Visualization of the Chinese academic web based on social network analysis

Bo Yang and Zhihui Liu
National Science Library of the Chinese Academy of Sciences, Beijing, P. R. China
Graduate University of the Chinese Academy of Sciences, Beijing, P. R. China

Joseph A. Meloche
Department of Library and Information Studies, Graduate School of Education, University at Buffalo, NY, USA

Abstract.

The top 100 university websites in mainland China have been selected as samples to explore the link relationship between their websites using principal component, multidimensional scale and social network analysis methods, such as co-link, betweenness, and k-core. These relationships reflect the characteristics of comprehensive strength, influential power and the subject types of the universities. Macroscopically, websites are typically linked for their influence, while subject/regional relations are discovered through sub-network analysis at the micro-level. The micro-level link analysis shows that subject relation, as well as geographical relation, are significant motivators for collaboration between universities.

Keywords: academic websites; link analysis; link motivation; social network analysis

1. Introduction

Social network analysis (SNA) is an important research field in sociometrics. It studies the relationship between or among groups and individuals. SNA assumes that people or groups are connected together by social relationships, forming a social network as a representation of relationships among them. In this kind of network, the node is an individual or a group, and the edge or arc stands for the relationship between couples of nodes. The hypothesis of link analysis is similar to that of SNA. Websites/domains are interconnected by links, which form a network of social groups in cyberspace or web communities. Therefore, in most cases the theory and methods of SNA are applicable to link analysis, such as in-degree and out-degree-based centrality analysis, and betweenness-based influence analysis [1, 2], etc.

Correspondence to: Bo Yang, Room 614, 33# Beisihuan Xilu, Zhongguancun, Beijing, 100190, P. R. China.
Email: boyang@gmail.com
Link analysis is one of the more important methods used to study the relationships among nations, domains, websites and institutions. The conclusions drawn from link analysis may be compromised greatly by the complexity of linking motivation and the instability of link quality. Furthermore, link analysis methods used hitherto are simple and not powerful enough to reveal the hidden structure of the network completely. The question asked here is how to collect link data scientifically so as to establish a reliable link network. Research that takes a multiple-method approach to reveal rules of linking is posited here as a field of great importance in link analysis research.

This study will explore the validity of SNA for use in link analysis on a national network rather than on an international network as was done in a previous study [2]. In this study the top 100 universities in mainland China were selected according to the generally accepted university ranking list published by China Academy of Management Science in 2008 [3]. Based on the linkage relationships, the authors try to reveal the linking rules of university websites in cyberspace as the basic indicators of SNA, and hope that these rules will reflect their characteristics, including the comprehensive strength, influence and subjects of these universities. At the same time, the link rules will be discovered by link mining and thus provide the opportunity to improve the methodologies of link analysis and provide argument for application of link analysis [4].

2. Related work

The research on link analysis dates back to the popularity use of hyper-text technology in the 1990s and has boomed considerably with the increased presence of search engines. The great success of the page rank-based search engine Google makes the link analysis algorithms well known. The aims of link analysis algorithms are to improve the efficiency of information retrieval and to find core web resources by assessing their linkage relationship. In scientometrics, the research exploring the rules of scientific production and communication by linkage relation among web entities began with ‘webometrics’ proposed by Almind and Ingwersen in 1997 [5]. Most webometric researches focused on websites of universities, institutions and scientific journals, with a few exceptions which typically focus on commercial websites [6]. This type of research has aimed to mine the scientific regulations and processes which exist in the conduct of the scientific research, communication and journal publishing, which now extend to cyberspace.

After Ingwersen put forward the concept of the web impact factor (WIF) [7], many explorations in webometrics have been conducted to evaluate universities, institutions and scientific journals, based on the WIF indicator. Thelwall [8] analysed quantitatively the linkage relation of universities in the UK, and found the aggregating pattern of links among UK universities by multidimensional scale analysis. In Onyancha and Ocholla’s research [9], the WIF of 16 Kenyan universities and 21 South African universities was calculated. Their research shows that the websites of Kenyan universities are at an initial stage of construction, while South African universities have made remarkable progress in their web presence. Though not comparable to their counterparts in developed countries, can South African universities have their websites evaluated webometrically? Ortega and Aguillo [1] analysed the convergence of disciplines by the web link graph of 699 websites, which were selected from a population of 2390 websites of departments and institutions in 52 Spanish universities, and explored the web presence of disciplines as well as research fronts, etc. The link structure analysis of 527 universities of European Higher Education Area (EHEA) indicates [2] that this network is composed of well-defined national networks, in which each one is connected to another through high level national universities in each country and the German and British networks are dominant. Referring to the case studies on WIFs, Lu and Qiu [10] found that the correlation between JIFs (journal impact factors) and WIFs was statistically significant; Qiu et al. [11] explored the relationship between in-link counts and WIFs for Chinese academic websites, and demonstrated that the in-link count can be a better evaluation measure for university websites than WIF. However, in Thelwall and Tang’s study [12], no significant result is shown about the relationship between in-link count and research productivity of universities in mainland China.
There are also studies on visualization of web links. Larson [13] applied co-link analysis to the websites of earth science, and drew the multidimensional scale (MDS) map of the linkage relations. The aggregating pattern of links among UK university websites is also presented by MDS map in Thelwall’s study [8]. Leydesdorff and Vaughan [14] extended the traditional author citation analysis (ACA) to cyberspace. Based on the data collected from Google Scholar, they visualized the information about 24 authors’ co-occurrence by Pajek, and found two cliques interconnected tightly. Ortega et al. [2] analysed the national academic networks of three Nordic countries (Denmark, Finland and Sweden) by Ucinet, finding that the Swedish network is the principal Nordic sub-network and the Finnish one is slightly isolated from Europe, with the exception of the University of Helsinki.

Most of the studies mentioned above analyse the international/federal networks, which are comprised of multiple nations or regions. These studies have focused on the linkage relationship among nations or relatively independent regions, and the conclusions are about the closeness of relationships or the web presence of universities from different countries. The sharing of samples in these studies shows the high cohesion inside the national network and the low coupling among them, which facilitates the analysis on the relationships among sub-networks by social network analysis. However, there are few studies on sub-units (nodes or cliques, etc.) in any national network for the lack of apparent cohesion/coupling relationship among the nodes in national networks. Mining linking patterns from the nodes without apparent linking characteristics makes an even higher demand on data quality, analysis methods, etc.

3. The research questions

In this paper, we have selected as the sample the websites of the top 100 universities in mainland China. Through analysis of the linkage relationship among these websites by SNA, we aim to establish the potential ability of web link to reflect university influence and their interconnection. The specific research questions are listed below.

1. Is the WIF as measured simply by in-link count really more reliable and stable than classic WIF for academic evaluation? As web creation is incredibly easy, so the number of pages on websites probably varies from one site to another. It is, however, unlikely that the research productivity of the affiliated institution associated with the websites would vary. Consequently, it seems that the method applied in classic WIF is influenced greatly by the number of pages for websites and may not be objective enough.

2. Do SNA measures work well on visualization of the Chinese academic web? Although most of the selected samples are national key universities, instinctively we can classify them into three levels by their comprehensive strengths, such as general key, prestige and excellence. For example, Peking University and Tsinghua University are rated as excellent ones.

3. What are the characteristics of the relationships as measured by SNA? The selected universities are located in different provinces in China and their interest focuses on different subjects. Is SNA capable of distilling the subject/regional structures from national academic networks as well as from international academic networks [2]?

4. Method

The reliability of the results deduced from link analysis mainly depends on the following three factors:

1. Reasonable sample selection.
2. Reliable data collection methods.
3. Scientific and multiple methods of data analysis.
Bo Yang, Zhihui Liu and Joseph A. Meloche

The former two aim to guarantee the quality of basic research data, which is vital in link analysis. Under the circumstances of too much noisy information, it is critical for this study to adopt a range of scientific methods to harvest reliable link data based on a reasonable sample set. The latter is to guarantee the reliability of the research result.

4.1. Sample selection

The top 100 universities in mainland China are selected as an original sample set in this study [3]. The ranking list is considered as the most authoritative university rating in China and was also used in Qiu et al.’s study [11], which is mostly based on research ability and educational indicators of universities [3]. Some cases are excluded for the reasons of domain names and script parsing problems and the number of the final available samples is 93. For example, BUPT (Beijing University of Posts and Telecommunications) is removed for that many third level domain names as its departments are not under the official top domain of the university (bupt.edu.cn). In the case of various domain names caused by university mergers, the statistical data are combined to ensure the integrality of research data.

4.2. Data collection

In the webometric study, the strategies or tools used in data collection for link analysis can be divided into the following groups:

- using popular internet search engines, such as AltaVista, Google, etc. as data collection tools;
- combining third-party software and self-developed tools, such as Binker [2] and Offline Explorer plus webStat [15];
- using self-developed tools alone, such as CheckWeb or Mike Thelwall’s crawler [16, 17].

In our opinion, the performance of a link-harvesting tool can be evaluated based on the following criteria:

- effectiveness in harvesting links to key resources in the sample set;
- compatibility of data organization and classification with data analysis tools;
- ability to customize different data harvesting strategies, such as choosing data harvesting depth and scope according to different purposes and objects.

Only data harvest tools satisfying the requirements mentioned above are considered well-equipped and capable of harvesting reliable data. However, few of the existing link-harvesting tools, either commercial or shared software, can satisfy all the above requirements completely. A self-developed data collection system for link analysis, LinkDiscoverer, is applied as a solution to the problems and errors caused by the poor performance of data harvest tools [18]. Data collection strategies include crawling depth, filtering rules, time-out, DNS cache and other options that can be set in LinkDiscoverer according to the characteristics of each sample. Link data will be harvested on the basis of the collection strategies and classified online, and preliminary statistical data will be obtained.

4.3. Data analysis

In Otte and Rousseau’s article [19], the theoretic rationality of SNA is demonstrated in detail and the theories are confirmed by several case studies. The statistical indicators and SNA measures used in this study are listed below:

- Inlink: An in-link is a directed link and is also called a back-link. If a page on site A links to a page on site B, we can say the link is an in-link received by site B. Generally, greater in-link numbers demonstrate greater popularity of a site in the specific web space.
• WIF: The web impact factor is a traditional webometric indicator that stems from the concept of the JIF. Accordingly, it adopts a similar algorithm. It is mentioned by Ingwersen [7] that institutional WIFs should be approached with caution. Qiu et al.’s and Yang and Qin’s studies show that the in-link measured WIF is more reliable than the classic WIF algorithm [11, 18]. So, in this study, the WIFs are measured straight with in-link numbers instead of the values of in-link numbers divided by page number.

• Degree: In graph theory, degree is defined as the number of edges adjacent to a node. In web space, the concept of degree has been widely used by search engines to rank web resources, such as page rank and hits. In SNA, degree is used to indicate how tightly a person or institution connects with its counterparts. The web network is similar to a social network, and we call the number of incoming links a site received in-degree and the number of outcoming links from a site out-degree.

• Co-link: The concept of co-link is analogous to the concept of co-citation. If pages X and Y are both linked to by page Z then X and Y are co-linked [20]. Co-link analysis can be carried out to mine relationships between objects at different granularity.

• Betweenness: This measure is based on the number of shortest paths passing through a node. Nodes with a high betweenness play the role of connecting different groups, as ‘middlemen’ [19].

• K-core: K-core is a highly cohesive area in the whole graph, in which each node is connected with fewer k-nodes. The tool of k-core allows the analyser to distil strongly connected nodes as groups. On the web, those groups may share the same interests.

An adjacency matrix representing the linking strength between a couple of samples was made from the out-link data harvested from 93 university websites by LinkDiscoverer, and other data analysis was executed on the transformed data of the adjacency matrix. Figure 1 shows the position of each sample in the linking network composed of the whole population, and the position represents the comprehensive influence of the institution that the sample (website) belongs to. After data reduction and binary transformation on the primitive adjacency matrix with Ucinet, principal component, co-link and k-core analysis can be performed effectively in terms of the strong relationship between samples.

5. Results

The 93 websites sampled were crawled from 20 May 2008 to 15 June 2008, and the raw link matrix was constructed on the 7,209,197 links that LinkDiscoverer harvested. The data were analysed and the results showed the following according to the key webometric and social network-based indicators in the following five subsections:

1. comprehensive strength evaluation based on WIF;
2. comparison among MDS, principal components and co-link;
3. in-link-based influence analysis;
4. betweenness-based analysis on core communicators;
5. k-core-based analysis on subject-centred and region-centred trends.

MDS is implemented with SPSS and other visual data analyses are performed with social network tool NetDraw [21].

5.1. Comprehensive strength evaluation based on WIF (in-link)

The ranking list applied here is also used in Qiu et al.’s study and is the most widely accepted by far in mainland China [11]. Tables 1 and 2 are the correlation coefficients between in-link numbers
Table 1
In-link/rank correlations in 2008.

<table>
<thead>
<tr>
<th>Correlations</th>
<th>in-link</th>
<th>rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman’s rho</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in-link</td>
<td>1.000</td>
<td>0.720**</td>
</tr>
<tr>
<td>rank</td>
<td>0.93</td>
<td>93</td>
</tr>
<tr>
<td><strong>Correlation Coefficient Significance (two-tailed)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>93</td>
<td>0.93</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (two-tailed).**

Table 2
In-link/rank correlations in 2005 [18].

<table>
<thead>
<tr>
<th>Correlations</th>
<th>in-link</th>
<th>rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman’s rho</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in-link</td>
<td>1.000</td>
<td>0.769**</td>
</tr>
<tr>
<td>rank</td>
<td>0.47</td>
<td>47</td>
</tr>
<tr>
<td><strong>Correlation Coefficient Significance (two-tailed)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>47</td>
<td>0.47</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (two-tailed).**

and university rankings based on the data in 2008 and 2005. Both of the results show that in-link numbers and university rankings are clearly correlative with each other at the significance level of 0.01. Comparing the results of the similar experiment in Yang and Qin’s study [18] and the result (about 0.5) obtained by the study by Qiu et al. [11], the correlation coefficients shown in Table 1 are much higher and stable at around 0.7.

5.2. **Comparison among MDS, principle components and co-link**

Multidimensional scale is a statistical technique used in information visualization for exploring similarities or dissimilarities between data. The graph in Figure 1 was drawn with an MDS tool to clearly show the distribution of the original data and to verify its overall validity. The goodness-of-fit measure of MDS shows how well a given data set maps into a specific space. In this study, the normalized raw stress is 0.13488, less than 0.2 (acceptable experientially), and the Tucker’s coefficient of congruence is 0.93012. The test indicators above demonstrate that the two-dimensional (2D) data map substantially fits the real data, which means the 2D MDS graph (Figure 1) generally reflects the linkage relationship between 93 websites. The data tags in Figure 1 are the third-level domain names of the websites and the dots are their positions on 2D planes, for example, ‘pku’ (near the centre of the map) stands for the Website of Peking University. The overall samples in Figure 1 can be separated into three groups manually by two circles. It is intuitively obvious that nearly two dozen of the samples except ‘dhu’ in the larger circle have the greatest comprehensive strength in mainland China, and most universities with greatest comprehensive strength do fall into the area, except ‘baua’. Two samples, ‘pku’ (Peking University) and ‘tsinghua’ (Tsinghua University), in the smaller circle are close to each other and located near the centre of the map. The situation reflects that Peking University and Tsinghua University are the most approved of among the 93 top-level universities in mainland China. Although no distinct clusters of universities are found in Figure 1, as a 2D data map tool the MDS graphs mapped from 93 dimensions show us the distribution and the meaning of the data is clearly that the comprehensive strengths of the universities decrease along the radius of the graph from the centre to the margin.

Multiple variables can be converted into several main components during principle component analysis and thus knowledge from the data is reflected by integrated components. The principle component analysis method helps us to separate the samples into several components. **Co-link is a new bibliometric-based approach to web-link analysis and is applied in many studies** [2, 22]. It is derived from the author co-citation analysis and shows how well the specified couples of websites are linked by the others [23]. Figures 2 and 3 drawn with NetDraw are the resultant maps of principle component analysis and co-link analysis. To show strong relationships between samples, the data set is reduced (threshold = 30) and converted into binary values before the principle component analysis. Two samples in the left circle of Figure 2 are ‘pku’ and ‘tsinghua’, and the samples in the middle rectangle are close in comprehensive strength to the ones in the larger circle in Figure 1. Moreover, the sample set in the circle of Figure 3 and the sample sets in the first two groups in Figures 1 and 2 have a quite high overlap rate.

Similar conclusions are reached in the MDS, principle component and co-link analyses with different data processing and analysis methods respectively. The core sample set in the link graph stands for most of the universities with the greatest comprehensive strength in mainland China being identified successfully.

5.3. **In-link-based influence analysis**

In-link (how many times a website is linked by other web pages) measures the acknowledgement from link source to linked target. The more times a website is linked, the greater influential power the affiliated institution is seen to have. The theory and hypothesis above are confirmed through in-link analysis. As shown in Figure 4, the size of node represents the in-link value of the sample. There are some samples in the centre of the in-link graph (Figure 4), such as ‘tsinghua’, ‘pku’, ‘zju’, ‘nju’ and ‘fudan’, which are the most influential universities, and ‘tsinghua’ and ‘pku’ are the overwhelming ones among them.
Fig. 2. Principle components.

Fig. 3. Co-link.
Bo Yang, Zhihui Liu and Joseph A. Meloche

5.4. Betweenness-based analysis on core communicators

In social network analysis, betweenness is a measure of the centrality of a node among other nodes. A node with high betweenness acts as an important medium, and thus locates at the centre of the network [24]. The theory is applied to finding universities which act as key communicators and play important roles in scientific communication. The samples in the centre of Figure 5, such as ‘pku’, ‘tsinghua’, ‘zju’ and ‘nju’, are notable in the link network, and the situation is consistent with the fact that Peking University and Tsinghua University bear greater responsibility than the rest in national and international scientific communications. The similar node sizes of ‘tsinghua’ and ‘pku’ in the influence analysis in Section 4.3 are explained as the similar influential power of the two universities, while the betweenness value (size of node) of ‘tsinghua’ is obviously less than that of ‘pku’. The possible reason for this is that Tsinghua University is a technical university with relatively narrower subject coverage, while Peking University is a comprehensive university and is committed to more scientific communication activities in more fields than the others.

5.5. K-core-based analysis on subject-centred and region-centred trends

The hypothesis of k-core analysis in this study is that 93 samples are located in different regions and belong to different subjects, so the linkage characteristics of the websites can show the subject relations or regional relations between samples to a certain degree. Maximal spanning trees are applied by Thelwall [8] to identify several geography-based groups among British universities; the subject relations are found among Nordic universities in Ortega and Aguillo’s study [25].

As shown in Figure 6, three cores are identified through k-core analysis (to show the substructures clearly, the isolated nodes and the cores without clear structure are omitted). The subject and regional concentration rules are reflected respectively in the four elliptic areas in Figure 6. The subject type of the samples in area 1 (‘snnu’, Shaanxi Normal University; ‘sdnu’, Shandong Normal
Fig. 5. Betweenness.

Fig. 6. K-cores.
University; ‘ecnu’, East China Normal University; ‘ccnu’, Huazhong Normal University.) is teacher training, the type in area 3 (‘cau’, China Agricultural University; ‘njau’, Nanjing Agricultural University; ‘hzau’, Huazhong Agricultural University.) is agriculture and the type in area 4 (‘nwpu’, Northwest Polytechnic University; ‘buai’, Beijing University of Aeronautics and Astronautics; ‘nuaa’, Nanjing University of Aeronautics and Astronautics) is aeronautics and astronautics. The samples in area 2 (‘scnu’, ‘sysu’, ‘scut’ and ‘jnu’) are located in Guangdong Province and this phenomenon explains the regional concentration rule well. The relations shown in closed areas in three cores are pure subject relations (areas 1, 3 and 4) or regional relations (area 2), while the branch nodes may be connected to the groups for the reasons of subject or regional relations. ‘Hbeu’ (Harbin Engineering University) attached to area 2 and ‘ouc’ (Ocean University of China) attached to area 3 follow the formal rule (subject relations), while another several cases follow the latter rule (regional relations), such as ‘snnu’ in area 1 and ‘xidian’ (in Shaanxi Province), ‘hzau’ in area 3 and ‘hust’ (in Hubei Province) and ‘njau’ in area 3 and ‘seu’ (in Jiangsu Province).

6. Discussion

Once a sample has more than one domain name, such as ‘cumt’ (cumt.edu.cn and cumtb.edu.cn), the statistical results from the raw data are combined to enhance the integrity of data collection as far as possible. However, there are still some errors in the data harvest of in-link and out-link. For example, omission in the in-link count will occur when some seldom-used domain names are not discovered, and the web pages with these domain names will not be crawled for they are classified as external pages by the collector, so the data collection on out-link is affected. With respect to link filtering, out-links are picked up on official websites of universities to improve the effects of link analysis, therefore errors induced by link filtering may cause statistical errors in out-link collection.

The thematic relationships are found in co-link analysis in the study of Ortega and Aguillo [20], while it is not in co-link analysis but in k-core analysis that the thematic relations are found in this study. Though subject and regional correlations are found, those relations are sometimes crossed. Both relationships reflect the collaborative relations between universities, but it is vital to distinguish one from the other. The subject-based link motivation shows the full-scale information about the distributions of the main research forces and related resources of the subject. In contrast, geography-based link motivation demonstrates the collaborative relationship within a specific region. Nevertheless, study on regional collaborative relationships is necessary, as most collaborations of this kind are at low level and are hard to represent at the true level of the field. Besides the division of the motivations discussed above, identification of abnormal link behaviours is also important to construct raw link matrix.

7. Conclusion

Arising from the research questions, the main contributions of this paper are as follows: the reliability of new in-link measured WIF has been proven; the SNA measures worked well on visualization of the Chinese academic web, especially on the representation of the comprehensive strength of the universities in mainland China; and the characteristics of the relationships among the universities were successfully worked out by SNA within a single national network rather than on the scale of an international/federal network.

Unlike the bibliometric indicator (JIF), the new indicator (WIF) has its own advantage on comprehensive strength evaluation on institution rather than research productivity. Several previous studies are dedicated to finding proper ways, including data collection methods and calculation formula, to evaluate academic institutions. Though most of these obtain significant conclusions, the reliability is relatively low with a correlation strength at around 0.5, and the stability of the methods applied is not proven. In this study, the strategy and the result were compared with that of a previous study, similar strategies were carried out and similar results were shown in a time window of three years. Both of the studies obtained much more significant results.
Principal component analysis (PCA), SNA measure (betweenness) and its derivatives with bibliometric (co-link, in-link) were applied by SNA tools to analyse the link network of the top 100 universities in mainland China. The conclusions drawn from the figures show that the method of mining hidden patterns in Chinese academic link network with SNA measures is workable. In most cases, the universities near the centre of the figure (the second group for PCA analysis) are all prestigious with a few exceptions. The majority of them are permitted by the Ministry of Education of China to enrol graduate students independently. Peking University and Tsinghua University stand out for their overwhelming influence. The authors argue that the patterns of the comprehensive strengths and the respect those universities received are successfully mined through link analysis with the help of SNA.

Most previous studies focused on websites between different countries, while studies on subject/regional trends of the university websites within a country or region without significant regional cultural differences are rare. The results of influence, betweenness, and k-core analysis, especially for k-core analysis, demonstrate that, besides macro analysis on country/region link behaviours, link analysis and identification of link patterns at the micro-level are feasible. During the k-core analysis it was shown that the attractions for one university linking to another on academic websites are influential power and academic strength, as well as geographic factors. At the macro-level, links to a website are known mostly for their influential power, such as ‘pku’ and ‘tsinghua’, while at the micro-level analysis on link sub-graph, subject/regional correlations are found. It is demonstrated through micro-link analysis that the motivations of collaboration between universities include both significant subject and geographic factors.

Although the statistical data were combined in the case of various domain names for one university caused by university mergers, not all domain names used by the selected universities were identified. Therefore the in-link and out-link calculations are affected to some extent. Furthermore, to get less biased evaluating results, we try to harvest relatively authoritative pages that reflect the official standpoints of universities. The ‘authoritative pages’ we defined for the filtering rules may not be always authoritative. The subject/regional sub-structures found are general at the level of the data granularity in this study. In a future study, more emphasis should be put on more specific information units, such as university departments, to get a more clear and detailed map of Chinese academic web.

Acknowledgement

We would like to thank Jian Qin at Syracuse University and the anonymous referees for their helpful comments on the draft of this paper.

References