Understanding the Diverging User Trajectories in Highly-related Online Communities during the COVID-19 Pandemic

Jason Shuo Zhang,[1] **Brian Keegan,** [1] **Qin Lv,** [1] **and Chenhao Tan** [2]

[1] University of Colorado Boulder
[2] University of Chicago
{jasonzhang,brian.keegan,qin.lv}@colorado.edu, chenhao@uchicago.edu

Abstract

As the COVID-19 pandemic is disrupting life worldwide, related online communities are popping up. In particular, two "new" communities, /r/China_flu and /r/Coronavirus, emerged on Reddit and have been dedicated to COVID-related discussions from the very beginning of this pandemic. With /r/Coronavirus promoted as the official community on Reddit, it remains an open question how users choose between these two highly-related communities.

In this paper, we characterize user trajectories in these two communities from the beginning of COVID-19 to the end of September 2020. We show that new users of /r/China_flu and /r/Coronavirus were similar from January to March. After that, their differences steadily increase, evidenced by both language distance and membership prediction, as the pandemic continues to unfold. Furthermore, users who started at /r/China_flu from January to March were more likely to leave, while those who started in later months tend to remain highly "loyal". To understand this difference, we develop a movement analysis framework to understand membership changes in these two communities and identify a significant proportion of /r/China_flu members (around 50%) that moved to /r/Coronavirus in February. This movement turns out to be highly predictable based on other subreddits that users were previously active in. Our work demonstrates how two highly-related communities emerge and develop their own identity in a crisis, and highlights the important role of existing communities in understanding such an emergence.

Introduction

In December 2019, a novel coronavirus strain (SARS-CoV-2) emerged in Wuhan, China. The disease (COVID-19) quickly spread all over the world and led the World Health Organization (WHO) to declare a pandemic (Times 2020). By the end of September 2020, 189 countries/regions reported over 34 million positive cases and 1.0 million deaths (Dong, Du, and Gardner 2020). The pandemic is not only exhausting public health resources but also causing social and economic disruption at an unprecedented speed and scale.

Social media are critical for people to access information and share experiences during this pandemic. Take Reddit as an example. User-created subreddits (communities) have popped up for discussions on coronavirus

and attracted millions of subscribers. Health professionals (Whalen 2016), essential workers (Winkie 2020), and recovered patients (Marcin 2020) are reportedly using Reddit as a primary source to share information. Despite the importance of these communities, it remains an open question as to how these highly-related communities emerge and form their identity in the context of a crisis.
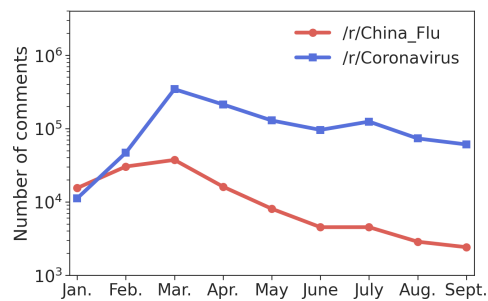
In this work, we focus on two highly-related communities on Reddit, /r/China_flu and /r/Coronavirus. Both communities are for *general* discussions about COVID-19 and were "founded" at the very beginning of the pandemic. There were more active users in /r/China_flu than in /r/Coronavirus in January 2020 (Figure 1). After that, /r/Coronavirus exploded, as the platform made it the official community for COVID-19 on February 17th (Figure 2). In comparison, user activity in /r/China_flu has gradually decreased. The emergence of these two communities during a once-in-a-lifetime pandemic provides a unique opportunity for understanding the formation of communities during a crisis, bridging the literature on highly-related communities (Hessel, Tan, and Lee 2016; Tan 2018; Zhu et al. 2014; Zhang, Tan, and Lv 2018; Zhu, Kraut, and Kittur 2014; Waller and Anderson 2019) and crisis informatics (Palen et al. 2007, 2009; Palen and Anderson 2016; Hagen et al. 2018; Reuter and Kaufhold 2018; Reuter, Hughes, and Kaufhold 2018; Maas et al. 2019).

We take a user-centered perspective, as communities are ultimately determined by their members. A key observation is that users in these two communities were similar from January to March and then diverge. To further understand this divergence, we examine two main questions: 1) Which users chose to start at /r/China_flu vs. /r/Coronavirus: We find that new users who started at /r/China_flu were already different from those who started at /r/Coronavirus in their use of language before joining either community, and this difference has grown over time. 2) How users move between these two communities: We find that a significant proportion of /r/China_flu (around 50%) members moved to /r/Coronavirus in February. Moreover, this movement is highly predictable based on other subreddits that users were previously active in.

Organization and highlights. We provide a detailed overview of /r/China_flu and /r/Coronavirus after discussing related work. We demonstrate how user activity is connected
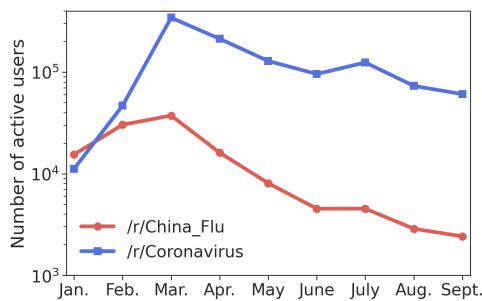
(a) Number of active users per month. (b) Number of comments per month.

Figure 1: The number of active users and comments per month, respectively, in /r/China_flu and /r/Coronavirus on a *log* scale. In January and February, user activity was at a similar level between these two subreddits. After that, /r/Coronavirus became much more popular than /r/China_flu.

with the important dates of this pandemic, including February 17th, when Reddit made /r/Coronavirus the official subreddit for COVID-19, and March 11th, when the WHO declared COVID-19 a pandemic. We show that both communities share similar founders who tend to be newsreaders or survivalists. In March, more than half of the /r/China_flu users were also active in /r/Coronavirus. After that month, the overlap decreases over time.

Inspired by the separation of active users in these two communities, we compare the differences between cohorts of *new* users who chose to start at /r/China_flu vs. at /r/Coronavirus, grouped by their starting month. Our results show that the language distance between users who started at /r/China_flu or /r/Coronavirus was small from January to March. After that, the language distance goes up, indicating an increasing distinction between users who chose to start at /r/China_flu vs. /r/Coronavirus. We observe a similar trend in a prediction task: Users' previous activity on Reddit can be used to predict which community they chose to start at, and the accuracy is better in later months than in earlier ones. These observations indicate that the separation between /r/China_flu and /r/Coronavirus users has widened as the pandemic unfolds.

In addition to the choice of where to start, we further examine whether users move between these communities. We find that users who started at /r/China_flu from January to March reduced their activity in this community over time, while those in later months tend to be highly "loyal". To further understand the movement between these two communities, we develop a framework to identify membership changes. Our framework shows that around 50% of /r/China_flu members moved to /r/Coronavirus in February, when /r/Coronavirus became the official community for COVID-19 on Reddit. This ratio is robust, even if we apply different membership definitions. In comparison, users rarely moved in the reverse direction. Moreover, we demonstrate that these users' movements from /r/China_flu to /r/Coronavirus are highly predictable based on other subreddits that users were previously active in. Donald Trump supporters, conspiracy theorists, and survivalists were more likely to stay in /r/China_flu. In contrast, Bernie Sanders (a

U.S. Democratic presidential candidate) supporters and science enthusiasts were more likely to leave /r/China_flu.

We offer concluding discussions in the end. Our work demonstrates the emerging process of two highly-related communities in a crisis through the perspective of their members. We show that these two communities resemble each other in the beginning and then gradually diverge. Despite the dominance of /r/Coronavirus, /r/China_flu forms its own identity and can still attract users with high loyalty as the pandemic unfolds. We also highlight the critical role of existing communities in this process.

Related Work

We review the literature in two areas that are most relevant to our work: highly-related communities and crisis informatics.

Highly-related Communities & Community Genealogy

When social media platforms give users the freedom to form interest groups, a series of highly-related communities can pop up. For example, during the 2016 election, a battery of Trump-related communities are created on Reddit, such as /r/The_Donald, /r/AskThe_Donald, and /r/AskTrumpSupporters. The creation, development, and lifecycle of highly-related communities have drawn considerable interest in the research community. The first line of work compares the characteristics of highly-related communities. Hessel, Tan, and Lee (2016) investigate the interactions between highly-related Reddit communities and identify patterns of affixes being used in their names. The work by Zhang, Tan, and Lv (2018, 2019) focuses on online NBA fan communities and analyzes how fans of different teams talk with each other and react to team performance. The second line of research studies the impact of membership overlap between highly-related communities. Haiyi Zhu, Robert Kraut, and collaborators investigate shared membership in online communities on a variety of platforms and propose strategies for long-term community survival (Zhu, Kraut, and Kittur 2014, 2013; Kraut and Resnick 2012; Zhu et al. 2014). The final line provides a global overview of how users move through the space of communities and how new

communities emerge from the old ones. For example, Tan and Lee (2015) use several temporal features across communities to predict users' activity levels and departure from Reddit. To understand how new communities are developed from old ones, Tan (2018) proposes a computational approach for building genealogy graphs between communities.

Crisis Informatics & Sociology of Disaster

Another closely related line of work is crisis informatics (Palen and Anderson 2016). Understanding disaster events and their impacts is a critical topic of societal relevance (United Nations Office 2018). In the early days of crisis informatics research, qualitative methods, such as descriptive surveys and interviews, were the main sources of data collection (Palen et al. 2007; Sutton, Palen, and Shklovski 2008; Palen et al. 2009). Recently, following the advancement of the Internet and mobile technology, social media have played a critical role in the flow of public information. A growing percentage of citizens frequently turn to these platforms for emergency updates (Lachlan et al. 2016; Rene 2016). To effectively analyze information on mass media, computational approaches have been widely adopted in crisis informatics. For example, Vieweg et al. (2010) analyze public responses to two disaster events, the Red River Floods and the Oklahoma Grassfires, using Twitter communications. Hagen et al. (2018) propose a network analysis approach to identify a number of distinct communities and influential actors using Zika-related tweets. Using Hurricane Sandy as a study subject, Stewart and Wilson (2016) characterize how citizens utilize social media to redistribute emergency updates and connect with family and friends.

Despite tremendous effort in these two areas, the phenomenon of highly-related communities emerging during a crisis is understudied. The substantial impact caused by COVID-19 provides an opportunity to understand this phenomenon. To the best of our knowledge, our work represents the first attempt towards this direction by unpacking the emerging process of two highly-related communities.

Recent studies on COVID-19. Meanwhile, there are recent studies that examine the public response to the COVID-19 pandemic (Van Bavel et al. 2020; Yin et al. 2021). In particular, sentiment and language usage in COVID-related conversations on social media are investigated from various angles (Budhwani and Sun 2020; Chen et al. 2020; Pei and Mehta 2020; Lyu et al. 2020). In this work, we also compare the language usage between users in /r/China_flu and /r/Coronavirus to investigate the separation of these two communities as the pandemic unfolds.

An Overview of /r/China_flu and /r/Coronavirus

Our main dataset is drawn from Reddit,[1] a community-driven forum for discussion, news consumption, and content rating. It was founded in 2005 and became the 21st most popular website globally in May 2020 (Alexa 2020). There are tens of thousands of communities (known as subreddits)

(A)

(B)

Figure 2: On February 17th, Reddit decided to make /r/Coronavirus the official community for COVID-19. All the comments submitted to /r/Coronavirus are heavily moderated. Meanwhile, the platform allows more relaxed discussions in /r/China_flu.

on Reddit dedicated to a wide variety of topics. Users can submit, comment on, upvote, and downvote content in each subreddit.

Since the outbreak of COVID-19 in January 2020, a series of COVID-related communities have emerged and drawn substantial public attention. In this study, we focus on two communities, /r/China_flu[2] and /r/Coronavirus,[3] which have been dedicated to general COVID-related discussions from the very beginning of this pandemic. Other COVID-related communities tend to focus on subtopics such as scientific discussions (e.g., /r/COVID19) and specific regions (e.g., /r/CanadaCoronavirus).

/r/China_flu was founded on January 20th, 2020, when news about the first breakout of COVID-19 emerged in Wuhan, China. /r/Coronavirus was founded on May 3rd, 2013, but remained inactive most of the time. The first post in /r/Coronavirus since 2017 was about COVID-19 on January 20th, 2020. Using the pushshift.io website (Baumgartner et al. 2020), we collect all the comments submitted to these two subreddits from January 20th to September 30th in 2020. Table 1 gives summary statistics about these two subreddits. In the rest of the paper, we focus on analyzing users' comments since posts are usually direct links to news articles and more formal, thus not comparable to comments.

Important Dates

On February 11th, the WHO named the new coronavirus disease "COVID-19". It was specifically named this way to avoid calling it the China virus or the Wuhan virus (SCHWEDEL 2020). On February 17th, the Reddit plat-

	#posts	#comments	#users
/r/China_flu	81K	1.3M	76K
/r/Coronavirus	312K	7.9M	664K

Table 1: Dataset statistics as of September 30th, 2020. Here #users refers to the number of unique users who have posted/commented in that subreddit.

Subreddit	prop.	Subreddit	prop.
/r/collapse	28%	/r/news	8%
/r/worldnews	18%	/r/worldnews	4%
/r/news	14%	/r/AskReddit	4%
/r/AskReddit	10%	/r/MMA	4%
/r/todayilearned	10%	/r/interestingasfuck	4%
/r/preppers	10%	/r/AskMen	4%
/r/conspiracy	8%	/r/perfectlycutscreams	2%
/r/aww	6%	/r/collapse	2%
/r/environment	6%	/r/preppers	2%
/r/CollapseSupport	6%	/r/nCoV	2%

Table 2: The top-10 subreddits that founders of /r/China_flu (left) and /r/Coronavirus (right) had commented on 30 days before they joined these two subreddits. Here we define founders of each community as the first 50 users who commented in that community. Prop. indicates the percentage of founders that were active in the parent community.

form decided to make /r/Coronavirus the official community for COVID-related updates. Users who searched for COVID-related keywords have since been recommended to check out /r/Coronavirus (Figure 2(A)). The subreddit is described as a place for high-quality discussions. All the posts and comments submitted are strictly moderated. Meanwhile, the platform allows for more relaxed discussions in /r/China_flu (Figure 2(B)). This policy shift is reflected by user activity in these two subreddits. As illustrated in Figure 1, the monthly number of posts and comments in these two communities was similar in January and February. After that, /r/Coronavirus became much more popular than /r/China_flu.

On March 11th, the WHO declared COVID-19 a pandemic (Times 2020). The number of comments generated in /r/China_flu and /r/Coronavirus peaked that month. User activity in both communities started to drop after that month. This may be due to people's fatigue of COVID-related topics. According to a study released by the Pew Research Center at the end of April, 71% of Americans said they needed to take a break from news about the coronavirus, and 43% said they felt worse emotionally as a result of following updates (Mitchell, Oliphant, and Shearer 2020). The user activity in /r/Coronavirus bounced back a bit in July, likely due to the second wave of coronavirus infections in the U.S. and many other countries (Times 2020), while user activity in /r/China_flu kept going down.

To summarize, the explosion of /r/China_flu and /r/Coronavirus happened from January to March, when the virus began to spread worldwide. Therefore, the period from
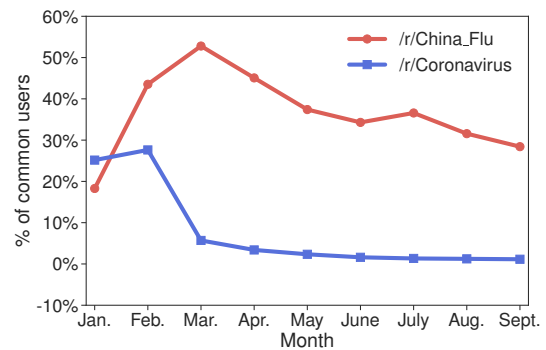

Figure 3: The proportion of monthly overlapping users between /r/China_flu and /r/Coronavirus. The overlap is measured by the proportion of commenters that are active in both communities.

January to March might be especially interesting for understanding the emergence of these two communities.

Membership in /r/China_flu and /r/Coronavirus

We present exploratory analyses of membership in these two communities. We start by looking at the early users ("founders") in each community and then examine the overlap of active users in these two communities over time.

Founders of /r/China_flu and /r/Coronavirus. A straightforward way to explore the beginning stage of these two new communities is to understand where the founders of /r/China_flu and /r/Coronavirus came from. Table 2 shows the top-10 parents of these two communities. These top-10 parents are ranked based on each community's first 50 commenters' activity on Reddit 30 days before joining /r/China_flu or /r/Coronavirus. Interestingly, the founders of /r/China_flu and /r/Coronavirus were both active in /r/collapse and /r/preppers, two subreddits that attract survivalists to discuss the potential collapse of global civilization. It suggests that people who paid attention to this virus at the very beginning tend to worry about the break down of our society in general (Davies 2020). The dominance of /r/collapse is more salient in /r/China_flu than /r/Coronavirus. Unsurprisingly, early members of /r/China_flu and /r/Coronavirus were also active in news-related subreddits (/r/worldnews and /r/news). This explains why they were the earliest to be aware of this new disease. Moreover, /r/conspiracy is one of the top-10 subreddits for /r/China_flu, indicating that COVID-19 topics attract conspiracy theorists (Muller 2020; Lee 2017).

The monthly overlap between /r/China_flu and /r/Coronavirus. Next, we examine the overlap between these two communities as the pandemic unfolds. We define the overlap as the proportion of commenters in one community that also commented in the other community in the same month (see Figure 3). In March, more than half of the /r/China_flu users were also active in /r/Coronavirus. After that month, the overlap goes down over time. This trend indicates that the separation between /r/China_flu and /r/Coronavirus has widened as the pandemic unfolds. As
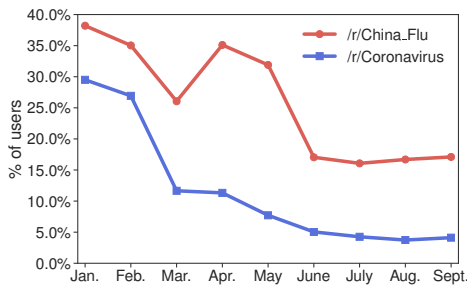
Figure 4: The proportion of users in /r/China_flu and /r/Coronavirus whose comments contain keywords "china" or "chinese" in that month.

Figure 1 shows, /r/Coronavirus has a much bigger user base since March. The fraction of its users who are also active in /r/China_flu is low.

Mention of "china"

The phrase "China Flu" is frequently used, especially by President Donald Trump, to blame China for the pandemic (Lee 2020; Joubin 2020; Schild et al. 2020). Even though the founders of /r/China_flu claim that the community "china flu" was created before the virus was officially named, it may nevertheless attract people who are anti-China. Figure 4 compares the proportion of comments that contain keywords "china" or "chinese". Indeed, /r/China_flu users talk more about China than /r/Coronavirus users in all the months. In January and February, when China was the epic center of coronavirus, users in both communities had a high rate of mentioning "china" or "chinese". After that, when China contained the virus, /r/China_flu users still maintain a high interest in discussing China-related topics, while users in /r/Coronavirus seldom bring up China. the blue line drops to a low level.

We also examine the most downvoted comments in these two communities. By design, Reddit communities allow users to upvote or downvote the comments they read. The difference between the number of upvotes and downvotes a comment receives is referred to as "score". The ranking system will display comments with the highest scores at the top of the page and hide the ones with the lowest scores. The most downvoted comments can help us understand what opinions that the community members dislike intensely.

Table 3 lists the top-5 most downvoted comments in /r/China_flu and /r/Coronavirus. All the comments in /r/China_flu are related to positive opinions about China or the Chinese government. In contrast, downvoted comments in /r/Coronavirus are mostly about downplaying the virus's seriousness. These comments suggest that users in /r/China_flu pay much closer attention to Chinese-related news and tend to show a negative opinion against China.

New Users Choosing between /r/China_flu and /r/Coronavirus

To further delineate the declining overlap in active users of these two communities, we investigate how *new* users
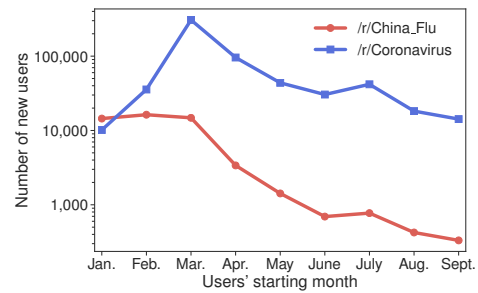

Figure 5: The number of users who start at /r/China_flu or /r/Coronavirus per month in *log* scale. Note that every user is only counted once here. If the user commented in /r/China_flu first, they will not be counted as a new user in /r/Coronavirus, and vice versa.

chose to start at /r/China_flu or /r/Coronavirus. We group users based on which month they started at /r/China_flu or /r/Coronavirus and which community they participated in first. This allows us to examine different cohorts of users and zero in on the declining overlap of the active users each month.

Figure 5 shows the number of new users who started at /r/China_flu and /r/Coronavirus each month (Note that every user is only counted once. If a user commented in /r/China_flu first, they will not be counted as a new user in /r/Coronavirus, and vice versa). Since June, /r/China_flu has become much less active: fewer than 1000 new users joined this community each month. In comparison, /r/Coronavirus attracts new users at a high level. In this section, we refer to users who chose to **start** at /r/China_flu as */r/China_flu users* and in /r/Coronavirus as */r/Coronavirus users*.

We also collect users' commenting history on the entire Reddit platform from one month before they started at /r/China_flu or /r/Coronavirus to the end of September 2020. In January and February, all the new users in both communities are included. From March to September, we include all the new users from /r/China_flu and randomly sample the same number of new users from /r/Coronavirus, as there are many more new users each month in /r/Coronavirus than in /r/China_flu.

Language Usage Difference

We characterize the behavioral differences between /r/China_flu users and /r/Coronavirus users through the lens of language usage, as posting comments is the major activity for users on Reddit. Similar to prior work (Tan, Friggeri, and Adamic 2016; Atkinson, Srinivasan, and Tan 2019; Althoff, Clark, and Leskovec 2016), we adopt the Jensen-Shannon Divergence (Manning and Schütze 1999) to measure the monthly unigram usage difference in comments as the language distance. A larger distance indicates a larger difference in language usage. For all text-related computations in this paper, we remove punctuation marks, urls, and stopwords. We also stem the words during preprocessing. The preprocessing is implemented using Gensim (Rehurek 2010). We also remove comments

	Score	Comment
/r/China_flu	-104	1. Focus less on China and more on your US. You are on brink of collapse.
	-103	2. You know what? You guys can talk about any government you would like but when you bring in a large group of people that is racism. That is not tolerated on this sub. Neither is Xenophobia, just because someone does something different from you does not make it wrong.
	-97	3. Thought you guys claiming he was kidnapped by the big bad wolf CCP. Turns out he was forced quarantined so he doesn't go out spreading the virus like a dumbass. Like what is happening in America.
	-93	4. Is there a real problem? China preferred money and they're getting what they want. Donation is giving away something for free without any requirement or expectation of reciprocity.
	-91	5. Western countries got hit hard because the gov disregard Chinese government's advice and refused to ask people to wear masks, not because they got news late.
/r/Coronavirus	-467	1. the peak has passed, open the country
	-452	2. The US response to the Corona Virus has been far better and more timely than most countries. The government will never look good no matter the response. Death never looks good. When the government did respond, they were attacked by the media for acting to harsh...
	-431	3. The Coronavirus is very much under control in the USA. We are in contact with everyone and all relevant countries. CDC and World Health have been working hard and very smart. Stock Market starting to look very good to me!
	-357	4. I mean it feels like 99% of the world downplayed the virus. The see-I-told-ya-so's can apply to lots of people. I don't see the point of it now.
	-324	5. A mix of medias have both downplayed the virus at some point. There is no use in making this into a political fight.. people make judgments on limited information and then shit hits the fan, it happens to everyone, there is no use in pointing fingers and wishing bad on anyone.

Table 3: Most downvoted comments in /r/China_flu and /r/Coronavirus.

from moderators or bots to focus on community members' responses during this pandemic.

We measure the language distance between /r/China_flu users and /r/Coronavirus users in two contexts. First, we measure their language distance before joining these two communities to characterize their differences before COVID-19. To do that, we aggregate the entirety of each user's comments on the Reddit platform one month before starting in /r/China_flu or /r/Coronavirus (all the communities). Second, we aggregate each user's comments in /r/China_flu and /r/Coronavirus to measure the language distance in these two communities.

Figure 6 summarizes the monthly language distance between users who started at /r/China_flu or /r/Coronavirus, as measured by the Jensen-Shannon Divergence. The language used by new users who started at /r/China_flu and /r/Coronavirus becomes more and more different, suggesting a widening separation between these two communities. Moreover, the language distance between /r/China_flu users and /r/Coronavirus users one month before joining these two communities also increases. This observation indicates that these two communities attract users who were already different before joining COVID-related discussions.

A closer look shows that the language distances of these two measures were low in January and February, especially in /r/China_flu and /r/Coronavirus. This suggests that the discussions at these two communities were similar, and users who joined these two communities at that time were also similar. Peak similarity appeared in February when the official announcement was made (Figure 2). The strict moderating rules implemented in /r/Coronavirus and allowing more relaxed discussions in /r/China_flu may have played roles in this shift. We further note a dip in July for both lines. One explanation could be due to the second wave of infected cases
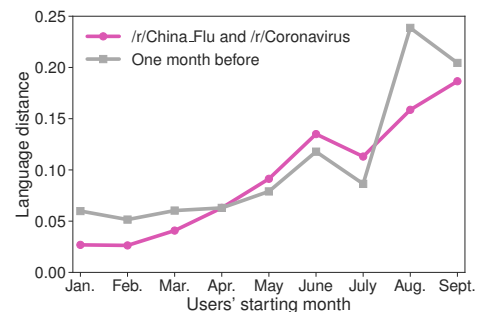

Figure 6: The language distance between /r/China_flu and /r/Coronavirus users who started in the same month. The gray line measures the language distance of users' comments on the entire Reddit platform one month before joining /r/China_flu or /r/Coronavirus. The pink line measures the language distance of users' entire comments (from the starting month to September) in /r/China_flu or /r/Coronavirus.

in the U.S. and other countries, more new users are looking for COVID-19 communities to follow, and they tend to be more similar (We also observe a peak of activity in July in Figure 1).

In Which Community will a User Start?

Another way to understand the differences between users who chose to start at /r/China_flu or /r/Coronavirus is to measure their predictability. We formulate a prediction task: given the users' activity on the Reddit platform one month before starting at /r/China_flu or /r/Coronavirus, can we predict which community they will choose? For each month, we
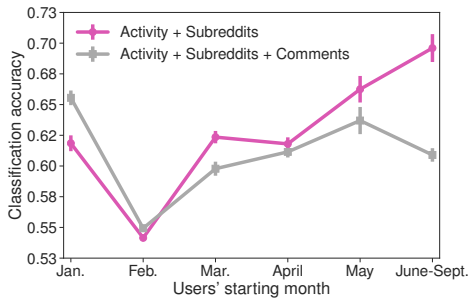
Figure 7: Monthly accuracy of determining which community (/r/China_flu or /r/Coronavirus) a new user will start at. The pink line shows the monthly accuracy using Activity and Subreddits features, while the gray line shows the accuracy using Activity, Subreddits, and Comments features.

randomly sample 1000 users who started at /r/China_flu or /r/Coronavirus, respectively. As there are not enough users from June to September in /r/China_flu, we combine users in these months and randomly sample 1000 users from each community. This task is thus a balanced prediction task, where the majority baseline accuracy is 50%. We extract the following features from users' comments one month before joining /r/China_flu or /r/Coronavirus:

- Activity: these features include the number of comments made by the user, average comment length, and the entropy of all subreddits where the user has been active.

- Subreddits: the bag of subreddits (similar to the bag of words (BOW)), based on the all subreddits where the user has been active one month before joining /r/China_flu or /r/Coronavirus. We remove subreddits that appeared less than five times in the training data.

- Comments: the bag-of-words (BOW) from users' comments in all the communities, one month before joining /r/China_flu or /r/Coronavirus.

To assess the prediction performance, we measure accuracy using stratified five-fold nested cross-validation with a standard ℓ_2-regularized logistic regression classifier. All the features are scaled to [0, 1] based on the training data.

Figure 7 presents the monthly accuracy of predicting which community (/r/China_flu or /r/Coronavirus) a new user is going to start at based on her activity on the Reddit platform one-month before. First, the accuracy for each month is much higher than 50%. It shows that users' previous activity history (also known as genealogy (Tan 2018)) matters, and using historical activities can outperform the random baseline by a significant margin. Second, we observe a similar trend as the language distance: The accuracy hits bottom in February, indicating that the users who joined /r/China_flu or /r/Coronavirus are hard to distinguish in that month. After that month, the accuracy goes up, indicating an increment of distinction. Interestingly, the best accuracies for months after March are achieved by only using Activity and Subreddits features. Adding features from comments in the model is not helpful. It is likely due to the sparsity of

words in our relatively small dataset for the prediction task, as the comments can come from many different communities. This observation further shows that genealogical relations may provide more robust signals than textual content.

Subreddit importance analysis. Figure 7 shows that the best prediction accuracy is achieved by using Activity and Subreddit features. A close examination of the coefficients can reveal the subreddits that are strong indicators. Here we rank subreddits month by month based on their coefficients in the prediction model, and the results are summarized in Table 4. First, /r/china is in the top-10 subreddits for /r/China_flu across all the months. This is a community for discussing China-related topics, and most of its posts and comments are anti-China.[4] It indicates that people who dislike China or China-related news are more likely to join /r/China_flu. Subreddits that focus on places near China geographically are also strong indicators (e.g., /r/singapore, /r/hongkong, and /r/taiwan). Second, some well-known extreme subreddits, such as /r/conspiracy and /r/wuhan_flu, have been strong predictors for /r/China_flu users since March. Interestingly, we do not see them in January or February. We also do not see them in subreddits that predict /r/Coronavirus users.

User Movement between /r/China_flu and /r/Coronavirus

We have identified clear signals of whether a new user chooses to start in /r/China_flu or /r/Coronavirus in their genealogy, but it remains unclear how stable this initial membership is. Hence, we investigate user movement between these two communities.

How do Users Split their Activity in /r/China_flu and /r/Coronavirus?

We first analyze how users split their activity in /r/China_flu and /r/Coronavirus, and how it differs over time. To illustrate this, we show the monthly activity split for users who started in different months in /r/China_flu or /r/Coronavirus, respectively. Each curve in Figure 8 is obtained by aggregating users who started in the same month in /r/China_flu or /r/Coronavirus and then calculating their proportion of activity in /r/China_flu among these two communities over time.

/r/China_flu users who started in January (the darkest red) have a sharp drop in the following months, from 75% in February to below 30% since June (Figure 8a). This indicates that the majority of these users' activity has shifted from /r/China_flu to /r/Coronavirus during this period. We observe a similar trend for /r/China_flu users who started in February and March. However, there is a clear separation between /r/China_flu users who started in the first three months and the remaining months. /r/China_flu users who started after March tend to maintain a high level of activity in /r/China_flu (above 50%) over time.

In comparison (Figure 8b), /r/Coronavirus users who started in January and February sustained a certain level of activity in /r/China_flu (around 15% for users who started

---

[4]https://www.reddit.com/r/china/.

	Jan.	Feb.	Mar.	Apr.	May.	June-Sept.
/r/China_flu	/r/collapse	/r/singapore	/r/wuhan_flu	/r/conspiracy	/r/china	/r/conspiracy
	/r/worldnews	/r/china	/r/conspiracy	/r/cringetopia	/r/conspiracy	/r/wuhan_flu
	/r/singapore	/r/collapse	/r/preppers	/r/wuhan_flu	/r/politicalcompassmemes	/r/conservative
	/r/preppers	/r/taiwan	/r/collapse	/r/china	/r/gifs	/r/covid19positive
	/r/news	/r/wallstreetbets	/r/wallstreetbets	/r/worldpolitics	/r/animalsonreddit	/r/china
	/r/china	/r/canada	/r/singapore	/r/politicalcompassmemes	/r/whatcouldgowrong	/r/thelastofus2
	/r/hongkong	/r/preppers	/r/entertainment	/r/historymemes	/r/libertarianmeme	/r/okbuddyretard
	/r/wallstreetbets	/r/italy	/r/the_donald	/r/hongkong	/r/gadgets	/r/art
	/r/nba	/r/presidentialracememes	/r/korea	/r/europe	/r/diwhy	/r/mildlyinteresting
	/r/canada	/r/bestoflegaladvice	/r/china	/r/iamatotalpieceofshit	/r/wuhan_flu	/r/anime_titties
/r/Coronavirus	/r/roastme	/r/breadtube	/r/nfl	/r/insaneparents	/r/askreddit	/r/politics
	/r/the_donald	/r/latestagecapitalism	/r/hockey	/r/askreddit	/r/politics	/r/nba
	/r/cringetopia	/r/epstein	/r/rocketleague	/r/politics	/r/news	/r/askreddit
	/r/mgtow	/r/patriots	/r/indieheads	/r/technology	/r/facepalm	/r/politicalhumor
	/r/natureisfuckinglit	/r/90dayfiance	/r/parenting	/r/nba	/r/whitepeopletwitter	/r/winstupidprizes
	/r/rickandmorty	/r/cfb	/r/nba	/r/amitheasshole	/r/nfl	/r/news
	/r/science	/r/reactiongifs	/r/soccer	/r/sandersforpresident	/r/xboxone	/r/oddlysatisfying
	/r/bad_cop_no_donut	/r/bodybuilding	/r/modernwarfare	/r/humansbeingbros	/r/animalcrossing	/r/aww
	/r/pan_media	/r/choosingbeggars	/r/wtf	/r/boston	/r/therewasanattempt	/r/humansbeingbros
	/r/meme	/r/woooosh	/r/bikinibottomtwitter	/r/science	/r/dundermifflin	/r/nintendoswitch

Table 4: The monthly top-10 subreddits with the highest coefficients for predicting users' starting community (/r/China_flu or /r/Coronavirus).

(a) /r/China_flu users' monthly activity.
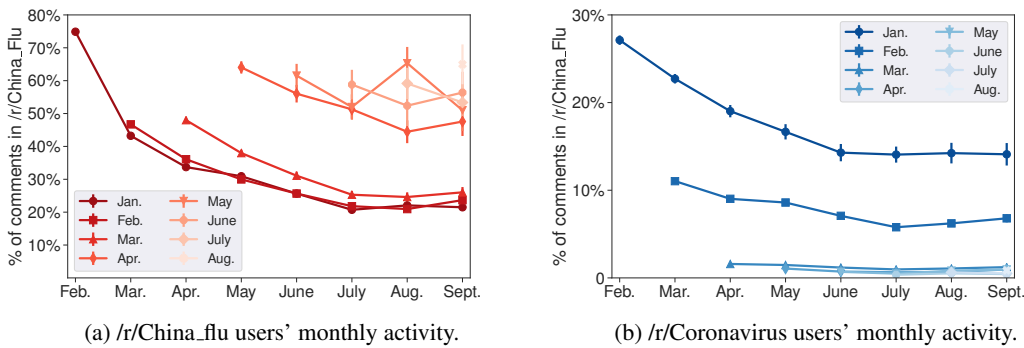

(b) /r/Coronavirus users' monthly activity.

Figure 8: The monthly proportion of activity in /r/China_flu for users who started in different months in /r/China_flu (Figure 8a) and /r/Coronavirus (Figure 8b), respectively. Each line represents users who started in a month from January to August.

in January and around 8% for users who started in February). However, users who joined later have little activity in /r/China_flu. These comparisons echo the clear separations between the two communities we observed previously, especially from April to September.

A Movement Analysis Framework

Given that early /r/China_flu users commented less in /r/China_flu and more in /r/Coronavirus over time, we further investigate users who decided to move from /r/China_flu to /r/Coronavirus and the potential factors that may influence their choices. To understand this behavior, an important question is to determine users' memberships in these two communities. One way is to define users who only commented in /r/China_flu as /r/China_flu members and apply the same rule for detecting /r/Coronavirus members (note that the definition differs from the previous section, which bases on a user's starting community. We thus use "mem-

bers" to signify the difference). However, as shown in Figure 3, a large proportion of /r/China_flu users also commented in /r/Coronavirus, especially in the early months (February to April). To provide a more complete picture, we develop a movement analysis framework and use an activity ratio threshold $k\%$ to determine a user's membership in these two communities.

A user is a member of /r/China_flu in a time window if more than $k\%$ of their comments *in these two communities* are posted in /r/China_flu. The same rule is used for determining /r/Coronavirus members. For example, when $k\% = 60\%$, users who made more than 60% of their comments in /r/China_flu are /r/China_flu members. Users who made more than 60% of their comments in /r/Coronavirus are /r/Coronavirus members. Users who do not fall into these two categories are discarded, as their memberships are deemed uncertain. In particular, when $k\% = 100\%$, users who only commented in /r/China_flu are /r/China_flu mem-
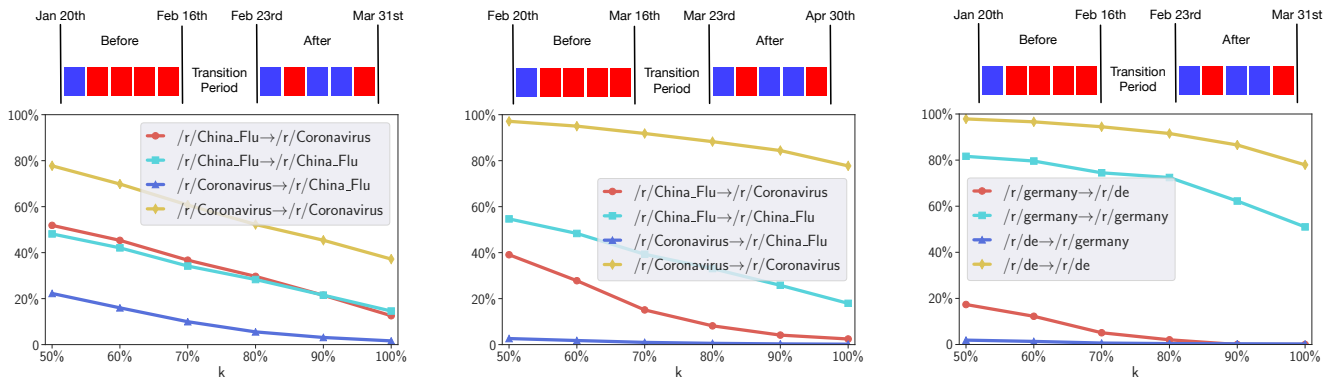
(a) Feb. 16th to Feb. 23rd as the transition window for /r/China_flu and /r/Coronavirus.

(b) Mar. 16th to Mar. 23rd as the transition window for /r/China_flu and /r/Coronavirus.

(c) Feb. 16th to Feb. 23rd as the transition window for /r/germany and /r/de.

Figure 9: Figure 9a shows the proportion of users who moved between /r/China_flu and /r/Coronavirus before and after the transition window (from February 16th to February 23rd). Figure 9b applies this movement framework to the transition window one month later (from March 16th to March 23rd). For further comparison, Figure 9c applies this movement framework to another two highly-related communities, /r/germany and /r/de (transition window from February 16th to February 23rd). The blue and red squares are only for illustrative purposes. They represent a user's activity in /r/China_flu and /r/Coronavirus, respectively.

bers, and users who only commented in /r/Coronavirus are /r/Coronavirus members.

To examine users who moved from one community to the other, we need to set a transition window and compare their memberships before and after that transition window. As discussed in Figure 2, on February 17th, the Reddit platform made /r/Coronavirus the official community, and more relaxed discussions are allowed in /r/China_flu. This decision may have motivated users to move between these two communities. Thus, we set the week of February 17th as a transition window (February 16th to February 22nd) and measure users' memberships before and after this transition window, illustrated in Figure 9a. The users' commenting activity before February 16th is used for determining their memberships before the transition window, and their activity after February 22nd but before March 31st is used for determining memberships after the transition window. We also require users to have at least five comments in these two communities before the transition and an additional five comments after the transition to be included in this study. There are in total 11,238 such users. If a user's memberships before and after the transition are different, then it means that she moved from one community to the other.

Figure 9a shows the movement analysis results. With varying $k\%$, the number of /r/China_flu users who moved to /r/Coronavirus (red line) during the transition window is similar to the number of users who stayed (cyan line). It shows that around 50% of /r/China_flu members have moved to /r/Coronavirus during this transition window, robust to the definition of membership ($k\%$). Specifically, 13% of users made a dramatic shift, from only commenting in /r/China_flu to only commenting in /r/Coronavirus ($k\% = 100$). The ratio of users who left is always around 50% regardless of k. In comparison, a small proportion of /r/Coronavirus members moved to /r/China_flu (blue line). Most of them stayed

in /r/Coronavirus (yellow line). When we choose $k = 100$, the number of users who went from 100% /r/Coronavirus to 100% /r/China_flu is close to zero.

To further examine such movement's robustness, we defer the transition window a month later and conduct the same analysis to find users who were moving between /r/China_flu and /r/Coronavirus in March (Figure 9b). The results demonstrate that few users have moved between these two communities. We also attempt to shift the transition window to the following months. The analyses show that the proportion of users who moved between these two communities in those months are even lower than in March.

In addition, we apply this movement analysis framework to another pair of highly-related communities, /r/germany and /r/de (Figure 9c). Both of these two communities are for discussions about Germany. The primary language in /r/germany is English, while the primary language in /r/de is German. When we set the transition window from February 16th to February 23rd, very few users moved from one to the other (similar to Figure 9b). This trend is consistent for /r/germany and /r/de when we defer the transition window to later months.

In summary, the movement analysis results in Figure 9 suggest that there is a significant number of /r/China_flu users moved to /r/Coronavirus in February. This movement may connect with Reddit's decision to make /r/Coronavirus the official community for COVID-19 and allow more relaxed discussion in /r/China_flu. We do not see movement at this frequency in the following months between these two communities or other highly-related communities. Moreover, compared with /r/China_flu users, many fewer /r/Coronavirus users moved to /r/China_flu in February.

Predicting the Movement from /r/China_flu to /r/Coronavirus

/r/China_flu is in flux in February: about half of /r/China_flu members moved to /r/Coronavirus. We formulate another prediction task to compare users who chose to stay or leave: given the members of /r/China_flu and their activity prior to the transition window in February, can we predict who will stay or leave? Here we define users who were /r/China_flu members before and after the transition window as "Stay" users, and users who moved from /r/China_flu to /r/Coronavirus as "Leave" users. To ensure a balanced prediction task, for each threshold $k\%$, we randomly sample the same number of Stay users from Leave users (except for $k\% = 100\%$ where we do it reversely as there are more Stay users than Leave users). This is thus a balanced prediction task, and the random baseline is 50%. The task allows us to understand the dynamic membership in /r/China_flu and its formation in the context of highly-related communities.

For this prediction task, we extract the following features from users' comments before the transition window (from January 20th to February 15th):

- Activity: these features include the number of comments made by the user, average comment length, mention of "china" or "chinese" in /r/China_flu. We also add the entropy of all subreddits where the user has been active.

- Subreddits: the bag-of-subreddits (similar to bag of words) based on all subreddits where the user has been active. /r/China_flu and /r/Coronavirus are excluded.

- Comments: the bag-of-words (BOW) from aggregated comments in all the communities, excluding /r/China_flu and /r/Coronavirus.

- /r/China_flu Comments: the bag-of-words (BOW) from aggregated comments in /r/China_flu.

For each $k\%$ threshold, we measure accuracy using stratified five-fold nested cross-validation with a standard ℓ_2-regularized logistic regression classifier, which is the same one we used in the previous section. All the features are scaled to [0, 1] based on the training data.

The prediction results are summarized in Figure 10. First, with a growing $k\%$, it is easier to predict users who will stay or leave. For example, when $k\% = 100\%$, the prediction accuracy is as high as 80%. We see a near monotonic trend between $k\%$ and the prediction accuracy for each feature set. It suggests that for highly active users in /r/China_flu, the users who chose to stay or leave are substantially different and easy to distinguish. Second, the subreddits at which users have also been active are the strongest indicators of users' movement. Only using the subreddits as the feature (Subreddit-only) can achieve comparable accuracies to using all the features (All), and outperforms other types of features by a significant margin.

Inspired by this observation, we analyze the coefficients of subreddits in the model. Table 5 summarizes the top-10 subreddits with the highest coefficients for predicting Stay and Leave users when choosing $k\% = 100\%$. First, the subreddit with the highest coefficient for Stay users is /r/the_donald. This community was dedicated to
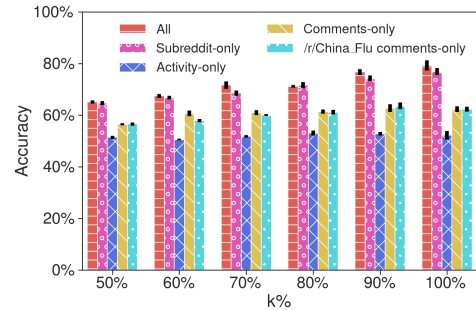

Figure 10: The performance of predicting users who would leave /r/China_flu for /r/Coronavirus during the transition window (February 16th to 23rd) with different threshold $k\%$.

Predicting Stay	Coefs	Predicting Leave	Coefs
/r/the_donald	-1.4	/r/sandersforpresident	1.1
/r/conspiracy	-1.4	/r/dataisbeautiful	1.0
/r/preppers	-1.3	/r/covid19	1.0
/r/china	-0.9	/r/books	0.9
/r/wellthatsucks	-0.9	/r/askscience	0.9
/r/collapse	-0.7	/r/idiotsincars	0.8
/r/facepalm	-0.7	/r/nba	0.8
/r/showerthoughts	-0.7	/r/leagueoflegends	0.7
/r/funny	-0.6	/r/coolguides	0.7
/r/explainlikeimfive	-0.6	/r/worldpolitics	0.6

Table 5: The top-10 subreddits with the highest coefficients for predicting Stay or Leave users ($k\% = 100\%$).

supporting the U.S. President Donald Trump and was banned in March 2020 due to racist content and hateful speech (Newton 2020). It suggests that a large proportion of Stay users are Donald Trump supporters. In contrast, /r/sandersforpresident, the community for Bernie Sanders has the highest coefficient for predicting Leave users, indicating that many Leave users are Bernie Sanders supporters. Besides, /r/conspiracy, /r/preppers, and /r/collapse are strong signs for users to stay in /r/China_flu. This is also reflected in founders' activity (Table 2). In contrast, some science-oriented subreddits, such as /r/askscience, /r/dataisbeautiful, and /r/covid19, appear on the table's right column. It indicates that science enthusiasts were leaving /r/China_flu for /r/Coronavirus.

Concluding Discussion

In this work, we study the emergence of two highly-related communities on Reddit, /r/China_flu and /r/Coronavirus, during the COVID-19 pandemic. We take a user-centered perspective and characterize user trajectories in these two communities from January to the end of September. We show that users who joined these two communities were similar in the first three months. After that, as the pandemic continues to unfold, their differences steadily increase. We further examine the user movement between these two communities. We find that users who started at /r/China_flu from January to March reduced their activity at /r/China_flu later,

while users who started in the following months remained highly "loyal". A newly designed movement analysis framework reveals that around 50% of /r/China_flu members moved to /r/Coronavirus in February, when /r/Coronavirus became the official COVID-19 community on Reddit. This movement turns out to be highly predictable based on other subreddits users were formerly active in.

Limitations. The findings in our work are subject to several limitations. First, the causal reasons for users to move from /r/China_flu to /r/Coronavirus are not established. Potential reasons include 1) making /r/Coronavirus the official community on the platform; 2) the strict moderation rules implemented in /r/Coronavirus, which may also attract users who seek serious discussions about COVID-19; 3) The anti-china sentiment of "China_Flu", which may drive people who have complementary views about China to leave. Our study is limited to revealing correlations. Second, even though our findings reveal many similarities between these two communities at the beginning stage, there are potentially other ways in which initial members of two communities differ that are not easy to capture from our data. Surveys or interviews with early members or moderators may help us better understand this development. Moreover, most of the data has been collected during the COVID-19 pandemic, a unique period in human history. The results reported in this study are highly connected to this unique context. For example, the speed of these two communities' explosion at the beginning is unusual and may not apply to other communities (Solon and Glaser 2020). Although the generalization of insights from this work to highly-related communities in other contexts requires further investigation, we believe that it is valuable to bridge the literature on highly-related communities and crisis informatics in this unique context.

Implications and future directions. Our observation that users in /r/China_flu and /r/Coronavirus were similar at the early stage but diverge later on provides implications for community organizers. This suggests that when highly-related communities emerge, they may follow different directions and each develops its own identity. In our example, even though /r/China_flu and /r/Coronavirus are both for general discussions about COVID-19, they attract people with different interests, and the separation grows over time. Understanding the mechanism behind this phenomenon could be potentially useful for designing online spaces.

Our work also demonstrates that online communities do not only exist in the virtual world. User activity in online communities can be heavily embedded in the offline context. COVID-related subreddits provide a unique opportunity for understanding the connections between the online and offline worlds in a crisis. These communities only exist as a result of the COVID-19 pandemic. Many activity changes identified in their users are highly correlated with how the epidemic unfolds in the real world. Such observations emphasize the necessity of connecting online and offline data resources to explain online communities' dynamics and their relationships with on-going offline events. An exciting future direction is to further understand the changes in online activities in the context of fine-grained offline events.

References

Alexa. 2020. The top 500 sites on the web. shorturl.at/mpsyI. Accessed 2021-01-30.

Althoff, T.; Clark, K.; and Leskovec, J. 2016. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *In TACL* 463–476.

Atkinson, D.; Srinivasan, K. B.; and Tan, C. 2019. What Gets Echoed? Understanding the "Pointers" in Explanations of Persuasive Arguments. In *EMNLP*, 2904–2914.

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The pushshift reddit dataset. In *ICWSM*, volume 14, 830–839.

Budhwani, H.; and Sun, R. 2020. Creating COVID-19 Stigma by Referencing the Novel Coronavirus as the "Chinese virus" on Twitter: Quantitative Analysis of Social Media Data. *Journal of Medical Internet Research* 22(5): e19301.

Chen, L.; Lyu, H.; Yang, T.; Wang, Y.; and Luo, J. 2020. In the eyes of the beholder: Sentiment and topic analyses on social media use of neutral and controversial terms for covid-19. *arXiv preprint arXiv:2004.10225* .

Davies, S. 2020. COVID-19 and the Collapse of Complex Societies. shorturl.at/xDU14. Accessed 2021-01-30.

Dong, E.; Du, H.; and Gardner, L. 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* 20(5): 533–534.

Hagen, L.; Keller, T.; Neely, S.; DePaula, N.; and Robert-Cooperman, C. 2018. Crisis communications in the age of social media: A network analysis of Zika-related tweets. *Social Science Computer Review* 36(5): 523–541.

Hessel, J.; Tan, C.; and Lee, L. 2016. Science, askscience, and badscience: On the coexistence of highly related communities. In *ICWSM*.

Joubin, A. A. 2020. Anti-Asian Racism during COVID-19 Pandemic, GW Today, April 20, 2020. *GW Today* .

Kraut, R. E.; and Resnick, P. 2012. *Building successful online communities: Evidence-based social design*. Mit Press.

Lachlan, K. A.; Spence, P. R.; Lin, X.; Najarian, K.; and Del Greco, M. 2016. Social media and crisis management: CERC, search strategies, and Twitter content. *Computers in Human Behavior* 54: 647–652.

Lee, B. Y. 2020. Trump Once Again Calls Covid-19 Coronavirus The 'Kung Flu'. https://www.forbes.com/sites/brucelee/2020/06/24/trump-once-again-calls-covid-19-coronavirus-the-kung-flu/. Accessed 2021-01-30.

Lee, D. 2017. NPR, CNN, Other News Sites Hit The Mute Button On Readers. shorturl.at/bdgkC. Accessed 2021-01-30.

Lyu, H.; Chen, L.; Wang, Y.; and Luo, J. 2020. Sense and sensibility: Characterizing social media users regarding the use of controversial terms for covid-19. *IEEE Transactions on Big Data* .

Maas, P.; Iyer, S.; Gros, A.; Park, W.; McGorman, L.; Nayak, C.; and Dow, P. A. 2019. Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery. In *ISCRAM*.

Manning, C. D.; and Schütze, H. 1999. *Foundations of statistical natural language processing*. MIT press.

Marcin, T. 2020. There's a community on Reddit where COVID-positive people share experiences. shorturl.at/lyFV6. Accessed 2021-01-30.

Mitchell, A.; Oliplant, J. B.; and Shearer, E. 2020. About Seven-in-Ten U.S. Adults Say They Need to Take Breaks From COVID-19 News. shorturl.at/buTZ0. Accessed 2021-01-30.

Muller, R. T. 2020. COVID-19 Brings a Pandemic of Conspiracy Theories. shorturl.at/aqB68. Accessed 2021-01-30.

Newton, C. 2020. Reddit bans r/The_Donald and r/ChapoTrapHouse as part of a major expansion of its rules. shorturl.at/uC057. Accessed 2021-01-30.

Palen, L.; and Anderson, K. M. 2016. Crisis informatics—New data for extraordinary times. *Science* 353(6296): 224–225.

Palen, L.; Vieweg, S.; Liu, S. B.; and Hughes, A. L. 2009. Crisis in a networked world: Features of computer-mediated communication in the April 16, 2007, Virginia Tech event. *Social Science Computer Review* 27(4): 467–480.

Palen, L.; Vieweg, S.; Sutton, J.; Liu, S. B.; and Hughes, A. 2007. Crisis informatics: Studying crisis in a networked world. In *Proceedings of the Third International Conference on E-Social Science*, 7–9. Citeseer.

Pei, X.; and Mehta, D. 2020. # Coronavirus or# Chinesevirus?!: Understanding the negative sentiment reflected in Tweets with racist hashtags across the development of COVID-19. *arXiv preprint arXiv:2005.08224* .

Rehurek, R. 2010. Software Framework for Topic Modelling with Large Corpora. In *LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. Valletta, Malta: ELRA.

Rene, P. L. 2016. The influence of social media on emergency management. *PA Times* .

Reuter, C.; Hughes, A. L.; and Kaufhold, M.-A. 2018. Social media in crisis management: An evaluation and analysis of crisis informatics research. *International Journal of Human–Computer Interaction* 34(4): 280–294.

Reuter, C.; and Kaufhold, M.-A. 2018. Fifteen years of social media in emergencies: a retrospective review and future directions for crisis informatics. *Journal of Contingencies and Crisis Management* 26(1): 41–57.

Schild, L.; Ling, C.; Blackburn, J.; Stringhini, G.; Zhang, Y.; and Zannettou, S. 2020. "Go eat a bat, Chang!": An Early Look on the Emergence of Sinophobic Behavior on Web Communities in the Face of COVID-19. *arXiv preprint arXiv:2004.04046* .

SCHWEDEL, H. 2020. It's Kind of Catchy Novel coronavirus, COVID, 'rona, and how we settle on names for deadly outbreaks. shorturl.at/jAZ48. Accessed 2021-01-30.

Solon, O.; and Glaser, A. 2020. How Reddit's coronavirus community became a destination. shorturl.at/anvOX. Accessed 2021-01-30.

Stewart, M. C.; and Wilson, B. G. 2016. The dynamic role of social media during Hurricane# Sandy: An introduction of the STREMII model to weather the storm of the crisis lifecycle. *Computers in Human Behavior* 54: 639–646.

Sutton, J.; Palen, L.; and Shklovski, I. 2008. Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires. *ISCRAM* .

Tan, C. 2018. Tracing community genealogy: how new communities emerge from the old. In *ICWSM*, 395–404.

Tan, C.; Friggeri, A.; and Adamic, L. 2016. Lost in propagation? Unfolding news cycles from the source. In *ICWSM*.

Tan, C.; and Lee, L. 2015. All who wander: On the prevalence and characteristics of multi-community engagement. In *WWW*, 1056–1066.

Times, T. N. Y. 2020. How the Coronavirus Pandemic Unfolded: a Timeline. shorturl.at/cgrEF. Accessed 2021-01-30.

United Nations Office. 2018. UN 20-year review: earthquakes and tsunamis kill more people while climate change is driving up economic losses. shorturl.at/joCE0. Accessed 2021-01-30.

Van Bavel, J. J.; Baicker, K.; Boggio, P. S.; Capraro, V.; Cichocka, A.; Cikara, M.; Crockett, M. J.; Crum, A. J.; Douglas, K. M.; Druckman, J. N.; et al. 2020. Using social and behavioural science to support COVID-19 pandemic response. *Nature human behaviour* 4(5): 460–471.

Vieweg, S.; Hughes, A. L.; Starbird, K.; and Palen, L. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *CHI*, 1079–1088.

Waller, I.; and Anderson, A. 2019. Generalists and specialists: Using community embeddings to quantify activity diversity in online platforms. In *WWW*, 1954–1964.

Whalen, A. 2016. Coronavirus Reddit Page Becomes Fastest Growing subreddit with almost 400,000 new users in a week. shorturl.at/lqHIX. Accessed 2021-01-30.

Winkie, L. 2020. Grocery workers are using Facebook and Reddit to swap stories and information. shorturl.at/hoELV. Accessed 2021-01-30.

Yin, Y.; Gao, J.; Jones, B. F.; and Wang, D. 2021. Coevolution of policy and science during the pandemic. *Science* 371(6525): 128–130.

Zhang, J. S.; Tan, C.; and Lv, Q. 2018. "This is why we play" Characterizing Online Fan Communities of the NBA Teams. *CSCW* 2(CSCW): 1–25.

Zhang, J. S.; Tan, C.; and Lv, Q. 2019. Intergroup Contact in the Wild: Characterizing Language Differences between Intergroup and Single-group Members in NBA-related Discussion Forums. *Proceedings of the ACM on Human-Computer Interaction* 3(CSCW): 1–35.

Zhu, H.; Chen, J.; Matthews, T.; Pal, A.; Badenes, H.; and Kraut, R. E. 2014. Selecting an effective niche: an ecological view of the success of online communities. In *CHI*.

Zhu, H.; Kraut, R. E.; and Kittur, A. 2013. Effectiveness of shared leadership in Wikipedia. *Human factors* 55(6): 1021–1043.

Zhu, H.; Kraut, R. E.; and Kittur, A. 2014. The impact of membership overlap on the survival of online communities. In *CHI*.