© *Media Watch* 6 (2) 238-254, 2015
ISSN 0976-0911 e-ISSN 2249-8818
DOI: 10.15655/mw/2015/v6i2/65670

Twittering Public Sentiments: A Predictive Analysis of Pre-Poll Twitter Popularity of Prime Ministerial Candidates for the Indian Elections 2014

KALYANI SURESH & CHITRA RAMAKRISHNAN
Amrita Vishwa VidyaPeetham, Coimbatore, India

Twitter is a useful tool for predicting election outcomes, effectively complementing traditional opinion polling. This study undertakes a volume, sentiment and engagement analysis for predicting the popularity of Prime Ministerial candidates on Twitter as a run-up to the Indian Elections 2014. The results from a survey of 2,37,639 pre-poll tweets finds tweet volume as a significant predictor of candidate vote share, and volume and sentiments as predictors for candidate engagement levels. Higher engagement rates evolve from the horizontality of conversations about the candidate, therefore indicating a high degree of interactivity, but do not translate into a higher vote share.

Keywords: Twitter analytics, Indian elections 2014, Modi, Kejriwal, Rahul Gandhi, sentiment analysis, twitter engagement rate

The 2014 general elections in India may be called the social media election of the world's largest democracy. The event saw political parties not only using social media platforms as the new battleground during the campaign, but also engaging the voter base with campaign related conversations. Tumasjan, Sprenger, Sandner and Welpe, 2010; Hong and Nadler, 2011; and O'Connor, Balasubramanyan, Routledge and Smith, 2010, have tried to measure the impact of the social media electoral landscape on the pre-poll popularity of political candidates in the United States and Europe. The studies analyzed collective trends in quality (sentiments) or quantity (volume) of conversations of twitter users.

Predicting pre-poll results is challenging in a socio-politically diverse country like India. India had 278 million registered Internet users as of October 2014 (IAMAI and IMRB International Report 2014). With voluminous amount of voter conversations on Twitter, and every other political party and politician joining the 'twittersphere,' the medium lends itself to opinion mining analysis in this context. Thus traditional exit polls need not be the only tool to predict electoral performance and popularity of political candidates.

Twitter allows for researchers to understand the political representativeness of its users by analyzing the nuances of public sentiments and consumption of political data. This study uses a data set from the Indian twitter landscape as a run–up to the Indian General Elections 2014, to predict the pre-poll popularity and engagement levels of Prime Ministerial candidates, using both tweet volume (tweet count) and sentiment (positive and negative emotions) analysis.

Literature Review

Related research in the area of mining of opinions and analysis of sentiments in the political arena show three emerging areas: Event monitoring (monitoring reactions in social media during

Correspondence to: Kalyani Suresh, Amrita School of Communication, Amrita Vishwa Vidyapeetham, Ettimadai, Coimbatore-641 112, Tamil Nadu, India. E-mail: s_kalyani@cb.amrita.edu

a specified event), real world continuous value predictions (predicting stock market values) and result forecasting (predicting election results).

According to Ceron, Curini, Iacus and Porro (2012), sentiment analysis of social networks throws up an accurate predictions about political leaders and parties which are popular. Previous research has shown the efficacy of Twitter in predicting election outcomes. Tumasjan et. al. (2010) suggest that election results can be forecast with greater accuracy with the help of candidate mentions. According to Bakliwal, Arora, Madhappan, Kapre, Singh and Varma (2013) identifying tweet sentiment can increase the accuracy of traditional opinion polling. O'Connor et al. (2010) have shown a significant relationship between President Barrack Obama's popularity ratings and the sentiments of Twitter users.

Metaxas, Mustafaraj and Gayo-Avello (2011) researched various state elections combining tweet volume analysis (volume of tweets, retweets, mentions etc.) and sentiment analysis, and posited that sentiment improved the precision of forecast rather than an analysis based solely on volume. Chung and Mustafaraj (2011) find that the number of candidate mentions or tweets does not appropriately provide a correct prescience into the electoral trend analysis, whereas Asur and Huberman (2010) find volume to be a strong predictor and sentiment to be a handy, yet weaker predictor.

All the studies mentioned above have focused either on volume or sentiment analysis, or in some cases both, to perform the prediction. But political candidates have different ways of engaging with the voter base and different levels of involvement. This study, groups the political candidates that have been analyzed, based on their twitter content volume, sentiment and engagement rates. Engagement rate measures the level of activity and interaction between platform members.

According to a joint research by the IMAI and IMRB International (IAMAI, 2014), there are 278 million Internet users in India. Most of them are young people from urban areas who are socially engaged, through networking sites such as Twitter and Facebook. A report from the IRIS Knowledge Foundation and IAMAI (2013) estimated that in almost 160 constituencies participating in the Indian General Elections 2014, there was a high likelihood of social media users playing a very important role in charting the road for shaping the way power-politics would be played out in India.

The Hindu (India's national newspaper) report dated May 13, 2014, based on the statistics shared by Twitter India, states that there were 56 million election-related tweets from January 1 to May 12, 2014. This is more than twice the election-related Indian tweets generated for the whole of 2013. Narendra Modi was found to be the most popular election-related term (11.1 million tweets) followed by the Aam Aadmi Party leader Arvind Kejriwal and the BJP. Rahul Gandhi was the fifth most tweeted about while the Congress party came in the tenth spot.

Context of the Study

Many newspapers and opinion polls narrowed down on three possible candidates who might have an impact on the national electorate in 2014: Narendra Modi, the BJP candidate; Rahul Gandhi, the Vice President of the Congress party and Arvind Kejriwal, the founder of Aam Aadmi Party (AAP). These candidates were and still are very popular in social media. While both Kejriwal and Modi have twitter accounts; Rahul Gandhi seems to be happy with Facebook and Whatsapp. All three candidates made concerted efforts to consolidate their social media presence and kept their campaign focused on mobilizing the younger vote bank.

Table 1. Candidate profiles on twitter (as on 8 April, 2014)

Candidate	Joined twitter	#tweets	#followers
Narendra Modi	January 2009	5275	4.86 million
Arvind Kejriwal	November 2011	3178	1.97 million

The study of twitter as a predictor of election outcomes is not without limitations. A McKinsey report, compiled by Gnanasambandam et. al. (2013) found that Internet usage in rural areas in India was just one twelfth of the urban usage. This digital divide coupled with conflicting claims regarding predicting election results by using Twitter due to the demographic bias (e.g Gayo-Avello et al., 2013) are some problem areas for the researchers. Added to this, a research study on the US national polls conducted by Mitchell and Hitlin (2013) for Pew Research Center has shown that tweeters' reactions to events are often deviant from the overall public opinion.

There has been a dramatic change in the twitter ecosystem since the time the studies mentioned above were conducted. Also since the volume of twitter usage for political opinions has substantially grown in the last few years in India, it is worth evaluating its predictive ability in the Indian elections 2014. This paper is also an attempt to fill the vacuum that exists in academic research, as far as Asian studies in social media predictive analysis is concerned.

Theoretical Framework

Rogers (1995) as quoted by Sahin (2006) defines diffusion as a systematic method which involves the use of certain channels for communicating innovation among the stakeholders of a social setup (p. 1). Direct contact between adopters or actors is a prerequisite for rapid diffusion of ideas to take place (McAdam & Rucht, 1993). However, Ahmed and Jaidka (2013) feel that Twitter's unique ecosystem might not lend itself to a simple interpretation based on diffusion of innovation (p. 31).

According to Enjolras, Steen-Johnsen and Wollebæk (2012), social media sites have inherent affordances and network functionalities that determine conditions of diffusion of civic and political information leading to participation. Affordances are actions enabled by the design of the medium used. Also according to the authors, "network effects transform individual action into collective action through collective mechanisms. The network's functionalities and affordances help disseminate information." Due to the 'small world effect' or the six degrees of separation phenomenon (i.e., the tendency of information bits in networks to be separated from each other only by a few steps) the information bits in social media sites are collectively disseminated to all parts of the network, and to a varied population within the network. These information cascades also have a motivational effect, since social media makes people's choices visible to all. Thus, social media affords a broader and more efficient mobilization processes than earlier forms of mobilization.

According to Sen (2012) and Habermas' (1989) public sphere is a setting that people use for political participation and the process of enactment is carried out through the medium of talk. It is an institutionalized space where citizens interact and discuss subjects of common interest. Twitter thus conforms to the definition of 'online' networked public sphere.

Granovetter (1985) was of the view that embeddedness emphasized the influence of inter-personal relationship in human behavior. The co-existence of both of their colleague politicians and the public in a common social networking platform for the politicians shapes two different contexts of embeddedness on Twitter - the civic and political (Yoon, 2011).

Tham and Zanuddin (2013) propose a model adapted from Lilleker (2006) to delineate levels of political communication. According to them, "the public sphere (made up of political actors such as president, prime minister and cabinet, national and local government and political parties) are no longer only transforming the messages to the ground (citizens and voters), but also communicating with other non-elective organizations (media, business sector, public organizations, etc). This two-way communication between elective political officials and non-elective organizations has made the communication vary, and both compete with one another to obtain the communication objective." (p.7)

McCombs and Shaw (1972) discovered the relationship between the agenda of the press and the public on major issues of importance by using the agenda setting theory. This was replicated by Dearing and Rogers (2007). To determine the agenda setting influence of Twitter, Leavitt, Burchard, Fisher, and Gilbert (2009) examined more than one lakh tweets and concluded that the content created by those who followed the news outlets were republished to other users on a bigger scale.

Using the theoretical framework of Diffusion of Innovation in social networks, and building on the study by researchers such as Smailoviæ, Grcar, Lavrac, and Znidarsic (2013) that use a combination of volume, sentiment analysis and engagement rate of tweets; this study examines two lakh thirty-seven thousand six hundred and thirty-nine tweets drawn over a period of six days prior to the Indian Lok Sabha (Parliament) elections 2014, and constructs a prediction model based on tweet volume, tweet engagement rate and lexicon-based sentiment polarity (positive and negative sentiments).

Volume analysis involves counting the number of candidate tweets, retweets, mentions and hashtags. Sentiment analysis attempts to study the voting intentions from tweets. Tweet engagement rate measures social media success. This study is based on the connection between the number, nature and success measure of prime ministerial candidate tweets vis-a-vis the number of votes that the candidate received.

Hypothesis

H1: Higher volume of tweets, retweets, hashtags, and mentions are linked to a large number of votes.

H2: Volume and sentiments expressed in tweets, retweets, hashtags and mentions are significant predictors for twitter engagement levels of candidates.

H3: Higher twitter engagement of a candidate in terms of volume of tweets, retweets, hashtags and mentions is linked to a large number of votes.

H4: Higher twitter engagement of a candidate in terms of positive tweets, retweets, hashtags and mentions is linked to a large number of votes.

Methodology

Data collection: pulling tweets

Tweets were imported from Twitter using Scraper Wiki. A list of keywords, that were not case-sensitive, was used to gather tweets containing username and @username. From these tweets, a list of frequently repeated hashtags were generated, and further tweets were drawn using these hashtags. The list is given in Table 1. Data thus generated contained tweets, retweets, hashtags and mentions from the candidates (except Rahul Gandhi, who does not have an @handle).

Table 1: Usernames, @username and #hashtags used to draw tweets

username	@username	#hashtags
Rahul Gandhi	narendramodi	Aap
Narendra Modi	ArvindKejriwal	Aapkamanifesto
Arvind Kejriwal	kejriwalarvind	Haryanaforaap
	bewithrg*	Bjp
		Modivision
		Abkibaarmodisarkar
		Namoforpm
		Namoteaparty
		Congress
		Amethifacts
		myvote2congress
		Incampaigntrail
		Latestnews

*No official twitter handle for Rahul Gandhi was found

According to Wu, M. (2012), Lithium's principal scientist of analytics, "social media sentiment for election outcomes cannot predict with any accuracy beyond a window of 1.5 weeks" (p. 1). Following this time frame, what Wu calls the 'predictive window', this study analyzed two lakh thirty seven thousand six hundred and thirty-nine tweets collected for a period of six days (from April 3 to April 8, 2014).

Variables for study

The following were the independent variables for the study:

(i) Tweets, retweets, hashtags and user mentions (user mentions classified as subsets: party-based, theme-based and name-based).

(ii) Retweets (named influencers for this study was classified into two subsets: media and non-media retweets). Retweets endorse and rebroadcast content by other users, raising content usability. Higher volume of retweets signify high influence levels.

(iii) @username and mentions (name-based, party-based and theme-based)

(iv) Hashtags

Wu, Hofman, Mason and Watts (2011) posit that volume of retweets and mentions give an estimation of the level of individual engagement in interactive political conversations. The number of retweets aid in keeping the momentum in messages from other users (Kelly et. al., 2012). Tweets and retweets from the media tend to be a good measure of the influence of the tweeter being talked about by them. Barbera and Rivero (2013) opine that high volume of candidate theme or party mentions can be considered a sign of the horizontality of conversations about that political actor, therefore indicating a high degree of interactivity.

Theme-based mentions can be located by parsing and analyzing the collected tweets for frequently occurring nouns, verbs and other parts of speech. (Ahmed and Jaidka, 2013) Ten broad themes were identified for the study, and tweets were drawn using these keywords: development and economy, law and order and internal affairs, caste and religion, corruption, defence, foreign policy, health care, youth and employment, women's safety and security, science and technology and education.

Hashtags in tweets identify a tweet as part of a particular conversation or topic. They can provide insights into specific cultural and socio-political conversations. The researchers extracted a dataset of tweets first based on the keywords mentioned in Table 1 and then extracted hashtags from that dataset to avoid sampling bias. The tone of voice expressed in tweets, retweets, hashtags and mentions differ, but certain concordances on positivity or negativity can be distinguished. For purposes of the study, the average tweet engagement rate was calculated for the candidates based on the formula by Social Bakers, in Coggins, S. (2012). This formula was used to calculate and define the dependent variable for the study.

$$\left\{ \frac{\dfrac{\text{Replies + Retweets on a given day}}{\text{\# of tweets made by the candidate on a given day}}}{\text{Total followers on a given day}} \times 100 \right\}$$

Data Analysis

Volume Analysis

The per day average volume of tweets, retweets, hashtags and mentions (@username, in reply to candidate screen name, party and theme-based mentions), positive and negative emotions expressed, overall sum of candidate's party mentions and tweet engagement rate for each candidate were calculated. Table 2 gives the details of the volume analysis.

Table 2: Volume analysis

Average/day (N)	Narendra Modi (Modi)			Arvind Kejriwal (Kej)			Rahul Gandhi		
Tweets (fromcandidate)	**9**			3			-		
Retweets (from candidate)	0.5			**3.3**			-		
Hashtags (from candidate)	0.3			0.3			-		
Mentions (from candidate)	0.8			**2.7**			-		
Tweets(from others)	**6575**			4792			2484		
Count of tweeters tweeting about candidate	**7642**			5479			3970		
Overall retweet count	2993952			**3401243**			3816		
Hashtags (used by others)	**549**			284			300		
@username	**1494**			955			759		
In reply to candidate screen name	**1570**			1142			-		
Party-based mentions	*BJP	**AAP	***Cong	BJP	AAP	Cong	BJP	AAP	Cong
	6257	989	2004	828	**4566**	296	980	336	**2099**
Theme-based mentions	**11608**			3054			7119		
Candidate name-based mentions	Modi	Rahul	Kej	Modi	Rahul	Kej	Modi	Rahul	Kej
	11332	137	481	260	127	**11950**	515	231	**5867**
Media tweets/retweets	52			**636**			405		
Followers count for candidate	**3672953**			1593789			-		
Positive emotions (sum)	**15602**			10983			3542		
Negative emotions (sum)	**11035**			6264			2619		
Candidate's overall Party mentions (sum)	BJP			AAP			Congress		
	8065			5891			4399		
Candidate overall name based mentions (sum)	12107			**18118**			495		
Overall Tweet Engagement Rate	2.3			**5.2**			-		

*BJP=BharatiyaJanata Party **AAP= AamAadmi Party ***Cong= Indian National Congress Party

From Table 2, it is clear that Modi leads over Kejriwal in terms of the following categories of variables: followers count, volume of tweets, count of tweeters, hashtags, mentions (@username, theme and party-based). Kejriwal scores over Modi in terms of the following categories of variables: volume of his retweets, mentions by him, overall retweet count, media tweets/retweets (influencers), candidate name-based mentions and overall tweet engagement rate. Kejriwal, a relatively newer entrant into politics, does not show a very big lag in volume of content in comparison to Modi.

(i) *Overall retweets*: Kejriwal had higher volumes for overall retweets followed by Modi and Rahul Gandhi, proving that content around Kejriwal was highly endorsed and rebroadcast, raising content usability.

(ii) *Media tweet/retweets*: With a very high media/retweet count, Kejriwal has mentions of one media organization and two media persons in his tweets: @abpnewstv, @sagarikaghose and @sardesairajdeep. He was retweeted by media persons/organizations: sardesairajdeep (Rajdeep Sardesai), Pritish Nandy, mediacrooks and India Today, with Rajdeep Sardesai, India Today and mediacrooks being among his top 20 retweeters. Modi had mentions of one media organization in his tweets: @NavbharatTimes and was retweeted by media persons/organizations: mediacrooks, IndiaBTL, Yuvai TV, Rajdeep Sardesai and SandeepWeb, with Rajdeep Sardesai, IndiaBTL and Yuvai TV being among his top retweeters. The overall high volume of media tweets/retweets for Kejriwal signifies the level of his interactivity with media in political conversations.

(iii) *Sentiments*: Modi had higher positive and negative sentiments overall in comparison to Kejriwal. While the difference between positive sentiments for both candidates was not very large, there was a significant amount of difference for the negative sentiments. Kejriwal had significantly fewer negative sentiments than Modi. It is significant that after the declaration of the exit polls, Modi had tweeted on 12 May, 2014 that social media helped him understand people's sentiments. This was retweeted 1,103 times and favourited 1,096 times.

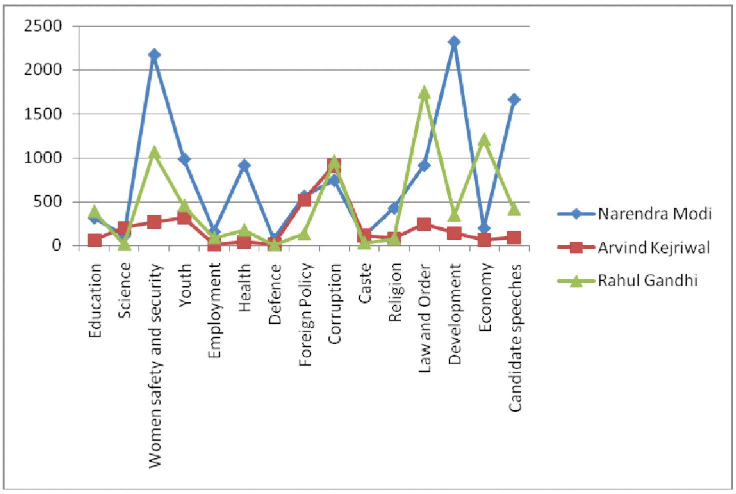
(iv) *In reply to screen name*: Modi gains over Kejriwal in terms of volume of in reply to screen name but the difference is not very huge.

(v) *Mentions*: In terms of the overall party-based mentions, the most-cited party overall was the BJP followed by the AAP and the Congress. Kejriwal had higher volumes for overall name-based mentions followed by Modi and Rahul Gandhi.

(vi) *Theme-based mentions*: The candidates differed in the topic genres, with Modi-related content drawing more volume of themed mentions than that of Rahul Gandhi and Kejriwal. The volume of election-related tweets around each subject was calculated. Women safety and security seemed to be the most talked about issue followed by law and order, development, corruption, and the candidate speeches. The least talked about was defence.

(vii) *Engagement rate*: The overall volume of twitter engagement rate for Kejriwal was higher than that of Modi. Figure 2 shows the engagement rate of both candidates across the time frame for the study.

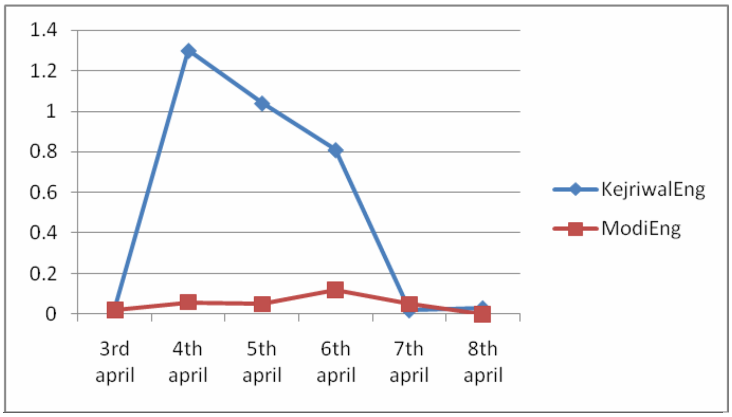
Discussions around Modi were typically high for development, women safety and security and his speeches. For Rahul Gandhi, discussions revolved around women safety and security, law and order and economy-based themes. For Kejriwal, the top conversation was around youth, corruption and law and order. Discussions around all three candidates seemed to converge around the theme of corruption.



Topical Themes

Figure 1. Relative volume of topical themes discussed around each candidate

(viii) *Engagement rate*: The overall volume of twitter engagement rate for Kejriwal was higher than that of Modi. Figure 2 shows the engagement rate of both candidates across the time frame for the study.



Time frame

Figure 2. Comparison of twitter engagement rates of Narendra Modi (ModiEng) and Arvind Kejriwal (KejriwalEng)

Tweet sentiment polarity detection

Date-wise, document level text analysis of twitter text found in the variable categories designed for the quantitative analysis of the study was done using LIWC 2007. The categories include tweets, retweet, tweets and retweets with hashtags, tweets and retweets with user mentions, theme-based and party- based mentions, tweets and retweets of media influencers. LIWC categorizes words into over 80 dimensions including descriptor categories like word count and words per sentence, linguistic, paralinguistic and punctuation words, psychological and personal category words. Nine significant dimensions that aligned with polarity dimensions (positive and negative emotions) were chosen for text analysis. The following were used to profile the candidates: positive emotion; negative emotion; negation; assent; optimism; tentative words; certain words; anger; and past, present and future words.

The study found instances of more positive than negative words in almost all categories in the study, supporting the Pollyanna hypothesis (Boucher & Osgood, 1969) that posits higher frequency and diversity among the use of positive words than negative ones in human communication. Garcia, Garas and Schweitzer (2012) found that positive words are higher than negative words in all written expressions, and this is especially true of the Internet.

(i) *Tweets from candidates*: Both Modi and Kejriwal have over 1 per cent of negation words, while assent words are non- existent. There is not much of a difference between usage of positive and negative terms contained in the tweets of both the candidates. On an average Modi's tweets have 3.74 per cent positive emotion words, and Kejriwal has 3.07 per cent positive emotion words. Negative emotion words in Modi and d Kejriwal's tweets are at 1.5 per cent and 1.7 per cent, respectively. Modi's tweets have much more optimism than Kejriwal's, and the respective percentages being 2.07 and 1.5 per cent. Anger words are at 0 .5 per cent in Kejriwal's tweets and much lower in Modi's tweets (0.14 per cent). Kejriwal uses more tentative words (1.5 per cent) whereas Narendra Modi uses only 0.7 per cent tentative words. Kejriwal uses more certain words than Modi and the percentages are 1.19 and 0.72, respectively. However, candidates use present and future words in almost the same percentages. But Narendra Modi uses more past words than Kejriwal. The percentages are 1.65 and 0.8, respectively.

(ii) *Tweets and retweets from others*: On 4 and 5 April 2014, tweets and retweets about Kejriwal have more emotion words and these numbers diminish by 7 and 8 April. Even though positive emotion words are more, there are a significantly large number of negative emotion words. According to Garcia et. al. (2012), the informativeness of words depends to a great extent on the emotional polarity, with negative emotion words correlating with higher informativeness. Media tweets show a high degree of positive emotion words for Modi. While the number of emotion words drops sharply on 7 April across all categories for Modi, the percentage of positive emotion words in tweets is 9.04 per cent for the day. Tweets about Rahul Gandhi contain a higher count of negative emotion words.

Steffans and Haslam (2013) find higher use of collective pronouns and lesser use of personal pronouns among the more successful political candidates in comparison to their lesser successful opponents in 80 per cent of all elections. Further, Mehl (2006) revealed the use of first-person singular to be a significant indicator of personal focus and reflection. Data in this study points to Modi using more of 'We' in his tweets than 'I'. This probably points towards Modi's successful political *avatar*. The average percentages for 'We' and 'I' used by Modi before the elections were 1.15 and 1.44 per cent, respectively. Kejriwal used 'I' at 3.25 per cent on average and 'We' only 0.396 per cent on an average day. This shows Kejriwal as the more inwardly reflective and personally focused than Modi.

(iii) *Hashtags*: Hashtags for Kejriwal have more number of emotion words on 4 and 5 April 2014. On 6 April, retweets with hashtags show a similar trend. Kejriwal's tweets pickup emotion on 7 and 8 April, and most significantly his tweets have more negative emotion words on 8 April. These two instances might be an indication of the contagion effect found to be true of social media networks in a study done by Coviello, Sohn, Kramer, Marlow and Franceschetti (2014).

(iv) *Party mentions*: For Kejriwal, positive emotions are twice as much as negative emotions in AAP party mentions. BJP mentions with respect to Kejriwal have slightly more positive emotions than negative emotions. Negative emotion words are more in Congress mentions for Kejriwal. In the party-based mentions referencing to Modi, positive emotion words are twice as much as negative emotion words in AAP mentions and four times as much as negative emotion words in BJP mentions. The average use of 'present' words per day amounted to 3.86 per cent for AAP, 5.35 per cent for the BJP and 6.84 per cent for the Congress. Tweets with party based mentions with respect to Kejriwal on an average use very less of anger, optimism, tentative, certain and future words. This is true for AAP, BJP and the Congress. The Congress and BJP party mentions use more of past words than AAP. Present words average slightly over 5 per cent for AAP and BJP. Average use of present words in Congress mentions is around 7.8 per cent. For Kejriwal, negation is greater than assent, and the percentages are insignificant. In the case of BJP mentions for Kejriwal, negation is twice as much as assent. For Rahul Gandhi, AAP mentions contain positive emotion words with respect to him. BJP mentions use negative emotion words a bit more than positive emotion words. In Congress mentions, positive emotions are twice as much as negative emotions. Negation words are greater than assent words for both AAP and BJP.

(v) *User mentions*: On 4 April when the user mentions tweet data for Kejriwal displays a sudden and steep rise in the number of positive and negative emotion words, Rahul Gandhi's user mentions data shows a high negative emotion count for the same day.

(vi) *Theme-based mentions*: For Kejriwal, the conversations surrounding education, employment, law, caste, development, foreign, religion and speech have dominated on 4 and 5 April in an emotionally charged atmosphere and fallen off after that, with the exception of employment and caste resurfacing on 8 April. Except for theme development, in which both positive emotion words and negative emotion words are equal and insignificant in terms of the average percentage, all other themes register greater positive emotion. Themes like women's safety and security, science and youth have more optimism words. Anger words are associated with economy and in one instance on 8 April, the theme employment registers 4.35 per cent anger words. The theme Caste has more negative emotion words on 4 April.

For Modi, caste-based mentions have both positive and negative emotion words at 1.78 per cent and 1.29 per cent, respectively. Theme-based tweets on employment and healthcare show a rise in the positive emotion words and caste shows a greater number of negative emotion words on 7 April. Corruption, defence and foreign policy themes show a considerable amount of negative emotion words. For Kejriwal, use of present words is high across all 15 themes. For Modi, present and future words are used together in themes like employment and foreign policy while past and present words are used together in themes like corruption, development and economy. All three words: past, present and future are used in themes like education and his speeches. Themes like defence, law and employment have a greater number of negative emotion words for Rahul Gandhi towards the 8 April. For the theme development, conversation starts off with negative emotion and turns positive towards the end of pre-election week. Conversation on economy loses emotion after the day one. Theme caste uses past, present and future words. Tentative words and negation words are also used, but there are no positive or negative emotion words. While there is no emotion associated with caste, religion has negative emotion words with respect to Rahul Gandhi. Theme religion uses 3.7 per cent anger words on 4 April.

(vii) *Media influencers*: For Kejriwal, media influencers use less of optimism words, tentative words and certain words. 7 to 8 per cent of all words used by media with reference to Kejriwal are present words. On 3 and 7 April, the percentage of 'present' words was 14.36 and 9.12 respectively. Anger words are used 1.03 per cent on an average day. On 7 April, the percentage of anger words used by media with respect to Kejriwal stands at 1.03 per cent. Negation words are used 1.1 per cent on an average day. Emotion words form 2.7 per cent of words used by media with respect to Kejriwal. Media influencers use present words and tentative words with respect to Modi. Both negation and assent are insignificant in terms of percentage for Modi, while media tweets on Rahul Gandhi and Kejriwal both show a strong use of positive and negative terms, only positive emotion words abound for Modi. For Rahul Gandhi, present words are used by media 4.18 per cent on an average day. On 6 April, present words are at 6.18 per cent.

(viii) *In reply to screen name*: For Modi, a trend is seen, wherein a surge in positive emotion from his tweets seems to have an effect on the 'in reply to status tweets'. The candidate tweets show a rise in the count of emotion words on 3rd and 5th April and the 'in reply to screen name' tweets show a rise in the count of emotion words on the 4th and 6th of April. In both cases, positive emotion words dominate.

Volume and sentiment as predictors of Twitter engagement rate

Tables 3 and 4 provide the significant Pearson's correlation (r) for the variables under study for the three candidates. Twitter engagement rates for Modi and Kejriwal are referred to as 'ModiEng' and 'KejriwalEng' respectively. Since there was no official handle for Rahul Gandhi, his twitter engagement rate was not calculated, and only correlations between dependent variables are shown. p-values are expressed in the following way: where $p < 0.001$, it is shown as (***), where $p < 0.01$, it is shown as (**), and where $p < 0.05$, it is depicted as (*).

Table 3. Narendra Modi Twitter engagement rate

Volume	Correlation
Retweets from others	.837 *
Theme-based mention (caste)	.863*
Theme-based mention (development)	.910*
Tweets from Narendra Modi	-.829*
In reply to candidate screen name	.838*
Sentiment	
Negative tweets from candidate	-.882*
Positive tweets from others	.832*
Positive retweets	.914**
Negative retweets	.881*
Positive hashtags	.900*
Negative hashtags	.845*
Positive user mentions	.904*
Positive In reply to candidate screen name	.912**

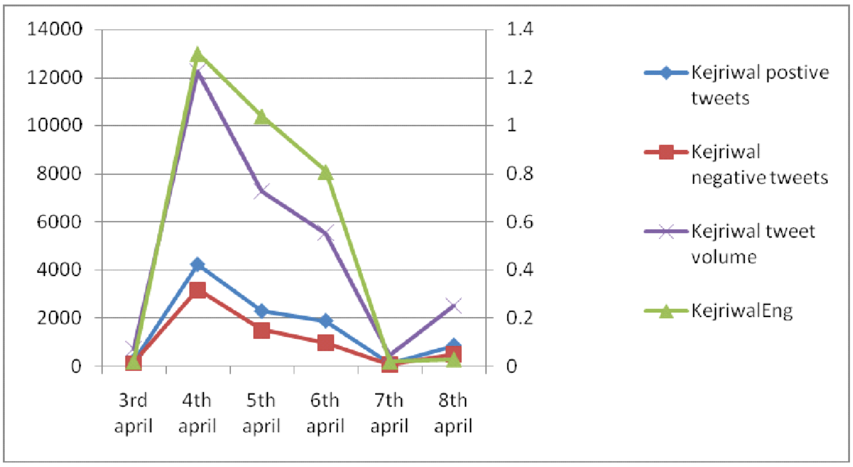
N= 2, 37,639 $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*)

Table 4. Arvind Kejriwal Twitter engagement rate

Volume	
Tweets from others	955**
Retweets from others	.993***.
Distinct count of tweeters tweeting about candidate	.978**
@username	.981**
Hashtags used by others	.965**
Theme-based mention (science)	.904**
Theme-based mention (employment)	.872*
Theme-based mention (defence)	.925*
Theme-based mention (corruption)	.989***
Theme-based mention (religion)	.866*
Theme-based mention (law and order)	.972**
Theme-based mention (development)	.921**
Party-based mention (BJP)	.908*
Party-based mention (AAP)	.960**
Tweets from Kejriwal	-.812**
In reply to candidate screen name	.879*
Media tweets/retweets	.823*
Sentiment	
Positive tweets	.943**
Negative tweets	.955**
Positive retweets	.885**
Negative retweets	.942**
Positive hashtags	.972**
Negative hashtags	.906*
Positive theme-based mention (science)	.974**
Negative theme-based mention (corruption)	.942***
Positive theme-based mention (defence)	.847*
Positive theme-based mention (development)	.886*
Negative theme-based mention (development)	.837*
Positive theme-based mention (foreign policy)	.897*
Positive theme-based mention (religion)	.830*
Positive user mentions	.914*
Negative user Mentions	.930**
Positive in reply to candidate screen name	.853*
Positive media tweets/retweets	.977**
Negative media tweets/retweets	.862*

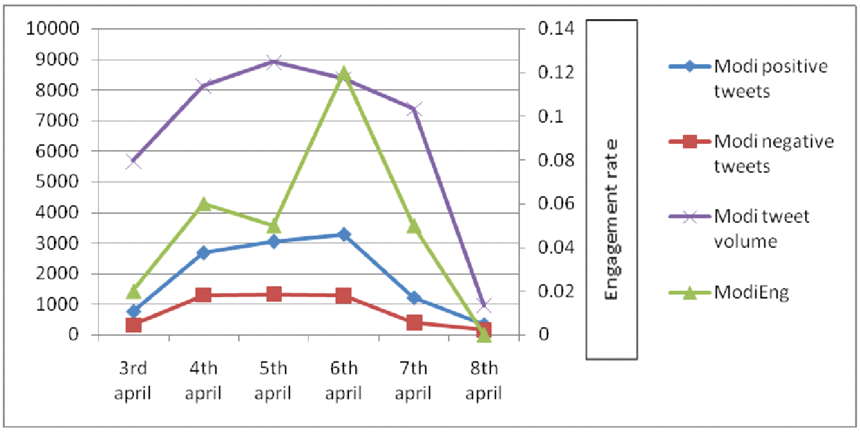
N= 2, 37,639 $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*)

From Table 4, it is evident that Kejriwal has higher measures of twitter engagement in relation to both volume of content and polarity of sentiments expressed. This may be due to the personalized approach adopted by the AAP on Twitter. Presented below are graphs depicting the content volume, positive/negative sentiment and twitter engagement rates of Kejriwal and Modi. Since Rahul Gandhi did not have an official twitter handle, the volume of content and sentiments expressed through a handle run by his fans was studied. Table 5 gives the significant Pearson's (r) values for volumes of mentions and sentiments.



Time frame

Figure 3.: Tweet volume, sentiment (positive and negative) and twitter engagement rates (KejriwalEng)



Time frame

Figure 4: Tweet volume, sentiment (positive and negative) and twitter engagement rates (ModiEng)

Volume of tweets, retweets and distinct count of tweeters for Rahul Gandhi's fan handle were significantly positively correlated with mentions of AAP. Retweets, distinct count of tweeters, hashtags, and @username was significantly positively correlated with themed mentions of caste. Hashtags and @username were significantly positively correlated with themed mentions of economy and Congress party-based mentions. Positive tweets were highly correlated with themed mentions of women, health, law and order, and corruption. More negative tweets were from tweets of fans and followers who had set up the @bewithrg fan handle. Positive hashtags correlated with the theme of law and order, and positive user mentions correlated with themed mentions of women and corruption.

Table 5: Significant correlations: Rahul Gandhi fan/followers' handle

Volume					
	Theme based mention- caste	Theme based mention- economy	Party based mention - AAP	Party based mention - Congress	
Tweets from fan handle	-	-	.855*		
Retweets from fan handle	.932**	-	.956**		
Distinct count of tweeters tweeting about candidate	.865*	-	.956**		
Hashtags used by fan handle	.965**	.885*		.971**	
@username (fan handle)	.849*	.812*		.963**	
Sentiment					
	Theme-based mention- women	Theme-based mention- health	Theme-based mention- law and order	Theme-based mention- corruption	Tweets from Rahul fans/ followers
Positive tweets	.897**	.930**	.851*		
Negative tweets					.875*
Positive hashtags		-	.827*		
Positive user mentions	.889*			.824*	

Findings

Volume analysis of candidates in this study strongly corroborates with the election results of the three candidates in their respective constituencies. The Election Commission of India data showed Modi in the lead with 5, 81,022 votes, followed by Rahul Gandhi with 4, 08,651 votes and Kejriwal with 2,09,238 votes. Sentiment polarity of candidates as found in this study corroborates with the CNN-IBN Live Elections Social Tracker (2014) figures. This study finds Modi garnering higher positive and negative emotions than Kejriwal overall. The difference between Kejriwal's positive and negative sentiment scores is much less than that of Modi.

The study findings support Hypothesis 1 that higher content volume is a predictor of higher share of votes. Tumasjan, et al. (2010) claim that twitter party mentions can reflect vote share with levels of accuracy comparable to traditional election results is supported by the study findings.

The correlation statistics obtained from the data support Hypothesis 2 that volume and sentiments expressed in tweets, retweets, hashtags and mentions are significant predictors for twitter engagement levels of candidates. The sample of Indian twitterati observed in this study has engaged in a depth of horizontal conversation about Kejriwal (therefore indicating a high degree of interactivity) but has lent a higher volume of content for Modi.

The win for Modi in terms of volume may be attributed to the reach of party and name-based mentions, resulting in widely dispersed interactions. However, this width of interaction seems to have lowered the depth and engagement levels for conversations around Modi. Kejriwal's engagement tallies with people, mostly youth who form a major population on twitter, expressing standpoints that may gather a lot of buzz (through name and theme-based mentions, overall retweets and media tweets/retweets), but this has not generated a lot of votes in the 2014 elections.

Sentiment analysis adjusts for this by showing his high engagement levels with positive and negative sentiments for the variables. Following Garcia et. al.(2012) and from the sentiment polarity data obtained on Kejriwal from this study, it can be said that high negative emotion words in discussions around him relate to higher informativeness and thus higher engagement levels for him. This high level of engagement on twitter cannot be looked at in isolation but needs to be explained in terms of his offline synergy with grassroots campaigns and online engagements across other social media platforms.

Hypothesis 3 and 4 are not supported in the study findings. Higher twitter engagement of a candidate in terms of volume of content and sentiments expressed is not related to greater number of votes for a political candidate. Kejriwal has managed to pull of his own against Modi on twitter on some levels. Having made an indelible stamp on Indian politics, he and his party need to strategize on how to translate his strong engagement levels on twitter to greater vote shares in the coming years.

Limitations

This study has various limitations. Political tweeting may be confined to a niche elite group. On twitter, unlike traditional poll studies, a sample of only those expressing their opinions on the medium can be studied, limiting the randomness of the sample. Also, tweeters are significantly younger, which can bias results. The demographics of tweeters are unknown. Also, Rahul Gandhi's engagement level could not be compared as he did not have an official twitter account.

In terms of sentiment analysis, except for present, negation, assent and emotion words, all other dimensions behave differently on different dates for different variables and candidates. These dimensions can be understood only through the context and through comparisons. The study has handled only six cases (i.e. six days leading up to the election day). With more data, these dimensions might show a definite pattern. Attempting to tame the LIWC dictionary and including more contextual words that suit the Indian environment will improve accuracy of results obtained.

This study fuels many questions, even as it seeks to answer some. The efficacy of Twitter as a poll outcome prediction tool has been studied and hypotheses proved and disproved. At the same time the dynamics of volume, sentiment and engagement trigger a multitude of questions on how the interaction between these three dimensions may affect the influence of political debates on Twitter. Our findings suggest important issues of future inquiry, while contributing towards filling up a vacuum in terms of lack of such research studies in the Indian and Asian context. A cross-sectional study involving more social media platforms for studying Indian voter behavior and poll predictions would significantly contribute to developing the literature in the Indian context.

References

Ahmed, S., & Jaidka, K. (2013). Protests against #delhigangrape on Twitter: Analyzing India's Arab Spring. *eJournal of eDemocracy and Open Government, 5*(1), 28-58. Retrieved from http://www.jedem.org

Asur, S., & Huberman, B. (2010). Predicting the future with social media. *Proceedings of ACM International Conference on Web Intelligence* (pp. 492-499). Toronto, Canada: IEEE. doi:10.1109/WI-IAT.2010.63

Bakliwal, A., Arora, P., Madhappan, S., Kapre, N., Singh, M., & Varma, V. (2012). Mining sentiments from tweets. *Proceedings of 3rd Workshop on Computational Approaches to Subjetivity and Sentiment Analysis (WASSA 2012), in conjunction with Association of Computational Lingusitics (ACL 2012)* (pp. 11-18). Jeju, Republic of Korea: Association for Computational Linguistics.

Barbera, P., & Rivero, G. (2014). Understanding the political representativeness of Twitter users. *Social Science Computer Review*. doi:10.1177/0894439314558836

Boucher, J., & Osgood, C. (1969). The Pollyanna Hypothesis. *Journal of Verbal Learning and Verbal Behavior, 8*(1), 1-8. doi:doi:10.1016/S0022-5371(69)80002-2

Ceron, A., Curini, L., Iacus, S. M., & Porro, G. (2012). Tweet your vote: How content analysis of social networks can improve our knowledge of citizen's policy preferences. An application to Italy and France. Retrieved from http://www.sisp.it/files/papers/2012/andrea-ceron-luigi-curini-e-stefano-iacus-1414.pdf

Chung, J., & Mustafaraj, E. (2011). Can Collective Sentiment Expressed on Twitter Predict Political Elections? *Proceedings of Twenty-fifth AAAI Conference on Artificial Intelligence.* San Francisco, USA: AAAI Press. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/view/3549

CNN-IBN. (2014). *IBN Live Elections Social Tracker*. Retrieved from CNN-IBN Live: http://ibnlive.in.com/general-elections-2014/social-tracker/

Coggins, S. (2012, October 2). *Twitter Metrics: Social Media Analytics (Web log Post)*. Retrieved from Vervely.com: http://vervely.com/twitter-metrics-social-media-analytics/2012/

Coviello, L., Sohn, Y., Kramer, A., Marlow, C., Franceschetti, M., & al., e. (2014). Detecting emotional contagion in massive social networks. *PLoS ONE, 9*(3). doi:10.1371/journal.pone.0090315

Dearing, J., & Rogers, E. (1988). Agenda-setting research: Where has it been, where is it going? *Communication Yearbook*, pp. 555-594.

ECI. (2014). *ECI Main Page*. Retrieved from ECI Web Site: eci.nic.in/eci_main1/index.aspx

Enjolras, B., Steen-Johnsen, K., & Wollebaek, D. (n.d.). How do social media change the conditions for civic and political mobilization. Retrieved from http://www.uio.no/english/research/interfaculty-research-areas/democracy/news-and-events/events/conferences/2012/papers-2012/steen-johnsen-elrojas-woolebaek-wshop%5D.pdf

Firstpost. (2014, May). *Why FB, Twitter and Google are betting big on India's Elections*. Retrieved from Firstpost.com: http://www.firstpost.com/politics/why-fb-twitter-and-google-are-betting-big-on-indias-elections-1510215.html

Garcia, D., Garas, A., & Schweitzer, F. (2012, May 10). Emotional persistence in online chatting communities. *Science Reports*. doi:10.1038/srep00402

Gayo-Avello, D., Schoen, H., Metaxas, P., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. *Internet Research, 23*(5), 528-543. doi:10.1108/IntR-06-2013-0115

Gnanasambandam, M; Madgavkar, A; Kaka, N; Manyika, J; Chui, M; Bughin, J; Gomes, M. (2013). *Online and Upcoming: The Internet's Impact on India.* McKinsey & Company. Retrieved from http://www.mckinsey.com/search.aspx?q=Internet%27s+impact+on+India

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology, 91*(3), 481-510. Retrieved from http://www.jstor.org/stable/2780199

Habermas, J. (1989). *The Structural Transformation of the Public Sphere: An Inquiry into a category of Bourgeois Society.* (T. Burger, Trans.) CC BY-SA 3.0. Retrieved from en.wikipedia.org

IAMAI. (2014). *Internet in India 2014.*. Retrieved from IAMAI: http://www.iamai.in/rsh_pay.aspx?rid=4hjkHu7GsUU=

India, E. (2014). *Election India Opinion Polls*. Retrieved from electionindia2014: http://electionindia2014.co.in

IRIS, & IAMAI. (2013). *Social Media and Lok Sabha Elections.* Retrieved from esocialsciences.org: http://www.esocialsciences.org/General/A2013412184534_19.pdf

Kelly, J., Barash, V., Alexanyan, K., Etling, B., Robert, G., & Palfrey, J. (2012). Mapping Russian Twitter. *Berkman Center Research Publication*(3). Retrieved from http://ssrn.com/abstract=2028158

Leavitt, A., Burchard, E., Fisher, D., & Gilbert, S. (2009). The influentials: New approaches for analyzing influence on twitter. *Web Ecology Project, 4*(2), 1-18. Retrieved from http://www.webecologyproject.org/2009/09/analyzing-influence-on-twitter/

Lilleker, D. (2006). *Key Concepts in Political Communication.* SAGE Publications.

McAdam, D., & Rucht, D. (1993). The Cross-national diffusion of movement ideas. *The Annals of the American Academy of Political and Social Science, 528*(1), 56-74. doi:10.1177/0002716293528001005

McCombs, M., & Shaw, D. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly, 36*(2), 176-187. doi:10.1086/267990

Mehl, M. (2006). The lay assessment of subclinical depression in daily life. *Psychological Assessment, 18*, 340-345.

Metaxes, P., Mustafaraj, E., & Gayo-Avello, D. (2011). How (not) to predict elections. *Proceedings of Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)* (pp. 165-171). Boston: IEEE Press. doi:10.1109/PASSAT/SocialCom.2011.98

Mitchell, A., & Hitlin, P. (2013). *Twitter Reaction to Events often at Odds with Overall Public Opinion.* Pew Research Centre. Retrieved from http://www.pewresearch.org/2013/03/04/twitter-reaction-to-events-often-at-odds-with-overall-public-opinion/

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM 2010)* (pp. 122-129). Washington, D.C: AAAI Press. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1536

Rogers, E. (1995). *Diffusion of Innovations* (Fourth Paperback ed.). New York: The Free Press Simon & Schuster Inc.

Sahin, I. (2006, Apr 3). Detailed review of Rogers' diffusion of innovations theory and educational technology - related studies based on Roger's theory. *Turkish Online Journal of Educational Technology, 5*(2), p. 1. Retrieved from http://eric.ed.gov/?q=technology+and+e ducation&ff1= souOnline+ Submission &id=ED501453

Sen, A. (2012). The social media as a public sphere: The rise of social opposition. *Proceedings of the International Conference on Communication, Media, Technology and Design* (pp. 490-494). Istanbul: ICCMTD.

Steffens, N., & Haslam, S. (2013). Power through 'Us': Leaders' use of we-referencing language predicts election victory. *PLoS ONE, 8*(10). doi:10.1371/journal.pone.0077952

Stieglitz, S., & Dang-Xuan, L. (2012). Social media and political communication: A social media analytics framework. *Social Network Analysis and Mining, 3*(4), 1277-1291. Retrieved from http://dx.doi.org/10.1007/s13278-012-0079-3

Tham, J., & Zanuddin, H. (2013). Malaysia's 13th general election: Political communication battle and public agenda in social media. *Proceedings from the Conference Organized by Asian Network for Public Opinion Research.* Seoul: ANPOR, p. 7. Retrieved from academia.edu: http://www.academia.edu/5433568/Malaysias_13th_General_Election_political_communication_battle_and_public_agenda_in_social_media

Toms, M. (2014, April 25). 20 million election tweets: India votes for Twitter. *Hindustan Times*. Retrieved from http://www.hindustantimes.com/business-news/20-million-election-tweets-india-votes-for-twitter/article1-1208135.aspx

Top Ten Indian Journalists to Follow on Twitter. (2013, July 30). Retrieved from Social Samosa: http://www.socialsamosa.com/2013/07/top-ten-indian-journalists-to-follow-on-twitter/

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of International AAAI Conference on Weblogs and Social Media; Fourth International AAAI Conference on Weblogs and Social Media.* Washington: AAAI Press. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441

Wu, M. (2012, March 19). *Science of Social Blog.(p. 1).* Retrieved from Lithium.com: http://community.lithium.com/t5/Science-of-Social-blog/Big-Data-Big-Prediction-Looking-through-the-Predictive-Window/ba-p/41068

Wu, S., Hofman, J., Mason, W., & Watts, D. (2011). Who says what to whom on Twitter. *Proceedings of the 20th international conference on World wide web* (pp. 705-714). Hyderabad: ACM.

Yoon, H., & Park, H. (2011). Social media information flow and public representation: A case of South Korean politicians on Twitter. *Proceedings of the 9th International Triple Helix Conference.* California: Triple Helix Association.

Dr. Kalyani Suresh is an assistant professor in the Department of Communication at Amrita Vishwa Vidyapeetham. Ettimadai, Coimbatore, India, She specializes in mediated communication, social media and youth studies.

Chitra Ramakrishnan is an assistant professor in the Department of Communication at Amrita Vishwa Vidya Peetham, Ettimadai, Coimbatore, India. Her current research interests include search marketing and data journalism.