Using Twitter to Observe Election Incidents in the United States*

Walter R. Mebane, Jr.† Logan Woods‡ Joseph Klaver§

Alejandro Pineda¶ Patrick Wu‖ Blake Miller**

August 26, 2016

*Prepared for presentation at the 2016 Annual Meeting of the American Political Science Association, Phildelphia, PA, September 1–4, 2016. Thanks to Josh Pasek for letting us use his script used to extract data from Sysomos, Catherine Morse for advice regarding use of Sysomos and Joseph Hansel for assistance. Work supported by NSF award SES 1523355.

†Professor, Department of Political Science and Department of Statistics, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: wmebane@umich.edu).

‡Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: ltwoods@umich.edu).

§Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: jrklav@umich.edu).

¶Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: ajpineda@umich.edu).

‖Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: pywu@umich.edu).

**Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: blakeapm@umich.edu).

Abstract

Election results are frequently disputed, and information about election disputes can usefully supplement data about electors and votes in election forensics analysis. In countries like Germany and Mexico institutions exist that aggregate and publicize such disputes, but no such institution exists in the United States. We propose infrastructure to accumulate information from Twitter that contains election observations by individuals all across the United States. In this paper we focus on the presidential primary elections and caucuses held across the country in 2016. To classify Tweets for relevance and by type of election incident, we use (or propose to use) automated machine classification methods (a support vector machine in an active learning framework). We obtain data from every state. We demonstrate that for election day in one state (California), the distribution of types of incidents revealed by data developed from Twitter roughly matches the distribution of complaints called in to a hotline run on that day by the state.

1 Introduction

Election forensics is the field devoted to using statistical methods to determine whether the results of an election accurately reflect the intentions of the electors. Many methods for trying to detect election frauds have been proposed (e.g. Myagkov, Ordeshook and Shaikin 2009; Levin, Cohn, Ordeshook and Alvarez 2009; Mebane 2010; Pericchi and Torres 2011; Cantu and Saiegh 2011; Beber and Scacco 2012; Mebane 2014; Montgomery, Olivella, Potter and Crisp 2015; Mebane 2016). Most such methods analyze information about voter participation or voters' choices, looking statistically for patterns that suggest frauds occurred. It would be useful to draw other information into statistical analysis, both generally to enhance diagnosis of what happened in an election and more specifically to help address the primary challenge for election forensics: trying to tell whether patterns in election results that may appear anomalous in statistical estimates and tests are the results of election frauds or of strategic behavior (Mebane 2013, 2016).

Mebane, Klaver and Miller (2016) and Mebane and Wall (2015) have begun such efforts using data respectively from Germany and from Mexico. In Germany the auxiliary data comes from citizen complaints about the federal election filed with a committee of the *Bundestag* (we say more about these below), and in Mexico the additional information comes from petitions parties filed to try to nullify the votes counted in particular ballot boxes. In both cases there are strong signs that election forensics statistics are responding to strategic behavior or to parties' tactical actions, as well as perhaps to frauds.

Performing such analysis for elections in the United States faces difficulties. Leaving aside the difficulty of obtaining suitable low-level (say precinct-level) information about voters' participation and choices in the United States, information about election complaints is difficult to obtain for several reasons. The complaint process in most states is convoluted and characterized by multiple possible channels for disputes, depending on which particular election law was supposedly violated. These channels could include submitting a complaint or dispute via an online portal, reporting an incident via phone, printing out a particular form and submitting a hard copy, or even simply emailing the relevant election authority. In many cases the process for filing a dispute, even once the appropriate channel has been determined, is itself burdensome, leading to few complaints being submitted. For instance, all complaints submitted in compliance with the Help America Vote Act of 2002 must be notarized. Consequently very few complaints are submitted via this process. Few states make what complaint data exist from other official channels publicly available.

To get information about American election administrative performance from individual observers throughout the country, we turn to data from Twitter. We describes steps we have taken to construct infrastructure to develop information from Tweets that contains election observations by individuals all across the United States. Here initially we focus on the presidential primary elections and caucuses held across the country in 2016 (we look forward to deploying a version of this machinery during the upcoming general election). The system involves collecting information about election officials' and other leading actors' Twitter accounts, techniques for searching for and extracting Tweets, and classifying Tweets for relevance and for type of incident. For the classification tasks we use automated machine classification methods. We are able to demonstrate that for election day in one state (California), the distribution of types of incidents revealed by data developed from Twitter roughly matches the distribution of complaints called in to a hotline run on that day by the state. In terms of clarity of type definition and in terms of number and geographic dispersion of incidents, the data derived from Tweets may be superior to the officially collected hotline data.

1.1 Modeling and Comparative Motivations

Information about elections from citizen observers—which is what "complaints" from citizens are—would be generally useful, but it would be particularly useful to supplement statistical analysis of eligible voter and vote-count data. In particular such information can be helpful to resolve the key challenge of distinguishing effects of election frauds from effects of strategic behavior.

Consider estimates for elections in three states (Arizona, California, Texas) in 2006 from the likelihood finite mixture model of Mebane (2016). Table 1 reports the estimated parameters. Likelihood ratio test statistics show that frauds as defined by the model are statistically significant in all the elections. Estimates of the frauds probabilities f_{i} and f_{e} show that in most of the elections the frauds are "incremental frauds," but in the race for governor in Texas there is appreciable "extreme fraud." By the criteria discussed in Mebane (2016) the diagnosis in all the elections that have no or negligible "extreme fraud" would most likely be that the estimates result from strategic behavior. But it would be good to be able to confirm or refute that interpretation with additional data such as what citizen observations might supply.

*** Table 1 about here ***

Mebane, Klaver and Miller (2016) use complaints filed with a committee of the *Bundestag* (Ziblatt 2009; Breunig and Goerres 2011) to develop information that they use to evaluate how much election forensics statistics measure frauds as opposed to strategic behavior. A summary of the kind of information used in that analysis appears in Table 2, which shows the frequency of different types of complaints in the German data. The type codes Mebane, Klaver and Miller (2016) use are modified versions of categories constructed as part of the Election Incident Reporting System (EIRS) originally developed for use in the United States (Verified Voting Foundation 2005; Hall 2005; Johnson 2005). See Mebane and Klaver (2015) for details regarding the adaptation to the German context. Across Germany in 2005 and 2009 there are fewer than 100 complaints of each type in each election, but Mebane, Klaver and Miller (2016) relate the distribution of complaints by type across *Wahlkreise* (election districts) to the distribution of statistics such as f_{i}. They find that in Germany f_{i} and other statistics respond significantly to strategic behavior as well as to whatever is prompting many of the complaints.

*** Table 2 about here ***

The analysis in Germany supports the case that, as Mebane (2016) argues, election forensics statistics respond to both strategic behavior and to frauds. Analogous analysis by Mebane and

Wall (2015), using nullification petition data from Mexico, has similar implications. Probably different kinds of statistics respond differently to various stimuli. Having detailed and localizable data from citizen observations can assist our ability to study the relationships and the variations.

2 Election Complaints in the United States

One problem is that nothing like the centralized system of election complaints that exists in Germany exists in the United States. Each state implements its own complaint process, or none. The Help America Vote Act (HAVA) mandates that each state create a complaint process (42 U.S.C. §402 (a) (1), (2002)), and state codes implementing HAVA frequently state that resolved complaints will be made available to the public online. But the HAVA infrastructure is moribund and useless. Using the methods described in Appendix section 5.1, we reviewed information about complaint procedures in each state. Table 3 summarizes information about the online presence of complaint procedures: whether links could be found to state HAVA forms or to other complaint-related forms; whether it is possible to submit complaints online; and whether some information about submitted complaints appears to be available online. As can be seen, few states allow complaints to be submitted online. Moreover the HAVA complaint process is burdensome, requiring the submission of notarized documents. The submission process in many states is equally burdensome. Submitted complaints are not actually available online for all states that have online links supposedly to the complaints, although some states do post some version of the official rulings on selected complaints.

*** Table 3 about here ***

The inutility of the HAVA complaint procedures is exemplified by the number of HAVA complaints that exist. Table 4 reports the number of HAVA complaints we were able to obtain after contacting all the states. We obtained HAVA complaints for five states. Most states did not respond substantively to written inquiries.[1] The numbers of HAVA complaints we received from

[1] Although the plan to implement HAVA in California states, "The determination must be posted on the Secretary of State's website, unless such posting might compromise a criminal investigation or other enforcement action" (Califor-

states is miniscule. For instance, Ohio sent the single HAVA complaint submitted regarding the 2012 election. Perhaps the reason for the general lack of responsiveness is that frequently a state had no HAVA complaints. The general lack of existence of HAVA complaints was confirmed by a commissioner of the Election Assistance Commission.[2] HAVA complaints cannot be used as a measure of the quality election administration or the existence of possible election frauds.

*** Table 4 about here ***

Many states have alternative avenues for election complaints outside of the HAVA-mandated procedures. These processes produce more extensive and more widely available documents, but even so the state-level complaints are not sufficient to provide citizen observation data across the whole of the United States. Table 5 lists states for which we have been able to obtain documents regarding state complaints. Despite what the "Complaints Online" column in Table 3 may suggest, to date we have obtained documents from only 11 states. In some cases documents are online, but in other cases we received documents (some electronic, some on paper) in response to written requests. Beyond the limited availability of any information regarding the complaints, the "complaints" we have obtained are often of limited utility. In a few cases we have the original text of complaints, but in most cases the documents reflect the results of administrative adjudication of each complaint. In Germany Mebane, Klaver and Miller (2016) also have only the reports of

nia Help America Vote Act State Plan 2010 Update, §254(a)(9)(4), (2010)), California refused to provide any HAVA complaints they may have. They wrote,

> "It is the longstanding policy and practice of the Secretary of State, like all agencies with investigative functions, to treat complaints and information associated with those functions as exempt from the requirements of disclosure pursuant to the California Public Records Act. Specifically, Government Code section 6254(f) allows the non-disclosure of investigative information because the disclosure of the information would impair an agency's ability to conduct investigations in the public interest.
>
> "Consistent with guidelines issued by the California Attorney General, material covered by the non-disclosure exemption of Government Code section 6254(f) includes, 'Records of complaints, preliminary inquiries to determine if a crime has been committed, and full-scale investigations, as well as closure memoranda...'
>
> "For these reasons, the Secretary of State is unable to provide the information you have requested. However, we are open to speaking to you to discuss ways in which we might be able to provide helpful information to you and your research." (Secretary of State, Constituent Affairs 2015)

We had requested copies of any HAVA complaints going back to 2006.

[2]Personal communication from Matt Masterson on March 10, 2016, at the Election Verification Network (EVN) Conference, Washington, DC.

official adjudications by the *Bundestag* committee, so such indirection is not necessarily a problem. More problematic is that in most states all or most complaints made available concern allegations of campaign finance violations or of procedural irregularities committed by parties or by candidates. Few complaints concern citizen observation of other particulars of the electoral process, including especially actions by elected or appointed election officials. The state complaints also cannot generally be used to assess election administration or to detect election frauds.

*** Table 5 about here ***

Some states operate election-day hotlines, usually phone hotlines. The states we identified as having such a hotline operation in the most recent election are listed in Table 6. Just fewer than half the states have a hotline.[3] Some states that operate hotlines say they do not record information about the calls they deal with.[4] Hotline data also cannot be used to generate data for the whole country. Below we will use data provided by California from the hotline that state operated for the 2016 primary election.

*** Table 6 about here ***

3 Using Twitter to Capture Election Observations

We begin to construct infrastructure to allow information in Tweets to be used to build data regarding election observations by individuals. We focus on the presidential primary and caucus elections in all states in 2016. We collect Tweets from within date windows beginning ten days before and ending ten days after each election day. For states that allow absentee (mail-in) voting in primaries, we begin collection on the first day that absentee ballots can be submitted as votes.

[3] In some states some counties operate hotlines, and in others voters are simply instructed to contact the elections division main phone number.

[4] We have distributed a survey to state election officials asking them about their voter hotline policies, and are still collecting data on this topic.

3.1 Collecting Twitter Data

We used two modalities for collecting Tweets: the Sysomos MAP (Sysomos 2016) search tool and archive and the Twitter API (Twitter, Inc. 2016*b*).

With Sysomos MAP (Sysomos 2016) we have so far used very extensive sets of search terms only for six states, while for the others we use more limited sets of search terms. We used the more extensive sets when downloading Tweets manually, while the more limited sets are used when downloading using a script in an automated process.[5] The states for which more extensive sets are used are Arizona, California, Colorado, Connecticut, Illinois and Washington. The search terms used in these cases are listed in Table 7. To define the search terms in the less extensive sets of terms, we first obtained a list of election official, party and other Twitter accounts ("handles") (see Appendix section 5.2 for details regarding compiling the list).[6] We combined the most productive kinds of keywords found in performing the manual searches (e.g., "azprimary") with search terms that would capture Tweets sent to officials (e.g., "to:CASOSvote"). The resulting sets of terms are listed in Table 8.

*** Tables 7 and 8 about here ***

Using the Twitter API (Twitter, Inc. 2016*b*) we downloaded the timelines of 493 Twitter accounts.[7] Use of the API gave more control over the data than the use of Sysomos, as Sysomos only returns certain fields, and the data returned is from a random sample (which we cannot be certain is truly random). The API returns more comprehensive (and more complete) data. To access the API, we registered an application with Twitter.com, giving us the security tokens necessary to query data from Twitter's database. Our goal was to pull entire timelines from 493 accounts (for perspective, one California account had over eleven thousand Tweets in their timeline). Further details about building the list of accounts and about the process of extracting

[5]The script is still running, downloading a few files each day. Here we use files downloaded as of July 30, 2016.

[6]The proportion of county election offices that have an affiliated Twitter account varies greatly across states.

[7]There are actually two main API's: REST API and the Streaming API (Twitter, Inc. 2016*a*). The REST API requires multiple pulls for access to historic data, while the Streaming API requires an open connection for up-to-date content. We used the REST API, as our content was from the first half of 2016.

Tweets using the API are in Appendix section 5.2.

Table 9 shows the number of unique Tweets downloaded from each state. Retweets are excluded. We use the `location` field in downloaded Tweets to determine the state for each Tweet.[8] In general the most Tweets are from states for which we used the larger sets of search terms, but not always. From California there are more than 60,000 Tweets, but for each other state there are less than 20,000 Tweets. South Dakota has the smallest number of Tweets (n = 25).

*** Tables 9 about here ***

3.2 Categorizing Twitter Data

To determine whether a downloaded Tweet includes any relevant observations of the electoral process and then to say what types of incidents are being reported, we augment, clean and classify the Tweets.

3.2.1 Text Augmentation

We augment the text "content" of each Tweet in two ways. In general we get the resource, if any, located at each URL the content contains. If that resource contains any text, we capture that text and append it to the original content.[9] If that resource contains an image, we capture the image's URL.[10] We plan to augment the original Tweet content also with a text description of the image, but this process has not yet been fully operationalized. See Appendix section 5.4 for a description of the process we have used to date with a sample of Tweets. As it is, images often affect human coders' judgments regarding any information Tweets may contain, but the machine classification algorithm currently does not have access to images or descriptions of images.

[8] In future versions we plan to use geolocation data to identify the location from which each Tweet was sent.
[9] Specifically, we capture any text in the `og:description` field in the resource's HTML code.
[10] Specifically, we capture any URL in the `og:image` field in the resource's HTML code.

3.2.2 Tweet Text Cleaning

Cleaning the augmented Tweet content involves removing nonprintable characters, stray HTML codes, internal quotation marks and the ‘*’ character. For the version of the contents used in machine classification and active learning processes, we also removed URLs and made some frequently occurring text strings generic instead of specific to each state. The latter changes replaced some state-specific strings with strings like “#XXvotes,” “#XXprimary,” “#XXcaucus” and “#XXvoterfraud,” where “XX” originally was the postal code abbreviation for a state. We did this to enhance the comparability of Tweets across states for the machine classification algorithm.

3.2.3 Classifying Tweets

To determine whether each downloaded Tweet includes relevant observations, we began by using humans[11] to examine the raw Tweets directly. A Tweet that contains relevant observations about electoral processes is coded to be a “hit.” Each hit was also classified into one or more categories based on the incident type categories in EIRS and in Mebane, Klaver and Miller (2016). Through several rounds of coding, discussion and recoding of random samples of Tweets from Arizona, California, Colorado, Connecticut and Washington[12] we developed consensus criteria for deciding that a Tweet is a hit and for what types to use to classify incidents. The procedure we developed for humans to use when making hit determinations is shown in Figures 1 and 2, the background for which is discussed in Appendix section 5.5.

*** Figures 1 and 2 about here ***

The coding rules for categorizing the incidents to which hits refer are described in Appendix section 5.6. For coding incidents by type there are 15 categories: Absentee, Mail-In, or Provisional Ballot Issue; Registration Issues; Disability/Accessibility Problem; Improper Outside Influence; Other Ballot Problems; Election Official Complaints/Incidents; Electoral System; Voter Fraud; Voter Identification Issues; Long Lines/Crowded Polling Place; Polling Place Problems;

[11]The human coders were subsets of this paper’s authors.

[12]In Washington Tweets come from both the Democratic caucuses and the Republican primary elections.

Voting Machine complaints; Unspecified Other; Positive; and Ambiguous. These categories collapse several EIRS categories into each other, and definitions of categories are modified accordingly. Categories are collapsed into each other when they are thematically related. For example, the categories regarding mail-in, provisional, and absentee ballots are combined. An additional category, Not Hit, is used when a human coding a Tweet the machine classification algorithm classified as a hit decides the Tweet is not a hit.

To produce a training set to use to start the machine classification algorithm, we used a stratified random sample[13] of Tweets from the manual Sysomos downloads from Arizona, California, Colorado, Connecticut and Washington. The Tweets in that sample were coded as "hit" or "not a hit" based on whether at least three of five human coders agreed (upon coding the Tweets again) that the Tweet is a hit or, for Tweets that did not attract such agreement, by using the flowchart. This produced an initial training set containing 192 hits and 806 not-hits.

We use a support vector machine (SVM) in an active learning framework to classify Tweets. Active learning allows us to build a training set with fewer labeled observations, and a better balance between classes, which is useful because of the scarcity of the positive "hit" class (Miller and Klaver 2016). Prior to classification, we first preprocess the text of each Tweet's augmented content. This involves removal of all duplicate texts, stemming, stop word removal and the eventual transformation of text data into a numeric matrix. After stemming, we remove stop words. Stop words are words we believe have low relative information value, or words we believe are uninformative to our classification task. Such words include "and," "the," etc. Once we have preprocessed the words in each document, we use a word n-gram model for Tweet text and a character n-gram model for hashtags to convert the Tweet corpus into a document term matrix: each row represents a Tweet and each column represents a unique character or word unigram, bigram or trigam. Cell values represent the count of each word in the document. Finally we do a TF-IDF transformation of the raw count matrix (Leopold and Kindermann 2002; Lan, Tan, Low and Sung 2005). Because the feature space is high dimensional, and we want to avoid overfitting,

[13]For a description of the sample see Appendix section 5.3.

we select features using the coefficients of a linear SVM with ℓ_1 norm penalty. Features with SVM coefficients lower than the mean of all coefficients are discarded (Rakotomamonjy 2003).

Our initial classificatory scheme to discriminate hits from not-hits is based on the initial training set. Because the initial training set is small, we use active learning, an iterative supervised machine learning technique (Settles 2010). This framework uses uncertainty sampling to identify observations that we should label by hand to provide the most useful new input to the next iteration of the classifier. At each iteration, we train a SVM on labeled Tweet texts. We use the distance from the SVM's separating hyperplane to measure model uncertainty. We then iteratively label the texts closest to the hyperplane and refit a model until acceptable average precision, recall and F-measure are achieved. To date humans have manually labeled 2,761 Tweet texts, which includes texts from the 998 Tweets in the initial training set. Among the human-labeled texts, 496 are hits and 2,265 are not-hits. Over all unlabeled Tweet texts the SVM classifies 39,663 texts as hits and 126,564 as not-hits. SVM performance measures, based on a weighted cross-validation method,[14] are shown in Table 10. Overall we achieve average precision, recall and F-measure of .75, .75 and .75, respectively.

*** Table 10 about here ***

3.3 Characteristics of Tweet Contents and of Incidents

According to the SVM, both hits and not-hits occur in every state. As Table 11 shows, in every state the number of not-hits exceeds the number of hits. In a few states the discrepancy between the numbers is not large. In Alaska there are 188 hits and 195 not-hits, while in Montana there are 51 hits and 55 not-hits. For California we also have a breakdown of hits versus not-hits by county, as is shown in Table 12. The number of Tweet texts and of hits is largest in Los Angeles, although overall both numbers seem roughly proportional to the population of each county. In every county except Alpine, Inyo and Mono counties the number of not-hits is greater than the number of hits,

[14]Because the number of hits is so much smaller than the number of not-hits, sample sizes for cross-validation are constrained so that the expected number of not-hits sampled is the same as the number of hits.

although the number of Tweet texts in those three counties is extremely small.

*** Table 11 and 12 about here ***

While in the future we plan to use machine classification to classify incidents by type, for the moment we have humans performing such classifications manually, according to the scheme described in Appendix section 5.6.[15] From the Tweet texts with locations in California that the SVM classifies as hits, we selected a simple random sample of $n = 600$ to classify by type manually. Table 13 shows these type frequencies. Among both the unique Tweet texts and the unique Tweets that have those texts (for which $n = 700$), Polling Place Problems are the most frequent type of incident, followed by Improper Outside Influence, Absentee Mail-in or Provisional Ballot Issues, Long Lines/Crowded Polling Place, and Electoral System concerns.

*** Table 13 about here ***

Notable is that human coders decided that 249 of the 600 sampled Tweet texts that the SVM had classified as hits were actually not-hits. A proportion of $.585 = 1 - 249/600$ is smaller than the .75 precision value reported in Table 10. It may be that such a discrepancy reflects variation in classifier performance across states, but in any case it suggests that the number of human-labeled texts should be increased.

Polling Place Problems remain the most frequent type of incident in California when we consider only the subsample of texts from Tweets on election day (June 7, 2016). Table 14 shows the election-day type frequencies. Omitting texts that express positive evaluations of the remarked situation, on election day Absentee Mail-in or Provisional Ballot Issues are second-most frequent in the subsample, while Long Lines/Crowded Polling Place and Improper Outside Influence are tied for third. If the sample size for the comparison between proportions is taken to be $n = 103$, then the proportion of Polling Place Problems among texts that are not Positive ($n = 34$) is significantly greater than the proportion of Absentee Mail-in or Provisional Ballot Issues

[15] Our active learning software is not currently designed to work when each text/document can have more than one label, but a single Tweet can refer to multiple type categories.

(n = 19), but the proportion of Absentee Mail-in or Provisional Ballot Issues is not significantly greater than the proportion of Long Lines/Crowded Polling Place or Improper Outside Influence incidents (n = 14).

*** Table 14 about here ***

3.4 Comparisons to the California Hotline

On primary election day in 2016 California operated a statewide voter hotline (Plummer 2016). The distribution of complaints recorded by hotline operators appears in Table 15. Because no codebook for the California categories is available to explain their meaning,[16] it is difficult to say how the distribution of hotline complaints compares to the distribution of election-day Tweet texts presented in Table 14. Nonetheless Poll Worker Problem alone is the most frequent hotline complaint, Polling Location is the second most frequent and Closed Polling Place is fifth. Perhaps those frequencies are a match for Polling Place Problems being the most frequent type of incident in the election-day Tweet texts. Voter Registration concerns are 11.4 percent of hotline complaints but Registration Issues describe less than five percent of election-day Tweet texts. Provisional Voting and Vote by Mail Ballot together are less than five percent of hotline complaints (Voting Process Issue complaints are another 3.9 percent), while Absentee Mail-in or Provisional Ballot Issues are 18.4 percent of election-day Tweet texts that are not Positive. On the whole there are many differences between the hotline complaints distribution and the distribution of election-day incidents that Tweet texts point to, but the distributions are not utterly unlike one another.

*** Table 15 about here ***

An important difference between the hotline complaints and the election-day Tweet text data is the latter have more extensive geographic coverage across the state. Table 16 shows that hotline complaints come from 31 counties, with most complaints coming from Los Angeles and other large population counties. A pattern in which large population counties have the most

[16]Codings were left to the discretion of the individual hotline operators (Pancharian 2016).

observations also occurred for the Tweet texts that are hits, as shown in Table 12 for a time period that includes but is not restricted to election day. Table 17 shows that on election day Tweet texts that are classified as hits occur in 41 counties as well as in the "Bay Area" (which includes "East Bay") and in "Silicon Valley" (without reference to a particular county). The tendency for more hits to occur in more populous counties continues to occur.

*** Tables 16 and 17 about here ***

Not all the instances classified as hits will prove to be hits on closer inspection—recall that only 58.5 percent of classified hits proved to be hits upon examination by a human (59.3 percent in Table 14, for election day). But the machine classification performance will very likely improve once a greater number of Tweets are labeled by a human in the active learning process. Even with likely reductions in the number of hits, more incidents and more widely dispersed incidents are likely to be identified by the Twitter data than there are complaints in the hotline data.

4 Discussion

Every indication is that Twitter can be used to develop data containing individuals' observations of how elections are conducted, data that cover the entire country. Observations for each day can be gathered, and observations can be even more finely resolved in time (using the timestamps on Tweets). The frequency and likely the diversity of observations varies depending on how salient an election is, and therefore depending on how many people care about an election and want to participate in it, observe it and comment on it. Some Tweets seem like shouts into the void (although maybe such a view underestimates the importance of "Twitter followers"), but others are messages directed specifically at election officials. One question is whether those two types of Tweets typically convey information about different kinds of election incidents.

An important immediate step for development is to exploit the geolocation information that exists for a small proportion of Tweets. The "location" data that is associated with each Tweet usually reflects the location associated generally with (and chosen by) the sender of the Tweet, not the place whence the Tweet originated. Perhaps in cases where voting happens in person, we can rely on selected locations to correspond both to where the sender lives and to the place where the sender is trying to vote—but clearly such is not a generally reliable assumption. Perhaps geolocation data can be used to develop models to estimate the likelihood that Tweets that do not have geolocation information actually come from the place the "location" indicates. "Location" information is also often vague, which makes it challenging to associate incidents with particular polling places or even particular legislative districts. That presents a challenge for the goal to combine such information with information about votes.

We don't know what observational biases affect the set of incidents measured using Twitter data. An obvious bias is that Tweets come only from individuals with a smartphone who use Twitter, and such individuals may not be be as frequently present at every place from which we would like to observe election incidents. Privacy settings in Twitter also limit the number of tweets we see, and incidence of (for us) adverse settings may vary across time and space. When we rely on Tweets at election officials we may be biasing our data to include more observations from states with high degrees of professionalization in their county governments.

In most cases we cannot know whether purported incidents actually occurred, although in a few cases incidents alleged in Tweets can be verified by information obtained from other channels such as news reports or official reports. Many other questions will arise regarding observations derived from Twitter, but at this point it seems better to get the data then critically appraise them rather than not obtain the data at all.

5 Appendices

5.1 State Election Resources

We are attempting to find and acquire copies of resolved official complaints filed with each state. Although state Help America Vote Act code frequently states that resolved complaints will be made available online to the public, this is rarely the case. In order to facilitate our searches for resources, we went to the official elections (or substantively similar) website for each state and collected what data we could regarding the available official complaint forms for each state, the contact email for each state election director, the mailing address for each state elections division, the phone number for each state elections division, the "voter hotline" phone number for each state as available, and the link to each state's online complaint portal as available.[17] We also obtained the election laws for each state. These data were collected over the span of the past year, and the accuracy of these data relies upon each state's official elections website. As is suggested by Table 3, we have collected many state-level resources for each state and the District of Columbia.[18]

We contacted states by reaching out to the state election division directors (or equivalent role) directly. We avoided general public e-mail accounts or hotlines. The e-mails of the state election directors were found in a publicly available National Association of State Election Directors (NASED) roster, last revised January 16, 2016. Below is an example of a generic e-mail sent out. E-mails were modified with links to state complaint forms (where available), state HAVA complaint forms (where available), and always referenced the publicly available state complaints of the State of Colorado. E-mails were almost always sent during the mornings or early afternoons on weekdays in an attempt to increase the probability that they will be seen by the state elections director. Any requests for formal FOIA requests were completed as well.

> Subject: [State Name] Voter Complaints Data
>
> Dear [Elections Director Name Here],
>
> I am a graduate student at the University of Michigan's Department of Political Science gathering data related to election administration in the United States. Specifically, I am collecting data on election and voter complaints. I was wondering if the State of [State Name Here] has any collection of voter complaints available on its website,

[17] The most recent update to the state election directors contact information was found through the National Association of State Election Directors at `https://www.nased.org/`.

[18] We plan to build an elections resources website to publicize the information we have collected, to be made available for the 2016 general elections.

similar to what the State of Colorado currently has (`http://www.sos.state.co.us/pubs/elections/complaints/index.html`).

If not, I am interested in receiving copies of complaints that were either submitted through state-complaint processes ([Link to state-complaint form]) or complaints that were filed as HAVA complaints ([Link to HAVA complaint form]). I am interested in all years where election complaints are available.

Thank you, in advance, for your assistance with my research project. I look forward to hearing from you very soon.

Best,

5.2 Twitter API Data

In order to use the Twitter API (Twitter, Inc. 2016*b*) to collect Tweets to and from election officials on and around the respective Election Days, we first had to find the Twitter accounts for those election officials. These Twitter accounts were found in two ways: first, the Election Assistance Commission has collected information regarding the social media accounts of election officials at both the state and county levels across the United States, with varying degrees of completeness of data across states.[19] The second way these Twitter accounts were obtained was by manually searching Twitter for terms associated with the office of election officials, such as "election official," "county clerk," "department of elections," and "county auditor." Along with manually searching for election officials, user-created lists of election officials were searched for previously not-found election officials.[20] We used similar methods to find the Twitter accounts of state-level Republican and Democratic Parties, state-level Leagues of Women Voters, and state-level ACLUs. In order to facilitate these searches, we created a Twitter account affiliated with this research project, and we will continue to use this Twitter account to improve our

[19]The list of resources can be found at `http://www.eac.gov/voter_resources/state_and_local_election_office_social_media_list.aspx`.

[20]An example of one of these user-created lists can be found at `https://twitter.com/EACgov/lists/us-election-officials/members`.

collection of the Twitter accounts of election officials.[21]

To access the API, we registered an application with Twitter.com, giving us the security tokens necessary to query data from Twitter's database. Our goal was to pull entire timelines from 493 accounts (for perspective, one California account had over eleven thousand Tweets in their timeline).

A few challenges arose in querying that much data. First, user timelines are not static: a user can post Tweets while our application queries the data, which would effect the results; we had to recursively pull Tweets twenty at a time, starting with the user's most recent Tweet and ending with the first Tweet posted (in some cases dating back to 2007). Second, the sheer size of the query would occasionally break the script, so we had to pull timelines in batches; that is, we could not pull all 493 accounts at once, but rather, pull them fifty at a time. For perspective, a single batch would return hundreds of thousands of Tweets. Finally, Twitter places rate limits[22] on applications that query data from the API, so we had to design the script to pause in between requests. This way, we would not exceed rate limits, and the script could complete each query.

The data returned are formatted in JSON[23], so we had to identify the specfic fields of interest (in this case, the unique identification number of each Tweet, its content, its timestamp, and the name and location of each user) and write them to a .csv file. Additionally, we were interested in obtaining geolocation data from each Tweet (returned in the form of coordinates) but Twitter's privacy settings are such that, this kind of data is not readily available for most users. The bulk of the content was from outside of our time range, so it was not used. We made sure that the data used from the Twitter API were from the same time frame as the data obtained via Sysomos. Part of the data collection was to identify Tweets by their unique identification number, allowing us to quickly identify and omit duplicate Tweets from our final dataset.

[21] The Twitter user name for this account can be found at `https://twitter.com/election_ballot`.

[22] Enforced on a "per access token" basis, Twitter limits users to fifteen requests per fifteen-minute window, although this number varies with the object being called; for more information on Twitter rate limits, see: `https://dev.twitter.com/rest/public/rate-limiting`.

[23] JavaScript Object Notation, a data format represented by simple text, used to transfer data objects that consist of atrribute-value pairs; for more information on the format of Twitter data, see: `https://dev.twitter.com/rest/reference/get/search/tweets`.

5.3 Stratified Sampling for Tweets in Training Set

The stratfied random sample used for the initial training set contained 1,001 Tweets of which n = 998 are unique Tweets (unique based on the 18-digit Twitter ID number). The population used for sampling was the union of the distinct samples drawn previously for use in developing the coding schemes. Strata were defined by state, by type of search terms used to find Tweets and by whether any human identified the Tweet as a hit in the initial round of coding (that is, before the flowchart of Figures 1 and 2 existed). The stratum labels derived from state and search terms are AZ, CA, CT, CO, WAd, WAr, CAeo, COeo and WAeo, where the first two letters are a state's postal code, "WAd" refers to the Democratic caucus, "WAr" refers to the Republican primary and the "eo" suffix means search terms focused on election officials. Table 18 shows the number of Tweets in each of the state-term strata in the full set of Tweets manually downloaded from Sysomos, as well as the breakdown by hit-or-not-hit preliminary classification. Because the hit strata are much smaller than the not-a-hit strata, sampling was weighted to include approximately 30 percent hits and 70 percent not-hits, with a minimum of two observations in the sample from each of the 18 strata. Stratum sample sizes appear in Table 18.

*** Table 18 about here ***

5.4 Image Classification

Images were often included as a part of the Tweet. Many of these Tweets with images often imply a complaint with the image. For instance, there are a large number of Tweets with images of crowded caucus locations or long polling lines. In order to classify these Tweets as hits, we uploaded the direct links to these images on Google's Reverse Image Search and classified the Tweet as a hit or not based on Google's guess of the subject of the image. Google's Reverse Image Search uses a proprietary image identification system that utilizes a mix of pre-existing images online and the contents of webpages containing similar images. At the moment, we have completed image classification of all images in our training set as well as a random sample of 500

Tweets in the non-training set data.

5.5 Flowchart Development

The hits flowchart (Figures 1 and 2) was developed over the course of several individual handcoding sessions. Tweets with three or more agreements as "hits" (among five coders) were designated core Tweets; a random sample of Tweets with two or fewer agreements as "hits" were reviewed and collectively discussed. After the discussion, we used both the core Tweets and the discussion of the marginal Tweets to create what we call the "hits flowchart." The flowchart was developed to standardize hits classification among the authors and avoided a simple definitional basis for classifying hits. The first half of the hits flowchart lays out what a hit is *not* (for instance, a hit is *not* an endorsement of a candidate); the second half of the hits flowchart engages with the substantive content of the Tweet and classifies the Tweet as a hit or not. This flowchart was used to create the training set, and coders currently use the flowchart to engage with the Tweets given by the active learning framework.

5.6 Coding Scheme for Tweets

Updated 8/21/16 (Version 5)

5.6.1 Instructions

After deciding whether the Tweet in question is a "hit" or not according to the flowchart, use the categories listed below to classify that hit. These categories and definitions also may help decide if a Tweet is a hit or not, if you are having trouble. A Tweet can be appropriately classified into multiple categories. For example, a Tweet that reads "For some reason there was a problem with my voter registration, but the workers at my polling place were very helpful!" would fall within the `registration problems` category and the `positive` category.

> 0 or blank: Tweet does not fit within this category
> 1: Tweet fits within this category

5.6.2 Categories for Coding

1. Absentee, Mail-In, or Provisional Ballot Issue: This category applies to hits relating to problems with absentee or mail-in ballots, including ballots not being received by the voter or ballots not being counted. This category also applies to incidents relating to provisional ballots, such as a voter having to vote provisionally (or not being allowed to). This category corresponds to the "provisional ballot abuse" and "Non-receipt of requested absentee ballots" EIRS category.

2. Other Ballot Problems: This category includes complaints or incidents regarding the design of the ballot, including layout and foldability. This category also applies to individuals being given the incorrect ballot, as well as a voter's preferred candidate or party not appearing on the ballot. This category corresponds to the other ballot problems EIRS category.

3. Disability/Accessibility Problem: Tweets that fall under this category would include complaints or observations about some aspect of the election that is not accessible for those with disabilities–for example, a polling place not offering special ballots or assistance to voters who are blind, or a polling place not being wheelchair accessible. This corresponds to the lack of disability access EIRS category.

4. Election Official Complaints/Incidents: Complaints that accuse governmental, election workers (including poll workers), or election officials of corruption, malfeasance, ignorance, being unhelpful or non-responsive, being rude, or some other complaint. This includes allegations of mis-managing the election. A Tweet that falls in this category and the positive category might not that a pollworker or election official was helpful, or the staff managed the polling place well. This category is analogous to the EIRS categories for 'pollworker malfeasance/ineptitude' as well as "other election worker problem."

5. Electoral System: This includes complaints relating to the specific aspects of the American electoral system, such as the first-past-the-post system, top-two electoral systems, caucuses, or open/closed primary elections. This category also includes complaints or hits that do not criticize a specific aspect of the American electoral system such as non-proportional representation. This also includes complaints about improper district boundaries and gerrymandering.

6. Improper Outside Influence: This category includes cases where the complainant encountered improper campaign advertising, such as advertising to close to a polling place. This category also includes complaints or observations alleging candidates', parties', or outside entities such as PACs' campaign practices violate the spirit or letter of the law. Also included in this category are allegations of police misconduct relating to the administration or outcome of the election, as well as complaints or incidents regarding the media. For example, an individual might complaint that the media called the election while people were still in line to vote, or reporters may be improperly interviewing voters. This category is in part analogous to the "Improper Outside Influence" EIRS category.

7. Long Lines/Crowded Polling Place: This category refers to a complaint, incident, or report that states a long line or crowded polling place, including statements about the polling place

being too small. Other examples of this category include a person referencing how long they have had to wait to vote, or reporting that their caucus has been moved outside due to crowding. This corresponds to the "polling place chaos and crowding" EIRS category.

8. Polling Place Problems: This category includes problems or incidents related to the polling place, such as the set-up of the voting booths and other election structures. Another example of a problem that would fit in this category is the presence of security cameras observing how individuals vote. Furthermore, this category includes voters being told an incorrect location for their polling place or precinct line. Finally, this category includes complaints or reports that allege intimidation by polling place officials or other persons (non-police) that occurred while the relevant person was casting his or her ballot, approaching the polling place, or in the polling place. This category does not include corruption, malfeasance, impropriety, or other comments regarding poll workers. It partially corresponds to the "Incorrect polling place/precinct information" and "Voter Intimidation" EIRS categories.

9. Registration Issues: Voters or prospective voters encountered difficulty registering to vote or had problems registering with their preferred party. It could also include instances of registration records being incorrect. This corresponds to "Incorrect registration lists/non-receipt of registration cards" EIRS category.

10. Voter Fraud: This category refers to instances or alleged instances of voter fraud, including a voter being told that he or she has already voted. This category is analogous to EIRS category "Voter fraud."

11. Voter Identification Issues: The voter or prospective voter had issues relating to voter identification requirements. This might include an election official improperly asking for identification, problems acquiring identification, or being rejected at the polls due to lack (or accused lack) of necessary identification. This corresponds to the "Improper ID requirements" EIRS category.

12. Voting Machine Complaints: This category includes voting machines being inoperable as well as unclear instructions regarding how to use the voting machines. Examples could include machines misreading scanned ballots, not printing receipts, or machines being difficult to use. This category is similar ot the "Machine malfunction/usage problem" EIRS category.

13. Unspecified Other: Includes complaints of which the nature is unclear as well as non-sequitur complaints. Analogous to the EIRS category "Other."

14. Positive: This category indicates that the complaint or incident was positive in nature: for example, complimenting an election official on being helpful, or there not being a long line to vote. In the latter case, it is appropriate to both mark the "Long Lines" category and the "Positive" category.

15. Ambiguous: This category notes that the wording of a Tweet or complaint is unclear and it is not possible to ascertain if it is complaint or hit. As such, it warrants further examination. For example, a Tweet might be worded such that it could be taken as a joke or as a serious comment on the election system, depending on the reader.

16. Not Hit: For the purposes of coding the machine-coded hits, mark this category if the Tweet in question is not a hit (that is, it was mistakenly defined as a hit by the machine classification algorithm).

References

Beber, Bernd and Alexandra Scacco. 2012. "What the Numbers Say: A Digit-Based Test for Election Fraud." *Political Analysis* 20(2):211–234.

Breunig, Christian and Achim Goerres. 2011. "Searching for Electoral Irregularities in an Established Democracy: Applying Benford's Law Tests to Bundestag Elections in Unified Germany." *Electoral Studies* 30(3):534–545.

Cantu, Francisco and Sebastián M. Saiegh. 2011. "Fraudulent Democracy? An Analysis of Argentina's Infamous Decade Using Supervised Machine Learning." *Political Analysis* 19(4):409–433.

Hall, Joseph Lorenzo. 2005. "Towards Work-Practice Modeling of Elections and the Election Incident Reporting System (EIRS)." URL `https://josephhall.org/papers/eirs_brahms_mims.pdf` (accessed July 5, 2006).

Johnson, Jeff. 2005. "The EIRS 2004 and 2005 Voting Problem Incident Data: A Valuable Public Resource." URL `http://triton.towson.edu/~jlazar/EIRS_factsheet.doc` (accessed July 10, 2014).

Lan, Man, Chew-Lim Tan, Hwee-Boon Low and Sam-Yuan Sung. 2005. A Comprehensive Comparative Study on Term Weighting Schemes for Text Categorization with Support Vector Machines. In *Special interest tracks and posters of the 14th international conference on World Wide Web*. ACM pp. 1032–1033.

Leopold, Edda and Jörg Kindermann. 2002. "Text Categorization with Support Vector Machines. How to Represent Texts in Input Space?" *Machine Learning* 46:423–444.

Levin, Inès, Gabe A. Cohn, Peter C. Ordeshook and R. Michael Alvarez. 2009. "Detecting Voter Fraud in an Electronic Voting Context: An Analysis of the Unlimited Reelection Vote in Venezuela." 2009 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections, Montreal, August 10–11.

Mebane, Jr., Walter R. 2010. "Fraud in the 2009 Presidential Election in Iran?" *Chance* 23:6–15.

Mebane, Jr., Walter R. 2013. "Election Forensics: The Meanings of Precinct Vote Counts' Second

Digits." Paper presented at the 2013 Summer Meeting of the Political Methodology Society, University of Virginia, July 18–20, 2013.

Mebane, Jr., Walter R. 2014. Can Votes Counts' Digits and Benford's Law Diagnose Elections? In *The Theory and Applications of Benford's Law*, ed. Steven J. Miller. Princeton: Princeton University Press pp. 206–216.

Mebane, Jr., Walter R. 2016. "Election Forensics: Frauds Tests and Observation-level Frauds Probabilities." Paper presented at the 2016 Annual Meeting of the Midwest Political Science Association, Chicago, April 7–10, 2016.

Mebane, Jr., Walter R. and Jonathan Wall. 2015. "Election Frauds, Postelection Legal Challenges and Geography in Mexico." Paper presented at the 2015 Annual Meeting of the American Political Science Association, San Francisco, CA, September 3–6.

Mebane, Jr., Walter R. and Joseph Klaver. 2015. "Election Forensics: Strategies versus Election Frauds in Germany." Paper presented at the 2015 Annual Conference of the European Political Science Association, Vienna, Austria, June 25–27, 2015.

Mebane, Jr., Walter R., Joseph Klaver and Blake Miller. 2016. "Frauds, Strategies and Complaints in Germany." Paper presented at the 2016 Annual Meeting of the Midwest Political Science Association, Chicago, April 7–10, 2016.

Miller, Blake and Joseph Klaver. 2016. "Active Learning for Costly Text Classification Problems and Exploratory Data Analysis." Unpublished paper draft.

Montgomery, Jacob M., Santiago Olivella, Joshua D. Potter and Brian F. Crisp. 2015. "An Informed Forensics Approach to Detecting Vote Irregularities." *Political Analysis* forthcoming.

Myagkov, Mikhail, Peter C. Ordeshook and Dimitry Shaikin. 2009. *The Forensics of Election Fraud: With Applications to Russia and Ukraine*. New York: Cambridge University Press.

Pancharian, Nick. 2016. "personal communication." phone conversation on August 22, 2016, with Nick Pancharian, Constituent Affairs, Secretary of State.

Pericchi, Luis Raúl and David Torres. 2011. "Quick Anomaly Detection by the Newcomb-Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela."

Statistical Science 26(4):502–516.

Plummer, Mary. 2016. "Numbers are in—poll workers top list of California voter complaints." *89.2 KPCC* . July 12, Internet Edition.

Rakotomamonjy, Alain. 2003. "Variable Selection Using SVM-based Criteria." *Journal of Machine Learning Research* 3(Mar):1357–1370.

Secretary of State, Constituent Affairs. 2015. "RE: HAVA Complaint Data Availability." Email message received November 19, 2015.

Secretary of State, Constituent Affairs. 2016. "Data Relating to the California Voter Hotline." File `CA Election Day Complaints Report _060716.xls`, August 9, 2016.

Settles, Burr. 2010. "Active learning literature survey." from Computer Sciences Technical Report 1648 `http://burrsettles.com/pub/settles.activelearning.pdf`.

Sysomos. 2016. "Social Technology: Turning Insights Into Action ... Before Anyone Else." URL `https://sysomos.com/`.

Twitter, Inc. 2016*a*. "Documentation." URL `https://dev.twitter.com/overview/documentation`.

Twitter, Inc. 2016*b*. "REST APIs." URL `https://dev.twitter.com/rest/public`.

Verified Voting Foundation. 2005. "Election Incident Reporting System." URL `https://voteprotect.org/index.php?display=EIRMapNation&tab=ED04` (accessed July 5, 2006).

Ziblatt, Daniel. 2009. "Shaping Democratic Practice and the Causes of Electoral Fraud: The Case of Nineteenth-Century Germany." *American Political Science Review* 103(1):1–21.

Table 1: Finite Mixture Model Parameter Estimates

Election	$\hat{f}_{\mathrm{i}}$	$\hat{f}_{\mathrm{e}}$	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\tau}$	$\hat{\nu}$	LR	n
Arizona 2006:								
Governor	.0024	0	1.6	.40	.58	.63	44.7	2,196
Atty Genl	.028	0	1.5	.35	.56	.62	233.6	2,196
US Senate	.043	0	1.3	.35	.57	.50	30 .2	2,196
California 2006:								
Governor	.097	0	.9	.32	.48	.55	11364.6	22,820
LT Governor	.072	0	1.7	.44	.48	.41	8444.2	22,820
Atty Genl	.073	0	1.7	.35	.48	.53	7704.2	22,819
Prop 1A	.016	.00011	3.3	.53	.48	.71	2323.2	22,806
Prop 1B	.040	.000001	1.7	.35	.48	.57	5143.4	22,817
Prop 1C	.070	0	1.7	.44	.48	.54	7643.8	22,813
Prop 1D	.066	0	1.7	.35	.48	.54	7442.2	22,814
Prop 1E	.046	.000046	1.7	.53	.48	.59	5856.6	22,814
Texas 2006:								
Governor	.0058	.0025	3.4	.49	.35	.42	1444.7	8,421
LT Governor	.022	.000067	1.7	.27	.37	.59	1497.5	8,429
Atty Genl	.025	.000003	1.7	.27	.39	.59	839.4	8,431
Comm of Ag	.031	.000004	1.7	.27	.38	.53	752.9	8,430
Comptroller	.013	.00012	1.7	.27	.38	.60	1151.1	8,431
US Senate	.038	.000001	1.7	.27	.39	.62	653.4	8,432

Note: LR is the likelihood ratio test statistic for the hypothesis that there are no frauds (i.e., that $f_{\mathrm{i}} = f_{\mathrm{e}} = 0$). n is the number of precinct observations.
Source: estimates using the model of Mebane (2016) with precinct data from the referent elections. For California, each absentee precinct is combined with its corresponding in-person precinct.

Table 2: Frequency of Postelection Complaint Types, Germany 2005 & 2009

Description	2005	2009
Absentee-ballot Related Problem	48	19
Electoral System	83	66
Polling Place Problem	24	17
Allegations of Official Corruption	8	3
Ballot Related Problem	6	3
Counting of the Votes	9	8
Criminal Status Related Problem	5	4
Disability Access Problem	2	2
Identification Related Problem	6	13
Improper Campaigning Influence	11	15
Improper District Boundaries	1	2
Improper Statistics	4	8
Party List Not on Ballot	21	26
Problems with the Creation of Party Lists	2	10
Registration Related Problem	26	6
Unspecified Other	10	5
Police Harassment	1	0
Voter Intimidation	1	0
Briefwahl in Dortmund	22	0
Nachwahl in Dresden	37	0

Note: Number of complaints each type. Each document can contain up to six types of complaints. Source: Compiled from archives of the Bundestag's website for the *Ausschuss für Wahlprüfung, Immunität und Geschäftsordnung*. Types are described in Mebane and Klaver (2015).

Table 3: Election Complaint Modes and Online Availability in Each State

		Modes						Modes			
State	Initial Contact[a]	HAVA Form	Other Form	Online Portal	Complaints Online	State	Initial Contact	HAVA Form	Other Form	Online Portal	Complaints Online
AL	7/21	Y	Y	Y		MT	5/13	Y[c]	Y[c]		Y
AK	7/21	Y			Y	NE	5/5	Y			
AZ	11/16[b]	Y				NV	4/29		Y		
AR	7/21		Y			NH	5/5	Y[c]	Y[c]		
CA	10/26[b]	Y	Y		Y	NJ	5/13				Y
CO	11/16[b]	Y	Y		Y	NM	5/2	Y[c]	Y[c]		
CT	11/16[b]		Y		Y	NY	5/9				
DE	7/21	Y			Y	NC	5/5	Y	Y		
FL	7/21	Y	Y		Y	ND	5/4	Y			
GA	7/21		Y	Y	Y	OH	5/13	Y	Y		Y
HI	5/5		Y			OK	5/4				
ID	3/24	Y				OR	3/24				
IL	5/10	Y	Y			PA	5/5	Y		Y	
IN	5/5	Y	Y			RI	5/5				
IA	5/4	Y	Y	Y		SC	5/9	Y			
KS	11/16[b]	Y	Y			SD	5/4				
KY	5/4	Y				TN	5/9	Y	Y		
LA	5/4	Y				TX	5/4	Y	Y		
ME	5/5	Y				UT	5/2				
MD	5/9	Y			Y	VT	5/9				
MA	5/5					VA	7/21	Y[c]	Y[c]	Y	
MI	5/13				Y	WA	3/24	Y		Y	Y
MN	5/13	Y			Y	WV	5/9	Y[c]	Y[c]		
MS	5/4	Y				WI	5/9			Y	
MO	5/13	Y	Y		Y	WY	5/4	Y[c]	Y[c]		

Note: An "Online Portal" is a dedicated online way to submit complaints; "Complaints Online" refers to an online repository of complaints available for viewing. [a] dates in 2016; [b] date in 2015; [c] dual HAVA and state forms.

Table 4: Number of HAVA Complaints from Each State

State	# HAVA Complaints
Illinois	4
Louisiana	1
New Hampshire	6
Ohio	1
Wyoming	4

Table 5: State Complaints Received

State	Years
Colorado	2013, 2014
Connecticut	1991, 1994, 2000, 2006, 2007–2016
Delaware	2009–2015
Florida	1993–2016
Indiana	2006–2007
Montana	1990–2016
New Hampshire	2012, 2013, 2014, 2015
New Jersey	1993, 1995–2015
New Mexico	2013–2016
Ohio	2000–2016
Washington	2011, 2012
Wyoming	2004, 2014

Table 6: State Voter Hotlines

Hotline	States
Yes	AZ, CA, CT, FL, GA, IL, IN, IA, KS, KY, LA, ME, MD, MS, MO, MT, NE, NJ, NC, TN, VT, VA
No	AL, AK, AR, CO, DE, HI, ID, MA, MI, MN, NV, NH, NM, NY, ND, OH, OK, OR, PA, RI, SC, SD, TX, UT, WA, WV, WI, WY

Table 7: Search Terms Used for Particular States with Sysomos

Arizona eleccion, spanishterms, eleccionprimaria, campaignfinance, corruption, azprimary, azvote, disenfranchised, election fraud, electionfraud, electionofficials, electionstealing, generalelection, linetovote, pollingplace, primaryelection, problemsvoting, righttovote, spanishtweets, voteaz, voterfraud, voterid, voteridentification, votingrights

California eleccion, spanishterms, eleccionprimaria, linetovote, caprimary, corruption, brokenvotingmachine, caelection, campaignfinance, cavote, disenfranchised, electioncomplaint, #electionfraud, electionfraud, electionstealing, generalelection, pollingplace, primaryelection, problemsvoting, righttovote, stoleelection, voteca, voterfraud, voterid, votingrights, caprimary, longlinetovote, caprimary, caprimaryANDNOTvote, caprimaryANDvote, caprimary, corruption

Colorado campaignfinance, caucus, corruption, disenfranchised, electionfraud, electionofficials, generalelection, linetovote, longlinetovote, pollingplace, primaryelection, probvoting, righttovote, spanishterms1, spanishterms2, statevote, voterfraud, voterid, voteridentification, votingrights

Connecticut campaignfinance, corruption, ctprimary, ctvote, disenfranchised, electionfraud, electionofficials, electionstealing, generalelection, linetovote, pollingplace, primaryelection, problemsvoting, righttovote, spanishtweets, votect, voterfraud, voterid, voteridentification, votingrights

Illinois campaignfinance, corruption, disenfranchised, electioncomplaint, electionfraud, electionofficials, ilprimary, ilvote, linetovote, longlinetovote, outofballots, pollingplace, primaryelection, problemsvoting, righttovote, stoleelection, twill, voteil, voterfraud, voterid, voteridentification, votersuppression, votingrights

Washington campaignfinance, corruption, disenfranchised, electioncomplaint, electionfraud, electionofficials, electionstealing, generalelection, longlinevote, pollingplace, primaryelection, problemvoting, righttovote, spanishterms, stolenelection, voterfraud, voteridentification, votingrights, wacaucus, wavoteetc

Note: examples of search terms used for a few states in searches using Sysomos within windows of ten days around each election/caucus day or election period (for states with absentee voting). "spanishterms" refers to a collection of election-related terms in Spanish.

Table 8: Search Terms Used to Cover All States with Sysomos

(akprimary, akprimary, akcaucus, akcaucus), (alprimary, to:alasecofstate),
(arprimary, to:Mark_Martin), to:ARSecofState, (azprimary, to:SecretaryReagan),
(caprimary+AND+vote, caprimary+AND+NOT+vote, to:CASOSvote),
(coprimary, cocaucus, to:ColoSecofState, to:juddchoate), (ctprimary, to:SOTSMerrill),
(dcprimary, dccaucus, to:DCBOEE, to:SecretaryofDC, dcprimary), (deprimary, to:SecretaryDE),
(flprimary+AND+vote, flprimary+AND+NOT+vote to:KenDetzner),
(gaprimary+AND+vote, gaprimary+AND+NOT+vote to:BrianKempGA),
(hicaucus, hiprimary, hicaucus, hiprimary),
(iacaucus, iaprimary, to:IowaSOS, to:PateforIowa), (idcaucus, idprimary, idprimary),
(ilprimary+AND+vote, ilprimary+AND+NOT+vote, to:ILSecOfState),
(inprimary, to:SecretaryLawson, to:IndianaSOS),
(kscaucus, ksprimary, to:BACaskey, to:KansasSOS),
(kycaucus, kyprimary, to:KySecofState, kyprimary, to:KySecofState),
(laprimary, to:Louisiana_sos), (maprimary, to:MrVoterReg),
(mdprimary, to:SOSMaryland, to:md_sbe), (mecaucus, meprimary, to:MaineSecOfState),
(miprimary+AND+vote, miprimary+AND+NOT+vote to:MichSoS, to:RJ4MI),
(mncaucus, mnprimary, to:MNSteveSimon, to:MNSecofState),
(moprimary, to:JasonKander, to:MissouriSOS),
(msprimary, to:DelbertHosemann, to:MississippiSOS), (mtprimary, to:SOSMcCulloch),
(ncprimary+AND+vote, ncprimary+AND+NOT+vote, to:Elaine4NC, to:NCSBE),
(ndcaucus, ndprimary, to:VoteND), (necaucus, neprimary, to:NESecJGale), nhprimary,
(njprimary, to:KimGuadognoNJ), nmprimary,
(nvcaucus, nvprimary, to:NVElect, to:NVSOS),
(nyprimary+AND+vote, nyprimary+AND+NOT+vote, to:NYSDOS, to:NYSBOE),
(ohprimary+AND+vote, ohprimary+AND+NOT+vote, to:JonHusted, to:OhioSOSHusted),
(okprimary, to:OKelections), (orprimary, to:oregonelections, to:OregonSoS),
(paprimary+AND+vote, paprimary+AND+NOT+vote, to:PAStateDept),
(riprimary, to:RI_BOE, to:RISecState, to:NellieGorbea),
(scprimary, to:scvotes), (sdprimary, to:shantelkrebs), (tnprimary, to:SecTreHargett),
(txprimary+AND+vote, txprimary+AND+NOT+vote, to:VoteTexas, to:TXsecofstate),
(utcaucus, utprimary, to:ElectionsUtah),
(vaprimary, to:vaELECT), (vtprimary, to:VermontSOS),
(waprimary, wacaucus, to:secstatewa), (wiprimary, to:Wisconsin_GAB, to:DougLaFollette),
(wvprimary, to:NatalieTennant),
(wycaucus, wyprimary, wycaucus, wyprimary, to:EdMurrayforWyo)

Note: search terms used in searches using Sysomos within windows of ten days around each election/caucus day or election period (for states with absentee voting). Parentheses group terms bearing on the same state.

Table 9: Number of Tweets by State

State	Count	State	Count	State	Count
AZ	11,212	KS	1,577	NV	3,379
CA	60,350	KY	2,793	NY	14,982
CO	10,187	LA	243	OH	1,791
CT	4,561	MA	962	OK	265
WA	15,599	MD	1,684	OR	312
IL	19,252	ME	321	PA	5,625
AK	411	MI	1,675	RI	562
AL	120	MN	3,144	SC	7,715
AR	192	MO	453	SD	25
DC	903	MS	325	TN	514
DE	118	MT	110	TX	3,004
FL	5,831	NC	1,264	UT	481
GA	716	ND	28	VA	939
HI	734	NE	428	VT	122
IA	12,593	NH	4,147	WI	6,250
ID	452	NJ	315	WV	1,294
IN	4,494	NM	204	WY	52

Note: Number of unique Tweets (excluding retweets) by State. For AZ, CA, CO, CT, WA and IL the larger sets of search terms shown in Table 7 were used.

Table 10: Machine Classifier (Support Vector Machine) Performance

Class	**Precision**	**Recall**	**F-Measure**	**Support**
Not a hit	.76	.77	.76	77
Hit	.75	.74	.74	72
Average/Total	.75	.75	.75	109

Table 11: “Hit” Classification of Tweets by State

Unique Tweet Texts[a]						All Tweets[b]					
	Hit?			Hit?			Hit?			Hit?	
State	no	yes	State	no	yes	State	no	yes	State	no	yes
AZ	4,989	3,583	MS	227	27	AZ	6,693	4,039	MS	258	28
CA	35,733	9,037	MT	55	51	CA	43,003	10,058	MT	55	51
CO	4,344	3,877	NC	891	210	CO	4,989	4,197	NC	922	212
CT	2,564	574	ND	18	9	CT	3,172	707	ND	18	9
WA	8,414	3,531	NE	262	128	WA	11,626	3,973	NE	263	129
IL	11,032	3,186	NH	3,129	441	IL	14,050	3,390	NH	3,267	452
AK	195	188	NJ	225	48	AK	196	188	NJ	233	48
AL	64	8	NM	137	36	AL	97	9	NM	143	36
AR	136	50	NV	1,991	831	AR	136	50	NV	2,117	857
DC	572	219	NY	10,058	1,785	DC	589	219	NY	10,460	1,809
DE	83	13	OH	1,199	484	DE	103	13	OH	1,203	485
FL	2,887	357	OK	196	42	FL	3,307	368	OK	196	43
GA	468	159	OR	225	45	GA	468	159	OR	229	45
HI	364	245	PA	4,083	793	HI	372	245	PA	4,172	796
IA	8,132	3,174	RI	373	123	IA	8,349	3,196	RI	391	127
ID	265	163	SC	5,532	767	ID	270	164	SC	5,948	785
IN	2,962	350	SD	16	5	IN	3,583	365	SD	16	5
KS	893	573	TN	388	115	KS	895	573	TN	389	115
KY	1,751	820	TX	2,159	393	KY	1,770	824	TX	2,208	396
LA	179	31	UT	396	73	LA	179	31	UT	396	73
MA	714	178	VA	659	229	MA	726	178	VA	661	229
MD	1,219	310	VT	87	34	MD	1,244	310	VT	87	34
ME	205	94	WI	4,185	1,065	ME	207	97	WI	4,499	1,081
MI	1,229	209	WV	935	224	MI	1,261	211	WV	938	225
MN	1,648	1,185	WY	40	7	MN	1,676	1,193	WY	40	7
MO	317	80				MO	318	80			

Note: Number of Tweets (excluding retweets) classified as “hits” by State. [a] Counts using the unique texts across all Tweets. [b] Counts using all unique (by 18-digit ID code) Tweets.

Table 12: "Hit" Classification of California Tweets by County

County	Hit? no	Hit? yes	County	Hit? no	Hit? yes
Alameda	1,200	323	Placer	117	46
Alpine	2	2	Plumas	6	3
Amador	9	3	Riverside	372	97
Butte	84	20	Sacramento	939	287
Calaveras	2	0	San Benito	22	6
Colusa	2	1	San Bernardino	352	99
Contra Costa	340	106	San Diego	2,781	674
Del Norte	2	1	San Francisco	3,926	873
El Dorado	97	9	San Joaquin	149	43
Fresno	228	52	San Luis Obispo	111	61
Glenn	15	3	San Mateo	223	37
Humboldt	103	23	Santa Barbara	179	38
Imperial	28	6	Santa Clara	660	179
Inyo	0	1	Santa Cruz	110	42
Kern	112	24	Shasta	49	18
Kings	3	0	Siskiyou	10	2
Lake	31	15	Solano	58	12
Lassen	5	2	Sonoma	184	31
Los Angeles	12,035	3,271	Stanislaus	91	19
Madera	13	8	Sutter	7	5
Marin	191	21	Tehama	20	6
Mariposa	2	0	Trinity	2	1
Mendocino	16	2	Tulare	56	15
Merced	27	6	Tuolumne	8	3
Modoc	1	0	Ventura	191	48
Mono	0	1	Yolo	219	65
Monterey	103	41	Yuba	1	0
Napa	22	3	Bay Area	265	131
Nevada	16	2	Silicon Valley	131	28
Orange	981	225			

Note: Number of Tweets (excluding retweets) classified as "hits" by county in California. Counts using the unique texts across all Tweets. "Bay Area" and "Silicon Valley" locations, which span multiple counties, are also shown.

Table 13: Frequency of Incidents by Type in Sample of California "Hits"

Type	Unique Tweet Texts[a] Count	Percent	All Tweets[b] Count	Percent
Absentee Mail-in or Provisional Ballot Issue	44	7.3	49	12.7
Ballot Problems	10	1.7	12	3.1
Disability/Accessibility	0	0.0	0	0.0
Election Official Complaints/Incidents	18	3.0	20	5.2
Electoral System	38	6.3	40	10.4
Improper Outside Influence	51	8.5	54	14.0
Long Lines/Crowded Polling Place	43	7.2	47	12.2
Polling Place Problems	64	10.7	66	17.1
Registration Issues	14	2.3	19	4.9
Voter Fraud	31	5.2	36	9.3
Voter Identification Issues	10	1.7	11	2.8
Voting Machine Complaints	4	.7	4	1.0
Unspecified Other	4	.7	4	1.0
Positive	34	5.7	37	9.6
Ambiguous	29	4.8	33	8.5
Not an Incident	249	—	314	—

Note: "Count" shows the number of sampled Tweets that are of the indicated type, and "Percent" shows the percentage of the 351 Tweet texts (or 386 Tweets) that refer to an incident that are of the indicated type.
The sample (n = 600) is of Tweets drawn from all California Tweets classified as "hits" either by a human or by a machine classification algorithm (n = 15 Tweets in the sample are human-coded as incidents). All Tweets are associated with California based on "California" (or a synonym) being included in search terms or by the `location` in the Tweet mentioning a place in California. Coding by type is performed directly by humans.
[a] Counts using a sample of the unique texts in California Tweets (n = 600). [b] Counts using all replicas of sampled Tweet texts in California (n = 700).

Table 14: California Election-day Incidents by Type, Tweet Sample

Type	Unique Tweet Texts[a] Count	Unique Tweet Texts[a] Percent	Omit Positive[b] Count	Omit Positive[b] Percent
Absentee Mail-in or Provisional Ballot Issue	19	15.7	19	18.4
Ballot Problems	1	0.8	1	1.0
Disability/Accessibility	0	0.0	0	0.0
Election Official Complaints/Incidents	12	9.9	12	11.7
Electoral System	8	6.6	8	7.8
Improper Outside Influence	14	11.6	14	13.6
Long Lines/Crowded Polling Place	18	14.9	14	13.6
Polling Place Problems	36	29.8	34	33.0
Registration Issues	5	4.1	5	4.9
Voter Fraud	2	1.7	2	1.9
Voter Identification Issues	0	0.0	0	0.0
Voting Machine Complaints	1	0.8	1	1.0
Unspecified Other	3	2.5	3	2.9
Positive	18	14.9	—	—
Ambiguous	2	1.7	2	1.9
Not an Incident	83	—	83	—

Note: "Count" shows the number of sampled Tweets that are of the indicated type, and "Percent" shows the percentage of the 351 Tweets that refer to an incident that are of the indicated type. The sample (n = 600) is of Tweets drawn from all California Tweets classified as "hits" either by a human or by a machine classification algorithm. This table shows only the subsample of election-day Tweets (n = 121). All Tweets are associated with California based on "California" (or a synonym) being included in search terms or by the `location` in the Tweet mentioning a place in California. Coding by type is performed directly by humans. [a] Counts using a sample of the unique texts in California Tweets (n = 121). [b] Omitting "positive" Tweets.

Table 15: California Election-day Hotline Complaints

Type	Count	Percent
Closed Polling Place	36	6.3
Electioneering	6	1.1
ID Issue	9	1.6
Other	10	1.8
Poll Worker Problem	234	41.1
Polling Location	71	12.5
Provisional Voting	17	3.0
SOS Election Day Observer Allegation	2	0.4
Vote by Mail Ballot	9	1.6
Voter Registration	65	11.4
Voting Materials	23	4.0
Voting Process Issue	22	3.9
Voting System Equipment	66	11.6

Note: "Count" denotes the number of complaints of a given type submitted to the hotline. The category of complaint was determined on a case-by-case basis by the individual hotline operators. Source: Secretary of State, Constituent Affairs (2016)

Table 16: California Election-day Hotline by County

County	Count	County	Count
Alameda	14	Sacramento	21
Butte	1	San Bernardino	14
Colusa	1	San Diego	25
Contra Costa	17	San Francisco	10
Fresno	7	San Joaquin	1
Humboldt	2	San Mateo	1
Imperial	1	Santa Barbara	1
Kern	3	Santa Clara	9
Kings	1	Santa Cruz	1
Los Angeles	367	Solano	5
Madera	1	Sonoma	3
Marin	1	Tulare	1
Mendocino	1	Tuolumne	1
Napa	1	Ventura	7
Orange	19	Yolo	5
Riverside	28		

Source: Secretary of State, Constituent Affairs (2016)

Table 17: "Hit" Classification of California Election-Day Tweets by County

County	Hit? no	Hit? yes	County	Hit? no	Hit? yes
Alameda	211	112	Riverside	55	34
Alpine	0	2	Sacramento	213	82
Amador	3	1	San Benito	2	1
Butte	12	10	San Bernardino	99	38
Contra Costa	36	21	San Diego	700	210
Del Norte	1	0	San Francisco	721	305
El Dorado	8	2	San Joaquin	45	12
Fresno	57	15	San Luis Obispo	18	4
Glenn	6	1	San Mateo	37	14
Humboldt	9	2	Santa Barbara	56	10
Imperial	4	2	Santa Clara	120	65
Kern	18	12	Santa Cruz	34	13
Lake	8	7	Shasta	8	5
Lassen	4	0	Siskiyou	3	0
Los Angeles	2,313	1,355	Solano	16	4
Madera	0	4	Sonoma	59	10
Marin	35	6	Stanislaus	23	11
Mariposa	1	0	Sutter	1	0
Mendocino	4	2	Tehama	10	1
Merced	10	3	Trinity	1	1
Monterey	28	11	Tulare	12	1
Napa	3	0	Ventura	31	20
Nevada	2	0	Yolo	35	11
Orange	231	64	Bay Area	79	46
Placer	25	16	Silicon Valley	29	13
Plumas	2	0			

Note: Number of election-day Tweets (excluding retweets) classified as "hits" by county in California. Counts use the unique texts across all Tweets. "Bay Area" and "Silicon Valley" locations, which span multiple counties, are also shown.

Table 18: Strata for Sampling Tweets to use in Initial Training Set

		population		sample	
state+	total[a]	not hit	hit	not hit	hit
AZ	9,890	1,607	478	62	109
CA	52,296	10,774	271	414	62
CT	3,537	712	24	27	6
CO	8,388	1,511	261	58	59
WAd	10,062	1,958	169	75	39
WAr	2,910	608	7	23	2
CAeo	3,041	558	72	21	16
COeo	177	68	6	3	2
WAeo	505	105	3	4	2

Note: [a] “total” values are the numbers of unique Tweets (no retweets) in each stratum in the set of Tweets manually downloaded using Sysomos.

Figure 1: Flowchart for Making Hits Decisions in American Twitter Election Comments, Part 1

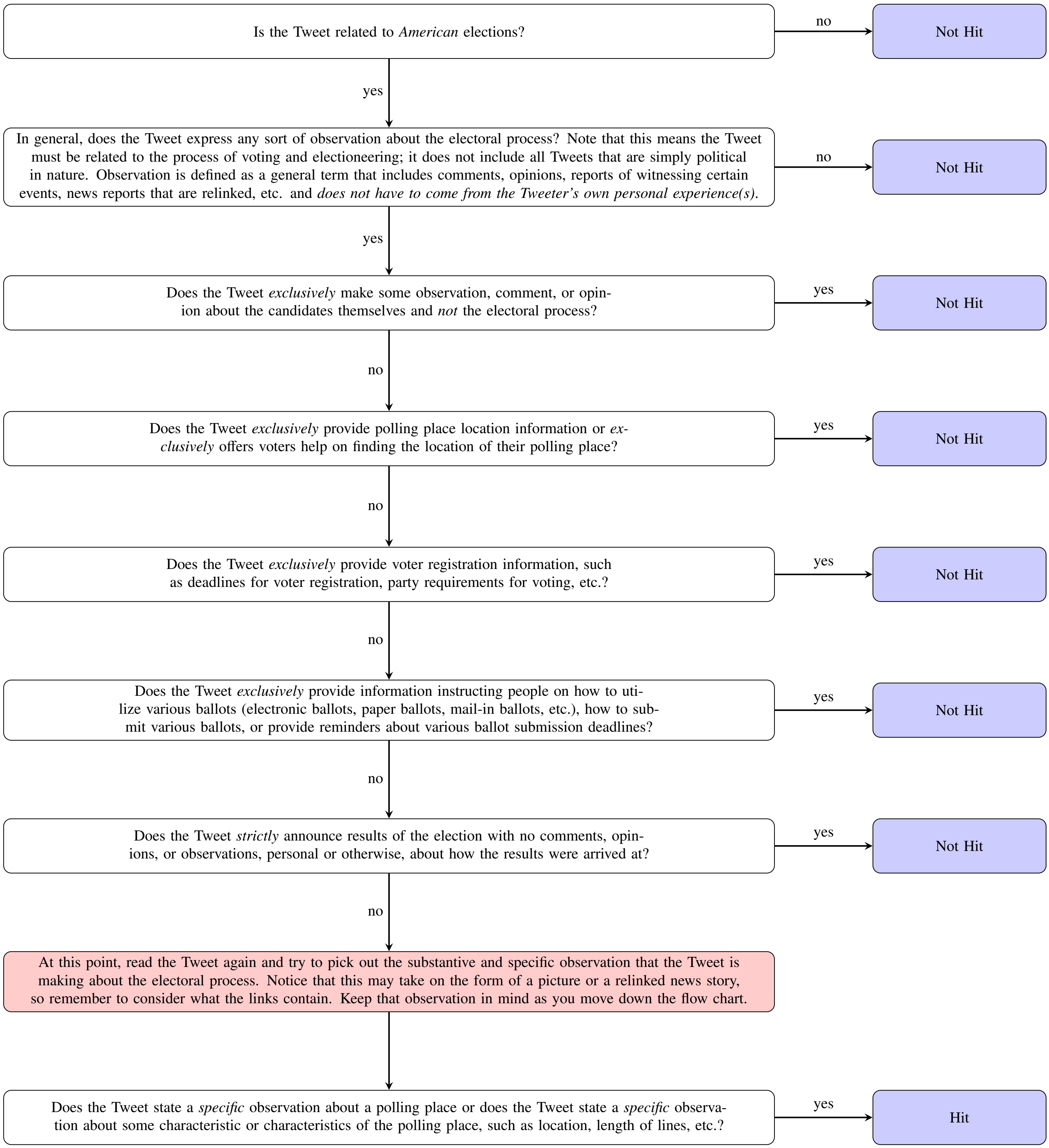

Figure 2: Flowchart for Making Hits Decisions in American Twitter Election Comments, Part 2