


Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support

Social Science Computer Review
2017, Vol. 35(3) 336-356
© The Author(s) 2016
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0894439316631043
journals.sagepub.com/home/ssc


Andreas Jungherr¹, Harald Schoen¹, Oliver Posegga²,
and Pascal Jürgens³

Abstract

In this article, we examine the relationship between metrics documenting politics-related Twitter activity with election results and trends in opinion polls. Various studies have proposed the possibility of inferring public opinion based on digital trace data collected on Twitter and even the possibility to predict election results based on aggregates of mentions of political actors. Yet, a systematic attempt at a validation of Twitter as an indicator for political support is lacking. In this article, building on social science methodology, we test the validity of the relationship between various Twitter-based metrics of public attention toward politics with election results and opinion polls. All indicators tested in this article suggest caution in the attempt to infer public opinion or predict election results based on Twitter messages. In all tested metrics, indicators based on Twitter mentions of political parties differed strongly from parties' results in elections or opinion polls. This leads us to question the power of Twitter to infer levels of political support of political actors. Instead, Twitter appears to promise insights into temporal dynamics of public attention toward politics.

Keywords

computational social science, digital trace data, mediation of politics, Twitter, electoral predictions, public opinion

Twitter Data as Indicator of Attention Toward Politics Rather than Political Support

The use of digital tools in the context of a growing number of social activities has provided scientists with a new reservoir of data documenting various aspects of social life (e.g., Lazer et al., 2009). Recently, researchers have attempted to unlock the potential of digital trace data—data produced by

¹ University Mannheim, Mannheim, Germany

² University of Bamberg, Bamberg, Germany

³ Johannes Gutenberg-University, Mainz, Germany

Corresponding Author:

Andreas Jungherr, University of Mannheim, A5, 6 Mannheim, 68131, Germany.
Email: andreas.jungherr@gmail.com

people while interacting with digital services (e.g., Howison, Wiggins, & Crowston, 2011)—in the study of public opinion (e.g., Murphy et al., 2014). Here, researchers focused predominantly on using data collected on Twitter either to infer current levels of support toward political actors or to predict their support in upcoming elections (cf. Gayo-Avello, 2013). The seeming success in linking Twitter data to metrics of political support has even led some scholars to suggest replacing or supplementing traditional survey-based techniques by social media-based strategies (e.g., Birmingham & Smeaton, 2011; Ceron, Curini, & Iacus, 2014, 2015; Franch, 2013; O'Connor, Balasubramanian, Routledge, & Smith, 2010; Thapen & Ghanem, 2013; Tjong, Sang, & Bos, 2012; Tumasjan, Sprenger, Sandner, & Welppe, 2010).

In this article, we will present evidence that the prospects of measuring political support through Twitter are highly unlikely to materialize and that the respective literature is very likely falling victim to a classic fallacy in the social sciences: using a quantitative indicator—Twitter-based metrics—to draw inferences on a latent target concept—political support—but instead measuring another concept—attention toward politics—which sometimes, but far from always, might be correlated with the target concept of interest. This mistake arises through the insufficient testing of indicator validity (cf. Adcock & Collier, 2001; Gerring, 2012; Hand, 2004).

While it is true that an ever increasing number of studies demonstrates the possibility to link Twitter-based metrics with metrics of political support in selected cases (e.g., Birmingham & Smeaton, 2011; Contractor & Faruque, 2013; DiGrazia, McKelvey, Bollen, & Rojas, 2013; Fink, Bos, Perrone, Liu, & Kopecky, 2013; Franch, 2013; Gaurav, Srivastava, Kumar, & Miller, 2013; Jensen & Anstead, 2013; McKelvey, DiGrazia, & Rojas, 2014; O'Connor et al., 2010; Skoric, Poor, Achananuparp, Lim, & Jiang, 2012; Soler, Cuartero, & Roblizo, 2012; Thapen & Ghanem, 2013; Tumasjan et al., 2010), it is also true that a just as quickly growing count of studies directly contradicts these findings (e.g., Diaz, Gamon, Hofman, Kıcıman, & Rothschild, 2014; Gayo-Avello, 2011; Huberty, 2015; Jungherr, 2015; Jungherr, Jürgens, & Schoen, 2012; Metaxas, Mustafaraj, & Gayo-Avello, 2011). This is a first indicator that the link between Twitter data and political support is far from stable.

Proponents should also be given pause by the fact that while reporting positive findings, the literature has not developed a consensus on which Twitter-based metric should be linked to which measure of political support. Suggestions range from simple counts of mentions or calculations of mention shares (e.g., Birmingham & Smeaton, 2011; DiGrazia et al., 2013; Fink et al., 2013; Gaurav et al., 2013; Skoric et al., 2012; Soler et al., 2012; Tjong et al., 2012; Tumasjan et al., 2010), mentions somehow identified as positive (e.g., Birmingham & Smeaton, 2011; Ceron et al., 2014, 2015; Fink et al., 2013; O'Connor et al., 2010; Thapen & Ghanem, 2013), or even advanced statistical procedures automatically linking Twitter-based time series with metrics of political support (e.g., Contractor & Faruque, 2013; Marchetti-Bowick & Chambers, 2012; Tsakalides et al., 2015). Again, this points to the instability of the proposed link between Twitter data and political support.

Also, up until now, proponents of using Twitter data to infer political support were content on basing their claim on the reporting of positive findings in selected cases. They have spent next to no effort in proposing and systematically testing a mechanism linking their target concept—political support—to their indicator—a Twitter-based metric of choice. Any discussion of this proposed link in the literature takes the form of ad hoc hypothesizing without systematic testing. In this, proponents fall victim to a positivistic fallacy and resemble what Sartori (1970) has termed “the unconscious thinker,” a scholar using advanced quantitative methods without linking them sufficiently to the phenomena of interest.

Finally, various studies have shown that political Twitter activity can be seen as an indicator of attention toward politics mediated by the interests and motivations of Twitter users (cf. Jungherr, 2015; Jungherr, Schoen, & Jürgens, 2016). While attention might be sometimes correlated with electoral chances—for example, in races with overwhelming odds for the favorite—this relationship

is unlikely to be stable as political attention reacts to scandals, gaffes, and unlikely surprises, all phenomena of little systematic connection with the public support of a candidate. So instead of measuring political support, we believe the literature reporting positive findings has identified cases in which—for some reason or other—their measurement of attention toward politics was correlated with public opinion polls or electoral results.

In the remainder of the article, we will show that the relationship between political support and various Twitter-based metrics is far from stable and instead that Twitter-based metrics correspond more closely to shifts in public attention toward politics. We do this, by analyzing mentions of political parties on Twitter during 3 months preceding the federal election 2013 in Germany and by comparing them to opinion polls over the same period and the results of the 2013 election. Our evidence raises doubts with regard to the validity of Twitter-based metrics as indicator for political support. In addition, trends in the daily mention counts of political parties showed no systematic link with trends in opinion polls. Instead, the dynamics in the daily mention counts of parties appear to correspond with media events, media coverage of politics, and controversies. It, therefore, appears far more likely that Twitter-based metrics measure public attention toward politics than political support. Sometimes attention might be a covariate of support, but this relationship is far from stable. Our analysis underscores the importance of approaching the use of digital trace data in the measurement and analysis of political and social phenomena cautiously and emphasizes the importance of using established standards of social science methodology in indicator validation to avoid premature conclusions.

Validating Twitter Metrics as Indicator of Political Support

With the growing popularity of the microblogging service Twitter, a series of studies claimed that mentions of political actors on Twitter correlate with their standing in opinion polls (e.g., O'Connor et al., 2010) or could even be used to predict election results (e.g., DiGrazia et al., 2013; Tumasjan et al., 2010). According to this line of reasoning, Twitter metrics provide a valid indicator of political support. This stance has met with strong critique (e.g., Diaz et al., 2014; Gayo-Avello, 2011; Huberty, 2015; Jungherr, 2015; Jungherr et al., 2012; Metaxas et al., 2011). Nevertheless, predicting election results or polling dynamics using social media data has proven to be highly popular.

Despite this heightened activity, prior research has spent not much effort examining the validity of Twitter-based metrics as indicator for political support following the prescripts of social science methodology (cf. Adcock & Collier, 2001; Gerring, 2012; Hand, 2004). Establishing the validity of an indicator requires clearly defining the target concept and its characteristics as well as spelling out differences to other concepts. Moreover, scholars should provide an auxiliary theory that links the indicator to the target concept (e.g., Carmines & Zeller, 1979; Cronbach & Meehl, 1955; Kane, 1992; Messick, 1989). Building on this theoretical analysis, a number of hypotheses, including some on convergent and discriminant validity, have to be put to empirical test before scholars should be willing to accept an indicator for a target concept.

Scholars in this field address two different target concepts of political support. Some studies aim at predicting future election results (e.g., Bermingham & Smeaton, 2011; Fink et al., 2013; Franch, 2013; DiGrazia et al., 2013; Gaurav et al., 2013; Jensen & Anstead, 2013; McKelvey et al., 2014; Skoric et al., 2012; Soler et al., 2012; Tumasjan et al., 2010). This kind of political forecasting resembles other forecasts with digital trace data, such as movies' box office results (e.g., Asur & Huberman, 2010) or stock market movements (e.g., Bollen, Mao, & Zeng, 2011). In another approach, scholars focus on correlations between Twitter-based time series and opinion polls (e.g., Contractor & Faruque, 2013; Fink et al., 2013; O'Connor et al., 2010; Thapen & Ghanem, 2013). This kind of political nowcasting conceives of Twitter as indicating current aggregate-level popularity of parties and politicians. It resembles attempts at using digital trace data as a sensor of

off-line phenomena in various areas, such as the spread of diseases (e.g., Ginsberg et al., 2009), earthquakes (e.g., Sakaki, Okazaki, & Matsuo, 2010), consumer prices (e.g., Choi & Varian, 2009), or stages in televised sports events (e.g., Chakrabarti & Punera, 2011).

Irrespective of target concepts, previous research utilized various Twitter-based metrics. Some studies use total counts or shares of mentions of political actors (e.g., Bermingham & Smeaton, 2011; DiGrazia et al., 2013; Fink et al., 2013; Gaurav et al., 2013; Skoric et al., 2012; Soler et al., 2012; Tjong et al., 2012; Tumasjan et al., 2010). Others suggest mentions classified as positive (e.g., Bermingham & Smeaton, 2011; Ceron et al., 2014, 2015; Fink et al., 2013; O'Connor et al., 2010; Thapen & Ghanem, 2013), while still others propose counting the number of users mentioning a political actor (e.g., Tjong et al., 2012). Beside these dominant approaches, others exist, such as using shares of mentions posted by specific users (McKelvey et al., 2014), mentions followed by a word indicating political success or failure (Jensen & Anstead, 2013) or trends in the mentions of candidates over time (Franch, 2013). Recently, scholars suggested more refined and complicated Twitter-based indicators. Ceron, Curini, and Iacus (2015) proposed to focus on all mentions accompanied by explicit declarations of voting intention, mentions explicitly supporting a candidate accompanied by a hashtag connected to his or her party or campaign and mentions explicitly opposing a candidate but using a hashtag connected to a different party or campaign. In contrast, Tsakalides et al. (2015) suggested combining various Twitter-based metrics with past opinion polls to predict future election results. Others use the occurrence of words statistically associated with polls or election results in the past (e.g., Contractor & Faruque, 2013; Marchetti-Bowick & Chambers, 2012).

This brief overview shows that this field grew in the number of contributions as well as in the multiplicity of indicators. Yet, looking at the literature with criteria from indicator validation in mind, little change and variation emerges. First, most available studies showing links between election results or public opinion indicators and Twitter-based metrics established that relationship after the fact. The studies thus do not provide forecasts but show that after the fact there is the possibility to identify a statistical relationship between two known variables in which one—Twitter-based metrics—may be interpreted as predicting the other—election results or opinion polls (cf. Gayo-Avello, 2013; Huberty, 2015; Metaxas et al., 2011). This ex-post strategy is somewhat at odds with hypothesis testing as proposed by social science methodology and might bias results in favor of the indicator claim, that is, lead to false positives.

Second, prior research focuses on convergent validity but does not consider discriminant validity. Virtually all studies claiming that Twitter-based metrics are useful as indicators of political support only explore statistical associations between the proposed indicator and established indicators of the respective target concept, that is, election results or opinion polls. They do not examine, however, the relationship between the proposed indicator and alternative concepts. They thus cannot rule out the possibility that a Twitter-based indicator measures a concept different but potentially correlated to political support and, therefore, to its Twitter-based indicator. Put differently, focusing on convergent validity while dismissing discriminant validity, once again, increases the likelihood of false positives.

Third, prior research has not spent much effort on establishing a theoretical link between the proposed target concept and the proposed indicator. From a methodological perspective this is a key element in the debate, and yet it is absent from most contributions. Still, a few studies illustrate potential theoretical linkages implicitly underlying proponents' arguments. According to one line of reasoning, mentions of a party or candidate on Twitter indicate political support (e.g., Tumasjan et al., 2010). In a nuanced version, this interpretation is limited to those mentions with positive sentiment (e.g., O'Connor et al., 2010). Aggregating mentions thus results in a valid estimate of support for parties and politicians. The second line of reasoning argues that mentioning a party or a politician on Twitter indicates attention, rather than support, with high shares of public attention pointing to success at the polls (e.g., DiGrazia et al., 2013; O'Connor et al., 2010).

Viewed in light of the discussion above, the coexistence of these two lines of reasoning is telling. While the first suggests that Twitter-based metrics are indicators of political support, the second implies that Twitter-based metrics measure political attention, which, the authors argue, might serve as a valid indicator of political support for specific parties or politicians. The latter notion suggests that prior research's ignorance of discriminant validity is more troubling than already suggested. It underscores the possibility that Twitter-based metrics capture political attention, rather than political support, which in some contexts might turn out to be correlated with political support. Yet, a stable relationship between political attention and political support is unlikely since negative events, such as political scandals, as well as positively evaluated events, such as accomplishments, can underlie attention for a party or candidate. In face of these competing arguments, it is crucial to determine which of the two measurement theories—Twitter as indicator of political support or Twitter as indicator of attention toward politics—is more convincing.

The tweets-as-support argument becomes even more strenuous once we examine a further underlying assumption. Political support is a latent attitude usually measured in surveys in a response to a direct stimulus getting a respondent to reflect on their evaluations of politicians or parties (e.g., Tourangeau, Rips, & Rasinski, 2000). Using Twitter data to infer political support implicitly assumes that the act of posting a message mentioning a politician or party in a specific way resembles the act of answering a well-specified survey question. This assumption is implausible (cf. Diaz et al., 2014; Schoen, Gayo-Avello, Metaxas, Strohmaier, & Gloor, 2013). It is true that Twitter data are observational data and might, therefore, be free of some specific biases that might afflict survey data. Still, it is conceptually hard to argue that Twitter mentions—which arise in reaction to diverse stimuli, such as messages posted by other users, mass media content, or real-world events—tell more than the simple fact that a user paid attention to a specific element of political reality at a given time (cf. Jungherr et al., 2016; Lin, Margolin, Keegan, & Lazer, 2013).

This argument receives additional support from prior research on day-to-day volatility in political mentions on Twitter. Various studies have shown that the daily volume of Twitter messages referring to candidates or parties fluctuates heavily depending on the events of the day—such as televised leaders' debates, high-profile interviews with candidates—or the coverage of political controversies and scandals (cf. Jungherr, 2014, 2015). These aggregate-level findings on Twitter communication do not fit with evidence on aggregate-level dynamics of support for political parties and candidates testifying to much less volatility than that Twitter time series exhibit (e.g., MacKuen, Erikson, & Stimson, 1989; Whiteley, Clarke, Sanders, & Stewart, 2013, pp. 55–89). In comparison, political attention exhibits considerably higher levels of variation. Thus, the characteristics of aggregate-level Twitter-based metrics suggest that they are more likely to be suitable as indicators of political attention to political actors than their overall popularity or vote shares on Election Day.

We also have to keep in mind that Twitter's user base is highly skewed and far from being representative of the population at large (e.g., Barberá & Rivero, 2014; Duggan & Smith, 2013; Huberty, 2015; Rainie, Smith, Schlozman, Brady, & Verba, 2012; Vaccari et al., 2013). Accordingly, even if Twitter messages were valid measures of political preferences at the individual level, aggregate-level distributions of political preferences would hardly reflect popular support in the population as a whole. Likewise, bias in the composition of Twitter users may lead to differences in the topics on which Twitter users and the general public focus their attention. Accordingly, it cannot be taken for granted that Twitter-based metrics lead to valid conclusions about public attention in the population at large. To be sure, advanced techniques may permit to draw valid inferences about public opinion from skewed samples (Wang, Rothschild, Goel, & Gelman, 2014). This approach, however, depends on knowledge about the characteristics of respondents included in the sample and a reasonably stable relationship between sample and population. Both preconditions are not met in the case of Twitter (Diaz et al., 2014). Thus, drawing inferences about political support and attention in the general public from Twitter-based metrics runs the risk of leading to flawed conclusions.

Building on this discussion, we suggest that research in this field should stick more closely to standards established in social science methodology, thereby integrating various perspectives voiced by critics of this research (e.g., Diaz et al., 2014; Gayo-Avello, 2011, 2013; Huberty, 2015; Metaxas et al., 2011). In this vein, our analysis of the characteristics of Twitter mentions implies some skepticism toward the notion that Twitter-based metrics are valid indicators of aggregate levels of political support in a society. Given the nature of political talk on Twitter, we expect Twitter-based metrics to reflect much more accurately objects of attention of politically vocal Twitter users mediated by the confounding factors listed above but also individual-level processes (cf. Jungherr, 2015; Jungherr et al., 2016). It is reasonable to assume that in some cases public attention toward politics might be correlated with political support. Still, we anticipate this relationship to be far from stable, given the fickle nature of public attention and the stability of political support. Having identified a potentially alternative target concept, we can examine both convergent and discriminant validity, thereby avoiding premature conclusions. For the same purpose, it is also warranted to include a number of Twitter-based metrics which, at least to some extent, reflect the variety of indicators proposed in prior research.

In the remainder of the article, we will employ this validation strategy while focusing on political support and Twitter mentions of political actors during the campaign for Germany's 2013 federal election. German elections have featured strongly in the relevant literature, with Germany being the first prominent case of seemingly successfully linking Twitter-based metrics to election results (Tumasjan et al., 2010) as well as providing a powerful example for this approach's fragility (Jungherr, 2013; Jungherr et al., 2012). Germany's system of proportional representation and its party system's high fluidity—with traditional actors dropping out of Parliament (i.e., Freie Demokratische Partei [FDP], 2013) or new political actors coming close to entering it (i.e., Pirate Party 2009 and 2013, Alternative für Deutschland [AfD] 2013)—make it a case in which the presupposed link between public attention and political support is strenuous at best. Germany, therefore, is an excellent case to assess if Twitter-based metrics are indeed suitable to measure political support.

Data Collection and Preparation

For data collection, we relied on the social-media-data vendor *Gnip* (<http://gnip.com>). To establish an initial data set covering politically relevant messages posted during the campaign for the German federal election 2013, we queried the *Gnip Historical Powertrack* for messages containing the names of political parties, candidates, campaign-related phrases, and key words related to campaign-related media events. This initial data set covering all public Twitter messages containing the queried key words (see Online Appendix 1)¹ includes 6,677,795 messages posted by 1,248,667 users mentioning at least one of the words on our list. We filtered these messages based on their propensity for referring to German politics. Unfortunately, there is no sure way to identify the language of a tweet or the nationality of a user. For this analysis, we decided to include all messages posted by users who had chosen German as interface language in interacting with Twitter. We deliberately chose to not employ a feature of Twitter automatically identifying the language of a tweet through machine learning (<https://dev.twitter.com/overview/api/tweets>), because the details of this proprietary approach are a black box to researchers not affiliated with Twitter and this filter could thus impact the data quality in uncontrollable ways (cf. Ruths & Pfeffer, 2014). While our approach might underestimate the total number of tweets referring to political parties during the campaign, it permits to create a robust data set of relevant messages. This filter resulted in a total of 1,390,571 messages posted by 98,149 users.

We included all mentions of eight political parties in Germany in our analysis. Six parties represented in Parliament from 2009 to 2013: the Conservatives Christlich Demokratische Union Deutschlands (CDU) and their regional Bavarian sister party Christlich-Soziale Union in Bayern

(CSU), the Social Democrats Sozialdemokratische Partei Deutschlands (SPD), the Socialists (Die LINKE), the Green party (Bündnis 90/Die Grünen), and the Liberals (FDP). Two parties with strong followings online but not represented in Parliament: the Euro-critical party AfD and the Pirate Party with a focus on Internet-related policy questions (Piraten).²

To calculate mention counts of parties in key words or hashtags, we identified specific character strings related to each party. We then counted each message including at least one character string related to a specific actor as one mention of that actor. For example, the use of the hashtags #grüne and #bündnis90 in one message is counted as one mention of the party Bündnis 90/Die Grünen although two hashtags refer to the party. Since there was significant variation in the spelling of parties, we collected the most prominent variations in character strings in encompassing concepts (e.g., grüne, gruene, bündnis, buendnis et al.; see Online Appendix 2). Provided a message contained at least one character string included in these collections, we interpreted this as one mention of the party. For the following analyses, we aggregated all mention counts of these concepts for each day between July 1 and September 22.

Some authors prefer to restrict their analyses to tweets mentioning a party or candidate in a positive context. To establish this, they use automated approaches to the sentiment analysis of Twitter messages. Given the low reliability of lexicon-based approaches to automated sentiment analysis of political tweets (cf. González-Bailón & Paltoglou, 2015) and the surprisingly sparse documentation of other procedures (e.g., Ceron et al., 2014, 2015), this essentially introduces a black box in the creation of indicators (cf. Gayo-Avello, 2012). As it stands, this approach replaces identifiable mention counts by probabilistically determined subsets of “positively” valenced mentions without clearly spelling out the auxiliary measurement theory that links positively valenced Twitter mentions of parties or politicians to political support. Given this added uncertainty introduced by this approach, we decided to test three different measures of sentiment. First, we hand coded 1% of all party mentions in key words or hashtags excluding retweets (6,479 tweets) on the thus identified tweet containing negative, neutral, or positive sentiment toward the mentioned party. Second, we used the Hopkins/King approach to automated content analysis (Hopkins & King, 2010), which has been used by some researchers in the past to identify sentiment in political tweets (Ceron et al., 2014, 2015; for details on these see Online Appendix 3). Third, we analyzed hashtags that German Twitterers used to identify the valence of a tweet toward a mentioned party. These hashtags consist of a party’s name followed by a “+” or a “-” (e.g., #cdu+ or #cdu-). For our analysis, we counted all occurrences of the character strings identified in Online Appendix 2 preceded by a hashtag (#) and followed by a “+” or “-” sign to measure positive sentiment. This procedure has the benefit that our indicator—positive mentions—is based on an exact count, not introducing additional untested assumptions into the measurement process. Yet, it only allows us to capture sentiment expressed in tweets through hashtags. This might introduce a considerable error given that supporters of specific parties—such as the Pirates or the AfD both of which emphasized the role of Twitter in their campaign—might be more prone to use them than supporters of parties putting less emphasis on Twitter. Still, by using three independent assessments of sentiment contained in tweets, our approach is a reasonable strategy to allow an assessment of the link between political support and Twitter mentions’ sentiment.

One could object that this selection does not cover all Twitter-based metrics suggested in the literature and, therefore, allows no general assessment of the validity of each Twitter-based metric. Yet, this objection is not valid. The selected measures allow a direct assessment of the relationship between central types of Twitter-based metric, that is, mentions, positive mentions, and users. These measures lend themselves to intuitive comparisons with election and polling results and permit a straightforward test of the relationship between Twitter and measures of political support. This relationship lies at the core of any approach used to infer political support from Twitter-based metrics, be it the simple counting of tweets or sophisticated machine learning approaches. The

Table 1. Sums of Party Mentions and Users.

Party	Number of Mentions (Key Words)	Number of Users (Key Words)	Number of Mentions (Hashtags)	Number of Users (Hashtags)	Number of Mentions (Positive Sentiment)	Number of Users (Positive Sentiment)	Number of Mentions (Negative Sentiment)	Number of Users (Negative Sentiment)
CDU	74,333	21,225	53,839	12,003	3,133	714	13,143	2,990
SPD	80,203	19,884	54,587	11,304	2,793	579	7,528	1,889
Die LINKE	51,024	15,767	24,754	5,408	4,033	542	740	314
Die Grünen	36,136	12,073	29,025	6,761	688	297	5,076	1,249
CSU	32,260	11,397	25,266	6,569	748	180	6,428	1,460
FDP	76,969	22,831	51,984	14,236	2,220	426	10,064	2,309
AfD	46,912	14,296	40,379	8,521	24,845	1,010	24,929	2,803
Piraten	64,558	14,172	114,710	10,497	23,154	3,115	2,261	479
Total number of tweets	344,520	49,723	304,772	30,312	60,341	5,630	53,873	6,169

Note. The table documents total counts of various Twitter-based metrics between July 1 and September 22, 2013. CDU = Christlich Demokratische Union Deutschlands; SPD = Sozialdemokratische Partei Deutschlands; CSU = Christlich-Sozialer Union in Bayern; FDP = Freie Demokratische Partei; AfD = Alternative für Deutschland.

general applicability of each of these approaches depends on the nature of this relationship illustrated here by the metrics included in the analysis.

During the period of observation, 49,723 users used at least one of the selected key words to refer to a political party in at least one of their messages. A total of 30,312 users used at least once one of the hashtags included in our analysis to refer to parties. In all, 344,520 messages contained key words referring to parties, while 304,772 included hashtags referring to parties (see Table 1); 5,630 and 6,169 users posted 60,341 and 53,873 messages referring positively or negatively to parties, respectively.

Table 2 shows that the distributions of the identified time series follow similar dynamics. All distributions have relatively high maxima—usually on days with high-profile political events such as the televised candidate debate or state elections—and comparatively modest median values. For example, on half of the days between July 1 and September 22, 2013, less than 341 users mentioned the CDU in a hashtag. Accordingly, despite some days with intense activity, on most days political Twitter activity was rather limited.

One way to assess the validity of an indicator in measuring a latent concept is by comparing it with established measures (cf. Adcock & Collier, 2001). With regard to political support, two comparative metrics offer themselves: First, official election results; second, results from opinion polls. To assess election results, we use the number of votes each party received as reported by the *Bundeswahlleiter* (<http://www.bundeswahlleiter.de>). We collected results of opinion polls on the website *Wahlen, Wahlrecht und Wahlsysteme* (<http://www.wahlrecht.de>). We included polls with field times between July 1 and September 22 published by seven different institutes *Forsa*, *Forschungsgruppe Wahlen*, *Gesellschaft für Markt- und Sozialforschung (GMS)*, *Infratest dimap*, *INSA-Consulting (INSA)-Meinungstrend*, *Institut für Demoskopie Allensbach*, and *TNS Emnid*. As a baseline for nowcasting, we assigned the published results to the respective field periods. In case of overlapping field periods, we averaged the results of the respective parties across the available polls. Using published results of opinion polls is problematic since institutes use various adjustment procedures to their polls before publishing them. Still, this is the best publicly available source

Table 2. Maxima and Medians of Party Mentions and Users.

Party	Counts Mentions (Key Words)		Counts Users (Key Words)		Counts Mentions (Hashtags)		Counts Users (Hashtags)		Counts Mentions (Positive Hashtags)		Counts Users (Positive Hashtags)		Counts Mentions (Negative Hashtags)		Counts Users (Negative Hashtags)	
	max	median	max	median	max	median	max	median	max	median	max	median	max	median	max	median
CDU	13,573	612.0	7,327	487.0	5,168	501.5	3,122	341.0	755	15.0	288	11.0	396	134.0	254	91.0
SPD	9,413	751.0	4,902	582.0	3,612	543.0	2,179	358.0	196	26.5	102	18.5	248	78.0	180	55.0
Die LINKE	7,715	431.0	4,530	347.5	2,393	202.5	1,451	139.0	223	34.0	82	18.0	64	7.5	58	6.0
Die Grünen	6,275	264.5	3,776	238.5	2,199	252.5	1,575	190.0	36	6.0	32	6.0	160	55.0	130	35.5
CSU	5,593	234.0	3,217	189.5	2,697	239.0	1,745	158.5	71	6.0	44	3.0	365	69.0	245	45.5
FDP	18,465	434.0	9,726	350.0	10,323	319.0	5,836	229.0	109	20.0	58	16.0	632	92.5	362	57.5
AfD	15,779	231.5	8,443	164.5	6,732	266.0	3,653	142.0	994	267.5	233	59.5	2,087	257.0	1,040	48.5
Piraten	5,231	639.0	3,399	517.5	5,878	1,232.5	2,833	615.0	1,095	211.5	572	147.0	99	20.5	77	12.0

Note. The table documents maximum and median values of time series of various Twitter-based metrics between July 1 and September 22, 2013. Maxima document the highest mention or user count for each time series, while median values document the number of messages or users that during half of the time was exceeded or undershot, respectively. CDU = Christlich Demokratische Union Deutschlands; SPD = Sozialdemokratische Partei Deutschlands; CSU = Christlich-Soziale Union in Bayern; FDP = Freie Demokratische Partei; AfD = Alternative für Deutschland.

on political support at any given time for a political party. Also, various studies have used poll results as baseline to test their indicators, making them a useful baseline for validating of Twitter metrics for the analysis of political support.

Findings

Forecasting Election Results: AfD and Pirates Take the Lead

In assessing the validity of an indicator in measuring a concept, one standard approach is comparing the indicator's measurements with those of established indicators for the same concept (cf. Adcock & Collier, 2001). Here, we compare seven Twitter-based metrics of parties with their respective vote shares (Table 3). To examine the accuracy of these indicators, we followed prior research in employing the Mean Absolute Error (MAE), that is, the average of the absolute errors of the forecasts for each variable (the difference between the actual vote share and the prediction based on various Twitter-based metrics; Hyndman & Koehler, 2006).

The results reported in Table 3 suggest that the overall validity of Twitter-based mentions as indicators of political support as expressed in votes is rather poor. Using the share of users mentioning a party—as suggested, for example, by Tjong, Sang, and Bos (2012)—would have led researchers to overestimate the results of all parties. Users regularly mentioned more than one party in their tweets. Aggregating the counts of users by party thus exceeds the total count of users mentioning political parties, leading to an overestimation of parties' vote shares. This is also true for the shares of users mentioning a party in a hashtag. This renders metrics based on comparing shares of users mentioning a party a dubious information source as they tend to significantly overestimate the relative importance of political actors offline. The literature offers no convincing proposition of a normalization process that might solve this problem. For example, Tjong et al. (2012) proposed counting only the first mention of a single party in the corpus by a user. As this approach appears arbitrary, we decided not to replicate this process.

Measurements based on aggregates of mention or hashtag counts also performed poorly. As the differences between shares in key word and hashtag mentions of all party mentions and actual vote share indicate, key word mentions exaggerated the vote shares of the Liberals, the Left, the Pirates, and the AfD while underestimating the Conservatives' and the Social Democrats' vote shares. The reason for the relative prominence in key word mentions of Liberals probably reflects the heightened media attention toward the party following their unsuccessful showing in the state elections in Bavaria, shortly before the federal election. In contrast, the relative prominence of the Pirates and the AfD in key word mentions is probably due to the Internet-savvy supporter group of the Pirates and the high levels of controversy the AfD created online. Hashtag mention share of political parties showed similar dynamics while more strongly overestimating the vote share of the Pirates, speaking again to the relative importance of the internet-savvy supporters of the Pirate Party. In general, the high levels of measurements' MAE based on raw mention counts of key word and hashtag mention shares just below 10% points indicate that raw mention counts of political actors had little if any direct potential to draw inferences about their subsequent electoral fortunes.

Measurements based on positive sentiment toward a party fail, as well. This is true for all three sentiment measures tested by us. Again, this approach leads to a massive overestimation of support for the AfD and the Pirates while significantly underestimating levels of support for CDU and SPD. Accordingly, including various sentiment measures does not improve the Twitter-based measurement of the result of the 2013 election.

The analyses covered an arbitrarily selected period, that is, July 1 to September 22. One may thus object that other periods might have resulted in more accurate measurements. To address this, we exemplarily report in Figure 1³ the MAEs of forecasts based on hashtag and key word shares for each

Table 3. User and Mention Shares Between July 1 and September 22, 2013 Compared to Vote Share.

Party	Vote Share	Vote Share (2009)—Vote Share (2013) ^a	Poll Results ^b —Vote Share	User Share, Key Words—Vote Share		User Share, Hashtags—Vote Share		Key Word Share—Vote Share		Hashtag Share—Vote Share		Positive Sentiment Share (Hand Coded)—Vote Share		Positive Sentiment Share (Hopkins/King)—Vote Share		Positive Sentiment Share (#+)—Vote Share
				Vote Share	Vote Share	Vote Share	Vote Share	Vote Share	Vote Share	Vote Share	Vote Share	Vote Share	Vote Share	Vote Share	Vote Share	
CDU	34.1	-5.67		+8.59	+5.50	-18.02	-20.45	-26.61	-27.24	-29.02						
CDU/CSU			-1.5													
SPD	25.7	-1.69	+0.2	+14.29	+11.59	-8.35	-11.86	-13.23	-18.61	-21.17						
Die LINKE	8.6	+3.79	-0.6	+23.11	+9.24	+2.43	-2.33	18.18	4.62	-2.05						
Die Grünen	8.4	+2.76	+4.6	+15.88	+13.90	-0.59	-1.04	-0.79	-5.35	-7.28						
CSU	7.4	-0.60		+15.52	+14.27	-0.42	-1.00	-1.75	-5.10	-6.19						
FDP	4.8	+10.38	+0.2	+41.12	+42.16	+11.85	+8.38	0.81	-0.80	-1.20						
AfD	4.7		-1.7	+24.05	+23.41	+5.44	+5.53	26.12	20.47	+35.62						
Piraten	2.2	-0.16	+0.8	+26.30	+32.43	+11.76	+26.87	28.95	36.11	+35.38						
MAE	—	3.58	1.07	21.11	19.06	7.36	9.68	13.08	14.79	16.77						

Note. The table reports vote shares of the parties included in the analyses in the federal election on September 22, 2013. Vote shares are compared to aggregates of various Twitter-based metrics to allow for an assessment of their predictive power. Vote share reflects the share of votes for each party on the total of all votes collected by all parties included in the analysis. This might lead the vote shares reported here to deviate from the official results. Mean absolute errors are calculated for each metric based on the comparisons between shares in Twitter-based metrics and vote share for each party. For an overview of the various mention shares the Online Appendix 4. CDU = Christlich Demokratische Union Deutschlands; SPD = Sozialdemokratische Partei Deutschlands; CSU = Christlich-Soziale Union in Bayern; FDP = Freie Demokratische Partei; AfD = Alternative für Deutschland; MAE = mean absolute error.

^aNot including the AfD as the party did not stand in the 2009 election. ^bWe calculated the expected vote share predicted by polls based on the median value of the polling results for each party between July and September 1.

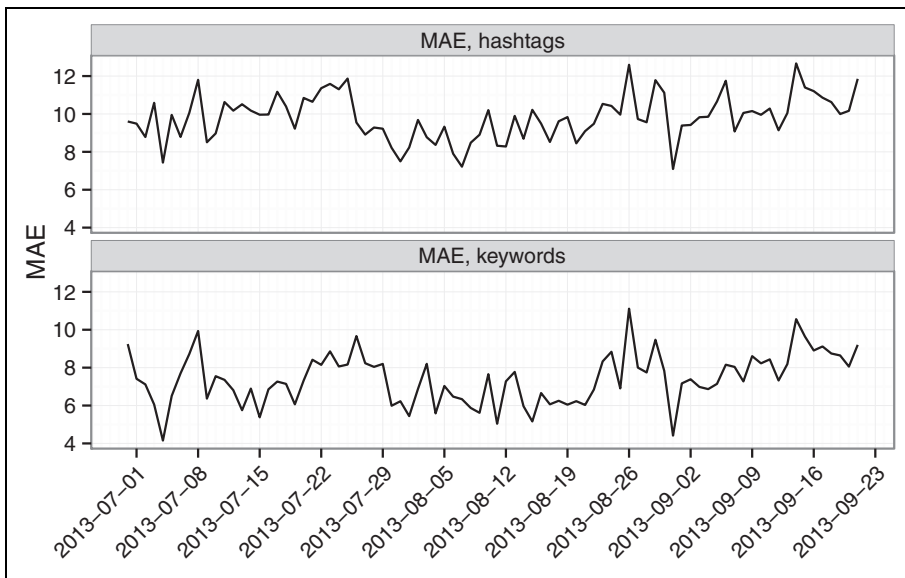


Figure 1. Mean absolute errors daily Twitter-based forecasts of election results over time.

day in this period. Both time series fluctuate noisily, indicating that Twitter-based measurements heavily depend on the time span chosen. They also raise some concerns whether these measures are valuable as reliable and valid indicators of concepts that do not presuppose high volatility. At the same time, the evidence shows that choosing a different time span would not have improved the quality of the forecasts.

These findings raise serious doubts about the validity of Twitter-based measures of political support. Absolute counts of various metrics of mentions of political parties on Twitter were skewed toward political parties at the center of controversies, media attention, and those with Internet-savvy supporters.

Nowcasting Opinion Polls: Stability Versus Attention-Based Fluctuations

Assessing the validity of Twitter-based metrics for the measurement of current levels of political support is possible by comparing daily levels in Twitter-based metrics with opinion poll results, that is, an established indicator of political support. Given previous findings, we expect these time series to diverge. Parties on which the media’s attention increasingly focuses, for example, because of controversies or surprising events, might witness an increase in Twitter mention volume, irrespective of their levels of public support as expressed in opinion polls. To explore the validity of these hypotheses we analyzed the dynamics of the time series of polling results and daily mention shares of political parties from July 1 to September 22. The analysis focuses on time series of mention shares (key words and hashtags). Given the overestimation of vote shares based on shares of users mentioning a party (see Table 3), we excluded these metrics from this analysis.

Starting with the variance of the time series of Twitter metrics and opinion polls, we find a striking difference (Table 4). The opinion polls vary mildly over time, whereas Twitter metrics exhibit considerable variation. Accordingly, the time series appear to follow different dynamics. This pattern suggests that the time series of Twitter-based metrics and opinion polls are unlikely to be strongly correlated.

To explore this expectation, we estimated three correlations between both kinds of time series, each representing specific assumptions concerning the relationship between polls and communication on Twitter. First, we analyzed the correlation between each Twitter-based metric and opinion polls with no time lag to find out whether both series evolved simultaneously. To find out whether

Table 4. Comparison Between Results of Opinion Polls and Various Twitter-Based Metrics Between July 1 and September 22, 2013.

Party	Metrics	Intercepts	Slopes (SE)	Variance	Corr. Polls (lag -1)	Corr. Polls (lag 0)	Corr. Polls (lag +1)
CDU/ CSU	Poll results	40.608*** (0.107)	-.016*** (0.002)	1.107	—	—	—
	Key word mentions	19.210*** (0.606)	-.013 (0.012)	7.584	.142	.025	.083
	Hashtag mentions	19.850*** (0.597)	-.062*** (0.012)	9.568	.027	-.031	.023
	Positive sentiment (#+) ^a	4.059*** (0.964)	.025 (0.025)	19.320	.081	.048	-.074
SPD	Poll results	24.320*** (0.171)	.013*** (0.004)	2.268	—	—	—
	Key word mentions	18.420*** (0.786)	-.058*** (0.016)	14.580	.174	-.053	.147
	Hashtag mentions	13.030*** (0.671)	-.020 (0.014)	9.402	.132	-.090	-.045
	Positive sentiment (#+)	5.056*** (0.507)	-.010 (0.010)	5.293	.203	-.181	.150
Die LINKE	Poll results	7.211*** (0.079)	.016*** (0.002)	0.821	—	—	—
	Key word mentions	9.506*** (0.460)	-.001 (0.009)	4.320	.172	.063	.267
	Hashtag mentions	3.702*** (0.283)	.030*** (0.006)	2.158	.138	.088	.177
	Positive sentiment (#+)	7.028*** (0.774)	-.009 (0.016)	12.260	-.174	-.047	-.089
Die Grünen	Poll results	14.520*** (0.178)	-.054*** (0.004)	2.843	—	—	—
	Key word mentions	5.218*** (0.324)	.019*** (0.007)	2.344	-.273	-.062	-.170
	Hashtag mentions	4.483*** (0.258)	.033*** (0.005)	2.001	-.564	-.294	-.222
	Positive sentiment (#+)	1.542*** (0.225)	-.007 (0.005)	1.057	.111	.261	.181
FDP	Poll results	5.062*** (0.069)	.007*** (0.001)	0.355	—	—	—
	Key word mentions	8.604*** (0.803)	.065*** (0.016)	15.640	-.004	.042	-.010
	Hashtag mentions	4.733*** (0.699)	.093*** (0.014)	14.330	-.017	.165	.080
	Positive sentiment (#+)	3.927*** (0.407)	-.005 (0.008)	3.389	-.067	-.203	-.328
AfD	Poll results	2.497*** (0.138)	.019*** (0.003)	0.534	—	—	—
	Key word mentions	1.384*** (0.511)	.106*** (0.010)	12.000	.175	.195	.279
	Hashtag mentions	1.345*** (0.573)	.131*** (0.011)	16.130	.336	.318	.291
	Positive sentiment (#+)	36.770*** (2.092)	.079* (0.043)	92.920	.212	.221	.212

(continued)

Table 4. (continued)

Party	Metrics	Intercepts	Slopes (SE)	Variance	Corr. Polls (lag -1)	Corr. Polls (lag 0)	Corr. Polls (lag +1)
Piraten	Poll results	2.699*** (0.062)	-.0001 (0.001)	0.322	—	—	—
	Key word mentions	18.490*** (0.759)	-.105*** (0.016)	18.270	-.089	-.109	-.067
	Hashtag mentions	33.110*** (1.003)	-.144*** (0.020)	32.830	-.171	.015	-.150
	Positive sentiment (#+)	41.620*** (1.662)	-.074** (0.034)	59.540	.120	-.001	-.120

Note. The table shows a comparison between descriptive metrics of time series of opinion poll results and various Twitter-based metrics. Variance documents the variance of each time series between July 1 and September 22, 2013. Correlation polls (lag 0) documents the correlation between opinion polls and each Twitter-based time series. Correlation polls (lag -1) documents the correlation between opinion polls and each Twitter-based time series with Twitter-based time series lagged by 1 day. Correlation polls (lag +1) documents the correlation between opinion polls and each Twitter-based time series with opinion polls lagged by one day. Intercepts and slopes document these metrics for each time series regressed over time. In this analysis, we combine mentions of Christlich Demokratische Union Deutschlands and Christlich-Soziale Union in Bayern to allow a comparison with opinion polls, which only reported results for both parties combined. Standard errors are given in parentheses.

^aHere, we only use tweet sentiment as indicated by the use of positive hashtags (i.e., #cdu+ or #spd+) as our other sentiment analysis approaches did not provide us with enough information to calculate sentiment on a daily basis. CDU = Christlich Demokratische Union Deutschlands; SPD = Sozialdemokratische Partei Deutschlands; CSU = Christlich-Soziale Union in Bayern; FDP = Freie Demokratische Partei; AfD = Alternative für Deutschland.

*p < .1. **p < .05. ***p < .01.

Twitter metrics mirrors preceding changes in offline support of political parties (e.g., O’Connor et al., 2010), we lagged opinion polls by 1 day. Finally, to account for the notion that Twitter users might serve as opinion leaders subsequently affecting off-line public opinion (Tumasjan et al., 2010), we lagged Twitter metrics by 1 day.

The evidence reported in the right-hand columns of Table 4 suggests that there is no systematic relationship between Twitter-based metrics and opinion polls, irrespective of the assumptions concerning the temporal and causal relationship. Accordingly, relying on Twitter as an indicator of the evolution of party support in the 2013 election campaign on a day-to-day basis would have led to flawed conclusions. Utilizing larger lags does not alter the substantive findings (see Online Appendix 5, documenting lags -4 to +4).

Yet, this criterion may be considered overly demanding. One may argue that Twitter-based metrics exhibited the same overall trend as opinion polls during campaign periods. Accordingly, once the differences in the parties’ online popularity are accounted for at the start of the campaign period, Twitter-based metrics might permit to capture the trend in party support in the campaign period (e.g., Franch, 2013; Jensen & Anstead, 2013). Provided similar overtime trends, future values of a time series of Twitter metrics might be used to predict unobserved future values in time series of opinion polls. To examine this hypothesis, we utilized time as a predictor of poll results and Twitter metrics.

The results reported in the left-hand columns of Table 4 indicate that the intercepts of poll-based and Twitter-based regression models differ considerably, implying some base line differences between offline support and Twitter metrics. However, the differences depend on party and Twitter-based indicator. For example, the key-word-based SPD intercept resembles quite closely the poll-based one, whereas for other parties like CDU/CSU, Die Grünen, FDP, and the Pirates this does not hold.

For the trend model, the regression slopes are more important. Still, the poll-based slopes reported in Table 4 do not closely resemble the Twitter-based estimates. To be sure, in some

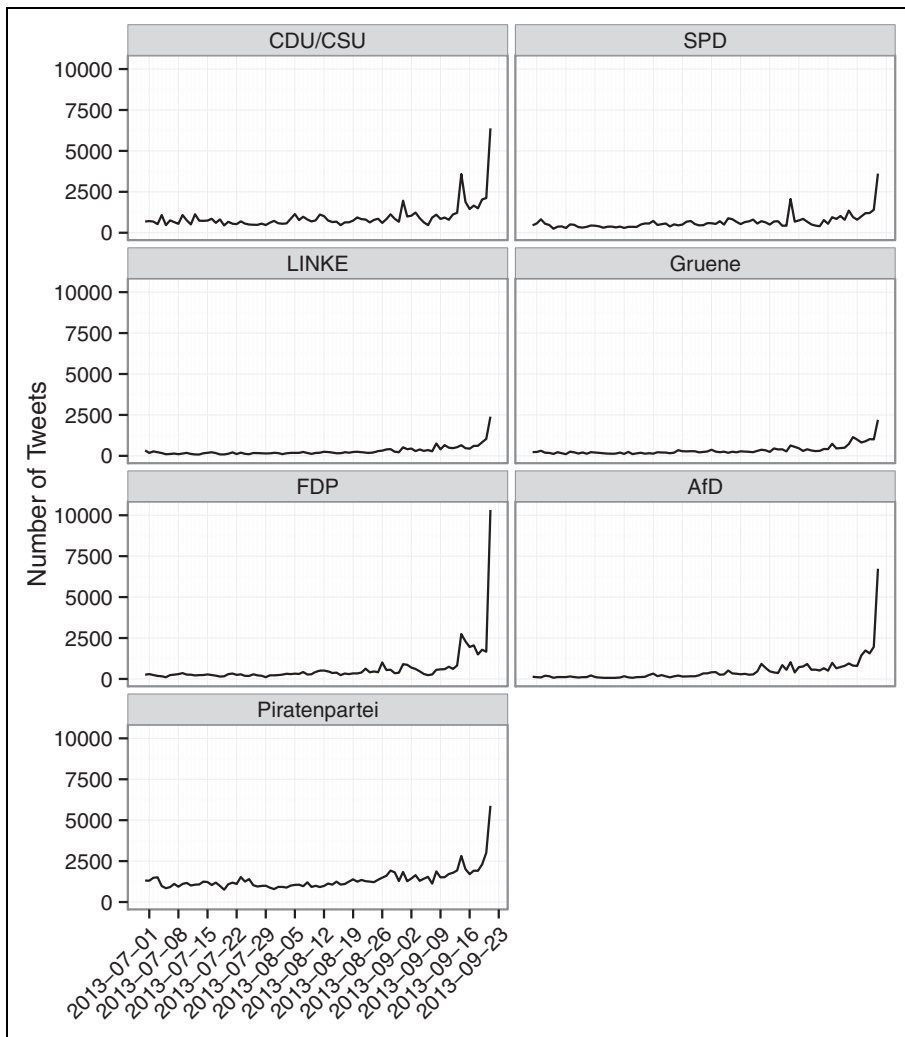


Figure 2. Party mentions, hashtags between July 1 and September 22, 2013.

instances a specific Twitter-based estimate is quite close to the poll-based one. But for other parties, the same Twitter-based slope is off the mark. We thus conclude that the evidence provides little, if any, support for the notion that the direction and dynamics of slopes of regression models based on Twitter metrics resemble those resulting from the analyses of opinion polls.

The evidence on convergent validity thus suggests that Twitter-based metrics are no valid indicator for public support of parties during the 2013 campaign. Twitter-based metrics did not correspond with opinion polls, irrespective of the criterion applied. As we argued above, the problems in establishing Twitter metrics as indicators of political support might reflect the fact that these metrics are better suited to capture political attention which at times might be correlated with political support. To explore this possibility, we take a closer look at the time series of Twitter-based metrics. For the sake of brevity, we focus on hashtag mentions of political parties. Relying on other Twitter-based metrics, however, leads to the same substantive conclusions (see Online Appendix 6).

The results reported in Figure 2 show similar dynamics for all parties. Until September, the time series fluctuate more or less stably around a base line. From early September onwards, the baselines

of mentions of all parties rise until Election Day, corresponding with the increasing intensity of the campaign. At the beginning of the time series politics is only one topic among many others and parties attract small but rather stable mention counts. Once the campaign draws closer to Election Day and thereby moves to the forefront of public attention, the mention levels of parties rise.

For virtually all parties, we observe days on which they received unusually high mention counts. These are typically days of particular high political significance. For example, the hashtag mentions of CDU/CSU and SPD spike strongly on September 1, the day of the televised leaders' debate. Likewise, the mentions of the FDP spike on September 15, the day of state elections in Bavaria. In this election, the Liberals surprisingly could not gather enough votes to enter the State Parliament, giving rise to speculations whether they would also fail in the federal election 1 week later. These ongoing speculations are mirrored by the high mention counts of the FDP during the last week of the campaign.

The evidence thus suggests that mention levels of political parties were largely driven by public attention to politics; in particular, media appearances of leading candidates, controversies, and public debate seem to have played a considerable role.

The evidence thus suggests that Twitter mention shares fluctuate because of various underlying phenomena which appear to be related to public attention toward politics but none of them related to political support. Based on these dynamics, it is doubtful to expect that in the 2013 German federal election campaign Twitter-based metrics of political parties mirrored their electoral fortunes or trends in public opinion polls. If anything, Twitter-based metrics appear to have reflected public attention to a given party, though even in this respect, they are not a perfect indicator (cf. Jungherr et al., 2016).

Conclusion

Twitter-based metrics have gained considerable popularity as indicators of public opinion toward political parties as well as indicators of future election results. Although prior research has produced valuable insights, it has not paid sufficient attention to the classical precepts of social science methodology for the validation of indicators that lend themselves for the validation of Twitter-based indicators for social phenomena. Looked at from this perspective, conceptual and theoretical issues and the role of hypothesis testing and discriminant validity did not receive sufficient attention in prior research on the suitability of Twitter-based measures as indicators of political support. Against this backdrop, we examined the validity of Twitter-based metrics as indicators of political attention and political support of candidates and parties. The evidence on Twitter-based metrics as indicator for political support over the course of the campaign for Germany's federal election provided little, if any, supporting evidence. Instead, the evidence lends considerable credence to the notion that Twitter-based metrics reflect users' attention to politics. Following our line of reasoning, political attention may be a covariate of political support in some cases but not in all circumstances—as our analysis shows.

While the analysis does not support the simple “more tweets, more votes” (DiGrazia et al., 2013, p. 1) formula, it suggests that Twitter holds some potential for public opinion research. Twitter mentions of political actors mirror parts of political reality mediated through the interests and attention of Twitter users (cf. Jungherr, 2015; Jungherr et al., 2016). Likely candidates for political phenomena to create digital traces are political media events, intense media coverage of politics, or public controversies. Accordingly, future research may focus on using Twitter data to analyze which kind of political information attracts Twitter users' attention and is distributed online. This gives rise to important questions concerning the sources, that is, the media, political elites, or social networks, and conditions successful in getting Twitter users to pay attention to political information. Thus,

Twitter has the potential to become a source of insight into conditions and dynamics of attention toward politics.

Our analysis suggests some caution drawing general conclusions about the inferential potential of Twitter-based metrics. We utilized data from a single case characterized by specific conditions—an election in a multiparty system, a rather uneventful campaign with next to no movement in opinion polls, and a comparably low adoption of Twitter in Germany. Our findings thus do not imply that Twitter-based metrics will never be correlated with political support for political parties and candidates. Under different conditions, they might have fared better as indicators of these concepts. To give an example, in 2013 Germany witnessed a nonnormalized political online sphere (Chaffee & Metzger, 2001) as parties dominating the online sphere attracted limited vote shares and public support. Provided a normalized political online sphere (Margolis & Resnick, 2000), that is, the same political actors dominate online as well as in elections, Twitter may appear as a more valuable indicator. The quality of Twitter-based inference may vary across countries as well as over time, be it in the short run, for example, campaign versus noncampaign periods, or in the long run, for example, fundamental changes in the online sphere. Our analysis suggests that their quality depends on the strength of the link between attention toward politics and political support in each given case. Future research might examine this and related hypotheses.

More generally, our analysis suggests that the debate about links between Twitter-based metrics and political phenomena revolves around theoretical issues concerning data-generating processes (see, e.g., Jungherr, 2015; Jungherr et al., 2016; Lazer, Kennedy, King, & Vespignani, 2014). Accordingly, future research may be well advised to focus on theoretical underpinnings, rather than exclusively on empirics. By implication, presenting empirical correlations between increasingly complicated Twitter-based metrics and political phenomena without a sound theoretical account of this link must be considered inconclusive. Looking at these issues from a theoretical perspective, furthermore, suggests manifold micro- and macro-level hypotheses that lend themselves to empirical testing. Future research on the validity of Twitter-based metrics as indicators of off-line phenomena may engage more in theory-driven hypothesis testing, rather than in data-fitting and ex-post rationalization (cf. Schoen et al., 2013). Scholars should acknowledge the conditional nature of findings more freely and be more careful in considering and analyzing the consequences of potentially varying data-generating mechanisms. This might lead Twitter-based research to free itself from inflated early expectations to find proxies of public opinion in Twitter-data and instead focus on the potential of digital trace data in yielding insights into public attention toward political information. Digital trace data may thus provide valuable information for public opinion research, though on different phenomena than those on which prior research focused.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

The online [appendices/data supplements/etc] are available at <http://journals.sagepub.com/doi/suppl/10.1177/0894439316631043>

Notes

1. We provide detailed additional results to some of the analyses presented here in an appendix which can be accessed online at <http://andreasjungherr.net/wp-content/uploads/2016/01/Jungherr-Schoen-Posegga-Jürgens-2016-Digital-Trace-Data-in-the-Study-of-Public-Opinion-An-Indicator-of-Attention-Toward-Politics-Rather-Than-Political-Support—Online-Appendix.pdf>
2. In our analysis, we focused on the mentions of parties instead of candidates as German Twitter users predominantly refer to politics using party names. Using Twitter mentions of candidates as basis for predictions would have performed significantly worse than using party mentions, as mentions focused predominantly on Angela Merkel (CDU) trailed by mentions of her challenger Peer Steinbrück (SPD). Candidates of smaller parties were nearly not mentioned at all. Using Twitter mentions (through key words as well as hashtags) of Angela Merkel to predict the CDU's vote share would have led us to overestimate her vote share by roughly 40 percentage points. Steinbrück's mentions would have led us to underestimate the SPD's share by roughly 10% points.
3. This and the following figures were prepared using *R* (R Core Team, 2014) and *ggplot2* (Wickham, 2009).

References

- Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *The American Political Science Review*, *95*, 529–546.
- Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. In X. J. Huang, I. King, V. Raghavan, & S. Rueger (Eds.), *WI-IAT 2010: Proceedings of the 2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology* (pp. 492–499). Washington, DC: IEEE.
- Barberá, P., & Rivero, G. (2014). Understanding the political representativeness of twitter users. *Social Science Computer Review*. Retrieved from <http://ssc.sagepub.com/content/early/2014/11/27/0894439314558836.abstract>
- Birmingham, A., & Smeaton, A. F. (2011). On using Twitter to monitor political sentiment and predict election results. In S. Bandyopadhyay & M. Okumura, *SAIIP: Proceedings of the workshop sentiment analysis where AI meets psychology at the international joint conference for natural language processing (IJCNLP)* (pp. 2–10). Chiang Mai, Thailand: Asian Federation of Natural Language Processing.
- Bollen, J., Mao, H., & Zeng, X. J. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*, 1–8.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Ceron, A., Curini, L., & Iacus, S. M. (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*, *16*, 340–358.
- Ceron, A., Curini, L., & Iacus, S. M. (2015). Using sentiment analysis to monitor electoral campaigns: Method matters—evidence from the United States and Italy. *Social Science Computer Review*, *33*, 3–20.
- Chaffee, S. H., & Metzger, M. J. (2001). The end of mass communication? *Mass Communication & Society*, *4*, 365–379.
- Chakrabarti, D., & Punera, K. (2011). Event summarization using tweets. In *ICWSM 2011: Proceedings of the 5th international AAAI conference on weblogs and social media* (pp. 66–73). Menlo Park, CA: Association for the Advancement of Artificial Intelligence (AAAI).
- Choi, H., & Varian, H. R. (2009). *Predicting the present with google trends*. Mountain View, CA: Google Inc. Retrieved from http://static.googleusercontent.com/media/www.google.com/de/googleblogs/pdfs/google_predicting_the_present.pdf
- Contractor, D., & Faruque, T. A. (2013). Understanding election candidate approval ratings using social media data. In D. Schwabe, V. Almeida, H. Glaser, R. Baeza-Yates, & S. Moon (Eds.), *WWW 2013: Proceedings of the 22nd international conference on World Wide Web* (pp. 189–190). Geneva, Switzerland: International World Wide Web Conferences Steering Committee.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

- Diaz, F., Gamon, M., Hofman, J., Kıcıman, E., & Rothschild, D. (2014). Online and social media data as a flawed continuous panel survey. Microsoft Research Working Paper. Retrieved from <http://research.microsoft.com/en-us/UM/redmond/projects/flawedsurvey/flawedsurvey.pdf>
- DiGrazia, J., McKelvey, K., Bollen, J., & Rojas, F. (2013). More tweets, more votes: Social media as a quantitative indicator of political behavior. *PLoS One*, 8, e79449.
- Duggan, M., & Smith, A. (2013). Social media update. Pew Internet & American Life Project. Retrieved from <http://pewinternet.org/Reports/2013/Social-Media-Update.aspx>
- Fink, C., Bos, N., Perrone, A., Liu, E., & Kopecky, J. (2013). Twitter, public opinion, and the 2011 Nigerian Presidential election. In R. Bilof (Ed.), *SocialCom 2013: The 5th IEEE international conference on social computing* (pp. 311–320). Washington, DC: IEEE.
- Franch, F. (2013). (Wisdom of the crowds)²: 2010 UK election prediction with social media. *Journal of Information Technology & Politics*, 10, 57–71.
- Gaurav, M., Srivastava, A., Kumar, A., & Miller, S. (2013). Leveraging candidate popularity on Twitter to predict election outcome. In R. L. Grossman & R. Uthurusamy (Eds.), *SNAKDD 2013: Proceedings of the 7th workshop on social network mining and analysis* (Article no. 7). New York, NY: ACM.
- Gayo-Avello, D. (2011). Don't turn social media into another 'Literary Digest' poll. *Communications of the ACM*, 54, 121–128.
- Gayo-Avello, D. (2012). No, you cannot predict elections with Twitter. *IEEE Internet Computing*, 16, 91–94.
- Gayo-Avello, D. (2013). A meta-analysis of state-of-the-art electoral prediction from Twitter data. *Social Science Computer Review*, 31, 649–679.
- Gerring, J. (2012). *Social science methodology* (2nd ed.). New York, NY: Cambridge University Press.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457, 1012–1014.
- González-Bailón, S., & Paltoglou, G. (2015). Signals of public opinion in online communication: A comparison of methods and data sources. *The ANNALS of the American Academy of Political and Social Science*, 659, 95–107.
- Hand, D. J. (2004). *Measurement theory and practice*. London, UK: Wiley.
- Hopkins, D., & King, G. (2010). A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, 54, 229–247.
- Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, 12, 767–797.
- Huberty, M. (2015). Can we vote with our tweet? On the perennial difficulty of election forecasting with social media. *International Journal of Forecasting*, 31, 992–1007.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679–688.
- Jensen, M. J., & Anstead, N. (2013). Psephological investigations: Tweets, votes, and unknown unknowns in the republican nomination process. *Policy & Internet*, 5, 161–182.
- Jungherr, A. (2013). Tweets and votes, a special relationship: The 2009 federal election in Germany. In I. Weber, A. M. Popesc, & M. Pennacchiotti (Eds.), *PLEAD 2013: Proceedings of the 2nd workshop politics, elections and data* (pp. 5–14). New York, NY: ACM.
- Jungherr, A. (2014). The logic of political coverage on Twitter: Temporal dynamics and content. *Journal of Communication*, 64, 239–259.
- Jungherr, A. (2015). *Analyzing political communication with digital trace data: The role of Twitter messages in social science research*. Heidelberg, Germany: Springer.
- Jungherr, A., Jürgens, P., & Schoen, H. (2012). Why the pirate party won the German election of 2009 or the trouble with predictions: A response to Tumasjan, A., Sprenger, T.O., Sander, P.G. & Welpe, I.M. "Predicting elections with Twitter: what 140 characters reveal about political sentiment." *Social Science Computer Review*, 30, 229–234.

- Jungherr, A., Schoen, H., & Jürgens, P. (2016). The mediation of politics through Twitter: An analysis of messages posted during the campaigns for the German federal elections 2009 and 2013. *Journal of Computer-mediated Communication, 21*, 50–68.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*, 527–535.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google flu: Traps in big data analysis. *Science, 343*, 1203–1205.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., . . . Alstyne, M. V. (2009). Computational social science. *Science, 323*, 721–723.
- Lin, Y. R., Margolin, D., Keegan, B., & Lazer, D. (2013). Voices of victory: A computational focus group framework for tracking opinion shift in real time. In D. Schwabe, V. Almeida, H. Glaser, R. Baeza-Yates, & S. Moon (Eds.), *WWW 2013: Proceedings of the 22nd international world wide web conference* (pp. 737–748). Geneva, Switzerland: International World Wide Web Conferences Steering Committee.
- MacKuen, M. B., Erikson, R. S., & Stimson, J. A. (1989). Macropartisanship. *American Political Science Review, 83*, 1125–1142.
- Marchetti-Bowick, M., & Chambers, N. (2012). Learning for microblogs with distant supervision: Political forecasting with Twitter. In W. Daelemans (Ed.), *EACL '12: Proceedings of the 13th conference of the European chapter of the association for computational linguistics* (pp. 603–612). Stroudsburg, PA: Association for Computational Linguistics.
- Margolis, M., & Resnick, D. (2000). *Politics as usual: The Cyberspace "revolution."* Thousand Oaks, CA: Sage.
- McKelvey, K., DiGrazia, J., & Rojas, F. (2014). Twitter publics: How online political communities signal electoral outcomes in the 2010 U.S. House election. *Information, Communication & Society, 17*, 436–450.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*, 5–11.
- Metaxas, P. T., Mustafaraj, E., & Gayo-Avello, D. (2011). How (not) to predict elections. In A. Vinciarelli, M. Pantic, E. Bertino, & J. Zhan (Eds.), *SocialCom 2011: The 3rd IEEE international conference on social computing* (pp. 165–171). Washington, DC: IEEE.
- Murphy, J., Link, M. W., Childs, J. H., Tesfaye, C. L., Dean, E., Stern, M., . . . Harwood, P. (2014). Social media in public opinion research: Executive summary of the aapor task force on emerging technologies in public opinion research. *Public Opinion Quarterly, 78*, 788–794.
- O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In M. Hearst, W. Cohen, & S. Gosling (Eds.), *ICWSM 2010: Proceedings of the 4th international AAAI conference on weblogs and social media*. Menlo Park, CA: Association for the Advancement of Artificial Intelligence (AAAI).
- Rainie, L., Smith, A., Schlozman, K. L., Brady, H., & Verba, S. (2012). Social media and political engagement. Pew Internet & American Life Project. Retrieved from <http://pewinternet.org/Reports/2012/Political-engagement.aspx>
- R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science, 346*, 1063–1064.
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. In M. Rappa, P. Jones, J. Freire, & S. Chakrabarti (Eds.), *WWW 2010: Proceedings of the 19th international conference on the World Wide Web* (pp. 851–860). New York, NY: ACM.
- Sartori, G. (1970). Concept misformation in comparative politics. *The American Political Science Review, 64*, 1033–1053.
- Schoen, H., Gayo-Avello, D., Metaxas, P. T., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. *Internet Research, 23*, 528–543.

- Skoric, M., Poor, N., Achananuparp, P., Lim, E. P., & Jiang, J. (2012). Tweets and votes: A study of the 2011 Singapore general election. In R. H. Sprague Jr (Ed.), *HICSS 2012: Proceedings of the 45th Hawaii international conference on system science* (pp. 2583–2591). Washington, DC: IEEE Computer Society.
- Soler, J. M., Cuartero, F., & Roblizo, M. (2012). Twitter as a tool for predicting elections results. In F. Karray & F. Polat (Eds.), *ASONAM 2012: IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 1194–1200). Los Alamitos, CA: IEEE Press.
- Thapen, N. A., & Ghanem, M. M. (2013). Towards passive political opinion polling using Twitter. In M. Cocea, M. Gabe, N. Wiratunga, & A. Göker (Eds.), *Proceedings of the BCS SGAI workshop on social media analysis 2013* (pp. 19–34). Aberdeen, UK: School of Computing Science and Digital Media, Robert Gordon University.
- Tjong, E., Sang, K., & Bos, J. (2012). Predicting the 2011 Dutch Senate election results with Twitter. In *Proceedings of the workshop on semantic analysis in social media* (pp. 53–60). Stroudsburg, PA: Association for Computational Linguistics.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, England: Cambridge University Press.
- Tsakalidis, A., Papadopoulos, S., Cristea, A. I., & Kompatsiaris, Y. (2015). Predicting elections for multiple countries using Twitter and polls. *IEEE Intelligent Systems*, 30, 10–17.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In M. Hearst, W. Cohen, & S. Gosling (Eds.), *ICWSM 2010: Proceedings of the 4th international AAAI conference on weblogs and social media* (pp. 178–185). Menlo Park, CA: Association for the Advancement of Artificial Intelligence (AAAI).
- Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., & Tucker, J. (2013). Social media and political communication: A survey of Twitter users during the 2013 Italian general election.” *Rivista Italiana di Scienza*, 43, 325–55.
- Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2014). Forecasting elections with non-representative polls. *International Journal of Forecasting*. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0169207014000879>
- Whiteley, P., Clarke, H. D., Sanders, D., & Stewart, M. C. (2013). *Affluence, austerity and electoral change in Britain*. Cambridge, England: Cambridge University Press.
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer.
- All links were accessible on October 18, 2015.

Author Biographies

Andreas Jungherr is a research fellow at the chair for political psychology at the University of Mannheim, Germany. andreas.jungherr@gmail.com

Harald Schoen holds the chair for political psychology at the University of Mannheim, Germany. harald.schoen@uni-mannheim.de

Oliver Posegga is a research fellow at the chair for social network analysis at the University of Bamberg, Germany. oliver.posegga@uni-bamberg.de

Pascal Jürgens is a research associate at the Department of Mass Communication at the University of Mainz, Germany. pjuergen@uni-mainz.de