



## Click to subscribe: interest group emails as a source of data

Zachary Albert<sup>1</sup> 

Published online: 21 July 2020  
© Springer Nature Limited 2020

### Abstract

Modern interest groups frequently utilize email communications with members as an organizational and informational tool. Furthermore, the nature of email communications—frequent, abundant, and simple to collect—makes them an excellent source of data for studies of interest groups. Nevertheless, despite the substantive importance and methodological possibilities of email communications, few interest group scholars have taken advantage of this data source due to the lack of a comprehensive, systematic database of email texts. This article makes the case for emails as a form of (big) data in the interest group field and discusses best practices for compiling and analyzing datasets of interest group emails. The article also introduces the Political Group Communication Database—the first large scale database of interest group and think tank email communications—and discusses the utility of this (and related) data for answering perennial and newly emergent questions in the interest group field.

**Keywords** Interest groups · Emails as data · Internet communication

Interest group communications represent an important data source for scholars. The technological innovations brought on by the Internet have made such data more abundant and observable to researchers, though to date studies of new group communication modalities have been limited due to the lack of large-scale, systematic data. Email communications in particular are a ubiquitous but underutilized form of group communication data, despite the fact that interested parties can often “click to subscribe” to email updates from a wide range of interest groups and other political organizations. Taking advantage of this natural source of communications data and new forms of textual analysis, scholars who systematically collect interest group

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1057/s41309-020-00097-7>) contains supplementary material, which is available to authorized users.

---

✉ Zachary Albert  
zalbert@brandeis.edu

<sup>1</sup> Department of Politics, Brandeis University, MS 058, 415 South Street, Waltham, MA 02453, USA



emails can gain new leverage on important questions about interest group agendas, strategies, biases, and objectives.

There are, to be sure, difficulties associated with the systematic collection and analysis of email communications, but these hurdles are by no means insurmountable. To illustrate this point, this article makes the case for email communications as a source of data and introduces the Political Group Communication Database (PGCD), a large-scale, continually updated dataset of email communications from interest groups and think tanks. While this dataset is not appropriate for all researchers, the data collection process should be illustrative for scholars seeking to take advantage of the abundance of data that interest groups make readily available in their electronic communications with members.

## Interest groups and the internet revolution

Interest group scholars have long recognized the importance of group communications as both an analytic topic and a source of data. For interest groups, direct communication with members is instrumental for group maintenance—attracting and retaining support—as well as direct and indirect political influence—educating, recruiting, and mobilizing members (e.g., Schlozman and Tierney 1986; Kollman 1998). For scholars, group communications can provide direct indicators of organizational agendas, strategies, messaging, and objectives.

Nevertheless, extant studies of interest group communications have been limited in their scope, with data and methodological limitations restricting the breadth and depth of insights gleaned from group communication. Much of this research, for example, has focused on the frequency of particular types of communication (e.g., through the mail or media outlets) and messages (e.g., fundraising, mobilization, educational) without analyzing the abundance of information found in the actual *content* of group messages. Others have focused more closely on message content but—due to the lack of large-scale data—only for select groups or topical areas. In both cases, research on group communications has relied primarily on the dominant methodological tools of the trade: content analysis, surveys, and interviews. Each of these methods—due to their costliness, restrictive formats, and limited scope—poses problems for the systematic analysis of large troves of group messages.

These data and methodological problems are compounded by the communication revolution brought on by the Internet. Interest groups increasingly communicate with the public through email and social networks like Facebook and Twitter (Karpf 2012). The rise in “organizations without members”—or professionally managed associations with passive supporters (Skocpol 2003)—is related to this trend and has increased the importance of mass electronic communication. For researchers, the data and methodological limitations inherent in interest group communications studies are compounded by the abundance of new online communication methods.



Despite the ever-increasing availability of interest group messages, until recently scholars have struggled to systematically collect and analyze this data.<sup>1</sup>

Perhaps the most obvious blind spot concerns email communication, which remains one of the most important tools available to interest groups despite the proliferation of newer mediums (Karpf 2013). To be sure, there have been a handful of studies employing emails as data. Drutman and Hopkins (2013), for instance, use rare internal Enron emails—made public during federal investigations—to study the firm’s lobbying efforts. Vining (2011) studies emails from ten interest groups sent to mobilize supporters during Supreme Court nomination hearings. And, in the most expansive study identified, Karpf (2012) collected emails from roughly 70 progressive organizations over six months to understand the activities of newly emergent “netroots” organizations. As these few studies show, however, examinations of interest group email communications have been limited in scope and timeframe. To date, no systematic, comprehensive, and continually updating database of interest group email communications has been developed.

## Email as (big) data

Email communications—both small and large scale—represent a substantively important, easily obtained, systematic, and analytically versatile form of data on interest group communications. First, emails are a ubiquitous form of communication for interest groups, with the vast majority of organizations distributing content via email to large mailing lists on a weekly or daily basis. In fact, email is the primary platform for interest group communications with members, allowing for rapid, frequent, and essentially costless interactions (Karpf 2013). Thus, the content and messaging found in these communications are *substantively* relevant for studies of interest group recruitment, mobilization, agendas, and strategies. Precisely because email communications form the backbone of modern interest group engagement with members, the messages contained within them are important not only as data but also as an arena for scholarly inquiry.

Furthermore, emails are delivered—with little effort on the part of the researcher—on a continually updating basis. Emails come to the researcher in real time and in a relatively consistent format. This data does not require insider access to obtain—in most cases interested parties can simply opt into receive email updates (but see the Appendix for exceptions). The medium is also far more stable than newer technological platforms like Facebook and Twitter, where groups can edit or delete content after posting. Perhaps most importantly, once an interest group sample is defined and enrollment is completed, the resulting dataset (i.e., email inbox) is

---

<sup>1</sup> A growing body of research has focused on Internet-based communication platforms like Facebook and Twitter, with increasingly sophisticated methods and expanded data (e.g., Perlmutter 2008; Williams and Gulati 2013). Nevertheless, as Karpf (2010: 11) notes, these studies demonstrate a “technocentric bias” by focusing on newer technologies like social networking sites. More mundane (but central) modes of communication, such as email, have received little attention.



compiled in a *systematic* and fairly *comprehensive* manner. Thus, email communications might help scholars address the “paucity of large-scale work” (Baumgartner and Leech 1998) in the interest group field.

Finally, emails provide researchers with an abundance of textual data which can be analyzed using traditional (e.g., content analysis) and newly emergent (e.g., automated text analysis) methods. As I discuss below, a wide range of tools have been developed specifically for the large-scale analysis of textual data. These methods enable scholars to ask and answer questions about message content, frequency, sentiment, and similarity across texts. For this reason, as I demonstrate in the concluding section, email communications can allow interest group scholars to address important questions in topics as wide ranging as agenda setting and issue attention, grassroots mobilization, coalition formation and maintenance, influence strategies, and latent ideological, partisan, or policy preferences.

For these reasons, a number of political scientists have turned to email texts as a form of data. Lindsay Cormack (2017), for example, has compiled the DCInbox, a dataset containing all official e-newsletters sent by members of Congress. These emails have been utilized to answer questions about partisan differences in issue attention (Cormack 2018) and gendered communication patterns (Cormack 2016), among other topics. Others have conducted more focused analyses of candidate email usage in campaigns (e.g., Trammell and Williams 2008). Outside legislatures and campaigns, scholars have explored how interest groups encourage members to send their own emails to regulatory agencies (Shulman 2009) or mobilize and coordinate activism via email (e.g., Pickerill 2003).

Still, the number of studies employing email as data—especially in the interest group field—is not commensurate with the role that emails play in collective political action and communication. One reason is the inherent difficulty of processing and analyzing email data once it has been collected. Email text is notoriously “noisy” in that it contains a range of substantively unimportant words and characters (see below). For any type of larger-scale, automated textual analysis, such noise must be processed out of the email texts. These procedures necessarily require some level of coding ability or access to text processing software.

There are additional concerns when it comes to employing email as data, though most simply require researchers to understand and acknowledge the particular type of data that emails provide. First, emails represent only one form of communication employed by interest groups. The content found in emails is written for a particular audience, namely those that decide to opt into the group’s updates. This means that, while the content is somewhat public-facing, it is also curated. Furthermore, groups have become increasingly sophisticated in their message testing and targeting, meaning that email content might be tailored not only for supporters but even for particular *subsets* of supporters. Many groups adopt a “culture of testing” and experiment with alternative messages, allowing them to identify best practices and develop segmented lists of recipients who receive specialized content based on geography, engagement, or other factors (Karpf 2012). In many cases, email analytics inform future interest group behavior (Karpf 2016), requiring the researcher to grapple with (or at least acknowledge) the fact that group communications represent, in many cases, a highly tailored presentation of information.



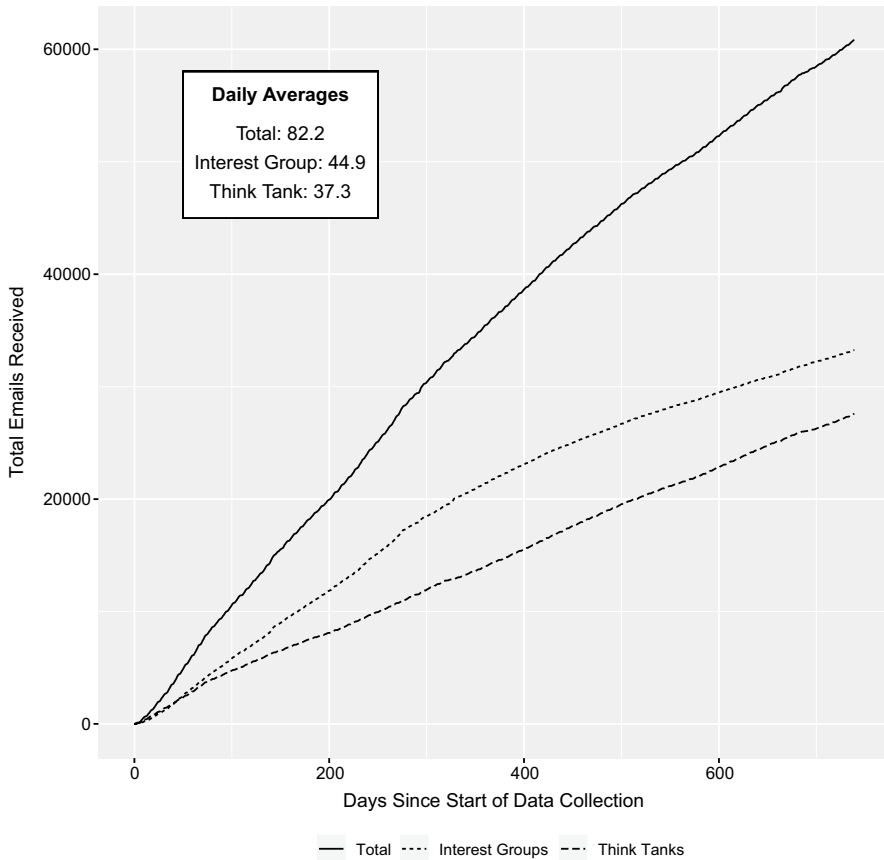


Fig. 1 Total and average emails collected over time

## Best practices for collecting email data

With these considerations in mind, I now turn to a discussion of best practices for email data collection, oriented around my own efforts to systematically collect email updates from a large sample of political organizations over an extended period of time. The result of these efforts is the Political Group Communication Database (PGCD), a continually updating dataset which (as of February 2020) includes more than 61,000 emails from 160 interest groups and 90 think tanks. Furthermore, the database receives, on average, 82 new emails per day, with roughly 55 percent of the daily emails sent by interest groups and the remainder sent by think tanks (see Fig. 1). The database currently contains emails from large, well-known organizations as well as smaller, niche groups involved in diverse issue areas including social welfare, health, national security, civil rights, and the economy, among others (see the Appendix for further information on the sample and issue areas).



In general, the pace and scope of this email collection effort highlights the abundance of data that can be amassed through relatively simple means. This data is publicly available<sup>2</sup> but may not be appropriate for all purposes. For this reason, I use this discussion to offer general guidance for scholars who might wish to collect email data for their own purposes (and refer readers interested in this particular dataset to the Appendix for further information).

### Sample definition and data collection

The first step in collecting email data is to define the sample of groups from which texts will be solicited. The choice of group samples is problem and project dependent: in some cases scholars might seek broadly representative samples—in which case they might look to comprehensive, third-party lists of groups like those used to create the PGCD sample (see the Appendix). In other cases they may seek a purposive sample focused on particular groups or issue areas—in which case their sample might be informed by subject area expertise, existing literature, or consultation with practitioners. One benefit of emails-as-data is that the relatively low costs associated with data collection allow for very large initial samples.

Another incidental benefit of collecting textual data through email subscriptions is that many groups share or sell their mailing lists to other organizations, meaning that the initial sample is often updated with new groups in a process akin to “snowball sampling”. Koger et al. (2009) creatively use this feature of interorganizational information sharing to map out partisan information networks. By signing up for mailing lists using unique email addresses, interest group scholars might similarly track the diffusion of their email subscription across organizations to study group collaborations and operationalize concepts like advocacy coalitions.

After defining an appropriate interest group sample, scholars should create a dummy email account to receive the emails. Dummy accounts—or email addresses which are used only to subscribe to email updates from organizations in the sample—allow researchers to easily download substantively significance messages and avoid (mostly) contamination by emails outside the scope of the research. Gmail is a commonly utilized email server for research purposes because it is free and allows for simple export of emails in “MBOX” format (see below).

Once an initial sample of interest groups has been defined and a dummy account has been created, subscribing to email updates is as simple as navigating to the interest groups’ websites, identifying whether and where email enrollment can be completed, and entering the dummy email address in the online form. Most modern organizations (roughly 73 percent in the PGCD) have active websites and forms where interested parties can “subscribe” or “sign up” for email updates, though there are some exceptions (see the Appendix). Organizations require varying levels of information to receive group communications, with some simply requiring an active email address and others soliciting additional information such as the name or

---

<sup>2</sup> The complete database and documentation can be found at [www.zackalbert.com/PGCD](http://www.zackalbert.com/PGCD).



location of the subscriber. Providing additional information introduces the possibility that emails will be tailored to specific geographic areas or subscriber types, so it is recommended that researchers provide only required information. Alternatively, researchers might design experiments whereby they provide different types of information (e.g., gendered or racialized names) to the same organization to study message testing and customization.

Following enrollment, email data collection is an automated process that requires limited researcher involvement, with one notable exception. Because of the culture of testing noted above, groups often monitor the “read rates” of their email messages and may identify non-responsive email addresses which, in the future, will receive fewer messages. Furthermore, following enrollment, some organizations send confirmation emails with links that subscribers must click to activate their subscription. For both these reasons, it is recommended that researchers monitor the flow of incoming messages—especially in the early stages of data collection—and “interact” with a handful of messages from each organization by opening the emails and clicking on at least one embedded link to the organization’s website. Such procedures will help prevent slippage in the group sample and email receipt rate.

At this stage, messages are stored on the server hosting the email account. In order to process and analyze these texts, researchers must download the emails to their hard drive.<sup>3</sup> Gmail has built in functionality to download entire inboxes in MBOX format, a common file type for storing email messages and meta data as a single, concatenated file. Other email programs, such as Microsoft Outlook or Apple Mail, allow for emails to be saved as individual EML files. When downloading emails, scholars can often choose various parameters (e.g., date range, email type) to define the sample of emails to extract. The resulting file (or files) includes the raw text and meta data for all emails within the parameters defined by the researcher. To enable automated text analysis, however, scholars must *process* these emails as clean text files that can be read and analyzed.

### Processing emails as text data

Perhaps the most difficult hurdle associated with using emails as text data concerns the process of converting the raw email text and meta data into a format that can be processed by human coders or computer software. Raw email files contain a great deal of noise, or substantively irrelevant and analytically distracting words and characters. The largest source of noise in email data is often the programming code that tells a computer how to lay out email elements (e.g., images), alter text (e.g., italicize), and reference external data (e.g., hyperlinks). Unfortunately, the process of converting emails into MBOX or EML format does not remove such information from the text. Thus, researchers must develop a procedure for “non-text filtering” in

<sup>3</sup> There are some methods (e.g., the “gmailr” R package) that enable interfacing with the email server without directly downloading emails, though for large scale analysis it is recommended to export the messages.



order to eliminate characters outside the main email content and relevant meta data (Tang et al. 2005).

Fortunately, a variety of software packages and scripts have been developed for the purpose of such non-text filtering. For example, more technically sophisticated scholars might utilize pre-built R packages or Python scripts, while less technically sophisticated researchers might utilize a range of free or paid online processors meant to strip HTML text from small numbers of emails.<sup>4</sup> In either case, the ultimate goal is to remove substantively irrelevant words and characters and, potentially, to “tokenize” texts into *n*-grams, or split up larger strings (i.e., complete email messages) into continuous sequences of *n* words. *N*-grams represent the inputs in automated textual analysis and, strung together, they comprise the full email message. Processing emails as *n*-grams represents a second type of pre-processing—text normalization—that enables consistent analysis across texts in the corpus (Tang et al. 2005).

At this stage, there are a number of more minor pre-processing decisions that might be made. First, despite the best efforts of the researcher, the corpus of email texts will likely include a small number of emails outside the scope of the project. The most common examples are bureaucratic emails sent by the hosting company. It is recommended that, prior to analysis, scholars examine the addresses from which emails were sent and remove any outside the scope of the project. Furthermore, the meta data extracted from the emails—which includes, among other information, the email address, author name, date sent, and subject line—does not clearly indicate the group that authored the email, as groups often send messages from multiple email addresses that do not necessarily indicate their organizational affiliation. Scholars might utilize automated or hand coding methods to connect email addresses to specific organizations. On the automated side, string extraction functions can be used to identify the text between the “@” symbol and “.com” or “.org” suffix. Still, these results will likely be incomplete, requiring researchers to hand-edit the output and connect it to named organizations, enabling the attachment of relevant external meta data to specific email texts.

## Methods of analysis

The analytical possibilities offered by email communications have expanded considerably with the advent of new, sophisticated textual and content analysis methodologies. At the same time, the data collected through email subscriptions lends itself to tried-and-true modes of quantitative and qualitative analysis. In this section I outline some of the most common methods researchers might utilize to analyze the trove of information found in interest group emails.

On the less technical side, scholars might rely on human coders or their own close readings of email texts to conduct content analysis or to construct qualitative

---

<sup>4</sup> It is worth noting that automated non-text filtering procedures will likely be imperfect, leaving some emails with noise that is not easy to extract on a large scale. Despite best efforts, for example, some PGCD emails still contain HTML programming code.





narratives. Because interest groups utilize email to fundraise, mobilize members, and discuss policies, content analysis of email communications can provide a direct measure of the activities in which groups engage. Furthermore, email communications provide a window into a particular type of group activity—grassroots engagement with members and supporters—which is often more difficult to observe and quantify than more visible activities like campaign contributions or lobbying. Similarly, the raw content found in emails provides a direct window into the messaging groups engage in and would be well-suited for process-tracing and other qualitative methods.

A range of more technically sophisticated modes of analysis fall broadly under the heading of *automated textual analysis*. In some cases, these methods are only semi-automated or “supervised” by the researcher. Machine-aided content analysis, for example, requires researchers to hand code and classify some subset of emails to act as a “training set” for the automated classifier. After this initial step, researchers can scale up their classification of texts by training an algorithm to automatically classify incredibly large numbers of documents into pre-defined categories. Scholars have used supervised content analysis to classify individual texts—for example, to assign congressional bills to topic areas (Hillard et al. 2008)—and to make generalizations about a population of documents—for example, to quantify partisan rhetoric in presidential stump speeches over time (Rhodes and Albert 2017). Interest group scholars could similarly utilize automated content analysis to understand individual and aggregate policy agendas (see below).

Alternatively, scholars might employ entirely autonomous or “unsupervised” methods, such as topic modeling, without having to pre-define keywords and categories or employ hand-coding to develop a training set. Quinn et al. (2010) develop a statistical model that estimates rather than assumes the keywords that define a topic and the division of topics within the observed data, allowing researchers to topic-code political texts with minimal assumptions. Grimmer (2010) utilizes a similar method to identify the issue agendas of individual US Senators by classifying press releases. These unsupervised methods are especially useful when there is uncertainty about the topical categories or their keyword correlates.

Another promising methodological tool is sentiment analysis, which allows researchers to measure the affect and tone of evaluative statements. Political scientists have utilized sentiment analysis to study topics like partisan polarization (Monroe et al. 2008) and media framing (Soroka et al. 2015). Interest group scholars might employ sentiment analysis to operationalize group affect toward, for example, policies, political actors, or parties.

Scholars might also utilize the natural linkages found in emails in conjunction with network analysis to identify and map advocacy coalitions or partisan networks. The hyperlinks found in emails, for example, allow scholars to identify references to content produced by other organizations. Such linkages can be used to create “citation networks” indicative of some form of engagement or collaboration. Relatedly, measures of “textual similarity” or outright plagiarism might be used to identify instances of information sharing within coalitions, as Albert (2019) did in a study of the spread of partisan language in cap-and-trade debates.



The methods discussed here are among the most common employed in the field of textual analysis, but scholars can and should be *question driven* in their methodological approach, seeking out tried-and-true as well as innovative methods that fit their specific problem. In the final section, I turn to some of these problems, highlighting the utility of interest group email data in answering both new and old questions in the field.

## Empirical applications

The texts found in interest group emails have the potential to shed new light on perennial and emerging questions in the study of interest groups. Because of the size and versatility of email text data, scholars can attempt to overcome some of the most common limitations found in studies of, for example, group agenda setting, mobilization, and influence strategies. Similarly, they can contribute to new questions about latent group preferences and ideologies as well as the role of interest groups in “extended party networks”.

First, email data combined with content analysis or topic modeling can contribute to studies of interest group agenda setting and issue priorities. Existing studies on these topics suffer from a simple problem of data availability. Scholars have largely satisfied by focusing on narrow time periods or data sources (e.g., *New York Times* coverage), but it is difficult to track issue attention and agenda setting over time and across many groups using this type of data. Larger-scale attempts to quantify agendas—such as the laudable Comparative Agendas Project—have focused on specific indicators of attention (e.g., congressional debates, hearings, and laws) but have not been able to operationalize the way that groups prioritize and frame issues to their own members. Large-scale email data, on the other hand, presents the opportunity to track issue attention from the interest group’s perspective. By analyzing groups’ public statements to their members, scholars can identify not only issue salience but also issue framing over extended periods of time. Furthermore, they can quantify the effect of focusing events on issue attention in real time and produce more nuanced estimates of groups’ effects on things like governmental and media agendas.

Second, large-scale email data allows scholars to more accurately quantify group mobilization efforts without having to rely on self-reported behavior in surveys or interviews. Interest groups often distribute “calls to action” to their members via email and, through content analysis, scholars might leverage such messages to operationalize groups’ attempts to mobilize members in elections—for example, to vote for endorsed candidates—and “outside lobbying” efforts—for example, to contact legislators or sign online petitions. This type of research would contribute to the study of electoral politics, public policymaking, and social movements, among other areas. Furthermore, group emails often reveal other strategies related to organizational maintenance—such as fundraising or advertising incentives—and influence efforts—such as inter-group collaborations and policy coalitions. Analyses of these types of messages could contribute to studies of organizational maintenance, incentive systems, and advocacy coalitions.



Relatedly, the messages sent by interest groups represent a new form of data that might aid in the identification and quantification of group support for (or opposition to) particular candidates, parties, and movements. Most simply, researchers might use some combination of keyword searches, content analysis, or sentiment analysis to identify instances where groups mention or make normative claims about people and policies. These methods could be used to identify or infer groups' policy preferences in the absence of public position-taking, allowing for more detailed understandings of their policy positions and latent ideological preferences. Sim et al. (2013), for instance, infer the proportions of various ideologies found in candidates' speeches. Applying a similar method to interest group texts might yield more accurate measures of group ideologies than survey data and less endogenous measures than financial records.

The same methods could also be used to identify the explicit or implicit partisan preferences of interest groups, contributing to emerging theory that frames interest groups as embedded within "extended party networks" (Bawn et al. 2012). Email messages offer the opportunity to operationalize partisan support (e.g., endorsements or normative statements) and informal ties across groups. Finally, scholars might utilize email data to investigate issues of interest group representation and potential biases by focusing on the issues and people that are discussed in these communications and the language used to talk about them. In general, the information found in group messages can enable creative researchers to *classify* organizations along a wide range of dimensions and *contextualize* organizational behavior, strategies, and preferences.

## Compliance with ethical standards

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

- Albert, Z. 2019. Partisan Policymaking in the Extended Party Network: The Case of Cap-and-Trade Regulations. *Political Research Quarterly*. <https://doi.org/10.1177/1065912919838326>.
- Baumgartner, F.R., and B.L. Leech. 1998. *Basic Interests: The Importance of Groups in Politics and Political Science*. Princeton, NJ: Princeton University Press.
- Bawn, K., M. Cohen, D. Karol, S. Masket, H. Noel, and J. Zaller. 2012. A Theory of Political Parties: Groups, Policy Demands and Nominations in American Politics. *Perspectives on Politics* 10 (3): 571–597.
- Cormack, L. 2016. Gender and Vote Revelation Strategy in the United States Congress. *Journal of Gender Studies* 6 (25): 626–640.
- Cormack, L. 2017. DCinbox—Capturing Every Congressional Constituent E-newsletter from 2009 Onwards. *The Legislative Scholar* 2 (1): 27–34.
- Cormack, L. 2018. *Congress and U.S. Veterans: From the GI Bill to the VA Crisis*. Santa Barbara, CA: Praeger.
- Drutman, L., and D.J. Hopkins. 2013. The Inside View: Using the Enron E-mail Archive to Understand Corporate Political Attention. *Legislative Studies Quarterly* 38 (1): 5–30.
- Grimmer, J. 2010. A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis* 18 (1): 1–35.



- Hillard, D., S. Purpura, and J. Wilkerson. 2008. Computer-Assisted Topic Classification for Mixed-Methods Social Science Research. *Journal of Information Technology & Politics* 4 (4): 31–46.
- Karpf, D. 2010. Online Political Mobilization from the Advocacy Group's Perspective. *Policy & Internet* 2 (4): 1–35.
- Karpf, D. 2012. *The Move On Effect: The Unexpected Transformation of American Political Advocacy*. New York, NY: Oxford University Press.
- Karpf, D. 2013. How Will the Internet Change American Interest Groups? In *New Directions in Interest Group Politics*, ed. M. Grossmann, 136–157. Abingdon: Routledge.
- Karpf, D. 2016. *Analytic Activism: Digital Listening and the New Political Strategy*. New York, NY: Oxford University Press.
- Koger, G., S. Masket, and H. Noel. 2009. Partisan Webs: Information Exchange and Party Networks. *British Journal of Political Science* 39 (3): 633–653.
- Kollman, K. 1998. *Outside Lobbying: Public Opinion and Interest Group Strategies*. Princeton: Princeton University Press.
- Monroe, B.L., M.P. Colaresi, and K.M. Quinn. 2008. Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis* 16 (4): 372–403.
- Perlmutter, D.D. 2008. Political Blogging and Campaign 2008. *International Journal of Press/Politics* 13 (2): 160–170.
- Pickerill, J. 2003. *Cyberprotest: Environmental Activism Online*. Manchester: Manchester University Press.
- Quinn, K.M., B.L. Monroe, M. Colaresi, M.H. Crespin, and D.R. Radev. 2010. How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science* 54 (1): 209–228.
- Rhodes, J.H., and Z. Albert. 2017. The Transformation of Partisan Rhetoric in American Presidential Campaigns, 1952–2012. *Party Politics* 23 (5): 566–577.
- Schlozman, K.L., and J.T. Tierney. 1986. *Organized Interests and American Democracy*. New York: Harper & Row.
- Sim, Y., B. Acree, J.H. Gross, and N.A. Smith. 2013. Measuring Ideological Proportions in Political Speeches. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA*.
- Skocpol, T. 2003. *Diminished Democracy: From Membership to Management in American Civic Life*. Norman: University of Oklahoma Press.
- Shulman, S.W. 2009. The Case Against Mass E-mails: Perverse Incentives and Low Quality Public Participation in US Federal Rulemaking. *Policy & Internet* 1 (1): 23–53.
- Soroka, S., L. Young, and M. Balmas. 2015. Bad News or Mad News? Sentiment Scoring of Negativity, Fear, and Anger in News Content. *The ANNALS of the American Academy of Political and Social Science* 659 (1): 108–121.
- Tang, J., H. Li, Y. Cao, and Z. Tang. 2005. Email Data Cleaning. In *Proceedings of SIGKDD 2005*. August 21–24, 2005, Chicago, IL. 489–499.
- Trammell, K.D., and A.P. Williams. 2008. Beyond Direct Mail: Evaluating Candidate E-Mail Messages in the 2002 Florida Gubernatorial Campaign. *Journal of E-Government* 1 (1): 105–122.
- Vining, R.L. 2011. Grassroots Mobilization in the Digital Age: Interest Group Response to Supreme Court Nominees. *Political Research Quarterly* 64 (4): 790–802.
- Williams, C.B., and G.J. Gulati. 2013. Social Networks in Political Campaigns: Facebook and the Congressional Elections of 2006 and 2008. *New Media & Society* 15 (1): 52–71.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

