Social Networks Analysis for Quantifying Students' Performance in Teamwork

Pedro Crespo
Instituto Superior Técnico
Av. Rovisco Pais 1
1049-001 Lisboa, Portugal
pedrocrespo@ist.utl.pt

Cláudia Antunes
Instituto Superior Técnico
Av. Rovisco Pais 1
1049-001 Lisboa, Portugal
claudia.antunes@ist.utl.pt

ABSTRACT

Teamwork has become an important part of the educational process, aiming for preparing students for their future performance, but also for making the learning process easier. Predicting students' performance in advance is one of the keys to prevent failure, but teamwork results are difficult to estimate and impair the global prediction results **[1]**.

Working in group is an inherent social activity, and measuring students' performance in it may be enhanced if understood in that context. In this paper, we propose to quantify the performance of students in teamwork by making use of the most effective techniques for social networks analysis. Teamwork is then represented as a network, where students interact with each other, achieving some results (their grades). We explore a network architecture and provide a strategy for quantifying the global contribution of each student through adaptations of the PageRank algorithm **[2]**.

Keywords

Social networks analysis, Teamwork, Failure prediction.

1. Introduction

The concept of Social Networks, typically seen as interactions between individuals, has become extremely popular in the last decade due to its huge application in the online domain in websites such as Facebook, Twitter or LinkedIn. The structure of Social Networks often encloses tremendous amounts of information in the linkage between individuals and content shared among them [3]. Ranking algorithms have already been applied to other domains [4], but seldom applied in the context of education.

In this paper, we show that it is possible to address the thematic of team working in the educational context, through the use of ranking techniques over social networks. Our main goal is to show how these techniques can be applied and what are the main drawbacks faced on trying to measure the value of each student as a team member.

2. Teamwork as a Social Network

Teamwork can be defined as social group where students are involved in social interactions with each other, share interests resulting from the terms of classes, and have the common goal of completing a project or assignment where labor can ideally be equally divided among all participants.

When a student agrees to participate in a social group in the context of some subject, three types of grades are achieved in that subject: the final grade, the group grade, and the individual grade.

According to [3], a *social network* is defined as a network of interactions or relationships, where the nodes consist of actors and the edges consist of the relationships or interactions between these actors. A social network is usually represented as a *graph*: a pair $G = (V, E)$ where V is a set of nodes and E a set of edges [5].

Social interactions within students' groups are mutual between all members, so the graph will be *undirected*. In terms of connections we chose to represent *unweighted* edges. Apart from the definition, it is possible to include a *content-based component* containing the students' grades, as seen in Table 1.

Table 1 – Content-based component: students' grades

Student	Final Grade Avg	Group Grade Avg	Individual Grade Avg
1	14	16	13
2	15	16	15
3	17	17	17
4	17	18	16
5	16	18	15
6	14	17	12

Figure 1 shows the structure of a social network composed by six students interacting among them. Students 1, 2 and 3 form a social group; students 3, 4 and 5 form a second one, and students 4, 5 and 6 form a third social group. The edges between students 4 and 5 should be seen as a single edge.

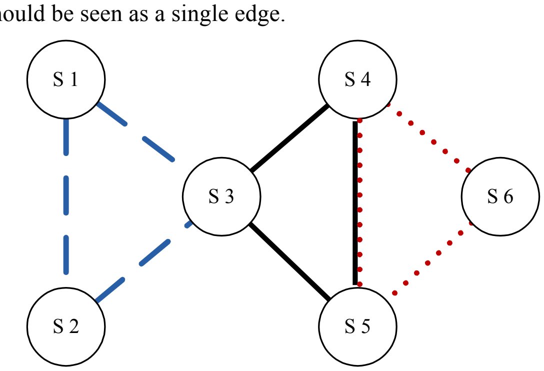

Figure 1 – Network Structure-based component

3. Link Analysis

In order to evaluate our argument, we conducted some experiments. The data sample contains approximately 1700 evaluations of over 550 unique students. This represents the data of 8 subjects during approximately 2 years combined in 17 evaluation terms. Each student record contains individual grades, group grades and final grades for each enrollment at a given

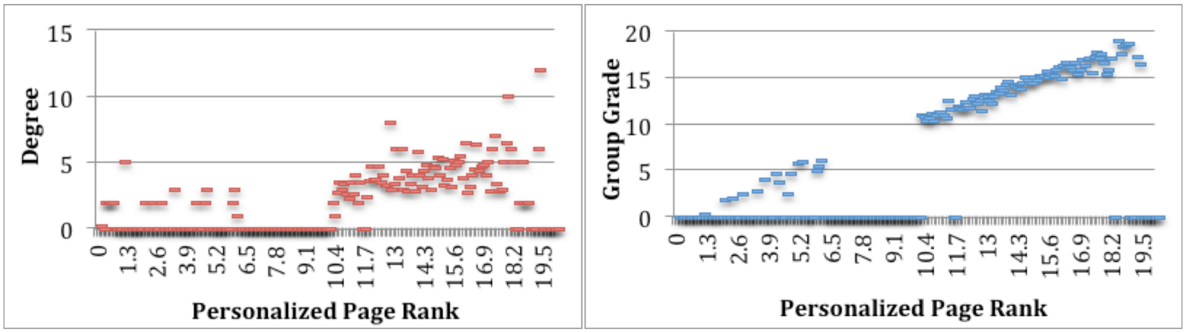

Figure 2 - Relation between Personalized PageRank and (a) Node Degree and (b) Average Group Grades

subject. The data was then modeled following the network structure described above.

The best-known algorithms for structure-based analysis fall under the category of ranking algorithms determining the rank of each node in the graph in terms of their similarity.

The PageRank algorithm [2] is known worldwide for ranking web-pages in order of their importance. Given an unweighted graph $G = (V, E)$ as defined above, and where N is the total number of nodes, if an edge exists from a node j to a node i then the node j is diffusing information to node i in terms of importance according to Expression 1:

$$PR(i) = d \sum_{j:j\to i} \frac{PR(j)}{D_{out}(j)} + (1-d)\, v(i)$$

Expression 1 – PageRank formula

where d is called the *damping factor* and can take values between 0 and 1. $D_{out}(j)$ represents the out-degree of node j. The last component $v(i)$ is part of a personalization vector that can be used to influence the ranking of a given node to better or worse. The Personalized PageRank [6] adaptation is similar to PageRank, differing only in the calculation of $v(i)$. The typical value for $v(i)$ in the Traditional PageRank is $\frac{1}{N}$ but it can vary in the Personalized PageRank.

In the context of this paper we do not see PageRank as a probabilistic distribution, but as the relative value of importance of each node in the graph.

3.1 Experimental Results

The PageRank algorithm allows a purely structural analysis based on the representation of a social network. The results from applying PageRank in the described network structure showed a strong proportionality with the degree of each node, and no relation with the group grade.

The Personalized PageRank algorithm allows using content-based data in order to influence the structural analysis by using Personalized PageRank vectors with different content regarding each one of the students. We used the average group grade of each student, present in Table 1 as the value for the PageRank vector but we could have used any other attribute.

Figure 2 (a) shows that the relationship of proportionality between the degree of a node is faded. We can still notice that there is a slight slope towards the highest-ranking values, together with a high dispersion rate. The Personalized PageRank and the average of group grades clearly show a relationship of proportionality in Figure 2 (b).

4. Conclusions

The pure structural analysis and the application of the traditional PageRank algorithm fall short from the desired objective of describing the teamwork value of each student. The fact that the ranking simply depends on the degree distribution of the nodes in the graph fails to capture the real value of each student.

Influencing the rankings by adding content-based data through the use of Personalized PageRank vectors seems to have improved the capture of the real teamwork value of each student, but we still need to determine how much of an added value this analysis brings in comparison to simply calculating the group grade average for each student.

Acknowledgments

This work is supported by Fundação para a Ciência e Tecnologia under research project *educare* (PTDC/EIA-EIA/110058/2009).

References

[1] Antunes, C. 2010. Anticipating student's failure as soon as possible. In *Handbook for Educational Data Mining*, Romero, C. et al., eds. Chapman & Hall/CRC Press, New York, NY.

[2] Brin, S. and Page, L. 1998. The anatomy of a large-scale hypertextual Web search engine. In *Proc. of Int'l Conf. World Wide Web* (Brisbane, Australia). Elsevier Science Publishers, 107-117.

[3] Aggarwal, C. 2011. Introduction to Social Network Data Analytics. In *Social Network Data Analytics*, Aggarwal, C., ed. Springer.

[4] Agirre, E. a.S.A. 2009. Personalizing PageRank for World Sense Disambiguation. In *Proc. of EACL '09 Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* .

[5] Gross, J. L. and Yellen, J. 2004. *Handbook of graph theory*. CRC Press.

[6] Liben-Nowel, D. and Kleinberg, J. 2003. The link prediction problem for social networks. In *Proc. of CIKM '03* .