Examining Demographic Trends in Political Opinion on Twitter: The Case of Gun Control

Authors: Nina Cesare[1], Tyler McCormick[1,2,3], Hedwig Lee[1,3], Ali Shojaie[4]

[1] Department of Sociology, University of Washington
[2] Department of Statistics, University of Washington
[3] Center for Studies in Demography and Ecology, University of Washington
[4] Department of Biostatistics, University of Washington

Abstract
This project seeks to leverage Twitter data to examine the relationship between individuals' political opinions and demographic characteristics. By applying semi-automated text analysis techniques and methods of extracting key demographic variables such as age, race, and sex to data gathered from Twitter, it will expand upon existing research that highlights the importance of Twitter as a tool for tracking the "pulse" of public opinion (Hubbard, 2011). Previous research exploring political opinion trends on Twitter has focused almost exclusively on the text of the tweets themselves. This project emphasizes that including demographic information of the Twitter users will help make Twitter a more effective policy making tool by providing a more nuanced view of political opinion trends and highlighting the role of marginalized demographic groups – such as African Americans, who are often underrepresented in survey research but compose a large portion of the Twitter universe - within broad political discourse.

Introduction

It is well known that in addition to its purely social influence, social media plays an important role in promoting civic engagement and facilitating political discourse (Bennett 2003, Benkler 2006; Conover et al 2011; Murthy 2007; Sunstein 2007; Tumasjan et al 2007). Twitter is viewed by many as a "microphone for the masses" (Murthy 2007), as its 500 million registered users provide updates on public reactions to noteworthy social political events in real-time. In addition to this, Twitter's unique network structure allows for direct communication between (political) institutions, political leaders and private individuals, which sets it apart as a tool for promoting active civic discourse (Parmalee and Bichard 2013). Finally, Twitter offers unique advantages to researchers as users' connections with one another and patterns of communication are well documented. This allows researchers to easily view associative and ideological splits within these networks following or during a particular politically-salient event (e.g., presidential election). Overall, research suggests that online spaces are not filled with "pointless babble," but are in fact a respected source for gathering substantive conversation on major social issues (boyd 2009).

Though a number of researchers have capitalized on the capacity of Twitter to examine trends in public opinion, these studies unanimously lack thorough consideration of the demographic trends underlying them. The addition of this dimension could greatly benefit researchers attempting to predict political outcomes or track potentially rapidly changing tides in political opinion or participation among particular demographic groups – particularly groups that are

generally underrepresented in the context of political engagement. Moreover, the addition of demographic information to projects that draw upon social media data for social science research could greatly expand the explanatory and predictive power of these analyses. Being able to track not only what is said in regards to a major social or political event, but what demographic characteristics are associated with which ideological stances or spheres of discourse will amplify the capacity of Twitter to provide insight into public opinion trends and render it a more useful tool for policy development and intervention.

Twitter as a tool for Political Analysis

Political discourse on Twitter is unique in that extends beyond citizen-to-citizen communication; political leaders and news agencies utilize Twitter as a resource for distributing information to and negotiating with private citizens as well (Parmalee and Bichard 2013). While other social media sites provide a forum for private citizens to exchange information, Twitter's open, undirected (or unreciprocated) network structure and use of "retweet" and "@" commands permits communication beyond each users' individual friendship network. Within this platform, direct conversation between those involved in policymaking and those who are affected by these policies is possible. In addition to this, Twitter's hashtags allow users to not only exchange information from person to person but to contribute to broad, open conversation centered on a specific issue or event.

Given Twitter's stand-out capacity in regards to its ability to facilitate political discourse and civic engagement, there are nonetheless barriers regarding the way in which Twitter users' opinions can be leveraged for change. Although those interested in mining Twitter for public opinion trends can use this tool to see what it being said given the content of individuals' tweets, it is difficult to tell *who* holds opinions of interest using Twitter's relatively sparse user profiles. This project, however, will employ methods for systematically extracting the demographic information of Twitter users as a means of understanding the composition of the groups behind the opinions represented.

By overcoming this challenge in gathering Twitter users' political opinions *and* demographic information, researchers will potentially be able to understand minority opinions in ways that are often not easy to evaluate in the context of standardized surveys. Researchers have noted that political activity – and to a large degree surveys of political behavior – fail to effectively capture the true distribution of public opinion (Verba 1995). Some have suggested that online spaces provide a more open and democratic platform that may "bring new individuals into the political process" (Kreuger 2002: 476), but there are nonetheless barriers associated with the use of social media data for social science research. Establishing a way to sample both opinions and key demographic variables through Twitter, however, may overcome this challenge. It may help researchers who use Twitter to understand shifts in popular opinion within short periods of time and may as well as better understand of the role of marginalized groups in the context of broad political debates. In turn, it may help both researchers and policymakers become more responsive to the needs and desires of specific demographic groups – particularly those that are not as well represented in the political arena – by allowing them to conduct a semi-ethnographic analysis of how particular issues impact these groups.

Application: Gun Control

Gun control is an ongoing and divisive issue within the American political sphere. A number of events have occurred within 2012-2013 – including (but not limited to) the Newtown shootings, Aurora shootings, Boston Marathon bombings, Manchin-Toomey bill failure, Chicago Park shootings, and Navy Yard shootings - that have brought private gun ownership to the forefront of media attention and invited many to reexamine the United States' current policies regarding the issue. Given the frequency of opinion-shaping events and extent of media coverage regarding gun ownership and usage, it is possible that Twitter provides a more convenient and affordable means of periodically tracking public opinion than traditional survey methods. In addition to this, Twitter allows researchers to not only view the opinions of those directly impacted by gun policies in the U.S., but gather an understanding of how the U.S. gun debate resonates with a global audience as well. In addition to this, current reports on reactions to U.S. gun policies do not address how public opinion trends vary according to demographic category. Given that events such as the Trayvon Martin case have brought issues of race to the forefront of debate regarding gun violence, and given that gun owners are overwhelmingly white males (Flock, 2013), it is possible that the gun issue, particularly after racialized events related to the gun control debate, divides itself along gender and/or racial lines. Gathering Twitter users' tweets *and* demographic information allows the researcher to break down unsolicited public opinion by demographic category and determine whether these factions exist.

Given the benefits of using demographic and text-based information from Twitter to examine an issue as volatile and potentially divisive as gun control, this project focuses on Tweets that include the hashtag identifier #guncontrol created from April 20th 2013 to April 22 2013, several months after the Newtown shootings and days after the Boston bombings and blockage of the Manchin-Toomey bill.

Methods:

To begin the data extraction process, information is first scraped from Twitter using the website's streaming API[1]. The streaming API creates an open web connection with the website that allows it to gather a "fire hose" supply of tweets based on particular keyword queries as they are created. This data is downloaded as a JSON file from which needed pieces of information, such as the text of the tweets, the names of the users, or the users' profile picture URLs can be extracted.

Because this project proposes to collect demographic information in tandem with reactions to particular social or political events, the first step in analyzing the data involves mapping spheres of conversation surrounding the event. Tweets are identified as displaying a pro-gun control stance, an anti-gun control stance, a neutral stance or are labeled as uncategorizable. One method of categorizing these tweets uses supervised machine learning. This technique involves developing a training set of tweets that are coded by hand as pro-gun control, anti-gun control, neutral or uncategorizable, running a portion of this training set against itself to validate the

[1] https://dev.twitter.com/docs/streaming-apis

accuracy of the algorithm chosen to match the structure of the coded data against the raw data, and using the *RTextTools*[2] package in *R*[3] to code the rest of the data based on the structure of pre-categorized tweets in the training set. Another method involves using the Amazon Mechanical Turk[4] (AMT) to conduct a qualitative analysis of individual Tweets. For this method three Turkers are asked to code each Tweet as pro-gun control, anti-gun control, neutral or uncategorizable; the final label is that which is agreed upon by at least two Turkers. The objective in using semi-automated means of coding the data is to ensure that the process is scalable and applicable to larger batches of Tweets.

Within this body of data, Twitter users' profile pictures are extracted and used as a source of information regarding their demographic characteristics. The characteristics of each user are determined based on crowdsourced evaluations from the AMT. Photos are shown to the Turkers alongside multiple choice questions that ask them to identify the Twitter user's age, age category (i.e. child, adolescent, adult, senior), sex and race. Because pictures can vary in clarity (multiple individuals may appear in the photo, the Twitter user's face may be somewhat obscured), pictures were shown to three Turkers; the demographic category chosen by two or more Turkers is assumed to accurately reflect the characteristics of the user. Previous research has confirmed that these methods indeed produce reliable assessments of Twitter users' age, race and gender (McCormick, Lee, Cesare and Shojaie 2013).

Preliminary results:

The text of the Tweets gathered was coded using both quantitative SMV classification and qualitative AMT evaluations. SVM classification performed relatively poorly (71% precision, 57% recall), likely due to the fact that tweets were domain specific, did not rely strictly on positive and negative vocabulary to express support for or opposition to gun control legislation, and often included links and references to specific individuals. Given this, qualitative AMT coding was used as the standard classification. Cases for which three Turkers could not agree were coded by hand (N=1213). Sensitivity analyses determined that this hand coding did not introduce any biases into the data. Demographic characteristics were also coded by the Turkers using the methods described above. Incomplete cases and duplicate usernames were removed. These classifications were then used to segment the total tweet corpus into corpuses that represent particular demographics groups and examine the opinions of these groups.

Preliminary analysis reveals that the majority of individuals discussing the issue of gun control on Twitter are white males ages 19 to 35 who are pro-gun control. Demographic ideological divides are present, but surprisingly minimal. The results indicate that there are slightly more white anti-gun control advocates than white pro-gun control advocates, but by a small margin (4%). Likewise, 57% of black and other nonwhite users are pro-gun control, versus only 50% of whites. In regards to gender, 56.9% of women are pro-gun control versus 47.4% of men. Because the racial and ethnic composition of the Twitter community does not parallel offline

[2] http://www.rtexttools.com/
[3] http://www.r-project.org/
[4] https://www.mturk.com/mturk/welcome

settings (for example, a higher proportion of African American internet users than white internet users are Twitter users as well) and because the data contain tweets from users outside the US, these results are not intended to be directly comparable to offline data. Within this context, however, they do signal demographic divisions in opinions regarding the issue of gun control.

In reference to the dialogue contained within the tweets, initial word frequency analyses reveal no obvious racial or gender, or age based divisions in the data. There are, however, some obvious ideological divisions within the corpus. Issues highlighted by pro-gun control and anti-gun control tweeters are chosen strategically. For example, “Boston” is mentioned more frequently by anti-gun control supporters and “Aurora” is mentioned more frequently by pro-gun control supporters. Likewise, CNN appears to be a more popular news source for anti-gun control advocates, whereas Gawker appears frequently as a source for pro-gun control supporters. These trends appear to be consistent across gender and race. Further semi-automated text analysis techniques such as examining word associations or topic modeling within demographic-specific corpuses and/or qualitative text analysis will be conducted in order to tease out potential opinion disparities among individuals of different racial, sex and age categories, as well as understand the presence of homogeneity in Twitter debate surrounding this topic.

References:

Benkler, Yochai. 2006. *The Wealth of Networks.* New Haven: Yale University Press.

Bennett, W Lance. 2003. "New Media Power : The Internet and Global Activism." in *Contesting Media Power*, edited by Nick Couldry and James Curran. Oxford: Rowman and Littlefield:17-38

boyd, danah. 2009. "Twitter: "pointless babble" or peripheral awareness + social grooming? ." vol. 2012. http://www.zephoria.org/thoughts/archives/2009/08/16/twitter_pointle.html.

Conover, MD, and J Ratkiewicz. 2011. "Political polarization on twitter." *Proc. 5th Intl. ….* Retrieved January 20, 2013 (http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2847/3275).

Dickerson, J. 2008. Don't Fear Twitter. *Nieman Reports*: 5-6.

Flock, Elizabeth. 2013. "Gun Owners Overwhelmingly White Males, but Groups Say the Demographics are Changing." *Usnews.com*. June 12. Retrieved September 23. http://www.usnews.com/news/articles/2013/06/12/gun-owners-still-overwhelmingly-white-males

Hubbard, Douglas J. 2011. "Pulse: The New Science of Harnessing Internet Buzz to Track Threats." Hoboken, NY: Wiley.

Krueger, B. S. 2002. "Assessing the Potential of Internet Political Participation in the United States: A Resource Approach." *American Politics Research* 30(5):476–498. Retrieved March 28, 2013 (http://apr.sagepub.com/cgi/doi/10.1177/1532673X02030005002).

Marwick, a. E., and D. Boyd. 2010. "I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience." *New Media & Society* 13(1):114–133. Retrieved October 26, 2012 (http://nms.sagepub.com/cgi/doi/10.1177/1461444810365313).

McCormick, Tyler, Hedwig Lee, Nina Cesare and Ali Shojaie. 2013. "Using Twitter For Demographic Research: Tools for Data Collection." Working Paper No. 127. University of Washington: Center for Statistics and the Social Sciences, Seattle, WA

Murthy, D. 2011. "Twitter: Microphone for the masses?" *Media, Culture & Society* 33(5):779–789. Retrieved November 3, 2012 (http://mcs.sagepub.com/cgi/doi/10.1177/0163443711404744).

Parmalee, John H. and Shannon L. Bichard 2013. “Politics and the Twitter Revolution: How Tweets Influence the Relationship between Political Leaders and the Public.” Lanham, MD: Rowman and Littlefield

Pew Research Center. November 1, 2012. “Nonvoters: Who They Are, What They Think.” Pew Research Center for the People & the Press. Retrieved February 23, 2013 (www.peoplepress.org/2012/11/01/nonvoterswhotheyarewhattheythink/ 4/)

Pew Research Center. February 14, 2013. “The Demographics of Social Media Users - 2012.” Pew Internet and American Life Project. Retrieved February 14, 2013

Sunstein, C. R. 2007. *Republic.com 2.0.* Princeton: Princeton University Press.

Tumasjan, a., T. O. Sprenger, P. G. Sandner, and I. M. Welpe. 2010. “Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape.” *Social Science Computer Review* 29(4):402–418. Retrieved November 4, 2012 (http://ssc.sagepub.com/cgi/doi/10.1177/0894439310386557).

Verba, Sidney. 2013. “and American Surveys Democracy The Citizen as Respondent : Sample Political Science Association , Presidential Address , American.” 90(1):1–7.

Williams, Christine R and Girish Gulati . 2008. "What Is a Social Network Worth? Facebook and Vote Share in the 2008 Presidential Primaries." Conference Papers -- American Political Science Association, p. 1