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.

**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**

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 130,908 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 ($\beta = 8.35, SE = 0.53, p < .001$, positive relationship) and average time interval between all CB comments and a session’s original post ($\beta = -3.02 \times 10^{-7}, SE = 1.03 \times 10^{-6}, p < .001$, 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 reached approximately 75%, with $\kappa = .38$.

### Classification Tree

- Variables used in the random forest analysis.

**Ranking of the Predictors**

- Proportion of CB comments to total comments $\geq 7.5$
- Number of CB Comments $\geq 2.0$
- Time between first and last CB comments
- Average time intervals between all CB comments
- Proportion of CB comments to total comments
- Number of Likes

**Variables Per Session**

<table>
<thead>
<tr>
<th>Variables Per Session</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB comments</td>
<td>0</td>
<td>94</td>
<td>3.755</td>
<td>5</td>
<td>10.02</td>
</tr>
<tr>
<td>Total Comments</td>
<td>7</td>
<td>147</td>
<td>63.03</td>
<td>51</td>
<td>41.94</td>
</tr>
<tr>
<td>Non-CB Comments</td>
<td>1</td>
<td>139</td>
<td>54.71</td>
<td>44</td>
<td>36.85</td>
</tr>
<tr>
<td>Proportion of CB comments to total comments</td>
<td>0</td>
<td>87.33%</td>
<td>14.51%</td>
<td>11.96%</td>
<td>51.22%</td>
</tr>
<tr>
<td>Time interval between first and last CB comments (minutes)</td>
<td>0</td>
<td>1,452,489</td>
<td>82,720.07</td>
<td>3.790</td>
<td>169,790.06</td>
</tr>
<tr>
<td>Average interval between all CB comments (minutes)</td>
<td>0</td>
<td>1,532,084</td>
<td>2,128.77</td>
<td>41,061.37</td>
<td>141,672.22</td>
</tr>
<tr>
<td>Likes</td>
<td>1</td>
<td>782,454</td>
<td>9,698</td>
<td>2,081</td>
<td>29,120.30</td>
</tr>
<tr>
<td>Session-level CB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comment-level CB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posting time of first CB comment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posting time of last CB comment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td># Likes</td>
<td>1</td>
<td>782,454</td>
<td>9,698</td>
<td>2,081</td>
<td>29,120.30</td>
</tr>
<tr>
<td># CB comments</td>
<td>0</td>
<td>87.33%</td>
<td>14.51%</td>
<td>11.96%</td>
<td>51.22%</td>
</tr>
<tr>
<td># Total Comments</td>
<td>7</td>
<td>147</td>
<td>63.03</td>
<td>51</td>
<td>41.94</td>
</tr>
<tr>
<td># Non-CB Comments</td>
<td>1</td>
<td>139</td>
<td>54.71</td>
<td>44</td>
<td>36.85</td>
</tr>
<tr>
<td>Time Interval to the Original Post in Minutes (First 21 Hours)</td>
<td>0</td>
<td>1,452,489</td>
<td>82,720.07</td>
<td>3.790</td>
<td>169,790.06</td>
</tr>
</tbody>
</table>

**Accuracy Level CB**

- The optimal prediction model was achieved when $mtry = 8$ 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.

This work was supported by National Science Foundation Award # 1719722