The Eurasia Proceedings of
Educational & Social Sciences (EPESS)
ISSN: 2587-1730

The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2016

Volume 5, Pages 196-202

ICRES 2016: International Conference on Research in Education and Science

CO-AUTHORSHIP NETWORK COMPARISON OF FOUR TURKISH UNIVERSITIES

Ilker Turker
Karabuk University, Department of Computer Engineering

Rafet Durgut
Karabuk University, Department of Computer Engineering

Abdullah Cavusoglu
Member of the Council for Higher Education (YÖK), Council for Higher Education

Abstract: Co-authorship networks are remarkable applications of complex networks, an interdisciplinary framework of exploring systems composed of smaller components with numerous interconnections. Generating connections (links) between collaborating authors, we constructed co-authorship networks of four Turkish universities as İstanbul Technical University, Selçuk University, Sakarya University and Karabük University. Along with the node and edge counts, we investigated the time evolution of network parameters like average degree, modularity, clustering coefficient and average path length in yearly resolution. We also outlined the effect of being a first-mover (or late) as a university in terms of the network parameters.

Keywords**:** Scientific collaboration networks, complex networks, co-authorship networks, scale free networks.

Introduction

Network science provide a comprehensive framework for understanding the self-organizing complex systems across many fields (Strogatz, 2001). Interdisciplinary studies of this topic span network-based approaches to various systems like biological systems, neural networks, spatial games, genetic networks, food webs, computer networks, power grid networks, linguistic networks, social networks, the network of film actors and many other self-organizing systems including scientific collaboration networks (SCNs) (Albert & Barabasi, 2002; Barabasi et al., 2002; Watts & Strogatz, 1998). These systems are modeled by considering the components as nodes, while the interconnections between these nodes define the links (Barabasi & Albert, 1999). Evaluating such systems as a network provides a broader approach to understand the behaviors of the whole system, by detailing the inner structure defined by the complex interconnections between numerous nodes. By the way, we can understand how information spreads in a computer network or predict how an epidemic spreads through the existing links, and also which possible link may be established between nodes (Girvan & Newman, 2002; Newman, 2002; Park & Newman, 2004).

Inspiring from various diverse systems, these networks display some generic properties like *small-world* or *scale-free* property. The former means, even if consisting of numerous nodes, in these systems one can find a short path from one node to another. The latter stands for the distribution of the number of neighbors for a node is consistent with power-law decay (Albert & Barabasi, 2002; Amaral, Scala, Barthelemy, & Stanley, 2000; Clauset, Shalizi, & Newman, 2009; Virkar & Clauset, 2014). Beyond these properties, these diverse systems display consistence in several network parameters that will be described in the next section, indicating that their inner structures are driven by the same organizing principles (Albert & Barabasi, 2002).

As a prototype of complex networks, SCNs provide an understanding of self-organizing systems mentioned above. SCNs also improve the bibliographic view on scientific collaborations between academicians, identifying the interactions between them in detail (Wagner & Leydesdorff, 2005).

- Selection and peer-review under responsibility of the Organizing Committee of the conference

*Corresponding author: Ilker Turker-Mail: iturker@karabuk.edu.tr

SCNs are dynamic systems, expanding by the addition of new authors. The existing authors also establish new links by coauthoring a new paper together. The scientific community as a network, exhibits clustering and preferential attachment properties as well as the other universal properties mentioned above. Clustering means that your friends (or collaborators) are also friends of each other. Preferential attachment means a new node (or author) likely prefers to attach (collaborate with) the more popular nodes (i.e. nodes having more neighbors, more degree), as a behavior that solidifies power-law degree distribution. These universal properties are observed in the previous studies about SCNs of several scientific disciplines like engineering, mathematics, physics, surgery etc. or of interdisciplinary view (Barabasi & Albert, 1999; Barabasi et al., 2002; Cavusoglu & Turker, 2013, 2014; Luzar, Levnajic, Povh, & Perc, 2014; Newman, 2001a, 2001b, 2001c, 2004). Nationwide analysis of SCNs together with the international analysis are also performed by several scientists (Cavusoglu & Turker, 2013; Ferligoj, Kronegger, Mali, Snijders, & Doreian, 2015; Hoekman, Frenken, & Tijssen, 2010; Luzar et al., 2014; Ma, Fang, Pang, & Li, 2014; Perc, 2010).

Having performed the analysis of Turkish SCN spanning a time window of more than 30 years (Cavusoglu & Turker, 2013, 2014), we focus on the differences in the network dynamics of four distinct universities of Turkey in this study. The four universities are İstanbul Technical University (İTU), Selçuk University (SU), Sakarya University (SAU) and Karabuk University (KBU). We picked these universities as representing a range from short to long established ones (KBU, SAU, SU and İTU respectively) as well as locating in the small to crowded cities (with the same order). The numeric and distributional network parameters are outlined distinctly in comparative view in the next section.

Methods and Results

We constructed a co-authorship network using ISI Web of Science Data collected from the search interface. We employed a filtering constraint to achieve the publications addressed to Turkey, starting from the year 1980 to 2015. An additional constraint of the university name is used to achieve four different datasets that belong to the four different universities mentioned above. Since the foundation dates of the two universities (SAU and KBU) are later than this date (1992 and 2007 respectively), their datasets span the time intervals after these foundation years. We used the cumulative downloading utility of Web of Science to achieve the data in sets of 500 publications' data in each bin.

Parsing the raw data using a computer program coded in C#, we constructed the *nodes* (authors) and *edges* (links) tables where each collaboration pair in a scientific paper defines a link connecting the two authors as *source* and *target* nodes. We performed the network analysis with Gephi (Bastian, Heymann, & Jacomy, 2009) and achieved the main network parameters for the four distinct universities in yearly resolution. The mentioned network parameters are the *average degree*, *modularity*, *clustering coefficient*, *average path length* and *network diameter* respectively, which are detailed separately below.

Number of Nodes and Links

We have 4 distinct datasets spanning the publication data for each university. A statistical view to these datasets is presented in Table 1. The publications in each dataset includes at least one author from that university but also includes some authors that collaborate with an author from that university. Thus, a dataset of that universities publications include the authors of that university plus a number of authors from the whole world other than that university. This is the reason for the difference between the number of nodes and the number of academicians for a specific university in Table 1. The table also includes the number of links, each of which are established by a collaboration in a scientific paper. A collaboration between 3 authors (let's say the authors A, B and C write a paper together) construct 3 undirected links A-B, A-C and B-C.

Table 1. Number of nodes and links for four universities

	İTU	**SU**	**SAU**	**KBU**
Num. of Nodes *(a)*	15,707	13,733	6,370	1,371
Num. of Academicians* *(b)*	2,226	2,677	1,964	917
Num. of Links *(c)*	83,825	80,220	35,541	3,468
a / b	7.06	5.13	3.24	1.5
c / a	5.34	5.84	5.58	2.53

* Data retrieved from Turkish Council of Higher Education (YÖK) reports (YÖK, 2015)

The a/b rate indicates the tendency of establishing inter-university links, while the c/a rate indicates the interconnectedness of the nodes for that university. From Table 1, we can say that the *a/b* rate is related with the establishment date of the universities, or the scientific history and impact briefly. İTU is the most attractive university, having nodes count 7 times of its academicians count. This rate is 1.5 for KBU, having a short scientific history of 8 years until this study. The *c/a* rate is about 5.5 for the first 3 university and 2.53 for KBU, having a small interconnectedness rate compared to the long-established universities.

Average Degree

With the time, the number of nodes (authors) participating in the network increase. In the other hand, the number of links increase by the new collaborations established between the both existing and new coming authors. Average degree ‹k› is a quantity that measures the number of links per author (Barabasi et al., 2002). As ‹k› increases, the network gets more interconnected. In the other words, the authors have more collaborators in average.

We present the time dependency of average degree ‹k› for the 4 universities in Fig.1. In this graph, the data is evaluated from 1980 to the indicated date that starts from 1990 to 2015 (i.e. the point 1995 corresponds to the data from 1980 to 1995 and so on). SU remains the most interconnected community from 1990 to 2015. The average degree values are not strictly effected from the foundation date that, İTU has a moderate average degree value compared to the remaining universities. The fluctuations in the graph, for example the steep increment between 2013 and 2014 for SAU, are especially driven by the papers of physics or surgery research studies that are collaborated by hundreds of researchers. For example, in our recent study (Cavusoglu & Turker, 2014), we encountered some papers having more than 500 authors in physics area. By this point of view, high average degree is an indication of physics and surgery studies performed in that university.

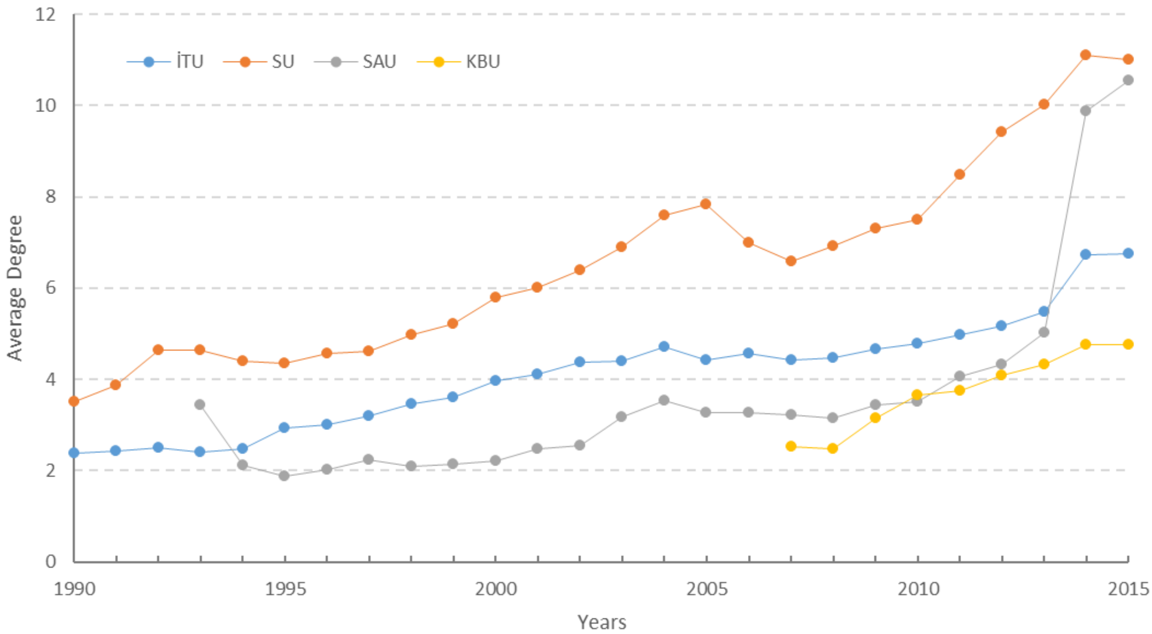


Figure 1. Average degree in years

Modularity

A measure for detecting the community structure in networks is modularity. It means the emergence of densely interconnected groups of nodes, having sparse connections to the other groups (Leicht & Newman, 2008). For a network having high modularity (i.e. close to 1), we can say that the network structure consists of subgroups named as communities. The evolution of modularity is plotted in Fig. 2 below, in which all the 4 networks converge to a modularity slightly above 0.9. This means that, the scientific communities dominantly consist of modules that perform inter-community studies and do not tend to collaborate with the other modules. The most modular networks in this study belong to İTU and KBU academicians.

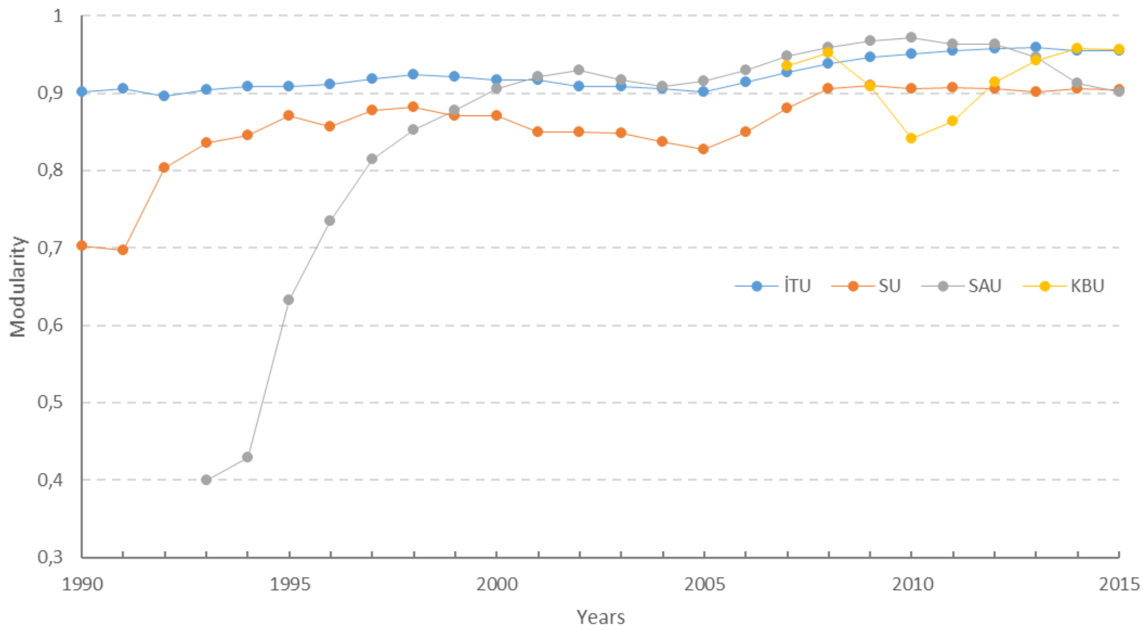


Figure 2. Modularity in years

Clustering Coefficient

The clustering coefficient of a node in a co-authorship network evaluates how much a node's collaborators are willing to collaborate with each other, and represents the probability that two of its collaborators wrote a paper together (Barabasi et al., 2002). This parameter is either calculated for each node and averaged for all the nodes in a network, or calculated as a global parameter counting all the triangles and triples in the whole network (Opsahl, 2013).

Unlike the random networks that the links are generated randomly, real networks display high clustering property. This generic property is also observed in nationwide scientific collaboration networks (Cavusoglu & Turker, 2013). To investigate that this rule is also evident in a narrow scope like university, we calculated the clustering coefficient values for all the datasets in yearly resolution, starting from 1990. Each years' value represents the cumulative data from 1980 to that year in Fig. 3.

Similar with the modularity measure, average clustering coefficients for the networks yield high values between 0.8 and 0.9, indicating that the collaborators of an author are also collaborators of each other at a high rate. The four distinct networks converge to similar trends and values, while the starting values vary by the effects of limited data in 1990s.

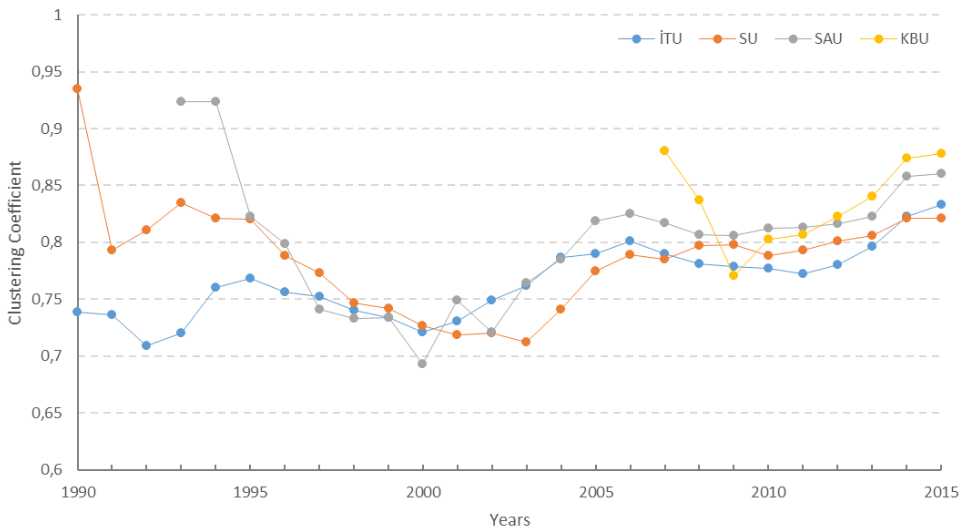


Figure 3. Clustering coefficient in years

Average Path Length

Real networks are similar with random networks in average path length (distance). The path length between two nodes is defined as the number of edges along the shortest path connecting them. In real networks, we can find a short path between one node to another and this rule is valid for the majority of the node pairs in the network, regardless from how large the network is. This generic property of real networks is known as "small-world phenomenon". The most popular manifestation of small worlds is the ''six degrees of separation'' concept, uncovered by Stanley Milgram, who concluded that the average path length between people as a friendship is typically about six in real world (Albert & Barabasi, 2002; Milgram, 1967).

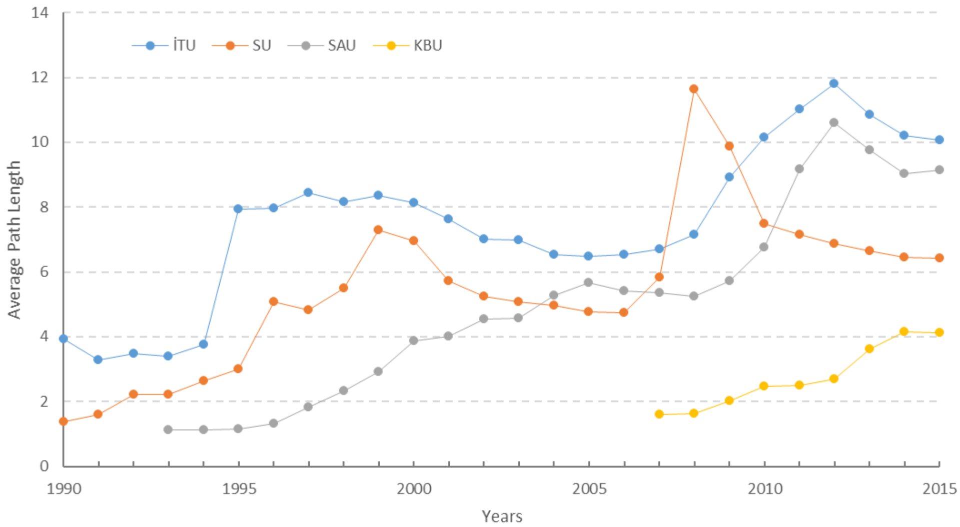
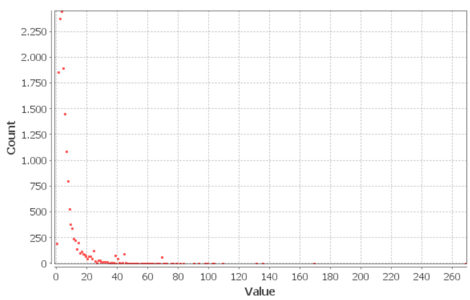


Figure 4. Average path length in years

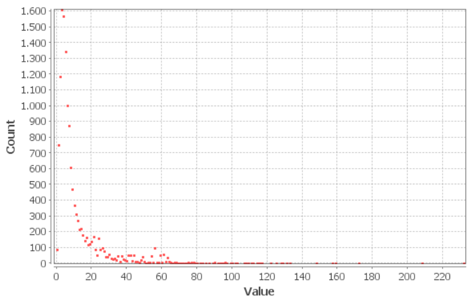
The average distance data is presented in Fig. 4, yielding increasing trends for all university networks. The resulting average distance is about 6 for SU, consistent with the "six degrees of separation" concept, while KBU underestimates and SAU and İTU overestimates this generic value. All the co-authorship networks have considerably short distances between the authors, thus can be classified as small-world networks. The high separation values for İTU and SAU may be driven by the collaborations with foreign authors, constructing specific paths that are only accessed via one author corresponding that university.

Degree Distribution

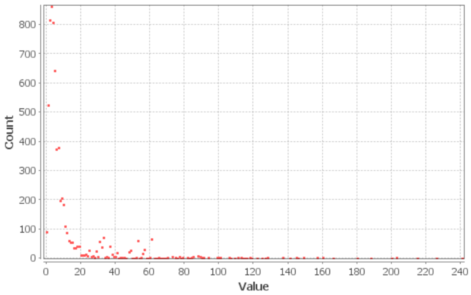
Degree distribution *p(k)*, states out the probability that a randomly selected node has k links. Networks for which *p(k)* has a power-law tail, are known as scale-free networks (Barabasi & Albert, 1999; Barabasi et al., 2002; Newman, 2003). Real networks are scale-free in general. This property of real networks diverges them from random networks, which have Poisson like degree distribution. We present the degree frequency graphs in Fig. 5 for the 4 university networks respectively.



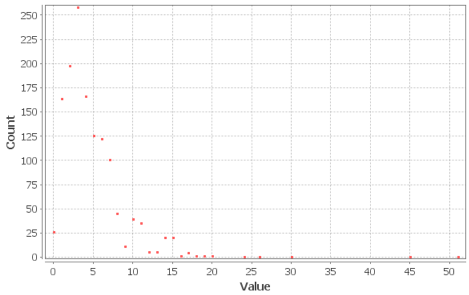
A



B



C



D

Figure 5. Degree occurrence frequencies for (A) İTU, (B) SU, (C) SAU and (D) KBU

In Fig. 5, the distributions indicate that the networks yield scale-free characteristics. The last plot (for KBU) spans a narrower degree range from 0 to 50, since the network is in a younger phase. Similar with the national (Cavusoglu & Turker, 2013; Ferligoj et al., 2015; Perc, 2010) or research area based (Barabasi et al., 2002; Cavusoglu & Turker, 2014; Hou, Kretschmer, & Liu, 2008; Newman, 2001a) scientific collaboration networks, university based networks display power-law distributions as well. The underlying mechanism behind a scale-free network is reported to be preferential attachment, that means a new node tends to connect with an existing node that has higher degree (or popularity) than the low degree ones (Vazquez, 2003). So we can say preferential attachment is the ingredient of the four network we investigated in this study.

Conclusion

The four co-authorship networks of distinct universities of Turkey display similar generic properties that are reported for national or disciplinal co-authorship networks before. SU and SAU authors are the most productive, having the highest average degrees (co-authors) around 11. All the university networks are highly modular, resulting modularity measures over 0.9. This fact indicates that inter-community studies dominate scientific networks. Similarly, clustering coefficient is very high for all of the networks, yielding values in the interval 0.8-0.9. This is an indicator that micro-modular cliques occur in co-authorship patterns that means the neighbors of an author are also neighbors of each other, regardless from the central author. As an indicator of small-world property, all the networks relatively small average path length measures. Affected from the limited data KBU has the smallest, while the others display average path length between 6 and 10. İTU has the longest average distance value 10, yielding that the scientists are less likely to form less unexpected collaborations that reduce path length. Another fact may be the absence of surgery departments, which result numerously collaborated papers like physics domain. İTU also has the most attractive author set, that 1 author approximately attracts 6 external authors (inspired from *a/b* ratios). This rate is about 4, 2.2 and 0.5 for SU, SAU and KBU respectively. Average links per nodes (*c/a*) values are about 5.5 for the three primal universities, while this rate is about 2.5 for KBU as a new and developing university. This rate is different from average degree rates because every link does not mean a new collaborator, since it may be a repetition of a previously established link to a specific author.

The effect of being a first-mover as a university is rather observed in links-per-nodes (*c/a*) and nodes per academicians (*a/b*) rates, average degree and average path length measures. Authors of a developing university are less distant, less productive and less collaborative that the old ones. But the modularity and clustering properties are identical with the old ones, yielding that clustering is a characteristic property for scientific communities.

References

Albert, R., & Barabasi, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74*(1), 47-97. doi: 10.1103/RevModPhys.74.47

Amaral, L. A. N., Scala, A., Barthelemy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America, 97*(21), 11149-11152. doi: 10.1073/pnas.200327197

Barabasi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509-512. doi: 10.1126/science.286.5439.509

Barabasi, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica a-Statistical Mechanics and Its Applications, 311*(3-4), 590-614. doi: 10.1016/s0378-4371(02)00736-7

Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi: an open source software for exploring and manipulating networks*. Paper presented at the International AAAI Conference on Weblogs and Social Media, San Jose, California.

Cavusoglu, A., & Turker, I. (2013). Scientific collaboration network of Turkey. *Chaos Solitons & Fractals, 57*, 9-18. doi: 10.1016/j.chaos.2013.07.022

Cavusoglu, A., & Turker, I. (2014). Patterns of collaboration in four scientific disciplines of the Turkish collaboration network. *Physica a-Statistical Mechanics and Its Applications, 413*, 220-229. doi: 10.1016/j.physa.2014.06.069

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *Siam Review, 51*(4), 661-703. doi: 10.1137/070710111

Ferligoj, A., Kronegger, L., Mali, F., Snijders, T. A. B., & Doreian, P. (2015). Scientific collaboration dynamics in a national scientific system. *Scientometrics, 104*(3), 985-1012. doi: 10.1007/s11192-015-1585-7

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 7821-7826. doi: 10.1073/pnas.122653799

Hoekman, J., Frenken, K., & Tijssen, R. J. W. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy, 39*(5), 662-673. doi: 10.1016/j.respol.2010.01.012

Hou, H., Kretschmer, H., & Liu, Z. (2008). The structure of scientific collaboration networks in Scientometrics. *Scientometrics, 75*(2), 189-202. doi: 10.1007/s11192-007-1771-3

Leicht, E. A., & Newman, M. E. J. (2008). Community structure in directed networks. *Physical Review Letters, 100*(11). doi: 10.1103/PhysRevLett.100.118703

Luzar, B., Levnajic, Z., Povh, J., & Perc, M. (2014). Community Structure and the Evolution of Interdisciplinarity in Slovenia's Scientific Collaboration Network. *Plos One, 9*(4), 5. doi: 10.1371/journal.pone.0094429

Ma, H. T., Fang, C. L., Pang, B., & Li, G. D. (2014). The Effect of Geographical Proximity on Scientific Cooperation among Chinese Cities from 1990 to 2010. *Plos One, 9*(11), 11. doi: 10.1371/journal.pone.0111705

Milgram, S. (1967). The small-world problem. *Psychology Today, 1*(1), 61-67.

Newman, M. E. J. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E, 64*(1), 8.

Newman, M. E. J. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E, 64*(1), 7.

Newman, M. E. J. (2001c). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 404-409. doi: 10.1073/pnas.021544898

Newman, M. E. J. (2002). Spread of epidemic disease on networks. *Physical Review E, 66*(1), 11. doi: 10.1103/PhysRevE.66.016128

Newman, M. E. J. (2003). The structure and function of complex networks. *Siam Review, 45*(2), 167-256. doi: 10.1137/s003614450342480

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*, 5200-5205. doi: 10.1073/pnas.0307545100

Opsahl, T. (2013). Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks, 35*(2), 159-167. doi: 10.1016/j.socnet.2011.07.001

Park, J., & Newman, M. E. J. (2004). Statistical mechanics of networks. *Physical Review E, 70*(6), 13. doi: 10.1103/PhysRevE.70.066117

Perc, M. (2010). Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics, 4*(4), 475-482. doi: 10.1016/j.joi.2010.04.003

Strogatz, S. H. (2001). Exploring complex networks. *Nature, 410*(6825), 268-276. doi: 10.1038/35065725

Vazquez, A. (2003). Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E, 67*(5), 15. doi: 10.1103/PhysRevE.67.056104

Virkar, Y., & Clauset, A. (2014). POWER-LAW DISTRIBUTIONS IN BINNED EMPIRICAL DATA. *Annals of Applied Statistics, 8*(1), 89-119. doi: 10.1214/13-aoas710

Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy, 34*(10), 1608-1618. doi: 10.1016/j.respol.2005.08.002

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440-442. doi: 10.1038/30918

YÖK. (2015). Akademisyen Başına Öğrenci Sayısı Listesi. from http://akademisyenler.org/wp-content/uploads/2015/05/Akademisyen-Ba%C5%9F%C4%B1na-%C3%96%C4%9Frenci-Say%C4%B1s%C4%B1-Listesi.pdf