

Exploring How Social Media Content Feeds Are Manipulated

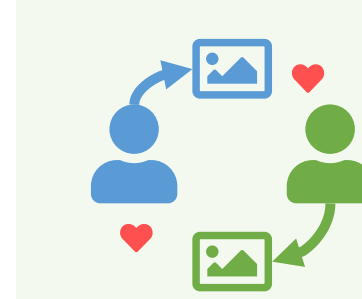
Janith Weerasinghe, Drexel University Cynthia Gill, Springfield High School/Drexel University
Jaime Richards, Sterling High School/Drexel University Damon McCoy, NYU Rachel Greenstadt, Drexel University

Overview

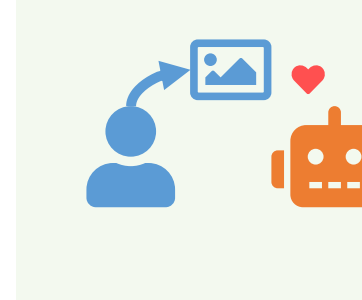
- Online Social Networks (OSNs) utilize curation algorithms to present relevant content to users.
- These algorithms can be manipulated by users with various intentions.
- We investigate common methods used by manipulators as part of a larger project looking to improve OSN defenses against manipulators.

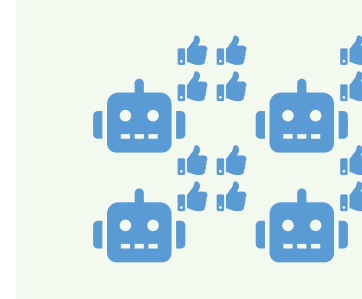
Manipulation Strategies

**Troll Farms**

**Manual Reciprocation**

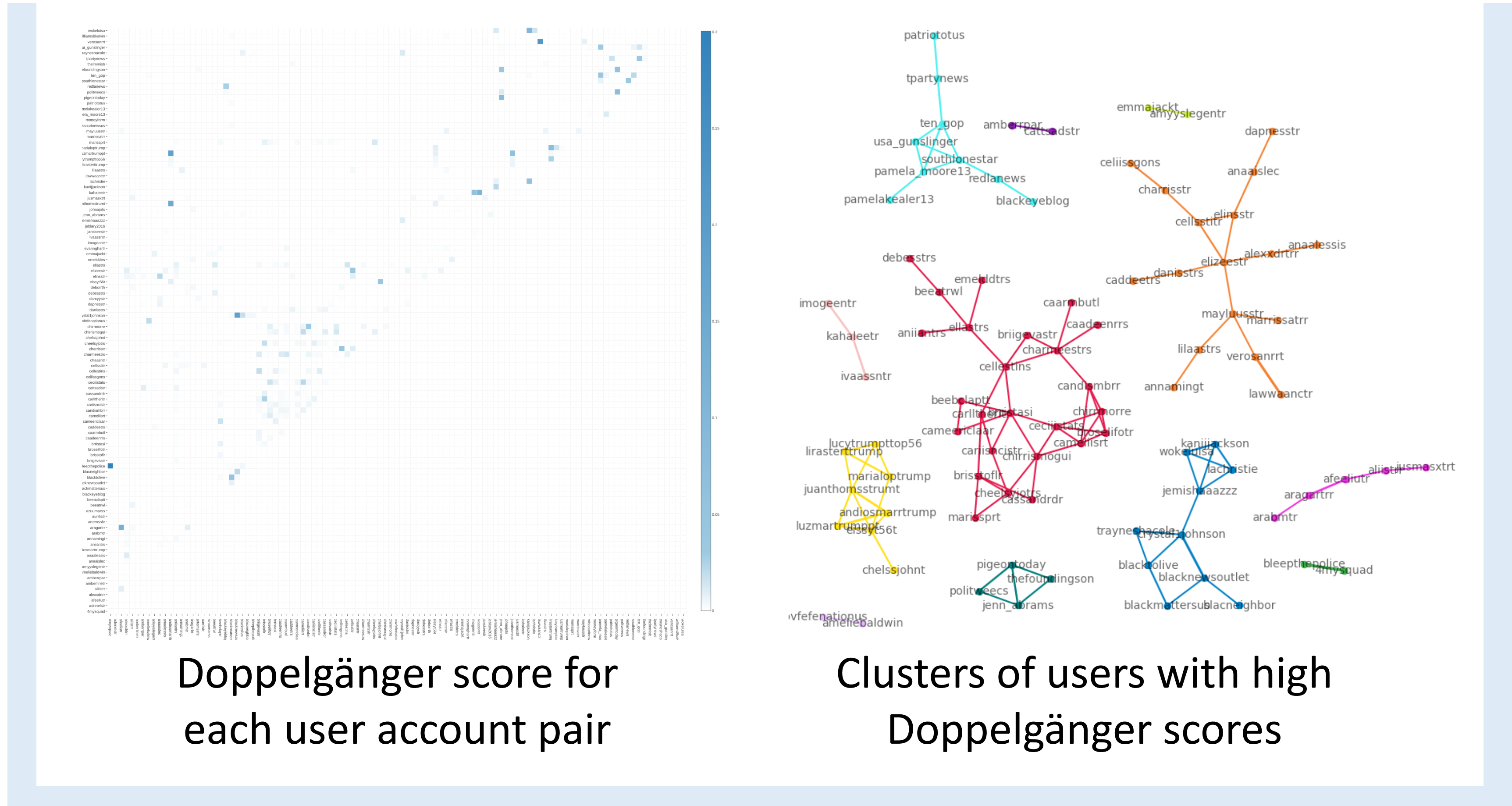
**Manual Manipulation Campaigns**

**Automated Actions**

**Bot-Based Manipulation**

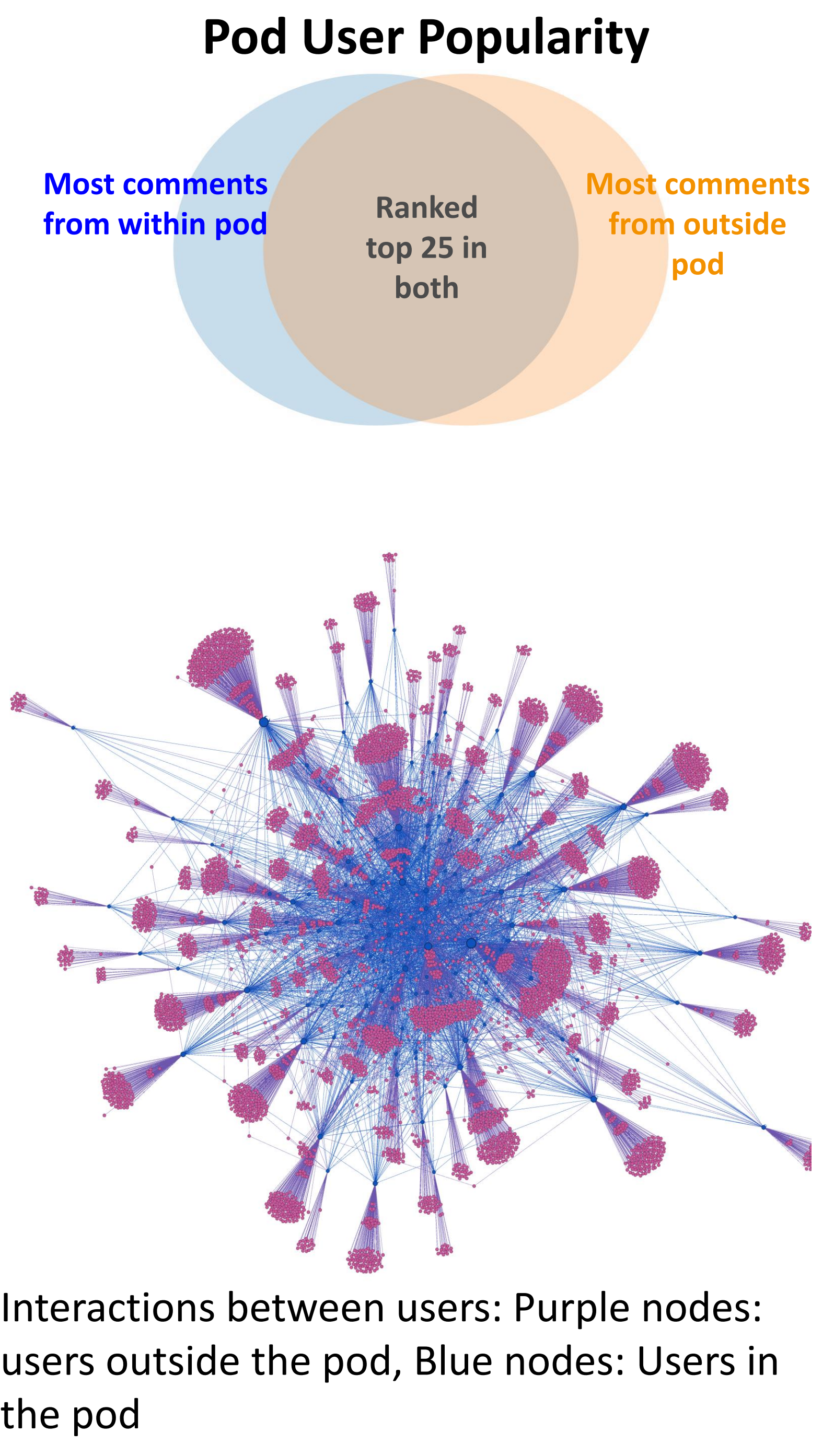
Analysis of Troll Farms

- Dedicated trolls produce content for a large number of accounts. These accounts may be identified via a shared writing style, or stylometry.
- Doppelgänger Finder [1] uses stylometry based authorship attribution:
 $\Pr(A \rightarrow B)$: Probability of A 's document being attributed to B
 $\Pr(A \rightarrow B) \times \Pr(B \rightarrow A)$ is a measure of how similar A & B are.
- We use Doppelgänger Finder to identify Russian Troll Accounts [2] that share similar writing styles.



Analysis of Manual Reciprocation Schemes

- Manual reciprocation schemes are prevalent among Instagram users, who join groups, or “pods,” where the group members like/comment on each other’s content.
- Data set: 2 Instagram pods, 17k users, and 30k interactions
- Observations
 - High connectivity within the pod members.
 - Users with large number of interactions within the pod have a large number of interactions outside the pod as well.
 - Content: fashion, photography, travel, fitness



Future Work

- Short Term: Understanding OSN Manipulation Ecosystem**
- Analysis of troll farms
 - Detect native languages of trolls.
 - Analysis of manual reciprocation schemes
 - Collect data from more pods.
 - Analyze engagement patterns and engagement quality.
 - Do users in pods get more interaction than other users?
- Long Term:**
- Investigate the defenses OSNs have deployed to mitigate manipulation.
 - Design more resilient defenses by leveraging algorithms and data internal and external to the OSN’s platform.

K-12 Education Integration

Curricular Unit: Social Media Networks and Privacy Implications

Goal Statement: When the order in which information presented by curation algorithms is manipulated adversarially, the trustworthiness of this information is eroded in such a way that is not apparent to the user. Students will learn to be skeptical and cognizant of potential misinformation.

References

[1] Afroz, Sadia, et al. "Doppelgänger finder: Taking stylometry to the underground." Security and Privacy (SP), 2014 IEEE Symposium on. IEEE, 2014.

[2] Oliver Roeder, “Why We’re Sharing 3 Million Russian Troll Tweets” FiveThirtyEight, August 6, 2018, <https://fivethirtyeight.com/features/why-were-sharing-3-million-russian-troll-tweets/>

Acknowledgements

REthink@Drexel
Research Experiences for Teachers Site for Machine Learning to Enhance Human-Centered Computing

This material is based upon work supported by the National Science Foundation under Grants No. CNS-1711773, CNS-1813697. Any opinions, findings, and conclusions expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

