International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES'18),
May 11-13, 2018 Safranbolu, Turkey

Evaluating Research Performance of the European Countries Through Social Network Analysis

İ. F. ŞENTÜRK[1] and B. TAŞ[2]

[1] Bursa Technical University, Bursa/Turkey, izzet.senturk@btu.edu.tr
[2]Bursa Technical University, Bursa/Turkey, burak.tas@btu.edu.tr

***Abstract* - Social network analysis investigates relationships in a networked structure in order to interpret the roles of individuals and evaluate their respective importance. The common approach is employing graph theory that models the network as a graph data structure. Graphs are mathematical abstractions to model pairwise relations between objects. A graph is comprised of nodes connected with edges. While nodes indicate individuals in a network, edges signify relationships or interactions. In this paper, we evaluate the research performance of the European countries considering the research activities within the framework of European Cooperation in Science and Technology (COST). Founded in 1971, COST is the oldest and widest scientific intergovernmental framework in Europe supporting 37 countries including Turkey. COST can be considered as an incubator to set up interdisciplinary research networks since it provides support for network activities such as meetings, training schools, short scientific exchanges, etc. but does not fund the research itself. Therefore, we believe that the research network under COST can be a good indicator to analyze relationships between countries in research activities and evaluate the research performance of the countries. In this paper, we considered research actions funded by COST between 2012-2017 and evaluated research performance of the countries according to their participations and interactions. To assess the performance, we modeled the relationships between countries on a directed graph and applied centrality analysis which is a common approach to evaluate the relative importance of a node within the network. Each action is coordinated by a management committee which is composed of a chair and at most two delegates per participant country. In the graph, each participant country is denoted with a node. We classify the countries according to their roles in the action. Since the proposer of the action becomes the chair, usually, we regard the country of the chair as a gateway to access the action. Therefore, to signify the relationship between two countries, a directed edge is added from the participant country to the respective country of the chair. Note that several projects were considered over a span of 6 years and multiple interactions between two nodes are indicated as the edge weight.**

***Keywords* - Centrality measures, betweenness centrality, closeness centrality, social network analysis, graph theory.**

I. Introduction

THE term social network denotes social relationships among actors. An actor may refer to a vast array of entities including organizations [1], diseases [2], sensor nodes [3], etc. Relationships imply pairwise connections between two actors. A connection can either express bidirectional link such as friendship between two individuals or a railroad between two cities, or it can also represent unidirectional interactions as well such as an event originating in one place and ending in another one. Transmissions of a pathogen propagating to a new geographical location or email communication among suspects of terrorist activities are some of the examples of directional interactions.

Social network analysis provides a means to study interactions and/or relationships within a social group. Statistical techniques [4] or graph-theoretical approaches can be employed to infer inherent network dynamics to analyze various features of actors and relationships within the network. In this paper, we use graph structure to model the network and exploit graph theory, namely centrality measures. Centrality reveals node level features such as relative prominence of a node within the network. Various approaches exist to assess centrality including degree, eigenvalue, betweenness, closeness, etc. based on the definition of the importance. Accessibility and expected force (influence) are some of the indicators to assess importance of a node [5]. While accessibility considers random walks in the network and measures probability of visiting a certain actor, expected force evaluates the likelihood of the spread of an infection in case of an outbreak of infectious diseases [6].

Walk can be defined as a series of actors visited following the connections among them. According to the considered application, various requirements can be defined on the walk. Visiting an actor multiple times can be restricted to model the application in a more realistic manner. Let us consider an application where a widespread foodborne outbreak is modeled as given in [2]. In this model, each actor represents a country where the obtained pathogens are associated and relationships denote unidirectional links connecting two countries where the isolates crosses towards the given direction. In this model, visiting an actor during the walk more than once can be limited since the corresponding country will be already infected during the first visit. Therefore, one can claim that eigenvalue centrality is not the suitable approach to measure centrality for this application and employ betweenness centrality instead.

Centrality approaches diversify based on various properties of the walk including the length and consideration of the start/end point of the walk [7]. For instance, while degree centrality considers a walk of length one (i.e., the number of immediate connections), eigenvalue centrality requires multiple iterations of walk until reaching convergence. On the other hand, unlike degree and eigenvalue centralities which count walks start/end on a certain actor, betweenness centrality considers the number of walks which passes from the given actor. Another classification is based on whether the volume or the length of the walk is more important. Unlike degree, eigenvalue, and betweenness centralities, closeness centrality considers the length of a walk.

In graph theory, node is the fundamental unit of the graph along with the edges. Nodes denote actors and edges represent relationships. As mentioned earlier, edges can be directed or undirected (bidirectional). Also, weights can be assigned to edges to denote the importance of the link, frequency of an event, or the cost of the transmission, etc. based on the considered application. When the network is modeled on the graph, one of the centrality approaches can be employed to assess relative importance of the nodes.

This paper aims to identify prominent countries in science across Europe by evaluating the research performance of the European countries. Despite availability of various framework programs for research, we selected COST framework program [7] to assess research performance of the countries considering the fact that COST is the longest running European framework supporting transnational cooperation across Europe [8]. COST program does not support the research itself but the networking activities, staff exchange, and training. Therefore, we believe the suitability of this research networking program to analyze research performance by employing social network analysis. We model the research network on a graph structure and exploit graph theory to identify prominent countries in research and their influence on other countries in this network. To understand the model, let us detail the participation procedure to COST actions. Each project funded by COST is referred to as an action. Each action has a chair and a management committee (MC). Action chair is usually the grant proposer and the MC should be represented by at least five countries. Each country can nominate up to two participant researchers. While each country follow a different approach to nominate the participants, TUBITAK, which is the representing institution of Turkey, requires the researcher to have an active project, funded by TUBITAK, related to the COST action [9]. There are 36 member countries of COST and one cooperating state. Considering the fact that a researcher (or a country) cannot participate to an action if there is no funded action, we classify the countries into two groups: providers and beneficiaries. Providers enable an action so that beneficiary countries may join. Corresponding country of the action chair is regarded as provider and the associated countries of the MC members are regarded as beneficiaries.

On the graph, each country is represented as a node and a directed edge is added originating from the beneficiary directed towards the provider. Provider country can be regarded as a gateway to participate COST and access the fund. If a county is represented in an action, there can be one or two members. Since we focus on whether a country is represented or not in an action, even if two participants exist from a country, edge weight is set to one on the graph for a single action. But since there are several actions funded each year, we count the links between countries and set the edge weight accordingly for the corresponding country pairs considering the edge directions. Once the graph is completed, we employ various centrality approaches to analyze the network and assess the research performance of the countries.

The rest of the paper is organized as follows. Previous work is discussed in Section II. Data collection and employed approaches are detailed in Section III. Findings are discussed in Section IV. The paper is concluded in Section V.

II. Related Work

In graph theory, centrality measures can be employed to determine how central a node is for the network, or in other words, to evaluate the importance of a vertex in a graph. However, the term, importance, can be vague which requires elaboration on the network type and application. In this paper, we are concerned with the interaction of countries while collaborating in the research activities. Since the action chair manages the budget, action's fund flows from the country represented by the action chair to other countries in the MC. Thus, in our case, importance infers centrality in accessing grants and relaying the grant money between the nodes.

A variety of works exist in the literature which employs centrality approaches to address different problems in various applications. In [10], trust issue in web-based platforms is considered. To evaluate reputation of the users, two different scores are calculated based on the contribution and centrality. Depending on the platform, contribution may refer to the number of reviews in an e-market platform or answers in a forum. But despite its importance on the success of the platform, calculating contribution can be difficult due to the sparseness of the data. On the other hand, centrality-based score can be calculated based on the who-trusts-whom network. The idea is trying to estimate contribution based score based on the centrality-based reputation. Another study in [11] aims to identify documents that are likely to have higher impact in the future by employing centrality measures. In this work, importance of a document is defined based on the number of citations it receives. [12] considers criminal networks and tries to identify key actors in a drug trafficking network by integrating degree and betweenness centrality measures. Another work presented in [13] evaluates financial institutions and analyzes their roles in financial crisis. It was shown that centrality measures perform well in identifying and monitoring systematically important financial institutions in case of a crash in the banking system. Determining neuronal activity is another application where centrality measures are used [14]. Activities between neurons in the neural network are investigated and importance of the nodes is identified. In the mentioned work, firing rate of the neuron is used to assess the importance. In another study, betweenness centrality is employed to understand the flow characteristics of the traffic within a city [15]. Closeness Centrality is exploited in [16] In to analyze the complaints. In [17], social network analysis was employed to analyze the image of various brands.

III. Method

A. Data Collection

The data that we have considered in this study was collected from the official website of COST [18] using Python 3. We have employed Beautiful Soup [19] for web scraping. Beautiful Soup is a python package to parse HTML and XML documents. Collected data was analyzed using Gephi [20]. Gephi was also employed to visualize the results. In this study we have considered the actions funded between 2012 and 2017. For each year, we counted the number of links between the pairs of countries considering the directionality of the link. To observe the progress, we classified the data into two groups. In the first group, we report the cumulative sum of the performance results for the years 2012-2014. In the second group, the results for the years 2015-2017 are reported.

B. Closeness Centrality

Closeness centrality of a node, u in graph G, is defined as the reciprocal of the farness [21] as given in Equation 1. $d(u,v)$ denotes the distance (farness) between nodes u and v. Farness can be defined as the total length of the shortest paths between node u and the rest of the nodes in the graph. If a node is located closer to other nodes, its reachability is expected to increase.

$$C(u) = \frac{1}{\sum_{v} d(u,v)} \tag{1}$$

C. Betweenness Centrality

Betweenness centrality considers the number of times a node appears on a shortest path between every pair of other nodes. If the number of shortest paths between a pair of nodes st, is $d(st)$ and node v exists $d(st)_v$ times in these shortest paths, then the betweenness centrality of node v can be expressed as in Equation 2.

$$C_B(v) = \sum_{s,t,v \in V} \frac{d(st)_v}{d(st)} \tag{2}$$

A node is associated with a higher centrality score, if the fraction of the shortest paths passing through the given node is higher.

D. PageRank

PageRank, a variant of the eigenvector centrality, measures centrality of a node based on the number of incoming edges. The idea is similar to a random walk where the probability of ending up on a specific node after a random walk is higher if the PageRank score for that node is higher. PageRank provides a probability distribution for the importance of the nodes and can be computed iteratively where the probability values always sum to one at each iteration. PageRank of node v can be expressed as in Equation 3. d refers to damping factor which is set to 0.85 as recommended [22]. $M(v)$ is the set of incoming edges that link to v. $L(s)$ is the number of outgoing links from node p_S, and N is the total number of nodes.

$$PR(v) = \frac{1-d}{N} + \sum_{s \in M(v)} \frac{PR(s)}{L(s)} \tag{3}$$

IV. Findings

In the rest of the paper, various node/edge colors and node/edge sizes are used in the graphs to improve the visualization in representing the results. Node size and color signify the prominence of the countries. Larger node size and darker color denotes increased importance. On the other hand, thickness of the edges represents frequency of interactions between the respective countries. Higher frequency is denoted with the thicker lines accompanied with a darker line color. To evaluate the performance of the countries, PageRank, betweenness centrality, and closeness centrality measures are used. Obtained results are normalized between 0 and 1 to adjust the values for a fair comparison. To analyze the correlation of our findings with the actual research performance of the countries, we have considered the COST country fact sheets report [23]. In the mentioned report, various facts are given for each member state including the amount of budget transferred to each country for respective years. In the report, the latest available data is from 2015 and therefore we have considered the amount of funds transferred between 2012-2014 and the centrality scores obtained for 2012-2014. The results of the analysis for the correlation between the transferred funds and the centrality scores for the period of 2012-2014 can be found in Table 1. It can be noticed that the betweenness centrality score is highly correlated with the research performance of the countries in terms of the grant money they received. PageRank is the next approach providing the best metric to assess the research performance. Closeness centrality is the least correlated approach in determining the research performance.

Closeness	Betweenness	PageRank
0.5109	0.8676	0.7493

Table 1. Correlation between the amount of transferred funds and the centrality scores for the period of 2012-2014.

A. Results for 2012-2014

The results can be found in Figures 1, 2, and 3 for PageRank, betweenness centrality, and closeness centrality respectively.

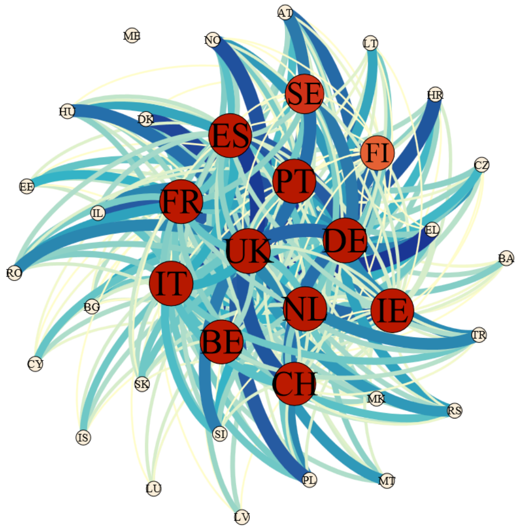

Figure 1. PageRank results for 2012-2014.

According to Figure 1, 11 countries are represented with nodes of almost the same size. Finland is denoted with a slightly smaller and lighter node. The rest of the countries have negligible importance.

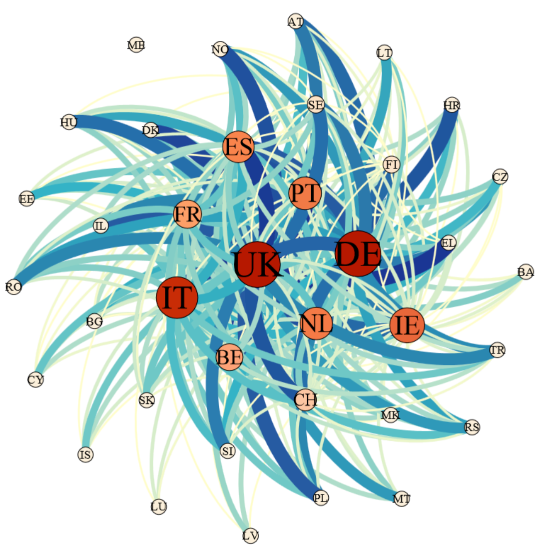

Figure 2. Betweenness centrality results for 2012-2014.

Compared to PageRank, betweenness centrality diversifies the performance of the countries better according to Figure 2. United Kingdom, Italy, and Germany have the highest scientific research performance within the network. Italy is slightly worse than United Kingdom and Germany.

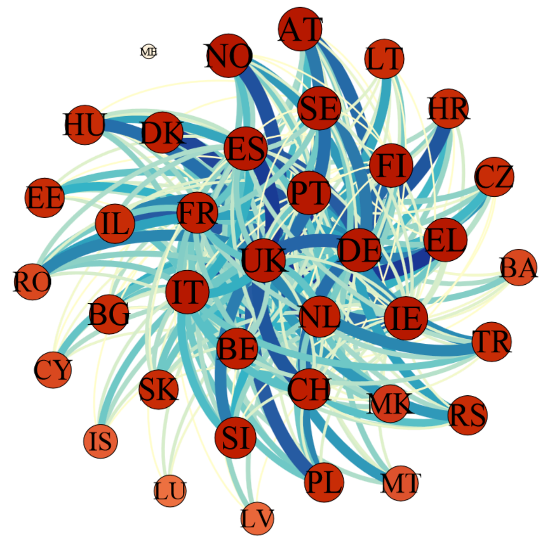

Figure 3. Closeness centrality results for 2012-2014.

According to Figure 3, almost all the countries have similar performance and the difference is negligible. As can be seen from Figures 1-3, three different results are obtained for the same data. The results suggest that betweenness centrality is the best approach to highlight the least number of prominent countries. While presenting the importance of the countries, PageRank emphasizes more countries compared to betweenness. Closeness centrality, on the other hand, provides scores closer to each other which makes it difficult to highlight the countries with the best performance.

B. Results for 2015-2017

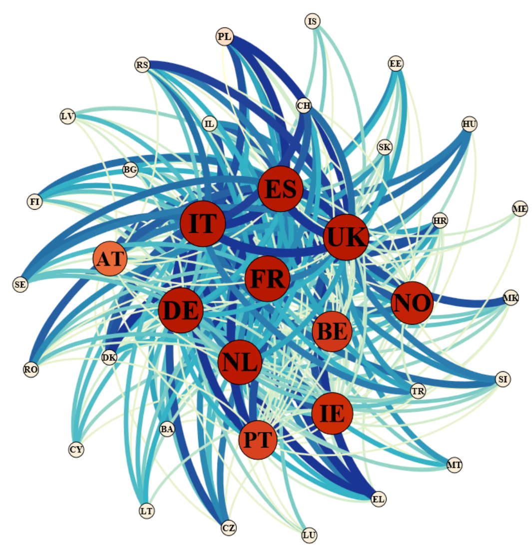

Figure 4. PageRank results for 2015-2017.

Again, 10 countries have similar performance. Austria is slightly worse than these countries. On the other hand, the rest of the countries have negligible importance according to PageRank algorithm as given in Figure 4.

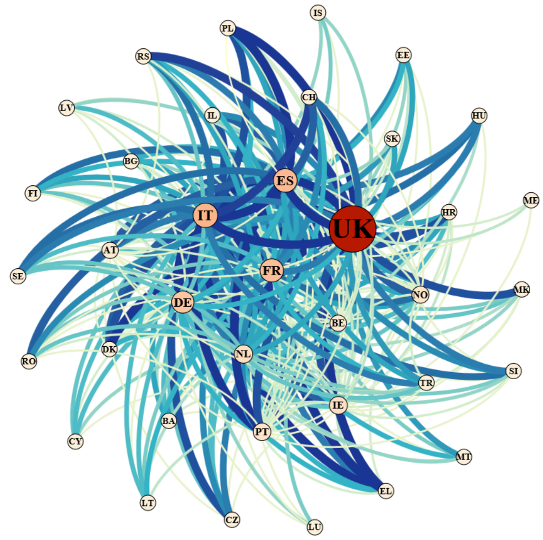

Figure 5. Betweenness centrality results for 2015-2017.

Figure 5 suggests that United Kingdom dominates the network in terms of the scientific performance. It can be claimed that this result avoids the contribution of other countries and may not represent the actual performance of the rest of the countries. But, to highlight only a few countries, betweenness centrality should be applied.

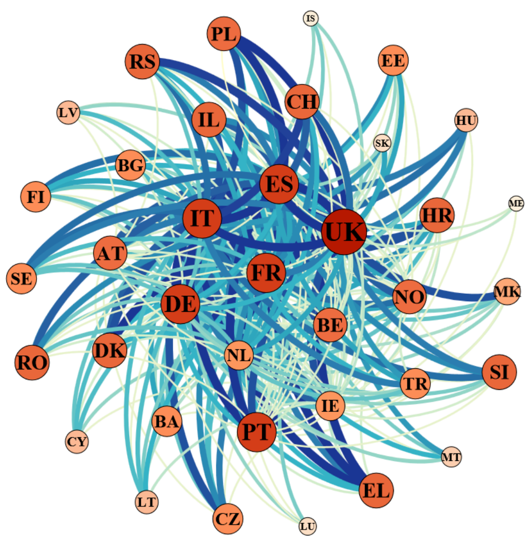

Figure 6. Closeness centrality results for 2015-2017.

According to Figure 6, again, closeness centrality emphasizes the most countries with similar scores.

V. Conclusion

In this study, we have analyzed the research performance of European countries between 2012-2017 through social network analysis. We have employed three different centrality approaches namely, PageRank, betweenness centrality, and closeness centrality to assess relative importance of each country in the COST transnational research network. To analyze the progress, we have considered 2012-2014 and 2015-2017 periods separately. The results are demonstrated on graphs visualized with various node/edge colors and node/edge sizes to represent the significance of respective countries and denote the frequency of interactions between countries. To compare our findings with the actual research performance of the countries, we have inspected the correlation of the obtained results with the amount of funds received by each country in the respective years. The results suggest that betweenness centrality is highly correlated with the actual research performance of the countries in terms of the amount of transferred funds. Centrality analysis show that betweenness centrality should be employed to limit the highlighted prominent countries While PageRank emphasizes more countries compared to betweenness, closeness centrality provides scores closer to each other and makes the results less distinctive.

Appendix

Country codes for 36 member states and 1 cooperating state in the COST framework can be found below.

Country Code	State
AT	AUSTRIA
BE	BELGIUM
BA	BOSNIA AND HER.
BG	BULGARIA
HR	CROTIA
CY	CYPRUS
CZ	CZECH REPUBLIC
DK	DENMARK
EE	ESTONIA
FI	FINLAND
FR	FRANCE
MK	MACEDONIA
DE	GERMANY
EL	GREECE
HU	HUNGARY
IS	ICELAND
IE	IRELAND
IL	ISRAEL
IT	ITALY
LV	LATVIA
LT	LITHUANIAN
LU	LUXEMBOURG
MT	MALTA
ME	MONTENEGRO
NL	NETHERLANDS
NO	NORWAY
PL	POLAND
PT	PORTUGAL
RO	ROMANIA
RS	RUSSIA

SK	SLOVAKIA
SI	SLOVENIA
ES	SPAIN
SE	SWEDEN
CH	SWITZERLAND
TR	TURKEY
UK	UNITED KINGDOM

Table 2. Country codes table.

REFERENCES

[1] C. Rob, S. P. Borgatti, and A. Parker. "Making invisible work visible: Using social network analysis to support strategic collaboration," *California management review*, 44, no. 2, pp. 25-46, 2002.

[2] M. Hoffmann, Y. Luo, S. R. Monday, N. Gonzalez-Escalona, A. R. Ottesen, T. Muruvanda, C. Wang, G. Kastanis, C. Keys, D. Janies, I. F. Senturk, U. V. Catalyurek, H. Wang, T. S. Hammack, W. J. Wolfgang, D. Schoonmaker-Bopp, A. Chu, R. Myers, J. Haendiges, P. S. Evans, J. Meng, E. A. Strain, M. W. Allard, and E. W. Brown, "Tracing Origins of the Salmonella Bareilly Strain Causing a Food-borne Outbreak in the United States," *The Journal of Infectious Diseases*, vol. 213, issue 4, pp. 502–508, 15 February 2016.

[3] I. F. Senturk and K. Akkaya, "Connectivity restoration in disjoint Wireless Sensor Networks using centrality measures," 39th Annual *IEEE Conference on Local Computer Networks Workshops*, Edmonton, AB, 2014, pp. 616-622.

[4] M. S. Handcock, "Statistical models for social networks: Inference and degeneracy," *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, 2003.

[5] Node influence metric, https://en.wikipedia.org/wiki/Node_influence_metric, Accessed: 03/29/2018.

[6] D. A. Janies, L. W. Pomeroy, C. Krueger, Y. Zhang, I. F. Senturk, K. Kaya, and Ü. V. Çatalyürek. "Phylogenetic visualization of the spread of H7 influenza A viruses," *Cladistics*, vol. 31, no. 6, pp. 679-691, 2015.

[7] Centrality, https://en.wikipedia.org/wiki/Centrality, Accessed: 03/29/2018.

[8] COST, http://www.cost.eu/about_cost, Accessed: 03/29/2018.

[9] COST participation rules of Turkey, https://www.tubitak.gov.tr/tr/kurumsal/uluslararasi/cok-tarafli-programlar/cost/icerik-aksiyonlara-katilim, Accessed: 03/29/2018.

[10] P.D. Meo, K. Musial-Gabrys, D. Rosaci, G.M. Sarne, and L. Aroyo, "Using centrality measures to predict helpfulness-based reputation in trust networks," *ACM Transactions on Internet Technology (TOIT)*, 17(1), p.8, 2017.

[11] P. Klimek, A. S. Jovanovic, R. Egloff, and R. Schneider, "Successful fish go with the flow: citation impact prediction based on centrality measures for term–document networks," *Scientometrics*, 107(3), pp.1265-1282, 2016.

[12] D. A. Bright, C. Greenhill, M. Reynolds, A. Ritter, and C. Morselli, "The use of actor-level attributes and centrality measures to identify key actors: a case study of an Australian drug trafficking network," *Journal of Contemporary Criminal Justice*, 31(3), pp.262-278, 2015.

[13] T. U. Kuzubaş, I. Ömercikoğlu, and B. Saltoğlu, "Network centrality measures and systemic risk: An application to the Turkish financial crisis," *Physica A: Statistical Mechanics and its Applications*, 405, pp. 203-215, 2014.

[14] J. M. Fletcher and T. Wennekers, "From structure to activity: Using centrality measures to predict neuronal activity," *International journal of neural systems*, 28(02), p.1750013, 2018.

[15] S. Gao, Y. Wang, Y. Gao, Y. Liu, "Understanding Urban Traffic-Flow Characteristics: A Rethinking of Betweenness Centrality," *Environment and Planning B: Urban Analytics and City Science*, vol. 40, issue 1, pp. 135-153, 2013.

[16] M. T. Argan, "A descriptive study of e-complaining: social network analysis on a website as complaint forum," *Journal of Internet Applications and Management*, vol. 5(1), 2014.

[17] Ç. Basfirinci, "Analysing Brand Image Through Social Network Analysis: A Research on Turkcell and Vodafone Brands," *İGÜSBD,* vol. 3, issue 2, October 2016.

[18] All COST actions, http://www.cost.eu/COST_Actions/all_actions, Accessed: 03/29/2018.

[19] Beautiful Soup, https://www.crummy.com/software/BeautifulSoup, Accessed: 03/29/2018.

[20] Gephi, https://gephi.org/, Accessed: 03/29/2018.

[21] A. Bavelas, "Communication patterns in task-oriented groups," *The Journal of the Acoustical Society of America*, vol. 22, no. 6, pp. 725-730, 1950.

[22] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1-7, pp. 107-117, 1998.

[23] COST country fact sheets, http://www.cost.eu/module/download/58026, Accessed: 03/29/2018.