© *Academy of Management Discoveries*
2018, Vol. 4, No. 1, 32–49.
Online only
https://doi.org/10.5465/amd.2015.0121

PROTESTING CORRUPTION ON TWITTER: IS IT A BOT OR IS IT A PERSON?

CAROLINA ALVES DE LIMA SALGE[1]
ELENA KARAHANNA
University of Georgia

In studying how activists use technology to express public dissatisfaction online, we discover that what we assumed to be human protestors were in some cases bots—automated accounts in online social networks. To explicate the discovery of bots, we problematize an implicit assumption of online social network research within Management and Information Systems as it pertains to the concept of actors. Our discovery takes place in the context of a 6-day inductive case study of a protest against government corruption in Brazil—the Mensalão scandal. We elaborate on how bots were detected and discuss how they are coded to amplify the magnitude of the protest on Twitter. Furthermore, we explore the application of bots beyond the context of our study by illustrating how they were used to increase revenue in the business of online dating and to manipulate public opinion during an election campaign. We also discuss how neglecting bots can threaten research validity and, as a result, we provide scholars investigating social phenomena online with a multi-method approach for detecting bots. Finally, we position bot as a crucial actor with implications for organizational theory and practice.

The *Mensalão* was a vote-buying case of corruption that almost collapsed the Brazilian government of Luiz Inácio Lula da Silva in 2005 (BBC, 2013). The scandal broke when Roberto Jefferson, the president of an allied party, announced in a newspaper interview that the Worker's Party (PT) was using public funds to buy political support for the then-Lula government and to pay off debts from election campaigns. Each congressman was receiving about R$30,000 a month (around $12,000 at the time) (The

Author's voice:
What motivated you personally to undertake this research? Why is it important to you?

[1] Corresponding author.

Economist, 2013). The allegation led to the downfall of several congressmen and senior members of the government, including José Dirceu, Lula's chief of staff and the alleged mastermind behind the case, Delúbio Soares, PT's treasurer, and José Genoino, PT's former president. In August 2007, the Supreme Federal Court, responsible for investigating cases against parliamentarians, accepted the indictments of 40 deputies involved in the Mensalão scandal. The trial began in August 2012, and roughly 2 months later, 25 of the 40 defendants were charged with several crimes ranging from embezzlement and corruption to conspiracy and misuse of public funds. Mr. Dirceu was among the 25 prosecuted deputies. He was sentenced to spend 10 years and 10 months in jail. The court's decision was celebrated by many and marked the beginning of a new era, where those involved in government corruption would be held accountable for their transgressions (Leahy, 2012). But the celebration was brief. Brazil's legal system is "a loophole-ridden oddity, allowing appeals even against supreme-court rulings" (The Economist, 2013). On September 18, 2013 (about a year after the sentence), the Supreme Federal Court accepted—in a 6-5 vote—a motion to hear a new round of appeals from 12 deputies charged in the Mensalão case. This result frustrated and angered many Brazilians (Lyons & Cowley, 2013; Singer, 2013) who were indignant with the justices for giving corrupted politicians a second chance. Motivated by rage, thousands started a corruption protest on Twitter.

Protests are "organized, collective, and public expressions of discontent" (King & Soule, 2007: 415). They reflect public action. Instead of reaching out to higher authorities with expressions of grievance and desires of maintaining conversations private, activists vent their dissatisfaction openly to a broader audience. Protests initiated on Twitter are significant because they can influence public discourse and, as a result, shape both civic and political engagement (Schumann, 2014). As Tufekci (2014) suggests, referring to the Ferguson protests, "what happens to #Ferguson [on Twitter] affects what happens to Ferguson" (Tufekci, 2014: para. 36). Yet, we know little about the central actors on Twitter protests. The purpose of this article, therefore, is to discover who these actors are and what they have in common. Central actors are the "most important" nodes of a network because they are "located in strategic locations within the network" (Wasserman & Faust, 1994: 169). We focus on the corruption protest that emerged after the Supreme Federal Court accepted to hear a new round of appeals for the Mensalão case.

In investigating this phenomenon, we discover that actors occupying central positions in online social networks are not always people but instead, they can also be *bots* (short for *social robots*)—automated accounts in online social networks (Morstatter, Wu, Nazer, Carley, & Liu, 2016). We reveal how these bots were detected and show how they were designed to amplify information on Twitter. Our discovery suggests that both humans and bots can be *central* when engaging in online activism. Existing Management and Information Systems literature does not account for this bot phenomenon. In light of this, we take a problematization approach (Alvesson & Sandberg, 2011) by questioning the implicit assumption that *actors* of online social networks are *people*. Our study contributes to the emerging literature in online social networks as we argue that the conceptualization of *actors* should not be constrained to people or organizations, but rather, it needs to be expanded to also include bots. We discuss how neglecting bots imposes threats to research validity and, as a result, we urge scholars investigating social phenomena online to carefully consider bot implications when designing their studies.[2] In addition, our problematization of a field assumption (Alvesson & Sandberg, 2011) opens to scrutiny the nature and relevance of bots for organizational research.

To guide the readers to our discovery, we start by identifying the assumption we problematize. We then explain our methodology and present our findings. Next, we introduce two mini-cases to show how our discovery extends beyond the context of this study to other organizational and social settings. We subsequently discuss how neglecting bots can impose threats to research validity and we develop a method that scholars can use to best detect them. Finally, we consider the implications of bots for theory and practice.

ONLINE SOCIAL NETWORKS

An online social network "consists of a set of actors or nodes along with a set of ties of a specified type (such as friendship) that link them" (Borgatti & Halgin, 2011: 1169) on a digital platform. Ties are connected via shared end points to form paths indirectly relating actors that are not directly tied with one another. The pattern of ties in online social networks generates a particular structure, and actors occupy positions within this structure. It is important to recognize that the term *actor* does not necessarily imply that these social entities have "the volition or ability to 'act'" (Wasserman & Faust, 1994: 17). In other words, an actor is not necessarily a person but instead

[2] In our article, we focus exclusively on social bots and use the term "bots" to refer solely to these. However, there are other types of bots on the Internet such as those used to scrape data from numerous websites (e.g., diffbot).

an entity. Also, it is the researcher who defines an online social network by choosing which type of ties and which set of actors to study (Borgatti & Halgin, 2011). Much of the online social network literature in Management and Information Systems defines one-mode networks illustrating structure (e.g., tie strength) and actor position (e.g., centrality) and relating these to either group-level (e.g., community growth) or actor-level (e.g., leadership) outcomes. In these studies, scholars implicitly assume that *actors* are *people*.[3] For instance, empirical work analyzing the impact of social influence on product adoption (Aral & Walker, 2014) and content generation (Zeng & Wei, 2013) on digital platforms such as Facebook and Flickr emphasize the human aspect of nodes by referring to them as individuals/people when presenting findings (*italics* emphasis added):

> "*Individuals* [on Facebook] exert 125% more influence on friends for each institutional affiliation they share in common ($p < 0.05$)" (Aral & Walker, 2014: 1362).

> "We found that *people* tend to upload more similar photos [on Flickr] around the time of the formation of a social tie" (Zeng & Wei, 2013: 72).

We observe this same type of assumption in research on *leadership*, where leaders of online communities are described as "*people* leading members of [a] newsgroup" (Faraj, Kudaravalli, & Waso, 2015: 400); *e-commerce*, where nodes purchasing or reviewing products on Amazon are defined as "*the individuals*" (Dhar, Geva, Oestreicher-Singer, & Sundararajan, 2014: 264) or seen as "*people*" (Kumar & Benbasat, 2006: 428); *social movements*, where Twitter users protesting authoritarian regimes are described as "*people*" (Oh, Eom, & Rao, 2015: 213); and *information diffusion*, where users receiving direct messages from other users aspiring to spread rumors on Twitter are seen as "specific *individuals* in the Tweeter's social network" (Oh, Agrawal, & Rao, 2013: 412). We even see this assumption in a conceptual article about knowledge collaboration:

> "Knowledge collaboration requires that *individuals* spend time contributing to the OC's [online community's] virtual workspace" (Faraj, Jarvenpaa, & Majchrzak, 2011: 1227).

[3] We identified 80 online social network studies published in Management and Information Systems journals on the Financial Times Top 45 list. Only one study accounted for the existence of bots. They did so through the CAPTCHA method to reduce "message volume and uninteresting messages generated by *spam programs*" (Butler, Bateman, Gray, & Diamant, 2014: 717). The examples we provide come from the remaining 79 studies.

In this case, the belief is that knowledge collaboration—"the sharing, transfer, accumulation, transformation, and cocreation of knowledge" (Faraj et al., 2011: 1224)—only occurs when *people* generate content to the online community's workspace. But, as we detail below in our discovery, this is not always the case. Actors collaborating knowledge, exerting or receiving influence, purchasing or reviewing products, protesting authoritarian governments, or diffusing information in online social networks do not need to be *people;* they can also be *bots*.

Post-Bot–Discovery Exploration

We did not begin this project expecting to find bots to be central actors protesting government corruption on Twitter. This was a discovery that emerged as part of our inductive approach. Rather, we began with two research questions: (1) *Who are the central actors in Brazil's anti-corruption protest on Twitter?* (2) *What do they have in common?* As is often the case with inductive research, we engaged in post-discovery exploration (Charmaz, 2006); as we dug deeper in the case—by iteratively comparing our existing data to emerging data—the prevalence of bots as central actors became apparent. As bots replicated specific messages on Twitter, we discovered that they were central partly because they amplified the magnitude of content embedded in those duplicated posts. Therefore, our study was refined to not only provide answers to our initial research questions but to also shed light on the implications of our discovery through a post-bot–discovery exploration phase.

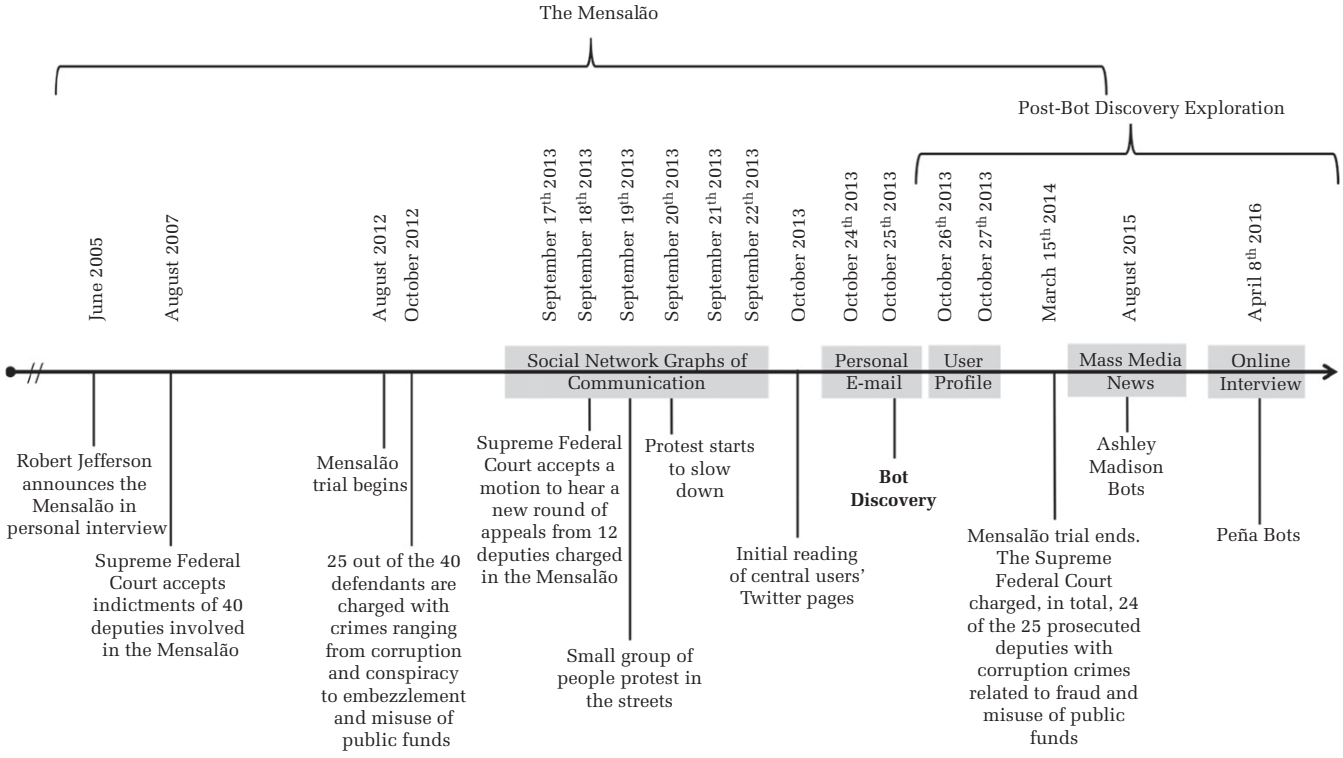
METHODS

We drew upon multiple data sources, including social network graphs of communications, personal e-mails, user profile reports, a four-part media article series, and an in-depth semi-structured interview. To preview our findings, and to serve as a blueprint for the discovery process described here, we present a timeline of our study in Figure 1. Details of each data source are explained next; an illustration of the data we collected is presented in Table 1.

We used NodeXL (Hansen, Shneiderman, & Smith, 2011) to collect social network data on Twitter. Our sample includes #ChangeBrazil messages because this was the most frequently used hashtag to protest government corruption in Brazil (Monroy-Hernández & Spiro, 2013). We gathered data from September 17,

Author's voice:
How did the paper evolve and change as you worked on it?

FIGURE 1
Study Timeline



2013, 1 day before the protest emerged, to September 22, 2013, 5 days after its start. This time period was selected because Twitter networks tend to be busiest during the first 5 days of a protest (see interaction networks in Monroy-Hernández & Spiro, 2013). Our dataset includes 259 unique Twitter users with 4,513 messages. For every message posted (tweet, retweet, reply, or mention), we have the username of the initiator, the URL of the message, and the time stamp.

We also engaged in two e-mail exchanges with a central actor in our sample. The e-mails confirmed the existence of bots and also included detailed information about their incentives, design, and usage. We began the post-bot–discovery data collection portion of the project by reading and analyzing user profile data related to the central actors (bots and humans). These analyses were specific and undertaken to further sensitize us as to how bots were different, yet similar, to people. To explore bot practices beyond the context of our study, we read and analyzed a series of articles delineating their use in a business organization. We also interviewed an activist studying bots in an election campaign in Mexico. We asked him questions dealing with the objective, motivation, and utilization of bots. The structured component of the interview allowed us to understand more details about the bots themselves, whereas the unstructured component allowed for contextual details to emerge. The interview, conducted in Spanish, was audio recorded, professionally transcribed and translated verbatim.

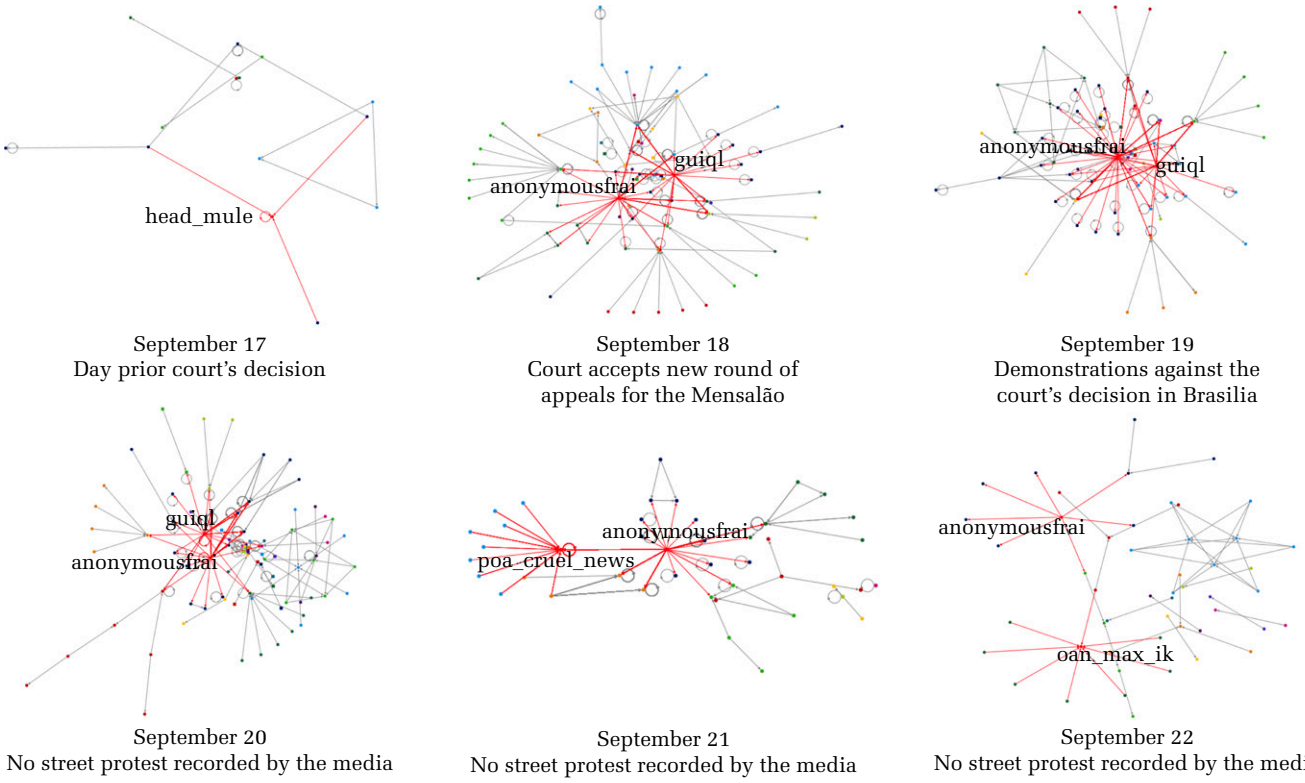
Pre-Bot–Discovery Data Analysis

To identify the central actors of the protest, we analyzed the social network data in two ways. First, because actor centrality may exhibit temporal patterns, we plotted six graphs using the Harel-Koren Fast Multiscale algorithm to visualize central actors

TABLE 1
Data Collection

Data Type	Study Phase	Quantity	Source
Social network analysis	Pre-bot discovery	4,513 messages from 259 unique Twitter users	NodeXL (Twitter)
Personal communication	Bot discovery	2 exchanges	E-mail
User profile	Post-bot discovery	4 reports	Simply measured
Mainstream mass media news reporting	Post-bot discovery	4 files	Gizmodo
In-depth semi-structured interview	Post-bot discovery	1 conversation (65 minutes in length)	Interview (Skype)

FIGURE 2
Central Actors in Brazil's Anti-Corruption Twitter Protest—#ChangeBrazil



September 17
Day prior court's decision

September 18
Court accepts new round of appeals for the Mensalão

September 19
Demonstrations against the court's decision in Brasilia

September 20
No street protest recorded by the media

September 21
No street protest recorded by the media

September 22
No street protest recorded by the media

for *each protest day* (see red ties in Figure 2). Second, for all users in our sample, we computed five centrality measures (defined in Table 2): in-degree, out-degree, betweenness, closeness, and eigenvector, which together "*cover the intuitive range of the concept of centrality*" (Freeman, 1979: 237). These two analyses enabled us to compare multiple types of data to check the robustness of our findings.

FINDINGS

Figure 2 shows the Harel-Koren Fast Multiscale graphs (an algorithmic method for drawing large weighted social network diagrams rapidly). *Head mule* was the most central actor on September 17, but the user becomes noncentral once the protest begins (September 18). Instead, two other actors (*anonymousfrai* and *guiql*) become and remain central through September 20 which is the day the protest starts to fade in the traditional media. The graphs also show that *anonymousfrai* and *guiql* are central because they bridge information across the network by connecting otherwise unconnected groups of users via tweets, retweets, mentions, or replies. Centrality computations corroborate this; *anonymousfrai* and *guiql* have high betweenness centrality scores between September 18 and 20. Although it may not be noticeable in Figure 2, we find that these actors are also connected to the same users—yet, they are not directly tied with each other. We observe a significant change on September 21. The network is now divided into two groups and linked through *poa_cruel_news* (instead of *guiql*) and *anonymousfrai*. We note that, on the fifth day of the protest (September 22), *anonymousfrai* remains central, whereas *poa_cruel_news*' bridging capability is replaced by that of *oan_max_ik*. In short, graph results indicate that the Twitter protest had four central actors during the 6-day period. These users are *anonymousfrai*, *guiql*, *poa_cruel_news*, and *oan_max_ik*.

TABLE 2
Actor Network Centrality

Dimension	Definition
In-degree	Number of incoming ties (tweets, retweets, replies, and mentions) of an actor
Out-degree	Number of outgoing ties of an actor
Betweenness	Number of times an actor bridges the shortest path between two other actors
Closeness	Average of the shortest path lengths from a certain actor to all other actors
Eigenvector	The degrees (in and out) of the actors that a certain actor is connected to

TABLE 3
Centrality Scores for Networks' Most Important Actors

	Betweenness	Eigenvector	Closeness	In-Degree	Out-Degree
Anonymousfrai	17,265.79	0.07	0.00	1.00	71.00
Guiql	6,910.79	0.05	0.00	1.00	41.00
Oan_max_ik	3,633.12	0.02	0.00	21.00	1.00
Poa_cruel_news	3,461.35	0.02	0.00	20.00	1.00
E_ditora	2,802.85	0.01	0.00	10.00	4.00

Table 3 shows centrality scores for the five most important actors in our sample. Betweenness, eigenvector, and degree centralities were used as measures of importance. We did not include closeness centrality given that 70.6 percent of actors have a score of nearly zero, making this not a useful metric for differentiating prominence across sampled users. Numeric findings are consistent with those observed in Figure 2. The most central actors in Table 3 are also *anonymousfrai*, *guiql*, *oan_max_ik*, and *poa_cruel_news.* We find that *anonymousfrai* has a significantly higher betweenness centrality score (BC = 17,265.79) than all other central actors. This user is nearly three times more central than *guiql* (BC = 6,910.79) and about five times more central than both *oan_max_ik* (BC = 3,633.12) and *poa_cruel_news* (BC = 3,461.35), implying that *anonymousfrai* is the most important *bridge* in our study. *Anonymousfrai* is also more central than all other actors based on eigenvector centrality (EC = 0.07), whereas *guiql* (EC = 0.05) is more central than both *oan_max_ik* (EC = 0.02) and *poa_cruel_news* (EC = 0.02). Intriguingly, we observe that both *anonymousfrai* and *guiql* score high in out-degree centrality—they mention various users in their tweets, reply to several other users in their tweets, or they retweet other users' tweets frequently—yet, they practically score zero for in-degree centrality; only one user mentions them in their tweets, replies to them in their tweets, or retweets their tweets. In contrast to *anonymousfrai* and *guiql*, both *oan_max_ik* and *poa_cruel_news* have high in-degree centrality and low out-degree centrality.

Our data analyses show robust evidence of four actors being structurally central in Brazil's anticorruption protest on Twitter. Indeed, *anonymousfrai* is the most important actor followed by *guiql*, *oan_max_ik*, and *poa_cruel_news.* However, graph and numeric data are lean and therefore not adequate for discovering who these users are and what underlies commonalities in betweenness and eigenvector centralities. In addition, the thinness of these types of data make them unsuitable for scrutinizing similarities in both in-degree and out-degree centralities for *anonymousfrai* and *guiql* and for *poa_cruel_news* and *oan_max_ik*, respectively. We therefore collected richer data (i.e., text and image) to learn more about each user and to discover what they have in common. The findings of these analyses are presented next.

Scrutinizing Central Actors: The Road to Discovering Bots

We examined each central user's Twitter page and gathered data that signaled information about their persona. We began with *anonymousfrai,* consistently the most important actor in our dataset. Personal textual information about this user was not found, besides what was presumed to be a blog address. However, *anonymousfrai*'s Twitter profile image—a Guy Fawkes mask—revealed that the user was part of *Anonymous*, a loosely associated international network of activists and hacktivists opposing Internet censorship and control. *Anons* (how members of Anonymous refer to themselves) supported the Occupy movement and the Arab Spring (Coleman, 2014) and thus it was not surprising that they were also opposing government corruption in Brazil. In addition, we discovered that *anonymousfrai* never tweeted a single message—instead, the user only retweeted other users' content. Retweeting can be considered a simple action, given that it is the forwarding of a message that has already been created and published by another user. Unmodified retweets—such as those spawned by *anonymousfrai*—are tweet copies with the exception that they include "RT @username" at the beginning of messages. This type of retweet does not require enactors to think or decide the next action. Furthermore, it had now become clear to us that *anonymousfrai's* out-degree centrality scores reflected the user's high retweeting activity rather than large sums of mentions or replies. Because *anonymousfrai* executed a simple action many times in a seemingly standardized manner, we developed a hunch that the user was an *Anon bot* coded to raise awareness for the protest by amplifying the volume of #ChangeBrazil messages via automated retweets.

Guiql was the next central actor that we examined. Unfortunately, only an e-mail address was located as a personal identifier upon initial review of the user's

Twitter page. We made several attempts to contact the account holder through the provided e-mail address, but we never received a response. Considering our hunch related to *anonymousfrai* and the similarity in out-degree centrality between the two, we suspected that *guiql* was also a bot. This speculation motivated us to further review the user's Twitter activity. Similar to *anonymousfrai*, *guiql* retweeted other users' tweets frequently and was almost never mentioned or replied to by others. Unlike *anonymousfrai*, however, *guiql* also tweeted. Yet, the user's messages focused exclusively on media news—not only political, but a variety. Interestingly, *guiql* only tweeted news from a media news portal called Terra.[4] We discovered that this news portal had an official Twitter account for publicizing its own content and were surprised to notice that *guiql* regularly tweeted Terra's articles before Terra did. This finding challenged our bot hunch. Thus, our refined supposition was that *guiql* was a journalist working for Terra who was also an activist supporting the protest via manual retweets.[5]

The third central actor's, *oan_max_ik*, Twitter page provided a valid blog account. On initial review of the blog page, we recognized that all posts revolved around politics. In addition, we discovered—by reading the "*about me*" section of the blog—that many other blog pages were associated with *oan_max_ik*'s account. No other data, however, were available from this user.

The last central actor's—*poa_cruel_news*—Twitter page also provided a valid blog account. Once we read the page, it became evident that *poa_cruel_news's* blog was one of the blogs linked to *oan_max_ik*'s account. Could these actors be the same person? Intrigued, the first author decided to question *poa_cruel_news* about it via e-mail:

> "Hi Eros Thanatos[6], how are you? I have been reading your blogs and I was wondering if you're the only person writing on them. The content is all great but very diverse. I particularly enjoy the 'metafísica do vento [the blog page related to *oan_max_ik's* Twitter account]', 'poema para a porto alegre chauvinista-elitista-machista [the blog page related to *poa_cruel_news's* Twitter account]', and 'opener media'."

Approximately 2 days later, Mr. Thanatos responded:

> "Yes. I am one person... but with diverse personalities, or better, I am a tribe of personalities... Very cool that you gave me feedback and for reading what I write. Thank you for reading and I am available for any clarification... Have a great week..."

Mr. Thanatos's response confirmed our suspicion that both accounts (*poa_cruel_news* and *oan_max_ik*) were managed by one person. His e-mail, however, did not provide any clues about our bot and journalist hunches associated with *anonymousfrai* and *guiql,* respectively. Because Mr. Thanatos was interacting with both of these users—*anonymousfrai* and *guiql* were retweeting his tweets—we thought that he might know *something* about them. Curious as to whether this was the case, the first author asked Mr. Thanatos, in a follow-up e-mail, to comment:

> "Hi Eros: That's quite helpful. Thank you for your reply. I also tried to contact two other users because I read their work on Twitter but I never heard back. Do you happen to know them? Their Twitter names are *anonymousfrai* and *guiql*."

A few hours later, Mr. Thanatos replied:

> "In reality, I do not know them since they are ***Bot***, short for ***robot***, also known as ***Internet bot*** or ***Web bot***, it is a software application conceived to simulate human actions in a repeated and standardized manner, in the same way a robot would. Both accounts were created to support protests in Brazil such as #VemPraRua and #ChangeBrazil, being that every time a Twitter user tweeted a message with these hashtags the bots would replicate it with a retweet... I hope I was helpful... Regards and Carpe Diem."

As this e-mail reveals, central actors in online social networks are not always *people*, but instead, they can also be *bots*.[7] Based on this discovery, we recommend reframing them as *entities,* as suggested by Wasserman and Faust (1994). Indeed, these entities can then be categorized in diverse ways (as people or bots or organizations).

[4] We obtained this information by selecting "view summary" within *guiql*'s tweets.

[5] We believed these were manual retweets because *guiql* did not retweet with the same high frequency as *anonymousfrai*. This user's retweets corresponded to nearly 58 percent of those published by *anonymousfrai*.

[6] Eros Thanatos was how the user described its persona on the blogs and it was also how he signed all e-mails.

[7] *Anonymousfrai* is a bot. We clicked on the user's blog address (http://anonymousfraiburgo.blogspot.se) which now says that the Twitter account represents a bot retweeting every message containing certain keywords or hashtags.

POST-BOT–DISCOVERY EXPLORATION

The e-mail exchange provided us with evidence that *oan_max_ik's* and *poa_cruel_news's* Twitter accounts were managed by one person (Mr. Thanatos) and that *anonymousfrai* and *guiql* were *bots* coded to support protests against corruption in Brazil. To learn more about the similarities and differences between bots and humans, we collected additional profile data for each central actor in our sample. We downloaded four Twitter reports (one per actor) from Simply Measured[8] that contain number of followers, following, and tweets, along with Klout Score and account creation date. There were no discernable patterns in terms of account history and activity. However, actors' *Klout Scores*—a measurement of overall online influence ranging from 1 to 100 (the higher the score, the more influential the actor)—were similar. The values ranged from 40.3 to 49.3 with a mean of 43.1 and a standard deviation of 4.2.

In summary, we found that although social network analysis suggested there were four distinct central actors protesting corruption on Twitter, the reality is that there were only three. Two of them are bots (*anonymousfrai* and *guiql*), whereas the third one is a blogger (Mr. Thanatos) managing two accounts (*poa_cruel_news* and *oan_max_ik*). Differences between bots and humans exist for *in-degree* and *out-degree* centrality. Bots score high on the latter, whereas humans score high on the former. In addition, we found that whereras both bots were programmed to retweet every message containing the hashtags #ChangeBrazil and #VemPraRua, only one of them was also coded to tweet (*guiql*). Finally, we discovered that central actors (bots and humans) have bridging importance and similar *Klout Scores*. Table 4 integrates our findings.

We now turn our attention to the bots themselves by reviewing the literature to learn how they are used and to discover how prevalent they are in online social networks. Because our discovery is based on a single case study, it is possible that the insights of this article do not apply broadly. Yet, based on our review, we suggest that Brazilian activists are not the only ones using bots, claiming that the generalizability of our discovery is extendable beyond both Brazil and protests. To further support this claim, we present two additional mini-cases detailing the use of bots to increase revenue in the business of online dating (Ashley Madison FemBots) and to manipulate public opinion during Mexico's 2012 presidential election campaign (Peña Bots).

[8] For more information, please see http://simplymeasured.com/about/#sm.00014ps91q1cmf7oxx814ulqzemsx.

Bot Prevalence and Usage

Certain bots are designed to behave benignly, often in ways that benefit society. For example, *SF QuakeBot* disseminates information about earthquakes, as they happen, in the San Francisco Bay area.[9] Other bots, however, are modeled to harm—they are coded to inflate support for a political candidate (Ratkiewicz, Conover, Meiss, Gonçalves, Flammini, & Menczer, 2011); spread false rumors (Ferrara, Varol, Davis, Menczer, & Flammini, 2016); damage an organization's reputation (Messias, Schmidt, Oliveria, & Benevenuto, 2013); and even limit free speech (Gallagher, 2015).

According to Lutz Finger, director of data science and engineering at LinkedIn, bots impose significant threats to organizations and society because they "are actually more common than you might think," and because they "can do things beyond our wildest dreams or nightmares."[10] As of the date of publication, Twitter and Facebook—the two most popular social networking sites[11]—contain as many as 23 million (about 8.5 percent) and 140 million (between 5.5 to 11.2 percent) bots, respectively (Goldman, 2014; Grant, 2014). Nearly 27 millions of Instagram users (close to 8.2 percent) are also bots (O'Reilly, 2015). Although LinkedIn is unaware of its bot statistical pervasiveness (Okalow, 2015), the company filed a complaint in California's federal court noting that an unknown number of bots are being used to "steal data about legitimate users, breaching the user agreement, and violating copyright law" (Lipkin, 2014: para. 1). Finally, Tumblr has also recognized that some of its users are bots (Perez, 2011).

Bots beyond the Mensalão Protest

Mini-case 1—Ashley Madison's female bots. Ashley Madison is a Canadian organization connecting users interested in pursuing extramarital affairs. In July 2015, a group of hackers named *Impact Team* gained unauthorized access to Ashley Madison's website. A few days after the security breach, they released personal data about the organization's users because Ashley Madison refused to terminate its business. When

[9] https://twitter.com/earthquakessf.

[10] http://www.lutzfinger.com/evil-business-social-media-bots/.

[11] http://www.alexa.com/topsites/category/Computers/Internet/On_the_Web/Online_Communities/Social_Networking.

TABLE 4
Major Findings

Research Question	Data Type	Study Phase	Findings
Who are the central actors in Brazil's anti-corruption protest on Twitter?	Social network analysis	Pre-bot discovery	There are ***four*** central actors (*anonymousfrai*, *guiql*, *oan_max_ik*, and *poa_cruel_news*)
What do these central actors have in common?	Personal e-mail	Bot discovery	There are, in fact, ***three*** central actors—***two bots*** (*anonymousfrai* and *guiql*) and ***one person*** (Mr. Thanatos) controlling ***two accounts*** (*oan_max_ik* and *poa_cruel_news*)
	User profile & social network analysis	Post-bot discovery	Bots have high ***out-degree*** centrality. People have high ***in-degree*** centrality. Central actors (*bots and people*) have high ***betweenness*** centrality and an average of 43 ***Klout Scores***

Impact Team began to release Ashley Madison's data, they stated that "the site is a scam with thousands of fake female profiles" and that "90–95 percent of actual users are male" (Reddit, 2015: para. 2), a fact corroborated by Newitz (2015a, 2015b, 2015c, 2015d) who published a detailed Gizmodo series describing her findings. She wrote:

> "What I *have* learned from examining the site's source code is that Ashley Madison's army of fembots appears to have been a sophisticated, deliberate, and lucrative fraud. The code tells the story of a company trying to weave the illusion that women on the site are plentiful and eager. . . . the company was clearly on a desperate quest to design legions of fake women to interact with the men on the site." (Newitz, 2015b: para. 3)

Inherent in this quote is a belief that *female* bots were designed to intentionally interact with *male* users. This was due to a "dramatic gender disparity"—only 5.5 million profiles were described as female in a database of about 37 million users (Newitz, 2015a: para. 5).

The female bots did not appear out of nowhere—"they were probably cobbled together from abandoned and fraudulent profiles in the company's massive member database" (Newitz, 2015c: para. 26). In addition, Newitz noticed patterns in the data revealing that bots often had ashleymadison.com e-mail addresses, although other accounts were also registered with Hotmail. Many of them also had IP addresses that suggested people located at the Ashley Madison headquarters created the accounts. She further discovered—by searching through the source code—a set of comments written by the developers explaining the behavior of the bots. Based on her examination, they were programmed to send simple initial phrases such as "hi there" and "u busy?" (Newitz, 2015b: para. 21). Once male users engaged in a dialog, bots responded with longer messages inducing them to pay for credits to carry on further conversations. The strategy, Newitz claims "worked marvelously—at least in 2012" (Newitz, 2015d: para. 3). She wrote:

> "When the engagers (i.e., female bots) were turned off in early 2011, the company's income took a nosedive. So did their conversion rate. When they were turned on again 14 months later, revenues and conversions skyrocketed. It appears that revenues went from roughly $60,000 per month, to $110, 500." (Newitz, 2015d: para. 3)

Mini-case 2—The Peña bots. On July 1, 2012 Mexico elected its current president Enrique Peña Nieto, "a dashing, disciplined campaigner who promised to bring peace and prosperity back to a country weary of drug violence and slow growth" (Miroff & Booth, 2012: para. 1). Nearly four years after Peña Nieto's triumph, a Colombian hacker named Andrés Sepúlveda announced, in a Bloomberg interview[12], that "he [Andrés] led a team of hackers that stole campaign strategies, manipulated social media to create false waves of enthusiasm and derision, and installed spyware in opposition offices, all to help Peña Nieto, eke out a victory" (Robertson, Riley, & Willis, 2016: para. 8). Specifically, Sepúlveda claims he built "an army of 30,000 Twitter bots" to create trends favoring Peña Nieto as a way to throw the preferences

[12] For more information, please see http://www.bloomberg.com/features/2016-how-to-hack-an-election/.

of voters (Robertson et al., 2016: para. 33)—a fact endorsed by Alberto Escorcia[13], a social activist who analyzed Twitter data during Mexico's 2012 election and whom we interviewed. We asked Escorcia to comment on the Peña bots, for example, how they were coded.[14] He responded:

> "I interviewed one of them, an engineer that [also] did that [what Sepúlveda did], and he told me that they made it [the Peña bots] with Python, and PHP. So what they did was create a system where they massively created many Twitter accounts. They changed the names on them, they changed their photos. And they used photos that they bought from databanks either on Facebook or Hive. . . . these programs used the accounts of real people, but at some point they used a 'bot network' and they put out messages in favor of Peña Nieto and turned [them] into trending topics."

In both of these cases—Sepúlveda's conversation with Bloomberg and Escorcia's interview with us—the use of bots to manipulate public opinion during Mexico's 2012 presidential election was reported. Designing bots to create and popularize messages favoring Peña Nieto on Twitter was one way in which engineers allegedly helped the then candidate become president.

DISCUSSION

We have examined the critical issue of bots on Twitter and how they were designed by activists to protest government corruption in Brazil. Our post-discovery data exploration phase provides deeper insight into bots. Embracing them offers new opportunities for theory development and refinement, which we discuss in the following paragraphs. But before any theoretical progress can be made, we discuss how neglecting bots threatens the research validity of online social network studies. As a result, it is important for us to understand how to detect them.

Neglecting Bots: Threats to Research Validity

Scholars in Management and Information Systems define—through an implicit assumption—actors of online social networks as *humans*. Our case study problematizes this assumption by discovering bots. We, therefore, pose that existing research imprecisely defines the concept of "actors" in online social networks. This lack of clarity can cause problems at the conceptual and operational levels (Podsakoff, Mackenzie, & Podsakoff, 2016).

At the conceptual level, sources of invalidity can originate from construct definition. A well-defined construct specifies what should be included and what should be excluded; if the domain is too broadly defined, extraneous factors other than the target construct may be included (Netemeyer, Bearden, & Sharma, 2003). Therefore, an imprecise *definition* of "actor" that embraces all online actors when it only intends to include humans (i.e., makes an error of inclusion), or excludes bots when it intends to include all actors (i.e., an error of exclusion) threatens construct validity. At the operational level, a lack of clarity threatens construct validity because it increases the likelihood that the operationalization of the concept will be contaminated and/or deficient (Mackenzie, 2003). For example, scholars operationalizing *actors* as people, and collecting online social network data without verifying whether they are bots, expose the construct to bot contamination. Consider leadership work as an illustration, where network centrality is a commonly used approach for identifying online leaders (Faraj et al., 2015; Huffaker, 2010). If the intent of these studies is to identify *human* leaders, then the existence of *bots* (which, as we have shown, can be central in online social networks) is a threat to the validity of a "*human leader*."

This potential contamination may explain some of the theoretical anomalies in research. Faraj et al. (2015), for example, hypothesize that online leadership is associated with actor sociability. Yet, they found no support for this relationship. Instead, central participants were more likely to be identified as leaders if they also exhibited sociable behavior. They explained:

> ". . . even though sociability does not predict identification as a leader, actors who are central in the communication network and exhibit greater sociability are more likely to be recognized as leaders. In other words, socially oriented behavior does not lead to someone being identified a leader but, all things being equal, sociability by highly central participants leads to increased recognition as a leader" (Faraj et al., 2015: 406).

Author's voice:
What was the most difficult or challenging aspect of this research project and paper?

[13] Escorcia has a blog (http://loquesigue.tv) in which he reports the findings of his analyses.

[14] We also asked him to describe the goals of the bots, how they were used, who used them, why they were developed, and if there were different types of bots.

Although this explanation is entirely possible, what if both bots and humans comprise the central actors in their study? Is it possible that they found what they did because *bots* exhibit high centrality and low sociability, whereas *humans* manifest high centrality and a combination of high and low sociability? Would their findings remain the same if bots were removed from the analysis or would sociability be predictive of leadership only if humans were considered? Is the bot/human distinction consequential to their theoretical arguments? Maybe, if their theorizing is about *human* behavior, and maybe not, if their theorizing generalizes to any type of actor.

More specifically, our bot discovery raises three implications for online social network research. First, when the validity of an actor hinges upon its *humanness*, and the concept is not clearly defined, *theoretical arguments* associated with the actor construct can be invalid. Second, scholars theorizing about *human* actors must not only define the concept but they must also identify bots and control for their potential effects. Third, if the humanness of actors is irrelevant to theoretical arguments of a study, researchers must be careful in how they describe these actors as to avoid assuming they are people. In short, scholars must consider what constitutes an "actor," define the construct accordingly, and devise a research design that eliminates confounding effects caused by contamination across actor types, if the distinction is necessary.

Because the notion of actors in online social networks is presently ill-defined, we provide a revised definition of the concept for those who want to take a broader view and not restrict actors to humans. In doing so, we strive to effectively and concisely capture essential conceptual properties and characteristics. We redefine *actors* of online social networks as "discrete entities populating socio-technical networks." These actors do not need to interact with one another to exist although it is likely that they will. Our definition is sufficiently narrow in that it sets an online boundary condition but is also broad enough to capture various types of actors (e.g., humans and bots) operating in distinct networks (e.g., Facebook and Twitter).

Detecting Bots: A Multi-Method Funnel Approach

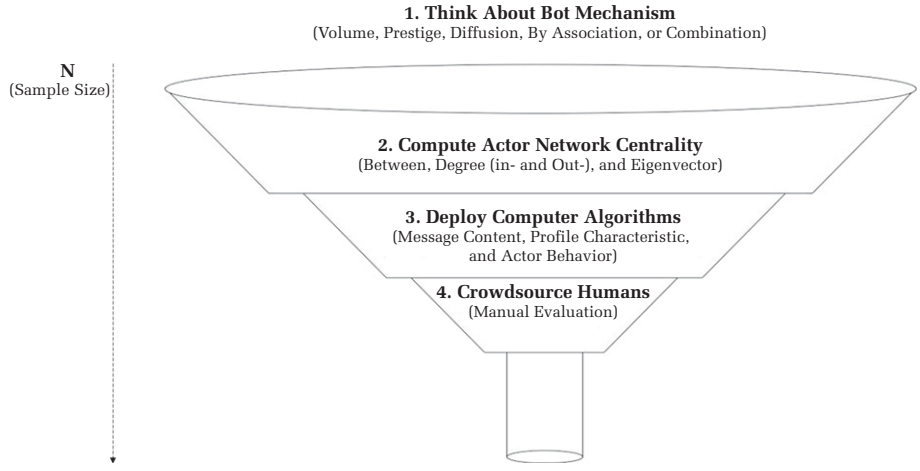
The prevalence of bots in online social networks coupled with the potential threats that they impose to organizations and society has sparked scholarly interest in distinguishing bots from humans (Cresci, Di Pietro, Petrocchi, Spognardi, & Tesconi, 2015; Davis, Varol, Ferrara, Flammini, & Menczer, 2016). Existing research often deploys feature-based machine-learning detection systems [e.g., the most popular method available for public use being *BotOrNot?*[15] (Davis et al., 2016)], but other common approaches involve the application of social networks (Paradise, Puzis, & Shabtai, 2014) and the crowdsourcing of human intelligence (Cao, Yang, Yu, & Palow, 2014). There is no consensus about which of these approaches is most effective, although it is evident that all have limitations. Existing social network techniques, for example, rely on the assumption that bots rarely have ties with humans and, therefore, tend to form their own communities (Wang et al., 2012). However, recent studies have shown this not to be the case; bots actually create links with people (Alvisi, Clement, Epasto, Lattanzi, & Panconesi, 2013) and, as a result, they do not form tight-knit groups (Yang, Wilson, Wang, Gao, Zhao, & Dai, 2014). Although human crowdsourcing can exhibit a near-zero false-positive rate, the method is not cost-effective for networks containing millions of users, such as Twitter. Finally, machine-learning techniques are problematic because of training sample dependency. Algorithms taught to detect bots generating content in English, for instance, produce less-accurate results in other idioms. Instead of one particular method, researchers suggest the adoption of complementary techniques (Alvisi et al., 2013; Ferrara et al., 2016) that explore multiple dimensions of actors' behaviors such as activity, timing information, and content (e.g., linguistic cues such as the frequency of verbs and nouns). Examples include the Renren Sybil (Wang, Konolige, Wilson, Wang, Zheng, & Zhao, 2013), the CopyCatch (Beutel, Xu, Guruswami, Palow, & Faloutsos, 2013), and the SynchroTrap (Cao et al., 2014).

We contacted the first authors of the three complementary techniques with questions about their applications. The Renren Sybil uses a clickstream methodology (i.e., a sequence of click events generated by users) and is therefore inappropriate for other types of data (e.g., numeric). We also learned that both CopyCatch and SynchroTrap are not yet available for public use. The lack of an existing publicly available approach that incorporates complementary techniques and uses traditional[16], graphical, and textual data to detect bots has compelled us to action. In the next few paragraphs, we present the *funnel process* (see Figure 3), a multi-method approach that is both theoretical and technical; we combine current techniques (social networks, machine-learning, and crowdsourcing of humans) with the goal of providing a holistic and precise, but also a feasible, approach scholars can use to best detect bots.

[15] http://truthy.indiana.edu/botornot/.

[16] Numerical, categorical, or binary (O'Neill & Schutt, 2013).

FIGURE 3
The Funnel Process: A Multi-Method Approach for Detecting Bots



We start with a simple assumption: not every bot poses a threat to research validity. Only those that actually have an impact on the online social network are likely to bias scholarly findings. Given our assumption, we advise researchers to first think about the *plausible sources (or mechanisms) of bot bias* by asking how might bots threaten the validity of their study's findings. In Table 5, we provide some guidance. For instance, political scientists may suspect that some of their sampled nodes are bots designed to create volume (i.e., noise) to manipulate political opinion online. To identify such bots, we recommend the computation of *actor network centrality* scores. Specifically, political bots designed to *create* noise are likely to have many *outgoing* ties, and so calculating out-degree centrality for all nodes in the sample to identify such important actors becomes appropriate here. To identify structural anomalies (i.e., prominent users who are potential bots), we recommend the median plus or minus two times the median absolute deviation (see Leys, Ley, Klein, Bernard, & Licata, 2013). Scholars can also think about a combination of mechanisms. For example, the *FemBots* contributing to Ashley Madison's revenue increases probably did not do so by only sending tons of messages to male users (out-degree) but also by receiving many responses back (in-degree). For those suspecting bots but who are unsure of their potential biasing mechanisms, we suggest the use of anomaly detection techniques in social networks (e.g., see Savage, Zhang, Yu, Chou, & Wang, 2014).

Once structural anomalies are identified, scholars can next *deploy two or more machine-learning algorithms* to further examine whether these are bots. Because every measure contains error, it is important to assess the reliability of user scores across different algorithms. *BotOrNot?* (Davis et al., 2016), BorOrNot (ABTO Software[17]), and *Boostor* (Morstatter et al., 2016) are options presently available for public use. They are feature based, meaning that they use the content of messages, characteristics of user profiles, and the behavior of actors to classify them as bots or humans. Different systems have different features and can generate different scores for the likelihood of a node being a bot. For nodes with non-converging scores (i.e., where different machine-learning algorithms do not agree whether the actor is a bot), we recommend *human crowdsourcing* as the next step. These human raters must be trained to identify and classify bots, humans, or even cyborgs (nodes exhibiting a mixture of human and bot features). Inter-rater reliability ought to be computed and disagreements resolved through discussion. With each stage of the process, we expect the actor sample size to decrease—identifying central actors based on theorized mechanisms should eliminate noncentral ones; algorithms should identify whether most actors are bots or humans; and what remains can be handled by human raters.

Embracing Bots: Opportunities for Theory Refinement and Development

Because bots can be used across a wide range of organizational phenomena, they create a variety of opportunities for theory development and refinement. We explore some possibilities by focusing on what we discovered—bots as *central* actors *amplifying and spreading content* about a *protest online*—to discuss specific ideas and directions for researchers in social movements, where protests matter (King & Soule, 2007); in leadership, where network centrality

[17] See http://botornot.co.

TABLE 5
Actor Network Centrality: Thinking About the Source of Bot Bias

Mechanism	Domain	Example	Centrality
Volume creation	Politics	Salud_dia1[a]	Out-degree
Prestige attainment	Gaming	Botgle	In-degree
Information diffusion	Ecology crisis	SFQuakeBot	Betweenness
By association	—	—	Eigenvector
Volume + prestige	Dating services	FemBots	*In + out*
Volume + diffusion	Protests	Anonymousfrai	*Between + out*
Prestige + diffusion	Leadership	—	*In + between*
Association + diffusion	—	—	*Eigenvector + between*

[a] *Note*: See http://emergencyjournalism.net/manipulation-of-public-opinion-in-venezuela-using-political-bots/

matters (Faraj et al., 2015); and in information diffusion, where content sharing matters (Berger & Milkman, 2012). We also encourage future research to systematically derive dimensions of social and technical context that make use of bots most relevant and effective.

Designing bots to mobilize collective behavior in social movements. Although we discovered that bots were designed to support activists in protesting corruption on Twitter, we do not have evidence to suggest that they influenced the court to charge 24 of the 25 prosecuted deputies in the Mensalão case with corruption crimes (Oliveira, 2014). But we know that to achieve social change, activists attempt to raise public awareness for their protest to mobilize collective action (Castells, 2012). Bots raised awareness for the Mensalão protest, by amplifying and diffusing #ChangeBrazil[18] messages, but it seems that mobilization was not achieved—only a small group of people joined in street demonstrations (BBC News, 2013). Thus, a clear question for future research is how we design bots to raise awareness of a social cause in a way that also mobilizes people to participate in offline collective action. This requires thinking beyond the embeddedness of a social network to ask questions concerning the content of bot-generated messages. For example, should they be designed to produce posts inciting an emotional state of shock? Although these can get people's attention and mobilize participation through a sense of urgency (Warren, 2010), what makes some audiences indignant and sympathetic may simply annoy the broader community (Jasper, 2014).

Another venue for future research is to examine the nuances of bot design across different social movement entities. For example, do social movement organizations (SMOs), such as Greenpeace, use bots differently than individual activists? SMOs may develop bots to mobilize people in signing digital petitions[19] over social media, whereas individual activists may develop bots to raise awareness of a protest on Twitter. Because SMOs are formally competing with one another for member support and resources in a social movement industry (McCarthy & Zald, 1977), they are also more concerned than individual activists with how their reputation affects their success in supporting a movement (Selander & Jarvenpaa, 2016). We therefore expect SMOs to pay close attention to how certain *e-tactics* (e.g., bot usage for mobilizing digital petitioning signatures) align with their core values as to not jeopardize support and resources from members. The difference in reputational concerns between the two entities raises questions about the underlying theories we can use to explain differences in how SMOs versus individual activists use bots, and the theories we can leverage to understand what constitutes an effective bot in each context.

Refining leadership theory and methods to consider "bot leaders" Considerable progress has been made in understanding the emergence of online leaders (Johnson, Safadi, & Faraj, 2015). Yet, our discovery points to the need for scholars to exploit what makes an effective bot leader, how bot leaders influence human behavior, and whether there are different styles of bot leadership. *Bot leader* attributes should not be assumed to be equivalent to those of *human leaders*. Such bots may incorporate characteristics that are essentially different, contain others that are similar, and either complement or substitute human leadership. For instance, to the extent that bots can influence people's behaviors, they seem to meet the definition of leadership, but to the extent that leaders need to exhibit independent thought and judgment in deciding *who* and *how* to

[18] To examine whether *anonymousfrai* and *guiql* raised awareness for the Mensalão protest, we plotted the network *with* and *without* these actors. We learned that, by removing them, we lose ~24 percent unique ties (337 to 255), ~51 percent duplicate ties (2,763 to 1,418), and ~2 percent unique users (from 259 to 252).

[19] Digital petitioning is one legitimate action that SMOs engage in with frequency (see Selander & Jarvenpaa, 2016).

influence and in *what direction* to influence, bots may not, at least for now, seem to be leaders[20]—they may just be tools of the human leaders deploying them.

Existing research articulates that we need to apply multiple theories to understand the emergence of leaders in online communities because no single theory "seems uniquely suited" (Johnson et al., 2015: 167). In the same spirit, we insinuate that a thorough study about the development of bot leaders also requires various theories. For example, a functional leadership view (Burke, Stagl, Klein, Goodwin, Salas, & Halpin, 2006) can help us understand which behaviors differentiate effective bot leaders from ineffective ones or even effective *bot* leaders from effective *human* leaders. In contrast, a shared leadership lens (Pearce & Sims, 2000) may explain the nuances of substitutive and complementary functions that bots and people perform. A substitution lens suggests identifying functions performed by human leaders that can be perfectly substituted by bots so that the performance of a community, as a whole, is improved. A complementary angle, on the other hand, suggests identifying distinct roles for bot and human leaders such that synergistic effects occur.

The interplay between actor type and network centrality is also worthy of attention. This work can provide a methodological contribution by clarifying which centrality measure should be used to predict online leadership contingent on the study's definition of an "actor" (i.e., online leaders in general, bot leaders, or human leaders). At the moment, existing literature ignores bots and finds itself debating whether leaders have high betweenness (Faraj et al., 2015; Fleming & Waguespack, 2007; Johnson et al., 2015), out-degree (Huffaker, 2010), or both high betweenness and out-degree centrality scores (Sutanto, Tan, Battistini, & Phang, 2011). Although our discovery shows that both bots and humans have high betweenness centrality, we found that only bots score high in out-degree centrality. Thus, it is possible that such measure is only valid for identifying *bot leaders*. We also recommend scholars to consider whether in-degree centrality should be used as a complementary metric in the recognition of human leadership because our findings show that it differentiated people from bots. Finally, because bots and humans have high betweenness centrality, and because this is a widely used measure to predict online leadership, it is likely fair to suspect that betweenness centrality is an appropriate measure for identifying online leaders in general (bots and humans). In answering these questions, scholars must also exploit *theoretical arguments* for these relationships. For instance, do bot leaders have out-degree centrality because they excessively post on a community's thread discussion as a way to potentially spark online dialogs? Similarly, do human leaders have high in-degree centrality because they are frequently mentioned by others for their tenure status and high quality knowledge contribution? Finally, do bot and human leaders have high betweenness centrality because they are both able to spread these types of (or other) information across the community?

Bots amplifying the diffusion of novel information through content sharing. Online content sharing is prevalent (Berger & Milkman, 2012) with 59 percent of people regularly retweeting messages, passing YouTube videos to relatives, and forwarding Amazon product reviews to colleagues (Allsop, Bassett, & Hoskins, 2007). Although we know that the *sharing* of online content is both frequent and relevant, less is known about *how* to actually engineer it as to amplify the diffusion of novel information. This is, we believe, an opportunity for bot research.

We know from prior work that weak ties are more likely to provide novel information (Granovetter, 1973) and promote, in a proactive way, content sharing (Shi, Rui, & Whinston, 2014). This knowledge along with our bot discovery suggests that bots can be built to *amplify* the diffusion of novel information if they are designed to simultaneously be weak ties and content sharers. They need to have both high betweenness (weak tie) and out-degree (content sharing) centrality. An interesting question for future research is how to design *weak-tie content-sharing* bots. To do so, scholars must think beyond bot-level attributes to also consider how *platform infrastructure* (e.g., Twitter's 140-character limit) and *social network properties* enable or constrain bot design. Equally important is to recognize that although most of the content diffused by bots today is created by humans, this does not need to be the case.[21] It is also important to examine the effects of bot usage. For instance, one of the objectives associated with the amplification and diffusion of novel information is to exert influence. We suspect that bots can be used to influence people to change their behaviors through *awareness* and *social learning* (Aral, 2011). For example, bots automatically sharing content about Pokémon Go's augmented reality feature with acquaintances may influence people to download the application by making them *aware* of

[20] Advancements in cognitive computing (e.g., IBM's Watson) may change this.

[21] Advancements in cognitive computing (e.g., IBM's Watson) are increasingly enabling the creation of new knowledge by bots, and therefore, we may see an increase in the number of bots sharing novel knowledge.

the feature. Similarly, bots sharing fun stories about Pokémon Go may impact the number of downloads by exposing non-adopters to the benefits associated with playing the game. An understanding of the mechanisms via which bot design enables the diffusion of novel information through content sharing is a necessary first step in our inquiry into bot influence.

Practical Bot Implications

Protests and corporations. Our discovery has an important implication for businesses targeted by protests and boycotts. Think about the potential use of bots for social appropriateness and boycott prevention. McDonnell and King (2013) noted that prosocial claims—expressions of the organization's commitment to socially acceptable norms, beliefs, and activities—function as an impression management strategy corporations use to neutralize reputational threats caused by boycotts. They found that a large increase in prosocial claims occurs when a boycott is more threatening (i.e., it receives more media attention), when a firm has a higher reputation, or when a company has a history of engaging in prosocial claims. We claim that these same businesses do not have to wait for boycotts and protests to gain popularity to engage in prosocial claims. Instead, they can use bots to prevent these social movement tactics from occurring. For example, they can design bots to frequently broadcast the firm's social appropriateness online. Another strategy involves coding bots to automatically engage in prosocial claims when negative images and grievances about the corporation begin to circulate on social media. This requires linking them to social listening technologies that monitor conversations about an organization's image. These technologies then need to prompt bots—once a negative keyword about the organization emerges—to immediately generate content countering grievances made by activists with positive claims that emphasize the firms' commitment to social norms. In doing so, they can maintain audience support without recognizing or legitimizing activists' declarations.

From an activist standpoint, our work has implications related to demand concession. Corporate targets are more likely to concede to boycotts that generate large amounts of media attention because "they see sustained media attention to a boycott as an indicator of public support for the boycotters' cause and a signal that the boycott, if not ended, could lead to revenue loss" (King, 2008: 400). Because many people use online social networks (e.g., Twitter) to consume news (Holcomb, Gottfried, & Mitchell, 2013), traditional media outlets (e.g., CNN) now actively report on nearly half of the most popular topics discussed in these social networks (Carrascosa, Cuevas, Gonzalez, Azcorra, & Garcia, 2015). Therefore, protestors can get traditional media to pay attention to their cause by designing bots to facilitate the popularity of their boycott online. Essentially, these bots need to amplify the channels in which activists transmit their grievances to obtain rapid and large public support for the boycott. To achieve this, they may want to design bots to constantly share and diffuse content that either explains the legitimacy of the boycott or exposes factual documents leading to immediate reputational harm of the targeted organization.

CONCLUSION

Our main discovery is that bots are central actors in online social networks. Our data show that these bots raise awareness of a protest on Twitter not because they frequently report on the status of street demonstrations or riots but because they automatically share content (via retweets) on specific topics (by targeting certain hashtags) related to the protest. The discovery enabled us to challenge the assumption that actors of online social networks are people, also leading us to more clearly define the concept. We hope this clarification improves the validity of future work in the field and that online social network scholars find our methodology to detect bots useful. We also hope our discovery spurs further research on bots in organizational and social contexts. In particular, we suggest that scholars consider bot design in the mobilization of collective action, revise existing theories and methods for identifying online leaders, explore the notion of *weak-tie content-sharing* bots, and examine the many different mechanisms in which bots may influence changes in human behavior. Finally, we hope that our work helps inform researchers, practitioners, and policymakers about Twitter's bot usage in online activism worldwide (both for protesting societal issues but also for protesting against certain organizational actions), as they move forward with decisions in this important domain.

REFERENCES

Allsop, D. T., Bassett, B. R., & Hoskins, J. A. 2007. Word-of-mouth research: Principles and applications. ***Journal of Advertising Research***, 47(4): 388–411.

Alvesson, M., & Sandberg, J. 2011. Generating research questions through problematization. ***Academy of Management Review***, 36(2): 247–271.

Alvisi, L., Clement, A., Epasto, A., Lattanzi, S., & Panconesi, A. 2013. ***Sok: The evolution of Sybil defense via***

social networks. In 2013 IEEE symposium on security and privacy. IEEE, 382–396.

Aral, S. 2011. Commentary—Identifying social influence: A comment on opinion leadership and social contagion in new product diffusion. ***Marketing Science***, 30(2): 217–223.

Aral, S., & Walker, D. 2014. Tie strength, embeddedness, and social influence: A large-scale networked experiment. ***Management Science***, 60(6): 1352–1370.

BBC. 2013. Q&A: Brazil's 'big monthly' corruption trial. http://www.bbc.com/news/world-latin-america-19081519. Accessed July 30, 2015.

Berger, J., & Milkman, K. L. 2012. What makes online content viral? ***Journal of Marketing Research***, 49(2): 192–205.

Beutel, A., Xu, W., Guruswami, V., Palow, C., & Faloutsos, C. 2013. ***Copy-catch: Stopping group attacks by spotting lockstep behavior in social networks***. In Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee, 119–130.

Borgatti, S. P., & Halgin, D. S. 2011. On network theory. ***Organization Science***, 22(5): 1168–1181.

Burke, C. S., Stagl, K. C., Klein, C., Goodwin, G. F., Salas, E., & Halpin, S. M. 2006. What type of leadership behaviors are functional in teams? A meta-analysis. ***Leadership Quarterly***, 17(3): 288–307.

Butler, B. S., Bateman, P. J., Gray, P. H., & Diamant, E. I. 2014. An attraction–selection–attrition theory of online community size and resilience. ***MIS Quarterly***, 22(3): 699–728.

Carrascosa, J. M., Cuevas, R., Gonzalez, R., Azcorra, A., & Garcias, D. 2015. Quantifying the economic and cultural biases of social media through trending topics. ***PLoS One***, 10(7): 1–14.

Cao, Q., Yang, X., Yu, J., & Palow, C. 2014. ***Uncovering large groups of active malicious accounts in online social networks***. In Proceedings of the 2014 ACM SIGSAC conference on computer and communications security. ACM, 477–488.

Castells, M. 2012. ***Networks of outrage and hope: Social movements in the Internet age***. Cambridge, UK: Polity Press.

Charmaz, K. 2006. ***Constructing grounded theory: A practical guide through qualitative analysis***. Thousand Oaks, CA: Sage.

Coleman, G. 2014. ***Hacker, hoaxer, whistleblower, spy: The many faces of Anonymous***. New York: Verso.

Cresci, S., Di Pietro, R., Pretocchi, M., Spognardi, A., & Tesconi, M. 2015. Fame for sale: Efficient detection of fake Twitter followers. ***Decision Support Systems***, 80: 56–71.

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. 2016. ***BotOrNot: A system to evaluate social bots***. Proceedings of the 25th International Conference Companion on World Wide Web, 273–274.

Dhar, V., Geva, T., Oestreicher-Singer, G., & Sundararajan, A. 2014. Prediction in economic networks. ***Information Systems Research***, 25(2): 264–284.

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. 2011. Knowledge collaboration in online communities. ***Organization Science***, 22(5): 1224–1239.

Faraj, S., Kudaravalli, S., & Wasko, M. 2015. Leading collaboration in online communities. ***MIS Quarterly***, 39: 393–412.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. 2016. The rise of social bots. ***Communications of the ACM***, 59(7): 96–104.

Fleming, L., & Waguespack, D. M. 2007. Brokerage, boundary spanning, and leadership in open innovation communities. ***Organization Science***, 18(2): 165–180.

Freeman, L. C. 1979. Centrality in social networks conceptual clarification. ***Social Networks***, 1(3): 215–239.

Gallagher, E. 2015. Mexican Twitter bots: Cyber-attacks on free speech & organizing. http://revolution-news.com/mexican-twitter-bots-cyber-attacks-free-speech-organizing/. Accessed July 30, 2015.

Goldman, D. 2014. 23 million Twitter users are fed by robots. http://money.cnn.com/2014/08/12/technology/social/twitter-bots/. Accessed May 17, 2016.

Granovetter, M. 1973. The strength of weak ties. ***The American Journal of Sociology***, 78(6): 1360–1380.

Grant, R. 2014. Facebook has no idea how many fake accounts it has—But it could be nearly 140 M. http://venturebeat.com/2014/02/03/facebook-has-no-idea-how-many-fake-accounts-it-has-but-it-could-nearly-140m/. Accessed May 17, 2016.

Hansen, D. L., Shneiderman, B., & Smith, M. A. 2011. ***Analyzing social media networks with NodeXL: Insights from a connected world***. Burlington, MA: Morgan Kaufmann.

Holcomb, J., Gottfried, J., & Mitchell, A. 2013. News use across social media platforms. http://www.journalism.org/2013/11/14/news-use-across-social-media-platforms/. Accessed July 29, 2016.

Huffaker, D. 2010. Dimensions of leadership and social influence in online communities. ***Human Communication Research***, 36(4): 593–617.

Jasper, J. M. 2014. ***Protest: A cultural introduction to social movements***. Cambridge, UK: Polity Press.

Johnson, S. L., Safadi, H., & Faraj, S. 2015. The emergence of online community leadership. ***Information Systems Research***, 26(1): 165–187.

King, B. G. 2008. A political mediation model of corporate response to social movement activism. ***Administrative Science Quarterly***, 53(3): 395–421.

King, B. G., & Soule, S. A. 2007. Social movements as extra-institutional entrepreneurs: The effect of protests on stock price returns. ***Administrative Science Quarterly***, 52(3): 413–442.

Kumar, N., & Benbasat, I. 2006. Research note: The influence of recommendations and consumer reviews on evaluations of websites. ***Information Systems Research***, 17(4): 425–439.

Leahy, J. 2012. Brazil gripped by anti-corruption battle. http://www.ft.com/cms/s/0/1b686ccc-21b1-11e2-b5d2-00144feabdc0.html#axzz3hOf8dRGJ. Accessed July 30, 2015.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. ***Journal of Experimental Social Psychology***, 49(4): 764–766.

Lipkin, M. 2014. LinkedIn sues bots over member data scraping. http://www.law360.com/articles/499109/linkedin-sues-bots-over-member-data-scraping. Accessed May 17, 2016.

Lyons, J., & Cowley, M. 2013. Brazil court allows corruption case appeals. http://www.wsj.com/articles/SB10001424127887323808204579083 571863389380. Accessed July 30, 2015.

MacKenzie, S. B. 2003. The dangers of poor construct conceptualization. ***Journal of the Academy of Marketing Science***, 31(3): 323–326.

McCarthy, J. D., & Zald, M. N. 1977. Resource mobilization and social movements: A partial theory. ***American Journal of Sociology***, 82(6): 1212–1241.

McDonnell, M., & King, B. 2013. Keeping up appearances: Reputational threat and impression management after social movement boycotts. ***Administrative Science Quarterly***, 58(3): 387–419.

Messias, J., Schmidt, L., Oliveira, R., & Benevenuto, F. 2013. You followed my bot! Transforming robots into influential users in Twitter. ***First Monday***, 18: 7.

Miroff, N., & Booth, W. 2012. Peña Nieto is winner of Mexican election. https://www.washingtonpost.com/world/the_americas/mexico-presidential-election-underway/2012/07/01/gJQAyd96FW_story.html. Accessed May 17, 2016.

Monroy-Hernández, A., & Spiro, E. 2013. How Brazilian protesters are using Twitter. http://www.theguardian.com/news/datablog/2013/jul/04/brazilian-protesters-twitter-microsoft. Accessed July 30, 2015.

Morstatter, F., Wu, L., Tahora, N. H., Carley, K., & Liu, H. 2016. A new approach to bot detection: Striking the balance between precision and recall. ***IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)***: 533–540.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. 2003. ***Scaling procedures: Issues and applications***. Thousand Oaks, CA: Sage.

Newitz, A. 2015a. Almost none of the women in the Ashley Madison database ever used the site. http://gizmodo.com/almost-none-of-the-women-in-the-ashley-madison-database-1725558944. Accessed November 8, 2015.

Newitz, A. 2015b. Ashley Madison code shows more women, and more bots. http://gizmodo.com/ashley-madison-code-shows-more-women-and-more-bots-1727613924. Accessed November 8, 2015.

Newitz, A. 2015c. How Ashley Madison hid its fembot con from users and investigators. The full title of the work it is part of http://gizmodo.com/how-ashley-madison-hid-its-fembot-con-from-users-and-in-1728410265. Accessed November 8, 2015.

Newitz, A. 2015d. One chart that shows how much money Ashley Madison made using bots. http://gizmodo.com/one-chart-that-shows-how-much-money-ashley-madison-made-1727821132. Accessed November 8, 2015.

Oliveira, M. 2014. Após um ano e meio e 69 sessões, STF conclui julgamento do mensalão. http://g1.globo.com/politica/mensalao/noticia/2014/03/apos-um-ano-e-meio-e-69-sessoes-stf-conclui-julgamento-do-mensalao.html. Accessed November 18, 2015.

Oh, O., Agrawal, M., & Rao, H. R. 2013. Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. ***MIS Quarterly***, 37: 407–426.

Oh, O., Eom, C. M., & Rao, H. R. 2015. Research note—Role of social media in social change: An analysis of collective sense making during the 2011 Egypt revolution. ***Information Systems Research***, 26(1): 210–223.

Okalow, S. 2015. Investigating social media's spam-bot problem. http://www.b2bnn.com/2015/08/investigating-social-medias-spam-bot-problem/. Accessed May 17, 2016.

O'Neill, C., & Schutt, R. 2013. ***Doing data science***. Sebastopol, CA: O'Reilly Media.

O'Reilly, L. 2015. 8% of Instagram accounts are fakes and 30% are inactive, study says. The full title of the work it is part of http://www.businessinsider.com/italian-security-researchers-find-8-percent-of-instagram-accounts-are-fake-2015-7. Accessed May 17, 2016.

Paradise, A., Puzis, R., & Shabtai, A. 2014. Anti-reconnaissance tools: Detecting targeted socialbots. ***Internet Computing***, 18(5): 11–19.

Pearce, C. L., & Sims, H. P. 2000. Shared leadership: Toward a multi-level theory of leadership. In M. M. Beyerlein, D. A. Johnson, & S. T. Beyerlein (Eds.), ***Advances in interdisciplinary studies of work teams***, vol. 7: 115–139. Greenwich, CT: Emerald Group Publishing Limited.

Perez, S. 2011. Tumblr acknowledges its growing spam problem, says it's doing everything it can. The full title of the work it is part of https://techcrunch.com/2011/10/31/tumblr-acknowledges-its-growing-spam-problem-says-its-doing-everything-it-can/. Accessed August 23, 2016.

Podsakoff, P. M., Mackenzie, S. B., & Podsakoff, N. P. 2016. Recommendations for creating better concept definitions in the organizational, behavioral, and social sciences. ***Organizational Research Methods***, 19(2): 159–203.

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Flammini, A., & Menczer, F. 2011. ***Detecting and tracking political abuse in social media***. In 5th international AAAI conference on weblogs and social media, 297–304.

Reddit. 2015. We are the impact team. We are releasing the Ashley Madison data. The full title of the work it is part of https://www.reddit.com/r/AnythingGoesNews/comments/3h71ar/we_are_the_impact_team_we_are_releasing_the/. Accessed November 8, 2015.

Robertson, J., Riley, M., & Willis, A. 2016. How to hack an election. http://www.bloomberg.com/features/2016-how-to-hack-an-election/. Accessed July 12, 2016.

Savage, D., Zhang, X., Yu, X., Chou, P., & Wang, Q. 2014. Anomaly detection in online social networks. ***Social Networks***, 39(1): 62–70.

Schumann, S. 2014. ***How the Internet shapes collective actions***. London: Palgrave Pivot.

Selander, L., & Jarvenpaa, S. L. 2016. Digital action repertoires and transforming a social movement organization. ***MIS Quarterly***, 40(2): 331–352.

Shi, Z., Rui, H., & Whinston, A. 2014. Content sharing in a social broadcasting environment: Evidence from Twitter. ***MIS Quarterly***, 38: 123–142.

Singer, S. 2013. Between executioners and gangsters. The full title of the work it is part of http://www1.folha.uol.com.br/internacional/en/ombudsman/2013/09/1346607-between-executioners-and-gangsters.shtml. Accessed July 30, 2015.

Sutanto, J., Tan, C. H., Battistini, B., & Phang, C. W. 2011. Emergent leadership in virtual collaboration settings: A social network analysis approach. ***Long Range Planning***, 44(5): 421–439.

The Economist. 2013. What is Brazil's "mensalão"? The full title of the work it is part of http://www.economist.com/comment/2216108. Accessed July 30, 2015.

Tufekci, Z. 2014. What happens to #Ferguson affects Ferguson: Net neutrality, algorithmic filtering and Ferguson. The full title of the work it is part of https://medium.com/message/ferguson-is-also-a-net-neutrality-issue-6d2f3db51eb0. Accessed November 18, 2015.

Wang, G., Konolige, T., Wilson, C., Wang, X., Zheng, H., & Zhao, B. Z. 2013. You are how you click: Clickstream analysis for Sybil detection. Proceedings of the 22nd USENIX conference on Security, 241–256.

Wang, G., Mohanlal, M., Wilson, C., Wang, X., Metzger, M., Zheng, H., & Zhao, B. Y. 2012. Social Turing tests: Crowdsourcing Sybil detection. Proceedings of Annual Network Distributed System Security Symposium (NDSS) San Diego, CA, February 2013.

Warren, M. R. 2010. ***Fire in the heart: How white activists embrace racial justice***. New York: Oxford University Press.

Wasserman, S., & Faust, K. 1994. ***Social network analysis: Methods and applications (structural analysis in the social sciences)***. Cambridge, UK: Cambridge University Press.

Yang, Z., Wilson, C., Wang, X., Gao, T., Zhao, B.Y., & Dai, Y. 2014. Uncovering social network Sybils in the wild. ***ACM Transaction on Knowledge Discovery from Data***, 8(1): 259–268.

Zeng, X., & Wei, L. 2013. Social ties and user content generation: Evidence from Flickr. ***Information Systems Research***, 24(1): 71–87.

Carolina Alves de Lima Salge (csalge@uga.edu) is a doctoral student in the MIS Department at the University of Georgia. Her research focuses on the use and impact of social bots in the context of online social movements as well as on social bot ethics.

Elena Karahanna (ekarah@uga.edu) is Distinguished Research Professor and L. Edmund Rast Professor of Business in the MIS Department, University of Georgia. She has published in *MIS Quarterly, Information Systems Research, Management Science*, *Organization Science*, and elsewhere and has served as senior editor for *MIS Quarterly* and *Information Systems Research* and as associate editor for *Management Science*.