Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Information propagation on cyber, relational and physical spaces about covid-19 vaccine: Using social media and splatial framework

Fuzhen Yin a,*, Andrew Crooks b, Li Yin a

a Department of Urban and Regional Planning, University at Buffalo, Buffalo, NY, USA
b Department of Geography, University at Buffalo, Buffalo, NY, USA

ARTICLE INFO

Keywords: Covid-19 Vaccination Social media Social network analysis Community detection Urban informatics

ABSTRACT

With the advent of social media, human dynamics studied in purely physical space have been extended to that of a cyber and relational context. However, connections and interactions between these hybrid spaces have not been sufficiently investigated. The “space-place (Splatial)” framework proposed in recent years allows capturing human activities in the hybrid of spaces. This study applies the Splatial framework to examine the information propagation between cyber, relational, and physical spaces through a case study of Covid-19 vaccine debates in New York State (NYS). Whereby the physical space represents the regional boundaries and locations of social media (i.e., Twitter) users in NYS, the relational space indicates the social networks of these NYS users, and the cyber space captures the larger conversational context of the vaccination debate. Our results suggest that the Covid-19 vaccine debate is not polarized across all three spaces as compared to that of other vaccines. However, the rate of users with a pro-vaccine stance decreases from physical to relational and cyber spaces. We also found that while users from different spaces interact with each other, they also engage in local communications with users from the same region or same space, and distance-based and boundary-confined clusters exist in cyber and relational space communities. These results based on the Splatial framework not only shed light on the vaccination debates but also help to define and elucidate the relationships between the three spaces. The intense interactions between spaces suggest incorporating people’s relational network and cyber presence in physical place-making.

1. Introduction

With the advent of social media, human dynamics studied in purely physical space have been extended to that of a cyber and relational context. This shift has been further intensified by the Covid-19 pandemic. During the early stages of the pandemic, while preventative measures such as social distancing and lockdowns attempted to reduce physical activities and interactions (Li, Zhao, He, Mansourian, & Axhausen, 2021), they also increased people’s use of social media to obtain health information and communicate with their friends and family while being physically apart (Gonzalez-Padilla & Tortolero-Blanco, 2020; Saud, Mashud, & Ida, 2020). For example, during the pandemic, the hashtags “#covid” and “#coronavirus” were heavily mentioned on social media platforms (Chen, Lerman, & Ferrara, 2020). Furthermore, healthcare agencies and government officials have also leveraged social media platforms to mitigate the spread of misinformation about Covid-19, especially that related to the vaccines (Lovari, 2020); for example, the New York State’s #GetTheVaxFacts campaign (New York State, 2021b) and Connecticut’s Long Term Care Ombudsman Program (CDC, 2021).

Despite the benefits of social media including its ability to efficiently reach a large population during a crisis (Saud et al., 2020), there have also been considerable public health concerns raised by the spread of online anti-vaccine messaging (Puri, Coomes, Haghbayan, & Gunaratne, 2020). Studies have found polarized vaccination debates on social media (Yuan, Schuchard, & Crooks, 2019; Schmidt, Zollo, Scala, Betsch, & Quattrociocchi, 2018), suggesting the lack of interactions between pro and anti-vaccine users. Such polarization has the potential to strengthen the ideological isolation of anti-vaccine users, fuel vaccine hesitancy, and potentially lead to a dramatic increase in disease outbreak probabilities (Salathe & Bonhoeffer, 2008). However, the polarized vaccination debates that were found before the pandemic might not apply to the current Covid-19 vaccine due to their new characteristics including the rising trend in positive sentiment towards them (Hu et al., 2021) and
health organizations’ active intervention to debunk vaccine misinformation on social media (New York State, 2021b; CDC, 2021). Therefore, it is important to gain a better understanding of the communication patterns of pro and anti-vaccine users in light of the current Covid-19 pandemic and specifically on social media in terms of whether they are polarized or not.

Coinciding with the lack of studies on Covid-19 vaccine polarization is that of studies on how information or different opinions (i.e., pro and anti-Covid-19 vaccine) are propagated across spaces. While some studies have attempted to establish connections between cyber and physical spaces. For example, through analyzing the spread of civil unrest during the Arab Spring across North Africa and the Middle East, AlSayyad and Guvenc (2015) found the reciprocal interactions between the social movements in the physical world and the media coverage on digital platforms. In another study, Zhao, Huang, Huang, Liu, and Lai (2014) found a correlation between online surfing behaviors and offline mobility patterns which suggested that human dynamics happening in one space could be used to predict those in another. While Gunaratne, Coones, and Haghbayan (2019) and Ahmed, Quinn, Hancock, Free, and Jamison (2018) argued that the pro or anti-vaccine messages trending in cyber space have the potential to link with the vaccine uptake or disease outbreak in the physical space. However, rarely have studies investigated how information and opinions are propagated from offline to online and from local to global.

To fill the gaps mentioned above, this study employs the recently proposed “space-place (Splatial)” framework to investigate the Covid-19 vaccine discourse on social media. The novelty of this Splatial framework lies in the fact that it provides a new way to capture and analyze human dynamics and interactions across multiple spaces (Shaw & Sui, 2020; Sui & Shaw, 2021). In addition to the cyber and physical spaces, this Splatial framework also adds the concept of relational space that infers networks between objects based on their relations and has the potential to bridge the cyber and physical gap (Croitoru, Wayant, Crooks, Radzikowski, & Stefanidis, 2015). However, this concept of relational space was relatively new and theoretical, and this concept needs further elaboration based on real-world events.

Using the Splatial framework, this paper investigates the propagation of different opinions (i.e., pro and anti-vaccine) between three spaces: cyber, relational, and physical spaces through a case study on the Covid-19 vaccine discussion in New York State (NYS) on Twitter. We aim to answer the following two research questions: (1) Are the pro and anti-Covid-19 vaccine social media users polarized in the three spaces? (2) How different opinions of pro and anti-vaccine were propagated between cyber, relational, and physical spaces? We employed several approaches including snowball sampling (Biernacki & Waldorf, 1981) to represent three spaces, sentiment analysis of Twitter data (Rathi, Pirolli, Pitkow, & Lukose, 1998; Chmiel, Kowalska, & Holyst, 2009; Sousa, Sarmento, & Mendes Rodrigues, 2010), using big data such as social media or smart card data to evaluate places in the physical world (e.g., Sulis, Manley, Zhong, & Batty, 2018; Long & Huang, 2019; Hamstead et al., 2018). There are only a few studies that explore and attempt to explain the mechanisms behind the cyber-physical system. For example, Croitoru et al. (2015) found that the interplay between cyber and physical spaces was transmitted through geosocial networks, networks tied to both cyber and physical spaces, and without such linkages, communities would not form and information would not propagate. Adding to this area of research, a new space-place (Splatial) framework has been recently proposed by Shaw and Sui (2020) which aims at developing a human-centered synergistic view of the space and place in the digital age. The Splatial framework allows for examining human dynamics across the cyber-physical system in a comprehensive way.

The Splatial framework includes not only the cyber and physical spaces but a hybrid of spaces including relational, relative, and mental spaces. Relational space highlights the relations between people and objects and provides a lens to view space as a network that promotes the flow of information, objects and materials (Castells, 2010). Relative space describes the corresponding locations of a moving or static object in the physical space (e.g., an autonomous vehicle detects the locations of nearby objects in its surroundings). Similar to Tuan (1977)”’s definition of space, mental space is an area in space to which people have given meanings. Mental space works with the cognitive and mental aspects of space and highlights the observers’ feelings, emotions and perceptions of space. Among these spaces, the relational space inferring networks between objects based on their relations, interactions and links has the potential to bridge the cyber and physical spaces (Croitoru et al., 2015). Therefore, this study incorporated the relational space into the center of focus for many research fields such as Urban Planning and Geography (Dodge & Kitchin, 2003). However over the last two decades, with the growing sophistication of information and communication technologies (ICTs), location-aware devices combined with machine learning and data science have greatly changed how we study and understand space and place (Shaw & Sui, 2020). The significant growth of social media has given rise to the expansion of cyber space, and has stimulated research in exploring the connections between cyber and physical spaces along with online-to-offline interaction (Sui & Shaw, 2018; Batty, 1997).

Despite its great dependency on the software and hardware in the physical space (Batty, 1997), cyber space is often considered as the opposite of physical space or a great “no place” (Herrera, 2016; Lessig, 1995). Cyber space shows metaphysical differences (i.e., place, size, distance, route) from physical space (Bryant, 2001). Furthermore, cyberspace is gradually replacing the role of physical space by altering the traditional human dynamic patterns in physical space (Sui & Shaw, 2018). Specifically, by erasing the barrier of physical distance, cyberspace enhances our ability to conduct activities in a more flexible and timely manner (Yu & Shaw, 2008) and allows us to undertake a wide range of daily activities online (Kwan, 2000). For example, with internet access, people can work and study remotely, and socialize with friends through social networking platforms along with carrying out other activities such as shopping for goods online (Line, Jain, & Lyons, 2011). Meanwhile, it is often suggested that there is a correlation between human behaviors and patterns in cyber and physical spaces (i.e., homophily) which suggests that the dynamics witnessed in one space could be used to predict the dynamics in another space (Zhao et al., 2014).

However, our current knowledge about the relationships between cyber and physical spaces and their implications for human dynamics is limited (Sui & Shaw, 2018; Sui & Sui, 2020; Porter, 2004). With much of the research being devoted to describing human mobility patterns in physical space (e.g., Gonzalez, Hidalgo, & Barabasi, 2008; Brockmann, Hufnagel, & Geisel, 2006), surfing and communication behaviors in cyber space (e.g., Huberman, Pirolli, Pitkow, & Lukose, 1998; Chmiel, Kowalska, & Holyst, 2009; Sousa, Sarmento, & Mendes Rodrigues, 2010), using big data such as social media or smart card data to evaluate places in the physical world (e.g., Sulis, Manley, Zhong, & Batty, 2018; Long & Huang, 2019; Hamstead et al., 2018).
cyber-physical system (1) to analyze the information propagation mechanism between three spaces, cyber, relational and physical spaces; (2) to enrich the Splatgal framework by providing concrete definitions of the three spaces based on the real-world event of the Covid-19 vaccine debates. This was at the expense of excluding mental and relative spaces whose focus is on the meaning of space or describing the physical world which we will revisit in our future work (see Section 6). Fig. 1 shows the schematic representation of the three spaces: cyber, relational and physical spaces by using NYS as a case study.

It has long been noted that traditional GIS methods based solely on notions of physical distance and proximity are inadequate for measuring human activities in cyber space (Kwan, 2000). With advances in ICTs and the proliferation of social media, a huge amount of data about human activities in a hybrid of spaces is now emerging (Sui & Shaw, 2018; Lazer et al., 2009). For example, researchers have explored e-mails (Eckmann, Moses, & Sergi, 2004), mobile phones (Grauwin et al., 2017; Vanhoof et al., 2017), smart cards (Sulis et al., 2018) and social media (Croitoru et al., 2015; Cvetojevic & Hochmair, 2021) data to study the digital traces of diverse human activities such as communication, mobility and information propagation. Social media data in particular has demonstrated several characteristics. First, social media enables efficient information transmission at a global scale (Stefanidis et al., 2013). By removing physical (or geographical) boundaries, social media allows individuals and organizations to disseminate and obtain information from anywhere and at any time (Kaplan & Haenlein, 2010). For example, during the Covid-19 pandemic, social media has been at the forefront of disseminating new scientific findings, publishing new protocols, and connecting the general public with their friends and family to reduce isolation and anxiety (González-Padilla & Tortolero-Blanco, 2020). In addition, social media data provides a rich geographic context that can be used to characterize human dynamics in physical space (Stock, 2018). For example, Long and Huang (2019) used social media data as a proxy of economic activity and found the positive impacts of design variables on urban vitality. Yin, Soliman, Yin, and Wang (2017) found that the mobility networks inferred from the geo-located tweets could yield geographically cohesive urban boundaries in the UK. In other studies, it has been shown how social media check-in data allows researchers to measure the physical co-presence in urban spaces and help to predict the socio-economic performance (Shen & Karimi, 2016; Shen, Karimi, Law, & Zhong, 2019). Furthermore, people’s activities on social media are part of a networking process whereby individuals share, reply and react selectively with other users based on their social ties or common interests (Sousa et al., 2010; Stefanidis et al., 2013). Users’ interaction on social media makes it possible to infer a relational space between users and analyze how information propagates. For example researchers have generated social networks based on users’ retweet activities (e.g., Yuan et al., 2019) and or the use of hashtags (e.g., Gunaratne et al., 2019). These characteristics of social media make it an ideal data source for investigating the interactions between the cyber, relational and physical spaces because not only does such data potentially bridge cyber and physical spaces, but also provides information about users’ interactions when it comes to generating relational spaces (Croitoru et al., 2015).

Turning to the current Covid-19 pandemic and the role of social media, it has had both a positive and negative impact on our society (González-Padilla & Tortolero-Blanco, 2020). The implementations of preventative measures such as social distancing and isolation intensified the use of social media as individuals try to stay connected while being physically apart. An example is how the number of tweets containing the keyword “coronavirus” spiked on the day when the U.S. reported the first Covid-19 related death (Chen et al., 2020). Moreover, Ahmed et al. (2018) demonstrated that the use of social media, specifically Twitter and Facebook as sources of health information, may promote vaccine uptake. Despite the benefits of social media including its ability to inexpensively reach a large population during a crisis (Saud et al., 2020), there are considerable public health concerns raised by the spread of harmful misinformation. This is especially the case for anti-vaccine messaging on social media platforms (Puri et al., 2020). In contrast to traditional media, social media allows anyone to generate rumors since the content posted need not undergo editorial scrutiny or scientific proof and thus can spread very rapidly (Massey et al., 2018). For example, amongst the top-ranked YouTube videos related to Covid-19, 27.5% of them contained non-factual information and currently have over 60 million views (Li, Bailey, Huynh, & Chan, 2020).

![Fig. 1. Schematic representation of the three spaces: cyber, relational and physical spaces.](image-url)
Meanwhile, the dissemination of anti-vaccination messaging on social media may generate more user engagements and has the potential to fuel vaccine hesitancy which could result in greater outbreaks in the physical world (Puri et al., 2020). For instance, Basch and MacLean, 2019 analyzed 150 HPV-related posts on Instagram and found that anti-vaccine posts generated significantly more than average likes than other posts. Additionally, through analyzing the temporal patterns of pro and anti-vaccine discourse on Twitter from 2010 to 2019, Gunaratne et al. (2019) found a significant surge in anti-vaccine discussion between 2015–2016 that coincided with the 2014–2015 measles outbreak in the real world. Furthermore, scholars have also expressed concerns about the “bubble filters” algorithms (Pariser, 2011) used by many social media platforms. Such algorithms work by pushing content to viewers based on their past click behavior and search history. Thus, these algorithms may result in the ideological isolation of anti-vaccine users and limit public health penetration to promote vaccination within social media (Puri et al., 2020; González-Padilla & Tortolero-Blanco, 2020). For example, through analyzing tweets related to the MMR vaccine, Yuan et al. (2019) found that anti-vaccine users mainly resided in their enclosed communities and were highly segregated from pro-vaccine users. However, such studies were largely based on the pre-Covid-19 events. We would argue that as Covid-19 has now become a focus of intense social media discourse, there is a need to update our understanding of online vaccine discourse by analyzing the pro and anti-vaccination debates of Covid-19 vaccines.

As one of the most important measures for preventing communicable infectious diseases (Andre et al., 2008), vaccination plays an important role to protect people against diseases and has been shown to save lives (Polack et al., 2020). As the vaccine discourse continues to evolve on social media with trends often tied to real-world events (Gunaratne et al., 2019), analyzing the pro and anti-vaccine discourse and the propagation of vaccine-related information on social media provides a lens to study the interactions between the cyber, relational and physical spaces. The next section explains the methodology in detail.

3. Study area and data collection

3.1. Study area

As discussed above, this study consists of three spaces: physical, relational, and cyber spaces. To ground this study in a physical space we chose the State of New York (NYS) for our case study. Although the level of vaccine sentiments (e.g., pro-vaccine rates and vaccine confidence) may vary in different states or even across different countries (e.g., Lyu et al., 2022; Larson et al., 2016), since the purpose of our study was to investigate the interactions between hybrid spaces (i.e., physical, relational and cyber) rather than the spatial disparities of vaccine sentiments, NYS demonstrates a relevant case study to fulfill our research purpose. We write this for several reasons, first, NYS is among the top-tier states that actively engage in online vaccine discussions (Chen & Crooks, 2022). In addition to this, NYS was the original epicenter of the pandemic in the U.S. (McMinn & Carlsen, 2022). For example, NYS as of December 15, 2020, lead the U.S. in Covid-19 deaths, at 35,427 (Dong, Du, & Gardner, 2020). Furthermore, NYS has played a proactive role in promoting the Covid-19 vaccine through both traditional media and digital platforms such as the State’s #GetTheVaxFacts campaigns (New York State, 2021b). To encourage vaccination in hard-hit communities, NYS governor allocated $15 million to promote vaccines (New York State, 2021a). Fig. 2 shows the map of the study area (NYS).

NYS has clear physical boundaries and we use these to screen out tweets sent from NYS. We then trace back to the users who posted these tweets and label them as NYS users. The regional boundaries in NYS and the locations of NYS users create the physical space of our study. The relational space describes the social networks of NYS users inferred from

Fig. 2. Map of study area (NYS) with the primary road system. Red dots denote collected vaccine-related tweets in NYS.
their tweet-reply activities. The relational space not only includes NYS users but also a small portion of cyber users who directly interact with NYS users. The relational space captures the tweet-reply activities of NYS users. Cyber space provides a larger conversational context to relational space. Cyber space includes NYS users and also cyber users in Covid-19 vaccine debates. Cyber users are those people who are either outside of NYS or without locational information. Unlike the relational space that only captures the tweet-reply activities of NYS users, the cyber space also includes the tweet-reply activities of cyber users. Fig. 1 illustrates the three spaces and their connections.

3.2. Data collection

The physical space consists of the boundaries of NYS and its nine regions collected from NYS GIS Clearinghouse (2020). Turning to the tweets, tweets were collected between December 1, 2020 until August 31, 2021, via repeated calls to the Twitter application programming interface version 2 (API v2). While discussions surrounding Covid-19 date back to the initial period of the pandemics, such discourse has continued to evolve (Tyson, Johnson, & Funk, 2020), even at the time of writing new discussions are emerging. Thus our study, like any study, may not capture the entire picture of vaccine debates. However, our selected time period (i.e., December 1, 2020 until August 31, 2021) does capture the upsurge in vaccine discussions on Twitter as seen in other longitudinal analysis studies (e.g., Chen & Crooks, 2022). We identified tweets that are related to Covid-19 vaccine by using a combination of keywords (i.e., vaccine, vax, Moderna, Pfizer, Johnson & Johnson, AstraZeneca). Our rationale for choosing these terms was that previous studies have used the term of vaccine and its variations, such as “vaccine, vax, vaccination” (e.g., Gunaratne et al., 2019 & Yuan et al., 2019). We also included other keywords such as “Moderna, Pfizer, Johnson & Johnson” because these were the Covid-19 vaccine that were approved by the FDA for emergency use in the US at the time of this study (FDA, 2020a, 2020b; FDA, 2021), and “AstraZeneca” in Canada and the U.K. (GOV.UK, 2022).

The NYS tweet dataset contained 39,204 tweets created by 11,443 distinct users. Each tweet has four types of information associated with it: identifier, content, location, and reference. Fig. 2 maps the locations of these tweets. The potential bias of this dataset is further discussed in Section 6. Table 1 shows the returned variables associated with each tweet. The variable of “conversation_id” was used as a key to snowball sample the entire conversation about Covid-19 vaccine. The variable of “in_reply_to_user_id” allowed to construct tweet-reply networks between users. For example, if user A (“author_id”=“A”) has replied to user B (“in_reply_to_user_id”=“B”)’s tweets twice, then an edge of weight two is created from A to B. Different from other studies that built retweet networks (Yuan et al., 2019; Croitoru et al., 2015; Bello-Orgaz, Hernandez-Castro, & Camacho, 2017) or hashtag networks (Gunaratne et al., 2019), this study built tweet-reply networks because tweet-reply activities are more driven by social motivation (Sousa et al., 2010) and require more user engagements (i.e., users do not just press the retweet icon).

To create the cyber and relational spaces, we collected a second round of tweets to capture the larger ongoing conversation about the Covid-19 vaccine. If a tweet is in reply to another tweet, then the two tweets are in the same conversation. By searching tweets based on their “conversation_id” as shown in Table 1, we could snowball sample all the other tweets that were in the same vaccine conversation as the NYS tweets. We first extracted a list of unique “conversation_id” from the NYS tweet dataset. Then we called the Twitter API v2 again and used the list of “conversation_id” as a key to search Covid-19 vaccine conversations. After obtaining the tweets in all conversations, we filtered out the tweets that are irrelevant to vaccines (i.e., using the same keywords used in the search above to filter out the tweets). The remaining tweets then constituted the final dataset for constructing cyber space. This cyber space dataset involved not only NYS users but also cyber users who do not share their locational information or live outside of NYS. The cyber space dataset contains 448,958 tweets created by 239,716 distinct users. We created the relational space based on the tweets that are directly related to NYS users where the NYS users are either the author or the receiver of a tweet. The relational space dataset has 55,484 tweets posted by 21,876 distinct users. The variables associated with each tweet are shown in Table 1.

### Table 1

| Categories   | Variables                  | Descriptions                                                                 |
|--------------|----------------------------|------------------------------------------------------------------------------|
| Identifier   | tweet_id                   | Unique id of the tweet                                                       |
|              | conversation_id            | Id of the original tweet that this tweet is directly or indirectly replied to |
|              | author_id                  | Unique id of the user who posted the tweet                                    |
| Content      | text                       | Posted text                                                                  |
|              | lang                       | Language of the tweet                                                        |
|              | created_at                 | Creation time of the tweet                                                   |
| Location     | place_id                   | Id of the place where the tweet is posted.                                   |
|              | place_geometry             | Geographical information of the place in the form of coordinates or bounding box |
| Reference    | in_reply_to_user_id        | Id of the user that this tweet replies to. N/A if none.                       |

4. Methodology

To understand the interactions between cyber, relational, and physical spaces, this study used a combination of techniques which are outlined in Fig. 3. After data collection, this study first conducted sentiment analysis of tweets (Section 4.1) through data processing, manual annotation, training classification model, and tweet labeling. By doing so, we classified tweets and users into three categories: pro, anti and neutral. Then, we constructed tweet-reply networks and performed community detection algorithms to reveal large communities in different spaces and examine users’ profiles (Section 4.2). Last, we built hybrid space networks to map the interactions between the three spaces (Section 4.3). For interested readers, we provide the code of our methodology and results at https://osf.io/mxq3k/?view_only=da5e78a3e0a54789b46067fac54d548e. We do this to allow for replication and for readers to extend our work as they see fit.

4.1. Sentiment analysis

One of the key aspects of implementing the Splatal framework is to understand the transformation and linkage between cyber, relational and physical spaces (Shaw & Sui, 2020). In our study, this relates to the spread of pro and anti-vaccine opinions between the three spaces. Therefore, the first step is to analyze the sentiments of both the vaccine tweets and Twitter users. In the sentiment analysis of tweets, opinions are often grouped into three categories, positive, negative and neutral (Rathi et al., 2018; Yuan et al., 2019; Pak & Paroubek, 2010). In our case, these three categories correspond to pro, anti and neutral. Pro-vaccine relates to tweets that demonstrate support to the Covid-19 vaccine. While anti-vaccine tweets can be considered those that tend to delay in acceptance or refusal of vaccines. Lastly, neutral tweets are those that report certain facts about the Covid-19 vaccine without showing any inclination of their opinion. For example, such as tweets might cite an academic study, or announce a new policy, or come from news agencies. Table 2 shows a selection of tweets from these three categories.

To automatically classify tweets into the three categories, we built a

---

This is a continuation of the text from the document. The above text is a natural representation of the document as per the guidelines.
sentiment classifier using a Support Vector Machine (SVM). Our rationale for choosing a SVM classifier is that they have been widely used in text-based classifications (e.g., Joachims, 1999) including those looking at the sentiment of tweets (e.g., Rathi et al., 2018; Yuan et al., 2019), and it has been noted that they are appropriate when a dataset is unbalanced, in the sense that the number of entities in each category is not equal (e.g., Tang, Zhang, Chawla, & Krasser, 2008; Zhou et al., 2015).

We also tested other classifiers including logistic regression, non-linear SVM and random forests. However, the linear SVM classifier outperformed the other classifiers in producing higher prediction accuracy and shorter execution time which we will come back to later. To train the linear SVM classifier, we first hand-labeled 2,065 unique tweets (Section 4.1.1). Then, we pre-processed all tweets to remove unnecessary information (cyber, relational and physical spaces).

Fig. 3. Research workflow to investigate the propagation of different opinions between three spaces: cyber, relational and physical spaces.

Table 2
Examples of pro, anti and neutral vaccine tweets.

| Sentiment | Tweet |
|-----------|-------|
| Pro-vaccine | "Vaccine appointment tomorrow because #science not stupidity" "Getting my vaccine" |
| Anti-vaccine | "Everyone who is taking the vaccine has an IQ of the average preschooler." "I’m in NO rush to get the vaccine." |
| Neutral | "NEW: More than 40 percent of New Yorkers have received at least one dose of the COVID-19 vaccine. Check out the full story here" "NYC to open COVID vaccine site in Queens via @nypmetro" |


4.1.1. Manual annotation

The labeling process was done by using a labeling questionnaire approach. Specifically, we first randomly selected 5,000 tweets from the whole data corpus to create a sample pool. We then generated 100 questionnaires from the sample pool that each contained 100 tweets. As such, each tweet will appear in two questionnaires and creates some overlapping across different participants. We then randomly distributed these questionnaires to 100 participants on campus to label the tweets and received 62 responses. Simultaneously, one researcher with domain knowledge also randomly gathered 15 tweets (15%) from each questionnaire, marked them independently and used these labels as a reference to measure participants’ reliability. We used both the percentage agreement and Cohen’s kappa statistic to calculate the inter-rater reliability between participants and the researcher (Cohen, 1960). To identify reliable participants, we only retained participants who has over 60% of agreement with the researcher and the kappa’s value higher than 0.4 (i.e., moderate level of agreement proposed by Landis & Koch, 1977). Through this process, we were able to identify 25 reliable participants and 2,065 unique tweets after text processing. Among the 2,065 unique tweets, 435 tweets had been labeled twice by participants, while 1,630 tweets had only been labeled once. To increase the robustness of the annotation, the researchers of this paper then labeled these 1,630 tweets manually to ensure each tweet had two annotators. The conflicting labeling results between participants and the
4.1.3. Classification model

After pre-processing, the tweets then needed to be vectorized and re-scaled in order to be learned by the machine. Previous studies, such as Yuan et al. (2019) and Rathi et al. (2018) have used Term Frequency-Inverse Document Frequency (tf-idf) for this purpose. The logic behind tf-idf is that it gives a high weight to a term that appears often in a particular text (in our case a tweet) but not in many other texts (Müller & Guido, 2016). By doing so, the terms that distinguish one text from others would receive a higher weight, but the terms that frequently appear in all texts (e.g., “vaccine”) will receive a lower weight. After rescaling the terms, tf-idf normalizes the representation of each text to have Euclidean norm of 1 so that the length of texts will not produce any bias. To do this, two parameters need to be tuned to construct a tf-idf vectorizer, N-gram and minimum document frequency (Min-df). N-gram refers to using a contiguous sequence of n words as a feature. Unigram means using a single word as a feature, bigram means using a two-word sequence as a feature and so on (Ahuja, Chug, Kohli, Gupta, & Ahuja, 2019). When building the vocabulary, Min-df is used to remove features that have a document frequency lower than a given threshold (Pedregosa et al., 2011).

Once we have vectorized and re-scaled the data, to train the tf-idf vectorizer and the SVM classifier, we split the labeled data into training (80%) and test (20%) datasets and perform the ten-fold cross-validation. The proportion of three classes (i.e., pro, anti and neutral) in the training and test datasets remained the same. Standard performance measures (i.e., precision, recall, accuracy, and F1-score) were used to compare the models of different combinations of parameters (Zhou et al., 2015). Table 3 shows the parameters that generated the best performance. The model constructed using these parameters achieved a cross-validation accuracy of 79.3% in the training dataset and the accuracy of 76.2% in the test dataset (see Table 4) which is in a similar range to other studies of sentiment analysis of vaccine-related tweets (e.g., Yuan et al., 2019; Du, Xu, Song, & Tao, 2017; Piedrahita-Valdés et al., 2021). Table 5 summarizes the precision, recall and F1-score for each class on the test dataset.

4.1.4. Tweets and user labeling

Once the linear SVM classifier was trained and tuned we then automatically classified all tweets in the data corpus into one of three classes: pro, anti and neutral. Table 6 shows the tweet labeling results, i.e. the number of tweets falling into the three classes in the physical, relational and cyber space datasets. Next, the labeled tweets with the same “author_id” were aggregated using the simple majority voting rule to decide users’ opinions. A simple majority voting rule is a decision rule that selects one from many alternatives based on the predicted classes that have the most votes (Lam & Suen, 1997) and it has been applied in other studies (e.g., Yuan et al., 2019; Gunaratne et al., 2019) to aggregate tweets or hashtags to derive users’ opinions. Specifically, if a user has the majority of their tweets labeled as one class, the user is also classified into that class. However, if the user has an equal amount of pro and anti-vaccine tweets, then this user is labeled as neutral.

4.2. Social network analysis

4.2.1. Reply network construction

After sentiment analysis of tweets and users, we then constructed two tweet-reply networks between Twitter users to represent relational and cyber spaces. In the tweet-reply network, nodes represent Twitter users,
and edges indicate reply activities between Twitter users. For instance, if user A has replied to user B two times on Twitter, then an edge with a value of two is drawn from user A to user B. User A has an out-degree of two and user B has an in-degree of two. To sift out influential users, to make the networks denser and to concentrate on the main conversation camps, past researchers have removed inactive users from their analysis, for example, Yuan et al. (2019) removed isolate users while Bello-Orgaz et al. (2017) removed users with a low degree. In our study we used a threshold of total degree less than two to eliminate inactive users. The size of the relational space network is smaller than the cyber space network in terms of nodes (active users) and edges (number of replies).

4.2.2. Community detection in reply networks

After constructing the tweet-reply networks, we then turn to community detection to group closely connected Twitter users based on their connections. Community detection is a widely used method to reveal the underlying structures in social networks (Bello-Orgaz et al., 2017). Communities in a social network indicate groups of users whose connections with each other are stronger than with users outside of their communities (Papadopoulos, Kompatsiaris, Vakali, & Spyridonos, 2012). In our case, communities represent groups of Twitter users who frequently reply to each other’s tweets to discuss the Covid-19 vaccine. Our interests in community detection stem from the community’s central role in information propagation and diffusion in social networks (Murata, 2010). We used the widely applied Louvain algorithm to detect communities and partition Twitter users because this algorithm outperforms other methods in producing higher modularity scores and shorter execution times (Blondel et al., 2008). The Louvain algorithm includes two iterative phases. The first phase is to calculate the modularity gain by adding and removing nodes to a new community. Once the modularity score cannot be increased by removing individual nodes, the algorithm aggregates the current communities into nodes and then repeat the first phase. We used the Louvain algorithm to detect communities only based on the edges (i.e., the number of tweet replies) between users and eliminate other information such as users’ locations and opinions. Therefore, the detected communities could be purely considered as well-connected online discussion communities or groups where users in the same community frequently reply to each other’s tweets to discuss the Covid-19 vaccine. By doing so, we could investigate how users with different opinions and from different locations (i.e., cyber, relational and physical spaces) interact with each other, and how the physical boundaries (i.e., regional boundaries in the NYS) impact the formation of online communities.

The software Gephi was used to perform the Louvain algorithm for community detection. When implementing the algorithm, the parameter of resolution controls the size of the smallest community (Bastian, Heymann, & Jacomy, 2009). Previous work has shown how setting the resolution to three resulted in robust findings (e.g., Yuan et al., 2019), and in our work, this resulted in 4,081 communities in the cyber space network and 1,328 communities in the relational space network. Fig. 4 shows the size (i.e., number of users) of all detected communities in the six top large communities (a)-(h) in the relational space network. Table 7 shows the attributes of the tweet-reply networks in relational and cyber spaces after removing inactive users. Table 7 shows the attributes of the tweet-reply networks in relational and cyber spaces after removing inactive users whose total degree is less than two.

| Networks | Nodes | Edges |
|----------|-------|-------|
|          | NYS users | Cyber users | Type | Count | Total Weights |
| Relational Space | 3,245 | 4,333 | Directed | 12,692 | 21,459 |
| Cyber Space | 3,542 | 90,658 | Directed | 193,310 | 251,376 |

In addition to the community detection in the tweet-reply networks, we also created hybrid space networks (Croitoru et al., 2015) between Twitter users to investigate the interactions between cyber, relational and physical spaces. By aggregating Twitter users based on their location, we created two hybrid space networks to investigate the information propagation in the cyber–relational and the physical–relational systems. The cyber-relational network is bipartite and depicts the interactions between cyber and relational space users, while the physical-relational network shows the communications between relational space users and users living in the nine regions in NYS. After constructing the hybrid networks, we investigated how the online vaccination debates, in the form of pro and anti-vaccine tweets, are propagated between physical, relational and cyber spaces.

5. Results

As mentioned in Section 4.2.2, the detected communities in our analysis represent groups of people who frequently reply to each other on Twitter to discuss the Covid-19 vaccine. This section analyzes the attributes of the communities in the cyber (Section 5.1), relational (Section 5.2) and physical spaces (Section 5.3). Next, Sections 5.4 and 5.5 show how different opinions (i.e., pro and anti-vaccine opinions) spread from physical to relational, and from relational to cyber spaces.

5.1. Cyber space communities

Fig. 5 visualizes the tweet-reply network in the cyber space. In this figure, a point represents a user. The users of the six top large communities I-VI are highlighted using colors. As mentioned in Section 4.2.2, large communities in cyber space are those that contain over 1% of total users. The community I colored in red is the largest community that contains 33.13% of the total users. Community II colored in green is the second large community that covers 26.72% of the total users. Then, communities III, IV, V and VI have 10.15%, 8.23%, 5.8% and 1.0% of the total users respectively. Communities I and II are the two main camps of vaccination debate in the cyber space.

Fig. 6 shows the users’ profiles in the six large communities. Fig. 6A illustrates the percentage of pro-vaccine, anti-vaccine and neutral users. In each community, more than a half (> 55%) of users are proponents of the Covid-19 vaccine, while there are still around 10–20% of anti-vaccine users. Community II and V have the largest proportion (17%) of anti-vaccine users. Users with neutral opinions constitute 10%–15% of total users in each community. Fig. 6B shows the distribution of users’ locations in each community. The top graph in Fig. 6B shows the percentage of cyber space and relational space users in each community. This graph suggests that the cyber space users dominate the online vaccination debates because they have reached over 64% of total users in all six large communities. Relational space users constitute 11–36% of total users in the six large communities. Among them, community VI has the largest proportion (35.2%) of relational space users. The bottom graph in Fig. 6B presents the proportion of physical space users (NYS users) in each community and depicts their locations at the regional level. Despite its small proportion (2.2%–6.5%), NYS users could be found in all six communities. Among them, communities III and VI have
the largest proportion (6.5%) of NYS users. Meanwhile, the NYS users in the six large communities tend to anchor in a particular region. For example, while communities I, II, III, IV have the majority of NYS users coming from New York City, communities V, VI have the most NYS users coming from Western New York and Western Finger Lakes respectively.

5.2. Relational space communities

After examining the large communities in cyber space, this section analyzes the eight large communities in relational space. Relational space represents the social networks of NYS users that are inferred from their tweet-reply activities (Section 3.1). As shown in Fig. 1, the social network in relational space includes NYS users and also the cyber users who are directly connected with NYS users to discuss the Covid-19

![Image of networks](image-url)
vaccine. In Fig. 7, nodes represent users and edges represent their interactions (i.e. tweet-reply activities). Fig. 7 illustrates the community structure in the relational space by highlighting its eight large communities (a)-(h). Large communities are communities with over 1% of total users (Section 4.2.2). The eight large communities all together cover more than 66% of users in the relational space. The size of communities decreases from (a)-(h). The largest community, community (a), covers more than 24% of total users in relational space. The community (h), which is the smallest, contains around 1.3% of total users in relational space.

Fig. 8A shows the distribution of users’ opinions in the eight large communities in relational space. Each community has more than 65% of users supporting the Covid-19 vaccine. Among these communities, the community (g) has the largest proportion (82.2%) of Covid-19 vaccine proponents, and the community (e) has the smallest proportion (65.9%) of vaccine proponents. Compared to the pro-vaccine users, the anti-vaccine users in each community constitute a smaller proportion ranging from 6.7% to 14.8%. Community (c) has the largest proportion (14.8%) of anti-vaccine users and community (g) has the smallest proportion (6.7%). The users with neutral opinions constitute the smallest...
group (4.4%–8.9%) in each community compared to the users with pro or anti-vaccine opinions. The grey bars represent users with no opinion because they only received tweet replies from other users but never posted tweets about Covid-19 vaccine.

Fig. 8B analyzes the spatial traces of the eight large communities (a)-(h) in relational space by showing the distribution of users’ locations in these communities. The length of color bars represents the percentage of users posting tweets from a particular region of NYS. Cyber users are labeled as “no location” and colored in white. Fig. 8B indicates that all eight communities have a group of users whose locations are anchored in a particular region in NYS. For example, in communities (a), (b), (d), (e), (g), (h), most NYS users are posting from New York City (i.e., yellow bar). Among them, the community (g) has the largest proportion (47.8%) of New York City users. In addition to New York City, Fig. 8B also shows the existence of another two local clusters in Western New York (i.e., grey bar) and Western Finger Lakes (i.e., pink bar). Community (c) has a majority of NYS users posting from Western New York, and community (f) has a majority of NYS users tweeting from Western Finger Lakes. Fig. 9 maps users’ locations of the eight large communities (a)-(h) into physical space. The red dots in Fig. 9 represent NYS users in the eight communities. Fig. 9 shows the similar results as Fig. 8B that while the communities (a), (b), (d), (e), (g), (h) have most NYS tweeting from New York City, communities (c), (f) shows a cluster of local users whose locations were anchored in Western New York and Western Finger Lakes. Fig. 8B and Fig. 9 together suggest the existence of distance-based and boundary-confined local clusters in the large communities in relational space.

5.3. Physical space communities

After examining the communities in cyber and relational spaces (Section 5.1 and 5.2), this section analyzes the communities in physical space. While the communities in cyber and relational spaces are detected based on the strengths of users’ interactions, the communities in the physical spaces are defined based on the regional boundaries. Physical space has nine communities that correspond to the nine regions in NYS.

Fig. 9. Mapping users’ locations of the eight large communities (a)-(h) from relational space into physical space. Red dots represent NYS social media users.
pro-vaccine users far outweighs that of anti-vaccine or neutral users. The majority (> 80%) of users supporting the Covid-19 vaccine, and this size of pro-vaccine users far outweighs that of anti-vaccine or neutral users. Anti-vaccine users constitute 6%–12% of users in these nine regions. Among these regions, Western New York has the largest proportion (11.6%) of anti-vaccine users. Neutral users constitute the smallest proportion ranging from 5.3% to 8.5%.

Communities in the three spaces (i.e., cyber, relational and physical) demonstrate similarities in terms of users’ opinions as shown in Figs. 6A, 8A and 10. First, pro-vaccine users dominate the vaccine debates in the large communities in all three spaces. Large communities (i.e., communities with over 1% of total nodes) in cyber and relational spaces and communities in physical space have more than half of users supporting the Covid-19 vaccine. Next, anti-vaccine users are fragmented in the communities in three spaces. The proportions of anti-vaccine users range from 10% to 18% in cyber space, and range from 7% to 15% in relational space and range from 6% to 12% in physical space. We did not observe highly clustered and segregated anti-vaccine users as found in other vaccination studies (e.g., Yuan et al., 2019; Schmidt et al., 2018). However, the communities in the three spaces also demonstrate differences. Communities in the physical space have larger proportions of vaccine proponents and smaller proportions of vaccine opponents than those in the relational and cyber spaces. The three spaces’ different rates of pro or anti-vaccine stances are consistent with the tweet-labeling results as shown in Table 6.

5.4. Interactions between physical and relational spaces

After examining the large communities in three spaces, we then investigate how different opinions (i.e., pro and anti-vaccine opinions) spread between physical, relational and cyber spaces. Fig. 11A–C show how all tweets, pro-vaccine tweets and anti-vaccine tweets spread between physical and relational spaces. In this figure, nodes R1–R9 represent users from the nine regions of NYS. The node of relational space indicates users in the relational space who are without locational information or posting outside of NYS. The size of nodes is proportional to the number of users. Edges are undirected and their thickness is proportional to the number of replied tweets between two nodes. Self-loop edges indicate replied tweets occurred between users of the same region. The number shows the percentage of replied tweets that happened between two nodes. The map in the bottom right of the Fig. 11 indicates the boundaries of the nine regions of NYS.

By comparing nodes’ sizes, Fig. 11 indicates that the region of New York City (R2) has the largest number of users participating in the vaccination debates and this number far outweigh the other regions. The region of Western New York (R9) has the second largest number of users. Moreover, the comparison of nodes’ sizes between Fig. 11B and C suggests that more users are involved in pro-vaccine than anti-vaccine discussions.

Fig. 11 also indicates that while the users from the nine regions intensively communicate with relational space users, they also engage in local conversations with users from the same region. In Fig. 11A, the thick edges linking the nine regions (R1–R9) and the relational space indicate a strong connection between physical and relational spaces. The self-loop edges suggest that the users of the nine regions also engage in local conversations. However, users’ connections with relational space users are stronger than with users from the same region. For example, New York City users (R2)’s interactions with relational space users constitute 64% of total tweets, but their interactions with New York City users only constitute 7%. Compared to the self-loop edges, the cross-region edges (i.e., communications between two different regions) are even fewer and weaker. Meanwhile, the comparison of the edges’ thickness between Fig. 11B and C indicates that NYS (i.e., physical space) users propagate more pro-vaccine than anti-vaccine opinions to relational space.

5.5. Interactions between relational and cyber spaces

Fig. 12 shows the spread of vaccine-related tweets between relational and cyber spaces. Nodes represent users in relational or cyber spaces and their size is proportional to the number of users in each space. Edges indicate tweet-reply activities. Edges are directed and drawn from authors to receivers of replied tweets. The thickness of edges is proportional to the number of replied tweets. Fig. 12 shows that the relational
space has a smaller user size compared to cyber space. In both cyber and relational spaces, there are fewer users in anti-vaccine than pro-vaccine discussions.

Fig. 12A shows that the relational space users have strong interactions with cyber space users because in total more than 45% of replied tweets happen between the two spaces. For example, 4.3% of replied tweets are sent from relational to cyber spaces and 42.4% of replied tweets are from cyber to relational spaces. The self-loop edges indicate that intense communications are happening within each space. For example, 38.8% of replied tweets are within cyber space users and 14.5% of replied tweets are within relational space users. This suggests that while relational space users frequently communicate with relational space users, they also actively interact with cyber space users through sending or receiving replies from cyber space users. Fig. 12B and C indicate that the pro and anti-vaccine discussions have a similar trend.

The comparison between Fig. 12B and C indicates that the self-loop edges of pro-vaccine tweets in cyber space is 35.9% while that of anti-vaccine tweets is 47.3%. This suggests that the cyber space users are experiencing stronger echo chamber effects in anti-vaccine than pro-vaccine debates.

6. Discussion and conclusion

By utilizing the Splatial framework and applying it to the study of the Covid-19 pandemic we present a novel way to study vaccination debates. Through analyzing users’ opinions in cyber, relational and physical spaces, our study did not observe the phenomenon of polarization which is commonly found in other political and vaccination debates (e.g., Yuan et al., 2019; Schmidt et al., 2018). Our results suggest that pro-vaccine users dominated the conversations in large communities (> 1% of total users), while anti-vaccine users were fragmented across these communities in all of the three spaces (i.e., cyber, relational and physical spaces). A possible explanation for this was that as the Covid-19 vaccine became available in December 2020, an increasing trend in positive sentiment toward vaccines took place on social media (Hu et al., 2021). For example, there have been many efforts to debunk vaccine misinformation on social media (New York State, 2021b; CDC, 2021) and there has been a growing trend for general users to actively engage in debates with anti-vaccine users to refute anti-vaccine arguments (e.g., Jamison et al., 2020).

In addition to the non-polarized vaccination debates, our study also provided a more nuanced view of the information propagation mechanism across various spaces (i.e., cyber, relational and physical spaces). Although social media was often considered as a great “no place” that

![Fig. 11. The hybrid space network shows the information propagation between physical and relational spaces. (A) shows the network of all tweets, (B) shows the pro-vaccine tweets, and (C) shows the anti-vaccine tweets.](image-url)
can erase the barrier of physical distance in human communication (Herrera, 2016), we found that people’s opinions towards a particular topic (e.g., Covid-19 vaccine) varied across spaces. In our case, physical space (NYS) users demonstrated a higher rate of pro-vaccine stance and a lower rate of anti-vaccine stance than relational space and cyber space users. Meanwhile, physical space users have propagated more pro-vaccine than anti-vaccine content to the relational space. One reason for this could relate to the on-ground efforts to promote the Covid-19 vaccine in physical spaces, such as allocating funds to hard-hit physical communities and employing both traditional and digital media to promote vaccines (e.g., New York State, 2021a).

Our findings also suggest the co-existence of cross-space and intra-space interactions. For example, while intense communications happened between users from different spaces (e.g., physical-relational or relational-cyber interactions), users were also engaged in local communications with users from the same region or the same space (as shown in the self-loop edges in Figs. 11 and 12). Meanwhile, large communities (i.e., that with >1% of total users) in both cyber and relational spaces contained a group of physical space users whose locations were anchored in a particular region of NYS, indicating the existence of distance-based and boundary-confined clusters within social media. This suggests that social media can help people to communicate across physical boundaries with a larger relational and cyber audience or within physical boundaries with local friends which has been argued by others (e.g. Bingham-Hall & Law, 2015; Arthur & Williams, 2019; Wellman, 2002).

The co-existence of cross-space and intra-space communications highlights the need to incorporate people’s relational networks and their cyber presence in the physical place-making since one objective of public place-making is to stimulate greater interactions among people and foster vitalized communities (Abdel-Aziz, Abdel-Salam, & El-Sayad, 2016). Meanwhile, the intense information flows that happen between cyber, relational and physical spaces indicate the inseparable nature of online-and-offline activities and strengthens the need to develop a hybrid-space perspective (i.e., the Splatial framework) for studying human dynamics and interactions (Shaw & Sui, 2020). As an early attempt to apply the Splatial framework for studying a real-world event (i.e., Covid-19 vaccine debate), our analysis approach can be applied in various domains for analyzing the interactions across a hybrid of spaces. Moreover, our definitions of the cyber, relational and physical spaces added realistic meanings to the three spaces and our snowball sampling methods provides a new way to construct three spaces using social media data. Next, our hybrid space networks help to visualize social interactions across different spaces and capture the spatial traces of online communities.

With respect to future works, while our focus here was on a proof of concept utilizing NYS as a case study, based on the findings presented here, it would be interesting to scale up to the whole of the U.S. or globally. However, this would require more data and computational resources than we currently have. But we believe our findings and the methodology presented here are generalizable to the other study areas. To this end, we have provided the code used in our methodology at OSF. However, as with all works, there are limitations related to our work. For example, one of our limitations is that our analysis only includes cyber, relational and physical spaces while excluding mental and relative spaces. As noted above (Section 2) this was done because relational space inferring networks between objects has more potential to bridge the cyber and physical spaces but a logical next step would be

![Fig. 12](image-url) The hybrid space network shows the information propagation between relational and cyber spaces. (A) shows the network of all tweets, (B) shows the pro-vaccine tweets, and (C) shows the anti-vaccine tweets.
Basch, C. H., & MacLean, S. A. (2019). A content analysis of HPV related posts on Instagram. Human Vaccines & Immunotherapeutics, 15(7–8), 1476–1478.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. Proceedings of the International AAAI Conference on Web and Social Media, 3, 361–362.

Batty, M. (1997). Virtual geography. Futures, 29(4–5), 337–352.

Berman, G., Hernandez-Camacho, D. (2017). Detecting discussion communities on vaccination in Twitter. Future Generation Computer Systems, 66, 125–136.

Berndt, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referential sampling. Sociological Methods & Research, 10(2), 141–163.

Bingham-Hall, J., & Law, S. (2015). Connected or informed?: Local Twitter networking in a London neighbourhood. Big Data & Society, 2(2), Article 2053951715597457.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008.

Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. Nature, 439(7075), 462–465.

Bryant, R. (2001). What kind of space is cyberspace. Minerva-An Internet Journal of Philosophy, 5(2001), 138–1.

Castells, M. (2010). Globalisation, networkizing, urbanisation: Reflections on the spatial dynamics of the information age. Urban Studies, 47(13), 2737–2745.

CDC. (2021). Connecticut uses social media to engage long-term care residents. Covid-19 Vaccine Community Features. Retrieved 2022-02-20, from https://www.cdc.gov/vaccines/covid-19/health-departments/features/connecticut-itcopp.html.

Chen, C.-F., Shi, W., Yang, J., & Fu, H.-H. (2021). Social bots’ role in climate change discussion on twitter: Measuring standpoints, topics, and interaction strategies. Advances in Climate Change Research, 12(6), 913–923.

Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. JMIR Public Health and Surveillance, 6(2). Article e19273.

Chen, Q., & Crooks, A. (2022). Analyzing the vaccination debate in social media data pre- and post-covid-19 pandemic. International Journal of Applied Earth Observation and Geoinformation, 110, Article 102783.

Chmiele, A., Kowalska, K., & Hojst, J. A. (2009). Scaling of human behavior during portal browsing. Physical Review E, 80(6), Article 066122.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46.

Croitoru, A., Waynati, N., Crooks, A., Radziukowski, J., & Stefanidis, A. (2015). Linking cyber and physical spaces through community detection and clustering in social media feeds. Computers, Environment and Urban Systems, 53, 47–64.

Dodge, M., & Kitchin, R. (2003). Mapping cyberspace. Routledge.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track Covid-19 in real time. The Lancet Infectious Diseases, 20(5), 533–534.

Du, J., Xu, J., Song, H.-Y., & Tao, C. (2017). Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. BMC Medical Informatics and Decision Making, 17(2), 63–70.

Eckmann, J.-P., Moses, E., & Sergi, D. (2004). Entropy of dialogues creates coherent structures in e-mail traffic. Proceedings of the National Academy of Sciences, 101(4), 13744–13749.

FDA (2020a). Coronavirus (Covid-19) update: December 22, 2020. U.S. Food & Drug Administration Newsroom. Retrieved 2022-01-17, from https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-december-22-2020.

FDA (2020b). FDA takes key action against Covid-19 by issuing emergency-use authorization for first Covid-19 vaccine. U.S. Food & Drug Administration Newsroom. Retrieved 2022-01-17, from https://www.fda.gov/news-events/press-announcements/fda-takes-key-action-against-covid-19-issuing-emergency-use-authorization-on-first-covid-19.

FDA (2021). Coronavirus (Covid-19) update: April 27, 2021. U.S. Food & Drug Administration Newsroom. Retrieved 2022-01-17, from https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-april-27-2021.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual and collective dynamics in mobile phone data. PLoS ONE, 3(6), Article e214466.

Hamstead, Z. A., Fisher, D., Ilieva, R. T., Wood, S. A., McPhearson, T., & Kremer, P. (2021). GOV.UK (2022). Decision: Information for U.K. recipients on Covid-19 vaccine astrazeneca. Retrieved 2022-01-17, from https://www.gov.uk/government/publications/regulation-174.

Gonzalez, Padilla, D. A., & Tortolero-Blanco, L. (2020). Social media influence in the Coronavirus pandemic. International Brazilian Journal of Urology, 46, 120–124.

GOV.UK (2022). Decision: Information for U.K. recipients on Covid-19 vaccine astrazeneca (Regulation 174). Medicines & Healthcare Products Regulatory Agency. Retrieved 2022-01-17, from https://www.gov.uk/government/publications/regulatory-approval-of-covid-19-vaccine-astrazeneca-information-for-uk-recipients-on-covid-19-vaccine-astrazeneca.

Grauwin, S., Szell, M., Sobolevsky, S., Holve, P., Simini, F., Vanhoof, M., Smoreda, Z., Barabasi, A.-L., & Ratti, C. (2017). Identifying and modeling the structural discontinuities of human interactions. Scientific Reports, 7(1), 1–11.

Guarnera, K., Coones, E. A., & Huyan, H. (2019). Temporal trends in anti-vaccine discourse on Twitter. Vaccine, 37(35), 4867–4871.

Hamstead, Z. A., Fisher, D., Ilieva, R. T., Wood, S. A., McPhearson, T., & Kremer, P. (2014). Geolocated social media as a rapid indicator of park visitation and equitable access. Computers, Environment and Urban Systems, 72, 38–50.

Herrera, G. L. (2016). Cyberspace and sovereignty: Thoughts on physical space and digital space. In Power and Security in the Information Age (pp. 81–108). Routledge.

Hu, T., Wang, S., Luo, W., Zhang, M., Huang, X., Yan, Y., et al. (2021). Revealing public opinion towards Covid-19 vaccines with Twitter data in the United States.
Papadopoulos, S., Kompatsiaris, Y., Vakali, A., & Spyridonos, P. (2012). Community detection in social media. Data Mining and Knowledge Discovery, 24(3), 515–554, Pacheco, W. J. (2011). The filter bubble: What the internet is hiding from you. Penguin UK.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Griesl, O., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Piedrahita-Valderrama, H., Piedrahita-Castillo, D., Bermejo-Hijuera, J., Guillen-Saiz, P., Bermejo-Hijuera, J. R., Guillemin-Saiz, J., et al. (2021). Vaccine hesitancy on social media: Sentiment analysis from June 2021 to April 2019. Vaccines, 9(1), 28.

Polack, F. P., Thomas, S. J., Kitchin, N., Absolon, J., Gurtman, A., Lockhart, S., et al. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. New England Journal of Medicine.

Porter, C. E. (2004). A typology of virtual communities: A multi-disciplinary foundation. Future research. Journal of Computer-Mediated Communication, 10(1), JCMC111.

Puri, N., Coomes, E. A., Haqbehayn, H., & Gunaratne, K. (2020). Social media and vaccine hesitancy: New updates for the era of Covid-19 and globalized infectious diseases. Human Vaccines & Immunotherapeutics, 16(11), 2586–2593.

Rathi, M., Malik, A., Vashnury, D., Sharma, R., & Mendiartt, S. (2018). Sentiment analysis of tweets using machine learning approach. In 2018 Eleventh International Conference on Contemporary Computing (IC3) (p. 1–3). doi: 10.1109/ICC3.2018.8530517.

Salathe, M., & Bonhoeffer, S. (2008). The effect of opinion clustering on disease outbreaks. Journal of The Royal Society Interface, 5(29), 1505–1508.

Saud, M., Mashud, M., & Ida, R. (2020). Usage of social media during the pandemic: Seeking support and awareness about Covid-19 through social media platforms. Journal of Public Affairs, 20(4), Article e2417.

Schmidt, A. L., Zollo, F., Scala, A., Betsch, C., & Quattrociocchi, W. (2018). Polarization of the vaccination debate on Facebook. Vaccine, 36(25), 3606–3612.

Schultz, L. (2019). Introducing New York’s rural economics (Vol. 11). Rockefeller Institute of Government Blog. Retrieved from https://rockinst.org/blog/introducing-new-york-rural-economics/.

Shaw, S. L., & Kurti, D. (2020). Understanding the new human dynamics in smart spaces and places: Toward a spatialial framework. Annals of the American Association of Geographers, 110(1), 339–353.

Shen, Y., & Karimi, K. (2016). Urban function connectivity: Characterization of functional urban streets with social media check-in data. Cities, 55, 9–21.

Shen, Y., Karimi, K., Law, S., & Zhong, C. (2019). Physical co-presence intensity: Measuring dynamic face-to-face interaction potential in public space using social media check-in records. PLoS One, 14(2), Article e0212004.

Sousa, D., Sarmento, L., & Mendes Rodrigues, E. (2010). Characterization of the Twitter squees network: Are users tie social or topical? In Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents (pp. 63–70). New York, NY, USA.

Stefanidi, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial data from social media feeds. GeoJournal, 78(2), 319–338.

Stock, K. (2018). Mining location from social media: A systematic review. Computers, Environment and Urban Systems, 71, 209–240.

Sui, D., & Sui, S.-L. (2018). Human dynamics in smart and connected communities. Computers, Environment and Urban Systems, 72, 1–3.

Sui, D., & Sui, S.-L. (2021). Outlook and next steps: Understanding human dynamics in a post-pandemic world—beyond mapping Covid-19 in space and time. In Mapping covid-19 in space and time (pp. 347–358). Springer.

Sridhar, S., Doshi, P., Monley, E., & Zhong, C. (2018). Using mobility data as proxy for measuring urban vitality. Journal of Spatial Science Information, 16, 137–162.

Tang, Y., Zhang, Y.-Q., Chawla, N. V., & Krasser, S. (2008). SVMs modeling for highly imbalanced classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(1), 281–296.

Tuan, Y.-F. (1977). Space and place: The perspective of experience. University of Minnesota Press.

Tynon, A., Johnson, C., & Funk, C. (2020). U.S. public now divided over whether to get Covid-19 vaccine. Pew Research Center Science & Society. Retrieved 2022-07-14, from https://www.pewresearch.org/science/2020/09/17/u-s-public-now-divided-over-whether-to-get-covid-19-vaccine/.

Vanhoof, M., Hendrickx, L., Puusaar, A., Verstraeten, G., Ploetz, T., & Smoreda, Z. (2017). Exploring the use of mobile phone data for domestic tourism trip analysis. NetCom. Reseaux. Communication et Territoires, 31-3/4, 335–372.

Vogels, E. A. (2021). Some digital divides persist between rural, urban and suburban America. Pew Research Center Science & Society. Retrieved 2022-07-14, from https://www.pewresearch.org/fact-tank/2020/03/08/some-digital-divides-persist-between-rural-urban-and-suburban-america/.

Wellman, B. (2002). Little boxes, globalcization, and networked individualism. In M. Tanehe, P. Van Den Besselaar, & T. Ishida (Eds.), Digital Cities &: Computational and Sociological Approaches (pp. 10–26). Berlin, Heidelberg: Springer. Berlin, Heidelberg.

Yin, J., Soliman, A., Yin, D., & Wang, S. (2017). Developing urban boundaries from a mobility network of spatial interactions: A case study of Great Britain with geo-located Twitter data. International Journal of Geographical Information Science, 31(7), 1293–1313.

Yu, H., & Shaw, S.-L. (2008). Exploring potential human activities in physical and virtual spaces: A spatio-temporal GIS approach. International Journal of Geographical Information Science, 22(4), 409–430.
Yuan, X., Schuchard, R. J., & Crooks, A. T. (2019). Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. *Social Media + Society, 5*(3), Article 2056305119865465.

Zhao, Z.-D., Huang, Z.-G., Huang, L., Liu, H., & Lai, Y.-C. (2014). Scaling and correlation of human movements in cyberspace and physical space. *Physical Review E, 90*(5), Article 050802.

Zhou, X., Coiera, E., Tsafnat, G., Arachi, D., Ong, M.-S., & Dunn, A. G. (2015). Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Studies in Health Technology and Informatics, 216*, 761–765.