

ORGANIZATIONAL AND SOCIAL NETWORK MEASURES AND RESEARCH CITATIONS

Marija Lukic¹, Mladen Cudanov², Ivan Matic³, Dejana Pavlovic⁴, Gheorghe Savoiu⁵

^{1,2,3}University of Belgrade, Faculty of Organizational Sciences, Jove Ilica 154, 11000 Belgrade, Serbia

⁴Economics Institute, Belgrade, Kralja Milana 16, 11 000 Belgrade, Serbia

⁵University of Pitești, Str. Targul din Vale, nr.1, Argeș, Romania

e-mail: ¹lukicmarija@gmail.com, ²mladenc@fon.bg.ac.rs, ³ivanmatic89@yahoo.com, ⁴dejana.pavlovic@ecinst.org.rs
⁵gsavoiu@yahoo.com

Abstract. *A network describes a correlation between individuals and how can they achieve common aims together. Correlation between them depends of the strength of links and nodes. In the paper the results can be seen of research based on the citation analysis of authors, connections between them and networking with other authors. We used a database of the National Library of Serbia, covering papers published in indexed journals from 2010 to 2014. Results have shown a positive correlation between the number of author's first neighbours and the number of citations.*

Key words: *Network theory, Node degree, Authors, Researchers, Publishing, Scientific research, Citation, Collaboration, Paper, Article*

1. INTRODUCTION

The organization consists of units that are in constant interaction with one another, creating necessary links to achieve the set goals as well as similar interests for mutual assistance and support. Therefore, the organization is a network of units and connections created in a mutual interaction under the influence of external and internal influences with a goal to enable an organization to respond more efficiently to new demands of the environment. The result of these activities is a creation of networks that can be complex and simple depending on the participants. The structure of complex network determines the type of dynamics and stability of the network (Kolasa, 2005, Strogatz, 2001, Namba and Tanabe, 2010). Network organization consists of nodes. The nodes can be individuals, teams, organizational units or organizations and links that represent the mechanisms of control, coordination and control of the participants' activities and their roles and positions in the organization within a network. The correlation between them will directly depend on the strength of the connections between the nodes. However, not all the nodes and links are equally important for network functioning. The importance of the nodes is determined based on the connection parameters by which various aspects related to the position of the nodes are calculated (Scott, 1991, De Nooy, Mrvar and Batagelj, 2005). Strong links tend to connect individuals with the same or similar interests and to group them in clusters. Examples of such networks with strong connections are the relations between friends, relatives, neighbors (Erickson and Yancey, 1980, Lin, Ensel and Vaughn, 1981). According to the research by Granovetter M.S. (1973) strong connections play an important role in organizations so they need to be closely looked into.

2. LITERATURE REVIEW

The appearance of the theory is related to the roots of the theory of social psychology and inter-organizational theory.

The network theory was originally applied to describe social relations between individuals. Recently, it has also been applied in modern theories of organization and strategic management, including some prominent studies that explore the structure of the organizations in the survey conducted by Almaas, Kulkarnir and Stroud (2002) and De Nooy, Mrvar and Batagelj, (2005).

Recent research shows that the structure of networks consists of parameters that are commonly used in research to understand the complex network structure:

- Degree distribution
- Average path length and diameter
- Clustering coefficient
- Centrality

Degree distribution

Degree distribution is one of the most important components. Given the direction within the network, three important parameters are the input level (number of connections to each node), the output level (number of outgoing links) and the total (the total number of links). According to Ghosh, Banerjee, Sharma, Agarwal and Ganguly (2011) structure of a network $p(k)$ is defined to be the path of nodes in the network with degree k . However, the degree distribution is often noisy and there are rarely enough nodes having high degrees to get good statistics in the tail of the distribution.

Average path length and diameter

Path length is defined as the average number of steps the shortest path takes to all possible nodes in the network. The shortest time is important in establishing and transmitting information in the organization, because it defines a small number of nodes and faster data transmission. Networks with shortest path length are more efficient than all the other networks of the organization.

Clustering coefficient

Clustering coefficient is a local measure used to describe a node group in a network representing one cluster (Almaas, Kulkarnir and Stroud, 2002). Grouping clusters in the network increase the probability that two randomly chosen neighbors are interconnected.

Centrality

Defining the relationship of the central nodes and links is a parameter used in complex organizations (Gonzalez Martin, Dalsgaard and Olesen, 2010). Node centrality concepts and measures help determine the importance of node in a network. Connection with the nodes is defined as a ratio of the shortest paths between the pairs. This parameter is used to specify a node that can have a crucial role as a mediator of the interaction between other nodes (Wambeke, Liu and Hsiang 2012). Because of the importance of this parameter,

various centrality measures have been proposed over the years (Scott, 1991) such as degree centrality, closeness centrality and betweenness centrality. The degree centrality of node k (i.e., p_k) is defined as follows:

$$C_D(p_k) = \sum_{i=1}^n a(p_i, p_k) \quad (1)$$

Where n is the number of nodes in the network and $a(p_i, p_k) = 1$ if and only if node i and k (i.e., p_i and p_k) are connected; $a(p_i, p_k) = 0$ otherwise.

Furthermore, nodes with high degree centrality could be identified as informal leaders of the group (Krackhardt, 2010).

Closeness centrality is defined by distance between nodes. According to Freeman L.C. (1979) closeness centrality of node k (i.e., p_k) is defined as follows:

$$C_C(p_k) = \sum_{i=1}^n d(p_i, p_k)^{-1} \quad (2)$$

Where $d(p_i, p_k)$ is the shortest path linking p_i and p_k .

Betweenness centrality (BC) describes the importance of a node as a connector between different parts of the network (Freeman, 1979). Nodes with $BC > 0$ connect areas of the network that would otherwise be sparse or not connected at all (Newman, 2004).

More precisely, the betweenness of node k (i.e., p_k) is formulated as follows:

$$C_B(p_k) = \sum_{i < j}^n \frac{g_{ij}(p_k)}{g_{ij}}; i \neq j \neq k \quad (3)$$

Where g_{ij} is the geodesic distance (shortest paths) linking p_i and p_j and $g_{ij}(p_k)$ is the geodesic distance linking p_i and p_j that contains p_k . Nodes with high betweenness centrality play the role of a broker or gatekeeper to connect the nodes and sub-group.

The network theory has grown to be a very popular field among the researchers all over the world. Researchers are committed to searching for appropriate network topology to capture the nature of complex networks (Ravasz and Brabasi, 2003, Buldryrev, Parshani, Stanley and Havlin, 2010).

3. PREVIOUS RESEARCH IN THE FIELD OF NETWORK THEORY

In Zhou, Irizarry and Qiming, (2014) research, the network theory is used to identify accident happening. They used network theory to investigate the complexity of subway construction accident network (SCAN). The five parameters including degree distribution, average path length and diameter, clustering coefficient and betweenness centrality were used to better understand and capture the structure of SCAN. Authors have concluded that an accident chain or network actually exists.

Another research in the field of network is based on collaboration networks by Abbasi, Hossain and Leydesdorff (2012). Their research analyzes whether preferential attachment in scientific co-authorship networks is different for authors with different forms of centrality. Using a complete database for the scientific specialty of research about "steel structures," they showed that betweenness centrality of an existing node is a significantly better

predictor of preferential attachment by new entrants than degree or closeness centrality. During the growth of a network, preferential attachment shifts from (local) degree centrality to betweenness centrality as a global measure. An interpretation of the supervisors of PhD projects and postdocs broker between new entrants and the already existing network, and thus become local to preferential attachment. Because of this mediation, scholarly networks can be expected to develop differently from networks which are predicated on preferential attachment to nodes with high degree centrality (Abbasi, Hossain and Leydesdorff, 2012).

4. SOCIAL NETWORK

Everything is connected in the world. According to this, the network theory started being used in safety management, among researchers, social networks etc. In academic world, scientific collaboration networks are a complex kind of social networks since the links between a number of authors (nodes) and co-authors are growing over time (Strogatz, 2001). Collaboration is defined as a group of people working together for a common goal sharing knowledge (Lin, Ensel and Vaughn, 1981). In the recent decades, modern science researchers have used methods to examine collaboration networks. Social network analysis has given many results regarding social influence, social groupings, inequality, disease propagation, communication of information and almost every topic that interested 20th century sociology (Newman, 2001).

Social networking has become one of the most important communication tools among people nowadays (Griffith and Liyanage, 2008). However, social networks exist on the Internet websites where millions of people share interests on certain disciplines, and make available to members of these networks various shared files and photos and videos, create blogs and send messages, and conduct real-time conversations.

These networks are described as social, because they allow communication with friends and colleagues study and strengthen the ties between members of these networks in the space of the Internet (Griffith and Liyanage, 2008, McNamara, Weininger and Lareau, 2013, Zaidieh, 2012). The most famous in the world of social networks are Facebook (Facebook.com) and Twitter (Twitter.com) and MySpace (myspace.com) and others Social networking sites (SNS) have infiltrated people's daily life with amazing rapidity to become an important social platform for computer-mediated communication provides a new method of communicating, employing computers as a collaborative tool to accelerate group formation and escalate group scope and influence (Griffith and Liyanage, 2008, McNamara, Weininger and Lareau, 2013, Zaidieh, 2012).

Social networking is built on the idea of how people interact with each other. It gives people the power to share, making the world more open and connected. Nowadays, social networking has a vital influence on our lives as it helps a lot in every field of life such as political field, economic field and educational field (McNamara, Weininger and Lareau, 2013, Zaidieh, 2012).

5. RESEARCH HYPOTHESIS AND METHODS

Research in this paper is based on the citation analysis of authors, connections between them and networking with

other authors. The database we used is the official report of the National Library of Serbia and it contained data for 2132 authors: authors' name, their published papers during the period 2010-2014 and number of citations per each paper. Due to poor structured data base, we had to adapt it to a form that corresponds to the software Cytoscape that we have used to create a network of authors. The final database contained an 1883 authors instead of 2132 of them. Cytoscape is an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. Although Cytoscape was originally designed for biological research, now it is a general platform for complex network analysis and visualization (Cytoscape consortium, 2014). Software Cytoscape was a logical choice for realization of this research, given the fact that this software creates network by interconnecting nodes. The network of authors was created by interconnecting authors who had joint published papers.

Taking into consideration the data, we had in our disposal and on basis of literature review, we have formulated following two hypotheses:

H1: There is a positive correlation between the number of author's first neighbours and the number of his/her citation.

H2: If the number of co-authors of an author increases by one unit, his/her citations will increase at least twice.

6. RESULTS AND DISCUSSION

The software created an undirected network of authors. An undirected network is type of network where direction between nodes is not defined. In order to increase the accuracy, we eliminated all duplicate edges and self-loops from the network, and then we got a network consisting of 1883 nodes and 2603 edges.

Figure 1 is a graphical representation of connectivity between authors. This network consists of 137 components. The biggest network is the major component that contains the largest number of related authors.

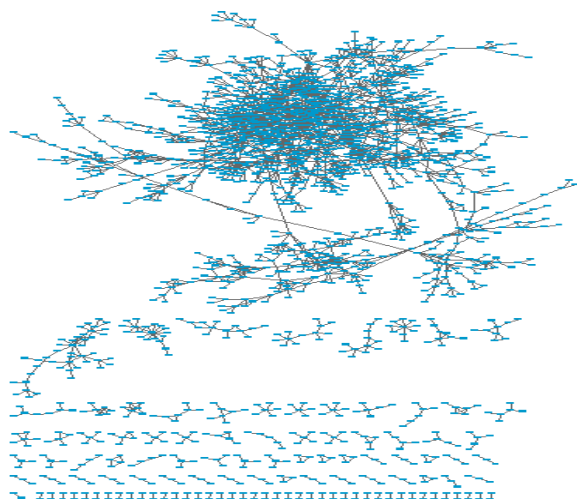


Fig. no. 1 Network of authors

Figure 2 presents the basic parameters of the network: the average number of neighbors in the network is 2.765 and also it can be seen that there is no self-loops and multi-edge node pairs.

Network Statistics of Podaci autori.csv (undirected)			
Betweenness Centrality		Closeness Centrality	
Shortest Path Length Distribution		Shared Neighbors Distribution	
Stress Centrality Distribution		Neighborhood Connectivity Distribution	
Simple Parameters	Node Degree Distribution	Avg. Clustering Coefficient Distribution	Topological Coefficients
Clustering coefficient :	0.190	Number of nodes :	1883
Connected components :	137	Network density :	0.001
Network diameter :	32	Network heterogeneity :	1.038
Network radius :	1	Isolated nodes :	0
Network centralization :	0.023	Number of self-loops :	0
Shortest paths :	1841440 (51%)	Multi-edge node pairs :	0
Characteristic path length :	9.776	Analysis time (sec) :	27.752
Avg. number of neighbors :	2.765		

Fig. no. 2 Network parameters

Taking into consideration the sample size, we assumed that there is a normal distribution of results in both populations. According to this assumption, we were able to calculate the Pearson Correlation Coefficient and the results of this analysis are presented in Table 1.

		Number of citations	Number of co-authors
Number of citations	Pearson Correlation	1	.324
	Sig.(2-tailed)		.000
	N	1883	1883
Number of co-authors	Pearson Correlation	.324	1
	Sig.(2-tailed)	.000	
	N	1883	1883

Table no. 1: Correlation between number of co-authors and number of citations

The Correlation Coefficient is positive and its value is 0.324. When correlation coefficient is positive, that means that both variables are changing in the same direction. In our study, this means that if the number of co-authors of an author increases, then its citation increases too, or if the number of co-authors decreases, then decreases its citation too. The significance level is 0.000, which means that our results are statistically significant ($p < 0.05$). Because of the fact that this correlation is statistically significant, it can be concluded that hypothesis H1 has been proved.

After we have proved a positive correlation between the number of co-authors and citations, we were able to take analysis of statistical linear regression. The purpose of this analysis is to calculate the level of increase in author's citations, on average, when the number of co-authors increases by one unit. The results are presented in Table 2.

The dependent variable is the number of citations and the independent variable is number of co-authors.

The value of non-standardized coefficient B is 2.52. This means that an increase of one point of independent variable - number of co-authors, corresponds to an increase of 2.52 points of number of citations, on average. This result is statistically significant, because the significance level is lower than 0.05 and we can conclude that hypothesis H2 has been proved too.

Model	Unstd. Coeff.		Std. coeff.	t	Sig.
	B	Std. Error	Beta		
(Constant)	1.192	.677		-1.760	.079
Number of co-authors	2.527	.170	.324	14.874	.000

Table no. 2: Linear regression

5. CONCLUSION

Network theory is becoming increasingly popular research field if we know the fact “on the world everything is connected and happens with the reason”. Organization or group is defined as a group of people working together and aiming to achieve a common goal. The aim of the network theory is to find connections between nodes in order to be done as efficiently as possible. Network theory is applied to describe the social cohesion between individuals. Social networks have become an important for communication between people. People have the opportunity to share information, opinions and attitudes and to be as much as possible connected together. If there is a strong connection between subjects the message will be sent as soon as possible and realise the target.

The importance of the nodes depends on parameters which are commonly used in research to understand complex network structure (degree distribution, average path length and diameter, clustering coefficient and centrality).

The study provided an overview of complex scientific co-authorship network. In our research we concluded that there is a positive correlation between the number of authors and citations. Our hypothesis show expected results.

This research will be used for the future research of scientific networks and analysis different parameters of cooperation between authors from research institution all over the world.

REFERENCES

[1]. Abbasi, A., Hossain, L. and Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks, *Journal of informetrics*. 6, 403-412

[2]. Almaas, E., Kulkarnir, V. and Stroud, D. (2002). Characterizing the structure of small world networks, *Phys. Rev. Lett.* 88, 9,982-101.

[3]. Buldryrev, S.V, Parshani, R., Stanley, G. and Havlin, H.E. (2010). Catastrophic cascade of failures in independent networks. *Nature* 464, 1025-1028.

[4]. Cytoscape consortium (2014) What is Cytoscape? [online] Available at: <http://www.cytoscape.org/what_is_cytoscape.html> [Accessed 9th December 2014].

[5]. Erickson, E. and Yancey, W. (1980). Class, Sector and Income Determination, Unpublished paper, Department of Sociology, Temple University.

[6]. Freeman, L.C. (1979). Centrality in social networks, conceptual clarification, *Soc. Networks* 1, 215-239.

[7]. Ghosh, S., Banerjee, A., Sharma, N., Agarwal, S. and Ganguly, N. (2011). Statistical analysis of the Indian railway network: a complex network approach, *Acta Phys. Polon. B Proc. Suppl.* 4, 123-137.

[8]. Gonzalez Martin, A.M., Dalsgaard, B. and Olesen, J.M. (2010). Centrality measure and the importance of generalist species in pollination networks, *Ecological Complexity*. 7, 36-43.

[9]. Granovetter, M.S. (1973). The strength of weak ties. *American Journal of sociology*. 78,1360-1380.

[10]. Griffith, S. and Liyanage, L. (2008). An introduction to the potential of social networking sites in education, *Emerging T echnologies* Los Angles, 75-81.

[11]. Kolasa, J. (2005). Complexity, system integration, and susceptibility to change: biodiversity connection. *Ecol. Complex*. 2, 431-442.

[12]. Krackhardt, D. (2010). Social networks. In J. M. Levine, & M. A. Hogg (Eds.), *Encyclopedia of group processes and intergroup relations* Los Angles, 817-82

[13]. Lin, N., Ensel, W.M. and Vaughn, J.C. (1981). Social Resources and Strength of ties: Structural Factors in Occupational Status Attainment, *American Sociological Review*. 46, 393-405.

[14]. McNamara, H., Weininger, E. and Lareau, A. (2013). Social Ties to Social Capital: Class Differences in the Relations between Schools and Parent Networks. *American Educational Research Association*.

[15]. Namba, T., Tanabe, K. and Maeda, N. (2010). Omnivory and stability of food webs. *Ecol. Complex*. 3, 73-85.

[16]. Newman, M.E.J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E*. vol. 64

[17]. Newman, M.E.J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl. 1), 5200.

[18]. Nooy, W.D., Mrvar, A. and Batagelj, V. (2005). *Exploratory social network analysis with Pajek* (Structural analysis in the social science). Cambridge University Press, Cambridge.

[19]. Ravasz, E. and Brabasi, A.L. (2003). Hierarchical organization in complex networks. *Phys. Rev. E*. 67

[20]. Scott, J. (1991). *Social network analysis: A handbook*. Sag

[21]. Strogatz, S.H. (2001). Exploring complex networks, *Nature*. 410, 268-276.

[22]. Strogatz, S.H. (2001). Exploring complex networks, *Nature*. 410, 268-276.

[23]. Wambeke, B., Liu, M. and Hsiang, S. (2012). Using Pajek and centrality analysis to identify a social network of construction trades. *J. Constr. Eng. Manage.* 138, 1192-1201

[24]. Zaidieh, J. (2012). The Use of Social Networking in Education: Challenges and Opportunities. *World of Computer Science and Information Technology Journal (WCSIT)*, 18-21.

[25]. Zhou, Z., Irizarry, J. and Qiming, L. (2014). Using network theory to explore the complexity of subway construction accident network (SCAN) for promoting safety management, *Safety Science*. vol. 64, 127-136.