Original Manuscript

Measuring political legitimacy with Twitter: Insights from India's Aadhaar program

Laura C Mahrenbach (iD)
and Jürgen Pfeffer (iD)
Technische Universität München, Germany

Abstract

As emerging powers forge ahead with big data initiatives, questions arise regarding the implications of these programs for governance in the Global South more broadly. One understudied aspect deals with how actors attribute legitimacy to governments' big data activities. We explore actors' agency in one crucial case: the world's largest demographic and biometric data program, India's Aadhaar. Analyzing roughly 250,000 tweets collected in the first 10 years of Aadhaar's operation, we find that both normative acceptance and cost–benefit calculations are crucial for legitimacy attribution. This finding challenges mainstream theoretical approaches, which prioritize normative factors and often fail to examine how normative and material factors interact during legitimacy attribution. In addition, our study demonstrates a new, mixed-methods approach to measuring legitimacy attribution using Twitter data, which overcomes traditional challenges. As such, we underline the viability of Twitter data as a tool for social measurement.

Keywords

Aadhaar, big data, India, legitimacy, social media, Twitter

Introduction

As some of the largest and most populous countries in the world and with a growing interest in big data and digital governance (Bauman et al., 2014), emerging powers' digital visions and activities are likely to affect how big data is used in the countries of the Global

Corresponding author:
Laura C Mahrenbach, Professorship of Computational Social Science and Big Data, School of Governance, Technische Universität München, Richard-Wagner-Str. 1, 80333 Munich, Germany.
Email: laura.mahrenbach@tum.de

South more broadly (see Bader, 2019; Mahrenbach and Mayer, 2020). While authoritarian states, such as China, have moved swiftly to establish uses for big data at home and contribute to data governance abroad (Arsène, 2012; Zheng, 2013), similar efforts by democratic states, such as India, have been hindered by the complications of domestic politics involving 1.3 billion people (Ebert and Maurer, 2013; Ghose, 2015). For emerging powers' data-driven programs to be more broadly applicable, governments must account for such contextual differences (Taylor and Schroeder, 2015). Likewise, scholars must move beyond viewing Southern citizens as beneficiaries of imposed development agendas to examining how they are themselves affecting big data outcomes (Arora, 2016; Kidd, 2019).

Joining a burgeoning literature examining actor agency vis-à-vis digital development (Jiang et al., 2019; Jiang and Okamoto, 2014), we contribute new insights by examining how one politically relevant group of actors—tweeters—are legitimating India's most recognizable big data policy initiative: the Aadhaar program. Aadhaar, which is run by the Unique Identification Authority of India (UIDAI), began operating in 2009. It is the world's largest biometric and demographic database. Indian residents register for an Aadhaar number by providing UIDAI or a certified affiliate with necessary biometric (e.g., fingerprints) and demographic (e.g., birthdate) information. They can subsequently use their Aadhaar number for diverse purposes, among them, providing medical information, identifying themselves, or transferring government subsidies into bank accounts (Abraham et al., 2018). As of June 2021, there were almost 1.3 billion distinct sets of information stored on UIDAI's servers and Aadhaar numbers had been used to authenticate over 56 billion transactions (Unique Identification Authority of India (UIDAI), 2019).

The usefulness of national identification projects like Aadhaar depends on widespread adoption, and adoption requires public acknowledgment that the program is legitimate (Bhat, 2013). As such, Aadhaar is a crucial case for understanding how subsets of a country's citizens attribute legitimacy to big data projects. In this study, we build on calls in the international relations (IR) literature to study legitimacy within the context of, rather than distinct from, the cost–benefit calculations of the actors attributing legitimacy (Hurd, 2019; Schlipphak, 2015). Drawing on a legitimacy typology from sociological institutionalism (Suchman, 1995), we investigate the role of normative and material factors in attributing legitimacy to Aadhaar and UIDAI. We do this using an innovative mixed-methods approach which employs quantitative sentiment analysis to identify trends in social media data and then qualitative directed content analysis to contextualize those trends and evaluate legitimacy attribution. The dataset itself comprises roughly 250,000 tweets collected in the first 10 years of Aadhaar's operation. We find that both normative acceptance and cost–benefit calculations are crucial for legitimacy attribution. This underlines the analytical value of complementing studies seeking to determine the causal impact of normative versus material factors with studies seeking to determine how these factors interact during legitimation processes. Moreover, while we are not the first to use Twitter data to measure public trust and government legitimation efforts (e.g., Brooker et al., 2018; Chatfield et al., 2015), our study demonstrates how employing a mixed-methods approach can overcome the challenges faced in applying social media data to political settings, including a tendency to overlook contextual details and difficulties linking social media data to concepts (Ruths and Pfeffer, 2014; Calderon et al., 2015).

Theoretical framework

In Easton's (1975) classic definition, something is legitimate if it conforms to an actor's "sense of what is right and proper in the political sphere" (p. 451). While recognizing the importance of structural factors in setting the boundaries for "what is right and proper" (Bernstein, 2011; Reus-Smit, 2007), we seek to shed light on how Southern citizens attribute legitimacy vis-à-vis big data projects. Consequently, we assume an agency perspective toward legitimacy. In so doing, we evaluate the evolving social attribution of legitimacy to Aadhaar and UIDAI rather than either's capacity to meet contextually determined criteria (Schneider et al., 2010).

We conceive an object's legitimacy as the sum of top-down efforts to cultivate legitimacy, for instance, by politicians, and bottom-up attribution of legitimacy by the people affected by an organization or its rules (Gronau and Schmidtke, 2016). *Legitimacy cultivation* can occur through legitimacy claims, that is, discursive justifications by an institution which aligns the institution with standards of appropriateness (Reus-Smit, 2007). In addition, actors can resort to symbolic actions, such as affiliating themselves with legitimate institutions, to increase acceptance of their own legitimacy (Rao, 1994; Zott and Huy, 2007). *Legitimacy attribution*, in turn, has been linked to the quality of outcomes an institution delivers, the extent to which the public can participate in decision-making, and the quality of decision-making processes (Schmidt, 2013). Actors can signal legitimacy attribution by complying with an organization's policies or by vocalizing support for the organization or its principles (Schneider et al., 2010).

While early contributions depict legitimacy as essentially dichotomous—either something is legitimate or it is not—recent contributions speak of gains and losses of legitimacy. This does not alter the bifurcated nature of legitimacy but, rather, highlights that organizations' legitimacy can be more or less certain over time due to changes in the target audience or in the resonance of organizations' legitimacy claims (Deephouse and Suchman, 2008). As Keohane (2011) argues, legitimacy is "a matter of degree [. . .] we do not merely assess whether it is legitimate but how far above or below the threshold of legitimacy it falls" (p. 101). Moreover, different legitimation strategies target different audiences and types of legitimacy (Gronau and Schmidtke, 2016; Zott and Huy, 2007).

While much previous work has focused on the normative criteria which contribute to legitimacy attribution (Reus-Smit, 2007; Schneider et al., 2010), there are analytical dangers in examining legitimation processes in the absence of actors' cost–benefit calculations (Hurd, 2019). For instance, scholars may fail to recognize that, in attributing legitimacy to an institution, actors create and/or preserve the material benefits those institutions provide. Similarly, they may oversee that the material benefits received from an institution can encourage actors to view an institution's normative framework as more legitimate. In other words, normative and material factors *interact* in legitimation processes. While not entirely absent from studies of legitimacy (e.g., Zürn and Stephen, 2010), this interaction is often overlooked and/or left unexamined in an attempt to identify causal factors affecting legitimacy attribution (e.g., Schmidtke, 2019). Doing so increases understanding of why legitimation occurs. However, it restricts understanding of the legitimation process.

Table 1. Legitimacy types (compiled from Suchman, 1995).

Legitimacy type	Results from . . .	Focus	Factors
Pragmatic	". . . self-interested calculations of an organization's most immediate audience" (578)	- Impact on legitimating audience - Extent of audience involvement in decision-making - Representation of audience preferences in policies	Material
Moral	". . . audience's evaluation that activities are "the right thing to do"	- Intended outcomes - Decision-making, managerial or operational processes - Structure - Leadership	Normative
Cognitive	. . . activities making sense to its audience	- Taken-for-grantedness, that is, whether activities are accepted/expected by the audience - Comprehensibility, that is, whether activities accord with broader social values and audiences' daily lives	Material and normative

Addressing this gap, we employ a schematic from sociological institutionalism that acknowledges the role of normative and material factors in legitimation processes. In his seminal study, Mark Suchman (1995) identifies three types of legitimacy. *Pragmatic legitimacy* refers to legitimacy attribution resulting from the "self-interested calculations of an organization's most immediate audience" (p. 578). This can refer to how an organization's policy affects that audience and to what extent the audience feels they are involved in and/or their preferences are represented in decision-making and policies. This type of legitimacy explicitly includes actors' material calculations by acknowledging the link between an institution's effectiveness in achieving stated goals and the audience's self-interested evaluations of its legitimacy (Lipset, 1959). *Moral legitimacy* exists when an organization's audience considers its activities "the right thing to do." This can refer to an organization's intended outcomes, processes, structure, or leadership. Here the focus is purely on the audience's normative perceptions of an organization. Finally, *cognitive legitimacy* refers to the extent to which an organization's activities make sense to its audience, whether because those activities are accepted and expected by the audience (taken-for-grantedness) or because the organization's account of its activities accords with broader social values and the audience's daily lives (comprehensibility). Unlike pragmatic and moral legitimacy, cognitive legitimacy is passive: people do not actively decide whether something is cognitively legitimate but, rather, fail to oppose something because it fits with their expectations or seems normal. As such, cognitive legitimacy encompasses both a material and a normative rationale: what is "normal" and whether something fits one's expectations can depend on one's self-interested calculations, one's perception of rightness, or some combination of both. Table 1 provides an overview of these types of

legitimacy. The next section summarizes how the Government of India (GOI) has sought to cultivate each type of legitimacy in relation to Aadhaar and UIDAI.

GOI legitimacy cultivation to date

Following implementation, it quickly became clear that gaining and maintaining public acceptance would make or break Aadhaar (Klitgaard, 2011; Polgreen, 2011). Without wide-scale adoption, the program had little hope of meeting the government's stated goals of social protection and financial inclusion (Abraham et al., 2017). As such, Indian residents, defined by UIDAI as nongovernmental actors residing in India for at least 182 days in a given year (Government of India, 2016), are the primary target audience for efforts to cultivate Aadhaar's legitimacy.

Public support initially appeared tenuous.[1] UIDAI had to convince residents to participate without any reliable information about how much the program would cost, nor how benefits would be distributed (Srinivasan and Johri, 2013). This, coinciding with early speculation—later confirmed—that people would sign up for Aadhaar not because they found it useful but because they feared not having it would be costly (Sathe, 2014). Initial support for the program by nongovernmental organizations (NGOs) also quickly evaporated as NGOs prioritized own goals over the politically complicated task of finding solutions to enrollment issues faced by India's poorest residents (Baxi, 2019). Moreover, Aadhaar faced numerous legal challenges, 30 of which were ultimately considered by the Supreme Court (see Abraham et al., 2018). These conditions are, at best, a weak starting point for legitimacy attribution by the Indian public, tweeters included.

In this context, UIDAI adopted a two-pronged strategy to cultivate legitimacy. First, UIDAI established and communicated diverse uses for Aadhaar numbers. Initially, Aadhaar was promoted as part of GOI's broader efforts to increase financial inclusion. As UIDAI's boss argued, Aadhaar would help the "invisible millions [. . .] take their rightful place as part of the formal financial sector" (Nilekani and Shah, 2016: 5) and help "masses of India's poor" access social services (quoted in Sathe, 2011). For bankers and businessmen, such rhetoric underlined the pragmatic legitimacy of the program, promising them more customers—and more profit—in return for relaxing rules and updating technology to accommodate new clientele. The rhetoric makes a pragmatic legitimacy claim to poorer residents too: UIDAI would focus on their interests and Aadhaar would bring material benefits. Over time, uses for Aadhaar multiplied, offering new pathways for cultivating legitimacy. For example, linking Aadhaar to e-Know Your Customer (eKYC) authentication enabled quicker opening of bank accounts and immediate verification of welfare benefits (Banerjee, 2015). This symbolic action demonstrated UIDAI's capacity to follow-through on its earlier promises of financial inclusion (see Zott and Huy, 2007) and underlined Aadhaar's pragmatic legitimacy claim vis-à-vis poorer residents. Expanded uses of eKYC, such as performing background checks on employees (Abraham et al., 2017), made pragmatic and cognitive legitimacy claims vis-à-vis wealthier residents too by highlighting Aadhaar's benefits and demonstrating the ease with which it could be incorporated into daily activities.

Second, UIDAI sought to make Aadhaar services accessible to diverse populations within India to maximize enrollment and ensure visible project successes. Regarding the

former, people were allowed to "enrol in any manner and at any location that was most convenient," thereby promoting Aadhaar's cognitive legitimacy (Nilekani and Shah, 2016). Regarding the latter, Aadhaar verification via micro-ATMs, that is, cell phone and fingerprint readers distributed to local leaders, enabled populations geographically distant from banking facilities to use banking services (Klitgaard, 2011). This represents a pragmatic legitimacy claim, demonstrating customers' and businesses' gains (e.g., access to banking services and more customers, respectively) from using the program. Efforts to demonstrate how Aadhaar could improve existing government programs (Abraham et al., 2017), such as enabling access to food subsidies outside one's home state, made a similar claim to poorer residents. Finally, convenience apps, such as e-Sign, which allows Aadhaar authentication of digital signatures, displayed the cognitive legitimacy of the program for Indian professionals.

Cultivation of the program's moral legitimacy is notably absent from these two strategies but was evident in other GOI activities. For instance, inviting Nandan Nilekani, the co-founder of Infosys and one of India's most successful entrepreneurs, to head UIDAI targeted moral legitimacy by associating the program with a well-respected, high-profile leader (Gerdeman, 2012). Similarly, advocates frequently spoke of using Aadhaar to fight corruption, for example, by cutting out middlemen who siphon away government benefits (see Daugman, 2014). Such claims highlighted the program's capacity to deliver the "right" outcomes to residents. The next section discusses the challenges and opportunities of using social media data to measure the success of these initiatives.

Measuring legitimacy attribution with social media data

We join a growing group of scholars using social media data to evaluate public trust and legitimacy attribution (Calderon et al., 2015; Chatfield et al., 2015; Liu and Lee, 2014). Social media offers new opportunities for political discussion and thus new means of affecting legitimation processes (Brooker et al., 2018; Etter et al., 2018; Maireder and Ausserhofer, 2014). In addition, a recent study found that social media activity reinforces, rather than disrupts or replaces, ongoing legitimation processes (Poell, 2020), thus making it a valuable—and accessible—tool for measuring legitimacy attribution.

Nonetheless, challenges abound, two of which seem relevant. First, many members of the voting population do not use social media, and this is particularly true in the Global South (Arora, 2016; McCarthy, 2016). This raises questions regarding the representativeness of social media data for social measurement. We address this challenge by validating our findings using traditional measures of political legitimacy (see Schlipphak, 2015). The results of this comparison, detailed in the supplemental online appendix, confirm the validity of our measurements. This provides new support for arguments that social media analyses can accurately represent public discussions of a political topic despite not fully representing the population itself (Schober et al., 2016).

There is substantial literature on Twitter's role in shaping and facilitating social movements in India (e.g., Barker-Plummer and Barker-Plummer, 2017; Belair-Gagnon et al., 2013; Poell and Rajagopalan, 2015). However, Twitter's political relevance extends beyond social movements, with the platform playing an increasingly important role in political discussions of national topics such as Aadhaar (Khursheed, 2014; Panda et al.,

2020; Rajput, 2014; Rodrigues and Niemann, 2017; *News India Times*, 2017). In fact, Indian citizens have been among the most likely worldwide to discuss politics via social media since at least 2011 (Pew Research Center, 2012). This tendency was amplified by the extensive use of the platform during the General Election in 2014 and continues today (Mohan, 2015; Panda et al., 2020; Rodrigues, 2020). Admittedly, use of Twitter has been largely localized to specific geographical regions, for instance, in coastal areas and along the northern border (Leetaru et al., 2013), and digital infrastructure challenges continue to prevent many citizens from using the platform (Banerjee, 2015). However, studies indicate that both urban and rural populations use the platform for political communication, with both male and female users participating, if not in equal measure (Arya, 2013; Barbera and Rivero, 2015; Mohan, 2015).[2]

Moreover, current events suggest that tweeters are a politically important segment of the Indian population. In April 2018, for instance, GOI issued a much-criticized tender seeking companies to build a social media monitoring platform to "gauge the sentiments amongst netizens' related to 'schemes run by the GOI'" (Government of India, 2018: 28). This was the culmination of several attempts to monitor public sentiment via social media (Ganjoo, 2018), and suggests GOI values input from digital platforms. Public statements by high-profile politicians confirm this expectation. For instance, in 2015, Prime Minister Narendra Modi equated social media posts with voting, saying, "We used to have elections every five years, now we have them every five minutes" (Constine, 2015). As such, in analyzing legitimacy attribution by tweeters, we provide a window into legitimacy attribution by a politically important demographic.

A second issue plaguing social media analysis of political legitimacy is the difficulty of connecting empirical data with broader concepts such as authority and governance. Some see this as an "inherent limitation" of social media data (Calderon et al., 2015: 1683), while others simply attribute it to the lack of a "computable definition of legitimacy" (Liu and Lee, 2014: 113). We join the optimists in suggesting that social media data is capable of measuring even broad concepts like legitimacy. Our approach uses quantitative sentiment analysis to identify trends in Twitter data and then contextualizes those trends via qualitative directed content analysis. These steps are described briefly below.

Sentiment analysis of daily tweets (2009–2019)

In line with previous studies examining legitimacy using social media data (e.g., Calderon et al., 2015), we first performed a sentiment analysis to identify changes in tweeters' feelings vis-à-vis Aadhaar and UIDAI in the first 10 years of Aadhaar's operation. To do this, we collected tweets separately for every calendar day from 1 January 2009 to 31 December 2019 using the search function on Twitter's website. Search terms included "Aadhaar," "UIDAI," and "Unique Identification Authority of India." After removing sponsored tweets, 216,351 tweets remained. Figure 1 shows the average number of tweets per day for every year. We recognize that this collection approach does not allow us to get all historic tweets related to Aadhaar. We are also aware that we do not know the actual sampling procedure that Twitter applies (Ruths and Pfeffer, 2014). However, since we are not interested in overall numbers but in tweet content, we believe our data are still sufficiently representative of the overall Twitter activity.
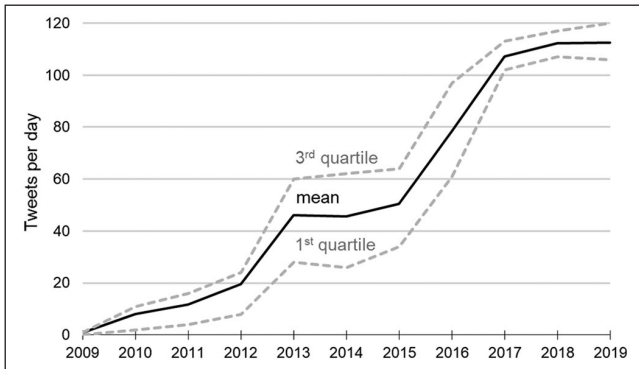
Figure 1. Average tweets per day (mean, 1st quartile, 3rd quartile) per year.

Before further analyzing the dataset, we removed all URL links from the tweets. We then utilized Linguistic Inquiry and Word Count (LIWC) 2015 (Pennebaker et al., 2015) to calculate sentiment scores per year. LIWC is a keyword-based approach to sentiment analysis which calculates the "sentiment" of a document by counting how frequently words from word lists associated with different emotions appear in a text. It has been widely used in analyzing political phenomena (e.g., elections; see O'Connor et al., 2010) and performs well in comparison to other sentiment analysis approaches (Gonçalves et al., 2013). It is especially useful for our study because the keyword-base enables detailed analysis and interpretation. Another advantage is classification into a wide variety of emotional categories rather than just positive and negative, allowing us to differentiate in particular among different negative emotions (e.g., anger and anxiety). Since the vast majority of this dataset consists of English-language tweets, we only analyzed sentiments based on the English-language LIWC 2015 dictionary.

Directed content analysis

As Grimmer and Stewart (2013) point out, while useful in identifying trends, dictionary-based methods of quantitative text analysis are incapable of contextualizing words and thus require external validation and explanation. Qualitative text analysis is helpful in this respect (Brooker et al., 2018). Turning to this task, we collected data for two time periods, each corresponding to developments in the Aadhaar program which were likely to provoke political discussion and, thus, tweets. Period 1 (2009–2010) reflects the economic and political machinations necessary to get Aadhaar and UIDAI established. A total of 2935 tweets referencing "Aadhaar" or "UIDAI" were extracted from an archive of a 1% random sample of all tweets from Period 1. Of these, 620 were deemed relevant.[3] Period 2 (2017–2018) represents a crucial period for both Aadhaar and UIDAI, as it corresponds to the Indian Supreme Court's consideration of whether Aadhaar could be made mandatory as well as the legality of the UIDAI and its activities. Data for this period were collected via a Python program searching for the same keywords and using the Tweepy API. This produced a relatively larger tweet population from Period 2 (32,589

Table 2. Measures of change.

Tweet type	Issue areas	Policy visions
- Informational	- Economy	- Government effectiveness
- Promotional	- Politics	- and efficiency
- Discussion	- Security	- Economic and social
- Joke	- Health	- development
- Irrelevant/	- Education	- Political liberation or
unclear	- Science and	repression
	technology	

tweets), which was inappropriate for manual coding. Consequently, we randomized the population of tweets and analyzed an equal number as in Period 1.[4]

We then performed a directed content analysis of the dataset (Hsieh and Shannon, 2005). Our intention was, first, to determine how discussion of Aadhaar and UIDAI changed from Period 1 to Period 2 and, second, to provide context for the growing dissatisfaction depicted in the sentiment analysis. Coded segments were limited to single tweets. However, any (still active) URLs included in the tweets were also coded, as we assumed users included links because they (dis)agreed with its contents and because the link is a part of the tweet text.

As is evident from Table 2, we coded three measures of change. In line with our agency perspective on legitimacy, these measures reflect both bottom-up (tweet type/issue area) and top-down (policy vision) elements (Gronau and Schmidtke, 2016). *Tweet type* enables us to gauge whether Aadhaar is on tweeters' radar and if so, in what capacity. This is important for evaluating pragmatic and moral legitimacy attribution, which require active knowledge and evaluation by the target audience. *Issue area* measures how and where tweeters expected Aadhaar to affect their lives. This is important for estimating pragmatic and cognitive legitimacy attribution, which requires audiences to connect an organization with their daily lives. Finally, *policy vision* addresses tweeters' awareness of GOI's big data visions, specifically big data as a tool for improving government efficiency and effectiveness, for facilitating development and/or for enhancing political liberation or repression (Mahrenbach et al., 2018). Promoting strategic visions is an overt legitimation strategy and one that can be linked with all three types of legitimacy. Finally, we controlled for the tone of tweeters' evaluations, distinguishing between explicitly negative and explicitly positive tweets. The next section presents our findings.

Findings

Identifying changes in Twitter discussions

Figure 2 presents the results of the sentiment analysis. The unbroken lines visualize the sentiment scores for positive (blue) and negative (red) emotions. To better analyze change over time, we set the scores for each category to 100% for the year 2009. The figure shows clearly that, while positive emotions in tweets about Aadhaar and UIDAI stay at roughly the same level, the score for negative emotions goes up tremendously. LIWC offers three subcategories for negative emotions—anxiety, anger, and
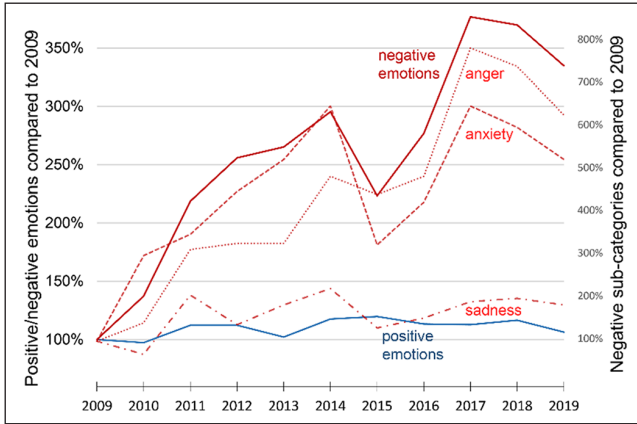
Figure 2. LIWC sentiment scores per year compared to sentiment scores of 2009.

sadness—that can help us better understand the driving forces behind this increase in negative emotions. The dotted and dashed red lines tell us that anger and anxiety seem to be the main cause for negative sentiments toward Aadhaar.

## *Contextualizing changes in Twitter discussions*

Figure 3 presents a chronological overview of coded segments. As is evident, the qualitative analysis confirms the trend indicated by the sentiment analysis.

Figure 4 presents a network of major topics and their co-occurrence in each period. The network from Period 1 is significantly denser than that of Period 2. This indicates a greater number of narratives and more diverse, variable connections among narratives in Period 2. More diffuse discussions could be interpreted as fewer tweeters being "on message," implying GOI efforts to cultivate legitimacy were less successful in Period 2 than Period 1. Figure 4 shows that, while economic discussions of Aadhaar were largely positive in Period 1, these discussions had become largely negative by Period 2. These factors underline the importance of qualitatively analyzing the data, as the network alone can explain neither the dramatic changes in tone nor why the structure of network discussions developed as it did.

Qualitative content analysis discovered three contextual factors relevant for evaluating legitimacy attribution. The first is a *growing dissatisfaction with the technological and privacy aspects of Aadhaar*. Discussion of operational issues increased substantially across the time periods. Tweets included reports of technical and logistical failures as well as UIDAI's responses to residents facing such challenges. Importantly, while tweets from Period 1 are primarily neutral in tone, tweets from Period 2 were nine times more likely to be negative than positive. Period 2 additionally featured a substantial number of "how-to" tweets, that is, tweets asking or describing how to use Aadhaar numbers. Combined, this suggests a high level of tweeter dissatisfaction with program operation after 10 years. Interestingly, we see UIDAI officials engaging with tweeters to overcome
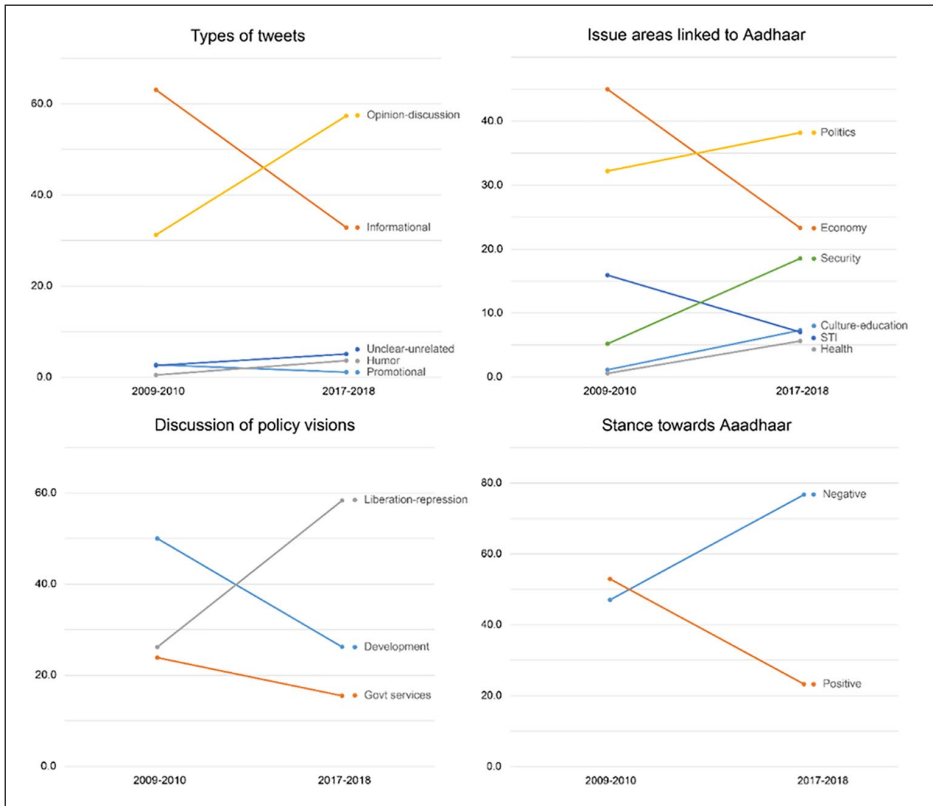
Figure 3. Changes in Twitter discussions of Aadhaar over time.
The vertical axes depict percentages of total tweets in a given category. Lines indicate differences between two points in time, not trends over an entire decade.
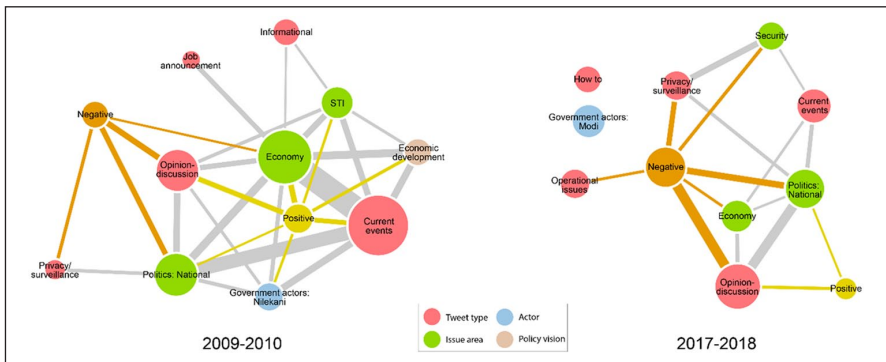
Figure 4. Major topics and their co-occurrence.
Networks represent coded topics and their co-occurrence in tweets. Topics are drawn as nodes if a topic occurs in at least 5% of all coded tweets in time period; node size = occurrence. A link (line) shows co-occurrence of two topics in at least 2% of all coded tweets; link width = co-occurrence.

such issues, answering questions in Hindi and English and asking individualized follow-up questions. This suggests UIDAI's awareness that addressing these issues was important and indicates GOI's continued commitment to cultivating pragmatic legitimacy, here, by minimizing participation costs and signaling responsiveness to user needs.

Discussion of privacy and surveillance is also markedly greater in Period 2. Negative tweets in Period 1 reflect anxiety about Aadhaar's capacity to secure user data (e.g., "Aadhaar is an impending privacy nightmare"), and demand stronger privacy and data protection laws in response. In contrast, negative tweets in Period 2 refer to actual, not potential, security issues and overwhelmingly attribute responsibility to government actors. Moreover, while tweeters continue to call for enhanced privacy laws, there appears to be a new sense that self-policing is the best—even only—solution to Aadhaar's privacy issues. Interestingly, UIDAI appears to promote this idea, for instance, by asking users not to share Aadhaar details via Twitter and reprimanding them when they do. However, much of it appears to arise from frustration with UIDAI's inability and/or unwillingness to protect privacy (e.g., "This Is Called Shamelessness By @UIDAI. We Are Not Responsible For Security Of #Aadhaar"). This evidences greater awareness of Aadhaar's privacy implications concurrent with less faith, even among UIDAI employees, regarding GOI's capacity to mitigate privacy concerns. Unsurprisingly, tweets linking Aadhaar with security were primarily negative and focused on UIDAI's inability to protect citizen data and GOI's (perceived) unwillingness to accept responsibility for data security.

The second contextual finding is tweeters' *ambivalence about Aadhaar's political role*. Politics was consistently linked with Aadhaar in both periods. Discussions in Period 1 connected Aadhaar with Indian legislative and electoral politics, while tweets in Period 2 extended this to the judicial realm. Political figures, including Congress party leader Rahul Gandhi and Prime Minister Modi, were praised and criticized in relation to Aadhaar. However, the majority of tweets linking Aadhaar and politics in both periods were neutral. This exemplifies both the expansiveness of Aadhaar in Indian political life and the tweeting public's ambivalence regarding this development. This ambivalence is striking when juxtaposed with tweeters' strong opposition to Aadhaar's encroachment in their personal lives, for instance, expressing anger and linking to news articles about how mandating Aadhaar would hinder access to medical services.

Ambivalence toward Aadhaar's political role was also evident in the discussion of the program's economic context. Period 1 tweets were overwhelmingly positive vis-à-vis Aadhaar and the economic opportunities on offer, including employment, bank accounts, and government subsidies (e.g., "Here's your opportunity to work with Mr.#Nilekani on #UIDAI—vacancies across India"). Fewer tweeters linked Aadhaar with economic activity in Period 2. When they did, opinions were mostly negative, highlighting Aadhaar's suboptimal social consequences, complaining about the program's invasiveness, and expressing anxiety about government decisions mandating an Aadhaar number. While tweeters generally held economic, not government, actors accountable, there was substantial criticism of the links between GOI and companies. This again illustrates tweeters' inability to evaluate how Aadhaar's political role affected the experienced reality of the program.

Finally, we identify *divergent discussions of big data policy visions*. Previous research found that GOI views big data as a force for liberation, useful for increasing policymaking transparency, advancing knowledge, and protecting human rights (Mahrenbach et al., 2018). A small group of tweeters agreed, underlining how Aadhaar could "reduce fake voters" and increase budgetary transparency, thereby improving the quality of democratic decision-making. Although small in number, these tweets increased substantially in Period 2 and were also more likely to be explicitly positive compared to the more neutral Period 1 tweets. Nonetheless, tweeters in both periods primarily fixated on Aadhaar's potential for repression, especially related to human rights. Most fears centered on misuse of data and a failure to protect residents' privacy. However, some feared even broader violations, with one tweeter arguing that "Aadhar And UIDAI Are Tools Of Social Oppression." This indicates clear divergence between the government's strongly positive vision of big data and tweeters' mostly negative appraisal of how that vision has been realized. It also suggests that suspicion of big data is decoupled from Aadhaar's performance, as tweeters saw the program as a tool for repression in both periods.

This divergence among popular and political visions is even more pronounced in relation to the development and government services visions. In Period 1, tweeters had high hopes for Aadhaar's capacity to improve government effectiveness: Aadhaar could "rule out fraud," "plug leakages in welfare delivery mechanism," and "change India n stop corruption." Comparable statements in Period 2 are essentially non-existent. A similar pattern appears regarding economic development, where hopes were even higher: "#Aadhaar leading a surge in bank accounts"; "UIDAI project will help IT cos innovate"; "Aadhaar offers hope for better job." Similar sentiments were almost absent in Period 2. More poignantly, tweets explicitly stating that Aadhaar would not advance development more than tripled from Period 1 to Period 2. Aadhaar was now "denying food to the most vulnerable" and had become "a tool to exclude." The prominence of empirical examples in Period 2 tweets (e.g., "Elderly woman dies after losing □90,000 to #Aadhaar fraud") further underlines decreased belief in Aadhaar's capacity to fulfill these two visions. The stark change from Period 1 to Period 2 in both cases suggests that experience with Aadhaar affected how tweeters evaluated its ability to fulfill GOI strategic visions.

Evaluating legitimacy attribution

This section links these findings with changes in legitimacy attribution by tweeters between 2009 and 2019. Starting with Aadhaar's *pragmatic legitimacy*, we find little legitimacy attribution in Aadhaar's first 10 years. Tweeters' dissatisfaction with Aadhaar's technological and operational aspects suggests that they saw more costs than benefits arising from the program. After all, promised benefits (e.g., improved background checks) cannot materialize if the technology upon which they depend malfunctions and/ or is unavailable. Tweeters' inability to situate Aadhaar in the domestic political context beyond basic acceptance of its political importance implies a similar incapacity to attribute legitimacy based on cost–benefit calculations.

In contrast, GOI's response to tweeters' frustration with UIDAI's handling of privacy concerns implies some attribution of pragmatic legitimacy. For example, GOI released a

White Paper on data protection in November 2017. This paper explicitly solicited public comment on the construction of privacy laws. It also facilitated participation by providing multiple response channels and includes a brief summary of the White Paper "for those who may not have either the time or the inclination to peruse the contents of the White Paper fully" (Srikrishna, 2017: i–ii). While tweets evaluating UIDAI's handling of privacy issues positively were relatively sparse, they nonetheless quadrupled after the release of report. This underlines UIDAI's agency—and success—in cultivating pragmatic legitimacy, both in reference to Aadhaar (by institutionalizing public input in a new privacy law) and to UIDAI (by reassuming responsibility for privacy protection rather than relying on self-policing).

Turning to *moral legitimacy*, the analysis indicates tweeters attributed little moral legitimacy to Aadhaar or UIDAI across the time periods. Consistent, negative emphasis on Aadhaar's privacy violations indicates initial skepticism and, later, disapproval of Aadhaar's outcomes. References to "data breach" and "a billion identities at risk" affirm that tweeters did not think these outcomes met the moral legitimacy criterion of being "the right thing to do." Tweeters' skepticism toward Aadhaar's capacity to support GOI's development vision of big data also indicates a lack of moral legitimacy attribution. This is apparent in the framing of these tweets, describing the program as "menace" and a "disaster for the poor."

UIDAI fared slightly better. On one hand, many factors prevented tweeters from attributing moral legitimacy to UIDAI. Dissatisfaction with UIDAI's concern for privacy was accompanied by a lack of faith among tweeters that UIDAI could—or would—address this issue (e.g., "Govt is forcing to link #Aadhaar everywhere but they are unable to stop data breach, compromising our data. #cybersecurity"). Tweeters compared UIDAI to a "police state," referenced George Orwell's 1984 (Orwell, 1949) and underlined UIDAI's illegality and unconstitutionality. Allocating responsibility for privacy protection to residents also breached the expectation, encouraged by GOI (Banerjee, 2015), that UIDAI procedures would protect user data. On the other hand, tweeters appreciated UIDAI's attention to attention to detail and customer service. This is evident in numerous direct requests for information (e.g., "@UIDAI is it possible to change name in aadhar card by gazette notification??") and in tweeters' thanks for UIDAI assistance (e.g., "@UIDAI Deleted it. Thank you"). Both imply tweeters attributed some moral legitimacy to UIDAI.

Finally, regarding *cognitive legitimacy*, our analysis indicates that Aadhaar has become an accepted feature of Indian life. Not only have people enrolled despite being demonstrably insecure about the program's benefits, they continue to enroll notwithstanding widespread awareness of Aadhaar's operational challenges. This suggests UIDAI's narrative that Aadhaar "connects every individual to the state" (Polgreen, 2011) has been effective in crossing the first hurdle for attributing cognitive legitimacy.

Beyond this, however, there is little evidence that tweeters attribute cognitive legitimacy to Aadhaar or UIDAI. Operational issues undercut GOI narratives about Aadhaar's usefulness. Tweeters described "pension related crises" arising from difficulties linking their Aadhaar number to government programs and expressed frustration about unsuccessfully "trying this link for many days." Such tweets evaluate Aadhaar as complicating tweeters' daily lives, not making it more "predictable, meaningful and inviting"

Table 3. Legitimacy attribution.

Legitimacy object	Legitimacy type		
	Pragmatic legitimacy	Moral legitimacy	Cognitive legitimacy
Aadhaar	−	−	−
UIDAI	+	o	−

UIDAI: Unique Identification Authority of India; +: indicates a clear legitimacy gain; o: balanced gains and losses; −: a clear legitimacy loss.

(Suchman, 1995: 582), thereby indicating a lack of cognitive legitimacy attribution. In addition, we saw a loss of cognitive legitimacy for Aadhaar regarding big data policy visions. Positive empirical examples given in Period 1 regarding Aadhaar's potential to facilitate economic development become negative by Period 2. Regarding government services, one tweeter summarizes tweeters' sentiment, writing Aadhaar ". . .was supposed to bolster welfare and check corruption. It has only created new problems." In other words, tweeters saw a connection between Aadhaar and their daily lives, but their experience(s) with Aadhaar failed to meet expectations. Language linking

UIDAI to tweeters' daily lives ("nightmare") suggests similar divergence between UIDAI's behavior, tweeters' belief system, and the experienced reality of Aadhaar, again impairing attribution of cognitive legitimacy. Table 3 summarizes the results.

Conclusion

This article has examined India's Aadhaar program in an effort to gain a more detailed understanding of how actors are cultivating and attributing legitimacy to big data projects in the Global South. In so doing, we have taken an innovative approach to Indian politics, to social media data, and to measuring legitimacy by performing a mixed-methods analysis of roughly 250,000 tweets collected during Aadhaar's first 10 years.

Aadhaar and UIDAI both appear to be cases of legitimacy lost. Tweeters' reluctance to attribute legitimacy stemmed both from normative failures (e.g., inadequate privacy protection) and material ones (e.g., technological issues). Importantly, sometimes it also stemmed from the interaction of material and normative factors, as in interpretations of daily hassles related to Aadhaar (material) as "nightmares" (normative). This highlights the importance of considering legitimacy attribution within the context of actors' cost-benefit calculations. It also suggests that, to gain an accurate understanding of legitimacy attribution, one must complement analyses delinking cost–benefit analysis and legitimacy analysis (e.g., to determine the causal mechanisms of legitimacy attribution) with studies explicitly linking these two factors (e.g., to determine how they interact during legitimation).

Our study additionally presents a mixed-methods approach to Twitter analysis to overcome one problem faced by scholars using social media data to examine political phenomena, namely, an inability to map social media data onto broad conceptual categories. By combining quantitative sentiment analysis and qualitative content analysis, we identified changes in public discussion of Aadhaar and UIDAI, contextualized these

findings within the Indian political context, and used the findings to evaluate changes in legitimacy attribution by tweeters over a 10-year period. Moreover, as the Supplemental online appendix illustrates, plausibility tests using survey and interview data confirmed our findings. This speaks not only of the usefulness of examining digital communities when evaluating legitimacy attribution vis-à-vis big data programs and administrators. In addition, it highlights the value of our innovative approach to social media data, namely, applying small *and* big data techniques when analyzing social media data.

Acknowledgements

Funding

ORCID iDs

Laura C Mahrenbach [iD] https://orcid.org/0000-0002-8622-3242
Jürgen Pfeffer [iD] https://orcid.org/0000-0002-1677-150X

Supplemental material

Supplemental material for this article is available online.

Notes

1. Government of India (GOI) acceptance of Aadhaar's legitimacy was also in question. Sathe (2014) writes that government actors constituted a "wall of opposition, cutting across political and ideological divisions," worrying the program would usurp bureaucratic territory, was technologically infeasible, and would bring limited, if any, electoral returns.
2. Discrimination toward Indian women using Twitter for political purposes reflects societal attitudes toward women and the male-dominated culture of Indian politics (Anderson, 2015). Similarly, the most powerful regional parties in India have been located along the northern border and in coastal regions since the 1990s (Bhatia, 2019). Consequently, demographic imbalances in tweeter distribution should not substantially alter our findings as they reflect norms in Indian politics.
3. Irrelevant tweets contained one of the search terms—most frequently "UIDAI"—but with a unrelated meaning (e.g., "colheita feliz ta lerda ASH**UIDAI**UDAIUDHIUAHDIUASHDIUH").
4. Details of the datasets, coding procedures and rules, and so on appear in the supplemental online appendix, included as supplemental material on Sage's website.

References

Abraham R, Bennett ES, Bhusal R, et al. (2018) *State of Aadhaar Report 2017-2018*. New Delhi, India: IDinsight.

Abraham R, Bennett ES, Sen N, et al. (2017) *State of Aadhaar Report 2016-17*. New Delhi, India: IDinsight.

Anderson S (2015) *India's Gender Digital Divide: Women and Politics on Twitter*. Available at: https://www.orfonline.org/wp-content/uploads/2015/12/ORFIssueBrief_108.pdf

Arora P (2016) The bottom of the data pyramid: big data and the Global South. *International Journal of Communication* 10: 1681–1699.

Arsène S (2012) The impact of China on global internet governance in an era of privatized control. In: *Chinese internet research conference*, Los Angeles, CA, 19–21 June.

Arya D (2013) Why are Indian women being attacked on social media? *BBC News*, 8 May. Available at: https://www.bbc.com/news/world-asia-india-22378366

Bader J (2019) To sign or not to sign: hegemony, global internet governance, and the international telecommunication regulations. *Foreign Policy Analysis* 15(2): 244–262.

Banerjee S (2015) Aadhaar: digital inclusion and public services in India. Available at: https://thedocs.worldbank.org/en/doc/655801461250682317-0050022016/original/WDR16BPAadhaarPaperBanerjee.pdf

Barbera P and Rivero G (2015) Understanding the political representativeness of Twitter users. *Social Science Computer Review* 33(6): 712–729.

Barker-Plummer B and Barker-Plummer D (2017) Twitter as a feminist resource: #YesAllWomen, digital platforms, and discursive social change. In: Earl J and Rohlinger DA (eds) *Social Movements and Media*. Bingley: Emerald Publishing Limited, pp. 91–118.

Bauman Z, Bigo D and Esteves P (2014) After Snowden: rethinking the impact of surveillance. *International Political Sociology* 8(2): 121–144.

Baxi P (2019) Technologies of disintermediation in a mediated state: civil society organisations and India's Aadhaar project. *South Asia: Journal of South Asian Studies* 42(3): 554–571.

Belair-Gagnon V, Mishra S and Agur C (2013) Reconstructing the Indian public sphere: newswork and social media in the Delhi gang rape case. *Journalism* 15(8): 1059–1075.

Bernstein S (2011) Legitimacy in intergovernmental and nonstate global governance. *Review of International Political Economy* 18(1): 17–51.

Bhat JM (2013) Legitimation of e-government initiatives: a study of India's identity project. In: *International working conference on transfer and diffusion of IT*, Bangalore, India, 27–29 June, pp. 343–358. Berlin: Springer.

Bhatia G (2019) *The Changing Face of Indian State Assemblies*. London: Reuters.

Brooker P, Barnett J, Vines J, et al. (2018) Researching with Twitter timeline data: A demonstration via "everyday" sociopolitical talk around welfare provision. *Big Data & Society* 5: 1–13.

Calderon NA, Fisher B, Hemsley J, et al. (2015) Mixed-initiative social media analytics at the World Bank: observations of citizen sentiment in Twitter data to explore "trust" of political actors and state institutions and its relationship to social protest. In: *2015 IEEE international conference on big data*, Santa Clara, CA, 29 October–1 November, pp. 1678–1687. New York: IEEE.

Chatfield AT, Reddick CG and Brajawidagda U (2015) Government surveillance disclosures, bilateral trust and Indonesia —Australia cross-border security cooperation: social network analysis of Twitter data. *Government Information Quarterly* 32(2): 118–128.

Constine J (2015) Indian prime minister tells Zuckerberg social media creates a new form of diplomacy. *TechCrunch*, 27 September. Available at: https://techcrunch.com/2015/09/27/modiberg/

Daugman J (2014) Large data analysis: automatic visual personal identification in a demography of 1.2 billion persons. In: *Proceedings of SPIE* (eds HH Szu and L Dai), Baltimore, MD, 22 May.

Deephouse DL and Suchman M (2008) Legitimacy in organizational institutionalism. In: Greenwood R, Oliver C, Suddaby R, et al. (eds) *The SAGE Handbook of Organizational Institutionalism*. Thousand Oaks, CA: SAGE, pp. 49–77.

Easton D (1975) A re-assessment of the concept of political support. *British Journal of Political Science* 5: 435–457.

Ebert H and Maurer T (2013) Contested cyberspace and rising powers. *Third World Quarterly* 34(6): 1054–1074.

Etter M, Colleoni E, Illia L, et al. (2018) Measuring organizational legitimacy in social media: assessing citizens' judgments with sentiment analysis. *Business and Society* 57(1): 60–97.

Ganjoo S (2018) Foreign ministry responds to reports of social media monitoring, says it does not monitor specific individuals. *India Today*, 15 August. Available at: https://www.indiatoday.in/technology/news/story/foreign-ministry-responds-to-reports-of-social-media-monitoring-1315020-2018-08-15

Gerdeman D (2012) India's ambitious national identification program. *Harvard Business School Working Knowledge*, 30 April. Available at: https://hbswk.hbs.edu/item/indias-ambitious-national-identification-program

Ghose I (2015) Indian government initiative for big data. In: Global Statement. Available at: https://globalstatement2015.wordpress.com/2015/08/12/big-data-in-india-the-governmental-intervention/ (accessed 2017).

Gonçalves P, Araújo M, Benevenuto F, et al. (2013) Comparing and combining sentiment analysis methods. In: *COSN'13: proceedings of the first ACM conference on online social networks*, Boston, MA, 7–8 October, pp. 27–38. New York: Association for Computing Machinery.

Government of India (2016) *Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Act* (ed Ministry of Law and Justice). New Delhi, India: Government of India Press.

Government of India (2018) *Request for Proposals invited for Selection of Agency for SITC of Software and Service and Support for Function, Operation and Maintenance of Social Media Communication Hub* (ed Ministry of Law and Broadcasting). New Delhi, India: Government of India.

Grimmer J and Stewart BM (2013) Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21(3): 267–297.

Gronau J and Schmidtke H (2016) The quest for legitimacy in world politics—international institutions' legitimation strategies. *Review of International Studies* 42(3): 535–557.

Hsieh H-F and Shannon SE (2005) Three approaches to qualitative content analysis. *Qualitative Health Research* 15(9): 1277–1288.

Hurd I (2019) Legitimacy and contestation in global governance: revisiting the folk theory of international institutions. *Review of International Organizations* 14: 717–729.

Jiang J, Meng T and Zhang Q (2019) From internet to social safety net: the policy consequences of online participation in China. *Governance Early View* 32: 531–546.

Jiang M and Okamoto K (2014) National identity, ideological apparatus, or panopticon? A case study of the Chinese national search engine Jike. *Policy and Internet* 6(1): 89–107.

Keohane RO (2011) Global governance and legitimacy. *Review of International Political Economy* 18(1): 99–109.

Khursheed R (2014) Governance, Twitter style. Available at: https://blog.twitter.com/official/en_in/a/2014/governance-twitter-style.html (accessed 2018).

Kidd D (2019) Extra-activism: counter-mapping and data justice. *Information, Communication & Society* 22(7): 954–970.

Klitgaard R (2011) Designing and implementing a technology-driven public-private partnership. *Innovations* 6(2): 67–72.

Leetaru KH, Wang S, Cao G, et al. (2013) Mapping the global Twitter heartbeat: the geography of Twitter. *First Monday* 18(5). Available at: https://firstmonday.org/article/view/4366/3654

Lipset SM (1959) Some social requisites of democracy: economic development and political legitimacy. *American Political Science Review* 53: 69–105.

Liu H and Lee D (2014) Quantifying political legitimacy from Twitter. In: Kennedy WG, Agarwal N and Yang SJ (eds) *Social Computing, Behavioral-Cultural Modeling, and Prediction*. Washington, DC: Springer, pp. 111–118.

Mahrenbach LC and Mayer K (2020) Framing policy visions of big data in emerging states. *Canadian Journal of Communication* 45(1): 129–141.

Mahrenbach LC, Mayer K and Pfeffer J (2018) Policy visions of big data: Views from the Global South. *Third World Quarterly* 39(10): 1861–1882.

McCarthy MT (2016) The big data divide and its consequences. *Sociology Compass* 10(12): 1131–1140.

Maireder A and Ausserhofer J (2014) Political discourses on Twitter: networking topics, objects, and people. In: Weller K, Bruns A, Burgess J, et al. (eds) *Twitter and Society*. New York: Peter Lang, pp. 305–318.

Mohan S (2015) Locating the "Internet Hindu": political speech and performance in Indian cyberspace. *Television & New Media* 16(4): 339–345.

News India Times (2017) Twitter sees 1.9 mn mentions of Gujarat Polls, Modi top personality. *News India Times*, 19 December. Available at: https://www.newsindiatimes.com/twitter-sees-1-9-mn-mentions-of-gujarat-polls-modi-top-personality/

Nilekani N and Shah V (2016) *Rebooting India: Realizing a Billion Aspirations*. London: Penguin Books.

O'Connor B, Balasubramanyan R, Routledge BR, et al. (2010) From tweets to polls: linking text sentiment to public opinion time series. *ICWSM: proceedings of the fourth international conference on weblogs and social media*, Washington, DC, 23–26 May, pp. 122–129. Menlo Park, CA: Association for the Advancement of Artificial Intelligence.

Orwell G (1949) 1984. London: Secker and Warburg.

Panda A, Gonawela A, Acharyya S, et al. (2020) NivaDuck—a scalable pipeline to build a database of political Twitter handles for India and the United States. In: *SMSociety'20: international conference on social media and society* (ed G Anatoliy, P Mai, R Recuero, et al.), Toronto, ON, Canada, 22–24 July, pp. 200–209, New York: Association for Computing Machinery.

Pennebaker JW, Booth RJ, Boyd RL, et al. (2015) *Linguistic Inquiry and Word Count*. Austin, TX: Pennebaker Conglomerates.

Pew Research Center (2012) 2012 Spring survey topline results. Available at: https://www.pewresearch.org/wp-content/uploads/sites/2/2012/10/Pew-Global-Attitudes-China-Report-FINAL-TOPLINE-October-10-2012.pdf

Poell T (2020) Social media, temporality, and the legitimacy of protest. *Social Movement Studies* 19(5): 609–624.

Poell T and Rajagopalan S (2015) Connecting activists and journalists: Twitter communication in the aftermath of the 2012 Delhi rape. *Journalism Studies* 16(5): 719–733.

Polgreen L (2011) Scanning 2.4 billion eyes, India tries to connect poor to growth. *New York Times*, 1 September. Available at: https://www.nytimes.com/2011/09/02/world/asia/02india.html

Rajput H (2014) Social media and politics in India: a study on Twitter usage among Indian political leaders. *Asian Journal of Multidisciplinary Studies* 2(1): 63–69.

Rao H (1994) The social construction of reputation: certification contests, legitimation, and the survival of organizations in the American automobile industry: 1895-1912. *Strategic Management Journal* 15: 29–44.

Reus-Smit C (2007) International crises of legitimacy. *International Politics* 44(2–3): 157–174.

Rodrigues U (2020) Political communication on social media platforms. In: Athique A and Parthasarathi V (eds) *Platform Capitalism in India*. Cham: Palgrave Macmillan, pp. 221–238.

Rodrigues UM and Niemann M (2017) Social media as a platform for incessant political communication: a case study of Modi's "Clean India" campaign. *International Journal of Communication* 11: 3431–3453.

Ruths D and Pfeffer J (2014) Social media for large studies of behavior. *Science* 346(6213): 1063–1064.

Sathe V (2011) The world's most ambitious ID project. *Innovations* 6(2): 39–65.

Sathe V (2014) Managing massive change: India's Aadhaar, the world's most ambitious ID project. *Innovations* 9(1–2): 85–111.

Schlipphak B (2015) Measuring attitudes toward regional organizations outside Europe. *Review of International Organizations* 10(3): 351–375.

Schmidt VA (2013) Democracy and legitimacy in the European Union revisited: input, output and "throughput." *Political Studies* 61(1): 2–22.

Schmidtke H (2019) Elite legitimation and delegitimation of international organizations in the media: patterns and explanations. *Review of International Organizations* 14: 633–659.

Schneider S, Hurrelmann A, Krell-Laluhová Z, et al. (2010) *Democracy's Deep Roots: Why the Nation State Remains Legitimate*. London: Palgrave Macmillan.

Schober MF, Pasek J, Guggenheim L, et al. (2016) Social media analysis for social measurement. *Public Opinion Quarterly* 80(1): 180–211.

Srikrishna BN (2017) *White Paper of the Committee of Experts on a Data Protection Framework for India* (ed Committee of Experts). New Delhi, India: Government of India.

Srinivasan J and Johri A (2013) Creating machine readable men: legitimizing the "Aadhaar" mega e-infrastructure project in India. In: *Sixth international conference on information and communication technologies and development*, Cape Town, South Africa, 7–10 December, pp. 101–112. New York: Association for Computing Machinery.

Suchman MC (1995) Managing legitimacy: strategic and institutional approaches. *Academy of Management Review* 20(3): 571–610.

Taylor L and Schroeder R (2015) Is bigger better? The emergence of big data as a tool for international development policy. *GeoJournal* 80(4): 503–518.

Unique Identification Authority of India (UIDAI) (2019) Aadhaar dashboard. Available at: https://uidai.gov.in/aadhaar_dashboard/ (accessed 13 February).

Zheng L (2013) Social media in Chinese government: drivers, challenges and capabilities. *Government Information Quarterly* 30(4): 369–376.

Zott C and Huy QN (2007) How entrepreneurs use symbolic management to acquire resources. *Administrative Science Quarterly* 52(1): 70–105.

Zürn M and Stephen M (2010) The view of old and new powers on the legitimacy of international institutions. *Politics* 30(S1): 91–101.

Author biographies

Laura C Mahrenbach is a research fellow and lecturer at the Technical University of Munich's School of Governance.

Jürgen Pfeffer is a professor of computational social science and big data at the Bavarian School of Public Policy of the Technical University of Munich.