

# Overview of Social Media Sites

IMT 547 - Social Media Data Mining and Analysis

11-Jan-2021 (Week 2, Day 3)

# Today's Plan

1. Paper reflections (summary, methods, new ideas, new questions,....)
  - *Why We Twitter: Understanding Microblogging Usage and Communities*
  - *Is It Really About Me?: Message Content in Social Awareness Streams*
2. Lab (*jupyter notebook*) — python intro

## **“Is it really about me? Message Content in Social Awareness Streams”, *naaman et al.***

- Analysis of communication patterns on Twitter (tweet content analysis)
- Both qualitative (hand coding of messages) and quantitative methods used
- Findings indicate the presence of two kinds of individuals: “meformers” and “informers”
  - Meformers garner more mentions, informers garner more retweets
  - Informers have more followers, Meformers are more women and mobile users.

# Summary

- *Why We Twitter: Understanding Microblogging Usage and Communities*

The *first* paper (Java et al.) talks about how

- Twitter's social network can be viewed as a graph where each node represents a user, and each edge represents the friend relation between the 2 users.
- By analyzing this link structure of users, the authors have been able to establish the topological and geographical properties of Twitter's social network,
- categorize users based on their activity and understand community intentions based on the interactions of similar or like-minded Twitter users.

The *second* paper (Naaman et al.) is an extension of the network-based analysis mentioned in the first paper by Java et al. wherein the authors have

- categorized Twitter messages based on their content to examine user activities as well as
- understand the difference in users based on the content they post on Twitter.

Ankita Naikdalal

Your personal reflections....

# Papers are more than a decade ago!!

“Both ...were written roughly over a decade ago and yet, some of the themes the papers discuss are still relevant. In particular, the three types of distinct user activities described by the latter 1) information seeking 2) information sharing and 3) social activity are still very accurate at a high level today.”

Jordan Bethea

# Differences between twitter of 2010 and Twitter of today

“How many of their findings would still be true today?

Would other categories emerge that were not initially identified by the authors?”

Stephen Prochaska

How microblogging has evolved with the additions of features like threads...users can now post multiple linked tweets as a workaround of the 140-character limit on Twitter.

Jordan Bethea

Features such as direct messaging, personalized feeds, verified accounts, and many more. With the addition of these features, , I believed it has altered the way users use the platform.

Julius Coleman

# Differences between twitter of 2010 and Twitter of today

Other categories mentioned in your reflections:

- Influencers
- Bots

where do **'Influencers'** fit in this spectrum....I would actually put Influencers in the Informers category.

Sakshi Agarwal



# Differences with other platforms

E.g. tiktok

TikTok assertively answers anyone's *what should I watch* with a flood. In the same way, the app provides plenty of answers for the paralyzing *what should I post?*

From the optional reading NYTimes article.

“TikTok is a little different from Twitter, in which you do not need to contact or follow anybody, and it will still share information to you. In contrast, the Twitter users need to select what they like or who they would like to follow or even connect to friends in order to access the information in 2010. ”

Yuanfeng Li (Jacky Li)

# Temporal evolution: Individual's social media usage over time

...the evolutions of one's engagement with the platform over time.

Jordan Bethea

**How Trump Reshaped the Presidency in Over 11,000 Tweets**

**<https://www.nytimes.com/interactive/2019/11/02/us/politics/trump-twitter-presidency.html>**

## How Trump Reshaped the Presidency in Over 11,000 Tweets

---

**By MICHAEL D. SHEAR, MAGGIE HABERMAN, NICHOLAS CONFESSORE,  
KAREN YOURISH, LARRY BUCHANAN and KEITH COLLINS**

Nov. 2, 2019

# Possible cultural factors

**Asia has the highest percentage of reciprocity** among North America and Europe, which makes me wonder there might be a cultural factor that plays a part. .... people in my home country (China) are interacting with each other in a certain form of reciprocity (in our culture, we call it 'Li Shang Wang Lai')....

Yajie Zhu, Zoe

## Reciprocity (*lishang-wanglai*): A Chinese model of social relationships and reciprocity – state and villagers' interaction 1936–2014

Xiangqun Chang

First Published February 16, 2016 | Research Article



<https://doi.org/10.1177/1440783315589151>

[Article information](#) ▾



### Abstract

In Kaixiangong village, south-east China, the ordinary villagers seek ways to cope with change and improve their lives by constantly maintaining and managing relationships not only between themselves but also with the state. Fieldwork carried out on social support from state or private sources (1979–96), has been supplemented with material from 1936 to the present. The findings show the majority of villagers' resource-seeking actions are rooted in implicit cultural models and patterns of social relationships, a Chinese model of social relationships and reciprocity; the current situation results from reciprocal influence and accommodation between villagers and the state over a long period.

It is observed that **Asia has more reciprocity** in terms of friendships than North America and Java et al. believe that ‘language’ plays an important role here. Considering that Asia speaks more languages than North America, and is in general, more diverse ethnically, this observation comes as a surprise.

Neeraj Kulkarni

# Communities within Twitter

Java et al describe the formation of communities within Twitter where users of a community share similar interests as well as talk about their personal feelings and experiences.

An interesting factor to determine the strength of a community could be **what impacts do such communities have on the real world?**

A recent example of this comes across from one of the biggest communities on Twitter called ARMY....K-pop boy band BTS. Shortly after news broke that the boy band had donated 1 million dollars to the racial justice organization Black Lives Matter,..., the ARMY community on Twitter started an online donation campaign based on the hashtag “**#MatchAMillion**” where they reached their goal of 1 million dollars within 25 hours.

Malvika Mohan



# Recent ACM CHI 2021 paper

## Armed in ARMY: A Case Study of How BTS Fans Successfully Collaborated to #MatchAMillion for Black Lives Matter

So Yeon Park  
Stanford University  
Stanford, CA, USA

Blair Kaneshiro  
Stanford University  
Stanford, CA, USA

Nicole Santero  
University of Nevada, Las Vegas  
Las Vegas, NV, USA

Jin Ha Lee  
University of Washington  
Seattle, WA, USA

### ABSTRACT

Music fans strategically support their artists. Their collective efforts can extend to social causes as well: In 2020 for example, ARMY—the fandom of the music group BTS—successfully organized the #MatchAMillion campaign to raise over one million USD to support Black Lives Matter. To better understand factors of fandoms’ collaborative success for arguably unrelated social goals, we conducted a survey focusing on ARMYs’ perceptions of their fandom and their social effort. Most ARMYs viewed the fandom as a community, loosely structured with pillar accounts. They reported trust in each other as well as high team composition, which mediated the relationship between their neutral psychological safety and high efficacy. Respondents attributed their success in #MatchAMillion to shared values, good teamwork, and established infrastructure. Our findings elucidate contextual factors that contribute to ARMY’s collaborative success and highlight themes that may be applied to studying other fandoms and their collaborative efforts.

### 1 INTRODUCTION

Fans of celebrities, movies, characters, sports teams, and other public figures are increasingly growing in size and impact, as evident from such tangible examples as the crowdsourced fan series “Star Trek Continues” [55, 63] and the work done by the Harry Potter Alliance, a non-profit run primarily by Harry Potter fans who fight for human rights.<sup>1</sup> Similarly, music fandoms have become prominent in the current landscape of social media, collaboratively accomplishing goals on these platforms on a regular basis [39, 47]. Due to fandoms’ massive size and influence, it is often the negative aspects, such as toxic behavior (e.g., bullying or doxxing), that garner media attention [12, 33, 51, 64]. Yet fandoms are also sources of positive influence, as showcased by numerous donations and other forms of support that fans provide not only for artists, but also for each other—at times for social causes that seem unrelated to their artists.

ARMY (Adorable Representative M.C. for Youth) is the fandom of BTS, a music band of seven members originating from South

# Communities within Twitter

Java et al describe the formation of communities within Twitter where users of a community share similar interests as well as talk about their personal feelings and experiences.

An interesting factor to determine the strength of a community could be **what impacts do such communities have on the real world?**

we have also seen communities work together to take a stand on several social movements trending over hashtags such as the “**#MeToo**”, “**#LoveWins**” which brings up another interesting question about how communities work together to tackle social and political issues?

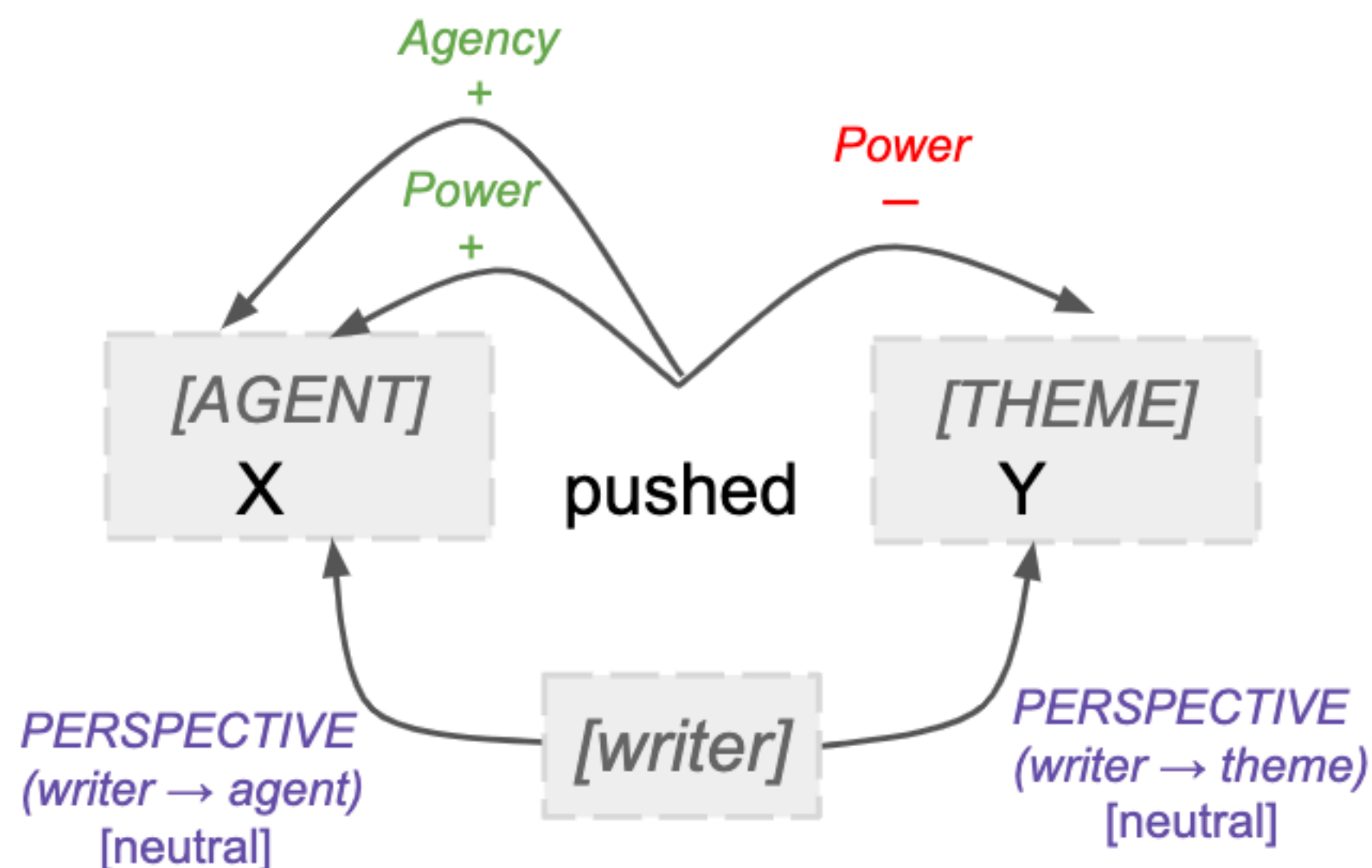
Malvika Mohan



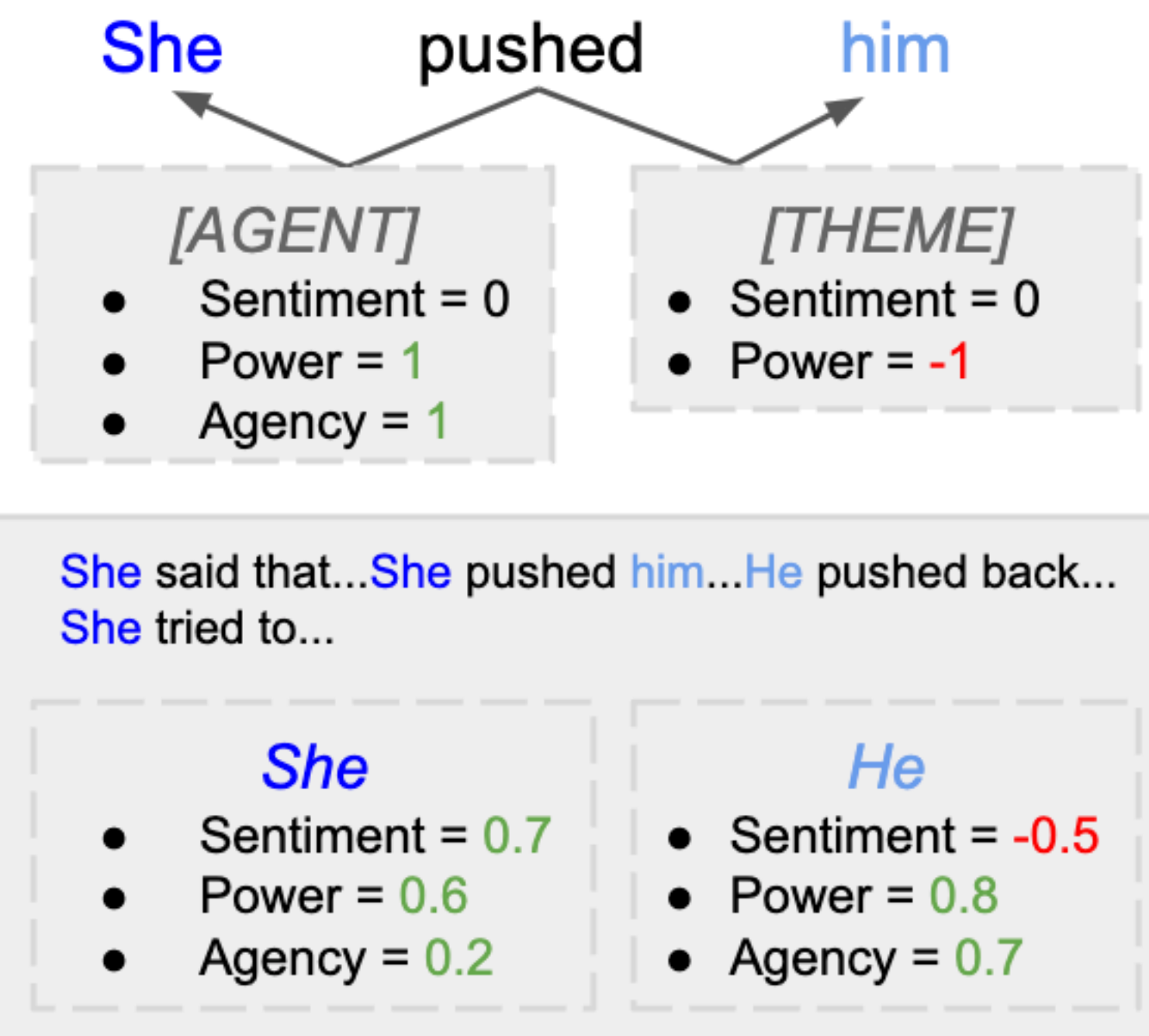
# Contextual Affective Analysis: A Case Study of People Portrayals in Online #MeToo Stories

Anjalie Field, Gayatri Bhat, Yulia Tsvetkov

Language Technologies Institute  
Carnegie Mellon University  
{anjalief, ytsvetko}@cs.cmu.edu



**Connotation Frames**



**Contextual Affective Analysis**

Figure 1: Left, we show off-the-shelf connotation frame annotations (Rashkin, Singh, and Choi 2016; Sap et al. 2017) for the verb “push”. Right, we show the proposed adaptation. We adapt connotation frames from a verb-centric formalism to an entity-centric formalism, transferring scores from verbs to entities using a context-aware approach (top right). We then aggregate contextualized scores over all mentions of entities in a corpus (bottom right). This new approach—*contextual affective analysis*—enables us to obtain sentiment, power, and agency scores for entities in unannotated corpora and conduct extensive



# Future research questions

Extending the analysis to other platforms

“In today’s time, this analysis can be extended to various social media platforms such as **Instagram, Snapchat or even LinkedIn** wherein users “follow” each other and in turn form communities based on similar user intentions or activities.”

- Ankita Naikdalal

In place of analyzing tweets, we can analyze image content, blog posts or even user subscriptions to better understand user behavior patterns and community trends on these platforms.

With the advent of advanced machine learning techniques, algorithms such as LDA can be used to cluster users....involve sentiment analysis using algorithms such as VADER, TextBlob etc

# Future research questions

**Why do “information sharers” have more friends, follows, and interaction with the community?**

Naaman et al. did not explore the “why” of this trend; they only identified its existence.

Kielan Lemoine-Kowalski

# Future research questions

Research has revealed the trends of political polarization in the US. Is Twitter playing a role, or what kind of role are platforms like Twitter playing in this phenomenon?

Lizzy Chen



Lada Adamic

Data Scientist

Data Science

CORE DATA SCIENCE (CDS)

## Advancing the frontiers of data science

RESEARCH | REPORTS

## POLITICAL SCIENCE

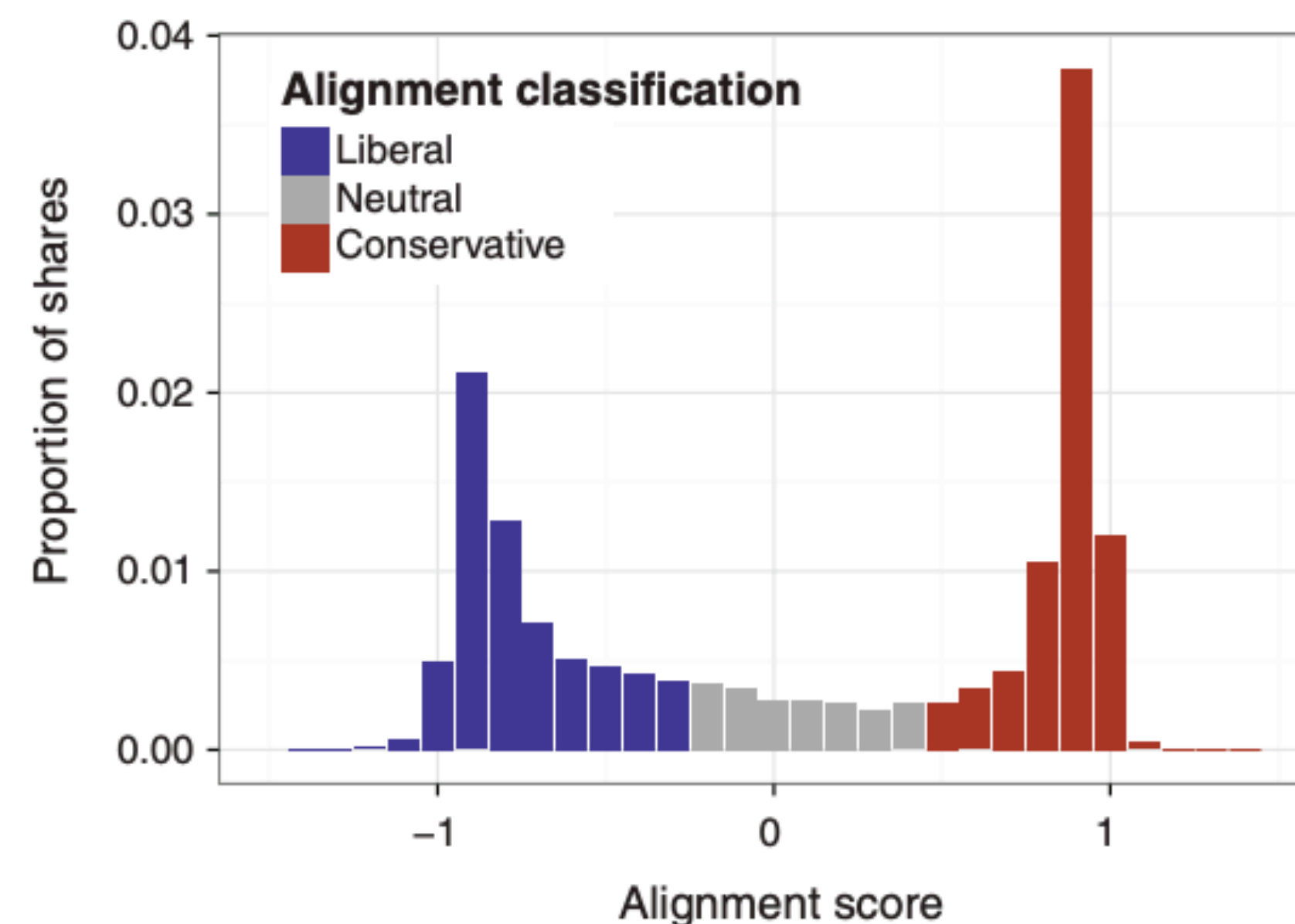
# Exposure to ideologically diverse news and opinion on Facebook

Eytan Bakshy,<sup>1\*</sup>† Solomon Messing,<sup>1†</sup> Lada A. Adamic<sup>1,2</sup>

Exposure to news, opinion, and civic information increasingly occurs through social media. How do these online networks influence exposure to perspectives that cut across ideological lines? Using deidentified data, we examined how 10.1 million U.S. Facebook users interact with socially shared news. We directly measured ideological homophily in friend networks and examined the extent to which heterogeneous friends could potentially expose individuals to cross-cutting content. We then quantified the extent to which individuals encounter comparatively more or less diverse content while interacting via Facebook's algorithmically ranked News Feed and further studied users' choices to click through to ideologically discordant content. Compared with algorithmic ranking, individuals' choices played a stronger role in limiting exposure to cross-cutting content.

## KEY FINDING:

**Compared with algorithmic ranking, individuals' choices played a stronger role in limiting exposure to cross-cutting content.**



**Fig. 1. Distribution of ideological alignment of content shared on Facebook measured as the average affiliation of sharers weighted by the total number of shares.** Content was delineated as liberal, conservative, or neutral on the basis of the distribution of alignment scores (details are available in the supplementary materials).

# Other extensions

The author has manually determined the possible intentions of a user on Twitter to categories like Daily Chatter, Conversations, Sharing information, Reporting News, etc. The author could also use machine learning techniques to classify the nature of the information being shared. Manually determining categories would not be possible with a huge dataset too.

Subhiksha Mukuntharaj

there is much more data we can use today which machine learning will be able to use more efficiently than those doing it by hand.

Julius Coleman

# Other extensions

## Analyzing Fake news spread

Some of the methods used in the paper such as Clique Percolation Method (CPM), HITS algorithm, and log-likelihood calculation can be incorporated by my team as we attempt to analyze how fake news is spread in the Twitter community.

Meghana Dinesh Parab

This research paper can also aid and contribute towards the development of a more resourceful and trusted source of providing news to users. For example – in the early days of COVID-19, there were a lot of incorrect facts and figures being spread on Twitter, and users found it difficult to trust a source. With the help of identifying a hub and the source of such incorrect information, Twitter would be able to cease the spread of such incorrect information.

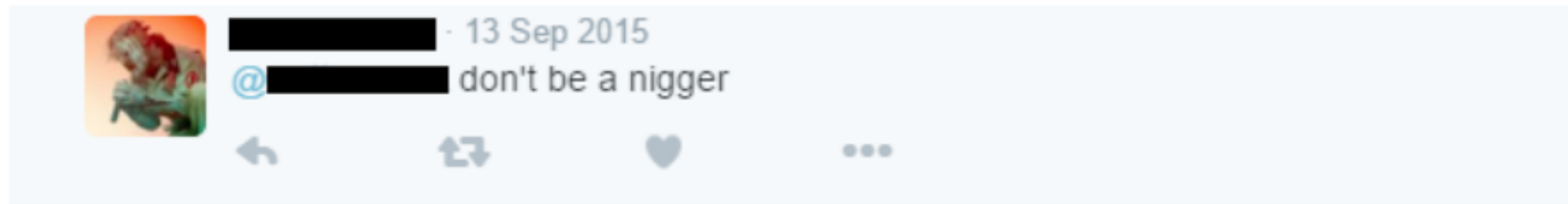
Saransh Arora

# Moderation & Free speech

“While not a topic of the paper, it relates to “informers” or “information providers” again. Should people disseminating information be moderated by Twitter. Should people disseminating information be moderated by Twitter. Is this a limitation on freedom of speech?”

Kielan Lemoine-Kowalski





**Rasheed Smith**  
@RasheedSmith45

Bot tweeting back

@ [redacted] Hey man, just remember that there are real people who are hurt when you harass them with that kind of language

Munger, Kevin. "Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment." *Political Behavior* (2016): 1-21.



# Notices on Twitter and what they mean

<https://help.twitter.com/en/rules-and-policies/notices-on-twitter>

As automations may result in mistakes, no suspensions will result from automatic warnings placed on potentially sensitive media.

This Tweet may include sensitive content.

The following media includes potentially sensitive content.  
[Change settings](#)

[View](#)

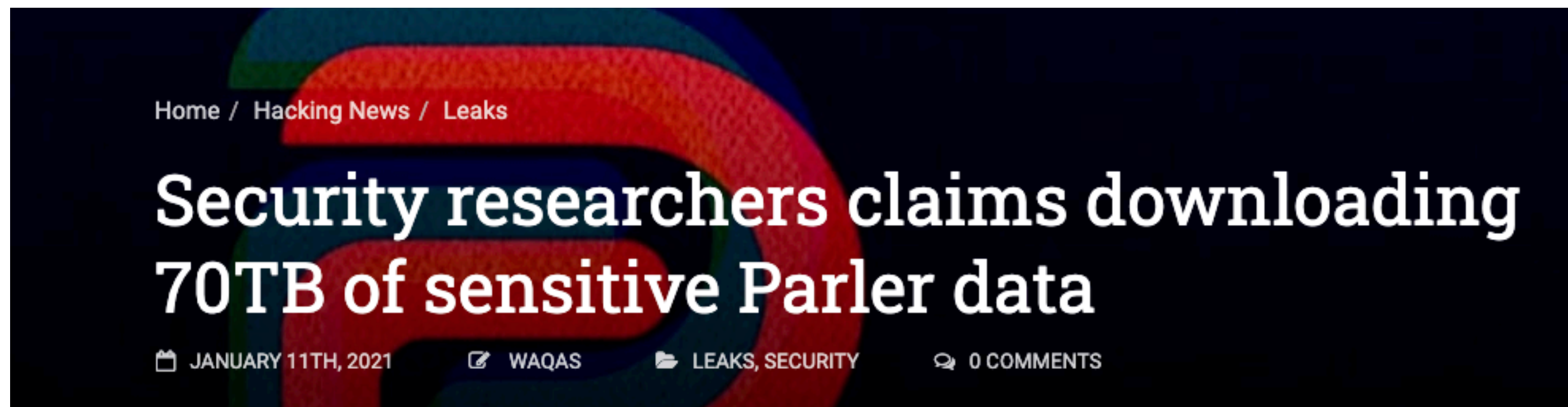
# Privacy Issues

## Privacy Issues

“I know some horror films are about how people get murdered because of the personal information they reveal on social media.....But I think the awareness to protect sensitive data is much higher than ten years ago. And more improved policies and measures have been implemented to protect the users. So I doubt if there are still so many "Me Now" posts on social media.”

Esther Yang

# Privacy Issues



## Millions of Videos, Posts, and Other Data Exposed

Donk\_enby wrote that more than a million video URLs, including private data, are part of the data dump. It also included deleted posts because instead of removing the post, Parler removed the pointer to that post to make it inaccessible to users.

## Verified Citizens' Data At Risk

Security researchers claim that the scrapped data is linked to the accounts which posted it. Some of the images and videos also contain geolocation information. Data from Parler's Verified Citizens refers to the users who had verified their identity by uploading their government-issued ID pictures, such as their driver's license.

Researchers revealed that images and video data still contain the EXIF data, including metadata of date, time, and location.



a sample of what's in there



Why is it useful/important after all, for us to learn about different uses of Twitter?

# Method: Defining active users

To operationalize the notion of active users, the paper selected users who had at least 10 friends, 10 followers, and had posted at least 10 messages. We randomly selected users that fit those criteria from our sampling frame.

We **manually examined** each user's profile details, and coded them for 'personal use', ruling out in the process commercial entities, as well as people solely promoting their services or businesses. The process left us with 911 users, out of which we randomly selected 350 users for analysis.

The paper uses a **small dataset** (350 users), what kind of **biases** can this potentially introduce in the categorization of “meformers” and “informers”?

Java et al. talk about Twitter's social network being clustered.

What benefits does a **clustered network** bring to its members? What are its pitfalls?

LAB



BREAK

Back by 9:55am

## Next Class

2nd reflection due by 5pm on Wednesday (no late days allowed on reflection, see late policy)



### Week 2 (Jan 11-15): Overview of Social Media Sites & Phenomena

#### TUE, JAN 12

##### Overview of social media sites

**Required Reading** (*due reading reflections by 5pm, Mon Jan 11*):

- [Why We Twitter: Understanding Microblogging Usage and Communities](#)
- [Is It Really About Me?: Message Content in Social Awareness Streams](#)

##### Optional Readings

#### THU, JAN 14

##### Social media phenomena: Identity & deception

**Required Reading** (*due reading reflections by 5pm, Wed Jan 13*):

- [Identity and Deception in the Virtual Community](#)
- [4chan and /b: An Analysis of Anonymity and Ephemerality in a Large Online Community](#)