Henley
Business School
UNIVERSITY OF READING

The Network Effect and Helpfulness of Electronic Word-of-Mouth: Understanding the online consumer reviews in social networking sites

Xue Pan

Submitted in partial fulfilment of the requirements of

University of Reading for

the degree of Doctor of Philosophy

Informatics Research Centre

Henley Business School

February 2019

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Name: Xue Pan

Certificate of Readiness to Be Included in Library

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Acknowledgements

Everything seems unreal, but I am here finally. Still remember the day when I set foot on this country. Looking back the journey of the last three years, I am filled with thousands of emotions. A long list that ever helped, supported, delighted me comes to my mind.

Firstly, the highest respect and sincerest gratitude is given to my supervisor, Professor Kecheng Liu for the continuous support of my PhD study as well as the life in the UK. He is not only a supervisor who has given me professional academic suggestions, but also a senior friend who shared a lot of experience for my life and career. I also thank Dr Lily Sun from the department of computer science who gave me a lot of guidance of my research.

Profound gratitude also goes to other members of staff in the IRC, especially, Professor Keiichi Nakata, Professor Yinshan Tang, Dr. Stephen Gulliver, Dr. Weizi Li, Dr. Yinleng Tan, Dr Liang Han, Ms Cindy Zhang for the support they have given me throughout my study. It has been a pleasure to work with many lovely colleagues and friends in the BISA.

Many thanks to my government, the China Scholarship Council who gives me financial support that enable me fully devoted to my research. I also acknowledge my department, BISA (Business Informatics, Systems and Accounting) that offers the scholarship during my PhD research.

Last but not least, I thank my beloved family: my parents, my brother and sister for supporting me spiritually throughout the whole research life in the UK and encouraging me constantly. I would like to specially thank my husband, Dr. Lei Hou, for the meticulous accompany and unconditional support in the past ten years.

Related Publications

Journal Articles

Pan, X., Hou, L., Liu, K., & Niu, H. (2018). Do reviews from friends and the crowd affect online consumer posting behaviour differently? *Electronic Commerce Research and Applications*, 29, 102-112.

Hou, L., **Pan, X**., & Liu, K. (2018). Balancing the popularity bias of object similarities for personalised recommendation. *European Physical Journal B*, 91, 47.

Pan, X., Hou, L., & Liu, K. (2017). Social influence on selection behaviour: Distinguishing local-and global-driven preferential attachment. *PLOS ONE*, 12(4), e0175761.

Liu, J., Hou, L., **Pan, X.,** Guo, Q., & Zhou, T. (2016). Stability of similarity measurements for bipartite networks. *Scientific Reports*, 6, 18653.

Conference Proceedings

Pan, X., Hou, L., & Liu, K. (2017). On the prediction of future helpfulness for online reviews. In: *Proceedings of International Conference on Logistics, Informatics and Service Sciences*. Kyoto, Japan.

Hou, L., **Pan, X.,** & Liu, K. (2017). Identifying informative objects for mining similar users in recommender systems. In: *Proceedings of International Conference on Logistics, Informatics and Service Sciences*. Kyoto, Japan. **(Best paper award)**

Working Papers

Pan, X., Hou, L., & Liu, K. Predicting the future increment of review helpfulness: an empirical study based on a two-wave dataset. (under review)

Pan, X., Hou, L., & Liu, K. The Effect of Product Distance on the WOM in Recommendation Network. (under review)

Abstract

The Internet has brought disruptive changes to people's life. Such drastic and profound changes demand much attention of research in understanding user's behaviour, especially the interactions between consumers and businesses or products. Nowadays consumers are enabled to search relevant information of products from social media, such as product review, which is normally referred as the electronic Word-of-Mouth (eWOM). Such eWOM conveying product descriptions and individuals' experience has shown significant influence on consumers' purchase behaviour. Focusing on the theory and practice of realising the substantial value of the digital resources, this thesis explores the impact of eWOM on consumer behaviour and help to understand the online consumer reviews on social networking sites.

Most e-commerce websites provide social networking services that allow users to add friends or follow trustworthy reviewers. However, in relevant studies, the source of eWOM, i.e. from friends or crowds, has not been fully addressed. It is thus an open question whether or not friend's reviews differently impact consumer's behaviour in comparison to the general crowd reviews. In this thesis, we develop a Probit model to study the impact of friend reviews and crowd reviews on the possibility of subsequent consumers' posting behaviour. Despite the common perception that the volume, valance and variance of reviews significantly impact the likelihood of following posting behaviour, we find that such influence comes from friend's reviews. Furthermore, a Monte Carlo simulation experiment is carried out to examine to what extent the consumers' decision is affected by the product popularity among friends and crowds respectively. The simulation confirms the econometric analysis, and it is shown that about 75% of the posting behaviour are affected by product popularity amongst one's friends. Such results imply the importance of trust in the process of eWOM diffusion.

Despite the huge value of online reviews, the overwhelming amount of them makes consumers impossible to access all of them before they find the most proper product. Thus, to uncover the most helpful reviews becomes a task of both practical and theoretical significance. This thesis applies a two-wave based dataset consisting of online reviews and their helpfulness information collected at two different time points and studies the increment of helpfulness between such two time points. Though previous studies have confirmed that the helpfulness of reviews can be largely explained by the features of reviewers and reviews, we show that these features have much less explanatory power for the increment of helpfulness in the future. Furthermore, the reviewer activeness and review disclosure information are shown to be more predictive. Product recommendation network is another way to help consumers find interesting products quickly. The demand and sale of products have been shown to be highly correlated to that of their neighbours. In this thesis, we employ an empirical book recommendation network of Amazon to investigate the effect of product distance on their eWOM in terms of online review volume and rating. The analysis indicates the connectivity between books has significant influence on the WOM.

To summarise, the thesis explores the interplay between consumers, products and online reviews in the context of social network and product recommendation network, and examines the factors that are contributing to the helpfulness of reviews. In theory, this thesis highlights the network effect on the formation and diffusion of eWOM. The influence of reviews on users can be enlarged through social network, and the product network also largely reshapes the eWOM of products. The developed econometric as well as computational methods also contribute to the knowledge providing new ways of studying consumer behaviour. The results reported in this thesis can inform practitioners on the design of online reviewing system, so to better aid consumers accessing information, and the possible online marketing strategies.

Table of Contents

List of Figures

List of Tables

List of Notions and Abbreviations

Symbol or Abbreviations	Meaning
CSS	Computational social science
WOM	Word-of-mouth
eWOM	Electronic word-of-mouth
GP	Global popularity
GPA	Global preferential attachment
LP	Local popularity
PL_{ij}	The average of shortest-path length between node i and node j
PRN	Product recommendation network
RA	Reviewer activeness
RC	Reviewer credibility
RHV	Reviewer historical votes
RDI	Review disclosure information
RR	Review readability
RS	Review sentiment
SNA	Social network analysis
SNS	Social network site
SIR	Susceptible-Infected-Recovered model
UGC	User-generated content

Chapter 1. Introduction

1.1 Background and Motivation

Due to the rapid development of Internet technology, human life has been massively changed in the recent decades. We communicate with friends online, post photos on social media, buy products on the e-commerce websites as well as share ideas, opinions or experience regarding specific person, product or subject in real time on the Internet. Thanks to the Web 2.0, all these activities become possible on the Internet, where people act as information recipients as well as creators because the low cost and wide availability. Consequently, traditional word-of-mouth (WOM) now has an electronic variant, namely electronic WOM (eWOM), which has caught huge attention in recent decades. Previously, when people need information for some targeted products, they turn to expert-generated sources like newspapers or seek suggestions from friends in the conversations "over the backyard fence". These methods are subsumed together in the communications of eWOM, where people can socially interact with each other, share product-related information, make informed purchasing decisions (Blazevic *et al.,* 2013).

With consumers' increasing reliance on online retailing and information seeking, eWOM has become more and more important. Thus, it has attracted considerable interest from researchers in the past decades. Many types of eWOM communications have been studied, such as discussion forums (Andreassen and Streukens, 2009; Cheung *et al*., 2009), blogs (Dhar and Chang, 2009; Kozinets *et al*., 2010; Thorson and Rodgers, 2006), UseNet groups (Godes and Mayzlin, 2004), and product reviews (Lee and Youn, 2009).

Online product reviews, being one of the most common type of eWOM, are the most accessible and prevalent (Chatterjee, 2001). According to a survey (Staff 2007a, b), 82% of the consumers believe that their decisions of buying products are directly influenced by online reviews and over 75% of them will consider the recommendations from buyers who have experience of a product. An online product review normally consists of information about a specific product or service, serving as a recommendation (Park *et al*., 2007), as well as delivering additional user-oriented information. Product review is of great importance, as it has strong influence on the process of decision-making for consumers (Engel *et al*., 1969; Gilly *et al*., 1998). Because product reviews are regarded as a more trustworthy information source than company-generated persuasive messages (Feick and Price, 1987), consumers often rely on reviews in online shopping. Prior research has examined the significant impact of product reviews on subsequent sales (Chevalier and Mayzlin, 2006; Goldsmith and Horowitz, 2006), consumer purchase decision making (De Bruyn and Lilien, 2008) and consumer attitude towards the brands (Lee *et al*., 2009).

The behaviour of consumers engaging in eWOM on social network sites (SNSs) has received mounting attentions from researchers, marketers and policy makers (Chu and Kim, 2011; Ellison *et al.,* 2007; Valenzuela *et al.,* 2009). The SNS has become an important component of integrated marketing communications (Mangold and Faulds, 2009). It is an ideal environment for eWOM communication, as consumers freely disseminate product-related information in their established social networks composed by their friends (Vollmer and Precourt, 2008). Given the social and commercial characteristics of SNSs, consumers are more likely to engage in social interactions like commenting or liking. Through these interactions, consumers spontaneously show their preferences along with their persona, which can engender more eWOM communications (Chu and Kim, 2011). Thus, understanding their engagement to eWOM on SNSs is of importance.

Given that social network connects users composing individuals' friend circles, consumers are normally exposed to information from both friends and crowds. Friends who have directed connections with a target person on social medias are perceived as more trustworthy and credible sources of product information for consumers than unknown strangers. In addition, due to the frequent interactions, eWOM from friends has strong influence over ones' engagement in eWOM (Centola, 2010; Crandall *et al*., 2008; Dellarocas and Narayan, 2006; Aral and Walker, 2011). Actually, such influence of social tie on one's behaviour has long been reported in social science (Brown and Reingen, 1987; Steffes and

Burgee, 2009; Kawachi and Berkman, 2001; Wang and Chang, 2013), and is normally recognised as a social contagion phenomenon (Aral and Walker, 2011; Aral *et al*., 2009). However, on the other hand, the crowd has a much larger population than that of a consumer's friend circle, which implies that the reviews and opinions would be of much greater diversity and richness. Therefore, reviews from crowds may provide much more information for consumers to make purchase decisions and further engage in eWOM discussions. Consequently, whether the friends' or the crowds' eWOM is more influential for a consumers' subsequent behaviour is still unknown.

Despite the convenience and value created by eWOM communications in the information age, the pressure of accessing such overwhelming information is still a severe issue. For e-commerce businesses, they naturally expect more and more consumers participating in product discussions, but the consequence that they are facing is how to detect the most helpful product reviews from the overwhelming ones for consumers. Recognising "review helpfulness", such problem has also been widely discussed (Zhu *et al*., 2014; Qazi *et al*., 2016; Pan and Zhang, 2011; Mudambi and Schuff, 2010).

In comparison to shopping at physical stores, online shopping is unique in its temporal and spatial separation of buyers and sellers (Luo *et al*., 2012). Consumers have to face more uncertainty and risk in online shopping than offline, because it is difficult for consumers to experience products before they buy them. Many e-commerce websites implement peer voting mechanism to decrease such uncertainty. For example, Amazon asks a question that "was this review helpful? (Yes or No)" following every review to collect votes from users. Those that get the most votes of "Yes" are regarded as the most helpful reviews and will be displayed in the most prominent place. According to the Spool, this simple mechanism could bring a large amount of revenue. However, user-specified feedback is too sparse to detect the helpfulness of reviews. A large proportion of reviews have few or no helpfulness feedback, in particular for the most recent ones (Lu *et al.*, 2010). Newly-generated reviews normally do not have enough time to accumulate helpful votes (Kim *et al*., 2006). Therefore, it is important to understand factors that are contributing to the helpfulness of reviews and as such to assess the helpfulness automatically.

To better help consumers find proper products quickly, recommender system is often utilised in many ecommerce websites. Relevant products are connected by hyperlinks for bundling sales or navigating consumers to appropriate products (Hou *et al*., 2017). Considering the products as nodes, and the recommendation hyperlinks as connections, such system is normally referred to as product recommendation network (PRN) that can be observed in a wide range of ecommerce websites. One of the most well-known examples is the “co-purchase” recommendation in Amazon, entitled “Customers who bought this item also bought”. Consumers can thereby quickly find the most suitable product by following the recommendation trails.

Research has suggested that the presence of PRN has significant economic impact on product performances such as sales and demands (Oestreicher-Singer and Sundararajan, 2012a). For example, the demand of a product measured by consumer willingness to purchase a product is closely associated with its position in the PRN, such as its in-degree that is the number of edges directed into a vertex, PageRank centrality that is a measure of importance of nodes (Oestreicher-Singer and Sundararajan, 2012b), as well as other centrality measures (Leem and Chun, 2014). Lin *et al.* (2017) showed that the network diversity and network stability have also significant influence over the product demands. In particular, the demand of products is shown to be spreading in the PRNs (Carmi *et al*., 2017). While external events, such as book reviews in the TV show, may boost the sales of the corresponding products, such boost can be surprisingly observed in the products that are up to three clicks away from the focal product.

Since both PRN and eWOM have significant value for consumers and ecommerce, research regarding eWOM in the context of PRN has caught some attentions. It is found that the directly-connected products in a PRN have similar eWOM rating with each other (Lin and Wang, 2018). However, relevant study is still scarce. Users normally surf on such network, from one product to another following the hyperlinks to find the appropriate products. Therefore, whether a pair of products with a certain distance have similar eWOM and how is the distance between them influencing such similarity are at the core to the understanding of the consumer behaviour and product performances in the PRN.

1.2 Research Questions and Objectives

Given the significance of eWOM, this PhD thesis aims to explore the impact of online product reviews on consumer behaviour and uncover the relationship between them in the context of social network and product recommendation network, as well as examining the factors that are related to the helpfulness of old reviews and new reviews.

To achieve the goals set above, the research follows four steps, each one focusing on a major research question described as follows.

Considering the significant value of product reviews and the popularity of social networks in ecommerce websites, we firstly investigate the effect of two types of product reviews: friend reviews and crowd reviews, which has been seldom investigated in prior literature. In this thesis, friend reviews that are exposed to a user only refer to the reviews posted by the users' directly connected users on the social medias. We explore:

Research Question 1 (RQ1): How friend's and crowd's reviews differently impact consumer engagement on posting behaviour? To answer the question, we aim to:

- examine factors of online product review that are related to consumer posting behaviour;
- develop a model to distinguish the different impact of friend and crowd review on consumer posting behaviour.

As described, social network connects consumers on ecommerce websites, review posting behaviour can thereby be "pried" by friends. Consequently, products or services that consumer would select may be undermined by social influence. Hence, we explore:

Research Question 2 (RQ2): How would social influence affect consumers' selection behaviour? To answer the question, we aim to

- confirm the existence of social influence on consumer selection behaviour;
- distinguish the effect of friends-based and crowds-based social influence;
- develop a model to quantity the intensity of friends-based and crowds-based social influence.

Review helpfulness largely support the needs of online consumers to quickly locate the most relevant product review. Thus, it has caught interests from website design and marketing. However, a very fundamental question is what makes a review helpful. Hence, this thesis explores:

Research Question3 (RQ3): Which factors can be used to predict review helpfulness? To do so, we aim to

- identify the influential factors for review helpfulness;
- examine such factors for the helpfulness increment for old and recent reviews respectively.

PRNs help consumers to explore different products and thereby make purchase decisions quickly, which have similar functions of eWOM. To date, the research on the relationship between eWOM and PRNs is scarce. Thus, we study:

Research Question 4 (RQ4): Do products which are close to each other in PRN have similar eWOM? The objectives include to:

- study the impact of direct and indirect connections between products on their eWOM information;
- develop models to systematically discuss the impact of distance between products on their eWOM similarities.

1.3 Organisation of the Thesis

The present PhD thesis is organised as shown in Figure 1.1, where the chapters demonstrate the overall flow of activities to address the aims of the research. Building on the introduction above, the rest of this thesis is structured as follows:

Chapter 2 presents the literature review. First, the chapter discusses some fundamental theory of information diffusion and social influence. Then the chapter introduces the theory of consumer search information, which provides antecedent for understanding the impact of information generated by consumers. Thereafter, the chapter introduces the research regarding online users-generated content, mainly laying out the studies of eWOM based on the view of antecedent and consequence of senders and receivers. At last, this chapter discusses social network analysis and introduces two kinds of network: friendship network and product recommendation network.

Chapter 3 introduces research methodology. The chapter firstly introduces two research philosophy in social science research: ontology and epistemology. Then the chapter present the methods of quantitative research. Following this, the changing of research methodology in social science to computational social science is presented. In the last, the chapter introduces the research mythology of the present thesis, which is multi-disciplinary by combing a series of approaches.

Chapter 4 is the data collection of this thesis. This thesis majorly applies two data sets. One is an open-source data from Yelp which will be applied in Chapter 5, 6 and 7, and the other one is collected by ourselves from Amazon which will be applied in Chapter 8.

Chapter 5 studies the impact of friend reviews and crowd reviews on consumer engagement on posting behaviour, tackling **RQ1**. We apply large-scale review data from Yelp.com, and compare the impact of friends' and the crowd's reviews on how consumers engage in posting reviews based on three aspects: volume, valance and variance. We develop a multilevel mixed-effect Probit model to examine posting behaviour in relation to review information and consumer characteristics. The chapter is largely based on our published paper:

- **Pan, X.,** Hou, L., Liu, K., and Niu, H. (2018). Do reviews from friends and the crowd affect online consumer posting behaviour differently? Electronic Commerce Research and Applications, 29, 102-112.

Chapter 6 explores the intensity of two social influence which is measured by the objects' popularity on consumer selection behaviour, tackling **RQ2**. We propose a network model to describe the mechanism of user-product interaction evolution with social influence. The chapter is largely based on our published paper:

- **Pan, X.,** Hou, L., and Liu, K. (2017). Social influence on selection behaviour: Distinguishing local-and global-driven preferential attachment. PloS One, 12(4), e0175761.

Chapter 7 explores the question that what features make reviews helpful and how to predict the helpfulness of reviews in the future, tackling **RQ3**. We adopt a dynamical method for data collection at two different time points to study the increment of helpfulness for both old and recent reviews. The chapter is based on a conference paper and a working paper:

- **Pan, X.,** Hou, L., and Liu, K. (2017). On the prediction of future helpfulness for online reviews. In: *Proceedings of International Conference on Logistics, Informatics and Service Sciences*. Kyoto, Japan.
- Pan, X., Hou, L., and Liu, K. Predicting the future increment of review helpfulness: an empirical study based on a two-wave dataset. (working paper)

Chapter 8 studies the impact of distance between products on their eWOM performance, tackling **RQ4.** This chapter employs an empirical book recommendation network collected from Amazon along with the eWOM information of every book to explore the effect of product distance on their eWOM on two levels: Neighbourhood Level and Dyadic Product-pair Level. The work is based on our working paper:

- Pan, X., Hou, L., and Liu, K. The Effect of Product Distance on the WOM in Recommendation Network. (working paper)

Chapter 9 summarises the major findings and contributions, and discusses the limitations of this work and the recommendations for future work.

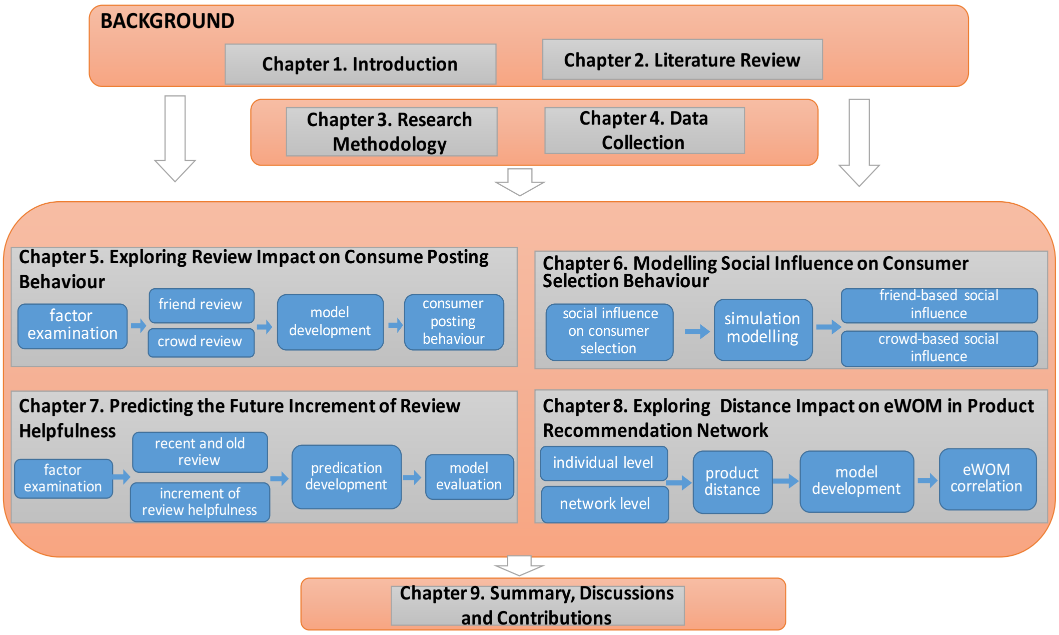

BACKGROUND
Chapter 1. Introduction
Chapter 2. Literature Review
Chapter 3. Research Methodology
Chapter 4. Data Collection
Chapter 5. Exploring Review Impact on Consume Posting Behaviour
factor examination
friend review
crowd review
model development
consumer posting behaviour
Chapter 6. Modelling Social Influence on Consumer Selection Behaviour
social influence on consumer selection
simulation modelling
friend-based social influence
crowd-based social influence
Chapter 7. Predicting the Future Increment of Review Helpfulness
factor examination
recent and old review
increment of review helpfulness
predication development
model evaluation
Chapter 8. Exploring Distance Impact on eWOM in Product Recommendation Network
individual level
network level
product distance
model development
eWOM correlation
Chapter 9. Summary, Discussions and Contributions

Figure 1.1|The research design of the thesis

Chapter 2. Literature Review

In this chapter, literature on some of the key aspects relating to this thesis is reviewed comprehensively. As suggested by the major objectives of the thesis, this chapter starts by introducing some fundamental theory of information diffusion and social influence. This is followed by the theory of consumer search information, that provides antecedent for understanding the impact of information generated by consumers. Section 2.3 introduces the research regarding online user-generated content, where the studies of eWOM based on the viewpoint of antecedent and consequence of senders and receivers are reviewed. In addition, the semiotics theory is reviewed in this section to discuss the interplay between eWOM, consumer behaviour, consumer selection and consumer purchased decisions. Section 2.4 discusses social network analysis and introduces two kinds of network: friendship network and product recommendation network, which will be studied in this thesis.

2.1 Information Diffusion and Social Influence

Diffusion is a social process through which new ideas, technologies or products spread among the members of a particular social system via specific communication channels over time (Kreps, 2017). It is a specialised form of communication that focuses on disseminating information. Information diffusion is quite important in such a society that new scientific findings, technologies, products and so on are continually coming out. A large amount of research in many fields such as agriculture (Fliegel, 1993), technologies (Palmer *et al.,* 1993), and policy innovations (Berry and Berry, 1992) and so on, has been conducted based on the diffusion of information over the past decades.

Initially, most of the research on information diffusion were from the sociologists, epidemiologists and economists focusing on the diffusion of innovation, epidemic and product respectively, through real networks. Previous studies investigated the flow of information based upon the analogy between information propagation and the spread of disease in networks. Thus, the questions of information diffusion are solved based upon the study of epidemiology (Bailey, 1975). One of the classical disease-propagation models in epidemiology is *Susceptible-Infected-Recovered (SIR)* model, which is based on the cycle of disease in a host. A person is first *susceptible (S)* to the disease. If this person is exposed to the disease by an *infectious (I)* contact, the person becomes infected with some probability. The disease then runs its course in that host, who is subsequently *recovered (R).* A recovered individual is immune to the disease for some period of time, but the immunity may eventually wear off. Such *SIR* model has been applied to describe the process of information diffusion among networks (Iribarren and Moro, 2009; Guille *et al.,* 2013). Initially all nodes are in the susceptible state except one infected node (namely the information source). The information source then infects its neighbours, and the information starts to spread in the network. Taking an example of blogspace in Gruhl *et al.,* (2004), a blogger who has not yet written about a topic is exposed to the topic by reading the blog of a friend. She has a probability to decide to write about the topic, becoming infected. The topic may then spread to readers of her blog. Later, she may revisit the topic from a different perspective, and write about it again.

2.1.1 Innovation Diffusion Theory

Diffusion of innovation theory is first introduced by Rogers (2010, the fifth edition), which explains the influence of communications on the adoption of new ideas, technologies and processes. According to the theory, the exposure to information through communication across social networks or via different media channels has a profound impact on the rate of adopting new ideas, behaviours or products. The area of innovation diffusion ranges quite broadly, from the adoption of hybrid seed corn by famers (Ryan and Gross, 1943) to applications of modern mathematic formulas by mathematicians, to the purchase of new cars by consumers, to the adoption of recommended health behaviours (Rogers, 2010; Haider and Kreps, 2004; Oldenburg and Glanz, 2008). Roger (2010) explains the process of innovation involves four main interacting elements: *Innovation, Communication, Social Systems* and *Time*.

Innovation can be an idea, practice, a new technology or even a set of behaviour such as health communicators help smoker to enrol in a specific new and promising tobacco control smoking cessation. *Communication* is the means by which message get from one individual to another. The nature of information-exchange determines the conditions whether a source will transmit the innovation to the receiver or not. Mass media channels, such as radio, television, newspaper, are the most rapid and efficient ways to inform a potential audience about the innovations. *Social System* is defined as a set of interrelated units that are engaged in joint problem solving to accomplish a common goal (Rogers, 2010). Innovation diffusion occurs within the social system where members cooperate seeking a common problem. *Time* refers to how long it is likely to take intend audience members to learn and adopt the innovations.

The model of innovation diffusion has been utilised to address different scenarios. For example, in the field of public health, the adoption of HIV/AIDS prevention behaviours (Bertand, 2004; Collins *et al.,* 2006), skin cancer prevention guidelines (Buller *et al.,* 2005; Escoffery *et al.,* 2007) and heart disease prevention (Scott *et al*., 2008) have applied the theory of innovation diffusion. Many others subjects that can be diffused, such as technologies (Venkatesh *et al*., 2003; Moore, 1991), knowledge (Kogut, 1992), services (Greenhalgh *et al.,* 2004) and so on have also been addressed base on innovation diffusion theory.

Since there are many literatures on diffusion research and a lot of variables have been defined influencing on an actor’s decision to adopt an innovation, Wejnert (2002) proposes a conceptual framework to integrate these variable. The framework is derived by grouping diffusion variables into three major components: innovations’ characteristics, innovators’ characteristics and environmental context.

The innovations’ characteristics consider two factors associated with innovations: 1) public versus private consequences (Strang and Meyer, 1993; Meyer and Rowan, 1977) which refer to the impact of an innovation’s adoption on entities (public consequences) versus that on the actor itself (private consequences); 2) benefits *versus* costs (Greve, 1998; James, 1993) which is caused in the process of adoption of innovation. The characteristics of innovators (actors) may influence the perception of an innovation's costs and benefits. Six sets of variables of innovators modulate the adoption of innovation, including societal entity of innovators (Bloom, 2002), familiarity with innovation (Meyer and Rowan, 1977; Mizruchi, 1993; Newel and Swan, 1995; Weimann and Brosius, 1994), status characteristics (Baerveldt and Snijders, 1994), socioeconomic characteristics (Mahajan and Muller, 1994), relative position in social network (Valente and Rogers, 1995), and personal characteristics (Weimann and Brosius, 1994). The successful transfer of innovation depends on suitability to the environment context where they encounter during diffusions (Ormrod, 1990). The environmental variables include geographic settings that influence the applicability of innovation to the ecological infrastructures of potential adopter (Fliegel 1993, Saltiel *et al.,* 1994); societal culture that reflects adopters' value, norms, language *etc*. (Tolnay, 1995; Straub, 1994); political conditions (Berry and Berry, 1990) and global uniformity that reflects the view of the contemporary world as one cultural community (Weimann and Brosiu, 1994; Mahajan and Muller, 1994).

2.1.2 Information Diffusion through Online Social Network

The research of information diffusion is broad. Most of the early research on information diffusion is based on the real networks in human society whereas the data collection is limited. The qualitative analysis towards small dataset were thereby commonly accepted. But nowadays, with the popularity of social media which has become the most pervasive channel to diffuse information (Bakshy *et al.,* 2012), a large amount of online data can be easily obtained. These data include not only the social network data, but also the massive information diffusion among networks, which enlarges research topics in the field of information diffusion. Since the field normally considers network structure analysis, textual analysis, it has attracted many computer scientists and physicists who come from the fields of complex network, natural language processing and information retrieval.

Online social network is one of the most commonly-discussed environments for information diffusion. Normally, online social network is composed by nodes and edges that represent users and relationships respectively. Users publish messages to share or forward information such as product recommendations, political opinions and so on. Thus, the messages are spreading through edges to other users. Figure 2.1 shows a graph that messages generated by four users spread in an online social network. As shown in this figure, users generate messages and spread them to others by network connections, for example, the user u_3 would receive the messages, m_1, m_2 and m_3 from u_1. The graph indicates that there is no one receives the messages from user u_4.

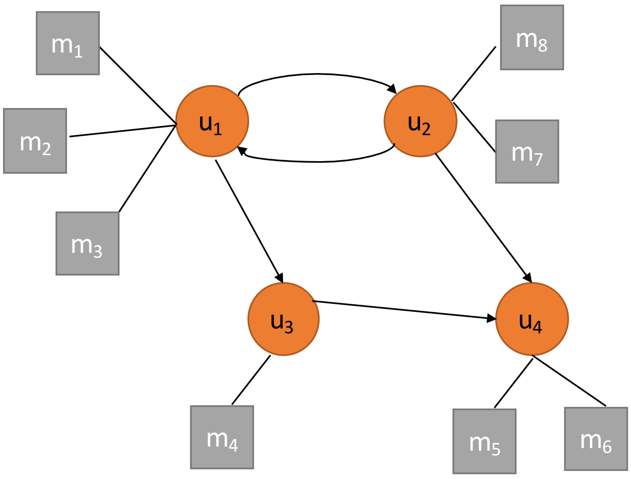


Figure 2.1|An example of online social network enriched by users' messages. The orange circles represent users and grey squares are messages generated by corresponding users. The directed arrow means a message is diffused from one person to another.

Research on information diffusion in online social network mainly raises the following questions: 1) which pieces of information or which kinds of topics are popular or can be diffused the most; 2) how to model the speed of diffusion or how to maximum the diffusion speed and 3) how to identify the most influential information spreaders. This section reviews relevant research from the three aspects.

The probability of information diffusion in term of topics have been studied (Liu *et al.,* 2010; Tang *et al.,* 2009). The results suggested that users in online social network are usually interested in many topics and the probability of diffusion of different topics is different (Granovetter, 1977; Krackhardt *et al.,* 2003). For example, the news of NBA diffuses faster among basketball fans than the news about technology. Detecting popular topics is one of the main tasks when studying information diffusion. A lot of methods were designed to detect topics that have drawn bursts of interest which means a topic gets extensive treats within a time period but rare before and after (Shamma *et al*., 2011; AlSumait *et al.,* 2008; Cataldi *et al.,* 2010). These methods are all based on frequencies of discretised messages.

The diffusion process is characterised by its structure, i.e. the diffusion graph that transcribes who influenced whom and its temporal dynamics that refers to the evaluation of diffusion rate. The models that capture and predict the speed of information diffusion are distinguished into two categories: explanatory models and predictive models. The explanatory models aim to infer the underlying spreading cascade and retrace the path taken by a message. Assuming the online social network remains static over time, researchers proposed several models to explore the correlations between nodes' infection times to infer the structure of the spreading cascades and assume that activated nodes influence their neighbours independently with some probability (Rodriguez *et al*., 2011; Gomez *et al.,* 2010). But the social network evolves quickly, thus a time-varying inference model was extended by Gomez *et al.* (2013) to estimate the structure and temporal dynamics of a network that change over time.

Identifying the most influential information spreader in a network is quite important for companies for ensuring maximisation of information diffusion. For example, a small company is going to launch a new product, but due to the limitation of marketing, they can only choose some influential users who try the product by free and future trigger large cascades of adoption by recommending to their friends or followers. Various methods for targeting influential individuals have been examined. Basically, the most efficient spreaders are regarded as the core of the network (Kitsak *et al.,* 2010) which is identified by the K-core decomposition (Seidman, 1983) or PageRank algorithm (Cataldi *et al*., 2010) based on only the topology of the network. Later, the models involving network topology, features of nodes, and the way they spread information have been developed. For example, Romero *et al.* (2011) develop a graph-based approach that assigns a relative influence and a passivity score to every user based on the ratio at which they forward information. Pal and Counts (2011) define a set of nodal and topical features for characterizing network members to develop a non-graph based, topic-sensitive method.

2.1.3 Social Influence

During the process of information diffusion in online social network, the phenomenon of social influence happens any time that actions of a user can induce his connectors to behave in a similar way (Anagnostopoulos *et al.,* 2008). Social influence plays a notable role in various domains of human behaviour, such as the inter-personal health (Christakis and Fowler, 2007; Christakis and Fowler, 2008), political attitude (Kenny, 1992; Bond *et al.,* 2012) and cultural product consumption (Salganik *et al.,* 2006; Loren *et al.,* 2011). Especially in the common context when selecting from countless products, such as books, movies, restaurants etc., people frequently look at others' decisions (Banerjee, 1992; Tucker and Zhang, 2011; Liu *et al.,* 2015).

It is very easy nowadays for people to access the information of objects such as qualities, ratings, popularities or even previous consumers' feedbacks from the mass media. Particularly, in many online systems, bestseller lists or highest-rated object lists are generally available for users to refer to. Those information aggregating the choices and opinions of the whole population of the system, can be recognised as the crowds-based information which has long been argued to be the key reference for human selection behaviour (*Salganik et al.,* 2006; Tucker and Zhang, 2011; Bikhchandani *et al*., 1998; Chen *et al.,* 2011) leading to the "rich get richer" phenomenon. A good example is the event that, two scholars secretly purchased 50,000 copies of their newly published book which consequently made the bestseller list. Then the book sold very good despite mediocre reviews, and was remained as a bestseller (Bikhchandani *et al*., 1998).

Another mainstream of the social influence research believes that people in the same social group act similarly to each other (Lewis *et al*., 2012) since individuals are always engaged in group activities. Such source of the influence which can be regarded as the friends-based information, also drives the human selection behaviour, i.e. people tend to select what their friends selected (Crandall *et al.,* 2008; Aral and Walker, 2011). It has long been argued that, objects are similar to viruses and ideas that could spread in the social network from an individual to his friends through the frequent interactions (Centola, 2010; Muchnik *et al.,* 2013). Accordingly, the friends-based social influence is also recognised as the social contagion phenomenon (Aral *et al.,* 2009; Christakis and Fowler, 2013).

2.2 Consumer Information Search Theory

The process of consumer purchase decision-making is a series of steps processing from problem recognition, to information search, to evaluation of alternatives, to purchase decision and finally purchase. To make potentially better purchase decision, collecting information in the information search stage is essential. Consumers require information searching even when they do not plan to buy products in the short term but sometime in the future. The stream of consumer information search is one of the most enduring field in consumer research (Beatty and Smith, 1987). As one of the key components during the process of consumer purchase decision-making, information search has been included in many information processing models (Payne *et al.,* 1991; Engel *et al.,* 1993).

2.2.1 Influential Factors for Information Search

Prior research attempted to use various variables to explain consumer information search behaviour including environmental-related factors like difficulty of choice task, alternative numbers, situational variables (e.g. perceived risk) and consumer characteristics (e.g. education, involvement).

Consumers' ability to search is one dimension of the determinants for information search activity. Perceived ability to search is defined as the perceived cognitive capability of searching and processing information (Schmidt *et al.,* 1996). The ability of searching positively influences information search (Duncan and Olshavsky, 1982; Srinivasan, 1987). Three main factors that determine one's perceived ability to search for information, namely educational level, objective product knowledge, and subjective product knowledge, have been studied. The research suggest that higher level of education lead to increased search activity (Schaninger and Sciglimpaglia, 1981; Kiel and Layton, 1981). Objective product knowledge is conceptualised as what consumers already know, whereas subjective product knowledge is defined as the consumers' perception of the amount they know about the product domain. Consumers with higher level of knowledge have well-developed knowledge structures and are able to comprehend and organise information more easily than consumer with lower knowledge level (Chase and Simon, 1973). Thus, high levels of consumer knowledge increase one’s ability to engage in search for information (Schmidt *et al.,* 1996; Miyake and Norman, 1979).

The motivation to search is another factor influencing information search activity, which is described as the desire to expend effort in the collection and processing of information. Schmidt *et al.* (1996) specify the motivation to search to be influenced by enduring involvement, need for cognition and shopping enthusiasm. The research suggested that consumers are more likely to engage in search information when the involvement is high (Engel *et al.,* 1993) and will spend more time in searching (Celsi and Olson, 1988). Need for cognition is defined as the tendency for individual to engage in and enjoy thinking. The research on need for cognition suggest that consumers have greater desire to process information and will gain more enjoyment in searching information (Schmidt *et al.,* 1996). Shopping enthusiasm shows similar impact that higher shopping enthusiasm increases one’s motivation to engage in information searching.

2.2.2 Online Consumer Information Search

The Internet provides consumers unprecedented benefits. It provides effortless and costless information that facilitates better decision making (Alba *et al.,* 1997). Underhill (2009) states that “limitless amounts of product information and other reading materials can be summoned and then saved, all in an instant, far beyond anything possible in the real world of brochures, manuals, and the memory and knowledgeability of salesclerks”. The Internet provides capability of inexpensively storing vast amount of information in different virtual locations and capabilities of interactivity. All these characteristics of Internet lead to the fact that consumers are increasingly relying on the Internet when searching information.

Research showed that the Internet is being searched when a consumer wants to gain information about a product or service as well as information of a brand (Breitenbach, 1998; Shim *et al*., 2001). A lot of factors have been examined to be influencing information search behaviour on the Internet, such as consumers’ characteristics (King *et al*., 1989; King *et al.,* 1994; Saarinen, 1996), consumers’ lifestyle (Kim and Kwon, 1999), demographic characteristics (King *et al*., 1989; King *et al.,* 1994; Saarinen, 1996) and consumers’ information search type (continuous or pre-purchase search) (Bloch *et al.,* 1986), as well as the design of websites (Jun and Cai, 2001; Muir and Douglas, 2001).

2.3 User-Generated Content and eWOM

2.3.1 Online User-Generated Content Website

Web 2.0 created an era of content generation in which online information can be easily created and shared by users. The information posted by users, which commonly referred as user-generated content (UGC) (Peck *et al*., 2008), can be assessed by different parties including firms, consumers, manufactures and so on. The websites, which allow users to post, share, spread ideas or opinions by different types of contents, such as blogs, product reviews, and videos and so on, are referred to as online user-generated content websites.

Firms become interested in mining consumer opinions and ideas from these UGC websites toward their products or brands. For instance, microblogs as a form of UGC which are posted on Twitter, have been investigated as an online tool for consumers' word of mouth communications and corporations' marketing strategy (Jansen *et al*., 2009).

YouTube, a video created platform, allows users to create their own channel and post content that can be shared almost instantaneously to a wide audience across the world. Research suggests that a user on YouTube can engage in self-expression (Raymond, 2001) as well as obtain peer recognition (Resnick *et al*., 2000) because of the ease of creating a personalised page or channel. Susarla *et al*. (2012) examine the impact of social contagion through the network on YouTube on the video diffusion. They found that social interactions between users are influential in determining which videos to become successful.

Facebook, as one of the most famous UGC websites, enables consumers to generate massive amount of brand-related information. For example, consumers join a group of a certain brand on Facebook where they can share brand experience and post-consumption evaluation. The information shared by users could influence brand perception and purchase intensions (Zhao *et al.,* 2008). Research shows that social media act as a check on brand credibility (Lee and Kim, 2011). Thus, advertising on UGC websites has occupied a reasonable proportion of budget in advertising for many firms (Falls, 2009).

Being one of the most well-known ecommerce websites all over the world, Amazon sells a large amount of products as well as allows consumers to submit product reviews after purchasing. These consumer-generated reviews have significant impact on consumer behaviour (Chevalier and Mayzlin, 2006; Schlosser, 2005) and subsequent product sales (Godes and Mayzlin, 2004; Chevalier and Mayzlin, 2006; Dellarocas *et al.,* 2004; Duan *et al*., 2008).

Besides websites described above, there are many other UGC websites, such as Tripadvisor, Yelp, Epinions and so on. Although they provide different types of functionalities for user social interaction, they have similar mechanism encouraging users to participate in generating information on their websites.

2.3.2 eWOM and Semiotics

eWOM, which is a typical type of user-generated content, has attracted considerable interest from researchers in the past decades. It is defined as “any positive or negative statement made by potential, actual, or former customer about a product or company, which is made available to a multitude of people and institutions via the Internet.” (Henning-Thurau *et al*., 2004).

The influence of eWOM on consumers can be theoretically explained via the theory of semiotics. Semiotics is the theory of signs or the “formal doctrine of signs” (Liu, 2000). A sign, which is defined by Peirce (1960), is something which stands to somebody for something in some respect or capacity. Signs may take various forms, e.g. words, text, sounds, images and objects or artefacts. The semiotics only concerns the meaningful use of signs that can be transferred to information (Saussure, 1916; Peirce,1960; Chandler, 2007). Saussure’s semiotics theory argues that language is a system of signs that express ideas (Saussure, 1916), and the science of signs is named by semiology. Given that WOM and eWOM are communications amongst humans and languages is the major means of humans’ communication (Mateas and Sengers, 2003), the theory of semiotics from the linguistic field may enhance the understanding the peer-to-peer interactions.

Peirceian model characterises a triadic relationship of a sign, composing by sign vehicle (or representamen), an object (or semiotic object) and an interpretant (Peirce,1960). Sign vehicle means the specific physical form of the sign. An object represents the aspect of the word that the sign carried meaning for. An interpretant is used to meditate the relationship and helps establish the mappings between the signs and objects. Liu (2000) states that the interpretant associated with the sign and with the object based on the knowledge or norms. Taking the traffic sign "STOP" as an example, only people who know English and traffic regulations can learn the meaning of sign "STOP." Thus, the interpretant is prone to be misunderstood and probably results in semiotics gap without extra information (Liu, *et al.,* 2005). Taking the semiotic theory into the study of eWOM, the information that consumers are exposed to can be regarded as signs; consumers can be objects, and the interpretants can be the understanding of eWOM for consumers.

The theory of semiotics has been developed into several branches, such as organisation semiotics, which aims to study the sign, information and human communication in an organisational context (Stmaper *et al.,* 2000; Liu and Li, 2015; Liu, 2000). Perceiving organisations as information systems made up of signs (Liu *et al.,* 2003), organisational semiotics focuses on the exploration of the complex nature and characteristics of information to embrace the digitalisation, innovation and transformation for organisations.

2.3.3 Research Agenda of eWOM

Basically, the major research questions related to the eWOM can be summarised by the framework shown Table 2.1 (Nyilasy 2005; King *et al.,* 2014). The framework assumes every WOM episode have several antecedents and consequences for two parties: the senders and the receivers. In this section, the progress of eWOM research will be reviewed according to this framework.

Antecedents of eWOM sender

In the early stage of WOM studies, the motivation of people to engage in the traditional WOM behaviour is widely discussed. Dichter (1966) identify four motivations that drive individuals to engage in it: perceived product involvement; self-involvement (e.g. gratification of emotional needs from the product); other involvement (e.g. the need to give something to the person receiving the WOM transmission); and message involvement (e.g. the way to present the product in the media). Sundaram *et al.* (1998) argue that the desire of consumers for altruism, product involvement and self-enhancement are the main factors leading to positive WOM, while negative WOM is normally due to anxiety reduction and vengeance.

Table 2.1|eWOM organizing framework (adopted from Nyilasy 2005).

		Study	
		Antecedents of eWOM (causes)	Consequences of eWOM (effect)
Unit of analysis	Sender of eWOM	Q1: Antecedents of eWOM sender---why do people talk online?	Q2: Consequences of eWOM sender--- what happens to the communicator?
	Receiver of eWOM	Q3: Antecedents of eWOM receiver---why do people listen online?	Q3: Consequences of eWOM receiver--- the power of eWOM

Considering the environment of WOM discussions becoming online, Hennig-Thurau, *et al.* (2004) conclude that social benefits, economic incentives, concern for others, and extraversion are the main motivations for consumer to participate in WOM on the Internet. Several other factors have been examined in the pervious research, such as self-enhancement (Angelis *et al.,* 2011; Fiske, 2002; Wojnicki and Godes, 2008), innovativeness and opinion leadership (Sun *et al.,* 2006), ability and self-efficacy (Gruen *et al.,* 2006; Huang *et al.,* 2009), individuation (Ho and Dempsey, 2010), neuroticism (Picazo-vela *et al.,* 2010).

Community engagement, enhanced volume, persistence, and observability determine why and how users create and transmit eWOM (King *et al*., 2014). Consumers are in various kinds of communities on the Internet, where they have social connections with each other. Research suggested that both social connection and community engagement drive the incidence and types of eWOM messages, even if the connections between eWOM senders and receivers are weak. Sohn (2009) examines how network strength and the valance of messages affect consumers' propensity to transmit eWOM. He finds that consumers are more likely to transmit messages to strong ties than to weak ties, but the effect can be moderated by perceived quality of information. It is more possible to pass on negative eWOM to weak ties for consumers, whereas they share both positive and negative eWOM with strong ties.

Prior research suggest that message volume could encourage consumers to provide their opinions. For instance, Moe and Schweidel's (2012) find that products with a greater number of reviews tend to attract even more reviews. This leads to the phenomenon of "richer get richer" that a few products and services tend to attract of substantial amount of reviews whereas the majority only have few or no reviews. As such, self-selection bias is significant in marketing implications (Hu and Li, 2011). However, users will avoid following the majority due to a desire to be different from others (Khare *et al*., 2011). The impact of volume on engagement in eWOM may be different based on the types of individual personality.

Valance, which is normally quantified by the average rating also effects the generation and transmission of eWOM. Dellarocas and Narayan (2006) find consumers typically post reviews for either very good or very bad movies (in other words, the ratings follow U-shaped distribution). Consumers' propensity to post eWOM is positively correlated with the level o44f disagreement among opinions (Dellarocas *et al*., 2010).

Consequences of eWOM sender

The question what eWOM senders would get after they post eWOM information seems to have got imitated attentions. Instead, researchers pay much more attentions on how does eWOM affect receivers' choices and behaviours. Social capital and reputation are the main outcome for eWOM senders (Chen *et al.,* 2010; Dholakia *et al*., 2009). People's status and reputation in a community are increased by the willingness to help other, high quality information and impressive technical details in one's answer (Kollock, 1999). Posting valuable information for eWOM seekers gives consumers high reputation, which in turn force consumers to make efforts to generate more for maintaining reputation. The positive reinforcing mechanism has been found in several studies. For example, Cheung and Lee (2012) find that reputation, a sense of belonging and enjoyment of helping others are the main drivers of intensions to generate eWOM. Chen *et al.* (2010) find that users on Movielens who are provided with outcome information like contributions' votes from other readers are more likely to post better-quality reviews for better recommendation.

Antecedents of eWOM receiver

Many research answers the question why consumers listen to or seek eWOM. A lot of factors that motivate consumers to engage to online opinion seeking have been measured. Firstly, searching the opinions of others online is an easy way to get information and secure lower prices, so that consumer can reduce their search efforts (Dabholkar, 2006; Goldsmith and David, 2006). Also, consumers would like to reduce evaluation efforts in both pre- and post- purchase (Bronner and de Hoog, 2010; Hennig-Thurau and Walsh, 2003). Secondly, online shopping is different from brick-and-mortar stores shopping, that consumers can not truly know the quality of goods. The uncertainty and risk can be reduced by learning the information shared by prior buyers (Kim *et al.,* 2011; Sweeney, Soutar *et al.,* 2008). Thirdly, finding social assurance is another reason for consumer to seek eWOM (Bailey, 2005; Kim *et al.,* 2011). A sense of belong to a community makes consumer feel much better, when they recognise that they are not the only ones concerning some problems. Lastly, consumers tend to communicate their online shopping experiences when the experience is negative and shows bias from previous eWOM, due to "concern for other customers" (Hennig-Thurau *et al.,* 2004). Thus enacting negativity bias (O'Reilly and Marx, 2011; Schlosser, 2005) is another motivation for eWOM receivers to generate new eWOM.

Beside the main factors shown above, message- and source-related factors have been examined. For example, Weiss *et al*. (2008) investigate the factors that lead to consumers' perceived of value and the subsequent adoption of information online. They found that the response speed of information senders and the extent to which its previous responses have been positively evaluated by others affect the judgements of information value. Mudambi and Schuff (2010) analyse perceived helpfulness of product reviews based on helpful votes that individuals receive from other consumers. They found that moderate review and the amount of message positively influence the perceived helpfulness of those reviews.

Consequences of eWOM receiver

The most significant impact of eWOM on receivers is that it facilitates consumers' purchase decisions because the information accessing helps them to find products that best meet their needs and preferences (Dellarocas, 2003). The large amount of eWOM reduces uncertainly and search cost, leading to a greater willingness to buy products (Brynjolfsson and Smith, 2000). Consequently, eWOM information has significant impact on products sales (Chevalier and Mayzlin, 2006; Li and Hitt, 2008; Duan *et al.,* 2008).

Research suggests that eWOM has impact on consumer trust and loyalty towards products (Awad and Ragowsky, 2008; Ba and Pavlou, 2002; Gauri *et al*., 2008). Awad and Ragowsky (2008) explore that the impact of gender on the effect of eWOM on trust, and found that the effect of eWOM on trust of an online vendor is significantly different in magnitude for women and men. Ba and Pavlou (2002) examine the extent to which trust can be induced by proper feedback mechanisms such as price premium.

Not only consumer engagement on eWOM belongs to the quadrant of "antecedents of eWOM sender", but also to the "consequences of eWOM receiver". It is shown that information readers are more likely to participate in their own opinions after they accessed many prior information posted by others (Algesheimer *et al*., 2010; Nambisan and Baron, 2007; Schau and Muniz, 2002). Algesheimer *et al.* (2010) use a field experiment on eBay to explore whether community participation leads to increased relational customer behaviours. They found consumer participation in firm's community can be increased by a simple e-mail invitation and community participation had mixed effects on customers' likelihoods of participating in buying and selling behaviours participating in buying and selling behaviours.

2.3.4 Volume, Valance and Variance of Online Reviews

Online reviews of product or service, as one of the most typical versions of eWOM, have attracted a large stream of research. Normally, three dimensions of online reviews, namely volume, valance and variance (Dellarocas and Narayan, 2006), have been widely addressed in this field.

Volume, measured by the number of reviews posted by consumers per product or service, is normally regarded as the popularity. The more popular a product is, the more likely it will be commented on by consumers. Liu (2006) and Duan *et al.* (2008) find that the volume of reviews has a positive impact on the box office sales of movies.

Valence, which is normally quantified by the average rating, can largely represent the quality of a product. Dellarocas *et al.* (2005) show that the valence of online ratings posted during a movie's opening weekend is the most important predictor of its revenue trajectory in the subsequent weeks. Similarly, Chintagunta *et al.* (2010) find that the valence of pre-release advertising is the main driver of box office performance. Studies also addressed the relation between consumer posting behaviour and the rating environment they are exposed to in term of the volume and valence of ratings. Moe and Schweidel (2012) find that a consumer is more inclined to share her experience when the volume and valence of previous ratings are high.

Variance, which is measured by statistical variance measures as well as other dispersion methods such as entropy, normally represents the fluctuation of user opinion. Clemons *et al.* (2006) find that the sales of beer grow faster for the brands with higher variance of ratings. Godes and Mayzlin (2004) find that the variance of opinions about weekly TV shows across the Internet communities positively related to the evolution of viewership of these shows. Sun (2012) investigates the interaction effect between the valence and variance of ratings.

2.3.5 Helpfulness of Online Reviews

As described in the Introduction, the problem of information overload plagues both consumers and businesses. For instance, a popular product may have thousands of reviews. It is impossible for consumers to go through each. Thus, detecting the helpful reviews from the large amount of them is of great practical significance.

In the field of review helpfulness, many factors determining whether a review is helpful or not have been examined. We review the past relevant literature from the perspectives of review level, reviewer level, and product related factors respectively (see Table 2.2).

Some studies believe that the statistical features of online reviews such as review length (Salehan and Kim, 2016; Liu and Park, 2015), review rating (Wu, 2013; Huang *et al.,* 2015), review extremity (Baek *et al*., 2012) are crucial factors to determine the review helpfulness. Mudambi and Schuff (2010) find that review depth has a positive effect on the helpfulness of reviews. Longer reviews normally cover more detailed information about the product and consumption experience, which could be valuable references for other consumers when making decisions. Therefore, consumers normally consider these longer reviews more helpful. Liu and Park (2015) suggest that long content of online reviews reflects the elaborateness, which could alleviate the uncertainty of consumers about product quality and help them to develop confidence during the process of decision making. Review rating and review extremity (Korfiatis *et al.,* 2012; Yin *et al.,* 2016; Pan and Zhang, 2011) are identified to have significant impact on review helpfulness. Typically, the numerical star of online reviews that normally ranges from one to five largely reflects consumer attitudes toward the product. Very high stars and very low stars are regarded as more helpful than indifferent stars (Liu *et al.,* 2008; Pan and Zhang, 2011; Liu and Park, 2015). Review age, which normally refers to the released time since this review was posted on the platform, is strongly associated with helpfulness as well (Racherla and Friske, 2012; Yin *et al.,* 2016; Pan and Zhang, 2011; Guo and Zhou, 2016). However, there are different findings. Yin *et al.* (2016) show that review age decreases the odds of helpful votes, but other articles (Racherla and Friske, 2012; Pan and Zhang, 2011; Guo and Zhou, 2016) report that review age is a positive factor, due to the accumulation mechanism of helpful votes on the platform.

Beside these statistical features, many studies also examine the textual characteristics such as linguistic features and writing styles (Krishnamoorthy, 2015). Such factors are closely related to readability (Liu and Park, 2015; Park and Nicolau, 2015; Ghose and Ipeirotis, 2007), which is defined as the level of comprehension of an individual for a piece of text. A review that can be easily understood would receive more attentions from interested buyers. In other word, the content with high readability may be considered helpful by a large amount of users. Besides the textual readability, stylistic choices are also found to affect the review usefulness. Kim *et al.* (2006) study how the writing style of reviews including lexical features, syntactic features and semantic features influence the review helpfulness. Each review may contain two types of words, one shows objective information such as product characteristics or product description; and the other shows sentimental information such as personal feelings and perceptions of the product. Many studies have been accordingly focusing on distinguishing sentimental opinions and statements of the fact, i.e. objective statements (Pang and Lee, 2004; Ghose and Ipeirotis, 2007). For the sentimental influence, some studies believe that affective words, both in terms of emotion variety and intensity increase the helpfulness of reviews (Martin *et al.,* 2014; Park and Lee, 2009a). Especially, some argue that the helpfulness of reviews increases when the review has a larger proportion of negative words (Baek *et al.,* 2012; Basuroy *et al*., 2003; Park and Lee, 2009a). They believe that people feel normative pressure to speak of only positive things, and thus those who talk about negative things are recognised to have persuasive effect (Ito *et al.,* 1998).

Reviewer characteristics have also been examined for the evaluation of review helpfulness. The disclosure of reviewer identity, expertise and reputation, as the most commonly discussed features, has positive effect on reviews' perceived usefulness (Liu and Park, 2015; Zhou and Guo, 2017). The reviewer identity disclosure could enhance the efficiency of information acquisition (Racherla and Friske, 2012). Meanwhile, the identifiable source of information reduces the uncertainty of reviewers (Kruglanski *et al.,* 2006) and increases the credibility of reviews (Fogg *et al.,* 2001). Baek *et al.* (2012) find that reviewer credibility plays a significant role for reviews to get helpful votes. Reviewer connectedness, which measures the established links between the reviewers and other users, is related to review helpfulness (Guo and Zhou, 2016; Liu and Park, 2015). Having more connecting users is beneficial for formulating and clarifying information needs because more information about products from other users is available. Connectedness will enhance the likelihood of accessing variance of product information and further help reviewers to make more helpful reviews. Moreover, reviewers who have more helpful votes in the past are believed to be able to produce helpful reviews in the future with higher possibilities (Guo and Zhou, 2016; Liu and Park, 2015; Baek *et al.,* 2012).

Some studies claim that the impact of those factors on review helpfulness may be different for different product types. Mudambi and Schuff (2010) find that product type could moderate the effect of review depth on helpfulness. Chua and Banerjee (2016) report that the relationship between information quality and review helpfulness varies across experience products and searching products. Hlee *et al.* (2016) show that the business type (e.g. restaurant) in the Yelp moderates the influence of presentation formats (text-based and image-based) of reviews on consumer evaluation on helpfulness.

Table 2.2|Summary of factors related to review helpfulness.

Category	Variables	Description
Review Level Factors	Review depth	number of words;
	Review rating	numerical star of the review;
	Review extremity	absolute difference between the rating and average rating;
	Review age	days since the review was posted;
	Number of concept	multiword expressions in the review;
	Review readability	ease of understanding reviews;
	Review writing style	linguistic features, review content, stylistic choices of review writing；
	Review sentiment feature	subjectivity, polarity of opinions or the ratio of positive, negative words
Reviewer Level Factors	Individual information disclosure	personal information, such as real name, image, location, reviewer identity;
	Review number	number of reviews posted by the reviewer historically;
	Friend number	number of friends or fans;
	Reviewer historic votes	helpful votes that the review got historically;
	Reviewer credibility	expert title or compliments the reviewer got.
Product Related Factors	Product type	Experience or searching goods

2.4 Social Network Analysis

The concept of “network” has been used in various scenarios and disciplines. A network normally refers to a system in which a number of actors (in particular, users and products in the thesis) interact with each other. Recently, the claim of “networks are everywhere” has become almost a routine that includes the Internet and other infrastructure networks, social, political and economic networks, as well as food webs and molecular-level biological networks (Brandes *et al.,* 2013). Barabási (2013) argues that we have seen the emergence of *network science* helping us to understand networks emerging in nature, technology and society using a unified set of tools and principles.

Network science can be applied to the description, analysis, understanding, design and repair of complex systems that are formed out of many components whose behaviour is emergent, but the behaviour of complex system cannot be easily inferred from the behaviour of its components (Bar-Yam, 1997). Thus, network science is also referred as the science of complex network. The field of complex network is big and multidisciplinary that includes Physics (Boccaletti *et al.,* 2006), Mathematics (Newman 2003), Biology (Thiery and Sleeman, 2006), Computer Science (Silva and Zhao, 2016), Social Science (Alvarez-Galvez, 2016), Political Science (Conover *et al*., 2011) and so on. Complex network theory has been applied to many complex systems such as human brain (Sporns, 2011), traffic flow (Tang *et al*., 2013), power grid (Ni *et al.,* 2007), online social network (Hu *et al.,* 2008) and so on.

One of the biggest application of complex network is Social Network Analysis (SNA). SNA is sometimes refers to as "structural analysis" (Wellman and Berkowitz, 1988). It focuses on the relationship between actors rather than individual properties. Wetherell *et al.* (1994) describe that SNA "conceptualises social structure as a network with ties connecting members and channelling resources, focuses on the characteristics of ties rather than on the characteristics of the individual members, and views communities as "personal communities", that is, as networks of individual relations that people foster, maintain and use in the course of their daily lives."

This section mainly introduces some measurements or notions of network from graph theory and reviews two kinds of social network: friendship network and product recommendation network.

2.4.1 Measurements and Notions of Network

A network is normally composed by a set of nodes $V = \{v_1, v_2, \dots v_N\}$ and links $L = \{l_1, l_2, \dots l_k\}$. In sociological research, nodes are often referred to as "actors" and a link is a pair (i, j) representing a connection between node i and node j. Figure 2.2 Shows a simple network that includes 7 nodes and 9 links. In this figure, node 5 has four connections with other nodes (2,3,7,6).

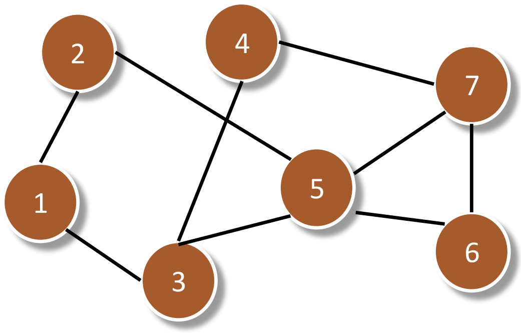


Figure 2.2| An example of network that is composed by 7 nodes and 9 links.

Neighbours

The neighbours refer to the nodes who have direct connections with the target nodes. For example, node 1 in Figure 2.2 has two neighbours: node 2 and node 3; node 5 has four neighbours: node 2, 3, 6 and 7. The second-order neighbours refer to nodes that have one connector with the targeted node. For example, node 2's neighbour, node 5 is the second-order neighbour of node 1.

Shortest-Path Length

Networks are connected, which means there normally are paths from one node to another one. Path length between two nodes refers to the number of links involved in the path. As shown in Figure 2.2, between node 1 and node 4, there are many paths such as {1,3,4}, {1,3,5,7,4}, {1,3,5,6,7,4}, {1,2,5,7,4}, {1,2,5,6,7,4}. Apparently, the length of shortest-path between node 1 and node 4 is $d_{1,4} = 2$, which is from node 1 to node 3 and then to 4. Finally, the average of shortest-path length PL_{ij}, is defined as the mean value of the shortest-path length between each pair of nodes (Watts and Strogatz 1998), which can be written as,

$$PL_{ij} = \frac{\sum_{i,j\in V; i\neq j} d_{ij}}{N(N-1)}, \tag{2.1}$$

where V is the set of nodes and N is number of nodes. The notion has great implications for real-world networks, describing how well the network is connected and how efficient the network accommodates diffusion dynamics. A famous application is "six degree of separation", which means for two random people, they can know each other though generally six intermediate friends (Thiery and Milgram, 1967). In recent years, the path length between two random people has been confirmed to be normally 4~5, such as in university email network (Kossinets and Watts, 2006), and in Facebook (Wilson *et al.,* 2009; Backstrom *et al.*, 2012).

Centrality

The centrality refers to the importance of a node in the network. There are three measures which have been most widely applied, namely degree centrality, closeness centrality and betweenness centrality.

Degree Centrality (DC) of a node is defined as the number of ties the node has. In mathematical term, DC is defined as,

$$DC_i = \sum_j m_{i,j}, \tag{2.2}$$

where $m_{i,j} = 1$ if there is a link between node i and node j.

Closeness Centrality (CC) is the summation of distances from all other nodes to the target one (Freeman *et al.,* 1979), which can be written as,

$$CC_i = \sum_j d_{ij}^{num}, \tag{2.3}$$

where d_{ij}^{num} is the number of links in a shortest path from node i to node j. A larger value of CC indicates the node is less central while a small value means it is more central.

Betweenness Centrality (BC) refers to the frequency of a target node to be the intermediate in the shortest path for all other node pairs. In other words, it is the number of shortest paths passing through a target node, which reads,

$$BC_i = \sum_{j,k} \frac{g_{ijk}}{g_{jk}}, \tag{2.4}$$

where g_{ijk} is the number of shortest paths from node j to node k passing through node i and g_{jk} is the number of shortest paths from node j to node k $(j, k \neq i)$.

Many research studied the effect of centrality of nodes in the communication structure on problem solving. For example, an experiment from MIT in 1950s studied the communication in a group of subjects under four structural forms: chain, Y shape, star and circle, and found that subjects' satisfaction and nomination leadership is related to the centrality.

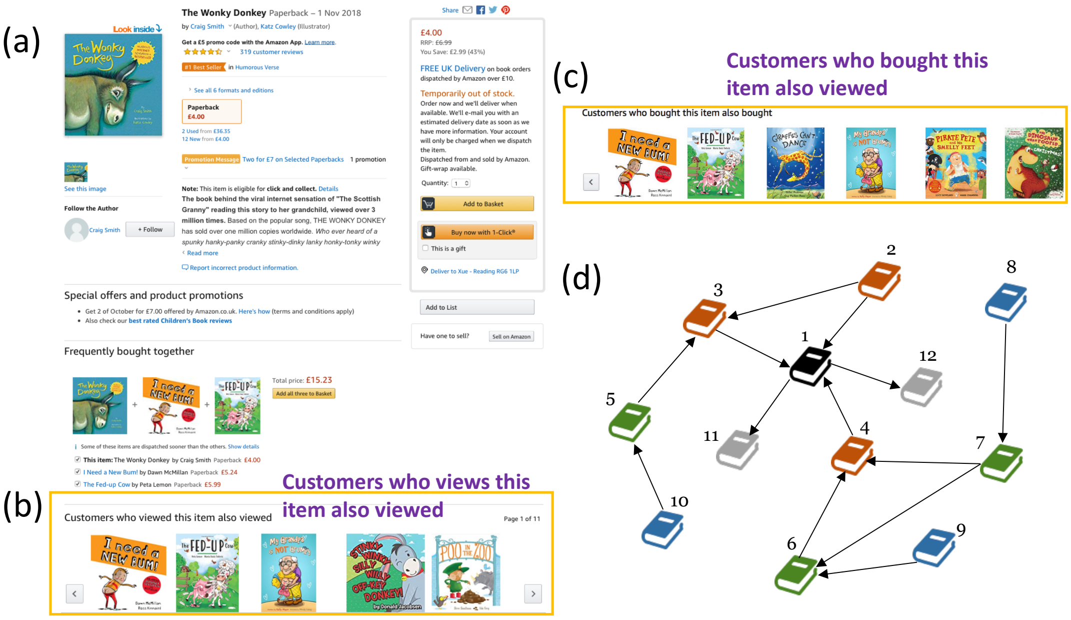
2.4.2 Friendship Network

Friendship network is a big component in the field of social network analysis, in which nodes represents persons and links represent relationships. Since friends may tend to have similar preference and frequent interactions, their behaviours may be influenced by each other. It is noted that the section 2.1.3 "Social Influence" focuses on the influence of social network that may come from the crowds. But this section reviews the impact of friends, that is also referred as peer influence.

In the field of social psychology, a large amount of research discussed the impact of friendship network on human behaviour. One of the most talk-about, the health behaviour such as smoking, alcohol use (Huang *et al*., 2014; Fisher and Bauman, 1988), drug use (Bauman and Ennett, 1994), mental health problem (Oliver *et al.,* 2005), is strongly influenced by their friends. Normally, most of research subjects in this field are youths or teenagers, because they are more susceptible, and friends play significant roles in their roads to growth. Valente *et al.* (2009) study the relationship between adolescent friendships and obesity and find that overweight youth were twice as likely to have overweight friends. Mouttapa *et al.* (2004) examine youth's bullying behaviour and find that friends' participation in aggressive behaviours is possibility associated with being a bully.

In recent decades, due to the success of social networking websites, such as Facebook and MySpace, people have their "second life" on the Web, which is a virtual environment to meet friends and share information. Thus, a considerable body of research is moving to study how online friendship network impact each other. Lewis *et al.* (2012) find that people within social network may share certain taste in music and movies based on the activities of Facebook. Ravi and Akhmed (2015) suggest peer influence among friends increases the odd of buying products. Aral and Nicolaides (2017) find that exercise is socially contagious among friends and the contagiousness varies with the relative activity and gender.

2.4.3 Product Recommendation Network



(a)
(b)
(c)
(d)
Customers who bought this item also viewed
Customers who views this item also viewed
1
2
3
4
5
6
7
8
9
10
11
12

The Wonky Donkey Paperback – 1 Nov 2018
by Craig Smith (Author), Katz Cowley (Illustrator)
Get a £5 promo code with the Amazon App. Learn more.
319 customer reviews
#1 Best Seller in Humorous Verse
See all 6 formats and editions
Paperback
£4.00
2 Used from £36.35
12 New from £4.00
Promotion Message Two for £7 on Selected Paperbacks 1 promotion

Note: This item is eligible for click and collect. Details
The book behind the viral internet sensation of "The Scottish Granny" reading this story to her grandchild, viewed over 3 million times. Based on the popular song, THE WONKY DONKEY has sold over one million copies worldwide. *Who ever heard of a spunky hanky-panky cranky stinky-dinky lanky honky-tonky winky*
Read more
Report incorrect product information.

Look inside
See this image
Follow the Author
Craig Smith + Follow

Share
£4.00
RRP: ~~£6.99~~
You Save: £2.99 (43%)
FREE UK Delivery on book orders dispatched by Amazon over £10.
Temporarily out of stock.
Order now and we'll deliver when available. We'll e-mail you with an estimated delivery date as soon as we have more information. Your account will only be charged when we dispatch the item.
Dispatched from and sold by Amazon. Gift-wrap available.
Quantity: 1
Add to Basket
Buy now with 1-Click®
□ This is a gift
Deliver to Xue - Reading RG6 1LP
Add to List
Have one to sell? Sell on Amazon

Special offers and product promotions
- Get 2 of October for £7.00 offered by Amazon.co.uk. Here's how (terms and conditions apply)
- Also check our best rated Children's Book reviews

Frequently bought together
Total price: £15.23
Add all three to Basket
Some of these items are dispatched sooner than the others. Show details
☑ This item: The Wonky Donkey by Craig Smith Paperback £4.00
☑ I Need a New Bum! by Dawn McMillan Paperback £5.24
☑ The Fed-up Cow by Peta Lemon Paperback £5.99

Customers who viewed this item also viewed
Page 1 of 11

Customers who bought this item also bought

Figure 2.3|(a). A screenshot of a book named "The Wonky Donkey" on Amazon. In the below, two recommendation lists display five and six books respectively, following "consumers who viewed this item also viewed" (b)and "consumers who bought this item also bought." (c). Regarding the books as nodes and recommendations as links, the products on the website can be transformed to a PRN (d).

In many ecommerce websites, there would be a list of recommendations consisting of objects with hyperlinks that the website considers to be similar to the current one. Accordingly, the massive objects are connected by the links composing a product recommendation network (PRN), in which consumers can browse many similar products in dedicated webpages. Figure 2.3 (a) shows a screenshot of a book from Amazon, on which we can see several other books are recommended below with a header of "consumers who bought this item also bought" and "consumers who viewed this item also viewed." These books can thereby be considered as a directed network, as shown in Figure 2.3 (d), in which the nodes are books and links are recommendation hyperlinks.

The PRNs allow users to browse a wide range of products and have been shown influencing the consumption pattern (Senecal and Nantel, 2004). Oestreicher-Singer and Sundararajan (2012a) study the network of books on Amazon and quantified the incremental correlation between book sales and the visibility of book network. They also found that the PageRank centrality, which is a measure of a node's position in a network, and some other quantities such as the in-degree, are closely associated with the books' demand measured by the sale rank in Amazon (Oestreicher-Singer and Sundararajan, 2012b). Leem and Chun (2014) further examine the other centrality measures, including degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, and confirmed that book demand is vastly influenced by its positions in PRNs. Assuming the revenue of a product as the summation of its intrinsic value and incoming value, Oestreicher-Singer *et al.* (2013) try to estimate the network value of products, which consists the value generated by itself, and the value it contributes to other products. Lin *et al.* (2018) explore the impact of product network attributes in terms of network diversity and stability on product demand. Hou *et al.* (2017) explore the users' surfing behaviour on Amazon and found that the recommendation network of Amazon tends to rapidly navigate users to very popular books, leading to the monopoly of traffic by the blockbusters. Goldenberg *et al.* (2012) integrates the social and product network as a dual-network and found it to be facilitating the process of content exploration.

2.5 Summary

This thesis aims to explore the impact of eWOM (also refers to online product reviews in the thesis) on consumer behaviours in the context of social network and PRN and examine the helpfulness of online reviews. Thus, this chapter firstly reviews the prior relevant literature regarding the theory of information diffusion that help us to understand the information spreading upon the interactions in human activities. Secondly, the theory of consumer information search including influential factors for searching and the reasons of online information search is reviewed, which correspond to the aims of the thesis that how consumer behaviour would be influenced during a series of actions like searching, adopting, purchasing and posting information. Thereafter, the chapter reviews eWOM studies considering antecedent and consequence of two parties, namely eWOM senders and receives, the dimensions for analysing online product reviews and influential factors of helpfulness. In the last, the chapter reviews social network analysis, which introduces some measurement of network and two main types of networks which will be studied in this thesis, i.e. friendship network and product recommendation network.

Chapter 3. Research Methodology

This chapter discusses the research methodology and selection of appropriate methodology adopted for this PhD research. Firstly, the quantitative research methodologies the have some relations to the thesis such as econometric modelling, simulation modelling and algorithmic modelling are introduced. Following this, a new research paradigm in social science, namely computational social science is presented. Finally, the chapter introduces the research methodology of this thesis, which can be characterised as multi-disciplinary by combing a series of approaches.

3.1 Quantitative Research

Quantitative methodology is based on the measurement of quantity or amount. It is applicable to phenomenon that can be expressed in terms of quantity.

The quantitative approaches involve the generation of data in quantitative form which can be subjected to rigorous quantitative analysis in a formal and rigid fashion. This can be classified into *inferential*, *experimental* and *simulation* approaches (Kothari, 2004). The inferential approach to research is to form a data set from which to infer characteristics or relationships of population, that normally refers to survey research. Experimental approach is characterised by much greater control over the research experiments and in this case some variables are manipulated to observe the effect to other variables. Simulation means the construction of an artificial environment in which relevant information and data can be generated. This section mainly describes three types of methodology in quantitative statistical modelling which will be applied in the thesis: *econometric modelling* (also refers to regression), *algorithmic modelling* and *Simulation Modelling.*

Both *econometric modelling* and *algorithmic modelling* regard data as being generated by a black box in which a vectors of input variables (independent variables) **X** are in one side and the other side is response variable (dependent variable) **Y.** In the box, the nature functions associate the input and response variables, which can be described as in Figure 3.1.

Figure 3.1|The culture of statistical modelling. A vector of input variables X goes in one side, and on the other side the response variable Y comes out. The inside in the grey box is nature function associating X and Y.

Econometric Modelling

In the data modelling, the analysis starts with assuming a stochastic data model, such as a function of Y=f (X, random noise, parameters), in the blackbox. The value of the parameters is estimated from the data and the model is then used to predict. The regression attempts to model the relationship between the dependent variable Y and the independent variables X by fitting an equation that may be linear, polynomial or logistic *etc.*

Figure 3.2|The culture of econometric modelling. The inside box that connects X and Y is filled with regression models.

Algorithmic Modelling

The analysis methods in the algorithmic modelling connecting X and Y is unknown. The approach is to find an algorithm or function f(X) to associate X with Y. The models in the blackbox are normally complex. A vector of **X** that goes in and a response **Y** comes out. The theory in this field focuses on the properties of algorithms. Some algorithmic approaches such as decision tree, support vector machine, neural networks have been developed.

Figure 3.3|The culture of algorithmic modelling. The algorithmic functions in the inside box are unknown, which maybe decision tree, neural net or anything else.

Simulation Modelling

Simulation is a process to design and conduct experiments for the purpose of understanding system behaviour or evaluating various strategies for the operation of the system. The term of "simulation" in business and social science refers to "the operation of a numerical model that represents the structure of a dynamic process. Given the values of initial conditions, parameters and exogenous variables, a simulation is run to represent the behaviour of the process over time (Meier *et al.,* 1969). The operation of simulation is represented as a chronological sequence of events in a discrete-event simulation. Each event occurs at an instance in time and marks a change of state in the system.

A simulation model that is commonly used in statistical physics is *Monte Carlo Simulation,* which is typically applied by latent structural equation to ascertain the robustness of statistical estimators (Chin and Newsted, 1999; Chou *et al.,* 1991). Monte Carlo simulation is a procedure of generating artificial data, based on a specific statistical model that is defined in terms of a stochastic generating mechanism (Noreen, 1989). In the Monte Carlo simulations, the implied covariance matrix of the observed variables is computed for given values on the parameters in the model and then the data are generated.

3.2 Research Paradigm in Social Science

Traditionally, the field of social science relies on the data collections such as surveys, questionnaires or interviews. With the rapid advances in data collection technologies, and business interactions generating vast amount of information, traditional social science inquiry is being shifted to computational social science (Chang *et al.,* 2014), as the availability of the large amount of data makes the social science more reliable.

The era of big data today has brought us new opportunities for researchers to achieve changes and transformations in how we study social phenomena. The new data collection technologies, new data mining methods, and new sources of contexts including blogging, online shopping and social advertising, allow business analysts and researchers to achieve frequent, controlled and meaningful observations of the real world. In this section, we mainly introduce the rising discipline: computational social science (CSS).

CSS is defined as an interdisciplinary approach to social science research through an information processing and complex adaptive systems paradigm, using computation as the key enabling scientific methodology (Cioffi-Revilla, 2014). CSS involves many exciting scientific research fields at the intersection of all social science disciplines, applied computer science and other related disciplines.

In the field of CSS, we rely on less construct artificial setting in the lab to control realism by mimicking the real world in a synthesised setting to collect data. Instead, we can capture data that represents the fundamental elements of human actions and interactions such as the tweets, shopping records, user opinions in the digital world. These objective factual data can tell us more about participants' characteristics, which is difficult to gain from the traditional data collection ways such as surveys. For example, subjects' preferences can be easily analysed from the digital traces of their behaviours rather than directly asking them questions.

CSS is, as an interdisciplinary field, composed of clusters of concepts, principles, theories and research methods. We mainly introduce four areas in CSS in this section: Automated Social Information Extraction, Social Networks, Social Complexity and Social Simulation Modelling.

3.2.1 Automated Social Information Extraction

Automated social information extraction refers to computational ideas and methodologies pertaining to the creation of scientifically useful social information based on raw data sources—all of which used to be done manually (Cioffi-Revilla, 2014), which also can be named computational content analysis or social data analysis.

Social information extraction through automated computational procedures is widely used in CSS. For example, the information concerning the political orientation of leader or other governmental actors can be extracted based on computational content analysis of speeches, testimony before legislative committees, or other public records (Tumasjan *et al.*, 2010; Zhou, 2009). All above studies illustrate that automated information extraction is regarded as a foundational methodology in CSS, that can be used for developing models and theories in other areas of CSS.

3.2.2 Social Networks

Social network analysis is another area of CSS, which has been introduced in section 2.4, we will not repeat it too much in this section.

There are many reasons for the success of the field. Firstly, social scientists have already developed a powerful set of concepts, statistical tools, and mathematical models and procedures, including formal theories of network analysis, by using computers being methodological toolkits, which enabled them to exploit computational approaches. Secondly, computational tools, especially the most recent generation of computer hardware and software systems, now enable efficient processing of high-dimensionality data and large matrices necessary for understanding complex social networks.

3.2.3 Social Simulation Modelling

Social simulation modelling is another area of CSS, which can be characterised as foundational, interdisciplinary, diverse based on many different methodologies in modelling and simulation disciplines. The simulation modelling begins in social science many decades ago during the earliest days of digital computing. There are several different kinds of social simulation modelling frameworks such as queuing models, equation-based models, agent-based models, and evolutionary computational models. Regardless of the specific type, all social simulation models share a set of common characteristics. A simulation model is always designed and built around a set of research questions, which may concern basic science or applied policy analysis, or sometimes both. Another is that they are developed through a set of developmental stages, not as a single methodological activity, especially in the case of complex modelling projects or those involving teams of investigators.

Each social simulation model can, at least in principle, include ideas and components from other areas of CSS, such as results from automated information extraction, social network analysis. Conversely, social simulation models can provide significant input and improvements pertinent to research in these other areas.

3.3 Research Methodology of the Thesis

The research in this thesis is, in nature, multi-disciplinary and being at the interface among information management, social science, mathematics and statistics. As a consequence, our methodology is a combination of a series of approaches. Figure 3.4 shows the process of the thesis.

The study mainly adopts quantitative methodology addressing the impact of eWOM, which follows the following procedures: 1) find the research questions and limitations based on the review literature; 2) develop models or algorithms to address the potential relationships between explanatory variables and response variable to solve the research questions; 3) gain the results based on the empirical data, and 4) evaluate the results and performance of models or algorithm. According to the literature and observation of consumer behaviours, we propose the hypothesis and define the research questions. Principles and techniques of the quantitative methodology are then applied to measure the impact and value of eWOM and collect data. The evaluation largely depends on the partition of empirical data. For example, to answer the third research question, we divide the data of review helpfulness into two parts, namely the training data and testing data. Training data will be applied to examine the relations between measured factors and review helpfulness, and the outputs of the models will be compared to the testing data to assess their performances.

The adopted approaches are basically empirical ones, including regression modelling such as Probit model, logistics model and OLS model, web crawling, data analytics, Monte Carlo simulation, and algorithm implementation via programming etc. For data collection, we crawl the data from Amazon book recommendation, which will be introduced in the next section. For all analysis in this study, we programme based on Python environment. To more clearly present the analytical results, many figures of plots in chapter 6 are presented in this thesis, which are produced via Matlab and Python (with matplotlib package). We clean and sample the raw data of Yelp with Python. The regression analysis in chapter 5 is conducted by Stata because of the vast volume of data and the calculation in Stata is the quickest. The rest of regression analysis is done by R language, but we also run the models by Stata and got the same results.

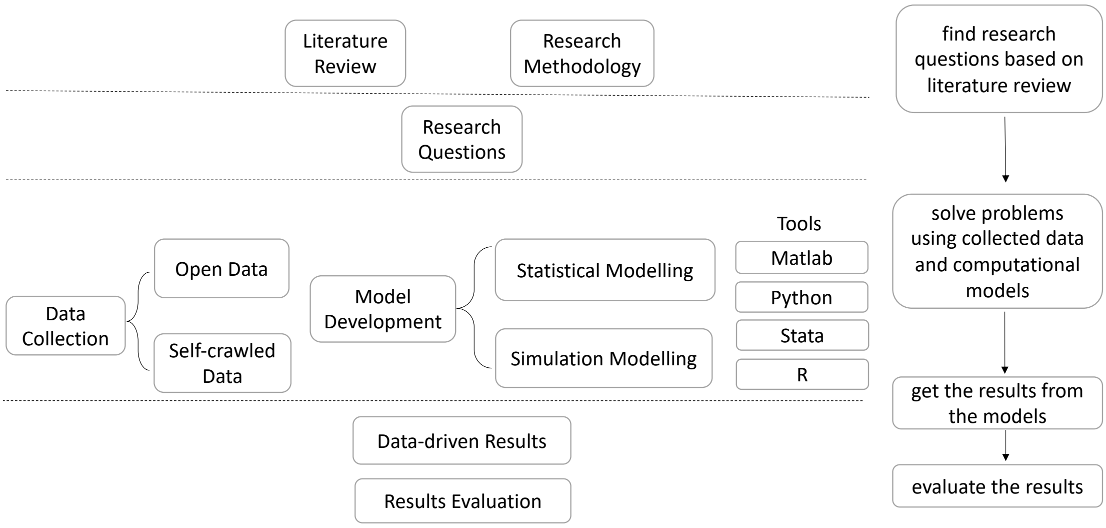


Figure 3.4| Research design of the PhD project.

3.3.1 Regression Analysis

One of the most common adopted methods in the studies of the eWOM and information system is the regression analysis (Bapna & Umyarov, 2015; Chen *et al.* 2011). The present thesis also largely relies on the regression analysis of the large-scale user-product interaction data.

Regression analysis is one of the most widely used statistical tools and has extensive applications in many research fields, because it indicates the significant relationships between variables and the strength of impact of multiple independent variables on a dependent variable. Driven by the shape of regression line, there are linear regression and nonlinear regression. The types of dependent variables (continuous, categorical or count data) also lead to different model selections such as Ordinal Logistic regression (OLS) or Poisson regression.

In the field of data science, many machine learning methods have been developed and suggested to have more prediction accuracy. However this thesis is not an engineering-based research, which may take the results (normally accuracy) as the priority. Being an explanatory research, this thesis aims to find patterns or factors that can explain and possibly predict the behaviours of consumers. Hence, achieving the highest prediction accuracy is not the priority in this thesis. The most important objective of this research is to uncover what factors have what (positive or negative) impact on the influence of eWOM on consumer behaviour. Being the most classic tool in data science, the regression is thus more suitable for such objectives.

3.3.2 Monte Carlo Simulation

The other technique that the thesis is using is Monte Carlo simulation, which also refers to probability simulation. It is a statistical probabilistic technique to design experiment to study the nondeterministic probability distribution of the factors and the responses. In a Monte Carlo simulation, a random value is selected for each of the tasks, based on the range of estimated. Then the model is calculated based on the random value. It is widely applied in the field of complex system like DNS sequence simulation (Rambaut & Grass,1997), transportation (Jacoboni & Reggiani, 1983).

Regression and simulation can be complementary methods when addressing the same research question. Regression helps to understand to what extent the independent variables can explain the dependent variable, but normally cannot infer the casual relations between independent variables and dependent variable. On the other hand, by assuming a certain casual structure, a simulation can validate whether such casual structure could lead to the empirical observation/phenomenon. Thus, simulation can uncover the possible mechanism or driver of systems. Seldom study of consumer behaviour think over using them both. We adopt the two techniques in chapter 5 and chapter 6 to examine whether a user's friends or crowds have different influence over his behaviour. The regression analysis in chapter 5 reveals the significant impact of friend reviews on consumer behaviour. The simulation result in chapter 6 convinces the finding that the intensity of friend influence is the possible underlying driver over consumer behaviour.

3.4 Summary

The chapter presents the research methodology underpinning the study. The chapter discusses the meaning of quantitative methodology. Furthermore, we introduce the traditional methodology in social science and the shift to a new discipline of computational social science from three areas: automated social information extraction, social networks and social simulation modelling. Lastly, the chapter introduces the methodology and approaches adopted in this thesis, mainly introduce the regression analysis and Monte Carlo simulation.

Chapter 4. Data Collection

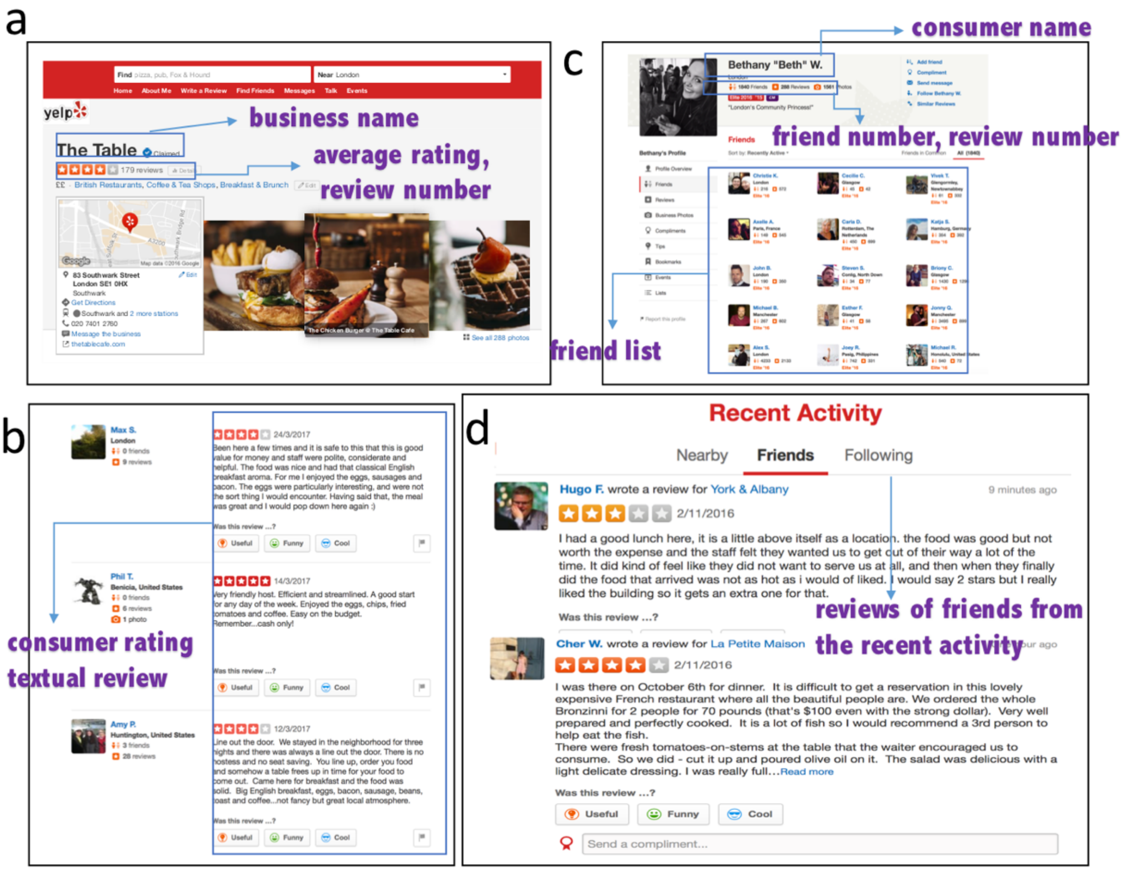
As we introduced in the last section, the thesis aims to explore and uncover the social phenomenon from vast volume of data by adopting computational models (Carley, 2002). The data used in this research is not collected by surveys or questionnaires, but from that generated by consumers and suppliers without the experiment settings, including user profiles, consumption records, objects information and consumer reviews toward objects from e-commerce websites.

As we suggested before, the crucial data are the consumer reviews relating to the objects, which also includes network connections such as friendship connections between consumers and recommendation connections between objects. The objects can be books, music, movies and restaurants depending on the services provided by e-commerce websites. In this chapter, we introduce the data sets which will be used in the thesis. One is an open-source data from Yelp that will be introduced in section 4.1, and the other one is collected by a web crawler from Amazon that will be introduced in section 4.2.

4.1. Open Data from Yelp

Thanks to the recent studies on big data analytics, network science and ecommerce, many datasets on e-commerce websites regarding the interactions between consumers and businesses have been published by the companies. We retrieve a dataset from one of the widely-investigated online systems, Yelp.

The Yelp is a business review website where users can check on countless businesses such as restaurants, cafes, theatres, or even clinics and hospitals. Besides the basic business information such as the addresses, opening hours, parking facilities, users can especially check on the ratings and reviews of prior consumers on a particular business. After gathering the opinions of others, a user may make his own decision accordingly, such as, whether or not to go to the business, or which one to go to. As a consequence, the opinions of others are very likely to influence the decision process of users. Yelp becomes an ideal scenario for the studies of social influence on user consumption, selection behaviour, and user preference. Especially, one of the most appealing features of Yelp is its social networking service. A user can establish friendships with other users either to be his real-word friends or those who write reviews s/he finds trustworthy in the system. On the homepage of the Yelp, there displays the list of your friends' recent activities (reviews) besides the list of non-friends' activities., as shown in Figure 4.1. Therefore, the friends' reviews are also influential factors for a user to make decision. Considering all the features and settings of the Yelp website, we believe it is very suitable for this study to explore the impact of eWOM on consumer behaviours in the context of social network.



24/3/2017
Been here a few times and it is safe to this that this is good value for money and staff were polite, considerate and helpful. The food was nice and had that classical English breakfast aroma. For me I enjoyed the eggs, sausages and bacon. The eggs were particularly interesting, and were not the sort thing I would encounter. Having said that, the meal was great and I would pop down here again :)

Was this review ...?

Useful Funny Cool

14/3/2017
Very friendly host. Efficient and streamlined. A good start for any day of the week. Enjoyed the eggs, chips, fried tomatoes and coffee. Easy on the budget. Remember...cash only!

Was this review ...?

Useful Funny Cool

12/3/2017
Line out the door. We stayed in the neighborhood for three nights and there was always a line out the door. There is no hostess and no seat saving. You line up, order you food and somehow a table frees up in time for your food to come out. Came here for breakfast and the food was solid. Big English breakfast, eggs, bacon, sausage, beans, toast and coffee...not fancy but great local atmosphere.

Was this review ...?

Useful Funny Cool

Recent Activity

Nearby Friends Following

Hugo F. wrote a review for York & Albany 9 minutes ago

2/11/2016

I had a good lunch here, it is a little above itself as a location. the food was good but not worth the expense and the staff felt they wanted us to get out of their way a lot of the time. It did kind of feel like they did not want to serve us at all, and then when they finally did the food that arrived was not as hot as i would of liked. I would say 2 stars but I really liked the building so it gets an extra one for that.

Was this review ...?

Cher W. wrote a review for La Petite Maison

2/11/2016

I was there on October 6th for dinner. It is difficult to get a reservation in this lovely expensive French restaurant where all the beautiful people are. We ordered the whole Bronzinni for 2 people for 70 pounds (that's $100 even with the strong dollar). Very well prepared and perfectly cooked. It is a lot of fish so I would recommend a 3rd person to help eat the fish.
There were fresh tomatoes-on-stems at the table that the waiter encouraged us to consume. So we did - cut it up and poured olive oil on it. The salad was delicious with a light delicate dressing. I was really full...Read more

Was this review ...?

Useful Funny Cool

Send a compliment...

Figure 4.1 | Introduction to Yelp system. (a) Homepage of a business in Yelp. (b) Reviews of the business shown in Fig. 4.1 (a) and (c) A user homepage in Yelp. (d) Recent friends' reviews.

Yelp, being enthusiastic on scientific research, has published their data and been holding big data challenges for many years. The data set used in this study was downloaded from

Yelp challenge website[1]. While they constantly update the published data set, the data in this study was accessed in January 2016, which consists of 1,569,264 comments on 61,184 businesses posted by 366,715 users. Although data with reasonable details is published, we only use the wiring patterns and the timestamps of the system, i.e. which user befriended with which users, and the posting behaviour and posting time. Therefore, the information we considered from the published data can be perfectly described by the user-business bipartite network with underlying social structure. The timing when each review is conducted and each user is registered are known, but the establishment timing of friendship is unknown in the data. Therefore, for the user-business connections, the timestamps are the exact time provided by the data, while for the user-user connections, the timestamps are estimated as the later date of the two connected users' registrations. In other words, if two users have established the friend connection in the data, we consider the connection was established when both of the users had registered to the system.

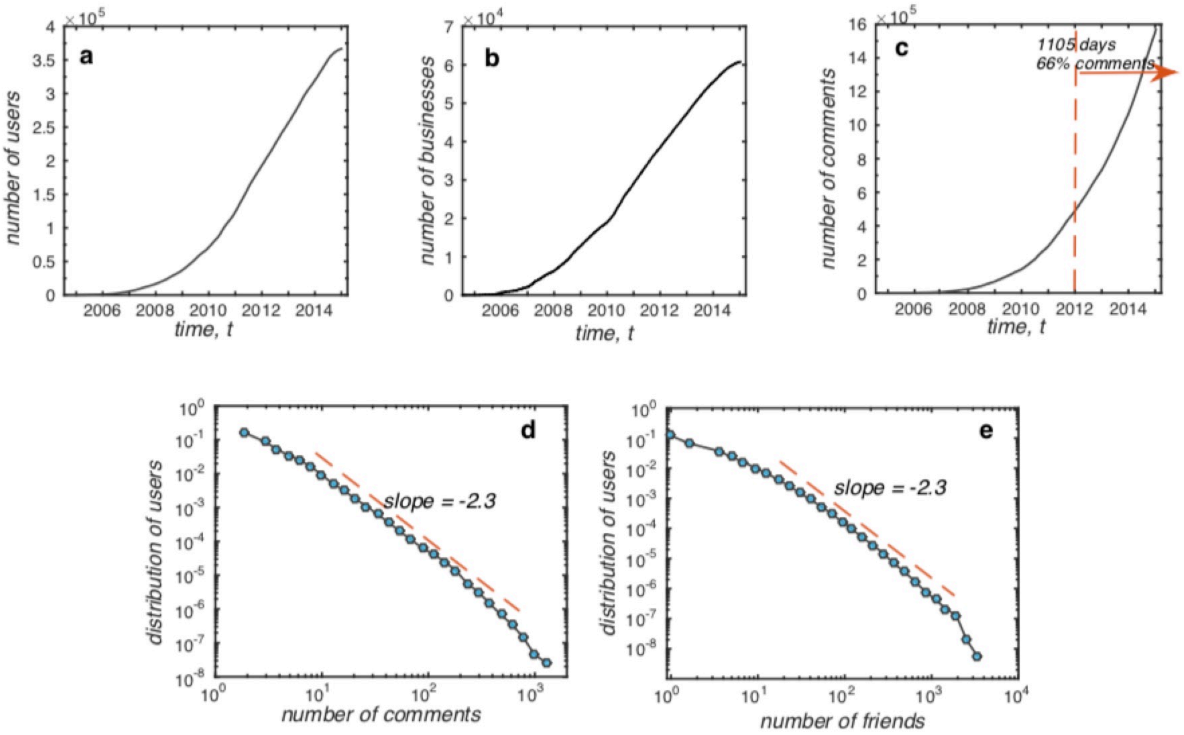


Figure 4.2 | Statistics of Yelp data.

[1] https://www.yelp.co.uk/dataset challenge

Figure 4.2 reports the detailed information and descriptive statistics of the data set. Figure 4.2 (a, b, c) show the growth of user population, number of businesses and number of comments respectively in the data over 11 years. Most of the data distributes in recent years. Note that, while the date of the user registration and user-business connection establishment are given by the data, the time that each business registered to the system is estimated as the first comment date. Figure 4.2 (d, e) show the distributions of users in terms of number of comments and number of friends respectively. The red dashed line in each of the subplots c and d has a slope of -2.3 in the log-log plot.

4.2 Data Collection from Amazon

To study the recommendation networks, we collect data from the "Customers who bought this item also bought" list in the Amazon, which is a retail website where users can buy products and leave comments. In the system, each product has a dedicated webpage displaying its basic information, user comments, and most importantly a list of similar other products as recommendations. In such recommendation lists, there are normally 100 similar products recommended, but displayed in pages. Hyperlinks are available for users to click on and surf to the corresponding recommended products. To collect the recommendation network is basically to collect such hyperlinks.

The empirical recommendation networks of Amazon have already been widely used in previous studies. However, most of these studies applied a depth-first searching strategy (Oestreicher-Singer and Sundararajan 2012a, 2012b; Carmi *et al.*, 2017). Therefore, the collected network does not have a unified out-degree, i.e. products have different number of recommended others. In Amazon system, and also most other similar systems, an apparent feature of recommendation networks is that the recommendation list length is fixed. Accordingly, previous strategies did not capture such feature. In the present thesis, we therefore adopt a width-first searching strategy to collect the hyperlinks so that every product would have the same number of recommendations.

The Amazon data was collected in two steps, namely the recommendation network collection and the user-object bipartite network collection respectively, over the January of

2016.

4.2.1 Recommendation network collection

In the Amazon system, each book has a unique ID, and the webpage of the book is composed as http://www.amazon.com/dp/ID, where the ID should be replaced by a real ID. Consequently, to collect the book network is actually to collect the corresponding webpages and the recommendation hyperlinks connecting them.

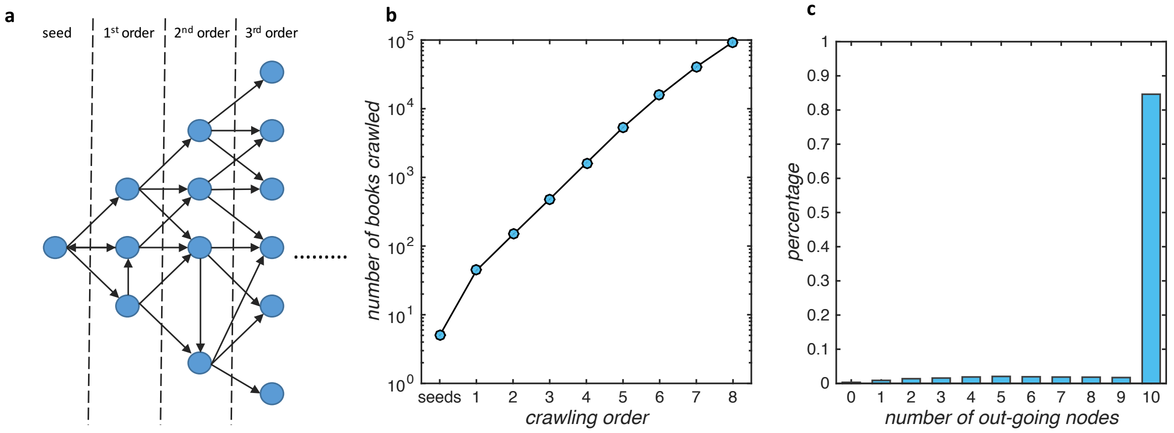


Figure 4.3 | Illustration for the collection of Amazon book recommendation network.

We firstly selected 5 books as the seeds of the crawling from the Amazon's bestseller list (www.amazon.com/gp/bestsellers/books). Note that, the list may change from time to time and in our collection, the seeds were collected on 1st January 2016. For each of the seed books, we collect books from its recommendation list known as “Customers who bought this item also bought” list. The books in the seeds' recommendation list are regarded as the 1st-order books (out-going nodes of the seeds) as shown in Figure 4.3a. While normally 100 similar objects are offered in each object’s full recommendation list, there are generally 5 to 10 recommendations in the first page depending on the window size of the web browser. Assuming the recommendations displayed in the first page would get most attentions, we collected 10 books ranking at the top of the list as the current one’s out-going nodes. As the crawling goes on and on, we then have the 2nd-order books, 3rd-order books and so on. The crawling continued for 8 steps. And for the 8th-order books, we collect the first 10 recommendations that have already been included in previous steps out from the list as their out-going books, so that there will be no 9th-order books. Figure 4.3b reports the number of books that are newly crawled at each order and in total of 157,856 books was collected. All of the collected books are reachable for the seeds within 8 steps. According to the crawling strategy, the out-degree of every node should be 10. However, some of the books may appear in others' recommendation list but somehow have no or just a few out-going books in the Amazon system. Additionally, those 8th-order books may have less than 10 out-going books that have been collected in the early steps. In summary, as shown in Figure 4.3c, 84.62% of all the crawled books have exactly 10 out-going books.

In this book recommendation network, the nodes are the Amazon webpages of the books, and the links are actual recommendation hyperlinks established by Amazon.

4.2.2 eWOM of Books

In addition to the book recommendation network, we also collect all the reviews for each of the collected 157,856 books. Note that, in Amazon system, different versions of a same book such as Kindle edition, hardcover edition and paperback edition etc., share the same review webpage. Considering that version selection is also a reflection of the users' interest and the different versions of the same book have different recommendation lists, we only collected the comments devoted to the very specific version of the crawled book. In total of 4,520,102 reviews are collected (after cleaning the reviews to the other versions) which are posted by 2,540,369 consumers. Furthermore, the Amazon system marks the comments that posted by users who actually have bought this book as "verified purchase". Accordingly, every review between a consumer and a book represents a purchase behaviour, and our data is basically a sample of the full sale record of these books.

4.3 Summary

The chapter presents the adopted two data sets and the ways of collecting them, explain the reasons why we choose the data sets and how they fit the scenario of the research. We firstly introduce the open data from the big data challenge of Yelp. The chapter provides the detailed introduction of Yelp including the homepages of both businesses and reviewers as well as the social structure between reviewers. Also, the statistics of Yelp data is presented. Secondly, the chapter introduces the process of self-crawling data from Amazon from the seed books to the neighbours as well as collecting the eWOM information of these books.

Chapter 5. Exploring Review Impact on Consumer Posting Behaviour

Online reviews, as we discussed above, have significant impact on consumer behaviour (Chevalier and Mayzlin, 2006; Schlosser, 2005), and thus are also appreciated by companies as valuable marketing resources (Jung *et al.,* 2013; Chen *et al.,* 2011). Consumers can access online reviews coming from both the crowd (non-friends) and their friends because of the social networking services.

Both friends and the crowd have been found to be influential for consumer engagement in eWOM in the literature. The crowd has a much larger population than that of a consumer's friend circle, which implies that the reviews and opinions would be of much greater diversity and richness. Therefore, the reviews from the crowd may provide much more information for consumers to make purchase decisions and further engage in post-purchase discussions. However, friends normally have closer relationships and similar preferences with the target consumer, resulting in trust and frequent interactions (McPherson *et al.,* 2001). Due to the frequent interactions, eWOM from friends has strong influence over one's engagement in posting (Centola, 2010; Crandall *et al*., 2008; Dellarocas and Narayan, 2006; Aral and Walker, 2011). Actually, such influence on one's behaviour has long been reported in social science (Brown and Reingen, 1987; Steffes and Burgee, 2009; Kawachi and Berkman, 2001; Wang and Chang, 2013), and is normally recognised as a social contagion phenomenon (Aral and Walker, 2011; Aral *et al.,* 2009).

As a consequence, a question rises that, (RQ1 of the thesis) how friend's and crowd's reviews differently impact consumer engagement on posting behaviour? This chapter adopts a social network perspective to examine how friend reviews and crowd reviews differently influence consumers' subsequent posting behaviour in an online context. Section 5.1 introduces the hypothesis development. Section 5.2 presents the way of data preparation based on the data set of Yelp and model specification. The results of review impact on consumer behaviour is presented in section 5.3.

5.1 Hypothesis Development

Inspired by the literature (Zhu and Zhang, 2010; Moe and Schweidel, 2012), we model the consumption process as shown in Figure 5.1. Before making a consumption decision, users first gather information about the product as well as previous reviews posted by other consumers. The target consumer's characteristics, product information as well as the review information posted by either friends or the crowd will jointly influence the decision on whether to buy this product. After purchasing items, consumers will decide either to be a poster to share their experience, such as via a numerical rating, text review, and even some photos, or to be a "lurker" who only read reviews but does not post them.

We aim to identify the influential factors in the pre- consumption stage for a consumer to post a review, and to detect if the impact of friends' reviews and the crowd's reviews are different for consumer posting behaviour. To do so, a number of hypotheses are developed. as shown in Figure 5.2.

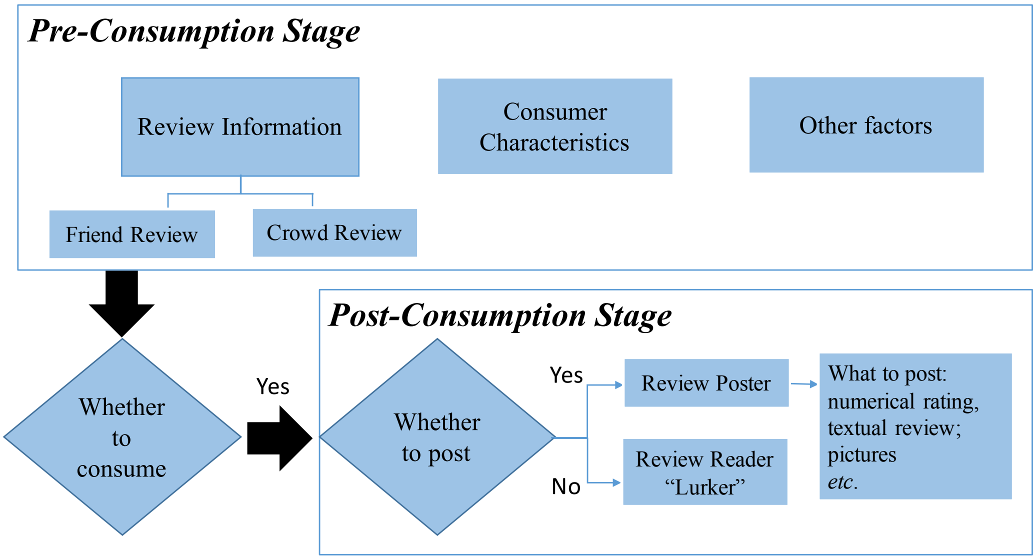


Figure 5.1| Process of online consumption and post-consumption evaluation.

5.1.1 Review Volume

Consumers tend to select products with more reviews. First, popular products get more attention because more consumers are aware of them and a larger volume makes the reviews more objective and trustworthy. Chen *et al*. (2004) show that the average rating converges to the true quality with the increase of review number. Therefore, reviews of popular products can more accurately reflect the true quality. Furthermore, consumers are more likely to access the information of popular products because they are exposed to these reviews more frequently. A favourable feeling can be created after sufficient exposure, which can be interpreted as an exposure effect (Bornstein, 1989; Zajonc, 1980). As popular products are reviewed more frequently and consumers are exposed to them repeatedly, the exposure effect makes consumers more likely to choose them and further engage in the discussions about them. Thus, we offer the following hypotheses:

- **Hypothesis 1a** (The Crowd's Review Volume Influence and Posting Engagement Hypothesis): The review volume for a product from the crowd positively influences the likelihood of subsequent consumer engagement in posting behaviour.
- **Hypothesis 1b** (The Friends' Review Volume Influence and Posting Engagement Hypothesis): The review volume for a product from a target consumer's friends positively influences the likelihood of engagement in posting behaviour.

Friends usually play an important role in several aspects of consumer purchase selection and posting behaviours (Lee *et al.,* 2015). Friends normally have similar tastes and preferences for product selection (McPherson *et al.,* 2001), as well as more frequent interactions. This may facilitate consumers in selecting products that are popular among friends and join them to engage in the post-consumption eWOM communications. Review information posted by friends is perceived as more credible, leading to significant influence over individual decision-making (Granovetter, 1973). A tendency toward transitivity is exhibited in a social network of friends according to Granovetter (1973): if person A is a friend with person B and person C, there is a high probability that B and C become friends too. Clark and Loheac (2007) suggest this tendency existed in brand preferences for consumers who belong to the same friend circle. In addition, our recent study (Pan *et al.,* 2017) developed a model describing whether consumers follow their friends or the crowd to make selections, and found that 75% of selection behaviour is driven by friends' opinions. Thus, we suggest:

- **Hypothesis 1c** (The Greater Effect of Review Volume for Friends Than the Crowd for Posting Hypothesis): The review volume of a target consumer's friends is more influential on her/his posting behaviour than the crowd's.

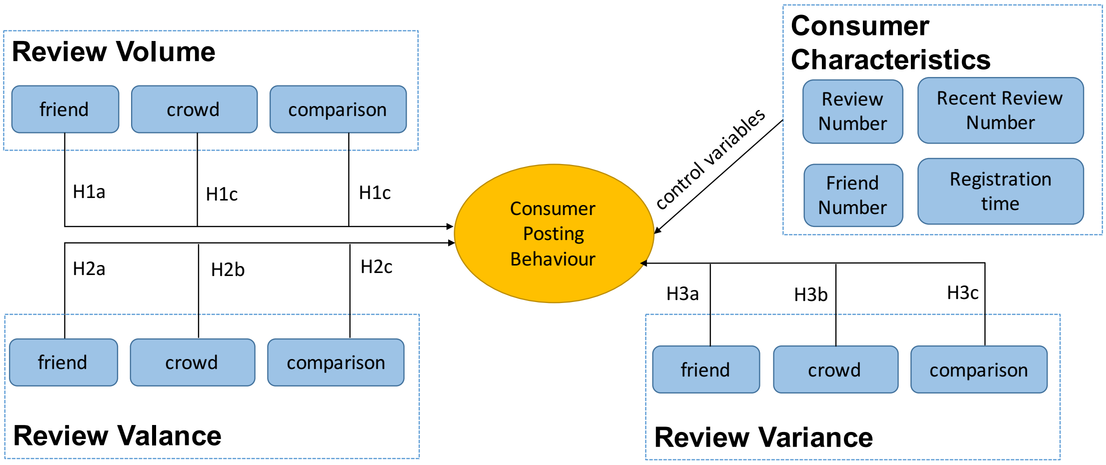


Figure 5.2| Research model in this chapter.

5.1.2 Review Valance

When users make purchasing decisions, a high valance can enhance the chance of a product to be selected. Highly-rated products are consequently more likely to be commented on. Ma *et al.* (2013) report that the average of previous ratings can serve as a signal to help consumers to form first impressions about the product, and positively impact subsequent consumer decision-making. Many studies have shown a positive link between the rating of products and sales (Godes and Mayzlin, 2004; Clemons *et al.,* 2006). They argue that the review valance may reflect the quality of items and that people rely more on positive cues (high valance) than negative ones. Doh and Hwang (2009) show that reviews with higher star ratings have a positive significant effect on purchase intention. Gershoff *et al.* (2003) suggest that positive reviews have a stronger impact than negative ones. Thus, we hypothesize that subsequent consumers may pay more attention to high-valance reviews:

- **Hypothesis 2a** (The Crowd's Review Valance Influence and Posting Engagement Hypothesis): The average of ratings posted by the crowd positively influences the likelihood of subsequent consumer engagement in posting behaviour.
- **Hypothesis 2b** (The Friends' Review Valance Influence and Posting Engagement Hypothesis): The average of ratings posted by a target consumer's friends positively

influences the likelihood of subsequent engagement in posting behaviour.

However, some studies argue that the review valance may not reflect the true quality of products (Li and Hitt, 2008) for two reasons. One is due to ‘‘forum manipulation” that firms employ paid reviewers to create high ratings. The other is that ratings may represent a mix of objective product quality and subjective assessments of value based on consumer fit. Therefore, ratings may be biased. For example, the early ratings of a start-up business may be very high, while it is possible that the firm created some fake reviews to appeal to consumers.

The fact that the crowd ratings may overestimate the product quality, makes friends’ reviews more credible to consumers. Driven by the intuition that ‘‘if I like that person, I may also be interested in his content,” a user’s friends normally provide better recommendations than others (Sinha and Swearingen, 2001). Thus, consumers are more likely to trust the average rating of friends’ reviews, rather than the crowd’s. Social networks are communities that can be sustained by a sense of participation (Zhang *et al.,* 2011). A well-established community makes people feel useful and have a sense of belonging (Zadeh *et al.,* 2010). Therefore, a forward loop may exist in the community that consumers have tendency to join friend discussions, and we assert:

- **Hypothesis 2c** (The Greater Effect of Review Valance on Posting for Friends Than the Crowd for Posting Hypothesis): The review valance of a target consumer’s friends is more influential on her/his posting behaviour than that of the crowd’s.

5.1.3 Review Variance

The variance of ratings is a common measure to capture the heterogeneity of consumer opinions. From a managerial perspective, this variance is also an easy way to monitor consumer preferences and predict potential purchase decisions. Normally, a low variance may suggest the product fits a broad range of interests, while a high variance is associated with a niche product suiting only a small group of consumer interests. Additionally, consumers tend to post extreme ratings when there is a big gap between their perceived quality and expectation (Anderson, 1973). The distribution of product rating thus has a right-skewed U-shape (McGlohon *et al.,* 2010; Anderson, 1998), which means consumers who are extremely satisfied or unsatisfied are more likely to post an opinion.

Literature study the effects of discordant opinions as well. There is evidence that inconsistent ratings have negative impacts on subsequent demand or sales because high variance may result in high risk for having a bad experience (Muchnik *et al.,* 2013). In contrast, high variance in ratings may trigger curiosity leading to higher demand and more discussions (Clemons *et al.,* 2006; Sun, 2012). Discordant findings thus are probably related to product types: search goods versus experience goods (Mudambi and Schuff, 2010). The quality of experience goods cannot be known before a consumer actually experiences or purchases them, which makes online ratings of prior buyers become useful information resources. For example, Ye *et al*. (2011) indicate that the high variance of prior ratings may decrease the sales of hotel rooms, which are typical experience goods. On the other hand, high-variance product reviews have been found to facilitate the likelihood of purchasing MP3 players (Park and Park, 2013) as search goods. As the target products to be studied in this chapter are experience goods like restaurant and hotel, we propose:

- **Hypothesis 3a** (The Crowd's Review Variance Influence and Posting Engagement Hypothesis - H3a): The variance of ratings posted by the crowd negatively influences the likelihood of subsequent consumer engagement in posting behaviour.

Most prior online review studies have focused on the impact of the variance of all ratings, while little is known about how the distribution of friends' ratings matters. The theory of innovation diffusion posits that new ideas, practices and objects would become known and spread quickly within communities (Gatignon and Robertson, 1985). Individuals within a friendship network act as WOM channels, inspiring others to imitate their behaviour and consumption experience (Flynn and Goldsmith, 1999). However, the long-developed spiral of silence theory (Noelle-Neumann, 1974) in social science has suggested that open deliberation may be impeded when friends disagree with each other in social discussions. An example is discussions of political elections. Hampton *et al.* (2017) find that disagreement among friends reduces the willingness of users in a social network to join a conversation. Due to posting reviews online is also a social discussion process, so we posit:

- **Hypothesis 3b** (The Friends' Review Variance Negative Influence on Posting Hypothesis): The variance of ratings posted by a target consumer's friends negatively influences the likelihood of her subsequent engagement in posting behaviour.
- **Hypothesis 3c** (The Greater Effect of Friends' Review Variance on Posting Than the Crowd Hypothesis): The review variance of a target consumer's friends is more influential on her posting behaviour than the crowd's.

5.2 Data Preparation and Model Specification

5.2.1 Data Preparation

To analyse the impact of friends' reviews and the crowd's reviews, as well as the consumer characteristics on the likelihood of a consumer to engage in the eWOM, we use a large set of review data from a user-generated content system, Yelp, as described in Chapter 4. As shown in Figure 4.1, Yelp provides social network service, which allows users to connect their friends in the system (Figure 4.1 c). A user can either go to the homepage of a business to go through its basic information (Figure 4.1a) and all reviews (Figure 4.1 b), or check her friends' recent reviews as a timeline displaying on her homepage (Figure 4.1 d). As a consequence, the reviews of both friends and the crowd may influence a user's consumption.

Table 5.1| Description of sampled data.

Data Level	Number	Description
Users	100	The number of users sampled from the Yelp dataset
Business	523	The number of businesses sampled from the Yelp dataset
		(Category of Business)
	414	Restaurant
	15	Arts and Entertainment (e.g., museums)
	14	Hotel
	7	Public Services (e.g., libraries, delivery)
	6	Shopping Centre
	5	Active Life (e.g., zoos, parking)
	4	Home Services (e.g., key, heat)
	4	Hair or Nail Salons
Duration	50	Number of weeks in the sampled data

The process of consumption posting reviews is shown in Figure 5.1. In Yelp, it is impossible to know for sure whether each reviewer has actually gone to the business, because Yelp is merely a platform for sharing experiences and opinions. However, the major purpose of posting reviews in Yelp, where there are mostly restaurants, is to share experience. Though there is a likelihood that a reviewer may fabricate her experience without really experiencing the business, we believe that this likelihood is fairly slim. So we assume that all reviews are posted after the reviewer indeed has consumption experience.

The data set of Yelp consists of 1,569,264 reviews posted by 366,715 users on 61,184 businesses during the year of 2005 to 2015. The relationships among the users, that is their social network, is known. Therefore, for a target user, we are enabled to identify whether a review is posted by the crowd or by his friends. To avoid sparsity in the data set (Pan and Zhang, 2011), we prepare the data for analysis as follows:

1) While over 70% of the reviews in the dataset is posted after the year of 2010, we focus only on 50 weeks from August 1, 2014 to July 15, 2015, when the posting behaviour is active.
2) We target at only businesses that locate in the City of Phoenix and have at least 100 reviews. Under the conditions, 523 businesses are selected. The reason why we choose the businesses in one city is that reviewers are more likely to be local residents, which means they are more likely to be friends on the Yelp. Thus, the social network connections could be more plentiful to benefit the research questions. The city of Phoenix is randomly selected.
3) For these selected businesses, the reviews are posted by more than 10,000 consumers. We randomly sample 1,000 of them who has at least ten friends, so that we could explore the different impacts of friends' and the crowd's reviews on their posting behaviour.
4) We regard each week as time t, and assume that each of the 1000 consumers is possible to post a review on each of the 523 businesses at each week. As a consequence, there are 1000*523*50 = 2.615* 10^7 data records for the regression analysis, and each record describes the posting behaviour of consumer u for

business b at time t.

5) The dependent variable is binary. For each data record, if user u posted a review on business b at time t, the corresponding value of the dependent variable was 1, and 0 otherwise.

6) The independent variables correspond to the hypotheses. The volume of reviews (popularity) is measured by the number of all reviews of business b before week t. Counting the number of reviews posted by crowd and friend, we then have the review volume among the crowd, CP_{ubt}, and among friends FP_{ubt}. Each review is associated with a numeric rating ranging from 1 to 5. The valance is measured by the average rating of reviews posted by the crowd and friends before week t, respectively. This is the average rating of the crowd reviews, CR_{ubt}, and the average rating of the friend's reviews, FR_{ubt}. The variance, CVR_{ubt} and FVR_{ubt}, are statistical variances of reviews on business b before week t posted by the crowd and consumer u's friends, respectively. The calculation of review volume, valance and variance is based on the whole data set, with 366,715 users rather than the sampled 1,000 users.

7) Drawing from the literature (Lee *et al.,* 2015; Moe and Schweidel, 2012), we consider four control variables. The number of historical reviews, NOR_{ut} and the review number in the recent week, $NORC_{ut}$ are extracted from the whole data set by week t. The number of friends, NOF_u is static, as the timing of establishment of friendship is not known from the data. The age of the user T_{ut} is the number of weeks since her registration time, RT_u by week *t,* that is, $T_{ut} = t - RT_u$.

In summary, the data for regression analysis in this chapter represents a dense sample, including 1,000 users and 523 businesses over 50 weeks, as described in Table 5.1. The independent variables, including consumer characteristics and the 3Vs of friend reviews and crowd reviews are calculated or extracted from the whole data set. A description of variables can be found in Table 5.2.

The sampled data selects consumers who have at least ten friends in Yelp. To address the impact of friends' reviews, we also extract the other group of data as a control, following the above steps. But in step 3), we sample 1000 consumers who have no friends in Yelp.

Table 5.2|Variables and their operationalization.

Dim.	Vari.	Description	Min	Max	Mean
User Char.	NOR_{ut}	# reviews submitted by user u till week t	0	447	14.89
	$NORC_{ut}$	Number of reviews submitted by user u at week $t-1$	0	48	0.44
	NOF_u	Number of friends of user u	0	1360	43.41
	T_{ut}	Weeks since user u registered	5	149	83.29
Crowd Reviews	CP_{ubt}	# reviews of business b in the whole system before week t	0	1186	150.47
	$PosCP_{ubt}$	# positive crowd reviews of business b before week t	0	832	148
	$NegCP_{ubt}$	# negative crowd review of business b before week t	0	354	27.78
	CR_{ubt}	Avg rating of business b in whole system before week t	1	5	3.86
	CVR_{ubt}	Variance of all ratings submitted until the week t	0	1.88	1.06
Friend Reviews	FP_{ubt}	# user u's friends commenting on business b before week t	0	104	0.45
	$PosFP_{ubt}$	# user u's friends commenting positively on business b before week t	0	64	1.82
	$NegFP_{ubt}$	# user u's friends commenting negatively on business b before week t	0	28	0.39
	FR_{ubt}	Avg rating user u's friends commenting on business b before week t	1	5	3.95
	FVR_{ubt}	Variance of ratings of business b from user u 's friends before week t	0	2	0.24

5.2.2 Model Specification

The dataset has a multilevel structure: consumer level and business level. Therefore, user or business heterogeneity can be appropriately controlled by an individual's or one business's observed characteristics. In addition, one week as a time stamp for posting reviews enables us to control for some unexplainable changes across time. Based on previous studies (Moe and Schweidel, 2012; Ying *et al.,* 2006; Lee *et al.,* 2015), here we develop a multi-level mixed-effect Probit model to describe the likelihood of posting behaviour. The reason why we choose the model is that the depend variable is binary, resulting in the linear regression model cannot be used. Following the literature which use probit model to study the possibility of consumer posting (Moe and Schweidel, 2012; Ying *et al.,* 2006; Lee *et al.,* 2015), thus we adopt the probit model as well to keep consistent. We also run the mixed-effect logistics model to the regression analysis, which indicates the worse fitting with higher BIC and AIC values, that will be discussed in the results.

The model separates friend reviews and crowd reviews, and also considers consumer characteristics and a random effect to reflect the varying baseline tendency of posters. We assume a consumer u would post a review for a business b at time t if

$$U_{ubt}^{*} = \delta_{u0} + b \cdot V_{ubt} + \mu_{ubt} > 0, \tag{5.1}$$

where V_{ubt} is the vector consisting of influential factors which may include the mentioned three aspects: consumer characteristics, $X_{ut} = \{NOF_u, NOR_{ut}, NORC_{ut}, \ T_{ut}\}$, crowd review information $C_{ubt} = \{CP_{ubt}, CR_{ubt}, CVR_{ubt}\}$ and friend review information $F_{ubt} = \{FP_{ubt}, FR_{ubt}, FVR_{ubt}\}$. The term μ_{ubt} is an idiosyncratic error and the term δ_{u0} represents the varying baseline tendencies for individuals to submit a review. The standard deviation of δ_{u0} is σ_{δ_0}.

To address and compare the impact of friend reviews and crowd reviews on consumer posting behaviour, we develop three models to analyse the data. As a control, model 1 specifically studies a group of consumers who have no friends at all, and thus only considers the consumer characteristics X_{ut} and crowds' review information C_{ubt}. Accordingly, the possibility of consumer u to post a review to business b at week t gives by the Probit model:

$$\text{Model1}: \ \Pr(z_{ubt} = 1) = \phi(b_{u0} + b_{1:4} \cdot X_{ut} + b_{5:7} \cdot C_{ubt}), \tag{5.2}$$

where $\phi(.)$ denotes the standard normal cumulative distribution function and z_{ubt} = 1 indicates that there is a review and vice versa. Since those consumers have no friends, all the possible impacts come from either themselves or the crowds. On contrast, we also analyse the behaviours of consumers with at least ten friends. In respect to model 1, we develop model 2 considering only the consumer characteristics X_{ut} and crowds' review information C_{ubt} as well.

Note that, the expressions for both model 1 and model 2 are the same, but are applied to different data sets. Model 1 studies consumers with no friends, while model 2 studies consumers with at least ten friends. Therefore, we have,

$$\text{Model2}: \ \Pr(z_{ubt} = 1) = \phi(b_{u0} + b_{1:4} \cdot X_{ut} + b_{5:7} \cdot C_{ubt}). \tag{5.3}$$

At last, we apply all the possible factors to study the impact of friends' and crowds' reviews on consumer posting behaviour in model 3 by using the data in which users have at least ten friends, which reads,

$$\text{Model3}: \ \Pr(z_{ubt} = 1) = \phi(b_{u0} + b_{1:4} \cdot X_{ut} + b_{5:7} \cdot C_{ubt} + b_{8:10} \cdot F_{ubt}). \tag{5.4}$$

5.3 Analytical Results

We use the three multilevel mixed-effect Probit models developed in the last section to analyse the prepared Yelp data. The results are shown in Table 5.3.

To summarise, Model 1 analyses the users with no friends, and thus the likelihood of posting is assumed to be influenced only by consumer characteristics and crowd reviews. On the other hand, both Model 2 and Model 3 analyse users with at least ten friends. However, Model 2 does not consider the influence of friends' reviews while Model 3 does. We normalise each variable using the min-max normalisation, and which makes the estimated coefficients comparable to each other.

From the results in table 5.3, first of all, we get the marginal and conditional R^2 of the mixed effect models developed by Nakagawa and Schielzeth (2013). The results are shown in the tables, where the full model gets the highest conditional R^2. However, the low marginal R-squared of the analytical results suggest the limitation of the proposed mixed models. Since the marginal R-squared for mixed models is not the same "variance explained" as for the linear models, we do not pay much attentions on the values.

The likelihood-ration (LR) tests show that $Prob >= chibar2$ is quite small in three models. The estimated variance of the random intercept at the user level, $\sigma_{\delta_0}^2$, is 0.141, 0.21, 0.21 in the three models, respectively. All these results indicate that there is enough variability between users to favour a mixed-effect probit regression over an ordinary probit regression. Secondly, the χ^2 test (Prob>chi2 is zero) rejects the null hypothesis suggesting significant impact of all independent variables. The values of AIC and BIC are suggested to determine which model is better (Kass and Raftery, 1995; Kyritsis *et al.,* 2018). The evidences given by the AIC and BIC in model 2 and 3 suggest model 3 is a better fit than model 2 ($\Delta BIC_{2.3} = 68, \Delta AIC_{2.3} = 109$).

Table 5.3|Estimates of consumer characteristics, friends' and crowd's impact on posting behaviour.

	Model 1		Model 2		Model 3	
Variables	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
b_0	-0.449	10.929	-14.528	9.770	-3.551***	0.168
b_1 NOR_{ut}	0.640**	0.252	0.515**	0.181	0.391*	0.183
b_2 $NORC_{ut}$	3.840***	0.155	4.589***	0.136	4.609***	0.137
b_3 NOF_u			0.366	0.349	-0.861*	0.383
b_4 T_{ut}	0.271	0.21	-0.350***	0.093	-0.299**	0.094
					(Crowd)	
b_5 CP_{ubt}	0.951***	0.074	0.612***	0.04	0.326***	0.064
b_6 CR_{ubt}	20.56	43.72	-31.31	30.02	-0.312*	0.064
b_7 CVR_{ubt}	-9.458	20.611	15.82	15	-0.784***	0.162
					(Friend)	
b_8 FP_{ubt}					2.035***	0.271
b_9 FR_{ubt}					0.264***	0.064
b_{10} FVR_{ubt}					0.229***	0.057
Variation of baseline tendency						
$\sigma^2_{\delta_0}$	0.141	0.031	0.21	0.025	0.21	0.025
LR test VS Probit model						
Chibar2	1645.89		1223.25		1191.67	
Prob>=chibar2	0.00		0.00		0.00	
Wald chi2	776.7		1308		1394	
Prob>chi2	0		0		0	
AIC	15812		15728		15619 (Logstics:15697)	
BIC	15925		15848		15780 (Logstics:15858)	
Marginal R^2	0.038		0.037		0.045	
Conditional R^2	0.267		0.235		0.683	

*p<0.05, **p<0.01, ***p<0.001

We attempt the mixed effect logistics model to the data in the model 3 as well. Here we only display the AIC and BIC values to compare the overall fit. According to the values in the model 3, the probit model that we select suggests a little bit better than the logistics model ($\Delta BIC = 78$ and $\Delta AIC = 78$), which is also another reason that we use probit rather than logistics.

5.3.1 Impact of Review Volume

For the crowd reviews, the volume CP_{ubt} is positive ($b_5 > 0$) with respect to influence on the likelihood of consumer posting, and this influence is significant in all three models. Thus, the Crowd's Review Volume Influence and Posting Engagement Hypothesis (H1a) is supported. This suggests a ''rich get richer" effect that businesses with a lot of reviews (high volume) tend to get more reviews. Similarly, the volume of friends' reviews, FP_{ubt}, has a positive impact as well ($b_8 > 0$) in Model 3, so the Friends' Review Volume Influence and Posting Engagement Hypothesis (H1b) is supported. Though the volume of reviews was significant in its association with the likelihood of consumer posting behaviour (Dellarocas and Narayan, 2006; Liu, 2006; Duan *et al.,* 2008), our results suggest that there is a different effect for friends and the crowd. While businesses that have been widely reviewed by either the crowd or friends tend to be further reviewed by the target consumer, the volume of friends' reviews is more influential than that of the crowd's reviews ($b_8 = 2.035,\ b_5 = 0.326$). This result supports the Greater Effect of Review Volume for Friends Than the Crowd for Posting Hypothesis (H1c).

As we discussed in section 5.1.1, a large volume of the crowd reviews may suggest that a business is of common interest for most consumers, while the volume of friends' reviews may indicate the interest of the target consumer's local social group. Since friends normally have similar interests, tastes and so on (Pan *et al*., 2017; Leskovec *et al.,* 2007), a consumer is more likely to select the businesses associated with her friends' interests (high volume of friend reviews), rather than those for which many share a common interest (high volume of crowd reviews). A higher possibility of consumption is associated with a higher likelihood of being reviewed. In addition, when deciding whether to post a review, a consumer may also want to behave similar to her friends due to the desire of maintaining the friendships and sharing a common experience with them (Schieman and Van Gundy, 2000), for example, friends would have similar music taste after chatter with friends (Dhar & Chang, 2009).

5.3.2 Impact of Review Valance

Previous studies show different opinions about the impact of review valance, as we present in Section 5.1.2. Our results regarding review valance may largely supplement the existing theories. First, the valance CR_{ubt} is not significant for Model 1 and 2, which shows differences from previous studies (Godes and Mayzlin, 2004; Chintagunta *et al.,* 2010). The reason may lie in the fact that these studies do not distinguish the friends from the crowds. The impact of valance is contributed by both friends and non-friends. In our study, on the other hand, the crowd does not include the target consumer's friends. Second, in our main model, the valance of crowd reviews is significant and negative ($b_6 < 0$) influencing the likelihood of consumer posting, while friends who highly evaluate the business enhance the likelihood for a target consumer to post ($b_9 > 0$). That means a consumer would like to select businesses with higher stars on average from friends, instead of the crowd. Therefore, the Friends' Review Volume Influence and Posting Engagement Hypothesis (H2b) is supported but the Crowd's Review Valance Influence and Posting Engagement Hypothesis (H2a) is not. The effect of the valance of crowd reviews is slightly stronger than that of the friends ($b_9 = 0.264$ less than $|b_6| = 0.312$). So the Greater Effect of Review Valance on Posting for Friends Than the Crowd for Posting Hypothesis (H2c) is not supported.

5.3.3 Impact of Review Variance

Review variance represents how different consumers evaluate the same product. The variance of the crowd reviews, CVR_{ubt}negatively influences the likelihood of consumer posting, confirming the Crowd's Review Variance Influence and Posting Engagement Hypothesis (H3a). However, the variance of friends' reviews, FVR_{ubt} is positive ($b_{10} > 0$) as an influencer of the posting likelihood. This result suggests that the variance of crowd reviews had stronger impact on posting behaviour than that of friends ($b_{10} = 0.229$ less than $|b_7| = 0.784$). Though the diversity of friends' opinions to some extent promotes more discussions, such an effect may be set back by the risk of having a bad experience, based on the high variance of the crowd's reviews. Consumers prefer to choose those businesses that others have consensus about in terms of their quality. Accordingly, H3a is supported, while the Friends' Review Variance Negative Influence on Posting Hypothesis (H3b) and the Greater Effect of Friends' Review Variance on Posting Than the Crowd Hypothesis (H3c) are not.

5.3.4 Consumer Characteristics

The consumer characteristics are control variables in this study. Our results suggest that they are of significant influence on the likelihood of engagement in eWOM. Both the historical number of reviews and the more recent number of reviews have positive effect on posting likelihood ($b_1 > 0, b_2 > 0$), and thus the H1 and H2 hypotheses are supported by all the three models. In particular, a user's recent activities are much more explanatory than her historical activities ($b_2 > b_1$). The number of friends, on the other hand, is significant in only the Model 3. However, the results suggest that the number of friends is negatively ($b_3 < 0$) correlated with posting likelihood.

Our finding here is different from what was obtained in a prior study (Lee *et al.,* 2015), and there may be two reasons for this. First, the number of friends, as discussed in Section 5.2.1, is a static number that we collect at the end of the timeline of our data. But the regression analysis considers a dynamic process, based on week *t* as the timestamp. Second, Model 3 considers the influence of friends' reviews, and the number of friends may be correlated with these variables. The age of user, measured by the weeks since the user registered in the system, is also found to be negative ($b_4 < 0$) influencing the likelihood of posting, but only statistically significant in Model 2 and 3. In other words, the users tend to post reviews in the early stage after registration, and become less active after a while.

5.3.5 Types of eWOM

To further discuss the different impacts of friend reviews and crowd reviews, we also study how positive or negative eWOM affects consumer posting behaviour. The type of eWOM is measured by the ratings. A review with more than three star is likely to be positive, and those with less than three will be negative. We also distinguish the impact of number of positive and negative friends' and the crowd's reviews on posting behaviour, to supplement our results for the different impact between friend and crowd. We derive four variables: $PosCP$ (number of positive crowd reviews), $NegCP$ (number of negative crowd reviews), $PosFP$ (number of positive friend reviews), and $NegFP$ (number of negative friend reviews). Thus, the Probit model is as follows,

Model4 : $\Pr(z_{ubt} = 1) = \phi(b_{u0} + b_{1:4} \cdot X_{ut} + b_5 \cdot PosCP_{ubt} + b_6 \cdot NegCP_{ubt}$

$$+b_7 \cdot CR_{ubt} + b_8 \cdot CVR_{ubt} + b_9 \cdot PosFP_{ubt} + b_{10} \cdot NegFP_{ubt}$$
$$+b_{11} \cdot FR_{ubt} + b_{12} \cdot FVR_{ubt}), \qquad (5.5)$$

Past research has offered different opinions about the role of positive and negative eWOM. Some suggest that positive eWOM can make consumers more willing to act on advice in their subsequent purchases based on the satisfaction expressed by others (East *et al.,* 2008; Goldenberg *et al*., 2007). However, others have argued that negative eWOM is more influential. This is because negative information is rarer than positive information, and therefore, is more diagnostic. This is referred to as the negativity effect (Fiske, 1980; Chevalier and Mayzlin, 2006).

Table 5.4|Estimates of positive and negative eWOM's impact on posting behaviour.

Variables	Estimate	S.E.
$PosCP_{ubt}$	0.026	0.078
$NegCP_{ubt}$	0.634***	0.113
CR_{ubt}	0.062	0.158
CVR_{ubt}	-7.623	20.42
$PosFP_{ubt}$	1.548***	0.207
$NegFP_{ubt}$	0.281	0.297
FR_{ubt}	0.155*	0.684
FVR_{ubt}	7.322	21.65
NOR_{ut}	4.612***	0.137
$NORC_{ut}$	0.422*	0.183
NOF_u	-0.748*	0.375
T_{ut}	-0.308**	0.093
Variation in baseline tendency		
$\sigma^2_{\delta_0}$	0.208	0.025
LR test VS Probit model		
Chibar2	1200	
Prob>=chibar2	0.00	
Wald chi2	1402	
Prob>chi2	0	

*p<0.05, **p<0.01, **p<0.001

The measures of fit in table 5.4 are similar to the above three models in the table 5.3. The estimated variance of the random intercept is 0.208 with the error of 0.025, indicating that the significant variability of users in the posting behaviour supported by the result of LR test as well. The test of χ^2 (Wald chi2 is 1402, Prob>chi2 is zero) means all selected variables have significant impact.

The results in Table 5.4 show that only the volume of negative crowd reviews and the positive friend reviews play a significant role, while the volume of positive reviews from the crowd and negative reviews from reviews have no influence.

The volume of negative crowd reviews $NegCP_{ubt}$ has a positive impact, which verifies the negativity effect (Chevalier and Mayzlin, 2006). But friend review regarding the type of eWOM has opposite result. The influence of friends' reviews on the likelihood of consumer posting in the presence of positive eWOM lead to a stronger impact, as shown by the coefficient of $PosFP_{ubt}$ (1.548). This implies that a consumer prefers to engage to evaluate businesses for which her friends had more positive altitudes. So friends' recommendations seem to be more influential for influencing consumer purchase intentions (Brown and Reingen, 1987), and further engagement with eWOM communications.

Such results show for the first time the interesting interplay between $PosCP, NegCP$, $PosFP, and\ NegFP$ with the likelihood of posting. Though it has been widely suggested that both positive and negative eWOM have significant impact, such impact largely rely on their sources. In particular, the results explain the finding about the negative influence of the valance of crowd reviews, and the positive finding of friend reviews. (See Table 5.3.) The volume of negative crowd reviews is shown in Table 5.4 to be positively correlated with the likelihood of posting, while the crowd's negative reviews is associated with lower valance values. The same logic applies to friends' reviews: that a large volume of positive reviews leads to a higher impact for their valance.

5.4 Summary

Consumers frequently look at the behaviour of others before making their own decisions and eWOM thus has attracted a lot of attention related to online marketing and the understanding of consumer posting behaviour. While reviews are widely known to be influential, we study in this chapter how friends' reviews and the crowd's reviews differently influence consumer posting behaviour after purchase. We hypothesize that likelihood of a consumer posting a review toward a particular business is influenced jointly by consumer characteristics, and the volume, valance and variance of friends' and the crowd's reviews. We carry out an analysis based on a multilevel mixed-effect probit model for a large dataset of Yelp. Our major findings are summarised in Table 5.5.

The volume of friends' reviews and the crowd's reviews both have positive impacts on the likelihood of consumer posting. The impact of friends' review volume is stronger though. The valance and variance of friends' reviews and the crowd's reviews show the opposite effect. The valance and variance of the crowd reviews have no significant influence when friend reviews are not considered, but negative influence when they are taken into account. Friends reviews always seem to have a positive impact, but the impact of the crowd reviews related to valance and variance is stronger than that of friend reviews. We also study the impact of sentimental types of eWOM on consumer posting behaviour. Positive and negative review volumes are separately analysed in our model. We find that the volume of negative reviews from the crowd is positively influence the likelihood of consumer posting behaviour, which verifies the negativity effect. But the influential type is the volume of positive friend reviews, and the impact is bigger than the volume of negative crowd reviews.

Table 5.5|Summary of findings for consumer engagement in posting behaviour.

Hypotheses	Description	Findings	Results
H1a	Volume of the crowd's reviews	Positive	Supported
H1b	Volume of friends' reviews	Positive	Supported
H1c	Comparison between the two	Friends' reviewing volume is more important	Supported
H2a	Valance of the crowd's reviews	Negative	Not Supp.
H2b	Valance of friends' reviews	Positive	Supported
H2c	Comparison between the two	Crowd's reviewing valance is more important	Not Supp.
H3a	Variance of the crowd's reviews	Negative	Supported
H3b	Variance of friends' reviews	Positive	Not Supp.
H3c	Comparison between the two	Crowd's reviewing variance is more important	Not Supp.

Chapter 6. Modelling Social Influence on Consumer Selection Behaviour

Social influence drives human selection behaviours when numerous objects competing for limited attentions, which leads to the “rich get richer” dynamics where popular objects tend to get more attentions. However, results from Chapter 5 indicate that, both the information from crowds and the information among one’s friends have significant influence over the one’s engagement to eWOM. In the field of social influence, evidences have been found that the global popularity of a merchandise may enhance the chance of it getting further attention. On the other hand, opinions and behaviours such as selection can also spread through social ties, implying local influence. Thus, how the social influence is shaping users’ selection behaviour and whether the local influence or the global influence is more determinative become a key question (RQ2 of the thesis). Consumer selection behaviour, in this chapter, essentially refers to consumer purchase behaviour, but the data of Yelp has no records regarding purchase, only has the posting records. Users who have purchased an merchandise, but do not post reviews are not known in the data. We ignore these users, only consider the purchase behaviour of those users who can be identified by posted reviews. Although the selection behaviour in this chapter represents the same meaning as the posting behaviour, we still use the term of “selection” to indicate the consumer purchase decision being the focus of this chapter, and to distinguish it from the previous chapter. More discussions of the two chapters can be found in the summary of this chapter.

This chapter explores the intensity of these two social influences that are measured by the objects’ popularity on consumer selection behaviour based on the data set from Yelp. Section 6.1 defines the two potential social influences, namely friend-based and crowd-based social influence respectively. Section 6.2 applies a bipartite network model to describe the Yelp dataset, which offers the context for further analysis. Section 6.3

empirically identifies and distinguishes the crowd-based social influence and the friend-based social influence. Section 6.4 proposes an evolutionary network model based on Monte Carlo to uncover the mechanism of the selection behaviour in Yelp and quantifies the intensities of the two influences.

6.1 Introduction of Friend-based and Crowd-based Social Influence

It is very easy nowadays for people to access the information of merchandise such as quality, rating, popularity or even previous consumers' feedbacks from the mass media. Particularly, in many online systems, bestseller lists or highest-rated object lists are generally available for users to refer to. Those information aggregating the choices and opinions of the whole population of the system, can be recognised as global information corresponding to crowd-based social influence in this thesis. It has long been argued to be the key reference for human selection behaviour (Bikhchandani *et al.,* 1998; Chen *et al.,* 2011) leading to the "rich get richer" phenomenon. A good example is the event that, two scholars secretly purchased 50,000 copies of their newly published book which consequently made the bestseller list. Then the book sold very good despite mediocre reviews, and was remained as a bestseller (Bikhchandani *et al.,* 1998).

Another mainstream of the social influence research believes that people in the same social group act similarly to each other (Lewis *et al.,* 2012), since individuals are always engaged in group activities. Such source of the influence which can be regarded as the local information corresponding to friend-based social influence in this thesis. The friend-based social influence is also shown to be driving the human selection behaviour, i.e. people tend to select what their friends selected (Lewis *et al.,* 2012). It has long been argued that, merchandises are similar to viruses and ideas that could spread in the social network from an individual to their friends through the frequent interactions (Centola, 2010; Muchnik *et al.,* 2013). Accordingly, the friend-based social influence is also recognised as the social contagion phenomenon (Aral *et al*., 2009; Christakis and Fowler, 2013).

Despite numerous evidences have been found that, both two types of influence can drive human behaviours, few of the previous investigations distinguish and compare these two sources of social influence. Onnela and Reed-Tsochas (2010) argue that it is important to distinguish the local and global sources of social influence. However, the method they apply in their study, namely the fluctuation scaling, though successfully reveals the emergence of the social influence in terms of popularity, could not efficiently distinguish the friend- and crowd-based influence. Lee *et al.* (2015) recently discuss the crowd's and friends' influence over users' behaviours of rating movies. For a specific movie that has already been selected by a user, the rating on it is very likely to be influenced by the previous one. However, they only reveal the social influence on how good would a user evaluate an already-selected movie. The social influence on the selection behaviour which is also a crucial reflection of user preference (Hou *et al.,* 2014; Pan *et al.,* 2014), still needs to be investigated. Additionally, while they focus on the nearest predecessor's influence over the subsequent user, how would the aggregated historical information influence the users' decision is still an open question.

6.2 Scenario Setting with the Yelp

This chapter aims to distinguish the friend-based influence from the crowd-based influence over the consumer selection behaviour from numerous merchandises. To achieve the objective, one needs to possess the social structure of a collective of people and the records of their sequential selection behaviour over a number of certain merchandises. Thus, we use a large-scale data set from Yelp.com where users can not only read reviews on various kinds of businesses such as restaurants, shopping centres, pubs, but also establish friendships with other users.

As we elaborated in the last section, when a user is looking for a business on Yelp, the system offers various kinds of ranking list of businesses on the user homepage, and number of reviews, average rating on the business homepage. Those are the global information that could possibly influence a user's selection decision. On the other hand, there is also a timeline displaying the reviews of their friends on businesses which can be recognised as the local information for the user to refer to. Such explicit data provide us the opportunity to explore the question that, whether the opinions of the local neighbourhood, i.e. friends, or that of the whole population matter most for a user to make selection decisions?

The Yelp data can be well described by a user-business bipartite network with social structure shown as Figure 6.1. In the user layer, two users will connect with each other if they are friends in Yelp; and between the layers, a user node will connect to a business node if s/he has selected it. Therefore, the local information for a target user i can be represented by the opinions and decisions of his or her local neighbourhood, i.e. those users who are connecting to him or her. On the other hand, the global information then is the opinions and decisions of the whole user layer, *i.e.* all users either connected or unconnected to user i. Please be noted that it is impossible for us to gain the information that which user actually select which business unless the consumer post a review, thus the selection link in this chapter only represents the posting activity.

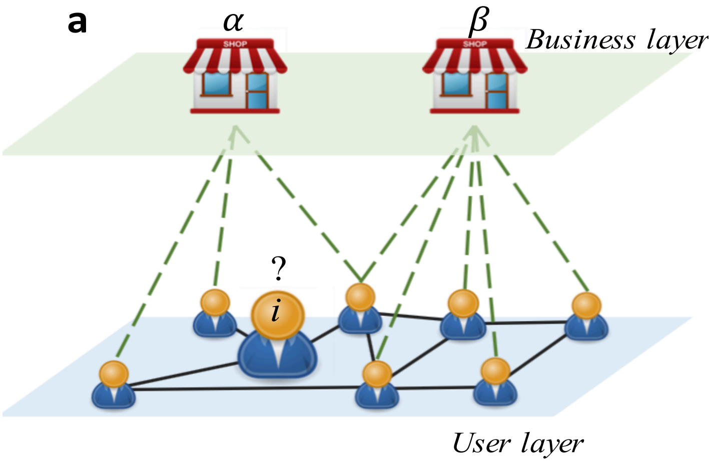


Figure 6.1| A network representation for the applied Yelp data.

As the popularity information is the most fundamental signal of social influence which has been argued to be self-reinforcing (Papadopoulos *et al.,* 2012), we take popularity to measure the local and global information in the system. Thus, the ***local popularity*** of a business α subjecting to a specific user i at time t, $LP_{i\alpha}(t)$, is defined as the number of user i's friends who have connected to the business α. The ***global popularity*** of a business α at time t, $GP_{\alpha}(t)$ is the number of users in the whole user layer who have connected to it. Take the network shown in Figure 6.1 as an example, the local popularity of business α and β for the target user i, $LP_{i\alpha}$ and $LP_{i\beta}$ are 3 and 1 respectively. The global popularity of business α and β, $GP_{\alpha}$and GP_{β} are 3 and 5 respectively. Given the fact that the business β is globally more popular than α, classical selection mechanism driven by global social influence (Barabasi and Albert, 1999; Fortunato *et al.,* 2006; Diaz *et al*., 2010) may predict that it is more likely for the target user i to connect to the business β. However, despite the low global popularity, the business α has been connected by all of the user i's three friends. Being locally more popular, the business α is more likely to be recommended by the target user's friends which may also enhance the possibility for user i to connect to it. Thus, a fundamental question the present study aims to answer could be simplified as: when making selection decisions, it is more likely for a user i to be influenced by his/ her friends (connect to business α), or by the crowd (connect to business β)?

6.3 Social Influence in the Yelp

6.3.1 Identifying the Social Influence

The confirmation of the existence of crowd-based (global) social influence is rather easy, for which methods are well-established. In the field of complex networks science, it has been uncovered both theoretically and empirically that, the preferential attachment would eventually lead to a power-law degree distribution (Barabasi and Albert, 1999). The so-called preferential attachment is actually the tendency that nodes (e.g. users) tend to connect to the popular nodes (e.g. businesses), and thus well represents the crowd-based social influence. Consequently, to examine whether the crowd-based social influence exists in the user selection behaviour in Yelp, one can explore the distribution of global popularity of businesses, which is expected to be a power-law distribution.

Similar to what has been observed from many systems, the global popularity of businesses in Yelp indeed follows the power-law distribution as shown in Figure 6.2, $i.e. p(GP(T) \sim GP(T)^{\gamma})$ with $\gamma = -1.7$, where $t = T$ is the end of the data. Such power-law distribution of the global popularity indicates that the popular objects could attract more connections. Thus, the crowd-based social influence in Yelp can be confirmed.

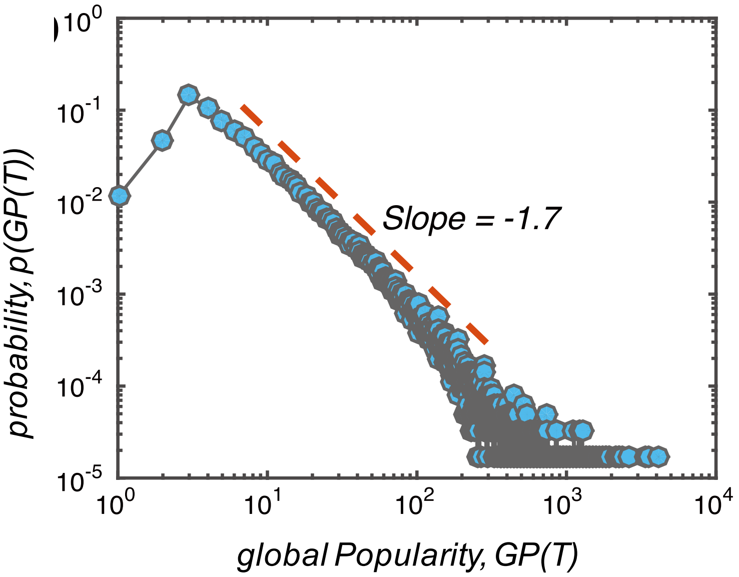


Figure 6.2| The distribution of the businesses' global popularity

To explore if there is the friend-based social influence, here we examine the real-time local popularity of the business α corresponding to the user i at time c when they are connected, $LP_{i\alpha}(c)$. Suppose a user i with k_i friends, connected to a business α with the global popularity $GP_\alpha(c)$, the ***expected local popularity*** should be $LP_{i\alpha}^{exp}(c) = k_i \cdot GP_\alpha(c)/M$, where M is the total number of consumers in the system. The expected local popularity, $LP_{i\alpha}^{exp}(c)$ describes the case that, the target consumer i's friends have no particular preferences in the business α when the consumer connects to it. Accordingly, if the actual local popularity $LP_{i\alpha}(c)$ when the user selected the businesses is significantly higher than the expected local popularity $LP_{i\alpha}^{exp}(c)$, i.e. $LP_{i\alpha}(c) > LP_{i\alpha}^{exp}(c)$, the existence of friend-based social influence can be confirmed.

We also compare the empirical data to two null models, so that the possibility that the resulted patters come from random mechanism or global social influence can be eliminated.

Global preferential attachment (GPA) model

Inspired by the preferential attachment mechanism in many social and technical systems (Barabasi and Albert, 1999; Fortunato *et al.,* 2006), various models (Ratkiewicz *et al.*, 2010; Diaz *et al.*, 2010) have been developed to describe the crowd-based influence, where the global popularity is the key indicator of an object's attractiveness for future selections. Therefore, we denote such method with global preferential attachment (GPA).

In order to make the simulated results comparable to empirical observations, we take the real size of the network, i.e. we consider a network consists of $N = 366{,}715$ users and a growing number of businesses. However, as the GPA model considers only the business's global popularity, the underlying social structure is irrelevant to the evolution and thusly not considered in this model. The growing rate of the businesses in the simulation is set to be the real rate of the system as shown in Figure 4.2(b). When a business enters the system, we suppose it to be connected by a random user, which means the popularity of each business at its entrance time is *GP = 1*. During the evolution, at each step of the simulation, a user i is randomly selected to establish a connection to an existing business α, that it is selected according to a probability proportional to its global popularity, i.e. the probability of business α being connected at time step t is $prob(\alpha, t) = {GP_\alpha(t-1)}\Big/{\sum_{\beta \in \Gamma_i} GP_\beta(t-1)}$, where Γ_i is the set of businesses that has not been connected by user i at the time t. The simulation continues until the number of user-business links reach 1,569,264 which is the real number of links in the Yelp data. Due to the pattern of connection of GPA, the target consumer's local popularity is expected to have no influence on the evaluation. In other words, we should not observe any friend-based social influence in the data generated by the GPA model.

<u>*Random experiment*</u>

In the random experiment, the global popularity of each business and the whole social structure are unchanged, while the wiring patterns between consumers and businesses are rewired. For example, if a business α is connected by GP_α consumers in the original data, we select GP_α consumers anew from the whole population uniformly at random and let them to connect to the business α. Meanwhile, we also keep the timestamp of each connection. In this way, the friend-based social influence would be removed because a consumer's selections would not be similar to his friends' anymore.

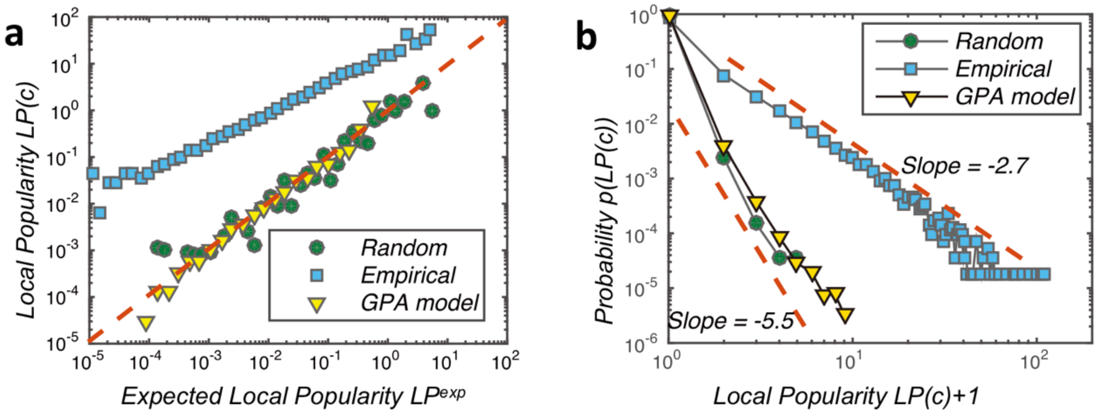


Figure 6.3|Local popularity of selection behaviours *LP(c)* versus the expected local popularity LP^{exp}. a, Local popularity of selection behaviours LP(c) versus the expected local popularity LP^{exp}. The red dashed line shows the condition that LP(c) = LP^{exp}. While the local popularity of the random experiments and global-driven preferential attachment (GPA) model are very similar to the expected value, the empirical local popularity is significantly higher which suggests that the users tend to select locally popular businesses. b, The distribution of real-time local popularity LP(c). For the empirical data, the local popularity follows the power-law distribution with slope of −2.7. On the other hand, the local popularity of the GPA model being very similar to the random experiment, cannot reproduce the empirical observation.

Figure 6.3 shows the comparison between the empirical results and the results of *GPA* model and random experiment. First of all, empirical local popularity, LP of those selection behaviours is much higher than the expected local popularity, LP^{exp}. Such result suggests that, businesses which are relatively more popular in the target consumer's friend circle are more likely to be selected. This is the evidence of the presence of the friend-based social influence, that the consumer may select what their friends have selected. Such phenomenon cannot be explained by traditional preferential attachment mechanism because *GPA* believes that the probability of new connections is determined by the global popularity rather than local popularity. Consequently, the real-time *LP(c)* generated by the *GPA* model coincide with that of the random rewiring, as shown in Figure 6.3 (a), which are very similar to the expected local popularity. Additionally, the empirical local popularity *LP(c)* exhibits a power-law distribution, i.e. $p(LP(c))\sim LP(c)^{\gamma}$ with $\gamma = -2.7$, as shown in Figure 6.3 (b). But the local popularity distribution of the *GPA* model and the random rewiring being similar to each other, are in very narrow ranges and the slopes are *−5.5*.

The analysis indicates that, the friend-based social influence is also a driver of consumers' selection behaviour in addition to the crowd-based social influence. Especially, we need to pay close attention on the mechanism of consumers' selection decision because the observations cannot be explained by neither the random mechanism nor the traditional preferential attachment mechanism.

6.3.2 Distinguishing the Friend-based influence from Crowd-based Influence

The power-law distribution of the global popularity as shown in Figure 6.2 implies that the evolution of the Yelp system could be characterised as "rich get richer" dynamics. Consequently, the global popularity must be driving the consumers' selection behaviours. On the other hand, we have also found the local popularity to be a notable driver for the system evolution. In other words, a business being either globally popular (large $GP(t)$) or locally popular (large $LP(t)$) enhances its probability to be selected by consumers. However, the global popularity and local popularity may be confounding factors because locally popular businesses are likely to be also globally popular. Therefore, here we try to distinguish the influence of global and local popularity on consumers' selection behaviour by analysing the probability of a business to be selected $P(s)$.

The conditional probability of a business being selected at a certain condition Θ, $P(s|\Theta)$ is calculated based on the empirical observation. The condition Θ could be any attributes of the business, but in this section, we only consider the popularity information, i.e. global popularity GP and local popularity LP. The probability is simply calculated as the fraction between the number of connections of the business and the possible connections, $P(s|\Theta) = N_{RS}(\Theta)/N_{PS}(\Theta)$. The real number of selection behaviour $N_{RS}(\Theta)$ is the number of established connections that satisfy the condition Θ. The calculation of the possible number of connections $N_{PS}(\Theta)$ is based on the assumption that each pair of consumer-business could be potentially connected to each other in each time interval t until the connection established. Figure 6.4 gives an example of the calculation using a toy evolution data with 3 consumers and 2 businesses over 3 time steps. After calculating the numbers of real and possible connections for all the possible conditions, one can get the probabilities for businesses with a certain condition Θ to be selected in a time interval t.

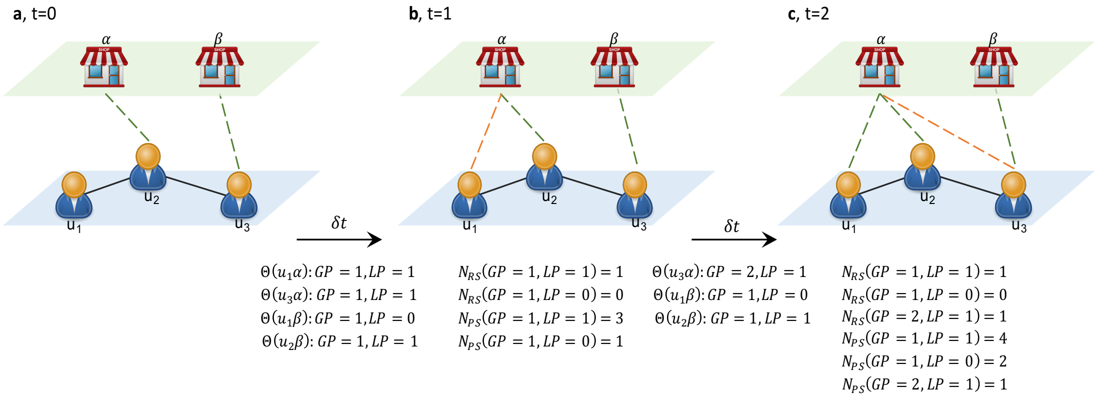

a, t=0
b, t=1
c, t=2
α
β
u1
u2
u3
δt
$\Theta(u_1\alpha): GP = 1, LP = 1$
$\Theta(u_3\alpha): GP = 1, LP = 1$
$\Theta(u_1\beta): GP = 1, LP = 0$
$\Theta(u_2\beta): GP = 1, LP = 1$
$N_{RS}(GP = 1, LP = 1) = 1$
$N_{RS}(GP = 1, LP = 0) = 0$
$N_{PS}(GP = 1, LP = 1) = 3$
$N_{PS}(GP = 1, LP = 0) = 1$
$\Theta(u_3\alpha): GP = 2, LP = 1$
$\Theta(u_1\beta): GP = 1, LP = 0$
$\Theta(u_2\beta): GP = 1, LP = 1$
$N_{RS}(GP = 1, LP = 1) = 1$
$N_{RS}(GP = 1, LP = 0) = 0$
$N_{RS}(GP = 2, LP = 1) = 1$
$N_{PS}(GP = 1, LP = 1) = 4$
$N_{PS}(GP = 1, LP = 0) = 2$
$N_{PS}(GP = 2, LP = 1) = 1$

Figure 6.4| A toy data to illustrate the calculation of probability of being selected, P(s)

In the toy data, there are two time intervals δt, where we could observe the evolution, i.e. from *t = 0* to *t = 1* and from *t = 1* to *t = 2*. During the first time interval, from *t = 0* to *t = 1*, there are only one connection $\mu_1 \rightarrow \alpha$ established, while there are in total four possible connections, which are $\mu_1 \rightarrow \alpha$, $\mu_3 \rightarrow \alpha$, $\mu_1 \rightarrow \beta$ and $\mu_2 \rightarrow \beta$. As the established connection $\mu_1 \rightarrow \alpha$ is with the condition $\Theta: GP = 1, LP = 1$, we then have $N_{RS}(GP = 1, LP = 1) = 1$. Additionally, among the four possible connections, three are with condition $\Theta: GP = 1, LP = 1$ and one is with condition $\Theta: GP = 1, LP = 0$. As a consequence, the possible numbers are $N_{PS}(GP = 1, LP = 1) = 3$ and $N_{PS}(GP = 1, LP = 0) = 1$. Given that the consumer u_1 has connected with business α at time *t=1*, this connection will not take into account for the following possible connections. Similarly, we could count the numbers for the second interval from *t = 1* to *t = 2*. After the statistics, there are in total of 3 conditions appeared in this toy data, i.e. $\Theta_1: GP = 1, LP = 0$, $\Theta_2: GP = 1, LP = 1$ and $\Theta_3: GP = 2, LP = 1$. Therefore, the probability of business with a certain condition Θ, $P(s|\Theta)$is estimated accordingly as $P(s|\mathrm{GP} = 1, \mathrm{LP} = 0) = 0/2 = 0$, $P(s|\mathrm{GP} = 1, \mathrm{LP} = 1) = 1/4 = 0.25$ and $P(s|\mathrm{GP} = 2, \mathrm{LP} = 1) = 1/1 = 1$. Although there are only limited possible conditions Θ in this toy data, the estimations in the full data set of Yelp would be much more accurate due to the abundant data amount. But the estimations of some extreme conditions such as very large LP and GP will be still inaccurate because such conditions may occur only for limited times.

We firstly examine the probability of being selected conditional to the global popularity $P(s|GP)$ and the local popularity $P(s|LP)$ separately.

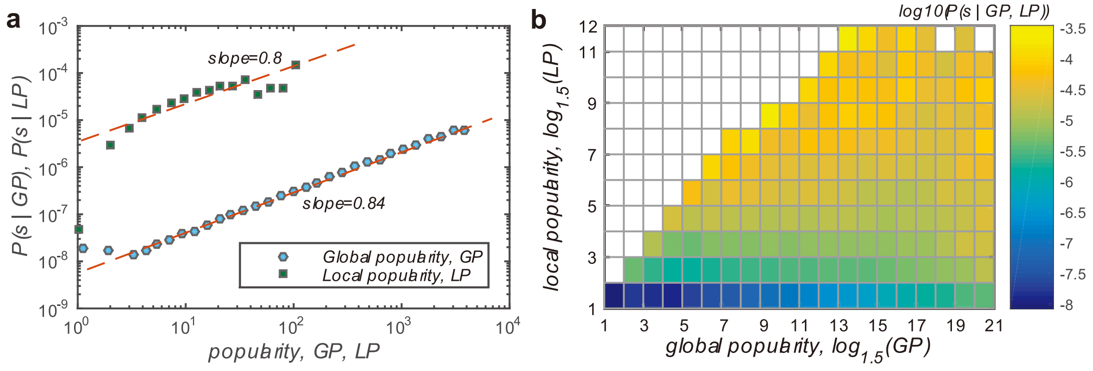


Figure 6.5| The probability of being selected conditional to the global popularity and local popularity.

As shown in Figure 6.5, both the global and local popularity has positive correlation with the probability. The larger a business's either global or local popularity is, the more likely it will be selected by consumers. The positive correlations indicate again that both the crowd- and friend-based social influence exist in the selection behaviour of Yelp users. Furthermore, the correlations could be well fitted by the power-law functions, $P(s|GP){\sim}GP^{\lambda}$ and $P(s|LP){\sim}LP^{\lambda}$. The parameter λ describes the increase speed of the probability as the business getting more and more popular either in the whole system or a consumer's friend circle. One can therefore use the parameter λ to quantify the intensity of social influence. As indicated by the results that $\lambda_{GP} = 0.84$ and $\lambda_{LP} = 0.8$, crowd-based social influence and friend-based social influence are of similar intensity.

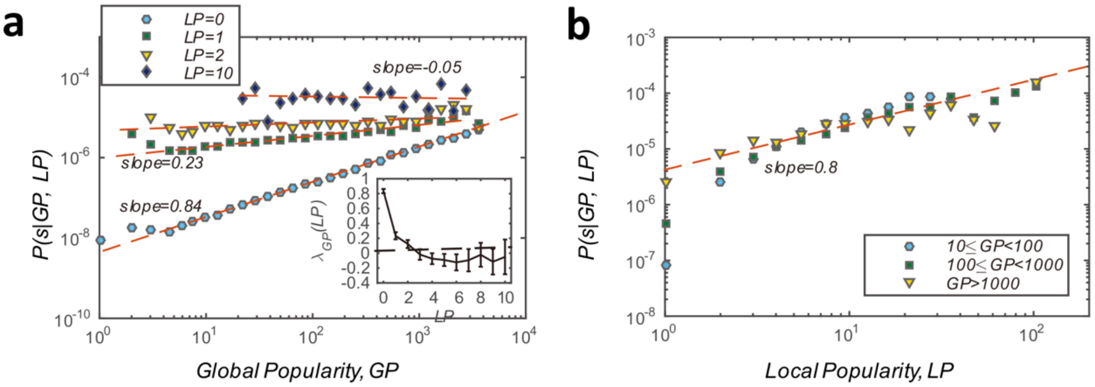


Figure 6.6|The probability of being selected conditional to a) the global popularity, with local popularity being controlled, and b) local popularity, with global popularity being controlled.

We further analyse how *GP* and *LP* jointly influence the probability of being selected $P(s|GP, LP)$. One can find from Figure 6.5 (b) that it is mainly the LP determining the probability. A locally unpopular business has very limited chance to be selected by consumers even if it is globally very popular. The LP is more effective in terms of the value of the probability. It can be observed from Figure 6.5(b) that, $P(s|GP = 1000) \approx P(s|LP = 2)$. To avoid the confounding effect of the whole and friend-based social influence, we take the GP and LP as control by turns.

When controlling *LP* at a fixed value *LP₀*, the correlation between the probability $P(s|GP, LP = LP_0)$ and *GP* could still be well fitted by the power-law functions as shown in Figure 6.6(a). For businesses with local popularity $LP = 0$, i.e. none of the target consumer's friends have selected it, *GP* is able to significantly enhance the selecting probability $(\lambda_{GP}(LP = 0) = 0.84)$. However, even only one of the target consumer's friends selected the business, the intensity of crowd-based social influence will drop to a quite low level, $(\lambda_{GP}(LP = 1)) = 0.23$. As the *LP* increases, the crowd-based social influence vanishes or even changes to weak, negative influence. The inset of the Figure 6.6(a) indicates that, there is no apparent crowd-based social influence for cases with *LP > 2*. On the other hand, the friend-based social influence is always very significant regardless of the global popularity level as shown in Figure 6.7(b) and the intensity $\lambda_{LP}(GP) \approx 0.8$.

Excluding the confounding effect among crowd-base and friend-based social influence, we could conclude that, the social influence from crowds on consumer selection behaviour exists only if there are not many friends' opinions to be referred to. It is the friend-based social influence always governing the consumer selection behaviour.

6.4 Modelling the Intensity of Social Influence

6.4.1 An Evolutionary Network Model based on Monte Carlo Simulation

To better understand the mechanisms of the friend- and crowd-based social influence, here we propose an evolutionary model to describe the users' selection behaviour. The fundamental mechanism of many systems can be described by the preferential attachment (Barabasi and Albert, 1999) where popular nodes have more chances to get new connections. Inspired by the models (Diaz *et al.,* 2010; Liu *et al.,* 2013) that have been trying to describe the evolution of bipartite networks based on the preferential attachment, we assume there may exist both the *local-driven preferential attachment* and the *global-driven* *preferential attachment.*

We consider a system with *N* consumers with a pre-defined social network among them, and a growing number of merchandises (in this case, businesses). When each business comes to the system, we suppose it will be connected by a random consumer. At each time step of the evolution t, a user i is chosen uniformly at random to connect to an object. With a probability μ, the user i will connect to the object according to the mechanism of local-driven preferential attachment, and the probability of each object α being connected, $prob^{local}(\alpha)$ is,

$$prob^{local}(\alpha) = \frac{LP_{i\alpha}(t)}{\sum_{\beta \in \Gamma_i} LP_{i\beta}(t)}, \tag{6.1}$$

where Γ_i is the set of objects that user i has not selected yet at the time t. Accordingly, the user i has a probability $1-\mu$ to perform a global-driven preferential attachment where the probability of each object α being connected, $prob^{global}(\alpha)$ reads,

$$prob^{global}(\alpha) = \frac{GP_{\alpha}(t)}{\sum_{\beta \in \Gamma_i} GP_{\alpha}(t)}\,. \tag{6.2}$$

Combing the local- and global-driven preferential attachment, we have the probability of a business α to be selected $prob(\alpha)$ reads,

$$prob(\alpha) = \frac{\mu \cdot LP_{i\alpha}(t)}{\sum_{\beta \in \Gamma_i} LP_{i\beta}(t)} + \frac{(1-\mu) \cdot GP_{\alpha}(t)}{\sum_{\beta \in \Gamma_i} GP_{\beta}(t)}. \tag{6.3}$$

In the model, μ is a tunable parameter ranging in [0, 1] which controls the influence of local- and global-driven attachment. The intensity of the friend-based social influence and crowd-based social influence could therefore be described by the parameter μ. The larger the parameter μ is, the stronger the friend-based social influence would be and at the same time, the weaker the crowd-based social influence would be.

6.4.2 Results of Simulation Modelling

We use the above model to simulate the evolution of the user-business bipartite network with underlying social structure to explore whether the model could reproduce the empirical observations of the local popularity distribution. To avoid the influence of other possible factors, we use the population (of both users and businesses) and the social structure of the Yelp data as the initial configuration of the model. With respect to the empirical data, we set the initial state of the simulation same with the data applied in this study, i.e. $N_{business}$ = 61, 184, N_{user} = 366, 715 and we use the empirical social structure as the pre-defined network among users. Furthermore, each simulation continues for 1,569,264 steps (same with the empirical data).

As shown in Figure 6.7(a), the simulations with different parameters μ can all reproduce the power-law global popularity distribution with slope same to the empirical observation. On one hand, such result suggests that, the traditional preferential attachment can indeed explain the emergence of the scaling phenomenon for the popularities. On the other hand, the local-driven preferential attachment may also be a driving mechanism of the power-law distribution. As to the local popularity in Figure 6.7(b), though the distributions with different parameters μ all exhibit linear pattern in the log-log plot, the slopes are different. For the parameter $\mu = 0$ (the traditional preferential model, which is totally driven by global popularity), the slope $\gamma \approx -5.5$, which is very similar to the random experiments shown in Figure 6.3(b) where the friend-based social influence has been removed. As the parameter of the model μ increases, the slope γ also gradually increases. For the experiment with $\mu = 1$ where the evolution is totally driven by the local-based preferential attachment, the slope could reach $\gamma = -2.1$.

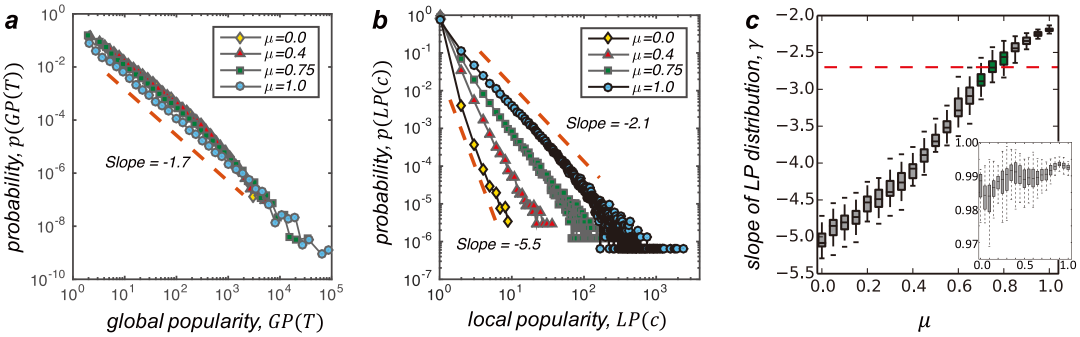


Figure 6.7|Results of the evolutionary model. **a**, Distributions of the simulated global popularity. **b**, Distributions of the real-time local popularity with different parameters μ. Each distribution exhibits a linear pattern in the log-log plot. **c**, The slope γ of the linear pattern for local popularity distributions with different parameter μ.

Figure 6.7(c) reports the correlation between the resulted slope of LP distribution and the parameter μ. For each parameter μ, the result is calculated based on 100 independent simulations. For each simulation, the fitting is based on a linear regression after taking logarithm for the simulated local popularity $LP(c)$ and the frequency (p.d.f.) of it $p(LP(c))$. The inset in the subplot (c) shows the coefficient of determination R^2 of corresponding fittings. The R^2 of the fittings are generally larger than 0.98 which indicates that the fittings can be considered good for all the experiments with different parameters μ. The red dashed line is the slope γ of the empirical local popularity distribution shown in Figure 6.3 (b), i.e. $\gamma^{em} = -2.7$. It is indicated by the green boxes that the parameter of the model should be $0.7 \leq \mu \leq 0.8$ to reproduce the results as same as the empirical result.

From the point of view of GP distribution, any combination of global-driven preferential attachment and local-driven preferential attachment could explain the empirical findings. However, from the point of view of LP distribution, the local-driven preferential attachment is responsible for about 75% of the evolution. In other words, 75% of the Yelp users' selection behaviours are driven by the friend-based social influence according to the consistency between the evolutionary model and the empirical observations.

6.6 Summary

The development of the modern world offers us numerous choices when we want to read a book, watch a movie or go out for a dinner. While making choices, the social influence has long been argued to be driving our behaviour. To distinguish the friend-based social influence and the crowd-based social influence is to explore whether it is our friends' or the whole population's opinion that matters most for us to make the decision.

By applying a large scale data from Yelp.com, where users could establish friendships with others and look for businesses, this chapter use local popularity LP, which is the popularity of a business in the users' local neighbourhood of friends, and global popularity GP, which is a business's popularity in the whole system, to represent the friend- and crowd-based signal for the social influence. We find the friend-based social influence driving the users' selection behaviour significantly in comparison with the random experiments. Additionally, the local popularity of the selection behaviour $LP(c)$ follows the power-law distribution, which means the evolution of such system could be described by the local-driven preferential attachment mechanism. On the other hand, while the crowd-based social influence is significant when the local popularity is low, it vanishes as the local popularity increases. Thusly, the crowd-based social influence only plays a supplementary role in the dynamics, and drives the evolution only if there are not much local opinions.

It is worth to be noted that in Chapter 5 and Chapter 6, we study very similar research questions, that whether a user's friends or the crowds have stronger influence over his behaviour, i.e. posting reviews or selecting businesses. However, the two chapters are different in many ways. First of all, Chapter 5 takes econometric approaches to examine a set of different variables, while Chapter 6 mainly takes the approach of statistical experiments where we focus on only the variable of popularity with more controlled manner. In addition, while Chapter 5 is more explanatory via the significance of the variables in determining the posting behaviour of users, in Chapter 6 we try to find direct evidence of the intensity of friends' and crowds' influence over one's behaviour and establish a network model to uncover the mechanism of such behaviour. Despite the different approaches and objectives, the results of these two chapters confirm each other, that the friends are indeed more influential over ones' behaviour.

Chapter 7. Predicting the Future Increment of Review Helpfulness

Chapter 5 and Chapter 6 address the review impact and social influence on consume behaviours. However, such impact or influence have normally been studied collectively, i.e. it is assumed that the massive amount of online reviews as a whole is influencing subsequent consumers' behavior. On the individual level, is one review different from another in terms of such influence, is another widely discussed problem. This chapter discusses such question by studying what makes a review helpful, tracking RQ3 of the thesis.

To answer such question, studies have examined the various features that are associated with the review helpfulness. It has been found that, some quantifiable features such as review length (Zhou and Guo, 2017; Yin *et al.,* 2016; Salehan and Kim, 2016; Mudambi and Schuff, 2010), review rating (Wu, 2013; Huang *et al.,* 2015; Korfiatis *et al*., 2012) and review extremity (Liu and Park, 2015; Baek *et al.,* 2012) are highly correlated with the helpfulness. Writing styles such as lexical, grammatical, semantic and stylistic features of reviews have also been regarded as influential factors for their helpfulness (Kim *et al.,* 2006; Krishnamoorthy, 2015). In addition, reviewer-based characteristics such as reviewer reputation (Baek *et al.,* 2012; Liu and Park, 2015), reviewer connectedness (Huang *et al.,* 2015; Racherla and Friske, 2012) and reviewer profile image (Karimi and Wang, 2017) have been suggested to associate with helpfulness. With the understanding of these influential factors, studies have been trying to predict the review helpfulness accordingly (Chua and Banerjee, 2016). Methods such as support vector machine (Zhang and Varadarajan, 2006), multilayer perceptron neural networks (Lee and Choeh, 2014), hybrid model (Ngo-Ye and Sinha, 2014) have been developed for such prediction.

To the best of our knowledge, the studies of understanding and predicting review helpfulness, have been only focusing on a certain time point, which is normally the time when the data is collected. In common practices, a dataset of reviews and the associated features are collected at a given time point, denoting with t_0. These studies accordingly explore the correlation between the number of helpful votes at t_0 and the features, or predict the votes at t_0. However, such framework for studying the question may lead to endogeneity problem that the total votes at t_0 may have causal relations with explanatory variables observed at the same time point. For example, the widely discussed reviewer reputation, though has been argued to be very determinative for helpfulness of reviews, could also be resulted from the helpful reviews s/he has wrote. In addition, whether a review could get more votes, and how to predict the increment of the votes over a time period in the future is still an open question. The question is crucial to the understanding of reviews' future helpfulness and uncovering the potential of reviews, especially for the recent ones which do not have enough collected votes to fully reflect their helpfulness.

To tackle such question, this chapter adopts a dynamic method for data collection to study the increment of helpfulness for both old and recent reviews. We collect two-wave of review data for the same businesses from Yelp. Thus, we have the number of helpful votes of every review at two different time points, denoting with t_0 (8th January 2016, the date set described in chapter 4) and t_1 (9th July 2017). The time interval between the two-wave data is about one and a half year, which makes it possible to examine the increment of helpful votes. Regarding the t_0 as "current time" and t_1 as the "future time", the task is accordingly transformed as predicting the increment of a review's helpful votes from t_0 to t_1, using only the information at t_0. In this way, the endogeneity problem would be avoided. We study the old reviews posted long time ago, and the recent review that are newly posted respectively. We also evaluate the importance of six classes of widely-discussed indicators, namely reviewer activeness RA, reviewer historic votes RHV, reviewer credibility RC, review disclosure information RDI, review readability RR and review sentiment RS, that will be introduced in section 7.1.2, in predicting the increment of helpful votes. It is found that RDI and RA are the most important indicator classes for the prediction.

The rest of the chapter is organised as follows. Section 7.1 introduces the method of two-wave data collection and the measurement of variables relating to review helpfulness. Section 7.2 presents the research model and the results are shown in section 7.3. Section 7.4 summarizes and discusses the chapter.

7.1 Data Preparation and Variable Measurements

7.1.1 The Design of Study

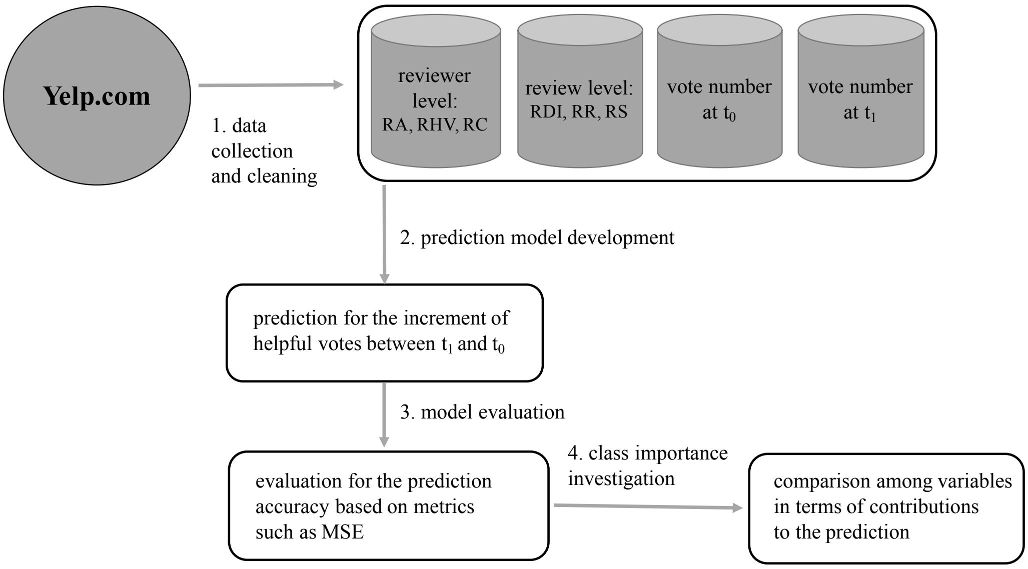

Yelp.com
1. data collection and cleaning
reviewer level: RA, RHV, RC
review level: RDI, RR, RS
vote number at t_0
vote number at t_1
2. prediction model development
prediction for the increment of helpful votes between t_1 and t_0
3. model evaluation
evaluation for the prediction accuracy based on metrics such as MSE
4. class importance investigation
comparison among variables in terms of contributions to the prediction

Figure 7.1|Research process in this chapter.

Figure 7.1 illustrates the research design of the chapter, which can be divided into four main steps: data collection and cleaning, prediction model development, prediction evaluation and sub-model development for the investigation of class importance. Firstly, the two-wave review data is collected from the website of Yelp. A total of 17 factors from both reviewer and review level, as independent variables, are extracted. Secondly, the relations between these factors and both review helpfulness at the current time and the increment of helpfulness in the future are explored respectively. The predictions are evaluated in the third step according to widely-used accuracy metrics. In the last, sub-models which exclude variable classes by turn are developed, to compare the contribution of each class to the prediction.

7.1.2 Two-Wave Dataset Preparation

The data used in this chapter is based on the review data published by Yelp which has been introduced in chapter 4. In Yelp, users can post reviews on various businesses, mostly restaurants, and vote others' reviews as "useful", "fun" and "cool". Since Yelp regularly updates their published data, we use two different rounds of data, in which the last dates of reviews are 8th January 2016 (denoting as t_0) and 09th July 2017 (denoting as t_1) respectively. Focusing on only the businesses with at least 100 reviews at t_0, we extract the reviews that appear in both waves of dataset. In this way, 443,702 reviews posted by 169,573 users on 1,737 businesses are extracted. While the information on the reviews such as review text and review rating, is identical in the two waves, the helpful votes of the reviews may increase over the period of gap time.

7.1.3 Variable Measurements

For each of the collected 443,702 reviews in our dataset, the features can be categorised as reviewer level and review level. For each level, we further classify the variables into three classes respectively (see Table 7.1).

On the reviewer level, we consider reviewer activeness (RA), reviewer historical votes (RHV) and reviewer credibility (RC). RA is measured by the number of reviews that a reviewer has posted previously, and the number of friends that s/he has. RHV is the historical votes that the reviewer has collected in all of his previous reviews, including the number of "useful" votes, "fun" votes and "cool" votes. We use the number of "Elite" and "compliments" a reviewer got as proxies for RC. Yelp prizes the outstanding reviewers who have great contributions with the title of "Elite" each year. The more "Elite" titles a reviewer has, the more credible s/he would be regarded. Additionally, there is also a "compliment" button that allows other users to send good verbs like "good writer", "cute pic" if they like the reviewer or the posted review. Therefore, the number of "compliments" could also reflect the credibility of reviewers.

On the review level, we consider the classes of review disclosure information (RDI), review readability (RR) and review sentiment (RS). *RDI* represents the most fundamental and straightforward information that consumers can get from a review such as the number of words ($WorCou$), numerical star (*Rating*), the difference between the rating and the average (*Extremity*) and released days (*Age*). These information forms the first impression

Table 7.1|Statistical descriptions for the applied variables.

Vari. Cate.	**Variable Class**	**Variable Name**	**Description**	**Mean**	**Min**	**Max**
Rev. Lev.	*Reviewer Activeness*	*Rev_num*	number of reviews the reviewer posted;	150.75	1	8843
		Fri_num	number of friends that the reviewer has;	49.77	0	3830
	Reviewer Historical Votes	*Hel_num*	number of useful vote the reviewer got in the history;	405.59	0	36474
		Fun_num	number of fun vote the reviewer got in the history;	221.66	0	32747
		Cool_num	number of cool vote the reviewer got in the history;	270.19	0	32517
	Reviewer Credibility	*Elite*	number of "Elite" titles that the reviewer got;	1.2	0	11
		Compliment	number of compliments the reviewer got;	220.06	0	175944
Rev. Lev.	*Review Disclosure Information*	*WorCou*	number of words	139.29	2	1580
		Rating	star (1 to 5) of the review;	3.83	1	5
		Extremity	the absolute difference between the rating and the average rating;	1.25	0	14.95
		Age	days since the review was posted	3150	286	4026
	Review Readability	*FRE*	Flesch Reading Ease	78	0.75	99.91
		SMOG	SMOG index	3	0	13
		GF	Gunning-Fox index	7.1	2.8	16.8
		ARI	Automated Readability Index	6.34	1.1	13.9
	Review Sentiment	*Subjectivity*	the extent of subjectivity of the review, where 0 is total objective and 1 is total subjective;	0.13	0	1
		Polarity	the extent of polarity of the review, where 0 is natural, 1 is total positive polarity and -1 is total negative polarity	0.21	-1	1
Vot. Info.	*Vote number at t_0*	$H(t_0)$	helpful votes the review gets by t_0;	1.13	0	166
	Vote number at t_1	$H(t_1)$	helpful votes the review gets by t_1;	1.21	0	168
	Increment of helpfulness	ΔH	the increasing number of helpful votes from t_0 to t_1	0.09	0	67

of a merchandise for consumers. To examine the readability of reviews, some of the commonly used metrics (DuBay, 2004) including *Gunning-Fox Index (GF)*, the *Automated*

Readability Index (ARI), *Flesch Reading Ease (FRE)* and *SMOG* are applied, which are defined as follows,

$$FRE = 206.835 - 1.015 \times \frac{TotalWords}{TotalSentences} - 84.6 \times \frac{TotalSyllables}{TotalWords}\ , \quad (7.1)$$

$$GF = 0.4 \times {}^{Words}/_{Sentences} + 100 \times {}^{ComplexWords}/_{Words'} \quad (7.2)$$

$$ARI = 4.71 \times {}^{Characters}/_{Words} + 0.5 \times {}^{Words}/_{Sentence} - 21.43, \quad (7.3)$$

$$SMOG = 1.043 \times \sqrt{PolysyllablesNumber} \times \sqrt{30/SentenceNumber} + 3.1291. \quad (7.4)$$

These metrics depend on parameters such as number of syllables per word, number of words per sentence and number of characters per word. RS refers to the emotional attitude of the review, which is determined by the implied meaning of the writing. We use the *Subjectivity* and *Polarity* to characterise the RS which are calculated by a Python package *"TextBlob"*. The subjectivity value describes to what extent is a piece of text objective or subjective. On the other hand, polarity describes the writers' emotions expressed through the text that whether s/he is positive or negative about the matter being talked about.

For the review helpfulness, we follow the previous studies (Zhou and Guo, 2017) and operationalise the dependent variable as the number of "useful" votes that a review receives, denoting with H. Since we have a two-wave data, the numbers of useful votes for each review at both time points are available, denoting with $H(t_0)$ and $H(t_1)$ respectively. The increment of the useful votes is thus $\Delta H = H(t_1) - H(t_0)$.

Note that, all the independent variables are max-min normalised, while the dependent variable ΔH is not, so that the prediction results are straightforward to be evaluated. Since we have a number of 17 predictor variables, the caused multicollinearity concerns us due to the high correlation coefficients between variables such as more than 0.9 between Fun_num, $Cool\_num$ and Hel_num, as shown in Table 7.2. Thus, we remove Fun_num, $Cool\_num$ in the predicted model for several reasons. On the one hand, the effect of multicollinearity could be reduced, that variance inflation factor (VIF) is 1.56 after removing them. According to Dielman (2001), VIF score of less than 10 suggests that multicollinearity will not significantly influence the stability of the parameter estimates. On the other hand, the work focus on helpful vote of reviews, thus the other type of vote number that reviewers get (“fun” or “cool”) seems to have less influence on the objectives. Therefore, the two variables will not be considered in the following.

7.2 Research Model Development and Evaluation Method

7.2.1 Prediction Model

Since the dependent variable is count number, that is, nonnegative integer values. For count data, the most widely used regression model is Poisson regression (Sellers & Shmueli, 2010). In this chapter, we use variance stabilising transformation to avoid heteroscedasticity by adopting Poisson Generalised Linear Mixed Model (GLMM), based on the decision tree of GLMM fitting and interface in the Bolker *et al.* (2009). GLMMs combine the properties of two statistical frameworks that are linear mixed models (with random effects) and generalized linear models (which handle nonmoral data) by using link functions. Because multiple reviews may come from the same business, to avoid the different influence of businesses on the dependent variable, we therefore add one random effect for businesses in the model as well. The full model is as follows,

$$\begin{aligned}\log(\Delta \mathrm{H}_{ij}) = b_0 + b_{1i} \cdot Rev_{num} + b_{2i} \cdot Fri_{num} + b_{3i} \cdot Hel_{num} + b_{4i} \cdot Compliment + \\ b_{5i} \cdot Elite + b_6 \cdot WorCou + \ b_7 \cdot Rating + \ b_8 \cdot Extremity + \\ b_9 \cdot Age + \ b_{10} \cdot FRE + b_{11} \cdot SMOG + \ b_{12} \cdot GF + \ b_{13} \cdot ARI + \\ b_{14} \cdot Subjectivity + b_{15} \cdot Polarity + a_j + \varepsilon,\end{aligned} \tag{7.5}$$

where a_j represents the random effect for business j.

To make the variables comparable to each other, we normalise the scale of them before we run the model. For each prediction, we divide randomly 80% reviews as the training set, and the remaining 20% as the testing set. Based on the training set, we firstly run a regression to explore the influence of the factors according to the model. By doing so, the estimated coefficient of each variable, $\widehat{\delta_0}$, $\widehat{\delta_1}$, $\widehat{\delta_2}$...$\widehat{\delta_{15}}$ can be obtained. These estimated coefficients can be brought back into the model to further predict the increment of the helpful votes $\widehat{\Delta H}$ of each review in the testing set.

Table 7.2| Correlation matrix between variables.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Rev_num	1																
2. Fri_num	0.49	1															
3. Hel_num	0.73	0.66	1														
4. Fun_num	0.61	0.58	0.95	1													
5. Cool_num	0.66	0.64	0.99	0.97	1												
6. Compliment	0.47	0.5	0.85	0.85	0.87	1											
7. Elite	0.69	0.4	0.48	0.4	0.43	0.31	1										
8. WorCou	0.16	0.13	0.17	0.15	0.15	0.14	0.23	1									
9. Rating	-0.01	0.01	0	0	0	0	0.01	-0.14	1								
10. Extremity	-0.1	-0.06	-0.06	-0.05	-0.06	-0.04	-0.12	0.07	-0.57	1							
11. Age	-0.21	-0.11	-0.14	-0.12	-0.13	-0.09	-0.25	-0.11	0.01	0.07	1						
12. FRE	-0.01	-0.02	-0.03	-0.03	-0.03	-0.03	-0.02	-0.04	-0.09	0.04	-0.01	1					
13. SMOG	0.05	0.03	0.04	0.03	0.03	0.03	0.08	0.16	-0.03	0.01	-0.04	-0.11	1				
14. GF	0.09	0.06	0.08	0.06	0.07	0.05	0.13	0.37	-0.12	0.04	-0.1	-0.22	0.14	1			
15. ARI	0.08	0.06	0.09	0.08	0.08	0.07	0.1	0.23	-0.01	-0.01	-0.07	-0.72	0.12	0.67	1		
16. Subjectivity	-0.05	-0.04	-0.04	-0.04	-0.04	-0.03	-0.06	-0.16	0.15	-0.08	0.05	-0.1	-0.05	-0.15	-0.02	1	
17. Polarity	-0.07	-0.04	-0.05	-0.05	-0.05	-0.03	-0.08	-0.24	0.52	-0.32	0.06	-0.05	-0.13	-0.22	-0.08	0.33	1

7.2.2 Model Evaluation

To examine whether, or to what extent are the estimated increments $\widehat{\Delta H}$ accurate, we apply five widely-used metrics for the evaluation. The metrics are Mean Standard Error (MSE), Root Mean Standard Error (RMSE), Mean Absolute Error (MAE), Relative absolute (RAE) and Root Relative Squared Error (RRSE) which are defined as,

$$MSE = 1/n \sum_{i=1}^{n} \left(\Delta H_i - \widehat{\Delta H_\imath}\right)^2, \tag{7.6}$$

$$RMSE = \sqrt{1/n \sum_{i=1}^{n} \left(\Delta H_i - \widehat{\Delta H_\imath}\right)^2}, \tag{7.7}$$

$$MAE = 1/n \sum_{i=1}^{n} \left|\Delta H_i - \widehat{\Delta H_\imath}\right|, \tag{7.8}$$

$$RAE = \sum_{i=1}^{n} \left|\Delta H_i - \widehat{\Delta H_\imath}\right| \Big/ \sum_{i=1}^{n} \left|\Delta H_i - \overline{\Delta H_i}\right|, \tag{7.9}$$

and

$$RRSE = \sqrt{\sum_{i=1}^{n} \left(\Delta H_i - \widehat{\Delta H_\imath}\right)^2 \Big/ \sum_{i=1}^{n} \left(\Delta H_i - \overline{\Delta H_i}\right)^2}\ , \tag{7.10}$$

respectively, where i represents the review ID, n is the total number of reviews in the testing set, and $\widehat{\Delta H_\imath}$ and ΔH_i are the estimated value and actual value of the increment of helpful votes respectively. For all of the metrics, smaller value represents better accuracy.

7.2.3 Class Importance

To further investigate the importance of each class of variables in predicting the increment of review helpfulness, six sub-models are developed, each of which removes one class of factors. For example, when examining the importance of the class $RA,$ we apply the following model,

$$\begin{aligned}\log\left(\Delta \mathrm{H}_{ij}\right) = b_0 + b_{3i} \cdot Hel_{num} + b_{4i} \cdot Compliment + b_{5i} \cdot Elite + \\ b_6 \cdot WorCou + b_7 \cdot Rating + b_8 \cdot Extremity + \\ b_9 \cdot Age + b_{10} \cdot FRE + b_{11} \cdot SMOG + b_{12} \cdot GF + \\ b_{13} \cdot ARI + b_{14} \cdot Subjectivity + b_{15} \cdot Polarity + a_j + \varepsilon,\end{aligned} \tag{7.11}$$

which removes the factors of Rev_num and Fri_num. Following the same procedure, we then have the accuracy of prediction without the consideration of RA. According to an arbitrary accuracy measure, e.g. *MSE*, of the full prediction model, and the RA-removed prediction model, the importance of the class RA is thereby defined as (MSE_{RA} − *MSE*)/*MSE*. Therefore, the importance of the RA describes by how much contribution can RA add to the prediction accuracy. Similarly, we examine the importance of the rest of five classes.

7.3 Analytical Results

7.3.1 Regression Results

All the independent variables in this chapter are selected with respect to the previous studies on the current review helpfulness, i.e. $H(t_0)$. We therefore firstly examine the relations between these variables and the current review helpfulness according to the model

$$\begin{aligned}\log\bigl(H(t_0)\bigr) = {} & b_0 + b_{1i} \cdot Rev_{num} + b_{2i} \cdot Fri_{num} + b_{3i} \cdot Hel_{num} + \\ & b_{4i} \cdot Compliment + b_{5i} \cdot Elite + b_6 \cdot WorCou + \\ & b_7 \cdot Rating + b_8 \cdot Extremity + \; b_9 \cdot Age + \; b_{10} \cdot FRE + \\ & b_{11} \cdot SMOG + \; b_{12} \cdot GF + b_{13} \cdot ARI + \; b_{14} \cdot Subjectivity + \\ & b_{15} \cdot Polarity + a_j + \varepsilon,\end{aligned} \tag{7.12}$$

Furthermore, we regard the reviews that are posted more than 100 days prior to t_0 as the old reviews, and others as recent reviews, to explore the different influence of the variables. Table 7.3 reports the regression results for all reviews, old reviews and recent reviews respectively, which basically confirm the observations of the previous studies. First of all, the table displays some measures of fit over the models such as BIC, AIC and R^2. Simply from the value of AIC and BIC, the model 3 gets the lowest value among the three models. The random effect shows the business level has a variance of 0.13 for old reviews, 0.48 for recent reviews. In addition with the results of LR test, there is indeed enough variability between businesses in determining the helpfulness. We get the marginal and conditional R^2 of the mixed effect models developed by Nakagawa and Schielzeth (2013). The results are shown in the tables, where Model 1 and 2 get the similar values. The model diagnostics can be found in the Appendix (Figure A.2).

For the reviewer level, reviewer credibility, historical votes and friend number all have significant positive impact on the review helpfulness in the three models. The results suggest that reviews that are posted by reviewers who have more honorary titles, compliments, or connections with other reviewers, get more helpful votes. But the variable

Table 7.3| Regression results for review helpfulness at the current time $H(t_0)$.

			Model 1: **All reviews**		**Model 2:** **Old reviews**		**Model 3:** **Recent reviews**	
Cate.	**Class**	**Name**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
		Intercept	-0.446***	0.029	-0.377***	0.030	-0.929***	0.127
Re.er Lev.	RA	Rev_num	-0.201	0.452	-0.216	0.439	2.904***	0.306
		Fri_num	2.186***	0.021	2.112***	0.021	3.403***	0.110
	RHV	Hel_num	3.010***	0.035	2.997***	0.035	2.728***	0.256
	RC	$Compliment$	2.932***	0.239	2.922***	0.239	2.724***	0.549
		$Elite$	1.384***	0.037	1.364***	0.037	1.117***	0.156
Rev Lev.	RDI	$WorCou$	1.525***	0.334	1.507***	0.335	2.653***	0.782
		$Rating$	0.013***	0.003	0.022***	0.003	-0.093***	0.017
		$Extremity$	0.117***	0.011	0.171***	0.011	-0.189***	0.051
		Age	-0.679***	0.064	-0.444***	0.065	-0.127*	0.065
	RR	FRE	1.376***	0.026	1.341***	0.027	1.094***	0.119
		$SMOG$	0.773***	0.014	0.772***	0.015	0.672***	0.058
		GF	0.017	0.019	-0.009	0.019	0.364***	0.090
		ARI	1.475***	0.018	1.448***	0.018	1.464***	0.092
	RS	$Subjectivity$	-1.877***	0.018	-0.445***	0.013	-0.735***	0.060
		$Polarity$	-0.465***	0.013	-1.873***	0.018	-1.928***	0.083
Random effect			Variance (Std. Dev)		Variance (Std. Dev)		Variance (Std. Dev)	
Business			0.13 (0.36)		0.13 (0.364)		0.48 (0.693)	
LR test								
Chibar2			12993		12454		1349	
Prob>=chibar2			0.00		0.00		0.00	
Observation			443702		410030		33672	
AIC			1496655.4		1405970.8		77474.1	
BIC			1496842.4		1406156.5		77617.3	
Marginal R^2			0.101		0.101		0.068	
Conditional R^2			0.215		0.221		0.262	

*$p<0.05$, **$p<0.01$, ***$p<0.001$.*

Rev_num in the class of RA only showing significant positive impact for new reviews in model 3, has no impact in the other two models, which indicates that these commonly-discussed factors may have different influence for reviews with different age.

For the review level, the disclosure information such as review length ($WorCou$), has positive impact on review helpfulness. Users are more likely to vote reviews that have more words as useful. While review age negatively influence the review helpfulness, that consumers prefer to vote the recent reviews. Review rating and extremity positively associate with the helpfulness. But differences can be observed when we separately analyse old and resect reviews. For recent reviews, those that have low ratings and stable rating (less extremity) are more likely to be regarded as helpful. However, for old reviews, review helpfulness are probably promoted by high and controversial ratings. Such results has seldom been observed in previous studies as those studies do not group reviews in terms of released days. The sentiment features of reviews show negative influence (*Subjectivity<0, Polarity<0*), which indicates that reviews that are less subjective and less polarity are more likely to get more helpful votes. Readability measured by *FRE*, *SMOG* and *ARI* shows positive impact in the three models.

We further explore whether these variables are able to explain the increment of the helpful votes, i.e. ΔH. According to Eq. (7.5), the model 4, 5 and 6 are reported in Table 7.4, addressing the regression for all reviews, old reviews and recent reviews respectively. The results of model fit are similar to that of table 7.3. The increment helpfulness of recent reviews gets smaller AIC and BIC. The random effect and LR test suggest that the businesses have variability in determining the increment of helpfulness. The model diagnostics can be found in the Appendix (Figure A.3).

In comparison to the results for the current helpfulness, most variables, including $Fri\_num,\ Hel\_num,\ Compliment,\ Extremity,\ FRE,\ SMOG,\ ARI,\ Polarity\ and\ Subjectivity$, show similar influences. However, the variables $Elite,\ WorCou,\ Rating$ and Age show different behaviours. Despite the significant influence on the current helpful votes $H(t_0),\ WorCou$ does not significantly influence the increment of the helpful votes ΔH for all three models, and $Elite$ only has significant influence for new reviews.

While Age shows negative significance describing the current helpful votes of all reviews and old reviews, it influences the increment of helpful votes in the future significantly and positively for all kinds of reviews. Reviews with lower ratings are shown to be more likely to get more increment of helpful votes in the future ($Rating < 0$ in model 4, 5 and 6), but reviews that get the most votes are those with higher ratings ($Rating > 0$ in model 1, 2).

Table 7.4|Regression results for the increment of review helpfulness ΔH.

			Model 4 **All reviews**		**Model 5** **Old reviews**		**Model 6:** **Recent reviews**	
Cate.	**Class**	**Name**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
		Intercept	-2.23***	0.10	-2.69***	0.12	-1.28***	0.18
Rev. Lev.	RA	Rev_num	-8.28	98.18	-8.54	11.74	1.25*	0.62
		Fri_num	3.59***	0.04	3.84***	0.04	3.50***	0.22
	RHV	Hel_num	2.42***	0.19	2.65***	0.19	0.05	1.05
	RC	$Compliment$	3.22***	0.72	3.51***	0.72	-5.26	39.19
		$Elite$	-0.01	0.29	0.22	0.29	1.72***	0.22
Review Level	RDI	$WorCou$	-7.59	151.67	-7.25	6.28	1.68	0.91
		$Rating$	-0.05***	0.01	-0.04**	0.01	-0.14***	0.02
		$Extremity$	1.60***	0.03	1.61***	0.04	0.81***	0.06
		Age	1.74***	0.06	1.00***	0.12	0.61***	0.07
	RR	FRE	0.25**	0.09	0.32**	0.11	0.87***	0.17
		$SMOG$	0.66***	0.05	0.68***	0.06	0.65***	0.08
		GF	0.46***	0.07	0.36***	0.08	0.81***	0.13
		ARI	0.42***	0.07	0.51***	0.08	0.80***	0.13
	RS	$Subjectivity$	-1.54***	0.06	-0.31***	0.05	-0.65***	0.08
		$Polarity$	-0.35***	0.04	-1.39***	0.07	-1.80***	0.11
Random effect			Variance (Std. Dev)		Variance (Std. Dev)		Variance (Std. Dev)	
Business			0.46(0.68)		0.51(0.71)		0.44(0.66)	
LR test								
Chibar2			10274		8962		695.8	
Prob>=chibar2			0.00		0.00		0.00	
Observation			443702		410030		33672	
AIC			258575		196558.3		49119.1	
BIC			258762		196744		49262.4	
Marginal R^2			0.013		0.009		0.058	
Conditional R^2			0.057		0.047		0.186	

**p<0.05, **p<0.01, ***p<0.001.*

In addition, when distinguishing the old and recent reviews, the Hel_num and

$Compliment$ of reviewers, though significant for old reviews, become insignificant for explaining recent reviews' increment of helpful votes. In summary, the widely-used factors to describe the current helpfulness are generally applicable for the increment of helpfulness in the future, but some may fail to explain the helpfulness increment of new reviews.

7.3.2 Evaluation Results

Based on the regression results from Table 7.4, we take the estimated coefficients to apply to the reviews in the testing set, and accordingly the increment of the helpful votes can be predicted. We carry out such predictions for all reviews, old reviews and recent reviews respectively, based on 10-fold cross-validation. Cross-validation entails a set of techniques that partition the dataset and repeatedly generate models and test their future predictive (Browne, 2000). In the K-fold cross-validation, the entire data is typically divided into K smaller observations. K-1 observations are used to generated to train a model. The validity and generalizability of the generated models is then tested on the kth observation. A value of K=10 is generally used as a rule of thumb for the number of folds. This method avoids the randomness emanating from estimates produced by splitting the data only once, and also can help making replication of the study (Koul *et al.,* 2018).

Table 7.5|Predication accuracy for all reviews, old reviews and recent reviews, respectively.

	All reviews	Old reviews	Recent reviews
MSE	0.164	0.157	0.668
RMSE	0.405	0.396	0.817
MAE	0.165	0.129	0.433
RAE	1.007	0.991	0.852
RRSE	0.958	0.969	0.931

The prediction accuracies based on the metrics introduced in section 7.2.2 are reported in Table 7.5. The accuracy of predicting old reviews is generally better than that for all reviews, while the accuracy for the recent reviews is less accurate. Taking the *MSE* as an example, the average error of the prediction for all reviews is 0.164. When predicting for the old reviews, the error decreases to 0.157. However, for the recent reviews, the error between predicted and the actual increment is more than four times bigger. The reason lies in the fact that old reviews are posted long time ago and their helpful votes have generally reached a stable level, leading to more precise regressions. On the other hand, recent reviews do not have sufficient time to collect enough votes to reflect their actual helpfulness, and as a results, the increment of such reviews may be much more fluctuated.

Moreover, we investigate the importance of the classes in determining the increment of helpfulness. As introduced in the section 7.2.3, we remove each class of factors by turn from the full model and examine the accuracies of the sub-models. Table 7.6 shows the results of the *MSE* of each prediction for all reviews, old reviews and recent reviews respectively. In general, the removal of any factor class will decrease the accuracy of the prediction, which means every class can contribute to the prediction. However, the reduction of accuracy when removing different factor class is of different degree. For example, when predicting for all reviews, the full model has *MSE*=0.164. With the class RA being removed, the accuracy becomes MSE_{RA}=0.170, while removing RHV makes the accuracy as MSE_{RHV}=0.165. Accordingly, one can regard RA as more important than RHV in the prediction because RA contributes more to the accuracy.

Table 7.6| Predication accuracy of sub-models for all reviews, old reviews and recent reviews, respectively.

	removal	All reviews	Old reviews	Recent reviews
MSE	None	0.164	0.157	0.668
MSE_{RA}	RA	0.17	0.162	0.682
MSE_{RHV}	RHV	0.165	0.158	0.670
MSE_{RC}	RC	0.164	0.157	0.674
MSE_{RDI}	RDI	0.17	0.16	0.705

To closely compare the importance of different classes, we measure the ratio of the change when removing each class as introduced in section 7.2.3. Figure 7.2 shows the class importance for predicting the increment of helpfulness for all reviews, old reviews and new reviews. Basically, the most important classes are RDI and RA, and the least important ones are RR and RS for all three models. While factors such as readability, sentiment expressions would take users some efforts to notice, the profile of reviewers and the fundamental factors of reviews such as review length and rating can make much more direct impression on users. As argued by the theory of *principle of least e*ffort (Bohner et al., 1994), people may be expected to conserve their cognitive resources and rely on heuristic cues, unless they have a strong motivation to process more nuanced information. In the online environment, both word count and rating are heuristic cues that can be gained without much efforts. Therefore, these factors play more influential roles in the process of perceived helpfulness of information for consumers. The historical votes of reviewers contribute to the prediction of helpfulness to some extent, as the value of RHV importance is moderate.

Some dramatic differences of class importance are suggested from the results, when predicting the increment of helpfulness of old reviews and recent reviews separately. RA and RDI do not show obvious difference of importance in predicting the helpfulness increment of old reviews. But the importance of the RDI for recent reviews is significantly larger than that of RA. Also, the value of RDI for recent reviews is the highest. Notably, though RC has basically no contributions to the prediction for all reviews and old reviews, it plays substantial role for recent reviews.

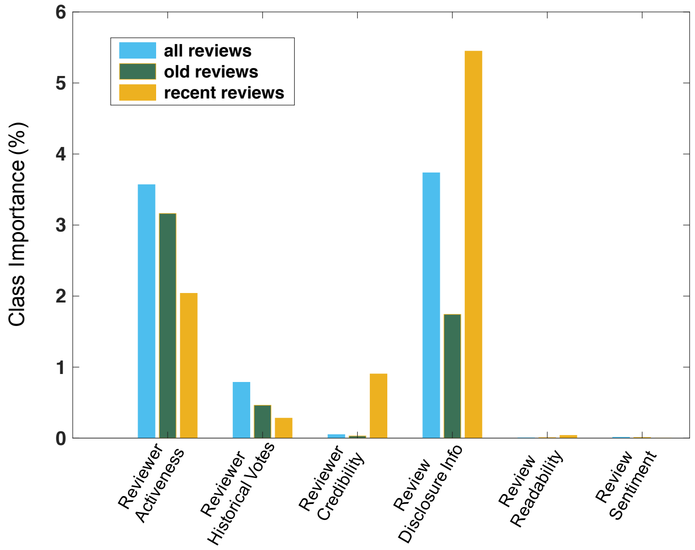


Figure 7.2| Importance of classes for predicting the increment of helpfulness for all reviews, old reviews and recent reviews.

7.4 Summary

This chapter investigates the performance of classical factors such as consumer characteristics, review textual factors in predicting the increment of helpfulness in the future. Instead of studying the review data of a static time point, we collect two-wave data at two different time points and focus on the change of helpful votes over the time period. It is found that the classical factors are able to significantly describe the increment of the helpfulness in general, but some factors such as review readability and sentiment show no significance for describing recent reviews. These factors are found with moderate accuracy when predicting the helpfulness increment in the future, and it is much harder to predict the recent reviews' helpfulness change. The investigation of the class importance indicates that, the most straightforward factors, such as the activeness of reviewers, the length and rating of reviews, contribute largely to the prediction of the helpfulness increment. In addition, review readability and sentiment do not significantly improve the accuracy of the prediction.

Chapter 8. Exploring Distance Impact on eWOM in Product Recommendation Network

The popularity of ecommerce services over the recent decades leads to the emergence of product recommendation networks (PRNs) where similar products are connected to each other by hyperlinks. Previous research find that the PRN significantly influence demands and sales of products. The reason may lie in the fact that the presence of PRNs shortened the distance between products, leading to the higher probability of joint purchase. That inspired us to think that whether the PRN could also influence the eWOM of products with each other (RQ4 of the thesis), since the connected products have similar attributions, which has not been explored in the literature.

With such a bold surmise, this chapter employs an empirical PRN collected from Amazon along with the eWOM information of every product (e.g. book) to explore the impact of product distance on their eWOM on two levels: 1) neighbourhood level and 2) dyadic product-pair level. On the neighbourhood level, regarding all books, which have shortest distance of n to a focal book, as the focal book's nth-order neighbours, we analyse the impact of the eWOM of the neighbourhood on the focal book's eWOM in terms of review volume and valance. On the dyadic product-pair level, we define the connectivity between two specific books as the number of paths from one book to another with the length of one, two and three respectively, and investigate the influence of connectivity on eWOM rating difference of the measured books. The results show that the products that are close to each other indeed have similar eWOM in the recommendation network. Specifically, eWOM of focal books is largely related to the neighbours' eWOM, and such impact can reach three clicks away. And on the product-pair level, not only the direct connection between products, but also the indirect connections could make two products having similar ratings.

The chapter is designed as follows, section 8.1 shows the network structure among products in Amazon recommendation system and introduces some terms including focal book, ordered neighbour and connectivity. Section 8.2 and section 8.3 present the analytics with two methods addressing the impact of product distance on eWOM. And the summary of this chapter is in section 8.4.

8.1 Product Recommendation Network on Amazon

As descried in section 4.2, the data that we are using in this chapter is the book recommendation network of Amazon. The data of recommendation network was collected on 1st January 2016. Considering that such recommendation network might be regularly updated, in this chapter we mainly focus on the book reviews posted during the period of 1st to 31st December 2015, and assume that the structure of the recommendation network in this month remains unchanged. The selected period includes 34,100 books that have at least one review and have been recommended at least once, and thus a total of 92,405 reviews are associated.

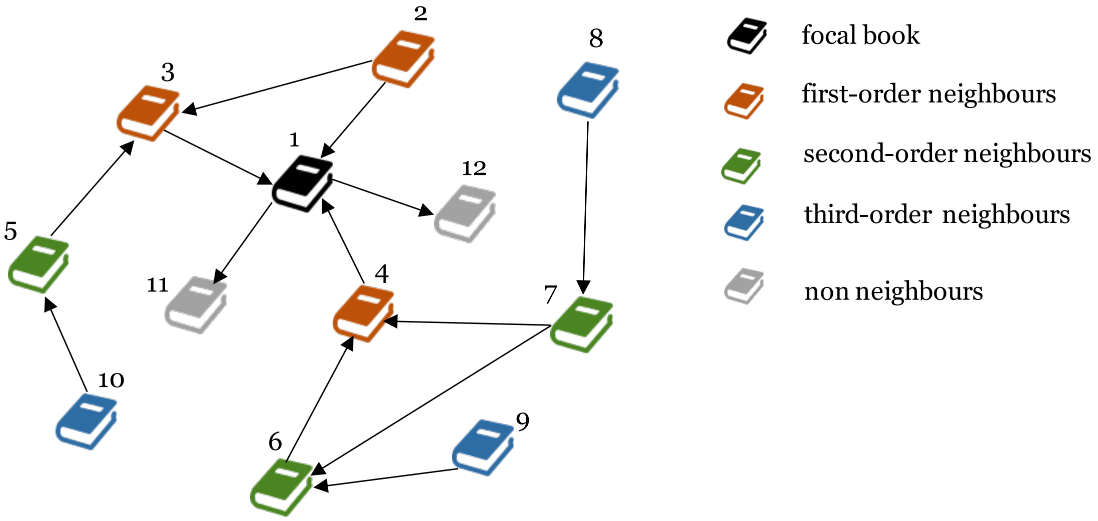


Figure 8.1|An illustration for Amazon book recommendation network.

The books are connected with each other via directed hyperlinks based on the recommendation of Amazon, as shown in Figure 8.1. Between every pair of books, there would normally be a certain distance, which represents how many clicks at minimum it takes a user starting from a book to visit another book. For example, book 3 needs one click to reach book 1, while book 10 needs three clicks to reach book 1. For a focal book, we define other books as its nth-order neighbours if this book has a distance of n to reach the focal book. Accordingly, as shown in Figure 8.1, taking the book 1 as the focal book, books 2, 3, 4 are its first-order neighbours; books 5, 6, 7 are its second-order neighbours; and books 8, 9, 10 are its third-order neighbours. Note that, we only consider books from which a user can reach the focal book as the neighbours, while ignore these books that can only be reached by the focal book, i.e. non-neighbours for the focal book. Accordingly, book 11 and 12 will not be considered as the focal book's neighbours.

We explore the impact of product distance on eWOM from two level: neighbourhood level and dyadic level, which will be introduced in the section 8.2 and section 8.3 respectively. On the neighbourhood level, we analyse the impact of the first-, second- and third-order neighbours' eWOM on the focal books' eWOM in terms of review volume and valance. On the dyadic product-pair level, we explore the number of paths from one book to another with length of one, two and three respectively, denoting as the connectivity, and investigate the influence of connectivity on the rating difference of the measured books.

8.2 Neighbourhood Level analysis

8.2.1 Variable Operationalisation for Neighbourhood Level Analysis

Dependent variables

The dependent variables in this section are the eWOM information of the focal books. We operationalise the eWOM with two most direct measurements known as review volume and average rating (valance). Accordingly, the dependent variables are the number of reviews at day t, denoting with $RV_i(t)$, and the average rating at day t, denoting with $AVE\_Rating_i(t)$, for each focal book i.

Independent variables and control variables

The independent and control variables are basically the eWOM information of the focal books' neighbours which are one, two and three clicks away in the PRN. Accordingly, there are three categories of variables namely the first-order neighbours, second-order neighbours and third-order neighbours, denoting with (1), (2) and (3) respectively.

Considering the connectivity of focal books largely varies from each other, we take the number of neighbours at each order as control variables. Hence, $In\_deg_i^{(1)}$, $In\_deg_i^{(2)}$

and $In_deg_i^{(3)}$represent the number of first-order neighbours, second-order neighbours and third-order neighbours of the focal book i respectively. In addition, we also control the analysis with the historical number of reviews of the focal book. To be specific, for a focal book i, the number of reviews that are posted before 1st December 2015, denoting with NOR_i is also a control variable. The control variables do not change over time, as we have assumed that the structure of recommendation network in the studied time period is fixed.

We consider daily review volume and daily average rating as the independent variables. For a neighbour book j, its review volume and average rating at the day t are denoted with $V_j(t)$ and $R_j(t)$. Note that, not every neighbour book has reviews every day in the studied period. If a neighbour book j does not have reviews at the day t, we regard its volume as $V_j(t) = 0$. But for the rating, a value of 0 would suggest an extremely low rating. Therefore, we regard the average rating of such books as the system average of all reviews $R_j(t) = 4.299$ so that it represents a neutral rating. In this way, for a focal book i, there are two independent variables for an order (n), namely the average review volume,

$$Ave\_RV_i^{(n)}(t) = \frac{1}{In\_deg_i^{(n)}} \sum_{j \in \mathbf{\Gamma}_i^{(n)}} V_j(t), \tag{8.1}$$

and the average rating,

$$Ave\_Rating_i^{(n)}(t) = \frac{1}{In\_deg_i^{(n)}} \sum_{j \in \mathbf{\Gamma}_i^{(n)}} R_j(t), \tag{8.2}$$

respectively, where $\mathbf{\Gamma}_i^{(n)}$ is the set of books that are the focal book i's nth order neighbours with a population of $In\_deg_i^{(n)}$. Accordingly, we have the independent variables $Ave\_RV_i^{(n)}(t)$ and $Ave\_Rating_i^{(n)}(t)$, with $n = 1, 2, 3$.

The operationalisation and descriptive statistics of all variables are shown in Table 8.1, while the Table 8.2 reports the correlations among these variables, in which the correlations between $In\_deg_i^{(1)}, In\_deg_i^{(2)}, In\_deg_i^{(3)}$ are significant and high, as shown in marked red.

Table 8.1|Description statistics of variables for neighbourhood level.

Variable Type	Variable Name	Descriptions	Mean	Std. Dev.	Min	Max
Dep. Var.	$RV_i(t)$	review volume of the focal book i at day t	1.205	0.996	1	107
	$AVE\_Rating_i(t)$	average rating of the focal book i at day t	4.449	1.043	1	5
Indep. Var.	$Ave\_RV_i^{(1)}(t)$	average review volume of first-order neighbours of the focal book i	0.073	0.150	0	7.1
	$Ave\_RV_i^{(2)}(t)$	average review volume of second-order neighbours of the focal book i	0.065	0.076	0	3
	$Ave\_RV_i^{(3)}(t)$	average review volume of third-order neighbours of the focal book i	0.028	0.035	0	2.22
	$Ave\_Rating_i^{(1)}(t)$	average review rating of first-order neighbours of the focal book i	4.428	0.083	1	5
	$Ave\_Rating_i^{(2)}(t)$	average review rating of second-order neighbours of the focal book i	4.423	0.061	2.71	5
	$Ave\_Rating_i^{(3)}(t)$	average review rating of third-order neighbours of the focal book i	4.422	0.016	3.74	4.81
Cont. Var.	NOR_i	number of historical reviews of focal book i	188.92	376.5	1	`5263
	$In\_deg_i^{(1)}$	number of first-order neighbours of the focal book i	29.601	51.02	1	911
	$In\_deg_i^{(2)}$	number of second-order neighbours of the focal book i	198.18	478.5	1	10204
	$In\_deg_i^{(3)}$	number of third-order neighbours of the focal book i	823.37	1842.7	1	26414

Table 8.2|Correlations among variables for neighbourhood level.

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1. $RV_i(t)$	1											
2. $AVE\_Rating_i(t)$	0.02	1										
3. $Ave\_RV_i^{(1)}(t)$	0.17	0.04	1									
4. $Ave\_RV_i^{(2)}(t)$	0.14	0.01	0.28	1								
5. $Ave\_RV_i^{(3)}(t)$	0.05	0.00	0.20	0.33	1							
6. $Ave\_Rating_i^{(1)}(t)$	0.00	0.12	0.16	0.06	0.01	1						
7. $Ave\_Rating_i^{(2)}(t)$	0.00	0.50	0.06	0.08	-0.01	0.14	1					
8. $Ave\_Rating_i^{(3)}(t)$	0.00	0.05	0.02	0.04	0.08	0.03	0.04	1				
9. NOR_i	0.00	0.02	0.16	0.00	0.06	0.00	0.01	0.01	1			
10. $In\_deg_i^{(1)}$	0.01	0.01	0.06	-0.13	-0.02	-0.02	-0.0	-0.0	0.47	1		
11. $In\_deg_i^{(2)}$	0.01	0.00	0.11	-0.09	-0.02	-0.02	-0.01	-0.01	0.44	0.92	1	
12. $In\_deg_i^{(3)}$	0.01	-0.01	0.12	-0.08	-0.01	-0.03	-0.01	-0.01	0.45	0.83	0.95	1

8.2.2 Model Specification for Neighbourhood Level Analysis

For the analysis of the neighbourhood level, we develop two regression models to explore how the eWOM of neighbours that are up to three clicks away, impact the focal books' eWOM. The following two models have the same equation considering the same variables, but they are analysed to the different depended variables. The detailed models can be described as,

$$\begin{aligned} RV_i(t) = b_0 + b_1 NOR_i + b_2 In\_deg_i^{(1)} + b_3 In\_deg_i^{(2)} + \\ b_4 In\_deg_i^{(3)} + b_5 Ave\_RV_i^{(1)}(t) + b_6 Ave\_Rating_i^{(1)}(t) \\ + b_7 Ave\_RV_i^{(2)}(t) + b_8 Ave\_Rating_i^{(2)}(t) + \\ b_9 Ave\_RV_i^{(3)}(t) + b_{10} Ave\_Rating_i^{(3)}(t) + \varepsilon_i. \end{aligned} \tag{8.4}$$

$$\begin{aligned} AVE\_Rating_i(t) = b_0 + b_1 NOR_i + b_2 In\_deg_i^{(1)} + b_3 In\_deg_i^{(2)} + \\ b_4 In\_deg_i^{(3)} + b_5 Ave\_RV_i^{(1)}(t) + b_6 Ave\_Rating_i^{(1)}(t) \\ + b_7 Ave\_RV_i^{(2)}(t) + b_8 Ave\_Rating_i^{(2)}(t) + \\ b_9 Ave\_RV_i^{(3)}(t) + b_{10} Ave\_Rating_i^{(3)}(t) + \varepsilon_i. \end{aligned} \tag{8.5}$$

Considering the scale, we use the logarithms of each control variable in the model.

8.2.3 Results for Neighbourhood Level Analysis

Review Volume of Focal Books

Table 8.3 shows the results for review volume of focal books. Four models are analysed. Model 1 only considers the control variables, while the rest three models are including the eWOM information of first-order, second-order and third-order neighbours are added progressively. The results of F test in these four models suggest the significant impact of selected variables. When the variables are stepwise included, the value of R squared becomes bigger and AIC, BIC become smaller. The inclusion of first-order and second-order neighbours could significant improve the explanatory power of the modes (R Squared is 0.068, 0.090, 0.128 in the model1 to model 3). However, the impact of the third-order neighbours is not as strong as the first- and second-order neighbours, as shown from the value of R^2, BIC and AIC in model 4 that remain unchanged. That means the impact of eWOM of neighbours on the focal books' review volume become slim when the distance between books reaches three steps.

The impact of control variables, as shown in model 1, suggests that NOR_i, $In\_deg_i^{(1)}$ and $In\_deg_i^{(2)}$ positively and significantly impact the review volume of focal books. However, the number of books at three clicks away, $In\_deg_i^{(3)}$ has negative influence.

Table 8.3| Regression results on review volume of focal books.

DV: $\boldsymbol{RV_i(t)}$				
	Model 1	**Model 2**	**Model 3**	**Model 4** **Full Model**
Variables	**Coef.** **(Std. err.)**	**Coef.** **(Std. err.)**	**Coef.** **(Std. err.)**	**Coef.** **(Std. err.)**
NOR_i	0.050*** (0.003)	0.039 *** (0.003)	0.027*** (0.003)	0.029*** (0.003)
$In\_deg_i^{(1)}$	0.051*** (0.007)	0.092*** (0.007)	0.067*** (0.007)	0.068*** (0.007)
$In\_deg_i^{(2)}$	0.227*** (0.012)	0.189*** (0.012)	0.338*** (0.012)	0.339*** (0.012)
$In\_deg_i^{(3)}$	-0.100*** (0.009)	-0.100*** (0.009)	-0.159*** (0.009)	-0.159*** (0.008)
$Ave\_RV_i^{(1)}(t)$		1.008*** (0.024)	0.530*** (0.025)	0.555*** (0.025)
$Ave\_Rating_i^{(1)}(t)$		-0.195*** (0.043)	-0.174*** (0.043)	-0.180*** (0.041)
$Ave\_RV_i^{(2)}(t)$			3.007*** (0.052)	3.156*** (0.053)
$Ave\_Rating_i^{(2)}(t)$			-0.205*** (0.057)	-0.233*** (0.055)
$Ave\_RV_i^{(3)}(t)$				-1.027*** (0.104)
$Ave\_Rating_i^{(3)}(t)$				0.150 (0.213)
Constant	0.450*** (0.017)	1.331*** (0.192)	1.792*** (0.294)	1.280 (0.993)
Observations	75800	75800	75800	75800
R^2	0.068	0.090	0.128	0.129
F-statistic Prob (F-statistic)	1387 0.00	1245 0.00	1396 0.00	1128 0.00
AIC	2.098e+05	2.080e+05	2.047e+05	2.047e+05
BIC	2.098e+05	2.081e+05	2.048e+05	2.048e+05

*** p<0.01, ** p<0.05, * p<0.1

When we include the eWOM information of the first-order neighbours into the model, as shown in Model 2, the average review volume of neighbour books, $Ave\_RV_i^{(1)}(t)$ has positive impact on the volume of focal books where the estimated coefficient is 1.008 and $p < 0.01$. However, the average rating of the first-order neighbour books, $Ave\_Rating_i^{(1)}(t)$, lowers the review volume of focal books where the estimated coefficient is -0.195 and $p < 0.01$. The identical impact also exists in the second-order neighbours as shown in Model 3. $Ave\_RV_i^{(2)}(t)$ is positively and $Ave\_Rating_i^{(2)}(t)$ is negatively influencing the review volume of focal books. Such impact becomes weaker for these neighbour books at three clicks away, as shown in Model 4. The impact of the average volume of the third-order neighbours, $Ave\_RV_i^{(3)}(t)$ is negative while the average rating, $Ave\_Rating_i^{(3)}(t)$ has no significant impact.

In summary, the review volume of focal books is positively related to the average volume of neighbours and negatively related to the neighbours' average rating. Such impact could come from the books that are three clicks away in the recommendation network, though their impact is weaker than that of the first- and second-order neighbours.

Review Rating of Focal Books

Table 8.4 shows the analytical results of the average ratings of focal books. Four models are studied with similar manner to the analysis of the review volume of focal books. The measures of model fit are basically as same as the results of last section. The results of F test in the four models suggest the significant impact of variables on the focal books' rating information. The R^2 becomes bigger when the first- and second- order neighbours are considered, but remains unchanged with the third-order neighbours. From above results, the impact of eWOM of neighbours on focal books' review rating become slim when the distance reaches three steps as well. According to AIC and BIC, the inclusive of eWOM information of neighbours can help to fit the model, leading to lower BIC and AIC. But the trend becomes slim when the third-order neighbour is added in model 8, that the difference of AIC, BIC of model 7 and model 8 is smaller than the previous models.

When only analysing the control variables (Model 5), only $In\_deg_i^{(1)}$ and NOR_i have significant and positive impact on the rating of focal books. The number of books at two or three clicks away, $In\_deg_i^{(2)}$ and $In\_deg_i^{(3)}$ do not have significant impact on the rating of focal books.

Table 8.4| Regression results on average rating of focal books.

DV: $AVE\_Rating_i(t)$				
	Model 5	**Model 6**	**Model 7**	**Model 8** **Full Model**
Variables	**Coef.** **(Std. err.)**	**Coef.** **(Std. err.)**	**Coef.** **(Std. err.)**	**Coef.** **(Std. err.)**
NOR_i	0.021***	0.018***	0.013***	0.013***
	(0.003)	(0.003)	(0.002)	(0.002)
$In\_deg_i^{(1)}$	0.016*	0.020*	0.009	0.009
	(0.008)	(0.008)	(0.007)	(0.007)
$In\_deg_i^{(2)}$	-0.010	-0.008	-0.013	-0.014
	(0.013)	(0.013)	(0.011)	(0.011)
$In\_deg_i^{(3)}$	-0.004	-0.006	0.009**	-0.009
	(0.009)	(0.009)	(0.008)	(0.008)
$Ave\_RV_i^{(1)}(t)$		0.112***	0.044***	0.040***
		(0.026)	(0.024)	(0.024)
$Ave\_Rating_i^{(1)}(t)$		1.474***	0.635***	0.627***
		(0.047)	(0.041)	(0.041)
$Ave\_RV_i^{(2)}(t)$			-0.458***	-0.489***
			(0.050)	(0.052)
$Ave\_Rating_i^{(2)}(t)$			8.555***	8.543***
			(0.055)	(0.055)
$Ave\_RV_i^{(3)}(t)$				0.141
				(0.102)
$Ave\_Rating_i^{(3)}(t)$				1.993***
				(0.209)
Constant	4.391***	-2.143***	-36.25***	-44.97***
	(0.018)	(0.208)	(0.283)	(0.955)
Observations	75800	75800	75800	75800
R^2	0.001	0.015	0.256	0.257
F-statistic	19.85	195	3260	2621
Prob(F-statistic)	0.00	0.00	0.00	0.00
AIC	2.211e+05	2.201e+05	1.988e+05	1.987e+05
BIC	2.212e+05	2.201e+05	1.988e+05	1.988e+05

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

When considering the eWOM information of the first-order neighbours in Model 6, both $Ave\_RV_i^{(1)}(t)$ and $Ave\_Rating_i^{(1)}(t)$ have positive impact on focal book's rating. This indicates that the feedback of books is strongly associated with the discussions of neighbours. However, the eWOM information of neighbours that are two clicks away in the recommendation network has different impact on the focal books' ratings. As shown in Model 7, the impact of $Ave\_RV_i^{(2)}(t)$ is negative where the estimated coefficient is -0.458, while the impact of $Ave\_Rating_i^{(2)}(t)$ is positive where the estimated coefficient is 8.555. In addition, the average rating of third-order books is positively related to the rating of focal books ($Ave\_Rating_i^{(3)}(t) = 1.993$ in Model 8), while the review volume $Ave\_RV_i^{(3)}(t)$ does not show significant effect.

In summary, the findings suggest that firstly the focal books' ratings are associated with the ratings of their neighbours. Such influence is positive, which means a positive feedback of one book may cause high ratings for its recommended books. Secondly, the relations between the review volume of neighbours and the focal books' ratings are mixed, where first-order neighbours' review volume has positive influence, the second-order neighbours' review volume has negative influence, while the third-order neighbours have no influence.

8.3 Dyadic Level Analysis

8.3.1 Variable Operationalisation for Dyadic Level Analysis

Dependent Variable

Beside the neighbourhood level, we also conduct the analysis at the product-pair level. In this analysis, we use the difference of eWOM rating for each product pair consisting of book i and book j as the dependent variable, denoting with $WOM\_diff_{ij} = \left|WOM_i - WOM_j\right|$. Accordingly, small (large) value of WOM_diff_{ij} indicates that the books i and j have more similar (different) ratings.

Independent variables and control variables

Since the two products being connected or not can only partially represents their connectivity, here on dyadic level, we use the number of paths between the studied two products as the independent variables. For two books i and j, we count how many paths of length n are connecting them, denoted with $NOP_{ij}^{(n)}$. Therefore, such value can well describe the traffic flow between the two products, i.e. how easy can consumers surf from one to another. Hence, larger value of $NOP_{ij}^{(n)}$ represents better connectivity between the two products. In respect to the neighbourhood level analysis, here we also consider distances up to three, i.e. the number of paths $NOP_{ij}^{(1)}, NOP_{ij}^{(2)}$, $NOP_{ij}^{(3)}$. While $NOP_{ij}^{(1)}$ can only take values of 1 or 0, $NOP_{ij}^{(2)}$ and $NOP_{ij}^{(3)}$ could be any integral values. Actually, Lin and Wang (2018) have already suggest that the direct connection ($NOP_{ij}^{(1)} = 1$) between two products can lead to the convergence of their rating. In this work, we also consider indirect connections, namely $NOP_{ij}^{(2)}$ and $NOP_{ij}^{(3)}$, to explore the impact of distance between two products in the network on their eWOM similarity. In addition, we define the connectivity between two products by combining $NOP_{ij}^{(1)}$, $NOP_{ij}^{(2)}, NOP_{ij}^{(3)}$ to measure the likelihood for consumers browsing from the homepage of product i to j within three clicks in the following equation,

$$Connectivity_{ij} = \frac{NOP_{ij}^{(1)}}{OutDegree} + \frac{NOP_{ij}^{(2)}}{OutDegree^2} + \frac{NOP_{ij}^{(3)}}{OutDegree^3} \ . \tag{8.3}$$

In this equation, $OutDegree$ is the number of products in each product's recommendation list and according to our data collection, equals to 10. Thus, the number of all paths with length n originating from a book is $OutDegree^n$.

We use the differences between the basic information of two products as the control variables, including: the difference of in-degree, $Indeg\_diff_{ij} = |Indeg_i - Indeg_j|$; the difference of rating at the beginning of the studied period (1st December 2016), $Rating\_diff_{ij} = |Rating_i - Rating_j|$; the difference of review volume at the studied time period, $RV\_diff_{ij} = |RV_i - RV_j|$.

The operationalisation and descriptive statistics of all variables for the dyadic level are shown in Table 8.5, while the Table 8.6 reports the correlations among these variables.

Table 8.5| Description statistics of variables for dyadic level.

Variable Type	Variable Name	Descriptions	Mean	Std. Dev.	Min	Max
Dependent Variables	WOM_diff_{ij}	Difference of WOM rating between product i and j	0.80	0.94	0	4
Independent Variables	$NOP_{ij}^{(1)}$	Number of paths from product i to j using one click	0.05	0.21	0	1
	$NOP_{ij}^{(2)}$	Number of paths from product i to j using two clicks	0.44	1.10	0	10
	$NOP_{ij}^{(3)}$	Number of paths from product i to j using three clicks	4.60	8.91	0	91
	$Connectivity_{i}$	Connectivity between product i and j within three clicks	0.013	0.036	0.001	0.281
Control Variables	$Indeg\_diff_{ij}$	Difference of in-degree between product i and j	44.58	85.4	0	910
	$Rating\_diff_{ij}$	Difference of in-degree between product i and j at the beginning time of data	0.38	0.38	0	4
	RV_diff_{ij}	Difference of review volume between product i and j	9.91	28.6	0	411

Table 8.6| Correlations among variables for dyadic level.

Variables	1	2	3	4	5	6	7	8
1. WOM_diff_{ij}	1							
2. $Rating\_diff_{ij}$	0.169	1						
3. $Indeg\_diff_{ij}$	-0.018	-0.012	1					
4. RV_diff_{ij}	-0.027	-0.001	0.551	1				
5. $NOP_{ij}^{(1)}$	-0.015	-0.028	-0.011	-0.007	1			
6. $NOP_{ij}^{(2)}$	-0.020	-0.034	0.029	0.028	0.574	1		
7. $NOP_{ij}^{(3)}$	-0.024	-0.037	0.066	0.065	0.667	0.901	1	
8. $Connectivity_{ij}$	-0.020	-0.036	0.019	0.021	0.910	0.852	0.890	1

8.3.2 Model Specification for Dyadic Level Analysis

For the analysis on the dyadic product-pair level, we model the effect of distance, which is measured by connectivity and path number in term of click number between each pair of products respectively, on the difference of eWOM ratings between two books. The following model is applied:

$$WOM\_diff_{ij} = b_0 + b_1 RV\_diff_{ij} + b_2 Indeg\_diff_{ij} + b_3 Rating\_diff_{ij} + b_4 X + \varepsilon_i, \tag{8.6}$$

where X is $NOP_{ij}^{(1)}$, $NOP_{ij}^{(2)}$, $NOP_{ij}^{(3)}$, or $Connectivity_{ij}$, which will be included in turns to analyse the impact of different distance separately.

Table 8.7| Regression results on difference of eWOM Rating at dyadic level.

DV: WOM_diff_{ij}					
	Model 9	**Model 10**	**Model 11**	**Model 12**	**Model 13**
Variables	**Coef. (Std. err.)**	**Coef. (Std. err.)**	**Coef. (Std. err.)**	**Coef. (Std.err.)**	**Coef. (Std. err.)**
$Rating\_diff_{ij}$	0.420*** (0.001)	0.420*** (0.001)	0.419*** (0.001)	0.419*** (0.001)	0.419*** (0.001)
$Indeg\_diff_{ij}$	-0.000** (6.58e-06)	-1.46e-05** (6.58e-06)	-1.06e-05 (6.58e-06)	-5.92e-06 (6.59e-06)	-1.19e-05* (6.58e-06)
RV_diff_{ij}	-0.001*** (1.96e-05)	-0.001*** (1.96e-05)	-0.001*** (1.96e-05)	-0.001*** (1.96e-05)	-0.001*** (1.96e-05)
$NOP_{ij}^{(1)}$		-0.046*** (0.002)			
$NOP_{ij}^{(2)}$			-0.011*** (0.000)		
$NOP_{ij}^{(3)}$				-0.002*** (5.29e-05)	
$Connectivity_{ij}$					-0.359*** (0.013)
Constant	0.651*** (0.000)	0.651*** (0.000)	0.654*** (0.000)	0.656*** (0.000)	0.653*** (0.000)
Observations	3,887,462	3,887,462	3,887,462	3,887,462	3,887,462
R-squared	0.029	0.029	0.029	0.029	0.029
F-statistic Prob(F-statistic)	3.887e+04 0.00	2.926e+04 0.00	2.935e+04 0.00	2.935e+04 0.00	2.935e+04 0.00
AIC	1.043e+07	1.043e+07	1.043e+07	1.043e+07	1.043e+07
BIC	1.043e+07	1.043e+07	1.043e+07	1.043e+07	1.043e+07

*** p<0.01, ** p<0.05, * p<0.1

8.3.3 Results for Dyadic Level Analysis

We firstly examine the effect of all control variables on the difference of eWOM ratings, as shown in model 9 in Table 8.7. Considering the overlap of $NOP_{ij}^{(1)}$, $NOP_{ij}^{(2)}$, $NOP_{ij}^{(3)}$, in the model 10 to model 12, they are examined by turns on the difference of eWOM rating rather than stepwise added. And model 13 examines the impact of connectivity measured by Eq (8.3). According to the results of F-statistic in these models, the studied variables (number of path and connectivity) significantly relate to the eWOM difference of two books at dyadic level. The estimated value of R-Squared, AIC and BIC are unchanged, suggesting that the studied variables may equally influence the eWOM difference.

Firstly, the results show that all control variables have significant impact on the rating difference between a pair of books. RV_diff_{ij} is positive and $Indeg\_diff_{ij}$, $Rating\_diff_{ij}$ are negative. The results are consistent with the work of Lin and Wang (2018). The significant estimates suggest that the difference of eWOM rating between two products could be explained by the differences of their attributions.

We further investigate the independent variables (i.e. $NOP_{ij}^{(1)}$, $NOP_{ij}^{(2)}$, $NOP_{ij}^{(3)}$ and $Connectivity_{ij}$) in turns as shown by model 10 to 13. All variables have significant negative effect on the difference of eWOM rating, which indicates that the both direct and indirect connections between two products are related to the similarity of the eWOM rating of two products, and the more paths connecting the two products, regardless of the length (1, 2, 3 or combined) the closer ratings (smaller WOM_diff_{ij}) the two products will have. The results not only confirm the finding of Lin and Wang (2018) that the eWOM of the directly connected products ($NOP_{ij}^{(1)}$) tend to converge to each other, but also extend such finding that indirect connections such as connections with length of 2 or 3 can also result in similar eWOM ratings. Additionally, the absolute coefficient of NOP_{ij} becomes lower as the distances of two books increase ($NOP_{ij}^{(1)}$=-0.046, $NOP_{ij}^{(2)}$=-0.011, $NOP_{ij}^{(3)}$=-0.002). This suggests that the effect of path numbers between products in the recommendation network on their eWOM rating difference become weaker when the distances become larger. The direct paths, $NOP_{ij}^{(1)}$ have the strongest impact, while the indirect paths $NOP_{ij}^{(2)}$ and $NOP_{ij}^{(3)}$ are weaker, though still significant. Model 13 examines the effect of

$Connectivity_{ij}$ which is the general metric combining the path numbers, has significant negative effect as well. In summary, products that have better connections to each other, will normally have similar eWOM ratings.

8.4 Summary

In this chapter, we apply an empirical book recommendation network from Amazon to investigate the impact of distance on eWOM from two levels: neighbourhood level and dyadic product-pair level.

On the neighbourhood level, the objective is to study the correlations of eWOM between the neighbours and the focal books in terms of daily review volume and daily average rating. We examine the impact of eWOM of first-order, second-order and third-order neighbours on the review volume of focal books. The results suggest that the average review volume of the first-order and second-order neighbours positively influences the focal books' review volume. But the impact of review volume of third-order neighbours is negative. On the other hand, the average rating of only the first-order and second-order neighbours are significantly, but negatively related to the focal books' review volume. Moreover, we explore the impact of neighbours' eWOM on the focal books' rating. We find that such impact is mixed depending on the distances between the neighbours and the focal books. Both the average review volume and average rating of the first-order neighbours have positive impact on the focal books' rating. But the second order neighbours' review volume is negatively associated with the focal books' rating, while the average rating of the second-order neighbours has positive correlations. For the third-order neighbours, only the average rating has significant effect on focal books' ratings.

On the dyadic product-pair level, we use number of paths connecting from book i to book j in one, two and three clicks respectively as the approximation for their connectivity, to examine its impact on the rating difference between book pairs. The results indicate that the connectivity between two products, describing the easiness for the users to surf from one to another, significantly influences the rating similarity between the studied books. Better connectivity normally associates with more similar eWOM rating. Such impact also depends on the distance. The number of shorter paths will have stronger influence while the number of the longer paths have weaker, yet significant influence.

In summary, the results suggest that both the direct and indirect connections between products in a product recommendation network can influence their eWOM, i.e. the eWOM of a certain product can influence not only its directly recommended products but also other products that are two or three clicks away, indicating a strong network effect on the formation of the eWOM.

Chapter 9. Summary, Discussions and Contributions

This chapter concludes the thesis by presenting a summary of the studies, the contributions, limitations, and the recommendations for the future work. Section 9.1 firstly presents concluding remarks which summarise the findings in the preceding chapters on how the research questions are addressed. Sections 9.2 presents the contributions from theoretical, methodological and practical aspects. Finally, the chapter discusses the limitations and recommendations for future work in section 9.3.

9.1 Concluding Remarks

The Internet continually generates information regarding people, products and businesses. Billions of digital footprints left by users such as posting photos with location, tagging brands of purchased products on social media, sharing opinions or reviews towards merchandises, arouse interests of researchers and businesses. However, the overwhelming and fast-growing information makes the proper usage of such kind of data of great importance. Data-driven decision-marking provides some insights to form a continuum in which data is transformed to information, and ultimately to knowledge that can be applied to make decisions (Ackoff, 1989; Light *et al.,* 2004). Focusing on the raw data that is generated by online users, this thesis tries to mine the usable knowledge from such large scale of raw data. In particular, the research explores the interplay between consumers, products and online reviews in the context of social networks and product recommendation networks, based on eWOM data in the ecommerce websites (Yelp and Amazon) and computational models.

In specific, the thesis formulates four research questions: (1) How friend's and crowd's reviews differently impact consumer engagement on posting behaviour? (2) How would social influence affect consumers' selection behaviour? (3) Which factors can be used to predict review helpfulness? (4) How would distance between a pair of products in the PRN influence their eWOM relationship? These research questions can be readily answered as follows according to the results in this thesis.

Chapter 5 addresses the first question of the thesis by examining the different impact of friend reviews and crowd reviews on consumer posting behaviour. Though it is widely argued that the volume, valence, and variance of online reviews largely influence subsequent consumer behaviours, we show that such influences have different sources. While the valence and variance of crowd reviews show stronger influence over one's behaviour, it is found that, for the review volume, friend review has more effect than that of general crowds. In addition, if further separate the positive reviews and negative reviews, the crowd and friends are shown to have another major difference. Only the negative reviews of crowds would influence one's behaviour, while the positive reviews of friends show significance. Such differences between crowd and friends may be caused by the trust between the users.

Chapter 6 further discusses friend-based and crowd-based social influence measured by a product's local and global popularity on consumer selection behaviour, which addresses the second research question. While both kinds of social influence are widely acknowledged, we confirm their existence in users' selection and posting behaviour and study their intensity separately. It is found that the friend-based social influence always has significant influence over one's selection behaviour, while the crowd-based social influence only plays a strong role when there is no much friends opinions to refer to. By modelling the users' behaviour based on the Monte Carlo simulation to reproduce the empirical data, we show that about 75% of users' decisions are made by followings their friends' opinions, while only 25% may be influenced by the crowd.

Chapter 7 addresses the third question of this study by adopting a dynamical method for data collection to study the determinants of review helpfulness for both old and recent reviews. We show that the reviewer activeness, reviewer historical votes, reviewer credibility, review disclosure information, review readability, and review sentiment all have significant influence on the helpfulness of reviews. However, such strong influence may be due to the endogeneity between these factors and the helpfulness, because normally the data is collected together for a same period of time. When using these factors to predict the future increment of helpfulness, most of them become much less predictive. In particular, the helpfulness of recent reviews is shown to be more difficult to be predicted. The factors with relatively high predictive power come from two aspects, namely reviewer activeness and review disclosure information.

Chapter 8 discusses the impact of distance between products in the PRN on their eWOM to address the fourth research question. The results indicate that the PRNs can not only influence the demand and sales of products, but also reshapes the eWOM. The products that are connected to each other in the PRN are found with similar eWOM. In addition to direct connections, the indirect connections within three clicks are also shown significance for the similarity between the eWOM of two products, but such influence is relatively weaker than the direct ones.

9.2 Contributions

The contributions of the research can be judged by originality in form of theory development, application of an existing theory or methodology in a novel way, developing new methods or approaches, as well as providing new insights for research or practice. Following this, we summarise the contributions from three aspects, i.e. theoretical, methodological and practical.

9.2.1 Theoretical Contributions

The study makes theoretical contributions to the fields of eWOM, social influence and consumer behaviour, as well as electronic commerce and marketing by filling several gaps in the literature.

Previous literature regarding the influence of eWOM on subsequent users' posting and selection behaviours have made great achievements, in which a number of key factors have been uncovered to be the key determinants. However, most of these studies only regard the eWOM as a whole to each user, leaving the effect of the source of eWOM an open question. In this thesis, we address the source of eWOM in terms of a user's friends or larger crowd. Applying econometric methods, simulations, and network modelling, we show that the source does significantly change the influence of eWOM. In general, the friends have stronger influence over one's behaviour. In addition, the developed network model describing the users' selection behaviour also contribute to the knowledge of network science. While the preferential attachment has been regarded as the main driver for the network evolution, we show that the local-driven attachment can better describe the observations.

In the field of review helpfulness, the normal practice of collecting data is based on a single time point. Thus, such studies are at risk of endogeneity, where some factors, such as reviewer reputation may actually be caused by the helpfulness. In addition, whether the identified factors in previous studies can actually help us to predict a review's future helpfulness votes, is rather unknown. With a dynamical data set collected at two different time point, this thesis examined whether the previously confirmed factors are able to predict the future helpfulness with endogeneity removed. The results do show many of these factors become insignificant. Hence, the field should be aware of such differences when designing future studies to avoid the endogeneity. Some promising classes of factors are also discussed in the thesis which are able to predict the future helpfulness for both new and old reviews.

Lastly, the thesis extends the study of PRN and eWOM considering the network effect on the formation of eWOM. While most previous studies focus only on the economic impact of the PRN (Carmi *et al.,* 2017; Lin and Wang, 2018; Oestreicher-Singer *et al.,* 2013), such the influence on sales and demand, we believe that the eWOM of the products in PRN is actually the intermediate variable. The thesis provides an possible explanation that the products' eWOM is highly correlated when they are near (short distance) to each other in PRNs, which further leads to the enhancement of demands and sales. In addition, for the first time, wo show that the eWOM is similar to ideas and innovations, which can spread over the product networks up to three clicks away.

To summaries, in theory, this thesis highlights the network effect on the formation and diffusion of eWOM, that the influence of reviews on users can be enlarged through social network, and the product network also largely reshapes the eWOM of products.

9.2.2 Methodological Contributions

The thesis contributes methodologically by providing new approaches and techniques to understand consumer behaviours. Following computational social science, this thesis brings together the traditional methodologies such as econometric models, and the computational methodologies such as web scraping for data collection, network modelling, semantic analysis, and simulations. These computational techniques not only provide new insights and new opportunities for understanding consumer behaviour, but also overcome the limitations of the traditional methodologies.

For example, though the influence of friend reviews and crowd reviews can be compared via the econometric model, as reported in Chapter 4, it is difficult to quantify the intensity of such two sources or observe some potential nonlinear correlations. By introducing the distributional analysis with the conditional probability, which is normal practice in statistical physics, we are enabled to study such intensity with each source of information being controlled in turns. Accordingly, an additional conclusion can be made that only when there are no friend reviews, the crowd reviews could influence the users' behaviour. Such nonlinear pattern is not as apparent to be only studied via traditional methodologies.

A significant amount of methodologies from the field of network science have also been introduced to the studies in this thesis, which helped to not only provide supplement analysis, but also generate potential variables to be analysed. It is an important part of network science studies to model the practical systems as networks. Accordingly, we proposed one such model in Chapter 5 to try to simulate the users' selection behaviour under the influence of both friends and crowds. Such modelling and simulation well math with the empirical observation, indicating that the mechanism of users' behaviour is as modelled, where 75% of the behaviours are influenced by the friends. Another application of network science methodology is in Chapter 8, where we use the concept of shortest path length and node connectivity as independent variables to explain the similarity between two products' eWOM.

Another innovative methodology we adopted in the thesis is the data collection. Instead of using second-handed data for our study, we collect web data by ourselves, such as in

Chapter 8, so that the data collection is better designed, and the data is more purposed to the targeted study. Most importantly, we also try to use data collection methodology to overcome limitations of the previous studies, such as in Chapter 7. While the one-time collection of data may result in endogeneity problem, we develop a data set which is collected at two time point with an eighteen-month gap in between. By doing so, we are enabled to study more closely on the causal relations between different factors and the future increment of the helpfulness votes of reviews.

Overall, this thesis is a good example for applying both econometric and computational methodologies to complement each other to gain better knowledge on the consumer behaviour. Such example should be able to shed some light on the future methodological design for new studies.

9.2.3 Practical Contributions

Online product review has become one of the most important channels to collect product information for consumers and advertise new products for business. Thus, the thesis has also significant practical implications.

First, the study offers implications for online marketing and the design of the online user-generated content systems. Prior reviews indeed have significant impact on the subsequent consumer decision of whether to post reviews after consumption. A large number of reviews normally associates with flourishing subsequent posting behaviour. On one hand, the system should make the reviews easy accessible for consumers to help them make decisions of consuming and posting. On the other hand, to have thriving reviews should be the one of the priorities for the online marketing, since the popularity of a product is normally self-reinforcing.

Second, the analysis in this thesis highlights the importance of the social networking. Friends' opinions are shown to be more determinative for the likelihood of consumer posting behaviour. The posted information is regarded as ‘‘sale assistant” (Chen and Xie, 2008) that can largely promote the business sales and consumer engagement on eWOM. Accordingly, social network services should be introduced to those online user-generated content systems to facilitate the eWOM. In addition, such finding suggests that it is possibly more efficient to seek for marketing in well-established social networks such as Facebook or Twitter.

Third, the findings in this work could help retailers to improve the ranking mechanism of consumer reviews. Having insightful consumer reviews could improve the shopping experience and the possibility for consumers to find proper products quickly. Though the cumulative helpful votes of reviews are important, retailers should not focus solely on it as recently-posted reviews may have not enough votes to reflect their true helpfulness. Accordingly, the review ecosystem should consider a proper ageing mechanism based on the pattern of the increment of helpful votes. People are generally described as cognitive misers, that prefer to process the information with less cognitive efforts than devoting mental resources to deliberate thinking (Evans, 2008). Thus, online retailers may need to emphasis more on the straightforward information of reviews instead of other complicated information when they design the ranking mechanism. For example, our findings show that the *review disclosure information* is much more important in determining the review helpfulness. The system designers should therefore carefully highlight the information to make them more easily accessible for consumers.

Last, the results show that the eWOM about a product is strongly related to that of its neighbours in the PRN. Therefore, ecommerce retailers can take advantage of the influence of the PRN to facilitate the discussions and drive the products' demands. For example, to achieve better eWOM, one should consider locating the target product at the optimised position where the neighbours have positive eWOM. Secondly, the study gives some insights for the design of PRNs. Our findings show that eWOM of products has impact on their recommended others, even the indirectly recommended ones. Thus, the platform designers shall carefully consider about the connection establishments in the PRN to optimally facilitate and make use of the advantage of the eWOM.

9.3 Limitations and Future Work

The thesis has many limitations as discussed in the following. Meanwhile some possible future directions can be followed to make such line of research more complete and comprehensive.

First, the study in this thesis is exploratory and data-driven. The developed econometric as well as computational methods in this work lack of evaluations and comparisons of benchmarks. For example, to examine whether the previously-studied factors are able to predict the increment of helpful votes, this study considers only a fundamental regression model for the prediction and does not explore for more efficient factors. Actually, some other models such as hybrid model (Ngo-Ye and Sinha, 2014) and neural networks (Lee and Choeh, 2014), have shown better accuracy. Their performance on predicting the increment of helpful votes should thus be further examined. How to select proper econometric models in data-driven research should be the focus of future work. In addition, the low marginal R-squared of the analytical results suggest the limitation of the proposed mixed models.

Second, the three research questions are answered by the data set of Yelp published by the company, leading to the limitation of product type. Different types of products, i.e. experience goods or searching goods, have been found with totally different behaviours on eWOM and review helpfulness (Mudambi and Schuff, 2010). However, most of "products" in our study are experience goods, e.g. restaurants. In addition, Yelp is the third party providing reviews toward businesses but not selling products. Thus, if other review websites such as Netflix, Amazon still gain similar results is still unknown. Thus, for the future work, more user-generated content platforms are needed to be explored.

Last, the use of a single PRN limits wide generalisation of the conclusions. We only examine one book recommendation network based on the co-purchase recommendations of Amazon. It is unknown whether our findings can be applied to other types of products. Thus, future work should investigate more types of product and generalise the findings. In this study, eWOM of neighbours regarding as independent variables are averaged over all the neighbours of the same category. While it shows significance explaining eWOM correlation, it is worthy to examine other operationalisations for the variables such as aggregated value over the neighbours. In addition, as the results have suggested that eWOM of a focal book is significantly determined by its neighbours, whether we can identify the optimal neighbourhood or position to best enhance a target book’s eWOM still need further efforts.

References

Ackoff, R. L. (1989). From data to wisdom. Journal of applied systems analysis, 16(1), 3-9.

Alba, J., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A., and Wood, S. (1997). Interactive home shopping: consumer, retailer, and manufacturer incentives to participate in electronic marketplaces. *The Journal of Marketing*, 38-53.

Algesheimer, R., Borle, S., Dholakia, U. M., and Singh, S. S. (2010). The impact of customer community participation on customer behaviors: An empirical investigation. *Marketing science*, 29(4), 756-769.

Alsumait, L., Barbará, D., and Domeniconi, C. (2008, December). On-line lda: Adaptive topic models for mining text streams with applications to topic detection and tracking. In: *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on (pp. 3-12). IEEE.* Washington, DC, USA

Alvarez-Galvez, J. (2016). Discovering complex interrelationships between socioeconomic status and health in Europe: A case study applying Bayesian Networks. *Social Science Research* 56, 133-143.

Anagnostopoulos, A., Kumar, R., and Mahdian, M. (2008, August). Influence and correlation in social networks. In: *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 7-15). Las Vegas, Nevada, USA.

Anderson, R.E., (1973). Consumer dissatisfaction: the effect of disconfirmed expectancy on perceived product performance. *Journal of Marketing Research*, 10 (1), 38–44.

Andreassen, T. W., and Streukens, S. (2009). Service innovation and electronic word-of-mouth: is it worth listening to? *Managing Service Quality: An International Journal*, 19(3), 249-265.

Angelis, M. D., Bonezzi, A., Peluso, A. M., Rucker, D. D., and Costabile, M. (2012). On braggarts and gossips: A self-enhancement account of word-of-mouth generation and transmission. *Journal of Marketing Research*, 49(4), 551-563.

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

Aral, S., Walker, D., (2011). Creating social contagion through viral product design: a randomized trial of peer influence in networks. *Management Science*, 57 (9), 1623–1639.

Aral, S., and Nicolaides, C. (2017). Exercise contagion in a global social network. *Nature communications*, 8, 14753.

Awad, N. F., and Ragowsky, A. (2008). Establishing trust in electronic commerce through online word of mouth: An examination across genders. *Journal of Management Information Systems*, 24(4), 101-121.

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS quarterly*, 243-268.

Backstrom, L., Boldi, P., Rosa, M., Ugander, J., and Vigna, S. (2012, June). Four degrees of separation. In: *Proceedings of the 4th Annual ACM Web Science Conference* (pp. 33-42). Evanston, Illinois, USA.

Baek, H., Ahn, J., Choi, Y. (2012). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce*, *17*, 99–126.

Bailey, A. A. (2005). Consumer awareness and use of product review websites. *Journal of Interactive Advertising*, 6(1), 68-81.

Bailey, N. T. (1975). *The mathematical theory of infectious diseases and its applications*.

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012, April). The role of social networks in information diffusion. In: *Proceedings of the 21st international conference on World Wide Web* (pp. 519-528). ACM.

Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics*, 107(3), 797-817.

Bankes, S., Lempert, R., & Popper, S. (2002). Making computational social science effective: epistemology, methodology, and technology. Social Science Computer Review, 20(4), 377-388

Bapna, R., & Umyarov, A. (2015). Do your online friends make you pay? A randomized field experiment on peer influence in online social networks. *Management Science*, 61(8), 1902-1920.

Barabási, A. L. (2013). Network science. *Phil. Trans. R. Soc. A*, *371*(1987), 20120375.

Baraba´ si A.L. & Albert R. (1999) Emergence of scaling in random networks. *Science* 286 (5439), 509–512.

Bar-Yam, Y. (1997). *Dynamics of complex systems* (Vol. 213). Reading, MA: Addison-Wesley.

Bauman, K. E., & Ennett, S. T. (1994). Peer influence on adolescent drug use. *American Psychologist*, 49(9), 820.

Beatty, S. E., & Smith, S. M. (1987). External search effort: An investigation across several product categories. Journal of consumer research, 14(1), 83-95.

Berry FS, Berry WD. (1990). State lottery adoptions as policy innovations: an event history analysis. *American Political Science Review*, 84, 395–415.

Berry FS, Berry WD. (1992). Tax innovation in the states: capitalizing on political opportunity. *American Journal Political Science,* 36, 715–42.

Bertrand, J. (2004). Diffusion of innovations and HIV/AIDS. *Journal of Health Communication*, 9(6), 113–121.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3), 151-170.

Blaikie, N. (2007). *Approaches to social enquiry: Advancing knowledge*. Polity.

Blazevic, V., Hammedi, W., Garnefeld, I., Rust, R. T., Keiningham, T., Andreassen, T. W. and Carl, W. (2013). Beyond traditional word-of-mouth: an expanded model of customer-driven influence. *Journal of Service Management*, 24(3), 294-313.

Bloch, P.H., Sherrell, D.L. and Ridgway, N.M. (1986). Consumer search: an extended framework. *Journal of Consumer Research*, 13,119–126.

Bloom J. (2002). A line of blood: how December 1980 prepared Polish workers for political transition in 1989. pp. 85–102. Westport, CT: Greenwood

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics reports*, 424(4-5), 175-308.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, *24*(3), 127-135.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E. and Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415), 295.

Bornstein, R.F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. Psychological Bulletin, 106 (2), 265–289.

Brandes, U., Robins, G., McCranie, A., and Wasserman, S. (2013). What is network science? *Network Science*, 1(1), 1-15.

Breitenbach, C. S., and Van Doren, D. C. (1998). Value-added marketing in the digital domain: enhancing the utility of the International *Journal of Consumer Marketing*, 15(6), 558-575.

Bronner, F., and De Hoog, R. (2010). Consumer-generated versus marketer-generated websites in consumer decision making. *International Journal of Market Research*, 52(2), 231-248.

Brown, J.J., Reingen, P.H., (1987). Social ties and word-of-mouth referral behaviour. *Journal of Consumer Research*, 14 (3), 350–362.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, *44*(1), 108-132.

Brynjolfsson, E., & Smith, M. D. (2000). Frictionless commerce? A comparison of Internet and conventional retailers. *Management Science*, 46(4), 563-585.

Buller, D. B., Buller, M. K., & Kane, I. (2005). Web-based strategies to disseminate a sun safety curriculum to public elementary schools and state-licensed child-care facilities. *Health Psychology*, 24, 470–476.

Bunell, G., & Morgan, G. (1979). *Sociological paradigms and organisational analysis*. Gower, Aldershot, UK.

Carley, K. M. (2002). Computational organization science: A new frontier. *Proceedings of the National Academy of Sciences*, 99, 7257-7262.

Carmi, E., Oestreicher-Singer, G., Stettner, U., and Sundararajan, A., (2017). Is Oprah Contagious? The Depth of Diffusion of Demand Shocks in a Product Network. *MIS Quarterly*. 41(1), 207-221.

Cassell, C., and Symon, G. (Eds.), (1994), *Qualitative Methods in Organizational Research*, Sage Publications, CA

Cataldi, M., Di Caro, L., & Schifanella, C. (2010, July). Emerging topic detection on twitter based on temporal and social terms evaluation. In: *Proceedings of the tenth international workshop on multimedia data mining* (p. 4). ACM.

Celsi, R. L., & Olson, J. C. (1988). The role of involvement in attention and comprehension processes. *Journal of Consumer Research*, 15(2), 210-224.

Centola, D., (2010). The spread of behaviour in an online social network experiment. *Science*, 329 (5996), 1194–1197.

Chandler, D. (2007). *Semiotics: the basics*. Routledge.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology*, 4(1), 55-81.

Chatterjee, P. (2001). *Online reviews: do consumers use them?*

Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management science*, 54(3), 477-491.

Chen, Y., Wang, Q., and Xie, J. (2011). Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of Marketing Research*, 48(2), 238-254.

Chen, P.Y., Wu, S.Y., Yoon, J., (2004). The impact of online recommendations and consumer feedback on sales. In: *Proceeding of 25th International Conference on Information Systems*. DC, USA, Washington, pp. 711–724.

Chen, J., Xu, H., and Whinston, A. B. (2011). Moderated online communities and quality of user-generated content. *Journal of Management Information Systems*, 28(2), 237-268.

Chen, Y., Harper, F. M., Konstan, J., & Li, S. X. (2010). Social comparisons and contributions to online communities: A field experiment on movielens. *American Economic Review*, 100(4), 1358-98.

Cheung, M. Y., Luo, C., Sia, C. L., & Chen, H. (2009). Credibility of electronic word-of-mouth: Informational and normative determinants of online consumer recommendations. *International Journal of Electronic Commerce*, 13(4), 9-38.

Cheung, C. M., & Lee, M. K. (2012). What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision support systems*, 53(1), 218-225.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354.

Chintagunta, P.K., Gopinath, S., Venkataraman, S., (2010). The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets. *Marketing Science*, 29 (5), 944–957.

Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4), 370-379.

Christakis, N. A., & Fowler, J. H. (2008). The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, 358(21), 2249-2258.

Christakis, N. A., & Fowler, J. H. (2013). Social contagion theory: examining dynamic social networks and human behavior. *Statistics in Medicine*, 32(4), 556-577.

Chu, S. C., & Kim, Y. (2011). Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites. *International Journal of Advertising*, 30(1), 47-75.

Chua, A.Y., Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior,* 54*,* 547–554.

Cioffi-Revilla, C. (2014). Introduction to computational social science. *New York*.

Clark, A.E., Loheac, Y., (2007). It wasn't me, it was them! Social influence in risky behavior by adolescents. *Journal of Health Economics*, 26 (4), 763–784.

Clemons, E.K., Gao, G.G., Hitt, L.M., (2006). When online reviews meet hyperdifferentiation: a study of the craft beer industry. *Journal of Management Information System*, 23 (2), 149–171.

Collins, C., Harshbarger, C., Sawyer, R., and Hamdallah, M. (2006). The diffusion of effective interventions project: Development, implementation, and lessons learnt. *AIDS Education and Prevention*, 18(4 Suppl. A), 5–20.

Conover, M., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., and Flammini, A. (2011). Political polarization on twitter. *ICWSM* 133, 89-96.

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S., (2008). Feedback effects between similarity and social influence in online communities. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Las Vegas, Nevada, USA, pp. 160–168.

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.*

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. Sage.

Dabholkar, P. A. (2006). Factors influencing consumer choice of a" rating Web site": An experimental investigation of an online interactive decision aid. *Journal of Marketing Theory and Practice*, *14*(4), 259-273.

De Bruyn, A. and Lilien, G.L. (2008) A multi-stage model of word-of-mouth influence through viral marketing. *International Journal of Research in Marketing*, 25(3), 151–163.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10), 1407-1424.

Dellarocas, C., Fan, M., Wood, C.A., (2004). Self-interest, reciprocity, and participation in online reputation systems. Working Paper, Smith School of Business. University of Maryland.

Dellarocas, C., Awad, N., Zhang, M., (2005). Using online ratings as a proxy of word-of-mouth in motion picture revenue forecasting. Working Paper, Smith School of Business. University of Maryland.

Dellarocas, C., Narayan, R., (2006). A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science*, 21 (2), 277–285.

Dellarocas, C., Gao, G., & Narayan, R. (2010). Are consumers more likely to contribute online reviews for hit or niche products? *Journal of Management Information Systems,* 27(2), 127-158.

Dhar, V., & Chang, E. A. (2009). Does chatter matter? The impact of user-generated content on music sales*. Journal of Interactive Marketing*, 23(4), 300-307.

Dholakia, U. M., Blazevic, V., Wiertz, C., & Algesheimer, R. (2009). Communal service delivery: How customers benefit from participation in firm-hosted virtual P3 communities. *Journal of Service Research*, *12*(2), 208-226.

Diaz M.B., Porter M.A. & Onnela J.P. Competition for popularity in bipartite networks. *Chaos,* 20(4), 043101 (2010)

Dichter, E. (1966). How word-of-mouth advertising works. *Harvard Business Review*, 44(6), 147-160.

Doh, S.J., Hwang, J.S., 2009. How consumers evaluate eWOM messages. *Cyber Psychology & Behaviour*, 12 (2), 193–197.

Duan, W., Gu, B., Whinston, A.B., (2008). Do online reviews matter? An empirical investigation of panel data. *Decision Support System*, 45 (4), 1007–1016.

Duncan, C. P., & Olshavsky, R. W. (1982). External search: The role of consumer beliefs. *Journal of Marketing Research*, 32-43.

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.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, *12*(4), 1143-1168.

Engel, J.F., Blackwell, R.D. & Kegerreis, R.J. (1969) How information is used to adopt an innovation. *Journal of Advertising Research*, 9(4), 3–8.

Engel, J. F., Blackwell, R. W., & Miniard, P. W. (1993). Understanding the consumer. *ESCO Public Relations for FD's*, 1-9.

Escoffery, C., Glanz, K., & Elliott, T. (2007). Process evaluation of the Pool Cool Diffusion Trial for skin cancer prevention across 2 years. *Health Education Research*, 23(4), 732–743.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. Annual Review Psychology, 59, 255-278.

Falls, J. (2009). *Public relations pros must be social media ready*. Social Media Explorer.

Feick, L.F. & Price, L.L. (1987) The market maven: a diffuser of marketplace information. *Journal of Marketing*, 51(1), pp. 83–97.

Fisher, L. A., & Bauman, K. E. (1988). Influence and Selection In the Friend-adolescent Relationship: Findings from Studies of Adolescent Smoking and Drinking 1. *Journal of Applied Social Psychology*, *18*(4), 289-314.

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), 889–906.

Fiske, A. P. (2002). Using individualism and collectivism to compare cultures--A critique of the validity and measurement of the constructs. *Psychological Bulletin*, 128, 78–88.

Fliegel FC. (1993). *Diffusion Research in Rural Sociology: The Record for the Future*. Westport, CT: Greenwood

Flynn, L.R., Goldsmith, R.E., (1999). A short, reliable measure of subjective knowledge. *Journal of Business Research*, 46 (1), 57–66.

Fogg, B., Marshall, J., Laraki, O., et al. (2001). What makes web sites credible? a report on a large quantitative study, in: *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM. pp. 61–68.

Fortunato S., Flammini A. & Menczer F. (2006). Scale-free network growth by ranking. *Physical Review Letters*, 96(21), 218701.

Freeman, L. C., Roeder, D., & Mulholland, R. R. (1979). Centrality in social networks: II. Experimental results. *Social Networks*, 2(2), 119-141.

Gatignon, H., Robertson, T.S., 1985. A propositional inventory for new diffusion research. J. Consum. Res. 11 (4), 849–867.

Gauri, D. K., Bhatnagar, A., & Rao, R. (2008). Role of word of mouth in online store loyalty. *Communications of the ACM*, *51*(3), 89-91.

Gershoff, A.D., Mukherjee, A., Mukhopadhyay, A., 2003. Consumer acceptance of online agent advice: extremity and positivity effects. J. Consum. Psychol. 13 (1– 2), 161–170.

Ghose, A., Ipeirotis, P.G. (2007). Designing novel review ranking systems: predicting the usefulness and impact of reviews, in: Proceedings of the ninth international conference on Electronic commerce, ACM. pp. 303–310.

Gilly, M.C., Graham, J.L., Wolfinbarger, M.F. & Yale, L.J. (1998). A dyadic study of interpersonal information search. *Journal of the Academy of Marketing Science,* 26(2), 83–100.

Godes, D., and Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science*, 23(4), 545-560.

Goldsmith, R.E. and Horowitz, D. (2006) Measuring motivations for online opinion seeking. *Journal of Interactive Advertising*, 6(2).

Goldenberg, J., Libai, B., Moldovan, S., Muller, E., (2007). The NPV of bad news. *International Journal of Research Marketing*, 24 (3), 186–200.

Goldenberg, J., Oestreicher-Singer, G., and Reichman, S., (2012). The Quest for Content: How User-Generated Links can Facilitate Online Exploration. *Journal of Marketing Research*, 49(4), 452-468.

Gomez Rodriguez, M., Leskovec, J., & Krause, A. (2010, July). Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1019-1028). ACM.

Gomez Rodriguez, M., Leskovec, J., & Schölkopf, B. (2013, February). 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). ACM.

Granovetter, M. S. (1973). *The strength of weak ties*. In Social networks (pp. 347-367).

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., and Kyriakidou, O. (2004). Diffusion of innovations in service organizations: systematic review and recommendations. *The Milbank Quarterly*, 82(4), 581-629.

Greve HR. (1998). Performance, aspirations, and risky organizational change. *Administrative Science Quarterly,* 43, 58–86.

Gruen, T. W., Osmonbekov, T., and Czaplewski, A. J. (2006). eWOM: The impact of customer-to-customer online know-how exchange on customer value and loyalty. *Journal of Business research*, 59(4), 449-456.

Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004, May). Information diffusion through blogspace. In: *Proceedings of the 13th international conference on World Wide Web* (pp. 491-501). ACM.

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.

Guo, B., Zhou, S. (2016). Understanding the impact of prior reviews on subsequent reviews: The role of rating volume, variance and reviewer characteristics. *Electronic Commerce Research and Applications,* 20, 147–158.

Haider, M., & Kreps, G. L. (2004). Forty years of diffusion of innovations: Utility and value in public health. *Journal of Health Communication*, 9, 3–11.

Hampton, K.N., Shin, I., Lu, W., (2017). Social media and political discussion: when online presence silences offline conversation. *Information Communication & Society*, 20 (7), 1090–1107.

Hennig-Thurau, T., Walsh, G., and Walsh, G. (2003). Electronic word-of-mouth: Motives for and consequences of reading customer articulations on the Internet. *International Journal of Electronic Commerce*, 8(2), 51-74.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, 18(1), 38-52.

Hlee, S., Lee, J., Yang, S.B., Koo, C. (2016). An empirical examination of online restaurant reviews (yelp. com): Moderating roles of restaurant type and self-image disclosure, in: *Information and communication technologies in tourism*, Springer, pp. 339–353.

Ho, J. Y., and Dempsey, M. (2010). Viral marketing: Motivations to forward online content. *Journal of Business research*, 63(9-10), 1000-1006.

Hou, L., Liu, K., and Liu, J., (2017). Navigated Random Walks on Amazon Book Recommendation Network. In: *International Workshop on Complex Networks and their Applications*, Lyon, France, pp. 935-945.

Hou L., Pan X., Guo Q. & Liu J.G., (2014). Memory effect of the online user preference. Scientific Report, 4, 6560.

Hu, Y., & Li, X. (2011). Context-dependent product evaluations: an empirical analysis of internet book reviews. *Journal of Interactive Marketing*, 25(3), 123-133.

Hu, H. B., Wang, K., Xu, L., and Wang, X. F. (2008). Analysis of online social networks based on complex network theory. *Complex Systems and Complexity Science*, 2, 1214.

Huang, C. C., Lin, T. C., and Lin, K. J. (2009). Factors affecting pass-along email intentions (PAEIs): Integrating the social capital and social cognition theories. *Electronic Commerce Research and Applications*, 8(3), 160-169.

Huang, A.H., Chen, K., Yen, D.C., Tran, T.P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior, 48*,17–27.

Huang, G. C., Unger, J. B., Soto, D., Fujimoto, K., Pentz, M. A., Jordan-Marsh, M., & Valente, T. W. (2014). Peer influences: the impact of online and offline friendship networks on adolescent smoking and alcohol use. *Journal of Adolescent Health*, 54(5), 508-514.

Hughes, J. and W. Sharrock (1997). *The Philosophy of Social Research*. Essex, Pearson

Iribarren, J. L., & Moro, E. (2009). Impact of human activity patterns on the dynamics of information diffusion. *Physical Review Letters*, 103(3), 038702.

Ito, T.A., Larsen, J.T., Smith, N.K., Cacioppo, J.T. (1998). Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology,* 75, 887–900.

Jacoboni, C., & Reggiani, L. (1983). The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. *Reviews of modern Physics*, *55*(3), 645.

James J. 1993. New technologies, employment and labor markets in developing countries. *Dev. Change* 24:405–37

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, *60*(11), 2169-2188.

Jun, M., & Cai, S. (2001). The key determinants of internet banking service quality: a content analysis. International journal of bank marketing, 19(7), 276-291.

Jung, T. H., Ineson, E. M., & Green, E. (2013). Online social networking: Relationship marketing in UK hotels. *Journal of Marketing Management*, *29*(3-4), 393-420.

Kawachi, I., Berkman, L.F., (2001). Social ties and mental health. *Journal of Urban Health*, 78 (3), 458–467.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, *90*(430), 773-795.

Kenny, C. B. (1992). Political participation and effects from the social environment. *American Journal of Political Science*, 259-267.

Khare, A., Labrecque, L. I., and Asare, A. K. (2011). The assimilative and contrastive effects of word-of-mouth volume: An experimental examination of online consumer ratings. *Journal of Retailing*, 87(1), 111-126.

Kiel, G. C., & Layton, R. A. (1981). Dimensions of consumer information seeking behavior. *Journal of marketing Research*, 233-239.

Kim, E. E. K., Mattila, A. S., and Baloglu, S. (2011). Effects of gender and expertise on consumers' motivation to read online hotel reviews*. Cornell Hospitality Quarterly,* 52(4), 399-406.

Kim, S. M., Pantel, P., Chklovski, T., and Pennacchiotti, M. (2006, July). Automatically assessing review helpfulness. In: *Proceedings of the 2006 Conference on empirical methods in natural language processing* (pp. 423-430). Association for Computational Linguistics.

Kim, H. and Kwon, S. (1999) An exploratory research on lifestyles and purchase decision making of internet users. *Korean Management Review*, 28 (2), 353–372.

King, R. A., Racherla, P., and Bush, V. D. (2014). What we know and don't know about online word-of-mouth: A review and synthesis of the literature. *Journal of Interactive Marketing*, 28(3), 167-183.

King, W. R., Grover, V., and Hufnagel, E. H. (1989). Using information and information technology for sustainable competitive advantage: Some empirical evidence. *Information & Management*, 17(2), 87-93.

King, W.R. and Teo, T.S.H. (1994) Facilitators and inhibitors for the strategic use of information technology. *Information and Management,* 27 (2), 71–88.

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., and Makse, H. A. (2010). Identification of influential spreaders in complex networks. Nature Physics, 6(11), 888.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3(3), 383-397.

Kollock, P. (1999). The economies of online cooperation. *Communities in cyberspace*, 220-239.

Koul, A., Becchio, C., & Cavallo, A. (2018). Cross-validation approaches for replicability in psychology. *Frontiers in Psychology*, *9*, 1117.

Korfiatis, N., Garc A-Bariocanal, E., Sanchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications,* 11, 205–217.

Kossinets, G., and Watts, D. J. (2006). Empirical analysis of an evolving social network. Science, 311(5757), 88-90.

Kothari, C. R. (2004). *Research methodology: Methods and techniques.* New Age International.

Kozinets, R. V., De Valck, K., Wojnicki, A. C., & Wilner, S. J. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, 74(2), 71-89.

Krackhardt, D., Nohria, N., and Eccles, B. (2003). The strength of strong ties. *Networks in the knowledge economy*, 82.

Kreps, G. L. (2017). Diffusion Theory in Integrative Approaches. In *Oxford Research Encyclopedia of Communication*.

Krishnamoorthy, S. (2015). Linguistic features for review helpfulness prediction. *Expert Systems with Applications,* 42*,* 3751–3759.

Kruglanski, A.W., Chen, X., Pierro, A., Mannetti, L., Erb, H.P., Spiegel, S. (2006). Persuasion according to the unimodel: Implications for cancer communication. *Journal of Communication,* 56, S105–S122.

Kyritsis, M., Gulliver, S. R., & Feredoes, E. (2018). Acknowledging crossing-avoidance heuristic violations when solving the Euclidean travelling salesperson problem. *Psychological research*, *82*(5), 997-1009.

Lee, S., Choeh, J.Y. (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications*, 41, 3041–3046.

Lee, Y.J., Hosanagar, K., Tan, Y., (2015). Do I follow my friends or the crowd? Information cascades in online movie ratings. *Management Science*, 61 (9), 2241–2258.

Lee, M., and Youn, S. (2009). Electronic word of mouth (eWOM) How eWOM platforms influence consumer product judgement. *International Journal of Advertising*, 28(3), 473-499.

Lee, M., Rodgers, S. and Kim, M. (2009) Effects of valence and extremity of eWOM on attitude toward the brand and website*. Journal of Current Issues and Research in Advertising*, 31(2), 1–11.

Lee, J., and Kim, J. K. (2011). The impact of online brand community type on consumer's community engagement behaviors: Consumer-created vs marketer created online brand community in SNW. *Cyberpsychology, Behavior and Social Networking,* 14(1), 59–63.

Leem, B., and Chun, H., (2014). An Impact of Online Recommendation Network on Demand. *Expert Systems with Application*, 41(4), 1723-1729.

Leskovec, J., Adamic, L.A., Huberman, B.A., (2007). The dynamics of viral marketing. *ACM Transaction on the Web*, 1 (1), 1–39.

Lewis, K., Gonzalez, M., and Kaufman, J. (2012). Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1), 68-72.

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19 (4), 456-474.

Light, D., Wexler, D., & Heinz, C. (2005). Keeping teachers in the center: A framework for data-driven decision-making. In: Society for Information Technology & Teacher Education International Conference (pp. 128-133). Association for the Advancement of Computing in Education (AACE).

Liu, K. (2000). *Semiotics in information systems engineering*. Cambridge University Press.

Liu, K. and Li, W. (2015). *Organisational Semitics for Business Informatics*. Routledge.

Liu, K., Sun, L., & Tan, S. (2007). Using Problem Articulation Method to Assist Planning and Management of Complex Project. In: *Project Management and Risk Management in Complex Projects* (pp. 3-13). Springer, Dordrecht.

Liu, K., Sun, L., Barjis, J. and Dietz, J. (2003). Modelling dynamic behaviour of business organisations—extension of DEMO from a semiotic perspective, *Knowledge-Based Systems* 16(2): 101–111.

Liu, D., Brass, D., Lu, Y., and Chen, D. (2015). Friendships in online peer-to-peer lending: Pipes, prisms, and relational herding.

Liu J.G., Hu Z. and Guo Q. (2013). Effect of the social influence on topological properties of user-object bipartite networks. *The European Physical Journal B*, 86(11), 1–11.

Liu, L., Tang, J., Han, J., Jiang, M., and Yang, S. (2010, October). Mining topic-level influence in heterogeneous networks. In: Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 199-208). ACM.

Liu, Y., Huang, X., An, A., Yu, X. (2008). Modelling and predicting the helpfulness of online reviews, in: 2008 Eighth IEEE International Conference on Data Mining, IEEE. pp. 443–452.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, 70(3), 74-89.

Lin, Z., Goh, K. Y., and Heng, C. S., (2017). The Demand Effects of Product Recommendation Networks: An Empirical Analysis of Network Diversity and Stability. *MIS Quarterly*, 41(2), 397-426.

Liu, Z., Park, S. (2015). What makes a useful online review? implication for travel product websites. *Tourism Management,* 47, 140–151.

Lin, Z., and Wang, Q., (2018). E-commerce Product Networks, Word-of-mouth Convergence, and Product Sales. *Journal of the Association for Information Systems*, 19(1), 23-39.

Lorenz, J., Rauhut, H., Schweitzer, F., and Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108 (22), 9020-9025.

Lu, Y., Tsaparas, P., Ntoulas, A., and Polanyi, L. (2010, April). Exploiting social context for review quality prediction. In: *Proceedings of the 19th international conference on World wide web* (pp. 691-700). ACM.

Luo, J., Ba, S., & Zhang, H. (2012). The effectiveness of online shopping characteristics and well-designed websites on satisfaction. *MIS Quarterly,* 1131-1144.

Ma, X., Khansa, L., Deng, Y., Kim, S.S., (2013). Impact of prior reviews on the subsequent review process in reputation systems. *Journal of Management Information Systems,* 30 (3), 279–310.

Madill, A., Jordan, A., and Shirley, C. (2000). Objectivity and reliability in qualitative analysis: Realist, contextualist and radical constructionist epistemologies. *British Journal of Psychology*, 91(1), 1-20.

Mahajan V, Muller E. (1994). Innovation diffusion in a borderless global market: Will the 1992 unification of the European Community accelerate diffusion of new ideas, products, and technologies? *Technological Forecasting Social Change,* 45(3), 221–235.

Mangold, W. G., and Faulds, D. J. (2009). Social media: The new hybrid element of the promotion mix. *Business Horizons*, 52(4), 357-365.

Martin, L., Sintsova, V., Pu, P. (2014). Are influential writers more objective? an analysis of emotionality in review comments, in: *Proceedings of the 23rd International Conference on World Wide Web*, ACM. pp. 799–804.

Mateas M., Sengers P. (2003). *Narrative intelligence*. John Benjamins Publishing Company, Amsterdam.

McPherson, M., Smith-Lovin, L., Cook, J.M., (2001). Birds of a feather: homophily in social networks. *Annual Review Sociology,* 27 (1), 415–444.

Meier, R. C., Newell, W. T., and Pazer, H. L. (1969). *Simulation in business and economics.*

Meyer JW, Rowan B. (1977). Institutionalized organizations: formal structure as myth and ceremony. *American Journal Sociology,* 83(2), 340–363.

Miles, M.B. and Huberman, A. M., (1994), *Qualitative Data Analysis: An Expanded Sourcebook*, Thousand Oaks, CA: Sage publications.

Miyake, N., & Norman, D. A. (1979). To ask a question, one must know enough to know what is not known. *Journal of verbal learning and verbal behavior*, 18(3), 357-364.

Mizruchi MS. 1993. Cohesion, equivalence, and similarity of behavior: a theoretical and empirical assessment. *Social Network,* 15(3), 275– 307.

Moe, W. W., & Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, 31(3), 372-386.

Moore, G. C., and Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, 2 (3), 192-222.

Mouttapa, M., Valente, T., Gallaher, P., Rohrbach, L. A., and Unger, J. B. (2004). Social network predictors of bullying and victimization. *Adolescence*, 39 (154), 315.

Muchnik, L., Aral, S., and Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, 341 (6146), 647-651.

Mudambi, S. M., and Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS Quarterly*, 185-200.

Muir, L., and Douglas, A. (2001). Advent of e-business concepts in legal services and its impact on the quality of service. *Managing Service Quality: An International Journal*, 11 (3), 175-181.

Muniz Jr, A. M., and Schau, H. J. (2005). Religiosity in the abandoned Apple Newton brand community. *Journal of Consumer Research*, 31(4), 737-747.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142.

Nambisan, S., and Baron, R. A. (2007). Interactions in virtual customer environments: Implications for product support and customer relationship management. *Journal of Interactive Marketing*, 21(2), 42-62.

Newell S, Swan J. (1995). Professional associations as important mediators of the innovation process. *Science Communication,* 16(4), 371–387.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM Review*, 45(2), 167-256.

Ni, X. P., Ruan, Q. T., Mei, S. W., and He, G. (2007). A new network partitioning algorithm based on complex network theory and its application in Shanghai power grid. *Power System Technology*, 9(31), 6-12.

Ngo-Ye, T.L., Sinha, A.P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 61, 47–58.

Noelle-Neumann, E., (1974). The spiral of silence a theory of public opinion. *Journal of Communication*, 24 (2), 43–51.

Oestreicher-Singer, G., and Sundararajan, A., (2012a). Recommendation Networks and the Long Tail of Electronic Commerce. *MIS Quarterly,* 36(1), 65-83.

Oestreicher-Singer, G., and Sundararajan, A., (2012b). The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets. *Management Science*, 58(11), 1963-1981.

Oestreicher-Singer, G., Libai, B., Sivan, L., Carmi, E., and Yassin, O., (2013). The Network Value of Products. *Journal of Marketing*, 77(3), 1-14.

Oldenburg, B., and Glanz, K. (2008). Diffusion of innovations. In K. Glanz, B. K. Rimer, & F. K. Viswanath (Eds.), Health behavior and health education: Theory, research, and practice, (4th ed., pp. 303–333). San Francisco: Jossey-Bass.

Oliver, M. I., Pearson, N., Coe, N., and Gunnell, D. (2005). Help-seeking behaviour in men and women with common mental health problems: cross-sectional study. *The British Journal of Psychiatry*, 186(4), 297-301.

Onnela J.P. and Reed-Tsochas F. (2010). Spontaneous emergence of social influence in online systems. *Proceedings of the National Academy of Sciences*, 107(43). 18375–18380.

O'Reilly, K., and Marx, S. (2011). How young, technical consumers assess online WOM credibility. *Qualitative Market Research: An International Journal,* 14(4), 330-359.

Ormrod RK. (1990). Local context and innovation diffusion in a well-connected world. *Economic Geography,* 66(2), 109–122.

Palmer DA, Jennings PD, Zhou X. (1993). Late adoption of the multidivisional form by large U.S. corporations: institutional, political, and economic accounts. *Administrative Science Quarterly*, 38, 100–131.

Pal, A., and Counts, S. (2011, February). Identifying topical authorities in microblogs. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 45-54). ACM.

Papadopoulos F., Kitsak M., Serrano M.A., Boguná M. and Krioukov D. (2012). Popularity versus similarity in growing networks. *Nature,* 489(7417), 537–540.

Pan X., Hou L., Stephen M. and Yang H. (2014). Long-term memories in online users' selecting activities. *Physics Letters A*, 378(35), 2591–2596.

Pan, X., Hou, L., Liu, K., (2017). Social influence on selection behaviour: distinguishing local- and global-driven preferential attachment. PloS ONE 12, e0175761.

Pan, Y., Zhang, J.Q. (2011). Born unequal: a study of the helpfulness of user- generated product reviews. *Journal of Retailing,* 87, 598–612.

Pang, B., Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, in: Proceedings of the 42nd annual meeting on Association for Computational Linguistics, Association for Computational Linguistics. pp. 271–279.

Park, C., Lee, T.M., (2009a). Information direction, website reputation and ewom effect: A moderating role of product type. *Journal of Business Research*, 62, 61–67.

Park, D. H., Lee, J., and Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce*, 11(4), 125-148.

Park, S., Nicolau, J.L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research, 50*, 67–83.

Park, S.B., Park, D.H., (2013). The effect of low-versus high-variance in product reviews on product evaluation. *Psychology & Marketing*, 30 (7), 543–554.

Payne, J., Bettman, J. R., and Johnson, E. J. (1991). Consumer decision making. *Handbook of Consumer Behaviour*, 50-84.

Peck, R. S., L. Y. Zhou, V. B. Anthony, K. Madhukar. (2008). *Consumer Internet, Bear Stearns equity research report*. Bear Stearns, New York.

Peirce, C. (1960). *Collected Papers of Ch. S. Peirce (1931-1935).* Cambridge, Mass, Harvard University Press.

Picazo-Vela, S., Chou, S. Y., Melcher, A. J., and Pearson, J. M. (2010). Why provide an online review? An extended theory of planned behavior and the role of Big-Five personality traits. *Computers in Human Behavior*, *26*(4), 685-696.

Rabjohn, N., Cheung, C. M., and Lee, M. K. (2008, January). Examining the perceived credibility of online opinions: information adoption in the online environment. In: Hawaii international conference on system sciences, proceedings of the 41st annual (pp. 286-286). IEEE.

Racherla, P., Friske, W. (2012). Perceived "usefulness" of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications,* 11, 548–559.

Rambaut, A., & Grass, N. C. (1997). Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics*, *13*(3), 235-238.

Ratkiewicz J., Fortunato S., Flammini A., Menczer F. and Vespignani A. (2010). Characterizing and modeling the dynamics of online popularity. *Physical Review Letter*, 105(15), 158701.

Raymond, E. (2001). *The Cathedral and the Bazaar.* O'Reilly, Sebastopol, Sebastopol, CA.

Resnick, P., R. Zeckhauser, E. Friedman, K. Kuwabara. (2000). Reputation systems. Comm. ACM 43(12) 45–48.

Ritchie, J., Lewis, J., Nicholls, C. M., and Ormston, R. (Eds.). (2013). *Qualitative research practice: A guide for social science students and researchers*. sage.

Rodriguez, M. G., Balduzzi, D., and Schölkopf, B. (2011). Uncovering the temporal dynamics of diffusion networks. arXiv preprint arXiv:1105.0697.

Rogers, E. M. (2010). *Diffusion of innovations. Simon and Schuster.*

Romero, D. M., Galuba, W., Asur, S., and Huberman, B. A. (2011, September). Influence and passivity in social media. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 18-33). Springer, Berlin, Heidelberg.

Ryan B. and Gross N. (1943). The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology,* 8(1), 15–24.

Saarinen, T. (1996). An expanded instrument for evaluating information system success. *Information and Management*, 31(2), 103–18.

Salehan, M., Kim, D.J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems,* 81, 30–40.

Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762), 854-856.

Saltiel J, Bauder JW, Palakovich S. (1994). Adoption of sustainable agricultural practices: diffusion, farm structure, and profitability. *Rural Sociology,* 59(2), 333–49.

Saussure, F. (1916). *Nature of the linguistic sign.*

Schau, H. J. (2002). Brand communities and personal identities: Negotiations in cyberspace. ACR North American Advances.

Schaninger, C. M., and Sciglimpaglia, D. (1981). The influence of cognitive personality traits and demographics on consumer information acquisition. *Journal of Consumer Research*, 8(2), 208-216.

Schieman, S., Van Gundy, K., (2000). The personal and social links between age and self-reported empathy. *Social Psychology Quarterly*, 63 (2), 152–174.

Schlosser, A.E., (2005). Posting versus lurking: communicating in a multiple audience context. *Journal of Consumer Research*, 32 (2), 260–265.

Schmidt, J. B., & Spreng, R. A. (1996). A proposed model of external consumer information search. *Journal of the academy of Marketing Science*, 24(3), 246-256.

Scott, S., Plotnikoff, R., Karunamuni, N., Bize, R., & Rodgers, W. (2008). Factors influencing the adoption of an innovation: An examination of the uptake of the Canadian Heart Health Kit (HHK). *Implementation Science*, 3(41).

Seidman, S. B. (1983). Network structure and minimum degree. *Social networks*, *5*(3), 269-287.

Sellers, K. F., & Shmueli, G. (2010). A flexible regression model for count data. *The Annals of Applied Statistics*, *4*(2), 943-961.

Senecal, S., and Nantel, J., (2004). The Influence of Online Product Recommendations on Consumers' Online Choices. *Journal of Retailing*, 80(2), 159-169.

Shamma, D. A., Kennedy, L., and Churchill, E. F. (2011, March). Peaks and persistence: modeling the shape of microblog conversations. In: *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 355-358). ACM.

Shim, S., Eastlick, M. A., Lotz, S. L., and Warrington, P. (2001). An online pre-purchase intentions model: the role of intention to search: best overall paper award—the sixth triennial AMS/ACRA retailing conference, 2000☆. *Journal of Retailing*, 77(3), 397-416.

Silva, T. C., and Zhao, L. (2016). *Machine learning in complex networks* (Vol. 2016). Switzerland: Springer.

Sinha, R.R., Swearingen, K., (2001). Comparing recommendations made by online systems and friends. In: Proceedings of the 2nd DELOS Network of Excellence Workshop on Personalisation and Recommender Systems in Digital Libraries, Dublin, Ireland pp. 18–20.

Sohn, D. (2009). Disentangling the effects of social network density on electronic word-of-mouth (eWOM) intention. *Journal of Computer-Mediated Communication*, 14(2), 352-367.

Sporns, O. (2011). The human connectome: a complex network. *Annals of the New York Academy of Sciences*, 1224(1), 109-125

Srinivasan, N. (1987). *A path analytic model of external search for information for new automobiles*. *ACR North American Advances*.

Stamper, R., Liu, K., Hafkamp, M. and Ades, Y. (2000). Understanding the Roles of Signs and Norms in Organisations, *Behaviour & Information Technology* 19(1): 15–27.

Staff, M., 2007a. Global survey world-of-mouth the most powerful selling tool. URL: http://www.marketingcharts.com/television/ global-survey-word-of-mouth-the-most-powerful-selling-tool -1884/.

Staff, M., 2007b. Most consumers read and rely on online reviews. URL: http://www.marketingcharts.com/online/most-consumers-read-and-rely-on-online-reviews-companies-must-adjust- 2234/.

Strang D, Meyer JW. (1993). Institutional conditions for diffusion. *Theory and Society*, 22(4), 487–511.

Straub DW. (1994). The effect of culture on IT diffusion: e-mail and FAX in Japan and the U.S. *Information System Research,* 5(1), 23–47.

Strauss, A. and Corbin, J., (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, Sage Publications, Thousand Oaks, CA

Steffes, E.M., Burgee, L.E., (2009). Social ties and online word of mouth. *Internet Research*, 19 (1), 42–59.

Sun, M., (2012). How does the variance of product ratings matter? *Management Science*, 58 (4),696–707.

Sun, T., Youn, S., Wu, G., and Kuntaraporn, M. (2006). Online word-of-mouth (or mouse): An exploration of its antecedents and consequences. *Journal of Computer-Mediated Communication*, 11(4), 1104-1127.

Sundaram, D. S., Mitra, K., and Webster, C. (1998). Word-of-mouth communications: A motivational analysis. *Advances in Consumer Research*, 25(1), 527-531.

Susarla, A., Oh, J. H., and Tan, Y. (2012). Social networks and the diffusion of user-generated content: Evidence from YouTube. *Information Systems Research*, 23(1), 23-41.

Sweeney, J. C., Soutar, G. N., and Mazzarol, T. (2008). Factors influencing word of mouth effectiveness: receiver perspectives. *European Journal of Marketing*, 42(3/4), 344-364.

Tang, J., Sun, J., Wang, C., and Yang, Z. (2009, June). 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). ACM.

Tang, J., Wang, Y., and Liu, F. (2013). Characterizing traffic time series based on complex network theory. *Physica A: Statistical Mechanics and its Applications*, 392(18), 4192-4201.

Thiery, J. P., and Sleeman, J. P. (2006). Complex networks orchestrate epithelial–mesenchymal transitions. *Nature Reviews Molecular Cell Biology,* 7(2), 131.

Thorson, K. S., and Rodgers, S. (2006). Relationships between blogs as eWOM and interactivity, perceived interactivity, and parasocial interaction. *Journal of Interactive Advertising*, 6(2), 5-44.

Tolnay SE. (1995). The spatial diffusion of fertility: a cross-sectional analysis of counties in the American South, 1940. *American Sociological Review,* 60, 299–308.

Travers, J., and Milgram, S. (1967). The small world problem. *Phychology Today,* 1(1), 61-67.

Tucker, C., and Zhang, J. (2011). How does popularity information affect choices? A field experiment. *Management Science*, 57(5), 828-842.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. Icwsm, 10(1), 178-185.

Underhill, P. (2009). *Why we buy: The science of shopping--updated and revised for the Internet, the global consumer, and beyond*. Simon and Schuster.

Valente, T. W., Fujimoto, K., Chou, C. P., and Spruijt-Metz, D. (2009). Adolescent affiliations and adiposity: a social network analysis of friendships and obesity. *Journal of Adolescent Health*, 45(2), 202-204.

Valenzuela, S., Park, N., and Kee, K. F. (2009). Is there social capital in a social network site? Facebook use and college students' life satisfaction, trust, and participation. *Journal of computer-mediated communication*, 14(4), 875-901.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

Vollmer, C., and Precourt, G. (2008). *Always on: Advertising, marketing, and media in an era of consumer control.* New York: McGraw-Hill.

Wang, J.C., Chang, C.H., (2013). How online social ties and product-related risks influence purchase intentions: a Facebook experiment. *Electronic Commerce Research and Applications*, 12 (5), 337–346.

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature*, 393(6684), 440.

Weimann G, Hans-Bernd B. (1994). Is there a two-step flow of agenda-setting? *International Journal of Public Opinion Research*, 6(4), 323–341.

Weiss, A. M., Lurie, N. H., and MacInnis, D. J. (2008). Listening to strangers: whose responses are valuable, how valuable are they, and why? *Journal of marketing Research*, 45(4), 425-436.

Wejnert, B. (2002). Integrating models of diffusion of innovations: A conceptual framework. *Annual review of sociology*, 28(1), 297-326.

Wellman, B., and Berkowitz, S. D. (Eds.). (1988). *Social structures: A network approach* (Vol. 2). CUP Archive.

Wetherell, C., Plakans, A., & Wellman, B. (1994). Social networks, kinship, and community in Eastern Europe. *The Journal of Interdisciplinary History*, *24*(4), 639-663.

Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P., and Zhao, B. Y. (2009, April). User interactions in social networks and their implications. In: *Proceedings of the 4th ACM European conference on Computer systems* (pp. 205-218) Nuremberg, Germany.

Wojnicki, A., & Godes, D. (2008). Word-of-mouth as self-enhancement. Woking paper.

Wu, P.F. (2013). In search of negativity bias: An empirical study of perceived helpfulness of online reviews. *Psychology & Marketing,* 30, 971–984.

Yin, D., Mitra, S., Zhang, H. (2016). Research note-when do consumers value positive vs. negative reviews? an empirical investigation of confirmation bias in online word of mouth. *Information Systems Research,* 27, 131–144.

Ying, Y., Feinberg, F., Wedel, M., (2006). Leveraging missing ratings to improve online recommendation systems. *Journal of Marketing Research*, 43 (3), 355–365.

Zajonc, R.B., (1980). Feeling and thinking: preferences need no inferences. *American Psychologist*, 35 (2), 151–175.

Zhang, L., Fang, H., Ng, W.K., Zhang, J., (2011). Intrank: Interaction ranking-based trustworthy friend recommendation. In: Proceedings of 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). IEEE, Changsha, China, pp. 266–273.

Zhao, S., Grasmuck, S., and Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior*, 24(5), 1816-1836.

Zhou, X. (2009). The political blogosphere in China: A content analysis of the blogs regarding the dismissal of Shanghai leader Chen Liangyu. *New Media & Society*, 11(6), 1003-1022.

Zhou, S., Guo, B. (2015). The interactive effect of review rating and text sentiment on review helpfulness, in: International Conference on Electronic Commerce and Web Technologies, Springer. pp. 100–111.

Zhu, F., Zhang, X., (2010). Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. *Journal of Marketing,* 74 (2), 133–148.

Appendix

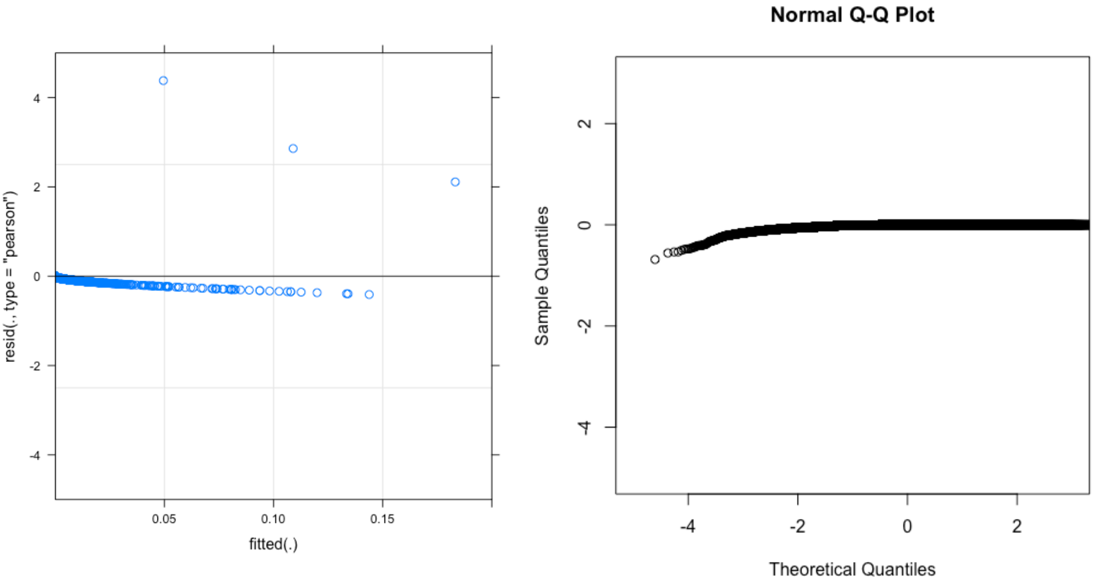


Figure A.1| The residual plot and QQ plot for model 3 in Chapter 5.

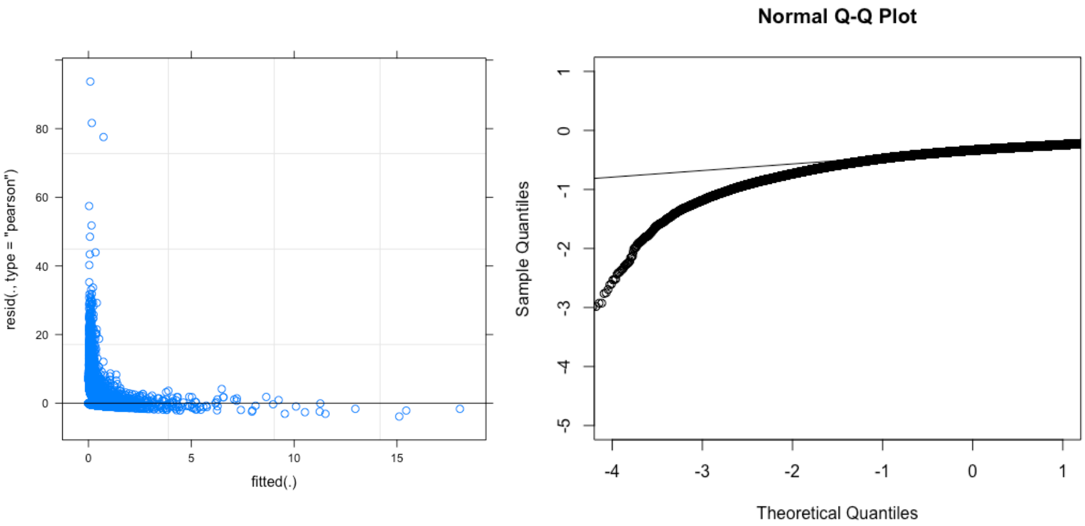


Figure A.2| The residual plot and QQ plot for model 1 in Chapter 7.

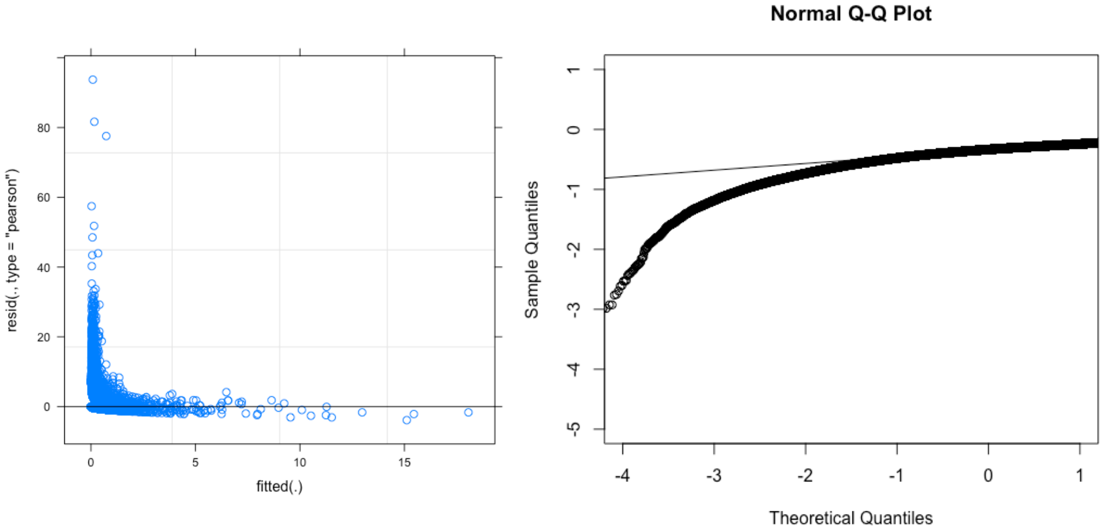


Figure A.3| The residual plot and QQ plot for model 4 in Chapter 7.

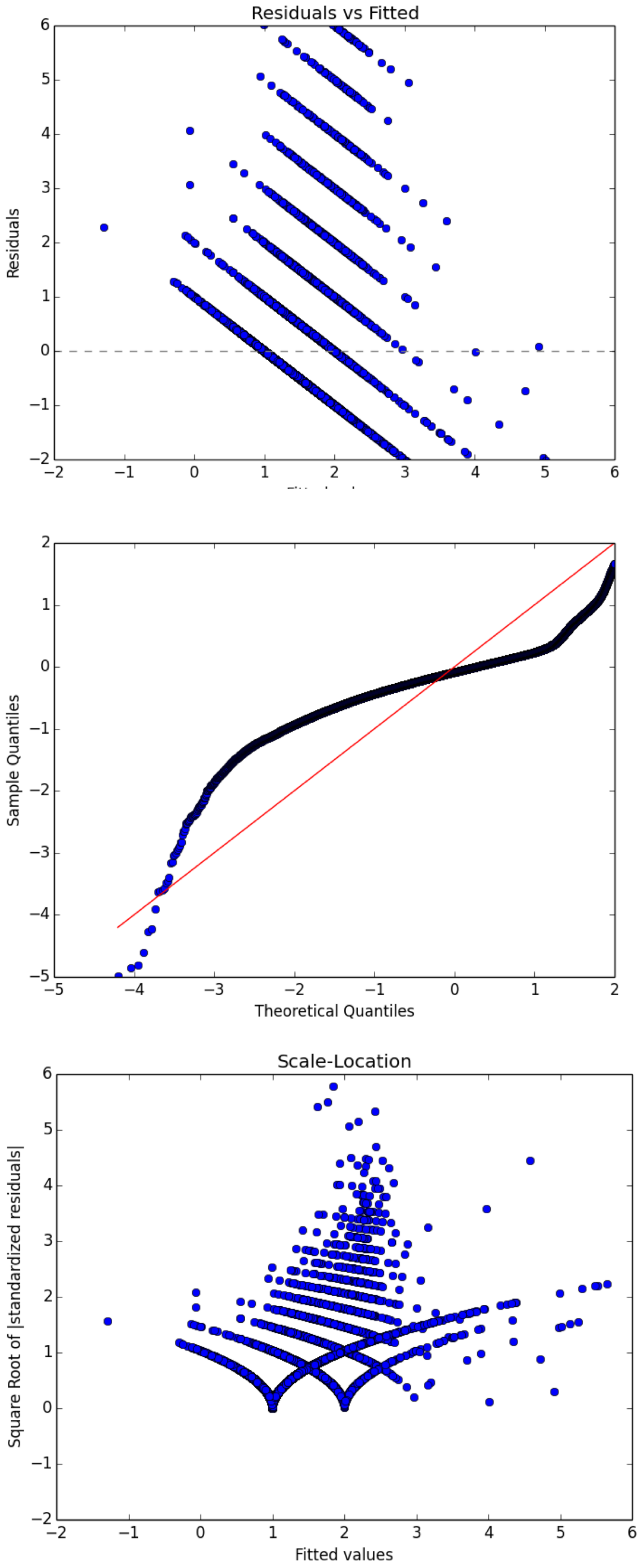


Figure A.4| The model diagnostics for model 4 in Chapter 8.

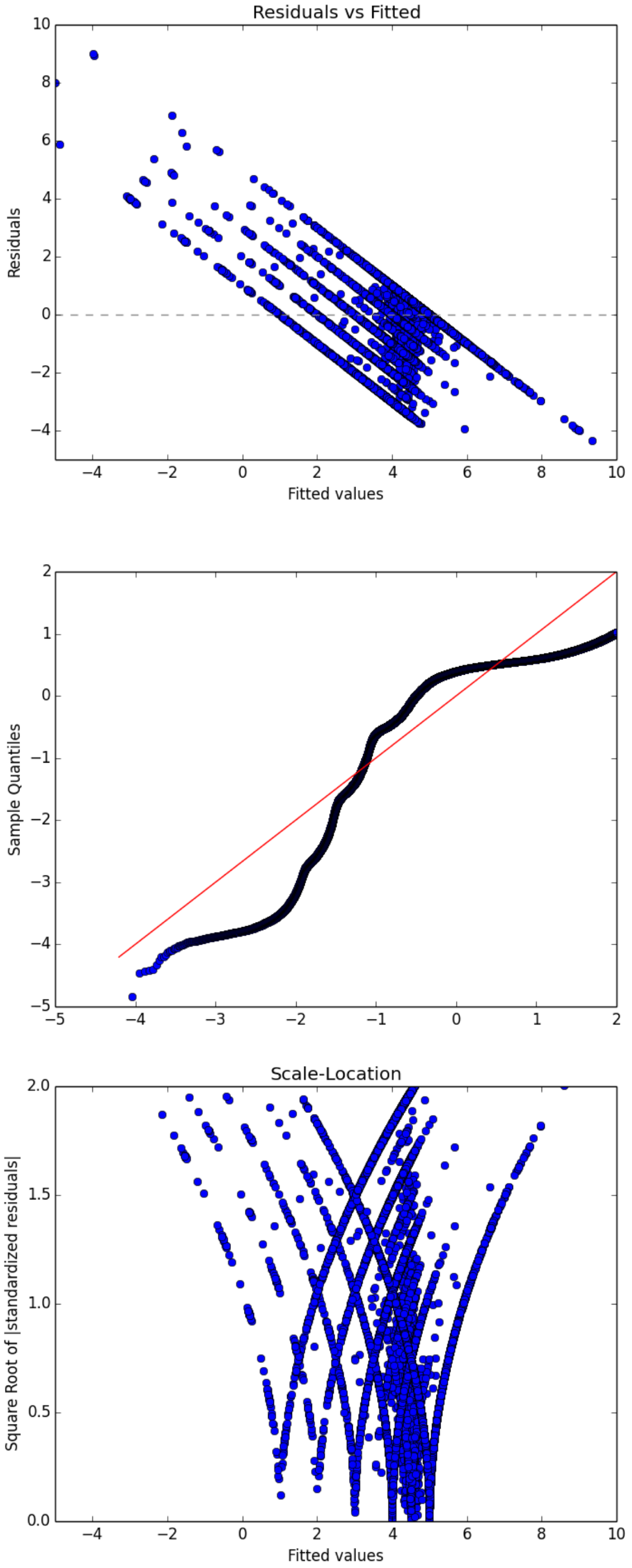


Figure A.5| The model diagnostics for model 8 in Chapter 8.