Article

Social Science Computer Review
2015, Vol. 33(1) 3-20

DOI: 10.1177/0894439314521983
ssc.sagepub.com

Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters—Evidence From the United States and Italy

Andrea Ceron[1], Luigi Curini[1], and Stefano M. Iacus[1]

Abstract

In recent years, there has been an increasing attention in the literature on the possibility of analyzing social media as a useful complement to traditional off-line polls to monitor an electoral campaign. Some scholars claim that by doing so, we can also produce a forecast of the result. Relying on a proper methodology for sentiment analysis remains a crucial issue in this respect. In this work, we apply the supervised method proposed by Hopkins and King to analyze the voting intention of Twitter users in the United States (for the 2012 Presidential election) and Italy (for the two rounds of the centre-left 2012 primaries). This methodology presents two crucial advantages compared to traditionally employed alternatives: a better interpretation of the texts and more reliable aggregate results. Our analysis shows a remarkable ability of Twitter to "nowcast" as well as to forecast electoral results.

Keywords

sentiment analysis, text mining, text analytics, social media, electoral forecast

Introduction

The exponential growth of social media and social network sites, like Facebook and Twitter, has started to play a growing role on real-world politics in recent years. Social networks have been used, for example, to organize demonstrations and revolts during the "Arab spring" (Cottle, 2011; Ghannam, 2011);[1] to engage individuals in mobilizations (Bennett & Segerberg, 2011; Segerberg & Bennett, 2011); and to build social movements and political parties, like the Pirate Party in Sweden and Germany or the Italian *Movimento 5 Stelle*, which use the web to set the party line and to select candidates.[2]

The diffusion of social media also raises the possibility to delve into the web to explore and track the political and electoral preferences of citizens (Madge, Meek, Wellens, & Hooley, 2009; Woodly, 2007). As a matter of fact, scholars have recently started to explore social media as a device to assess

[1] Università degli Studi di Milano, Milano, Italy

Corresponding Author:
Luigi Curini, Università degli Studi di Milano, Via Conservtorio 7, Milano 20122, Italy.
Email: luigi.curini@unimi.it

the popularity of politicians (Gloor, Krauss, Nann, Fischbach, & Schoder, 2009), to track the political alignment of social media users (Barberá, 2012; Conover, Goncalves, Ratkiewicz, Flammini, & Menczer, 2011), and to compare citizens' political preferences expressed online with those caught by polls (O'Connor, Balasubramanyan, Routledge, & Smith, 2010). Analyzing social media during an electoral campaign can indeed be a useful supplement/complement of traditional off-line polls for a number of reasons (Xin, Gallagher, & Cao, 2010). Besides being cheaper and faster compared to traditional surveys, a social media analysis allows to monitor an electoral campaign day by day (at the extreme, hour by hour). Through that, the possibility to *nowcast* the campaign, that is to track in real-time trends and capture any sudden change (so-called "momentum": Jensen & Anstead, 2013) in public opinion well before of what can be done via traditional polls (as a result, e.g., of a TV debate: see subsequently), becomes a reality.[3] Some scholars, however, go even further than that, claiming that analyzing social media allows a reliable *forecast* of the final result (Tjong Kim Sang & Bos, 2012). This is quite fascinating cause forecasting an election is one of the *few* exercises on social events where an independent measure of the outcome that a model is trying to predict is clearly and indisputably available, that is, the vote share of candidates (and/or parties) at the ballots.

Some of these works rely on very simple techniques, focusing on the volume of data related to parties or candidates. For instance, Véronis (2007) proved that the number of candidate mentions in blog posts is a good predictor of electoral success and can perform better than election polls. Along the same line, some scholars claimed that the number of Facebook supporters could be a valid indicator of electoral fortunes (Upton, 2010; Williams & Gulati, 2008), while Tumasjan, Sprenger, Philipp, and Welpe (2011) compared party mentions on Twitter with the results of the 2009 German election and argued that the relative number of tweets related to each party is once again a good predictor for its vote share.

By noting that the mere count of mentions or tweets or the number of followers or "like" can be a rather crude way to provide an accurate foresight (Chung & Mustafaraj, 2011), other studies have tried to improve this stream of research by means of sentiment analysis.[4] Lindsay (2008), for example, built a sentiment classifier based on lexical induction and found correlations between several polls conducted during the 2008 presidential election and the content of wall posts available on Facebook. O'Connor, Balasubramanyan, Routledge, and Smith (2010) show similar results displaying correlation between Obama's approval rate and the sentiment expressed by Twitter users. In addition, sentiment analysis of tweets proved to perform as well as polls in predicting the results of both the 2011 (Tjong Kim Sang & Bos, 2012) and the 2012 legislative elections in the Netherlands (Sanders & den Bosch, 2013), while the analysis of multiple social media (Facebook, Twitter, Google, and YouTube) was able to outperform traditional surveys in estimating the results of the 2010 U.K. Election (Franch, 2013).

Still not all enquiries succeeded in correctly predicting the outcome of the elections (Gayo-Avello, Metaxas, & Mustafaraj, 2011; Goldstein & Rainey, 2010; Huberty, 2013). For instance, it has been shown that the share of campaign weblogs prior to the 2005 federal election in Germany was not a good predictor of the relative strength of the parties insofar as small parties were overrepresented (Albrecht, Lübcke, & Hartig-Perschke, 2007). In a study on Canadian elections, Jansen and Koop (2005) failed in estimating the positions of the two largest parties. Finally, Jugherr, Jürgens, and Schoen (2012) criticized the work of Tumasjan et al. (2011), arguing that including the small German Pirate Party into the analysis would have yielded a negative effect on the accuracy of the predictions.

Gayo-Avello (2011, 2012) pinpoints several theoretical problems with predicting elections based on tweets. First, he stresses how several of the quoted works are not predictions at all, given that they generally present post hoc analysis after an election has already occurred. This, inter alia, also increases the chances that only good results are published, inflating the perceived ability of using social media to correctly forecast election (see also Lewis-Beck, 2005). Second, he underlines the difficulty to catch the real meaning of the texts analyzed, given that political discourse is plagued with humor, double meanings, and sarcasm. Third, he highlights the risk of a spamming effect: Given the presence of rumors and misleading information, not all the internet posts are necessary trustworthy. Finally, in most of the previous studies, demographics are neglected: Not every age, gender, social, or racial group is in fact equally represented in social media.

In the present article, we show how the second and (at least partially) the third of the above concerns can be addressed by relying on a proper methodology for sentiment analysis. With respect to the first concern, on the other side, all the analyses that we discuss here have been conducted (and published on media) before the day of the election, so they can be considered as real predictions. More in detail, by employing the method proposed in Hopkins and King (2010; HK, from now on), we will analyze two different scenarios: the American electoral campaign for the 2012 Presidential election and the first and second round of the primary elections held by the Italian centre-left coalition in November 2012.[5] By focusing on these two cases, we have followed the most different system design setting. Indeed, the cases differ both in terms of the type of election considered (a national election devoted to select the head of state vs. an election aimed to select the leader of a political coalition running in the next national election) and in the type of competition involved (a "single-issue" election in which the preference eventually expressed by Internet users involved mainly a choice among two options vs. an election in which Internet users could choose to express a preference among a larger number of potential viable *targets*, at least in the first round). Moreover, in the United States, the rate of penetration of Twitter,[6] that is, the social network on which we will focus here, is considerably larger than in Italy (11% vs. 5%), implying an Internet community not necessarily identical in the two cases.[7] Such dissimilarities are clearly important for exploring the robustness of our results and for controlling the potentiality of a method that, for a number of different reasons explained subsequently, seems to be an advance compared to other available methods. In the conclusion, we present some suggestions for future research.

How to (Effectively) Scrutinize Voters' Preferences Through Social Media

Nowadays, Internet access is available to a wider audience of citizens (and voters). In turn, the usage of social media is growing at very fast rates. Around 35 of the 100 people got access to the Internet-users community web, all over the world, in 2011 (approximately 2.5 billion people).[8] Among them, 72% of the Internet population is active on at least one social network, like Facebook (over 1.100 million of users) or Twitter (over 500 million of users).[9]

Given the wide amount of data on public opinion available online (and its growing relevance), monitoring this flow of preferences becomes a relevant task. The problem is to select the kind of method more appropriate in this regard. While earlier studies, as already discussed, focused mainly on the volume of data (related, for instance, to each party or candidate), here we aim to catch the attitude of Internet users going beyond the mere number of mentions. To this aim, we will employ the method recently proposed in Hopkins and King (2010).

The HK method presents *two* specific advantages compared to other methods, especially when it comes to relate social media and (the results of) elections: a better *interpretation* of the texts and more *reliable* aggregate results. The former advantage is mainly linked to the fact that HK performs a supervised sentiment analysis. The traditional approach to sentiment analysis is in fact based on the use of ontological dictionaries: This means that a text is assigned to a specific opinion category if some predetermined words or expressions appear (or do not) in the text (see Grimmer & Stewart, 2013, on this point). The benefit of this approach is, of course, the possibility to implement a totally automated analysis (once the dictionary has been defined). The strong drawback, on the other side, is the difficulty in classifying opinions expressed through ironic or paradoxical sentences, or in appreciating all the language nuances (specific jargons, neologisms, etc.): The informal expression "*what a nice rip-off!*", for instance, is quite ambiguous from the viewpoint of an ontological dictionary, because it includes both a positive and a negative term.

The HK method, on the contrary, is based on a two-stage process. The first step involves human coders and consists in reading and coding a subsample of the documents downloaded from some Internet source. This subsample—with no particular statistical property, see subsequently—represents a *training set* that will be used by the HK algorithm to classify all the unread documents in the second stage. Human coders are of course more effective and careful than ontological dictionaries in recognizing all the previously discussed language specificity and the author's attitude toward the subject (Hopkins & King, 2010). Moreover, human coding is better suited to identify the (ever-present) problem of spamming in social communication. This is of course important, given that spamming can have an impact on the accuracy of the final result. At the second stage, the automated statistical analysis provided by the HK algorithm extends such accuracy to the whole population of posts, allowing for properly catching the opinions expressed on the web.

The methodology relies on the assumption that the opinion of people posting on social networks can be deduced by all the terms they use: not only the terms explicitly related to the topic they talk about but also the "neutral" part of the language commonly used. Therefore, in order to characterize the different opinions, the single units (internet posts) in the data set are decomposed into their own single words: Consequently, each unit is represented by the vector of the terms used, which we call "word profile" of the unit.[10]

The formal background of the method is rather simple. The word profiles used in the text units are indicated by **S** and people's opinions expressed in the texts are indicated by **D**. The target of estimation is $P(\mathbf{D})$, that is, the frequency distribution of the opinions over the posting population.

The standard statistical approach is to decompose $P(\mathbf{D})$ in the following way:

$$P(\mathbf{D}) = P(\mathbf{D}|\mathbf{S})\,P(\mathbf{S}). \tag{1}$$

$P(\mathbf{S})$ corresponds to a tabulation of frequencies of word profiles in the whole population of texts. $P(\mathbf{D}|\mathbf{S})$ is estimated from the training set as $P_{\mathrm{T}}(\mathbf{D}|\mathbf{S})$, that is, the conditional frequency distribution of word profiles inside the training set, using any standard classifier (multinomial regression, classification trees, random forests, support vector machines, etc.).[11] Through this approach, each individual classification of posts in the *test set* (i.e., post belonging to the corpus of texts but not to the training set) is assigned to some category $\mathrm{D_i}$ with some probability, that is, for a text j in the test set, with word profile $\mathrm{S_j}$, its category is estimated through $P_{\mathrm{T}}\left(\mathbf{D}_i|\mathbf{S}_j\right)$ for $i = 1, 2,\ldots, k$. Then, the aggregated distribution of opinions $P(\mathbf{D})$ of all texts in the corpus is obtained by aggregating individual classifications, each with its own misclassification error. As a result, the individual misclassification error does not vanish due to aggregation but may easily propagate up to the extent that, in many applications with thousands or millions of texts, one could see the error rising up to 15–20%. This is clearly *quite* problematic if one is mainly interested in estimating some kind of aggregate measure through the analysis of social media, as it happens with all the researches that want to map, somehow, tweets into votes.

HK theory is effective in that it reverses the previous approach and, instead of estimating the individual opinion and the aggregating, it aggregates all word profiles and estimates the aggregated distribution of opinion directly, leading to an error of the order of 2–3%.[12] Accordingly, being an approach for analyzing texts which does not aim to classify the individual documents into categories, but to measure directly the proportion of documents in each category, represents the second (statistical) advantage of the HK method (see also the discussion in Grimmer & Stewart, 2013). More in detail, the frequency distribution of the terms $P(\mathbf{S})$ can be expressed as:

$$P(\mathbf{S}) = P(\mathbf{S}|\mathbf{D})\,P(\mathbf{D}). \tag{2}$$

Table 1. A Typology for Classifying Texts Posted on Social Network Sites.

	Method to Estimate the Distribution of Opinions	
	Individual	Aggregate
Method to classify texts		
Unsupervised	Counting mentions, ontological dictionaries	=
Supervised	=	HK method

Note. HK = Hopkins and King.

The frequency distribution $P(\mathbf{S})$ can be evaluated tabulating all the texts posted, and it requires only some computer time and no debatable assumption. The conditional distribution $P(\mathbf{S}|\mathbf{D})$ cannot be observed and must be estimated by the hand coding of the training set of texts.

The hand coding of the training text, in fact, allows for calculating $P_{\mathrm{T}}(\mathbf{S}|\mathbf{D})$, that is, the conditional frequency distribution of word profiles inside the training set. The assumption—and the reasonable requirement—of the method is that the texts of the training set are homogeneous to the whole data set, that is, they come from the same "world" the rest of the data set comes, such that one can assume that:

$$P_{\mathrm{T}}(\mathbf{S}|\mathbf{D}) = P(\mathbf{S}|\mathbf{D}). \tag{3}$$

If this is the case, the frequency distribution of the opinions can be consistently estimated, because both $P(\boldsymbol{S})$ and $P_{\mathrm{T}}(\mathbf{S}|\mathbf{D})$ are observable. Therefore, by Equation 2 and noticing that $P_{\mathrm{T}}(\boldsymbol{S}|\boldsymbol{D})$ and $P(\boldsymbol{S}|\boldsymbol{D})$ are both matrixes, we have

$$P(\mathbf{D}) = P(\mathbf{S}|\mathbf{D})^{-1}P(\mathbf{S}) = P_{\mathrm{T}}(\mathbf{S}|\mathbf{D})^{-1}P(\mathbf{S}), \tag{4}$$

where $P_{\mathrm{T}}(\mathbf{S}|\mathbf{D})^{-1}$ is the inverse matrix of $P_{\mathrm{T}}(\mathbf{S}|\mathbf{D})$, similarly for $P(\mathbf{S}|\mathbf{D})^{-1}$. It is worth remarking that—while the homogeneity of the training set to the data set is required—no statistical property must be satisfied by the set: In particular, the training set is not a representative sample of the population of texts.

Table 1 summarizes the methods available in the literature to analyze social media (as well as any kind of text in digital forms: see Grimmer & Stewart, 2013) according to the two criteria discussed previously: first, the method employed to classify texts (unsupervised vs. supervised one); second, the method adopted to estimate the overall distribution of opinions of the classified texts (aggregation of the individual classification of all posts vs. direct measurement of the aggregate distribution). Traditionally, the works on social media and elections have focused on the upper-left cell by counting the number of mentions related to a candidate, for example, or by employing ontological dictionaries. The HK method, on the other side, focuses on the opposite cell, that is, on the lower-right cell. The remaining two cells of the table have been up to now scarcely visited by scholars.

In the two empirical cases on which we focus in the present article, we analyze social media (more precisely Twitter: see subsequently) in an effort to monitor the ongoing electoral campaign as well as to predict the final electoral result. In this respect, $P(\mathbf{D})$ refers to the (aggregate) propensity to vote for each candidate/party.[13] In particular, we considered a tweet as casting a "vote" for a candidate/party only if at least one of the following three conditions is satisfied: (a) the tweet includes an explicit statement related to the intention to vote for a candidate/party; (b) the tweet includes a statement in favor of a candidate/party together with an hashtag[14] connected to the electoral campaign of that candidate/party; (c) the tweet includes a negative statement opposing a candidate/party together with an hashtag connected to the electoral campaign of *another* candidate/party. Considering not simply a generic positive statement, but a positive statement *plus* an hashtag permits to focus on those signals that by being more "costly" in terms of self-exposition by the Twitter user (including an hashtag in a message denotes after all a clear stance) are also more credible (on this point, see the large literature on signaling games: Banks, 1991). On the other hand, condition (c) allows to reduce the arbitrariness in the "supervised" stage of the analysis. This applies also to a (largely) two-candidate case such as the U.S. Presidential race. For example, if a tweet says "*do not vote for Romney*," this does not necessarily imply that the person who wrote that post will then vote for Obama. He could decide to vote for a third candidate or also to abstain. In a multiparty race, of course, this problem just strengthens. However, going back to the previous example, a hypothetical tweet such as "*do not vote for Romney. #fourmoreyears*" would be counted, according to our classification, as a vote in favor of Obama, given that *#fourmoreyears* has been one of most largely used hashtag supporting the Obama's electoral campaign.

Similarly, we counted all the retweets (i.e., a message rediffusion by a Twitter user of a message posted by another user) that satisfy the previous conditions as a "vote" for a candidate/party. Although retweeting, strictly speaking, does not imply the production of new information, it implies that someone else thought a communication was valuable for herself (Jensen & Anstead, 2013). On the other side, if it is true that the act of retweet does not necessarily imply an "endorsement" by the user who retweets, it is also true that when the retweet includes a text in which an intention to vote a given candidate/party is clearly expressed or where an identifiable hashtag connected to a candidate/party is presented, it becomes a costly act, exactly for the same reasons already noted previously. As a consequence, it should happen only when a Twitter user shares to a large extent the content of the tweet and the underlying connected vote.

Broadly speaking, there are several social media that could be analyzed. Here, we will focus, as already noted, on Twitter, a social network for microblogging (Jansen, Zhang, Sobel, & Chowdury, 2009) that experienced a sharp growth in the last years. When we come to the countries analyzed in this work, we observe that in 2012, Twitter was the second most used social network both in the United States (around 23 millions of users) and in Italy (more than 4 millions). A further crucial advantage of Twitter, which makes it so popular in the literature on social media analysis, is that all the posts by users can be freely accessible, contrary to other social networks. Moreover, it also allows to geolocalize the origin of the tweet, therefore permitting a more fine-grained analysis (as we will see subsequently). To download the data employed in this article, we have relied on the social media monitoring engine Voices from the Blogs (http://voicesfromtheblogs.com/),[15] while the analysis have been run in R.

Electoral Campaign and Social Media (1): The 2012 U.S. Presidential Election. Too Close to Call?

The first political context in which we explore the usefulness of a social media analysis concerns the 2012 U.S. Presidential election. From September 28 to November 6, we monitored the voting intention expressed on Twitter toward the four main candidates: Barack Obama (Democratic Party), Mitt Romney (Republican Party), Gary Johnson (Libertarian Party), and Jill Stein (Greens). In this lapse of time, we estimated the political preferences of American voters on a daily basis by analyzing more than 50 million of tweets, a bit more than 1 million of tweets per day. The data are calculated as a moving average along those 7 days (following what suggested in O'Connor et al., 2010).[16] These results are summarized in Figure 1.

The fluctuation of the preferences expressed online closely follows the main events that happened during the electoral campaign. For instance, while in September Obama retained a wide margin over his main opponent, the wind started to change at the beginning of October, when Romney was able to reduce the distance from the Democratic candidate. In particular, as Obama performed poorly during the first TV debate, Romney overcame the former President in the voting intention of Twitter

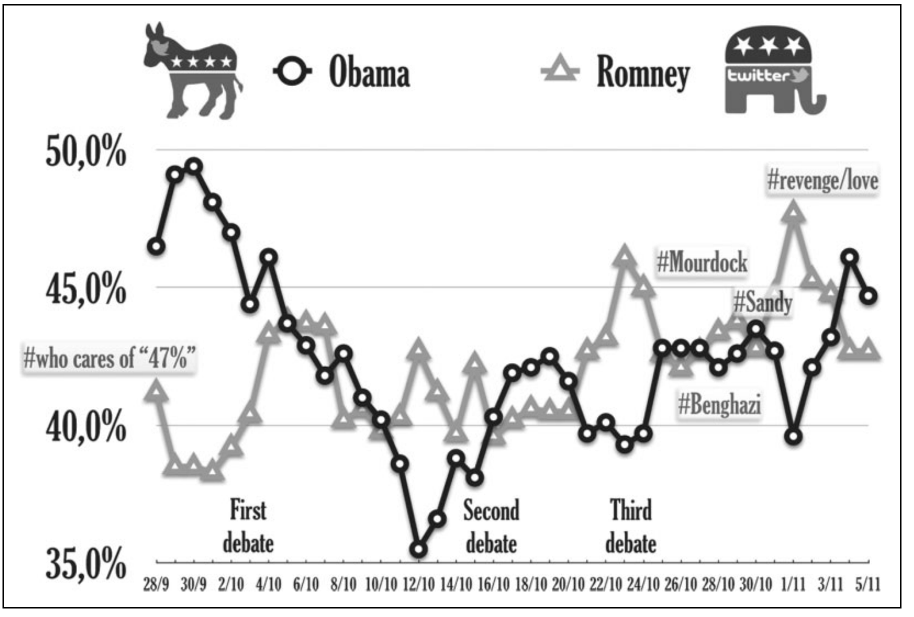

Figure 1. U.S. Presidential 2012: The trend of Twitter votes for Obama and Romney.

users. In this respect, it is interesting to note that traditional survey polls revealed a similar trend only in the following days of the first TV debate, given the relatively slow pace of polling. This highlights what was previously noted about the ability of social media analysis to nowcast any "momentum" during an electoral campaign. The second debate represents another turning point that arrested the loss of support for Obama. In the last days before the election, some political scandals (like the "Benghazi gate" or the statement against the abortion in case of rape, pronounced by the member of the "Tea Party", Richard Mourdock)[17] and exogenous events (the Sandy hurricane) jumped into the campaign wielding advantages to one candidate or the other.

In the very last days, the online sentiment highlighted a positive trend toward Obama, and our final prediction made on November 6 forecasted a victory for Obama in the popular vote with a clear and safe margin of 3.5%. As the real gap in the share of votes was 3.9%, our forecast proved (surprisingly) to be more accurate than those made by traditional survey polls that on average assigned only a narrow margin in favor of Obama (+0.7), claiming that the race was "too close to call".[18]

Besides the popular vote, we also tried to predict the results in the "swing states," that is, those where the race is usually very close and few thousands of votes can alter the balance between the candidates and the outcome of the whole Presidential election.[19] In 2012, the surveys focused on 11 "swing states" and among these they considered Florida, Ohio, and Virginia as the main battlegrounds. We therefore paid attention to those races. In Table 2, we report the gap between Obama and Romney in each state according to three different measures. The first one (labeled HK) consists in the method of sentiment analysis discussed so far. For the "swing states" estimates, we replicated our method considering the pools of tweets geotagged in each State only. The second (R) is the gap displayed on November 6 on the website www.realclearpolitics.com, which recorded the average of the survey polls issued on the last week before the election. The third one (V) represents the actual gap between the two candidates after votes have been counted. Then, we display the difference between the forecasts (made either through sentiment analysis or surveys) and the actual votes. Finally, we highlight which prediction has been the best one according to the ability to correctly

Table 2. U.S. Presidential 2012: Accuracy of the Predictions. Comparison Between HK Method and Survey Polls (R) Estimates With the Actual Results (V).

State	Gap (HK)	Gap (R)	Gap (V)	|HK-V|	|R-V|	Best prediction
Popular vote	Obama +3.5	Obama +0.7	Obama +3.9	0.4	3.2	HK
Florida	Obama +6.1	Romney +1.5	Obama +0.9	5.2	2.4	HK
Ohio	Obama +2.9	Obama +2.9	Obama +3.0	0.1	0.1	=
Virginia	Obama +3.5	Obama +0.3	Obama +3.9	0.4	3.7	HK
Colorado	Romney +1.3	Obama +1.5	Obama +5.4	4.1	3.0	R
Iowa	Obama +4.8	Obama +2.4	Obama +5.8	1.0	3.4	HK
Nevada	Obama +3.3	Obama +2.8	Obama +6.7	3.4	3.9	HK
New Hampshire	Obama +3.8	Obama +2.0	Obama +5.6	1.8	3.6	HK
North Carolina	Romney +3.0	Romney +3.0	Romney +2.0	1.0	1.0	=
Michigan	Obama +5.5	Obama +4.0	Obama +9.5	4	5.5	HK
Pennsylvania	Romney +2.5	Obama +3.8	Obama +5.4	2.9	1.6	R
Wisconsin	Obama +7.4	Obama +4.2	Obama +6.9	0.5	2.7	HK

Note. HK = Hopkins and King.

predict the winner. When both sentiment analysis and survey polls predicted the same winner, we discriminate by measuring the difference between the expected and the actual gap.

Overall, analyzing social media correctly predicts the winner in 9 of the 11 swing states, Colorado and Pennsylvania being the only two exceptions. Furthermore, in a plurality of states (7 against 2), our data (HK) proved to be more accurate than the average of polls (these states are Florida, Iowa, Virginia, Nevada, New Hampshire, Michigan, and Wisconsin), while in the remaining two swing states (Ohio and North Carolina), the different forecasts (social media vs. surveys) performed in a similar manner.

The most interesting results concern the three main battlegrounds: Ohio, Virginia, and Florida. In Ohio, both tools predicted similar results. On the contrary, in Virginia, our data could catch the trend pro-Obama that emerged in the last days (and in the last hours) when the Democratic staff mobilized the partisan voters (even during the electoral night they pushed voters to stay in line by means of Twitter messages). The same happened in Florida, where our prediction was claiming a victory for Obama, with a safe margin. Eventually, these results could be explained by our ability to measure the voting intention of the Hispanic voters who could be less likely to answer to survey polls.

This goes back once again to the specific method we employed to analyze social media. In fact, while the HK-supervised method for sentiment analysis was remarkably able to catch the voting intention of U.S. citizens, other methods adopted to analyze social media in the same occasion were less successful. First of all, it is interesting to note that at the beginning of the electoral campaign, Obama had almost 16.8 million followers on Twitter, while Romney had not even hit 600,000. Despite such (huge) disparity, our results underlined a different story that was not only remarkably in line with the actual votes, as we have discussed, but that also illustrated a social media support for the two main competitors which was much more volatile compared to what we could have expected by looking at the number of followers only (see Figure 1). In this sense, our results confirm that the number of Facebook friends or Twitter followers on their own is largely misleading as predictors of election outcomes (see Cameron, Barrett, & Stewardson, 2013, on this point). This happens also because Twitter users are often divided between those who follow leaders they agree with and those who also follow political figures they disagree with (see Pamelee & Bichard, 2011).

The same unsatisfactory prediction could have happened if we had counted the number of mentions (M) related to the different candidates running in the 2012 U.S. presidential election (see Washington, Parra, Thatcher, LePrevost, & Morar, 2013). Once again, this result is not surprising, given that the sheer number of mentions related to a candidate gives just a measure of the notoriety

Table 3. Comparison of the Accuracy of Twitter Forecast Made Through Mentions, Automated Sentiment Analysis, and HK Method (U.S. 2012, Popular Vote).

Analysis	Source	MAE
M	Washington, Parra, Thatcher, LePrevost, and Morar (2013)	17.90
T	Topsy	3.63
W1	Washington, Parra, Thatcher, LePrevost, and Morar (2013)	1.80
W2	Washington, Parra, Thatcher, LePrevost, and Morar (2013)	16.00
C1	Choy, Cheong, Laik, and Shung (2012)	1.29
C2	Choy, Cheong, Laik, and Shung (2012)	0.47
HK	—	0.02

Note. HK = Hopkins and King; MAE = mean absolute error.

in the web of such candidate (either for positive *or* negative reasons), without any necessary connection with her (expected) voting share. Techniques of automated sentiment analysis were sometimes more useful in the U.S. 2012 elections, although their accuracy remains lower compared to the HK method. For instance, Twindex (T), the Twitter Political Index developed by Topsy and Twitter itself, estimated a wide margin for Obama the day ahead of the election in terms of positive sentiment compared to Romney (74% against 59%).[20] Washington, Parra, Thatcher, LePrevost, and Morar (2013), on the other side, found contrasting evidence since they were able to get accurate estimates only when using the algorithm provided by the social media marketing platform Radian6 (W1), while their results get much worse when applying other dictionary-based techniques (W2).[21] Finally, Choy, Cheong, Laik, and Shung (2012) show that automated sentiment analysis can wield good predictions (C1) even though this accuracy approaches the one reached by the HK technique only when the results are weighted by some census information, such as preexisting party affiliations of the American voters (C2).[22]

Table 3 provides an evaluation of the different techniques in the U.S. 2012 Presidential election case. To allow a comparison between these studies, we considered only the two main candidates, Obama and Romney, and normalized the Twitter results and the popular vote data accordingly. We then measured the mean absolute error (MAE) of each prediction. Table 3 simply shows the difference, in terms of MAE, between the HK sentiment analysis and the other automated techniques discussed previously. The table illustrates that the HK sentiment analysis clearly performs better than the alternatives, at least in the U.S. 2012 case.

Electoral Campaign and Social Media (2): The Selection of the Centre-Left Coalition Leader in Italy

We double checked the predictive skills of a social media analysis by applying this technique to monitor the first and second rounds of the primary elections held by the centre-left coalition "Italia Bene Comune," to select the leader of the alliance. This is an intriguing exercise, given the particularly complex environment of a primary election, that makes the possibility of an analysis particularly burdensome (American Association of Public Opinion Research, 2009): Primary elections are indeed a contest that involve typically a partisan electorate, a larger number of viable candidates, with less ideological differentiation than one finds in the general election (a fact that makes the electoral choices by voters more costly), and lower turnout (Jensen & Anstead, 2013). The fact that in Italy the primary elections are not legally recognized as such makes things just harder.

From October 6 to November 25 (the day the election was held), we analyzed no less than 500,000 tweets on several points in time to assess the voting intention of Internet users toward the five different candidates: Pierluigi Bersani (head of the Democratic Party), Matteo Renzi (Mayor of

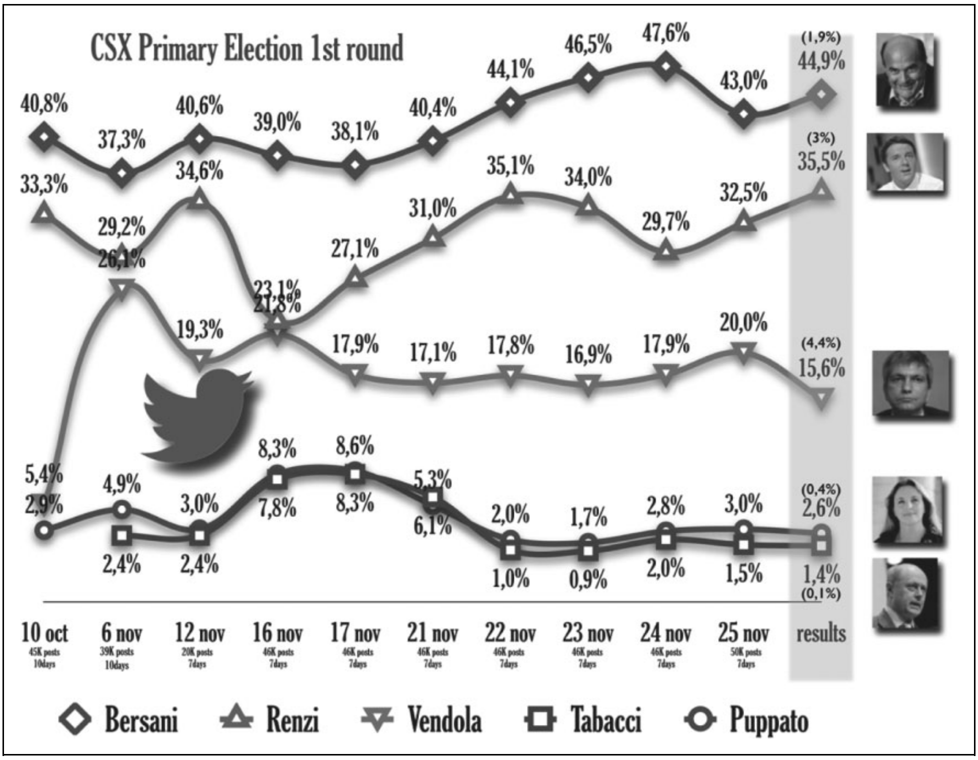

Figure 2. Candidates share of votes according to Twitter forecast. Comparison with actual results.

Florence), Nichi Vendola (governor of Apulia and leader of the left-wing party Left Ecology and Freedom), Laura Puppato (Democratic Party whip in Veneto), and Bruno Tabacci (Milan budget councillor). Figure 2 reports the fluctuation of voting intentions according to our analysis. We reported our estimates measured in 10 different days by analyzing around 40,000–50,000 tweets released in a time span that ranges from 10 days (at the beginning of the electoral campaign) to 1 week (since November 12).

Figure 2, as it happened with the American case, allows once again to monitor how the distribution of preferences changed over the campaign, as well as the different "momentums" that characterized such campaign. To start with, Bersani's expected vote share has been always around 40%, while we observed an increase during the last week, when the voting intentions grew up to 47.6% (note that a similar trend was also reported by several polls). However, in the last 2 days before the elections, this value shrank back to 43%, closer to the actual result (we mistook by 1.9% only). Renzi was always ranked second since the beginning. His support has been on average around 31%, with a peak on November 12, when the candidates were involved in a debate on SKY television, followed by a loss after the debate. Then his expected votes share started to rise again after the convention (called "Leopolda 2012") held in Florence by Matteo Renzi and his supporters. Nichi Vendola, who actually ranked third, started his campaign only in November after he was discharged from a prosecution.[23] This combination of events enabled him to win an initial large share of support, which declined in a few days when such effect vanished, bringing Vendola's expected vote share to around 18%. Finally, the two minor candidates, Puppato and Tabacci, retained only a low share of votes (according to our forecast) during the whole campaign, except after the debate when they took advantage from an outstanding public visibility.

Table 4. Comparison of the Accuracy of Twitter Forecast and Survey Polls in the Italian Primary Election of the Centre-Left Coalition (first round).

Survey Polls	Day of Publication of the Survey	MAE	Gap Bersani–Renzi
Popular vote	—	—	Bersani +9.4
HK	November 25, 2013	1.96	Bersani +10.5
Ipr	November 19, 2013	1.64	Bersani +5.0
Piepoli	November 19, 2013	2.16	Bersani +11.0
Ipsos	November 21, 2013	1.06	Bersani +8.42
CISE	November 22, 2013	2.48	Bersani +10.6
SWG Agorà	November 22, 2013	1.80	Bersani +14.0
Tecnè	November 25, 2013	2.20	Bersani +16.9

Note. HK = Hopkins and King; MAE = mean absolute error.

Table 5. Comparison of the Accuracy of Twitter Forecast and Surveys Polls in the Italian Primary Election of the Centre-Left Coalition (Second Round).

	Day of Publication of the Survey	Bersani	Renzi	Gap
Popular Vote	—	61.1	38.8	Bersani +22.3
HK	December 1, 2012	58.4	41.6	Bersani +16.8
Piepoli	November 25, 2012	59	41	Bersani +18
Ips	November 26, 2012	56	44	Bersani +12
ISPO	November 27, 2012	56.5	43.5	Bersani +13
SWG	November 28, 2012	55	45	Bersani +10
COESIS	November 28, 2012	54	46	Bersani +8
Quorum	November 28, 2012	56.4	43.6	Bersani +12.8
Ipsos	November 29, 2012	57.5	42.5	Bersani +15

Note. HK = Hopkins and King.

Figure 2 also reports the actual votes share and highlights, per each candidate, the absolute difference between our last forecast and the final results. The gap between our estimates and the results is very narrow being on average below 2%. This error is in line with the average error provided by the surveys polls issued in the last week, which is 1.9%. This also appears from Table 4 where we display the mean absolute error of our prediction along with those of polls. What is more, our technique succeeded in predicting the gap between the two foremost candidates, Bersani and Renzi, better than the traditional survey polls. Indeed, according to our results, the gap between the two candidates was 10.5% while Bersani was leading by 9.4 points after votes have been counted. This means a difference of 1.1% while on average the polls mistook the magnitude of the gap by 3%.

We were able to produce a similar accurate forecast of the Italian centre-left Primaries also in the second round. Table 5 reports the results of our forecast by using Twitter and the actual results according to several surveys published in the last week before the second round. In this case, our analysis on almost 25,000 tweets posted between Thursday, November 29 and Saturday, December 1 (the night ahead of the second round) predicted a clear victory for Bersani (58.4% vs. 41.6% for Renzi). At the ballot, Bersani won with 61.1% of votes against 38.8% for Renzi.

According to these results, social network sites confirm themselves as sources of valuable information that can be exploited to carry out electoral forecasts. However, the goodness of such forecasts once again seems to depend on the technique adopted. Table 6 portrays a comparison of the MAE obtained by the HK method versus the ones arising from several other social media analyses

Table 6. Comparison of the Accuracy of Twitter Forecast Made Through Mentions, Automated Sentiment Analysis, and HK Method (Italian Primary Election of the Centre-Left Coalition, First Round).

Analysis	Source	MAE
Mentions 1	http://seigradi.corriere.it/2012/11/25/le-primarie-del-centrosinistra-su-twitter-vincono-renzi-e-vendola/	6.36
Mentions 2	http://www.chefuturo.it/2012/11/twitter-la-tv-e-i-voti-reali-analisi-del-primo-round-delle-social-primarie/	9.72
Automated SA	http://vincos.it/2012/11/25/primarie-centro-sinistra-citazioni-e-performance-online-dei-candidati/	8.65
HK	—	1.96

Note. HK = Hopkins and King; SA = sentiment analysis; MAE = mean absolute error.

conducted on the first round of the Italian Primary elections that considered either the volume of mentions or the positive sentiment measured through dictionary-based methods.[24] As can be seen, the MAE increases a lot in the latter types of analyses. What is worst, such analyses were clearly (and consistently) highlighting a strong advantage for the actual second-ranked candidate, Matteo Renzi.

Conclusion

In the last years, we have witnessed a dramatic increase in the number of works that analyzes social media in order to assess the opinions of Internet users and to check whether the attitudes expressed online can be eventually used to nowcast and forecast the voting behavior of the whole population of voters. Therefore, being able to rely on techniques apt to measure online public opinion becomes a pressing topic.

In this article, in two (very) different political scenarios we have applied a statistical method recently introduced in the literature that performs a supervised sentiment analysis on social networks. Our analyses, all conducted before elections occurred, show a remarkable ability of social media to nowcast an electoral campaign as well to forecast electoral results. Moreover, Tables 3 and 6 illustrate that the MAE of the HK prediction is always lower when compared to measures based either on the volume of data (mentions) published on social networks or on other techniques of automated sentiment analysis.

Employing the HK method seems, therefore, a promising way to analyze social media with respect to electoral campaigns.[25] Still, a traditional puzzle arises here. To nowcast and forecast election, we need to rely on a representative sample, and there is no guarantee that this is something that can be obtained by analyzing social media.[26] On the contrary, socioeconomic traits of social media users do not exactly match the actual demographics of the whole population (Bakker & de Vreese, 2011; Tjong Kim Sang & Bos, 2012; Wei and Hindman, 2011): People on social media are generally younger (albeit the percentage of elderly people is rapidly increasing)[27] and more highly educated, concentrated in urban areas, as well as more politically active overall (Conover et al. 2011; Jensen, Jorba, & Anduiza, 2012). But do we need a representative sample when, for example, 22% of voters spontaneously declared their voting behavior on social network sites, as it happened during the U.S. Presidential campaign (Pew Research Center, 2012b)? Perhaps the sheer magnitude of data available on social media, that is, the “wisdom of crowds” (Franch, 2013), may compensate for this partly unrepresentative information. After all, the crowd to be wise needs to be diverse, independent, while its decisional procedure has to be decentralized (Surowiecki, 2004). And this is something that is usually attained in the Big Data world.

Moreover, to cast an accurate forecast, we should be more worried about the distribution of political preferences on the web. Previous (albeit quite dated) analyses showed that left-leaning is

Table 7. Distribution of Ideological Self-placement of Italian Voters Versus Subsample Active on Social Media.

Self-Ideological Placement	All Sample	Subsample of Social Media Users
Left	10.50	10.22
Centre-left	15.75	15.25
Centre	14.63	13.81
Centre-right	11.25	11.51
Right	4.13	4.32
None	37.50	38.42
Do not know/do not answer	6.25	6.47

Source. IPSOS, February 2012.

overrepresented, though only marginally (Best & Krueger, 2005). We clearly need more (updated) analysis in this regard. Accordingly, one way to improve the social media forecast would be to develop an appropriate set of weights based on the representativeness of certain groups of users (Choy, Cheong, Laik, & Shung, 2011, 2012), or, even better, according to the political preferences of social media users, provided this type of information is available (and reliable). Still, some of the potential bias that arise from social media analysis may be softened in the medium (short?) run, as the usage of social network increases (see Lenhart, Purcell, Smith, & Zickuhr, 2010, on this point). For example, Table 7 compares the distribution of the ideological self-placement of a sample of Italian voters in February 2012 and the subsample that declares to be active on social media. As can be seen, the difference among the two samples is quite trivial.

Finally, although the social media population, so far, is not still always representative of one country's citizenry, there are still some doubts about whether such bias could affect the *predictive skills* of social media analysis. Indeed, the latter aspect (the predictive skills of social media analysis) does not necessarily need the previous factor (i.e., the issue of representation) to hold true to effectively apply. This can happen, for example, if we assume that Internet users act like opinion makers who are able to influence (therefore often anticipating) the preferences of a wider audience (O'Connor, Balasubramanyan, Routledge, & Smith, 2010), including the ones of the broader media ecosystem (Farrell & Drezner, 2008).[28] The same applies if social media discussions are able to reproduce all the (more general) public opinions. For example, this can be true if Twitter communications are considered to function like a critically engaged interaction system (Ampofo, Anstead, & O'Loughlin, 2011) whose communications about specific issues (such as electoral contests) are thematically representative of larger currents of conversations and preference distributions (Jensen & Anstead, 2013). This is clearly another fascinating topic that deserves a further investigation.

Summing up, despite the well-known limits and the troubles faced by social media analysis (Gayo-Avello, Metaxas, & Mustafaraj, 2011; Goldstein & Rainey, 2010), our results provide reasons to be optimistic about the capability of sentiment analysis to become (if not to be already) a useful supplement of traditional off-line polls. But the method does matter: Shi, Agarwal, Agrawal, Garg, and Spoelstra (2012) noted that "merely using the volume of tweets [...] is not enough to capture public opinions. We need to come up with some sophisticated algorithm and model to make the prediction successfully." As we have argued in this article, aggregate supervised techniques, such as the one advanced by HK, seem to grant more accuracy than old-fashioned sentiment analysis, precisely in this last respect.

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.

Notes

1. For a more skeptical view on the role that social media can play in organizing revolts, with respect in particular to protests related to the 2009 Iranian elections, see Morozov (2009).
2. During the European Union (EU) elections held in 2009, the Pirate Party won 7.1% of votes in Sweden, gaining one seat in the EU parliament. In Germany, it received 2% of votes in the 2009 German Federal Election. It subsequently obtained positive results in German regional elections. In Italy, the *Movimento 5 Stelle* also reported surprising results during local elections held between 2009 and 2012, before obtaining a striking 25.1% in the 2013 general elections.
3. In addition, traditional surveys pose solicited questions and it is well known that this might inflate the share of strategic answers (Payne, 1951). Conversely, sentiment analysis does not make use of questionnaires and just focus on listening to the stream of unsolicited opinions freely expressed on Internet. In other words, it adopts a *bottom-up* approach, at least if compared with the *top-down* approach of off-line surveys.
4. Sentiment analysis consists in analyzing texts to extract information.
5. The results of both the American and the Italian analysis throughout the respective electoral campaign have been published on a daily basis on the home page of the newspaper Corriere della Sera.
6. The rate of penetration is defined as the number of monthly active Twitter users relative to the number of Internet users.
7. *Source:* Peerreach.com (http://blog.peerreach.com/2013/11/4-ways-how-twitter-can-keep-growing/)
8. *Source:* International Telecommunications Unions (http://www.itu.int/ITUD/ict/facts/2011/material/ICTFactsFigures2011.pdf).
9. http://expandedramblings.com/index.php/resource-how-many-people-use-the-top-social-media/
10. In other terms, a "word profile" is a vector made of 0's and 1's: 0 when a term does not appear in the unit (but it is used in some other units) and 1 when a term appears in the unit.
11. The fact that the training set, in this case, is manually codified or based entirely on an ontological dictionary does not make any difference: As long as the classification of texts is done on an individual base, we will produce severely biased aggregate estimates (see subsequently).
12. From our replications, the root mean square error of the estimates drops until 1.5% when the number of hand-coded documents increases up to 500.
13. In all the analyses, we also considered the categories "Others" and "Uncertain."
14. An hashtag is a word or a phrase prefixed with the symbol # that provides a means in Twitter of grouping all the messages including that word or phrase. Through that, one can search online for the hashtag and get the set of messages that contain it.
15. Data have been downloaded through the Twitter search application programming interface. The population of tweets collected consists of all the tweets posted during the temporal period considered (see subsequently) that include in their text at least one of a set of key words (the name of the political leaders/parties covered by each of our analysis in Italy and the United States, as well as the most popular hashtags characterizing each candidate/party's electoral campaign). Duplicated tweets have been removed.
16. Applying a shorter moving average (i.e., 3 days) does not change qualitatively any of our results.
17. The Benghazi gate concerns to the murder of the U.S. ambassador in Libya. According to the Republicans, Obama may have omitted the truth when talking about this event in public speeches. On the other side, Richard Mourdock, who was running for the Senate in Indiana for the Republican party, said that abortion should be neglected in case of rape, because even that birth is a "God's will."
18. See, for instance, the article by Andrew Gelman, director of the Applied Statistics Center at Columbia University: http://campaignstops.blogs.nytimes.com/2012/10/30/what-too-close-to-call-really-means/?smid=tw-share

19. The percentage of geotagged tweets is usually a portion of the overall number of tweets posted everyday. As such, the sample we drew upon was necessarily global in nature, while not being necessarily representative of those using Twitter.
20. Twindex, a fully automated index based on a dictionary-based method, was constructed for the two candidates as the proportion of the total number of positive tweets (measured by means of ontological dictionaries) versus the total number of positive *and* negative tweets. See http://usatoday30.usatoday.com/news/politics/twitter-election-meter
21. In their analysis, Washington, Parra, Thatcher, LePrevost, and Morar (2013) focus only on the proportion of positive sentiment messages.
22. Choy, Cheong, Laik, and Shung (2012) measured support by considering both positive and negative sentiment.
23. Note that on October 10, the campaign for the primary elections was not officially started yet. Therefore, our estimates also included some potential candidates who later decided to do not run for nomination.
24. Unfortunately, to our knowledge, no social media analysis besides ours has been conducted on the second round of the Italian Primary elections.
25. Note that with the same methodology discussed in this article, we conducted an analysis on both the French Presidential and Legislative elections 2012 once again producing notable good results in terms of both nowcasting and forecasting (Ceron, Curini, Iacus, and Porro, 2013).
26. This is a problem *also* for standard off-line surveys as the poll rates keep falling dramatically in recent years, thanks to mobile phones, caller identification and a rise in phone solicitation, while the difficulties of reaching many population segments still persist (Goidel, 2011; Hillygus, 2011; Pew Research Center, 2012a; Tourangeau & Plewes, 2013).
27. For example, the percentage of population aged 55–64 has increased on Twitter by 79% in the last year. See Global Web Index, "SOCIAL PLATFORMS GWI.8 UPDATE: Decline of Local Social Media Platforms." URL: https://www.globalwebindex.net/social-platforms-gwi-8-update-decline-of-local-social-media-platforms/.
28. The fact that quite often journalists are among the most active consumers of social media (Spieringse & Jacobs, 2013 Lasorsa, Lewis & Holton, 2012) could provide an empirical ground to this hypothetical claim. Indeed someone compares the talk of the Internet to the 18th-century salons: "the conversation of the salons reflected and shaped the culture of France and much of Western Europe and ignited the revolutions that would change our world forever. We do not take the salons lightly now; they are invaluable to historians. And we should treat the internet in precisely the same way" (Herbst, 2011, p. 95).

References

Albrecht, S., Lübcke, M., & Hartig-Perschke, R. (2007). Weblog campaigning in the German bundestag election 2005. *Social Science Computer Review*, *25*, 504–520.

American Association of Public Opinion Research. (2009). *An evaluation of the methodology of the 2008. Pre-election primary polls*. Lenexa, KS: Author.

Ampofo, L., Anstead, N., & O'Loughlin, B. (2011). Trust, confidence, and credibility. *Information, Communication & Society*, *14*, 850–871.

Bakker, T. P., & de Vreese, C. H. (2011) Good news for the future? Young people, internet use, and political participation. *Communication Research*, *20*, 1–20.

Banks, J. S. (1991). *Signaling games in political science*. New York: NY: Harwood Academic.

Barberá, P. (2012). Birds of the same feather tweet together. Bayesian ideal point estimation using twitter data. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2108098

Bennett, W. L., & Segerberg, A. (2011). Digital media and the personalization of collective action: Social technology and the organization of protests against the global economic crisis, *Information Communication and Society*, *14*, 770–799.

Best, S. J., & Krueger, B. S. (2005) Analyzing the representativeness of internet political participation. *Political Behavior*, *27*, 183–216.

Cameron, M. P., Barrett, P., & Stewardson, B. (2013). *Can social media predict election results? Evidence from New Zealand* (Working Paper in Economics 13/08). University of Waikato, New Zealand.

Ceron, A., Curini, L., Iacus, S.M, & Porro, G. (2013). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' policy preferences. An application to Italy and France. *New Media & Society*, doi: 10.1177/1461444813480466.

Choy, M., Cheong, M., Laik, M. N., & Shung, K. P. (2011). A sentiment analysis of Singapore presidential election 2011 using twitter data with census correction. Retrieved from http://arxiv.org/abs/1108.5520

Choy, M., Cheong, M., Laik, M. N., & Shung, K. P. (2012). US presidential election 2012 prediction using census corrected twitter model. Retrieved from http://arxiv.org/ftp/arxiv/papers/1211/1211.0938.pdf

Chung, J., & Mustafaraj, E. (2011) Can collective sentiment expressed on twitter predict political elections? Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, San Francisco, CA.

Conover, M., Goncalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011). Predicting the political alignment of twitter users. Proceedings of 3rd IEEE Conference on Social Computing SocialCom. Retrieved from http://cnets.indiana.edu/wp-content/uploads/conover_prediction_socialcom_pdfexpress_ok_version.pdf

Cottle, S. (2011). Media and the Arab uprisings of 2011: Research notes. *Journalism*, *12*, 647–659.

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

Franch, F. (2013). (Wisdom of the Crowds)^2: 2010 UK election prediction with social media. *Journal of Information Technology & Politics*, *10*, 57–71. doi:10.1080/19331681.2012.705080

Gayo-Avello, D. (2011). A warning against converting social media into the next literary digest. *CACM*.

Gayo-Avello, D. (2012). No, you cannot predict elections with twitter. *IEEE Internet Computing*, *16*, 91–94.

Gayo-Avello, D., Metaxas, P., & Mustafaraj, E. (2011). Limits of electoral predictions using social media data. Proceedings of the International AAAI Conference on Weblogs and Social Media, Barcelona, Spain.

Ghannam, J. (2011). *Social media in the Arab world: Leading up to the uprisings of 2011*. Washington, DC: Center for International Media Assistance.

Gloor, P. A., Krauss, J., Nann, S., Fischbach, K., & Schoder, D. (2009). *Web science 2.0: Identifying trends through semantic social network analysis* (pp. 215–222.). *CSE* 4. 2009 International Conference on Computational Science and Engineering, Vancouver, BC.

Goidel, K. (2011). Political polling in the digital age: The challenge of measuring and understanding public opinion. New Orleans, LA: LSU Press.

Goldstein, P., & Rainey, J. (2010). The 2010 elections: Twitter isn't a very reliable prediction tool. Retrieved from http://lat.ms/fSXqZW.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*, 1–31. doi:10.1093/pan/mps028

Herbst, S. (2011). Un(Numbered) Voices? Reconsidering the meaning of public opinion in a digital age, in goidel (2011). *Political Polling in the Digital Age* (pp. 85–98). Baton Rouge, LA: Louisiana State University Press.

Hillygus, D. S. (2011). The evolution of election polling in the United States. *Public Opinion Quarterly*, *75*, 962–981.

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

Huber, J., & Hauser, F. (2005). Systematic mispricing in experimental markets evidence from political stock markets. In 10th Annual Workshop on Economic Heterogeneous Interacting Agents, Essex, England.

Huberty, M. (2013). Multi-cycle forecasting of Congressional elections with social media, Workshop on Politics, Elections, and Data, CIKM 2013 Oct. 27–Nov. 1, 2013, ACM Press, San Francisco, CA, pp. 23–30.

Jansen, B. J., Zhang, M. M., Sobel, K., & Chowdury, A. (2009) Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, *60*, 2169–2188.

Jansen, H. J., & Koop, R. (2005). Pundits, ideologues, and ranters: The British Columbia election online. *Canadian Journal of Communication*, *30*, 613–632.

Jensen, M. J., & Anstead, N. (2013). Psephological investigations: Tweets, votes, and unknown unknowns in the republican nomination process. *Policy & Internet*, *5*, 161–182.

Jensen, M. J., Jorba, L., & Anduiza, E. (2012). Introduction. In E. Anduiza, M. Jensen, & L. Jorba (Eds.), *Digital media and political engagement worldwide: A comparative study* (pp. 1–15). New York, NY: Cambridge University Press.

Jugherr, 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 TO, Sander PG and Welpe IM ''Predicting elections with twitter: What 140 characters reveal about political sentiment''. *Social Science Computer Review. 30*, 229–234.

Lasorsa, D. L., Lewis, S. C., & Holton, A. E. (2012). Normalizing twitter: Journalism practice in an emerging communication space. *Journalism Studies*, *13*, 19–36.

Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K. (2010). *Social media and young adults*. Washington, DC: Pew Internet and American Life Project.

Lewis-Beck, M. S. (2005). Election forecasting: Principles and practice. *The British Journal of Politics and International Relations*, *7*, 145–164.

Lindsay, R. (2008). Predicting polls with Lexicon. Retrieved from http://languagewrong.tumblr.com/post/55722687/predicting-polls-with-lexicon

Madge, C., Meek, J., Wellens, J., & Hooley, T. (2009). Facebook, social integration and informal learning at university: It is more for socialising and talking to friends about work than for actually doing work. *Learning, Media and Technology*, *34*, 141–155.

Morozov, E. (2009). Iran: Downside to the 'Twitter revolution'. *Dissent*, *56*, 10–14.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. Proceedings of the International AAAI Conference on Weblogs and Social Media, Washington, DC.

Pamelee, J. H., & Bichard, S. L. (2011). *Politics and the twitter revolution: How tweets influence the relationship between political leaders and the public*. Lanham, MD: Lexington.

Payne, S. (1951). *The art of asking questions*. Princeton, NJ: Princeton University Press.

Pew Research Center. (2012a). Assessing the representativeness of public opinion surveys. Retrieved from http://www.people-press.org/files/legacy-pdf/Assessing%20the%20Representativeness%20of%20Public%20Opinion%20Surveys.pdf

Pew Research Center. (2012b). Social media and voting. Retrieved from http://www.pewinternet.org/~/media//Files/Reports/2012/PIP_TheSocialVote_PDF.pdf

Sanders, E., & den Bosch, A. V. (2013). *Relating political party mentions on Twitter with polls and election results* (Unpublished manuscript). Retrieved from ceur-ws.org/Vol-986/paper_9.pdf

Segerberg, A., & Bennett, W. L. (2011). Social media and the organization of collective action: Using twitter to explore the ecologies of two climate change protests, *Communication Review*, *14*, 197–215.

Shi, L., Agarwal, N., Agrawal, A., Garg, R., & Spoelstra, J. (2012). *Predicting US primary elections with twitter* (unpublished manuscript). Retrieved from http://snap.stanford.edu/social2012/papers/shi.pdf

Spierings, N., & Jacobs, K. (2013). Getting personal? The impact of social media on preferential voting. *Political Behavior*. doi:10.1007/s11109-013-9228-2

Surowiecki, J. (2004). *The wisdom of crowds*, New York, NY: Doubleday.

Tjong Kim Sang, E., & Bos, J. (2012). Predicting the 2011 Dutch senate election results with twitter. Proceedings of SASN 2012, the EACL 2012 Workshop on Semantic Analysis in Social Networks, Avignon, France.

Tourangeau, R., & Plewes, T. J. (2013). *Nonresponse in social science surveys: A research agenda*. Washington, DC: The National Academies Press.

Tumasjan, A., Sprenger, T.O., Philipp, G.S., & Welpe, I. M. (2011). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Social Science Computer Review*, *29*, 402–418.

Upton, G. Jr. (2010). Does attractiveness of candidates affect election outcomes? Retrieved from http://com/lib/files/AttractivePoliticians.pdf

Véronis, J. (2007). Citations dans la presse et résultats du premier tour de la présidentielle 2007. Retrieved from http://aixtal.blogspot.com/2007/04/2007-la-presse-fait-mieux-que-les.html.

Washington, A. L., Parra, F., Thatcher, J., LePrevost, K., & Morar, D. (2013). What is the correlation between twitter, polls and the popular vote in the 2012 presidential election? APSA 2013 Annual Meeting Paper. Retrieved from http://ssrn.com/abstract=2300363

Wei, L., & Hindman, D. B. (2011). Does the digital divide matter more? Comparing the effects of new media and old media use on the education-based knowledge gap. *Mass Communication and Society*, *14*, 216–235.

Williams, C., & Gulati, G. (2008). What is a social network worth? Facebook and vote share in the 2008 presidential primaries (pp. 1–17). Boston, MA: Annual Meeting of the American Political Science Association.

Woodly, D. (2007). New competencies in democratic communication? Blogs, agenda setting and political participation. *Public Choice*, *134*, 109–123.

Xin, J., Gallagher, A., & Cao, L. (2010). *The wisdom of social multimedia: Using flickr for prediction and forecast*. ACM Multimedia 2010 International Conference, New York, NY.

Author Biographies

Andrea Ceron is an assistant professor in political science at the Università degli Studi di Milano. His research interests include intraparty politics, coalition governments, public opinion, and social media analysis. He may be reached at andrea.ceron@unimi.it.

Luigi Curini is an associate professor in political science, Università degli Studi di Milano. His major interests are electoral competition, legislative behavior, automated content analysis, quantitative methods, and game theory. He may be reached at luigi.curini@unimi.it.

Stefano M. Iacus, is a statistician and probabilist, member of the R Core Team for the development of the R statistical environment, Università degli Studi di Milano. Interest of research span from inference for stochastic processes to causal inference in observational studies, mathematical finance, and computational statistics. He may be reached at stefano.iacus@unimi.it.