A Computer-in-the-Loop Approach for Detecting Bullies in the Classroom

Juan F. Mancilla-Caceres, Wen Pu, Eyal Amir, and Dorothy Espelage

University of Illinois at Urbana-Champaign, Urbana IL 61801, USA,
{mancill1,wenpu1,eyal,espelage}@illinois.edu

Abstract. Bullying is a social phenomenon that is highly prevalent within the school population. To study this phenomenon, social scientists traditionally use questionnaires that are costly to administer and that cannot provide detailed information about children's interactions without causing a large amount of fatigue to the participants. An on-line computer game has been developed to aid social scientists in observing, in a non-intrusive way, children's behaviors and roles within their peer group. Participants solve a collaborative and an adversarial task, and are allowed to communicate only through a chat system. Observable data from the game, such as the amount of messages sent and received and points transactions, correlates well with questionnaire data while providing more detailed information about participants' interactions. The online game is a new tool that alleviates the cost of obtaining data and considerably reduces the fatigue of the participants while providing sound results.

Keywords: Group interaction and collaboration, Influence process and recognition, Methodological innovation

1 Introduction

Computer involvement in adolescent networks has been growing in recent years, giving rise to new group dynamics and new opportunities to study group interactions and individual preferences. Reseach in computer science already takes advantage of the availability of this information, but is biased towards visible computer-mediated social networks when friendship nominations are assumed equal and meaningful. Social science research, on the other hand, applies well founded psychological techniques, but finds difficulty using the data available to computer scientists because of limited computational tools and the partial nature of the data.

In this paper, we present a new method for gathering indirect information about friendships in classrooms and for observing real interactions between adolescents using an on-line social game. The game is used as a tool for observing the behavior of the participants in a non-intrusive way and for gathering information about friendships and roles that each participant plays within their social groups. Current research suggests that peer socialization is one of the key

factors that determines risky behavior in adolescent social networks [6]. However, in order to understand the roles and behaviors of adolescents, traditional methods involve surveys and questionnaires that can yield limited information due to their cost and potential participant fatigue. These surveys are usually time-consuming and in some cases, they are simply not feasible to use with large school populations. In addition, real social-interaction data (e.g., adolescent manipulations) are difficult to collect, evaluate, and validate, and it is difficult to use available data (e.g., Facebook friendships) to draw sound conclusions.

Instead, we propose that the roles of the participants in the classroom (*bully*, *victim*, or *bystander*) are observable through social interactions within a game designed to simulate common real-world situations, such as collaboration and competition within groups. The data obtained using the game consists on teammate nominations of participants, text messages amongst the players (which can be either public or private), and point transactions. We show that game outcomes are correlated with survey data gathered previously from the same participants, and that they can be eventually used to predict the values of the surveys, and to provide updates to the roles found using survey data (as will be shown in a case analysis).

2 Related Work

Game-based methods for data collection have been previously used as an alternative method to crowdsourcing. One of the first successful examples of what has been called *Games with a Purpose* (GWAPs) was the ESP game [8] used for labeling images by volunteer contributors. In the context of social networks, GWAPs have been created to take advantage of people's engagement on such networks. For example, Collabio [1] is a social tagging game within the online social network Facebook, which encourages friends to tag one another. Collabio's encourages members of the social network to generate information about each other by producing tags that describe another individual. Also, the Turing Game [5] is a Facebook multiplayer game that encourages players to verify common-sense knowledge by carefully designing the rules and stages of the game so that the players have an incentive to generate the appropriate data.

The game described in this paper differs from those previously described in the sense that the desired outcome of the game is not the specific action the players take on the game (i.e., the classification of the image or text by the player), but the interactions amongst the players themselves. This is a novel approach to gather data and information because, although the design of the game is important (in order to encourage participation, engagement, and messages), the actual game task is not crucial (in this case, answering trivia questions). Most of the other games for data collection can be played by a single person, but our game is played in teams, which is necessary to gather information about social group dynamics. Our data directly reflects the actual interactions of the players, thus reducing the noise in the data. This kind of game is especially useful for non-intrusive data-collection in the social sciences because the games provide a

natural data-intensive interaction with participants. The game is applied here to learn about social communications, manipulation methods, and reactions to natural occurrences of friendliness or aggression.

With respect to using automatic methods to detect bullying, previous research has focused on cyber-bullying and not on physical bullying as is our case. Some examples are [4] and [7], which focus exclusively on observable behavior (e.g., insulting or racist messages), whereas our approach is aimed at finding the latent behavior associated to children's roles within the classroom, which may or may not be explicit in the game.

3 Game Design

The proposed data collection platform is a multiplayer game played in teams of 3 or 4 members. The game is designed to answer the following questions: 1. What are the friendship relationships amongst the participants, and what kind of interaction do they have (cordial, aggressive, polite, etc.)? 2. How are the loyalties and trust placed amongst the participants, and how do the rules of the game encourage leadership, competitiveness, etc.?

The game is played via a computer network and follows the steps described below:

1. Each user has the opportunity to nominate other participants whom they would or would not like to have on their team. In this stage of the game, we obtain information about task-directed peer nomination. Each team is currently created using a priori information gathered through surveys, ensuring that on each team there is at least one bully and one victim on each team.
2. The second stage of the game consists on collaborating to answer a set of trivia questions, ensuring that all members of the team submit the same answer in order to obtain a reward (in this case, points in the form of coins).
3. The third stage is competitive or adversarial. During this task, each member of the team must provide a different answer to the question while one team member must choose a clearly wrong answer, effectively losing points while the rest of the team wins points. In contrast to the collaborative task, in which the entire team must work together to maximize their individual reward, in the competitive task players are encouraged to directly oppose other members by convincing (or coercing) them to pick the wrong answer.
4. There are two winners in each game: a winning team (summing up all the individual rewards) and a winning player (the one with the larger amount of coins).

During both tasks, participants are encouraged to use the chat system to coordinate their answers and to trade coins (points) amongst themselves. Figure 1 shows a screen shot of the nomination screen (left) and of the current interface of the game (right).

During the collaborative stage of the game, team members work together to answer a set of 5 trivia questions (about topics such as history, geography, pop
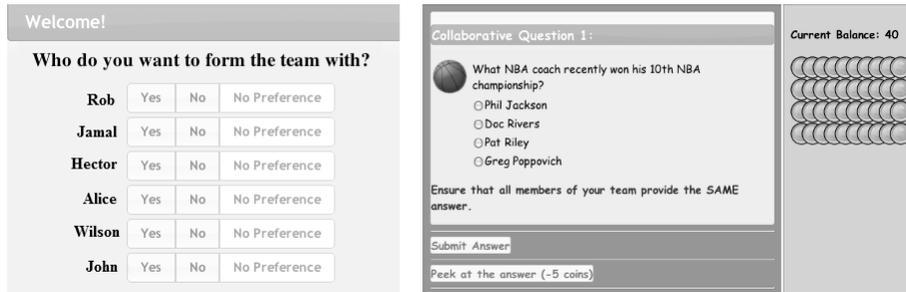
Fig. 1. On the left: Nomination screen shot. Each player has the opportunity of stating with whom they want or don't want to play; or alternatively, of stating that they have no preference. On the right: Screen shot of one question during the collaborative stage.

culture, etc.). Each question has four possible answers, and only one of them is correct. For each question, all the team members must agree on the right answer or otherwise no points are awarded to anyone in the team. These rules ensure that the players in the team must communicate and collaborate to agree on an answer, or everybody loses. The fact that the team shares the payoffs and outcomes guarantees that everyone shares the same interest. For each question, team members have the option to peek at the correct answer in exchange for points from one of the players. In order to maximize the utility of the game, players must balance the peek penalty by sharing points amongst themselves. It is in the best interest of each player to share their knowledge about the question and not to let other players peek at the answers (in order to retain points).

During the competitive stage, each team receives a set of 5 trivia questions with four possible answers. Three of the four answers are correct and one of them is incorrect. In this task, each player on a team must choose a different answer for each question, with the added constraint that at least one member of the team must pick the wrong answer. Only the players that choose a correct answer get points. The wrong answer is marked explicitly (written in bold letters) in order to make it obvious and to encourage players to discuss who will pick such answer. It is in the best interest of each player not to pick the wrong answer, but also to ensure that someone else in their team picks it. This can only be accomplished by negotiating (either aggressively or non-aggressively) through the chat channel.

The game is intended to emulate the circumstances of natural interactions amongst participants. Its features include the presence of limited resources (i.e., points and coins), a collaborative task, and a competitive task. By restricting the channel of communication to text messages, we have a non-intrusive way to monitor and analyze the interactions of participants with their peers (a method previously approved by an IRB).

Each game provides the following output:

– Users' team preferences: friends/rivals nominations and the order in which they are selected.

– Raw messages from users: all chat messages both in public and private channels along with the time at which the message was sent.
– Points transactions: transfer and forfeiture of points (e.g. in exchange for information).

4 Correlating Game Behaviors with Questionnaire Data

4.1 Data Collected through Surveys

Three different surveys were administered to ninety-six students from six different 5th grade classrooms in two Midwestern middle schools. These surveys are aimed at measuring aggression and delinquency, as described in [2] and [3]. The surveys include the following scales:

– The *Bully Scale*, which measures the frequency of teasing, name-calling, social exclusion, and rumor spreading.
– The *Fight Scale*, which measures the frequency of physical fighting.
– The *Victimization Scale*, which assesses verbal and physical peer victimization.
– The *Positive Attitude Toward Bullying and Willingness to Intervene*, which evaluates participants attitudes toward bullying, and the extent to which they are willing to assist a victim.
– The *Need for Control and Dominance*, which assesses self-perceptions of dominance and control within one's peer group.

Using the values of these scales, an expert labeled each participant as either a *bully* or a *non-bully*. We hypothesized that there is a subset of students who are bullies that have a need for control and dominance, will engage in coercive tactics directed toward non-friends, and will solicit support for these tactics from friends within and outside their group. The survey data was used to evaluate the psychometric properties of variables yielded by the computer game with the ultimate goal of being able to develop measures of interaction patterns solely through the computer game.

4.2 Data from the Game

Using a 2(bully/non-bully) x 2(collaborative/competitive) ANOVA we studied the interactions and differences in the behaviors of Bullies and Non-Bullies (classified as such according to the data obtained through the surveys and analyzed by an expert) during both Competitive and Collaborative tasks. Results show that those two kinds of players behave differently during both tasks.

The features used for this analysis were (abbreviated name of the features is shown in parenthesis):

– The amount of private messages sent during the collaborative and the competitive stage (*prsent*).
– The number of private messages received (*prrec*).

- The number of public messages sent and received (*pusent* and *purec*).
- The number of times a player peeked at the answer (*peeked*).
- The number of points sent and received (*credsent* and *credrec*).
- Number of positive nominations sent and received, i.e., stating with how many people they want to play and how many want to play with them (*pnsent* and *pnrec*).
- Number of negative nominations sent and received, i.e., stating with how many people they don't want to play and how many do not want to play with them (*nnsent* and *nnrec*).
- Reciprocated nominations, i.e., number of people that nominated each other positively or negatively (*bpn* and *bnn*).
- Unreciprocated nominations, i.e., number of positive nominations to people that nominated the player negatively (*un*).

Table 1 shows the average of the variables per type of player (i.e., Bully or Non-Bully), and the average per stage. There was a significant main effect of bully/non-bully on the amount of private messages sent (*prsent*), and the amounts of times peeked at the answer (*peeked*). Participants labeled as bullies sent more private messages and peeked at the answer a greater number of times than non-bullies. There was also a significant effect of bully/non-bully on the amount of negative nominations sent (*nnsent*); bullies sent more than non-bullies.

There was a significant main effect of collaborative/competitive on all variables. All players sent and received more messages (both public and private), sent and received more coins and peeked at the answer more during the collaborative stage. There were no significant interactions between bully/non-bully and collaborative/competitive. Taken together, the results suggest that bullies tend to send more private messages, to peek at answers more, and to send more negative nominations than non-bullies. Future analysis of the contents of those private messages will provide more insight into the reasons for these behavioral differences.

The previous analysis shows that the game can effectively be used to detect bullies in the classroom by observing those behaviors in which they differ from other players that have different social roles.

5 Observations of Informative Behavior

Several interesting interaction patterns can be observed during the game. We present one example, which we consider specially interesting because it clearly shows a type of bullying behavior that cannot be detected using traditional research methods. Figure 2 shows a subgraph of the nomination network of one classroom. Solid arrows are positive nominations, doted arrows are negative nominations. According to the survey, the individuals labeled as *211* and *214* are bullies, *203* is a victim, and *208* and *216* are bystanders. We can observe that participant *216* positively nominated *211*, *203* and *208*, but was negatively nominated by all of them. It is important to notice that that *211*, *203*, *208*, and *214* almost form a clique. By observing their chat messages, it is clear that *216* is

Table 1. Results of 2 x2 ANOVAs of Bully/Non-Bully, Collaborative/Competitive and the interaction variables. *$p<0.1$, ** $p<0.05$.

	prsent	pusent	prrec	purec	credsent	credrec	peeked
Bullies	23.12	20.69	19.79	50.86	2.57	3.5	1.55
Non-Bullies	15.77	17.93	16.71	50.82	3.39	3.25	1.06
p-value	0.037**	0.221	0.326	0.994	0.302	0.735	0.023**

	prsent	pusent	prrec	purec	credsent	credrec	peeked
Collaborative	20.18	22.35	20.22	61.62	3.98	4.01	2.01
Competitive	14.61	14.72	14.56	40.04	2.43	2.6	0.32
p-value	0.057*	<0.001**	0.03**	<0.001**	0.019**	0.022**	<0.001**

	pnsent	pnrec	nnsent	nnrec	bpn	bnn	un
Bullies	6.19	6.05	5.05	3.71	3.05	1.33	1.33
Non-Bullies	6.02	5.97	3.31	3.79	3.26	1.08	1.18
p-value	0.791	0.925	0.09*	0.914	0.671	0.566	0.685

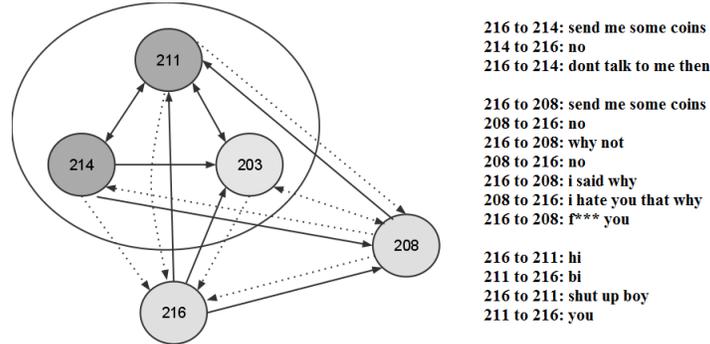experiencing some kind of victimization, but the survey cannot detect this type of pattern.

Fig. 2. Example of victimization observed during game, but not captured by survey. Solid arrows show positive nominations, doted arrows show negative nominations. Chat messages show aggressiveness towards *216*.

6 Conclusions and Future Work

The ultimate goal of this research is to be able to create a tool that can be used broadly to help social scientists and educators better understand and prevent bullying. Currently, our system has shown that bullies, bystanders and victims behave differently while playing the game, and therefore it can be used to identify

bullies in classrooms that have not been surveyed yet. For this, a reliable model to identify individual bullies using only data from the game is being developed.

The data collected in the game includes a large amount of text messages sent amongst the participants. So far, the messages have been classified by a team of 20 raters (each message by at least two raters), but these features have not yet been included in the model. In the future, NLP techniques for discourse analysis will be used to increase the efficiency of bully/victim behavior analysis.

The game itself might still be improved. The rules of the games determine the kind of information that is obtained. New game tasks will be developed in order to increase participants' engagement in the game and the amount of information gathered. By changing the rules, modifying the tasks, and changing the way teams are created, it is possible to observe the key interactions that provoke or invoke bullying. Possible variations include: (1) changing the order of the tasks and determining how the change affects the way participants interact with one another; (2) having a finite number of points for an entire team and only changing the distribution of the points according to the performance in the game; and (3) forcing roles onto the members of the team to determine how quickly participants lead or follow others.

References

1. Bernstein, M., Tan, D., Smith, G., Czerwinski, M.: Collabio: A Game for Annotating People within Social Networks. Proceedings of ACM UIST 2009 (2009)
2. Espelage, D., Holt, M.: Bullying and victimization during early adolescence: Peer influences and psychosocial correlates. Journal of Emotional Abuse, 2, 123–142 (2001)
3. Espelage, D., Mebane, S., Adams, R.: Empathy, caring, and bullying: Toward an understanding of complex associations. In D. Espelage, and S. Swearer (Eds.), Bullying in American Schools: A social-ecological perspective on prevention and intervention. 37–61 (2004)
4. Lieberman, H., Dinakar, K., Jones, B.: Let's Gang Up on Cyberbullying. Computer. vol. 44, no.9, 93-96 (September, 2011)
5. Mancilla-Caceres, J.F., Amir, E.: Evaluating Commonsense Knowledge with a Computer Game. INTERACT'11: Proceedings of the 13th IFIP TC13 International Conference on Human-Computer Interaction. vol 1. 348-355 (2011)
6. Mayberry, M.L., Espelage, D.L., Koenig, B.: Multilevel modeling of direct effects and interactions of peers, parents, school, and community influences on adolescent substance use. Journal of Youth Adolescence. 38, 1038–1049 (2009)
7. Ptaszynski, M., Dybala, P., Matsuba, T., Masui, F., Rzepka, R., Araki, K., Momouchi, Y.: In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis. International Journal of Computational Linguistics Research. vol 1, no.3, 135-154 (September, 2010)
8. von Ahn, L., Dabbish, L.: Labeling Images with a Computer Game. CHI'04: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. 319–326 (2004)