Noveck, Beth Simone (2003). 'Designing deliberative democracy in cyberspace: the role of the cyber-lawyer'. *Boston University Journal of Science and Technology Law* 9: 1.
Okada, Alexandra, Simon J. Buckingham Shum and Tony Sherborne (2008). *Knowledge Cartography: Software Tools and Mapping Techniques*. Amsterdam: Springer.
Parkinson, John and Jane Mansbridge (eds) (2012). *Deliberative Systems: Deliberative Democracy at the Large Scale*. New York: Cambridge University Press.
Pickard, Victor W. (2008). 'Cooptation and cooperation: institutional exemplars of democratic internet technology'. *New Media and Society* 10(4): 625–645.
Pingree, Raymond J., T. Davies and S.P. Gangadharan (2009). 'Decision structure: a new approach to three problems in deliberation'. In T. Davies and S.P. Gangadharan (eds), *Online Deliberation: Design, Research and Practice*. San Francisco, CA: CSLI Publications.
Rhee, June W. and Eun-mee Kim (2009). 'Deliberation on the net: lessons from a field experiment'. In T. Davies and S.P. Gangadharan (eds), *Online Deliberation: Design, Research and Practice*. San Francisco, CA: CSLI Publications.
Rittel, Horst W.J. and Melvin M. Webber (1973). 'Dilemmas in a general theory of planning'. *Policy Sciences* 4(2): 155–169.
Schlosberg, David, Stephen Zavestoski and Stuart W. Shulman (2007). 'Democracy and e-rulemaking: web-based technologies, and the potential for deliberation'. eRulemaking Research Group, Paper 1, University of Massachusetts – Amherst.
Schuler, Douglas (2009). 'Online civic deliberation with E-Liberate'. In T. Davies and S.P. Gangadharan (eds), *Online Deliberation: Design, Research and Practice*. San Francisco, CA: CSLI Publications.
Sobieraj, Sarah and Jeffrey M. Berry (2011). 'From incivility to outrage: political discourse in blogs, talk radio, and cable news'. *Political Communication* 28(1): 19–41.
Steiner, Jürg (2012). *The Foundations of Deliberative Democracy: Empirical Research and Normative Implications*. New York: Cambridge University Press.
Sunstein, Cass R. (2002). 'The law of group polarization'. *Journal of Political Philosophy* 10(2): 175–195.
Thimm, Caja, Mark Dang-Anh and Jessica Einspänner (2014). 'Mediatized politics – structures and strategies of discursive participation and online deliberation on Twitter'. In Andreas Hepp and Friedrich Krotz (eds), *Mediatized Worlds: Culture and Society in a Media Age*. London: Palgrave.
Tucey, Cindy Boyles (2010). 'Online vs. face-to-face deliberation on the global warming and stem cell issues'. Western Political Science Association 2010 Annual Meeting Paper.
Upadhyay, Meenakshi (2014). 'Political deliberation on Twitter: is Twitter emerging as an opinion leader?', International Conference on People, Politics and Media (ICPPM), 25–26 April, Jagran Lakecity University.
Wilhelm, Anthony G. (2000). *Democracy in the Digital Age: Challenges to Political Life in Cyberspace*. New York: Routledge.
Wright, Scott (2006). 'Government – run online discussion fora: moderation, censorship and the shadow of control'. *British Journal of Politics and International Relations* 8(4): 550–568.
Wright, Scott (2009). 'The role of the moderator: problems and possibilities for government-run online discussion forums'. In T. Davies and S.P. Gangadharan (eds), *Online Deliberation: Design, Research and Practice*. San Francisco, CA: CSLI Publications.
Wright, Scott and John Street (2007). 'Democracy, deliberation and design: the case of online discussion forums'. *New Media and Society* 9(5): 849–869.

16. Computational approaches to online political expression: rediscovering a 'science of the social'

Dhavan V. Shah, Kathleen Bartzen Culver, Alexander Hanna, Timothy Macafee and JungHwan Yang

It is a curious fact that the empirical study of political talk, particularly online exchanges, is increasingly traced back to the social interactionism of nineteenth-century French sociologist Gabriel Tarde. As Terry Clark (1969) and Elihu Katz (2006) remind us, Tarde argued for conversation's place at the center of sociological inquiry, articulating a complex theory of 'inter-mental activity' concerning how people influence one another. In so doing, he developed the concepts that later became known as the two-step flow of communication and opinion leadership, among other propositions of interpersonal influence (Lazarsfeld et al., 1944; Berelson et al., 1954; Katz and Lazarsfeld, 1955). As Tarde writes about the late nineteenth century (1969 [1898]: 313), 'newspapers have transformed . . . the conversations of individuals, even those who do not read papers but who, talking to those who do, are forced to follow the groove of their borrowed thoughts. One pen suffices to set off a million tongues'. His thesis still holds true, with televised events such as the first 2012 presidential debate in the USA generating more than 10 million tweets in just a few hours, including many retweets of major accounts (Hanna et al., 2013).

Tarde was particularly concerned with the relationship between mass communication and interpersonal conversation for the formation of publics and their opinions. Katz (2006: 267) describes this mediated process as follows: 'To the press, he assigned the role of creating a public . . . The press, then, sets an agenda for the conversation of the cafes. Opinions are clarified and crystallized in these conversations, and then translated into actions in the world of politics'. The central tenets of this ordered model – press, conversation, opinion, and action – are supported by research on multi-step flow (Rogers and Shoemaker, 1971), opinion leadership (Shah and Scheufele, 2006), and communication mediation (Lee et al., 2013). In digital media environments, the sources of information and sites of conversation have begun to converge – amplified and reinforced within seemingly polarized ecologies – suggesting important directions for research on opinion and action in networked societies.

The Laws of Imitation (*Les lois de l'imitation*) (1903 [1890]), Tarde's most widely known work in English, speaks to communication and social influence within such settings. It also marks him as the 'founding father of innovation diffusion research' (Kinnunen, 1996), charting processes of communicative invention, reproduction, and opposition. Methodologically, his calls for attention to observable interpersonal interactions, particularly within conversational processes, presage the approaches central to computational social science, where each interaction 'leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies' (Lazer et al., 2009: 721). The technological affordances of social media permit tracking of message creation and expression within a network, as well as reception and diffusion through systems (Namkoong et al., 2010; Han et al., 2011). Tarde's insights about 'invention' and 'imitation' provide ways to study online talk as it intersects with deliberative democracy, informed opinion, and participatory citizenship (Schudson, 1978; Barber, 1984; Habermas, 1984; Kim et al., 1999; Price and Cappella, 2002; Mutz, 2006).

SITUATING POLITICAL TALK

Tarde, often presented as the foil to Emile Durkheim's efforts to distinguish the study of society from that of human psychology,[1] was profoundly concerned with the interplay of mental and social forces, particularly as seen in the locus of conversational exchanges. As Berelson et al. wrote in *Voting* (1954: 300), Tarde 'was convinced that opinions are really formed through the day-to-day exchanges of comments and observations which goes on among people . . . by the very process of talking to one another, the vague dispositions which people have are crystallized, step by step, into specific attitudes, acts, and votes'. From this perspective, conversation is not simply a site of networked information exchange, but also an opportunity for the composition and clarification of one's own views (see Pingree, 2007).

In the present day, this position seems all the more correct, especially in online environments, where political expression and conversation are increasingly common and visible. These environments also lend themselves to the sort of large-scale, highly detailed interactional analysis that Tarde's approach advocated. As Bruno Latour (2010) recently recognized, 'it is indeed striking that at this very moment, the fast expanding fields of "data

visualisation," "computational social science," or "biological networks" (Lazer et al., 2009; Wimsatt, 2007) are tracing, before our eyes, just the sort of data Tarde would have acclaimed'. 'Big data' provide a way to understand everyday political talk online, its triggers, content, and structures.

Such close analyses of political talk, whether face-to-face or online, have typically been restricted to ethnographic or content analytic research. Works such as William Gamson's (1992) classic study, *Talking Politics*, used analysis of small-group discussions around affirmative action, nuclear power, the Arab–Israeli conflict, and US industry to counter the conventional wisdom of an uninformed and inactive electorate. Taking a similarly granular approach, Papacharissi (2004) examined the quality of online political talk by studying the level of civility in 287 discussion threads drawn randomly from 147 political newsgroups, concluding that discussions were civil but heated. Somewhat similarly, Walsh (2012) used participant observation of 37 reoccurring groups from 27 distinct communities across the state of Wisconsin to show how class- and place-based identity is linked with perceptions of relative deprivation.

Complementing this work on the actual content of conversations, research has also employed cross-sectional and panel survey methods to examine the causes and consequences of political talk (Huckfeldt and Sprague, 1995). Most notably, work by Jack McLeod et al. has examined the role of interpersonal conversation for community activism and political engagement (McLeod et al., 1999). Emphasizing the mediating roles of news and talk for participation in public life, communication behaviors are thought to shape and amplify the impact of background characteristics on citizens' engagement in democratic societies (McLeod et al., 1996). Various inquiries about the roles of media and conversation have coalesced into 'communication mediation models', which conclude that news consumption and political talk largely channel the effects of demographics, ideology, and social structure on outcome orientations and participatory responses (Sotirovic and McLeod, 2004; Cho et al., 2009).

This process has been further specified in the form of a 'citizen communication mediation model' (Shah et al., 2005; Shah et al., 2007), which theorizes and finds, consistent with Tarde's framework, that media influences are strong, but largely indirect, shaping opinion and action through effects on face-to-face and online discussion about news. Informational use of media, particularly newspaper and online news use, is found to stimulate expression and discussion through interpersonal and computer-mediated political talk, channeling effects on to engagement. The power of digital pathways to participation – from both conventional and online news sources through digital messaging – was particularly strong among the youngest generational groups (Lee et al., 2013).

POLITICAL TALK VIA DIGITAL MEDIA

Face-to-face and online talk share many virtues. Both spur compositional processing, mental elaboration, attitude crystallization, cross-cutting exposure, media reflection, knowledge gain, and mutual understanding (Shah et al., 2005; Mutz, 2006; Lee et al., 2013; Valenzuela et al., 2012). But political conversation via digital platforms may have certain advantages over face-to-face talk, as well. First, digital media often provide a source of political information and a sphere for political expression, readily facilitating their interplay and participatory consequences (Dahlgren, 2000, 2005). Second, the functionality of many online media emphasize discursive elements such as walls posts, online chat, photo sharing, and social networking, highlighting opportunities for public-spirited talk (Bennett et al., 2011; Boyd and Ellison, 2007). Third, generating user-created content, whether images, videos, or posts, may demand deeper forms of reflective and compositional processing (Freelon, 2010; Ekström and Östman, 2013). Regardless of the setting, 'conversation provides people with the opportunity to think through their "idea elements" and reduce cognitive inconsistency' (Kim et al., 1999: 363).

Less is known about the specific consequences of online political talk, such as whether deeper dialogue contributes to the enhancement of opinion quality and the development of efficacy (Kim et al., 1999). There is also some question as to the value of measuring behaviors in the digital world using analogue tools such as survey instruments, which rely on self-reported measures of network composition, structural heterogeneity, and discussion frequency (see Sotirovic and McLeod, 2001; Eveland and Hively, 2009; Kwak et al., 2005), rather than precise measures of content-specific message expression, reception, response, and repetition (see Han et al., 2011; Namkoong et al., 2010). In addition, it may be important to look across different conversational settings, such as messaging platforms (Shah et al., 2007; Hardy and Scheufele, 2005), political blogs (Gil de Zúñiga et al., 2009), and social networking sites (Bode et al., 2014), all of which have been linked to civic engagement and political action.

Social networking sites bring together the most powerful features of online interaction, messaging in real time and asynchronously, posting to smaller circles and to full networks, exchanging information and providing emotional support, creating new content and sharing the ideas of others, starting a group and joining those created by others; all seemingly strengthening social ties, albeit in unique registers to differing degrees (Boyd and Ellison, 2007; Pasek et al., 2009; Gil de Zúñiga et al., 2012). One recent study examines the consequences of political social networking behaviors, such as displaying a political preference on a profile, becoming

a fan or a friend of a politician, joining a cause or political group, and using a news or politics application (Bode et al., 2014). It finds that such uses shape political participation above and beyond the effects of offline talk and online expression.

SOCIAL NETWORKING AROUND POLITICS

There is little doubt that social media provide a forum for the discussion of political issues, a system through which to recruit individuals to participate in pertinent political issues, and a means to find people who share similar political opinions (Papacharissi, 2010). In this sense, political engagement through social media may represent a shift from top-down political communication and participation to one that is interpersonal or bottom-up (Thackery and Hunter, 2010). It is also characterized by feedback mechanisms to elites, providing opinion leaders with an opportunity to shape the thinking of political and media elites if they can 'set off a million tongues'. Social media have certainly increased candidates' digital exposure, permitting communication with volunteers, donors, and constituents (Gueorguieva, 2008).

Alongside Facebook (Vitak et al., 2011) and YouTube (Dylko et al., 2012), Twitter has emerged as a major location of political interaction (Hanna et al., 2013; Tumasjan et al., 2010). Much of this work emphasizes the use of social media channels by politicians during elections (Bruns and Highfield, 2013). Studies of members of the US Congress have found that they use Twitter primarily for information dissemination, with little retweeting or hashtag use, suggesting elite efforts to lead opinion rather than echoing the ideas of others (Golbeck et al., 2010; Gainous and Wagner, 2014). Efforts to examine whether Twitter can predict election results are more mixed. Based on a sentiment analysis of more than 100,000 messages referencing a political party or politician in the 2009 German federal election, Tumasjan et al. (2010) found that Twitter 'is used extensively for political deliberation and that the mere number of party mentions accurately reflects the election result'. However, an alternative analysis of the same election found no link and argued that the predictive power of the prior study was a consequence of the 'arbitrary choices of the authors' (Jungherr et al., 2012: 229).

Regardless of its predictive power, it is widely acknowledged that political talk on Twitter is triggered by media happenings and news coverage (Hanna et al., 2013; Graham and Hajru, 2011). Only a handful of studies have examined the types of political expression that occur within social networking sites, with even fewer considering how citizens self-organize under certain banners or hashtags (that is, user-generated keywords organized around the # symbol). Given that 'hashtags are used to bundle together tweets on a unified, common topic, and that the senders of these messages are directly engaging with one another', they provide a potentially powerful way to track everyday political expression (Bruns and Burgess, 2011: 5). Hashtags can become identified with specific issue positions and protest efforts. Along these lines, Segerberg and Bennett (2011) analyzed hashtag usage around the 2009 United Nations Climate Change Conference, observing the strategic deployment of certain terms to organize action.

As noted above, Twitter also allows for the tracking of 'inventive' message expression and 'imitation' by others in the network as indicators of their recirculation and diffusion through systems (Tarde, 1903 [1890]; Rogers and Shoemaker, 1971). At the most basic level, this can be understood in terms of tweets and retweets (Boyd et al., 2010), although more sophisticated analyses have examined the flow of influence within social networking systems. These studies conclude that Twitter is an excellent medium for message propagation, finding that 37.1 percent of message flows spread more than three degrees of separation away from the original sender (Ye and Wu, 2010). Of course, not all issues are created equal. Tracing the diffusion of hashtags on Twitter, Romero et al. (2011: 695) find significant variation across topics, with hashtags on politically controversial topics 'particularly persistent, with repeated exposures continuing to have unusually large marginal effects on adoption'. Influence in online social networks is not simply a function of size of follower networks, at least as measured in retweets or user mentions (Cha et al., 2010). In fact, a recent analysis of 74 million diffusion events by 1.6 million Twitter users questioned the emphasis on online elites as influencers, concluding that 'word-of-mouth information spreads via many small cascades, mostly triggered by ordinary individuals' (Bakshy et al., 2011: 73).

COMPUTATIONAL APPROACHES

As these studies suggest, the next phase of research on political talk online may benefit from moving beyond ethnographic, content-analytic, or survey-based assessments of these phenomena, instead exploring the value of computational approaches to questions of issue attention, social influence, opinion formation, and political mobilization. The use of 'big data' may provide unique insights into how political elites and their followers deploy particular language around controversial political issues, and how these forms of expression splinter into polarized factions, coalesce into action, and get recirculated through networks.

Using such approaches to help illuminate how particular events or controversies spur different types of political talk and social activism, we explore two news controversies that erupted in 2012 in the United States: (1) Rush Limbaugh's statements about Sandra Fluke, a law student and women's rights activist who rose to national prominence when advocating for a contraception mandate on health care plans; and (2) the shooting of Trayvon Martin, an unarmed African-American teen killed by neighborhood watch volunteer George Zimmerman, who claimed self-defense under Florida's 'stand your ground' law.

The data used to examine these cases began with a purposive sample of politically active users. We identified the most prominent political user accounts in five categories: political advocacy groups; politicians and candidates; political party operatives; journalists and pundits; and political satirists and celebrities. Recency and frequency of activity, size of follower network, ideological diversity, and political prominence were considered when creating lists within each category. From this process, we compiled a final list of 165 political elites, including more than 40 candidates for national office, 40 major advocacy groups, and more than 50 journalists and pundits. We call these the 'top-level' users (see supplemental Online Appendix, available at http://bit.ly/shah-etal-2014).

We then collected follower lists for each of these users and drew a random sample of 80 of their followers, called the 'second-level' users. This resulted in 13,200 user accounts that we tracked. In practice, given the nature of Twitter user accounts, this number shrank slightly because some users' accounts were suspended or deleted. In fact, we began with the assumption that political elites gain and lose followers given their shifting prominence and visibility. Accordingly, we re-collected follower information at three distinct points during the US presidential election cycle: the initial collection in late 2011, mid-June 2012, and September 2012. Follower information was collected using the RESTful Twitter application programming interface (API) (https://dev.twitter.com/docs/api).

Using an account with expanded ('whitelisted') access to Twitter data, we gathered tweets from these users using the Streaming Twitter API (https://dev.twitter.com/docs/streaming-apis).[2] This collection gave us the following information for each user we specified: tweets created by the user, tweets which were retweeted by the user, replies to any tweet created by the user, retweets of any tweet created by the user, and 'manual' replies to the user created without using Twitter's 'Reply' button.

This resulted in a collection of more than 431 million tweets between December 30, 2011, and January 22, 2013, the day after President Obama's second inauguration. Given this sampling methodology, it is inevitable that we would pick up some 'elite' users among our follower

Table 16.1 Follower and following counts at level 1 and 2 of collection

		Level 1	Level 2
Followers			
	Mean	275,483	3,832
	Median	36,324	398
Following			
	Mean	10,551	2,243
	Median	578	689
Ratio			
	Mean	26.1	1.7
	Median	62.8	0.6

sample. For example, our second-level sample included four users with more than 1 million followers: rhythm and blues singer John Legend, Whole Foods, the *Huffington Post*, and National Public Radio (NPR)'s Scott Simon. Still, the vast majority of those included at this level were not political elites.

Table 16.1 shows the means and medians of these two levels, comparing number of follower accounts, number of accounts following, and the ratio of these two numbers. Given the timing of the cases considered in this chapter, this analysis centers on the first wave of data collection. Table 16.1 compares the two levels. Focusing on median values, which avoid inflated means due to outliers, we see that level 1 users (political elites) have nearly 100 times more followers than level 2 users (follower network). Level 1 users also follow fewer users than level 2 users, resulting in a ratio of followers to following that differs dramatically between our elite and follower groups. While it may be true that follower counts do not denote influence, per se (Cha et al., 2010), this certainly differentiates our elite users from those who followed them.

UNDERSTANDING THE CONTROVERSIES

To analyze these data, we turn to computational methods, namely computer-aided content analysis of keywords and hashtags, along with social network mapping of these online tokens. This chapter will proceed as follows: First, we examine the frequency of the appearance of particular keywords. We then identify which level of users are using these keywords. Next, we focus on the co-occurrence of hashtags as an indicator of public sentiment and political organizing, tracking these among elites and their

followers. Finally, we map networks of message retweeting to understand the social structure of everyday political talk. Before moving to this computational analysis, we first provide some context on the two cases, the Sandra Fluke–Rush Limbaugh controversy and the Trayvon Martin–George Zimmerman shooting.

Sandra Fluke: A Washington, DC Conflict

Georgetown law student and women's health activist Sandra Fluke landed in the center of US national controversy – and a Twitter firestorm – in February 2012 shortly after testifying in a hearing staged by Congressional Democrats. An original hearing before the House Oversight and Government Reform Committee focused on the Affordable Care Act ('Obamacare') and provisions mandating coverage of contraceptives. Committee Chairman Darrell Issa (Republican, California) refused to allow Fluke to speak in a three-hour hearing concerning the contraceptive mandate during which only men testified, all opposing this policy.

Led by former House Speaker Nancy Pelosi (Democrat, California), Congressional Democrats convened an unofficial hearing February 23, 2012, to take Fluke's testimony. She spoke about the high cost of contraceptives and adverse effects on women's health, drawing praise from Democrats. Her appearance at the unofficial hearing drew derision shortly afterward, beginning with attacks from the conservative blogosphere, such as, 'Sex-Crazed Co-Eds Going Broke Buying Birth Control, Student Tells Pelosi Hearing Touting Freebie Mandate', from *CNS News*. The issue bubbled in those circles for five days before coming to a boil with conservative talk show radio host Rush Limbaugh. On his February 29, 2012, show, Limbaugh called Fluke a 'slut' and a 'prostitute', saying she wanted to be paid to have sex.

Swift and critical response to the statement came immediately from the left, and built over days to statements from Republicans calling the remarks inappropriate, including House Speaker John Boehner (Republican, Ohio) and National Republican Senatorial Committee Vice-Chair Carly Fiorina. Limbaugh was initially unmoved, saying on his March 1 show, 'If we are going to pay for your contraceptives, thus pay for you to have sex, we want something for it, and I'll tell you what it is: We want you to post the videos online so we can all watch'.

His intransigence proved costly. Responding to online outrage directed at the program's sponsors, advertisers began pulling their spots from the show on March 2, with Sleep Train Mattress Centers announcing its decision on Twitter: 'We don't condone negative comments directed toward any group. In response, we are currently pulling our ads from Rush with

Rush Limbaugh'. As online outrage and pressure on advertisers grew, Limbaugh apologized on his March 3 show: 'My choice of words was not the best, and in the attempt to be humorous, I created a national stir'; though some found the *mea culpa* lacking. The apology and a subsequent clarification did nothing to stem the sponsor losses, with some outlets reporting nearly 100 advertisers backing out of Limbaugh's show, an effect that stretched into 2014, more than two years later.[3] Pressure on advertisers was a clear trend on Twitter, as users called on Sleep Train, Netflix and others to pull their spots from the show.

Trayvon Martin: A Grassroots Firestorm

The Florida shooting of unarmed African-American teen Trayvon Martin in February 2012 slowly built into a national conversation on race and the US justice system, discourses that resurfaced to frame the August 2014 killing of Michael Brown in Ferguson, Missouri. Martin, 17, was shot and killed by George Zimmerman, a neighborhood watch volunteer, in Sanford, Florida, the night of February 26. Zimmerman was taken into custody immediately following the shooting but was released after claiming he acted in self-defense. Florida is a so-called 'stand your ground' state, providing legal protection for individuals who use deadly force to defend themselves when they feel their lives are in danger outside of their homes.

About ten days after Zimmerman's release, Martin's parents posted a petition to Change.org, urging Zimmerman's arrest, and the first mainstream news coverage was published the following day, when the parents filed suit to get records in the case. National media attention followed on March 13 to 15. A week later, the Sanford police chief stepped down from the investigation and a special prosecutor was assigned to the case. Scrutiny of the case ranged from the actions of the police and prosecutors to the media's inattention to the story. The controversy was charged both racially and politically, showing considerable polarization. In-person protests and demonstrations accompanied intense social media attention to the case in the early weeks.

Attention quickly narrowed to Martin's attire and questions of whether youths wearing 'hoodie' sweatshirts are inherently menacing. On March 23, speaking on *Fox and Friends*, Geraldo Rivera said, 'I think the hoodie is as much responsible for Trayvon Martin's death as George Zimmerman was'. Conservative blogs published excerpts from Martin's Twitter feed, including references to marijuana use (Graeff et al., 2014). Shortly after, a hacker broke into Martin's e-mail and social networking accounts and released personal information, which was subsequently covered by mainstream media. At the same time, Zimmerman's legal team set up a

social media presence for their client, just as he earned scrutiny for posts about Mexicans made on his abandoned MySpace account. Nearly a month later, on April 11, 2012, the prosecutor announced second-degree murder charges against Zimmerman.

EXPLORING POLITICAL TALK ONLINE

Keyword Volume

To analyze the relative volume of keywords on Twitter, we graphed their concentration relative to all other content. Each line graph represents the proportion of the tweets within that level mentioning a keyword or hashtag. This allowed some insight into the intensity of political talk on these topics at these different levels and its flow between elites and their followers.

As can be seen in Figure 16.1, our analysis of the top-trending keywords on Twitter concerning the Sandra Fluke–Rush Limbaugh controversy

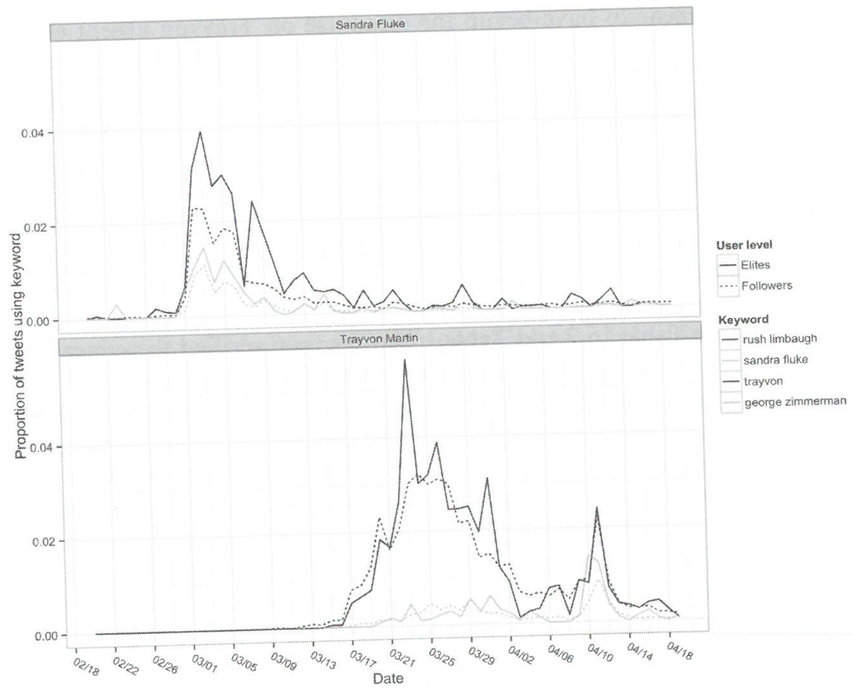

Figure 16.1 *Proportional volume of keyword use for Sandra Fluke and Trayvon Martin cases*

indicates a political conversation dominated by elites. The level 1 pool of candidates, party operatives, advocacy groups, media and pundits had a far greater within-level proportion of tweets, as volume relative to all other content, with the keywords 'Sandra Fluke' or 'Rush Limbaugh' than the randomly selected sample of their followers at level 2. This is particularly true for the keyword 'Rush Limbaugh', where elite attention appears to spur considerable attention among their followers.

In contrast, the Trayvon Martin shooting was initially discussed among the follower networks with somewhat more intensity than at the elite level. When the story initially gained attention in national news outlets between March 13 and 15, 2012, Twitter activity soon followed, with use of the keywords 'Trayvon' and 'George Zimmerman' proportionally higher for followers relative to the political and media elites. This was true through its first major spike of activity on March 20. Starting the next day and continuing through its peak on March 23, the date of both the special prosecutor's appointment and President Obama's public statement on the case, a greater proportion of elite posts mentioned Trayvon, suggesting a shift toward indexing. In contrast to the Sandra Fluke case, followers in aggregate appear to have begun the discourse on Trayvon absent elites, before major media outlets had started to pay attention to the issue.

Hashtag Clusters

Of course, keywords are only the starting point for any tracking over time of language use among elites and followers. To understand the specific uses and associations of different hashtags, we tracked the most prominent hashtags used to identify each case, and then conducted a principal component analysis of the use of particular hashtags by monitored accounts (see Table 16.2 and Table 16.3). We then verified these meaning clusters by comparing hashtag use relative to the other content of the tweet. That is, human coders verified whether these hashtags are used in consistent fashion in a random sample of actual tweets, and whether this use was consistent with the interpretation. This combination of scale building and content verification allowed us to capture the associations among hashtags quantitatively and qualitatively.

For the Fluke–Limbaugh case, we picked the hashtags that reflected support for Sandra Fluke and Rush Limbaugh, as well as the hashtags that represented calls to action. Principal component analysis in Table 16.2 suggests that there are two hashtag factors: a broad set discussing and supporting actors on both sides of the controversy (for example, #standwithsandra, #sandrafluke, #istandwithrush, #standwithrush, and #slutgate) and a distinct set calling for activism against Limbaugh (for example,

Table 16.2 Principal component analysis of hashtag use in Sandra Fluke case

	Factor 1	Factor 2	Communality
#boycottrush	0.87		0.76
#stoprush	0.87		0.76
#flushrush	0.78		0.62
#standwithrush		0.80	0.65
#standwithsandra	0.40	0.58	0.50
#istandwithrush		0.43	0.19
#sandrafluke		0.35	0.13
#slutgate		0.28	0.14

Note: Standardized loadings over 0.25 of varimax rotation.

Table 16.3 Principal component analysis of hashtag use in Trayvon Martin case

	Factor 1	Factor 2	Communality
#trayvon	0.92		0.89
#trayvonmartin	0.88		0.81
#zimmerman	0.60		0.37
#georgezimmerman	0.53		0.30
#iamtrayvon		0.82	0.68
#millionhoodiemarch		0.70	0.52

Note: Standardized loadings over 0.25 of varimax rotation.

#stoprush, #boycottrush, and #flushrush). Notably, the activism factor is more pronounced, with three clear loadings. The second factor, focusing on the actors and expressions of support for them, is more mixed, with relatively low communality estimates for a number of the hashtags, suggesting low fragmentation and considerable contestation within this cluster.

Human coders checked hashtag contexts by looking at 1,023 randomly selected tweets related to Sandra Fluke, a subset of which contained the relevant hashtags. The results suggest that tweets using 'activism' hashtags are overwhelmingly opposed to Limbaugh, and 37 percent of those make explicit calls for boycotting Limbaugh's radio program. In addition, most of the tweets that use #istandwithrush, #standwithrush, and #slutgate hashtags express sentiment positive toward Limbaugh or critical of Fluke, whereas 82 percent of tweets using #standwithsandra are supportive of Fluke. Based on this analysis and the principal component analysis, we distinguished the pro-Fluke, represented by #standwithsandra and #sandrafluke, from the pro-Limbaugh, represented by #istandwithrush, #standwithrush, and #slutgate, with these two distinct from the call to action against Rush represented by #stoprush, #boycottrush, and #flushrush.

A parallel process was used for the Trayvon Martin case. We began by selecting a number of hashtags that expressed key ideas, such as a dominant set possibly reflecting polarization and another set possibly expressing the idea of solidarity. Principal component analysis in Table 16.3 reveals that the hashtags can be grouped in two factors: hashtags debating the actors at the center of the controversy, Martin and Zimmerman (for example, #trayvon, #trayvonmartin, #zimmerman, and #georgezimmerman) and hashtags related to solidarity expression (for example, #iamtrayvon and #millionhoodiemarch). Again, the relatively low commonality estimates for the George Zimmerman hashtags within the actor factor suggest a more complex picture.

To confirm this classification, human coders checked hashtag contexts by looking at 1980 randomly selected tweets related to Trayvon Martin, a subset of which contained the relevant hashtags. Each hashtag was coded for whether it was an expression of support for Martin, opposition to Martin, or a call to action. Results suggest that 84 percent of tweets that use #trayvon and #trayvonmartin show overall support for Martin. For the hashtags #zimmerman and #georgezimmerman, 51 percent of tweets support Trayvon Martin, and 40 percent of them support Zimmerman, reflecting greater ambivalence. As we expected, a majority of the tweets that use solidarity hashtags support Martin over Zimmerman, with 24 percent of tweets that use #millionhoodiemarch and #iamtrayvon explicitly expressing calls to action. Since we can see clear distinctions in the use of hashtags to solicit support for Martin and Zimmerman, we grouped #trayvon and #trayvonmartin as pro-Trayvon and #zimmerman and #georgezimmerman as an alternate grouping, given that it was more mixed. We also built a solidarity grouping using #millionhoodiemarch and #iamtrayvon, a type of call to action.

Expression Intensity

Returning to the case of Sandra Fluke, when the volume of top relevant tweets is separated into the expression of particular hashtags, the elite group generates a greater proportion of tweets related to the controversy on all keywords and hashtags except the activism grouping of #stoprush, #boycottrush, and #flushrush (see Figure 16.2). These hashtags emerged March 2, the day after Limbaugh doubled down on his Fluke comments

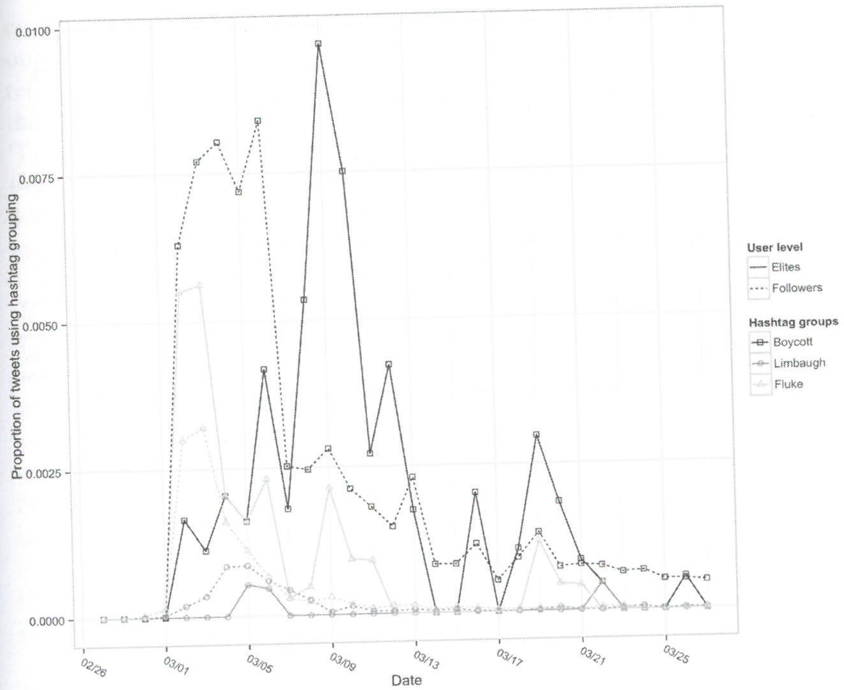

Figure 16.2 Proportional volume of hashtag clusters for Sandra Fluke case

and Sleep Train announced its withdrawal of advertisements. This call to action trailed off more quickly among the followers than the elites, who discussed the developing story, carrying the theme in about 1 percent of all tweets well into the next week. Proportionally, the elite group tweeted more about the case, with spikes in their activity marked by spikes among followers shortly afterward in most other ways.

These patterns may point to a number of possible influences. The Fluke case was a particularly Washington, DC-centric affair, beginning with Congressional hearings and quickly emerging as a flashpoint in the presidential campaign. The Washington, DC insider nature of the elite panel makes the group more likely to focus a greater proportion of Twitter activity on issues in their own arena. We also suspect Limbaugh's national profile may have led to quick and expansive focus among level 1 elites, perhaps indicated by the far greater use of Limbaugh's name than Fluke's (nearly 4.5 percent of tweets to 1.5 percent of tweets at the peak on March 3, 2012). Conversation among elites may also have been influenced by efforts from the left to highlight the case as an example of the 'war on

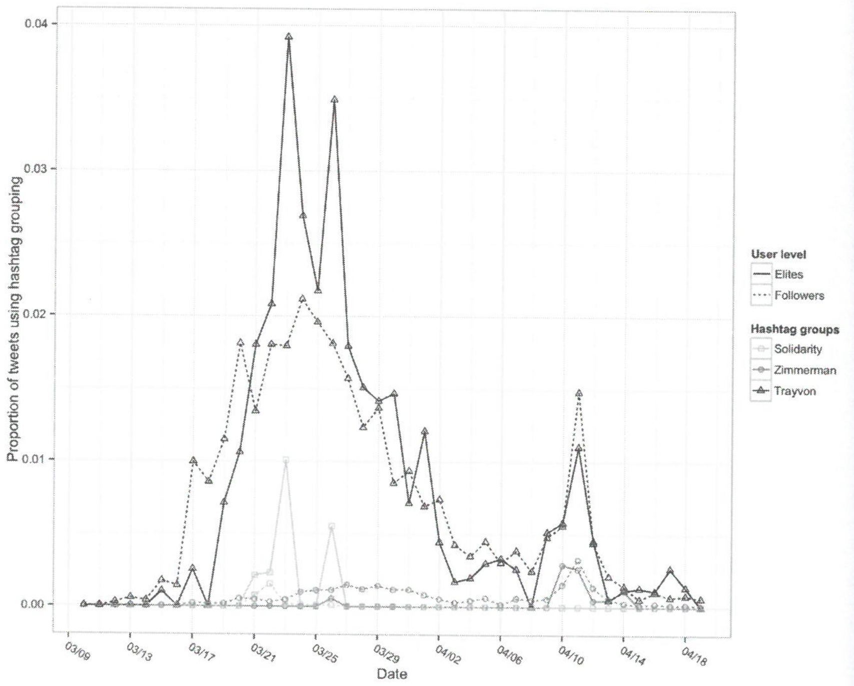

Figure 16.3 Proportional volume of hashtag clusters for Trayvon Martin case

women' and fundraise on the point. In contrast, it is notable that the followers were ahead of the elites in pointing toward a boycott, suggesting grassroots leadership on this matter.

Turning to the Trayvon Martin case, parallel analysis finds that among elites and their followers, #trayvon and #trayvonmartin were the most frequently used hashtags, followed by #georgezimmerman and #zimmerman (see Figure 16.3). In addition, a set of call-to-action hashtags appeared, including #iamtrayvon and #millionhoodiemarch, but was used far less frequently (less than 1 percent at its peak). It is surprising, then, that reporters reviewed the importance and effects of Internet activism through social media tools for this case (Graeff et al., 2014).

As an example of everyday political talk, the timing and volume of activity in the Martin case differed from what we found with Limbaugh and Fluke. Where elites quickly and clearly led and dominated the conversation surrounding Fluke, the Martin case shows non-elite followers picking up the story earlier and discussing it with more intensity before elites reacted and came to lead the conversation. Again, without a more detailed

content analysis and message tracking – issues we return to below – we can only speculate on the reasons for this. It seems likely that the difference reflects early grassroots efforts to urge media and political actors to notice the case and direct attention to it. Followers began tweeting the keyword 'Trayvon' by March 8 and began circulating the hashtags #trayvon and #trayvonmartin by March 11, in the days following Martin's parents posting their petition on Change.org. The elite group first tweeted these hashtags on March 15, after national media coverage had begun.

However, once the elite group began discussing the case on Twitter, those users paid far more attention proportionally than the follower networks, coming close to double the percentage of tweets with #trayvon and #trayvonmartin on the March 23 spike. Although the elites spiked higher in tweets related to the case, their attention dropped off more quickly. The non-elite followers displayed more sustained conversation about the case, with drops that were less precipitous than those observed among the elites. Notably, the call to action implicit in the #iamtrayvon and #millionhoodiemarch hashtags were proportionally a very small part of the online conversation among elites and even less so among followers, indicating that this grassroots effort was somewhat smaller than the one behind the Rush Limbaugh boycott.

Network Mapping

For each of the topics, we generated a retweet network based upon tweets that included the major keywords and hashtags. Retweets are tweets in which users share content produced by another user, a near perfect reflection of Tarde's 'invention' and 'imitation'. Visualizations of the retweet network in both cases show clear distinctions between core groups discussing the issues. The graph depicts users and their centrality within the retweet network, with each node (bubble) in the network representing a user; the size of the node denotes how much that person retweets and is retweeted within the network; and the edges denoting a retweet. These visualizations represent the largest connected components of activity for each case. Users outside these components are not represented because of minimal interaction with this core group.[4]

Upon visual inspection of the Fluke network, this case shows two core clusters speaking almost entirely separately from each other (see Figure 16.4). One cluster is focused heavily on the boycott-focused hashtags of #stoprush, #boycottrush and #flushrush, represented in black. Retweets containing these hashtags clearly dominate that core, showing a group of users heavily focused on calls to action. This cluster has fewer instances of the grouping of #standwithsandra and #sandrafluke (denoted

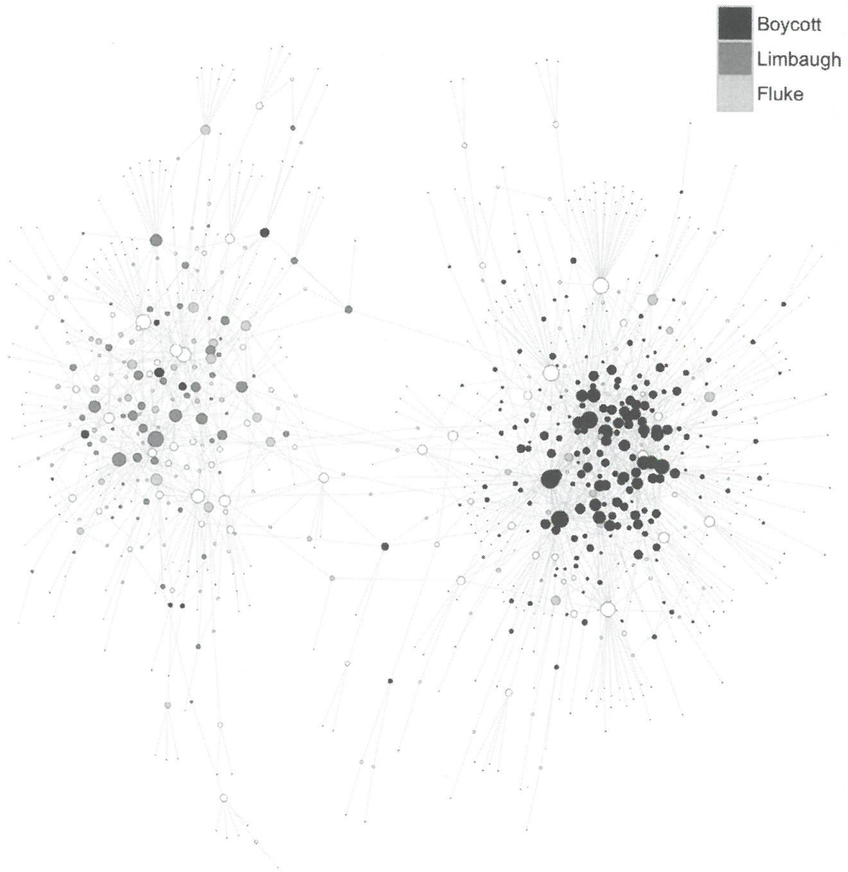

Figure 16.4 Retweet network for Sandra Fluke case with major hashtags

with light gray), as well as activity with keywords without hashtag (white). The calls to boycott are denser at the center of the network than the non-activist tweets, indicating activity focused more squarely on a set of central users. The activity among the nodes representing discussion of the actors in the event appears more disperse and less hierarchical.

The contrasting core of users tweeting about the Fluke case has a retweet network far less focused on political action. The most dominant hashtag use is the cluster for #standwithrush, #istandwithrush and #slutgate (dark gray). This network cluster also has relatively heavy use of the hashtag grouping of #standwithsandra and #sandrafluke, also apparent from our factor analysis, suggesting a spirited discussion about the

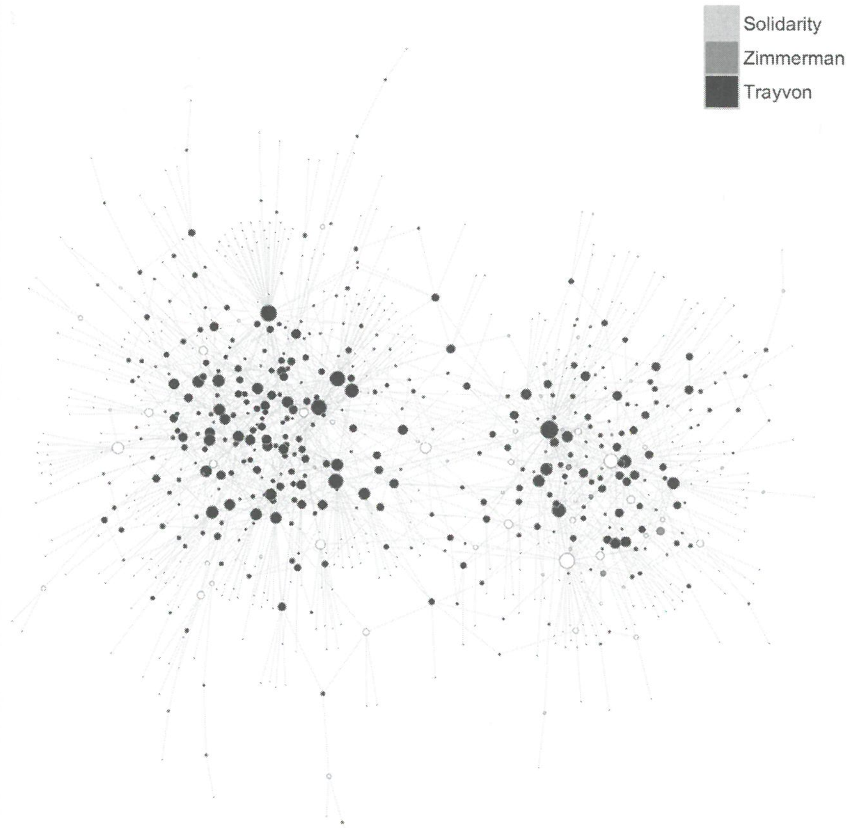

Figure 16.5 Retweet network for Trayvon Martin case with major hashtags

various players. This cluster was considerably less dense and active than the boycott-dominated core.

The network in the Trayvon Martin case also split into two distinct clusters (see Figure 16.5). Both were far less activist in their hashtag use and focused almost entirely around the use of #trayvon and #trayvonmartin (black), with contending characterizations of the case and its players. Each node represents a user who used this hashtag or retweeted a message containing the hashtag. The grouping for #zimmerman and #georgezimmerman (dark gray) appears in one cluster but with nowhere near the dominance of the Trayvon-focused grouping. The solidarity-focused grouping of #iamtrayvon and #millionhoodiemarch (light gray) is virtually invisible in the network, showing almost no retweet activity among the hashtags.

The two clusters, both relying on the centrality of the Trayvon hashtags, have substantially more between-cluster interaction than we found in the Fluke case. Hashtag use within the network does not focus on calls to action though further analysis may reveal these in tweet content, rather than hashtag use. Overall, the cluster with the Zimmerman-focused grouping (dark gray) is less dense than the cluster containing the Trayvon-focused grouping.

As observed with the use of keywords and the clustering of hashtags, these two cases functioned quite differently, despite the fact that both addressed controversies regarding inequalities. Both apparently featured social influence by elites and followers – suggesting both a two-step flow and a two-way flow of communication – though who appears to exercise influence differs by topic and stages in each controversy's evolution, especially as it related to calls to action. The visualizations of these networks confirm these differences, and suggest that measuring message propagation, information diffusion, and interpersonal influence across these fluid levels is the future of research on everyday political talk.

LOOKING FORWARD

Rather than reflecting on the specifics of these cases, which were mainly deployed to illustrate the most basic potential of computational approaches to understanding online political talk, we close by using them to offer some broader suggestions about the future of online politics research and the implications of big data for this coming age of computational social science. In doing so, we look to the past, returning to the work of Gabriel Tarde that began this chapter. The digital traces contained in a single tweet provide a myriad of directions for research on interpersonal interaction and social influence linked directly to Tarde's main theoretical contributions. Yet the contemporary environment also complicates this linear formulation of elite influence through the press on to the constituted public, setting the conversational agenda and thereby distilling opinion into action. Rather, as we have observed here, within the feedback loops provided by the new information environment and social networking technologies, elite agendas appear to be influenced by their followers under certain circumstances, substantially complicating the framework Tarde proposed.

It may be that the most useful insights from Tarde are not derived from his early formulation of two-step flow or communication mediation, but

instead from his assertion about the types of data that should be collected as part of a 'science of the social'. His approach, its full potential unrealized at the time, refuses the schism between individual and society as we have come to understand them after Durkheim. Moreover, technology and methodology has caught up to Tarde's theory, for as Latour writes (2010: 160):

> What we are witnessing, thanks to the digital medium, is a fabulous extension of [Tarde's] principle of traceability. It has been put in motion not only for scientific statements, but also for opinions, rumors, political disputes, individual acts of buying and bidding, social affiliations . . . What has previously been possible for only scientific activity – that we could have our cake (the aggregates) and eat it too (the individual contributors) – is now possible for most events leaving digital traces, archived in digital databanks.

As this suggests, computational approaches that stress syntactical coding such as natural language processing may prove critical to understanding the nature of social influence and interpersonal dynamics in everyday political talk (Bird et al., 2009; Jurafsky and Martin, 2009; Shah et al., 2002; Han et al., 2011).[5] In addition, the growth of supervised and unsupervised machine learning tools focusing on the statistical co-occurrence of words or phrases may also prove powerful for managing large volumes of political messages (Blei and Lafferty, 2009; Hopkins and King, 2010). Such language processing allows for the coding of vast amounts of communications with considerable subtlety and qualification.

Complementing these language-processing approaches are tools for tracking their movement through social networks. Efforts to examine the role of audience size, pass-along value, and conversational ability on social influence now have ways of examining these questions at a large scale (Cha et al., 2010) and can do so alongside measures of whether the message is deemed interesting or elicits positive feelings (Bakshy et al., 2011), blending the structural and the psychological. Indeed, social networking sites like Twitter and Facebook provide 'a setting where many different kinds of information spread in a shared environment' (Romero et al., 2011: 695), permitting new insights regarding invention and expression of ideas and their imitation and diffusion with communication networks. These methods will provide new vistas onto political talk as it increasingly occurs through online channels, calling into question the accepted wisdom about message flows and offering new accounts of emergent forms of civic and political expression.

NOTES

1. As Terry Clark writes in the introduction of *Gabriel Tarde On Communication and Social Influence: Selected Papers* (1969: 25): 'Durkheim refused to accept that sociological principles should be grounded in psychology. Sociology as a distinctive science, he held, must take as its object of study social facts; and these social facts must find their causes as well as their consequences in other distinctly social facts'.
2. A brief tutorial of how to collect Twitter data using a simple Python script can be found at http://badhessian.org/2012/10/collecting-real-time-twitter-data-with-the-streaming-api/.
3. Online activists have maintained detailed databases of Rush Limbaugh sponsors, tracking continued supporters (who remain targets of a boycott), and noting which brands have pulled advertisements. At the time of writing in Fall 2014, these campaigners claim more than 2,700 local and national advertisers have withdrawn support (http://stoprush.net/rush_limbaugh_sponsor_list.php#current_a).
4. The Fluke network overall comprises 13,365 users, with this largest component including 870. For Trayvon, the overall network comprises 13,365, with 856 users in the largest component. The vast majority of users in both are 'isolates', nodes not tied with any other users.
5. Two main lines of modeling that have been developed do not rely so heavily on word counts and the manual curation of specific words: language modeling and statistical modeling (Monroe and Schrodt, 2008). Language modeling attempts to leverage as much information as it can out of a single document; it attempts to identify parts of speech in a given document and allows us to see the who, what, when, where, and how of a message.

REFERENCES

Bakshy, E., Hofman, J.M., Mason, W.A. and Watts, D.J. (2011). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, February (pp. 65–74). ACM.

Barber, B. (1984). *Strong Democracy: Participatory Politics for a New Age*. Berkeley, CA: University of California Press.

Bennett, W.L., Wells, C. and Freelon, D. (2011). Communicating civic engagement: contrasting models of citizenship in the youth web sphere. *Journal of Communication*, 61, 835–856.

Berelson, B., Lazarsfeld, P.F. and McPhee, W.N. (1954). *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago, IL: University of Chicago Press.

Bird, S., Klein, E. and Loper, E. (2009). *Natural Language Processing with Python*. Sebastopol, CA: O'Reilly Media.

Blei, D.M. and Lafferty, J.D. (2009). Topic models. *Text Mining: Classification, Clustering, and Applications*, 10, 71.

Bode, L., Vraga, E K., Borah, P. and Shah, D.V. (2014). A new space for political behavior: political social networking and its democratic consequences. *Journal of Computer-Mediated Communication*, 19(3), 414–429.

Boyd, D.M. and Ellison, N.B. (2007). Social network sites: definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13, 210–230.

Boyd, D., Golder, S. and Lotan, G. (2010). Tweet, tweet, retweet: conversational aspects of retweeting on twitter. In *2010 43rd Hawaii International Conference on System Sciences (HICSS)* (pp. 1–10). January, IEEE.

Bruns, A. and Burgess, J.E. (2011). The use of Twitter hashtags in the formation of ad hoc publics. Paper presented at the 6th European Consortium for Political Research General Conference, August 25–27, University of Iceland, Reykjavik.

Bruns, A. and Highfield, T. (2013). Political networks on twitter: tweeting the Queensland state election. *Information, Communication and Society*, 16(5), 667–691.

Cha, M., Haddadi, H., Benevenuto, F. and Gummadi, P.K. (2010). Measuring user influence in Twitter: the million follower fallacy. *ICWSM*, 10, 10–17.

Cho, J. Shah, D.V., McLeod, J.M., McLeod, D.M., Scholl, R.M. and Gotlieb, M.R. (2009). Campaigns, reflection, and deliberation: advancing an O-S-R-O-R model of communication effects. *Communication Theory*, 19, 66–88.

Clark, Terry (ed.) (1969). *Gabriel Tarde On Communication and Social Influence: Selected Papers*. Chicago, IL: University of Chicago Press.

Dahlgren, P. (2000). The Internet and the democratization of civic culture. *Political Communication*, 17, 335–340.

Dahlgren, P. (2005). The Internet, public spheres, and political communication: dispersion and deliberation. *Political Communication*, 22(2), 147–162.

Dylko, I.B., Beam, M.A., Landreville, K.D. and Geidner, N. (2012). Filtering 2008 US presidential election news on YouTube by elites and nonelites: an examination of the democratizing potential of the internet. *New Media and Society*, 14(5), 832–849.

Ekström, M. and Östman, J. (2013). Information, interaction, and creative production: the effects of three forms of Internet use on youth democratic engagement. *Communication Research*, online pre-publication, 0093650213476295, 1–22.

Eveland, W.P. and Hively, M.H. (2009). Political discussion frequency, network size, and 'heterogeneity' of discussion as predictors of political knowledge and participation. *Journal of Communication*, 59(2), 205–224.

Freelon, D.G. (2010). Analyzing online political discussion using three models of democratic communication. *New Media and Society*, 12(7), 1172–1190.

Gainous, J. and Wagner, K.M. (2014). *Tweeting to Power: The Social Media Revolution in American Politics*. New York: Oxford University Press.

Gamson, W.A. (1992). *Talking Politics*. New York: Cambridge University Press.

Gil de Zúñiga, H.G., Jung, N. and Valenzuela, S. (2012). Social media use for news and individuals' social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication*, 17(3), 319–336.

Gil de Zúñiga, H.G., Puig-I-Abril, E. and Rojas, H. (2009). Weblogs, traditional sources online and political participation: an assessment of how the internet is changing the political environment. *New Media and Society*, 11(4), 553–574.

Golbeck, J., Grimes, J. and Rogers, A. (2010). Twitter use by the UC Congress. *Journal of the American Society for Information Science and Technology*, 61(8), 1612.

Graeff, E., Stempeck, M. and Zuckerman, E. (2014). The battle for 'Trayvon Martin': Mapping a media controversy online and off-line. *First Monday*, 19(2). http://firstmonday.org/ojs/index.php/fm/article/view/4947.

Graham, T. and Hajru, A. (2011). Reality TV as a trigger of everyday political talk in the net-based public sphere. *European Journal of Communication*, 26(1), 18–32.

Gueorguieva, V. (2008). Voters, MySpace, and YouTube the impact of alternative communication channels on the 2006 election cycle and beyond. *Social Science Computer Review*, 26(3), 288–300.

Habermas, J. (1984). *The Theory of Communicative Action: Vol. 1. Reason and the Rationalization of Society*. McCarthy, T. (trans.). Boston, MA: Beacon.

Han, J.Y., Shah, D.V., Kim, E., Namkoong, K., Lee, S.Y., Moon, T.J., Cleland, R., McTavish, F.M. and Gustafson, D.H. (2011). Empathic exchanges in online cancer support groups: distinguishing message expression and reception effects. *Health Communication*, 26(2), 185–197.

Hanna, A., Wells, C., Maurer, P., Friedland, L., Shah, D. and Matthes, J. (2013, October). Partisan alignments and political polarization online: a computational approach to understanding the French and US presidential elections. In *Proceedings of the 2nd Workshop on Politics, Elections and Data* (pp. 15–22). ACM.

Hardy, B.W. and Scheufele, D.A. (2005). Examining differential gains from Internet use: comparing the moderating role of talk and online interactions. *Journal of Communication*, 55(1), 71–84.

Hopkins, D.J. and King, G. (2010). A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, 54(1), 229–247.

Huckfeldt, R. and Sprague, J. (1995). *Citizens, Politics, and Social Communication: Information and Influence in an Election Campaign*. New York: Cambridge University Press.
Jungherr, A., Jürgens, P. and Schoen, H. (2012). Why the Pirate Party won the German election of 2009 or the trouble with predictions: a response to Tumasjan, A., Sprenger, T.O., Sander, P.G. and Welpe, I.M. 'Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment'. *Social Science Computer Review*, 30(2), 229–234.
Jurafsky, D. and Martin, J.H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition and Computational Linguistics*, 2nd edn. Upper Saddle River, NJ: Prentice Hall.
Katz, E. (2006). Rediscovering Gabriel Tarde. *Political Communication*, 23(3), 263–270.
Katz, E. and Lazarsfeld, P. (1955). *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Glencoe, IL: Free Press.
Kim, J., Wyatt, R.O. and Katz, E. (1999). News, talk, opinion, participation: the part played by conversation in deliberative democracy. *Political Communication*, 16, 361–385.
Kinnunen, J. (1996). Gabriel Tarde as a founding father of innovation diffusion research. *Acta Sociologica*, 39(4), 431–442.
Kwak, N., Williams, A., Wang, X. and Lee, H. (2005). Talking politics and engaging politics: an examination of the interactive relationships between structural features of political talk and discussion engagement. *Communication Research*, 32, 87–111.
Latour, B. (2010). Tarde's idea of quantification. In Candea, M. (ed.), *The Social After Gabriel Tarde: Debates and Assessments*. London: Routledge.
Lazarsfeld, P.F., Berelson, B. and Gaudet, H. (1944). *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. New York: Duell, Sloan & Pearce.
Lazer, D., Pentland, A.S., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., . . . and Van Alstyne, M. (2009). Life in the network: the coming age of computational social science. *Science*, 323(5915), 721.
Lee, N.J., Shah, D.V. and McLeod, J.M. (2013). Processes of political socialization a communication mediation approach to youth civic engagement. *Communication Research*, 40(5), 669–697.
McLeod, J.M., Daily, K., Guo, Z., Eveland, W.P., Bayer, J., Yang, S. and Wang, H. (1996). Community integration, local media use, and democratic processes. *Communication Research*, 23(2), 179–209.
McLeod, J.M., Scheufele, D.A. and Moy, P. (1999). Community, communication, and participation: the role of mass media and interpersonal discussion in local political participation. *Political Communication*, 16, 315–336.
Monroe, B.L. and Schrodt, P.A. (2008). Introduction to the special issue: the statistical analysis of political text. *Political Analysis*, 16(4), 351–355.
Mutz, D.C. (2006). *Hearing the Other Side: Deliberative Versus Participatory Democracy*. New York: Cambridge University Press.
Namkoong, K., Shah, D.V., Han, J.Y., Kim, S.C., Yoo, W., Fan, D., . . . and Gustafson, D.H. (2010). Expression and reception of treatment information in breast cancer support groups: how health self-efficacy moderates effects on emotional well-being. *Patient Education and Counseling*, 81, S41–S47.
Papacharissi, Z. (2004). Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media and Society*, 6(2), 259–283.
Papacharissi, Z. (ed.). (2010). *A Networked Self: Identity, Community, and Culture on Social Network Sites*. New York: Routledge.
Pasek, J., More, E. and Romer, D. (2009). Realizing the social Internet? Online social networking meets offline civic engagement. *Journal of Information Technology and Politics*, 6(3–4), 197–215.
Pingree, R.J. (2007). How messages affect their senders: a more general model of message effects and implications for deliberation. *Communication Theory*, 17(4), 439–461.
Price, V. and Cappella, J.N. (2002). Online deliberation and its influence: the electronic dialogue project in campaign 2000. *IT and Society*, 1(1), 303–329.

Rogers, E.M. and Shoemaker, F.F. (1971). *Communication of Innovations: A Cross-Cultural Approach*. New York: Free Press.
Romero, D.M., Meeder, B. and Kleinberg, J. (2011, March). Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter. In *Proceedings of the 20th International Conference on World Wide Web* (pp. 695–704). ACM.
Schudson, M. (1978). The ideal of conversation in the study of mass media. *Communication Research*, 5(3), 320–329.
Segerberg, A. and Bennett, W.L. (2011). Social media and the organization of collective action: using Twitter to explore the ecologies of two climate change protests. *Communication Review*, 14(3), 197–215.
Shah, D.V., Cho, J., Eveland, W.P. and Kwak, N. (2005). Information and expression in a digital age: modeling Internet effects on civic participation. *Communication Research*, 32, 531–565.
Shah, D.V., Cho, J., Nah, S., Gotlieb, M.R., Hwang, H., Lee, N., Scholl, R.M. and McLeod, D.M. (2007). Campaign ads, online messaging, and participation: extending the communication mediation model. *Journal of Communication*, 57, 676–703.
Shah, D.V. and Scheufele, D.A. (2006). Explicating opinion leadership: nonpolitical dispositions, information consumption, and civic participation. *Political Communication*, 23(1), 1–22.
Shah, D.V., Watts, M.D., Domke, D. and Fan, D.P. (2002). News framing and cueing of issue regimes: explaining Clinton's public approval in spite of scandal. *Public Opinion Quarterly*, 66(3), 339–370.
Sotirovic, M. and McLeod, J.M. (2001). Values, communication behavior, and political participation. *Political Communication*, 18, 273–300.
Sotirovic, M. and McLeod, J.M. (2004). Knowledge as understanding: the information processing approach to political learning. In Kaid, L. (ed.), *Handbook of Political Communication Research*. Mahwah, NJ: Lawrence Erlbaum Associates.
Tarde, G. (1969 [1898]). *Gabriel Tarde On Communication and Social Influence: Selected Papers*, Vol. 334. Chicago, IL: University of Chicago Press.
Tarde, G. (1903 [1890]). *The Laws of Imitation*, trans. E.C. Parsons. New York: Henry, Holt.
Thackeray, R. and Hunter, M. (2010). Empowering youth: use of technology in advocacy to affect social change. *Journal of Computer-Mediated Communication*, 15(4), 575–591.
Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M. (2010). Predicting elections with Twitter: what 140 characters reveal about political sentiment. *ICWSM*, 10, 178–185.
Valenzuela, S., Kim, Y. and de Zúñiga, H.G. (2012). Social networks that matter: exploring the role of political discussion for online political participation. *International Journal of Public Opinion Research*, 24(2), 163–184.
Vitak, J., Zube, P., Smock, A., Carr, C., Ellison, N. and Lampe, C. (2011). It's complicated: Facebook users' political participation in the 2008 election. *Cyberpsychology, Behavior and Social Networking*, 14(3), 107–114.
Walsh, K.C. (2012). Putting inequality in its place: rural consciousness and the power of perspective. *American Political Science Review*, 106(03), 517–532.
Wimsatt, W.C. (2007). *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.
Ye, S. and Wu, S.F. (2010). Measuring message propagation and social influence on Twitter.com. In IEEE (ed.), *Social Informatics*. Berlin and Heidelberg: Springer.

23. Visibility and visualities: 'ways of seeing' politics in the digital media environment

Katy Parry

On contemplating how to approach writing a chapter on visual politics online, my initial thoughts turned to the potential slipperiness of each of these terms. How to think about politics online as distinct from its offline manifestations? How narrowly or broadly to define politics and the political? In terms of the recognized political actors and institutions of official politics and policy-making, or more broadly, as a public space in which meanings, identities and values are contested? How productive is it to separate the visual dimension from other qualities or modalities (text, sound) across a range of digital media forms? Indeed 'the visual' is about more than images alone, so how to place notions of the visual within wider concerns of visibility and visuality?

Such questions around indistinct boundaries and instability form the basis for this chapter, then, and in exploring the slipperiness of these terms I address old and new concerns about visibility, vision and visuality in political communication and culture. Drawing on relevant insights from political studies, media and communications and social movement studies, the present chapter places what has come to be known as 'visual culture studies' at the heart of its approach and spirit, for reasons further explained below. It is in the interplay of online and offline political practices, serious and comedic mediations, and their authoritative or subversive purposes, that senses of convergence and collision exist, and through which new opportunities for analysis emerge. But it is also in the productive sharing of resources and tools from across the academic disciplines that we might better scrutinize, understand and appreciate the varied forms of visual politics online.

The chapter is organized into five sections: first, there is a brief outline of how varied disciplinary perspectives have informed the discussion of politics and mediated imagery, tracing the tensions and anxieties associated with emergent media technologies and the concerns over the corrupting influences identified with such media. Second, elaborating on my argument for a central role of 'visual culture studies' in exploring online politics, the subsequent section outlines notions of visibility and visuality. The following two sections are broadly split into politics 'from below' and politics 'from above'. Arguably, the separation of protest from 'official politics' can create another division of research agendas which belies the multiple civic activities and practices each of us chooses to undertake (or not) in our everyday lives. In the hybrid media environment through which many of us experience politics, both established politicians and protesters appear to be adopting more personalized and expressive forms of engagement (Bennett, 2012; Chadwick, 2013). The sections are designed to be illustrative and exploratory, as it is beyond the scope of the current chapter to provide detailed empirical analysis. The final section sets out future research questions and my concluding comments. Before proceeding to the next scene-setting section, I outline two guiding stipulations.

First, as indicated above, studying visual politics online requires openness to a variety of approaches, based on the research questions and methods that most intrigue and provoke us. In his article entitled 'There are no visual media', W.J.T. Mitchell warns against 'visual culture as the "spectacle" wing of cultural studies', noting that its very promise is in its insistence 'on problematizing, theorizing, critiquing and historicizing the visual process as such' (Mitchell, 2005: 264). A broad interdisciplinary interest in varied cultural practices and artefacts does not equate to a flattening out or homogeneity of approach, or a misrecognition of how one medium's affordances are qualitatively different to another's. As Mitchell has argued elsewhere, 'the opening out of a general field of study does not abolish difference, but makes it available for investigation, as opposed to treating it as a barrier that must be policed and never crossed' (Mitchell, 2002: 173).

Second, the fast-evolving role of the Internet in our everyday lives has further disrupted many of the traditional parameters for studying images in mediated contexts. Such disruptions are not entirely new: for example, Elihu Katz (1988) recognized many years ago how the technological theory of 'disintermediation', or cutting out the middleman, could be applicable for media sociology. Social media use and peer-to-peer sharing online merely represent the latest technological and cultural practices embodying a reinvigorated sense of connectivity and direct communication. The promiscuity of digital media images in a 'cut-and-mix' culture (van Zoonen et al., 2010) creates both possibilities and risks for producers, depicted subjects and audiences who view and share such images. The processes of upload and display add to the ephemeral quality, as images circulate and become divorced from original captions or audio, remediated in ever-mutating 'circuits of culture' (du Gay et al., 1997), seemingly unbounded by shared viewing contexts or clearly defined 'imagined communities' (Anderson, 1991). However the often celebratory claims for disintermediation and a sense of less mediation can be misleading and obfuscate the shifts that digital technologies enable; new and old intermediaries may be adapting and altering their role but they remain 'vitally' important (Thumim, 2012). All digital images encountered via the Internet are mediated in some form; the contexts may be increasingly varied, with images constituted as a mix of both amateur and professional in origin, but this multifarious jumble of image-text circulates within a discursive public space of framing practices, semiotic recipes, rhetorical challenges and ironic gestures.

In an inevitably brief review, the next section provides contextual background to the broad fields of study which inform the chapter, setting out the traditional anxieties, valuable perspectives and emergent tensions through which we might incorporate the study of visual culture in understanding politics and digital media.

THE SPECTRE OF THE SPECTACLE: THE HAUNTING ANXIETIES AROUND THE VISUAL IMAGE IN POLITICAL COMMUNICATION

The role of the image in mediated political communication has long provoked suspicion and unease. With its concerns for governance, democracy and citizenship, political studies provides theoretical tools for assessing the health of the polity and public sphere, often bringing a defensive posture to guarding the integrity of politics against debasing forces (Crick, 1968; Flinders, 2012). Although a concern of political philosophers and critical theorists in earlier centuries, fears of a distracted and passive public or citizenry have more recently been associated with the dominant role of television in political campaigning and as the main source for public knowledge. Throughout the twentieth century, as politicians increasingly addressed publics through the mass medium of television, concerns were raised over the diminishing of political life into a spectacle, distorted by a media-driven shift promoting conflict, sensationalism and inauthentic celebrity politicians. In such 'audience democracies' (Manin, 1997) citizen-viewers are characterized as passive, apathetic spectators, monitoring the actions of political leaders but merely reactive to the theatre of political life performed by a set of interchangeable elites, rather than socially engaged actors with a significant decision-making role (see also Edelman, 1988; Meyer and Hinchman, 2002; Postman, 1987; Putnam, 2000). Writers concerned with the role of media in democracy note trends towards a politics evermore shaped by 'media logic' or 'mediatization', accompanied by an overemphasis on stylization, presentation, performance and image (Blumler and Kavanagh, 1999; Mazzoleni and Schulz, 1999; Strömbäck, 2008).

Where some authors see politics tarnished by the blurring lines of information and entertainment that such 'media logic' brings about, with a public service ethos crowded out by consumer-driven 'infotainment' and lifestyle programming (McChesney, 1999; Thussu, 2007), others see an enhancement for democratic life in the popular engagement with politics encouraged by a variety of formats offering a mix of serious and more playful generic recipes (Corner and Pels, 2003; van Zoonen, 2005; Richardson et al., 2012). The shifting interests and concerns reflect a cultural turn across humanities and social sciences, in which other cultural factors and contested sites of meaning and identity are considered alongside political structures, for example, in debating the role of comedy as a catalyst for civic engagement (Baym and Jones, 2012), or the intersections of celebrity culture and political culture (Street, 1997). Alongside this, there is a turn to the emotional or personal aspects of political life, where the private lives and personal-psychological qualities of political leaders are emphasized in discussion of their political performances across the broader media environment (Corner, 2003; Langer, 2011; Stanyer, 2012). In short, many of the discussed trends in political communication and culture – whether on the appeal to the emotions and entertainment, a personalization of politics, or the role of humour and satire – often include an implicit or explicit concern with the visual or symbolic dimension.

The nature of politics as spectacle is often central to these perspectives, despite proponents rarely engaging with the visual or symbolic forms or properties in any detail. The concerns for a healthy and vibrant public sphere are laudable, but such perspectives are in danger of oversimplifying the identified problem. The 'iconophobia' or 'iconoclasm' at the heart of the Western tradition has now attracted critical attention from authors calling for a rethink towards spectatorship and democracy (Finnegan and Kang, 2004; Green, 2010). In summary, I suggest three potential pitfalls. First, there is often a sense of the emotive image in contest with the rational word; that the expressive, symbolic or affective dimensions work to degrade the crucial rationality of the political realm. This negates the interplay of word and image in mediated communications and the constitutive role of both in interpreting our social worlds. Second, an idealized notion of the citizen is contrasted against the deficiencies of the spectator in an unproductive dichotomization which fails to explore how watching and interpreting are also active and creative: 'The spectator also acts, like the pupil or scholar. She observes, selects, compares, interprets' (Rancière, 2011: 13). Third, a traditional emphasis on official politics which narrowly defines political engagement or political actions is in danger of overlooking alternative roles for citizens outside of recognized political structures (see Coleman and Blumler, 2009: 156). This final point brings

the discussion back to the distinctions between the top-down politics of institutions and policy-makers against the 'from-below' character of grassroots politics and dissent.

Where the visual performance of political leaders has been greeted with unease and suspicion (if examined at all), those writing on protest and dissent provide a much more fruitful scholarly engagement with the sense of the visual and the artistic as powerful cultural tools in political action (for a recent collection see McLagan and McKee, 2012). For those struggling to gain attention on the political stage, 'image politics' (DeLuca, 1999) offer a striking way to promote a cause and spark imagination. Operating with a freedom of expression that traditional party politicians are unlikely to risk, the politics of dissent, or contentious politics, can embrace the symbolic and theatrical, and even be cheered for combating the 'pseudo-events' (Boorstin, 1962) of the political elite with heartfelt and humorous image events and political artwork. Choice of imagery and expression can, of course, have repercussions for those hoping to move from 'outsider' to 'insider' status. Fighting the 'image of power with the power of imagery' (Doerr and Teune, 2012) not only identifies you as part of a collective '99%' (to use the Occupy movement's slogan), but, through an adoption of the symbols of contentious politics, can reinforce your status as a heckler rather than an orator; just as notions of 'the political class' can reinforce a sense of an unreachable elite in the distant echelons of high society.

Finally, visual culture studies or 'image studies', with its ancestry in art history, museum studies and media studies, provides the younger sister to the current melange for this chapter. It is through the visually inflected framing of visual culture studies that we can better recognize the rhetorical, aesthetic, expressive and ironic qualities of politically themed imagery. One particular strength of this perspective is that it reflects on its own analytical usefulness, with key theorists sensitive to the limitations of even labelling media texts as visual media (Bal, 2003; Mitchell, 2005). It is in studying the interplay and interdiscursivity of sound, word and image in a variety of media platforms, formats and genres that we can consider visual images as resources for making meaning within a mix of modalities; that is, as resources for producers, viewers or users. Placing an emphasis on the visual here is an attempt to redress the traditional text-based research bias in political communication.

Given the above summary, this chapter aims to 'scope out' the 'ways of seeing' and ways of understanding politics online. The visual metaphors in the preceding sentence are deliberate: conflations of seeing and believing, and seeing and understanding ('Ah, I see') are thought to betray a favouring of vision as the superior sense through which to access the world (see Jay, 1993 and Mirzoeff, 2011 for historical and critical accounts), and yet it is only more recently that social scientists have embraced an iconic turn or 'pictorial turn' within their research agendas (Mitchell, 1994). Visual experience encompasses a great deal more than pictures or imagery and, indeed, includes the written word in graphic form; yet the study of images has been separated from scientific and literary enquiry historically, marked with a rationalist suspicion dating back to Plato's allegory of the cave (see Jay, 1993: 27; Mitchell, 2005: 86). It is not only material cultural artefacts that are of interest here, it is the way such images interact with the 'pictures in our heads' (Lippmann, 1922), the mental images of our mind's eye which also guide how we see the world and how we place ourselves within social spaces and political structures. Our ways of seeing the world are not only about vision (what the eyes observe), but 'visuality' or 'visualities'; 'what is made visible, who sees what, how seeing, knowing and power are interrelated' (Bal, 2003: 19). Notions of control, knowledge and power are crucial here: visuality is also about a political struggle, the 'right to look' and be seen as citizens (Mirzoeff, 2011). It is necessary then to further outline the intersecting notions of vision, visuality and visibility in relation to both mediated political performances and our ways of seeing the political world.

RECONCILING NOTIONS OF VISIBILITY AND VISUALITY IN POLITICAL COMMUNICATIONS

John B. Thompson's work on the visibility of politicians is central to thinking about how visibility and politics link to visuality and vision, and is often cited by those interested in how our politics has become more mediatized, personalized or intimate. Thompson writes of how the development of mass media technologies in the twentieth century, and especially television as the central platform, linked issues of visibility with vision (Thompson, 1995). In the age of mediated visibility, the field of vision is shaped 'by the distinctive properties of communication media, by a range of social and technical considerations (such as camera angles, editing processes and organizational interests and priorities) and by the new types of interaction that these media make possible' (Thompson, 2005: 35–36). Thompson characterizes mediated visibility as a double-edged sword for politicians; a 'source of a new and distinctive kind of *fragility*' (2005: 42) in a more complex and less controllable information environment, and where public–private boundaries are rewritten.

More recently, Daniel Dayan has argued that historically visibility was desired as an enviable right enjoyed by the few; a form of attention-gaining which publics, as spectators, have been denied: 'Being anonymous has become a stigma, and visibility has become a right frequently and sometimes violently claimed; a right that all sorts of people feel entitled to obtain. The exclusive visibility once conferred upon some is perceived by the anonymous as an injustice in need of redress' (Dayan, 2013: 139). Dayan proposes a 'paradigm of visibility', which, as distinct from the media effects paradigm, 'stresses the role of media in coordinating collective attention' (p. 139). This offers a narrative of 'deprivation followed by a conquest' (p. 139) initially acquired by citizens as a form of visibility offered on a conditional basis (as reality TV contestants, for example), or by violent means, such as terrorist acts designed with media in mind. But this is often 'the wrong type of visibility' (p. 141).

Further along in Dayan's narrative, new media technologies and platforms emerge, such as Facebook and YouTube, in which the quest for visibility can be taken a step further: 'Not only do such media allow publics to acquire visibility, and to acquire visibility on their own terms, but they also allow them to define the visibility of others, to become organizers of visibility' (p. 143). Such 'visibility entrepreneurs' can challenge the narrative offered by mainstream media, promoting debate and even scandal, but they can also encourage a battlefield mentality in which professionals assert their legitimacy against 'uninvited intruders' (p. 145). In either case, this is 'an attempt at interfering with a silencing process' (p. 150), in which various forms of expression can be heard or seen, and compared on an equal footing. New representational practices offer innovative ways for users and spectators to find social and political meanings while negotiating issues of truth, trust and credibility. At the same time, the traditional facilitators of the public sphere face increasing economic pressures to perform such a role. In some cases this means the loss of established newspapers, or in a recent example specific to news images, the *Chicago Sun-Times* sacking its entire photography department (BBC, 2013).

Rather than thinking about conditions of visibility from the politicians' perspectives, Dayan's paradigm of visibility helps us to conceptualize online and physical spaces as contested sites of political meaning, values and identity, in dialogue with other more traditional mediated forms. The question then becomes one of investigating the different patterns of involvement and the kinds of visual display produced and circulated across the mediascape.

PUBLICS AND SPECTATORS: BECOMING VISIBLE CITIZENS IN VISUAL DISPLAYS

Dayan's paradigm of visibility, with its emphasis on the media coordination of collective attention, offers a useful way to think about how citizenship is performed and the reactions of publics to different attempts to disrupt the status quo. In her book, *Revolting Subjects*, Imogen Tyler (2013) explores how certain groups in society are figured as 'revolting' and how such stigmatization can be resisted through aesthetic and political strategies. In the chapter on the August 2011 riots in the UK, feelings of abandonment and alienation expressed by the rioters are intricately linked to their sense of invisibility in the social body. Tyler cites the Guardian–London School of Economics (LSE) research project, Reading the Riots, which interviewed 270 people who had participated in the riots: 'This sense of being invisible was widespread' (Lewis et al., 2011: 25, cited in Tyler, 2013: 197). But not all attempts to combat societal invisibility – and especially those of a chaotic and violent nature – lead watching publics to reassess their prior attitudes; indeed in the case of the riots, the outraged media attention, public fear and harsh sentences handed down to participants all suggest that rather than effecting an 'alternative aesthetics', the rioters 'became the abjects they had been told they were' (Tyler, 2013: 204). One way in which the rioters and bystanders attempted to make sense of the disturbances and looting in their own locales was through the sharing of camera images via Twitter; interestingly, these shared images included captured TV screenshots in addition to on-the-street user-generated images, indicating a sense of involvement through 'practices of remote witnessing' alongside the impulse to capture the here and now (Vis et al., 2013: 396). The sharing of such images also records the 'second screen' phenomenon, whereby users participate via websites whilst also watching television; live media events and elections tend to attract this activity but a regular UK example is the high levels of tweeting during the BBC political panel show *Question Time* (using the #bbcqt hashtag) (Anstead and O'Loughlin, 2011).

This struggle for visibility is often most compelling in the 'image politics' (DeLuca, 1999) of political protest and social movements, with those less powerful utilizing the rhetoric of the visual in order to gain the eyes and ears of the public and, ultimately, society's decision-makers. With the diverse groups under the anti-globalization banner dominating the scene in the later twentieth century, the truly transnational character of protest coalesced in 2011 as a mixture of movements with revolutionary aims (for example, Occupy, 15-M and the 'Arab Spring' demonstrations) took to the streets in spectacular style, claiming their right to be visible and vocal

in, at times, carnivalesque displays (see Castells, 2012; Gerbaudo, 2012; Khatib, 2012).

In their analysis of YouTube videos uploaded and shared over the 18 days of the Egyptian uprising in 2011, starting with the mobilization of demonstrators on 25 January, Mohamed Nanabhay and Roxane Farmanfarmaian write of an 'amplified public sphere' created through the complex interplay of the 'inter-related spaces of the physical (protests), the analogue (satellite television and other mainstream media) and the digital (internet and social media)' (Nanabhay and Farmanfarmaian, 2011: 573). Warning against simplistic separations between amateur and mainstream image-making, or singling out social media such as Facebook or Twitter, their study points to a symbiotic relationship between journalists and activists, producers and consumers. Their point on the importance of physical place is reinforced in Paulo Gerbaudo's (2012) *Tweets and the Streets*, with the significance of assembly, solidarity and corporeality emphasized in the reappropriation of public space, such as the 15 May 2011 demonstrations in cities across Spain (and in what became known as the 15-M movement, or *los indignados*). Crucially, the *indignados* of Spain were also expressing their collective indignation, and as Gerbaudo argues, while social media was central in mobilizing the demonstrations and protest camps that followed, it was in the symbolic and material concentration in physical spaces, such as in Puerta del Sol in Madrid, that protesters rediscovered 'a sense of physical communion' (Gerbaudo, 2012: 96). In adopting the chant of 'We are not on Facebook, we are on the streets', this was about appearing as a tangible public, and signals how visibility is central to this debate (p. 96).

In the summer of 2013, Taksim Square in Istanbul, and the adjoining Gezi Park, became the latest symbol of an extended protest and clash with authority, with the initial protests against the square's development mutating into widespread demonstrations against the authoritarian nature of the government. Critically, it is the imposition of a commercial redevelopment in an iconic public space that first sparked the protests, while the reaction of Turkish Prime Minister Recep Tayyip Erdogan was to take a dismissive stance towards the protesters and the solidarity expressed via social media, characterizing Twitter as a 'menace' to society (Shafak, 2013).

In setting up camps and demonstrations in the heart of cities around the world, such 'spectacles of dissent' (D'Arcus, 2006) are both about mobilizing local people and embodying revolutionary zeal in the immediate physical place, but also about amplifying their countercultural message through the documentation and circulation of images and videos online. As Tina Askanius argues, we are seeing an 'aesthetization of public protest' in the theatrical displays and vast body of images and videos created: 'This emerging audio-visual repository of interconnected narratives stages popular contestation within a coherent framework and constitutes the basis from which collective identity formation is forged among activists scattered around the world' (Askanius, 2010: 341). The hope of such activists is to challenge the legitimacy of existing power relations through alternative forms of organization and political identity, adopting visually strong practices that range from the raw and antagonistic to the absurdly humorous. Mobile media devices also allow protesters to subvert the surveillance tactics of police, playfully mimicking their recording practices and even capturing violent police behaviour which has led to public inquiries and arrests (Archibald, 2011; Shaw, 2013). The 15-M and Occupy movements are emblematic of the challenge to democratic authority through networked action, but also of the desire to create an experiential and sensorially rich public space for imaginative reworking of what 'the political' might mean.

POLITICIANS ONLINE: SHARING A HUG AND GETTING THE JOKE

When Barack Obama realized he had secured victory in the 2012 US presidential election, his team chose to announce his second term by tweeting an image of the President hugging his wife Michelle with the simple statement: 'Four more years'. The image became the most retweeted in Twitter's history, signalling the desire of more than 816000 people who shared the image in those first few days (Ries, 2012) to join in some kind of communicative political action or affinity; but also signalling a political leader comfortable with announcing his victory via social media, before his televised speech. The photograph had been taken months earlier in mid-August by Scout Tufankjian, a campaign photographer, and was selected by digital staffer Laura Olin working as part of Obama's social media team. Olin puts Obama's successful social media campaign down to choosing diligent staff who also 'knew their social-media shit', and letting those staff 'talk to people like people' (cited in Ries, 2012). In this way, Obama's campaign harnessed the visual and verbal potential of online mediation, and achieved the much-sought-after 'virality' (the copying and multiplying of a message or phenomenon through social networks) which generally eludes professional political marketing. The particular expressive qualities of the photograph are significant and strategic: its selection speaks to the personalized and emotional appeal of a president who also appears as a loving husband, embedded in a political style that 'weaves

together matter and manner, principle and presentation, in an attractively coherent and credible political performance' (Pels, 2003: 57).

Effective use of social media, as with other media genres, is contingent on embracing socio-technical knowhow and competences in mediated visibility. This enables a projected image of authenticity and integrity and is a recipe that few politicians accomplish, at least with any consistency. Images especially thrive in a digital world of mash-up, montage, juxtaposition, repetition and manipulation. While Obama's pictured embrace of his wife went viral and became an iconic image of the campaign, Limor Shifman would distinguish this viral image from the 'memetic video' or image which '*lures extensive creative user engagement* in the form of parody, pastiche, mash-ups or other derivative work' (Shifman, 2012: 190). These images are much more indicative of the participatory culture of the Internet, according to Shifman. The most popular Internet memes, whether user-generated or popular culture-related, are humorous and playful but not necessarily political. Those dealing directly with politics can express a range of expressive modes, from light-hearted mockery to oppositional fervour. Visual memes are particularly suited to travelling across national borders, whether mocking in tone (such as the 'Pepper Spray Cop', http://pepperspraying cop.tumblr.com/, PhotoShopped after Lieutenant John Pike casually sprayed peaceful Occupy protesters in California), or symbolic of a struggle against corruption and state violence, (for example, the 'We are all Khaled Said' Facebook page and the later appropriations of Khaled's photograph as a 'visual injustice symbol' by activists during the Egyptian uprising; Olesen, 2013; see also Khatib, 2012).

The success of the viral dance video by South Korean pop-star, PSY, 'Gangnam Style', subsequently inspired its own political parodies, with 'Mitt Romney Style' and 'Kim Jong Style' examples of 'prosumer'-generated material with political intent (Müller and Kappas, 2013). In some cases, politicians refer to their own memes, signalling that they are in on the joke and conversant in the socio-technological practices of online platforms: On 26 August 2012 Barack Obama signed off from his Reddit 'Ask me anything' session with a reference to the 'Not Bad' meme. The 'Not Bad' meme is based on a photograph of the US President pulling a strange expression known as a 'sturgeon face' during a visit to the UK in May 2011, and which is thought to convey general satisfaction. The combination of media platform and intertextual playfulness signal a rare mix of approachability and self-assurance. UK Deputy Prime Minister, Nick Clegg, showed he too could get the joke when a video he recorded apologising to university students for breaking his election pledge on tuition fees was subsequently set to music and released on satirical website 'The

Poke' (http://www.thepoke.co.uk): Clegg gave his consent for the video to be released as a charity single, seemingly content to join in the mocking of his own insincerity. Nevertheless, emergent hostility against a strong cult of visual iconography can signal more than satirical or light-hearted ridiculing of political leaders. The destruction of material posters, murals and statues, along with the subsequent circulation of the images and videos depicting such protests online, provides a symbolic rejection of the visual legacy of a regime; as happened in Syria in March 2011, when footage of the ruling Assad family posters being ripped from buildings was shared across YouTube, Twitter and Facebook (Caldwell, 2011).

FUTURE RESEARCH QUESTIONS AND CONCLUDING COMMENTS

This chapter has so far attempted to set the scene for researching visual politics online. In: (1) outlining the traditional tensions between politics, popular culture and images; (2) emphasizing the interrelated notions of visibility and visuality (including how we appear as political actors and the representational forms employed); and (3) providing some recent illustrations of visual politics 'from below' and 'from above', a number of questions emerge for further consideration and examination:

- How might we better understand imaging practices and online activity (posting, viewing, commenting) as meaningful political participation?
- What kinds of political performance are best suited to the hybrid, potentially global, political information environment? And how do we analyse their effectiveness?
- Are the most visible and visual elements of mediated politics online representative of our political cultures, or do they offer a distortive perspective?

Such questions are best approached with keen attention to representational forms, alongside contexts of production, mediation and consumption. Similarly to the old debates on the spectacle of politics, there is a danger that the current fascination with how images or videos circulate online and their role in mobilizing support leads to a lack of attention given to the actual images as expressive forms: 'Rather remarkably, the systematic study of the structure and the expressive means of the image itself ("image studies") is relatively rarely practised . . . there is a persistent misunderstanding that one can go without insight into the structure of

images or other visual artifacts' (Pauwels, 2008: 84). As indicated earlier, pulling together approaches and methods from a variety of research fields can offer an enriched perspective from which to question and problematize the nature of visibility and visuality as encountered in the political realm.

Levels of attentiveness and interest are not assured by the expanse of networked political information available, and participation remains unequal across different communities and socio-economic groups, but the affordances of the Internet-based technologies undoubtedly enable access to an abundance of visual display from around the world, and offer cheap and easy ways to generate new material. Paradoxically, the motivations behind the patterns and practices of re-presentation, linking and sharing might work to question the digital images' supposed inherent ambiguity and the 'post-photographic' disruption to truth claims. Our investment in digital images as forms of communicative expression suggests a more complex picture than a simple characterization of shallow naivety or post-photography scepticism. Whether perceived as compelling evidential material, or as profound and transformative expressions of solidarity or affinity, the role of digital images in political discourse has undoubtedly been enhanced through Internet-based presentation and the resulting collective (and at times disorderly) debates they inspire on meanings, values and identities.

FURTHER READING

Dayan, D. (2013). Conquering visibility, conferring visibility: visibility seekers and media performance. *International Journal of Communication*, 7, 137–153.
Finnegan, C.A. and Kang, J. (2004). 'Sighting' the public: iconoclasm and public sphere theory. *Quarterly Journal of Speech*, 90 (4), 377–402.
Gerbaudo, P. (2012). *Tweets and the Streets: Social Media and Contemporary Activism*. New York: Pluto Press.
Khatib, L. (2012). *Image Politics in the Middle East: The Role of the Visual in Political Struggle*. London: I.B. Tauris.
McLagan, M. and McKee, Y. (2012). *Sensible Politics: The Visual Culture of Nongovernmental Activism*. New York: Zone Books.
Mitchell, W.J.T. (2005). There are no visual media. *Journal of Visual Culture*, 4 (2), 257–266.
Pels, D. (2003). Aesthetic representation and political style: re-balancing identity and difference in media democracy. In Corner, J. and Pels, D. (eds), *Media and the Restyling of Politics* (pp. 41–66). London: Sage.
van Zoonen, L., Vis, F. and Mihelj, S. (2010). Performing citizenship on YouTube: activism, satire and online debate around the anti-Islam video Fitna. *Critical Discourse Studies*, 7 (4), 249–262.

REFERENCES

Anderson, B. (1991). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London, UK and New York, USA: Verso.

Anstead, N. and O'Loughlin, B. (2011). The emerging Viewertariat and BBC *Question Time*: television debate and real-time commenting online. *International Journal of Press/Politics*, 16 (4), 440–462.

Archibald, D. (2011). Photography, the police and protest: images of the G-20, London 2009. In Cottle, S and Lester, L. (eds), *Transnational Protests and the Media* (pp. 129–139). Oxford: Peter Lang.

Askanius, T. (2010). Video Activism 2.0 – space, place and audiovisual imagery. In Hedling, E., Hedling, O. and Jönsson, M. (eds), *Regional Aesthetics: Locating Swedish Media* (pp. 337–358). Stockholm: Kungliga Biblioteket.

Bal, M. (2003). Visual essentialism and the object of visual culture. *Journal of Visual Culture*, 2 (1), 5–32.

Baym, G. and Jones, J. (2012). News parody in international perspective: politics, power and resistance. *Popular Communication*, 10 (1–2), 2–13.

BBC (2013). *Chicago Sun-Times* sacks entire photo department. *BBC News US and Canada*, 30 May. Retrieved 7 June 2013 from http://www.bbc.co.uk/news/world-us-canada-22723725.

Bennett, W.L. (2012). The personalization of politics: political identity, social media, and changing patterns of participation. *Annals of the American Academy of Political and Social Science*, 644 (1), 20–39.

Blumler, J.G. and Kavanagh, D. (1999). The third age of political communication: Influences and features. *Political Communication*, 16 (3), 209–230.

Boorstin, D.J. (1962). *The Image*. New York: Atheneum.

Caldwell, L. (2011). The new face of President Asad on YouTube. *Arab Media and Society*, Issue 13, retrieved from http://www.arabmediasociety.com/index.php?article=776&p=0.

Castells, M. (2012). *Networks of Outrage and Hope: Social Movements in the Internet Age*. Cambridge, UK and Malden, MA, USA: Polity.

Chadwick, A. (2013). *The Hybrid Media System: Politics and Power*. Oxford: Oxford University Press.

Coleman, S. and Blumler, J.G. (2009). *The Internet and Democratic Citizenship: Theory, Practice and Policy*. Cambridge: Cambridge University Press.

Corner, J. (2003). Mediated persona and political culture. In Corner, J. and Pels, D. (eds), *Media and the Restyling of Politics* (pp. 67–84). London: Sage.

Corner, J. and Pels, D. (2003). *Media and the Restyling of Politics: Consumerism, Celebrity and Cynicism*. London: Sage.

Crick, B. (1968). *In Defence of Politics*. London: Penguin.

D'Arcus, B. (2006). *Boundaries of Dissent: Protest and State Power in the Media Age*. New York: Routledge.

Dayan, D. (2013). Conquering visibility, conferring visibility: visibility seekers and media performance. *International Journal of Communication*, 7, 137–153.

DeLuca, K. (1999). *Image Politics*. New York: Guilford Press.

Doerr, N. and Teune, S. (2012). The imagery of power facing the power of imagery: towards a visual analysis of social movements. In Fahlenbrach, K., Klimke, M., Scharloth, J. and Wong, L. (eds), *The 'Establishment' Responds: Power and Protest During and After the Cold War* (pp. 43–55). Cambridge, UK and New York, USA: Berghahn Books.

Du Gay, P., Hall, S., Janes, L., Mackay, H. and Negus, K. (1997). *Doing Cultural Studies*. London: Sage/The Open University.

Edelman, M. (1988). *Constructing the Political Spectacle*. Chicago, IL: University of Chicago.

Finnegan, C.A. and Kang, J. (2004). 'Sighting' the public: iconoclasm and public sphere theory. *Quarterly Journal of Speech*, 90 (4), 377–402.

Flinders, M. (2012). *Defending Politics: Why Democracy Matters in the Twenty-First Century*. Oxford: Oxford University Press.

Gerbaudo, P. (2012). *Tweets and the Streets: Social Media and Contemporary Activism*. New York: Pluto Press.

Green, J.E. (2010). *The Eyes of the People: Democracy in an Age of Spectatorship*. New York: Oxford University Press.

Jay, M. (1993). *Downcast Eyes: The Denigration of Vision in Twentieth-Century French Thought*. Berkeley, CA: University of California Press.

Katz, E. (1988). Disintermediation: cutting out the middleman. *Inter Media*, 16 (2), 30–31.

Khatib, L. (2012). *Image Politics in the Middle East: The Role of the Visual in Political Struggle*. London: I.B. Tauris.

Langer, A.I. (2011). *The Personalisation of Politics in the UK: Mediated Leadership from Attlee to Cameron*. Manchester: Manchester University Press.

Lippmann, W. (1922). *Public Opinion*. New York: Macmillan.

Manin, B. (1997). *The Principles of Representative Government*. New York: Cambridge University Press.

Mazzoleni, G. and Schulz, W. (1999). 'Mediatization' of politics: a challenge for democracy?. *Political Communication*, 16 (3), 247–261.

McChesney, R. (1999). *Rich Media, Poor Democracy: Communication Politics in Dubious Times*. Urbana, IL: University of Illinois Press.

McLagan, M. and McKee, Y. (2012). *Sensible Politics: The Visual Culture of Nongovernmental Activism*. New York: Zone Books.

Meyer, T. and Hinchman, L. (2002). *Media Democracy: How the Media Colonize Politics*. Cambridge: Polity.

Mirzoeff, N. (2011). *The Right to Look: A Counterhistory of Visuality*. Durham, NC, USA and London, UK: Duke University Press.

Mitchell, W.J.T. (1994). *Picture Theory: Essays on Verbal and Visual Representation*. Chicago, IL, USA and London, UK: University of Chicago Press.

Mitchell, W.J.T. (2002). Showing seeing: a critique of visual culture. *Journal of Visual Culture*, 1 (2), 165–181.

Mitchell, W.J.T. (2005). There are no visual media. *Journal of Visual Culture*, 4 (2), 257–266.

Müller, M.G. and Kappas, A. (2013). Politics Mitt Romney style: Gangnam style as a cross-cultural visual meme – online citizen creativity and the power of digitally facilitated political prosumer participation. Presented at 63rd Annual International Communication Association Conference, 21 June, London.

Nanabhay, M. and Farmanfarmaian, R. (2011). From spectacle to spectacular: how physical space, social media and mainstream broadcast amplified the public sphere in Egypt's 'revolution'. *Journal of North African Studies*, 16 (4), 573–603.

Olesen, T. (2013). 'We are all Khaled Said': visual injustice symbols in the Egyptian revolution, 2010–2011. In Doerr, N., Mattoni, A. and Teune, S. (eds), *Advances in the Visual Analysis of Social Movements*, Research in Social Movements, Conflicts and Change, Vol. 35 (pp. 3–25). Bingley: Emerald Group.

Pauwels, L. (2008). Visual literacy and visual culture: reflections on developing more varied and explicit visual competencies. *Open Communication Journal*, 2, 79–85.

Pels, D. (2003). Aesthetic representation and political style: re-balancing identity and difference in media democracy. In Corner, J. and Pels, D. (eds), *Media and the Restyling of Politics* (pp. 41–66). London: Sage.

Postman, N. (1987). *Amusing Ourselves to Death*. London: Methuen.

Putnam, R.D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Rancière, J. (2011). *The Emancipated Spectator*. London, UK and New York, USA: Verso.

Richardson K., Parry, K. and Corner, J. (2012). *Political Culture and Media Genre: Beyond the News*. Basingstoke: Palgrave.

Ries, B. (2012). The story behind the most viral photo ever. *Daily Beast*, 19 November. Retrieved from http://www.thedailybeast.com/articles/2012/11/19/the-story-behind-the-most-viral-photo-ever.html.

Shafak, E. (2013). The view from Taksim Square. *Guardian*, 4 June. G2 supplement, pp. 6–8.

Shaw, F. (2013). 'Walls of seeing': protest surveillance, embodied boundaries, and counter-surveillance at Occupy Sydney. *Transformations*, 23. Retrieved from http://www.transformationsjournal.org/journal/issue_23/article_04.shtml.
Shifman, L. (2012). An anatomy of a YouTube meme. *New Media and Society*, 14 (2), 187–203.
Stanyer, J. (2012). *Intimate Politics: The Rise of the Celebrity Politician and the Decline of Privacy*. Cambridge: Polity.
Street, J. (1997). *Politics and Popular Culture*. Philadelphia, PA: Temple University Press.
Strömbäck, J. (2008). Four phases of mediatization: an analysis of the mediatization of politics. *International Journal of Press/Politics*, 13 (3), 228–246.
Thompson, J.B. (1995). *The Media and Modernity*. Stanford, CA: Stanford University Press.
Thompson, J.B. (2005). The new visibility. *Theory, Culture and Society*, 22 (6), 31–51.
Thumim, N. (2012). *Self-Representation and Digital Culture*. Basingstoke: Palgrave.
Thussu, D.K. (2007). *News as Entertainment: The Rise of Global Infotainment*. London: Sage.
Tyler, I. (2013). *Revolting Subjects*. London, UK and New York, USA: Zed Books.
van Zoonen, L. (2005). *Entertaining the Citizen: When Politics and Popular Culture Converge*. Lanham: Rowman & Littlefield.
van Zoonen, L., Vis, F. and Mihelj, S. (2010). Performing citizenship on YouTube: activism, satire and online debate around the anti-Islam video Fitna. *Critical Discourse Studies*, 7 (4), 249–262.
Vis, F., Faulkner, S., Parry, K., Manyukhina, Y. and Evans, L. (2013). Twitpic-ing the riots: analysing images shared on Twitter during the 2011 UK riots. In Weller, K., Bruns, A., Puschmann, C., Burgess, J. and Mahrt, M. (eds), *Twitter and Society* (pp. 385–398). New York: Peter Lang.

24. Automated content analysis of online political communication

Ross Petchler and Sandra González-Bailón

INTRODUCTION

Content analysis has a long tradition in the social sciences. It is central to the study of policy preferences (Budge, 2001; Laver et al., 2003), propaganda and mass media (Krippendorff, 2013 [1980]; Krippendorff and Bock, 2008), and social movements (Della Porta and Diani, 2006; Johnston and Noakes, 2005). New computational tools and the increasing availability of digitized documents promise to push forward this line of inquiry by reducing the costs of manual annotation and enabling the analysis of large-scale corpora. In particular, the automated analysis of online political communication may yield insights into political sentiment which offline opinion analysis instruments (such as polls) fail to capture. Online communication is constantly pulsating, generating data that can help us uncover the mechanisms of opinion formation – if the appropriate measurement and validity methods are developed.

Several linguistic peculiarities distinguish online political communication from traditional political texts. For a start, it is often far less formal and structured. In addition, automated content analysis techniques are not always as reliable or as valid as manual annotation, which makes measurements potentially noisy or misleading. With these challenges in mind, we provide an overview of techniques suited to two common content analysis tasks: classifying documents into known categories, and discovering unknown categories from documents (Liu, 2012; Blei, 2013). This second task is more exploratory in nature: it helps to identify topic domains when there are no clear preconceptions of the topics that are discussed in a certain communication environment. The first task, on the other hand, can help to label a large volume of text in a more efficient manner than manual annotation; for instance, when the research question requires identifying the emotional tone of political communication (as positive, negative or neutral) or its ideological slant (liberal or conservative). This chapter focuses on the application of these automated techniques to online political communication, and suggests directions for future research in this domain.

METHODS FOR AUTOMATED CONTENT ANALYSIS

The application of automated text analysis techniques requires the prior acquisition and preprocessing of data. This section discusses the logic of preprocessing texts to then provide an overview of techniques to classify documents in known categories or discover topics when no categories are known.

Acquiring and Preprocessing Online Political Texts

Political scientists have applied automated content analysis techniques to many kinds of offline political texts, including newspaper articles (Young and Soroka, 2012), presidential and legislator statements (Grimmer and King, 2011), legislature floor speeches (Quinn et al., 2010), and treaties (Spirling, 2012). Until recently, though, political texts remained relatively understudied because they were difficult to parse and process for analysis.

Acquiring online political texts is becoming simpler as more sites store and transmit them in machine-readable formats such as Extensible Markup Language (XML) and JavaScript Object Notation (JSON) or make them publicly available via application programming interfaces (APIs). When such options are unavailable, researchers familiar with statistical software or scripting languages can use new packages for automated HyperText Markup Language (HTML) scraping (Python and R, for instance, have built-in packages and libraries). Finally, when neither machine-readable nor easy-to-parse HTML data are available, researchers can crowdsource data acquisition and parsing via sites like Amazon Mechanical Turk (Berinsky et al., 2012). Overall, these new technologies enable communication scholars to access and study previously unavailable indicators of public opinion.

In order to perform automated content analysis researchers must transform texts into structured data that can be quantified (Franzosi, 2004). Prior to the advent of new computational tools this was performed by human coders using a pre-determined scheme (Krippendorf, 2013 [1980]; Neuendorf, 2001). Initially, a codebook is written guided by a research question and a theoretical context. It is iteratively improved until coders no longer notice ambiguities, at which point it is applied to the data set. Automated approaches preprocess the text to reduce the complexity of language, often using a bag-of-words model to eliminate the most frequent words and to reduce words to their morphological roots (Jurafsky and Martin, 2009; Hopkins and King, 2010; Porter, 1980). After preprocessing, documents are represented as a document-term matrix in which rows

correspond to documents, sentences, or expressions (depending on the unit of analysis), and columns correspond to words or tokens. Cells can contain continuous values (representing how frequently each term occurs in each document) or binary values (representing whether each term occurs in the document).

Which of the two approaches is more appropriate (to code documents manually or to apply automated preprocessing) depends on the complexity of the research question at hand, the number of documents collected, and the tolerance for error. Although manually annotated data remain the gold standard for content analysis, the sections that follow focus mostly on cases in which data are automatically preprocessed, since this is more common when dealing with large volumes of text.

Once online political texts are converted to a structured form, several methods for automated content analysis can be applied. We divide these methods into two groups to reflect the two most common content analysis tasks: classifying documents into known categories, and discovering theoretically important categories from the content. The former task encompasses techniques such as lexicon-based classification and supervised learning. The latter task encompasses unsupervised learning and the analysis of text as networks of concepts. The following sections explain the details of these techniques and highlight their relative strengths and weaknesses to help communication researchers choose the approach best suited to their specific data and research question.

Classifying Documents into Known Categories

The goal of supervised content analysis techniques is to classify documents into a number of known categories. For example, news articles may have left-leaning or right-leaning ideological biases (Gentzkow and Shapiro, 2010) or have positive or negative coverage (Eshbaugh-Soha, 2010). This section offers an overview of the techniques that allow that sort of classification. There are two main methods. The first is a lexicon-based approach, which uses relative keyword frequencies to measure the prevalence of each category in a document. The second is supervised learning, which uses a training data set of manually annotated documents to classify new, unlabeled documents.

Lexicon-based classification

The lexicon (or dictionary)-based approach to document classification is the simplest automated content analysis technique (Liu, 2012). It is based on a list of words and phrases and their associated category labels. For example, a lexicon for classifying micro-blog posts according to sentiment may map the words 'good' and 'beautiful' to the positive category and the words 'bad' and 'ugly' to the negative category. A lexicon for classifying blog posts according to ideological subject may map the words 'health-care' and 'environment' to the left-leaning category and 'foreign policy' and 'taxes' to the right-leaning category.

Off-the-shelf lexicons include the Linguistic Inquiry and Word Count, or LIWC (Pennebaker et al., 2001), and the General Inquirer (Wilson et al., 2005). Not all lexicons are based on binary categories. Some sentiment lexicons have positive, neutral, and negative terms, measured on a several points scale. The Affective Norms for English Words (ANEW) lexicon, for instance, labels words and phrases according to psychometric categories which rate words on three emotional dimensions: valence, arousal, and dominance (Bradley and Lang, 1999; Osgood et al., 1957). This lexicon helps to analyze documents by counting the relative frequency with which words appear and averaging the scores associated to each word in each dimension, from 0 to 9. This approach has been applied effectively to extract sentiment measures from a number of online data sources (Dodds and Danforth, 2009; Dodds et al., 2011).

The success of a lexicon-based content analysis relies on the quality of the lexicon; that is, how appropriate it is in the context of the specific research question and data being analyzed (González-Bailón and Paltoglou, 2015). Using 'off-the-shelf' lexicons compiled with generic research goals may produce poor results when applied to specific types of political communication (Loughran and McDonald, 2011). It is always best to generate lexicons specific to a research question, and there are three main approaches for doing so. The first is to manually annotate the sentiment of all adjectives in a dictionary of all the words in a corpus, in line with the information domain under scrutiny (that is, 'warming' can be labeled differently if used in environmental policy or foreign affairs communication). This is time-consuming but tunes the lexicon to specific communication contexts. Researchers concerned with efficiency as well as accuracy have used online crowdsourcing platforms such as CrowdFlower, Amazon Mechanical Turk, and Taskcn to quickly and accurately label large sentiment lexicons. For example, Dodds et al. (2011) created a lexicon of 10222 words by merging the 5000 most frequently occurring words in a Tweet corpus, Google Books, music lyrics, and the *New York Times*; they then used Amazon Mechanical Turk to obtain 50 sentiment ratings of each word on a nine-point scale from negative to positive. They found that the sentiment lexicon labeled by crowdsourcing workers was highly correlated with the ANEW lexicon.

The second way to generate a sentiment lexicon is dictionary-based. The general approach is to manually label the sentiment of a small set of seed words, and then search a dictionary (the most frequently used is WordNet; see Miller et al., 1990) for their synonyms and antonyms; these snowballed terms are then labeled with the same or opposite sentiment as the corresponding seed word and then are added to the set of seed words. The process is iterated until no words remain unlabeled. For example, the seed word 'excellent' is labeled positive; synonyms such as 'beautiful', 'fabulous', and 'marvelous' are labeled as positive as well; while antonyms such as 'awful', 'rotten', and 'terrible' are labeled as negative. An example of a lexicon generated using the dictionary-based approach is Sentiment Lexicon, constructed by Hu and Liu (2004). This dictionary-based approach quickly generates a large list of labeled sentiment words, but requires manual cleaning and ignores ambiguity due to context, which is particularly important in the analysis of political communication.

The third way to generate a sentiment lexicon is corpus-based. The general approach is to manually label the sentiment of a small set of seed words and then define linguistic rules to identify similar or dissimilar sentiment words. A seed word may be 'beautiful' and its label 'positive'; linguistic rules based on connective words (such as 'and' or 'but') help to assign labels to subsequent words. For instance, if a document in a corpus contains the phrase 'The car is beautiful and spacious' then the term 'spacious' could be assigned the label 'positive' based on the connective word 'and'. Conversely, if a document in a corpus contains the phrase 'The car is spacious but difficult to drive' then the term 'spacious' could be assigned the label 'negative' based on the connective word 'but'. This methodology requires clearly defined linguistic rules in order to achieve good results; and linguistic rules assume sentiment consistency across documents, which is not necessarily the case for most empirical domains: the same word can express opposite sentiments in different communication contexts (Liu, 2012). Overall, though, the corpus-based methodology to lexicon generation is useful in two cases: to discover other sentiment words and their orientations on the basis of a hand-made seed list; and to adapt a general-purpose lexicon to a specific communication domain. The corpus-based approach is less useful for building a general-purpose sentiment lexicon than the dictionary-based approach because dictionaries encompass more words.

These three techniques are based on different assumptions that affect the results they produce. None of these sentiment lexicons is perfect because they are too general to suit the specific needs of different communication domains. In addition, certain words and phrases in online political communication are too informal, specific, or novel (and therefore infrequent) to be contained in existing lexicons. A corpus-based technique can capture and label these distinct words; for instance, Brody and Diakopoulos (2011) find that lengthened words in microblog posts (for example, 'looove') are strongly associated with subjectivity and sentiment; and Derks et al. (2007) find that emoticons (for example, ':)') strengthen the intensity of online communication. Researchers have already incorporated the peculiarities of online communication into their sentiment models (Paltoglou et al., 2010; Paltoglou and Thelwall, 2012), but often additional manual labeling is needed to add other novel words to the seed list. These limitations make validation a crucial component of automated content analysis (Grimmer and Stewart, 2013). Having the appropriate validation strategies in place is necessary to increase confidence in measurement.

Supervised learning

The second main approach to document classification using pre-existing categories is supervised learning. Supervised algorithms learn from a training data set of manually annotated documents how to classify new, unlabeled documents. The supervised learning approach has three steps. First, it constructs a training data set. Second, it applies an automated algorithm to determine the relationships between features of the training data set and the categories that are used to classify documents. And third, it predicts (or assigns) categories for unlabeled documents and validates that classification. The remainder of this section reviews these three steps in turn.

The first step in supervised learning is to construct a training data set. As described above, this involves transforming unstructured textual data into structured quantitative data. In addition to preprocessing, it is common for researchers to manually code documents for features that the bag-of-words model ignores; for instance, they may add features accounting for the source or the author of a document. The larger the training data set, the more information supervised learning algorithms have with which to make predictions, but scaling up can be computationally costly. The specific research question and data source inform the balance between the need for a large training data set and the costs of compiling training data.

The second step in supervised learning is to apply an algorithm that will associate text features to each category in the classification scheme. There are many different algorithms and the field of machine learning and natural language processing is quickly growing in this area; Hastie et al. (2009) offer a good overview of the techniques available. Each model has specific characteristics and parameters, which makes a general discussion difficult, but popular algorithms include (multinomial) logistic regression, the naive Bayes classifier (Maron and Kuhns, 1960), random forests (Breiman, 2001), support vector machines (Cortes and Vapnik, 1995), and neural networks (Bishop, 1995). Each of these algorithms uses the information gathered from the training data to assign new examples of text into the classification categories.

This assignment takes place in the third and final step, where supervised approaches predict the categories for unlabeled documents and validate the results. A model that performs well will replicate the results of manual coding, which still offers the gold standard; a model that performs poorly will fail to replicate these results. The standard method to validate models is cross-validation. This entails splitting the labeled documents into equally sized groups (usually about ten) and then predicting the categories of the observations in each group using the pooled observations in the other groups. This method avoids overfitting to data because it focuses on out-of-sample prediction. Overall, the supervised approach systematically performs better than unsupervised approaches in the analysis of online communication because it is able to capture more accurately the contextual features of the text and language used (González-Bailón and Paltoglou, 2015).

Discovering Categories and Topics from Documents

Unsupervised learning

In contrast to supervised approaches, unsupervised techniques do not require manually annotated training data; consequently, they are much less costly to implement. They are good exploratory techniques but their results can be difficult to evaluate: concepts such as validity and consistency compared to human labeling do not immediately apply because these techniques are used, in part, to overcome the lack of predefined labels or categories – hence their exploratory nature. This section briefly discusses three categories of unsupervised techniques: cluster analysis, dimensionality reduction, and topic modeling.

The goal of cluster analysis is to partition a corpus of documents into groups of similar documents, where 'similar' is measured in terms of word frequency distributions. The most widely used clustering algorithm is k-means (MacQueen, 1967), which partitions documents into k disjoint groups by minimizing the sum of the squared Euclidean distances within clusters; distance is measured as the number of words that any two documents share. Other clustering algorithms use different distance metrics or objective functions (which are used to optimize or find the best clustering classification out of all possible classifications). Given that few papers provide guidance on which similarity metrics, objective functions, or optimization algorithms to choose, Grimmer and Stewart (2013) caution social scientists from importing clustering methods developed in other, more technical fields like machine learning. The computer-assisted cluster analysis technique suggested by Grimmer and King (2011) offers a more intuitive tool for the task of fully automated cluster analysis.

The goal of dimensionality reduction is to shorten the number of terms in the term-document space while maintaining the structure of the corpus. One dimensionality reduction technique is principal component analysis, which transforms a document-term matrix into linearly uncorrelated variables that correspond to the latent semantic topics in the data set. The technique is not different from more conventional uses in multivariate modeling where a subset of variables are selected to represent a larger data set (Dunteman, 1989). A related dimensionality reduction technique is multidimensional scaling, which projects a corpus of documents into *N*-dimensional space such that the distances between documents correspond to dissimilarities between them. These methods provide good intuition of the topics that characterize a corpus of text but are best used as exploratory techniques; principal component analysis, in particular, is a typical data reduction step performed prior to subsequent, more substantive analysis.

Finally, the goal of topic modeling is to represent each document as a mixture of topics. Each topic is a probability mass function over words that reflect a distinct information domain. For instance, the topic 'foreign policy' may assign high probabilities to words such as 'war', 'treaty', and 'Iraq'; while the topic 'economy' may assign high probabilities to words such as 'unemployment', 'GDP', and 'labor'. The most widely used topic model is called latent Dirichlet allocation (LDA) (Blei et al., 2003). This technique has recently been applied to the analysis of newspaper content to dissect the framing of policies (DiMaggio et al., 2013). The method provides a new computational lens into the structure of texts and, as the authors state:

> finding the right lens is different than evaluating a statistical model based on a population sample. The point is not to correctly estimate population parameters, but to identify the lens through which one can see the data more clearly. Just as different lenses may be more appropriate for long-distance or middle-range vision, different models may be more appropriate depending on the analyst's substantive focus. (ibid.: 20)

Again, the crucial step in the analysis comes with validation; that is, with the substantive interpretation of the themes identified.

Network Representations of Text

As the sections above have illustrated, content analysis is essentially a relational exercise: words that relate to the same topic are associated by co-appearing frequently in documents and they tend to cluster; likewise, positive words tend to be connected to other positive words, and as shown

above, language connectors might change the affective tone of words by setting them in a different linguistic context. Networks offer a mathematical representation of the relational nature of language, and provide yet another tool for the analysis of its structure. Networks have been used to model narratives, and to analyze identity formation (Bearman and Stovel, 2000); to represent mental models (Carley and Palmquist, 1992); and to map semantic associations (Borge-Holthoefer and Arenas, 2010). A network approach has also been used with Twitter data to identify entities by looking at the co-occurrence of words and the clusters that emerge from those connections (Mathiesen et al., 2012). The nodes in these networks are words; what changes depending on the approach is the definition of the links that connect those words: co-occurrence is one of the options, but links can also be used to track the temporal evolution of narratives, as when political movements change their framing or candidates change their positions during an election campaign. These networks can be constructed and visualized using standard network analysis tools.

One of the by-products of generating a dictionary-based lexicon (discussed above) is that the method also creates a network of words that researchers can use to label the strength as well as the sign of the sentiment expressed. For example, Kamps et al. (2004) determined the strength and sentiment of words according to their distances in WordNet from labeled seed words; in this case, two words are linked if they are synonyms, and distance is measured as the number of links that need to be crossed to go from one word to another. Blair-Goldensohn et al. (2008) also used WordNet to construct a network of positive, negative, and neutral sentiment words, and then labeled the strength of the words using a propagation algorithm: starting from a seed word, its sign (positive, negative, or neutral) is propagated to all its neighboring words in the network (its synonyms); following a majority rule, that sign is further propagated to the neighbors of the neighbors, and so on, recursively, until all words have a sign assigned – the valence of which gets weaker the further apart the word is from its seed. These network-based techniques help to extend the dictionary-based approach by suggesting measures of sentiment strength.

APPLICATIONS TO THE ANALYSIS OF ONLINE POLITICAL COMMUNICATION

Sentiment in Online Political Talk

When applying sentiment analysis to political communication, it is important to remember that different methods inherit different assumptions from psychological theories of emotions. The ANEW lexicon, for instance, derives from now classic psychological research suggesting that three dimensions account for variance in the expression of emotion: valence (which ranges from pleasant to unpleasant), arousal (which ranges from calm to excited), and dominance (which ranges from domination to control; Osgood et al., 1957). Neurological research, on the other hand, suggests that five emotional dimensions underlie most brain activity: fear, disgust, anger, happiness, and sadness (Murphy et al., 2003). Reducing the breadth of human emotions to just a few dimensions is arguably a crude simplification, but necessary to make problems tractable; however, it also introduces measurement error that has to be taken into consideration when operationalizing research questions about the affective tone of political communication.

Sentiment analysis of online political communication must take into account not only measurement error but also sampling bias. Internet users, and in particular those present in social media, are typically not representative of the population: they tend to be female, young, and urban (Duggan and Brenner, 2013); in addition, the bias might be more or less important depending on the context and subject of communication. For some dimensions of public opinion, the bias might not matter, but for others it can be crucial. Again, it is only through validity tests that the measures of public opinion extracted from online communication can be relied upon (Grimmer and Stewart, 2013). The increasing number of Internet users who join social media sites and discuss politics means that the volume of online political communication is growing, and the profile of users involved is changing. Analyses of how online sentiment changes over time must therefore account for these non-stationary characteristics, typically by comparing short, adjacent periods of online communication rather than the entire history of communication on a given site.

The assumptions made by automated methods about emotional mechanisms and the nature of the samples analyzed demand a thoughtful research design when studying online communication. In many cases basic methods produce useful results that rival more sophisticated approaches; in particular, simple word frequencies and analysis of how the volume of communication fluctuates over time often yield good insights while preserving efficiency. Carvalho et al. (2011), for instance, found that in some cases these basic descriptive statistics predict sentiment as accurately as more advanced statistical techniques. This suggests that exploratory analysis can be crucial to avoid rushing into the implementation of more complex solutions when a simpler, more intuitive approach would perform as well.

In addition to the lexicons introduced above, a number of alternative

approaches have also been developed to facilitate the study of online communication. These include OpinionFinder (OF), which rates expressions as strongly or weakly subjective (Wilson et al., 2005); and the Profile of Mood States (POMS) questionnaire (Lorr et al., 2003), in which respondents rate each of 65 adjectives on a five-point scale. The questionnaire produces emotion scores in six dimensions: Tension–Anxiety, Anger–Hostility, Fatigue–Inertia, Depression–Dejection, Vigor–Activity, and Confusion–Bewilderment. Like ANEW, the POMS lexicon is suited for analyzing more complex emotions in online communication; the OF lexicon, like LIWC, is used for simpler tasks such as the identification of polarity in sentiment analysis. Other prominent lexicons optimized for the analysis of online communication include SentiWordNet (Adrea and Sebastiani, 2006) and SentiStrength (Thelwall et al., 2010). SentiStrength is particularly useful for online political communication because it includes misspellings and emoticons which abound in online talk.

Recent empirical applications of these approaches include Connor et al. (2010), Bollen et al. (2011) and Castillo et al. (2013). Connor et al. (2010) derive sentiment valence from Twitter posts using a subjectivity lexicon based on a two-step polarity classification. They compare Twitter sentiment to consumer confidence and election polling data. They find high correlations (between 0.7 and 0.8) and evidence that smoothed Twitter sentiment predicts consumer confidence (but not election) poll results with relatively high accuracy. However, Bollen et al. (2011) find that the intersection of a tweet corpus and their subjectivity lexicon is not a good leading indicator of the direction of shifts in the Dow Jones Industrial Average. This highlights how sentiment analysis of online communication may not work in all contexts: some lexicons are better suited to particular problem domains, such as consumer confidence, but not financial markets. Finally, Castillo et al. (2013) apply the SentiStrength lexicon to measure sentiment in cable news coverage; although this is traditional media content, the data were accessed through a software company that develops applications for smartphones and tablets that display extra information about TV shows, including captions of content.

Unsupervised Learning Applications

As explained above, many unsupervised learning methods are used as exploratory tools rather than testing techniques, and thus are less common in published literature on online communication. Nevertheless, a few prominent examples exist, although many are still peripheral to the core research questions of political communication.

Turney (2002), for instance, classifies online reviews as positive or negative by estimating the semantic orientation of sentences containing adjectives or adverbs. Specifically, the paper makes use of the pointwise mutual information–information retrieval (PMI-IR) algorithm to measure the number of co-occurrences between words and the seed words 'excellent' and 'poor' on AltaVista search engine results. This co-occurrence frequency determines the semantic orientation of words, and thus can be used to rate online reviews as positive or negative.

Quinn et al. (2010) use a technique similar to LDA in order to analyze the daily legislative attention given to various topics in 118000 United States Senate floor speeches from 1997 to 2004. They found 42 topics, the most prominent being legislative procedures, armed forces, social welfare, environment, and commercial infrastructure. Yano et al. (2009) use LDA in order to model topics in political blog posts and their corresponding comments sections. They found five topics: religion, (election) primary, Iraq War, energy, and domestic policy. Associated with each topic are a set of words that appeared in blog posts and a set of words that appeared in comments. Additionally, the authors predict which users are likely to comment on which blogs. Finally, another recent example applies the same method to the analysis of issue salience in the Russian blogosphere (Kolstova and Koltcov, 2013). The authors use the method to identify a shift in topics during the political protests that took place during the parliamentary and presidential elections in late 2011 and early 2012.

In sum, unsupervised methods are less frequently used because they are exploratory techniques employed to charter communication domains that lack predefined boundaries. They are good for estimating the structure of a corpus of text when no a priori classifications exist, but they still require a posteriori theoretical and subjective labeling of categories. This stands in contrast to supervised techniques: whereas manual annotation is the starting point for supervised techniques, it is the ending point for the unsupervised approach.

FUTURE LINES OF WORK

This chapter has given an overview of techniques for the automated analysis of large-scale texts, especially as they are generated in online communication. Although this is a massive area of research, and is fast evolving, a few facts have already been established. One is the consistent evidence that the effectiveness of automated classifiers is not independent from the communication domain being analyzed: the meaning of words or their emotional load varies with the context in which they are used. More work is required to build tailored dictionaries that can capture the nuances of

political communication as it takes place in different information contexts; for this, supervised-learning approaches offer the most accurate (and promising) solutions. Likewise, more work is needed to consolidate validation strategies, for instance by measuring the strength to which online measures of public opinion are correlated with more traditional measures, such as polls and surveys. A more systematic account of the efficiency and robustness of different algorithms is also needed: some corpora of text are better analyzed by certain techniques than others. Supervised methods, for instance, are more appropriate for content expressed in Twitter messages, whereas for longer communication, such as blog entries, unsupervised methods might be more appropriate. More research is needed to assess the robustness of each method for different data sources, as facilitated by online communication. In any case, the appropriateness of each technique has to be assessed in the light of each particular research question.

Validation is a crucial step in the application of automated content analysis, and this implies finding ways of assessing the accuracy, precision, and reliability of automated classifiers as compared to human coding. For instance, researchers who choose a lexicon-based approach face several design considerations. The first is what type of lexicon to generate or adapt. Some sentiment lexicons have binary categories (positive and negative), some have ternary categories (positive, negative, and neutral), and some have ordinal categories (−5 to +5, for example). A second design consideration is what word features to include in a lexicon. Some lexicons simply have word valence (ranging from positive to negative), while others have additional features such as arousal (ranging from calm to excited). The type of sentiment lexicon a researcher chooses should be based on the features a lexicon offers and the specific research question they seek to answer.

Researchers who choose a lexicon-based approach also face several implementation considerations. Most of these have to do with how to detect and resolve the complexities of text. The algorithm that implements a lexicon-based approach should often not just naively match words but also be sensitive to their local context. For instance, it should be aware of negating words (such as 'no', 'not', and 'none') and strengthening punctuation (such as exclamation marks, question marks, and ellipses). Some of the lexicons revised in this chapter, such as SentiStrength, already take these language modifiers and intensifiers into account. Good lexicon-based approaches to sentiment detection do not just rely on word matching: they are also sensitive to how the local context of each word affects the overall sentiment.

The advantage of automated content analysis is that it helps to scale up the amount of text analyzed by lowering the costs of coding and the efficiency of document classification; but it still needs to be reliable. Many sentiment lexicons are based on psychological theories of language use but it is still unclear whether these psychometric instruments work for written communication and large-scale text analysis. In addition, these techniques are still not very good at capturing essential features of political talk, such as sarcasm. A document may contain many strong sentiment words but the author might actually have intended the opposite sentiment to that captured by the automated approach. This means that automated methods might be more appropriate when applied to text in which sarcasm and figurative language are rarely used, for instance news reports; communication through social media, on the other hand, might be more vulnerable to measurement error. As the tools for automated content analysis become more prevalent in communication research, more unified standards for evaluation and assessment will have to be consolidated. The advantages of automated methods are, overall, too great to dismiss.

ACKNOWLEDGEMENTS

We would like to thank the participants of the workshop 'Extracting Public Opinion Indicators from Online Communication', sponsored by the Oxford John Fell Fund under project 113/365 while the authors were based at the University of Oxford. We are especially grateful to Scott Blinder, Javier Borge-Holthoefer, Andreas Kaltenbrunner, Patrick McSharry, Karo Moilanen, and Georgios Paltoglou for insightful discussions.

LEARNING MORE

Methods for automated content analysis are fast evolving, and any list of available resources is likely to be soon outdated. What follows are a few recommendations on where to start to learn more about the methods and applications of automated tools. Rather than an exhaustive list, these references offer entry points to what is a vast and quickly expanding area of research.

FURTHER READING

Krippendorff (2013 [1980]). Now in its third edition, this book is a classic in content analysis, a long-standing reference that precedes the explosion of automated methods for the

analysis of large-scale data. Even though the book does not consider emerging methods, the discussion on validity and reliability still applies.

Liu (2012). This monograph is one of the most up-to-date reviews of opinion mining methods. It offers an accessible discussion of state-of-the art tools for automated content analysis, and it defines basic terminology as well as research standards.

Däubler et al. (2012). This research note offers an interesting comparison of the validity of automated versus human coding in identifying basic units of text analysis. The discussion considers how automated methods offer an improvement to human coding schemes without loss of validity.

Grimmer and Stewart (2013). This article offers an interesting overview of methods that analyze text at the document level. In addition to discussing in an accessible way the basic features of different approaches, the article also emphasizes the need to develop new validation methods.

Dodds and Danforth (2009). One of the first examples that used unsupervised methods to extrapolate opinion measures from large-scale communication. It offers a good schematic example of how unsupervised methods work, and how it can be applied to several data sets to track aggregated sentiment dynamics.

Tools for Content Analysis

- R packages:
  - ReadMe: http://gking.harvard.edu/readme
  - TextMining: http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf
  - LDA Topic Modeling: http://www.cs.princeton.edu/~blei/topicmodeling.html
  - TextTools: http://www.rtexttools.com/
- Other software:
  - LexiCoder: http://www.lexicoder.com
  - SentiStrength: http://sentistrength.wlv.ac.uk
  - LIWC: http://www.liwc.net.

REFERENCES

Adrea, E. and Sebastiani, F. (2006). SentiWordNet: a publicly available lexical resource for opinion mining. Paper presented at the 5th Conference on Language Resources and Evaluation, Genoa, Italy.

Bearman, P. and Stovel, K. (2000). Becoming a Nazi: a model for narrative networks. *Poetics*, 27(1), 69–90.

Berinsky, A.J., Huber, G.A. and Lenz, G.S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3), 351–368. doi: 10.1093/pan/mpr057.

Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. New York: Oxford University Press.

Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G.A. and Reynar, J. (2008). Building a sentiment summarizer for local service reviews. WWW Workshop on NLP in the Information Explosion Era, Beijing.

Blei, D. (2013). Topic modeling and digital humanities. *Journal of Digital Humanities*, 2(1).

Blei, D., Ng, A. and Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning and Research*, 3, 993–1022.
Bollen, J., Mao, H. and Zeng, X.-J. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.
Borge-Holthoefer, J. and Arenas, A. (2010). Semantic networks: structure and dynamics. *Entropy*, 12(5), 1264–1302.
Bradley, M.M. and Lang, P.J. (1999). *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings*. Gainesville, FL.
Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. doi: 10.1023/A:1010933404324.
Brody, S. and Diakopoulos, N. (2011). Cooooooooooooooollllllllllllll!!!!!!!!!!!!!!: using word lengthening to detect sentiment in microblogs. Paper presented at the Conference on Empirical Methods in Natural Language Processing, Stroudsburg, PA.
Budge, I. (ed.). (2001). *Mapping Policy Preferences. Estimates for Parties, Electors and Governments 1945–1998*. Oxford: Oxford University Press.
Carley, K. and Palmquist, M. (1992). Extracting, representing, and analyzing mental models. *Social Forces*, 70(3), 601–637. doi: 10.2307/2579746.
Carvalho, P., Sarmento, L., Teixeira, J. and Silva, M.J. (2011). Liars and saviors in a sentiment annotated corpus of comments to political debates. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics.
Castillo, C., de Francisci Morales, G., Mendoza, M. and Khan, N. (2013). Says who? Automatic text-based content analysis of television news. Paper presented at the MNLP Workshop, ACM International Conference on Information and Knowledge Management, San Francisco, CA.
Connor, B.O., Balasubramanyan, R., Routledge, B.R. and Smith, N.A. (2010). From Tweets to polls: linking text sentiment to public opinion time series. *Fourth International AAAI Conference on Weblogs and Social Media*. Menlo Park, CA: AAAI.
Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. doi: 10.1023/A:1022627411411.
Däubler, T., Benoit, K., Mikhaylov, S. and Laver, M. (2012). Natural sentences as valid units for coded political texts. *British Journal of Political Science*, 42(04), 937–951.
Della Porta, D. and Diani, M. (2006). *Social Movements: An Introduction*, 2nd edn. London: Wiley-Blackwell.
Derks, D., Bos, A.E.R. and von Grumbkow, J. (2007). Emoticons and online message interpretation. *Social Science Computer Review*, 26, 379–388. doi: 10.1177/0894439307311611.
DiMaggio, P., Nag, M. and Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: application to newspaper coverage of government arts funding in the US. *Poetics*, 41(6), 570–606.
Dodds, P.S. and Danforth, C.M. (2009). Measuring the happiness of large-scale written expression: songs, blogs, and presidents. *Journal of Happiness Studies*, DOI: 10.1007/s10902-009-9150-9.
Dodds, P.S., Harris, K.D., Kloumann, I.M., Bliss, C.A. and Danforth, C.M. (2011). Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. *PloS ONE*, 6(12), e26752. doi: 10.1371/journal.pone.0026752.
Duggan, M. and Brenner, J. (2013). The demographics of social media users. http://www.pewinternet.org/Reports/2013/Social-media-users.aspx.
Dunteman, G.H. (1989). *Principal Components Analysis*. London: Sage.
Eshbaugh-Soha, M. (2010). The tone of local presidential news coverage. *Political Communication*, 27(2), 121–140. doi: 10.1080/10584600903502623.
Franzosi, R. (2004). *From Words to Numbers. Narrative, Data, and Social Science*. Cambridge: Cambridge University Press.
Gentzkow, M. and Shapiro, J.M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, 78(1), 35–71.
González-Bailón, S., Banchs, R.E. and Kaltenbrunner, A. (2012). Emotions, public opinion and US presidential approval rates: a 5-year analysis of online political discussions. *Human Communication Research*, 38, 121–143.

González-Bailón, S., Paltoglou, G. (2015). Signals of public opinion in online communication: a comparison of methods and data sources, *The Annals of the American Academy of Political and Social Science*, in press.
Grimmer, J. and King G., (2011). General purpose computer-assisted clustering and conceptualization. *Proceedings of the National Academy of Sciences*, 108(7), 2643–2650.
Grimmer, J. and Stewart, B. (2013). Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297.
Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. New York: Springer.
Hopkins, Daniel and King, Gary (2010). Extracting systematic social science meaning from text. *American Journal of Political Science*, 54(1), 229–247.
Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*. New York City: ACM Press.
Johnston, H. and Noakes, J.A. (eds) (2005). *Frames of Protest. Social Movements and the Framing Perspective*. Lanham, MD: Rowman & Littlefield.
Jurafsky, Dan and Martin, James (2009). *Speech and Natural Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.
Kamps, J., Marx, M., Mokken, R.J. and de Rijke, M. (2004). Using WordNet to measure semantic orientations of adjectives. Paper presented at the LREC. http://dblp.uni-trier.de/db/conf/lrec/lrec2004.html#KampsMMR04.
Kolstova, O. and Koltcov, S. (2013). Mapping the public agenda with topic modelling: the case of the Russian LiveJournal. *Policy and Internet*, 5(2), 207–227.
Krippendorff, K. (2013 [1980]). *Content Analysis. An Introduction to its Methodology*. Los Angeles, CA: Sage.
Krippendorff, K. and Bock, M.A. (eds) (2008). *The Content Analysis Reader*. Thousand Oaks, CA: Sage.
Laver, M., Benoit, K. and Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2), 311–331.
Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Chicago, IL: Morgan & Claypool.
Lorr, M., McNair, D.M., Heuchert, J.W.P. and Droppleman, L.F. (2003). *POMS: Profile of Mood States*. Toronto: MHS.
Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35–65.
MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1: 281–297. London: Cambridge University Press.
Maron, M.E. and Kuhns, J.L. (1960). On relevance, probabilistic indexing and information retrieval. *Journal of the ACM*, 7(3), 216–244. doi: 10.1145/321033.321035.
Mathiesen, J., Yde, P. and Jensen, M.H. (2012). Modular networks of word correlations on Twitter. *Scientific Reports*, 2.
Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K.J. (1990). Introduction to wordnet: an on-line lexical database. *International Journal of Lexicography*, 3(4), 235–244.
Murphy, F.C., Nimmo-Smith, I. and Lawrence, A.D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cognitive, Affective and Behavioral Neuroscience*, 3, 207–233.
Neuendorf, K. (2001). *The Content Analysis Guidebook*. London: Sage.
Osgood, C.E., Suci, G.J. and Tannenbaum, P.H. (1957). *The Measurement of Meaning*, Vol. 47. Urbana, IL, University of Illinois Press.
Paltoglou, G., Gobron, S., Skowron, M., Thelwall, M. and Thalmann, D. (2010). Sentiment analysis of informal textual communication in cyberspace. *Proc. ENGAGE*, 13–23.
Paltoglou, G. and Thelwall, M. (2012). Twitter, MySpace, Digg: unsupervised sentiment

analysis in social media. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(4):66: 1–66: 19.
Pennebaker, J.W., Booth, R.J. and Francis, M.E. (2001). *Linguistic Inquiry and Word Count: LIWC*. Mahwah, NJ: Erlbaum Publishers.
Porter, M. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130–137.
Quinn, K.M., Monroe, B.L., Colaresi, M., Crespin, M.H. and Radev, D.R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1), 209–228.
Spirling, A. (2012). US treaty making with American indians: institutional change and relative power, 1784–1911. *American Journal of Political Science*, 56(1), 84–97. doi: 10.1111/j.1540-5907.2011.00558.x.
Thelwall, M., Buckley, K. and Paltoglou, G. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544–2558.
Turney, P.D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Paper presented at the *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Philadelphia, PA.
Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., . . . Patwardhan, S. (2005). OpinionFinder: a system for subjectivity analysis. *Proceedings of HLT/EMNLP on Interactive Demonstrations*, Vancouver, British Columbia, Canada.
Yano, T., Cohen, W.W. and Smith, N.A. (2009). Predicting response to political blog posts with topic models. *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL* (pp. 477–485). Association for Computational Linguistics.
Young, L. and Soroka, S. (2012). Affective news: the automated coding of sentiment in political texts. *Political Communication*, 29, 205–231. doi: 10.1080/10584609.2012.671234.

25. On the cutting edge of Big Data: digital politics research in the social computing literature

Deen Freelon

Most of this volume's chapters review studies rooted in political science, communication, and closely related disciplines. Indeed, many reference a small clique of foundational authors in agreement and/or disagreement, including Castells, Benkler, Hindman, Jenkins, Morozov, and Shirky. In the current chapter I diverge from this norm to examine a body of literature only rarely acknowledged by mainstream digital politics scholarship. This literature contains politically relevant research by computer scientists and information scientists and is published under a variety of disciplinary labels, but will be referred to here as 'social computing research'. As its name implies, social computing research includes any aspect of human behavior involving both digital technology and more than one person (Parameswaran and Whinston, 2007; Wang et al., 2007). Politics accounts for a small but thriving subset of this literature, which also encompasses health, business, economics, entertainment, artificial intelligence, and disaster response, among other topics.

Social computing research on politics holds relevance for scholars of digital politics and political communication for two related reasons, one methodological and the other theoretical. First, social computing researchers have for many years led the vanguard in computational and 'Big Data' methods (sometimes in combination with other methods), in which the disciplines of political science and communication have both expressed great interest of late.[1] Reviewing how social computing researchers have applied such methods to politically relevant datasets will help digital politics readers to consider how the methods could be applied to their own research. The field's methods and findings also hold a number of theoretical implications, but its researchers devote only sporadic attention to such concerns. For the benefit of those with a more theoretical scholarly orientation, and perhaps also for some social computing researchers with social science leanings, I explore major theoretical trends in the literature. I conclude with suggestions for future research, focusing on how digital politics researchers can best adapt the insights of social computing research to their own ends.

Before proceeding to these sections, however, it is necessary to more thoroughly describe social computing and its goals, which differ in key ways from those of the social science mainstream. The following section is devoted to this task.

SOCIAL COMPUTING: A BRIEF INTRODUCTION

A caveat before I begin: this section is written from the perspective of one who was trained in and still operates within a social science-based research orientation that emphasizes abstract theory as a guide and justification for empirical work (Fink and Gantz, 1996). My participation in social computing research up to this point in my career has been minimal. Accordingly, the description of social computing I offer here is intended as an introduction for those of a similar scholarly orientation to me, which I imagine will include many if not most of this volume's audience.

In a widely cited overview article, Wang et al. (2007) define social computing as 'computational facilitation of social studies and human social dynamics as well as the design and use of ICT [information and communication technology] technologies that consider social context' (p. 79). A similar characterization by Parameswaran and Whinston (2007) establishes social computing as a highly ubiquitous activity of study:

> Social computing shifts computing to the edges of the network, and empower [*sic*] individual users with relatively low technological sophistication in using the Web to manifest their creativity, engage in social interaction, contribute their expertise, share content, collectively build new tools, disseminate information and propaganda, and assimilate collective bargaining power. (p. 763)

Both of the above quotes emphasize the two essential elements of social computing: digital tools ('computing', broadly construed) and social interaction. Of course, researchers in communication, sociology, anthropology, and other social-scientific disciplines have explored topics such as 'computer-mediated communication' and 'cyberculture' for decades. This similarity in subject matter invites the question of how social computing research differs from approaches with which we are more familiar.

The main difference between social computing research and research traditions grounded in social science is as paradigmatic as that between social science and critical theory (Fink and Gantz, 1996; Potter et al., 1993). Whereas social science's goals are to explain empirical outcomes while promulgating theory, and critical theory's is to foment social change, social computing research is devoted to the development of new techniques for organizing, analyzing, and improving the user experience and

output of social computing software. As such, social computing studies are usually published in highly technical articles that focus on methods, analysis, and evaluation at the expense of what we would consider 'theory' (Freelon, 2014). The call for papers for the 2014 conference on Computer-Supported Collaborative Work (CSCW), a prominent social computing publication venue, expresses this in its introduction: 'We invite submissions that detail existing practices, inform the design or deployment of systems, or introduce novel systems, interaction techniques, or algorithms' (CSCW, n.d.). Further evidence for this claim can be seen in the strong presence of employees of well-known technology companies such as Google, Microsoft, and Yahoo on major social computing conferences' program committees. Of course, theory is not always entirely absent: some articles include a few theoretical references of relevance to the project at hand, but the discussions tend to be much shorter than in most social science fields. And articles are often accepted without referencing any social science theories at all.

In addition to downplaying theory, social computing research relies heavily on computational methods such as social network analysis, machine learning, computational linguistics, and algorithmic preprocessing of raw web data. These methods are common in computer science and information science, disciplines which many (though by no means all) social computing researchers call home. Programming serves at least two major purposes in social computing: (1) to develop and improve digital platforms for social interaction; and (2) to evaluate their performance efficiently and at scale. Qualitative methods such as ethnography and in-depth interviews are occasionally seen, often as part of a multi-method approach with one or more computational methods. However, such studies are relatively rare, as the next section will demonstrate. The field places the highest value on research techniques and metrics that can be implemented algorithmically. The ability to visualize quantitative results in intuitive and innovative ways is also highly prized.

The final characteristic of social computing research of relevance to the digital politics researcher may seem rather obvious: the field is not principally concerned with politics per se, but rather with social computer use. In other words, social computing researchers typically analyze political cases to make broader points about social computing systems and affordances rather than about politics. Matters of system development and algorithm optimization almost always come first, and broader implications for politics are discussed secondarily if at all. Consequently, the results sections of social computing research papers often leave many implications of theoretical interest unexplored. Later in this chapter I will attempt to reclaim some of these implications in order to clarify their value for students of political science and communication. But first, I will examine in detail the most common methods social computing researchers employ.

METHODS IN SOCIAL COMPUTING RESEARCH ON POLITICS

Social computing research is sometimes published in journals, but many of the most relevant studies for our purposes are published in the archived proceedings of prominent conferences in computer science, information science, and human–computer interaction. Haphazardly selecting papers from these conferences would bias my discussion, so instead I chose them using a systematic and replicable method. First, I focused on conferences and publications sponsored by the Association of Computing Machinery (ACM) and the Institute of Electrical and Electronic Engineers (IEEE), the premier professional organizations in computer science and computer engineering. Using Google Scholar, I searched for the term 'political' exclusively within such outlets. I then ranked the results in descending order by number of citations in order to capture the most widely referenced articles. Being interested only in articles that address politics as a central concern, I qualitatively assessed the most-cited items in each list, flagging articles that empirically analyze political messages, opinions, attitudes, and/or other content as their main focus. (Thus, for example, I excluded articles that analyzed political content as only one of three or more other content categories.) I continued this process until I had flagged 20 articles within each group, for a total of 40 articles (see Table 25.1). These form the basis of the discussions in this section and the next.

After finalizing the sample, I informally classified the articles based on the methods they employed.[2] Three methodological categories were used: traditional quantitative, qualitative, and computational. The traditional quantitative category includes long-established quantitative methods in social science such as surveys, experiments, content analysis, and statistical analysis of secondary data. The qualitative category includes in-depth interviews, field observations, and close readings of texts, among others. The computational category includes any method that entailed the creation of original source code whose purpose was to collect, preprocess, or analyze data. The rest of the chapter will focus mainly on this last category, as the others are much more familiar to scholars of digital politics.

Unsurprisingly, the most prevalent methodological category throughout the sample was by far computational (29/40, 72.5 percent), followed by traditional quantitative (19/40, 47.5 percent) and then qualitative (5/40, 12.5 percent). All but one of the qualitative studies used a mixed

Table 25.1 The methods of 40 highly cited social computing research papers

Authors	Title	Traditional quantitative methods	Qualitative methods	Computational methods
Adamic and Glance, 2005	The political blogosphere and the 2004 US election: divided they blog			x
Awadallah et al., 2010	Language-model-based pro/con classification of political text			x
Baumer et al., 2009	MetaViz: visualizing computationally identified metaphors in political blogs	x		x
Baumer et al., 2010	America is like Metamucil: fostering critical and creative thinking about metaphor in political blogs	x		x
Bélanger and Carter, 2010	The digital divide and internet voting acceptance	x		
Conover et al., 2011	Predicting the political alignment of Twitter users	x		x
DeNardis and Tam, 2007	Interoperability and democracy: a political basis for open document standards		x	
Diakopoulos and Shamma, 2010	Characterizing debate performance via aggregated Twitter sentiment	x		
Diaz-Aviles et al., 2012	Taking the pulse of political emotions in Latin America based on social web streams			x
Fang et al., 2010	Mining contrastive opinions on political texts using cross-perspective topic model			x

Table 25.1 (continued)

Authors	Title	Traditional quantitative methods	Qualitative methods	Computational methods
Fisher et al., 2010	Egovernment services use and impact through public libraries: preliminary findings from a national study of public access computing in public libraries	x	x	
Furuholt and Wahid, 2008	E-government challenges and the role of political leadership in Indonesia: the case of Sragen		x	
Garcia et al., 2010	Political polarization and popularity in online participatory media: an integrated approach			x
Golbeck and Hansen, 2011	Computing political preference among Twitter followers			x
Gulati et al., 2012	Understanding the impact of political structure, governance and public policy on e-government	x		
Hong and Nadler, 2011	Does the early bird move the polls?: The use of the social media tool 'Twitter' by US politicians and its impact on public opinion			x
Jiang and Argamon, 2008	Exploiting subjectivity analysis in blogs to improve political leaning categorization	x		x
Jürgens et al., 2011	Small worlds with a difference: new gatekeepers and the filtering of political information on Twitter			x
Kannabiran and Petersen, 2010	Politics at the interface: a Foucauldian power analysis		x	
Kaschesky and Riedl, 2011	Tracing opinion-formation on political issues on the Internet: a model and methodology for qualitative analysis and results	x	x	
Kaschesky et al., 2011	Opinion mining in social media: modeling, simulating, and visualizing political opinion formation in the web			x
Kim et al., 2007	Toward a model of political participation among young adults: the role of local groups and ICT use	x		
Kim et al., 2012	Automatic detection of conflicts in spoken conversations: ratings and analysis of broadcast political debates	x		x
Mascaro et al., 2012	Tweet recall: examining real-time civic discourse on Twitter	x		x
Munson and Resnick, 2010	Presenting diverse political opinions: how and how much	x		x
Nahon and Hemsley, 2011	Democracy. com: a tale of political blogs and content			x
Park et al., 2011	The politics of comments: predicting political orientation of news stories with commenters' sentiment patterns	x		x
Ratkiewicz et al., 2011	Truthy: mapping the spread of astroturf in microblog streams			x
Sarmento et al., 2009	Automatic creation of a reference corpus for political opinion mining in user-generated content	x		x
Singh et al., 2010	Mining the blogosphere from a socio-political perspective			x
Skoric et al., 2012	Tweets and votes: a study of the 2011 Singapore general election			x
Stieglitz and Dang-Xuan, 2012	Political communication and influence through microblogging: an empirical analysis of sentiment in Twitter messages and retweet behavior			x

Table 25.1 (continued)

Authors	Title	Traditional quantitative methods	Qualitative methods	Computational methods
Ulicny et al., 2010	Metrics for monitoring a social-political blogosphere: a Malaysian case study			x
Vallina-Rodriguez et al., 2012	Los twindignados: the rise of the indignados movement on Twitter	x		x
Wallsten, 2011	Beyond agenda setting: the role of political blogs as sources in newspaper coverage of government	x		x
Weber et al., 2012	Mining web query logs to analyze political issues			x
Wei and Yan, 2010	Knowledge production and political participation: reconsidering the knowledge gap theory in the Web 2. environment	x		
Younus et al., 2011	What do the average Twitterers say: a Twitter model for public opinion analysis in the face of major political events	x		
Zhang et al., 2009	Gender difference analysis of political web forums: an experiment on an international Islamic women's forum			x
Total		19	5	29

methodology which combined either multiple qualitative methods or one qualitative method with one of the other types. The authors used a variety of traditional quantitative methods, with surveys and content analyses being the two most popular. A large minority of studies employing traditional quantitative methods complemented them with computational methods (8/19, 42.1 percent). Based on this highly cited sample, it would seem that social computing publication venues welcome political research that is methodologically traditional, although computational methods are more common.

I classified an extremely heterogeneous collection of methods as 'computational' in accordance with the operational definition given above. These fall into three general subcategories: data collection, preprocessing, and analysis.

Data Collection

Nearly all major social media services, including Twitter, Facebook, and YouTube, offer application programming interfaces (APIs) through which large amounts of data can be harvested computationally. A variety of options exist for doing so, from simple desktop-based solutions such as NodeXL to Linux-based servers such as 140dev and yourTwapperKeeper to writing a script in the programming language of one's choice. Some articles that analyzed social media content briefly described their data collection process, including such details as the language and specific API used (Mascaro et al., 2012; Ratkiewicz et al., 2011; Skoric et al., 2012), while others did not (Diaz-Aviles et al., 2012; Garcia et al., 2009; Golbeck and Hansen, 2011; Jürgens et al., 2011; Vallina-Rodriguez et al., 2012). Interfacing with APIs to extract data is evidently such a routine activity in social computing research that documenting it in detail is optional. Studies that examined content from sources without APIs – blogs, for example – usually used their own custom web-scraping scripts (Adamic and Glance, 2005; Nahon and Hemsley, 2011; Ulicny et al., 2010).

For researchers in disciplines such as political science and communication that are relatively new to computational methods, this lack of detail on data collection methods is unfortunate. I do not intend to imply that it is the responsibility of social computing researchers to educate outsiders on the elementary aspects of social media data collection, but only observe that beginners will not learn much about how to collect social media data from articles in the field. Textbooks on social media analysis (Leetaru, 2012; Russell, 2013) are more helpful in this regard, but their utility will inevitably decrease with time due to the rapid developmental pace of social media platforms. Some enterprising political science and communication researchers will be able to teach themselves effectively using such resources, but until computational methods become a disciplinary priority, social scientists' ability to collect and analyze social media data will remain limited.

Preprocessing

Preprocessing encompasses a miscellany of techniques to convert raw text and other content collected from the web into research-grade data suitable for quantitative and qualitative analysis. Examples include manipulating social media posts into formats suitable for calculating descriptive statistics (Mascaro et al., 2012), social network analysis (Adamic and Glance, 2005; Conover et al., 2011; Jürgens et al., 2011; Ratkiewicz et al., 2011), simple time-series plots (Vallina-Rodriguez et al., 2012), statistical analysis incorporating non-social media data (Golbeck and Hansen, 2011; Skoric et al., 2012), automated content analysis (Diaz-Aviles et al., 2012; Stieglitz and Dang-Xuan, 2012), and analysis of metadata such as 'likes' or star ratings (Garcia et al., 2012). Like data collection, computational preprocessing requires programming skills by definition, but while the former is a rote task that rarely changes substantially between projects, the latter is completely open-ended. Indeed, creativity in preprocessing determines the kinds of analyses that can be applied to one's data; as such it is more akin to an art than a science.

The articles in the sample exemplify the dizzying range of choices researchers face when preprocessing their data. For example, in using social network methods to analyze relationships between social media users, a preprocessing script may count '@-mentions', retweets relationships, and/or follow relationships as tie indicators, among other features (Conover et al., 2011; Golbeck and Hansen, 2011; Jürgens et al., 2011; Ratkiewicz et al., 2011). The findings of the ensuing social network analysis will obviously differ based on which tie indicators are used. Similarly, most types of automated text analysis require some preprocessing to allow the algorithms to output intelligible results. In a sentiment analysis of political tweets, Stieglitz and Dang-Xuan (2012) imported Twitter-specific jargon and emoticons from their dataset into a dictionary of positively and negatively valenced terms which they used to classify tweets as positive or negative in tone. Using a similar dictionary-based technique, Diaz-Aviles et al. (2012) assembled 'profiles' of tweets and blog posts mentioning 18 Latin American presidents to analyze the online sentiments associated with each. More sophisticated automated techniques such as supervised and unsupervised learning require even more extensive preprocessing. After removing very common words that contain little informational value (called stopwords), raw documents are often disaggregated into clusters of one-, two-, or three-word phrases called 'n-grams' which learning algorithms analyze directly. The choice of which stopwords, types of n-grams, and algorithms to use will all influence the end results. For example, Fang et al. (2012) attempted to quantify the ideological distance between differing political opinions in newspapers and in statements by US senators. To prepare their data for analysis, they used verbs, adjectives, and adverbs as opinion descriptors and retained certain opinion-relevant terms such as 'should' and 'must' that would otherwise be considered stopwords. In a very different research context, Zhang et al. (2009) extracted unigrams and bigrams from an Islamic women's web forum to examine gender differences in content and writing style using supervised learning.

Analysis

Programming usually plays some part in the analysis phase of studies that use computational methods. Complex and creative visualizations produced using specialized code libraries often appear in the results. Most of these tools are applied to communication content – tweets, blog posts, video transcripts, news articles – that do not require direct interaction with participants. The most common computational analytical methods for texts among the sample are dictionary (or corpus)-based approaches, unsupervised learning, supervised learning, and network analysis.[3]

Dictionary-based approaches use either predefined or custom word collections representing different concepts to classify texts. For example, a dictionary of positive emotions might include terms such as 'love', 'awesome', 'happy', and 'best', and the software might measure positivity as the number of such terms within each text. This technique was used in several articles to analyze social media users' positive and negative feelings toward political issues and politicians (Diaz-Aviles et al., 2012; Garcia et al., 2012; Sarmento et al., 2009; Stieglitz and Dang-Xuan, 2012). Unsupervised learning approaches attempt to detect latent structure in texts inductively and automatically; one of its applications to politics research is the identification of topics mentioned in political texts (Fang et al., 2012). Supervised learning, in contrast, is a deductive method whose goal is to identify pre-established content categories automatically. It often begins with a traditional content analysis, the results of which the algorithm uses as exemplars to classify previously unexamined texts. Several social computing research teams have used supervised learning to predict the political leanings of social media users (Conover et al., 2011; Jiang and Argamon, 2008; Park et al., 2011). Finally, network methods have proven themselves quite versatile, with applications in the study of political spam (Ratkiewicz et al., 2011), communication patterns among political bloggers (Adamic and Glance, 2005; Nahon and Hemsley, 2011; Ulicny et al., 2010), and political gatekeeping in social media (Jürgens et al., 2011).

This very brief survey was intended to highlight some of the ways in which computational methods have been used to study political topics. The kinds of research questions social computing scholars pursue using these methods are limited by their field-specific concerns; thus, there are many opportunities for innovative work by enterprising scholars in other fields with different concerns. The following section substantiates this point more fully.

THEORY IN SOCIAL COMPUTING RESEARCH ON POLITICS

There is a great deal of variation in how social computing research addresses theoretical concerns. Two broad approaches to theory are apparent in the current sample. The first is an *explicit* approach that closely resembles the norm in social science: relevant theoretical contributions from prior research are explored in an in-depth literature review, and then empirical research questions and/or hypotheses are derived from them. The depth of these literature reviews varies widely, as we shall see. The second approach is *implicit* in that theoretical concerns about politics are not discussed at all, but the methods or findings could be integrated into theory-based research by innovative authors. This section will first discuss the theoretical implications of explicitly theoretical papers, and then offer suggestions as to how implicitly theoretical work can inform existing theoretical traditions.

Explicitly Theoretical Work

Social computing research that explicitly incorporates theory does so in a similar fashion to social science. In fact, some such papers are comparable in their theoretical rigor to traditional political science and communication fare (for example, Munson and Resnick, 2010; Nahon and Hemsley, 2011; Wei and Yan, 2010). However, most mention theoretical concerns only in passing: these will typically cite a small number of classic theoretical pieces without exploring much or any of the recent empirical work

they have inspired (for example, Adamic and Glance, 2005; Baumer et al., 2009; Kaschesky and Riedl, 2011; Weber et al., 2012). I do not intend to fault the less theoretical pieces here; as explained earlier, social computing and social science have different goals. But observing trends in how the former field uses prior research is important for social scientists who may be interested in building on its studies or in submitting papers to social computing publication venues.

Only one cluster of theories attracted attention from more than one or two papers: online political polarization, homophily, and selective exposure. The research on this topic fell into two categories: studies of online content and evaluations of design interventions. The content-based research analyzed text and metadata from YouTube, the American political blogosphere, Twitter, online newspaper comments, and Yahoo!'s search query logs. Most of these studies found clear evidence of online homophily: for example, that the blogosphere is divided in terms of hyperlinking patterns (Adamic and Glance, 2005); liberal blogs tend to link primarily to liberal election videos and *mutatis mutandis* for conservatives (Nahon and Hemsley, 2011); the Twitter followers of media outlets tend to skew liberal or conservative (Golbeck and Hansen, 2011); and liberals and conservatives tend to use ideologically distinctive queries in search engines (Weber et al., 2012). The design intervention studies evaluated the effects of systems designed to promote exposure to opinion-challenging content (Munson and Resnick, 2010) and critical thinking about politics (Baumer et al., 2009; Baumer et al., 2010). Unsurprisingly, all three of these studies reported some degree of success in their stated goals.

The remaining explicitly theoretical pieces covered a hodgepodge of theoretical concerns. Kaschesky and Riedl (2011) justified their research examining how opinions form and diffuse online partly by reference to the public sphere and deliberation. Along somewhat similar lines, Wei and Yan (2010) grounded their survey-based study of online knowledge production in the knowledge gap and political participation literatures. Bélanger and Carter (2010) invoked the digital divide in a study of US attitudes toward Internet voting, finding that younger and more affluent citizens are more favorably disposed toward it. DeNardis and Tam (2007) offered a legalistic analysis of global ICT standards based on democratic theory, ultimately recommending open document formats for public institutions. In the sole study grounded in critical theory, Kannabiran and Petersen (2010) presented a Foucauldian reading of Facebook's interface.

Implicitly Theoretical Work

Most of the studies reviewed for this chapter did not discuss theory in any substantial way (although some of these cited social science papers to discuss their empirical results). A few lacked literature reviews altogether (Jiang and Argamon, 2008; Jürgens et al., 2011; Ratkiewicz et al., 2011). Those that included them tended to focus on previous studies' methodological efficiency and range of application, and they generally framed their contributions in those terms as well (Diakopoulos and Shamma, 2010; Diaz-Aviles et al., 2012; Fang et al., 2012; Garcia et al., 2012; Kaschesky et al., 2011; Kim et al., 2012; Sarmento et al., 2009; Skoric et al., 2012; Younus et al., 2011; Zhang et al., 2009). In a representative example, Awadallah et al. (2010) presented a new method for classifying political debate arguments as pro or con. Much previous work in the area had at that point been context-independent; for example, judging a statement as inherently positive or negative, whereas pro/con judgments depend upon how the debate position is phrased. Further, previous work had also required manually classified training data, which is time-consuming and expensive. Awadallah's approach was both context-sensitive and fully automatic, which constitute substantive contributions in the social computing research tradition.

Perhaps the best way to demonstrate the value of implicitly theoretical work is to describe its attempted goals, many of which fall into one or more of three categories: classification, forecasting, and description. Classification, the largest category, includes studies that aim to fully or partially automate the process of labeling digital content (mostly but not exclusively text). Some of the classification tasks in this sample include labeling political texts as positive or negative (which is also known as sentiment analysis) (Diakopoulos and Shamma, 2010; Diaz-Aviles et al., 2012; Garcia et al., 2012; Sarmento et al., 2009), pro or con (Awadallah et al., 2010), subjective or objective (Younus et al., 2011), and liberal or conservative (Conover et al., 2011; Fang et al., 2012; Golbeck and Hansen, 2011; Jiang and Argamon, 2008). Forecasting studies seek to predict patterns or outcomes in the digital realm or offline; examples include elections (Skoric et al., 2012), public opinion polls (Diaz-Aviles et al., 2012; Hong and Nadler, 2011), and the diffusion of political opinions online (Kaschesky and Riedl, 2011; Kaschesky et al., 2011). Descriptive studies are similar to their counterparts in social science except that they use very little or no theory (and sometimes no prior research at all) to guide them. As a result, their attempts to discover how platforms such as Twitter were used in particular contexts vary widely in their methodological specifics (Mascaro et al., 2012; Vallina-Rodriguez et al., 2012).

Each of these categories is implicitly theoretical in its own way. Classification studies do not go quite far enough to qualify as social science; their goal is typically to optimize algorithmic performance rather than to contribute to theory. From a social science perspective they resemble extended method sections, full of details on each of the classification task's steps and the results of various evaluation metrics. This simile clarifies the theoretical implications of advanced classification studies to social science: any theory that requires classification could potentially make use of their methodological innovations. For example, the ability to classify political ideology algorithmically could enable theoretically-oriented studies of political polarization and deliberation to analyze sample or population sizes in the millions. Similarly, an automated system for quantifying political sentiment in social media posts could help researchers better theorize how voters react to targeted political messages outside of experimental settings (for more on the uses of sentiment analysis in digital politics research, see Petchler and González-Bailón, Chapter 24 in this volume). Forecasting is more the province of natural scientists and economists than most of social science, which is more concerned with explanation.[4] That said, we should recall that forecasting encompasses within it correlation and time precedence, which are two of Babbie's (2012) three essential components of causation. The remaining component, the elimination of potential alternative causes, then becomes the task of the social scientist. In the rush to build models that can predict elections based on user-generated data (for example), it is the social scientist rather than the social computing researcher who will be interested in why the model works. Finally, most descriptive studies would not pass muster in most social science journals because of their long-standing bias against atheoretical work. Nevertheless, they can still offer the social scientist a sense of the methodological possibilities afforded by new social computing platforms, which could then be incorporated into research questions and/or hypotheses that build theory.

CONCLUSION AND FUTURE WORK

As I have shown, social computing research has produced much of interest to the digital politics researcher. The field has employed computational methods and Big Data since the 1990s, and still conducts much of the cutting-edge research in these areas. In contrast, political science and communication are still very firmly invested in their traditional methods, which are not always optimally suited for analyzing digital data.

Engagement with the best social computing research studies has been and will continue to be essential for all social scientists interested in applying computational methods in their home disciplines. The field's theoretical contributions are not always as obvious, but with a bit of work, students of digital politics will be able to profitably draw upon them for inspiration.

I close this chapter with two general recommendations for social scientists who find this sort of work valuable. The first is simply to learn a programming language suitable for manipulating and analyzing large datasets. While researchers can conduct a few descriptive analyses on large datasets without knowing how to program, most research-grade operations require the ability to work directly with code. Collaborating with social computing researchers may work well for some projects, but as we have seen, they have different standards for what constitutes a contribution (and corresponding publication incentives). Moreover, social scientists can recognize theoretically relevant patterns in data that computer scientists cannot; thus it greatly benefits the former to know how to explore large-scale datasets firsthand. (Imagine having to rely on statisticians for all your statistics!) For the beginning computational researcher I recommend learning the Python programming language, both because it offers a number of libraries and modules specifically for collecting, preprocessing, and analyzing data; and also because its growing popularity in academic circles offers critical support for new learners. The statistical language and programming environment R offers a wider variety of statistical models than Python, but also has a steeper learning curve.

As computational research becomes more accepted in the disciplines in which digital politics research is conducted, graduate faculties should strongly consider how best to teach its methods to their students. Very few communication departments in the USA currently teach computational methods in any systematic fashion, and I suspect the situation is not substantially different in political science. Few US communication departments have any experts in computational methods on faculty, and fewer still have more than one. Some of these experts, such as Benjamin Mako Hill (University of Washington) and Sandra Gonzalez-Bailón (University of Pennsylvania), received their graduate departments in fields other than communication. Others, such as Drew Margolin (now at Cornell) and me, trained in communication departments that do not emphasize computational methods as a core teaching strength. In light of the paramount importance and ubiquity of digital communication data, I submit that computational methods should become one of communication's premier research methods, on par with survey methods, content analysis, experiments, and in-depth interviews. And just as every doctoral student need not learn how to conduct and analyze surveys, not everyone needs to learn computational methods; but it ought to be one of communication's major areas of methodological specialization. A detailed explanation of how to achieve this outcome lies beyond the scope of this chapter, but at a minimum, committed departments will need to thoroughly revise their hiring practices, tenure guidelines, graduate curricula, and departmental resources (including purchasing appropriate hardware, software, and data subscriptions).

My second recommendation pertains to the construct validity of digital traces. Construct validity is the extent to which an operationalized metric actually measures the underlying concept it is intended to measure (Babbie, 2012). As I have documented elsewhere (Freelon, 2014), social computing research studies do not always amply demonstrate the construct validity of the traces they use as metrics. To take an example from the current sample, Ulicny et al. (2010) purport to measure four concepts of academic and practical relevance in the Malaysian blogosphere: relevance, specificity, timeliness, and credibility. Without any reference to prior literature, they define these concepts in terms of manifest digital traces, including use of a real name, network authority, number of comments, and number of unique nouns, among others. Not only are these metrics biased in favor of what can be collected and measured easily, but there is no discussion of whether the metrics are comprehensive, and if not, which aspects of the underlying concepts might be omitted. While a lack of attention to construct validity is by no means universal in social computing research, it is common (Fang et al., 2012; Garcia et al., 2009; Jürgens et al., 2011; Mascaro et al., 2012; Younus et al., 2011).

Social science research on politics is ultimately concerned with abstract concepts such as power, influence, preference, ideology, and homophily, among many others. Traces such as retweets, Facebook 'likes', social media follow relationships, and hyperlink patterns are only interesting inasmuch as they faithfully relate to such concepts. Yet just as we should avoid studying traces for their own sake, so should we also refrain from simply assuming that retweets are always endorsements, hyperlinks always signify authority, and 'likes' always imply approval. Credible arguments for these positions should be articulated and substantiated. In some cases, it will be possible to make logical arguments on the basis of a trace's inherent properties, as in the observation that retweets represent peer-to-peer information propagation. But whenever possible, a trace's imputed meaning should draw on empirical observation. Close qualitative observation of how traces are used can help to fulfill this purpose (Boyd et al., 2010).

The rise of computational techniques in social science has barely begun, and digital politics scholars (myself included) still have much to learn. Social computing researchers offer some of the most methodologically sophisticated work currently available, and many of them are interested in very familiar subject matter. For these reasons, we would do well to learn what we can from them.

NOTES

1. Consider for example the panels held on the topic of Big Data and/or computational methods at the 2013 annual meetings of the International Communication Association (ICA) and the Association of Educators in Journalism and Mass Communication (AEJMC), as well as the conference theme of the 2014 annual meeting of the American Political Science Association (APSA); 'Politics After the Digital Revolution'.
2. I chose not to conduct a formal content analysis here mainly due to the great diversity of methods comprising the 'computational' category, which proved difficult for a non-expert coder to identify consistently.
3. Readers interested in more in-depth discussions of these methods than I offer here are recommended to consult Graesser et al. (2010) and Petchler and González-Bailón (Chapter 24 in this volume).
4. For more on the differences between scientific prediction and explanation, see Shmueli and Koppius (2011).

FURTHER READING

On Social Computing

Parameswaran, M. and Whinston, A.B. (2007). Social computing: an overview. *Communications of the Association for Information Systems*, 19(37), 762–780.
Tian, Y., Srivastava, J., Huang, T. and Contractor, N. (2010). Social multimedia computing. *Computer*, 43(8): 27–36.
Wang, F.-Y., Carley, K.M., Zeng D. and Mao, W. (2007). Social computing: from social informatics to social intelligence. *Intelligent Systems, IEEE*, 22(2), 79–83.

Learning Computational Methods

Cogliati, J., Aikens, M., Morante, K., Cogliati, E., Brown, J.A., Oppengaard, J. and Hell, B. (2013). *Non-Programmer's Tutorial for Python 3*. Wikibooks. Available at http://en.wikibooks.org/wiki/Non-Programmer's_Tutorial_for_Python_3.
Russell, M.A. (2013). *Mining the Social Web*. Sebastopol, CA: O'Reilly Media.
Stanton, J. (2013). *Introduction to Data Science*. Available at http://jsresearch.net/.

REFERENCES

Adamic, L.A. and Glance, N. (2005). The political blogosphere and the 2004 US election: divided they blog. In Proceedings of the 3rd International Workshop on Link Discovery (pp. 36–43).
Awadallah, R., Ramanath, M. and Weikum, G. (2010). Language-model-based pro/con classification of political text. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 747–748). New York: ACM. doi:10.1145/1835449.1835596.
Babbie, E. (2012). The Practice of Social Research. Belmont, CA: Cengage Learning.
Baumer, E.P.S., Sinclair, J., Hubin, D. and Tomlinson, B. (2009). metaViz: Visualizing Computationally Identified Metaphors in Political Blogs. In International Conference on Computational Science and Engineering, 2009. CSE '09, Vol. 4 (pp. 389–394). doi:10.1109/CSE.2009.482.
Baumer, E.P.S., Sinclair, J. and Tomlinson, B. (2010). America is like Metamucil: fostering critical and creative thinking about metaphor in political blogs. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1437–1446). New York: ACM. doi:10.1145/1753326.1753541.
Bélanger, F. and Carter, L. (2010). The digital divide and internet voting acceptance. In Fourth International Conference on Digital Society, 2010. ICDS'10 (pp. 307–310). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5432779.
Boyd, D., Golder, S. and Lotan, G. (2010). Tweet, tweet, retweet: conversational aspects of retweeting on twitter. In 43rd Hawaii International Conference on System Sciences (HICSS) (pp. 1–10).
Conover, M.D., Goncalves, B., Ratkiewicz, J., Flammini, A. and Menczer, F. (2011). Predicting the political alignment of Twitter users. In Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on Social Computing (SocialCom) (pp. 192–199). doi:10.1109/PASSAT/SocialCom.2011.34.
CSCW (n.d.). Call for participation papers. ACM. Retrieved from http://cscw.acm.org/participation_papers.html.
DeNardis, L. and Tam, E. (2007). Interoperability and democracy: a political basis for open document standards. In 5th International Conference on Standardization and Innovation in Information Technology, 2007. SIIT 2007 (pp. 171–180). doi:10.1109/SIIT.2007.4629327.
Diakopoulos, N.A. and Shamma, D.A. (2010). Characterizing debate performance via aggregated Twitter sentiment. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1195–1198). New York: ACM. doi:10.1145/1753326.1753504.
Diaz-Aviles, E., Orellana-Rodriguez, C. and Nejdl, W. (2012). Taking the pulse of political emotions in Latin America based on social web streams. In 2012 Eighth Latin American Web Congress (LA-WEB) (pp. 40–47). doi:10.1109/LA-WEB.2012.9.
Fang, Y., Si, L., Somasundaram, N. and Yu, Z. (2012). Mining contrastive opinions on political texts using cross-perspective topic model. In Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (pp. 63–72). Retrieved from http://dl.acm.org/citation.cfm?id=2124306.
Fink, E.J. and Gantz, W. (1996). A content analysis of three mass communication research traditions: social science, interpretive studies, and critical analysis. Journalism and Mass Communication Quarterly, 73(1), 114–134. doi:10.1177/107769909607300111.
Fisher, K.E., Becker, S. and Crandall, M. (2010, January). eGovernment services use and impact through public libraries: preliminary findings from a national study of public access computing in public libraries (pp. 1–10). IEEE.
Freelon, D. (2014). On the interpretation of digital trace data in communication and social computing research. Journal of Broadcasting and Electronic Media, 58(1), 59–75.
Furuholt, B. and Wahid, F. (2008, January). E-government challenges and the role of political leadership in Indonesia: the case of Sragen. In Proceedings of the 41st Annual Hawaii International Conference on System Sciences, IEEE.

Garcia, A., Standlee, A., Beckhoff, J. and Cui, Y. (2009). Ethnographic approaches to the Internet and computer-mediated communication. Journal of Contemporary Ethnography, 38(1), 52–84.
Garcia, D., Mendez, F., Serdült, U. and Schweitzer, F. (2012). Political polarization and popularity in online participatory media: an integrated approach. In Proceedings of the first Edition Workshop on Politics, Elections and Data (pp. 3–10). Retrieved from http://dl.acm.org/citation.cfm?id=2389665.
Golbeck, J. and Hansen, D. (2011). Computing political preference among Twitter followers. In Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (pp. 1105–1108). New York: ACM. doi:10.1145/1978942.1979106.
Graesser, A.C., McNamara, D.S. and Louwerse, M.M. (2010). Methods of automated text analysis. In Kamil, M.L., Pearson, D., Moje, E.B. and Afflerbach, P. (eds), Handbook of Reading Research, Vol. 4 (pp. 34–53). New York: Routledge.
Gulati, G.J., Yates, D.J. and Williams, C.B. (2012, January). Understanding the impact of political structure, governance and public policy on e-government. In 2012 45th Hawaii International Conference on System Sciences (HICSS) (pp. 2541–2550). IEEE.
Hong, S. and Nadler, D. (2011). Does the early bird move the polls?: The use of the social media tool 'Twitter' by US politicians and its impact on public opinion. In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times (pp. 182–186). New York: ACM. doi:10.1145/2037556.2037583.
Jiang, M. and Argamon, S. (2008). Exploiting subjectivity analysis in blogs to improve political leaning categorization. In Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 725–726). New York: ACM. doi:10.1145/1390334.1390472.
Jürgens, P., Jungherr, A. and Schoen, H. (2011). Small worlds with a difference: new gatekeepers and the filtering of political information on Twitter. Proceedings of the ACM WebSci'11 (pp. 14–17).
Kannabiran, G. and Petersen, M.G. (2010). Politics at the interface: a Foucauldian power analysis. In Proceedings of the 6th Nordic Conference on Human–Computer Interaction: Extending Boundaries (pp. 695–698). New York: ACM. doi:10.1145/1868914.1869007.
Kaschesky, M. and Riedl, R. (2011). Tracing opinion-formation on political issues on the Internet: a model and methodology for qualitative analysis and results. In 2011 44th Hawaii International Conference on System Sciences (HICSS) (pp. 1–10). doi:10.1109/HICSS.2011.456.
Kaschesky, M., Sobkowicz, P. and Bouchard, G. (2011). Opinion mining in social media: modeling, simulating, and visualizing political opinion formation in the Web. In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times (pp. 317–326). New York: ACM. doi:10.1145/2037556.2037607.
Kim, B.J., Kavanaugh, A. and Pérez-Quiñones, M. (2007). Toward a model of political participation among young adults: the role of local groups and ICT use. In Proceedings of the 1st International Conference on Theory and Practice of Electronic Governance (pp. 205–212). ACM.
Kim, S., Valente, F. and Vinciarelli, A. (2012). Automatic detection of conflicts in spoken conversations: ratings and analysis of broadcast political debates. In 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5089–5092). doi:10.1109/ICASSP.2012.6289065.
Leetaru, K. (2012). Data Mining Methods for the Content Analyst: An Introduction to the Computational Analysis of Content. London: Routledge.
Mascaro, C., Black, A. and Goggins, S. (2012). Tweet recall: examining real-time civic discourse on Twitter. In Proceedings of the 17th ACM International Conference on Supporting Group Work (pp. 307–308). Retrieved from http://dl.acm.org/citation.cfm?id=2389233.

Munson, S.A. and Resnick, P. (2010). Presenting diverse political opinions: how and how much. In Proceedings of the 28th International Conference on Human Factors in Computing Systems (pp. 1457–1466).
Nahon, K. and Hemsley, J. (2011). Democracy.com: a tale of political blogs and content. In 2011 44th Hawaii International Conference on System Sciences (HICSS) (pp. 1–11). doi:10.1109/HICSS.2011.140.
Parameswaran, M. and Whinston, A.B. (2007). Social computing: an overview. Communications of the Association for Information Systems, 19(37), 762–780.
Park, S., Ko, M., Kim, J., Liu, Y. and Song, J. (2011). The politics of comments: predicting political orientation of news stories with commenters' sentiment patterns. In Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (pp. 113–122). New York: ACM. doi:10.1145/1958824.1958842.
Potter, W.J., Cooper, R. and Dupagne, M. (1993). The three paradigms of mass media research in mainstream communication journals. Communication Theory, 3(4), 317–335. doi:10.1111/j.1468-2885.1993.tb00077.x.
Ratkiewicz, J., Conover, M., Meiss, M., Goncalves, B., Patil, S., Flammini, A. and Menczer, F. (2011). Truthy: mapping the spread of astroturf in microblog streams. In Proceedings of the 20th International Conference Companion on World Wide Web (pp. 249–252). New York: ACM. doi:10.1145/1963192.1963301.
Russell, M.A. (2013). Mining the Social Web. Stebastopol, CA: O'Reilly Media.
Sarmento, L., Carvalho, P., Silva, M.J. and de Oliveira, E. (2009). Automatic creation of a reference corpus for political opinion mining in user-generated content. In Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion (pp. 29–36). New York: ACM. doi:10.1145/1651461.1651468.
Shmueli, G. and Koppius, O. (2011). Predictive analytics in information systems research. Management Information Systems Quarterly, 35(3), 553–572.
Singh, V.K., Mahata, D. and Adhikari, R. (2010). Mining the blogosphere from a socio-political perspective. In 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM) (pp. 365–370). IEEE.
Skoric, M., Poor, N., Achananuparp, P., Lim, E.-P. and Jiang, J. (2012). Tweets and votes: a study of the 2011 Singapore general election. In 2012 45th Hawaii International Conference on System Science (HICSS) (pp. 2583–2591). doi:10.1109/HICSS.2012.607.
Stieglitz, S. and Dang-Xuan, L. (2012). Political communication and influence through microblogging: an empirical analysis of sentiment in Twitter messages and retweet behavior. In Hawaii International Conference on System Sciences (Vol. 0, pp. 3500–3509). Los Alamitos, CA: IEEE Computer Society. doi:http://doi.ieeecomputersociety.org/10.1109/HICSS.2012.476.
Ulicny, B., Kokar, M.M. and Matheus, C.J. (2010). Metrics for monitoring a social-political blogosphere: a Malaysian case study. IEEE Internet Computing, 14(2), 34–44. doi:10.1109/MIC.2010.22.
Vallina-Rodriguez, N., Scellato, S., Haddadi, H., Forsell, C., Crowcroft, J. and Mascolo, C. (2012). Los Twindignados: the rise of the indignados movement on Twitter. In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on Social Computing (SocialCom) (pp. 496–501). doi:10.1109/SocialCom-PASSAT.2012.120.
Wallsten, K. (2011, January). Beyond agenda setting: the role of political blogs as sources in newspaper coverage of government. In 2011 44th Hawaii International Conference on System Sciences (HICSS) (pp. 1–10). IEEE.
Wang, F.-Y., Carley, K.M., Zeng, D. and Mao, W. (2007). Social computing: from social informatics to social intelligence. Intelligent Systems, IEEE, 22(2), 79–83.
Weber, I., Garimella, V.R.K. and Borra, E. (2012). Mining web query logs to analyze political issues. In Proceedings of the 3rd Annual ACM Web Science Conference (pp. 330–334). New York: ACM. doi:10.1145/2380718.2380761.
Wei, L. and Yan, Y. (2010). Knowledge production and political participation: reconsidering the knowledge gap theory in the Web 2.0 environment. In 2010 The 2nd IEEE International