

A Fuzzy Social Network Analysis Method and a Case Study on *Tianya* Tourism Forum in China

Zi Lu, Ruiling Han, Weilu Du, Dianshuang Wu¹

Abstract. Social networking service (SNS) has become online service platforms that focus on facilitating the building of social networks among people who share interests, activities, backgrounds or real-life connections, and has had a rapid development in China in the past few years. This paper aims to develop a fuzzy social network service analysis method, which combines graph-theory with related fuzzy approach, to analyze the social network structural features and the distribution characteristics of interpersonal nodes in SNS community. A case study on a very famous Chinese tourism BBS – *Tianya* is conducted to illustrate and validate the proposed approach. The research findings are: 1) the attraction degrees of various areas in the forum are significantly different; 2) interpersonal nodes in the forum are concentrated relatively; 3) the fuzzy out-degrees and the fuzzy in-degrees of interpersonal nodes in the forum conflict with each other; 4) the distribution of interpersonal nodes is influenced by geographical relations. These findings can directly support social network service management and particularly tourism online service developments.

Keywords: Social networking service. Fuzzy systems. *Tianya* tourism forum.

1 Introduction

Social networking services (SNS) realize the internet application services in social networks which establish a virtual interpersonal network with the help of the network virtual platform [1]. They are based on the real interpersonal relationships, and turned into “social networking service community” (SNS community) gradually. As a continuation of real communications, the users in a SNS community

¹ Z. Lu, R. Han, W. Du

School of Tourism, Hebei Normal University, Shijiazhuang, Hebei 050024, P.R. China

D. Wu (✉)

Decision Systems & e-Service Intelligence Lab, Centre for Quantum Computation & Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia

e-mail: Dianshuang.Wu@student.uts.edu.au

have the same hobby, belief and status, which will form a virtual social relationship after a period of evolution [2]. SNS take advantages of both the virtual and real communities. They mainly contain four types: social-oriented, campus-oriented, business-oriented and entertainment-oriented.

The studies on SNS community were mainly focused on SNS marketing development, SNS community member interaction models and the enlightenments to other sharing platforms, involving all types of SNS. The users' social network characteristics can be mined by deep analysis of the user relationship network. Recently, SNS community researches began to focus on the geographic factors. Lu and Wang [3] applied two methods, namely, information entropy and degree distribution to study the interpersonal node spatial distribution characteristics of SNS community and the significance of geographical factors. Some social network theoretical analysis has also been done. For example, Abbasi et al. [4] developed a theoretical model based on social network theories and analytical methods, using measures from social network analysis (SNA).

The social network is a method for quantitative analysis of social relations, which mainly includes two fundamental elements, namely, actors and relations. The nodes of social network graph represent the information mediator and recipients. The arrows indicate the direction of information transfer. The thickness of the connections indicates the frequency of the information transmission or the amount of information transmitted. The overall reflects the statistical features of the information flow between the members of the group. The methods to analyze the social network include: graph-theory method, matrix method, social metrology method, algebraic method, etc. Among these methods, the graph-theory method is suitable for describing the relations in small groups and expressing the structure characteristics of network intuitively. So far, in geographic fields, the graph-theory method has been used in many aspects, such as regional structural characteristics and development direction [5], inbound tourism flow of network frame and structure features of tourism space [6].

With the in-depth research into the nature of online networking community, it is found that there are non-balanced characteristics in network structure, that is, network consists of several of "nodes" and "connections", and the connections between some nodes are very close while the connections between others are sparse [7] relatively. It is believed that the relationship between actors in social network is fuzzy, and the relationship between actors could not be simply divided into binary relation – "Yes (1)" and "No (0)". Hence, Nair and Sarasamma [8] proposed the definition of fuzzy social networks. They took the nodes and edges in fuzzy graphs as actors and the relationships between actors in social networks respectively, and took that fuzzy social network as a result of giving practical meanings to the fuzzy graph. Ciric and Bogdanovic [9] defined social network as a fuzzy relational structure. They pointed out that the social network is a special case of fuzzy social network. Although social networks has been widely studied, and fuzzy social network has been proposed [10], there are very limited research in

SNS, particularly, no report about the application developments of fuzzy social network analysis in tourism online services.

This paper proposes a fuzzy social network analysis method, which combines graph-theory method with fuzzy approach. As online services in tourism industry have been developed rapidly [11, 12], a case study on a famous Chinese tourism BBS – *Tianya* (<http://travel.tianya.cn/>) is conducted based on the proposed method. The structural features of SNS community network and the distribution of community interpersonal nodes are analyzed. It is a significant exploration to use fuzzy approach in social network analysis and study spatial pattern of community members and their communications, which expands SNS research methods and improves the applications and research of SNS community structure in geography. The research findings in this study can support social network service management and particularly tourism online service developments, and can also help local governments to make decisions on the development of tourism industry.

The rest of the paper is organized as follows. The fuzzy social network analysis method is described in Section 2. In Section 3, a case study on *Tianya* tourism forum is conducted and illustrates the applications of the proposed method. A set of conclusions are presented in Section 4.

2 Fuzzy Social Network Analysis Method

This section will first give some important concepts of fuzzy social networks. It will then present our fuzzy social network centrality analysis method.

2.1 Basic concepts of fuzzy social networks

Definition 1. Fuzzy social network can be defined as a fuzzy relational structure $\tilde{G} = (V, \tilde{E})$, where $V = (v_1, v_2, \dots, v_n)$ is a non-empty set of actors, $\tilde{E} = \sum_{i=1}^n \sum_{j=1}^n \tilde{E}(e_{ij})/e_{ij}$ is the fuzzy relation on V .

For directional fuzzy social network, the fuzzy node degree can be categorized into fuzzy node in-degree and fuzzy node out-degree. The former refers to the number of head endpoints adjacent to a node and the latter represents the number of tail endpoints adjacent to the node. The in-degree represents the popularity degree of a node, which is a measurement of the “power” or the attraction degree when an individual releasing or acquiring information. The out-degree of a node illustrates the activity level of the nodes in the network, and measures the individual influence.

Definition 2. The fuzzy node degree of a node is the sum of membership degrees of the edges that attach to the node, denoted as: $\tilde{d}(n_i) = \sum_{j=1}^n e_{ij}$. The average node degree of the fuzzy social network is: $\bar{d} = \sum_{i=1}^n \tilde{d}(n_i) / n$, where n is the number of nodes.

Definition 3. The scale of a fuzzy social network refers to the number of actors in fuzzy social network.

The scale affects the relation complexity between actors in a fuzzy social network. The larger scale means more members, thus more complex of their relationship structure.

2.2 Social network centrality and fuzzy social network centrality

The following concepts are given.

- Node centrality

Node centrality is the central level of a node's position in the network. Node degree represents the number of nodes which directly connect to a node, namely, the number of lines that connect to it. If the node centrality is greater, the actor's position will be more central, and more other actors will connect to the actor. The node centrality is represented as: $C_D(n_i) = d(n_i)$.

In the formula above the centrality is calculated with the absolute number, and the center is the actor that has the maximum number. Another way to calculate the centrality is based on the relative number, which means the ratio of the node degree of a specific node to the maximum degree of any node, which can be expressed as: $C'_D(n_i) = d(n_i) / (N - 1)$ where N is the scale of network, and the maximum degree of any node is $N-1$.

The following two issues when measuring node centrality based on node degree should be noticed:

A. this measurement is based on direct relationship, and the indirect relationship is not taken into account;

B. whether there is a unique "center" in the entire network is not considered in the measurement.

The group central potential will solve these two issues.

- Group central potential

Central potential is a concept to reflect the concentration degree of the entire network or graph. It is an important mutually complementary measurement with the density. According to the absolute centrality, the formula of central potential can be expressed as: $C_D = \sum_{i=1}^n (C_{D_{\max}} - C_{D_i}) / \max \sum_{i=1}^n (C_{D_{\max}} - C_{D_i})$.

According to the relative centrality, the formula can be expressed as:

$$C_D = \sum_{i=1}^n (C_{D_{\max}} - C_{D_i}) / (N-1).$$

- Fuzzy node centrality

The fuzzy node centrality of an actor is the central degree of the node in a fuzzy social network, which is expressed by the fuzzy node degree of the node. The fuzzy node degree of a node is the sum of membership degrees of the edges that attach to the node. The greater is the fuzzy node centrality, the more central is the actor in the fuzzy social network. The fuzzy node centrality is denoted as: $\tilde{C}_D = \tilde{d}(n_i)$, where $\tilde{d}(n_i)$ means membership sum of the fuzzy relations connected with n_i .

In the formula above the fuzzy centrality is calculated with the absolute number, and the center is the actor that has the maximum number. Another way to calculate the fuzzy centrality is based on the relative number, which means the ratio of the fuzzy node degree of a specific node to the maximum degree of any node. It can be expressed as: $\tilde{C}_D(n_i) = \tilde{d}(n_i) / (N-1)$, where N is the scale of network, and the maximum degree of any node is $N-1$.

- Fuzzy group central potential

Fuzzy group central potential is a concept to reflect the concentration degree of the fuzzy social network. According to the absolute centrality, its formula can be expressed as: $\tilde{C}_D = \sum_{i=1}^n (\tilde{C}_{D_{\max}} - \tilde{C}_{D_i}) / \max \sum_{i=1}^n (\tilde{C}_{D_{\max}} - \tilde{C}_{D_i})$.

According to the relative centrality, the formula can be expressed as:

$$\tilde{C}_D = \sum_{i=1}^n (\tilde{C}_{D_{\max}} - \tilde{C}_{D_i}) / (N-1).$$

From this model, we can obtain the centrality degree of each node in the fuzzy social network and evaluate the concentration degree of the fuzzy social network.

3 A Case Study on *Tianya* Tourism Forum

3.1 Data sources

Tianya community is the most influential social network platform in China, which has more than 70 million registered users. The function of *Tianya* tourism forum (<http://www.tianya.cn/bbs/index.shtml>) is to provide communication platforms for all forum users. The communication is mainly completed in the mode of posting and replying posts by users. The tourism forum contains four themes: general, hot, regional and thematic topic. It is further subdivided into 24 travel parts based on

spatial relations, such as “Caiyunzhinan”, “Bashanshushui”, “Zoujin Tibet”, “Meizai Guangxi”, etc. This paper is mainly based on the 22 regional sections (except “Changyou Taiwan” and “Huancai Xiangjiang”). The number of main posts and the total number of posts of each forum section from September 2012 to November 2012 are aggregated. In each region, the article with the most repliers is seen to represent the whole situation of the region. The articles with the most repliers in each forum section are extracted, the regions of posters and repliers are reviewed, and the numbers of the replies are counted. When taking statistics, the repeat replies are excluded.

In this case study, taking the areas where the poster and the replier locate as the nodes, and taking the connections between them as the lines between nodes, the replying relationship diagram is drawn in Fig. 1.

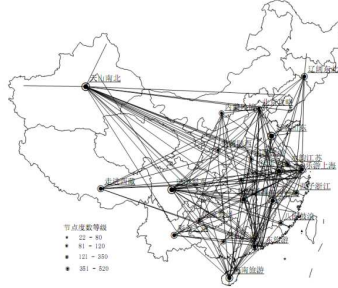


Fig. 1 Locations of the posters and the repliers

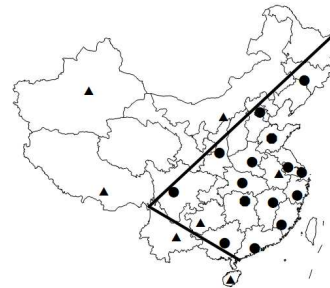


Fig. 2 The spatial distribution of the two types of nodes

3.2 Nodes Centrality

Table 1 The absolute centrality and relative centrality of regions

| Areas | NC | FNC | Areas | NC | FNC | Areas | NC | FNC | Areas | NC | FNC |
|----------------|-----|-------|----------|-----|-------|----------|-----|-------|-----------------|-----|-------|
| Xinjiang | 520 | 91.68 | Sichuan | 417 | 86.34 | Shandong | 362 | 79.71 | Hunan | 190 | 61.09 |
| Hainan | 168 | 65.81 | Anhui | 155 | 44.74 | Jiangxi | 139 | 35.11 | Guangdong | 115 | 55.33 |
| Yunnan | 112 | 58.20 | Shanghai | 110 | 51.22 | Tibet | 87 | 41.65 | Northeast China | 85 | 52.77 |
| Inner Mongolia | 77 | 62.34 | Fujian | 64 | 41.12 | Shanxi | 64 | 52.67 | Guizhou | 58 | 31.29 |
| Beijing | 53 | 20.92 | Hubei | 41 | 27.53 | Guangxi | 35 | 22.25 | Zhejiang | 32 | 22.70 |
| Jiangsu | 29 | 26.23 | Henan | 22 | 20.41 | | | | | | |

Note: NC – Node centrality, FNC – Fuzzy node centrality.

Statistics are made on the number of participants and the locations of posts with the most replies in the selected 22 districts. The results are ranked according to the node centrality and fuzzy node centrality (Table 1). When calculating the fuzzy

node centrality, the membership of each node is calculated as the ratio of the number of local replied posts to the number of total replied posts of the node. As a comparison, the rankings of the tourist number of various regions are shown in Table 2. It can be seen that the online community ranking of the 22 areas is not the same with the actual annual tourist reception ranking.

Table 2 The rankings of the tourist number of regions in 2010

| Areas | TRR | Areas | TRR | Areas | TRR | Areas | TRR | Areas | TRR |
|----------|-----|-----------------|-----|----------------|-----|---------|-----|----------|-----|
| Xinjiang | 26 | Sichuan | 22 | Shandong | 7 | Hunan | 13 | Hainan | 25 |
| Anhui | 12 | Jiangxi | 21 | Guangdong | 1 | Yunnan | 9 | Shanghai | 2 |
| Tibet | 28 | Northeast China | 15 | Inner Mongolia | 18 | Fujian | 6 | Shanxi | 11 |
| Guizhou | 27 | Beijing | 5 | Hubei | 14 | Guangxi | 10 | Zhejiang | 3 |
| Jiangsu | 4 | Henan | 17 | | | | | | |

Note: TRR – Tourist receipts ranking.

The interpersonal node network of tourism forum is a directed network. The fuzzy node in-degree of a region is the number of the replies of “good-quality articles” in the regional section, and the fuzzy node out-degree is the sum of the replies from the region. Among the 22 regional nodes, except Shandong, Hunan and Jiangxi, the other 19 nodes have significantly different fuzzy node in-degree and fuzzy node out-degree. For the areas of Xinjiang, Hainan, Anhui, Yunnan, Tibet, Inner Mongolia and Guizhou, the in-degree is significantly larger than out-degree. These nodes have high degrees of consideration and attractiveness in the network, but low activity degree. On the contrary, for the areas of Sichuan, Guangdong, Shanghai, Northeast China, Fujian, Shanxi, Beijing, Hubei, Guangxi, Zhejiang, Jiangsu and Henan, the out-degree is larger than in-degree. These nodes have a low consideration degree but high activity degree. The node of Xinjiang has the highest difference between fuzzy node in-degree and out-degree, which has the highest consideration and attractiveness, but is not active in the entire network. Conversely, Guangdong has the highest out-degree, but the lowest in-degree. It is the most active in network, but has the low degree of consideration and attractiveness (Fig. 3).

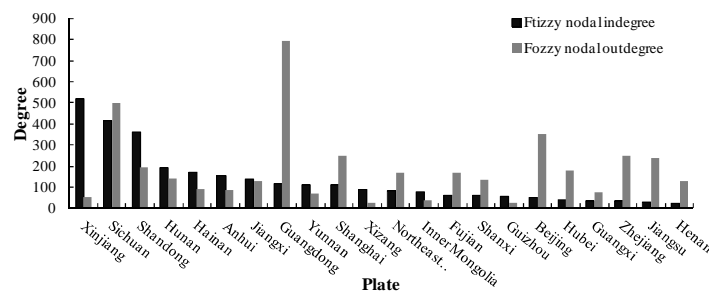


Fig. 3 The comparison between fuzzy node in-degree and out-degree of each node

The 22 nodes are divided into the following two types: for one type, the fuzzy node in-degree is greater than the fuzzy node out-degree; for the other type, the fuzzy node out-degree is greater than the fuzzy node in-degree. They are labeled in the map (Fig. 2). It is found that nodes of the former type are widely distributed in northwest and southwest of China, and nodes of the latter are distributed in northeast, mid-east and southeast of China.

The distribution of the repliers is also affected by geopolitical factors and presents the local concentration. In the tourism forum, the regions are basically divided according to provincial-level administrative units. Therefore, the associated region of the replier can be determined by the province where the replier locates. The number of “locals” among the repliers can then be determined. In the “high-quality articles” of each area section, there are more “local” repliers. The locals are more active. Except Xinjiang and Tibet, the proportions of “local” repliers are the largest in other sections (Fig. 4). It is concluded that the communication in the tourism forum is still subject to the impact of geospatial.

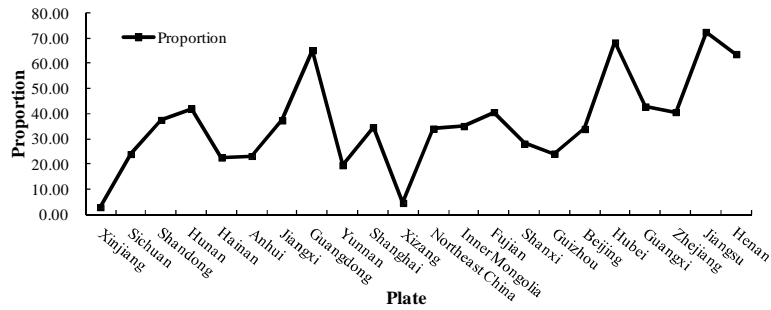


Fig. 4 The proportions of the regional locals in total number of people

3.3 Central potential

Central potential reflects the degree of concentration of the entire network and whether there is an absolute center of the network. The value range of central potential degree is 0 to 1. The network will be more centralized as the central potential approaches 1. There will be no absolute central point in the network and it will present a state of decentralization if the central potential approaches 0. According to the formulas of group central potential, the group central potential of the network is 0.776, and the fuzzy group central potential is 0.616. The central potential is moderate, which proves that the network presents a centralized state and has outstanding convergences and central points. The sections of tourist hotspots such as Xinjiang, Sichuan, Shandong, Hunan and Hainan have a high attraction.

4 Result Analysis and Conclusions

Through the case study and related data analysis, we obtained the following conclusions.

- The attraction degrees of various areas in the forum are significantly different
The nodes with great centrality and fuzzy centrality obtain high attention in the network, which are often referred to “opinion leaders” with larger “power”. They play a large part in the information dissemination, and guide the public opinion in a certain extent. By surveying the number of posters and repliers of those “articles” in each area and analyzing the results of node centrality and fuzzy node centrality, it is found that there exist great differences in the centrality of various areas, which indicates that the attraction degrees of nodes or the activity degrees of interpersonal nodes in different regions are significantly different. These degrees are greatly related to the attracting degrees of the section, quality of posts in the section and the tourist attraction degree where the section locates. However, the ranking of node centrality and the ranking of real tourism attraction is not consistent with each other, which indicates that the virtual community breaks the actual geographical space limitations and changes people’s concern points and tourist trend.

- Interpersonal nodes in the forum are concentrated relatively

The central potential degree of the network is around 0.5, which indicates that there is no obvious network concentration and there is not an absolute “power center” in the network. The network presents a decentralized feature. The interpersonal nodes of the forum present a decentralized state. There is no absolute hotspot. The users are not limited to focus or reply one regional section, but to consider multiple forum sections.

- The fuzzy out-degrees and the fuzzy in-degrees of interpersonal nodes in the forum conflict with each other

The high fuzzy node out-degree means that it is more convenient for the node to access information. The high in-degree illustrates that the node has a high attraction degree and great “power” in information releasing. Among the 22 regional nodes, except Shandong, Hunan and Jiangxi, the other 19 nodes have significantly different fuzzy node in-degree and fuzzy node out-degree. For the areas of Sichuan, Guangdong, Shanghai, Northeast China, Fujian, Shanxi, Beijing, Hubei, Guangxi, Zhejiang, Jiangsu and Henan, the fuzzy out-degree is larger than the fuzzy in-degree. For the areas of Xinjiang, Hainan, Anhui, Yunnan, Tibet, Inner Mongolia and Guizhou, the fuzzy in-degree is significantly larger than the fuzzy out-degree. In general, the nodes in the network are divided into two types: one type has high degrees of consideration and attractiveness in the network but low activity degree, the other type has a low consideration degree but high activity degree. The nodes of the former type are widely distributed in northwest and southwest of China, and nodes of the latter are distributed in northeast, mid-east and southeast of China.

- The distribution of interpersonal nodes is influenced by geographical relations
For the division of sections in the tourism forum, on one hand, users concern more to local or close areas; on the other hand, the local concentration of friends in the virtual community influence the users' concern scopes. Once their friends post, they will concern about them, which provides the possibility and convenience for replying. Hence, the number of interpersonal nodes is reduced with the distance of the regional section increased, which indicates that the distribution of interpersonal nodes affected by geopolitical factors presents local concentration. The virtual community is still constrained by the geographical space, and the distribution of interpersonal nodes is still influenced by geopolitical factors.

Acknowledgment The work presented in this paper was supported by the National Natural Science Foundation of China (Grant No.41271142).

References

1. Boyd, D.M., Ellison, N.B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.
2. Heidemann, J., Klier, M., Probst, F. (2012). Online social networks: A survey of a global phenomenon. *Computer Networks*, 56(18), 3866-3878.
3. Lu, Z., Wang, W.T. (2011). The spatial distribution characteristics of interpersonal node in social networking services community and the analysis of geopolitical factors. *Scientific Geographica Sinica*, 31(11), 1293-1300.
4. Abbasi, A., Altmann, J., Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, 5(4), 594-607.
5. Liu, H.Y., Wei, L.L., Zhang, J. (2012). The research on the characteristics of network structure of area tourist flows based on the tourism routine. *Human Geography*, 27(4), 131-136.
6. Wang, Y.M., Ma, Y.F., Wang, M.X. (2012). Network structure of multicity inbound tourists to China. *Process in Geography*, 31(4), 518-526.
7. Fu, L.L., Lv, B.F., Pei, R.M. (2009). Research on the user participation in relation-based virtual communities. *Economic Management Journal*, 31(5), 134-139.
8. Nair, P.S., Sarasamma, S.T. (2007). Data mining through fuzzy social network analysis. *Proceedings of the 26th Annual Meeting of the North American Fuzzy Information Processing Society*, San Diego, CA. 251-255.
9. Ignjatovic, J., Ciric, M., Bogdanovic, S. (2010). On the greatest solutions to weakly linear systems of fuzzy relation inequalities and equations. *Fuzzy Sets and Systems*, 161(24), 3081-3113.
10. Tseng, M.L. (2010). Implementation and performance evaluation using the fuzzy network balanced scorecard. *Computers & Education*, 55, 188-201.
11. Lu, Z., Lu, J., Zhang, C. (2002). Website development and evaluation in the Chinese tourism industry. *Networks and Communication Studies*, 16(3-4), 191-208.
12. Lu, J., Tang, S., McCullough, G. (2001). An assessment for internet-based electronic commerce development in businesses of New Zealand. *Electronic Markets*, 11(2), 107-115.