Local Governance, Civil Discourse, and Social Media: Charting Incivility From and Directed at Tucson-Area Elected Officials

Making Action Possible in Southern Arizona (MAP Dashboard)
White Paper #13
October 2019

Prepared by
Steve Rains, PhD
Professor, Department of Communication, University of Arizona

Yotam Shmargad, PhD
Assistant Professor, School of Government and Public Policy, University of Arizona

Kate Kenski, PhD
Professor, Department of Communication, University of Arizona

Kevin Coe, PhD
Associate Professor, Department of Communication, University of Utah

Bulut Ozler, Visiting Scholar
School of Information, University of Arizona

Steve Bethard, PhD
Assistant Professor, School of Information, University of Arizona

Table of Contents

Executive Summary

Civility has long been understood as a central element of an effective democracy. From the ancient Athenian forums to modern calls for deliberation and dialogue, a commitment to civil discourse has been viewed as a democratic ideal. In the United States, growing political polarization and the widespread adoption of social media during the past decade have coincided with increased concerns about incivility.

In this white paper, we examine incivility by and directed at elected officials representing the Tucson area across three levels of government on social media. We collected messages (i.e., "tweets") authored by and directed at 33 elected officials on Twitter from January 2018 through June 2019. The Tucson-area elected officials came from the city councils of Oro Valley, Marana, Tucson, Sahuarita, and South Tucson as well as state and federal representatives whose districts included Pima County. A total of 24,778 original tweets made by local elected officials and 71,638 directed at one or more of these same officials were collected. We then developed a machine learning classifier to detect the presence of incivility -- operationally defined as name-calling. The classifier was used to evaluate all tweets in the sample. Analyses were conducted to examine the prevalence of incivility among different groups and over time.

Key findings:

- Incivility was most common in tweets directed at officials representing the Tucson area in federal government. Fifteen percent of all tweets directed at this group contained incivility. Uncivil tweets directed at federal officials representing the Tucson area peaked during December 2018.
- Although officials in city government were more likely to be the target of incivility than in state government, the volume of uncivil tweets directed at these two groups was modest. Less than 5% of tweets directed at officials representing the Tucson area in city and state government contained incivility.
- The discrepancy between incivility directed at federal officials and state or city officials could indicate that incivility increases when people perceive greater distance between themselves and their elected officials. State and city level officials are more accessible to the average citizen, and thus potentially less appealing as targets of incivility. The typically weightier stakes of national level politics may also be a factor, with citizens lashing out with name-calling around particular national issues about which they have strong feelings.
- Incivility was uncommon in tweets authored by elected officials representing the Tucson area. Although officials in federal government were most likely to be uncivil, only 3% of their tweets contained incivility. Officials representing the Tucson area in state government were least likely to be uncivil with only 1.5% of their tweets containing incivility. Trends in incivility from state level officials

followed the most discernable pattern over time. Incivility peaked among state officials just prior to the 2018 election.

Recommendations:

- Federal officials should learn from the Twitter behavior of state level officials. State level officials used half the amount of incivility that federal officials did, presumably with no loss of useful information transmission. In other words, state level officials were able to pursue the same objectives as federal officials without the need to use incivility as frequently.
- Citizens using Twitter for political exchanges should reflect upon the fact that they may engage in more name-calling with federal officials than they would with city or state level officials.

Introduction to the Problem and Project Overview

Civility has long been understood as a central element of an effective democracy. From the ancient Athenian forums to modern calls for deliberation and dialogue in a variety of sectors, a commitment to civil discourse has been viewed as a democratic ideal. Civility is not without problems—the concept has at times been used as way to continue to exclude from public discourse voices that have historically been marginalized—but is something that, at least in the abstract, people usually support.

Civility is particularly likely to be talked about and held up as an ideal during times of social strife and political conflict—that is, during periods when incivility seems to be omnipresent. In the United States, growing political polarization and the widespread adoption of social media during the past decade have coincided with increased concerns about incivility. For example, a 2018 "Civility in America" survey found that 93% of the U.S. public felt that the nation had a "civility problem," with 69% viewing the problem as "major."[1] On average, those surveyed reported encountering 10.6 instances of incivility each week, a number that had increased from 6.2 in 2016.

As public concern over incivility has risen in recent years, researchers have increasingly wrestled with several of its key elements, including its content, causes, and consequences. This research has demonstrated that, although incivility is not wholly negative in its effects, it does have some outcomes that might limit the kind of productive public discussions often viewed as necessary for the smooth functioning of democracy (e.g., incivility can undermine evaluations of speakers and their arguments). What remains to be seen is the scope of the threat posed by incivility, and how it arises in key democratic contexts.

The present research contributes to emerging understandings of incivility by evaluating the prevalence of uncivil discourse in social media messages from and about officials representing the Tucson area in city, state, and federal government.

Defining and Studying Incivility

In research on incivility, it is popular to talk about the difficulty of defining this concept. One person's incivility might be another person's heroic antagonism. Incivility directed at an entity one abhors might not sound as uncivil as the same remark directed at an entity one admires. We build on insights from our prior work to inform our definition of incivility. In a previous study, we had trained human coders evaluate more than 6,000 comments posted during a three-week period during 2011 in response to news stories on the *Arizona Daily Star* website.[2]

Working with a general understanding of incivility as "features of discussion that convey an unnecessarily disrespectful tone toward the discussion forum, its participants, or its topics," the coders tracked specific forms of incivility, such as name-calling, vulgarity, and accusations of lying. Far and away, the most common form was name-calling. In a follow-up study, we found that name-calling was also among the forms of incivility most consistently identified as uncivil by members of the public.[3] Taking guidance from this past research, we narrowed our focus in the present research to instances of uncivility in the form of name-calling (i.e., mean-spirited/disparaging words directed at another person or group of people). As the most common form of incivility and form most recognized by the lay public, name-calling represented the best candidate for exploration in our project.

Even having defined it, measuring incivility is no easy task. For example, in an early effort by the Engaging News Project (now part of the Center for Media Engagement at the University of Texas at Austin) a researcher spent three months experimenting with different ways of coding incivility in online comments (e.g., trained coders, untrained "workers" via Amazon's Mechanical Turk, the computer program WordStat) but had difficulty achieving reliability.[4] The measurement strategy we employ here is to start with trained human coders who received extensive training to identify incivility in the form of name-calling and use their coding to develop and validate a machine learning algorithm that can detect name-calling in large data sets. This approach harnesses the human capacity to draw nuanced distinctions in textual content with the capacity of computers to "learn" from exemplars and rapidly process vast amounts of text.

Key Contexts: Local Politics and Social Media

Research on politics too often neglects the local. This happens for a myriad of reasons: local data are often harder to gather, variation across localities makes generalizations (a goal of many forms of research) more difficult, national level research attracts more funding and interest, and so on. But as the saying goes, "all politics is local"—and research that focuses on a specific locality has an opportunity for nuance sometimes missed in broader research contexts.

Incivility is an especially good topic for local-level research, because regional differences in cultural and social norms are likely to influence how incivility unfolds in different parts of the country. Isolating how incivility is employed in a specific location—such as a city or state—removes this hurdle and thus allows for different kinds of questions to be asked.

This study pairs an interest in the local with a focus on social media, which provide an ideal context in which to track incivility. Social media data are vast and (usually) accessible, which is a particular advantage when employing machine learning. Just as important, social media are a venue where people often encounter incivility. In fact, the aforementioned "Civility in America" survey found that respondents reported encountering slightly more instances of incivility online than offline.

The goal of our work, then, is to examine the prevalence of Twitter-based incivility by, and directed at, local elected officials and to investigate differences in incivility by Tucson-area officials elected to the three levels of government (local, state, and national).
We formalize these objectives in the following two aims:

Aim 1: Evaluate the prevalence of incivility (a) by local elected officials representing the Tucson area and (b) compare it with incivility among Tucson-area representatives in the state and federal government.

Aim 2: Evaluate the prevalence of incivility (a) directed at local elected officials in the Tucson area and (b) compare it with incivility directed at Tucson-area representatives in the state and federal government.

Following these aims, this white paper offers insights into the state of incivility on social media among elected officials representing the Tucson area as well as in discussions between those officials and the lay public. Beyond helping to better understand the nature of the problem posed by incivility for local governance in the Tucson area, this research helps generate strategies for promoting civility.

Data Acquisition and Analytic Procedures

Identifying Elected Officials

The data for this project were acquired in a series of steps. Elected officials representing the Tucson area in city, state, and federal government during 2018 were first identified using the state and city records. The sample included 62 total officials from the US Senate (n = 3), US House of Representatives (n = 3), Arizona State Senate (n = 7), Arizona State House of Representatives (n = 14), and officials serving as mayors or city council members for the cities of Oro Valley, Marana, Tucson, Sahuarita, and South Tucson (n = 35). Twitter was then manually searched to determine whether or not each official had an account and identify the official’s username.

The final sample included in the analysis consisted of the 33 officials who had created a Twitter account. The percentage of officials who had accounts from the different levels of government were as follows: US Senate (100%), US House (100%), Arizona Senate (71%), Arizona House (79%), Mayor and City Councils (31%). The use of Twitter at the city level was not evenly distributed, with only mayors and council members from Tucson (86%), Oro Valley (43%), and Sahuarita (33%) using Twitter. None of the officials from Marana or South Tucson had a Twitter account. Figure 1 identifies the complete list of elected officials and their Twitter usernames.

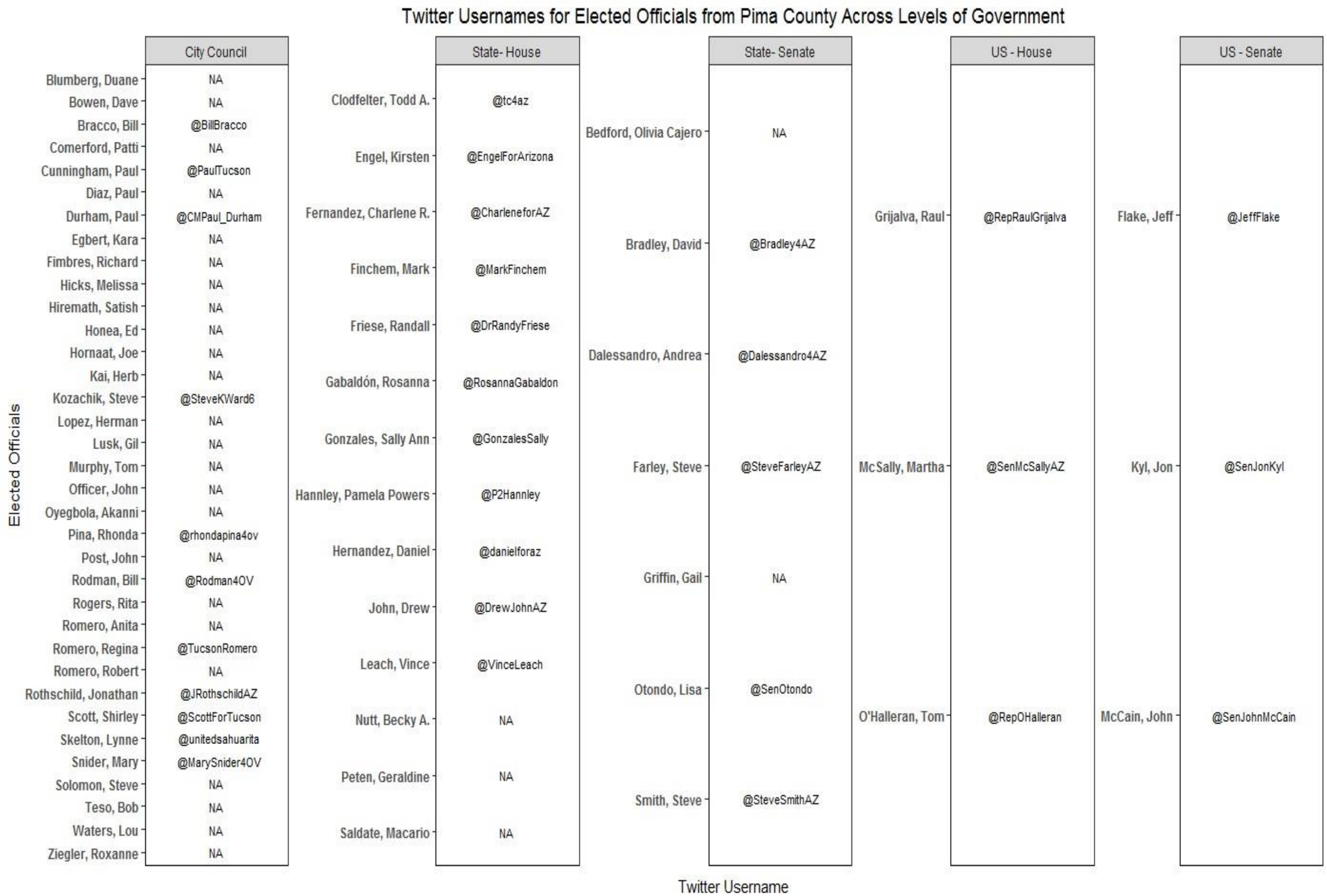
Collecting Tweets to and from Elected Officials

For each of the 33 elected officials with a Twitter account, Twitter's application programming interface (API) was used to acquire two sets of tweets. First, we identified tweets made by each official. Second, we extracted tweets in which at least one of the 33 officials was mentioned. This latter set of Tweets were posted by people other than the official but explicitly mentioned the official using their Twitter username (e.g., @JeffFlake).

Tweet collection was constrained by two limitations of Twitter's API. First, only the 3,200 most recent tweets were available for each official. This limit includes the official's replies to other users' tweets (which we include in our analysis) as well as other users' tweets that the official shared (which we exclude). Second, tweets mentioning officials are constrained to the past 10 days at the time of data collection.

Three accommodations were made to overcome these limitations. First, we used the API to collect data at two time points, which were selected based on the timeline for this project: April 10, 2019 and June 7, 2019. Second, we used a paid, "premium" version of the API to collect additional tweets. With the premium API, we collected the 100 most recent tweets mentioning officials, which provided additional tweets for officials that were mentioned relatively infrequently. Third, we supplemented the API data with data obtained through web scraping. This allowed us to overcome the 10-day limit of the free API and the 100 tweet limit of the premium API to collect tweets directed at officials in 2018. Notably, this method breaks Twitter's terms of service, and so was used sparingly only to obtain data that was unavailable through the API. While these accommodations helped to supplement the data obtained with Twitter's API, the data should not be considered a census or representative sample with respect to time or level of government. In particular, tweets directed at officials are likely missing in a systematic way for earlier parts of our timespan and for highly-mentioned accounts, which tend to be associated with officials serving at the Federal level.

Figure 1. *Twitter Usernames for Officials Representing*



Prior to conducting the analyses, we limited the sample to original tweets made by or directed at elected officials. Retweets of messages, in which a user forwards a tweet authored by another party, were excluded. We also excluded tweets sent by and directed at local officials after they left office. The final sample included in the analyses consisted of 96,416 original tweets made by and directed at elected officials representing the Tucson area between January 1, 2018 and June 8, 2019. A total of 24,778 tweets were authored by elected officials and 71,638 were directed at officials. The number of collected tweets made by and directed at elected officials for each of the three levels of government can be found in Figure 2.

Figure 2. *Tweets Extracted to and from Elected Officials Across Levels of Government.*

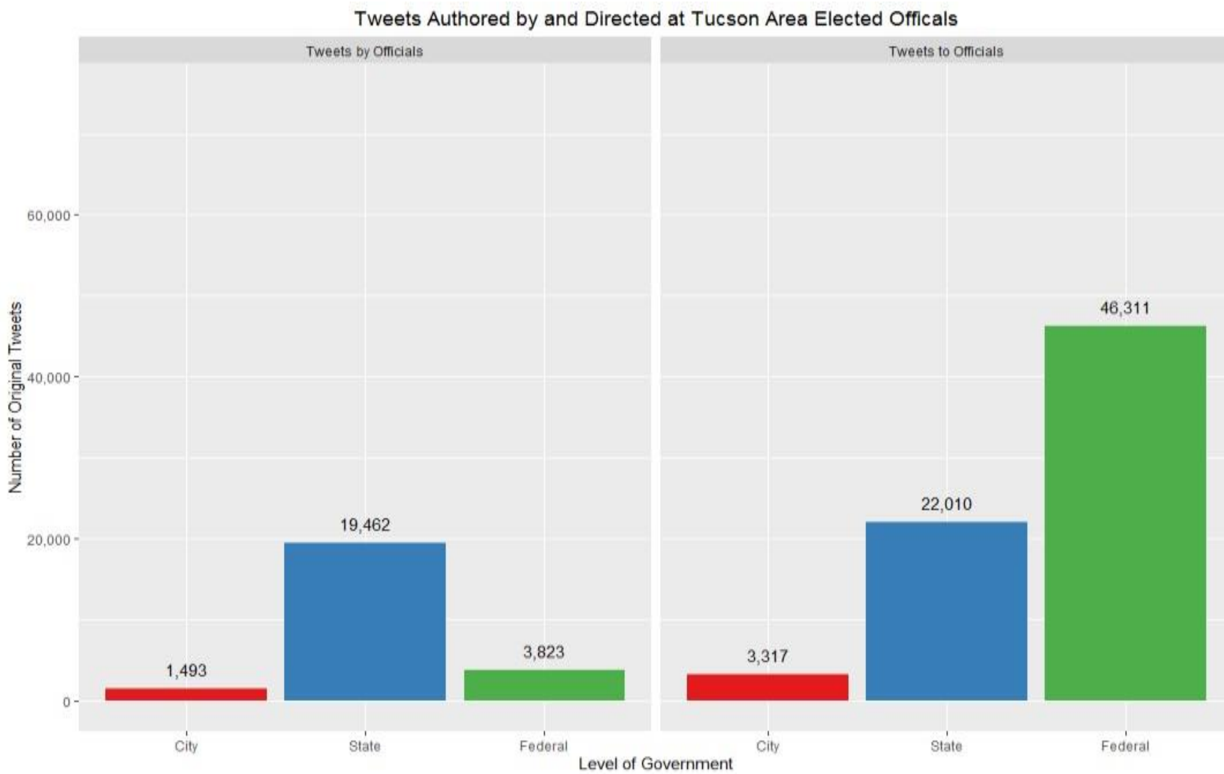
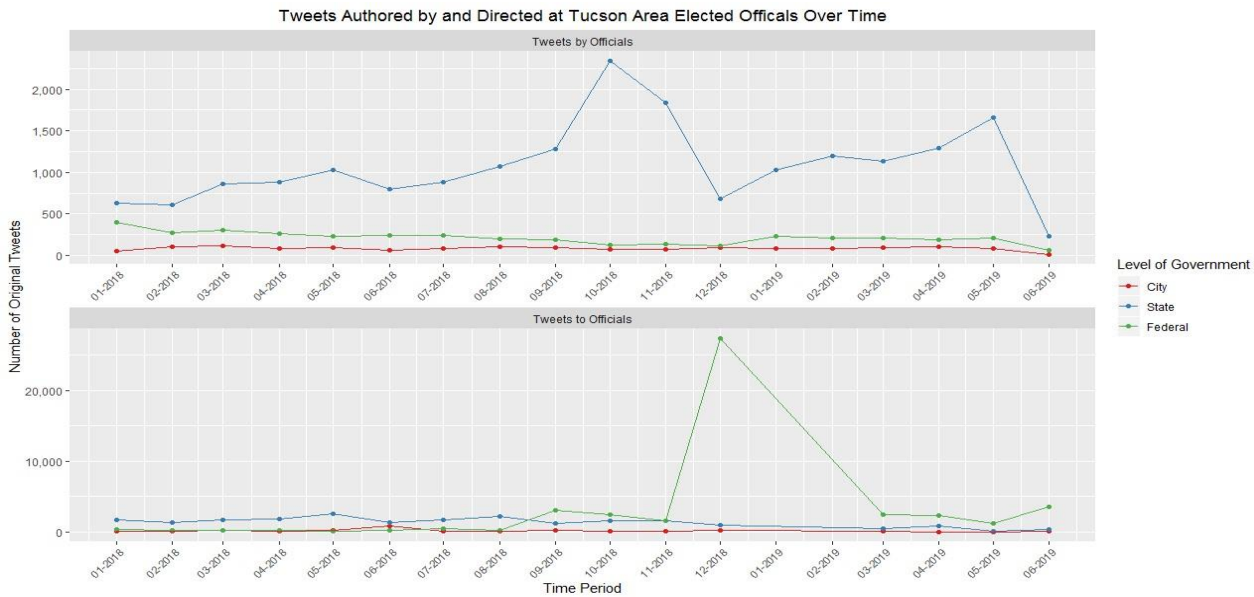


Figure 3 shows the number of collected tweets extracted across time. Although the number of tweets made by city and state officials was fairly constant, the number of tweets collected from state officials were markedly larger during October and November 2018. The increased volume of tweets likely reflects election-related activity. The volume of extracted tweets that were directed at city and state officials was consistent across the sampling frame. For federal officials, however, a substantial number of tweets were authored during December 2018. This increased volume likely reflects limitations of the web scraping procedure, which was unable to go far back in 2018 for popular accounts (which were over-represented at the federal level).

Figure 3. *Tweets Extracted by and from Elected Officials Over Time*



Data Annotation

A random sample of 3,800 tweets were annotated by human coders. The purpose of this process was to develop a set of tweets to serve as the "ground truth" for training and evaluating the machine learning classifier. The human coders were trained to objectively identify the presence of incivility in the form of name-calling, which was evaluated using the guidelines from a previously-developed coding scheme for incivility [2]. It was defined as:

> *Mean-spirited/disparaging words directed at another person or group of people, including derogatory nicknames. Name-calling can go beyond the words used and be implied in stylistic features. Name-calling is directed at another symbol producing entity (person/group/organization).*

Four coders were trained to annotate the tweets. Coders received approximately 20 hours of training in which they were informed about what constitutes name-calling and practiced being able to identify name-calling in tweets. Intercoder agreement was determined by having all four coders evaluate the same set of 400 tweets. Intercoder agreement, determined using Krippendorff's alpha, was acceptable (α = .86). This result indicates that the coders were able to consistently recognize name-calling in the data.

Once intercoder agreement had been established, an additional 3,400 tweets were evaluated by the coders. Each tweet was annotated by two coders. At the conclusion of the coding process, disagreements in annotation were resolved through discussion. The

3,800 tweets annotated by human coders were used to develop the machine learning classifier to automatically detect incivility in the remaining 90,000+ tweets we collected.

Classifier Development

We extended our previous research, where we trained machine learning models to detect incivility in Arizona Daily Star comments [5], to take advantage of a new machine-learning model: bidirectional encoder representations from transformers (BERT), a state-of-the-art neural network classifier pre-trained on more than 3 billion words of English [6]. BERT is not trained to detect incivility, but it can be fine-tuned for the task using manually annotated incivility data such as the 3,800 tweets annotated as part of the current project or the 6,175 comments from the Arizona Daily Star newspaper annotated as part of a previous project.

We fine-tuned the BERT model with 3,040 of the 3,800 Tweets. We explored a variety of hyperparameters including batch size, learning rate, and whether or not to combine the newspaper comments with the tweets. A set of 760 tweets were held out from the tuning process and used to evaluate the fine-tuned model. The best hyperparameter settings yielded a model that achieved 76% precision (how often the classifier's prediction of incivility was also judged as uncivil by humans) and 76% recall (how many of the human-judged incivilities the classifier was able to find) on the held-out 760 Tweets.

To apply the classifier to the 92,616 tweets that had not been annotated by human coders, we first combined all available incivility annotations (including the previously reserved 760 tweets), retrained the BERT model using the hyperparameter settings chosen on the development set, and then applied the classifier to predict the presence of incivility in each of the 92,616 tweets.

Uncivil tweets detected by the classifier included obvious instances of name-calling such as "[elected official] you are a moronic idiot", "you are a turncoat," and "he is just creepy and untruthful." They also included more subtle efforts such as "our unstable [lawmaker]", "cantankerous senator", and "climate-denier advisors." Readers should note that all of the sample tweets presented here have been anonymized to protect the identity of the authors and targets.

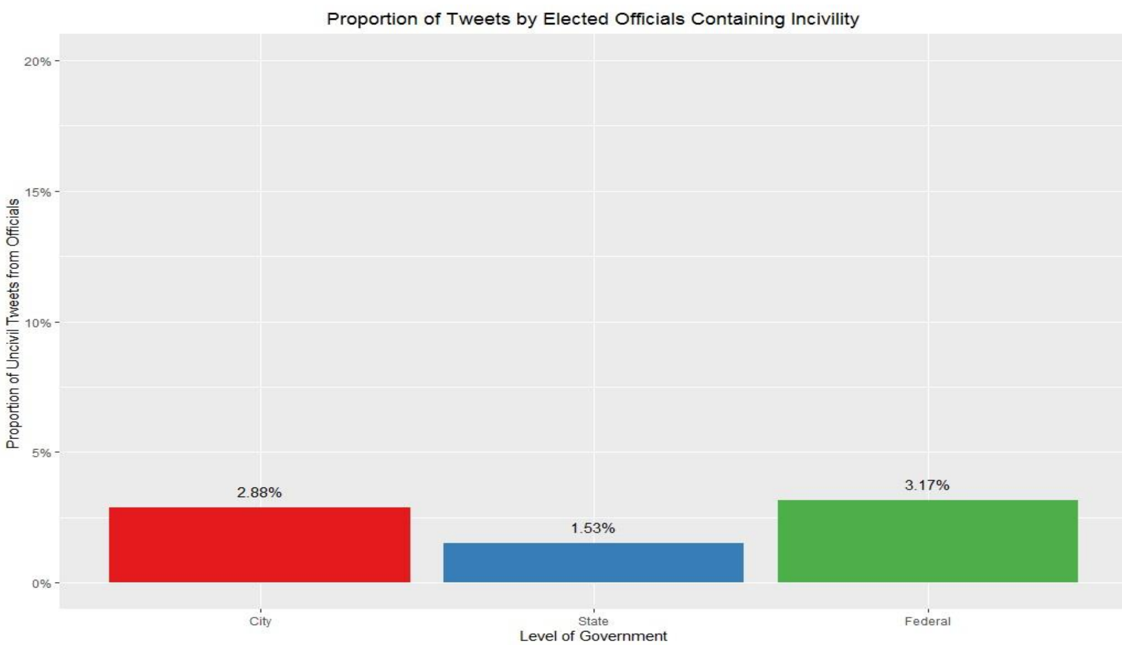
Summary of Results

The two aims animating this project focused on evaluating the presence of incivility on social media by and directed at officials representing the Tucson area. Aim 1 addressed the prevalence of incivility by local elected officials on Twitter. Aim 2 involved incivility directed at these same officials. A series of analyses were conducted in order to address these two aims.

Incivility by Tucson-Area Elected Officials

The prevalence of incivility by elected officials representing the Tucson area in different levels of government can be found in Figure 4. As can be seen in the figure, incivility was rare among elected officials. Although elected officials representing the Tucson area in federal government were most likely to use incivility, they did so in only 3% of their tweets. Officials representing the Tucson area in state government were least likely to be uncivil, engaging in incivility in only 1.5% of their tweets.

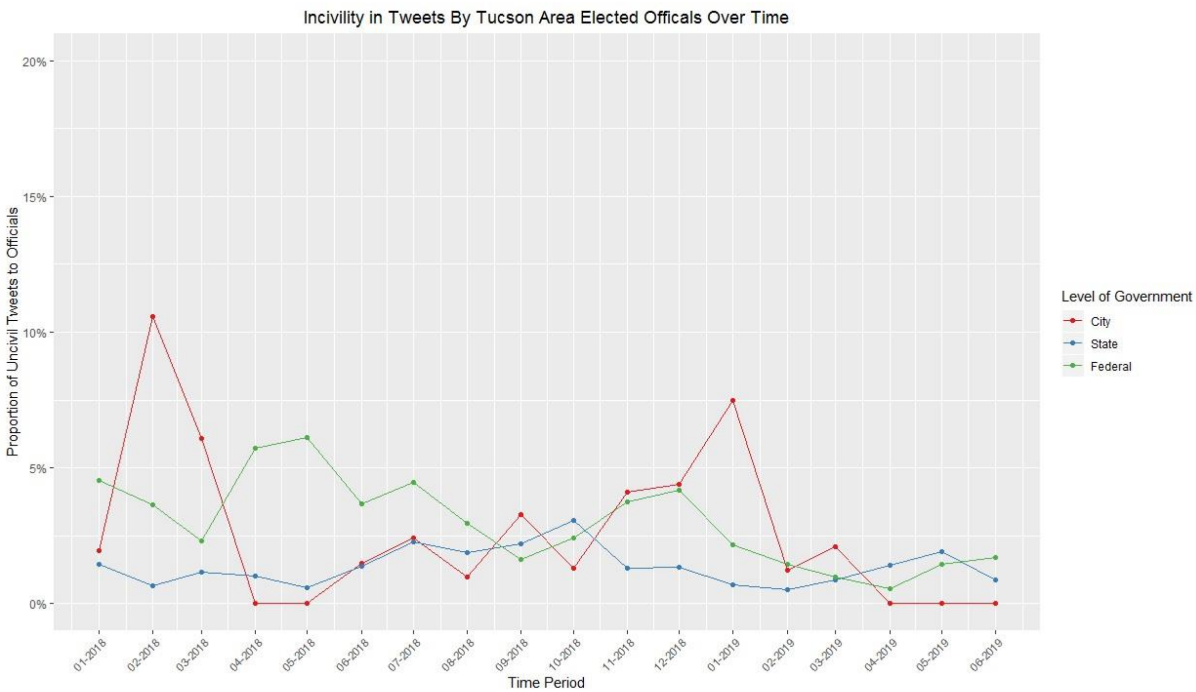
Figure 4. *Proportion of Tweets Containing Incivility by Elected Officials*



Considering the raw number of uncivil tweets offers additional context. Of the approximately 1,500 tweets made by officials representing city government, a little over 40 were uncivil. For officials representing the Tucson area in state government, approximately 300 of the over 19,000 tweets were uncivil. At the national level, our machine learning algorithm indicated that approximately 120 of the almost 4,000 tweets made by officials representing the Tucson area in federal government contained incivility. Over the 18-month period we evaluated, approximately 450 of almost 25,000 tweets collectively authored by elected officials representing the Tucson area were uncivil.

Figure 5 reports the prevalence of incivility in tweets made by Tucson-area elected officials over time. This figure should be interpreted cautiously because the absolute number of tweets being evaluated for each time period was fairly small for representatives in city and federal government. City level officials authored fewer than 250 tweets per month, and federal officials authored fewer than 500 tweets per month.
State officials, in contrast, typically authored between 500 and 1,500 tweets per month.

Figure 5. *Incivility in Tweets by Elected Officials Over Time*

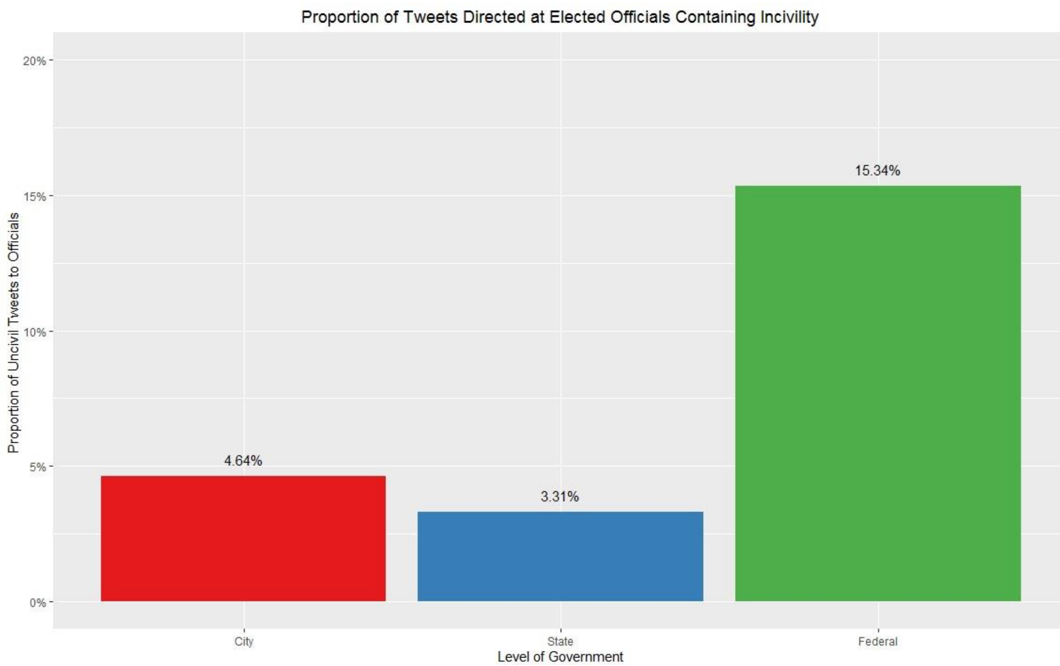


One trend evident in Figure 5 is the noticeable increase in incivility among state level officials during the months prior to the 2018 election. Incivility peaked among this group as a percentage of monthly tweets during October 2018. Incivility among city and federal officials followed a different trend. Incivility among these two groups as a percentage of monthly tweets peaked during early 2018 and again at the end of the year and beginning of 2019. Readers should be cautious, however, in extrapolating from this trend and keep in mind the previously noted caveat about the limited volume of monthly tweets from these city and federal officials.

Incivility Directed at Tucson-Area Elected Officials

The prevalence of incivility directed at elected officials representing the Tucson area across different levels of government can be found in Figure 6. Although a slightly greater percentage of uncivil tweets were directed at officials in city government than state government, neither exceeded 5% of all tweets directed at these two groups. Federal officials, in contrast, were at least three times more likely to receive uncivil tweets than other levels of government. Fifteen percent of all tweets directed at Tucson area officials in federal government contained incivility.

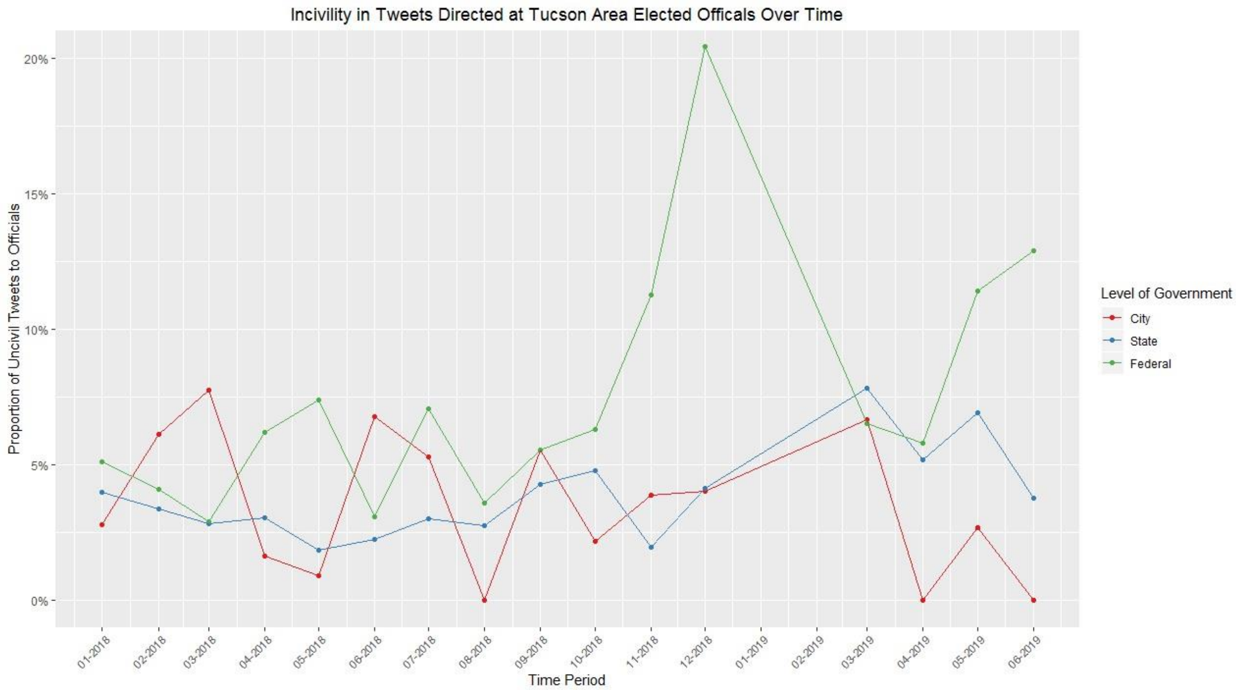
Figure 6. *Proportion of Tweets Containing Incivility Directed at Elected Officials*



As with tweets by local officials, considering the raw number of uncivil tweets directed at officials representing the Tucson area is worthwhile. Of the more than 3,000 tweets directed at city officials, approximately 150 contained incivility. For officials in state government, approximately 725 of the more than 22,000 tweets they received were uncivil. Federal officials received over 7,000 uncivil tweets among more than 45,000 total tweets.

Figure 7 displays the proportion of uncivil tweets directed at officials representing the Tucson area over time. Perhaps the most obvious trend is the spike in incivility directed at federal officials in the final months of 2018. Twenty percent of all tweets directed at federal officials during December 2018 contained incivility. With the exception of November, the proportion of uncivil tweets directed at this group from January through October 2018 did not exceed 7.5%. The proportion of uncivil tweets directed at state level officials were highest during early 2019 and peaked during March. Uncivil tweets directed at city officials were most variable, rising and falling every few months. However, this final trendline should be interpreted cautiously due to the relatively small total number of tweets each month directed at local officials representing the Tucson area.

Figure 7. *Incivility in Tweets Directed at Elected Officials Over Time*



Discussion and Conclusion

Incivility is routinely cited as a problem by large swaths of the public. In a political environment in which messages are increasingly transmitted by and at public officials via social media, it is important to understand how incivility circulates. By using trained human coders to validate a machine learning algorithm that analyzed more than 90,000 tweets to detect name-calling, this white paper provides insight into the presence of incivility by and directed at elected officials representing the Tucson area across three levels of government.

Summary of Key Trends

Our findings illustrate several key points. First, public officials representing the Tucson area rarely used incivility on Twitter. State level officials (1.5%) were the lowest in use of incivility and federal officials the highest (3%), indicating in all cases a hesitance to employ incivility. This finding is consistent with the idea that, notable exceptions aside, elected officials are generally disinclined to engage in public incivility. Especially at the local level, where fewer votes decide election outcomes, there is little to be gained and much to be lost by making direct personal attacks.

Second, incivility directed at public officials varied dramatically by level of representation. Most notably, tweets directed at federal officials employed name-calling 15% of the time—more than three times as often as at the city and state level, neither of which exceeded 5%. Two points should be made about this trend. For one, it signals that Tucson-area political Twitter use is somewhat more civil than at least one other major local discussion forum for two of the three levels of government. In our previous work, we found that 14% of comments to articles on the *Arizona Daily Star* website included name-calling. [2] For city and state levels of government, then, the Tucson area Twitter-using public is more civil than commenters to the *Arizona Daily Star*. Additionally, the discrepancy between incivility directed at federal officials and state or city officials may well indicate that incivility increases when people perceive greater distance between themselves and their elected officials. State and city level officials are more accessible to the average citizen, and thus potentially less appealing as targets of incivility. The typically weightier stakes of national level politics may also factor in here, with citizens lashing out with name-calling around particular national issues about which they have strong feelings.

Finally, our results provide insight into the timing of Tucson-area incivility on Twitter. Keeping in mind the aforementioned caveats about interpreting our over-time findings, the key point that can be made is that incivility among state level officials increased in advance of the election. Here we see potential evidence of the rising tensions that can occur as an election draws near, with a rare behavior such as name-calling arising more often when the stakes feel greater.

Recommendations

The findings summarized above provide the opportunity to derive two recommendations for managing online incivility in the Tucson area. First, granting that overall incivility in this context is low, we nonetheless suggest that federal officials could learn from the Twitter behavior of state level officials. State level officials used half the amount of incivility that federal officials did, presumably with no loss of useful information transmission. In other words, state level officials were able to pursue the same objectives as federal officials without the need to use incivility as frequently. Second, citizens using Twitter for political exchanges should be aware that they may engage in more name-calling with federal officials than they would otherwise. Knowing this might be cause for reflection: If people are not comfortable employing incivility when addressing a mayor or city councilperson, should they feel the same discomfort addressing a U.S. Senator that way?

References

[1] Poll results are available at https://www.webershandwick.com/wpcontent/uploads/2018/06/Civility-in-America-VII-FINAL.pdf

[2] Study is available at https://onlinelibrary.wiley.com/doi/10.1111/jcom.12104

[3] Study is available at https://journals.sagepub.com/doi/10.1177/0093650217699933

[4] Report is available at https://mediaengagement.org/research/new-approaches-tocomment-moderation/

[5] Reference is available at https://www.aclweb.org/anthology/S19-1031.bib

[6] Reference is available at https://www.aclweb.org/anthology/N19-1423.bib

Author Notes

The authors would like to thank Dr. Talia Stroud for serving as an external reviewer for this white paper and Dr. Farig Sadeque for cleaning some of the training data we used in developing the machine learning classifier.