Check for
updates

Sampling Twitter users for social science research: evidence from a systematic review of the literature

Paula Vicente[1]

Abstract

All social media platforms can be used to conduct social science research, but Twitter is the most popular as it provides its data via several Application Programming Interfaces, which allows qualitative and quantitative research to be conducted with its members. As Twitter is a huge universe, both in number of users and amount of data, sampling is generally required when using it for research purposes. Researchers only recently began to question whether tweet-level sampling—in which the tweet is the sampling unit—should be replaced by user-level sampling—in which the user is the sampling unit. The major rationale for this shift is that tweet-level sampling does not consider the fact that some core discussants on Twitter are much more active tweeters than other less active users, thus causing a sample biased towards the more active users. The knowledge on how to select representative samples of users in the Twitterverse is still insufficient despite its relevance for reliable and valid research outcomes. This paper contributes to this topic by presenting a systematic quantitative literature review of sampling plans designed and executed in the context of social science research in Twitter, including: (1) the definition of the target populations, (2) the sampling frames used to support sample selection, (3) the sampling methods used to obtain samples of Twitter users, (4) how data is collected from Twitter users, (5) the size of the samples, and (6) how research validity is addressed. This review can be a methodological guide for professionals and academics who want to conduct social science research involving Twitter users and the Twitterverse.

Keywords Sampling plan · Social science research · Twitter · User-level sampling

1 Introduction

Social science research is the activity of gathering, analyzing, and interpreting information to understand the why, when, where, what and how of social relationships between humans and their interactions within society. Surveys are a key method when conducting cross-sectional quantitative research in the social science field and permit the study of

✉ Paula Vicente
   paula.vicente@iscte-iul.pt

1   Business Research Unit (bru_ISCTE), ISCTE-Instituto Universitário de Lisboa, Av. Forças Armadas, 1649-026 Lisboa, Portugal

populations that are too large to observe exhaustively and that have individual people as the unit of analysis (Babbie 2017, p. 270). Surveys involve the administration of standardized questionnaires to a sample of respondents for the systematic collection of a wide variety of unobservable data. It is extremely important to select a truly representative sample from the population of interest so that the inferences derived from the sample can be generalized back to the population.

Nowadays, surveys are up against the significant challenges of declining response rates and increased measurement errors, both with consequences on the validity of study results. The low response rate is due in part to the response burden; more specifically, surveys are reportedly boring, redundant, and frustrating for the respondents, and questionnaires are too long with questions people do not want to answer. Regarding measurement error, there is increasing concern about the quality of data because many respondents tend to avoid negative opinions and try to portray themselves in a socially desirable manner, which impedes the researchers from getting truthful responses and prejudices the validity of the outcomes of the survey research. Additionally, responses to survey questions often depend on subjects' motivation, memory, and ability to respond. Particularly when dealing with events that happened in the distant past, respondents may not have a clear recollection of their own motivations or behaviors or perhaps their memory of the events has evolved with time and is no longer retrievable (Bhattacherjee 2012, p. 90; Couper 2013; Salganik 2018, p. 86). These weaknesses make it pertinent to try out novel approaches to social science research.

Social media provide exciting opportunities that can ''open up a new era'' of social science research (Salganik 2018, p. 2). The ability to aggregate vast amounts of digital traces of human behavior through social media platforms represents a new data collection paradigm for social science research (Salganik 2018, p. 13). Powerful computational resources combined with the availability of massive datasets have given rise to a growing body of work that uses a combination of machine learning, natural language processing, network analysis, and statistics for the measurement of population structure and human behavior on an unprecedented scale (Gilbert 2010; Stieglitz et al. 2014).

Twitter is one of the social media platforms that social scientists rely on to conduct research. With more than 400 million active monthly users that post 500 million tweets per day (Statista 2022c), Twitter is a huge database—both in number of users and amount of data—for conducting large-scale studies of human behavior. Twitter allows access to its data via several Application Programming Interfaces (APIs). Whereas academic researchers generally rely on one of these freely available data sources, social analytics industries and government entities buy in to get elevated access, e.g., to 10% of the overall Twitter data, also known as the Decahose API (Twitter 2022c). However, analyzing Twitter data to describe human behavior is complicated by several challenging factors, related above all to the representation of human populations and human behavior, as well as by methodological issues, such as coverage (e.g., Gayo-Avello 2011), measurement (e.g., Cohen and Ruths 2013; Lazer et al. 2014) and generalizability (e.g., Tufekci 2014).

Almost all of the analyses relying on Twitter depend on access to samples retrieved by resorting to a free or costly API. As Twitter does not reveal details about how sampling is handled by their APIs, the use of Twitter data is regarded as highly problematic from a data quality point of view, especially in the social sciences (Pfeffer et al. 2018). The prevalence of sampling means that researchers need, on one hand, to understand platform mechanisms and possible biases in the resulting data, and on the other hand, to question research validity—Is the data retrieved representative and reliable for phenomenon analysis?. There is evidence that the sampling strategy affects not only the representativeness of

the data collected but also the substantive conclusions of analyses (Rafail 2018). Moreover, it is only recently that researchers have begun to question whether tweet-level sampling—in which the tweet is the sampling unit—should be replaced by user-level sampling—in which the sampling unit is the user (Zhang et al. 2018). The major rationale for this shift is that tweet-level sampling does not consider the fact that some core discussants on Twitter are much more active tweeters than other less active users, which means that discussion trends are dominated by these active users. On the other hand, user-level sampling can incorporate users' sociodemographic characteristics as well as their engagement level on social media, thus providing more accurate insights into the phenomena under study (Zhang et al. 2018).

Though scarce, sound knowledge on how to design and implement a user-based sampling plan to obtain a sample of Twitter users with the desirable properties of representation and replication is of indisputable relevance, as reliable and valid information can only be achieved when based on methodologically sound research. Berzofsky et al. (2018) conducted research on the efficiency of probability sampling, but the focus was on a narrow population (persons aged 14–21). Rafail (2018) discusses the impact of non-random sampling strategies for different types of Twitter populations, and Hino and Fahey (2019) propose a strategy to archive a representative database of Twitter data by sampling Twitter accounts. However, a sampling plan is more than just the choice of a sampling method. It comprises a set of interrelated stages that includes population definition, constructing the sampling frame, decision about sample size, and choice of a data collection strategy.

The present work seeks to contribute to this topic by conducting a systematic literature review of sampling plans designed and executed to obtain samples of Twitter users for social science research. The remainder of the paper is structured as follows: Sect. 2 provides a literature review and a summary of related work; Sect. 3 describes the methodology used for the systematic literature review; Sect. 4 presents the results obtained; finally, Sect. 5 discusses the results and presents the conclusions.

2 Background

2.1 Surveys sampling plan

The sampling plan is the process used to select a sample of units from the population for the purpose of making measurements and inferences about that population (Groves et al. 2009, p. 42). Figure 1 presents the typical sampling plan adopted in surveys. Despite the linear sequence suggested by the diagram, a change in the order of the stages or even the simultaneity of two stages is not uncommon since decisions are strongly interrelated (for

Fig. 1 Sampling plan stages in surveys
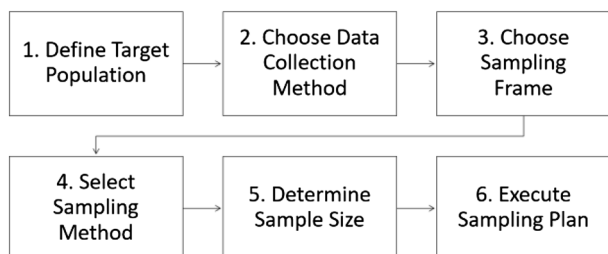
example, the choice of the sampling frame and the sampling method, or the choice of the sampling method and the data collection method). A brief discussion of the issues in each stage is presented below.

Social science researchers interested in estimating the prevalence of a certain condition in the population would start by formally defining the target population, i.e., the set of statistical units about which information is to be sought.

Regarding data collection, the questionnaire is the key instrument used in surveys and it can be administered in person, via fixed or mobile phones or online. The adoption of multiple modes of data collection has become increasingly popular. Both concurrent and sequential multi-mode systems can be used, such as (De Leeuw 2005): (i) One sample, one time period, one questionnaire but different modes for different sample persons; (ii) One sample, one time period, but different modes for different parts of the questionnaire (for the same person); (iii) One sample, multiple time points but the same persons measured with different modes at different time points, and (iv) Different samples, different modes, sometimes even different times and questionnaires.

Another stage in the sampling plan is the target population enumeration, which can be accomplished with a list of names, numbers, or other representation of the target population. That population enumeration is called the sampling frame because it is the list of units from which the sample is drawn. A perfect sampling frame is one in which each statistical unit in the population is separately listed once, only once, and no other irrelevant or extraneous elements are listed. However, not all sampling frames are perfect, and it takes effort and attention to review potential sampling frames to ensure that they are free from error or that the errors in the frames can be addressed. The sampling frame is ideally a full census of the target population but, in practice, it is typically a subset. When a material list does not exist, a conceptual list can be created by randomly generating numbers representing the units of the target population. Random Digit Dialling is an example of a conceptual sampling frame, built by randomly generating phone numbers, frequently used in telephone surveys (Waksberg 1978).

Sampling methods for drawing samples in the context of social sciences can be separated into random (emphasizing the representative relationship between the entire population and the selected sample), and non-random. In the first group, sampling eliminates the element of personal choice in selection and therefore removes subjective selection bias. Random sampling strengthens the belief that the sample is representative of the target population because there is an absence of selection bias. Non-random methods are a set of techniques where the researcher has some element of choice in the process. As inclusion is determined by a subjective criterion, it can be more difficult to extrapolate whether the sample accurately represents the larger population than when random sampling is used. The following are examples of non-random methods. Convenience sampling is an approach where units are selected based on the ease of access to the available group. Purposive sampling is an approach where units conform to certain predefined criteria for selection. In snowball sampling, the researchers usually start with a small number of initial contacts (seeds) who fit the research criteria and are invited to become participants within the research. The willing participants are then asked to recommend other contacts who fit the research criteria and who might also agree to be participants, who then in turn recommend other potential participants, and so on. Sampling usually finishes once either a target sample size or saturation point has been reached. Quota sampling is used to improve the representativeness of a sample to ensure that the relevant characteristics are present in sufficient quantities. If the sample has the same distribution as the target population, then it is likely to be representative of the target population (Babbie 2017, p. 198).

Determining the sample size entails considering several factors, namely: (i) the type of analysis being conducted—there are statistical procedures (e.g., regression analysis) that require a certain number of observations per variable—, (ii) population diversity—if the target population exhibits large variability in the behaviors and attitudes being researched, a larger sample is needed—, (iii) the tolerance for risk and the level of desired precision in the project—the more precision and the less risk of error we are willing to take, the larger the sample size must be—, and (iv) the available budget.

Finally, the execution of the sampling plan comprises implementing what was planned and designed in previous stages. In the end, we have a sample of respondents from which the data needed for the research project is, hopefully, successfully obtained.

2.2 Populations in Twitter data

Three distinct statistical units are observed in Twitter data and they correspond to three distinct, but interrelated, populations (Eurostat 2018): the population of tweets, the population of accounts and the population of users. A tweet is a post on Twitter containing up to 280 characters that can include URLs and hashtags. The tweet may also include up to 4 photos, a GIF, or a video. Users must sign up to the platform to use it, and this can be done by creating a free account or buying an account. To keep an account active, the user must log on to the platform at least every 6 months (Twitter 2022d). One user can have several accounts (e.g., Joe Biden has his personal account—@JoeBiden—and an institutional account as President of the United States—@POTUS), and one account can have multiple users (e.g., business accounts may allow different staff members to access and post contents). Users may subscribe to other users' tweets, which is known as "following" and subscribers are known as "followers".

Twitter Inc. defines tweeting as a behavior "performed by people" as Twitter was created to be "a service for friends, family, and co-workers to communicate and stay connected through the exchange of quick, frequent messages." (Twitter 2022e). However, the term "Twitter user" has expanded and today one can distinguish between two broad categories of users: (i) real-users, and (ii) digital-actors. Real-users represent human-beings, and digital-actors represent automated computer programs. Real-users can be described as (Uddin et al. 2014):

- *Personal users* casual home users who create their Twitter profile for fun, learning, or to acquire news, etc. These users do not strongly advocate any type of business or product, and their profiles are not affiliated with any organization. Generally, they have a personal profile and show a low to mild behavior in their social interaction.
- *Professional users* these are home users with professional intent on Twitter. They share useful information about specific topics and get involved in healthy discussion related to their area of interest and expertise. Professional users tend to be highly interactive; they follow many and are followed by many.
- *Business users* these users are different than personal/professional users in that they follow a marketing and business agenda on Twitter. The profile description strongly depicts their motive and a similar behavior can be observed in the way they tweet. Frequent tweeting and less interaction are two key factors that distinguish business users from both personal and professional users.

Digital actors are usually characterized by highly frequent tweeting, less or no inter-activity, and their followers generally either increase (e.g., in case of feed/news users) or decrease (e.g., in case of spam users) over time. This type of user can be grouped into three different classes (Uddin et al. 2014):

- *Spam users* spammers mostly post malicious tweets at a fast rate, automated computer programs (bots) run behind a spam profile, and randomly follow users, expecting a few users to follow back. Sometimes, personal users can also behave as spammers, but they seldom get caught because their spamming behavior does not follow a pattern, which can be easily seen in the case of an automated spam profile. Moreover, followers of spam users decrease over time.
- *Feed/news* these profile types represent automated services that post tweets with information taken from news websites such as CNN, BBC, etc. or from different RSS (Really Simple Syndication) feeds. Like spammers, tweets posted by these profiles are often controlled by bots. The key difference between spammers and these profiles is the increase in the follower count over time. Moreover, these users are not interactive at all (i.e., zero replies).
- *Viral/marketing services* viral marketing, or advertising, refers to the marketing tech-niques that marketers use with the help of technologies/social networks to increase their brand awareness or sales, or to achieve other marketing objectives. People use a viral process, which is an advanced type of bot (i.e., an intelligent bot that spreads informa-tion and produces fake likes, followers, etc.), to accomplish their marketing tasks.

2.3 Twitter data retrieval

Researchers' interest in Twitter has been increasing partly because the platform makes its data easily available: any Twitter user can apply for a Developer Account which allows him/her to gain access to Twitter data using an API. Although Twitter itself offers APIs to allow the public to access their data, changes have been made to the terms of service to become more restrictive, especially since the Cambridge Analytica scandal (Bruns 2019). Whereas it was previously relatively straightforward to register as a Twitter developer and request one or more of the authentication keys required for Twitter apps that access the public Twitter API, the company now imposes a registration process requiring develop-ers "to provide detailed information about how they use or intend to use Twitter's APIs so that we can better ensure compliance with our policies. (…) Applications submitted with incomplete or insufficient information may be delayed while we request further infor-mation from a developer. Applications that do not comply with Twitter's policies will be rejected." (Roth and Johnson 2018).

Nevertheless, Twitter keeps two different families of APIs to allow the public to access their data: the Streaming API and the REST API. They differ in the functionality offered and the constraints on users (Bruns and Stieglitz 2013; Schwitter and Liebe 2020):

- *Streaming API* Twitter data can be accessed as a constant real-time data stream. It gives access to (a sample of) all tweets as they are published on Twitter. After sending a cer-tain request to the API, specific data will be sent to the user continuously. The Stream-ing API only sends out real-time tweets. The Streaming API needs a search term to filter the results, such as a hashtag, a specific user ID or a geographical area defined by coordinates. Public statuses that match one or more of the filter predicates are returned.

- *REST API* retrieves data through the Search API (searching for tweets containing certain words, using specific hashtags, etc.) and the User API (collecting a user's tweets, followers, etc.). In contrast to the Streaming API, which collects real-time data in an ongoing fashion, REST APIs are suitable for single searches of historical data. The most common cases include searching historic tweets and reading user profile information. The Search API retrieves past tweets that match the criteria within the search window available for Twitter searches (which covers a period ranging from a few days to several weeks depending on the frequency with which the search term occurred in recent tweets). Furthermore, rate limits must be considered. This API returns a maximum of 100 tweets per request with a limit of 180 requests in 15 min. Regarding user profile information, Twitter allows the collection of up to 3200 of the most recent tweets from a user's individual timeline if the developer account accessing the API has access to that user (this either means the user must have a public profile or that the developer follows that user; in both cases, the developer cannot have been blocked by the user).

Although any registered researcher can access Twitter data for free via Twitter's public API, there are companies licensed to access Twitter data who sell databases containing tweets and metadata at the request of researchers. Additionally, there are premium Twitter APIs to access historical archives of tweets that are only available on payment. The pricing however is prohibitive for most research and academic institutions (Bruns 2019).

All Twitter APIs that return tweets provide data and metadata. Data and Metadata are both forms of data, but they have different uses and different specifications. Data is simply the content that can provide a description, measurement, or even a report on anything relative to a person, event, or topic. Metadata is data about data. It means it is the description and context of the data which helps to organize, find, and understand data. Therefore, in addition to the text content itself, a tweet can have over 150 attributes associated with it such as (Twitter 2022b):

- "*Tweet Object*" the tweet object includes fundamental attributes such as *id, created_at,* and *text*. Tweet objects are also the 'parent' object to several child objects. Tweet child objects include *user*, *entities*, and *extended_entities*. Tweets that are geo-tagged will have a *place* child object.
- "*User Object*" the user object contains Twitter user account metadata that describes the Twitter user referenced. Users can author tweets, retweet, quote other users tweets, reply to tweets, follow users, be @mentioned in tweets and can be grouped into lists. In general, these *user* metadata values are relatively constant. Some fields never change, such as the user's *id* and when the account was created—*created_at.* Other metadata can occasionally change, such as the *screen_name,* display *name, description, location*, and other profile details. Some metadata frequently change, such as the number of tweets the account has posted *statuses_count* and its number of followers *followers_count*.
- "*Geo Object*" Tweets can be associated with a location, generating a tweet that has been 'geo-tagged.' Tweet locations can be assigned by using the Twitter user-interface or when posting a tweet using the API. Tweet locations can be an exact 'point' location—*coordinates*—or a Twitter place with a 'bounding box' that describes a larger area ranging from a venue to an entire region—*place*.

For years, users who chose to geotag tweets with any location, even somewhere as geographically broad as "New York City", also automatically gave their precise GPS

coordinates. Neither the users nor their followers would see the coordinates displayed on Twitter but the GPS information would still be included in the tweet's metadata and was accessible through Twitter's API. Twitter did not change this policy across its apps until April 2015. Users must now opt-in to share their precise location, but the GPS data people shared before the update still remains available through the API (Lapowsky 2019; Twitter 2022f).

2.4 Sampling the Twitterverse: a parallel with surveys

While many features of online data collection are similar to surveys, a number of unique properties and new features account for methodological differences. Surveys produce self-report data, i.e., data that comes from asking respondents to recall wide-ranging aspects such as personality traits, moods, thoughts, attitudes, preferences, and behaviors, retrospectively with limited scope. In fact, much of social science knowledge and theory is based largely on self-report data. Social media provide the opportunity to record personal expression and human interaction in real time and on a large scale, without people actively doing anything. When using surveys, researchers have comparatively few respondents but a great deal of control over what information respondents provide. Although respondents provide information of interest to the researchers under these conditions, the limited sample size may not produce enough variability to study less commonly observed phenomena in their entirety. Social media data, in many ways, are precisely the opposite. They are completely unsolicited but offer unprecedented volume and variability (McCormick et al. 2017). In addition, social media data is quite often unable to answer specific questions posed by researchers and data users. This is because social media data, unlike survey data, are "found data" not "designed data." (Couper 2013).

Despite their differences, both survey research and social media research face the challenge of data quality. Within the survey research tradition, the Total Survey Error (TSE) framework is the guidance for data quality taken by all researchers when designing and implementing the two dimensions of a survey: (a) the representation dimension, comprising all stages of the sampling plan, and (b) the measurement dimension, comprising all stages of the data measurement and collection process (Biemer 2010). As social media does not yet have a similar framework devoted to enumerating the error sources and the error generating processes, the data quality approach shaped by the TSE framework can be leveraged for this context (Biemer 2014; Callegaro and Yang 2018; Amaya et al. 2020). The sampling plan in Fig. 1 can be transposed to the Twitterverse, thus helping researchers to design and implement the process leading to a sample of Twitter users and the corresponding dataset. However, some adaptations must be made to address the main issues in a sampling plan, i.e., target population, sampling frame, sampling method, sample size and data collection design. Finally, a word must be said about research validity.

2.4.1 Target population

Figure 2 shows the relationship between the target population of a social science project (e.g., persons aged 18–69 years) and the population observed on Twitter.

One can distinguish between the internet population—those with access to the internet—, the population with social media, and the Twitter population. The social media population suffers from under-coverage since not all people in the target population are social media users, and the under-coverage is even greater for the Twitter population since not all

Fig. 2 From target to Twitter population (adapted from Eurostat 2018)
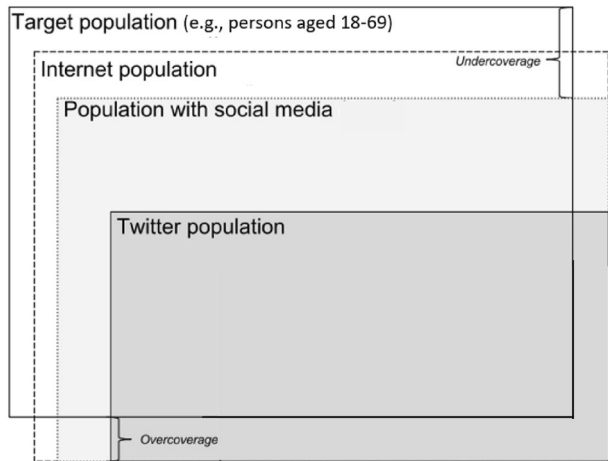
social media users are Twitter users. Likewise, over-coverage can occur to the extent that businesses and organizations maintain Twitter accounts (e.g., @mydeltaq or @FCGulbenkian). When the focus of the research is on individuals, businesses are not eligible and are a cause of over-coverage. Moreover, there are the digital-actors that do not correspond to human-beings and are for that reason out of the scope of studies targeted at populations of persons. Varol et al. (2017) estimate that the bot population on Twitter may range between 9 and 15%.

2.4.2 Sampling frames

In a study targeting the population of Twitter users, the ideal sampling frame would be a list of Twitter users—exhaustive, updated, and informative—from which a sample of tweeters is selected. However, Twitter, like other electronic populations, lacks a public membership list or a central registry like a phone directory (Andrews et al. 2003; Berzofsky et al. 2008; Couper and Miller 2008) and such a list is not made available by the platform either. As this makes it impossible to know the full population of Twitter users, creating a sampling frame to represent the general population of Twitter users is problematic. In such circumstances, researchers may resort to proxy populations for which a sampling frame is available, based on the presence or absence of critical attributes in the target population that must be mimicked (Lu and Franklin 2018), and from which a sample of Twitter users can be approached (e.g. a representative panel of the general adult population from which a sample of Twitter users is retrieved by imposing the condition "having an active Twitter account").

2.4.3 Sampling methods

Random selection of a sample of Twitter users can only be accomplished when a sampling frame is available. If not, the selection must be non-random. However, in both cases there is the risk of selectivity bias caused by the decisions of the Twitter users that impact the likelihood of selection. As noted in Fig. 2, Twitter users are a subset of social media users, who are a subset of Internet users, who are in turn a subset of the whole target population.

But it is not just a question of the percentage of users and non-users of internet, social media or Twitter: the frequency of usage and the level of proficiency in using these platforms are different across users and may have a non-ignorable effect on sample selection.

Figure 3 identifies three phases in the self-selection process present in Twitter data. Phase I and phase II are connected to the coverage error, i.e., the extent to which the frame population adequately covers the target population. As of January 2021, only 59.5% of the global population used the internet (Statista 2022a). According to DataReportal (2022), well over 9 out of 10 internet users now use social media but only 7% of social media users are Twitter users (Statcounter 2022). In this scenario, this restricted sampling frame may not be problematic if the research conclusions and the questions that generated them are limited to the Twitter universe. However, limiting the sampling frame to the users of Twitter in a study in which the interest is in generalizing findings to the offline population may raise issues of validity.

The population represented by Twitter data is Twitter users, who are disproportionately younger adults—more than 80% of users are under the age of 50 (Hootsuite 2020), therefore, if inferences are to be made to the general population, older adults will be underrepresented, and the parameter of interest will be over or underestimated, i.e., the outcomes will be biased (Amaya et al. 2020). Beside the age bias, Twitter's audience tends to be more educated and wealthier than the rest of the population, at least in the UK and US (Blank 2017; Pew Research Center 2019). Additionally, the use of online social media, and Twitter, is dependent on the socioeconomic background of individuals and their skills to use, first, the Internet and, secondly, social media platforms (Hargittai 2020), which is likely to make the sample of users of social media and of Twitter unbalanced. Moreover, tweeters' profiles are not equal across countries: Pew Research Center (2019, 2021a) found no significant gender difference in Twitter users in the US although the Twitter's global audience skews male—62% male versus 38% female (Hootsuite 2020).

Phase III relates to any decision of individuals that affects their presence in the sample. Social media is self-selective in the sense that not everybody posts messages on social media platforms, and those who do, do so at varying rates, from an occasional message from time to time to many messages a day (Nielsen 2006). In the US, 53% of Twitter users access the platform only a few times a week or less, and a minority of extremely active tweeters produce almost all tweets: the top 25% of users (the top is measured by tweet volume) produces 97% of all tweets. It is also important to note that those 25% of users do not exclusively produce original tweets—80% of tweets from this group are either direct retweets (49%) or replies to other tweets (33%). Replies and retweets are also what
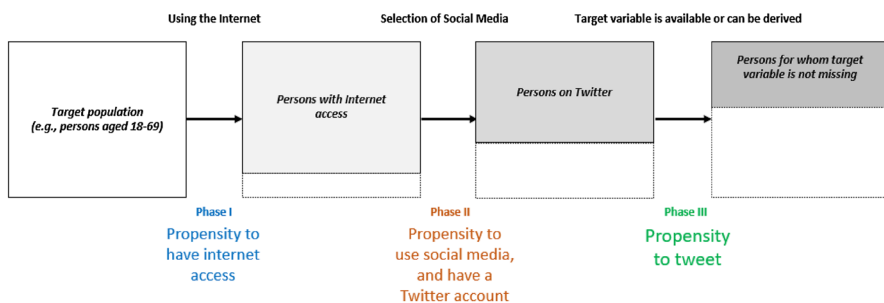

Fig. 3 Self-selection mechanism in Twitter (Adapted from Eurostat 2018)

less-active tweeters tend to do (Bruns and Stieglitz 2013; Pew Research Center 2021b). Moreover, there is evidence that the content generated by users who tweet very often and users who rarely tweet is significantly different, which should be a cause of concern when creating predictive models based on aggregated data (e.g., Mustafaraj et al. 2011). Individuals who never or rarely post information may be invisible to certain sampling techniques. Those users might be systematically under-sampled, and this process can bias the results of data collection toward the heaviest users. In addition, not everybody reveals the same kind of information in their profile—some biography profiles are information-rich, but others are not—which means that it is difficult to execute quota sampling or to assess sample bias when the value of a target variable and/or an auxiliary variable is missing (Daas et al. 2016; Eurostat 2018).

2.4.4 Sample size

The decision about sample size when sampling the Twitterverse raises less concern than in surveys. While cost and time constraints pose limits on the number of subjects in the sample in surveys, these kinds of limitation do not exist (or are not so prominent) in social media. Social media generates datasets with millions of users or millions of data points about thousands of users and the access to these resources is wide and easy in most cases (there may be exceptions due to the API system used, for instance). It might be thought that results can be generalized due to the sheer number of observations. However, this can be a mistake, depending on methodological details independent of sample size. Numbers are not the only factor to consider when establishing the generalizability of a study and bigger is not always better (Anderson 2008; Hargittai 2015).

2.4.5 Data collection

The data that is generally of interest to social science researchers when investigating the Twitterverse is posted in the platform, i.e., the text of the tweets, or images, or files, uploaded and shared, or even metadata associated with tweets or users. This data is accessed using API systems, but this poses significant problems with regard to access to representative, high-quality data for analysis. Cheap, publicly available data such as that obtained from Twitter's public APIs is often of low quality, while high-quality data is expensive both financially and computationally. The free Streaming API provides only real-time data, is limited to about 1% of Twitter traffic and does not provide a representative sample of tweets in many cases (Hino and Fahey 2019). In addition, Twitter is quite often unable to answer specific questions posed by researchers and data users—because it is not "designed data" (Couper 2013)—which makes resorting to other data sources mandatory to compensate for the weaknesses of the Twitter datasets (Callegaro and Yang 2018).

2.4.6 Research validity

Researchers must scrutinize the research undertaken on Twitter for a variety of possible methodological pitfalls. The design and execution of the sampling plan largely determine the validity of the research findings, above all the external validity. Checking external validity requires focus on the ways in which the findings may not represent the broader population or context (Olteanu et al. 2019), therefore social scientists must inquire: (a) to

what extent does an artificial situation like Twitter correctly reflect a broader real-world phenomenon?; (b) to what extent can the effects observed on the Twitter platform manifest differently on other platforms due to different functionalities, communities, or cultural norms?; (c) to what extent do the chosen sampling method and sample size affect sample representativeness, and (d) to what extent may constructs change over time and invalidate previous conclusions about societal and/or platform phenomena? The last question is only relevant in longitudinal studies but the other three are mandatory in all research.

3 Method

3.1 Research questions

The main goal of this systematic literature review is to disclose the design and execution of sampling plans in the context of applied social science investigation on Twitter. For this purpose, the main research questions were as follow:

RQ1   How is the target population defined?
RQ2   Which sampling frames exist and are used to represent the target population?
RQ3   Which sampling methods and sampling strategies are implemented to select Twitter users?
RQ4   How is data collected from Twitter users?
RQ5   What is the size of the samples?
RQ6   How is research validity addressed?

Additionally, some further information was extracted from the studies to complement the results:

- The purpose of the studies (exploratory, descriptive, or explanatory)
- The domain of the studies (e.g., political science, sociology, …)
- The geographic spread of the studies

3.2 Search strategy

A standard systematic quantitative literature review methodology (Pickering and Byrne 2014) was followed with the aim of identifying all the relevant articles published in peer-reviewed journals in the subject area of social science research involving samples of Twitter users. Books, theses, dissertations, editorials, and conference papers were excluded. The searches covered any part of the paper (title, abstract and full text).

3.2.1 Databases searched

Eight databases were searched, including some of the main databases for social sciences. Specifically, the search was performed in: ACM Digital Library, Oxford Academic, Sage, Science Direct, Scopus, Springer, Taylor and Francis, and Web of Science.

3.2.2 Search terms

To perform the search in the databases, the focus was set on two main terms of interest: "Twitter" and "sampling". A preliminary search based solely on these two terms yielded more than 20 thousand articles in the set of eight databases. An analysis of the first twenty results retrieved in each database revealed that most of the publications were off the goal, and further filtration was needed to ensure that included articles serve the objectives of this survey. Therefore, the search terms were reformulated to be more specific. This was done by using the term "sample of Twitter users" and several alternative terms to define the search queries, namely:

- "sample of twitter users" OR "sample of tweeters"
- "sampling twitter users" OR "sampling tweeters"
- "selection of twitter users" OR "selection of tweeters"

Additionally, inclusion/exclusion criteria were employed, namely:

Inclusion criteria (i) applied research papers containing evidence of Twitter users (i.e., individual persons) as the target population, and (ii) papers with the full text available (institutional access).
Exclusion criteria (i) methodological papers, i.e., studies that despite resorting to sampling had the main purpose of investigating computational or engineering issues (e.g., network analysis, bot detection, deep learning applications) and (ii) literature reviews.

Although no time restrictions were set, the papers are necessarily posterior to 2006, the year when Twitter was launched.

3.3 Data analysis

For each paper selected for the literature review, several pieces of information were recorded on a database. First, general information was extracted to contextualize each paper, namely paper title, journal title, publication year, author(s) affiliation (i.e., university, organization, …), geographical information of author(s), publisher, keywords. Then data was collected on each of the research questions. When available, any additional information providing a more in-depth review of the application of sampling strategies in the Twitterverse was collected. The data collected for the selected studies was classified according to the following criteria:

- The characteristics adopted to define the target population (addressing RQ1)
- The lists of Twitter users adopted to select samples (addressing RQ2)
- Sampling methods (random vs non-random) and sampling strategies (addressing RQ3)
- Modes of data collection and data collection strategies (addressing RQ4)
- The number of users selected (addressing RQ5)
- The discussion about the generalizability of the findings and the strategies to overcome limitations (addressing RQ6)
- The domain of the research, evaluated through corresponding author's affiliation (university, faculty, department, research centre)

- The purpose of the research
- The country of the corresponding author

4 Results

4.1 Papers identified by the search

Papers were retrieved in March 2022 using the abovementioned search strategy. This first search identified 196 papers. After removing duplicates, the title and abstract were scanned: 20 papers were excluded either for not meeting the inclusion criteria (7 could not be fully accessed, 11 were not related to the Twitter platform) or for meeting the exclusion criteria (2 were literature reviews).

An additional full text review—with special focus on the Method sections—was performed to ensure the suitability of the papers. In this final review, 67 papers were excluded in line with the exclusion criteria: 39 methodological papers (not applied research); additionally, 23 studies did not meet inclusion criteria for having as the target unit organisations/companies or the tweets instead of the user (individual). Moreover, a decision was made on excluding 5 papers using experimental designs since sample selection in these cases is addressed differently from that of cross-sectional studies.

The final sample consists of 73 papers. Figure 4 summarizes the full selection process.

Table 1 shows the total number of papers identified in the search and that were retained for analysis from each database considered.

4.2 Papers' spread

Most of the papers are published by corresponding authors from the United States (34), followed by United Kingdom (7), Hong Kong, Japan, and Spain (4 paper each), Germany and Israel (3 papers each), Belgium, and Chile (2 papers each) and with a single publication from Australia, Canada, China, Indonesia, Italy, Kingdom of Saudi Arabia, Singapore, South Korea, Switzerland and United Arab Emirates.

Most of the corresponding authors (68) are affiliated to universities, and the areas or domains of the respective faculties, schools, or departments are mostly Information & Communication (16 studies), Information & Computer Science (11 studies), and Social & Behavioral Sciences (11 studies) (Fig. 5). It was not possible to identify the area/domain of the corresponding author in 10 papers due to a lack of information in the affiliation description.

Regarding the year of publication of the selected papers, Fig. 6 shows the number of papers selected for the literature review for each year of publication. The figure reveals an upward trend in the number of papers published involving applied research on the Twitterverse with samples of tweeters. Despite the low frequency in 2017 vis-a-vis the previous years, there is evidence of increased interest in Twitter as a resource among social scientists. The slight decrease in 2020–2021 may be the consequence of the COVID-19 pandemic which affected all activities. The 2022 bar covers only the 1st quarter of the year.
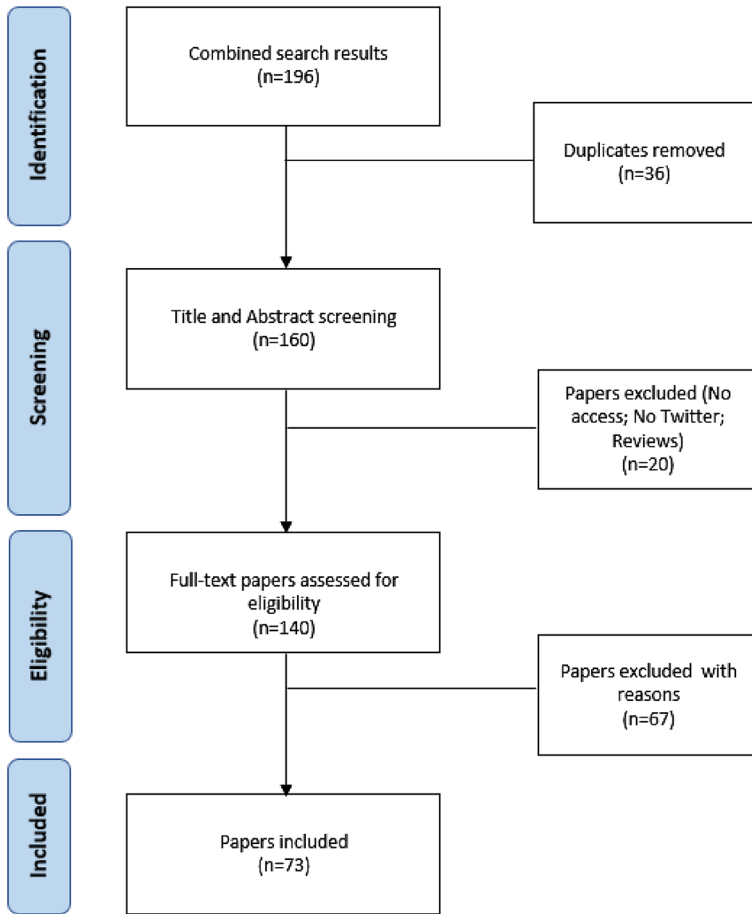
Fig. 4 Flow chart of paper selection process for the systematic literature review

Table 1 Number of papers identified in search and retained from each database

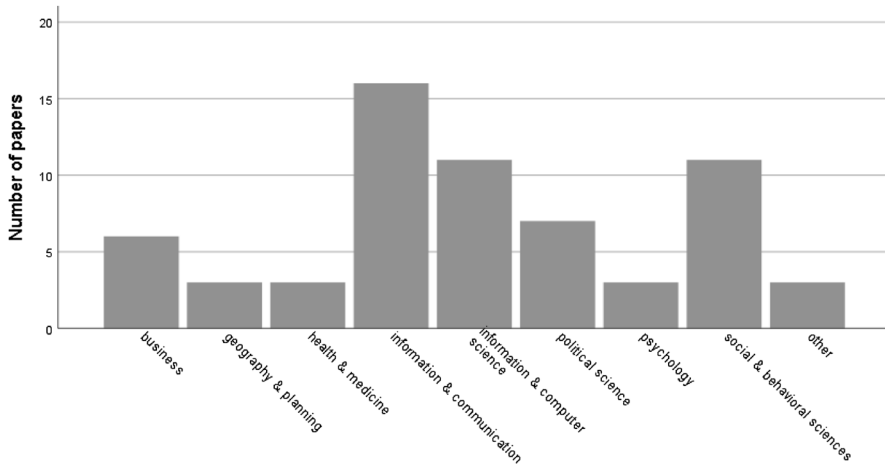| Database	Number of papers identified in search	Number of papers retained
ACM	9	1
Oxford Academic	14	2
Sage	34	17
Science Direct	51	20
Scopus	26	5
Springer	14	3
Taylor and Francis	22	14
Web of Science	26	11
Total	196	73

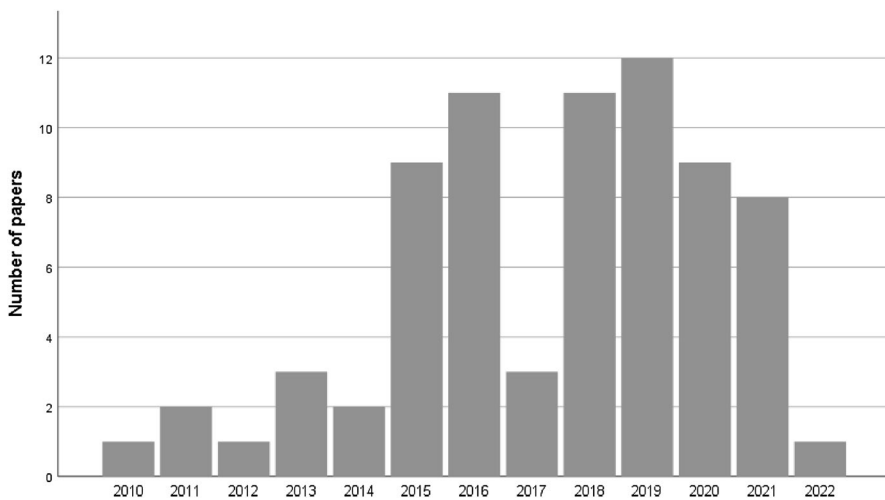**Fig. 5** Scientific areas or domains of corresponding authors

Fig. 6 Number of selected papers per year of publication

4.3 Main purpose of the studies

The purpose of a study is a statement of the goal or why the study is being conducted. After considering all qualified papers, their purpose was mapped into 3 categories: exploratory, descriptive, and explanatory (Babbie 2017, p. 97):

- *Exploratory* information gathering aimed at discovering, uncovering, exploring the phenomena.
- *Descriptive* information gathering aimed at describing, summarizing the phenomena.

- *Explanatory* understanding and explaining the phenomena by looking at the relationships between them, and patterns among variables.

The two main research purposes found in the reviewed papers are explanatory (29 papers, 40%) and descriptive (28 papers, 38%). Table 2 presents two example studies, with quotations from the respective paper text, for each purpose category.

4.4 Definition of the target population

This subsection responds to RQ1: How is the target population defined? Twitter users have the option to make their profile private; therefore, only the data of users without private profiles can be collected for research in any study involving the "population of Twitter users" with or without boundaries (Twitter 2022a).

Table 3 summarizes the features used to bound the target population and the number of studies that use each feature type and presents two examples for each.

Eighteen of the reviewed studies define the target population simply as the general population of "Twitter users". The remaining 55 studies use specific Twitterverse features to delimit the target population: tweet object features are found in 18 studies; user object features are found in 18 studies and geo object is the least used—only 4 studies delimit the population resorting to geographic criteria. Socio-demographic characteristics, such as age or country, or other specific characteristics (e.g., being a psychologist, being an asylum seeker) are used to bound the population in 23 studies. Age and region/country are the most frequently used (13 studies). Note that the definition of the target population involves more than one feature type in 7 studies; for example, in the case of Kobayashi et al. (2019) (Table 3) the population is bounded using a socio-demographic variable—country ("Japanese")—and a user object feature—"follows at least one media account and at least one member of the Japanese Diet."

Table 2 Categories of research purposes for the selected papers and two example studies

Purpose category	Definition of purpose (examples)
Exploratory	Adopting an **exploratory approach**, we distinguish between health institutions, specialists, and advocates, and we assess key topics and framings promoted online by these actors. (Reveilhac and Lupton 2022)
	The current study was **exploratory** in nature with the aim to isolate a group of self-identified parents (i.e., stay-at-home parents) and examine their publicly available tweets concerning discipline and spanking. (Lee et al. 2020)
Descriptive	This study focuses on Twitter use during an eminent Belgian current affairs television (TV) programme to investigate **how** people talk about TV on Twitter. (D'Heer and Verdegem 2015)
	The current study was designed to **determine the use patterns and characteristics** of African American, Hispanic and White young adult Twitter users who reported past month blunt use. (Montgomery et al. 2018)
Explanatory	In this study, (…) we contribute to the extant literature (…) by **investigating the impact** of Twitter trust on both the intentions to continue using the platform and the intentions to follow and purchase other brands that are "hosted" on the Twitter platform. (Pentina et al. 2013)
	The aim of this paper is to approach concurrently these **various factors that influence** users' behaviour on social media during crises. (Mohammed and Ferraris 2021)

Table 3 Feature types to delimit the population of Twitter users and two example studies

Feature types	Number of papers	Target population (examples)
No feature	18	This study's primary interest is in **Twitter users.** (Fiesler and Proferes 2018) This study examined the ideological media consumption patters of **Twitter users.** (Shin 2020)
Tweet Object	18	Twitter users who **tweet about depression** (Cavazos-Rehg et al. 2016) Twitter users (…) who **posted tweets with #NBCFail** (O'Hallarn and Shapiro 2014)
User Object	17	Japanese Twitter users who **follow at least one media account and at least one member** of the Japanese Diet. (Kobayashi et al. 2019) Twitter users with **3 or more tweets** supporting ISIS (Torregrosa et al. 2020)
Geo Object	4	Twitter users who **tweeted from Los Angeles County** (Lerman et al. 2018) Twitter users whose **geo-tagged tweets were sent from the catchment area** (Pourebrahim et al. 2019)
Socio-demographic or other specific features	23	(…) Twitter users **with schizophrenia** (Hswen et al. 2017) **Israeli** Twitter users (Moshkovitz and Hayat 2021)

4.5 Sampling frames

This subsection addresses RQ2: Which sampling frames exist and are used to represent the target population? Fig. 7 displays the distribution of the sampling frames adopted in the reviewed studies. The difficulty of building a sampling frame to list the Twitter users of the target population in an exhaustive and updated manner is evident. Most of the studies (46) were conducted without a sampling frame to support the sample selection of Twitter users.

In the remaining studies, which used a list to represent the target population, four different sampling frames can be distinguished:

- *Online panels* in 11 studies researchers resorted to research companies owning online panels (e.g., YouGov, Qualtrics), representative of the general adult population of a country, and requested a sample of Twitter users (e.g., Han et al. 2015; Sasaki et al. 2015).
- *Crowdsourcing services* services like Amazon Mechanical Turk or Yahoo! Crowd Sourcing Service are marketplaces for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform virtually any task. In 4 studies, researchers resorted to these services to request a sample of individuals interested in collaborating with the investigation (being a Twitter user was mandatory) (e.g., Tominaga et al. 2015; Fiesler 2018).
- *Customized list of Twitter users* in 8 studies the researchers built or acquired a list representing the target population of the study, such as a list of candidates to elections (Bouteca et al. 2017), a list of followers of 12 Twitter accounts (Kearney 2019) and a list of scientific tweeters (Yu et al. 2019).
- *Random Generated ID numbers* 4 studies adopted this kind of sampling frame, which consisted of creating a "list of Twitter users' IDs" using a software system that randomly generates numbers, within (or not) a pre-specified range of numbers (e.g., Liang and Fu 2019). This process resembles Random Digit Dialling in telephone surveys (Waksberg 1978) and allows everyone with an active Twitter account to be represented,
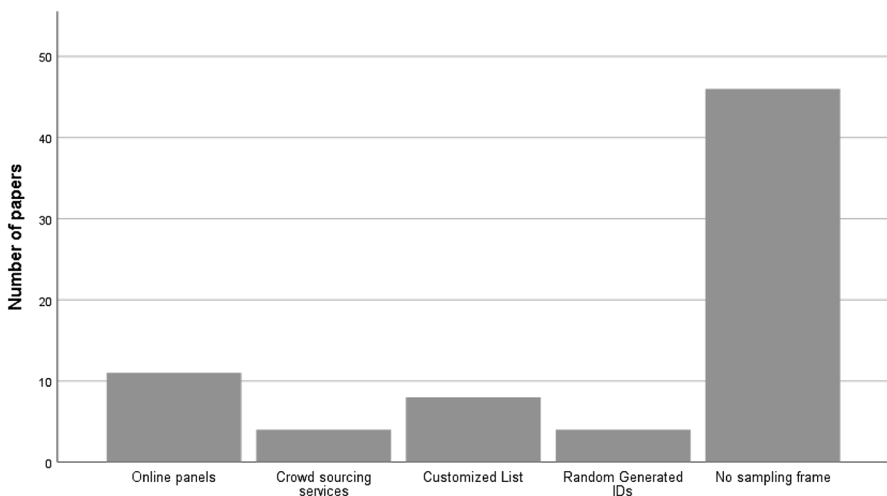

Fig. 7 Sampling frames adopted in the selected papers

rather than only those who are in a list acquired to serve as a sampling frame. As in Random Digit Dialling, this approach is suitable for general population studies, which in the Twitterverse means targeting the "general population of Twitter users".

Overall, most of the studies reviewed (63%) were conducted without a sampling frame and the remaining resorted to sampling frames that are external sources of the Twitterverse, either panels of research companies or recruitment from crowdsourcing platforms.

4.6 Sampling methods and strategies

This subsection addresses RQ3: Which sampling methods and sampling strategies are implemented to select Twitter users?

The sampling methods in the reviewed papers can be mapped into two distinct categories: (i) methods supported by a sampling frame, and (ii) methods designed and implemented without a sampling frame. In the first category, when the sampling frame is a panel or a list of crowdsourcing members, sample selection is designed by setting conditions that sample members must meet, such as "being a Twitter user with a public profile", "being a regular Twitter user". These conditions are necessary since neither the panels nor the crowdsourcing platforms are made up solely of Twitter users. The researcher has no control over or intervention in the selection process. In studies with customized sampling frames, the sample of Twitter users is obtained by random selection (e.g., Bouteca et al. 2017) or by setting filters such as number (minimum or maximum) of followers or number of statuses posted (e.g., Kearney 2019). Finally, when the sampling frame comes from randomly generated numbers, the sample automatically includes those Twitter users whose ID numbers match the randomly generated numbers (e.g., Liang and Fu 2015; Liang et al. 2016).

When no sampling frame was available, the sample of users was selected either by non-random sampling methods or resorting to an API system search. The non-random sampling methods found in the reviewed papers were grouped in three main categories:

- *Convenience sampling* focuses on gaining information from Twitter users who are 'convenient' for the researcher to access. In the reviewed papers, convenience sampling was adopted in 6 studies, using one of the following approaches: (i) selection via researchers' followers (e.g., Akyuz et al. 2021); (ii) selection via promoted or ad tweets (e.g., Montgomery et al. 2018); and (iii) selection of students on campus (e.g., Mohammed and Ferraris 2021).
- *Snowball sampling* this method is a kind of convenience sampling in which, after agreeing to cooperate, the Twitter users initially selected for the study are solicited to retweet the invitation to the study to their followers. In the revised studies, snowball sampling was adopted in 4 studies (e.g., Chen 2011; Pentina et al. 2013).
- *Purposive sampling* sample members are chosen based on the researcher's sound judgment of what is a representative sample. Only 3 studies employed purposive sampling: in Sharples (2021), the researcher intentionally chose a sample of four Twitter handles based on the frequency of posts; in Jünger and Fähnrich (2020), sample members are the followers of a specific Twitter account and in Brady (2016), Twitter users are chosen based on specified profile features.

A significant number of studies (31) designed and implemented sample selection based on an initial search in Twitter data (Table 4). Keywords, hashtags, or geo-tags were used

Table 4 Sampling methods and sampling strategies in studies without a sampling frame

Method or strategy	Number of studies
Convenience sampling	6
Snowball sampling	4
Purposive sampling	3
Search via API system→Twitter users	25
Search via API system→Twitter users + Random sampling	4
Search via API system→Twitter users + Snowball sampling	2

to query Twitter's API and identify those Twitter users who posted messages related to the topic or event under study. The Twitter users were pinpointed based on topic-related keywords or topical hashtags contained in the messages they posted during a delimited period. This strategy is used mostly in studies intended to target an audience narrowed down to those who tweeted during a certain event or regarding a certain topic, such as smoking (Hswen et al. 2017), health conditions (Zhang and Ahmed 2019), ISIS support (Torregrosa et al. 2020), and natural disasters (Bica et al. 2021). Instead of hashtags or keywords, an API search can focus on tweets from specific locations (Murty et al. 2016, Osorio-Arjona and García-Palomares 2019) or followers of specific accounts (Hayat et al. 2016). The strategy of reaching a sample of Twitter users by simply searching data via an API system was adopted in 25 studies (Table 4).

Another sampling strategy based on API systems consisted of locating an initial sample of Twitter users via an API search and, in a subsequent stage, selecting a random subsample of those users because the initial set was found to be too large for the analysis in question (e.g., Vaccari et al. 2016; Baik et al. 2021). In another two studies, the opposite strategy was found: after locating an initial sample of Twitter users via an API system search, snowball sampling was adopted to increase sample size since the initial sample was found not to be large enough for the study objectives (O'Hallarn and Shapiro 2014; Lee et al. 2020).

Table 4 summarizes the sampling methods and the sampling strategies adopted in the reviewed studies when no sampling frame is available. Note that two of these studies did not have enough information to allow a full understanding of how the sample was selected.

4.7 Data collection

This subsection provides answers to RQ4: How is data collected from Twitter users? Despite the richness of Twitter data, this data may not be enough for the study purposes, and it may be necessary to resort to other data sources to supplement and enrich the data gathered on the Twitterverse.

The reviewed papers include solely cross-sectional studies which means that data refer to "one time point". It is important to note that it is necessary to specify the search's time range when retrieving data and metadata from the Twitterverse. Although the time range of the studies varied from a few seconds to several months, the data is regarded as "one time point" as a trend evaluation of the phenomenon was never the study objective.

The data collection designs found in the reviewed studies are classified distinguishing between single-mode designs and mixed-mode designs. Three studies did not provide enough information to allow an understanding of how data collection was implemented.

4.7.1 Single mode, one sample

Under this design, data is collected for all sample members using a single mode of data collection. In the reviewed papers, this design was implemented using one of the following strategies:

- *Data collection via an API system* tweet texts and associated metadata are gathered for one sample of Twitter users resorting to an API system. The collection is guided by keywords, hashtags, or geo-tags to identify which tweets and tweeters are relevant for the study (e.g., Huang and Wong 2016; Storer and Rodriguez 2020).
- *Self-report data collection* data is collected from one sample of Twitter users resorting to inquiry modes, such as online surveys, or in-depth interviews (e.g., Visser et al. 2014; Álvarez-Bornstein and Montesi 2019).

4.7.2 Single mode, different samples

This kind of design uses the same mode of data collection for different samples or subgroups. Studies using this approach involve international and regional comparison. Data are collected either using an API system or a self-report mode (e.g., Hayat et al. 2016; Vaccari et al. 2016).

4.7.3 Mixed-mode

In the reviewed papers, mixed-mode designs were found in 22 studies. In all these studies, the mixed-mode design was implemented using a sequential approach. More than half of these studies combine a self-report mode − mostly surveys to gather data on Twitter users' subjective perceptions and opinions − in a first stage of data collection, with data collection via an API system in a subsequent stage. The motivation for mixed-mode designs is either to avoid biases inherent in self-reported data − the researcher gets objective data by collecting data via an API system −, or to allow richer data sets − social media data may not be enough for the study purpose (Salganik 2018, p. 118). These designs produce two matched databases. Specifically, the following strategies (for simplicity are named as A, B and C) were found:

4.7.3.1 Strategy A: self-report data followed by API search This strategy requires the researcher to conduct a survey on a sample of people that have a Twitter account and then request all survey participants to provide their Twitter ID. Upon acceptance of the participants, the researcher will retrieve their Twitter contents resorting to an API. This strategy can be resumed in the following stages:

- *First stage* Twitter users provide self-reported data by means of a survey questionnaire or in-depth interview.

- *Second stage* the members of the sample volunteer their usernames or IDs for researchers to access their Twitter account and retrieve data and/or metadata via an API system (e.g., Tominaga et al. 2018; Abney et al. 2019).

4.7.3.2 Strategy B: API search followed by self-report data This strategy starts with a search of the Twitterverse based on keywords, hashtags, or other search term related to the topic under investigation. Twitter users whose data is retrieved are then contacted via the Twitter platform soliciting them to participate in a survey (e.g., sending a message with the invitation and the link to the survey questionnaire). This strategy is common in projects that need to link respondents to their respective auxiliary information (e.g., gender, age, income, education). This kind of information is usually insufficient or is not reliable in big data sources (Daas et al. 2016) making it necessary to implement this kind of data collection strategy. This strategy can be resumed in the following stages:

- *First stage* Twitter data and/or metadata is collected for Twitter users via an API system.
- *Second stage* the same sample of Twitter users provide self-reported data by means of a survey questionnaire or in-depth interview (e.g., D'Heer and Verdegem 2015; Majmundar et al. 2019).

4.7.3.3 Strategy C: API search followed by secondary data This strategy is adopted in studies in which the geographic location of Twitter users when posting contents needs to be taken in consideration in the analysis. Secondary sources provide supplementary information about the geographical areas under research. This strategy can be resumed in the following stages:

- *First stage* Twitter data and/or metadata is collected for Twitter users via an API system.
- *Second stage* secondary data is collected, mostly from Census or Official statistics, at a spatial or regional (aggregated) level (e.g., Liang et al. 2017; Osorio-Arjona and García-Palomares 2019).

Figure 8 summarizes the designs of data collection in the studies covered in the reviewed papers.

In short, 48 studies (69%) relied on a single mode of data collection: 29 studies resorted to an API system applied to a unique sample of Twitter users while 14 studies (19%) resorted to a self-reported mode. Five studies involved more than one sample of Twitter users, but the mode of data collection was unique in each study. A non-ignorable number of studies (18 studies, 26%) combined self-reported data with data gathered via an API system collected for the same sample of Twitter users. Surveys, the most widely used method of data collection in social science research, appears in 50% of the studies either exclusively as the single mode of data collection or in combination with data gathering via API systems.
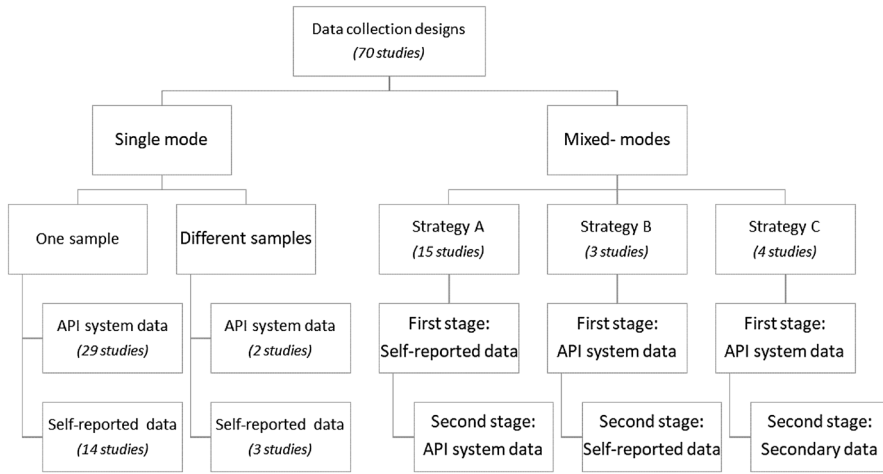
Fig. 8 Data collection designs and respective number of studies

4.8 Sample size

This subsection addresses RQ5: What is the size of the samples? Regarding sample size in terms of number of Twitter users in the reviewed studies, it was found that: 16 studies (22%) used fewer than 200 Twitter users, while 36 studies (49%) used between 200 and 2000 users. Seventeen studies (23%) used more than 2000 users (3 studies reported more than 1 million participants). Four studies did not report the Twitter user sample size.

The size of the sample is closely connected to the data collection strategy. Table 6 presents statistics summarizing the size of the sample of Twitter users splitting the studies into two categories: (i) "API system data" studies—includes studies in which the data collection design relies exclusively on API systems or involve gathering data via an API system in the first stage of a mixed-mode design;

ii) "Self-reported data" studies—includes studies in which the data collection strategy relies exclusively on self-reported modes or involve self-reported data gathering in the first stage of a mixed-mode design.

The sample sizes of the "self-report data" studies tend to be lower (average = 557 users, median = 330 users) than the sizes of the "API system data" studies (average = 297,893

Table 6 Descriptive statistics of Twitter user sample size by data collection strategy	Statistic	"API system data" studies (n = 34)	"Self-report data" studies (n = 32)
	Average	297,893	557
	Std. deviation	846,179.4	503.5
	Minimum	4	12
	1st quartile	346	181
	Median	2234	330
	3rd quartile	36,648	956
	Maximum	3,800,000	1,496

users, median = 2,234 users). The distribution of the sample sizes of the "API system data" studies is strongly skewed and highly dispersed which can be explained by the 6 outlier studies that have sample sizes of ≥ 340,000. Excluding these studies from the calculations, the average sample size of the remaining studies drops to 7,380 users (Std deviation = 16,505.8).

4.9 Research validity

This section addresses RQ6: How is research validity addressed? The discussion on the limitations of the study due to methodological shortcomings is only present in 49% of the reviewed papers. As mentioned in §2.4, the Twitter population is likely to suffer from coverage error since not all people use Twitter. Although all the reviewed studies identify their target or study population as Twitter users (either with or without boundaries) (§ 4.4), some warn of the non-generalizability of the findings to other populations and acknowledge that Twitter users may differ either from other social network users (11 studies) or from other general or specific off-line populations (12 studies). Explanations for these issues are provided mostly in the respective Discussion section.

The generalizability of the findings may also be compromised by a non-representative sampling frame, a sampling method that does not guarantee sample representativeness, a sample size inadequate to the study's objectives, and self-selection bias caused by the pattern of Twitter usage. Details about how the reviewed studies acknowledge these issues are presented as follows.

4.9.1 Twitter users differ from other network users

Table 7 presents the citations of the explanations given in the 9 studies that highlight the limitation that the outcomes cannot be generalizable beyond the Twitter network because Twitter users may differ from users of other platforms. The characteristics of Twitter users and the specificities of the Twitter platform are likely explanations for this issue in 4 studies (Gómez-Zará and Diakopoulos 2020; Liang and Fu 2019; Liang and Shen 2018; Tominaga et al. 2018). Three studies postpone investigating whether the findings are valid in other social media platforms for future research (Schaarschmidt and Könsgen 2020; Tominaga et al. 2018; Vaccari et al. 2015).

4.9.2 Twitter users differ from off-line or other specific populations

Fourteen studies highlight the limitation that the outcomes are not generalizable beyond the Twitter network because Twitter users may differ from off-line or other specific populations. Table 8 presents the citations of the explanations for this shortcoming.

4.9.3 Sampling frame lacks representativeness

Only 3 papers acknowledge that a sampling frame lacking representativeness may affect the study quality. Baik et al. (2021) mention that "(…) *without a full roster of all Twitter users (…) within a certain period, it is impossible to sample a random sample of Twitter users, which would have been the gold standard (…)."* The other two papers recognise the limitations of resorting to crowdsourcing platforms: Schaarschmidt and Könsgen (2020) refer that "(…) *this study included only U.S. employees recruited* via *a crowdsourcing*

Table 7 Papers and citations of the impact of differences between Twitter users and other social media users

Paper identification	Citation
Amin et al. (2019)	(…) given our work requires the users to have been active both on Twitter and Foursquare, the findings can only be generalized to the population of users who have joined both platforms.
Fiesler and Proferes (2018)	These findings with respect to Twitter may or may not be generalisable for other platforms or contexts [such as] Reddit, Tumblr, Instagram or Facebook.
Gómez-Zará and Diakopoulos (2020)	(…) the Twitter demographic may influence our results. (…) Our results may have been different considering other social media platforms and audiences' demographics.
Liang and Fu (2019)	Twitter is a social media platform focusing on information sharing. However, there are platforms emphasizing social networking functions, like Facebook.
Liang and Shen (2018)	Twitter could have very different features from other types of social media platforms, such as Facebook and Instagram.
Majmundar et al. (2019)	(…) these findings might not generalize to people with very high social media use, and/or to users of other social media platforms.
Schaarschmidt and Könsgen (2020)	(…) we tested our hypotheses for Twitter accounts only. Further research should scrutinize the observed results for other social media platforms.
Tominaga et al. (2018)	(…) there is still a possibility that characteristics of people depend on the platforms; therefore, we need to verify this issue in the future work.
Vaccari et al. (2015)	(…) Twitter is only one of several such platforms, it is crucial to investigate whether our conclusions can be generalized to broader populations than the one featured in this study.

platform." and Tominaga et al. (2018) recognise that "*(…) our study cannot address the issue because we gathered our subjects by crowdsourcing services in which people under 18-year-old cannot use the system.*"

4.9.4 Sampling method yields biased samples

Nineteen studies recognise that the sampling method is far from yielding a representative sample of the target population. In 7 studies researchers acknowledge that the sampling methods restrict or compromise the possibility of generalizing the outcomes beyond the scope of the study (Chen 2011; Fischer and Reuber 2011; Hswen et al. 2018b; Majmundar et al. 2018, 2019; Sasaki et al. 2015; Watson 2016). Two studies suggest that random sampling would be the gold standard of sampling in their research (Akyuz et al. 2021; Sashittal and Jassawalla 2019). Table 9 presents citations that illustrate the explanations found about this issue.

4.9.5 Insufficient sample size

Only five studies mention that sample size or the achieved response rate were not good enough to grant validity to the findings (Table 10). Two studies postpone replicating the research with bigger samples sizes for future research (Qiu et al. 2012; Shin 2020).

Table 8 Papers and citations of the impact of differences between Twitter users and off-line or other specific populations

Paper identification	Citation
Hswen et al. (2017)	(…) these findings may not generalize to Twitter users with private accounts or to people who do not use Twitter.
Hswen et al. (2018a)	(…) these findings may not generalize to Twitter users with private accounts or to people who do not use Twitter.
Hswen et al. (2018b)	(…) these Twitter users likely differ from individuals with schizophrenia who do not disclose their illness online or who do not use social media. (…) our findings likely do not generalize to individuals who do not use social media.
Jünger and Fähnrich (2020)	(…) limits the generalization of findings because the users of social media sites do not necessarily represent the whole group of interest.
Lee et al. (2020)	Our results cannot be generalized to the broader group of stay-at-home parents who do not participate in this form of social media.
Majmundar et al. (2019)	This sample comprises Twitter users with public profiles limiting generalizability to those with private accounts.
Mihunov et al. (2020)	(…) the results derived from this survey cannot be used to generalize to the general population including, for example non-Twitter users who requested for rescue or Twitter users who did not request for rescue using Twitter.
Mohammed and Ferraris (2021)	As the data were gathered only from students [with Twitter account], the results cannot be generalised to other populations.
Montgomery et al. (2018)	(…) findings from this study are not generalizable to other racial/ethnic groups.
Schaarschmidt and Könsgen (2020)	(…) we tested our hypotheses for Twitter accounts only. Further research should scrutinize the observed results for other social media platforms, such as Facebook, Xing, LinkedIn or Glassdoor.
Storer and Rodriguez (2020)	(…) this sample of Twitter users is not representative of the general population.
Storer et al. (2018)	(…) there are likely significant differences between Twitter users and the general population.
Thomas et al. (2019)	(…) the use of social media data, specifically Twitter, limits generalizability to the broader population.
Vaccari et al. (2015)	(…) we studied a specific population of individuals who posted at least one election-related message, and we cannot determine whether our findings can be generalized to other populations.

4.9.6 Self-selection bias

A small number of studies (3) acknowledged self-selection bias—caused by the fact that the pattern of Twitter use can negatively impact sample representativeness—as a potential problem for outcome validity. Baik et al. (2021) refer that "*one clear bias in using this method* [API system search] *is that any user that tweets more frequently will be more likely to be selected*.". Storer et al. (2021) stress that the behavior of Twitter users when in the platform is not uniform: "*individuals use this platform in a variety of different ways (i.e., consuming news, connecting with work colleagues,* etc.*) (…). This sample is generalizable only to those who respond to and follow social issues on Twitter.*" Hswen et al. (2017) mention that the information users chose (or not) to make public restricts the scope of the

Table 9 Papers and citations concerning sampling methods deficiencies

Paper identification	Citation
Akyuz et al. (2021)	A convenience sample allows anyone who wants to participate in the study to do so. A random sample would be expected to better represent the target population because neither the researcher nor the participants decide who will participate in the study.
Baik et al. (2021)	(…) one clear bias in using this [nonrandom selection] method is that any user that tweets more frequently will be more likely to be selected.
Cavazos-Regh et al. (2017)	(…) the "typical Twitter user" (…) is not an exact match of our sample of Tweeters.
Chen (2011)	(…) a nonprobability sample was obtained using convenience snowball sampling, (…) this sampling method does not offer generalizability of results (…).
Fischer and Reuber (2011)	(…) the sample [of entrepreneurs] we chose to use included only business-to-business firms, and no business-to-consumer firms. (…) This sample limitation does mean, however, that the theoretical insights should not automatically be assumed to generalize to all contexts.
Hswen et al. (2017)	(…) given that we generated a convenience sample of Twitter users who self-identified as having schizophrenia through manually searching the Twitter platform, these individuals also may not be representative of the broader group of Twitter users who self-identify as having a schizophrenia spectrum disorder."
Hswen et al. (2018b)	(…) we employed a convenience sampling approach to generate the group of Twitter users who self-identify as having schizophrenia This sampling method further limits generalizability of these findings.
Kobayashi et al. (2019)	(…) the respondents in the present study were not randomly sampled from the population (i.e., they were Japanese Twitter users who follow at least one media account and at least one member of the Japanese Diet), so we cannot preclude potential systematic sampling bias.
Lee et al. (2020)	(…) snowball sampling can be prone to bias because of the correlations between individuals. That is, Twitter users in the subsequent round of data collection are not independent of Twitter users in the first round of data collection, which could ultimately yield biased or inconsistent results. For these reasons, results of the study should be interpreted with caution.
Majmundar et al. (2018)	This sample comprises Twitter users with public profiles limiting generalizability to those with private accounts.
Majmundar et al. (2019)	(…) the study sample is non-representative of U.S. population, which limits generalizability of the findings.
Montgomery et al. (2018)	(…) the use of a convenience sample limits the ability to generalize the findings to the national population.
Pentina et al. (2013)	The method of snowball sampling could have introduced selection bias in the data collection.
Sasaki et al. (2015)	(…) our sample was limited to users in only one country (…) Therefore, we recommend caution in generalizing the results to users in other countries.
Sashittal and Jassawalla (2019)	Random samples of all Twitter users are left to future research.
Storer et al. (2018)	Twitter users are not a representative sample. (…) This sample is generalizable only to those who respond to and follow social issues on Twitter.

Table 9 (continued)

Paper identification	Citation
Visser et al. (2014)	This survey was originally disseminated via tweets from the authors' Twitter accounts, and our tweets were then retweeted many times by many people, including some influential educators. Due to the inability to track how many people received or read these tweets, the researchers were unable to obtain the precise numbers necessary to calculate and provide an exact response rate
Yu et al. (2019)	Because of the large number of scientific tweeters, it is difficult to analyze a statistically representative sample (…) so the results may overrepresent the highly active scientific tweeters.
Watson (2016)	Neither the journalists nor Twitter users in this study represent a scientific random sample (…). Findings should not be generalized beyond the Gulf Coast newspaper journalists and most followed Twitter users in the sample.

Table 10 Papers and citations concerning sample size insufficiencies

Paper identification (Authors)	Citation
Brady (2016)	The sample size for this project was small and posed a limitation to the strength of the results.
Gómez-Zará and Diakopoulos (2020)	Given the low response rate (…)
Majmundar et al. (2018)	(…) sample size of this study also limits findings' generalizability (…)
Qiu et al. (2012)	While our sample size is comparable to other studies on personality and social media (…) future studies should include more participants in order to verify our findings.
Shin (2020)	(…) the findings were drawn from a relatively small sample of Twitter users (…). Future studies are needed to expand the sample size to more accurately assess the demographics and media behaviors of Twitter users

findings: "*they may differ from individuals with schizophrenia who choose not to disclose their illness online*."

5 Discussion

The general information of the studies revealed that most of the research concentrated in the United States (34 studies) and is conducted in an academic context (68 studies). Most research falls into the domains of Information & Communication (16 studies), Information & Computer Science (11 studies), and Social & Behavioral Sciences (11 studies), which are key areas of social science and computational social science (Cioffi 2010; Investopedia 2021). Much of the research is either descriptive (28 studies) or explanatory (29 papers), which aligns with the pattern of social science investigation—describing the phenomenon under study and identifying its causes and effects are most the common research interests of social scientists (Babbie 2017, p.22). However, four of the reviewed studies acknowledge that the cross-sectional nature of the data did not allow cause-effect relationships to

be proved, but merely identified associations between variables; they claimed the research would benefit from adopting either a longitudinal or a before-and-after experimental design (Hofer and Aubert 2013; Majmundar et al. 2019; Thomas et al. 2019; Moshkovitz and Hayat 2021).

This section highlights the factors that represent the challenges in sampling plans for applied social science research in Twitter, namely target populations, sampling frames, sampling methods/strategies, sample size and data collection methods/strategies. Additionally, the impact of these factors on research generalizability is discussed in the light of the literature review.

5.1 Target populations

The target population of Twitter users is in most of the studies bounded resorting to Twitter-specific features that describe what people write on Twitter ("tweet object) (18 studies), what people do or are on Twitter ("user object") (17 studies), or where people are when they tweet ("geo-object") (4 studies).

Even though the target population involves, in all studies, Twitter users—a consequence of papers' inclusion criteria—in 23 studies the respective authors acknowledged that the findings are likely to suffer from coverage error and for that reason could not be generalised outside the Twitterverse. This alert can be seen, in a first moment, as unnecessary since in a study targeted at "Twitter users" it is not expected to extrapolate conclusions beyond that universe. However, the decision to disclosure it as a potential limitation of the study is a sign of social science researchers' prudence by recognising that Twitter cannot fully replace a study conducted in the "real-world" or even in other platforms.

5.2 Sampling frames

Most studies (46) were found to be conducted without a sampling frame and some others resorted to lists that are imperfect representations of the Twitter population, e.g., online panels (11 studies) or crowdsourcing lists (4 studies). The reason for such a scenario is that there is no public and easily available list of Twitter users, thus forcing researchers to either create their own lists—which is difficult to accomplish for large populations—or use "proxy" lists to approach Twitter users; however, these cases suffer from a serious risk of lack of coverage (Ruths and Pfeffer 2015). Online panels or platforms of crowdsourcing services allow easy access to Twitter users, but these lists are hardly representative of the Twitterverse because these sampling frames are not built with the purpose of matching that population. Although online panels are representative of the general adult population, they cannot be assumed to be representative of the Twitter population as previous investigation suggests that this population is socio-demographically different from the general offline population (Blank 2017; Pew Research Center 2019; Hootsuite 2020). In the few studies in which researchers were able to build a tailored sampling frame (8 studies), the research covered either small or very specific populations—e.g., psychologists who were directors of professional associations (Brady 2016), scientific tweeters (Yu et al. 2019) or educators (Visser et al. 2014). Only 4 studies used a conceptual sampling frame "created" by random generation of Twitter users' IDs; this is likely to be explained firstly by the fact that not all social scientists possess the demanding computational knowledge required to implement such a procedure, and

secondly because while creating a list of Twitter users' IDs by random number generation is feasible for the "population of all Twitter users", it is difficult to implement for bounded populations (which was the case of most studies). However, Berzofsky et al. (2018) present promising results of this approach to select probability-based samples on Twitter if the range of ID numbers that correspond to active Twitter users can be identified in advance.

5.3 Sampling methods and strategies

The sampling strategies undertaken when no sampling frame is available are based either on non-random methods (convenience, purposive or snowball sampling) or data search-based. On one hand, this is a consequence of the lack of sampling frames which impedes random selection (Groves et al. 2009, p. 94; Couper 2013); on the other, it is a choice to guarantee an efficient sampling procedure: selecting a sample of Twitter users by first locating tweets they posted containing the topic or event of interest is an efficient strategy to locate hard-to-reach populations or topic/event specific population. The selection is based on the dependent variable which guarantees the users are selected within the scope of the study; however, there is a risk of oversampling the users who are more active on the platform and under sampling those who rarely use or seldom post contents (Bruns and Stieglitz 2013; Zhang et al. 2018).

The randomness of a sample—each element has a nonzero probability of being chosen—is of the utmost importance for the social scientific methodological integrity as a sample selected randomly is regarded as a valid representation of the total population. There is no known method of obtaining a random and representative sample of all Twitter users, so sample selection based on an initial data search using keywords or hashtags was the most common approach namely in studies that were topic- or event- centered. Even though Twitter promises "random" samples of their data when using their API systems, Twitter samples are not free of criticism. Pfeffer et al. (2018) conducted an experiment to test the sampling procedure of Twitter's API by inducing tweets into the feed in such a way that they appear in the sample with great certainty. The authors found that 100 accounts were enough to manipulate the data stream for a globally important topic. Despite sharing (parts of) its data, Twitter does not reveal details about its sampling mechanisms (Ruiz-Soler 2017), which leaves researchers with no control over sample selection and makes it difficult to design strategies capable of dealing with potential bias. The reviewed papers that based sample selection on data search resorting to Twitter's APIs (31 studies) are sparce in methodological details about sample selection, which is a likely sign of the lack of information that the researchers themselves have of the sampling process. Researchers must trust Twitter to supply them with methodologically sound samples while dealing with all kinds of other problems, such as bias and ethical issues (Bruns 2019).

Ground theory from survey methodology states that random sampling must be encouraged in applied research. This is also desirable for Twitter context research and can be implemented by random generation of users' ID numbers. Although few studies used this approach (only 4), it is the most promising for implementing random selection when no list of Twitter users is available. Despite the risk of generating numbers that do not correspond to active users or that are out of the scope of the study—which brings some inefficiency to the process—, it has the advantage of giving all users the same probability of selection and thus avoids samples biased towards the most active and most proficient users. It would be interesting to transpose the Random Digit Dialling of phone numbers—which is also

based on randomly generated numbers—to the sampling of social media users as this is a highly regarded method of sampling frame creation that is commonly used by research organizations around the globe. However, this strategy will bring the challenge of detecting out-of-the-scope users such as bots (Alothali et al. 2018) since random number generation is likely to generate a lot of numbers that do not match researchers' interest.

5.4 Sample size

Samples sizes in the reviewed papers ranged from "small" (up to 2000 users) to "big" but there is a strong skewness towards smaller samples: 70% of the studies used samples of less than 2000 Twitter users; only 4% reported samples sizes over 1 million users. Given that Twitter is a big data source providing access (at least theoretically) to huge datasets, one would expect bigger samples sizes. That was not the case though. Likely explanations for this might lie in the fact that some research was focused on very restricted domains or populations (e.g., Watson 2016; Lee et al. 2020) or had an exploratory purpose where the precision and accuracy requirements are less demanding so there is no need for large sample sizes (e.g., Brady 2016; Baik et al. 2021).

The size of the sample is to a great extent also closely connected to the data collection strategy, namely whether the strategy relies on gathering data via an API system or mostly on self-report data. When splitting the analysis by type of data collection strategy, it is evident that the sample size in studies employing surveys or other self-reported mode (32 studies) tends to be lower (average = 557 users) than in studies that use exclusively (or rely heavily on) API systems (average = 297,893 users) (34 studies). The average sample size of self-reported studies is in line with social science practice since sample size in surveys targeted at specific populations with no stratification is usually dimensioned between 200 and 500 cases (Sudman 1983). However, the size of samples relying on data search is low given the huge dataset sizes that the Twitterverse potentially offers. Despite the impressive scale of Twitterverse data, researchers start looking at large datasets as a means to an end and not as an end in themselves (Callegaro and Yang 2018; Salganik 2018, p. 17).

5.5 Data collection methods

The data collection designs found in the reviewed papers were classified as: (i) Single mode, one sample, (ii) Single mode, different samples, and (iii) Mixed-mode. The most frequent design is "single mode, one sample" in which data is collected from one sample, using a single mode of data collection. In 29 studies, Twitter data and/or metadata is gathered for one sample of Twitter users resorting to an API system; on the other hand, data were collected in 14 studies through the exclusive use of a self-report mode, specifically surveys. In mixed-mode designs, data are collected for one sample of Twitter users, with different modes for different stages of the study. More specifically, 15 studies collected data for a sample of Twitter users by means of a survey questionnaire in a first stage, and the same sample is asked to volunteer their Twitter usernames or IDs in a second stage so that researchers can access their Twitter data and/or metadata via an API system. Despite the huge amount of data that can be retrieved from social media platforms, social media data is not enough in some studies and other data sources must be used to supplement and enrich the data gathered within the Twitter platform. This is demonstrated by the 22 studies that employed mixed-mode designs of data collection. Even though some argue that sampling is "an artefact of a period of information scarcity" (e.g., Mayer-Schönberger and Cukier

2013), the systematic review presents evidence that survey sampling is a central necessity even in times of information abundance.

Regarding data collection, research would benefit from exploring mixed-mode designs instead of focusing solely on Twitter data retrieval. Twitter data may not be enough for the research goals since it is mostly composed of events and usually provides very little or no information on the user that generated the data. Social media studies are frequently limited by not reporting demographic data because of the difficulty in retrieving this information from publicly available online data sources. Twitter suffers from this problem, which is a severe limitation since variables such as age, sex, income, education are crucial in social science research for multivariate analyses or subgroup comparisons. From a social science perspective, Twitter data is case rich (big samples) but variable poor (few covariates), which is why the complementarity of other modes is essential in some studies (Couper 2013; Callegaro and Yang 2018; Salganik 2018, p. 24). This suggests that contrary to what some have predicted (e.g., Savage and Burrows 2007), Twitter will not replace surveys in social science research; however, it opens new avenues of investigation that will enrich the understanding of social issues if both approaches work together instead of separately. Surveys are the best mode to collect data at an individual level and are the most used tool to supplement Twitter data in social science research. Researchers should give further attention to refining the design to articulate data from these two different sources (Callegaro and Yang 2018).

5.6 Research validity

Less than half the studies (36 studies) acknowledge that factors related to the design and implementation of the sampling plan may impact the outcomes and compromise findings' generalizability. Twenty-three studies highlight coverage error as a major problem when conducting applied social science research using Twitter. This concern is justified by the limited penetration of the Twitter platform—nearly 400 million active users worldwide—much behind Facebook, YouTube, or WhatsApp which each have more than 2000 million active users worldwide (Statista 2022b). With such a low use, it is hard to accept that the Twitter population is a reliable representation of the offline population or of the population using other platforms.

Sample selection is noted as a major problem in 19 studies, but no details are provided about the magnitude of the impact on the findings of using specific sampling methods. The development and implementation of strategies to deal with these issues, such as adopting random sampling instead of non-random sampling (e.g., Pentina et al. 2013; Kobayashi et al. 2019; Akyuz et al. 2021) or sample weighting (Kobayashi et al. 2019), are postponed to future research.

Self-selection bias is acknowledged in three papers. The usage pattern of social media in general and of the Twitter platform in particular is highly diverse across users, with a few contributing with many tweets and the remaining posting hardly anything (Bruns and Stieglitz 2013; Nielsen 2017; Pew Research Center 2019). This impacts sample selection since the most active users are more likely to be selected—namely in a sampling strategy based on data search—and there is no guarantee that they are representative of the Twitterverse. This problem must be addressed not only by designing sampling strategies to reduce its occurrence but also by developing weighting schemes to attenuate its effect. Just as in surveys, paradata (i.e., data about the process of answering the survey itself, collected by systems and third parties, before, during, and after the administration of a questionnaire)

are used to build weights to correct for nonresponse bias (Nicolaas 2011; Olson 2013) in social media, metadata, such as number of posts and number of followers, can be used to build weights aimed to account for different likelihoods of the selection of Twitter users. The future of Twitter-based research must look for answers to these issues so that a solid body of knowledge is acquired, thus contributing to reliable social science and computational social science research (Ruths and Pfeffer 2014).

6 Conclusion

This paper presents a systematic quantitative literature review of the design and implementation of sampling plans to select samples of Twitter users in applied social science studies. Seventy-three papers were identified as reporting evidence of the methodology adopted to select samples of Twitter users, and were analysed to find empirical answers to questions such as which sampling frames are best to represent Twitter users, which sampling strategies should be employed to select representative samples of Twitter users and which strategies of data collection should be adopted.

Results show that applied social science studies are conducted either without a sampling frame or with a sampling frame that does not adequately represent the population of Twitter users. Sampling strategies are to a great extent conditioned by this fact. Twitter users are selected either resorting to non-random methods—mostly convenience or snowball sampling—or by searching the tweets they posted containing specific keywords or hashtags. Regarding data collection designs, most of the studies relied on a single mode of data collection—namely, via an API system—applied to a unique sample of Twitter users. A non-ignorable number of studies combined self-reported data, namely surveys, with data gathered via an API system collected for the same sample of Twitter users. Surveys, the most widely used method of data collection in social science research, appears in 50% of the studies either exclusively as the single mode of data collection or in combination with data gathering via API systems. This is a significant sign that many research questions cannot be fully answered exclusively through social media data and require mixed-mode approaches to obtain a deep understanding of the phenomenon.

As in any other literature review, this work is limited by the search terms used and the databases searched (Pickering and Byrne 2014). To try to minimize this limitation, similar and commonly interchanged terms were included as well as a search on databases focused on social sciences and related fields. However, there may still be some relevant works that were not captured by the search criteria. Despite this fact, the collected evidence suggests that sampling considerations should become central components of a project's research design, and the strengths and limitations of different sampling plans should be explicitly included in the discussion of results. Based on the review undertaken, it is evident that we are still far from that scenario since less than half the studies discussed the limitations and the impact on research validity caused by using Twitter.

The research conducted on Twitter is vast and has been applied to a variety of domains, including social sciences. However, the use of this platform raises new quality issues that must be properly addressed because only when the limitations and challenges inherent to a data source are acknowledged, can a phenomenon be understood and the research findings validly interpreted (Callegaro and Yang 2018).

Declarations

References

Alothali, E., Zaki, N., Mohamed, E., Ashwal, H.: Detecting social bots on Twitter: a literature review. Int. Conf. Innov. Inf. Technol. (IIT) **2018**, 175–180 (2018)

Amaya, A., Biemer, P., Kinyon, D.: Total error in a big data world: adapting the TSE framework to big data. J. Surv. Stat. Methodol. **8**(1), 89–119 (2020). https://doi.org/10.1093/jssam/smz056

Anderson, C.: The end of theory: The data deluge makes the scientific method obsolete. Wired Mag, 16(7). Available at: http://statlit.org/pdf/2008EndOfTheory-DataDelugeMakesScientificMethodObsolete-WiredMagazine.pdf. Accessed 20th Sept 2022. (2008)

Andrews, D., Nonnecke, B., Preece, J.: Electronic survey methodology: a case study in reaching hard-to-involve Internet users. Int. J. Hum. Comput. Interact. **16**(2), 185–210 (2003)

Babbie, E.: The Basics of Social Research, 7th edn. Cengage Learning, Boston (2017)

Berzofsky, M., McKay, T., Hsieh, P., Smith, A.: Probability-based samples on Twitter: methodology and application. Surv. Pract. (2018). https://doi.org/10.29115/SP-2018-0033

Bhattacherjee, A.: Social Science Research: Principles, Methods, and Practices. Textbooks Collection. 3. http://scholarcommons.usf.edu/oa_textbooks/3 (2012)

Biemer, P.: Total survey error: design, implementation, and evaluation. Public Opin. Q. **74**(5), 817–848 (2010). https://doi.org/10.1093/poq/nfq058

Biemer, P.: Dropping the 's' from TSE: applying the paradigm to big data. In: Paper presented at the 2014 International Total Survey Error Workshop 23 The Role of Surveys in the Era of "Big Data" 187 (ITSEW 2014). Washington, DC: National Institute of Statistical Science. https://www.niss.org/sites/default/files/biemer_ITSEW2014_Presentation.pdf. (2014)

Blank, G.: The digital divide among Twitter users and its implications for social research. Soc. Sci. Comput. Rev. **35**(6), 679–697 (2017)

Bruns, A.: After the 'APIcalypse': social media platforms and their fight against critical scholarly research. Inf. Commun. Soc. **22**(11), 1544–1566 (2019). https://doi.org/10.1080/1369118X.2019.1637447

Bruns, A., Stieglitz, S.: Towards more systematic Twitter analysis: metrics for tweeting activities. Int. J. Soc. Res. Methodol. **16**(2), 91–108 (2013)

Callegaro, M., Yang, Y.: The role of surveys in the era of "Big Data." In: Vannette, D.L., Krosnick, J.A. (eds.) The Palgrave Handbook of Survey Research, pp. 175–192. Springer, Berlin (2018)

Cioffi, C.: Computational social science. Wiley Interdiscip. Rev. Comput. Stat. **2**, 259–271 (2010). https://doi.org/10.1002/wics.95

Cohen, R., Ruths, D.: Political orientation inference on Twitter: It's not easy! Proc. ICWSM **13**, 91–99 (2013)

Couper, M.: Is the sky falling? New Technology, changing media, and the future of surveys. Surv. Res. Methods **7**(3), 145–156 (2013). https://doi.org/10.18148/srm/2013.v7i3.5751

Couper, M., Miller, P.: Web survey methods: introduction. Public Opin. Q. **72**(8), 831–835 (2008)

Daas, P., Burger, J., Le, Q., Bosch, O., Puts, M.: *Profiling of Twitter users: a big data selectivity study*. Discussion paper 6. Statistics Netherlands. (2016)

DataReportal (2022). Global social media stats. Available at: https://datareportal.com/social-media-users. Accessed 22nd Apr 2022. (2022)

De Leeuw, E.: To mix or not to mix data collection modes in surveys. J. Off. Stat. **21**(2), 233–255 (2005)

Eurostat. An overview of methods for treating selectivity in big data sources. Statistical Working Papers. European Union. (2018)

Gayo-Avello, D.: Don't turn social media into another "Literary Digest" poll. Commun. ACM **54**(10), 121–128 (2011). https://doi.org/10.1145/2001269.2001297

Gilbert, G.: Computational Social Science. Sage, Los Angeles (2010)

Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., Tourangeau, R.: Survey Methodology, 2nd edn. Wiley, New York (2009)

Hargittai, E.: Is bigger always better? Potential biases of big data derived from social network sites. Ann. Am. Acad. Pol. Soc. Sci. **659**(1), 63–76 (2015). https://doi.org/10.1177/0002716215570866

Hargittai, E.: Potential biases in big data: Omitted voices on social media. Soc. Sci. Comput. Rev. **38**(1), 10–24 (2020). https://doi.org/10.1177/0894439318788322

Hino, A., Fahey, R.: Representing the Twittersphere: Archiving a representative sample of Twitter data under resource constraints. Int. J. Inf. Manage. **48**, 175–184 (2019). https://doi.org/10.1016/j.ijinfomgt.2019.01.019

Hootsuite. Top Twitter demographics that matter to social media marketers. Available at: https://blog.hootsuite.com/twitter-demographics/. Accessed 1st Apr 2022. (2020)

Investopedia. Social sciences. Available at: https://www.investopedia.com/terms/s/social-science.asp Accessed 31st Mar 2022. (2021)

Lapowsky, I. Your old tweets give away more location data than you think. Available at: https://www.wired.com/story/twitter-location-data-gps-privacy/. Accessed 21st Apr 2022. (2019)

Lazer, D., Kennedy, R., King, G., Vespignani, A.: The parable of Google flu: traps in big data analysis. Science **343**(6176), 1203–1205 (2014)

Lu, T., Franklin, A.: A protocol for identifying and sampling from proxy populations. Soc. Sci. Q. **99**(4), 1535–1546 (2018). https://doi.org/10.1111/ssqu.12519

Mayer-Schönberger, V., Cukier, K.: Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt, Boston (2013)

McCormick, T., Lee, H., Cesare, N., Shojaie, A., Spiro, E.: Using Twitter for demographic and social science research: tools for data collection and processing. Sociol. Methods Res. **46**(3), 390–421 (2017). https://doi.org/10.1177/0049124115605339

Mustafaraj, E., Finn, S., Whitlock, C., Metaxas, P.: Vocal minority versus silent majority: discovering the opinions of the long tail. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, 103–110, doi: https://doi.org/10.1109/PASSAT/SocialCom.2011.188 (2011)

Nicolaas, G.: Survey paradata: a review. ESRC National Centre for Research Methods Review paper (NCRM/017). Available at: https://eprints.ncrm.ac.uk/id/eprint/1719/1/Nicolaas_review_paper_jan11.pdf. Accessed 5th April 2022. (2011)

Nielsen. 2016 Nielsen social media report-social studies: a look at the social landscape. Available at: https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/2016-nielsen-social-media-report.pdf. Accessed 27th April 2022. (2017)

Nielsen, J.: The 90-9-1 rule for participation inequality in social media and online communities. Available at: https://www.nngroup.com/articles/participation-inequality/. Accessed 17th Jan 2022. (2006)

Olson, K.: Paradata for nonresponse adjustment. Ann. Am. Acad. Pol. Soc. Sci. **645**, 142–170 (2013)

Olteanu, A., Castillo, C., Diaz, F., Kıcıman, E.: Social Data: biases, methodological pitfalls, and ethical boundaries. Front. Big Data (2019). https://doi.org/10.3389/fdata.2019.00013

Pew Research Center. Sizing up Twitter users. Available at: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/. Accessed 1st Apr 2022 (2019).

Pew Research Center. Social media use in 2021a. Available at: https://www.pewresearch.org/internet/2021a/04/07/social-media-use-in-2021a/. Accessed 20th Jan 2022. (2021)

Pew Research Center. The behaviors and attitudes of U.S. adults on Twitter. Available at: https://www.pewresearch.org/internet/2021b/11/15/2-comparing-highly-active-and-less-active-tweeters/. Accessed 1st Apr 2022. (2021b)

Pfeffer, J., Mayer, K., Morstatter, F.: Tampering with Twitter's sample API. EPJ Data Sci. **7**, 50 (2018). https://doi.org/10.1140/epjds/s13688-018-0178-0

Pickering, C., Byrne, J.: The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. High. Educ. Res. Dev. **33**(3), 534–548 (2014). https://doi.org/10.1080/07294360.2013.841651

Rafail, P.: Nonprobability sampling and Twitter: strategies for semibounded and bounded populations. Soc. Sci. Comput. Rev. **36**(2), 195–211 (2018). https://doi.org/10.1177/0894439317709431

Roth, Y., Johnson, R. (2018). *New developer requirements to protect our platform.* Available at: https://blog.twitter.com/developer/en_us/topics/tools/2018/new-developer-requirements-to-protect-our-platform. Accessed 21st September 2022.

Ruiz-Soler, J.: Twitter research for social scientists: a brief introduction to the benefits, limitations, and tools for analyzing Twitter data. Revista Dígitos **3**, 17–31 (2017)

Ruths, D., Pfeffer, J.: Social media for large studies of behavior. Science **346**(6213), 1063–1064 (2014)

Salganik, M.: Bit by Bit- Social Research in the Digital Age. Princeton University Press, New Jersey (2018)

Savage, M., Burrows, R.: The coming crisis of empirical sociology. Sociology **41**(5), 885–899 (2007)

Schwitter, N., Liebe, U.: Going Digital: Web data collection using Twitter as an example. Oxfam GB. (2020). https://doi.org/10.21201/2020.5686

Statcounter. Social Media Stats Worldwide. Available at: https://gs.statcounter.com/social-media-stats Accessed 1st Apr 2022. (2022)

Statista. Global digital population as of January 2021. Available at: https://www.statista.com/statistics/617136/digital-population-worldwide/. Accessed 1st Apr 2022a (2022a)

Statista. Most popular social networks worldwide as of January 2022b, ranked by number of monthly active users. Available at: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/. Accessed 24th Mar 2022b (2022b).

Statista (2022c). *Twitter – Statistics & Facts.* Available at: https://www.statista.com/topics/737/twitter/#topicHeader__wrapper. Accessed 21st April 2022c.

Stieglitz, S., Dang-Xuan, L., Bruns, A., Neuberger, C.: Social media analytics—An interdisciplinary approach and its implications for information systems. Bus. Inf. Syst. Eng. **6**(2), 89–96 (2014)

Sudman, S.: Applied sampling. In: Rossi, P., Wright, J., Anderson, A. (eds.) Handbook of Survey Research, pp. 145–194. Academic Press Inc, Cambridge (1983)

Tufekci, Z.: Big questions for social media big data: representativeness, validity and other methodological pitfalls. In: Proceedings of the eight international AAAI conference on weblogs and social media. AAAI Press, Menlo Park, pp. 505–514. (2014)

Twitter. About public and protected Tweets. Available at: https://help.twitter.com/en/safety-and-security/public-and-protected-tweets. Accessed 24th Apr 2022 (2022a)

Twitter. Data dictionary: Standard v1.1. Available at: https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview. Accessed 7th Mar 2022 (2022b)

Twitter. Decahose API. Available at: https://developer.twitter.com/en/docs/twitter-api/enterprise/decahose-api/overview/decahose. Accessed 4th Oct 2022 (2022c)

Twitter. Inactive account policy. Available at: https://help.twitter.com/en/rules-and-policies/inactive-twitter-accounts. Accessed 26th Sept 2022d (2022d).

Twitter. New user FAQ. Available at: https://help.twitter.com/en/resources/new-user-faq. Accessed 26th Sept 2022e (2022e)

Twitter. Tweet location FAQs. Available at: https://help.twitter.com/en/safety-and-security/tweet-location-settings. Accessed 21st Apr 2022f (2022f).

Uddin, M., Imran, M., Sajjad, H. Understanding types of users on Twitter. Available at: https://arxiv.org/abs/1406.1335. Accessed 22nd Apr 2022. (2014)

Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: detection, estimation, and characterization. Available at: https://arxiv.org/pdf/1703.03107.pdf. Accessed 22nd Apr 2022. (2017)

Waksberg, J.: Sampling methods for random digit dialing. J. Am. Stat. Assoc. **73**(361), 40–46 (1978). https://doi.org/10.2307/2286513

Zhang, H., Hill, S., Rothschild, D.: Addressing selection bias in event studies with general-purpose social media panels. ACM J. Data Inf. **10**(1), 24 (2018). https://doi.org/10.1145/3185048

Coded papers

Abney, A.K., Cook, L.A., Fox, A.K., Stevens, J.: Intercollegiate social media education ecosystem. J. Mark. Educ. **41**(3), 254–269 (2019). https://doi.org/10.1177/0273475318786026

Akyuz, K., Akbas, H., Onat, I.: Evaluating the impact of Syrian refugees on fear of crime in Turkey. Eur. J. Criminol. (2021). https://doi.org/10.1177/14773708211010665

Álvarez-Bornstein, B., Montesi, M.: Who is interacting with researchers on Twitter? A survey in the field of Information Science. JLIS. It **10**(2), 87–106 (2019). https://doi.org/10.4403/jlis.it-12530

Baik, J.M., Nyein, T.H., Modrek, S.: Social media activism and convergence in tweet topics after the initial #metoo movement for two distinct groups of twitter users. J. Interpers. Violence (2021). https://doi.org/10.1177/08862605211001481

Barbera, P., Jost, J.T., Nagler, J., Tucker, J.A., Bonneau, R.: Tweeting from left to right: Is online political communication more than an echo chamber? Psychol. Sci. **26**(10), 1531–1542 (2015). https://doi.org/10.1177/0956797615594620

Bekafigo, M.A., McBride, A.: Who Tweets about politics?: political participation of twitter users during the 2011gubernatorial elections. Soc. Sci. Comput. Rev. **31**(5), 625–643 (2013). https://doi.org/10.1177/0894439313490405

Bica, M., Palen, L., Henderson, J., Spinney, J., Weinberg, J., Nielsen, E.R.: "Can't think of anything more to do": public displays of power, privilege, and surrender in social media disaster monologues. Hum. Comput. Interact. (2021). https://doi.org/10.1080/07370024.2021.1982390

Bouteca, N., D'heer, E., Lannoo, S.: How second-order is the regional level? An analysis of tweets in simultaneous campaigns. Pol. Stud. **65**(4), 1021–1039 (2017). https://doi.org/10.1177/0032321717697346

Brady, L.L.C.: Canaries in the ethical coal mine? Case vignettes and empirical findings for how psychology leaders have adopted Twitter. Eth. Behav. **26**(2), 110–127 (2016). https://doi.org/10.1080/10508422.2014.994064

Cavazos-Rehg, P.A., Krauss, M.J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., Bierut, L.J.: A content analysis of depression-related Tweets. Comput. Hum. Behav. **54**, 351–357 (2016). https://doi.org/10.1016/j.chb.2015.08.023

Chadwick, A., Vaccari, C., O'Loughlin, B.: Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing. New Media Soc. **20**(11), 4255–4274 (2018). https://doi.org/10.1177/1461444818769689

Chen, G.M.: Tweet this: a uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. Comput. Hum. Behav. **27**(2), 755–762 (2011). https://doi.org/10.1016/j.chb.2010.10.023

D'Heer, E., Verdegem, P.: What social media data mean for audience studies: a multidimensional investigation of Twitter use during a current affairs TV programme. Inf. Commun. Soc. **18**(2), 221–234 (2015). https://doi.org/10.1080/1369118x.2014.952318

Eady, G., Nagler, J., Guess, A., Zilinsky, J., Tucker, J.A.: How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. SAGE Open **9**(1), 215824401983270 (2019). https://doi.org/10.1177/2158244019832705

Ferrara, E., Yang, Z.: Measuring emotional contagion in social media. PLoS One **10**(11), e0142390 (2015). https://doi.org/10.1371/journal.pone.0142390

Fiesler, C., Proferes, N.: "Participant" perceptions of twitter research ethics. Soc, Media Soc. **4**(1), 205630511876336 (2018). https://doi.org/10.1177/2056305118763366

Fischer, E., Reuber, A.R.: Social interaction via new social media: (How) can interactions on Twitter affect effectual thinking and behavior? J. Bus. Ventur. **26**(1), 1–18 (2011). https://doi.org/10.1016/j.jbusvent.2010.09.002

Gómez-Zará, D., Diakopoulos, N.: Characterizing communication patterns between audiences and news-bots. Digit. J. **8**(9), 1093–1113 (2020). https://doi.org/10.1080/21670811.2020.1816485

Han, S., Min, J., Lee, H.: Antecedents of social presence and gratification of social connection needs in SNS: a study of Twitter users and their mobile and non-mobile usage. Int. J. Inf. Manage. **35**(4), 459–471 (2015). https://doi.org/10.1016/j.ijinfomgt.2015.04.004

Hayat, T., Samuel-Azran, T., Galily, Y.: Al-Jazeera Sport's US Twitter followers: sport-politics nexus? Online Inf. Rev. **40**(6), 785–797 (2016). https://doi.org/10.1108/oir-01-2016-0033

Hofer, M., Aubert, V.: Perceived bridging and bonding social capital on Twitter: differentiating between followers and followees. Comput. Hum. Behav. **29**(6), 2134–2142 (2013). https://doi.org/10.1016/j.chb.2013.04.038

Holiday, S., Hayes, J.L., Britt, B.C., Lyu, Y.: The cause effect: the impact of corporate social responsibility advertising on cause consumer engagement behavior after brand affiliation ceases. Int. J. Advert. **40**(2), 199–224 (2020). https://doi.org/10.1080/02650487.2020.1769408

Hswen, Y., Naslund, J.A., Brownstein, J.S., Hawkins, J.B.: Monitoring online discussions about suicide among Twitter users with schizophrenia: exploratory study. JMIR Ment. Health **5**(4), e11483 (2018). https://doi.org/10.2196/11483

Hswen, Y., Naslund, J.A., Brownstein, J.S., Hawkins, J.B.: Online communication about depression and anxiety among twitter users with schizophrenia: preliminary findings to inform a digital phenotype using social media. Psychiatr. Q **89**(3), 569–580 (2018b). https://doi.org/10.1007/s11126-017-9559-y

Hswen, Y., Naslund, J.A., Chandrashekar, P., Siegel, R., Brownstein, J.S., Hawkins, J.B.: Exploring online communication about cigarette smoking among Twitter users who self-identify as having schizophrenia. Psychiatr. Res. **257**, 479–484 (2017). https://doi.org/10.1016/j.psychres.2017.08.002

Huang, Q., Wong, D.W.S.: Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? Int. J. Geogr. Inf. Sci. **30**(9), 1873–1898 (2016). https://doi.org/10.1080/13658816.2016.1145225

Jünger, J., Fähnrich, B.: Does really no one care? Analyzing the public engagement of communication scientists on Twitter. New Media Soc. **22**(3), 387–408 (2020). https://doi.org/10.1177/1461444819863413

Kearney, M.W.: Analyzing change in network polarization. New Media Soc. **21**(6), 1380–1402 (2019). https://doi.org/10.1177/1461444818822813

Kobayashi, T., Ogawa, Y., Suzuki, T., Yamamoto, H.: News audience fragmentation in the Japanese Twittersphere. Asian J. Commun. **29**(3), 274–290 (2019). https://doi.org/10.1080/01292986.2018.1458326

Lee, J.Y., Grogan-Kaylor, A.C., Lee, S.J., Ammari, T., Lu, A., Davis-Kean, P.: A Qualitative analysis of stay-at-home parents' spanking Tweets. J. Child Fam. Stud. **29**(3), 817–830 (2020). https://doi.org/10.1007/s10826-019-01691-3

Lerman, K., Marin, L.G., Arora, M., de Lima, L.H.C., Ferrara, E., Garcia, D.: Language, demographics, emotions, and the structure of online social networks. J. Comput. Soc. Sci. **1**(1), 209–225 (2018). https://doi.org/10.1007/s42001-017-0001-x

Levordashka, A., Utz, S.: Ambient awareness: from random noise to digital closeness in online social networks. Comput. Hum. Behav. **60**, 147–154 (2016). https://doi.org/10.1016/j.chb.2016.02.037

Liang, H., Fu, K.-W.: Network redundancy and information diffusion: the impacts of information redundancy, similarity, and tie strength. Commun. Res. **46**(2), 250–272 (2019). https://doi.org/10.1177/0093650216682900

Liang, H., Fu, K.W.: Testing propositions derived from twitter studies: generalization and replication in computational social science. PLoS One **10**(8), e0134270 (2015). https://doi.org/10.1371/journal.pone.0134270

Liang, H., Shen, F.: Birds of a schedule flock together: social networks, peer influence, and digital activity cycles. Comput. Hum. Behav. **82**, 167–176 (2018). https://doi.org/10.1016/j.chb.2018.01.016

Liang, H., Shen, F., Fu, K.-W.: Privacy protection and self-disclosure across societies: a study of global Twitter users. New Media Soc. **19**(9), 1476–1497 (2017). https://doi.org/10.1177/1461444816642210

Majmundar, A., Allem, J.P., Boley Cruz, T., Unger, J.B.: The why we retweet scale. PLoS One **13**(10), e0206076 (2018). https://doi.org/10.1371/journal.pone.0206076

Majmundar, A., Chou, C.P., Cruz, T.B., Unger, J.B.: Relationship between social media engagement and e-cigarette policy support. Addict. Behav. Rep. **9**, 100155 (2019). https://doi.org/10.1016/j.abrep.2018.100155

Marchetti, R., Ceccobelli, D.: Twitter and television in a hybrid media system. J. Pract. **10**(5), 626–644 (2016). https://doi.org/10.1080/17512786.2015.1040051

Marwick, A.E., Boyd, D.: I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New Media Soc. **13**(1), 114–133 (2010). https://doi.org/10.1177/1461444810365313

Mascaro, C.M., Goggins, S.P.: Technologically mediated political discourse during a nationally televised GOP primary debate. J. Inform. Tech. Polit. **12**(3), 252–269 (2015). https://doi.org/10.1080/19331681.2015.1071687

Mihunov, V.V., Lam, N.S.N., Zou, L., Wang, Z., Wang, K.: Use of Twitter in disaster rescue: lessons learned from Hurricane Harvey. Int. J. Digital Earth **13**(12), 1454–1466 (2020). https://doi.org/10.1080/17538947.2020.1729879

Mohammed, A., Ferraris, A.: Factors influencing user participation in social media: evidence from twitter usage during COVID-19 pandemic in Saudi Arabia. Technol. Soc. **66**, 101651 (2021). https://doi.org/10.1016/j.techsoc.2021.101651

Montgomery, L., Heidelburg, K., Robinson, C.: Characterizing blunt use among Twitter users: racial/ethnic differences in use patterns and characteristics. Subst. Use Misuse **53**(3), 501–507 (2018). https://doi.org/10.1080/10826084.2017.1341926

Moshkovitz, K., Hayat, T.: The rich get richer: extroverts' social capital on twitter. Technol. Soc. **65**, 101551 (2021). https://doi.org/10.1016/j.techsoc.2021.101551

Murthy, D., Gross, A., Pensavalle, A.: Urban social media demographics: an exploration of Twitter use in major American cities. J. Comput.-Mediat. Commun. **21**(1), 33–49 (2016). https://doi.org/10.1111/jcc4.12144

Naskar, D., Singh, S.R., Kumar, D., Nandi, S., Rivaherrera, E.O.D.L.: Emotion dynamics of public opinions on Twitter. ACM Trans. Inf. Syst. **38**(2), 1–24 (2020). https://doi.org/10.1145/3379340

O'Hallarn, B., Shapiro, S.: #NBCFail: a qualitative review of the shared experience as a social movement. First Monday (2014). https://doi.org/10.5210/fm.v19i1.4760

Osorio-Arjona, J., García-Palomares, J.C.: Social media and urban mobility: using twitter to calculate home-work travel matrices. Cities **89**, 268–280 (2019). https://doi.org/10.1016/j.cities.2019.03.006

Pentina, I., Zhang, L., Basmanova, O.: Antecedents and consequences of trust in a social media brand: A cross-cultural study of Twitter. Comput. Hum. Behav. **29**(4), 1546–1555 (2013). https://doi.org/10.1016/j.chb.2013.01.045

Pourebrahim, N., Sultana, S., Edwards, J., Gochanour, A., Mohanty, S.: Understanding communication dynamics on Twitter during natural disasters: a case study of Hurricane Sandy. Int. J. Disaster Risk Reduct. **37**, 101176 (2019). https://doi.org/10.1016/j.ijdrr.2019.101176

Qiu, L., Lin, H., Ramsay, J., Yang, F.: You are what you tweet: personality expression and perception on Twitter. J. Res. Pers. **46**(6), 710–718 (2012). https://doi.org/10.1016/j.jrp.2012.08.008

Reuschke, D., Long, J., Bennett, N.: Locating creativity in the city using Twitter data. Environ. Plann. b: Urban Anal. City Sci. **48**(9), 2607–2622 (2021). https://doi.org/10.1177/2399808320980745

Reveilhac, M., Blanchard, A.: The framing of health technologies on social media by major actors: Prominent health issues and COVID-related public concerns. Int. J. Inf. Manag. Data Insights **2**(1), 100068 (2022). https://doi.org/10.1016/j.jjimei.2022.100068

Samuel-Azran, T., Hayat, T.: The geography of the Arab public sphere on Twitter. Technol. Soc **62**, 101327 (2020). https://doi.org/10.1016/j.techsoc.2020.101327

Sasaki, Y., Kawai, D., Kitamura, S.: The anatomy of tweet overload: how number of tweets received, number of friends, and egocentric network density affect perceived information overload. Telemat. Inform. **32**(4), 853–861 (2015). https://doi.org/10.1016/j.tele.2015.04.008

Sasaki, Y., Kawai, D., Kitamura, S.: Unfriend or ignore tweets?: A time series analysis on Japanese Twitter users suffering from information overload. Comput. Hum. Behav. **64**, 914–922 (2016). https://doi.org/10.1016/j.chb.2016.07.059

Sashittal, H., Jassawalla, A.: Brand entification as a post-anthropomorphic attribution among Twitter-using Millennials. Mark. Intell. Plan. **37**(7), 741–753 (2019). https://doi.org/10.1108/mip-10-2018-0446

Schaarschmidt, M., Könsgen, R.: Good citizen, good ambassador? Linking employees' reputation perceptions with supportive behavior on Twitter. J. Bus. Res. **117**, 754–763 (2020). https://doi.org/10.1016/j.jbusres.2019.04.004

Sharples, R.: Disrupting state spaces: asylum seekers in Australia's offshore detention centres. Soc. Sci. **10**(3), 82 (2021). https://doi.org/10.3390/socsci10030082

Shin, J.: How do partisans consume news on social media? A comparison of self-reports with digital trace measures among Twitter users. Soc. Media Soc. **6**(4), 205630512098103 (2020). https://doi.org/10.1177/2056305120981039

Storer, H.L., Rodriguez, M.: #Mapping a movement: social media, feminist hashtags, and movement building in the digital age. J. Commun. Pract. **28**(2), 160–176 (2020). https://doi.org/10.1080/10705422.2020.1757541

Storer, H.L., Rodriguez, M., Franklin, R.: "Leaving Was a Process, Not an Event": the lived experience of dating and domestic violence in 140 characters. J. Interpers. Violence **36**(11–12), NP6553–NP6580 (2021). https://doi.org/10.1177/0886260518816325

Thomas, J., Al-Shehhi, A., Al-Ameri, M., Grey, I.: We tweet Arabic; I tweet English: self-concept, language and social media. Heliyon **5**(7), e02087 (2019). https://doi.org/10.1016/j.heliyon.2019.e02087

Tominaga, T., Hijikata, Y., Konstan, J.A.: How self-disclosure in Twitter profiles relate to anonymity consciousness and usage objectives: a cross-cultural study. J. Comput. Soc. Sci. **1**(2), 391–435 (2018). https://doi.org/10.1007/s42001-018-0023-z

Torregrosa, J., Thorburn, J., Lara-Cabrera, R., Camacho, D., Trujillo, H.M.: Linguistic analysis of pro-ISIS users on Twitter. Behav. Sci. Terror. Pol. Aggress. **12**(3), 171–185 (2020). https://doi.org/10.1080/19434472.2019.1651751

Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J.T., Nagler, J., Tucker, J.A.: Political expression and action on social media: exploring the relationship between lower- and higher-threshold political activities among twitter users in Italy. J. Comput.-Mediat. Commun. **20**(2), 221–239 (2015). https://doi.org/10.1111/jcc4.12108

Vaccari, C., Valeriani, A., Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A.: Of echo chambers and contrarian clubs: exposure to political disagreement among german and Italian users of twitter. Soc. Media Soc. **2**(3), 205630511666422 (2016). https://doi.org/10.1177/2056305116664221

Valenzuela, S., Puente, S., Flores, P.M.: Comparing disaster news on Twitter and television: an intermedia agenda setting perspective. J. Broadcast. Electron. Media **61**(4), 615–637 (2017). https://doi.org/10.1080/08838151.2017.1344673

Visser, R.D., Evering, L.C., Barrett, D.E.: #TwitterforTeachers: the implications of Twitter as a self-directed professional development tool for K–12 teachers. J. Res. Technol. Educ. **46**(4), 396–413 (2014). https://doi.org/10.1080/15391523.2014.925694

Watson, B.R.: Is Twitter an alternative medium? comparing Gulf Coast Twitter and newspaper coverage of the 2010 BP oil spill. Commun. Res. **43**(5), 647–671 (2016). https://doi.org/10.1177/0093650214565896

Yu, H., Xiao, T., Xu, S., Wang, Y.: Who posts scientific tweets? An investigation into the productivity, locations, and identities of scientific tweeters. J. Informet. **13**(3), 841–855 (2019). https://doi.org/10.1016/j.joi.2019.08.001

Zhang, Z., Ahmed, W.: A comparison of information sharing behaviours across 379 health conditions on Twitter. Int. J. Public Health **64**(3), 431–440 (2019). https://doi.org/10.1007/s00038-018-1192-5