Projects Collaboration Network for Faculty of Computer Science and Engineering in Skopje

Bojan Ilijoski, Kire Trivodaliev, Emil Stankov and Ilinka Ivanoska
Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, Skopje, N. Macedonia

bojan.ilijoski@finki.ukim.mk, kire.trivodaliev@finki.ukim.mk,
emil.stankov@finki.ukim.mk, ilinka.ivanoska@finki.ukim.mk

***Abstract*—This paper presents the process of building of the collaboration network of employees at the Faculty of Computer Science and Engineering (FCSE), Ss. Cyril and Methodius University in Skopje. The source for constructing the network are the scientific collaborations by joint projects participations of the FCSE employees, extracted from the institutional database. The internal structure of the acquired model of the FCSE's collaboration network is analyzed to investigate the hidden knowledge it possesses. The structure of the network affects the information available to individuals and their opportunities to cooperate. The structure of the network also affects the overall flow of information and the nature of the scientific community. Several centrality metrics were computed for identifying important individuals in the institution. Moreover, community detection algorithm was employed for determining the community structures in the collaborative network. The resulting structures show that such a division corresponds to the real connections between researchers.**

***Index Terms*—collaboration graph, centrality analysis, community structure, scientific project collaborations.**

I. Introduction

In most areas of academic science, collaboration in research and papers publication is essential, leading to a collaboration expansion in wide range of fields [1]–[3]. Collaboration can be seen as a process in which knowledge is transferred in scientific communities, and where individual scientists have the opportunity to enhance their knowledge. As collaboration networks grow, scientists can gain access to information directly (from their collaborators) and indirectly (through their collaborators' collaborators). The structure of larger networks can influence the work done by an individual scientist in ways that are not obvious to them, and additionally, the structure of the entire collaboration network can affect scientific productivity. Some network structures promote diverse and creative work, while other network structures create separation and creativity retention.

In complex networks, a network is said to have a community structure if nodes in the network can be easily grouped into (potentially overlapping) node sets so that each of them is internally dense [4], [5]. In the case of nodes that do not overlap in the community, it follows that the network naturally splits into node clusters with intrinsically dense connections and rare inter-group links. However, overlapping communities are also allowed. The more general definition is based on the principle that a pair of nodes is more likely to be related to each other if they are also members of the same communities and less likely to be connected if they do not share communities. A related, but different problem is community search, where the goal is to find a community that belongs to a particular topic [6], [7].

Studying networks, such as computer and information networks, social networks, and biological networks, often discloses a number of different network characteristics, including communities in a "small world", heavy tail distributions, or community structure [8]–[10]. Community structure refers to the appearance of a network's node clusters that are more closely connected internally than to other clusters in the network. This edges' inhomogeneity indicates that the network has some natural divisions within it. Communities are often defined in terms of each node being placed in one and only one community, and most community discovery methods find this type of community structure. In some cases, however, a better representation may be where foundations are found in more than one community. In social networks, for example, each node can represent a person, and communities can represent different groups of friends: one community for family, another for colleagues, for same sports club etc. [11]–[13].

Identifying a significant community structure in social networks is a difficult task. As the number of real-world datasets increases, there is an increasing demand for more effective algorithms. Existing methods are limited by their computational requirements and rely heavily on network topology, which fails in scalable networks. However, many databases with graphs also include attributes of individual nodes, but current methods cannot incorporate this data.

Community structure is an important aspect of networks and has many implications for various fields, including science, social science, and computer science. Identifying communities within a network can provide insight into the structure and function of the network, and can also facilitate more efficient communication and collaboration within the community. This is particularly substantial in collaboration networks, which are becoming increasingly important in many fields of research as they allow individual scientists to share knowledge and expertise, leading to more efficient and effective research

Supported by Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, Skopje, N. Macedonia

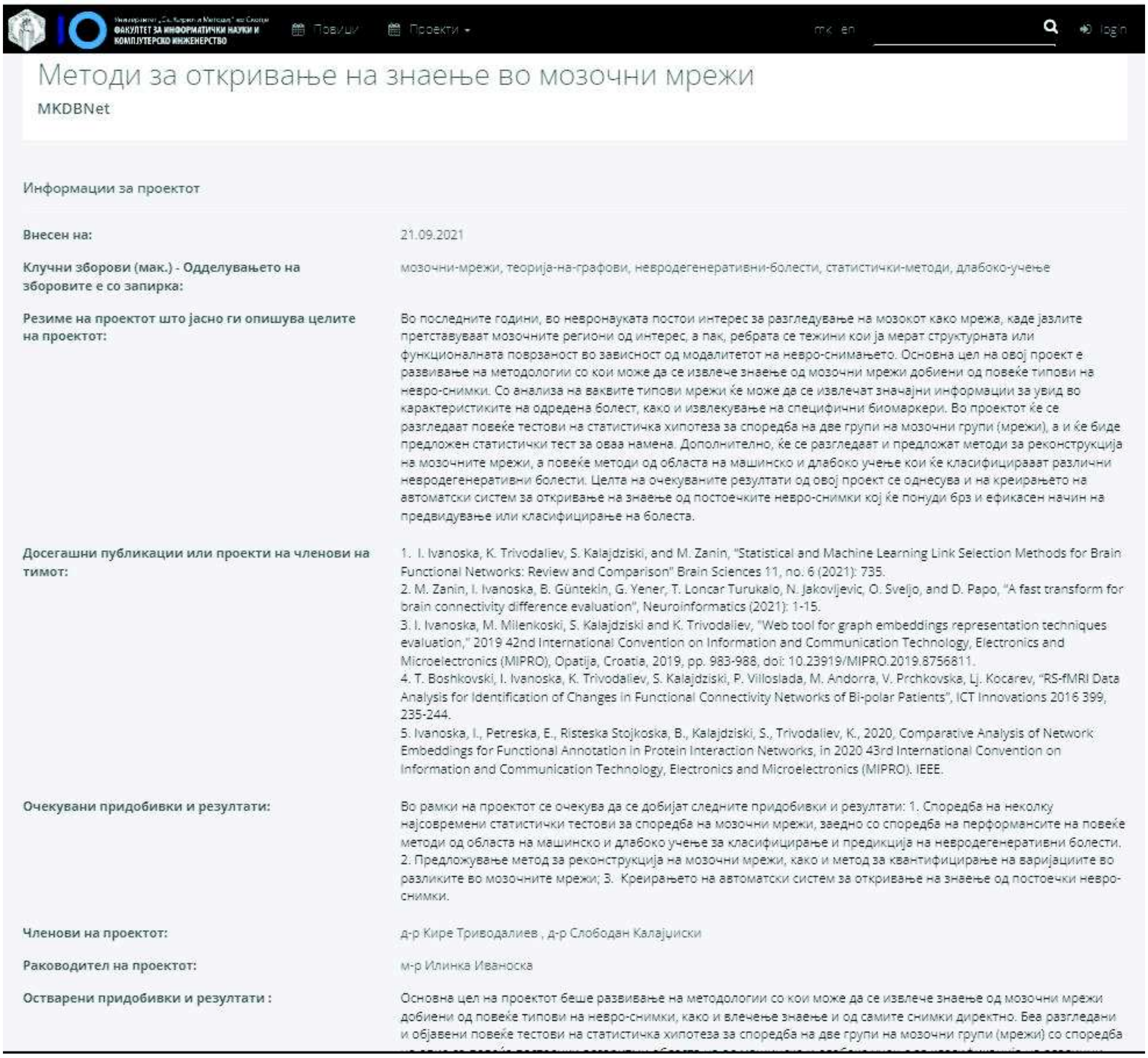

Fig. 1. FCSE projects.finki.ukim.mk project web interface.

outcomes [2], [14]. Collaboration networks can help to build and strengthen professional relationships among researchers, leading to potential future collaborations and collaborations across different disciplines. Thus, collaboration networks are an essential component of modern scientific research and play a vital role in advancing scientific knowledge.

In this paper we analyse the structure of the scientific projects collaboration network of the employees at the Faculty of Computer Science and Engineering (FCSE), Ss. Cyril and Methodius University in Skopje, coming from the portal projects.finki.ukim.mk (see Fig. 1 for web interface of one project). The FCSE network is studied as a social network, examining the social relationships of collaboration among researchers in their ”affiliation” network, in which actors (researchers, FCSE employees) are ”associated” with collaborations on projects. Centrality analysis is performed and community structures within the network are obtained.

The rest of the paper is structured as follows: the dataset used, the architecture of the system that we have built, as well as the centrality metrics and community detection algorithms are explained in detail in section II. The experimental setup with results and discussion of the ran experiments are laid out in section III. Finally, section IV concludes the findings.

II. Materials and Methods

In this section we present the dataset used for building the collaboration network, the system architecture, and the analysis algorithms used.

A. System Architecture

The system architecture for the FCSE collaboration network analysis built is presented in Fig. 2. It contains 3 modules: data source, network extraction module and a network analysis module. The data source used for building the FCSE projects collaboration network is the portal http://projects.finki.ukim.mk. The network extraction module extracts the data from this portal and generates the graph representing the projects collaboration network. The network analysis module then conducts a centrality analysis of the obtained network, and applies an appropriate algorithm for community structure detection.

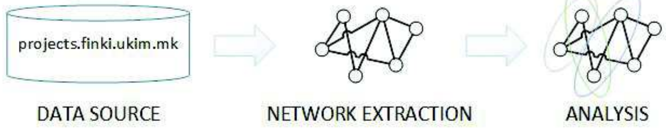

Fig. 2. FCSE system architecture for projects collaboration network analysis.

B. Dataset and Building the Projects Collaboration Network

The data source used to build the FCSE projects collaboration network is the portal http://projects.finki.ukim.mk. The collaboration network is built based on participation(s) in the same project(s), from the data source for FCSE's employees (and their collaborators). The network is generated with its adjacency matrix according to direct collaboration and joint collaborations in FCSE projects. http://projects.finki.ukim.mk is a portal developed at FCSE with the intention for its employees to use the portal to store information about FCSE's national and international projects. For each year, all employees' FCSE projects are stored as in Fig. 1.

The projects' information is parsed from the portal and the collaboration network is built based on memberships in projects and collaboration with the project's principal investigator (PI). The collaboration between a PI and a project member on each project is weighted with 1 in the adjacency matrix of the network, while collaborations between one project member and second/other project member are weighted with 0.5, due to nondirect collaboration. Each collaboration is summed based on all projects and the PI and projects members and the final projects collaboration network is computed.

The resulting graph consists of a total of 264 nodes (FCSE employees and their collaborators, out of which 67 are PIs) with a dense structure and only one connected component (see Fig. 3).

C. Centrality Measures for Network Analysis

Centrality measures are used to identify which nodes/researchers are the most collaborative and strongest influencers in the collaboration network [15], [16]. The different types of centrality in analyzing the network are given as follows:

- Degree centrality: it measures the number of collaborations (connections) of each node [17];
- Betweenness centrality: it measures a node's importance by the number of times it occurs in the shortest paths between other nodes [18]. High betweenness centrality implies high influence over other nodes in the network;
- Closeness centrality: it measures the quickness (minimum number of steps) for one node to connect to others in the network [19]. Higher closeness centrality implies shorter distances to others (faster to make direct collaboration);

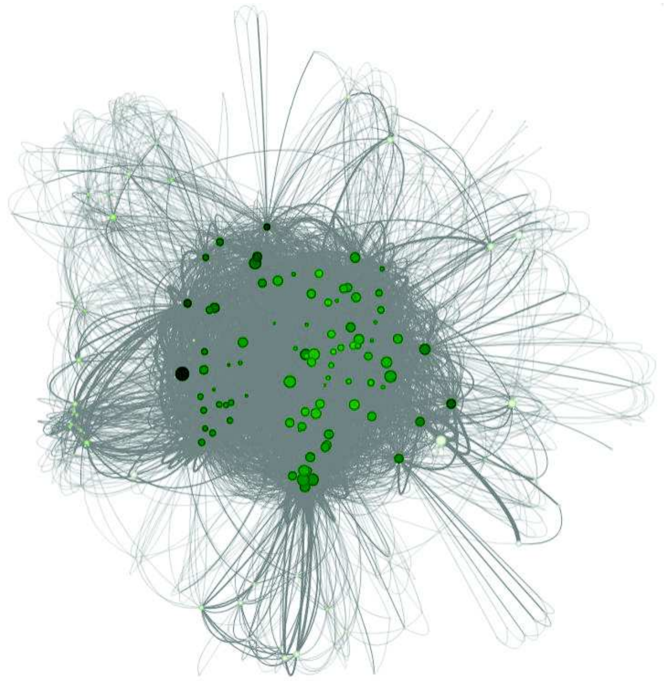

Fig. 3. FCSE projects collaboration network.

- Eigenvector centrality: it measures a node's connection to those who are highly connected [20]. High eigenvector centrality has someone who is influencing several highly connected others and has a hidden control in the collaboration network.

D. Community Structure Detection

As explained previously, community structure refers to the occurrence of clusters of network nodes that are more interconnected than the rest of the network. The number of communities (if any) within a network is usually unknown, and communities are often of unequal size and / or density. Despite these difficulties, several approaches and many methods for finding communities have been developed: minimum-cut method [5], hierarchical clustering [21], [22], Girvan-Newman algorithm [23], Kernighan-Lin algorithm [24], Walktrap algorithm [25], modularity maximization [26], [27], statistical inference [28], k-clique percolation [11], etc.

Here, we use the famous Louvain community structure detection algorithm [26] for discovery of relevant communities between employees/researchers from the FCSE projects collaboration network. The Louvain algorithm is a representative of the graph (component) modularity optimization [9] approach. It compares the edges/connections frequency within a component to other components in the same graph, adding a node to a community if it improves its modularity, and finishing when maximum modularity is achieved.

III. Experiments, Results and Discussion

Following the construction of the FCSE projects collaboration network, we performed a centrality analysis of the whole

TABLE I
SUMMARY TABLE OF TOP 8 EMPLOYEES/RESEARCHERS IN THE FCSE PROJECTS COLLABORATION NETWORK BASED ON CENTRALITY AND PROJECTS ANALYSIS, SORTED IN DESCENDING ORDER BY EACH METRIC OBSERVED.

Degree centrality		Betweennes centrality		Closeness centrality		Eigenvector centrality	
Person	Value	Person	Value	Person	Value	Person	Value
K. Mishev	0.475	K. Mishev	0.081	K. Mishev	0.655	K. Mishev	0.11
S. Gievska	0.448	A. M. Bogdanova	0.078	S. Gievska	0.644	B. Jakimovski	0.109
B. Jakimovski	0.444	K. Zdravkova	0.054	B. Jakimovski	0.642	K. Kjiroski	0.109
B. Koteska	0.432	S. J. Sarknjac	0.053	B. Koteska	0.638	G. Petkovski	0.108
A. M. Bogdanova	0.429	S. Gievska	0.042	A. M. Bogdanova	0.636	V. Bidikov	0.108
G. Velinov	0.421	B. Jakimovski	0.041	G. Velinov	0.633	I. Cvetanovski	0.108
B. Ilijoski	0.417	A. Tenev	0.038	B. Ilijoski	0.631	E. Panovska	0.108
K. Kjiroski	0.417	B. Koteska	0.038	K. Kjiroski	0.631	K. Nakov	0.108

Number of projects			
Person	Total	Person	As PI
K. Mishev	25	B. Jakimovski	18
N. Ilievska	23	N. Ackovska	12
A. Mishev	22	Lj. Antovski	12
V. Dimitrova	22	L. Basnarkov	12
M. Kostoska	22	S. Gievska	12
B. Ilijoski	22	V. D. Ristova	12
S. Loshkovska	21	B. Tojtovska	12
V. Trajkovikj	20	K. Mitreski	12

network. Fig. 4 presents the FCSE projects collaboration network analysis for all 4 centrality measures: degree centrality (top left), betweenness centrality (top right), closeness centrality (bottom left), and eigenvector centrality (bottom right). In each of the centrality graphs, the more projects collaborations there are between two employees/researchers, the stronger the line for that edge is. Moreover, in Fig. 4 the graph nodes' color and size are determined by the centrality metric value, i.e., the higher the centrality, the darker and bigger the node is. For visualization purposes, since the whole graph is dense, it has been thresholded in such a way that collaboration edges with weight values less than 10 are not shown (only the more significant collaborations can be noticed in Fig. 4).

Table I provides a comprehensive summary of the top 8 employees/researchers in the FCSE projects collaboration network, based on the centrality analysis. The centrality metrics values for the top researchers in descending order of the corresponding centrality metric are shown, as well as the top projects numbers that they have worked on, both in total and as a PI. K. Mishev has worked on the most projects in total - 25, followed by N. Ilievska, and A. Mishev. B. Jakimovski has led (has been a PI of) the most projects - 18, followed by N. Ackovska, and Lj. Antovski.

The top 3 researchers by degree centrality are K. Mishev, S. Gievska, and B. Jakimovski, and as these people have the highest number of connections to other people in the network, they correspond to the 3 darkest/biggest nodes of the graph shown in the top left of Fig. 4. Similarly, K. Mishev, A. M. Bogdanova, and K. Zdravkova have the highest influence over others according to the betweenness centrality (see top right of Fig. 4). K. Mishev, S. Gievska, and B. Jakimovski are the fastest in making direct collaborations (according to the closeness centrality analysis - bottom left of Fig. 4. Moreover, K. Mishev, B. Jakimovski, and K. Kjiroski have the highest hidden control in the network (eigenvector centrality analysis - bottom right of Fig. 4).

From the centrality analysis, it can be superficially concluded that in the FCSE projects collaboration network several nodes are always located at the top of many measures of influential position. ”Key” or ”star” actors such as K. Mishev, B. Jakimovski, A. M. Bogdanova, are leading figures across all measures, which makes them important collaborators in FCSE projects.

The calculated modularity of the whole network is 0.17429 yielding the existence of a small dense connected communities structure. Using the Louvain community detection algorithm, 9 communities have been identified. Fig. 5 shows the communities detected in the collaboration network, where nodes with different colors are in different communities. The graph in Fig. 5 is thresholded for visualization purposes in the same way as in Fig. 4. As can be seen from the figure, the detected communities vary in size: from 2 nodes (people) in the smallest community, to more than 10 nodes (people) in the largest. Furthermore, the stronger influence a person has in its own community in the FCSE projects collaboration network, the bigger the node for that person is in the graph in Fig. 5. We can see that the leading researchers do not form a ”closed” elite, they are scattered across the network, which makes the network stable with losing its most important nodes. The detected communities reflect the real connections and groups situation at the institution.

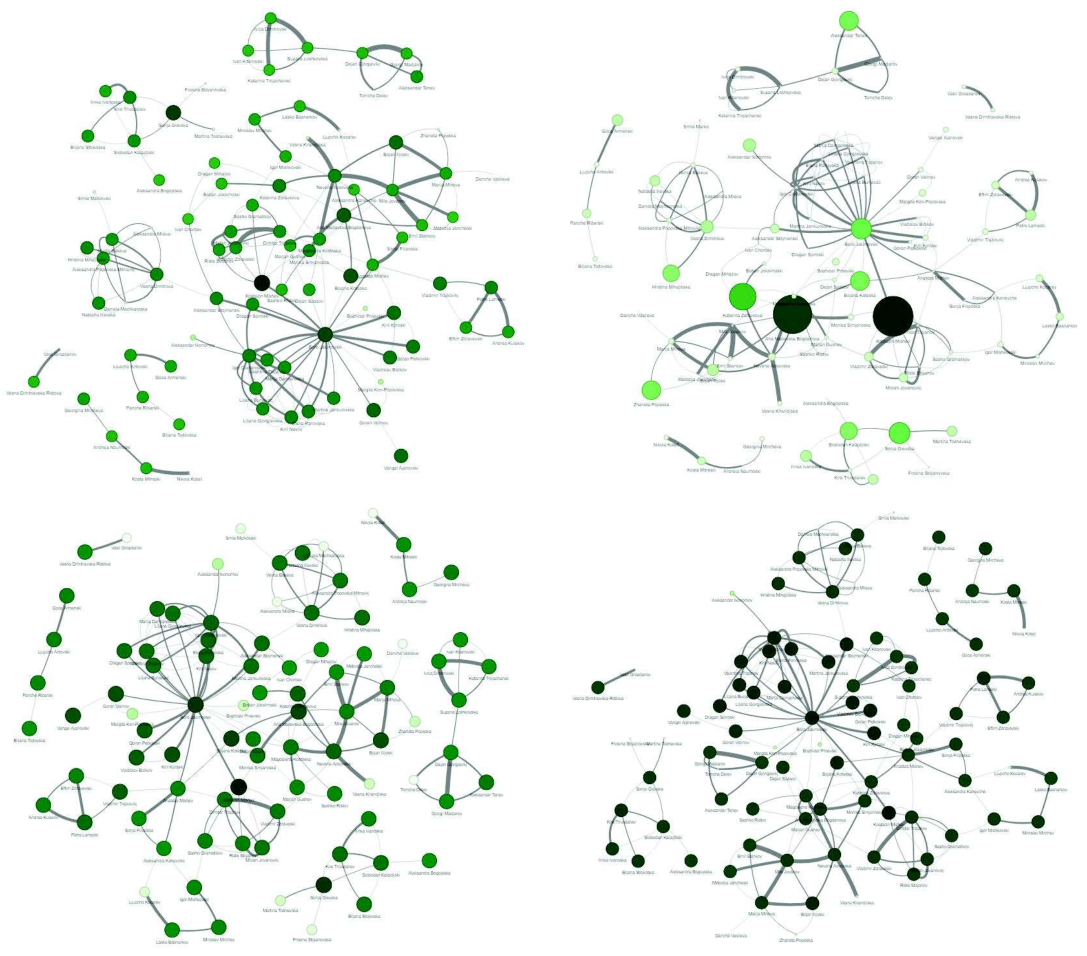

Fig. 4. FCSE projects collaboration network centrality visualization: top left) degree centrality; top right) betweenness centrality; bottom left) closeness centrality; bottom right) eigenvector centrality.

IV. Conclusion

In this paper we have presented a centrality and community structure based analysis approach for the FCSE projects collaboration network. The centrality and community detection results, yield that the communities globally do not undergo major changes. The leading FCSE employees do not form a ”closed” elite” and several of them are always at the top of many influential position measures.

Although the FCSE collaboration network is not large, the results are useful to provide future possible illustration of approaches for measuring structural network characteristics, identifying hubs and peripheral nodes and scale-free and small world in-depth network investigation. In future research we want to explore the diploma thesis collaboration network of the same institution and see whether if an employee is in a graduate thesis jury together with other colleagues than that yields to future together paper publication. Additionally, the network can be expanded semantically to investigate whether a single professor’s cluster is elected in a jury together because they work together in the same field or for other reasons. Natural language processing techniques can be used for graduation thesis titles to predict future collaborations and future publishing papers in a jury member’s or mentor’s field of work. Questions that may be asked are, e.g., what is specific about these individuals that occur in different clusters? Is it because they have many weak connections with other characters, or do they have a strong connection with a certain character and more with other weakly connected ones?

Acknowledgment

This work was partially financed by the Faculty of Computer Science and Engineering at the Ss. Cyril and Methodius University in Skopje.

References

[1] M. E. J. Newman, “Scientific collaboration networks. i. network construction and fundamental results,” *Physical review E*, vol. 64, no. 1, p. 016131, 2001.

[2] ——, “The structure of scientific collaboration networks,” *Proceedings of the national academy of sciences*, vol. 98, no. 2, pp. 404–409, 2001.

[3] A.-L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, “Evolution of the social network of scientific collaborations,” *Physica A: Statistical mechanics and its applications*, vol. 311, no. 3-4, pp. 590–614, 2002.

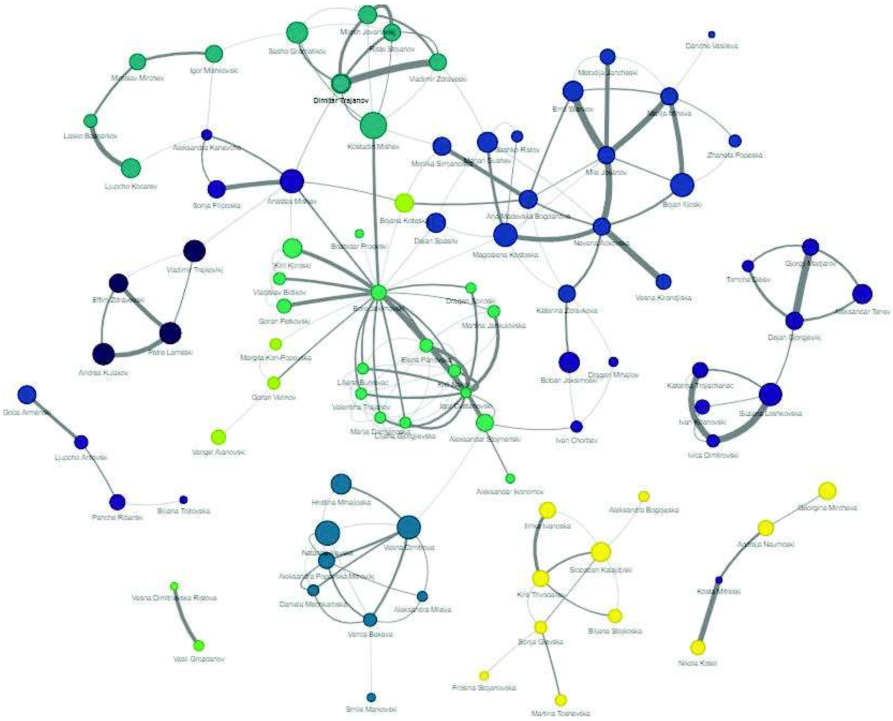

Fig. 5. FCSE projects collaboration network communities discovered with Louvian community detection algorithm.

[4] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical review E*, vol. 69, no. 2, p. 026113, 2004.

[5] M. E. J. Newman, “Detecting community structure in networks,” *The European physical journal B*, vol. 38, pp. 321–330, 2004.

[6] I. Himelboim, M. A. Smith, L. Rainie, B. Shneiderman, and C. Espina, “Classifying twitter topic-networks using social network analysis,” *Social media+ society*, vol. 3, no. 1, p. 2056305117691545, 2017.

[7] Z. Zhao, S. Feng, Q. Wang, J. Z. Huang, G. J. Williams, and J. Fan, “Topic oriented community detection through social objects and link analysis in social networks,” *Knowledge-Based Systems*, vol. 26, pp. 164–173, 2012.

[8] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, “Defining and identifying communities in networks,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 9, pp. 2658–2663, 2004.

[9] M. E. J. Newman, “Modularity and community structure in networks,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[10] M. A. Porter, J.-P. Onnela, and P. J. Mucha, “Communities in networks,” *Notices of the American Math. Soc.*, vol. 56, no. 9, pp. 1082–1097, 2009.

[11] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, “Uncovering the overlapping community structure of complex networks in nature and society,” *nature*, vol. 435, no. 7043, pp. 814–818, 2005.

[12] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, “Comparing community structure to characteristics in online collegiate social networks,” *SIAM Review*, vol. 53, no. 3, pp. 526–543, 2011.

[13] E. M. Airoldi, D. Blei, S. Fienberg, and E. Xing, “Mixed membership stochastic blockmodels,” *Advances in neural information processing systems*, vol. 21, 2008.

[14] P. Paraskevopoulos, C. Boldrini, A. Passarella, and M. Conti, “The academic wanderer: structure of collaboration network and relation with research performance,” *Ap. Net. Science*, vol. 6, no. 1, pp. 1–35, 2021.

[15] S. P. Borgatti and M. G. Everett, “A graph-theoretic perspective on centrality,” *Social networks*, vol. 28, no. 4, pp. 466–484, 2006.

[16] A. Landherr, B. Friedl, and J. Heidemann, “A critical review of centrality measures in social networks,” *Business & Information Systems Engineering*, vol. 2, no. 6, pp. 371–385, 2010.

[17] J. Nieminen, “On the centrality in a graph,” *Scandinavian journal of psychology*, vol. 15, no. 1, pp. 332–336, 1974.

[18] L. C. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

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

[20] P. Bonacich, “Factoring and weighting approaches to status scores and clique identification,” *Journal of mathematical sociology*, vol. 2, no. 1, pp. 113–120, 1972.

[21] S. C. Johnson, “Hierarchical clustering schemes,” *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.

[22] G. Gan, C. Ma, and J. Wu, *Data clustering: theory, algorithms, and applications*. Society for Industrial and Applied Mathematics, 2020.

[23] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[24] B. W. Kernighan and S. Lin, “An efficient heuristic procedure for partitioning graphs,” *The Bell system technical journal*, vol. 49, no. 2, pp. 291–307, 1970.

[25] P. Pons and M. Latapy, “Computing communities in large networks using random walks,” *J. Graph Alg.s Appl.*, vol. 10, no. 2, pp. 191–218, 2006.

[26] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, pp. 1–12, 2008.

[27] S. Fortunato, “Community detection in graphs,” *Physics reports*, vol. 486, pp. 75–174, 2010.

[28] M. B. Hastings, “Community detection as an inference problem,” *Physical Review E*, vol. 74, no. 3, p. 035102, 2006.