Exploratory Social Network Analysis with Pajek: Case Study on Student Group Performance

Lionel Khalil, Marie Khair, Tina Daaboul, Marie-Joelle El Hajje

Abstract— **A**ssessing, evaluating, and predicting student's performance has always been a major research part of the academic workers aiming for academic excellence. One of the main issues related to this is identifying the major factors influencing it positively or negatively. A major related question, and not yet enough researched, is whether students taking courses together has any influence on their outcome. To answer this question, this paper presented a study of the student's behavior in course selection. We described the population considered under study, and explained the methodology followed to accomplish it. Finally, we analyzed student's performance of individual versus common courses taking several dimensions like gender, pairs or tribes, passing level (very good, good or failing). The main finding is that student cooperation in groups improves the probability of passing courses. Nevertheless, good performing students are affected slightly negatively by their contribution to the group.

Keywords—Social Network Analysis, Pajek, GPA, Student Performance.

I. Introduction

THIS paper presents a study performed on university students for observing their behavior in courses they have taken. The aim is to determine the level of performance of the students when they took their courses individually in contrast to when they took courses with their friends. Accordingly, the difference in GPA of the students in courses taken individually is analyzed and compared to that of courses taken in common. In order to reach a reliable conclusion, the relations of students between each others were mapped using Social Network Analysis and a series of steps have been followed to explore these relations, in addition to several tools such as Pajek and R, which were used to visualize the results and help us better understand the impact of the course selection on the student performance.

L. Khalil is with Notre Dame University – Louaize, Lebanon (phone: +961 9 208 118; fax: +961 9 218 771; e-mail: lkhalil@ndu.edu.lb).

M. Khair is with Notre Dame University – Louaize, Lebanon (e-mail: mkhair@ndu.edu.lb).

T. Daaboul is with Notre Dame University – Louaize, Lebanon (e-mail: tdaaboul@ndu.edu.lb).

MJ. El Hajje is with Notre Dame University – Louaize, Lebanon (e-mail: mahajje02@ndu.edu.lb).

.

II. Literature review

Social Network Analysis SNA is the study of relationships of individuals or groups of individuals. SNA has been heavily researched and successfully tested in several fields mainly related to social sciences, human disease, scientific collaboration, business, medicine and many others. However, very little work has been done to study the SNA in the educational sector [1, 2, 3].

In general SNA involves two major directions: one direction that seeks to understand what influences the formation of relational ties in a given population, and another direction seeks to understand the impact of the relations within a SNA on a specific outcome either at an individual or at the population level [1].

When it comes to the higher educational sector, there are several disparate publications related to it. Some papers try to overlook on the subject and to show the number and the diversity of the research in it [4], others study the influence of network association on the success, or on student's research potentials, or on student integration and persistence, or even on the distribution of knowledge [1, 3, 6, 7].

In our study, a cohort of architecture students was traced. The architecture major was chosen for several reasons. First, the bachelor of architecture is very selective (out of 500 applicants only 300 are enrolled) compared to other majors (3 or 4 years). Third, the success rate within major courses is low with an average of 20% of failing students. 11% of the courses are repeated two times and 4% are repeated three times and more.

To obtain the degree of bachelor of architecture, a student must complete about 57 courses with an overall GPA of at least 2.0/4.0 and a minimum cumulative GPA of 2.3/4.0 in the Core and Major requirements.

Among the 57 courses, 43 courses are compulsory courses and 14 are elective courses. The number of sections of the same course given during the same semester is given in the Table 1 shown below. Only 16% of the compulsory courses are given in only one section. This gives the student bigger opportunity to choose among multiple sections.

In principle, students majoring in architecture tend to collaborate between each other more than other majors [8, 9]. Our hypothesis is to evaluate how much the work in groups is

ISBN: 978-1-61804-320-7

affecting the whole group GPA as well as the students' individual GPA.

Number of Sections per semester	Compulsory Courses	Elective Courses
One section	16%	32%
Between 2 and 3	34%	30%
More than 4	50%	38%

Table 1 Number of Sections per Semester

III. METHODOLOGY

In order to be able to evaluate, assess, and compare the performance of the students between courses taken individually and courses taken in common, several steps were taken. The rest of this paper will cover first the procedures and steps taken to accomplish the study. Next, the treatment of the data using mainly Pajek and R in addition to other tools will be mentioned. In addition, the sample size that was taken and its validity will be presented. Finally, we will illustrate the findings and identify the conclusion.

A. Procedure and Task

Upon choosing architecture students as a population, the study will be based on evaluating whether the performance of architecture students at a private university in taking courses individually is better/worse/ or same than taking common courses with friends. Accordingly, the list of common students in architecture that have taken courses during all the six semesters inclusively from fall 2012 till spring 2015 was taken. The next step was to manipulate the population in order to get a valid sample.

B. Treatment of Data Two-mode network:

In social network analysis, matrices have been used as an efficient tool for representing a small social network and for computing results on its structure. In addition, matrices offer visual clues on the structure of small and dense networks. A matrix is a two-way table containing rows and columns. The intersection of a row and a column is called a cell of the matrix.

The selected population consisted of 311 architecture students. This data was presented using a binary matrix for each semester which shows the courses each student has taken for each semester separately. The six matrices were added to obtain a large matrix of all the courses the students have taken in the six semesters (Student-Course relation). After obtaining the matrix, R and Pajek were used for manipulating and visualizing the matrix [10]. Thus, the initial matrix is obtained which includes all the 311 student-course relation for the six consecutive semesters.

In Pajek language, affiliation networks consist of at least two sets of vertices such that affiliations connect vertices from different sets only. In the initial matrix presented (student-course relation) there are two sets, which are Students and Courses. Affiliations connect Students to Courses, not directly Students to Students. Figure 1 shows a fragment of the Students/Courses network. This type of network is also called a two-mode network or a bipartite network, which is structurally different from the one-mode networks.

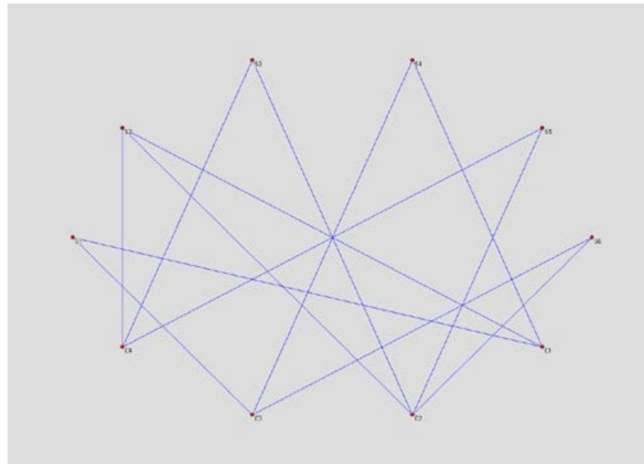

Figure 1 Students/Courses Network Example

We followed the solution commonly used to change the two-mode network into a one-mode network of students that attend common courses, which can be analyzed with standard techniques.

C. Treatment of Data: One- mode network:

Pajek has special facilities to derive a one-mode network from the two-mode network. The submenu Net>Transform>2-Mode to 1-Mode contains commands for translating two-mode into one-mode networks. A one-mode network can be created on each of the two subsets of vertices. There is an alternative way using "R" on the matrix form of the net.

By convention, vertices of the first subset Students are called rows, whereas columns refer to the second subset Courses. These terms are derived from matrix notation. Let M be the matrix having Students as rows and Courses as columns. The one-mode network for students is $M*M^T$ and the one-mode network for Courses is M^T*M.

In $M* M^T$ representing the one-mode network for students, called student-student matrix each row and column represents one vertex of the network, for instance, the first (highest) row and the first (left) column feature on Student. Social cohesion is linked to the structural concepts of density and connectedness. Density refers to the number of links between Vertices meaning how many courses students take together. A number in the cell indicates the number of courses shared with another student and a zero cell means that there is no course in common. "R" programming language was used in order to obtain the student-student matrix by calculating M^T*M.

D. Threshold and Sample Size:

The Network has a density range from 0 up to 26 courses in common. Figure 2 shows the log-linear relation between the density and the number of vertices (students in relation).

Any Student sharing a minimum number of courses with another student is probably a friend of him or her. This hypothesis can be validated using Facebook where it was checked whether students belonging to a certain tribe or pair are really friends.

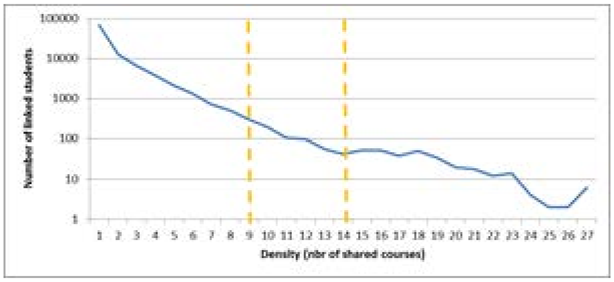

Figure 2 log representation of density per vertices

As a density threshold to deciding whether a group of students are friends or not, the number of shared courses between them was specified. Two density thresholds have been chosen, the first was 14 or more courses in common and the second was 9 or more courses in common. The density threshold 9 or more and 14 or more were chosen because students are allowed to take a maximum of 5 to 7 courses each semester. Thus, a student can take around 10 to 14 courses in a year; so students that share minimum 9 or 14 courses have been together for a minimum of a year.

Two matrices have been designed: a high density matrix with a range of density from 9 to 13 and a very high density matrix with a threshold at 14. 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. The high density matrix is reduced to 94 students and the very high density matrix is reduced to 81 students. To avoid overlap between the two Networks, we first consider the tribes and pairs in the very high density network and then only the pairs from the matrix with high density that share from 9 up to 13 courses was taken into consideration. Note that one student-student relation is only shown once. This means that if this relation is shown in a very high density matrix (more than 14 common courses), it will not be shown in the high density matrix (9 to 13 common courses). In addition, if two students are part of a tribe, they will not be shown in a pair.

Hence, the high and very high density matrix (14 courses or more) were upload on Pajek to get the pairs and tribes.

E. Validity of the Sample:

The sample size required for a specified level of confidence in the result with a specified degree of sampling error is calculated based on a formula in relation to a population of a specified size [11]. But Cook et al [12] and Draugalis et al [13] point out that response representativeness can compensate a low sample. Several case studies [14,15,16] have presented low samples and identified a small effect of error indicating adequate representativeness of their sample. With 26% ratio for the sample of minimum 14 courses in common and 57% for the sample of 9% in common, the two samples fulfill and are far above liberal conditions of 10% sampling error and 80% confidence level as per defined by Nulty [17].

F. Obtaining Tribes:

The student-student matrix was placed in Pajek that transforms it into a network where it was manipulated into groups of students i.e. showing the students sharing the same courses as tribes. There are several techniques to detect cohesive subgroups based on density and connectedness. We identified cliques or complete subnetworks based on the Pajek function Layout>Energy>Kamada-Kawai>Separate Components. Pajek divided the students into Pairs which means they share two courses together and into Tribes where they share three or more courses together.

Figure2 shows the one-mode network of students that is derived from the network in Figure 3. It is constructed in the following way. Whenever two Students share a Course in the two-mode network, there is a line between them in the one-mode network.

When students share multiple courses multiple lines are replaced by a valued single line indicating the original number of lines between two vertices, in other words the number of courses in common of two students or the density of the relation.

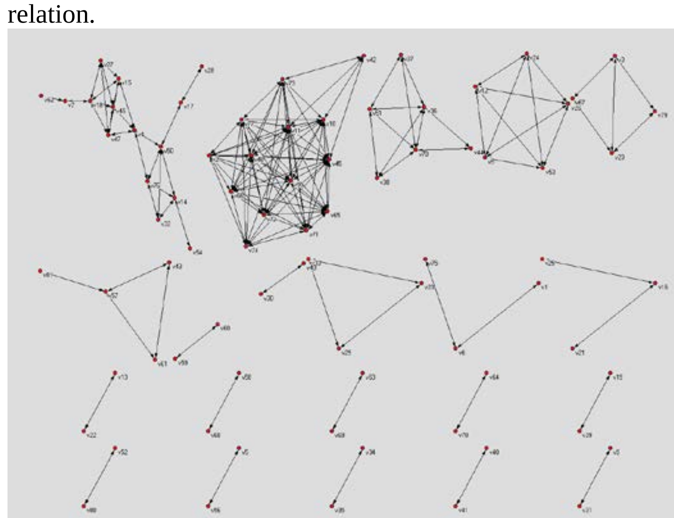

Figure 3 Tribes and Pairs in high density Network

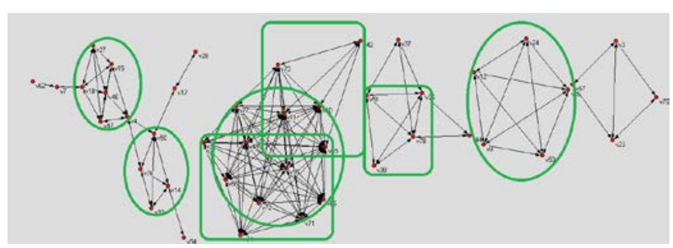

Figure 4 Cliques in the High Density network

Independent students are the ones connected with less than 14 courses (respectively 9) with other students. Pairs of students are students who are connected with bidirectional arcs with a density of more than 14 (respectively 9). Tribes are group of students who are connected with more than 14 courses two by two (respectively 9). Cliques are group of students who are all connected with more than 14 courses (respectively 9). A clique is a set of vertices in which each vertex is directly connected to all other vertices In Figure 4 there is only one example of clique (clique of 5 nodes); other tribes (clusters) are often cliques with additional students linked to some members of the tribes.

D. Validity of Tribes:

As mentioned above, a tribe holds set of students that have taken courses together which means that there might be a friendship relation between them. In order to make sure that our assumption is right, we have checked the friendship relation of the tribes on Facebook. Below is the result of our analysis. Table 2 shows 19 pairs out of 22 that have a 100% match with Facebook network and an overall match of 94%. Some tribes were identified on Facebook as friends from their friend list and some were identified from their common pictures. The 9 pairs denoted by “undetermined” are the ones that were either not found on Facebook or their friend list isn’t viewable.

Tribe Name	Number of Stud in Tribe	FB Friends Match	Percentage of Match
Tribe 1	3	3	100%
Tribe 2	3	3	100%
Pair 1	2	2	100%
Pair 3	2	2	100%
Pair 4	2	2	100%
Pair 9	2	2	100%
Pair 10	2	2	100%
Pair 11	2	2	100%
Pair 12	2	2	100%
Tribe 4	4	4	100%
Tribe 7	6	6	100%
Tribe 9	15	15	100%
Pair 13	2	2	100%
Pair 14	2	2	100%
Pair 15	2	2	100%
Pair 16	2	2	100%
Pair 17	2	2	100%
Pair 18	2	2	100%
Tribe 8	5	4	80%
Tribe 3	3	2	67%
Tribe 5	14	9	64%
Tribe 6	4	2	50%
Percentage of Match			94%
Pair 2	2	UNDETERMINED	N.A
Pair 5	2	UNDETERMINED	N.A
Pair 6	2	UNDETERMINED	N.A
Pair 7	2	UNDETERMINED	N.A
Pair 8	2	UNDETERMINED	N.A
Pair 19	2	UNDETERMINED	N.A
Pair 20	2	UNDETERMINED	N.A
Pair 21	2	UNDETERMINED	N.A
Pair 22	2	UNDETERMINED	N.A

Table 2 Validity of Tribes and Pairs

G. Adding demographic data

Lastly, demographic data for each student in a tribe and pair was collected such as gender, grade, campus, and others. For each pair and tribe we calculated the average grade for individual and common courses, standard deviation, and percentage of courses taken. The data was also analyzed depending on gender. The following section will present the findings of the study.

IV. FINDINGS

A. Gender Inference:

1) *Performance in Individual and Common Courses According to Gender:*

To begin with, the first finding concerns the performance of students regarding their gender. The goal is to identify whether female and male architecture students perform better in individual than common courses. Accordingly, the average and standard deviation of grades were calculated. The results show that females perform slightly better in individual work rather than common, having an average GPA of 2.9 and 2.86 for individual and common courses respectively. As for males the result showed that they perform better in common courses than in individual with an average GPA of 2.41 and 2.56 in individual and common respectively. The standard deviation of the GPA is slightly the same for females in both types of courses (0.98 and 0.96) and for males the standard deviation of the GPA is higher for individual courses (1.11) than in common courses (0.94).

2) *Performance in Tribes According to Gender:*

Another finding concerning the gender is the level of performance of students in tribes versus students in pairs according to the gender. The results show that females in tribes perform better in individual courses than females in pairs (3.09 and 2.83 for tribes and pairs respectively). As in common courses for females the average of grade is better in pairs than in tribes (2.82 and 2.95 for tribes and pairs respectively). The result for males is better in pairs (2.43) than in tribes (2.36) for the average of individual courses and the result is the same in common courses (2.56).

B. *Percentage of Passing in Individual and Common Courses*:

The percentage of passing and failing a course was calculated for both individual and common courses to see where the students are more successful. There are three categories of grades an architecture student can get on a course, either a passing grade (A and B range), a non-passing grade (C and D range), or a failing grade (F or withdrawal). Accordingly, the percentages of the three types of categories of grades were calculated for both individual and common courses. The results show a 79% of passing in common courses which is slightly higher than percentage of passing individual courses (75%). As for the non-passing grades the percentages are almost the same (17% and 16% for individual and common courses respectively). The percentage of failing is slightly higher in individual than common courses (8% and 5% respectively).

C. *GPA Difference of Common minus Individual Courses:*

Figure 5 below shows the GPA difference of common minus individual courses. The blue line represents the linear evolution of the GPA difference, the red line represents the outliers and the green line is the boundary. The results show that the average difference of GPA is negative 0.09. Figure 6 shows that regardless the percentage of common courses, GPA difference is between -0.04 and -0.02 excluding the outliers.

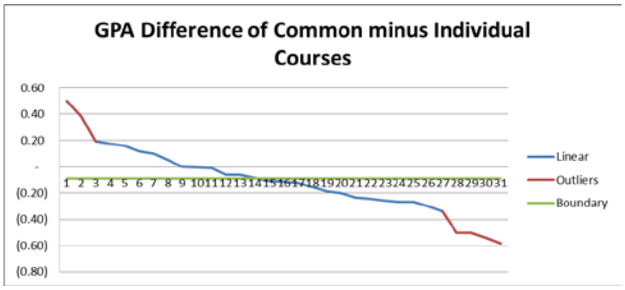

Figure 5 GPA Difference of common minus Individual Course

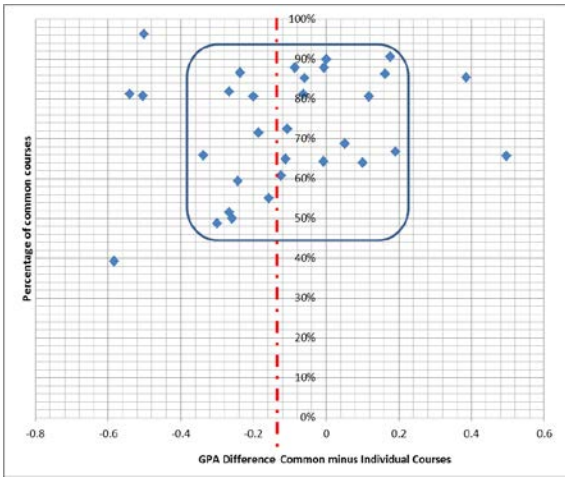

Figure 6 GPA Difference with percentage of common courses

D. Difference between Standard Deviation (Common minus Individual courses):

Figure 7 shows the difference between standard deviation of common and individual courses. Students' grades are closer to each other when they work together in 78% of cases rather than when they work alone. This result is independent of the percentage of common courses.

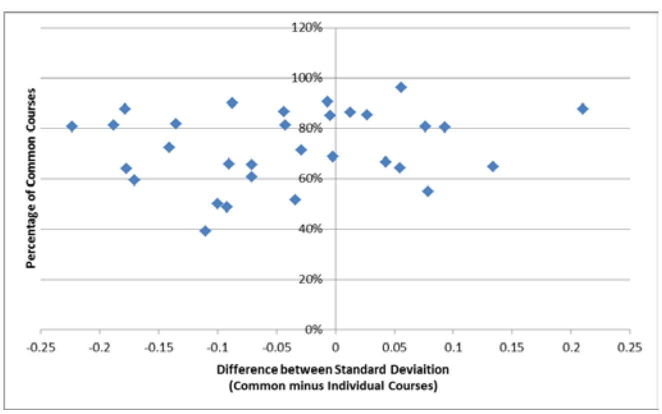

Figure 7- Difference in the Standard Deviation of GPA

V.CONCLUSION

The main conclusions which we can draw are: first that males perform better in groups while females perform the same or worse; the standard deviation for the grades in group course is lesser than that in individual courses meaning that the students, once in groups help each other; third the number of students belonging to groups seems low (only one fourth of the students) despite the fact that they have the choice to take common courses; and finally the percentage of students passing the courses is higher in common courses than in individual courses. The main finding is that student cooperation in groups improves the probability of passing courses. Nevertheless, good performing students are affected slightly negatively by their contribution to the group.

The limitations of this study are that application of subsequent six semesters of architecture students in one university can be applied to a local context. Findings report new factors on the GPA with implications to only undergraduate students. The generalizability of findings is limited because of small sample size and area selected for sampling.

In further works, the study has to be generalized to all majors of the university. This study has not examined the decision making abilities of students within a tribe, offers an area for future research information gathered from this study and conclusions made might need further research in other university as well. The study is quantitative in nature; therefore requires further exploratory analysis in order to address remaining research questions on elements which have a significant influence on students' choice of a course.

REFERENCES

[1] D. Grunspan, B. Wiggins, and S. Goodreau. "Understanding classrooms through social network analysis: A primer for social network analysis in education research." CBE-Life Sciences Education 13.2, 2014, pp. 167-178.

[2] K. Akers, and K. Bradley. "Examining graduate committee faculty compositions-A social network analysis example.", http://www.uky.edu/~kdbrad2/Kate.pdf, last visited May 27, 2015.

[3] J. Hommes, et al. "Visualising the invisible: a network approach to reveal the informal social side of student learning." Advances in Health Sciences Education 17.5, 2012, pp. 743-757.

[4] S. Biancani, and D. McFarland. "Social networks research in higher education." Higher education: Handbook of theory and research, Springer Netherlands, 2013, pp.151-215.

[5] X. Liu, and H. Zhu. "The Influence of Friendship Network on Graduate Student's Research Potential.", in International conference on social and technology education (ICSSTE 2015),2015.

[6] T. Scott. "Ties that bind: A social network approach to understanding student integration and persistence." Journal of Higher Education,2000, pp. 591-615.

[7] D. Rulke, and J. Galaskiewicz. "Distribution of knowledge, group network structure, and group performance." In Management Science 46.5,2000, pp. 612-625.

[8] O. Demirbas, and H. Demirkan. "Learning styles of design students and the relationship of academic performance and gender in design education." Learning and Instruction 17.3,2007, pp. 345-359.

[9] M. Mills, and C. Fullagar. "Motivation and flow: Toward an understanding of the dynamics of the relation in architecture students." The Journal of psychology 142.5,2008, pp.533-556.

[10] W. Nooy, A. Mrvar, and V. Batagelj, (2011) Exploratory Social Network Analysis with Pajek. Cambridge University Press, New York

[11] A. Astin. Preventing Students from Dropping Out, San Francisco: Jossey-Bass Publishers, 1975.

[12] C. Cook, F. Heath, and RL. Thompson . “A meta-analysis of response rates in web- or internet-based surveys.” Educ and Psychol Meas,2000, pp. 60(6):821–36.

[13] J. Draugalis, S. Coons, C. Plaza, “Best Practices for Survey Research Reports: A Synopsis for Authors and Reviewers”. Am J Pharm Educ, 2008, 2008;(1):72. Article 11.

[14] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2254236/pdf/ajpe11.pdf

[15] L. Khalil, J. Draiby and N. Abi Karam, “Author rights awareness to promote an inter-university open-access repository for theses and memoires” , in SEAAIR 2014 Conference Cross-Cultural Education for AEC 2015: Realizing Possibilities, Defining Foundations.http://www.seaairweb.info/Conference/index.aspx

[16] S. Sivo, C. Saunders, Q. Chang, and J. Jiang "How Low Should You Go? Low Response Rates and the Validity of Inference in IS Questionnaire Research," Journal of the Association for Information Systems: Vol. 7: Iss. 6, Article 17. 2006. http://www.bus.ucf.edu/faculty/csaunders/file.axd?file=2011%2F2%2F How+Low+Should+You+Go..Low+Response+Rates+and+the+Validity +of+Inference+in+IS+Questionnaire+Research.pdf (accessed August 15, 2014).

[17] J. Fincham. “Response Rates and Responsiveness for Surveys, Standards, and the Journal”, Am J Pharm Educ. 2008 April 15; 72(2): 43. PMCID: PMC2384218, 2008, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2384218/pdf/ajpe43.pdf (accessed August 15, 2014).

[18] D. Nulty . “The adequacy of response rates to online and paper surveys: what can be done?”, Assessment & Evaluation in Higher Education Vol. 33, No. 3, June 2008, 301–314, 2008. http://www.uaf.edu/files/uafgov/fsadmin-nulty5-19-10.pdf (accessed August 15, 2014).

[19] J. Scott. Social network analysis: A handbook. London: Saga Publications. ISBN 0-8039-8480-4, 1991.