M.S. Dissertation in Engineering

A Promising Structural Importance Heuristic for Evaluating Online Reviews

February 2018

Graduate School of Seoul National University

Technology Management, Economics and Policy Program

LIU YU-LIN

A Promising Structural Importance Heuristic for Evaluating Online Reviews

지도 교수 Prof. Jörn Altmann

이 논문을 공학석사 학위논문으로 제출함
2018 년 2 월

서울대학교 대학원
협동과정 기술경영경제정책전공
LIU YU-LIN

리우위린의 석사 학위논문을 인준함
2018 년 2 월

위 원 장 이정동 (인)

부위원장 Jörn Altmann (인)

위 원 Netsanet Haile (인)

Abstract

A Promising Structural Importance Heuristic for Evaluating Online Reviews

LIU YU-LIN

Technology Management, Economics and Policy Program

College of Engineering

Seoul National University

Representing a dramatic increase in the number of online customers and the volume of reviews, online e-commerce markets with the most successful business models continue to expand. Apart from the information offered by sellers about their products, online customer reviews are the only other information source available. Therefore, online customer reviews are a particularly useful source of information to evaluate the product quality. However, two issues arise here though: First, due to limited time, it is difficult for users to read all of the reviews. Second, the credibility of online customer reviews is frequently problematic. As demonstrated by the extant research, some sellers' recognition of the power of online customer reviews for their business growth leads those sellers to recruit people to issue fake reviews on their products or services. To address this issue, in the present study, we propose a solution that combines the structural importance of reviewers and their responses towards existing reviews. Structural importance shows the importance of reviewers within social networks, enabling users to build online relationships and process information (e.g. create, read, share, re-share, comment) among them. In order to quantify structural importance of

reviewers in a network, various centrality measures are used. Reviewers' responses point to their reactions towards other reviewers' comments. Combining these two sources of information is expected to be a promising structural importance heuristic that is based not solely on the connectedness within a network, but also considers node's attributes. Finally, we investigate how the proposed solution can help end-users in identifying and ranking the most relevant and accurate reviews within online e-commerce platforms. The proposed solutions are planned to be run and evaluated on extensive data (i.e., millions of nodes and links and published crowd-sourced reviews) collected from Yelp.com, one of the most widely used and successful online ecommerce platforms.

Keywords: Social Network Analysis, Centrality Measures, Recommender Systems, Online Shopping System, Reviews, Data Mining, Graph Mining

Student Number: 2016-22088

Contents

List of Tables

List of Figures

Chapter 1. Introduction

The Internet changed the way how we communicate with each other, enabling us to post texts, photos, and videos to express our feelings and opinions. The information directly created by users is referred to as user-generated content (Jindal & Liu, 2008). This content differs from the one provided by website owners (e.g., companies or organizations). Since the appearance of e-commerce, this user-generated content has acquired a paramount importance.

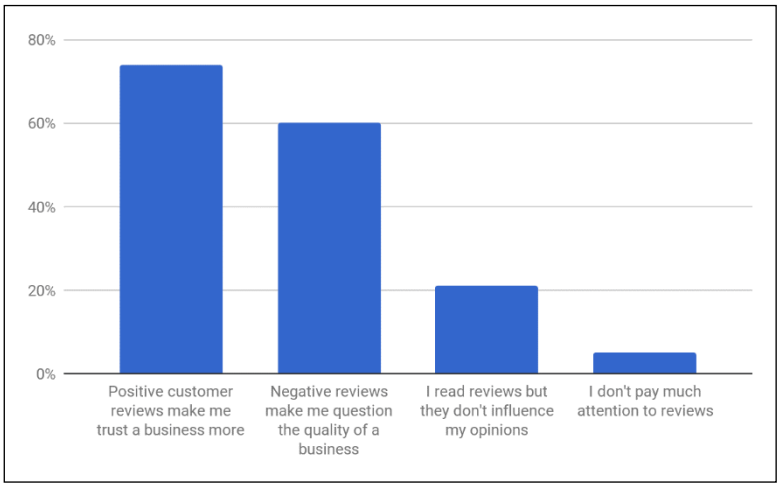
With the rapid development of e-commerce, online reviews have become more and more valuable for customers to make decisions. Based on the 2016 survey report from Statistia, a famous German company that focuses on online statistics, market research and business intelligence, 74 percent of the respondents thought that positive customer reviews made them trust a business more. Therefore, reviews play an important role in customer purchasing decisions.

1.1 Research Motivation

Representing a dramatic increase in the number of online customers and the volume of reviews, online e-commerce markets with the most successful business models continue to expand. Besides the information offered by sellers about their products, online customer reviews are the only other information available. This makes online customer reviews a

particularly useful information source to evaluate the product quality. According to the results of the October 2017 survey by Statistia (see Figure 1), 73 percent of the respondents felt that positive customer reviews made them trust a business more. Furthermore, 60 percent of the respondents felt that negative customer reviews made them question the quality of a business. Therefore, the reviews appear to be one of the most important and objective sources information for consumers' purchasing decisions. However, due to limited time, it is difficult for users to read all of the reviews. Therefore, finding a high-quality review among an extensive body of reviews has become another serious problem that needs to be solved. Moreover, the credibility of online customer reviews is another issue that needs be addressed.

Figure 1. How do online customer reviews affect your opinion of a local business?



(Source: Statistia, 2017)[1]

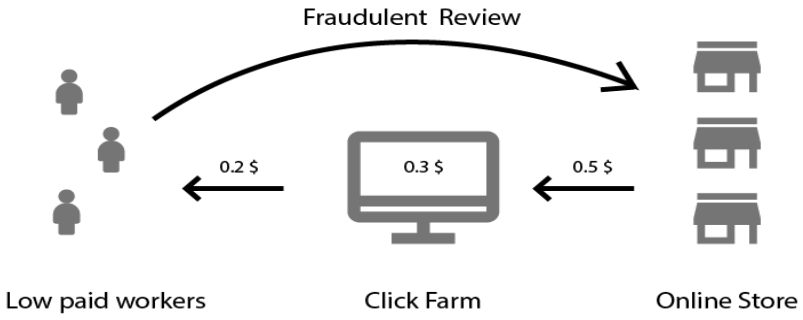
1.2 Problem description

Previous studies have demonstrated that spammers have more financial incentives to write review spam rather than email spam and web spam. Acknowledging the importance of reviews and their ability to generate significant revenues, more and more reviewers are paid or receive other bonuses (e.g., coupons) to write fake reviews to boost a business. As a result, review spams has become a much more serious problem than ever before. Previous research has shown that one more star rating (in the evaluation system with the maximum of 5 and the minimum of 0 stars) will increase the revenue by 5-9% (Luca, 2016). Accordingly, writing spam reviews has become an industry all over the world; in developing countries, a phenomenon called “Click Farm”, which is a form of click fraud, has emerged where a large group of low-paid workers are recruited to click on paid advertising links for the click fraudster. In Dhaka, Bangladesh, the price of per thousand "likes" at "click farm" is only $15. Workers might be on a three-shift working system, and only be paid as little as $120 a year (Arthur, 2013). Figure 2 shows how this click farm works. Online stores pay the click

[1] https://www.statista.com/statistics/315751/online-review-customer-opinion/

farms for recruiting low paid workers to write fraudulent reviews in order to attract customers or to slander competitors.

Figure 2. How click farm works?



In response to this problem, several studies have investigated opinion spam and trustworthiness of online opinions (e.g., Jindal & Liu, 2008). Since then, different methodologies with different perspectives—including, among others, text mining, reviews behaviors' analysis, and the review-centric approach—have been used to detect review spams.

1.3 Research objective and research question

The main aim of the present study is to propose a new solution that combines the structural importance of reviewers and their behaviors. In order to achieve this goal, the following three main research questions need be

addressed:

(1) How can one measure the structural importance of reviewers?

(2) How can one judge whether a review is true or fake?

(3) Is there a correlation between structural features of reviewers and the authenticity of reviews?

1.4 Methodology

To answer the three research questions outlined in Section 1.3, we will first find an appropriate way to measure the structural importance of reviewers based on previous research. Based on existing research in the area of social network analysis and in view of the computation complexity of very large networks, we selected the four key measurement variables that may have an effect on detecting fake reviews. These variables include clustering coefficient (Watts & Strogatz, 1998), closeness centrality (Bavelas, 1950), PageRank (Page, Brin, Motwani, & Winograd, 1999), and HITS (Kleinberg, 1999). Secondly, the dataset of Yelp, which has not only its own review recommendation system to detect deceptive reviews, but also its own social network platform, was chosen However, the problem is that Yelp does not specify how its review recommendation system works, as it is a business secret. Therefore, we had to create our own review spam detection model.

On an extensive literature review and deriving from experience, we used "Yelp-CHI" dataset (Mukherjee, Venkataraman, Liu, & Glance, 2013) to train and built our own spam review detection model by using three different classification methodologies in the area of machine learning—namely, SVM, Bayesian and Logistic. All of them show precision rates above 80%. Finally, based on this detection model, we used it to analyze the Yelp's official and real dataset and to get not only both true and deceptive reviews' information, but also their social network structural information. Since fake vs. non-fake is a binary dependent problem, we decided to use probit statistical model to find the relativity between fake reviews and reviewers' social network structural information.

1.5 Contribution

The statistical results showed that reviewers' social structural importance could be one of important factors that have an influence on detecting whether or not a review is fake. Although there are some other limitations in this research, such as, for instance, the limitation of dataset, the lower calculation ability of huge and complex network, we believe that our research results provide a promising structural importance heuristic solution and ideas for evaluating online reviews.

The remainder of this paper is structured as follows. Section 2 overviews extant research on review spam. In Section 3, we describe our model and present essential variables that will be used in the later experiment. In Section 4, the analysis results are presented and the theoretical relationship between reviewers' network structure and fake reviewers is discussed. Section 5 draws conclusions from the research findings and discusses future research directions.

Chapter 2. Literature Research

2.1 Incentives of Review Spams

The more consumers rely on online reviews and online ratings, the more incentives a company has to hire people to produce fake reviews to promote their goods or services. It has been estimated that more than 20% of Yelp's reviews are fake (Donfro, n.d.). There are lots of different incentives for consumers to write fake reviews. In prior research, two kinds of factors have been identified: self-benefiting (discount in future purchasement) and other-benefiting (contribution charity) (Choi, Mattila, Van Hoof, & Quadri-Felitti, 2016). Moreover, it has been concluded that promotional reviewing is sufficiently economically important. (Mayzlin, Dover, & Chevalier, 2014). Figure 3 shows one example posted on the Internet for hiring people with a high reward to write fake reviews on Yelp.

Figure 3. Get Paid To Write Fake Reviews For Yelp?


CL > new york > manhattan > all jobs > writing/editing jobs

Reply flag: miscategorized prohibited spam best of

Yelp review $25 / $50

We are looking for established Yelp accounts with over 50 reviews (please link Yelp account) to write well-written reviews for a restaurant. Many of these restaurants have a bi-polar review history (mostly positive 4's and 5's but a couple unfiltered 1's dragging them down, either from competitors or disgruntled ex-staff) and need a few 5's to rebuild their rating back. If this is something you'd be interested in, let us know.

The price is a Paypal transfer of $25 for the review, and another $25 to cut and paste that same review onto a couple other social media websites.

(Source: Evan V. Symon, 2016[2])

2.2 Types of Review Spams

Since 2008, researchers started to address this problem, and define this kind of reviews as opinion spams. (Jindal & Liu, 2008). They firstly summarized that there were three main types of spam reviews. The first type is untruthful opinions. Its objective is to mislead readers or consumers by giving underserving positive reviews, in order to promote some business services or products. The second type is reviews on brands only. This kind of reviews does not comment on specific product or business service but on the whole brands. The third type is named non-reviews, which includes two main sub-types, advertisements and other irrelevant reviews expressing no

[2] http://www.cracked.com/personal-experiences-2376-i-get-paid-to-write-fake-reviews-amazon.html

opinions. By analyzing the dataset of Amazon with logistic regression model, Jindal & Liu showed the effectiveness of this model and give one possible solution to solve this problem (Jindal & Liu, 2008).

2.3 Detection of Review Spams

Most of previous research has explored the detection of deception in online communities, and studies in computer science and information technology have developed automated identifier systems to filter out fake reviews (Mukherjee, Liu, & Glance, 2012). Current researchers mainly adopt three approaches to detect reviews: First, objective indicators are defined. For instance, researchers consider scores given by specific users and others. If there are big differences between a user's score and the average scores, it is defined to be a fake review (Li, Huang, Yang, & Zhu, 2011). The second approach conducts text analyses. Yoo and Gretzel (2009), who used traditional statistical analysis methodologies, conclude that deceptive and truthful reviews are different in terms of lexical complexity, the use of first person pronouns, the inclusion of brand names, and the expression of their sentiments (Yoo & Gretzel, 2009). Besides, based on semantic similarity, researchers also gave three possible solutions for detecting single review spammers. One extends the semantic similarity by using WordNet to

compute the relatedness between words. The other is based on topic modeling and exploits the similarity of the reviews topic instead of simple words. Sometimes, we cannot judge duplicated reviews by only counting how many words they used are identical, because different words could also express same or similar meaning (Sandulescu & Ester, 2015). The third perspective analyzes user behaviors. Rayana and Akoglu (Rayana & Akoglu, 2015) thought the users usually followed four principles below. (i) Benign users often write positive reviews to good-quality products and negative reviews to bad-quality products (ii) Spammers more likely write positive reviews to bad-quality products for promoting them and write negative reviews to good-quality products for misleading consumers (iii) It is still possible for spammers to write genuine-looking reviews to camouflage their activities (iv) However, it is less likely for benign users, to write genuine-looking reviews to camouflage their activities, because they have no reason to do that. Based on these assumptions, they made their SPEAGLE holistic framework and show its effectiveness in review spam detections (Rayana & Akoglu, 2015).

2.4 Comparative Analysis of Previous Review Spams Detection Methodologies

Table 1 shows some of the famous review spam methodologies and their precision rates, summarized by (Dixit & Agrawal, 2013). Some of them used review centric methods which mostly used text analysis as explained in the previous section, and others used reviewer centric methods by surveying users' behaviors. For instance, Lim et al. model spammers' rating behaviors and get an effective review spam detection way. Due to the popularity of Ecommerce, growing interest in mining opinions from reviews appears and researchers keep finding effective ways to solve the problem of review spams. The methods to detect review spams differ widely. Some assume that some of spammers work together to write fake reviews, so it must be effective to detect spammers groups (Mukherjee, Liu, Wang, Glance, & Jindal, 2011). Some researchers also thought spammers would not spend much time in writing fake reviews, they usually copy others' reviews and give target score. So they thought detecting duplicates reviews could be one of effective ways (Jindal & Liu, 2007). One of the most interesting perspectives is the "singleton review". Xie, Wang, Lin, & Yu found that more than 90% reviewers usually write only one review. In order to avoid being caught, spammers always post many reviews by using different accounts or names. This strategy makes most reviewers contribute only one review. Xie, Wang, Lin, & Yu thought these enormous singleton reviewers

were so enormous that they can almost determine a store's rating and impression. So they make a multi-scale anomaly detection algorithm based on multi-dimensional time series and prove its effectiveness (Xie, Wang, Lin, & Yu, 2012). One of the most famous research articles in the area of spam detection must be the paper "Opinion Spam and Analysis" by (Jindal & Liu, 2008). They use the dataset from Amazon including 5.8 million reviews and 2.14 million reviewers. It is almost the first time for researchers to use a large-scale real-world dataset. They detect three types of reviews, duplicate reviews, reviews on brands only, and non-reviews (like advertisement). They adopt logistic regression model with three types of features: review centric feature, reviewer centric feature and product centric feature and get a high precision rate, which means the fraction of relevant instances among the retrieved instances (Jindal & Liu, 2008).

Table 1. Comparative Analysis Summarize by Dixit & Agrawal in 2013

No.	Method	Reference	Precision
Review Centric Methods			
1	Opinion spam and analysis	(Jindal & Liu, 2008)	85%
2	Conceptual level Similarity Measure Based Review Spam Detection	(S. P. Algur, Patil, Hiremath, & Shivashankar, 2010)	43.64%
3	Spam Detection of customer Reviews from Web Pages	(S. Algur, Hiremath, Patil, & Shivashankar, 2010)	75.04%
4	Finding deceptive opinion spam by any stretch of the imagination	(Ott, Choi, Cardie, & Hancock, 2011)	83.30%
5	A method for sorting out the spam from Chinese product reviews	(Liu & Wang, 2012)	88.30%
Reviewer Centric Methods			
6	Detecting product review spammers using rating	(Lim, Nguyen, Jindal, Liu, & Lauw, 2010)	78%
7	Review Graph based Online Store Review Spammer Detection	(G. Wang, Xie, Liu, & Philip, 2011)	49%

Besides these, researchers also tried text mining. For instance, researchers detect online review spam by using text mining and semantic language model (Lau et al., 2011), or KL divergence and probabilistic language modeling based computational model (Lai, Xu, Lau, Li, & Jing, 2010). What is more, other researchers also tried the perspective of

reviewers' behavior pattern, like rating behaviors (Lim et al., 2010).

2.5 Business Application of Review Spam Detection

Even though a lot of research on review spam detection is out, only a few e-commerce websites consider this problem. Yelp is an exception. One of the most famous business applications of review spam detection methods is the Yelp Review Recommendation System. In order to ensure the effectiveness of user opinions posted on Yelp, it uses a filtering algorithm to filter those fake or deceptive reviews and puts them in a filtered list after the normal reviews' list. Yelp's review recommendation system has been proved to be highly accurate (Weise, 2011). As the recommendation system is classified to protect the business, the algorithm is not public. However, Mukherjee, Venkataraman, Liu and Glance figured out its possible methodology in their paper of "What Yelp Fake Review Filter Might Be Doing?" They showed that the recommendation system did a reasonable job at filtering review spam. They found that Yelp might be using a behavioral based approach to filter review spams by using five main variables including maximum number of reviews, percentage of positive reviews, review length, reviewer deviation, and maximum content similarity. But they also pointed out that Yelp can employ many other behavior features extracted from its

internal data. (Mukherjee et al., 2013) For instance, researchers referred to these factors below, IP addresses, geographical location, network logs, mouse gestures, click behaviors and social network interaction (friend and fan relations) (Z. Wang, 2010).

2.6 Social Network Analysis and Review Spam Detection

There are very little research articles about whether social network structure could be one of behavioral features to detect review spam. One we could find is the paper of Identify Online Store Review Spammers via Social Review Graph (G. Wang, Xie, Liu, & Yu, 2012). This paper considers reviewers, reviews, and stores as three different nodes and offers complementary models to existing approaches to find more difficult and subtle spamming activities. The other one is Learning to Identify Review Spam (Li et al., 2011). The authors adopt three perceptive features. In the part of reviewer related features, one of them is behavior feature including authority score. They use dataset from Epinions.com where reviewer can trust another reviewer on it. So, they have relationship data. But the results show that the precision rate will go up and the recall rate will go down if dropping the behavior feature.

Social network analysis and graph models' analysis become famous

research topics since the rapid development of the Internet. Moreover, this methodology is also used in different research areas. The most successful application is a social recommendation system. Traditional recommender systems assume that all users are independent and identically distributed. They do not consider the effects of social interactions or connections among them. What is more, social network analysis could also be used to solve the two common problems of data sparsity and poor prediction accuracy which was recognized as the most serious challenges in traditional recommendation system (Ma, Yang, Lyu, & King, 2008).

Table 2. Social Network Analysis and Existing Review Spam Detection

No.	Article	Methodology and Remarks
1	Learning to Identify Review Spam (Li et al., 2011)	● By using Graph Analysis among users' trust network ● However, this article shows its insignificance on review spams detection
2	Identify Online Store Review Spammers via Social Review Graph (G. Wang, Xie, Liu, & Yu, 2012)	● By using Graph Analysis among reviewers, reviews and stores to survey spammers' behaviors and proved its effeteness in review spam detection

Besides recommendation systems, social network analysis is also widely used in other areas of research such as economics and management. For instance, Koohborfardhaghighi & Altmann conduct agent-based modeling and survey the knowledge management in various networks with differernt connectivity patterns and they found that organizations need a more flexible hybrid structure to allow mobility and lateral connection between the individuals.(S Koohborfardhaghighi & Altmann, 2017) What is

more, Somayeh Koohborfardhaghighi, Romero, Maliphol, Liu, & Altmann adopted network analysis methodology in surveying the tradition game theory model – Prisoners' Dilemma. game. They investigate the effect of bounded rationality of individuals on the networking topology (i.e., the individuals' personal networks). Finally, they found that while the Prisoner's Dilemma evolutionary game simulations tend towards full cooperation, graph topologies with shorter average path lengths and low clustering coefficients increase the amount of steps needed to reach equilibrium.(Somayeh Koohborfardhaghighi et al., 2017)

We suppose that the information among users' interactions and connections could help us to detect review spam, especially those reviews that are hard to be detected through text and sentiment analysis. Moreover, business providers started to buy high quality accounts. That means, accounts of owners who have more than 50 friends and fans. This way, the spammers could avoid being detected by the detection system and would be recognized as truthful reviewers.

Chapter 3. Model

Since Yelp updated and improved it system (for instance, encrypted user-id so that people cannot visit their profile page by using html link and user-id) in order to prevent people crawling their data, a crawling program cannot be used to generate current dataset from Yelp. Therefore, we use a dataset, which has been generated before the update of the Yelp system. Actually, there are two real world datasets, which could be used in our research. One is called Yelp-CHI, which has been collected from Yelp and includes reviews for a set of hotels and restaurants in the Chicago area. But it is not completely suitable for our research since it does not include social interaction information (e.g., friend relationships). But, fortunately, Yelp is holding a dataset challenge and sharing their partial real-world dataset, which is beneficial not only for researchers but also themselves to improve their platform. It includes a huge network including millions of nodes and trillions of edges.

Figure 4. Research Models

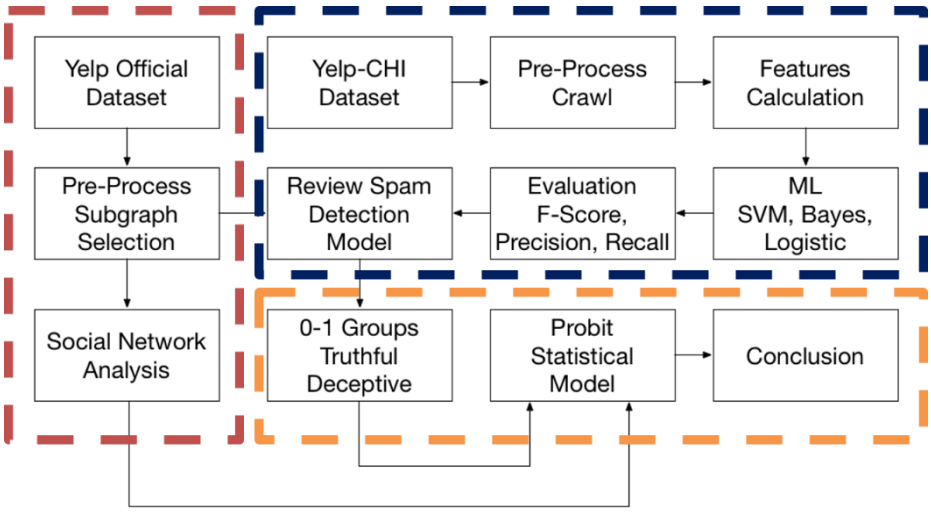


Figure 4 shows the total structure of our research models. The research model includes three main parts: First, the review spam detection model that uses the Yelp-CHI dataset (note, as the Yelp official dataset does not include the information about which review might be deceptive or fake, we need to make our own detection model first) for solving 0-1 classification problem of reviews; Second, the social network analysis that measures the centrality of each node in the selected graph (note, since the Yelp official dataset is too huge, we cannot analyze the whole graph because of limited calculation power); Final, the probit statistical model that is used for checking the relationship between the 0-1 classification of reviews and the centrality of reviewers.

3.1 Review Spams Detection Model

3.1.1 Selection of Dimensions

A complete behavior of posting a review includes three entities, user, review and product. So, we choose diverse dimensions in these three perspectives. We will talk about each dimension one by one in following part. And finally, after that, we talked about the supervised learning way which is widely used to solve classification problem (fake review or truthful review)

By considering about the case of Yelp, we decide to use the following dimensions:

User Features

(1) Evaluation Votes

We suppose that there might be relationship between evaluation votes and activity degree of users. If users have even given lots of evaluation votes to others, it usually means that they spend much time reading and making comments on others' reviews at Yelp platform (high activity degree). If a user is willing to spend much time doing that on this platform, probably he or she will not make a fake account to write fake reviews.

$$\text{Evaluation Votes} = \text{Useful} + \text{Funny} + \text{Cool}, \quad \text{(All are given to others)}$$

In Yelp system, users could evaluate others' review by voting "Useful", "Funny" and "Cool". For instance, if a user has ever voted others by giving 3 "useful", 2 "funny" and "4" cool. Then his evaluation votes will be 9 (=3+2+4).

(2) Elite

Elite is a kind of honor system in Yelp, which is the way of recognizing people, who are active in the Yelp community and have big influences on others. Elite-worthiness is based on a number of things, including well-written reviews, high quality tips, a detailed personal profile, an active voting and complimenting record, and a history of playing well with others. Yelp's selection of Elite is quite difficult and mysterious. Scientists even made a research on how it works to help users become an elite (Kim, Lin, & Bang, n.d.). Through Yelp's strict and tough selection, we believe that those elite users are not review spam writers.

$$\text{Elite} = \sum \{x = 0|1\}, (if\ \ user\ get\ the\ honor\ of\ Elite, then\ 1)$$

In Yelp, Elite is a honor system. Here if a user is Elite member, then its elite is 1, otherwise is 0.

(3) Compliment

Compliment is another evaluation system in Yelp. It is for users instead of reviews and includes 11 categories, which are “Thank You, Cute Pic, Good Writer, Hot Stuff, Just a Note, Like Your Profile, Write More, You're Cool, Great Photo. Great Lists, You're Funny”. We suppose that the more compliments a user received from others, the lower probability that a user will write fake or deceptive reviews.

$$\text{Compliment} = \sum_{i=1}^{11} C_i \quad (i = 1,2,\ldots,11)$$

$$C_i \; is \; the \; number \; of \; compliments \; from \; others$$

In Yelp, Compliment is an honor system to user instead of review. For instance, if a user gets 3 compliments of “Thank you”, 4 compliments of “Like Your Profile”, then its compliment is 7 (=3+4).

(4) Yelping Since

It measures how long a user has been a Yelp user. We thought it might be reasonable that there is a lower probability for elder users to write a fake review, because as we mentioned, review spammers do not want to be caught, so they will keep changing accounts to write review spams. One way to do that is by registering new accounts. So usually we assume that elder

account is more like as a truthful user

$$\text{Yelping\_Since}_i = \text{Y}_i - 2000$$

$$Y_i \; is \; the \; registration \; date \; (Year)$$

For instance, if a user is created in 2013, then his Yelping_Since will be 13 (=2013 – 2000)

(5) Fans

Even though Yelp is a kind of website, which offers business evaluation service, it built its own social network in a different way than other SNS websites (e.g., Facebook or Twitter). Yelp's social network system is different as people cannot know, who are their fans except fans themselves. In other words, we can only know the number of our fans, but we do not know who they are. Other users know only the number of fans of a reviewer. We think that those people, who have few social activities, are more likely to be spam review writers.

$$\text{Fans} = \text{Total Amount of Fans}$$

(6) Friends

As other traditional and famous SNS platform, Yelp allows users to make friends and show social network in their own profile page. As we

explained before, a fake user usually is not an active social person. Moreover, although they are sometimes even active in social activities, they usually prefer to create an account to write fake reviews to get benefit instead of using their real account directly.

$$\text{Friends} = \text{Total Amount of Friends}$$

Review Features

(7) Squared of text length

The purpose of fake reviewers is for getting some benefit like coupon or slander to competitors. What they would like to do is giving low scores to them. Usually they would not pay much time and attention to write a long review. So, we suppose that the fake reviews usually are shorter than normal reviews. Lots of researchers have adopted it as one dimension to detect review spams (Li et al., 2011; Yoo & Gretzel, 2009). However, there are also some researchers have different ideas about its effects. Some of them thought the more review content will be likely as a fake or deceptive review (Li et al., 2011).

$$\text{Text Length} = \textit{The length of review text}$$

(For instance, if a review content is “It is delicious!”, then its text length is 16 (=2(it)+1+2(is)+1+9(delicious)+1(!)))

(8) First Person v.s. Second Person Phrases

Lots of users write fake reviews, in order to help business owners to attract customers. So, this kind of review would be like “advertisement”. Those reviews use phrases like “You should have a try” instead of “I think it is really good”. Even though researchers have different ideas about the effect of fake review detection written in ‘first person’ and ‘second person’ pronouns, they still agree with that it must have influence on fake review detection. Some researchers thought that the frequent use of first person pronouns reveals deceptive reviews. Others pointed out that faked reviews use second person more often than truthful reviews (Li et al., 2011).

$$\text{Person Pronuns} = (number\ of\ first\ person\ words - number\ of\ second\ person\ words)^2$$

For instance, if a review content is “You should have a try!”. The number of first person words is 0. The number of second person words is 1. So, its person pronouns score is 1.

(9) Subjective v.s. Objective

Researchers also point out that truthful reviews usually express both sides of sentiments including subjective and objective. Spam reviews, however, express only objective information. It is more like an

advertisement instead of a truthful review from a customer. We also agreed with that and select it as one of our factor (Li et al., 2011).

$$\text{Subjective} = (number\ of\ subjective\ words - number\ of\ objective\ words)^2$$

For instance, if a review content is “It is delicious!”. The number of subjective words is 1. The number of objective words is 0. So, its subjective score is 1.

(10) Positive v.s. Negative

Based on the previous research, researchers tested the ratio of the number of positive and negative words to total words (Li et al., 2011). It has been proved that deceptive reviews contain more positive words on average than truthful reviews. Contrary, deceptive reviews contains less negative words on average than truthful reviews (Yoo & Gretzel, 2009).

$$\text{Postive} = (number\ of\ postive\ words - number\ of\ negative\ words)^2$$

For instance, if a review content is “It is delicious!”. The number of postive words is 1. The number of negative words is 0. So, its positive score is 1.

(11) Evaluation Score (Useful + Funny + Cool)

This is the most intuitional variable for evaluating reviews' helpfulness directly from other users. Users can give evaluation (useful, funny or cool) about other users' reviews. Since Yelp does not allow users to give bad evaluation, we imply that a review is evaluated by more users (Given Useful, Funny or Cool), , it must be more like a truthful review.

$$\text{Evaluation Score} = \text{Useful} + \text{Funny} + \text{Cool}, \quad \text{(All receive from others)}$$

(In Yelp system, users could evaluate others' review by voting "Useful", "Funny" and "Cool". For instance, if a review gets 3 "useful", 2 "funny" and "4" cool. Then his evaluation score will be 9 (=3+2+4))

(12) Brand's Appearing

Yoo & Gretzel also point out that 90.5% of the deceptive reviews mentioned the brand name, however, only 62.5% of the truthful reviews include a reference to the brand. What is more, only 22.5% of the truthful reviews mentioned the brand more than once, however, 47.6% of the deceptive reviews mention the brand name at least twice (Yoo & Gretzel, 2009).

$$\text{Brands} = \text{the amount of brand } name's\ appearing$$

For instance, if a review content is "Coca-Cola is good, You

Should also Buy a Coca-Cola's Beverage!". The number of brand name is 2. So its Brands score is 2.

Product Features

(13) Product Average Scores v.s. Review Score

This factor is the most common and famous used in previous research. As a kind of statistical method, it shows that it could be abnormal, if a user gives a score which is quite different from another reviewers' score. Since Yelp also offers this kind of data, we also adopt this factor in our research.

$$\text{Score Diff.} = (\text{User's Score} - \text{Business's Average Score})^2$$

For instance, if a user gives a restaurant 0, but the average score of that restaurant is 5, then its score Diff. is 25.

3.1.2 Classification Methodology – Supervised Learning

After deciding these dimensions above, we need to adopt appropriate models for classifying reviews. Recently, machine learning becomes more and more famous and shows its effectiveness. For classification problems, people usually use supervised learning. For solving our 0-1 classification problem, we adopt three different methods: SVM,

Bayesian, and Logistic.

(1) SVM

The basic idea of SVM (Support Vector Machine) is to find the best separating hyperplane, which is the farthest from the support vector (i.e., the set of nodes which is closest to the separating hyperplane).

If we use mathematical formula and define the hyperplanes as $(w^T x + b)$

We can express the problem as:

$$\arg\max_{w,b}\left\{\min_{n}\left(label \cdot (w^T x + b)\right) \cdot \frac{1}{\|w\|}\right\}, label = \{-1,1\}$$

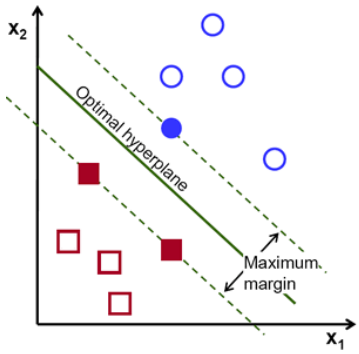
$$w\ means\ coefficient\ of\ independent\ variable\ x;\ b\ is\ constant;$$

$$label\ means\ direction\ of\ the\ hyperplance$$

$$in\ order\ to\ make\ sure\ the\ distance\ value$$

$$will\ be\ more\ than\ 0.$$

Figure 5. SVM



(Source: Wiki[3])

(2) Naïve Bayes

The idea of Naïve Bayes classification is from the famous conditional probability theory:

$$p(c|x) = \frac{p(x|c) \cdot p(c)}{p(x)}$$

$$There\ are\ two\ events\ (x\ and\ c), p(c|x)\ means$$

$$the\ probability\ of\ event\ c\ happens$$

$$if\ event\ x\ alreadyy\ happened.$$

With respect to our classification problem, we assume that there are two categories C_1 and C_2.. If we consider the category of point (x, y), the classification problem can be changed to:

$$p(c_i|\mathrm{x,y}) = \frac{p(x,y|c_i) \cdot p(c_i)}{p(x,y)}$$

If ($p(c_1|x, y)$> ($p(c_2|x,y)$, then (x,y) belongs to category c_1

If ($p(c_1|x, y)$< ($p(c_2|x,y)$, then (x,y) belongs to category c_2

$$There\ are\ i\ categories;\ (x,y)\ is\ the\ point;$$

$$p(c_i|x,y) means\ the\ porbability\ of\ (x,y)\ belongs\ to\ c_i;$$

$$p(x,y|c_i) means\ the\ probablity\ that\ get\ point\ (x,y) in\ catergory\ i$$

$$p(x,y) means\ the\ porblility\ of\ point(x,y);$$

[3] https://en.wikipedia.org/wiki/Support_vector_machine

$$p(c_i) means\ the\ porbibility\ of\ catergory\ i)$$

(3) Logistic

Logistic is another regression model, to solve the classification problem. If a function can be found that outputs 1 or 0, then we can solve the classification problem. This kind of function is called Heaviside step function.

However, since this kind of function is not continuous, sometimes it is hard to use. Another similar function, which has the approximate characteristics, is called Sigmoid function.

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

So, for lots of different features, we need to do the regression and solve the formula below,

$$z = w_0 x_0 + w_1 x_1 + w_2 x_2 \ldots + w_n x_n$$

$$w\ means\ the\ coefficient\ of\ independent\ feature\ variable\ x;$$

$$e\ is\ natural\ logarithm.$$

3.2 Reviewers' Network Structure Analysis Model

3.2.1 Selection of Centrality Dimension

(1) Clustering Coefficient

The clustering coefficient of a node in a graph quantifies how close its neighbors are to being part of a clique, which is first published in the research paper "Small World Network" (Watts & Strogatz, 1998). In social network, the clustering coefficient measures how much your friends know each other. If a vertex v has k_v neighbors, then the most edges exist between them is $k_v \cdot (k_v -1) / 2$. Then, the clustering coefficient C(p) is defined as follows:

$$\text{clustering coefficient} = \frac{the\ number\ of\ really\ existing\ edges}{\frac{\text{kv} \cdot (\text{kv} - 1)}{2}}$$

(2) Closeness Centrality

In a connected graph, closeness centrality of a node is another measure of a node's centrality in a network. It is calculated as the sum of the length of the shortest paths between the node and all other nodes in the graph. Thus, the more central a node is, the closer it is to all other nodes (Bavelas, 1950).

If a vertex i in a connected network, we can calculate average shortest path length of this vertex to another node j. The closeness centrality

is defined as follows:

$$d_i = \frac{1}{N} \left(\sum_{j=1}^{N} d_{ij} \right)$$

Then we define closeness centrality is the reciprocal of a vertex's average shortest path length,

$$\text{Closeness Centrality} = \frac{1}{d_i}$$

d_{ij} is 0 or 1. If it is 0, it means node i and node j are connected, otherwise not

.

(3) PageRank

PageRank is a link analysis algorithm and it assigns a numerical weighting to each element of a hyperlinked set of documents (e.g., World Wide Web), with the purpose of "measuring" its relative importance within the set (Page et al., 1999).

The basic idea of PageRank is also about analyzing the website's structure. The importance of a specific page is based on the amount and the importance of those pages, which point to it.

Algorithm of PageRank: By using PageRank algorithm, we can measure the importance of nodes (users) just like websites on the Internet.

I. Initial Step: Give each node i an initial PR score $PR_i(0)$. i=1, 2,…,N, and

$$\sum_{i=1}^{N} PR_i(0) = 1$$

$$(Graph\ includes\ N\ nodes(i = 1 \dots N);$$

PR means PageRank Score; $PR_i(0)$ means initial PageRank Score)

II. Iteration (step k>=1)

$$PR_i(k) = \sum_{j=1}^{N} a_{ji} \frac{PR_j(k-1)}{k_j^{out}}, \quad i = 1,2, \dots, N$$

a_{ji} is 0 or 1. If node i and node j are connected , a_{ji} is 1, otherwise 0;

k_j^{out} means in step k, the out degree of node j .

(4) HITS

Hyperlink-Induced Topic Search (HITS), which is also known as Hubs and Authorities, is a link analysis algorithm that rates Web pages, developed by Jon Kleinberg. The idea behind Hubs and Authorities stemmed from a particular insight into the creation of web pages, when the Internet was originally formed.; That is, certain web pages, known as hubs, served as large directories. They were not actually authoritative in the information that they held but were used as compilations of a broad catalog of information that led users directly to other authoritative pages. In other words, a good hub represented a page that pointed to many other pages, and a good

authority represented a page that was linked by many different hubs (Kleinberg, 1999).

The basic idea of HITS comes also from the analysis of websites' graph structure. For measuring the importance of a website, two factors have to be captured: One is the authority of a webpage. The other is the hub of webpage. For instance, if someone wants to search for the information of Seoul National University, considering about the content authority, the homepage of Seoul National University has the most importance. In other perspective, if there are some university ranking websites, like QS University Ranking[4], then people can find its homepage at that website. Both QS websites and Seoul National University are important but with a different perspective.

Algorithm of HITS:

I. Initial Step: Give each node i an initial authority score $X_i(0)$ and hub score $Y_i(0)$. i=1, 2,…,N

$$(Graph\ has\ totally\ N\ nodes.$$

$$Each\ Node\ i\ has\ a\ authority\ score\ X\ and\ hub\ Score\ Y\ )$$

II. Iteration (step k>=1)

a) Adjust Authority Score

[4] https://www.topuniversities.com/university-rankings

Each node's authority score equals to summation of hub scores from all nodes that point to it

$$x_i'(k) = \sum_{j=1}^{N} a_{ji} y_j(k-1) \ , i = 1,2, \dots, N$$

$x_i'(k)$ is updated authority score;

a_{ji} is 0 or 1. If node i and node j are connected , a_{ji} is 1, otherwise 0

y_j means the hub score of node j.

b) Adjust Hub Score

Each node's hub score equals to summation of authority scores from all nodes that point to it

$$y_i'(k) = \sum_{j=1}^{N} a_{ji} x_i'(k) \ , \qquad i = 1,2, \dots, N$$

$y_i'(k)$ is updated authority score;

a_{ji} is 0 or 1. If node i and node j are connected , a_{ji} is 1, otherwise 0

$x_i'(k)$ is updated authority score of node i.

c) Normalization for making X and Y in the interval of [0,1]…

$$x_i(k) = \frac{x_i'(k)}{\|x'(k)\|}, \quad y_i(k) = \frac{y_i'(k)}{\|y'(k)\|}, \quad i = 1,2, \dots, N$$

3.2.2 Sub-Network Construction

As Yelp Official Data includes millions of nodes and trillions of

edges, it is hard to calculate centrality in this kind of huge network. As the size of the entire network from Yelp Official Dataset is too big, it is impossible for us to calculate the network centrality variable we talked about before. So, we have to adopt some strategy to shrink the size of the entire network reasonably. We consider about two construction methodology, four degrees of separation network and six degrees of separation network.

Firstly, I choose the research target as the process below:

(1) I select all the business services (e.g., restaurants) (1556) in Illinois State, because Yelp-CHI is also from Chicago.

(2) Based on those business services (e.g., restaurants), I get all of those reviews (29874) related to them.

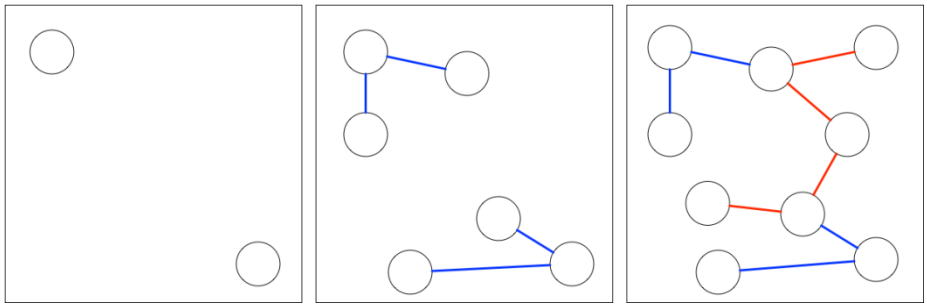
(3) There are totally 10022 reviewers, who wrote those 29874 reviews.

But now these 10022 reviewers are not connected. In order to connect them, we need to find their friends, even their friends' friends and so on. Then we can create the social graph successfully as figure 6 and figure 7 shows. But if we keep finding friends' friends' friends …., it will become the initial graph which includes too many nodes. So which layer or depth should we stop? We adopt two strategies. One is four degrees of separation network and the other is six degrees of separation network.

Four Degrees of Separation Network

Lars Backstrom in his 2012 research article shows that the world is becoming smaller than expected. Based on his data analysis of the entire Facebook network of active users, which are about 721 million users and 69 billion friendship links, the average distance they observed has been 4.74, corresponding to 3.74 intermediaries. So, it got defined as a four-degrees-of-separation network. The first experiment we made is also based on this theory. We adopt those 10022 target users, those users' friends and those users' friends' friends. Through this way, we can build a sub-network graph of the entire network graph. Nodes are those 10022 Yelp users and links are friendships among them. After this way, we can create a sub-graph, and then calculate these 10022 Yelp users' network centrality scores. We suggest that if we have enough calculation power and can calculate these 10022 Yelp users' network centrality scores in the entire big graph (4.7 million nodes), they will be approximately equal.

Figure 6. Four Degrees of Separation Network



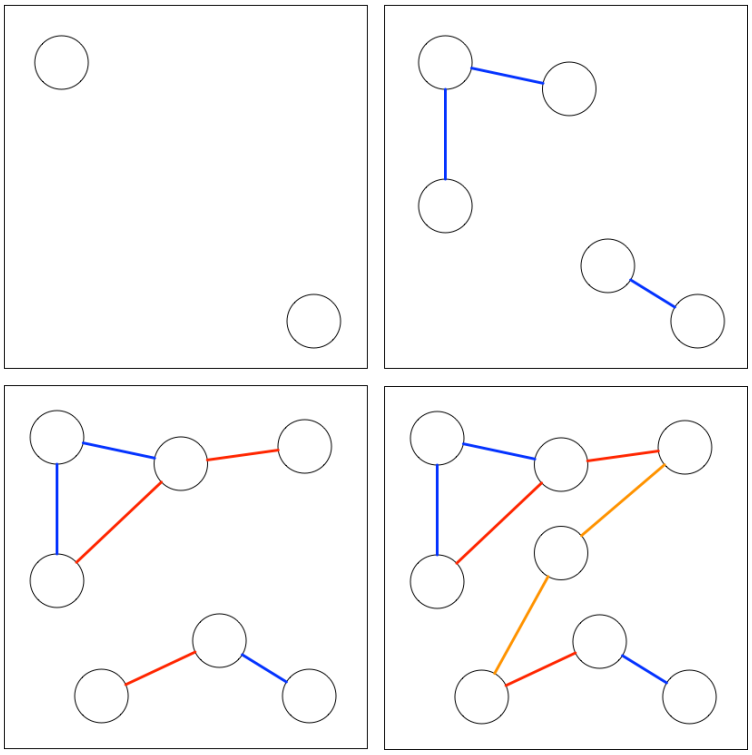
(If we want to construct a sub-graph which connects two strangers, we can try to build four degrees of separation network. Firstly, we find these two friends' friends, where the blue links show the friend relationship. Secondly, we continue to find their friends' friends, which are represented through the red links show. Based on the theory of four degrees of separation network, it is possible and reasonable that they can be connected through this way. The length of this path between them is 4 now.)

The initial target users who we selected from the Yelp Official Dataset are 10022 reviewers. Through the four-degree separation network theory, we keep adding their friends and their friends' friends into our graph. Then the current sub-network includes 454325 nodes and 735466 edges.

Six Degrees of Separation Network

Even though scientists have pointed out that our world is becoming smaller and smaller, it has still limitations because this research is based on the dataset of Facebook. As we all known, Facebook is the largest social network platform in the world. The average path length could be lower to 4. However, Yelp is a platform focusing on evaluation of businesses instead of social networks. Therefore, it is doubtful that the average path length could be 4. As a consequence, we decide to adopt the six degrees of separation theory, which is also called small world theory (Watts & Strogatz, 1998). We explore more friends and build a higher depth of network. We receive 10022 targets' friends, their friends' friends and also their friends' friends' friends.

Figure 7. Six Degrees of Separation Network



(If we want to construct a sub-graph which connects two strangers, we can try to build four degrees of separation network. Firstly, we find these two friends' friends, where the blue links show this kind of relationship. Secondly, we continue to find their friends' friends, where the red links show. Thirdly, different from previous construction method, we continue to find their friends' friends' friends, where the orange links show the relationship. Based on the theory of six degrees of separation network, it is possible and reasonable that they can be connected through this way. The length of this

path is 6 now)

Through this way (Figure 7), we get a much bigger network comprising 2,868,405 nodes and 10,545,238 edges.

3.3 Final Statistical Analysis

By inputting the essential information (users' features, reviews' feature and products' features), we can know which review is deceptive (fake) or truthful. And then through the social network analysis above, we know the foure network centrality scores of each user (node). Finally, we use a probit model which is a type of famous regression model in statistics, where the dependent variable can take only two values , for example married or not married (for our case, fake or not fake)

$$\mathrm{Y}\{0,1;\text{Truthful or Deceptive}\} = w_1 x_{clustering\ coefficient} + w_2 x_{closeness\ centrality} + w_3 x_{PageRank} + w_4 x_{HITS} + \varepsilon$$

(w means the coefficient of

four independent network centrality variables)

Related analysis result will be detail explained in the next section

Chapter 4. Analysis

4.1 Data Description

Yelp-CHI Dataset

Lots of Internet services, especially e-commerce service, allow users to write reviews. However, only a few of them have built the social network among users. One of the most successful cases is Yelp. Until 2016, Yelp.com has almost 135 million monthly visitors and 95 million reviews (Mukherjee et al., 2013). Yelp launched the automated software to recommend the reviews they think will be the most helpful to the Yelp users based primarily on quality, reliability and the reviewer's activity on Yelp. Yelp filtered that information they thought are not helpful. They still retain them in the folded part (initially hidden) instead of deleting them directly. Even though it is not perfect, it is still proved as an effective review spams detection system to produce accurate results (Weise, 2011).

Table 3. Yelp-CHI Dataset Description

No.	Variable	Explanation	Example (Data Format)
1	Date	Review Written Date	9/22/2012
2	Review ID	Identification of Review	GtwU21YOQn-wf4vWRUIx6w
3	Reviewer ID	Identification of Reviewer	bNYesZ944s6IJVowOnB0iA
4	Product ID	Identification of Product	pbEiXam9YJL3neCYHGwLUA
5	Fake_Label	N means genuine review and Y means fake reviews	N
6	Useful	Amount of people thought it is useful	2
7	Funny	Amount of people thought it is funny	1
8	Cool	Amount of people thought it is cool	3
9	Star rating	The evaluation given to this product by this reviewer	5
10	Review Content	Review Content	Service was impeccable. Experience and presentation was cool. Eating a balloon was fun. Trying to make a reservation was ridiculous. I appreciate delicious food, so I don't get the hype here.

We can find deceptive review information in the hidden parts of Yelp website, but it is still difficult to crawl all information we need, especially huge social network information. For instance, people are prevented from visiting those "fake" reviewers' homepage directly by the latest Yelp system (like they cancel the link to their own homepage web address). But fortunately, previous researchers have ever done related research by using Yelp dataset (Mukherjee & Venkataraman & Liu & Glance, 2013). They shared those data (called Yelp-CHI) to other researchers. It includes 63000 Reviews for a set of hotels and restaurants in the Chicago in 2014 to 2015. But since they are not focus on the area of social network, they did not collect social network information at that time. Detail data description is shown in table 3 below,

As we mentioned before, we need more information to make our own detection model. The extra information is listed below (table 4),

All of the information below can be easily crawled by making a Python crawler program in their profile page at Yelp website including the number of friends and fans. But it is quite difficult to get the whole network structure information (follows and followed) among them. But Yelp offered us another choice. Yelp is holding a competition game and share a part of their dataset including 4,700,000 reviewers, 156,000 businesses.

What is more, it includes the relationship among these 4,700,000 reviewers.

Table 4. Supplement Information for Yelp-CHI Dataset

No.	Variable	Explanation	Example
1	Useful	Amount of votes (useful) user give to others	2
2	Funny	Amount of votes (funny) user give to others	1
3	Cool	Amount of votes (cool) user give to others	3
4	Compliment	Amount of total compliments received from others. Compliment is a kind of Yelp evaluation system to users. It includes 11 categories (Thank You, Cute Pic, Good Writer, Hot Stuff, Just a Note, Like Your Profile, Write More, You're Cool, Great Photo. Great Lists, You're Funny)	13
5	Elite	Times of getting the honor of Elite, which is a kind of honor system in Yelp, which is the way of recognizing people who are active in the Yelp community and role models on and off the site.	N
6	Yelping_since	Period since this user joined Yelp	June 2015
7	Fans	Amount of this user's fans	99
8	Friends	Amount of this user's friends	213

Since the review recommendation system in Yelp is a business secret, we have no idea about how it works. So we need to make our own review spam detection model, which will be explained in next sub-section.

4.2 Analysis of Review Spams Detection Model

Combing the dataset and those variables referred in section 3, we use SVM, Bayesian and Logistic in separate to make our own detection model and classify those reviews' data which are fake or deceptive in the official Yelp dataset.

It is so surprised that all of them show effectiveness in classifying the Yelp-CHI data. Then we move each variable from the model in separate. We found that there were no big differences except the variable of Elite (especially for Bayesian). It can be demonstrated that Yelp's own evaluation system is an effective way to classify review spam writers. However, the amount of Elite members is quite low. By using this way, we can only detect the limited users who are not spam writers. Another point which we want to refer here is that the amount of friends or fans does not show their effectiveness in classification. As table 5 shows, there is no big difference even though we remove both amount of friends and amount of fans.

The maximum precision rate among previous review spam

detection methodology is 88.30% (as table 1 shows), however our maximum precision rate is 89%~90%. Our review spam detection model shows its effectiveness well.

However, we also feel curious about the issue if different dataset shows different difficulty on detection, so we continue compare with the previous research of review spam detection methodology on same dataset. The model from "What Yelp Fake Review Filter Might Be Doing ?(Mukherjee et al., 2013)" shows significant effectiveness. It has higher precision rate, recall rate and F1-Score (Maximum: Precision Rate: 84.1%, Recall Rate: 87.3%, F1-Score: 85.7). Its overall evaluation metric is much better than our methodology. But what an important difference is our precision rate is much higher than it. Our precision rate could be almost 90%. And in our perspective, precision rate is much important in a ecommerce website. Because it is allowed for website owners to miss the review spammers, but it is not allowed for them to mistake regular or honest users for a review spammer. It harms users' experience.

Table 5. Evaluation Results of Three Classification Models

	F-Score	Precision	Recall
SVM - All	0.732060361	0.887255232	0.687250188
SVM - Except Friends and Fans	0.7271730187	0.8926111835	0.6761939861
SVM - Except Elite	0.742629393	0.8831862775	0.6946142893
SVM - Except Others' Evaluation	0.7115683477	0.8815181005	0.6577770333
Bayesian - All	0.73660866	0.901592169	0.687375827
Bayesian - Except Friends and Fans	0.7372461364	0.9024373797	0.6881330095
Bayesian - Except Elite	0.6495197881	0.8949566919	0.5896959544
Bayesian - Except Others' Evaluation	0.5613986382	0.8850417793	0.499907865
Logistic - All	0.809102358	0.82442786	0.865707346
Logistic - Except Friends and Fans	0.806867581	0.8326500084	0.8653773348
Logistic - Except Elite	0.8085102117	0.830889747	0.8655498785
Logistic - Except Others' Evaluation	0.8061862649	0.8254989456	0.8649384371

Those variables are ignored if there is no big difference after removing it. In order to figure out the importance of each variable, w

e remove each variable and do the three classification models again. S ome of results are ignored since there is no big difference. The most meaningful variables are listed in the table 5. As previously explained, elite has the most significant influence on review spam detection. An d then others' evaluation also shows little significant influence. But wh at makes our supersized is the number of friends or fans does not sho w helpfulness on review spam detection.

4.3 Analysis regarding Reviewers' Network Structure

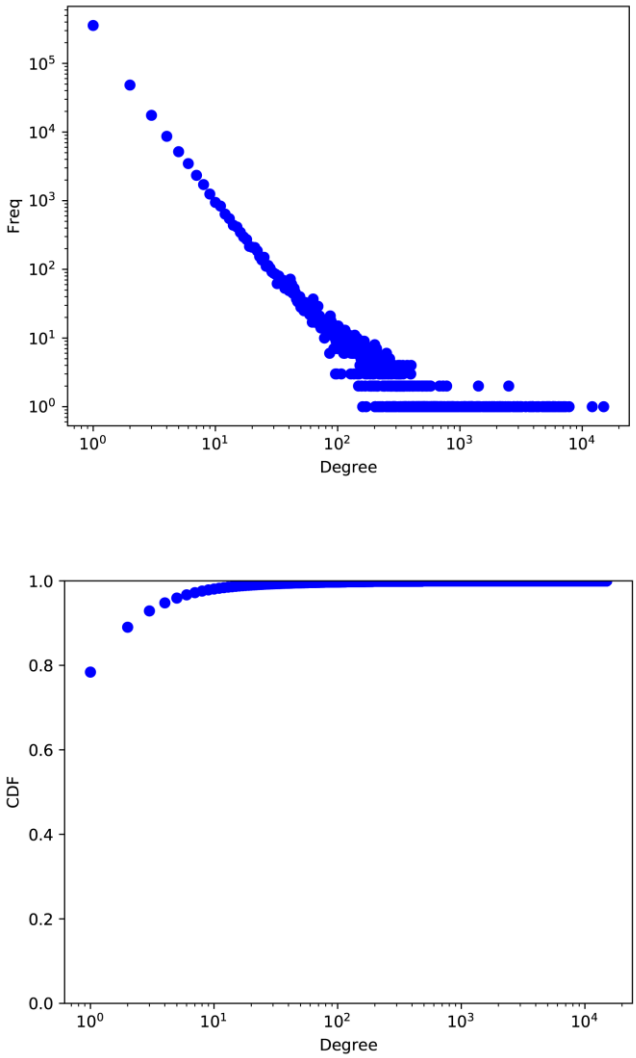
As explained before, Yelp's official dataset includes millions of nodes and trillions of edges; it is quite difficult to analyze the whole graph. In order to draw the representative graph of 10022 reviewers, we adopt two kinds of different theories, four-degrees separation (Backstrom, Lars, et al 2012) and six-degrees separation theory (Watts, D. J., & Strogatz, S. H. 1998).

4.3.1 Four Degrees of Separation Network

Because of the way we construct this sub-network, it is reasonable that most of nodes' degree is one (Because we stop extending network, if the average depth arrives to four). However, it is still matched to power

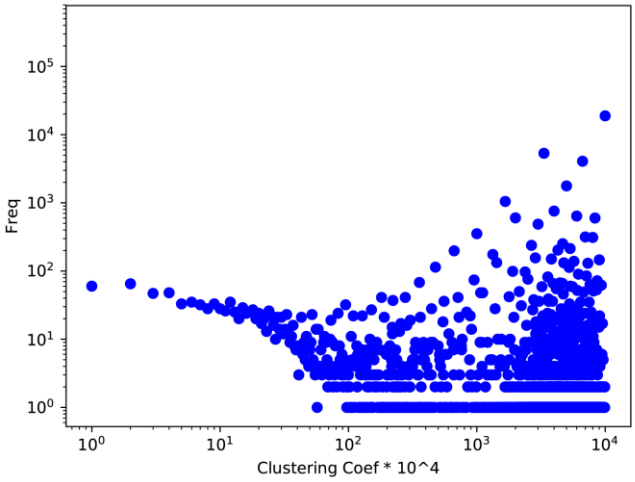
distribution as the figure 8 shows. There are still lots of nodes have larger degrees (maximum value has been more than ten thousand)

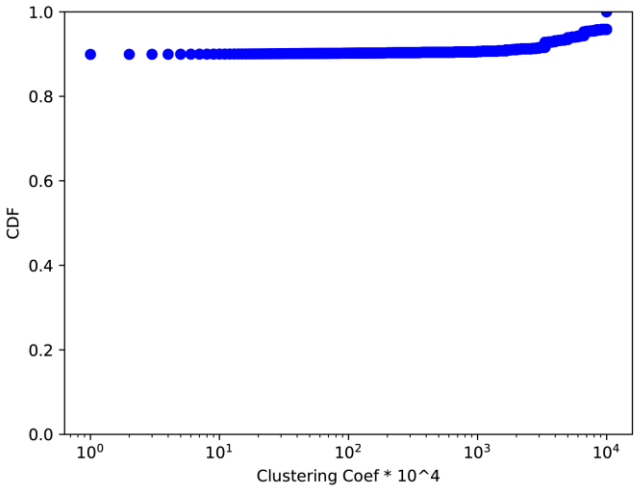
Figure 8. Degrees Distribution of Four Separation Network



Through figure 9, we can see it is so supervised that the clustering coefficient is much lower than we expected. It is almost 80% of nodes' coefficient is zero. It is probably same reason as above. Most of nodes are at the periphery of the network. We ignore their friends, so their degrees and clustering coefficients are low. …

Figure 9. Clustering Coefficient Distribution of Four Separation Network





After adopting our previous review spams detection model, we classify all of the nodes into two groups. One is deceptive reviewer and the other is truthful reviewer. The reason why we call them deceptive instead of fake is, as explained before, there is no model which can 100% detect review spammer correctly. So even though we cannot conclude its truth, but we can describe them as deceptive review spammers. Then, we calculate average score of each variable (Clustering Coefficient, Closeness Centrality, PageRank and HITS) in two groups and summarize them in the table 6.

Table 6. Basic Statistical Comparison between Deceptive and Truthful Reviews in Four Separation Network

Network Structural Property	**Four Degrees of Separation Network**					
	Mean			**Std.Dev.**		
	D	T	Differences	D	T	Differences
Clustering Coefficient	0.0049281	0.0101078	-0.0051797	0.0421791	0.0652612	-0.0230821
Closeness Centrality	0.0875734	0.149659	-0.0620856	0.0983514	0.1102118	-0.0118604
PageRank	5.56E-06	0.0000112	-5.64E-06	0.000133	0.0000309	0.0001021
HITS	1.14E-07	3.77E-06	-3.66E-06	8.64E-07	0.0000134	-1.25E-05
	Minimum			**Maximum**		
	D	T	Differences	D	T	Differences
Clustering Coefficient	0	0	0	1	1	0
Closeness Centrality	0	0	0	0.278459	0.3315558	-0.0530968
PageRank	0	0	0	0.0001473	0.0008151	-0.0006678
HITS	0	0	0	0.000023	0.0004335	-0.0004105

(D: Deceptive Reviews; T: Truthful Reviews)

Through the basis statistical results (means, standard deviation,

minimum and maximum) in table 6, we can find that deceptive reviews always have lower value of clustering coefficient, closeness centrality, PageRank and HITS than truthful reviews.

Then we do the probit model analysis (Bliss, 1934) by choosing Y (means fake or not, 0 means truthful, 1 means fake) as dependent variable and choose Clustering Coefficient, Closeness Centrality, PageRank and HITS as independent variables.

Table 7. Probit Analysis in Four Separation Network

Network Structural Property	Four Degrees of Separation Network		
	P-Value (Coefficient Sign)		
	SVM	Bayesian	Logistic
Clustering Coefficient	0.052 (-)	0.004 (-)	0.579 (-)
Closeness Centrality	0.000 (-)	0.000 (-)	0.022 (-)
PageRank	0.487 (-)	0.000 (-)	0.889 (-)
HITS	0.000 (-)	0.000 (-)	0.047 (-)

(SVM, Bayesian and Logistic all adopted all the variables since there is no huge differences comparing other ways. “-“means negative, small er than 0)

Results shows as below, as we can see that, closeness centrality and

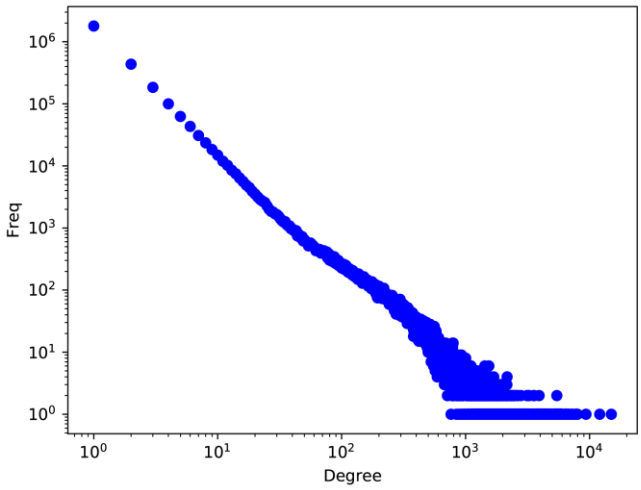
HITS are significant. It is also reasonable that the larger closeness centrality and HITS mean the node exist in a key position, the lower probability that they will become review spams writers.

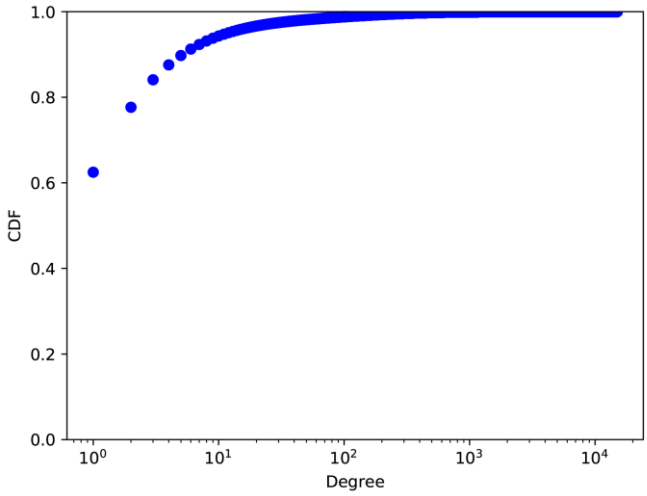
4.3.2 Six Degrees of Separation Network

We still feel worried about that whether there is enough social information in the four degrees of separation network or not. After confirming the feasibility of our calculation workload, we continue to do the six degrees of separation network analysis.

As the figure 9 and figure 10 shows that both four and six-degree separation network has almost same shape.

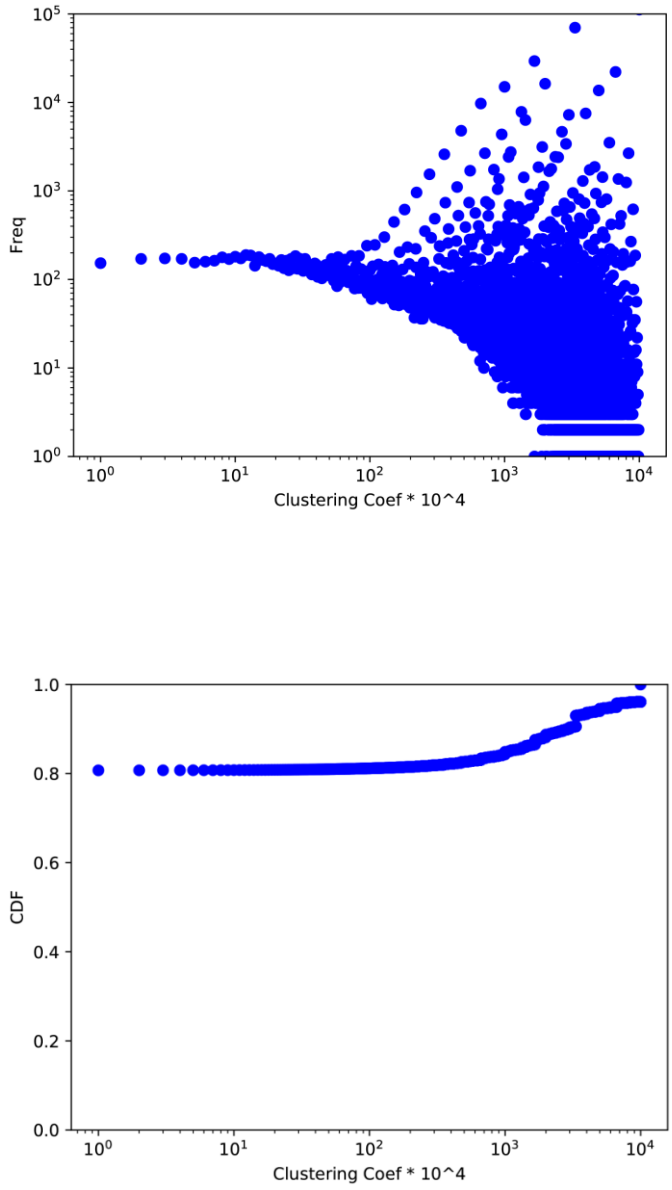
Figure 10. Degrees Distribution of Six Separation Network





The most outstanding difference is the average degree and average clustering coefficient in six-degree separation network is much higher than those in four-degree separation network. For instance, the percentage of nodes who has only one friend is only 60% in six-degree separation network than 80% in four-degree separation network and the percentage of nodes whose clustering coefficient is zero is 80% in six-degree separation network than 90% in four degree separation network.

Figure 11. Clustering Coefficient Distribution of Six Separation Network



The basis statistical results (means, standard deviation, minimum and maximum) in table 8, as almost same as the previous four-degree separation network, also shows that deceptive reviews always have lower value of clustering coefficient, closeness centrality, PageRank and HITS than truthful reviews. This part is same as the case of four separation network

As same as previous case, we do the probit model analysis (Bliss, 1934) by choosing Y (means fake or not, 0 means truthful, 1 means fake) as dependent variable and choose Clustering Coefficient, Closeness Centrality, PageRank and HITS as independent variables.

The results of Closeness Centrality and HITS are same. They also show the significant. But the difference in this part is PageRank. It shows significant, but their effects of influence are different. The reasonable explanation should be like Bayesian way shows that the larger PageRank value means the node exist in a key position, the lower probability that they will become review spams writers. We cannot demonstrate the effectiveness of PageRank in our research.

Table 8. Basic Statistical Comparison between Deceptive and Truthful Reviews in Six Separation Network

Network Structural Property	**Six Degrees of Separation Network**					
	Mean			**Std.Dev.**		
	D	T	Differences	D	T	Differences
Clustering Coefficient	0.0049281	0.0101078	-0.0051797	0.0421791	0.0652612	-0.0230821
Closeness Centrality	0.0942884	0.1599144	-0.065626	0.1005532	0.1116725	-0.0111193
PageRank	4.40E-06	0.0000108	-6.40E-06	0.0000104	0.000036	-0.0000256
HITS	1.82E-08	9.33E-07	-9.15E-07	1.92E-07	4.56E-06	-4.37E-06
	Minimum			**Maxium**		
	D	T	Differences	D	T	Differences
Clustering Coefficient	0	0	0	1	1	0
Closeness Centrality	0	0	0	0.2883473	0.3476995	-0.0593522
PageRank	0	0	0	0.0001249	0.0013581	-0.0012332
HITS	0	0	0	7.17E-06	0.0001826	-1.75E-04

(D: Deceptive Reviews; T: Truthful Reviews)

But for clustering coefficient, even though only two ways (SVM

and Bayesian) shows significant. But their effects of influence are same and reasonable. The higher clustering coefficient value means the their friends know each other and be more like a real account instead of a fake account, the lower probability that they will become review spams writers.

Table 9. Probit Analysis in Four Separation Network

Network Structural Property	Six Degrees of Separation Network		
	P-Value (Coefficient Sign)		
	SVM	Bayesian	Logistic
Clustering Coefficient	0.047 (-)	0.001 (-)	0.691 (-)
Closeness Centrality	0.000 (-)	0.000 (-)	0.000 (-)
PageRank	0.000 (+)	0.002 (-)	0.016 (+)
HITS	0.000 (-)	0.000 (-)	0.041 (-)

(SVM, Bayesian and Logistic all adopted all the variables since there is no huge differences comparing other ways. “-“means minus, smaller than 0)

Chapter 5. Discussion and Conclusions

5.1 Summary

While there is extensive research on detecting deceptive or fake reviews, most relevant studies focus on text analysis, sentiment analysis, or synthesis methodology. In this context, our research gives a promising structural importance heuristic to evaluate online reviews from a different perspective of social network analysis. Reviewers' social network structure could be one of the most important factors that have an effect on the review spam detection model.

5.2 Discussion

As demonstrated by the present research findings, our review spam detection model shows a high precision rate. In addition, based on the prediction results, we found a significant relationship between reviewers' structure importance (or network centrality) and reviews' authenticity. Graph link information has been widely applied in many different scenarios like recommendation system. For example, Wang et al. (2011) also adopted the social network analysis to do review spam detection. Yet, the main difference between their and our research findings is the construction of

graph. While the present study focuses more on the review and reviewers' connection, Wang et al. (2011) consider the graph widely, including all of entities (i.e., reviews, reviewers and stores). However, one point of disagreement between the present study and the one by Wang et al. (2011) is that the latter study assumed that it was entirely groundless that a prestige node had a larger chance to be a spammer or a benign reviewer. In other words, the centrality of a node, which usually describes a node's prestige in the graph, cannot be a useful evaluation metric for detecting review spams. This point is what we cannot agree with, for the following two reasons:

(1) More and more people use SNS (Social Network Service) to stay in touch with their friends on the Internet. These friends include "real" friends in real life. Therefore, review spammers usually do not want to be found by their real friends. Besides, review spammers do not want to be caught by websites. For these two reasons, they will usually create another account and use another name to write spam.

(2) When people use real and personal accounts to write review spams, others will recognize what they wrote does not make sense, so they will not be followed or trusted anymore. As a result, they will eventually lose their centrality.

Our research has shown the significant effect of Closeness Centrality and HITS. Our methodology to detect review spam writers is based on two assumptions. The first assumption is that, as explained above in (1), review spammers usually create a fake account to write fake reviews instead of using their real accounts. In this case, the problem of detecting fake reviews could be transferred into the problem of detecting fake accounts. It is obvious that real accounts will have a more important position in the social network than fake accounts. Moreover, clustering coefficient for real accounts is another distinct feature to compare the real and the fake accounts. If it is a real account, it usually has a higher clustering coefficient, as people in real life usually have a higher clustering coefficient. People's friends are more likely to know each other in real life than in virtual life (Online World). Our second assumption is that fake reviews are usually not useful and mislead consumers. Therefore, users will not follow spam reviews writers as friends anymore, since they find their reviews do not make sense or will feel like that reviews are biased. In that case, spam review writers do are unlikely to have a key position in a social network. Our research has demonstrated that the centrality analysis of a node could be a possible solution to detect review spams.

We also believe that the social network analysis can be

meaningfully used for spam detection. It is widely known that text or review content is easier to be fabricated; however, trust from others is more difficult to fabricate. If those reviews written by users do not make sense and are not helpful, they will naturally lose the key role position. In other words, we propose that the social network information is more objective than other types of information. Accordingly, we believe that the social network analysis could be used in review spam detection and the experiment results based on Yelp large-scale dataset provide support to our assumption. To the best of our knowledge, we the present study the first to prove the effectiveness of the social network analysis for review spam detection.

5.3 Limitations

However, the present study has several limitations. First of all, the network we built based on either the theory of four degrees of separation or the theory of six degrees of separation cannot fully represent the entire network. Especially for the measurement variable PageRank, the more a network is complete, the more exact PageRank Score could be calculated. Therefore, in this case, analyzing this huge graph by using the big data analysis tool, such as Spark or Hadoop, would be necessary in the future. Secondly, since the review spams detection model is still an emerging

research area, there is no perfect model for different areas of reviews. Therefore, there is no guarantee that our review detection model is the most perfect one and that the detected fake reviews are completely deceptive. However, but we can still prove that these nodes are suspicious and thus have a higher probability to be review spam. Finally, since the fake reviews' dataset is actually difficult to get, researchers also point out that, while Yelp recommendation system shows effectiveness, no evidence is available to prove it is 100% correct. Therefore, in the future, further search for a more accurate dataset to train our detection model is necessary and the relationship between reviews' quality and reviews' network structure features should be surveyed.

Bibliography

Algur, S., Hiremath, E., Patil, A., & Shivashankar, S. (2010). Spam detection of customer reviews from web pages. In *Proceedings of the 2nd International Conference on IT and Business Intelligence* (pp. 1–13).

Algur, S. P., Patil, A. P., Hiremath, P. S., & Shivashankar, S. (2010). Conceptual level similarity measure based review spam detection. In *Signal and Image Processing (ICSIP), 2010 International Conference on* (pp. 416–423). IEEE.

Arthur, C. (2013). How low-paid workers at "click farms" create appearance of online popularity. *The Guardian*, *2*.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, *22*(6), 725–730.

Bliss, C. I. (1934). The method of probits. *Science*, *79*(2037), 38–39.

Choi, S., Mattila, A. S., Van Hoof, H. B., & Quadri-Felitti, D. (2016). The Role of Power and Incentives in Inducing Fake Reviews in the Tourism Industry. *Journal of Travel Research*, 47287516677168.

Dixit, S., & Agrawal, A. J. (2013). Survey on review spam detection. *Int J Comput Commun Technol ISSN (PRINT)*, *4*, 975–7449.

Donfro, J. (n.d.). A whopping 20% of yelp reviews are fake.

Jindal, N., & Liu, B. (2007). Review spam detection. In *Proceedings of the 16th international conference on World Wide Web* (pp. 1189–1190). ACM.

Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 219–230). ACM.

Kim, C., Lin, G., & Bang, H. (n.d.). Discovering Yelp Elites: Reifying Yelp Elite Selection Criterion.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment.

Journal of the ACM (JACM), *46*(5), 604–632.

Koohborfardhaghighi, S., & Altmann, J. (2017). How Organizational Structure Affects Organizational Learning. *Journal of Integrated Design and Process Science*, *21*(1), 43–60.

Koohborfardhaghighi, S., Romero, J. P., Maliphol, S., Liu, Y., & Altmann, J. (2017). How bounded rationality of individuals in social interactions impacts evolutionary dynamics of cooperation. In *Proceedings of the International Conference on Web Intelligence* (pp. 381–388). ACM.

Lai, C. L., Xu, K. Q., Lau, R. Y. K., Li, Y., & Jing, L. (2010). Toward a language modeling approach for consumer review spam detection. In *e-Business Engineering (ICEBE), 2010 IEEE 7th International Conference on* (pp. 1–8). IEEE.

Lau, R. Y. K., Liao, S. Y., Kwok, R. C. W., Xu, K., Xia, Y., & Li, Y. (2011). Text mining and probabilistic language modeling for online review spam detecting. *ACM Transactions on Management Information Systems*, *2*(4), 1–30.

Li, F., Huang, M., Yang, Y., & Zhu, X. (2011). Learning to identify review spam. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence* (Vol. 22, p. 2488).

Lim, E.-P., Nguyen, V.-A., Jindal, N., Liu, B., & Lauw, H. W. (2010). Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 939–948). ACM.

Liu, L., & Wang, Y. (2012). A method for sorting out the spam from Chinese product reviews. In *Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on* (pp. 35–38). IEEE.

Luca, M. (2016). Reviews, reputation, and revenue: The case of Yelp. com.

Ma, H., Yang, H., Lyu, M. R., & King, I. (2008). Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings*

of the 17th ACM conference on Information and knowledge management (pp. 931–940). ACM.

Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *The American Economic Review*, *104*(8), 2421–2455.

Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st international conference on World Wide Web* (pp. 191–200). ACM.

Mukherjee, A., Liu, B., Wang, J., Glance, N., & Jindal, N. (2011). Detecting group review spam. In *Proceedings of the 20th international conference companion on World wide web* (pp. 93–94). ACM.

Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. S. (2013). What yelp fake review filter might be doing? In *ICWSM*.

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 309–319). Association for Computational Linguistics.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web.* Stanford InfoLab.

Rayana, S., & Akoglu, L. (2015). Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining* (pp. 985–994). ACM.

Sandulescu, V., & Ester, M. (2015). Detecting singleton review spammers using semantic similarity. In *Proceedings of the 24th international conference on World Wide Web* (pp. 971–976). ACM.

Wang, G., Xie, S., Liu, B., & Philip, S. Y. (2011). Review graph based online store review spammer detection. In *Data mining (icdm), 2011 ieee 11th international conference on* (pp. 1242–1247). IEEE.

Wang, G., Xie, S., Liu, B., & Yu, P. S. (2012). Identify online store review spammers via social review graph. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *3*(4), 61.

Wang, Z. (2010). Anonymity, social image, and the competition for volunteers: a case study of the online market for reviews. *The BE Journal of Economic Analysis & Policy*, *10*(1).

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world"networks. *Nature*, *393*(6684), 440–442.

Weise, K. (2011). A lie detector test for online reviewers. *Bloomberg Business Week*.

Xie, S., Wang, G., Lin, S., & Yu, P. S. (2012). Review spam detection via time series pattern discovery. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 635–636). ACM.

Yoo, K.-H., & Gretzel, U. (2009). Comparison of deceptive and truthful travel reviews. *Information and Communication Technologies in Tourism 2009*, 37–47.