ABSTRACT

Earlier this year, from April 7 to May 12, 540 million voters cast votes in the Indian general elections, the largest democratic exercise in human history. It was hugely influential: the right-wing Bharatiya Janata Party (BJP) won the prime ministership and a majority of seats in the Lok Sabha, the lower house of Indian congress. This ended 10 years of rule by the left-wing Indian National Congress party. A new prime minister, Narendra Modi, swept into office with 31% of the vote (a strong plurality in the highly fractured Indian political landscape). The Indian stock market jumped 5% on the news of his victory. The 2014 elections were also the first national elections in India that saw heavy use of social media, particularly Twitter.

We performed an analysis of the Twitter network in India in the months leading up to and including the election. Our subject of analysis are the politicians and the political parties they belong to. One of the major talking points of these elections and their results have been the effective use of social media by the winning party and the winning prime minister Narendra Modi. Some commentators have even gone as far as calling it a #TwitterElection. This project aims to evaluate these claims to draw a comparative landscape of the effectiveness of the use of social media by members of different political affiliations in the public landscape in India. By exploring network parameters and various data analysis techniques on Twitter Data specific to the Indian elections, we want to evaluate the social media strategies between members of the different major political formations in India for the elections. The methodology used to do this involves looking closely at both the direct role of the politician/formation as well as their influence and an approximation of their catalyzing role in the network. To quantitatively be able to compare these, we looked at network parameters such as the Contagion modeling of information via retweets, Network parameters such as clustering, diameter, pageRank, analysis of strength of followers, sentiment analysis based clustering of the network.

1. DATA

The dataset has been obtained by a combination of using Twitter’s public API and a dataset specific to Indian Elections obtained by the help of Jure Leskovec and Rok Sosic.

The data obtained represents a substantial portion of the Twitter network from India and contains 15.5 million Unique user accounts (The nodes in our network) and up to 5000 followers for each of these accounts (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 dataset are tweets pertaining to the elections and have been filtered on keywords. The entire set of keywords is listed in the Appendix. This dataset comprises a substantial portion of the ≈ 33 million unique Twitter users in India.

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-Politicians. 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.

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.
2. PRELIMINARY ANALYSIS OF GRAPH PROPERTIES

We initially computed the fraction of nodes contained in the largest weakly connected component (Wcc) of the graph. Unsurprisingly, we found that 99% of all nodes were a part of the largest Wcc. This means that almost everyone who discussed political ideas via twitter is connected to a political leader.

<table>
<thead>
<tr>
<th>Network Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
</tr>
<tr>
<td>Edges</td>
</tr>
<tr>
<td>Tweets</td>
</tr>
<tr>
<td>Politicians Tweets</td>
</tr>
<tr>
<td>Largest WCC</td>
</tr>
</tbody>
</table>

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 interesting 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.
3. OVERVIEW OF PAST 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 \(\theta_u\). 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.

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 4.

Direct analysis on the applicability of Twitter to elections has also been done, most notably by McKelvey, DiGrazia, and Rojas\(^2\). They perform a sentiment analysis of Tweets in a given district to estimate the Democratic or Republican voter share and find that that is a good predictor or a party’s electoral success. Our paper builds explores additional metrics and network attributes that may drive further research in this area.

4. MATHEMATICAL MODELS

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

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 \(f : V \times V \times N \rightarrow \mathbb{R}\) as

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

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

\[
\sum_{v \in \text{InNodes}(u)} \sum_{i=1}^{n_v} f(v, u, i)
\]

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

\[
\sum_{v \in \text{InNodes}(u)} \frac{1}{\text{InDeg}(u)^i}
\]

\[
= \sum_{v \in \text{InNodes}(u)} \frac{\text{InDeg}(u)^n_v - 1}{\text{InDeg}(u)^n_v (\text{InDeg}(u) - 1)}
\]

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

If we consider only tweets by the nodes in \(\text{InNodes}(u)\) 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.

5. ALGORITHMS

Due to the large scale of our dataset (\(\approx 7\)GB 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:
- \( \text{numTweets} = \{} \);
- \( \text{resultTemp} = \{} \);
- \( \text{result} = \{} \);
- \( \text{count} = \{} \);

For each node in array do
  - If node in count then
    - \( \text{count}[\text{node}] += 1 \);
  - Else
    - \( \text{count}[\text{node}] = 1 \);
  - End

For each follower in OutNodes(node) do
  - If follower in resultTemp then
    - \( \text{resultTemp}[\text{follower}] += \frac{1.0}{(\text{inDeg} \times \text{count}[\text{node}])} \);
  - Else
    - \( \text{resultTemp}[\text{follower}] = \frac{1.0}{(\text{inDeg} \times \text{count}[\text{node}])} \);
  - End
  - If node in resultTemp then
    - \( \text{result}[\text{node}] = \text{resultTemp}[\text{node}] \);
  - Else
    - \( \text{result}[\text{node}] = 0 \);
  - 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 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

6. 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 any other coalition, including all the others outside 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.
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.

7. 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 engagement 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 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. 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 behavior. 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 significant 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) supporting.

8. ANALYSIS OF RETWEET HOPS

As discussed earlier, retweets are a useful way to gauge how many people agree with the views of a particular po-
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.

9. ANALYSIS OF SENTIMENT THRESHOLD

![Threshold for Positive Sentiment by Party](image)

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 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 exposure 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.

10. 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 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

11. CITATIONS


3. "BJP may return to power in Rajasthan, Hung assembly in Delhi: India TV-CVoter projection." India TV. 18 September 2013.

12. APPENDIX

Below is a table of all political handles and their corresponding political affiliations. A handle not associated with the NDA, UPA, or AAP is classified as OTH for “other.”

<table>
<thead>
<tr>
<th>Political Handle</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>quizderek</td>
<td>OTH</td>
</tr>
<tr>
<td>KalrajMishra</td>
<td>NDA</td>
</tr>
<tr>
<td>M_Lekhi</td>
<td>NDA</td>
</tr>
<tr>
<td>PawarSpeaks</td>
<td>UPA</td>
</tr>
<tr>
<td>prithvraj</td>
<td>UPA</td>
</tr>
<tr>
<td>PMOIndia</td>
<td>OTH</td>
</tr>
<tr>
<td>AITCoofficial</td>
<td>OTH</td>
</tr>
<tr>
<td>meerasanyal</td>
<td>AAP</td>
</tr>
<tr>
<td>SudheenKulkarni</td>
<td>OTH</td>
</tr>
<tr>
<td>jayantsinha</td>
<td>NDA</td>
</tr>
<tr>
<td>GujPanag</td>
<td>AAP</td>
</tr>
<tr>
<td>Swamy39</td>
<td>NDA</td>
</tr>
<tr>
<td>Tarunvijay</td>
<td>NDA</td>
</tr>
<tr>
<td>KLDSingh_India</td>
<td>OTH</td>
</tr>
<tr>
<td>SushmaSwaraj</td>
<td>NDA</td>
</tr>
<tr>
<td>BJP4India</td>
<td>NDA</td>
</tr>
<tr>
<td>NCPspeaks</td>
<td>UPA</td>
</tr>
<tr>
<td>ThinkwithGoogle</td>
<td>OTH</td>
</tr>
<tr>
<td>narendramodi</td>
<td>NDA</td>
</tr>
<tr>
<td>INCIndia</td>
<td>UPA</td>
</tr>
<tr>
<td>asadowanasi</td>
<td>OTH</td>
</tr>
<tr>
<td>ncbn</td>
<td>NDA</td>
</tr>
<tr>
<td>_SanddeepDikshit</td>
<td>UPA</td>
</tr>
<tr>
<td>yadavakhilesh</td>
<td>UPA</td>
</tr>
<tr>
<td>PandaJay</td>
<td>OTH</td>
</tr>
<tr>
<td>dilipkpandey</td>
<td>AAP</td>
</tr>
<tr>
<td>ianuragthakur</td>
<td>NDA</td>
</tr>
<tr>
<td>smritiirani</td>
<td>NDA</td>
</tr>
<tr>
<td>JKNC_</td>
<td>UPA</td>
</tr>
<tr>
<td>shaziaiimi</td>
<td>AAP</td>
</tr>
<tr>
<td>KapilSibal</td>
<td>UPA</td>
</tr>
<tr>
<td>ShahnawazBJP</td>
<td>NDA</td>
</tr>
<tr>
<td>JP_LOKSATTA</td>
<td>OTH</td>
</tr>
<tr>
<td>jaluprasadrajd</td>
<td>UPA</td>
</tr>
<tr>
<td>buzzindelhi</td>
<td>OTH</td>
</tr>
<tr>
<td>PriyalDutt_INC</td>
<td>UPA</td>
</tr>
<tr>
<td>AnanthKumarBJP</td>
<td>NDA</td>
</tr>
<tr>
<td>BJP_RajnathSingh</td>
<td>NDA</td>
</tr>
<tr>
<td>ShainaNC</td>
<td>NDA</td>
</tr>
<tr>
<td>digvijaya_28</td>
<td>UPA</td>
</tr>
<tr>
<td>supriya_rule</td>
<td>UPA</td>
</tr>
<tr>
<td>sambatosh33</td>
<td>AAP</td>
</tr>
<tr>
<td>abdullah_omar</td>
<td>UPA</td>
</tr>
<tr>
<td>DrShahFaisal</td>
<td>OTH</td>
</tr>
<tr>
<td>DrKumarVishwas</td>
<td>AAP</td>
</tr>
<tr>
<td>ManishTewari</td>
<td>UPA</td>
</tr>
<tr>
<td>AUThackeray</td>
<td>NDA</td>
</tr>
<tr>
<td>KirronKherBJP</td>
<td>NDA</td>
</tr>
</tbody>
</table>
Political Handle | Affiliation
--- | ---
manoharparrikar | NDA
mayankgandhi04 | AAP
Pallamrajam | UPA
milinddeora | UPA
varungandhi80 | NDA
vasundharaBJP | NDA
rajeev_mp | OTH
DrJitendraSingh | NDA
ShashiTharoor | UPA
PrakashJavdekar | NDA
ChennaiConnect | OTH
SalmanSoz | UPA
AamAadmiParty | AAP
SinghRPN | UPA
usitharaman | NDA
priyankac19 | UPA
MPNaveenJindal | UPA
NKSingh_41 | NDA
ShuklaRajiv | UPA
ChouhanShivraj | NDA
ajaymaken | UPA
PiyushGoyal | NDA
harsimrat_badal | NDA
jkpd | OTH
ArvindKejriwal | AAP
AapYogendra | AAP
mkstalin | UPA
arunjaitley | NDA
naqvimukhtar | NDA
drharshvardhan | NDA
RamJethmalani5 | OTH
SushilModi | NDA

13. APPENDIX

\[
\frac{1}{x} + \frac{1}{x^2} + \ldots + \frac{1}{x^n} = \sum_{i=0}^{n} \frac{1}{x^i}
\]  \hspace{1cm} (6)

Recall the formula for an infinite geometric series

\[
\sum_{i=1}^{\infty} x^i = \frac{x}{1-x}
\]  \hspace{1cm} (7)

for \( x < 1 \).

So we have

\[
\sum_{i=1}^{\infty} \frac{1}{b^i} = \frac{\frac{1}{b}}{1 - \frac{1}{b}}
\]  \hspace{1cm} (8)

Thus,

\[
\frac{n}{b} = \sum_{i=1}^{n} \frac{1}{b^i} = \frac{\frac{1}{b} - \frac{1}{b^n}}{1 - \frac{1}{b}}
\]  \hspace{1cm} (9)

so

\[
\frac{1}{b^n} - \frac{1}{b} \frac{1}{b^n} = \frac{n}{b} \frac{1}{b^n}
\]  \hspace{1cm} (10)

Thus,

\[
\frac{n}{b^n} - \frac{1}{b} \frac{1}{b^n} = \frac{n}{b^n(b - 1)}
\]  \hspace{1cm} (11)

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

Keywords captured within Tweet Data

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