Macro-level collaboration network analysis and visualization with Essential Science Indicators: A case of social sciences

Dong-Hui Yang[1]*, Yan Wang[1] , Tian Yu[2]* and Xueyu Liu[1]
[1] School of Economics and Management,
Southeast University, Nanjing 210096, CHINA
[2] School of Economics and Management,
Harbin Engineering University, Harbin 150001, CHINA
e-mail: *dhyang@seu.edu.cn (corresponding author); happyyanyan003@163.com;
*yutian@hrbeu.edu.cn (corresponding author); xyliu@gmail.com

ABSTRACT
Cross-national collaboration has been shaped by internationalization of scientific relationships. To study the synergic network of high quality research patterns, this paper collects a total of 300 top 50 items, in each indicator from the big database, Essential Science Indicators, which lists top-ranking papers, scientists and institutions from 2005 to 2015. First, the country level relations of co-authorship addresses in five indicator variables are extracted in the field of social sciences to build international collaboration networks. The social network analysis (SNA) method was applied to calculate the metrics of vertices, edges, average degree, average shortest path, diameter, clustering coefficient and betweenness centrality to illuminate the structural characters and collaboration patterns. Based on the international collaboration similarities, this paper also visualizes the endemic clustering groups of six networks, as cluster dendrograms, using Hierarchical Clustering (HC) method. Findings illustrate that USA, England and Canada are outstanding countries in the international collaboration networks of five indicators. There are geographical groups in European countries in the collaboration networks of scientists, institutes and countries/territories. It is also found that international collaboration contributes to both highly cited papers in the recent 10 years and hot papers in the recent 2 years in this field, rather than geographical similarity does. Those conclusions are critical for policy makers to produce guidelines on how to encourage researchers to build collaboration networks with high-level scholars in different countries.

Keywords: International collaboration; Scientometrics; Social network analysis; Hierarchical clustering; Essential Science Indicators.

INTRODUCTION

International collaboration in research activities is an important structure to provide the panorama of scientific activity among nations. It is more a need than a choice for scientific productivity over the long term (Mason 2020; Sarwar and Hassan, 2015; Ulnicane 2015). From analyzing the collaboration network, one can gain the knowledge diffusion and scientific pattern of a particular research field (Chen, Zhang and Fu 2019; Yang, Hu and Liu 2015). In the last few decades, there has been a growing interest in the analysis of international scientific collaboration (Hsiehchen, Espinoza and Hsieh 2016). International collaboration is classified into three levels: individual/co-authorship (Coccia and Wang 2016;

https://doi.org/10.22452/mjlis.vol25no1.7

da Silva and Muscolo 2012; Frame and Carpenter 1979; Jalal 2019; Liu et al. 2015; Luukkonen, Persson, and Sivertsen 1992; Olmeda-Gomez et al. 2015), subject categories (Butrous 2015; Guo, Zhang and Guo 2016; Jonsen et al. 2013; Palacios-Callender, Roberts and Roth-Berghofer 2016; Peterson 2001; Wagner 2005; Wang, Thijs and Glänzel 2015) and nations (Almeida, Pais and Formosinho 2009; Arunachalam and Doss 2000; Barrios et al. 2019; da Silva and Muscolo 2012; Hayati and Didegah, 2010; He 2009; Hsiehchen, Espinoza and Hsieh 2016; Zitt, Bassecoulard and Okubo 2000). There have been some concerted works in both subject and nation levels by collecting and contrasting large scale data sets (Ronda-Pupo and Katz 2016; Zdravkovic, Chiwona-Karltun and Zink 2016; Zhao et al. 2016). Exploring international co-authorship patterns and analyzing collaboration network are the most widespread method to study the structure of scientific collaboration in the literature (Hou, Kretschmer and Liu 2008). International collaboration is becoming more and more essential to scientific success especially to young researchers because of its effect on (a) learning new techniques and coming up with new ideas; (b) impact and visibility; and (c) greater capacity to carry out research (Francisco 2015). Yet, some researchers question the generalization and propose that not all international collaboration is beneficial (Didegah and Thelwall 2013; Rousseau and Ding 2016; Schmoch and Schubert 2008; Sud and Thelwall 2016). Rousseau and Ding (2016) found that international collaboration did not always yield a higher citation potential for US scientists publishing in highly visible interdisciplinary journals. At times, collaborations with some other nations seem to decrease impact of a certain subject category. Hsiehchen et al. (2016) advocated a greater emphasis on the qualitative aspects of collaborations with efficient mechanisms of assessing research and policy outcomes.

The purpose of this study is to understand what good scientific pattern is and to identify the best peers in a particular field around the globe. Previous literature on such studies use three Web of Science (WoS) scientific databases, particularly the Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) and Arts & Humanities Citation Index (A&CHI) (Zhao et al. 2016). These databases provide retrieval of all journal articles indexed without classifying the quality of the paper. Nevertheless, finding good international collaboration needs to consider highly cited papers that coincide with having a high impact. This requires high quality data to anatomize top papers in many ways. As the main macro-level research tool of Clarivate Analytics, the Essential Science Indicators (ESI) built on the foundation WoS afford a good opportunity to figure out the pattern of international collaboration with high impact and visibility (Wang, Yu and Yu 2011). The ESI assigns each of the science or social science articles included in the WoS database into one of 20 science and 2 social sciences field categories, which covers a rolling 10-year period. These big data concerning top-performing research attracts a great deal of attention. More scholars have put forward to study what the main national and international characteristics and patterns in the ESI database are since 2002. For example, Csajbók et al. (2007) calculated Hirsch-index (*h*-index) for countries (mainly focus on European countries) on various science fields with data obtained from ESI database. Chuang, Wang and Ho (2011) studied the high impact papers presented in the subject category of water resources. Others analyzed the characteristics, network structure, distribution and the growth of scientific research collaboration in China to find international collaboration patterns (Fu et al. 2011; Niu and Qiu 2014). In most of these works, researchers focused on subject categories to compare the numbers of authors cited, numbers of institutes cited, numbers of countries cited, and numbers of subject areas cited. Few scholars, such as Harzing and Giroud (2014) consider the country level collaboration or relations.

Bibliometric analysis is well-utilized especially in presenting scientific output and making comparison with various indicators, models and network theory. Nevertheless, only few of the literature pays attention to collaboration network or groups of countries and fields in the field of social sciences based on ESI database. In several works, co-authorship networks and cluster groups are extracted to discuss the collaboration patterns and country-to-country relationships (Almeida et al. 2009; Perianes-Rodríguez et al. 2009). Although these researchers do not attempt ESI database, the methods are effective to reflect the collaboration strategies. Therefore, in this paper, international collaboration is further examined using network analysis and clustering methods with ESI, focusing in one of the two social sciences categories, i.e. Social Sciences, General. The research questions posed are:

(a) Which countries could be hubs, or more important, in the international collaboration networks?

To address this research question, the degree average shortest path, betweenness and clustering coefficient of each subset are calculated to figure out the key node and their collaboration relationships.

(b) What are the features that could lead to close international collaboration?

To address this research question, the method of hierarchical clustering helps to illustrate the cluster groups of six macro-level collaboration networks.

MATERIALS AND METHOD

As a big data set and relatively new part of WoS, ESI was proposed in 2002, covering top quality papers during a period of successive 10 years in 22 disciplines. Among eight sub-categories of ESI, four belonged to citation rankings according to the total number of citation in the last decade, while two are under the category of most cited papers, one of which is defined as highly cited paper ranking in the top 1% with 10 years citations, and the other is named hot papers based on the recent two years citations with ranking in the top 0.1% when compared to peer papers. Researchers use the ESI database to look up and download information containing influential researchers, institutions, papers, publications, and countries in their field of study - as well as publication performance and emerging research areas that can impact their work.

In this paper, to analyze the international collaboration network and patterns, data of macro-level variables were collected, i.e. country-level data of five variables, namely scientists (researchers), institutions, countries/territories, journals, and most cited papers. Citation performance of papers are measured by two indicators i.e highly cited papers and hot papers. The data of the top 50 items in each variable, in total 300 items, was obtained in the field of Social Sciences, General for 10 consecutive years from 2005 to 2015. Focusing on country-level analysis, the collaboration networks were built according to the co-author relationships. A total of 57, 65, 110, 84, 18 and 68 countries were involved in the collaboration network of scientists, institutions, countries/territories, journals, highly cited papers (last 10 years) and hot papers (last 2 years) respectively. The data that were downloaded mainly include the titles of publications and scientists' affiliation (addresses) information. Each collaborative paper was co-authored with at least 2 countries and had collaboration value in every 2 countries. This gave rise to a symmetric matrix in which heading on the rows and columns were countries with the value of 1 to embody "one time" collaboration relationship. If they collaborated much times, the values were accumulated. The values on the main diagonals without collaboration relationship were set to 0.

Social network describes the relationships between participants. The vertex and edge are two essential elements of a network (Borgatti et al. 2009; Scott 1988). Generally, in the network of international collaboration, the papers published by at least two scientists are considered. The authors and their different country-level addresses form the network reflect the collaboration relationships between countries. There exist two types of social network analysis - static and dynamic (Kossinets and Watts 2006; Yang and Yu 2013). This paper focuses on dynamic analysis to discover the structural regularities of international collaboration network for ten years (from 2005 to 2015). Social network analysis can use nodes and their links to describe their relationship and structure. It is shown that the growth of international co-authorship can be explained based on the organizing principle of preferential attachment (Wagner and Leydesdorff 2005; Wang, Yu, and Yu 2008; 2009). The network of international collaboration feeds back into the national, regional, and local levels, influencing the organization of science. When analyzing the network structure, indicators to reflect features of international collaboration in the ESI database are calculated. Indicators, involving degree, clustering coefficient, density and betweenness centrality (described accordingly below), are usually computed to embody the network characteristics.

Degree: Through calculating the links every node has with the others to measure a node's importance in a certain network. The degree for a node is the number of nodes link with it directly.

Clustering coefficient: Is defined as the probability that a node's neighbors are all connected with each other. It is used to measure the strength of sub-group formation and the density of the network. For an undirected network, it can be expressed as:

$$C_i = \frac{2E_i}{K_i(K_i - 1)}$$

Where K_i is the degree of node *i* and E_i is the total number of links among node's neighbors.

Density: The ratio of exist links and maximum links is carried out to measure whether a network linked closely.

$$Density = \frac{L}{N(N-1)}$$

Where *L* is the total links in a network and *N* presents aggregated vertices.

Betweenness centrality: Is defined as the probability that a node's neighbors are all connected to each other.

To check whether international collaboration has a relation with geographical proximity, the method of hierarchical clustering is adopted to observe the significant groups identified in this paper. Hierarchical clustering is a method for finding the underlying structure of objects through an iterative process that associates or dissociates object by object (Almeida et al. 2007; Murtagh and Contreras 2012). In the first step, two most similar countries make the significant group with maximizing the cosine similarity.

$$\mathrm{sim}(\mathrm{S_i},\mathrm{S_j}) = \max_{\mathrm{d_i} \in \mathrm{S_i}, \mathrm{C_j} \in \mathrm{S_j}} \{\cos(\mathrm{d_i},\mathrm{C_j})\}$$

Where *S* is the cluster, *d* is the element of cluster *S*, C_j is the centroid of S_j.

The next step is to test the distance between the countries with the centroid of existing group, and link the new country into the group. This step is iterated until no country can be added into. Through this method, the scattered objects can be clustered gradually and their collaboration groups are assembled at the end.

RESULTS AND DISCUSSION

In order to understand the collaboration structure more intuitively, network structure and cluster groups are gained using Pajek and R-tools together (Jalal 2019). Collaboration networks are drawn in accordance with the symmetric matrices of five variables in the analysis. The common metrics covering degree, average shortest path, diameter, clustering coefficient and betweenness centrality of international collaboration network were calculated in each variable listed in the top 50 of ESI database i.e., scientist, institutes, countries/territories, highly cited papers and hot papers (Table 1). The vertices are the countries involved in each network, and the edges means the synergic relationships in the same paper referring two or more countries.

Table 1: The Network Metrics Top 50 Items of Five Variables in the ESI Database.

Metrics	Scientists	Institutions	Countries/ Territories	Highly cited papers	Hot papers
Number of vertices	57	65	110	18	68
Number of edges	531	531	1721	34	219
Average degree	18.63	8.17	31.29	3.78	3.22
Average shortest path	1.653	1.663	1.763	2.096	1.442
Diameter	3	4	4	4	3
Betweenness centrality	0.211	0.566	0.102	0.36	0.112
Clustering coefficient	0.700	0.397	0.602	0.513	0.355
Density	0.333	0.126	0.287	0.222	0.474

When considering the variables, all the publications in the ranking top 50 are counted and the addresses used in the collaborative papers were identified. The vertices consist of all participating countries in the collaboration addresses. England, Scotland, North Ireland and Wales are treated as four separate territories and it is the same with Hong Kong and the People's Republic of China. The edges consist of various country-level addresses in the same paper without considering duplications in other papers. However, the link between two countries are computed in the clustering method and shown as cluster dendrogram.

Macro-level International Collaboration Network of Top 50 Scientists

In the collaboration network of scientists as shown in Figure 1, 57 countries form a network with 531 relationships linked with each other, with a ratio 33.27 percent compared with complete utilization. Density characterizes a less relevance network, having a potential 3 times links to reach complete station. While the degree varies from 0 to 48, the network owns an average degree of 18.63, making up 32.68 percent of all the vertices, results in a relatively close collaboration. Scientists with USA addresses have 48 degree and collaborate with 84.21 percent of all countries listed, followed closely by England (46). Among all the vertices, 8.77 percent nodes link with 30 and 23 other nodes, holding the same percentage with 2 degree. The average shortest path is 1.65, meaning that any 2 nodes link with less than 2 path on average. The diameter is 3, which can conduct the countries of top 50 scientists collaborate frequently and only it only need 2 countries to contact at most. The clustering coefficient is 0.7; a node's neighbors seem to be closely cooperated that certain small groups exist in this network.

However, scientists from Austria, Indonesia and Malaysia tend to study solely without any international collaboration in the top 50 list. Except those three countries, the scientists in the top 50 list cooperate with the scientists in other countries more than once. Scientists in three countries, Cyprus, Guatemala and Iran, have collaborative relationship with American scientists and garner higher citations of their papers in the field of social sciences. Figure 1 depicts that USA, England and Canada are the targets of cooperative countries for other countries to edge in top 50 list.

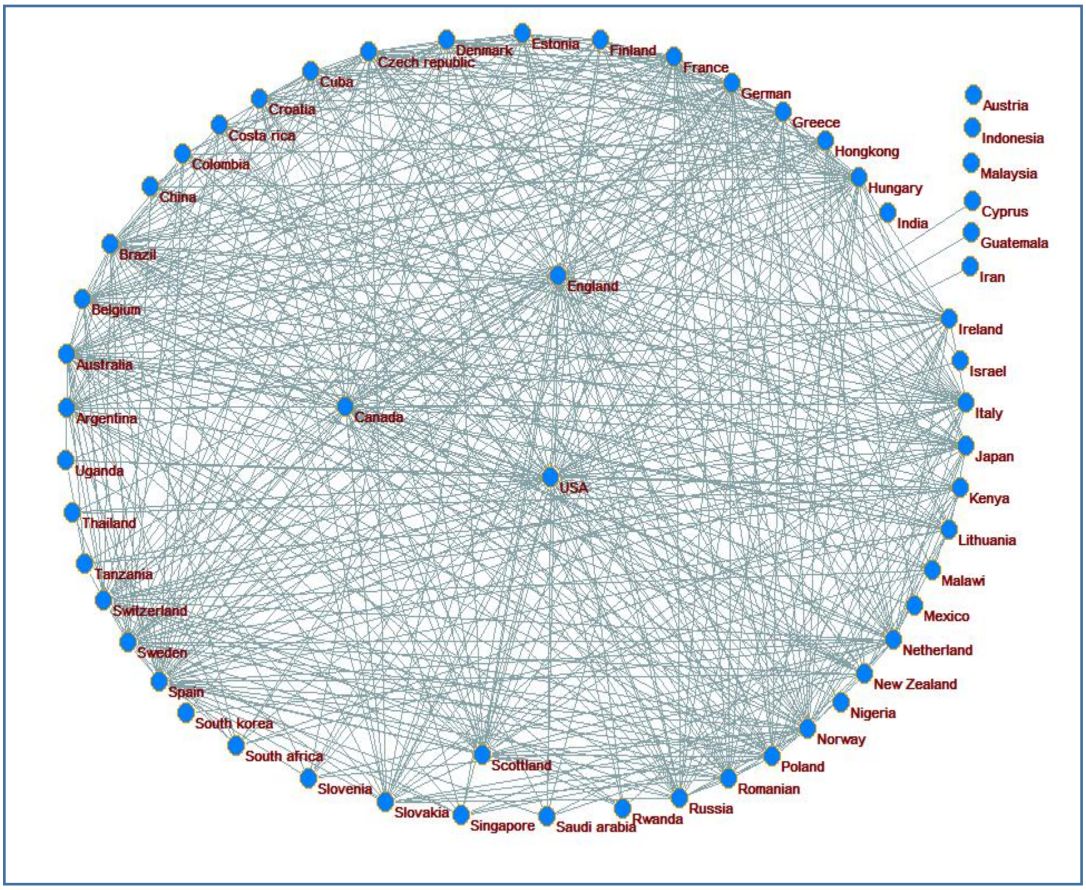

Figure 1: The Macro-level International Collaboration Network of Top 50 Scientists.

To observe the top scientists' community structure in the network requires further clustering research. Hierarchical clustering method is used to illuminate the similar nodes among international collaboration network. The cluster dendrogram (Figure 2) shows the pedigree clustering graph. Scientists from Belgium joined those from Finland first, followed by Greece and then accumulated with another branch involving those from Scotland, Italy and Sweden. Analogously, Australia gathered Norway and Switzerland, the same with Germany, Denmark and France. The two groups initially formed emerged with Spain and later Netherland. The third group gathered England and Canadian-based scientists. The remaining countries almost all from Asia and Africa, compose the fourth group. USA finally joined other two groups. Since there exist no core vertex, small groups get together primarily geographically especially European countries which is consistent with the findings of Almeida et al. (2009) and Harzing and Giroud (2014). This may be caused by regional cooperation between countries which have similar culture.

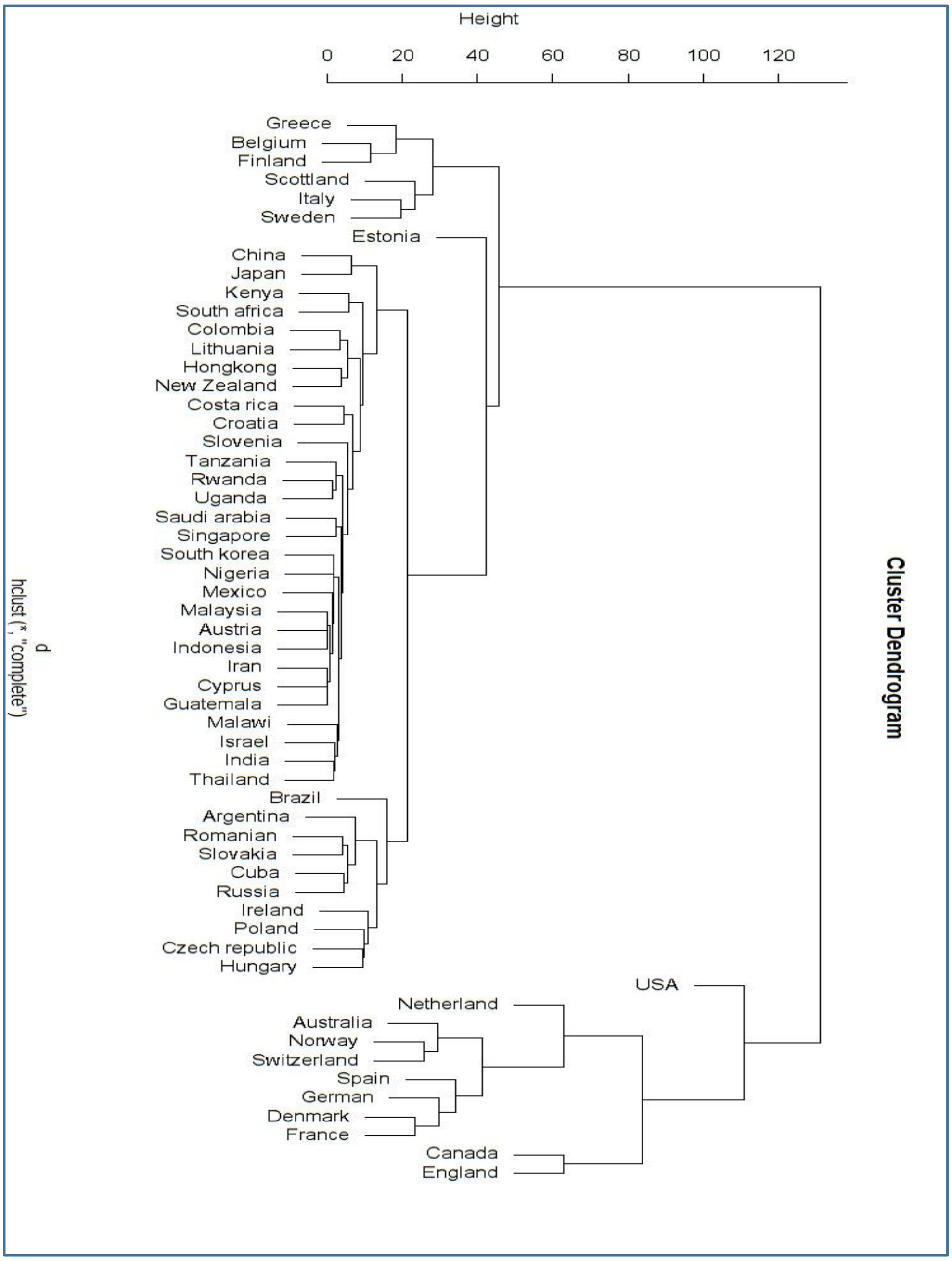

Figure 2: Dendrogaram of Macro-level International Collaboration of Top 50 Scientists.

Macro-level International Collaboration Network of Top 50 Institutions

Next, the metrics of international collaboration network formed by the top 50 papers' addresses of institutions were analyzed. This network contains 65 vertices and 531 edges. For these 65 nodes, the biggest degree of the node representing USA is 57, and it means that the institutions based in USA have the most collaborators among the countries of top 50 institutions. This is followed by England and Australia, with a vertice of 50 and 40 respectively

However, there are some countries' institutions that work with only one collaborator i.e., those from Portugal, Peru, Taiwan, South Korea, Kenya, Gambia, India, Lebanon and Lithuania. The average shortest path is 1.66 that is similar with the network of scientists. The betweenness centrality is 0.0566, showing that this network does not have an apparent central node. It was found that the density of the network is relatively low, with a clustering coefficient of 0.40; that is to say the vertices in the collaboration of institutions are free to link to each other.

The top three institutions consist of two institutions from USA and one from England. Among the top 50 institutions, there are 39 institutions from USA; five from England; two from Canada and Netherlands respectively; and one each from Swedish and French institutions. Based on the findings, it can be concluded that institutions in USA have superior research achievements in the field of social sciences during the past 10 years. The American institutions have built international collaborations with the institutions from England, Canada, and other European countries, which can be seen from the macro-level international collaboration network of top 50 institutions depicted in Figure 3.

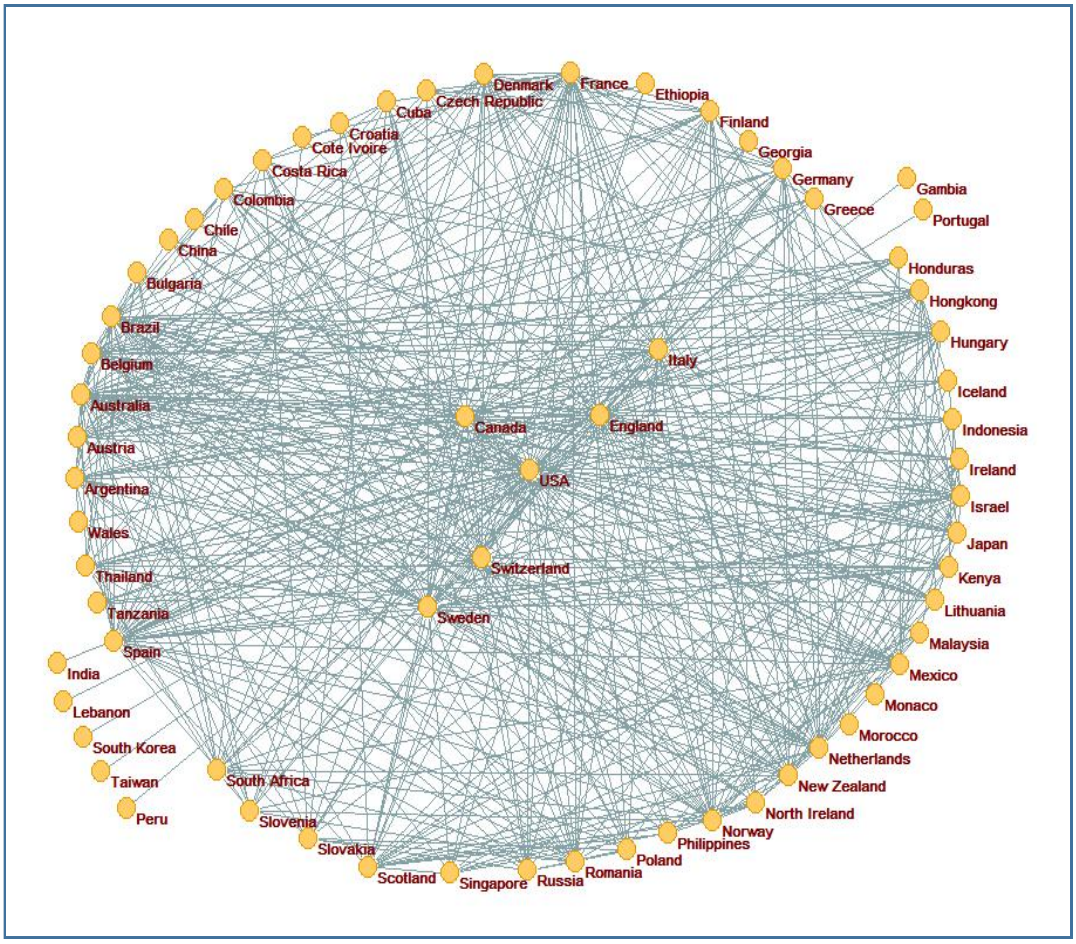

Figure 3: The Macro-level International Collaboration Network of Top 50 Institutions.

Although there is no institution in the top 50 list publishing papers without international collaboration, several institutions tend to rarely collaborate with overseas institutions. For example, the institution in India only has collaboration with that of Spain. The institutions of two countries, Gambia and Portugal, only cooperate with the institutions in

England. Other institutes whose countries also have one collaboration relationship with other countries, such as Lebanon, South Korea, Taiwan and Peru, only establish connection with the institutions based in USA.

The cluster analysis of collaboration between different institutions in Figure 4 shows that USA connects the group involving all the countries of the top 50 institutions. According to the clustering calculation function, USA has no similar institutions with other countries. This illuminates that the American institutions have strong scientific research abilities and sustainability, similar to England which comes second in terms of collaborating with the top 50 institutions. Institutions in Canada, Italy and Switzerland and Sweden linked with other countries' institutions more frequently than others.

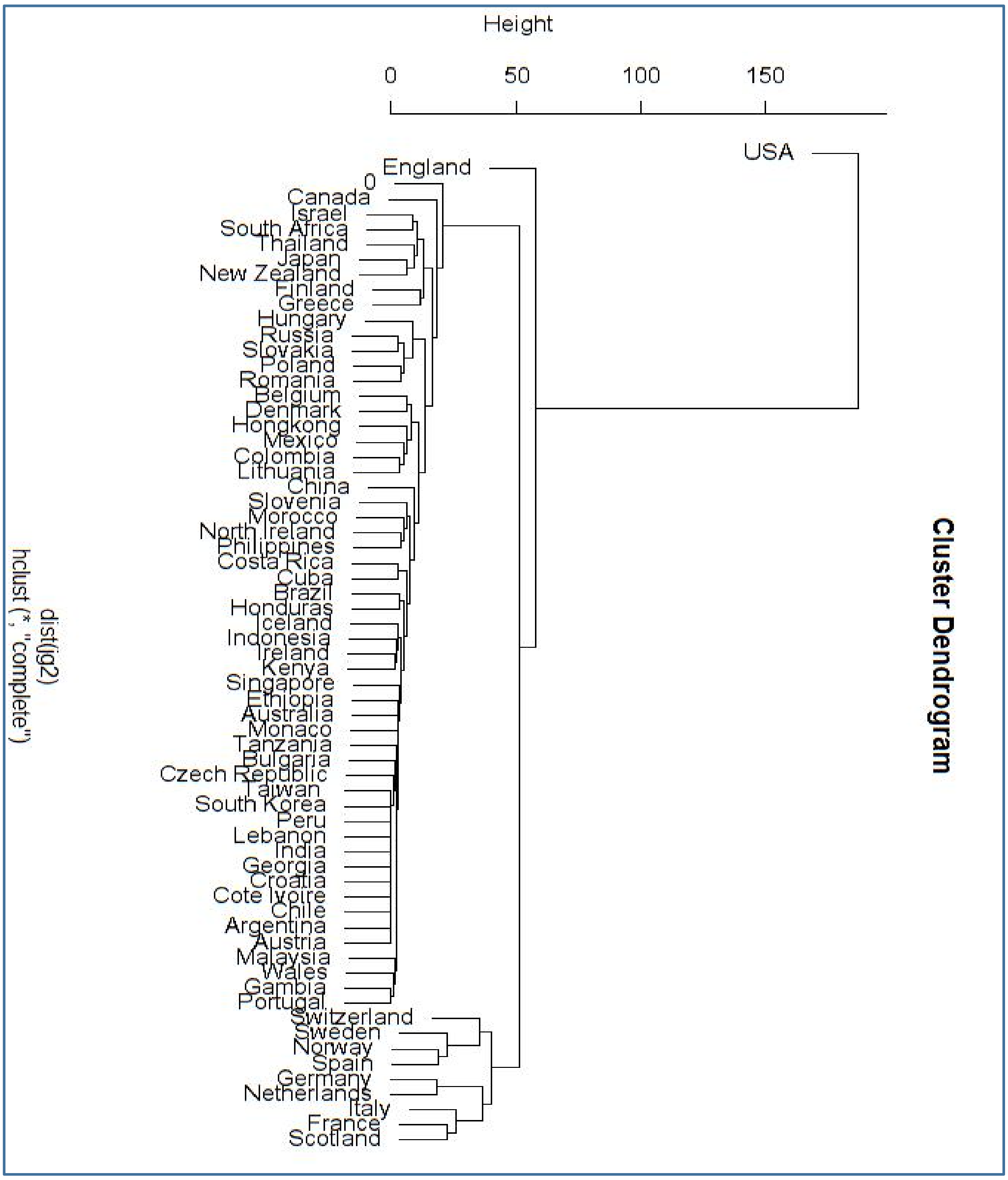

Figure 4: Dendrogaram of International Collaboration of Top 50 Institutions.

Network of Top 50 Countries/Territories

Data set abstracted from the top 50 countries/territories in ESI database consists of 110 nodes which connect 1721 times aggregately. Density with a value of 0.2871 can express the utilization of this network, showing much collaboration potential among those papers of top 50 countries/territories culminated in the last 10 years. Each node has 31.29 relationships on average ranging from 1 to 90. USA gains the highest value of 90, kept in step by England once again, while Yugoslavia is merely linked to Taiwan.

With an average shortest path of 1.76 and diameter 4, international collaboration of countries/territories is frequent and widely spread. It is much higher than the metric values for top 50 scientists and institutions, reflecting that the country-level collaboration is not easy as organization and individual level in the ESI database. Betweenness centrality with a value of 0.102 describes a static network without a certain center node. This value is the lowest among all indicator variables which implies that the node's neighbors of this network are connected to each other with the lowest extent. Most of the countries/territories have at least two connections with others countries/territories except for Yugoslavia (Figure 5).

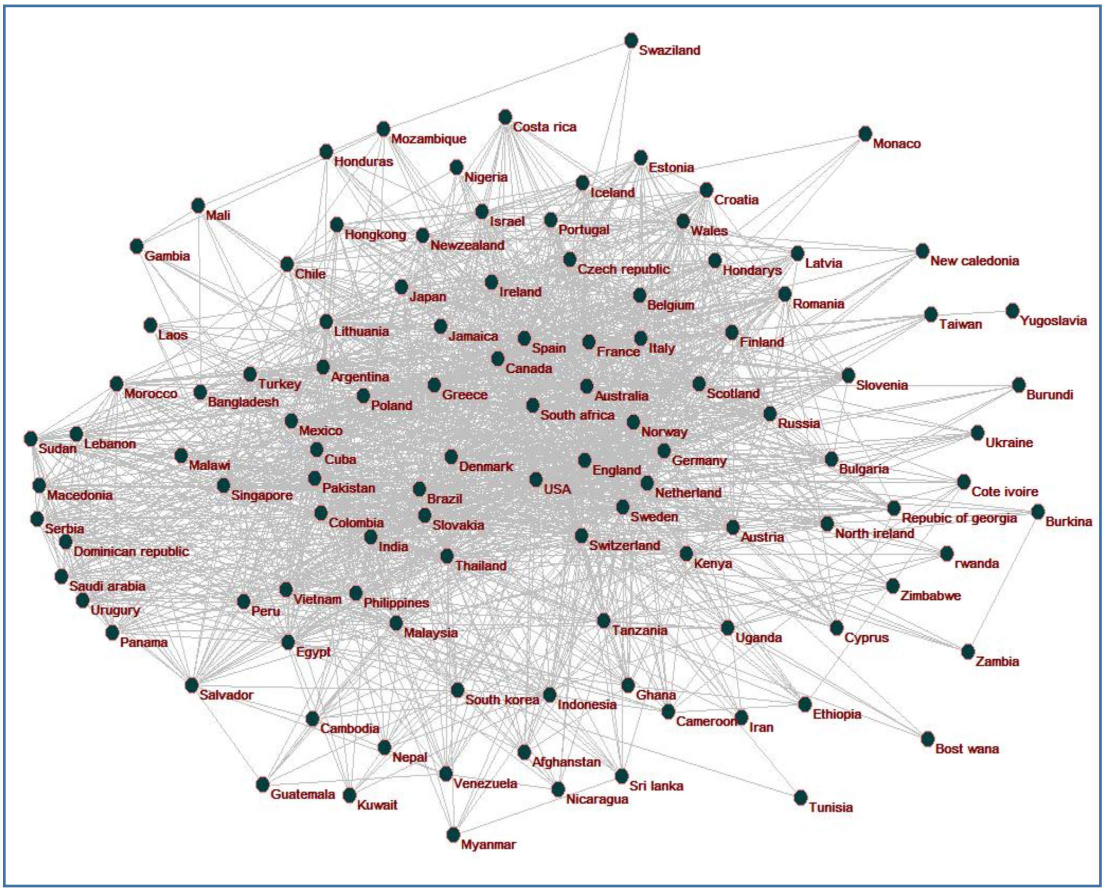

Figure 5: The Macro-level International Collaboration Network of Top 50 Countries/Territories.

From the cluster dendrogram of top 50 countries/territories, a palpable cluster groups could not be found in Figure 5. Similar with the findings on top 50 scientists, USA and England collaborated with most countries due to their strong academic research ability and high quality scientific output among the 110 countries/territories. USA ranks first with 289100 papers; a distant second is England with 84140 papers, which is less than one third of that of the USA; then followed by Canada, Australia and the Netherlands (Table 2). Interestingly, 24 out of 50 countries ranked based number of papers are located in Europe, which also shows strong international collaboration with each other. That can verify the supposition of geographically collaborative groups among European countries.

Table 2: Top 20 Countries/Territories in the Field of Social Sciences

No	Countries/territories	Citations	Papers	CPP	HCP	TP
1	USA	2255828	289100	7.8	4188	4207
2	ENGLAND	637838	84140	7.58	1291	1301
3	CANADA	315910	41946	7.53	634	635
4	AUSTRALIA	258125	42306	6.1	544	548
5	NETHERLANDS	228281	26145	8.73	528	530
6	GERMANY	180935	29400	6.15	472	473
7	SWEDEN	120247	15468	7.77	251	252
8	FRANCE	110012	16723	6.58	301	304
9	CHINA	107037	19227	5.57	203	203
10	SPAIN	98424	22144	4.44	215	217
11	SWITZERLAND	93187	9427	9.89	283	284
12	ITALY	89134	13297	6.7	201	202
13	SCOTLAND	80821	10788	7.49	175	177
14	DENMARK	75006	8365	8.97	208	209
15	NORWAY	71827	9470	7.58	163	165
16	BELGIUM	65087	9807	6.64	132	133
17	BRAZIL	56529	14916	3.79	85	88
18	SOUTH AFRICA	54871	10886	5.04	136	136
19	FINLAND	54848	7111	7.71	118	118
20	JAPAN	52033	8617	6.04	80	80

Note: CPP = Cites/Paper; HCP = Highly cited papers; TP = Top papers.

Macro-level International Collaboration Networks of Top 50 Highly Cited Papers and Hot Papers

The top 50 highly cited papers and top 50 hot papers that most researchers are concerned with in the field of Social Sciences, General were also examined. After collecting their vertices and edges from ESI database, the information in each category of the metrics were obtained. The international collaboration network of top 50 highly cited papers consists of 18 countries and a total of 34 links. These two values are the lowest among all five indicator variables. Only 18 countries whose papers enter into the top 50 list as highly cited papers. England with the highest degree value of 11 is most likely to be in international collaboration, while New Zealand works alone in social sciences research. The value of average degree also demonstrates that this network has the potential to link with each other when compared with other indicators. Betweenness centrality is used to manifest the central node in a network which values according to the times one vertex go-between. It can judged from the value that this network is less likely to block up. And simultaneously, they would be listed in ESI database without cooperating with a certain core node (Barrios et al. 2019). Clustering coefficient complains the density of a network with the value of 0.5, indicating that vertices are relatively free to link with each other.

Nevertheless, highly cited papers in social sciences at the macro-level cover only 18 countries; it can be seen that only adequate academic research capability can keep the countries in the leading position of highly cited papers no matter how closely they collaborate with other countries. Among the top 50 highly cited papers, three are a collaboration between USA and England, and another three between USA and Canada. It is clear that the international collaborations of USA, England and Canada produce high quality papers, especially England has a total of 15 connections as shown in Figure 6. Meanwhile, only one paper with a New Zealand address publishes highly cited papers without international collaboration. Hence, to improve the impact of research paper, international collaboration is a significant strategy.

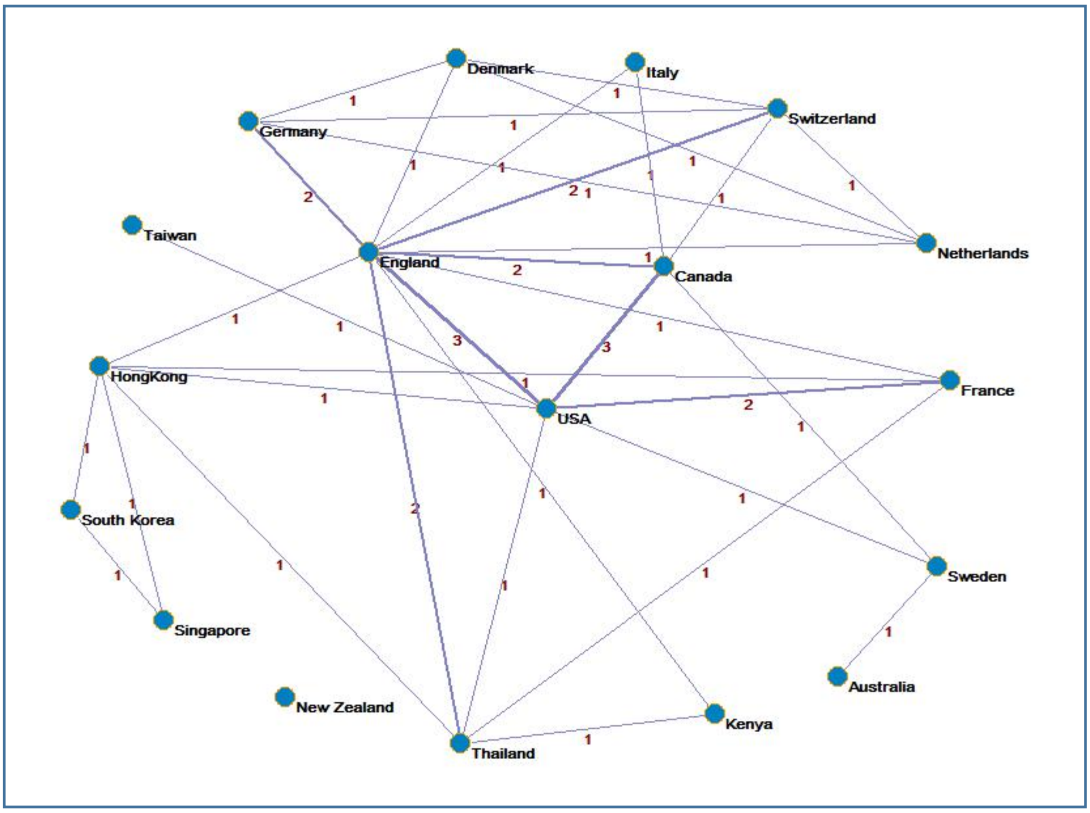

Figure 6: The Macro-level International Collaboration Network of
Top 50 Highly Cited Papers.

The country level data of top 50 hot papers are used to draw the international collaboration network as shown in Figure 7. The network has 68 countries and 219 links. The countries that had the most collaboration in social sciences are USA, England and Netherlands with the degree of 31, 26, and 24 respectively. This network has the lowest average shortest path with the value of 1.44 among all five indicator variables. The diameter is 3, showing that the longest path between two nodes in the network is 3, i.e. from Portugal to Finland.

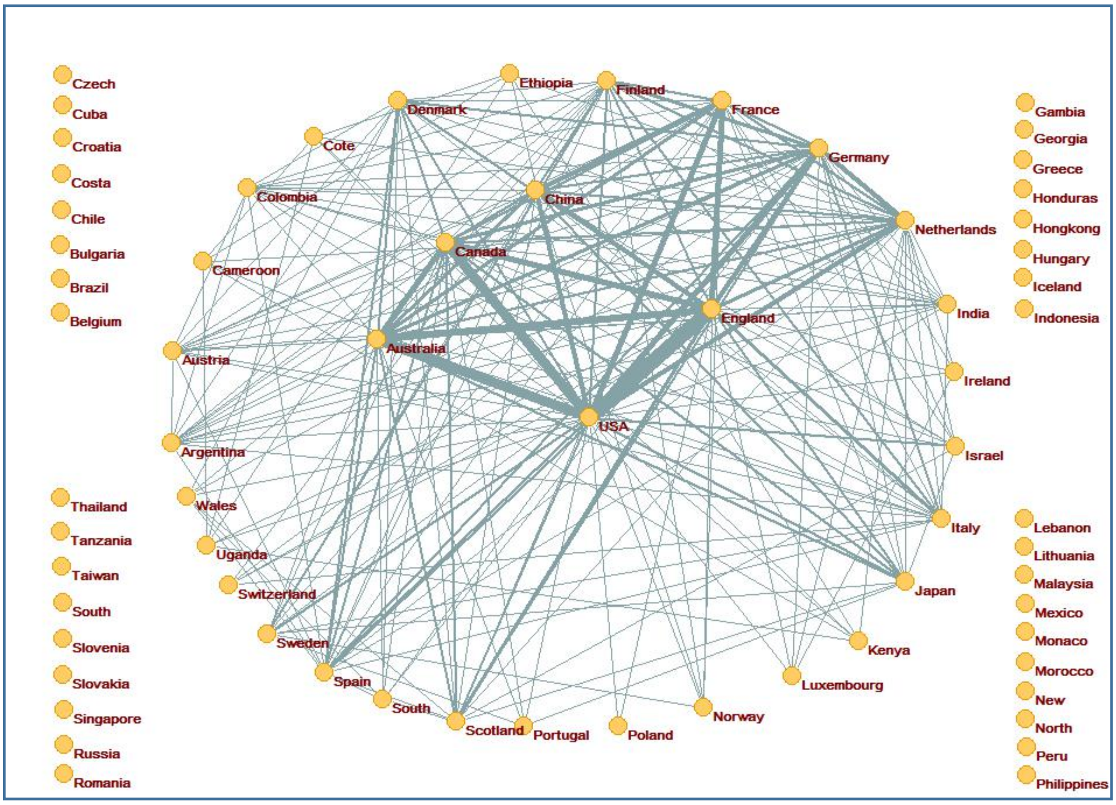

Figure 7: The Macro-level International Collaboration Network of Top 50 Hot Papers.

These metrics mean that multinational collaborated papers will induce higher citations in the recent 2 years. The co-authors have the impact in their own countries that results in larger citations and enter into hot paper list. The international collaborations between USA and England, USA and Canada, USA and Australia are three frequent links among the top 50 hot papers. It means that these three countries contribute significantly in the recent 2 years in the field of social sciences. The papers of some countries that have no connections with other countries also gain higher citation in the recent 2 years. These countries of hot papers with relatively low citations have similar impacts, which is illustrated in Figure 7 as isolated nodes.

This phenomenon can also be seen from the cluster dendrogram in Figure 8. There is a big cluster from Thailand to Belgium joining into one group in which the countries are involved only once. However, when the cluster height increases, the results of Canada, Netherlands, England and USA are gained with different weights and collaboration links. Therefore, international collaboration has the superiority to gain higher citations in the recent two years as hot papers.

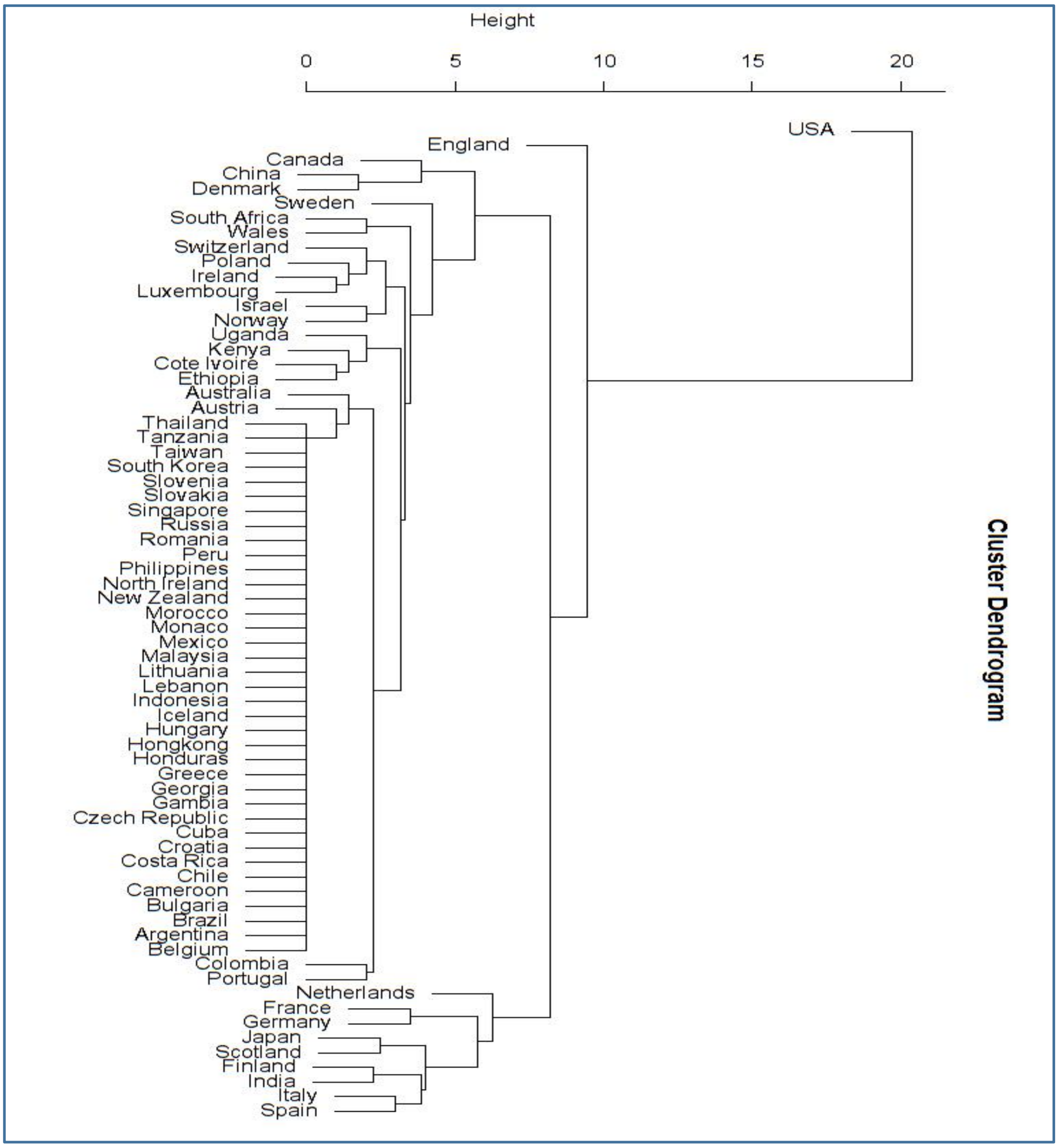

Figure 8: Dendrogaram of Macro-level International Collaboration of Top 50 Hot Papers

CONCLUSION

To study the international collaboration network of higher quality research in a certain subject category, social network analysis and hierarchical clustering are used in this paper. With original data extracted from ESI database in the field of social sciences, the authors analyze the international collaboration network of top 50 items for five variables. The countries in each indicators are the vertices of the networks, and the links among different countries appeared in the same paper consist of the edges of those networks. The authors illustrate these networks based on calculating six metrics associated with the variables and discuss the international collaboration in each network. Various network visualizations in the five variables are realized to present structural analysis directly. Based on the findings, the following conclusions are made:

(a) In order to be ranked top 1% in each indicator variable in the field of social sciences, one should improve collaboration with scientists especially from USA, England, Canada, Netherland and Australia, as well as the collaboration times (number of collaboration).

Although analysis demonstrates that there exists no core vertex in each subset, along with the concrete data downloaded from ESI database, data show that USA, England, Canada, Netherland and Australia head in each top list and also hold the most value in metric degree. Strong academic research ability and frequent scientific collaboration lead these countries to the top.

(b) USA's openness to international collaboration attracts the cooperation with many countries. The American-based scientists and institutions have built international collaborations with the institutions in England, Canada, and other European countries. Findings reveal that USA, England and Canada are the targets of cooperative countries for other countries to edge in the top 50 list. In most cases of five indicator variables, USA has the largest links with other countries in various international collaboration networks.

(c) Europe is another important research group. The clustering graphs confirm the findings that small groups existed especially in Europe. Hierarchical clustering method is used to reveal the groups of similar collaboration model. European countries formed 1 or 2 groups which collaborated frequently due to strong scientific research ability and geographically adjacency. The majority of countries with high total publications are located in Europe, which also reflects their strong international collaboration with each other. In addition, belonging to Anglophone sphere strengthen their global connection and collaboration.

(d) To improve the impact of high quality research paper, international collaboration is a significant approach as a measure of scientific excellence. International collaboration has the superiority to gain papers to be top of their field for citation impact - as highly cited papers and hot papers. The authors reach to the conclusion that publication ranked in the top list in social sciences also needs international collaboration with the scientists from the top 50 countries and institutes for the social science researchers to produce better results rather than working alone in a research project.

Therefore, strengthening national academic research capability through international collaboration and attempting to form a group in which members could make progress and share achievements jointly is necessary. Highly-cited and hot papers published in medical and health sciences journals with hotspot issues are found to publish popular topics in the field of social sciences. For future work, the international collaboration network analyses of five indicator variables in specific social science subject categories could be done to find out important collaboration patterns.

ACKNOWLEDGEMENT

Many thanks to two reviewers for giving useful suggestions to refine this paper. This work was supported by the National Natural Science Foundation of China (Grant No. 71871053), the China Scholarship Council and the Fundamental Research Funds for the Central Universities.

REFERENCES

Almeida, J. A., Pais, A., and Formosinho, S. J. 2009. Science indicators and science patterns in Europe. *Journal of Informetrics*, Vol. 3, no. 2: 134-142.

Almeida, J. A. S., Barbosa, L. M. S., Pais, A., and Formosinho, S. J. 2007. Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering. *Chemometrics and Intelligent Laboratory Systems*, Vol. 87, no. 2: 208-217.

Arunachalam, S., and Doss, M. J. 2000. Mapping international collaboration in science in Asia through coauthorship analysis. *Current Science*, Vol. 79, no. 5: 621-628.

Barrios, C., Flores, E., Martinez, M. A., and Ruiz-Martinez, M. 2019. Is there convergence in international research collaboration? An exploration at the country level in the basic and applied science fields. *Scientometrics*, Vol. 120, no. 2: 631-659.

Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. 2009. Network Analysis in the Social Sciences. *Science*, Vol. 323, no. 5916: 892-895.

Butrous, G. 2015. International research collaboration: the key to combating pulmonary vascular diseases in the developing world. *Pulmonary Circulation*, Vol. 5, no. 3: 413-414.

Chen, K. H., Zhang, Y., and Fu, X. L. 2019. International research collaboration: An emerging domain of innovation studies? *Research Policy*, Vol. 48, no. 1: 149-168.

Chuang, K.-Y., Wang, M.-H., and Ho, Y.-S. 2011. High-impact papers presented in the subject category of water resources in the essential science indicators database of the institute for scientific information. *Scientometrics*, Vol. 87, no. 3: 551-562.

Coccia, M., and Wang, L. 2016. Evolution and convergence of the patterns of international scientific collaboration. *Proceedings of the National Academy of Sciences*, Vol. 113, no. 8: 2057-2061.

Csajbók, E., Berhidi, A., Vasas, L., and Schubert, A. 2007. Hirsch-index for countries based on Essential Science Indicators data. *Scientometrics*, Vol. 73, no. 1: 91-117.

da Silva, J. A. T., and Muscolo, A. 2012. International collaboration, co-operation and partnerships in science writing: Focus on Italy. *Romanian Biotechnological Letters*, Vol. 17, no. 2: 7043-7048.

Didegah, F., and Thelwall, M. 2013. Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, Vol. 7, no. 4: 861-873.

Frame, J. D., and Carpenter, M. P. 1979. International research collaboration. *Social Studies of Science*, Vol. 9, no. 4: 481-497.

Francisco, J. S. 2015. International Scientific Collaborations: A Key to Scientific Success. *Angewandte Chemie-International Edition*, Vol. 54, no. 50: 14984-14985.

Fu, H.-Z., Chuang, K.-Y., Wang, M.-H., and Ho, Y.-S. 2011. Characteristics of research in China assessed with Essential Science Indicators. *Scientometrics*, Vol. 88, no. 3: 841-862.

Guo, S. S., Zhang, G. Z., and Guo, Y. F. 2016. Social network analysis of 50 years of international collaboration in the research of educational technology. *Journal of Educational Computing Research*, Vol. 53, no. 4: 499-518.

Harzing, A.-W., and Giroud, A. 2014. The competitive advantage of nations: An application to academia. *Journal of Informetrics*, Vol. 8, no. 1: 29-42.

Hayati, Z., and Didegah, F. 2010. International scientific collaboration among Iranian researchers during 1998-2007. *Library Hi Tech*, Vol. 28, no. 3: 433-446.

He, T. W. 2009. International scientific collaboration of China with the G7 countries. *Scientometrics*, Vol. 80, no. 3: 571-582.

Hou, H., Kretschmer, H., and Liu, Z. 2008. The structure of scientific collaboration networks in Scientometrics. *Scientometrics*, Vol. 75, no. 2: 189-202.

Hsiehchen, D., Espinoza, M., and Hsieh, A. 2016. Hypoallometric scaling in international collaborations. *Physica a-Statistical Mechanics and Its Applications*, Vol. 444: 188-193.

Jalal, S. K. 2019. Co-authorship and co-occurrences analysis using Bibliometrix R-package: a case study of India and Bangladesh. *Annals of Library and Information Studies*, Vol. 66, no. 2: 57-64.

Jonsen, K., Butler, C. L., Makela, K., Piekkari, R., Drogendijk, R., Lauring, J., Lervik, J. E., Pahlberg, C., Vodosek, M., and Zander, L. 2013. Processes of international collaboration in management research: A reflexive, autoethnographic approach. *Journal of Management Inquiry*, Vol. 22, no. 4: 394-413.

Kossinets, G., and Watts, D. J. 2006. Empirical analysis of an evolving social network. *Science*, Vol. 311, no. 5757: 88-90.

Liu, J., Li, Y., Ruan, Z., Fu, G., Chen, X., Sadiq, R., and Deng, Y. 2015. A new method to construct co-author networks. *Physica a-Statistical Mechanics and Its Applications*, Vol. 419: 29-39.

Luukkonen, T., Persson, O., and Sivertsen, G. 1992. Understanding patterns of international scientific collaboration. *Science, Technology and Human Values*, Vol. 17, no. 1: 101-126.

Mason, S. 2020. Adoption and usage of Academic Social Networks: a Japan case study. *Scientometrics*, Vol. 122, no. 3: 1751-1767.

Murtagh, F., and Contreras, P. 2012. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, Vol. 2, no. 1: 86-97.

Niu, F. G., and Qiu, J. P. 2014. Network structure, distribution and the growth of Chinese international research collaboration. *Scientometrics*, Vol. 98, no. 2: 1221-1233.

Olmeda-Gomez, C., Antonia Ovalle-Perandones, M., and de Moya-Anegon, F. 2015. Analysis of research collaboration between universities and private companies in Spain based on joint scientific publications. *Information Research - an International Electronic Journal*, Vol. 20, no. 4, paper 692. Available at: http://informationr.net/ir/20-4/paper692.html#.XpLdX8gzZPY.

Palacios-Callender, M., Roberts, S. A., and Roth-Berghofer, T. 2016. Evaluating patterns of national and international collaboration in Cuban science using bibliometric tools. *Journal of Documentation*, Vol. 72, no. 2: 362-390.

Perianes-Rodríguez, A., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Gómez, C. O., and Moya-Anegón, F. 2009. Synthetic hybrid indicators based on scientific collaboration to quantify and evaluate individual research results. *Journal of Informetrics*, Vol. 3, no. 2: 91-101.

Peterson, M. F. 2001. International collaboration in organizational behavior research. *Journal of Organizational Behavior*, Vol. 22 no. 1: 59-81.

Ronda-Pupo, G. A., and Katz, J. S. 2016. The scaling relationship between citation-based performance and international collaboration of Cuban articles in natural sciences. *Scientometrics*, Vol. 107, no. 3: 1423-1434.

Rousseau, R., and Ding, J. L. 2016. Does international collaboration yield a higher citation potential for US scientists publishing in highly visible interdisciplinary Journals? *Journal of the Association for Information Science and Technology*, Vol. 67, no. 4: 1009-1013.

Sarwar, R., and Hassan, S. U. 2015. A bibliometric assessment of scientific productivity and international collaboration of the Islamic World in science and technology (S&T) areas. *Scientometrics*, Vol. 105, no. 2: 1059-1077.

Schmoch, U., and Schubert, T. 2008. Are international co-publications an indicator for quality of scientific research? *Scientometrics*, Vol. 74, no. 3: 361-377.

Scott, J. 1988. Social network analysis. *Sociology*, Vol. 22, no. 1: 109-127.

Sud, P., and Thelwall, M. 2016. Not all international collaboration is beneficial: The Mendeley readership and citation impact of biochemical research collaboration. *Journal of the Association for Information Science and Technology*, Vol. 67 no. 8:

1849-1857.
Ulnicane, I. 2015. Why do international research collaborations last? Virtuous circle of feedback loops, continuity and renewal. *Science and Public Policy*, Vol. 42, no. 4 433-447.
Wagner, C. S. 2005. Six case studies of international collaboration in science. *Scientometrics*, Vol. 62, no. 1: 3-26.
Wagner, C. S., and Leydesdorff, L. 2005. Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, Vol. 34, no. 10: 1608-1618.
Wang, L., Thijs, B., and Glänzel, W. 2015. Characteristics of international collaboration in sport sciences publications and its influence on citation impact. *Scientometrics*, Vol. 105, no. 2: 843-862.
Wang, M. Y., Yu, G., and Yu, D. R. 2008. Measuring the preferential attachment mechanism in citation networks. *Physica a-Statistical Mechanics and Its Applications*, Vol. 387, no. 18: 4692-4698.
Wang, M. Y., Yu, G., and Yu, D. 2009. Effect of the age of papers on the preferential attachment in citation networks. *Physica a-Statistical Mechanics and Its Applications*, Vol. 388, no. 19: 4273-4276.
Wang, M. Y., Yu, G., and Yu, D. R. 2011. Mining typical features for highly cited papers. *Scientometrics*, Vol. 87, no. 3: 695-706.
Yang, D. H., and Yu, G. 2013. Static analysis and exponential random graph modelling for micro-blog network. *Journal of Information Science*, Vol. 40, no. 1: 3-14.
Yang, G.-Y., Hu, Z.-L., and Liu, J.-G. 2015. Knowledge diffusion in the collaboration hypernetwork. *Physica a-Statistical Mechanics and Its Applications*, Vol. 419: 429-436.
Zdravkovic, M., Chiwona-Karltun, L., and Zink, E. 2016. Experiences and perceptions of South-South and North-South scientific collaboration of mathematicians, physicists and chemists from five southern African universities. *Scientometrics*, Vol. 108, no. 2: 717-743.
Zhao, Y., Li, D., Han, M. J., Li, C. Y., and Li, D. M. 2016. Characteristics of research collaboration in biotechnology in China: evidence from publications indexed in the SCIE. *Scientometrics*, Vol. 107, no. 3: 1373-1387.
Zitt, M., Bassecoulard, E., and Okubo, Y. 2000. Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics*, Vol. 47, no. 3: 627-657.