1 Trustworthy Elections

1.1 Introduction

In his classic study of electoral politics in the American South, V. O. Key, Jr. wrote: "If a democratic regime is to work successfully it must be generally agreed that contestants for power will not shoot each other and that ballots will be counted as cast" (Key, 1984, p. 443). That was true back in 1949, when Key wrote *Southern Politics*, and it is certainly true today. For a representative democracy to function well, it needs to have a reliable process for administering elections, that process should be free and fair, and all concerned should be convinced of the integrity of the administration of elections. That is, stakeholders, candidates, and voters should all have a high degree of confidence that each and every election is conducted with great integrity – as Key said, "that all ballots will be counted as cast."

Consider this question from the perspective of a registered voter who lives in Costa Mesa, California. Costa Mesa is a community in Orange County, tucked between the beautiful coastal town of Newport Beach (and the Pacific Ocean), and the larger inland cities of Irvine, Santa Ana, and Anaheim. In the 2018 general election in Costa Mesa, there was a hotly contested election for the 48th Congressional seat. This Congressional seat had been held for three decades by a conservative Republican, Dana Rohrabacher. As a Congressman, Rohrabacher supported conservative causes his entire career, getting a score of 94 of 100 by the American Conservative Union (ACU) in 2017, and controversially flaunted his close connections to Russia. In the general election, the incumbent Rohrabacher faced Democrat Harley Rouda, who received financial support from Democrats across the nation in his bid to flip this Republican seat in Orange County.

On election night, it was clear that the race between Rohrabacher and Rouda would be very close: in the first tally reported by the OCROV, Rohrabacher had a slim lead, with 52,451 votes to Rouda's 52,370. As additional ballots were counted on election night, the lead bounced back and forth between the two candidates; by the time the Registrar released the first tally report early in the morning after the election, Rouda had a slight lead of just over 2,200 votes. In the days and weeks following the election, Rouda's lead steadily grew. In the final official results,[1] Rouda had accumulated 157,837 votes, to Rohrabacher's 136,899 – an impressive win by a challenger over a long-time incumbent.

[1] Neal Kelley, Orange County Registrar of Voters, *Orange County 2018 General Election, November 6, 2018 Official Results for Election*, `www.ocvote.com/fileadmin/live/gen2018/results.htm`, last retrieved March 15, 2020.

But our hypothetical voter, who does not know much about election administration, might wonder exactly what went on. Why did such a close election become a strong win for Rouda? Why did the ballots counted days and days after the election favor the Democratic candidate so heavily? Was this a fair process, and were all of the ballots counted as cast? Did Rouda win, fair and square?

Students of election administration may recognize this phenomenon: typically, the first set of ballots counted right away on election night are the absentee by-mail ballots that have been received by the election officials in the days before the election, with those by-mail ballots often coming from Republican-leaning voters in California. Election Day voters, and also provisional voters, often tend to be more Democratic in their preferences, so ballots that get counted in the days and weeks following an election will frequently skew Democratic. There is nothing nefarious about this – but these details are not obvious to most voters.

So how can we help convince this concerned voter in Costa Mesa that the election was conducted fairly, and that the process had a high degree of integrity from start to finish?

Measuring and confirming the integrity of elections is precisely what this Element is about. How can a team of data and political scientists monitor a major and important federal election like the one held in 2018 in Orange County, and produce analytical materials that can be made available to the public in near real-time during an election? How can we quickly produce studies covering all aspects of an election's conduct, and package these analyses for an interested public to read in order to determine whether to trust an election's reported result?

These were the goals of an innovative project in Orange County that our team launched in 2018, and we report the results of this project in this Element. In the sections that follow we will present in detail the types of studies we undertook in Orange County, what we found, and from there outline best practices for future projects like these. To jump to the conclusion, we find that the election was conducted with a high degree of integrity – our data and analyses show that Rouda won the election, and that we can be very confident in the outcome of this and the rest of the elections on the 2018 Orange County ballot.

1.2 Performance Auditing of Elections

In many American states, election officials routinely do some type of post-election auditing – procedures that seek to confirm that the methods and

technologies used to tabulate votes operated as expected and that the tabulation process led to reproducible results. These routine postelection ballot audits have been defined by the National Council of State Legislatures (NCSL):[2]

> While the phrase "post-election audits" can be used to mean a variety of election validation efforts, as a term of art it refers to checking paper ballots or records against the results produced by the voting system to ensure accuracy. 34 states + DC currently have a post-election audit as defined here. Paper records used in an audit may include voter-marked paper ballots, voter-verified paper audit trails produced by direct-recording electronic voting machines (DREs) or paper ballot records produced by ballot-marking devices. Typically only a sample of the paper records are examined, so in effect a post-election audit is a partial recount of results to verify that the voting system is accurately recording and counting votes.

Postelection ballot audits are powerful evaluative methodologies. By comparing an independent tabulation of the ballots cast in an election, postelection ballot audits can help confirm that the technologies and procedures used for ballot tabulation worked as expected, and if conducted in a public and transparent manner, can help buttress public confidence in the integrity of the ballot tabulation process (Alvarez et al., 2012). There are many different types of post-election ballot audits. The NCSL study shows that the most prevalent type is the "traditional" fixed-percentage audit: a set percentage of ballots cast or voting precincts are selected, and either the ballots selected or the ballots from the sampled precincts are included in the auditing tabulation. Another type is the so-called "risk-limiting" audit (RLA) (Stark, 2009), currently used in a limited (but growing) number of election jurisdictions in the United States. The RLA samples ballots for the postelection audit based on the reported election outcome: the closer the election contest, the greater the number of ballots that will be included in the audit – this helps the RLA confirm the outcome of very close election contests (a property that traditional fixed-percentage audits do not have).

However, as we will discuss in detail in Section 4, all postelection ballot audits are limited to confirming the initial tabulation of ballots. In other words, they are useful for confirming that the initial tabulation was correct, or for confirming that the procedures and technologies used for the initial ballot tabulation worked as expected. Thus, postelection ballot audits are an important tool for confirming the integrity of aspects of an election's administration, but

[2] NCSL, *Post-Election Audits*, `www.ncsl.org/research/elections-and-campaigns/post-election-audits635926066.aspx`, last retrieved March 15, 2020.

they do not help us understand the integrity of the voter registration data, specific problems that voters or poll workers had on Election Day, and whether there were anomalies or problems with the conduct of a jurisdiction's election.

Thus, instead of focusing solely on postelection ballot auditing, in this Element we argue that we should take a broader perspective, and try to evaluate the performance of the complete administration and conduct of an election. This type of holistic and ecological auditing seeks to examine the performance of an election process from end to end, so that a skeptical voter (like our hypothetical voter in Costa Mesa) can have confidence that all aspects of an election, from the registration of voters well before ballots are cast, to the vote-by-mail process, to the postelection tabulation of all ballots, are performing as expected. Our general approach for evaluating the integrity of an election is performance auditing of elections, which builds off of recent work on comprehensive audits of elections, and audits of components of the election process other than ballot tabulation (Alvarez, Atkeson, & Hall, 2012a, 2012b; Selker, 2005). In the remaining sections of this Element, we will present different methodologies and tools (including postelection ballot auditing) that can help us evaluate an election from a holistic and ecological perspective, to gain a broader vision of the integrity of an election from start to finish.

1.3 The 2018 Orange County Project

In the 2018 election cycle, we proposed an ambitious project to examine the utility of different types of quantitative election forensics during a major federal election, in a large election jurisdiction, that would be useful to both the public and to election administrators. For this comprehensive study, we were lucky to be able to work closely with the Orange County Registrar of Voters (OCROV), Neal Kelley, and his team.

We choose to focus on Orange County for a variety of reasons. Orange County is a large and diverse area of Southern California. Located south of Los Angeles and north of San Diego, Orange County is home to a wide array of different businesses, colleges, and universities, and of course, Disneyland. The county currently has a total population of almost 3.2 million residents, and in the 2016 presidential election, Orange County had just over 2 million voting-eligible citizens, with 1.5 million registered voters.[3] In that same election, 1.2 million of those registered voters participated (80.71% of registered

[3] Data from the California Secretary of State, `http://elections.cdn.sos.ca.gov/sov/2016-general/sov/02-voter-reg-stats-by-county.pdf`.

voters).[4] Orange County's population is also quite diverse, as the US Census Bureau's most recent estimates show that 72% of the county's population is White, 21% Asian, 2% Black, and 3.5% two or more races. The Census Bureau's recent data estimates that 34% of the Orange County's population is Hispanic or Latino.[5] Thus, one reason we focus on Orange County for this study is that it is one of the largest and most diverse election jurisdictions in the United States.

Secondly, Orange County is widely viewed as innovative in the administration of elections. The County's Registrar of Voters, Neal Kelley, participates widely in state and national professional organizations, and has been recognized for his innovative administrative practices. Under his administration, Orange County has developed many administrative processes and tools that are viewed as best practices for election administration, for example, building transparency by webcasting in real-time virtually all aspects of the process of administering an election, or more recently, pilot testing risk-limiting audits.

Because of these factors, in 2018 we established a unique collaboration between researchers from the California Institute of Technology and OCROV. The collaboration, which continues today, focuses on developing applications and analytical tools for documenting the integrity of the county's elections (primary and general), with the use of quantitative and qualitative methodologies. One of the important components of this project is the development of the quantitative methodologies that we report on in this Element – methods for quickly analyzing daily snapshots of Orange County's voter registry, in order to efficiently and effectively audit the voter registration database. As part of this component of our collaboration, OCROV agreed to provide *daily* snapshots of their entire voter list, excluding only a small set of fields that contain highly personal and sensitive information (in particular, the voter's California driver's license or identification number, and their Social Security Number). As far as we are aware, this is the first time that a county election jurisdiction in the United States has provided daily voter-file snapshots, with an extensive array of information about each registered voter, over a long period of time, to academic researchers.[6]

[4] Data from the California Secretary of State, `http://elections.cdn.sos.ca.gov/sov/2016-general/sov/03-voter-participation-stats-by-county.pdf`.

[5] These figures are from the Census Bureau's July 1, 2017 population estimates, `www.census.gov/quickfacts/fact/table/orangecountycalifornia/PST045217`.

[6] Scholars of election administration know that certain states make some of their voter registration data available, either by mail (e.g., Florida), or online (e.g., North Carolina or Ohio). While our techniques could be used on data from those jurisdictions, the data we have is more granular (available daily) and has a very extensive set of features for each voter, allowing deeper analyses of record change and of potential duplicates.

1.4 Roadmap

In the sections that follow, we present each of the quantitative methodologies that we have developed and used in our election performance auditing research. Our Element is structured in two parts. The first part presents methodologies that seek to obtain direct evaluative data from participants and observers in elections: social media monitoring of election reports and discussion (Section 2), and surveys of voters and poll workers (Section 3). In the second part of the Element, we turn to statistical forensic methods for evaluating the performance and integrity of an election, using postelection ballot audits and forensic studies of turnout and voting statistics (Section 4), as well as the statistical analysis of voter registration data (Section 5).

We begin in Section 2 with our work that uses social media monitoring, specifically monitoring Twitter, to collect data on election experiences, and concerns about elections, at large scale and in real-time. In this section, we present a number of results that demonstrate the potential utility of social media monitoring for studying election integrity; we also discuss many of the current challenges raised by the collection and analysis of social media data for election monitoring.

Section 3 turns to ways to best measure voter and poll worker experiences. In this section, we focus on directly measuring these experiences using micro-level surveys. Over the past decade, a number of research groups have developed some relatively standard ways to measure voter experience and confidence, and we use many of those measurement approaches in our Orange County work. In this section, we discuss the pros and cons of direct measurement of a voter's experience with online surveys, and we also present some data from poll worker surveys collected by the OCROV.

Moving to Section 4, there we shift the discussion to the use of postelection ballot audits, and other statistical anomaly detection methodologies, that can be used to assess the integrity of components of the electoral process. Postelection ballot audits are used in many states, some using auditing approaches that seek to confirm that the election technology and procedures for balloting worked as expected, while others try to assess whether the election outcome was correct. In 2018, Orange County implemented both types of postelection ballot audit, so we discuss what we can learn from this methodology in the first part of this section. Then we turn to the topic of independent third-party forensic techniques, focusing on the analysis of precinct-level turnout and vote share data, using graphical methods. We present some of the results from the use of these forensic tools in Orange County, and discuss their strengths and weaknesses.

In Section 5, we turn to the important question of monitoring one of the major components of election administration in the United States: evaluating the integrity, accuracy, and security of voter registration databases. In this section, we present a methodology that we have developed to audit and monitor large voter registration databases, a methodology that we have used in Orange County since early 2018 and are now beginning to implement in other counties and states. This section outlines the approach we take to monitoring voter registration databases, shows the information we can obtain using our method, and discusses what the future may hold for quantitative methods that can audit voter registration data.

In the final section, we bring these various methods together to examine, in a holistic and ecological way, what inferences and conclusions we can draw from the data and analysis about the integrity of the 2018 elections in Orange County. We also discuss in this final section some important lessons we have learned in our 2018 research, and some next steps in the development of methods for assessing the integrity and security of American elections.

Finally, our code and data are all publicly available and hosted on GitHub (`https://github.com/monitoringtheelection`). Our project website will also continue to be the host for updates and public reporting about our election integrity projects (`https://monitoringtheelection.us/`).

2 Social Media Monitoring

2.1 Why Monitor Social Media?

The advent of social media has granted voters access to a modern "political forum," a place to discuss the voting process with others in real-time during an election. By tracking voter discussions of elections on social media as they occur, we can begin to observe, highlight, and address specific problems as they arise in the electoral process.

In this way, tracking citizen-provided descriptions of problems and concerns with the voting process is a potential replacement for in-person election monitoring, which many consider the current "gold standard" for detecting election problems. In-person election monitoring, where trained and experienced poll watchers physically observe the election process on-site, allows researchers to gain a highly detailed, qualitative sense of exactly where and when problems in the voting process occur. However, this methodology requires a large investment of time and resources; even with a trained team of poll watchers, only a minuscule fraction of polling places can realistically be observed in a given election. This limited coverage makes in-person election monitoring unfeasible

if the goal is to detect problems in a large election jurisdiction, especially when trying to detect low-incidence problems.

Analyzing social media data, on the other hand, can provide wide and detailed coverage across the United States, in specific states, and perhaps in counties. In addition to being far less costly than in-person monitoring, social media data might detect election problems at the scale of a large federal election, an important distinction to previous qualitative efforts. This is true even for low-incidence and geographically concentrated election issues, as discussed later in this section.

Thus, in our effort to develop a well-rounded, ecological approach to monitoring and securing elections, our team created and implemented a social media election monitor during the 2018 midterm elections. Building on previous work analyzing Twitter discussions during electoral cycles (e.g. Adams-Cohen et al., 2017, Lin et al., 2013, McKinney, Houston, & Hawthorne, 2013), we developed a series of scripts and algorithms to collect and store a large volume of social media data. We then analyzed the overall trends in conversations about the 2018 voting process in the days before and after Election Day.

While we collected several months of data, we focus most of this section on our analysis of data collected on November 6, 2018, the date of the midterm election. We first analyze conversations during this 24-hour period to detect problems with the election at the national scale, before using location inference techniques to monitor voting problems at a state and local level. As in the rest of this Element, we focus our local analysis on the Orange County elections.

Even though our approach had some limitations, which we discuss in the conclusion of this section, our work provides a strong foundation for future methods that leverage social media data to better understand and address issues with the electoral process.

2.2 Our Methodology

Our team developed and implemented a methodology to track discussions concerning elections and the voting process by collecting and analyzing Twitter data. With an average of 326 million daily active users[7] and emphasis on sharing immediate reactions to events, Twitter is a rich potential source of information about elections and voting.[8]

[7] According to the Q3 2018 Twitter investor statement. See `https://investor.twitterinc.com/files/doc_news/archive/4ad1fd92-0dea-4c13-9a71-8674acf154cc.pdf`.

[8] See Steinert-Threkeld (2018) for a primer on the best practices in using social media data in the social sciences.

Political scientists use social media datasets to study various aspects of the electoral process, including the way politicians use social media in their political campaigns (Golbeck, Grimes, & Rogers, 2010; Graham, Jackson, & Broersma, 2016; Theocharis, Barberá, Fazekas, Popa, & Parnet, 2016), how the public discusses presidential debates and other major campaign events (Lin et al., 2013; McKinney et al., 2013; Murthy, 2015), and what Twitter networks reveal about political polarization (Barberá, 2015; Conover et al., 2011).

Given Twitter's emphasis on sharing immediate reactions to trending events, many researchers have leveraged the platform as a way to gauge and measure public opinion (Beauchamp, 2017; O'Connor, Balasubramanyan, & Routledge, 2010; Sajuria & Fabrega, 2016). A large body of this work specifically tracks Twitter conversations of politicians and political parties, analyzing these conversations in order to forecast election results (Burnap, Gibson, Sloan, Southern, & Williams, 2016; Ceron, Curini, & Iacus, 2015; Murthy, 2015).

We follow this literature by first collecting Twitter data during a campaign to make inferences about the election process. However, instead of attempting to forecast election results, we turn our attention to measuring the integrity of the election. That is, instead of forecasting the electoral success of specific politicians or political parties, we collect messages in which users describe the voting process, in an attempt to locate potential issues with federal, state, and local elections.

Our collection of Twitter data roughly consists of the following steps:[9]

1. Defining a set of keywords broadly associated with common electoral and voting issues.
2. Setting up a series of Python scripts that interact with the Twitter Streaming API, granting access in real-time to tweets that contain one or more of our track words.
3. Storing each message, including the associated metadata and user information, in a MySQL database.
4. Analyzing the text and metadata of these messages over the election period.

Our team designed this process and implemented early versions in the 2014 election cycle, and we have been working to refine and improve the process in each successive election cycle (Adams-Cohen et al., 2017).

For the 2018 midterm election, we focused on five broad topics: "Election Day voting," "election fraud," "remote voting," "polling places," and "voter identification." Table 1 displays the keywords we monitored for each category.

[9] This methodology was initially developed in Adams-Cohen et al. (2017), which contains additional technical details on each specific step in our process.

Table 1 Tracked keywords

Category	Keywords
Election Day Voting	provisional ballot, voting machine, ballot
Voter Fraud	election fraud, election manipulation, illegal voters, illegal votes, noncitizen voting, noncitizen votes, illegal voting, illegal vote, illegal ballot, illegal ballots, dirty voter rolls, vote illegally, voting illegally, voter intimidation, voter suppression, rigged election, vote rigging, voter fraud, voting fraud, ballot destruction, vote flipping, flipped votes, voter coercion, ballot stuffing, ballot box stuffing, vote buying, voting machine tampering, rigged voting machines, voter impersonation, election integrity, election rigging, duplicate voting, duplicate vote, ineligible voting, ineligible vote, dead voters
Remote Voting	absentee ballot, mail ballot, vote by mail, voting by mail, early voting
Voter ID	voter identification, voting identification, voter id
Polling Places	polling place line, precinct line, pollworker, poll worker

While clearly not an exhaustive set of terms related to each category, we leveraged domain expertise to choose a set of terms broad enough to collect the majority of conversations concerning an election topic but narrow enough to stay within a particular issue domain.[10] We solicited input from a number of election administration experts in selecting these keywords.

For each message we collect, we also obtain a rich set of metadata associated with a tweet, including information about the user sending the message. One particularly useful piece of information contained in a tweet's metadata is the user's location. By detecting the origin of a message, it is possible to pinpoint specific states or municipalities experiencing an abnormally large volume of

[10] A risk in following a static list of keywords is missing potentially important words and phrases that may develop dynamically during the data collection period (King, Lam, & Roberts, 2017). As we iterate and develop our social media monitor for future elections, we hope to introduce dynamically evolving keywords along the lines of Liu et al. (2019).