Measuring #GamerGate: A Tale of Hate, Sexism, and Bullying

Despoina Chatzakou[†], Nicolas Kourtellis[‡], Jeremy Blackburn[‡]
Emiliano De Cristofaro[♯], Gianluca Stringhini[♯], Athena Vakali[†]

[†]Aristotle University of Thessaloniki [‡]Telefonica Research [♯]University College London

International Workshop on Computational Methods in CyberSafety, 2017

Social Networking Services

Aggressive & Bullying behavior

- **Cyberbullying.** Repeated and hostile behavior by a group or an individual, using electronic forms of contact.
- **Cyber-aggression.** Intentional harm delivered by the use of electronic means to a person or a group of people who perceive such acts as offensive, derogatory, harmful, or unwanted.

Gamergate controversy

- A coordinated campaign of harassment in the online world.
- It started with a blog post by an ex-boyfriend of independent game developer Zoe Quinn, alleging sexual improprieties.
- It quickly evolved into a polarizing issue, involving sexism, feminism, and "social justice," taking place on social media like Twitter.

Gamergate controversy provides us a unique point of view into online harassment campaigns.

Our goals

- Proposal of a principled methodology to collect content related to aggressive and bullying activities.
- Gamergate specific:
  - Quantification of this controversy.
  - Exploration of the existing differences between Gamergaters and random Twitter users.

Outline

1. Abusive dataset building
   - Data collection
   - Data processing
2. Measurement results
   - How Active are Gamergaters?
   - How Social are Gamergaters?
   - Are Gamergaters Suspended More Often?
3. Conclusions

Outline

1. Abusive dataset building
   - Data collection
   - Data processing
2. Measurement results
   - How Active are Gamergaters?
   - How Social are Gamergaters?
   - Are Gamergaters Suspended More Often?
3. Conclusions

Outline

1. Abusive dataset building
    - Data collection
    - Data processing
2. Measurement results
    - How Active are Gamergaters?
    - How Social are Gamergaters?
    - Are Gamergaters Suspended More Often?
3. Conclusions

Overall process

Steps.

- 1. Select seed keyword(s).
- 2. Create dynamic list of keywords.
- 3. Crawl tweets.
- 4. Collect a random sample*.

* Complements the abusive-related dataset with cases that are less likely to contain abusive content.

Seed keyword(s)

Select seed keyword(s) which are likely to relate with abusive incidents.

- E.g., #GamerGate, #BlackLivesMatter, #PizzaGate.
- Set of hate- or curse-related words, e.g., Hatebase database.

At the time, t_i,

the lists of words to be used for filtering posted texts includes only the seed word(s): $L(t_1) =< seed(s) >$.

Dynamic list of keywords (I)

- Filter keywords list to select abusive-related content.
- Update dynamically - in consecutive time intervals - the filtering list.
- Depending on the topic under examination: update the filtering list at different time intervals.

Keywords list, $L(T)$

In $T = \{t_1, t_2, ..., t_n\}$ the $L(T)$ equals to:

$L(t_i) =< seed(s), kw_1, kw_2, kw_N >$,
where kw_j is the jth top keyword in time period $\Delta T = t_i - t_{i-1}$.

Dynamic list of keywords (II)

Update dynamic list for $t_i \rightarrow t_{i+1}$

- Step 1. Investigate the texts posted at $t_{i-1} \rightarrow t_i$.
- Step 2. Extract the top N keywords based on their frequency of appearance.
- Step 3. Update $L(t_i)$ with the up-to-date top N keywords along with the seed word(s).

Use of the updated list at the time period: $t_i \rightarrow t_{i+1}$.

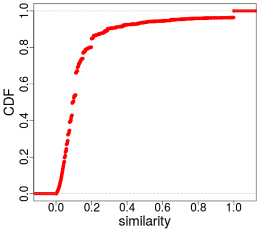
Outline

Preprocessing

– **Cleaning.** Removal of stop words, URLs, punctuations marks, normalization (repetitive characters elimination).

– **Spam removal.** Based on the number of hashtags, and duplications.

- Study of hashtags and duplication distributions to find optimal cutoffs.
- Avg. # hashtags: 0 to 17.
- Hashtags: we set the limit to 5.
- Similarity of tweets: Levenshtein distance.
- About 5% of the users have a high percentage of similar posts.



Final dataset: 659k GG-related tweets, 1M random tweets.

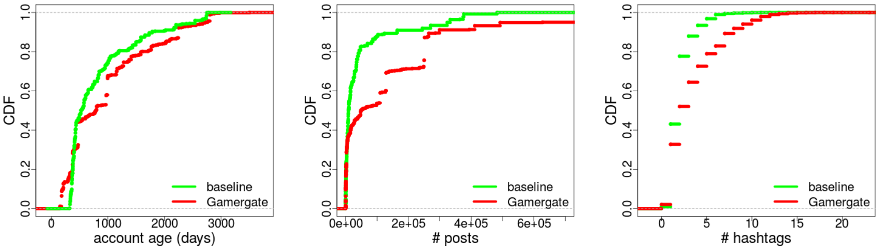
Outline

1. Abusive dataset building
   - Data collection
   - Data processing
2. Measurement results
   - How Active are Gamergaters?
   - How Social are Gamergaters?
   - Are Gamergaters Suspended More Often?
3. Conclusions

Outline

1. Abusive dataset building
   - Data collection
   - Data processing
2. Measurement results
   - How Active are Gamergaters?
   - How Social are Gamergaters?
   - Are Gamergaters Suspended More Often?
3. Conclusions

Account age, posts, hashtags



- GGers tend to have older accounts $->$ They are not bots.
- GGers are significantly more active than random Twitter users (more posts and hashtags).

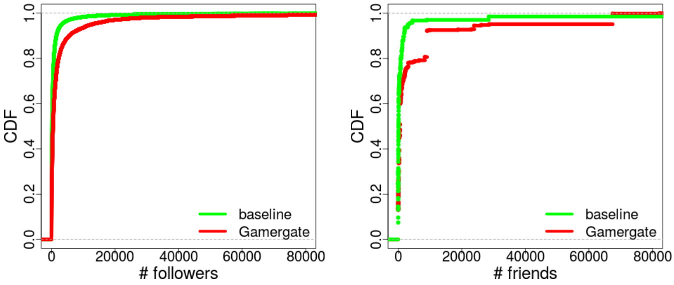
Favorites, lists, URLs, mentions

- GGers have more favorites and topical lists declared than random users.
- GGers post more URLs in an attempt to disseminate information about their “cause”.
- GGers make more mentions within their posts $- >$ higher number of direct attacks compared to random users.

Outline

1. Abusive dataset building
   - Data collection
   - Data processing
2. Measurement results
   - How Active are Gamergaters?
   - How Social are Gamergaters?
   - Are Gamergaters Suspended More Often?
3. Conclusions

Followers, friends



- GGers tend to have more friends and followers than random users.
- The controversy appears to be a clear "us vs. them" situation.
- Existence of in-group membership $->$ heightens the likelihood of relationship formation.

Emoticons, uppercases, sentiment, emotion

Emoticons and "shouting" by using all capital letters: two common ways to express emotion.

- GGers and random users use emoticons at about the same rate.
- GGers tend to use all uppercase less often than random users.

Sentiment, Offense, & Emotion

- GGers post tweets with a generally more negative sentiment $->$ large proportion of offensive posts.
- GGers use more hate words than random users (Hatebase database).
- GGers and random users do not differ substantially in a variety of emotions: anger, disgust, fear, sadness, surprise.
- GGers are less joyful $->$ they are not necessarily angry, but they are apparently not happy.

Outline

Twitter Reaction to Aggression

	active	deleted	suspended
Random users	67%	13%	20%
Gamergate	86%	5%	9%

- Focus on a sample of 33k users from both the GG and random datasets.
- Users tend to be suspended more often than deleting their accounts by choice.
- Random users are more prone to be suspended or delete their accounts than GGers.

Outline

1. Abusive dataset building
   - Data collection
   - Data processing
2. Measurement results
   - How Active are Gamergaters?
   - How Social are Gamergaters?
   - Are Gamergaters Suspended More Often?
3. Conclusions

Summary

- GGers use Twitter as a mechanism for broadcasting their ideals (hashtags, mentions).
- GGers appear to be Twitter savvy users and quite engaged with the platform (posts, participating lists, favorites).
- GGers are more well-connected within their network (followers, friends).
- GGers express with more negative sentiment overall, but they only differ significantly from random users with respect to joy.
- GGers are less likely to be suspended due to the inherent difficulties in detecting and combating online harassment activities.

Future work

- Conduction of a more in-depth study of Gamergate controversy, focusing on how it evolved over time.
- Consideration of additional features, e.g., network-based, to further examine the differences among the GGers and random users.
- Automatically detect abusive users (upcoming HyperText paper: stay tuned!)

Questions?

This work has been funded by the European Commission as part of the ENCASE project (H2020-MSCA-RISE), under GA number 691025.