J Syst Sci Syst Eng (Sep 2013) 22(3): 257-282
DOI: 10.1007/s11518-013-5225-6
ISSN: 1004-3756 (Paper) 1861-9576 (Online)
CN11-2983/N

SOCIAL MEDIA RESEARCH: A REVIEW*

Junjie WU[1] **Haoyan SUN**[2] **Yong TAN**[2]

[1] *School of Economics and Management, Beihang University, Beijing 100191, China*
wujj@buaa.edu.cn

[2] *Michael G. Foster School of Business, University of Washington, Seattle, Washington 98195-3226, U.S.A.*
haoyan@uw.edu, *ytan@uw.edu* (✉)

Abstract

Social media is fundamentally changing the way people communicate, consume and collaborate. It provides companies a new platform to interact with their customers. In academia, there is a surge in research efforts on understanding its effects. This paper aims to provide a review of current status of social media research. We discuss the specific domains in which the impacts of social media have been examined. A brief review of applicable research methodologies and approaches is also provided.

Keywords: social media, empirical models, experimental methods, analytical approaches, predictive analytics

1. Introduction

There is a growing consensus that social media is fundamentally changing the way people communicate, consume and collaborate. Within just a few years, social media has integrated itself into almost every aspect of our personal and professional lives, and as a result, has brought a lot of changes on businesses in promoting sales, gaining customers and establishing trusts. Managers try to identify ways in which firms can make profitable use of social media applications such as Blogs, YouTube, Facebook, and Twitter.

Different social media platforms play different roles in sharing information, connecting people, and creating knowledge. However, roles have evolved over time to suit the needs of social interactions and business strategies. Facebook and Twitter, traditionally designed to connect people, express opinions and re-kindle real life social relationships, have now been widely used as a marketing tool to gain brand awareness or obtain customer feedbacks. Indeed, Kumar et al. (2013) show that social media can be used to generate growth in sales, return on investment, and positive word

* This research was partially supported by the National Natural Science Foundation of China (NSFC) under Grants 71322104,71171007,70901002 and 71031001, by the National Information Security Research Plan of China under Grant 2012A137, by the Foundation for the Author of National Excellent Doctoral Dissertation of PR China under Grant 201189, and by the Program for New Century Excellent Talents in University under Grant NCET-11-0778. Dr. Yong Tan was supported in part by NSFC under Grants 71328103 and 71231002.

of mouth, and can spread brand knowledge further. Luo et al. (2013) find a significant predictive relationship between social media and firm equity value. Social media-based metrics (Web blogs and consumer ratings) are observed to be significant leading indicators of firm equity value, stronger than conventional online behavioral metrics (Google searches and Web traffic).

As technology and innovation are moving forward, there are a lot more social and physiological patterns that can be observed from social media. Besides helping organizations to guide, promote, and shape online conversations, social media has also worked as a useful tool to consumers when making decisions or gaining knowledge. For example, consumers have proved to learn from reviews more effectively than from their own experience with other books of the same genre (Zhao et al. 2013). Customer influences, through social media platforms, grow exponentially and extend beyond geographic locations and individual preferences. Therefore, it is prudent for firms to engage customers in their social media brand communities. Such customer involvements have shown to lead to a positive increase in purchase expenditure (Goh et al. 2013). Rishika et al. (2013) find that customer participation in a firm's social media efforts leads to an increase in the frequency of customer visits.

The social media phenomena have also attracted attention from researchers. For example, Information Systems Research journal has recently published a special issue on social media (Aral et al. 2013). In this paper, we aim to provide a review of recent studies related to social media applications, based on the primary way in which information is created and exchanged on those applications, such as:

- E-commerce sites with user-generated content feature (e.g., Amazon.com)
- Content sharing sites (e.g., YouTube, Flicker, blogs)
- Social networking sites (e.g., Facebook, Twitter, Linkedin)
- Virtual community (e.g., online gaming, Second Life)
- Collaborative projects (e.g., Wikipedia, Dropbox, Google Docs)

Through examining major streams in social media research, we hope to provide a broad research agenda on social media and social networks that highlights the contributions made till now and stimulates interest in future research endeavors.

The rest of the paper is organized as follows. In next section, we discuss the contexts under which social media research have been conducted. This is followed by discussions on empirical, experimental, and analytical models for descriptive analytics. Then, we review the research methodologies in the area of predictive analytics. Finally, we offer some concluding remarks.

2. Genres of Social Media Research

Social media has given companies access to unprecedented volumes of information about their clients and purchasing behaviors on an aggregate level. It has also integrated itself into uses that go way beyond the familiar applications for sales, marketing and community building. The way social media is rolled out has been evolving towards different goals; internal networks, mobile applications, social-care,etc.

have all started embracing social media tools. The challenge, which confronts everyone from small businesses to researchers on social media, is how to leverage all of the information and turn them into actionable policies. In this section, we introduce some popular research topics on social media.

Word-of-Mouth. The availability of data from fast changing e-commerce environment has created exciting opportunities for researchers to answer questions that have direct and critical impact on businesses and consumers. A number of researches have studied the network impact brought by social media. One question that has been addressed by many scholars in various contexts is whether the word-of-mouth effect has impact on sales and how the effect brings the changes to the sales figure. Traditionally, the impact of word-of-mouth was limited to the people close to each other and interacting with on a daily basis; but nowadays, word-of-mouth has endowed more power with social media and has shaken off the boundary completely. Therefore, the accessibility of online word-of-mouth information has made it possible to study online conversations in all aspects. The numerous experimental and empirical studies in the information systems and marketing literature have shown the importance of social media in this respect.

The most direct indicator of word-of-mouth – online ratings and reviews – has been well acknowledged to have impacts on sales, while the dynamics of impacts brought by the ratings/reviews are still debatable. Many of the researchers studying the impact of online conversations are limited to the data they obtain. Review information on certain products is widely available from online stores like Amazon.com or any sites carrying user generated content. People have found different stories regarding the factors that influence the sales. Chevalier and Mayzlin (2006) find that reviews with positive valence have positive impact on sales. By contrast, Liu (2006) states that this explanatory power of sales comes from the volume of WOM but not from its valence. Sonnier et al. (2011) demonstrate a significant effect of positive, negative, and neutral online communications on daily sales performance, and find that aggregating online comments over valence masks the effect on sales. Later on, many studies started to focus on the nature of products. Sun (2012) studies the impact of rating variance on product sales and finds that a higher variance is usually associated with a niche product and will lead to a higher demand if and only if the average rating is low. Gu et al. (2012) discover that external WOM has a significant impact on the sales of high-involvement products that consumers spend considerable time searching for before purchasing.

Even though a great amount of previous studies have been discussing the impact of social media or what is shaped by social network effects, there are also more and more researches talking about the processes or factors that drive those consequences, especially the drivers of the word-of-Mouth effects. Godes and Mayzlin (2004) show that a measure of the dispersion of online conversations across communities has explanatory power in a dynamic model of TV ratings. Moreover, Godes and Mayzlin (2009) initiate a field study to evaluate the effectiveness of the commonly used opinion leader designation and find that although opinion leadership is useful in identifying potentially effective spreaders of WOM among very loyal customers, it is less useful for the sample of less loyal customers. Berger and Schwartz (2011) examine how product characteristics shape immediate and ongoing WOM from a perspective of product characteristics and physiological factors, and suggest that rather than being driven by motivation, WOM, particularly ongoing WOM, is driven by accessibility. People tend to talk about whatever comes to their minds, regardless of how mundane it may be. Moe and Schweidel (2012) find significant heterogeneity with respect to consumers' desire to post in high-consensus versus high-variance environments (Godes and Silva 2012).

Although word-of-mouth is the most direct way for us to see the impact of social media on sales, there are other innovative ways people have developed based on the digital structure of online networks in order to push for more sales. Jing and Xie (2011) argue that a "Group Buying" strategy, under which consumers enjoy a discounted group price if they are willing and able to achieve a required group size and coordinate their transaction time, actually boosts sales for sellers. Oestreicher-Singer and Sundararajan (2012) find that co-purchase relationship leads to higher demand. However, the tremendously increasing comments and reviews over time also suggest that too much information may lead to customer information overload and thus reviews may exhibit a declining pattern over time (Godes and Silva 2012). Gopinath et al. (2013) examine how local geographic markets affect the impact of blogs on movie performances.

A recent study by Lee et al. (2013) examines how users generate online reviews in the presence of social influence. Using a data set on movie box office and user online social networking behaviors, they find the existence of herding behaviors for popular movies, that is, a user is more likely to give a higher rating if the prior aggregated user rating is higher. In doing so, they have carefully removed the confounding effect stemming from homophily (that users may have similar preferences and tastes). Ho et al. (2013) discover that a user is more likely to provide a rating for a product or service if the gap between her expected and realized perceptions of quality is larger, and thus explained the commonly observed U-shape distribution of online ratings.

Information Diffusion and Network Dynamics. Emotion can often spread rapidly through large crowds, and thus, individual ideas or behaviors are very much influenced by someone else. Social networking sites like Twitter, Facebook, Pinterest, and etc. have been used more and more often to broadcast people's opinion and ideas with the world, through ways such as "like", "re-tweeting", or "re-pinning". Understanding and predicting the outcomes of those complex interactions is, therefore, a topic of great interest.

In the business world, people care about social contagion because they want to understand how their products can go viral. Diffusion models can be useful for managers to forecast demand and to plan a variety of pre-launch and post-launch strategic decisions such as the optimal level of product sampling, optimal pricing, and optimal timing of successive generations of a product (Hosanagar et al. 2010) There is an emerging stream of literature in which the impact of diffusion on individuals' adoption is assessed. For instance, Iyengar et al. (2011) find that the amount of contagion is moderated by both the recipients' perception of their opinion leadership and the sources' volume of product usage. Because of the importance of opinion leaders, identifying opinion leaders in a community is an essential job for the purpose of marketing. Trusov et al. (2010) develop an approach to determine the opinion leaders in a social network by using the longitudinal records of members' log-in activity. Besides general information diffusion, various questions have been asked such as the factors in determining peer influence and contagion. Aral and Walker (2011) design a field experiment to examine how firms can create word-of-mouth peer influence and social contagion by designing viral features into their products and marketing campaigns. They find that active-personalized viral messages are more effective in encouraging adoption per message, while passive-broadcast messaging is used more often, generating more total peer adoption in the network.

On the other hand, it is also important to understand the dynamics of digital content diffusion structured through a network. Susarla et al. (2012) study the networked structure of social influence through a Youtube data set and find the evidence that interactions in networked structure can guide opinion formation and direct product discovery. Zeng and Wei (2013) observe from a Flicker dataset that around the time of the formation of a social tie, members of dyads began to upload more similar photos than they did before that time. After a social tie was formed, this similarity evolved in different ways in different subgroups of dyads. Fang et al. (2013) take a data-driven approach to examine user adoption in a social network. They identify key factors that underlie a social entity's adoption decision and operationalize them using Naïve Bayesian learning for accurate prediction.

Internal Social Network. There are more and more organizations using social media tools as part of their internal communications, with the aim to help engage employees on key topics such as performance, collaboration, culture and values. As the popularity of social media grows tremendously, companies are getting more and more interests in understanding the value of internal social network. IBM, for instance, has been doing research to see how social networking can impact enterprise networks, and found that many companies are encouraging employees to use their social networks so they can connect with other employees, help people socialize when they take a break, or even help contribute to other work-related issues (DiMicco et al. 2008). A few studies have been conducted in understanding the true impact of internal social network on employee behaviors and internal network structure. For example, Wu (2013) studies the change in employees' network positions before and after the introduction of a social networking tool, and finds that information-rich networks enabled through the use of social media can drive both work performance and job security, although there is a trade-off between engaging in social communication and gathering diverse information. Sasidharan et al. (2012) examine the roles that the social network structures play in influencing the ERP post-implementation success. They find that centralized structures inhibit implementation success, while employees with high in-degree and betweenness centrality revealed high task impact and information quality.

Mobile. Social networking once began in the online space, but it has rapidly spread to mobile platforms. More people are using smartphones and tablets to access social media so that consumption of mobile data usage is being driven by mobile social networking. According to Nielson's 2012 social media report, 46% of social media users use their smartphones to access social media, and the time spent on mobile apps and mobile web account for 63% of the year-over-year growth in overall time spent on social media. Besides native mobile social networks, such as Foursquare, Instagram, and Path that are built around mobile functionality, almost all the social networking websites have created mobile apps to give their users instant and real-time access from their devices. Mobile social media makes use of its GPS location feature and time flexibility to influence an upward trend in the popularity and accessibility of e-commerce, or offline spending. So far, the line between mobile and web is becoming very blurred. Given the current trend of instant communication, location-based services, and augmented reality, everything requires mobile devices and technology. It is foreseeable that mobile and web-based social networking systems will work symbiotically to spread content, increase accessibility and connect users from wherever they are. In their recent research, Ghose et al. (2012) find that ranking effects and the benefit of browsing for geographically close matches are both higher on mobile phones compared to computer-based internet users. Ghose and Han (2011) analyze whether there is a positive or negative interdependence between the mobile-phone-based content generation behavior and content usage behavior.

Healthcare. Information technology has yet to revolutionize healthcare the way it has changed our daily lives. Nowadays a cadre of social platforms, such as PainentsLikeMe and Patient Fusion, aim to disrupt the way people share information about personal health, physicians, and treatments. The intersection of healthcare and social media represents a promising space for future information systems research (Fichman et al. 2011). Although social media community has grown tremendously in every industry, there are still very few researches that systematically study the social influence or peer-to-peer sharing among patients in online healthcare communities. As social media is utilized more and more by hospitals and medical professionals as a means to convey general health information, it is of interest to see the network impact or diffusion model to be applied in the healthcare community. There exist a few studies conducted from different perspectives to confirm the social impact on healthcare. For example, Miller and Tucker (2013) collect data on whether the hospital is actively managing its social media presence by customizing their Facebook page. They find that when an organization actively manages its social media presence, it predominantly succeeds in increasing user-generated content as a function of number of their employees, rather than the number of clients. Angst et al. (2010) use a social contagion lens to study the dynamic, temporal process of the diffusion of electronic medical records in the population of U.S. hospitals and find that diffusion can be accelerated if specific attention is given to increasing social contagion effects in the case of hospitals' likelihood of adopting electronic medical records when subjected to the influence of prior adopters. Yan and Tan (2013) analyze an online healthcare community and find the evidence that social support exchanged among patients has a positive impact on improving their health conditions. Yan et al. (2013) study how patients identify and select proper users to communicate in the healthcare community to reduce their information overload and to receive expected social gains from the online activities.

Crowdsourcing. In general, all different kinds of user-generated content can be considered as crowdsourced content as these content provides useful information and wisdoms to some extent. A few researches have been conducted to access the value of mining consumer reviews and ratings. As user content generation platforms have been widely developed, many organizations are now using the platform to outsource their ideation efforts in facilitating their innovation process. This "crowdsourcing" approach has been receiving lots of attention ever since, and several organizations have implemented online crowdsourcing systems that gather ideas for new products and services from the "crowd" of nonexperts. One prominent example of the crowdsourcing sites is Dell's IdeaStrorm. Bayus (2013) collects two years' data from Dell's IdeaStorm community to study the nature of a crowdsourced idea generation process over time. They find that serial ideators are more likely than consumers with only one idea to generate an idea the organization finds valuable enough to implement but are unlikely to repeat their early success once some of their ideas are implemented.

Kleemann et al. (2008) characterize different types of crowdsourcing given different levels of knowledge companies expect the consumers to contribute. For example, companies may ask for comments and suggestions on product development and configuration, or advices to develop a whole new product, or any new information or documentation, and etc. Among those levels, the most common one is what nowadays adopted by almost all online vendors – product ratings and reviews. For instance, Archak et al. (2011) capture textual data of Amazon reviews, and show how it can be used for learning consumers' relative preferences for different product features and for predictive modeling of future changes in sales.

A recent study by Hildebrand et al. (2013) points out that social media could be bad for crowdsourcing. The negative influence of consumer feedback on their satisfaction could result from increased decision uncertainty and perceived process complexity.

Crowdfunding. Crowdfunding can be viewed as an element of crowdsourcing. In particular, it deals with financial contribution to an entity, instead of contributing ideas to a company. There is an emerging stream of research that has studied the new micro-financing method of peer-to-peer lending, or crowd-funding. Crowd-funding has been defined as "the financing of a project of a venture by a group of individuals instead of professional parties" (Schwienbacher and Larralde 2010). Basically, people who want to borrow are matched with those that want to lend on the web-based microlending platform. This out-of-the-box, non-traditional investing option has become a very compelling alternative to traditional investing, and quite a few crowd-funding platforms have arisen over the last few years. The earliest successful example is Sellaband.com, a Dutch-based marketplace in which musical artists raise funds to produce and sell albums (Burtch et al. 2013). Prosper.com, founded in 2006, has grown quickly and become a well-known crowd-funding site in the United States. As the big competitor of Prosper, Lending Club has teamed up with Google and has grown really fast to compete with institutional lenders. A few studies have done based on the data obtained from these prominent sites. Lin et al. (2013) examine Prosper.com in an attempt to identify the network effects in assisting individual's crowd-funding decision making. They find that a borrower with higher social capital is more likely to gain trusts from the lender and thus has higher probability of raising fund successfully. However, Zhang and Liu (2012) find, counter to intuition, that herding effect happens among lenders when the borrower exhibits signals of low quality. The possible explanation to this is that lenders assume there is some private information about the borrower that they are not privy to. Burtch (2011) also provides the evidence in his research, stating that herding behavior is a network externality in a crowd-funding market. In additional, Burtch et al. (2013) empirically study social influence in a crowd-funded marketplace for online journalism projects. They examine both the antecedents and consequences of the contribution process and find that there is a clear linkage between marketing effort and the success of crowd-funded projects.

3. Descriptive Research

Descriptive Research aims to uncover the underlying mechanisms in the problem under study. The emphasis of descriptive research is on identifying and accurately describing causal relationships. In general, we can categorize the methodologies into three areas: econometric or empirical, experimental, and analytical modeling. In the following, we describe each area and give some examples related to social media from existing literature.

3.1 Empirical Methods

Given that various forms of data exist in the context of social media, econometric methods are often the natural choice employed to derive insights. For empirical research, there are typically two major issues to consider: specification and identification. Model specification concerns the selection of the variables included in the model. Very often, we need to provide justifications, from the related economic and social theories, why certain variables or measures are included or excluded. Non-experimental researches, whether using reduced forms or structural models, are all subject to an identification problem. A model has to be properly specified so that different characteristics are identifiable and distinguishable given the data. That is, the model parameters for all the variables (or covariates) can be properly estimated.

It is often the case that the data is imperfect. The deficiencies in the data cause challenges in model identification. One example is the omitted variable problem resulting from the fact that certain information is unobservable. If not properly handled, omitted variables can introduce biased estimates of key model parameters, and hence lead to wrong, or at least inaccurate implications. Another consequence is the problem of endogeneity, that is, the error term is correlated with some independent variables. This will happen when an omitted variable is, for example, a decision variable which drives some independent variables. For cross sectional data, most identification strategies rest on identifying plausible instruments. When measuring the impact of on-line movie reviews with aggregated market level data, Chintagunta et al. (2010) use the exogenous variables from previous markets as instruments for user ratings in subsequent market, in order to account for the sequential release of movies across geographic markets. Their approach can be generalized in dealing with endogeneity concerns to other settings where products are sequentially rolled out.

The availability of panel data provides a better solution to the endogeneity problem. With panel data one can control product heterogeneity through fixed effect approach, or control unobserved error via random effect approach. Both of the two approaches serve the role of picking up common components of a certain variable. Gu et al. (2012) collect a panel data to investigate the role of internal and external WOM sources on retailer sales of high-involvement products. Besides adopting a fixed-effect model to control product heterogeneity and brand effect, they also use an instrumental variable approach to control the endogeneity between the WOM volume and product sales. The difference-in-difference strategy has also been used widely in the literature to control for common unobservable heterogeneity. Chevalier and Mayzlin (2006) adopt the difference-in- difference strategy to eliminate a possible website-specific fixed effect, when studying the effect of WOM on book sales across different online bookstores. Similarly Sun (2012) employs a difference-in-difference strategy to a data from both Amazon and Barnes & Noble to control for unobserved book characteristics that may influence both sales and ratings.

Social Effects. There are additional challenges in dealing with network data to identify endogenous social effects. This is commonly known as the reflection problem, first pointed out by Manski (1993). This problem arises when a researcher is observing the distribution of behavior and tries to infer where the average behavior in some group influences the behavior of the individuals that comprise the group. A linear specification where the individual behavior is linearly related to the group average will result in a tautology and the social influence is not identifiable. To resolve this problem, there are many approaches, for example, using a nonlinear relationship which of course needs to be justified by underlying theories. Another approach is to explicitly incorporate social structures and measures in the model. Instead of using group average, individuals in social relationships are modeled to contribute differently.

This actually leads to a natural integration of econometric and social networks analyses. Social networks analysis (Wasserman and Faust 1994) provides a comprehensive set of measures, for example, various centralities, structural equivalence, structural holes, cohesive subgroups, and etc. at both individual and network levels. As a social network is mapped with respect to a specific social relationship such as friendship, subscription, co-worker, and so on, the estimated effects can be better interpreted in the proper context.

In general, it is agreed that it is difficult, if not impossible, to cleanly identify social influences. There could exist many confounding effects, most notably, homophily and exogenous shocks. For example, if we observe two users who downloaded the same mobile app, we can say that maybe the second user was influenced socially by the first user, and hence followed. However, it could also be the case that both users had similar preferences or tastes, or they are homophilous. The third reason could be that they were exposed to the same exogenous shock, such as a marketing promotion. In order to tease out each cause, one has to exert effort to identify and make use of their unique characteristics which may be contextual. Bramoullé et al. (2009) suggest a way to identify the true effect of social influence by accounting for the homophily. A network specific unobservable is introduced to the individuals that belong to the same social network structure. Then, we use group means of the variables and subtract the variables from corresponding group means, to remove the effect stemming from similar preferences.

Latent Variable Models. There are numerous econometric models that can be applied to analyze social media data. A few of them are particularly suited, for example, a class of generalized latent variable models (Rabe-Hesketh and Skrondal 2004). While some variables are unobservable, they describe needed individual heterogeneity and also play very important roles in driving some social behaviors. Hence, it is especially useful to model them and, if possible, make an inference to uncover them. The latent space model, developed by Hoff et al. (2002), is a perfect example. In this model, each actor is assumed to have a position in a social space. The positions are unknown, so the social space is a latent one. However, the probability of a link between two actors depends on the distance between them in the social space. If we can observe how actors communicate, form or sever links, we can estimate their hidden social positions.

Similarly, the Hidden Markov Model (HMM) (Singh et al. 2011) is a useful model to understand latent drivers of social or economic behaviors. It consists of two Markov chains; one is observable while another is not. But these two chains are interrelated – the behavior observed in one chain is driven or influenced by the latent one. Once the model parameters are estimated, either by maximum likelihood or Bayesian, posterior analysis can be conducted to uncover the hidden chain. POMDP (partially observed Markov decision process) is a variant of HMM which allows some parts of the otherwise completely hidden chain to be observed. This model has also found applications in social media research (Yan and Tan 2013).

Recently, there is a new development in applying latent instrument variable to resolve potential endogeneity biases (Ebbes et al. 2009). This method takes the advantage of the distribution of data, and separates an endogenous variable into exogenous and endogenous parts respectively.

Dynamic Models. Snijders (2006) has proposed a dynamic model of a dichotomous social network based on Markov Chain Monte Carlo (MCMC), to study the formation and evolution of a social network. The continuous time Markov transition probabilities are modeled as a function of a set of covariates derived from individual and dyadic characteristics, for example, degree of ties, reciprocity, transitivity, and etc. With the observation of some snapshots of the network at some time points, the dynamic model can estimate the parameters associated with the covariates or factors which drive the network evolution.

Shriver et al. (2013) conduct an empirical study to confirm that users' online content generation activity is codetermined by their social ties. In another study by Zeng and Wei (2013), the interaction between social ties and user content generation is also examined. They find that members of dyads were more likely to upload similar photos when they formed a tie; however, the dyads of users with similar popularity began to upload less similar photos.

Structural Modeling. We have observed an increase in applying structural models in marketing and information systems research (Chintagunta et al. 2006). Structural models rely on economic or related theories to derive model specification. This approach is especially suited to capture the strategic behaviors of rational users or firms that maximize their respective objects such as utility or profit. A dynamic model allows describing behaviors such as forward-looking decision making. Structural models can resolve the potential endogeneity problem. Another advantage is its capability to conduct policy analyses and produce counterfactuals to answer some important "what-if" questions.

3.2 Experimental Methods

Field experiments have been used in social media research to exogenously vary the treatment across subjects. This helps to solve the potential endogeneity problem. Another common problem in empirical research is that sometimes what the data and/or model show is merely correlation rather than causality. It is therefore useful to carefully design experiments to identify causal relationships.

Aral and Walker (2011) design a field experiment to examine how firms can create word-of-mouth peer influence and social contagion by designing viral features into their products and marketing campaigns. Toubia and Stephen (2013) analyze and contrast two types of user utility: intrinsic and image-related, in motivating them to post content. They conducted a field experiment to address an issue that the number of followers on Twitter could be endogenous by exogenously adding followers to a set of users. Zhang and Zhu (2011) take advantage of a natural experiment to examine the causal relationship between group size and incentives to contribute in the setting of Chinese Wikipedia, the Chinese language version of an online encyclopedia that relies entirely on voluntary contributions. Claussen et al. (2013) make use of a rule change by Facebook to study the effects of rewarding user engagement.

3.3 Analytical Methods

The analytical models in social media research typically originate from microeconomic theory and/or operations research. Recently, there are applications of the methodologies in other disciplines (for example, physics) in social media, especially social networks, research. In the following, we discuss a few papers with different modeling approaches.

Game Theoretical Model. Mayzlin and Yoganarasimhan (2012) formulate game theoretical model to explain why a rational blogger may choose to link to another blog. It is assumed that bloggers differ by their abilities to break news and to find news in other blogs, respectively. By linking, a blog signals to the reader that it will be able to direct her to news in other blogs in the future. However, the tradeoff is that linking generates a positive signal on the rival's news-breaking ability. They show that linking will be in equilibrium when the heterogeneity on the ability to break news is low relative to the heterogeneity on the ability to find news in other blogs. Overall, it is found that the activity of linking enhances readers' learning. Dou et al. (2012) investigate how software firms should optimize the strength of network effects by adjusting the level of embedded social media features. This paper analyzes the right market seeding and pricing strategies in the presence of seeding disutility, with a focus on the interaction between building more social media features to increase the strength of network effects and seeding the market.

Operations Research Model. Dawande et al. (2012) formulate optimization problems to search for useful structures in social networks. They consider two problems: EGP (the elite group problem) and PP (the portal problem) derived, respectively, from the notions of influence and centrality. It is shown that for a variety of social networks, EGP is polynomially solvable whereas PP is strongly NP-hard. They find the solution of PP on several special networks: bi-cliques, balanced and full d-trees, paths, cycles, and cliques, and proposed a heuristic for general trees.

Ising Spin Glass Model. The Ising model originated in statistical physics as a means of modeling the interactions among electron spins to explain the emergence of magnetization. It is often used to demonstrate phase transition and critical phenomena. The Ising model has also been used to explore the emergence of critical dynamics such as bubbles and crashes in various economic settings. Recently, this model has been applied to examine herding in open source software development community (Oh and Jeon 2006). Less attention has been paid, however, to how Ising-like structures can be used to model the dynamics of opinion formation and product adoption. The simplicity of the Ising methodology can help articulate models that are analytically tractable and empirically verifiable. There exists a simple mapping in that an electron can be considered as an individual in a social network, and the spin which could be positive or negative represents the sentiment of this individual's opinion. The temperature in Ising model can be interpreted as the strength of social interaction, whereas the external magnetic field can be attributed to exogenous marketing effort. As such, some results in Ising model can be readily applied to social media. For example, the characteristics of the critical point where the consensus of opinions is formed can be obtained in relation with social interactions and exogenous shocks.

4. Predictive Analytics

The data generated by various online social media is of massive amounts, unstructured, and dynamic in nature, which poses great challenges to the so-called big data analytics (Chen et al. 2012). As an inter-discipline springing up in recent years, data mining is often regarded as one of the best solutions to cope with these challenges. It tries to discover useful or actionable rules from big data (Tan et al. 2005), by establishing heterogeneous (may be distributed) databases to store and manipulate data of various kinds using both SQL and noSQL techniques (Stonebraker 2010), and designing predictive algorithms suitable for parallel or distributed computing on platforms such as Hadoop (White 2012).

As a methodology for predictive analytics, data mining can also boost the descriptive analytics of social media. On one hand, data mining enables the objective characterization of social behaviors of an individual, a group, or an organization upon the rich online interactive data. This is particularly valuable for studying how, for example, sentiments or authorities of individuals, can influence the performances of online businesses or finance. On the other hand, the deployment of various data mining techniques on social media sites, such as recommender systems and community detection, provides precious opportunities to probe exciting new research topics. For instance, researchers are getting more and more interested in understanding the major factors that lead to the real success of social recommendation, or the underlying forces that drive the dynamic evolution of specific communities. In what follows, we briefly introduce the advances of social media mining research in some subfields most related to descriptive analytics.

Sentiment Analysis. As massive opinion data emerges, there is a pressing demand for sentiment analysis that could aggregate the overall public opinions and characterize the sentiment variations over a given period of time. This information allows for perceiving the brand reputation or the-word-of-mouth for some products, and predicting the box office revenues (Asur and Huberman 2010) or stock market (Bollen et al. 2011).

The lexicon-based method is one solution to sentiment analysis. With a comprehensive and authoritative emotional lexicon, this method is easy to apply and works well in some domains, like finance (Das and Chen 2007). Moreover, many experimental results show that the machine learning approach could also achieve very high accuracy (Pang et al. 2002), which promoted the combination of the two methods (Melville et al. 2009).

In the context of social media, more information such as emoticons, hashtags and network structures is available, which provides new thoughts for sentiment analysis. For example, Read (2005) shows that using training data labeled by emoticons can greatly reduce the dependency of domain, topic and time in the machine learning techniques. Zhao et al. (2012) propose an emoticon-based method for multi-sentiment analysis of Chinese tweets, which shows potential for abnormal event detection. Wang et al. (2011) use graph models based on hashtags co-occurrence relationship to perform sentiment analysis. Guerra et al. (2011) use the endorsement network to measure the bias of social media users toward a topic, and adopt a transfer learning approach to compute the polarity of posted contents.

Social Influence Computing. Social influence computing aims to find influential nodes such as authorities, opinion leaders, gatekeepers or mavens, which could activate other nodes in social networks as much as possible. Motivated by applications in viral marketing (Domingos and Richardson 2001), personalized recommendation (Song et al. 2007), feed ranking (Ienco et al. 2010), and the analysis of online social network (Weng et al. 2010), the study of influence propagation has caught tremendous attention in the last decades.

In the early research, Krackhardt (1992) and Granovetter (1973) reveal the impact of the strength of strong and weak ties to social works, respectively. Later, some topological properties and network metrics like the number of followers, PageRank value and the number of retweets were used to rank and find the most influential users (Kwak et al. 2010). Based on these, Bakshy et al. (2011) propose diffusion tree to quantify the user influences. Note that the rankings by different influence measures are often inconsistent, and PageRank usually fails to produce high quality target sets dispersing across the whole network (Pandit et al. 2012).

The methods above only take the network topology into account. Recently, Domingos and Richardson (2001) introduce propagation factors to identify influential users in a marketing area as a learning problem. Subsequently, Kempe et al. (2003) formulate the problem as a discrete optimization problem and prove that it is NP-hard if the influence propagation is based on the independent cascade model or the linear threshold model. Several follow-up studies propose different heuristics to speed up the computation including shortest-path (Kimura and Saito 2006), cost-effective lazy forward (Leskovec et al. 2007), maximum influence path (Chen et al. 2010), LDAG (Chen et al. 2010), and SPIN (Narayanam and Narahari 2010).

Despite many algorithms proposed for influence modeling, few of them can work for very large social networks. Several studies that extract influence cascade model parameters from real datasets to generate influence graphs (Anagnostopoulos et al. 2008, Tang et al. 2009, Saito et al. 2010, Goyal et al. 2010) can be regarded as the initial works along this line. Zhao et al. (2013) point out that the whole network topology can usually be unavailable and hence introduce an approach based on graph sampling.

Recommender Systems. Recommender systems suggest a few items from many possible choices to users by learning their profiles. In the last two decades, many different types of recommender systems have been developed like the collaborative-filtering based (Schafer et al. 2007), content based (Debnath et al. 2008), and hybrid (Symeonidis et al. 2008) systems, most of which exploit user-item rating matrix to generate recommendations.

Recently the development of recommender systems has been advanced by the rapid growth of social media. In addition to user ratings, social media further provides social signals and rich contexts like friend networks, social tags, trust, and locations. Recommender systems research has advanced into how to use heterogeneous information to improve recommendation accuracy. For instance, Jamali and Ester (2010) incorporate the mechanism of trust propagation into the matrix factorization model, and led to a substantial increase in recommendation accuracy, especially for cold-start users. The work by Konstas et al. (2009) takes into account both the social annotations and friendships established among users, items and tags, and adopts the generic framework of random walk with restarts to provide a more natural and efficient way to represent social networks. Ye et al. (2010) exploit the social and geographical characteristics of users and locations to realize location-based recommendation services.

Besides the real item recommendation, link or key node recommendation on social networks also attracts much attention. For example, Saez-Trumper et al. (2012) combine temporal attributes of nodes and edges of network with a PageRank based algorithm to find the trendsetters for a given topic. Backstrom and Leskovec (2011) study the problem of inferring interactions among existing members that are likely to occur in the near future. The work by Dong et al. (2012) also studies the problem of link recommendation across heterogeneous networks where a ranking factor graph model is used.

Online recommender systems are vulnerable to malicious users who inject biased ratings to manipulate online product recommendations, i.e., the "shilling attacks" or "profile injection attacks" (Burke et al. 2005, Lam and Riedl 2004, O'Mahony et al. 2004). The related studies mainly focus on the three subareas: the shilling attack generation models (Mobasher et al. 2007), the detection metrics (Burke et al. 2006, Williams 2006), and the classification methods (Mehta and Nejdl 2009). Considering the shortage of manually labeled data in shilling attack classification, Wu et al. (2012) provide a semi-supervised method to detect shilling attacks, and demonstrate its effectiveness on enhancing the product recommendations in Amazon.

Community Detection. People within an online social community usually hold similar opinions, have common interests, and exert more influences to each other. So discovering such social communities can enhance the analysis of opinions, the understanding of influences among users, and the performance of intelligent recommender systems. The community detection problem is initially formulated as finding a good K-way crisp partition (Kernighan and Lin 1970, Slater 2008, Newman 2006, Newman 2004, Flake et al. 2002, Girvan and Newman 2002). Since networks keep evolving with new events and individuals, much attention is then paid to community evolution to reveal community structures at sequential timestamps (Yang et al. 2009, Lin et al. 2009, Tang and Liu 2012). Furthermore, people notice that communities are nested in nature, and thus have great interests in finding overlapping communities (Palla et al. 2005, Lancichinetti et al. 2009, Zhang et al. 2007, Nepusz et al. 2007). We here briefly review basic methods in these three areas.

The crisp community detection methods can be classified into two categories, in terms of whether or not a global optimization objective is available. The methods with global optimization objectives typically consider the global topology of a network, and aim to optimize the objective function defined over a network partition. The differences between these methods ultimately come down to the global criteria and the algorithmic heuristics. For instance, the cut criterion and its variations have been used by Kernighan-Lin algorithm (Kernighan and Lin 1970) and spectral methods (Slater 2008), and the modularity (Q function) proposed by Newman has been used in a great deal of algorithms such as FastNewman (Newman 2006, Newman 2004). The methods without global optimization objectives typically employ a bottom-up strategy to find communities. The maximum flow community algorithm (Flake et al. 2002) and the GN algorithm (Girvan and Newman 2002) are just two examples. In recent years, with the emergence of super-large social networks from Facebook, Twitter, and etc., community extraction rather than partition becomes a very hot topic in this area (Zhao et al. 2011, Wu et al. 2013). The efficiency and evaluation issues are the two main concerns.

The studies on community evolution can be also divided into two types. One employs a straightforward two-step method, where static analysis is applied firstly to the snapshots of the network, and then community evolutions are introduced to interpret the change of communities over time. The other one attempts to unify the processes of community detection and evolution. Dynamic stochastic block model is widely used for this purpose, aided by some optimization methods such as Gibbs sampling (Yang et al. 2009) and Expectation-Maximization (Lin et al. 2009). Meanwhile, community evolution on multi-mode networks, consisting of different types of nodes and relations, has also attracted some attention (Tang and Liu 2012).

In real life, a person usually has connections to multiple social groups. Overlapping community detection thus gains increasing interests, which aims to discovery partitions that are not necessarily disjoint. The famous Clique Percolation Method (CPM) (Palla et al. 2005) is based on the concept of k-clique, with CFinder (http://www.cfinder.org/) being a successful implementation. Many other methods have been presented including the link partitioning method (Ahn et al. 2010), the local expansion and optimization methods (Lancichinetti et al. 2009, Shen et al. 2009), the fuzzy clustering based methods (Zhang et al. 2007, Nepusz et al. 2007), and etc.

5. Conclusions

This paper provides a concise, and hence inherently incomplete, review of research on social media. We have discussed extant prior studies which cover a wide spectrum of topics, ranging from word of mouth to the applications of social media in healthcare and mobile platforms. A summary of data driven approaches in the form of data and text mining has also been provided. We have also reviewed quantitative research methods and discussed the challenges researchers may face when applying these methodologies.

There are many opportunities for researchers to advance our understanding of social media and its effects on businesses. For example, existing analytical studies on diffusion have often used mean-field approximation which ignores the underlying network topology. Given the importance of network structure, it is imperative to develop a framework which explicitly accounts for it. It is also interesting to identify and investigate novel business models and social phenomena emerging in the online platforms such as microblog sites. Many of them may become test beds, unavailable before, for researchers to re-examine existing social and economic theories.

References

[1] Ahn, Y.Y., Bagrow, J.P. & Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. Nature, 466(7307): 761-764

[2] Anagnostopoulos, A., Kumar, R. & Mahdian, M. (2008). Influence and correlation in social networks. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008: 7-15

[3] Angst, C.M., Agarwal, R., Sambamurthy, V. & Kelley, K. (2010). Social contagion and information technology diffusion: the adoption of electronic medical records in U.S. hospitals. Management Science, 56(8): 1219-1241

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

[5] Aral, S., Dellarocas, C. & Godes, D. (2013). Social media and business transformation: a framework for research. Information Systems Research, 24(1): 3-13

[6] Archak, N., Ghose, A. & Ipeirotis, P.G. (2011). Deriving the pricing power of product features by mining consumer reviews. Management Science, 57(8): 1485-1509

[7] Asur, S. & Huberman, B.A. (2010). Predicting the future with social media. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 1: 492-499

[8] Backstrom, L. & Leskovec, J. (2011). Supervised random walks: predicting and recommending links in social networks. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, 2011: 635-644

[9] Bakshy, E., Hofman, J.M., Mason, W.A. & Watts, D.J. (2011). Everyone's an influencer: quantifying influence on twitter. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, 2011: 65-74

[10] Bayus, B.L. (2013). Crowdsourcing new product ideas over time: an analysis of the dell ideastorm community. Management Science, 59(1): 226-244

[11] Berger, J. & Schwartz, E.M. (2011). What drives immediate and ongoing word of mouth? Journal of Marketing Research, 48(5): 869-880.

[12] Bollen, J., Mao, H. & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1): 1-8

[13] Bramoullé, Y., Djebbari, H. & Fortin, B. (2009). Identification of peer effects through social networks. Journal of Econometrics. 150(1): 41-55

[14] Burke, R., Mobasher, B., Williams, C. & Bhaumik, R. (2006). Classification features for attack detection in collaborative recommendation systems. Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining, 2006: 542-547

[15] Burke, R., Mobasher, B., Zabicki, R. & Bhaumik, R. (2005). Identifying attack models for secure recommendation. Beyond Personalization: Workshop on Next Generation Recommender Systems

[16] Burtch, G. (2011). Herding behavior as a network externality. ICIS 2011 Proceedings, 2011: 28

[17] Burtch, G., Ghose, A. & Wattal, S. (2013). An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets. Information Systems Research, 24(3): 499-519

[18] Calais Guerra, P.H., Veloso, A. & Meira, J.W. (2011). From bias to opinion: a transfer-learning approach to real-time sentiment analysis. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011:150-158

[19] Chen, H., Chiang, R.H.L. & Storey, V.C. (2012). Business intelligence and analytics: from big data to big impact. MIS Quarterly, 36(4): 1165-1188

[20] Chen, W., Wang, C. & Wang, Y. (2010). Scalable influence maximization for prevalent viral marketing in large-scale social networks. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010: 1029-1038

[21] Chen, W., Yuan, Y. & Zhang, L. (2010). Scalable influence maximization in social networks under the linear threshold model. Proceedings of the 10th IEEE International Conference on Data Mining, 2010: 88-97

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

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

[24] Chintagunta, P., Erdem, T., Rossi, P.E. & Wedel, M. (2006). Structural modeling in marketing: review and assessment. Marketing Science, 25(6) : 604-616

[25] Claussen, J., Kretschmer, T. & Mayrhofer, P. (2013). The effects of rewarding user engagement: the case of facebook apps. Information Systems Research, 24(1): 186-200

[26] Das, S.R. & Chen M.Y. (2007). Yahoo! for Amazon: sentiment extraction from small talk on the web. Management Science, 53: 1375-1388

[27] Dawande, M., Mookerjee, V., Sriskandarajah, C. & Zhu, Y. (2012). Structural search and optimization in social networks. INFORMS Journal on Computing, 24(4): 611-623

[28] Debnath, S., Ganguly, N. & Mitra, P. (2008). Feature weighting in content based recommendation system using social network analysis. Proceedings of the 17th International Conference on World Wide Web, 2008: 1041-1042

[29] DiMicco, J., Millen, D.R., Geyer, W., Dugan, C., Brownholtz, B. & Muller, M. (2008). Motivations for social networking at work. Proceedings of the 2008 ACM

Conference On Computer Supported Cooperative Work, 2008: 711-720

[30] Domingos, P. & Richardson, M. (2001). Mining the network value of customers. Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001: 57-66

[31] Dong, Y., Tang, J., Wu, S. & Tian, J. (2012). Link prediction and recommendation across heterogeneous social networks. Proceedings of the 12th IEEE International Conference on Data Mining, 2012: 181-190

[32] Dou, Y., Niculescu, M.F. & Wu., D.J. (2013). Engineering optimal network effects via social media features and seeding in markets for digital goods and services. Information Systems Research, 24(1): 164-185

[33] Ebbes, P., Wedel, M. & Böckenholt, U. (2009). Frugal IV alternatives to identify the parameter for an endogenous regressor. Journal of Applied Econometrics, 24(3): 446-468

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

[35] Fichman, R.G., Kohli, R. & Krishnan, R. (2011). Editorial Overview — The Role of Information Systems in Healthcare: Current Research and Future Trends. Information Systems Research, 22(3): 419-428

[36] Flake, G.W., Lawrence, S., Giles, C.L. & Coetzee, F.M. (2002). Self-organization and identification of web communities. Computer, 35(3): 66-70

[37] Ghose, A., Goldfarb, A. & Han, S.P. (2012). How is the mobile internet different? search costs and local activities. Information Systems Research, forthcoming

[38] Ghose, A. & Han, S.P. (2011). An empirical analysis of user content generation and usage behavior on the mobile internet. Management Science, 57(9): 1671-1691

[39] Girvan, M. & Newman, M.E. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99(12): 7821-7826

[40] Godes, D. & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. Marketing Science, 23(4): 545-560

[41] Godes, D. & Mayzlin, D. (2009). Firm-created word-of-mouth communication: evidence from a field test. Marketing Science, 28(4): 721-739

[42] Godes, D. & Silva, J.C. (2012). Sequential and temporal dynamics of online opinion. Marketing Science, 31(3): 448-473

[43] Goh, K.Y., Heng, C.S. & Lin, Z. (2013). Social media brand community and consumer behavior: quantifying the relative impact of user- and marketer-generated content. Information Systems Research, 24(1): 88-107

[44] Gopinath, S., Chintagunta, P.K. & Venkataraman, S. (2013). Blogs, advertising, and local-market movie box office performance. Management Science, forthcoming

[45] Goyal, A., Bonchi, F. & Lakshmanan, L. (2010). Learning influence probabilities in social networks. Proceedings of the 3rd ACM International Conference on Web

Search and Data Mining, 2010: 241-250

[46] Granovetter, M.S. (1973). The strength of weak ties. American Journal of Sociology, 1973: 1360-1380

[47] Gu, B., Park, J. & Konana, P. (2012). Research note — the impact of external word-of-mouth sources on retailer sales of high-involvement products. Information Systems Research, 23(1): 182-196

[48] Hildebrand, C., Häubl, G., Herrmann, A. & Landwehr, J.R. (2013). When social media can be bad for you: community feedback stifles consumer creativity and reduces satisfaction with self-designed products. Information Systems Research, 24(1): 14-29

[49] Ho, C., Wu, J. & Tan, Y. (2013). Disconfirmation effect on online opinion expression. Working paper

[50] Hoff, P.D., Raftery, A.E. & Handcock, M.S. (2002). Latent space approaches to social network analysis. Journal of the American Statistical Association, 97(460): 1090-1098

[51] Hosanagar, K., Han, P. & Tan, Y. (2010). Diffusion models for peer-to-peer (p2p) media distribution: on the impact of decentralized, constrained supply. Information Systems Research, 21(2): 271-287

[52] Ienco, D., Bonchi, F. & Castillo, C. (2010). The meme ranking problem: Maximizing microblogging virality. IEEE International Conference on Data Mining Workshops, 2010: 328-335

[53] Iyengar, R., Van den Bulte, C. & Valente, T.W. (2011). Opinion leadership and social contagion in new product diffusion. Marketing Science, 30(2): 195-212

[54] Jamali, M. & Ester, M. (2010). A matrix factorization technique with trust propagation for recommendation in social networks. Proceedings of the 4th ACM Conference on Recommender Systems, 2010: 135-142

[55] Jing, X. & Xie, J. (2011). Group buying: a new mechanism for selling through social interactions. Management Science, 57(8): 1354-1372

[56] Kempe, D., Kleinberg, J. & Tardos, É. (2003). Maximizing the spread of influence through a social network. Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003:137-146

[57] Kernighan B.W. & Lin S. (1970). An efficient heuristic procedure for partitioning graphs. Bell Systems Technical Journal, 49: 291-307

[58] Kimura, M. & Saito, K. (2006). Tractable models for information diffusion in social networks. Knowledge Discovery in Databases: PKDD 2006. Springer Berlin Heidelberg, 2006: 259-271

[59] Kleemann, F., Voß, G.G. & Rieder, K. (2008). Un(der)paid Innovators: the commercial utilization of consumer work through crowdsourcing. Science, Technology & Innovation Studies, 4(1): 5-26

[60] Konstas, I., Stathopoulos, V. & Jose, J.M. (2009). On social networks and collaborative recommendation. Proceedings of the 32th ACM SIGIR International Conference on Research and Development in Information Retrieval, 2009: 195-202

[61] Krackhardt, D. (1992). The strength of strong ties: The importance of philos in organizations. Networks and Organizations: Structure, Form, and Action, 1992: 216-239

[62] Kumar, V., Bhaskaran, V., Mirchandani, R. & Shah, M. (2013). Practice prize winner–creating a measurable social media marketing strategy: increasing the value and ROI of intangibles and tangibles for hokey pokey. Marketing Science, 32(2): 194-212

[63] Kwak, H., Lee, C., Park, H. & Moon, S. (2010). What is twitter, a social network or a news media? Proceedings of the 19th ACM International Conference on World Wide Web, 2010: 591-600

[64] Lam, S. & Riedl, J. (2004). Shilling recommender systems for fun and profit. Proceedings of the 13th International Conference on World Wide Web, 2004: 309-402

[65] Lancichinetti, A., Fortunato, S. & Kertész, J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. New Journal of Physics, 11(3): 033015

[66] Lee, M., Kim, M. & Peng, W. (2013). Consumer reviews: reviewer avatar facial expression and review valence. Internet Research, 23(2): 116-132

[67] Leskovec, J., Krause, A., Guestrin, C. & Faloutsos, C. (2007). Cost-effective outbreak detection in networks. Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2007: 420-429

[68] Lin, M., Prabhala, N.R. & Viswanathan, S. (2013). Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. Management Science, 59(1): 17-35

[69] Lin, Y.R., Chi, Y., Zhu, S., Sundaram, H. & Tseng, B.L. (2009). Analyzing communities and their evolutions in dynamic social networks. ACM Transactions on Knowledge Discovery from Data, 3(2): 8

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

[71] Luo, X., Zhang, J. & Duan, W. (2013). Social media and firm equity value. Information Systems Research, 24(1): 146-163

[72] Manski, C.F. (1993). Identification of endogenous social effects: the reflection problem. The Review of Economic Studies, 60(3): 531-542.

[73] Mayzlin, D. & Yoganarasimhan, H. (2012). Link to success: how blogs build an audience by promoting rivals. Management Science 58(9): 1651-1668

[74] Mehta, B. & Nejdl, W. (2009). Unsupervised strategies for shilling detection and robust collaborative filtering. User Modeling and User-Adapted Interaction, 19(1-2): 65-97

[75] Melville, P., Gryc, W. & Lawrence, R.D. (2009). Sentiment analysis of blogs by combining lexical knowledge with text classification. Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009: 1275-1284

[76] Miller, A.R. & Tucker, C. (2013). Active

social media management: the case of health care. Information Systems Research, 24(1): 52-70

[77] Mobasher, B., Burke, R., Bhaumik, R. & Williams, C. (2007). Toward trustworthy recommender systems: an analysis of attack models and algorithm robustness. ACM Transactions on Internet Technology, 7(4): 1-40

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

[79] Narayanam, R. & Narahari, Y. (2010). A shapley value-based approach to discover influential nodes in social networks. IEEE Transactions on Automation Science and Engineering, 2010(99): 1-18

[80] Nepusz, T., Petróczi, A., Négyessy, L. & Bazsó, F. (2008). Fuzzy communities and the concept of bridgeness in complex networks. Physical Review E, 77(1): 016107.

[81] Newman, M.E. (2004). Fast algorithm for detecting community structure in networks. Physical review E, 69(6): 066133

[82] Newman, M.E. (2006). Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23): 8577-8582

[83] O'Mahony, M., Hurley, N., Kushmerick, N. & Silvestre, G. (2004). Collaborative recommendation: a robustness analysis. ACM Transactons on Internet Technology, 4(4): 344-377

[84] Oestreicher-Singer, G. & Sundararajan, A. (2012). The visible hand? demand effects of recommendation networks in electronic markets. Management Science, 58(11): 1963-1981

[85] Oh, W. & S. Jeon. (2007). Membership herding and network stability in the open source community: the ising perspective. Management Science, 53(7): 1086-1101

[86] Palla, G., Derényi, I., Farkas, I. & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. Nature, 435(7043): 814-818

[87] Pandit, S., Yang, Y. & Chawla, N.V. (2012). Maximizing information spread through influence structures in social networks. Proceedings of the 12th IEEE International Conference on Data Mining Workshops, 2012: 258-265

[88] Pang, B., Lee, L. & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. Conference on Empirical methods in natural language processing, 10: 79-86

[89] Rabe-Hesketh, S. & Skrondal, A. (2004). Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. Boca Raton: CRC

[90] Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. Proceedings of the ACL Student Research Workshop, 2005:43-48

[91] Rishika, R., Kumar, A., Janakiraman, R. & Bezawada, R. (2013). The effect of customers' social media participation on customer visit frequency and profitability: an empirical investigation. Information Systems Research, 24(1): 108-127

[92] Saez-Trumper, D., Comarela, G., Almeida, V., Baeza-Yates, R. & Benevenuto, F. (2012). Finding trendsetters in information networks. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012: 1014-1022

[93] Saito, K., Kimura, M., Ohara, K. & Motoda, H. (2010). Selecting information diffusion models over social networks for behavioral analysis. Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, 2010: 180-195.

[94] Sasidharan, S., Santhanam, R., Brass, D.J. & Sambamurthy, V. (2012). The effects of social network structure on enterprise systems success: a longitudinal multilevel analysis. Information Systems Research, 23(3): 658-678

[95] Schafer, J.B., Frankowski, D., Herlocker, J. & Sen, S. (2007). Collaborative filtering recommender systems. Springer Berlin Heidelberg, 2007: 291-324

[96] Schwienbacher, A. & Larralde, B. (2010). Crowdfunding of small entrepreneurial ventures. Handbook of entrepreneurial finance, forthcoming

[97] Shriver, S.K., Nair, H.S. & Hofstetter, R. (2013). Social ties and user-generated content: evidence from an online social network. Management Science, 59(6): 1425-1443

[98] Singh, P.V., Tan, Y. & Youn, N. (2011). A Hidden Markov Model of Developer Learning Dynamics in Open Source Software Projects. Information Systems Research, 22(4): 790-807

[99] Slater, P.B. (2008). Established clustering procedures for network analysis. arXiv preprint arXiv:0806.4168

[100] Snijders, T.A.B. (2006). Statistical methods for network dynamics. Proceedings of the XLIII Scientific Meeting, 2006 (1994): 281-296

[101] Song, X., Chi, Y., Hino, K. & Tseng, B.L. (2007). Information flow modeling based on diffusion rate for prediction and ranking. Proceedings of the 16th International Conference on World Wide Web, 2007: 191-200

[102] Sonnier, G.P., McAlister, L. & Rutz, O.J. (2011) A dynamic model of the effect of online communications on firm sales. Marketing Science, 30(4): 702-716

[103] Stonebraker, M. (2010). SQL databases v. NoSQL databases. Communications of the ACM, 53(4): 10-11

[104] Sun, M. (2012). How does the variance of product ratings matter? Management Science, 58(4): 696-707

[105] Susarla, A., Oh, J.H. & Tan, Y. (2012). Social networks and the diffusion of user-generated content: evidence from youtube. Information Systems Research, 23(1): 23-41

[106] Symeonidis, P., Nanopoulos, A. & Manolopoulos, Y. (2008). Providing justifications in recommender systems. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 38(6): 1262-1272

[107] Tan, P.N., Steinbach, M. & Kumar, V. (2005). Introduction to Data Mining. Addison-Wesley

[108] Tang, J., Sun, J., Wang, C. & Yang, Z. (2009). Social influence analysis in large-scale networks. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009: 807-816

[109] Tang, L., Liu, H. & Zhang, J. (2012). Identifying evolving groups in dynamic multimode networks. IEEE Transactions on Knowledge and Data Engineering, 24(1): 72-85

[110] Toubia, O. & Stephen, A.T. (2013). Intrinsic vs. Image-related utility in social media: why do people contribute content to twitter? Marketing Science, 32(3): 368-392

[111] Trusov, M., Bodapati, A.V. & Bucklin, R.E. (2010). Determining influential users in internet social networks. Journal of Marketing Research, 47(4): 643-658

[112] Wang, X., Wei, F., Liu, X., Zhou, M. & Zhang, M. (2011). Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. Proceedings of the 20th ACM International Conference on Information and Knowledge Management, 1031-1040

[113] Wasserman, S. & Faust, K. (1994). Social network analysis: methods and applications. Cambridge University Press

[114] Weng, J., Lim, E.P., Jiang J. & He, Q. (2010). Twitterrank: finding topic-sensitive influential twitterers. Proceedings of the 3rd ACM International Conference on Web Search and Data Mining, 2010: 261-270

[115] White, T. (2012). Hadoop: the definitive guide. O'Reilly

[116] Williams, C. (2006). Profile injection attack detection for securing collaborative recommender systems. Technical Report, DePaul University

[117] Wu, L. (2013). Social network effects on productivity and job security: evidence from the adoption of a social networking tool. Information Systems Research, 24(1): 30-51

[118] Wu, Z., Cao, J., Wu, J., Wang, Y. & Liu, C. (2013). Detecting genuine communities from large-scale social networks: a pattern-based method. The Computer Journal, forthcoming

[119] Wu, Z., Wu, J., Cao, J. & Tao D. (2012). HySAD: a semi-supervised hybrid shilling attack detector for trustworthy product recommendation. Proceedings of the 18th ACM SIGKDD International Conference On Knowledge Discovery And Data Mining, 2012: 985-993

[120] Yan, L., Peng, J. & Tan, Y. (2013). Network dynamics: how do we find patients like us? Working paper

[121] Yan, L. & Tan, Y. (2013). Feel blue so go online: an empirical study of online social support among patients. Working paper

[122] Yang, T., Chi, Y., Zhu, S., Gong, Y. & Jin, R. (2009). A bayesian approach toward finding communities and their evolutions in dynamic social networks. In SDM, 2009: 990-1001

[123] Ye, M., Yin, P. & Lee, W.C. (2010). Location recommendation for location-based social networks. Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2010: 458-461

[124] Zeng, X. & Wei, L. (2013). Social ties and user content generation: evidence from flickr.

Information Systems Research, 24(1): 71-87
[125] Zhang, J. & Liu, P. (2012). Rational herding in microloan markets. Management Science, 58(5): 892-912
[126] Zhang, S., Wang, R.S. & Zhang, X.S. (2007). Uncovering fuzzy community structure in complex networks. Physical Review E, 76(4): 046103
[127] Zhang, X. & Zhu, F. (2011). Group size and incentives to contribute: a natural experiment at chinese wikipedia. American Economic Review, 101(4) : 1601-1615
[128] Zhao, J., Dong, L., Wu, J. & Xu, K. (2012) Moodlens: an emoticon-based sentiment analysis system for chinese tweets. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012: 1528-1531
[129] Zhao, J., Lui, J., Towsley, D., Guan, X. & Wang, P. (2013). Social sensor placement in large scale networks: a graph sampling perspective. arXiv preprint arXiv:1305.6489
[130] Zhao, Y., Yang, S., Narayan, V. & Zhao, Y. (2013). Modeling consumer learning from online product reviews. Marketing Science, 32(1): 153-169

Junjie Wu received his Ph.D. degree in Management Science and Engineering from Tsinghua University in 2008. He also holds a B.E. degree in Civil Engineering from the same university. He is currently an Associate professor in Information Systems Department, School of Economics and Management, Beihang University. His general area of research is data mining and complex networks, with a special interest in solving the problems raised from the emerging big-data applications. His research was supported by over 20 grants from NSFC, MOE, MOST and MIIT. He has published one monograph in Springer and over fifty papers in refereed conference proceedings and journals, such as KDD, SCIENCE, DMKD and TKDE. He is the recipient of the NSFC Excellent Young Scholars award (2013), the National Excellent Doctoral Dissertation award (2010), and the New Century Excellent Talents in University award (2011). He is a member of ACM, IEEE, INFORMS, AIS, and CCF.

Haoyan Sun is a doctoral student in Information Systems at the Michael G. Foster School of Business, University of Washington. Her research interests include online trust, social networks, and electronic commerce. She has published in International Conference on Information Systems.

Yong Tan is the Neal and Jan Dempsey Professor of Information Systems at the Michael G. Foster School of Business, University of Washington. He received his Ph.D. in Physics and Ph.D. in Business Administration, both from the University of Washington. He was a postdoctoral fellow at the University of Strathclyde and a visiting scientist at the Laboratoire de Physique Quantique, Université Paul Sabatier. His research interests include electronic, mobile and social commerce, economics of information systems, social and economic networks, and software engineering. He has published in Management Science, Information Systems Research, Operations

Research, Management Information Systems Quarterly, Journal of Management Information Systems, INFORMS Journal on Computing, IEEE/ACM Transactions on Networking, IEEE Transactions on Software Engineering, IEEE Transactions on Knowledge and Data Engineering, IIE Transactions, European Journal on Operations Research, Decision Support Systems, among others. He served as an associate editor of Information Systems Research, and is an associate editor of Management Science and a senior editor of Journal of Electronic Commerce Research. He served as a co-chair of Conference on Information Systems and Technology (CIST 2010) and the cluster chair of 2012 INFORMS Information Systems Cluster, and is a track chair of International Conference on Information Systems (ICIS 2013).