Collaboration Networks Analysis: Combining Structural and Keyword-Based Approaches

Ana Meštrović(✉)

Department of Informatics, University of Rijeka,
Radmile Matejčić 2, 51000 Rijeka, Croatia
amestrovic@inf.uniri.hr

Abstract. This paper proposes a method for the analysis of the characteristics of collaboration networks. The method uses social network analysis metrics which are especially applicable to directed and weighted collaboration networks. By using the proposed method it is possible to investigate the global structure of the collaboration networks, such as *density*, *centralisation*, *assortativity* and the dynamics of network growth. Furthermore, the method proposes appropriate network centrality measures (*degree* and its variations for directed and weighted networks) for ranking the nodes. In addition the proposed method combines a keyword-based approach and Louvain algorithm for the community detection task. Next, the paper describes a case study in which the proposed method is applied to the collaboration networks emerged from STSMs on the KEYSTONE COST Action.

Keywords: Social networks analysis · Collaboration networks
Keyword-based community detection

1 Introduction

Collaboration networks are a special case of social networks in which nodes represent actors/individuals who collaborate in certain projects, jobs or scientific publications. The collaboration environment can be for example, any organisation or institutions, academic communities, or international project. Collaboration networks can represent collaborations on different levels: between individuals, between institutions or between countries as project participants. These networks can be (but not necessarily) weighted and/or directed. Weights (if they exist) may denote the number of interactions, projects or publications. Directions (if they exist) may denote direction of communication, institutional exchange, knowledge sharing, etc. In this paper the focus is on the directed and weighted collaboration networks. Social networks analysis (SNA) in general offers a wide range of metrics for data analysis on the global, middle and local network level. However not all the measures are appropriate in all networks and cases. A variety of these measures has been applied to various researches of collaboration networks.

J. Szymański and Y. Velegrakis (Eds.): IKC 2017, LNCS 10546, pp. 111–122, 2018.
https://doi.org/10.1007/978-3-319-74497-1_11

The analysis of the collaboration networks provides an insight into the quality of the relations among actors in the network. It may identify crucial actors in the network and closely related communities. This is all of great importance for studying knowledge sharing among actors and for proposing future steps and actions.

Collaboration networks have received much attention in the research for at least the past forty years. The first attempts of collaboration networks analysis were focused on the networks based on scientific publications. The simple reason is that this data is freely available to the public. Newman [17] was one of the first who intensely studied the structure of scientific collaboration networks in terms of SNA. Following this approach many authors have combined various network measures aiming to analyse the structure of the collaboration networks constructed from scientific publications [1,2,7,9,14,17]. There were less attempts to analyse scientific project collaboration networks such as FP7 or ERASMUS, as for example in [3,19,20]. Furthermore, collaboration networks are of much interest in the organisations. For example in [8,21] the authors explored how the structure of the collaboration network can influence the innovation.

Still, there is no standardised/universal methodology nor framework proposed for the project collaboration networks analysis. Each study proposes its own set of measures and approaches. The goal of this study is to propose a method that provides an integral approach that unifies the all important measures on the global, middle and local level. It proposes a network analysis in four steps, one step for each aspect. Firstly, it analyses network on the global level. Secondly, network centrality measures are used for ranking the nodes and revealing the crucial nodes in the network. Thirdly, the network is analysed on the middle level in terms of community detection. And lastly, the network dynamics is analysed based on various network stages during the network growth.

The rest of the paper is organised as follows. The following section describes the networks, network measures and methodology of the proposed approach. The third section presents a case study based on the STSMs networks analysis. The last section provides concluding remarks and plans for future work.

2 Research Methodology

2.1 Network Types

A network or graph $G = (V, E)$ is a pair of a set of nodes V and a set of edges E, where N is the number of nodes and K is the number of edges. A network is directed if the edges have a direction associated with them. A network is weighted if there is a weight function w that assigns value (real number) to each edge. In weighted networks, S denotes the sum of all weights in the network; that actually refers to all realised relations.

2.2 Network Measures

We now review some of the standard network measures [18]. Most of them can be applied to the directed and weighted networks, but some of them are suitable only for undirected and/or unweighted networks and in that case of analysed collaboration networks the directions and/or weights are omitted.

The *degree* of a node i, k_i is the number of edges incident to the node. In a directed network, *in/out-degree* of a node is the number of incoming and outgoing links. *Weighted degree* is called *strength.* The *strength* of the node i is the sum of the weights of all the links incident with the node i:

$$s_i = \sum_j w_{ij}. \tag{1}$$

In the directed network, the i*n/out strength* s of the node i is defined as the number of its incoming and outgoing links, that is:

$$s_i^{in/out} = \sum_j w_{ji/ij}. \tag{2}$$

Average degree, $\langle k \rangle$ is the sum of the degree over all nodes divided by the number of nodes. Analogously, the *average strength*, $\langle s \rangle$ is the sum of the strength over all nodes divided by the number of nodes.

Network *centralisation*, defined in [6], reflects the extent to which interactions are concentrated in a small number of individuals rather than distributed equally among all members.

$$cent = \frac{\sum_i (k_{i^*} - k_i)}{max \sum_v (k_{v^*} - k_v)}, \tag{3}$$

where i^* is the node with largest degree in a network and $max \sum_v (k_{v^*} - k_v)$ refers to network of the same size with the maximal possible centralisation which is a star network. The value of the *cent* lies between 0 and 1 and obviously, values close to 1 denote highly centralised networks. In the collaboration networks, high value of *centralisation* shows that network relations are organized around one group of actors. In general, *centralisation* may refer to the power and control structure of the network.

Network *density* represents a fraction of existing connections and the number of all possible connections. In directed networks it is calculated as:

$$d = \frac{K}{N(N-1)}. \tag{4}$$

The network *connected component* is a subgraph in which any two nodes are connected to each other by paths. The *number of connected components* is denoted by ω. When a network has a property that it has no disconnected parts, we say it is connected; otherwise it is disconnected. Each piece is usually called a component (or connected component). The largest connected component is called the giant connected component (GCC). In directed networks the components can be strongly or weakly connected. Weakly connected components refer to the same components as if the network were undirected. Furthermore, it is possible to measure how well the network is connected in a sense that we measure the percentage of network that belongs to a giant component. Here this measure is defined as *connectedness* and defined as follows:

$$conn = \frac{N_{GCC}}{N}, \tag{5}$$

where N_{GCC} is the number of nodes in the giant component.

The network is said to show assortative mixing by degree if nodes tend to be connected to other nodes with a similar degree. If the opposite is true then we say that network shows disassortative mixing by degree. The degree of *assortativity* is defined as follows:

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2}. \tag{6}$$

In general, values of *assortativity* lie between 1 and 1. When r close to 1, the network is said to be assortative when r is close to 0 the network is non-assortative, while when r is close to 1 the network is disassortative.

The *clustering coefficient* of a node measures the density of edges among the immediate neighbours of a node. For weighted networks the *clustering coefficient* of a node i is denoted by c_i and defined as the geometric average of the subgraph edges weights:

$$c_i = \frac{1}{k_i(k_i - 1)} \sum_{j,k} (\hat{w}_{ij}\hat{w}_{ik}\hat{w}_{jk})^{1/3}, \tag{7}$$

where k_i is the degree of the node i, and the edges weights $\hat{w}_{ij}$ are normalized by the maximum weight in the network $\hat{w}_{ij} = w_{ij}/\max(w)$. If $k_i < 2$, then the value of c_i is 0.

The *average clustering* of a network, C, is defined as the average value of the *clustering coefficients* of all nodes in an undirected network:

$$C = \frac{1}{N} \sum_i c_i. \tag{8}$$

A *path* in a network is a sequence of edges, which connect a sequence of nodes that are all distinct from one another. The *shortest path* between two nodes i and j is a path with the shortest length and it is called the distance between i and j and is denoted as d_{ij}. The *average path length*, L, of a directed network is given by the equation:

$$L = \sum_{i,j} \frac{d_{ij}}{N(N-1)}. \tag{9}$$

Note that the *average path length* can be calculated only for a connected network. More precisely, if $\omega > 1$, L is computed for the GCC.

The *eccentricity* of the node i is the maximum distance from i to all other nodes in the network. The *diameter* D is the maximum eccentricity.

Network *modularity* m measures the quality of the network partition in the communities. The *modularity* of a network partition is a scalar value between -0.5 and 1 that measures the density of links inside communities as compared to links between communities. Communities are groups of densely interconnected nodes within a network. In other words, nodes in a community have a greater amount of connections amongst each other than with other nodes in the network. There are many algorithms proposed for this task. We chose the most commonly used and implemented Louvain algorithm [11], a greedy optimization method that optimizes the modularity of a network's partitions.

2.3 Method for Analysing Collaboration Networks

This section proposes a method for the structural investigation of the directed and weighted collaboration networks. Besides network structure, the additional step of the whole method is to analyse the semantic aspect of the network communication. The whole approach can be performed in four main steps as follows.

The first step is to analyse the global network measures that are adequate for directed and weighted networks and which have sense in the given context. The measures are described in the previous section and the final list of measures in focus is as follows: *average (weighted) degree*, *centralisation*, *density*, *number of (weakly) connected components*, *percentage of connectedness*, *assortativity*, *average clustering coefficient*, *distance measures*, and *modularity*.

In the second step, the goal is to find centrally positioned nodes in the network. Network *degree* and its directed and weighted variations are the most appropriate centrality measures in this case. There are many other centrality measures. However, *degree* and *strength* shows how important the node is according to the established relations. *Degree* shows the number of different relations and strength shows the number of overall relations. Additionally, the *in-* and *out- degree* and *strength* take into account whether the direction is in-going or out-going.

In the third step the focus is on the middle level. The community detection is based on the Louvain algorithm. The result is a set of communities. Furthermore, it is possible to identify additional communities according to the semantic background of the relationships if it exists. The most straightforward way to do that is to associate a list of keywords/topics to each node. The general idea is to group nodes according to the topics they share. Two nodes are in the same community if two actors have more than t common keywords. Therefore, the results of the Louvain algorithm are extended with communities identified based on keywords.

The fourth step includes the structural analysis of the network dynamics. This step refers to structural analysis of all the previously mentioned measures and their changes over time. This is how it is possible to monitor new trends in the network growth and predict future actions.

3 Case Study: STSM Networks

This section presents a case study in which proposed method is applied to the STSM collaboration networks. The initial phase is network construction and after that the integral results of the analysis are presented.

Network construction and analysis are performed in Python packages NetworkX [22] and LaNCoA [12]. Visualisations were prepared using the tool Gephi [4].

3.1 KEYSTONE-STSM Dataset and Networks Construction

COST is the European framework supporting trans-national cooperation among researchers, engineers and scholars. Short-Term Scientific Missions (STSMs) are one of the COST networking tools that allow participants to visit an institution or laboratory in another country in order to foster collaboration, share new techniques and infrastructure [5].

For the purpose of this experiment, a collaboration network is constructed, based on the data collected from the STSMs on the KEYSTONE (semantic KEYword-based Search on sTructured data sOurcEs) COST Action [10]. There were 50 STSM grants in total: 11 in the first grant period (year 2014), 14 in the second grant period (year 2015), 8 in the third grant period (year 2016) and 17 in the fourth grant period (years 2016 and 2017).

A general $STSM_{gp4}$ network is constructed from all data collected during the four grant periods. In this network, countries are nodes and a directed link between two nodes A and B exists if there was an STSM application from country A to country B. The weight denotes how many STSM applications were realised between these two countries. Furthermore, three additional networks $STSM_{gp1}, STSM_{gp2}, STSM_{gp3}$ for the first three grant periods were constructed on the same principle. The final set of four networks serves as an examination of the network growth and dynamics.

3.2 Results

Network Structure on the Global Level. The results of the first step - the global network measures of $STSM_{gp4}$ are given in the first row of Table 1. According to the results, the value of *average strength* is close to the value of *average degree*. The same consideration holds for S and K. Consequently, the impact of weights in the network is not so significant. This can be explained by the fact that STSM agreements between two countries are rarely repeated and the links are rather uniformly distributed. The network *density* is low, however this is the usual property of such networks. The value of *centralisation* measure is around 0.3. This property is indicator that collaborations of the KEYSTONE participants is not centralised. Furthermore, the network has only two components that do not belong to the giant component, more precisely 86% of nodes belong to the GCC which indicates that network is well connected. This network is not assortative, nor disassortative which means that nodes have no property to tend to connect with mostly similar nodes (with high degree), neither opposite. The *clustering coefficient* is relatively high and *distance measures* are relatively low which may indicate that this is a small-world network [23]. The *modularity* is equal 0.429 which is on the lower limit to say that the nodes tend to group into separate communities. Actually, the whole network is overall well connected

Table 1. Global network measures for networks based on the STSM agreements, *number of nodes* (N), *number of links* (K), *network strength* (S), *average degree* ($\langle k \rangle$), *average strength* ($\langle s \rangle$), *centralisation* ($cent$), *density* ($dens$), *number of components* (ω), *connectedness* ($conn$), *assortativity* (r), *average clustering* (C), *average path length* L, *diameter* D, *modularity* m

Network	N	K	S	$\langle k \rangle$	$\langle s \rangle$	$cent$	$dens$	ω	$conn$	r	C	L	D	m
$STSM_{gp4}$	28	43	50	1.536	1.786	0.32	0.057	3	86%	−0.27	0.102	3.054	3	0.429
$STSM_{gp3}$	23	30	33	1.304	1.435	0.368	0.059	4	70%	−0.16	0.078	2.405	5	0.385
$STSM_{gp2}$	19	23	25	1.211	1.316	0.408	0.067	3	79%	−0.350	0.042	2.57	5	0.348
$STSM_{gp1}$	11	10	11	0.909	1	0.144	0.091	1	100%	−0.102	0	2.172	4	0.475

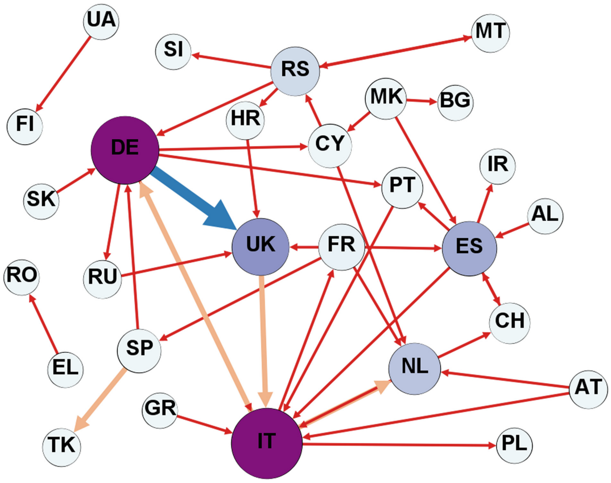

Fig. 1. $STSM_{gp4}$ directed and weighted collaboration network. Node labels are given as country acronyms based on the international 2-letters coding, ISO Alpha-2. The size of node is proportional to the node strength. The thickness of the link is proportional to the link weight

and there is no hierarchy. Obviously, there is no intentions of participants to separate in the closed communities. The overall collaboration network, $STSM_{gp4}$ is illustrated in Fig. 1.

Network Structure on the Local Level. This section presents the results of applying centrality measures to the node ranking task. The node degree and strength for undirected and directed links are examined, and therefore six variations of nodes ranking are presented. In Table 2 the first 12 of 28 nodes in total are ranked. All other nodes have centrality values 0, 1 or 2 and are not included in the table due to the limited space. In Fig. 1 node sizes are proportional to their strength. According to the results node IT has the highest centrality value except in the case of the *out degree* and *out strength* in which the node DE has the highest values. Highly ranked nodes are almost the same for all six chosen measures.

Table 2. Ranked nodes in the $STSM_{gp4}$ network

	Degree						Strength					
	Node	k_i	Node	k_i^{in}	Node	k_i^{out}	Node	s_i	Node	s_i^{in}	Node	s_i^{out}
1	IT	11	IT	7	DE	5	IT	14	IT	8	DE	8
2	DE	9	DE	4	FR	4	DE	13	UK	7	IT	6
3	ES	8	NL	4	RS	4	UK	9	DE	5	FR	4
4	NL	6	ES	4	IT	4	ES	8	NL	5	RS	4
5	RS	6	UK	4	ES	4	NL	7	ES	4	ES	4
6	FR	5	PT	2	MK	3	RS	6	PT	2	SP	3
7	UK	5	RS	2	NL	2	FR	5	RS	2	MK	3
8	CY	4	CH	2	CY	2	CY	4	TK	2	NL	2
9	PT	3	CY	2	AT	2	SP	4	CH	2	CY	2
10	CH	3	FR	1	SP	2	PT	3	CY	2	AT	2
11	SP	3	BG	1	HR	1	CH	3	FR	1	UK	2
12	MK	3	HR	1	RU	1	MK	3	BG	1	HR	1

Communities Identification. According to the Louvain algorithm there are six communities in the STSM network. Approximately 40% of the network is in the largest community which contains 11 nodes. All communities are shown in Table 3. The visualisation of all communities in the STSM network is illustrated in Fig. 2.

Table 3. Six communities for $STSM_{gp4}$ network identified using the Louvain algorithm

	Nodes	%
1	AL, AT, CH, ES, FR, GR, IT, IR, NL, PL, PT	40%
2	BG, CY, HR, MK, MT, SI, RS	25%
3	DE, RU, SK, UK	14%
4	SP, TK	7%
5	EL, RO	7%
6	FI, UA	7%

An additional insight into the existing communities is possible to be gained analysing the semantic aspect of relations/connections. The Louvain set of communities is extended by taking into account keywords from the STSM titles. The titles are adequately preprocessed: stopwords are removed and lemmatised. Certain words with too general meanings such "approach", "method", etc. are also removed from the titles. Next, the algorithm proposes that two or more

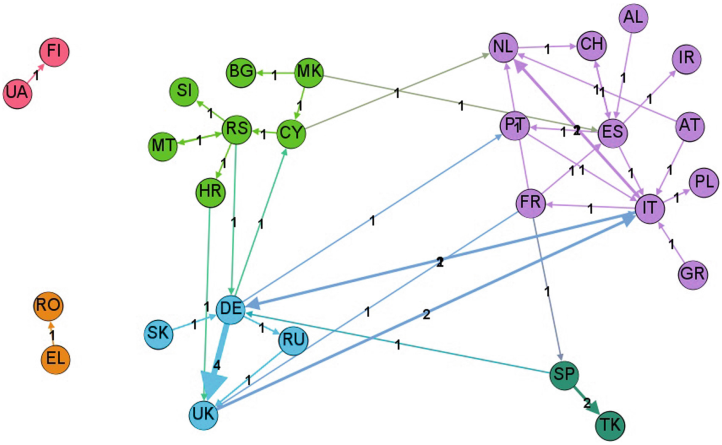

Fig. 2. Communities in network

countries are in the same community if they have more than 2 keywords in common. The final result is a set of possible overlapping communities that can be mapped onto the existing set of communities.

SC_1 = {MK, ES, NL, IT}, with a set of keywords:
KS_1 = {data, keyword-based, search, source}
SC_2 = {RS, DE, ES, IR}, with a set of keywords:
KS_2 = {resource, semantic, sharing}
SC_3 = {IT, PL, RS, SI}, with a set of keywords:
KS_3 = {experimenting, search, structured, techniques}
SC_4 = {MK, BG, UK, IT}, with a set of keywords:
KS_4 = {data, keyword, search}

Network Dynamics. Global network measures for networks constructed for the previous three grant periods are given in the last three rows in Table 1. As expected, as the network grows, the values of some measures tend to increase (k, s, C, L, m). Besides, the network growth exhibits an interesting property for this network - the participants form one connected component from the beginning. In the first grant period there is no disconnected components, while later there are only 2 or 3 components that do not belong to the GCC. This can be explained by the assumption that there has been an excellent connection and communication among participants in KEYSTONE community so they formed one coherent group from the beginning. The network growth is illustrated in Fig. 3.

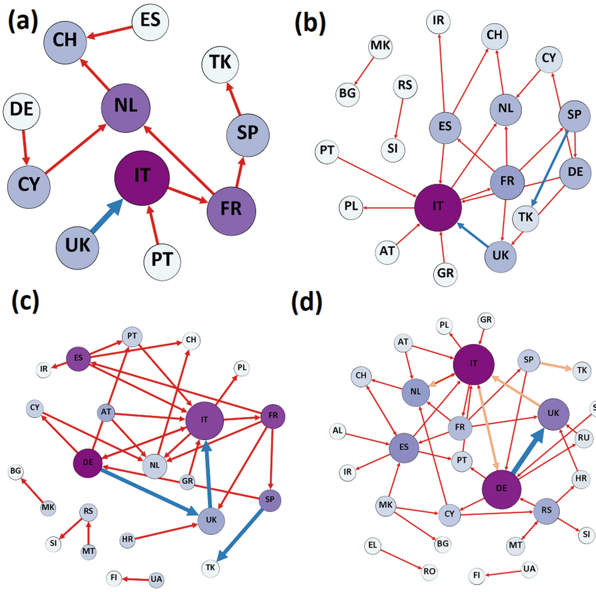

Fig. 3. Network growth in 4 grant periods: $STSM_{gp1}$ (a), $STSM_{gp2}$ (b), $STSM_{gp3}$ (c) and $STSM_{gp4}$ (d)

4 Conclusion and Discussion

This paper presents an integral method for collaboration networks analysis. The approach proposes four steps of the analysis. In the first step it proposes a set of global network measures. The chosen measures have meaningful interpretations in the context of collaboration networks. For example, they can measure how well the network is connected, whether the network is centralised or not, do nodes tend to group into communities or not, etc. In the second step, six centrality measures are chosen on the local network level for the node-ranking task. In the third step, the analysis is moved onto the middle level where communities are focused upon. The chosen algorithm is the Louvain algorithm for community detection. In addition, the proposed approach extends these algorithms with the keyword-based approach. This way, extra communities can be identified by analysing keywords that describe collaboration. In the last step, the network growth is analysed by using the same set of global measures proposed in the first step.

The approach is well suited for the directed and weighted networks with some exceptions when directions and/or weights are omitted aiming to achieve a global overview of the network from different aspects. More precisely, on the node level, the proposed approach takes into account various centrality measures some of which are applicable to the undirected networks (*degree*, *strength*), some are applicable only to unweighted networks (*degree*, *in-degree* and *out-degree*), and only two measures are applicable strictly to directed and weighted networks (*in-strength* and *out-strength*). However, all these node-level measures may be interesting for the node-ranking task. This is because from certain aspects the relations are interesting when we omit directions. Sometimes is not necessary to know the directions of scientific exchanges, only the number of exchanges. In addition, sometimes we can be interested only in exchanges with different actors - in that sense, we are not interested in the *strength*, only in the *degree*.

The network structure analysis provide information about the nature of participants and relations in the network. This knowledge can be used to improve the quality of the communication, relations, etc. As well, insight into the structure of the network can be used to foreseen future trends.

The proposed approach is tested on the STSM collaboration network of the KEYSTON COST Action. The results show that this network is well connected, not centralised. There are six communities that are well related among themselves. The network grows as one giant component from the beginning.

Note that in the presented case study only interactions between countries are considered. This could be a potential limitation because more precise results can be examined on the institutional level and on the individual level. However, in this particular case it has sense to focus only to the country level to get information how various countries collaborate in this COST Action. To get more precise insight it would be good to include a multilayer approach and involve institutional and individual perspective. Moreover, instead of number of STSMs, the duration of STSMs can be assigned as a link weight. All these ideas will be examined in the future research. There are two limitations of this study that will also be resolved in the future. First, there is only one example as a case study. The second drawback is absence of any kind of evaluation. This both limitations will be resolved in the future work by introducing, analysing and comparing more collaboration networks.

The proposed approach has two novelties. The first novelty is integrability in a sense that the proposed method combines network analysis over all the possible levels, while existing studies are mostly focused on one or two levels. The second novelty is that the approach extends the standard community detection algorithm with the keyword-based approach. The presented idea is very simple and this is still only a preliminary suggestion that need to be improved upon in future work. There are few possible directions of improvements. One possible improvement of the keyword-based analysis is to utilise ontologies [15,16]. Another direction is to analyse the collaboration network as multilayer network. Semantic in texts can be represented as suggested in [13] and it can be combined with collaboration layers.

References

1. Abbasi, A., Hossain, L., Uddin, S., Rasmussen, K.J.: Evolutionary dynamics of scientific collaboration networks: multi-levels and cross-time analysis. Scientometrics **89**(2), 687–710 (2011)
2. Abbasi, A., Altmann, J., Hossain, L.: Identifying the effects of co-authorship networks on the performance of scholars: a correlation and regression analysis of performance measures and social network analysis measures. J. Informetrics **5**(4), 594–607 (2011)

3. Balland, P.A.: Proximity and the evolution of collaboration networks: evidence from research and development projects within the global navigation satellite system (GNSS) industry. Reg. Stud. **46**(6), 741–756 (2012)
4. Bastian, M., Heymann, S., Jacomy, M.: Gephi: an open source software for exploring and manipulating networks. ICWSM **8**, 361–362 (2009)
5. Cost glossary. http://www.cost.eu/service/glossary/STSM
6. Freeman, L.C.: Centrality in social networks conceptual clarification. Soc. Netw. **1**(3), 215–239 (1978)
7. Guan, J., Yan, Y., Zhang, J.J.: The impact of collaboration and knowledge networks on citations. J. Informetrics **11**(2), 407–422 (2017)
8. Guan, J., Zhang, J., Yan, Y.: The impact of multilevel networks on innovation. Res. Policy **44**(3), 545–559 (2015)
9. Hou, H., Kretschmer, H., Liu, Z.: The structure of scientific collaboration networks in Scientometrics. Scientometrics **75**(2), 189–202 (2007)
10. KEYSTONE STSMs. http://www.keystone-cost.eu/keystone/outreach/short-term-scientific-missions-stsms/stsms-approved/
11. De Meo, P., Ferrara, E., Fiumara, G., Provetti, A.: Generalized Louvain method for community detection in large networks. In: 11th International Conference on Intelligent Systems Design and Applications, pp. 88–93. IEEE (2011)
12. Margan, D., Meštrović, A.: LaNCoA: a Python toolkit for language networks construction and analysis. In: MIPRO 2015, pp. 1628–1633 (2015)
13. Martinčić-Ipšić, S., Margan, D., Meštrović, A.: Multilayer network of language: a unified framework for structural analysis of linguistic subsystems. Phys. A Stat. Mech. Appl. **457**, 117–128 (2016)
14. Meštrović, A., Grubiša, Z.: Preliminary analysis of co-authorship networks at The University of Rijeka. Zbornik Veleucilista u Rijeci **3**(1), 159–178 (2015)
15. Meštrović, A.: Semantic matching using concept lattice. In: Proceedings of Concept Discovery in Unstructured Data, Katholieke Universiteit Leuven, pp. 49–58 (2012)
16. Meštrović, A., Calì, A.: An ontology-based approach to information retrieval. In: Calì, A., Gorgan, D., Ugarte, M. (eds.) KEYSTONE 2016. LNCS, vol. 10151, pp. 150–156. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-53640-8_13
17. Newman, M.E.: The structure of scientific collaboration networks. Proc. Nat. Acad. Sci. **98**(2), 404–409 (2001)
18. Newman, M.: Networks: An Introduction. Oxford University Press, Oxford (2010)
19. Roediger-Schluga, T., Barber, M.J.: R&D collaboration networks in the European Framework Programmes: Data processing, network construction and selected results. Int. J. Foresight Innov. Policy **4**(3–4), 321–347 (2008)
20. Savic, M., Ivanovic, M., Putnik, Z., Tütüncü, K., Budimac, Z., Smrikarova, S., Smrikarov, A.: Analysis of ERASMUS staff and student mobility network within a big European project. In: IEEE Mipro (2017)
21. Schilling, M.A., Phelps, C.C.: Interfirm collaboration networks: the impact of large-scale network structure on firm innovation. Manag. Sci. **53**(7), 1113–1126 (2007)
22. Schult, D.A., Swart, P.: Exploring network structure, dynamics, and function using NetworkX. In: Proceedings of the 7th Python in Science Conferences (SciPy 2008), pp. 11–16 (2008)
23. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684), 440–442 (1998)