



# Identifying Bullies with a Computer Game

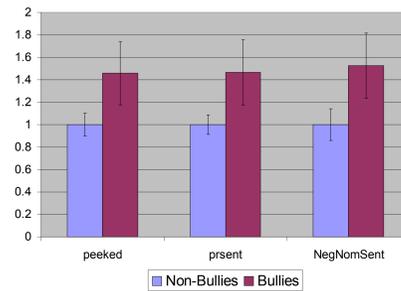
Juan F. Mancilla-Caceres<sup>1</sup>, Wen Pu<sup>1</sup>, Eyal Amir<sup>1</sup>, and Dorothy Espelage<sup>2</sup>

<sup>1</sup>Department of Computer Science <sup>2</sup>Department of Educational Psychology  
University of Illinois at Urbana-Champaign

## 1. Introduction

- Computer game designed to collect information about friendships and real-world interactions within classrooms.
- The game serves as a non-intrusive tool for observing participants' behavior, providing data to infer each participant's most likely role within the classroom.
- The game is evaluated by comparing its results with those of a traditional method for data collection (i.e., self-report surveys).
- A Bayesian model is used to reason about the players interactions.

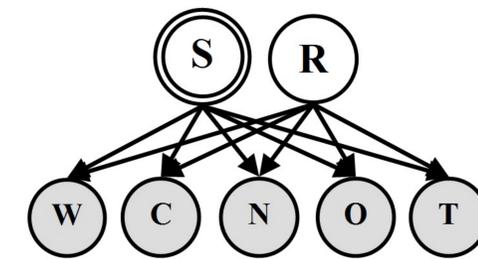
## 4. Differences in Gameplay



Feature	Non-Bullies	Bullies	p-value
Peeked	1.06	1.55	0.023*
Prsent	15.77	23.12	0.037*
NegNomSent	3.31	5.05	0.090*

**Peeked:** Number of times peeked at the answer.  
**Prsent:** Number of chat messages sent through private channel.  
**NegNomSent:** Number of negative nominations sent during Stage 1.

## 7. Two-Layer Bayesian Model

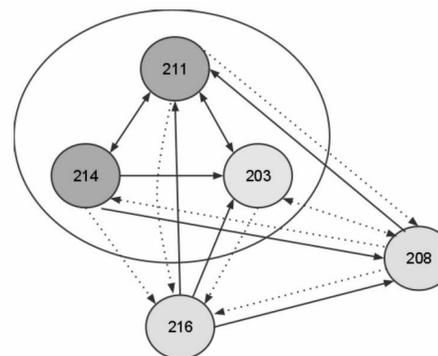


- S is the Sender of messages
- R is the Recipient of messages
- Features:
  - Number of Wins (W)
  - More Coercive than Prosocial (C)
  - More Negative than Positive Affect Messages (N)
  - Positive, Neutral or Negative Nomination (O)
  - Same Team? (T)

## 2. Experimental Design

- Ninety-six students from six different 5th grade classrooms completed a psychological survey, prior to playing the computer game.
- The survey measures the frequency of behaviors such as name-calling, rumor spreading, and social exclusion.
- This information was used to generate labels (Bully or Non-Bully) for each participant by a field expert.
- Approximately two weeks after completing the survey, participants played the game in a computer laboratory for approximately 30 minutes.

## 5. Detailed Data on Interactions



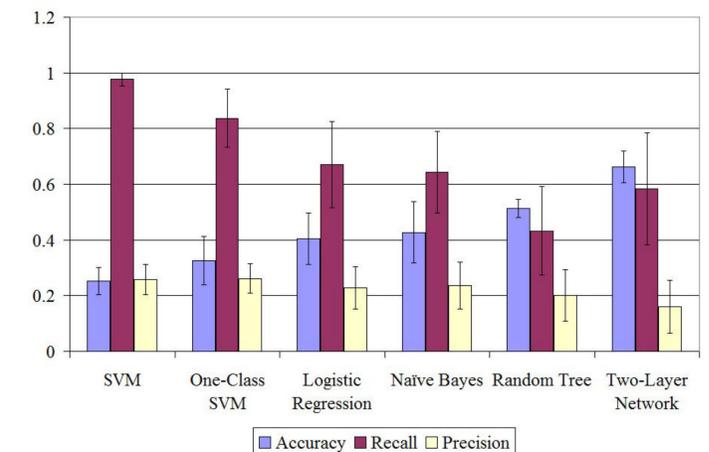
216 to 214: send me some coins  
 214 to 216: no  
 216 to 214: dont talk to me then

216 to 208: send me some coins  
 208 to 216: no  
 216 to 208: why not  
 208 to 216: no  
 216 to 208: i said why  
 208 to 216: i hate you that why  
 216 to 208: f\*\*\* you

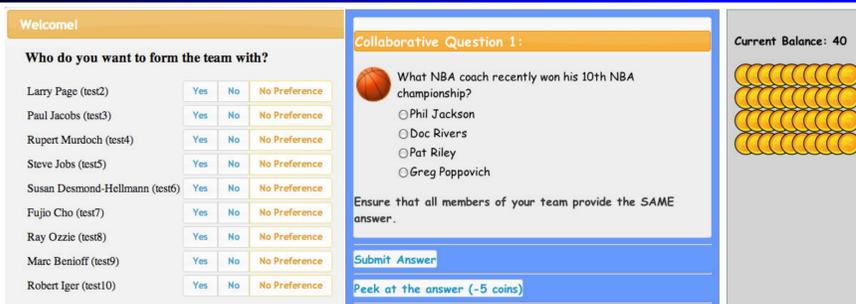
216 to 211: hi  
 211 to 216: bi  
 216 to 211: shut up boy  
 211 to 216: you

→ Positive Nomination      ○ Non-Bully  
 → Negative Nomination      ● Bully

## 8. Prediction Results



## 3. The Game and its Output



- Amount of private (and public) messages sent and received during the collaborative and competitive stages.
- The number of coins sent and received.
- Positive (and Negative) nominations sent and received.
- Reciprocated positive/negative nominations, and unreciprocated nominations.
- Times peeked at the answer.

## 6. Predicting Bullies

- Analysis of the content of the messages.
  - NLP techniques are out of our present scope. We used 20 independent raters to classify the messages according to their content in two different categories: Prosocial/Coercive Message and Positive/Negative Affect Message.
- Bullies do not behave in the same way with every classmate. Therefore, we considered each pair-wise interaction to identify when a given player *S* is behaving as a *bully* towards another player *R*, and aggregate all the interactions of the given player to generate a single label (*Bully* on *Non-bully*).
- We modeled each pair-wise interaction as a Two-Layer Bayesian Network where each of the features is independent of each other (given the type of players interacting).

## 9. Conclusion and Future Work

- Data from the game can be used to identify bullies in the real world and to alleviate the fatigue on both participants and psychologists.
- Two-Layer Network Bayesian Classifier produces the best results (so far) when identifying bullies.
- Future work will focus on:
  - Formal representation of the game (MAIDs).
  - Improving the game and making it adaptable.
- References:
  - Mancilla-Caceres, J.F., Pu, W., Amir, E., and D. Espelage. *Identifying Bullies with a Computer Game*. In the Proceedings of the 26<sup>th</sup> International AAAI Conference on Artificial Intelligence (AAAI'12). 2012
  - Mancilla-Caceres, J.F., Pu, W., Amir, E., and D. Espelage. *A Computer-in-the-Loop Approach for Detecting Bullies in the Classroom*. In the proceedings of the 5<sup>th</sup> International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP'12). 2012