Project CASSI: A Social-Graph Based Tool for Classroom Behavior Analysis and Optimization

Robert Olson, Zachary Daily, John Malayny, Robert Szkutak
Department of Computer & Information Sciences
State University of New York at Fredonia
Fredonia, NY 14063
olsonr@fredonia.edu, dail9583@fredonia.edu, mala3598@fredonia.edu, szku7989@fredonia.edu

ABSTRACT
Although educational data mining is a well-established field, it has not yet sought to provide serious, actionable intelligence that can be used by teachers to address bullying in a reasonable amount of time. This paper seeks to propose a system that will streamline the processing and storage of bullying data in social graph form so that it will be available to be mined by expert systems that can help educators in the classroom. In addition, one such expert system will be proposed demonstrating how this data may be used to automate a common classroom management task that may improve students' classroom experiences.

Keywords
Behavior modeling, Implicit social graphs, Classroom behavior optimization, Seating chart generation

1. INTRODUCTION
Although bullying has long been a significant problem, increased awareness has brought the matter to the attention of legislatures. States across the country are passing laws aimed at preventing bullying. Some pieces of legislation, like New York's *Dignity for All Students Act*, include provisions for information sharing and increased data retention [3], creating an environment ripe for innovation in the fight against bullying.

In this paper, we will propose CASSI (Classroom Assisting Social Systems Intelligence), an open-source system aimed at being inexpensive and easily integrated into existing educational practices that will allow for the collection, modeling, and analysis of student behavioral data. Specifically, the collected behavioral data will be used to construct a social graph that represents how dysfunctional the directed relationship between each pair of students is. This social graph can then be used to inform the behavior of a variety of expert systems.

It is hoped that this system, in due time, will be implemented in educational institutions to adhere to both the letter and the spirit of new pieces of anti-bullying legislation. A single central repository for an educational institutions behavioral data would allow for the data to be more easily shared amongst educators and formatted into reports for administrators. This repository of data would also allow for the implementation of expert systems which educators can use in day-to-day classroom management tasks that influence bullying [1]. One expert system will be described that makes use of the data stored in the social graph to improve classroom management by allowing teachers to automatically generate seating charts likely to reduce negative classroom behavior. Other possible expert systems will also briefly be discussed.

2. BACKGROUND
Romero and Ventura[7] provide a detailed overview of early educational data mining projects. The authors describe a wealth of educational data mining projects aimed at improving pedagogy. However, there is a lack of data mining projects intended to improve students' educational experiences through means other than pedagogy. Romero and Ventura also note that early data mining tools do seem to be designed with data mining experts, not educators in mind. Their implication seems to be that, if provided with the right tools, educators would be able to make better use of the information gathered from data mining.

Dawson[2] recounts a study utilizing social network analysis to draw some conclusions about how a student's position in a social network influences the student's perceptions regarding the sense of community they experience. Although this is a narrow application of social network analysis in the classroom, it provides an excellent justification for using social network analysis as a means of evaluating classroom behavior. While this study doesn't describe a system easily used by educators, one can easily imagine how such a system could be extended to provide a more detailed level of analysis by integrating more observations containing the sort of behavioral data that Hung and Lockard [4] used to create their Behavioral Matrix software.

There has been a recent push for software to address cyber-bullying through the analysis of social networking sites. Nahar, Unankard, Li, and Pang [5] describes a method for using sentiment analysis to develop a graph-based method for identifying cyber bullies and their victims while Sanchez and Kumar [1] describes a method for integrating this style of sentiment analysis with Twitter. Although cyber-bullying is a significant problem, neither of these papers address problems associated with completeness. It is easy for students to restrict educators' access to their social media accounts and such restrictions may skew the decisions of an expert system which integrates social media data with observed classroom data. However, both of these methods appear to accurately recognize the asymmetric nature of bullying discussed by Allen [1].

It also is important to note the distinction between social networks and social networking. As noted in Purtell et al. [6], it is possible to extract implicit or inferred social topographies [8] from other sources other than social media.

3. DATA MODEL AND VISUALIZATION
One of the strengths of CASSI is that much of the data that the system collects is likely to already be collected as a part of routine classroom management practices. At the moment, the only data

the system requires educators to record is: the victim, the bully, and a ranking of the bullying incident on a score from 1 to 10. It is expected the system will also record additional book-keeping data such as a time-stamp, the educator filing the incident report, and a detailed description of the even that may be utilized in the future. This data, collected via a web-form when an incident occurs and stored in a local database, is similar to the directed data described by Purtell et al. [6] as suitable for use in the construction implicit social graphs of the type described in Roth et al. [8].

There are some flaws in this data collection mechanism. The incident score necessitates the development of an objective standard against which incident seriousness may be compared to ensure consistency. Another flaw is that bullying often takes place in spaces not observable by educators. This may be addressed in the future through mechanisms such as student-driven incident reporting tools. However, this flaw may also be addressed by educator training aimed at expanding definitions of bulling to include include what Allen [1] refers to as "relational bullying."

Once the data is collected via web-form, it is processed to form the social graph. This amounts to populating a matrix S, indexed by student, of ordered pairs (T, V) where T for $S_{i,j}$ represents the sum of the seriousness ratings from incidents where student i is engaging in bullying behavior targeting student j. V, meanwhile, represents the number of incidents added to produce T.

This two dimensional information can be easily analyzed. For example, the student relationships furthest from the origin should be red flagged as those most in need of immediate intervention. Relationships where T / V is high while V is low may indicate an emerging bully. Finally, the case where the Euclidean distance between points Si,j and Sj,i is low but the distance from both of these points to the origin is high may represent a rivalry in need of serious intervention.

4. CLASSROOM OPTIMIZATION

One expert system that has been developed to make use of the behavioral social graph aims at automating the task of finding behaviorally optimal seating charts for rectangular seating arrangements of an arbitrary size. This task is accomplished by comparing the social graph to a particular arrangement of students. If two students are seated adjacent to one another in the classroom, their relationship information is extracted from the social graph and added to the ranking of the classroom. This reduces the task of finding an optimal classroom down to a simple minimization problem – the lower the ranking, the better the classroom.

Iterating through all possible arrangements of students guarantees that the best seating arrangements is found. However, this is extremely computationally expensive at $O(N!)$ for classrooms with asymmetric physical properties that may influence behavior, such as windows that might provide a distraction.

Fortunately, it is also relatively easy to develop a heuristic that exploits the physical properties of the classroom to find good, but not necessarily optimal, arrangements. Specifically, this heuristic exploits the number of adjacencies that each seat has. Corner seats only have three direct adjacencies in a rectangular classroom, for example, making them ideal locations for the students most likely to disrupt others. Using these physical classroom properties and sorting the student in order of the probability of causing a disruption reduces the time to $O(N)$.

There was some concern that the formation of networks of friendships along racial, gender, academic performance, or economic class boundaries may cause the seating charts generated this way to, unintentionally, segregate classrooms. This was addressed by scoring the classroom using a weighted average of the social-graph adjacencies within the arrangement of students and the number of homogenous demographic adjacencies within the arrangement of students.

5. CONCLUSION AND FUTURE WORK

Project CASSI is a tool that should allow educators to share behavioral information more easily and serve as the foundation on which useful classroom-management expert systems may be built. However, Project CASSI is in its infancy. Although CASSI has been tested extensively on simulated data, it is most immediately in need of testing with authentic behavioral data.

Once the system has been tested on genuine behavioral data, there are a number of additional expert system modules that may extend its usefulness. In particular, the researchers of Project CASSI expect that the system may be extended to support behaviorally informed scheduling and time-series forecasting. The former task – grouping students into non-disruptive classes – may be accomplished a similar selection algorithm to the one proposed for seating chart generation but operating on combinations of students rather than permutations of students. The later task, time-series forecasting using the time-stamps of the incident report, may be conducted with the ultimate goal of predicting bullying trends and predicting bullying events before they actually occur.

6. REFERNECES

[1] Allen, K. P. (2010). Classroom Management, Bullying, and Teacher Practices. *Professional Educator, 34*(1), 1-15.

[2] Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Educational Technology & Society, 11*(3), 224-238.

[3] Dignity for All Students Act. 2010. N.Y. Educ. Law §§ 10-18

[4] Hung, W.-C., & Lockard, J. (2007). Using an Advance Organizer Guided Behavior Matrix to Support Teachers' Problem Solving in Behavior Classrom Management. *Journal of Special Education Technology, 22*(1), 21-36.

[5] Nahar, V., Unankard, S., Li, X., & Pang, C. (2012). Sentiment Analysis for Effective Detection. In *Web Technologies and Applications* (Vol. 7235, pp. 767-774). Springer Berlin Heidelberg.

[6] Purtell, T. J., McLean, D., Teh, S. K., Hangal, S., Lam, M. S., & Heer, J. (2011). An Algorithm and Analysis of Social Topologies from Email and Photo Tags. *Workshop on Social Network Mining & Analysis, ACM KDD.* San Diego, CA.

[7] Romero, C., & Ventura, S. (2007). Educational Data Mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*, 135-146.

[8] Roth, M., Ben-David, A., Deutscher, D., Flysher, G., Horn, I., Leichtberg, A., et al. (2010). Suggesting Friends Using the Implicit Social Graph. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* , (pp. 233-242).

[9] Sanchez, H., & Kumar, S. (2011). *Twitter Bullying Detection.* UCSC ISM245 Data Mining course report.

Content analysis is a method of researching communication patterns. It can focus on words, subjects, and concepts in texts or images.Â Content analysis can be used to quantify the occurrence of certain words, phrases, subjects or concepts in a set of historical or contemporary texts. Quantitative content analysis example.Â Based on your research question, you have to categorize based on gender and the concept of trustworthiness. To get more detailed data, you also code for other categories such as the age, political party, and marital status of each politician mentioned. 3. Develop a set of rules for coding. Graph analysis, statistics, data visualization, optimization, image recognition. CSV, DOT, GraphML, JSON, Pajek, XLS and multiple other non-network formats. CSV, DOT, GraphML, JSON, Pajek, XLS and multiple other non-network formats.Â NetMiner is a software tool for exploratory analysis and visualization of large network data. NetMiner 4 embed internal Python-based script engine which equipped with the automatic Script Generator for unskilled users. Then the users can operate NetMiner 4 with existing GUI or programmable script language.Â Bergm provides tools for Bayesian analysis for exponential random graph models; hergm implements hierarchical exponential random graph models Data analysis is a complex and intricate process. It involves collecting and structuring data, forming and testing hypotheses, identifying patterns, and drawing conclusions. Data analysts are essential in business, administration, and science.Â What youâ€™ll get. Over the course of this six-month, 20-hour-per-week program, you will master the skills required to become a data analyst and build a portfolio of projects on topics such as these: Streaming on-demand media: user preferences. Find out how the popularity of musical genres depends on the time of day and day of the week. Title:A Graph-Based Platform for Customer Behavior Analysis using Applications' Clickstream Data. Authors:Mojgan Mohajer. Download PDF. Abstract: Clickstream analysis is getting more attention since the increase of usage in e-commerce and applications. Beside customers' purchase behavior analysis, there is also attempt to analyze the customer behavior in relation to the quality of web or application design. In general, clickstream data can be considered as a sequence of log events collected at different levels of web/app usage. The analysis of clickstream data can be performed direct