Organizational Perspective on Academic Trajectories: Exploring STEM Programs through MMSNA

Gustavo Medina

Universidad Austral de Chile, Valdivia, Chile

Abstract. Low graduation rates and high dropout rates are continually part of the overall problems in STEM careers, usually associated with a variety of demographic and academic factors. In this study, we explore these issues from an organizational perspective, taking into account stakeholders' viewpoints. We leverage that by first working with database technologies, visualization tools, and statistical methods to make structures inside cohorts of STEM programs more visible and interpretable. The data is obtained from anonymized student records (2004–2015) from the Universidad Austral de Chile.
Initial interviews reveal known aspects from our literature review, while our initial SNA studies show us different realities from each STEM program and recurrent patterns associated with every engineering career. Further interviews with SNA visualizations provide deeper insight into student desertion and retention causes. However, the study highlights the need for additional research, particularly due to the non-intuitive SNA visualizations.

Keywords: dropout · MMSNA · academic trajectories. · learning analytics · STEM

1 Introduction

Students' academic trajectories have been a topic of interest for educators, researchers, and policymakers alike. However, an examination of these paths solely from an individual standpoint may overlook crucial contextual factors that shape students' educational experiences. To thoroughly understand students academic journeys in university, it is essential to adopt an organizational perspective that takes into account the wider educational environment they are immersed in. In this article, we propose a novel approach, leveraging the power of mixed-methods social network analysis (MMSNA), to explore and analyze students academic trajectories from an organizational perspective. This study aims to bring a new way to look at the complex student experience and dropout phenomenon by examining the interplay between students, their peers, teachers, and the institutional structures surrounding them, looking for factors influencing student outcomes, and inform evidence-based interventions for educational improvement.

In the literature, we can find different studies related to dropouts, from different perspectives of the university experience. One way to look at them is through three categories: factors related to students perceptions, the course and teachers, and those related to students demographic [1]. In the perception factors, we can consider aspects like unfulfilled expectations, insufficient motivation, mental health disorders, identity crises, substance abuse, and difficulty fitting in. In the course and teacher aspects, we can find academic demands, dissatisfaction with advising, language barriers, etc. Finally, in the demographic factors, we find aspects like age, gender, initial academic performance, and socioeconomic status [1] [2] [3] [4] [5].

According to Tinto [9], students who are capable of fully integrating themselves into the community have a greater likelihood of succeeding in the university. Homophily based on gender, age, and academic performance is a consistent finding in student networks [10], and the integration of students to their cohort have a positive impact on them, specially when they become sophomores [2]. As a result, comprehending the influence of relationships on students over time is crucial to elucidating the causes of both academic attrition and low graduation rates.

In order to study the integration of students into their cohorts, we use a social network analysis (SNA) approach, where relations and social structures are studied through networks and graph theory [14]. But, one aspect that this approach lacks is an understanding of the reasons behind the structures and patterns it finds [14]. For this reason, it is recommended to employ the mixed methods approach with the SNA study, in order to leverage a thorough understanding of our quantitative data through the mixing of qualitative studies [13] [14]. Examples of MMSNA studies include investigating national and foreign students in an organizational behavior class [15] and analyzing data from an online summer course [16]. One important aspect lacking in most studies is the inclusion of feedback from stakeholders to validate the "emergent" findings [17] [18].

In this study, we will analyze the data from undergraduate STEM programs at Universidad Austral de Chile (UACh) with the aim of describing the evolution of academic integration among students over time. Specifically, we will investigate the characteristics that distinguish students who drop out from those who successfully graduate, from an organizational perspective. To accomplish this, we will gather anonymized data from academic records and interviews with various stakeholders. These stakeholders are chosen to represent the majority of different perspectives inside the student context at all organizational levels, from directors of careers to secretaries to students.

Based on the research question and methodology described, we propose the following hypotheses for this study: Firstly, it is expected that there will be differences in social network structures among students enrolled in different academic programs. These differences could be attributed to cultural factors, such as varying levels of competitiveness or a stronger sense of community within certain programs. Secondly, we expect that students who face academic delays but establish stable social connections within their academic environment may have a lower likelihood of dropping out in comparison to those who struggle to form connections. Further analysis of the data collected in this study is necessary to test these hypotheses and draw meaningful conclusions.

This article consists of the following sections. First, a literature review that explores the existing research and theories relevant to the topic. The Methods sections follow, where we describe the data sample, the other standards we take into account, and finally, an explanation of the methodology and the principles followed. Then there is the results section, where we cover any important detail in the execution and reveal our findings. Finally, the discussion section summarizes the study's main findings, outlines its limitations, and concludes with a final statement.

2 Literature review

There are multiple factors that contribute to student dropout rates in engineering programs, including an unwelcoming academic climate, conceptual difficulties with core courses, a lack of self-efficacy or self-confidence, inadequate high school preparation, insufficient interest or commitment, and experiences of racism and/or sexism [8]. Additionally, economic factors such as lower perceived returns to tertiary education and insufficient integration into academic and social components of the university can also lead to dropouts [4]. Addressing these factors is crucial for improving retention rates and requires a comprehensive understanding of individual and organizational factors [4]. Inadequate preparation and understanding, poor course planning, and individual and institutional factors such as poor quality of instruction and mentoring, inadequate high school preparation, and an unwelcoming culture in engineering departments contribute to high dropout rates [19] [20].

Curricular analytics (CA) has emerged as a sub-field of Learning Analytics (LA), aiming to use evidence to drive curriculum decision-making and program improvement [21] [22]. CA uses data both at course-level and program-level, to analyze the interactions between students and the curriculum over time [21]. There are examples of CA research analyzing late dropout using Process Mining [24], which aims to extract knowledge from event logs obtained from information systems, in order to discover process models, verify conformance, and suggest improvements. Our case is similar, where the data is obtained from information systems, but with a focus on cohorts and relationships among students rather than a process-centric model in mind.

When implementing MMSNA research, there are several reasons for combining SNA and other methods, which can be categorized according to the approach as triangulation, complementarity, or development [25]. Triangulation involves comparing results from the different methods for convergence and divergence, while complementarity involves combining results from both qualitative and quantitative methods to address the complexity of the phenomenon [25]. In order to explain the overall methodology, it is useful to use the terminology of concurrent and sequential Quan (Quantitative) and Qual (Qualitative) combinations, representing the order and way of mixing the methods. For example, a triangulation approach using a concurrent Quan + Qual design could be understood as taking both quantitative and qualitative studies at the same time in order to compare the results at the end.

One set of good practices when formulating an MMSNA research project can be encompassed with the acronym PRICE [23]. These acronyms identify five themes: participatory, reflexive, integrated, critical, and ethical (PRICE). Each theme contributes to a comprehensive and dynamic characterization of MMSNA, with participatory approaches emphasizing active stakeholder involvement, reflexivity highlighting the complex relationship between knowledge production and context, integration promoting the formation of two approaches into a whole, and criticality promoting stakeholder empowerment and adherence to ethical standards [23]. These themes are significant as they provide guidelines for researchers to conduct MMSNA in a systematic and ethical manner while empowering stakeholders and producing comprehensive research outcomes.

3 Methods

This section explains the methodology used to investigate the causes of low graduation and high dropout rates in STEM programs at UACh. Our methodology takes as references the PRICE set of good practices, applying them to themes like ethical principles of data collection and the overall process that looks for the involvement and active participation of stakeholders in each stage of the study.

3.1 Data sources

The main data source for this study is a collection of anonymized student records from 2004 to 2015, compiled from the STEM programs of the Universidad Austral de Chile. The records include information such as student ID, gender, program, mathematics PSU (Chilean university selection exam), percentile rank in mathematics PSU, place of origin, year of admission, final status (early dropout, late dropout, graduation in time, graduation out of time), funding source, and income quintile. The study focuses on three engineering programs: Computer Engineering (ICI), Civil Engineering (OOCC), and Electronic Engineering (ICE), as we have more data on these programs.

The secondary data source for this study is a series of interviews with various stakeholders who are involved in the student dropout and retention phenomenon. These stakeholders include teachers, students, secretaries of careers, directors of schools of engineering, and other relevant actors at the organizational level such as psychologists and counselors. The interviews were recorded and transcribed, or taken on paper, with the initial permission and confirmation of every stakeholder. The aim was to elicit their opinions, perceptions, experiences, and suggestions regarding the causes and consequences of student dropout and retention in at least two stages of this study.

Regarding the results of this paper, they are not exhaustive of the material used but illustrative of the overall work and capacities of the approach formulated. We specifically center our final cases on those carriers that allow us to follow up in the long run and give us a good amount of information.

3.2 Data collection procedures

We made two principal phases in our study, where the first one was with a triangulation approach and a concurrent Quan + Qual model. In this way, we gather initial information that helps us formulate our second complementary approach phase, combining the results of the first phase with a qualitative study in a sequential way.

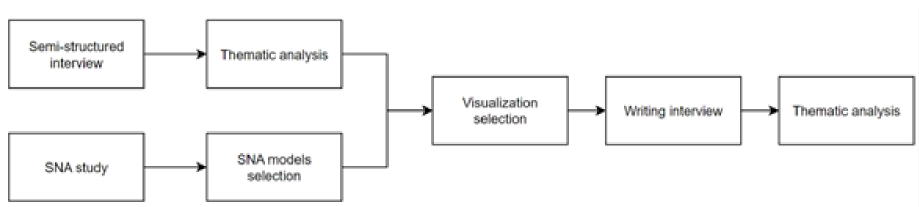

Fig. 1. Methodology approach

The initial semi-structured interview phase involved interviewing a sample of stakeholders from each group (teachers, students, secretaries, directors, etc.) using a set of open-ended questions that were designed to obtain their initial conceptions of the reasons for student dropout and retention. The questions also asked about their roles, responsibilities, expectations, challenges, and recommendations regarding student success in engineering education. The interviews were recorded and transcribed for later analysis.

In order to obtain the necessary quantitative data for our SNA analysis, we sourced it from the university data lake, which was then ingested into an Oracle SQL database management system. We leveraged Oracle SQL functions, views, and packages to query the data from the data lake and extract the specific information we needed. The resulting data was then transformed into CSV format, organized by career, and added to with every required characteristic for the SNA stage. It is worth mentioning that prior to the analysis, the initial data underwent an anonymization process to ensure privacy and confidentiality. Finally, we utilized the internal tools provided by Gephi to convert the CSV files into GEXF format, which is compatible with Gephi's graph analysis features. This multi-step approach enabled us to access and analyze the data effectively while maintaining data integrity and privacy. The second interview phase involved a writing assessment with the majority of previous and new stakeholders from each group, using a set of graphs and diagrams to illustrate the network structures and patterns selected in the SNA stage. The questions aimed to validate, clarify, explain, or challenge the results obtained from the first phase, as well as obtain their feedback and suggestions for improvement.

3.3 Data analysis methods

The SNA phase involved analyzing the quantitative data from the student records using database technologies, visualization tools, and statistical methods. The analysis aimed to test the hypotheses proposed in the introduction section as well as explore other possible relationships and patterns among the variables. At the same time, it involved constructing network graphs based on different criteria such as program, gender, place of origin, funding source, income quintile, mathematics PSU (Chilean university selection exam), percentile rank in mathematics PSU, final status (early dropout, late dropout, graduation in time, graduation out of time), etc. The network graphs were created using Gephi software and were used to measure and compare various network properties such as centrality, modularity, and homophily.

We formulated at least six different models of visualization and defined some restrictions for every case. The two principal restrictions were that every student needs to have a PSU score not equal to 0, in order to get a better visualization of PSU in general, and on the other hand, to only show a relationship between two nodes (students) if they have a weight equal to or adove 3 (meaning that two students need to have more than two classes together to show a relation in the graph).

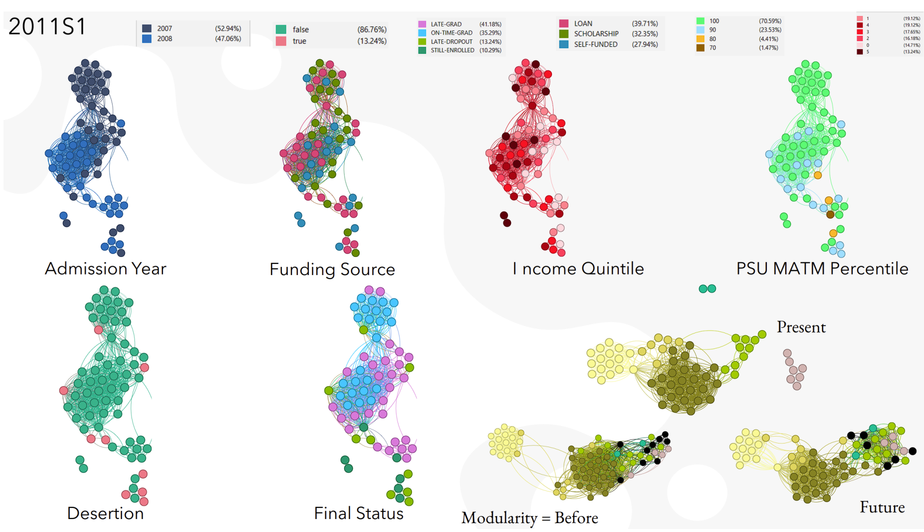

Fig. 2. SNA on the data

4 Results

In this section, we present the main results from each segment of the implemented MMSNA stages. Each phase is given its own subsection, where we detail pertinent findings. In particular, we've narrowed down the results from the first phase to align with those of the second phase. More extensive details, although not central to the main findings, will be available in the appendix.

4.1 Phase One: Thematic Analysis Results from Semi-Structured Interviews

Through the course of our study, qualitative data was accumulated from semi-structured interviews and examined through thematic analysis. To illustrate which stakeholders played a significant role in shaping each theme, we will initially present a table detailing their contributions. Following this, we will delve into each theme, outlining the overarching findings as they emerge from the interviews.

Table 1. Principal Informants for Each Theme

Theme	**Principal Informants**
Economic Factors	Secretary, Director, Psychologist
Benefits	Secretary, Director, Educational Psychologist
Performance Indicators	Director, Basic Sciences Teacher, Educational Psychologist
Course Failures	Basic Sciences Teacher, Educational Psychologist, Psychologist, Student
Social Skills	Psychologist, Educational Psychologist, Basic Sciences Teacher, Student
Entrance Scores	Director, Secretary, Basic Sciences Teacher, Student

Economic Factors: Participants frequently alluded to students' economic status as a critical aspect of academic performance. Notably, students from lower socioeconomic backgrounds faced multifaceted challenges, yet these were not always determinative of their performance or continuation in university.

Benefits: Participants perceived students admitted through the PACE program as being advantaged due to the tools provided to them. Loss of tuition-free education was identified as a significant event, particularly impacting lower quintile students.

Performance Indicators: First-year performance was frequently linked to a higher likelihood of university persistence. It was identified as an essential indicator of a student's capacity to handle academic challenges.

Course Failures: Course failure rates, especially in algebra and geometry during the first semester, were recognized as pivotal in forecasting further academic difficulties.

Social Skills: Participants affirmed that strong social skills positively impact university persistence by creating a comfortable and conducive environment. Conversely, students lacking such skills often reported distress and lower performance.

Entrance Scores: Entrance scores, although not always perfectly correlated, were recognized as a general indicator of a student's academic capacity and potential success in their chosen field.

Additional qualitative insights from the interviews that were not directly linked to our research questions but may be beneficial for understanding the broader context are included in the Appendix (see Appendix A).

4.2 Phase One: SNA selection methods and approach

In the development of this study, we created different models, trying to encompass the majority of visualization tools in Gephi, like the timeline for semester-to-semester filtering on data or different layout options like Force Atlas 2 or the ordered graph layout for a dimensional categorization of information. In general, we formulated at least four different types of visualizations, but in the end we decided to stick with a single approach to the data.

Our final visualization approach took into account the final purpose of integrating the results with the second phase of interviews, so it couldn't be too complex or simple in a way that conveyed minimal information. That's why the final decision was to use the Force Atlas 2 for the layout, with the lin-log option (which makes the clusters more tight) and the prevent overlap option, in order to get a better identification of singular nodes. At the same time, it was used with the Gephi Create Time Interval option to formulate a semester-to-semester study in a dynamic way. We decided to use a selection of two consecutive cohorts, in a way to see the evolution of relationships in the students through time, of those that are in their ideal semester courses, and those that get stuck in first or second semester courses.

Two specific restrictions were raised: not allowing students with PSU equal to 0 and the edge weight to be equal or more than three. The nodes without relationships were left out of the graphs before the layout. These decisions were made in order to reveal more relevant data in the graphs and to be consistent in our overall methodology.

We chose two models with four cases each, taking aspects like the knowledge of the stakeholders that we acquired in our first interview into account and the amount of information the visualizations should convey.

First Model: Fourth Semester Analysis Our first model emphasizes two aspects: the phenomenon of sophomore desertion and the shared and contrasting tendencies of desertion across STEM programs. We chose four cases to illustrate these aspects: OOCC 2010–2011, ICE 2011–2012, ICI 2004–2005, and ICI 2009–2010.

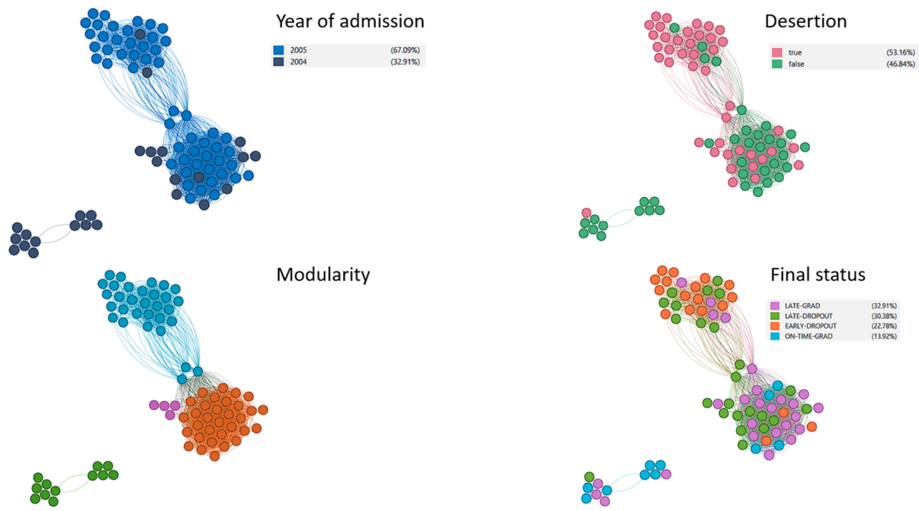

Fig. 3. Fourth semester of ICI 2004-2005

As can be seen in the example, for each case, we have four graphs. The first one, at the top left, shows the distribution of students according to their entrance cohort. In this case, 2004 students are sophomores, while 2005 students are in their first year of university. Then, at the top right, we have the desertion, where false (green) means that the student could graduate and true (pink) means that the student dropped out at a certain point in their career. The Final status graph at the bottom right shows if the dropout was early (first two years) or late (after two years) or if the graduation was on time or late (past the eleven semesters of the STEM careers selected). Another option in the final status is still enrolled, as seen in other cases (given that our database only shows information until 2015). The Modularity graph at the bottom left shows clusters defined by the Gephi community detection tool (called Modularity) and is an optional way to define clusters mathematically.

The graphical analysis reveals the separation of students according to their academic semester (more clear in some careers then others). Each cluster reveals different tendencies in terms of desertion and progression rate, bringing to light some key observations.

The first cluster (light blue in modularity) represents students (principally first-year students) taking their first academic semester courses (Typically failing to pass geometry or algebra), showing the highest desertion rates. Students stuck in their first semester during their second year, disconnected from their cohorts, show a 99 percent probability of desertion, mostly characterized as late dropouts.

The second cluster (orange and pink in modularity) represents students progressing normally through their first year. Although desertion is less common than in the first cluster, it remains a significant concern. A prominent trend is that sophomores lagging by two semesters present the highest desertion rate, despite their smaller proportion.

The third and fourth clusters (green in modularity) primarily consist of sophomores, most of whom persist in their university studies. Determining which is the third or fourth semester becomes challenging if it is completely separated from the second-semester cluster.

Second Model: PSU-Quintile-funding The second model of our analysis focuses on the relevance of PSU scores, quintile classification, and financial support in understanding student desertion during the initial three years of Informática 2009–2010.

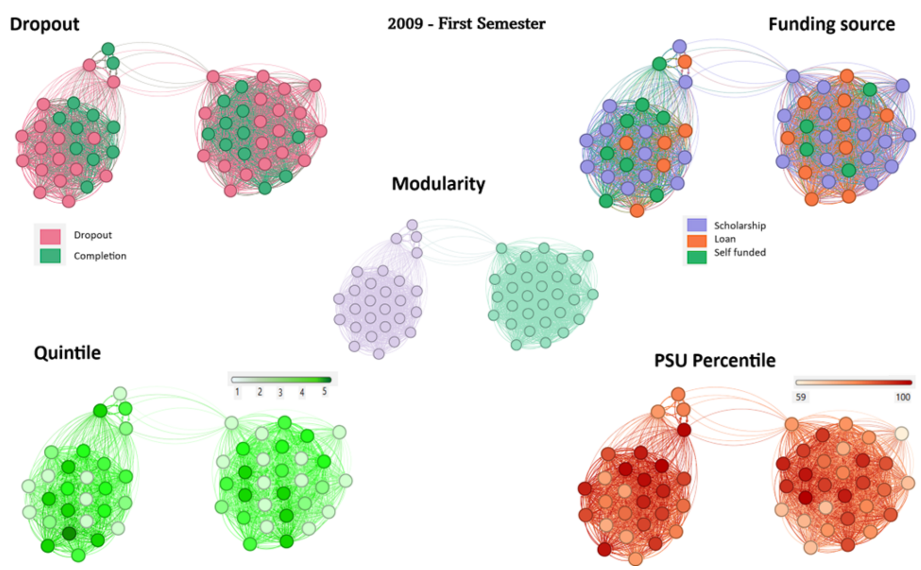

Fig. 4. First semester Informática 2009-2010

PSU scores, particularly in the first semester of 2009, indicate that students in areas with higher red concentrations (indicative of high PSU scores) tend to persist. However, by the second year, low PSU percentile students in the yellow cluster exhibit total dropout. Interestingly, high PSU score students also desert, emphasizing that a high PSU percentile does not necessarily guarantee persistence.

Regarding financial support, we find a high desertion rate among loan students in the first semester. At the same time, we find a high graduation rate for

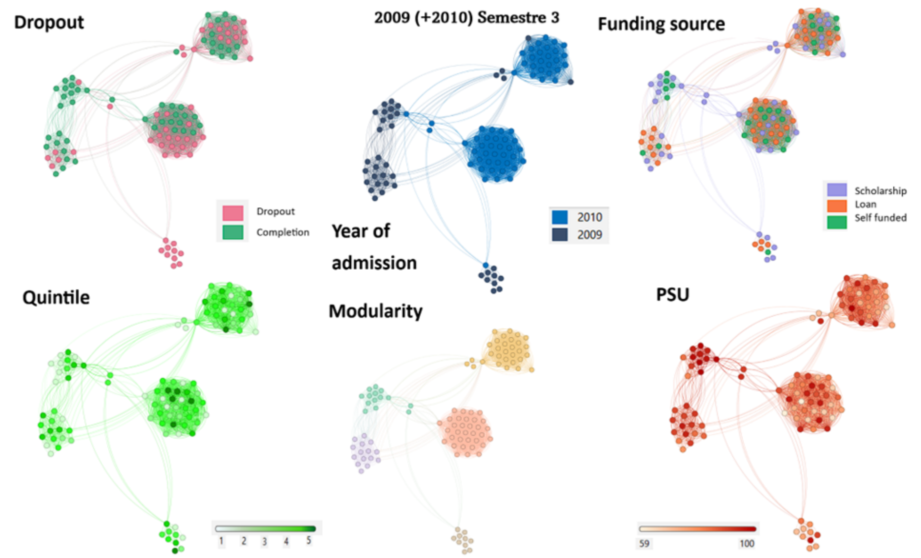

Fig. 5. Third semester ICI 2009-2010

self-funded students, even though they have a smaller proportion in contrast to scholarship or loan students.

Although the quintile may suggest higher PSU scores, this is not categorical. In the second year, we find that low PSU percentile students in clusters with high PSU concentration trends tend to persist, while high PSU percentile students in clusters with low PSU concentrations tend to desert.

Our models, therefore, reveal intricate patterns of student desertion and progression within STEM programs. The early identification of these patterns can provide important cues for taking preventive measures.

4.3 Phase two: Final writing assessment interview

In our final interview, we decided to change our previous semi-structured interview format, in part because we thought that the complexity of the insights couldn't be captured in a semi-structured interview form. This guided us to formulate a writing assessment where they could express and synthesize every idea and insight they could think of.

In general, we formulate a three-part writing assessment, with the first part being an exercise to introduce stakeholders to graphs and how to analyze them in our context. And then we presented each model with their cases for them to evaluate.

Part one: Introductory exercises Our objective in this introduction was to show the stakeholders a first glance of the potential cases that they would evaluate in an example graph with few nodes and relations, based on real cases but smaller in size.

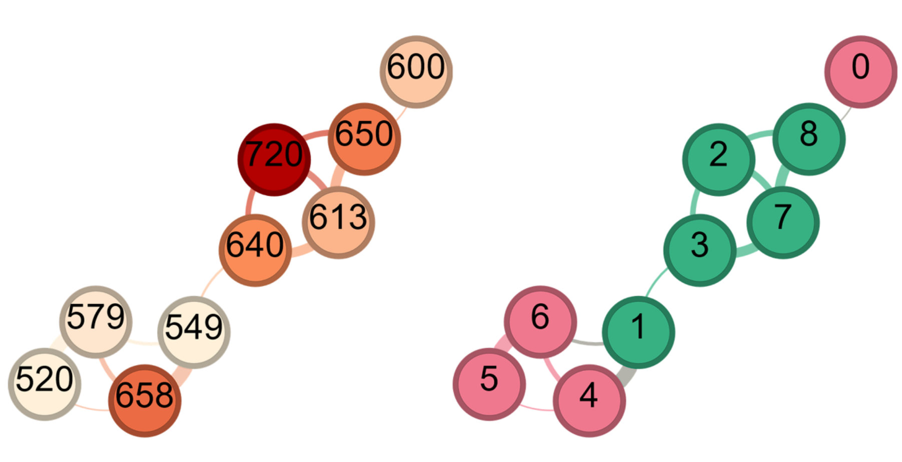

Fig. 6. Introductory example

What we showed them was a fake example of a third semester of a STEM program, with an explanation of the colors (green=graduation, pink=dropout, dark shade of red = high PSU, light shade of red = low PSU), the weight of the relations (how many courses two students take together), and the number over them (In the left is the PSU score, and in the right is the ID of a student).

Firstly, they needed to identify clusters, characterize them, and then interpret what this means in real-life scenarios. In the first graph of desertion (the one on the right), we proposed three questions:

1. How many clusters (groups with high interrelation) do you find in the graph?
2. What can you say about each cluster?
3. What do you think of cases 1 and 3? What about the one with ID 0?

Then for the PSU graph (the one on the left), we try to make them realize how to compare both graphs with the color differences in a way to better understand the reality of each student's. For this graph, we formulated another three questions:

1. How does it contribute to your previous findings or conclusions?
2. What can you say about each cluster?
3. What speculation can be made regarding node ID 4? Does it remind you of any real-life cases you have encountered?

There were two stakeholders with little knowledge of this kind of study, but in general, stakeholders could make use of these examples to delve deeper into how to obtain information from these visualizations and interpret them in their context as students of STEM programs. Finally, it proved to be essential for many of them to address real-life scenarios.

First Model: Fourth Semester Analysis

Intercluster Connectivity and Dropout: A basic science teacher and an educational psychologist observe that students who serve as bridges between two clusters tend to drop out. These students might focus more on major-specific courses, neglecting the basic science courses.

Descriptive Patterns of Deserter Groups: A basic science teacher highlights that in ICI, it is not clear which case has a higher dropout rate compared to OOCC. In ICE, there is isolation between advanced and delayed students, and clusters related to desertion can be observed. Also highlights that in some cases, modularity proves to be more useful.

Variability in Retention: An ICI Teacher states that there seems to be more variability in 2009–2010 compared to 2004–2005. In OOCC, there is less delay, and in ICE, the dropout rate is clearly seen among delayed students.

It also describes specific retention patterns for different cohorts, such as the 2010–2011 OOCC group, which advances together with fewer dropouts.

Student Characteristics: The ICI Secretary discussed the characteristics of students according to their career and the common conditions that prevent the continuity of studies.

An educational psychologist identifies different patterns within cohorts, like the 2005 ICI students divided between repeaters and those who pass.

Impact of Study Habits and Course Burden: An educational psychologist suggests that smaller groups with fewer subjects may have less desertion because students can focus on the subjects they can pass.

An ICI secretary points out the importance of dedicating time to studies rather than distractions.

Social Factors and Personal Characteristics: An ICI secretary raises issues related to personal characteristics and social factors, such as a lack of tolerance for frustration, conditions like Asperger's, ADHD, and ASD that can hinder continuity of studies, addiction to games, and the need to balance work and studies.

Second Model: PSU-Quintile-funding

Group Separation and Heterogeneity: A basic science teacher and the director of the pedagogy center of basic sciences (CDCB) both observe natural separation into heterogeneous clusters in the first and third semesters.

Impact of Financing: An ICI teacher and two educational psychologists highlight the influence of financing on student abandonment. Students with credit financing tend to abandon school more quickly.

An ICI teacher and an educational psychologist mention the role of income and scholarships. Students from higher-income backgrounds or from lower-income backgrounds with scholarships tend to have better emotional and psychological handling and prior academic preparation.

Relationship between PSU Scores and Retention: An ICI teacher, a basic science teacher, and two educational psychologists discuss the correlation between PSU scores and student retention. Higher PSU scores indicate better persistence in some cases.

Abandonment Patterns: An educational psychologist notes a group of repeaters who abandon the third semester.

Academic Preparation and Study Habits: An ICI secretary and an educational psychologist emphasize the importance of academic preparation and study habits. Students with better academic preparation and study habits tend to persist, regardless of financing or PSU scores.

Suggestions for future investigation Some stakeholders show concern for different aspects, like the lack of importance in the axis in our graphs, which could be used to gain stakeholders space intuition.

We find that there is a tendency for stakeholders related to mathematics, from teachers to students, to understand the graphs more quickly than those related to administration or counseling areas. We couldn't comprehend the reason behind it, even though the final conclusions of both of them were good, regardless of the initial complexity. This could be something to take into account in future studies, especially when selecting the visualizations for the interviews.

Finally, there is a need for solid visualizations with standard sizes, places, and colors for each graph. These differences added complexity to our visualizations. In general, the average time for the writing interview was 1:30 to 2 hours, which hopefully could be improved by better visualization standards.

5 Discussion

This study presents an innovative methodological approach, employing mixed methods social network analysis (MMSNA), to explore the complex reality of students trajectories within STEM programs, especially in cases of dropout. The aim of this methodology is to bridge the gap between complex graph theory concepts and diverse stakeholders in a way to enable communication and interpretation of the multifaceted nature of student retention and dropout phenomena with powerful tools of data exploration. Through this investigation, we focus on refining, simplifying, and validating our proposed MMSNA approach, especially through the creation of intuitive visual representations of network data. The involvement of stakeholder groups like educational psychologists, career secretaries, institutional directors, and educators from the University of Austral de Chile played a significant role in this process, reaffirming the practicality and effectiveness of MMSNA in our educational context. The conclusive outcomes of our research, namely the Fourth Semester Analysis and the PSU-Quintile-Funding Model, exemplify the efficacy of this approach, providing comprehensive and accessible insights into student progression patterns within STEM education.

We formulated different models of visualization that reduce the gap between complex graph theory concepts and the different stakeholders, like educational psychologists, secretaries of careers, directors of institutions in STEM programs, and different teachers at Universidad Austral de Chile. This task proved to be of utter importance in the overall thesis, given that our final results were based on the stakeholders' perceptions of our SNA visualizations. We emphasize the importance of finding a balance between complex and simple visualizations in a way that shows as much information as intuitively as possible. Our final models were the Fourth Semester Analysis and the PSU-Quintile-Funding Model.

The Fourth Semester Analysis shed light on the distribution of students based on their progression through the semesters. The clusters identified in this analysis unveiled a range of dropout and progression tendencies among students, signifying that the year of study and course progression are critical elements to consider in dropout prevention strategies. Notably, students stuck in their first semester during their second year exhibited a high probability of dropping out. Future interventions could target these students specifically, providing them with additional support to enhance their progression and retention.

The PSU-Quintile-Funding model underscored the relevance of previous preparation and financial support for students. We find that PSU scores were relevant only in the first year but that their significance diminished over time when trying to understand dropout. This finding suggests that the impact of entry-level academic capabilities is essential to their initial advancement, but after the first year, other factors, such as their actual academic semester or their external academic support, become more pertinent.

Students with loan financing tended to drop out more quickly (before their third year), highlighting the potential stress and pressure associated with financial concerns. There is a great concern among students who have to sustain themselves financially and tend to dropout. Universities might consider providing additional financial guidance and support to these students to improve their retention rates.

The insights obtained from our study highlight important improvements that could be informed by the university to provide better policies and practices aimed at improving student retention, particularly in STEM programs. However, it is important to consider that the present study only takes into account data from 2004 to 2015 and that actual events, like COVID, and changes in educational aspects, like the new PAES national test (replacing PSU), reinforce the need for an update of the data for future studies.

In conclusion, the present study demonstrates the utility of MMSNA in elucidating complex patterns of student dropout and retention in STEM programs. Further research is needed to take on more actualized data and to explore more STEM programs and other academic programs.

6 Acknowledgments

This study was conducted using anonymized data, provided by Universidad Austral de Chile.

A Additional Interview Insights

Other themes obtain through the interviews, but not related to the overall findings of this essay.

A.1 Demographics

Student's place of origin influences their adaptation to new environments, affecting academic performance. The provenance is a key factor, with students from higher-income schools tending to perform better. The place of origin is not necessarily a significant indicator itself, with other factors like daily commute and climate variations also impacting the academic outcome.

A.2 Family

A strong support network of friends and family decreases the likelihood of student dropout. Family factors, including socioeconomic level and internal dynamics, significantly influence academic performance. Dropout is a complex, multi-factorial phenomenon, with problematic family relationships identified as a notable contributing factor.

A.3 Means of Aid

The percentage of students participating in tutorials and assistant classes might have a correlation with academic improvement. Further data required for in-depth analysis.

A.4 Miscellaneous

A reasonable dropout rate is below 5 percent, but it has risen during the pandemic. High student retention is a problem, as it affects the completion of the curriculum. Anomalies like 'ghost students' - those who enroll but never attend classes - skew the data and contribute to a misleading dropout rate.

A.5 Mental Health

The prevalence of mental health problems, including high stress levels, can significantly affect academic performance, with some students maintaining a high failure rate throughout their academic career.

References

1. Santos, R., Ponti, M. & Hora Rodrigues, K. The Use of Digital Reports to Support the Visualization and Identification of University Dropout Data. *Human Interface And The Management Of Information: Visual And Information Design: Thematic Area, HIMI 2022, Held As Part Of The 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part I.* pp. 308-323 (2022)
2. Mauldin, R., Barros-Lane, L., Tarbet, Z., Fujimoto, K. & Narendorf, S. Cohort-based education and other factors related to student peer relationships: A mixed methods social network analysis. *Education Sciences.* **12**, 205 (2022)
3. Meyer, M. & Marx, S. Engineering dropouts: A qualitative examination of why undergraduates leave engineering. *Journal Of Engineering Education.* **103**, 525-548 (2014)
4. Aina, C., Baici, E., Casalone, G. & Pastore, F. The determinants of university dropout: A review of the socio-economic literature. *Socio-Economic Planning Sciences.* **79** pp. 101102 (2022)
5. Höhne, E. & Zander, L. Belonging uncertainty as predictor of dropout intentions among first-semester students of the computer sciences. *Zeitschrift Für Erziehungswissenschaft.* **22**, 1099-1119 (2019)
6. Salas-Morera, L., Molina, A., Olmedilla, J., García-Hernández, L. & Romero, J. Factors affecting engineering students dropout: A case study. *The International Journal Of Engineering Education.* **35**, 156-167 (2019)
7. Sithole, A., Chiyaka, E., McCarthy, P., Mupinga, D., Bucklein, B. & Kibirige, J. Student attraction, persistence and retention in STEM programs: Successes and continuing challenges.. *Higher Education Studies.* **7**, 46-59 (2017)
8. Geisinger, B., Raman, D. & Raman, D. Why they leave: Understanding student attrition from engineering majors. (2013)
9. Tinto, V. Dropout from higher education: A theoretical synthesis of recent research. *Review Of Educational Research.* **45**, 89-125 (1975)
10. McPherson, M., Smith-Lovin, L. & Cook, J. Birds of a feather: Homophily in social networks. *Annual Review Of Sociology.* **27**, 415-444 (2001)
11. Crossley, N. & Edwards, G. Cases, mechanisms and the real: The theory and methodology of mixed-method social network analysis. *Sociological Research Online.* **21**, 217-285 (2016)
12. Crossley, N. Towards relational sociology. (Routledge,2010)
13. Tashakkori, A. & Creswell, J. The new era of mixed methods. *Journal Of Mixed Methods Research.* **1** pp. 3-7 (2007)
14. Froehlich, D., Rehm, M. & Rienties, B. MMSNA: An introduction of a tale of two communities. *Mixed Methods Social Network Analysis.* pp. 1-9 (2019)
15. Rienties, B., Héliot, Y. & Jindal-Snape, D. Understanding social learning relations of international students in a large classroom using social network analysis. *Higher Education.* **66** pp. 489-504 (2013)

16. Rienties, B., Tempelaar, D., Bossche, P., Gijselaers, W. & Segers, M. The role of academic motivation in Computer-Supported Collaborative Learning. *Computers In Human Behavior*. **25**, 1195-1206 (2009)
17. Toraman, S., Cox, K., Clark, V. & Dariotis, J. Graduate Students' Current Practices for Writing a Mixed Methods Research Study Abstract: An Examination of Doctoral Dissertation and Master's Thesis Abstracts in the ProQuest Dissertations and Theses Global Database.. *International Journal Of Multiple Research Approaches*. **12** (2020)
18. Rienties, B. Critical reflections and moving forward. *Mixed Methods Social Network Analysis: Theories And Methodologies In Learning And Education*. (2019)
19. Salas-Morera, L., Molina, A., Olmedilla, J., Garcıa-Hernández, L. Romero, J. Factors affecting engineering students dropout: A case study. *The International Journal Of Engineering Education*. **35**, 156-167 (2019)
20. Meyer, M. & Marx, S. Engineering dropouts: A qualitative examination of why undergraduates leave engineering. *Journal Of Engineering Education*. **103**, 525-548 (2014)
21. Hilliger, I., Aguirre, C., Miranda, C., Celis, S. & Pérez-Sanagustın, M. Design of a curriculum analytics tool to support continuous improvement processes in higher education. *Proceedings Of The Tenth International Conference On Learning Analytics Knowledge*. pp. 181-186 (2020)
22. Pinnell, C., Paulmani, G. & Kumar, V. Curricular and learning analytics: A big data perspective. *Big Data And Learning Analytics In Higher Education: Current Theory And Practice*. pp. 125-145 (2017)
23. Onwuegbuzie, A. The PRICE of mixed methods social network analysis: Toward an ethical process for MMSNA. *Mixed Methods Social Network Analysis*. pp. 245-262 (2019)
24. Salazar-Fernandez, J., Sepúlveda, M., Munoz-Gama, J. & Nussbaum, M. Curricular analytics to characterize educational trajectories in high-failure rate courses that lead to late dropout. *Applied Sciences*. **11**, 1436 (2021)
25. Toraman, S. & Clark, V. Reflections about intersecting mixed methods research with social network analysis. *Mixed Methods Social Network Analysis*. pp. 175-188 (2019)
26. Creswell, J. & Clark, V. Designing and conducting mixed methods research. (Sage publications,2017)
27. Leech, N. & Onwuegbuzie, A. Guidelines for conducting and reporting mixed research in the field of counseling and beyond. *Journal Of Counseling Development*. **88**, 61-69 (2010)
28. Tashakkori, A., Johnson, R. & Teddlie, C. Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences. (Sage publications,2020)
29. Morse, J. Mixed method design: Principles and procedures. (Routledge,2016)