

Background and Goals

Social media are an important context in the lives of youth. Social interactions that occur through social media may facilitate peer relationships, yet they may also place youth at risk for cyberbullying. Cyberbullying is usually studied by asking youth to report on their experiences. This study used social media data to identify instances of bullying (both those that occur online and offline) and bullying role players. We addressed two goals:

- (1) Can bullying and teasing be distinguished in social media posts?
- (2) Who are most likely to tease rather than bully in their social media posts?

Methods

Participants and Measures

We used methods derived primarily from computer science to identify social media posts about bullying episodes (i.e., bullying interactions that involve at least one bullying role player). Beginning in August 2011, we collected posts about bullying episodes from the public Twitter streaming API (see <https://dev.twitter.com/docs/streaming-apis>) on a daily basis. We started by screening for tweets that contained at least one of the following keywords: “bully,” “bullied,” and “bullying.” Yet, keyword filtering is not sufficient. For example, the post “band playing bully wully next door” does not refer to a bullying episode. To correct for this, we used standard machine learning and natural language processing methods to accurately identify the traces. These methods were informed by a “training set” that was created by human coders. The training set was comprised of 1762 tweets that were sampled from the enriched dataset of bullying traces generated on August 6, 2011. In terms of reliability, two researchers coded each post for these 1762 tweets. The Kappa for “teasing or not” was .81 and for “author role” was .52. After reaching an agreement between different coders, we implemented this training setting to other datasets. In this study, the dataset we used was from August 10, 2011 to August 9, 2012.

Here are sample tweets that researchers coded.

1) Bullying samples

John kicked me again in school. Why does he always bully me?

=> Bullying trace: Y / Teasing: N / The author's role: victim

I used to bully Anna when we were in the same class.

=> Bullying trace: Y / Teasing: N / The author's role: bully

2) Teasing samples

Ha Ha~ I like to call her b*!\$ all day long since we are good friends.

=> Bullying trace: Y / Teasing: Y / The author's role: bully

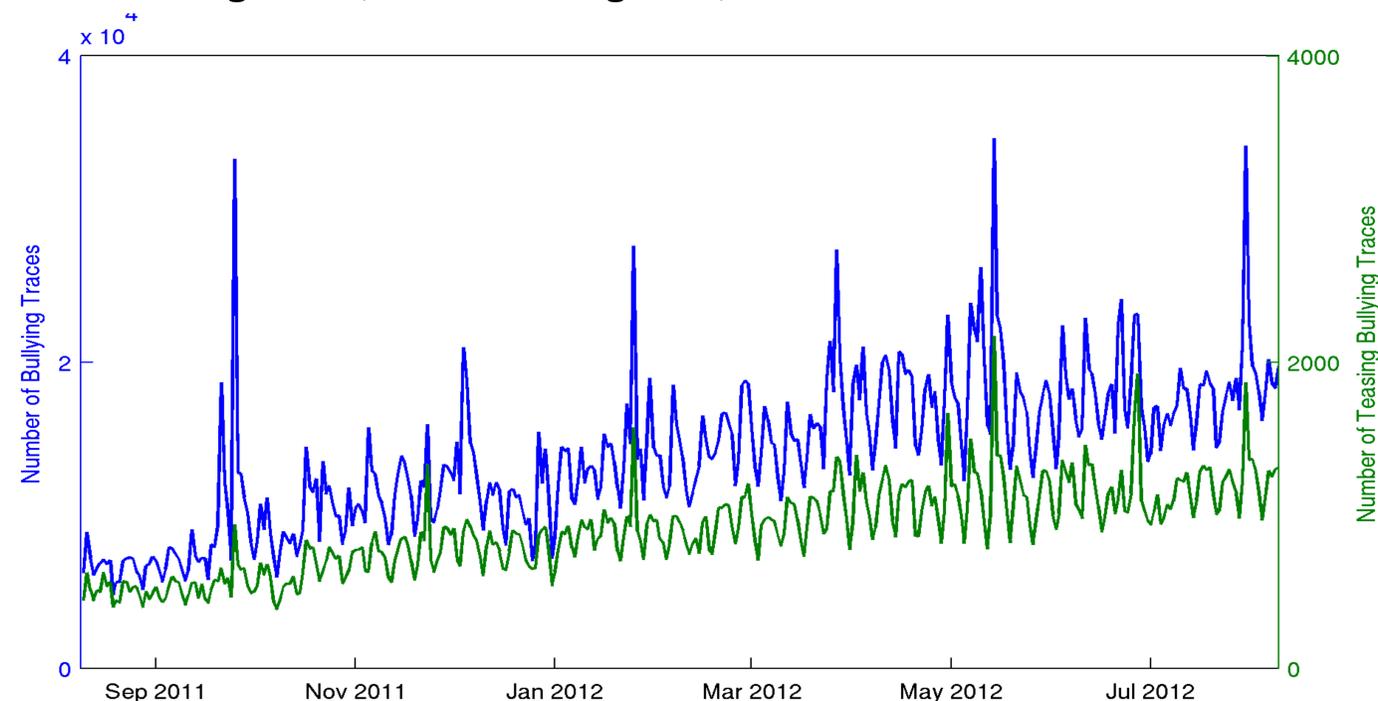
Data Analysis

1) To address our first goal, we evaluated whether posts about bullying episodes could be differentiated from posts about teasing. We used the machine learning data (August 10, 2011 - August 9, 2012) to distinguish bullying posts from teasing bullying posts.

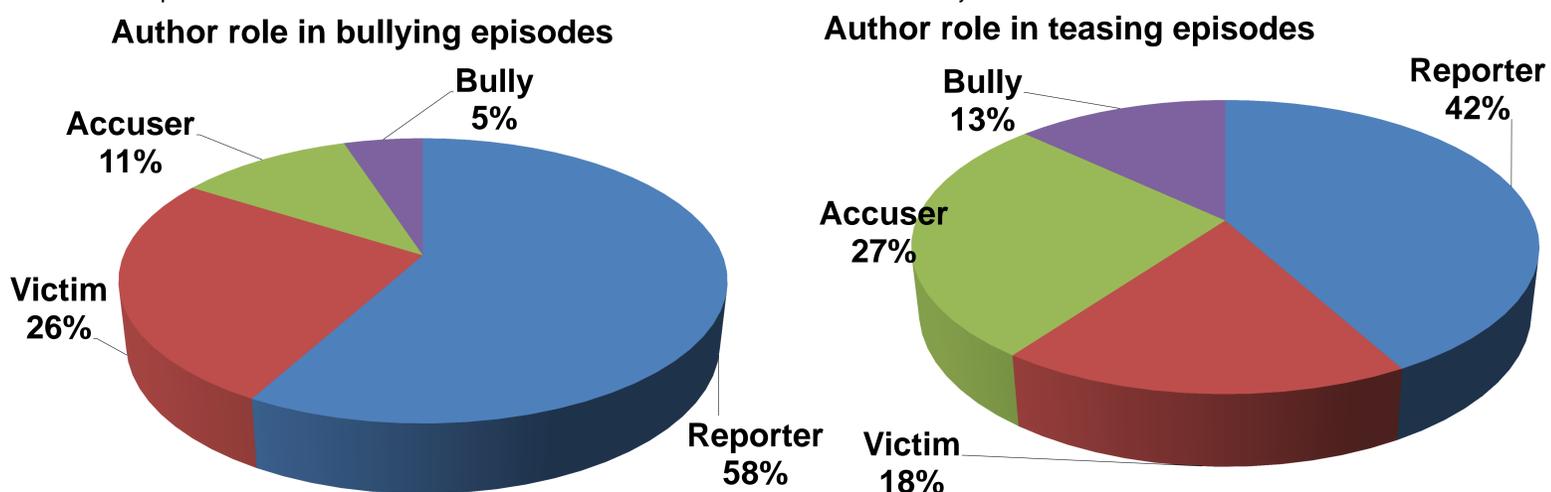
2) To address our second goal, we used the machine learning data (August 10, 2011 - August 9, 2012) to investigate who are most likely to tease in their social media posts. That is, we evaluated the associations between author roles and teasing or not.

Results

Question 1. Number of daily bullying episodes identified from teasing bullying posts between August 10, 2011 and August 9, 2012



Question 2.



Number and percentage of author types in social media posts about bullying and teasing episodes

	Bullying Episode Role Player			
	Accuser n (%)	Bully/Perpetrator n (%)	Reporter n (%)	Victim/Target n (%)
Bullying Episode	531243 (86%)	267233 (87%)	2761480 (95%)	1229549 (95%)
Z	-64.25	-38.8	28.67	20.51
Teasing Episode	89002 (14%)	40716 (13%)	140191 (5%)	60812 (5%)
Z	244.51	147.66	-109.1	-78.05