Summer 9-21-2015

Cross-time Analysis of Countries Co-authorship Networks in Library & Information Science Research

Elaheh Hosseini
Alzahra university, elahehosseini65@gmail.com

Amin Erfanmanesh
Shahid Beheshti University, amin.erfanmanesh@gmail.com

Cross-time Analysis of Countries Co-authorship Networks in Library & Information Science Research

Mohammadamin Erfanmanesh[1] and Elaheh Hosseini[2]

[1] *amin.erfanmanesh@gmail.com*

Assistant Professor, Department of Information Science, Shahid Beheshti University, Tehran, Iran

[2] *elahehosseini65@gmail.com*

Ph.D. Student, Department of Information Science, Alzahra University, Tehran, Iran

Abstract

This paper studies the role of world countries in Library and Information Science research during 1963 to 2012 using scientometric and social network analysis (SNA) approaches. A total of 58757 papers which published by 83 Information Science and Library Science journals in JCR 2013 and indexed in the Web of Science were selected as the sample of the study. In this paper, the overall structure and evolution of the collaboration network of countries were investigated using macro-level SNA metrics. Additionally, scientometric and micro-level SNA metrics were adopted to analyze the performance of countries in the network. UCINET and VOSVIEWER software were utilized for data analysis and visualization. Findings of the study show that the co-authorship network of countries in LIS research contains 151 vertices which connected together through 3121 links (co-authorships). The collaboration network of countries seems to exhibit "scale-free" and "small world" network properties and the theory of "six degrees of separation" is valid in this network. Moreover, the results of clustering analysis show that this network comprises 39 clusters. Amongst them, the eleventh and ninth clusters which contain US and UK, have the highest density.

Keyword: Library and Information Science, Scientometrics, Social Network Analysis, Countries Collaboration Network.

Introduction

Scientific collaboration among individuals, research organizations and countries has been increased over the past decades. Sharing of knowledge, expertise, equipment, resources and funds, obtaining prestige and visibility as well as providing intellectual companionship are potential factors which motivate research collaboration (Katz & Martin, 1997). Several studies

have reported that collaboration may increase reserach productivity (Barjak & Robinson, 2007). Moreover, associations between scientific collaboration and citation impact have been widely examined; the results generally suggest that the higher the number of authors, the higher the citation impact (Beaver, 2004). Multiple authorship or co-authored publication has been used as the most visible and accessible indicator to measure scientific collaborative activities. Katz and Martin (1997) discussed that accessibility of data, the ease of measurement and stability during the time are as advantages of co-authorship. According to Crane (1972) "the co-authorship of papers creates a social network which can be studied in order to understand the characteristics of a particular field and its invisible colleges". A co-authorship network consists of researchers who have connected to each other if they have co-authored one paper at-least. Such a network can be represented as a set of nodes denoting co-authors joined by links denoting co-authorship.

Co-authorship network analysis have been studied in different fields, such as economics (Krichel & Bakkalbasi, 2006), sociology (Moody, 2004), computer science and information systems (Takeda, 2010), energy (Sakata, Sasaki & Inoue, 2011), health care (Godley, Baron & Sharma, 2011), medicine (Yu, Shao & Duan, 2012) and tourism (Benckendorff 2010). This method is also widely used in library and information science (Pluzhenskaia, 2007; Hou, Kretschmer & Liu, 2008; Yan, Ding & Zhu, 2010; Erfanmanesh, Abrizah & Rohani 2012). Although a few previous studies have studied co-authorship networks of LIS researchers, they are limited in their targeted regions, studied metrics, time span and sample size. Additionally, no previous study analyzes the collaboration network of countries in LIS research. Therefore, comprehensive studies are required to understand the characteristics of co-authorship networks in LIS. This study utilize social network analysis (SNA) to depict scientific collaboration among countries based on 58757 papers published during 1963- 2012 by 83 Information Science and Library Science journals in JCR 2013 and indexed in the Web of Science. This paper aims to utilize the social network analysis method to identify the features of the co-authorship network of countries in LIS research. We will analyse this network with macro-level metrics which capture the global features of the networks as well as micro-level metrics which illustrate the local features of countries in the networks.

Research Methodology

The present research is conducted using scientometric and social network analysis (SNA) methods. We select 83 Information science and library science journals from the JCR 2013 with the time span of 50 years (1963-2012) as the sample of study. During this period, there were

58757 research articles published in IS & LS journals from 151 countries. First, bibliometric data of aforementioned articles were retrieved from the Thomson-Reuters' Web of Science database. Then the dataset was converted into a recognized format of SNA software using a C# application. UCINET and VOSVIEWER software were utilized for data analysis and visualization. The co-authorship network of countries in LIS research was analysed using both macro-level and micro-level metrics. Macro-level metrics studies the overall characteristics of a social network to show its structure; while micro-level metrics focuses on the evaluation of nodes to capture the features of each actor in a network (Yan, Ding & Zhu, 2010). In this study we will focus on the following metrics:

Density: Network density is defined as the total number of observed ties in a network, divided by the total number of possible ties in the same network (Benckendorff, 2010).

Clustering Coefficient: Clustering coefficient indicates the probability that nodes with the same neighbor tend to cluster together (Newman, 2003).

Component: A component is a set of vertices that can be reached by paths running along links of the network (Newman, 2003).

Giant Component: Giant component represent the largest group of nodes who are connected to each other either directly or indirectly (Newman, 2003).

Mean distance: Mean distance is the mean length of the shortest path between two vertices in a network (Yan, Ding & Zhu, 2010).

Diameter: The diameter of a network is the length (edges) of the longest path between any two nodes (Newman, 2003).

Degree Centrality: The degree centrality is defined as the number of an actor's links divided by the maximum possible number (Benckendorff, 2010; Abbasi, Hossain & Leydesdorff, 2012).

Closeness centrality: Closeness centrality is the vertex's average geodesic distance from every other vertex in the network (Benckendorff, 2010; Abbasi, Hossain & Leydesdorff, 2012).

Betweenness centrality: Betweenness centrality is an indicator of an actor's potential control of communication within the network (Benckendorff, 2010; Abbasi, Hossain & Leydesdorff, 2012).

Moreover, the performance of the countries in LIS research was investigated using some scientometric indicators such as the total number of publications, total number of citations received, mean citations per paper, self-citation percentage, citedness rate as well as h-index.

Results and Analysis

An Overview of the Network

The co-authorship network of countries in LIS research consists of nodes and links: nodes represent countries and links connect countries in the form of co-authorships. There is a link between two countries if their researchers have co-authored one IS&LS paper at-least. The size of a node is proportional to the number of co-authorships of that country. The size of the total network in 50 years' time span denotes by the number of unique countries (151) with 3121 international co-authorships (Figure 1).

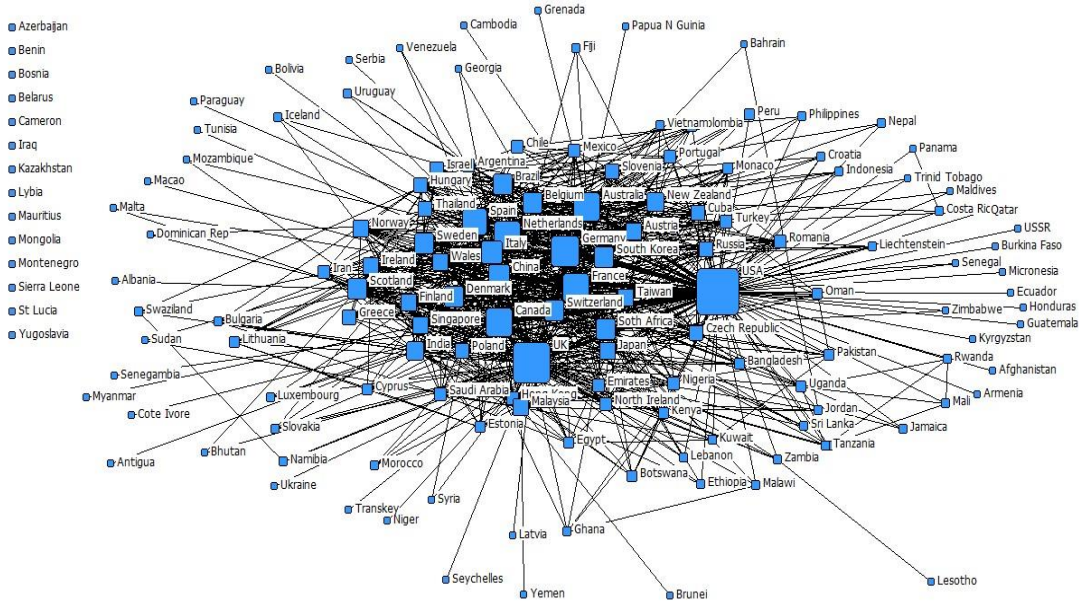

Figure 1. The co-authorship network of countries in LIS during 1963-2012

Macro-level Structure Analysis

Five key elements of the network include density, clustering coefficient, components, mean distance and diameter studied in this paper. Network density shows the relationship between the numbers of actual links against all possible linkages. The density of the co-authorship network of countries in LIS research is 0.082, which indicates only 8.2% of all possible links being present. Another network topology attribute, the clustering coefficient, indicates the extent to which nodes in a network tend to cluster together (Newman, 2003). Considering all nodes of the network, the total clustering coefficient is 0.427, which indicates that the network is clustered. The co-authorship network of countries in LIS research is composed of one large and many small components. This network consists of 15 components, the largest yielded a

ratio of 90.7% of the whole network. It indicates that there is a large group of countries who are interconnected in a cohesive network. Additionally, there are 14 isolate components with size 1 in the network. In fact, they are 14 countries that do not have any co-authorship with the other countries. These countries are Azerbaijan, Benin, Bosnia, Belarus, Cameroon, Iraq, Kazakhstan, Libya, Marie-Tooth, Mongolia, Montenegro, Sierra Leone, St. Louis, and the former Yugoslavia. Studying the average shortest path show that the mean distance between countries in the network is 2.178, suggesting that there are less than three degrees of separation between most countries in the network. Moreover, the network diameter is 4, which means that the farthest countries in the giant component of the network are reachable through 4 steps (Table 1).

Table 1. Macro-level characteristics of countries collaboration network in LIS

Network Parameter	Value
Network Size (No. of Nodes)	151
No. of Links (Co-authorship)	3121
Mean Co-authorship per Country	20.66
Network Density	0.082
Network Connectedness	0.189
Network Fragmentation	0.811
Clustering Coefficient	0.427
Average Mean Distant	2.178
Network Diameter	4
No. of Components	15
Size of Main Component	137 (90.7%)
Isolated Nodes	14
No. of Clusters	39

Evolution of the Countries Collaboration Networks Over 50 Years

As can be seen in Table 2, the countries co-authorship network in the first time span (1963-1967) is made of 10 nods and 1 co-authorship between the United States and Peru. In the second time span (1968-1972) the number of nodes (20) was exactly double that of the previous span and the number of links increased to 4. During next 35 years, the number of nodes grew from 20 to 124 and the number of co-authorship grew to 707. Finally, the size of the network in the last time span (2008-2012) has been increased to 124 countries which are connected through 1320 co-authorship (Table 2). Figures 2-11 show the evolution of the countries collaboration network in LIS research over 50 years.

Table 2. Number of nodes and links in countries collaboration network by 5 year time spans

Time Span	No. of Countries	No. of Co-authorship
1963-1967	10	1
1968-1972	20	4
1973-1977	41	30
1978-1982	57	26
1983-1987	70	98
1988-1992	76	131
1993-1997	85	320
1998-2002	90	484
2003-2007	103	707
2008-2012	124	1320
1963-2012	151	3121

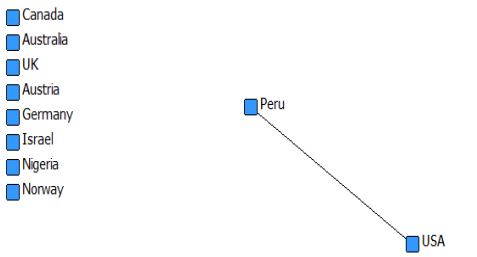

Figure 2. Co-authorship network of countries in LIS research during 1963-1967

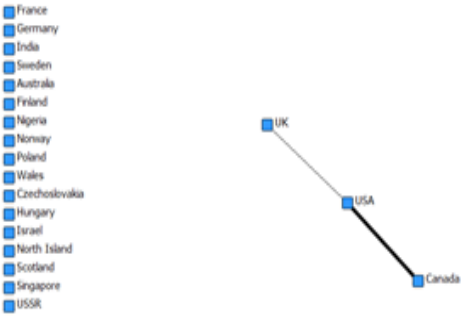

Figure 3. Co-authorship network of countries in LIS research during 1968-1972

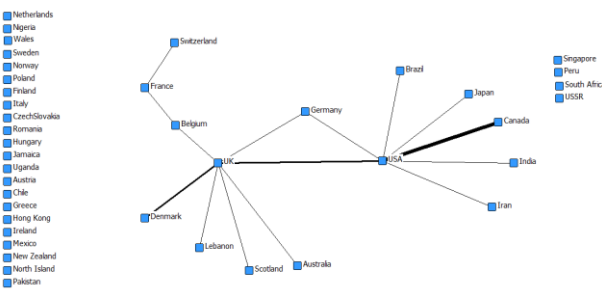

Figure 4. Co-authorship network of countries in LIS research during 1973-1977

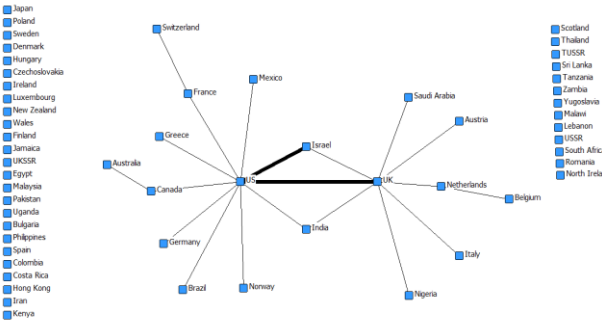

Figure 5. Co-authorship network of countries in LIS research during 1978-1982

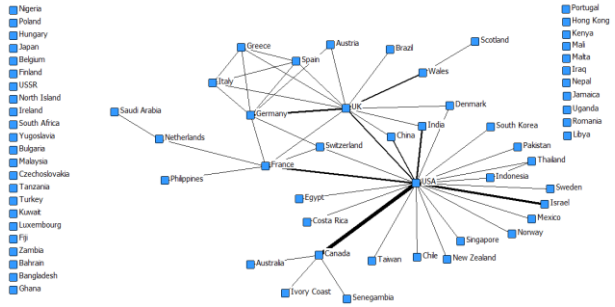

Figure 6. Co-authorship network of countries in LIS research during 1983-1987

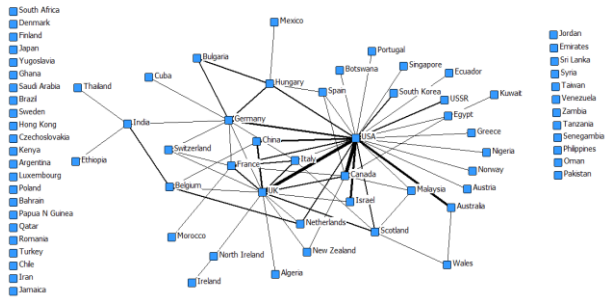

Figure 7. Co-authorship network of countries in LIS research during 1988-1992

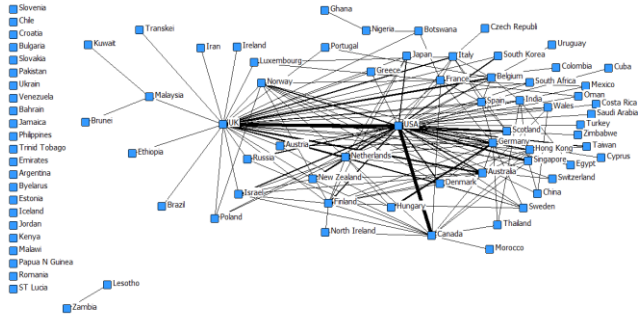

Figure 8. Co-authorship network of countries in LIS research during 1993-1997

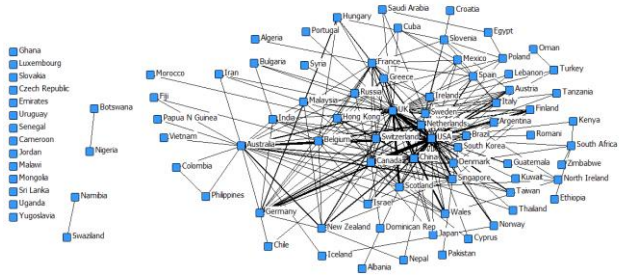

Figure 9. Co-authorship network of countries in LIS research during 1998-2002

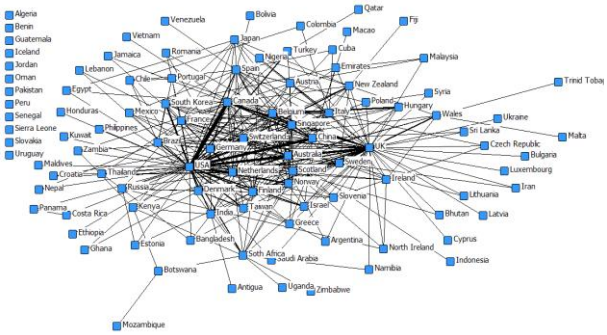

Figure 10. Co-authorship network of countries in LIS research during 2003-2007

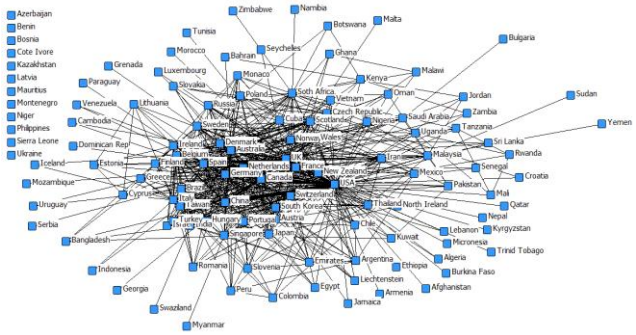

Figure 11. Co-authorship network of countries in LIS research during 2008-2012

Cluster Analysis of the Network

The cluster density visualization of the network in 50 years is shown in Figure 12. A cluster is a set of closely related nodes. Usually cluster analysis can be used to find subgroups in a network. Each node in a network is assigned to exactly one cluster. In this map, each vertex has a color that depends on weight of vertex in the network, number of vertices in the neighborhood and the importance of the neighboring vertices. This colour ranges from red to blue which indicates highest density to lowest density. Moreover, nodes are located closer if they have more co-authorship. Figure 12 shows that U.S.A and the U.K. (red ones) as well as Canada, China, Netherlands, Australia and German (yellow ones) have the highest density in the network. Additionally, cluster analysis shows that the network is formed from 39 different clusters, most of which are fairly small (Table 3).

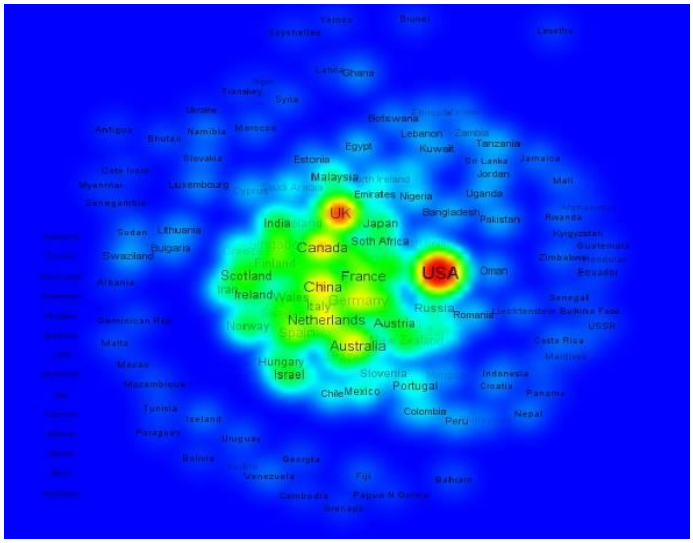

Figure 12. Cluster density map of countries collaboration network in LIS during 1963-2012

Table 3. Cluster of countries in LIS during 1963-2012

Cluster	Size of the Cluster	Countries
1	17	Netherlands, Botswana, Ghana, Kenya, Lesotho, Mali, Malawi, Mozambique, Namibia, Nigeria, Rwanda, South Africa, Swaziland, Tanzania, Uganda, Zambia, Zimbabwe
2	14	Hungary, Bahrain, Brunei, Bulgaria, Egypt, Japan, Jordan, Kuwait, Malaysia, Oman, Pakistan, Saudi Arabia, Sudan, Yemen
3	11	Spain, Argentina, Bolivia, Chile, Colombia, Dominican Republic, Mexico, Paraguay, Peru, Philippines, Venezuela
4	11	Croatia, Cuba, Czech Republic, Denmark, Monaco, Northern Ireland, Poland, Russia, Slovakia, Slovenia, Syria
5	10	Italy, Sweden, Cyprus, Finland, Estonia, Iceland, Ireland, Lithuania, Luxembourg, Norway
6	8	Belgium, Brazil, Fiji, Nepal, New Zealand, Portugal, Romania, Uruguay
7	7	India, Antigua, Bangladesh, Ethiopia, Indonesia, Taiwan, Thailand
8	7	Wales, Scotland, Albania, Bhutan, Malta, Sri Lanka, Trinidad and Tobago
9	6	China, Germany, Georgia, Switzerland, Serbia, Macau
10	5	England, Latvia, Seychelles, Transkei, Ukraine
11	5	United States, Afghanistan, Ecuador, Micronesia, Kyrgyzstan
12	5	Australia, Cambodia, Grenada, Vietnam, Papua New Guinea
13	5	France, Algeria, Jamaica, Senegal, Tunisia
14	4	Canada, Senegambia, Ivory Coast, Lebanon
15	4	South Korea, Austria, Maldives, Liechtenstein
16	4	Iran, Turkey, UAE, Qatar
17	3	Hong Kong, Greece, Morocco
18	3	Singapore, Myanmar
19	2	Panama, Costa Rica
20-39	1	Armenia, Azerbaijan, Benin, Bosnia, Belarus, Burkina Faso, Cameroon, Guatemala, Iraq, Honduras, Kazakhstan, Libya, Mauritius, Montenegro, Mongolia, Israel, Sierra Leone, the Soviet Union, Yugoslavia, St. Louis

Micro-level Structure Analysis

The micro-level analysis of the network involves examining the characteristics and roles of individual countries in the network. United States is the most productive country in LIS research (26915), followed by UK, Canada, Germany and Spain with 5473, 2223, 1682 and 1551 publications, respectively. Top 10 most productive countries are responsible for almost 74% of the world scientific production of LIS. The citations received by LIS publications were also analyzed by country. Based on the findings, American researchers have the most cited publications among researchers from other countries (220681). UK, Canada, Netherlands and Australia came second to fifth with 42329, 28041, 15661 and 11396 citations, respectively. The citation per paper (CPP) in Hungary is the highest with the value of 18.33. Netherlands (13.24) and Hong Kong (13.12) contribute the second and the third followed by Canada (12.61) and

Denmark (11.87). In all, the average rate of CPP for all countries was 4.32. The highest rate of self-citation was belong to Spanish researchers (20.42% of total citations), followed by Iran (18.78%), Nigeria (17.25%), Pakistan (16.37%) and Belgium (14.92). Results of the study showed that Hong Kong clearly has the highest citedness rate (85.07%), followed by Hungary (84.79%) and Denmark (83.28%). Additionally, referring to the h-index value, the United States (81), UK (69), Netherlands (55), Australia (43) and China (42) have the highest value (Table 4).

Scientific collaboration of countries was also studied using the SNA approach, with the aim of capturing the features of each actor in the network using centrality metrics. Centrality measures indicate how central the actor is to the network (Benckendorff, 2010). Three centrality metrics (degree centrality, closeness centrality and betweenness centrality) were used to analyse the co-authorship network of countries in LIS research. The most prolific countries in terms of degree centrality are: United States (2457), UK (1216), Canada (697), China (649) and Netherlands (485). Moreover, table 4 shows the top 20 countries ranked on the standardized closeness centrality measure. The top scorers in terms of closeness are: US (0.06591), UK (0.06548), Canada (0.06472), Germany and Australia (0.06469). In regard to standardized betweenness centrality scores, the most influential countries in this co-authorship network are: US (0.2658), UK (0.1532), Australia (0.0698), Canada (0.0527) and Spain (0.0469). The total number of countries with whom a country collaborated directly was also calculated. The most connected country in the network is United States which has collaboration with 96 different countries, followed by UK (79), Germany (56), Australia (55) and Canada (54). Table 4 presents the top 20 countries based on productivity, citation impact and centrality.

Table 4. Micro-level characteristics of top 20 countries in LIS research

Country	TP	TC	SC	CPP	CP	HI	DC	BC	CC	IN
US	26915	220681	10.46	8.2	70.97	81	2475	0.2658	0.0659	96
UK	5473	42329	12.56	7.73	76.21	69	1216	0.1532	0.0654	79
Canada	2223	28041	6.56	12.61	75.21	27	697	0.0527	0.0647	54
Germany	1682	8261	8.44	4.91	59.03	38	453	0.0376	0.0646	56
Spain	1551	5711	20.43	3.68	59.5	28	350	0.0469	0.0644	48
Australi	1363	11396	6.07	8.36	70.8	43	471	0.0698	0.0646	55
China	1303	9887	10.44	7.59	72.37	42	649	0.0220	0.0642	38
Netherlands	1183	15661	8.96	13.24	83	55	485	0.0414	0.0645	49
France	1048	6152	5.98	5.87	59.82	33	282	0.0383	0.0646	51
Taiwan	791	6429	7.69	8.13	72.18	35	181	0.0206	0.0637	23
S. Korea	647	5420	5.73	8.38	73.57	31	250	0.0033	0.0639	32
Belgium	629	6648	14.92	10.57	81.24	35	183	0.0090	0.0639	32
Scotland	601	3446	6.44	5.73	75.87	24	191	0.0244	0.0640	32
Italy	598	3538	7.71	5.92	68.89	27	185	0.0110	0.0641	37
India	566	2847	13.06	5.03	75.26	21	121	0.0194	0.0639	29
Singapore	499	5314	4.76	10.65	78.75	35	242	0.0143	0.0637	27

Brazil	498	1362	10.64	2.73	37.35	18	124	0.0086	0.0640	32
Finland	487	5406	6.99	11.1	78.43	34	182	0.0019	0.0638	27
Japan	422	1694	8.38	4.01	61.8	18	90	0.0013	0.0636	23
Sweden	393	3065	5.97	7.8	70.73	26	177	0.0066	0.0639	32

TP: Total Number of Publications / **TC**: Total Number of Citations / **SC**: Percentage of Self-citations / **CPP**: Citation per Publication / **CP**: Percentage of Cited Publication / **HI**: H-index / **DC**: Degree Centrality / **BC**: Betweenness Centrality / **CC**: Closeness Centrality / **IN**: Immediate Neighbors

Conclusion

This study investigates the performance of world countries in library and information science research over a 50-years period. Using the data from 58757 papers, we construct the co-authorship network of countries. The key findings of the study are:

a) The collaboration network of countries in LIS research is a "small world network" by demonstrating its short mean distance and scale free properties. A "small world" is a network in which any two nodes are only a few steps apart, regardless of network size.

b) The mean geodesic distance of the network is 2.178, suggesting that the famous notion of "six degree of separation" can be valid in this network.

c) The network also possesses the characteristics of "scale-free networks" in which a few countries collaborate widely while others collaborate with limited number of countries.

d) Two measures (density and clustering coefficient) which have been used to investigate the cohesion of the network indicate relatively loose structure of the countries collaboration network.

e) The co-authorship network of countries in LIS research appears to be quite connected, with a giant component which contains 90.7% of the nodes.

f) Prolific countries like US, UK, Australia, Canada, Germany, Netherlands, Spain and China are ranked high in most of the studied measures, indicating their critical role in LIS research.

It is one of the first studies to analyse collaboration in the field of LIS using co-authorships network of countries. The study has included a time span of five decades for the LIS co-authorship network. The positive evolutions of the network coupled with the presence of a number of key players are evidence of the healthy status of the LIS research community. The results allow scholars in the field of LIS to step back and look at international research collaboration patterns over a relatively long period of time. An overview of the field and the connections between countries provides a useful schematic of invisible colleges for new researchers.

REFERENCES

Abbasi, A., Hossain, L. & Leydesdorff, L. 2012. Betweenness centrality as a driver of referential attachment in the evolution of research collaboration networks. *Journal of Informetrics*, 6 (3), 403-412.

Barjak, F., & Robinson, S. 2007. International collaboration, mobility and team diversity in the life sciences: impact on research performance. *Social Geography Discussions, 3*, 121–157.

Beaver, D. 2004. Does collaborative research have greater epistemic authority? Scientometrics, 60 (3), 399-408.

Benckendorff, P. 2010. Exploring the limits of tourism research collaboration: A social network analysis of co-authorship patterns in Australia and New Zealand tourism research. *Paper presented at the Tourism and Hospitality: Challenge the limits conference,* Tasmania, Australia. 8-11 February 2010.

Crane, D. 1972. *Invisible colleges: Diffusion of knowledge in scientific communities*. Chicago: University of Chicago Press.

Erfanmanesh, M., Rohani, V.A. & Abrizah, A. 2012. Co-authorship network of scientometrics research collaboration. *Malaysian Journal of Library & Information Science*, 17 (3), 73-93.

Godley, J., Barron, G. & Sharma, A. M. 2011. Using social network analysis to assess collaboration in health research. *Journal of Healthcare, Science & the* Humanities, 1 (2), 99-116.

Hou, H., Kretschmer, H. & Liu, Z. 2008. The structure of scientific collaboration networks in Scientometrics. *Scientometrics*, 75 (2), 189-202.

Katz, J. S. & Martin, B. R. 1997. What is research collaboration?. *Research policy*, 26 (1), 1-18.

Krichel, T. & Bakkalbasi, N. 2006. A social network analysis of research collaboration in the economic community. *Paper presented at the International Conference on Webometrics, Informetrics & Scientometrics*, Nancy, France. 10-12 May 2006.

Martinez-Romo, J., Robles, G., Gonzalez-Barahona, M. & Ortuno-Perez, M. 2008. Using social network analysis techniques to study collaboration between a FLOSS community and a company. In Russo, B. Et al. (eds.), *IFIP International Federation for Information Processing*, Volume 275: Open Source Development, Communities and Quality; (Boston: Springer), 171-186.

Moody, J. 2004. The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69 (2), 213-238.

Newman, M. E. J. 2003. The structure and function of complex networks. *SIAM Review*, 45, 167-256.

Pluzhenskaia, M. 2007. Research collaboration of Library and Information Science (LIS) school's faculty members with LIS and non-LIS advanced degrees: multidisciplinary and interdisciplinary trends. *Paper presented at the 8th Conference on Interdisciplinarity & Transdisciplinarity in the Organization of Scientific Knowledge*, Leon, Spain. 18-20 April 2007.

Sakata, I., Sasaki, H. & Inoue, T. 2011. Structure of international research collaboration in wind and solarenergy. *Paper presented at the International Conference on Industrial Engineering & Engineering Management*, Singapore. 6-9 December 2011.

Takeda, H. 2010. A social network analysis of the IS field: A co-authorship network study. *Proceedings of the Southern Association for Information Systems Conference*, Atlanta, USA. 26-27 March 2010.

Yan, E., Ding, Y. & Zhu, Q. 2010. Mapping library and information science in china: a coauthorship network analysis. *Scientometrics*, 83 (1), 115-131.

Yu, Q., Shao, H. & Duan, Z. 2012. The research collaboration in Chinese cardiography and cardiovasology field. *International Journal of Cardiography*. 2012 Mar 26, 1-6.