

# Using Twitter for Demographic and Social Science Research: Tools for Data Collection and Processing

Tyler H. McCormick<sup>1,2,3,4</sup>, Hedwig Lee<sup>1,4</sup>,  
Nina Cesare<sup>1</sup>, Ali Shojaie<sup>2,3,5</sup> and Emma S. Spiro<sup>1,3,6</sup>

## Abstract

Despite recent and growing interest in using Twitter to examine human behavior and attitudes, there is still significant room for growth regarding the ability to leverage Twitter data for social science research. In particular, gleaning demographic information about Twitter users—a key component of much social science research—remains a challenge. This article develops an accurate and reliable data processing approach for social science researchers interested in using Twitter data to examine behaviors and attitudes, as well as the demographic characteristics of the populations expressing or engaging in them. Using information gathered from Twitter users who state an intention

---

<sup>1</sup> Department of Sociology, University of Washington, Seattle, WA, USA

<sup>2</sup> Department of Statistics, University of Washington, Seattle, WA, USA

<sup>3</sup> Center for Statistics and the Social Sciences, University of Washington, Seattle, WA, USA

<sup>4</sup> Center for Studies in Demography and Ecology, University of Washington, Seattle, WA, USA

<sup>5</sup> Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA, USA

<sup>6</sup> Information School, University of Washington, Seattle, WA, USA

## Corresponding Author:

Nina Cesare, Department of Sociology, University of Washington, 211 Savery Hall, Box 353340, Seattle, WA 98105, USA.

Email: ninac2@uw.edu

to not vote in the 2012 presidential election, we describe and evaluate a method for processing data to retrieve demographic information reported by users that is not encoded as text (e.g., details of images) and evaluate the reliability of these techniques. We end by assessing the challenges of this data collection strategy and discussing how large-scale social media data may benefit demographic researchers.

### **Keywords**

Twitter, demographics, data collection, attitudes, population

## **Introduction**

Social media, such as Twitter and Facebook, provide exciting opportunities that can “open up a new era” of social science research (Golder and Macy 2012). These new communication platforms afford the ability to examine social data on a variety of topics, on a massive scale, and over short periods of time. As such, they have sparked the interest of many researchers seeking to better understanding social relationships and behavior (Brickman Bhutta 2012; Golder and Macy 2011; Heavilin et al. 2011; Lowe et al. 2012; Moreno et al. 2012; Valkenburg, Peter, and Schouten 2006). Although some researchers have begun to use data from social media spaces such as Twitter to document changing moods and other sentiments on the aggregate level (Diakopoulos and Shamma 2010; Golder and Macy 2011; Naaman, Becker, and Gravano 2011; Reips and Garaizar 2011; Yardi and Boyd 2010), the potential of such data for demographic research has yet to be fully realized.

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 (Cambria, Wang, and White 2014; Lazer et al. 2009). In some ways, data produced online parallel more well-researched methods of data collection such as surveys or structured observations, but they also contain unique properties and new features. Surveys, for example, ask respondents to recall behaviors or sentiments retrospectively with limited scope, whereas social media afford the opportunity to observe personal expression and human interaction in real time and on a large scale. With appropriate infrastructure, social scientists can analyze social media data and begin presenting results within a matter of months (or sooner), rather than the years typically required for survey-based data collection.

Social media data also share some characteristics of observational or ethnographic work. Specifically, social media data allow researchers to collect reports of behaviors that are unsolicited and unprompted by the researcher. One might even argue that these data provide a better reflection of day-to-day social experiences than ethnographic work. Indeed, Twitter activity has been described as persons “want[ing] to know what the people around them are thinking and doing and feeling, even when co-presence isn’t viable” and “shar[ing] their state of mind and status so that others who care about them feel connected” (Boyd 2009). Unlike previous observational work, however, activity on social media platforms can be captured and stored. This is in contrast to face-to-face interactions, which once passed cannot be reconstructed. A researcher conducting off-line observational work is left with only his or her perceptions (and notes) of what transpired. Interaction on social media, however, is preserved and can be reviewed multiple times and passed to other interested researchers.

Despite these exciting advantages, social scientists often see social media data as inaccessible for social science research and solely relevant to computer scientists and others in related fields. Golder and Macy (2012:7) lament, “most of the social and behavioral science using online data is coming from computer and information scientists who do not always have the training required to ask the right questions, or to recognize unfounded assumptions and socially unjust ramifications.” This perception of inaccessibility is perpetuated by the fact that social scientists often do not gain the computational skills necessary to collect and analyze online data, nor do access points in online platforms—Twitter application programming interface (API) queries, for example—lend themselves to the purposive sampling of users rather than content.

Beyond technical skills, using social media data—in particular Twitter data—for social science research is challenging due to the fact that there are numerous differences between Twitter data and surveys collected using traditional sampling techniques. When using traditional surveys, for example, researchers have comparatively few respondents but have a great deal of control over what information respondents provide. Under these conditions, respondents provide information of interest to the researchers, but the limited sample size may not produce enough variability to study less commonly observed phenomena in their entirety (e.g., self-reports of suicide attempts, eating disorders, or HIV-positive status). Data from Twitter, in many ways, are just the opposite. They are completely unsolicited but offer unprecedented volume and variability.

In addition to challenges associated with specific qualities of the data, it is difficult to gather demographic data from text-based blogs and microblogs

such as Twitter. Indeed, in many online platforms, users give few or no explicit, self-reported characteristics. Demographic information is at the heart of many, if not most, social science analyses. It is often vital that researchers are able to utilize information on race, age, and gender to examine differential patterns in attitudes and behaviors. Removing the actual and perceived barriers that prevent social scientists from fully utilizing social media data expands research opportunities for social scientists and increases the potential for interdisciplinary research between computer scientists or statisticians and social and behavioral scientists, thus increasing the potential of studying complex social problems.

Focusing on the above-mentioned challenges, this article will describe an accurate and reliable data processing infrastructure that facilitates access to demographic information encoded as images within Twitter data. Furthermore, it seeks to encourage social scientists to consider Twitter as a valuable source of demographic information to answer relevant social science questions. To help illustrate this process, we examine a specific behavior—reporting an intention to not vote in the 2012 presidential election. This case study demonstrates the validity of the proposed methods as a toolkit that can be modified and applied to different questions and contexts.

In the following sections, we outline our approach to processing data from Twitter for demographic research and offer verification of the accuracy and reliability of these methods. We begin with an introduction to Twitter and a discussion of the resources and challenges associated with using Twitter data for social science research. Next, we propose a data-processing infrastructure that allows us to gather demographic information not directly reported within Twitter users' profiles. At the heart of the strategy is a framework for using Amazon's Mechanical Turk (AMT) to efficiently code large volumes of users' profile images, which we validate based on reliability of the AMT workers (Turkers), and comparisons between the evaluations of the Turkers and those of expert, trained coders. We illustrate the possible use of these methods by examining the racial composition of a small sample of Twitter users who state an intention to not vote in the 2012 presidential election. We conclude by addressing the benefits and challenges associated with this method of data collection, as well as the potential for future research that uses demographic data obtained from Twitter. It is our intention to present a flexible framework that can be used and adapted by researchers with varying experience regarding the collection and management of online data. The platforms we feature to collect and analyze data are not the only tools available to researchers, and while we outline our processes, we seek to highlight other potential pathways for analysis.

It is important to note that while social media data are often categorized as big data, these methods intend to focus on the research capacities of social media data that are not directly tied to the scale of the data. Although we seek to propose methods that are semiscalable and can be applied to data sets that would be time consuming for small teams of individual coders to assess, our primary goal is to leverage user metadata in a way that helps make Twitter data more accessible to the field of social science and demographic research. We believe that social media pose a number of benefits to social science research aside from their scale and we seek to highlight these capacities within this study.

### *Applications of Twitter Data in the Social Sciences*

Twitter provides a convenient source of data on users' opinions, interactions, and reported behaviors. It may be utilized, for example, by researchers who seek to examine large-scale processes of contagion, track preferences, and/or opinions among broad audiences; examine behaviors and attitudes where social desirability bias in an official survey may occur (e.g., racist attitudes, voting behavior or anti-immigrant sentiments; Belli et al. 1999; Holbrook and Krosnick 2010; Janus 2010; Tourangeau and Yan 2007); analyze collective experiences based on a timely event (e.g., terrorist attacks or natural disasters; Cassa et al. 2013; Sutton et al. 2012); gather large amounts of data on hard-to-reach populations (Koepfler and Fleischmann 2012); and potentially to pretest to see if attitudes and behaviors not present in current surveys are evident among particular population subgroups.

In addition to attitude and trend tracking, Twitter data can prove useful in the field of population health. Achrekar and colleagues (2011), for example, track Twitter posts containing mentions of influenza in order to create a real-time illustration of the spread of the illness. Heavilin and colleagues (2011) use Twitter data as a means of gathering information on the prevalence of oral health problems and the actions taken to remedy them. Golder and Macy (2011) approach Twitter from a mental health perspective and use data collected from this platform as a means of tracking how sleep patterns and day length impact individuals' moods. As these authors suggest, the candidness of Twitter users in discussing personal matters—such as oral health or emotional status—suggests that health-care providers may begin to use this platform as a tool for monitoring the public's health and communicating with patients.

Regardless of its specific application, these studies and others suggest that Twitter provides a convenient means of developing a broad understanding of

a population's activities and attitudes. In other words, the content of users' tweets provides insight into what Naaman et al. (2011:902) call "social awareness streams." This source of organically created and automatically archived human data allows researchers to see what people are doing, what they are saying, and how they feel about particular issues as these actions and thoughts arise. This is an unprecedented form of data for social scientists with broad research potential, but its use presents new methodological challenges (Lazer et al. 2009). The ability to systematically gather demographic data from this source would greatly expand the potential for research that seeks to capitalize on its availability.

## **Using Twitter for Political Analysis**

One popular application of Twitter data is political analysis and election forecasting. Prompted by promising analyses of political opinion trends within the blogosphere and other social media outlets, some scholars have explored whether Twitter provides a useful outlet for examining political preferences as they develop. There are a number of interesting applications of Twitter data in this burgeoning body of research. Tumasjan and colleagues (2010), for example, collected 100,000 Twitter messages prior to the 2009 German federal election, analyzed these tweets for mentions of party affiliation and positive or negative sentiment, and were able to effectively conclude that the opinion trends reflected in their data paralleled the results of the election. Politicians themselves have noted the popularity of Twitter as a tool for political exchange and many now use this platform as a means of reaching out to potential constituents, though the precise relationship between electoral vulnerability and adoption is open to further exploration (Lassen and Brown 2011). Some researchers, such as Conover and colleagues (2011), have found ways to better understand the dynamics of political discussion within Twitter by mapping interaction and tweet sharing along partisan lines. While these studies carry limitations—such as the ability to gather information only from those who volunteer their opinion or the representation of politically active individuals on Twitter—they nonetheless do highlight the potential uses of Twitter data within this analytical arena.

Although many researchers have found ways to examine political trends using Twitter data, these studies unanimously lack thorough consideration of the demographic trends underlying such phenomena. Adding dimension would greatly benefit researchers looking to predict political outcomes or track changing tides in political opinion or participation among individuals of particular groups, because it would allow for a more nuanced view of

behavior. This trend is not unique to political analysis, and the same can be said of other research that makes use of Twitter data. The addition of demographic data to social media data sets could greatly expand the explanatory and predictive power of analyses that use these data. In the following sections, we propose a method for gathering demographic data in conjunction with information on collective voting intentions as a means of expanding upon existing applications of social media data within social science research.

## **Case Study: Using Twitter to Examine Intention to Not Vote**

The objective of this study is to establish a systematic approach to gathering demographic information about Twitter users, thus overcoming an important limitation of this rich data source. The first step we take is to choose a focal population. This analysis will focus specifically on extracting the demographic characteristics of individuals who express an intention to not vote in the 2012 presidential election. Subsequent steps involve (1) collecting relevant user data from Twitter, (2) organizing and cleaning the resulting data, and (3) extracting demographic information about the users by utilizing profile content. Following these steps, we assess the reliability and accuracy of the demographic information obtained.

The analysis presented here stands out from similar work in two important ways. First, it focuses not on *what* is said (i.e., the content of the tweets) but *to whom* these opinions belong (i.e., the Twitter user who report an intention to not vote). The methods presented here utilize the personal content contained within individuals' Twitter profiles to systematically add a layer of previously unavailable demographic information. Doing so will allow social scientists to not only track trends and opinions using Twitter but also to examine the demographic characteristics of opinion-based networks and to better predict behaviors and attitudes based on social connections.

### ***Twitter: An Overview***

Twitter is a microblogging platform that allows users to record their thoughts in 140 characters or less. The text-based content of these messages may include personal updates, humor, or thoughts on media and politics. This concise format allows users to update their "blogs" multiple times per day, rather than every few days, as is the case with traditional blogging platforms (Java et al. 2007). Besides projecting their thoughts independently, users can

communicate with one another through private messages, by re-tweeting one another's tweets, or by using the *@reply* command to flag content for or about specific users. Twitter users may also contribute to broader conversations channels by including a # (*hashtag*) identifier in their tweet. Twitter allows users to articulate "following" relationships, such that the tweets from those whom the user follows are displayed as a sequential feed that is updated in real time. Twitter was originally intended to be used on mobile devices to facilitate frequent updating, but tweets can also be sent using other Internet-capable devices including tablets and personal computers. Self-presentation (Goffman 1959) on Twitter is negotiated through active conversation as well as the maintenance of personal profiles. To generate this conversation, Twitter users project their thoughts toward an imagined audience of networked individuals (Marwick and Boyd 2010), some of whom bear reciprocal ties to the users themselves and some of whom do not. This is different from other social networking sites such as Facebook, in which all users are reciprocally tied to one another and disclose information mutually. This interesting mix of public and private attention requires users to maintain a balance between transparency and authenticity in the material they choose to tweet (Marwick and Boyd 2010). It is important to note, however, that strong considerations of disclosure do not apply to the entire Twitter population. According to the social media analytics platform Beevolve (2012), only 11.8 percent of all Twitter users choose to "protect" their accounts—meaning the tweets associated with these accounts are only viewable by approved followers. Nonetheless, the strong majority of Twitter users (88.2 percent) manage a public presence.

Presently, many analyses of Twitter data focus on the text of the tweets. This study utilizes other metadata encoded and/or displayed in the Twitter user's profile. Of particular importance to our objective of gathering demographic data are the users' profile pictures—the photograph that the user chooses to represent himself or herself within the site. These pictures, which can be easily mined and stored by the researcher, provide the primary source of information for data collection methods outlined in this article.

### *Who Uses Twitter?*

As of 2014, Twitter reports having approximately 271 million monthly active users (Twitter user statistics can be found at <https://about.twitter.com/company>). According to the Pew Internet and American Life Project (Duggan and Smith 2014), the percentage of Internet users who are on Twitter stood at as of December 2013. This population is dominated by younger individuals (i.e.,



those under the age of 50). African Americans Internet users are more likely than whites or Hispanics Internet users to use Twitter, as are urban dwelling Internet users as opposed to Internet users who live in rural or suburban areas. The same Pew study finds that 31 percent of Internet users between the ages of 18 and 29 use Twitter, compared to 19 percent of Internet users between the ages of 30 and 49, 9 percent of Internet users between the ages of 50 and 64, and 5 percent of Internet users over 65. Likewise, about 29 percent of non-Hispanic black Internet users are on Twitter, compared with 16 percent of white, non-Hispanic Internet users and 16 percent of Hispanic Internet users. Gender is approximately evenly distributed on Twitter: 17 percent of male Internet users are on Twitter and 18 percent of female Internet users also use Twitter.

## **Collecting Data From Twitter**

### *Querying Data Using the Twitter API*

Web-based data collection has gained prominence among social science researchers as a means of collecting large amounts of data to explore topics such as election forecasting, tracking social trends, and time usage (Golder and Macy 2011; Naaman et al. 2011; Tumasjan et al. 2010). This term refers to the process of using an external computer program to extract data from a web platform—which is usually coded in HTML—and organize the data into a readable form. In order to automate communication with a web platform, the program can obtain the platform's API, which is a standardized system of programming instructions that allows web platforms to access and share information from one another. In the same way that the web page's interface provides the user directives for interaction, the API helps guide communication between web programs. When applied to web-based data collection, the API allows the researcher to specify which elements of information he or she wishes to retrieve from the primary platform. Like many web tools, the web platforms often release their API for researchers to use. API-based commands are then embedded within an additional coding language—such as python or Personal Home Page (PHP) as a means of refining the search to include specific key words or queries.

Twitter maintains multiple free, public options for accessing data. Each method has unique advantages and disadvantages. The objectives of the researcher dictate which method is best within a particular context. A common approach to collecting Twitter data involves access through Twitter's streaming API, which provides researchers a queryable sample (max 1 percent of all content) of tweets created in near real time. Data can be gathered

via the streaming API according to specific key words or at random (without key words). Advantages of using this version of the API include its speed and the volume of data available. Disadvantages include a lack of transparency regarding how tweets are sampled within the stream and the potential for missing data if query limits are reached and the connection to the API must pause temporarily.

This article describes an approach that utilizes Twitter's representational state transfer (REST) API (for more information on the Twitter REST API, see <https://dev.twitter.com/rest/public>) to collect new tweets matching key word(s), created within approximately the past nine days. Due to the growing popularity of social media-based analyses by both academic and industry researchers, there are a number of user-friendly tools that facilitate API access. This project uses the online platform ScraperWiki (<https://scraperwiki.com/>) to connect to and gather information from the Twitter REST API. At the time of data collection, ScraperWiki served as a platform through which researchers could share preconstructed segments of code designed to access the Twitter API. Site managers still offer assistance to researchers looking to get started with the API, but programs such as R also offer useful, user-friendly wrappers that help enable access to the Twitter API and automatically parse results into readable data frames.

Twitter's API allows the researcher to gather information using structured queries (for information on how to build a Twitter API query, see <https://dev.twitter.com/rest/public/search>) as a means of accessing information regarding specific topics or user attributes. In this case study, queries were designed to capture information on individuals who express the intention to not participate in the 2012 U.S. presidential election. Using the REST API, the researcher can also exclude individuals who disagree with a particular candidate (e.g., who might say "I'm not voting for Romney") by excluding tweets containing words or phrases that reflect this phenomenon ("for Romney" or "Romney" in this example). We can also exclude many users who are discussing voting in other contexts (e.g., for a contestant on a television show) using key words. Note that this exclusion element requires the researcher to familiarize himself or herself with the nature of the behavior or characteristic at hand in order to develop a preliminary understanding of the terms necessary to exclude (in this case, the intention to not vote in the 2012 U.S. presidential election such as "homecoming" or names of popular, contestant-based television shows).

Due to the idiosyncratic and temporal nature of text information on Twitter, tweets were loosely monitored during the initial data collection process and some exclusion terms were added as they arose within the data. These irrelevant tweets—which compose the minority of the total body of data

**Table 1.** Not Voting Search Queries.

Query	# Tweets
"I am not voting"	1,952
"I'm not voting"	4,583
"I will not vote"	1,965
"I won't vote"	783
"I am not going to vote"	199
"I'm not going to vote	238
"I'm not gonna vote"	159
"I am not gonna vote"	22
"I refuse to vote"	1,149
"I don't plan to vote"	5
"I do not plan to vote"	2
"I didn't register to vote"	21
"I will never vote"	925
"I ain't voting"	994
"I ain't registered"	51
"I did not register to vote"	27
"I'll never vote"	156

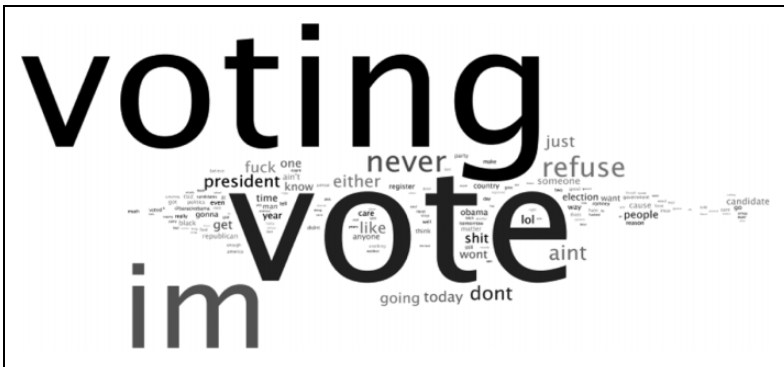
Note: Exclusion terms are "EMA," "AMA," "Romney," "Obama," "xfactor," "x-factor," "x factor," "#xfactor."

collected—were systematically removed during subsequent data preprocessing steps. The process used to clean this information will be discussed in the following paragraphs. A complete list of the queries and exclusion terms used is shown in Table 1.

ScraperWiki was used to collect data from October 16, 2012, until November 9, 2012. During this time, a total of 13,426 tweets from 12,898 unique users were collected. From this data set, we randomly sampled 1,000 unique users. Our data set is relatively small in relation to current research using big social media data, however, working with a small but sufficient random subsample allows us to validate aspects of the data processing—such as removing irrelevant tweets to assess the demographics of intended non-voters—more easily. In addition to this, our work seeks to highlight the capabilities of using social media data for social science research aside from data scale.

### *Cleaning Twitter Data*

In addition to designing queries to help filter out potentially irrelevant tweets to effectively capture the demographics of intended non-voters,



**Figure 1.** Wordle Cloud. This figure illustrates the use of Wordle for preliminary text analysis. Larger terms signify that the words occur more frequently within the document. All words are normalized and stop words are removed from the corpus prior to visualization.

we use a two-pronged approach to clean our data. The first step requires one team member to read through and designate whether the tweets displayed indicate someone's intention to not vote. Second, we clean irrelevant tweets using a key word list assembled using a combination of data-driven techniques, including familiarity with the content of the data through casual reading of the tweets, and the discovery of potentially irrelevant and frequently occurring key words through the use of automated text analysis tools.

We initially used the online text visualization service Wordle ([www.wordle.net](http://www.wordle.net)) to view frequently occurring words, but found this to be fairly unhelpful (Figure 1). We then used the *tm* R package (Feinerer and Hornik 2014) to organize the text of the tweets into a single corpus, stem, normalize and remove stop words from the text, and organize unique terms into a document-term matrix (treating each tweet as a single document). We then used this document-term matrix to view words that appear with designated frequency as a means of detecting potentially irrelevant terms. Reducing our word list to terms that appear at least five times, for instance, allows us to detect terms that seemed not to relate to the 2012 U.S. presidential election, such as “IrvineElection,” or “#healthconscious” which referred to a specific California ballot proposition intended to regulate the labeling of genetically modified organism (GMO) crops. Similar to Wordle, this method did not yield a significant amount of insight and help with data cleaning. However, these methods may prove helpful as a means of auto-

**Table 2.** Key Words Used to Remove Irrelevant Tweets.

Term	Justification
Obama/Romney	“I won’t vote for Romney/Obama” does not indicate intention to not vote at all.
Him/her	Often preceded by “for”—refers to specific candidate, possibly outside of presidential race.
Color/black	Appears often within tweets refuting intention to vote for a particular candidate because of his skin color.
Irvine	Found within hashtag associated with specific University of California—Irvine campaign.
Health	Found within hashtag associated with GMO labeling bill.
Third	Appears in references to intentions to vote for third party candidates.
Voice	May reference television show “The Voice”.
Republican/democrat	May reference decision to not vote for one of two parties.
School	May reference homecoming elections.
Woman	No female presidential candidates; may be irrelevant. May also refer to possible future Hillary Clinton candidacy.

Note: GMO = genetically modified organism.

detecting large portions of off-topic tweets within other corpora and should not be discounted based solely on their application to this analysis. The final key word list used to automatically clean the data is shown in Table 2. These key words were assembled primarily through familiarity with the data content developed by reading through portions of the tweets.

Ultimately, it was decided that due to the nuance of the topic used for this study, hand cleaning provides the most effective means of removing irrelevant tweets. As text analysis methods advance, however, researchers may be able to predict with high accuracy the presence of irrelevant and relevant tweets using machine learning techniques. It is important to note that the text cleaning process is especially important for this project, given that it samples Twitter users who report the intention to engage in a specific activity related to a particular event. A user who does not plan to vote in his or her high school class elections can easily fall into a collection of tweets intended to reflect users who do not plan to vote in the 2012 presidential election. Researchers who plan to gather data using a single-word query that reflects conversation surrounding a particular topic or event (e.g., gathering all tweets that reflect users’ opinion of Mitt Romney using the hashtag #Romney) may collect very few irrelevant tweets and may not need to clean the tweets at all. It is also important

to note that because our unit of analysis for this project is the user rather than the tweet, some users had multiple tweets in the original data collected via the Twitter API, prior to subsampling. If the user's tweet included in the subsample indicated a desire to not vote, it was assumed that the user did in fact intend to not vote.

### *Coding Data From Twitter: Amazon Turkers*

The following section addresses the problem of extracting demographic information from Twitter data. This stage of the data collection process is perhaps the most important methodological contribution of this study. Our approach—crowdsourcing human intelligence—is an essential step in extracting demographic information encoded as images rather than text on users' profiles. The following section describes AMT—a platform through which individuals can pay workers to perform short tasks for small fees—and the way in which this tool was used as a means of coding demographic data. We discuss the details of this data collection procedure as well as the ways in which AMT has been successfully used/implemented as a resource in previous studies.

### *AMT*

AMT is a marketplace for work that requires human rather than artificial intelligence. Within this platform, individuals, known as *requesters*, post brief tasks that can be performed in minutes in exchange for a dollar or less. These small assignments—called human intelligence tasks (HITs)—typically involve requests that are difficult or impossible for artificial intelligence systems to complete. Examples include tagging images, transcribing text from images, or answering questions about website content. Requesters have the ability to customize the price, format, and duration of their HITs as well as set qualifications for the employees—referred to as Turkers within this study—who are permitted to view and/or complete these HITs. Turkers are anonymous, independent contractors who are identifiable only by their unique ID numbers. Each Turker's work history and overall approval rating are also available for view and can be used by requesters as a qualification on which requesters can filter their HIT accessibility. Despite their anonymity, some demographic information about the Turkers is known as the result of past survey research efforts. In addition to this, the survey instrument used for this study gathered administrative data about the Turkers themselves.

## *Use of Amazon Turkers for Social Science Research*

Previous studies have indicated that Turkers can be highly reliable experimental research subjects (Mason and Suri 2012). It has been shown that Turkers behave and react similarly to purposively sampled research subjects within laboratory settings and produce results of comparable quality (Bhurmester, Kwang and Gosling 2011; Mason and Suri 2012). Furthermore, using the AMT platform often allows experimental researchers to quickly and easily reach out to a larger, more stable and more diverse population than they might have been able to otherwise. In research experimenting directly with Turkers, Snow et al. (2008) found that in regard to many language processing tasks such as affect determination, Turkers are just as effective as and less expensive than expert labelers. Marge and colleagues (2010) affirm the ability of Turkers to transcribe audio files. Of the 20,116 words transcribed by the Turkers in this study, only 997 (4.96 percent) contained errors. Urbano and colleagues (2010) asked Turkers to categorize pieces of music based on similarity, and again found that the Turkers performed approximately as well as experts for a lesser price.

In addition to serving as a successful platform for experimental research, AMT provides a valuable space for survey distribution. Researchers have expressed positive attitudes toward the potential accuracy and representativeness of the Turkers as survey subjects. Behrend et al. (2011), for example, distributed a short survey to both the Turkers and a sample of university students as a means of comparing the psychometric properties of each. This study found that Turkers and university students behaved similarly and displayed similar judgment but that the Turkers held a significant advantage for survey research in that they comprise a significantly more diverse respondent pool.

In terms of demographic representation, Ipeirotis (2010) find that populations of Turkers are concentrated within two primary locations—approximately 50 percent are from the United States and 40 percent are from India. Turkers are overwhelmingly female (approximately 70 percent) and younger than the general population (54 percent of Turkers are between the ages of 21 and 35). Turkers also have a slightly lower yearly income than the general population of U.S. Internet users. Over 60 percent of U.S.-based Turkers have incomes below US\$60 K. They also have small families (55 percent have no families).

The characteristics of the Turkers who participated in this study generally parallel those surveyed by Ipeirotis (2010). In regard to Turker selection, we administered our HITs to U.S.-based Turkers only in order to render assessments of race—a socially constructed characteristic—

**Table 3.** Turker Characteristics.

HIT Type	Master's Required ( <i>n</i> = 44; percent)	Master's Not Required ( <i>n</i> = 111; percent)
MTurk is main income source	40.9	37.8
Education		
High school	4.5	9.9
Some college	29.5	27.0
Associate's degree	13.6	9.9
Bachelor's	43.2	43.2
Graduate degree, master's	6.8	7.2
Graduate degree, doctorate	2.3	1.8
Age		
19–25	6.8	20.7
26–35	65.9	54.1
36–45	27.3	22.5
Over 60	0.0	1.8
Sex		
Male	54.5	60.3
Female	45.5	38.7

Note: HIT = human intelligence tasks

comparable to existing studies of U.S. voting intentions and habits. We also issued two batches of HITs. One batch was administered to Turkers who had acquired “master’s status”—or the documented completion of 1,000 HITs or more with at least a 95 percent approval rating (referred to as the “masters required” batch in subsequent sections) as well as the general Turker population (referred to as the “master’s not required batch” in subsequent sections). The Turkers who ended up participating in our study are highly educated (52 percent of both master’s required and “masters not required” Turker participants have a bachelor’s degree or higher). They are also young. The majority of Turkers who participated in both the master’s required and the master’s not required batches are 35 years old or younger. Many report that AMT is their main source of income as well (40.9 percent for master’s required and 37.8 percent for master’s not required batch). Finally, it is interesting to note that a relatively small number of Turkers completed the HITs administered. Only 44 unique Turkers participated in the master’s required batch, for instance, which equates to an average of approximately seven HITs per Turker (see Tables 3 and 4).

Despite promising research findings, it is important to note that there are challenges associated with the use of the Mechanical Turk. Mechanical Turk



**Table 4.** Turker HIT Activity.

HIT Type	Master's Required ( <i>n</i> = 44; percent)	Master's Not Required ( <i>n</i> = 111; percent)
Mean # of HITs completed	7	3
Hours per week on Turk website		
1–2 Hours	0.0	3.6
2–4 Hours	0.0	7.2
4–8 Hours	6.8	15.3
8–20 Hours	25.0	21.6
20–40 Hours	34.1	27.9
40 Hours or more	34.1	24.3

Note: HIT = human intelligence tasks

is a dynamic, evolving platform with an ever-changing workforce. It is important that researchers keep up-to-date with the affordances and requirements of the platform. In addition to this, they must be careful to design HITs that are easy for the Turkers to complete, as well provide sufficient pay to ensure Turker cooperation. Researchers may find themselves managing hundreds or thousands of independent, anonymous coders when using this platform. Therefore, it is important that researchers assemble, pretest, and monitor their HITs carefully. Turkers can also communicate with requesters via e-mail, and these messages often provide valuable feedback for those designing the HIT(s).

### *Description of Methodology*

In order to gather demographic information on the Twitter users who report the intention to not vote in the 2012 presidential elections, Turkers were asked to view each user's profile picture and evaluate their sex, age, race, grooming, and attractiveness. Categories for sex included male and female. For age, Turkers were asked to identify Twitter users according to both a numeric age range (from below 12 to 60+ years old) and a general age categories (child, adolescent, adult, and senior). Evaluations of attractiveness and grooming were both measured on a 5-point, ascending Likert-type scale ranging from very unattractive/very poorly groomed to very attractive/very well groomed. These questions were drawn from the National Longitudinal Study of Adolescent to Adult Health. Since a large body of research has linked attractiveness to many aspects of success in off-line social settings (Anderson, Adams, and Plaut 2008; Umberson and Hughes 1987), this trait

may be important to study online as well. Therefore, it may be important to consider the reliability and accuracy of the Turkers in measuring these traits. Finally, users were asked to identify whether the photo displayed a person, a group of people, something other than a person (a pet, a cartoon, a logo, etc.), or whether the user's profile image link was broken (indicating that they had changed their profile photo between the time of data collection and metadata evaluation). Also included in the survey were questions regarding the characteristics of the Turkers. Turkers were asked to state their sex, age, education level, the amount of time they spend per week on the AMT, and whether the AMT provides their primary source of income. The full survey instrument is included in Online Appendix B.

Each HIT required Turkers to assess 10 photos. Completion of one HIT yielded US\$0.60, which equated to an average wage of US\$6.97 per hour for the master's not required batch and US\$6.45 per hour for the master's required batch. In order to test their reliability, three separate Turkers completed each HIT. The completion time for the master's required HIT batch was 49 minutes, and the completion time for the master's not required batch was 44 minutes.

### *Assessing Turker Reliability and Accuracy*

Table 5 explores the reliability of the Turkers in the master's required and master's not required batches. This table displays the number of HITs for which zero, two, or three of the three Turkers agree on the evaluated characteristics of the user. Overall, the Turkers prove to be very reliable in regard to their assessment of sex, age, age categorization, and race. In regard to user demographics, sex is the most reliably coded—between 98.1 and 98.6 of all users' sex was agreed upon by at least two Turkers. Even more subjective questions—such as attractiveness and grooming yield fairly high reliability measures. Finally, we note that the reliability measures do not seem to be very different for the master's required and master's not required batches.

In addition to assessing the reliability of the Turkers' evaluations, we test accuracy by comparing their evaluations to evaluations from expert human coders trained by the research team to complete identical HITs. These expert coders were given the same metadata as the Turkers—namely, the users' profile photos—as well as the same survey instrument administered to the Turkers. Because there was a time lag between when the Turkers coded the photos and when the expert coders were able to code the photos, we assess accuracy based solely on photos for which the expert coders indicate the

**Table 5.** Turker Reliability for Individual Photo Assessment ( $n = 1000$  user photos).

	% Three Agree	% Two of the Three Agree	% None Agree
Age			
Master's required	55.9	40.7	3.4
Master's not required	54.6	42.0	3.4
Age category			
Master's required	62.6	35.3	2.1
Master's not required	59.9	38.0	2.1
Race			
Master's required	71.1	20.0	4.7
Master's not required	70.6	24.7	4.7
Sex			
Master's required	83.3	14.8	1.9
Master's not required	82.0	16.6	1.4
Attractiveness			
Master's required	29.3	56.1	14.6
Master's not required	31.5	53.5	15.0
Grooming			
Master's required	28.1	54.4	17.5
Master's not required	29.8	55.0	15.2
Is a person?			
Master's required	92.1	7.8	0.1
Master's not required	86.4	13.0	0.6

**Table 6.** Assessing Turker Accuracy Against Expert Coders.

Characteristic	Percentage of User Photos Experts Coders/Turkers Agreed Upon	
	Master's Required	Master's Not Required
Sex ( $N = 635$ user photos)	91.8	91.2
Race	82.5	81.6
Age	68.0	67.9
Age category	77.0	76.5
Attractiveness	50.2	49.0
Grooming	53.4	53.5
Is person?	97.0	96.1

image link is not broken (i.e., the user has changed their profile photo since the date of the Turker evaluation). Table 6 displays the proportion of photos evaluated by both the Turkers and the expert coders for which they agreed

**Table 7.** Cohen's  $\kappa$  Scores for Expert Coder/Turker Agreement.

Characteristic	Master's Required	Master's Not Required
Sex	0.90	0.87
Race	0.80	0.77
Age	0.53	0.54
Age category	0.59	0.59
Attractiveness	0.42	0.39
Grooming	0.29	0.32
Is person?	0.94	0.92

about the sex, race, grooming, attractiveness, and age of the user. This table informs us that the ratings of the Turkers and expert coders are very similar. Examining Cohen's  $\kappa$  values for the agreement between the master's required and the master's not required Turker evaluations and expert coder evaluations indicate that agreement for sex and race is substantial to almost perfect ( $\kappa = 0.87$ – $0.90$  and  $\kappa = 0.77$ – $0.80$ , respectively), and agreement for numeric age and age category is moderate to substantial ( $\kappa = 0.53$ – $0.54$  and  $\kappa = 0.60$ – $0.59$ , respectively; Table 7). Confusion matrices for each rating (found in Online Appendix C) suggest that expert coders are more likely to categorize photos of group members or nonpeople (perhaps avatars or cartoon characters) as belonging to particular demographic categories than the Turkers. This difference is important to note for researchers interested in using the Turkers to code these characteristics.

### **Application: Examining Demographic Characteristics of Intended Non-voters**

Given our confidence in the reliability of the Turkers, we then consider the results of the Turkers' evaluations of Twitter user profile pictures. Tables A1–A5 in Online Appendix A display the demographic characteristics of the intended nonvoters sampled for this study as determined by the Turkers. These tables are broken down to display evaluations for data that were (a) hand coded to remove irrelevant tweets and (b) cleaned based on the presence of irrelevant key words for both the master's required and the master's not required batches. Note that for a Twitter user to be categorized in a particular way for any given question two or more Turkers had to agree upon that categorization. In this section, we consider only the objective characteristics coded by the Turkers—not attractiveness and grooming. Note that while we assessed the reliability and accuracy of the Turker assessment using

**Table 8.** Pew Institute Data on Nonvoters.

Sex	
Men	52 percent
Women	48 percent
Race/ethnicity	
White, non-Hispanic	59 percent
Black, non-Hispanic	10 percent
Hispanic	21 percent
Age	
18–29	36 percent
30–49	35 percent
50–64	20 percent
65+	8 percent

the full 1000 user subsample, assessing the demographics of intended nonvoters requires us to filter out irrelevant tweets either by hand (full filter) or using keywords (partial filter).

According to these results, intended nonvoters on Twitter tend to be male (45.51–48.49 percent) adults between the ages of 19 and 35 (65.68–69.23 percent). The majority of intended nonvoting Twitter users are reportedly white (41.97–43.89 percent) followed in descending frequency by black intended nonvoters (26.95–29.10 percent), Hispanic intended nonvoters (5.01–7.02 percent), and Asian intended nonvoters (1.84–2.80 percent).

Although the intention of this article is not to provide empirical evidence regarding the demographic characteristics of individuals who state an intention to not vote in the 2012 presidential election, we attempt to contextualize the results by comparing them to data collected for a 2012 Pew Center report on nonvoters (Kohut et al. 2012; see Table 8). This comparison is not intended to suggest that, given the current state of statistical modeling in this area, Twitter should be used to estimate population proportions or the sizes of certain populations. Instead, we currently find that the most compelling uses of Twitter data lie in studying real-time dynamics of social interactions, as we discuss below.

As expected, the data presented in Tables 5–7 do not exactly parallel the data on national nonvoters gathered by Pew. There are a number of reasons why we would not expect these values to line up directly. For one, the demographic composition of the Twitter population is different than that of the U.S. population. In addition to this, discrepancies between the demographic data on nonvoters gathered in this study may be attributable to social desirability effects in previous surveys. According to Belli et al. (1999), many

researchers believe that traditional surveys have a tendency to underrepresent the total number of nonvoters in a given election, as nonvoters often refuse to disclose this information for reasons of self-presentation. Given this systematic bias in survey design, it is possible that collecting information on deviant behaviors through Twitter—a space characterized by high levels of self-disclosure—provides *more* information on the individuals who engage in these behaviors than traditional surveys (Marwick and Boyd 2010). Future research may determine whether this is the case. Finally, the Twitter population of interest in this article is composed of individuals whose stated intention is to not vote, and research suggests that there may not be a strong relationship between voting intentions and voting outcomes (Rogers and Aida 2011).

We do see some parallels between our data and the nonvoters examined by Pew. Both the Pew and Twitter estimates suggest that nonvoters tend to be white, male, and young (near or below the age of 35). It is clear that the Twitter estimates far exceed the Pew estimates in regard to the proportion of nonvoters who are black and young. However, these specific inconsistencies are likely attributable to two factors. For one, the population composition of Twitter does not align with the national population. As mentioned previously, Twitter is dominated by younger (ages 18–29), non-Hispanic black users (Duggan and Smith 2014). Overall, when discrepancies between the data collected from Twitter on those who state an intention to not vote and existing data on the nonvoting population are considered, these estimates support the face validity of our intended-nonvoter demographic estimates.

## Discussion

It is becoming widely acknowledged that “social media offers us the opportunity for the first time to both observe human behavior and interaction in real time and on a global scale” (Golder and Macy 2012:7). Currently, the majority of researchers who are taking advantage of social media data for social science research are not social scientists but rather computer scientists and others in related fields. Perhaps one reason for this trend is the fact that key pieces of information for sociological—and specifically demographic—research, such as age, race, and gender, are currently difficult to extract from social media sites such as Twitter. Adding demographic information to Twitter data increases the breadth of social science research to which these data may be applied. Doing so opens the possibility for engaging in research that examines not only collective attitudes and opinions but also the composition of the groups driving these trends. This information could also be

incorporated into network data and used as a means of examining the structure of groups that display particular behaviors or opinions, as well as how this structure changes over time.

In this article, we present an approach to extracting, processing, and analyzing the data from Twitter that can be modified to suit many topics of research. We believe that social media data, such as Twitter, present an opportunity to develop a fundamentally different approach to social science research. As with all new data collection, Twitter has certain limitations to overcome. Although many capacities of Twitter data parallels that of existing data collection, it does not replicate the results of these methods and poses new challenges. Our goal, however, is to suggest reliable and accurate methods for overcoming these challenges in order to make Twitter an accessible resource for a larger number of social scientists. While there are limits to the scalability of these methods, we believe that the use of the Mechanical Turk allows researchers to process larger amounts of data in a timelier and potentially more cost-effective way than they would using hired coders. Furthermore, we believe this analysis framework can be adapted in ways that allow social science researchers to make strong contributions in the growing field of big social data analysis.

Our preliminary analysis indicates that it is possible to collect demographic information on Twitter users using a combination of available, easy-to-use technologies. Gathering raw data from Twitter using the website's API is simple and quick. Although a small sample of 1,000 users was used in this exploratory study, the data collection platform used in this study managed to collect a fairly large sample of individuals who publicly report an intention to vote in the 2012 presidential election ( $N = 12,898$  unique users). Obtaining evaluations of the Twitter user's demographic characteristics—including sex, age, and race—using the AMT proved efficient and effective. Similar to previous research (Behrend et al. 2011; Bhurmester, Kwang, and Gosling 2011; Mason and Suri 2012), our results indicate that Turkers are a reliable resource for coding information. Although the data cleaning techniques suggested in this study require further exploration and may not apply to queries of all types, the results of these data collection efforts are nonetheless promising. Though the resultant demographic breakdown of intended nonvoters presented in this article do not parallel existing, national data on nonvoters exactly, they nonetheless yield results that make sense given the known biases of the Twitter environment. We are confident that the data yielded through these methods could be used to develop more complex models for social analysis.

### *Advantages of Using Twitter for Demographic Data Collection*

There are a number of advantages associated with the use of Twitter as source of social data. To begin, Twitter data are abundant and relatively easy to access. Among the approximately 500 million currently registered Twitter users, approximately 88.2 percent are not protected, meaning that all published content is available for view to all web users. This published material is considered public data. Twitter users do not need to issue approval for researchers to use their profile information. Although laws regarding the use of Twitter information as public data may change in the future, social scientists have the opportunity to capitalize on the availability of Twitter data as predocumented insight into the collective attitudes, opinions, and behaviors of Internet users. In addition to this advantage, microblogging websites such as Twitter are often updated multiple times per day, which allows the researcher to track opinions and actions as they emerge and develop. While traditional surveys accomplish a similar task, they are nonetheless time consuming and costly to administer and cannot provide the same minute-to-minute insight that Twitter data can.

Beyond availability, Twitter data are often easy and free to collect. There are a number of tools available that allow researchers to collect archived information within social media sites such as Twitter without requiring extensive coding knowledge. The platform used for this study is Scrapperwiki, an open-source platform for REST API-based data collection in which members can develop and share the code to gather information from particular websites. Researchers may also use wrappers available through R—a free, collaborative computing language and software environment used primarily for statistical and graphical analyses—to query the Twitter API and parse resultant data.

Finally, Twitter provides ready access to certain populations that are difficult to reach using other means due to the voluntary nature of the information disclosed within the site. As discussed earlier, representation of self on Twitter is unique from other social networking platforms. Although users must sign in via a password-protected web portal to post tweets, the majority of Twitter profiles (88.2 percent) are visible to all Internet users (Beevolve 2012). In addition to this, networks within Twitter are directed and often contain a mixture of familiar and unfamiliar connections. Given these conditions, norms for disclosure on Twitter are ambiguous. Twitter users must maintain an online presence that is simultaneously polished and genuine (Marwick and Boyd 2010). Occasionally, these users utilize Twitter as a platform for unfiltered personal expression and admit to nonnormative ideas and actions. In addition to



individuals who express refusal to vote in the 2012 presidential election, preliminary analyses for this study found a number of individuals who engage in deviant behaviors such as drunk driving or expressing racial slurs.

Although Twitter data are not suitable for all research questions, there are particularly interesting applications that may serve to expand our knowledge about social processes. There are also ways to leverage apparent weaknesses of Twitter for scientific purposes. For example, because of the open nature of Twitter, individuals can follow a profile without being a “friend” of the person tweeting. This may allow researchers to examine the influence of weaker social network ties for behaviors and opinions.

While Twitter is not representative of the total U.S. population, this does not negate the use of Twitter to examine social questions or for theory generation. In fact, the overrepresentation of African Americans and young adults on Twitter can be used to better understand populations that are often underrepresented in most surveys. In addition, similar to the case study approach, Twitter can be used for developing theoretical generalizations, if not statistically generalizable conclusions (Small 2009). In other words, although these data cannot be used to draw conclusions about the actions and behaviors of any population beyond that of Twitter users specifically, it can nonetheless be used to make statements about social processes in general.

In addition to this, Twitter allows researchers to examine the impact of short-term events on behaviors and attitudes in ways we would not be able to do on such a large scale with a survey. These changes could presumably be analyzed on a scale as minute as week to week or day to day. Some researchers—such as De Loungueville, Smith, and Luraschi (2009)—have addressed this capacity by using Twitter data as a means of tracking reactions to time sensitive events such as forest fire outbreaks. Nonetheless, adding demographic data to these analyses would expand possibilities for model building and the capacity for making predictions.

### *Challenges of Using Twitter for Demographic Data Collection*

One major challenge associated with the use of Twitter data for social science research is the idiosyncratic nature of the data. Unpredictable content such as bots and tweets unrelated to a specific query may find their way into the data and skew results. In this article, the removal of irrelevant tweets is important. This case study attempts to provide a method for removing irrelevant tweets that allows the researcher to forego coding each tweet by hand, as this can prove costly and time consuming for a research team. Future research may address this challenge of data cleaning by utilizing automated

text analysis techniques or subsetting and hand coding a portion of a larger body of data and using the demographic information garnered from this subset when building predictive models. It is important to note, however, that for this study results remained fairly consistent regardless of the filtering approach used (i.e., tweets removed using a semiautomated word search, and irrelevant tweets removed after hand coding).

In addition to handling the unpredictability of user-generated data, analyses that use Twitter data must be careful to consider issues of representation when interpreting results. It is important to state that this article's data collection approach provides information about a very particular respondent pool: individuals who report the intention to not vote on Twitter. As indicated in the application portion of this article, it is clear that Twitter users are not representative of the national population. In addition to this, researchers often need to sample based on the tweet when their desired unit of analysis is the user, which adds further complication to the issue of representative user sampling and requires forethought regarding how to handle the tweets and metadata for users who appear multiple times in the data.

Furthermore, this analysis relies on voluntary information. Its findings represent only those individuals who offer information about their voting intentions. It does not reflect individuals who did not vote *and* did not report these intentions or individuals who *claimed* to have no intention to vote but did nonetheless. Indeed, there is likely a group of individuals who are making false claims and may be providing erroneous profile information. Identifying these individuals will require the use of network information and other advanced verification techniques. We believe that the targeted sampling strategy (collecting profiles based on specific behaviors/attitudes) we use reduces the likelihood we are drawing from fake accounts or avatars, however future research will be necessary to better handle this issue.

Finally, in regard to the possibility of using Twitter user metadata for the sake of demographic research—not all users include in their profiles identifiable photos. Similarly, Twitter users' racial and gender identities explain in may not align perfectly with the way in which others assess them. This study does not directly address these limitations. Rather, it focuses on whether—in general—Turkers provide accurate and reliable evaluations of characteristics not directly encoded within user profiles. Future work may delve further into the mechanisms that link users' self-reported demographic characteristics and the way in which they choose to disclose or not disclose this information through their Twitter profiles.

Overall, we acknowledge that researchers should exercise caution when studying social phenomena using online data—including data from Twitter.

This study seeks to highlight challenges associated with conducting user-centered analyses of Twitter data. Many other studies—such as Mislove, Lehman, and Ahn (2011), Wong, Sen, and Chiang (2012), and Burgess and Brunson (2012)—explain in detail challenges such as representativeness and data accessibility when using Twitter data for social science research. Similar to the way in which Lazer et al. (2014) caution that big data are not better data, it may also be said that widely available data—regardless of scale—are not a panacea for social research. Nonetheless, we maintain that Twitter provides valuable insight into human behavior when harnessed properly and that adding demographic dimensionality to these data can open new pathways for social science research.

## Conclusion

Twitter is arguably the largest observational study of human behavior to date. Not only is this source of data large and easily accessible by social scientists, we contend that there is a tremendous opportunity for sociologists to use Twitter data for social science research. However, we recognize that currently a barrier exists regarding the use of these data for demographic research. The purpose of this article is to suggest a reliable and accurate means of gathering demographic data from Twitter—including age, race, and gender—as a means of overcoming this challenge. Supplementing textual data from Twitter with this additional information could open up brand new opportunities for social research and could allow demographers to model and predict behaviors and attitudes on a large scale and/or among difficult to reach populations. Restated, the potential for Twitter in social science research is yet to fully be articulated. However, we believe there are exciting opportunities to use this research to investigate social problems and other phenomena, but as a research community we cannot fully explore these opportunities without (a) widespread access and familiarity with the data by social scientists and (b) reliable information about users' demographic characteristics. These are tremendous challenges, but overcoming them is worthwhile if doing so allows social science to play a role in utilizing one of the largest sources of social information available.

## Acknowledgments

This work is supported by the U.S. Army Research Office under project 62389-CS-YIP. We thank the reviewers, associate editor, and editor for helpful comments.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors received funding from the U.S. Army Research Office under project 62389-CS-YIP to complete this research.

### Supplemental Material

The online appendices are available at <http://smr.sagepub.com/supplemental>.

### References

- Achrekar, Harshavardhan, Avinash Gandhe, Ross Lazarus, Ssu-Hsin Yu, and Benyuan Liu. 2011. "Predicting Flu Trends Using Twitter Data." First International Workshop on Cyber-Physical Networking Systems (CPNS) 2011, IEEE Infocom, Shanghai, China.
- Anderson, Stephanie L., Glenn Adams, and Victoria C. Plaut. 2008. "The Cultural Grounding of Personal Relationship: The Importance of Attractiveness in Everyday Life." *Journal of Personality and Social Psychology* 95:352-68.
- Beevolve. 2012. "An Exhaustive Study of Twitter Users across the World." Retrieved March 1, 2013 (<http://www.beevolve.com/twitter-statistics/>).
- Behrend, Tara S., David J. Sharek, Adam W. Meade, and Eric N. Wiebe. 2011. "The Viability of Crowdsourcing for Survey Research." *Behavior research methods* 43: 800-13.
- Belli, Robert F., Michael W. Traugott, Margaret Young, and Katherine A. McGonagle. 1999. "Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring." *Public Opinion Quarterly* 63:90-108.
- Boyd, Danah. 2009. "Twitter: 'Pointless Babble' or Peripheral Awareness + Social Grooming?" Vol. 2012. Retrieved May 17, 2012 ([http://www.zephorias.org/thoughts/archives/2009/08/16/twitter\\_pointle.html](http://www.zephorias.org/thoughts/archives/2009/08/16/twitter_pointle.html)).
- Brickman Bhutta, Christine. 2012. "Not by the Book: Facebook as a Sampling Frame." *Sociological Methods & Research* 41:57-88.
- Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-quality, Data?" *Perspectives on Psychological Science* 6:3-5.
- Burgess, Jean and Axel Bruns. 2012. "Twitter Archives and the Challenges of 'Big Social Data' for Media and Communication Research." *M/C Journal* 15:1-7.

- Cambria, Erik, Haixun Wang, and Bebo White. 2014. "Guest Editorial: Big Social Data Analysis." *Knowledge-Based Systems* 69:1-2.
- Cassa, Christopher A., Rumi Chunara, Kenneth Mandl, and John S. Brownstein. 2013. "Twitter as a Sentinel in Emergency Situations: Lessons from the Boston Marathon Explosions." *PLoS Currents*: 1-5.
- Conover, M., J. Ratkiewicz, M. Francisco, B. Gonçalves, A. Flammini, and F. Menczer. 2011. "Political Polarization on Twitter." Paper presented at the 5th International Conference on Weblogs and Social Media (ICWSM), July 17-21, Barcelona, Spain.
- De Longueville, Bertrand, Robin S. Smith, and Ginaluca Luraschi. 2009. "Omg, from Here, I Can See the Flames!: A Use Case of Mining Location Based Social Networks to Acquire Spatio-temporal Data on Forest Fires." Pp. 73-80 in *Proceedings of the 2009 International Workshop on Location Based Social Networks*, November 3, Seattle, WA: Association for Computing Machinery (ACM).
- Diakopoulos, Nicholas A. and David A. Shamma. 2010. "Characterizing Debate Performance Via Aggregated Twitter Sentiment." Pp. 1195-98 in *Proceedings of the 28th International Conference on Human Factors in Computing Systems*. Atlanta, GA: Association for Computing Machinery (ACM).
- Duggan, Maeve and Aaron Smith. 2014. "Social Media Update 2013." Pew Internet and American Life Project. Retrieved September 1, 2014 (<http://www.pewinternet.org/2013/12/30/social-media-update-2013/>).
- Feinerer, Ingo and Kurt Hornik. (2014). tm: Text Mining Package. R package version 0.6. Retrieved May 13, 2013 (<http://CRAN.R-project.org/package=tm>).
- Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Doubleday.
- Golder, Scott A. and Michael W. Macy. 2011. "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength across Diverse Cultures." *Science* 333:1878-81.
- Golder, Scott A. and Michael Macy. 2012. "Social Science with Social Media." *ASA Footnotes* 40:7.
- Heavilin, N., B. Gerbert, J. E. Page, and J. L. Gibbs. 2011. "Public Health Surveillance of Dental Pain via Twitter." *Journal of Dental Research* 90:1047-51.
- Holbrook, Allyson L. and Jon A. Krosnick. 2010. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." *Public Opinion Quarterly* 74:37-67.
- Ipeirotis, Panagiotis G. 2010. *Demographics of Mechanical Turk (Tech. Rep. No. CeDER-10-01)*. New York: New York University. Retrieved December 21, 2012 (<http://hdl.handle.net/2451/29585>).
- Janus, Alexander L. 2010. "The Influence of Social Desirability Pressures on Expressed Immigration Attitudes." *Social Science Quarterly* 91:928-46.

- Java, Akshay, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. "Why We Twitter: Understanding Microblogging Usage and Communities." Pp. 56-65. in *Proceedings of the 9th WebKDD and 1stSNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*. San Jose, CA: Association for Computing Machinery (ACM).
- Koepfler, Jes A. and Kenneth R. Fleischmann. 2012. "Studying the Values of Hard-to-reach Populations: Content Analysis of Tweets by the 21st Century Homeless." Pp. 48-55 in *Proceedings of the 2012 iConference*: Association for Computing Machinery (ACM).
- Kohut, Andrew, Carroll Doherty, Michael Dimock, and Scott Keeter. 2012. *Nonvoters: Who They Are, What They Think*. Pew Research Center for the People & the Press. Retrieved February 23, 2013 ([www.people-press.org/2012/11/01/nonvoters-who-they-are-what-they-think/4/](http://www.people-press.org/2012/11/01/nonvoters-who-they-are-what-they-think/4/)).
- Lassen, David S. and Adam R. Brown. 2011. "Twitter: The Electoral Connection?" *Social Science Computer Review* 29:419-36.
- Lazer, David, Devon Brewer, Nicholas Christakis, James Fowler, and Gary King. 2009. "Life in the Network: The Coming Age of Computational Social Science." *Science* 323:721-23.
- Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "Big Data. The Parable of Google Flu: Traps in Big Data Analysis." *Science (New York, NY)* 343:1203-05. Retrieved June 2, 2014 (<http://www.ncbi.nlm.nih.gov/pubmed/24626916>).
- Lowe, John B., Margaret Barnes, Cynthia Teo, and Stephanie Sutherns. 2012. "Investigating the Use of Social Media to Help Women from Going back to Smoking Post-partum." *Australian and New Zealand Journal of Public Health* 36:30-32.
- Marge, M., S. Banerjee, and A. I. Rudnicky. 2010. "Using the Amazon Mechanical Turk for Transcription of Spoken Language." Pp. 5270-73 in *Proceedings of the 2010 IEEE Conference on Acoustics, Speech and Signal Processing*, edited by J. Hansen. Dallas, TX: Institute of Electrical and Electronics Engineers (IEEE).
- Marwick, Alice E. and Danah Boyd. 2010. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience." *New Media & Society* 13:114-33.
- Mason, Winter, and Siddharth Suri. 2012. "Conducting Behavioral Research on Amazon's Mechanical Turk." *Behavior Research Methods* 44:1-23.
- Mislove, Alan, S. Lehmann, and Y. Y. Ahn. 2011. "Understanding the Demographics of Twitter Users." *ICWSM*. Retrieved August 1, 2014 (<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2816/3234>).
- Moreno, Megan A., Allison Grant, Lauren Kacvinsky, Katie G. Egan, and Michael F. Fleming. 2012. "College Students' Alcohol Displays on Facebook: Intervention Considerations." *Journal of American College Health* 60:388-94.

- Naaman, Mor, Hila Becker, and Luis Gravano. 2011. "Hip and Trendy: Characterizing Emerging Trends on Twitter." *Journal of the American Society for Information Science and Technology* 62:902-18.
- Reips, Ulf-Dietrich and Pablo Garaizar. 2011. "Mining Twitter: A Source for Psychological Wisdom of the Crowds." *Behavior Research Methods* 43: 635-42.
- Rogers, Todd and Masa Aida. 2011. "Why Bother Asking? The Limited Value of Self-reported Vote Intention." HKS Working Paper No. RWP12-001. Available at SSRN. Retrieved February 23, 2015 (<http://ssrn.com/abstract=1971312>).
- Small, Mario Luis. 2009. *Unanticipated Gains: Origins of Network Inequality in Everyday Life*. Oxford, UK: Oxford University Press.
- Snow, Rion, B. O'Connor, Daniel Jurafsky, and A. Y. Ng. 2008. "Cheap and Fast—But Is It Good?: Evaluating Non-expert Annotations for Natural Language Tasks." Pp. 254-65 in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii: Association for Computational Linguistics.
- Sutton, Jeannette N., Emma S. Spiro, Britta Johnson, Sean M. Fitzhugh, Mathew Greczek, and Carter T. Butts. 2012. "Connected Communications: Network Structures of Official Communications in a Technological Disaster." in *Proceedings of the International Systems for Crisis Response and Management (ISCRAM) 2012*, edited by Leon Rothkrantz Jozef Ristvej and Zeno Franco. Vancouver, Canada: Simon Fraser University. Retrieved January 26, 2015 ([http://www.iscramlive.org/ISCRAM2012/proceedings/ISCRAM2012\\_proceedings.pdf](http://www.iscramlive.org/ISCRAM2012/proceedings/ISCRAM2012_proceedings.pdf)).
- Tourangeau, Roger and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133:859-83.
- Tumasjan, A., T. O. Sprenger, P. G. Sandner, and I. M. Welp. 2010. "Election Forecasts with Twitter: How 140 Characters Reflect the Political Landscape." *Social Science Computer Review* 29:402-18.
- Umberson, Debra and Michael Hughes. 1987. "The Impact of Physical Attractiveness on Achievement and Psychological Well-being." *Social Psychology Quarterly* 50:227-36.
- Urbano, Julián, Jorge Morato, M. Marrero, and Diego Martín. 2010. "Crowdsourcing Preference Judgments for Evaluation of Music Similarity Tasks." Pp. 9-16 in *Proceedings of the ACM SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE 2010)*, edited by M. Lease, V. Carvalho, and E. Yilmaz. Geneva, Switzerland: Association for Computing Machinery (ACM).
- Valkenburg, Patti M., Jochen Peter, and Alexander P. Schouten. 2006. "Friend Networking Sites and Their Relationship to Adolescents' Well-being and Social Self-esteem." *CyberPsychology & Behavior* 9:584-90.

- Wong, Felix Ming Fai, Soumya Sen, and Mung Chiang. 2012. "Why Watching Movie Tweets Won't Tell the Whole Story?" Pp. 61-66 in *Proceedings of the 2012 ACM Workshop on Online Social Networks*, August 13–17, Helsinki, Finland: Association for Computing Machinery (ACM).
- Yardi, Sarita and Danah Boyd. 2010. "Dynamic Debates: An Analysis of Group Polarization over Time on Twitter." *Bulletin of Science, Technology & Society* 30:316-27.

## Author Biographies

**Tyler H. McCormick** is an assistant professor of statistics and sociology of the University of Washington. He is a core faculty member in the Center for Statistics and Social Sciences and a Senior Data Science Fellow of the eScience Institute, both at the University of Washington.

**Hedwig Lee** is an associate professor of sociology at the University of Washington in Seattle. She is also a faculty affiliate of the Center for Research on Demography and Ecology and Center for Statistics and the Social Sciences and coleads the Northwest Region Scholars Strategy Network. She is broadly interested in the social determinants and consequences of population health and health disparities, with a particular focus on race/ethnicity, poverty, race-related stress, and the family.

**Nina Cesare** is a graduate student in the Department of Sociology the University of Washington. Her work focuses on using digital data for social science research and exploring the dynamics of online social life.

**Ali Shojaie**, PhD, is an assistant professor of Biostatistics at the University of Washington. He is also an adjunct faculty at the Department of Statistics and an affiliate member of the University of eSciences Institute and the Center for Statistics and Social Sciences. Dr. Shojaie's research focuses on developing statistical machine learning methods for analyzing diverse high dimensional data from biomedical and social sciences, as well as statistical methods for large social and biological networks.

**Emma S. Spiro**, PhD, is an assistant professor at the Information School, University of Washington. She is also an adjunct assistant professor in the Department of Sociology, and an affiliate of the Center for Statistics and the Social Sciences at UW. She studies online communication and information-related behaviors in the context of emergencies and disaster events. Her work also explores the structure and dynamics of interpersonal and organizational networks in both online and off-line environments.