2528-9705

Örgütsel Davranış Araştırmaları Dergisi
Journal Of Organizational Behavior Research
Cilt / Vol.: 3, Sayı / Is.: S2, Yıl/Year: 2018, Kod/ID: 81S2325

MEDICAL GENETICS GATEKEEPERS (OPINION LEADERS): A CO-AUTHORSHIP SURVEY OF MEDICAL GENETIC SCIENTISTS IN IRAN

Mitra ZAREI[1*], Mahmoud SALARI[2], Zahra JAFARZADEH[3]

[1] PHD Student in Information Sciences, International University of Imam Reza, Mashhad, Iran,
[2] Faculty Member of International University of Imam Reza, Information Sciences Departments, Mashhad, Iran,
[3] Faculty Member of Imam Reza University, Information Sciences Departments Mashhad, Iran.

***Corresponding Author**
Email: mitrazarei@yahoo.com

ABSTRACT

Introduction: In the course of scientific communication in the production of a scientific work there are always key positions occupied by key people. Identifying these positions and the impact of each, whether positive or negative, in the research process is the main objective of this research. Methods: In this research, using the network structure analysis techniques and using the co-authoring method, the authors have discovered the pattern of scientific collaboration between the authors of the field of Medical Genetics. The choice of this field is due to its interdisciplinary nature in medical research, with many scientific contacts going on. Since the main issue of this research is the study of scientific communication, co-authored techniques as one of the known methods in this field were used as a research method. Results: The results of this analysis showed that many relationships in the network are remained potential. There is a high willingness for scholars to cooperate in this field. The information on this network is moving at an acceptable rate between the researchers. The availability of information on this network has been reported weak. According to the Iegener Factor, 70 nodes have more significant links. In spite of the results, none of the researchers are in high position on the basis of structural hole indices. Conclusion: Based on the findings of this research, there were no significant differences between the structural hole indices as goalkeeper nodes in this study. Therefore, the use of central nodes as gatekeepers based on the findings of this research is more logical. The results of this research can be generalized to many emerging areas of medical science, as this area is highly interdisciplinary and common themes are common in its fields.

Keywords: *Information Gatekeepers, Structural Holes, Centrality Index, Medical Genetics*

INTRODUCTION

According to the definitions, gatekeepers work with the organization to filter and control the inputs of information and technology to the organization. Despite the good functioning of the gatekeepers in terms of the quality of information expected, gatekeepers sometimes control flow with specific goals that do not match the initial goal of their formation. In such a situation, checking the control points in the scientific communication cycle as gatekeepers and their effects on the production of science seems necessary. One of the functions of network studies, such as the coauthorship network, is the discovery of effective individuals who play the gatekeepers role in the production and distribution of information.

Problem Statement

Universities and research institutes are the most important institution for distributing credible information. Individuals in these institutions from different fields of science communicate

Geliş tarihi/Recieved: 10.12.2017 – Kabul tarihi/Accepted: 09.02.2018 – Yayın tarihi/Published: 21.08.2018

through interaction and exchange scientific information and collaborate in the production of science. Nowadays, in the fields of medical sciences, from one hand, there are new fields of study and interdisciplinary courses emerges, On the other hand, scientific production is the result of the collaboration of dozens in one or more research institutes. This requires a more detailed and in-depth study of scientific communication and its reinforcing and destructive factors. The ultimate aim of gatekeeping research is to always use gatekeepers and opinion leaders to prevent the infiltration of information into the flow of information and enhance the dissemination of qualitative information. Overall, using the results of this research, the gatekeepers and their techniques are identified and their positive and negative effects are measured. Finally, an effective model is presented to enhance the effect of effective factors on the quality of the information produced.

METHODOLOGY

A survey is an original research article that has been conducted to be practical research. We do this research with the aid of social network analysis techniques. The study was carried out in 4 phases:

Phase1: Identification of the group of documents under study

The first step in every co-authorship study is retrieval of scientific publications. Data should be collected from structured bibliographic databases witch provide correct information on the affiliations and authors, allow the exportation of data in suitable formats, provide the full name of the authors over a large number of academic journals, allowing the construction of organizational networks, and systematic cleaning and standardization of data (1).

Most of bibliographic research with social network approach is done based on WOS data because of many software and tools witch simply extract and standardize raw data for latter analysis. In between, Scopus data is located obsolete. Whereas searching both databased shows that, as well as great overlap in document coverage, Scopus is more comprehensive that WOS. According to this, we choose Scopus data in order to do analysis.

Scopus is chosen as the study database as its comprehensiveness and large overlap with WOS and Medline. We first retrieved documents conforming to the below search strategy:

Table 1:

Query	Number of Records
Mashhad University of Medical Sciences	
(AF-ID ("Mashhad University of Medical Sciences" 60001500) OR AF-ID ("Imam Reza Hospital" 60013544) OR AF-ID ("Ghaem Hospital" 60001853) OR AF-ID ("Eye Research Center of Mashhad University of Medical Sciences" 60077831) OR AF-ID ("Mashhad University of Medical Sciences Faculty of Medicine" 60001553)) AND (LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012) OR LIMIT-TO (PUBYEAR , 2011)) AND (LIMIT-TO (SUBJAREA , "BIOC"))	881
Tehran University of Medical Sciences	
(AF-ID("Tehran University of Medical Sciences" 60027708)) AND (LIMIT-TO(PUBYEAR,2016) OR LIMIT-TO(PUBYEAR,2015) OR LIMIT-TO(PUBYEAR,2014) OR LIMIT-TO(PUBYEAR,2013) OR LIMIT-TO(PUBYEAR,2012)) AND (LIMIT-TO(SUBJAREA,"BIOC"))	2885

Isfahan University of Medical Sciences	
(AF-ID ("Isfahan University of Medical Sciences" 60020609) OR AF-ID ("Alzahra University Hospital" 60021768)) AND (LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012)) AND (LIMIT-TO (SUBJAREA , "BIOC"))	640
Tehran genetic center	
(AF-ID ("National Institute for Genetic Engineering and Biotechnology Iran" 60019580)) AND (LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012)) AND (LIMIT-TO (SUBJAREA , "BIOC"))	393
Royan	
(AF-ID ("Royan Institute" 60017182)) AND (LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012)) AND (LIMIT-TO (SUBJAREA , "BIOC"))	129
Tarbiat Modarres University	
(AF-ID("Tarbiat Modares University" 60032053)) AND (LIMIT-TO(PUBYEAR,2016) OR LIMIT-TO(PUBYEAR,2015) OR LIMIT-TO(PUBYEAR,2014) OR LIMIT-TO(PUBYEAR,2013) OR LIMIT-TO(PUBYEAR,2012)) AND (LIMIT-TO(SUBJAREA,"BIOC"))	1606
Shahid Beheshti University of Medical Sciences	
(AF-ID("Shahid Beheshti University of Medical Sciences" 60018934) OR AF-ID("Labbafinejad Medical Center" 60007658) OR AF-ID("Loghman-Hakim Hospital" 60102239) OR AF-ID("Shahid Beheshti Medical University Modarres Hospital" 60089278) OR AF-ID("Infectious and Tropical Diseases Research Center Tehran" 60018219)) AND (LIMIT-TO(PUBYEAR,2016) OR LIMIT-TO(PUBYEAR,2015) OR LIMIT-TO(PUBYEAR,2014) OR LIMIT-TO(PUBYEAR,2013) OR LIMIT-TO(PUBYEAR,2012)) AND (LIMIT-TO(SUBJAREA,"BIOC"))	1572
TOTAL	8106

The queries prepared to find documents affiliated under 7 universities and institutions working specifically on medical genetics. The results, then, were refined by subject area "biochemistry, genetics and molecular biology" witch was nearest to medical genetics. The chronological limitation was set for recent 5 years as well. Total number of 8106 documents retrieved and saved in CSV format.

Phase 2: Data preparation and cleaning

For preparing data for coauthorship survey, a standardization of entries for authors and organizations need to be done. The objective is to consolidate names of a particular author or organization in order to ensure the correct acknowledgement of their scientific production. This step can be done manually or using specific software (1). Once the documents bibliographic information were saved we need to standardize the variations in single author name. With regard to this aspect Scopus prepare algorithms to collect different author names under a unique entry. To ensure, we manually review all names, checking institutional affiliation. Another data cleaning process was to deleting isolates and duplicated witch was done by SCI2 tool.

Phase 3: Determination of bibliometric indicators and social network analysis.

In this phase we begin calculation of metrics. The analysis in co-authorship networks includes quantitative metrics that may reflect the properties of the network as a whole or of its individual nodes. The overall structure of the network' common measurement are the number of nodes and links, centralization, density … The choice of which metric to use for calculating individual nodes depends upon the context and subjects of the analysis.

With specific regard to author-based indicators, we examined productivity of authors, according to different thresholds (authors with 1 document, 2–9 documents, and 10 documents)

With regard to the social network analysis, the following indicators were obtained to characterize the evolution of the overall size of the network and the patterns of scientific collaboration observed therein: the number of authors making up the network (number of vertices), the number of co-authorship links (number of edges), average number of collaborators per author (average degree), proportion between the number of real links in the network and the maximum number of links that are theoretically possible (network density), Clustering coefficient calculated according to the measure proposed by Watts and Strogatz, as the average of the local clustering coefficients of all the nodes, where the local clustering coefficient of each node is the proportion of real connections between it and its neighbors, compared with the number of all links that could possibly exist between them, assesses the extent to which each node plays the role of intermediary or bridge due to its location along the shortest path between other nodes of the network (Betweenneess), evaluating the distance between a given node and the rest of the nodes in the network; it is calculated as the sum of the shortest distances from the node being analyzed to all of the other nodes (closeness)

As co-authorship requires reciprocal cooperation among the participants, all connections are considered to be nondirectional (Fonseca, 2016).

Phase 4: visualization

We used the Gephi program for visualization using the Kamada-Kawai algorithm for the visual representation of the co-authorship network with the main active research clusters that exist in the area (in 2011-2016). At the end we can use metrics and graphs to interpret results.

RESULTS

Iran Medical Genetics Coauthorship Network

According to a search made in Scopos on December 2017, 5047 authors were identified and extracted. In order to find the central authors, the SCI2 software is used. It is a weighted network with 5047 authors as nodes and 29,736 edges as the number of times they have collaborated. After drawing the network, duplicate nodes and isolated nodes were deleted. The nodes' information along with the extracted value for each index was calculated for each node in Sci2 and extracted in the Excel file. The data was refined in several steps. The file was entered into the SPSS software to analyze the results. The number of articles produced by the authors is from 1 to 154 articles.

The information gathered from the database was for all researchers at the centers

After entering the information into the SCI2 software and extracting the indexes with the help of the Gephi Package, faculty information was seperated and analyzed from the entire data . Two types of analyzes were conducted to investigate the characteristics of the network. The

first analysis examined the overall network indexes under the topic of macro analysis. The macro social network analysis indicators examine the configuration and general characteristics of the networks.

Macro analysis

Density

Density is the collection of possible nodes within the network. In other words, the degree of connectionness is measured using density. The higher the density, the greater the cohesion of the network. Since the present network density (0.18) is closer to zero, the network does not have good consistency. In other words, only 0.18% of the total potential network connection has been activated.

Micro indicators

In addition to analyzing the entire network based on macro indicators, the performance of each node was also calculated using micro indicators. Among micro indicators are central indicators. The high value of this measure means that the node has a high potential as a watch or interface in social network.

Degree centrality

This indicator measures the potentiality of a node's activity, meaning that as much as it is easier for a person to communicate directly with others, that person has a higher degree of centrality. In the coathorship network, the authors and the number of links (degrees) are the number of times a person has been cited with other people. The most effective measure of the centrality is degree. The degree is calculated for individuals (micro index) as well as for the whole network (macro index). The average grade level for the compilation network was calculated 25. The degree of each person was also calculated in the study of the micro-centrality index in the network. The degree centrality of each person reflects the degree of co-authorship with other network members. The value of this indicator varies between 0 and 1272 for network nodes. A person with a high degree rank can create experiences, skills and organizational memory for others. It is essential that they identify those individuals who can act as a bottleneck for the flow of information and also be able to potentially overload with information requests.

Betweenness centrality

Betweenness centrality is the node or interface point that communication paths pass through elsewhere, and through which a large number of nodes directly connect to other nodes. Betweenness centrality is the number of times that an agent interacts with other agents in the network and links them. An agent with a high degree of betweenness has the potential to link the other factors with greater power. The low betweenness centrality indicates that most links and connections in the network are made directly and without the use of an intermediary. This is not desirable because people are expected to interconnect in the intermediary network to increase network power. High-betweeness nodes play an important role in network connectivity. In the current network, 3455 nodes do not play any role in the network and their betweenness index is zero. The calculated betweenness among the remaining nodes varies from 0/1428577 to 3645672.

Closeness centrality

The centrality of proximity refers to the distance between a node to other nodes, or the shortest distance between each agent and other factors on the network. The lesser the distance between the nodes, the lesser the role of the intermediary and the more the node becomes aware of its

data. This measure indicates the availability, health and safety of the agents. The nodes with high closeness index have more effective power in the network and play a more central role in the network and have more accessibility to other nodes. This index also shows some degree of independence of a node. When a node has the ability to communicate with more nodes and uses the least number of intermediaries in these connections, it will have more autonomy and, therefore, greater independence. The value of this numeric index is between zero and one. This network has 9 nodes with closeness degree 1. The average centrality of the nearby network is 0/266661522. The lowest value of this index is calculated to be 0/17684161944566.

Eigenfactor centrality

The eigenfactor is the importance of the node in the network. This index allocates relative scores to all nodes. The basis of this indicator is that communication with nodes with higher scores will result in higher nodal value . Special vector focus emphasizes that the center of a node is calculated with respect to the center of other nodes associated with it. It is stated in this measure that all edges do not have the same value, but the edges that are linked to the effective vertices of the network are more valuable than other edges. In other words, a high-level vertices is linked to the more influential vertices. This index is calculated between zero and one. In the network, there are 70 central centers close to one (more than 0.9) which are the central nodes.

Table 5: Frequency and percentage of nodes centrality

Closeness centrality			Eigenfactor centrality		
	Frequency	percentage		Frequency	percentage
0-0/2	21	0/4	00050-0/	4231	83/8
0/21-0/3	4741	93/9	001-0//00060	89	1/8
0/31-0/35	262	5/2	0/0900-0/001	621	12
0/36-1	22	0/4	-109010/	101	2
Degree centrality			Betweenness centrality		
	Frequency	percentage		Frequency	percentage
200-	4428	87	0-100000	4930	97
20-50	403	8	100001-300000	81	1/6
50-100	88	1/7	300001-500000	17	0/3
100+	128	2/5	500001+	19	0/4

Structural holes

Cutpoints

In Morel's words, the cut-off point in a network is called actor, whose elimination increases the number of network components by dividing the sub-graphs into several sections. The cutting points are the connecting points between the elements forming a component Cutting points are important in communication networks, because if the node is removed from the network, the network is divided into two parts that there is no connection between them. The cutoff concept can be assigned to a node or set of nodes. A set of nodes that communicate in a graph is called the set of cutting points. Cutoffs are introduced as key players on the network because they are responsible for connecting multiple nodes on the network. Hence, identifying cutting points in the network, especially in terms of capacity building and institutional strengthening, is an

important analytical tool for program management. Such a node plays the role of a communications broker among others. (5) According to Ortiz, cutting points are the same nodes that have a high betweenness centrality. It also considers the concept of structural holes according to cutting points in network. (6)

HITS

The HITS algorithm was used to extract structural holes in the network. This algorithm measures two hub scores and an authority score for each node. The entities that many tendencies tend to be oriented towards or affect the performance of other derivatives are called hubs. The entities that move from their direction to the hubs are authorities. So relationships are from the authorities to hubs. The hubs are oriented towards the authorities and the authorities are oriented towards the above, but the authorities are amplifiers compared to the reference ones. Based on the information obtained from a total of 5047 nodes, 5036 nodes have a hubs and authority score. Since the current network is non-directional, both scores are consistent. The highest score in this network is 0/00123 and the lowest score is 0/000000000000000164.

Pagerank

In addition, the indicator can be referenced as Pagerank, which is used to calculate structural holes. This number varies between 0/00003078344113 and 9 in the current network. A total of 6 records have a Top Rank Index of 1 and the rest of the nodes are less than this. For key people and cutting points, it is suggested that these people be considered in strategic planning for research in this area and will be provided to enhance the scientific cooperation network of research facilities. This index seems to be the most important indicator of key persons in network.

Table 3: structural holes

Authority Score			PageRank		
	Frequencies	Percentage		Frequencies	Percentage
0.001-0.00999	71	1/4	0.001-0.999	36	0/7
0.0001-0.000999	29	0/6	0.0001-0.0005	2509	49/7
0.00001-0.000001	503	10	0.00051-0.000999	94	1/9
0-0.0000011	4444	88	0.00001-0.0000999	2400	47/6

Opinion leaders

For extracting key authors two structural hole indicators and 4 centrality indicators calculated. After data entry into the Spss software and the Kolmogorov-Smirnov test, normality of the data was rejected. Therefore, in order to study the correlation of these indicators, information was ranked and Kruskal Wallis test were performed between these variables. The results of this test have been meaningful about the difference in the value of the structural holes with the centrality scores. It means that there is a significant difference between the five variables of importance score, Page Rank and centrality scores. In order to ensure the validity of this result, the result of Mann-Whitney test was also performed between variables two by two and there was no significant correlation between the indices. In order to select opinion leaders inspired by the necessary precondition of the eigenfactor, all indicators were considered and those whose grades in each index were much higher than the others are selected. betweenness index

was the most important indicator in identifying key people. In this way, people with a visitor center of over 100,000 were selected, including 118. In the following, the authority index as the other key indicator in identifying key people was the basis of choice. Since the score for the authority of the people was close to zero, only a few scored more than 0.001 in this index, so Individuals who scored more than 0.001 percent of significance included a total of 72 nodes chosen for the lead list of opinion leaders. Similarly, people with a centrality score of more than 100 including 129 nodes, those with a centrality score of 1 includng 10, people with an eigen factor of more than 0.9, including 72, and those with a page rank score More than 0.001 including 37 people were selected. In this way, people who have scored a high score at least in one index (a range that is significantly different from the average and low scores) were chosen. Obviously, the names of some extracted people were repeated based on an indicator in the names of the top names of other indicators. By removing repetitions, the 125-person list was extracted. Demographic data show that the majority of people (about 43%) had a full professor's degree. The selected subjects are taught in 23 fields. Most people (about 18%) were specialized in genetics. The fields of pharmacy, immunology, physiology, and biostatistics are at the forefront of opinion leaders. Altogether about 60% of the selected people have published more than 10 articles in the last five years. Very few have produced only one scientific article. Investigating the organizational affiliation of opinion leaders showed that Tehran University of Medical Sciences ranked first with 46 people. Mashhad University of Medical Sciences was ranked second with 26 leading opinion leaders. Beheshti University of Medical Sciences, Tarbiat Modarres, Isfahan and National Genetics Research Institute were ranked respectively. Alborz University of Medical Sciences, Bushehr, Iran, Kashan, Urmia Each one has one and of the medical universities in Kermanshah, Hamedan, Jihad Research Institute, Pasteur Institute of Genetics each had two faculty members in the list of leaders. A researcher was also identified from Harvard University.

Chart 1 sets the scores earned by 126 opinion leaders in each of the 6 indicators surveyed

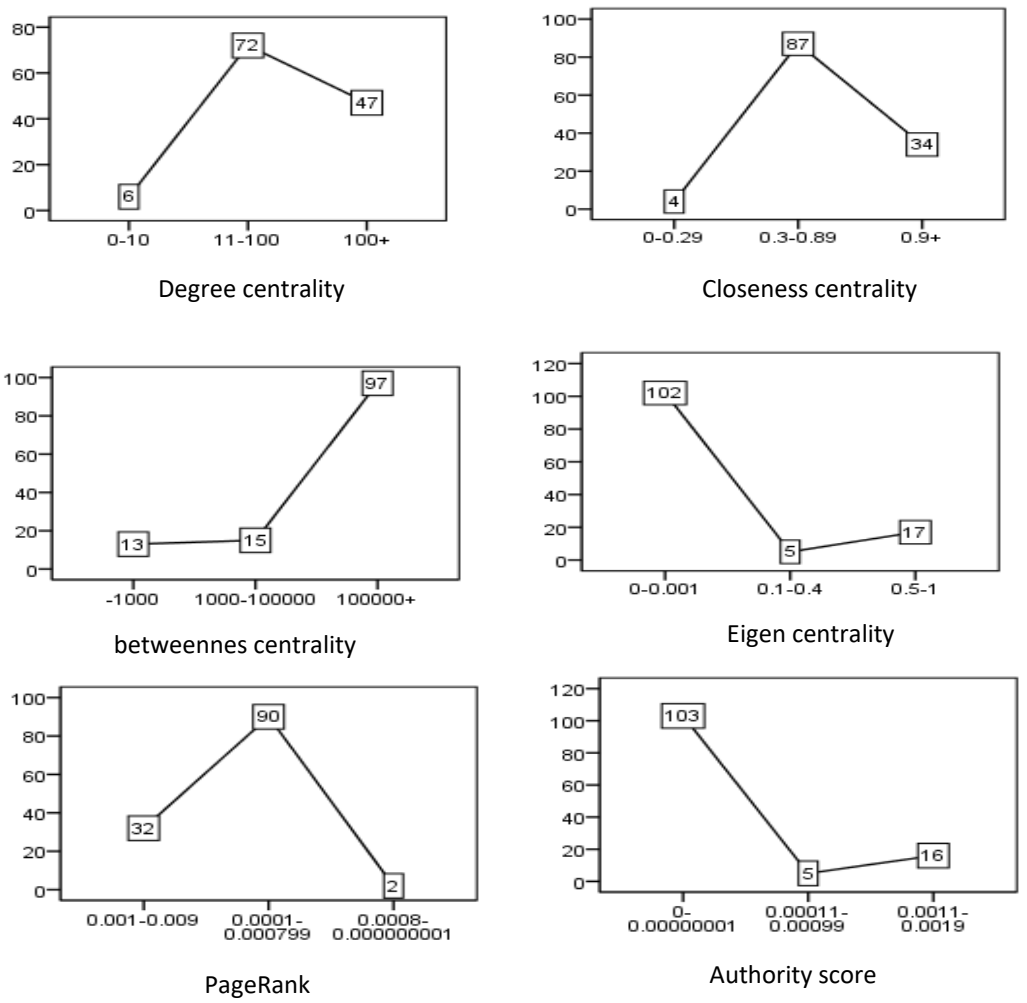

Figure 1: Degree centrality; Closeness centrality, betweennes centrality; Eigen centrality; PageRank and Authority score charts

An investigation of these figures indicates that the authority score, the eigenfactor and the Page Rank of these individuals are low compared to all known nodes. So that a significant number of well-known nodes have near-zero authority. In contrast, the betweenness of opinion leaders in the entire network is of desirable value. So that most of the people selected had over 100,000 betweenness. In the degree and closeness index, average scores are earned. In this way, individuals selected on the basis of each indicator might have scored very low in other indicators. It may also be a random indicator for an individual. For example, the result of his collaboration with a person influenced a period of his scientific activity. In some studies, this dichotomy points out the results of ranking nodes based on each centrality and hub index Therefore, selecting key people does not seem to be based on an indicator of reliable results. Therefore, out of 126 people were selected who scored at least two higher scores. In other

words, those who were repeated in more than one index were named. These people make up the list of 55 opinion leaders.

CONCLUSION

An analysis of the network of medical geneticists reveals the scientific connection in this field. The coherence of the researchers' network in this field is indicative of the low figure. This means that many of the possible relationships within the network are not activated. On the other hand, the acceptive clustering coefficient obtained shows high tendency to cooperate among researchers. The gap between the authors on this network is less than 4, and this helps to quickly transfer information across the network. Examining the degree centrality as the most effective measure of centrality at both micro and macro levels showed an average of 25 relationships to each node. Of course, given the relatively high deviation from the mean of this index (107), it can be estimated that the dispersion of this value among the nodes is relatively high. As a result, there is a large dispersion of the links generated by the co-authorship. This also applies to the betweenness index to some extent. But since more than half of the nodes have zero betweenness, it can be inferred that the dispersion between nodes whose zero betweenness y index is not high. In other words, the data in this index are less dispersed than the degree index. The relatively low variance (9.28) also confirms that the calculated figures for this index are more consistent. In contrast, the calculated values for the degree index (with a variance of 11491) have a low uniformity. The dispersion indexes of closeness centrality are lower than the other two centralities due to the different nature of the data (being between zero and one). A look at the data of this index indicates that the data obtained are largely dispersed around the average (0.27) and are well-matched. Since the index can be assumed to be desirable (1), A relatively high average distance with the optimal number in the network indicates that Access to information on the network is slow due to the multiplicity of factors existing between them. Based on the calculated eigenactor in the current network, 70 people clearly have more significant links. Another point about the eigenfactor is the condition that there should be a big difference between nodes with high centrality and other nodes. This condition is provided in the current network. In the sense that there is a significant difference between the nodes that have the highest eigenfactor (0.09 or close to one) with later nodes (0.5 or less than half). Therefore, the credibility of this index can be assured in current data.
Structural holes in the network are in some ways central nodes. Of course, there are no special measures to extract structural holes in the network. In other words, in fact, none of the network analysis measures is sufficient to extract the structural holes. However, some measures and algorithms have so far been used in research for this purpose. Several methods are currently used to evaluate the importance of a node in studies. Degree and betweenness centralities (2)(3) closeness, … In the meantime, betweenness is referred to as a measure of the structural holes due to the ability to control the resources. (2) Some network analysis pioneers also agree on this issue (4). Among other network analysis indicators, two authority and Page Rank are also helpful in identifying structural holes. In general, the researchers did not score any significant score on these two indicators.

ACKNOWLEDGEMENT

This is thanks to the cooperation of the Medical Sciences Universities of Mashhad, Shahid Beheshti, Tehran, Isfahan, Royan Institute, National Center for Genetics of Iran, and Tarbiat Modarres University for cooperation in providing information. All material rights of this work belongs to Imam Reza International University

Author's contribution's statement

Mitra Zarei and Mahmoud salari Participated in study design, data collection and evaluation, drafting and statistical analysis, interpretation of the data and the conclusion. All authors performed editing and approving the final version of this paper for submission, also participated in the finalization of the manuscript and approved the final draft.

Acknowledgement

This article is affiliated by International University of ImamReza.

References

Fonseca, B. d. P. F. e., R. B. Sampaio, et al. (2016). "Co-authorship network analysis in health research: method and potential use." health research policy and systems 14(34).

Habiba, Yintao Yu, Tanya Y. Berger-Wolf, and Jared Saia. "Finding Spread Blockers in Dynamic Networks." In Advances in Social Network Mining and Analysis, edited by Lee Giles, Marc Smith, John Yen and Haizheng Zhang. Las Vagas: Springer, 2008.

He, Lifang, Chun-Ta Lu, Jiaqi Ma, Jianping Cao, Linlin Shen, and Philip S Yu. "Joint Community and Structural Hole Spanner Detection Via Harmonic Modularity." ACM 2016.

Kosorukoff, Alex. Social Network Analysis Theory and Application. Passmore, D. L, 2011.

Morel, Carlos Medicis, Suzanne Jacob Serruya, Gerson Oliveira Penna, and Reinaldo Guimara˜es. "Co-Authorship Network Analysis: A Powerful Tool for Strategic Planning of Research, Development and Capacity Building Programs on Neglected Diseases." PLOS 3, no. 8 (2009): 1-7.

Ortiz-Arroyo D. Discovering sets of key players in social networks. Computational social network Analysis, computer communications. 2010. pp. 27–47. doi:10.1007/978-1-84882-229-0.

Scott, J. (2012). What is Social Network Analysis? Bloomsbury Academic.

Wasserman, Stanely, and Katherine Faust. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press, 1994.

Zhongati, C., M. Ning, et al. (2015). "Key nodes identify in the Peasants' social network based on structural hole theory." International Journal of Hybrid Information Technology 8(4): 251-258.