Studying political communication on Twitter: the case for small data
Joyojeet Pal and A’Ndre Gonawela

Big data has dramatically changed the study of political communications online as researchers access massive feeds of data on social media behavior, networks, and language. However, the nature political communication remains inherently message-driven, where the composition, timing, and metaphor are necessary components of the overall message. This article surveys research on political communication on Twitter and classifies it into seven subjective domains of research. The methodological approaches that have been applied toward these domains include quantitative technique studying the size, shape, profile of the networks and their nodes; large-scale data mining techniques applied to study the contents of Twitter messaging; and qualitative methods for in-depth study of messages. Showing that qualitative research methods have extended our understanding of political communications domains, we propose that small data approaches, through interpretive analysis and commentary by human readers, can be coupled with large-scale data analysis for deeper, contextual understanding of political messaging.

Address
University of Michigan, 105 S State Street, Ann Arbor, MI 48104, USA
Corresponding author: Pal, Joyojeet (joyojeet@umich.edu)

The problem of context
On May 5, 2016, Donald J. Trump, who had not yet secured the Republican presidential nomination, tweeted “Happy #CincoDeMayo! The best taco bowls are made in Trump Tower Grill. I love Hispanics!” alongside a picture of himself eating a taco bowl. The message became one of Trump’s most retweeted, with a combined reach of over 200,000 likes and retweets.

The tweet also immediately made the news, like several other of Mr. Trump’s social media messages in the preceding months, and as with many of those tweets — was read differently by different commentators. Trump had just ended a heated primary campaign, with the withdrawal of his key opponents in the preceding days. Reading the tweet, however, gives us little insight into what drove it viral. Mention of an ethnic category — Hispanics; a festival — Cinco de Mayo; a popular food — tacos; a recognized edifice — the Trump tower; or any combination of those factors could have influenced the relative popularity of the message. Several of these factors could be captured by machines; however, a human commentator versed in U.S. political news could examine this tweet through the lens of Trump’s inimical characterizations of Latinos in the preceding months, the prevalent discourse of The Wall and immigrants, and read into the nuances of the composition.

Big data analysis techniques allow for any combination of words, hashtags, and punctuation to be pulled for analysis in combination with other information about the social media artifacts. Small data, on the other hand, refers to individual artifacts that can be interpreted through deep human reading — whether a book, a film, an observation, or in this case a tweet. Small data, in this sense, beyond its empirical definition, refers to a methodological approach of human qualitative reading.

Read literally, Trump’s tweet is a festival greeting, followed by a plug for food at his restaurant. Read as innuendo, it mocks an ethnic group. Read for its rhetoric style, it sounds like awkward phrasing. Likewise, a retweet could be an expression of approval, irony or execration. While one may never arrive at a definitive meaning on the semantics of the composition or its intent, the human reader offers interpretation, which is at the heart of all political communication.

There is a long history of qualitative, interpretive research that drives theory in political outreach, despite the range of perspectives on what constitutes political communication [1]. The recent growth of social media use for direct outreach by politicians as well as the emergence of Twitter and Facebook among others as part of the public sphere for political discourse has raised new questions for how political communication can be studied. There are many post-facto interpretations of what makes a politician’s message appealing. Politicians benefit from having direct access to communication channels without mediation by the mainstream media, and when such messages find wide purchase, it offers a powerful form
of endorsement [2**]. The growing reliance on social media for political outreach [3] has led to the increasing personalization of politics [4,5], a facet of political communication that lends itself naturally to qualitative, interpretive analysis [6]. The online behavior of politicians like Donald Trump, who have rewritten the rules of media engagement, shows that it is not just in the reverberation of individual messages through an online space, but the very composition of political communication that needs careful study. These messages say something about both the politician communicating and the citizenry reacting.

Three broad methodological approaches, and combinations thereof, have been used to interpret political tweets. First, the network traversed by the tweet is a source of data — we can quantitatively study the size, shape and characteristics of the network [7**], as well as the individual nodes that populate it [8,9]. Second, the language and tone of the messages, when aggregated across a set of users, topics, and classes of messages, can be indicators of the appeal of the message on social media in general [10] or a means of learning more about the preferences of the people who populate the network [11]. Finally, contextual, qualitative analysis of tweets has been used to explain communication within the larger social and political setting in which it exists [12*].

Much work in social media studies uses the first two methodologies, which are broadly within the frame of big data analytics. Traditional political science, typically concerned with issues of electoral outcomes, has not engaged with qualitative analysis of political communication, consequently most ‘small data’ qualitative analysis has taken place in media theory, or in niche circles within media studies. While these approaches are distinct in the kinds of questions they answer, as we see with a large number of works in this field (Table 1), there is a significant amount of cross-referencing across methodologies and domains.

### Typologies of political social media studies and their research approaches

The matrix in Table 1 shows the themes of political communication-related studies on Twitter. Studies of subjects of discourse have identified or analyzed the range of topics discussed directly by or about certain politicians or parties. Qualitative approaches have explored the discursive style of political public relations [14]; semantic analyses have helped understand the topical spread of political speech, whereas network dissemination studies have shown distinctions between what appeals to citizens on the ground versus online [18].

Research on polarization and preference identification has examined ways in which Twitter can lend insight into the political preferences of stakeholders, their inclinations toward certain media, or their levels of ideological polarization. Qualitative analyses offered descriptive understand of the language of polarization and ideological

### Table 1

<table>
<thead>
<tr>
<th>Methodological spread of studies of Twitter and political communication</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject</strong></td>
</tr>
<tr>
<td>Subjects of online discourse</td>
</tr>
<tr>
<td>Communicative style</td>
</tr>
<tr>
<td>Stakeholder likelihood of social media participation</td>
</tr>
<tr>
<td>Network Influence</td>
</tr>
<tr>
<td>Electoral outcome</td>
</tr>
</tbody>
</table>
| Meta-studies                                               | Boulianne [61], Jungherr [62**] | Chung and Mustafaral [63] |}
affiliation [22], whereas linguistic analyses offer indicators of people’s politics [23], and the network characteristics help understand political homophily [26].

Studies of communicative style explain the ways in which politicians, parties or their supporters self-represent or craft a discourse of an issue or opponent online. Much work in this space is close-read qualitative discourse analysis of personal style of political agents on social media [28], data-mining to detect figures of speech [33], and network studies that examine the role of a politician’s discursive style and its impact on the politician’s acceptance among audiences [34*].

Studies of stakeholder participation have considered various characteristics of political actors and how these relate to their participation online. These include qualitative studies of politicians and commentators’ engagement with mainstream media coverage [36], semantic analyses of the textual content and emotional tone of politicians across sub-regions on social media [37], and demographic analyses to study politicians’ likelihood of participating on social media [41].

Studies of network influence have examined the interconnections between various actors on Twitter that help understand the spread of messages and influence online. These studies include interview-based analysis of motivations and online style of political network influencers [67], coding of messages to identify the framing of issues in social networks, and studies of the dynamics of temporal user influence across various topics [51*].

Meta-studies seek patterns in the ways certain subject matters or methodologies are employed in studies of social media and mainstream politics. Several important meta-studies have tackled one or another specific sub-domain, such as electoral prediction [65**], or spanned wider disciplines of scholarly progress on various facets of political communication [62**].

Finally, on studies of electoral outcomes, researchers have sought to look at ways in which ‘Twitter activity relates to political candidates’ or parties’ performance in elections. This sub-field has had some of the most contested findings within the realm of social media research [65**,68]. Studies have been dominated by sentiment analysis approaches to study factors such as valence [54] or frequency of mention [60] to examine likelihood of electoral outcome, though researchers have also sought to understand the likelihood of electoral success by profiling various political actors and their social media performance [33].

The challenges with predicting this critical outcome variable have been a defining element of social media studies of political communications, with repeated erroneous signals [58,63], despite significant investment into supervised machine-learning techniques and sentiment analyses [57,69]. And yet, since political communication through social media is clearly here to stay, the broader impacts on political culture are important subjects of study.

The case for small data
Online political communication as a field of real-world practice is booming, with politicians all over the world, at various levels of government, investing heavily in directly managing or outsourcing their social media output. And while President Trump’s has been an important case study for its colorful use of social media, his use of Twitter is by no means unique. Political speech is known for its heavy use of innuendo and figurative messaging, which has presented a notoriously intractable problem for automated semantics analysis [70]. The gold standard remains a human reader.

Political messages are by nature highly nuanced and are driven by marketing a specific crafted message [71]. The change of the medium from public speeches or televised appearances to social media does not change the fundamental nature of political outreach [15], where the message serves as a means of telling a story as well as a means of reinforcing a candidate’s credibility [72]. The actual practice of political communication is consequently very detail-driven, and while major campaigns are frequently staffed by entire teams paying attention to facets such as the shape and size of the online network, the influencers, and the means for online propagation of messages, many studies have shown [73,74,75*] there is likewise evidence that language is carefully used and framed [76,77].

The challenge for the study of political communication in the academy is the distance between the intellectual communities working on different facets of this work. This is partly motivated by the differing questions that drive the various communities, and often the monocultures that inhabit methodological silos that are primarily trained to evaluate similar research. While the draw for ‘big data’ has had important consequences on our ability to understand several key aspects online communications, the intractability of some of the big questions into bite-sized variables necessitates innovation in research methods. Likewise, scholars working with qualitative, interpretive methods have stayed within their communities, despite evidence presented here that both sets of approaches have been widely applied to similar questions. There is a need for scholars with different perspectives on communications to open up to different approaches those outside of the affordances of their methodological worlds.

The success of political campaigns on Twitter such as those of Barack Obama, Recep Tayyip Erdoğan, Donald
Trump, and Narendra Modi all feature a mix of factors wherein both the nature of social media networks and the text of what was being said on them were responsible for their eventual success with voters. Presenting well-integrated studies of political communications requires not just that individual authors expand their methodological lens or reach out to collaborators with different perspectives, but that research communities as a whole open themselves to diverse approaches and the benefits they offer.

While machine-managed analysis of political communications undoubtedly offers cheap, expansive insights into political communications, there will always be important political artifacts — whether key actors, issues, or messages — that merit an in-depth human commentator. Small data do not replace big data — they make each other more complete.

Conflict of interest statement
Nothing declared.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest


This article examines the various ways in authorship, attribution, and communicative fidelity are negotiated in diverse was in Tweeting and retweeting.


This article studies information diffusion online to support the notion of complex contagion of hashtags on politically controversial topics — they persist online and are likely to be spread by users.


This article examines online behavior against political style to propose that UK Liberal Democrats use weak interactivity online as their goal is to promote ideas than co-create a product with citizens.


Small data on political communications Pal and Gonawala


This article studies Donald Trump’s 2016 campaign uses both conventional public relations to find space on traditional media and ‘tweetstorms’ to create stories on social media.


This article argues that large scale action networks are influence by collective action associated with high organizational resources and connective action based on sharing across media networks.


This paper finds that influence on twitter needs to be cultivated through concerted effort rather than by accident, and that influencer play a role influencing various topics rather than just one.


This paper provides a wide literature review of the various methodological approaches to studying the use of Twitter in politics.


This paper provides a meta review of the literature on using Twitter feeds for electoral prediction, highlighting the shortcomings and various approaches.


This book highlights the changes in political campaigning since the 1990s including the use of mass-mediated internet campaigning, networked campaigning, and data-driven campaigning.
