Evolutionary Analysis of Collaboration Networks in *Scientometrics*

Yuehua Zhao[1], Rongying Zhao[2]

[1]*yuehua@uwm.edu*
University of Wisconsin-Milwaukee Milwaukee, School of Information Studies, P. O. Box 413 Milwaukee, WI 53201 (United States)

[2]*zhaorongying@126.com*
Wuhan University, School of Information Management, Research Center for China Science Evaluation, The Center for the Studies of Information Resources, Wuhan, Hubei 430072 (China)

Abstract
The research area of scientometrics began during the second half of the 19th century. After decades of growth, the international field of scientometrics has become increasingly mature. The present study intends to understand the evolution of the collaboration network in *Scientometrics*. The growth of the discipline is divided into three stages: the first time period (1978-1990), the second period (1991-2002), and the third period (2003-2014). Both macro-level and micro-level network measures between the studied time periods were compared. Macro-level analyses show that the degree distribution of the collaboration in each timespan are consistent with power-law, and both the average degree and average distance steadily increase with time. Micro-level structure analyses illustrate the authors with high performance in raw degree measure, degree centrality measure, and betweenness measure are dynamic in different timespans. From three dimensions (raw degree, degree centrality, and betweenness centrality), the collaboration dominators are identified in each time span. In addition, the visualization methods are applied to display the evolution of the collaboration networks for each of the three stages of scientometrics' development.

Conference Topic
Journals, databases and electronic publications

Introduction
Scientometrics is an interdisciplinary field that uses mathematical, statistical, and data-analytical methods and techniques to perform a variety of quantitative studies of science and technology (Chen, Börner, & Fang, 2013). In short, it can be defined as the science of science. The term "Scientometrics" has been first used as a translation of the Russian term "naukometriya" (measurement of science) coined by Nalimov and Mulchenko (1969). The research area of scientometrics began during the second half of the 19th century. This paper proposed a macro- and micro-level overview of the author collaboration patterns in journal *Scientometrics* to study the evolution of the field of scientometrics. The present study intends to understand the evolution of the collaboration network in *Scientometrics*. In this study, social network analysis methods are employed to describe the evolution of scientometrics over nearly 40 years after entering the development stage of this field. Both macro-level and micro-level network measures between the studied time periods were compared. Then, visualization methods were applied to display the evolution of the collaboration networks in three periods: the first time period (1978-1990), the second period (1991-2002), and the third period (2003-2014).

Related Works and Research Questions
Scientometrics has been studied for more than 100 years. Over the past years, scientists' studies of scientometrics shifted from the unconscious to consciousness, from qualitative research to quantitative research, and from external description to detailed study revealing the inherent properties of scientific production. Previous scholars (Pang, 2002; Yuan, 2010) tend to divide the development of scientometrics into three stages: embryonic period (from the

second half of the 19th century to early 20th century), the founding period (from the beginning of the 20th century to the 1960s), and development period (after the 1970s). In order to study the development period of scientometrics, Schubert (2002) indicated that as the representative communication channel of its field, the journal *Scientometrics* reflects the characteristic trends and patterns of the past decades in scientometric research. Therefore, in this study, we employed the publications in *Scientometrics* over the past 37 years to detect the evolution of the scientific collaboration networks in this field.

Previous research has provided some insight into the author collaboration network analysis in different disciplines. Barabasi et al. (2002) investigated the collaboration network in mathematics and neuroscience articles published between 1991 and 1998. Newman (2001) compared the co-authorship networks of in physics, biomedical research, and computer science, and found the differences of the collaboration networks between experimental and theoretical disciplines. By using the bibliometric methods, Ardanuy (2012) analyzed the level of co-authorship of Spanish research in Library and Information Science (LIS) until 2009, and found a significant increase in international collaboration. Given the advanced visualization techniques, Franceschet (2011) represented a collaboration picture of computer science collaboration including all papers published in the field since 1936.

These studies have investigated the collaboration networks in different disciplines and compared their differences. However, few studies investigated the field of scientometrics over the past 37 years. There is a need for researchers to identify and compare both the macro-level and micro-level characteristics of the scientific collaboration network in *Scientometrics* through different time periods.

This paper intended to address the following two research questions:

RQ1. What are the macro-level features of the collaboration networks in *Scientometrics* in each time period?

RQ2. What are the micro-level features of the collaboration networks in *Scientometrics* in each time period?

Method

Data collection

For the development period of scientometrics, the foundation of the journal *Scientometrics* (in September, 1978) is a landmark event. Following some of the predecessors (Schoepflin & Glänzel, 2001; Hou, 2006), this study used the journal as a representative model of scientometrics research. The research data involves 3627 documents published in *Scientometrics* during 1987 to 2014 retrieved from the Web of Science on December 10th, 2014, and the other 347 articles published from 1978 to 1986 retrieved on April 20th, 2013. The total of 37 years were divided into three periods: the first time period (1978-1990), the second period (1991-2002), and the third period (2003-2014).

The raw data extracted from Web of Science database that consisted of the bibliometric information of each paper. Microsoft Excel was applied to build the 2-mode author-to-paper matrices for each time period. In order to produce the collaboration networks, the 2-mode author-to-paper matrices were transferred to 1-mode author-to-author matrices based on the formula proposed by Breiger (1974): $P=A(A^T)$. In this case, the matrix A was the 2-mode author-to-paper matrix and the matrix AT was the transposition of the matrix A, and the 1-mode author-to-author matrix was generated by multiplying these two 2-mode matrices. In the produced author-to-author matrix, each row and column represented an author, the intersection cells contained the cumulative number of the co-authored papers by two authors, and the diagonal cells demonstrated the total number of papers written by each author.

Data analysis

Two social network analysis software packages (Ucinet and Netdraw) (Borgatti, Everett, & Freeman, 2002) were adopted in the data analysis to calculate the network measures and draw the networks. Ucinet is a software package which mainly deals with the social network analysis, and Netdraw, the network visualization tool, can be used to display the networks generated by Ucinet.

Results and Discussion

An overview

Over the 37 years, a total of 4,211 authors published 3,974 papers in *Scientometrics*. Figure 1 indicates the distribution of the number of articles and the number of scholars in each time period. In Figure 1, the *X*-axial represented the 3 time periods, and the *Y*-axial represented the frequencies, and the 2 bars in each period showed the number of authors and articles separately, and the line showed the trend of the differences between the two bars. Separately, 626 papers were contributed to by 435 authors from 1978 to 1990, 1,106 papers were published by 1,029 authors from 1997 to 2005, and 2,242 papers were written by 3,102 authors from 2006 to 2014. Based on Figure 1, both the number of articles and the number of authors increased over the three time spans. When we compared the two frequencies in each period, the number of articles was greater than the number of authors at the first two stages, but the number of authors boomed at the third stage which resulted in the number of authors being much greater than that of the authors. The increases of the total number of articles and authors suggested the rises of the collaboration opportunities through the three time periods.

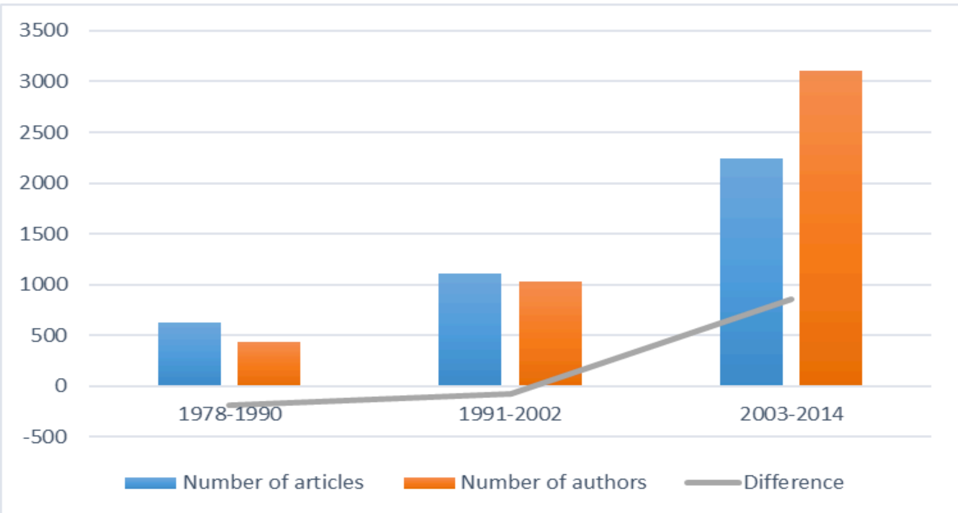


Figure 1. Distribution of the number of articles and authors in three time periods.

Macro-level structure analysis

In order to study the evolution of the scientific collaborations through three time periods, three 1-mode author-to-author matrices were plugged in Ucinet to calculate a variety of network measurements. There are a number of measures which can be used to evaluate the structure of a network. In this study, we will mainly focus on four elements to approach: degree distribution, average degree, average distance, and cluster coefficient.

The number of collaborators that each author has in a collaboration network is the degree of a node (Ding, Rousseau, & Wolfram, 2014). In Figure 2, three lines illustrated the distributions of the node degree in each time span, respectively. The *X*-axial represented the number of authors, and the *Y*-axial represented the degree of the authors. From Figure 2, it can be seen that most authors held the low degree in all three periods. Based on the locations of three distribution lines, more authors tended to join more collaborations from 1978 to 2014 with the increase of the number of total authors published on the journal.

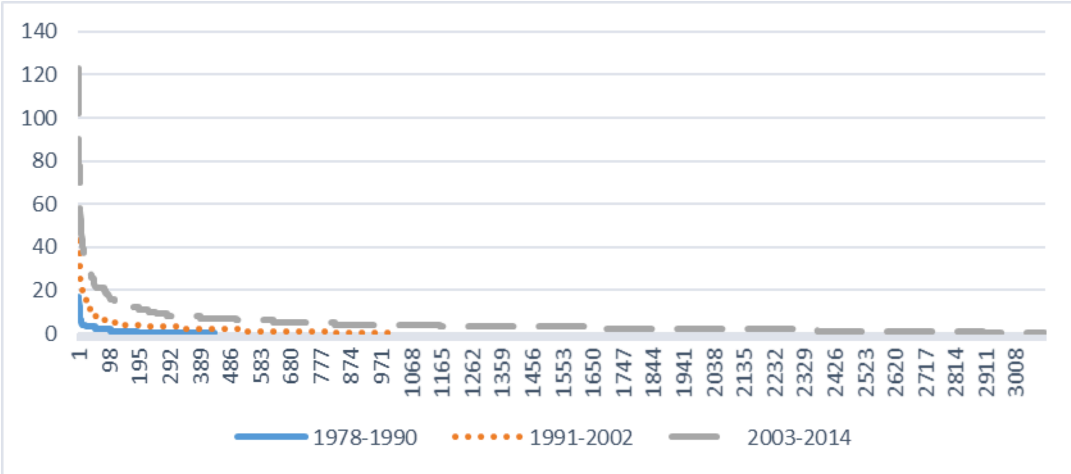
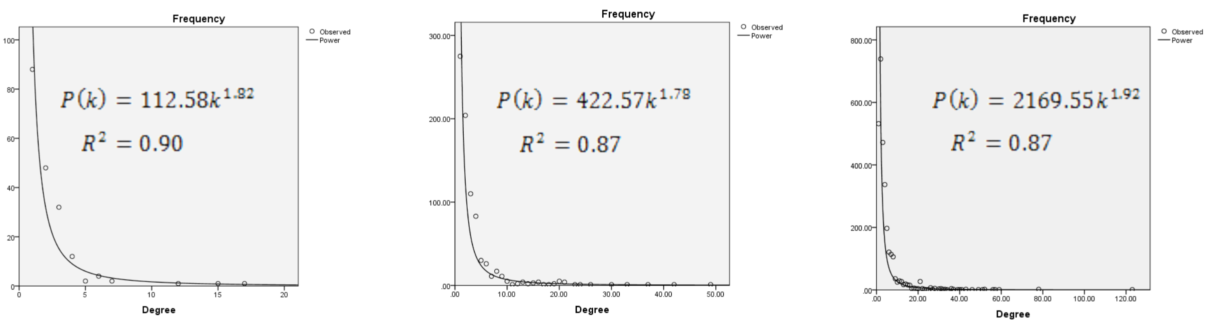


Figure 2. Degree distribution for authors in three time periods.

The degree distribution characterizes the spread of the edges each node has in a network. Although the degree distribution of a random graph is a Poisson distribution, Albert and Barabási (2002) have discovered that, for most large networks, the degree distribution has a power-law tail: $P(k) \sim k^{-\gamma}$, where $P(k)$ is the distribution function. In this study, the distributions of the collaboration network in each period were calculated and drawn in Figure 3. Power-law regression model was used to detect the degree distribution patterns in different timespans (Albert & Barabási 2002). Figure 3 illustrated the modeling results for the three periods, and the x-axis plots low degree nodes on the left and high degree nodes on the right; the y-axis indicates their probability. In both cases, power-law model performed the good fits to the observed data. In relationship between the degree of the authors and the corresponding frequencies can be estimated by: $P(k) = 112.58k^{1.82}$ with $R^2 = 0.90$ in 1978-1990, $P(k) = 422.57k^{1.78}$ with $R^2 = 0.87$ in 1991-2002, and $P(k) = 2169.55k^{1.92}$ with $R^2 = 0.87$ in 2003-2014. As discussed by Albert and Barabási (2002), the degree distribution of the collaboration network of high-energy physicists reach the almost perfect power-law



1978-1990 1991-2002 2003-2014

Figure 3. Degree distribution plots for collaboration networks.

with an exponent of 1.2, while the collaboration networks of mathematicians and neuroscientists between 1991 and 1998 held the degree exponents 2.1 and 2.5 (Barabasi et al., 2002). Comparing with those previous studies in different disciplines, the degree distribution of the collaboration of *Sicentometrics* in each timespan were consistent with power-law with degree exponents 1.82, 1.78, and 1.92, respectively. In addition to degree distribution, previous studies proved that there were several other useful indicators to feature a social network. Table 1 represented the four key measures for each time periods. Figure 3 describes the changes of each measure between 1978 and 2014.

Table 1. Four key measures of the collaboration networks in each time periods.

	1978-1990	*1991-2002*	*2003-2014*
Average Degree	0.794	2.101	3.435
Average Distance	1.412	4.673	7.106
Clustering Coefficient	0.941	0.873	9.014
Components	309	420	701
Diameter	4	11	19

Average degree is calculated by counting the average number of links per author (Barabasi et al., 2002). In the collaboration network, the average degree characterizes the interconnectedness between authors. Yin, Kretschmer, Hanneman, and Liu (2006) identified that the higher the average degree, the tighter the network. From Table 1, we can see that the average degree steadily increased with time, which demonstrated that authors cooperated more often. This results confirmed Barabasi et al.'s (2002) observations in Mathematics and Neuroscience. One possible reason might be the sharp increase of the total number of authors led to more possible connections between the new authors and also between the new authors and the existing authors.

The distance between two nodes is measured by the length of the shortest path between those two nodes. Average distance in a network is calculated by the average length of the geodesic paths between all reachable pairs of nodes (Borgatti, Everett, & Freeman, 2002). From Table 1, the average distance of the collaboration networks started form 1.412 (in 1978-1990), grew to 4.673 (in 1991-2002), and finally reached 7.106 (in 2003-2014). Watts and Strogatz (1998) examined that many social networks show a “small world” phenomenon that have small characteristic path lengths. According to Yin et al. (2006), short average distance allows authors to share information more rapidly. In this case, the average distance of the collaboration network enlarged with time, but actors were still able to reach the others within short paths in all periods. The cluster coefficient for the co-authorship network in *Scientometrics* appeared to have increased sharply: rising from 0.941 in 1978-1990 to 9.014 in 2003-2014.

Micro-level structure analysis

Micro-level structure analysis was adopted to measure the individual authors. One of the main purpose of social network analysis is to identify the core actors in a network. We applied four measures (raw degree, degree centrality, betweenness centrality, and closeness centrality) to investigate the structural characteristics of each author in each timespan.

Table 2 summarized the top 10 authors with highest degrees in each time period. Freeman (1978) defined the degree of a point as the number of other points to which a given point is adjacent. In the collaboration networks, the degree of an author represents the number of authors a given author co-authored with before. Schubert A held the highest degree with 17 in the first period, which showed he cooperated with 17 authors between 1978 and 1990. In both second and third timespan, Glänzel W. achieved the first place with 49 and 123 collaborators in 1991-2002 and 2003-2014, respectively.

Table 2. Raw degree (top 10 authors) in each time period.

1978-1990		*1991-2002*		*2003-2014*	
Schubert, A	17	Glänzel, W	49	Glänzel, W	123
Braun, T	15	Schubert, A	42	Chen, DZ	78
Zsindely, S	12	Braun, T	37	Huang, MH	78
Moed, HF	7	Moed, HF	33	Debackere, K	59
Vanraan, AFJ	7	Gupta, BM	30	Zhang, X	57
Burger, WJM	6	Gomez, I	26	Rousseau, R	56
Courtial, JP	6	Courtial, JP	24	Gorraiz, J	52
Frankfort, JG	6	Rivas, AL	23	Thijs, B	52
Lepair, C	6	Dore, JC	21	Abramo, G	51
Lancaster, FW	5	Miquel, JF	21	D'Angelo, CA	49

Apart from the raw degree of the actors, the centrality is one of the most important structural attributes of social networks (Freeman, 1978). Over the past years, a number of centrality measures have been proposed by sociologists. In the case of co-authorship network, each centrality measure demonstrate special characteristics of the author cooperation. The centrality indicators are designed to identify the "core" authors from different perspectives. The degree centrality can be seen as an index of its potential communication activity. For the co-authorship network, the authors with high degree centrality may result in the status of "elite" (Yin et al., 2006). Freeman's (1978) betweenness centrality is based upon the frequency with which a point falls between pairs of other points on the shortest or geodesic paths connecting them. Regarding to the collaboration, betweenness centrality can be used to assess the potential of an author for control of communication in the knowledge flow network. Tables 3 and 4 summarized the top 10 authors with the highest degree and betweenness centralities in each time period, respectively.

From Table 3, we can see that authors with high degree centrality were dynamic in different timespans. New authors arrived in a field and gathered more collaborations, whereas the existing authors decayed, to some extent, with time. No author ranked in the top 10 in all three time periods. From the perspective of potential communication ability, the "star" of the collaboration networks changed over time. When it comes to the betweenness centrality, Glänzel W was no doubt the core author in both the second and third time periods. Interestingly, from both dimensions (degree centrality and betweenness centrality), Glänzel W occupied the genuine dominator (or "star") position from 2003 to 2014, which suggests that he possesses potential communication ability as well as the possible ability to control the communication between other authors in recent years.

Collaboration network visualization

Figures 4 to 6 present the evolution of the collaboration network in the three stages. Clearly, both the number of the authors and the collaborations boosted, which also illustrated the expansion of this field. With the time advanced, the collaborations between authors were strengthened. To highlight the changes in collaboration, we removed removed isolated nodes in the network in both Figures and displayed only the collaborating authors and their connections. The size of both the nodes and the labels indicated the degree of the authors. The strength of the collaboration was shown by the thickness of the ties between nodes. The authors with high degree in Table 2 were outstanding in the networks.

Table 3. Degree centrality (top 10 authors) in each time period.

1978-1990		*1991-2002*		*2003-2014*	
Courtial, JP	1.379	Moed, HF	1.846	Glänzel, W	1.419
Lepair, C	1.379	Courtial, JP	1.652	Rousseau, R	1.387
Lancaster, FW	1.149	Gupta, BM	1.458	De Moya-Anegon, F	0.967
Braun, T	0.92	Rousseau, R	1.458	Ho, YS	0.935
Dobrov, GM	0.92	Tijssen, RJW	1.458	Borner, K	0.903
Krebs, M	0.92	Glänzel, W	1.361	Park, HW	0.838
Nagy, JI	0.92	Gomez, I	1.263	Thelwall, M	0.838
Plagenz, K	0.92	Rivas, AL	1.263	Chen, DZ	0.838
Porta, MA	0.92	Deshler, JD	1.166	Wu, YS	0.806
Schubert, A	0.92	Gonzalez, RN	1.069	Debackere, K	0.806

Table 4. Betweenness centrality (top 10 authors) in each time period.

1978-1990		*1991-2002*		*2003-2014*	
Braun, T	0.017	Glänzel, W	1.408	Glänzel, W	5.478
Nagy, JI	0.016	Kretschmer, H	1.1	Rousseau, R	3.918
Courtial, JP	0.012	Moed, HF	1.017	Park, HW	2.17
Lepair, C	0.01	Gupta, BM	0.855	Leydesdorff, L	1.661
Schubert, A	0.007	Rousseau, R	0.489	Kretschmer, H	1.478
Dobrov, GM	0.005	Tijssen, RJW	0.397	Ho, YS	1.423
Inhaber, H	0.005	Gomez, I	0.351	Chen, J	1.374
Narin, F	0.005	Luwel, M	0.262	Meyer, M	1.284
Lancaster, FW	0.004	Braun, T	0.261	Huang, JS	1.219
Studer, KE	0.004	Schubert, A	0.259	Aguillo, IF	1.218

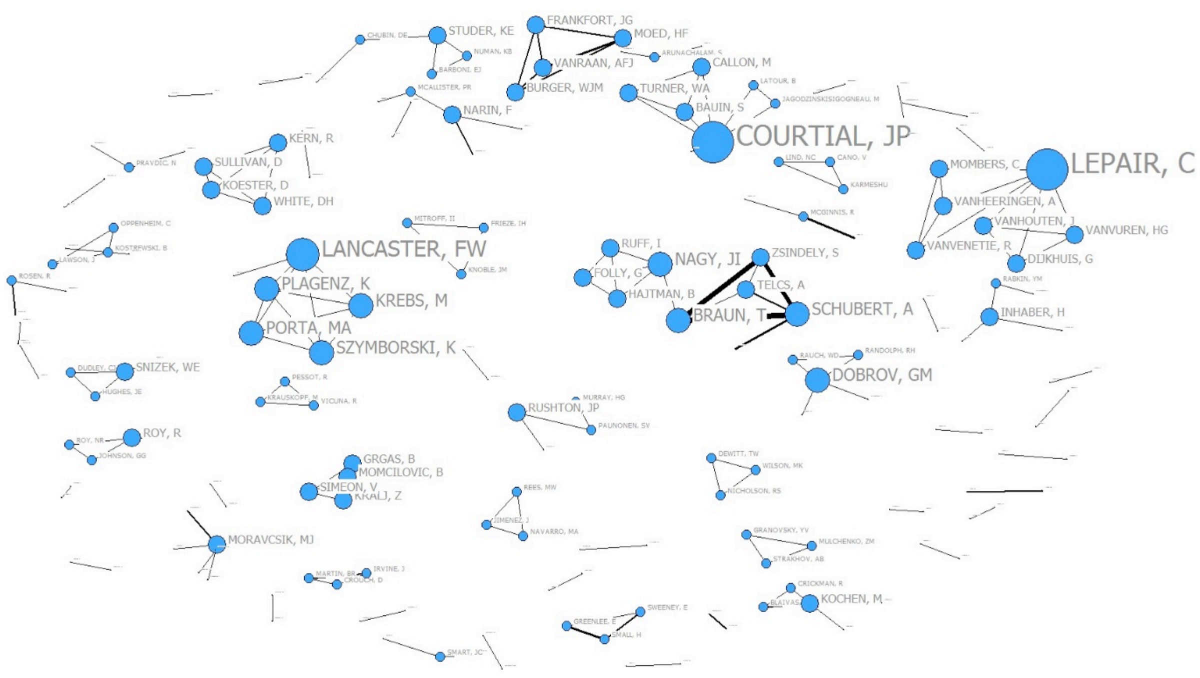


Figure 4. The collaboration networks in 1978-1990.

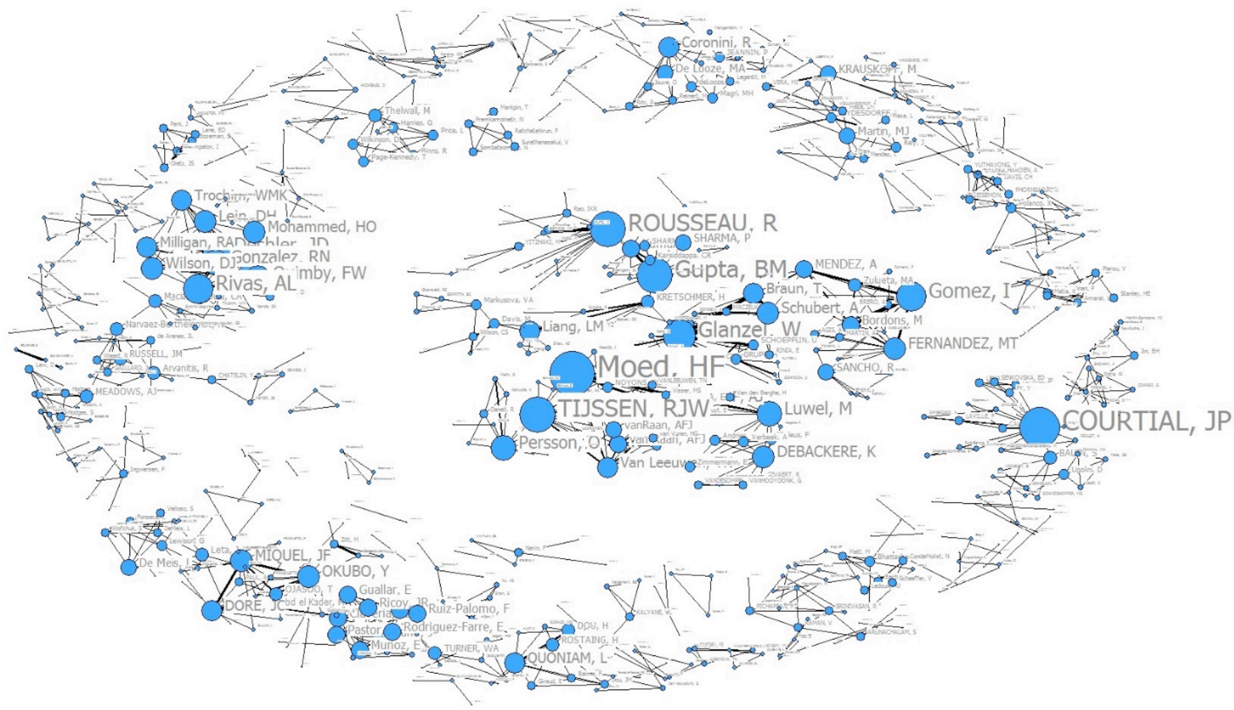


Figure 5. The collaboration networks in 1991-2002.

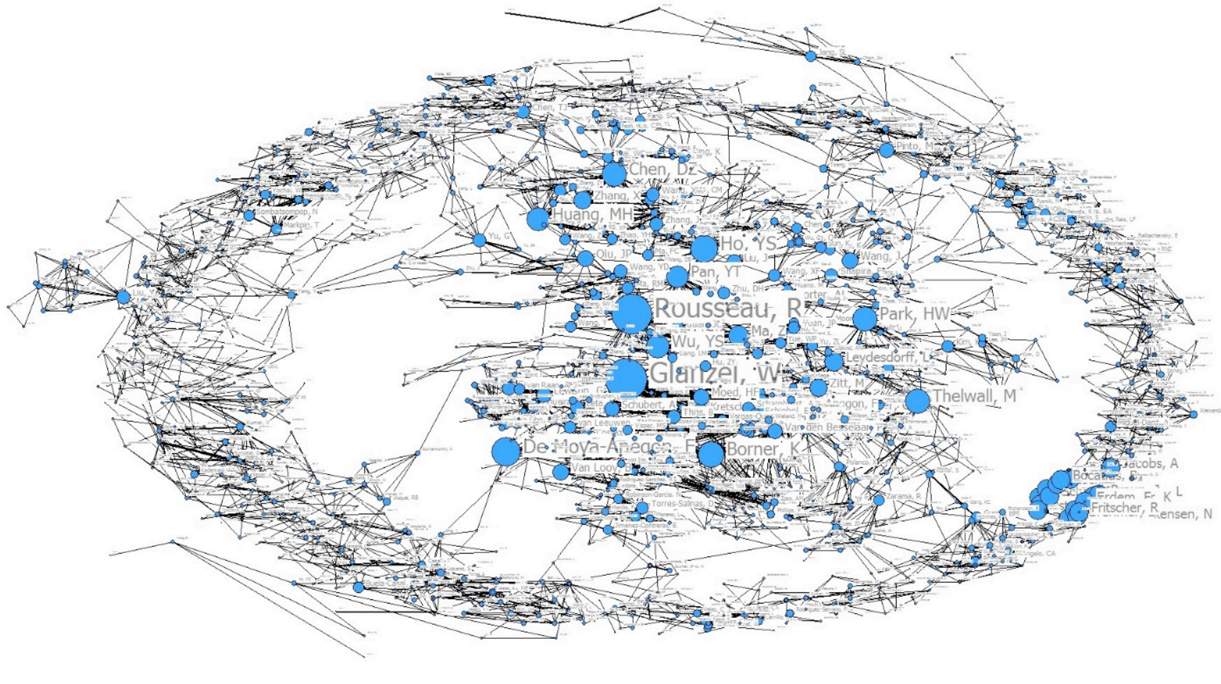


Figure 6. The collaboration networks in 2003-2014.

Conclusion

This paper approached the evolution of the scientific collaboration networks of scientometrics based on the publications in *Scientometrics*. The past 37 years were divided into three timespans: the first time period (1978-1990), the second period (1991-2002), and the third period (2003-2014). Based on the macro-level structure analyses, the degree distribution of the collaboration of *Scientometrics* in each timespan were consistent with power-law, and both the average degree and average distance steadily increased with time, which demonstrated that the cooperation between authors was getting more frequent. Micro-level structure analyses illustrated the authors with high performance in raw degree measure, degree centrality measure, and betweenness measure were dynamic in different timespans. Interestingly, on each dimension, Glänzel W became the genuine dominator (or "star") in the most recent period: 2003-2014. Finally, the visualization of the evolution of the collaboration network in three stages was presented, and the boosts of the number of authors and their collaborators were displayed in the network graphs.

References

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*(1), 47–97. doi:10.1103/RevModPhys.74.47

Ardanuy, J. (2012). Scientific collaboration in Library and Information Science viewed through the Web of Knowledge: the Spanish case. *Scientometrics*, *90*(3), 877–890.

Barabasi, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and Its Applications*, *311*(3-4), 590–614.

Borgatti, S.P., Everett, M.G. & Freeman, L.C. (2002). *Ucinet 6 for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

Breiger, R. L. (1974). The duality of persons and groups. *Social Forces*, *53*(2), 181–190.

Chen, Y., Börner, K., & Fang, S. (2013). Evolving collaboration networks in Scientometrics in 1978–2010: a micro–macro analysis. *Scientometrics*, *95*(3), 1051–1070.

Ding, Y., Rousseau, R., & Wolfram, D. (Eds.). (2014). *Measuring Scholarly Impact: Methods and Practice*. New York: Springer.

Franceschet, M. (2011). Collaboration in computer science: A network science approach. *Journal of the American Society for Information Science and Technology*, *62*(10), 1992–2012. doi:10.1002/asi.21614

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Hou, H. (2006). Study on the evolution of scientometrics based on the scientific map. Retrieved from http://www.cnki.net/

Nalimov, V.V. & Mulchenko, B.M. (1969). *Scientometrics*. Moscow: Nauka (in Russian).

Newman, M. (2001). Scientific collaboration networks. Network construction and fundamental results. *Physical Review E*, *64*(1), 016131.

Pang, J. (2002). *The Research Methodology of Scientometrics*. Beijing, China: Scientific and Technical Documentation Press.

Schoepflin, U., & Glänzel, W. (2001). Little Scientometrics, Big Scientometrics…and Beyond? *Scientometrics*, 30: 375-384.

Schubert, A. (2002). The Web of Scientometrics: A statistical overview of the first 50 volumes of the journal. *Scientometics*, 53(1):3-20.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*(6684), 440–442.

Yin, L., Kretschmer, H., Hanneman, R. A., & Liu, Z. (2006). Connection and stratification in research collaboration: An analysis of the COLLNET network. *Information Processing & Management*, *42*(6), 1599–1613.

Yuan, J. (2010). *The Advanced Tutorial of Scientometrics*. Beijing, China: Scientific and Technical Documentation Press.