

Saha, K., Chandrasekharan, E., & De Choudhury, M. (2019). Prevalence and Psychological Effects of Hateful Speech in Online College Communities. In *Proceedings of the 11th ACM Conference on Web Science.*,
<https://dl.acm.org/citation.cfm?id=3326032>

PREVALENCE AND PSYCHOLOGICAL EFFECTS OF HATEFUL SPEECH IN ONLINE COLLEGE COMMUNITIES

Koustuv Saha, Eshwar Chandrasekharan, Munmun De Choudhury

@kous2v | koustuv.saha@gatech.edu | koustuv.com







Hate Incidents on Campus Still Rising

Results of a new survey show that administrators are still reporting incidents targeting minority students.

By **Jeremy Bauer-Wolf** // February 25, 2019

27 COMMENTS 

HATEFUL SPEECH ON COLLEGE CAMPUSES

“..verbal and written words, and symbolic acts, that convey a grossly negative assessment of particular persons or groups based on their race, gender, ethnicity, religion, sexual orientation, or disability, which is not limited to a face-to-face confrontation or shouts from a crowd, but may also appear on T-shirts, on posters, on classroom blackboards, on student bulletin boards, in flyers and leaflets, in phone calls, etc.” (Kaplin, 1992)

The background is a dark blue space filled with white stars and a grid of glowing white lines. These lines connect various white-outlined icons representing digital technology and communication. The icons include: a lowercase 'i' in a circle, a speech bubble with a lowercase 's', a Pinterest 'P' logo, a keyhole, a YouTube logo, an envelope, a smartphone, a camera, a Twitter bird, a Facebook 'f' logo, a microphone, a play button, a speaker, a smartphone, an Apple logo, an Android robot, a LinkedIn 'in' logo, a camera, a gear, a play button, a double left arrow, a double right arrow, a crossed-out airplane, and a document with a checkmark.

ONLINE HATEFUL SPEECH

ONLINE HATEFUL SPEECH AND COLLEGE CAMPUSES

- Difficult to mitigate negative effects of online hateful speech
- Need to assess the pervasiveness of online hateful speech
- Need to understand the vulnerability in psychological wellbeing to marginalized communities on college campuses
- Present methods to do the above are limited:
 - Anecdotal / covered by popular media / discrete events
 - Social confounds, such as stigma

RESEARCH QUESTIONS

How **prevalent** is hateful speech in online college communities, across the **demographic categories** such as *gender, religion, race*, etc.?

How does **exposure** to online hate **affect** an individual's **expression** of their psychological state on social media, particularly **stress**?



/r/gatech chat here

↑
15
↓



Missed Julie Gordon White's AMA? Catch up here and learn how to grow your small business into an empire.

Promoted by Intuit_QuickBooks

[promoted](#) [save](#) [report](#)

↑
67
↓



Congrats! You just got in! 2018 Regular Decision Megathread

Submitted 1 month ago by [\[redacted\]](#) [CS - 0x7E3](#) [M](#) - announcement

[224 comments](#) [share](#) [save](#) [hide](#) [report](#) [crosspost](#)

↑
24
↓



Phase One Megathread: Summer/Fall 2018 Edition

Submitted 12 days ago by [\[redacted\]](#) [ME - 2019?](#) [M](#) - announcement

[244 comments](#) [share](#) [save](#) [hide](#) [report](#) [crosspost](#)

↑
30
↓



What do i do about a professor who threatened me? [self.gatech](#)

Submitted 6 hours ago * by [\[redacted\]](#)

[32 comments](#) [share](#) [save](#) [hide](#) [report](#) [crosspost](#)

↑
↓



Are there any body fat percentage monitors on campus? [self.gatech](#)

Submitted an hour ago by [\[redacted\]](#) [ME - 2018](#)

[comment](#) [share](#) [save](#) [hide](#) [report](#) [crosspost](#)

↑
7
↓



Question for cyclists on campus [self.gatech](#)

Submitted 3 hours ago by [\[redacted\]](#)

[4 comments](#) [share](#) [save](#) [hide](#) [report](#) [crosspost](#)

↑
↓



Do you think you would have missed out going to your state school instead of GT? [self.gatech](#)

Submitted 37 minutes ago [\[redacted\]](#)

[1 comment](#) [share](#) [save](#) [hide](#) [report](#) [crosspost](#)

↑
0
↓



Changing Threads in CS [self.gatech](#)

Submitted 3 hours ago by [\[redacted\]](#)

[2 comments](#) [share](#) [save](#) [hide](#) [report](#) [crosspost](#)

Online College Communities on Reddit

Submit new content



gatech

14,196 readers, 272 here

[\(edit flair\)](#)

Hey there Techies, join the rally for procrastination. Long live reddit.

Useful Links

- [GT FAQ](#)
- [Official GT iOS App](#)
- [GT IRC](#)
- [GT Buses for iOS for Android](#)
- [GT Homepage](#)
- [GT Protips](#)
- [/r/Atlanta](#)
- [/r/ACC](#)
- [/r/RamblinWreck for GT Sports](#)

Subreddit Rules:

- Don't be a jerk.
- Don't troll or flame
- In order to avoid redundant posts: 7

DATA

- 174 U.S. college subreddits
- 5,884,905 comments posted by 453,781 unique users
- 4,144,161 comments posted by 425,410 unique users who never cross-post across college subreddits

OPERATIONALIZING ONLINE HATEFUL SPEECH

Hateful speech lexicon⁺: 157 keywords in hate categories:

- Behavior
- Class
- Disability
- Ethnicity
- Gender
- Physical
- Race
- Religion
- Sexual Orientation
- Other

Gender: *If you disagree with us, then you are an anti-consumerist c*nt. I told you I'd call you a c*nt twice.*

Racial: *Damn n*ggah youze is just a little dude with a litte ole baby d*ck.*

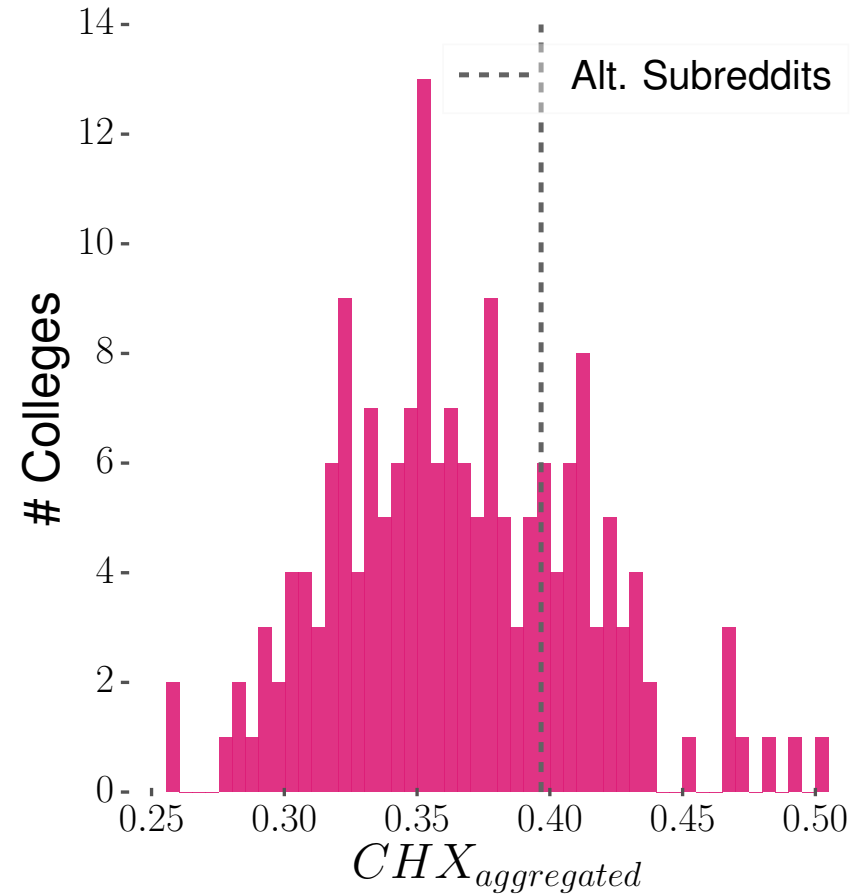
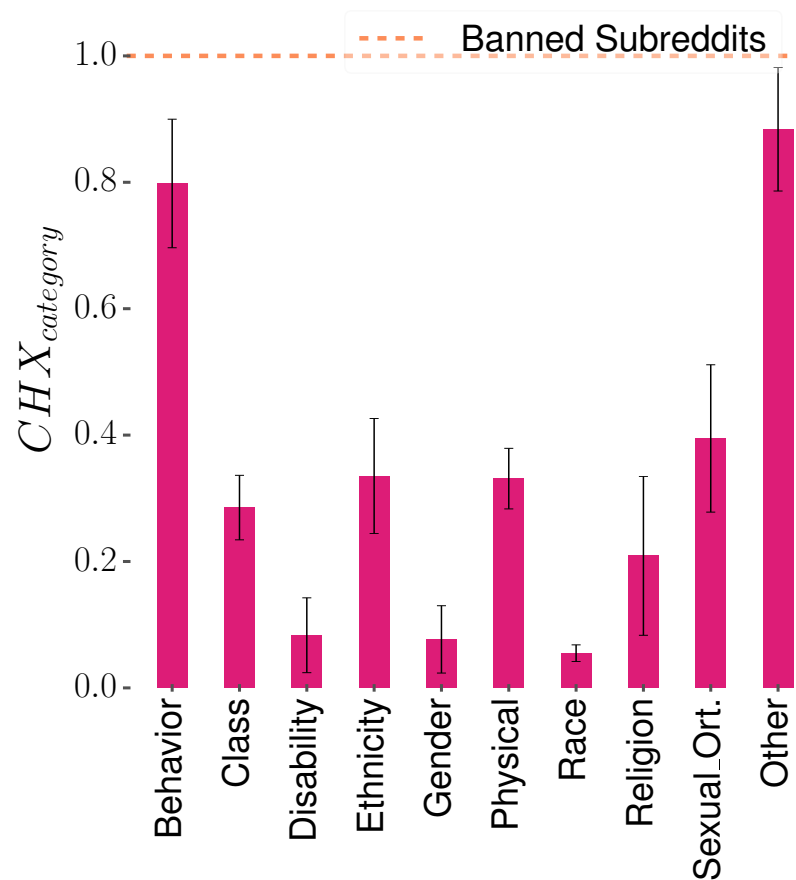
Sexual Ort.: *BIG SHOT, U WANNA FIGHT? U WANNA BOX F*GGOT?*

⁺ Mondal, M., Silva, L. A., & Benevenuto, F. (2017, July). A measurement study of hate speech in social media. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media* (pp. 85-94). ACM.

PREVALENCE OF HATEFUL SPEECH

College Hate Index (CHX)

- Collect 1,436,766 comments from five banned subreddits (r/CoonTown, r/fatpeoplehate, r/KikeTown, r/nazi, r/transf*gs)
- CHX of an online college community to be the ratio of the normalized hate measure in the college subreddit to the same measure in banned subreddits



They admit gifted students with bright futures but produce uneducated hobos who can't get a job and rely on State alumni for welfare.” [OTHER]

PSYCHOLOGICAL EFFECTS OF HATEFUL SPEECH

Whether and how **hateful speech affects** the **psychological state** of the online community members?

Hate exposure per *user* as an aggregated percentage of hateful words used by others on all the threads the *user* has participated in

Online stress expression, as measured using **transfer learning** classification on the *user's* posts


ONLINE STRESS EXPRESSION

Transfer learning methodology to classify social media (Reddit) posts with contains **expression of stress** vs. **no expression of stress** ⁺

- + *r/stress*
- Non-Mental Health subreddits (*r/movies*, *r/music*, *r/aww*, *r/Jokes*, etc.)

“Feel like I am burning out (again...) Help: what do I do?”

COMMUNITY DETAILS

 r/Stress

4.0k

Subscribers

3

Online

r/stress is a subreddit dedicated to discussion of the causes of stress and stress management techniques

⁺ Saha, K., & De Choudhury, M. (2017). Modeling stress with social media around incidents of gun violence on college campuses. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 92.

ONLINE STRESS EXPRESSION: CLASSIFICATION

- 5000 n -gram & 3 sentiment Features
- SVM Classifier
- Accuracy: 0.82 k -fold cross-validation ($k=5$)
- Machine labeled stress in college dataset
- Validated on college subreddit dataset: 82% accuracy

⁺ Saha, K., & De Choudhury, M. (2017). Modeling stress with social media around incidents of gun violence on college campuses. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 92.

STUDY DESIGN

- Conduct **Observational Study**
- Adopt a **Causal Inference Framework** based on **Matching**
- Use **propensity score** analysis
- Compare the **outcomes** (stress expression) of **similar (matched) individuals**, those exposed to a **treatment** (exposure to hate), and those who were not.

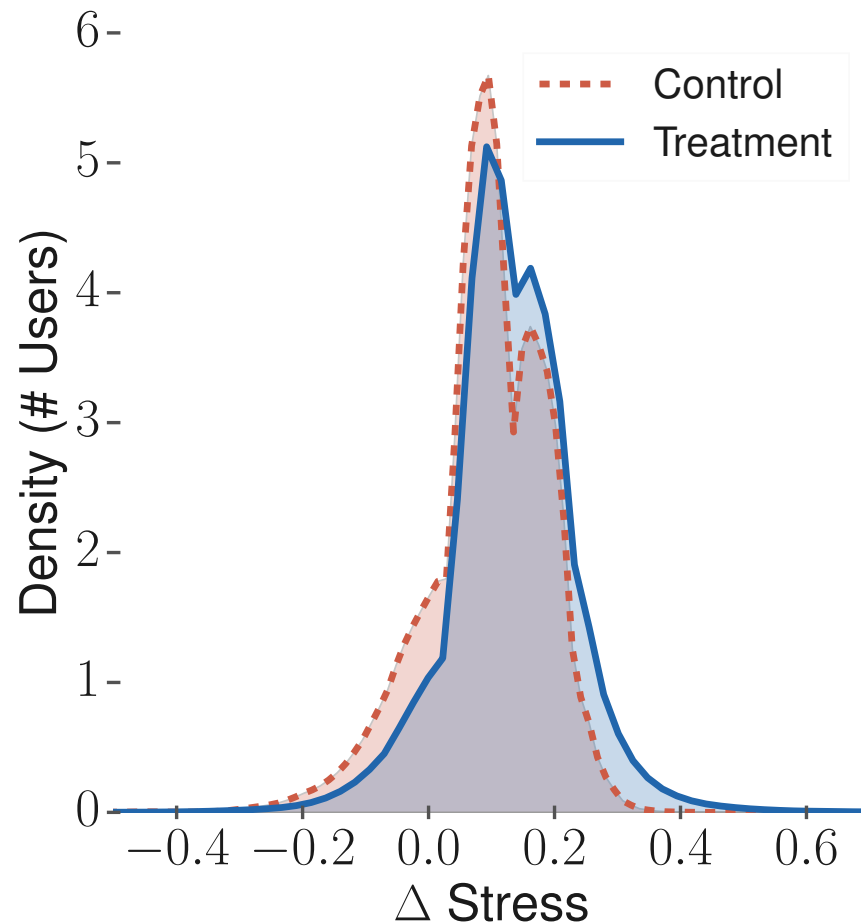
MATCHING FOR CAUSAL INFERENCE

Covariates:

- Social Attributes (#posts, #karma, duration in the community, frequency of posting)
- Offline and local factors (same college subreddit)
- Language and psycholinguistics (Normalized psycholinguistic occurrence of LIWC categories)
- Baseline exposure to hate and baseline stress expression

Propensity score matching for 143K Treatment and 74K Control users

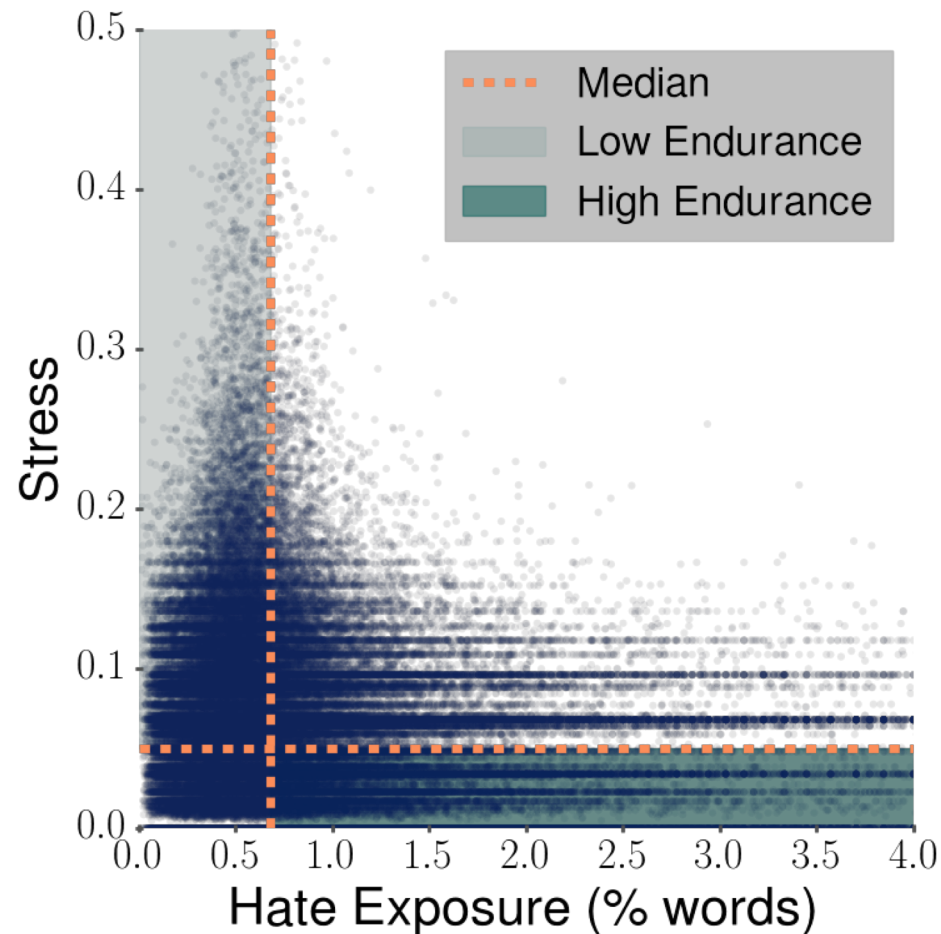
DOES HATE IMPACT STRESS LEVEL?



Stress change of the Tr. users (~139%) is higher than the Ct. users (~106%). Effect size (Cohen's $d=0.40$) and paired t -test shows **statistically significant difference**

Individuals **exposed** to hate speech show **greater likelihood of stress** compared to their matched controls

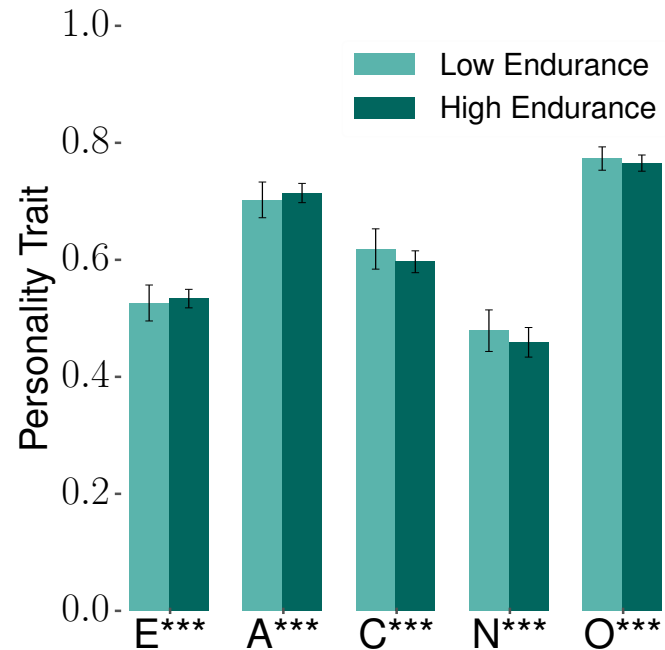
PSYCHOLOGICAL ENDURANCE TO HATE



Not all Treatment users are affected equally: *differences in psychological endurance*

Analyze high and low endurance users' personality traits, and language (SAGE)

LINGUISTIC MARKERS OF PSYCHOLOGICAL ENDURANCE



Personality and Linguistic Differences (SAGE)

Low Endurance		High Endurance	
<i>n</i> -gram	SAGE	<i>n</i> -gram	SAGE
education code	1.99	guitar	-1.19
prerequisite course	1.68	discord	-1.15
general education	1.64	baylor	-1.15
classes mentioned	1.26	esports	-1.09
mathematics	1.07	smash	-1.00
credit course	1.03	bath	-0.98
upper division	0.95	bands	-0.97
prerequisite	0.95	tournament	-0.93
preassessment	0.89	phi	-0.92
therapy	0.88	garden	-0.90
senate	0.85	shots	-0.89
task	0.84	pokemon	-0.89
cse	0.83	delicious	-0.86
immigrants	0.82	temple	-0.85
prereq	0.81	wings	-0.84
cs1	0.79	used work	-0.84
mathematical	0.78	players	-0.83
irrelevant	0.78	song	-0.83
mentor	0.78	jazz	-0.80
tasks	0.76	anime	-0.80
division	0.75	basement	-0.79
comptia	0.75	rock	-0.79
assessment	0.73	student section	-0.78
anxiety	0.73	yo	-0.78
sql	0.73	sublease	-0.77

- High endurance individuals show greater agreeableness and extraversion, whereas Low endurance individuals show greater Neuroticism and Openness
- High endurance individuals show relaxed-, leisure-, and non-academic interests
- Low endurance individuals use more classroom- and academic- related topics

DISCUSSION

TAKEAWAYS

- Leverage social media data to understand the prevalence of hateful speech in online college communities
- Examine the psychological effects of hateful speech in the online community members
- Analyze the differences in personality traits and language of individuals differing in psychological endurance to hate

SOCIO-POLITICAL AND POLICY IMPLICATIONS

- Findings help to account for a previously under-explored, but a critical facet of the speech debate, especially in the context of college campuses
- New dimensions to the college speech debate centering around legal, ownership, and governance issues
- New policy challenge: How to decipher when online and offline hate reinforce each other.
- Opens up discussions about “counterspeech” to tackle the psychological effects of hateful speech

TECHNOLOGICAL IMPLICATIONS

- Mental health support provisions on college campuses:
 - Empirical information to guide informed decisions and tailored interventions
 - Proactive detection of the psychological ramifications of online hate
- Moderation efforts in online college communities
 - Call-to-action to moderators to adopt measures beyond blanket banning/censorship
 - Provide assistance and support via peer-matching and redirection to resources

ETHICAL IMPLICATIONS

- Any **decision** and **policy** making based on our findings requires **careful** and **in-depth supplemental ethical analysis**
- **Caution** against **surveillance of student speech** on college campuses, or as a guideline to **ensor speech** on campus.
- **Not intended** to be used to inadvertently **marginalize** or influence prejudice against those groups who are already marginalized

Saha, K., Chandrasekharan, E., & De Choudhury, M. (2019). Prevalence and Psychological Effects of Hateful Speech in Online College Communities. In *Proceedings of the 11th ACM Conference on Web Science.*,
<https://dl.acm.org/citation.cfm?id=3326032>

Prevalence and Psychological Effects of Hateful Speech in Online College Communities



Thank You

@kous2v | koustuv.saha@gatech.edu | koustuv.com