DETECTING SOCIAL COMMERCE: AN EMPIRICAL ANALYSIS ON YELP

Jiangtao Qiu
School of Economic Information Engineering, Southwestern University of Finance and Economics
Key Laboratory of Financial Intelligence and Financial Engineering of Sichuan, Chengdu, China
555, Liutai Avenue, Chengdu, Sichuan, P.R. China, 611130
Qiujt_t@swufe.edu.cn

Yinghong Li*
School of Humanities, Southwestern University of Finance and Economics
555, Liutai Avenue, Chengdu, Sichuan, P.R. China, 611130
abstract@swufe.edu.cn

Zhangxi Lin
The Rawls College of Business Administration, Texas Tech University
Key Laboratory of Financial Intelligence and Financial Engineering of Sichuan, Chengdu, China
Box42101, Lubbock, TX, USA, 79409
Zhangxi.lin@ttu.edu

ABSTRACT

Social commerce refers to the use of social media (or social network) to facilitate user participation in online businesses. The major configurations of social commerce websites fall into two classes, the social media with commercial features and the e-commerce websites with social networking capabilities. This paper is concerned with the question of whether the social networking functions can definitely help the e-commerce websites establish their social commerce. To answer this question, this study develops an analysis framework that incorporates four tasks: (1) explore the small-world and power-law properties of the friendship networks constructed by users of e-commerce websites, (2) detect homophily, (3) investigate social influence occurring among users, and (4) build a profile for the influencers in the network. Employing the framework to conduct an empirical analysis on the Yelp dataset, we gained insights into social commerce on Yelp, which reveals that the social networking functions do not work well and social commerce occurs only in a small user group. This study also shows a clear route by which Yelp can improve its social commerce, i.e., by encouraging more users to build friendships and by promoting users to use the social networking functions.

Keywords: Social commerce; Social network analysis; Yelp; E-commerce

1. Introduction

Recent years have witnessed an increasing popularity of social commerce, which is generally regarded as a new development of e-commerce. Broadly speaking, social commerce refers to the use of social media (or social network) to facilitate user participation in both online and offline marketplaces [Zhou et al. 2013]. Users share shopping information with their online friends, and would make the purchasing decisions based on the recommendations from online friends. Social commerce provides a brand new type of online business.

While social networking websites incorporate commercial features that allow for advertisements and transactions, the e-commerce websites, including Amazon, eBay and Yelp, are also getting interested in harnessing the power of social networking functions to promote their businesses [Curty et al. 2011; Liang et al. 2011]. For instance, Amazon provides social networking capabilities to help B2C websites better understand and serve their customers [Liang et al. 2011]. Yelp, an e-commerce website that assists people in finding great local businesses, incorporates social networking functions to improve user engagement. These functions enable the registered user to add friends, follow other users, share reviews with friends, vote on reviews, and receive compliments. Figure 1 exhibits a part of the social elements of Yelp.

* Corresponding author

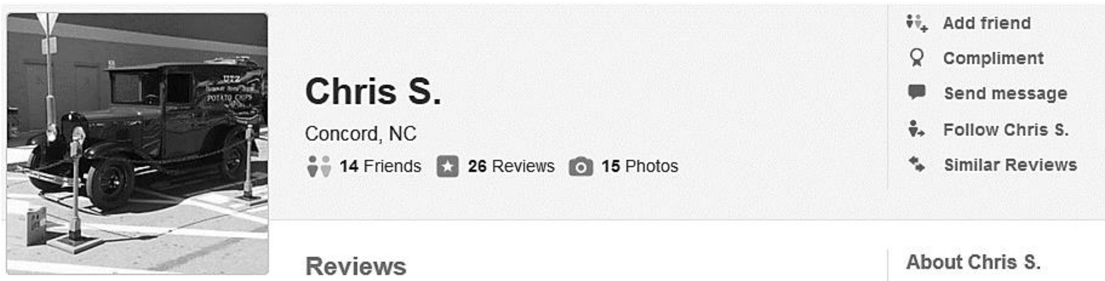

Figure 1: Social elements of Yelp

In this paper, we are concerned with the question of whether the social networking functions are able to help the e-commerce websites establish their social commerce. To the best of our knowledge, previous studies are based upon an assumption that social commerce has been well established [Lin et al. 2017; Zhang & Benyoucef 2016]. However, if an e-commerce website attempts to build the social commerce by providing the social networking functions whereas the social commerce does not work at all, the research efforts on investigating social commerce in this website would lead to the biased or even wrong conclusions.

In addition, prior work pays much less attention to the e-commerce websites that incorporated social networking functions but rather focused on the social media-based social commerce [Kim & Kim 2018]. Previous research efforts on social commerce do not address our concern.

To answer this question is an interesting and challenging task. We define the task as "detecting social commerce". In this paper, we seek to develop a general analysis framework to deal with the task. Having investigated the attributes of social commerce, including social interactions and commerce activities (see Section 3 for details), the analysis framework incorporates four tasks: (1) explore the small-world and power-law properties of the friendship networks constructed by users of e-commerce websites, (2) detect homophily, (3) investigate the social influence occurring among users, and (4) build a profile for the influencers in the network. When employing the framework to detect social commerce in an e-commerce website, if positive results are obtained, we can affirmatively provide an answer, i.e., "Social commerce works well". This analysis framework can effectively help firms improve their social commerce strategies or assist decision-making for those who plan to establish social commerce in the future.

Yelp has hosted the "Yelp dataset challenge" since 2012. In each challenge, a new dataset is issued. The Yelp datasets incorporate a rich set of information, including users, businesses[1], reviews, and social relationships. This study employs the developed analysis framework to conduct an empirical analysis on the Yelp dataset. Consequently, we gain insights into social commerce in Yelp, which reveal that the social networking functions do not work well and that the social commerce occurs only in a small user group.

The main contributions of this study include the following:

(1) Develop an analysis framework for detecting social commerce, which is an interesting problem but has received much less attention from the academic community.
(2) An empirical analysis conducted on the Yelp dataset obtains deep insights into social commerce on Yelp. We also suggest the route by which Yelp can improve its business.

The remainder of this paper is organized as follows. Section 2 reviews the related work on social commerce from both firm-side and consumer-side. Section 3 introduces the theoretical background of this study and the analysis framework. Section 4 introduces the Yelp dataset and presents the details of the empirical analysis. Section 5 offers discussions and concluding remarks.

2. Related Work

Following the development of social media, a new marketing pattern, called social commerce, has emerged. In social commerce, social media can be employed as a tool to build brand loyalty [Labrecque et al. 2014; Laroche et al. 2013] and develop new customers [Hajli et al. 2014]. Additionally, consumers are able to actively interact with consumer peers within social media to aid in their decision-making [Gabriela et al. 2014; Bapna & Umyarov 2015]. Social commerce is a popular and important research field in recent time. According to a survey [Lin et al. 2017], the current social commerce research focuses on three major research themes: organization, advertisement, and word-of-

[1] Yelp gives items, products and services in e-commerce a general name, businesses.

mouth. However, we can also find research efforts on social commerce that fall into two categories, i.e., firm-side and customer-side.

From the firm perspective, certain work investigates how social factors influence the marketing performance of firms. Kankanhalli et al. [Ha et al. 2016] demonstrated that interactions between firms and customers on social media, more specifically customer engagement behaviors through their SMM (short for social media marketing) messages, exert positive impacts on increasing sale performance and building trust and reputation with customers. [Kumar et al. 2016] found that the numbers of SMM messages posted by the firms have a positive correlation with the expenditure of customers, and even can affect their cross-buying behavior. [Bai & Yan 2020] indicated that firm social media marketing has a significant positive impact on firm performance.

Certain studies focus on leveraging social commerce to build brand loyalty. [Laroche et al. 2013] indicated that the social media accounts of brands actively influence relationships between customers and products, and even firms, and the positive impact can help build brand trust and brand loyalty. [Labrecque 2014] designed successful social media strategies to solidify the place of brands in social media environments. Other studies investigate factors that affect brand loyalty in social commerce. [Hew et al. 2016] found two important factors that can positively influence brand loyalty, including satisfaction regarding social commerce and continued intention to use social commerce. [Herrando et al. 2019] indicated that hedonic stimulus and utilitarian stimulus affect users' flow experience to positively impact emotional and behavioral loyalty.

Basic behavioral psychology drives consumers trust their friends more than anonymous users. Therefore, the purchase behaviors of their friends play an important role in the decision-making of consumers. For example, the odds of a user adopting a paid subscription increase by 50% due to peer influence when her friend adopts it [Bapna & Umyarov 2015]. From the consumer perspective, many research efforts explore how social networks help customers shape their purchase decisions and shopping intentions. [Bai et al. 2015] demonstrated that social factors significantly enhance users' purchase intentions in social shopping.

All the above results are derived under the assumption that social commerce has been well established. If this assumption was not true, however, the results would lose credibility. Therefore, detecting social commerce is an important and noteworthy task.

3. Theoretical background and analysis framework

Liang et al. summarized three major attributes of social commerce: social media technologies, social interactions[2], and commercial activities [Liang & Turban 2011]. For our purposes, i.e., detecting social commerce in the e-commerce websites that incorporate social networking functions, we conduct a quantitative analysis of both social interactions and commercial activities attributes.

3.1. Social interactions

A social network, which leads to frequent interactions among friends, is a key component for establishing social commerce. Social networks generally exhibit small-world properties [Mislove et al. 2007; Adamic et al. 2003; Leskovec & Horvitz 2008; Watts & Strogatz 1998; Fleming & Marx 2006], i.e., the network have a small average shortest path length and a high clustering coefficient so that any two nodes can be connected in a small number of steps. For example, MSN messenger network is reported that the degrees of separation is 7.8 [Leskovec & Horvitz 2008]. A patent network constructed by inventors presents a small-world property, which affects how innovation is realized [Fleming & Marx 2006]. Cho et al. examined the social network structure of the United States Congress from 1973 to 2004. The results showed that Congress exemplifies small-world properties [Adamic et al. 2003].

The power-law degree distribution is also an important property of social networks [Mislove et al. 2007; Stephen & Toubia 2009; Adamic et al. 2003], called the power-law property. Formally, a network is said to have a power-law degree distribution if the degree of a node k can be expressed mathematically as $\mathrm{p(k)} \propto \mathrm{k}^{-\alpha}$, where the parameter α is called the power-law coefficient. Most real networks including social networks have a power-law coefficient between 2 and 3 [Kwak et al. 2010]. Mislove et al. [Mislove et al. 2007] confirmed that online social networks have power-law properties by examining the structure of multiple online social networks, such as Flickr, YouTube, LiveJournal, and Orkut. Stephen et al. [Stephen & Toubia 2009] developed a statistical model that explains the emergence of power-law degree distribution in social commerce networks.

Users of e-commerce websites build a friendship network by using social networking functions. Because social commerce generally refers to the delivery of e-commerce activities and transactions via the social networks [Liang &

[2] Instead of "community interactions" used in the original paper, we argue that social interaction is a better term for depicting the features of social commerce.

Turban 2011], we argue that if social commerce works well, the social network properties including the small-world and power-law properties should can be detected in the friendship network.

Both small-world and power-law properties describe the features of a social network at a certain time point (i.e., a static social network). Homophily provides an approach to measure the tendency of network evolution. Homophily can be observed when similar individuals become friends due to their high similarity. Halberstam et al. [Halberstam & Knight 2016] found strong evidence of homophily on Twitter, i.e., conservatives and liberals are more likely to link to users with the same faith. The results provide evidence that homophily is a property of dynamic social networks.

In social commerce, homophily refers to a tendency for individuals to choose friends with similar tastes and preferences [Aral & Walker 2012]. Building friendship means that high trust occurs among users. The trusts further promote interactions among friends [Hajli et al. 2014]. Hence, homophily helps promote social interactions. Homophily is one of the key indicators of successful social commerce.

3.2. Commercial activities

As [Iyengar et al. 2010] pointed out, the success of network-based marketing depends on whether peers actually influence one another. In the context of social commerce, social influence refers to the behavior of consumers influencing their friends to purchase businesses [Ma et al. 2014], and commercial activities refer to the purchase behaviors resulting from social influence. We thus argue that social influence has the potential to shape commercial activities and becomes one of the cornerstones of building social commerce. Hence, detecting purchase behaviors under social influence can assist the task of detecting social commerce.

Iyengar et al. also indicated that opinion leaders are able to affect the adoption and diffusion of new products [Iyengar et al. 2010]. Hence, if we can find out influencers (opinion leaders) who influenced the purchase decisions of other users in the friendship network of a website, the results can provide evidence that social commerce works on the website.

According to the above discussions, we build an analysis framework incorporating four subtasks to answer the question of whether social commerce works in an e-commerce website. Figure 2 exhibits the inner logic of the analysis framework that follows a logical flow from a task decomposition perspective. Starting from the left side in Figure 2 and moving right, having finished every subtask, the question can be answered.

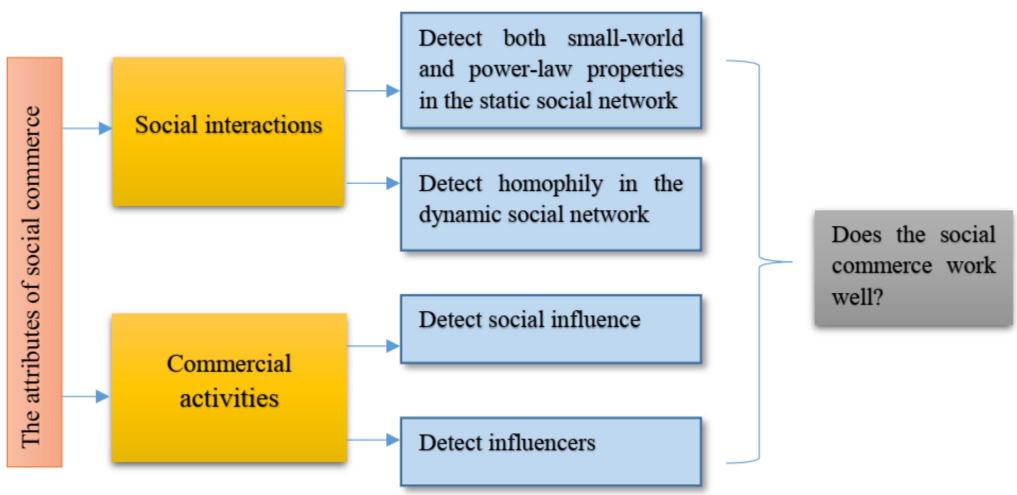

Figure 2: An analysis framework for detecting social commerce in a website

4. Empirical Analysis[3]

4.1. Data

Since 2012, Yelp has hosted the "Yelp dataset challenge". By the end of 2019, they had launched 13 rounds of this challenge. In each of the rounds, a new dataset is issued. The datasets are incremental updates, i.e., the dataset in the *i*th round comprises the dataset in the *i*-1th round and new data. As discussed in Section 3, the analysis framework requires exploring both static and dynamic social networks. This study employs the 6th round Yelp dataset (Yelp6) , the 8th round dataset (Yelp8), the 11th round dataset (Yelp11), and the 13th round dataset (Yelp13) for the empirical analysis.

[3] See source codes in https://github.com/Allen-Qiu/social_commerce

The Yelp datasets provide a rich set of contents, including users, reviews, businesses, and friendships. We build the user table (see Table 8), review table, business table, and friendship networks from the original Yelp dataset for this study. The attributes of reviews and businesses are listed in Table 1.

Table 1: The attributes of reviews and businesses

Reviews	
Id	Identifier of one review
User	Reviewer of the review
Stars	Star rating given by the reviewer
Contents	Contents of the review
Business	Target of the review (business identifier)
Business	
Id	Identifier of one business
Star	Star rating of the business

We build the networks using users and friendship as nodes and edges, respectively, which are the undirected and unweighted networks. However, the original networks are not the connected networks. We detect the largest connected graphs from the original networks as the friendship networks.

Every network is a snapshot of the Yelp website. For example, Yelp6 presents the status of users, the relationship of users, and the status of the businesses on the Yelp website at the end of July 2015. Likewise, Yelp8 presents the information at the end of July 2016. Based on four static networks, we can investigate the evolution of the friend network of the Yelp.

Some statistics of four the datasets and the networks are listed in Table 2.

Table 2: Statistics of four Yelp datasets and their friendship networks.

	Yelp6	Yelp8	Yelp11	Yelp13
Date of the latest review	2015-01-08	2016-07-19	2017-12-11	2018-11-14
Reviews	1.57M	2.54M	5.26M	6.69M
Users	367K	687K	1.32M	1.63M
Businesses	61K	85K	174K	192K
Nodes of friendship network	168K	289K	593K	765K
Edges of friendship network	1.29M	2.10M	5.32M	7.39M
Percentage of users not having friends	53%	57%	43%	42%
Average degree	14.80	14.09	17.55	18.99
Average shortest path length	4.43	4.37	4.75	4.74
Clustering coefficient	0.12	0.11	0.10	0.09

We can observe that 53%, 57%, 43%, and 42% of users have no friends at all in four datasets, respectively. The statistics reveal that users tend not to be interested in the social networking function provided by Yelp.

This study regards a review in the Yelp dataset as a purchase instance.

4.2. Small-world and power-law properties

According to the proposed analysis framework, this subsection seeks to investigate both small-world and power-law properties in friendship networks in the Yelp dataset. We first compute the average shortest path length and the clustering coefficient of four friendship networks, which are listed in Table 2. To help understand these measures, we also compare them with those of the existed networks, including social networks and physical networks, which are listed in Table 3.

Table 3: The average shortest path length and the clustering coefficient of networks

Network	Clustering coefficient	Average shortest path length
Power grid [Watts & Strogatz 1998]	0.080	18.7
WWW [Mislove et al. 2007]	0.081	16.12
YouTube [Mislove et al. 2007]	0.136	5.10
Flicker [Mislove et al. 2007]	0.313	5.67
SPIRES [Newman 2001]	0.726	4.0

Table 3 shows that both the power grid and WWW have a clustering coefficient lower than 0.1. The networks in social networking sites, such as YouTube and Flicker, have values higher than 0.1. SPIRES is a scientific collaboration network built from papers published in high-energy physics. The clustering coefficient of the network reaches the highest value of 0.726. We further explore the average shortest path length and find that all three social networks such as YouTube, Flicker, and SPRIES have a short average shortest path length, which exhibits a small-world phenomenon well. However, the two physical networks, power grid and WWW, have a much larger average shortest path length.

Compared with these networks, the clustering coefficients of Yelp are lower than those of the social networks but higher than those of the physical networks. The average shortest path lengths of Yelp are almost identical to those of social networks. Yelp presents weak small-world property.

As discussed in Section 3, we further determine whether four friendship networks exhibit the power-law property. Visualizations is a typical approach of qualitative appraisal for the purpose by drawing log–log plots of the degree distribution [Stephen & Toubia 2009]. If the log–log plots are close to linear, suggesting that the network appears to have a power-law degree distribution. Figure 3 illustrates the log-log plot of the degree distributions of the four networks. Denote the degrees by the x-axis, cumulative distribution function $p(X \geq x)$ of degrees by the y-axis and the power law fit by the dash line in Figure 3. We can observe the degree distribution (blue circle line) roughly fits the power law distribution (dash line).

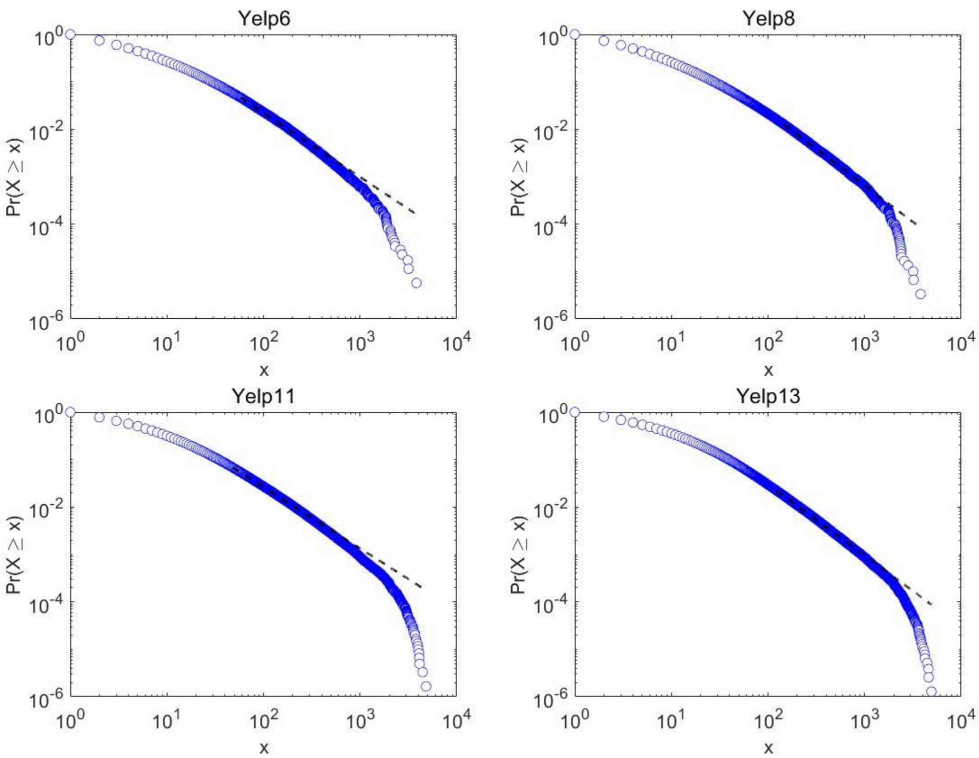

Figure 3: Log-Log plots for degree distribution in the friendship networks

According to a work in [Clauset et al. 2009], the log-log plot can be deceptive and can lead to a claim of power-law behavior that does not hold up under closer scrutiny. The goodness-fit-test is required to quantify the plausibility of a hypothesized power-law distribution.

We thus employ the approach of goodness-fit-test developed in [Clauset et al. 2009] to test the power-law hypothesis quantitatively for the four networks. The approach first fits the degree distributions of the networks to the power law model, and subsequently generates a large number of synthetic data sets from the power-law model. Each synthetic data set is individually fitted to its own power-law model. Additionally, the Kolmogorov-Smirnov (KS) statistic [Clauset et al. 2009] for each one relative to its own model are also calculated. After counting what fraction of the time the resulting statistic is larger than the value for the empirical data, this fraction is regarded as the p-value, which varies in the range from 0 to 1. If the p-value is close to 1, the result indicates the degree distribution can be perfectly fit to a power law distribution. Generally, if p-value is less than 0.1, we would reject the power law hypothesis.

The power-law coefficient α, the lower bound of the distributions x_{min}, and p-values of four friendship networks are listed in Table 4. We can find the p-values of four networks are all zero. We thus conclude that the four friendship networks do not exhibit the power-law property.

Table 4: The fitting of power-law distribution

	α	x_{min}	p-value
Yelp6	2.35	58	0
Yelp8	2.54	147	0
Yelp11	2.29	48	0
Yelp13	2.50	123	0

We can draw the following conclusion based on the above analysis.

Conclusion 1. The friendship networks of Yelp exhibit a weak small world property but not power law property. The social networks in Yelp have not established well yet.

4.3. Homophily

Social commerce becomes deeply involved in, and dependent on, the social relationship among consumers. The aforementioned discussions show that the friendship networks in the Yelp dataset present weak social characteristics. In this subsection, we seek to detect homophily, which is one of the indicators of social commerce, by exploring the evolution of friendship networks.

To calculate homophily first requires determining the assortativity of the friendship network. A network is called (dis)assortative when nodes in the network preferentially connect to nodes with (dis)similar properties [Newman 2003]. In an assortative network, links are more likely to occur among similar nodes than dissimilar nodes. For example, in social networks, friends always present high similarity on interests, activities, and preferences.

Assortativity is a property of a static social networks, whereas homophily occurs in a dynamic social network. Homophily of a network can be measured by investigating the change of the assortativity of a network over time. Consider two snapshots of a network G at times t1 and t2, i.e., $G_{t1}(V, E_{t1})$ and $G_{t2}(V, E_{t2})$, where $t2 > t1$. The homophily can be measured using an equation written in the form of

$$H = Q^{t2} - Q^{t1} \tag{1}$$

where Q^{t1} and Q^{t2} are the assortativity values of G_{t1} and G_{t2}, respectively. This study calculates the assortativity of the friendship network of four Yelp datasets and then measures the homophily in Yelp.

The assortativity of a network can be calculated if every node in the network has a categorical label, such as race, gender or language. In this study, we assign a label to every node and then calculate the assortativity of both friendship networks regarding the categorical label. The corresponding equation is written in the form of

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} \tag{2}$$

where e denotes a matrix of types, i.e., e_i refers to a categorical type. e_{ij} indicates the fraction of edges in a network that connect the nodes of type i to the nodes of type j. They satisfy sum rules

$$\sum_{ij} e_{ij} = 1, \sum_j e_{ij} = a_i, \sum_i e_{ij} = b_j \tag{3}$$

The value of r varies in a range from -1 to 1. r = 1 informs a perfect assortativity whereas r =-1 informs a perfect disassortativity. If every node has a numeric property, the assortativity can also be calculated by the numeric value. A study [Newman 2003] provide a detailed discussion of assortativity.

To explore the homophily of Yelp, we calculate three types of assortativity: labels, degrees of nodes, and the number of purchases.

We first attempt to assign a label to each user according to their favorite business type. This is rooted in the idea that friends usually have similar interests and have common favorites. We summarize the most frequently purchased business category for every user and then assign the business categories as the labels to the users. Table 5 lists the assortativity of four friendship networks. We find that the four networks do not exhibit assortativity.

Furthermore, we calculate assortativity according to the numeric value. We first calculate the degree-assortativity of four networks. The degree-assortativity denotes an extent to which nodes are connected with other nodes of a similar degree. The Yelp6 has a small negative value, whereas Yelp8 has a small positive value. Although the larger values appear in Yelp11 and Yelp13, they are not sufficient to inform a good assortativity. We further calculate assortativity regarding the number of purchases. We obtain the largest assortativity in Yelp13 with a value of 0.227, which informs a mild assortativity. When exploring the latest status of Yelp (Yelp13), two out of three type of measures in Yelp13 show that friendship networks do not present assortativity. The result indicates that users still are not active.

We calculate homophily using the gap of assortativity between Yelp13 and Yelp6. The positive H occurs in the all three type of assortativities. They indicate that the homophily can be detected in Yelp, and there exists a trend that user are becoming gregarious.

Table 5: Assortativities of the four friendship networks and Homophily

	Assorativity				Homophily
	Yelp6	Yelp8	Yelp11	Yelp13	
Label by Business type	0.033	0.037	0.053	0.045	0.012
Degree (numeric)	-0.0074	0.0042	0.074	0.097	0.104
Purchase (numeric)	0.026	0.030	0.205	0.227	0.201

We can draw the following conclusion based on the above discussions.

Conclusion 2: Yelp presents a trend such that the social network is building, but social interactions are still not active enough.

4.4. Social influence

Because Yelp dataset maintains incremental updates, we conduct experiments on Yelp13 that contains all the purchase data in the previous datasets. We examine every purchase instance on the Yelp13 dataset and partition them into two groups: the affected purchase instances and the unaffected purchase instances, which result from the affected purchase behaviors and the unaffected purchase behaviors, respectively.

Definition 1 (The affected purchase behavior and the unaffected purchase behavior). Given a consumer A, if A purchases a business P and an A’s friend B bought P before, we call the purchase behavior of “A purchases P” an affected purchase behavior. If A purchases a business P and none of his/her friends bought P before, we call this purchase behavior an unaffected purchase behavior.

We should stress that ‘the unaffected purchase behavior’ refers to the purchases definitely not being affected by friends. However, ‘the affected purchase behavior’ does not have to indicate that the purchase was truly affected by friends. We provide this definition for the purpose of investigating the impact of social relationships on purchase behaviors on the “the affected purchase behaviors”. This subsection designs an experiment to perform the task.

Table 6 lists the statistics of Yelp13 regarding Definition 1. We can observe that the majority of purchases are unaffected purchase behaviors because only 14.2% of the purchases fall into the affected purchase category.

Table 6: The statistics of purchase instances Yelp13

The affected purchase instances	14.2%
Users who have the affected purchase behavior	7.6%
Users who affected friends on purchase	7.1%

This subsection explores social influence from the perspective of the impact of friend relationships on purchase decisions. If we are able to determine that friend relationships can definitely influence the purchase decisions of users in the dataset, the result will confirm the existence of social influence.

Having explored the Yelp website, we summarize three factors that influence the purchase behavior of consumers, including the ratings of businesses, social factors and other unobserved factors. Every business in Yelp have an overall star-rating assigned by Yelp and every consumer can assign a star-rating to the purchased business. A previous work has proved that the ratings of business have a significant impact on the purchase decisions of consumers in Yelp and consumers prefer to purchase the high-rated business [Qiu et al. 2018]. Based on this conclusion, if we can find the affected purchase behaviors in businesses that do not have the high overall ratings but the friends of the consumer give higher ratings than the overall ratings, we can conclude that the social relationships make the impacts on the purchase behaviors.

Based on this fact, we build the following analysis steps. We first extract the affected purchase instances to build a dataset D_A. Given a consumer A and a business P purchased by A, we examine the difference between the overall ratings of business P, i.e., $\mathrm{R}_{business}$ (it is provided by Yelp), and the average ratings that A’s friends assign to P before A’s purchase, i.e., R_{friend}. If the friend relationship does influence the decision-making of purchases, the error of R_{friend} and $\mathrm{R}_{business}$ should be significantly larger than zero. Employing statistical tests, if we can reject the null hypothesis $\mathrm{R}_{friend} - \mathrm{R}_{business} \leq 0$, we would conclude that R_{friend} is significantly larger than $\mathrm{R}_{business}$, i.e., friend relationships do have the impacts on the purchase decisions of consumers.

We build a dataset D_B from D_A. Each instance in D_B is a purchase behavior that contains two fields, including R_{friend} and $\mathrm{R}_{business}$. Because the star ratings of the Yelp dataset vary in the range from 1 to 5, we perform Student's t-Test on all businesses and on businesses with 1~5 star ratings, respectively. The results are listed in Table 7.

Table 7: The results of the t-test on DB

Ratings	All	5-star	4-star	3-star	2-star	1-star
The mean of $\mathrm{R}_{friend} - \mathrm{R}_{business}$	0.352	-0.127	0.169	0.584	0.927	0.997
The number of instances	950,001	16,279	541,716	344,749	44,547	2710
p-value	<2e-16	1	<2e-16	<2e-16	<2e-16	<2e-16

Table 7 shows that the p-value of "All businesses" is far less than 0.05. Hence, we reject the null hypothesis $\mathrm{R}_{friend} - \mathrm{R}_{business} \leq 0$. This result means that when consumer A purchases business P, the average ratings assigned by A's friends are significantly higher than the overall ratings of the business. We can conclude that the friend relationship influences the purchase decisions of consumers when investigating all businesses in D_B. We further explore the businesses relative to each star-rating. For businesses with 5-star ratings, the average ratings of friends are less than 5. The results do not surprise us because the highest rating is 5-star. For businesses with 4-star, 3-star, 2-star, and 1-star, their p-values are all less than 0.05. We thus reject the null hypothesis. The results indicate that R_{friend} is significantly higher than $\mathrm{R}_{business}$, in each of star ratings except for 5-star. We can draw the following conclusion from the above analysis.

Conclusion 3: The majority of purchases belong to the unaffected purchase behaviors category in Yelp. The friend relationships have significant impacts on consumers who have at least one affected purchase. Social influence can be detected in a small user group in Yelp.

4.5. Building a profile for influencers in social commerce

Conclusion 3 shows that "The friend relationships significantly influence consumers who have at least one affected purchase". In this subsection, we seek to build a profile for influencers who impacted others in making decisions of purchases. If we can successfully build this profile, the result will provide empirical evidence to support the claim that social commerce works in Yelp.

This study employs a logistic regression (LR) to build the profile for the influencers in Yelp13. According Definition 1, we assign a label "yes" to those who influenced his/her friends at least once; otherwise, we assign the label "no". Consequently, a dataset is built with 115,930 and 1,521,208 users falling into the two groups, respectively. However, this is an imbalanced dataset. If we train a LR model on the dataset, the majority class would dominate the training process. Consequently, the model would lead to a biased or even wrong analysis results. We employ the randomly oversampling on the minority class to make a balanced dataset D_C and then train an LR model on D_C. By exploring the coefficients of the LR model, the important features of influencers can be derived. Table 8 lists all attributes of users.

Table 8: A list of attributes of users in the dataset Dc

Attributes	Description
v	The number of 'votes' sent by the user
c	The number of 'compliments' received by the user
reviewcount	The number of reviews
fans	The number of fans the user has
avgstars	The average star rating
friends	The number of friends the user has
pageRank	The PageRank score of users
elite	A label that indicates the users is ' elite' in Yelp
label	"yes" or "no"

Yelp provides three types of votes: 'funny votes', 'useful votes', and 'cool votes'. They are summed up in the variable *v*. Likewise, *c* is the sum of the eleven types of compliments. PageRank is originally developed to produce a global "'importance" ranking for every web pages in search engine. Nowadays, PageRank and its variants have been widely applied on ranking nodes in the network. This study employs PageRank to calculate a PageRank score for every user in the friendship network, which informs the importance of each user in the friendship network.

Before training the model, we must detect multicollinearity because it can limit the analysis results. Variance inflation factors (VIFs) are a popular measure of the amount of multicollinearity in a set of multiple regression variables. We use a stepwise VIF method to exclude the highly correlated variables from the model. The approach works in the following steps. In each of the iterations, the VIF is calculated for every variable, and then the variable with the highest VIF is removed. The process is repeated until the all VIFs are lower than a specific threshold.

If the VIF is equal to 1, the result informs that the variable does not have multicollinearity. A VIF between 1 and 5 indicates moderately correlated. In practice, the threshold is generally set to 5. Table 9 shows the VIF of every variable is less than 5 except for PageRank. When calculating the correlation of variables, PageRank and Friends present a high Pearson correlation coefficient with R=0.67. When removing PageRank, The LR model does not have the multicollinearity problem.

Table 9: VIF of variables

Attributes	VIF
v	4.98
c	3.14
reviewcount	4.57
Fans	2.54
Avgstars	1.05
Friends	1.17
PageRank	6.30
Elite	1.547

Furthermore, we measure the goodness of fit of the model because an analysis of selecting important features seems credible only when the model fits the data well.

Table 10: Goodness of fit of the model

Goodness of fit	Value
Pseudo R2 (McFadden)	0.482
Pseudo R2 (Maximum likelihood)	0.487
Pseudo R2 (Cragg and Uhler)	0.650
Accuracy	83.5%

When employing Pseudo R2 (pR2 for short) to evaluate goodness of fit, the value varies in the range from 0 to 1. A value closer to 1 indicates a perfect fit. There is not an exact line between a good fit and a bad fit. In practice, a model satisfying pR2 >0.2 can be regarded as an acceptable model. Table 10 shows that all three types of pR2 have a value far larger than 0.2. We also employ the trained LR model to classify the training set, and subsequently use the accuracy to evaluate the goodness of fit. Consequently, the model reaches a high accuracy of 83.5%. We can conclude that the model has a good fit to the dataset.

Following the conclusion, we further explore features of influencers. All variables are normalized within a range of [0, 1]. Accordingly, we can compare the importance of all variables according to their coefficients. The coefficients of the LR model are listed in Table 11.

Table 11: Coefficients of the LR model

Variables	Coefficients	P-value
(intercept)	-4.545	<2e-16
Log(c+1)	2.452	<2e-16
Log(reviewcount+1)	2.473	<2e-16
Log(fans+1)	-3.481	<2e-16
Avgstars	0.204	<2e-16
Log(Friends+1)	9.487	<2e-16
Elite	-0.333	<2e-16

Table 11 shows that every variable has a p-value less than 0.05 and is therefore significant for the model. Obviously, friends has the largest positive value. *c* and *reviewcount* also have a positive value. We can build a profile for the influencers of social commerce in Yelp: (1) They have many friends; (2) they are keen on writing reviews for businesses; and (3) they received many compliments from others. On the other hand, because we successfully build a profile for the influencers, the following conclusion can be drawn.

Conclusion 4: There exists influencers who can definitely affect the purchase decisions of other users in Yelp.

5. Discussion and Implications

This study develops an analysis framework for the task of investigating whether social commerce works on an e-commerce website with social networking functions, and subsequently conducts an empirical analysis on the Yelp dataset. From the analysis results, we gain deep insights into social commerce in Yelp, described as follows.

(1) The majority of users on Yelp are not gregarious. Only 47%, 43%, 57%, and 58% of users have friends in four Yelp datasets, respectively. This insight means that the efforts of building social commerce in Yelp could be weakened by fewer social relationships.

(2) The friendship networks in Yelp exhibit weak small-world property and bad power-law property. This insight indicates that the social network has been not formed well, social interactions among users are not active in Yelp.

(3) The majority of purchase behaviors are not affected by others. In Yelp13 dataset, only 7.6% users have the affected purchase behavior.

With the above three insights, Yelp seems to be pessimistic for building its social commerce. However, we still derive positive insights from the following analysis results.

(4) Homophily is detected. Yelp presents a trend such that social network is building.

(5) From the affected purchase instances, we view a phenomenon in which friend relationships do have an impact on making purchase decisions.

(6) From the affected purchase instances, we successfully build a profile for influencers. The result shows that there exists users who can definitely affect the purchase decisions of other users in Yelp.

Combining all insights, we can conclude that the social networking functions in Yelp do not work well, and social commerce occurs only in a small user group.

This study identifies the salient features of influencers in social commerce. (1) They have many friends; (2) they are keen on writing reviews on businesses; and (3) they have received many compliments from others. These findings provide support for implementing social commerce strategies.

The derived insights also show a clear route by which Yelp can improve its social commerce: (1) encourage more users to build friendships with others and (2) promote users to use social networking functions to make users more gregarious. These efforts will make the social commerce more active and consequently lead to its success.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 71571145), and the Humanities and Social Science Foundation of Ministry of Education of China (No. 14YJAZH063).

REFERENCES

Adamic, L. A., O. Buyukkokten, and E. Adar, “A social network caught in the Web,” *First Monday*, Vol. 8, No. 6, 2003.

Aral, A. and D. Walker, “Identifying Influential and Susceptible Members of Social Networks,” *Science*, Vol. 337, No. 6092: 337-341, 2012.

Bai, L. and X. Yan, “Impact of Firm-generated Content on Firm Performance and Consumer Engagement: Evidence from Social Media in China,” *Journal of Electronic Commerce Research*, Vol. 21, No. 1: 56-74, 2020

Bai, Y., Z. Yao, and Y. Dou, “Effect of social commerce factors on user purchase behavior: An empirical investigation from renren.com,” *International Journal of Information Management*, Vol. 35, No.5: 538–550, 2015.

Bapna, R. and A. Umyarov, “Do Your Online Friends Make You Pay? A Randomized Field Experiment on Peer Influence in Online Social Networks,” *Management Science*, Vol. 61, No. 8: 1741-2011, 2015.

Clauset, A., C. R. Shalizi, M. E. J. Newman, “Power-law distributions in empirical data,” *SIAM Review*, Vol. 51, No.4: 661-703, 2009.

Curty, R.G. and P. Zhang, “Social commerce: looking back and forward,” *Proceedings of the American Society for Information Science and Technology*, Vol. 48, No. 1: 1–10, 2011.

Fleming, L. and M. Marx, “Managing Innovation in Small-worlds,” *Management Review*, Vol. 48, No. 1: 8–9, 2006.

Gabriela, L. P. A., L. F. Hor-meyll, and de Paula Pessôa, L. A. G., “Influence of virtual communities in purchasing decisions: The participants’ perspective,” *Journal of Business Research*, Vol. 67, No. 5: 882–890, 2014.

Ha, S., A. Kankanhalli, S. J. Kishan, K. W. Huang, "Does Social Media Marketing Really Work for Online SMEs?: An Empirical Study," *Proceedings of 37th International Conference on Information Systems*, 2016.

Hajli, N., X. Lin, M. S. Featherman, and Y. Wang, "Social word of mouth: how trust develops in the market," *International Journal of Market Research*, Vol. 56, No. 5: 673–689, 2014.

Halberstam, Y. and B. Knight, "Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter," *Journal of Public Economics*, Vol. 143, November: 73–88, 2016.

Herrando, C., J. Jiménez-Martínez, M. J. M. Hoyos, "Social Commerce Users' Optimal Experience: Stimuli, Response and Culture," *Journal of Electronic Commerce Research*, Vol. 20, No. 4, 2019

Hew, J. J., V. H. Lee, K. B. Ooi, B. Lin, "Mobile social commerce: The booster for brand loyalty?," *Computers in Human Behavior*, Vol. 59, June: 142-154, 2016.

Iyengar, R., V. Christophe, and T. W. Valente, "Opinion Leadership and Social Contagion in New Product Diffusion," Marketing Science, Vol. 30, No. 2: 195-388, 2010.

Kim, N. and W. Kim, "Do your social media lead you to make social deal purchases? Consumer-generated social referrals for sales via social commerce," *International Journal of Information Management*, Vol. 39, April: 38–48, 2018.

Kumar, A., R. Bezawada, R. Rishika, R. Janakiraman, and P. k. Kannan, "From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior," *Journal of Marketing*, Vol. 80, No. 1: 7–25 2016.

Kwak, H., C. Lee, H. Park, and S. Moon, "What is Twitter, a Social Network or a News Media?," *Proceedings of the 19th international conference on World Wide Web*, 2010.

Labrecque, L. I., "Fostering Consumer–Brand Relationships in Social Media Environments: The Role of Parasocial Interaction," *Journal of Interactive Marketing*, Vol. 28, No. 2: 34–148, 2014.

Laroche, M., M. R. Habibi, and M. R. Richard, "To be or not to be in social media: How brand loyalty is affected by social media?," *International Journal of Information Management*, Vol. 33, No. 1: 76–82, 2013.

Leskovec, J. and E. Horvitz, "Planetary-scale views on a large instant-messaging network," *Proceedings of the 17th international conference on World Wide Web*, 2008

Liang, T. P. and E. Turban, "Introduction to the special issue Social Commerce: A Research Framework for Social Commerce," *International Journal of Electronic Commerce*, Vol. 6, No. 2: 5–13, 2011.

Lin, X., Y. Li, X. Wang, "Social commerce research: Definition, research themes and the trends," *International Journal of Information Management*, Vol. 37, No. 3: 190–201, 2017.

Ma, L., R. Krishnan, and A. L. Montgomery, "Latent Homophily or Social Influence? An Empirical Analysis of Purchase Within a Social Network," *Management Science*, Vol. 61, No. 2: 454-473, 2014.

Mislove, A., M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 2007.

Newman, M. E. J., "The structure of scientific collaboration networks," *PNAS*, Vol. 98, No. 2: 404-409, 2001.

Newman, M. E. J., "Mixing patterns in networks," *Physical Review E*, Vol. 67, No. 2: 026126-026138, 2003.

Qiu, J., Y. Li, and Z. Lin, "Does Social Commerce Work in Yelp? An Empirical Analysis of Impacts of Social Relationship on the Purchase Decision-making," *Proceedings of Pacific Asia Conference on Information Systems*, 2018.

Stephen, A. T. and O. Toubia, "Explaining the power-law degree distribution in a social commerce network," *Social Networks*, Vol. 31, No.4: 262–270, 2009.

Watts, D. J. and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, Vol. 393: 440–442, 1998.

Zhang, K. and M. Benyoucef, "Consumer behavior in social commerce: A literature review," *Decision Support Systems*, Vol. 86, June: 95–108, 2016.

Zhou, L., P. Zhang, and H. Zimmermann, "Social commerce research: An integrated view," *Electronic Commerce Research and Applications*, Vol. 12, No. 2: 61–68, 2013.