Soc. Netw. Anal. Min. (2013) 3:1277–1291
DOI 10.1007/s13278-012-0079-3

ORIGINAL ARTICLE

Social media and political communication: a social media analytics framework

Stefan Stieglitz · Linh Dang-Xuan

Received: 2 February 2012 / Revised: 24 May 2012 / Accepted: 13 July 2012 / Published online: 25 August 2012
© Springer-Verlag 2012

Abstract In recent years, social media are said to have an impact on the public discourse and communication in the society. In particular, social media are increasingly used in political context. More recently, microblogging services (e.g., Twitter) and social network sites (e.g., Facebook) are believed to have the potential for increasing political participation. While Twitter is an ideal platform for users to spread not only information in general but also political opinions publicly through their networks, political institutions (e.g., politicians, political parties, political foundations, etc.) have also begun to use Facebook pages or groups for the purpose of entering into direct dialogs with citizens and encouraging more political discussions. Previous studies have shown that from the perspective of political institutions, there is an emerging need to continuously collect, monitor, analyze, summarize, and visualize politically relevant information from social media. These activities, which are subsumed under "social media analytics," are considered difficult tasks due to a large numbers of different social media platforms as well as the large amount and complexity of information and data. Systematic tracking and analysis approaches along with appropriate scientific methods and techniques in political domain are still lacking. In this paper, we propose a methodological framework for social media analytics in political context. More specifically, our framework summarizes most important politically relevant issues from the perspective of political institutions and corresponding methodologies from different scientific disciplines.

Keywords Social media · Social media analytics · Framework · Political communication

S. Stieglitz (✉) · L. Dang-Xuan
Department of Information Systems, University of Muenster, Muenster, Germany
e-mail: stefan.stieglitz@uni-muenster.de

L. Dang-Xuan
e-mail: linh.dang-xuan@uni-muenster.de

1 Introduction

In the past few years, social media have shown a rapid growth of user counts and have been object of scientific analysis (Wigand et al. 2010; McAfee 2006). For example, more than 800 million people worldwide are members of the Facebook network (Facebook 2011) while Twitter counts more than 200 million accounts in total (HuffPost Tech 2011). This mainstream adoption of social media applications has changed the physics of information diffusion. Until a few years ago, the major barrier for someone who wanted a piece of information to spread through a community was the cost of the technical infrastructure required to reach a large number of people. Today, with widespread access to the Internet, this bottleneck has largely been removed. In this context, personal publishing modalities such as social network sites (SNS), microblogging, and weblogs have become prevalent (Kaplan and Haenlein 2010). The process by which people locate, organize, and coordinate groups of individuals with similar interests, the number and nature of information and news sources available, and the ability to solicit and share opinions and ideas across various topics have all undergone dramatic change with the rise of social media (Agrawal et al. 2011).

As a result, the growing relevance of communication in social media implies a fundamental change in traditional public communication, which has usually been exclusively initiated and managed by specific actors, e.g., politicians,

companies as well as journalists (Chadwick 2006). This phenomenon is currently observed by numerous disciplines such as sociology, information communication studies, information systems, political science, and linguistics. Among other fields of interest, it is a common goal to better understand modes of communication such as agenda setting or opinion making in social media.

Given the tremendous growth of social media, in particular Twitter and Facebook, social media are increasingly used in political context recently—both by citizens and political institutions (e.g., politicians, political parties, political foundations, think tanks etc.). From the perspective of political institutions, it is important to actively participate in the political communication based on the use of social media, especially during election campaigns. Social media thereby represents the ideal vehicle and information base to gauge public opinion on policies and political positions as well as to build community support for candidates running for public offices (Zeng et al. 2010). It has been observed that in a very short space of time, politicians in modern democracies across the world have eagerly adopted social media for engaging their constituents, entering into direct dialogs with citizens and enabling vivid political discussions (Hong and Nadler 2011). In this regard, US politicians are said to have a leading role with the most prominent example of Barack Obama being able to successfully employ social media within his last election campaign (Wattal et al. 2010). On the other hand, social media are said to have the potential for increasing political participation and discussions among citizens. Twitter, Facebook, and others provide ideal platforms for users to spread not only information in general but also political opinions through their networks.

Studies have shown that from the perspective of political institutions and government agencies, there is a need to continuously gather, monitor, analyze, summarize, and visualize politically relevant information from online social media with the goal to improve communication with citizens and voters (e.g., Zeng et al. 2010; Kavanaugh et al. 2011; Paris and Wan 2011; Stieglitz et al. 2012). For example, it is important to identify influential users or (political) opinion leaders, and follow the discussions taking place within their peers, particularly during periods of election campaigns. Other relevant issues might concern the identification of emergent issues and trends as well as the ability to make prediction of potential rising topics. The final goal is then to get a compact and comprehensive summary (e.g., in regular reports or real-time dashboards) which aggregates and visualizes analysis results from different social media platforms.

However, this is considered a challenging task due to a large numbers of different social media platforms and vast amount as well as complexity of information and unstructured data. One reason for this is that information of this kind is not considered by means of classic information retrieval as done by common search engines. Identifying distinct subjects, gathering and analyzing information and aggregate them is therefore still a challenge, which, however, is being tackled by "social media analytics" (Zeng et al. 2010; Agrawal et al. 2011; Leskovec 2011; Nagarajan et al. 2011). According to Zeng et al. (2010), social media analytics is supposed to provide tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data in an automated way due to the massive amount of (mostly unstructured) social media data.

Social media analytics has become increasingly relevant not only for political institutions and government sector (e.g., Kavanaugh et al. 2011; Paris and Wan 2011; Stieglitz et al. 2012) but also in the context of business and marketing (e.g., Gruhl et al. 2010; Larson and Watson 2011). For profit businesses are tapping into social media as both a rich source of information and a business execution platform for product design and innovation, consumer and stakeholder relationship management, and marketing. For them, social media represent an essential component of the next-generation business intelligence platform (Gruhl et al. 2010).

In business and marketing context, there exist already a variety of open-source as well as proprietary social media analytics tools, which provide from simple standard analyses to customer-tailored solutions. However, this is not the case for the political domain. In particular, there is a lack of frameworks describing systematic approaches and appropriate methods and techniques required for tracking, monitoring and analyzing content from social media in political context. The goal of this paper is to propose such a framework, which outlines most important politically relevant analyses from the perspective of political institutions and various methodologies from multiple scientific disciplines for conducting those analyses.

The remainder of this paper is structured as follows. In the next section, we review related work regarding social media and political communication. The subsequent section addresses social media analytics and its relevance for political institutions. We then present and discuss our framework for social media analytics in political context. Subsequently, we provide some illustrative examples of data tracking and analysis described by our framework. Finally, the paper ends with concluding remarks and an outlook for further research.

2 Social media and political communication

Social media are defined as "a group of internet-based applications that build on the ideological and technological foundations of Web 2.0 that allow the creation and

exchange of user-generated content" (Kaplan and Haenlein 2010). Recently, social media such as SNS, weblogs, microblogging, and wikis play an increasingly important role in shaping political communication in the US and around the world (e.g., Aday et al. 2010; Benkler 2006; Bennett 2003; Farrell and Drezner 2008; Sunstein 2002; Tumasjan et al. 2011). The potentials of social media appear to be most promising in political context as they can be an enabler for more participation and democracy. Creighton (2005) defines public participation as the process by which public concerns, needs and values are incorporated into governmental and corporate decision making. The so-called e-participation focuses not only on this process but also on using the Internet as an additional or exclusive instrument to create dialogs between the elected and the electorate. Related to that, Karpf (2009) introduces the notion of "politics 2.0," which can be understood as the harnessing of the Internet's lowered transaction costs and its condition of information abundance, toward the goal of building more participatory, interactive political institutions.

There is a growing body of research focusing on the role of social media in political deliberation. The recent US presidential campaign in 2008 has shown that social media technologies have become increasingly important for political communication and persuasion (Wattal et al. 2010). It became obvious, that social media could be successfully adapted to contact and discuss with voters as well as to disseminate important information to them. Especially young people were inspired to political topics using social media as communication platform (Chen et al. 2009; Kushin and Kitchener 2009). In this section, we give an overview of related work on social media in political context. In particular, we focus on previous studies dealing with the three major types of social media: microblogging, SNS, and weblogs.

2.1 Microblogging

A number of other studies focus on political microblogging (particularly Twitter) use, with studies focusing on either non-parliamentary or parliamentary uses of the service. As for parliamentary uses, previous literature has dealt with the USA. For example, Golbeck et al. (2010) focus on the US Congress and analyze the contents of more than 6,000 tweets from Congress members. They find that Congress members consider Twitter rather a vehicle for self-promotion as they are primarily using Twitter to share information, particularly links to news articles about themselves and to their blog posts, and to report on their daily activities. Similarly, another study by Ammann (2010) focuses on the use of Twitter by US Senate candidates and the content of their tweets during the 2010 midterm election season. Results show that candidates use Twitter as a part of their political campaigns. However, the amount of use significantly varies by the level of resources a candidate possesses, state size, and the competitiveness of the congressional race. Also, the content of the tweets is largely related to candidate type and in some cases political affiliation of the candidate. In a study of Twitter adoption and use by US Congress members, Lassen and Brown (2011) find that members are more likely to use Twitter if they belong to the minority party, if their party leaders urge them to, if they are young, or if they serve in the Senate. Hong and Nadler (2011) estimate the impact of the use of Twitter by American politicians on changes in public opinion of those politicians over time and find little evidence that the political use of Twitter has either a positive or negative impact on public opinion.

Other works focus on the use of Twitter by citizens in political context. Recently, the notion of "Twitter revolutions" in totalitarian countries has been introduced, although the exact contents and effects of these uprisings are disputed. For example, Gaffney (2010) studies Twitter use during the 2009 Iran elections by tracking the use of the #IranElection hashtag. Although Twitter helped protesters in Iran and around the world in organizing their efforts, the author claims that "it is difficult to say with any certainty what the role of Twitter was." In a study of approximately 100,000 messages containing a reference to either a political party or a politician in the context of the 2009 German federal election, Tumasjan et al. (2011) show that Twitter is extensively used for the dissemination of politically relevant information and that the mere number of party mentions accurately reflects the election result suggesting that microblogging messages on Twitter seem to validly mirror the political landscape offline and can be used to predict election results to a certain extent. Similarly, Larsson and Moe (2011) study Twitter use during the 2010 Swedish general election and find that Twitter serves as a channel for disseminating political contents and not for political dialog. Furthermore, Twitter also serves as just a new outlet for speakers already belonging to elite, or at least affiliated with prominent positions in mainstream media or political debate in general. Conover et al. (2011) examine two networks of political communication on Twitter with more than 250,000 tweets from the 6 weeks leading up to the 2010 US congressional midterm elections. Using a combination of network clustering algorithms and manually annotated data, the authors demonstrate that the network of political retweets exhibits a highly segregated partisan structure, with extremely limited connectivity between left- and right-leaning users. Surprisingly, this is not the case for the user-to-user mention network, which is dominated by a single politically heterogeneous cluster of users in which ideologically opposed individuals interact at a much higher rate compared to the network of retweets.

Similarly, Yardi and Boyd (2010) find that, in political context, Twitter users are more likely to interact with others who share the same views as they do in terms of retweeting, but they are also actively engaged with those with whom they disagree. In addition, replies between like-minded individuals would strengthen group identity, whereas replies between different-minded individuals would reinforce in-group and out-group affiliation.

2.2 Social network sites

Analyzing publicly available content on SNS such as Facebook has become an increasingly popular method for studying socio-political issues. Such public-contributed content, primarily available as Wall posts and corresponding comments on Facebook pages or Facebook groups, let people express their opinions and sentiments on a given topic, news or persons, while allowing social and political scientists to conduct analyses of political discourse.

Recently, previous studies have specifically focused on SNS and their use by politicians. Williams and Gulati (2007, 2009) investigate the extent of Facebook use by Congressional candidates during election campaigns. They find that the number of Facebook supporters can be considered a valid indicator of electoral success. In the context of the 2006 Dutch elections, Utz (2009) shows that SNS provide an opportunity for politicians to reach individuals less interested in politics. Thereby, viewing a candidate's profile further strengthens existing attitudes. On the other hand, politicians who react on the comments of users are perceived more favorable.

On the other hand, Kushin and Kitchener (2009) explore the use of Facebook for online political discussion by citizens. Their results indicate that Facebook is a legitimate location for discussion of political issues and, to some extent, the discussion appears to have succeeded in overcoming polarization of online discussion that has pervaded online political discussion in the past. Baumgartner and Morris (2010) examine the political uses of SNS by young adults in context of the early stages of the 2008 presidential primary season. Using a survey of over 3,500 18- to 24-year olds contacted immediately prior to the Iowa caucuses, they show that although SNS are recognized by youth as a possible source of news and that many receive some of their news from these sites, the types of news gathered probably do little to inform them or add to democratic discourse. Moreover, in spite of the promise SNS hold for increasing political interest and participation among a chronically disengaged cohort, users are no more inclined to participate in politics than are users in other media. In line with these findings, Vitak et al. (2011) also show that undergraduate students tend to engage in rather lightweight political participation both on Facebook and in other venues. The level of engagement indicated suggests that while young voters may be participating in political activity, the degree of this participation is somewhat superficial. The most common forms of general political participation tended to be informational and low in resource intensity, whereas political actions that required a greater commitment of resources (e.g., volunteering) were less frequent. This finding in isolation lends credibility to the concern that young citizens are becoming "slacktivists," engaging in feel-good forms of political participation that have little or no impact on effecting change.

By examining 2 years of posts on the Facebook walls of the three major contenders for the US Presidency in 2008: Barack Obama, Hillary Clinton, and John McCain, Robertson et al. (2010) analyze participation patterns of usage along dimensions of breadth and frequency, and interpret them in terms of the concept of the "public sphere." They conclude that SNS are currently the socio-technical environments that most closely enable public sphere discourse for those who choose to enter the online "salons" of political candidates. However, it remains an open question whether the outcome of this dialog is informed political decision making in the form of votes. Similarly, Zhang et al. (2010) show that reliance on SNS such as Facebook and MySpace was positively related to civic participation but not to political participation or confidence in government. Rather, interpersonal discussion would foster both civic participation and political activity.

2.3 Weblogs

There are different strands of literature focusing on blogosphere in political context. The first one deals with the uses and benefits of political blogs. In a survey of 3,747 blog readers, who were generally young in age and highly educated, Kaye (2005) finds the trend that blogs were basically a medium for political information seeking and participation. McKenna and Pole (2007) report that the contents of political blogs are composed of information about news articles from the mass media, introductions to other blogs' postings, and criticism about mass media coverage on political affairs. In another paper, McKenna (2007) also reports that political bloggers construct their blogs from the motivation of voluntarism and not for commercial purposes. Hacker et al. (2006) explore the potential of political blogs to provide discourse which contributes to uses of computer-mediated communication for depolarizing political discourse. They argue that political blogs might provide spaces for securitization discourse that might not be found in traditional mass media and older genres like websites. In the light of the 2008 US presidential election, Wattal et al. (2010) investigate the contingent impact of political blogs on the campaign process.

Their results show that in particular the blogosphere can influence the campaign process and the election outcomes. In another study, Munson and Resnick (2011) provide evidence of the prevalence of political discourse even in non-political blogs. Using a random sample from Blogger.com, they find that 25 % of all political posts are from blogs that post about politics less than 20 % of the time. In particular, non-political spaces devoted primarily to personal diaries, hobbies, and other topics represent a substantial place of online political discussion.

Another strand of blog research concerns structural or network analysis of political blogosphere, particularly in terms of social network analysis (Rosen et al. 2011). Farrell and Drezner (2008) argue that the blogosphere is of significant importance for political communication. By conducting case studies, they show that political blogs which attract much attention—so-called "focal points"—have an influence on the agenda setting as they attract a high level of attention from journalists who act as multipliers. In an empirical work, Adamic and Glance (2005) study the linking patterns and discussion topics of US political bloggers and provide evidence for a divided blogosphere: liberals and conservatives primarily link within their separate communities, with far fewer cross-links exchanged between them. Regarding political discussions, liberal and conservative blogs focus on different news, topics and political figures. Also, conservative bloggers tend to link to other blogs (mostly other conservative blogs) than their liberal counterparts. As Adamic and Glance (2005) indicate, the benefit of hyperlink analysis is the finding of the relational patterns in cyberspace. Other studies find that political blogs run by individuals are hyperlinked with organizational blogs as well as other individuals. Barabasi (2002) shows that the network of these relationships tends to show an unequal distribution where a small number of blogs occupy a majority of hyperlinks. This concentration also means that a majority of blogs do not link or are not linked by other blogs (Adar and Adamic 2005; Adar et al. 2004; Wu and Huberman 2004). In addition, political blogs tend to be clustered along with their interest or affiliation. For instance, social activist blogs tend to link other activist blogs with similar movement themes. Further, they link to the political parties they support or criticize. However, clusters found from blogosphere centered around a small number of key nodes (Herring et al. 2005; Schmidt 2007). Schmidt (2007) reports that blog communities are established with informal rules such as "blog etiquette," including "crediting the source of a link." He adds that the bloggers who share implicit social norms and meet the group's expectations are considered members regardless of the existence of the official member registration process for the group.

3 Relevance of social media analytics

Social media provide a connection between social networks, personal information channels, and the mass media. Social media data in the form of user-generated content on blogs, microblogs, SNS, discussion forums, and multimedia sharing websites (e.g., YouTube) present many new opportunities and challenges to both producers and consumers of information. Although there is a vast quantity of data available, the consequent challenge is to be able to analyze the large volumes of user-generated content and often implicit links between users in order to gain meaningful insights into the diffusion of information, opinions and sentiments as well as emergent issues and trends (Leskovec 2011; Agrawal et al. 2011; Nagarajan et al. 2011). This is referred to as "social media analytics." According to Zeng et al. (2010), "social media analytics is concerned with developing and evaluating informatics tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data, usually driven by specific requirements from a target application." However, as Zeng et al. (2010) also point out, social media analytics faces several challenges such as an enriched set of data or metadata (e.g., tags, user-expressed subjective opinions, ratings, user profile etc.), human-centered computing with their own unique emphasis on social interactions among users, semantic inconsistency/inaccuracies, misinformation and lack of structure as well as dynamic nature of social media data and their sheer size. Nevertheless, recent advances in different scientific disciplines, particularly computer science, statistics, computational linguistics, etc., provide a variety of analysis techniques and methods to tackle those problems (Nagarajan et al. 2011). For example, in the business-to-customer context, Larson and Watson (2011) propose a "social media ecosystem" framework, explicating the social-media-enabled relationships among stakeholder group. The framework deconstructs the phenomenon of social media into multiple layers of firm-initiated and customer-initiated actions and provides a theoretical understanding of what firms and customers accomplish using social media. It sets the stage for developing measures of those firm/customer social media activities with a critical bearing on firm performance (Larson and Watson 2011).

Recent studies have shown an emerging need of political institutions as well as government services for leveraging social media resources to improve services and communication with citizens and voters (e.g., Kavanaugh et al. 2011; Paris and Wan 2011). It becomes increasingly important to stay updated about current discussions and to manage one's own reputation in virtual communities, particularly regarding emerging topics that can end up in a scandal or crisis for a specific politician or party (e.g., Zeng et al. 2010; Stieglitz et al. 2012).

In an exploratory study of social media use by government officials in Virginia, USA, Kavanaugh et al. (2011) find that social media aggregation tools are needed to make sense of the overwhelming amount of data that is being generated, to model the flow of information, and to identify patterns over time. There might be already some tools for these purposes, however, these tools are designed to support businesses and not government, so they are not optimal for civic needs. Moreover, government officials would prefer digital libraries to archive and curate user-generated content, especially for crisis and social convergence situations, but also for analyses that cover longer time frames.

By conducting interviews with responsible people of an Australian government agency, Paris and Wan (2011) identify three main reasons for social media monitoring activities for government services: measuring campaign effectiveness, measuring the impact of and reaction to content produced by the institution, and offering improved services by interacting with the online community regarding specific social media postings.

In a survey of members of the German parliament ("Bundestag"), Stieglitz et al. (2012) find that the majority of parliament members wish to have concepts and instruments to identify trending political topics in an early stage. Furthermore, many politicians using social media also attempt to look for feedback, suggestions and new ideas from their group members, followers and others for their political work. However, this is a task that requires much effort, appropriate tools and particularly a systematic approach, which most politicians cannot afford due to limited resources and capacities.

In sum, Fig. 1 illustrates the role of social media analytics in social media-based political communication between political actors/institutions and citizens/voters as well as among citizens/voters themselves.

4 A social media analytics framework in political context

As discussed above, there is a lack of frameworks outlining politically relevant analyses from the perspective of political institutions and appropriate methods and techniques for conducting those analyses. We seek to fill this research gap by proposing such a framework for social media analytics in the political domain. From the practical perspective, the framework should serve as a kind of guideline for the development of toolsets aiming at collecting, storing, monitoring, analyzing, and summarizing politically relevant user-generated content from social media for political institutions. From the research perspective, the framework gives a comprehensive overview of different specific methodological approaches from different disciplines, which can be employed for social media analytics. As Zeng et al. (2010) already point out, social media analytics has a research agenda which is multidisciplinary in nature and has drawn attention from research communities in major disciplines. Consequently, in our framework we consider methodological approaches from various disciplines such as computer science, statistics, computational linguistics as well as communication studies and sociology.

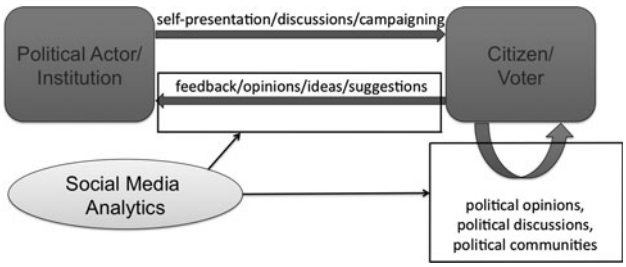

Fig. 1 Political communication and social media analytics

The framework presented in this paper consists of two major parts: data tracking and monitoring, and data analysis. While data tracking and monitoring is concerned with different approaches for how and what kind of politically relevant user-generated data from different social media platforms can be tracked and monitored, the data analysis part deals with various analysis methods for different analysis purposes and approaches. Furthermore, the framework considers three major types of social media: (1) microblogging, (2) SNS, and (3) weblogs. Although there exist currently many platforms which can be categorized as microblogging or SNS, they are, however, often characterized by significantly different functionalities, target groups, or purposes. In this paper, we focus on Twitter and Facebook as the most prominent public platforms of their kind.

4.1 Data tracking and monitoring

4.1.1 Tracking sources

The very first step regarding data collection is to determine the sources of the data. With skyrocketing growth in the user base in recent years, social network sites and microblogging services with their most prominent examples Facebook and Twitter offer a huge amount of (politically relevant) user-generated content in an unprecedented scope. As a longer established social media type, weblogs are not to be neglected either. In particular, blogs are considered a very important communication platform for political discussions. Therefore, we decide to build up our framework with focus on Twitter (microblogging), Facebook (SNS), and weblogs as main sources of social media data.

Regarding Twitter, the data to be tracked and monitored are in the form of public "tweets" to which access can be easily obtained. For Facebook, the most important kind of data represents the content of the "Wall" including "status updates" and corresponding comments. It is important to note that data on Facebook are, contrary to Twitter, of semi-public nature, i.e., one can collect data only from publicly made profiles (or more practically, groups or "pages," since most individual profiles are not public). More specifically, in political context, one should predefine a list of politically relevant Facebook groups and pages which serves as sources of data. Similarly, one should also identify influential political blogs in advance (i.e., before starting to track blog entries and corresponding comments). Various rankings of political blogs would facilitate this task (e.g., Technorati, BlogPulse or PostRank). It is important to note that Facebook and blog messages, in contrast to tweets, are not limited regarding their length. Therefore, Facebook and blog postings might be more eligible for in-depth discussions, while Twitter rather provides an ad-hoc short message-based platform for information dissemination.

4.1.2 Tracking methods

Both Twitter and Facebook offer application programing interfaces (API) for data tracking. The most frequently used API for Twitter are "Search API" and "Streaming API", while Facebook's "Graph API" allows programmers to conveniently track Wall postings (i.e., status updates and comments). In contrast, weblogs do not provide such standardized way to get access to the data. However, most blogs do offer RSS feeds which can be easily tracked. For those without RSS function, manual web-crawling techniques such as HTML parsing are necessary.

4.1.3 Tracking outputs

4.1.3.1 Data types Within each data source, there will typically be elements of both structured and unstructured data. While structured data (or more precisely metadata) comprise profile/user demographics, spatial, temporal, and thematic data as well as attention-related data (e.g., number of 'likes,' comments, retweets, mentions etc.), unstructured data include user-generated textual content ranging from relatively context-sparse microblogs, Facebook comments, to context-rich blogs. Those data can be transformed into common formats and ingested into databases. In general, typical data to be tracked and stored may include, for instance, ID of the posting, time stamp, username (of the author), content of the posting, and possibly the type of the posting (i.e., status update, blog entry, or retweet, comment).

4.1.3.2 Tracking approaches There are different approaches of data tracking for a political institution which depend on the specific intentions of that institution: (1) self-involved, (2) keyword/topic-based, (3) actor-based, (4) random/exploratory, and (5) URL-based.

- Self-involved approach: The first approach is applicable when, for example, individual politicians or political parties want to know explicitly how people are talking about them in social media. In such case, the politicians or parties can have all tweets collected that contain their name either as simple keyword or hashtag. If they have an own Facebook presence in terms of a page or group, they should track all posts and corresponding comments published by users or fans/members of their own page or group. Likewise, if they also have an own blog, all comments to their blog entries should be gathered for analysis. Furthermore, it might be useful to collect all Facebook and blog postings that contain their name from external predefined Facebook groups/pages and blogs, respectively.
- Keyword/topic-based approach: Political actors are usually highly interested in the feedback or opinions of social media users to certain political topics. Here, the second tracking approach seems to be eligible where tweets as well as Facebook and blog postings that involve keywords related to topics of interest can be tracked. To attain a high level of data completeness, relevant keywords representing the topic of interest have to be carefully and systematically chosen in advance. The broader the topic to be analyzed, the more keywords should be taken into account.
- Actor-based approach: In political communication, particularly in the blogsphere and recently on SNS and microblogging platform, there are usually a number of actors who can be considered as more influential or more popular than most other users. These actors are said to have the power to influence (online) opinion-making processes. Therefore, politicians or political parties are also interested in monitoring such important users in terms of their generated content. For that, an actor-based tracking approach might be employed to track tweets, Wall postings, and blog entries as well as corresponding comments specifically contributed by those influential users who should also be identified in advance.
- Random/exploratory approach: Contrary to the first three tracking approaches which are rather of targeted nature, the fourth approach supports exploratory inductive content mining. The idea behind this tracking approach is to randomly select one or several sets of data (tweets, Facebook or blog postings) for different time periods for analysis. Based on these random

datasets, particularly content analysis might be conducted to identify major political topics and detect users' opinions or sentiment associated with those topics.

- URL-based approach: Given that social media platforms are widely used, among other purposes, to disseminate information, particularly by means of posting URL, political actors might also apply a URL-based approach to selectively track contents behind hyperlinks shared in tweets, Facebook and blog postings. This might provide additional meaningful insights, especially in case of tweets with a limited length of 140 characters.

4.1.3.3 Data Preprocessing To prepare textual data for further analysis, a preprocessing step needs to be carried out. For example, stop words (i.e., words that only have grammatical significance) should be eliminated. In addition, stemming (i.e., reducing inflected or sometimes derived words to their stem, base or root form) as well as lemmatization (i.e., grouping together the different inflected forms of a word, so they can be analyzed as a single item) might also be performed to facilitate further analyses.

A comprehensive overview of the described data tracking and monitoring process is provided in Fig. 2.

4.2 Data analysis

4.2.1 Analysis purposes

From the perspective of political institutions, findings by previous studies suggest that there are two major purposes of conducting social media analytics. The first concerns the *management of own reputation and impression* in the sphere of social media while the second relates to a more *general monitoring* of user-generated content from social media (Zeng et al. 2010; Stieglitz et al. 2012).

4.2.1.1 Reputation and impression management Previous studies have shown that many politicians and parties

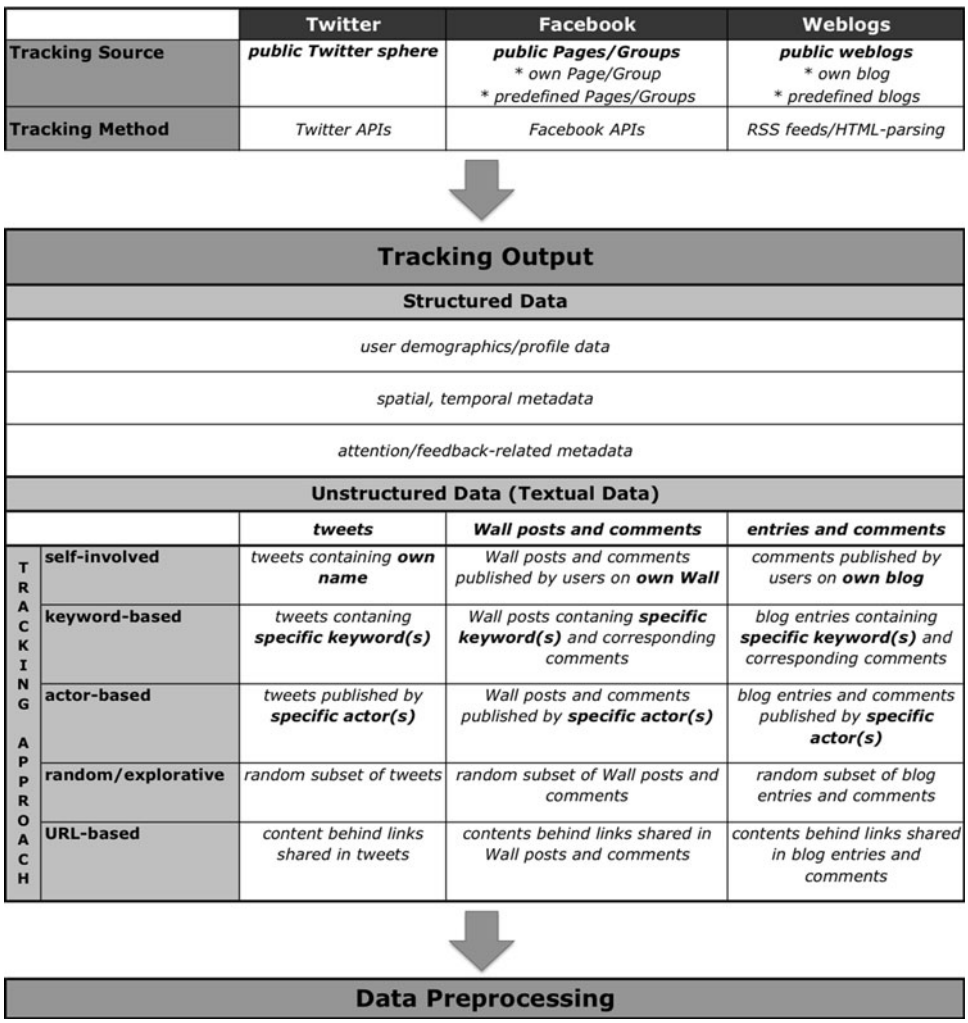

Fig. 2 Data tracking and monitoring

may have the need to be updated about their own reputation and impression in public social networks. In particular, three following aspects might be relevant. First, political institutions are interested in investigating how people (i.e., voters) are talking about them, particularly regarding specific characteristics such as trustworthiness, determination, success etc. Second, they are also interested in emerging topics that have the potential to trigger a crisis or scandal that might harm their reputation. Finally, political institutions seek to measure the degree of influence they have in social networks (e.g., the ability to spread messages and political opinions widely and quickly).

4.2.1.2 General monitoring Besides management of reputation and impression, political institutions might employ social media analytics to monitor the social web in an explorative way. For example, they might be interested in knowing about what kind of political topics or issues are discussed and how such discussions take place in social media. In addition, early detection of upcoming "hot" topics or issues might enable political institutions to react timely to such trends (e.g., by elaborating appropriate political strategies). More generally, trends can also be viewed as a reflection of societal concerns or even as a consensus of collective decision making. Understanding how a community decides that a topic is "trendy" can help better understand how ad-hoc communities are formed and how decisions are made in such communities (Agrawal et al. 2011). Compared to reputation and impression management-based monitoring, general monitoring is more challenging because a main task is to predefine appropriate data sources for tracking and analyses (e.g., Twitter messages posted by specific users or containing specific keywords, specific Facebook pages or groups, specific weblogs etc.). Furthermore, it is needed to continuously update or extend the list of data sources.

4.2.2 Analysis approaches and corresponding analysis methods

Besides analysis purpose, another dimension of the framework represents the different analysis approaches. In our framework, we consider three major approaches: (1) topic/issue-related, (2) opinion/sentiment-related, and (3) structural.

4.2.2.1 Topic/issue/trend-based approach For politicians and parties, it is important to identify and monitor political topics or issues, particularly those that might have a direct or indirect association with themselves as issues contain conflict potential and may evolve to a crisis (Ingenhoff and Röttger 2008; Wartick and Mahon 1994). In communication studies, such topic scanning and monitoring activities are referred to as issue management, which can serve the prevention of potential crisis or scandal that might lead to damages of own reputation.

In order to identify topics, content analysis, or more specifically, text mining techniques, might be applied. Content analysis is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use (Krippendorff 2004). Given the massive amounts of social media data, automated quantitative methods of content analysis or text mining are necessary. These methods are suitable for providing answers to a broad variety of questions, among which are the classification of texts and the identification and modeling of recurring topics (Krippendorff 2004). There are different approaches of text classification. Besides descriptive and explorative approaches, such as word frequency analysis or co-occurrence analysis (Doerfel and Barnett 1996), as well as deductive approaches including dictionary-based coding or rule-based parsing (Stone et al. 1966), there are a number of new approaches to automated text analysis that have recently found their way into the social sciences (Hillard et al. 2007; Monroe and Schrodt 2008), most notably text classification based on unsupervised and supervised learning (Sebastiani 2002; Liu 2011) as well as semantic network analysis (van Atteveldt 2008). In particular, supervised text classification, which uses statistical algorithms from machine learning (e.g., support vector machine (SVM) or naïve Bayesian classifier), has the potential to become a standard method for automated text mining (Scharkow 2011). In addition, documents might be clustered based on unsupervised learning using techniques such as hierarchical and k-means clustering. Regarding topic modeling, recent advances in natural processing language provide more sophisticated statistical models for discovering abstract topics that occur in documents. These include, for example, the probabilistic latent semantic indexing (Hofmann 1999) or latent Dirichlet allocation models (Blei et al. 2003) along with algorithms such as singular value decomposition or non-negative matrix factorization (Blei 2011). Despite many advantages of automated approaches, manual text analysis is nevertheless needed to back up findings by automated analysis and to get a more fine-grained picture, as it defines a set of practices that enable human coders to define reproducible categories for qualitative features of text more reliably (Krippendorff 2004). As previous studies have indicated, many politicians and parties have also expressed their wish to have trending political topics predicted. Also, trends in social networks have recently been a major focus of interest among both research and industry (Agrawal et al. 2011). Recent advances in computer science and statistics have proposed a variety of algorithms to predict emerging topics. For example, many trend-detecting

algorithms are based on so-called hidden Markov models where observations of topics are trained by such models which in turn are saved in a library for the topic's prediction. Topics with similar life cycle are recorded and share the same model (e.g., Zeng et al. 2007; Liu and Guo 2011). In a recent study, Budak et al. (2011) introduce new methods for identification of important topics in social networks that utilize the network topology. They propose two novel trend definitions called coordinated and uncoordinated trends that detect topics that are popular among highly clustered and distributed users, respectively. A novel information diffusion model called "independent trend formation model" (ITFM) has also been introduced to distinguish viral diffusion of information from diffusion through external entities, such as news media, and to capture the diffusion of an arbitrary number of topics in a social network.

Following another approach, Kasiviswanathan et al. (2011) propose a dictionary learning-based framework for detecting emerging topics in social media and related streams. The overall framework consists of two stages—first, determining novel documents in the stream, and subsequently identifying cluster structure among the novel documents. Finally, Mathioudakis et al. (2010) consider the problem of early online identification of items that gather a lot of attention in social media. They model social media activity using a stochastic model for "interacting streaming information sources" (ISIS) that intuitively captures the concept of attention gathering information items. Given the challenge of the information overload characterizing digital social activity, they present sequential statistical tests that enable early identification of attention gathering items. This effectively reduces the set of items to be monitored in real time in order to identify pieces of information attracting a lot of attention.

4.2.2.2 Opinion/sentiment-based approach Given the rapid growth of social media, people are enabled to express their views, opinions or emotions on almost anything in forums, blogs, and on SNS more than ever before. This applies particularly to political communication, which is assumed to be of polarizing controversial nature. Opinions are important because whenever we need to make a decision we want to hear others' opinions. This is not only true for individuals but also for organizations. It becomes thus increasingly important for political institutions to get a feel of prevalent sentiment (positive or negative emotions) or opinions expressed by others about themselves as person or organization as well as on certain political topics.

In recent years, sentiment analysis or opinion mining has emerged as a distinct method to study people's opinions in terms of views, attitudes, appraisals and emotions towards entities, events and their attributes in a more thorough way (Liu 2010; Pang and Lee 2008). Until now, it is difficult for people to find relevant sites, extract related sentences with opinions, read them, summarize them, and organize them into usable forms. Automated opinion discovery and summarization systems are thus needed, which can be accomplished by sentiment analysis (Liu 2010). Basically, sentiment analysis can be performed based on two different approaches. The first one is the traditional dictionary-based classification of sentiment orientation including polarity (positive, negative, neutral) and strength, i.e., dictionaries of words annotated with their sentiment orientation are used to extract sentiment from text. Recently, sentiment analysis makes use of another approach based on machine learning where the classification of sentiment can be formulated as a learning problem with three classes: positive, negative and neutral. Here, sentiment classification can be performed based on supervised or unsupervised learning (Liu 2011). Most common supervised learning methods include naïve Bayesian and SVM classification, while unsupervised learning methods mostly make use of the natural language processing technique called part-of-speech (POS) tagging (see Liu 2011 and Takacs et al. 2007 for more details). Despite many advances, sentiment analysis still faces many problems regarding the nature of informal texts with emoticons, acronyms, amplifications, slang, and sarcasm or irony (particularly on Twitter and Facebook). In addition, negation and different contexts or domains might exacerbate the accurate classification of sentiment.

4.2.2.3 Structural approach Political institutions might be interested in identifying influential users or opinion leaders, in particular those who are ideologically or politically opposed to them. By monitoring those users, political actors might be able to have certain influence on those users' opinion making by some forms of intervention such as directly seeking dialogs with them. Sometimes, it might be even more helpful to not only to identify certain political opinion leaders, but also to detect influential politically relevant "communities", particularly those whose members frequently mention or talk about them. To find such influential individuals or communities, we suggest employing social network analysis methods. Social network analysis is a research field that studies the relations linking persons, organizations, interest groups, states, etc., by analyzing the structure of these relations (Scott and Carrington 2011). To identify influential users, one has to measure influence. There are a number of different measures of influence of an actor in a network. Basically, influence is determined by many factors, such as the novelty and resonance of their messages with those of their audience and the quality and frequency of the content they generate (Romero et al. 2011). In particular, resonance in

terms of retweets (Twitter) and comments (Facebook and blogs) can be modeled as edges connecting nodes that represent users within a social network. In this way, different metrics of the concept of centrality and prestige can be applied to measure influence (e.g., degree, betweenness or eigenvector centrality; degree, proximity or rank prestige) (Freeman et al. 1989; Wasserman and Faust 1994; Scott and Carrington 2011). Regarding detection of politically relevant communities, social network analysis might also be useful with different methods and algorithms. Community detection in social networks has attracted lots of attention in the domain of sociology (Gilbert et al. 2011). Some graph theoretical approaches such as the Girvan–Newman algorithm (Girvan and Newman 2002; Auber et al. 2003; Newman and Girvan 2004) or other clustering methods such as hierarchical, k-means and fuzzy c-means clustering (see Prabhu et al. 2010) have performed better than the others for the discovery of communities in social networks. Researchers have also shown interest in discovering changing clusters in dynamic data (Kalnis et al. 2005) and clustering the evolving data streams (Aggarwal et al. 2003).

4.2.2.4 Combination of methods In practice, there are many other analytical issues that are covered not only by a single, but rather multiple analysis approaches, which require a combination of different analysis methods. For example, besides identifying influential actors or opinion leaders, politicians and parties might also be interested in revealing political preferences of those actors as well as their opinions or sentiments on certain political topics. In such cases, text mining and sentiment analysis techniques might be applied in addition to social network analysis methods. As another example, sentiment analysis can be conducted on topics identified from the topic/issue-related analysis. However, analyses can also be performed the other way round. One can first collect emotionally charged postings and then try to identify topics from that dataset, i.e., extraction of emotionally charged topics.

The data analysis process as well as the whole framework described above is illustrated in Figs. 3 and 4, respectively.

5 Tracking and analysis examples

Given the growing relevance of social media, a large number of social media analytics tools have been developed for commercial use. However, little is known about the adopted methodological approaches. The proposed framework might help tool developers as well as researchers to systematically identify and implement appropriate methods for specific issues of analysis and research questions, respectively. For each of the most relevant analysis methods proposed in the framework (e.g., content analysis/text mining, opinion mining/sentiment analysis, social network analysis etc.), there exist already many tools. A large number of them provide programing interfaces to integrate functions into custom tools.

Some of the most important tools are given in the following:

- Automated content analysis/text mining: WordStat, LIWC, General Inquirer, etc.
- Manual content analysis/text mining: ATLAS.ti, QDA-Miner, The Ethnograph, etc.
- Opinion mining/sentiment analysis: SentiStrength, PolArt, SentiWordNet, etc.
- Social network analysis: Gephi, UCINET, Pajek, etc.

For illustrative purposes, this section provides some tracking and analysis examples as exemplary practical implementation of some aspects of our framework described above. We developed a software prototype, which is able to gather, store and analyze data from Twitter, Facebook and specific weblogs (for a detailed specification of the software, see Stieglitz and Kaufhold (2011)). We applied a simple test scenario to the real-world environment of political

Analysis Approach	Analysis Purpose		Analysis Method
	Reputation Management	General Monitoring	
topic/issue/trend-related	scanning and monitoring of political topics/issues relevant for own reputation	scanning and monitoring of political topics/issues	content analysis/ text mining
	detection of emerging political topics or trends relevant for own reputation	detection of emerging political topics or trends	trend analysis
opinion/ sentiment-related	detection of prevalent sentiment or perception related to own reputation	detection of prevalent sentiment or perception related to specific topics/events	opinion mining/ sentiment analysis
structural	identification of political opinion leaders relevant for own reputation	identification of political opinion leaders	social network analysis
	detection of politically relevant communities relevant for own reputation	detection of politically relevant communities	

Fig. 3 Data analysis

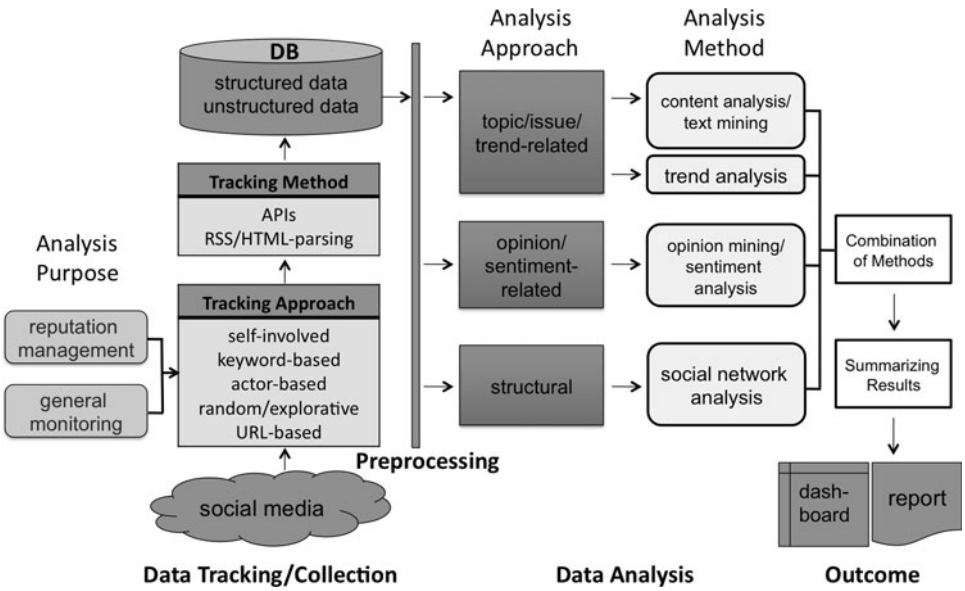

Fig. 4 Social media analytics framework in political context

Fig. 5 Tag cloud showing most relevant keywords occurring in tweets containing the tracking keyword "merkel"

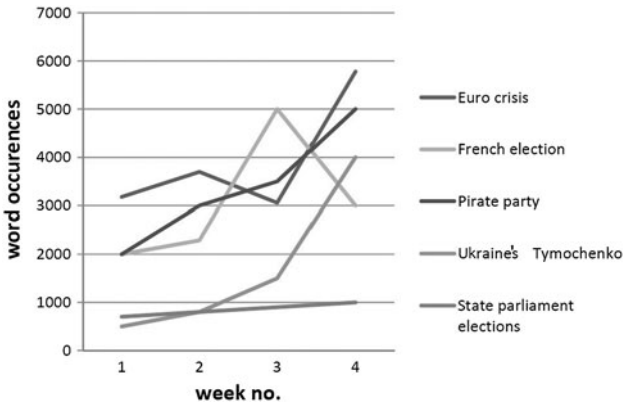

Fig. 6 Development of the top five identified topics during analysis period

communication in public social networks in Germany. More specifically, we chose to track and analyze data related to the current German chancellor Angela Merkel and her political party CDU, of which she is also the chairwoman, as well as other biggest political parties in Germany (SPD, FDP, The Greens, The Left, and Pirate Party). As an example of the integration of functionalities of other open-source or commercial tools into custom tool, for tracking and storing tweets we used the codes of "yourTwapperKeeper" (an open-source version of "TwapperKeeper.com", a free, publicly available online tool that allows users to download and archive tweets according to a variety of specifications).

First, during a time period of 1 month (April 2012), Twitter data based on the name of the chancellor as tracking keyword ("merkel") have been gathered and analyzed automatically. As a first analysis example, Fig. 5 shows a tag-cloud visualization of the most relevant identified keywords occurring in Twitter messages that contain the tracking keyword "merkel". We employed the engine of another software ("Wordle.net") to implement the tag-cloud visualization functionalities for our prototype.

Next, by applying automated content analysis techniques such as word occurrence and co-occurrence analysis on the dataset described above, our prototype could identify the top five topics and illustrate their dynamics during the analysis period (see Fig. 6).

As a final example of analysis, Fig. 7 shows a visual representation of a retweet network, in which users are connected if one has rebroadcast content produced by another. Users who are often retweeted are identified by larger node sizes (i.e., higher indegree). The darker the color of the node, the more active the user is at retweeting others' tweets (i.e., the higher the outdegree). We implemented the functionality of network visualization by means of the API provided by the open-source social network analysis software "Gephi".

In sum, the illustrative analysis examples shown above might all be parts of a regular report or real-time dashboard that aggregates analysis results from different social media

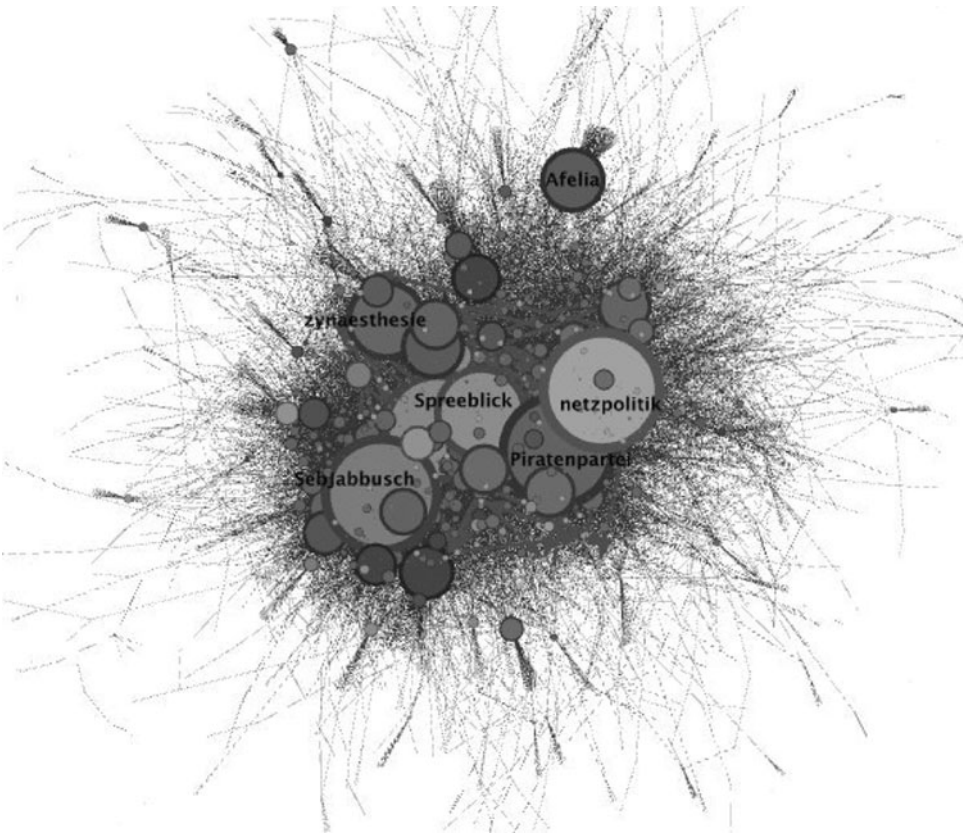

Fig. 7 Retweet network based on the dataset of tweets containing the tracking keyword "merkel"

platforms. Results of analysis should be presented in a way that the user (i.e., political institutions in our case) can make sense of it. The goal is to enable users to rapidly grasp the evolution in the social landscape and explore complex data without having to understand all the subtleties of the underlying system.

6 Conclusion

As previous studies have shown, in the last few years social media have become an important political communication channel. It enables political institutions and voters to directly interact with each other. Therefore, political activities might gain more transparency and citizens might be more involved into political decision-making processes. However, until now the potentials of political discussions in social media involving political institutions could not be exploited sufficiently. One reason for that is a lack of knowledge of politicians about current topics and discourses on different social media platforms. Based on an extensive literature review, we could observe the increasing relevance of and the need for analyzing political discussions on different social media platforms such as Twitter, Facebook, and weblogs.

As a main contribution, we proposed a framework for social media analytics in political context. More specifically, we outlined various approaches of data tracking and data analysis as well as corresponding analysis methods that might help gain a deeper insight into political discussions in social media. From the practical perspective, the framework should serve as a guideline for the development of toolsets aiming at collecting, storing, monitoring, analyzing, and summarizing politically relevant user-generated content from social media for political institutions. Even though our framework focuses on analyzing public data exclusively, we suggest that political actors should also reflect on ethical issues which might become relevant when investigating communication in social media. From the research perspective, the framework is of high relevance for the academic discussion because, to our knowledge, it is the first comprehensive summary of different social media analytics approaches and according analysis methods within the political sphere. Finally, we believe that this framework could also be adapted to other contexts such as business and marketing.

References

Adamic L, Glance N (2005) The political blogosphere and the 2004 U.S. election: divided they blog. In: Proceedings of the 3rd international workshop on link discovery, pp 36–43

Adar E, Adamic L (2005) Tracking information epidemics in blogspace. In: Proceedings of the IEEE/WIC/ACM international conference on web intelligence, pp 207–214

Adar E, Adamic L, Zhang L, Lukose R (2004) Implicit structure and the dynamics of blogspace. Paper presented at the 13th international World Wide Web conference

Aday S, Farrel H, Lynch M, Sides J, Kelly J, Zuckerman E (2010) Blogs and bullets: New media in contentious politics. Technical report, U.S. Institute of Peace

Aggarwal CC, Han J, Wang J, Yu PS (2003) A framework for clustering evolving data streams. In: Proceedings of VLDB'03, pp 81–92

Agrawal D, Budak C, El Abbadi A (2011) Information diffusion in social networks: observing and influencing societal interests. In: Proceedings of VLDB'11

Ammann SL (2010) A political campaign message in 140 characters or less: the use of Twitter by U.S. Senate Candidates in 2010. http://ssrn.com/abstract=1725477

Auber D, Chiricota Y, Jourdan F, Melancon G (2003) Multiscale visualization of small-world networks. In: North SC, Munzner T (eds) Proceedings of IEEE information visualization symposium, Seattle, USA. IEEE Computer Press, San Francisco, pp 75–81

Barabasi A (2002) Linked: the new science of networks. Perseus, Cambridge

Baumgartner JC, Morris JS (2010) Myfacetube politics: social networking web sites and political engagement of young adults. Soc Sci Comp Rev 28:24–44

Benkler Y (2006) The wealth of networks: How social production transforms markets and freedom. Yale University Press, New Haven

Bennett L (2003) New media power: the Internet and global activism. In: Couldry N, Curran J (eds) Contesting media power: alternative media in a networked world. Rowman & Littlefield, Littlefield

Blei DM, Ng AY, Jordan MI, Lafferty J (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022

Blei DM (2011) Introduction to probabilistic topic models. Comm ACM 55(4):77–84

Budak C, Agrawal D, El Abbadi A (2011) Structural trend analysis for online social networks. Technical Report UCSB/CS-2011-04, UCSB

Chadwick A (2006) Internet politics: states, citizens, and new communications technologies. Oxford University Press, New York

Chen D, Tang J, Li J, Zhou L (2009) Discovering the staring people from social networks. WWW'09: Proceedings of the 18th international conference on World Wide Web. ACM, New York, pp 1219–1220

Conover MD, Ratkiewicz J, Francisco M, Gonalves B, Flammini A, Menczer F (2011) Political polarization on Twitter. In: Proceedings of the 5th international conference on weblogs and social media

Creighton JL (2005) The public participation handbook: making better decisions through citizen involvement. Jossey-Bass, San Francisco

Doerfel M, Barnett G (1996) The use of Catpac for text analysis. Cult Anthropol Methods J 8(2):4–7

Facebook (2011) Facebook Official Statistics. http://www.facebook.com/press/info.php?statistics

Farrell H, Drezner D (2008) The power and politics of blogs. Public Choice 134(1):15–30

Freeman LC, White DR, Romney AK (eds) (1989) Research methods in social network analysis. Transaction books, New Brunswick

Gaffney D (2010) Iran election: quantifying online activism. In: Proceedings of WebSci'10: extending the frontiers of society online. USA, Raleigh, NC

Gilbert F, Simonetto P, Zaidi F, Jourdan F, Bourqui R (2011) Communities and hierarchical structures in dynamic social networks: analysis and visualization. Soc Netw Anal Min 1:83–95

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci USA 99:8271–8276

Golbeck J, Grimes JM, Rogers A (2010) Twitter use by the U.S. Congress. J Am Soc Inf Technol 61(8):1612–1621

Gruhl D, Nagarajan M, Pieper J, Robson C, Sheth A (2010) Multimodal social intelligence in a real-time dashboard system. Int J Very Large Data Bases 19(6):825–848

Hacker K, Coombs M, Weaver C, McCulloh G (2006) Possible uses of blogs and computer-mediated communication (CMC) for depolarizing political discourse. Paper presented to the Communication and Technology division (debate panel), Dresden, Germany

Herring SC, Scheidt LA, Bonus S, Wright E (2005) Weblogs as a bridging genre. Inform Technol People 18(2):142–171

Hillard D, Purpura S, Wilkerson J (2007) An active learning framework for classifiying political text. In: Annual Meeting of the Midwest Political Science Association, Chicago

Hofmann T (1999) Probabilistic latent semantic indexing. In: Proceedings of the 22nd annual international SIGIR conference on research and development in information retrieval

Hong S, Nadler D (2011) Does the Early Bird Move the Polls? The use of the social media tool 'Twitter' by U.S. politicians and its impact on public opinion. In: Proceedings of the International Conference on Digital Government Research

HuffPost Tech (2011) Twitter: we now have over 200 million accounts (update) http://www.huffingtonpost.com/2011/04/28/twitter-number-of-users_n_855177.html

Ingenhoff D, Röttger U (2008) Issues management. Ein zentrales Verfahren der Unternehmenskommunikation. In: Schmidt B (ed) Unternehmenskommunikation. Kommunikationsmanagement aus der Sicht der Unternehmensführung, pp 323–354

Kalnis P, Mamoulis N, Bakiras S (2005) On discovering moving clusters in spatio-temporal data. In: SSTD, pp 364–381

Kaplan AM, Haenlein M (2010) Users of the world, unite! The challenges and opportunities of Social Media. Bus Horiz 53(1):59–68

Karpf D (2009) Blogosphere research: a mixed-methods approach to rapidly changing systems. IEEE Intell Syst 24(5):67–70

Kasiviswanathan SP, Melville P, Banerjee A, Sindhwani V (2011) Emerging topic detection using dictionary learning. In: Proceedings of CIKM'11

Kavanaugh A, Fox EA, Sheetz S, Yang S, Li LT, Whalen T, Shoemaker D, Natsev P, Xie L (2011) Social media use by government: from the routine to the critical. In: Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, College Park, Maryland, 12–15 June, 2011

Kaye BK (2005) It's a blog, blog, blog world. Atl J Commun 13(2):73–95

Krippendorff K (2004) Content analysis. An introduction to its methodology, 2nd edn. The Sage Commtext Series. Sage Publications Ltd., London

Kushin M, Kitchener K (2009) Getting political on social network sites: Exploring online political discourse on Facebook. First Monday 14(11)

Larson K, Watson RT (2011) The value of social media: toward measuring social media strategies. In: Proceedings of ICIS 2011, Shanghai, China

Larsson A, Moe H (2011) Who tweets? Tracking microblogging use in the 2010 Swedish election campaign. In: ECIS 2011 Proceedings, Paper 251

Lassen DS, Brown AR (2011) Twitter: The electoral connection? Soc Sci Comp Rev 29(4):419–436

Leskovec J (2011) Social media analytics: tracking, modeling and predicting the flow of information through networks. In: Proceedings of WWW (Companion Volume) 2011, pp 277–278

Liu B (2010) Sentiment analysis: a multifaceted problem. IEEE Intell Syst 25:76–80
Liu B (2011) Web data mining: exploring hyperlinks, contents, and usage data. Springer, Heidelberg
Liu R, Guo W (2011) HMM-based state prediction for Internet hot topic. In: Proceedings of the IEEE International Conference on Computer Science and Automation Engineering (CSAE)
Mathioudakis M, Koudas N, Marbach P (2010) Early online identification of attention gathering items in social media. In: Proceedings of the third ACM international conference on Web search and data mining, New York, 4–6 February, 2010, USA
McAfee A (2006) Enterprise 2.0: the dawn of emergent collaboration. MIT Sloan Management Rev 47(3):20–28
McKenna L (2007) "Getting the word out:" policy bloggers use their soap box to make change. Rev Policy Res 24(3):209–229
McKenna L, Pole A (2007) What do bloggers do: an aver-age day on an average political blog. Public Choice 134(1):97–108
Monroe BL, Schrodt PA (2008) Introduction to the special issue: the statistical analysis of political text. Political Anal 16(4):351–355
Munson S, Resnick P (2011) The prevalence of political discourse in non-political blogs. In: Proceedings of 5th International Conference on Weblogs and Social Media (ICWSM 2011)
Nagarajan M, Sheth A, Velmurugan S (2011) Citizen sensor data mining, social media analytics and development centric web applications. In: Proceedings of the 20th international conference companion on world wide web (WWW'11), pp 289–290
Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys 69(2 Pt 2):026113
Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr 2(1–2):1–135
Paris C, Wan S (2011) Listening to the community: social media monitoring tasks for improving government services. Proc Ext Abstr CHI ACM 2011:2095–2100
Prabhu J, Sudharshan M, Saravanan M, Prasad G (2010) Augmenting rapid clustering method for social network analysis. Int Conf Adv Soc Netw Analysis Min 2010:407–408
Robertson SP, Vatrapu RK, Medina R (2010) Off the wall political discourse: Facebook use in the 2008 U.S. Presidential election. Information Polity 15:11–31
Romero DM, Galuba W, Asur S, Huberman BA (2011) Influence and passivity in social media. In: Proceedings of the 20th international conference companion on World wide web, pp 113–114
Rosen D, Barnett G, Kim JH (2011) Social networks and online environments: when science and practice co-evolve. Soc Netw Analysis Min 1(1):27–42
Scharkow M (2011) Thematic content analysis using supervised machine learning. An empirical evaluation using German online news. Quality Quantity (online first)
Schmidt J (2007) Blogging practices: an analytical framework. J Comput Mediat Commun 12(4):1409–1427
Scott J, Carrington PC (eds) (2011) Handbook of social network analysis. Sage, London
Sebastiani F (2002) Machine learning in automated text categorization. ACM Comput Surv 34(1):1–47
Stieglitz S, Kaufhold C (2011) Automatic full text analysis in public social media - adoption of a software prototype to investigate political communication. Procedia Comput Sci 5:776–781
Stieglitz S, Dang-Xuan L, Brockmann T (2012) Usage of social media for political communication in Germany. In: Proceedings of PACIS 2012, forthcoming
Stone P, Dunphy D, Smith M, Ogilvie D (1966) The general inquirer: a computer approach to content analysis. The MIT Press, Cambridge
Sunstein C (2002) The law of group polarization. J Political Philos 10(2):175–195
Takacs G, Pilaszy I, Nemeth G, Tikk D (2007) On the gravity recommendation system. In: Proceeding of KDD cup and workshop
Tumasjan A, Sprenger T, Sandner P, Welpe L (2011) Election forecasts with Twitter: how 140 characters reflect the political landscape. Soc Sci Comput Rev 29(4):402–418
Utz S (2009) The (potential) benefits of campaigning via social network sites. J Comput Mediat Commun 14:221–243
van Atteveldt W (2008) Semantic network analysis: techniques for extracting, representing, and querying media content. BookSurge Publishers, Charleston
Vitak J, Zube P, Sfmock A, Caleb T, Ellison N, Lampe C (2011) It's complicated: Facebook users' political participation in the 2008 election. Cyberpsychol Behav Soc Netw 14(3):107–114
Wartick S, Mahon J (1994) Toward a substantive definition oft he corporate issue construct—a review and synthesis of the literature. Business Soc 33:293–311
Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, New York
Wattal S, Schuff D, Mandviwalla M, Williams C (2010) Web 2.0 and politics: the 2008 U.S. presidential election and an e-politics research agenda. MIS Q 34(4):669–688
Wigand RT, Wood JD, Mande DM (2010) Taming the social network jungle: from web 2.0 to social media. In: Proceedings of the Americas Conference on Information Systems, Paper 416
Williams C, Gulati G (2007) Social networks in political campaigns: Facebook and the 2006 midterm elections. Annual Meeting of the American Political Science Association
Williams C, Gulati G (2009) Facebook grows up: an empirical assessment of its role in the 2008 congressional elections. Annual Meeting of the Midwest Political Science Association
Wu F, Huberman B (2004) Social structure and opinion formation. HP Labs Research Paper, Palo Alto
Yardi S, Boyd D (2010) Dynamic debates: an analysis of group polarization over time on twitter. Bull Sci Technol Soc 20:1–8
Zeng J, Zhang S, Wu C, Xie J (2007) Predictive model for internet public opinion. In: Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)
Zeng D, Chen H, Lusch R, Li S (2010) Social media analytics and intelligence. IEEE Intell Syst 25(6):13–16
Zhang W, Johnson TJ, Seltzer T, Bichard S (2010) The revolution will be networked: the influence of social networking sites on political attitudes and behavior. Soc Sci Comput Rev 28:75–92