THE EFFECTS OF COMMUNICATION NETWORKS ON STUDENTS’ ACADEMIC PERFORMANCE: THE SYNTHETIC APPROACH OF SOCIAL NETWORK ANALYSIS AND DATA MINING FOR EDUCATION

Yasaman Mastoory¹, Saeedeh Rajaee Harandi², and Neda Abdolvand²

¹ Faculty of Computer and Information Technology, Qazvin Islamic Azad University, Qazvin, Iran
² Faculty of Social and Economics Sciences, Alzahra University, Tehran, Iran

ABSTRACT

Social networks, as the most important communication tools, have had a profound impact on social aspects of community user interactions and they are used widely in various fields, such as education. Student interaction through different communication networks can affect individual learning and leads to improved academic performance. In this study, a combined approach of social network analysis and educational data mining (decision tree method) was used to study the impact of communication networks, behavior networks and the combination of these two networks on students’ academic performance considering the role of factors such as computer self-efficacy, age, gender and university. The results of this study, which included 139 students, indicate gender is highly prioritised in all three models. Moreover, according to the results all three models had enough confidence level that among them communication networks with higher confidence, accuracy and precision had significant impacts on the prediction of academic performance.

KEYWORDS

Social Networks, Social Network Analysis, Educational Data Mining (EDM), Decision Tree, Academic Performance

1. INTRODUCTION

Development of information and communications technology (ICT), especially the uptake of the internet, has brought about many changes in education. The increasing use of ICT in higher education is strongly related to the student experience [1]. The internet provides extensive information regarding learning tools; these tools facilitate learning processes at all levels [2]. Widespread use of the internet in education has required teachers to acquire the skills that are necessary in order to use the internet and to have an appropriate attitude to learning in the classroom [3]. Use of the internet in training has created e-learning and a web-based training environment that provides a wealth of information regarding the interactions between learning and teaching [4]. Communication and behavior networks between individuals can be effective in both traditional classroom learning (face-to-face learning) and internet-based environments, which can be used in a complementary manner in order to support traditional learning [5]. It is clear that social networks are effective on learners’ performance; therefore, analyzing the different aspects of these networks and studying the status of each student in social networks by comparing their academic performance will determine the effective factors on academic performance. Moreover, providing constructive solutions in this area will increase learning efficiency further. While prior research has made considerable progress in understanding the characteristics of different types of communication media and their relationship towards effectiveness under different situations (e.g. [6][7][8]), there is limited research related to the
effects of technologies on students’ academic performance. With regard to students’ improved achievement through the increasing use of internet in higher education, there is no agreement in the literature on the benefits of internet usage [9]. Moreover, most of the recent studies in this area examined the impact of online social networks and individual factors on academic performance. Despite the importance of social networks in learning environments, not enough attention has been paid to the effects of students’ networks on their learning [10]. Hence, given the importance of communication and networking between students on academic performance, with regard to the role and importance of learning in higher education, more research is needed in this area. Therefore, this question will arise that how communication and behaviour networks will affect the students’ academic performance? Hence, this study aims to measure the effectiveness of social networks (communication and behaviour networks) between students on their academic performance using Data Mining for Education to support reflection on teaching and learning.

With regard to the purpose, this study begins with a brief review of the literature. This is then followed by a brief review of the research methodology chosen for the study. The last part of the research includes the conclusion and recommendation of the study.

2. Literature Review

Information technology (IT) advances have changed the way people record history; this change has also had a considerable impact on human interaction. Social networks, as the most important tools, not only have a profound impact on the social aspects of users, but they also have many usages in various fields such as education. Because of the ability to publish information that the internet provides, developing students’ skills of dealing with digital information is essential [11]. Students’ digital information and communication skills, including different aspects of knowledge, beliefs, attitudes, and ICT values, have become important during the last decade [11]. Network analysis can be used as a method for evaluating interaction and relationships between students and the impact of these relationships on learning [12]. Since social media tools help information sharing, collaboration, and participation, they are necessary technologies for Higher Education [13]. Social network is a set of nodes and a social structure that includes a group of people and their relations, such as friendship, cooperation, and exchange of information [14]. The social networking approach describes the patterns of individuals’ interaction as a relationship diagram [15]. Nodes or members of the network are often individuals or organizations. In fact, each unit that is connected to another can be studied as a node [16]. The social network analysis provides both a strategy and a valuable method for examining the social structure and patterns of communication and interaction between people, organizations, and institutions [17]. Communication networks and behavioural networks are effective on learning in both Traditional classes and Web-based environments. The behavioral network including friendship network and advisory network assesses the hybrid characteristics of people [18]. This network determines the status of each person based on their relationships with others. Friendship networks often seek access to information and knowledge. Student at the center of a friendship network is more likely to help others and get help from them. If students do well in the traditional classroom and in the Web-based community, they have a greater chance for friendship with other students [19]. The advisory networks are more useful than Friendship networks. Advisory networks, include communications between individuals for sharing resources such as information, assignments and guidance that is needed to complete the task. Therefore, centrality in these networks is effective in exchange of resources in order to solve the problems [19]. The social network analysis provides four features of structural insight, that is, the systematic relationship between data, graphics, and mathematical or computational models [20]. Its aim is to connect people with social capital, which thereby helps the development and promotion of science [21]. The changing demographic characteristics of students have led researchers to study new ways to improve the learning process [13]. Rienties and Kinchin [22] examined the interaction and motivation of learners in online
environments that could affect their communication patterns. Martinez et al. [23], by combining qualitative evaluation and social network analysis, studied social interactions in the classroom and indicated that instructional design could force students to shape a new form of cooperation that despite being counterproductive can improve the collaborative learning environment. Dawson, Tan, and McWilliam [24] have also explored social networks as a social activity between students, and have indicated their effect on school performance. Moreover, Yang and Tang [25] investigated the effect of social networks on student academic performance in online education, and indicated that Advice network centrality was positively related to student performance. Cho et al. [26] studied the relationships between communication styles, social networks, and learning performance in a computer-supported collaborative learning (CSCL) community, and suggested that communication and social networks are the main factors in a distributed learning environment. Conley et al. [27] examined the effects of network structure on student characteristics and academic achievement, and indicated that spending time studying with friends has a strong effect on both study time and achievement. Alwagait et al. [28] have studied the impact of social media usage on students’ academic performance and demonstrated that use of social media is not directly related to the Grade Point Average. Vitale et al. [29] investigated the relationship between student relations and their performances at universities, and indicated that informal contacts are related to performance, while formal groups have no such effect. Cen et al. [30] examined the effectiveness of Learning Factors Analysis on learning efficiency and indicated that there was no significant difference in the learning performance of different groups. Zheng and Warschauer [31] studied students’ participation and interaction in an online discussion environment and the relationship between their participation and academic achievements in reading and writing. The researchers suggested that well-designed online discussions among linguistically different upper elementary students can result in increased participation and interaction, thus leading to language and literacy development. From a network perspective, learning is a communication network, which means the exchange of knowledge and information between trainer and learner. E-learning is a technology-mediated learning approach that allows learners to interact through a technology platform. Real-time learning, cross boundary interaction, convenience, cost effectiveness, and open learning are the suggested advantages of e-learning [32]. The learning process is facilitated through participation in social interactions [33]. Increased attention to social network analysis has led to the emergence of several methods for analyzing network members [34]. In learning, graph theory is used to visualize interaction patterns among learners, and data mining is used to enhance the quality of education and decisions in training systems [35]. Educational data mining techniques are another method of knowledge discovery. This method is used to discover unknown and potentially useful patterns from large data repositories [36]. Data mining is used in various contexts and is related to the different methods of discovering the variety of data in learning environments in order to improve students’ understanding and learning [35].

According to Baker studies, data mining methods include predictive techniques (classification, regression and density estimation), clustering, association mining (association rule mining, sequential pattern mining and causal data mining) and data distillation based on human judgment and discovery through modeling [37][38]. Merceron and Yacef [39] using data mining approaches indicated that how the discovery of different patterns through the use of different data mining algorithms and visualization techniques provides a simple pedagogical policy. Ba-Mohammed et al. [40] used the data mining approach to design their Social E-Learning Recommender System (SELRS) for different e-Learning contents based on user interests and social activities; the authors indicated that SELRS suggests the best learning objects to the learners and promotes e-learning. Moreover, Vilaradi et al. [21] demonstrated how using data mining algorithms can help to obtain relevant knowledge from databases obtained from web-based educational systems. Moreover, they suggested that a timely warning to students at risk of poor exam performance could help to improve success in final exams. Minaei-Bidgoli and
Punch [41] used genetic algorithms for data mining optimization in an educational web-based system for classifying students in order to predict their final grade; this study demonstrated that prediction accuracy can be optimized through weighting the feature vectors. Similarly, Yadav et al. [35] used data mining methodologies to study student academic performance and the authors suggested CART to be the best algorithm for data classification.

Prior key studies on social network, data mining and academic performance are briefly listed in Table 1.

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Major Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>[26]</td>
<td>To study how communication styles, social networks can influence the learning performance in CSCL community.</td>
<td>Communication and social networks should be central elements in a distributed learning environment.</td>
</tr>
<tr>
<td>[35]</td>
<td>To study students’ academic performance using data mining methodologies.</td>
<td>CART is the best algorithm for classification of data.</td>
</tr>
<tr>
<td>[40]</td>
<td>To design a Social E-Learning Recommender System (SELRS) for different e-Learning contents using data mining approach.</td>
<td>SELRS enhance e-Learning by recommending the best learning objects to learners.</td>
</tr>
<tr>
<td>[41]</td>
<td>Data Mining optimization in an Educational Web-Based System for classifying students Using Genetic Algorithms.</td>
<td>Prediction accuracy can be optimized through weighting the feature vectors.</td>
</tr>
<tr>
<td>[39]</td>
<td>To synthesize and share the experiences of using data mining for education</td>
<td>Discovery of different patterns through different data mining algorithms and visualization techniques provide a simple pedagogical policy.</td>
</tr>
<tr>
<td>[30]</td>
<td>To examine the effectiveness of Learning Factors Analysis (LFA) on improving the learning efficiency in the Cognitive Tutor curriculum.</td>
<td>There is no significant difference in the learning performance of the two groups in either an immediate post-test or a two-week-later retention test.</td>
</tr>
<tr>
<td>[24]</td>
<td>To Measure creative potential Using social network analysis.</td>
<td>Social network as a social activity between students is effective on school performance.</td>
</tr>
<tr>
<td>[23]</td>
<td>To study the social interactions in the classroom by combining qualitative evaluation and social network analysis.</td>
<td>Instructional design could force student to shape a new form of cooperation.</td>
</tr>
<tr>
<td>[31]</td>
<td>To investigate the effect of social networks on student academic performance in online education.</td>
<td>Advice network centrality was positively related to student performance.</td>
</tr>
<tr>
<td>[25]</td>
<td>To study the Participation, interaction, and academic achievement in an online discussion environment, and the relationship between their participation and academic achievements in reading and writing</td>
<td>Well-designed online discussions among linguistically diverse upper elementary students can result in increased participation and interaction, leading to language and literacy development.</td>
</tr>
</tbody>
</table>
According to various studies on social networks, it is obvious that the rank of student in class has significant effects on individual and organizational results. The structure of social interactions improves access to valuable resources such as career information, etc. [26]. Since student interaction through various communication networks can affect learning and potentially improve their academic performance, study and research on student networks and the role of technology is important. Hence, the purpose of this study is to evaluate how communication networks affect student academic performance.

3. Method: Participants and Data Collection

This study aims to answer this question that how communication and behaviour networks will affect the students’ academic performance to improve people's perception of the effects of network relationships, including communication networks, behavior networks and the combination of these two networks, on students’ academic performance. The population used in this study included students of information technology (IT) engineering and IT management from two universities. A questionnaire, which was adapted from [42][43], was used for data collection. The questionnaire, which was completed by each participant, consisted of five parts: (i) demographic information, (ii) computer experience, (iii) computer self-efficacy, (iv) online and offline network ties and (v) advice and friendship networks based on 7-point Likert scale (ranging from very good to very bad). Data on academic performance was obtained by questioning the teachers as well as from the average scores in the final exams, class presentation and student involvement, which included participation in class and web-based activities. Both the reliability and validity of the questionnaire were examined using expert concepts and Cranach’s alpha, which was 0.739, for the whole scale of the questionnaire (see Table.2). Moreover, the factor loading coefficient was used to examine the reliability and validity of the scales. As indicated in Table 2, loadings of all factors were higher than 0.7 and had a desirable factor loading of higher than 0.6, as stated by Klein [44].
Table 2. Descriptive Statistics, Cranach’s Alphas & Factor loading

<table>
<thead>
<tr>
<th>Bartlet Test</th>
<th>KMO Index</th>
<th>Cronbach’s Alpha</th>
<th>Factor Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>K²</td>
<td>122.189</td>
<td>0.739</td>
<td>Use of the system without involving others 0.814</td>
</tr>
<tr>
<td>F</td>
<td>6</td>
<td>0.735</td>
<td>Use of the system with help of others 0.705</td>
</tr>
<tr>
<td>SIG</td>
<td>.00</td>
<td></td>
<td>Long-term use of the system 0.742</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Use of the system with built-in help facilities for assistance 0.749</td>
</tr>
</tbody>
</table>

3.1. Data Analysis

The data matrix is entered into excel as either zeros or ones (tie = 1, no tie = 0) according to the student’s response to the questionnaire for each class. The most simple and most common metric in social network analysis is composed of many rows and columns, and this is the starting point of almost all network analysis. In a directed graph, the the horizontal axis is the sender of a tie and the the vertical axis is the target of the tie. The UCINET.6 software was used to calculate the intensity of online and offline direct ties for each student. Two numbers, which were the intensity of online direct ties and of offline direct ties for all students, were calculated using Formula (1):

\[
C(P_k) = \sum_{i=1}^{N} a([p_i, p_k])
\]

(Where \((P_i, P_k) = 1, \text{ there is one tie, and there is no tie otherwise})

Using data for direct ties, the distance matrices of online and offline ties were calculated separately for each class. The greater the distance value, the further the person is in terms of their relationship with classmates; in other words, the number of intermediaries between the two is greater. The intensity of indirect ties is calculated through Formula (2):

\[
\text{Intensity of indirect ties} = \sum_{d \geq 2} N_d/d
\]

Decision tree classification algorithms, using Rapid Miner 6.4 software, were used for categorizing variables in order to predict academic performance. Decision tree classification has a hierarchical structure, including internal nodes, leaves and branches. The main advantage of decision trees is that they are easy to create and trees are readily interpreted. In this study, the CART decision tree algorithm and the Gini coefficient were used to create a decision tree. The first step in using decision trees is to determine the number of effective parameters depending on the size of the resulting tree. Preferred parameters are selected according to the Gini index. Table 3 describes the determined parameters and the decision trees are shown in Figure 1.

Table 3. Decision Tree Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of pruning</td>
<td>Based on sub-tree raising</td>
</tr>
<tr>
<td>Confidence factor for tree pruning</td>
<td>0.25</td>
</tr>
<tr>
<td>Binary splits or multiple splits</td>
<td>Binary splits</td>
</tr>
<tr>
<td>Minimum number of instances per leaf</td>
<td>2</td>
</tr>
<tr>
<td>Sampling type</td>
<td>Stratified sampling</td>
</tr>
<tr>
<td>Method has been applied to operate on Data Model</td>
<td>k-fold cross validation</td>
</tr>
<tr>
<td>Operators on Data Model</td>
<td>x-validation</td>
</tr>
</tbody>
</table>
As indicated in Figure 1, all three decision trees are divided into two branches at the first level based on gender; this demonstrates the important role of gender on academic performance. At the second level, model (a) is divided based on online direct ties and university type, model (b) is divided based on the university type and advisory relationships, and model (c) is divided based on online direct tie*friendship and university type.

At the third level, in model (a) Females for whom the online direct tie is greater than 5.5, job performance is high and ends at 1, while Females for whom the online direct tie is smaller or equal to 5.5, the online indirect tie factor is important in decision making. For Azad University Males, online direct tie is an important factor in decision making. In model (b), friendship is important for Males. In model (c) offline direct tie * friendship is important for Females and online direct tie * friendship is important for Males.
At the fourth level, for decisions made by Females, in model (a) online indirect tie is important and in model (c) offline indirect tie * advisory relationship is important; age is important for decisions by both Females and Males. According to the results, computer self-efficacy is the least important for decision making. The results of the decision tree evaluation of all three networks using the recall measure are summarized in Table 4. As indicated in Table 4 given the acceptable confidence level of all three models the communication networks had higher confidence, accuracy and precision than two other models.

$$\text{Table 4. The Decision Tree Results of All Three Networks}$$

<table>
<thead>
<tr>
<th>Model</th>
<th>True positive rate (%)</th>
<th>False positive rate (%)</th>
<th>F-measure (%)</th>
<th>Class</th>
<th>Classification recall (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication Network</td>
<td>94%</td>
<td>28%</td>
<td>83.76%</td>
<td>positive</td>
<td>92.27%</td>
</tr>
<tr>
<td>Behavioural Network</td>
<td>88%</td>
<td>19%</td>
<td>83.72%</td>
<td>positive</td>
<td>86.03%</td>
</tr>
<tr>
<td>Communication and Behavioural Network</td>
<td>90%</td>
<td>28%</td>
<td>81.59%</td>
<td>positive</td>
<td>88.25%</td>
</tr>
</tbody>
</table>

5. DISCUSSION AND CONCLUSION

The development of communication technologies has led to many changes in various fields, including learning. Many traditional learning methods are outdated, and new methods are being adopted to replace them at the individual level and the group level. This has led to a rise in the importance of learners’ social networks [45]. The impact of ICT on learning has generated research that seeks to identify the extent of ICT usage and its effects on students’ academic performance [1]. Hence, this study aims to improve people's perception of how network relationships, including communication networks, behavior networks and the combination of these two networks will affect students’ academic performance considering the role of age, gender and type of university. Moreover, it discusses the role of relationships with peers in determining student success. The results indicate that gender is highly prioritised in all three models and friendship and advisory networks are equally important variables in the prediction of academic performance. Moreover, the impact of all three networks (communication network, behaviour network and the combination of these two networks) on the prediction of students’ academic performance was confirmed. According to the results all three models had enough confidence level that among them communication networks with higher confidence, accuracy and precision had significant impacts on the decision making and prediction of academic performance. Regarding gender differences, our findings are inconsistent with those of Siddiq, Scherer, and Tondeur [11] and the Teaching and Learning International Survey (TALIS) (2013), which indicated an insignificant relationship between gender and ICT use [46]. Since undergraduates have to do more than engage in mere information searches, it has been argued that they need more training in sourcing and evaluating digital information. In fact, evaluating and using information are considered to be important competencies that students should acquire in order to solve tasks, present results, and collaborate with others [11]. Hence, teachers and schools are expected to integrate ICT into their classroom practices to provide students with opportunities to acquire these skills.

The novelty of this study is the use of educational data mining (decision tree method) to analyze students' academic performance. This study goes beyond prior research in that it examines the effects of students’ communication networks on academic performance. Using communication technologies and meeting students at university, in addition to improving students’ interactions,
will enhance communication with other universities. Moreover, it can help researchers seek to understand the impact of technology on academic performance. This study provides a better understanding of the role of technology in explaining academic performance by using the social network theory and the complementary theory and distinction between online and offline communication networks, as well as the impact of the behaviour network. This study describes the impact of ICT on academic performance through communication networks and helps educational institutions to understand the difference between networks and the strengths and the weaknesses of different types of network connections in terms of access to sources. The results of this research can help teachers and students better in grouping students. Moreover, the study will help identify students who need special attention in order to decrease errors and improve their performance in the next semester exams. This study was conducted among students of engineering and information technology management; therefore, its results can be effective in educational programs and teaching students of these majors. Given that this is the first study which examines the effects of students’ communication networks as well as behaviour networks for academic performance in Iran may represent useful information for decision-making in higher education institutes. In addition, with regard to the situation and according to the surveys, the results of this study could be useful in educational programs at other universities; the findings can be useful in fostering students’ communication skills.

According to the results, universities are suggested to develop e-learning management systems in order to promote student academic performance in regular hours. Moreover, it is suggested to create a culture of collaborative learning among students and using this critical thinking in the classroom. Also, teachers should consider a score for using group learning in order to increase the students’ motivation to learn.

Future research should collect data on the way students leverage to the social media, especially those who are at the center of online and offline networks. This will help better understanding the role of supplements or supersedes of two networks. The relationships between students can also be used to design a recommender system based on the frequency of relations.

Moreover, personality and individual characteristic of students such as family status, students IQ, as well as educational facilities and academic classes have not been studied. Furthermore, the teachers and teaching methods were considered equal, therefore, it is suggested that future studies consider these issues and do the same study.

Gathering real data from networks and technology-based educational platforms was impossible that force us to use the questionnaire to collect required data, so future studies should use the integrated system for gathering data.

REFERENCES


Authors

Yasaman Mastoory holds her master’s degree in information technology engineering from Qazvin Islamic Azad University, Qazvin, Iran. She is interested in research in the field of information systems, and ICT, including business intelligence, and educational technology.

Saeedeh Rajaee Harandi holds her master’s degree in information technology management from Alzahra University, Tehran, Iran. She is interested in research in the field of information management systems, and ICT, including business intelligence, knowledge management, and educational technology.

Neda Abdolvand is an assistant professor in Alzahra University and was a postdoctoral research fellow at Tarbiat Modares University. She holds her PhD and MS in information technology from Tarbiat Modares University and a post graduate certificate in information systems from Melbourne University. She is interested in research in the field of information systems & technology, including educational technology, business intelligence, and big data analytics. She has been a member of the Association for Information Systems since 2009.