Political Alliance prediction

BroCoders

Anuj kumar, Nihar Shah, Pratik Kulkar, Preet Gandhi

Date: 19/04/21

Content

1 Introduction 3
1.1 Motivation 3
1.2 Abstract 3
2 Data and Methods 4
2.1 Preprocessing 4
2.2 Data Statistics 4
2.3 Network Visualization 5
2.4 Implementation 8
3 Innovation 11
3.1 Politician to Party alliance prediction 11
3.2 Party to Party alliance prediction 12
4 Results 13
5 References 14

1 Introduction

1.1 Motivation

Now a days Politicians from new and established parties worldwide have embraced social media. They use these platforms to explain their political views and connect with voters. Many studies have examined issues related to the uptake of social media by politicians. As anticipated the age of Internt has come. Every people are getting connected to Internet. There has been an exponential growth of data in this decade. For example back in 2008 Facebook had only 100 million users now it has more than 2.6 billion users, Twitter had only 2 million users and now it has more than 330 million users. This growth of online data has caught the attention of many reaserchers and it arose many research questions of many distributed fields. Also the studies shows that MPs' use of social media affects communications in parliaments [1]. This is very crucial information for this topic. Also in a big nation like India which has more than 60 parties, collaboration between parties is must to win the elections. That's why the problem of predicting the alliance is so popular right now yet there is no significant research in this area. Existing prediction for alliance formation are more biased from a perception of particular person or ideology of a group. For ex. Newspapers, Media etc. To address this particular problem we need an unbiased predictor which solely makes decision on validated and authenticated data.

1.2 Abstract

This report focuses on the study of various network parameters of Indian Politicians Twitter mention and retweet network. To visualize the complex relationships between huge amount of elements this report also includes the results of visualization of the network by changing different network parameters. Also later using this information of network parameters it predicts the alliance formation between two parties and also the alliance of an individual politician to any particular party with very

high accuracy (90.54%). The dataset consists of tweets of all the Indian politicians from the past two years and that includes three main politicaly debatable topics i.e Covid-19, Farmers bill and India China standoff. There are total 14019 tweets in the dataset.
Keywords : Political Polarization, Homophily, Eco chamber, E-I Index, Modularity, Reciprocity, Betweenness centrality.

2 Data and Methods

2.1 Preprocessing

First and foremost we need to clean the data, we need to preprocess the data such that it will be useful to visualize the network parameters as well as to predict the alliances. At first we separated out the tweets of three different topics from our main dataset using filtering on the hashtags. After separating out the three topics we removed the unnecessary data columns from our dataset. To do the same we have made a python script which takes the author name, the name of the politician which he has mentioned or retweeted and the sentiment of this tweet. Rest all the data is filtered out. We have used this file as a edges file for our network visualization tool gephi. Also the given politician-party file needed some cleaning. For that we have written a c++ code which gives us the csv file that contains the name of the politician and the party name of that politician. We used this file as a nodes file for gephi.

2.2 Data Statistics

The statistics of the data after the preprocessing step is following.

Topic	*Total Tweets*
Covid-19	11300
Farmers bill	2108
India-China standoff	125

The total tweets here represents the total number of tweets done by all the politicians combined on each topics. These tweets are later separated into two different networks namely mention network and retweet network. The statistics of the mention network for different topics is following.

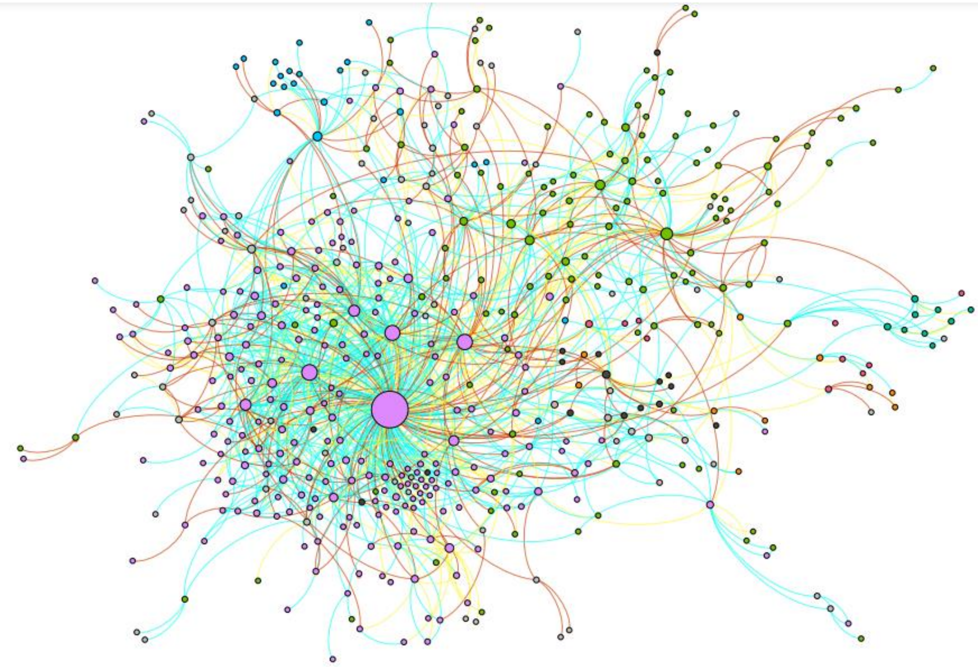
Topic	*Nodes*	*Edges*	*Unmerged Edges*	*Total Parties*
Covid-19	507	1314	4458	36
Farmers bill	194	316	978	15
India-China standoff	36	38	66	5

Here the edges represents the merged edges. Which means if there are two edges between same nodes then the weight of that edge is changed accordingly and both edges will be merged. The statistics of the retweet network for different topics is following.

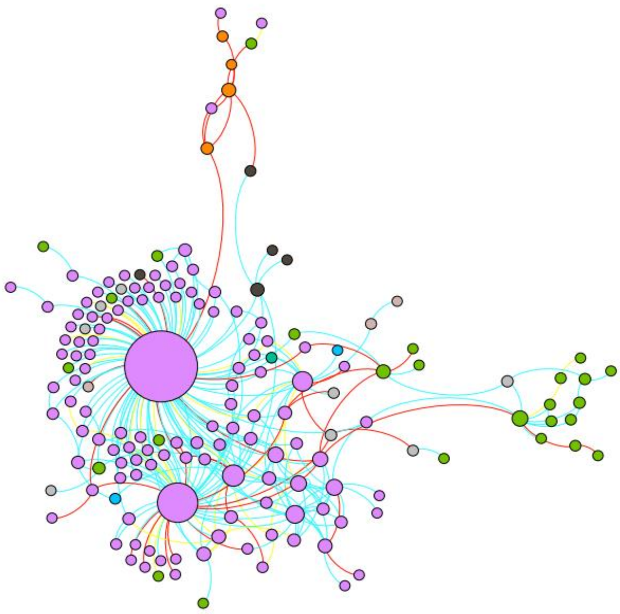
Topic	*Nodes*	*Edges*	*Unmerged Edges*	*Total Parties*
Covid-19	467	863	2159	26
Farmers bill	127	156	232	7
India-China standoff	16	15	16	3

2.3 Network Visualization

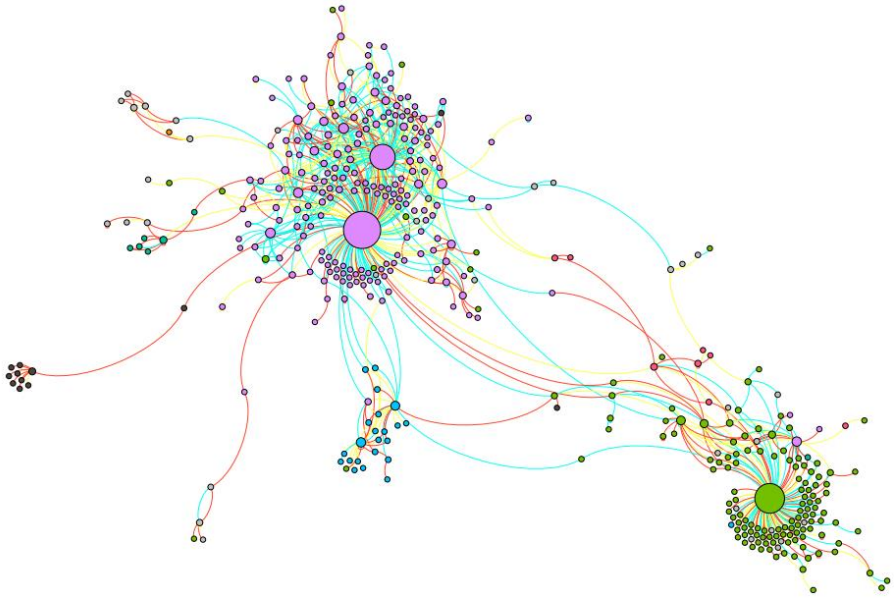
Now to understand the nature of communication between two parties and politicians we generated the network in gephi using the files we created in data cleaning step. we changed the size of the node according to their in-degree and colored the edges using sentiments to visualize the network clearly. The red edges represents the negative sentiment, blue represents the positive sentiment while yellow is for neutral sentiment. We observed that for the mention network of all the topics the political polarization is less. The distance of the politicians from the different parties is minute. Also from the graph we can see that there is no formation of echo chamber. The intercommunication between parties is high. Later we will show this using code aswell but from the graph we can visualize it clearly. The following is the Covid-19 mention network.



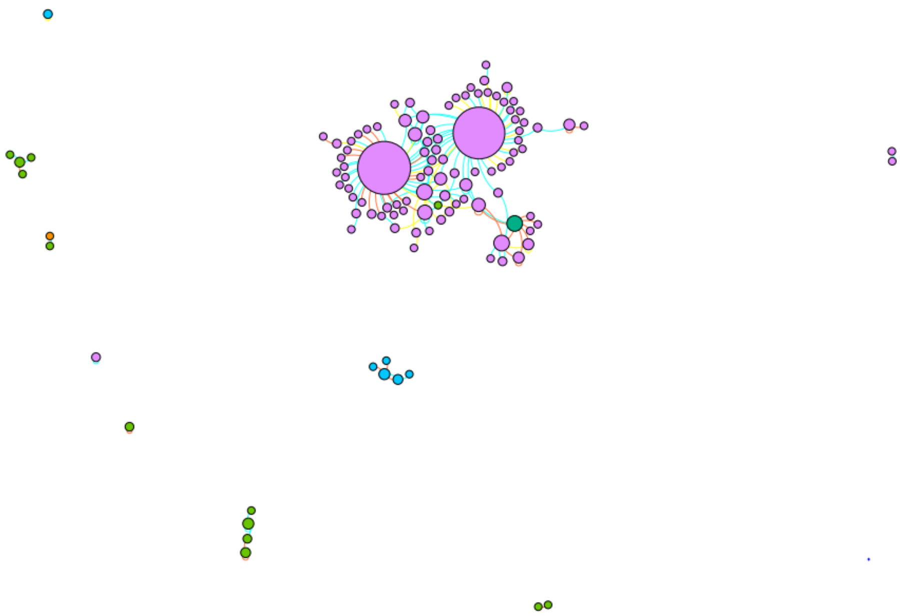
The following graph is for Farmers bill mention network



The similar results are happening here as well. Contrary to mention network, In retweet network we can clearly see that the phenomenon of echo chamber is occurring here. The politicians tends to retweet the party in which they belong and this also does make sense because the retweet is basically the token of agreement. The following graph is for Covid-19 retweet network.



The same results are seen for Farmers bill retweet network as well.



Few more network parameters like modularity, density and reciprocity are also found using this tool. The statistics of the same for the mention network is following.

Topic	*Modularity*	*Density*	*Reciprocity*
Covid-19	0.524	0.06	0.045
Farmers bill	0.338	0.011	0.031
India-China standoff	0.514	0.041	0

The statistics of the retweet network is following.

Topic	*Modularity*	*Density*	*Reciprocity*
Covid-19	0.524	0.06	0.045
Farmers bill	0.338	0.011	0.031
India-China standoff	0.514	0.041	0

2.4 Implementation

To check the visual results of political polarization in mention and retweet network we have found external-internal index i.e E-I index across and between two parties using c++ code. The E-I index of a party is defined by (total outgoing edges from a particular party – total internal edges of that party) / total edges of that party. E-I index can range between -1 and 1. The low value of E-I index shows that there exists an echo chamber effect in the network. The high value of E-I index shows that there is a lot of interparty communication going on and it refutes the claim of political polarization. The results of across party E-I index and between party E-I index for both mention and retweet network for all the topics are following.

Following is the E-I index statistics of the Covid-19 mention network.

	BJP	*Congress*	*AAP*	*AITC*	*CPI(M)*
BJP	-	-0.958	-0.983	-0.988	-1
Congress	-0.092	-	-0.943	-0.982	-0.949
AAP	-0.593	-0.889	-	-1	-1

AITC	-0.272	-0.842	-1	-	-1
CPI(M)	0.467	-1	-1	-0.334	-

E-I index statistics of the Farmers bill mention network.

	BJP	*Congress*	*AAP*	*AITC*	*Akali Dal*
BJP	-	-0.971	-0.997	-1	-1
Congress	-0.275	-	-1	-1	-1
AAP	-0.583	-1	-	-1	-1
AITC	0.23	-1	-1	-	-1
Akali Dal	-0.23	-0.454	-1	-0.778	-

E-I index statistics of the India-China standoff mention network.

	BJP	*Congres s*	*AITC*	*NCP*	*DMK*
BJP	-	-0.885	-1	-1	-1
Congres s	-0.25	-	-1	-1	-1
AITC	-0.5	-1	-	-1	-1
NCP	-1	-1	-1	-	-1
DMK	-1	-1	-1	-1	-

E-I index statistics of the Covid-19 retweet network.

	BJP	*Congress*	*AAP*	*AITC*	*CPI(M)*
BJP	-	-0.950	-0.987	-1	-1

Congress	-0.706	-	-0.989	-0.994	-1
AAP	-0.787	-0.843	-	-1	-1
AITC	-0.667	-1	-1	-	-1
CPI(M)	-1	-1	-1	-1	-

E-I index statistics of the Farmers bill retweet network.

	BJP	*Congress*	*AAP*	*RJD*	*RLP*
BJP	-	-0.99	-1	-1	-1
Congress	-0.882	-	-1	-1	-1
AAP	-1	-1	-	-1	-1
RJD	-1	0	-1	-	-1
RLP	-1	-1	-1	-1	-

E-I index statistics of the India-China standoff retweet network.

	BJP	*Congress*	*AIMIM*
BJP	-	-1	-1
Congress	-0.5	-	-1
AIMIM	0	-1	-

Analysis of average sentiment is also necessary because it represents the ideology of a particular party for a retweet network and it is useful feature to predict the alliances between two parties and also the alliance of a politician to a particular party. Following are the results of average sentiment of different parties for both mention and retweet network of all the three topics.

Covid-19 mention Average sentiment:

BJP : 0.5527, Congress : 0.24, AAP : 0.1285, AITC : -0.4313, CPI(M) : 0.475

Farmers bill mention Average sentiment:

BJP : 0.7578, Congress : 0.6562, AAP : 0.8333, AITC : 0.6153, Akali Dal : -0.6470

India-China standoff Average sentiment:

BJP : -0.0285, Congress : -0.0416, AITC : 0, NCP : -1, DMK : 1

Covid-19 retweet Average sentiment:

BJP : 0.1824, Congress : 0.0729, AAP : -0.1571, AITC : -0.6470, CPI(M) : -0.3888

Farmers bill retweet Average sentiment:

BJP : 0.4168, Congress : 0.2941, AAP : -0.2, RJD : -0.3333, RLP : 0

India-China standoff retweet Average sentiment

BJP : -0.3333, Congress : 0.2857, AIMIM : 0

3 Innovation

3.1 Politician to Party alliance prediction

$Idea$: For each party we are considering top K influential politicians from retweet network, currently influential nodes are found based on weighted indegree of politician. We calculate average sentiment of each politician participating from retweet network. Now, based on sentiment and interaction between the party we are predicting the party for current politician. Basically, for each party we are checking if current politician has ever interacted with the party or not, if yes then we find similarity based on sentiment between them. Thus, for a politician we will collect similarity with all influential politicians. Now, we check top T politicians who has maximum similarity with current politician and based on top T politicians we will check their parties and whichever party leads we will declare that current politician should belong to the party.

Example: K = 2, Current Politician P has average sentiment: 2.7. P2 and P3 are politicians belonging to Party A and their corresponding average sentiment 2.6 and 2.8. P4 and P5 are politicians belonging to Party B and their corresponding average sentiment 3.4 and 3.1. Considering T as 3, we can clearly see that P2, P3 and P4 are competing candidates but max time party coming is Party A so Party A is predicted.

Hypermaters: T and K, Similarity is found using mean squared error.

Accuracy: 90.54% with K=7 and T=7.

3.2 Party to Party alliance prediction

First thing that we came up with is to generate a vector which will help us compare the politicians' similarity. So, we have used DeepWalk technique on retweet network to generate a walk of length 4 and then we have generated several walks. Now, since our corpus of walks have been generated we have taken a window of 3 on a walk of length 4 out of which first 2 are inputs and last word is expected output. Once we had done this we have created a code which generated node embeddings i.e generated embedding for each politician. After this we have appended average sentiments of politician towards each party at the end of its corresponding embedding. There are two idea we have came up with to predict party alliance and they are as follows

Clustering Politicians: Since we have generated node embeddings of each politicians then we will apply a clustering algorithm to create a clusters but we will create n_clusters which is a hyperparameter and signifies the number of clusters. In our case we have use K-Means clustering to find the 5 clusters. Once we have clusters we will find the number of points belonging to each party for each cluster and then find the top 2 parties for each cluster, store them in the set and then declare the parties which are forming an alliance. For example, suppose we have 2 clusters and (BJP, Shivsena),(Congress,AAP) have the highest counts in corresponding clusters so we will declare them as an alliance.

Politician Similarity: In this method we will first compute the similarity between each politician pair. Here, we can use any kind of similarity and in our project we have used cosine similarity. Then we have parameter sim which signifies a limit upto which two node embeddings or politicians has to be similar. Now, if two politicians have their similarity >= sim then check if they belong to two different parties. If they do belong to different parties then increase the number of cross party interaction for the two parties to which the two parties belong. For example, if P1 belongs to BJP and P2 belongs to Congress and their cosine_sim(P1,P2) >= sim then cross_party_interaction[BJP,Congress]++. Now, if cross party interaction >= T then declare that those parties are going to form an alliance.

Intuition: Politicians will have the same ideology and will tend to form a group and they will interact with each other upto certain threshold because less interactions signifies that they have less chance of forming an alliance.

Link: https://jovian.ai/kulkarpratik/sm-project

4 Results

For Politician to party prediction with hyperparameter K=7 and T=7 accuracy achieved is 90.54%. Some of the politicians our model predicted correctly according to their current party but failed on the dataset. Which shows the power of this method, using this method we can surely predict the alliance accurately.

Party of politician: ashokgehlot51 as per data is: BJP and predicted is: Congress. Party for politician: msisodia as per data is: BJP and predicted is: AAP. Party for politician: misabharti as per data is: RJD and predicted is: Congress. And there were chances of RJD and Congress making alliance during Bihar elections. These are some of data we were able to found, there might be more.

For Party to Party prediction we didn't have enough data to compare our results with. Hence we just printed out our prediction. But if well annotated data is available our model should work properly.

5 References

1. Esteve Del Valle, Marc et al. "Political Interaction Beyond Party Lines: Communication Ties And Party Polarization In Parliamentary Twitter Networks". *Social Science Computer Review*, 2021, p. 089443932098756. *SAGE Publications*, doi:10.1177/0894439320987569.
2. Sluban, B., Smailović, J., Battiston, S., & Mozetič, I. (2015). Sentiment leaning of influential communities in social networks. *Computational Social Networks, 2*(1). doi:10.1186/s40649-015-0016-5
3. Alamsyah, A., & Adityawarman, F. (2017). Hybrid sentiment and network analysis of social opinion polarization. *2017 5th International Conference on Information and Communication Technology (ICoIC7).* doi:10.1109/icoict.2017.8074650