A survey on political event analysis in Twitter

Michalis Korakakis
Department of Informatics
Ionian University
Corfu, Greece
E-mail: michalis.korakakis@gmail.com

Evaggelos Spyrou
Institute of Informatics and Telecommunications
National Center for Scientific Research – "Demokritos"
Athens, Greece
E-mail: espyrou@iit.demokritos.gr

Phivos Mylonas
Department of Informatics
Ionian University
Corfu, Greece
E-mail: fmylonas@ionio.gr

***Abstract*—This short survey paper attempts to provide an overview of the most recent research works on the popular politics domain within the framework of the Twitter social network. Given both the political turmoil that arouse at the end of 2016 and early 2017, and the increasing popularity of social networks in general, and Twitter, in particular, we feel that this topic forms an attractive candidate for fellow data mining researchers that came into sight over the last few months. Herein, we start by presenting a brief overview of our motivation and continue with basic information on the Twitter platform, which constitutes two clearly identifiable components, namely as an online news source and as one of the most popular social networking sites. Focus is then given to research works dealing with sentiment analysis in political topics and opinion polls, whereas we continue by reviewing the Twittersphere from the computational social science point of view, by including behavior analysis, social interaction and social influence identification methods and by discerning and discriminating its useful types within the social network, thus envisioning possible further utilization scenarios for the collected information. A short discussion on the identified conclusions and a couple of future research directions concludes the survey.**

I. Motivation and Introduction

From interacting with friends and family, to performing online purchases and consuming news, social media play a predominate role in people's everyday lives [1]. Moreover, a plethora of recent important political events have highlighted the capacity in which social media, and Twitter, in particular, are able to influence and even guide political discourse. We can safely characterize the 2016 US presidential election as a defining moment for the utilization of Twitter within the political domain. During the aforementioned election period the active use of Twitter by the presidential candidates has led to a number of note-worthy moments,[1][2] acting essentially as a medium through which the candidates were able to communicate directly with their voter base, thus influencing significantly the outcome of the election.[3]

In this paper, we present a short overview of recent studies focusing on exploiting Twitter for a multitude of politically related tasks. Given the ever-increasing importance of Twitter within the political domain, our motivation is to identify the main trends in this particular research area and provide a concise analysis in order to facilitate a better understanding of the field for future studies.

[1] https://twitter.com/HillaryClinton/status/740973710593654784
[2] https://twitter.com/realDonaldTrump/status/741007091947556864
[3] https://www.theguardian.com/us-news/2017/mar/15/donald-trump-twitter-fox-news-interview-wiretapping

978-1-5386-0756-5/17/$31.00

The structure of the rest of this paper is as follows: in Section II we present the Twitter platform, detailing its history and characteristics. Section III deals with recent research efforts aiming to provide election forecasts through the utilization of Twitter, whereas Section IV details the application of sentiment analysis upon the aforementioned platform for the purpose of tracking public opinion. In Section V we present various studies aiming to take advantage of Twitter within the context of computational social science, and more specifically investigate its applications in the political domain. Final comments on the topic and relevant conclusions are drawn in Section VI.

II. The Twitter Platform

Twitter[4] is a vast social networking and microblogging online service, created in March 2006 and rapidly became one of the most popular websites. Twitter-registered users post short messages (limited to 140 characters) which are called "tweets," and are inspired by the Short Message Service (SMS) text-based communication between mobile phone users. Registered users are also able to accompany their tweet with a photo, a short video, a url or their "whereabouts," i.e., the precise or vague location where they are at a given moment. In addition, they may post their tweets publicly, within their profile or directly send them to other users as messages. Moreover, they are able to "follow" other users' updates. Unlike typical social networks, a user may follow any other user, no approval is needed. Following is not a bidirectional relationship, such as e.g., "friendship" in other networks. Registered users are also able to "like" or "retweet" (i.e., include) other users' tweets on their own feed.

Within her/his tweet, a user may use a "hashtag" (which is a word or a small phrase denoted by the # sign), in order to provide a simple annotation of its content. She/he is also capable of "mentioning" another user by her/his username (prefixed with the @ sign). Hashtags, people and frequent words are tracked by Twitter. This way, the most trending ones are harvested, tailored based on the users' location and on the accounts they follow and are presented to the users' feed. Users are also able to see trends of other locations. When a user selects such a trend, she/he is able to see all tweets that contain this specific phrase or hashtag, which when added to her/his tweets lets her/him participate within this trend.

Although Twitter is much simpler than the majority of the other social networking sites, it quickly became popular. The first ever tweet was created and posted in March 2006,

[4] http://twitter.com

by Jack Dorsey, Twitter's founder. In 2007, 5,000 tweets were posted every day [2]. 6 years later, in 2013, more than 500,000,000 tweets were posted every day, i.e., 5,000 tweets were posted in less than a second [3]. Apart from the aforementioned simplicity of use, one possible reason for this tremendous growth may be the fact that many celebrities, politicians, athletes, musicians, etc., newspapers, magazines or even companies and public or private organizations have their own "verified" accounts and many Twitter users choose to follow them so as to get updated on their news. Also, Twitter is accessible from smartphones and tablets, allowing users to continuously tweet or access their feed. However, the aforementioned growth has started to significantly decrease by 2014 [4]. According to the official Twitter statistics,[5] there are currently (as of June 30, 2016) about 313M active users, 82% of which are using its mobile app.

Twitter provides powerful APIs, in order to expose its data. More specifically, the Twitter Streaming API[6] provides real time access on all tweets as they are posted. However, normal developers are able to acquire only a small portion of it. On the other hand, the Twitter REST API[7] provides access to historical data, again with limitations to normal developers. The enterprise platform of Twitter, namely GNIP[8] does not have such strict limitations. Nevertheless, Twitter data have been used for several research purposes, such as sentiment analysis [5], stock market prediction [6], trend detection [7], information credibility [8], event detection [9], etc. Finally, Twitter is used in several real-life occasions, such as emergencies, politics and campaigning, news and sports reporting, public relations, education, business, etc.

III. Predicting the Elections

The unexpected and disruptive outcome of recent political events with far-reaching global implications (e.g., the Greek bailout referendum, Brexit and the 2016 US presidential election) has sparked a heated debate concerning the ability of polls to provide accurate election forecasts. Towards this end, a number of studies have tried to take advantage of the large-scale human behavioral data generated by social media such as Twitter, in order to utilize them as a means to provide unbiased and accurate election forecasts.

A recurring theme within this section is the application of sentiment analysis techniques in order to detect and correlate the so called *public mood* with the notion of *vote intention*. Among the pioneers in this field, Tumasjan et al. [10] demonstrated that it is possible to predict the outcome of an election by simply looking at tweet volume. Moreover, the authors stated that Twitter constitutes an ideal platform for monitoring public opinion. Burnap et al. [21] aimed to predict ahead of time the outcome of the 2015 UK general election. In order to conclude the winner of a particular electorate seat, the authors performed sentiment analysis upon a collection of tweets, while also considering results from the previous general election, in order to detect fluctuations in voting behavior. By relying on the admission that "more tweets equals more votes," Anuta et al. [17] performed sentiment analysis for the purpose of ascertaining whether social media can pave the way for less biased results than those generated by conventional polls. Their findings indicated that even though numerical bias is common in both approaches, relying on social media alone for forecasts might lead to less accurate predictions. By trying to monitor in real time the dissemination of public support for a candidate within Twitter, Kagan et al. [11] were able to predict in advance the outcome of the 2013 Pakistani and 2014 Indian elections, respectively. Finally, in contrast with the aforementioned studies who focused solely on the industrialized world, Prasetyo and Hauff [16] opted to provide election forecasts for developing countries, focusing particularly on Indonesia. They showed that their sentiment analysis model was able to outperform the majority of national polls.

In the same application domain, Cameron et al. [19] investigated whether a candidate's online presence can impact her/his chances of being elected. Through the application of two regression models, they concluded that a statistical significance, albeit small, exists between the amount of people following and/or befriending a politician in social media and the outcome of an election. Beauchamp [15], motivated by the lack of polls in non-swing US states, utilized textual information mined from geo-tagged tweets for the generation of forecasts. The proposed linear feature selection model was able to outperform various baselines along with actual polls, thus proving the suitability of social media for tracking public opinion and inferring voter intention. Eom et al. [13] contemplated whether the daily number of tweets about a political party can be indicative of its successes during an election. Their analysis concluded that in short-term this information can have a positive, although limited, predictive ability.

Bovet et al. [20] aimed to determine popular support for the 2016 US presidential candidates on Twitter. Using as a ground-truth an aggregate of traditional polls they were able to show that their supervised classifier was able to generate comparable results. Tsakalidis et al. [14] tried to predict the results of the 2014 EU election for three countries by exploiting a multi-feature time-series model upon a Twitter dataset. Their proposed methodology was able to outperform various baselines as well as conventional polls. Dokoohaki et al. [12] applied a link analysis approach upon a dataset from two different election periods in Sweden. Their results confirmed that it is possible to deduce the outcome of an election by focusing on user interaction within a social network. Sanders et al. [18] presented a prediction model that integrated demographics in order to predict two different election instances in Netherlands. To achieve this, they relied on an unsupervised approach that was able to inference a Twitter user's age and gender. Although the results of the study were promising, the authors stated that the process of automatically detecting the age of a Twitter user may obstruct any further improvements in the prediction model, due to its overall complexity.

Conversely, a number of studies have been critical of the accuracy and robustness of such prediction models. Among them, Metaxas et al. [22] were the first to try to replicate a multitude of forecast models as well as provide a set of guidelines for future reference. Their findings showcased that prediction models based on social media offer accuracy that is often comparable to that of performing random predictions. In a more recent survey, Huberty [23] stated that the main problems

[5] https://about.twitter.com/company
[6] https://dev.twitter.com/streaming/overview
[7] https://dev.twitter.com/rest/public
[8] https://gnip.com

TABLE I. ELECTION PREDICTION APPROACHES.

Work	Task(s)	Method(s)	Pros	Cons	Dataset(s)
[21]	predict 2015 UK election	sentiment analysis, prior elections' information	pioneer work	under-estimation of seats	13,899,073 tweets
[17]	compare social media with polls	sentiment analysis	innovative work	limited dataset	8 polls, 3M tweets, 750K unique users
[11]	predict 2013 Pakistani and 2014 India elections	AVA algorithm adaptation	beats traditional polling	short list of topics	31 topics, TIEN, IET-DB
[16]	predict the 2014 Indonesia presidential elections	sentiment analysis	focused on the developing world	the experiments were conducted after the elections	7,020,228 tweets, 20 conventional polls
[19]	forecasting the 2011 New Zealand general election	OLS regression models	proved statistically significant influence of social media	small impact	candidate's no. of Twitter followers & FB friends per day
[15]	predict state-level polls for the 2012 US presidential election	time-series & sentiment analysis	appropriate modeling	long-term viability of fitted model	120M tweets, 1200 polls
[10]	predict 2009 German federal elections	sentiment analysis	seminal paper	small dataset	104,003 tweets
[13]	provide forecasts for national and EU elections	stochastic differential equations geometric Brownian motion	indicated statistical fluctuations of Twitter activity	lack on combined validity checking	12,535,469 + 7,755,063 + 16,077 tweets for 3 election periods
[20]	determine popular candidate support	supervised hybrid classifier	hybrid (no-sentiment) tweet classification	not directly comparable results	6.7M users, 73M tweets
[14]	predict 2014 EU election	multivariate time-series forecasting task	outperforms baseline models	short time period considered	361,713+452,348+263,465 tweets 74,776+74,469+19,789 users
[12]	Swedish 2014 elections	link mining/analysis	innovative approach	extremely focused no textual analysis	7M tweets from Sweden
[18]	Netherlands 2012/2015 elections	prediction model, unsupervised approach	rather innovative approach	demographic complexity issues	TwiNL, kiesraad.nl
[22]	evaluate social media-based election predictions	forecast models testing, accuracy evaluation	provided guidelines	limited application	
[23]	evaluate election forecasting	survey paper	recent approach	no solution proposition	

within this particular domain is the lack of representative data along with the inefficiency of sentiment analysis methods, two rather important issues to be neglected by fellow researchers. Table I provides a detailed overview of the discussed research efforts by categorizing them accordingly and illustrates their advantages and disadvantages and reasons on their suitability within the broader research field.

IV. Sentiment Analysis in Political Topics and Opinion Polls

It is also true that discussions on political topics and/or opinion polls have been widely spread over the Internet. Consequently, their existence in the framework of Twitter came as a natural selection. Approaches dealing with sentiment analysis in politics focus on topics such as approval ratings and tracking public perception about legislative acts. Towards this end, Oliveira et al. [24] examined if social media can be utilized to adequately capture the public's support for a political candidate. By applying sentiment analysis upon a Twitter dataset concerning the 2014 Brazilian presidential election, they were able to achieve comparable results to that of traditional opinion polls. Similarly, Smailovi et al. [25] presented a SVM-based sentiment classifier for manually annotated tweets in order to detect in real time the public's attitude towards the political parties participating in the 2013 Bulgarian parliamentary elections. Among their observations, the authors stated that public mood remained negative throughout the elections and closely resembled the election results. Through the exploitation of the large-scale human behavioral data contained within Twitter, Cody et al. [31] examined if social media can replace traditional opinion polls. By inferring and subsequently quantifying the "happiness" contained within a tweet, they were able to monitor public opinion. The experimental evaluation of their proposed methodology against traditional opinion polls highlighted that social media are ideal for the aforementioned task. By exploiting a crowd-sourced annotated Twitter dataset, Mohammad et al. [32] presented a set of classifiers able to extract sentiment and purpose from a tweet. Among their findings, the authors stressed that sentiment alone cannot provide accurate detection for the latter. In order to uncover the main reasons behind the unexpected outcome of two critical Greek political events, Antonakaki et al. [26] performed a series of NLP-related tasks upon two Twitter datasets. Their findings highlighted the public's negative view of the government and its economic policies, accompanied by an eurosceptic attitude. By implementing three different search strategies for data collection in Twitter, Llewellyn and Cram [33] investigated whether public opinion differs between social groups. In contrast with their main hypothesis, the authors concluded that bias was apparent between the examined groups. Finally, Hürlimann et al. [34] created an annotated dataset containing information about the sentiment expressed in a tweet. Their main motivation for this was to give the opportunity to fellow researchers to study the Twittersphere during the Brexit referendum.

Besides using Twitter to track public mood, a number of studies have focused on discovering the underlying reasons that influence the political discussion in the Twittersphere. Using a linear regression model on a time-series mined from Twitter, Lansdall-Welfare et al. [27] aimed to deduce the root causes for public mood swings before and after the Brexit referendum. They concluded that they were three distinct occasions that influenced negatively the public, adding that for the next two days after the announcement of the results public mood aligned with fluctuations in the financial markets. In the same application domain, Giachanou et al. [28] aimed to discover the origins of sudden swings in public sentiment within Twitter. More specifically, through the application of the LDA algorithm and the Kullback-Leibler divergence, they were able to successfully detect and classify the most probable causes for sudden changes in public mood. Rill et al. [29] developed PoliTwi, a system capable of dynamically recognizing trending political topics within Twitter. Furthermore, their system was able to monitor public sentiment and detect the main causes that gave birth to a particular topic. Using as a ground-truth a Twitter dataset from the 2013 German parliament elections, the proposed system was able to outperform Google Trends. Bhattacharya et al. [30] aimed to retrieve tweets in order

TABLE II. SENTIMENT ANALYSIS IN POLITICAL TOPICS APPROACHES.

Work	Task(s)	Method(s)	Pros	Cons	Dataset(s)
[24]	detect voter preference	sentiment analysis	compared results to six voting surveys	dataset does not accurately reflect the voter population	37,817 + 32,501 + 27,839 + 34,526 + 54,037 + 92,441 tweets, election polls
[25]	real time tracking of public opinion	SVM-based sentiment classifier	novel work	costly manual annotation	29,433 + 10,300 tweets
[31]	assess public opinion polling in social media	calculate "hapinness" in a tweet	identified advantages of using tweets for public opinion polling	biased dataset	number of tweets not stated
[32]	extract sentiment & purpose from a tweet	SVM-based classifier	emotional state detector classifier	did not consider users' behavioral model	170,000 tweets
[26]	detect prevalent topics concerning political events	sentiment analysis & tweets volume analysis	sarcasm detection	no emotion utilization	301,001 tweets
[33]	assess differences in political discussion between groups	sentiment analysis	novel visualization	extremely focused domain	8,916,733 + 31,106 + 11,752 tweets
[34]	tweet sentiment-annotation	manual annotation	real-life annotation	small dataset	2000 tweets
[27]	detect swings in public mood	LARS algorithm	method can be adapted to other domains	increased complexity	over 10M tweets
[28]	identify sentiment spikes	LDA algorithm & KL-divergence	performance results	no comparative evaluation	1,076,732 + 1,265,001 + 1,369,756 tweets
[29]	detect emerging political topics	calculation of Topic Value	polarity calculation	no sentiment analysis	4M tweets
[30]	extract personality-traits for politicians	measuring a "trait" score	novel, complement to traditional sentiment analysis methods	mediocre recall values	81,200,065 + 41,860,086 tweets

to discover characteristics that are attributed to a politician by the public. Moreover, they examined whether temporal changes and/or major events can affect these traits. Among their observations, the authors indicated that even though people's opinions on the 2012 US presidential candidates largely differed, their proposed methodology was able to single-out distinct personality traits for each candidate. Finally, Table II presents the herein discussed approaches according to their type and illustrates each one's main features and characteristics.

V. TWITTER AND COMPUTATIONAL SOCIAL SCIENCE

A. Behavior Analysis

The first work that exploited Twitter in order to analyze human behavior, was probably the one of Larsson and Moe [35]. They collected tweets from the 2010 Swedish election and identified user and tweet types. They also showed that Twitter activity is closely linked to several campaign-related events. Howard and Kollanyi [36] investigated the use of bots during the UK-EU referendum. They found that the majority of relevant tweets originated from two bots which did not produce any content and instead retweeted tweets with hashtags relevant to their goals. Barbera and Rivera [37] studied the behavior of users during the 2011 Spanish legislative election and the 2012 US presidential election. They concluded that male users with clear ideological leaning generate the majority of content. Moreover, the geographical distribution of tweets closely follows the population distribution. Finally, Hosch-Dayican et al. [38] investigated the behaviour of Twitter users during the electoral campaign of the Dutch elections of 2012 and more specifically, whether and how they actively participate. By automatically analyzing a large set of collected tweets, they concluded that indeed citizens are significantly active, however they tend to use the negative campaigning approach, while politicians prefer the persuasive one. Moreover, citizens tend to use Twitter as a means to express their dissatisfaction.

B. Social Interaction

It is true that Twitter has significantly changed the way of interaction between politicians and citizens. However, several studies indicate that old, traditional parties have not yet taken advantage of its capabilities of social interaction, while this is not the case for the newer ones. Jürgens and Jungherr [39] aimed to assess whether Twitter has led into new patterns of political interaction during the 2009 German national election by statistically analyzing a corpus of collected tweets. They showed that traditional parties failed to find a successful strategy, thus their influence was limited. On the contrary, a new party and several individuals dominated the political discourse within Twitter. Ramos-Serano et al. [40] studied the extent to which Spanish political parties used Twitter to interact and with whom did they interact, during the European elections of 2014. Their findings showed that Twitter was used mainly as a means of unidirectional communication and broadcasting by traditional political parties, yet the younger ones focused more on interactivity. Similarly, Ahmed et al. [41] assessed that younger parties used Twitter for self promotion, while traditional ones to supplement their offline strategies. Additionally, they showed that the success of the winning party was associated with using Twitter for engaging first-time voters. Having similar goals, Graham et al. [42] compared the behaviour of Dutch vs. UK parliamentary candidates. By measuring tweets and investigating how they are distributed during a campaign, they showed that Dutch candidates were much more active than the British ones and also interacted more often. They also observed that the distribution of tweets followed a similar pattern. Larsson and Ihlen [43] witnessed a shift in the way politicians handle Twitter. Working with data from the 2013 Norwegian elections, they showed that politicians now reply with a higher rate to citizens than ever. Ianelli et al. [44] studied the connection between traditional and newer media, using political talk shows and Twitter to represent the aforementioned categories, respectively. They concluded that the audience that used simultaneously both media (i.e., a "hybrid" approach as referred to by the authors) engaged mainly in expressing their opinion and requesting interaction, although the latter was missing by the other side. Similarly, Trilling [45] analyzed how a TV debate during the German election campaign of 2013 was discussed on Twitter. He tried to extract topics that were emphasized by the candidates and by Twitter users and showed that comments not favourable by the users tend to create negative publicity, while favourable comments do not have any positive publicity effect.

C. Identifying Social Influence

Twitter has also been used to measure the social influence of politicians. Several studies have been conducted mainly measuring tweets, replies and retweets. Their goal is typically to assess whether a politician is the center of a network and if her/his tweets are influencing her/his followers. Amaral et al. [46] empirically studied the flows of communication of Portuguese and Spanish politicians aiming to assess whether and to what extent did these politicians become real opinion leaders during the corresponding campaigns of the European elections of 2014. Their statistical findings indicated that in neither country did politicians become really influential through Twitter. McGregor and Murão [47] investigated the effect of gender in politicians and their influence. Their research, conducted within the context of the US 2014 Senate elections and showed that female candidates had more interactivity with voters and also were more central to conversations than their male opponents. Grcar et al. [48] presented a measure of influence that was based on the Hirsch index [49] and tailored to the needs of Twitter. This measure is calculated based on a user's tweets and their retweets from other users. They showed that the proponents of Brexit were significantly more active, having also a higher influence. Xu et al. [50] examined the characteristics of opinion leaders (activists) and showed that centralized users, such as organizations have a greater influence, compared to simple Twitter users, using data from the 2012 Wisconsin recall election. Wang et al. [51] analyzed the tweets of the followers of H. Clinton and D. Trump, during the 2016 US presidential election. Among their findings, they showed that in terms of social influence, supporters of the former are less polarized than than those of the latter. More specifically, supporters of Trump tend to have either a lot or little influence. Finally, Shapiro and Hemphill [52] investigated the influence of the US congress' Twitter account to the content of the New York Times newspaper. They showed that Twitter may substitute traditional means of information sources, e.g., press releases. A summary of the aforementioned studies is provided in Table III, which classifies them accordingly.

VI. Conclusion

In this work, we presented several studies focusing on utilizing social media, and more specifically Twitter, for improving and/or gaining insight into a multitude of different tasks within the political domain. We organized the aforementioned studies into three major categories, namely providing election forecasts through social media, using sentiment analysis for monitoring the public opinion, and finally, using Twitter for politically charged computational social science tasks. Having in mind the impact of social media on the outcome of recent political events, our motivation was to identify the main trends within this particular application domain in order to facilitate a better understanding of the field for future studies. We believe and hope that based on the tabular organization and interpretation of each category we provided, future useful research directions may be identified by interested fellow researchers and that they may be able to use this survey work as a point of reference. A clear trend is to be identified and this may be summarized into the active utilization of social media dynamics in crucial everyday life activities. Among our future work is the extension of this study into other popular platforms, while in addition exploring the relationship between journalism, politics and social media.

References

[1] S. Greenwood, A. Perrin and M. Duggan, *Social Media Update 2016*, https://http://www.pewinternet.org/2016/11/11/social-media-update-2016/ (retrieved: 2017-03-18).

[2] K. Weil, *Measuring Tweets*, Official Twitter Blog, https://blog.twitter.com/2010/measuring-tweets (retrieved: 2017-03-18).

[3] R. Krikorian, *New Tweets per second record, and how!*, Official Twitter Blog, https://blog.twitter.com/2013/new-tweets-per-second-record-and-how (retrieved: 2017-03-18).

[4] B. Popper, *Twitter doesn't have explosive growth, and investors aren't happy*, The Verge, http://www.theverge.com/2014/10/27/7079277/twitter-q3-2014-earnings (retrieved: 2017-03-18).

[5] E. Kouloumpis, T. Wilson and J.D. Moore, *Twitter sentiment analysis: The good the bad and the omg!*, in Proc. of Int'l Conf. on Weblogs and Social Media, Barcelona Spain, 2011.

[6] J. Bollen, H. Mao and X. Zeng, *Twitter mood predicts the stock market*, Journal of computational science, vol. 2, no. 1, pp.1–8, 2011.

[7] M. Mathioudakis and N. Koudas, *Twittermonitor: trend detection over the twitter stream*, in Proc. of Int'l Conf. on Management of data, Indianapolis, Indiana, USA, 2010.

[8] C. Castillo, M. Mendoza and B. Poblete, *Information credibility on twitter*, in Proc. of Int'l Conf. on World Wide Web, Hyderabad, India, 2011.

[9] H. Becker, M. Naaman and L. Gravano, *Beyond Trending Topics: Real-World Event Identification on Twitter*, in Proc. of Int'l Conf. on Web and Social Media, Barcelona, Spain, 2011.

[10] A. Tumasjan, T. O. Sprenger, P. G. Sandner, I. M. Welpe, *Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment*, in Proc. of Int'l Conf. on Weblogs and Social Media, Washington, DC, USA, 2010.

[11] V. Kagan, A. Stevens, V. S. Subrahmanian, *Using Twitter Sentiment to Forecast the 2013 Pakistani Election and the 2014 Indian Election*, IEEE Intelligent Systems, vol. 30, no. 1, pp. 2–5, 2015.

[12] N. Dokoohaki, F. Zikou, D. Gillblad, M. Matskin, *Predicting Swedish Elections with Twitter: A Case for Stochastic Link Structure Analysis*, in Proc. of Int'l Conf. on Advances in Social Networks Analysis and Mining, Paris, France, 2015.

[13] Y-H. Eom, M. Puliga, J. Smailovi, I. Mozeti, G. Caldarelli, *Twitter-Based Analysis of the Dynamics of Collective Attention to Political Parties*, PLoS ONE, vol. 10, no. 7, 2015.

[14] A. Tsakalidis, S. Papadopoulos, A. I. Cristea, Y. Kompatsiaris, *Predicting elections for multiple countries using Twitter and polls*, IEEE Intelligent Systems, vol. 30, no. 2, pp. 10–17, 2015.

[15] N. Beauchamp, *Predicting and Interpolating State-Level Polls Using Twitter Textual Data*, American Journal of Political Science, 2016.

[16] N. D. Prasetyo, C. Hauff, *Twitter-based Election Prediction in the Developing World*, in Proc. of the ACM Conference on Hypertext & Social Media, New York, USA, 2015.

[17] D. Anuta, J. Churchin, J. Luo, *Election Bias: Comparing Polls and Twitter in the 2016 U.S. Election*, arXiv:1701.06232, 2017.

[18] E. Sanders, M. de Gier, A. van den Bosch, *Using Demographics in Predicting Election Results with Twitter*, in Proc. of Int'l Conf. on Social Informatics, Bellevue, WA, USA, 2016.

[19] M. P. Cameron, P. Barrett, B. Stewardson, *Can Social Media Predict Election Results? Evidence From New Zealand*, Journal of Political Marketing, vol. 15 , no. 4, 2016.

[20] A. Bovet, F. Morone, H. A. Makse, *Predicting election trends with Twitter: Hillary Clinton versus Donald Trump*, arXiv:1610.01587, 2016.

[21] P. Burnap, R. Gibson, L. Sloan, R. Southern, M. Williams, *140 characters to victory?: Using Twitter to predict the UK 2015 General Election*, Electoral Studies, vol. 41, pp. 230–233, 2016.

[22] P. T. Metaxas, E. Mustafaraj, D. Gayo-Avello, *How (Not) To Predict Elections*, in Proc. of PASSAT/SocialCom, Boston, MA, USA, 2011.

[23] M. Huberty, *Can we vote with our tweet? On the perennial difficulty of election forecasting with social media*, International Journal of Forecasting, vol. 31, no. 3, pp. 992-1007, 2015.

[24] D. J. S. Oliveira, P. H. de Souza Bermejo, P. Aparecida dos Santos, *Can social media reveal the preferences of voters? A comparison between*

TABLE III. COMPUTATIONAL SOCIAL SCIENCE APPROACHES.

Work	Task(s)	Method(s)	Pros	Cons	Dataset(s)
[35]	user type identification	descriptive statistical analysis	pioneer work	empirical approach	99,832 tweets
[36]	use of bots during elections	statistical analysis	novel work	introduces heuristics	1.5M tweets
[37]	patterns of behavior towards content generation	statistical analysis	biases identification	simple approach	2.8M + 62M tweets
[38]	citizens' behavior during electoral campaign	supervised learning approach with manual annotations	automatic	requires significant human effort	363,855 tweets
[39]	investigating patterns of interaction	statistical analysis	analytical study	empirical approach	>10M tweets
[40]	Twitter usage from political parties	statistical analysis	novel work	limited dataset	6316 tweets
[41]	Twitter usage from political parties	statistical analysis	comparative strategies	empirical approach	˜100K tweets
[42]	Twitter usage from political parties	statistical analysis	emerging models of election tweeting practices	empirical approach	26,282 + 28,045 tweets
[43]	Twitter usage from political parties	statistical analysis	provides insights on Twitter practices	limited dataset & empirical approach	846 tweets
[44]	connection between new and traditional media	statistical analysis	novel work	lack on comparative techniques wrt connected audiences	2,489,669 tweets
[45]	connection between new and traditional media	statistical analysis	novel work	requires significant human effort	120,557 tweets
[46]	studying of flows of communication of politicians	statistical analysis	innovative flow of communication study	empirical approach	1645 tweets, 427 re-tweets, 374 mentions
[47]	studying the effect of gender to the influence of politicians	statistical analysis	novel and systematic work	lack on evaluation	773,038 tweets
[48]	influence measure	statistical evaluation	novel work	limited evaluation; only one use case	>4.5M tweets
[50]	examination of opinion leaders	network analysis	combined network & content analysis	not applied on a real electoral campaign	3546 tweets
[51]	analysis of party supporters	multimodal analysis	novel work	empirical approach	98,291 followers
[52]	influence of Twitter to news	descriptive statistical analysis	novel work	many assumptions	275,816 tweets

sentiment analysis and traditional opinion polls, Journal of Information Technology & Politics, 2016.

[25] J. Smailovi, J. Kranjc, M. Grar, M. nidari, I. Mozeti, *Monitoring the Twitter sentiment during the Bulgarian elections*, Proc. of Int'l Conf. on Data Science & Advanced Analytics, Paris, France, 2015.

[26] D. Antonakaki, D. Spiliotopoulos, C. V. Samaras, S. Ioannidis, P. Fragopoulou, *Investigating the Complete Corpus of Referendum and Elections Tweets*, in Proc. of Int'l Conf. on Advances in Social Networks Analysis and Mining, San Francisco, CA, USA, 2016.

[27] T. Lansdall-Welfare, F. Dzogang, N. Cristianini, *Change-point Analysis of the Public Mood in UK Twitter during the Brexit Referendum*, in Proc. of Int'l Conf. on Data Mining Workshops, 2016.

[28] A. Giachanou, I. Mele, F. Crestani, *Explaining Sentiment Spikes in Twitter*, in Proc. International on Conference on Information and Knowledge Management, New York, NY, USA, 2016.

[29] S. Rill, D. Reinel, J. Scheidt, R. V. Zicari, *PoliTwi: Early detection of emerging political topics on twitter 4 and the impact on concept-level sentiment analysis*, Knowledge-Based Systems, vol. 69, pp. 24-33, 2014.

[30] S. Bhattacharya, C. Yang, P. Srinivasan, B. Boynton, *Perceptions of Presidential Candidates Personalities in Twitter*, Journal of the Association for Information Science and Technology, 2015.

[31] E. M. Cody, A. J. Reagan, P. Sheridan Dodds, C. M. Danforth, *Public Opinion Polling with Twitter*, arXiv:1608.02024, 2016.

[32] S. M. Mohammad, X. Zhu, S. Kiritchenko, J. Martin, *Sentiment, emotion, purpose, and style in electoral tweets*, Information Processing and Management, vol. 51, no. 4, pp. 480–499, 2015.

[33] C. Llewellyn, L. Cram, *Brexit? Analyzing Opinion on the UK-EU Referendum within Twitter*, in Proc. of Int'l Conf. on Web and Social Media, Cologne, Germany, 2016.

[34] M. Huerlimann, B. Davis, K. Cortis, A. Freitas, S. Handschuh, S. Fernndez, *A Twitter Sentiment Gold Standard for the Brexit Referendum*, Proc. of Int'l Conf. on Semantic Systems, Leipzig, Germany, 2016.

[35] A.O. Larsson and H. Moe, *Studying political microblogging: Twitter users in the 2010 Swedish election campaign*, New Media & Society, vol. 14, no. 5, pp. 729–747, 2011.

[36] P.N. Howard, B. Kollanyi, *Bots, #StrongerIn, & #Brexit: computational propaganda during the UK-EU Referendum*, arXiv:1606.06356, 2016.

[37] P. Barbera and G. Rivero, *Understanding the political representativeness of Twitter users*, Social Science Computer Review, vol. 33, no. 6, pp. 712–729, 2015.

[38] B. Hosch-Dayican, C. Amrit, K. Aarts and A. Dassen, *How do online citizens persuade fellow voters? Using Twitter during the 2012 Dutch parliamentary election campaign*, Social science computer review, 34(2), pp.135–152, 2016.

[39] P. Jürgens, and A. Jungherr, *The use of Twitter during the 2009 German national election*. German Politics, vol. 24, no. 4, pp. 469–490, 2015.

[40] M. Ramos-Serrano, J.D.F Gmez and A. Pineda, *Follow the closing of the campaign on streaming: The use of Twitter by Spanish political parties during the 2014 European elections*, New media & Society, 2016.

[41] S. Ahmed, K. Jaidka and J. Cho, *The 2014 Indian elections on Twitter: A comparison of campaign strategies of political parties*. Telematics and Informatics, vol. 33, no. 4, pp. 1071–1087, 2016.

[42] T. Graham, D. Jackson and M. Broersma, *New platform, old habits? Candidates use of Twitter during the 2010 British and Dutch general election campaigns*, New Media & Society, vol. 18, no. 5, pp. 765-783, 2014.

[43] A.O. Larsson and O. Ihlen, *Birds of a feather flock together? Party leaders on Twitter during the 2013 Norwegian elections*. European Journal of Communication, vol. 30, no. 6, pp. 666–681, 2015.

[44] L. Iannelli and F. Giglietto, *Hybrid spaces of politics: the 2013 general elections in Italy, between talk shows and Twitter*, Information, Communication & Society, vol. 18, no. 9, pp. 1006–1021, 2015.

[45] D. Trilling, *Two different debates? Investigating the relationship between a political debate on TV and simultaneous comments on Twitter*. Social Science Computer Review, vol. 33, no. 3, pp. 259–276, 2015.

[46] I. Amaral, R. Zamora, M.D.M. Grandio and J.M. Noguera, *Flows of communication and influentials in Twitter: A comparative approach between Portugal and Spain during 2014 European Elections*. Observatorio (OBS*), vol. 10, no. 2, pp. 111–128, 2016.

[47] S.C. McGregor and R.R. Mourao, *Talking Politics on Twitter: Gender, Elections, and Social Networks*, Social Media+ Society, vol. 2, no. 3, 2016.

[48] M. Grcar, D. Cherepnalkoski and I. Mozetic, *The Hirsch index for Twitter: Influential proponents and opponents of Brexit*. In Proc. of Intl. Workshop on Complex Networks and their Applications, Lyon, France, 2016.

[49] J.E. Hirsch, *An index to quantify an individual's scientific research output*, Proceedings of the National Academy of Sciences, USA, vol. 102, no. 46, pp. 16569-16572, 2005.

[50] W.W. Xu, Y. Sang, S. Blasiola and H.W. Park, *Predicting opinion leaders in Twitter activism networks: The case of the Wisconsin recall election*. American Behavioral Scientist, vol. 58, no. 10, pp. 1278–1293, 2014.

[51] Y. Wang, Y. Li and J. Luo, *Deciphering the 2016 US Presidential campaign in the Twitter sphere: A comparison of the Trumpists and Clintonists*, arXiv:1603.03097, 2016.

[52] M.A. Shapiro and L. Hemphill, *Politicians and the Policy Agenda: Does Use of Twitter by the US Congress Direct New York Times Content?*, Policy & Internet, 2016.