IDENTIFYING SCIENTIFIC COLLABORATION TRENDS AT THE UNIVERSITY OF MONTENEGRO

SAVO TOMOVIĆ[*], IGOR IVANOVIĆ[†]

[*] University of Montenegro, Faculty of Mathematics and Natural Sciences
Podgorica, Montenegro
e-mail: savot@ac.me, web page: http://www.pmf.ac.me

[†] University of Montenegro, Faculty of Mathematics and Natural Sciences

Summary. Social Network Analysis is often used to examine the scientific collaboration patterns defined by co-authorship relationships. In this paper we examine and quantify the collaborative relations between researchers at the University of Montenegro - UoM. We show the UoM collaboration network does satisfy main characteristics of real world networks: degree distribution follows power law along with coexistence of high clustering coefficient and small-world phenomena. At the moment, the UoM collaboration network is in super-critical regime with a giant component that absorbs about 50% nodes. Although the authors' connectedness at the UoM has been constantly increasing throughout past fifteen years, the network is still in growing phase.

1 INTRODUCTION

Mining social-network graphs is a rapidly growing new area. It is an interdisciplinary field at the crossroad of distinct disciplines deeply rooted in computer science and social sciences[1,2,7,9,13].

Today one of the best-known examples of social networks is the "friends" relation found on sites like Facebook. There is a lot of information to be gained by analyzing big data that is derived from social networks.

Social networks are usually modeled as undirected graphs. The nodes represent entities, and an edge connects two nodes if the nodes are related by the relationship that characterizes the network. In this paper we analyze collaboration network at the University of Montenegro - UoM. The UoM consists of twenty faculties with more than four hundred study programs covering wide variety of research areas: natural sciences, economy, law, languages, engineering, social sciences… During the last reaccreditation period, that will be over in September 2017, the question about collaboration degree between scientists at the UoM has emerged. Although there are remarkable scientists at the UoM whose publications and research work are recognized all around the world, according to the Academic Ranking of World Universities the UoM is positioned very lowly. It is believed that this ranking can be significantly improved by increasing connectedness and collaboration between researchers at the UoM.

In this paper we examine and quantify the collaborative relations between researchers at the UoM using graph based methods. The collaboration network is viewed as a graph in which the nodes are individuals from the UoM and there is an edge between two individuals who published one or more papers jointly. In this study, we do not consider authors outside of the UoM and their possible effect on connectedness between researchers at the UoM, i.e. it is

2010 Mathematics Subject Classification: 62H17, 62J05, 91C20
Key words and Phrases: social networks, degree distribution, power law, small-world property.

possible that two authors from UoM are connected via one or more abroad authors, but for now we ignore these connections.

It comes out that the UoM collaboration network satisfies all characteristic properties of real networks, albeit it is very small, ”local”, still growing network. Also, we prove that high degree nodes do exist in the UoM collaboration network. These nodes are “hubs” representing authors with plenty collaboration at the UoM, but at the same time the study shows that their “neighbors” are still not connected.

2 THE UNIVERSITY OF MONTENEGRO COLLABORATION NETWORK

The University of Montenegro is the only public university in the country, founded in 1974. It has an average number of students of about 21,000. It enrolls more than two thirds of the Montenegrin high school students. The educational process at the University is designed according to the Bologna principles in all three cycles according to the 3+2+3 scheme. The Bologna process has been fully operational at the University since 2003. All aspects dealing with ECTS, the diploma supplement, academic recognition, computerized follow-up of students’ progress etc. have been already implemented. The two first cycles are also fully operational in all faculties, while the third cycle is operational at half of the faculties.

At the UoM there are about 1400 researchers or PhD students, teaching assistants, associate and full professors. Although their work is part of a large number of projects with different funding schemas, there is an impression that there are not enough connections between people interested in different research areas or people from different faculties or even between people from the same faculty and the same research field.

Figure 1. Collaboration network at the University of Montenegro

In this study, the UoM collaboration network is viewed as an undirected graph in which the nodes are individuals from UoM who have published research papers in journals or conferences indexed in SCI, SCIE, SSCI, A&HCI and SCOPUS categories over the period 2000-2015. There is an edge between two individuals who published one or more papers jointly.

We do not consider authors outside the UoM, and consequently neither the links between authors from the UoM and authors beyond; our goal in this study is to examine collaboration degree between researchers exclusively from the UoM. The SCI, SCIE, SSCI, A&HCI and SCOPUS categories are selected because the number of publications in these categories is of great impact on assessing a professor career progress at the University of Montenegro. The data source for this study is portal www.nastava.ucg.ac.me, where one can find publications of researchers at the UoM since 1975.

The UoM collaboration network is presented in Figure 1. The UoM collaboration network consists of 345 nodes and 504 edges. Average node degree is 2.922. Average clustering coefficient is 0.696. Average path length is 6.133. Network density is Δ = 0.0043, which indicates only 0.43% of all possible links being present. Low density indicates low overall cohesion of the network[15]. Network density is calculated as the number of links, L, divided by the $N \times (N-1)$ total possible links, with N being the total number of vertices in the network:

$$\Delta = \frac{L}{N \times (N-1)}.$$

3 REAL NETWORK PROPERTIES DO HOLD IN THE UOM COLLABORATION NETWORK

Real networks share common characteristics. In particular, three networks' attributes exhibit consistent measurements across real-world networks: degree distribution, clustering coefficient, and average path length[3,4,25].

Degree distribution denotes how node degrees are distributed across a network. The clustering coefficient measures transitivity of a network. It represents a measure of the likelihood that two associates of a node are associates themselves. A higher clustering coefficient indicates a greater 'cliquishness'[5,6].

Average path length denotes the average distance (shortest path length) between pairs of nodes. These three properties: power-law degree distribution, high clustering coefficient and small average path length are consistently observed in real-world networks[7].

In this section we show that the previous does hold in the UoM collaboration network.

Also, as it is common, we make parallels between random network model[8] and the UoM collaboration network.

We start with the degree distribution, providing basic results known from the literature, but without any proof due to techniques beyond the scope of this paper. Let k denote the degree of a node, i.e. the number of edges attached to the node. Let p_k denote the fraction of individuals with degree k. Then, in the power-law distribution:

$$p_k = k^{-\gamma}.$$

where γ is a power-law exponent[8]. Taking the logarithm from both sides, we get:

$$\ln(p_k) = -\gamma \cdot \ln(k).$$

The previous equation shows that the log-log plot of a power-law distribution is a straight line with slope $= -\gamma$ and intercept $\ln(k)$. This also reveals a methodology for checking whether a network exhibits a power-law distribution: if a power-law distribution exists, we should observe the straight line in the plot[23, 24].

Degree distribution in the UoM collaboration network is presented in Figure 2. According to the previous considerations, we can conclude that the distribution follows power law.

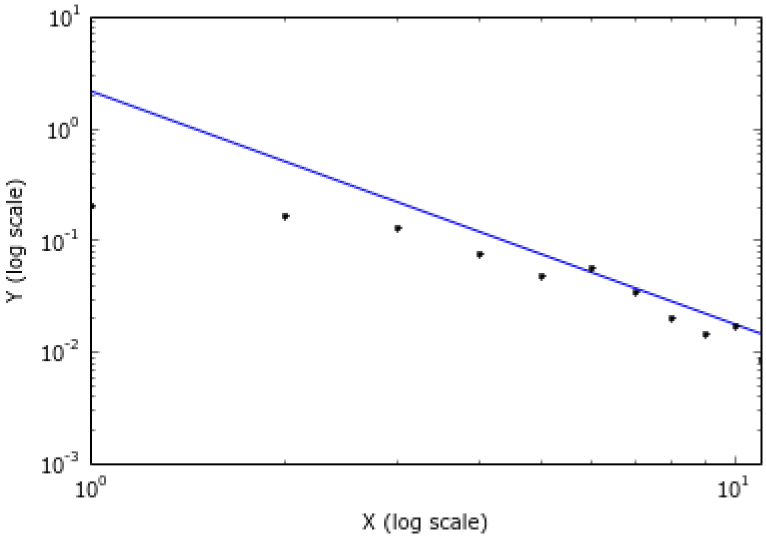

Figure 2. Degree distribution

Networks exhibiting power-law degree distribution are in the literature often called scale-free networks.

Here we will just mention that random graphs generate Poisson degree distribution[3].

The degree of a node does not contain any information about the relationship between the node's neighbors. Local clustering coefficient C_i measures the density of links in node *i*s immediate neighborhood: $C_i = 0$ means that there are no links between *i*s neighbors; $C_i = 1$ implies that each of *i*s neighbors link to each other[10, 11, 12].

The local clustering coefficient of a random network is:

$$C_i = \frac{\langle k \rangle}{N},$$

where $\langle k \rangle$ is average degree in the network with N node[8, 20, 21]. Random network model is considered generally incapable of generating networks with high clustering coefficients without compromising other required properties[9, 22, 25].

We find that the UoM collaboration network has a clustering coefficient 0.696. The observed value is much higher than expected value for a random network model given by the formula $\frac{\langle k \rangle}{N} = 0.0076$ for $\langle k \rangle = 2.922$ and $N = 345$.

According to the evolution scheme of the random network given in[8], we find that the UoM collaboration network has passed the critical point $\langle k \rangle = 1$, and at the moment it is in supercritical regime ($\langle k \rangle = 2.922 > 1$) that has the most relevance to real systems[8].

In summary in the supercritical regime numerous isolated components coexist with the giant component - cluster. These small components are trees, while the giant component contains loops and cycles. The supercritical regime lasts until all nodes are absorbed by the giant component. Additionally, the UoM collaboration network is still far away from the connected regime defined by the condition $\langle k \rangle > lnN = 5.843$.

In real-world network the average path length is small. In other words distance between two randomly chosen nodes in a network is short. This is known as the small-world phenomenon[17,19,24].

We provide mathematical formulation of the small world phenomenon in random networks without proof due to techniques beyond the scope of this paper:

$$\langle d \rangle \approx \frac{lnN}{ln\langle k \rangle}$$

describing the dependence of the average distance in the network on N and $\langle k \rangle$. In general $lnN \ll N$, hence the previous formula implies that the distance in a random network depends logarithmically on the network size. In other words, "small world property" means that $\langle d \rangle$ is proportional to lnN.

In real networks systematic deviations from the previous formula are encountered, but still offering a reasonable approximation to the measured $\langle d \rangle$. Results reported in [3] demonstrate that despite the diversity of systems considered and the significant differences between them in terms of N and $\langle k \rangle$, formula ⟨d⟩≈lnN/ln⟨k⟩ offers a good approximation to the empirically observed $\langle d \rangle$.

In the case of the UoM collaboration network, $\langle d \rangle \approx \frac{lnN}{ln\langle k \rangle} = 5.45$, and it is, reasonable approximation to the empirically observed $\langle d \rangle = 6.133$. In the following sections we give even better approximation, based on the results from[8, 16, 18].

The presence of hubs in scale free networks and their affection to the small world property have been considered in [2, 4]. Hubs are nodes with an exceptionally large degree. In a random network most nodes have comparable degrees and hence nodes with a large number of links are absent. Generally, the presence of hubs decreases average path lengths because they act as joins between small-degree nodes. The previous implies that most scale free networks are not only "small", but "ultra-small"[8, 16, 18].

In the Figure 3, we display degree distribution in the UoM collaboration that indicates hubs presence.

The UoM collaboration network, according to the results from[8, 16, 18], still follows small world rather than ultra-small world property. There are hubs in the network, but they are not sufficiently frequent and numerous to have significant impact on the average distance between the nodes (figure 3). The estimation of $\langle d \rangle$ in networks that follow small world property is given by $\langle d \rangle \approx lnN$ [8, 16, 18]. In the case of the UoM collaboration network $lnN = 5.84$, and this estimation is closer to empirically observed $\langle d \rangle = 6.133$ than the one calculated from the formula ⟨d⟩≈lnN/ln⟨k⟩.

Another aspect of hub presence is illustrated in Figure 4. We plot average clustering coefficient for nodes with different degrees $2 \leq k \leq 17$. As it is expected in real networks, nodes with the highest degrees have the smallest clustering coefficient[14, 25].

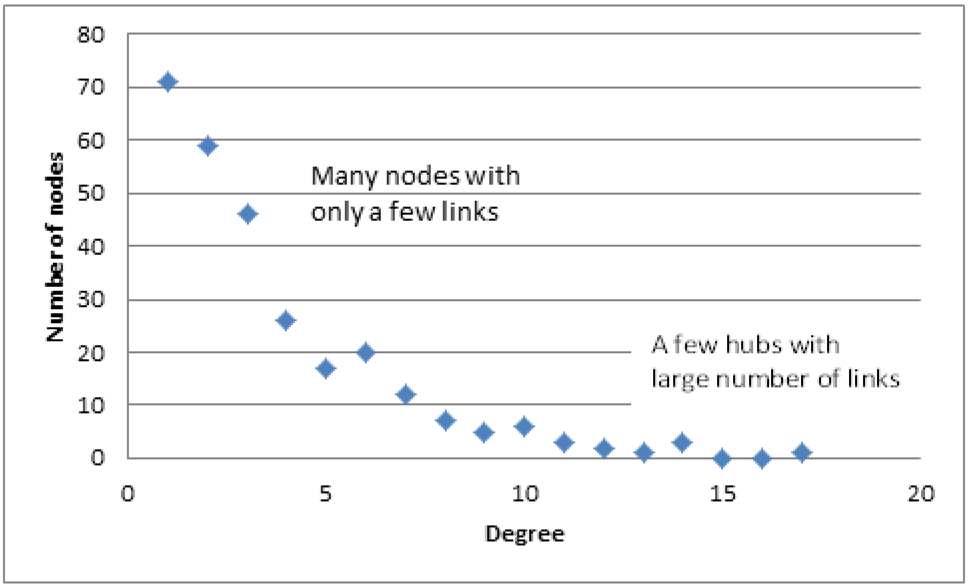

Figure 3. Power law

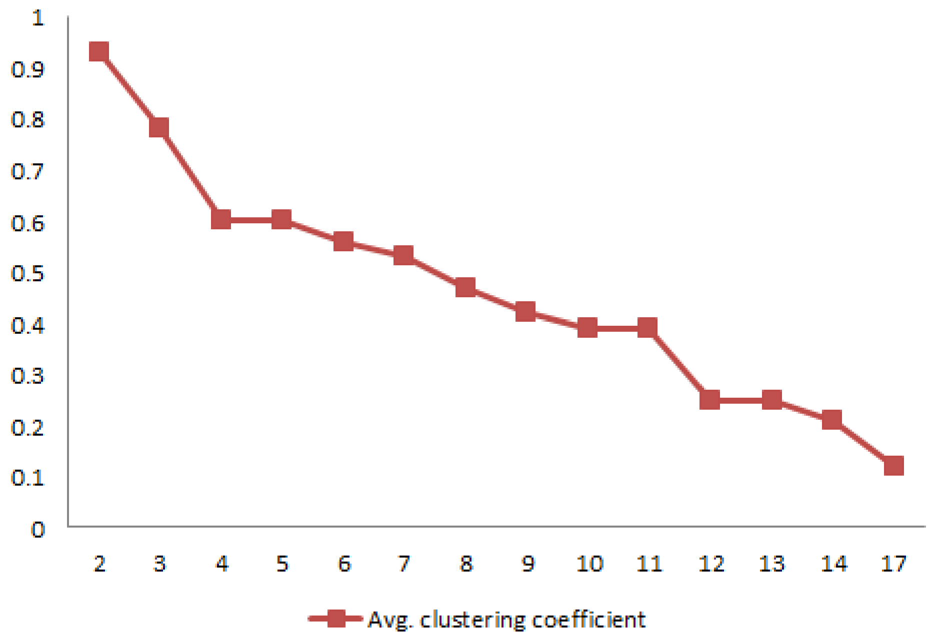

Figure 4. Hubs in the UoM collaboration network

4 EVOLUTION OF THE UOM COLLABORATION NETWORK

In this section we examine the UoM collaboration network's evolution over the period 2000-2015. The following empirical measurements allow us to uncover the time dependence of the topological measures explained in the previous section.

The UoM collaboration network, as it is expected, constantly expands by the addition of new researchers to the database, as well as by the addition of new internal links presenting papers co-authored by authors that have already been part of the database, as Figures 5 and 6 indicate.

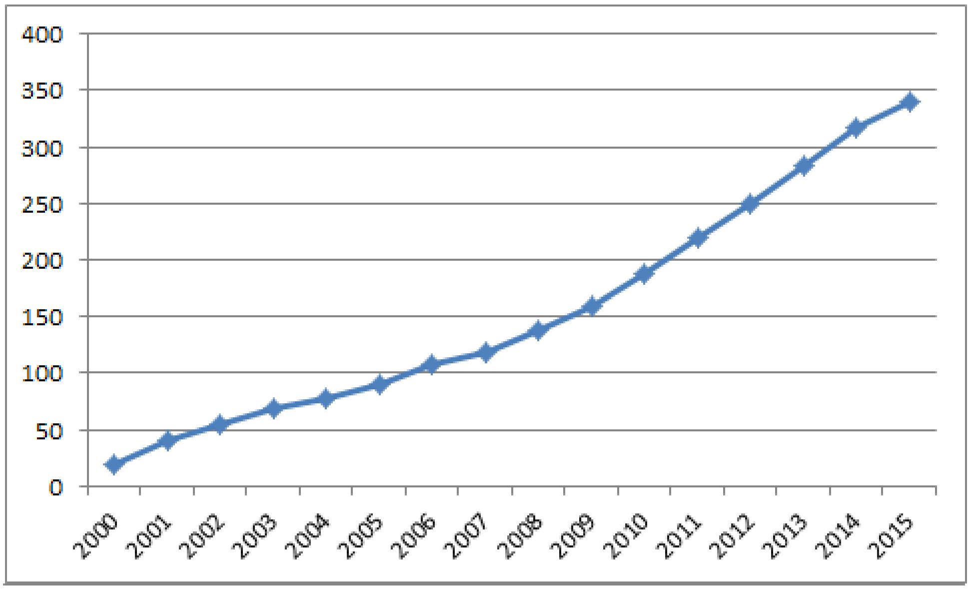

Figure 5. Cumulative number of authors up to the indicated year

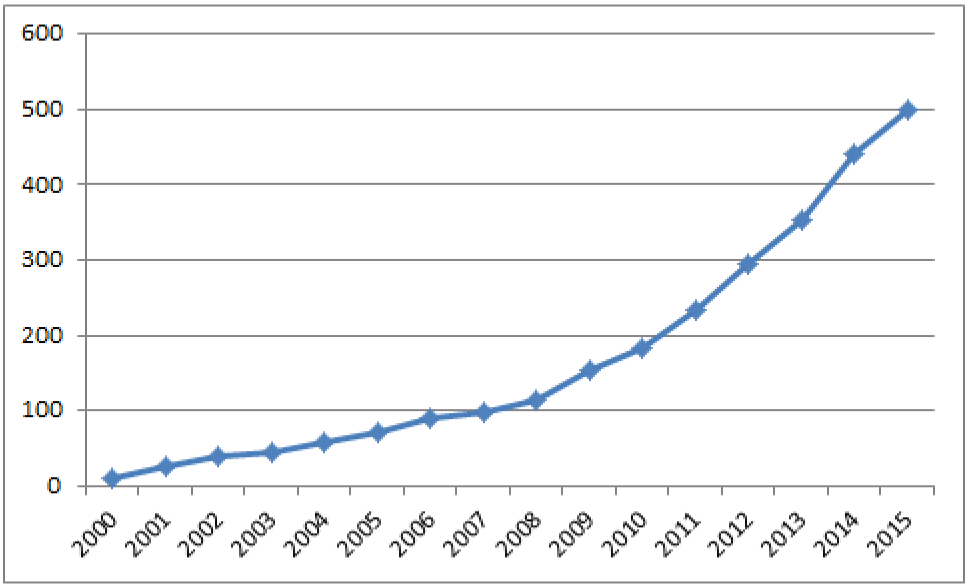

Figure 6. Cumulative number of links up to the indicated year

In the case of the UoM collaboration network the average path length increases with network size. This could have origin in the fact that papers written by ”new” authors exclude the authors that were previously part of the database thus increasing the average path length. The increasing trend in average path length is also predicted by several network models[4, 5, 6].

The clustering coefficient for the UoM collaboration network as a function of time is shown in Figure 8. In simple terms, the clustering coefficient of a node in the co authorship network tells us how much a node's collaborators are willing to collaborate with each other, and it represents the probability that two of its collaborators wrote a paper together [3, 4, 5, 6].

Figure 8 shows that clustering coefficient in the UoM collaboration network during the investigation period is in interval $(0.62, 0.76)$ and constantly and significantly is greater than the expected value in random network model with the same N and $\langle k \rangle$.

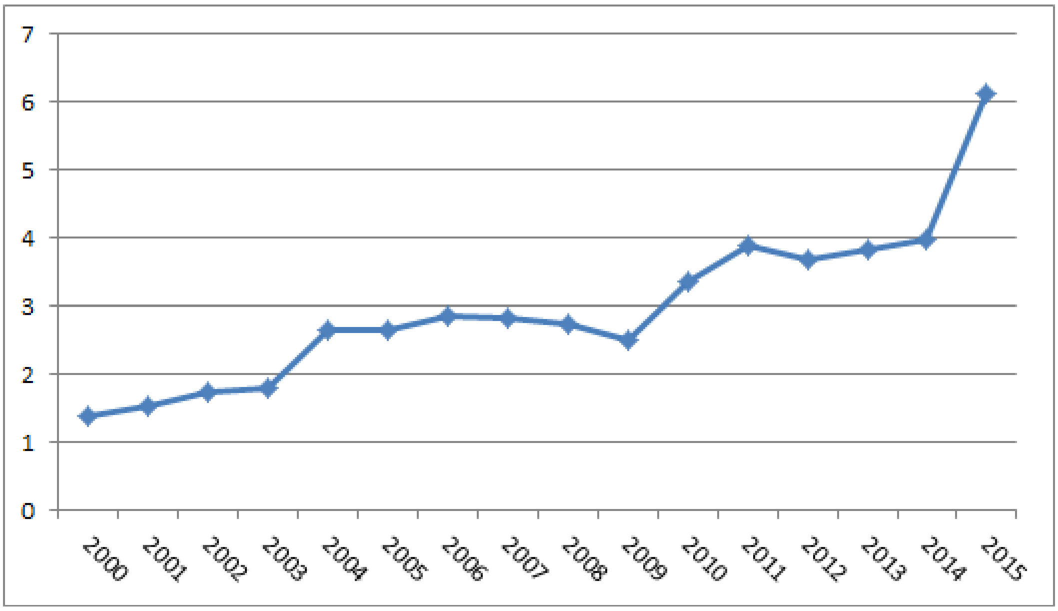

Figure 7.Average path length

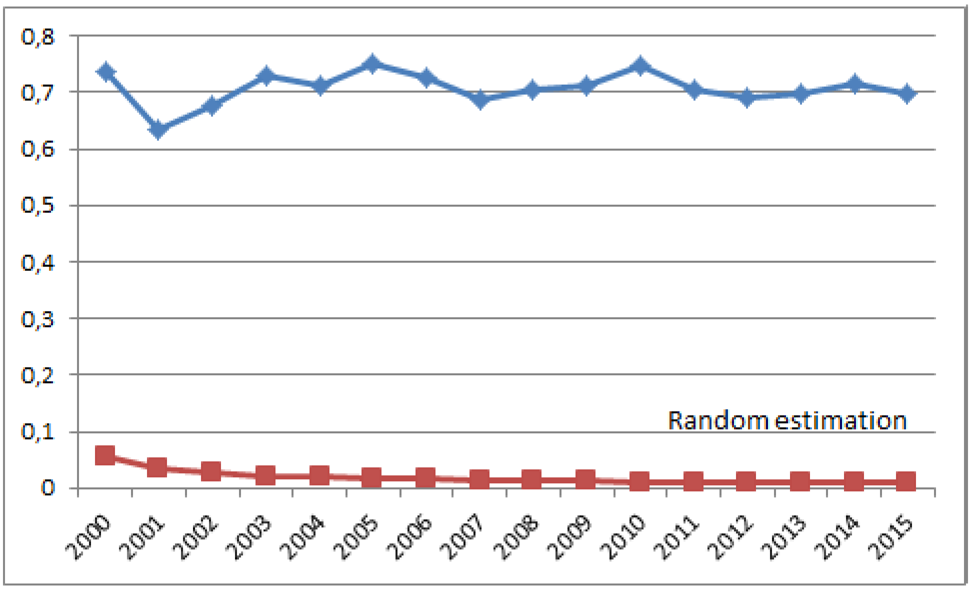

Figure 8. Clustering coefficient

It is clear from Figures 5 and 6 that the number of nodes and links in the network increases due to arrival of new authors as well as connections made by new authors with old ones and due to new connections between old authors. The time dependence of the average degree $\langle k \rangle$ in the UoM network, giving the average number of links per author, is shown in Figure 9. It can be seen that $\langle k \rangle$ approximately linearly increases with time that is important difference with respect to the majority of evolving network models that assume a constant $\langle k \rangle$ as the network grows[4]. As the average degree $\langle k \rangle$ is usually considered as a quantity characterizing the network's interconnectedness[3,4,5,6], we can conclude that the UoM collaboration network shows increasing tendency towards collaboration.

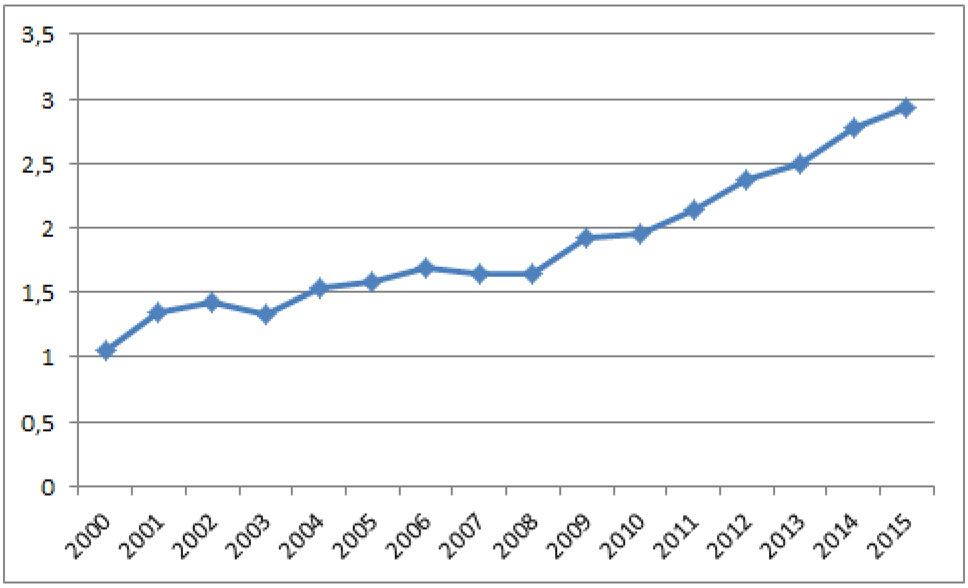

Figure 9. Average degree

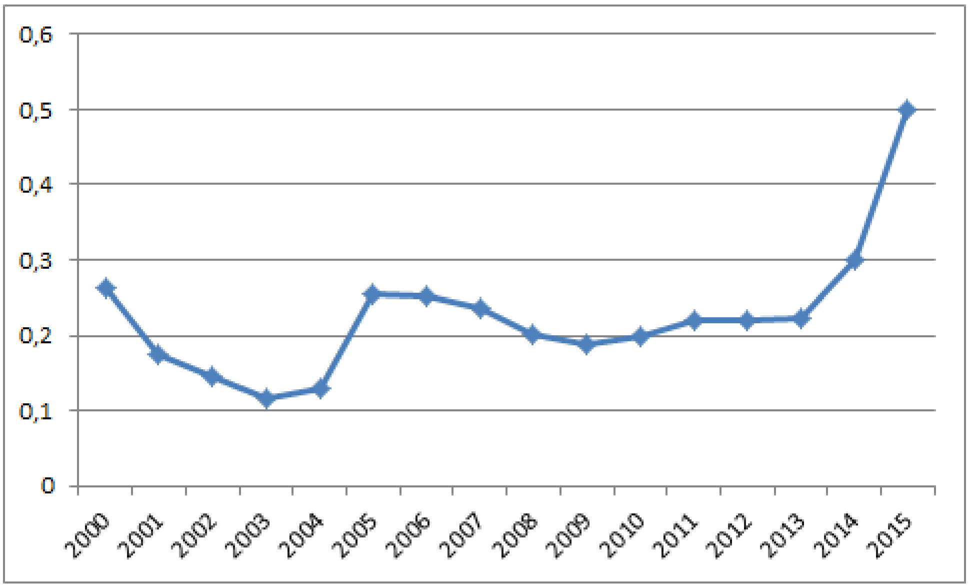

Figure 10. Relative size of the largest cluster on the cumulative data

Finally, we investigated relative size of the largest cluster throughout the period 2000-2015. A cluster is defined as a subset of nodes interconnected by links, while the relative size of the largest cluster is the ratio between the number of nodes in the largest cluster and the total number of nodes in the network[4]. Figure 10 shows that the largest cluster during past five years has been constantly growing, implying that majority of authors do collaborate. This is consistent with the conclusion from the previous section, where it is stated that the UoM collaboration network is in supercritical regime, meaning that there is a giant component (cluster) but far away from connected regime.

5 CONCLUSION

In this paper we present the results of the first study of science collaboration network at the University of Montenegro.

In our study we investigate three major network's measures, i.e. degree distribution, clustering coefficient, and average path length in order to characterize the network. At the moment, the UoM collaboration network is scale-free network, and is residing small world property with a quite high clustering coefficient.

Also, we investigated changes in average path length, clustering coefficient and average degree during the period 2000-2015. Experimental results show that these key measures are time dependent. Again, we emphasize that the UoM collaboration network is analyzed throughout the papers published by authors from the University in SCI, SCIE, SSCI, A&HCI and SCOPUS categories. We do not consider authors outside the UoM, and consequently neither the links between authors from the UoM and authors beyond. Observed trends are comparable to trends in bigger and more famous networks, implying that the UoM collaboration network despite its size and locality still represents a realistic model of a complex evolving real world network.

Also, we provide how scientific collaboration at the UoM has evolved during past fifteen years and identify some trends. The study shows that there are increasing trends towards collaborative research work, but still, we have many researchers that work on they own or collaborate with a small number of colleges.

As a future work we plan to investigate study programs at the University of Montenegro in order to examine their connections and potentials for creating new, modern and interdisciplinary ones. Here we will mention just one curiosity: there is no connection between scientists from different faculties in the UoM collaboration network. This surprising fact says that there is a lot of work to be done in order to achieve necessary connectedness degree in scientific community in Montenegro and avoid embeddedness of authors within their affiliations.

REFERENCES

[1] A.E. Abraham, *Computational Social Networks: Mining and Visualization*, Springer, SpringerVerlag, London (2012).

[2] C. Aggarwal, K. Subbian, "Evolutionary network analysis: A survey", *ACM Computing Surveys,* **47** (1), (2014).

[3] A.L. Barabasi, *Network Science*, Cambridge University Press, Cambridge, UK, (2016).

[4] A.L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, T. Vicsek, "Evolution of the social network of scientific collaborations", *Physica A*, 590 -614 (2002).

[5] A.L. Barabasi, A. Reka, "Emergence of scaling in random networks", *Science,* **286**, 509 -512 (1999).

[6] S.P. Borgatti, M.G. Everett, "A graph-theoretic perspective on centrality", *Social Networks*, **28** (4), 466 -484 (2006).

[7] D. Braha, A. Minai, Y.E. Bar-Yam, *Complex Engineered Systems: Science Meets Technology,* Springer, Springer-Verlag, London (2006).

[8] Capocci, V.D.P. S ervedio, G. Caldarelli, F. Colaiori, " Detecting communities in large networks", *Physica A: Statistical Mechanics and its Applications,* **352**, 669 -676 (2005).

[9] D. De Stefano, D. Giordano, G. Vitale, M. Prosperina, "Issues in the analysis of co-authorship networks", *Quality and Quantity,* **45** (5), 1091 -1107 (2011).

[1] J. Duch, A. Arenas, "Community detection in complex networks using extremal optimization", *Physical Review E* , **72** (2), 371 -385 (2005).
[2] M. Erfanmanesh, V. Rohani, A. Abrizah, "Co-authorship network of scientometrics research collaboration", *Malaysian Journal of Library and Information Science*, **17** (3), 73 -93 (2012).
[3] S. Fortunato, V. Latora, M. Marchiori, "Method to find community structures based on information centrality", *Physical Review E*, **70** (5) (2004).
[4] C.L. Freeman, *The development of social network analysis - with an emphasis on recent events,* The SAGE Handbook of Social Network Analysis, (2011).
[5] M. Girvan, M.E.J. Newman, "Community structure in social and biological networks", *Proceedings of the National Academy of Sciences of the United States of America,* 7821 -7826 (2002).
[6] S. Hennemann, D. Rybski, I. Liefner, "The myth of global science collaboration - collaboration patterns in epistemic communities", *Journal of Informetrics,* **6** (2), 217 -225 (2011).
[7] Landherr, B. Friedl, J. Heidemann, "A critical review of centrality measures in social networks", *Business Information Systems Engineering*, **2** (6), 371 -385 (2010).
[8] J. Leskovec, J. Kleinberg, C. Faloutsos, "Graph evolution: Densification and shrinking diameters", *ACM Transactions on Knowledge Discovery from Data,* **1**(1), (2007).
[9] P. Pons, M. Latapy, "Computing communities in large networks using random walks", *Journal of Graph Algorithms and Applications,* **10** (2), 191 -218 (2006).
[10] F. Radicchi, C. Castellano, F. Cecconi, V. Loretto, D. Parisi, " Defining and identifying communities in networks", *Proceedings of the National Academy of Sciences of the United States of America,* 2658 -2663 (2004).
[11] U.N. Ragnaven, R. Albert, S. Kumara, " Near linear time algorithm to detect community structures in large-scale networks", *Physical Review E*, **76,** (2007).
[12] J. Reichardt, S. Bornholdt, "Statistical mechanics of community detection", *Physical Review E* , **74**, (2006).
[13] C.L. Staudtm, *Analysis of scientific collaboration networks: Social factors, evolution, and topical clustering,* Diplomarbeit, Institut fr Theoretische Informatik Prof. D. Wagner, KIT (2011).
[14] S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications,* Cambridge University Press, (1994).
[15] D.J. Watts, S.H. Strogatz, "Collective dynamics of "small-world" networks", *Nature,* **393**, 409 - 410 (1998).
[16] R. Zafarani, M.A. Abassi, H. Liu, *Social Media Mining,* Cambridge University Press, Cambridge, UK (2014).

Received December, 15 2015