Research on Link Structure of Academic Blogs Based-on Social Network Analysis

Xinjin Fu[1,2], Shu Fang[1], Hongshen Pang[1,2]

1 The Chengdu Branch of the National Science Library, Chinese Academy of Sciences, Chengdu, P.R.C.

2 Graduate University of Chinese Academy of Sciences, Beijing, P.R.C.

fuxinjin@mail.las.ac.cn

Abstract-**Academic blogs in field of information and library science are chosen as a sample. Claw their links to generate a link matrix. By using social network analysis and visualize softwares, analyze the link matrix with betweenness centrality, structural holes and core/periphery model to find the core area of the sample structure and explain why there are some exception.**

Keywords-academic blogs; link; social network analysis

I. INSTRUCTION

Blog is an offspring of web2.0. It makes web users posting information anytime anywhere and enriches the information of the internet. Blog's easy using gains the popularity among a lot of scholars who open their blogs one after another for publishing their own new ideas. Academic blogs have their own subject, independence, flexibility and personality. This informal academic communication becomes popular, reinforcing the traditional way effectively. The bloggers of the academic blogs using their real names and their background and knowledge make sure the quality of these blogs' content.

Crawl these academic blogs and use social network analysis to mine the structure of the blog links and find the core blogs.

II. DATA CAPTURE AND TOOLS CHOOSE

Based on the ref. [2] referring the list of blogs in the field information and library science, we check these links connective or not. Then complement friend links from these available blogs. After deleting the unconnected blogs, there are 19 blogs shown in table I.

TABLE I. ACADEMIC BLOGS SAMPLE LIST

No.	Name	Blog Url
1	oldhuai	www.oldhuai.name
2	kevenlw	www.kevenlw.name
3	mingzhiguwen	www.mingzhiguwen.net
4	tuyouqibiao	www.tuyouqibiao.com
5	yyiyatou	yy.iyatou.com
6	libseeker	libseeker.bokee.com
7	wujianzhong	www.wujianzhong.name
8	infokm	infokm.bokee.com
9	dqdx	dqdx.blogbus.com
10	sogg	www.sogg.name
11	yizii	www.yizii.info
12	guofu	guofu.us
13	catwizard	catwizard.net
14	liuliu	liuliu.bokee.com
15	solopro	solopro.cn
16	beefsteak	beefsteak.blogbus.com
17	blueyye	hi.baidu.com/blueyye
18	cunfu	cunfu.info
19	gsls	gsls.info

The free academic software SocScibot is used to crawl and form the link matrix shown in table II. For example, oldhuai has links to kevenlw, so count as 1; oldhuai has no link to mingzhiguwen, so count as 0. We do not study the link to itself, so count as 0.

TABLE II.   ACADEMIC BLOGS LINK MATRIX

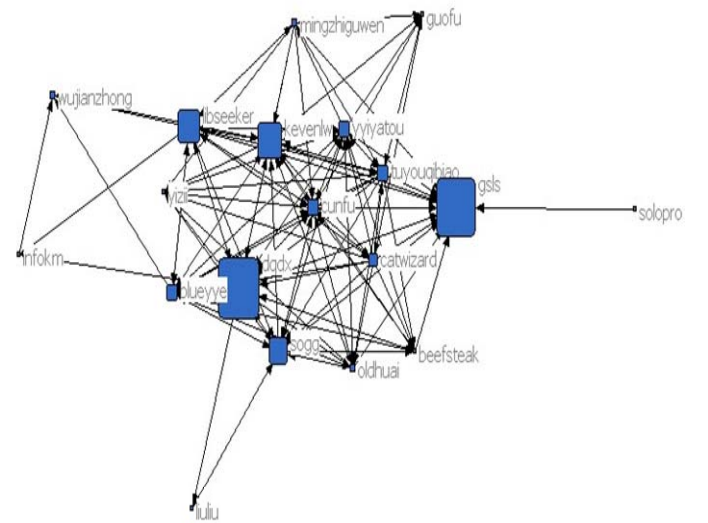|		o k m t y l w i d s y g c l s b b c g
1	oldhuai	0 1 0 0 1 0 0 0 0 1 1 0 0 0 0 0 1 1 1
2	kevenlw	1 0 0 0 1 1 1 0 1 0 1 1 0 0 0 1 1 1 1
3	mingzhiguwen	0 1 0 1 0 1 0 0 1 0 1 1 0 0 0 0 0 1 1
4	tuyouqibiao	1 1 0 0 1 1 0 0 1 1 1 1 0 0 0 1 1 1 1
5	yyiyatou	1 1 1 1 0 1 0 0 1 1 1 1 0 0 0 1 1 1 1
6	libseeker	0 1 1 1 1 0 1 1 1 1 0 0 1 0 0 0 1 1 1
7	wujianzhong	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
8	infokm	0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
9	dqdx	1 1 0 0 0 1 0 1 0 1 1 0 0 1 0 1 1 1 1
10	sogg	0 1 0 0 1 0 0 0 1 0 1 0 0 1 0 1 1 1 1
11	yizii	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
12	guofu	0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
13	catwizard	1 1 0 1 1 1 0 0 1 1 1 1 0 0 0 1 1 1 1
14	liuliu	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
15	solopro	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
16	beefsteak	0 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 1
17	blueyye	1 1 0 0 1 1 1 0 1 1 0 0 0 0 0 1 0 1 0
18	cunfu	1 1 0 0 1 0 0 0 1 1 1 1 1 0 0 0 1 0 1
19	gsls	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
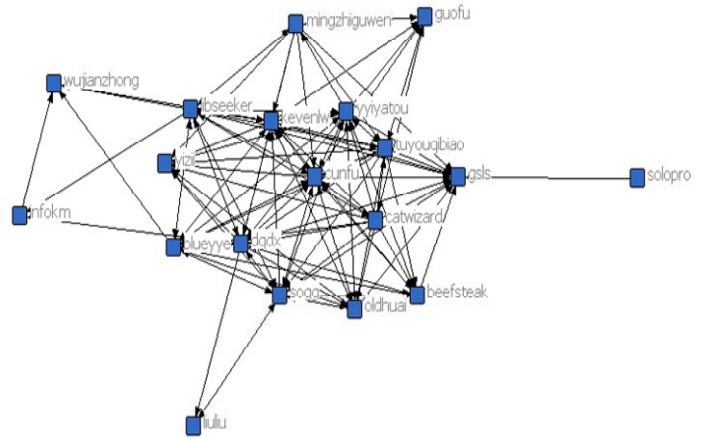
Figure1   Academic blog network chart

Use the social network tool Ucinet and visualization tool Netdraw to analyze and draw. Fig.1 shows the sample network chart.

III.   DATA ANALYSIS

A.   Betweenness Centrality of Every Node

Betweenness centrality measures the degree of actor's ability of controlling the resource. Using Ucinet to analyze the link matrix, we get the betweenness centrality of blogs shown in table III columns 3 and 7. Draw this chart by Netdraw shown in fig.2. Shape and size of node can be observed directly.

Figure 2.   Betweenness centrality of academic blog network

Based on table III and fig.2, we can see betweenness centrality of solopro and liuliu is 0, meaning this kind of node cannot control any actor with the position of periphery. While dqdx has the highest betweenness centrality, it means dqdx has the superiority to control and very powerful with the position of core. From the data of table II, we can see dqdx has 11 links to other blogs, and receives 10 links, so it also proves that dqdx own the maximal control superiority. Some other blogs which have comparatively high control ability are gsls, kevenlw, libseeker.

TABLE III.   BETWEENNESS CENTRALITY OF ACADEMIC BLOGS

No.	Name	Betweenness Centrality	Constraint	No.	Name	Betweenness Centrality	Constraint
1	oldhuai	0.547	0.294	11	yizii	0.379	0.275
2	kevenlw	10.609	0.253	12	guofu	0.111	0.311
3	mingzhiguwen	1.425	0.275	13	catwizard	3.117	0.261
4	tuyouqibiao	3.83	0.264	14	liuliu	0	0.654
5	yyiyatou	3.83	0.262	15	solopro	0	1
6	libseeker	9.433	0.242	16	beefsteak	0.2	0.298
7	wujianzhong	0.744	0.398	17	blueyye	3.536	0.281
8	infokm	0.393	0.46	18	cunfu	3.83	0.267
9	dqdx	19.106	0.247	19	gsls	17.879	0.237
10	sogg	8.032	0.26				

B. Structural Holes

In structural holes, the third one who can connect the other two no direct links has superiority of owning information and control.

Therefore, organizations and individuals in an organization all try to be the third one, and keep the superiority not allowing the other two connected. There are many indicators of measuring structure holes, including the constraint which is the most important one. Constraint means the one has ability to use structure holes in the network.

To analyze the structure holes in table II, we can get the constraint value shown in table III columns 4 and 8. Solopro has the value of 1 which has the highest constraint, and it locates at the periphery of the network associating fig.1. Gsls has the lowest constraint, because the periphery node solopro connects to other nodes only by gsls. Thus there are some nodes having low constraint: gsls, libseeker, dqdx, kevenlw.

It is clear that a node which locates closer to the core, it has more structure holes, lower constraint and higher betweenness centrality.

C. Core/Periphery Analysis

Core/Periphery model can separate a series of actors having high density (core) and ones having low density (periphery). Core actors have superiority in the commutation relation with periphery actors.

By using Ucinet to do Core/Periphery model, we get Core/Periphery Class Memberships:

- oldhuai kevenlw tuyouqibiao yyiyatou libseeker dqdx sogg catwizard blueyye cunfu

- mingzhiguwen wujianzhong infokm yizii guofu liuliu solopro beefsteak gsls
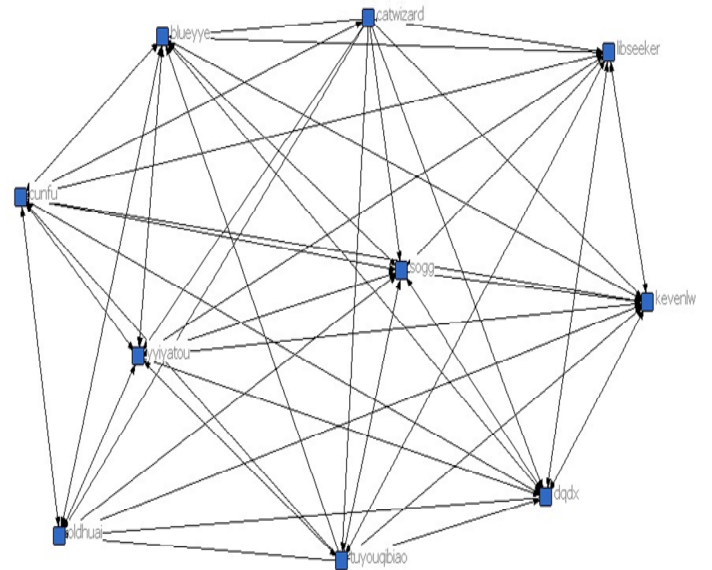

Figure 3.   Network of core nodes

The final fitness is 0.667. This value is closer to 1, the fitness is higher. We also can judge the result better or worse according to the density matrix. Top left block reflects the core members, and its density is 0.767, meaning more information concentrate on this block. From this, the core

nodes showing in this partitioned matrix are 10. Extract the 10 nodes and draw their network chart shown in fig.3. Apparently, this chart is compliant with stable net structure. This net structure has higher complete degree, reflecting each node has high tightness.

IV. RESULT ANALYSIS

Choose the top 10 nodes in betweenness centrality analysis and structure holes analysis separately to compare with core/periphery analysis. We find that:

- Gsls is in the very front in betweenness centrality and constraint analysis, but is not in the core area. From table 2, this node has no links to other nodes, but receives 12 links. Node solopro only connect to gsls, so gsls receive link from solopro making its control resource ability increase and constraint decrease.

- Oldhuai is in core area, but not in the top 10 of betweenness centrality and constraint analysis. We count the number of in-link and out-link, and get the average value, finding oldhuai has amount of link counts, and in-link number is nearly as much as out-link. Perhaps it can be in the core area associating with the link counts.

- Mingzhiguwen is in top 10 of Constraint analysis, but is not in the core area. Its link counts levels center bottom, maybe causing it not in the core.

This paper analyzes the nodes based on the whole network, so the node analysis result will change if the other node changes, and gsls is the good example. Ref. [5] also refers that the big number of out-link may cause high centrality and low core.

We just choose the blogs of the field of information and library science, and only check there has or not links, not considering the link counts, so it just reflects academic blogs link structure partly. In addition, some bloggers changed their blog address, and others who link to them didn't update, so it causes the in-link decreases, influences the result. For the future, we will choose academic blogs in different field to analyze, thereby reflect the net structure of academic blogs more comprehensively.

REFERENCES

[1] Shi Xinyan, Xiao Xiantao, "Link Structure of Academic Blog and Characteristics of Communication Analysis," Document,Information & Knowledge, vol 5, pp.79-83+89, 2009 (In Chinese).

[2] Qiu Junping, Ma Ruimin, Yu Changfu, "The Social Network Analysis of Blogs about Library Research," Library and Information Service, vol 11, pp. 6-9, 2008 (In Chinese).

[3] SocSciBot: Link crawler for the social sciences. http://socscibot.wlv.ac.uk/.

[4] Liu Jun, "Letures on Whole Network Approach," Shanghai: Gezhi Publishing House, 2009 (In Chinese).

[5] Zhang Yue, Zhu Qinghua, "An Empirical Study of Communication Network in Academic Blogs Using Core—Periphery Analysis," Library and Information Service, vol 12, pp. 25-29, 2009 (In Chinese).