A comparative study of information diffusion in weblogs and microblogs based on social network analysis*

Yang ZHANG† & Wanyang LING

School of Information Management, Sun Yat-sen University, Guangzhou 510006, China

Abstract

Purpose: This paper intends to explore a quantitative method for investigating the characteristics of information diffusion through social media like weblogs and microblogs. By using the social network analysis methods, we attempt to analyze the different characteristics of information diffusion in weblogs and microblogs as well as the possible reasons of these differences.

Design/methodology/approach: Using the social network analysis methods, this paper carries out an empirical study by taking the Chinese weblogs and microblogs in the field of Library and Information Science (LIS) as the research sample and employing measures such as network density, core/peripheral structure and centrality.

Findings: Firstly, both bloggers and microbloggers maintain weak ties, and both of their social networks display a small-world effect. Secondly, compared with weblog users, microblog users are more interconnected, more equal and more capable of developing relationships with people outside their own social networks. Thirdly, the microblogging social network is more conducive to information diffusion than the blogging network, because of their differences in functions and the information flow mechanism. Finally, the communication mode emerged with microblogging, with the characteristics of micro-content, multi-channel information dissemination, dense and decentralized social network and content aggregation, will be one of the trends in the development of the information exchange platform in the future.

Research limitations: The sample size needs to be increased so that samples are more representative. Errors may exist during the data collection. Moreover, the individual-level characteristics of the samples as well as the types of information exchanged need to be further studied.

* This work is supported by Sun Yat-sen University Cultivation Fund for Young Teachers (Grant No.: 20000-3161102) and the National Social Science Fundation of China (Grant No.: 08CTQ015).
† Corresponding author: Yang Zhang (E-mail: zhyang2@mail.sysu.edu.cn).
Practical implications: This preliminary study explores the characteristics of information diffusion in the network environment and verifies the feasibility of conducting a quantitative analysis of information diffusion through social media. In addition, it provides insight into the characteristics of information diffusion in weblogs and microblogs and the possible reasons of these differences.

Originality/value: We have analyzed the characteristics of information diffusion in weblogs and microblogs by using the social network analysis methods. This research will be useful for a quantitative analysis of the underlying mechanisms of information flow through social media in the network environment.

Keywords Weblog; Microblog; Information diffusion; Social network analysis (SNA); Library and information science (LIS)

1 Introduction

The emerging technologies surrounding social media such as weblogging and microblogging have led to a revolution in the way people communicate with each other in the networked environment and also elicited widespread interest of researchers in weblogs and microblogs. Different from weblogging which provides a complex user interface, focuses on in-depth discussion of a topic, and features a one-way communication, microblogging is attracting more and more users with such characteristics as simple user interface, instant publishing of content and interactive communication with others in the online community[1].

Weblogs have come into being for more than 10 years and have been comprehensively studied. Researchers both from China and abroad studied the mechanism and the mode of information diffusion in blogging systems. By comparison, microblogging is still a relatively new research area and current researches focus on information dissemination mechanism[2], applications of microblogs[3,4], the extended functions[5], user characteristics[6], network characteristics[7] and content features[8], etc. In terms of information diffusion in microblogging systems, researchers studied the advantage of microblogs in promoting information dissemination[9,10], revealed the information flow mechanism and characteristics[11] and explored the interactive structural patterns of microblogs[12].

Social network analysis (SNA) is a series of theories and techniques to examine the structure of social relationships and attributes, as well as qualities of these relationships[13]. Researchers used the social network analysis techniques to describe the structure of group relations so as to analyze the impact of group relationships on individual members of the group[14]. They have attempted to use the SNA techniques to analyze the characteristics of information exchange in blogging and microblogging systems. For example, it was observed that the blogging network has
A comparative study of information diffusion in weblogs and microblogs based on social network analysis

Yang ZHANG & Wanyang LING

National Science Library, Chinese Academy of Sciences

http://www.chinalibraries.net

With respect to the comparative study of blogging and microblogging, there are some preliminary researches published, such as the study of their similarities and differences\cite{18}, their merits and demerits and their respective development trends\cite{19}. There are relatively rare research results published on the comparative analysis of the information exchange characteristics of blogging and microblogging using the social network analysis techniques.

This paper will explore a quantitative method for analyzing characteristics of information diffusion through social media. We will use social network analysis techniques and take the Chinese blogs and microblogs in the field of library and information science (LIS) as the research samples to study the characteristics of information diffusion in blogs and microblogs as well as to analyze the possible reasons of the differences.

2 Methodology

2.1 Sample collection

We selected LIS blogs and microblogs as the research samples. To collect the sample data, the following steps have been taken. First of all, by making reference to the list of the most valuable blogs in 2008 announced in Journal of Academic Libraries\cite{20}, we selected the top three blogs in library science as well as the most outstanding blog in archival science as the starting nodes of sampling. Using snowball sampling technique we chose the four blogs as the seed user set. Because blogs allow a user to “follow” updates from other bloggers who are added as “friends”, we then expand from this seed set based on their “friend” links. This process was repeated in turn for each of the contacts (i.e. friends) and again for each friend of these contacts. Finally, we checked whether or not these bloggers had microblog accounts and we selected only bloggers who had microblog accounts, and then we got the microblog sample for our research.
Because of the instability of the network and the different quality of data collected from the Internet, the sample data need to be filtered to ensure quality. During the process, we paid more attention to blogs. The following steps were taken:

- Exclude blogs that were not in the LIS field.
- Exclude blogs that paused or stopped updating content.
- Exclude duplicate blogs and dead links.
- Select microblog accounts registered at Sina Weibo. When determining the final sample list, we excluded those blogs which did not have corresponding microblog accounts or we could not get the corresponding microblog accounts.

Sina Weibo is currently the largest microblogging service provider in China. As Sina Weibo allows users to “follow” only registered users of Sina Weibo, we selected Sina Weibo as the source to collect more data.

After completing the above steps, we finally selected 34 blogs and their corresponding microblogs as the samples. We coded the sample blogs from A1 to A34 and the microblogs from B1 to B34. We created tables to reflect relationships between individual blogs and those of individual microblogs, respectively, based on the “friends” link or the “follower” relationships between them. Table 1 shows the relationships between blogs, in which “0” indicates that the two bloggers are “followers” of each other and “1” means they are not linked with each other. To ensure reliability of the data, we collected all data on the same day.

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>A7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

2.2 Statistical indicators of social networks

Social network indicators such as network density, core/peripheral structure, network centrality, and tie strength will be used to analyze the characteristics of information diffusion in blogging and microblogging networks.
Network density measures closeness between network nodes\(^{[21]}\). It can be calculated as the ratio of the existing connections over all possible connections in the network\(^{[22]}\). The density formula is

\[
D(n) = \frac{L}{n(n-1)}
\]

where \(n\) is the number of nodes and \(L\) is the number of lines present.

Core/periphery structure is used to indicate the positions occupied by individual actors (nodes) in the network. Core nodes and nodes in the periphery will be identified and their characteristics will be analyzed.

Network centrality is used to describe the locations of individual nodes in terms of their positions close to the “center” of the action in a network. It is a quantitative indicator to measure the power of actors, aiming at finding the core actors in the network. Network centrality mainly involves three kinds of measures: Degree centrality, betweenness centrality and closeness centrality\(^{[23]}\).

Weak ties are loose connections between individuals who seldom interact but may provide useful information or new perspectives for one another, such as alumni and neighbors, etc.

Granovetter put forward “the strength of weak ties” theory in 1973\(^{[24]}\). He argued that weak ties are more powerful than strong ties because information was far more likely to be “diffused” through weaker ties. Therefore, weak ties can be used to reflect the strength of connections between the individuals in and outside a community. The number and length of weak ties can be used to evaluate a social network in terms of the ability to connect to an external network. In addition, this paper also examines whether there exists a “small-world effect” in the blogging and microblogging network.

3 Result and analysis

3.1 Network density

In this study, UCINET 6.232 is used for data processing and analysis. Table 2 shows the statistics of the blog and microblog samples. The theoretical value of the density of a network is in the interval \([0, 1]\), but Mayhew and Levinger indicated that the actual social network density is generally not more than 0.5\(^{[25]}\). The data in Table 2 is consistent with their conclusion. Comparatively speaking, the microblogging network has a higher density.
social network density is significantly higher than that of the blogging social network, which means microblog users are more closely connected than blog users.

### 3.2 Core/peripheral structure

It can be clearly observed from Fig. 1 that the vast majority of individuals contact more or less with other individuals directly or indirectly, except for a few individuals who drift away or remain at the periphery of the contact circle. By comparison, groups of nodes in microblogging network are more densely connected than in blogging network. The number of connections in microblogging network (Fig. 1(b)) is 278 in total and every node has 8.18 links on average, whereas every node has only 3.02 links in the blogging network (Fig. 1(a)). This is consistent with the result of our analysis of network density.

In addition, the core of the blogging network consists of 8 actors and that of the microblogging network 14 actors. The former takes up about 20% of the total number of the blog samples while the latter about 40% of the microblog ones. This indicates that the microblogging social network displays a clearer tendency towards “decentralization”, in other words, microblog users are more equal than blog users.

### 3.3 Network centrality

#### 3.3.1 Degree centrality

Degree centrality measures how many connections each node has and it can be divided into out-degree and in-degree centrality, or absolute centrality and relative centrality[26]. If a node has a high degree centrality, it means that the node is in the core of the network and it has direct contact with many other nodes. As a result, it can produce a high impact on the network[23], therefore, the degree centrality is generally used to identify the influential or popular nodes of a network.
A comparative study of information diffusion in weblogs and microblogs based on social network analysis

The values of each degree centrality measure and corresponding rankings of blogs and microblogs are summarized in Tables 3–6.

**Table 3  Centrality of the partial LIS blogging network**

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Node</th>
<th>Degree</th>
<th>NrmDegree</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A3</td>
<td>16.000</td>
<td>48.485</td>
<td>0.091</td>
</tr>
<tr>
<td>2</td>
<td>A7</td>
<td>16.000</td>
<td>48.485</td>
<td>0.091</td>
</tr>
<tr>
<td>3</td>
<td>A6</td>
<td>14.000</td>
<td>42.424</td>
<td>0.080</td>
</tr>
<tr>
<td>4</td>
<td>A4</td>
<td>12.000</td>
<td>30.303</td>
<td>0.068</td>
</tr>
<tr>
<td>5</td>
<td>A19</td>
<td>10.000</td>
<td>27.273</td>
<td>0.057</td>
</tr>
</tbody>
</table>

Note: Network centralization = 34.85%. NrmDegree refers to the normalized degree.

**Table 4  In- and out-degree of the partial LIS blogging network**

<table>
<thead>
<tr>
<th>Node</th>
<th>OutDegree</th>
<th>InDegree</th>
<th>NrmOutDeg</th>
<th>NrmInDeg</th>
</tr>
</thead>
<tbody>
<tr>
<td>A7</td>
<td>12.000</td>
<td>6.000</td>
<td>36.364</td>
<td>18.182</td>
</tr>
<tr>
<td>A4</td>
<td>11.000</td>
<td>3.000</td>
<td>33.333</td>
<td>9.091</td>
</tr>
<tr>
<td>A6</td>
<td>9.000</td>
<td>11.000</td>
<td>27.273</td>
<td>33.333</td>
</tr>
<tr>
<td>A21</td>
<td>8.000</td>
<td>1.000</td>
<td>24.242</td>
<td>3.030</td>
</tr>
<tr>
<td>A11</td>
<td>8.000</td>
<td>4.000</td>
<td>24.242</td>
<td>12.121</td>
</tr>
<tr>
<td>A19</td>
<td>7.000</td>
<td>1.000</td>
<td>21.212</td>
<td>12.121</td>
</tr>
<tr>
<td>A32</td>
<td>7.000</td>
<td>1.000</td>
<td>21.212</td>
<td>3.030</td>
</tr>
<tr>
<td>A31</td>
<td>7.000</td>
<td>1.000</td>
<td>21.212</td>
<td>3.030</td>
</tr>
</tbody>
</table>

Note: Network centralization: Out-degree = 28.007%; in-degree = 31.129%. NrmOutDeg refers to the normalized out-degree, and NrmInDeg, the normalized in-degree.

**Table 5  Centrality of the partial LIS microblogging network**

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Microblog</th>
<th>Degree</th>
<th>NrmDegree</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B1</td>
<td>22.000</td>
<td>66.667</td>
<td>0.061</td>
</tr>
<tr>
<td>2</td>
<td>B6</td>
<td>22.000</td>
<td>66.667</td>
<td>0.061</td>
</tr>
<tr>
<td>3</td>
<td>B18</td>
<td>21.000</td>
<td>63.636</td>
<td>0.058</td>
</tr>
<tr>
<td>4</td>
<td>B27</td>
<td>20.000</td>
<td>60.606</td>
<td>0.056</td>
</tr>
<tr>
<td>5</td>
<td>B3</td>
<td>19.000</td>
<td>57.576</td>
<td>0.053</td>
</tr>
<tr>
<td>6</td>
<td>B17</td>
<td>19.000</td>
<td>57.576</td>
<td>0.053</td>
</tr>
</tbody>
</table>

Note: Network centralization = 36.74%. NrmDegree refers to the partial normalized degree.

**Table 6  In- and out-degree of the partial LIS microblogging network**

<table>
<thead>
<tr>
<th>Node</th>
<th>OutDegree</th>
<th>InDegree</th>
<th>NrmOutDeg</th>
<th>NrmInDeg</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>20.000</td>
<td>17.000</td>
<td>60.606</td>
<td>51.515</td>
</tr>
<tr>
<td>B27</td>
<td>20.000</td>
<td>13.000</td>
<td>60.606</td>
<td>39.394</td>
</tr>
<tr>
<td>B6</td>
<td>20.000</td>
<td>15.000</td>
<td>60.606</td>
<td>45.455</td>
</tr>
<tr>
<td>B18</td>
<td>20.000</td>
<td>15.000</td>
<td>60.606</td>
<td>45.455</td>
</tr>
<tr>
<td>B17</td>
<td>18.000</td>
<td>15.000</td>
<td>54.545</td>
<td>45.455</td>
</tr>
<tr>
<td>B21</td>
<td>16.000</td>
<td>14.000</td>
<td>48.485</td>
<td>42.424</td>
</tr>
<tr>
<td>B24</td>
<td>15.000</td>
<td>14.000</td>
<td>45.455</td>
<td>42.424</td>
</tr>
<tr>
<td>B26</td>
<td>15.000</td>
<td>11.000</td>
<td>45.455</td>
<td>33.333</td>
</tr>
<tr>
<td>B3</td>
<td>14.000</td>
<td>17.000</td>
<td>42.424</td>
<td>51.515</td>
</tr>
</tbody>
</table>

Note: Network centralization: Out-degree = 36.915%; in-degree = 27.548%. NrmOutDeg refers to the normalized out-degree, and NrmInDeg, the normalized in-degree.
Taking the entire network as a whole, the centralization value of the blogging and microblogging network was 34.85 and 36.74, respectively. There is not much difference between the two networks in network centralization. This means the concentration distribution of the two networks is similar on the whole. In-degree and out-degree centralities reflect the number of incoming and outgoing directed ties for a given node. Nodes with high in-degree centrality have many incoming links\[27\]. The in-degree centrality of the blogging network (31.129\%) is higher than that of the microblogging network (27.548\%). This shows that more nodes have many incoming links in the blogging network and they are regarded as opinion leaders in the network. This is in line with the prior research findings that there is a high concentration of opinion leaders in the blogging network\[5\].

As far as the in- or out-degree centrality is concerned, individual nodes in the blogging network all have smaller values than those in the microblogging network in our analyses. This implies that individuals in the microblogging network are more interconnected than those in the blogging network, and the microblogging network has a higher density. This is in consistent with our previous analysis in network density and core/peripheral network structure.

After comparing the results of the centrality values, we found that some individuals had both high blog and microblog rankings such as A1/B1, A3/B3, A6/B6 and A8/B8\[5\]. This indicates that these users stay in the core position in both blogging and microblogging networks. They play a crucial role in promoting content contribution and propagation in the network. After analyzing the related information, we found that they are famous researchers and have made notable achievements in LIS field. They have drawn a great deal of public attention and the “Matthew effect” is more likely to occur. As a result, they will gain more and more attention and become the opinion leaders of the community. This finding is consistent with previous reports\[7,11\].

3.3.2 Betweenness centrality

Betweenness centrality quantifies the number of the shortest path between two other nodes\[28\]. It is a measure to indicate whether a person is an intermediary between two others that are not linked to each other. Individuals with high betweenness centrality will gain more opportunities to control the flow of information and also occupy a key position to manipulate the flow of information. Obvious differences have been observed between the blogging network and the microblogging network in betweenness centrality in our study.

\[5\] Please refer to their rankings in Tables 3–6. Because of page limitations, the tables only show part of their ranking lists.
First of all, the betweenness centrality of the microblogging network is 7.42% while that of the blogging network is 32.10%. Both of the two networks have a low value for betweenness centrality, but compared with the blogging network, the betweenness centrality of the microblogging network is even lower. This indicates that nodes in microblogging network depend less on intermediate nodes to help transmit information to other nodes.

Secondly, there is a significant gap between the two networks in betweenness centrality as illustrated in Tables 7 and 8.

Table 7  Betweenness centrality of the partial LIS blogging network

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Node</th>
<th>Betweenness</th>
<th>nBetweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A3</td>
<td>145.631</td>
<td>35.870</td>
</tr>
<tr>
<td>2</td>
<td>A13</td>
<td>78.000</td>
<td>19.212</td>
</tr>
<tr>
<td>3</td>
<td>A7</td>
<td>61.874</td>
<td>15.240</td>
</tr>
<tr>
<td>4</td>
<td>A15</td>
<td>55.000</td>
<td>13.547</td>
</tr>
<tr>
<td>5</td>
<td>A6</td>
<td>53.114</td>
<td>13.082</td>
</tr>
</tbody>
</table>

Note: Network centralization = 32.10%. nBetweenness refers to the normalized betweenness centrality.

Table 8  Betweenness centrality of the partial LIS microblogging network

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Node</th>
<th>Betweenness</th>
<th>nBetweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B1</td>
<td>37.481</td>
<td>8.616</td>
</tr>
<tr>
<td>2</td>
<td>B24</td>
<td>29.330</td>
<td>6.743</td>
</tr>
<tr>
<td>3</td>
<td>B6</td>
<td>26.131</td>
<td>6.007</td>
</tr>
<tr>
<td>4</td>
<td>B3</td>
<td>17.622</td>
<td>4.051</td>
</tr>
<tr>
<td>5</td>
<td>B4</td>
<td>15.451</td>
<td>3.552</td>
</tr>
</tbody>
</table>

Note: Network centralization = 7.42%. nBetweenness refers to the normalized betweenness centrality.

In absolute terms, the blogging network has a higher value in betweenness centrality. In relative terms, there are more obvious differences between nodes of the blogging network and those in the microblogging network in terms of the value of the betweenness centrality. Therefore, a few nodes in the blogging network act as intermediaries and they play a critical role in maintaining the relationships among nodes in the entire network. If we have these intermediary nodes removed, the network will be more decentralized and connected locally.

However, in the microblogging network, there are not such intermediary nodes. This also implies that nodes in the blogging network depend more on intermediate nodes to help transmit information to other nodes, whereas nodes in the microblogging network can communicate with each other more directly, resulting in a highly interactive network.

3.3.3  Closeness centrality

Closeness centrality is a measure that indicates how close a node is to all the other nodes in a network[29] and it also describes the extent of influence of a node on the...
network. A node that has high closeness centrality is likely to be less dependent on others in receiving or sending information. Closeness centrality can be measured by the sum of geodesic distances from a given node to all others, where geodesic distance refers to the length of the shortest path between two points\cite{30}.

The values of closeness centrality and corresponding rankings of nodes in blogging and microblogging network are summarized in Tables 9 and 10.

Table 9 Closeness centrality of the partial LIS blogging network

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Node</th>
<th>Farness</th>
<th>nCloseness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A3</td>
<td>44.000</td>
<td>65.909</td>
</tr>
<tr>
<td>2</td>
<td>A7</td>
<td>47.000</td>
<td>61.702</td>
</tr>
<tr>
<td>3</td>
<td>A6</td>
<td>49.000</td>
<td>59.184</td>
</tr>
<tr>
<td>4</td>
<td>A4</td>
<td>54.000</td>
<td>53.704</td>
</tr>
<tr>
<td>5</td>
<td>A19</td>
<td>56.000</td>
<td>51.786</td>
</tr>
</tbody>
</table>

Note: Network centralization = 44.46%. The farness of a node is the sum of geodesic distance between this node and all other nodes in the network. nCloseness refers to the normalized closeness centrality.

Table 10 Closeness centrality of the partial LIS microblogging network

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Node</th>
<th>Farness</th>
<th>nCloseness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B1</td>
<td>30.000</td>
<td>86.667</td>
</tr>
<tr>
<td>2</td>
<td>B6</td>
<td>30.000</td>
<td>86.667</td>
</tr>
<tr>
<td>3</td>
<td>B18</td>
<td>31.000</td>
<td>83.871</td>
</tr>
<tr>
<td>4</td>
<td>B27</td>
<td>32.000</td>
<td>81.250</td>
</tr>
<tr>
<td>5</td>
<td>B17</td>
<td>33.000</td>
<td>78.788</td>
</tr>
</tbody>
</table>

Note: Network centralization = 41.34%. The farness and nCloseness are the same mentioned in Table 9.

Table 11 shows that the value between the maximum and the minimum of closeness centrality of the two networks is close (41.3 and 42.6, respectively), but the closeness centrality value of each node in the blogging network is smaller than that in the microblogging network and the average difference is about 20. This indicates that the average number of intermediaries is relatively less in the microblogging network than in the blogging network and nodes in the microblogging network are more interconnected. Therefore, the microblogging network is denser than the blogging network on the whole, which confirms that nodes of the microblogging network depend less on intermediaries to send information than those of the blogging network.

Table 11 Comparison of closeness centrality of blogging network and microblogging network

<table>
<thead>
<tr>
<th>Closeness centrality</th>
<th>Blog</th>
<th>Microblog</th>
<th>D-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>65.9</td>
<td>86.7</td>
<td>−20.8</td>
</tr>
<tr>
<td>Minimum</td>
<td>24.6</td>
<td>44.1</td>
<td>−19.5</td>
</tr>
<tr>
<td>Span</td>
<td>41.3</td>
<td>42.6</td>
<td>−1.3</td>
</tr>
</tbody>
</table>
3.4 The existence of weak ties phenomenon

Inbound links are links pointing to one website from pages of various other external websites and outbound links are those pointing to other websites from this website. Based on the following/follower relationships between the users, we counted the outbound links of each sample blog pointing to other sample blogs and their outbound links pointing to other websites separately. For convenience, we call the outbound links of each sample blog pointing to the other sample blogs “internal links” and the outbound links of each sample blog pointing to other websites “external links”.

We also counted the internal links and external links of sample microblogs. Table 12 shows the results. We found that external links are much more than internal links. Those external links are all weak ties, which indicate the existence of a large number of weak ties both in blogging and microblogging communities.

<table>
<thead>
<tr>
<th>Table 12</th>
<th>Links statistics of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of sample</td>
</tr>
<tr>
<td>Blog</td>
<td>34</td>
</tr>
<tr>
<td>Microblog</td>
<td>34</td>
</tr>
</tbody>
</table>

We observe that microbloggers have more weak ties than bloggers. There are several possible reasons. One is that people use blogs mainly to express their opinions as well as record their daily activities while people use microblogs mainly to exchange views and share resources. Microblogs encourage users to connect and maintain relationships with weak ties to obtain more information. In addition, both the amount of information exchanged and the rate of information flow are different, which may also affect the number of outbound links.

3.5 Small-world effect

In social network analysis, it can test the small-world effect based on the shortest path of the network through which one can reach any other actors of the network. According to previous studies, the analysis of small-world effect has become a hotspot in SNA related studies[31].

In this paper, after removing the isolated points, the shortest path of the blogging and microblogging network is about one on average. In other words, a node in the blogging and microblogging network can send or receive messages to any other node in the network through one intermediary on average. This indicates that the two networks exhibit the small-world effect, which is consistent with prior researches[32,33].
4 Discussion

By analyzing blog and microblog samples using social network analysis, we can detect different characteristics of information diffusion in the two networks:

- In terms of the network density, the microblogging network is much denser than the blogging network and information diffusion in the microblogging network is more efficient. Both the blogging network and the microblogging network exhibit a small-world effect, which indicates that interactive communication occurs frequently in the two networks. However, from the view of network density and centrality, the microblogging network is denser. Individuals in the microblogging network are more closely connected and rely less on intermediaries, which can reduce the information erosion during the information transmission process and the impact of noise on information fidelity. Thus the microblogging community is more efficient in interactive communications. This is probably due to the emphasis of the microblogging platform in encouraging users to establish relationships. For example, the function of “one click to follow” provides a simple and convenient way for microblog users to establish relationships, especially having fixed contacts, in order to facilitate the communication. In addition, the limit to the number of words for each post makes it possible for one microblogger to have a lot of followers or to follow a lot of microbloggers. These design differences potentially are creating a dense network.

- With respect to network structure, individuals in the microblogging community are more equal than those in the blogging community and the microblogging network is more decentralized and connected locally. The results of our core/peripheral structure and centrality analyses show that there are obvious “opinion leaders” in the blogging network. The greater power these opinion leaders have, the smaller influence the users with a peripheral position in the user community, which often leads to the “Matthew effect” and the “spiral of silence” effect. This will result in the phenomenon that opinion leaders will shape the views of those around them and even “control” public opinions, which will hinder the communication process. By contrast, there is little core in the microblogging network and individuals in the network are more equal, which indicates a flatter social structure.

\[^{2}\] Spiral of silence theory describes the process by which one opinion becomes dominant as those who perceive their opinion to be in the minority do not speak up because they fear isolation from society. In a social network, the users in a peripheral position will remain silent when they feel the opinion leaders are powerful and influential.
The reasons for a decentralized microblogging network are probably the low entry threshold to use microblogs or publish messages so that the user can have a more equal right to speak and have the same opportunities to be followed by other microblog users. It also weakens the impact of the hub users and result in a flatter network structure. The content of a microblog is focused more on the daily lives, entertainment and news and advocates open discussion and interactive communication while the content of a blog is focused more on the intellectual and autobiographical memory. A blog is used as a record to inform and update others of bloggers’ activities, a tool to communicate internally within a certain group and a knowledge management platform. The different positioning of the two social media in information content may also be an important cause of the different features of the two networks.

- In the light of concentricity, the blogging network is more cohesive. The microblog users are more capable of developing relationships with people outside their own social networks. Weak ties contribute a lot in communication that strong ties cannot match. Weak ties play an important role in obtaining the unrelated information of community topics, which may inspire innovation and develop new ideas. It is shown that actors establish weak ties in both blogging and microblogging networks. By comparison, microbloggers have more weak ties. Microblog users may create links to various kinds of content and follow different types of users out of their needs for a variety of information. This results in a large number of weak ties and a decentralized network structure. On the other hand, blog users, especially when bloggers are from one specific academic field, pay more attention to other users in the same field. As a result, the users that they follow and the topics they communicate with each other are more focused on the users who work in the same field and the topics of this field, respectively. The bloggers’ online community is more cohesive and these users are closely connected. Compared with of the common users, famous scholars are usually in the center of public attention because of their reputation. These famous researchers will be more likely to become the star nodes surrounded by other nodes. This also proves the high concentration of opinion leaders, which is a prominent feature of the blogging network.

In short, the microblogging community is more conducive to the interactive communication than the blogging community. This is closely related to the differences of these two social media in the positioning of their functions and their different communication mechanisms. In terms of the positioning of functions, blogging emphasizes a record to inform and update others of a user’s activities and knowledge management, while microblogging focuses on talking about daily
activities and sharing information. In communication mechanism, compared with blogging, microblogging provides a set of unique mechanisms that mainly include: Following other users, information retrieving, linking, reminder, information disseminating, real-time communicating, etc., which ensures timely and smooth communication between users and encourages interaction between users as well. So the communication mode emerged with microblogging, with characteristics of micro-content, multi-channel information dissemination, dense and decentralized network and content aggregation, will be the trend in the development of the information communication platform in the network environment.

5 Conclusion

The characteristics of information diffusion through social media and the effect of the communication are mainly determined by the information flow mechanism and the positioning of functions of the media. This can be reflected in the properties of different networks as shown in our study. Our preliminary study also proves that using the social network analysis techniques to investigate the characteristics of information diffusion in the user community network is one effective way to conduct a quantitative analysis of the community. From the perspective of the future trends, the communication mode emerged with microblogging will prevail for a longer period.

It should be noted that our sample size is small because we selected the Chinese LIS blogs and the corresponding microblogs as samples. Making reference to the related studies, our sample can basically reflect the real situation of LIS blogs and microblogs in China. If we select blogs and microblogs in other fields, we might need a much larger sample. Nevertheless, how social media will affect the information flow mechanism, which indicators can reflect what characteristics of the information flow process and how to use those quantitative indicators to evaluate the effect of information diffusion, these are some questions that we will delve into in our future studies.

References

A comparative study of information diffusion in weblogs and microblogs based on social network analysis

Contents Index to Volume 5

Perspective
Transforming the academic library for the future

Research Papers
An empirical research on strategic planning in public libraries of Mainland China

Information behaviors of Chinese K-12 physical education teachers:
A survey study

Kewen WU, Qinghua ZHU, Julita VASSILEVA & Yuxiang ZHAO 2012,5(1):34–50

Investigating the relationships between facets of work task and selection and query-related behavior

A preliminary study on exploratory search behavior of undergraduate students in China
Yunqiu ZHANG, Wenxiu AN & Jia FENG 2012,5(1):70–84

Defining an open access resource strategy for research libraries:
Part II—Challenges and responses

Academic institutional repositories in China: A survey of CALIS member libraries
Yongchao ZHAO, Xiaoxia YAO & Chengfu WEI 2012,5(2):18–32

Cross-efficiency DEA model-based evaluation of allocative efficiency of rural information resources in Jiangsu Province, China

Integrating innovative library e-reserves service into teaching and learning at Shanghai Jiao Tong University
Xiaohua SHI, Xiaowei YU & Wei PAN 2012,5(2):45–58

Information literacy course design based on student survey:
The practice of subject librarians at NSL, CAS
Ming WU, Li WANG & Yanli LIU 2012,5(2):59–73
A study of the effects of students’ cognitive styles on their learning behaviors of online database search strategies
Ying FENG, Liren GAN & Deyi QIAO 2012,5(2):74–86

Defining an open access resource strategy for research libraries: Part III—The strategies and practices of National Science Library
Xiaolin ZHANG, Xiwen LIU, Lin LI, Yan ZENG & Li-Ping KU 2012,5(3):1–11

A view on big data and its relation to Informetrics
Ronald ROUSSEAU 2012,5(3):12–26

Person-specific named entity recognition using SVM with rich feature sets
Hui NIE 2012,5(3):27–46

A comparison of mapping strategies from DDC to CLC

A study of information exchange through social networks in rural China
Ya LIU 2012,5(3):62–75

Is it time for wider acceptance of e-textbooks? An examination of student reactions to e-textbooks
Ziming LIU 2012,5(3):76–87

A knowledge discovery method based on analysis of multiple co-occurrence relationships in collections of journal papers

Designing an automated FAQ answering system for farmers based on hybrid strategies

Factors influencing knowledge contribution: An empirical investigation of social networking website users
Shengli DENG, Ting ZHOU & Min ZHANG 2012,5(4):37–50

A comparative study of information diffusion in weblogs and microblogs based on social network analysis

An index model for measuring microblog users’ influence
Fuyong YUAN, Jing FENG & Qianqian FU 2012,5(4):67–76

A method for improving the accuracy of automatic indexing of Chinese-English mixed documents
Yan ZHAO & Hui SHI 2012,5(4):77–92

Research Progress
Digital reading: An overview

National Science Library, Chinese Academy of Sciences
http://www.chinalibraries.net
Chinese Journal of Library and Information Science

(中文刊名:《中国文献情报》)

1 基本信息

期刊名称: Chinese Journal of Library and Information Science (CJLIS)
创刊时间: 2008年
主管单位: 中国科学院
主办单位: 中国科学院国家科学图书馆（即中国科学院文献情报中心）
出版单位: 《中国文献情报》（英文刊）编辑部
主 编: 张晓林教授（中国科学院国家科学图书馆馆长）
刊 号: ISSN 1674-3393; CN 115670/G2
每期定价: 国内: 200元/期（800元/年、卷）; 国际: 199美元/年、卷（含邮资）

2 期刊定位与报道范围

国内图书情报领域唯一一份面向国内外公开征订与发行的学术性英文期刊，是国内图书、情报与档案学领域的国际化高端学术交流平台与工具。刊载内容涵盖图书馆学、情报学、档案学与新技术在图书馆中的应用，包括图书馆学情报学教育等相关研究领域的各种理论研究与技术实践。

3 论文类型与评审

CJLIS注重论文的前沿性、学术性与实用性。秉持理论与实践并重，采用同行匿名评议。其编委会由海内外23位知名专家与学者组成，其中，境外编委约占2/3。CJLIS主要发表原创性研究，重点为实证性研究论文，包括研究综述与实践研究等学术论文。

4 读者对象

CJLIS是国内外图书情报机构从业者、广大图情研究管理人员、教师、硕博研究生和本科生的重要参考工具，是了解国内外文献情报事业发展现状和最新研究成果的重要窗口。于每季末25日出版。年出版4期。如欲订阅，可通过以下方式办理:

1. 邮局订阅: 全国各地邮局均可办理
   - 刊 号: ISSN 1674-3393
   - CN 11-5670/G2
   - 邮发代号: 82-563

2. 国际代理
   Ebsco International, Inc. (USA)

3. 《中国文献情报（英）》编辑部
   - 地址: 北京海淀区中关村北四环西路33号5D07房间
   - 邮编: 10019
   - 电话: 010-82624454; 010-82626611转6628
   - 传真: 010-82624454
   - 联系人: 唐欣
   - E-mail: chinalibraries@mail.las.ac.cn

4. 北京中科期刊出版有限公司
   - 地址: 北京海淀区德胜门外大街83号德胜国际中心B座3层301室
   - 邮编: 100088
   - 电话: 010-59790736 / 8028
   - 传真: 010-59790736-8092
   - 联系人: 孙志蓬
   - E-mail: sunzhipeng@mail.sciencep.com

5. 国科国家科学图书馆编辑出版中心发行组
   - 地址: 北京海淀区中关村北四环西路33号
   - 邮编: 10019
   - 电话: 010-82622346
   - 传真: 010-82626611转6155
   - 联系人: 苗志刚
   - E-mail: miaozg@mail.las.ac.cn

http://www.chinalibraries.net
Chinese Journal of Library and Information Science
(CJLIS, Chinese Title:《中国文献情报》)

ISSN 1674-3393, CN 11570/G2;  Price: $US 199/Volume

Chinese Journal of Library and Information Science (CJLIS) is the first and only English-language academic journal in the field of Library and Information Science (LIS) that has been published in Mainland China so far. It aims to establish a platform for students, researchers and professionals in LIS and archive science to engage in intellectual dialogs and to share best professional practices of their experience.

CJLIS adopts a double-blind peer-review process for each published article. The Editorial Board of CJLIS is composed of well-known scholars both from national and international LIS fields. The main aim of CJLIS is to serve librarians, information professionals, researchers and practitioners in the LIS field for their various purposes and keep them well informed of the latest development of LIS field both in China and the world.

CJLIS mainly publishes original research articles, including research papers, library practice papers, progress reports, research reviews and perspectives. For subscription, you can use the following channels:

1) Via EBSCO International, Inc (USA)

2) To order at any post office inside China, with a distribution code 82-563.

3) To contact with the Editorial Office
   Address: No.33 Beisihuan Xilu, Zhongguancun, Haidian District, Beijing 100190, P.R. China
   Tel: 86-10-82624454 or 86-10-82626611 ext. 6628
   Fax: 86-10-82624454
   Email: chinalibraries@mail.las.ac.cn
   Website: http://www.chinalibraries.net

4) Via Beijing China Science Journal Publishing Co., Ltd.
   Address: Room 301, Building B, Desheng International Center, No. 83 Deshengmenwai Street, Beijing 100088, P.R. China
   Tel: 86-10-59790736 ext. 8028
   Fax: 86-10-59790736 ext. 8092
   E-mail: sunzhipeng@mail.sciencep.com
Submission Guidelines

Aims
Chinese Journal of Library and Information Science (CJLIS), being sponsored by the Chinese Academy of Sciences (CAS) and published quarterly by the National Science Library of CAS, is a scholarly journal in the field of library and information science (LIS). Its aim is to provide an international communication link between researchers, educators, administrators, and information professionals.

With the publication of the research results both from China and from other foreign countries, the Journal CJLIS strikes a balance between theory and practice. With its goal to provide an open forum for Chinese and international scholars in this field to exchange their research results, CJLIS also offers new possibilities in the advancement of Chinese library operations. The CJLIS tries to establish a platform for LIS students, researchers and library staff all over the world to engage in intellectual dialog and also to improve library services so as to promote even more quickened and substantial development of LIS in China.

Scope
Striving toward academic excellence, innovation, and practicality, the CJLIS mainly includes research papers both on the theoretical as well as on the practical fronts in all aspects of the field. More specifically, it includes but not limited to informatics, library management, information technology application, knowledge organization system, knowledge management, archives, permanent preservation of library resources, LIS education, and so on.

Referencing Process
Articles and papers covering the topics or themes mentioned above will be refereed through a double-blind peer review process.

Editorial Advisory Board
The Editorial Board is composed of the nationally and internationally well-known scholars and researchers in the LIS field and the high quality of this Journal is thus reasonably assured.

Manuscripts Categories
As the first English-language academic journal on LIS published in Mainland China, the CJLIS will take a proactive attitude to trace and report the prevailing hot issues in the field around the globe as well as the more serious scholarly communications. As such, the submitted manuscripts are classified into constant categories and unfixed categories. In the former category, research papers, library practice and progress reports are the essential components. In the latter, book reviews, biographical sketches, anecdotes, reminiscence of prominent librarians and brief communications will appear occasionally.

Research papers represent original research work or a comprehensive and in-depth analysis of a topic. More than 3,000 words are considered as a proper length for such manuscripts, with a structured abstract ca. 200 words.

Library practice covers the latest development and application in any segment of library field work and information service. The length of the manuscript is preferred to be more than 3,000 words, with a structured abstract ca. 200 words.

Progress reports reflect the projects result or research progress on the key topics of the library and information science. Submissions of articles in this section are expected to be comprehensive and analytical, which may deepen the understanding of the discussed issue and stimulate further researches on the topics, or give a new perspective on future technological applications. The manuscript length should be within 5,000 words, with a structured abstract ca. 200 words.

Manuscripts Requirements
All papers can be submitted either in English or in Chinese (or both) with a double line space. For the assurance that all the materials of the to-be-submitted are included, please check the following:

Title. Please give a brief biographical introduction to all contributing authors and their research background on a separated paper. For a better organization of the paper, please use the headings and subheadings.

Authors and affiliations. Please do not forget to write down the mailing address of each and every article contributors.

References. Be sure all the references used should be cited properly in both in-text and in bibliography. Particular attention should be paid to the proceedings. Do not forget adding the name(s) of editors of the compilation, as well as the name of the publishers. For the detailed information, please request a copy of Reference Citation Format.

Copyright
All submitted papers normally should not have been previously published nor be currently under consideration for publication elsewhere. For all the materials translated or obtained from other published resources, they should be properly acknowledged. All copyright problems should be cleared without any legal entanglements prior to the publication.

Notes for Intending Submissions
A guide for authors and other relevant information, including submitting papers online, is available at the website of the Editorial Office of the CJLIS (http://www.chinalibraries.net). For any questions, you can e-mail the Office or directly to:

Prof. ZHANG Xiaolin
Editor-in-Chief of CJLIS
The CJLIS Editorial Office
National Science Library, Chinese Academy of Sciences
No.33 Beishihu Xiliu, Zhongguancun, Haidian District,
Beijing 100093, P.R. China
Tel: 86-10-82624454 or 86-10-82626611 ext. 6628
Fax: 86-10-82624454
E-mail: chinalibraries@mail.las.ac.cn
Website: http://www.chinalibraries.net

Subscription
For single-copy subscription in China, RMB ¥ 200/issue. For subscription outside of China, US $ 199/Volume yearly (including air shipping)
Contents

■ Perspective
1 Transforming the academic library for the future
   Gary E. STRONG

■ Research Papers
9 A knowledge discovery method based on analysis of multiple co-occurrence relationships in collections of journal papers
   Hongshen PANG
21 Designing an automated FAQ answering system for farmers based on hybrid strategies
   Junliang ZHANG, Xuefang ZHU & Guang ZHU
37 Factors influencing knowledge contribution: An empirical investigation of social networking website users
   Shengli DENG, Ting ZHOU & Min ZHANG
51 A comparative study of information diffusion in weblogs and microblogs based on social network analysis
   Yang ZHANG & Wanyang LING
67 An index model for measuring microblog users’ influence
   Fuyong YUAN, Jing FENG & Qianqian FU
77 A method for improving the accuracy of automatic indexing of Chinese-English mixed documents
   Yan ZHAO & Hui SHI
93 Reference Citation Format

■ Content Indexes
94 Contents Index of Volumes 1–4
101 Contents Index of Volume 5

■ Enlightenment to Subscribe
103 Print Subscription