ALL ONLINE FRIENDS ARE NOT CREATED EQUAL: DISCOVERING INFLUENCE STRUCTURE IN ONLINE SOCIAL NETWORKS

Xi Chen, School of Business and Management, Hong Kong University of Science and Technology, Hong Kong, chenxinh@ust.hk

Chong (Alex) Wang, College of Business, City University of Hong Kong, Hong Kong, alex.wang@cityu.edu.hk

Xiaoquan (Michael) Zhang, School of Business and Management, Hong Kong University of Science and Technology, Hong Kong, zhang@ust.hk

Abstract

Online social network generates an online mapping of socially connected individuals. These interlinked relationship networks enable new forms of marketing activities that take advantage of the embedded interpersonal influence. However, all online friends are not created equal. Despite possible similarity of positions in the network structure, two different friends of an individual may exert significantly different influence. In the context of an online social network, we examine the influence structure on top of the network structure. A Bayesian model with a reversible jump Markov Chain Monte Carlo procedure is proposed to estimate (1) individual susceptibility to social influence from different relationship categories, (2) dyad-level tie strength within each relationship category, and (3) across-category complexity of social influence among online contacts. Our results suggest that there is significant heterogeneity in individual level social influence structure with respect to category complexity and susceptibility. We identify distinctively different patterns with respect to how activity intensity interacts with individual connectedness to determine influence structure across contact categories. These results offer insights into why seemingly similar viral marketing efforts may have markedly different outcomes. The methodology can be useful for managers to improve the effectiveness of their targeted marketing plans.

Keywords: Online social network, Social influence, Network Structure, Category complexity, Reversible jump MCMC.
1 INTRODUCTION

Social media and online social networks are hot pursuits for marketers. With Internet social networking platforms such as Facebook, Twitter, YouTube and Google+, consumers are connected like never before. On these sites, product-related contents are produced, consumed and shared rapidly and globally among consumers, making consumers themselves the most powerful marketing conduit. The power of social marketing has been widely preached and accepted. Companies from all kinds of business, Burger King, Coca-Cola, Ford, Comcast, Dell, just to name a few, are all actively engaging themselves in social media and social marketing campaigns. With the quickest growth rate in marketing mix, Forrester projected that by 2014, spending on social media marketing will exceed 3.1 billion in the US. However, according to a recent survey by Harvard Business Review and SAS, only 25% of the 2,100 companies surveyed know whether their most valuable customers were talking about them; 23% of them are using media analytic tools; and only 7% are able to integrate social media into their marketing activities. The report concludes that the companies were “yet to capitalize on the ability to not only listen to, but to analyze, consumer conversations and turn the information into insights.”

The vast marketing potential that social media brings is based on two well recognized facts. First, people influence each other. Social influence is ubiquitous, from product adoption decisions (Aral et al. 2009; Iyengar et al. 2011), website participation (Trusov et al. 2010), to post-consumption evaluation (Moe and Schweidel 2012; Wang et al. 2010). Even emotions and obesity are believed to be socially contagious (Christakis and Fowler 2007, 2009). Second, social influence is facilitated and transmitted on social networks. Social networks exhibit some structural regularities that are proven to be robust and efficient (Barabasi 2003; Watts and Strogatz 1998). To take advantage of social media platforms and to evaluate and optimize social marketing campaign activities, researchers need to be able to describe and analyze the structure of the social networks and the embedded social influence. Only with a good understanding of the network structure and the social influence structure, could social marketing campaign be effectively designed, implemented, and evaluated.

In the literature, there has been quite some discussion about how to target the influential customers - the potential promoters of products. Conventional targeting activities rely heavily on demographical and transactional information. Later, social-network-related observations were included. Network marketing has been shown to out-perform traditional ways by taking into account the social interactions among consumers (Hill et al. 2006). Hartmann (2010) finds that 10% to 50% of customer value is attributable to the effect of partners. In other words, effectively exploiting customer social connections create significant value for firms. Taking advantage of (online) social connections becomes very attractive with the development of social media and online social networking services. Interestingly, researchers in information systems and marketing are integrating network theory into the studies of online social networks (e.g. Aral 2011; Aral and Walker 2012; Trusov et al. 2010; Watts and Dodds 2007; Yoganarasimhan 2012). The literature in social-network analysis provides guidance on how network data can be used to identify influential actors and to measure performances (e.g., Wasserman and Faust 1994) However, most of the analyses in the literature are solely based on descriptive measures (self-reported social contacts or observation of social connections). One disadvantage of this type of data (and measure) is that researchers have limited knowledge about the observed ties. For example, Bob may list both Jimmy and Justin on a questionnaire about friendship. However, while Bob and Jimmy may live together, Bob and Justin may only be pen-friends who never see each other. Similarly, among Bob's 200 friends on Facebook, only a few can exert influence on him. Merely observing the existence of connections is insufficient to determine how Bob can be influenced. With social media sites and online social networking functions, detailed documentation of network topology and behaviors are available. The vast amount of data makes it possible to analyze

---

the influence structure in online social networks. With a map of influence structure in online social networks, network-based marketing could benefit tremendously. Understanding this new dimension of how social influence takes place may enable new forms of marketing campaigns.

In this paper, we propose a model that recovers a fine structure of interpersonal influence in online social networks, taking into account both observations of the social network structure and observations of behavior pattern. Current online networking services typically allow the use of tags to define social connections. It provides researchers with observations about categorization of users' online social connections. Our model takes advantage of the observation and differentiates between categories of online contacts. Based on our observation about the complexity of online connections (in the same category), our model allows for different levels of social influence within each category. The social influence coefficients are identified based on observed behavior of users, which gives a good description of the influence network structure. Research on social influence in the literature typically focuses on a single connection type and often allows no individual or dyad-level heterogeneity. The complexity of online social connections and the variation across connections are often over-looked. Our model fills this gap in the literature. Our empirical model is based on Trusov et al. (2010)'s work in this direction. They propose and implement a model that identifies the influence between users of a web site with respect to their log-in behavior. Our model is based on the same idea about parameter shrinkage but is different in several ways. First, we allow for multiple types of observed connections, thus taking advantage of the data collected through social tagging. Second, we model individual heterogeneities with a hierarchical structure and allow for cross-category correlations. Third, we allow for flexible number of levels of within-category influence strength. To estimate the model with all these features, we adapt a reversible jump Markov chain Monte Carlo procedure that is new to the literature. It allows an efficient estimation of multiple categories of social contacts. The detail of the model will be described in Section 2.

We test implemented the proposed model on a data set collected from Douban.com, a Chinese social networking website. We differentiate between two types of online social connections, namely, follow relationship and friend relationship. We estimate the model with 50,000 iterations. The estimation showed good convergence properties. Based on the model estimation, several interesting observations emerge. First, there is significant variation with respect to interpersonal social influence within each category of online contacts (friends vs. follows). Second, on average, online friends are more influential than followees (users that are followed by a focal user) with respect to site participation. These two observations suggest that there is significant variation in social influence across online contacts. In other words, “all friends are not equal” in online social networks. Not only does susceptibility to social influence depends on contact category, it also varies within each category.

Third, users' activity levels and network connectedness have different implications on category complexity and susceptibility to social influence for different connection categories.

## 2 THE MODEL

Our target is to recover the structure of influence among socially connected individuals, exploiting data available in online social networks. In the model, we consider both observed network data and observations of actual behavior (of the focal users and their social contacts). One of the most prominent characteristics of online social networks is their multiplexity, i.e., the overlapping of social circles. Internet eliminates the temporal and spatial boundaries and helps individuals to join their friends in large social circles. Contacts from different aspects of one's life are connected simultaneously online. This creates problems for marketers who are trying to understand the influence structure from the observation of social ties. The multiplexity of different friendship ties also creates problems for the users. To help users managing their online contacts, social tagging is a function that is commonly implemented by popular social media sites. With social tagging, users can categorize their online contacts into groups by assigning customizable tags. As a by-product, marketers can thus categorize online contacts of the users. Social tagging or contact categorization alone is not enough to
form a refined view about online social network structure. Within each category, there is significant variation in the strength of social connections. For example, there could be both expressive and instrumental connections under the tag “colleague” (Umphress et al. 2003). When the perceived tie strength differs, social influence differs as a result. Trusov and his coauthors find that only one-fifth of her online friends actually influence a user’s log-in frequency (Trusov et al. 2010). It is thus important to identify within-category variation in social influence.

2.1 Model Specification

Suppose the researcher is interested in recovering the structure of social influence in an activity. The target activity could be any observable actions taken by the users, for example, log-in record, sharing of information, or purchasing decisions. Denote the level of target activity of user $u$ in period $t$ as $y_{ut}$. We assume that $y_{ut}$ is a count variable and follows a Poisson distribution with rate parameter $\lambda_{ut}$, that is, $y_{ut} \sim $ Poisson$(\lambda_{ut})$.2

Assume that there are $M$ (mutually exclusive) categories of social connections observed. There are $f_{um}$ contacts of user $u$ in category $m$. Denote activity level in period $t$ of the $f$th contact of user $u$ in category $m$ as $y_{umf,t}$ ($m$ and $f$ together identify a contact). Under the Poisson regression assumption, social influence model states that:

$$log(\lambda_{ut}) = f(X_{ut}, Y_{ut-1}),$$

where $X_{ut}$ is a vector of user-specific covariates, and $Y_{ut-1}$ is a vector that contains the activity levels of $u$’s social contacts up to time $t - 1$, i.e.

$$Y_{ut-1} = (y_{u11}, ..., y_{u1,f-1}; y_{u21}, ..., y_{u2,f-1}; ..., y_{um1}, ..., y_{um,f-1}).$$

In other words, the activity level of user $u$ in period $t$ depends both on his characteristic (e.g. experience, innate preference, etc.) and the activity levels of his social contacts. We lag the activity levels of the social contacts to avoid the reflectivity problem (Manski 1993). Further, to facilitate identification, we use weighted averages of contacts’ activity levels, $z_{umf,t} = f(y_{umf,t}, ..., y_{umf,t-1})$ to denote the social influence effect. Specifically, following Trusov et al. (2010), we use the following exponential smoothing function:

$$z_{umf,t} = \sum_{d=1}^{D} w_u(d) \times y_{umf,t-d},$$

where $\sum_{d=1}^{D} w_u(d) = 1$ and $w_u(d) = \frac{\exp(-d \times \rho_u)}{\sum_{k=1}^{D} \exp(-k \times \rho_u)}$. While it is a special case of the general social influence model described by Equation 1, the model is still quite flexible as it allows for learning of peers’ activities to take effect over time and allows for heterogeneity in the speed that users react to the peers’ influence, i.e., individual discounting factor $\rho_u$.

With these additional assumptions, the general model reduces to the following Poisson regression model:

$$log(\lambda_{ut}) = \alpha_u X_{ut} + \sum_{m=1}^{M} \left( \beta_{mu} \sum_{f=1}^{D} y_{umf} \times z_{umf,t-1} \right).$$

$\lambda_{ut}$ is the parameter for the Poisson distribution, $X_{ut}$ is user-specific covariates, and $z_{umf,t}$ is the social influence variable as defined in Equation 2.

2 This assumption could be easily extended to other distribution families without significant change to the estimation procedure. For example, when investigating social influence effect in adoption behavior, one could use Probit specification or a Hazard Rate Model.
2.1.1 Model Coefficients

In model 3, $\alpha_u$ are the coefficients of user specific covariates. Social influence in the model is captured by two coefficients, $\beta_u = (\beta_u^1, ..., \beta_u^M)$ is a vector of the individual level coefficients that capture a user’s susceptibility to social influence from each contact group. To model the heterogeneity in influence strength within each contact category, we allow for discrete levels of social influence within each category of contacts. $\gamma_{uf}^m$’s are auxiliary variables that capture the relative strength of connections. We assume that there are $K_u^m$ different levels of influence strength in contact category $m$ of user $u$. With $K_u^m$ fixed, $\gamma_{uf}^m$ follows a categorical distribution and takes an integer value from 0 to $K_u^m$, i.e. $\Pr(\gamma_{uf}^m = i) = p_{ui}^m$, and $\Pr(\gamma_{uf}^m = 0) = 1 - \sum_{i=1}^{K_u^m} p_{ui}^m$. Identification of $p_{ui}^m$’s is based on the correlation in observed behavior.

$K_u^m$ captures the complexity of contact category $m$ of user $u$. When category complexity ($K_u^m$) is large, significant variation in social influence exists in category $m$. On the contrary, when a contact category is simple (for example category under the tag of “parents”), there should be little variation in the strength of connections. In that case, $K_u^m$ takes a small value. Although it would be possible to assume $K_u^m$ exogenously, researchers typically have little idea about how complex each contact category is, let alone how complex they are compared to each other. As $K_u^m$ is of theoretical interest, we allow $K_u^m$ to vary across individuals and categories and treat it as a part of the estimation. The only assumption needed is that $K_u^m$ is less than the number of contacts that user $u$ has in category $m$. Estimation of $K_u^m$ is not conventional in the literature, it involves a model selection process (because of changes in model dimensionality). To address this modeling difficulty, we implement a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm that can efficiently estimate category complexity.

2.1.2 Hierarchical Structure

To model the heterogeneity among users with respect to their responses to user-specific and social-influence covariates, we impose a hierarchical structure on the coefficients. Specifically, we assume that $(\alpha_u, \beta_u)^T \sim MVN((\bar{\alpha}, \bar{\beta})^T, \Sigma)$, where $\Sigma$ describes the covariance structure between different coefficients. In the model, we allow the individual level coefficients to correlate. The correlations between $\beta_u^m$’s are of special interest. They indicate how the susceptibility to influence from one contact category co-varies with the susceptibility to the influence from another category. For example, if a user is sensitive to his classmates’ behavior, will he be more (or less) likely to be influenced by the actions of his siblings?

It is also possible to include additional covariates to this hierarchical structure to explain individual’s susceptibility to social influence from different groups of social contacts. These moderating factors can help distinguish among multiple social mechanisms of social influences (Iyengar et al. 2011a). For the purpose of the current study, we focus on this simple structure.

2.2 Estimation Procedure

To estimate the model, we propose a Bayesian MCMC procedure to draw from posterior distribution of parameters. In general, our MCMC procedure employs Gibbs-sampling nested with Metropolis-Hasting steps. The procedure could be partitioned into two sub-routines: the individual level MCMC and the aggregate level MCMC. To estimate the within-category influence levels, we implement RJMCMC steps in individual-level draws. For exposition clarity, we illustrate the procedure with $M = 2$, with “type-1 friend” and “type-2 friend”. The procedure can process situations with any value of $M > 1$. 

Individual Level MCMC Procedure

With fixed $K^m_{u_i}$’s, individual-level coefficients’ updating follows conventional procedures as detailed in the technical appendix. Discount coefficient $\rho_{u}$’s are drawn by a random walk Metropolis-Hasting algorithm. $\alpha_{u}$’s and $\beta_{u}$’s are drawn by a Metropolis-Hasting algorithm with an independent chain sampler where the proposal density is a normal approximation to the posterior density. $y_{u_t}$’s are drawn from categorical distributions, and coefficients for the categorical distribution $p^c_{u_t}$’s are drawn directly from the conditional distribution with a Gibbs sampler. Next, we describe the reversible jump steps for updating $K^m_{u_i}$’s as it is new to the literature.

In estimating the optimal influence level $K^1_u$ and $K^2_u$, the widely-used model fitting method (e.g. Deviance Information Criterion, Bayes factor, etc.) fails to apply. To understand this, suppose that a user has 20 contacts in category 1 and category 2. To compare between different combinations of $(K^1_u$ and $K^2_u)$, one has to perform the within-model MCMC routine with fixed-level numbers for nearly 400 times. Suppose that, for each MCMC routine, 20,000 iterations in total are needed for the burn-in and posterior distribution evaluation, one has to run 8,000,000 iterations. With 10 categories of contacts and 20 contacts in each, this number grows to $20^{10} \times 20,000$ for a single user. The computational complexity will increase exponentially with the number of categories.

To tackle this problem, we introduce reversible jump moves in our MCMC procedure. The reversible jump move is a variable dimension algorithm that is capable of dealing with model selection with changing dimensionality. It has been implemented in different model selection scenarios where candidate models alter in dimensionality. Green (1995) uses this algorithm to estimate a mixture model when the number of components is unknown. Huerta and West (1999) apply it in autoregressive time-series models to choose the optimal lag number. Brooks et al. (2003) extend the use this algorithm to deal with model uncertainty in an autoregressive moving average (ARMA) process.

The jumping algorithm used in this paper is adapted from Green (1995)’s jumping scheme for mixture model. To understand the jumping scheme, let us consider a user $u$ with $j^1_u$ type-1 friends and $j^2_u$ type-2 friends. We start with a randomly generated $(K^1_u, K^2_u) = (k^1, k^2)$ and update the individual level coefficients once. With the updated individual coefficients $\{y_{u_t}^1, y_{u_t}^2, p^1_{u_t}, p^2_{u_t}\}$, we first go to the “death elimination” step. In this step, if there is no type-1 friends whose $y_{u_t} = k$, where $k < k^1$, we eliminate $k$ from the distribution and set $K^1_u = k^1 - 1$. If there are $n$ such $k$’s, we eliminate all of them and set $C = k^1 - n$. After the elimination step, we go to the “split move” step. In the split move, we randomly choose a category, for example type-1, and decide whether to move $K^1_u$ to $k^1 + 1$. To make this decision, we calculate the likelihood under the case of $K^1_u = k^1 + 1$ by randomly assigning some type-1 friends with $y_{u_t} = k^1$ to $y_{u_t} = k^1 + 1$. If the move is allowed (as detailed in the technical appendix), we move on to other categories (type-2 friends in the current case). If the split move is rejected, we go to the “combined move” step. In the combine move, we decide whether to combine level $k^1$ (the highest level for type-1) and level $k^1 - 1$ together. Similar to the split move, we evaluate the likelihood under both cases based on the most recent update of coefficients. If the combine move is accepted, we set $K^1_u = k^1 - 1$ and move to other categories. Otherwise, $K^1_u$ will not be changed in this iteration. After all $K^m_{u_i}$’s are updated, another round of coefficient updating will be carried out, and the process continues by iterating between the within-model updating and the jump steps.

The reversible jump algorithm moves the model towards a better fit. At the same time, it deals with the tendency for over-fitting. The split moves shift the chain towards more flexible models while the death elimination and the combine moves take care of the redundancy. It is an efficient way for optimal model selection within MCMC steps. A nice feature of the algorithm is that its complexity increases linearly with the number of categories.

---

3 The technical appendix is available upon request. It is not included in the submission due to page limit.
2.2.2 Aggregate Level MCMC Procedure

Following each round of individual-level coefficient updating, aggregate-level coefficients (grand mean and variance covariance matrix) are updated. We follow conventional routines to carry out the aggregate-level coefficients updating (Allenby and Rossi 1998; Albert and Chib 1993). The individual level coefficient updating and the aggregate-level coefficient updating together are termed as "within-model MCMC steps". Updating of $K_H$'s is termed as the "between-model jump steps". The complete description of the MCMC procedure and details about the RJMCMC algorithm is available in the technical appendix to this paper.

3 EMPIRICAL ANALYSIS

We apply the model to a data set collected from a major online social networking-site in China, Douban.com (http://www.douban.com).

3.1 Empirical Context

Douban focuses on user reviews and discussions about cultural products and events. According to alexa.com, Douban is ranked 113 in the world and 20 in China among the most visited web sites.4 Established in March of 2005, Douban accumulated over 8 million registered users by 2010. While unregistered users can browse the contents on the site, a registered user can keep track of his/her collection of items (books, movies, and music) and write reviews and notes to share with others (henceforth, a user refers to a registered user). Through the search or the browse function, a user can collect items in his/her profile. Collected items form a portfolio for each user. Douban's collaborative-filtering algorithms then use this information to suggest new items and calculate the distance between the interests of any two users. Registered users review an item by giving a star rating from one to five, writing the opinion in a short sentence, and (or) accompanying the star rating with a longer review. Other users would then be able to comment on the reviews and mark the reviews as either useful or not. All ratings and reviews are public.

Douban is also an online social network. Soon after its inception, in addition to offering functions to leave ratings, Douban allowed its users to create and join various groups to discuss a diverse spectrum of topics related to arts and entertainment. By mid-2010, there are more than 250,000 discussion groups. Some of the groups are private while others are public. To view contents of a private group, a user needs to become a member of that group which requires an invitation from group administrators. For public groups, any registered user can view the content and join the group to participate in the discussion.

Besides discussion groups, Douban allows its users to follow each other. Follow relationship on Douban is similar to the follow function on Twitter. If a user (the follower) follows another user (the followee), the follower will be updated about activities of the followee (e.g. collecting an item, writing a review). This makes the followee’s actions salient and influential to the follower. Follow relationships can be initiated by clicking the follow button on a user's profile page. A user can follow any other registered user without the need for consent. Users will be notified when they are followed. They can then block the follower if they prefer not. Starting from January 31, 2008, Douban further allowed users to become online friends. Similar to the follow relationship, friend requests on Douban can be initiated with a click of a button on users' profile pages. However, to become online friends, both parties need to consent, which is similar to Facebook's friend relationship. Unlike on Facebook, where users may search their email lists to connect with others they know elsewhere, friends on Douban typically do not know each other offline. It is encouraged that users make friends based on the

common interests. The site provides information about common interests when a user visits the page of other users. If two users form a friend relationship, each will be updated about the other's activities.

Douban provides an ideal setting for us to apply the proposed model and examine the influence structure. We can observe the activity levels of each user over time, the social network connections, as well as the categories of social contacts. The multiplex social network structure fits our model, with two distinct relationship categories, the follow relationship and the friend relationship. With the proposed model, we investigate the structure of social influence in users' activity levels.

3.2 Data Description

Our data cover detailed information about user activities as well as social network connections. For privacy considerations, demographical information was not provided and user identification was encoded. In preparing the data set for this study, we focus on a nine-week period starting from March 1st, 2010. To select the sample users, we randomly picked a Douban group with 376 members as of March 1st, 2010. With these 376 focal users, we then retrieved all their direct social contacts with which relationships had been established before the starting date. This procedure gave us 28,089 additional users. We then counted the daily activity levels of the 28,465 users for during the nine weeks' window. The activities in our data include rating and reviewing activities for books, movies, and music, as well as votes (useful or not) and comments on others reviews. After this calculation, we removed the users who were not active in the sampling period, the users who did not connect to any one, and the users whose social contacts were not active. This procedure left us with 181 focal users and their 13,565 social contacts.

For the 181 focal users, we calculated the average activity level during the observation period. The average of user daily activity levels is 5.98 with a standard deviation of 4.48. There is significant variation in user participation. The most active user has a daily average activity count of 35.76, while the least active user has only 1.08. Users typically exhibit consistency in participation over time.

Degree distributions for follow and friend relationship are shown in Figure 1. On average, a focal uses follows 40.58 others, the standard deviation is 62.92. The average out-degree in friend relationship is 56.53 with a standard deviation of 93.66. The correlation between user degrees in the two networks is 0.55. This suggests that sociable users tend to have more contacts in both categories.

3.3 Covariates

To control for individual heterogeneity in the activity levels, we include individual fixed effects (inner preference factor) (Chintagunta 1993). We also include the lagged activity level ($y_{u,t−1}$) as a covariate to control for the effect of habitual persistence (Guadagni and Little 1983).
4 RESULTS

Our research objective is to discover the “fine structure” of social influence with observation of social network structure and the users’ behavior patterns. We estimate the model using Bayesian methods with the RJMCMC procedure described previously. The complete model specification (including prior settings) and estimation procedure are described in the technical appendix. To accelerate convergence of the RJMCMC procedure, we performed 5 iterations of within-model MCMC steps between two between-model jump steps. Only the last iteration of the 5 iterations was kept. This helps the coefficients draws to stabilize before the model jump (Al-Awadhi et al. 2004).

Our estimation exhibited good convergence properties (Brooks and Giudici 1999). The results reported here are based on 50,000 draws (the actual iteration is thus 250,000 times) from the posterior. The first 30,000 draws were discarded as burn-in and the remaining 20,000 draws are used for analysis.\(^5\)

4.1 Results about Social Influence

4.1.1 Within Category Differentiation, \(\gamma_{uf}^m\) and \(K_u^m\)

The auxiliary variables \(\gamma_{uf}^m\)’s measure the social influence intensity on the dyad-level among contacts in the same category. It is assumed that, for each user \(u\), \(\gamma_{uf}^m\) takes integer value from 0 to \(K_u^m\). Estimates of \(\gamma_{uf}^m\)’s can be used for prediction and simulation purposes. The estimation results show that the variation of \(\gamma_{uf}^m\)’s across users and contact categories is in a limited range.

\(K_u^m\)’s capture the complexity (multiplexity) of connections in contact category \(m\). Table 1 reports basic statistics of the estimates of individual \(K_u^m\)’s for the follow and the friend relationships. Compared to the friend relationship, variability of influence level is smaller for the follow relationship (as reflected by the means). It also appears that the between-individual variation is higher for the friend relationship. This implies that, compared to the follow relationship, online friendship is more complex (in terms of social influence). Further, users tend to have different understandings about what online friendship means. This difference in understanding is not as significant for the follow relationship. It relates to the purpose that different social connection categories serve on social media platforms. The follow relationship serves as an information channel where little emotional attachment is placed. In contrast, online friends involve higher level of intimacy, which results in more within-category differentiability. The correlation between the maximum (mode) of posterior draws for \(K_u^m\)’s and \(K_u^m\)’s are 0.12 (0.18). This indicates that higher complexity in one type of relationship does not translate to higher complexity in another.

<table>
<thead>
<tr>
<th>Max of Posterior ((K_u^m))</th>
<th>Mode of Posterior ((K_u^m))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Follow</td>
<td>Friend</td>
</tr>
<tr>
<td>Mean</td>
<td>5.33</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>3.19</td>
</tr>
<tr>
<td>Maximum</td>
<td>12</td>
</tr>
<tr>
<td>Minimum</td>
<td>2</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Table 1. Estimation Result - Level of Influence Intensity

A closer look at the posteriors of \(K_u^m\) for each individual reveals some distinctive patterns. In Figure 2, we plot the posterior density for four representative users in the data set. Several observations could be

---

\(^5\) Generally, for RJMCMC procedure, there is no well accepted rule for choosing the burn-in period. Previous works normally use the first 10% to 50% of draws as burn-ins. See Green and Hastie (2009) for more details.
made from this figure. First, our RJMCMC algorithm navigates across different models nicely. This is desirable for models with changing dimensionality. A closer look at the MCMC iterations reveals that, at the later stage (iterations after 20,000 times), \( K_{m} \) s jump in a limited range with only a few states. Second, users have different category complexities. Significant heterogeneity exists among users with respect to the social influence structure. While user 114 attains a maximum level number of 14, user 136 has only 4 levels at the maximum. Third, although in general the follow relationship is less complex than the friend relationship, there are idiosyncratic cases. For example, user 92 has significantly higher category complexity in the following relationship.

![Figure 2. Posterior Densities Examples, \( K_{m} \)](image)

### 4.1.2 Susceptibility to Social Influence, \( \beta_{m} \)’s

Table 2 reports the estimation results for coefficients of susceptibility to social influence. Grand mean of susceptibility to social influence from friend (\( \bar{\beta}^{2} \)) is significantly positive. However, susceptibility to social influence from the follow relationship (\( \bar{\beta}^{1} \)) is not significant. This is consistent with the conjecture that online friends represent stronger connections than online followings. The later focuses mostly on the information content, while the former involves a certain level of psychological attachment (friend identity). We note here that the results also depend on the type of behaviors that the researchers are interested in. For participation activities studied here, social influence is mostly based on normative imitations. For other types of behaviors, such as the adoption of a new product, informational influence might play a more important role (Burnkrant and Cousineau 1975)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Posterior Mean</th>
<th>Posterior Std. Dev.</th>
<th>5% Quantile</th>
<th>95% Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{\beta}^{1} )</td>
<td>0.0057</td>
<td>0.0082</td>
<td>-0.0076</td>
<td>0.0192</td>
</tr>
<tr>
<td>( \bar{\beta}^{2} )</td>
<td>0.0148</td>
<td>0.0067</td>
<td>-0.0023</td>
<td>0.0366</td>
</tr>
<tr>
<td>( \sigma_{\beta^{1},\beta^{2}} )</td>
<td>-0.0008</td>
<td>0.0016</td>
<td>-0.0030</td>
<td>0.0020</td>
</tr>
</tbody>
</table>

Table 2. Estimation Result - Susceptibility to Social Influence
Figure 3 plots the estimates of individual susceptibility to social influence ($\hat{\beta}^1, \hat{\beta}^2$). Not all individual-level estimates of susceptibility are significantly different from zero. Online social connections are generally perceived as weak ties, it is thus expected that some online contacts exert no social influence at all. However, there are some users who are strongly influenced by their online peers. Interestingly, our estimates suggest, for some individuals, the social influence estimates are negative. This indicates that, for some users, action (content creation) of their online peers may crowd out their participation. The correlation between influence susceptibilities is not statistically significant and exhibits a negative sign (as illustrated in the figure). There might be some substitution effects between category-level social influences. In other words, if a user is strongly influenced by online friends, she should pay less attention to the actions taken by the people she follows. It is thus critical to differentiate between online social contacts of the focal user. In some viral marketing plans, managers tend to use the social network information collected elsewhere to identify influential customers. As our analysis suggests, such migration of influence network might be inefficient and may lead to wrong conclusions.

Figure 3. Individual Susceptibility to Social Influence across Categories

4.2 The Relationship between Usage, Category Size, and Social Influence

4.2.1 Category Complexity

First, we examine how category complexity relates to category size and activity intensity of the user. We use the mass points of the posteriors as the estimates of the user’s category complexity ($K^m$). We then implement medium split with respect to average activity level (number of contacts in the category) to categorize users into heavy users and light users (well-connected users and poorly-connected users). Multivariate Analysis of Variance (MANOVA) is carried out to find the main and the interaction effects of activity intensity and connectedness on category complexity. We use MANOVA to allow category complexity measures to correlate. Separate Analysis of Variance (ANOVA, reported in Table 3) analyses are also implemented. The main effects of connectedness in both categories are significant (friend relationship: Wilks' lambda = 0.813, p-value < 0.001; follow relationship: Wilks' lambda = 0.790, p-value < 0.001) Well connected individuals have significantly higher complexity in both categories. Separate ANOVA analyses reveal that category complexity only depends on the users' connectedness in the same category. The main effect of activity intensity is also significant (Wilks lambda = 0.961, p-value = 0.033). We also find that activity intensity has a positive effect on the complexity of the friend relationship (F statistics = 4.61, p-value = 0.033). But the effect is non-significant for the follow relationship (F statistics = 1.33, p value = 0.250). This implies that users with higher activity level differentiate more among their online friends, and the differentiation does not extend to the follow relationship.

An interaction effect between activity intensity and connectedness in the friend category also emerges from the MANOVA analysis (Wilks' lambda = 0.914, p-value < 0.001). However, connectedness in
the follow relationship does not interact with activity intensity (Wilks' lambda = 0.987, p-value = 0.323). Light users with high connectedness in the friend network have the highest category complexity in the friend relationship. These users focus solely on social-networking activities, thus differentiate more among their online friends. For the follow relationship, category complexity does not change much with a user’s activity level. Another interaction effect we find is between the connectedness in both categories (Wilks' lambda = 0.962, p-value = 0.036). Users who are well connected in both categories exhibit the highest complexity in both.

### DV: Follow Relationship Complexity

<table>
<thead>
<tr>
<th>Source</th>
<th>F Stat.</th>
<th>p-value</th>
<th>Source</th>
<th>F Stat.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>11.96</td>
<td>0.000</td>
<td>Model</td>
<td>4.52</td>
<td>0.000</td>
</tr>
<tr>
<td>ConnFollow</td>
<td>35.86</td>
<td>0.000</td>
<td>ConnFollow</td>
<td>9.28</td>
<td>0.003</td>
</tr>
<tr>
<td>Usage</td>
<td>1.33</td>
<td>0.250</td>
<td>ConnFriend</td>
<td>0.00</td>
<td>0.958</td>
</tr>
<tr>
<td>ConnFollow*Usage</td>
<td>0.01</td>
<td>0.928</td>
<td>ConnFollow*ConnFriend</td>
<td>3.21</td>
<td>0.075</td>
</tr>
</tbody>
</table>

### DV: Friend Relationship Complexity

<table>
<thead>
<tr>
<th>Source</th>
<th>F Stat.</th>
<th>p-value</th>
<th>Source</th>
<th>F Stat.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>19.48</td>
<td>0.000</td>
<td>Model</td>
<td>6.99</td>
<td>0.000</td>
</tr>
<tr>
<td>ConnFollow</td>
<td>43.95</td>
<td>0.000</td>
<td>ConnFollow</td>
<td>2.44</td>
<td>0.120</td>
</tr>
<tr>
<td>Usage</td>
<td>4.61</td>
<td>0.033</td>
<td>ConnFriend</td>
<td>16.94</td>
<td>0.000</td>
</tr>
<tr>
<td>ConnFollow*Usage</td>
<td>14.80</td>
<td>0.000</td>
<td>ConnFollow*ConnFriend</td>
<td>3.90</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Table 3. Results for ANOVA Analyses - Category Complexity

### 4.2.2 Susceptibility to Social Influence

Similar to our analysis on category complexity, we study the relationship between susceptibility to social influence, connectedness, and activity intensity. MANOVA results suggest that only the connectedness with respect to the follow relationship has a marginally significant main effect (Wilks' lambda = 0.971, p-value = 0.081). The interaction between activity intensity and connectedness in the friend relationship is significant (Wilks' lambda = 0.966, p-value = 0.049). However, the interaction between activity intensity and connectedness in the follow relationship is not significant (Wilks' lambda = 0.988, p-value = 0.360). Separate analysis reveals that the interaction between activity intensity and connectedness in the follow relationship is marginally significant (Table 4).

### DV: Susceptibility to Influence from Followings \( \beta_1^1 \)

<table>
<thead>
<tr>
<th>Source</th>
<th>F Stat.</th>
<th>p-value</th>
<th>Source</th>
<th>F Stat.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2.02</td>
<td>0.112</td>
<td>Model</td>
<td>2.96</td>
<td>0.034</td>
</tr>
<tr>
<td>ConnFollow</td>
<td>2.08</td>
<td>0.151</td>
<td>ConnFollow</td>
<td>1.79</td>
<td>0.183</td>
</tr>
<tr>
<td>Usage</td>
<td>1.82</td>
<td>0.180</td>
<td>ConnFriend</td>
<td>0.27</td>
<td>0.600</td>
</tr>
<tr>
<td>ConnFollow*Usage</td>
<td>2.74</td>
<td>0.100</td>
<td>ConnFollow*ConnFriend</td>
<td>7.09</td>
<td>0.009</td>
</tr>
</tbody>
</table>

### DV: Susceptibility to Influence from Followings \( \beta_2^2 \)

<table>
<thead>
<tr>
<th>Source</th>
<th>F Stat.</th>
<th>p-value</th>
<th>Source</th>
<th>F Stat.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1.46</td>
<td>0.223</td>
<td>Model</td>
<td>1.10</td>
<td>0.352</td>
</tr>
<tr>
<td>ConnFollow</td>
<td>1.79</td>
<td>0.183</td>
<td>ConnFollow</td>
<td>0.16</td>
<td>0.687</td>
</tr>
<tr>
<td>Usage</td>
<td>1.43</td>
<td>0.233</td>
<td>ConnFriend</td>
<td>3.03</td>
<td>0.083</td>
</tr>
<tr>
<td>ConnFollow*Usage</td>
<td>0.05</td>
<td>0.830</td>
<td>ConnFollow*ConnFriend</td>
<td>0.24</td>
<td>0.622</td>
</tr>
</tbody>
</table>

Table 4. Results for ANOVA Analyses - User Susceptibility to Social Influence

Upon a closer look at the group means, we find that while susceptibility to influence from the follow relationship is not different between well connected and poorly connected users when their activity levels are low. The difference becomes more significant when users activity levels are high. Heavy users who are not well connected are negatively influenced by those they follow. It suggests the crowding out effect is most significant for engaged users who connect to only a few others for
information purposes. On the other hand, light users and users who are less connected are subject to stronger social influence from friends.

The findings here contrast sharply with those about category complexity. Together, they suggest the two types of connections are different with respect to social influence patterns. For the follow relationships, there is weaker social influence in general. Activity levels do not interact with connectedness for category complexity, while they interact when predicting susceptibility to social influence. In contrast, friend relationship has stronger social influence. With respect to category complexity, heavy users are similar when they are poorly- or well- connected. Well-connected light users have significantly higher complexity in friend connections. For susceptibility to friend influence, well connected users are less influenced on average. There is no interaction between activity intensity and connectedness.

Another interesting finding is that while most well-connected users experience the highest influence from those they follow, susceptibility to social influence from friends is stronger when the users are well connected in the follow network and not well-connected in the friend network. Again, as users are making more friends online, the category complexity rather than their susceptibility to influence increases. On average, compared to poorly-connected users, well connected users actually exhibit weaker susceptibility to friend influence after controlling for within-category variation. The follow relationship, on the other hand, does not show such a pattern.

Overall, we find distinctive patterns both across influence coefficients (category complexity and susceptibility to social influence) and contact categories (friend and follow relationship). These findings underscore the necessity to analyze both aspects of social influence separately for the different types of online social connections.

5 CONCLUSION

To discover the influence structure of users in online social networks, we propose a Bayesian model which takes both the observable user-generated categorization and behavior-revealed social influence into consideration. The resulting model estimation help the managers better understand and utilize online social connections to design and manage social marketing campaigns. We implement the model on a data set collected from a popular Chinese social networking website where registered users could follow or become online friends with each other. We find significant variation with respect to interpersonal social influence within each category of online contacts. On average, online friends are more influential than followees. These findings indicate that “all friends are not equal” in online social networks. It is thus critical for managers to understand the “fine structure” of social influence between online contacts. We further investigate the effects of connectedness in each category and the activity levels of users on category complexity and susceptibility to social influence. We find that users’ activity levels and network connectedness have different implications on category complexity and susceptibility to social influence depending on the types of social connections. Our model is general in the sense that it could be easily extended to analyze other types of behaviors and it allows other categories of online social contacts to be included. It is also possible to add covariates that are expected to moderate social influence and test related theoretical predictions.

In the current implementation, we only consider the total activity level although separate observations about different types of online activity are available. Again, for different types of activities (content creation, product adoption, socializing), the salience of different underlying mechanisms for social influence differs. As a result, different conclusions could be drawn with respect to both category complexity as well as susceptibility to influence from different categories of social contacts. It is interesting to consider different activities separately and compare the estimations. This empirical model could also be combined with field experiments to offer more insights about the underlying mechanism through which social influence is transmitted.
References


