

# Election campaigning on social media: Politicians, audiences and the mediation of political communication on Facebook and Twitter

Sebastian Stier, Arnim Bleier, Haiko Lietz, Markus Strohmaier

*GESIS – Leibniz Institute for the Social Sciences*

[sebastian.stier@gesis.org](mailto:sebastian.stier@gesis.org)

## Abstract

Although considerable research has concentrated on online campaigning, it is still unclear how politicians use different social media platforms in political communication. Focusing on the German federal election campaign 2013, this article investigates whether election candidates address the topics most important to the mass audience and to which extent their communication is shaped by the characteristics of Facebook and Twitter. Based on open-ended responses from a representative survey conducted during the election campaign, we train a human-interpretable Bayesian language model to identify political topics. Applying the model to social media messages of candidates and their direct audiences, we find that both prioritize different topics than the mass audience. The analysis also shows that politicians use Facebook and Twitter for different purposes. We relate the various findings to the mediation of political communication on social media induced by the particular characteristics of audiences and sociotechnical environments.

**Keywords:** online campaigning; social media; cross-media analysis; text analysis; language models

Forthcoming in *Political Communication*

DOI: <https://doi.org/10.1080/10584609.2017.1334728>

## Introduction

Social media have become ubiquitous communication channels for candidates during election campaigns. Platforms like Facebook and Twitter enable candidates to directly reach out to voters, mobilize supporters and influence the public agenda. These fundamental changes in political communication therefore present election candidates with a widened range of strategic choices. Should candidates address the topics most important to a mass audience? Should they tailor their messages to the specific habits and audiences on social media platforms? Although academic research on social media campaigning has flourished in the last several years (Boulianne, 2016; Jungherr, 2016b), it is still unclear which topics politicians address on these platforms, since previous research mostly concentrated on meta data generated by the use of communication conventions such as retweets, @-mentions, likes or hashtags. Understanding the ways in which politicians adapt the contents of their messages to the peculiarities of different platforms generates deeper insights into how political communication is shaped by social media.

Much research revealed a continuation of the status quo in online campaigning, as politicians mostly replicated traditional messages and campaign modes on their web presences while limiting engagement with users (Gibson, Römmele, & Williamson, 2014; Larsson, 2015; Lilleker et al., 2011; Stromer-Galley, 2000). New media notwithstanding, it seems imperative for candidates to tailor their online messages towards the preferences of the majority of voters, in line with models of mass communication (Downs, 1957; Druckman, Kifer, & Parkin, 2010). However, the observation of McQuail (2010, p. 140) that “the audience member is no longer really part of a mass, but is either a member of a self-chosen network or special public or an individual” is especially true for interactive social media. On these platforms, politicians get directly exposed to users with rather specific demographic characteristics and political interests (Diaz, Gamon, Hofman, Kiciman, & Rothschild, 2016; Nielsen & Vaccari, 2013; Schoen et al., 2013) and have to adapt to unique affordances of social media sites (Hoffmann & Suphan, 2017; Jungherr, Schoen, & Jürgens, 2016). Candidates might

therefore tailor their communication to the sociotechnical environments of platforms like Facebook and Twitter. To infer if campaign messages are aimed at a mass audience or more particular sets of audiences, we analyze “the distribution of salience across a set of issues” (Iyengar, 1979, p. 396) and make a novel contribution by (1) deriving these distributions from survey responses to proxy the preferences of a mass audience, as well as (2) from different social media and (3) from multiple content layers (politicians and audiences on social media).

The empirical analysis focuses on political communication on Facebook and Twitter by candidates during the German federal election (*Bundestagswahl*) campaign 2013. The baseline is a representative survey of the German population conducted during the election campaign in which participants were asked to describe in free text the most important contemporary political problem. These responses were assigned to 18 topic categories by the survey administrators. We present a human interpretable Bayesian language model which allocates social media messages to the known topic categories from the survey responses, but creates additional social media specific topics, if necessary. Applying the model to social media messages of candidates and their social media audiences, we find that although their topic agendas converge to some extent with the survey, both prioritize topics like campaign-related events that are different from the concerns of a mass audience. Furthermore, the focus of candidates on social media is more similar to the topics discussed by the audiences to which they are most directly exposed. An analysis of the language used in messages adds methodological robustness to the results and confirms that politicians primarily use Facebook for campaign-related purposes like the promotion of their activities while preferring Twitter to comment on contemporary political events. We relate these differences to the mediation of political communication on social media induced by diverging characteristics of audiences and sociotechnical environments.

## **Related literature and research gaps**

Our study is located at the intersection of cross-media and social media research. There is an established research tradition relating the use of different media to outcomes and processes like political

knowledge, participation and voting (Prior, 2007), news consumption (Althaus & Tewksbury, 2002) and political communication (Druckman et al., 2010).

In terms of election campaigning on the Internet, Druckman et al. (2010) presented several relevant findings for our study. The campaign officials the authors surveyed revealed that even though they were aware that supporters are the most frequent visitors of candidate websites, these formats were still designed for a mass audience. In a comparison of websites and TV ads, the authors showed that candidates are equally likely to use both media for negative campaigning, implying that the medium and different user groups do not matter much in campaign strategy. Other studies mirrored the rather conservative use of the web by politicians (e.g. Gibson et al., 2014; Larsson, 2015; Lilleker et al., 2011; Stromer-Galley, 2000).

In terms of audience behavior, Althaus and Tewksbury (2002) established that whether news are consumed offline or online affects perceptions of issue importance. The selection of news at the individual level is “filtered” by different media, even when the original source is the same. While the newspaper edition of the New York Times heavily guides the readers by providing journalistic cues and a steady diet of “public affairs coverage”, users of its online edition were navigating contents in a manner that suits their personal preferences. Others have observed an audience fragmentation and turn towards entertainment formats in high-choice media environments as well (e.g. Nielsen & Vaccari, 2013; Prior, 2007).

Taken together, these findings posit that although the citizens who actually use the Internet for political purposes have rather specific political interests, politicians have so far used the web with a mass audience in mind. The skewed perceptions of topic importance by Internet users that Althaus and Tewksbury (2002) observed should be even more pronounced when using social media, given the social and algorithmic cues that these platforms provide. Accordingly, Jungherr et al. (2016) showed that during the German federal election campaign 2013, topic priorities of Twitter audiences deviated from a survey and mass media coverage. Strategic campaigns should adapt to these environments and narrowcast their messages to the particular audiences they encounter (Kreiss, 2016).

Election campaigning on social media has been studied extensively, as researchers examined

how election campaigns unfold, how candidates are embedded in communication networks and how they interact among themselves and with the public (cf. [Boulianne, 2016](#); [Jungherr, 2016b](#)). Still, in terms of cross-media research, this literature is limited in several regards: first, most studies focused on one isolated platform, overwhelmingly Twitter and – less often – Facebook. Second, only a fraction of this work concentrated on the actual contents of communication going beyond meta data, i.e., digital traces left by communication artifacts like @-messages, retweets, likes or hashtags. While several studies coded contents of social media posts by U.S. parliamentarians ([Bronstein, 2013](#); [Gainous & Wagner, 2014](#); [Golbeck, Grimes, & Rogers, 2010](#)), these efforts mostly consisted of smaller samples and/or did not specifically categorize the topics politicians talk about. Third, most research is confined to the boundaries of election campaigning on a given social media platform. The few cross-platform analyses either restricted themselves to main accounts of party organizations ([Larsson, 2015](#); [Rossi & Orefice, 2016](#)) or metrics of attention like the number of views or followers on multiple platforms ([Nielsen & Vaccari, 2013](#)). In recent advances, [Karlsen and Enjolras \(2016\)](#) linked a candidate survey with Twitter data to uncover candidates' strategies and determinants of Twitter success. [Bode et al. \(2016\)](#) compared television advertisements with Twitter data and identified deviations indicating that the two media represent distinct modes of campaigning. Building on that, a comparison between multiple social media platforms might reveal even more fine-grained affordances of different media. Such platform specific mediation effects urgently need to be taken into account in models of political communication ([Jungherr et al., 2016](#)).

We add to this research by integrating features from multiple spheres of political communication in one research design. Recent advances in quantitative text analysis ([Grimmer & Stewart, 2013](#)) enable us to perform text analysis at a larger scale. We regard responses to an open-ended question in a representative survey as reflections of the topic priorities of a mass audience and take these as an empirical base for the analysis of social media messages by candidates and their audiences.

## Strategic election campaigning on social media

Politicians seeking election need to be responsive to the political preferences of their constituencies (Downs, 1957). However, it is an open question if politicians tailor their online messages to the topic priorities of a mass audience or particular social media audiences. In contrast to Druckman et al. (2010) who revealed rather traditional strategies on campaign websites, we argue that social media poses a yet again different communication constellation: politicians are embedded in an interactive context which skews their messages to the topic preferences of their immediate communication network (see also Bode et al., 2016). This might be due to strategic reasoning in order to increase the success of messages or an unwitting outcome of the uses and gratifications of politicians themselves.

Our empirical study relies on survey and social media generated during the German federal election campaign 2013. Given the German electoral system where party identification is still rather strong (Arzheimer, 2006), district-level topics are of minor importance in election campaigns and public agendas mostly converge to the topics salient at the national level. Therefore, topic salience expressed in politicians' social media messages can reasonably be compared to topic salience in public opinion polls. To the best of our knowledge, no such research has been undertaken so far, which naturally makes our study an exploratory one. Yet we derive useful propositions from previous research to develop several theoretical expectations.

### *Social media as part of multifunctional online campaigns*

When theorizing on the topics of politicians' campaign messages, at least three arguments indicate that social media indeed pose campaign environments distinct from mass communication arenas. First, we have to consider that social media are not only used to address political topics important to a mass audience, but perform several other functions in election campaigns. Kobayashi and Ichifuji (2015), for instance, identified three functions: promoting issue positions, demonstrating beneficial personality traits and improving name recognition. Jungherr (2016a) proposed a four-fold typology distinguishing between organizational uses, active campaigning in information spaces, resource collection and allocation as well as symbolic purposes. A considerable share of

online campaigning should therefore be devoted to the mobilization of supporters, organization of campaigns (Lilleker et al., 2011; Nielsen & Vaccari, 2013) and representational/symbolic purposes. Contrary to the “electronic brochures” of websites (Hoffmann & Suphan, 2017; Stromer-Galley, 2000), a campaign has to internalize a whole set of platform-specific affordances on social media in order to demonstrate that it represents the “state of the art”.

Second, the demographic composition as well as the political preferences and interests of social media audiences are much different from a representative sample of citizens (Diaz et al., 2016; Schoen et al., 2013). In the context of the German election 2013, political audiences on Twitter rarely talked about core political issues like the Euro crisis or energy policy, but predominantly addressed NSA surveillance and campaign-related events like the televised debate between the leading national candidates (Jungherr et al., 2016). Strategic, “micro-targeted” (Kreiss, 2016) campaigns should tailor their messages to specific audiences and successful marketing strategies on social media. While the German political system and privacy regulations certainly pose barriers to data-driven micro-targeting (Stier, 2015), mediation effects of social media platforms should still be felt in online campaigning. In consequence, candidates might be inclined to address topics like internet policy, which are more important to social media than mass audiences, and unfolding campaign events.

Third, candidates themselves have different uses and gratifications of online media (Hoffmann & Suphan, 2017; Marcinkowski & Metag, 2014). One of the central features of the Internet is its coalescence of private and public functions (McQuail, 2010, p. 41), and many German politicians indeed operate their Twitter accounts personally (Spiegel Online, 2015). Studying individual-level predictors of online campaigning, Marcinkowski and Metag (2014) revealed that German state-level candidates with a more positive attitude towards the Internet use its applications more intensively. Hoffmann and Suphan (2017) reported that although the motives self-promotion and information dissemination are dominant among Swiss politicians, which is consistent with traditional modes of campaigning, they also use social media for more personal uses like information seeking and entertainment. The authors conclude that “specific motives—possibly based on varying levels of understanding of and experience with new media—can lead to more or less avid and strategic ICT

adoption” (Hoffmann & Suphan, 2017, p. 251).

Given the multiple functions of campaigns as well as an adoption of typical practices and engagement with specific audiences by a considerable share of candidates, it can be assumed that the topics salient in politicians’ social media messages do not necessarily reflect the topic priorities of a mass audience.

**H1:** Topic saliences in messages by candidates and audiences on social media are more similar to each other than to topic salience among a mass audience.

### *The use of Facebook and Twitter by election candidates*

Due to the different architectures of social media platforms, topic saliences could also differ between the two networks we look at, Facebook and Twitter. Previous studies have shown that different media logics influence the strategic considerations underlying election campaigns (Bode et al., 2016; Kreiss, 2016). On Twitter, most user accounts are publicly visible and accessible even for non-registered audiences. Its usage is centered around topics and the retweet feature facilitates the diffusion of political information beyond the direct follower network via two-step flow processes (Wu, Hofman, Mason, & Watts, 2011). In contrast, most accounts on Facebook are private and its usage is based on one-way or reciprocal friendship ties. Information travels less fluidly through this medium, also due to the extensive algorithmic filtering of contents. Thus, the audience for Facebook posts mostly consists of people already “liking” a candidate page (see for similar audience-centered arguments: Nielsen & Vaccari, 2013; Norris, 2003).

Among the most active Twitter users are prime targets of campaigns like political elites and influentials. Journalists, for instance, regard Twitter of higher value for news reporting while using Facebook primarily for private purposes (Parmelee, 2014). Campaign officials interviewed by Kreiss (2016) emphasized that journalists use Twitter as an “index of public opinion”, which implies that targeted campaign messages on Twitter had the potential to create spillover effects to other media. Similarly, Twitter is used more intensively by Swedish political parties than Facebook due to the greater potential to reach opinion leaders via the former medium (Larsson, 2015). Disaggregating campaigning by political arena, Larsson and Skogerbo (2016) provided evidence from Norway that



local politicians valued Facebook higher for political communication, whereas politicians at the national level preferred Twitter. The former medium seems to be particularly suited for local politics, which election campaigning is to a large extent about, while the latter is being used to connect to national audiences.

Considering these findings, we expect that candidates use Twitter to contribute to contemporary national debates about policies or high-attention campaign events like TV debates. In contrast, politicians “preach to the converted” on Facebook (using the words of [Norris, 2003](#)), mostly party supporters and local constituents. Since these users already show a considerable interest in a candidate and her political stances, the strategic value of this friendly audience lies in mobilization, i.e., persuading followers to take part in the election campaign as volunteers or turn out to vote, rather than in convincing them of policy propositions.

**H2:** Topic salience among a mass audience is more similar to topic salience by candidates on Twitter than to topic salience by candidates on Facebook.

**H3a:** Candidates prefer Twitter to discuss policies and unfolding campaign events like TV debates.

**H3b:** Candidates prefer Facebook for campaigning and mobilization purposes.

## Methodology

**Language model.** We begin with some considerations regarding the joint modeling of survey responses and social media messages. Classification is a supervised task of identifying to which group, here topics, a document (a survey response, Facebook message or tweet) belongs on the basis of labeled training data. Clustering is an unsupervised task of grouping observations that are more likely to follow the same word distributions as those in other groups ([Bishop, 2006](#)). Our model occupies a middle ground between classification and clustering. We use labeled survey responses as training data, but then cluster those social media messages that use distinct words or word combinations than survey responses into new topics that capture the specifics of online communication.

To this end, we introduce a Bayesian semi-supervised nonparametric single-membership language model. The model allows for the grouping of short messages into known survey classes and, if necessary, additional social media specific topics. The underlying assumptions of the model are that documents are generated in a probabilistic process of (1) drawing a topic for a document and then (2) drawing the words of that document from the selected topic. Such a content analysis identifies differences in language as practiced culture and is therefore particularly suited to distinguish the different communication styles of survey and social media contexts (McFarland et al., 2013).

The model is semi-supervised, allowing for the incorporation of survey responses coded with a topic label. The labeled documents are responses to an open-ended survey question on the most important contemporary political problem in Germany (MIP). Moreover, the model is nonparametric, i.e., we do not set the total number of topics in advance (Teh & Jordan, 2010). Instead, we use a Dirichlet process prior on the number of topics. While the allocation of survey responses to topics is known, the allocation of social media messages to topics is unobserved and has to be inferred. The nonparametric prior allows for new topics specific to Facebook and Twitter communication to be established when messages cannot plausibly be explained by the topics used in the survey. Finally, we follow the single-membership assumption that each document is generated from a single topic. This assumption handles short documents such as social media messages or survey responses well (Yin & Wang, 2014). In fact, the short length of documents effectively prohibits a mixed-membership approach, such as the one put forward by Roberts et al. (2014). A formal description of this model, specifically tailored to our needs, is given in Appendix A.

The inference step of the model described in Appendix B reverses assumptions for document generation to an automated reading of otherwise latent semantic topics from the social media datasets. For this, we ran the Gibbs sampling procedure for 100 iterations. For the evaluation of model outputs, the most important criterion was the substantive fit as determined by human judgment, i.e., the quality of topics in the context of our research question (Grimmer & Stewart, 2013, p. 286). As parameters, we chose  $\alpha = 1.0$  and  $\beta = 1.5$  which produce well-interpretable topics that cluster messages unique to social media while allocating topically related social media messages to the known survey topics. The final output provides an aggregated and unified macroscopic view

on signals in political communication that would be impossible to gather with conventional methods.

**GLES survey data.** The MIP we use was administered in the Rolling Cross-Section of the German Longitudinal Election Study (GLES) (Rattinger, Roßteutscher, Schmitt-Beck, Weßels, & Wolf, 2014) accompanying the federal election 2013.<sup>1</sup> The use of such an MIP has been criticized in terms of conceptualization and measurement in previous research. Some respondents relate their answer to their individual situation, some to the political situation at the societal level, and some consider the question based on their voting considerations (Wlezien, 2005). People therefore have different understandings of topic importance, of the extent to which a topic should be considered as a problem and also of the perspective that should be taken when making these judgments. These methodological considerations notwithstanding, our application of the MIP is less problematic than its predominant application in survey research, in which open-ended responses are reduced to one quantitative scale of issue salience. In contrast, we take the entire textual contents of each response into account and therefore capture all the semantic facets related to the *importance* of topics and citizens' perception of them as being *problematic* (arguing similarly: Mellon, 2014). Therefore, the MIP question serves our methodology well as a “catch all category” soaking up the nuances in the vocabulary used by the respondents, which will in turn also be reflected in the inference stage when the model is applied to social media messages.

The GLES survey was in the field from 8 July until 3 November 2013, surveying 7,882 unique participants twice, in a pre-election and a post-election wave. We pooled all responses on the first and second MIP during this time period, which gives us a total of 23,604 observations.<sup>2</sup> Pooling implicates that observations are not independent of each other, since most participants responded multiple times in the survey.<sup>3</sup> Thus, the sample loses its initial representativeness when we aggregate responses to GLES topic categories for our empirical analysis. Yet the practical benefits of pooling to construct a sufficiently large training dataset outweigh its theoretical disadvantages, since we primarily use the survey responses for the extraction of textual information.

The open-ended responses were coded by GLES experts according to a hierarchical categorization of topics (GLES, 2013) which consists of the three traditional higher-level dimensions used in political science (politics, polity and policy). 188 responses (0.8%) were concerned with polit-

ical processes (politics) while 1,063 observations (4.5%) were concerned with political structures (polity). The overwhelming share of survey responses thus located the most important problem in a policy area. The classes for these responses required additional manual filtering and reassembling in order to arrive at discriminative and more equally sized training classes. As a result, we obtained 18 topics for training (see Appendix C).

**Social media data.** We use a dataset covering political communication on Facebook and Twitter during the election campaign for the German Bundestag 2013 gathered via the Facebook Graph API and the Twitter Streaming API (Kaczmirek et al., 2013). From this dataset, we extracted all social media messages posted by candidates, their incoming @-mentions on Twitter and the comments to their posts on Facebook during the research period analogous to the survey.<sup>4</sup> We only selected posts by parties with a realistic chance of passing the electoral threshold of 5% required for representation in the Bundestag, namely CDU, CSU, FDP, Grüne, Linkspartei and SPD. The AfD almost made it into the Bundestag in 2013, however, its rise to prominence in public opinion polls came late and thus the party was not incorporated in the data collection.

The dataset contains 49,573 Facebook posts and 134,462 tweets by candidates. As audience data, we preselected all 180,214 comments on candidates' Facebook posts and all 282,118 tweets that @-mentioned at least one candidate. These numbers at first seem to reflect a low interest by the public. However, it should be kept in mind that election-related topics are also discussed without commenting on candidates' Facebook posts or mentioning them in tweets. We chose these two particular metrics since they are the most direct exposure of a candidate to messages by her audience (account holders per default get a notification via the platform interface and via email that they received a mention/comment).

**Preprocessing.** We preprocessed our five corpora (survey, politicians Facebook, politicians Twitter, audience Facebook, audience Twitter) before we applied our model, which is a necessary reduction of linguistic complexity for text analysis (Grimmer & Stewart, 2013). The preprocessing of documents involved removing punctuation, standard German stop words, URLs and Twitter user handles. In addition, we removed the names of all political parties and candidates to prevent inferring topics that are party-specific or centered around social relations. Customized stop words,

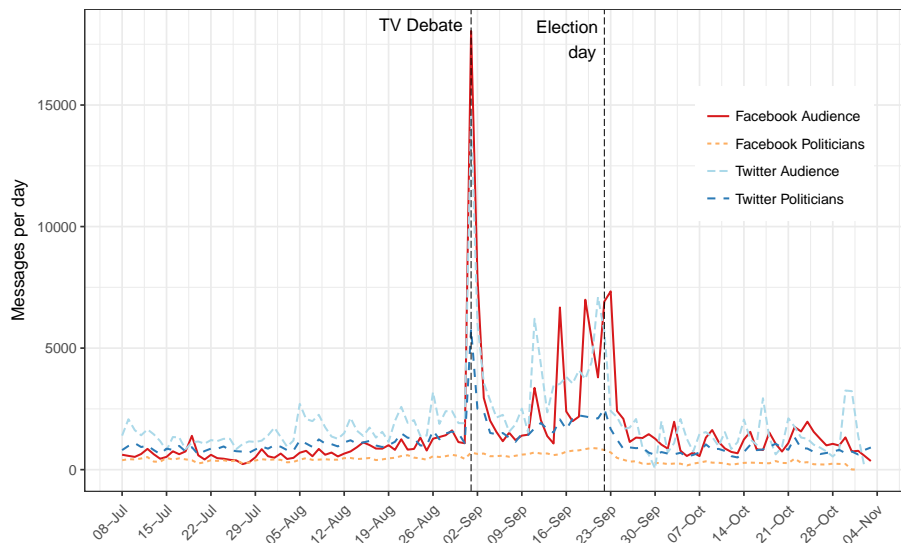
compiled inductively while refining the model, were also removed. This customized stop word list contains ambivalent words with high probabilities in many distinct topics (e.g., “Politik”) as well as very frequent and very infrequent words. Finally, only those social media messages were kept that contain at least three words in order to incorporate sufficient information for the clustering procedure of the model. Survey responses were allowed to consist of only one word since the topic was already known for these documents.

In addition to text cleaning, we made sure that the four social media corpora are comparable in size and number of words used so that they have a comparable influence on topic construction. To achieve this, we operationalized the Facebook and Twitter audience corpora as random samples of equal size to politicians’ messages, stratified by political party. For instance, the 5,902 tweets by FDP candidates were mirrored by the same number of audience tweets mentioning an FDP candidate and their 1,777 Facebook posts were matched by as many randomly sampled accompanying Facebook comments.<sup>5</sup> In total, 22,186 survey responses, 17,546 Facebook posts by politicians, 17,546 Facebook comments, as well as 54,093 tweets by politicians and 54,093 @-mentions were used as inputs for the model.

## Results and discussion

### *Analysis at the document level*

We start our analysis with a description of social media activity by politicians and audiences over time using the raw counts of messages from the unprocessed dataset. This gives us first indications about specific focus points and the topics predominantly addressed on social media. The vertical lines in Figure 1 represent the TV debate between the party leaders Angela Merkel and Peer Steinbrück (1 September 2013) and election day (22 September 2013). Especially the televised debate drew attention (see also [Jungherr et al., 2016](#); [Lietz, Wagner, Bleier, & Strohmaier, 2014](#)), since social media users are particularly active during these high-attention periods ([Diaz et al., 2016](#); [Kreiss, 2016](#)). Interestingly, the Facebook politicians time series barely reacts to the TV debate and not nearly in the order of magnitude like the other three time series. This indicates



**Figure 1:** Social media messages over time.

that politicians prefer Twitter over Facebook to comment on unfolding events (supporting H3a). We will follow up on these hints with an extensive analysis of the topics discussed by politicians and their audiences on social media.

The text analysis model described above generates 51 topics; the 18 known topics from the survey and 33 additional ones based on the social media corpora. We removed 23 new topics with less documents than the smallest known topic from the survey ( $< 468$  documents, aggregated across all social media corpora) in order to improve interpretability. Those 23 residual topics in total only covered 1,387 of 143,278 social media documents. The first 18 topics were labeled using the known topic labels from GLES (cf. Table A1). The ten remaining social media topics were labeled by inspecting the documents for each topic qualitatively. The few ambiguous cases were discussed among the authors and decided upon consensually.

**Table 1:** Top words per topic

---

*Known topics from the survey*

---

**Budget & Debt:** schulden, schuldenabbau, staatsverschuldung, verschuldung, haushalt, finanzen, ueberschuldung, staatsschulden, abbau, neuverschuldung (*debt, debt reduction, national debt, debt, budget, finances, debt overload, national debt, reduction, new debt*)

**Currency & Euro:** euro, eurokrise, schulden, europa, griechenland, krise, geld, milliarden, banken, schuldenkrise  
(*euro, euro crisis, debt, europe, greece, crisis, money, billions, banks, debt crisis*)

**Economy:** wirtschaft, finanzkrise, wirtschaftliche, wirtschaftspolitik, wirtschaftskrise, stabilitaet, wirtschaftlichen, banken, versteht, wachstum (*economy, financial crisis, economic, economic policy, economic crisis, stability, economic, banks, understand, growth*)

**Education:** bildung, bildungspolitik, mitte, schulen, jaehrige, schule, geld, lehrer, kinder, schueler (*education, education policy, middle, schools, annual, school, money, teacher, children, pupils*)

**Environment:** umwelt, umweltschutz, umweltpolitik, wohlstand, klimawandel, klima, klimaschutz, aufbruch, oekologie, energiewende (*environment, environmental protection, environmental policy, prosperity, climate change, climate, climate protection, start, ecology, energy transformation*)

**Family Policy:** kinder, kita, frauen, betreuungsgeld, familie, familienpolitik, eltern, familien, gleichstellung, kind  
(*children, kita, women, child care subsidy, family, family policy, parents, families, equalization, child*)

**Foreign Policy (Defense):** syrien, fluechtlinge, krieg, lampedusa, frieden, russland, aussenpolitik, muessen, europa, waffen (*syria, refugees, war, lampedusa, peace, russia, foreign policy, must, europe, weapons*)

**Foreign Policy (Europe):** europa, europapolitik, europaeische, europaeischen, zusammenhalt, integration, europas, stabilitaet, laender, euro (*europe, european policy, european, solidarity, integration, europe's, stability, countries, euro*)

**General Fiscal Policy:** finanzen, finanzpolitik, finanzielle, finanzlage, sicherheit, geld, finanzmarkt, finanzierung, situation, ordnung (*finances, fiscal policy, financial, financial situation, security, money, capital market, funding, situation, order*)

**General Social Policy:** soziale, gerechtigkeit, altersarmut, reich, armut, schere, ungerechtigkeit, sozialen, wandel, sozialpolitik (*social, justice, poverty of the elderly, rich, poverty, gap, injustice, social, change, social policy*)

**Health Care & Pensions:** renten, rente, pflege, buergerversicherung, gesundheitspolitik, tvduell, rentenpolitik, gesundheit, gesundheitswesen, medizin (*pensions, pension, caregiving, citizen insurance, health policy, tvduell, pension policy, health, health sector, medicine*)

**Infrastructure:** energiewende, energie, erneuerbare, strom, energien, kohle, energiepolitik, volksentscheid, umlage, klimaschutz (*energy transformation, energy, renewable, electricity, energies, coal, energy policy, referendum, contribution, climate protection*)

**Labor Market:** mindestlohn, arbeitslosigkeit, arbeit, euro, muessen, steuern, geld, leben, maut, tvduell (*minimum wage, unemployment, work, euro, must, taxes, money, live, toll, tvduell*)

**Law & Order:** nsa, prism, snowden, ueberwachung, datenschutz, affaere, bnd, freiheit, daten, buerger (*nsa, prism, snowden, surveillance, data protection, affair, bnd, freedom, data, citizens*)

**Migration & Integration:** integration, auslaender, auslaenderpolitik, zuwanderung, migranten, einwanderung, asylbewerber, asylanten, migration, einwanderungspolitik (*integration, foreigners, policy towards foreigners, immigration, migrants, asylum seeker, migration, asylum policy*)

**Politics:** euro, muessen, buerger, jahr, unternehmen, arbeit, stadt, land, region, leben (*euro, must, citizens, year, businesses, work, city, country, region, live*)

**Polity I:** geld, ehrlichkeit, glaubwuerdigkeit, buerger, bevoelkerung, demokratie, uneinigkeit, ausland, volk, vertrauen (*money, honesty, credibility, citizens, population, democracy, disagreement, foreign countries, people, trust*)

**Taxes:** steuern, steuerpolitik, progression, steuererhoehung, steuer, kalte, kalten, abbau, steuergerechtigkeit, steuererhoehungen (*taxes, fiscal policy, progression, tax increase, tax, cold, reduction, fiscal justice, tax increases*)

---

*New topics found on social media*

---

**Campaigning (Events):** veranstaltung, talk, podiumsdiskussion, wahlkreis, gast, einladung, interview, unterwegs, bundestagswahl, gespraech (*event, conversation, panel discussion, constituency, guest, invitation, interview, on the road, federal election, talks*)

**Campaigning (Local):** danke, dank, super, guten, stimmung, spass, infostand, wahlkreis, unterwegs, aktion (*thanks, thank, super, good, atmosphere, fun, info booth, constituency, on the road, action*)

**Coalition Formation:** koalition, grosse, mehrheit, waehler, waehlen, grossen, muessen, bundestag, demokratie, opposition (*coalition, grand, majority, voter, vote, grand, must, bundestag, democracy, opposition*)

**Demonstrations:** nazis, demo, baden, wuerttemberg, angst, innen, wasser, freiheit, platz, protest (*nazis, demo, baden, wuerttemberg, fear, inside, water, freedom, square, protest*)

**Misconduct:** paedophilie, zurueck, treten, debatte, verfassungsschutz, tritt, unfassbar, arbeit, ruecktritt, aufloesen (*pedophilia, back, step, debate, internal intelligence service, step, inconceivable, work, resignation, close*)

**NSA Surveillance:** snowden, edward, moskau, treffen, trifft, brief, nsa, respekt, germany, russland (*snowden, edward, moscow, meeting, meet, letter, nsa, respect, germany, russia*)

**Parliamentary Procedures:** bundestag, nsu, sitzung, fraktion, bundestages, fraktionssitzung, landesgruppe, bundestagsfraktion, rheinland, rede (*bundestag, nsu, session, faction, bundestag, caucus, regional group, bundestag faction, rhineland, speech*)

**Political Debates:** tvduell, waehlen, danke, bayern, dreikampf, zeit, richtig, beide, duell, kanzlerin (*tvduell, vote, thank, bavaria, triathlon, time, right, both, duel, chancellor*)

**Polity II:** leben, muessen, land, geld, buerger, wissen, richtig, volk, kinder, freiheit (*live, must, country, money, citizens, know, right, people, children, freedom*)

**Post Election:** glueckwunsch, herzlichen, bundestag, ergebnis, erfolg, geburtstag, dank, danke, nsa, gewaehlt (*congratulation, cordial, bundestag, result, success, birthday, thank, thanks, nsa, elected*)

---



Table 1 lists the ten most probable words per topic in German and English. The topics are very well interpretable. Overall, the model successfully assigned the social media documents to the known classes when appropriate. With the exception of *NSA Surveillance* (which is assigned to *Law & Order* in the survey), no new topic mirroring an existing one was created. An inspection of the documents shows that the new topic focuses on Edward Snowden and Russia, while the topic from the survey concentrates on the implications of NSA spying for citizens. One new topic, *Polity II*, revolves around the role of citizens in, and criticism of, Germany’s political structure. The topic is distinct from *Polity I* in the survey, which is more about the role of politicians in the polity. Besides these two topics, all other new topics are related to political processes. Campaign events leave particularly strong marks in the topic *Political Debates*, which is the largest new topic and also the most ambiguous one, as it contains a great variety of exchanges between politicians and audiences. Especially the televised debate is prominently featured, which concurs with [Kreiss \(2016\)](#) who showed that Twitter is used by campaigns for the engagement with audiences and journalists during high-attention events.

Table 2 shows topic saliences in the five corpora (in percent). The first clear finding is that core policy areas like *Labor Market* and *General Social Policy* have a much higher salience in the survey than on social media. This resembles the results of [Althaus and Tewksbury \(2002\)](#) who showed that readers of the paper version of the New York Times were more likely to expose themselves to “public affairs coverage” than the users of its online version. Among policies, *Infrastructure*, *Family Policy* and *Foreign Policy (Defense)* are all discussed frequently by politicians and audiences on social media. Aspects regarding NSA surveillance in the topic *Law & Order* are clearly overrepresented in messages by the audience and politicians on Twitter (see also [Jungherr et al., 2016](#)). The latter finding demonstrates that the text analysis model performs well in allocating social media messages to known topics from the survey. Besides the discussed policy areas, politicians and audiences alike mostly address different topics than the ones salient in the survey, which lends support to H1.

It also becomes apparent that politicians use Facebook and Twitter in different ways. Their summed share of messages related to both *Campaigning* topics is 42.3% on Facebook compared to 26,1% on Twitter. In contrast, *Political Debates* take up a higher share of politicians’ tweets,

**Table 2:** Topic salience per corpus (in percent)

	Survey	Politicians		Audience	
		Facebook	Twitter	Facebook	Twitter
<i>Known topics from the survey</i>					
Labor Market	19.1	4.9	6.3	8.3	6.3
General Social Policy	12.9	1.1	1.5	1.4	1.4
Currency & Euro	12.5	1.6	2.6	1.9	2.2
Education	7.2	1.1	1.4	0.5	1.1
Economy	7.0	0.5	0.6	0.5	0.6
Infrastructure	6.8	3.0	5.7	1.5	5.1
Health Care & Pensions	5.7	0.9	1.0	0.5	0.7
Migration & Integration	4.0	0.2	0.2	0.1	0.3
Polity I	4.0	0.4	0.3	0.3	0.2
Family Policy	3.3	2.4	2.4	2.3	3.0
Law & Order	3.3	3.9	7.5	3.9	9.4
Foreign Policy (Defense)	2.9	2.4	3.4	2.3	3.2
Budget & Debt	2.7	0.2	0.2	0.1	0.1
Taxes	2.6	0.2	0.3	0.1	0.2
Foreign Policy (Europe)	2.2	0.1	0.1	0.2	0.1
General Fiscal Policy	1.7	0.1	0.1	0.0	0.1
Environment	1.5	0.1	0.1	0.0	0.1
Politics	0.6	3.8	0.8	1.3	0.7
<i>New topics found on social media</i>					
Campaigning (Local)		21.2	13.7	5.6	7.4
Campaigning (Events)		21.1	12.4	1.9	4.6
Political Debates		8.5	13.6	21.1	18.8
Polity II		5.7	7.2	22.0	13.2
Coalition Formation		5.6	6.7	12.0	8.1
Post Election		3.8	3.8	6.9	4.8
Parliamentary Procedures		3.3	3.2	0.5	1.4
Demonstrations		2.3	2.8	0.5	2.0
NSA Surveillance		0.4	1.0	0.4	3.1
Misconduct		0.3	0.5	0.3	1.4

Note: Ranked by first given column.

and the latter medium is also used more extensively to discuss various policies like *Infrastructure* and *Law & Order*. This indicates that politicians tailor their messages to different media logics and audiences. In line with our medium specific hypotheses, candidates use Twitter for the commentary of policies and unfolding public events (H3a), while trying to mobilize Facebook users to attend campaign events (H3b).

Political audiences use social media overwhelmingly for *Political Debates*, to scrutinize the relationship between state and citizens (*Polity II*) and comment on *Coalition Formation*. While politicians and audiences are more in sync on Twitter in that regard, there is a considerable disconnect between politicians and their audiences on Facebook. Although politicians devoted 42.3% of their messages to campaigning, their audiences mostly talked about other topics.

**Table 3:** Rank correlations of topic salience in all corpora

	Survey	Facebook politicians	Twitter politicians	Facebook audience
Facebook politicians	0.43			
Twitter politicians	0.59*	0.95***		
Facebook audience	0.53*	0.91***	0.90***	
Twitter audience	0.58*	0.89***	0.95***	0.92***

Note: Spearman’s rho. For pairs including the survey, N=18. For social media pairs, N=28. \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

A systematic way to analyze topic salience at the document level is to correlate the ranks of topics in each corpus using Spearman’s rho. Several findings can be inferred from Table 3. First, comparing the topic ranks in the survey with social media reveals varying results. Topic saliences in messages by politicians on Twitter ( $p = 0.011$ ), the Facebook audience ( $p = 0.024$ ) and the Twitter audience ( $p = 0.012$ ) are all rather similar to topic salience among a mass audience. Second, topic ranks of all social media corpora are nevertheless correlated more strongly with other social media corpora than with the survey ( $p < 0.001$ ).<sup>6</sup> Third, correlations between the topic ranks in the survey and politicians’ Facebook posts are weaker than in case of the other social media corpora ( $p = 0.075$ ). This lends support to H2 which postulates that candidates address topics relevant to a mass audience on Twitter while discussing such topics more sparsely on Facebook.

We learn from these distributions that the messages of social media users are shaped by considerable mediation effects, which provides further evidence for H1. On social media, a specific subset of politically engaged citizens discusses specific topics via specific sociotechnical transmission mechanisms. Politicians seem to adjust to these mediated environments and thus have remarkably similar topic ranks like the personal networks they are most directly exposed to. They adopt the public communication practices of the Twittersphere (Wu et al., 2011) and use the medium primarily for political commentary, while trying to mobilize their interested followers on Facebook for campaign purposes. At the same time, correlations between the different content layers on social media – except for the Facebook politicians corpus – and the survey are still strong. Therefore, the stark differences in the sizes of topics per corpus revealed in Table 2 are reduced considerably when comparing the ranks of topics. When politicians and audiences talk about policies on social media, they tend to prioritize similar topics like respondents in a representative survey. This indicates that the

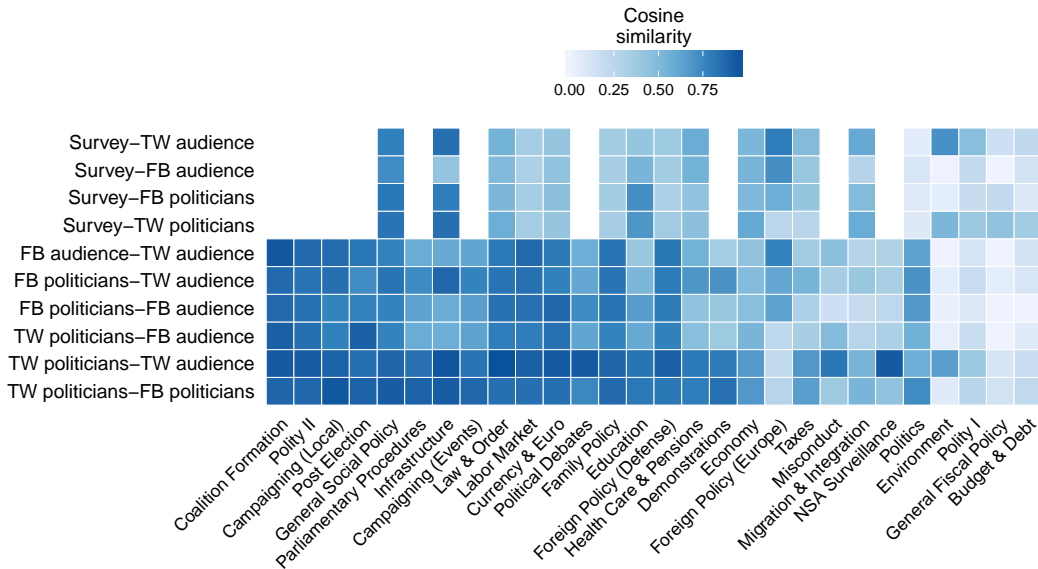
public agenda is still rather cohesive during election campaigns – independent of the medium.

### *Analysis at the word level*

The analyses at the document level revealed the salience of political topics in different media and content layers. But with the textual data at hand, we can move our analysis to the word level, which serves two purposes. First, it adds methodological robustness to the previous results. Given different types of media, it is to be expected that the language encountered in corpora is distinct, since specific conventions and space restrictions apply to interview situations, Facebook messages or tweets. The following analysis will rule out that differences in salience are mere artifacts of platform specific talk. Second, we can also investigate how different topics are perceived and talked about by a mass audience, politicians and their social media audiences. By that, we move beyond the salience of topics toward an identification of similar perspectives regarding political topics in different corpora.

Since the underlying vocabulary in our model is the same across all five corpora, we can compare the topic-word distributions across media and content layers. For this, we calculate cosine similarities (scale 0 to 1) between all corpus pairs in each topic. Figure 2 displays the results, organized in decreasing order by the average cosine similarity per topic. The darker a cell, the more similarly a corpus pair discusses a topic. Cells of social media specific topics remain blank for the survey since the topics are not featured in this corpus. It becomes clear from the visualization that similarities between corpora vary considerably depending on the topic.

In our final analysis, we take the values of the 240 cells in Figure 2 as the dependent variable in Ordinary Least Squares regressions.<sup>7</sup> This allows us to assess if our hypotheses regarding similarities between different media and content layers hold at the level of words used in political communication. As independent variables, we construct a dummy indicating that the survey is part of a given corpus pair and, similarly, if the politicians' Facebook or Twitter corpus is featured in a corpus pair. We also take control variables into account that could influence the relationships between cosine similarities and the independent variables. First, we include a logged variable counting the combined number of tokens (the aggregate word count) in each corpus pair per topic. By that



**Figure 2:** Cosine similarities between corpora by topic.

we take the size of corpus pairs into account which could systematically affect cosine similarities. We added dummies indicating that a corpus pair is from the same social media, i.e., Facebook or Twitter, or that the group (politicians, audiences) in a corpus pair is the same. Moreover, we use the information from GLES on topic types, i.e., whether a topic belongs to the categories policy, politics or policy and classified the ten new social media topics analogously. Based on that, we construct a dummy signaling that a topic is related to politics aspects, which we assume are discussed more heterogeneously than the polity or policies.<sup>8</sup> Finally the 10 new topics identified on social media are marked by a dummy variable to account for differences in topic origin (*New topics*).

Table 4, Model 1 shows that the corpus pairs including the survey have an approximately 12% lower cosine similarity, other things being equal. This demonstrates that in addition to the differences in topic salience identified before (H1), the description of political problems in the survey is also markedly different from the average topically related social media message. In Model 2 and Model 3, we see that contrary to their posts on Facebook, the language in tweets by politicians is significantly more similar to other corpora than the average corpus pair. This means that on Twitter, candidates emphasize aspects of topics important to survey respondents, but also use

**Table 4:** Models of cosine similarities between corpus pairs

	(1)	(2)	(3)	(4)	(5)
Survey	-0.12*** (0.03)	-0.12*** (0.03)	-0.10** (0.03)	-0.11*** (0.03)	-0.09** (0.03)
Facebook politicians		0.00 (0.02)	0.02 (0.02)		
Twitter politicians			0.05* (0.02)		0.06** (0.02)
Twitter audience				0.03 (0.02)	0.05* (0.02)
Number of tokens (logged)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)
Same medium	0.08** (0.03)	0.08** (0.03)	0.08** (0.03)	0.08** (0.03)	0.08** (0.03)
Same actor	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)
Topic type (politics=1)	-0.10* (0.05)	-0.10* (0.05)	-0.10* (0.05)	-0.10* (0.05)	-0.10* (0.04)
New topics	-0.02 (0.05)	-0.02 (0.05)	-0.02 (0.04)	-0.02 (0.05)	-0.01 (0.04)
Constant	-0.59*** (0.07)	-0.59*** (0.07)	-0.61*** (0.07)	-0.60*** (0.07)	-0.62*** (0.07)
R <sup>2</sup>	0.63	0.63	0.63	0.63	0.64
Adj. R <sup>2</sup>	0.62	0.61	0.62	0.62	0.63
N	240	240	240	240	240

Note: OLS with robust standard errors in parentheses.

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

language similar to other content layers on social media. Such a hybrid communication strategy by politicians can be interpreted as a synergy of mass communication strategies and messages targeted at individual follower networks on social media, much in line with the notion “masspersonal” ascribed to Twitter (Wu et al., 2011).

Several more factors could influence these results. First, throughout Table 4 the number of underlying tokens is highly correlated with the cosine similarity between the corpus pairs. This is to be expected by the law of large numbers as the empirical word distributions in both corpora become more stable with an increasing number of tokens and cosine similarities increase. Controlling for the size effect ex post in a multivariate regression model allows for a systematic comparison of similarities between corpus pairs. Second, politicians’ tweets might reflect a typical communication style on Twitter and thus not be indicative of strategic considerations. To test this, we include a Twitter audience dummy in Model 4 without finding a significant effect. When including the Twitter audience dummy in combination with the Twitter politicians dummy in Model 5, it becomes

significant, but the Twitter politicians variable still has superior explanatory power.<sup>9</sup> Third, there is nonetheless evidence for platform effects, as *Same medium* is a significantly positive predictor ( $p \leq 0.009$  in all models). This demonstrates that politicians and audiences not only emphasize similar topics (H1) but also address similar aspects when talking about a topic on the same social media platform. Fourth, the results also hold when controlling for the insignificant variable *Same actor*, which means that audiences and politicians adopt distinct communicative practices when using different media. Fifth, the topics primarily related to politics are more semantically diverse than the topic types polity and policy. Finally, similarities do not differ systematically depending on whether a topic is known from the survey or newly created from social media (*New topics*). This is of particular methodological importance, as it demonstrates that semantic differences and the measurement of topic salience at the document level are not just artifacts of stark differences in the two communication situations interview and social media message. Finding a systematic effect of the two data generation modes would have questioned the validity of applying a model based on survey responses to social media data.<sup>10</sup>

In sum, Table 4 reveals similarities in political communication between corpora that resemble the ranks of topic salience at the document level. Still, problem descriptions in the survey and social media are more distinct than to be expected from Table 3. This shows that overlaps in topic salience can still mask *how* topics are talked about and perceived by different audiences.<sup>11</sup> We also learn that the distinct uses of Facebook and Twitter by politicians are not just byproducts of media specific language, but are rather driven by the strategic considerations we discuss in our theory. Moreover, the analysis at the word level makes the topic generation process more transparent and increases our understanding of mediation processes in political communication.

## Conclusion

This study provides insights into how political communication is shaped by social media. We compared the topics of most importance to a mass audience in a representative survey to the topics discussed on Facebook and Twitter. Based on a text analysis model developed specifically for this

purpose, we show that politicians and their audiences discuss different topics on social media than those salient among a mass audience. Moreover, politicians use Facebook and Twitter for different purposes which we relate to the distinct target groups candidates encounter. Taken together, our findings suggest that campaign strategies and political communication in general are mediated by varying sociotechnical affordances of social media platforms.

These results challenge previous findings in political communication which indicated that politicians use the web rather conservatively and in a non-interactive manner (Druckman et al., 2010; Gibson et al., 2014; Larsson, 2015; Lilleker et al., 2011; Stromer-Galley, 2000). However, our study does not necessarily contradict previous research, as each Internet application has specific affordances and politicians demonstrate a strategic awareness of various communication arenas. They complement the “masspersonal” communication (Wu et al., 2011) in the quasi-public sphere of Twitter with the more direct communication practices on Facebook for organizational and mobilization purposes. Our cross-media study also shows that relevant differences in political communication exist between social media platforms. This underscores the need to argue with the utmost caution when trying to infer findings from one platform to “social media” as a whole, as it has often been done.

Several findings indicate that the high-choice media environments of social media contribute to a fragmentation of the mass audience (Althaus & Tewksbury, 2002; Prior, 2007; McQuail, 2010, p. 140). Politicians and their audiences talk about policies sparsely, but rather discuss campaign-related events and topics specific to social media. Their salient topics resemble each other remarkably, compared to the priorities of survey respondents. Furthermore, the language used indicates that other aspects of similar topics are addressed on social media than in survey responses. Those differences notwithstanding, when candidates and audiences actually discuss policies on Facebook and Twitter, they prioritize those similarly like respondents in a representative survey. In this regard, the public agenda is still rather integrated. This points towards persistent – although probably diffuse and mediated – agenda-setting effects between mass media and social media (Neuman, Guggenheim, Jang, & Bae, 2014) as well as within social media. A fruitful research avenue would be to focus on the interactions of politicians and audiences with accounts of journalists and legacy



media which bridge the gap between the general public and more particular sets of audiences on social media.

We also want to discuss the limitations of this study. The article focused on an election campaign with several high-attention events that were most commented-on on social media. A follow-up study during non-election periods could possibly reveal more issue-related communication in line with the preferences of a mass audience. Demographic information on the follower graphs of politicians on Facebook and Twitter would certainly enhance our findings, however, such data is not readily available to researchers. While our single-membership model is well suited for the short nature and narrow focus of social media messages, its downside is that the more multidimensional and textually rich messages necessarily also have to be allocated to just one topic. Moreover, due to restrictions in survey data containing an MIP and the considerable efforts required to mine social media data of all election candidates, we only investigated one campaign in one country. On the one hand, the mediation effects we found should travel well, as German social media campaigning was still in its infancy in 2013 and severe regulatory restrictions on micro-targeting apply (Stier, 2015). More sophisticated campaigns elsewhere might be better at addressing the particular needs of digital media. On the other hand, the rather ad hoc campaign style and personal use of accounts by several candidates in our case might be more suited for the peculiarities of social media than, for instance, the highly professionalized U.S. campaigns. Therefore, it remains to be seen to which extent the findings are applicable to online campaigning in other countries and the (near) future.

The paper makes a methodological contribution to the field of political text analysis (Grimmer & Stewart, 2013) by proposing a novel methodology to apply labeled text to a test corpus while also allowing for the introduction of additional topic categories. The method is applicable in a vast range of research contexts in political communication in which a test dataset deviates from a predefined coding scheme. Substantively, the paper shows that mediation effects induced by social media platforms and their sociotechnical environments are strongly felt in political communication. This means that social media is not an ideal data source for citizens seeking clearly structured information on policies or researchers using textual information to locate parties in an ideological space.

## Acknowledgements

We thank three reviewers, the editors and participants at the ECPR General Conference 2016 for helpful comments.

## Notes

<sup>1</sup>The wording is: “In your opinion what is the most important political problem facing Germany at the moment?” (Rattinger et al., 2014).

<sup>2</sup>Before the election, 7,249 participants (92%) stated a MIP and 6,673 (85%) stated a second MIP. After the election, 5,016 respondents (64%) stated a MIP and 4,666 (59%) stated a second MIP.

<sup>3</sup>4,220 (54%) stated four problems, 543 (7%) stated three, 2,364 (30%) stated two and 367 (5%) stated one problem. 388 (5%) did not know a problem or did not respond.

<sup>4</sup>The data mining of Facebook posts ended three days earlier, on 31 October 2013.

<sup>5</sup>We took audience tweets in which only one politician was mentioned for the sampling.

<sup>6</sup>The correlations between all social media corpora are very similar and highly significant ( $p < 0.001$ ) when only using the 18 known topics from the survey as input.

<sup>7</sup>Since most models were heteroscedastic, we use robust standard errors.

<sup>8</sup>This is not a hard distinction but rather an exploratory application of the GLES labels. In fact, it is difficult to distinguish policy aspects from politics aspects related to an issue.

<sup>9</sup>We do not include all corpus dummies at once because as part of corpus pairs, the dummies are not mutually exclusive. Therefore, we cannot exclude one dummy as the reference category and interpret results accordingly, as it is usually done with multiple-category variables.

<sup>10</sup>The main results are robust when rerunning the regression models using only known topics from GLES (with  $N=180$  cells in Figure 2) and also when solely taking the new social media topics into account ( $N=60$  cells). The exception is the model with only known topics, in which the Twitter politicians and audience dummies lose their significance. Since some of the dummy variables overlap considerably, we also ran broader and smaller models with varying constellations of included variables, which confirm the main findings.

<sup>11</sup>Although agenda-setting is outside the scope of our study, this distinction resembles the conceptualization of first- and second-level effects found in the related literature (McCombs, Llamas, Lopez-Escobar, & Rey, 1997).

## References

- Althaus, S. L., & Tewksbury, D. (2002). Agenda setting and the “new” news. *Communication Research*, 29(2), 180–207. doi: 10.1177/0093650202029002004
- Arzheimer, K. (2006). ‘Dead men walking?’ Party identification in Germany, 1977-2002. *Electoral Studies*, 25(4), 791–807. doi: <http://dx.doi.org/10.1016/j.electstud.2006.01.004>
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Heidelberg: Springer.
- Bode, L., Lassen, D. S., Kim, Y. M., Shah, D. V., Fowler, E. F., Ridout, T., & Franz, M. (2016). Coherent campaigns? Campaign broadcast and social messaging. *Online Information Review*, 40(5), 580–594. doi: 10.1108/OIR-11-2015-0348
- Boulianne, S. (2016). Campaigns and conflict on social media: A literature snapshot. *Online Information Review*, 40(5). doi: 10.1108/OIR-03-2016-0086
- Bronstein, J. (2013). Like me! Analyzing the 2012 presidential candidates’ Facebook pages. *Online Information Review*, 37(2), 173–192. doi: 10.1108/OIR-01-2013-0002
- Diaz, F., Gamon, M., Hofman, J. M., Kiciman, E., & Rothschild, D. (2016). Online and social media data as an imperfect continuous panel survey. *PLoS ONE*, 11(1), 1–21. doi: 10.1371/journal.pone.0145406
- Downs, A. (1957). *An economic theory of democracy*. New York, NY: Harper & Row.
- Druckman, J. N., Kifer, M. J., & Parkin, M. (2010). Timeless strategy meets new medium: Going negative on congressional campaign web sites, 2002-2006. *Political Communication*, 27(1), 88–103. doi: 10.1080/10584600903502607
- Gainous, J., & Wagner, K. M. (2014). *Tweeting to power: The social media revolution in American politics*. Oxford University Press.
- Gibson, R., Römmele, A., & Williamson, A. (2014). Chasing the digital wave: International perspectives on the growth of online campaigning. *Journal of Information Technology & Politics*, 11(2), 123–129. doi: 10.1080/19331681.2014.903064
- GLÉS. (2013). *Kategorienschema und Hinweise für die Codierung der Agendafragen*. Retrieved from <https://dbk.gesis.org/dbksearch/download.asp?id=52526>

- Golbeck, J., Grimes, J. M., & Rogers, A. (2010). Twitter use by the U.S. Congress. *Journal of the American Society for Information Science and Technology*, *61*(8), 1612–1621. doi: 10.1002/asi.21344
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. doi: 10.1093/pan/mps028
- Hoffmann, C. P., & Suphan, A. (2017). Stuck with ‘electronic brochures’? How boundary management strategies shape politicians’ social media use. *Information, Communication & Society*, *20*(4), 551–569. doi: 10.1080/1369118X.2016.1200646
- Iyengar, S. (1979). Television news and issue salience: A reexamination of the agenda-setting hypothesis. *American Politics Quarterly*, *7*(4), 395–416. doi: 10.1177/1532673X7900700401
- Jungherr, A. (2016a). Four functions of digital tools in election campaigns: The German case. *The International Journal of Press/Politics*, *21*(3), 358–377. doi: 10.1177/1940161216642597
- Jungherr, A. (2016b). Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, *13*(1), 72–91. doi: 10.1080/19331681.2015.1132401
- Jungherr, A., Schoen, H., & Jürgens, P. (2016). The mediation of politics through Twitter: An analysis of messages posted during the campaign for the German federal election 2013. *Journal of Computer-Mediated Communication*, *21*(1), 50–68. doi: 10.1111/jcc4.12143
- Kaczmarek, L., Philipp Mayr, Ravikiran Vatrappu, Arnim Bleier, Manuela Blumenberg, Tobias Gummer, . . . Christof Wolf (2013). Social media monitoring of the campaigns for the 2013 German Bundestag elections on Facebook and Twitter. *GESIS-Working Papers*.
- Karlsen, R., & Enjolras, B. (2016). Styles of social media campaigning and influence in a hybrid political communication system: Linking candidate survey data with Twitter data. *The International Journal of Press/Politics*, *21*(3), 338–357. doi: 10.1177/1940161216645335
- Kobayashi, T., & Ichifuji, Y. (2015). Tweets that matter: Evidence from a randomized field experiment in Japan. *Political Communication*, *32*(4), 574–593. doi: 10.1080/10584609.2014.986696
- Kreiss, D. (2016). Seizing the moment: The presidential campaigns’ use of Twitter during the 2012

- electoral cycle. *New Media & Society*, 18(8), 1473–1490. doi: 10.1177/1461444814562445
- Larsson, A. O. (2015). Green light for interaction: Party use of social media during the 2014 Swedish election year. *First Monday*, 20(12).
- Larsson, A. O., & Skogerbø, E. (2016). Out with the old, in with the new? Perceptions of social (and other) media by local and regional Norwegian politicians. *New Media & Society*. doi: 10.1177/1461444816661549
- Lietz, H., Wagner, C., Bleier, A., & Strohmaier, M. (2014). When politicians talk: Assessing online conversational practices of political parties on Twitter. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (pp. 285–294). Palo Alto, CA: AAAI Press.
- Lilleker, D. G., Koc-Michalska, K., Schweitzer, E. J., Jacunski, M., Jackson, N., & Vedel, T. (2011). Informing, engaging, mobilizing or interacting: Searching for a European model of web campaigning. *European Journal of Communication*, 26(3), 195–213. doi: 10.1177/02673231111416182
- Marcinkowski, F., & Metag, J. (2014). Why do candidates use online media in constituency campaigning? An application of the theory of planned behavior. *Journal of Information Technology & Politics*, 11(2), 151–168. doi: 10.1080/19331681.2014.895690
- McCombs, M., Llamas, J. P., Lopez-Escobar, E., & Rey, F. (1997). Candidate images in Spanish elections: Second-level agenda-setting effects. *Journalism & Mass Communication Quarterly*, 74(4), 703–717. doi: 10.1177/107769909707400404
- McFarland, D. A., Ramage, D., Chuang, J., Heer, J., Manning, C. D., & Jurafsky, D. (2013). Differentiating language usage through topic models. *Poetics*, 41(6), 607–625. doi: 10.1016/j.poetic.2013.06.004
- McQuail, D. (2010). *McQuail's mass communication theory* (6th ed.). London: Sage publications.
- Mellon, J. (2014). Internet search data and issue salience: The properties of Google trends as a measure of issue salience. *Journal of Elections, Public Opinion and Parties*, 24(1), 45–72. doi: 10.1080/17457289.2013.846346
- Neal, R. M. (2000). Markov Chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2), 249–265.

- Neuman, W. R., Guggenheim, L., Jang, M., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, *64*(2), 193–214. doi: 10.1111/jcom.12088
- Nielsen, R. K., & Vaccari, C. (2013). Do people “like” politicians on Facebook? Not really. Large-scale direct candidate-to-voter online communication as an outlier phenomenon. *International Journal of Communication*, *7*.
- Norris, P. (2003). Preaching to the converted? Pluralism, participation and party websites. *Party Politics*, *9*(1), 21–45. doi: 10.1177/135406880391003
- Parmelee, J. H. (2014). The agenda-building function of political tweets. *New Media & Society*, *16*(3), 434–450. doi: 10.1177/1461444813487955
- Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge, UK: Cambridge University Press.
- Rattinger, H., Roßteutscher, S., Schmitt-Beck, R., Weßels, B., & Wolf, C. (2014). *Rolling cross-section campaign survey with post-election panel wave (GLES 2013). ZA5703 data file version 2.0.0*. Cologne: GESIS Data Archive. doi: 10.4232/1.11892
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., . . . Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082. doi: 10.1111/ajps.12103
- Rossi, L., & Orefice, M. (2016). Comparing Facebook and Twitter during the 2013 General Election in Italy. In A. Bruns, E. Skogerbø, Christensen, Christian, A. O. Larsson, & S. E. Gunn (Eds.), *The Routledge Companion to Social Media and Politics* (pp. 434–446). London: Routledge.
- Schoen, H., Gayo-Avello, D., Metaxas, P. T., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. *Internet Research*, *23*(5), 528–543. doi: 10.1108/IntR-06-2013-0115
- Spiegel Online. (2015). *Die-140-Zeichen-Macht*. Retrieved 31 January 2017, from <http://www.spiegel.de/politik/deutschland/bundestagsabgeordnete-auf-twitter-wer-wie-viel-schreibt-und-mit-wem-a-1041402.html>
- Stier, S. (2015). Strukturbedingungen im Online-Wahlkampf: USA und Deutschland im Vergle-

ich (Structural conditions in online campaigning: Comparing the U.S. and Germany). In C. Bieber & K. Kamps (Eds.), *Die US-Präsidentenschaftswahl 2012* (pp. 363–382). Wiesbaden: VS Verlag für Sozialwissenschaften.

Stromer-Galley, J. (2000). Online interaction and why candidates avoid it. *Journal of Communication*, 50(4), 111–132. doi: 10.1111/j.1460-2466.2000.tb02865.x

Teh, Y. W., & Jordan, M. I. (2010). Hierarchical Bayesian nonparametric models with applications..

Wlezien, C. (2005). On the salience of political issues: The problem with ‘most important problem’. *Electoral Studies*, 24(4), 555–579. doi: <http://dx.doi.org/10.1016/j.electstud.2005.01.009>

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Who says what to whom on Twitter. In *Proceedings of the 20th International Conference on World Wide Web* (pp. 705–714). New York, NY, USA: ACM.

Yin, J., & Wang, J. (2014). A Dirichlet multinomial mixture model-based approach for short text clustering. In *Proceedings of the 20th ACM SIGKDD* (pp. 233–242). New York, NY: ACM.

## Appendix A: Description of the model

Following the common notation for probabilistic language models, we treat a GLES response  $d$  as a tuple of a topic indicator variable  $z_d^G$  and a vector  $\mathbf{w}_d^G$  of observed words. The indicator  $z_d^G \in [1, \dots, \hat{K}]$  is an assignment to one of  $\hat{K}$  initial GLES topics, and each of the  $n_d^G$  observed words  $w_{d_i}^G$  in  $\mathbf{w}_d^G$  is chosen from a vocabulary of  $V$  terms, leading to the collection of GLES responses  $\mathcal{D}^G = \{(\mathbf{w}_d^G, z_d^G), \dots, (\mathbf{w}_{D^G}^G, z_{D^G}^G)\}$ . Analogously, we treat a social media message  $(\mathbf{w}_d^M, z_d^M)$  as a tuple of words and a topic assignment. While the  $n_d^M$  words  $\mathbf{w}_d^M$  are chosen from the same vocabulary as in the case of GLES responses, the topic assignments  $\mathbf{z}^M$  are unobserved and not restricted to the initial  $\hat{K}$  topics. The collection of social media messages is  $\mathcal{D}^M = \{(\mathbf{w}_d^M, z_d^M), \dots, (\mathbf{w}_{D^M}^M, z_{D^M}^M)\}$ . We refer to the collection of all documents (GLES responses and social media messages) by  $\mathcal{D} = \mathcal{D}^G \cup \mathcal{D}^M$ . The generative storyline for the documents in our model, i.e. the mechanism assumed to create responses and messages, can be described by the following steps.

1. For the unrestricted global topic popularity a distribution  $\theta$  is drawn from an infinite dimen-

sional Dirichlet distribution (i.e.,  $K \rightarrow \infty$ ) with parameter  $\alpha$ :

$$\boldsymbol{\theta} \sim Dir\left(\frac{\alpha}{K}\right).$$

2. For each topic  $k = [1, \dots, \infty[$ , a distribution  $\phi_k$  over the vocabulary is drawn from a  $V$ -dimensional Dirichlet distribution parametrized by  $\beta$ :

$$\phi_k \sim Dir(\beta).$$

3. The GLES responses  $\mathcal{D}^{\mathcal{G}}$  and social media messages  $\mathcal{D}^{\mathcal{M}}$  are then generated by the same mechanism of first drawing a topic index  $z_d$  from the topic popularity  $\boldsymbol{\theta}$  and subsequently drawing multinomially distributed words  $\mathbf{w}_d$  from a topic indexed by  $z_d$ :

$$\mathbf{w}_d \sim Mult(\phi_{z_d}), \quad z_d \sim Cat(\boldsymbol{\theta}),$$

where we omitted the indexes  $\mathcal{G}$  and  $\mathcal{M}$  for readability, leading to the joint probability for the model

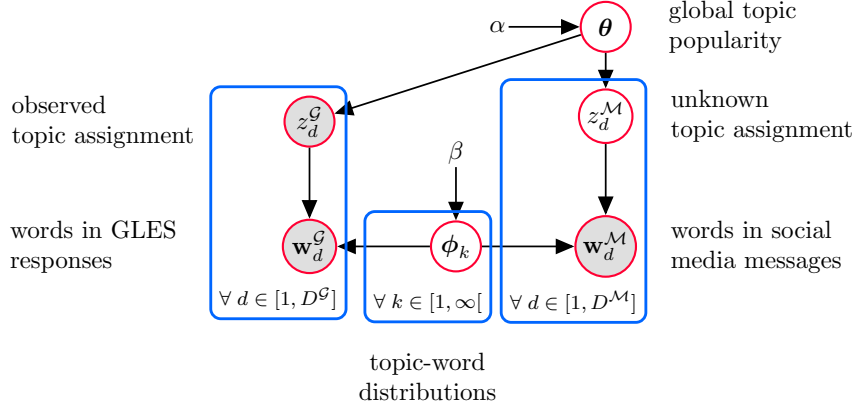
$$p(\mathcal{D}, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\phi}) = Dir(\boldsymbol{\theta} \mid \frac{\alpha}{K}) \prod_{k=1}^{\infty} Dir(\phi_k \mid \beta) \prod_{d=1}^{|\mathcal{D}^{\mathcal{G}}|} Mult(\mathbf{w}_d^{\mathcal{G}} \mid \phi_{z_d^{\mathcal{G}}}) Cat(z_d^{\mathcal{G}} \mid \boldsymbol{\theta}) \prod_{d=1}^{|\mathcal{D}^{\mathcal{M}}|} Mult(\mathbf{w}_d^{\mathcal{M}} \mid \phi_{z_d^{\mathcal{M}}}) Cat(z_d^{\mathcal{M}} \mid \boldsymbol{\theta}).$$

Figure A1 gives a graphical representation of our model.

## Appendix B: Description of the inference

For inference, we resort to collapsed Gibbs sampling of the latent indicator variables  $\mathbf{z}^{\mathcal{M}}$ , with all other variables integrated out. The full conditional posterior probability for a topic  $k$  is then given





**Figure A1: Model in Plate Notation:** Random variables are represented by nodes. Blank nodes are used for the (unknown) hidden variables  $\theta$ ,  $\phi$  and  $\mathbf{z}^M$ . Shaded nodes denote the observed words  $\mathbf{w}^M$  and  $\mathbf{w}^G$  as well as the observed labels  $\mathbf{z}^G$ , bare symbols indicate the fixed priors  $\alpha$  and  $\beta$ . Directed edges between the nodes then define conditional probabilities, where the child node is conditioned on its parents. The plate surrounding  $\phi_k$  indicates the replication of the node for each mixture component and the plates surrounding  $\mathbf{w}_d^G$  and  $\mathbf{w}_d^M$  indicate the replication of the nodes over the  $\mathcal{D}^G$  labeled and  $\mathcal{D}^M$  unlabeled documents, respectively.

by

$$p(z_d^M = k) \propto \begin{cases} \frac{\neg^d n_k}{\neg^d n_{\cdot} + \alpha} p(\mathbf{w}_d^M \mid \neg^d \mathbf{w}_{z=k}^M, \mathbf{w}_{z=k}^G, \beta) & \text{if } k \leq K \\ \frac{\alpha}{\neg^d n_{\cdot} + \alpha} p(\mathbf{w}_d^M \mid \beta) & \text{if } k = K + 1, \end{cases}$$

where  $n_k = \sum_{d=1}^D \mathbb{1}[z_d = k]$  is the number of documents and  $\mathbf{w}_{z=k}$  are the words assigned to the  $k^{\text{th}}$  topic. The superscript  $\neg d$  indicates that the current document  $d$  is excluded from consideration.

We further use a approximation for the likelihood of message  $d$  under topic  $k$

$$\begin{aligned} p(\mathbf{w}_d^M \mid \neg^d \mathbf{w}_{z=k}^M, \mathbf{w}_{z=k}^G, \beta) &= \int_{\phi_k} \text{Mult}(\mathbf{w}_d^M \mid \phi_k) \text{Dir}(\phi_k \mid \hat{\beta}_k^{\neg d}) d \phi_k \\ &\approx \prod_{i=1}^{n_d^M} \frac{\neg^d n_{kw_{di}}^M + n_{kw_{di}}^G + \beta}{\neg^d n_{k\cdot}^M + n_{k\cdot}^G + V\beta}, \end{aligned}$$

with  $n_{kw}$  being the number of times term  $w$  is used in topic  $k$ . Likewise, the likelihood of  $\mathbf{w}_d^{\mathcal{M}}$  being generated from a yet unseen topic is

$$p(\mathbf{w}_d^{\mathcal{M}} | \beta) = \int_{\phi} Mult(\mathbf{w}_d^{\mathcal{M}} | \phi) Dir(\phi | \beta) d\phi \\ \approx \frac{1}{V^{n_d^{\mathcal{M}}}} .$$

Given the priors  $\alpha$  and  $\beta$ , Gibbs sampling in this model reduces to iteratively sampling the indicator variables  $\mathbf{z}^{\mathcal{M}}$  from their full conditional posterior probability. Once a sufficient number of iterations is completed, the distribution of the indicator variables and as such the number of topics  $K$  with associated observations is independent from its initialization and the sampler has converged. For a more in-depth discussion of Gibbs sampling for Dirichlet process mixture models the interested reader is referred to [Neal \(2000\)](#).

## Appendix C: Survey dataset

Topics useful for the training of the model should be sufficiently discriminative and relatively equal in size. The Rolling-Cross-Section survey of the German Longitudinal Election Study (GLES) allows for answers to be coded as politics, polity, or one of 13 policy areas. From the latter, three were removed (206 answers on “other” problems, 102 on East Germany, and 1 on cultural leisure policy). The sizes of the resulting areas ranged from 188 (“politics”) to 6,010 (“social policy”) observations. For our purposes, some areas had to be split by testing different subtopic combinations and qualitatively assessing topic overlaps, while others could be used as they are.

In essence, six areas were kept as they are. “Social policy” was split into *General Social Policy*, *Family Policy*, *Health Care & Pensions* and *Migration & Integration*. “Fiscal policy” was split into *General Fiscal Policy*, *Currency & Euro*, *Taxes* and *Budget & Debt*. GLES topics on foreign policy and defense policy were merged into a training topic on *Foreign Policy (Defense)*, but the general subtopic on Europe was extracted to stand alone. As a result, we obtained 18 discriminative topics (totalling 23,295 responses in one politics class, one polity class and 16 policy classes). The coding is reproducible from [Table A1](#).

**Table A1:** Construction of training topics.

Label	Observations	GLES Codes
Politics	188	1XXX
Polity I	1,063	2XXX
Budget & Debt	629	431X
Currency & Euro	2,830	433X
Economy	1,618	39XX, 40XX
Education	1,624	41XX
Environment	352	36XX
Family Policy	776	371X
Foreign Policy (Defense)	701	310X, 312X, 316X, 317X, 318X, 329X, 330X, 331X, 332X, 333X, 339X
Foreign Policy (Europe)	561	311X
General Fiscal Policy	389	430X, 439X
General Social Policy	2,950	370X, 372X, 373X, 378X, 379X
Health Care & Pensions	1,333	374X, 376X, 377X
Infrastructure	1,580	35XX
Labor Market	4,358	38XX
Law & Order	793	34XX
Migration & Integration	951	375X
Taxes	599	432X

Note: Topic codes from [GLES \(2013\)](#).