

# Analysis of the NETSCIENCE Social Network

X.B. Li

College of Software Engineering  
Lanzhou Institute of Technology  
China

Jozef Stefan International Postgraduate School  
Slovenia

N. Lavrac

Department of Knowledge Technologies  
Jozef Stefan Institute  
Slovenia

V. Podpečan

Department of Knowledge Technologies  
Jozef Stefan Institute  
Slovenia

**Abstract**—This paper presents a basic analysis of the Net Science database together with a description of the techniques for social network analysis. The database was provided by M. Newman in May 2006 and contains a network of scientists who together co-authored the network theory and experiment. The analysis was performed using the Pajek software. Numerous pictures and tables have been created to illustrate the structure and show the most important parts of this network. Some of the notable findings include important groups of actors whose collaboration is important for the discoveries and development of the theory of networks, and those actors which are important with respect to the whole graph.

**Keywords**—social network; pajek software; core; clique; island; centrality

## I INTRODUCTION

Social network analysis (SNA) is the mapping and measuring of relationships and flows between people, groups, organizations, computers, web sites, and other information/knowledge processing entities. The nodes in the network are the people and groups while the links show relationships or flows between the nodes. SNA provides both a visual and a mathematical analysis of human relationships [2-6].

The Pajek software was used with the network NetScience, which was obtained from the Pajek datasets website <http://vlado.fmf.uni-lj.si/pub/networks/pajek> [2-9].

## II GENERAL INFORMATION ABOUT THE DATA

The file, NetScience, was written by a group of scientists working on network theory and experiment, as compiled by M. Newman in May 2006. The network was compiled from the bibliographies of two review articles on networks, M.E.J. Newman, SIAM Review 45, 167-256 (2003) and S. Boccaletti, et al., Physics Reports 424, 175-308 (2006), with a few additional references added by hand. NetScience.gml was created by Mark Newman on July 22, 2006, and was transformed into the Pajek format by Vladimir Batagelj on March 1, 2007[2-9].

The network is weighted, with varying weights assigned as described in M.E.J. Newman, Phys. Rev. E 64, 016132 (2001). The weights represent the strength of scientific collaboration.

NetScience.net valued the undirected network with 1589 vertices and 2742 edges. Each node represents one author; the edges represent both authors' jointly written papers. The value is the MEJ Newman weight. There are no loops or multiple edges included.

### A. Line Weights Distribution

Table 1 shows the line weights distribution in the NetScience network, the lowest value is 0.0526 and the highest value is 4.7500. Figure 1 shows the line weights distribution visually.

TABLE I: LINE WEIGHTS DISTRIBUTION.

line values	frequency
( ... 0.0526]	0
( 0.0526 ... 0.5746]	2256
( 0.5746 ... 1.0965]	371
( 1.0965 ... 1.6184]	57
( 1.6184 ... 2.1404]	22
( 2.1404 ... 2.6623]	22
( 2.6623 ... 3.1842]	7
( 3.1842 ... 3.7061]	4
( 3.7061 ... 4.2281]	2
( 4.2281 ... 4.7500]	1

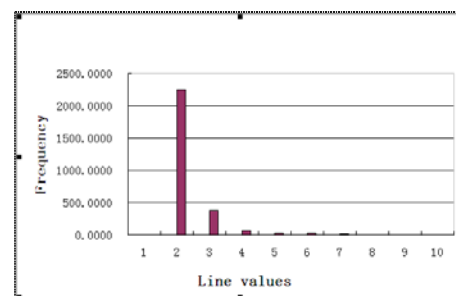


FIGURE I: LINE WEIGHTS DISTRIBUTION.

### B. Density

Density is the quotient of the amount of connections of a given network and the maximum possible amount of connections in the same network [1-3]. It is obvious that a complete network has maximum density. However, density is not the best measure because it depends on the size of the

network, i.e. the amount of possible connections. Density of NetScience network is 0.0021719.

Figure 2 shows the full graph of the NetScience network. Note: several authors did not collaborate at all, which resulted in isolated nodes. For the sake of space, the isolated nodes that represented those authors were deleted.

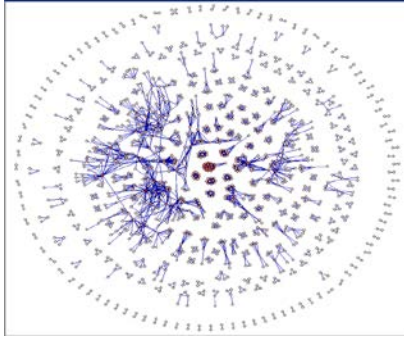


FIGURE II: THE FULL PICTURE OF THE NET SCIENCE NETWORK.

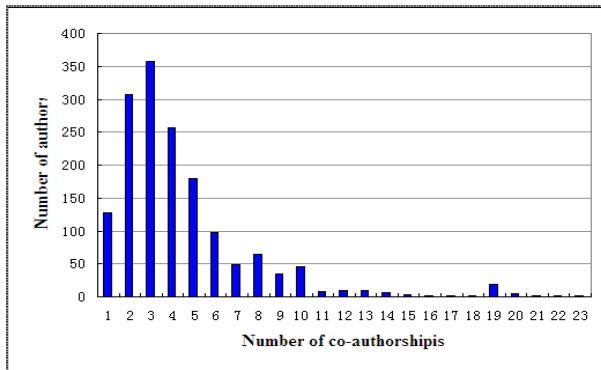


FIGURE III: DEGREE DISTRIBUTION.

### C. Degree Distribution

The degree of a vertex is the number of edges incident to the vertex. Figure 3 shows the distribution of the degree of co-authorships.

## III ANALYSIS OF THE NETSCIENCE DATABASE

In this paper, the techniques of Connectivity, Cohesion and Centrality are applied to the NetScience network analysis.

### A. Connectivity

Connectivity is a concept used to describe the link relation between vertices in the social network. Generally, social networks contain groups of people who are tightly related. Those people who are connected closely are considered as a social subgroup [4-6]. The purpose of the analysis here is to investigate if the specific subgroups differ to the others with respect to social characteristics, e.g. norms, behavior. The Components are used to detect the connectivity in the NetScience database.

#### 1) Components

In an undirected graph, a component is a maximal connected sub graph [7]. Since the NetScience network is undirected, weak and strong components are the same. Figure

4 shows the components in the NetScience network which include at least 20 nodes.

#### 2) Hubs and Authorities

The more important authors can be determined by computing the hubs and authorities of the graph. A good hub represents an author that points to many other authors, and a good authority represents an author that is linked by many different hubs [8].

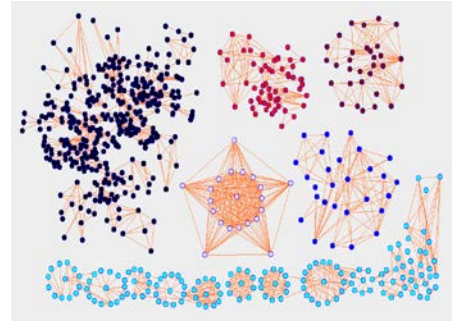


FIGURE IV: COMPONENTS OF THE NET SCIENCE NETWORK.

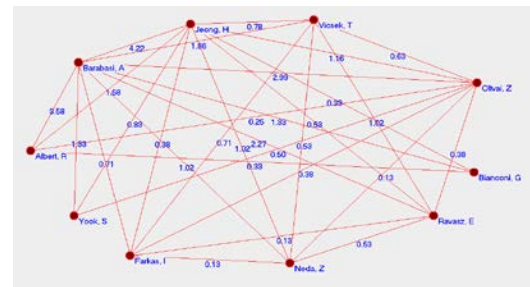


FIGURE V: HUBS AND AUTHORITIES.

For the undirected graph, using Pajek, it is only possible to draw those vertices as both are hubs and authorities. Figure 5 shows ten authors who constitute both hubs and authorities.

### B. Cohesion

Cohesion is a concept to describe the attractive “force” between individuals. In social networks, the dense pockets of people who stick together are called cohesive subgroups [10-11]. The cores are used to detect the cohesion in NetScience database.

#### 1) Cores

A k-core is a maximal sub network in which each vertex has at least degree k within the sub network.

Cores are nested – a vertex in a k-core is also part of (k-1) core, but not all members of a (k-1) core belong to a k-core.

After searching the cores in NetScience network, a hierarchy of cores (0, 1, 2, ..., 19) of the NetScience network were obtained, which is shown in the Table 2.

Whilst searching cores with Pajek, we found that the strongest core is a 19-core, which is shown in Figure 6.

This core is also known as a clique (maximal complete sub network) with a size of 20 (the size of a clique is the number

of vertices in it). This is a very rare coincidence, and means that these 20 authors have connected with each other.

Moreover, Figure 7 shows 9-cores; 19-cores from Figure 6 is also including in the 9-cores (the upper left corner of the figure).

### 2) Islands

Islands are the concept for describing the numerical property of vertices or lines. They represent locally important subnet works on different levels.

Figure 8 shows islands which include 10 to 15 vertices detected by Pajek in the Net Science network. The bottom right red islands and dark red islands are also 9-cores.

TABLE II: HIERARCHY OF CORES OF THE NET SCIENCE NETWORK.

core	frequency	representative
0	128	Agrawal, H
1	320	Adar, E
2	390	Kuperman, M
3	281	Adamic, L
4	223	Acebron, J
5	89	Albert, R
6	21	Fortuna, L
7	60	Alon, U
8	27	Barabasi, A
9	30	Barjoseph, Z
19	20	Giot, L

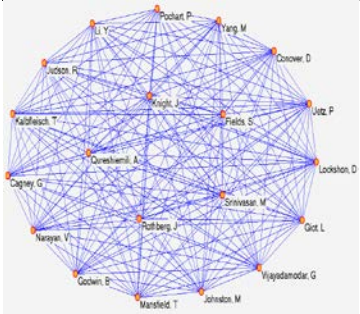


FIGURE VI: 19-CORES OF THE NET SCIENCE NETWORK WHICH IS ALSO A CLIQUE.

### C. Centrality

Centrality is a measure which gives a rough indication of the social power of a vertex based on how well they "connect" the network [12]. Obviously, the most efficient network structure in this case is a star network. "Degree", "closeness", and "betweenness " are three measures of centrality.

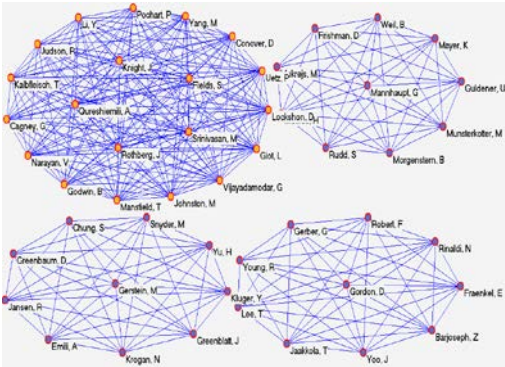


FIGURE VII: 9-CORES OF THE NET SCIENCE NETWORK VERTICES.

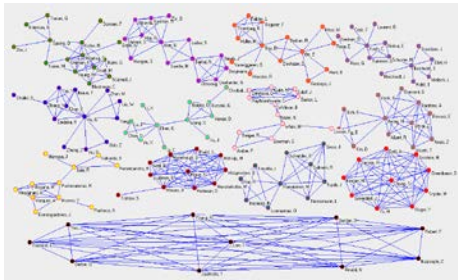


FIGURE VIII: ISLANDS OF THE NETSCIENCE NETWORK WHICH INCLUDE 10 TO 15.

### 1) Degree Centrality

Social network researchers measure network activity for a node by using the concept of degrees - the number of direct connections a node has. It can be defined as the quantity of links occurring upon a node (i.e., the number of ties that a node has). Figure 9 shows the degree centrality of the NetScience network.

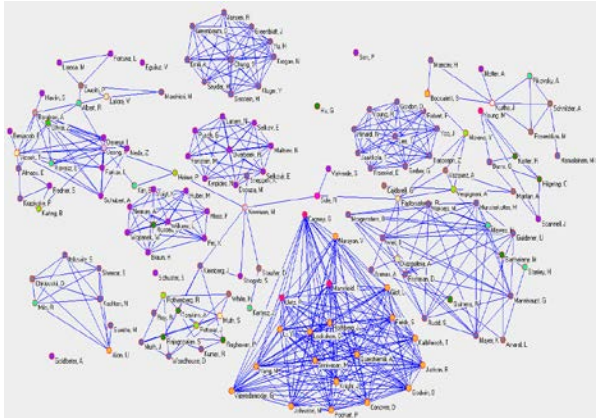


FIGURE IX: DEGREE CENTRALITY OF THE NETSCIENCE NETWORK.

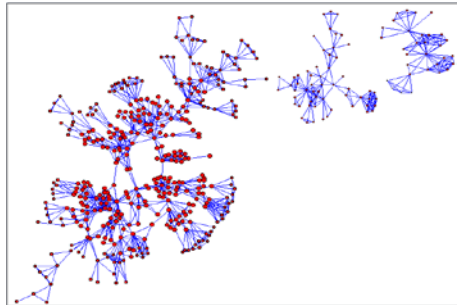


FIGURE X: CLOSENESS CENTRALITY OF THE NETSCIENCE NETWORK.

### 2) Closeness Centrality

In social networks, closeness centrality can be interpreted as how easy is for a node to reach the other nodes in the whole network. Figure 10 shows the closeness centrality of the NetScience network.

Degree and closeness centrality measures were calculated on NetScience network. The top 10 authors in terms of centrality are shown in Table 3.



TABLE III: TOP 10 CENTRAL PERSONS IN THE NETSCIENCE NETWORK.

degree centrality	closeness centrality
Barabasi, A	Newman, M
Newman, M	Sole, R
Oltvai, Z	Pastorsatorras, R
Young, M	Holme, P
Uetz, P	Caldarelli, G
Cagney, G	Vespignani, A
Mansfield, T	Jeong,
Alon, U	Bianconi,
Boccaletti, S	Ghoshal, G
Giot, L	Leicht, E

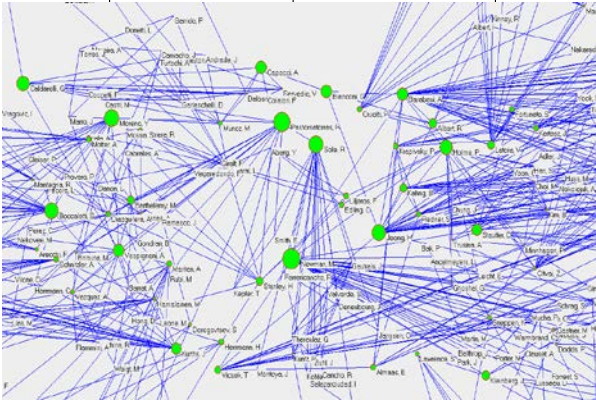


FIGURE XI: BETWEENNESS CENTRALITY OF THE NETSCIENCE NETWORK.

### 3) Betweenness

The concept of betweenness is the extent to which a nodes lies between other nodes in the social network. This measure considers the connectivity of the node's neighbors. It reflects the number of nodes which a node is connecting indirectly through their direct links. Figure 11 shows betweenness of the NetScience network.

From the three pictures of figure 9, 10 and 11, it can be found that the author Newman, M is a very important vertex because of his vast work on the network theory, and his strong power of connecting the different authors.

## IV CONCLUSION

In this paper, we briefly analyzed the NetScience network by using the Pajek software tool. We created abundant of pictures and tables to summarize our findings about the structure and show the most important parts of this network. With the aim of helping the reader to understand discovered patters, a few basic definitions from the graph theory and social network analysis were also presented. We found out some groups of actors whose collaboration is important for the discoveries and development of the theory of networks, and those actors which are important with respect to the whole graph. These patterns were discovered by searching for cores, cliques, islands, and by computing various measures of centrality, respectively.

## REFERENCES

[1] Bowler, W.M. & Brass, D.J., Relational correlates of interpersonal citizenship behavior: A social network perspective. *Journal of Applied Psychology*, **91**(1), pp. 70-82, 2011.

[2] De Nooy, W., Mrvar, A. & Batagely, V., *Exploratory Social Network Analysis with Pajek*, Cambridge University Press: Cambridge, 2011.

[3] Huang, R. & Sun. X., Weibo network, information diffusion and implications for collective action in China. *Information, Communication & Society*, **77**(1), pp. 86-104, 2014.

[4] Jan, N., Levina, A. & Timme, M., Impact of single links in competitive percolation. *Nature Physics*, **7**, pp. 265-270, 2011.

[5] Jones, C. & Volpe, E.H., Organizational identification: Extending our understanding of social identities through social networks. *Journal of Organizational Behavior*, **32**, pp. 413-434, 2011.

[6] Kadushin, C., *Understanding Social Networks: Theories, Concepts, and Findings*, Oxford University Press: New York, 2012.

[7] Kilduff, M. & Daniel J.B., Organizational social network research: Core ideas and key debates. *The Academy of Management Annals*, **4**(1), pp. 317-357, 2010.

[8] Lovejoy, K. & Gregory S. Information, community, and action: How nonprofit organizations use social media. *Journal of Computer Mediated Communication*, **17**(3), pp. 337-353, 2012.

[9] Networks/ Pajek: Program for Large network Analysis, <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>.

[10] Papachristos, A., Murder by structure: dominance relations and the social structure of gang homicide. *American Journal of Sociology*, **115** (1), pp. 74-128, 2009.

[11] Skyler, G. & Desmarais, B., Inferential network analysis with exponential random graph models. *Political Analysis*, **19**(1), pp. 66-86, 2011.

[12] Wang, H. & Barry, W., Social connectivity in America: Changes in adult friendship network size from 2002 to 2007. *American Behavioral Scientist*, **53** (8), pp. 1148-69, 2010.