ESHWAR CHANDRASEKHARAN

Presented by Srividhya Chandrasekharan and Anu Yadav

ABOUT HIM

- Social Computing, NLP, Machine Learning and Social Networks
- *Currently working on : Combating Abusive Behavior in Online Communities* with Dr.Eric Gilbert

RESEARCH INTERESTS

FAMOUS FOR

You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech

ESHWAR CHANDRASEKHARAN, Georgia Institute of Technology
UMASHANTHI PAVALANATHAN, Georgia Institute of Technology
ANIRUDH SRINIVASAN, Georgia Institute of Technology
ADAM GLYNN, Emory University
JACOB EISENSTEIN, Georgia Institute of Technology
ERIC GILBERT, University of Michigan

In 2015, Reddit closed several subreddits—foremost among them r/fatpeoplehate and r/CoonTown—due to violations of Reddit's anti-harassment policy. However, the effectiveness of banning as a moderation approach remains unclear: banning might diminish hateful behavior, or it may relocate such behavior to different parts of the site. We study the ban of r/fatpeoplehate and r/CoonTown in terms of its effect on both participating users and affected subreddits. Working from over 100M Reddit posts and comments, we generate hate speech lexicons to examine variations in hate speech usage via causal inference methods. We find that the *ban worked for Reddit*. More accounts than expected discontinued using the site; those that stayed drastically decreased their hate speech usage—by at least 80%. Though many subreddits saw an influx of r/fatpeoplehate and r/CoonTown "migrants," those subreddits saw no significant changes in hate speech usage. In other words, other subreddits did not inherit the problem. We conclude by reflecting on the apparent success of the ban, discussing implications for online moderation, Reddit and internet communities more broadly.

CCS Concepts: • **Human-centered computing** → *Empirical studies in collaborative and social computing*;

Additional Key Words and Phrases: online communities, hate speech, moderation, banning, causal inference.

ACM Reference format:
Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech. *Proc. ACM Hum.-Comput. Interact.* 1, 2, Article 31 (November 2017), 22 pages.
https://doi.org/10.1145/3134666

THE NEW REDDIT JOURNAL OF SCIENCE

comments other discussions (26) filter by field

Trending: **Harvard and MIT researchers have developed smart tattoo ink capable of monitoring dehydration and blood…**

46.9k

Reddit's bans of r/coontown and r/fatpeoplehate worked--many accounts of frequent posters on those subs were abandoned, and those who stayed reduced their use of hate speech ▸ comp.social.gatech.edu

18 days ago by asbruckman Professor | Interactive Computing x2

6814 comments share report

Top 200 Comments show 500

sorted by: new (suggested)

[+] ***Comment removed*** *17 days ago (4 children)*

[–] **Stuck_In_the_Matrix** 20 points 17 days ago

/u/asbruckman

Thanks for your wonderful work! I'm glad my data dumps have led to some very cool academic studies. A few things:

1) redditanalytics.com is no more -- it became pushshift.io (You can take out redditanalytics.com if possible and just give credit to pushshift.io)

2.) Are you using the updated monthly data at all? I'll be honest and admit that I haven't read the entire paper, but you appear to have used the 2015 Corpus. I'm also adding monthly updates to https://files.pushshift.io

RECENT PUBLICATIONS

- Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, Eric Gilbert. You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech**,** *CSCW 2018*
- Eshwar Chandrasekharan, Mattia Samory, Anirudh Srinivasan, Eric Gilbert. The Bag of Communities: Identifying Abusive Behavior Online with Preexisting Internet Data , *CHI 2017*
- Ari Schlesinger, Eshwar Chandrasekharan, Christina Masden, Amy Bruckman, W Keith Edwards, Rebecca Grinter. Situated Anonymity: Impacts of Anonymity, Ephemerality, and Hyper-Locality on Social Media**,** *CHI 2017*

You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech (2018)

- Reddit's decision to close r/fatpeoplehate and r/CoonTown
- "You f*cking f*tass, you made the decision to be a fat f*ck after you decided to stuff your fat f*cking face instead of acting like a normal human being." - a highly-upvoted r/fatpeoplehate comment
- Research focus -
    - Effect of ban on contributors to banned subreddits
    - Effect of ban on subreddits that saw influx of banned subreddit users

CAUSAL EFFECTS OF THE BAN?

- Hate Speech lexicons made public @ https://tinyurl.com/hatewords
- Users of banned subreddits:-
  - Left
  - Active and migrated; decrease by >80% in their hate speech usage
- Invaded subreddits - NO significant changes in hate speech use
- Banning - cut down outlets to propagate hate speech
- Reddit banned copycats
- Subreddits and members - didn't want to attract attention of site admins
- Reddit’s ban - made hateful people migrate to darker parts of internet
- Implications - 1,536 r/fatpeoplehate users have exact match usernames on Voat.com.

Footprints on Silicon: Explorations in Gathering Autobiographical Content (2015)

- Analyzed email to extract content that could be of autobiographical nature
- Built the classifier by mining discriminating features like textual keywords, threads, labels and mail network properties
- Naive Bayes, Random Forest and LibSVM
- Accuracy and Precision

Results

- Textual keywords and labels were most effective during classification when considered by themselves
- Email network properties and thread counts were not very good indicators on their own, but when augmented with textual keywords and labels, they were observed to give improved performances


That's all Folks!


Presentermedia