The predictive power of SNA metrics in education

Diego García-Saiz
Department of Mathematics, Statistics and Computation
University of Cantabria
garciasad@unican.es

Camilo Palazuelos
Department of Mathematics, Statistics and Computation
University of Cantabria
camilo.palazuelos@unican.es

Marta Zorrilla
Department of Mathematics, Statistics and Computation
University of Cantabria
zorrillm@unican.es

ABSTRACT

Nowadays, centrality measures from social network analysis are being used for discovering underlying relationships among different actors or elements whose connections can be modeled as a graph. Their application in the educational domain has not been studied thoroughly, but the information that these metrics provide, as well as their predictive power, justify their use to model the students' social profile, as this paper shows.

Keywords

SNA, centrality measures, student performance

1. INTRODUCTION

Technology has provoked a great change in the way of teaching and learning. Its use in classrooms has supposed a change in the learning paradigm, from a teacher-centered model to a student-centered one, boosting students to construct their own knowledge whilst only being guided by the teacher. Also, social media is currently contributing, in an informal way, to train new skills to look for information, discuss, and elaborate new knowledge collaboratively. Therefore it is conceivable that, in a not very far future, social networks might be re-thought as a support for learning [2].

This has led us to study metrics that help us measure student social behavior in order to assess their predictive power to build student performance classifiers. In particular, we review centrality measures provided by social network analysis (SNA) and evaluate those that are suitable to build the student social profile and those that contribute to achieve better student performance models.

2. METHODOLOGY

First of all, we extracted all measurable activity variables for each student from the Moodle database. Next, we built a social network with the answers given in the forums. That means, we designed a graph in which students and instructors were defined as nodes and the answers given to questions or answers written by students were gathered as directed and weighted edges, being the weight the number of times that a student answers questions initiated by the student to whom is connected. Secondly, we utilized the ORA social network application [1] to calculate 10 node level SNA metrics [3]. Next, we interpreted the meaning of these SNA metrics in the educational context and associated each one to a social behavior. Then, we added them to previous datasets and carried out a feature selection process using Weka. This process allowed us to find out the subset of input variables of each dataset that had relevant predictive information for the pursued objective.

3. DATASETS

For this study, we chose two virtual courses taught at the University of Cantabria, entitled "Calculus" and "Gender Equality in Institutions" (GEI) with 115 and 48 students enrolled, respectively. These were selected because the activity in the forum was remarkable. In the former, the forum was used by students to ask for help and suggestions to their peers and instructor, and in the GEI course, it was used as a discussion tool.

4. RESULTS AND DISCUSSION

Table 1 shows the top 5 scoring nodes side-by-side for those centrality measures of the "Calculus" course. From its analysis, we can discover the different educational roles and the social behavior of each individual.

Node 3254 leads almost all the ranking categories. This is usually associated with the instructor of the course since it is often the one that starts the threads and makes others intervene. Node 5036, though not as prominent as 3254, can be thought as a co-instructor or a teaching assistant: it asks and responds noticeably (high indegree and outdegree), happens to be well-connected (high betweenness and low closeness) and seems to be both an authority and a hub. Node 4046 also presents an interesting behavior: somewhat in between the teacher mode of operation and the students' one. It is an active node (high degree) but its reputation is not as high as the previous nodes' (low eigenvector, hub, and authority values). It could be a not-so-active co-instructor or a highly collaborative student.

Regarding the students' behavior, node 12722 seems to have a remarkable attitude towards the course as a student. It answers a high amount of questions (high outdegree) and

Table 1: Top 5 scoring nodes for the centrality measures

Degree	Indegree	Outdegree	Eigenvector	Closeness	Information	Betweenness	Hub	Authority
3254	3254	3254	3254	3254	3254	3254	12722	3254
5036	4046	5036	12722	5748	5036	4046	6837	6826
4046	5036	12722	6837	5036	12722	5036	5036	5036
12722	6837	4046	5036	4046	4046	6685	5077	6821
6837	5630	6837	5077	6821	6837	5630	6821	6837

its responses are of great value (high hub). This permits it to reach a majority of the people through its answers (high eigenvector). It does not seem to ask too much. It would rather respond than ask. Finally, node 6837 presents the behavior of a good student, but participating in the forum in a radically different way than the previous node. This node tends to ask more than answer (high indegree) and the questions it makes are appreciated by the rest of students (it is in the top 5 authority ranking).

We used ClassifierSubSetEval and SubSetEval techniques for the feature selection process. Both were run using 10-cross fold validation. Therefore, the relevance of each attribute for the classification task is measured from 0 to 10 (no relevant to highly relevant). Table 2 shows the attributes selected by both techniques in the GEI course. The relevance of some SNA metrics is exceptional, such as degree and hub (students who answer those who receive many answers, learn, and have a higher probability to pass), as well as activity metrics, such as the number of initiated discussions in the forum and the number of visits to the course resources, though to a lesser extent.

Table 2: SubSetEval in the GEI course

	Attribute	Relevance (0-10)
SubSetEval	Degree	7
	Hub	3
ClassSubSet NaïveBayes	Hub	10
	Authority	3
	InformationCentrality	6
	ClickMembership	4
	DegreeClustering	3
	N_initiated_discussions	3
	N_views_resources	9
ClassSubSet J48	Hub	5
	ClickMembership	3
	Betweeness	3
	N_Initiated_discussions	4
	N_read_discussions	3

Table 3 displays the result obtained in "Calculus" course. It can be observed that the SNA measures have also an outstanding importance in order to build the prediction models.

Finally, we built classification models from these datasets with and without including SNA measures as attributes. Table 4 shows the accuracy (Acc.), sensitivity (Sens.) and specificity (Spec.) of the models built, as well as the improvement obtained using SNA attributes.As can be observed, the models obtained using SNA measures are more accurate in 75% of cases and present significant improvements. Due to the fact that students' involvement in the

Table 3: SubSetEval in the "Calculus" course

	Attribute	Relevance (0-10)
CfsSubSet	Degree	4
	N_attempts_quizzes	6
ClassSubSet NaiveBayes	Eigenvector	3
	Authority	4
	ClusteringDegree	3
	N_read_discussions	8
	N_view_resources	8
	N_attempt_quizzes	9
ClassSubSet J48	Betweeness	3
	N_view_resources	7
	N_read_discussions	7
	N_attempt_quizzes	8

"Calculus" course is 40%, whereas in the GEI is 100%, it is to be expected that SNA measures show a higher predictive power in the latter, as it can be confirmed by our results.

Table 4: Accuracy, Sensitivity and Specificity obtained with J48 and NaïveBayes in both courses

			SNA	No SNA	Improv.
GEI	J48	Acc.	76.74%	62.79%	13.95%
		Sens.	76.9%	65.4%	11.5%
		Spec.	76.5%	58.8%	17.7%
	NB	Acc.	51.16%	48.84%	2.32%
		Sens.	57.7%	53.8%	3.9%
		Spec.	41.2%	41.2%	0.0%
"Calculus"	J48	Acc.	76.52%	70.43%	6.09%
		Sens.	86.0%	80.2%	5.8%
		Spec.	48.3%	41.4%	6.9%
	NB	Acc.	64.34%	65.21%	-0.87%
		Sens.	82.6%	83.7%	-1.1%
		Spec.	10.3%	10.3%	0.0%

These results allow us to conclude that SNA measures, extracted from the interactions of the students in forums from e-learning courses, are very informative to predict the students' performance and help to improve the classification models. Of course, the more the forum is used in a course, the more useful SNA measures are for this purpose.

5. REFERENCES

[1] K. Carley, J. Pfeffer, J. Reminga, J. Storrick, and D. Columbus. Ora user's guide 2013. Technical Report CMU-ISR-13-108, Carnegie Mellon University, 2013.

[2] C. Greenhow. Online social networks and learning. *On the Horizon*, 19(1):4–12, 2011.

[3] M. Newman. *Networks: an introduction.* Oxford University Press, 2010.