Politically Predictive Potential of Social Networks: Twitter and the Indian General Election 2014

Anunay Kulshrestha
Stanford University
anunay@stanford.edu

Arpan Shah
Stanford University
ashah29@stanford.edu

Devin Lu
Stanford University
devinlu@stanford.edu

Keywords

Indian General Election, Twitter, Clustering, Sentiment Analysis, Augmented Contagion Model, Information Spread, Social Network Analysis, Sentiment Threshold

ABSTRACT

The Indian General Election 2014 witnessed the casting of 540 million votes, making it the largest democratic exercise in human history. The center-right Bharatiya Janata Party (BJP) single-handedly won a majority of seats in the lower house of the parliament, a feat emulated after 30 years in India's vibrant multiparty democracy where coalition governments have long been the norm. A new prime minister, Narendra Modi, swept into office with 31% of the vote riding on an extensive social media campaign – a significant first in Indian polity. Some commentators have even gone as far as calling it a Twitter election[2]. We investigate these claims by analyzing the Twitter network in India in the months leading up to and including the election. We study the use of social media by different political actors using an augmented contagion model of information dissemination. We look closely at both the direct role of the actors as well as their catalyzing role and influence in the network. Sentiment analysis based clustering is used to gauge the public opinion. Drawing on these sources, we compare the efficacy of the social media strategies of important political actors. We find that the BJP and its coalition partners pursued a more rigorous and effective social media strategy than those of other political actors. Furthermore, they were able to not only establish but also maintain a robust network of supporters that eventually translated into a significant electoral victory.

1. MOTIVATION

In the last decade, the advent of social media has changed significant parts of how people communicate. Political leaders use social media more than ever to communicate with the electorate. Between the 2009 and 2014 elections, the number of internet users in India had quadrupled to almost 240 million people. The 2014 elections were the first ones where media and journalists hypothesized that social media would play a significant role. It became an interesting question to understand whether there was any material difference between the social network strategy of the various political players. Furthermore, with polls often not capturing a representative sample due to various constraints; we wanted to determine if social network analysis may offer a credible way to determine poltiical sentiment. When looking at future elections in India, the US and across the world doing structured analysis of social networks may present us novel insights into people's views and positions and offer more credible strategies for political communication.

2. PRIOR WORK

There are two basic models for contagion to consider: Linear Threshold and Independent Cascade[1]. In the Linear Threshold model, nodes are activated based on whether the sum of edge weights to activated neighbors exceeds a randomly drawn activation parameter θ_v. Both edge weights and activation parameter are drawn randomly i.i.d from some underlying distribution.

In the Independent Cascade model, nodes again have a particular weight. However, the method of propagation is different: active neighbors cause probability of flipping according to the edge weight, where edge weights are again drawn i.i.d from some underlying distribution. Lerman and Rumi explore the cascading of information via the contagion model and compare and contrast this spread at a general level between nodes in the twitter graph and in Digg [4]. However they do not explore this information and its predictive power on elections.

We adapt some of these ideas in formulating our contagion model. Since it is likely that people will tweet multiple times in support or opposition to a particular political party, We present an augmented contagion model that accounts for repeated stimuli from the same source. This is further explained in section 5.

Direct analysis on the applicability of Twitter to elections has been looked into by prior researchers. McKelvey, DiGrazia, and Rojas demonstrate that to be able to link tweets to political outcomes it is very important to select for the right twiter messages [5]. We limit our research and data set only to political tweets.By doing this based on keyword relevant only to elections and political activities we make sure that the tweets are good candidates for analysis. Hao Wang et al look at how a real time system to categorize and analyze sentiment for a real time stream of twitter data could be used to analyze election outcomes but they do not actually perform the analysis involved [8]. Tumasjan et al look closely at german elections by analyzing tweets that mention politicians or political parties[7]. This analysis demonstrates that simply looking at number of tweets and mentions is sufficient to indicate sweeping trends that predict outcomes [7]. Our analysis uses aspects of this study since our tweets filter on similar criteria but the analysis we perform goes further in the network parameters explored. The analysis presented doesn't just look at the tweets but also the network over which these tweets are being shared. Meckel and Stanoevska-Slabeva [6] discuss the important of social media strategy for politicians to succeed especially in the German context; but do not explore what aspects determine the strength of such a strategy. Wong et al demonstrate the methods by which convex optimization and streams of tweets can be used to explore the tweet and retweet relationships to understand political leanings [9]. Wong et al make numerous interesting findings and motivated some of the research we did with retweet analysis. While Wong et al focus on the tweets themselves without exploring the static network graph; our analysis focuses on the graph and network properties as a means to establish politician strength. Amongst research that found limited efficacy of social media in mattering in electoral results, Kushin and Yamamoto utilized surveys to understand how much social media influenced electoral outcomes[3]. Their methodology didn't specifically look at posts, content or network parameters and our study differs in this key facet. Additionally the study was limited in that it only looked at college students and not a wider electorate [3].

Looking at work performed in this domain, Our study looks to expand on this work and answer specific questions related to the Indian elections and from a network analysis and tweet sentiment perspective.

3. DATA

The data obtained represents a substantial portion of the Twitter network from India and contains 15.5 million unique users. These users form the nodes of the network in this study. The data also contains up to 5000 followers for each of these account, which define the edges of our network. The number of political tweets in this entire network is 10.6 million, with $18,000$ of the tweets originating directly from politicians and political commentators in India. Each of the tweets in our data pertain to the elections and have been filtered by keywords (see appendix). Note that this data spans 47% of the $\sim$ 33 million unique Twitter users in India. [citation needed]

The data corpus consists of Users, Tweets and Sentiments. However some key concepts used throughout this paper with regards to the dataset need to be explained below. We build a directed graph representing the following relationship of the users as edges in the graph. We model the users as nodes and consider an edge to exist from node u to v if v follows u.

For the purpose of the analysis, we often separate users into Politicians and Non-Politicans. The set of **Politicians** that were active on twitter during the elections were identified from a twitter source. The entire list of political twitter handles analyzed are listed in the Appendix. All other nodes in the graph are said to be **Non Politicians** for the purpose of the analysis. In addition, each tweet has been subject to a sentiment analysis by derived from the words the tweet includes. The sentiment is expressed as a numerical value ranging from -1 to 1 indicating the degree of positivity/negativity towards the parties addressed or discussed in that tweet

One of the key tools for analysis of the strength of the network are the tweets themselves. In particular there are some important pieces of information in addition to sentiment and keywords captured by the tweets. We use the Tweet-time to do a time-series analysis of the graph. In addition to this, tweets that are retweets of others have the useful field signifying which user this tweet retweeted.

Nodes	15,181,851
Edges	39,548,702
Tweets	10,595,729
Politicians Tweets	17,815
Largest WCC	0.99

Table 1: Network Statistics

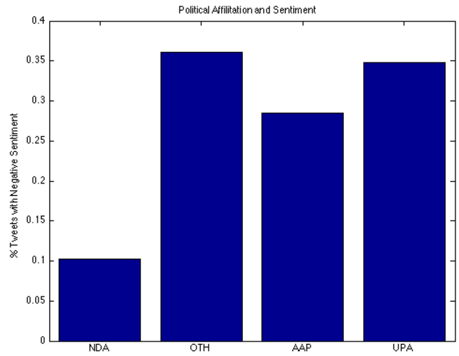

Figure 1: Percentage of Tweets with negative sentiment by political actor.

Finally since the elections in India had defined political leanings embodied by the 4 primary formations: NDA, UPA, AAP, OTHER(3rd Front/4th Front). We have split the political handles on twitter into their their associations with one of the above buckets.This split has also been listed in the appendix. We focus most of the attention of our analysis on NDA, UPA and AAP as they represent the three primarily discussed contenders in the media leading to the elections.

4. PRELIMINARY ANALYSIS OF GRAPH PROPERTIES

Table 1 displays some statistics of the network constructed using the data as described in preceding sections. This network is studied in the remainder of this paper. We begin by computing the largely weakly connected component (WCC) of the graph. Unsurprisingly, we found that 99% of all nodes were a part of the largest WCC. This implies that almost every user in our network who tweeted about the election is connected directly or indirectly to a political actor.

Interestingly, the sentiment of the tweets seems to reflect the outcome of the elections. The NDA had far less negatively classified tweets than the UPA overall. However, the time distribution of these negative tweets was not uniform: the BJP saw a relative burst of negative tweeting in March before settling into considerably more positive and neutral tweeting in the following months. Combined with the fact that April and May were the highest tweeting months of the campaign may have played a factor in lending them momentum in the final weeks of the campaign.

We then explored the average clustering coefficients of the politicians as categorized by their affiliation and found in-

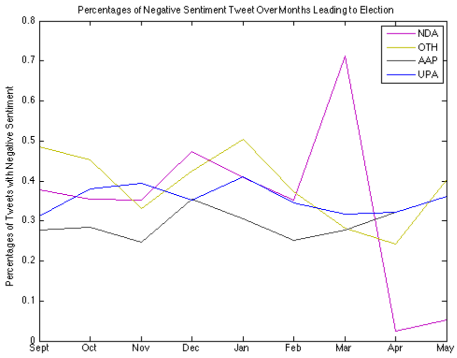

Figure 2: Percentage of Tweets with negative sentiment over time.

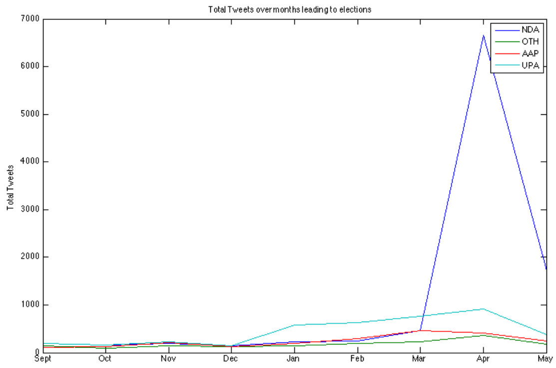

Figure 3: Total Tweets in each month leading up to the elections

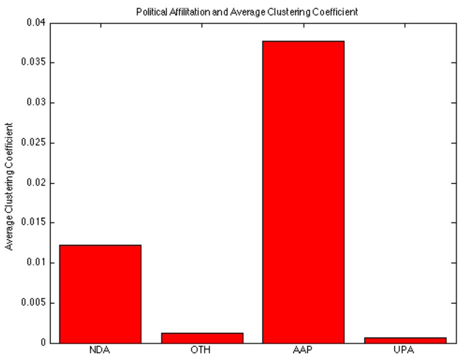

Figure 4: Clustering cofficient by political affiliation. AAP having the greatest network clustering affinity

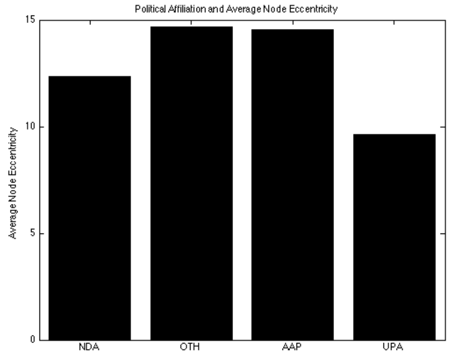

Figure 5: Average Node Eccentricities

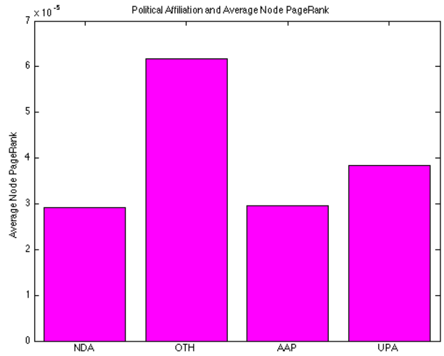

Figure 6: Average Node PageRank

teresting results here as well (specifically, the graphs under consideration here consist of the subgraph of a politician and all his/her followers). While the social network of the AAP is quite strong, from a contagion/influence perspective, between the two primary formations: NDA and UPA, the UPA's network is not as well connected. This could imply that contagions/ideas may find difficulty in spreading through the followers of the UPA politicians since the immediate network around the politicians does not reinforce each other as strongly.

Eccentricity Centrality (depicted in the graph below) is defined as the largest shortest-path distance from a given node to any other node in the graph. We found that the average value of eccentricity centrality for UPA is lower than that of NDA and AAP. This is surprising because this means it was easier for the UPA to reach the masses yet it fared poorly in the polls.

PageRank is another metric that produced some interesting results. As OTH includes political analysts as well as independent candidates, the high average PageRank for this group is expected. This is explained by the larger influence of its members. But we find that the average PageRank for the UPA is higher than that of the NDA and AAP. While this means that members of the UPA have a larger influence, they may have failed to capitalize on it.

5. MATHEMATICAL MODELS

For any node $v \in G$, let $OutDeg(v)$ denote the out-degree of v. Similarly, let $InDeg(v)$ denote the in-degree of v. We also define $OutNodes(v) = \{u|(v,u) \in E\}$ and $InNodes(v) = \{u|(u,v) \in E\}$.

In this paper, we introduce an augmented contagion analysis model which accounts for the impact of repeated stimuli from adjacent nodes. This is different from the usual contagion model in that it considers the number of times an infected node tries to spread the infection and thus, can be effectively used to model the flow of information.

We define $f : V \times V \times \mathbb{N} \rightarrow \mathbb{R}$ as

$$f(v,u,i) = \begin{cases} \frac{1}{InDeg(u)^i} & : v \in InNodes(u) \\ 0 & : v \notin InNodes(u) \end{cases} \quad (1)$$

This function f models the impact that the i-th stimuli (due to v) has on u. Applied to the problem at hand, this represents the impact on u by v's i-th tweet about a particular keyword k.

Thus, for a fixed keyword k, the total impact on u by its adjacent nodes is

$$\sum_{v \in InNodes(u)} \sum_{i=1}^{n_v} f(v,u,i) \quad (2)$$

where n_v represents the number of times v tweets about keyword k. Note that only adjacent nodes can impact u in this contagion model.

(2) simplifies to

$$\sum_{v \in InNodes(u)} \sum_{i=1}^{n_v} \frac{1}{InDeg(u)^i} \quad (3)$$

$$= \sum_{v \in InNodes(u)} \frac{1}{InDeg(u)} + \frac{1}{InDeg(u)^2} + ... + \frac{1}{InDeg(u)^{n_v}} \quad (4)$$

$$= \sum_{v \in InNodes(u)} \frac{InDeg(u)^{n_v} - 1}{InDeg(u)^{n_v}(InDeg(u) - 1)} \quad (5)$$

(To see the derivation of this expression, see the appendix).

If we consider only tweets by the nodes in $InNodes(u)$ about keyword k before node u's first tweet about k, we can somewhat gauge the 'threshold' of node u for tweeting about k. Informally, it is a measure of how many tweets (about k) u needs to see before u itself starts tweeting about k. This is the motivation behind using the augmented contagion model.

6. ALGORITHMS

Due to the large scale of our dataset ($\approx$ 7GB of Tweet data alone), we were forced to develop highly optimized algorithms for analyzing the data and generating results.

We made sure that most algorithms we employed had an amortized runtime of $O(n)$. We further used multiprocessing on a 16-core Amazon AWS EC2 instance to run these algorithms and compute the presented metrics.

Furthermore, we memoized most data - including the social graph and the total number of tweets by a given user etc.

Perhaps one of the most important algorithm we devised is the one used to compute the impact on every node using the augmented contagion model described above.

```
Data: An array of users (nodes) who tweeted about
      keyword k in chronological order
initialization;
numTweets = {};
resultTemp = {};
result = {};
count = {};
for node in array do
    if node in count then
        count[node] +=1;
    else
        count[node]= 1;
    end
    for follower in OutNodes(node) do
        if follower in resultTemp then
            resultTemp[follower] += (1.0 / (inDeg **
            count[node]));
        else
            resultTemp[follower] = (1.0 / (inDeg **
            count[node]));
        end
    end
    if node in resultTemp then
        result[node] = resultTemp[node];
    else
        result[node] = 0
    end
end
```

Algorithm 1: Augmented Contagion Model

Another set of analyses involved understanding sentiment within tweets and classifying parameters and trends based on tweets with specific keywords. Two specific objectives were outlined:
1) Measuring Subgraph parameters after classification based on sentiment for specific keywords linking tweets to the major political formations
2) Visualizing the graph based on these keywords
The results are outlined later in the paper.

To enable this analysis in an efficient manner, due to the large size of the data, the tweets were all read just once after which readable maps and lists were maintained on disk which summarized the data by classifying the tweets based on keyword[see appendix for set of keywords].

After processing the keyword $\rightarrow$ Tweets Mappings, To obtain a sentiment level understanding as well, another map was created on disk which consisted a map of nested keyword $\rightarrow$ Tweet Mappings, mapping from Positive/Negative Sentiment.

To gather insight from the data in the context of these elections, the keywords were further separated based on the political affiliations to do be able to visualize sentiment across the network towards the affiliations. In addition to the original calculations, The visualization was a non-trivial problem due to the large number of nodes with sentiment information. Filtering was done to obtain colorings based on users

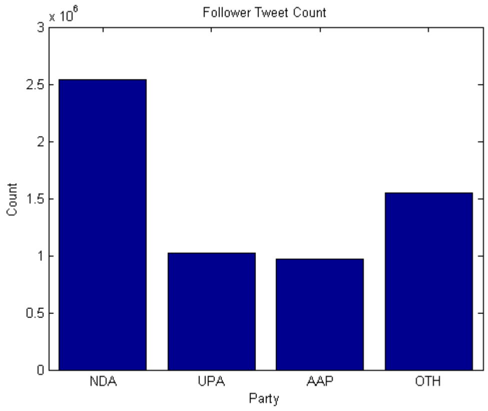

Figure 7: Total number of tweets by followers.

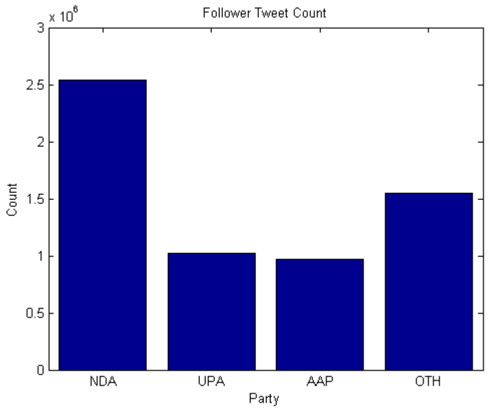

Figure 8: Number of retweets by supporters.

with **Uniquely** positive sentiment tweets towards a particular affiliation to be able to discern concrete support. The above limits the nodes, and edges exist if an edge exist in the real twitter network . The final visualizations involved converting to JSON format and using the scalable javascript rendering library: D3

7. ANALYSIS OF FOLLOWER STRENGTH

To disseminate information and be persuasive through social media, effective strategies rely on the power of the network and other individuals to be active propagators of ideas. The concept of virality stems from this notion. To understand how active and effective politicians were at building a strong support based we look at the number of political tweets written by the followers of political formations as well as the number of retweets for these followers:

From the graphs we start to see real differences in the strength of the supporters of the different coalitions. The followers of NDA leaders are almost twice as active at political tweeting during the election compared to to any other coalition, including all the others outisde of AAP and UPA combined.

Since these followers are so active, it would lend a significant edge to the NDA coalition in terms of bringing more attention to their message and campaign.

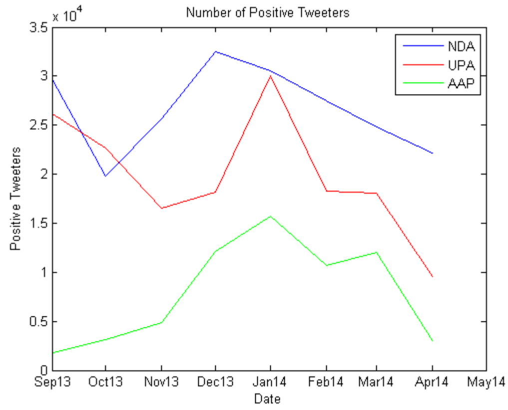

Figure 10: Number of positive tweeters over time.

The second graph is equally interesting in this scenario. A retweet is a mechanism by which people within the Twitter network can voice their support or agreement to a particular tweet. In the context of political tweets, a retweet is an especially interesting metric since it not just signals agreement but is also a way to further propagate the message through the network. This makes it a very good indicator of the strength of the follower. The followers of politicians who are most highly retweeted are therefore most successful not just at getting people to like what they say but also spread their message.

We see that the NDA dominates this metric as well: its retweet count is double that of both the UPA and the AAP and is even doing better than all the other fourth parties combined.

This corroborates the idea that the NDA led by Modi were able to build a strong network on social media.

8. ANALYSIS OF SENTIMENT BASED SUPPORTER SUBGRAPHS

Throughout the election season, the NDA maintained a better Twitter presence than both the UPA and the AAP. The following figures show the number of users who made a positive tweet about the NDA, UPA, and AAP as a function of month, followed by the average degree of the positive tweeters in over the same time period. Both indicate generally high engagment on Twitter with the NDA compared to the UPA and AAP.

Both these trends are not surprising, given the conventional wisdom and analysis above that the NDA was particularly strong in social media compared to the other two parties.

However, we did make the following interesting observation: During September to November of 2013, the clustering coefficient of the subgraph of Twitter users who made positive tweets about the AAP was markedly higher than that of the corresponding subgraph for both the NDA and the UPA (note that this is distinct from the clustering coefficient considered in section 2, which was formed of a subgraph of a

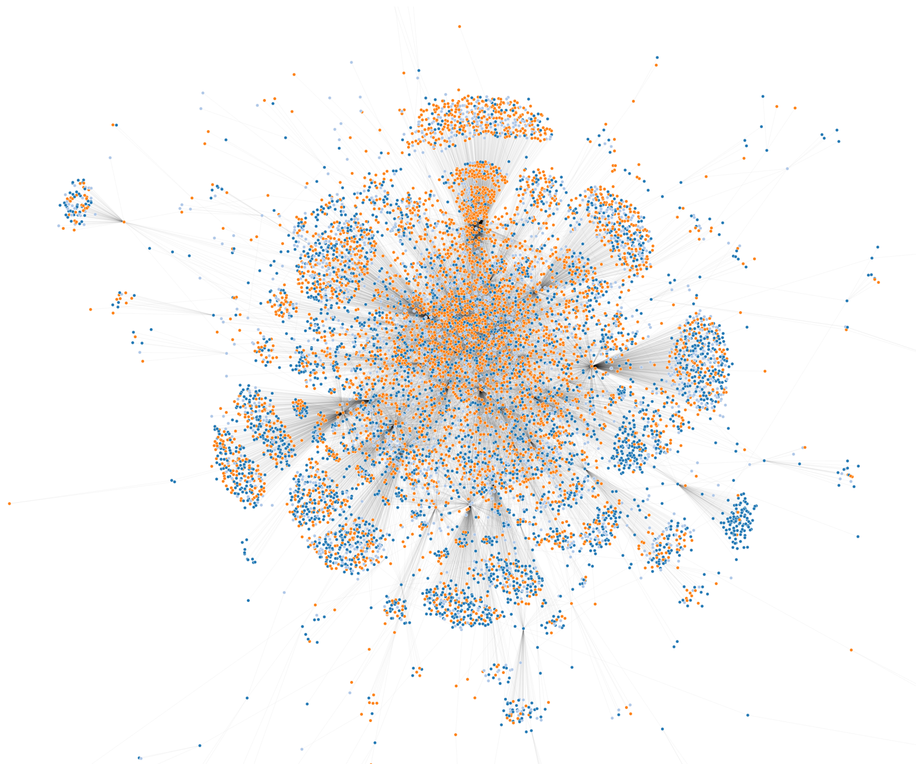

Figure 9: Visualization of a random subsampling of 10,000 politically active nodes. Orange nodes are supporters of the NDA, dark blue are supporters of the UPA, and light blue are supporters of the AAP. Edges are derived from follower relationships.

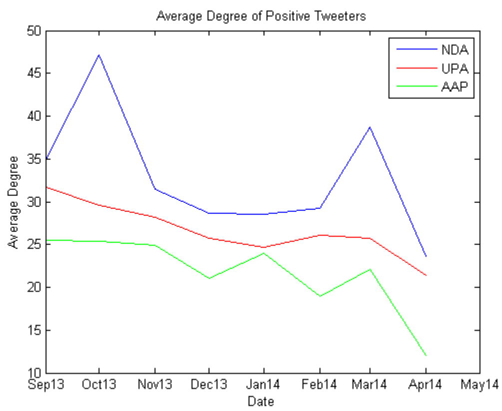

Figure 11: Average degree of positive tweeters over time.

politician and his/her followers). This could be an indicator of tighter and more concentrated clusters of interest, which would be a source of political strength that could be missed by opinion polls alone.

The AAP is a brand new party. It was founded only in November 2012 and the 2013 Delhi legislative assembly elections was its first electoral test. It is a small third party competing against the two factions that have dominated Indian politics for decades. In September and October, the conventional wisdom was that the AAP had no chance. Opinion polls generally supported this view[3]. In the end, however, conventional wisdom was wrong: The AAP made an unexpectedly strong showing, winning 28 out of 70 seats, becoming the second largest party in the assembly and eventually the majority partner in the ruling coalition.

In the national elections the following year, the clustering coefficient for the AAP supporter subgraph had fallen in line with that of the NDA and UPA. Indeed, in the national elections it did not perform particularly well.

These results are obviously preliminary and we will have to wait for more elections to take place for further analysis. However, the results from this past election cycle suggest clustering coefficient may be a good leading indicator of electoral performance, even when pundit opinion or polls are not.

The last analysis done was to understand whether there is polarization reflected in the support for the political affiliations contesting the elections. If nodes with a political leaning are clustered together, this would signify such behav-

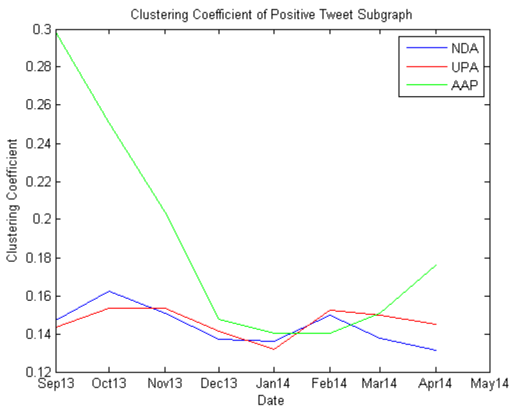

Figure 12: Clustering Coefficient of the subgraph constructed using only tweets with positive sentiment.

ior. Using the sentiment information, a rendering of nodes with uniquely positive sentiment towards at least one of the keywords associated with NDA, UPA or AAP leaning was done. This rendering has been shown in the graph above (Figure 1).

The results here are also interesting, while there are clusters of signficiant NDA support and clusters of UPA support. For the most part, the nodes are intermixed with significant follow relationships present between people with opposing political viewpoints. It is also clear that there are sum significant hub nodes within the graph. It is noteworthy that in corroboration with other findings, these hubs are primarily Orange(NDA) suppo

9. ANALYSIS OF RETWEET HOPS

As discussed earlier, retweets are a useful way to gauge how many people agree with the views of a particular political leader. A fair assumption, in this case, is that those who retweet a political leader's tweet are almost certainly his/her supporters.

To measure a leader's clout, apart from looking at the number of retweets, it may also be useful to look at the average hop distance between the political leader and people who retweet his/her tweets.

The graph presented above was generated by calculating this average hop distance of top leaders from every party and then aggregating the result.

It is evident that even here, the NDA was able to reach a wider audience as even distant nodes in the social graph retweeted NDA leaders' tweets.

10. ANALYSIS OF SENTIMENT THRESHOLD

We use the augmented contagion model to estimate the average number of tweets a user has to see (i.e. tweets by the people he/she follows) before he/she tweets something positive or negative about a keyword.

Here the social media edge of the NDA is starkly illustrated: firstly, the threshold for posting positive sentiments

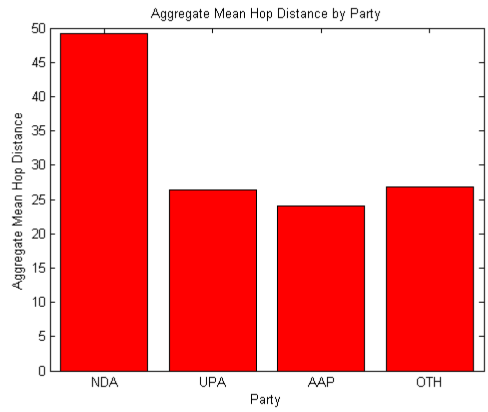

Figure 13: Average hop distance between political actor and retweeters.

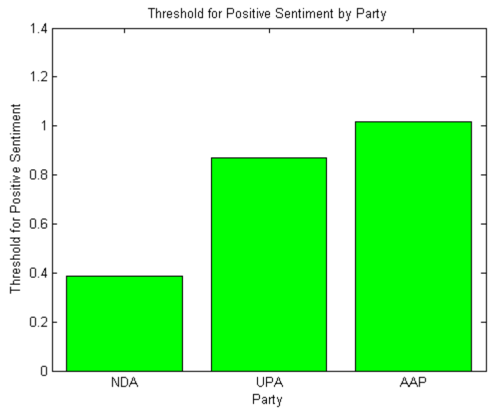

Figure 14: Average threshold for a tweet with positive sentiment.

was much lower for the NDA, at approximately 0.39 vs. 0.87 for the UPA and 1.02 for the AAP. Secondly, the advantage in resisting negative contagion was even more dramatic: anti-NDA sentiments required an average augmented contagion expoure of 3.63 to spread, as opposed to 0.97 and 0.99 for UPA and AAP, respectively.

Part of this discrepancy is likely due to the greater efforts of the NDA to saturate social media with their message, thus providing some "inoculation" against negative messages. Furthermore, the general population of NDA supporters may be more tech-savvy than the supporters of the other parties, providing a natural bias when calculating these metrics over the entire Twitter network.

11. CONCLUSIONS AND FURTHER WORKS

Upon performing the analyses above, ranging from Network parameters, Supporter Strength analysis, Sentiment analysis, Retweet analysis, and the augmented contagion analysis; Our results show that on almost every metric of evaluation, the NDA outperformed the UPA, AAP as well as other political parties. The strength of their network as well as the rigor with which they have pursued their social media strategy seems to have paid off as visible in the

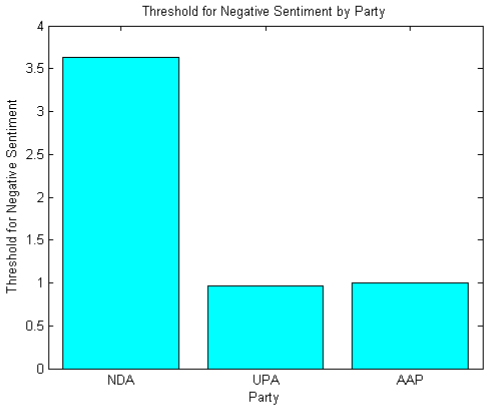

Figure 15: Average threshold for a tweet with negative sentiment.

general elections of 2014. They were able to convey their campaign messages both more effectively as well as numerously and the nature of the followers they managed to get on board were also beneficial towards that objective. It is also important to note that on a large number of these parameters, the NDA not just outperformed the AAP and the UPA marginally but substantially so. The performance of the AAP and the UPA was almost the same in a lot of the metrics where the NDA took significant leads.

In an age where social networks such as Twitter are a very effective tool for broadcast based communications, other parties have not have succeeded at utilizing this resources to the extent that the NDA was able to in these elections. With 65% of the population under the age of 35 and growing connectivity and use of the internet, Social Media and Social Networks will continue to be relevant in political discourse within India and across the world.

The goal of attempting to quantify an abstract notion as a successful Social Network Political Campaign is of course quite a complicated task. Quantifying it however, enables a more objective, holistic understanding of the phenomena that guide successful communication in such Networks and Network Analysis techniques provide tools to enable this.

There is a lot of interesting further work that is possible in the field of understanding effective use of social networks in a political or rhetorical context. Some such questions unexplored in this paper but within the scope of further work are:

1. Attempting to understand the compositions of the most successful political Tweets in the context of Indian Elections

2. Attempting to understand similarities and differences amongst such networks and elections across the world

3. Understanding which events and time periods within an election cycle are the most crucial to engage via social media and social networks

12. ACKNOWLEDGMENTS

We would like to thank Dr. Jure Leskovec and Dr. Rok Sosic for their invaluable help in procuring data for this study.

13. REFERENCES

[1] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[2] R. Khursheed. India's 2014 twitterelection | twitter blogs, May 2014.

[3] M. J. Kushin and M. Yamamoto. Did social media really matter? college students' use of online media and political decision making in the 2008 election. *Mass Communication and Society*, 13(5):608–630, 2010.

[4] K. Lerman and R. Ghosh. Information contagion: An empirical study of the spread of news on digg and twitter social networks. *ICWSM*, 10:90–97, 2010.

[5] K. McKelvey, J. DiGrazia, and F. Rojas. Twitter publics: How online political communities signaled electoral outcomes in the 2010 us house election. *Information, Communication & Society*, 17(4):436–450, 2014.

[6] T. Plotkowiak. Auch zwitschern muss man üben: Wie politiker im deutschen bundestagswahlkampf ■twitterten■. *NZZ*, (10. November):1, 2009.

[7] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10(1):178–185, 2010.

[8] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*, pages 115–120. Association for Computational Linguistics, 2012.

[9] F. M. F. Wong, C. W. Tan, S. Sen, and M. Chiang. Quantifying political leaning from tweets and retweets. *ICWSM*, 13:640–649, 2013.

14. APPENDIX I

A handle not associated with the NDA, UPA, or AAP is classified as OTH for "other."

14.1 Keywords captured within Tweet Data

Listed Below are the keywords that encompass all the tweets in the data:

"Bharatiya Janata Party", "BJP","Bahujan Samaj Party", "BSP","LK Advani","Mayawati","Narendra Modi","Congress Party","Rahul Gandhi","Manmohan Singh","Sonia Gandhi","Mamata Banerjee","Trinamool Congress","P Chidambaram","Rajnath Singh","Nitish Kumar","Janata Dal United","JDU","Naveen Patnaik","Sharad Pawar","Laloo Prasad","Rashtriya Janata Dal","Mulayam Singh Yadav","Jayalalitha","United Progressive Alliance","UPA","Dravida Munnetra Kazhagam","DMK","Nationa Democractic Alliance","Shiv Sena","Rashtriya Swayamsevak Sangh","Arvind Kejriwal"

quizderek	OTH
KalrajMishra	NDA
M_Lekhi	NDA
PawarSpeaks	UPA
prithvrj	UPA
PMOIndia	OTH
AITCofficial	OTH
meerasanyal	AAP
SudheenKulkarni	OTH
jayantsinha	NDA
GulPanag	AAP
Swamy39	NDA
Tarunvijay	NDA
KDSingh_India	OTH
SushmaSwaraj	NDA
BJP4India	NDA
NCPspeaks	UPA
ThinkwithGoogle	OTH
narendramodi	NDA
INCIndia	UPA
asadowaisi	OTH
ncbn	NDA
_SandeepDikshit	UPA
yadavakhilesh	UPA
PandaJay	OTH
dilipkpandey	AAP
ianuragthakur	NDA
smritiirani	NDA
JKNC_	UPA
shaziailmi	AAP
KapilSibal	UPA
ShahnawazBJP	NDA
JP_LOKSATTA	OTH
laluprasadrjd	UPA
buzzindelhi	OTH
PriyaDutt_INC	UPA
AnanthKumar_BJP	NDA
BJPRajnathSingh	NDA
ShainaNC	NDA
digvijaya_28	UPA

Table 2: Twitter Handles of Political Actors I

manoharparrikar	NDA
mayankgandhi04	AAP
Pallamrajumm	UPA
milinddeora	UPA
varungandhi80	NDA
VasundharaBJP	NDA
rajeev_mp	OTH
DrJitendraSingh	NDA
ShashiTharoor	UPA
PrakashJavdekar	NDA
ChennaiConnect	OTH
SalmanSoz	UPA
AamAadmiParty	AAP
SinghRPN	UPA
nsitharaman	NDA
priyankac19	UPA
MPNaveenJindal	UPA
NKSingh_41	NDA
ShuklaRajiv	UPA
ChouhanShivraj	NDA
ajaymaken	UPA
PiyushGoyal	NDA
harsimrat_badal	NDA
jkpdp	OTH
ArvindKejriwal	AAP
AapYogendra	AAP
mkstalin	UPA
arunjaitley	NDA
naqvimukhtar	NDA
drharshvardhan	NDA
RamJethmalani5	OTH
SushilModi	NDA
abdullah_omar	UPA
DrShahFaisal_	OTH
DrKumarVishwas	AAP
ManishTewari	UPA
AUThackeray	NDA
KirronKherBJP	NDA
supriya_sule	UPA
ashutosh83B	AAP

Table 3: Twitter Handles of Political Actors II

15. APPENDIX II

$$\frac{1}{x}+\frac{1}{x^2}+...+\frac{1}{x^n}=\sum_{i=0}^{n}\frac{1}{x^i} \tag{6}$$

Recall the formula for an infinite geometric series

$$\sum_{i=1}^{\infty}x^i=\frac{x}{1-x} \tag{7}$$

for $x<1$. So

$$\sum_{i=1}^{\infty}\frac{1}{b^i}=\frac{\frac{1}{b}}{1-\frac{1}{b}} \tag{8}$$

So we have

$$\sum_{i=1}^{\infty}\frac{1}{b^i}=\frac{\frac{1}{b}}{1-\frac{1}{b}} \tag{9}$$

Thus,

$$\sum_{i=1}^{n}\frac{1}{b^i}=\sum_{i=1}^{\infty}\frac{1}{b^i}-\sum_{i=n+1}^{\infty}\frac{1}{b^i} \tag{10}$$

$$=\sum_{i=1}^{n}\frac{1}{b^i}=\sum_{i=1}^{\infty}\frac{1}{b^i}-\frac{1}{b^n}\sum_{i=1}^{\infty}\frac{1}{b^i} \tag{11}$$

$$=\frac{\frac{1}{b}}{1-\frac{1}{b}}-\frac{1}{b^n}\frac{\frac{1}{b}}{1-\frac{1}{b}} \tag{12}$$

$$=(1-\frac{1}{b^n})\frac{\frac{1}{b}}{1-\frac{1}{b}} \tag{13}$$

$$=\frac{b^n-1}{b^n}\frac{\frac{1}{b}}{\frac{b-1}{b}} \tag{14}$$

$$=\frac{b^n-1}{b^n(b-1)} \tag{15}$$

Plugging this back into the original equation gives us the desired result.