Exploring study profiles of Computer Science students with Social Network Analysis

Nidia Guadalupe López Flores
Department of Computer Science,
Reykjavík University, Iceland
nidia20@ru.is

Anna Sigridur Islind
Department of Computer Science,
Reykjavík University, Iceland
islind@ru.is

María Óskarsdóttir
Department of Computer Science,
Reykjavík University, Iceland
mariaoskars@ru.is

Abstract

Information technology is widely adapted in all levels of education. The extensive information resources facilitate enhanced human capacity and the social environment to support learning. In particular, Social Network Analysis (SNA) has been broadly used in teaching and learning practices. In this paper, we perform community detection analysis to identify the learning behavior profiles of undergraduate computer science students in a Nordic university. The social network was created using 273 responses to an online survey. The students themselves provided their social connections at the university, and node attributes were created based on responses to questions regarding Educational Values, Goals Orientation, Self-efficacy, and the university teaching methods. We analyze the biggest communities to identify the factors that characterize the learning strategy and preferences of undergraduate computer science students.

1. Introduction

Information technology has become an essential tool of education. It is rich in information resources and can extend human capacity and the social environment to support learning. As part of its rapid growth, Social Network Analysis (SNA) has been broadly used in teaching and learning practices [1]. Recent literature suggests that the potential of Learning Analytics and Educational Data Mining offer benefits through the use of educational data both for teachers and students to further understand the way students approach their learning [2]. With the increase in online learning brought on by the COVID-19 pandemic, there is a greater need for understanding students' social structures in relation to study preferences and motivation so that universities can better accommodate the needs of more students, especially of underrepresented students [3].

This paper analyzes and describes the study profile of undergraduate students of four Computer Science related programs through Social Network Analysis and community detection. In particular, we are interested in knowing their learning preferences regarding group working, physical attendance of lectures, self-efficacy perception and goal orientation. We aim to answer the research question: Can Community Detection algorithms applied to the social network of students identify undergraduate study profiles at a Computer Science Department? To answer the question, we analyzed the social connections of undergraduate students and their study preferences to outline their study profiles. The social network is built based on an online survey created and distributed in 2019 to 717 undergraduate computer science students to investigate their learning patterns and behaviors. The students' social network is assortative and has a high clustering, common features in social networks. We discover five communities of students, where each of them is characterized by a different study profile.

The rest of this paper is organized as follows. In the next section, we discuss related research on learning analytics, educational data mining, social network analysis and study profiles. In Section 3 we present the methodology used in this research followed by the results in Section 4. The paper concludes with a discussion on the implications and limitations of our work and directions for future work.

2. Related work

2.1. Learning Analytics and Educational Data Mining

Learning Analytics (LA) and Educational Data Mining (EDM) have emerged as impactful research fields that draw on educational data in the last decades [4]. LA is defined as the "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" [4]. The implementation of LA strategies

HĮCSS

has been highlighted among the priorities of higher education institutions [5]. However, EDM is defined as "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in" [6], it employs data mining theories and techniques to analyze educational data. Both LA and EDM aim to improve and create methods that enhance education at all levels. They revolve around personalization, adaptive learning, predictive analysis and user behavior profiling [7, 8]. Furthermore, LA and EDM methods had been widely applied to address a large set of concerns, e.g., predicting students' performance, retention analytics, intelligent feedback provision and course recommendation [9]. Notwithstanding, limited research on learning styles like the personalization of learning, learning style identification and its application in teaching, learner motivation, and student profiling has been carried out [6].

Among the data mining techniques used in EDM, prediction methods, like classification and regression, and structure discovery methods, like clustering and factor analysis, are the most commonly used. Most recent research in EDM had been focused on the use of two or more methods [10]. Usual clustering and classification problems in EDM and LA can be extended to Social Network Analysis [11, 12]. However, Social Network Analysis has been used less frequently to examine educational data [10].

2.2. Network Science in educational context

Among SNA applications in the educational context, recent research has been focused on a wide range of areas of interest. Homophily is a fundamental property of social networks; it establishes that people with similar properties are more likely to connect [13]. Nguyen et al. [14] analyzed homophily regarding gender, ethnic minority identity, family income, and academic performance using WIFI log data. Their studies confirmed homophily concerning demographics and academic performance and showed that gender-based homophily increases over time. Shirvani et al. [15] applied SNA modeling techniques to analyze the social dimension and learners' roles on MOOCs discussion forums and their changes over time. Their research found that activity level can be predicted one week in advance based on the course structure, forum activity and properties of the communication network.

Community detection is one of the most significant problems in Social Network Analysis; the analysis of closely linked social groups is one of the comprehensible methods of describing social structures [11]. In higher education, community detection algorithms have been successfully implemented to address varied topics about learning processes. Sturludóttir et al. [16] identified fields of interest in the courses offered in undergraduate education programs. Xu et al. [17] analyzed discussion forum data of MOOCs courses to gain insights on the creation of social structures and how they change over time. Finally, Yassine et al. [18] used community detection algorithms to study users' engagement patterns on online learning networks.

2.3. Learning style, study patterns, and study profile

Learning style theories are used in an educational context to improve learners' learning strengths and instructors' teaching abilities. EDM methods had been used to investigate learning styles [6]. Ahmad and Tasir [19] used log files of online learning activities to analyze the behavior patterns of engineering students; they concluded that the course structure, students' previous experience, and subject influence the thresholds defined for learning style identification. Costaguta and Menini [20] studied the relationship between learning style and performance to improve group creation. More recently, Shobbrook et al. [21] implemented elements of EDM to validate the Fedler and Silverman's Index of Learning Styles (ILS) developed for engineering education. In their research, no correlation supporting the validity of the ILS was found, except for Lecture attendance. The research about learning styles had been controversial due to the limitations in measuring and determining the learners' learning styles individually [22].

EDM methods are also applied to investigate study patterns in varied contexts. Shirvani et al. [23] research presented a data-driven approach to identify and trace study patterns in an unsupervised manner and a hypothesis-driven approach to extract predefined patterns from learners' interactions. Casey and Azcona [24] used the student activity pattern for early detection of poor performers and to identify topics that the students found less interesting or more difficult to understand. Regarding using SNA to investigate study patterns, Leeet al. [25] analyzed click-stream data using hierarchical clustering analysis to identify behavior patterns concerning the use of a video discussion platform. They analyzed the transition pattern between consecutive activities in a video discussion platform.

Considering the difficulties related to the individual identification of learning styles, this paper applies

Social Network Analysis and community detection methods to analyze the undergraduate students' study profiles. For our purposes, we define the study profiles through a set of attributes related to educational values, goals, self-efficacy perception, and teaching methods preferences. Our approach relies on homophily to explore the characterization of the study profiles by analyzing the structure of the student community instead of focusing on the individual characteristics of each student.

3. Methods

3.1. Dataset

The data in this study were collected from an online survey distributed to 717 undergraduate students enrolled in the four computer science bachelor programs at Reykjavík University; BSc Computer Science, BSc Computer Science research-based, BSc Software Engineering, and BSc Discrete Mathematics and Computer Science. The survey was created and distributed in 2019, before the pandemic. It included 42 questions related to Institutional Support, Educational Values, Goals, Self-efficacy and Academic apathy, based on a students' readiness survey, the Academic Readiness Questionnaire [26]. The survey was initially distributed to study and understand undergraduate students' learning patterns and behaviors. The Cronbach's alpha coefficient of this questionnaire in this sample is 0.70, indicating good internal consistency reliability [27].

Among the 42 questions included in the survey, the first two questions asked for age and gender. Questions three to five were related to Institutional support; the students were asked about the amount of information they had regarding the university, their degree program and their career possibilities. In questions six to eight, the students' Educational values were evaluated; to measure the priority degree assigned to the university studies and the grades obtained. The following ten questions were Goal-related; in this section, the students were asked about their drivers for goal definition, organization, learning behavior, methodological preferences, and long-term expectations. The following twelve questions addressed the student's Self-efficacy perception; expected performance, skills and abilities, self-motivation, confidence and capacity of adaptation were addressed in those questions. In the next section, Academic apathy was measured in four questions asking for the student's effort, work avoidance, and scheduling level when planning their study sessions. The following seven questions were related to the university teaching methods, their preferences about attendance to lectures and practical sessions, as well as honors achievements. The last three questions were about the students' willingness to work in groups and their social network size. Except for age, gender and honors achievement, the questions' answers were on a 5 point Likert scale. Additionally, the students were asked to provide the list of students they most communicate with at the university; the maximum length of the list was 10. The survey's response rate was 38%, with 273 students answering it. Nonetheless, among those responses, only 218 students provided a list of connections at the university.

In the light of the COVID-19 pandemic and the sudden change to Emergency Remote Teaching (ERT) [28], and in line with recent research developed to analyze and understand its impact on teaching and learning processes, we decided to use the data obtained from the survey previously implemented to analyze the preferred study style of Computer Science students before the pandemic, to understand the impact of ERT in the undergraduate student community. Questions in Table 1 were selected to perform pre-pandemic study profile identification; these questions relate to the self-perception of motivation, adaptation to different teaching styles, preference to attend or not to lectures, and their willingness to work in groups instead of working alone. The list of students with whom the respondent communicated at the university was used to build a social network of students by creating a link between the respondent and everyone that they named. The data pre-processing was performed in R and R-studio, while the Social Network Analysis and community detection were performed in Python with NetworkX.

3.2. The Girvan-Newman Algorithm

The Girvan-Newman algorithm was used to identify the communities in the student network. This algorithm successively removes the edges with the highest betweenness as those edges tend to connect different clusters [11]. Betweenness is a centrality measure helpful to identify the most influential people in a social network. To calculate it, the times a node (edge) is crossed by the shortest path between any other pair of nodes in the network are quantified. The higher the betweenness coefficient, the more essential the node (edge) is to connect with the rest of the network [29]. The Girvan-Newman algorithm returns a set of partitions where each of them represents the clusters identified from the connected components after each edge is removed. As there is no natural benchmark

Table 1. Questions selected to perform study profile identification

Topic	Question/Statement
Individual Background	Age in years
Individual Background	Gender
Educational Values	Getting good grades is important to me
Goals	I'm a very methodical person
Self-Efficacy	I can easily adapt to different styles of teaching
Self-Efficacy	I can motivate myself to study when I need to
University teaching methods	I like the way of teaching (the methods) used at the university
University teaching methods	I do not usually attend lectures at the university
University teaching methods	I watch the lectures online, on Echo360 in Canvas, rather than attend class
University teaching methods	I always attend problem solving classes
Social Networks	I prefer to work in groups (arranged by the teacher), rather than work on my own
Social Networks	I prefer to work in groups (chosen by students), rather than work on my own

for the identified clusters, each partition returned by the algorithm was evaluated in its modularity to select the partition that maximizes it. The modularity coefficient compares the edges among nodes in the same cluster and the edges among nodes belonging to different clusters [13].

4. Results

4.1. Network Description

Nodes in the network represent a student who either answered the survey or was mentioned by someone who did. Directed edges were created from the student (source) who mentions another student (target). Questions' answer values were included as node attributes in the network. The final network displayed in Figure 1 includes 615 nodes with 806 edges. About 22% of the students who answered the survey did not provide a list of friends at the university. Those students, 59 in total, are included in the network as singletons representing 9.5% of the total nodes in the network. There are several reasons explaining the singletons: (1) the students do not have connections in the university, (2) the students do not feel comfortable sharing information about their connections, or (3) as the survey was not mandatory, the student skipped the last part of the questionnaire.

The density and the clustering coefficient are measures commonly used to describe the structure of a network. The density is defined as the fraction of connected nodes among all the possible pairs in the network. With a maximum value of 1, the higher the value, the more connected the network is [13]. The average clustering coefficient measures, on average, the extent to which the neighbors of each node in the network link to each other [30]. Both measures are helpful to outline characteristics of the
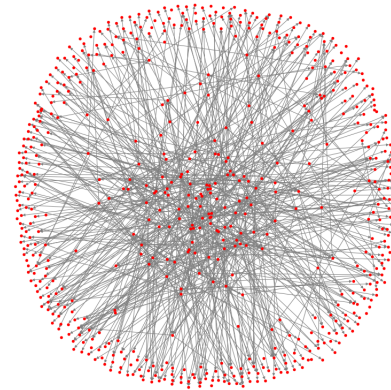

Figure 1. Friendship Network of undergraduate students.

network, such as its completeness and connectedness. The friendship network constructed has a density of 0.00213, and its average clustering coefficient is 0.1213. Networks with low density are told to be sparse; real-world networks are commonly characterized by sparsity [13]. Real-world networks with comparable densities coefficients were found in yeast protein interactions (2,277 links and a density of 0.001) and US air transportation data (18, 617 links and a density of 0.004) [13]. An assortative network is defined by [31] as networks with "a preference for high-degree vertices to attach to other high-degree vertices". The assortativity coefficient of a network is calculated as the correlation among the degrees of each pair of nodes in the network. Networks with positive coefficients are known as assortative, whereas negative values lead to disassortative networks [13]. With an assortativity coefficient of 0.24, we can say this network is assortative; students with many

connections tend to frequent other students with a high number of connections. The network has 94 connected components, the biggest with 358 nodes, while the smaller ones are singletons.
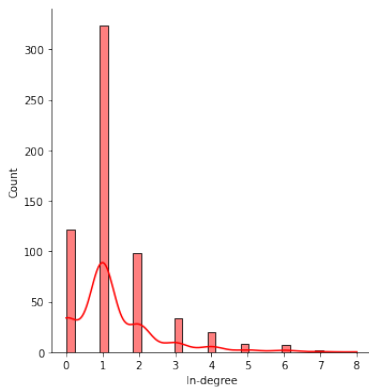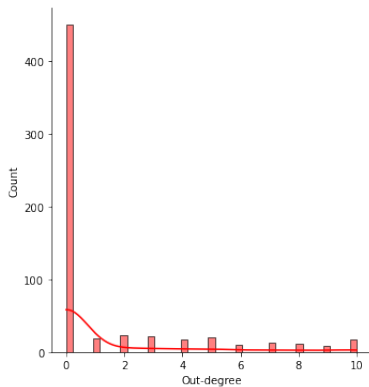

Figure 2. In-degree distribution.

Figure 3. Out-degree distribution.

Figures 2 and 3 display the in-degree and out-degree distribution of the nodes in the network. The in-degree value represents the number of times the student appears in others' list of connections, whereas the out-degree is the number of friends or connections declared by each student. The in-degree distribution is right-tailed; most students have a low in-degree (are mentioned by fewer people) than the maximum in-degree of the network who has an in-degree of 8. For the out-degree distribution, most of the nodes have an out-degree of zero. The reason for that is, among the 615 nodes in the network: (1) only 273 answered the survey, those nodes that were mentioned by someone but did not answer the survey will have out-degree zero; (2) among those who answered, 59 were singletons with no connections. Figure 4 displays the distribution when the zero out-degree nodes are not considered. It is remarkable the number of nodes with a high out-degree.
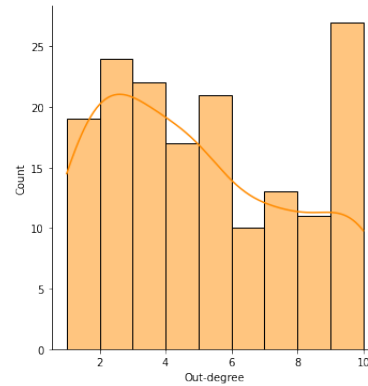

Figure 4. Out-degree distribution for degrees greater than zero.

4.2. Community Detection

The directed network was transformed into an undirected network before applying the community detection algorithms. With this change, the final amount of edges decreased to 739, and the average degree of the nodes is 2.40. Figure 5 displays the modularity coefficient for each of the partitions returned by the Girvan-Newman algorithm, the partition with the highest modularity, 0.89, has 110 communities. The number of communities is high due to the 59 singletons in the network; each singleton is a single community. We analyze the attributes of the five largest communities and present their preferred study profile. The communities were named 'Star pupils', 'Independent students', 'Team players', 'Female power', and 'Versatile students'. In those communities, 162 students are included. Figure 6 displays them colored by the community they belong to. 'Star pupils' is the only community disconnected from the others. In the following subsections, the study profile that characterizes the communities will be outlined based on the distribution of the answers to the attributes in the Academic Readiness Questionnaire [26] in Table 1. Figures 7 and 8 display the average and median of the responses to each question by converting the Likert scale into: Strongly Agree=5, Agree=4, Neutral=3, Disagree=2,and Strongly Disagree=1.

4.2.1. Community No. 1: 'Star pupils'. The first community identified includes 35 students. Among them, 69% are males, 20% female and 11% unknown. The response rate of the students in this community is 37%.The study profile is characterized by:
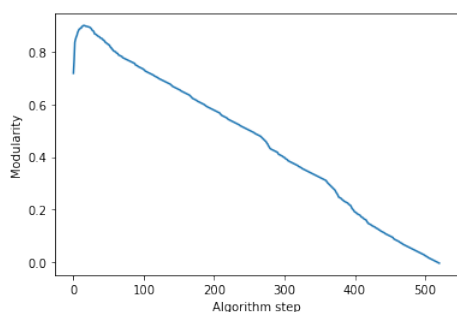
- Its members are mostly younger than 22 years.

Figure 5. Modularity coefficient in each step of the Girvan-Newman algorithm.
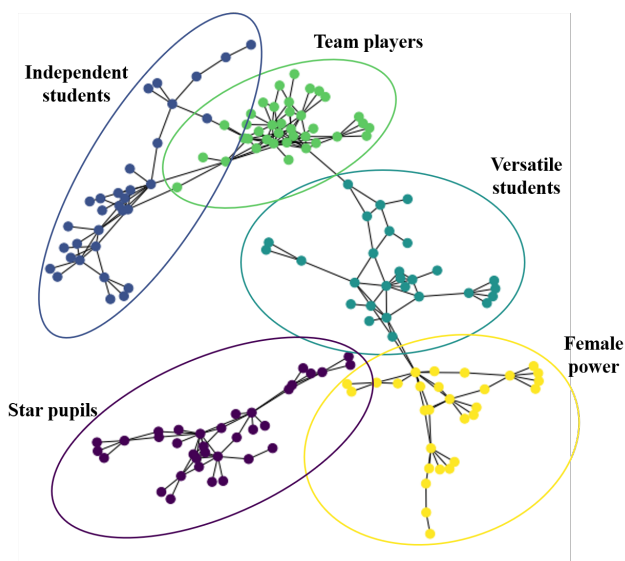

Figure 6. The five biggest communities identified using the Girvan-Newman algorithm.

- They declare that getting good grades is important to them.
- They are highly methodical, and they can highly adapt to different styles of teaching.
- They also declare being always able to motivate themselves when needed.
- These students also say they do like the teaching methods at the university.
- They always attend lectures, and they prefer to attend rather than to watch recordings. They always attend practical classes.
- These students prefer to work alone rather than working in groups arranged by the teacher.

4.2.2. Community No. 2: 'Independent Students'. This community includes 32 students. The gender distribution is quite different from the first community

analyzed; 56% males, 38% females, and 6% unknown. In this community, the response rate was 38%. The study profile of this community is featured by:

- Most of the students are around 25 years old, but the ages range from 23 to 38.
- Getting good grades is important, but they are not very methodical.
- They declare they could adapt to different teaching styles, but they mostly like the teaching methods at the university.
- What makes this community special is that they do prefer to watch the lecture recordings.
- Finally, this community prefers to work in groups chosen by themselves rather than working alone.

4.2.3. Community No. 3: 'Team players'. This community consists of 38 students. A 63% of them are males, 20% females, and 8% unknown. Similar to the previous communities, its response rate is 40%. Among its features is possible to identify:

- The students are around 23 years old, with ages from 20 to 26 years.
- Getting good grades is important, and they declared themselves to some extent methodical.
- They agree they can adapt to different teaching methods, and they mostly like those used at the university.
- The students usually attend lectures at the university and also problem-solving classes.
- In contrast to the other communities, these students do prefer to work in groups chosen by themselves rather than working alone.

4.2.4. Community No. 4: 'Female Power'. This community has 29 students. It is the only community with a higher percentage of females, 55%, whereas 34% are males and 10% unknown. Additionally, this community has the lowest response rate among the five communities analyzed, 24%. The features characterizing this community are:

- Their age goes from 20 up to 30 years with a uniform distribution.
- In this community, getting good grades is essential.
- They declare they could adapt to different teaching methods.
- In addition, they could prefer to work alone rather than in groups chosen by the teacher, but also prefer to work in groups rather than alone.

Figure 7. Average responses of the communities based on the Likert scale: Strongly Agree (5), Agree (4), Neutral (3), Disagree (2), Strongly Disagree (1).

4.2.5. Community No. 5: 'Versatile Students'. Twenty-eight students are allocated to this community. It is the community with the highest percentage of males, 82%, while 7% are females, and 11%unknown. The response rate was 42%. This community shares most of its features with the previous communities presented:

- Its members are around 27 years old.
- Grades are important, and they declare to be able to adapt to different styles of teaching.
- They declare themselves to be able to motivate themselves when needed and mostly like the teaching methods at the university.
- These students attend lectures almost always.
- Regarding group work, they prefer to work in groups chosen by themselves, or alone if the groups are chosen by the teacher.

4.2.6. Singletons. There are 59 singletons in the network, most of them have less than 30 years. The gender distribution is similar to the distribution of the whole network; almost 65% are males. Among their features; getting good grades is important to them, but there is no evident definition of being methodical when studying; they could adapt to different styles of teaching, motivate themselves when needed to, and they like the teaching methods at the university; regarding attendance to lectures, the distribution of the answers is uniform among the statements, but most of them agree on prefer watching the lecture recordings instead; finally,

regarding group working, there is no clear definition when the teacher chooses the groups, but they tend to slightly prefer to work alone rather than groups chosen by the students themselves.

4.3. Statistical comparison of attributes' distribution

Differences among the distribution of the communities' attributes were evaluated using the Kruskal-Wallis test. Only four attributes showed differences in their distributions considered statistically significant; (i) Question 4 in Table 1: *Being methodical*, (ii) Question 5 in Table 1: *Adaptation to teaching styles*, (iii) Question 8 in Table 1: *Lecture attendance*, and (iv) Question 9 in Table 1: *Watching recorded lectures*. Among those attributes, the community 'Star pupils' has significant differences in *Being methodical* for 'Independent students', 'Versatile students', and 'Female power' communities. Regarding *Lecture attendance* and *Watching recorded lectures*, a significant difference was found between 'Star pupils' and 'Independent students'. In contrast, for *Adaptation to teaching styles*, a significant distribution difference was found between 'Star pupils' and 'Female power'.

5. Discussion & Conclusion

This study presents five different study profiles among the undergraduate students in the Computer Science department at Reykjavík University. The data were gathered from an online survey distributed in 2019 as an initial approach to understanding undergraduate

Figure 8. Median of the communities' responses based on the Likert scale: Strongly Agree (5), Agree (4), Neutral (3), Disagree (2), Strongly Disagree (1).

students' learning patterns and behaviors. Relevant features of the student network in this university before the pandemic are; students with many connections tend to interact with other students who also have many connections, but also, the network has a high percentage of singletons, students without connections. As the second step in this analysis, the Girvan-Newman algorithm was used to identify the communities; the optimal partition was selected based on its modularity. The largest communities were analyzed to identify the features and learning preferences that characterize the study profile of its members. In the third section, we identified five communities with an evident and particular profile. (1) 'Star pupils' is featured by being those who always behave as expected and have many of the *best* attributes. Being the community with the youngest members (mostly less than 22 years), first-year students are likely allocated to this community and keeping most of the study habits they used to have at high school. (2) With students slightly older than the first community presented (around 23), 'Team players' members consider the group work an essential factor in their learning preferences, as long as they can choose their groups. (3) The third community, 'Independent students', has students around 25 years. This community is featured by preferring lecture recordings instead of always being at the university's venues. (4) 'Versatile Students' community has, on average, the oldest students (around 27 years). This community share features with the previous communities. From the first four communities, we can infer that the year of study and maturity level play an important role in determining the study style

profile and preferences. As the students move on in their undergraduate studies, they become able to adapt their learning strategy to the needs and requirements of each course, presumably more complex in the last terms of their studies, becoming a 'Versatile Student'. Last but not least, the fifth community, 'Female Power', is characterized by being the only community with more females than males. In line with the results obtained by [14] gender-based homophily is present. Nevertheless, besides gender, the attribute that features this community is how important getting good grades is. Regarding singletons, most of them prefer lecture recording, and as it could be expected, they prefer working alone; if the teacher or the students choose the groups, it does not matter.

The sample of the students examined in this paper falls under STEM, which stands for Science, Technology, Engineering and Mathematics. Within the literature on STEM, there has been an ongoing discussion on the issues related to students' sense of belonging. On that note, an extensive body of literature has focused on solutions targeted towards developing an increased sense of belonging, which is thought to lower the impact of identity-related issues on education; one such identity-related issue can be due to skewed distribution between genders in STEM [32]. This paper is a contribution to that literature through the five communities.

Among the limitations of this study, the data was gathered with an online survey. It does not allow linking the students with the average grade, the number of credits earned, or the year of study. That information could be helpful to analyze how the communities

evolve through the years, to what extent the students interact with and provide support to peers from other years, and how the study style selection relates to academic performance. Other drawbacks of this data collection method are the (i) response rate [33], affecting the performance of the Kruskal-Wallis tests due to small group sizes and (ii) response biases related to social desirability and the tendency to always select extreme ends in the Likert scale [34]. Regarding the sample used in this analysis, the students surveyed belong to the Computer Science department; students in other departments may have different study styles, so the identified communities' presence should not be generalized to students in other departments. This paper focused on the biggest communities; 30% of the students with at least one connection are part of the communities and profiles presented. Therefore, the other communities should be analyzed to identify if their study profiles are similar to those identified.

The results regarding the profiles of undergraduate students lead to relevant implications and future work. Before the pandemic arose, only one of the five identified study profiles preferred lecture recordings over the university venues. The students in the community 'Independent students' could have fewer difficulties during the ERT as they had previously interacted with the lecture recording platform. In contrast, most students were used to attending face-to-face lectures, and their transition to distance learning could be more complicated. Furthermore, the impact of the forced distance learning period could be bigger for the youngest students, 'Star pupils', as they highly prefer university venues. Another affected community was 'Team players', who highly valued interacting with their peers; during the pandemic, their interactions were limited due to the pandemic restrictions to contain the spread of the virus. Recent studies on the effects of the pandemic in education provide insights into how the new normality in Higher Education will be [35, 36, 37, 3]. On one side, the use of the technology and the transition to hybrid learning are expected to be present in the new normality of higher education [36]. In contrast, the impact of ERT on the students' experience with online learning may modify their preferences about enrolling in online courses in the future [35, 38]. The communities and study profiles analyzed in this study correspond to the pre-pandemic period. Consequently, once the pandemic finishes, it is essential to study and understand the changes in the learning profile and preferences of the students in the communities highly affected by the ERT. Furthermore, understanding to what extent the pre-pandemic communities' structure remained or not

after the pandemic will allow redesigning teaching strategies to provide better support to the students according to their specific profiles.

The past term, Spring 2021, ran still with restrictions due to the COVID-19 pandemic. Most of the schools around the world remained closed. The term Autumn 2021 will be, maybe, the first term of Education adaptation after the pandemic. Future work of this research relates to analyzing the undergraduate students' new connections, study profiles, self-efficacy perception, and goals during the adaptation to a new normality in higher education. In addition, other university departments and other data sources, such as data generated from LMSs or external tools, such as forum activity or recorded lectures, will be included to enrich the SNA and student communities' profiling. The integration of those data sources will allow linking the results obtained in this study with similar research performed using SNA on LMSs data, making our conclusions more generalizable. Finally, more research is also needed in investigating the evolution of students' social networks through their years of study at the university and how their modifications relate to their study profile and performance.

References

[1] R. Huang, J. M. Spector, and J. Yang, *Educational Technology*. Springer, 2019.

[2] A. Nguyen, L. Gardner, and D. Sheridan, *A Design Methodology for Learning Analytics Information Systems: Informing Learning Analytics Development with Learning Design*. 2020.

[3] N. G. López Flores, A. S. Islind, and M. Óskarsdóttir, "Effects of the covid-19 pandemic on learning and teaching: a case study from higher education," *arXiv preprint arXiv:2105.01432*, 2021.

[4] D. Gašević, S. Dawson, and G. Siemens, "Let's not forget: Learning analytics are about learning," *TechTrends*, vol. 59, no. 1, p. 64–71, 2015.

[5] M. Gaebel, T. Zhang, H. Stoeber, and A. Morrisroe, "Digitally enhanced learning and teaching in european higher education institutions," tech. rep., European University Association absl., 2021.

[6] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," *IEEE Access*, vol. 5, 2017.

[7] A. Peña-Ayala, "Learning analytics: A glance of evolution, status, and trends according to a proposed taxonomy," *WIREs Data Mining and Knowledge Discovery*, vol. 8, May 2018.

[8] M. Brown, "Learning analytics: Moving from concept to practice," 2012.

[9] S. Roy and S. N. Singh, "Emerging trends in applications of big data in educational data mining and learning analytics," in *2017 7th International Conference on Cloud Computing, Data Science Engineering - Confluence*, pp. 193–198, 2017.

[10] A. Aleem and M. M. Gore, "Educational data mining methods: A survey," in *2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 182–188, 2020.

[11] C. C. Aggarwal, *Data Mining*. Springer International Publishing, 2015.

[12] G. Deeva, S. Willermark, A. S. Islind, and M. Oskarsdottir, "Introduction to the minitrack on learning analytics," 2021.

[13] F. Menczer, S. Fortunato, and C. A. Davis, *A first course in network science*. Cambridge University Press, 1 ed., 2020.

[14] Q. Nguyen, O. Poquet, C. Brooks, and W. Li, "Exploring homophily in demographics and academic performance using spatial-temporal student networks," in *Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Fully virtual conference, July 10-13, 2020* (A. N. Rafferty, J. Whitehill, C. Romero, and V. Cavalli-Sforza, eds.), International Educational Data Mining Society, 2020.

[15] M. S. Boroujeni, T. Hecking, H. U. Hoppe, and P. Dillenbourg, "Dynamics of mooc discussion forums," in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, LAK '17, (New York, NY, USA), p. 128–137, Association for Computing Machinery, 2017.

[16] E. G. Sturludóttir, E. Arnardóttir, G. Hjálmtýsson, and M. Óskarsdóttir, "Gaining insights on student course selection in higher education with community detection," 2021.

[17] Y. Xu, C. Lynch, and T. Barnes, "How many friends can you make in a week?: evolving social relationships in moocs over time," in *EDM*, 2018.

[18] S. Yassine, S. Kadry, and M.-A. Sicilia, "Application of community detection algorithms on learning networks. the case of khan academy repository," *Computer Applications in Engineering Education*, vol. 29, no. 2, pp. 411–424, 2020.

[19] N. Ahmad and Z. Tasir, "Threshold value in automatic learning style detection," *Procedia - Social and Behavioral Sciences*, vol. 97, p. 346–352, 2013.

[20] R. Costaguta and M. de los Angeles Menini, "An assistant agent for group formation in cscl based on student learning styles," in *Proceedings of the 7th Euro American Conference on Telematics and Information Systems*, EATIS '14, (New York, NY, USA), Association for Computing Machinery, 2014.

[21] R. Shobbrook, P. Branch, and P. Ling, "Using educational data mining to test the validity of learning style theory," in *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pp. 490–496, 2020.

[22] P. A. Kirschner, "Stop propagating the learning styles myth," *Computers & Education*, vol. 106, p. 166–171, 2017.

[23] M. Shirvani Boroujeni and P. Dillenbourg, "Discovery and temporal analysis of mooc study patterns," *Journal of Learning Analytics*, vol. 6, no. 1, p. 16–33, 2019.

[24] K. Casey and D. Azcona, "Utilizing student activity patterns to predict performance," *International Journal of Educational Technology in Higher Education*, vol. 14, no. 1, 2017.

[25] S. Y. Lee, H. S. Chae, and G. Natriello, "Identifying user engagement patterns in an online video discussion platform," in *EDM*, 2018.

[26] J.-C. Lemmens *et al.*, *Students' readiness for university education*. PhD thesis, University of Pretoria, 2011.

[27] K. S. Taber, "The use of cronbach's alpha when developing and reporting research instruments in science education," *Research in Science Education*, vol. 48, Dec 2018.

[28] C. Hodges, S. Moore, B. Lockee, T. Trust, and A. Bund, "The difference between emergency remote teaching and online learning," 2020.

[29] L. Igual and S. Seguí, *Introduction to Data Science*. Springer International Publishing, 2017.

[30] A.-L. Barabási and M. Pósfai, *Network science*. Cambridge: Cambridge University Press, 2016.

[31] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, no. 20, 2002.

[32] R. Robnett, "The role of peer support for girls and women in stem: implications for identity and anticipated retention," *International Journal of Gender, Science and Technology*, vol. 5, no. 3, pp. 232–253, 2013.

[33] B. Blumberg, D. R. Cooper, and P. S. Schindler, *Business research methods*. McGraw-Hill Higher Education, 3 ed., 2011.

[34] *Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR), May 2017.

[35] E. M. Aucejo, J. French, M. P. Ugalde Araya, and B. Zafar, "The impact of covid-19 on student experiences and expectations: Evidence from a survey," *Journal of Public Economics*, vol. 191, p. 104271, 2020.

[36] A. Benito, K. Dogan Yenisey, K. Khanna, M. F. Masis, R. M. Monge, M. A. Tugtan, L. D. Vega Araya, and R. Vig, "Changes that should remain in higher education post covid-19: A mixed-methods analysis of the experiences at three universities," *Higher Learning Research Communications*, vol. 11, no. 0, 2021.

[37] S. Roy and B. Covelli, "Covid-19 induced transition from classroom to online mid semester: Case study on faculty and students' preferences and opinions," *Higher Learning Research Communications*, vol. 11, no. 0, 2020.

[38] K. O'Neill, N. Lopes, J. Nesbit, S. Reinhardt, and K. Jayasundera, "Modeling undergraduates' selection of course modality: A large sample, multi-discipline study," *The Internet and Higher Education*, vol. 48, p. 100776, 2021.