
#Hashtagging hate: Using Twitter to track racism online

by Irfan Chaudhry

Abstract

This paper considers three different projects that have used Twitter to track racist language: 1) Racist Tweets in Canada (the author's original work); 2) Anti-social media (a 2014 study by U.K. think tank DEMOS); and, 3) The Geography of Hate Map (created by researchers at Humboldt University) in order to showcase the ability to track racism online using Twitter. As each of these projects collected racist language on Twitter using very different methods, a discussion of each data collection method used as well as the strengths and challenges of each method is provided. More importantly, however, this paper highlights why Twitter is an important data collection tool for researchers interested in studying race and racism.

Contents

[Introduction](#)

[Understanding Twitter](#)

[Collecting data using Twitter](#)

[Case one: Racist tweets in Canada](#)

[Case two: "Anti-social media"](#)

[Case three: The Geography of Hate Map](#)

[Using Twitter to track racism online](#)

[Twitter and social research — Future considerations](#)

Introduction

Under our current social context, discussing issues related to race are often very difficult and perceived as impolite. As Malinda Smith notes, "there is a belief that to talk openly about race matters is an affront to good manners." [1]. As a result, there is

strong sentiment from people to feel that race (and consequently racism) is a thing of the past. While it is important to acknowledge this may be a byproduct of living in a multicultural and pluralistic society such as Canada, not being able to talk openly about issues related to race makes it difficult for Canada to become a place that is diverse and inclusive of all people, as both overt and covert forms of racism are able to persist. Although overt forms of racism in a public setting are less frequent than in the past (for the most part), one can shift focus to the online world, where overt forms of racism are rampant on social media sites, such as Twitter. A recent report released by Demos (a U.K.-based think tank), for example, found that on average, there are roughly 10,000 uses (per day) of racist and ethnic slurs in English being used on Twitter (Bartlett, *et al.*, 2014).

While this appears to be a high number, it is important to note there are no comparative figures which this finding can be contrasted with. For example, is this figure any higher or lower than what one might find on sites such as Facebook or Instagram? Although we currently do not have the information to make this comparison, it is important to remember that “new modes of communication mean it is easier than ever to find and capture this type of language” [2].

In light of new communication technology, social media sites like Twitter allow us to view and track racist language like we have never been able to do before. In recent years, racist graffiti sprawled on the sides of businesses or homes would have been the most overt text-based form of seeing racist language in a public area, however, with the rise and growth of communication technology (and social media specifically), the online realm has turned into a space where racist language is used openly. As Manuel Castells points out, “the fundamental change in the realm of communication has been the rise of self communication — the use of Internet and wireless networks as platforms of digital communication” [3]. The rise of digital communication tools (like Twitter) has given anyone with something to say a ‘digital soapbox’, where they can tweet their thoughts, values, and opinions on a variety of issues. While most Twitter users will tweet about news stories (Tao, *et al.*, 2014), some users may take to Twitter to espouse hateful sentiment. The older Twitter gets, the more its service (like the rest of the Web) becomes a vehicle for trolls [4] to challenge the social contract in a way that they might not be able to on the street (Greenhouse, 2013).

Although the use of racist language online is not a new phenomenon (see Foxman and Wolf, 2013), what is new is the ability for users to strategically track and monitor racism online. Due to Twitter’s “free speech” ideal (Greenhouse, 2013), it does not filter out terms or threads that are racist in nature. As a result, users can easily track and monitor racist language. The ability to track racist language on Twitter provides researchers interested in examining race and racism with a unique way to collect research data. While there are a number of paid services that can provide Twitter data

for users (such as Gnip or Datasift), these services are often costly, focus on large sets of data, and require added expertise with different data formats for users to utilize. As a result, researchers have been hesitant to utilize Twitter as a data gathering source. Due to the structure of Twitter, however, users can still collect data in an efficient and strategic manner, without the need to rely on costly data providers or learning a new data format.

In this paper, I will consider three different projects that have used Twitter to track racist language: 1) Racist Tweets in Canada (the author's original work); 2) Anti-social media (a 2014 study by DEMOS); and, 3) The Geography of Hate Map (created by researchers at Humboldt University) in order to showcase the ability to track racism online using Twitter. As each of these projects collected racist language on Twitter using very different methods, I will provide a discussion of each data collection method used as well as the strengths and challenges of each method. More importantly, however, I will highlight why Twitter is an important data collection tool for researchers interested in studying race and racism. Before discussing these projects, however, I will provide a brief genealogy of Twitter and how it is transforming from a social media platform to a useful space for researchers.



Understanding Twitter

Twitter is one of the largest and most popular social media Web sites (Murthy, 2013). Founded in 2006, Twitter brought together two subcultures: new media coding culture and radio scanner/dispatch enthusiasm (Rogers, 2014), forming what some refer to as first generation Twitter (or Twitter I), an urban lifestyle tool for friends to provide each other with updates on their activities and whereabouts (Akcora and Dembris, quoted in Rogers, 2014). As Jack Dorsey (the founder of Twitter) explains, Twitter was conceived as part of a long line of squawk media, dispatch, short messaging, as well as citizen communications services (Rogers, 2014) to be used as a public instant messaging system. Twitter allows users to “maintain a public Web-based asynchronous conversation through the use of 140 characters” [5]. Twitter's aim is to have users answer the question of “what is happening” through tweets which are posted and publically accessible via the users Twitter profile. As Murthy explains, tweets are a public version of the Facebook status update function (Murthy, 2013). One key difference, however, is that the tweets are viewable to all users of Twitter (unless your account is set to private), rather than being limited to your friends list on Facebook.

Due to their vast reach, tweets have the capacity to connect with a larger audience (either real or imagined), as the platform does not restrict content to one's friends list. Rather than "friending" someone (as you would on Facebook), Twitter users "follow" people. Usually, people will follow users who share a similar interest with them on a variety of topics (such as sports, fashion, or current events). As a result, Twitter has the capacity to create virtual communities where people collectively follow each other out of mutual interest (without ever having to have met the person off-line). This dynamic is what makes Twitter quite unique. As Dorsey explains,

"On Twitter, you are not watching the person, you are watching what they produce. It is not a social network, so there is no real social pressure inherent in having to call them 'friend' or having to call them a relative, because you are not dealing with them personally, you are dealing with what they are putting out there" (Dorsey, quoted in Sarno 2009a).

In light of this dynamic, early Twitter researchers deemed Twitter to be more of a news media site, rather than a social network (Rogers, 2014) since users broadcasted messages to followers (rather than writing on each other's walls, 'poking' them, or commenting on recently posted vacation pictures as some may do on Facebook or Instagram). By broadcasting or sharing personal information with many others, Twitter users can reinforce connections within the community or network (Humphreys, *et al.*, 2013).

Although tweets are limited to 140 characters, Twitter has "simple yet powerful methods of connecting tweets to larger themes, specific people, and groups" [6] through the use of hashtags — any word or phrase preceded by a hash sign "#" (Murthy, 2013). As Murthy explains, hashtags are an integral part of Twitter's ability to link conversations and strangers together (Murthy, 2013). For example, during the time of writing, the 2014 Winter Olympics Games in Sochi, Russia had a number of hashtags being used in tweets as a way to collect conversations related to the Olympic games. The official hashtag was #Sochi2014, used by many athletes, news agencies, and spectators to share their experiences via Twitter and also allow users (who are not present at the games) to receive up-to-date information.

While official hashtags are commonly created by event organizers as a way to track user activity as well as protect the brand image of a product, company, or service, unofficial hashtags also emerge, which sometimes receive more attention than the official hashtag itself. For example, even though #Sochi2014 was the official used during the Olympic games, in the days before the games had started, the hashtag #SochiProblems emerged and was being used by athletes and journalists to convey stories related to the "horrific" accommodation conditions in some hotels in Sochi,

Russia. By including a hashtag in one's tweet, the message becomes included into a larger conversation consisting of all tweets with that hashtag (Murthy, 2013), facilitating impromptu interactions between Twitter users (who are often strangers) into conversation. Due to its structure, rather than social circles, Twitter users have audiences (Rogers, 2014). As Murthy explains:

“Because hashtags represent an aggregation of tagged tweets, conversations are created more organically. Just because people are tweeting under the same hashtag, this does not mean they are conversing with each other in the traditional sense. Rather, the discourse is not structured around directed communication between identified interactants. It is more of a stream which is composed of a polyphony of voices all chiming in. Either serendipitously or by reading through scores of tweets appearing second by second, individuals and groups interact with each other after seeing relevant tweets” [7].

In 2009, Twitter's tagline changed from “What are you doing?” to “What's happening?” (Rogers, 2014). This simple shift was a significant move for the social media platform and was a strategic effort to further differentiate themselves from Facebook (whose initial popularity stemmed from the user's ability to provide a status update on what the user was doing). Internally, Twitter's tagline shift was a nudge for both users and researchers to consider information sharing through tweets (Rogers, 2014). As Biz Stone (another co-founder of Twitter) discussed shortly after this tagline shift, “Twitter is a state of affairs machine, or a discovery engine for finding out what is happening right now” (Stone, 2009, quoted in Rogers, 2014). In light of this shift, the second generation of Twitter emerged (Twitter II) as a source for news gathering and event-following. As Dorsey explains, Twitter usage always appeared to do well during natural disasters, man-made disasters, events, conferences, presidential elections, or as he calls them, “massively shared experiences” (Sarno, 2009b).

As a result, Twitter has become a powerhouse in sharing news information (both locally and globally). Users are able to search for key words or hashtags related to an event, moment, or experience, allowing them to feel like they are experiencing the event in real time. As Murthy explains, “part of Twitter's seductive power is the perceived ability of users to be important contributors to an event” [8]. These contributions, however, provide rich data for marketers as well as researchers to consume because it allows those interested in a certain topic to track, capture and analyze Twitter users' responses and activities, leading to what some are calling Twitter III — Twitter as a data set which requires both contractual access as well as technical infrastructure to take tweets, store them, and analyze them (Rogers, 2014).



Collecting data using Twitter

Although Twitter III is still in its infancy, it is clear that the company has a strong desire to make Twitter more appealing for researchers [9]. Initially, Twitter emerged as a way for users to share their interests and update their followers on what was occurring “right now”. Now, its power to capture data has gained the attention of researchers from a variety of disciplines to help answer a variety of questions, ranging from “simple information about a particular user or events (How many followers does a given user have?) to complex queries (How does information propagate among groups of users?)” [10]. Twitter’s power as a space for collecting social data was perhaps best seen during the Arab Spring uprisings, where numerous researchers took to Twitter to study how social media impacted the revolutions in Egypt, Libya, and other Arab Spring countries (Lotan, *et al.*, 2011). As a result, other researchers are now turning to Twitter to understand public sentiment in order to generate further social analysis on various topics. Depending on the aim of the study, different tools can be used to collect data — from Web-based analytical services to directly mining the Twitter API (application programming interface) (Gaffney and Puschman, 2014) and interpreting the data manually (or through different dedicated statistics packages such as QDA Miner or WordStat).

One of the ways Twitter allows users to search for data is through a streaming API. The streaming API is likely the most used data source for Twitter research as large-scale quantitative data are based on raw data collected through this source (Gaffney and Puschman, 2014). The streaming API is a unique way of gathering data as it is “push” based, meaning “data is constantly flowing from the requested URL (the end point) and it is up to the researcher to develop or employ tools that maintain a persistent connection to this stream of data while simultaneously processing it” [11]. Although it is a useful way to collect data on Twitter, one of the main obstacles is that the researcher must capture the data in “real time”, since stream data is provided as a live poll, meaning that it becomes available the moment a tweet is tweeted (Gaffney and Puschman, 2014). As a result, this form of capturing data works quite well in trying to gather information on an event that is occurring “right now”, but does not do a good job in collecting historical data.

Another obstacle in collecting streaming data (either in real time or historical) using Twitter is due to the structure of Twitter itself. The streaming API permits access to data in three bandwidths: “spritzer”, “gardenhose”, and “firehose” which delivers one percent, ten percent, and 100 percent of all tweets posted on the system, respectively

(Gaffney and Puschman, 2014). By default, regular accounts on Twitter have access to sprtizer data, which for the most part, will suffice for research. The gardenhose is granted occasionally to users with definable and compelling reasons for increased access, while the firehose is only available as a component of a business relationship with Twitter directly (Gaffney and Puschman, 2014). Users can contract third-party service providers such as Gnip or Datasift in order to gain more access to Twitter data (especially historical data), however, each service is quite costly and only deals with large sets of data.

Researchers who wish to use Twitter for small to medium sized projects may be better off trying to gather the data themselves. A current project of mine (www.twiterracism.com), for example, would certainly fit into the category of a small project using Twitter as the main site of data gathering and is a good starting point to examine the first way users can collect data on racist language using Twitter.



Case one: Racist tweets in Canada

Between June, July, and August of 2013, I collected data on Twitter using the streaming API method discussed above via Hootsuite, a social media management platform, to track the amount of times certain racist words were used on Twitter originating from users located in certain Canadian cities [12]. These cities include Edmonton, Calgary, Vancouver, Toronto, Montreal, and Winnipeg. They were chosen for two main reasons: 1) these cities are major metropolitan areas in Canada; and, 2) these cities experienced some of the highest reported race-based hate crimes in 2010 [13]. Initially, I was interested in understanding how racist Canadians are on Twitter and how the online level of racial intolerance connects to reported incidents of racism in the off-line world (if at all), however, as the project progressed, it became clear to me from colleagues that a quantitative approach for this study was not feasible. As a result, my research focus shifted qualitatively towards understanding the context these racist terms were being used on Twitter in these six Canadian cities.

I used Hootsuite to collect data since the platform allows users to search for tweets by both keywords and geographic location (something which cannot be done using the basic Twitter search API). Locations provide an ideal access point for researchers interested in geographically bounded research, however, it is still fairly limited since only about one percent of all traffic on Twitter is “geo-tagged” (Gaffney and Puschman, 2014), meaning only a small number of users are opting in to have their geographical location shown with their tweet. This data, however, carries immense

commercial value (Wilken, 2014) and it is likely that the proportion of geo-tagged tweets will increase in the future (Gaffney and Puschman, 2014).

Despite the small number of geo-tagged tweets, researchers can still gather unique data related to their topic of interest. For this project, I inputted each cities geographical coordinates (which I obtained from Google) in hootsuite, looking for any tweets containing the following racist words in a negative context: native(s); white trash; nigger(s); paki(s); and chink(s). These words were chosen because they represent the most common racist terms associated with specific racialized groups. The term “native”, which is not usually associated with being a racist word (due to its use in everyday English language), was surprisingly used in a negative way on twitter in the aforementioned Canadian cities, and as such, was included as one of the key terms for this data set.

In order to capture this data using Hootsuite, I created a search query to reflect the key word being searched originating from the cities latitude and longitude coordinates. For example, when searching for the keyword *paki* originating from the city of *Edmonton*, I typed the following into the Hootsuite search bar: *paki geocode:53.5381965637,-113.5029678345,100km*. Essentially, what this syntax is looking for is the word (*paki*) originating from the coordinates of 53.3 and -113.5 (which is Edmonton’s geo-location) and the surrounding 100 kilometers. I used the same search query for Calgary, Vancouver, Winnipeg, Montreal, and Toronto, and over the span of three months, each tweet that came up containing one of the above racist words from these cities was read by the researcher to determine the context of the tweet. For the purpose of this project, only tweets which contained the above words in a negative context were captured in the data and saved in an MS Excel spreadsheet for further content analysis. The total number of tweets collected over the three month period was a nominal 776. A few examples of tweets captured include the following:

“Your not much diff from natives when it comes to drinking ...
Except your clean”.

“Dad cuts up Water Melon for my lunch, looks at me and says
‘Here’s your nigger food’.” #Hahaha #LoveYou

“This group of chinks on the bus f***ing wreak I’m gagging and
covering my face in my shirt”.

“I think I hate cardio as much as I hate paki’s”.

This data offers some initial results that generate a sense of a) where the most racist tweets come from and b) what racialized groups are most commented on in a negative way on Twitter in the six Canadian cities (see [Table 1](#)).

TORONTO		VANCOUVER		WINNIPEG	
Nigger(s)	190	Nigger(s)	29	Natives	33
Paki(s)	100	White Trash	23	Paki(s)	20
White Trash	88	Natives	17	Nigger(s)	10
Chink(s)	40	Paki(s)	15	Chink(s)	8
Natives	16	Chink(s)	15	White Trash	7
Totals	434	Totals	99	Totals	78

CALGARY		EDMONTON		MONTREAL	
Natives	30	Natives	27	Nigger(s)	3
White Trash	11	Nigger(s)	11	White Trash	3
Nigger(s)	10	Paki(s)	11	Chink(s)	3
Paki(s)	9	Chink(s)	6	Paki(s)	3
Chink(s)	2	White Trash	5	Natives	3
Totals	62	Totals	60	Totals	4

Table 1: Racist tweets in six Canadian cities.

Although this data is very tentative, it is interesting to see that the cities with the highest number of Aboriginal residents (Winnipeg, Edmonton, and Calgary) have the term natives used most often in a negative context in the tweets gathered for this study. Toronto and Montreal — two cities with the highest black population in the country — also have the term nigger(s) used most often in a negative context in the tweets gathered for this study. As the number of tweets collected for this study is quite small (776), the tweets are hardly reflective of overall Twitter activity, and work better as a case study of a small sample of racist tweets in Canada.

After conducting a content analysis of the tweets, interesting themes emerged in relation to how racist language is being used on Twitter. Following similar usage categories as the DEMOS study (discussed below), the following seven categories highlight how racist language is being used on Twitter in these six Canadian cities: 50 percent “Real time” response; 28 percent Negative Stereotype; 12 percent Casual use of slur; four percent Responses to racism; two percent Targeted abuse (online), Appropriated, and “Non-Derogatory” [14].

One of the most striking findings from the content analysis is the amount of times tweets are used as a “real time” response, meaning that the person is tweeting about something as it is happening to them at the current moment. Consider the following set of tweets as examples of this category:

“Man these pakis need to leave the gym it’s starting to smell like stale Burger King in here”.

“Drunk natives brushing his pony tail on me #help #gross”

“When my friends make me sit next to the nigger in the theater ... #thanksbitches”

The real-time response tweets are interesting to consider because they reveal thoughts and opinions that most people would want to remain private, and expose a level of racial intolerance that we are not used to publically seeing in Canada. This finding showcases the type of rich data researchers can collect using Twitter and capture information which would not have been available before.

Strengths and limitations of tracking racist language using this method

By using this method, I was able to develop a strategy to capture streaming data and archive it myself for later analysis. It also provided me with interesting data which I would not have been able to see had I not used Twitter as the site for data collection, particularly after conducting a content analysis of the tweets to reveal patterns of usage. This process, however, is somewhat time consuming as I would login to Hootsuite once every few days (since Hootsuite only shows historical data up to five days before it does not show up in the search perimeter), in order to read and sift through the tweets. Reading the tweets to determine context is also fairly time-consuming, as often I would have to further delve into certain tweets in order to gather the correct context of the tweet. (This is also a highly subjective process, as something I would classify as negative, another user may not and vice versa).

The biggest limitation of this research method, however, is the sample size itself. By initially only including negative tweets, I lost an opportunity to contextualize the tweets. For example, had I collected all tweets which included the racist words, and later on given a breakdown of how many of the overall tweets were used in a negative way, my data analysis would have been a lot richer, and I would have had a better sense of the amount of times racist language was being used on Twitter in these six Canadian cities. As a result, the current project only reflects a very small sample of racist tweets in Canada.

This method would work well for small-scale projects, but would not be useful for trying to collect and manage large sets of Twitter data, unless there was more than one researcher. Using geo-coded Twitter information is also fairly limited, as you are only getting a very small number (one percent) of Tweets with geo-coded location information, again limiting your sample size. However, since most basic Twitter searches provide users with spritzer data access (one percent), it is unclear as to how much information one may be missing [15].



Case two: "Anti-social media"

While utilizing geo-coded search perimeters might limit your sample size, a 2014 study by DEMOS provides another way users can track racist language online, resulting in a higher amount of racist tweets. The study, entitled "Anti-social media", aimed to find in what ways [racial] slurs are being used on Twitter and in what volume (Bartlett, *et al.*, 2014). To collect the data, similar to the previous study, researchers scraped the publically available live Twitter feed (the streaming API) containing one or more candidate slurs over a nine-day period in November (19 November to 27 November 2012). The list of terms judged candidate slurs was crowd sourced from Wikipedia (Bartlett, *et al.*, 2014). The ten most common found in the data set (in order of prevalence) were: *white boy*, *paki*, *whitey*, *pikey*, *nigga*, *spic*, *crow*, *squinty*, and *wigga*.

In order to collect the data, the tweets were filtered to ensure that the slurs were contained in the body of the tweet and were not part of a user's account name (Bartlett, *et al.*, 2014), something the previous study also took into account. The tweets were also passed through an English language filter to exclude non-English tweets. In total, 126, 975 tweets were collected: an average of 14,100 tweets per day (Bartlett, *et al.*, 2014). As the authors of the study note, "all of the tweets in our samples were publically available to any user of Twitter" [16]. Due to the high

volume of tweets gathered for this study, automated machine analysis was utilized to help sift through the data. This study used a software platform called Agile Analysis Framework (AAF), which is designed to help the researcher isolate tweets of interest and then identify and quantify the different ways in which language is used in those tweets (Bartlett, *et al.*, 2014). As the study authors explain, "AAF allows the researcher to construct standard and bespoke filters (or classifiers) which automatically places tweets into certain (human defined) categories, allowing the researcher to iteratively sort very large data sets into separate categories for further study" [17].

In addition to the automated analysis, researchers also used manual human analysis, where analysts would validate the automated system and also do in-depth, iterative analysis of both small and large random samples of the data to reveal a stable set of categories (Bartlett, *et al.*, 2014). Through the combined effort of both automatic and manual analysis methods, the study found the following as the most common used racial slurs during the nine-day study period: white boy (48.9 percent); Paki (11.7 percent); Whitey (7.9 percent); Pikey (4.1 percent); Coon (3.2 percent); nigga (3.2 percent); Spic (3.0 percent); Crow (2.1 percent); squinty (1.8 percent); and, Wigga (1.7 percent). Some examples of tweets found include:

"fUcK yOu In ThE aSs PuNk wHiTe BoY"

"Fucking paki shops that charge you for cash withdrawals or paying by card need nuked"

"Dear waiter: about the dirty rag on your head to hide your dreds: YOU ARE WHITE. Dreds on a white boy just makes you look homeless"

On the basis of their analysis, the DEMOS study identified eight categories of patterns of usage: Negative stereotype; Causal use of slurs, Targeted abuse; Appropriated; Non-derogatory; Off-line action; Impossible to determine; and error [18]. What the study found overall was that racist slurs are used in a variety of ways on Twitter (both offensive and non-offensive), and that there were very few cases that presented an imminent threat of violence (Bartlett, *et al.*, 2014).

Strengths and limitations of tracking racist language using this method

The DEMOS study is a useful template to consider, mainly due to the way that content analysis is used to develop the racist tweets categories. Careful analysis of each tweet reveals how significant context is in determining meaning and intent (Bartlett, *et al.*, 2014). Additionally, since both automated and manual analysis

techniques were used, the research team was able to capture and analyze a larger data set ($n=126,975$). As the researchers suggest, “machine classifiers were extremely useful to identify and filter data sets into more manageable data sets” [19].

Furthermore, by having a larger data set, qualitative analysis was useful to determine specific categories (Bartlett, *et al.*, 2014) and “allowed the analysts to find more nuanced categories, such as appropriated use” [20].

As with the previously mentioned Twitter study, the DEMOS study highlights the use of social media as a data collection tool to understand trends and changes in language use. At the same time, however, because the use of language is heavily dependent upon context of a situation, studies using Twitter to generate data need to be aware of this constraint and work around it. For example, if we consider the terms used for the DEMOS study, certain terms may only be considered offensive in the context of Great Britain (such as “Crow” or “squinty”) and would not really translate well to other countries (such as Canada). As the research team from DEMOS explain,

“Twitter sampling works on the basis of key word matches. This type of analysis automatically creates systemic bias into the research method. Our use of a crowd sourced word cloud was a simple way around this problem, but there are no doubt several other, fast changing, terms that we may have missed. It is certainly the case that automated key word matches are of limited power in respect of finding genuine cases of serious ethnic slurs or hate speech. Each case is highly contextual; and often will depend on approximations about the individuals involved” [21].

In this regard, language classifiers are useful tools to filter and manage large data sets, however, also need to be considered within the contextual nature of the tweet itself. This calls for the researcher to not only consider the initial tweet but also understand its context, content, and construct. What this means is the tweet cannot be taken only at face value. It must be considered in relation to the characteristic of the Twitter user (*i.e.*, is it offensive if a black Twitter user tweets something which contains the term ‘nigger’?), the context of the tweet itself (*i.e.*, is it a conversation between friends?) and, among other things, the purpose of the tweet (*i.e.*, is it someone using a term in a very casual way). Doing so allows for greater analysis of the data but also requires greater time from the researcher, as manual analysis (namely reading each tweet and understanding its sentiment) is something that only a human researcher can do.



Case three: The Geography of Hate Map

Similar to the Racist Tweets in Canada project, another project that has relied on using geo-coded tweets to track racist (as well as homophobic and ableist) language on Twitter is the *Geography of Hate Map* [22] (*GHM*). According to the researchers, “the prominence of debates around online bullying and the censorship of hate speech prompted us to examine how social media has become an important conduit for hate speech and how particular terminology used to degrade a given minority group is expressed geographically” [23]. Inspired by a map created by researchers at Humboldt University in 2012 (which looked at racist responses to the re-election of U.S. President Barack Obama), the *GHM* “analyzed a broader swath of discriminatory speech in social media, including the usage of racist, homophobic and ableist slurs” [24] using DOLLY [25] (Digital OnLine Life and You currently only available to those affiliated with the University of Kentucky) to search for all geo-tagged tweets in North America between June 2012 and April 2013. The DOLLY Project is a repository of billions of geo-located tweets that allows for real-time research and analysis. Building on top of existing open source technology, the research team created a backend that ingests all geo-tagged tweets (~eight million a day), does basic analysis, indexing and geo-coding to allow real-time search throughout the entire database (three billion tweets since Dec 2011). Similar to the Canadian study, researchers from this study read each tweet individually and only included tweets which contained the following key hate words in a negative context: Nigger, Gook, Chink, Wetback, and Spick.

Overall, researchers found over 150,000 geo-tagged tweets with a hateful slur to be negative (Stephens, 2013). This number, however, includes racist, homophobic, and ableist tweets [26]. Although an exact figure of how many times each word was used in a negative context is not available on the *GHM* Web site, viewing the online “heat map”, it appears that the most common racist terms used (in order of prevalence) are: Nigger, Chink, Wetback, Spick, and Gook. To determine the context of the tweet, research assistants would manually read through each tweet and classify them as positive, neutral, or negative.

Similar to the Canadian study, certain areas showed higher negative tweets related to a specific racialized group, reflecting demographic trends in each respective area. For example, when looking at the map, the term “wetback” (which is often used as a derogatory way to refer to Mexican immigrants) was used most often in different areas close to Texas “showing the state’s centrality to debates about immigration in the US” [27]. Researchers caution, however, that “many of the slurs included in our analysis display little meaningful spatial distribution. For example, tweets referencing ‘nigger’ are not concentrated in any single place or region in the United States; there are a number of pockets of concentration that demonstrate heavy usage of the word” [28].

Strengths and limitations of tracking racist language using this method

What is interesting about this collection of racist data (that differs from the other two methods above) is that researchers aggregated and normalized the data to reflect Twitter use based on county. As lead researcher Dr. Monica Stephens explains,

“Hateful tweets were aggregated to the county level and then normalized by the total number of tweets in each county. This then shows a comparison of places with disproportionately high amounts of a particular hate word relative to all tweeting activity. For example, Orange County, California has the highest absolute number of tweets mentioning many of the slurs, but because of its significant overall Twitter activity, such hateful tweets are less prominent and therefore do not appear as prominently on our map.”

By aggregating the data, the researchers were able to also map the geo-tagged tweets to create a heat map (See [Figure 1](#)). To produce the map all tweets containing each ‘hate word’ were aggregated to the county level and normalized by the total Twitter traffic in each county (Stephens, 2013). Counties were reduced to their centroids and assigned a weight derived from this normalization process. This was used to generate a heat map that demonstrates the variability in the frequency of hateful tweets relative to all tweets over space (Stephens, 2013). Where there is a larger proportion of negative tweets referencing a particular ‘hate word’ the region appears red on the map, where the proportion is moderate, the word was used less (although still more than the national average) and appears a pale blue on the map. Areas without shading indicate places that have a lower proportion of negative tweets relative to the national average (Stephens, 2013).

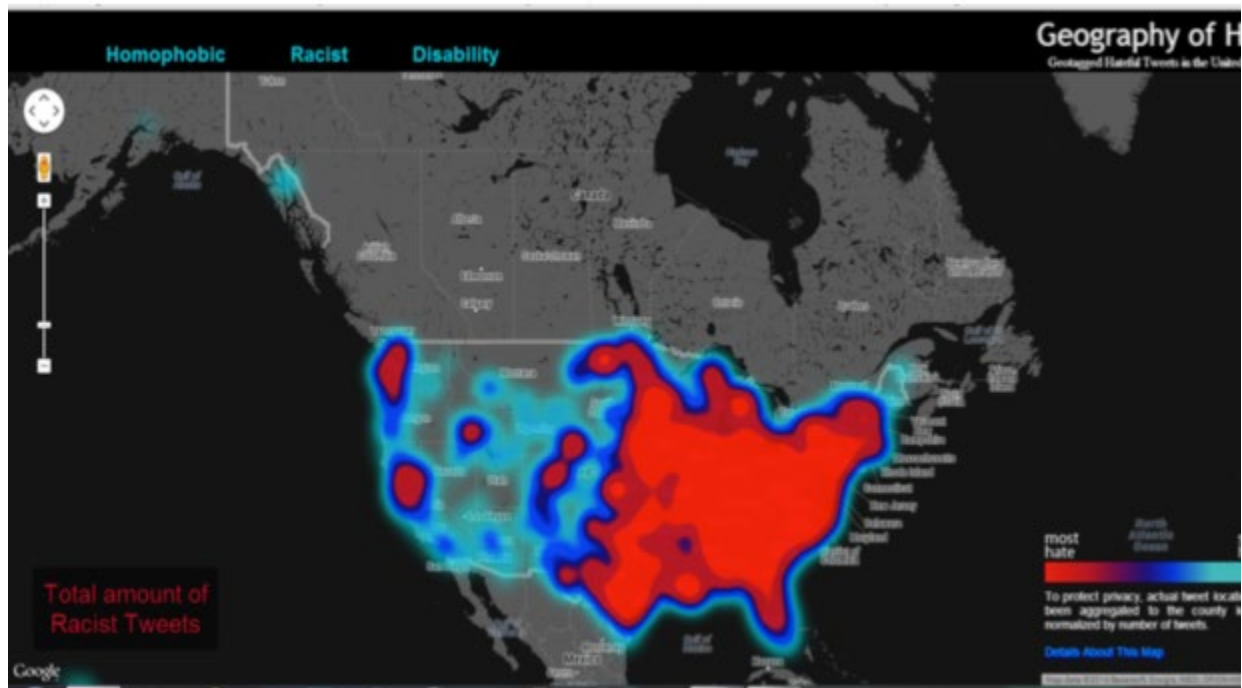


Figure 1: Geography of Hate Map.

Source: http://users.humboldt.edu/mstephens/hate/hate_map.html#.

While this method is highly technical and is proprietary to users associated to Humboldt University, it highlights another important factor to consider when using Twitter to collect data, namely, how prevalent certain language usage is within the context of overall Twitter activity (something which the previous two studies do not do). By considering the amount of times racist terms are used in relation to overall Twitter activity during a given time frame, researchers can develop an accurate context to consider tweets. For example, are racist terms used any more or less than other terms on Twitter? Unless there is some form of contextual reference, this question is impossible to answer. What this and the other two studies do show, however, is that open social media platforms, such as Twitter, allows some users to express racist sentiment, highlighting the fact that hateful messages are still propagated in society.



Using Twitter to track racism online

As noted earlier, “new modes of communication mean it is easier than ever to find and capture this type of language” [29]. The transition of Twitter from a social networking site to a space where researchers can track and collect data from highlights the power of this new form of communication technology. Each of the three projects highlighted in this paper captured racist language on Twitter in very different ways, however, each one revealed that social media is a space where individuals can convey hateful messages that they would not normally “off-line”. What these data collection methods allow us to do is track information on racism in a way which we have never been able to do before. In Canada, for example, unless complaints about racism reach various human rights tribunals in the country, there is no real mechanism of tracking racism. As a result, using Twitter as a data collection tool allows us to generate data which we would not have been able to do otherwise.

Critics might argue that using Twitter to track racism does not represent overall racist activity within society — a fair criticism. However, what Twitter does allow for is access to information that can assist our understanding of how different forms of communication within society allows users to express racist sentiment.

One can potentially develop a research instrument to either survey or interview people about instances of racism, however, these methods are costly, time-consuming, and perhaps do not reveal a deeper level of racial bias (as interview subjects may not want to appear as racist to the interviewer). The city of Edmonton, for example, contracted the Population Research Lab (PRL) at the University of Alberta in 2011 to conduct 400 telephone surveys with Edmontonians’ regarding experiences and perceptions about discrimination (including racism) in the city [30]. While some of the survey findings showcased what Edmontonians think about discrimination in the City, the survey did not uncover any deeper sentiments or particular instances of racial discrimination in the city.

Using Twitter to track racism, however, allows us to develop a sense of the type of racist terms most commonly used and more importantly, in what context they are being used. The Racist Tweets in Canada project as well as the DEMOS study both showcase the importance of conducting a content analysis on tweets in order to generate context around how language is being used online. While the findings from both studies are snapshots of overall Twitter activity, the results serve as a reminder that racism still overtly manifests itself in some realms within society. It is also important, however, to further contextualize how racist Twitter activity fits in to overall Twitter activity (as showcased by the *Geography of Hate Map*).

Utilizing Twitter as a data collection tool is something researchers who are interested in studying race and racism must pay attention to. The access to discussions about race and racism on Twitter is extremely rich and one just has to delve into the platform to see the ways Twitter users openly discuss race and racism — something which is not often seen in the “off-line” world [31].

While it might appear to be impolite to discuss matters of race “off-line” in society (race manners, as Malinda Smith notes), the politeness disappears online. For researchers interested in studying the topic, it is precisely these moments of impoliteness that reveal a level of racial bias which is integral in assisting our understanding of race and racism in society.



Twitter and social research — Future considerations

These case studies highlight how researchers can use Twitter to track and capture data related to racism online. While there are a number of far more complex methods researchers can use to capture data on Twitter, the methods mentioned earlier are fairly straightforward and require a basic understanding of how Twitter works. Additionally, these methods only consider collecting data using the streaming API, which as mentioned above, only yields about one percent of all tweets. For small projects (such as the one tracking racist tweets in Canada), this access to data is sufficient. However, as Twitter III starts to take shape and broadens the access to Twitter data beyond spritzer access, projects such as the ones mentioned in this paper can receive further support and capture larger data sets to assist in understanding the intersection between Twitter, race and society. Twitter resembles much more a news media platform than a social network, making it a significant source of information (Tao, *et al.*, 2014).

As Dhiraj Murthy highlights,

“Twitter shares similarities with blogs, albeit the posts are much shorter. However, once one’s tweets are aggregated, a new structure emerges. This is not merely a technical consideration, but rather the organization of communication as a series of short communiqués is qualitatively different from examining tweets individually. As a corpus, they begin to resemble a more coherent text. Granted the corpus is disjointed, but narratives can and do emerge.” [32]

Herein lays the power of Twitter as a source for social research. The potential and capacity is there for researchers to sift through open data sources that provide interesting context and insight on a variety of social issues. As Twitter becomes more engrained as a tool for data collection, it will be up to the researcher to develop strategies to capture the data and further connect the disjointed narratives that emerge on Twitter, one tweet at a time.

About the author

Irfan Chaudhry is currently a sessional instructor at MacEwan University, Department of Sociology and a Ph.D. candidate (provisional) with the Department of Sociology, University of Alberta. He received an M.A. in criminal justice at the University of Alberta (Department of Sociology).
E-mail: irfan [at] ualberta [dot] ca

Notes

- [1.](#) Smith, 2003, p. 122.
- [2.](#) Bartlett, *et al.*, 2014, p. 5.
- [3.](#) Castells 2012, p. 6.
- [4.](#) According to Kelly Bergstrom, an internet troll is an identity game where the troll attempts to pass as a legitimate participant, sharing the group's common interests and concerns, but, however, is really there to disrupt discussion and evoke reaction from users; Kelly Bergstrom, 2011. "'Don't feed the troll' Shutting down debate about community expectations on Reddit.com," *First Monday*, volume 16, number 8, at <http://firstmonday.org/article/view/3498/3029>.
- [5.](#) Murthy, 2013, p. 2.
- [6.](#) Murthy, 2013, p. 3.
- [7.](#) Murthy, 2013, p. 4.
- [8.](#) Murthy, 2013, p. 33.

9. In February 2014, Twitter introduced a pilot project called Twitter Data Grants, through which they will give a handful of research institutions access to public and historical data. More information can be found here: <https://blog.twitter.com/2014/introducing-twitter-data-grants>.

10. Gaffney and Puschman, 2014, p. 55.

11. Gaffney and Puschman, 2014, p. 56.

12. The project Web site is located here: <http://www.twiterracism.com>.

13. <http://www.statcan.gc.ca/pub/85-002-x/2012001/article/11635/tbl/tbl05-eng.htm>.

14. Apart from the “real time response” category, the remainder of the categories used for the content analysis were inspired by the 2014 DEMOS study (discussed below). A full list of the definition of each category can be found here: http://www.demos.co.uk/files/DEMOS_Anti-social_Media.pdf?1391774638, p. 24.

15. See Morstatter, *et al.* (2013) for a detailed discussion comparing Twitter’s streaming API with Twitter’s firehose.

16. Bartlett, *et al.*, 2014, p. 6.

17. Bartlett, *et al.*, 2014, p. 15.

18. A full list of the definition of each category can be found at http://www.demos.co.uk/files/DEMOS_Anti-social_Media.pdf?1391774638, p. 24.

19. Bartlett, *et al.*, 2014, p. 8.

20. Bartlett, *et al.*, 2014, p. 9.

21. Bartlett, *et al.*, 2014, p. 10.

22. Available at <http://www.floatingsheep.org/2013/05/hatemap.html>.

23. Stephens, 2013, paragraph 2.

24. Stephens, 2013, paragraph 4.

25. <http://www.floatingsheep.org/p/dolly.html>.

[26.](#) An individual breakdown between racist, homophobic, and ableist tweets is not available, however, the researchers note that 41,306 tweets contained the word ‘nigger’.

[27.](#) Stephens, 2013, paragraph 8.

[28.](#) Stephens, 2013, paragraph 7.

[29.](#) Bartlett, *et al.*, 2014, p. 5.

[30.](#) A full copy of the report can be found at <http://www.racismfreedmonton.ca>.

[31.](#) One just has to look at a number of high profile incidents involving race and racism such as Donald Sterling, Trayvon Martin, Justine Sacco, Paula Deen, etc. where the “off-line” world was fairly silent, but the online world — Twitter specifically — was rampant with users expressing opinions about race and racism.

[32.](#) Murthy, 2013, p. 8.

References

J. Bartlett, J. Reffin, N. Rumball, and S. Williamson, 2014. “Anti-social media,” at http://www.demos.co.uk/files/DEMOS_Anti-social_Media.pdf?1391774638, accessed 4 June 2014.

K. Bergstrom, 2011. ““Don’t feed the troll’ Shutting down debate about community expectations on Reddit.com,” *First Monday*, volume 16, number 8, at <http://firstmonday.org/article/view/3498/3029>, accessed 29 January 2015.

M. Castells, 2012. *Networks of outrage and hope: Social movements in the Internet age*. Cambridge: Polity Press.

A.H. Foxman and C. Wolf, 2013. *Viral hate: Containing its spread on the Internet*. New York: Palgrave Macmillan.

D. Gaffney and C. Puschman, 2014. “Data collection on Twitter,” In: K. Weller, A. Bruns, J. Burgess, M. Mahrt, and C. Puschmann (editors). *Twitter and society*. New York: Peter Lang, pp. 55–68.

E. Greenhouse, 2013. "Twitter's free-speech problem," *New Yorker* (1 August), at <http://www.newyorker.com/tech/elements/twitters-free-speech-problem>, accessed 4 June 2014.

L. Humphreys, P. Gill, B. Krishnamurthy, and E. Newbury, 2013. "Historicizing new media: A content analysis of Twitter," *Journal of Communication*, volume 63, number 3, pp. 413–431.
doi: <http://dx.doi.org/10.1111/jcom.12030>, accessed 29 January 2015.

G. Lotan, E. Graeff, M. Ananny, D. Gaffney, I. Pearce, and d. boyd, 2011. "The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions," *International Journal of Communication*, volume 5, pp. 1,375–1,405, at <http://ijoc.org/index.php/ijoc/article/view/1246/613>, accessed 29 January 2015.

F. Morstatter, J. Pfeffer, H. Liu, and K.M. Carley, 2013. "Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose," *Seventh International AAAI Conference on Weblogs and Social Media*, <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6071>, accessed 29 January 2015.

D. Murthy, 2013. *Twitter: Social communication in the Twitter age*. Cambridge: Polity Press.

R. Rogers, 2014 "Debanalising Twitter: The transformation of an object of study," In: K. Weller, A. Bruns, J. Burgess, M. Mahrt, and C. Puschmann (editors). *Twitter and society*. New York: Peter Lang, pp. ix–xxvi.

D. Sarno, 2009a. "Twitter creator Jack Dorsey illuminates the site's founding document. Part 1," *Los Angeles Times* (18 February), at <http://latimesblogs.latimes.com/technology/2009/02/twitter-creator.html>, accessed 4 June 2014.

D. Sarno, 2009b. "Jack Dorsey on the Twitter ecosystem. Part 2," *Los Angeles Times* (19 February), at <http://latimesblogs.latimes.com/technology/2009/02/jack-dorsey-on.html>, accessed 4 June 2014.

M. Smith, 2003. "Race matters and race manners," In: J. Brodie and L. Trimble (editors). *Reinventing Canada: Politics of the 21st century*. Toronto: Prentice Hall, pp. 108–130.

M. Stephens, 2013. "The Geography of Hate Map," at http://users.humboldt.edu/mstephens/hate/hate_map.html#, accessed 4 June 2014.

K. Tao, C. Hauff, F. Abel, and G.J. Houben, 2014. "Information retrieval for Twitter data," In: K. Weller, A. Bruns, J. Burgess, M. Mahrt, and C. Puschmann (editors). *Twitter and society*. New York: Peter Lang, pp. 195–206.

R. Wilken, 2014. "Twitter and geographical location," In: K. Weller, A. Bruns, J. Burgess, M. Mahrt, and C. Puschmann (editors). *Twitter and society*. New York: Peter Lang, pp. 155–168.

Editorial history

Received 17 July 2014; accepted 24 January 2015.

Copyright © 2015, *First Monday*.
Copyright © 2015, Irfan Chaudhry.

#Hashtagging hate: Using Twitter to track racism online
by Irfan Chaudhry.

First Monday, Volume 20, Number 2 - 2 February 2015

<https://firstmonday.org/ojs/index.php/fm/article/download/5450/4207>

doi: <http://dx.doi.org/10.5210/fm.v20i2.5450>