An Interdisciplinary Investigation of Temporal Aspects of Cyberbullying on Instagram

Wenxi Yang, Divya Prakash Sivakumar, Lu Cheng, Kaitlyn Schodt, Chun Shao, Deborah Hall, Yasin Silva
Arizona State University

Motivation

Cyberbullying, the use of online digital media to communicate false, embarrassing, or hostile information about another person is the most common online risk for adolescents. A key characteristic of cyberbullying is the repetitive nature, yet little is known about temporal aspects of cyberbullying. Drawing on a range of interdisciplinary techniques, the purpose of this study was to (1) identify the core temporal cyberbullying (CB) trends and properties in a large, real-world Instagram dataset and (2) investigate how temporal factors predict whether the media session was perceived as CB in this dataset.

Dataset

Figure 1. A sample Instagram social media session, from Hosseinmardi et al. (2015)

The dataset, initially used by Hosseinmardi et al. (2015), consisted of 2,218 Instagram social media sessions that had been coded (by humans) based on whether each session (the original Instagram post and its associated comments) was a CB or non-CB session, as a whole. Roughly 20% of the sessions had been coded as CB sessions.

The previous research did not, however, include information at the individual comment-level about CB. To address this, we employed an eXtreme Gradient Boosting Model (XGBoost, a tree-based model) to predict comment-level CB. The three features that were used in the prediction model are Word Count Vectors, Word Level TF-IDF, and Linguistic Inquiry & Word Count (LIWC). After integrating the three features to train the model, the accuracy level was about 91%.

After removing the sessions with temporal inconsistencies (e.g. comments with timestamps prior to their main post's timestamp), the final dataset consisted of 130,900 comments across 1,980 Instagram social media sessions, with 17,245 (15%) of the comments identified as CB by the prediction model.

Analysis and Results

Logistic regression

- Proportion of CB comments to total comments in a session (b = 8.35, SE = 0.53, p < .001, positive relationship) and average time interval between all CB comments and a session's original post (b = -3.02 * 10^{-8} , SE = 1.03 * 10^{-8} , p < .001, negative relationship) emerged as significant predictors of a media session being perceived as CB overall.
- Proportion of CB comments within a media session was the most influential predictor.

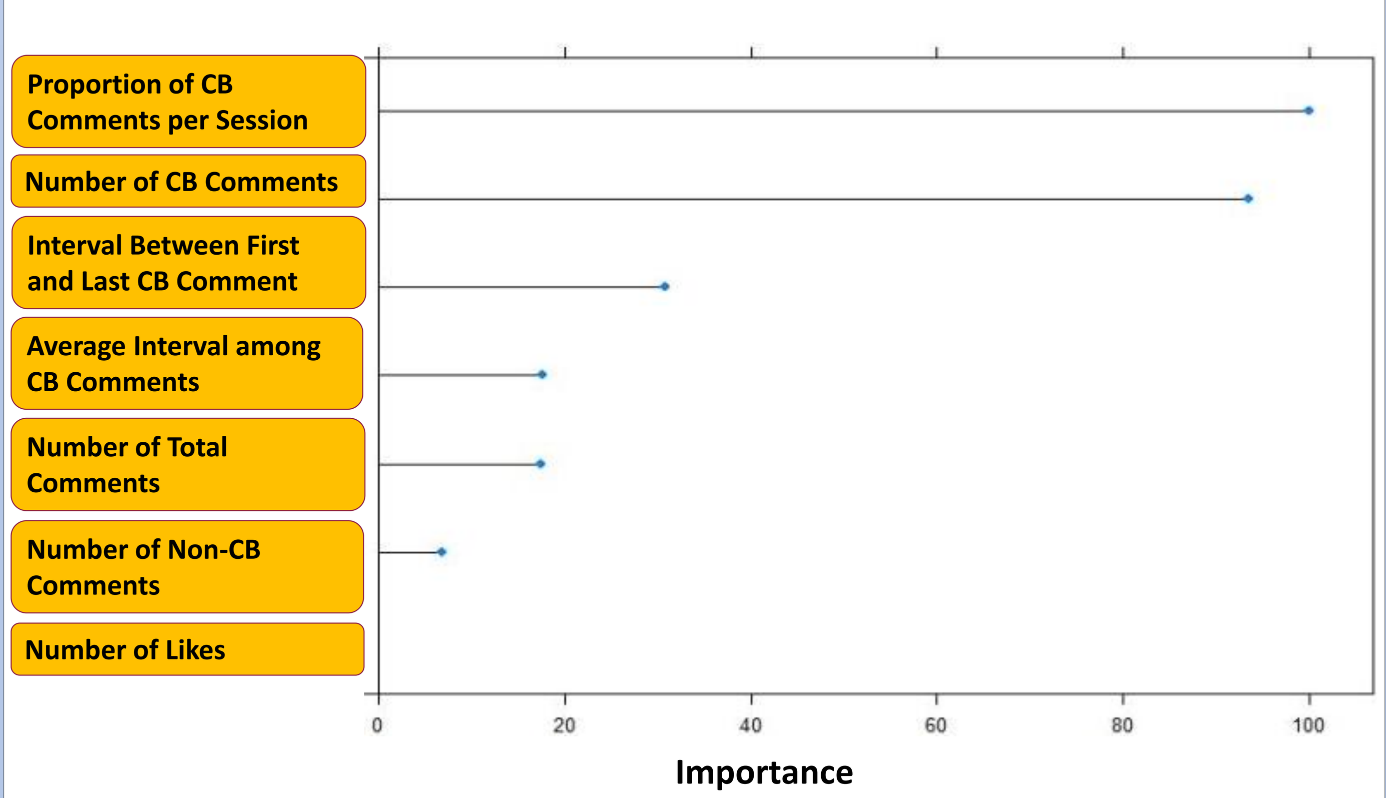
Random Forest

- A random forest analysis using the variables presented in Table 1 was performed to (1) predict session-level CB identification, and (2) indicate the importance of each of the variables listed in predicting session-level CB. The model was trained and tested using a 10-fold cross-validation method. The highest accuracy level was used to select the optimal model.
- The optimal prediction model was achieved when mtry (number of variables randomly sampled at each split) was 2 and ntree (number of trees to be grown) was 200. The final value used for the model was cp = 0.01015038. Accuracy level reached approximately 75%, with $kappa$ = .38.

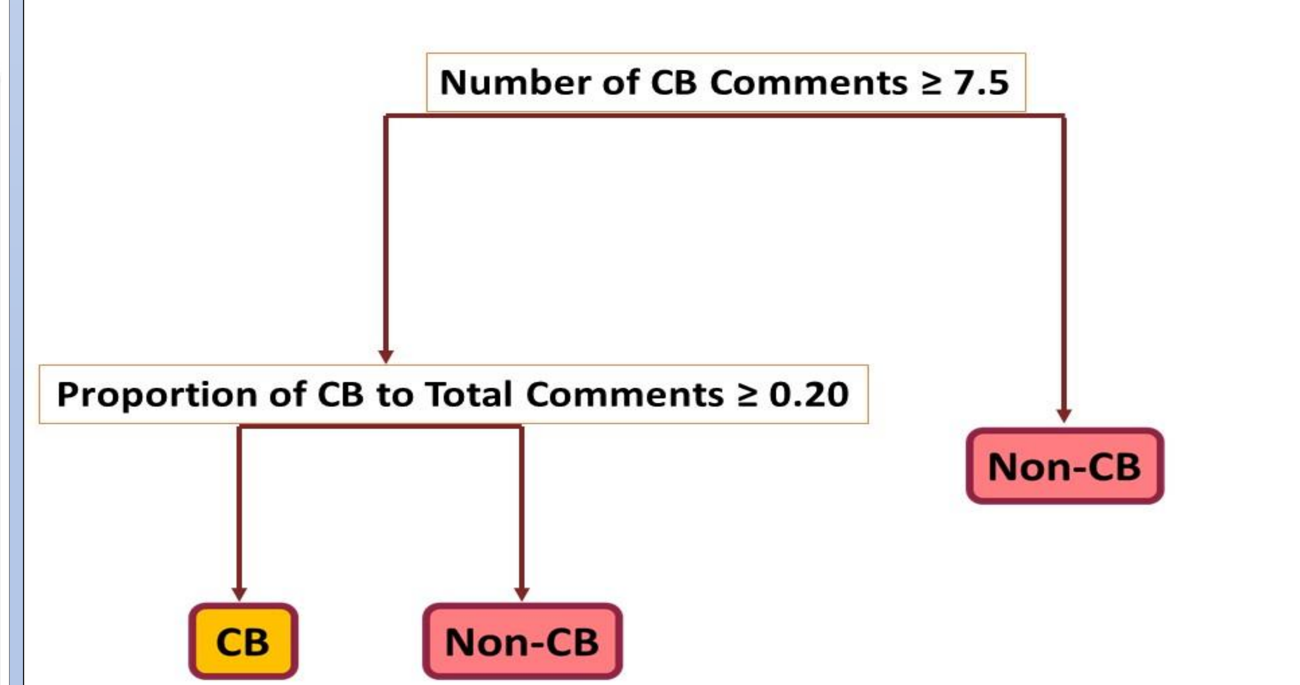
Variables Per Session	Min	Max	Mean	Median	SD
#CB comments	0	94	8.705	5	10.02
#Total Comments	7	147	63.03	51	41.94
#non-CB Comments	1	139	54.33	44	36.85
Proportion of CB comments to total comments	0	93.33%	14.51%	11.96%	11.22%
Time interval between first and last CB comments (minutes)	0	1,452,409 (1008.62 days)	82,726.07 (57.45 days)	3,300 (2.29 days)	169530.88 (117.73 days)
Average interval between all CB comments (minutes)	0	1,532,085 (1063.95 days)	2,128.57 (1.48 days)	41,061.37 (28.51 days)	114187.22 (79.30 days)
# Likes	1	782,434	9,698	2,001	29120.36
Session-level CB					
Comment-level CB					
Posting time of first CB comment					
Posting time of last CB comment					

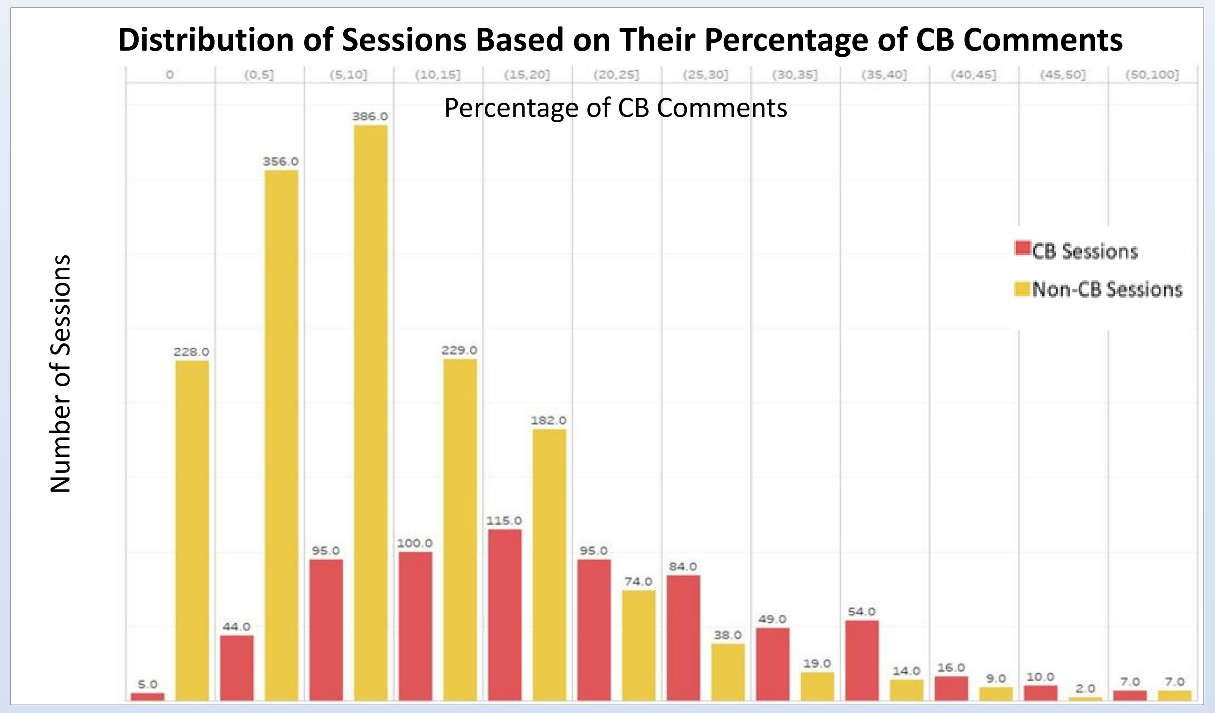
Table 1. Variables used in the random forest analysis

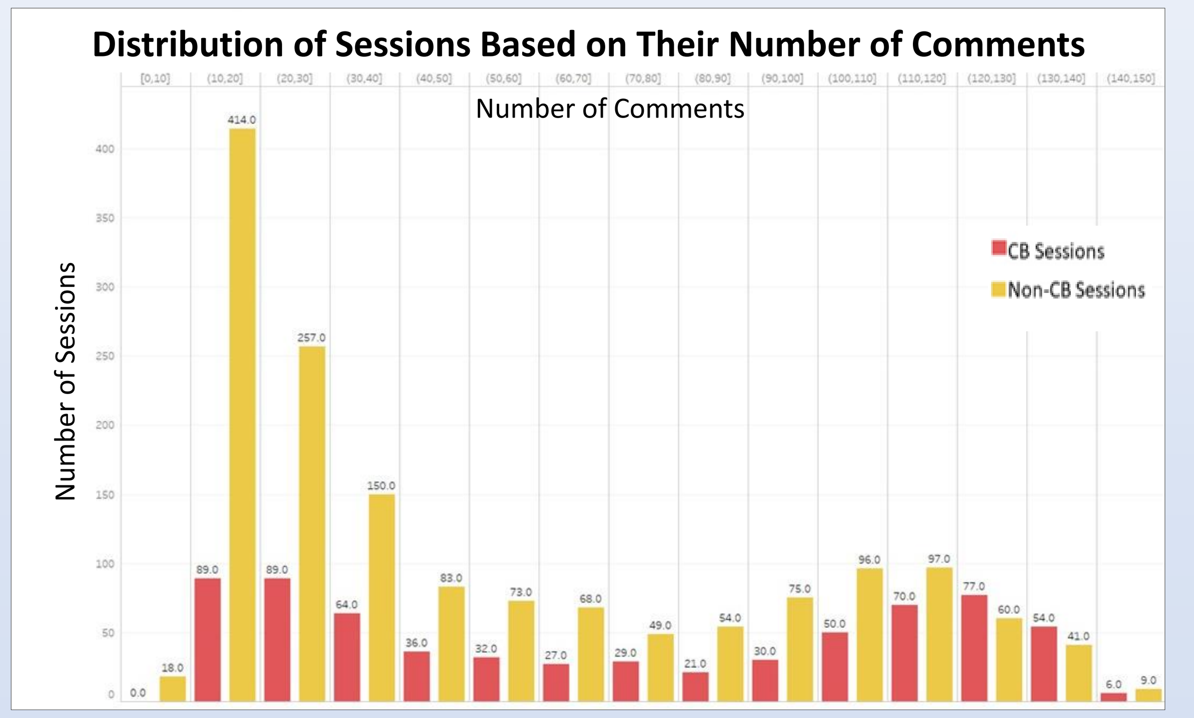
Ranking of the Predictors

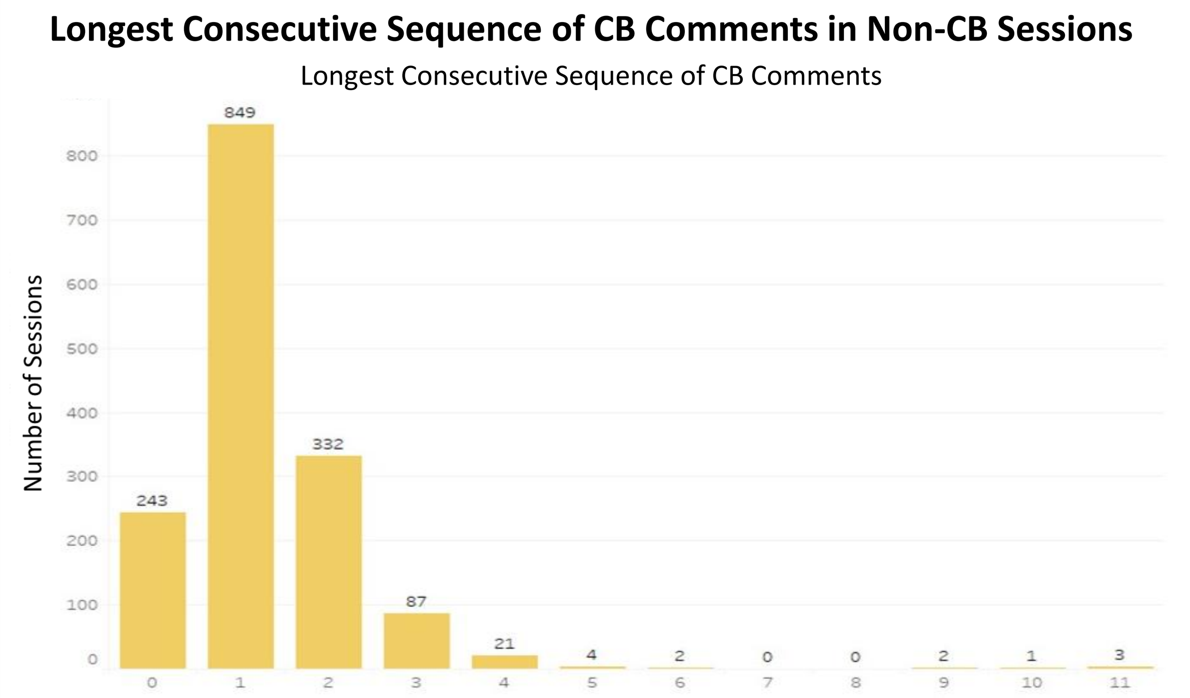


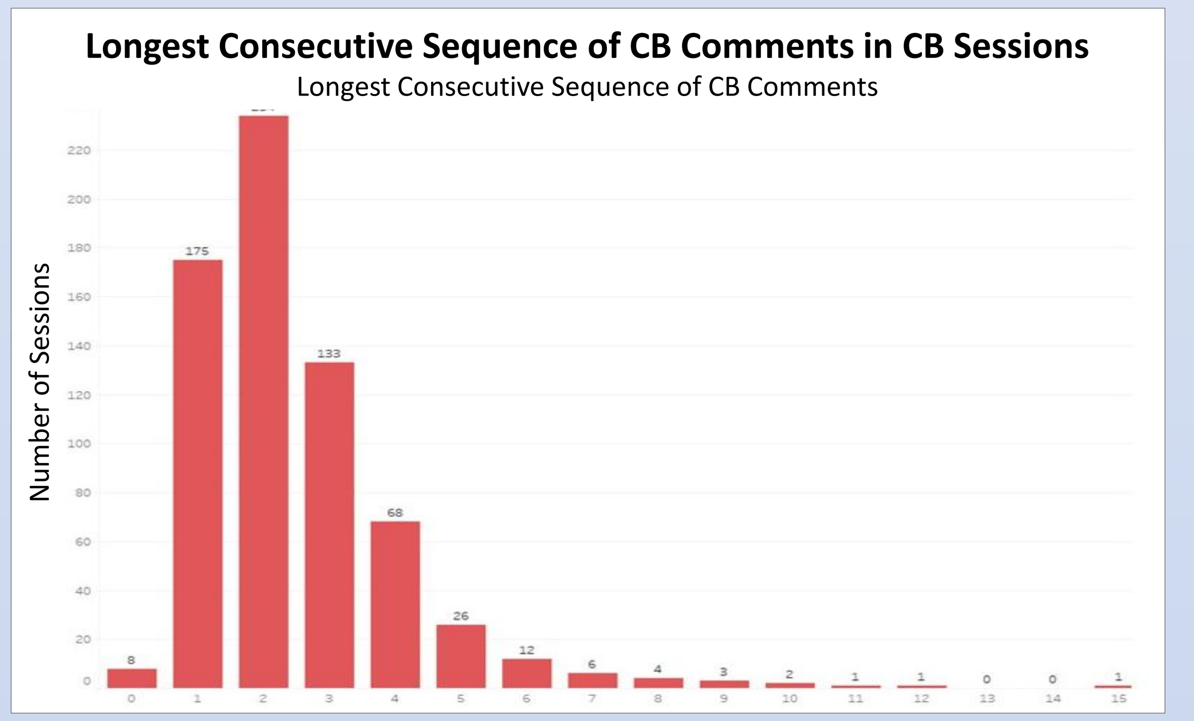
Classification Tree

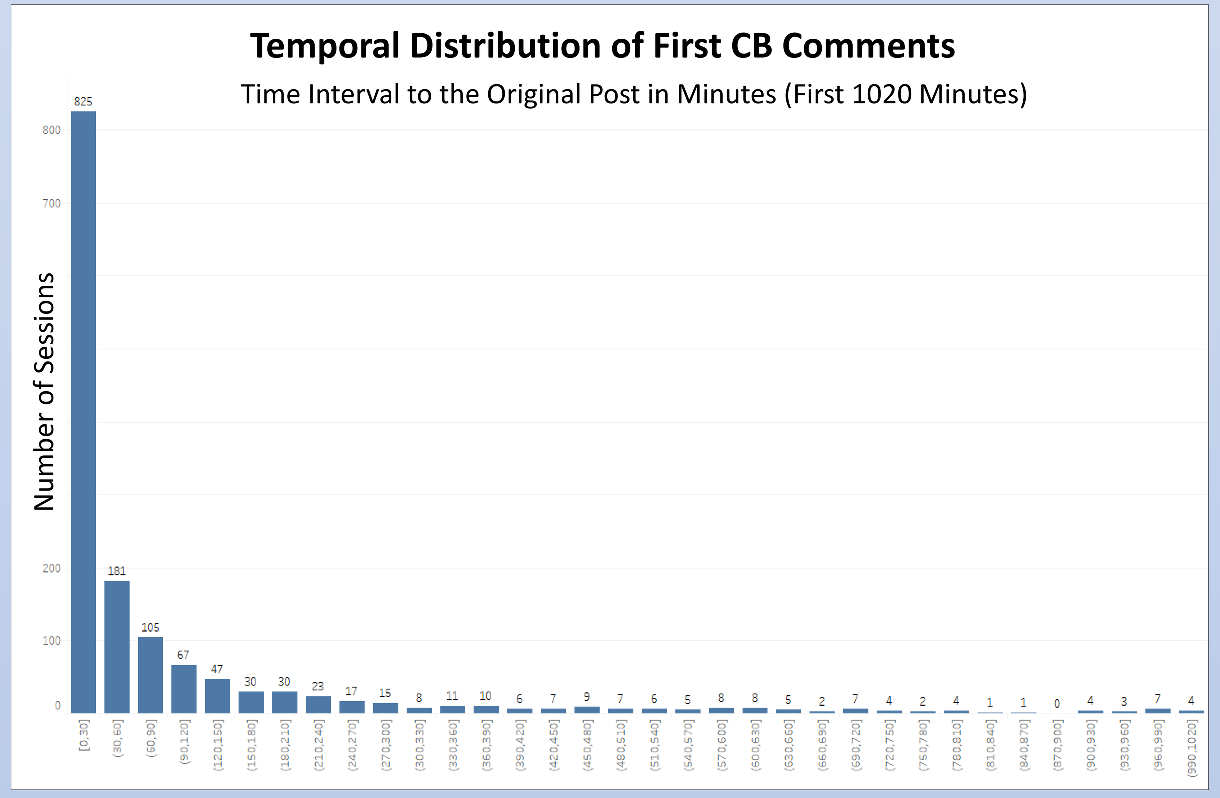


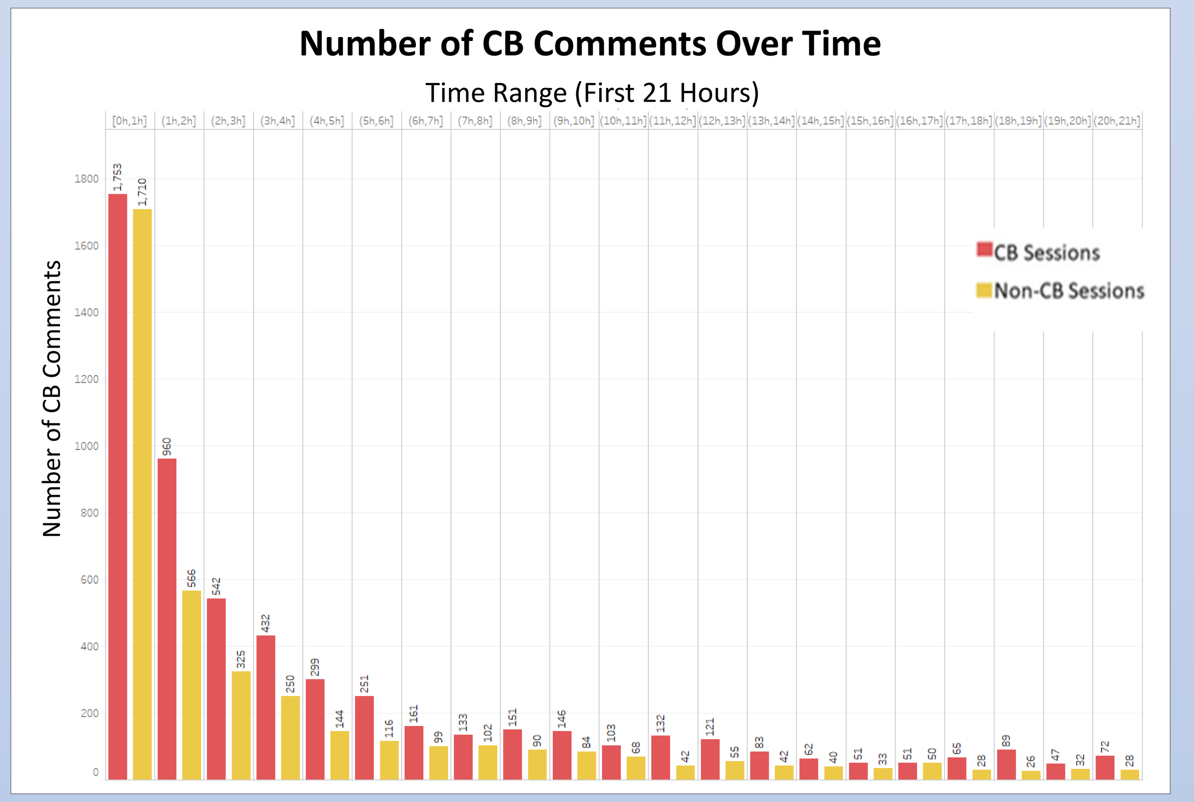


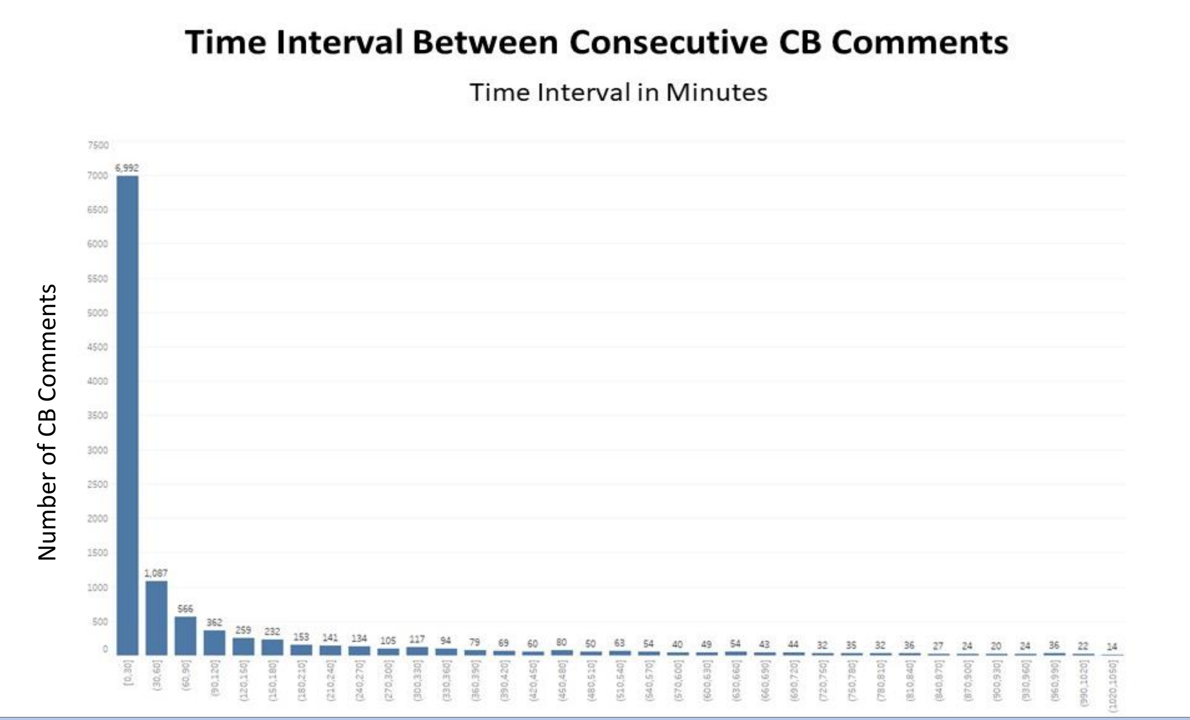


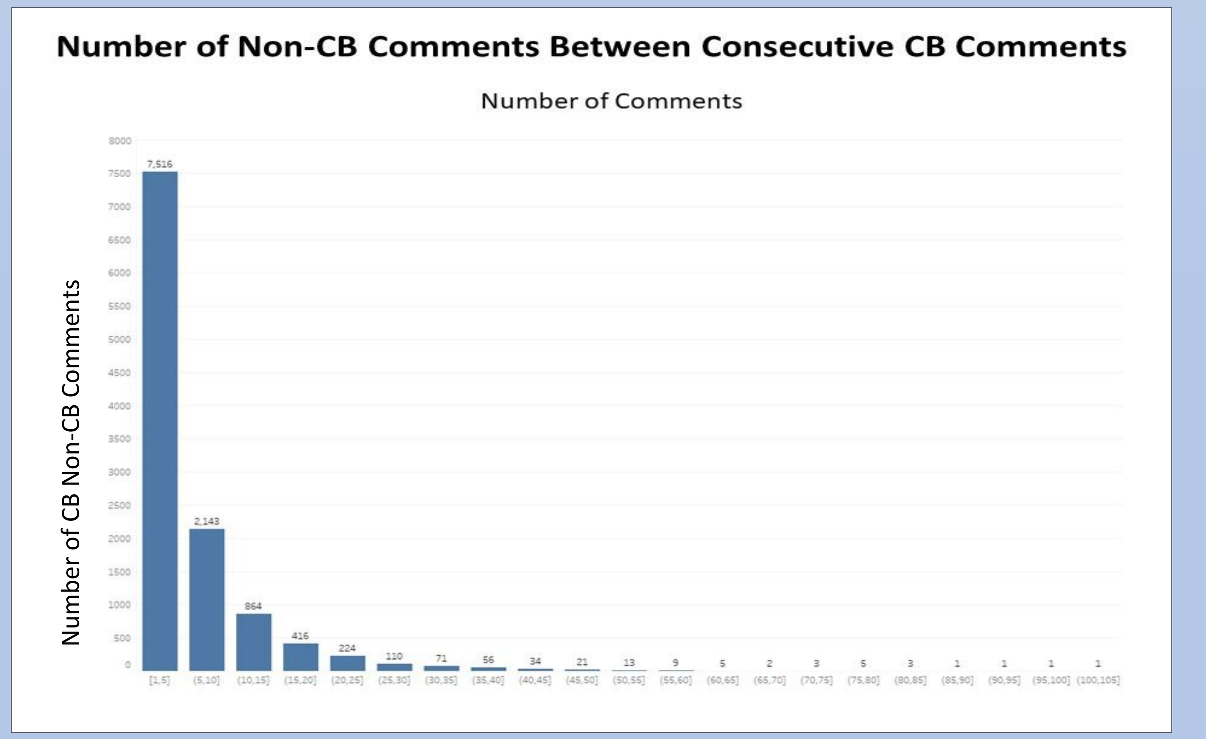












http://bullyblocker.asu.edu

This work was supported by National Science Foundation Award # 1719722