Computers and Education: Artificial Intelligence 8 (2025) 100430

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Predicting student achievement through peer network analysis for timely personalization via generative AI

Ivica Pesovski [a,*], Petar Jolakoski [a,b], Vladimir Trajkovik [c], Zuzana Kubincova [d], Michael A. Herzog [e]

[a] Brainster Next College, Vasil Gjorgov 19, Skopje, 1000, North Macedonia
[b] Research Center for Computer Science and Information Technologies, Macedonian Academy of Sciences and Arts, Krste Petkov Misirkov, Skopje, 1000, North Macedonia
[c] Faculty of Computer Science and Engineering, Ss Cyril and Methodius University, Ruger Boskovik 16, Skopje, 1000, North Macedonia
[d] Comenius University, Šafárikovo námestie 6, Bratislava, 814 99, Slovakia
[e] Magdeburg-Stendal University, Breitscheidstraße 2, Magdeburg, 39114, Germany

ARTICLE INFO

Keywords:
Personalized learning
Peer nomination
Student network centrality
AI for learning
Automated content generation
LLMs in education
Innovative teaching methods
Learning management systems
Learning analytics
Student performance

ABSTRACT

Background: Peer influence is a significant determinant in shaping students' academic performance, yet it is often overlooked in traditional educational strategies. The ability to analyze peer influence and collaboration is an important piece in personalizing student educational experiences.

Objective: This study aims to investigate how peer interactions can be used to predict students' achievement levels and create clusters based on these predictions. Building on these clusters, a novel AI-driven approach is introduced to personalize the learning experiences for each group, providing content aligned with students' predicted achievement levels and ensuring that students receive the right resources throughout their entire education.

Methods: Bi-weekly surveys were conducted throughout the academic year with 45 students enrolled in a computer science program, where students nominated up to five peers they collaborated with the most. The data were used to construct a weighted directed graph (also known as directed network), a type of directed graph allowing multiple edges between the same nodes. Centrality measures were calculated from this directed graph to classify students into low-achieving and high-achieving groups for two-class classification, and into low-achieving, high-achieving, and neutral groups for three-class classification. Generative AI was then utilized to create tailored learning scenarios for each student cluster. For high achievers, the scenario includes providing advanced topics, critical-thinking quizzes and supplementary materials such as research papers and real-world projects. For low achievers, these materials include foundational course materials, multiple-choice quizzes and supplementary materials like videos and step-by-step visualizations. The neutral group was given the option to request specific learning materials and self-assessment quizzes on demand to address their individual learning needs.

Results: Eigenvector centrality of the peer collaboration network was identified as the best predictor of student achievement. To assess students' perceived acceptance of the proposed AI-driven personalization, a questionnaire was distributed. The responses indicated that students found the idea appealing and believed it could enhance their learning experience. Several suggestions for improvement were also provided, which will be used to refine the proposed solution in future iterations. Several improvements to the classification mechanism that can enhance its accuracy and effectiveness were also identified during the process.

Conclusion: The findings of this research provide educators, administrators, and policymakers with essential information that can be used to build targeted interventions and support systems for students early in their studies. The findings contribute to the broader discourse on educational strategies, enhancing positive peer interactions and developing a learning environment that provides support and motivation to all students.

* Corresponding author.
E-mail addresses: ivica.pesovski@gmail.com, ivica@next.edu.mk (I. Pesovski).

https://doi.org/10.1016/j.caeai.2025.100430
Received 25 November 2024; Received in revised form 12 May 2025; Accepted 26 May 2025
Available online 30 May 2025

1. Introduction

Artificial Intelligence (AI) has offered new opportunities for higher education, one of which is an efficient and automated course creation mechanism (Khan, 2024). Yet, creating more educational content alone is not enough; it is crucial to know for whom the content is created and why it is needed.

Conventional teaching methods have mostly focused on the academic achievements of individual students. Even in group projects, schools seek methods to assess each member's contributions efficiently (Alden, 2011). The rise of digital learning platforms provides a distinct chance to explore how students interact and collaborate (Timotheou et al., 2023). Understanding collaborative patterns is essential for assessing student progress, guiding teaching methods, developing course syllabi, and fostering a conducive learning environment (Han & Ellis, 2021). Student network analysis can identify key individuals among students who may be suitable for roles like teaching assistants. It can also help detect student isolation and loneliness, allowing educators to prevent dropouts and enhance student performance and satisfaction through early intervention (Phillips et al., 2022). Such interventions were hard to implement because teachers and teaching assistants could devote limited resources to the individual student. The emergence of generative artificial intelligence in the form of large language models (LLMs) will enable these interventions to reach a new level of effectiveness. Institutions are employing these technologies to develop role-based learning by generating learning materials in different styles (Pesovski et al., 2024c). Others have used generative AI to assess and evaluate students as well as assist in teaching (Kasneci et al., 2023; Zawacki-Richter et al., 2019). Personalized learning is going through its AI/LLM renaissance period, with many researchers examining and reporting ways to use such capabilities for a personalized approach to learning (Jauhiainen & Guerra, 2023; Pataranutaporn et al., 2021; Bernius et al., 2022). The missing part that we identified is the answer to the question of how to know what kind of materials and help, in general, to provide to each student in the effort to offer personalized learning experience. In this paper, we propose an approach for early classification of students in either two or three clusters and then providing each cluster with different learning materials, automatically generated by LLM and tailored to the needs of these specific groups of students. The two-way classification categorizes students into low-achieving and high-achieving groups, while the three-way classification adds neutral group. The students were not explicitly labeled or informed of their achievement group to avoid potential demotivation. Instead, all students were transparently provided with materials that best matched their current learning needs, ensuring that the support was perceived as constructive rather than stigmatizing.

The motivation behind our research comes from the necessity to examine whether if we manage to classify students early in the semester into at least two groups (those likely to excel and those likely to struggle academically) we can use AI-generated content to support these student groups separately. We propose a method by which artificial intelligence can be used to assist students with low academic performance in acquiring the knowledge they need to pass exams, while simultaneously providing additional learning opportunities to students with excellent academic performance in order to reduce boredom in the classroom.

The secondary aim of this paper is to investigate the correlation between peer influence and academic performance using a network analysis framework to examine the intricate details of student collaborations. The complex dynamics of student interactions and their predictive power over academic success are examined by leveraging data from regular assignments and bi-weekly surveys of student collaborations. Specifically, this part of the research focuses on employing network centrality measures to identify influential nodes within the student collaboration network and examine how these nodes, indicative of peer influence, correlate with academic performance. In doing so, we want to provide actionable insights into how educational institutions can use the positive aspects of peer influence, mitigate negative effects, and, ultimately, develop a supportive and productive learning environment.

Having explained the motivation and current state, this paper explores the following research questions:

- How can students be classified into high-achieving and low-achieving clusters early in the study process based on their peer collaboration?
- How can AI-generated content and assessments be tailored and distributed to these clusters to enhance student motivation and satisfaction?

The paper is organized as follows: the next section is the background section, where we examine the existing literature in six distinct subsections for each related discipline. The data and methods section follows where we describe the methodology used in our study as well as the data collection process. The results, discussion and conclusion sections follow elaborating the novelty of the presented study and the plans for future work.

2. Background and related work

The background section is subdivided into six distinct subsections, each focusing on key elements relevant to our research. These subsections offer an overview of the existing literature and research in the areas of peer influence, collaborative learning, student classification, early identification of students at risk, AI-enhanced personalized education and AI-generated learning materials.

2.1. Peer influence

In the literature, student-to-student relations have been a topic of extensive research for many years. Recently, academic achievement and its relationship to peer influence have been the subject of considerable scholarly interest. Various studies have provided insights into the mechanisms underlying this relationship. In 2013, using empirical evidence from students enrolled in a second-level degree course, Celant (2013) showed that peer effects are determinants of student performance, with the explanatory power of student achievement greater than that of common regressor techniques. Vitale et al. (2016) showed that student interactions in informal groups, analyzed among participants with an average age of 25, are strongly related to performance, while collaboration in formal groups instructed by the teacher has no such effect on student performance. Molloy et al. (2011) went a step further and examined different types of peer relations among nearly 500 elementary and middle school students, finding that distinct relationship types (e.g., reciprocated friendships, frequent interactions and shared group membership) affected student outcomes in unique ways. Skipper and Keup (2017) in a study of over 1200 undergraduate students, found that performance varied depending on group roles, with peer leaders demonstrating higher levels of academic understanding. Howard (2004) examined the impact of peer pressure on academic performance among adolescents, emphasizing the influence of peer groups on individual learning outcomes. Her study highlighted the importance of social relationships in shaping academic behaviors and achievement. According to Ryan (2000), peer groups have an impact on changes in intrinsic attitudes toward school and, also on student performance among adolescents. It has been found that associating with friends who have a positive attitude towards school increases students' level of comfort with school while socializing with students who have a negative attitude towards school decreases it. Brouwer et al. (2022) found that college students are more likely to help their friends, and students who help each other in their studies are more likely to develop friendships. The higher a student's academic performance, the more often that student is picked as a friend or study assistant. These findings are consistent with the study by Lomi et al. (2011), which provides evidence of peer effects based on longitudinal data from a cohort of 75 students enrolled in a full-time residential MBA program. Golsteyn et al. (2021) conducted their study on a sample of over 4000 students enrolled across bachelor's and master's degree programs, demonstrating that peer activities at the beginning of studies have a sustained positive impact on student performance throughout subsequent semesters.

Jain and Langer (2019), on the other hand, reported findings that an increase in closeness centrality negatively reflects on student productivity as measured by students' grade point average (GPA). Their research focused on postgraduate business studies, where professional goals such as executive remuneration, job search, and promotion take precedence over academic performance. Peza (2015) researched the pressure in peer-group settings among adolescents and found that it decreases academic motivation in adolescents. Reang and Kaipeng (2022) reason that peer collaboration can go both ways, and the teachers should provide suitable guidance in order to steer it toward enhanced academic performance. Understanding students' perceptions of collaboration is equally crucial as understanding how they interact, especially since Pesovski et al. (2023) researched college students and found that student performance is significantly impacted by differences in teachers' and students' perceptions of learning. This emphasis on perception is not new; earlier studies by Tavakoli (2009) and van de Watering and van der Rijt (2006) also highlighted the importance of aligning instructional approaches with students' perceptions in order to enhance academic outcomes and engagement.

These studies highlight the importance of considering peer dynamics in understanding differences in academic performance and designing effective interventions to promote student success. Although there is increasing interest in understanding the significance of peer-to-peer relations in the learning process, there is a scarcity of studies on the correlation between collaboration patterns and academic performance, especially in identifying influential individuals within these networks. This study addresses this gap by comprehensively examining the relationships between academic success, peer collaboration, and the development of influential individuals within a group of first-year students. Teachers can gain highly valuable perspectives by acknowledging and comprehending the impact of key individuals in educational communities. This knowledge encompasses not just the prominent individuals themselves but also their impact on the overall learning dynamics of the wider group. Equipped with these understandings, educators could create and execute specific actions with the goal of using the power of these important individuals. In addition, there is increasing recognition of the capacity for peer-learning analysis to uncover complex findings regarding the emotional and social welfare of students (Thomas et al., 2020). More precisely, analyzing the relationships between peers can provide insights into indications of social isolation or loneliness among students (Oakley, 2019; Diehl et al., 2018). Educators might get a useful viewpoint by detecting individuals experiencing marginalization or disconnection within the school community. These perspectives are essential because they enable prompt and focused interventions, guaranteeing students who are at risk receive the necessary support to succeed academically and emotionally.

2.2. Collaborative learning

In modern education, collaborative learning is increasingly utilized to cultivate a range of skills, enhance subject knowledge, and elevate student engagement. A variety of studies investigating various aspects of this pedagogy have emerged. These studies include analysis on the time and processes by which collaboration affects results, (Nokes-Malach et al., 2015), the most effective strategies for implementing collaborative learning across diverse educational settings (Douville & Wood, 2020), and the role of technology in facilitating collaborative learning experiences (Jeong & Hmelo-Silver, 2016), among others. Educational practitioners often highlight the positive influence of peer collaboration on student achievement (Chandra, 2015).

The study authored by Bouton et al. (2021) examines students' utilization of social networking technologies (SNTs). It reveals that students not only share knowledge through social network tools (SNTs) but also perceive this collaboration as beneficial to their learning outcomes. Similarly, the study by Serra et al. (2023) introduced Student-Led Tutorials, resulting in a 20% increase in course pass rates and improved grade distribution.

Some studies have explored students' preferences in collaborative partnerships. Hoffman et al. (2020) examined students' choices of collaborators, revealing a preference for group members, particularly those demonstrating higher levels of centrality, success, and commitment. Pulgar et al. (2022) looked at how students collaborated in high school physics and found that friendships led to better grades than working with less familiar peers.

Previous studies have highlighted differing collaboration patterns across student groups (Yeh et al., 2018; Feng et al., 2023). Existing studies predominantly focus on behavioral disparities rather than students' preferences regarding the gender of their collaborators. Various approaches regarding assessment strategies for student collaboration are evident in the literature. Student self-reporting is a commonly employed method (Kubincová & Kolcak, 2022; Kubincová et al., 2017), offering versatility across disciplines and classroom modalities. While some researchers advocate for limiting the number of collaborators (Ellis & Han, 2021), others impose no such restrictions. Alternative methods include analyzing server logs and specialized collaboration software (Leeder & Shah, 2016), as well as interviewing teachers to gauge classroom collaboration dynamics (Pathak & Intratat, 2012). Real-time experiments, such as those conducted by Lazonder (2005), involve measuring students' task completion times individually and in groups, offering insights into collaboration dynamics.

2.3. Student classification

Student classification is vital in education to address diverse student needs Alam (2023). Researchers have employed various methods for this purpose. Mythili and Shanavas (2014) investigated factors such as parental education, gender, socioeconomic status, and locality. They applied different algorithms like "J48", "Random Forest", "Multilayer Perceptron", "IB1", and "Decision Table" to predict student performance. Sunday et al. (2020) analyzed completed assignments, lab work, and class attendance, finding class attendance to be the most influential factor in programming subjects. The "C4.5" algorithm was also noted for its effectiveness in student classification (Saheed et al., 2018). Roy and Garg (2017) utilized classification algorithms to categorize students into groups such as potential dropouts, low-performing students, good students who lately deteriorated, and high-achieving students. Dynamic Student Classification on Memory Networks (DSCMN) and Deep Knowledge Tracing models have been proposed by Minn et al. (2019) to assess students' knowledge states and predict their performance. These approaches capture temporal learning abilities and dynamically assign students to groups with similar capabilities. Machine learning techniques have also been applied to classify students into different learner categories based on cognitive abilities, with the k-Nearest Neighborhood (k-NN) algorithm showing high accuracy in predicting learning outcomes (Vital et al., 2020). Molinari et al. (2013) used fuzzy logic techniques to classify students and devise individual study plans in e-learning environments.

Network centrality measurements have also been used to classify students into different clusters, especially when collaboration and peer learning have already been set in place. A study of PhD students found a reversed U-shaped relationship between centrality and academic performance (Zhang et al., 2009). Another study by Vignery and Laurier (2020) proposed a methodology for selecting appropriate centrality measures for student networks, identifying six latent dimensions of centrality. Further research demonstrated positive impacts of geodesic k-path and closeness centralities on GPA, along with a positive effect of cluster density on performance (Vignery, 2022). In a physics classroom context, centrality was found to be predicted by gender and incoming GPA, and was itself a predictor of concept evaluation gain (Williams et al., 2015). Eigenvector centrality was used to identify influential students in collaborative work, with high eigenvector centrality indicating students who act as information sources (Mansur et al., 2016; Pesovski et al., 2024a). These studies highlight the complex relationship between network centrality and student outcomes, suggesting that centrality measures beyond the commonly used—degree, closeness, betweenness and eigenvector—may be valuable in understanding student networks and their impact on academic performance.

2.4. Early identification of at-risk students

A subset of student classification is identifying students at-risk by classifying them as such. Various scholars have highlighted challenges related to quality management in detecting at-risk students (Duarte et al., 2014; Cohen, 2017; Valentine et al., 2011). Data gathered through diverse collaborative learning methodologies can be utilized in models for student classification to effectively pinpoint students who are at risk of dropping out. Santos and Henriques (2023) have demonstrated the robust predictive capabilities of such models, emphasizing the utilization of logs from learning management systems (LMSs) for accurate and timely identification. Na and Tasir (2017) similarly found that data analytics techniques are proficient in identifying at-risk students. Other researchers have employed predictive modeling techniques with notable success in detecting at-risk students (Hung et al., 2019; Santos et al., 2024). Hlosta et al. (2017) devised a methodology relying solely on subject-specific data to train predictive models for student success, demonstrating remarkable accuracy.

2.5. AI-enhanced personalized education

Generative AI models are algorithms specifically created to detect patterns and rules within their training data and produce new observations that follow similar rules (Mondal et al., 2023). These algorithms have progressed from basic statistical methods like the Naive Bayes classifier (Ng & Jordan, 2001) to complex deep learning models with billions of parameters, such as the Generative Pre-trained Transformer (GPT) or Meta's Llama-2 which are deep learning models that have been trained on large volumes of text. These models are mostly used to generate new text based on a given prompt. While the research on the use of LLMs in education is still relatively new, there is an increasing amount of data that suggests that these tools will likely have significant effects on the future of education.

One of the hot topics following the mass-availability of generative AI is personalized education. The current setting in most educational institutions is such that groups of tens or hundreds of students follow any given course at the same pace, delivered by the same professor with the same examples etc. The problem with this approach is that the delivery of the material in such a way is suitable only for a certain number of students, while it is too fast-paced or too slow-paced for most of them (Pritchett & Beatty, 2015; Marzano et al., 2000). Generative AI offers a solution for this problem, by enabling tailoring not only the materials themselves but the delivery method as well to the needs of each and every student. Artificial intelligence technologies facilitate adaptive content delivery, personalized feedback, or learning resources and materials that are aligned with students' preferences, learning goals, and skill levels (Kaswan et al., 2024; Hashim et al., 2022; Jian, 2023). Maghsudi et al. (2021) provide an overview of how AI and machine learning can enhance personalized education by acquiring student characteristics, recommending content, advising curricula, and connecting learners. Researchers have also examined the use of artificial intelligence in personalized learning, promoting better student outcomes, greater engagement, and improved learning experiences (Harry, 2023; Xu, 2024). Recent publications analyze LLM's ability to provide personalized learning experiences as a result of their teaching and grading skills (Kasneci et al., 2023). Tailoring LLMs also accomplished providing personalized feedback on programming questions, according to research by Bernius et al. (2022) and Sailer et al. (2023). Feedback from LLMs is generally well-received and helps learners develop. Lesson creation presents new and exciting possibilities (Pesovski et al., 2024b). Jauhiainen and Guerra (2023) demonstrate how LLMs may be used to tailor a course to diverse audiences' knowledge levels. Customizable lesson creation enables the emergence of virtual educators that will deliver the learning materials in a more engaging way (Pataranutaporn et al., 2021). Ayeni et al. (2024) point out that in addition to AI applications for personalized learning, AI also offers intelligent tutoring systems that are based on the ability to target particular gaps in learning, amplify concepts, and react flexibly to the changing needs of each learner. With regard to the application of artificial intelligence in higher education, Putra Pratama et al. (2023) investigate the perceptions, obstacles, and choices that students anticipate encountering.

2.6. AI-generated learning materials

Generative AI has emerged as an effective means of personalizing student learning experiences and tailoring learning materials to individual readiness levels Pesovski et al. (2024c). Jauhiainen and Guerra (2023) used the content generation capabilities of generative AI to develop multiple iterations of a history course, adapting content to accommodate varying levels of student knowledge. The AI-generated lessons received positive reactions, with students reporting high levels of enjoyment and improved understanding.

In the field of coding education, Sarsa et al. (2022) found that Large Language Models excel at generating introductory programming exercises and providing detailed explanations for code lines. Macneil et al. (2022) demonstrated that ChatGPT can comment on code clearly and comprehensively. Teye et al. (2024), in their study, compared instructor-generated and AI-generated learning materials for teaching the principles of Universal Design for Learning. The results showed that the textual content created by artificial intelligence was highly congruent with instructor-created learning content in terms of relevance, comprehensibility, and engagement with instruction. However, significant differences in syntax and depth were noticed. The study therefore highlights the potential of AI as a complementary resource in education.

Denny et al. (2023) observed that students perceived the quality of AI-created resources to be of equal quality to those created by their peers, suggesting their viability as complementary materials. Rakovac Bekes and Galzina (2022) further highlighted the popularity and acceptability of content, created using intelligent technologies, indicating positive acceptance of AI-created materials. Pesovski et al. (2024c) have recently published their work in which students receive access to the same learning materials delivered in different styles, completely generated by artificial intelligence, ChatGPT in particular, and completely automated. The different styles included ones inspired by pop-culture icons like Batman and Wednesday Addams. Their research reports findings that such practice of making different versions of reading materials increased the time students spend learning and was positively accepted by students.

On the other hand, some studies do not favor AI-generated materials. For example, Darko et al. (2024), in their study compared the design quality of ChatGPT- and Google Bard-generated materials with the materials created by the instructor. They came to the conclusion that AI should be used to support instructors' design of instructional materials, but AI-generated materials should not be used on their own. Instead, they should be used along with the traditional way of instructional design.

Chiu et al. (2023) in their systematic literature review explored the opportunities and challenges of artificial intelligence in education. They identified 13 roles of AI technologies in core educational domains, 7

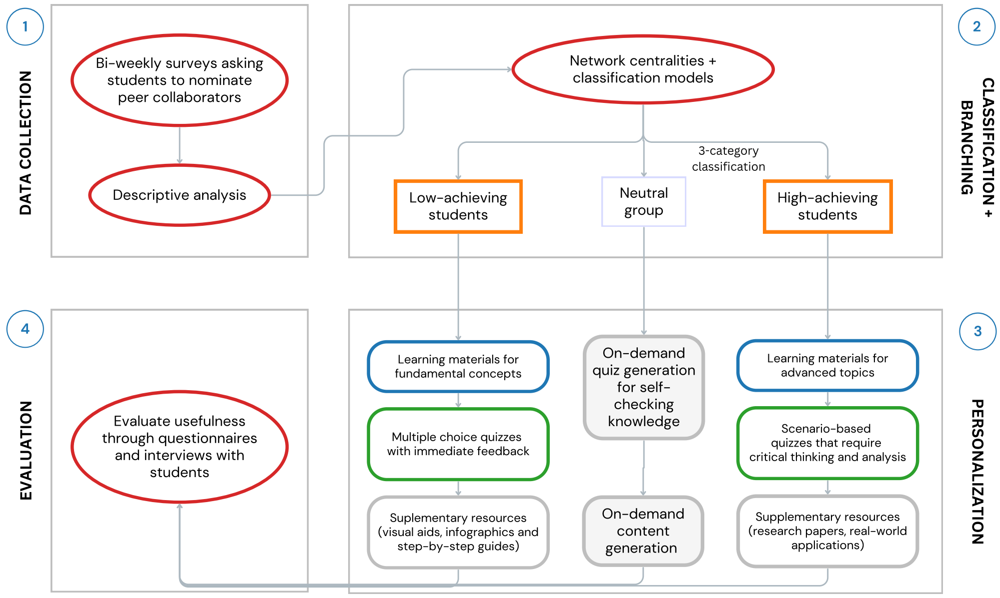

Fig. 1. Visual representation of the steps involved in the study's methodology.

educational outcomes, and 10 top challenges. However, they only mentioned AI-generated educational materials in the context of teachers' complaints about the single format of content provided by AI agents, making it impossible to satisfy efforts to accommodate different learning styles.

Ahmad et al. (2024) in a comprehensive review discuss the applications of data-driven artificial intelligence in education. It provides a detailed overview of existing tools and applications using AI in areas such as student assessment, dropout rates, sentiment analysis, intelligent tutoring systems, and classroom monitoring. However, it only categorizes the area of AI-generated learning materials as part of future research directions and open issues.

3. Data and methods

The methodology of this study was explicitly designed to address the two primary research questions outlined in the introduction: firstly, identifying students' achievement levels early through their patterns of peer collaboration; and secondly, tailoring AI-generated educational materials and assessments based on these classifications to enhance student outcomes.

To answer the first question, we conducted bi-weekly surveys where 45 students nominated peers with whom they collaborated most closely. These nominations were used to build directed weighted graphs representing student collaboration patterns. Network centrality measures derived from these graphs allowed us to classify students early in their studies into high-achieving, low-achieving, and neutral clusters. This classification process is critical, as timely identification enables targeted interventions, potentially improving student performance.

Addressing the second research question, we proposed a generative AI workflow integrated with our learning management system that generates tailored content and quizzes specifically designed for the identified student clusters. Low-achieving students received foundational materials and quizzes intended to solidify essential concepts, while high-achieving students received advanced materials, challenging questions and supplementary resources to sustain their engagement. Neutral students had on-demand access to learning resources.

Finally, the methodology includes evaluating the usefulness of this approach through student questionnaires, evaluating their perception toward the approach. A visual representation of the methodology is given in Fig. 1 and more detailed information about each step follows in distinct subsections.

3.1. Surveys

The study's data was gathered using a customized Learning Management System (LMS) used by students at a software engineering college. This LMS incorporates distinctive functionality specifically designed for this study. The LMS software had a survey system that allowed students to assess teachers and subjects. This mechanism was enhanced by adding an extra stage, allowing students to nominate each other for collaboration.

Beginning in October 2022, students were invited to participate in a bi-weekly survey structured into three distinct parts to capture various aspects of their academic experience and interactions. A total of 14 surveys were delivered to the students throughout the academic year. The questions presented to students in their original form are shown in Table 1. The first part of the survey collected student evaluations of their professors using a 5-star scale, based on teaching experiences from the previous two weeks. The LMS filtered professors to show only those with whom students had direct interactions, ensuring relevant feedback. The second part asked students to rate subjects they attended within the same period, using the same 5-star scale to capture their engagement and interest, allowing for a more personalized assessment of their academic experience. The final part was designed to map the network of student collaborations. Students were presented with a list of all their peers within the same cohort, regardless of shared classes or direct academic interactions. They were instructed to select up to five colleagues with whom they had most frequently collaborated or who had been most helpful over the past two weeks. The data was stored in format from_student_id:to_student_id:timestamp indicating which student gave the nomination to which other students. There could be more records for the same students on subsequent surveys, meaning if in week 1, student A nominated student B, in week 3, student A could again nominate student B. The data is not commutative, meaning that if student A nominates student B, it does not necessarily mean that student B will nominate student A. Due to the design limitation that student nominations do not associate the connections with any particular subjects, yet they still provide valuable information on the patterns and intensity of student collaboration across the program, this step of the survey was most important for performing the network analysis on which this study is based.

Table 1
Survey Questions.

Survey part	Question formulation
Part 1: Professor evaluation	Reflect on the lectures, exercises, and laboratory in the past two weeks. Give a rating based on your experience, their teaching approach and knowledge transfer, the methodology and tools they used in their classes, their communication skills, the constructiveness of the feedback they give, etc. (1 - much improvement needed, 5 - exquisite teaching approach)
Part 2: Subject evaluation	Reflect on the learning outcomes in the past two weeks. Give a rating based on the level of complexity of the subject, the ratio between theory and practice, the examples and exercises, the quality of learning materials, etc. (1 - very hard and incomprehensible, 5 - completely comprehensible)
Part 3: Peer nomination	Which colleagues did you work most closely with in the past two weeks while learning, doing homework, solving exercises and challenges, etc.?

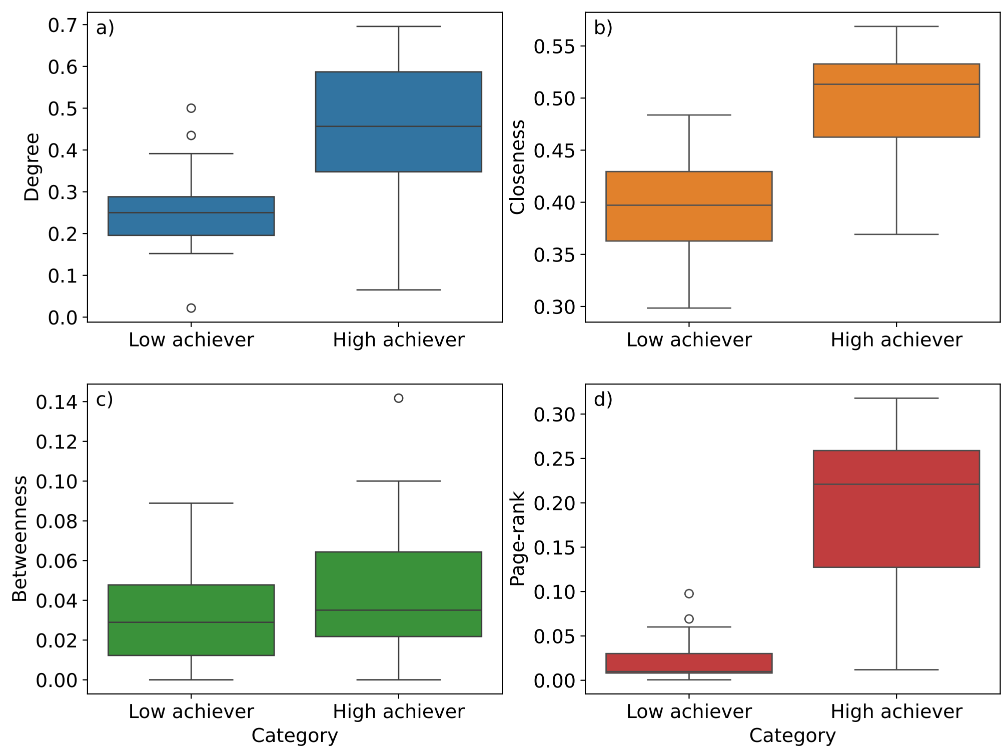

Fig. 2. Box-plots of the network metrics a)-d) as a function of ground-truth two categories classification. The left box plot in each subplot refers to students classified below the median of the total score and the right for the students above the median. The total score is the score each student achieved at the end of the academic year.

3.2. Dependent variable, centrality measures and control variables

In the first step of model specification, we generate the categorical dependent variable as follows; at the end of the academic year each student achieved a certain total score which varies from 0 to 100. This score is calculated by multiplying the student's end-of-year GPA by the number of subjects they passed. This method ensures that the total score reflects not just the average performance in the subjects they passed but also the breadth of their academic success. For example, if a student passed only one out of ten subjects with a GPA of 8, their score would be calculated as 8 (GPA) multiplied by 1 (the number of subjects passed), resulting in a total score of 8 out of 100, rather than appearing as an 8 out of 10. This approach more accurately represents the student's overall academic achievement for the year. Then, we assign each student a ground-truth category according to his quartile membership in the empirical total score distribution. In other words, if a student's score is below the first quartile, then that student will be classified as a low achiever. Thus, this qualitative variable has three levels, namely: Low achiever (below 25%), Neutral (middle 50%) and High achiever (top 25%). On the other hand, this procedure is identical to the case of two-category classification. As an example, a student is a high achiever above the median, and a low achiever otherwise.

In the second step, we include a network centrality measure that is calculated from the empirical networks generated from the survey data: first, given that a student could choose to collaborate with the same peers over multiple surveys, our data contains multiple rows indicating collaboration between the same students. Thus, when presented in a graph structure, we consider a directed weighted graph with the weights derived from the frequency of nominations a student received. Next, we consider four centrality measures: Degree centrality, Closeness centrality, Betweenness centrality and Page-rank (Eigenvector) centrality (Das et al., 2018; Freeman, 2002; Bonacich, 2007; Jackson, 2008).

Finally, we also include two control variables that may have an impact on the probability that a student is classified as a low achiever, high achiever or neutral. These are gender and admission test scores. The latter represents a score that students achieved as part of a test in the college admission process.

In terms of outliers detection, we removed a student from the dataset if at least one quantitative variable had a standardized z-score, $|z| > 3$.

3.3. Descriptive analysis

In this section, we present descriptive analysis in order to motivate the main statistical analysis. Fig. 2 provides box-plots that depict the differences in the magnitudes of the four network centrality measures between the students classified below the median of their total score (Low achiever) or above the median (High achiever). It can be observed that high-achieving students have a consistently higher centrality mea-

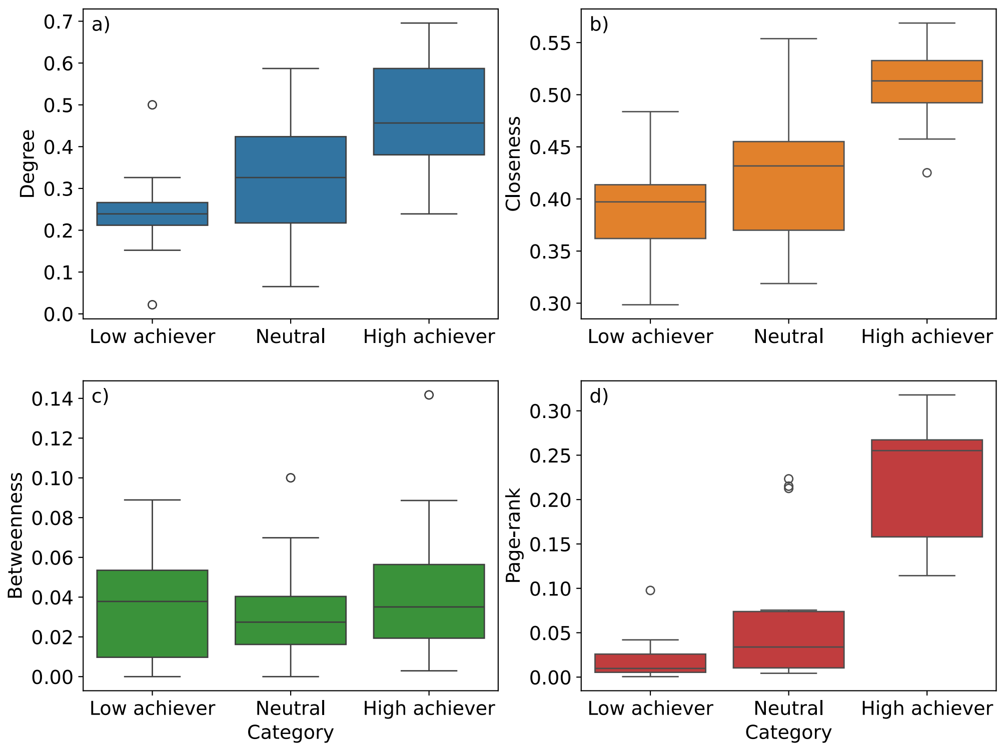

Fig. 3. Box-plots of the network metrics a)-d) as a function of ground-truth three categories classification. The left box plot in each subplot refers to students classified below the first quartile of the total score distribution, the second is for the middle 50%, and the rightmost is reserved for the students in the top 25% of the distribution. The total score is the score each student achieved at the end of the academic year.

sure statistics compared to the low achievers. However, this is not the case for betweenness centrality where the differences are negligible. The largest difference between the two groups is for PageRank centrality indicating that it may be a good predictor in a model for two-class classification. In Fig. 3 we show the differences in the network centralities when using three-class classification based on students' quartile membership in terms of the empirical distribution of the total score variable. We can deduce that PageRank centrality is the best factor that discriminates between these three groups, while betweenness centrality is the worst for the data in hand. For further analysis in this regard, see Sec. 4, the PageRank centrality network in Fig. 5 and the three networks in Appendix A.

3.4. Model

In order to apply a framework for student classification, we first start with the simplest modeling choice, a logistic regression model of the form:

$$p(X;b,W) = \frac{e^{b+W\cdot X}}{1+e^{b+W\cdot X}} = \frac{1}{1+e^{-(b+W\cdot X)}} \tag{1}$$

where X represents the design matrix consisting of a network centrality measure and control variables, W are the weights (coefficients) to be estimated and b is the intercept known as the bias term. The model is based on the sigmoid function, $\sigma(x) = [1-e^{-x}]^{-1}$, which converts continuous variables into probability. In our case, we use Eq. (1) to quantify the probability that a certain student is a "High achiever" and with complementary probability $1-p(X;b,W)$ that the student is a "Low achiever".

In the above case of two categories (binomial logistic regression), the classes were labeled as "Low achiever" and "High achiever", and we had two probabilities: the probability that the outcome was in the category "High achiever" given by $p(x)$ and "Low achiever" with $1-p(x)$. However, in general we can use a generalized classification procedure for multiclass problems called Multinomial logistic regression. It is a model that is used to predict the probabilities of different possible outcomes of a categorically distributed dependent variable as a linear combination of the predictor variables. In our case, the dependent variable has three levels, namely "Low achiever" (bottom 25% of the score distribution), "Neutral" (middle 50%) and "High achiever" (top 25%). The log odds here are modeled relative to some baseline level (here we use the category "Neutral"). In general, one out of C possible outcomes of the categorical variable is chosen as a baseline and then the rest $C-1$ outcomes are individually regressed on the baseline outcome. If the last category is chosen as the baseline, then the $C-1$ regression equations are:

$$\ln\frac{P(Y=c)}{P(Y=C)} = b_c + W_c \cdot X \tag{2}$$

where c is some other outcome s.t. $c < C$.

Exponentiating both sides and rearranging terms we get:

$$P(Y=c) = P(Y=C)\, e^{b_c+W_c\cdot X} \tag{3}$$

Using the complementary probability for $P(Y=C)$ and substituting Eq. (3) for probability of each category $P(Y=c)$, we find:

$$\begin{aligned} P(Y=C) &= 1-\sum_{j=1}^{C-1} P(Y=j) = 1-\sum_{j=1}^{C-1} P(Y=C)e^{b_j+W_j\cdot X} \\ \implies P(Y=C) &= \frac{1}{1+\sum_{j=1}^{C-1} e^{b_j+W_j\cdot X}} \end{aligned} \tag{4}$$

Finally, we use this to find the probability for a particular category c:

$$P(Y=c) = \frac{e^{b_c+W_c\cdot X}}{1+\sum_{j=1}^{C-1} e^{b_j+W_j\cdot X}} \tag{5}$$

For more information about the models see (Hosmer et al. (2013); Agresti (2012); Venables and Ripley (2013)).

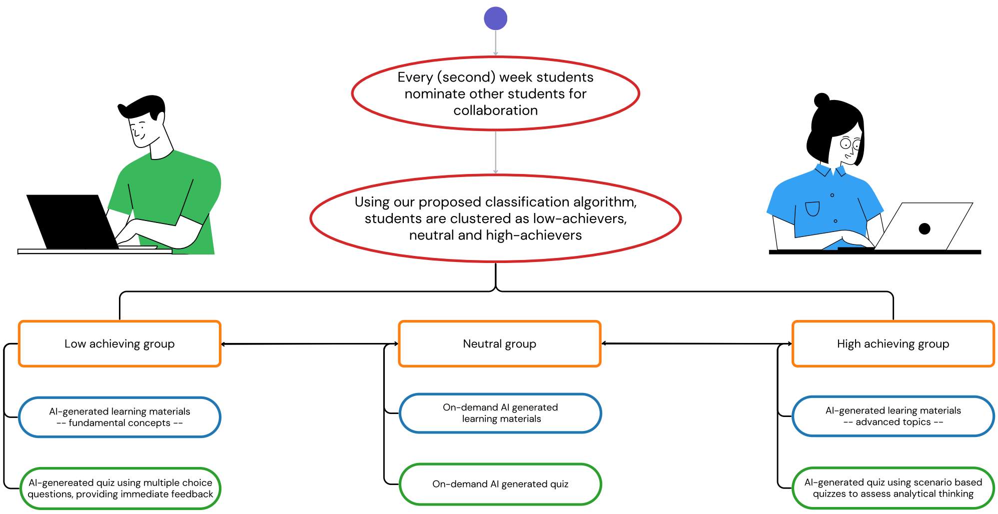

Fig. 4. Workflow methodology.

3.5. Personalization

Following the identification of low and high-achieving students through network analysis of peer nominations, the following imperative is to address their unique learning needs with customized educational content. Hence, the following step in our methodology involves the design and implementation of AI-driven systems that generate tailored learning materials for the distinct groups. The differentiation in educational content aims to improve learning outcomes by addressing specifically the needs of each group.

The suggested approach is presented in Fig. 4.

After every class, the professor enters into the LMS the learning outcomes from the class. Since the students are already classified into high-achieving, low-achieving and neutral clusters on the learning management system, as discussed previously, an API call to OpenAI is made in order to generate appropriate content for both groups. For the low-achieving group, materials explaining the foundational concepts are generated. Mastering these concepts should help these students understand the subject since they are generated based on the learning outcomes entered by the professor. A quiz containing single-choice questions derived from the generated materials is also generated. OpenAI's API returns JSON result with questions and answers that are stored in the existing LMS database and are then distributed to students. For the high-achieving group, the call to OpenAI's API is made with intent to generate advanced materials, identify research papers and come up with real-world examples. A quiz consisting of both single choice and multiple choice answers is also generated and stored in the LMS's database. Contrary to the quizzes generated for the low-achieving group, these quizzes don't need to contain questions from the generated materials, but rather the questions should boost the curiosity and problem solving skills of the high-achieving students. If the three-fold approach is used, with some students classified as neutral, additional steps are included. For these students, the LMS system doesn't automate anything but presents on-demand learning materials and quiz creation. If any of the students belonging to this group ask for more materials, the LMS connects to OpenAI to generate appropriate content and quiz to test the knowledge.

3.6. Evaluation of perceived acceptance and anticipated benefits

To evaluate the perceived usefulness and acceptance of the proposed approach, a questionnaire was distributed to the participating students in this research. The questions are detailed in Table 2. The design of this questionnaire was influenced by established educational technology acceptance models, particularly drawing from frameworks such as those developed by Teo (2011), which focus on factors influencing the acceptance and use of technology in educational contexts. The results are discussed in Sec. 4.

Table 2
Questionnaire items evaluating students' perceived usefulness and acceptance.

Question
1. How do you prefer to learn new material? (options: reading, video, interactive activities, practical exercises, other)
2. How well do you understand the proposed AI-generated materials tool? (options: 1-5)
3. How useful do you think this tool would be for your studies? (options: 1-5)
4. How likely are you to use this tool if it were available? (options: 1-5)
5. To what extent do you believe personalized materials will help you perform better academically? (options: 1-5)
6. What benefits do you anticipate from using this AI tool? (options: improved understanding of material, better grades, more efficient study time, increased motivation, other)
7. What concerns do you have about using this tool? (options: accuracy of AI-generated content, data privacy, dependency on technology, lack of personal interaction)
8. What suggestions do you have for improvement? (open-ended)

4. Results

Fig. 5 shows the network constructed from the responses to Part 3 from the survey questions shown in Table 1. The node size is proportional to the PageRank (eigenvector) centrality measure, whereas color intensity corresponds to the final score the student achieved. As previously argued in section 3.3, here we observe that students with higher scores are located centrally in the network. The student networks for the remaining network centrality measures (degree centrality, betweenness centrality and closeness centrality) are shown in Appendix A. Concretely, the graphs corroborate the results obtained through the simple statistical modeling. For instance, when considering the network in terms of betweenness centrality (Fig. A.10), we can conclude that this metric does not provide meaningful insights into student performance and is not a good predictor for students' gpa scores (for our dataset) because of its weak explanatory power.

Fig. 6 visually summarizes the results for the binomial logistic regression and multinomial logistic regression models, respectively. These models utilize degree, closeness, betweenness, and PageRank centrality

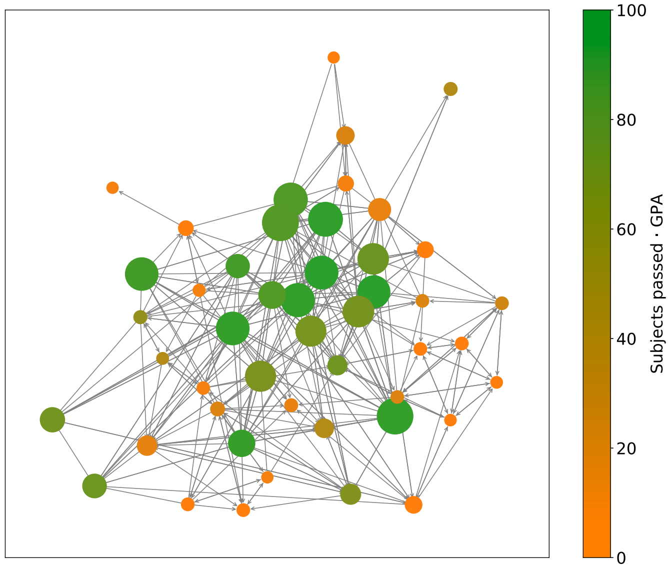

Fig. 5. A student network. Node color represents their score calculated as the product of number of subjects passed and GPA, and the node size is proportional to their PageRank centrality (eigenvector centrality).

as the main predictor variables to assess their influence on the outcomes. In addition to these centrality measures, all models included control variables, such as admission test scores and gender, to ensure the robustness and accuracy of the findings.

Table 3 shows the results for the two-category classification models (1-4), contains each network centrality metric separately and also includes the two control variables. The network metric is the only independent variable in models (5-8). The results show that the log odds for a student to be classified as a "High achiever" is positively related to each centrality measure. The estimated coefficient for Betweenness centrality is the only one which is not statistically significant. A simple inspection of the lower left subplot in Fig. 6 reveals that this result is expected, meaning that betweenness centrality cannot discriminate between the two groups of students in our dataset. On the other hand, the regression coefficients of degree, betweenness and eigenvector centrality are statistically significant even after controlling for admission test score and gender. The coefficients in the binomial logistic regression represent the change in the log odds of the outcome ("High achiever") associated with a one-unit increase in the predictor (independent) variable.

In our case, for every unit change in Page-Rank centrality, the log odds of the student being "High achiever" (versus "Low achiever") increases by 4.2 before and 5.65 after controlling for admission test score and gender. We also report Pseudo-R^2 as a measure of goodness-of-fit which is used when the outcome variable is nominal or ordinal. In particular, the model that includes PageRank centrality as a predictor, along with the control variables, emerges as the best one among those tested. This model achieves a Pseudo-R^2 value of 0.82, indicating a strong explanatory power when the control variables are included. Even without the control variables, the model maintains a high Pseudo-R^2 of 0.80, underscoring the significant predictive capacity of PageRank centrality on its own. This suggests that students with higher centrality scores are not only more connected but also more likely to succeed academically. Following this, the models incorporating closeness and degree centrality as predictors also demonstrate considerable effectiveness, with Pseudo-R^2 values ranging from 0.48 to 0.66, though the model using PageRank centrality remains the most accurate.

We proceed with the multinomial logistic regression model and report the results for each network centrality measure, except for betweenness centrality due to non-significant results. Tables 4, 5 and 6 report the results for three-category classification models for degree, closeness and PageRank centrality, respectively, together with the control variables.

As expected, the results show that larger centrality measure (any of degree, closeness and eigenvector centrality) leads to a larger probability for a student to be classified as an "High achiever", versus the middle 50% group ("Neutral"), indicating that students with stronger positions in the network tend to achieve higher academic performance. In particular, the fitted model (see Table 4) for the log-odds of being classified as a "High achiever" (versus "Neutral") when the network predictor is the eigenvector (PageRank) centrality, reads:

$$\ln\left(\frac{p(\text{'High achiever'})}{p(\text{'Neutral'})}\right) = -9.78 + 24 \cdot \text{pagerank}+ \\ 3.19 \cdot (\text{gender} = \text{m}) + 0.15 \cdot \text{Admission test score} \tag{6}$$

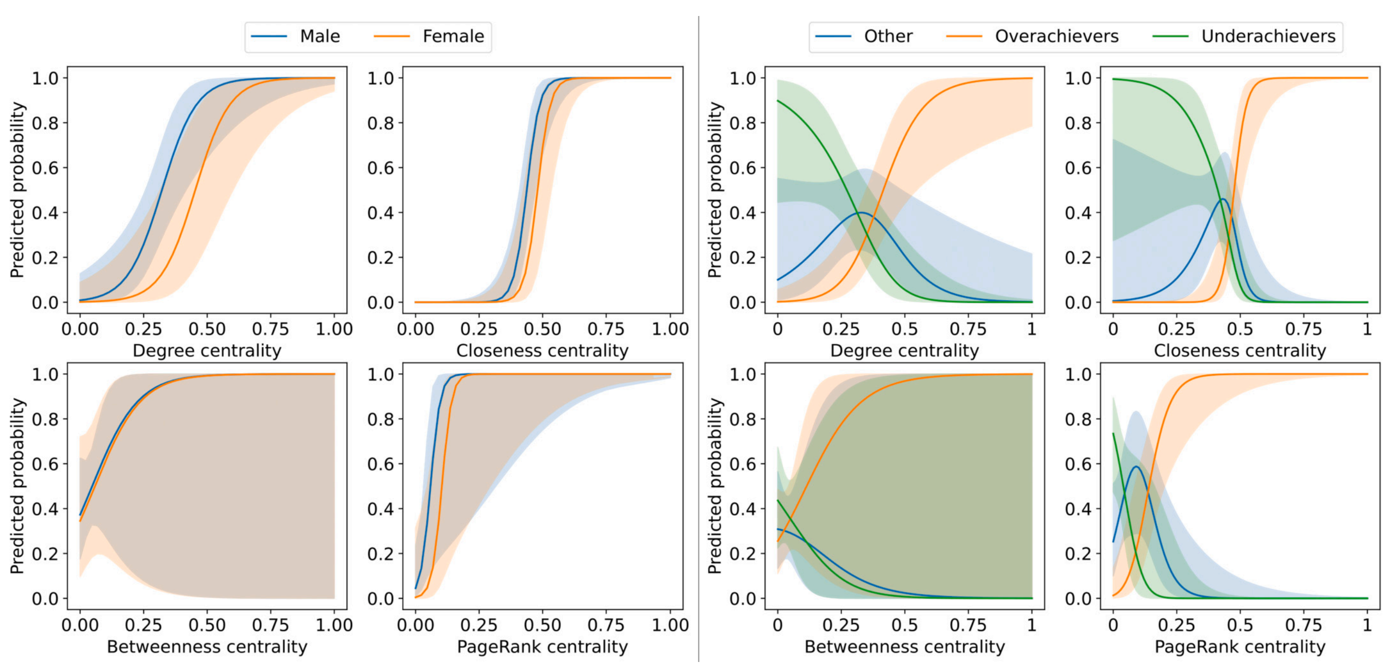

Fig. 6. Logistic regression results (left), Multinomial logistic regression results (right).

Table 3
Logistic regression results.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
(Intercept)	-1.55	-0.23	-1.60	-0.01	-0.1	-0.13	-0.26	0.89
	(1.26)	(0.76)	(1.41)	(1.83)	(0.38)	(0.32)	(0.47)	(0.66)
Degree centrality	2.36 **				1.85 **			
	(0.90)				(0.66)			
Gender (male = 1)	1.93	0.12	1.69	2.34				
	(1.39)	(0.85)	(1.48)	(2.06)				
Admission test score	0.05	0.28	-0.41	-0.51				
	(0.38)	(0.37)	(0.37)	(0.48)				
Betweenness centrality		0.33				0.40		
		(0.35)				(0.32)		
Closeness centrality			2.93 **				2.48 ***	
			(0.89)				(0.74)	
Eigenvector centrality				5.65 ***				4.20 ***
				(1.37)				(1.25)
N	45	45	45	45	45	45	45	45
AIC	46.88	67.79	39.27	27.43	46.11	64.58	37.84	25.39
BIC	54.11	75.01	46.49	34.66	49.72	68.19	41.45	29.00
Pseudo R^2	0.54	0.07	0.66	0.82	0.48	0.05	0.62	0.80

All continuous predictors are standardized.
The brackets contain heteroskedasticity robust standard errors.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 4
Multinomial logistic regression results (eigenvector centrality).

	$\log(OR)$	SE	95% CI
Low achievers			
Eigenvector centrality	-23	14.1	(-51, 4.5)
Gender			
F	-	-	-
M	0.41	1.17	(-1.9, 2.7)
Admission test score	0.07	0.112	(-0.15, 0.29)
High achievers			
Eigenvector centrality	24**	8.67	(6.8, 41)
Gender			
F	-	-	-
M	3.2	1.91	(-0.55, 6.9)
Admission test score	0.15	0.137	(-0.12, 0.42)
N	45		
AIC	69.7		
Pseudo R^2	0.46		

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.
OR = Odds Ratio, SE = Standard Error, CI = Confidence Interval.

Table 6
Multinomial logistic regression results (degree centrality).

	$\log(OR)$	SE	95% CI
Low achievers			
Degree centrality	-6.5	4.17	(-15, 1.7)
Gender			
F	-	-	-
M	0.49	1.10	(-1.7, 2.6)
Admission test score	0.05	0.095	(-0.13, 0.24)
High achievers			
Degree centrality	10*	4.17	(1.9, 18)
Gender			
F	-	-	-
M	2.6*	1.29	(0.04, 5.1)
Admission test score	0.10	0.102	(-0.10, 0.30)
N	45		
AIC	88		
Pseudo R^2	0.27		

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.
OR = Odds Ratio, SE = Standard Error, CI = Confidence Interval.

Table 5
Multinomial logistic regression results (closeness centrality).

	$\log(OR)$	SE	95% CI
Low achievers			
Closeness centrality	-13	8.08	(-28, 3.3)
Gender			
F	-	-	-
M	0.87	1.02	(-1.1, 2.9)
Admission test score	0.08	0.103	(-0.12, 0.28)
High achievers			
Closeness centrality	37**	12.8	(12, 63)
Gender			
F	-	-	-
M	2.9*	1.45	(0.06, 5.7)
Admission test score	0.01	0.117	(-0.22, 0.24)
N	45		
AIC	78.5		
Pseudo R^2	0.37		

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.
OR = Odds Ratio, SE = Standard Error, CI = Confidence Interval.

The estimated coefficient for PageRank is the only significant one and is interpreted as follows: a one-unit increase in eigenvector centrality is associated with an increase of 24 in the log odds of being an "High achiever" versus the group "Neutral". This suggests that PageRank centrality is a strong predictor of student performance.

On the other hand, the equation for the log odds of "Low achiever" versus "Neutral" reads:

$$\ln\left(\frac{p(\text{'Low achiever'})}{p(\text{'Neutral'})}\right) = -0.99 - 23 \cdot \text{pagerank} + 0.41 \cdot (\text{gender} = \text{m}) + 0.07 \cdot \text{Admission test score} \tag{7}$$

In the above case, there are no statistically significant estimated coefficients.

In terms of the remaining models, (see Table 5 and Table 6) for the log-odds of "High achiever" (versus "Neutral") when the network predictors are Closeness and Degree centrality, respectively, show statistical significance of both network metrics. Concretely, the coefficient of closeness centrality is significant on level $p < 0.01$ and degree centrality on $p < 0.05$. In addition, the coefficient for the gender variable (Male = 1) is positive and statistically significant on level $p < 0.05$ in both models.

Table 7
Model accuracy.

	Two-class		Three-class	
	Training	Test	Training	Test
Degree	82%	57%	64%	43%
Closeness	84%	51%	69%	38%
Betweenness	60%	55%	42%	38%
Page-rank	93%	62%	71%	45%

Statistical analysis was carried out using R version 4.3.1 (The R Core Team (2020)) and the *nnet* (v7.3-19, Venables and Ripley (2013)) packages. In particular, for the multinomial logistic model, we used the *multinom* function from the *nnet* package, and for the binomial logistic models, we used the *glm* function from the *stats* package.

4.1. Model accuracy

In this section, we evaluate the accuracy of the models by measuring the agreement between the ground-truth classifications of students and the classifications predicted by two distinct models. The first model is a logistic regression that categorizes students into two groups: low achiever and high achiever. The second model is a multinomial logistic regression, which expands the categorization into three groups: low achiever, neutral, and high achiever. The ground-truth classifications used in the out-of-sample analysis for both the two-category and three-category models were determined *a priori* through a qualitative assessment conducted by college professors who closely observed and worked with the students throughout the academic year. This approach ensures that the ground-truth data is based on informed, expert evaluations of student performance, providing a reliable benchmark against which the model predictions can be tested. Comparing model outputs to this expert-informed classification offers a meaningful measure of the models' practical value in real educational settings. In particular, we report:

- The classification accuracy relative to the ground truth of the training set (in sample) which is set according to the quartile membership. Here, we fit two models on the training dataset and then check the accuracy with respect to the students' ground truth in the training set.
- The performance of the model with respect to the ground-truth categories in a test set (out of sample). We fit two models on the training set and check the accuracy with respect to the ground truth of the second-generation students (test set). Note that the training set in this part is limited only to the time period that coincides with the test set.
- The first two bullet points are repeated for each network centrality measure included in the models as an independent variable.
- Temporal analysis of model accuracy.

The accuracy of the models is summarized in Table 7. The best-performing model is the one with eigenvector centrality (PageRank) as a predictor variable. Thus, to further understand the total accuracy in detail we report the confusion matrix for each model with eigenvector centrality (see Fig. 7).

Finally, the temporal properties of the classifications are of significant interest. For instance, the time when the classification accuracy is maximum or saturates is of great importance because educators can take appropriate actions early in the semester to improve the learning of all groups. In our case, the temporal accuracy shown in Fig. 8 indicates that the model's maximum accuracy is reached within the first two months of the academic year. This implies that the earliest intervention can be implemented within this two-month time frame.

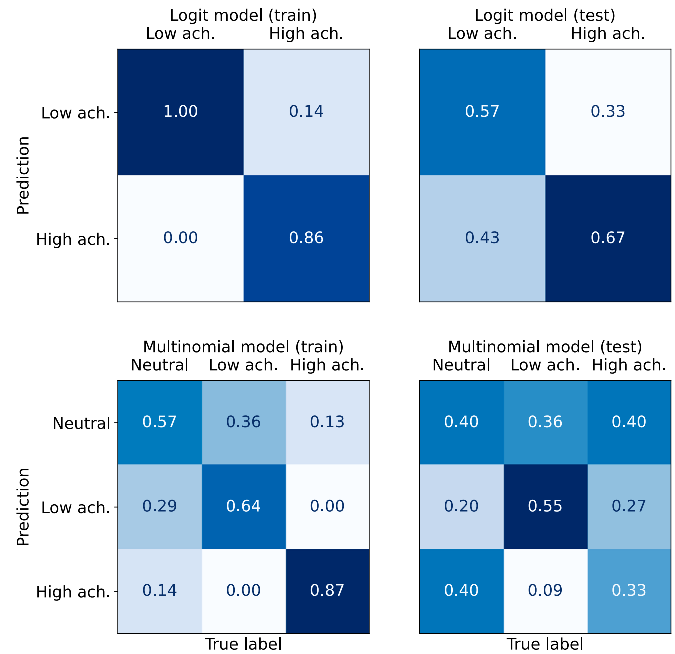

Fig. 7. Confusion matrices (PageRank - Eigenvector centrality).

4.2. Analysis of questionnaire responses on projected student acceptance

As discussed in the data and methods section, we delivered a questionnaire to participating students in order to evaluate their perceived acceptance of the proposed approach. We were interested in finding their preferences and perceptions regarding AI-generated materials and the concept of our proposal in general. Prior to completing the questionnaire, students were presented with the functionality of the AI-generated materials tool, which involves grouping students into low-achieving and high-achieving groups and personalizing their learning materials accordingly. Following the introduction, a questionnaire was delivered to 14 student participants to assess their understanding and perceptions of the tool.

The results indicate a generally positive reception. The students' understanding of the proposed functionality was rated highly, with a mean score of 4.5 (SD = 0.65) on a 5-point scale. This suggests that most students understand the concept of categorizing learners and personalizing materials based on their achievement levels. The perceived usefulness of the tool for their studies also received a mean score of 4.5 (SD = 0.65), indicating that students believe the tool would significantly benefit their academic success. The likelihood of students using the tool was even higher, as indicated by a mean score of 4.71 (SD = 0.47), which demonstrates the students' high level of enthusiasm and preparedness to include this technology in their learning process.

In addition to the quantitative questions, the students were asked to give feedback on the benefits, concerns and suggestions for improving the AI-generated materials tool via open-ended questions. The qualitative responses were analyzed to identify commonalities. Students generally expected the tool to improve their understanding of course material and make their study time more efficient. The primary concerns raised by students revolved around the "lack of personal interaction", the "accuracy of AI-generated content" and the potential "dependence on technology". These concerns highlight the importance of ensuring that the tool complements rather than replaces human interaction in the learning process. Students provided several constructive suggestions for improving the tool, including ensuring that the tool adapts flexibly to changes in students' performance levels. One student mentioned the importance of continuous feedback to further refine the tool's output.

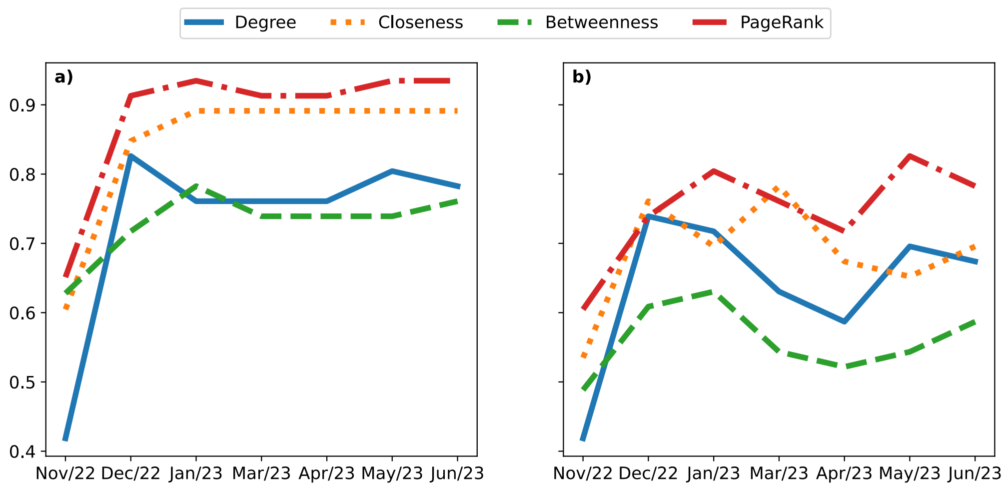

Fig. 8. In-sample accuracy over time. **a)** shows the evolution of the accuracy of the simple logistic model with two categories classification. On the other hand **(b)** shows the accuracy over time for the multinomial logit model with three categories. A particular month on the x-axis represents measurements until and including that month.

5. Limitations

In this section we focus on the limitations of the employed statistical methodology. The first limitation to the generalization of the results is the sample size of 45 students (with 14 students participating in the qualitative analysis). This number is especially important for the case of multinomial logistic regression, where multiple equations are involved in the fitting procedure, which requires an even larger sample size than binary logistic regression. Also, even though the results for binomial logistic regression are satisfactory, a potential limitation is the fact that a student is “Low achiever” if he or she is below the median, and “High achiever” otherwise (in terms of the ground-truth category). A more realistic classification would be to consider only those students that are below the first quartile as low achievers and high achievers those with scores above the third quartile. However, this also requires a larger sample size due to the removal of 50% of students located in the interquartile range of the score distribution. To avoid this and in particular to address the issue of misclassification, future research should consider implementing control materials for neutral students, for example a functionality that offers these students the option to engage with two different sets of materials—one designed for low-achieving and the other for high-achieving students—and then asking students to rate their understanding of each. This strategy could result in deeper insights into their actual proficiency levels and provide additional data points, enabling us to refine the classification and ensure more targeted interventions for this group.

Second, we classify the students based on their network centrality measure derived from their collaborations with other students. One underlying assumption here is that *every* high-achieving student is central in the social network of the college. In reality, this is not the case for introverted high achievers who may falsely appear on the outskirts of the graph and will be incorrectly classified as average or low achievers. To mitigate this bias, one can control for the Myers-Briggs Type Indicator by adding an additional qualitative variable. In addition, we observe inflated centrality scores which stem from favoritism based on personal relationships between students rather than merit, as expected. In future work, one can think of a reasonable control variable that can remedy this phenomenon. Furthermore, controlling for students’ previous education type or grades is something that is absent in our model, albeit we use a proxy for this, namely their score on the admission test. Nevertheless, using a direct measurement of past education could improve the model.

Third, the ground-truth category in the training phase is set according to a students’ quartile membership in the empirical total score distribution (derived from GPA). This approach was chosen to ensure equal membership across all groups, but it has limitations due to the use of a strict cutoff point, which may inaccurately assign students to certain groups as ground-truth. A better alternative is using a qualitative assessment (as we did in the test phase due to non-availability of GPA data for the test group) conducted by the professors who closely observe and work with the students throughout the academic year.

Finally, the perceived usefulness of the proposed approach to integrating AI into the LMS was tested on a small sample of only 14 students. Although the initial results are promising, the small sample size limits the generalizability of these findings. Following the implementation of the tool, its real-world usage will be closely monitored and assessed, with findings to be reported in future studies to evaluate its effectiveness and practical value.

6. Conclusion and future work

This study demonstrated the potential of using network analysis for early classification of students into achievement-level groups based on their peer activities and using AI-driven personalization to provide targeted support for these groups. We found that eigenvector centrality was the most effective predictor for classifying students into low-achieving, high-achieving, and neutral groups, while closeness and degree centralities were moderately effective, and betweenness centrality proved to be a poor predictor of student success. These findings demonstrated that peer collaboration data can serve as a powerful tool for predicting academic outcomes.

Our generative AI-driven approach to providing personalized learning content at the group level has been well-received by students, leading to higher reported motivation and comprehension among both low- and high-achieving groups. Notably, the benefits observed do not require content personalization at the individual level but can be effectively realized through targeted content for identified achievement clusters.

Furthermore, our analysis suggests an additional potential application beyond personalization: the network-based classification could be leveraged strategically to disrupt low-performing collaboration circles. By intentionally mixing student groups—curricularly through project assignments or extra-curricularly via team-building activities—we can help break patterns of persistent underachievement and facilitate improved peer interactions. Such an approach may lead to enhanced overall student performance and a more balanced peer-learning environment.

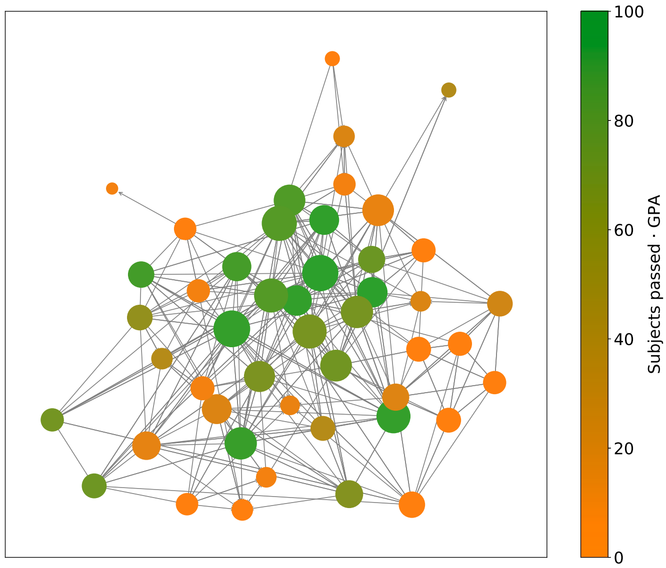

Fig. A.9. A student network. Node color represents their score calculated as the product of number of subjects passed and GPA, and the node size is proportional to their degree centrality.

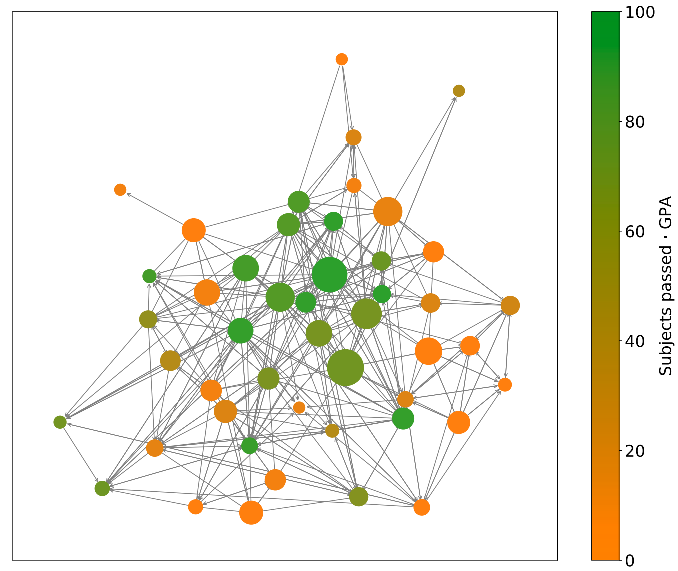

Fig. A.10. A student network. Node color represents their score calculated as the product of number of subjects passed and GPA, and the node size is proportional to their betweenness centrality.

Future work will focus on refining classification mechanisms, controlling for relevant missing factors and validating the effectiveness through long-term studies. Although this study identifies early predictors of student achievement and proposes tailored generative-AI educational materials, future research will include practical implementation and longitudinal evaluation. Specifically, upcoming studies will involve real-world deployment of these personalized resources, monitoring students' actual engagement, help-seeking behaviors, and measurable improvements in their academic performance.

CRediT authorship contribution statement

Ivica Pesovski: Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Petar Jolakoski:** Writing – original draft, Visualization, Resources, Investigation, Formal analysis, Conceptualization. **Vladimir Trajkovik:** Writing – review & editing, Validation, Supervision, Project administration, Methodology. **Zuzana Kubincova:** Writing – review & editing, Validation, Supervision, Investigation. **Michael A. Herzog:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Network visualizations

In this appendix section we show the networks for the remaining centrality measures, namely: degree centrality, betweenness centrality and closeness centrality. See Figs. A.9, A.10, A.11.

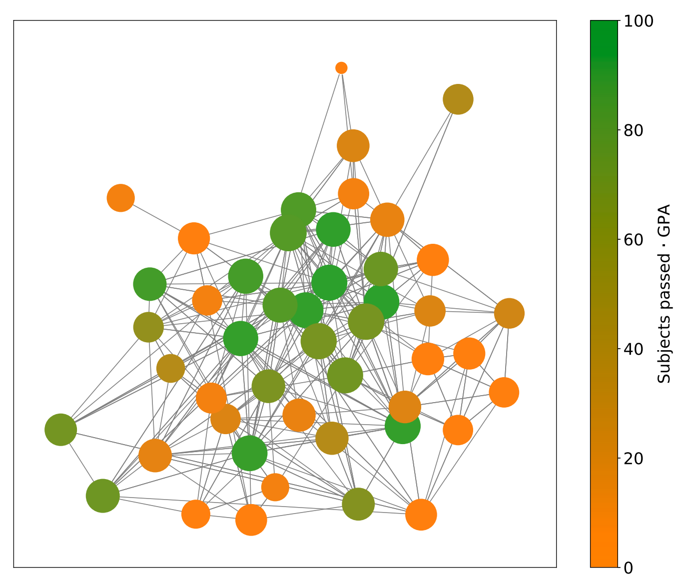

Fig. A.11. A student network. Node color represents their score calculated as the product of number of subjects passed and GPA, and the node size is proportional to their closeness centrality.

References

Agresti, A. (2012). *Categorical data analysis* (3rd edition). John Wiley & Sons.

Ahmad, K., Iqbal, W., El-Hassan, A., Qadir, J., Benhaddou, D., Ayyash, M., & Al-Fuqaha, A. (2024). Data-driven artificial intelligence in education: A comprehensive review. *IEEE Transactions on Learning Technologies*, *17*, 12–31.

Alam, A. (2023). *Harnessing the power of ai to create intelligent tutoring systems for enhanced classroom experience and improved learning outcomes. Lecture notes on data engineering and communications technologies: Vol. 171*. Springer (pp. 571–591).

Alden, J. (2011). Assessment of individual student performance in online team projects. *Journal of Asynchronous Learning Networks*, *15*(3), 5–20.

Ayeni, O. O., Hamad, N. M. A., Chisom, O. N., Osawaru, B., & Adewusi, O. E. (2024). Ai in education: A review of personalized learning and educational technology. *GSC Advanced Research and Reviews*, *18*(2), 261–271.

Bernius, J. P., Krusche, S., & Bruegge, B. (2022). Machine learning based feedback on textual student answers in large courses. *Computers and Education: Artificial Intelligence*, *3*, Article 100081.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*(4), 555–564.

Bouton, E., Tal, S. B., & Asterhan, C. S. C. (2021). Students, social network technology and learning in higher education: Visions of collaborative knowledge construction vs. the reality of knowledge sharing. *The Internet and Higher Education*, *49*, Article 100787.

Brouwer, J., de Matos Fernandes, C. A., Steglich, C. E. G., Jansen, E. P. W. A., Hofman, W. H. A., & Flache, A. (2022). The development of peer networks and academic performance in learning communities in higher education. *Learning and Instruction*, *80*, Article 101603.

Celant, S. (2013). The analysis of students' academic achievement: The evaluation of peer effects through relational links. *Quality and Quantity*, *47*(2), 615–631.

Chandra, R. (2015). Collaborative learning for educational achievement. *Journal of Research & Method in Education*, *5*(3), 4–07.

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *4*, Article 100118.

Cohen, A. (2017). Analysis of student activity in web-supported courses as a tool for predicting dropout. *Educational Technology Research and Development*, *65*(5), 1285–1304.

Darko, C. O., Karki, R., Teye, E., & Liu, L. (2024). *Assessment of the quality of design of human created and ai-based learning*.

Das, K., Samanta, S., & Pal, M. (2018). Study on centrality measures in social networks: A survey. *Social Network Analysis and Mining*, *8*(1), 1–11.

Denny, P., Khosravi, H., Hellas, A., Leinonen, J., & Sarsa, S. (2023). Can we trust ai-generated educational content? Comparative analysis of human and ai-generated learning resources. arXiv preprint. Retrieved from arXiv:2306.10509.

Diehl, K., Jansen, C., Ishchanova, K., & Hilger-Kolb, J. (2018). Loneliness at universities: Determinants of emotional and social loneliness among students. *International Journal of Environmental Research and Public Health*, *15*(9), 1865.

Douville, P., & Wood, K. D. (2020). Collaborative learning strategies in diverse classrooms. In *Collaboration for diverse learners* (pp. 123–151). Routledge.

Duarte, R., Ramos-Pires, A., & Gonçalves, H. (2014). Identifying at-risk students in higher education. *Total Quality Management & Business Excellence*, *25*(7–8), 944–952.

Ellis, R., & Han, F. (2021). Assessing university student collaboration in new ways. *Assessment & Evaluation in Higher Education*, 1–16.

Feng, Q., Luo, H., Li, W., Chen, T., & Song, N. (2023). Effects of gender diversity on college students' collaborative learning: From individual gender to gender pairing. *Heliyon*, *9*(6), Article e16237.

Freeman, L. C. (2002). *Centrality in social networks: Conceptual clarification*In *Social network: Critical concepts in sociology, Vol. 1* (pp. 215–239). Routledge.

Golsteyn, B. H. H., Non, A., & Zölitz, U. (2021). The impact of peer personality on academic achievement. *Journal of Political Economy*, *129*(4), 1052–1099.

Han, F., & Ellis, R. A. (2021). Patterns of student collaborative learning in blended course designs based on their learning orientations: A student approaches to learning perspective. *International Journal of Educational Technology in Higher Education*, *18*(1), 1–16.

Harry, A. (2023). Role of ai in education. *Injurity: Interdisciplinary Journal and Humanity*, *2*(3).

Hashim, S., Khaizer Omar, M., Ab Jalil, H., & Bt Mohd Sharef, N. (2022). Trends on technologies and artificial intelligence in education for personalized learning: Systematic literature review. *International Journal of Academic Research in Progressive Education and Development*, *12*(1), 884–903.

Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017). Ouroboros: Early identification of at-risk students without models based on legacy data. In *ACM international conference proceeding series* (pp. 6–15).

Hoffman, A. S., Hamm, J. V., Lambert, K., & Meece, J. (2020). Peer-collaborator preferences of African American and white sixth-graders in mathematics classrooms. *The Journal of Experimental Education*, *88*(4), 559–577.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd edition). Wiley.

Howard, N. M. (2004). *Peer influence in relation to academic performance and socialization among adolescents: A literature review*. Doctoral dissertation, University of Wisconsin–Stout.

Hung, J. L., Shelton, B. E., Yang, J., & Du, X. (2019). Improving predictive modeling for at-risk student identification: A multistage approach. *IEEE Transactions on Learning Technologies*, *12*(2), 148–157.

Jackson, M. O. (2008). *Social and economic networks*. Princeton University Press.

Jain, T., & Langer, N. (2019). Does whom you know matter? Unraveling the influence of peers' network attributes on academic performance. *Economic Inquiry*, *57*(1), 141–161.

Jauhiainen, J. S., & Guerra, A. G. (2023). Generative ai and chatgpt in school children's education: Evidence from a school lesson. *Sustainability*, *15*(18), Article 14025.

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist*, *51*(2), 247–265.

Jian, M. J. K. O. (2023). Personalized learning through ai. *Advances in Engineering Innovation*, *5*(1), 16–19.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). Chatgpt for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274.

Kaswan, K. S., Dhatterwal, J. S., & Ojha, R. P. (2024). Ai in personalized learning. In *Advances in technological innovations in higher education: Theory and practices* (pp. 103–117). CRC Press.

Khan, S. (2024). *Brave new words: How AI will revolutionize education (and why that's a good thing)*. Crown Publishing Group.

Kubincová, Z., Homola, M., & Dropčová, V. (2017). *Peer reviews in a web design course: Now students like them too*. *Lecture notes in computer science: Vol. 10108*. Springer (pp. 624–634).

Kubincová, Z., & Kolcak, K. (2022). Team assessment and self-assessment – can high school students do it? In *2022 20th international conference on information technology based higher education and training (ITHET)*.

Lazonder, A. W. (2005). Do two heads search better than one? Effects of student collaboration on web search behaviour and search outcomes. *British Journal of Educational Technology*, *36*(3), 465–475.

Leeder, C., & Shah, C. (2016). Strategies, obstacles, and attitudes: Student collaboration in information seeking and synthesis projects. *Information Research: An International Electronic Journal*, *21*(3).

Lomi, A., Snijders, T. A. B., Steglich, C. E. G., & Torló, V. J. (2011). Why are some more peer than others? Evidence from a longitudinal study of social networks and individual academic performance. *Social Science Research*, *40*(6), 1506–1520.

Macneil, S., Tran, A., Mogil, D., Bernstein, S., Ross, E., & Huang, Z. (2022). Generating diverse code explanations using the gpt-3 large language model. In *Proceedings of the 2022 ACM conference on international computing education research, Vol. 2* (pp. 37–39).

Maghsudi, S., Lan, A., Xu, J., & van der Schaar, M. (2021). Personalized education in the artificial intelligence era: What to expect next. *IEEE Signal Processing Magazine*, *38*(3), 37–50.

Mansur, A. B. F., Yusof, N., & Basori, A. H. (2016). The analysis of student collaborative work inside social learning network analysis based on degree and eigenvector centrality. *International Journal of Electrical and Computer Engineering*, *6*(5), 2488–2498.

Marzano, R. J., Gaddy, B. B., & Dean, C. (2000). *What works in classroom instruction*.

Minn, S., Desmarais, M. C., Zhu, F., Xiao, J., & Wang, J. (2019). *Dynamic student classification on memory networks for knowledge tracing*. *Lecture notes in computer science: Vol. 11440*. Springer (pp. 163–174).

Molinari, A., Trevisan, I., & Bodganova Akhmetova, D. (2013). Classifying students based on fuzzy logic to devise individual study plans. *Systems Engineering*, *17*(59), 43–46.

Molloy, L. E., Gest, S. D., & Rulison, K. L. (2011). Peer influences on academic motivation: Exploring multiple methods of assessing youths' most "influential" peer relationships. *The Journal of Early Adolescence*, *31*(1), 13–40.

Mondal, S., Das, S., & Vrana, V. G. (2023). How to bell the cat? A theoretical review of generative artificial intelligence towards digital disruption in all walks of life. *Technologies*, *11*(2), 44.

Mythili, M. S., & Shanavas, A. R. M. (2014). An analysis of students' performance using classification algorithms. *IOSR Journal of Computer Engineering*, *16*(1), 63–69.

Na, K. S., & Tasir, Z. (2017). Identifying at-risk students in online learning by analysing learning behaviour: A systematic review. In *2017 IEEE conference on big data and analytics (ICBDA)* (pp. 118–123).

Ng, A., & Jordan, M. (2001). On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. *Advances in Neural Information Processing Systems*, *14*.

Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review*, *27*(4), 645–656.

Oakley, L. (2019). *Exploring student loneliness in higher education: A discursive psychology approach*. Palgrave Macmillan.

Pataranutaporn, P., Danry, V., Leong, J., Punpongsanon, P., Novy, D., Maes, P., & Sra, M. (2021). Ai-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence*, *3*(12), 1013–1022.

Pathak, A., & Intratat, C. (2012). Use of semi-structured interviews to investigate teacher perceptions of student collaboration. *Malaysian Journal of ELT Research*, *8*(1).

Pesovski, I., Jolakoski, P., Kubincová, Z., Herzog, M., & Trajkovik, V. (2024a). *Digital learning platforms and peer influence: Analysis of performance and collaboration*. *Lecture notes in computer science: Vol. 14607*. Springer (pp. 195–209).

Pesovski, I., Santos, R., Henriques, R., & Trajkovik, V. (2024b). Chatgpt in exam rooms: Preliminary insights into student performance with and without ai assistance. In *EDULEARN24 proceedings, Vol. 1* (pp. 5908–5914).

Pesovski, I., Santos, R., Henriques, R., & Trajkovik, V. (2024c). Generative ai for customizable learning experiences. *Sustainability*, *16*(7), 3034.

Pesovski, I., Santos, R., & Kagan, S. (2023). Perspective matters: A preliminary look into the differences between professor and student perception of test difficulty and their relationship with test performance. In *EDULEARN23 proceedings, Vol. 1* (pp. 5163–5169).

Peza, I. L. (2015). The impact of peer relations in the academic process among adolescents. *Mediterranean Journal of Social Sciences*, *6*(1S1), 127–132.

Phillips, R., Seaborne, K., Goldsmith, A., Curtis, N., Davies, A., Haynes, W., McEnroe, R., Murphy, N., O'Neill, L., Pacey, C., Walker, E., & Wordley, E. (2022). Student loneliness through the pandemic: How, why and where? *Geographical Journal*, *188*(2), 277–293.

Pritchett, L., & Beatty, A. (2015). Slow down, you're going too fast: Matching curricula to student skill levels. *International Journal of Educational Development*, *40*, 276–288.

Pulgar, J., Ramírez, D., Umanzor, A., Candia, C., & Sánchez, I. (2022). Long-term collaboration with strong friendship ties improves academic performance in remote and hybrid teaching modalities in high school physics. *Physical Review Physics Education Research*, *18*(1), Article 010146.

Putra Pratama, M., Sampelolo, R., & Lura, H. (2023). Revolutionizing education: Harnessing the power of artificial intelligence for personalized learning. *Klasikal: Journal of Education, Language Teaching and Science*, *5*(2), 350–357.

Rakovac Bekes, E., & Galzina, V. (2022). Utilizing smart digital technology and artificial intelligence in education for transforming the way content is delivered. In *2022 45th jubilee international convention on information, communication and electronic technology (MIPRO)* (pp. 573–578).

Reang, J. J., & Kaipeng, R. (2022). A study on the influence of peer group on academic performance of students. *International Journal of Creative Research Thoughts*, *10*(11), 2320–2882.

Roy, S., & Garg, A. (2017). Predicting academic performance of student using classification techniques. In *2017 4th IEEE Uttar Pradesh section international conference on electrical, computer and electronics (UPCON)* (pp. 568–572).

Ryan, A. M. (2000). Peer groups as a context for the socialization of adolescents' motivation, engagement, and achievement in school. *Educational Psychologist*, *35*(2), 101–111.

Saheed, Y. K., Oladele, T. O., Akanni, A. O., & Ibrahim, W. M. (2018). Student performance prediction based on data mining classification techniques. *Nigerian Journal of Technology*, *37*(4), 1087–1091.

Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., Gurevych, I., & Fischer, F. (2023). Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning. *Learning and Instruction*, *83*, Article 101620.

Santos, R., Pesovski, I., Henriques, R., & Trajkovik, V. (2024). Predicting bootcamp success: Using regression to leverage preparatory course data for tech career transitions. In *EDULEARN24 proceedings, Vol. 1* (pp. 6181–6190).

Santos, R. M., & Henriques, R. (2023). Accurate, timely, and portable: Course-agnostic early prediction of student performance from lms logs. *Computers and Education: Artificial Intelligence*, *5*, Article 100175.

Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022). Automatic generation of programming exercises and code explanations using large language models. In *ICER 2022: Proceedings of the 2022 ACM conference on international computing education research, Vol. 1* (pp. 27–43).

Serra, R., Martinez, C., Vertegaal, C. J. C., Sundaramoorthy, P., & Bentum, M. J. (2023). Using student-led tutorials to improve student performance in challenging courses. *IEEE Transactions on Education*, *66*(4), 339–349.

Skipper, T. L., & Keup, J. R. (2017). The perceived impact of peer leadership experiences on college academic performance. *Journal of Student Affairs Research and Practice*, *54*(1), 95–108.

Sunday, K., Ocheja, P., Hussain, S., Oyelere, S., Samson, B., & Agbo, F. (2020). Analyzing student performance in programming education using classification. *International Journal: Emerging Technologies in Learning*, *15*(2), 127–144.

Tavakoli, P. (2009). Investigating task difficulty: Learners' and teachers' perceptions. *International Journal of Applied Linguistics*, *19*(1), 1–25.

Teo, T. (2011). Factors influencing teachers' intention to use technology: Model development and test. *Computers and Education*, *57*(4), 2432–2440.

Teye, E., Tandogan, F., & Liu, L. (2024). Instructor-created versus ai-generated learning contents: Evaluation and suggestions. In *Society for information technology & teacher education international conference, Vol. 2024* (pp. 868–874).

The R Core Team (2020). *R: A language and environment for statistical computing*.

Thomas, L., Orme, E., & Kerrigan, F. (2020). Student loneliness: The role of social media through life transitions. *Computers and Education*, *146*, Article 103754.

Timotheou, S., Miliou, O., Dimitriadis, Y., Sobrino, S. V., Giannoutsou, N., Cachia, R., Monés, A. M., & Ioannou, A. (2023). Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review. *Education and Information Technologies*, *28*(6), 6695–6726.

Valentine, J. C., Hirschy, A. S., Bremer, C. D., Novillo, W., Castellano, M., & Banister, A. (2011). Keeping at-risk students in school. *Educational Evaluation and Policy Analysis*, *33*(2), 214–234.

van de Watering, G., & van der Rijt, J. (2006). Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment items. *Educational Research Review*, *1*(2), 133–147.

Venables, W. N., & Ripley, B. D. (2013). *Modern applied statistics with S-PLUS* (4th edition). Springer Science & Business Media.

Vignery, K. (2022). From networked students centrality to student networks density: What really matters for student performance? *Social Networks*, *70*, 166–186.

Vignery, K., & Laurier, W. (2020). A methodology and theoretical taxonomy for centrality measures: What are the best centrality indicators for student networks? *PLoS ONE*, *15*(12), Article e0244377.

Vital, P. R., Sangeeta, K., & Kumar, K. K. (2020). Student classification based on cognitive abilities and predicting learning performances using machine learning models. *International Journal of Recent Technology and Engineering*, *8*(6), 3554–3569.

Vitale, M. P., Porzio, G. C., & Doreian, P. (2016). Examining the effect of social influence on student performance through network autocorrelation models. *Journal of Applied Statistics*, *43*(1), 115–127.

Williams, E., Brewe, E., Zwolak, J. P., & Dou, R. (2015). Understanding centrality: Investigating student outcomes within a classroom social network. In *2015 physics education research conference* (pp. 375–378).

Xu, Z. (2024). Ai in education: Enhancing learning experiences and student outcomes. *Applied and Computational Engineering*, *51*(1), 104–111.

Yeh, Y.-L., Lan, Y.-J., & Lin, Y.-T. R. (2018). Gender-related differences in collaborative learning in a 3d virtual reality environment by elementary school students. *Journal of Educational Technology & Society*, *21*(4), 204–216.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 1–27.

Zhang, Y., Rajabzadeh, I., & Lauterbach, R. (2009). *Student network centrality and academic performance: Evidence from United Nations university*. Technical report, UNU-MERIT.