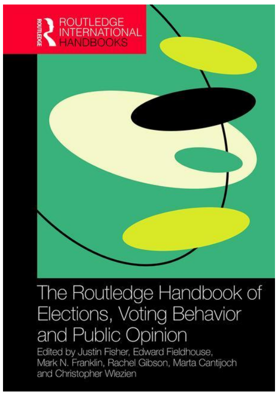

The Routledge Handbook of Elections, Voting Behaviorand Public Opinion

Justin Fisher, Edward Fieldhouse, Mark N. Franklin, Rachel Gibson, Marta Cantijoch, Christopher Wlezien

Making inferences about elections and Public oPinion using incidentally collected data

Publication details
https://www.routledgehandbooks.com/doi/10.4324/9781315712390.ch41
Jonathan Mellon
Published online on: 26 Sep 2017

41

MAKING INFERENCES ABOUT ELECTIONS AND PUBLIC OPINION USING INCIDENTALLY COLLECTED DATA

Jonathan Mellon

Introduction

This chapter discusses the use of large quantities of incidentally collected data (ICD) to make inferences about elections and public opinion. ICD is data that was created or collected primarily for a purpose other than analysis (Sjoberg, Mellon, and Peixoto 2017). The internet has expanded the availability and reduced the cost of ICD with data sources including internet searches, social media data, and civic platforms. This chapter focuses on the uses of ICD in elections and public opinion (EPOP) research and the challenges that researchers face in using it effectively.

ICD is often categorized as "big data." This chapter doesn't use that term for several reasons. First, the term big data has at least six commonly used definitions that are often incompatible with each other (Monroe 2012; Ward and Barker 2013). The most common definitions focus on the amount of data (for instance, Intel uses a cutoff of 300 terabytes of weekly data). Data size is an important criterion when looking at storage or technological requirements for running an analysis (e.g., which database software or cloud service to use), but is less important when considering questions of what types of inference can be made from data. Looking merely in terms of size, both anonymized census records and large sets of tweets are big data. However, census records are some of the least problematic data to draw inferences to the general population from (because they are near complete information on the population), whereas drawing valid inferences from social media data is much more difficult.

Other big data definitions focus on the velocity of the data (how quickly it is produced), on the computing power required to analyze the data, on the extent to which the data is structured in a complex way, or whether particular tools are used to analyze the data (e.g., machine learning, NoSQL databases, or Hadoop). A particular dataset or analysis can easily fit different sets of these definitions, meaning the term big data is uninformative without further explanation. Rather than trying to solve this semantic debate, this chapter focuses on the analysis and inferential issues that political scientists face when studying data that was collected for purposes other than analysis.

This chapter proceeds in four sections. The first focuses on the different types of inferences that elections and public opinion researchers have made from ICD: point estimates of public opinion or party behavior, election forecasts, and estimates of causal relationships. The second discusses how researchers should think about representativeness and validation when using ICD. The third section discusses other common problems that researchers face when analyzing ICD, including dealing with spam and automation. The fourth section applies the ICD framework to a paper analyzing the ideology of Twitter users.

Inferences with ICD

Several uses of ICD are especially relevant to the study of elections and public opinion: making point estimates, election forecasting, and estimating causal relationships. This section reviews the ways in which researchers have used ICD to do each of these.

Point estimates

One way in which ICD EPOP researchers use ICD is making point estimates of public opinion within particular populations. The most commonly used form of ICD for this purpose is internet search data, particularly from Google Trends. Google Trends provides aggregated time series counting the number of searches for a particular search term within a specified geographic area. The time series are available daily or weekly.

Google data has been used in public opinion research to study issue salience. These studies include studying agenda setting by the media (Weeks and Southwell 2010; Ripberger 2011; Granka 2010; Ragas and Tran 2013) and identifying trends in public interest in various environmental issues (Wilde and Pope 2012; Oltra 2011; Anderegg and Goldsmith 2014). Other studies using search data have shown how interest in candidates affects fundraising for that candidate (Ellis, Swearingen, and Ripberger 2011), and the effect of racist attitudes on Barack Obama's 2008 vote share (Stephens-Davidowitz 2012).

Some studies have also used text analysis of tweets as proxies for public opinion. Studies using Twitter have studied the reaction of the public to presidential debates (Wang et al. 2012), including using spikes in tweets to identify key moments within a debate, and analyzed the effect of political events on the public mood (Bollen and Pepe 2011) and the public's engagement with politicians (Raynauld and Greenberg 2014).

Most of the studies using ICD to track public opinion do not conduct any validation of the measures they use to track public opinion and it is therefore not clear whether the results of these studies are valid. The representativeness and ICD section of this chapter discusses some studies which have conducted validation and the extent to which unvalidated data is likely to lead to false inferences.

ICD has also been used to describe the behavior of parties. This analysis tends to be much less problematic as it is relatively straightforward to sample either all the relevant online behavior of parties or a representative subset of it. Examples of this type of analysis include Gibson and Ward's (2003) analysis of Australian party websites using automated content analysis. During that period, this could be seen as a reasonably representative sample of how the political parties used the internet. If the aim had been to make an inference about party behavior in general, the inference would have required more assumptions or validation, but the approach was well suited for the narrower question of how parties use the internet. Similar studies have looked at elections in Germany and Austria (Russmann 2011), Norway (Enli and Skogerbø 2013), and other Australian elections (Bruns and Highfield 2013). Recent studies have also documented the rapid rise in political parties' use of social media across many electoral contexts (Jungherr 2015; Bode and Epstein 2015; Karlsen 2011; Van Dalen et al. 2015; Larsson 2015).

Forecasting

Another major use of ICD is election forecasting. Many papers have used data on the number of times different candidates are mentioned on Twitter to forecast elections, with the assumption that more Twitter mentions is associated with a higher vote share (Tumasjan et al. 2010; Sang and Bos 2012; O'Connor et al. 2010; McKelvey, DiGrazia, and Rojas 2014; Marchetti-Bowick and Chambers 2012; Choy et al. 2012; Digrazia et al. 2013). Election forecasting has also been attempted using Google Trends data on the number of searches for candidate and party names (Graefe and Armstrong 2012; Granka 2013; Polykalas, Prezerakos, and Konidaris 2013a, 2013b). While all of these papers claimed success in this process, they are all based on retrospective forecasts of elections.

Subsequent research has suggested that these "forecasts" succeeded only due to arbitrary decisions (Gayo-Avello 2012) and that when their methods are applied to elections other than the one where success is claimed they perform no better than chance (Gayo-Avello, Metaxas, and Mustafaraj 2011; Metaxas, Mustafaraj, and Gayo-Avello 2011). Additionally, a pre-registered Twitter forecast of the 2015 UK election did not replicate the success of retrospective "forecasts" (Burnap et al. 2016).[1]

Causal relationships

Another form of inference that is sometimes used with ICD is to argue that even though the sample is unrepresentative, the social mechanisms that the authors are testing are not likely to be affected by the sample's unrepresentativeness. This is essentially the same logic that governs external validity in laboratory experiments. As with experiments, the extent to which this is a convincing argument will vary dramatically across studies. This logic of inference is rarely convincing when trying to get exact point estimates of a proportion (e.g., the proportion of voters who will vote for the Democrats), but can be more convincing when trying to understand how two variables will be correlated (e.g., whether consumption of left-wing media content correlates with voting democrat).

One example of this logic is a 61 million person get-out-the-vote experiment that was run on Facebook users by showing them a message about their friends' voting (Bond et al. 2012). Given the size of Facebook, this is interesting in its own right, but it is also plausible that such subtle social nudges are influential outside of Facebook, helping to justify the wider conclusions that the authors drew.

In another example, Mellon, Sjoberg, and Peixoto (2016) examine predictors of petition success on the change.org platform. They argue that the mechanisms tested are sufficiently broad (level of mobilization, institutional support, and regime type) that they are likely to apply to settings beyond the change.org platform. While the analysis itself makes inferences about petitions on the platform, the conclusions are drawn more widely.

In another example, Reddit data was used to examine the types of arguments that are most convincing to other people (Tan et al. 2016). While the data is specific to the Reddit platform, the authors explicitly make wider claims about the mechanisms behind persuasion. In this case, the generalizability of the findings is more difficult to assess. The data the paper uses is based on the ChangeMyView Reddit forum where users specifically ask for people to try and change their minds. It is therefore unclear whether data from this setting is relevant to opinion change more generally. In each of these cases, the mechanisms are argued to be sufficiently similar in the available ICD, that wider conclusions can be drawn.

Drawing useful inferences from ICD

There are two main ways that scholars can draw inferences from ICD: representativeness and proxy validation. Representativeness makes sure that the sample is sufficiently representative of the population that a researcher wants to make inferences about. Proxy validation takes a black box approach, where the key question is whether we can be confident that, for whatever reason, a trend in ICD reliably tracks a real-world phenomenon of interest.

Representativeness

Researchers using ICD need to consider representativeness in two ways. First, whether the platform the ICD is taken from is representative of the wider population that the research is interested in (voters, politicians, the general population, etc.). Second, regardless of the population of interest, researchers have to consider whether the ICD is organized at the correct unit of analysis for the research question.

Representativeness of users

When looking at ICD, the first threat to representativeness is the composition of the platform's users. In the case of social media data, Facebook users tend to be more demographically and politically representative of the general population than Twitter users (Mellon and Prosser 2017), but both groups would take considerable adjustment to make them representative of the general population. As with survey data, when an initial sample is not representative, it is sometimes possible to achieve representativeness using weighting; however, this still assumes the sample is representative within the weighting strata.

Most studies that attempt to make point estimates do not assume that ICD is a sample of the general population of a country, but many studies do try to make inferences about subpopulations. On the other hand, many studies that focus on causal relationships implicitly assume that social media users are sufficiently representative of the general population to draw inferences outside of social media users.

Choosing the correct unit of analysis

Most datasets used by scholars of elections and public opinion are either directly collected by them or by someone who collected it to make inferences about political phenomena. In standard social science data collection, the sampling procedure will generally reflect the analysis to be conducted. If a research question is about countries, data will be collected at the country level and if a question is about the behavior of individual voters, then data will be sampled at the individual level.

By definition, ICD is not collected in this way and instead reflects the priorities of the platform. Consequently, researchers may need to adjust the data in order to make inferences about the phenomena of interest. With ICD, the data will generally be organized at whatever level was most useful for the original purpose. Often this is in the form of event logs, which take an event as the unit of analysis. However, making inferences about the universe of events is often not the aim of a political scientist.

In public opinion research and elections research, the individual is generally treated as the fundamental unit of analysis. That is, we want to know something about the average individual in a population. In most survey research, we are interested in knowing something about the distribution of a variable across individuals. In the case of media analysis, we are usually ultimately interested in understanding the distribution of exposure to possible influences across individuals – for example, how much pro-Labour media is a typical voter exposed to?

Twitter data can be used to make inferences about these different populations. If we are, for instance, interested in using Twitter data in the run-up to a UK election, we could be interested in making inferences about any of the following (even before we consider making inferences beyond Twitter):

1 UK tweets
2 The consumption of Twitter content in the UK
3 The behavior of UK Twitter users

These choices are non-trivial because Twitter usage is highly skewed. The median Twitter account has just one follower (Bruner 2013). The first option is often the default way in which researchers receive Twitter data: a chronological stream of tweets written that match certain criteria (such as location, time, and topic), gathered by storing tweets matching a certain criteria in the streaming API for a certain time period. However, it is not immediately clear why we should care about tweets as a population to make inferences about. If we think that Twitter is politically relevant because it is an important source of campaign information, then we should be focused on the second option: what Twitter users consume, and if we are interested in Twitter as a source of data on the political behavior of individuals, then we should be interested in the third option. A stream of tweets is essentially a measure of individual behavior (tweeting) weighted proportionally to the level of activity of each individual. However, many articles using Twitter data (Jungherr 2014; Raynauld and Greenberg 2014; Jungherr, Jurgens, and Schoen 2011; Christensen 2013; Caldarelli et al. 2014) take it is as given that the content of Twitter as a whole is the most relevant analysis frame.

The second option (the consumption of Twitter content in the UK) is most useful for research looking at Twitter as a medium for media consumption. Obtaining a representative sample of what content is consumed on Twitter is possible using weighting: a researcher simply needs to capture a stream of tweets fitting particular criteria and then subsequently reweight or resample according to the number of followers the creator of each tweet has. This means that a tweet seen by 10,000 followers is weighted 500 times as highly as a tweet seen by 20 followers. While there are some simplifying assumptions[2] in this process, it will create a collection of tweets that much more closely resembles what people see on Twitter. If we are interested in Twitter as a source of information, then this is the most relevant universe.

The difference between what is consumed and what is tweeted is likely to be important. While Twitter users as a whole are numerous enough that they span many sections of society, popular Twitter users tend to be more reflective of existing sources of political influence: for example, celebrities, media figures, academics, and political figures. Focusing on everything that is tweeted would be likely to give the impression that Twitter consumption looks less like traditional media than is actually the case.

The third potential population of interest is Twitter users themselves or a particular subset of the users. Samples of Twitter users can be obtained in a number of ways. It is possible to scrape a random sample of users from the Twitter API by randomly sampling ID numbers from a uniform distribution (Bruner 2013), as user IDs are assigned more or less sequentially over time. As of 2013, around 63 percent of randomly chosen ID numbers resolved to a Twitter user. The downside of this approach is that it is not possible to filter users by particular criteria, so researchers would have to sample the whole of Twitter and then discard all non-relevant users.

Alternatively, a researcher could obtain a representative sample of active users (i.e., users who tweeted at least once in a time period), by collecting all tweets matching particular criteria (e.g., in the UK and mentioning political terms). The researcher can then use the Twitter search API to collect the full tweeting behavior of these users in this time period. This approach is relatively rare in the literature, although Boyadjian and Neihouser (2014) do demonstrate how to collect a panel of Twitter users.

Another approach that has been taken is to define a core set of political Twitter users such as politicians from other sources. Politically interested users can then be further identified by looking at the followers of these core political users (Barberá 2014). Similarly, other studies have looked at all users who tweeted using a particular hashtag (Larsson and Moe 2011).

In light of this discussion, it is perhaps not surprising that Twitter election forecasts have a poor track record, given that: (1) almost all papers on this topic use tweet counts, which (as noted above) neither track Twitter user behavior nor what Twitter users are exposed to, and (2) Twitter users are highly unrepresentative of the general public in every country studied (Vaccari et al. 2013; Mellon and Prosser 2017; Barberá and Rivero 2014).

This is not to say that researchers should never analyze a stream of tweets, simply that they should articulate why doing so answers their research question. While this section has focused primarily on Twitter, these same concerns apply to any ICD analysis.

Proxy validation

Another form of logic that researchers use to make inferences about a population on the basis of ICD is proxy validation. In this case, it is considered sufficient to use the ICD to measure public opinion if we can be confident that there is a strong relationship between the underlying variable in the population and a particular measurement using ICD, even if the mechanism driving the link is not necessarily clear.

An example of the validation logic comes from work using Google Trends. Mellon (2013b) outlines a three-step procedure for determining the extent to which a Google Trends time series can be considered a valid proxy for the salience of a particular issue: face validity, content validity, and criterion validity.

Face validity simply refers to whether or not a Google Trends term initially looks plausible as a proxy for a given variable. For instance, the search term "council housing" seems plausible as a measure of the issue salience of housing in the UK.

Content validity goes a step further and examines the actual search terms used within searches that make up the trend for a keyword. For instance, are Google Trends for "jobs" about searches for employment or the new Steve Jobs biopic? Google Trends allows the top terms for a trend to be downloaded and examined. Problematic terms can then be iteratively removed, to leave only relevant searches.

Criterion validity refers to the extent to which a measure can be shown to correlate with an existing gold standard measure. Given the widespread concerns about traditional data collection techniques such as polling data (Sturgis et al. 2016; Mokrzycki, Keeter, and Kennedy 2009), it is doubtful whether we truly have a gold standard for many public opinion measures, but traditional techniques at least have established standards for assessing their likely quality and unambiguous tests of their accuracy around elections.

When applying these steps to Google Trends series in the US, just 5 out of 20 trends with face validity were shown to possess both content validity and criterion validity (Mellon 2013b). A similarly low validation rate was seen in Spain (5 out of 12) and the United Kingdom (14 out of 39) (Mellon 2013a). In none of the three countries was an initially plausible Google Trends series more likely to turn out to be valid than not. While these steps are designed around using internet search data, many other sources of ICD could potentially benefit from similar steps of validation.

A limited amount of work has been conducted validating trends in Twitter data against public opinion (O'Connor et al. 2010). While the authors of this work are optimistic about the potential for tracking public opinion using Twitter, their results show that the strength of the relationship between public opinion and Twitter sentiment varies greatly over time, sometimes reaching as high as a 0.8 correlation but often showing zero or even negative correlations. This study also therefore casts doubt on the efficacy of Twitter data for making point estimates of public opinion.

Measurement challenges associated with ICD

In addition to the issues about making inferences from the sample themselves, researchers using ICD also need to consider other measurement concerns that are inherent to analyzing these forms of data.

Inferring attitudes from behavior

Survey research generally gathers a large quantity of systematic data about respondents' attitudes, but has relatively limited directly observed behavioral data (vote validation is a notable exception in electoral research). With ICD, the situation is usually reversed: there are large quantities of non-systematic behavioral data with little systematic attitudinal data. The issue for analyzing ICD is therefore how to interpret behaviors.

One example of this problem is Google Trends data. Researchers observe normalized counts of searches for a given term, but have to assume the reasons why these people searched for the terms they did. Are people searching for "Trump" because they plan to vote for Donald Trump in the Republican primary, because they want to find negative information about him, or because they plan to stay at a Trump hotel? With Google Trends data, it is possible to see what other terms are being combined with a search term which can help to disambiguate these meanings (Mellon 2013b), but the meaning of behavioral data will often be ambiguous.

Non-behavior and self-selection

While any survey response is technically behavioral, the fact that the survey respondents are proactively collected by the researcher reduces the impact of self-selection. By contrast, ICD events are proactively generated by the research subjects themselves. Consequently, we only observe any behavior (such as tweeting, posting, or even reading) for people who are sufficiently motivated to take this action. Making inferences about what people in general think on the basis of the actions of the most motivated can therefore be potentially misleading. Consequently, researchers need to make explicit how they are considering the large number of potential subjects in their study who did not take an action. While there may be unprecedented numbers of people searching for Donald Trump, the vast majority of people on any given day will not be doing so. Researchers need to consider whether it is valid to make inferences about people who did not take an action on the basis of the behavior of those people who did take an action.

Artifacts of the platform

Another potential problem with ICD is the extent to which certain behaviors are encouraged or even automated by the platform itself. Google auto-completes searches with suggestions, Facebook and Twitter suggest possible people to connect with, and change.org emails users with suggested petitions to sign. Even email clients will automatically include all previous recipients in a message when a user clicks "reply all." Consequently, it is easy for research to conflate the design of the platform itself with the behavior of users on that platform.

This problem has become more acute with the introduction of algorithmic timelines on several platforms. This means that users are exposed to content on the basis of a proprietary algorithm rather than chronologically. This has led to controversy when the Facebook algorithm was alleged to reduce the visibility of conservative-leaning news outlets (LaCapria 2016) and further highlights the role that a platform's algorithms play in the behavior of its users.

Spam and fake data

Since the collection of ICD is generally not determined by the researcher, there are fewer protections against fake or duplicate information. Twitter and Facebook are both frequently targeted by advertising bots. These may even end up contaminating political data if they make use of popular hashtags, or retweet political information to help hide their tracks. Researchers should proactively look for this kind of contamination when using social media data.

Case study

Despite the challenges that using ICD presents to the researcher, it can and has been used to conduct novel and important analyses of political phenomena. This section briefly describes one such successful attempt conducted by Pablo Barberá, who studied the interactions of politicians and citizens on Twitter with a view to testing the extent to which such ICD could be used to infer ideological orientations. Here we concentrate particularly on how he navigated the concerns of representativeness and drawing valid inferences (Barberá 2014).

Estimating ideology from Twitter

In this study, Barberá aims to make inferences about the ideology of elite political actors (politicians, media outlets, and think tanks) based on the composition of users they interact with and who follows them. While legislators have long been classified on the basis of ideology, Barberá points out that other types of political actors have generally not been able to be rated on the same scale. Developing a method that can estimate ideology for any political actor (providing they are on Twitter) thus offers a potentially very useful new resource for political scientists.

Barberá correctly samples at the level of the user by first choosing several hundred political Twitter accounts in each of the countries in the study. He then downloads the information for each Twitter user who follows at least one of these target accounts. These users then allow the position of the elites and general users to be simultaneously estimated on the basis of their connections to each other. Barberá uses the logic of validation at both the standard user and the elite level. For normal users, he validates the ideal point estimates against matched data on users' campaign contributions and party registration. At the elite level he validates the ideal point estimates against DW-nominate scores in the US and expert survey measures of ideology in other countries. Barberá also accounts for potential contamination of measurement by spam on Twitter by excluding accounts with low levels of activity.

Barberá is also clear in outlining what population he intends to make inferences about: political actors in general at the elite level and Twitter users at the mass level, finding that most exchanges on Twitter take place between users with similar ideological positions and that a small cohort of highly engaged right-wing users disproportionately drive the public conversation on Twitter. This analysis exemplifies good practice for analyzing ICD by combining ICD with traditional data sources, appropriately choosing the unit of analysis, and accounting for potential biases in the data caused by the platform.

Conclusions

This chapter has outlined the sources of incidentally collected data (ICD) that have been used in public opinion and elections research. The nature of the data necessitates a careful consideration of what population is being researched and how the behavior on these platforms can be interpreted.

While ICD sources are highly varied, researchers would be advised to consider the following questions when deciding whether to use ICD in their research. The first question is whether the research question is best answered using ICD or is there another data source that would work better? The second question researchers should ask is how the ICD they are using was collected and make sure that this process is accounted for in the analysis process. Finally, researchers should ask what population they want to make inferences about and whether the data they have is structured appropriately to make these inferences.

While this chapter has emphasized the limitations of ICD in political analysis, this should not distract from the substantial research possibilities that ICD opens up. There are very limited possibilities for collecting large-scale network data outside of ICD, for instance. It fact it is precisely because of the increasing use of ICD in political science that it is important for researchers to understand how to best make use of these data sources and understand their limitations.

Notes

1 It should be noted that all the forecasts in the 2015 UK forecasting symposium performed poorly (Fisher and Lewis-Beck 2015), so Twitter forecasting was certainly not the only method called into question.

2 In particular, we assume that all Twitter followers are equally likely to read a tweet and that the follower count attached to a tweet that matches a certain criterion (such as originating from the UK) is representative of the number of UK followers.

References

Anderegg, W. R. L. and Goldsmith, G. R. (2014) "Public Interest in Climate Change over the Past Decade and the Effects of the 'Climategate' Media Event," *Environmental Research Letters*, vol. 9, no. 5, May, 1–8.

Barberá, P. (2014) "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data," *Political Analysis*, vol. 23, no. 1, January, 76–91.

Barberá, P. and Rivero, G. (2014) "Understanding the Political Representativeness of Twitter Users," *Social Science Computer Review*, vol. 33, no. 6, December, 712–729.

Bode, L. and Epstein, B. (2015) "Campaign Klout: Measuring Online Influence During the 2012 Election," *Journal of Information Technology and Politics*, vol. 12, no. 2, April, 133–148.

Bollen, J. and Pepe, A. (2011) "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena," Paper presented at the 2011 Conference on Weblogs and Social Media, Barcelona, July 17–21.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., and Fowler, J. H. (2012) "A 61-Million-Person Experiment in Social Influence and Political Mobilization," *Nature*, vol. 489, no. 7415, September, 295–298.

Boyadjian, J. and Neihouser, M. (2014) "Why and How to Create a Panel of Twitter Users," in Skoric, M. M., Parycek, P., and Sachs, M. (eds.) *CeDEM Asia 2014: Conference for E-Democracy an Open Government*, Krems: Donau-Universität Krems, 247–252.

Bruner, J. (2013) "Tweets Loud and Quiet," *O'Reilly Radar*, December 18, 2013, available online: www.oreilly.com/ideas/tweets-loud-and-quiet [accessed February 17, 2017].

Bruns, A. and Highfield, T. (2013) "Political Networks on Twitter: Tweeting the Queensland State Election," *Information, Communication and Society*, vol. 16, no. 5, June, 667–691.

Burnap, P., Gibson, R., Sloan, L., Southern, R., and Williams, M. (2016) "140 Characters to Victory?: Using Twitter to Predict the UK 2015 General Election," *Electoral Studies*, vol. 41, March, 230–233.

Caldarelli, G., Chessa, A., Pammolli, F., Pompa, G., Puliga, M., Riccaboni, M., and Riotta, G. (2014) "A Multi-Level Geographical Study of Italian Political Elections from Twitter Data," *PLoS ONE*, vol. 9, no. 5, May, e95809.

Choy, M., Cheong, M., Laik, M. N., and Shung, K. P. (2012) "US Presidential Election 2012 Prediction Using Census Corrected Twitter Model," *ArXiv.org: Applications; Computers and Society*, November.

Christensen, C. (2013) "Wave-Riding and Hashtag-Jumping: Twitter, Minority 'Third Parties,' and the 2012 US Election," *Information, Communication and Society*, vol. 16, no. 5, June, 646–666.

Digrazia, J., McKelvey, K., Bollen, J., and Rojas, F. (2013) "More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior," *PLoS One*, vol. 8, no. 11, November, e79449.

Ellis, C., Swearingen, C., and Ripberger, J. T. (2011) "Examining the Impact of Public Attention on Fundraising in U.S. Senate Elections," Paper presented to the 2011 Annual Meeting of the American Political Science Association, Seattle, September 1–4.

Enli, G. S. and Skogerbø, E. (2013) "Personalized Campaigns in Party-Centered Politics," *Information, Communication and Society*, vol. 16, no. 5, June, 757–774.

Fisher, S. D. and Lewis-Beck, M. S. (2015) "Forecasting the 2015 British General Election: The 1992 Debacle All Over Again?," *Electoral Studies*, vol. 41, March, 225–229.

Gayo-Avello, D. (2012) "'I Wanted to Predict Elections with Twitter and All I Got Was This Lousy Paper' – A Balanced Survey on Election Prediction Using Twitter Data," *ArXiv.org: Computers and Society; Computation and Language; Physics and Society*, April, 13.

Gayo-Avello, D., Metaxas, P. T., and Mustafaraj, E. (2011) "Limits of Electoral Predictions Using Twitter," *Association for the Advancement of Artificial Intelligence*, 490–493, available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.221.5177&rep=rep1&type=pdf [accessed November 1, 2016]. In Proceedings of the International AAAI Conference on Weblogs and Social Media. Barcelona, Spain, AAAI.

Gibson, R. and Ward, S. (2003) "Letting the Daylight in? Australian Parties' Use of the World Wide Web at the State and Territory Level," in Gibson, R., Nixon, P. and Ward, S. (eds.) *Political Parties and the Internet: Net Gain?*, London: Routledge: 161–174.

Graefe, A., and Armstrong, J. S. (2012) "Predicting Elections from the Most Important Issue: A Test of the Take-the-Best Heuristic," *Journal of Behavioral Decision Making*, vol. 25, no. 1, July, 41–48.

Granka, L. (2010) "Measuring Agenda Setting with Online Search Traffic: Influences of Online and Traditional Media," Paper presented at the 2010 Annual Meeting of the American Political Science Association, Washington, DC, September 2–5.

Granka, L. (2013) "Using Online Search Traffic to Predict US Presidential Elections," *PS: Political Science and Politics*, vol. 46, no. 2, April, 271–279.

Jungherr, A. (2014) "The Logic of Political Coverage on Twitter: Temporal Dynamics and Content," *Journal of Communication*, vol. 64, no. 2, April, 239–259.

Jungherr, A. (2015) "Twitter Use in Election Campaigns: A Systematic Literature Review," *Journal of Information Technology and Politics*, vol. 13, no. 5, December, 72–91.

Jungherr, A., Jurgens, P., and Schoen, H. (2011) "Why the Pirate Party Won the German Election of 2009 or The Trouble with Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G. and Welpe, I. M. 'Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment'," *Social Science Computer Review*, vol. 30, no. 2, May, 229–234.

Karlsen, R. (2011) "A Platform for Individualized Campaigning? Social Media and Parliamentary Candidates in the 2009 Norwegian Election Campaign," *Policy and Internet*, vol. 3, no. 4, December, 1–25.

LaCapria, K. (2016) "The Algorithm is Gonna Get You," *The Technosceptic*, January 21, 2016, available online: https://thetechnoskeptic.com/algorithm-is-gonna-get-you/ [accessed February 17, 2017].

Larsson, A. O. (2015) "The EU Parliament on Twitter – Assessing the Permanent Online Practices of Parliamentarians," *Journal of Information Technology and Politics*, vol. 12, no. 2, April, 149–166.

Larsson, A. O., and Moe, H. (2011) "Studying Political Microblogging: Twitter Users in the 2010 Swedish Election Campaign," *New Media and Society*, vol. 14, no. 5, August, 729–747.

Marchetti-Bowick, M. and Chambers, N. (2012) "Learning for Microblogs with Distant Supervision: Political Forecasting with Twitter," Paper presented at the 2012 Annual Conference of the European Chapter of the Association for Computational Linguistics, Avignon, April 23–27.

McKelvey, K., DiGrazia, J., and Rojas, F. (2014) "Twitter Publics: How Online Political Communities Signaled Electoral Outcomes in the 2010 US House Election," *Information, Communication and Society*, vol. 17, no. 4, April, 436–450.

Mellon, J. (2013a) "Where and When Can We Use Google Trends to Measure Issue Salience?," *PS: Political Science and Politics*, vol. 46, no. 2, April, 280–290.

Mellon, J. (2013b) "Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience," *Journal of Elections, Public Opinion and Parties*, vol. 24, no. 1, February, 45–72.

Mellon, J. and Prosser, C. (2017) "Twitter and Facebook Are Not Representative of the General Population: Political Attitudes and Demographics of British Social Media Users," Research and Politics, Jul–Sep: 1–9.

Mellon, J., Sjoberg, F. M., and Peixoto, T. (2016) "Connective Action: The Prevalence and Effects of Large Scale Transnational Activism," Working Paper, The World Bank.

Metaxas, P. T., Mustafaraj, E., and Gayo-Avello, D. (2011) "How (Not) to Predict Elections," Paper presented at the 2012 International Conference on Social Computing, Amsterdam, September 3–5.

Mokrzycki, M., Keeter, S., and Kennedy, C. (2009) "Cell-Phone-Only Voters in the 2008 Exit Poll and Implications for Future Noncoverage Bias," *Public Opinion Quarterly*, vol. 73, no. 5, December, 845–865.

Monroe, B. L. (2012) "The Five Vs of Big Data Political Science: Introduction to the Virtual Issue on Big Data in Political Science," *Political Analysis*, vol. 21, no. 5, April, 1–9.

O'Connor, B., Balasubramanyan, R., Routledge, B., and Smith, N. (2010) "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," Paper presented at the 2010 International AAAI Conference on Weblogs and Social Media, Washington, DC, May 23–26.

Oltra, C. (2011) "Stakeholder Perceptions of Biofuels from Microalgae," *Energy Policy*, vol. 39, no. 3, March, 1774–1781.

Polykalas, S., Prezerakos, G., and Konidaris, A. (2013a) "A General Purpose Model for Future Prediction Based on Web Search Data: Predicting Greek and Spanish Election," Paper presented at the 2013 International Conference on Advanced Information Networking and Applications Workshops, Barcelona, March 25–28.

Polykalas, S., Prezerakos, G., and Konidaris, A. (2013b) "An Algorithm Based on Google Trends' Data for Future Prediction. Case Study: German Elections," Paper presented at the 2013 International Symposium on Signal Processing and Information Technology, Athens, December 12–15.

Ragas, M. W. and Tran, H. (2013) "Beyond Cognitions: A Longitudinal Study of Online Search Salience and Media Coverage of the President," *Journalism and Mass Communication Quarterly*, vol. 90, no. 3, September, 478–499.

Raynauld, V. and Greenberg, J. (2014) "Tweet, Click, Vote: Twitter and the 2010 Ottawa Municipal Election," *Journal of Information Technology and Politics*, vol. 11, no. 4, 412–434.

Ripberger, J. T. (2011) "Capturing Curiosity: Using Internet Search Trends to Measure Public Attentiveness," *Policy Studies Journal*, vol. 39, no. 2, May, 239–259.

Russmann, U. (2011) "Targeting Voters via the Web – A Comparative Structural Analysis of Austrian and German Party Websites," *Policy and Internet*, vol. 3, no. 3, September, 26–48.

Sang, E. T. K. and Bos, J. (2012) "Predicting the 2011 Dutch Senate Election Results with Twitter," Paper presented at the 2012 Workshop on Semantical Analysis in Social Media, Avignon, April 23.

Sjoberg, F. M., Mellon, J., and Peixoto, T. (2017) "The Effect of Bureaucratic Responsiveness on Citizen Participation," *Public Administration Review*, vol. 77, no. 3, May/June, 340–351.

Stephens-Davidowitz, S. I. (2012) "The Effects of Racial Animus on a Black Presidential Candidate: Using Google Search Data to Find What Surveys Miss," Working Paper, Harvard University.

Sturgis, P., Baker, N., Callegaro, M., Fisher, S. D., Green, J., Jennings, W., Kuha, J., Lauderdale, B. E., and Smith, P. G. (2016) *Report of the Inquiry into the 2015 British General Election Opinion Polls*, London: Market Research Society and British Polling Council.

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., and Lee, L. (2016) "Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-Faith Online Discussions," Paper presented at the 2016 Conference on the World Wide Web, Montreal, April 11–15.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010) "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," Paper presented at the 2010 Conference on Weblogs and Social Media, Washington, DC, May 23–25.

Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., and Tucker, J. A. (2013) "Social Media and Political Communication: A Survey of Twitter Users during the 2013 Italian General Election," *Rivista Italiana di Scienza Politica*, vol. 43, no. 3, December, 381–409.

Van Dalen, A., Fazekas, Z., Klemmensen, R., and Hansen, K. M. (2015) "Policy Considerations on Facebook: Agendas, Coherence, and Communication Patterns in the 2011 Danish Parliamentary Elections," *Journal of Information Technology and Politics*, vol. 12, no. 3, July, 303–324.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., and Narayanan, S. (2012) "A System for Real-Time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle," Paper presented at the 2012 Annual Meeting of the Association for Computational Linguistics, Jeju, July 8–14.

Ward, J. S. and Barker, A. (2013) "Undefined by Data: A Survey of Big Data Definitions," *ArXiv.org: Databases*, September, 2.

Weeks, B. and Southwell, B. (2010) "The Symbiosis of News Coverage and Aggregate Online Search Behavior: Obama, Rumors, and Presidential Politics," *Mass Communication and Society*, vol. 13, no. 4, September, 341–360.

Wilde, G. R. and Pope, K. L. (2012) "Worldwide Trends in Fishing Interest Indicated by Internet Search Volume," *Fisheries Management and Ecology*, vol. 20, no. 3, April, 211–222.