

# CS 4873: Computing, Society & Professionalism

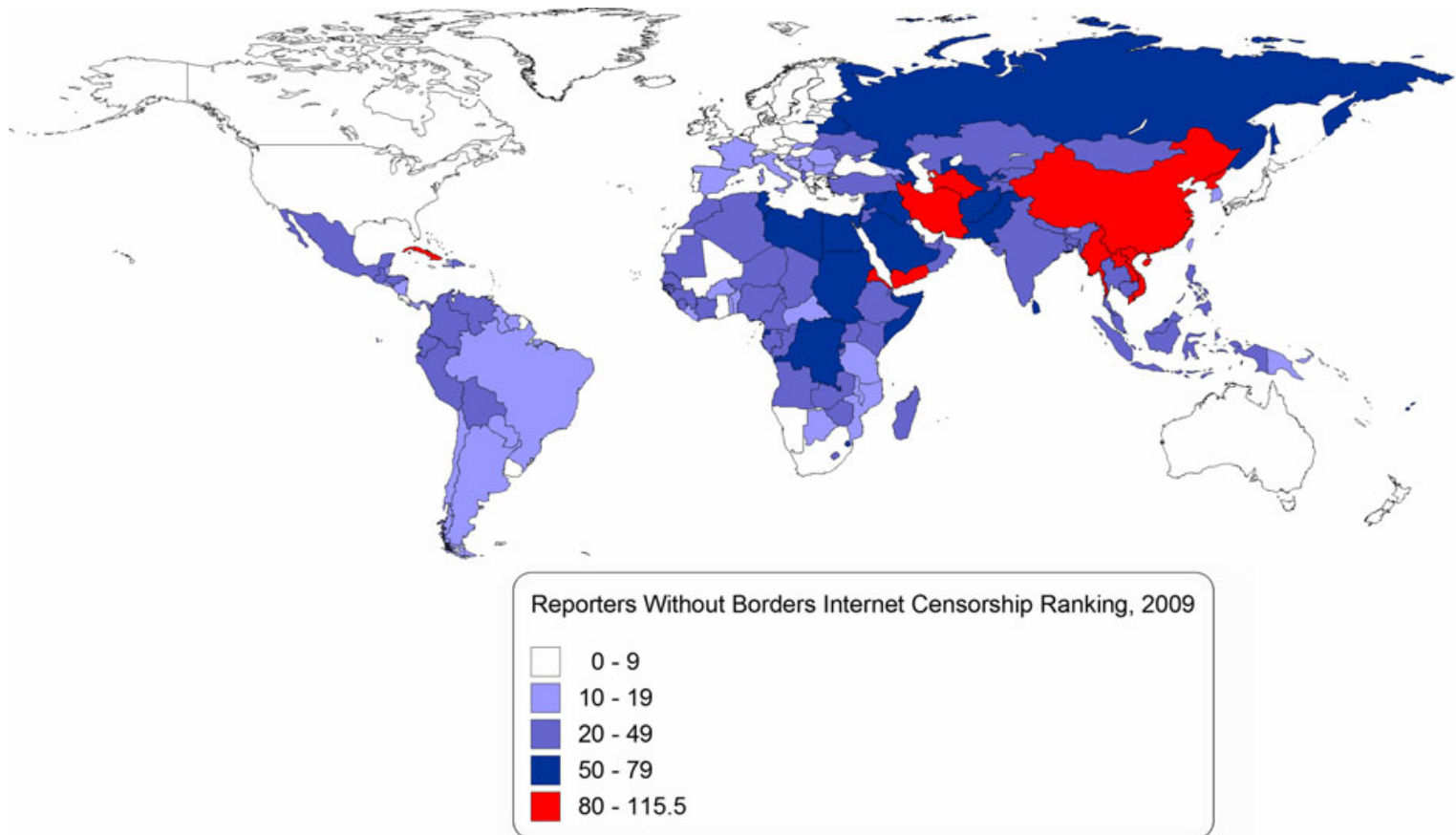
Munmun De Choudhury | Assistant Professor | School of Interactive Computing

Week 10: Regulating Online Speech  
March 11, 2020



# Is Censorship Ethical?

# \* Censorship and Internet (Warf 2010)



**Fig. 2** Reporters Without Borders Internet Censorship Ranking 2009. *Source:* data drawn from <http://www.rsf.org/en-classes/ment1003-2009.html>



# Freedom of Expression

# \* First Amendment


- It constrains what the US government can do
  - Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the government for a redress of grievances
- The primary purpose is political – allow an open discussion of public issues; allow minority perspectives to be represented
  - But extends to other domains
- Not an absolute right
  - The private right must be balance against the public good

# First Amendment

- Private groups can censor all they want and in some cases may seem to violate the first amendment
  - E.g., Facebook's Terms of Service disallow the sharing of explicit or violent content
  - In general written with "offensive"/controversial speech in mind
- Covers spoken & written words, pictures, art, and other forms of expression of ideas (like wearing an arm band)

# First Amendment

- Existing restrictions on speech – **balancing private right and public good**
  - Advocating illegal acts can be acceptable (e.g., civil disobedience)
  - Libel (making false and damaging statements) is illegal
  - Making specific threats is illegal
  - Inciting violence can be illegal
  - Perjury
  - Obscenity
  - False advertising




*Discussion point 1:* In the US, television commercials for cigarettes are banned. Should there be a ban on commercials for violent video games too? (Describe in the light of “private right versus public good”)





# Platform enforced content regulations



# Online Hate/Abusive Speech

# Free Speech and Hate Speech

- Social media and the Internet have opened up for many new arenas for exchanging opinions.
- Freedom of speech is an absolute value in any democracy, both for the public and for the media.
- At the same time, opinions and debates challenge us as hate speech are spread widely and frequently on new platforms for publishing.
- Hate speech may cause fear and can be the reason why people withdraw from the public debate
  - Women and minority groups most affected

# Reddit Banned FatPeopleHate

- Reddit introduced a new anti-harassment policy
- Reddit banned a handful of its worst communities, saying: "Our goal is to enable as many people as possible to have authentic conversations and share ideas and content on an open platform"
- Prominent is the removal of r/fatpeoplehate, a forum devoted to posting (among other things) pictures of overweight people for ridicule.
- They said "They are banning behaviors or ideas"
  - Banning policy was based on specific instances of harassment rather than general offensiveness

# Reddit Banned FatPeopleHate

- Deciding which subreddits engage in harassment is going to be a difficult problem, especially because it's hard to even agree on what constitutes harassment — and that assumes everyone involved is acting in good faith, which almost certainly won't be the case.

Without Steve Wynn, Casino Empire Risks Losing More Than a Name

Twitter Reports First Quarterly Profit, Despite User Stagnation

The Stock Market Is Worried About Inflation. Should It Be?

PepsiCo Dips Its Toes Into the Sparkling Water Market

Rising Hummus Prices? Blame a Drought Half a World Away

PAID POST: DESTINATION CANADA Why Quebec City Is North America's Culture Capital

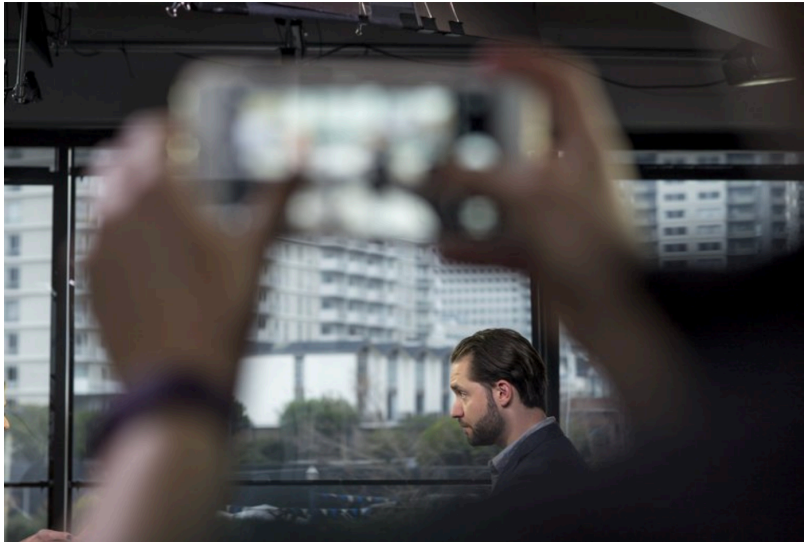
CANADA KEEP EXPLORING

STATE A Cr Loca for It

BUSINESS DAY

# Reddit Limits Noxious Content by Giving Trolls Fewer Places to Gather

The Shift  
By KEVIN ROOSE SEPT. 25, 2017



Alexis Ohanian, a founder of the website, which banned several forums in 2015 as part of a broad crackdown on poisonous behavior. David Paul Morris/Bloomberg

RELATED COVERAGE



THE SHIFT  
This Was the Alt-Right's Favorite Chat App. Then Came Charlottesville. AUG. 15, 2017



Facebook and Twitter Could Face Fines in Germany Over Hate Speech Posts  
MARCH 14, 2017

There are — and always have been, and probably always will be — trolls, scoundrels and reprobates on the internet.

It is a problem that has vexed multibillion-dollar corporations and the smartest computer programmers in the world. Facebook, Twitter and YouTube have all declared war on abuse and harassment, spent years training sophisticated algorithms and hired vast armies of moderators to root out hateful content.

And yet, the trolls persist.

But what if a better way of combating online toxicity were right under our

- 1 44 ↑ Any women have experience with having a baby while in grad school? (self.GradSchool)  
submitted 12 hours ago by HigHog  
29 comments share
- 2 10 ↑ Hitler doesn't get a postdoc (youtube.com)  
submitted 5 hours ago by Epistaxis (PhD, genetics)  
5 comments share
- 3 6 ↑ I think it's time to quit (self.GradSchool)  
submitted 5 hours ago \* by Amelorn  
11 comments share
- 4 ↑ Advice: Masters or go straight for PhD (self.GradSchool)  
submitted 12 minutes ago by Cagedcrab  
comment share
- 5 ↑ For Profit or Not University (self.GradSchool)  
submitted 19 minutes ago by lilfermat  
comment share
- 6 ↑ First time writing an state of purpose !! pls some help (self.GradSchool)  
submitted 25 minutes ago by nautiluz92  
comment share
- 7 ↑ Does it make sense to take a mortgage (if I have the downpayment) than rent out a studio for my 5 year PhD program? (self.GradSchool)

MY SUBREDDITS FRONT - ALL - RANDOM | ASKREDDIT - FUNNY - PICS - WORLDNEWS - GIFS - VIDEOS - TODAYILEARNED - NEWS - MOVIES - CREEPY - GAMING - AWW - SHOWERTHOUGHTS - IAMA - MILDLYINTERESTING



From the SW Mods | [If you see abuse, trolling, or guideline violations, click here to message us!](#)

- 156 ↑ Update on PM trolls, self-appointed enforcers, and SW Moderation Practices. If you got a PM that was abusive, read, thx! (self.SuicideWatch)  
submitted 11 months ago \* by SQLwitch - - stickied post  
39 comments share
- 45 ↑ Something New - an automod message to helpers (don't panic!) (self.SuicideWatch)  
submitted 3 months ago \* by skyqween - [M] - stickied post  
57 comments share
- 1 ↑ No one will want to read the note, so I'm leaving it here (self.SuicideWatch)  
submitted 2 hours ago \* by spacedoughnut  
5 comments share
- 2 ↑ I need some help here (self.SuicideWatch)  
submitted an hour ago \* by imightneedhelphere  
2 comments share
- 3 ↑ Tired of the pain (self.SuicideWatch)



# Pro Self-Harm Content



TECH

# Why Eating Disorders Are So Hard For Instagram And Tumblr To Combat

Over the last four years, the social media platforms have done a lot to curb content that promotes self-injury. But they'll never fully succeed. Is it worth trying?

Posted on April 14, 2016, at 2:01 p.m.



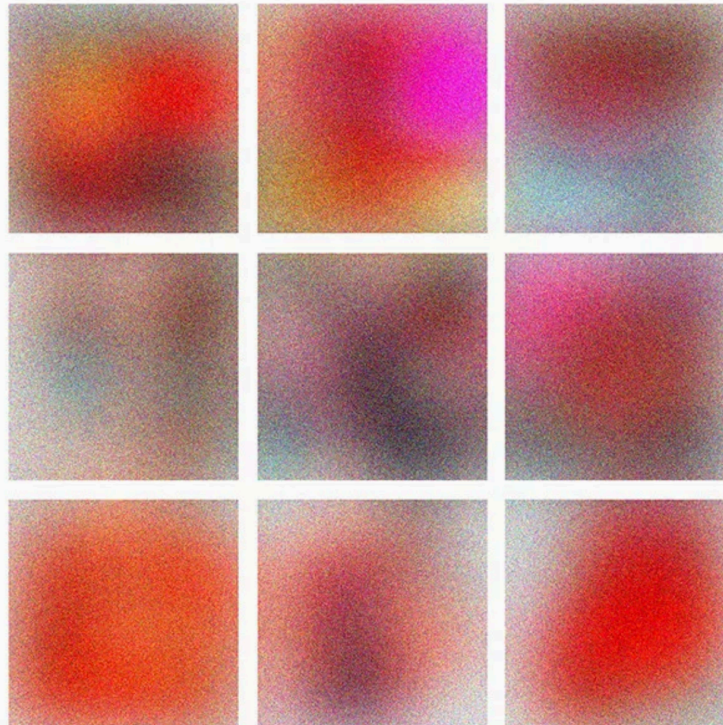
**Stephanie M. Lee**  
BuzzFeed News Reporter



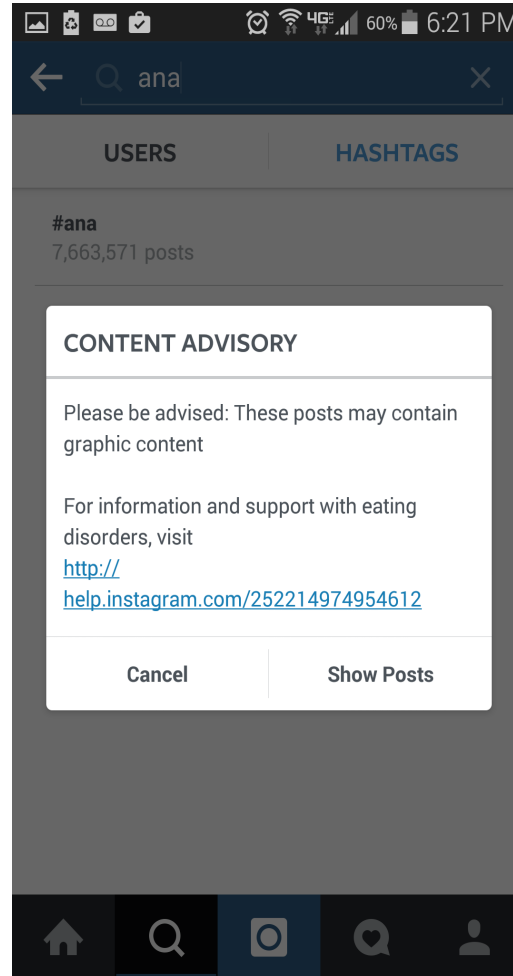
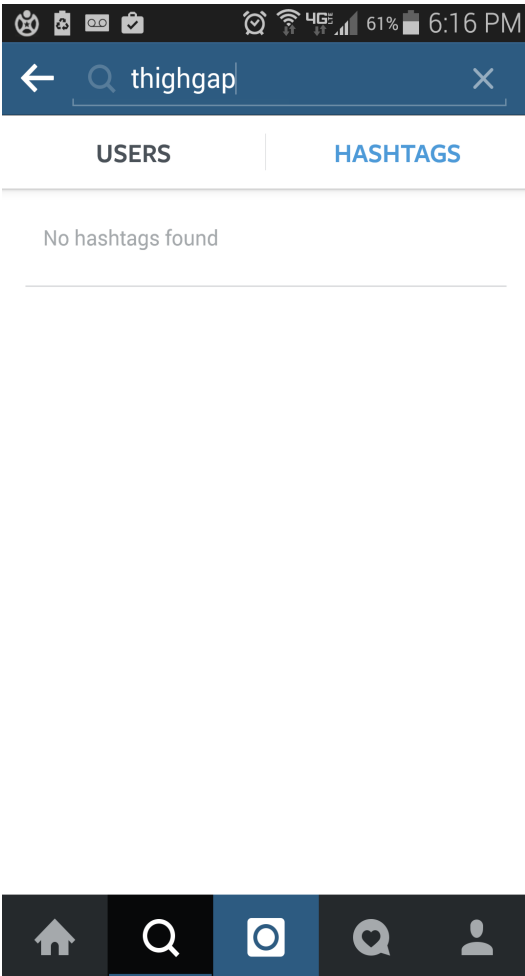
#anorexia

5,170,983 posts

TOP POSTS



# Interventions



## Everything okay?

If you or someone you know is struggling with thoughts of suicide, the Lifeline is here to help: call 1-800-273-8255

If you are experiencing any other type of crisis, consider chatting confidentially with a volunteer trained in crisis intervention at [www.imalive.org](http://www.imalive.org), or anonymously with a trained active listener from 7 Cups of Tea.

And, if you could use some inspiration and comfort in your dashboard, you should consider following the Lifeline on Tumblr.

[Go back](#)

[View search results](#)

# *But* deviant behavior subverts attempts to intervene

anorexic, anorexie, anoressia, anorexi, anorexia, anorexique, anorexica, anorectic, anorexia, anorectic

eatingdisorders, eatingdissorder, eatingdisoder, eatingdis, eatingdisorter, eatingdisoreder, eatingdisorde,  
eatingdisorderrr, eatingdisordered, eating\_disorder

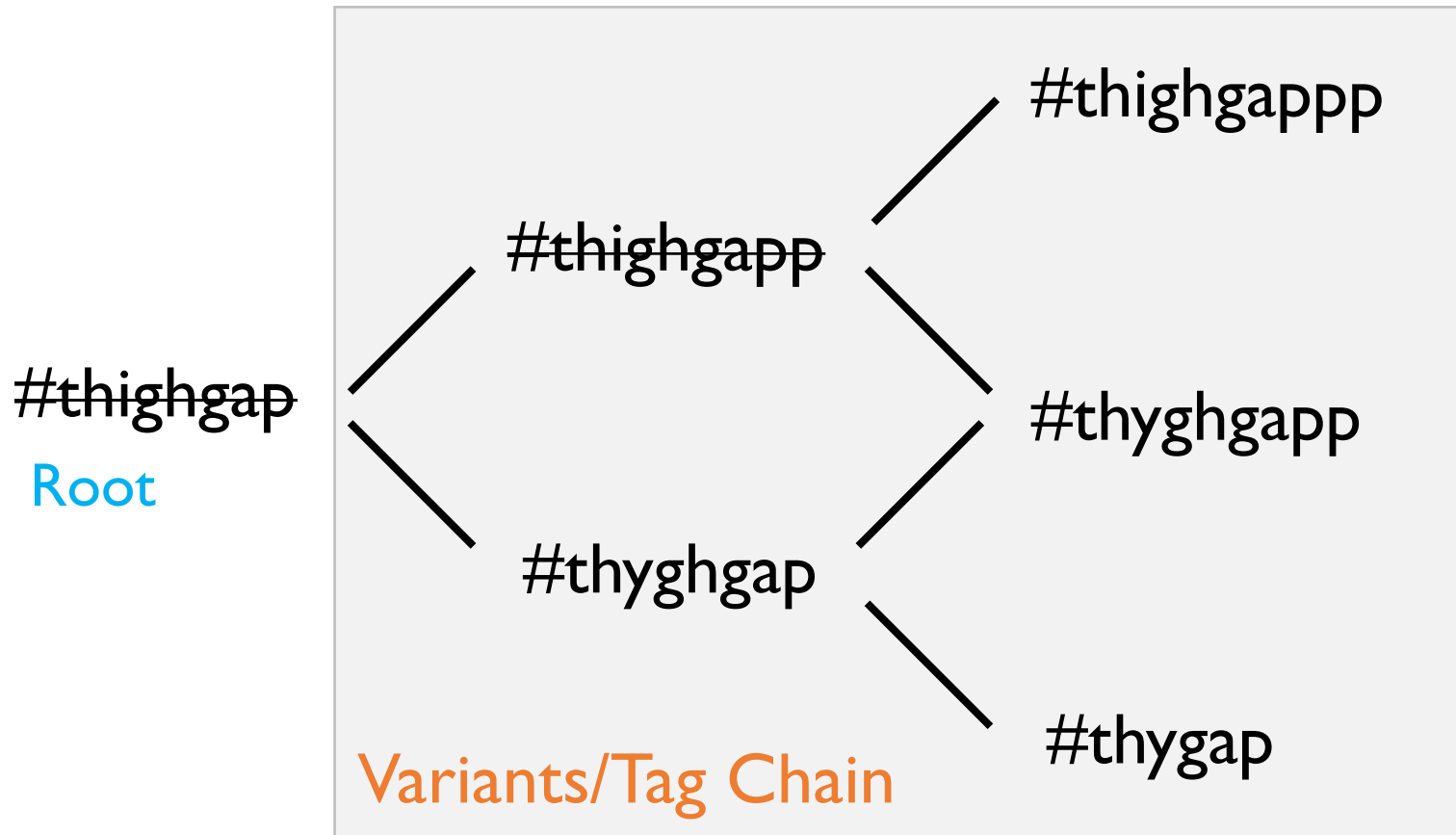
thighgaps, thygap, thighgapp, thigh\_gap, thightgap, thyghgap, thighgappp, thegap, thigap, thighgapss

thinspoooo, thynspo, thynspoo, thynspooo, thinspoo, thinspooo, thynspoooo, thinnsपो, thinspoooooo

Increasing and more complex lexical variations have emerged since Instagram enforced moderation of pro eating disorder content in 2012

Automatically detect lexically variant tags that emerged out of moderated tags using edit distance computation + regular expression matching + semantic annotations via crowdsourcing (Amazon's mechanical turk)

“thighgap” > “th\*g\*p\*”



# *But* deviant behavior subverts attempts to intervene

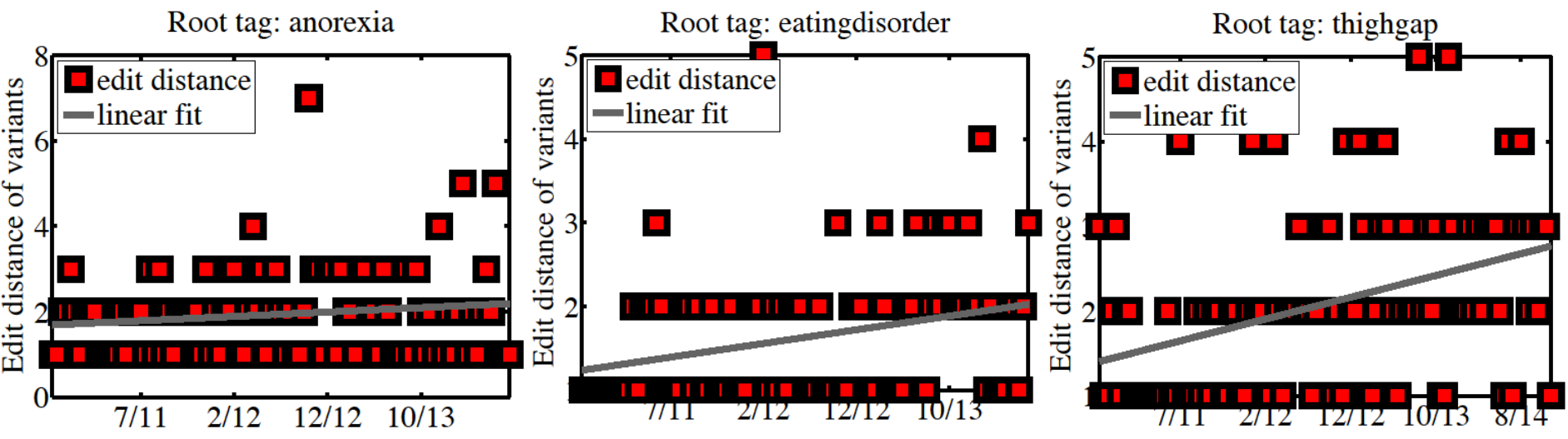
anorexic, anorexie, anoressia, anorexi, anorexia, anorexique, anorexica, anorectic, anorexia, anorectic

eatingdisorders, eatingdissorder, eatingdisoder, eatingdis, eatingdisorter, eatingdisoreder, eatingdisorde, eatingdisorderrr, eatingdisordered, eating\_disorder

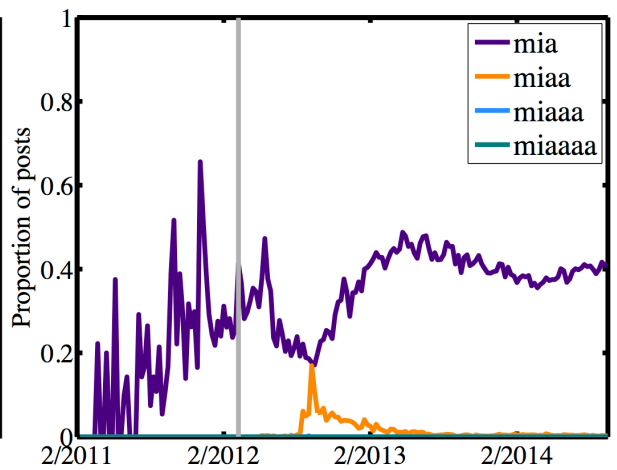
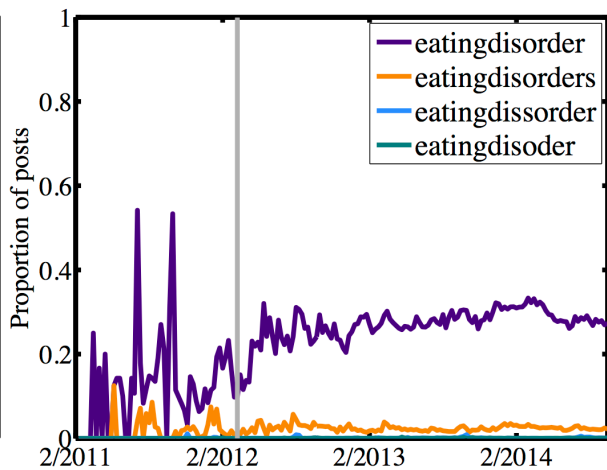
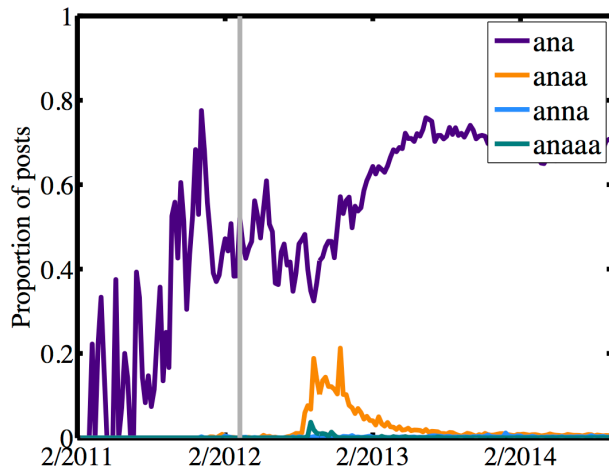
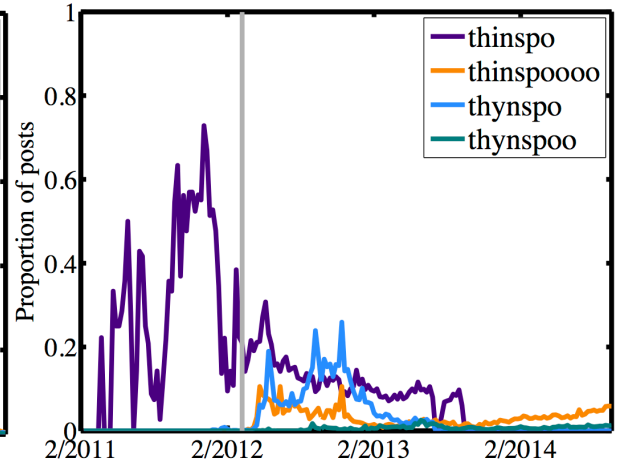
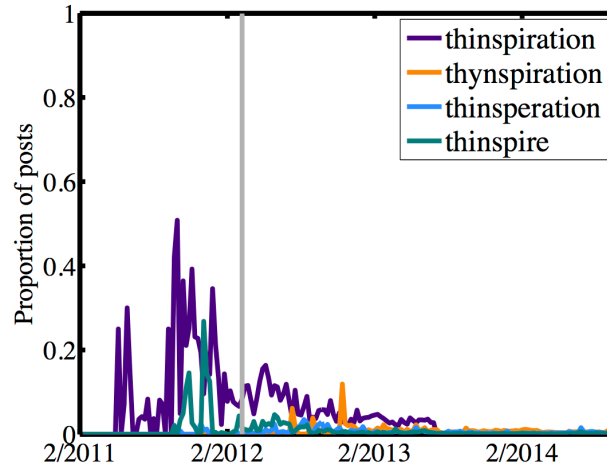
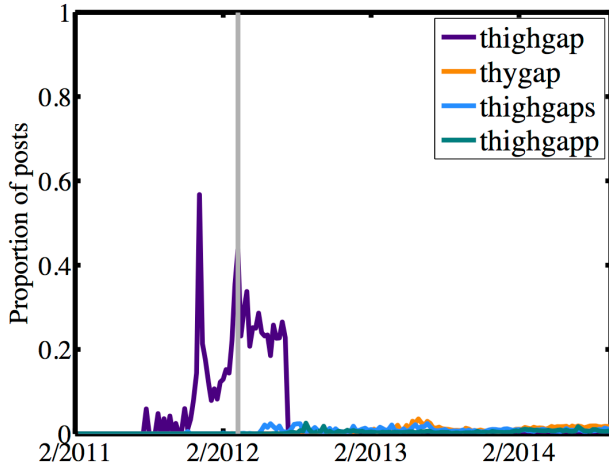
thighgaps, thygap, thighgapp, thigh\_gap, thightgap, thyhgap, thighgapppp, thegap, thigap, thighgapss

thinspoooo, thynspo, thynspoo, thynspooo, thinspoo, thinspooo, thynspoooo, thinnsपो, thinspoooooo

Increasing and more complex lexical variations have emerged since Instagram enforced moderation of pro eating disorder content in 2012



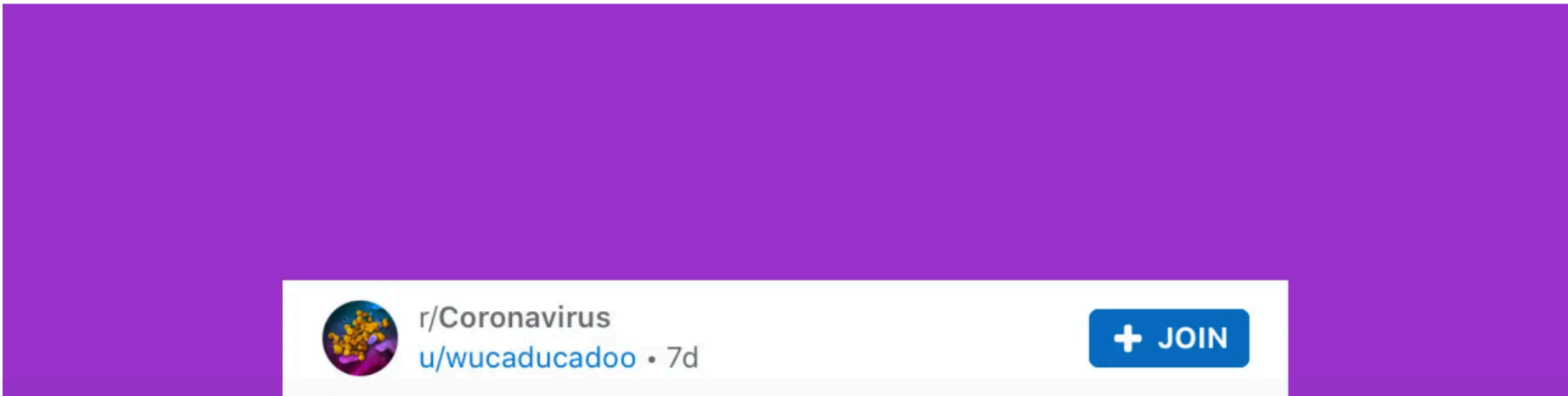
# Moderation was followed by increased social engagement





# Reddit ran wild with Boston bombing conspiracy theories in 2013 and is now an epicenter for coronavirus misinformation. The site is doing almost nothing to change that.

Paige Leskin Mar 6, 2020, 9:14 AM



r/Coronavirus  
u/wucaducadoo • 7d

+ JOIN



Spam




# \* Spam

- What is spam?
- With ease of internet access, businesses looked for ways to capitalize on market opportunities associated with Internet communications – easier/cheaper to send emails than physical mails
  - How to find email addresses though?
  - Crawling the web; scrape address books with viruses; listen to chatroom conversations; sneaky way to sign up; dictionary attacks on ISPs
- This entrepreneurial behavior has given rise to a new set of legal and ethical problems

# Spam

- Cyber Promotions versus American Online
- The scenario:
  - AOL blocks Cyber Promotions email
  - Cyber promotions changes “from” address
  - Cyber promotions sues AOL for violating first amendment rights
    - Loses
  - AOL defends self by claiming violation of property rights; carrying spam is a cost to AOL



*Discussion point 2:* Why is “cold calling” considered to be an acceptable sales practice, but spamming isn’t?