

Analysis of Science Collaboration Network of Top Universities in The World

Xianwen Wang¹ Zeyuan Liu² Yue Chen³ Xiaoyu Zhu⁴

16 October 2008

Abstract

In this paper, the authors try to analysis and visualize the collaboration situation of universities in the world and the influence of distance to collaboration. Firstly the authors select 100 famous research-oriented universities in the world, and search the collaboration of scientific literature from 1975 to 2007 in Web of Science, to get the collaborative matrices of universities. Using methods of social network analysis, the authors study the structure of collaborative network among 100 universities, and a new index is designed to measure and rank the collaboration of universities.

1 Introduction

This article explores the research collaboration among famous universities in the world on the basis of millions of articles indexed in Web of Science published from 1975 to 2007. The article's bibliographical data have been used to analyze or learn about:

1. *The structure of collaborative network of the universities;*
2. *The ranking of collaboration of the universities.*

There are many studies about the scientific collaboration among scientists. Davidson Frame and Carpenter(Davidson Frame & Carpenter, 1979) studied international collaborative behaviour among scientists. Newman constructed

networks of collaboration between scientists in each of these disciplines using computer databases of scientific papers in physics, biomedical research, and computer science(Newman, 2001). Luukkonen et al studied the networks of international scientific collaboration among countries, and patterns of international collaboration in scientific fields(Luukkonen, Persson, & Sivertsen, 1992). Liming Liang researched the pattern of China's inter-regional research collaboration, and analyzes how the collaborative pattern was formed(Liang & Zhu, 2002).

However, studies about the collaboration at universities are seldom. Persson et al researched the collaboration at 22 Nordic universities(Persson, Melin, Danell, & Kaloudis, 1997).

2 Data and methodology

2.1 Method

As Table 1 shows, When C_{11} and C_{22} the amounts of published articles in cell C_1 and C_2 , and C_{21} or C_{12} represents the amount of collaborative articles between cell C_1 and C_2 .

Table 1: Collaboration matrix between two units

	C_1	C_2
C_1	C_{11}	C_{12}
C_2	C_{21}	C_{22}

Then the collaborative ratio of region C_1 and C_2 can be calculated as:

¹WISE Lab, Dalian University of Technology, Dalian, China, wangxianwen at gmail dot com

²WISE Lab, Dalian University of Technology, Dalian, China

³WISE Lab, Dalian University of Technology, Dalian, China

⁴WISE Lab, Dalian University of Technology, Dalian, China

$$C_{12} = C_{21} = C_{11} \cap C_{12}, I_{12} = \frac{C_{12} \times C_{21}}{C_{11} \times C_{22}} \quad (1)$$

For the collaborations among N units, as Table 2 shows, the ratio of collaboration of any two cells can be calculated as follows.

$$C_{ij} = C_{ji} = C_{ii} \cap C_{jj}, I_{ij} = \frac{C_{ij} \times C_{ji}}{C_{ii} \times C_{jj}} \quad (2)$$

Table 2: Collaboration matrix among N units

	C ₁	C ₂	C ₃	...	C _n
C ₁	C ₁₁	C ₁₂	C ₁₃	...	C _{1n}
C ₂	C ₂₁	C ₂₂	C ₂₃	...	C _{2n}
C ₃	C ₃₁	C ₃₂	C ₃₃	...	C _{3n}
...
C _n	C _{n1}	C _{n2}	C _{n3}	...	C _{nn}

2.2 Top research-oriented universities

The data of our study is from The Academic Ranking of World Universities – 2007, published by the Institute of Higher Education, Shanghai Jiao Tong University. The authors select top 100 research-oriented universities according to the ranking.⁵

Table3: Statistics of top 100 universities by Country

Country USA	UK	Japan	Germany
Amount 54	11	6	6
Country Canada	France	Sweden	Switzerland
Amount 4	4	4	3
Country Netherlands	Australia	Israel	Denmark
Amount 2	2	1	1
Country Norway	Finland	Russia	
Amount 1	1	1	

2.3 Collaboration matrix

The primary source of data is collected from Web of Science (WOS), including SCI-E, SSCI, A&HCI. The authors choose the retrieval time as 1975 to 2007, and then got the collaboration matrix as Table 4 shows. In this study, the authors only analysis the articles each university participated, regardless of the order of authors, so the collaboration matrix among universities is asymmetrical.

Table 4: The collaboration matrix of top 100 universities

	Harvard	Stanford	Berkeley	Cambridge ...
Harvard	284505	3141	1895	652 ...
Stanford	3141	149359	3119	391 ...
Berkeley	1895	3119	148355	601 ...
Cambridge	652	391	601	123379 ...
...

2.4 Ratio of collaboration

Then we calculated the matrix of collaboration ratio of the 100 universities, as Table 5 shows.

Table 5: The ratio of collaboration among top 100 universities

	Harvard	Stanford	Berkeley	Cambridge ...
Harvard	0	0.00023	8.5E-05	1.2E-05 ...
Stanford	0.00023	0	0.00044	8.3E-06 ...
Berkeley	8.1E-05	0.00044	0	2.0E-05 ...
Cambridge	1.2E-05	8.3E-06	2.0E-05	0 ...
... 0

3 The structure of collaboration network of 100 universities

When imported the matrix of collaboration ratio to Ucinet, we dichotomize the matrix into 0-1 matrix by comparing each cell value with a specified cutoff, which is set according to the concrete structure of network(Borgatti & Everett, 2000; Borgatti, Everett, & Freeman, 2007). Then the program of Ucinet can assign a 1 to any cell whose value is greater than the threshold while assign a 0 to the cells less than the threshold.

3.1 The cutoff is 1255

Here we set the cutoff as 1255, which indicates that any cell in the collaboration matrix whose value greater than 1255 is assigned 1, and the cells whose value less or equal than 1255 are assigned 0. Then the collaboration matrix is transformed into 0-1 matrix.

If the amount of collaborative articles between two universities is greater than 1255, there exists a link. As Fig.1 shows, there are several main clusters in the network, including 87 nodes. One dominant cluster consists of universities in

⁵cf. <http://www.arwu.org/rank/2007/ranking2007.htm>

America, another smaller cluster, nodes in which includes universities mainly from England and German. These two clusters are connected by the link between Univ Manchester and Univ Maryland - Coll Park, which means between the universities in America and universities in England/Germany, only the value of collaboration ratio of Univ Manchester and Univ Maryland is greater than 1255.

There are also some other smaller clusters in the network, such as the cluster of Japanese universities, the cluster of French and Sweden universities.

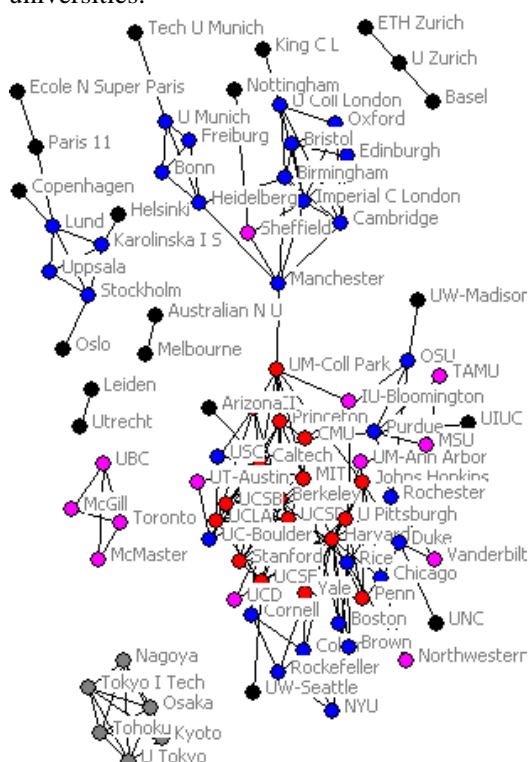


Figure 1: The structure of network of collaboration among 100 universities (The cutoff is 1255)

3.2 The cutoff is 822

When we reduce the cutoff to 822, the network becomes more complex, there are more nodes and links emerged in the network, and the most crucial link for the cutoff 822 is the link between Univ Paris 11 and Univ Heidelberg, which made the cluster of French and Sweden universities connected with the large cluster of American, British and German universities. And the links of 3 Swiss universities, ETH Zurich, Univ Zurich and Univ Basel, is connected with the large

cluster through the links from German universities. The cluster of American universities is still the most dense, and the cluster of European universities is more close to American universities. And the cluster of Japanese universities is still separate with the main cluster in the network, which indicates that the collaboration between universities of Japan and America/Europe is not so closer. There are also other individual links in the network, such as Univ Melbourne and Aus Nat Univ, link of Univ Leiden and Univ Utrecht.

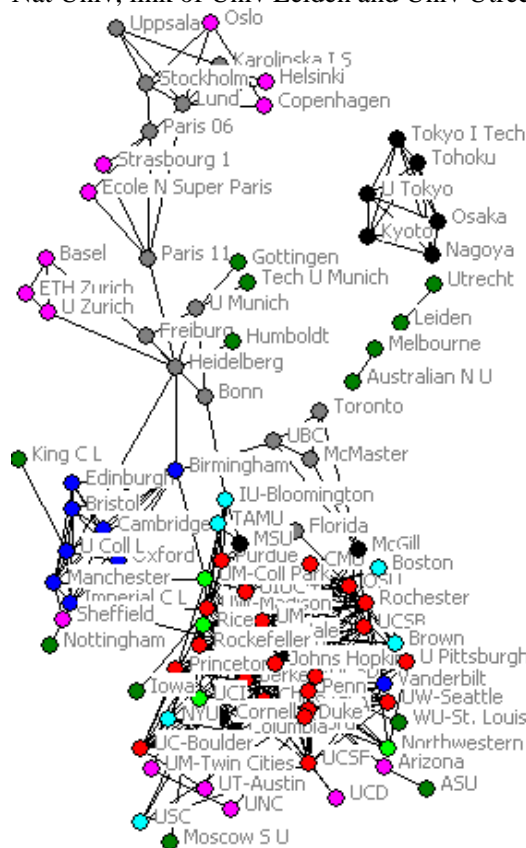


Figure 2: The structure of network of collaboration among 100 universities (The cutoff is 822)

3.3 The cutoff is 438

When the cutoff is reduced to 438, all the nodes appear in the network, except Rutgers State Univ and Hebrew Univ Jerusalem, as Fig. 3 shows. Most of the nodes are in the main cluster, especially, the cluster of Japanese universities is connected to the large cluster through the link of Univ Tokyo and Univ Munich.

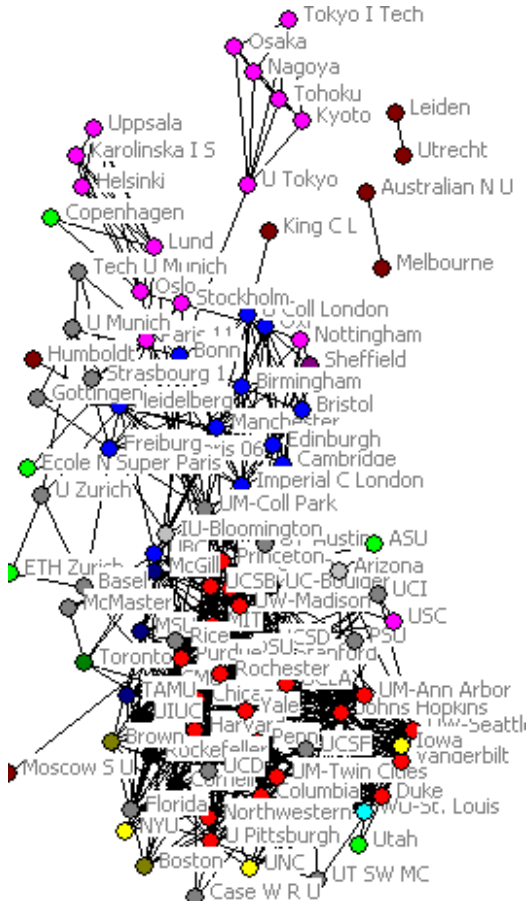


Figure 3: The structure of network of collaboration among 100 universities (The cutoff is 438)

4 Collaborative index

4.1 Design of Collaborative index

A new index is designed in this paper to measure the collaboration of universities. For the collaborative matrix of universities as Table 2 show, the collaborative indices can be calculated as following.

$$\begin{aligned}
 I_j &= \lambda_1 I_{j1} + \dots + \lambda_{j-1} I_{j(j-1)} + \lambda_{j+1} I_{j(j+1)} + \dots + \lambda_n I_{jn} \\
 &= \frac{C_{11} + C_{jj}}{\sum_{i=1}^n C_{ii}} \times \frac{C_{1j} \times C_{j1}}{C_{11} \times C_{jj}} + \dots + \frac{C_{(j-1)(j-1)} + C_{jj}}{\sum_{i=1}^n C_{ii}} \\
 &\times \frac{C_{j(j-1)} \times C_{(j-1)j}}{C_{(j-1)(j-1)} \times C_{jj}} + \frac{C_{(j+1)(j+1)} + C_{jj}}{\sum_{i=1}^n C_{ii}} \times \frac{C_{j(j+1)} \times C_{(j+1)j}}{C_{(j+1)(j+1)} \times C_{jj}} \\
 &+ \dots + \frac{C_{mn} + C_{jj}}{\sum_{i=1}^n C_{ii}} \times \frac{C_{nj} \times C_{jn}}{C_{mn} \times C_{jj}} \quad (3)
 \end{aligned}$$

4.2 Ranking of collaborative indices

Then the collaborative indices of the 100 universities are calculated, as Table 6 shows. The index of Harvard Univ is the highest (0.4346), when MIT is second and Univ Tokyo third. The others in top ten are Univ California – Los Angeles, Univ California – Berkeley, Stanford Univ, Caltech, Boston Univ, Johns Hopkins Univ, Univ Michigan – Ann Arbor.

Table 6: Ranking of collaborative indices of 100 universities

Institution	Index	Rank
Harvard U	0.4346	1
MIT	0.2232	2
U Tokyo	0.1977	3
U C-Los Angeles	0.1833	4
U C-Berkeley	0.1630	5
Stanford U	0.1340	6
Caltech	0.1289	7
Boston U	0.1277	8
Johns Hopkins U	0.1259	9
U Michigan-Ann Arbor	0.1217	10
U Penn	0.1188	11
Yale U	0.1172	12
Duke U	0.1171	13
U C-San Diego	0.1160	14
Kyoto U	0.1144	15
U C-San Francisco	0.1094	16
U Wisconsin-Madison	0.1040	17
U Pittsburgh	0.0984	18
U Washington-Seattle	0.0976	19
Columbia U	0.0925	20
U Toronto	0.0924	21
Osaka U	0.0909	22

U Chicago	0.0896	23	U Edinburgh	0.0295	73
Ohio S U-Columbus	0.0854	24	U Bonn	0.0286	74
Tohoku U	0.0769	25	Tech U Munich	0.0275	75
U Paris 06	0.0748	26	U Freiburg	0.0269	76
U Rochester	0.0735	27	ETH Zurich	0.0256	77
UNC-Chapel Hill	0.0721	28	U Utrecht	0.0214	78
Cornell U	0.0708	29	U Utah	0.0210	79
Princeton U	0.0699	30	Penn S U	0.0208	80
McGill U	0.0691	31	Uppsala U	0.0195	81
Imperial C London	0.0686	32	Case Western Reserve U	0.0184	82
U C-Santa Barbara	0.0685	33	U Leiden	0.0183	83
Northwestern U	0.0680	34	Lund U	0.0182	84
U Minnesota-Twin Cities	0.0645	35	U Basel	0.0160	85
Carnegie Mellon U	0.0644	36	U Sheffield	0.0140	86
Purdue-West Lafayette	0.0630	37	U Helsinki	0.0132	87
U Maryland-Coll Park	0.0625	38	Ecole N Super Paris	0.0131	88
U C-Davis	0.0618	39	U Texas SW Med Center	0.0116	89
U Oxford	0.0618	40	Moscow S U	0.0102	90
Nagoya U	0.0605	41	U Oslo	0.0094	91
U Illinois-Urbana	0.0596	42	Arizona S U	0.0080	92
USC	0.0571	43	Australian N U	0.0079	93
U C-Irvine	0.0562	44	U Nottingham	0.0078	94
U Colorado-Boulder	0.0551	45	U Melbourne	0.0077	95
U Paris 11	0.0519	46	Humboldt U	0.0075	96
Tokyo I Tech	0.0509	47	Hebrew U Jerusalem	0.0065	97
Vanderbilt U	0.0506	48	U Copenhagen	0.0064	98
U British Columbia	0.0487	49	U Gottingen	0.0055	99
Rice U	0.0482	50	U Strasbourg 1	0.0054	100
New York U	0.0465	51	Rutgers S U	0.0047	101
Rockefeller U	0.0464	52	King's C London	0.0026	102
U Florida	0.0461	53			
U Cambridge	0.0461	54			
Karolinska I Stockholm	0.0459	55			
U Manchester	0.0458	56			
Michigan S U	0.0451	57			
U Birmingham	0.0435	58			
U Arizona	0.0428	59			
U Coll London	0.0424	60			
Stockholm U	0.0423	61			
McMaster U	0.0420	62			
U Heidelberg	0.0416	63			
Washington U-St. Louis	0.0413	64			
U Munich	0.0407	65			
U Zurich	0.0399	66			
U Texas-Austin	0.0383	67			
U Iowa	0.0378	68			
Indiana U-Bloomington	0.0352	69			
Brown U	0.0327	70			
Texas A&M U-Coll Station	0.0324	71			
U Bristol	0.0310	72			

5 Discussion

The structure of collaboration among the 100 universities could be separated into several main clusters, the cluster of American universities, cluster of European universities and cluster of Japanese universities. This could be seen by setting different cutoff when transferring the collaborative matrix to 0-1 matrix. The collaboration in the cluster is very common, and rare between clusters, especially the cluster of Japan universities and American Universities.

A new method is used to calculate the collaborative index and evaluate the collaborative status of universities in the network. According to the ranking, Harvard Univ, MIT and Univ Tokyo are the highest, when Univ Strasbourg 1, Rutgers State Univ and King's Coll London are the lowest.

Acknowledgement

This research was supported by the National Natural Science Foundation of China under Grant 70773015, 70431001, 70620140115, National Social Sciences Foundation under Grant 07CTQ008, Project of DUT under Grant DUTHS1002, Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20070141059, projects of League of Social Sciences in Liaoning Province under Grant 2007lslktglx-52.

References

- Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, 21(4), 375-395.
- Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2007). *Ucinet for Windows: Software for Social Network Analysis*, Harvard: Analytic Technologies).
- Davidson Frame, J., & Carpenter, M. P. (1979). International Research Collaboration. *Social Studies of Science*, 9(4), 481.
- Liang, L., & Zhu, L. (2002). Major factors affecting China's inter-regional research collaboration: Regional scientific productivity and geographical proximity. *Scientometrics*, 55(2), 287-316.
- Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding Patterns of International Scientific Collaboration. *Science, Technology & Human Values*, 17(1), 101.
- Newman, M. E. J. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, 64(1), 16131.
- Persson, O., Melin, G., Danell, R., & Kaloudis, A. (1997). Research collaboration at Nordic universities. *Scientometrics*, 39(2), 209-223.