Social network analysis: friendship inferred by chosen courses, commuting time and student performance at university

Lionel J. Khalil and Marie G. Khair*

Notre Dame University – Louaize, Zouk Mosbeh Main Road, Zouk Mosbeh, Lebanon
Email: lionel.khalil@polytechnique.org
Email: mkhair@ndu.edu.lb
*Corresponding author

Abstract: Our social network analysis (SNA) evaluates the performance of students taking courses with a group of friends versus students used to take courses alone. We evaluate the probability to befriend by comparing the number of courses shared by students with the probability to be assigned in the same classroom randomly based on curriculum constraints. A minimum of courses taken in common is used as a criterion to identify students belonging to a tribe of friends. The main findings are that students in tribes over perform other students by about half point of GPA, and are dropping and repeating fewer courses. Considering student without friends, we measured the impact of the commuting distance on GPA and drop off rate: students with very low GPA and high drop off are mostly students with significantly higher commuting time.

Keywords: social network analysis; SNA; friendship; student performance; GPA; drop off; commuting time.


Biographical notes: Lionel J. Khalil is the Director of Institutional Research and Assessment at Notre Dame University – Louaize, Lebanon. As a veteran of the real-time algorithmic volatility trading with ten years experience in financial modelling and experienced in managing in-house quantitative team. His research agenda has been extensive, ranging from financial quantitative methods to IT and privacy law by developing, implementing and extending algorithms for real-time pricing for sociological analysis and marketing. Alternatively, as a legal advisor, he is heavily involved in associations for data protection legislation, promotion of e-learning and in the combat of online fraud. His current research interests include information systems and social network modelling, and strategic management of technology.

Marie G. Khair is an Associate Professor of Computer Science in the Faculty of Natural and Applied Sciences at Notre Dame University – Louaize, Lebanon. She has a BS in Informatics from the Lebanese University, and a PhD in Informatics from the Aristotle University of Thessaloniki with the title ‘Design and implementation of a secure database system in the health environment’. She has published and presented a number of publications in journal and conferences related to medical database security, distributed database design, and data mining. Her research interests include the areas of information systems design, database technology, medical database systems, and information systems security, data mining and new trends in databases.

This paper is a revised and expanded version of a paper entitled ‘The relation between friendship, commuting time, and student performance: a social network analysis’ presented at International Conference on Communication, Management and Information Technology, ICCMIT 2016, Cosenza, Italy, 26–29 April 2016.

1 Introduction

Social networks and social network analysis (SNA) has been gaining extensive importance lately and this is obvious due to the increased usage of the social network and to the opportunities for analysis and visualisation that the SNA opens. SNA has not been extensively used in the educational sector (Radwan, 2014; Wang et al., 2015; Della Ventura, 2015). In this paper, SNA is used in order to visualise, analyse, simulate, and suggest possible
intervention to enhance students’ performance and reduce students at risk. For this reason, we need to evaluate the impact of friendship as well as the commuting time on the students’ performance and the student retention.

In general, students’ performance is related both to individual-level attributes (such as gender, age and socio-economic status) and to the roles of the relationships with others (such as peers, instructors or friends). Several studies have been conducted to evaluate the impact of instructor on student’s performance (Kim and Sax, 2011; Umbach and Wawrzynski, 2005; Gasiewski et al., 2012). Recently, there has been a study on how to use critical SNA in order to understand course enrollment pattern and the influence of peers in those patterns (González Canché and Rios Aguilar, 2015).

This study aims to determine the impact of the relations of the students on their performance by comparing their results when they took their courses individually in contrast to when they took courses with their friends. In addition, it studied the impact of the distance needed to commute to the university on the students’ performance and students drop out. Being able to visualise the student network, analysis of the sub network allows administrators to identify tribes as well as loosely connected students. Changes to the course offering can be simulated by the administrators in such a way to enhance student connections, and proper interventions can be made to encounter possible negative impact on the student academic progress.

Based on a randomised model we separated students who are randomly allocated in courses versus those who have decided to take their courses with friends. Using SNA we inferred friendship between classmates based on a minimum of courses taken in common. We studied the difference in GPA of the students taking the courses alone versus those who are taking courses with friends. In addition, the impact on the drop off and the repetition of courses was evaluated. From another perspective, we studied the impact of the commuting time on both the drop off and the students’ performance.

This work identified sub-populations at risk to either drop out or change major in the university. This study highlighted the importance of the social integration on the students’ performance. The group at risk is composed by students with heavy commuting time ending up in a university with few high school classmates (i.e., less contacts to have a social integration) and tired by long commuting times. In addition, the work identified a number of means to increase the retention rates such as changing the course offering, increasing group work to raise friendship within students in the same major, or even personal follow up by advisors and students counsellors.

2 Definition and literature review

SNA aims to understand the determinants, the structure as well as the consequences of the relationships of individuals or groups of individuals.

2.1 Social network analysis

The foundations of the SNA are introduced within the framework of our study, and the basic concepts as well as the terms used will be explained.

Ties connect actors among each other’s. There are usually three metrics to be considered in a SNA. The first studies the links which are formed as opposed to the total number of links which can be formed which is known as density. There may be actors with dense ties and others with sparse ties which represents the relationship connections among actors (Scott, 2012; Hannemon and Riddle, 2005; Prell, 2012). Second we can analyse the architecture of the network by identifying sub networks. Third we can study the degree of centrality for specific nodes.

Actors are the social units under study, which we are trying to visualise and analyse their associated connections (Scott, 2012; Hannemon and Riddle, 2005; Prell, 2012). In our application, these social units are students and courses. Saying that two students have taken the same course means that those two students have been in the same classroom with the same instructor during the same schedule.

A social network based on links between actors of the same type is called a one-mode network. A one mode network such as a student network consists of ties between students whenever a student is taking the same course with another student. A two-mode network consists of two different types of actors. In our application, a two mode network consists of linking students and their courses.

There are four types of SNA applications (Sie et al., 2012). Network visualisation which can be seen by sharing the nodes and their connections. Network analysis which is related to the mathematical analysis of interactions and relationships. Network simulation aims at showing and extrapolating interactions in case of possible changes. Network intervention is applying possible changes to enhance some metrics.

Matrices have been used as an efficient tool for representing a small social network. The student-course social networks were presented using a binary matrix having students as rows and courses as columns for all the semesters, and all majors. Let M be the matrix having students as columns and courses as rows representing the two-mode network. The one-mode network for students is $M^tM$ and the one-mode network for courses is $M^tM^t$.

Groups of actors are named pair when a tie exists between two actors, and named tribe when the number of connected actors is more or equal than three. We identified tribes based on the Pajek function Kamada-Kawai energy (Nooy et al., 2011). The function divides the students into pairs, and tribes.

2.2 The blockmodelling of friendship analysis

The goal of blockmodelling is to reduce a large incoherent network to a smaller comprehensible structure that can be interpreted more rapidly. It is based on the idea that units in a network can be grouped according to the extent to which they are equivalent, according to some meaningful
definition of equivalence. One of the main goals of blockmodelling is to identify, in a given network \( N = (U, R) \), clusters or classes of units that share structural characteristics defined in terms of \( R \). They form a clustering \( C = \{ C_1, C_2, \ldots, C_k \} \) which is a partition of the set \( U \). Each partition determines an equivalence relation. A clustering \( C \) partitions the relation \( R \) into blocks: \( R(C_i, C_j) = R \cap C_i \times C_j \) where each such block consists of units belonging to clusters \( C_i \) and \( C_j \) and all arcs leading from cluster \( C_i \) to cluster \( C_j \).

For blockmodelling over our student’s network, we start with the data from the registrar’s office: the detailed list of courses including the classroom, the section, the schedule (MWF or TTh) and the time. The data are presented using a matrix with the students in the rows and the courses in the columns for all of the semesters.

For every combination of a pair of students, we can record which courses were taken by each of the two students and by both. Table 1 illustrates this scenario. In Table 1, the absolute numbers are shown. Alex, and David’s courses heavily overlap; Alex and Ed’s courses do not overlap at all; and Alex and Bob’s courses are somewhere in between. Table 1 is symmetrical in terms of the diagonal (the number of courses that are taken by Alex that are also taken by Bob is the same as the number of the courses that taken by Bob that are also taken by Alex).

<table>
<thead>
<tr>
<th></th>
<th>Alex</th>
<th>Bob</th>
<th>Cindy</th>
<th>David</th>
<th>Ed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alex</td>
<td>15</td>
<td>4</td>
<td>1</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Bob</td>
<td>4</td>
<td>12</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Cindy</td>
<td>1</td>
<td>0</td>
<td>15</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>David</td>
<td>10</td>
<td>9</td>
<td>3</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>Ed</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>4</td>
<td>14</td>
</tr>
</tbody>
</table>

In Table 1 the numbers in the diagonal cells show that ‘all of the courses of Alex are also taken by Alex’ and are therefore meaningless to this discussion about shared courses between students.

The diagonal value should be almost the same because it counts the number of courses that are taken by a student during six semesters. Variations are due to the number of credits per course (1, 2 or 3 credits per course) and the variation of the number of courses that are taken by a full-time student during a semester (between 12 and 18 credits per semester).

This type of table can be easily extended to produce a student-student adjacency matrix that covers all of the students over six semesters. In practice, there is more structure to the matrix in a university than is shown by this theoretical example.

2.3 The structure of course sharing based on blockmodelling

Confronted with such diversity of numbers, two very simple transformations will clearly reveal the main patterns in the data and with them the main message of the structure of the student relationships. Table 2 shows the columns of Table 1 rearranged in a way that the students that share courses are grouped in blocks. The \( \subseteq \) blockmodelling procedure moves rows and corresponding columns around so that one may obtain the substrates or blocks in which the values tend to be either very small (from 0 up to 5) or high (from 6 up to 18). There are some exceptions, which are discussed below; however such exceptions are minor in the context of the whole table. The diagonal cell entries (which indicate the total number of courses that are taken by the students) have been omitted altogether to aid in the visual scanning. The resulting image matrix provides an overview of the overall structure of the network. This also facilitates the calculation in a spreadsheet in which, the diagonal values are not included in the calculation of the threshold.

<table>
<thead>
<tr>
<th></th>
<th>Alex</th>
<th>David</th>
<th>Bob</th>
<th>Ed</th>
<th>Cindy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alex</td>
<td>-</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>David</td>
<td>10</td>
<td>-</td>
<td>9</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Bob</td>
<td>4</td>
<td>9</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ed</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>-</td>
<td>9</td>
</tr>
<tr>
<td>Cindy</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>9</td>
<td>-</td>
</tr>
</tbody>
</table>

2.4 Interpreting the student relationships

In Table 3, we apply a transformation to Table 2 where all values below a threshold of nine are set at 0, and all values of nine and above are set at one. Clear and discernible patterns are revealed in Table 3. Because of the structure that we have imposed, we can easily interpret them as follows:

1. Alex, David and Bob constitute a tribe and not a clique: the pair Alex and David and the pair David and Bob share most of their courses, but Alex and David do not.
2. The pair that is composed of Cindy and Ed shares most of their courses.

<table>
<thead>
<tr>
<th></th>
<th>Alex</th>
<th>David</th>
<th>Bob</th>
<th>Ed</th>
<th>Cindy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alex</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>David</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bob</td>
<td>1</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ed</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Cindy</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>
These patterns of course sharing provide an insight into how students share their courses with their peers. The type of result is stable across majors with hundreds of students.

2.5 Impact of friendship on performance

According to Carolan and Brian (2013), while the individual-level attributes (such as gender, age and socio-economic status) are responsible for 70% to 90% of the variation in the educational outcome, less appreciated are the roles of the relationships with others in shaping the outcomes. Some papers try to overlook the subject and to show the number and the diversity of the research in it (Hommes et al., 2012; Biancani and McFarland, 2013). Nevertheless it is important for the development of 18-22 years olds to have friends (Fullerton and Ursano, 1994).

This period is the end of the time that adolescents spend with their parents and spend time with their peers (Crosnoe, 2000). Upon being in peer groups, students begin developing new identities, roles, and social skills (Strauss and Terenzini, 2007). Those peer groups have emotional functions, as they provide social support and give a sense of belonging (Fullerton and Ursano, 1994). Best friends provide acceptance, trust, intimacy, stability as a material, psychological, and emotional support (Cole and Bradac, 1996; Richey and Richey, 1980; Scholte et al., 2001).

The relationship investment model (Rusbult, 1980, 1983) shows the relationship commitment as the construct of relationship satisfaction and relationship investments. Relationship satisfaction is a function of the rewards minus the costs of a relationship. Relationship investments are the things that the individual has put into a relationship minus the alternatives to the relationship. Carrell et al. (2013) proposes an alternative model based on a mechanism of self-selection into friendship subgroups that are mainly formed of peers with similar abilities: students more often choose to become friends with similarly able peers. Benhabib et al. (2010, p.1159) identifies the impact of learning, group formation, discrimination, and peer dynamics on student performance.

Studies have shown that the change to a university has a significant impact on friendship. Relationships that are built before attending a university tend to decline in quality and quantity during the first year (Shaver et al., 1985). Most students report finding a new close friend within the first month of attendance at a university, although the majority of such friendships do not last more than one year (Paul and Brier, 2001; Paul and Kelleher, 1995).

Recently disparate work has been done to study in depth the SNA in the educational sector (Sacerdote, 2011; Radwan, 2014; Wang et al., 2015; Della Ventura, 2015; González Canché and Rios Aguilar, 2015; Khalil and Khair, 2016) in comparison to the studies performed in other sectors. Several authors studied the influence of network association on the success, or on the student’s research potentials, or on student integration and persistence, or even on the distribution of knowledge (Liu and Zhu, 2015; Scott, 2000; Rulke and Galaskiewicz, 2000; Demirbas and Demirkan, 2007; Mills and Fullagar, 2008).

2.6 Impact of the commuting time on student integration

Commuting time can be determined as the time spent by a person to move forth and back between home and work (Choudhary et al., 2015). This determinant has been studied to evaluate its impact by Kobus et al. (2015), Tigre et al. (2017), Mathews and Mulkeen (2005) and Blaney and Mulkeen (2008). They found a strong and consistent evidence that duration of commuting time has a negative causal effect on academic achievement. In addition, within the Eurostudent Project (Orr et al., 2012), a student survey was performed in order to compare the socio-economic background and the living conditions of European students. The main output of the survey was to evaluate and compare the students commuting time and its implications.

It was noted that the commuting distance of students to college has an impact on their integration, and the long travel times meant they could not wait around after lectures and therefore were more likely to find it harder to make friends. In addition, it was found that university students with long commuting times were the most likely group to be non-completers when compared with other residence-types.

In addition, it was noted that the more the commuting time, the less frequent the student visits the university and the lower the students’ GPA.

3 Inferring friendship based on courses scheduling

SNA is used to determine the impact of the relations of the students on their performance by comparing their results when they took their courses individually in contrast to when they took courses with their friends. In addition, the impact of the commuting time on the student’s performance and students’ retention is evaluated.

3.1 Aim of the study

The aims of this study are to visualise sub networks (pairs, tribes) within the network of students and to compare their performance when they take their courses alone versus when they take courses with friends. Based on the distinction between tribe and non-tribe students’ performance, analysis of the positive impact of friendship on their academic progress is evaluated. This analysis aims at evaluating the importance of several criteria (major, gender, campus, GPA...), and identifies the density of the subnetworks. Hence, administrators and academic consultants can use the visual network as well as its network metrics for further analysis and simulations in coming semesters.

We have defined a model of friendship and by applying the model, we defined a minimum of number of shared courses that gives a high probability (90%) to specify whether a student is taking courses with friends or not. This threshold is about half courses taken in common based on the distribution of common courses. We study the difference in
GPA among students taking courses with friends compared to those taking courses alone. Then considering students taking course alone, we studied the impact of proximity to home on the students’ success by comparing the commuting time vs. the students’ GPA.

3.2 Modelling friendship with classmates

We defined a friendship model to separate students who are randomly in courses together and those who have decided to be with friends. We inferred friendship between two classmates when those two students have taken a high number of courses (i.e., same classroom, same schedule, and same section) together above the constraints imposed by the curriculum. The probability to be in the same course is decomposed in the probability to be in the same classroom due to curriculum constraints, and the choice to be in the same course. We have evaluated the probability to be friend by subtracting from the number of courses shared by students the probability to be assigned in the same classroom randomly based on curriculum constraints.

To measure the distribution of the number of courses taken by two students without the intent to be together, we build a theoretical model. We have considered the network of the undergraduate students linked by the number of courses they have taken in common during six consecutive semesters. We have decided to keep the real list of students and the real list of courses (including section and semester) taken, then we have allocated randomly the student in a course according to the ratio of the number of students initially in the course (including section and semester) over the total number of students enrolled in the course during the six semesters. The created random matrix of course-student has the following properties:

- the students are the real number of students
- the courses are the real courses taken by the students
- the order of the courses taken by students is random
- the session of a course (i.e., section and semester) with p students will have p students in the classroom.

Using a Monte Carlo methodology we have created several random matrices of courses-students and we have calculated the distribution of the number of links between students.

Our study is based on a social network of 1,248 students with a number of courses (density) that ranges from 0 up to 41 shared courses. The number of students sharing more than nine courses in reality is ten times higher than the number of students sharing more than nine courses in the average of the random matrix distribution. To ensure a probability of 90% that the students are not in class randomly we fixed a minimum of 14 shared courses between students during six consecutive semesters. In practice students sharing more than 14 courses during those six semesters have on average 57% of their courses in common.

4 Methodology

4.1 Background on course scheduling

This university typically schedules courses in two ways: those that meet two days a week (TTh) for approximately 75 minutes and those that meet three days a week (MWF) for approximately 50 minutes. Both types of courses offer the same number of credit hours and class time; the three day per week schedule spreads out the materials over more days. The course schedule is published before advising and registration begin for each semester and summer session. It lists each class that is offered, its time, location, and instructor. Students schedule their own courses using different sources of information such as word-of-mouth, social networks, friends or advisors and according to personal priorities: the reputation of the instructor, course content, time schedule, and curriculum constraints (Dills, 2007).

4.2 Characteristics of the selected cohort for the case study

We have selected a cohort of undergraduate students from six different majors in a 7000-student university. The six majors selected are chosen from different faculties and are selected out of the top 15 enrolled majors for the years 2012–2014. The first major selected is the top one major, Bachelor of Architecture from the Faculty of Architecture. From the Faculty of Engineering, two majors are chosen, BE in Civil Engineering and BE in Mechanical Engineering. As from the Faculty of Business Administration, the Bachelor of Business Administration is chosen. From the Faculty of Natural and Applied Sciences, the BS in Computer Science is chosen. Lastly, BA in Communication Art-Radio TV is chosen from the Faculty of Humanities. During the six semesters of the study a full time student is expected to complete 24–40 courses. The number of sections of the same course given during the same semester varies from one up to 16 sections per course per semester with an average of around two sections per course. Students have bigger opportunities to define their schedule by choosing among multiple sections.

<table>
<thead>
<tr>
<th>Major</th>
<th>Average credits</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arch</td>
<td>85</td>
<td>12</td>
</tr>
<tr>
<td>BBA</td>
<td>78</td>
<td>14</td>
</tr>
<tr>
<td>CARTV</td>
<td>81</td>
<td>14</td>
</tr>
<tr>
<td>CIV</td>
<td>88</td>
<td>11</td>
</tr>
<tr>
<td>CS</td>
<td>77</td>
<td>13</td>
</tr>
<tr>
<td>Mech</td>
<td>88</td>
<td>11</td>
</tr>
</tbody>
</table>

The number of courses that are taken each semester depends on the number of credits that are given per course. In the Architecture, Civil, and Mechanical degrees, several lab courses are worth only one credit; thus, the distribution is approximately 85, 88 and 88 respective credits that are
taken for six semesters; while in Business, Computer Sciences and Arts, the majority of the courses are at three credits, so the distribution is centred approximately to 78, 77 and 81 respective credits. The standard deviation ranges from 11 to 14 (see Table 4).

Approximately 30% of the courses are taken from a pool of general education requirements (GERs) courses that are available for all students in the university. The core courses of a given major are chosen in a pool of specialisation courses. Furthermore, some students must take some remedial courses.

Considering the respective cohort of students in the studied majors along six semesters, we found that few courses are taken in the same time by the selected cohort of students in the same major; about 10% of the courses are taken by all students and repeated, and 60% of the courses (from 0% up to 60% percentile) are taken by less than 20% of the students. In principle, students majoring in the majors mentioned above, tend to collaborate among each other more than other majors (Paul and Brier, 2001; Paul and Kelleher, 1995). Our hypothesis is to evaluate how much the work in groups affects the whole group GPA as well as students’ individual GPA. Upon choosing students from the six selected majors as a population, we selected for each major, students that have taken courses during all the six semesters inclusively from fall 2012 until spring 2015. The selected population consisted of a total of 1,248 students distributed on the six majors: 343 in Civil Engineering, 311 in Architecture (Arch), 259 in BBA, 199 in Mechanical Engineering, 87 in CS and 49 in Communication Arts-Radio/TV (CARTV) students.

4.3 Extracting friendship network from the selected cohort

The next step was to manipulate the population in order to extract the friends’ network. To determine the elements of the tribes, the treatment of the data was done as mentioned previously on Pajek (Nooy et al., 2011), and R. Based on the student-course matrix, we build the student network by applying a threshold at 14 shared courses in six semesters: the initial matrix has been cleaned with the diagonal value down to 0 and density relations with less than the threshold have been put at 0. Only 243 students out of 1,248 students share more than 14 courses with some classmates.

We identified tribes based on the Pajek function Kamada-Kawai energy (see Figure 1). The function divides the students into pairs, and tribes.

Figure 1  Tribes and pairs in architecture
5 Findings

Tribes characteristics were analysed depending on GPA, gender, type of courses. The following section will present the findings of the study.

5.1 Impact of tribes characteristics on students’ performance

Cooperation looks more important in CARTV, Mechanical Engineering, Civil and Architecture with a ratio of about 25% of students being identified in a tribe or a pair of friends. BBA and CS have a lower number of tribes and pairs. 86% of the tribes are from different high schools especially in small campuses (goes up to 96%). This confirms the hypothesis of Shaver et al. (1985) that most of the friends at university are new friends.

Gender impact in tribes is strong. The results show that females in tribes perform better in individual courses than females in pairs in all majors except CS and Civil Engineering. The difference is mostly significant in Mechanical Engineering where it is 0.30 for individual courses. As in common courses for females, the grade average is better in pairs than in tribes in all majors except CARTV and CS majors (0.15 and 0.41 respectively). The result for males is similar in individual and common. In all the majors, except Architecture and Mechanical Engineering, males are performing better in tribes. This difference is very significant in BBA and CS majors where the difference of average GPA is 0.94 and 0.81 for individual and common respectively in BBA and 0.80 and 0.67 for individual and common in CS major.

Comparing in-tribe and non-tribe students we see an impact on withdrawal and repetition of courses. The percentage of withdrawal is higher in students of non-tribes, having a percentage of 15% while tribes have 6%. The difference of percentages shows a higher difference in CS (16%) than other majors. The percentage of repeating courses three times or more for all students is higher than that of students belonging to tribes. The average of semester GPA of students in tribe is over performing the one of non-tribe students. As an overall comparison of the average semester GPA between tribes and non-tribe students, we observe a higher GPA of students in tribes (2.86 / 4) than students in non-tribe (2.31 / 4). According to gender, we observe a higher average GPA in females in tribes (3.09 / 4) than males (2.73 / 4) in tribes, and the same for non-tribe students (2.65 in females and 2.19 in males). Moreover, the difference in average GPA between tribes and non-tribes, according to gender, is higher in males (0.54) than in females (0.45). This shows that although females perform better in tribes, males in tribes are more efficient than females in tribes.

5.2 Impact of commuting time on students’ performance

Looking at students without friends, we have identified a sub population of students at risk. There is a strong correlation between students that have no friends as classmates, and high commuting time (46 min, i.e., 30% more than the average) and low GPA (GPA 0–1) and high drop off rate.

For very low GPA (0–1) there is a 30% dropping rate, for low GPA (1–2) there is a 12% dropping rate. Drop off students with very low and low GPA have higher average commuting time (40–50 min) than average students. For high GPA (3–4) there is 8% dropping rate. Those drop off students with high GPA have slightly the same commuting time (35 min) than average GPA students (see Table 5). As a result we can assume only a correlation between dropping off students with very low and low GPA and commuting time.

As further work, we might need to analyse deeper the reasons of this correlation: one hypothesis is that they are tired commuting and it is impacting the GPA.

Table 5 Relation between the GPA, commuting time and drop off rate

<table>
<thead>
<tr>
<th>GPA</th>
<th>Drop-off rate</th>
<th>Average commuting time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–1</td>
<td>30%</td>
<td>46</td>
</tr>
<tr>
<td>1–2</td>
<td>12%</td>
<td>37</td>
</tr>
<tr>
<td>2–3</td>
<td>4%</td>
<td>34</td>
</tr>
<tr>
<td>3–4</td>
<td>8%</td>
<td>34</td>
</tr>
</tbody>
</table>

6 Conclusions

SNA was used in order to infer friendship between classmates based on their course distribution. We define a friendship model to identify the friendship effect in a social network as a long tail distribution of the real distribution of courses compared to a random distribution of the courses among students. As a case study we consider that students are friends if they take about 50% of their courses together which means friends are students sharing more than 14 out of 30 courses during six semesters. This work identified a population of about 20% of the students which consider friendship as a fundamental criterion to choose a course.

The main conclusions which we can draw are the following. We confirm an already known result that most of the friends’ tribes are formed by new friends who are not coming from the same high school (Shaver et al., 1985; Paul and Brier, 2001; Paul and Kelleher). All students tend to have better GPA, to drop lesser and to repeat courses lesser whenever they are used to take courses with their friends. In the same time we have identified a population at risk to drop off, those students who have no friends, and long commuting time are more likely to have very low GPA and a high drop off rate.

As further work it may be of high importance to further analyse the relationships among the tribes of classmates and whether they are only in classroom or even outside by referring to their social networks and checking whether classmates’ tribes continue to be friends.
We found that those drop off students with high GPA have slightly the same commuting time (35 min) than average GPA students. As a result we can assume only a correlation between dropping off students with very low and low GPA and commuting time.

This research can conduct to more sophisticated questions regarding the impact of the relationships among students on their performance: additional studies can be conducted to identify the reasons for the formation of tribes and to follow the life of the tribe over time. The work could analyse the relationships among the tribes and whether they only exist in the classroom or outside of the classroom by referring to the social networks and determining whether classmates’ tribes continue to be friends.

We could also study the impact on performance of the relationships between students and instructors on one side and university administrators on the other.

References


