedy selection. Moreover, it is able to handle the influential users mining tasks in dynamic networks without any re-calibrations. Furthermore, the proposed decentralised approach is suitable for many real-world networks, as it is applicable to large-scale networks and even functions without a global view.

This chapter mainly answers the Research Question 1 mentioned in Chapter 1. The model and results of this chapter have been published in (W. Li, Bai, Jiang & Zhang, 2016) and in the journal of IEEE Transactions on Big Data (W. Li, Bai & Zhang, 2018b).

In the next chapter, based on the generic agent-based influence diffusion model proposed in Chapter 3, I will further extend the influence maximization problem by investigating how to achieve long-term marketing by maintaining a particular influence.

Chapter 5

Automated Influence Maintenance in Social Networks

As introduced in Chapter 1, most existing studies in the field of influence maximization concentrate on how to maximize positive social impact to promote product adoptions based on static network snapshots. Such approaches can only increase influence in a social network in short-term, but cannot generate sustainable or long-term effects.

In this chapter, I study on how to maintain long-term influence in a social network and propose an agent-based influence maintenance model, which can select influential nodes based on the current status in dynamic social networks in multiple time-frames Within the context of my investigation, the experimental results indicate that multiple-time seed selection is capable of achieving more constant impact than that of one-shot selection. I claim that influence maintenance is crucial for supporting, enhancing and assisting long-term goals in business development. The proposed approach can automatically maintain long-lasting impact and achieve influence maintenance.

5.1 Overview

With the prevalence and advancement of the Internet, on-line social networks have become an important and efficient channel for information propagation. The propagation relies on one of the social phenomena, i.e., social influence, indicating that one's opinions or behaviours are affected by his or her contactable neighbours in the social network (Turner, 1991; Raven, 1964). Influence message is a common and concrete representation of social influence, which 'travels' rapidly through the network topologies via users' sharing experiences or behaviours.

In recent years, influence maximization draws tremendous attention to both researchers and domain experts. From a business perspective, influence maximization corresponds to short-term marketing effects, which tend to cause sudden profit spikes that rarely last (Valencia, 2013). Whereas, long-term marketing is typically more beneficial since it emphasises on long-term and sustainable business goals. Specifically, long-term influence can establish brand awareness and continually produce results even years down the road; thus, without having long-term marketing strategies, short-term success may be short-lived (Marketing, 2015). Motivated by this background, in this chapter, I aim to achieve constant impact for long-term marketing by investigating the preservation of a particular type of influential situation or status, called *influence maintenance*.

There are many limitations for short-term (or even one-shot) influence maximization when being utilised in real business cases. First, it focuses on how to maximize the influence of one-shot investment. Based on the risk management theory and best practice (Bender et al., 2010), with the same budget, the multiple-time investment could enable a better business strategy. In this way, the next action can be planned and carried out based on the outcome of the previous investment. For example, in a stock market, very few investors purchase stocks with all their money at only one time. Second, a great many business owners intend to expand the lifespan of influence, so that the brand awareness can be enhanced and increased in the long run (Aaker, Aaker & Biel, 1993). Influence maintenance not only cares about the quantity of users being affected but also considers constant influence impact.

Influence maintenance needs to be supported by a formal influence diffusion model which possesses two attributes: **(1)** the model is capable of capturing the temporal feature of a social network; **(2)** the model can monitor the status of a particular influence. On the other hand, in most existing on-line social media applications, information cannot be delivered to the users directly, but cached in individual's message repository, pending for users to access. The timeliness of a particular influence message becomes an important factor to be considered. More specifically, an individual reading list in on-line social networks, such as Weibo[1], is typically presented as a stack, which turns out to be last-post-first-read. Thus, the accessing priority of a particular message keeps decreasing over time, and posting or sharing behaviours are not supposed to be triggered without reading it.

In this chapter, I systematically elaborate and formulate the influence maintenance problem, which tends to maximize the constant impact of a particular influence by considering time-series. Meanwhile, a decentralised influence propagation model, i.e., the Agent-based Timeliness Influence Diffusion (ATID) model, is proposed. In the ATID model, the diffusion process is considered as a networked evolutionary phenomenon, users are modelled as autonomous agents, and each maintains its local information incorporating friendship affiliation list, message repository and posting histories. Furthermore, I propose the Timeliness Increase Heuristic (TIH) algorithm for solving the influence maintenance problem. Extensive experiments are conducted by using three real datasets. The experimental results show that:

[1]http://www.weibo.com

- Multiple-time selection can maintain influence better than one-shot selection.
- The TIH algorithm outperforms the other traditional seed selection algorithms regarding maintaining influence in social network.
- Seed-set variation is associated with both selection approaches and network properties.

The remainder of this chapter is structured as follows. Section 5.2 reviews the literature related to this research work. Section 5.3 introduces the preliminaries, formal definitions and problem description. Section 5.4 systematically elaborates the influence maintenance using the proposed decentralised diffusion model, and the TIH algorithm is also described. In Section 5.5, experimental results are presented to evaluate the performance of the proposed model. The summary of this chapter is given in Section 5.6.

5.2 Related Work

5.2.1 Adaptive Influence Maximization

A rich body of research works has been devoted to the influence maximization problem over the past ten years (Kempe et al., 2003; Domingos & Richardson, 2001). The majority of these studies fall into either full-feedback or non-feedback models (Golovin & Krause, 2011). In the former, all the seeds are committed based on the networked features or specific heuristics. Namely, there is no adaptive seed selection policy applied. Whereas, the latter utilises the observations during the seeding process, where the rules for identifying influencers are also known as adaptive policies. Based on the full-feedback model, some researchers extend the influence maximization problem by exploring the adaptive budget allocations (Golovin & Krause, 2011; Alon et al., 2012;

Soma et al., 2014). Typical research works tackling adaptive influence maximization problem have been reviewed in Section 2.2.3 of Chapter 2.

My research work departs from the body of these studies mainly in two aspects. First, the existing studies focus on investigating adaptive policies on the basis of the concept of adaptive submodularity (Golovin & Krause, 2011). Whereas, I concentrate on modelling the influence maintenance, achieving a constant impact by considering the timeliness degrees, though adaptive seeding algorithms are proposed to accommodate to the model. Second, these research works do not give a clear concept of time-series, and the networked evolutionary trend driven by influences is not captured. While, in this chapter, the time-series can be presented and the global observation of a social network status can be captured since ABM has been applied in the proposed model.

5.2.2 Influence Diffusion Modelling

As mentioned in Subsection 2.1.3 that the IC model and the LT model are two popularly adopted influence-diffusion models and many studies are conducted under various extended influence diffusion models.

However, most of the existing research works oversimplify the influence diffusion process, and the traditional propagation models concentrate on the activation state of each individual. Whereas, the users' features and behaviours affecting the influence acceptance have not been considered. Moreover, the dynamic status of influence messages over time is neglected. By contrast, the proposed ATID model for influence maintenance is decentralised, focusing on modelling individuals' personalised traits and behaviours. Furthermore, ATID model is capable of capturing the evolutionary network trend based on time-series, as well as the status of influence messages.

5.2.3 Agent-based Modelling for Influence Diffusion

As introduced in Section 2.3.1 of Chapter 2 that some studies are conducted to model the influence diffusion in a social network by leveraging ABM. Different from the works in this field, the proposed ATID model captures the properties of the influences existed in the same environment as that of the individuals'. Therefore, the observations of the disseminated influence messages can also be reflected from ATID model.

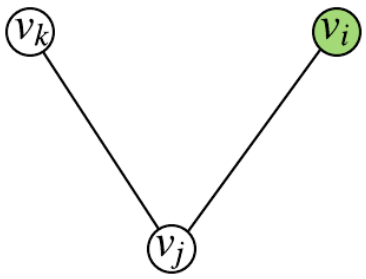
5.3 Preliminaries and Problem Formulation

5.3.1 Social Networks and Newsfeed

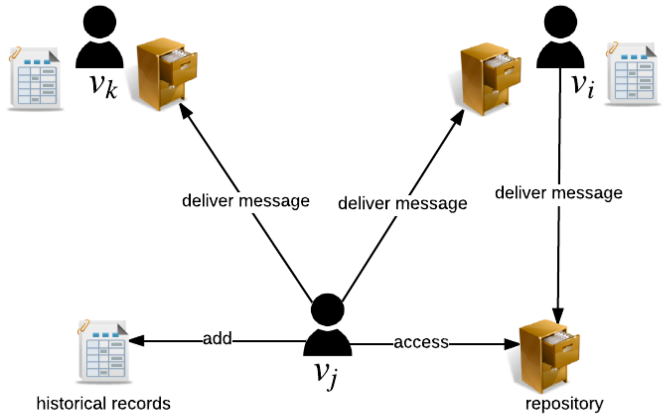
Most on-line social networks can be classified into two categories, depending on whether the newsfeed is re-organised.

First, some popular on-line social networks, such as Facebook[2], create personalised activity feeds for increased interactions and content contributions (Berkovsky & Freyne, 2015). For example, Facebook previously employed the *EdgeRank* algorithm[3] to determine which stories appear as newsfeed for each user by considering three original elements, i.e., affinity, weight and time decay (Carlton, 2015). Therefore, to maximize the impact of a particular influence, social media marketers need to stay informed of the changes to the latest newsfeed algorithms. Nowadays, newsfeed algorithms have become much more sophisticated. For example, Facebook has begun to employ a more complex ranking algorithm based on machine learning (Berkovsky & Freyne, 2015; Carlton, 2015; McGee, 2013). In this sense, it is nearly impossible for researchers to investigate the influence diffusion modelling in such social networks, as the outcome is much dependent on the newsfeed algorithms.

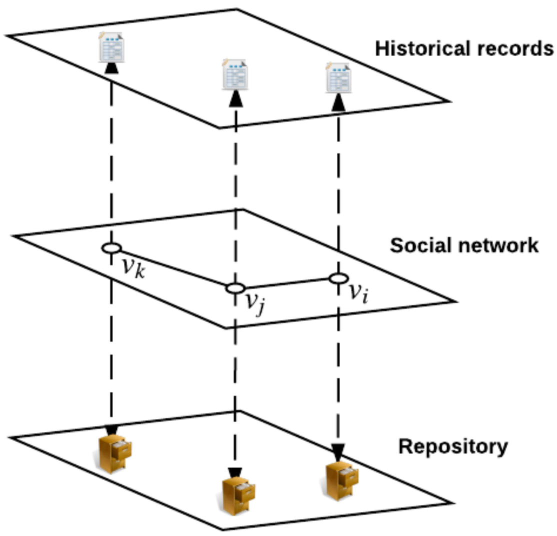
[2]https://www.facebook.com/
[3]http://edgerank.net/



(a) Social Network Graph



(b) Agent-based Influene Diffusion Model



(c) Influence Diffusion Facets

Figure 5.1: General Idea of the Proposed Model

Second, on-line social networks, like WeChat[4], enable users to share daily moments with friends. The newsfeed is generated instantly based on the timeline. Moreover, the social interactions among the individuals, such as 'comments' and 'like', are only visible if friendship connections are established. Different from Facebook, such kind of social networks allow duplicate messages propagating through the network, and

[4]http://www.wechat.com/en/

no newsfeed algorithms are applied. Moreover, ‘posting a message’ or ‘forwarding a message’ can be regarded an influential behaviour, while ‘like’ and ‘comments’ weigh less due to the visibility and privacy restrictions.

In this research, I mainly focus on the second category of social networks and investigate influence maintenance, where the timeliness degree of a message plays a pivotal role in organising the newsfeed.

5.3.2 Agent-based Influence Diffusion

ABM simulates the influence diffusion process by emphasising individualised features and behaviours. Based on the influence theory, homophily and influence are driven by the users’ preferences. Thus, individuals have different tendencies of reading and posting different types of topics (L. Wu et al., 2016). The messages wrapped with influence are supposed to be delivered to the repository of corresponding recipients, where the repository is filled with the influence messages from the neighbours. Each agent has a different frequency of accessing its repository. Based on agent’s preference and message timeliness degree (see Definition 5.4), the agent determines whether the information is to be shared with its adjacent neighbours. If an agent is influenced (*activated*), i.e., posting action is triggered, then, the influence message reaches its neighbours’ repositories. Whereas, in the recipient’s repository, the timeliness degree of this message keeps decreasing over time, but this will be refreshed if the repository owner is activated or the same message has been received again.

Figure 5.1 demonstrates a toy example, which represents the general idea of the proposed model. Figure 5.1a shows an ordinary social network graph in traditional influence diffusion models. Let v_i be an initial influencer who attempts to activate v_j with a certain success rate. If v_j is activated by v_i, it will then intend to influence the adjacent neighbour v_k. Figure 5.1b describes the proposed model from a microscopic

Table 5.1: Frequently Used Notations

Notation	Description
v_i	user agent
msg_p	influence message
φ	timeliness degree
r	influence attenuation constant
λ	the speed of influence decay
R_{v_i}	incoming message repository of v_i
H_{v_i}	historical records of v_i
A	seed set
ϱ_{msg_p}	global activation coverage (GAC) of msg_p
$\xi^{t_n}_{msg_p}$	global timeliness degree (GTD) of msg_p at t_n
Ω_{msg_p}	global cumulative timeliness degree (GCTD) of msg_p
$\mathit{\Delta\Omega}$	incremental timeliness contribution
$P(.)$	probability
$g(.)$	timeliness gain

point of view. Individual's influence activation is achieved by accessing the repository. More specifically, if a user is influenced (activated), the influence message is supposed to be delivered to all the neighbours' repositories. Meanwhile, this message is archived as one of the sender's historical records. From a macroscopic viewpoint, apart from the topological structure of a social network, two more factors are affecting the influence propagation, i.e., the historical records and the repository, which is illustrated in Figure 5.1c. The detailed modelling will be elaborated in Section 5.4.

Before moving on to the technical parts of this chapter, the frequently used notations are summarised in Table 5.1.

5.3.3 Formal Definitions

Definition 5.1: A user agent $v_i, (v_i \in V)$ is defined as a vertex in a social network $G = (V, E)$, where $V = \{v_1, ..., v_n\}$ denotes a set of agents and E represents a set of edges, $E = \{e_{ij} | 1 \leq i, j \leq n\}, i, j \in \mathbb{N}^{+}, \{v_i, v_j\} \subseteq V$. v_i has a neighbour set $\mathit{\Gamma}(v_i)$. If agent v_j is a neighbour of v_i, then $\{e_{ij}\} \subseteq E, v_j \in \mathit{\Gamma}(v_i)$. While, E_{v_i} indicates the edge set connected with v_i, where $E_{v_i} = \{e_{ij} | v_i \neq v_j \wedge v_j \in \Gamma(v_i)\}$. $|V|$ and $|E|$ denote the cardinality of agents and edges respectively. The affiliation information is maintained by each agent locally. In addition, agent v_i has a binary state $s_{v_i}^{msg_p}$ towards a particular influence message msg_p (see Definition 5.3), where $s_{v_i}^{msg_p} \in \{0, 1\}$, representing inactive and active, respectively.

Definition 5.2: Environment ε_{v_i} is an ego network representing the local influence diffusion context or the local view of a particular agent v_i. The environment of v_i is denoted by using a four-tuple, $\varepsilon_{v_i} = (\Gamma(v_i), E_{v_i}, R_{v_i}, H_{v_i})$, where R_{v_i} and H_{v_i} represent the repository and historical records of v_i, respectively (see Definitions 4 and 5). Each agent is capable of accessing all the resources in its environment.

Definition 5.3: Influence message msg_p is defined as a particular piece of information sent from one person to his or her contactable recipients, affecting their opinions or behaviours. It is a common and concrete representation of social influence in on-line social networks. In the current settings, each influence message msg_p belongs to a particular topic τ_x, i.e., $msg_p \in \tau_x$. If agent v_i is influenced after accessing msg_p, then $s_{v_i}^{msg_p} := 1$; meanwhile, v_i attempts to deliver the influence message to the repositories of neighbours $\Gamma(v_i)$.

Definition 5.4: Timeliness degree of an influence message is a real value, describing the position of an influence message in a user's repository at a particular time. Timeliness degree not only reflects the status of the influence message, but also implies whether a specific piece of news arrives at a suitable time. In reality, it happens more than often that users check the friend-circle or moments update right after a message has been posted. Subsequently, this influence message has a higher chance to draw the user's attention than that of the others. Mathematically, the timeliness degree of message

msg_p in v_i's repository at time t_m is defined using the notation $\varphi(v_i, msg_p, t_m)$.

Inspired by the behaviour analysis approach introduced in (Benevenuto, Rodrigues, Cha & Almeida, 2009; Fang, Hu, Li & Tsai, 2013), I assume that the effect of influence satisfies the principle of natural decay; thus, the exponential decay, i.e., e^{-r}, can be leveraged to describe the attenuation of influence, where r denotes the attenuation constant. Suppose message msg_p has been delivered to v_i's repository at t_b, then the timeliness degree is formulated in Equation 5.1.

$$\varphi(v_i, msg_p, t_m) = e^{-r\cdot(m-b)} \tag{5.1}$$

The timeliness degree of any message equals to 1 when arriving at the repository, i.e., $m = b$, and starts to decrease over time. Therefore, the speed of influence decay λ is described in Equation 5.2, which shows the speed is gradually slowing down.

$$\begin{aligned}\lambda &= \varphi(v_i, msg_p, t_{m-1}) - \varphi(v_i, msg_p, t_m)\\ &= e^{-r\cdot(m-1-b)} - e^{-r\cdot(m-b)}\\ &= (e^r - 1)\cdot e^{-r\cdot(m-b)}, m \geq b\end{aligned} \tag{5.2}$$

I assume msg_p is supposed to be ignored by agent v_i after time t_e, subject to $e \in \mathbb{N}$, $e > m - b$ and $\varphi(v_i, msg_p, t_e) \geq \sigma(msg_p)$, where $\sigma(msg_p)$ denotes the valid timeliness degree threshold of msg_p. Likewise, the higher timeliness degree, the greater probability that the influence message can be accessed by the user when visiting the repository.

Definition 5.5: Repository $R_{v_i}^{t_m} =< r_1, r_2, ..., r_n >$ refers to a cached container of agent v_i at time step t_m. It incorporates all the valid incoming messages from neighbours $\Gamma(v_i)$ to agent v_i. Each agent has a different frequency of accessing the repository. An element in $R_{v_i}^{t_m}$ can be represented as a three-tuple, i.e., $r_k = (v_j, msg_p, \varphi)$, where v_j denotes the agent who posts the influence message msg_p, $v_j \in \Gamma(v_i) \cup \{v_i\}$ and

$\varphi \geq \sigma(msg_p)$. For simplification purposes, φ is regarded as the timeliness degree of the corresponding message at t_m, which is equivalent to $\varphi(v_i, msg_p, t_m)$.

Definition 5.6: Historical records refer to past outgoing influence messages delivered from a particular user to the neighbours. Historical records $H_{v_i} = \{txn_1, txn_2, ..., txn_n\}$ is defined as a collection of user v_i's past sharing transactions, i.e., posted messages. An element of H_{v_i} can be denoted by a three-tuple, i.e., $txn_n = (msg_p, \varphi, t_m)$, where φ represents the message timeliness degree when posted (clarified in Definition 5.5), $\varphi \geq \sigma(msg_p)$. While, t_{now} refers to the current time step, and Δt describes the valid lifespan of a transaction, $t_{now} - t_m \leq \Delta t$. Given $t_{now} - t_m > \Delta t$, the corresponding transaction is supposed to be removed from the collection. Historical records H_{v_i} is also an implication of agent v_i's interests or preferences.

5.3.4 Problem Description

Influence maintenance in this thesis is defined as the process of preserving a particular type of influential situation or the status of influence being preserved. The concept is derived from influence maximization. Specifically, given a finite budget k (seed set size) and a limited time span $[t_0, t_m]$, an investment (seed selection) occurs once every n time steps, thus, the investment time steps $I = \{t_{N\times n} | N \in \mathbb{N} \wedge N \times n < m\}$, where $t_{N\times n}$ represents a particular seed selection point. There are $|I|$ times of investment considered for maintaining the influence.

Influence maintenance aims to find a solution of identifying the seed set $A_{t_{N\times n}}$ for each time step $t_{N\times n}$ to maximize the influence lifespan of msg_p. Thus, the selected seed set A is a collection of seeds identified from each investment time step, i.e., $A = \{A_t | t \in I\}$ and

$$\sum_{t \in t_{N \times n}} |A_t| = k \tag{5.3}$$

I assume that the same amount of seeds are supposed to be selected for each selection point, and any seeds cannot be selected more than once. In other words, given $\{A_i, A_j\} \subseteq A$, $|A_i| = |A_j|$, $A_i \cap A_j = \varnothing$.

The *Global Timeliness Degree (GTD)* of msg_p at a particular time step t_n is represented as $\xi^{t_n}_{msg_p}$, which can be calculated by using Equation 5.4. The popularity trend of a particular influence message can be reflected by connecting the GTD of the corresponding influence in each time step.

$$\xi^{t_n}_{msg_p} = \sum_{v_i \in V} \varphi(v_i, msg_p, t_n) \tag{5.4}$$

The overall effective influence lifespan of msg_p in the entire social network is evaluated by using *Global Cumulative Timeliness Degree (GCTD)* of a specific time span $[t_0, t_m]$, i.e., Ω_{msg_p}, which can be derived by using Equation 5.5.

$$\Omega_{msg_p} = \sum_{t_0}^{t_m} \xi^{t_n}_{msg_p} = \sum_{t_0}^{t_m} \sum_{v_i \in V} \varphi(v_i, msg_p, t) \tag{5.5}$$

The objective of influence maintenance is to maximize Ω_{msg_p}. Furthermore, the traditional influence effectiveness evaluation metrics, i.e., Global Activation Coverage (GAC), is taken into consideration as well. GAC of influence message msg_p is denoted using the notation ϱ_{msg_p}, indicating the number of users in the social network getting affected or activated by msg_p. ϱ_{msg_p} can be calculated by using Equation 5.6.

$$\varrho_{msg_p} = \sum_{v_i \in V} |\{v_i | s^{msg_p}_{v_i} = 1\}| \tag{5.6}$$

5.4 Influence Maintenance Model

5.4.1 The Agent-based Timeliness Influence Diffusion (ATID) Model

ATID model is a decentralised influence diffusion model which utilises the advantages offered by ABM. The influence propagation in social networks demonstrates a networked evolutionary pattern driven by individuals' actions. In this model, each agent maintains its ego-network and makes decisions of performing social activities based on both timeliness degree of the influence message and its preference.

There are many reasons to make a user to carry out a social behaviour, such as influence from neighbours in the same social networks, affected by any external events, or the user actively posts some messages without getting influenced by anybody (A. Goyal et al., 2010). In the proposed model, I assume users deliberately post messages after influenced by the neighbours, and each individual's repository and historical records contain enough evidence for statistical analysis. Furthermore, each user agent (e.g., v_i) has a different frequency of accessing its repository, i.e., $freq(v_i)$, which can be calculated by using Equation 5.7. It can be seen that $freq(v_i)$ is equivalent to the probability of v_i accessing a particular message msg_p in its repository at time t_m, i.e., $P_f(v_i, msg_p, t_m)$.

$$\begin{aligned} freq(v_i) = & P_f(v_i, msg_p, t_m), \\ & subject\ to\ \varphi(v_i, msg_p, t_m) \geq \sigma(msg_p) \end{aligned} \tag{5.7}$$

One important task of influence diffusion modelling is to identify the probability of getting activated after reading message msg_p of topic τ_x at time t_m, where the influence probability may not remain constant independently of time (A. Goyal et al., 2010). Therefore, in the proposed model, a user agent has the capability of adapting its probability of posting message msg_p based on two major factors, i.e., the attention degree of influence message msg_p and the user preference derived from the latest k posts. Therefore, the probability of user agent v_i posting message msg_p at time t_m can be estimated in Equation 5.8:

$$P(msg_p|R_{v_i}^{t_m}, H_{v_i}) = P(msg_p|R_{v_i}^{t_m})P(\tau_x|H_{v_i}, msg_p \in \tau_x) \tag{5.8}$$

In Equation 5.8, $P(msg_p|R_{v_i}^{t_m})$ represents the attention degree of influence message msg_p in v_i's repository at time t_m, i.e., the probability of getting attracted by msg_p, which is associated with the message timeliness degree $\varphi(v_i, msg_p, t_m)$. While $P(\tau_x|H_{v_i}, msg_p \in \tau_x)$ denotes the probability of sharing topic τ_x at time t_m on the basis of v_i's past behaviours.

Thus, the attention degree of influence message msg_p in v_i's repository at time t_m is formulated in Equation 5.9.

$$P(msg_p|R_{v_i}^{t_m}) = \frac{\sum\limits_{r_n \in R_{v_i}^{t_m} \wedge r_n.msg = msg_p} \varphi(v_{r_n}, r_n.msg, t_{r_n})}{\sum\limits_{r_n \in R_{v_i}^{t_m}} \varphi(v_{r_n}, r_n.msg, t_{r_n})}, \tag{5.9}$$

where $t_{r_n} = t_m - t_n$, t_n denotes the time when the message r_n arrives the repository.

According to v_i's historical records, the probability of sharing topic $\tau_x, msg_p \in \tau_x$ at time t_m can be derived from the weighted average of topic τ_x's timeliness difference. Specifically, if msg_p has been posted when its timeliness degree $msg_p.\varphi$ is low, this implies that the user is very interested in the topic of msg_p (i.e., $msg_p.\tau$), and the message timeliness degree will not significantly impact the chances of posting such messages. Hence, $P(\tau_x|H_{v_i}, msg_p \in \tau_x))$ is represented in Equation 5.10.

$$P(\tau_x|H_{v_i}, msg_p \in \tau_x)) = \frac{\sum\limits_{msg_p \in \tau_x} (1 - msg_p.\varphi)}{\sum\limits_{msg_q \in H_{v_i}} (1 - msg_q.\varphi)} \tag{5.10}$$

5.4.2 Diffusion Process under ATID Model

Benefited from ABM, individual's features, behaviours and the local environment can be considered in ATID model. As ATID model is a decentralised influence propagation model, the diffusion algorithm under ATID model corresponds to an agent's response when accessing its repository. The diffusion process in ATID model is described in Algorithm 5.

Algorithm 5 The Influence Diffusion Algorithm under ATID Model

Input: $v_i, t_m, msg_p, msg_p \in \tau_x$

Output: v_i's social behaviour (posting / not)

1: Generate random decimal $rand_1$
2: **if** $rand_1 \leq freq(v_i) \wedge \Phi(msg_p|H_{v_i}) = 0$ **then**
3:     Compute $P(msg_p|R_{v_i}^{t_m})$ using Equation 5.9
4:     Compute $P(\tau_x|H_{v_i}, msg_p \in \tau_x))$ using Equation 5.10
5:     Compute $P(msg_p|R_{v_i}^{t_m}, H_{v_i})$ using Equation 5.8
6:     Generate random decimal $rand_2$
7:     **if** $P(msg_p|R_{v_i}^{t_m}, H_{v_i}) \leq rand_2$ **then**
8:         **for** $\forall v_j \in \Gamma(v_i) \cup \{v_i\}$ **do**
9:             $R_{v_j}^{t_{m+1}} := R_{v_j}^{t_m} \cup \{(v_i, msg_p, 1)\}$
10:         **end for**
11:         $H_{v_i} := H_{v_i} \cup \{(msg_p, \varphi, t_m)\}$
12:     **end if**
13: **end if**
14: **for** $\forall r_n \in R_{v_i}^{t_{m+1}} \setminus \{(v_i, \tau_x, \varphi)\}$ **do**
15:     $r_n.\varphi := r_n.\varphi - \lambda$
16: **end for**

In Algorithm 5, the inputs incorporate user agent v_i, time t_m, the influence message msg_p and msg_p's corresponding topic τ_x; while the output is v_i's social behaviour, i.e., post msg_p at time t_m or not. Line 2 checks the precondition of sharing msg_p, where $\Phi(msg_p|H_{v_i})$ is an indicator function, which returns 0 if msg_p is not posted by v_i before, and 1 otherwise. Lines 3-5 aim to compute the probability of posting msg_p by v_i at t_m. Lines 8-11 update the repositories of agents in v_i's ego-network, as well as its own historical records. Lines 14-16 demonstrate that the message timeliness attenuation occurs in v_i's repository.

5.4.3 The Timeliness Increase Heuristic (TIH) Algorithm

There are some classic seed selection algorithms, such as degree-based, greedy, random and Degree Discount Heuristic (DDH) selections (Kempe et al., 2003; W. Chen et al., 2009). These algorithms are developed based on either the node features or influence diffusion models. More specifically, degree-based approach identifies the influencers by considering the node degree; greedy algorithm attempts to reach the maximum influence marginal gain in each selection, but it is not scalable; DDH extends the rank-based algorithm that once a node is selected, the degree of corresponding neighbours is deducted by one; random selection does not follow any heuristics, which selects seeds randomly.

DDH is developed based on the fact that many of the most central nodes may be clustered; thus, it is not necessary to target all of them (W. Chen et al., 2009). A similar concept can be utilised for maintaining an influence, that is, the influence fading-out zone should be first targeted to achieve influence maintenance.

Inspired by DDH, I have developed the TIH algorithm (i.e., Algorithm 6). For selecting each seed, the TIH tends to search for the user v^*, who can bring the maximum message timeliness gain, which is calculated in Equations 5.11 and 5.12.

$$v^*_{t_m} = \underset{v_i}{\operatorname{argmax}} \sum_{v_j \in \{v_i\} \cup \Gamma(v_i)} g(v_j, msg_p, t_m) \tag{5.11}$$

$$g(v_j, msg_p, t_m) = 1 - \varphi(v_j, msg_p, t_m) \tag{5.12}$$

In Equations 5.11 and 5.12, $g(v_j, msg_p, t_m)$ denotes v_j's message timeliness gain if v_i is selected as a seed. The selection of the next seed is based on the assumption that if previously identified seeds are selected. Thus, the TIH selection is described in Algorithm 6.

Algorithm 6 The TIH Algorithm

Input: $G = (V, E), k_m, t_m, msg_p$
Output: A_m

Initialise $A_m := \varnothing$
for $\forall v_i \in V$ **do**
    $v_i.\varphi' := v_i.\varphi$
end for
while $|A_m| < k_m$ **do**
    for $\forall v_i \in V$ **do**
        $g_{sum}(v_i, msg_p, t_m) := 0$
        for $\forall v_j \in \{v_i\} \cup \Gamma(v_i)$ **do**
            $g(v_j, msg_p, t_m) = 1 - v_j.\varphi'$
            $g_{sum}(v_i, msg_p, t_m) += g(v_j, msg_p, t_m)$
        end for
    end for
    Find v^* using Equation 5.11
    $A_m := A_m \cup \{v^*\}$
    $v^*.selected := true$
    for $\forall v_j \in \{v^*\} \cup \Gamma(v^*)$ **do**
        $v_j.\varphi' := 1$
    end for
end while

In Algorithm 6, the inputs include the social network G, the number of seeds to be selected k_m, the time step t_m and influence message msg_p; the output is the selected seed set at t_m. Lines 2-4 replicate all the user agents' current timeliness degree of msg_p to a temporary variable. Lines 6-11 calculate the global timeliness gain for all the users in G, in other words, this evaluates the influence impact of each individual. Lines 12-13 aim to find the most 'beneficial' user. Lines 15-16 update the temporary timeliness variables of all the users in v^*'s ego network with the assumption that if v^* is activated and selected as a seed. The worst-case time complexity of the TIH algorithm is determined by Lines 5-8. As k_m is a constant, the complexity is $O(n^2)$.

It can be seen that the seed set selected by TIH algorithm is the local optimal solution, following the heuristic that the largest timeliness fading-out zone should be firstly targeted. Moreover, the TIH demonstrates its advantages in maintaining the influence of a hypothesis message.

$$
\begin{aligned}
\varDelta\varOmega &= \sum_{v_j \in \{v_i\} \cup \varGamma(v_i)} \sum_{t=t_m}^{t_{m+n}} \varphi'(v_j, msg_p, t) - \varphi(v_j, msg_p, t) \\
&= \sum_{v_j \in \{v_i\} \cup \varGamma(v_i)} \left(\sum_{t=t_m}^{t_{m+n}} \varphi'(v_j, msg_p, t) - \sum_{t=t_m}^{t_{m+n}} \varphi(v_j, msg_p, t) \right) \\
&= \sum_{v_j \in \{v_i\} \cup \varGamma(v_i)} \left(\sum_{i=0}^{n} e^{-i \cdot r} - \sum_{i=0}^{n} e^{-(m_j+i) \cdot r} \right) \\
&= \sum_{v_j \in \{v_i\} \cup \varGamma(v_i)} \left(\frac{1-e^{-(n+1) \cdot r}}{1-e^{-r}} - \frac{1-e^{-(n+1) \cdot r}}{1-e^{-r}} \cdot e^{-m_j \cdot r} \right) \\
&= \sum_{v_j \in \{v_i\} \cup \varGamma(v_i)} \frac{1-e^{-(n+1) \cdot r}}{1-e^{-r}} \cdot (1-e^{-m_j \cdot r}) \\
&= \frac{1-e^{-(n+1) \cdot r}}{1-e^{-r}} \sum_{v_j \in \{v_i\} \cup \varGamma(v_i)} (1-e^{-m_j \cdot r}) \\
&= \frac{1-e^{-(n+1) \cdot r}}{1-e^{-r}} \sum_{v_j \in \{v_i\} \cup \varGamma(v_i)} (1-\varphi(v_j, msg_p, t_m)) \\
&= \frac{1-e^{-(n+1) \cdot r}}{1-e^{-r}} \sum_{v_j \in \{v_i\} \cup \varGamma(v_i)} g(v_j, msg_p, t_m)
\end{aligned}
\tag{5.13}
$$

Theorem 4. *In TIH algorithm, to obtain the node at t_m with maximum timeliness gain* *using Equation 5.11, is equivalent to get the node with maximum GCTD increment, i.e.,* $v^*_{t_m} = \underset{v_i}{\operatorname{argmax}} \Delta\Omega$

Proof. Given current time step t_m, and $\varphi(v_j, msg_p, t_m) = e^{-m_j \cdot r}$, where m_j denotes the time difference between when msg_p arrives and t_m. If node v_i has been selected as a seed, the corresponding timeliness degree of node set $\{v_j | v_j \in \{v_i\} \cup \Gamma(v_i)\}$ is supposed to be reset back to 1, i.e., $\varphi'(v_j, msg_p, t_m) = 1$. Therefore, the incremental timeliness contribution of activating v_j, i.e., $\Delta\Omega$ can be derived using Equation 5.13.

In Equation 5.13, $n \in \mathbb{N}$, representing the difference between the total time steps and the current time step, and $e^{-m_j \cdot r}$ denotes the timeliness degree of a particular message in v_j's repository at t_m according to Equation 5.1. It is obvious that $\frac{1-e^{-(n+1)\cdot r}}{1-e^{-r}}$ is a coefficient, $\sum_{v_j \in \{v_i\} \cup \Gamma(v_i)} g(v_j, msg_p, t_m)$ exactly corresponds to the objective function of TIH algorithm in Equation 5.11. Therefore, TIH is a kind of greedy algorithm. □

Lemma 5. *Let S be the seed set selected by TIH and S^* be the seed set that maximizes Ω_{msg_p}. $\Omega_{msg_p}(S)$ be the GCTD of msg_p with seed set S. Then $\Omega_{msg_p}(S) \geq (1 - 1/e) \cdot \Omega_{msg_p}(S^*)$. In other words, the theoretical guarantee for TIH in the influence maintenance problem is $1 - 1/e$.*

Proof. Let A be the initial seed set and $X =< v_1, v_2, ..., v_h >$ be one of the paths activated by A. $f(A)$ represents the GCTD of msg_p caused by A. $f_X(A)$ denotes the GCTD accumulated by path X. Similar to the calculations in Equation 5.13, we have:

$$f_X(A) = \sum_{j=1}^{h} \sum_{t=t_{m+j}}^{t_{m+n}} g(v_j, msg_p, t)$$

, where $g(v_j, msg_p, t_m)$ is defined in Equation 5.12, and $0 \leq g(v_j, msg_p, t) \leq 1$. It is easy to proof that $f_X(A)$ is sub-modular. Hence:

$$f(A) = \sum_{outcomes\ x} Prob|X| \cdot f_X(A)$$

, which is also sub-modular since the non-negative linear combination of sub-modular functions is sub-modular. As is clarified in Theorem 4 that TIH is a kind of greedy algorithm. According to Theorem 2.4 in (Kempe et al., 2003), we have $f(A) \geq (1 - \frac{1}{e})f(A^*)$, where A^* denotes the set that maximizes $f(.)$ over all k-element sets.

Let $f(S) = \Omega_{msg_p}(S)$, the lemma is proofed.

□

5.4.4 Influence Maintenance Analysis

I analyse the influence maintenance by considering the timeliness gain contributed by two seeds v_a and v_b under both scenarios, i.e., one-shot selection and multiple-time selection, where the time discrepancy of selecting both users is denoted by using m_0. In the former, no time discrepancy is presented, i.e., $m_0 = 0$, while in the latter, $m_0 \neq 0$.

Suppose that enough time is given for the influence decay, i.e., $n \to \infty$, if any node is activated, the theoretical timeliness gain would be $1/(1 - e^{-r})$ according to Equation 5.13. If all the influence-diffusion paths of active users fail to overlap with each other, the global timeliness gain of one-shot selection would be the same as that of the multiple-time selection. Whereas, in reality, this rarely happens. Therefore, I consider the situation when the influence-propagation paths cover same partial nodes with each other.

Suppose the influences disseminated from v_a and v_b can reach each other. In other words, path $\overrightarrow{v_{a,b}} =< v_a, v_1, v_2, ..., v_n, v_b >$ exists in the network, which is illustrated in Figure 5.2. Moreover, for simplification purpose, I assume that the influence propagation probability remains the same.

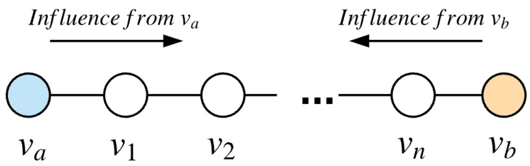


Figure 5.2: One of the Overlapped Influence Diffusion Paths

Table 5.2: Influence Maintenance Parameters

	v_1	v_2	...	v_n
T_{v_a}	d_0	$d_0 + 1$	...	$d_0 + n - 1$
T_{v_b}	$m_0 + d_0 + n - 1$	$m_0 + d_0 + n - 2$	...	$m_0 + d_0$
$\varDelta T$	$\lvert m_0 + n - 1\rvert$	$\lvert m_0 + n - 3\rvert$	...	$\lvert m_0 + n - (2n - 1)\rvert$
$\varDelta\Omega$	$\sum\limits_{i=0}^{\lvert m_0+n-1\rvert-1} e^{-ir}$	$\sum\limits_{i=0}^{\lvert m_0+n-3\rvert-1} e^{-ir}$	...	$\sum\limits_{i=0}^{\lvert m_0+n-(2n-1)\rvert-1} e^{-ir}$

For any node v_x in path $\overrightarrow{v_{a,b}}$, the corresponding $\varDelta\Omega$ is explicitly determined by the time discrepancy of the influences $\varDelta T$ from two sources, v_a and v_b, where $\varDelta T$ is associated with the number of nodes in between, i.e., n and the time difference in activating v_a and v_b, i.e., m_0. Hence, values of influence maintenance related parameters are described in Table 5.2.

In Table 5.2, T_{v_a} is an n-tuple, having a finite ordered list of n elements, where each element denotes the time step when the influence initiated from v_a arrives at the corresponding node, and the sequence implies the influence-diffusion path. Meanwhile, $\varDelta T = (T_{v_b} - T_{v_a})$, where each element indicates the absolute value of the difference between the elements in T_{v_b} and T_{v_a} at the same position.

If $\varDelta T$ is odd, $\varDelta T$ starts from $m_0 + n - 1$, decreasing by 2 further down the influence-diffusion path, and begins to increase by 2 for each hop when the value reaches 1. Similarly, if $\varDelta T$ is even, $\varDelta T$ drops by 2 and then is added by 2 after reaching 0. For example, given $n = 6$, we can obtain the data in Table 5.3.

Table 5.3: Example: value variation of ΔT (n=6)

m_0 (even)	ΔT (odd)	m_0 (odd)	ΔT (even)
0	$(5,3,1,1,3,5)$	1	$(6,4,2,0,2,4)$
2	$(7,5,3,1,1,3)$	3	$(8,6,4,2,0,2)$
4	$(9,7,5,3,1,1)$	5	$(10,8,6,4,2,0)$
6	$(11,9,7,5,3,1)$	7	$(12,10,8,6,4,2)$

Apparently, in both scenarios where ΔT is even or odd, merely one different element can be seen when m_0 increases by 2.

Lemma 6. *$\forall k \in \mathbb{N}$, $\Delta\Omega(m_0 = k+2) > \Delta\Omega(m_0 = k)$.*

Proof.

$$\begin{aligned}\Delta\Omega' &= \Delta\Omega(m_0 = k+2) - \Delta\Omega(m_0 = k)\\ &= \sum_{i=0}^{|k+2+n-1|-1} e^{-ir} - \sum_{i=0}^{|k-n+1|-1} e^{-ir}\\ &= \sum_{i=|k-n+1|}^{|k+n+1|-1} e^{-ir} > 0, \{n,k\} \in \mathbb{N}, n \geq 1\end{aligned}$$

□

According to Lemma 6, $\forall k \in \mathbb{N}$, we have:

$$\begin{aligned}&\Delta\Omega(m_0 = 0) < \Delta\Omega(m_0 = 2) < \ldots < \Delta\Omega(m_0 = 2k)\\ &\Delta\Omega(m_0 = 1) < \Delta\Omega(m_0 = 3) < \ldots < \Delta\Omega(m_0 = 2k+1)\end{aligned} \tag{5.14}$$

Theorem 7. *Multiple-time selection maintains a particular influence more effectively than that of one-shot selection, i.e., for any $t > 0$, $\Delta\Omega(m_0 = t) > \Delta\Omega(m_0 = 0)$*

Proof. Based on Equation 5.14, I only need to proof $\Delta\Omega(m_0 = 1) > \Delta\Omega(m_0 = 0)$. Assume that the path length between two active nodes has an equal chance to be even

Table 5.4: Value Comparison for ΔT and $\Delta \Omega$ (n = 2h)

	v_1	...	v_h	v_{h+1}	v_{h+2}	...	v_{2h}
$\Delta T(n=2h, m_0=0)$	$2h-1$		1	1	3		$2h-1$
$\Delta \Omega(n=2h, m_0=0)$	$\sum_{i=0}^{2h-2} e^{-ir}$		$\sum_{i=0}^{0} e^{-ir}$	$\sum_{i=0}^{0} e^{-ir}$	$\sum_{i=0}^{2} e^{-ir}$		$\sum_{i=0}^{2h-2} e^{-ir}$
$\Delta T(n=2h, m_0=1)$	$2h$		2	0	2		$2h-2$
$\Delta \Omega(n=2h, m_0=1)$	$\sum_{i=0}^{2h-1} e^{-ir}$		$\sum_{i=0}^{1} e^{-ir}$	0	$\sum_{i=0}^{1} e^{-ir}$		$\sum_{i=0}^{2h-3} e^{-ir}$

Table 5.5: Value Comparison for ΔT and $\Delta \Omega$ (n = 2k+1)

	v_1	...	v_k	v_{k+1}	v_{k+2}	...	v_{2k+1}
$\Delta T(n=2k+1, m_0=0)$	$2k$		2	0	2		$2k$
$\Delta \Omega(n=2k+1, m_0=0)$	$\sum_{i=0}^{2k-1} e^{-ir}$		$\sum_{i=0}^{1} e^{-ir}$	0	$\sum_{i=0}^{1} e^{-ir}$		$\sum_{i=0}^{2k-1} e^{-ir}$
$\Delta T(n=2k+1, m_0=1)$	$2k+1$		3	1	1		$2k-1$
$\Delta \Omega(n=2k+1, m_0=1)$	$\sum_{i=0}^{2k} e^{-ir}$		$\sum_{i=0}^{2} e^{-ir}$	$\sum_{i=0}^{0} e^{-ir}$	$\sum_{i=0}^{0} e^{-ir}$		$\sum_{i=0}^{2k-2} e^{-ir}$

or odd. In other words, $P(n=2h) = P(n=2k+1)$, where $k, h \in \mathbb{N}$. The values of ΔT and $\Delta \Omega$ under different parameters are listed and compared in Tables 5.4 and 5.5. Then, Equations 5.15 and 5.16 can be obtained.

$$\begin{aligned} &\Delta \Omega(n=2k+1, m_0=1) - \Delta \Omega(n=2h, m_0=0) \\ &= 2\sum_{i=0}^{2h} e^{-ir} + \ldots + 2\sum_{i=0}^{2k-2} e^{-ir} + \sum_{i=0}^{2k} e^{-ir} \end{aligned} \tag{5.15}$$

$$\begin{aligned} &\Delta \Omega(n=2h, m_0=1) - \Delta \Omega(n=2k+1, m_0=0) \\ &= -(\sum_{i=0}^{2h-1} e^{-ir} + 2\sum_{i=0}^{2h+1} e^{-ir} + \ldots + 2\sum_{i=0}^{2k-1} e^{-ir}) \end{aligned} \tag{5.16}$$

Suppose $h > k$, then by adding Equation 5.15 to Equation 5.16, we can obtain:

$$
\begin{aligned}
&\left(\Delta\Omega(n=2h, m_0=1) + \Delta\Omega(n=2k+1, m_0=1)\right) \\
&- \left(\Delta\Omega(n=2k+1, m_0=0) + \Delta\Omega(n=2h, m_0=0)\right) \\
= {}& e^{-2hr} + \ldots + e^{-2(2k-2)r} + e^{-2kr} \\
&- \left(e^{-(2h+1)r} + \ldots + e^{-(2k-1)r}\right) \\
= {}& \left(e^{-2hr} - e^{-(2h+1)r}\right) + \ldots \\
&+ \left(e^{-(2k-2)r} - e^{-(2k-1)r}\right) + e^{-2kr} \\
&> e^{-2kr} > 0
\end{aligned}
\tag{5.17}
$$

The same proof can be applied when $h \leq k$. Therefore, $\Delta\Omega(m_0 = 1) > \Delta\Omega(m_0 = 0)$. □

5.5 Experiments and Analysis

Three major experiments are conducted for this research work. The first one aims to compare the difference in influence impact between one-shot and multiple-time investment. The second experiment evaluates the performance of the TIH algorithm. In the third experiment, I further compare one-shot selection against multiple-time selection by exploring the variations of selected seeds based on ATID model.

5.5.1 Experiment Setup

Datasets. In the experiments, the following three datasets are used.

- **Ego-Facebook**[5] dataset, collected by McAuley et al. (2012) using a Facebook application, which is archived in Stanford Large Network Dataset Collection. It contains profile and network data from 10 ego-networks, consisting of 193 circles, 4,039 users and 88,234 edges.

[5] http://snap.stanford.edu/data/egonets-Facebook.html

- **Email-Enron**[6] dataset, which covers all the email communication. It has been posted to the web by the Federal Energy Regulatory Commission (Klimt & Yang, 2004). The Enron email network has 36,692 nodes and 367,662 Edges. To diminish the computing time, a sub-graph with 10k nodes is captured for the experiment.

- **Wiki-Vote**[7] dataset, which incorporates administrator elections and votes history data from 3 January 2008. There are 2,794 elections with 103,663 total votes and 7,066 users participating in the elections. Nodes refer to Wikipedia users and edges represent votes from one user to another (Leskovec et al., 2010b).

System Setup. The social context is simulated by creating a number of user agents based on the public datasets. Each user agent manages its local information, including a friendship list, a repository and historical records. I assume a hypothesis influence message is supposed to be maintained and each agent has a different tendency of posting this message. In the meanwhile, the reporting agent is responsible for monitoring the entire multi-agent system and collecting global information. The system has three types of states as follows:

- **Evolve**: user agents perform actions, incorporating accessing the repository, reading the message and making decisions (share the post or not) based on both past experiences and timeliness degrees.

- **Pause**: the entire system pauses, and stops functioning temporarily. This state allows seed selection algorithms to identify influential users and select seeds based on the current network status. In other words, further investment happens at this point. The system evolution resumes as soon as the seed selection is completed.

[6]https://snap.stanford.edu/data/email-Enron.html
[7]https://snap.stanford.edu/data/wiki-Vote.html

Table 5.6: Experiment Parameters

Parameter	**Value(s)**
Fixed time steps for seed selections	100
Fixed time steps in total	150
Number of seeds to be selected for each selection points	25, 5, 1
The interval (time steps) of seed selection	100, 20, 4
Seed set size	25
attenuation constant r	0.1
General action frequency of user agents (times per second)	5

- **Stop**: All the user agents decompose, and the system terminates.

By setting up the system, the parameters for the experiments are given in Table 5.6. I assume that the observations of network evolution are within a fixed interval, and the same amount of seeds are supposed to be selected at each seed selection point. To reduce the bias of measuring the performance of different strategies, additional time steps, i.e., 50 time steps in the experiments, are given after the final seed selection for the influence dissemination and attenuation. Furthermore, the budget is limited, in other words, the seed set size is limited. The overall action frequency of user agents controls the speed of network evolution.

Evaluation Metrics. As introduced in Section 5.3.4, three major evaluation metrics are taken into consideration, i.e., GTD, GCTD and GAC, which have been explained and formulated in Equations 5.4, 5.5 and 5.6, respectively. GCTD and GAC were applied in both Experiment 1 and Experiment 2 for comparing the performance of different selection strategies. GTD has been mainly utilised in Experiment 1 for tracking the variation of timeliness degree of a particular influence message in different time steps. In Experiment 3, some distance indices were facilitated to measure the variation of seed sets.

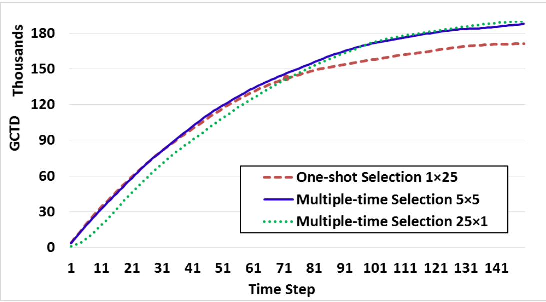


Figure 5.3: Rank-based (GCTD)

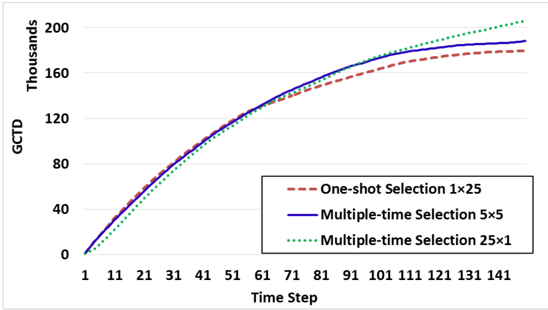


Figure 5.4: DDH Selection (GCTD)

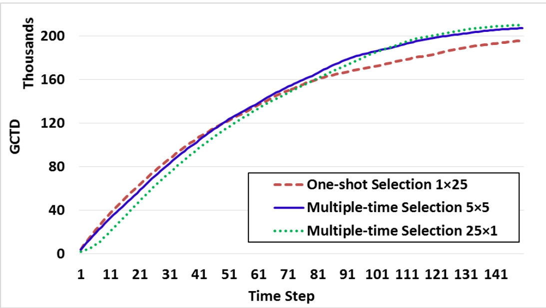


Figure 5.5: TIH Selection (GCTD)

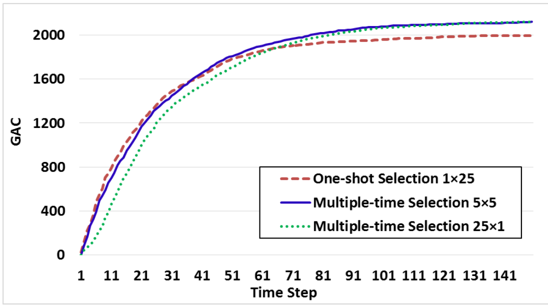


Figure 5.6: Rank-based (GAC)

5.5.2 Experiment 1: One-shot vs multiple-time selection

Experiment 1 compares one-shot investment against the multiple-time by facilitating different seed selection algorithms, i.e., rank-based, DDH and the TIH selection. In this experiment, the Ego-Facebook dataset is applied for the explorations. The notations of selection approaches are listed in Table 5.7.

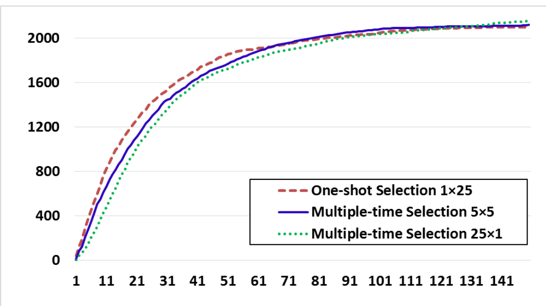


Figure 5.7: DDH Selection (GAC)

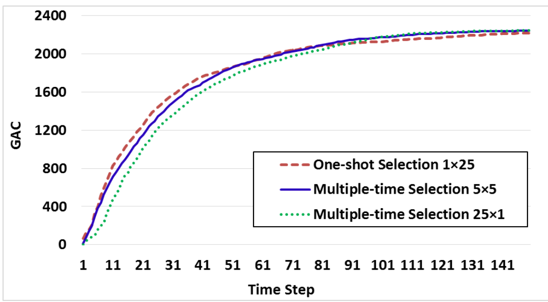


Figure 5.8: TIH Selection (GAC)

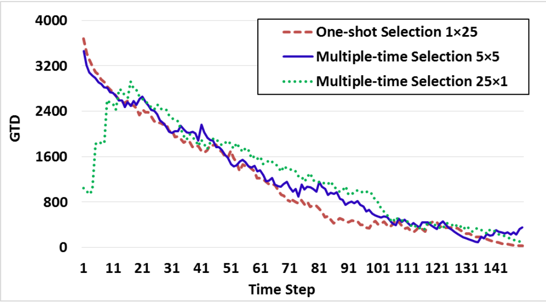


Figure 5.9: Rank-based (GTD)

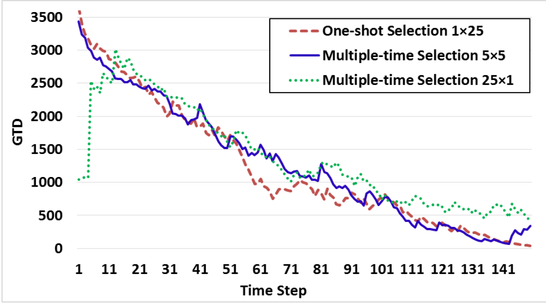


Figure 5.10: DDH Selection (GTD)

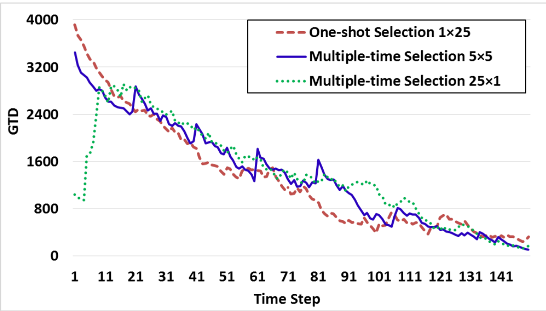


Figure 5.11: TIH Selection (GTD)

Table 5.7: Notations of Selections

Notation	**Meaning**
1×25	One-shot selection, 25 seeds
5×5	5-time selection, select 5 each time
25×1	25-time selection, select 1 each time

As we can observe from Figures 5.3, 5.4 and 5.5 that multiple-time selections can produce higher GCTD. The gap between one-shot selection and multiple-time selection turns out to be evident over time. 5×5 and 25×1 give pretty close performance, but 25×1 shows slightly better, especially after 100 time steps when the selections are completed. By comparing the GAC in Figures 5.6, 5.7 and 5.8, the multiple-time selection also outperforms the one-shot selection. One-shot selection demonstrates a rapid influence activation coverage, but unfortunately it loses the leading position halfway.

Based on the results in Figures 5.3 - 5.8, we can observe that with the same budget, increasing the frequency of investments generally carries out higher GCTD and GAC, subject to additional time for influence diffusion and attenuation provided. This is due to the reason that the last-round investment in multiple-time selections is not supposed to give much credit without additional time for influence spread and decay. The results also explicitly reveal that multiple-time selections target the reward in the long-run, but may yield short-term performance. If an organisation intends to maintain an influence by considering both effectiveness and time required for GCTD to reach a certain level, selection strategies with extremely high frequencies (25×1 in the experiment) may not be advocated, since it takes longer time to reach the maximum GCTD. Whereas, 5×5 balances the trade-off between time and GCTD, which is a better option under such a scenario. The same rule also applies to GAC.

To drill down into the details, I explore timeliness variations of the influence message after adopting different selection strategies in Figures 5.9, 5.10 and 5.11. One-shot selection has the highest starting point, but it declines faster than that of 5×5 and generally falls behind the others after around 20 time steps. Obvious spikes can be observed in multiple-time selections and appear to be more prominent in the TIH 5×5. Whereas, 25×1 demonstrates a different pattern. It climbs to the peak point, which is higher than that of the other two selection approaches, then falls gradually. The organisation expects a sharp upward trend after each investment. However, this is not guaranteed based on the results. For example, no obvious increase can be observed at time steps 60 and 20 of rank-based 5×5 and DDH 5×5, respectively. In contrast, the TIH 5×5 sees an evident spike after each investment.

5.5.3 Experiment 2: The TIH Seed Selection Evaluation

Experiment 2 aims to evaluate the performance of the TIH algorithm. I compare the proposed TIH algorithm against state-of-the-art algorithms. Since the diffusion model

Table 5.8: Seed Selection Performance Comparison

Social Network	Algorithm	Metrics	One-shot Selection 1×25	Multiple-time Selection 5×5	Multiple-time Selection 25×1
Ego-Facebook	TIH	GCTD	**195,951**	**207,115**	**209,675**
Ego-Facebook	TIH	GAC	**2,222**	**2,242**	**2,249**
Ego-Facebook	RANK	GCTD	170,966	187,696	189,968
Ego-Facebook	RANK	GAC	1,996	2,124	2,118
Ego-Facebook	DDH	GCTD	181,994	190,653	205,774
Ego-Facebook	DDH	GAC	2,097	2,123	2,151
Ego-Facebook	Random	GCTD	168,733	175,899	188,300
Ego-Facebook	Random	GAC	1,889	1,988	2,003
Email Eron	TIH	GCTD	**341,418**	**358,722**	**384,861**
Email Eron	TIH	GAC	**4,307**	**4,445**	**4,331**
Email Eron	RANK	GCTD	328,026	352,744	362,992
Email Eron	RANK	GAC	4,082	4,391	4,365
Email Eron	DDH	GCTD	338,803	355,452	373,218
Email Eron	DDH	GAC	4,227	4,255	4,492
Email Eron	Random	GCTD	324,994	337,380	338,269
Email Eron	Random	GAC	4,181	4,189	4,196
Wiki Vote	TIH	GCTD	**254,710**	**267,810**	**272,292**
Wiki Vote	TIH	GAC	**2,868**	**3,001**	**2,953**
Wiki Vote	RANK	GCTD	247,659	264,417	267,213
Wiki Vote	RANK	GAC	2,826	2,944	2,878
Wiki Vote	DDH	GCTD	249,977	265,950	270,906
Wiki Vote	DDH	GAC	2,843	2,954	2,829
Wiki Vote	Random	GCTD	247,626	253,349	257,599
Wiki Vote	Random	GAC	2,813	2,843	2,824

is probabilistic based, the results are obtained by averaging multiple trials. To reduce the bias, I evaluate the TIH algorithm by using the three datasets mentioned previously, i.e., Ego-Facebook, Email-Enron and Wiki-Vote.

The experimental results are demonstrated in Table 5.8. It can be seen that the TIH outperforms the others in all the three datasets. By using any selection strategy, the TIH performs the best in terms of GCTD and GAC.

Another intriguing finding from the experimental results is concerning the relationship between GCTD and GAC. More specifically, given the same budget, GCTD rises with the increment of selection trails. In general, GAC gains when GCTD increases. However, by adopting rank-based 25×1, the GAC yields that of the 5×5, though GCTD rises. This phenomenon implicitly shows that the outcome of influence maintenance is not always in accordance with that of the influence maximization. Whereas, the relationship between GCTD and GAC tends to be affected by the applied business strategies. In other words, the strategies created for long-term marketing can possibly

Table 5.9: Network Properties and Seed Sets Variations

Dataset	Average Path Length	one-shot selection vs. multiple-time selection 5×5			one-shot selection vs. multiple-time selection 25×1			multiple-time selection 5×5 vs. multiple-time selection 25×1		
		Jaccard	Dice	Sequence	Jaccard	Dice	Sequence	Jaccard	Dice	Sequence
Email Enron	3.123	0.442	0.284	0.365	0.498	0.332	0.424	0.094	0.051	0.113
Wiki Vote	3.247	0.622	0.452	0.540	0.730	0.576	0.650	0.424	0.270	0.394
Ego-Facebook	3.693	0.768	0.624	0.631	0.792	0.656	0.668	0.477	0.315	0.383

suppress the short-term growth of the product adoptions.

5.5.4 Experiment 3: Seed Set Variation Analysis

With the same budget, different selection approaches inevitably produce different seed sets. To understand the outcome of various strategies, in this experiment, I further compare one-shot selection against multiple-time selection by exploring the variations of selected seeds based on ATID model. The TIH algorithm has been applied for the seeding procedures in three social networks mentioned previously.

Three evaluation metrics are adopted for measuring the distance (*referring to variation or dissimilarity*) between any two seed sets, i.e., Jaccard distance $d_{jcd}(A_1, A_2)$, Dice dissimilarity $d_{dic}(A_1, A_2)$ and sequential distance considering the index of the elements $d_{sqc}(A_1, A_2)$, which are formulated in Equations 5.18, 5.19 and 5.20, respectively. In these three equations, A_1 and A_2 denote two different seed sets, having the same cardinality, i.e., $A_1 \neq A_2, |A_1| = |A_2|$. $I(c|A_1)$ refers to the index of element c in set A_1.

$$d_{jcd}(A_1, A_2) = 1 - \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|} \tag{5.18}$$

$$d_{dic}(A_1, A_2) = 1 - \frac{2|A_1 \cap A_2|}{|A_1| + |A_2|} \tag{5.19}$$

$$d_{sqc}(A_1, A_2) = \frac{1}{|A_1|}\Big(\sum_{c \in A_1 \cap A_2} \frac{|I(c|A_1) - I(c|A_2)|}{|A_1|} + |A_1 \setminus A_2|\Big) \tag{5.20}$$

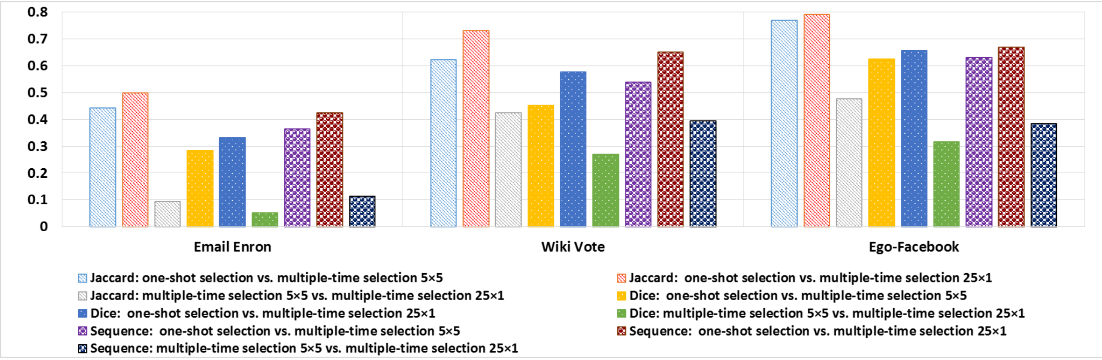


Figure 5.12: Seed Set Variation Comparison under Different Strategies

As the influence diffusion appears to be probabilistic-based, different sets of the nodes could be selected by using the same algorithm. To reduce the bias, results are averaged over multiple trials. Figure 5.12 compares the variations of the seed sets produced by using different strategies. It explicitly shows that the seed-set variations between one-shot selection and multiple-time selection appear to be more prominent when having a higher frequency of selections. Two multiple-time selection approaches, i.e., 5×5 and 25×1, share larger overlapping seeds than that of one-shot selection.

To investigate the correlations between network properties and seed-set variations, I list the detailed results in Table 5.9, where *"Average Path Length" (APL)* refers to the average number of steps along the shortest paths for all possible pairs of nodes. APL is one of the key metrics to measure the transitivity of the network (Peres, 2014). A shorter APL generally indicates that less time is required for any influence travelling from one node to another.

It can be seen from Table 5.9 that a greater average path length corresponds to a higher seed-set variation. The reason behind is that in shorter APL networks, influences become relatively easier to reach any node, thus $e^{-r\cdot(m-b)}$ in Equation 5.1 appears to be lower as b shrinks. Subsequently, timeliness gain turns out to be less prominent. Therefore, based on Equations 5.11 and 5.12, the TIH algorithm has a higher chance to carry out similar seed sets under such circumstances.

5.5.5 Discussion

I simulated a social environment and the process of influence maintenance in a social network. Through the experiments, the advantages of applying ATID to model the influence propagation process are demonstrated. Two critical factors required by the influence maintenance can be presented clearly in ATID model, i.e., the temporal feature of the social network and the status of a particular influence. Furthermore, the seed-set variations are compared after applying different selection approaches. I also evaluate the effectiveness of various seed-selection algorithms in maintaining an influence. The TIH algorithm surpasses some selected traditional selection algorithms on three different datasets.

More importantly, three empirical laws can be drawn from the experimental results.

- Given the same budget, the multiple-time investment is generally more beneficial for achieving the long-lasting influence of a particular product than that of the one-shot investment.

- Influence maintenance is not always in accordance with that of the influence maximization. In other words, sustaining a long-term impact of a particular influence cannot ensure a large fraction of activation coverage; the long-term marketing strategies may hinder the profit spikes.

- Seed-set variation is not only associated with the frequency of selections, but also affected by the network property. A greater average path length of social networks leads to a higher seed-set variations.

5.6 Summary

In this chapter, I systematically studied the influence maintenance problem, which targets the long-term and sustainable business goals. To the best of my knowledge, this is the first full research work that characterises the influence maintenance in social networks. The distributed influence diffusion model, i.e., ATID model, can also pave the way in exploring influence propagation social pheromone, since it concentrates on modelling the agent's personalised traits and behaviours, tracking the temporal feature of a social network, as well as the status of influence messages. Many features of both individuals and influences can be enabled in ATID model when analysing the social influence diffusion phenomenon. I have also proposed a novel seed selection algorithm, i.e., the TIH, which is capable of maintaining long-term influence effectively. Extensive experiments are conducted, and the empirical results show that the proposed model is capable of enhancing long-term influence. Given the same budget and limited time frame, multiple-time investment is superior to one-shot investment in terms of influence maintenance. Moreover, the experimental results also explicitly show that the TIH performs better than the other traditional selection algorithms by considering GCTD and GAC. I believe that the findings can shed light on the understanding of influence maintenance for long-term marketing.

This chapter mainly answers the Research Question 2 mentioned in Chapter 1. The research work of this chapter has been published in (W. Li et al., 2017) and (W. Li, Bai, Zhang & Nguyen, 2018a).

The next chapter aims to model multiple influences diffusion and investigate adverse influence minimisation problem from an agent-based perspective.

Chapter 6

Modelling Multiple Influences Diffusion in On-line Social Networks

In the real-world, various influences normally coexist in the same context and have subtle relations, such as supportive, contradictory and competitive relations, affecting the users' decisions of adopting any innovations. Therefore, modelling diffusion process of multiple influences is an important, yet challenging research question.

By extending the generic agent-based framework proposed in Chapter 3, in this chapter, a distributed approach has been proposed to model the diffusion process of multiple influences in social networks. The proposed model has been applied in the undesirable influence minimisation problem, where the time series is taken into consideration. The experimental results show that the model can be utilised to minimise the adverse impact of a certain influence by injecting other influences. Furthermore, the proposed model also sheds light on understanding, investigating and analysing multiple influences in social networks.

6.1 Overview

In real-world, multiple influences of various topics normally coexist within the same context, and their divergent relationships impact each other regarding individual's influence acceptance. Intuitively, influences of the same topic can be either supportive or contradictory to each other. Influences usually propagate in the presence of a wide variety of rich content, i.e., influence messages, which can be images, videos, long articles or even short comments, conveying the opinions or ideas towards the one or more innovations. When multiple influence messages with the same opinion flood into one's friend circle, he or she has a high tendency of adopting the opinion. Whereas, individuals usually struggle with taking a side when adverse opinions of the corresponding topic emerge. In addition, different influences appear to be associated with each other indirectly by competing for the 'common resources', i.e., the users' attention. More specifically, nobody can take care of all the influence messages due to the limited vigour of human nature. Instead, individuals usually get attracted by the information that they care most. In other words, each individual possesses a finite capacity of considering and absorbing the impact of influences, and the corresponding attention is always focused on particular influence messages. Meanwhile, the existing information keeps fading out of the public attention, especially when other significant influences are injected into the same context. This feature becomes more prominent in time-sensitive social networks, such as microblogging platforms (Castillo, Mendoza & Poblete, 2013).

There are several motivations to model and analyse multiple influences diffusion in social networks. A non-trivial incentive is to investigate effective approaches for rationally alleviating or even suppressing the impact of a particular undesirable influence message, e.g., a rumour, or negative opinions towards a social event. Based on the contemporary research work, when any adverse opinions are propagating through a social network, some researchers recommend blocking a particular group of nodes (S. Wang et al., 2013) or a bunch of links (Kimura et al., 2008) from on-line social networks to control the influence contaminations. However, these approaches can only be facilitated to a few types of networks, such as virus or epidemic networks (Basaras, Belikaidis, Maglaras & Katsaros, 2016). As for those ordinary or customer-based social networks, any user or affiliation link is not supposed to be blocked or removed. Furthermore, the topological structure of a network is out of control in most cases. Therefore, approaches without restricting users' behaviours or altering the networked structure are highly recommended. This scenario frequently arises in the real world: when a piece of sensational news disseminates fast, public attention tends to be diverted by other news eventually. Inspired by this social phenomenon, the subtle relationships among multiple influences and the individualised features of users can be utilised to achieve the undesirable influence minimisation.

In this chapter, I proposed an Agent-based Multiple Influences Diffusion (AMID) model to analyse multiple influences propagation in social networks by considering their relationships. Each user's personalised traits, preferences, behaviours and social context have been taken into consideration. Influential relationships among the entities, including *user and user, user and influence, influence and influence,* have been considered in the proposed model. Furthermore, undesirable influence minimisation is utilised as a typical application of the proposed model. Extensive experiments have been conducted, and the results suggest that by using the proposed model, introducing external influences can suppress the adverse influence effectively.

The remainder of this chapter is structured as follows. Section 6.2 reviews the literature related to this research work. Section 6.3 introduces the modelling of multiple influences diffusion using ABM and the formal definitions. Section 6.4 systematically elaborates the influential relationships modelling. In Section 6.5, experiments and experimental results are presented by using a typical application, i.e., undesirable influence minimisation. The summary of this chapter is given in Section 6.6.

6.2 Related Work

6.2.1 Influence Diffusion

As clarified in Chapter 2 that Domingos and Richardson (2001) attempt to mine the value of customers in social networks by considering influence diffusion. The seminal research works by Kempe et al. (2003) have been extensively extended.

However, in nearly all the extended research works in the field of influence maximization, only a single influence is considered. In other words, most studies focus on the adoption of a particular product or opinion, while other influences in the same context have been ignored. With an exception, Tang et al. (2009) propose topical affinity propagation to model the topic-level social influence, which can identify the experts in different topics and measure the strength quantitatively. Nevertheless, Tang's work is developed based on the assumption that no dependencies are presented among the various topical influences. Different from the aforementioned research work, I model the influence propagation by considering the impacts and relationships among the multiple influences.

6.2.2 Competitive Influence

Competitive influence turns out to be one of the extended works in the field of influence maximization. Substantial related studies have been reviewed in Subsection 2.2.2. Three major limitations are spotted as follows.

- The studies focus on the influential competitive relationships among the social influence and ignore other factors, such as supportive influences and the impact of other innovations. For example, the introduction of Samsung phone competes with that of the Apple phone for the public attention (H. Wu, Liu, Yue, Huang & Yang, 2015). Whereas, the emergence of Samsung 3D Glasses tends to be a supportive influence for Samsung phones due to its compatibility, but becomes a subtle force of discouraging the adoption of Apple phones.

- A major assumption in most current approaches is that each user possesses a single adoption when given multiple choices, e.g., various products from different firms. Whereas, users may adopt the multiple products or innovations, e.g., a customer can purchase both Samsung phone and Apple phone.

- Nearly all of the research work extends the IC or LT model to accommodate the competitive influence dissemination. However, due to the nature of both models, i.e., centralised influence diffusion models (W. Li, Bai, Jiang & Zhang, 2016), the extended IC and LT models can neither capture the dynamics of social networks nor track the long-term trend of a social network driven by influence propagation (W. Li, Bai & Zhang, 2016a; W. Li et al., 2017).

To cover the limitations mentioned above, my study models multiple influences diffusion by considering the various corresponding relationships. An agent-based diffusion model is utilised to capture the evolutionary trend of a social network, as well as the individual's features and behaviours. Thus, the multiple adoptions of different innovations by a particular user at different time steps can be enabled.

6.2.3 Negative Influence Minimisation

Many researchers explore the approaches to minimise the adverse impact of a particular existing influence in a social network. As mentioned in Subsection 2.2.2 that most studies attempt to block an influence in a very straightforward way, i.e., altering the structure of a social network. Such kind of approaches can only be applied based on the

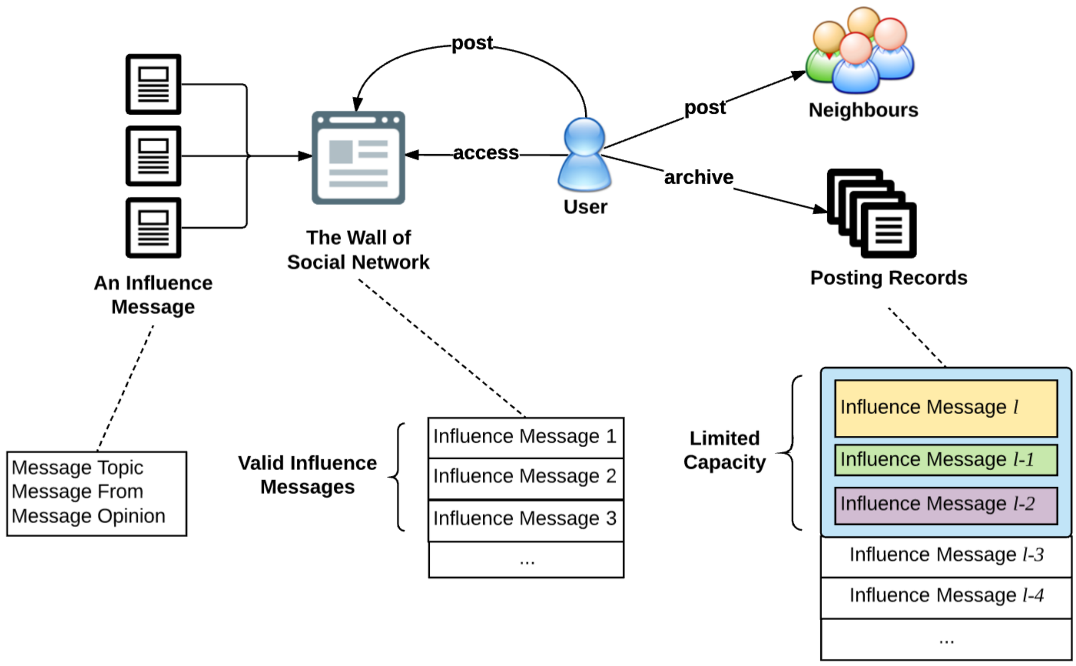


Figure 6.1: The Framework of an Agent-based Multiple Influences Diffusion Model

assumption that the organisation is authorised to manage network topological structures. However, in reality, the modifications are generally not applicable.

Some other studies propose competitive influence models to suppress the adverse impact by introducing the opposite influence only. However, the influential effects originated from other influences are neglected, and these ‘irrelevant’ influences can be even more powerful in distracting users from focusing one opinion. Moreover, the individual’s features, such as preference and information intake capacity, are not taken into consideration. These factors can affect a user’s influence acceptance to a large extent.

By contrast, I attempt to alleviate the negative influence minimisation problem in a real situation when multiple influences coexisted in the same social context. Furthermore, three possible relationships among the influences, i.e., support, competitiveness and irrelevance, are taken into consideration.

6.3 Multiple Influences Diffusion

6.3.1 An Agent-based Multiple Influences Diffusion Model

To analyse and model multiple influences diffusion in social networks, traditional propagation models, such as the IC model and the LT model, only concentrate on the diffusion process and activation status of each node, ignoring the interactions between users and influence messages, as well as the co-actions among influences. Motivated by this background, a novel propagation model is necessarily required.

The AMID proposed in this thesis models the propagation process in a decentralised manner. In the AMID model, users have been modelled as a set of interactive autonomous agents that possess their own personalised traits and behaviours. Meanwhile, influence messages appear to be another type of entities in the same context which can be interacted with the agents directly. From a macroscopic point of view, the influence diffusion demonstrates a networked evolutionary pattern driven by the individuals' actions, i.e., interactions with various influences.

Figure 6.1 shows the framework of the proposed AMID model. An ordinary influential behaviour of a particular user agent incorporates only two simple sequential steps, i.e., reading messages from the wall of on-line social networks and getting influenced by posting an influence message. More specifically, in time-sensitive social networks, such as Twitter, various influence messages of different topics are constantly posted to a user's wall. He or she tends to be influenced by the received influence messages based on the interests and peer trust relationships (Hsu & Lin, 2008). Subsequently, an influence message is not only posted to the adjacent neighbours, but also archived as one of the posting records, reflecting the user's latest interests. In this model, three major attributes of an influence message are taken into consideration, including the topic, delivered from and the opinion.

I assume that only a particular number of the latest messages are regarded as valid, accessible information, and each user has a limited and different-size capacity (vigour) for taking care of the influence messages. Once a new message has been posted, a certain amount of space is occupied. The space it takes depends on the peer trust and the user's interests. In addition, the old ones are fading out of the user's attention.

6.3.2 Formal Definitions

Definition 6.1: A User Agent $v_i, (v_i \in V)$ refers a node in a time-sensitive social network $G = (V, E)$, where $V = \{v_1, ..., v_n\}$ denotes a set of agents and E represents a set of edges, $E = \{e_{ij} | 1 \le i, j \le n\}, i, j \in \mathbb{N}^+, \{v_i, v_j\} \subseteq V$. User agent v_i has a set of neighbours $\mathit{\Gamma}(v_i)$, and such affiliation information is maintained by the agent locally. If v_j is a neighbour of v_i, then $\{e_{ij}\} \subseteq E, v_j \in \mathit{\Gamma}(v_i)$. While E_{v_i} indicates the edge set connected with v_i, where $E_{v_i} = \{e_{ij} | v_i \neq v_j \wedge v_j \in \mathit{\Gamma}(v_i)\}$. In addition, each user agent has a local view, which covers all its neighbours and the corresponding posting records (refer to Definition 6.4).

Definition 6.2: An Influence Message $msg_p, (msg_p \in M)$ in general refers to a communication containing some information, which potentially affects users' opinions and behaviours, where $M = \{msg_1, msg_2, ..., msg_k\}$ denotes the influence message set in a social network. $msg_p^{(v_j \to v_i)}$ refers to msg_p delivered from v_j to v_i, subject to $v_j \in \mathit{\Gamma}(v_i)$.

Given a finite number of n influence topics $T = \{\tau_1, \tau_2, ..., \tau_n\}$, each influence message is associated with all the topics with different membership degrees. Therefore, the relationships among influence message msg_p and the topics T can be represented as a fuzzy set:

$$
\begin{aligned}
S_{msg_p} &= (T, m_p) \\
&= m_p(\tau_1)/\tau_1 + m_p(\tau_2)/\tau_2 + ... + m_p(\tau_n)/\tau_n,
\end{aligned} \tag{6.1}
$$

where $m_p(.)$ is a membership function, and $m_p(\tau_k) \in [0,1], k \in [1,n]$ quantifies τ_k's membership degree of topic τ_k in the fuzzy set. An influence message msg_p can be expressed by using a two-tuple: $msg_p = (S_{msg_p}, o_p)$, where $o_p \in \{0,1\}$ refers to the general opinion of msg_p, $o_p = 1$ means positive, and negative otherwise.

Definition 6.3: Social Network Wall $W_{t_m}^{(v_i)}$ refers to a dynamic area on a time-sensitive social network profile or home page of user agent v_i at time step t_m, displaying the latest n influence messages posted by $\Gamma(v_i)$ in a reverse chronological order. $W^{(v_i)}$ generally represents v_i's wall in a predefined context. Mathematically, $W_{t_m}^{(v_i)} = \langle msg_p^{(v_j \to v_i)} | v_j \in \Gamma(v_i), msg_p \in M \rangle$ describes a sequential vector, incorporating n messages delivered to v_i. User agent accesses the messages from $W_{t_m}^{(v_i)}$ at time t_m and determines which message to be posted.

Definition 6.4: Posting Records $PR_{t_m}^{(v_i)}$ describes a collection of historical influence messages delivered by user agent v_i. Similar to social network wall, the posting records also can be represented by using a sequential vector $PR_{t_m}^{(v_i)} = \langle msg_p^{(v_i \to v_j)} | v_j \in \Gamma(v_i), msg_p \in M \rangle$, which reflects v_i's preferences. For simplification purpose, $PR^{(v_i)}$ denotes v_i's posting records in a predefined context.

Definition 6.5: Capacity $c^{(v_i)}$ is defined as v_i's capability to take care of the influence messages, which implies the limited vigour or attention of a user agent. When an influence message $msg_p^{(v_j \to v_i)}$ arrives or pre-exists, a particular amount of capacity $\widehat{A}(msg_p^{(v_j \to v_i)})$ is supposed to be occupied if the message has been accepted (see Relationship 3). In addition, the old influence messages are suppressed and fading out of the user's attention.

6.4 Influential Relationship Modelling

Relationship 1: User and User. Users are far more likely to be influenced by the people they know and trust, rather than from strangers or systems (Sinha & Swearingen, 2001). In the current setting, $TR(v_i, v_j)$ describes the trust relationship established between two users, i.e., truster v_i and trustee v_j. In this chapter, the definition of trust of (Jøsang, Hayward & Pope, 2006) is borrowed, and it can be interpreted truster's engagement probability respected to the influence messages posted by the trustee.

When user v_i accesses influence message $msg_p^{(v_j \rightarrow v_i)}$, there is a possibility that $msg_p^{(v_j \rightarrow v_i)}$ will be posted (or shared) by v_i. If v_i posts the same message, we say v_i trusts v_j on the topics of $msg_p^{(v_j \rightarrow v_i)}$. Therefore, the trust value of v_i to v_j can be estimated from the number of times that v_i shares v_j's posting records. As each user agent is able to access the posting records of its neighbours, the trust relationships are obtained by individuals locally.

There are two possibilities of an action towards an influence message, i.e., post or not post. Therefore, the probability density over these binary events can be expressed as Probability Density Function (PDF), i.e., $beta(\alpha, \beta)$. A simplified subjective logic approach in (Jøsang et al., 2006) can be applied to estimate the trust degree. Here, the transitive trust is not considered. I denote s, u, a as the number of posted, unshared messages, and the priori, which is the default value that can be assigned to users.

Then α and β can be determined as:

$$\alpha = s + 2a, \qquad \beta = u + 2(1 - a) \tag{6.2}$$

As only two possible responses exist in the environment, a can take 0.5. With s shared and u unshared messages, the a posteriori distribution is beta PDF with $\alpha = s + 1$ and $\beta = u + 1$. To capture the dynamic sharing behaviours, a forgetting factor λ is used to weight a message at time t_{now}:

$$f_{msg_p} = \lambda^{(t_{now} - t_{msg_p})}, \tag{6.3}$$

where $0 \leq \lambda \leq 1$, t_{msg_p} is the time at which the message was posted. After that, the trust relationship is measured using all posts related to v_i and v_j. I denote the cumulative post and not post rate as $\bar{s}$ and $\bar{u}$. They can be aggregated by summing up the weights of the posted and not posted messages, respectively, using the following equations.

$$\bar{s}_{v_i,v_j} = \sum_{msg_h \in PR^{(v_i)}} f_{msg_h^{(v_j \to v_i)}} \tag{6.4}$$

$$\bar{u}_{v_i,v_j} = \sum_{msg_h \in W^{(v_i)} \setminus PR^{(v_i)}} f_{msg_h^{(v_j \to v_i)}} \tag{6.5}$$

The trust relationship between v_i and v_j can be estimated by aggregating the evidence from both users, while the base trust value a is involved in the case that both users have never interacted before. The trust values of v_i to v_j can be obtained by calculating the mean of their distribution:

$$TR(v_i, v_j) = E[beta(\bar{s}_{v_i,v_j} + 1, \bar{u}_{v_i,v_j} + 1)] \tag{6.6}$$

Apply the mean value of beta distribution, Equation 6.6 can then be normalised to:

$$TR(v_i, v_j) = \frac{\bar{s}_{v_i,v_j} + 1}{\bar{s}_{v_i,v_j} + \bar{u}_{v_i,v_j} + 2} \tag{6.7}$$

Relationship 2: User and Influence. User's influence acceptance of a particular influence mainly depends on two major factors, i.e., the peer trust relationships and individual's interests. Similar to an influence message, a user agent's topical level interests can also be expressed as a fuzzy set:

$$\begin{aligned} S^{(v_i)} &= (T, m^{(v_i)}) \\ &= m^{(v_i)}(\tau_1)/\tau_1 + m^{(v_i)}(\tau_2)/\tau_2 + ... + m^{(v_i)}(\tau_n)/\tau_n, \end{aligned} \tag{6.8}$$

where the membership degree $m^{(v_i)}(\tau_k)$ represents v_i's interest towards influence topic τ_k, which can be evaluated by user agents locally based on the past posting records $PR^{(v_i)}$. Thus, $m^{(v_i)}(\tau_k)$ can be formulated in Equation 6.9.

$$m^{(v_i)}(\tau_k) = \frac{1}{|PR^{(v_i)}|} \sum_{msg_p \in PR^{(v_i)}} \frac{m_p(\tau_k) \cdot f(t)}{\sum_{\tau_x \in T} m_p(\tau_x)}, \tag{6.9}$$

where $|PR^{(v_i)}|$ denotes the cardinality of posting records, $m_p(\tau_k), \tau_k \in T$ refers to the membership degree of msg_p, and $f(t)$ is an attenuation function formulated in Equation 6.10.

$$f(t) = e^{-t \cdot k}, k > 0 \tag{6.10}$$

The relationship between user agent v_i and the message msg_p is presented as the Cartesian product of the topical fuzzy set of msg_p and user's interest fuzzy set, which is described in Equation 6.11.

$$\begin{aligned} R(v_i, msg_p) = S_p \times S^{(v_i)} &= (T, \mu_R^{(v_i)}) \\ &= \mu_R^{(v_i)}(\tau_1)/\tau_1 + \mu_R^{(v_i)}(\tau_2)/\tau_2 + ... + \mu_R^{(v_i)}(\tau_n)/\tau_n, \end{aligned} \tag{6.11}$$

The fuzzy relationship $R(v_i, msg_p)$ is a mapping from Cartesian space to the interval, and the strength of the mapping can be expressed by using the membership function $\mu_R^{(v_i)} : S_p \times S^{(v_i)} \to [0, 1]$. Therefore, we can derive the user agent v_i's acceptance to message msg_p sent from neighbour v_j by using Equation 6.12.

$$\begin{aligned} A(msg_p^{(v_j \to v_i)}) &= g(R(v_i, msg_p), TR(v_i, v_j)) \\ &= \gamma R(v_i, msg_p) + (1 - \gamma) TR(v_i, v_j), \end{aligned} \tag{6.12}$$

where $g(.)$ is a weighted average function and γ represents a trade-off factor balancing the peer trust relationship and the individual's interests.

Relationship 3: Influence and Influence. Different from user agents, influences are not capable of interacting with each other directly, but their relations and impacts are mediated by user agents. Individuals have high chances to adopt the opinion strongly supported by most of the adjacent neighbours, which complies a common social phenomenon, i.e., social conformity (Tang et al., 2013). In other words, messages of similar topics with the same opinion are supportive to each other, and contradictory otherwise. As aforementioned, fuzzy set S_p represents the degree of topical belongingness of msg_p. Therefore, to obtain the topical similarity between msg_x and msg_y, i.e., $Sim_T(msg_x, msg_y)$, is equivalent to measure the similarity between fuzzy sets S_x and S_y. The most obvious way of calculating fuzzy sets similarity is based on the distance of their membership degrees (Beg & Ashraf, 2009). Thus, $Sim_T(msg_x, msg_y)$ is formulated in Equation 6.13 by using normalised Hamming distance, namely, one of the most widely used distances for fuzzy sets (Szmidt & Kacprzyk, 2000).

$$Sim_T(msg_x, msg_y) = 1 - \frac{1}{|T|} \sum_{\tau_k \in T} |m_x(\tau_k) - m_y(\tau_k)| \tag{6.13}$$

The comprehensive strength exerting on v_i to accept the opinion of msg_p is formulated in Equation 6.14, where θ denotes the similarity threshold.

$$\begin{aligned}\varphi(msg_p) &= \sum_{msg_q \in W^{(v_i)}} A(msg_q^{(v_j \to v_i)}) \\ &\text{subject to } Sim_T(msg_p, msg_q) \geq \theta, msg_p.o_p = msg_q.o_q\end{aligned} \tag{6.14}$$

Similarly, the comprehensive strength of declining the opinion of msg_p, i.e., $\varphi'(msg_p)$, stems from the similar messages with adverse opinions, thus $\varphi'(msg_p)$ can be formulated in the same way as $\varphi(msg_p)$, but with a different constraint, i.e., $msg_p.o_p \neq msg_q.o_q$. We can derive the probability that v_i accepts the opinion of msg_p by using Equations 6.15 and 6.16.

$$p(msg_p) = 0, \varphi(msg_p) \leq \varphi'(msg_p) \tag{6.15}$$

Otherwise:

$$\begin{aligned}p(msg_p) &= \frac{\varphi(msg_p) - \varphi'(msg_p)}{\varphi(msg_p)} \cdot \frac{\varphi(msg_p) + \varphi'(msg_p)}{\sum_{msg_q \in W^{(v_i)}} \varphi(msg_q) + \varphi'(msg_q)} \\ &= \frac{\varphi(msg_p)^2 - \varphi'(msg_p)^2}{\varphi(msg_p) \cdot \sum_{msg_q \in W^{(v_i)}} \varphi(msg_q) + \varphi'(msg_q)} \\ &\leq \frac{\varphi(msg_p)^2 - \varphi'(msg_p)^2}{\varphi(msg_p)^2} \leq 1\end{aligned} \tag{6.16}$$

As mentioned previously, the influence competitive relations are reflected from the limited capacity of each user agent. Once an influence message $msg_p^{(v_j \to v_i)}$ has been accepted, the amount of occupied capacity can be represented as the normalised value of user acceptance to $msg_p^{(v_j \to v_i)}$, i.e., $\widehat{A}(msg_p^{(v_j \to v_i)})$. In addition, $C(PR^{(v_i)})_{t_m}$ denotes the influence message set drawing v_i's attention at time t_m, which appears to be a subset of $PR_{t_m}^{(v_i)}$, i.e., $C(PR^{(v_i)})_{t_m} \subseteq PR_{t_m}^{(v_i)}$, subject to:

$$\sum_{msg_n^{(v_j \to v_i)} \in C(PR^{(v_i)})_{t_m} \wedge v_j \in \Gamma(v_i)} \widehat{A}(msg_n^{(v_j \to v_i)}) \leq c^{(v_i)} \tag{6.17}$$

Algorithm 7 describes a user agent's response towards an incoming influence message. The inputs include $msg_p^{(v_j \to v_i)}$ and wall $W_{t_m}^{(v_i)}$ at the time step t_m, while the output of the algorithm produces an influence message set that attracts user agent v_i's attention at the following step t_{m+1}. Lines 1-2 calculate the probability of accepting the $msg_p^{(v_j \to v_i)}$ and the influence message set drawing v_i's current attention. Lines 3-4 initialise the variables. Lines 5-7 determine if $msg_p^{(v_j \to v_i)}$ is posted by v_i, and update the posting records by adding msg_p to the head of $C(PR^{(v_i)})_{t_m}$. Lines 8-14 tend to construct the influence message set drawing v_i's attention in time step t_{m+1} by replicating the influence messages until user agent's capacity reaches the limit.

Algorithm 7 Multiple Influences Diffusion Algorithm

Input: $msg_p^{(v_j \to v_i)}, W_{t_m}^{(v_i)}$
Output: $C(PR^{(v_i)})_{t_{m+1}}$

1: Calculate $p(msg_p)$ by using Equations 6.15 and 6.16.
2: Obtain $C(PR^{(v_i)})_{t_m}$ by using In-equation 6.17.
3: Initialise $C(PR^{(v_i)})_{t_{m+1}} = \varnothing$
4: Generate a random decimal $rand$
5: **if** $rand \le p(msg_p)$ **then**
6:     $PR^{(v_i)} := PR^{(v_i)} \cup \{msg_p^{(v_j \to v_i)}\}$
7:     $C(PR^{(v_i)})_{t_m} := \{msg_p\} \cup C(PR^{(v_i)})_{t_m}$
8:     Initialise temp variable $c_{temp}^{(v_i)} := \widehat{A}(msg_p^{(v_j \to v_i)})$
9:     **for** $\forall msg_q \in C(PR^{(v_i)})_{t_m}$ **do**
10:         **if** $\widehat{A}(msg_q) + c_{temp}^{(v_i)} \le c^{(v_i)}$ **then**
11:             $C(PR^{(v_i)})_{t_{m+1}} := \{msg_q\} \cup C(PR^{(v_i)})_{t_{m+1}}$
12:             $c_{temp}^{(v_i)} := c_{temp}^{(v_i)} + \widehat{A}(msg_q)$
13:         **end if**
14:     **end for**
15: **end if**

6.5 Experiments and Analysis

Two experiments have been conducted to evaluate the proposed model. In both experiments, the AMID model has been applied in an extended version of the influence maximization problem (Kempe et al., 2003) by considering how to suppress and minimise the constant impact of a particular influence message or opinion within a fixed time-span (Yao, Zhou, Xiang, Cao & Guo, 2014). The objective of the experiments is set to suppress an undesirable influence by utilising various strategies based on the AMID model. In the experiments, three major types of influences are involved:

- *Irrelevant Influence:* the topics of the influences are not relevant to any of the existing influences. In other words, the influence messages are not topically related at all.
- *Opposite Influence:* the topics of the influences are close to the existing ones but with an adverse opinion.
- *Relevant Influence:* the topics of the influences are strongly related the existing ones, and the opinion appears to be supportive.

The differences between the two experiments are reflected as follows: In the first experiment, I intend to explore and analyse the trend of the undesirable influence after adopting different strategies. Whereas, the second experiment tends to measure and compare the effectiveness of different approaches, including blocking nodes (Kimura et al., 2008), by varying the seed set size and aggregating the results in each time step.

6.5.1 Problem Formulation

Assume that an undesirable influence msg_p is spreading across the social network $G = (V, E)$, and new messages with same opinion keep emerging over time. An organisation aims to suppress the impact of such opinions as much as possible in a fixed time-span, i.e., $[t_0, t_m]$. I regard the targeting influence message/opinion msg_p has been suppressed successfully if msg_p has been faded out of users' attention. Specifically, I leverage **Active Influence Coverage Degree (AICD)** as the evaluation metric, which implies how much the users care about a particular opinion at a specific time step, and the value can be derived from the users' latest posting records. Furthermore, **Cumulative AICD** measures the influence impact within a timespan. Therefore, the problem can be represented as an optimisation problem, expecting to minimise the objective function:

$$min \sum_{t=t_0}^{t_m} \sum_{v_i \in V} \sum_{msg_p \in C(PR^{(v_i)})_t \wedge v_j \in \Gamma(v_i)} \widehat{A}(msg_p^{(v_j \to v_i)}) \tag{6.18}$$

6.5.2 Experiment Setup

Dataset and settings. The experiments have been conducted by using the Facebook-like social network, which is originated from an on-line community for students at the University of California, Irvine. The public dataset is collected by Opsahl and Panzarasa (Opsahl & Panzarasa, 2009), which incorporates 1,899 users and 20,296 directed links.

Since how to estimate individuals' topical interests and how to generate fuzzy sets for a particular influence message are not part of the major purpose of the experiments, thus, to make it simple, dataset is extended by giving the following settings and assumptions:

- The individuals' capacities are randomised by following the Gaussian distribution.
- There are ten pre-defined topics in the social network, i.e., $T = \{\tau_1, \tau_2, ..., \tau_{10}\}$.
- Users' interests towards these ten topics are randomly generated. Peer trust and user's interests are equally important for an individual to accept any influence, i.e., $\gamma = 0.5$.

- There are four pre-existing influence messages in the context, which are not topically related to each other.
- Among the four influence messages, *Influence Message 2* in Figures 2-7 is targeted to be suppressed.
- Each individual's social network wall is initialised by filling with randomised influence messages.
- Measures are not supposed to be taken until the evolution of the network reaches the 30th time step.

Comparison methods. Based on the settings mentioned above, given such a social network with several pre-existing influence messages, users interact and exert influences on each other by disseminating influence messages to the adjacent neighbours. The evolution of the network pauses after some time steps. Next, based on this state, I attempt various strategies to navigate the direction of the networked evolution. Two scenarios are involved in the experiments. **(1)** The social network is under the control of this organisation, having the privileges to manipulate the topological structure of the social network. **(2)** The organisation does not possess any control to the social network. Therefore, any nodes or links are not supposed to be blocked or removed. In the former, I attempt to identify the most negative influencers and block their capabilities of spreading the designated undesirable influence. While, in the latter, three types of influences are supposed to be injected into the same environment to suppress the existing undesirable influence, which are

- the influences topically irrelevant to any of the existing influences
- the influences holding the opposite opinion towards the undesirable influence

- the influences strongly associated with the existing influences but excluding the undesirable influence

The process of identifying influential users are named as *seed selection*; the selected users are called *seed set*; the size of seed set refers to the *budget*. In the experiments, the greedy selection algorithm (Kempe et al., 2003) has been applied for all the approaches. The approach which can minimise the Cumulative AICD of undesirable influence (refer to Equation 6.18) with less budget is regarded as the optimal solution.

6.5.3 Experiment 1

In the first experiment, my objective is to explore and analyse the trend of the undesirable influence after adopting different strategies. As aforementioned, among four pre-existing influence messages, *Influence Message 2* is undesirable and supposed to be minimised. In Figures 6.2 - 6.9, the x-axis represents the networked evolving time steps, and the y-axis denotes the AICD. Various strategies are only supposed to be adopted after the 30th time step when the adverse influence does not fully dominate the network.

Figure 6.2 demonstrates the evolutionary trend of the social network without taking any measures. As we can observe that the undesirable influence message spreads rapidly and dominates the entire social network after 50 time steps. Whereas, others keep fading out of context gradually. During the evolving process, Influence Message 1 seems competitive and shows a spike around the 20th time step, but loses the public attention eventually.

Next, a new influence message is injected into the social network to compete with the existing ones for the resources, expecting that the undesirable influence message could be suppressed. The injected influence is totally independent and not associated with any existing influences in terms of topics. Unfortunately, as we can see from Figure 6.3 that given investment budget as 20 (seed set size), the undesirable influence still attracts most users' attention, though an upper shaking trend is spotted. As illustrated in Figure 6.4, by increasing the seed set size (up to 30) of the same injected influence message, the undesirable influence demonstrates a sharp downward trend and is fading out of the users' attention eventually. In addition, the injected influence dominates the entire social network.

Moreover, I attempt to inject an influence message, which is topically associated with two of the existing influence messages, i.e., Influence Messages 1 and 3. In Figure 6.5, the expected outcome can be achieved with merely ten seeds. In addition, an interesting phenomenon can be observed from Figure 6.5 that the associated influences, i.e., both Influence Messages 1 and 3 rise up when the new influence message has been injected into the social network, and this is due to topical similarities among the three influence messages.

Another ordinary strategy is introducing influences with opposite opinions. According to Figures 6.6 and 6.7, undesirable influence stops expanding and demonstrates a sharp downward trend after injecting an opposite influence. However, the injected message shows different growing tendencies when varying the seed set size. In Figure 6.6, the injected message increases and starts to oscillate when reaching the 45th time step. Meanwhile, Influence Message 3 shows an upper trend from the same point and steadily rises to 400. A higher budget in Figure 6.7 can ensure a relatively smooth increase, though other influences still show a slightly upper trend.

6.5.4 Experiment 2

Experiment 2 tends to measure and compare the effectiveness of different approaches in suppressing an undesirable influence, including injecting irrelevant influences, opposite opinions, relevant influences and blocking nodes, by employing the proposed AMID model.

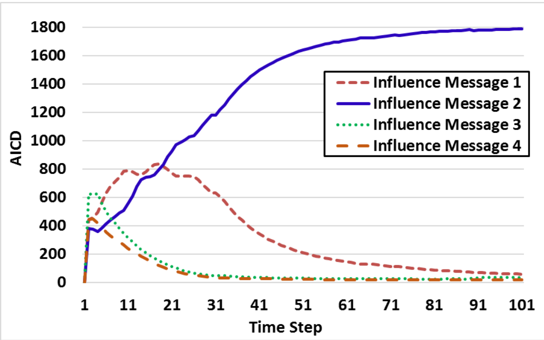


Figure 6.2: No Strategies Applied - Without Injecting any Influences

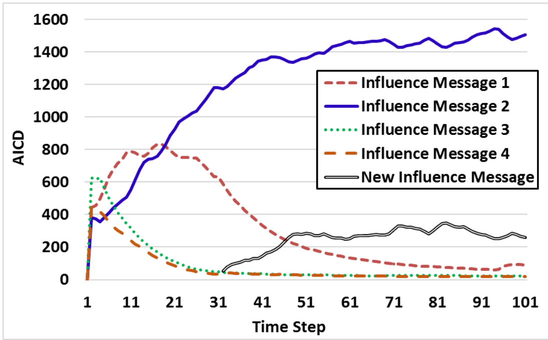


Figure 6.3: Inject Irrelevant Influence (Seed Set Size = 20)

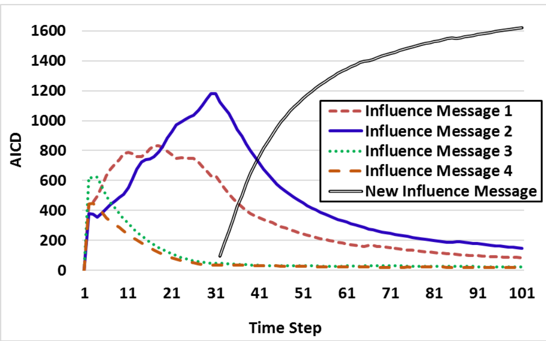


Figure 6.4: Inject Irrelevant Influence (Seed Set Size = 30)

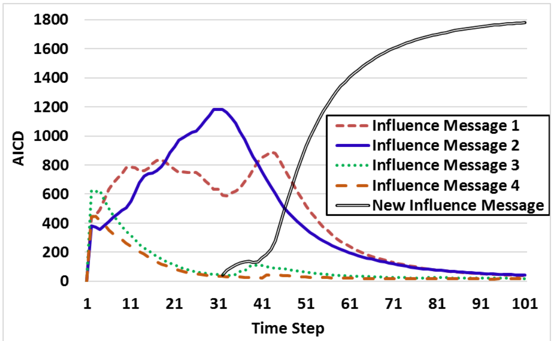


Figure 6.5: Inject Relevant Influence (Seed Set Size = 10)

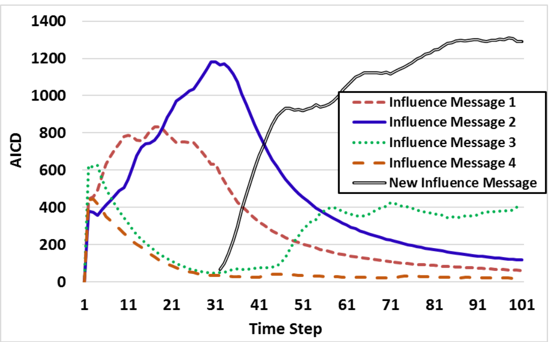


Figure 6.6: Inject Opposite Influence (Seed Set Size = 10)

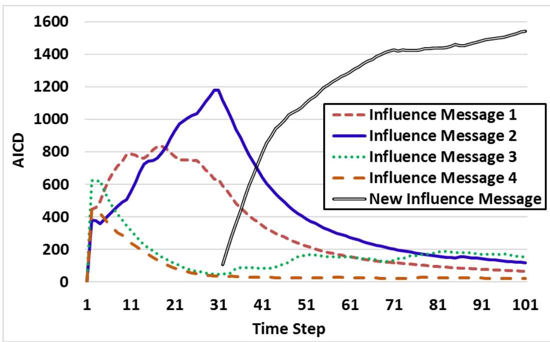


Figure 6.7: Inject Opposite Influence (Seed Set Size = 30)

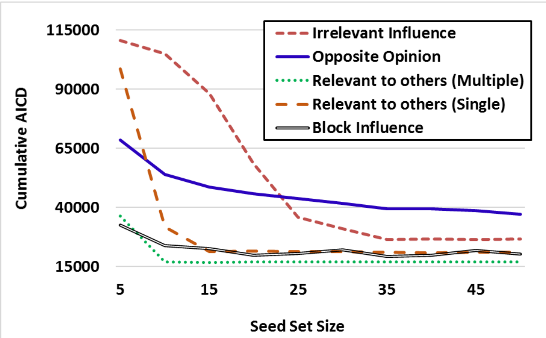


Figure 6.8: Undesirable Influence (Cumulative AICD)

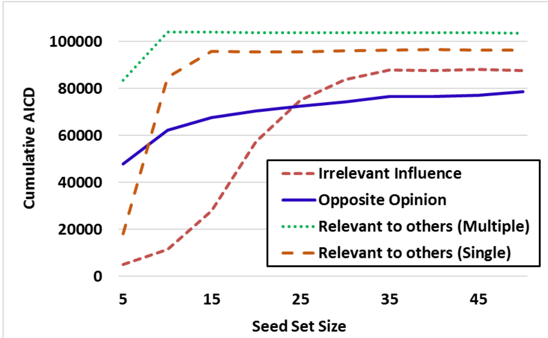


Figure 6.9: Injected Influence (Cumulative AICD)

Figure 6.8 describes the trend of the undesirable influence by applying various strategies. The AMID model employs probabilistic methods. Therefore, the results are averaged over 100 trials in the experiment. By varying the seed set size, the traditional approach, i.e., blocking influential nodes, performs very well, especially when the budge is limited. However, the administrative privileges of the social network must be granted to adopt this approach. By contrast, without any authorisations, injecting a new influence topically associated with multiple existing influences can produce an even better performance than that of blocking nodes. Overall, injecting a relevant influence topically associated with one or more existing influences appears more effective than that of bringing in opposite opinions or irrelevant influences. Furthermore, utilising irrelevant influences is not cost-efficient compared with others, but it outperforms that of adopting the adverse-opinion influence when the seed set size increases up to 25.

I also measure and compare the dissemination of newly injected messages when any strategy has been adopted. It can be seen from Figure 6.9 that the influence message topically relevant to multiple existing ones can easily dominate the social network, and a low budget of approximate ten seeds can almost achieve the maximum spread. Whereas, a new message requires a much higher budget and appears not cost-efficient.

6.5.5 Discussions

Based on the experimental results, I can derive that the fast dissemination of the newly injected influence message can generally suppress the expansion of the undesirable influence effectively. The results also suggest that to suppress an undesirable influence, introducing new influences topically associated with the existing ones appears more cost-efficient than that of injecting an influence message of brand new topics. Meanwhile, the supportive strength of the new influence message encourages the spread of the pre-existing influence messages with similar topics. On the other side, involving influences with opposite opinions does not carry out a desirable result unless the budget reaches a certain threshold. If the opposite influence appears not strong enough, i.e., limited budget, such strategy may cultivate the growth and spread of other influences, since their competitions and the contradicting opinions reduce the probability of being shared. As a standard approach, blocking nodes has been widely acknowledged as an effective method to alleviate the spread of any undesirable influence, especially when having a low budget. However, the targeting nodes are usually those influences of the social network, and such approaches are not applicable in most of the scenarios. From the above analysis and discussions, I can conclude that the undesirable influence can be suppressed by injecting other influences based on the AMID. The experiments also prove the rationality of applying AMID in analysing influence propagation when multiple influences involve. The proposed model can shed light on understanding, investigating and analysing multiple influences in social networks.

6.6 Summary

In this chapter, I studied the problem of multiple influences diffusion in social networks and proposed an Agent-based Multiple Influences Diffusion (AMID) model to describe the problem by using the concepts from multi-agent systems. In this model, I precisely formulated three types of influential relationships among different entities, i.e., user and user, user and influence, influence and influence. A distributed multiple influences diffusion algorithm was presented to show the user agent's response towards an influence message, where the personalised features, behaviours and social context were considered. To evaluate the proposed model, it is applied to the undesirable influence minimisation problem. The experimental results revealed that the proposed model was capable of alleviating the adverse impact of a particular influence by injecting other influences. The approach is also applicable in cases where the organisation does not possess the control of the social network.

This chapter mainly answers the Research Question 3 mentioned in Chapter 1. The model and results of this chapter have been published in (W. Li, Bai, Zhang & Nguyen, 2018b).

The conclusions of the thesis and future research directions will be given in the next chapter.

Chapter 7

Conclusion

7.1 Introduction

This chapter summarises the findings in influence diffusion modelling for complex networks using agent-based approaches, as well as the newly extended real-world problems addressed in this thesis. There are four major agent-based models proposed in this thesis, and each of them focuses on a particular aspect of influence diffusion in complex networks.

The research contributions are summarised in Section 7.2. The limitations of the models and possible directions for the future work are discussed in Section 7.3.

7.2 Research Contributions

This thesis contributes to the field from the following four aspects.

7.2.1 Agent-based Influence Diffusion Model

- I studied influence diffusion in social networks by using ABM, where the sophisticated features of social influence are considered. I systematically articulated and simulated influence diffusion process through defining the characteristics and behaviours of micro-level individuals.

- I further proposed a generic agent-based influence diffusion model, focusing on the individual's personalised traits and behaviours. Moreover, agents are granted capabilities of conducting training to evaluate the personalised parameter.
- I produced a specific range of dynamical behaviours based on various parameters by leveraging the proposed model and analyse the implications.
- I proposed Evolution-Based Backwards (EBB) and Enhanced Evolution-Based Backwards (2E2B) algorithm for effectively mining influencers under the proposed model.

7.2.2 Stigmergy-based Influencers Miner (SIMiner)

- I first systematically articulated the influence maximization using the stigmergy-based approach in a distributed environment. In other words, to the best of my knowledge, SIMiner is the first distributed model in this field.
- I leveraged SIMiner to tackle the challenging issues, i.e., the large-scale and dynamic environment, from a microscopic level using multi-agents.
- I explored the convergence of SIMiner, as well as the impact of varying parameters, showing that the efficiency of the mining capability can be improved by simply adding the same ant agents to the environment.
- I analysed and evaluated the performance of SIMiner. The empirical results reveal that SIMiner can give excellent performance, and even better than that of greedy algorithm in some datasets. The greedy algorithm in the influence maximization problem outperforms most of the proposed algorithms but is not scalable.

7.2.3 Agent-based Timeliness Influence Diffusion Model

- I formally defined the influence maintenance problem. To the best of my knowledge, this is the first literature describing the maintenance of influence in on-line social networks, which is significantly different from the adaptive influence maximization problem (clarified in the related work of Chapter 5).

- I proposed a novel decentralised influence diffusion model to accommodate to the influence maintenance problem. The proposed model is capable of capturing two primary elements for maintaining long-lasting influence, i.e., the temporal feature of a social network and the status of particular influence.

- I proposed a novel timeliness-based seed selection algorithm to maximize the influence lifespan.

7.2.4 Agent-based Multiple Influences Diffusion Model

- I formally defined a multiple influences diffusion model. To the best of my knowledge, this is the first literature systematically articulating the multiple influences and their relationships.

- I proposed a novel decentralised multiple influences diffusion model by considering the influential relationships, as well as individual's personalised traits, such as interests and trusts.

- I explored the intriguing discoveries and insights through modelling the relationships of different influences and evaluate the effectiveness of different approaches to minimise the undesirable influences by facilitating the proposed model.

7.3 Limitations and Future Directions

In my PhD study, I developed a basic model for modelling the hybrid effects of influence diffusion in complex networks. There are various potential directions to investigate influence diffusion by leveraging this generic approach. (1) Capture the dynamics of influence diffusion using HSN. Hybrid social network implies the decomposition of influence effects, which gives high extensibility and flexibility. Specifically, when other available influential factors are added or the existing factors are changed, the model can be extended by granting adaptation capabilities, which is able to update a particular influence facet with the evolution of social networks. (2) Analyse major channels of influence diffusion. The hybrid social network model can be extended by considering the impact factor in each influence facet. A particular influence can be diffused through various channels with different chances/possibilities. Further research works can be set to analyse the major channels for influences.

Second, in SIMiner, the selection still relies on the attempts of different combinations but I would like to set the investigation of parameter choices for SIMiner as one of the future work. It is possible to improve the effectiveness of SIMiner by granting the adaptation capabilities to each ant agent, so that they can explore the influential nodes more effectively. In addition, I may exploit a hybrid model by leveraging both centralised and decentralised approaches, where both local views from ant agents and limited global information are available.

Third, there are quite a few assumptions when I investigate the influence maintenance problem by leveraging ATID model. I plan to free up some of them. In particular, I will attempt to explore the solutions for the situations, where (1) the time step is not fixed; (2) for each investment, the seed set size is not fixed; and (3) the seed selection point can be a variant.

Last but not least, in the studies of multiple influences diffusion, I would like to set the theoretical explorations as one of the future works. There are also some possible extended research questions based on my studies: (1) What kind of influence message is most suitable to suppress an undesirable influence in a hypothesis situation (multiple influences exist)? (2) What's the impact of injecting multiple influence messages?

References

Aaker, D., Aaker, D. & Biel, A. (1993). Are brand equity investments really worthwhile. *Brand equity and advertising: Advertising's role in building strong brands*, 333–341.

Ahluwalia, R., Burnkrant, R. E. & Unnava, H. R. (2000). Consumer response to negative publicity: The moderating role of commitment. *Journal of marketing research*, *37*(2), 203–214.

Ahmed, S. & Ezeife, C. (2013). Discovering influential nodes from trust network. In *Proceedings of the 28th annual acm symposium on applied computing* (pp. 121–128).

Ahn, H. J. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences*, *178*(1), 37–51.

Alon, N., Gamzu, I. & Tennenholtz, M. (2012). Optimizing budget allocation among channels and influencers. In *Proceedings of the 21st international conference on world wide web* (pp. 381–388).

Anthony. (2009, March). *The nature of influence.* Retrieved from `http://apowerpoint.blogspot.co.nz/2009/03/nature-of-influence.html` ([Online; posted 02-March-2009])

Aral, S., Muchnik, L. & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, *106*(51), 21544–21549.

Arora, N. & Allenby, G. M. (1999). Measuring the influence of individual preference structures in group decision making. *Journal of Marketing Research*, 476–487.

Axelrod, R. (1997). The dissemination of culture a model with local convergence and global polarization. *Journal of conflict resolution*, *41*(2), 203–226.

Bai, Q., Zhang, M. & Zhang, H. (2005). A colored petri net based strategy for multi-agent scheduling. In *Rational, robust, and secure negotiation mechanisms in multi-agent systems* (pp. 3–10). doi: 10.1109/RRS.2005.1

Bailey, N. T. et al. (1975). *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.

Bao, Y., Wang, X., Wang, Z., Wu, C. & Lau, F. (2016). Online influence maximization in non-stationary social networks. In *Proceedings of ieee/acm international symposium on quality of service.* Beijing, China.

Barbieri, N., Bonchi, F. & Manco, G. (2012). Topic-aware social influence propagation models. In *Ieee 12th international conference on data mining (icdm)* (pp. 81–90).

Basaras, P., Belikaidis, I., Maglaras, L. & Katsaros, D. (2016). Blocking epidemic propagation in vehicular networks. In *12th annual conference on wireless on-demand network systems and services (wons)* (pp. 1–8).

Beg, I. & Ashraf, S. (2009). Similarity measures for fuzzy sets. *Appl. and Comput. math*, *8*(2), 192–202.

Bender, J. et al. (2010). Best practices for investment risk management. *Journal of Financial Transformation*, *28*, 37–43.

Benevenuto, F., Rodrigues, T., Cha, M. & Almeida, V. (2009). Characterizing user behavior in online social networks. In *Proceedings of the 9th acm sigcomm conference on internet measurement conference* (pp. 49–62). New York, NY, USA: ACM. doi: 10.1145/1644893.1644900

Berkovsky, S. & Freyne, J. (2015). Personalized social network activity feeds for increased interaction and content contribution. *Frontiers in Robotics and AI*, *2*, 24.

Bhagat, S., Goyal, A. & Lakshmanan, L. V. (2012). Maximizing product adoption in social networks. In *Proceedings of the fifth acm international conference on web search and data mining* (pp. 603–612).

Bharathi, S., Kempe, D. & Salek, M. (2007). Competitive influence maximization in social networks. In *International workshop on web and internet economics* (pp. 306–311).

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics reports*, *424*(4), 175–308.

Bonabeau, E. (1999). Editor's introduction: stigmergy. *Artificial Life*, *5*(2), 95–96.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *99*(suppl 3), 7280–7287.

Bonato, A. & Tian, Y. (2012). Complex networks and social networks. In *Advances in network analysis and its applications* (pp. 269–286). Springer.

Borgs, C., Brautbar, M., Chayes, J., Khanna, S. & Lucier, B. (2012). The power of local information in social networks. In *Proceedings of the 8th international conference on internet and network economics* (pp. 406–419).

Borodin, A., Filmus, Y. & Oren, J. (2010). Threshold models for competitive influence in social networks. In *International workshop on internet and network economics* (pp. 539–550).

Boyd, D. M. & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, *13*(1), 210–230.

Brautbar, M. & Kearns, M. J. (2010). Local algorithms for finding interesting individuals in large networks. In *Innovations in theoretical computer science* (pp. 188–199).

Breese, J. S., Heckerman, D. & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the fourteenth conference on uncertainty in artificial intelligence* (pp. 43–52).

Cannarella, J. & Spechler, J. A. (2014). Epidemiological modeling of online social network dynamics. *arXiv preprint arXiv:1401.4208*.

Carlton, M. . (2015, March). *Which social media platforms use an algorithm and what does that mean for marketing?* Retrieved from https://thedigiterati.com/social-media-platforms-use-algorithm-mean-marketing/ ([Online; posted 5-March-2015])

Castillo, C., Mendoza, M. & Poblete, B. (2013). Predicting information credibility in time-sensitive social media. *Internet Research*, *23*(5), 560–588.

Centola, D., Gonzalez-Avella, J. C., Eguiluz, V. M. & San Miguel, M. (2007). Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution*, *51*(6), 905–929.

Chantemargue, F., Krone, O., Schumacher, M., Dagaeff, T. & Hirsbrunner, B. (1998). Autonomous agents: from concepts to implementation. In *Proceedings of the fourteenth european meeting on cybernetics and systems research (emcsr'98)* (Vol. 2, pp. 731–736).

Chen, S., Fan, J., Li, G., Feng, J., Tan, K.-l. & Tang, J. (2015). Online topic-aware influence maximization. *Proceedings of the VLDB Endowment*, *8*(6), 666–677.

Chen, W., Collins, A., Cummings, R., Ke, T., Liu, Z., Rincon, D., ... Yuan, Y. (2011). Influence maximization in social networks when negative opinions may emerge and propagate. In *Proceedings of the 2011 siam international conference on data mining* (Vol. 11, pp. 379–390).

Chen, W., Wang, Y. & Yang, S. (2009). Efficient influence maximization in social networks. In *Proceedings of the 15th acm sigkdd international conference on knowledge discovery and data mining* (pp. 199–208). Paris, France.

Chen, W., Yuan, Y. & Zhang, L. (2010). Scalable influence maximization in social networks under the linear threshold model. In *Proceedings of the 10th international conference on data mining (icdm)* (pp. 88–97). Sydney, Australia.

Chen, Y.-C., Zhu, W.-Y., Peng, W.-C., Lee, W.-C. & Lee, S.-Y. (2014). Cim: community-based influence maximization in social networks. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *5*(2), 25.

DiMaggio, P., Evans, J. & Bryson, B. (1996). Have american's social attitudes become more polarized? *American journal of Sociology*, *102*(3), 690–755.

Doerr, B., Fouz, M. & Friedrich, T. (2012). Why rumors spread so quickly in social networks. *Communications of the ACM*, *55*(6), 70–75.

Domingos, P. & Richardson, M. (2001). Mining the network value of customers. In *Proceedings of the seventh acm sigkdd international conference on knowledge discovery and data mining* (pp. 57–66). San Francisco, California, USA.

Dorigo, M., Bonabeau, E. & Theraulaz, G. (2000). Ant algorithms and stigmergy. *Future Generation Computer Systems*, *16*(8), 851–871. doi: 10.1016/S0167-739X(00)00042-X

East, R., Hammond, K. & Lomax, W. (2008). Measuring the impact of positive and negative word of mouth on brand purchase probability. *International Journal of Research in Marketing*, *25*(3), 215–224.

Fang, X., Hu, P. J.-H., Li, Z. & Tsai, W. (2013). Predicting adoption probabilities in social networks. *Information Systems Research*, *24*(1), 128–145.

Feng, L., Hu, Y., Li, B., Stanley, H. E., Havlin, S. & Braunstein, L. A. (2015).

Competing for attention in social media under information overload conditions. *PloS one*, *10*(7), e0126090.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of personality and Social Psychology*, *38*(6), 889.

Galam, S. & Moscovici, S. (1991). Towards a theory of collective phenomena: Consensus and attitude changes in groups. *European Journal of Social Psychology*, *21*(1), 49–74.

Ganguly, N., Deutsch, A. & Mukherjee, A. (2009). *Dynamics on and of complex networks: applications to biology, computer science, and the social sciences*. Springer.

Garland, T. (2015, March). *The scientific method as an ongoing process.* Retrieved from `http://idea.ucr.edu/documents/flash/scientific_method/story_html5.html` ([Online; posted 20-March-2015])

Gayraud, N. T., Pitoura, E. & Tsaparas, P. (2015). Diffusion maximization in evolving social networks. In *Proceedings of the 2015 acm on conference on online social networks* (pp. 125–135).

Gilbert, G. N. (2008). *Agent-based models* (No. 153). Sage.

Golovin, D. & Krause, A. (2011). Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, *42*, 427–486.

Gomez-Rodriguez, M., Balduzzi, D., Schölkopf, B., Scheffer, G. T. et al. (2011). Uncovering the temporal dynamics of diffusion networks. In *28th international conference on machine learning (icml 2011)* (pp. 561–568).

Gomez Rodriguez, M., Leskovec, J. & Krause, A. (2010). Inferring networks of diffusion and influence. In *Proceedings of the 16th acm sigkdd international conference on knowledge discovery and data mining* (pp. 1019–1028).

Gomez Rodriguez, M., Leskovec, J. & Schölkopf, B. (2013). Structure and dynamics of information pathways in online media. In *Proceedings of the sixth acm international conference on web search and data mining* (pp. 23–32).

Gomez-Rodriguez, M., Song, L., Du, N., Zha, H. & Schölkopf, B. (2016). Influence estimation and maximization in continuous-time diffusion networks. *ACM Transactions on Information Systems (TOIS)*, *34*(2), 9.

Goyal, A., Bonchi, F. & Lakshmanan, L. V. (2010). Learning influence probabilities in social networks. In *Proceedings of the third acm international conference on web search and data mining* (pp. 241–250). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/1718487.1718518` doi: 10.1145/1718487.1718518

Goyal, A., Bonchi, F. & Lakshmanan, L. V. (2011). A data-based approach to social influence maximization. *Proceedings of the VLDB Endowment*, *5*(1), 73–84.

Goyal, A., Lu, W. & Lakshmanan, L. V. (2011a). Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th international conference companion on world wide web* (pp. 47–48). Hyderabad, India.

Goyal, A., Lu, W. & Lakshmanan, L. V. (2011b). Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *2011 ieee 11th international conference on data mining (icdm),* (pp. 211–220).

Goyal, S., Heidari, H. & Kearns, M. (2014). Competitive contagion in networks. *Games and Economic Behavior*.

Guille, A., Hacid, H., Favre, C. & Zighed, D. A. (2013). Information diffusion in online social networks: A survey. *ACM SIGMOD Record*, *42*(2), 17–28.

Gustafsson, L. & Sternad, M. (2010). Consistent micro, macro and state-based population modelling. *Mathematical biosciences*, *225*(2), 94–107.

Hadeli, Valckenaers, P., Kollingbaum, M. & Van Brussel, H. (2004). Multi-agent coordination and control using stigmergy. *Computers in industry*, *53*(1), 75–96. doi: 10.1016/S0166-3615(03)00123-4

Harper, F. M. & Konstan, J. A. (2015). The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *5*(4), 1–19. doi: 10.1145/2827872

Hatano, D., Fukunaga, T. & Kawarabayashi, K.-i. (2016). Adaptive budget allocation for maximizing influence of advertisements. In *Proceedings of the 25th international joint conference on artificial intelligence ijcai-16* (p. 3600-3608).

He, X., Song, G., Chen, W. & Jiang, Q. (2012). Influence blocking maximization in social networks under the competitive linear threshold model. In *Proceedings of the 2012 siam international conference on data mining* (pp. 463–474).

Holland, J. H. (1995). *Hidden order: How adaptation builds complexity*. Basic Books.

Holme, P. (2015). Modern temporal network theory: a colloquium. *The European Physical Journal B*, *88*(9), 1–30.

Hsu, C.-L. & Lin, J. C.-C. (2008). Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. *Information & management*, *45*(1), 65–74.

Jamali, M. & Ester, M. (2010). A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth acm conference on recommender systems* (pp. 135–142). doi: 10.1145/1864708.1864736

Jiang, C., Li, W., Bai, Q. & Zhang, M. (2015). Preference aware influence maximization. In *Proceedings of international joint agents workshop and symposium (ijaws).* Springer.

Jiang, Y. & Jiang, J. (2015). Diffusion in social networks: A multiagent perspective. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, *45*(2), 198–213. doi: 10.1109/TSMC.2014.2339198

Jin, Y., Wang, W. & Xiao, S. (2007). An sirs model with a nonlinear incidence rate. *Chaos, Solitons & Fractals*, *34*(5), 1482–1497.

Jøsang, A., Hayward, R. & Pope, S. (2006). Trust network analysis with subjective logic. In *Proceedings of the 29th australasian computer science conference - volume 48* (pp. 85–94). Darlinghurst, Australia, Australia: Australian Computer Society, Inc. Retrieved from `http://dl.acm.org/citation.cfm?id=1151699.1151710`

Jung, K., Heo, W. & Chen, W. (2012). Irie: Scalable and robust influence maximization in social networks. In *2012 ieee 12th international conference on data mining (icdm)* (pp. 918–923).

Karimi, F. & Holme, P. (2013). Threshold model of cascades in empirical temporal networks. *Physica A: Statistical Mechanics and its Applications*, *392*(16), 3476–3483.

Kelman, H. C. (1961). Processes of opinion change. *Public opinion quarterly*, *25*(1), 57–78.

Kempe, D., Kleinberg, J. & Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth acm sigkdd international conference on knowledge discovery and data mining* (pp. 137–146). Washington, DC, USA.

Kiesling, E., Günther, M., Stummer, C. & Wakolbinger, L. M. (2012). Agent-based simulation of innovation diffusion: a review. *Central European Journal of Operations Research*, *20*(2), 183–230.

Kimura, M., Saito, K. & Motoda, H. (2008). Minimizing the spread of contamination by blocking links in a network. In *Aaai* (Vol. 8, pp. 1175–1180).

Kimura, M., Saito, K. & Motoda, H. (2009). Blocking links to minimize contamination spread in a social network. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *3*(2), 9.

Klimt, B. & Yang, Y. (2004). Introducing the enron corpus. In *Ceas*.

Kostka, J., Oswald, Y. A. & Wattenhofer, R. (2008). Word of mouth: Rumor dissemination in social networks. In *International colloquium on structural information and communication complexity* (pp. 185–196).

Kumar, P. & Raghavendra, G. (2011). A note on the parameter of evaporation in the ant colony optimization algorithm. In *International mathematical forum* (Vol. 6, pp. 1655–1659).

Kumar, R., Novak, J. & Tomkins, A. (2010). Structure and evolution of online social networks. In *Link mining: models, algorithms, and applications* (pp. 337–357). Springer.

Kutzkov, K., Bifet, A., Bonchi, F. & Gionis, A. (2013). Strip: stream learning of influence probabilities. In *Proceedings of the 19th acm sigkdd international conference on knowledge discovery and data mining* (pp. 275–283).

Leskovec, J., Huttenlocher, D. & Kleinberg, J. (2010a). Predicting positive and negative links in online social networks. In *Proceedings of the 19th international conference on world wide web* (pp. 641–650).

Leskovec, J., Huttenlocher, D. & Kleinberg, J. (2010b). Signed networks in social media. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1361–1370).

Leskovec, J., Kleinberg, J. & Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *1*(1), 2.

Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J. & Glance, N. (2007). Cost-effective outbreak detection in networks. In *Proceedings of the 13th acm sigkdd international conference on knowledge discovery and data mining* (pp. 420–429).

Leskovec, J. & Krevl, A. (2014, Jun). *SNAP Datasets: Stanford large network dataset collection.* http://snap.stanford.edu/data.

Leskovec, J., Lang, K. J., Dasgupta, A. & Mahoney, M. W. (2009). Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, *6*(1), 29–123.

Lewis, T. G. (2013). Cognitive stigmergy: A study of emergence in small-group social networks. *Cognitive Systems Research*, *21*, 7–21. doi: 10.1016/j.cogsys.2012.06.002

Li, D., Xu, Z.-M., Chakraborty, N., Gupta, A., Sycara, K. & Li, S. (2014). Polarity related influence maximization in signed social networks. *PloS one*, *9*(7), e102199.

Li, M., Wang, X., Gao, K. & Zhang, S. (2017). A survey on information diffusion in online social networks: Models and methods. *Information*, *8*(4), 118.

Li, M. Y., Graef, J. R., Wang, L. & Karsai, J. (1999). Global dynamics of a seir model with varying total population size. *Mathematical biosciences*, *160*(2), 191–213.

Li, W., Bai, Q., Jiang, C. & Zhang, M. (2016). Stigmergy-based influence maximization. In *Proceedings of pacific rim international conference on artificial intelligence (pricai)* (p. 750-762). Phuket, Thailand: Springer. doi: 10.1007/978-3-319-42911-3_63

Li, W., Bai, Q., Nguyen, T. D. & Zhang, M. (2017). Agent-based influence maintenance in social networks. In *Proceedings of the 16th conference on autonomous agents and multiagent systems (aamas 2017) (extended abstract)* (pp. 1592–1594).

Li, W., Bai, Q. & Zhang, M. (2016a). Agent-based influence propagation in social networks. In *Proceedings of ieee international conference on agents* (pp. 51–56).

Li, W., Bai, Q. & Zhang, M. (2016b). Comprehensive influence propagation modelling for hybrid social network. In *Australasian joint conference on artificial intelligence* (pp. 597–608).

Li, W., Bai, Q. & Zhang, M. (2018a). A multi-agent system for modelling preference-based complex influence diffusion in social networks. *The Computer Journal*.

Li, W., Bai, Q. & Zhang, M. (2018b). Siminer: A stigmergy-based model for mining influential nodes in dynamic social networks. *IEEE Transactions on Big Data*. doi: 10.1109/TBDATA.2018.2824826

Li, W., Bai, Q., Zhang, M. & Nguyen, T. D. (2018a). Automated influence maintenance in social networks: an agent-based approach. *IEEE Transactions on Knowledge and Data Engineering*.

Li, W., Bai, Q., Zhang, M. & Nguyen, T. D. (2018b). Modelling multiple influences diffusion in on-line social networks. In *Proceedings of the 17th conference on autonomous agents and multiagent systems (aamas 2018).*

Li, Z. & Tang, X. (2012). Group polarization: connecting, influence and balance, a simulation study based on hopfield modeling. In *Pricai 2012: Trends in artificial intelligence* (pp. 710–721). Springer.

Li, Z. & Tang, X. (2015). From global polarization to local social mechanisms: A study based on abm and empirical data analysis. In *Smart modeling and simulation for*

complex systems (pp. 29–40). Springer.
Lien, C. H. & Cao, Y. (2014). Examining wechat users motivations, trust, attitudes, and positive word-of-mouth: Evidence from china. *Computers in Human Behavior*, *41*, 104–111. doi: 10.1016/j.chb.2014.08.013
Linden, G., Smith, B. & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, *7*(1), 76–80.
Liu, H., Hu, Z., Mian, A., Tian, H. & Zhu, X. (2014). A new user similarity model to improve the accuracy of collaborative filtering. *Knowledge-Based Systems*, *56*, 156–166.
Liu, S., Ying, L. & Shakkottai, S. (2010). Influence maximization in social networks: An ising-model-based approach. In *The 48th annual allerton conference on communication, control, and computing (allerton)* (pp. 570–576).
Liu, W., Yue, K., Wu, H., Li, J., Liu, D. & Tang, D. (2016). Containment of competitive influence spread in social networks. *Knowledge-Based Systems*, *109*, 266–275.
Ma, H., Yang, H., Lyu, M. R. & King, I. (2008). Mining social networks using heat diffusion processes for marketing candidates selection. In *Proceedings of the 17th acm conference on information and knowledge management* (pp. 233–242).
Macal, C. M. & North, M. J. (2009). Agent-based modeling and simulation. In *Winter simulation conference* (pp. 86–98).
Macy, M. W., Kitts, J. A., Flache, A. & Benard, S. (2003). Polarization in dynamic networks: A hopfield model of emergent structure. *Dynamic social network modeling and analysis*, 162–173.
Marketing, L. (2015, August). *What are your short-and long-term marketing strategies?* Retrieved from `http://www.launch-marketing.com/articles/what-are-your-short-and-long-term-marketing-strategies/` ([Online; posted 26-August-2015])
McAuley, J. J. & Leskovec, J. (2012). Learning to discover social circles in ego networks. In *Nips* (Vol. 2012, pp. 548–56).
McGee, M. (2013, August). *Edgerank is dead: Facebook's news feed algorithm now has close to 100k weight factors.* Retrieved from `http://marketingland.com/edgerank-is-dead-facebooks-news-feed-algorithm-now-has-close-to-100k-weight-factors-55908` ([Online; posted 16-August-2013])
McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415–444.
Mnih, A. & Salakhutdinov, R. R. (2008). Probabilistic matrix factorization. In J. C. Platt, D. Koller, Y. Singer & S. T. Roweis (Eds.), *Advances in neural information processing systems 20* (pp. 1257–1264). Curran Associates, Inc.
Mostafa, A., Zhang, M. & Bai, Q. (2014, January). Trustworthy stigmergic service composition and adaptation in decentralized environments. *Service Computing, IEEE*, *9*(2), 317 - 329. doi: 10.1109/TSC.2014.2298873
Nemhauser, G. L., Wolsey, L. A. & Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions-i. *Mathematical Programming*, *14*(1), 265–294.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, *45*(2), 167–256.

Nguyen, D. T. (2017). *Trust management for complex agent groups* (Unpublished doctoral dissertation). Auckland University of Technology.

Ohsaka, N., Akiba, T., Yoshida, Y. & Kawarabayashi, K.-i. (2016). Dynamic influence analysis in evolving networks. *Proceedings of the VLDB Endowment*, *9*(12), 1077–1088.

Opsahl, T. & Panzarasa, P. (2009). Clustering in weighted networks. *Social networks*, *31*(2), 155–163.

Pastor-Satorras, R. & Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Physical review letters*, *86*(14), 3200.

Peres, R. (2014). The impact of network characteristics on the diffusion of innovations. *Physica A: Statistical Mechanics and Its Applications*, *402*, 330–343.

Raven, B. H. (1964). *Social influence and power* (Tech. Rep.). CALIFORNIA UNIV LOS ANGELES.

Rodriguez, M. G. & Schölkopf, B. (2012). Influence maximization in continuous time diffusion networks. *arXiv preprint arXiv:1205.1682*.

Rozin, P. & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and social psychology review*, *5*(4), 296–320.

Saito, K., Nakano, R. & Kimura, M. (2008). Prediction of information diffusion probabilities for independent cascade model. In *International conference on knowledge-based and intelligent information and engineering systems* (pp. 67–75).

Serazzi, G. & Zanero, S. (2004). Computer virus propagation models. In *Performance tools and applications to networked systems* (pp. 26–50). Springer.

Shakarian, P., Bhatnagar, A., Aleali, A., Shaabani, E. & Guo, R. (2015). *Diffusion in social networks*. Springer.

Shiffman, D., Fry, S. & Marsh, Z. (2012). The nature of code. In (pp. 323–330). D. Shiffman.

Sinha, R. R. & Swearingen, K. (2001). Comparing recommendations made by online systems and friends. In *Delos workshop: personalisation and recommender systems in digital libraries* (Vol. 106).

Soma, T., Kakimura, N., Inaba, K. & Kawarabayashi, K.-i. (2014). Optimal budget allocation: Theoretical guarantee and efficient algorithm. In *Proceedings of the 31st international conference on machine learning* (p. 351-359).

Song, G., Li, Y., Chen, X., He, X. & Tang, J. (2017). Influential node tracking on dynamic social network: an interchange greedy approach. *IEEE Transactions on Knowledge and Data Engineering*, *29*(2), 359–372.

Su, X., Zhang, M. & Bai, Q. (2015). Decentralised task allocation under space, time and communication constraints in disaster domains. In *Smart modeling and simulation for complex systems* (pp. 41–57). Springer.

Subbian, K., Aggarwal, C. C. & Srivastava, J. (2016). Querying and tracking influencers in social streams. In *Proceedings of the ninth acm international conference on web search and data mining* (pp. 493–502).

Sunstein, C. R. (2002). The law of group polarization. *Journal of political philosophy*, *10*(2), 175–195.

Sycara, K. P. (1998). Multiagent systems. *AI magazine*, *19*(2), 79.

Szmidt, E. & Kacprzyk, J. (2000). Distances between intuitionistic fuzzy sets. *Fuzzy sets and systems*, *114*(3), 505–518.

Takahashi, J., Kanamori, R. & Ito, T. (2012). A preliminary study on anticipatory stigmergy for traffic management. In *Proceedings of the 2012 ieee/wic/acm international conference on intelligent agent technology(iat 2012)* (Vol. 3, pp. 399–405). Macau, China. doi: 10.1109/WI-IAT.2012.121

Tang, J., Chang, Y. & Liu, H. (2014). Mining social media with social theories: A survey. *ACM SIGKDD Explorations Newsletter*, *15*(2), 20–29.

Tang, J., Sun, J., Wang, C. & Yang, Z. (2009). Social influence analysis in large-scale networks. In *Proceedings of the 15th acm sigkdd international conference on knowledge discovery and data mining* (pp. 807–816).

Tang, J., Wu, S. & Sun, J. (2013). Confluence: Conformity influence in large social networks. In *Proceedings of the 19th acm sigkdd international conference on knowledge discovery and data mining* (pp. 347–355).

Theraulaz, G. & Bonabeau, E. (1999). A brief history of stigmergy. *Artificial life*, *5*(2), 97–116. doi: 10.1162/106454699568700

Tong, G., Wu, W., Tang, S. & Du, D.-Z. (2017). Adaptive influence maximization in dynamic social networks. *IEEE/ACM Transactions on Networking (TON)*, *25*(1), 112–125.

Trpevski, D., Tang, W. K. & Kocarev, L. (2010). Model for rumor spreading over networks. *Physical Review E*, *81*(5), 056102.

Turner, J. C. (1991). *Social influence*. Thomson Brooks/Cole Publishing Co.

Valencia, A. (2013, August). *Short-term vs. long-term online marketing.* Retrieved from `https://www.wedowebcontent.com/blog/set-short-and-long-term-goals-and-objectives-for-your-online-marketing/` ([Online; posted 22-August-2013])

van Maanen, P.-P. & van der Vecht, B. (2013). An agent-based approach to modeling online social influence. In *Proceedings of the 2013 ieee/acm international conference on advances in social networks analysis and mining* (pp. 600–607).

Veblen, T. (2017). *The theory of the leisure class*. Routledge.

Wang, B., Chen, G., Fu, L., Song, L. & Wang, X. (2017). Drimux: Dynamic rumor influence minimization with user experience in social networks. In (Vol. 29, pp. 2168–2181). IEEE.

Wang, C., Chen, W. & Wang, Y. (2012). Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, *25*(3), 545–576.

Wang, K., Huang, L., Zhou, C., Pang, W. et al. (2003). Traveling salesman problem. In *Proceedings of conference on machine learning and cybernetics* (Vol. 3, pp. 1583–1585). Xi'an, China.

Wang, Q., Jin, Y., Lin, Z., Cheng, S. & Yang, T. (2016). Influence maximization in social networks under an independent cascade-based model. *Physica A: Statistical*

Mechanics and its Applications, *444*, 20–34.
Wang, Q., Lin, Z., Jin, Y., Cheng, S. & Yang, T. (2015). Esis: emotion-based spreader-ignorant-stifler model for information diffusion. *Knowledge-Based Systems*, *81*, 46–55.
Wang, S., Zhao, X., Chen, Y., Li, Z., Zhang, K. & Xia, J. (2013). Negative influence minimizing by blocking nodes in social networks. In *Workshops at the twenty-seventh aaai conference on artificial intelligence.*
Wang, Y., Cong, G., Song, G. & Xie, K. (2010). Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In *Proceedings of the 16th acm sigkdd international conference on knowledge discovery and data mining* (pp. 1039–1048). Washington, DC, USA..
Wang, Y., Fan, Q., Li, Y. & Tan, K.-L. (2017). Real-time influence maximization on dynamic social streams. *Proceedings of the VLDB Endowment*, *10*(7), 805–816.
Watts, D. J. & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of consumer research*, *34*(4), 441–458.
Webb, G. (1981). A reaction-diffusion model for a deterministic diffusive epidemic. *Journal of Mathematical Analysis and Applications*, *84*(1), 150–161.
Wooldridge, M. (2009). *An introduction to multiagent systems*. John Wiley & Sons.
Wu, H., Liu, W., Yue, K., Huang, W. & Yang, K. (2015). Maximizing the spread of competitive influence in a social network oriented to viral marketing. In *International conference on web-age information management* (pp. 516–519).
Wu, L., Ge, Y., Liu, Q., Chen, E., Long, B. & Huang, Z. (2016). Modeling users preferences and social links in social networking services: A joint-evolving perspective. In *Thirtieth aaai conference on artificial intelligence.*
Xiong, F., Liu, Y., Zhang, Z.-j., Zhu, J. & Zhang, Y. (2012). An information diffusion model based on retweeting mechanism for online social media. *Physics Letters A*, *376*(30-31), 2103–2108.
Yang, Y., Mao, X., Pei, J. & He, X. (2016). Continuous influence maximization: What discounts should we offer to social network users? In *Proceedings of the 2016 international conference on management of data* (pp. 727–741).
Yao, Q. & Guo, L. (2015). Minimizing the social influence from a topic modeling perspective. In *International conference on data science* (pp. 6–15).
Yao, Q., Zhou, C., Xiang, L., Cao, Y. & Guo, L. (2014). Minimizing the negative influence by blocking links in social networks. In *International conference on trustworthy computing and services* (pp. 65–73).
Ye, D., Zhang, M. & Vasilakos, A. V. (2016, January). A survey of self-organization mechanisms in multiagent systems. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, *PP*(99), 1 - 21. doi: 10.1109/TSMC.2015.2504350
Zhang, W., Wang, S., Han, G., Yang, Y. & Wang, Q. (2015). Hiprank: Ranking nodes by influence propagation based on authority and hub. In *Smart modeling and simulation for complex systems* (pp. 1–14). Springer.
Zhu, Y., Li, D. & Zhang, Z. (2016). Minimum cost seed set for competitive social influence. In *The 35th annual ieee international conference on computer communications* (pp. 1–9).

Zhuang, H., Sun, Y., Tang, J., Zhang, J. & Sun, X. (2013). Influence maximization in dynamic social networks. In *Ieee 13th international conference on data mining (icdm)* (pp. 1313–1318).

Appendix A

Glossary

2E2B Enhanced Evolution-Based Backward Seed Selection

ABM Agent-Based Modelling

ACOS Adjusted Cosine measure

AICD Active Influence Coverage Degree

AMID Agent-based Multiple Influences Diffusion

APL Average Path Length

ATID Agent-based Timeliness Influence Diffusion

ATID Agent-based Timeliness Influence Diffusion model

CA Cellular Automata

CDHKcut Community and Degree Heuristic with Kcut

CDH-SHRINK Community and Degree Heuristic with SHRINK

CIACD Cumulative Influence Activation Coverage Difference

CKR Common Knowledge Repository

CPS Common Preference Similarity

CSI Comprehensive Social Influence

DDH Degree Discount Heuristics

DWA Degree Weighted Activation

EBB Evolution-Based Backwards Seed Selection

ESIS model Emotion-based Spreader–Ignorant–Stifler model

GAC Global Activation Coverage

GCTD Global Cumulative Timeliness Degree

GT model General Threshold model

GTD Global Timeliness Degree

HAM Hopfield Attractor Model

HSN Hybrid Social Network

IC model Independent Cascade model

ICMPS Independent Cascade Model with Pyramid Scheme

IDMAS Influence Diffusion Multi-Agent System

IPP Influence Propagation Probability

irSIR model infection recovery SIR model

LT model Linear Threshold model

MAE Mean Absolute Error

MAS Multi-Agent Systems

MC Monte-Carlo Simulations

MIA model Maximum Influence Arborescence model

PCL Prior Commitment Level

PDF Probability Density Function

PMF Probabilistic Matrix Factorisation

PTIC model Preference-based Trust IC model

RMSE Root Mean Squared Error

SEIR model Susceptible-Exposed-Infected-Removed model

SI model Susceptible-Infected model

SIM Stream Influence Maximization

SIMiner Stigmergy-based Influencers Miner

SIR model Susceptible-Infected-Removed model

SIRS model Susceptible-Infected-Removed-Susceptible model

SIS model Susceptible-Infected-Susceptible model

TIC model Topic-aware Independent Cascade model

TIH Timeliness Increase Heuristic

TLT model Topic-aware Linear Threshold model

TSP Travelling Salesman Problem

VM Voter Model

WoM effect Word-of-Mouth effect