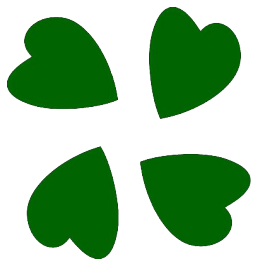


# Radicalization of 4chan throughout the years

Group 4

Benedikt Geitner  
Vincent Halasz  
Irene Lopez Gutierrez



**4chan**

What is it?

- Imageboard
- Launched on 2003
- Originally about anime
- Organized into different boards
- Anonymous, largely unmoderated & ephemeral
- Origin of many controversies

# How has radical content increased throughout the years?

- Number of posts
- Two subcategories:
  - Hate speech
  - Conspiracy theories (in particular: antivax conspiracies)
- Identify patterns, e.g. correlation with socioeconomic factors?

# Related work

- Research on 4chan is relatively sparse.

Colley T, Moore M. *The challenges of studying 4chan and the Alt-Right: 'Come on in the water's fine.'* New Media & Society. September 2020.

[doi:10.1177/1461444820948803](https://doi.org/10.1177/1461444820948803)

- Challenges:
  1. Fast evolving language (1)
  2. Users' anonymity
  3. Ethical concerns of researchers (2)

<sup>1</sup>Ludemann D. */pol/emics: Ambiguity, scales, and digital discourse on 4chan*. Discourse, Context & Media. 2018. 24: 92-98. <https://doi.org/10.1016/j.dcm.2018.01.010>

<sup>2</sup>Phillips W. *The House That Fox Built: Anonymous, Spectacle, and Cycles of Amplification*. Television & New Media. 2013;14(6):494-509. doi:10.1177/1527476412452799

# Related work

## First measurement study

Hine G., et al. *Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan's Politically Incorrect Forum and Its Effects on the Web*. Proceedings of the Eleventh International AAAI Conference on Web and Social Media (2017)  
<https://arxiv.org/abs/1610.03452>

- Limited to /pol/.
- 2 months of data
- Conclusion: hate speech is predominant

## Largest public dataset

Papasavva A., et al. *Raiders of the Lost Kek: 3.5 Years of Augmented 4chan Posts from the Politically Incorrect Board*. Proceedings of the Fourteenth International AAAI Conference on Web and Social Media (2020).  
<https://arxiv.org/abs/2001.07487>

- Limited to /pol/
- Dataset augmented with labels about toxicity of posts
- Plotted statistics about activity in the forum.

# Data: what we have

We collect title, text, timestamp and replies.

Board	# threads in archive	# threads we collected	# posts collected (including replies)	Archive since	Board exists since
/pol/	> 8,000,000	3076	110358	Nov 2013	Oct 2013
/adv/	~ 900,000	2788	35234	Jan 2014	Jan 2010
/sci/	~ 600,000	2980	62026	May 2010	Jan 2010
/news/	~ 30,000	3156	97055	Nov 2015	Nov 2015

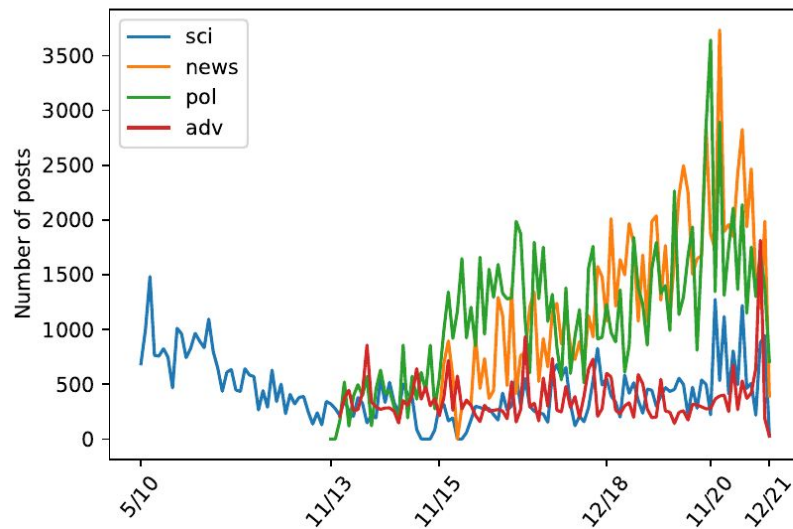
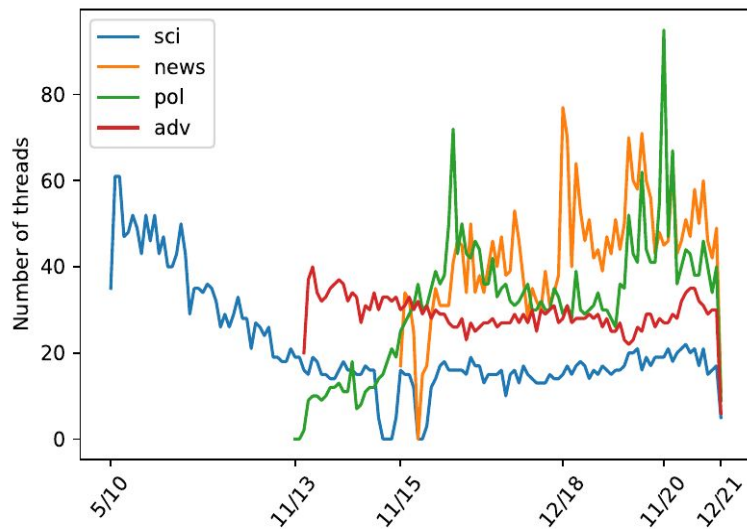
# Data: what we have

We collect title, text, timestamp and replies.

Board	# threads in archive	# threads we collected	# posts collected (including replies)	Archive since	Board exists since
/pol/	> 8,000,000	3076	110358	Nov 2013	Oct 2013
/adv/	~ 900,000	21569	293262	Jan 2014	Jan 2010
/sci/	~ 600,000	20860	404025	May 2010	Jan 2010
/news/	~ 30,000	3156	97055	Nov 2015	Nov 2015

# Results

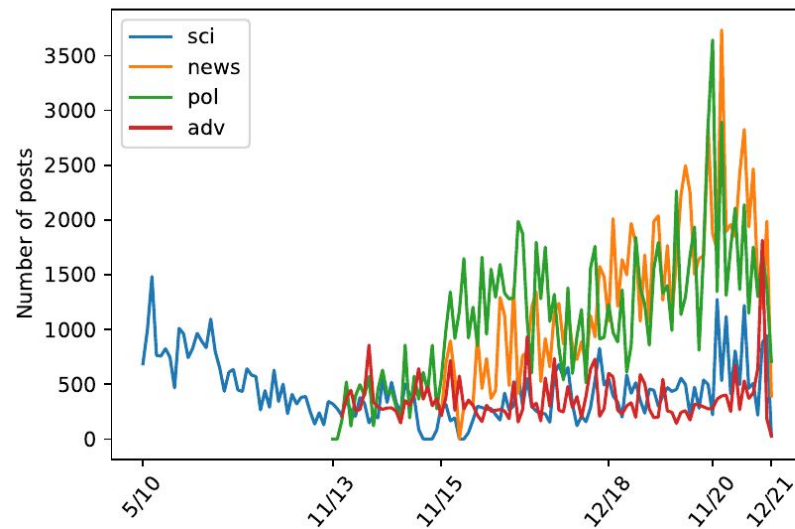
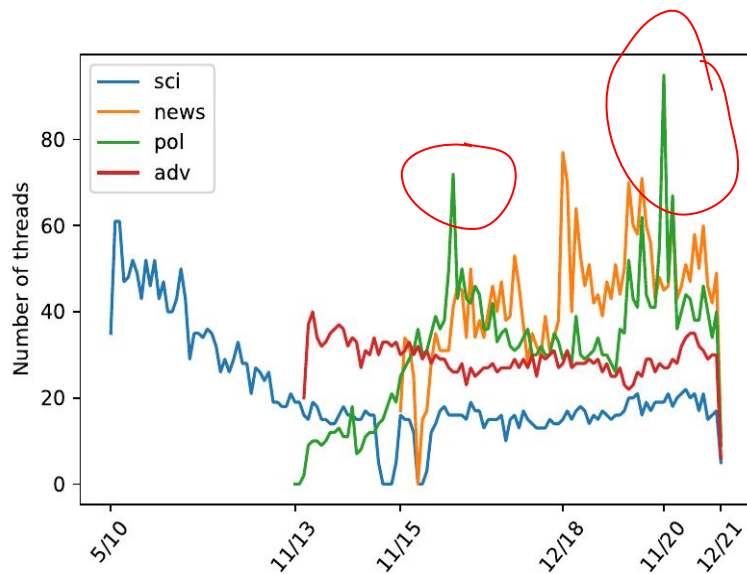
Plotting posts as a function of time





# Results

Plotting posts as a function of time



# Methods

To flag hate speech:

- Vocabulary database from HATEBASE
- Classify as related to: ethnicity, religion, gender, sexuality, etc.



For anti-vax sentiment:

- Flag posts with variations of the word vaccine.
- Sentiment analysis to classify into negative or positive

Regression analysis to find patterns:

- E.g. compare with US GDP, unemployment...

# Thank you!

Group 4

Benedikt Geitner

Vincent Halasz

Irene Lopez Gutierrez

Check out our git repo:

<https://github.com/irenelopez42/4chanAnalysis>