Gaining Insights on Student Course Selection in Higher Education with Community Detection

Erla Guðrún Sturludóttir*
Reykjavík University
Menntavegi 1
102 Reykajvík, Iceland
erlas13@ru.is

Eydís Arnardóttir
Reykjavík University
Menntavegi 1
102 Reykajvík, Iceland
eydis13@ru.is

Gísli Hjálmtýsson
Reykjavík University
Menntavegi 1
102 Reykajvík, Iceland
gisli@ru.is

María Óskarsdóttir†
Reykjavík University
Menntavegi 1
102 Reykajvík, Iceland
mariaoskars@ru.is

ABSTRACT
Gaining insight into course choices holds significant value for universities, especially those who aim for flexibility in their programs and wish to adapt quickly to changing demands of the job market. However, little emphasis has been put on utilizing the large amount of educational data to understand these course choices. Here, we use network analysis of the course selection of all students who enrolled in an undergraduate program in engineering, business or computer science at a Nordic university over a five year period. With these methods, we have explored student choices to identify their distinct fields of interest. This was done by applying community detection (CD) to a network of courses, where two courses were connected if a student had taken both. We compared our CD results to actual major specializations within the computer science department and found strong similarities. Analysis with our proposed methodology can be used to offer more tailored education, which in turn allows students to follow their interests and adapt to the ever-changing career market.

Keywords
Community detection, higher education, Louvain method, bipartite networks, student network, course selection

1. INTRODUCTION
University students enter higher education with a plethora of courses to choose from on their path to graduation. Gaining insight into student choices holds significant value for universities, especially those who aim for flexibility in their programs and those who wish to adapt quickly to changing demands of the job market. For example, the fast rise in popularity of machine learning over the past years could impel universities to make machine learning and related courses readily available to their students. In contrast, more subtler trends could be directly identified by the students' choices rather than an obvious shift in the job market.

Numerous studies based on questionnaires and surveys have found that there are various components that contribute to a student's course selection [2, 19, 20]. These are factors such as learning value, workload, age and academic performance [2]. Of these, the learning value of the course (which refers to factors such as intellectual level and interest in the topic) has been found to be the most influential factor in course selection. Course selection has also been a target in studies aiming to understand the gap between student mindsets and career demands [20]. Maringe [19] found that although intrinsic interest was important, course choices depend mainly on future career goals. According to the author, universities may need to adapt their strategies to the idea that students' course choices now seem to reflect their expectations of future employment rather than simply interests. Thus, universities would benefit greatly from a deeper understanding of the path their students choose towards their degree.

Educational data mining (EDM) has risen as a new field to answer these and other questions about students and their learning environment. It utilizes a variety of analytical methods and applies them to the vast amounts of data that has become available with increased digitization of administrative educational information. For example, EDM methods have already been applied to try to accurately predict college success using common classification algorithms with different feature sets [31]. They have also been used to analyze student clicking behavior in online courses to determine students' learning strategies and how those strategies can have an impact on their learning outcomes [1], as well as to predict student dropout [10]. One area of educational stud-

*EGS and EA contributed equally.
†Corresponding author.

Erla Guðrún Sturludóttir, Eydís Arnardóttir, Gísli Hjálmtýsson and María Óskarsdóttir "Gaining Insights on Student Course Selection in Higher Education with Community Detection". 2021. In: Proceedings of The 14th International Conference on Educational Data Mining (EDM21). International Educational Data Mining Society, 367-374. https://educationaldatamining.org/edm2021/
EDM '21 June 29 - July 02 2021, Paris, France

ies that has not received much attention is student course selection, despite its importance in understanding student interests and preparing them for a future career [28].

In this paper, we aim to reveal patterns in course selection through EDM, providing a new data-driven technique based on institutional analytics to gain insight into students' interests that would otherwise be difficult to discern. This knowledge can then be used for monitoring student interests and ensuring that courses reflecting those interests are available. We examine whether network analysis applied to students' course data, with a focus on community detection (CD), can effectively be used to identify university students' fields of interest. To accomplish this, we use a weighted projection network in combination with CD to explore student course selection. We focus on communities of elective courses for different majors and compare them to some of the official specializations the university already has to offer. Deeper understanding of students' choices is a stepping stone into allowing students to take more control over their studies, improve flexibility in the curricula, and facilitate students' pursuit of their interests.

2. RELATED WORK

A promising method for EDM is to represent educational data as networks. In general, networks consist of nodes and edges, where the nodes can for example represent people, countries or cells, and edges represent connections between nodes based on factors such as spatial and temporal proximity or social connections such as friendships [12, 8]. Network analysis is used to look at internal characteristics and the connections and patterns of nodes and edges, providing the ability to better understand the fundamental structure of networks and the real-life phenomena they model [29]. Different methods can be used to analyze networks, for example by looking at structural characteristics such as centrality, which indicates the importance of any given node in the network by assuming nodes that are more central have higher control over information passed through the network [8]. Community detection is another common way of analyzing networks which allows for the aggregation of different nodes into communities based on shared characteristics by identifying groups of nodes that have a high number of edges within themselves but fewer edges to other groups [12].

A common application of network analysis in educational settings is to understand social connections between students. This has helped reveal the negative effects of student interdependence in music education programs and its relationship to the program's friendship networks [26], as well as identifying how positive and negative friendship ties emerge [27]. Network analysis has also helped clarify the relationship between students' social networks and the development of their academic success [6, 14]. Furthermore, looking at students' social networks over time, close coequal communities are typically formed early on [30], although in some cases, students enhance their performance due to social relations outside their assigned group [24].

Although students' social networks have been studied, the exploration of students' course choices through network analysis has few precedents. Within the EDM field, Kardan et al. [16] used neural networks to predict course enrollment based on various factors such as course and instructor characteristics, and course difficulty. Further, Turnbull and O'Neale [28] used network analysis with CD and entropy measures to explore enrollment in STEM courses at the high school level. Among other results, they revealed that indigenous populations showed higher levels of entropy in their enrollment patterns, which was moderated by adolescent socioeconomic status. Neither of these studies focused on detecting student interests from course selection patterns.

3. METHODS

3.1 Data Source

Here, we use student and course data from Reykjavík University (RU). The university offers many different areas of study, including preliminary studies, undergraduate and graduate degrees. Most RU students are undergraduate students, and the RU undergraduate programs also offer the most variety of courses. Generally, the majority of RU undergraduate programs' courses are mandatory. These are the core courses each department decides is essential to their study program. The rest of the courses are either free choice electives, which can be any course in the university that the student qualifies for, or restricted elective courses from a selection tailored to the specific major.

We sample data from all graduated RU students that enrolled in the year 2014 or later and completed undergraduate programs in engineering, business, or computer science (CS) before 2021 (the total number of students was 1481). The university offers other programs as well, but we left them out since they have fewer students. The variables we look at include the student's registration ID and registration semester, the name and semester of each course a student has completed, and whether they passed or failed the course. We also include each student's department, major, and type of study (undergraduate, graduate, etc.).

To anonymize the data, we remove anything that could identify students, specifically their social security number and a numerical registration ID and give them a unique random sequence of numbers to replace both original numbers. For each student, we also remove any courses that they had de-registered from early in the semester. Further, for each major, courses taken by fewer than 5% of students are considered outliers and removed.

3.2 Network Analysis

3.2.1 Bipartite networks

We apply network analysis to the data to explore the fields of interests of RU students from a data driven perspective. Many real-world networks have a bipartite structure, where nodes belong to one of two groups or divisions and edges connect nodes of opposite groups without within-group edges [3]. In our bipartite network, the students make up one division of the nodes, and courses the other. If a student has taken a course, an edge is created between the respective nodes. Since edges represent that a student has taken a course, there is no edge between two students nor between two courses (see Figure 1, left).

Although bipartite networks give a more realistic and detailed representation of the system, analyzing them can be

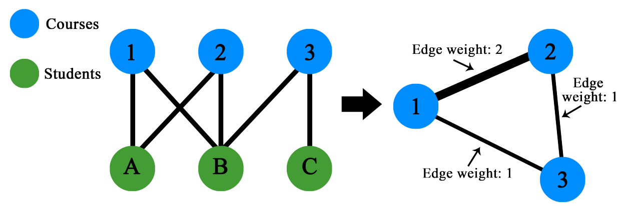

Figure 1: From bipartite network to weighted projected network. Left: a bipartite network, where the blue nodes represent courses and the green nodes, students. Right: a unipartite network has been obtained from the bipartite network, where the nodes are courses and the edges have weights that determine how many students have taken both courses.

complex. Therefore we project the bipartite network onto its unipartite counterpart (see Figure 1, right) [3]. This leaves a network with one type of nodes that can be analyzed with typical network methods. The resulting projected network consists of nodes representing the courses and edges between two nodes indicating that a student has taken both courses. We assign weights to the edges to represent the number of students who have taken both courses (see Figure 1).

A base problem with projection of bipartite networks is that a lot of important information in the original bipartite network is lost. Thus, we may end up connecting all courses in the network to each other –and form a clique– as long as they have at some point been taken by the same student, without taking into account how many students connected the two courses in the original bipartite network. Here, we address this by assigning weights to the edges in the projected network [3], where the weights represent the number of students who have taken both courses (see Figure 1).

3.2.2 Community detection

Building on the weighted projected networks, we use CD with the objective of inferring fields of interests in students' course selection. To identify fields of interest, we want to emphasise electives. However, in our data set, the information on which courses are mandatory and which are electives is incomplete. Mandatory courses along with very popular electives appear in the network as hubs, which usually occur in real-world networks as nodes with much higher degrees and edge weights than the other nodes [4]. We therefore define hubs in a data driven way, where a node is a hub if its total edge weight is at least one standard deviation above the mean edge weight of all nodes. We remove hubs from the network based on this definition.

Next, we apply the Louvain algorithm for CD [7]. This is an established, computationally efficient, fast converging method that produces accurate communities with high network modularity, especially in smaller networks [7, 17, 12, 23]. It has been successfully applied to identify communities of intrinsic brain systems [9], and to help create friend lists for Facebook users [18]. Modularity, is a measure of edge density within a partition (or proposed community) as opposed to edge density between partitions, whereby a higher modularity suggests a more cohesive community, separate from the others in the network. Importantly for our analysis using weighted projected networks, the Louvain algorithm can be used both with weighted and unweighted edges. The method starts by assigning each node to its own community [7], as seen in Figure 2. It then iterates over all nodes of the network and assesses the modularity gain obtained by assigning the node to the same community as each of its neighboring nodes. Next, the node is assigned to the community that yields the largest positive modularity gain, or maintains its current community if no positive modularity gain can be achieved by switching communities. This way, each new community assignment brings us closer to optimal modularity. The nodes are usually considered multiple times and the final iteration is determined when no switch leads to a gain of modularity, resulting in optimal partitioning of the network. This optimal partitioning is a local maxima, as the result is influenced by which node is considered first and the order in which nodes are visited. For some communities, we re-apply the Louvain algorithm for more detailed results, while using the inter/intra weight density ratio described below to ensure our communities maintain high quality.

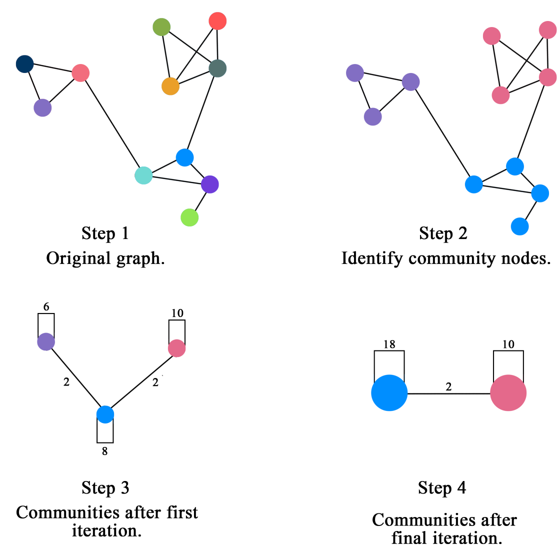

Figure 2: The Louvain algorithm. The first step of the algorithm is to assign each node to its own community. In step 2, a random node is selected to start the community aggregation process. All nodes are visited and allocated to the community of one of their neighbors or maintain their current community, depending on which choice gives the highest gain in modularity for the network. When no more modularity gain is possible in the network, step 3 is to aggregate the nodes of each community into new super-nodes. Here, the numbers given show the sum of node edges within and between supernodes. Steps 2 and 3 are then repeated until modularity has been optimized, as seen in step 4.

3.2.3 Community validation

Although the objective of CD is to split nodes into groups based on their connections within versus outside the group, there are many more aspects to consider [12]. One important factor is intra-cluster density, which refers to how many edges there are within the community as a ratio of how many possible edges there could be if all nodes of the community were connected to each other. This is contrasted by inter-cluster density, which shows how many edges go from the community to the rest of the network as a proportion of the maximum possible connections. High intra-cluster density may suggest a strong and cohesive community, however if it coincides with equally high inter-cluster density, it may simply suggest a strong and cohesive overall network.

To assess the quality of our communities, we use intra and inter weight density [13]. This is the same as intra and inter edge density previously described, but now accounting for weighted edges. The two are defined as follows:

$$WD_{\text{inter}} = \frac{w_C^{\text{ext}}}{\bar{w} n_C (n - n_C)} \text{ and } WD_{\text{intra}} = \frac{w_C^{\text{int}}}{\bar{w} n_C (n_C - 1)/2},$$

where w_C^{ext} is the sum of edge weights connecting the community to the rest of the network, or external community edges. We divide this by the estimated total edge weight of the network, which shows the edge weight going from the community to the rest of the network as a proportion of the maximum possible edge weight (assuming that the average edge weight of the fully connected network were unchanged). Here, $\bar{w}$ is the average edge weight of the network, n is the total number of nodes in the network and n_C is the total number of nodes within the community. Similarly, w_C^{int} refers to the sum of edge weights inside the community, which is divided by the expected total edge weight within the community. We then use a ratio of these two measures (WD_{inter} / WD_{intra}) to obtain the community strength on a scale where 0 is the strongest value, indicating a community that is disconnected from the rest of the network, and a value of 1 indicates a community equally connected within itself as to the rest of the network. We call this measure *density ratio* and use it not only to determine the community strength, but also to ensure that as we create smaller and more focused communities, community strength is not compromised.

3.2.4 Comparing communities and specializations

To further assess the real-world application of the communities we detect, we compare them to specializations within RU's Computer Science (CS) department, described in Table 4 in the Appendix. Any student who pursues an undergraduate degree in CS at RU has the option to graduate with a specialization in a certain field. The specializations do not need to be declared at enrollment but any student who fulfills the requirements can choose to add this to their graduation certificate. The specializations offered are Artificial Intelligence, Law, Web- and User Experience (UX) Design, Sports Science, Game Development and FinTech. Each specialization has 2-4 core courses that students need to complete, along with 1-3 courses from a pool of specialization-specific electives. Our approach to defining fields of interest is purely through data driven CD. Comparing the detected communities with these specializations helps validate the results and perhaps provide a reference for the creation of new specializations. We compare both the courses in each community and specialization, and the number of students belonging to a specialization versus those belonging to the corresponding community. We define a student as belonging to a community if they have taken at least 50% of the community's courses, with a special case of two course communities where both courses have to be completed.

3.3 Tools

Aside from the initial retrieval and anonymization of data, which we do using C# and SQL, all code for the data analysis was written in Python 3.9. We use multiple Python libraries to help with the data analysis. For our network analysis, we mainly utilize the NetworkX library [21]. For more general data manipulation, we use the pandas library [22]. We used Gephi for the majority of our network visualization [5], along with the Matplotlib library [15].

4. RESULTS

4.1 Communities that Reflect Interest Fields

We conducted CD with the Louvain algorithm on three undergraduate majors: engineering, business, and computer science. These majors have quite different program structures and emphases on electives, with the business major having the lowest number of elective courses allowed in their study plan (four electives). This is followed by the CS major with 11 electives and finally engineering, which offers only four free electives but nine "guided electives" (that is, nine electives must be specific to engineering), depending on the chosen engineering specialization.

We first look at the communities for the engineering department, see Figure 4 and Table 2 in the Appendix, which after hub removal consisted of 81 courses taken by 496 undergraduate students. Reykjavík University offers various undergraduate engineering programs such as biomedical engineering, financial engineering, and mechatronics engineering. These engineering majors all fulfill the same core courses in addition to some additional major-specific requirements. These majors are quite structured and offer few free elective courses. Due to the similarity in the core courses of these programs, we group them together into a more general engineering major. This means that the hub removal method removed general core engineering courses but leave most specialty-specific courses in the network. The resulting engineering network has 81 course nodes and 2614 edges. The weighted average inter/intra weight density ratio is 0.24. This suggests that hub removal was effective and the average community is relatively strong. The communities we have detected were eight in total as seen in Table 2. Note that communities are named after common characteristics between the majority of the courses, even though rarely all courses of a community fall within that definition. As expected, these communities mainly correspond to the official engineering majors such as financial, biomedical, and electrical engineering, with electrical engineering being our strongest community (WD_{inter} / WD_{intra} = 0.05). However, we also observe unrelated communities that supersede the official majors, such as a community of applied design and another for business related courses not mandatory in the financial engineering major. Courses in these communities are commonly taken together by engineering undergraduates, suggesting a common interest not credited to the specialized majors.

There are 334 undergraduate students in our data set who majored in business. For this major, the network consists of 36 course nodes and 504 edges, with a weighted average WD_{inter} / WD_{intra} of 0.25, again suggesting strong communities, see Figure 5 in the Appendix. This is not unexpected, as the business major only allows electives in the final year, giving business students less room to pursue distinct interests outside their core subjects. Table 3 in the Appendix shows the five communities identified within the business major. The strongest community is that of popular courses,

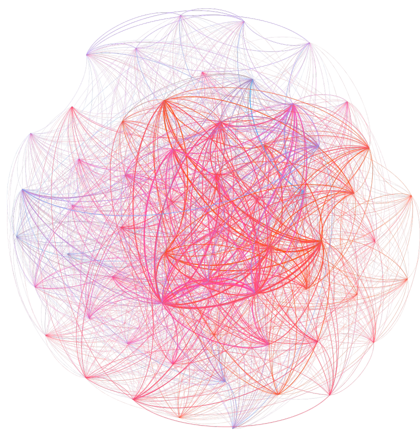

Figure 3: The network with communities for the BSc program in CS.

Table 1: Community detection results for BSc in CS.

	Community	No. courses	Density ratio
●	UX and Business	15	0.25
●	Engineering	13	0.17
●	Web and Software	10	0.20
●	Artificial Intelligence	7	0.39
●	Deprecated Courses I	6	0.08
●	Game Development	4	0.10
●	Deprecated Courses II	4	0.23
	Weighted average		**0.21**

which includes the most common electives in the business majors along with a handful of newer core courses (WD_{inter} / $WD_{\text{intra}} = 0.07$). These core courses were recently added to the study plan, meaning that they were only mandatory for a minority of the students in our data set. This is why these core courses were not identified as hubs and removed during hub removal. The business major also contains the weakest community of all the majors, management (WD_{inter} / $WD_{\text{intra}} = 0.71$). As the name suggests, this community includes various courses on management, such as service management and project management. The low inter and intra weight density ratio is interesting, as intuitively these courses would seem very connected. This is why measuring community strength is vital in determining the importance of the detected communities. The other business communities are both strong and reflect more specific interests, suggesting that there are students of the business major who actively seek distinct interests despite the program having no official specializations. The last major we explore is CS, with 377 students. Computer science has the least structured study plan of the three majors, as it puts a higher emphasis on unstructured flexibility and free electives. The CS course network consists of 59 nodes and 1492 edges. The communities (see Figure 3 are the strongest we found, with a weighted average WD_{inter} / WD_{intra} of 0.21. Most, but not all, detected communities seem to reflect an interest in a CS sub-field. However, the strongest community we have discovered was Deprecated Courses I (see Table 1), which represents older courses that may have been core courses at some point but are no longer being offered (WD_{inter} / $WD_{\text{intra}} = 0.08$). We conjecture that this community exists as some older students re-register to complete their undergraduate degree, for example after previously completing a CS diploma or taking a longer study break. It is therefore very intuitive that this specific sub-field is combined into our strongest community. Aside from communities based on deprecated courses, the other communities suggest that there is in fact an underlying pattern of interest fields present in the CS major, as observed for the other majors explored here.

4.2 RU Communities and Specializations

As a final validation of the communities we have detected for the CS undergraduate major, we now cross-reference our results with the actual specializations available for CS students. Unlike the other majors, CS offers a number of specializations meant to aid students in pursuing a specific subfield (see Table 4 in the Appendix for a short description of each specialization). However, only a subsection of students choose to do this. Of the students who graduated between 2014 and 2020, inclusive, only 9.5% fulfilled the requirements for a specialization. A further 13% partially fulfilled a specialization's requirements, by completing at least 60% of the specialization's core courses and 60% of the restricted electives needed.

Comparing the specializations and the communities we detected (shown in Table 1), we find interesting similarities. Our CD reveals that some communities are consistent with the specializations, but there is no absolute match. For the AI specialization (taken by 11 students, or 29% of those who graduated with a specialization), there is a partially corresponding community that includes both of the AI core courses (Artificial Intelligence and Machine Learning). There are 28 students who belong to this community, making it more popular than the official AI specialization. Although this community does not include any of the other courses from the specialization, it does include more theoretical and academically demanding courses than most other communities, suggesting a reflection of interest in theoretical computer science in general rather than specifically AI.

To fulfill the official AI specialization requirement, students must complete two core courses and three or more courses from a list of specialization-specific electives. However, in our data set most of these other electives were removed during either data cleaning (where we removed courses taken by fewer than 5% of students) or during hub removal and are therefore not part of any community. Interestingly, two of the remaining electives overlap between the AI specialization and that of Game Development. Both these courses have been sorted by our algorithm into a community that reflects Game Development much more strongly than AI, with 67 students. This is intriguing, as we know that students are much more likely to specialize in Artificial Intelligence than Game Development (only one student in our data set fulfills the requirements for Game Development), but this indicates that the gaming sub-field of Artificial Intelligence may be the biggest area of interest for these students.

The final specialization for which we discovered a similar community is Web and UX design, which was by far the most popular specialization taken by students (with 23 students, or 64% of all students who had a specialization). While this specialization encompasses both web programming and user experience, the corresponding community of Web and Software Development (with 84 students) is much more web than UX specific. Most of the UX related courses belong to a separate community of 21 students that unites UX and business rather than UX and web design. This suggests that dividing the Web and UX design specialization into two distinct specializations (Web design and UX design) might be more appealing to students. Interestingly, the remaining four official specializations have no corresponding community in our results. This was to be expected, as these remaining specializations are very rarely pursued by students. That is, the communities we have detected are able to represent the specializations that students are actually choosing, but did not reflect other specializations. This is exactly what we expect of CD, with the added bonus of identifying fields of interests that may not have been previously considered.

5. DISCUSSION AND CONCLUSION

With this project, we aimed to find whether CD could be used to effectively identify students' fields of interest at RU. To maintain the scope of the results, we have presented only the findings for undergraduate majors in engineering, business, and CS. Our resulting communities vary slightly in strength and size, yet almost all of them contain courses of a general theme that seem to indicate that they do in fact reflect fields of interest. This builds on the results found by Turnbull and O'Neale [28], who performed CD on a similar school course network, but without hub removal. This resulted in much more general course communities that demonstrated important but slight differences in the overall majors. In focusing on fields of interests, removing the hubs has allowed us to increase the granularity of the resulting communities while still maintaining community strength and cohesion. However, one of the commonalities between these majors is that the largest community detected usually included the major's most popular courses, be that electives or new mandatory courses our hub removal does not consider. As Fortunato [13] suggested, using the inter/intra weight density, we were able to evaluate the quality of the communities that were detected with the Louvain algorithm.

The communities we have discovered encapsulate various distinct areas of interests for the different undergraduate majors RU has to offer. Additionally, for the CS department, we have verified that the detected communities also reflect the main areas students choose to specialize in, which further validates our findings. To our knowledge, applying CD in this way and for this purpose has not been done before. This provides an exciting new tool for universities to better understand their students' aspirations.

In improving knowledge of student course selection, we provide academic institutions with more tools to increase study flexibility for their students. This knowledge can then be used to decide which courses the university wants to offer. This knowledge is also useful for academic counselors when helping students to discover their own field of interest. Based on previous studies, we assume that interest is the main motivation behind course choices [2, 19]. However, these communities may be based on other factors. Examining the characteristics of courses that make up different communities might reveal other factors that contribute to course selection, such as course difficulty, grading, teacher characteristics, and more [25, 2, 19].

Although we were able to successfully apply network analysis to our student and course data, there were a few setbacks. One drawback in our analysis is the fact that although RU's administrative data has largely been digitized, this has not always been done in the most structured and data-mining friendly way. For example, all information on specializations was retrieved directly from RU's website and formatted manually, as this information is not stored in the university's data warehouse. Reliable information on the mandatory courses of each major was also not available, which was why we decided to use data driven hub removal. Improving data availability, centrality and consistency is currently a priority at RU, but should also be considered by other universities wanting to take full advantage of EDM methods.

Our findings show that network analysis with CD is a useful tool in understanding students' course selection. The course choice patterns found here can still be explored further. For example, the current results are based on data from students who enrolled in the same program at different times. Thus any small changes in the program structure between years can introduce noise in the data. Looking at individual registration years, perhaps including a larger university with more students, could give clearer results. Further, it would be interesting to repeat the same analysis over separate periods to discover changes in interest fields over time. Finally, it was out of the scope of the current paper to analyze trends based on more detailed characteristics such as gender, age or grades. Augmenting the communities with these factors could for instance provide a tool to identify differences in choices made by students who graduate successfully and those who struggle more with their studies, perhaps yielding an opportunity for early intervention.

Educational data mining is an exciting new field with the potential to greatly influence educational institutions and their students going forward [11]. This project aimed to reveal how network analysis could be used to enhance student course selection by improved understanding of students' academic interests. Our analysis has successfully led to meaningful results that could easily be replicated by most interested universities with digitized information. Coupling this increased understanding of student interests with added academic support gives universities the tools to raise flexibility within majors while maintaining educational quality. Hopefully, this and other research in the field can be used to offer more tailored and student-led education, which in turns allows students to follow their interests and easily adapt to the ever-changing demands of the job market.

6. ACKNOWLEDGMENTS

We appreciate the contributions made by Kolbrún Eir Óskarsdóttir who had a big impact on the conceptualization and development of this research.

7. REFERENCES

[1] N.-J. Akpinar, A. Ramdas, and U. Acar. Analyzing student strategies in blended courses using clickstream data. In *Proceedings of the 13th International Conference on Educational Data Mining*, 2020.

[2] E. Babad and A. Tayeb. Experimental analysis of students' course selection. *British Journal of Educational Psychology*, 73(3):373–393, 2003.

[3] S. Banerjee, M. Jenamani, and D. K. Pratihar. Properties of a projected network of a bipartite network. In *2017 International Conference on Communication and Signal Processing (ICCSP)*, pages 0143–0147. IEEE, 2017.

[4] A.-L. Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.

[5] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks, 2009.

[6] D. Blansky, C. Kavanaugh, C. Boothroyd, B. Benson, J. Gallagher, J. Endress, and H. Sayama. Spread of academic success in a high school social network. *PLoS ONE*, 8(2):e55944, 2013.

[7] 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*, 2008(10):P10008, 2020.

[8] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca. Network analysis in the social sciences. *Science*, 323(5916):892–895, 2009.

[9] M. Cole, D. Bassett, J. Power, T. Braver, and S. Petersen. Intrinsic and task-evoked network architectures of the human brain. *Neuron*, 83(1):238–251, 2014.

[10] G. Deeva, J. De Smedt, P. De Koninck, and J. De Weerdt. Dropout prediction in moocs: a comparison between process and sequence mining. In *International Conference on Business Process Management*, pages 243–255. Springer, 2017.

[11] G. Deeva, S. Willermark, A. S. Islind, and M. Oskarsdottir. Introduction to the minitrack on learning analytics. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, page 1507, 2021.

[12] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.

[13] S. Fortunato and D. Hric. Community detection in networks: A user guide. *Physics Reports*, 659:1–44, 2016.

[14] N. Gitinabard, F. Khoshnevisan, C. F. Lynch, and E. Y. Wang. Your actions or your associates? predicting certification and dropout in moocs with behavioral and social features. *International Educational Data Mining Society*, 2018.

[15] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[16] A. A. Kardan, H. Sadeghi, S. S. Ghidary, and M. R. F. Sani. Prediction of student course selection in online higher education institutes using neural network. *Computers & Education*, 65:1–11, 2013.

[17] A. Lancichinetti and S. Fortunato. Community detection algorithms: A comparative analysis. *Physical Review E*, 80(5):056117, 2009.

[18] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference - IMC '11*, page 61. ACM Press, 2011.

[19] F. Maringe. University and course choice: Implications for positioning, recruitment and marketing. *International Journal of Educational Management*, 20(6):466–479, 2006.

[20] V. C. Milliron. Exploring millennial student values and societal trends: Accounting course selection preferences. *Issues in Accounting Education*, 23(3):405–419, 2008.

[21] NetworkX developer team. Networkx, 2014.

[22] T. pandas development team. pandas-dev/pandas: Pandas, Feb. 2020.

[23] X. Que, F. Checconi, F. Petrini, and J. A. Gunnels. Scalable community detection with the louvain algorithm. In *2015 IEEE International Parallel and Distributed Processing Symposium*, pages 28–37. IEEE, 2015.

[24] B. Rienties and D. Tempelaar. Turning groups inside out: A social network perspective. *Journal of the Learning Sciences*, 27(4):550–579, 2018.

[25] R. Sabot and J. Wakeman-Linn. Grade inflation and course choice. *Journal of Economic Perspectives*, 5(1):159–170, 1991.

[26] M. Sarazin. Can student interdependence be experienced negatively in collective music education programmes? a contextual approach. *London Review of Education*, 2017.

[27] M. A. Sarazin. Disliking friends of friends in schools: How positive and negative ties can co-occur in large numbers. *Social Networks*, 64:134–147, 2021.

[28] S. M. Turnbull and D. O'Neale. Entropy of co-enrolment networks reveal disparities in high school stem participation. *ArXiv*, abs/2008.13575, 2020.

[29] B. Wellman. Network analysis: Some basic principles. *Sociological Theory*, 1:155, 1983.

[30] Y. Xu, C. F. Lynch, and T. Barnes. How many friends can you make in a week?: Evolving social relationships in moocs over time. *International Educational Data Mining Society*, 2018.

[31] R. Yu, Q. Li, C. Fischer, S. Doroudi, and D. Xu. Towards accurate and fair prediction of college success: Evaluating different sources of student data. In *13th International Conference on Educational Data Mining*, 2020.

APPENDIX

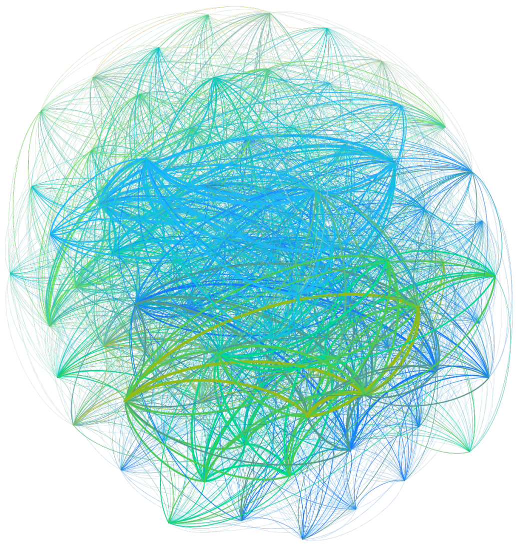

Figure 4: The network with communities for the BSc program in engineering.

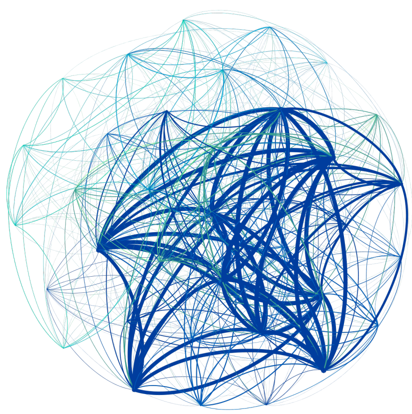

Figure 5: The network with communities for the BSc program in business.

Table 2: Community detection results for BSc in engineering.

Community	No. courses	Density ratio
Comp Sci and Mechatronics	25	0.37
Engineering Management	15	0.16
Finances and Management	10	0.25
Biomedical Engineering	10	0.10
Financial Engineering	9	0.21
Electrical Engineering	5	0.05
Applied Design	4	0.29
Business	3	0.32
Weighted average		**0.24**

Table 3: Community detection results for BSc in business.

Community	No. courses	$WD_{\text{inter}}/WD_{\text{intra}}$
Popular Courses	15	0.07
Management	6	0.71
Finance	6	0.29
Operations	5	0.10
Asset Management	4	0.36
Weighted average		**0.25**

Table 4: Official specializations in the CS program.

Name	Description
Artificial intelligence	Core courses reflecting an interest in AI and machine learning, with electives focused on game development and analytical skills.
Game design	Core courses encompass game development in general, computer graphics and game engine architecture. Electives reflect more general programming skills and AI.
FinTech	Both core courses and electives focus on the financial part of the Financial Technology discipline, as all students taking these courses gain software development skills from the core courses of the CS major.
Web and UX design	As the name suggests, most courses for this specialization directly relate to either web programming (such as the courses Web Programming II and Web Services) or user experience (User-Focused Software Development, Human-Computer Interaction).
Psychology	Core courses in psychology that emphasize cognitive processing and research methodology. Any other psychology courses can then be chosen as electives.
Law	General law courses with some emphasis on intellectual property rights and negotiations.
Sports science	General sports science courses.