Social media phenomena: News &

Misinformation

+ collect Reddit Data

IMT 547 - Social Media Data Mining and Analysis

Today's Topics

Twitter data collection (left-over)

Paper discussion

Reddit data collection

Leftover Twitter lab

Paper

This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News

Benjamin D. Horne and Sibel Adalı

Rensselaer Polytechnic Institute 110 8th Street, Troy, New York, USA {horneb, adalis}@rpi.edu

SUMMARY

Key Question:

• Is there any systematic stylistic & other content differences between fake & real news?

Data Sets - 3 datasets, 3 categories (fake, real, satire)

Data Analysis: feature construction + classifier

Results:

- The content of fake and real news articles are substantially different.
- Titles are a strong differentiating factor between fake and real news
- Fake content is more closely related to satire than to real.
- · Real news persuades though arguments, while fake news persuades through heuristics.

Characterizing the Social Media News Sphere through User Co-Sharing Practices

Mattia Samory*

Computational Social Science GESIS, Cologne, Germany mattia.samory@gesis.org

Vartan Kesiz Abnousi

Department of Computer Science Virginia Tech, VA, USA vkesizab@vt.edu

Tanushree Mitra

Department of Computer Science Virginia Tech, VA, USA tmitra@vt.edu

What **communities of news sources** emerge, when considering the sharing practices of their social media audience?

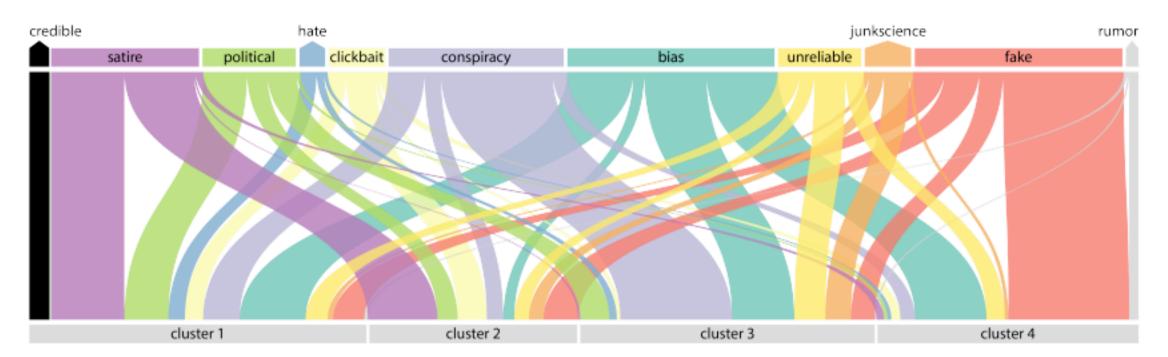


Figure 1: Cluster composition in terms of expert assessments of credibility. We map the primary category of opensources.co for each source to the corresponding cluster. Cluster 1 is the destination for all of the credible sources, as well as for political misinformation sources; cluster 2 is home to satirical and clickbait sources; cluster 3 hosts most conspiratorial, unreliable, and junk science sources; cluster 4 is largely comprised of fake news and biased sources.

widespread	satirical/clickbait	conspiratorial	right-wing/fake
washingtonexaminer.com	usuncut.com	activistpost.com	angrypatriotmovement.com
thegatewaypundit.com	countercurrentnews.com	21stcenturywire.com	usasupreme.com
wnd.com	occupydemocrats.com	thedailysheeple.com	truthandaction.org
Newsmax.com	attn.com	russia-insider.com	prntly.com
heatst.com	rawstory.com	blacklistednews.com	subjectpolitics.com
dailycaller.com	bipartisanreport.com	washingtonsblog.com	usanewsflash.com
dailywire.com	dailykos.com	veteranstoday.com	christiantimesnewspaper.com
americanthinker.com	addictinginfo.org	investmentwatchblog.com	ilovemyfreedom.org
conservativereview.com	politicususa.com	ronpaulinstitute.org	departed.co
wsj.com	alternet.org	wearechange.org	supremepatriot.com

Table 1: Top 10 central news sources in each cluster—i.e., sources that are the most cosine-similar to each KMeans centroid.

sentiment		clickbait		conspir.		fake	
sentiment	intercept	-2.1	*	-1.36	*	-2.92	*
	compound	-0.04	*	0.00		-0.31	*
	neu	-0.33	*	-0.38	*	-0.39	*
psycholinguistic		clickbait		conspir.		fake	
	intercept	-2.17	*	-1.58	*	-3.01	*
pronouns	we	0.01	*	0.02	*	0.02	*
	they	0.01	*	-0.01	*	0.01	*
	you	0.00		-0.04	*	-0.04	*
	i	-0.01	*	0.02	*	-0.08	*
	shehe	0.03	*	-0.18	*	-0.02	*
	ipron	0.03	*	-0.00		0.08	*
social	friends	0.01	*	-0.10	*	-0.06	*
words	family	0.05	*	-0.10	*	0.04	*
emotion words	anger	0.10	*	0.17	*	0.24	*
	sad	0.02	*	-0.01	*	-0.01	*
	anx	0.04	*	0.00		0.05	*
	posemo	0.06	*	0.31	*	-0.09	*
cognitive mech.	discrep	-0.04	*	-0.06	*	-0.03	*
	cause	0.13	*	0.25	*	-0.02	*
	tentat	0.09	*	0.02	*	0.10	*
	excl	0.01	+	0.02	*	-0.01	*
	certain	0.23	*	0.23	*	0.21	*
	insight	0.19	*	0.32	*	0.11	*

What kind of language and engagement attributes characterize communities of news sources?

Definitions

"What is the definition of the fake news here? Is it completely fake? Or does the news article contain some amount of truth along with false claims? Another concern would be fake news articles from the so called "legitimate sources."

- Vidyashree

"..... that **intentional disinformation** has been around for decades and has evolved immensely since its inception (originally called "active measures" by Soviet agents)(Ridd, 2020)."

- Stephen

...body section of real news articles is longer than that of the fake news articles

- Prathamesh

...nugget stories

- Vishwa



Inshorts - 60 words News summary

Inshorts News & Magazines

★★★★ 498,487 👤

€ Everyone

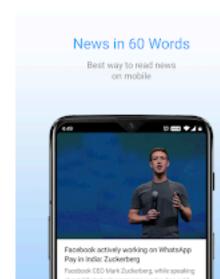
Contains Ads

▲ You don't have any devices

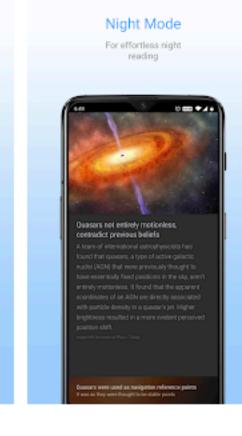
Read Offline

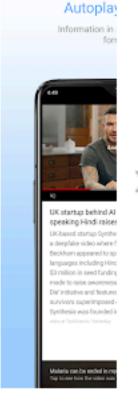
Add to Wishlist

Install









Inshorts is a news app that selects latest and best news from multiple national and international sources and summarises them to present in a short and crisp 60 words or less format, personalized for you, in both, English or Hindi. All summarised stories contain only headlines and facts, no opinions, to help you stay informed of the current affairs. Whether it's the latest government policies or shakeups in bollywood, we get them covered and delivered super fast! Get updated with the latest news and current affairs in a jiffy!

Timing as a factor

"the **timing** when the news is published is also a significant factor to consider. For example, before or just around the election, loads of misinformation campaigns emerged, which were manipulated by different parties with the respective purposes to confuse people on social media."

- Esther Yang

"We can further extend these questions to consider the factor of time: how did these common user behaviors develop over time? (2020 paper)

- Tony Chu"

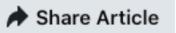
Spread of misinformation

"To inform users of potential falsehood, one solution I can think of could be a plug-in to label the category of each news source, or preferably the assessment results in terms of factuality for each news article shared. The label leverages the work from Samory et al.(2020) and serves as a mini fact-checker to help people be more mindful when they choose to retweet.

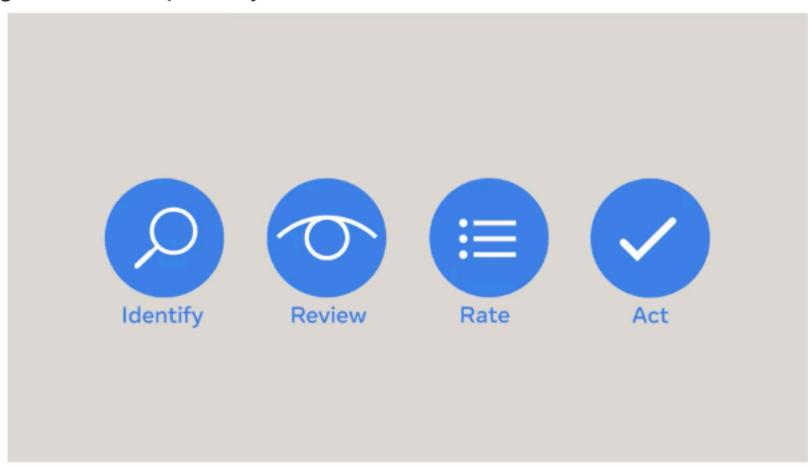
- Yajie"

https://fb.me/thirdpartyfactchecking

How is Facebook addressing false information through independent fact-checkers?



We're committed to fighting the spread of false information on Facebook. We use both technology and human review to remove fake accounts, promote information literacy and disrupt the financial incentives of spammers. In certain countries, we also work with independent fact-checkers who are certified through the non-partisan International Fact-Checking Network to help identify and review false information.



Reducing the Distribution of False Information

- Identifying false information: We identify information that may be false using signs like feedback from people on Facebook. Fact-checkers may also identify stories to review on their own.
- Reviewing stories: Fact-checkers will review stories, check their facts, and rate their accuracy.

https://www.facebook.com/help/1952307158131536

IFCN Fact-checking database

https://www.poynter.org/ifcn-covid-19-misinformation/



January 21, 2021

Search

WS TRAINING ETHICS & LEADERSHIP FACT-CHECKING MEDIA LITERACY EVENTS

The CoronaVirusFacts/DatosCoronaVirus Alliance Database

Here is the database that gathers all of the falsehoods that have been detected by the <u>CoronaVirusFacts/DatosCoronaVirus alliance</u>. This database unites fact-checkers in more than 70 countries and includes articles published in at least 40 languages. Para búsquedas en **español**, haz <u>clic aquí</u>. Faça buscas em **português** aqui.

			SEARCH		
Select a country	Select a rating	Select an organization	•	VIEW	X

Improving the study

Method level improvement

"In place of ANOVA and Wilcoxon test, use different feature reduction methods like singular value decomposition (SVD), principal component analysis (PCA), or use covariance matrix to finds related features and remove them."

- Aftab Alam

"incorporating deep learning models such as DeepMoji or sentiment analysis models such as VADER that understand the emotion behind emojis can be used to add another layer of truth behind text analysis.... BERT"

- Ankita

Future Directions

Several new directions listed in your reflections. Too many to fit in one slide!!!

"Where were these users geographically located? What were the demographics of the users?... how do user sharing practices evolve over time, if at all? (2020 paper) "

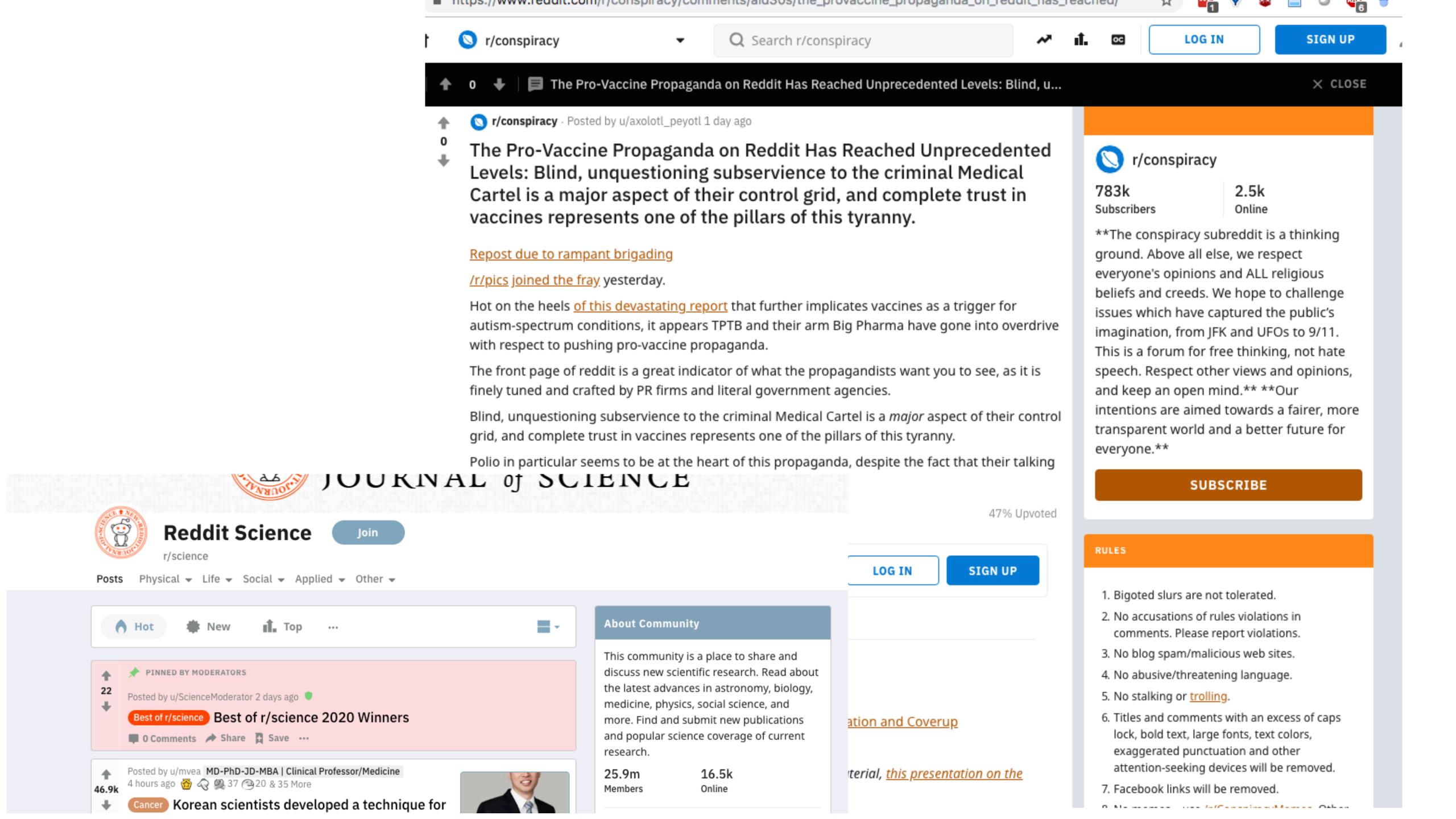
-Jordan

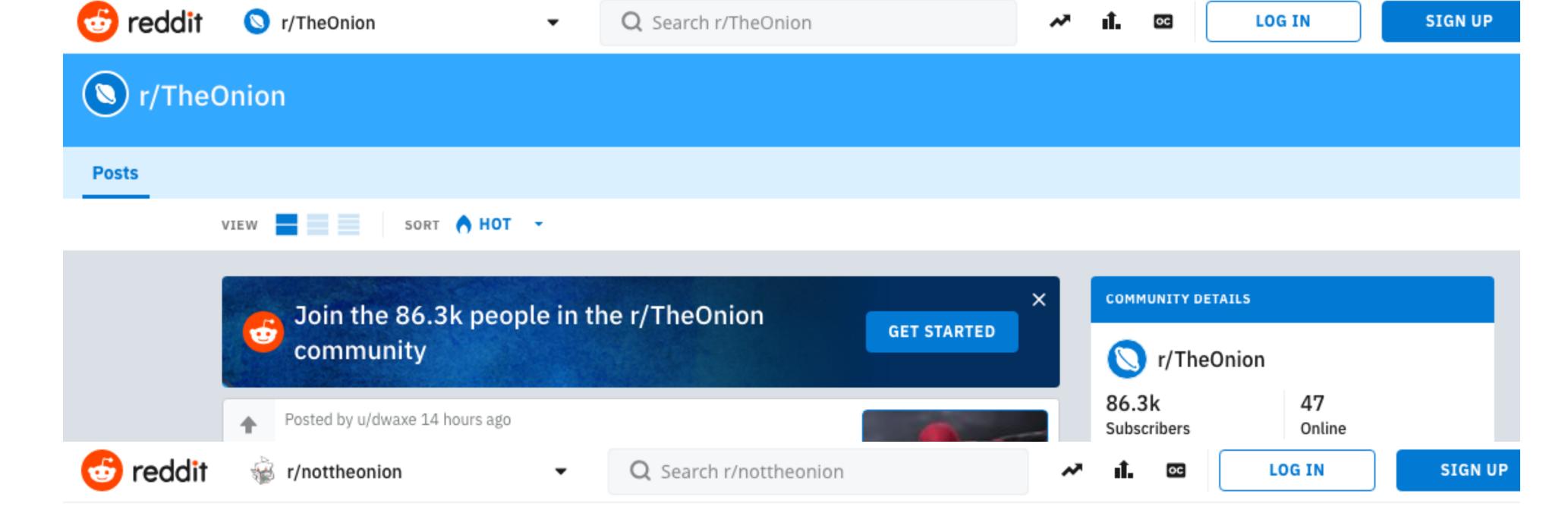
"understand the distinctions in the fake versus real news surrounding the pandemic and whether similar patterns of fake news mentioned by Horne and Adali apply to this scenario"

Anusha

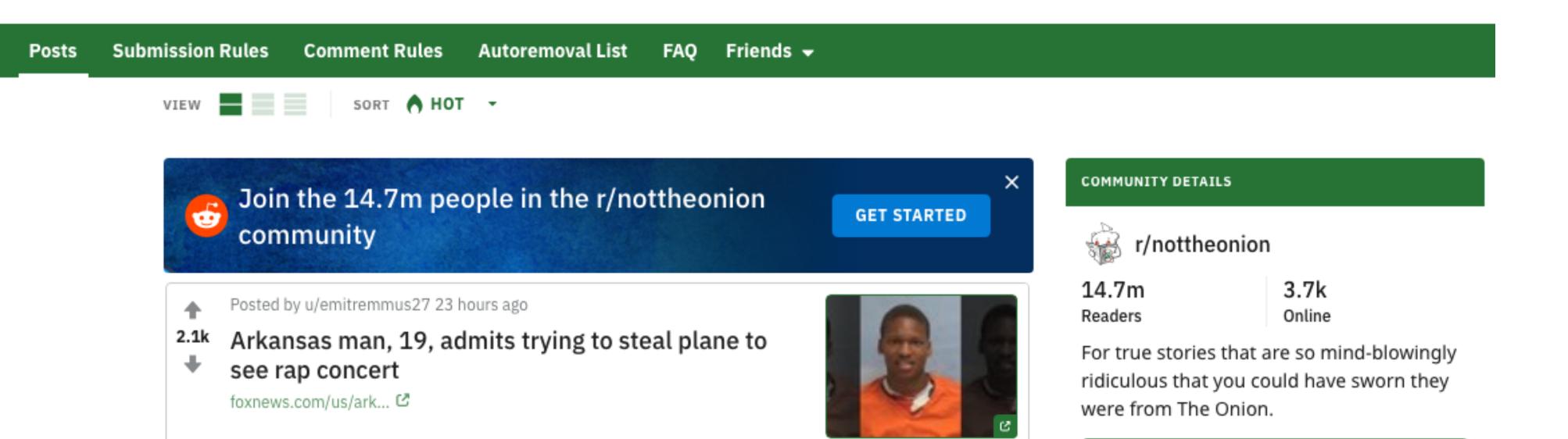
"Are the news communities found by Samory et al. easier or harder to distinguish over time?"

- Julius









Think & Share

BREAK

Will be back at 9:45 am

Lab - collect Reddit data

Next week at a glance

v

Week 4 (Jan 25-29): Getting Started with Studying Social Media Phenomena

TUE, JAN 26

Social media phenomena: Hate & conspiracies

Asking DS questions with social media

Required Reading (due reading reflections by 5pm, Mon Jan 25)

- Early Public Responses to the Zika-Virus on YouTube: Prevalence of and Differences Between Conspiracy Theory and Informational Videos
- Automated hate speech detection and the problem of offensive language

Optional material:

hate speech classifier github repo

THU, JAN 28

Asking DS questions with social media

Exploratory Data Analysis

Required Reading

Python Plotting for Exploratory Data Analysis

Lab: EDA

Idea generation writeup (due by Fri, 29-Jan)