Title
Predicting elections from social media: a three-country, three-method comparative study

Journal
Asian Journal of Communication, 29(3)

ISSN
0129-2986

Authors
Jaidka, Kokil
Ahmed, Saifuddin
Skoric, Marko
et al.

Publication Date
2019-05-04

DOI
10.1080/01292986.2018.1453849

Peer reviewed

Predicting elections from social media: a three-country, three-method comparative study

Article *in* Asian Journal of Communication · March 2018

DOI: 10.1080/01292986.2018.1453849

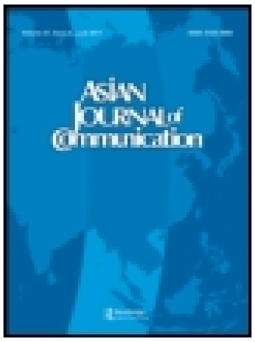
4 authors:

Kokil Jaidka
University of Pennsylvania
41 PUBLICATIONS 152 CITATIONS

Saifuddin Ahmed
University of California, Davis
19 PUBLICATIONS 142 CITATIONS

Marko M. Skoric
City University of Hong Kong
52 PUBLICATIONS 1,023 CITATIONS

Martin Hilbert
University of California, Davis
61 PUBLICATIONS 1,935 CITATIONS



Asian Journal of Communication

ISSN: 0129-2986 (Print) 1742-0911 (Online) Journal homepage: http://www.tandfonline.com/loi/rajc20

Predicting elections from social media: a three-country, three-method comparative study

Kokil Jaidka, Saifuddin Ahmed, Marko Skoric & Martin Hilbert

To cite this article: Kokil Jaidka, Saifuddin Ahmed, Marko Skoric & Martin Hilbert (2018): Predicting elections from social media: a three-country, three-method comparative study, Asian Journal of Communication, DOI: 10.1080/01292986.2018.1453849

To link to this article: https://doi.org/10.1080/01292986.2018.1453849

Published online: 24 Mar 2018.

ASIAN JOURNAL OF COMMUNICATION, 2018
https://doi.org/10.1080/01292986.2018.1453849

Routledge
Taylor & Francis Group

Predicting elections from social media: a three-country, three-method comparative study

Kokil Jaidka[a,b], Saifuddin Ahmed[c], Marko Skoric [d] and Martin Hilbert[c]

[a]Department of Computer and Information Science, University of Pennsylvania, PA, USA; [b]Adobe Research, Bangalore, India; [c]Department of Communication, University of California, Davis, CA, USA; [d]Department of Media & Communication, City University of Hong Kong, Hong Kong

ABSTRACT
This study introduces and evaluates the robustness of different volumetric, sentiment, and social network approaches to predict the elections in three Asian countries – Malaysia, India, and Pakistan from Twitter posts. We find that predictive power of social media performs well for India and Pakistan but is not effective for Malaysia. Overall, we find that it is useful to consider the recency of Twitter posts while using it to predict a real outcome, such as an election result. Sentiment information mined using machine learning models was the most accurate predictor of election outcomes. Social network information is stable despite sudden surges in political discussions, for e.g. around elections-related news events. Methods combining sentiment and volume information, or sentiment and social network information, are effective at predicting smaller vote shares, for e.g. vote shares in the case of independent candidates and regional parties. We conclude with a detailed discussion on the caveats of social media analysis for predicting real-world outcomes and recommendations for future work.

ARTICLE HISTORY
Received 1 August 2017
Revised 12 February 2018
Accepted 12 March 2018

KEYWORDS
Computational methods; election prediction; social media; sentiment analysis; India; Pakistan; Malaysia

Introduction

The widespread use of social media platforms for self-expression, communication, and social participation has resulted in an abundance of voluntarily disclosed personal information online, which can be aggregated to gauge public opinion unobtrusively. As compared to traditional methods of public opinion measurement, social media allows time- and cost-effective data collection and analysis with less human effort. Scholars analyzing social media data to gauge public opinion have supported the idea that the predictive validity of social media analysis does not necessarily rely on how representative the users are of the general population (Ceron, Curini, Iacus, & Porro, 2014).

One of the most frequently analyzed platforms is the microblogging site, Twitter. Twitter is a public forum for self-expression, communication and community participation worldwide, with approximately 330 million active monthly users and over 100 million daily active users. Several scholars have identified significant associations between Twitter activity, such as the frequency or sentiment of Twitter posts (tweets),

CONTACT Kokil Jaidka jaidka@sas.upenn.edu

and civic or electoral outcomes (Boutet, Kim, & Yoneki, 2012; Gayo-Avello, 2013; Livne, Simmons, Adar, & Adamic, 2011; Tumasjan, Sprenger, Sandner, & Welpe, 2010). However, others are skeptical about the rigor of these studies and reproducibility of the results (Gayo-Avello, 2013).

Prior research on inferring political opinion from social media has focused on economically developed, technologically advanced and politically stable democracies, which comprise a two-party or a multi-party system with low fragmentation – such as the United States (Livne et al., 2011), United Kingdom (Boutet et al., 2012), Germany (Tumasjan et al., 2010) and Ireland (Bermingham & Smeaton, 2011). From their findings, there is no clear agreement on which approach, whether volumetric, sentiment or social network analysis, would yield the most accurate predictions of election outcomes from social media, or why (Skoric, Liu, & Lampe, 2015).

There is also a need to examine whether these findings are replicable in the context of developing Asian democracies, where as few as one in ten citizens may have access to the Internet, and/or the political environment may be violent and highly fragmented. A few studies in this region have published encouraging results about the growing role of social media in voter mobilization and election campaigning (Brajawidagda & Chatfield, 2014; Song, Kim, & Jeong, 2014). In more sophisticated analyses, scholars have attempted to correlate social media attention with actual voting decisions and vote share, in Pakistan, Indonesia, Singapore, Malaysia, and India. In a study of the 2013 Pakistan General Elections, Ahmed and Skoric (2014) explored the differentiating Twitter campaigning characteristics of an emerging political party and its potential role in their subsequent electoral success. In the case of India, Khatua, Khatua, Ghosh, and Chaki (2015) analyzed geo-located tweets mentioning the general elections over a span of three months, and found that at the individual level, sentiment score was an effective predictor of vote swing. In the case of Indonesia, Satria, Kurnia, and Nurhadryani (2014) identified significant relationships between website utilization, social media attention, and voting decisions. Skoric, Poor, Achananuparp, Lim, and Jiang (2012) found that predicting election results from tweets in Singapore was possible, but did not replicate the accuracy of similar studies conducted in Western democracies. The authors suggested that contextual issues such as media freedom, competitiveness of the election and the structure of the party system could influence the success of estimating vote share from tweets.

The present comparative study compares the performance of three major approaches for predicting the vote share in the general elections of three Asian countries. It also assesses the utility of social media models for such predictions, by benchmarking them against previous election results and traditional opinion polls. It implements all three commonly used approaches found in previous work: volume-based models, sentiment analysis based on the lexicon and probabilistic models, and social network analysis. The dataset comprises 3.4 million tweets related to the general elections in Malaysia, India, and Pakistan. It also evaluates the impact of different pre-processing and temporal weighting steps on prediction accuracy.

We have chosen to study Malaysia, India, and Pakistan to study the role of social media in disparate political environments and technological set-ups. By comparing the same approaches in different contexts, we hope that our insights will be conclusive and generalizable. Malaysia (66.9% internet connectivity in 2013) is technologically advanced. The internet connectivity in India (15.1%) and Pakistan (10.9%) is poorer and localized to urban areas. Malaysian citizens, political parties, and leaders were familiar with the utilization of Web 2.0 and social media technologies during the 2013 general elections – but for India and Pakistan, the 2014 and the 2013 election was their first exposure to the use of social media technologies for electoral campaigning. In many cases, it was the first time that these parties were setting up official social media accounts, which they used to publish campaign updates and engage with the public (Ahmed & Skoric, 2014; Ahmed, Jaidka, & Cho, 2016).

Inferring political preferences from Twitter

Social media can be used as a real-time complement to traditional surveys, to monitor the day-by-day sentiments of voters towards electoral candidates and to identify trends in users' political preferences, which take time and effort to collect and collate from survey responses (Diaz, Gamon, Hofman, Kiciman, & Rothschild, 2014). Social media users voluntarily disclose their voting preferences when they discuss a political party or candidate. Pollsters can mine this information without introducing biases from survey questions.

Scholars have suggested that social media users may be less susceptible to social desirability biases than survey participants when they are discussing their political preferences (Payne, 2010). However, it should be noted that the opinions of the Twitter user base may not be representative of the entire population but some argue that it is still worthwhile to mine opinion from Twitter because the Twitter user base comprises 'a certain segment' of the public, whose opinion can confirm influence (or 'anticipate') the preference of a wider audience (Ceron et al., 2014, p. 345).

There is an open debate about whether Twitter can be used to infer political opinions and predict the results of the elections (Boutet et al., 2012; Bravo-Marquez, Gayo-Avello, Mendoza, & Poblete, 2012; Tumasjan et al., 2010). Among the studies seeking to validate this claim, one or more of the following approaches have been used to predict elections from Twitter:

- Volumetric analyses (*Vol*): frequency of mentions online (e.g. the frequency of mentions, retweets, supporters, likes etc.) measured by simple counts,
- Sentiment analyses (*Sen*): aggregate positive or negative sentiment in online posts, expressed emotions towards certain candidates or political parties, and
- Network analyses (*Net*): the characteristics of the network of social media users supporting or discussing certain candidates or political parties.

The scholars that criticize the utility of social media for predicting elections have pointed to a few drawbacks in the current body of literature, which this study hopes to address. The first drawback is that there is no clear agreement about which approach, whether volumetric, sentiment or social network would yield the most accurate predictions of election outcomes from social media. In their meta-analysis, Skoric et al. (2015) identified that volumetric approaches have been used in over half of all studies conducted in this area, but a combination of sentiment analysis with other approaches was usually more successful.

The second drawback is that model performance tends to fluctuate based on the data filtering, cleaning and processing methods followed, but the impact of individual methodological choices is not known (Jungherr, Jürgens, & Schoen, 2012). For instance, a predictive model based on specific counts of tweets mentioning only a single party or the actual numbers of authors involved (Gaurav, Srivastava, Kumar, & Miller, 2013; Lui, Metaxas, & Mustafaraj, 2011) outperforms predictive models based on raw counts of tweets mentioning political parties (Tumasjan et al., 2010). Positive sentiment models, or models combining sentiment with volumetric information, have been found to outperform negative sentiment models (Gayo-Avello, Metaxas, & Mustafaraj, 2011).

The third and final drawback is that few scholars have demonstrated the importance of incorporating temporal information in developing predictive models from social media (Bovet, Morone, & Makse, 2016; Khatua et al., 2015). Time could be an important factor in assessing the sustaining preference for a political party over a long period of time; but it may also weigh down the rising popularity of a new incident (MacDonald & Mao, 2015).

This study addresses these drawbacks by comparing the performance of volumetric, sentiment, and social network methods with and without temporal weights. It tests the replicability of the results in three Asian democracies, all of which have varying technological and political environments, variable internet penetration, and high fragmentation of political parties.

Method

We calculated the percentage share of volumetric (*Vol*), sentiment (*Sen*) and social network influence (*Net*) for Malaysia, India, and Pakistan – in the period preceding their general elections. First, we pre-process each party's tweets in one of four ways: (i) general mentions with no filtering, (ii) specific mentions or retaining only those tweets that mentioned a single party at a time, (iii) positive tweets or retaining only those specific mention tweets that had positive sentiment, and (iv) temporally weighing tweets that were closer to the election. We then compare the different pre-processing methods and approaches in terms of their ability to predict the actual vote share of each party. These methods are detailed in the following sections.

Data collection

Tweets were collected from Twitter's streaming API by using Tweet Archivist to track the mentions of political parties and their top two leaders. Approximately 3.4 million tweets were collected between the candidate nomination date and voting day, for 14 parties in Malaysia (1.1 m), 15 parties in India (1.2 m) and 11 parties in Pakistan (1.1 m).

Data cleaning

The following paragraphs detail the steps followed to filter out spam tweets, Twitter bots, and non-English tweets, to ensure the credibility of the datasets.

Filtering

There were several unrelated and spam tweets in the datasets; e.g. the abbreviation 'INC' (referring to an Indian political party) created ambiguity in the Indian dataset.

Accordingly, the dataset was filtered with the top election hashtag (#GE13, #GE2013, #GE14, #GE2014, #GE15 or #GE2015) to retain valid tweets.

Language detection

The Natural Language Toolkit package in Python was used to separate English tweets and discard non-English tweets before any analyses were conducted. English was the dominant language in the Twitter posts from two out of three countries, with the India and Pakistan datasets comprising over 90% English tweets. On the other hand, 77% of the Malaysia dataset was in Malay, and only 23% was in English. In this case, Malay tweets were excluded only from the sentiment analysis, because of the unavailability of a Malay sentiment lexicon.

Pre-processing

Before implementing the approaches, we pre-process our tweets in four different ways which are described in Table 1:

- Specific mentions only (when a tweet mentioned just one party) denoted by the subscript *m*,
- Authors of tweets (which may generally or specifically mention a party) denoted by the subscript *a*,
- Positively-labeled tweets denoted by the subscript *p* and
- Temporally-weighting of tweets in order of decreasing distance from the election, denoted by the subscript *t*.

Analytical approach

The following section describes the high-level volume, sentiment, and network approaches adopted. The actual steps followed to generate the different models are described in Table 2, which can be understood in conjunction with Table 1. In Table 2, the *Vol*, *Sen*, or *Net* prefixes denote the volumetric, sentiment or social network approaches respectively.

Volumetric analysis (Vol)

The volumetric analysis was aimed at measuring the volume of attention or support (i.e. the frequency of mentions, supporters, likes etc.), based on previous work (Gaurav et al., 2013; Lui et al., 2011). Volume-based measures are calculated as the proportional share of party mentions in a set of tweets for a given time:

Table 1. Pre-processing methods.

Model Suffix	Pre-processing method	Description
X_m	Specific mentions	A tweet mentioning a single party or its two most prominent political candidates
X_a	Authors of tweets	Counts of the unique users
X_p	Positively labeled tweets	Filtering to retain only the tweets with positive sentiment
X_t	Temporally-weighted tweets	Weighting each tweet according to its daily distance from the election, with the least distant tweets having the highest weights

Table 2. Description of prediction models.

Name	Method	More details
(VOL) Volumetric models		
Vol_{none}: General mention tweets	Volume with no pre-processing	Any mention of a party or its two most prominent political candidates (Tumasjan et al., 2010)
Vol_{m}: Specific Mention tweets	Volume with pre-processing to filter out non-specific mentions	Tweets which specifically mention only a single party or its most prominent political candidates (Bermingham & Smeaton, 2011;Response:-Resolved"> Burnap, Gibson, Sloan, Southern, & Williams, 2015; Lui et al., 2011; Sang & Bo,2012)
Vol_{a}: General Mention User $Vol_{m,a}$: Specific Mention User	Volume with pre-processing to count authors of general mentions. Volume with pre-processing to count authors of general mentions	This reflects the uniqueness of the social media polling. It is calculated as the unique numbers of authors identified from the sets of general and specific mentions for a party, respectively (Gaurav et al., 2013; Lui et al., 2011; Skoric et al., 2012)
Vol_{t}	Volume with pre-processing to weight daily tweets differently	Temporally weighted variants of the above models
(SEN) Sentiment models		
$Sen_{u,none}$: Net Sentiment	Net sentiment following an unsupervised approach and no pre-processing	Sum of total positive and negative valences for a party (Bermingham & Smeaton, 2011; González-Bailón et al., 2012)
$Sen_{u,p}$: Positive Sentiment	Pre-processing to retain only the positive sentiment tweets, labeled an unsupervised approach	This is calculated as the normalized share of all positive tweets mentioning a party (Bermingham & Smeaton, 2011; Ceron & D'Adda, 2013; Pimenta, Obradovic, & Dengel, 2013)
$Sen_{u,p,a}$: Positive Unique Users	Pre-processing sentiment tweets to count the authors of positive tweets	The number of unique authors posting positive tweets, for a party (Pimenta et al., 2013)
$Sen_{u,rch}$: Net sentiment reach	The reach of tweets (labeled for sentiment) based on their followers	The net sum of the positive and negative sentiments, weighted by the number of followers of the author
$Sen_{s,none}$, $Sen_{s,p}$, $Sen_{s,p,a}$, $Sen_{s,rch}$		Supervised variants of the above models
$Sen_{u,t}$, $Sen_{s,t}$: Temporally weighted models	Pre-processing tweets for all the models above, to weigh daily tweets differently	Temporally weighted variants of the above models
(NET) Social network models		
$Net_{dens,p}$: positive density	Graph density after pre-processing to retain only positively labeled tweets	It measures the overall connectivity of the party's positive sentiment social network. It is the ratio of the number of edges in a graph, to the total number of possible edges
$Net_{betw,p}$: positive betweenness	Betweenness after pre-processing to retain only positively labeled tweets	It measures the popularity of the party relevant to the overall positive sentiment social network, in terms of the number of shortest geodesic paths on which the party node lies $C_B(x) = \sum_{s,t \in V} \frac{\sigma(s,t\|x)}{\sigma(s,t)}$ where $\sigma(s,t)$ is the number of shortest geodesic paths and $\sigma(s,t\|x)$ is the number of all such paths which pass through the party node x (Brandes, 2008; Cuzzocrea et al., 2012; Newman, 2010)
$Net_{eig,p}$: positive eigenvector centrality	Eigenvector centrality after pre-processing to retain only positively labeled tweets	is calculated as a reciprocal process, in which the centrality of the party node x is proportional to the sum of the centralities of those users with which it is connected in the positive sentiment graph. In general, vertices with high eigenvector centralities are those which are connected to many other vertices which are, in turn, connected to many others (and so on). $C_e(x) = \frac{1}{\lambda} \sum_{t \in M(x)} x_t$ where $M(x)$ refers to the neighbors of the vertex x

(*Continued*)

Table 2. Continued.

Name	Method	More details
$Net_{pag,p}$: Positive Page Rank	Page Rank after pre-processing to retain only positively labeled tweets	It measures the number of times the party's node is encountered in a random walk over the positive sentiment social network. $C_p(x) = \alpha \sum_{j=1}^{n} A_{j,i} \frac{C_p(v_j)}{d_j^{out}} + \beta$ where A is the adjacency matrix of the graph and $d = \mathrm{diag}(d_1^{out}, d_2^{out}, d_3^{out}, d_4^{out} \ldots)$ is a diagonal matrix of degrees

$$Vol_x = \frac{C_x}{\sum_{j=1}^{n} C_j} \% \tag{1}$$

Here Vol_x represents the volumetric share of tweets for a party x, in a system of n parties, and C_x is the count of the tweets in the dataset which is relevant to party x. Table 2 describes how the eight volumetric models were calculated for each country by following different pre-processing methods: party-level general mentions (Vol_{none}), specific mentions (Vol_{m}), unique authors of general (Vol_{a}) and specific mentions ($Vol_{\mathrm{m,a}}$) and the temporal variants of these four models.

*Sentiment analysis (*Sen*)*

The sentiment analysis was aimed at measuring the positive, negative, and net sentiment impressions of each party on social media, based on simple counts of the number of tweets with the positive and negative sentiment.

This study implemented two approaches to estimate the positive and negative sentiments from Twitter. The unsupervised approach used a standard lexicon to look up words in the tweets, and then score the overall tweet for positive and negative sentiment. The supervised approach learned the features predicting sentiment from a set of hand-annotated tweets discussing politics, before automatically annotating a set of unseen tweets for sentiment, based on the words that comprise it (González-Bailón, Banchs, & Kaltenbrunner, 2012; Monti et al., 2013). In previous work, attention to a comparison of supervised and unsupervised methods on a data set of political tweets is rare (see González-Bailón & Paltoglou, 2015) and our study would facilitate an understanding of the benefits and limitations of either approach in different contexts.

First, in the unsupervised approach, the SentiStrength sentiment lexicon was applied to look up each word in each tweet, and then score the overall tweet for positive and negative sentiment. SentiStrength (Thelwall, Buckley, Paltoglou, Cai, & Kappas, 2010) draws from three lexica – Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007), General Inquirer (Stone, Bales, Namenwirth, & Ogilvie, 1962) and ANEW (Bradley & Lang, 1999); it also contains modules to account for negation, idioms, emoticons and other features inherent in social media (González-Bailón et al., 2012). It returns a positive and a negative score, on a scale of 0–5, for each input tweet, based on the positive, negative or slang words and emoticons that were found to match with its fixed vocabulary.

From these scores, the following model was used to calculate the net sentiment score senti_t for each tweet t as:

$$Sen_t = \{\ 1,\ pos_t > |neg_t| \quad -1,\ pos_t < |neg_t| 0,\ pos_t = |neg_t|\ \} \tag{2}$$

where pos_t and neg_t are the positive/negate Sentistrength scores for the tweet t.

The supervised approach followed was to apply a sentiment classifier to automatically classify each tweet as positive or negative. The Naïve Bayes classifier used in this paper was trained on a dataset of hand-annotated tweets about the Ireland's ection in 2011 (Bermingham & Smeaton, 2011) and a corpus of sentences from political news articles from the New York Times (Sanders, 2011), to identify the top 6000 words and features predictive of positive and negative sentiment. When a labeled corpus is provided to a classifier, it first decomposes labeled tweets into single words. The word profile used in text units (such as Twitter posts) are indicated by *S*, and the opinions expressed by people posting the tests are indicated by *D*. *P(S)* or the probability of a positive or negative label for a word, is obtained by tabulating all the tweets to identify their word profiles, and can be expressed as:

$$P(S) = P(S|D)*\ \ P(D) \tag{3}$$

The trained classifier can then predict tweet-level sentiment for any input tweet as a positive or negative class, which is mapped to binary numeric scores, as:

$$f{:}[\,pos_t,\ \ neg_t] \rightarrow [1,\ -1] \tag{4}$$

The unsupervised model was evaluated on a held-out sample of 669 instances and resulted in an overall accuracy of 89%. The precision and recall for identifying positive sentiments were 79% and 93% and the same for negative sentiments was 96% and 87%. After the tweets are labeled, Table 2 describes how the sentiment models were calculated: net sentiment ($Sen_{\text{u,none}}$ and $Sen_{\text{s,none}}$), and positive tweets ($Sen_{\text{u,p}}$ and $Sen_{\text{s,p}}$), unique authors of positive tweets ($Sen_{\text{u,p,a}}$ and $Sen_{\text{s,p,a}}$) and net sentiment reach ($Sen_{\text{u,reach}}$ and $Sen_{\text{s,reach}}$) and the temporal variants of these models.

*Social network analysis (*Net*)*

Social network analysis (*Net*) measures the strength of the online community supporting each political party. SNA can identify the central position played by a party in the overall online community by measuring its centrality – the way the party is connected to others in the community, as a function of its incoming, outgoing or bidirectional links. Several studies have found a connection between the centrality of political candidates on social networks and their electoral standing (Brandes, 2008; Cuzzocrea, Papadimitriou, Katsaros, & Manolopoulos, 2012; Newman, 2010). Weingert and Sebastian (2015) followed a supervised approach trained on Page Rank and network centrality of donor relationships to predict the outcome of primary elections in the US.

Each party's tweets were first modeled as incoming edges from various authors, to a single node depicting the political party. Other edges constituted @-mentions of Twitter users. Hence, edges among users existed if one author had mentioned another Twitter user in their tweet.

Each party's graph was measured in terms of its graph density. The betweenness and closeness centrality for each party node relative to its graph (Freeman, 1978) was calculated by using the *igraph* package in the R programing language (Csardi & Nepusz, 2006). The relative centrality scores were subsequently expressed as percentage proportions, by calculating ratios of raw score to the sum of the raw scores for all the parties:

$$Net_x = \frac{s_x}{\sum_{j=1}^{n} s_j}\% \tag{5}$$

Here, Net_x represents the centrality score for a party x, in a system of n political parties, calculated by using the raw centrality scores s_i for each of the n parties. In the case of graph density, Equation 5 was applied to the inverse of the raw graph density scores for each party. A higher 'density' on this metric reflected a decentralized social network graph, with many active users.

It is hypothesized that a denser graph would reflect a localized influence on Twitter, while a larger, more diffuse network would enable the political party would signify a wider outreach to a larger potential voter base. Betweenness centrality considers the importance of the party in terms of how many shortest information pathways it lies on, which measures the bottleneck influence on other users. Eigenvector centrality reflects how well-connected a party is to other influential users. In this model, influential users would be those who frequently mention a political party and frequently mention and interact with other users. A node with a high betweenness score (indicating it connects disparate parts of a network) could have a low eigenvector centrality score if it is still some distance from the centers of connectivity in the network. Finally, similar to eigenvector centrality, PageRank identifies influential nodes by taking the influence of adjacent nodes into account and is useful for visualizing network activity. Table 2 provides the mathematical intuition behind the network models (*Net*), which were based on the edge graph of the positively-labeled tweets identified via the supervised approach. They comprise positive graph density ($Net_{\mathrm{dens,p}}$), positive betweenness ($Net_{\mathrm{betw,p}}$), positive eigenvector centrality ($Net_{\mathrm{eig,p}}$) and positive PageRank centrality ($Net_{\mathrm{pag,p}}$).

Although several other models were calculated and compared, some of them resulted in high MAEs – for instance, sentiment models of negative sentiment share, and social network models of all mentions rather than positive tweets and therefore have not been reported in this study.

Performance evaluation

Model performances are compared in terms of their Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and the normalized Kendall's tau rank distance (Kendall, 1938).

MAE

This provides the average error in terms of the difference between a set of predicted values (tweets) and actual values (votes):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |e_i| \tag{6}$$

where n is the number of forecasts and e_i is the difference in actual result and predicted the result for the ith forecast. MAE focuses on the errors centered on the mean and overlooks infrequent big errors.

RMSE

RMSE regulates model performance, keeping in mind the larger errors. Unlike MAE which gives the same weight to all errors, RMSE penalizes variance as it assigns errors with large absolute values more weight than errors with smaller absolute values (Chai & Draxler, 2014). Such aberrations are especially important when studying political uses of Twitter, because studies have shown that minor political parties are typically more active on social media platforms as compared to majority parties (Ahmed & Skoric, 2014), and some parties are disproportionately discussed as compared to their seat or actual vote share (Khatua et al., 2015).

Normalized Kendall's Tau distance

Kendall's tau is a distance measure which then calculates the pairwise disagreement between the actual ranks and the predicted ranks (Kendall, 1938). In this paper, it is used to compare the different predictive models, in terms of their accuracy in ranking political parties according to their predicted success. It is calculated as:

$$\tilde{K}(\tau_1, \tau_2) = \frac{|\{(i, j): i < j, (\tau_1(i)\langle\tau_1(j) \wedge \tau_2(i)\rangle\tau_2(j)) \vee (\tau_1(i) > \tau_1(j) \wedge \tau_2(i) < \tau_2(j))\}|}{n(n-1)/2}$$

where $\tau_1(i)$. and $\tau_2(i)$. are the ranked positions of party i among n parties, in the actual rankings τ_1 and the predicted rankings τ_2 respectively. A larger normalized Kendall's tau reflects that the predicted ranking is far from the actual ranking of the parties.

Results

The following sections first present the ovl performance of the predictive models for the three countries, in terms of the MAEs and RMSEs. Next, the results drill into to the party-wise performance analysis for each country. We end with a comparison of the best Twitter models against previous election results and traditional opinion polls.

Overall model performance

Tables 3–5 provide the country-wise evaluation of the volumetric, sentiment and social network models respectively, in terms of their MAE, RMSE and the difference between the two when compared against the actual election results. Additionally, a column corresponding to each country provides the performance of the temporal variants of the same models (Vol_t).

In the case of volumetric models, all the volumetric models showed the same performance for Malaysia, but the inter-model differences are more marked for India and Pakistan. Among the pre-processing steps, filtering tweets to retain only the tweets which mention a single party (x_{m}, specific mentions) and counting unique authors instead of all the tweets (x_{a}, authors) helped to reduce the prediction errors for India and Pakistan.

Table 3. Volumetric analyses MAEs and RMSEs summary.

		MY	IND	IND VOL$_t$	PAK	PAK VOL$_t$
General mention tweets (Vol_{none})	MAE	1.63	3.58	1.87	6.38	6.37
	RMSE	2.29	5.64	2.84	10.34	8.31
	RMSE − MAE	(.66)	(2.06)	(0.97)	(3.96)	(1.94)
Specific mention tweets (Vol_m)	MAE	1.64	3.09	1.62	5.88	7.18
	RMSE	2.27	4.45	2.62	9.14	9.12
	RMSE − MAE	(.63)	(1.36)	(1.00)	(3.26)	(1.94)
General mention user (Vol_a)	MAE	1.83	2.77	1.35	7.85	5.55
	RMSE	2.86	4.05	2.03	10.93	7.82
	RMSE − MAE	(1.03)	(1.28)	(0.68)	(3.07)	(2.27)
Specific mention user ($Vol_{m,a}$)	MAE	1.69	2.23	1.69	5.94	7.25
	RMSE	2.38	3.38	2.39	8.53	10.08
	RMSE − MAE	(0.70)	(1.15)	(0.70)	(2.58)	(2.83)

The results after temporal weighting for Malaysia have been omitted as they offered no real improvements over the original models. The best predictions (or the lowest MAE) for India and Pakistan are for the temporally weighted variant of general mentions (Vol_a). For India, the temporal weighting of tweets (Vol_t models) leads to a marked improvement in prediction for all four models. All the models had large errors for Pakistan. The possible reasons for these high errors are investigated through a party-level evaluation in the next section.

In Table 4, all the sentiment models performed about the same for Malaysia. Supervised, temporally weighted sentiment models best predicted the actual vote shares for India and Pakistan, as compared to the volumetric models. Overall, supervised sentiment approaches appear to work better than unsupervised approaches, especially for inferring sentiment for a specialized corpus such as tweets discussing an election. Once again,

Table 4. Sentiment analyses MAEs and RMSEs summary.

		Unsupervised models				
		MY	IND	SENu,t	PAK	SENu,t
Net sentiment ($Sen_{u,none}$)	MAE	2.59	3.57	2.44	3.82	6.14
	RMSE	3.66	5.16	3.67	5.55	8.69
	RMSE − MAE	(1.07)	(1.59)	(1.23)	(1.73)	(2.55)
Positive sentiment ($Sen_{u,p}$)	MAE	2.12	4.37	2.22	6.68	5.64
	RMSE	2.82	7.48	3.39	9.53	8.11
	Diff	(0.70)	(3.11)	(1.17)	(2.85)	(2.47)
Positive unique users ($Sen_{u,p,a}$)	MAE	2.01	3.94	2.22	6.71	5.32
	RMSE	2.82	6.07	3.47	11.06	8.20
	Diff	(0.81)	(2.13)	(1.25)	(4.34)	(2.88)
Net sentiment reach ($Sen_{u,rch}$)	MAE	2.58	3.88	1.84	4.29	5.42
	RMSE	3.52	6.66	3.2	6.11	7.17
	Diff	(0.94)	(2.78)	(1.36)	(1.81)	(1.75)
		Supervised models				
		MY	IND	IND SEN$_{s,t}$	PAK	PAK SEN$_{s,t}$
Net sentiment ($Sen_{s,none}$)	MAE	2.12	3.00	1.71	3.44	5.10
	RMSE	3.31	3.89	2.49	4.05	6.68
	RMSE − MAE	(1.19)	.89	(0.78)	(.61)	(1.58)
Positive sentiment ($Sen_{s,p}$)	MAE	2.01	3.49	1.97	5.39	6.23
	RMSE	2.93	5.48	2.96	7.94	8.38
	RMSE − MAE	(0.92)	1.99	(0.99)	(2.56)	(2.15)
Positive unique users ($Sen_{s,p,a}$)	MAE	2.10	3.80	1.82	6.58	5.47
	RMSE	3.05	6.08	3.11	9.00	7.95
	RMSE − MAE	(.95)	2.28	(1.29)	(2.42)	(2.48)
Net sentiment reach ($Sen_{s,rch}$)	MAE	2.14	3.66	1.98	6.69	6.03
	RMSE	2.95	5.24	2.91	10.26	8.47
	RMSE − MAE	(.81)	1.58	(0.93)	(3.58)	(2.44)

Table 5. Social network analysis MAEs and RMSEs summary.

		MY	IND	PAK
Positive density ($Net_{dens,p}$)	MAE	4.57	1.51	5.84
	RMSE	5.91	2.45	8.97
	RMSE − MAE	1.34	0.94	3.13
Positive betweenness ($Net_{betw,p}$)	MAE	4.59	1.93	5.68
	RMSE	6.15	3.01	8.99
	RMSE − MAE	1.56	1.08	3.31
Positive eigenvector ($Net_{eig,p}$)	MAE	5.48	1.62	5.39
	RMSE	6.37	2.14	7.17
	RMSE − MAE	0.89	0.52	1.78
Positive page Rank ($Net_{pag,p}$)	MAE	2.96	1.56	5.46
	RMSE	3.64	2.12	8.48
	RMSE − MAE	0.68	0.56	3.02

temporal weighting improves the performance of all the sentiment models in the case of India, and only offers improvements about half the time in the case of Pakistan.

Table 5 shows that different social network models work best across the three countries. In the case of Malaysia, the best predictor was Positive Page Rank ($Net_{pag,p}$), implying that parties with higher vote shares had influential authors talking about them. In the case of India, the density of the party graph was the best predictor (less density = higher vote share). In the case of Pakistan, positive eigenvector centrality ($Net_{eig,p}$) is the best predictor for Pakistan, but it does not outperform the net sentiment share model from Table 4. This implies that for Pakistan, it is important to consider the net sentiment of people who are talking about a party as well as their social influence.

A cursory examination of the MAEs would suggest that social network models did not perform as well as expected. However, when we calculate Kendall's Tau distances between estimated and actual party rankings, we obtain a detailed perspective. Table 6 shows that volumetric models were indeed the best-performing models in the case of Malaysia. In the case of India and Pakistan, social network models provide the closest approximation to actual vote share rankings of the political parties. Eigenvector centrality demonstrated the most accurate rank estimation of parties among all three countries. Volumetric and sentiment models were about the same at predicting party ranks in India, but the variance in performance was large in the case of Pakistan.

Table 6. Normalized $\tilde{K}$ distances for party-wise performance analysis.

Prediction approach	Model	MY	IND	PAK
Volumetric	General mentions (Vol_{none})	0.09	0.12	0.19
	Specific mentions (Vol_{m})	0.09	0.07	0.19
	General mention Author (Vol_{a})	0.10	0.09	0.25
	Specific mention Author ($Vol_{m,a}$)	0.10	0.08	0.25
Unsupervised sentiment	Net sentiment ($Sen_{u,none}$)	0.16	0.12	0.22
	Positive sentiment ($Sen_{u,p}$)	0.10	0.12	0.19
	Positive unique user ($Sen_{u,p,a}$)	0.07	0.09	0.22
	Sentiment reach ($Sen_{u,rch}$)	0.10	0.10	0.16
Supervised sentiment	Net sentiment ($Sen_{s,none}$)	0.07	0.12	0.19
	Positive sentiment ($Sen_{s,p}$)	0.12	0.12	0.19
	Positive unique user ($Sen_{s,p,a}$)	0.10	0.09	0.22
	Sentiment reach ($Sen_{s,rch}$)	0.09	0.10	0.22
Social network analysis	Density ($Net_{dens,p}$)	0.18	0.09	0.16
	Betweenness ($Net_{betw,p}$)	0.10	0.09	0.19
	Eigenvector centrality ($Net_{eig,p}$)	0.10	0.10	0.16
	Page rank ($Net_{pag,p}$)	0.13	0.10	0.19

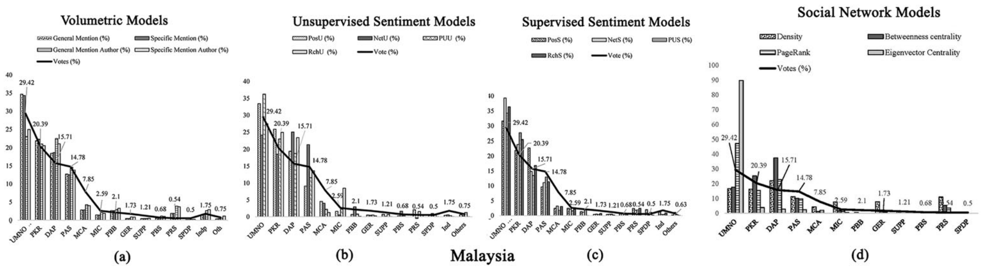


Figure 1. Party-level performance evaluation for Malaysia.

Party-level performance evaluation

The following paragraphs provide a fine-grained analysis of the predicted vote shares for individual parties against the actual outcomes, for all the countries and models. These results are depicted in Figures 1(a–d), 2(a–d), and 3(a–d).

Malaysia

Figure 1(a–d) depicts the party-wise predictions against the actual vote share trendline for the political parties under focus. The volumetric representation of political parties on Twitter correspond closely with their actual order and share in votes awarded. The leading party, UMNO, is marginally over-represented and the minority party, MCA, is marginally under-represented. In Figure 1(a), these errors were best resolved by authors of general and specific mentions (Vol_a, $Vol_{m,a}$), which suggests that a few users might have been tweeting disproportionately about UMNO in the dataset. The spikes in the predicted vote share for DAP and PRS reflect that there were many unique authors, posting only a few tweets about them. The sentiment models in Figure 1(b,c) shows no real improvement over the volumetric models. In Figure 1(d), graph density ($Net_{dens,p}$) and betweenness centrality ($Net_{betw,p}$), to a certain extent, were able to dampen the over-representation of UMNO and estimate the percent share of MCA.

India

Figure 2(a–d) depicts the party-wise predictions against the actual vote share, for the fourteen parties/coalitions which contested India's election. Figure 2(a) shows that the models exceptionally overestimated the vote shares of the election winners, NDA (14% to 20% error), and the political debutant, AAP (7% to 10% error). This is likely because NDA and AAP were the most active on Twitter before the election – as such, they had a greater Twitter audience than the other parties (Ahmed et al., 2016). Volumetric models underestimated the outcome for three regional political parties – BSP, the Left coalition and the YSCRP, which went on to win a large vote share that was disproportionate to their Twitter following. Figure 2(c) shows that supervised sentiment models performed better than the other volumetric and unsupervised models (Figure 2(b)) in assessing vote shares. The net sentiment (Sen_{none}) and sentiment reach ($Sen_{s,rch}$) models were the best predictors of the actual results. In Figure 2(d), social network models dampen the over-estimation of NDA's and AAP's vote shares only up to an extent.

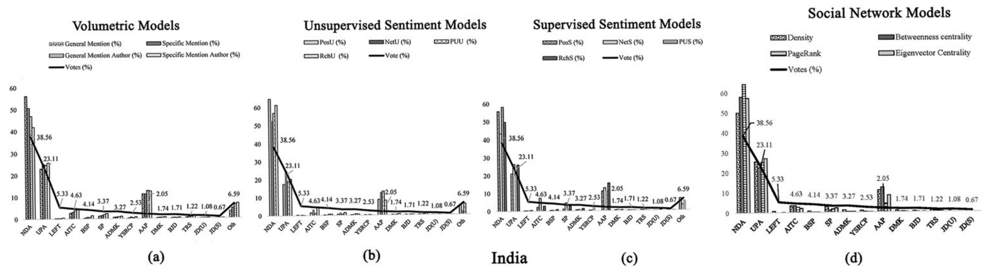


Figure 2. Party-level performance evaluation for India.

Pakistan

Figure 3(a–d) depicts the party-wise predictions against the actual vote share, for the parties which contested Pakistan's election. The election winner, PML (N) was under-represented on Twitter (−9% to −14%) and the fringe party, PTI, was largely over-represented (50% to 8% error). In fact, PTI is the factor which throws off all the social media models discussed in the previous section. Almost two-thirds of all the collected tweets mentioned PTI, so its popularity on Twitter was disproportionate to its actual election performance. Its inordinate Twitter presence was related to the way in which it heavily used Twitter to connect and mobilize potential voters (Ahmed & Skoric, 2014, 2015). The MAEs all of the models discussed in the previous section drops to around 3.5% when PTI is excluded from the evaluation.

Figure 3(b,c) shows that in the sentiment models, the net sentiment ($\mathrm{Sen}_{u,none}$ and $Sen_{s,none}$) and sentiment reach models ($Sen_{u,rch,}$ and $Sen_{s,rch}$) were the best predictors of the actual results. The errors were dampened for both PML(N) (down to 2% error for unsupervised models and 9% for supervised models) and PTI (down to 11% error for unsupervised models and 30% for supervised models).

Finally, the social network models in Figure 3(d) shows that PageRank centrality ($Net_{pag,p}$) was the closest to realizing PML(N)'s actual vote share and damping the overestimation of PTI's and PPP's vote share.

Comparison against traditional polls and previous election results

Finally, we compare social media predictions against the previous election results and opinion polls for each of the countries. The results are provided in Table 7. We see that

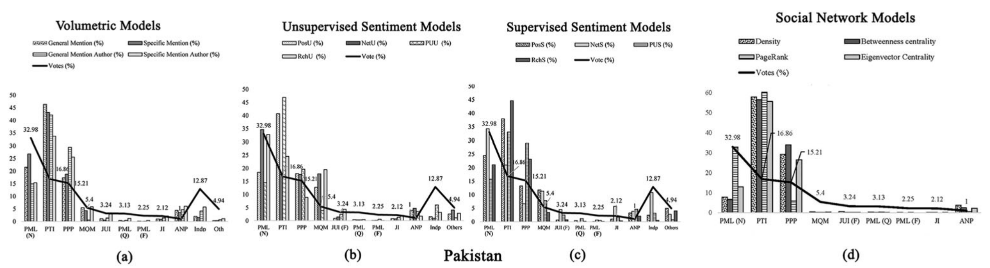


Figure 3. Party-level performance evaluation for Pakistan.

Table 7. Comparison against previous election results and opinion polls.

Country	Source	Sample size	Month conducted	MAE	RMSE	RMSE-MAE
Malaysia	Last election	8 m	March 2008	0.73	3.65	2.91
	Merdeka	1,600	May 2013	7.25	10.50	3.25
	Twitter	1.1 m	May 2013	1.96	2.80	0.84
India	Last election[a]	426 m	April 2009	3.54[a]	22.63	19.09
	India Today-CVoter	21,792	December 2013	10.50	24.15	13.64
	NDTV- Hansa Research	24,000	May 2014	11.51	16.67	5.15
	CNN-IBN-Lokniti-CSDS	18,591	January 2014	9.76	20.47	10.70
	ABP News-Nielsen	64,006	December 2013	9.02	23.44	14.42
	Twitter	1.2 m	May 2014	4.31	8.71	4.40
Pakistan	Last election[a]	35 m	February 2008	6.26[a]	24.11	17.85
	Gallup Pakistan	9,600	February 2013	3.91	10.59	6.69
	IRI	4,997	December 2012	3.09	11.21	8.12
	SDPI	5,700	May 2013	2.08	7.59	5.51
	Twitter	1.1 m	May 2014	3.44	4.05	0.61

[a]Did not include a projection for all parties (see Results section).

in Malaysia, where the incumbent party has ruled since independence, the results of the previous election were able to predict the outcome for 2013 with under 1% MAE as compared to the survey poll. The opinion poll results reported by the Merdeka opinion poll ('Public Opinion Survey', 2013) only asked whether voters would prefer BN or PKR to rule, so we recalculated the errors for all our models by taking a sum of party predictions according to party affiliations. The best performing Twitter model outperformed the Merdeka opinion poll but not the previous election results.

In the cases of India and Pakistan, previous election results are not a perfect baseline for a few reasons. Firstly, in India, some parties such as the JD(U) shifted their allegiance as compared to the previous year's election. Secondly, parties such as AAP and YSRCP contested the general election for the first time in 2014. Thirdly, in Pakistan, some parties such as PTI and Jamaat-e-Islaami had boycotted the 2009 general election amidst concerns that it would not be free or fair. Nevertheless, we have still used previous election results to calculate mean average errors, based on the percentage vote shares for the parties that were common across the two elections. However, these results should be approached with caution.

All the opinion polls in India grouped individual parties according to their coalitions, into one of four categories – NDA, UPA, Third Front or Others. We recalculated the errors for all our models based on these new categories. In the case of India, the best performing Twitter model (net sentiment supervised) outperformed the major opinion polls – India Today-CVoter ('NDA may win', 2014), NDTV-Hansa Research ('NDTV's opinion poll', 2014), CNN-IBN-Lokniti-CSDS ('Methodology of Lokniti', 2014) and ABP News-Nielson ('Modi-led NDA way ahead', 2014). An ideal comparison would have been at the party level, to compare the errors in predicted vote share against those from opinion poll results for each party – but unfortunately, those numbers were not available.

In the case of Pakistan, the best Twitter model (net sentiment supervised) seems to perform at par with two of the three polls – the Gallup Pakistan conducted in February 2013 ('Political weather forecast', 2013) and the IRI poll conducted in December 2012 (Ahmad, 2013) which incidentally did not provide voting preference information for MQM, JUI-F and ANP at the country level. The best Twitter model did not perform as well as the SDPI poll (Suleri, 2013) conducted in May 2013.

Discussion

We have assessed the utility of predicting elections from social media by benchmarking our models against predictions based on previous election results and traditional opinion polls. We find that in the case of Malaysia, simple volumetric approaches show low errors, but overall, social media may offer little advantage over traditional polling. In the case of India and Pakistan, predictions using social media are at least at par with traditional opinion polls but we were restricted in conducting an exhaustive analysis due to lack of data availability or the changes in political dynamics (compared to last elections). In these cases, pre-processing steps such as weighting tweets according to the nearness of the elections or according to the social influence of the authors work well in combination with the sentiment to capture the recency and outreach of positive messaging about a political party. However, in the case of Pakistan, where most parties are under-represented on Twitter, applying many pre-processing steps together appears to lead to the loss of valuable signal.

Tables 8 and 9 summarize the key insights, benefits, and challenges observed as a result of the (a) pre-processing methods and (b) approaches followed in this study. Table 8 highlights the need to recalibrate tweet share according to the actual number of participating users, such as by counting specific mentions and unique authors. While heavy users could monopolize a party's presence on Twitter, equally dangerous are a larger number of unique users, which may reflect the presence of bot accounts (Ratkiewicz et al., 2011), or astroturfing, where an organization simulates organic popularity on a social medium by employing 'click farms' for political campaigning (De Cristofaro, Friedman, Jourjon, Kaafar, & Shafiq, 2014; Ratkiewicz et al., 2011). In future work, social media users who could potentially represent bots or click farms should be identified and discarded prior to predictive model building.

The biggest observed impact is of temporal weighting, which can immediately and greatly improve the predictive performance of all models – as was observed in the case of India. Although it is possible that older tweets would have a greater number of impressions, and thus could influence more readers and be more impactful than newer tweets, the findings suggest that recent tweets are more important for the election

Table 8. Insights from pre-processing.

Pre-processing method	Justification	Main benefit	Main challenge
X_m: Specific mentions	To filter out spam tweets which tend to mention a lot of trending words such as political parties altogether	Reduces noise in estimating vote shares	It may inadvertently discard non-spam tweets
X_a: Authors of tweets	To count the actual number of engaged users	Reduces the effect of broadcasters – heavy users indulging in one-way information dissemination	If a political party is in the news, several users may mention it without intending to vote for it
X_{pos}: Positively-labeled tweets	Negative tweets are not an intention to vote	Helps to reduce the noise because of the overwhelmingly negative nature of political discussions on Twitter	It suffers from low recall
X_t: Temporal weighting	Tweets closer to an election are more important than tweets two weeks ago	Better captures the pulse of voters close to an election	If a political party is in the news, it would be unduly favored by this method

Table 9. Insights from different approaches.

Approach	Justification	Main benefit	Main challenge
Volumetric	Simple approaches can work well in countries with high internet access	Easy to obtain	Needs to be normalized to reflect actual author count; otherwise it can be gamed by astro-turfing or heavy users
Unsupervised Sentiment	Positive tweets help to reduce the noise because of news-sharing	Works well in countries with native English language speakers	Vocabularies in non-English countries; misses context even in English tweets in other cultures
Supervised Sentiment	Better adapted to mining voting intentions on Twitter; Good at capturing inferred sentiment from seemingly neutral text	Predictive models based on positive tweets can improve upon raw volumetric shares	Needs text parsing and labeled foreign language lexica
Network	A connected community reflects a stronger voter base than isolated tweets	Removes the over-estimation and under-estimation effects. A diffused network means greater outreach and greater electoral success	Computationally expensive. Can be gamed by bot accounts because it counts at-mentions

outcome. However, temporal weighting is liable to further over-represent a party with an inordinate social media presence, such as Pakistan's PTI. Although this study using a simple linear function to weigh tweets based on the distance from the election date, in future work, it would be fruitful to explore different weighting mechanisms to find the one that is best suited to this problem.

Table 9 discusses that rather than the party occupying a central position in the social network graph, it is the first-degree connections, or the users who talk about the party, who are the most important, in terms of whether they leverage their influence by interacting with others in the social network. Social network models can uncover the underlying connections and structural pattern of discussions related to political parties, and remove the over-estimation and under-estimation effects encountered in volumetric and sentiment analysis, by focusing on the position and interaction patterns of political parties in the overall network of election discussions on social media.

In order to predict the vote shares of independent candidates and regional parties, information about the community structure, such as the number of users or the number of followers, can help to refine the raw projections. Methods combining sentiment and network analysis work well together in these cases.

Conclusion

This study has made two important contributions to the state of the art in predicting elections through social media, Firstly, it has provided a full-scale comparison of predictive models across three Asian countries to explore the generalizability of its results. Secondly, it has identified the social media characteristics that are predictive of electoral success in different contexts, specifically demonstrating the benefits of incorporating the temporal nature of social media in any estimates, and the ability of social networking models in overcoming overestimation and underestimation errors.

Our findings show that mining Twitter for predicting election outcomes is a promising direction for India and Pakistan, but it did not offer any advantage over traditional polls in Malaysia. The findings corroborate previous work in that different pre-processing methods make can make all the difference in the final accuracy of the model in predicting election outcomes (Jungherr et al., 2012). In terms of approaches, the findings suggest, unsurprisingly, that a political party garnering positive sentiment closer to an election is likely to have a higher vote share than a political party which could have the same number of positive mentions spread out over a longer period. On Twitter, network breadth, rather than depth, is important for electoral success – so a political party with a wider community of users mentioning it in a positive light, would be likely to have a higher vote share than a party with a centralized social network and a fewer number users tweeting more actively about it.

A limitation of this work is the unavailability of multilingual resources for detecting Malay, Urdu, and Hindi sentiment, which could have improved our models' performance. Generic English sentiment lexica were found to be inadequate in capturing the sentiment in countries where English is not the native language, because of the inherent cultural differences in language usage, or because users prefer to tweet in other languages. Filtering out non-English tweets may have penalized regional parties such as BSP and YSRCP in India, which demonstrated a preference for local dialects over English to connect locally with their voter base. Another limitation is that this study was conducted on a 1% sample provided by the Twitter streaming API in the one-month period leading up to each country's election – this choice of data collection method may have inadvertently affected the balance of the dataset (Huberty, 2015).

Although the focus of studies like ours has traditionally been to predict election results, we posit a flipped question – could geo-located social media posts be more representative of citizens' preferences than physical votes? In countries where citizens are subjugated – either by the government or by terrorist threats, it is likely that the online medium would afford a 'safe space' (Castells, 2015, p. 81) for voters to express their voting preferences. We thus anticipate that online voting could lead to different results and interesting insights if it were suitably implemented. However, considering that only one in five tweets are ever geo-located (Wu, 2013), discarding the remaining tweets could result in the irrevocable loss of signal and power for predictive models.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Kokil Jaidka is a postdoctoral fellow at the Department of Computer and Information Science, University of Pennsylvania. Her research interests lie in computational approaches to explore the role of online communication in health and political outcomes.

Saifuddin Ahmed is a doctoral candidate at the Department of Communication, University of California, Davis. His research interests lie in new and emerging media, political communication, election studies, and public opinion.

Marko Skoric is Associate Professor at the Department of Media and Communication. His research interests are focused on new media and social change, with particular emphasis on civic and political implications of new communication technologies.

Martin Hilbert is Associate Professor at the University of California, Davis. He pursues a multidisciplinary approach to understanding the role of information, communication, and knowledge in the development of complex social systems.

ORCID

Marko Skoric http://orcid.org/0000-0001-6578-9872

References

Ahmad, Z. (2013, January 28). IRI survey: PML-N's popularity on the rise. *The Express Tribune*. Retrieved from https://tribune.com.pk/story/499599/iri-survey-pml-ns-popularity-on-the-rise/

Ahmed, S., Jaidka, K., & Cho, J. (2016). The 2014 Indian elections on Twitter: A comparison of campaign strategies of political parties. *Telematics and Informatics*, *33*(4), 1071–1087.

Ahmed, S., & Skoric, M. (2014). My name is Khan: The use of Twitter in the campaign for 2013 Pakistan general election. In *Proceedings of the 47th Hawaii international conference on system sciences* (pp. 2242–2251). Waikoloa, HI: Institute of Electrical and Electronics Engineers.

Ahmed, S., & Skoric, M. (2015). Twitter and 2013 Pakistan general election: The case of David 2.0 against goliaths. In *Case studies in e-government 2.0* (pp. 139–161). Cham: Springer International Publishing.

Bermingham, A., & Smeaton, A. F. (2011). On using Twitter to monitor political sentiment and predict election results. In *Sentiment analysis where AI meets psychology (SAAIP)* (pp. 2–10). Chiang Mai: ACL.

Boutet, A., Kim, H., & Yoneki, E. (2012, June). What's in your tweets? I know who you supported in the UK 2010 general election. In *Proceedings of the international AAAI conference on weblogs and social media* (pp. 411–414). Dublin: Association for the Advancement of Artificial Intelligence.

Bovet, A., Morone, F., & Makse, H. A. (2016). Predicting election trends with Twitter: Hillary Clinton versus Donald Trump. *arXiv preprint arXiv:1610.01587*.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Report No. C-1). Miami, FL: The Center for Research in Psychophysiology.

Brajawidagda, U., & Chatfield, A. T. (2014). Roles of social media in open data environments: A case study of the 2014 Indonesian presidential election voting results. In *Proceedings of the 25th Australasian conference on information systems* (pp. 1–14). New Zealand: Auckland University of Technology.

Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, *30*(2), 136–145.

Bravo-Marquez, F., Gayo-Avello, D., Mendoza, M., & Poblete, B. (2012). Opinion dynamics of elections in Twitter. In *Proceeings of the Eighth Latin American Web Congress* (pp. 32–39). Cartagena de Indias,Colombia: IEEE.

Burnap, P., Gibson, R., Sloan, L., Southern, R., & Williams, M. (2015). 140 characters to victory?: Using twitter to predict the UK 2015 general election. *Electoral Studies*, *41*, 230–233.

Castells, M. (2015). *Networks of outrage and hope: Social movements in the Internet age*. New York, NY: John Wiley & Sons.

Ceron, A., Curini, L., Iacus, S. M., & Porro, G. (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*, *16*(2), 340–358.

Ceron, A., & D'Adda, G. (2013, August). Enlightening the voters: The effectiveness of alternative electoral strategies in the 2013 Italian election monitored through (sentiment) analysis of Twitter posts. In *Proceedings of the ECPR conference* (pp. 4–7). Bordeaux: ECPR.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *Inter Journal Complex Systems*, *1695*(5), 1–9.

Cuzzocrea, A., Papadimitriou, A., Katsaros, D., & Manolopoulos, Y. (2012). Edge betweenness centrality: A novel algorithm for QoS-based topology control over wireless sensor networks. *Journal of Network and Computer Applications*, *35*(4), 1210–1217.

De Cristofaro, E., Friedman, A., Jourjon, G., Kaafar, M. A., & Shafiq, M. Z. (2014). Paying for likes? Understanding facebook like fraud using honeypots. In *Proceedings of the Internet Measurement Conference* (pp. 129–136). Vancouver: ACM.

Diaz, F., Gamon, M., Hofman, J., Kiciman, E., & Rothschild, D. (2014). *Online and social media data as a flawed continuous panel survey*. Retrieved from Microsoft website: http://research.microsoft.com/flawedsurvey

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Gaurav, M., Srivastava, A., Kumar, A., & Miller, S. (2013, August). Leveraging candidate popularity on Twitter to predict election outcome. In *Proceedings of the 7th workshop on social network mining and analysis* (pp. 7–16). Chicago, IL: ACM.

Gayo-Avello, D. (2013). A meta-analysis of state-of-the-art electoral prediction from Twitter data. *Social Science Computer Review*, *31*(6), 649–679.

Gayo-Avello, D., Metaxas, P. T., & Mustafaraj, E. (2011). Limits of electoral predictions using twitter. In *Proceedings of the fifth international AAAI conference on weblogs and social media* (pp. 490–493). Barcelona: Association for the Advancement of Artificial Intelligence.

González-Bailón, S., Banchs, R. E., & Kaltenbrunner, A. (2012). Emotions, public opinion, and US presidential approval rates: A 5-year analysis of online political discussions. *Human Communication Research*, *38*(2), 121–143.

González-Bailón, S., & Paltoglou, G. (2015). Signals of public opinion in online communication a comparison of methods and data sources. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 95–107.

Huberty, M. (2015). Can we vote with our tweet? On the perennial difficulty of election forecasting with social media. *International Journal of Forecasting*, *31*(3), 992–1007.

Jungherr, A., Jürgens, P., & Schoen, H. (2012). Why the pirate party won the German election of 2009 or the trouble with predictions: A response to tumasjan, A., sprenger, T. O., sander, P. G., & welpe, I. M. "predicting elections with twitter: What 140 characters reveal about political sentiment". *Social Science Computer Review*, *30*(2), 229–234.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1/2), 81–93.

Khatua, A., Khatua, A., Ghosh, K., & Chaki, N. (2015). Can Twitter_trends predict election results? Evidence from 2014 Indian general election. In *Proceedings of the 48th Hawaii international conference on System sciences (HICSS)* (pp. 1676–1685). Manoa, HI: IEEE.

Livne, A., Simmons, M. P., Adar, E., & Adamic, L. A. (2011). The party is over here: Structure and content in the 2010 election. In *Proceedings of the international AAAI conference on weblogs and social media* (pp. 17–21). Barcelona: Association for the Advancement of Artificial Intelligence.

Lui, C., Metaxas, P. T., & Mustafaraj, E. (2011). *On the predictability of the US elections through search volume activity* (Report no. 23). Retrieved from http://repository.wellesley.edu/scholarship/23/

MacDonald, R., & Mao, X. (2015). *Forecasting the 2015 general election with internet big data: An application of the TRUST framework*. Glasgow: Business School-Economics, University of Glasgow.

Methodology of Lokniti-IBN Tracker Round II. (2014). *Lokniti,org*. Retrieved from Lokniti website: http://www.lokniti.org/tracker-poll-round-2.php

Modi-led NDA way ahead than UPA, to win 236 seats in LS polls 2014: ABP News-Nielsen Opinion Poll. (2014). *ABP news*. Retrieved from http://www.abplive.in/india/2014/02/22/article267206.ece/Modi-led-NDA-way-ahead-than-UPA-to-win-236-seats-in-LS-polls-2014-ABP-News-Nielsen-Opinion-Poll

Monti, C., Rozza, A., Zappella, G., Zignani, M., Arvidsson, A., & Colleoni, E. (2013). Modelling political disaffection from Twitter data. In *Proceedings of the second international workshop on issues of sentiment discovery and opinion mining* (pp. 3–11). Chicago, IL: ACM.

NDA may win over 200 seats as Modi's popularity soars further: India Today Mood of the Nation opinion poll. (2014). *India Today*. Retrieved from http://indiatoday.intoday.in/story/nda-may-win-over-200-seats-as-modis-popularity-soars-further-india-today-mood-of-the-nation-opinion-poll/1/339374.html

NDTV's opinion poll: latest all-India forecast. (2014). *NDTV*. Retrieved from https://www.ndtv.com/india-news/ndtvs-opinion-poll-latest-all-india-forecast-733104

Newman, M. (2010). *Networks: An introduction*. New York, NY: Oxford University Press.

Payne, J. G. (2010). The Bradley effect: Mediated reality of race and politics in the 2008 US presidential election. *American Behavioral Scientist*, *54*(4), 417–435.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *LIWC2007: linguistic inquiry and word count*. Austin, TX: liwc. net.

Pimenta, F., Obradovic, D., & Dengel, A. (2013). *A comparative study of social media prediction potential in the 2012 us republican presidential preelections*. Cloud and green computing (CGC), 2013 third International conference, IEEE, Karlsruhe, Germany, 30 Sept.-2 Oct. 2013.

Political weather forecast for general election 2013. (2013). *PILDAT*. Retrieved from the PILDAT website: http://www.pildat.org/publications/publication/elections/PoliticalWeatherForecastforElection2013.pdf

Public Opinion Survey. (2013). Public opinion survey 2013 Peninsular Malaysia voter survey (2013). Retrieved from Merdeka website: at http://www.merdeka.org/v2/download/Survey%20Release%20May%203%202013.pdf

Ratkiewicz, J., Conover, M., Meiss, M. R., Gonçalves, B., Flammini, A., & Menczer, F. (2011). Detecting and tracking political abuse in social media. In *Proceedings of the 20th international conference companion on world wide web* (pp. 297–304). Hyderabad: ACM.

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). Truthy: Mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th international conference companion on world wide web* (pp. 249–252). Hyderabad: ACM.

Sanders, N. (2011). *Twitter sentiment corpus*. Retrieved from http://www.sananalytics.com/lab/twitter-sentiment/

Sang, E. T. K., & Bos, J. (2012). *Predicting the 2011 Dutch senate election results with twitter*. In Proceedings of the workshop on semantic analysis in social media. (pp. 53-60), Association for Computational Linguistics, Stroudsburg, PA, 2012, April.

Satria, I., Kurnia, A., & Nurhadryani, Y. (2014). Influence of presidential candidates e-campaign towards voters in 2014 presidential election in Bogor City. In *Proceedings of the Advanced computer science and information systems (ICACSIS)* (pp. 90–95). Jakarta: IEEE.

Skoric, M., Liu, J., & Lampe, C. (2015). Gauging public opinion in the age of social media. In *Proceedings of the ICA annual meeting, San Juan, Puerto Rico*. San Juan: ICA.

Skoric, M., Poor, N., Achananuparp, P., Lim, E. P., & Jiang, J. (2012). Tweets and votes: A study of the 2011 Singapore general election. In *Proceedings of the 45th Hawaii international conference on System science (HICSS)* (pp. 2583–2591). Manoa: IEEE.

Song, M., Kim, M. C., & Jeong, Y. K. (2014). Analyzing the political landscape of 2012 Korean presidential election in Twitter. *IEEE Intelligent Systems*, *29*(2), 18–26.

Stone, P. J., Bales, R. F., Namenwirth, J. Z., & Ogilvie, D. M. (1962). The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Behavioral Science*, *7*(4), 484–498.

Suleri, A. (2013, May 9). *Political barometer – a study of public opinion – voters' preferences and political parties' popularity across Pakistan (Phase 2)*. Retrieved from Pakvoter website: http://www.pakvoter.org/content/political-barometer-study-public-opinion-voters%E2%80%99-preferences-and-political-parties%E2%80%99

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*(12), 2544–2558.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proceedings of the international AAAI conference on weblogs and social media* (pp. 178–185). Washington, DC: Association for the Advancement of Artificial Intelligence.

Weingert, M., & Sebastian, E. (2015). *Predicting primary election results through network analysis of donor relationships*. Retrieved from Stanford website: http://web.stanford.edu/class/cs224w/projects_2015/Predicting_primary_election_results_through_network_analysis_of_donor_relationships.pdf
Wu, S. (2013). *Twitter and Privacy: Nearly one-in-five Tweets divulge user location through geotagging or metadata*. Retrieved from USC website: at https://pressroom.usc.edu/twitter-and-privacy-nearly-one-in-five-tweets-divulge-user-location-through-geotagging-or-metadata/