Using Social Network Analysis to Investigate Students' Collaboration Patterns in eMUSE Platform

Alex Becheru, Elvira Popescu
Department of Computers and Information Technology
University of Craiova
Craiova, Romania
becheru@gmail.com, popescu_elvira@software.ucv.ro

***Abstract*—Social network analysis can be used to investigate collaboration in learning networks, which can be modeled as social graphs. We have already proposed a conceptual framework for knowledge extraction and visualization from a social media-based learning environment, starting from specific educational needs identified by the instructors. In the current paper, we experimentally validate this framework on our eMUSE social learning platform. In particular, we investigate students' collaboration patterns in a project-based learning scenario. Social network analysis techniques are used to extract knowledge regarding specific differences in blog vs. microblog support, as well as intra-team vs. inter-team collaboration; several salient students and teams are also analyzed in more detail.**

***Keywords*— social network analysis; collaboration patterns; social learning environment; knowledge extraction and visualization**

I. Introduction

Social Network Analysis (SNA) is a cross-disciplinary research field, which explores social environments that can be modeled as graphs [2]. In particular, SNA can be used to investigate students' interactions and collaboration patterns taking place in educational environments. This could prove very valuable to the instructor, by offering an insight into the structure of online learning communities, various communication aspects and patterns of academic collaboration [17]. In addition, collaborative filtering techniques could be used to generate personalized recommendations for students, in terms of learning objects, relevant links, relevant courses or most appropriate study partners [17].

In recent years, several studies have used SNA methods and metrics for technology-enhanced learning. For example, Maglajlic and Gutl [13] employed techniques such as *cliques, centrality* and *density* to measure and enhance collaboration in an educational environment and as an early detection method of potential weak trainees. Crespo and Antunes [8] used diverse variants of the *PageRank* algorithm [15] for ranking learners and focused on exploring and predicting teamwork results. Haythornthwaite [10] also relied on *number of ties, density, centrality* and *cliques* metrics to answer specific educational needs.

Continuing this line of work, in a previous paper [5] we proposed a *conceptual framework for knowledge extraction and visualization from a social media-based learning environment*. A set of knowledge extraction methods based on SNA techniques were proposed, starting from specific educational needs identified by the instructor. More specifically, the following seven pedagogical knowledge needs were proposed: *Determine the collaboration network (KN1), Determine methods to quantify collaboration over various time intervals and various network granularities (KN2), Determine salient students or communities of students (KN3), Determine if the course environment adequately supports collaboration (KN4), Determine external factors that influence the evolution of collaboration (KN5), Determine the impact of instructors' actions on collaboration (KN6), Determine clear and comprehensive methods of presenting the information extracted (KN7)* [5].

In this paper we aim to experimentally validate the proposed conceptual knowledge extraction framework on our eMUSE social learning environment [16]. More specifically, we will investigate students' collaboration patterns in a project-based learning scenario and use SNA to extract knowledge that answers instructors' pedagogical needs. More details regarding the eMUSE platform and the context of study are included in section 2. An overview of the collaboration process and social learning graph is presented in section 3. Various issues are subsequently explored in more detail: differences in collaboration patterns supported by blog vs. microblog (in section 4), intra-team vs. inter-team collaboration (in section 5) and in-depth student / team analysis (in section 6). Finally, conclusions and future work directions are included in section 7.

II. Context of Study

The context of our study is a social learning environment, called eMUSE [16]. The platform integrates several social media tools (wiki, blog, microblogging tool) that students use for communication and collaboration support. All students' social media traces are monitored and recorded by eMUSE; in addition, the platform provides basic administrative services, data visualizations, as well as peer evaluation and grading support.

The social learning environment has been used at the University of Craiova, Romania, for the past seven years, with various incremental improvements. For the current study, we take into account the latest installment of a course on "Web Applications Design", taught to 4th year undergraduate students in Computer Science in the first semester of the 2016-2017 academic year. A project-based learning (PBL) approach was used in which students collaborated in teams of 4 peers in order to build a web application of their choice (e.g., a virtual bookstore, an online auction website, a professional social network, an online travel agency, etc.).

The PBL scenario was implemented in blended mode, with weekly face-to-face meetings between each team and their instructor. These meetings were complemented by the use of three social media tools for online communication and collaboration. MediaWiki was used for collaborative writing tasks, for gathering and organizing team knowledge-base and resources, and for documenting the project. Blogger was used for reporting the progress of each project similar to a "learning diary", for publishing ideas and resources, as well as for providing feedback and solutions to peer problems. Each team had its own blog, however inter-team cooperation was encouraged as well. Twitter was introduced for fostering additional connections to peers and for posting short news, announcements, questions, and status updates regarding each project [16].

In addition, students had to create four compulsory intermediary presentations in order to be actively engaged throughout the semester and to discourage the practice of activity peak at the end. Each student's performance assessment took into account both the final product delivered at the end of the semester and the continuous collaborative work carried out on the social media tools made available in eMUSE.

A total of 32 students participated in the study, being split in 8 teams (4 students per team). The number of student actions on the social media tools, recorded by eMUSE at the end of the semester, included 1686 tweets, 271 blog entries (159 posts and 112 comments) and 1696 wiki page revisions and file uploads.

For the current study, we are interested in students' communication and collaboration patterns; since the wiki actions log did not include explicit information on students' interactions, we excluded them from this analysis. We therefore take into account students' communication on the blog and Twitter, as described next.

III. Collaboration Overview: Building the Social Graph

First of all, a custom tool was designed and implemented for processing the raw data collected by eMUSE; Python 3.5 programming language with *NetworkX* graph analysis library [9] were employed for building the social graph. More specifically, a directed graph was built starting from students' interactions on the blog and microblogging tool; vertices represent learners and links represent messages exchanged through the two social media tools integrated in eMUSE. The types of interactions (collaborations) taken into account on Blogger and Twitter respectively are detailed next.

Regarding the blog, since each team has one common blog space, we assume that each post is addressed to the corresponding team members. Hence, for each blog post we consider a collaboration between the author of the post (source student) and each respective team peer (target students). Thus, 477 collaborations (i.e., 159 posts * 3 team members) are considered. In addition, each blog comment is directly addressed to the initial blog post author. Hence, a collaboration between the author of the comment (source student) and the author of the initial post (target student) is considered. Thus, for each blog comment we can extract one interaction, resulting in 112 additional collaborations.

As far as Twitter is concerned, collaboration among students is associated to the *reference* ("*@username*") and *retweet* mechanisms. This was encountered in 1043 tweets, i.e., 62% of all Twitter actions. For every action presenting the referencing mechanism, we considered a collaboration between the author of the tweet (source student) and all the referenced students in it (target students). Thus, a number of 1635 interactions were extracted from Twitter actions.

Overall, a total of 2224 collaborations were extracted, yielding 263 distinct source-target pairs. Therefore, the social graph includes 32 vertices and 263 links. In order to capture the quantitative aspects of the collaboration process, we added a *weight* attribute to each link; this is equal to the number of collaborations between the source student and target student. The range of values taken by the weight was found to be between 1 and 100. A visual representation of the obtained social graph is included in Fig. 1.

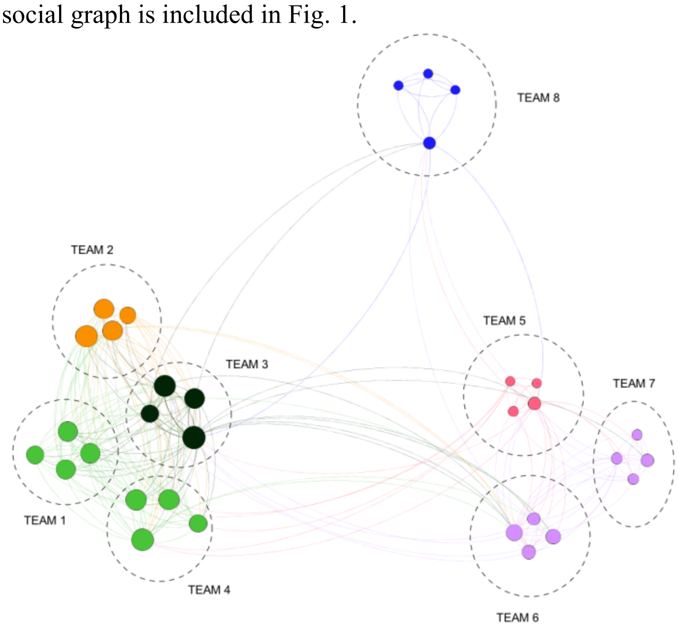

Fig. 1. Graph depicting all social media interactions among students (*Base graph*). Vertices represent students and links represent collaborations among them. The size of each vertex is proportional to its respective *eigenvector* centrality value. The colour of the vertices depicts students' clustering as determined by the *modularity* algorithm [14]. Links are coloured according to their respective source vertex. Graph plotting was done with *Gephi's Force Atlas 2* algorithm [11].

While this graph includes students' interactions on both blog and Twitter, in the next section we explore the different collaboration patterns supported by each social media tool.

IV. Twitter and Blog Collaboration Patterns

By applying filtering on the above presented *Base graph*, we were able to obtain two subgraphs corresponding to the blog collaborations (as depicted in Fig. 2) and Twitter collaborations (as depicted in Fig. 3). Please note that the union of these two graphs is the *Base graph*. Subsequently, each of these graphs was analyzed through various SNA metrics, some of which are presented in Table 1; all computations were done with *Gephi* network analysis tool [1]. In what follows, we discuss the collaboration patterns for each social media tool, as reflected in the metrics and visual representations.

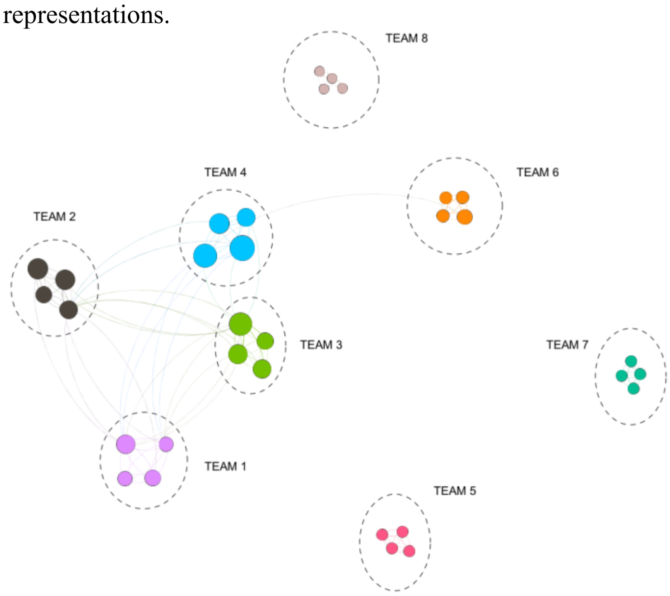

Fig. 2. Graph depicting students' collaborations taking place on the blog (*Blog graph*); graph plotting conventions are the same as in Fig. 1.

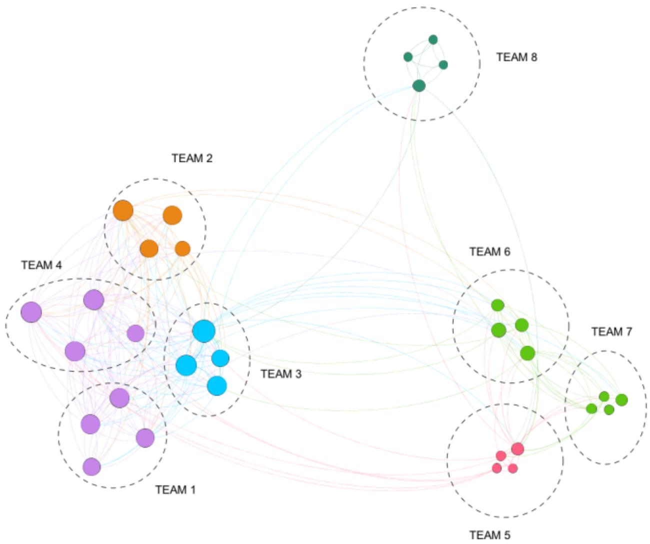

Fig. 3. Graph depicting students' collaborations taking place on Twitter (*Twitter graph*); graph plotting conventions are the same as in Fig. 1.

TABLE I. SNA DIRECTED WHOLE-GRAPH METRICS. NO. COMMUNITIES IS DETERMINED THROUGH THE MODULARITY ALGORITHM.

Metric	***Base graph***	***Twitter graph***	***Blog graph***
No. links	263	253	117
Avg. degree	16.43	15.81	7.31
Avg. weighted degree	69.31	51.09	18.21
Diameter	5	5	5
Avg. path length	2.13	2.13	2.21
Density	0.26	0.25	0.11
Modularity	0.49	0.43	0.74
Avg. clustering coefficient	0.65	0.63	0.73
No. WCC (weakly connected components)	1	1	4
No. SCC (strongly connected components)	1	2	6
No. communities	6	6	8

As seen in Table 1, the *Twitter graph* and *Base graph* have very similar results. This is especially due to the fact that 96% of the collaboration links present in the *Base graph* are also present in the *Twitter graph* (see *No. links* metric). By comparison, the *Blog graph* contains only 44% of the total collaboration links, positioning Twitter as a more comprehensive tool for collaboration. The *Avg. degree* metric stands for the number of peers a student has collaborated with (on average). The results depict that students engage in collaborations with other team members also, since a value of 3 would have been an indication of intra-team collaboration only. As shown by the *Avg. weighted degree* metric, collaborating students exchange multiple messages, with an average of more than 4 with each peer. Moreover, students engage in collaborations with many more peers on Twitter than on blog. The *Diameter* and *Avg. path length* metrics represent markers of information diffusion in the social graph. Based on the complex networks' classification method introduced in [4] and the metrics' values obtained, we can state that the *Base* and *Twitter graphs* are of type *core-periphery*. In our scenario, this categorisation represents evidence of a learning environment that nurtures information exchange.

Density indicates the tendency of students to collaborate with peers from other teams. The results show that Twitter better supports such collaboration patterns, as compared to the blog. *Modularity* [14] and *Avg. clustering coefficient* address the division of communities that arise in social graphs. A higher *modularity* (range [-1/2,1]) depicts dense collaborations among community members and sparse collaborations between members of different communities. Graphs with high *Avg. clustering coefficient* (range [0,1]) characterize inner communities that are close to forming complete graphs/cliques. As it can be seen in Table 1, the blog provides better support for this type of collaborations. Furthermore, *No. WCC* and *No. SCC* also depict blog collaborations as intra-team interactions, i.e., highly coupled communities of collaboration arise on the blog. By comparing Fig. 1-3, we can easily notice that blog collaborations are mainly established among members of the same team, as intended through the instructional scenario. Conversely, *No. communities* metric shows that collaboration among teams is better supported by Twitter, i.e., some teams reach a high level of collaboration and act as one larger community, as seen in Fig. 3.

Based on the above discussion, we argue that both social media tools provide adequate support for collaboration; Twitter supports both inter-team and intra-team interactions, while Blogger mainly supports intra-team collaboration. Moreover, both tools seem to meet the educational purposes intended in the instructional scenario, as mentioned in section 2.

V. Intra-Team and Inter-Team Collaboration

In what follows, we discuss in more detail the intra-team and inter-team collaboration patterns, while highlighting salient students and teams. By using a *reduction transformation* of the *Base graph* (such that vertices depicting students are aggregated into their respective team vertex, while keeping inter-teams links), we were able to construct the *Teams graph*, as shown in Fig. 4. Hence, we observe two large communities of collaboration: the first is formed by teams 1 to 4 (red community), while the second is formed by teams 5 to 8 (green community). This can be explained by the fact that teams in the red community were scheduled in a different face-to-face session with the instructor than teams in the green community. This means that teams 1-4 attended each other's intermediate presentations, but not the presentations of teams 5-8 (and the other way around); hence, closer ties were formed inside each community. No clear explanation arose regarding the lower level of inter-team collaboration inside the green community, compared to the red community; it should be mentioned however that students in the green community had lower average marks than students in the red community. We can thus argue that various external factors may influence the collaboration patterns.

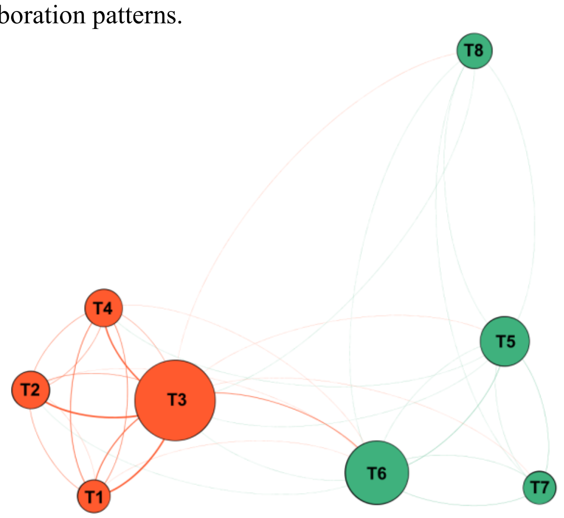

Fig. 4. Graph depicting collaborations established between the teams, both on blog and Twitter (*Teams graph*). Vertices represent teams and links represent inter-team collaborations. The size of each vertex is proportional to its respective *betweenness* centrality value. The colour of the vertices depicts teams' clustering as determined by the *modularity* algorithm. Links are coloured according to their respective source vertex. Links' thickness is proportional to their respective weight.

With the exception of Team 8 members, all students engage in inter-team collaboration, each team having two highly active (dominant) students. Sample graphical representations for Team 3 and Team 8 can be seen in Fig. 5 and Fig. 6 respectively.

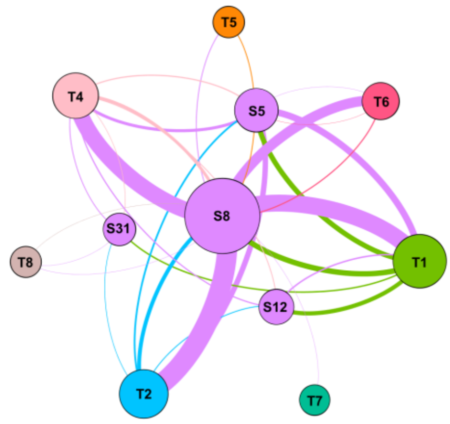

Fig. 5. Inter-team collaborations of students from Team 3. Vertices S5, S8, S12, S31 represent Team 3 students, while the other vertices represent teams (T1, T2, T4-T8). Links depict collaborations among vertices and are coloured according to their source vertex. The size of each vertex is proportional to its respective weighted degree centrality. Links' thickness is proportional to their weight. Graph plotting was done with Gephi's double circular layout.

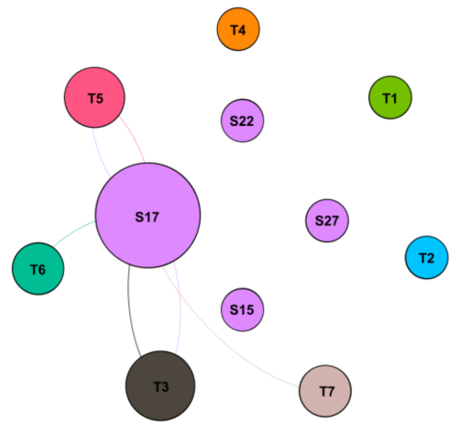

Fig. 6. Inter-team collaborations of students from Team 8. Vertices S15, S17, S22, S27 represent Team 8 students, while the other vertices represent teams (T1-T7). Graph plotting conventions are the same as in Fig. 5.

Nevertheless, the majority of collaborations (1598 out of 2224 or ~72%) take place among members of the same team. These results were expected, as intra-team collaboration was required by the instructors, while inter-team collaboration was only recommended. Hence, the difference can be explained by the pedagogical scenario and the corresponding instructors' actions. Furthermore, we found that all teams display consistent collaboration among their members, especially for the *red community*. Also, every team has two dominant members for this type of collaboration, the same students that are dominant for the inter-team collaborations (except Team 8). Graphical representations of the intra-team collaborations for Teams 3 and 8 can be seen in Fig. 7 and Fig. 8 respectively.

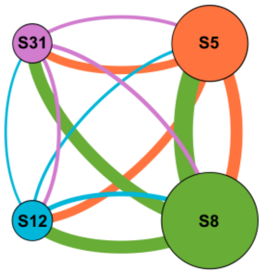

Fig. 7. Intra-team collaborations for Team 3. Vertices represent students and links collaborations among them. The size of each vertex and is proportional to its respective weighted degree centrality. Links are coloured according to their source vertex and links' thickness is proportional to their weight. Graph plotting was done with Gephi's circular layout.

Fig. 8. Intra-team collaborations for Team 8. Graph plotting conventions are the same as in Fig. 7.

By studying the intra-team and inter-team collaboration patterns, we were able to discover salient teams (T3 & T8) and students (S8 & S15). These teams and students were chosen as they stand on opposite sides of the collaboration spectrum and play key roles in the overall collaboration environment. In the next section, we discuss their particularities in detail.

VI. Student / Team Level Analysis

In order to discover salient students and teams we employed the graph metrics listed in Table 2. These were chosen as they are representative for both the topological aspects of the social graph and for the quantitative aspects of the collaborations. *Hubs* and *Authorities* [12] refer in our case to students that initiate many collaborations (source students), respectively to students that are involved in many collaborations initiated by others (target students); i.e., students that share knowledge versus those that receive knowledge. *Betweenness* [7] emphasizes students that act as bridges of collaboration, i.e., they facilitate collaboration among peers that otherwise are part of different communities. As seen in Fig. 1, 4 and 5, members of Team 3 appear to act as a communication bridge between other teams. Furthermore, in our scenario we interpret *Closeness* [3] as a marker of involvement in the overall collaboration of the respective graph. *Eigenvector* [6] and *PageRank* [15] determine the influence of a student's position in the collaboration graph. *Weighted in-degree* and *out-degree* count for the number of collaborations received, respectively for the number of collaborations initiated by a student/team. Highly ranked students and teams are included in Table 2.

As mentioned in the previous section and also shown in Table 2, student S8 can be considered as the top ranked student (collaboration-wise). In order to explain her almost unchallenged top ranking we conducted further investigations. University records show she was involved in two long-duration foreign exchange projects in different locations, which could suggest a highly social person. Her peers confirmed that S8 is indeed a very sociable and talkative learner. By conducting an interview with the student, we determined that she was actively engaged in social media, managing a globetrotter blog. Furthermore, she has participated in formal and informal training on social media techniques. Hence, we reached a potential explanation of the high collaboration rankings achieved.

Moreover, student S8 is part of team T3, which can also be considered as the *leading* team from a collaboration point of view. The importance of this team can be explained by two factors. First, by looking at *Out-degree* metric in Table 2, we can see that students in T3 initiated the highest number of collaborations; however, the team is not placed in the top 3 according to *In-degree* metric. Hence, we argue that T3 is actively sharing more knowledge than it receives, supposition confirmed to some extent by Fig. 5. Second, the other metrics in Table 2 show that T3 is situated at the very center of the collaboration paths, as can be seen also in Fig. 4. Hence, the quantitative and graph positioning factors make this team the highest ranked collaboration-wise.

TABLE II. SNA DIRECTED GRAPH METRICS FOR VERTICES. SX STANDS FOR STUDENT X, TX STANDS FOR TEAM X. CLOSENESS ON THE *BLOG GRAPH* WAS CONSIDERED ONLY FOR THE GIANT COMPONENT. ALL COMPUTATIONS WERE DONE WITH GEPHI NETWORK ANALYSIS TOOL.

	Base Graph				***Twitter Graph***				***Blog Graph***				***Teams Graph***			
Metric / Rank	**#1**	**#2**	**#3**	**#4**	**#1**	**#2**	**#3**	**#4**	**#1**	**#2**	**#3**	**#4**	**#1**	**#2**	**#3**	**#4**
Authority	S9	S8	S10	S11	S8	S10	S9	S12	S10	S8	S9	S5	T6	T3	T4	T1
Hub	S8	S11	S7	S5	S8	S11	S7	S5	S11	S3	S8	S1	T3	T5	T6	T1
Betweenness	S8	S17	S32	S16	S8	S17	S32	S16	S8	S9	S11	S1	T3	T6	T5	T2
Closeness	S8	S32	S11	S5	S8	S32	S11	S5	S8	S1	S3	S11	T3	T6	T5	T2
Eigenvector	S8	S9	S11	S5	S8	S5	S9	S11	S10	S9	S8	S4	T3	T6	T8	T5
PageRank	S8	S15	S5	S16	S8	S9	S1	S5	S10	S16	S5	S7	T3	T4	T1	T2
Weighted In-degree	S5	S8	S12	S31	S8	S5	S12	S31	S31	S5	S12	S6	T4	T2	T1	T3
Weighted Out-degree	S8	S5	S1	S11	S8	S5	S1	S32	S8	S5	S11	S15	T3	T1	T4	T2

Conversely, T8 stands as the team with the lowest number of inter-team collaborations. Even if in Fig. 4 it appears to be part of the *green community,* at a more in-depth analysis on the *Base graph,* the *modularity* algorithm depicts it as a separate community. In addition, student S17 was the only member of T8 who engaged in collaborations with members of other teams. However, S17 is not the top ranked collaborator inside T8, but rather S15. Therefore, a good strategy for the instructors would have been to encourage S15 to cross the team boundaries and collaborate also with students from other teams. This way, T8 might have been better integrated in the course environment and more engaged in collaborations.

VII. Conclusion and Future Work

Through the current study, we were able to satisfy, to some extent, the following pedagogical knowledge needs mentioned in the introduction. In section 3, we managed to determine the collaboration network and quantify collaborations over several graph granularities (*KN1* and *KN2*). By analyzing various traits of the collaboration process through SNA metrics, we emphasized salient students and teams (*KN3*). Through detecting communities of collaboration, we were able to better comprehend the collaboration patterns (*KN2-6*). Our analyses showed that the course environment adequately supports both intra-team and inter-team collaboration (*KN4*). We argued that students' knowledge level, background in using social media and face-to-face meetings might have an impact on the collaboration patterns, as proof of the influence of external factors (*KN5*). We also observed the impact of the instructional scenario and teachers' instructions on promoting intra-team collaborations (*KN6*). Finally, by providing social graph plots and ranked lists of students, we managed to present the extracted knowledge in a clear and comprehensive way (*KN7*). Therefore, we successfully validated the proposed conceptual knowledge extraction framework on our eMUSE social learning environment.

Nevertheless, we acknowledge that further research is required in order to fully satisfy instructors' knowledge needs. Hence, as future work, we will focus on capturing qualitative aspects of the collaboration process, by developing additional link weighting methods and by considering the textual complexity of students' content. In addition, methods for capturing the collaboration over various time intervals can be proposed, so that instructors will be able to follow students' progress. We also plan to study and develop new graph plotting methods suitable for our context. Finally, the social network analysis can be extended to all available datasets, collected by eMUSE platform over the course of seven years.

Acknowledgment

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-2604.

References

[1] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: an open source software for exploring and manipulating networks", Proc. ICWSM 2009 (3rd AAAI Conference on Weblogs and Social Media), pp. 361-362, 2009.

[2] V. Batagelj, P. Doreian, A. Ferligoj, and N. Kejzar, Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. John Wiley & Sons, 2014.

[3] A. Bavelas, "Communication patterns in task-oriented groups", The Journal of the Acoustical Society of America, vol. 22(6), pp. 725-730, 1950.

[4] A. Becheru and C. Badica, "Complex networks' analysis using an ontology-based approach: initial steps", Knowledge Science, Engineering and Management, Springer, pp. 326-337, 2014.

[5] A. Becheru and E. Popescu, "Design of a conceptual knowledge extraction framework for a social learning environment based on social network analysis methods", Proc. ICCC 2017 (18th International Carpathian Control Conference), pp. 177-182, 2017.

[6] P. Bonacich, "Some unique properties of eigenvector centrality", Social Networks, vol. 29(4), pp. 555-564, 2007.

[7] U. Brandes, "A faster algorithm for betweenness centrality", Journal of Mathematical Sociology, vol. 25(2), pp. 163-177, 2001.

[8] P.T. Crespo and C. Antunes, "Predicting teamwork results from social network analysis", Expert Systems, vol. 32(2), pp. 312-325, 2015.

[9] A. Hagberg, D. Schult, and P. Swart, "Exploring network structure, dynamics, and function using NetworkX", Proc. SciPy2008 (7th Python in Science Conference), pp. 11-15, 2008.

[10] C. Haythornthwaite, "Social network methods and measures for examining e-learning", Technical Report, 2005.

[11] M. Jacomy, "ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software", PLoS ONE 9(6), e98679, 2014.

[12] J.M. Kleinberg, "Hubs, authorities, and communities", ACM Computing Surveys (CSUR), vol. 31.4es, 1999.

[13] S. Maglajlic and C. Gutl, "Efficiency in e-learning: Can learning outcomes be improved by using social networks of trainees and tutors?", Proc. ICL 2012 (15th International Conference on Interactive Collaborative Learning), 2012.

[14] M.E.J. Newman, "Modularity and community structure in networks", Proceedings of the National Academy of Sciences, vol. 103(23), pp. 8577-8582, 2006.

[15] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web", Technical Report, Stanford InfoLab, 1999.

[16] E. Popescu, "Providing collaborative learning support with social media in an integrated environment", World Wide Web, vol. 17(2), pp. 199–212, 2014.

[17] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 40(6), pp. 601-618, 2010.