Analysis of Political Leanings of News Organisations in India

Rishab Ketan Doshi*, Madhavi Srinivasan†, Akash S Kumar‡ and G Ram Mohana Reddy§
Information Technology Department, National Institute of Technology Karnataka Surathkal, Mangalore - 575025, India
Email: *rishabketandoshi@gmail.com, †madhavi94@gmail.com, ‡akashskumar@gmail.com, §profgrmreddy@nitk.ac.in

Abstract—**Most of the News and Media organizations in the world have some direct or indirect political affiliations. Hence the content generated by these organizations is biased towards the respective ideologies of the political affiliations that the news organization is affiliated with. In this paper we aim to find the political leanings of few of the leading newspapers in India. This holds relevance as India is the worlds largest democracy with more than 1700 political parties and also the largest market for newspapers in the world with over 100 million copies sold everyday. A hybrid approach which involves the integration of Sentiment Analysis and Social graph concepts to determine the bias is proposed in this paper.**

Index Terms—**Social Graph, Social Network Analysis, Bias Estimation, News Analysis, Sentiment Analysis**

I. Introduction

In today's world the content being generated by news and media organizations has wide reaching effects as they are not only restricted to print and television media but they also have an online presence in the form of websites, Facebook Pages, Twitter Handles etc. With around 3.17 billion people on the internet of which around 1.19 billion use Facebook and 651 million use Twitter, the content generated by these news organizations has wide reaching effects and plays a major role in shaping public opinion.

Since the news and media organizations have the potential to shape public opinion, they attract political parties and political entities to tap into this potential and use it for their own propaganda. So these political entities influence the news and media organizations to generate content that will popularize their respective propaganda and shape public opinion to their favor. Hence we see that the news content being generated by different organizations is increasingly biased towards the respective political party / political ideology / political entity that the news organization is either directly or indirectly affiliated to.

In addition to being the largest and most diverse democracy, India has a number of political parties with their own political ideologies. Such a varied political landscape, makes it even more challenging to determine bias.

Table I gives us a comparison of some of the previous approaches used for bias estimation of news organizations.

TABLE I
Comparison of Bias Estimation Methods

Work	*Advantages*	*Limitations*
Mahmood et al. [3]	Used number of graph concepts like betweenness, clustering coefficient, closeness	Lacks semantic understanding
Sonal Gupta [4]	Prediction accuracy more than supervised algorithms	Restricted to only memes on news websites in US

Political orientation of news organisations has been in existence in a variety of forms. An apt illustration for this would be the joint study done by Harvard university and Project for Excellence in Journalism where specific analysis revealed that CNN was more liberal than Fox news channel[1].

The presence of bias in content generated by news and media organizations has always been known. In the past, research work has been carried out to find the political leanings of news organizations and websites in USA. Social Network Analysis (SNA) is an efficient way to discover the implicit relations by modelling information as a network [2]. A number of graph methods have been proposed to quantify bias of news organisations. For example, Mahmood et al. [3] validated the known political leanings of two newspapers in New York by performing an analysis on the entities present in the newspaper articles. The nodes in the graph were restricted only to politicians and hence the insights generated were limited. Sonal Gupta [4] attempted to find the bias of political news and blog websites by examining the memes(political phrases) that appear on the website. A bipartite graph was created with websites and memes' as nodes and an approach similar to Hubs and Authorities approach [5] was used to weigh the edges in the graph and hence determine the bias. This work is restricted only to memes appearing on the news websites and is carried out again for USA specifically. A similar meme analysis was carried out by Jure et al. [6] where memes are viewed as short units of text that act as signatures of topics and events that create buzz in and around media. They sculpted a framework that tracks the traversal of memes and eventually clusters their textual variants across news articles. A directed acyclic graph called the Phrase Graph is constructed in which each node represents the multiple variants of quotes that appear in news articles and there exists an edge from the shorter phrase to the longer phrase if the shorter phrase has a certain threshold number of overlapping words with the longer phrase. Consequently, a Directed Acyclic Graph partitioning algorithm is applied to discover phrase clusters. Each of the phrase clusters are then tied to a series of news articles and blog posts that contain any item from the phrase cluster. The approach reflects a hard bound on the parameters used for constructing the graph.

Communication patterns within a classroom can also be designed as a network to reveal interesting insights. Infact, Saltz et al. [7] undertook a similar student-centered analysis within an online classroom. Each student is mapped to a node and the interactions between the students and instructors are encoded as directed edges. Graph metrics of indegree, outdegree and degree centrality are used to understand the student's involvement within the class. However, there has been no use of advanced SNA concepts like betweenness centrality, closeness centrality etc. Wilson et al. [8] carried out SNA by using the graph metrics of social degree, clustering coefficient along with user interaction analysis to discover the implicit meaning of relations within interaction graphs. The research in this paper helps to validate the presence of social links as an indicator of real user interaction. An interesting research carried out specifically for blogs was done by Adamie and Glance [9] , where they presented analysis on the political blogosphere. Popular political blogs in the US were modelled as a citation network with each blog having either the liberal or conservative flavour. This paper presented a set of compulsive patterns to look for in blogs and news articles with specific political orientation. Mori et al. [10] suggested an interesting method to derive relationships for a social network. Given the superficial relations between different entity pairs belonging to the same social network, this paper aimed to determine the underlying relations between each pair of entities such that all the relationships confirm to the same collective context. One of the key things about this approach is the fact that it is entirely unsupervised and can be easily incorporated into any social network domain. Park et al. [11] proposed a novel social annotation analysis approach to uncover bias of news articles by using the sentiment patterns of commenters'. The effectiveness of commenters' with strong political inclination is then leveraged to quantify bias.

There have been many studies regarding the measurement of political polarization of news organisations on the social media. Wong et. al [12] proposed a novel method to quantify political orientation from tweet and retweet pattern analysis. They based their hypothesis on the fact that tweeters tend to follow and retweet those who share similar political views. While Wong et al. [7]'s research followed an event based method of data collection, Kokil and Saifudeen [13] collected tweets posted by genuine accounts of major political parties. They carried out extensive analysis on these tweets for a two month period just before the general elections in India to gauge the extent to which political parties exploit social media platforms for their campaigning.

To get a numerical estimation of the leverage of media by political organisations in the Indian context is the main motivation of our proposed work. A compound approach of sentiment analysis and social graph is applied to get the same.

The core idea behind our method is to numerically estimate the implicit bias of news organisations by using a social graph approach. We understand that the entities in the articles have a natural inclination to form a graph, and we use this property along with the graph edge weight assignment using the sentiment scores churned out by the API, to come up with a hybrid framework for bias determination.

Some key points about the proposed work are:

- To the best of our knowledge, this is the first paper that employs a hybrid approach of sentiment analysis and social graph in the determination of political bias, unlike other approaches that use just one of the two.
- To the best of our knowledge, we are the first to carry out research on implicit political bias present in news organisations in India.

The rest of the paper is organized as follows: Section II describes a step wise process of the proposed approach. Section III validates the obtained results with the real time scenario. Section IV concludes our paper with the calculated bias for each target news organisation and discusses the scope for future enhancements of the proposed work.

II. Proposed Work

The proposed system uses data gathered from the news and media websites. The gathered data is persisted into a NoSQL Database as a buffer. The buffered data is then processed using Alchemy API [15] to find the entities and the sentiment associated with each of the articles. These are then plotted on a social graph and graph metrics [2] are applied on the nodes to determine bias.

The high level overview of the proposed framework is shown in Figure 1.

Data Collection:

The following news and media organizations were identified for the study. Articles from these organization were manually crawled from the following news archives and used for the examination of bias.

1) Hindustan Times
2) The Hindu
3) The Daily Pioneer
4) The Indian Express
5) The Times of India

Identification of Political Events

Political events which were controversial with respect to the two major political parties in India, Indian National Congress(INC) and Bharatiya Janata Party(BJP) were identified and articles were gathered based on these events(ex:Coal-Gate scam,riots...). Since these events qualify as news and also are controversial with respect to some political party or other, it reflects the leanings of the news organization. In addition to this, all the political articles appearing in the respective paper belonging to a specific period of time were also collected. A good mixture of events and articles from both the ruling periods of UPA(2009-2014) and NDA(2014-present)

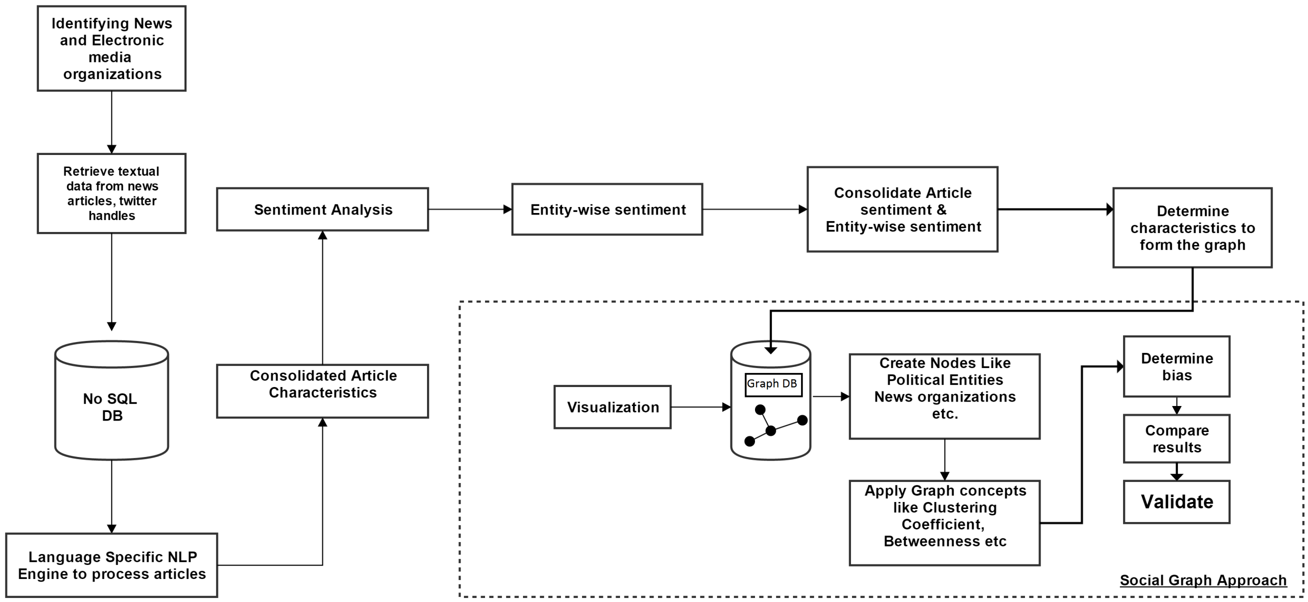

Fig. 1. System Architecture of the Proposed Work

governments is taken to account for the uniform distribution of opinions. Totally, around 300 articles specific to chosen events were collected across these 5 newspapers.

Algorithm 1: Pseudo code for Consolidating Article Characteristics

```
while articles to be processed do
    a=cleanArticle();
    entities=findEntities(a);
    articleSentiment=findSentiment(a);
    entitySentiment=findEntityWiseSentiment(a);
    while es in entitySentiment do
        pe=findPoliticalEntity(es);
        pp=findPoliticalParty(pe) ;
        res=mapPoliticalSentiment(pe,pp,es) ;
        articleChars.append(res);
    end
end
```

Entity Mapping,Sentiment Analysis

In the next set of steps we identify the entities in the article and gauge the sentiment associated with every article. To achieve this we use IBM Watson's AlchemyAPI [15]. Alchemy language is a web based REST-API which provides us with entities, sentiment scores and entity wise sentiments for a given article. Algorithm 1 explains the steps applied to obtain the consolidated article characteristics. Every article is passed through a data cleansing station which performs basic functions like removing dirty URL's, dirty characters. These articles are then processed by Alchemy's Language API to give us entities and sentiment associated with each of the entities. The entities obtained in the previous steps are now mapped to an existing list of politicians and political parties. The entities for which a match is not found in the political list are discarded. These entities along with the political party and sentiment are then appended to a list which then constitute the consolidated article characteristics.

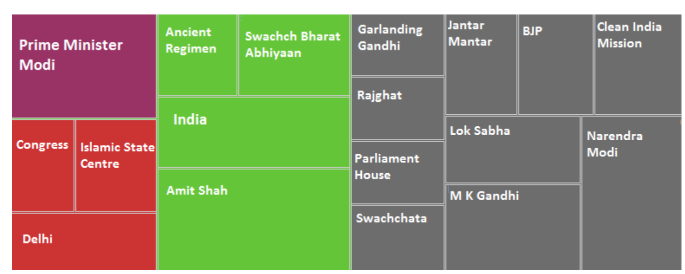

Fig 2. The Entity Wise sentiment obtained from AlchemyAPI for a given article

Figure 2 shows the entity wise sentiments as received from Alchemy Language API [15]. The entities in :

- Gray have a neutral sentiment associated with them.
- Green have a positive sentiment associated with them.
- Red have a negative sentiment associated with them.
- Purple have a mixed sentiment associated with them.

```
{
  "news-org":"The Daily Pioneer",
  "article":"2",
  "text_entities" : [
        {
            "mapped_party" : "BJP",
            "alch_sentiment" : "positive",
            "mapped_entity" : "Narendra Modi",
            "alch_sentiment_score" : "0.6"
        },
        {
            "mapped_party" : "BJP",
            "alch_sentiment" : "positive",
            "mapped_entity" : "Rajnath Singh",
            "alch_sentiment_score" : "0.13"
        }
    ]
}
```


Fig 3. Sample of Consolidated Article Characteristics for an article appearing in The Daily Pioneer

Building the graph

A graph is built for every news organization by first constructing the news organization node. A node for every article is then created and linked to the news organization node that it belongs to. To ensure equal contribution from all articles in a given newspaper network, equal weights are assigned to edges connecting the newspaper node to the article node. The entities in the article which appear in the consolidated article characteristics are then created as nodes and linked to the article nodes created in the previous step. For similar reasons, the weight assigned to the edge from the article node to the entity node is split equally among all entities appearing in the article. These entities are also mapped to the political party and a link is made from the entity nodes to one of the following party-sentiment nodes : *Pro-INC, Pro-BJP, Anti-INC, Anti-BJP, Neutral* with a weight equal to the sentiment score received from the Alchemy Language API [15]. To the constructed graph, graph metrics are applied on the party sentiment nodes mentioned above.

Figure 3 shows a sample of consolidated article characteristics that was obtained for an article that appeared in The Daily Pioneer. Let us illustrate the construction of nodes for this sample. A node is constructed for article-2 and an edge is created from 'The Daily Pioneer' node to the article-2 node. 'Narendra Modi'(NM) and 'Rajnath Singh'(RS) nodes are now constructed if they don't already exist in the graph. Links are made from article-2 node to NM and RS nodes, with the weight of each edge being 0.5 ($\frac{1}{no.of entities}$). Further we create an edge from NM and RS nodes to the Pro-BJP node. This is because the mapped party is BJP and the sentiment is positive. The weights assigned to these edges are +0.6 and +0.13(the sentiment scores) respectively. This process is repeated for all the articles appearing in The Daily Pioneer. At the end of the process we have a multi level graph with different kinds of nodes at every level. The graphs were built using Python-igraph[17], an agglomeration of network and graph analysis tools in python.

Graph Metrics

The below graph metrics are applied on the graph constructed for every news organization.

Degree: Among the given nodes higher degree to the party-sentiment nodes represents more connections being made to a particular party sentiment, reflecting bias of the news organization towards a particular party-sentiment [14].

Betweenness: Total number of shortest paths passing through a particular node. Applying this metric on the entity nodes gives us the most important political entities with respect to news organization network [2],[14].

Closeness: Defined as the inverse of farness, which is the sum of distances to all nodes. Applying this centrality measure on the party sentiment nodes reflects the closeness of the party sentiment nodes to other nodes in the network. A large value of closeness centrality of a particular party-sentiment node reflects a bias towards that party and its associated sentiment [14].

It may be argued that the news reported by the news organizations may be genuine facts and not represent bias. But since the articles are taken across the same time period and for the same set of political events an aggregation of sentiments by using graph metrics will depict bias as news organizations may publish more news that aligns with their bias. For example: A Pro-INC newspaper will publish more articles about an event which has a positive sentiment about INC as compared to a NEUTRAL newspaper.

III. Results and Discussion

Networks for all the news organization are constructed in the above manner and graph metrics [14] like Degree, Closeness and Betweenness are applied on these networks. Results have been discussed below.

Figure 4 shows closeness centrality values of major political entities with respect to every news paper. As discussed above high Closeness centrality values indicate influential positions in the network. We notice that Narendra Modi has the highest closeness centrality values and is followed by Rahul Gandhi, Sonia Gandhi and Rajnath Singh. These values reflect the relative importance of each of these nodes in all the newspapers. High value for Narendra Modi can be attributed to the fact that he is an influential leader of the ruling party and that he is vocal about all of his policy making decisions and active on all forms of media.

Figure 5 shows the degree of party-sentiment nodes in each of the newspaper network. We can clearly see that the degree of the Neutral Node is mostly high in almost all of the papers. Also PRO-BJP node has a high degree in The Daily Pioneer Network and PRO-INC node has a high degree in Hindustan Times network indicating a bias towards the respective party sentiment.

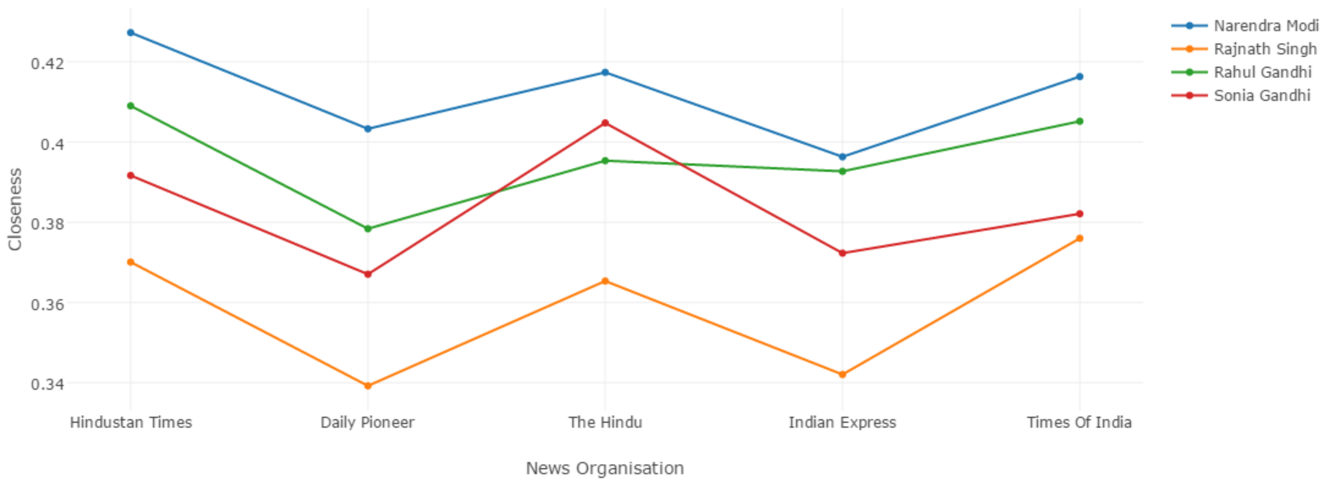

Fig 4. Closeness Centrality of major political entities with respect to every newspaper

TABLE II
CLOSENESS CENTRALITY VALUES FOR EACH OF THE PARTY SENTIMENT NODES IN THE NEWS ORGANIZATION NETWORK

News Organization	PRO-BJP	PRO-INC	ANTI-BJP	ANTI-INC	NEUTRAL
Hindustan Times	0.36	0.63	0.55	0.37	0.47
The Daily Pioneer	0.52	0.33	0.37	0.45	0.45
The Hindu	0.38	0.34	0.38	0.35	0.53
The Indian Express	0.35	0.34	0.36	0.34	0.47
Times Of India	0.4	0.39	0.41	0.38	0.56

Table II shows the closeness values as obtained for each of the party sentiment nodes in the individual networks. On comparing the values for each of the news organization we notice that the values lie in in the range 0.3-0.65. The PRO-INC, ANTI-BJP nodes in the Hindustan Times network have a relatively high closeness centrality value as compared to other newspaper networks and other party sentiment nodes in the same network and so does the PRO-BJP node in The Daily Pioneer. This establishes the fact that Hindustan Times has a slight PRO-INC bias and The Daily Pioneer has a slight PRO-BJP bias. In the rest of the newspaper networks, all the party sentiment nodes have similar values and are lesser than the NEUTRAL node. This establishes the fact that these newspapers are neutral with respect to INC and BJP.

Table III shows the betweenness values as obtained for each of the party sentiment nodes in the individual networks. We notice that PRO-INC and ANTI-BJP nodes have relatively high values in the Hindustan Times network and so does the PRO-BJP node in The Daily Pioneer. In the rest of the newspaper networks the NEUTRAL node has a much higher value than other party sentiment nodes, establishing the fact that these newspaper networks are neutral with respect to INC and BJP. These observations concur with those made by inspecting the closeness centrality values in Table II.

Figures 6, 7 and 8 show the network visualizations of The Daily Pioneer, Hindustan Times and The Indian Express respectively.

The node colors in these figures signify the below party sentiment:

- Yellow : PRO-INC
- Purple : ANTI-INC
- Green : PRO-BJP
- Red : ANTI-BJP
- Blue : NEUTRAL

A node is marked into one of the above colors. Depending on the political entity and the party it is mapped to and the sentiment associated with the particular entity.

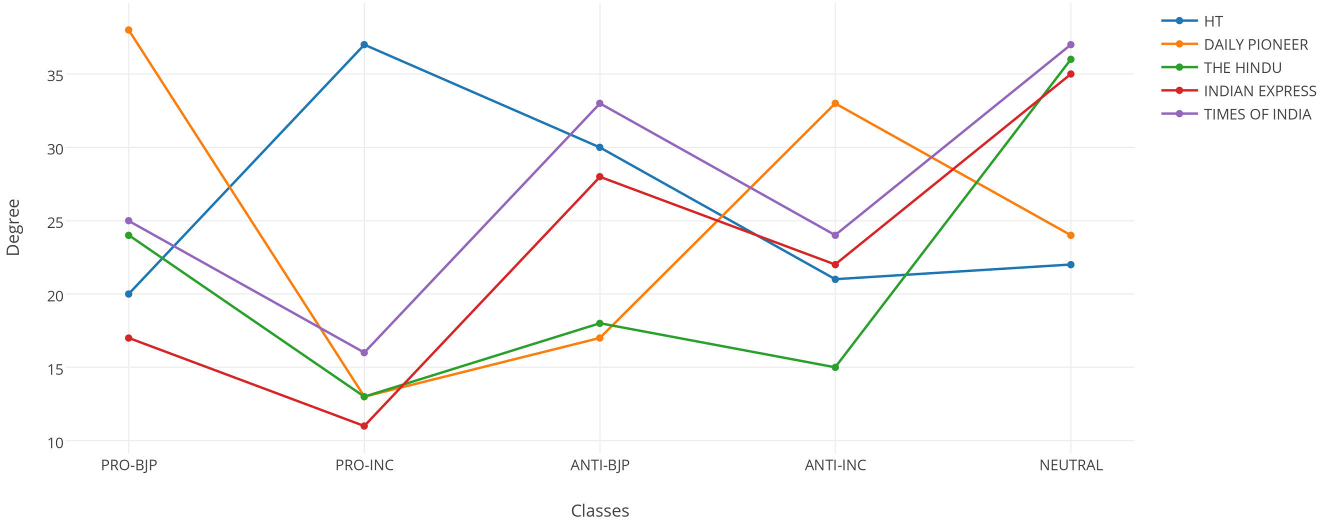

Fig 5. Degree of party sentiment nodes with respect to every newspaper

TABLE III
BETWEENNESS VALUES FOR EACH OF THE PARTY SENTIMENT NODES IN THE NEWS ORGANIZATION NETWORK

News Organization	PRO-BJP	PRO-INC	ANTI-BJP	ANTI-INC	NEUTRAL
Hindustan Times	159.65	532.28	426.74	164.2	526.9
The Daily Pioneer	556.145	105.55	79.29	376.28	436.47
The Hindu	93.95	45.63	22.05	59.6	1524.28
The Indian Express	103.3	59.47	174.52	124.98	926.507
Times Of India	145.32	78.32	212.43	149.76	1701.17

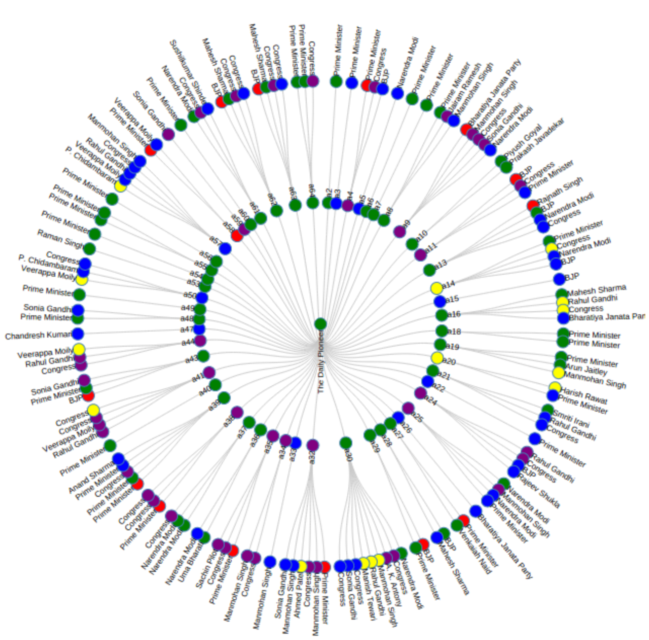

Fig 6. Network Visualization of The Daily Pioneer

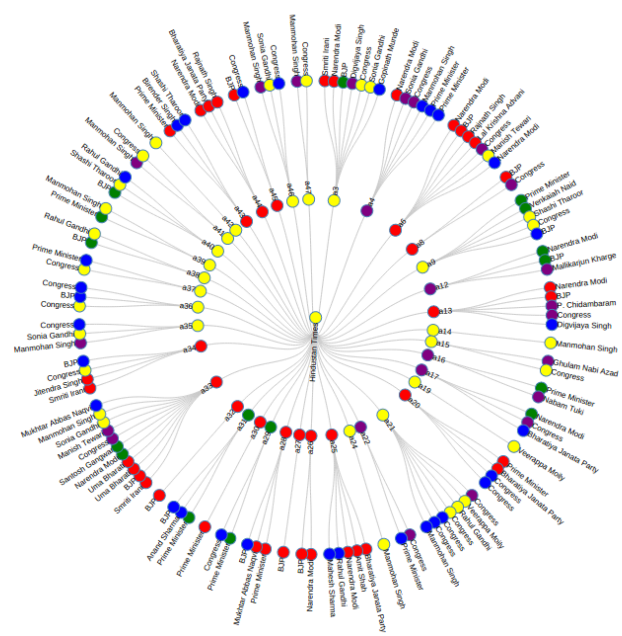

Fig 7. Network Visualization of Hindustan Times

The network visualization of The Daily Pioneer(Fig 6) shows the large number of green nodes indicating a large number of articles having a PRO-BJP sentiment. The large number of yellow and red nodes in the network visualization of Hindustan Times(Fig 7) indicate a PRO-INC, ANTI-BJP bias. The network visualization of The Indian Express(Fig 8) has a mix of all type of party sentiment nodes signifying that The Indian Express is neutral with respect to INC and BJP.

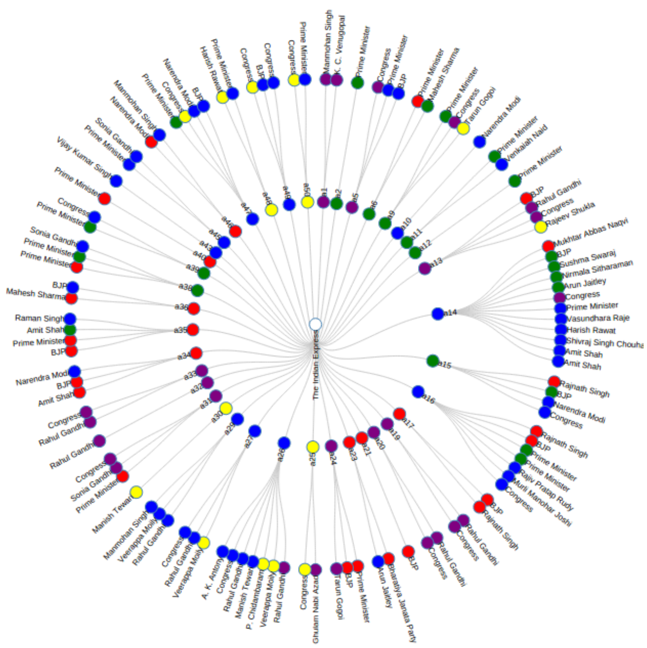

Fig 8. Network Visualization of The Indian Express

IV. Conclusion and Future Work

In this paper we have proposed a hybrid approach to estimate bias. In our work, news articles were gathered based on politically controversial events and sentiment analysis was applied on the collected articles. Further entity-wise sentiments were collected using Watsons Alchemy Language API and hence the sentiments towards political parties were found. To get a clear picture of the bias of the news organizations and to aggregate the obtained sentiments, social graph concepts were applied on the obtained sentiments. We conclude that The Daily Pioneer has a slight PRO-BJP bias, Hindustan Times has a slight PRO-INC bias and the rest of the news organizations were found to be NEUTRAL with respect to BJP and INC. We observe that these results are consistent with the political bias the news organization are known to have. Having performed our bias analysis on two major political parties, we now look forward to working with the analysis of more political parties which will provide us with bias determination of greater accuracy and a distinct political party to news organisation influence determination model. Currently, we work with data only from the websites of the respective news organizations i.e. the news articles they generate on a daily basis which is similar to the printed digests. News organisations nowadays publish their content in different forms over social networking websites (Facebook, Google Plus...) and also microblogging websites like Twitter. Including these sources for bias determination can be a part of the future work. Finally, more relevant insights can be obtained by creating more kinds of nodes(Political Ideology...) in the graph and applying social graph concepts on the same.

References

[1] DellaVigna, Stefano, Kaplan, Ethan. 2007. The Fox News effect: Media bias and voting. Quarterly Journal of Economics, August.

[2] R. Aviv, Z. Erlich, G. Ravid, and A. Geva, Network analysis of knowledge construction in asynchronous learning networks, Journal of Asynchronous Learning Networks, vol. 7, no. 3, pp. 123, 2003.

[3] B. Mahmood and R. Menezes. United States congress relations according to liberal and conservative newspapers. 2013 IEEE 2nd Network Science Workshop (NSW), pages 98-101, 2013.

[4] Sonal Gupta. Finding bias in political news and blog websites. http://snap.stanford.edu/class/cs224w-2010/proj2009/report Sonal Gupta.pdf, 2009.

[5] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. J. ACM, 46(5), 1999.

[6] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In Proc. of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009.

[7] Saltz, J., S. R. Hiltz, and M. Turoff. Student Social Graphs: Visualizing a Students Online Social Network. Proceedings, CSCW 04, 2004.

[8] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In Proc. of the 4th ACM European conference on Computer systems. ACM, 2009.

[9] Lada Adamic and Natalie Glance. The political blogosphere and the 2004 u.s. election: Divided they blog. In In LinkKDD 05: Proceedings of the 3rd international workshop on Link discovery, 2005.

[10] Mori, J., Tsujishita, T., Matsuo, Y., Ishizuka, M.: Extracting relations in social networks from the Web using similarity between collective contexts. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 487500. Springer, Heidelberg (2006)

[11] Park, S.; Ko, M.; Kim, J.; Liu, Y.; and Song, J. 2011. The politics of comments: Predicting political orientation of news stories with commenters sentiment patterns. In Proc. CSCW.

[12] Felix Ming Fai Wong, Chee Wei Tan, Soumya Sen, and Mung Chiang. 2013. Quantifying political leaning from tweets and retweets. In Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM).

[13] Kokil Jaidka and Saifuddin Ahmed. The 2014 Indian general election on Twitter: an analysis of changing political traditions. ICTD '15 Proceedings of the Seventh International Conference on Information and Communication Technologies and Development ,2015.

[14] M. Newman, Who is the best connected scientist? a study of scientific coauthorship networks, Complex networks, pp. 337370, 2004.

[15] AlchemyApi-For sentiment analysis of news articles http://www.alchemyapi.com/

[16] Plotly Technologies Inc. Collaborative data science. Montral, QC, 2015. https://plot.ly.

[17] Csardi G, Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. http://igraph.org

[18] www.hindustantimes.com/

[19] http://timesofindia.indiatimes.com/

[20] http://indianexpress.com/

[21] http://www.thehindu.com/

[22] http://www.dailypioneer.com/