THE ESTABLISHMENT OF FRIENDSHIP TIES IN THE ONLINE HEALTH FORUM BASED ON EXPONENTIAL RANDOM GRAPH MODEL

Lin-Wei Wang  
*East China University of Science and Technology, leeven92@163.com*

Xuan Liu  
*East China University of Science and Technology, xuanliu@ecust.edu.cn*

Jia Li  
*East China University of Science and Technology, jiali@ecust.edu.cn*

Follow this and additional works at: [http://aisel.aisnet.org/pacis2016](http://aisel.aisnet.org/pacis2016)
THE ESTABLISHMENT OF FRIENDSHIP TIES IN THE ONLINE HEALTH FORUM BASED ON EXPONENTIAL RANDOM GRAPH MODEL

Lin-Wei Wang, Department of Management Science and Engineering, School of Business, East China University of Science and Technology, Shanghai, China, leeven92@163.com
Xuan Liu, Department of Management Science and Engineering, School of Business, East China University of Science and Technology, Shanghai, China, xuanliu@ecust.edu.cn, corresponding author
Jia Li, Department of Management Science and Engineering, School of Business, East China University of Science and Technology, Shanghai, China, jiali@ecust.edu.cn

Abstract

The increasingly booming online social networks promote the development of online health forum. People tend to solve kinds of health issues through the Internet. However, it is still unclear what the main characteristics of friendship networks are and how the users’ attribution affect the friendship ties establishment in the online health forum. This study uses the Exponential Random Graph Model to examine the characteristics of friendship networks and explores the effect of user attributions on the friendship ties establishment in the online health forum. The results indicated that friendship networks exhibit transitivity characteristics; meanwhile, user homophily, in terms of social identify, social activeness, age as well as geographical location have positive effects on the friendship ties establishment. The study could help patients uncover the underline mechanisms of friendship networks formation, as well shed light on how to facilitate users to make friends and finally promote prosperity effectively for the online health platform.

Keywords: online health forum, friendship networks, ERGM, friendship ties establishment.
1. Introduction

Online social networks refer to the networks that show the relationships of users in social activities based on computer networks (Yu & Fei 2008). Nowadays, Online Social Network Service (SNS) has attracted much attention because of growing groups of users and increasing social influences (Xu & Liu 2010). Users establish relations and share information through online social networks such as Facebook, Twitter, Blog and forums.

With the rapid development of Health 2.0 technologies (Van et al 2010), people are tending to utilize the internet to learn knowledge about health, seek health information or resources, make friends with patients like them, share their health status with their online friends, and seek emotional support from the online forums. Online health forums, one of the most popular online social services, become increasingly popular because they provide an open platform where users interact with each other expeditiously and solve their problems effectively. Human beings, the species who are used to being gregarious, would like to build relationships even in the online health communities.

A lot of researches have been done about the relationships in the online forum. The group dynamics theory (Maloney & Preece 2002) is useful to help understand social interaction and social relationships. Tang and Yang (2010) proposed the UserRank algorithm which combined link structure, content similarity and responding order as well as time of repliers to identify the influential users. Similar to those in the real world, some people are enthusiastic about social contacts while others are not. Niu (2010) called those who are respected by the others in the group “opinion leaders”, and referred to those who did not desire to build relationships with others, seldom publish articles or made little contribution to the forum as “isolative members”. Benevenuto et al (2012) found that interactions like browsing friends’ pages would increase the possibility of interaction among users and they also found that friends who requesting contents from others are often within close geographical location. And it was demonstrated by Song et al (2014) that health homophily and health-status homophily would increase the likelihood of collaborative friendship formation.

Former researches provide much enlightenment about the relationships in the online health forum, but only a few have studied simultaneously how the network configurations and the attributions of users affect the relationship establishment among users. From the previous literature, we can see that structural properties (Wang 2013) and individual attributes (Zhang et al 2012) would affect the network formation in social media. Our goal is to explore the main characteristics of friendship ties in the health forum and to identify determinants of friendship establishment in the health forum. In order to find a proper model to incorporate both network structure and node attributes, we reviewed models which have been applied in the previous researches. Musolesi et al (2004) proposed a new mobility model based on social network theory to describe human relations. Phuc and Phung (2007) applied naive bayes model to classify the messages in the forum. Niu (2010) used software UCINET and ID3 algorithm to explore the network relationships in a university online forum. Hill et al (2010) introduced the susceptible–infected–susceptible disease model into the social network research and investigated the emotion diffusion in social networks. Also Usage Decision Online Social Network (UDOSN) model was presented in Shahpasand et al (2012) research in order to protect users’ private data in online social networks. Meanwhile the generalized Markov Graph Model was applied to the social network for synthesis and classification (Wang et al 2013). In a summary, models above are all...
suitable for investigate one dimensionality of the social network, but they can’t be applied in
examining both of the network structure and node attributions. In our study, we applied the
exponential random graph model (ERGM) to examine which factors affect the establishment of users’
relationships. ERGM is a family of statistical models that tests whether the observed networks exhibit
theoretically hypothesized structural tendencies, and it can also deal with nodes with complex network
configurations and multiple attributions (Heaney 2014; Yan et al 2015). Our research tries to identify
the determinants of friendship establishment in the health forum based on ERGM model. It will enrich
the network literature and provide evidence about the underline mechanisms why users in the health
forum establish relationships. Meanwhile, from the perspective of e-health reality, this research will
help the platform managers know how to manage numerous forum users and boom the forum
effectively.

The rest of this paper is organized as follows. In section 2 we propose research hypothesis. In section 3
we describe the data and methodology. Section 4 presents the results analysis and the goodness of fit
tests. Finally in section 5 we conclude our research and propose some suggestions for the future
research.

2. Theory and Research Hypothesis

To explore how the network configurations and the attributions of users affect the friendship
establishment, several hypotheses are proposed based on the characteristics of network configurations
and homophily theory (McPherson et al 2001).

2.1 Network Transitivity

Actually there are kinds of configurations in directed and non-directed networks (see Figure 1) and
former literatures have explored those different types of relations (Kolaczyk & Csárdi 2014; Robinsona
et al 2007). Within those types of relations, transitivity has been well known as a key property
(Newman & Park 2003). In social networks, where entities are represented as nodes and relationships
are represented as edges, transitivity is defined as the tendency of two nodes to be connected if they
share a common node (Newman & Park 2003). Friendship is a kind of non-directed relations, and it is
usually thought that two individuals with more overlapped friends tend to build a new friendship
(Zhao et al 2015), which delivers us the enlightenment to explore the transitivity in friendship
networks. ERGM models consider network transitivity as a triangle configuration, which is
measured by the extent of transitivity pro rate to the number of triangles (Aghagolzadeh et al 2012).
To seek whether friendship ties exhibit transitivity characteristics in the forum, we proposed the
following hypothesis:

• Hypothesis 1.1: Users who have only one friend exist rarely.
• Hypothesis 1.2: Two individuals who share a common friend tend to build ties.
2.2 Homophily in social networks

Homophily is another important dimension in social networks. McPherson et al. (2001) proposed the principle of homophily, which is that connects between similar persons appear at a higher rate than those persons who are not similar with each other. Thelwall (2009) argued the existence of universal homophily specified demographic homophily to be race homophily, religion homophily, age homophily, gender homophily, marital status homophily, and so on.

As mentioned above, attribution homophily contains various aspects. Actually some former researches tried to classify different types of homophily. Choudhury (2011) measured the attribution homophily among individuals from three dimensions: demographic attributes, activity-specific attributes and content based attributes. Manavopoulos (2014) classified the homophily into demographic, political orientation and attitudinal characteristics. And Song et al. (2014) incorporated homophily into e-health research and they studied health related homophily based on two dimensions: treatments and health condition. In our paper, we classify node attributions into two types of homophily: social homophily and demographic homophily. The former focused on the social identity and social activity in the health forum; while the latter focused on the users’ demographic attributions.

2.2.1 Social homophily

Individuals showed different preferences when they establish friendship ties. Goodreau et al. (2009) specified “sociality” as one of those preferences, which is formed as the result of random process shaped by sociality. People with similar social characteristics will attract each other psychologically, which is called social homophily (McCrea 2009). Heath et al. (2007) enumerate some characteristics about social homophily, such as similar jobs, similar career paths and similar hobbies. To explore the effect of social homophily on friendship ties establishment, in our study, we focused on two attributions, the first is social identity” (here we mean user’s social roles in the health forum, like patients, doctors, relatives and so on) and the other is “social activeness” (people in the same forum exhibit different levels of activeness: some contribute a lot and are tending to publish posts and answer other’s questions frequently, while others like to be silent). Hence we proposed the following two
hypotheses:
• Hypothesis 2.1: Users with the same social identity groups are more likely to make friends with each other.
• Hypothesis 2.2: Users with the same level of social activeness are more likely to build friendship ties.

2.2.2 Demographic homophily

According to previous papers, the demographic homophily represents the similar demographic attributions of human beings such as location, gender, race, marital status and so on (Choudhury 2011). Rusmin et al (2014) also summarized some demographic variables, like gender, age, education level, income level, employment concentration, and population growth. In our research, we explored three demographic attributions, including gender, age and geographical location. In order to explore how demographic homophily affects the friendship ties establishment, we proposed the following hypotheses:
• Hypothesis 2.3: The friendship establishment is more likely to occur between users with the same gender.
• Hypothesis 2.4: Users in the same age group are more likely to make friends.
• Hypothesis 2.5: Users within close geographical location are more likely to establish friendship ties.

3. Data and methodology

3.1 Data description

In order to test our hypotheses, we collected the data from a diabetes forum (http://bbs.tnbz.com/), a well-known online diabetes community in China which provides patients, doctors and relatives to interact with each other. It was established in around year 2005 and up to now has 177,240 memberships in total. It is also the largest website among all those Chinese oriented diabetes forums. As mentioned above, diabetes is a chronic disease which still can’t be cured thoroughly so that it would lead the patients to pay much attention to their health conditions persistently. Thus the patients need to obtain the relevant information and trace the latest therapeutic methods, which ensure the vitality of users in the forum. The forum is functioned as a platform where patients could seek better therapeutic schedule and diffuse their helpful experiences, doctors could offer professional suggestion and relatives could exchange helpful information and seek emotion from anyone else. People can also visit someone’s homepage if they are interested in him, and even friendship relation with him.

We collected all those posts and user homepages from the website. Posts from the forum span from October 1, 2005 to May 22, 2015, and the total users in the forum are included. Since our research focuses on the effect of the individual attributes on friendship formation, we then filtered those users with completed attributions and friendship networks. Finally, our research testbed included a dataset of 693 friendship ties among 279 users, accompanying with their attributions including social identities in the forum (categorical variable: doctor, relatives, customer service, and different types of diabetes), social activeness (ordinal variable, there is an original variable in the platform represents each user’s
activeness in the forum and we transferred it into 7 scales), gender (binary variable, 1 represents male and 0 represents female), age group (binary variables, 1 represents above average age and 0 represents below average age), geographical location (categorical variable, we used area codes to represent users’ residence). The network structure is shown as Fig.2.

Figure 2. The friendship networks

3.2 ERGM

As mentioned in section 1, exponential random graph model (ERGM) was applied in our study. ERGM combines both the network configuration and node attributions. In this model, a network graph is represented as a mathematical relation, or a binary network. The number of nodes in the graph is n. The random variable $Y_{ij}$ indicates whether there exists a tie between nodes i and j ($Y_{ij} = 1$) or not ($Y_{ij} = 0$) (Snijders et al 2006). The ERGM generates the random networks which are based on the observed networks and compare the estimated networks with observed networks. The more they are similar, the better the estimations are. Then we incorporate the nodes’ attributions into the model. We use the term “node main” which represents the main effects to examine the sociality of nodes. Meanwhile we use the term of “node match” which is called second-order effects to examine various attribution homophily (Kolaczyk & Csárdi 2014). The ERGM will estimate the parameters which indicate that the frequency of the configurations are observed and how significance the node attributions’ main effects and second-order effects affect the formation of ties (Kolaczyk & Csárdi 2014; Robinsa et al 2007).

The general mathematical formulation of ERGM is as follow:

$$Pr(Y = y) = \left(\frac{1}{k}\right) exp\{\sum_A \theta_A g_A(y)\}$$

(1)
Where the summation is over all configurations A; y represents one kind of particular network graph y; \( \theta_A \) is the parameter corresponding to the configuration A; \( g_A(y) \) is the network statistic corresponding to configuration A; \( g_A(y) = 1 \) if the configuration is observed in the network y, and \( g_A(y) = 0 \) otherwise; k is a normalizing number which ensures that (1) is a proper probability distribution (Robinsa et al 2007).

To specify how ERGM applies to test our hypotheses, we listed the hypotheses with ERGM configurations. The research hypotheses, their corresponding ERGM configurations, and graphical illustration are represented in table 1.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Level of Analysis</th>
<th>Illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1.1: Users who have only one friend rarely exist.</td>
<td>Dyadic level</td>
<td>Friendship tie Node A Node B</td>
</tr>
<tr>
<td>H1.2: Two users who share a common friend tend to build ties.</td>
<td>Triadic level</td>
<td>Friendship tie Node A</td>
</tr>
<tr>
<td>Hypothesis 2.1: Users with the same social identity groups are more likely to make friends with each other.</td>
<td>Dyadic level</td>
<td>Friendship tie User A with group X User B with group X</td>
</tr>
<tr>
<td>Hypothesis 2.2: Users with the same level of social activeness are more likely to build friendship ties.</td>
<td>Dyadic level</td>
<td>Friendship tie User A with active level X User B with active level X</td>
</tr>
<tr>
<td>Hypothesis 2.3: The friendship establishment is more likely to occur between users with the same gender.</td>
<td>Dyadic level</td>
<td>Friendship ties Male User A Female User B</td>
</tr>
<tr>
<td>Hypothesis 2.4: Users in the same age group are more likely to make friends.</td>
<td>Dyadic level</td>
<td>Friendship tie User A in age Group X User B in age Group X</td>
</tr>
<tr>
<td>Hypothesis 2.5: Users within close geographical location are more likely to establish friendship ties.</td>
<td>Dyadic level</td>
<td>Friendship tie User A at location X User B at location X</td>
</tr>
</tbody>
</table>

Table 1. Research Hypotheses, Levels of Analysis, and Graphical Illustrations

4. Result analysis and goodness of fit tests

4.1 The analysis of results

Table 2 reports the results of ERGM estimation. The estimated parameter represents the appearance frequency of specific configuration in the observed network. According to former studies, a parameter is significant if the estimate is at least twice as the standard error (Pahor et al 2008; Su & Contractor 2011).
<table>
<thead>
<tr>
<th>Type</th>
<th>Hypothesis</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>P-value</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network transitivity</td>
<td>H1.1</td>
<td>-5.8993</td>
<td>0.16867</td>
<td>&lt;1e-04</td>
<td>Supported</td>
</tr>
<tr>
<td></td>
<td>H1.2</td>
<td>0.74827</td>
<td>0.03982</td>
<td>&lt;1e-04</td>
<td>Supported</td>
</tr>
<tr>
<td>Social homophily</td>
<td>H2.1</td>
<td>0.39793</td>
<td>0.07862</td>
<td>&lt;1e-04</td>
<td>Supported</td>
</tr>
<tr>
<td></td>
<td>H2.2</td>
<td>0.07980</td>
<td>0.01357</td>
<td>&lt;1e-04</td>
<td>Supported</td>
</tr>
<tr>
<td>Demographic homophily</td>
<td>H2.3</td>
<td>-0.15592</td>
<td>0.08825</td>
<td>0.0773</td>
<td>NOT Supported</td>
</tr>
<tr>
<td></td>
<td>H2.4</td>
<td>0.20708</td>
<td>0.05231</td>
<td>&lt;1e-04</td>
<td>Supported</td>
</tr>
<tr>
<td></td>
<td>H2.5</td>
<td>0.89665</td>
<td>0.09877</td>
<td>&lt;1e-04</td>
<td>Supported</td>
</tr>
</tbody>
</table>

Table 2. Results for ERGM estimation

From table 2, we can see that H1.1 was supported, the negative parameter indicates rare friendship ties only exists within two nodes, in other words most users have more than one friend. H1.2 was supported which provided evidence that there’s significant transitivity pattern in the friendship networks. Users tend to establish friendship ties if they share common friends. We can conclude that people in the health forum are desired to seek more friends just like what they do in the true world; meanwhile transitivity is the important mechanism for friendship tie formation (Aghagolzadeh et al 2012).

H2.1 was supported, indicating that users prefer to build friendship ties with those having the same social identity. This may be because users in the online health forum tend to share their situation and exchange information with those who own similar cognition with them. H2.2 was supported, indicating that activeness homophily has a positive effect on friendship ties’ establishment, and users tend to make friends with those individuals who have similar personality with them. H2.3 was not supported, and the negative parameter indicates that users of the same gender seldom establish ties with each other. At the same time, the non-significant parameter shows that gender has little influence on friendship formation. This result is consistent with the findings in previous studies -- Thelwall (2009) also found that gender homophily maybe has no effect in the social networks. H2.4 and H2.5 were supported, providing evidences that users with similar age group and geographic homophily are more likely to establish friendship ties with each other. Users with these two kinds of homophily among users may indicate the higher possibility to transfer users’ virtue friendship into the real life relations. So users with similar age group and geographic homophily will promote the formation of friendship. In summary, most attribution homophily could promote the friendship ties establishment (McPherson et al 2001) except for one specific attribution: gender.

4.2 Goodness of Fit Tests

To examine whether the model fit the data well, we plotted the goodness of fit plots shown as Figure 3. The plots presented the comparisons between original friendship networks with estimated parameters gained from ERGM. The comparisons include three dimensionalities which contain degree, edge-wise shared partners, and geodesic distance. In each part, respectively, the bold solid lines represent the original networks; the dashed lines represent variance; boxplots cover the median and interquartile range. The light gray curves represent the range in 10th and 90th quantiles. Small differences between the original networks and the observed networks indicate that our ERGM models fit the observed data.
well.

5. Conclusion

Our study used the approach ERGM to explore the effect of network structure and node attributions on the friendship ties establishment. Using data from the largest diabetes forum in China, we conducted empirical researches and tested our seven proposed hypothesis. We found that friendship ties in the online health forum show significant transitivity pattern that users tend to make more friends especially when two individuals share common friends. We also found that attributions homophily such as social identity, social activeness, age, geographic location had positive effect on the friendship ties establishment, while gender homophily effect was distinguished from those previous attributions. These conclusions enlighten us to selectively design mechanisms for friend recommendation and promote the e-health platform.

Our study innovatively incorporated the network structure and users’ attributions into the network research simultaneously. Meanwhile this research provides new supporting evidences for former literature and contributes to both social network and electronic health research. Practically, this study could help patients uncover the underline mechanisms of friendship networks formation, as well shed light on how to facilitate users to make friends and finally promote boom effectively for the online health platform.

Some limitations of this study provide avenues for future research. First, this study considered only one type of network -- friendship network. Besides friendship relations, there are also browsing relations, ask-and-answer relations and so on, further researches can be extended to examine multiple types of networks (i.e. directed and non-directed networks) and various types of relations (i.e. friendship vs knowledge diffusion vs information diffusion) and explore how those networks interact with each other. Secondly, this paper emphasized on the antecedents of friendship ties. Future analysis could be investigated about the consequence of friendship establishment. Through establishing friendship ties, people can exchange useful information with others and seek emotional supports from them. Of course, the platform should also pay much attention to the dark side of social media. Forums,
besides offering emotional support, might also lead to depression or beyond due to the diffusion of erroneous information and biased opinions. It’s also worthwhile exploring this kind of issues in our further researches.

Acknowledgements

This study was supported by the Fundamental Research Funds for the Central Universities and the National Natural Science Foundation of China (Grants Number: 71471064, 71101053, 61104139, 71171131, and 71103021), and was also sponsored by Shanghai Pujiang Program (Grant Number 15PJCO19).

References

Mccrea R. (2009). Explaining sociospatial patterns in South East Queensland, Australia: social

