

Dynamic Analysis of Twitter and Facebook Through Social Network Analysis

¹RUKMANI P, ¹S. GRACELINE JASMINE, ¹VERGIN RAJA SAROBIN, ¹SATHIYA NARAYANAN, ²MODIGARI NARENDRA, ³DHANYA D, ⁴BENSON EDWIN RAJ,

¹School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India

²Department of Computer Science and Engineering, Vignan Deemed to be University, India

³Mar Ephraem College of Engineering & Technology, Marthandam, India

⁴Higher Colleges of Technology, Fujairah Women's Campus Fujairah, UAE

rukmani.p@vit.ac.in, graceline.jasmine@vit.ac.in, verginraja.m@vit.ac.in,
sathiyarayanan.s@vit.ac.in, narendramodigari@gmail.com, dhanvis@gmail.com, braj@hct.ac.ae

Abstract: In recent years, the use of social media has seen a significant surge. Having begun on a very small scale as a socialization forum, it has now helped itself in every possible sphere. Advertising goods, exchanging news, driving companies and much more are among them. More people are drawn to social networking sites such as Facebook, Google+, Twitter, so we propose an approach for estimating the lifetime and retweet times of tweets. Similarly the popularity of the tweets are also computed using the social network analysis. In order to create prediction knowledge bases, the data are extracted from retweet graphs, such as posting times, content information's, and area of interest etc. Tweets are sequentially extracted from the knowledge base with a similar topic, retweet pattern and properties and then used to make a prediction. The user-user relationship is followed and user-user relationships are modeled to better analyze the data. Different indices of centrality are determined to assess the prominent people who led the election campaign. In order to better understand whether social media has any major impact on tweet data, the outcomes are then compared with ground-truth evidence.

Keywords: Twitter Tweets, Centrality measures, Retweet pattern, Similarity Index

1. Introduction

On today's network, Twitter and other microblogging services have become indispensable sources of knowledge. For the study of opinion forming and many other views, knowing the key factors that make these pieces of knowledge spread rapidly on these channels may be decisive. Tweets from June 2012 to October 2012 on Twitter is collected to assess information's like a) why retweeting a tweet b) Is it possible to predict its frequency c) effect on the spread of tweets. Facebook has become one of the important platforms for these kinds of social interactions. Facebook allows people to have their own account to convey their feelings with the help of emoticons, comments

and reactions on particular topics. Today there are many burning topics and the people express their feelings about the topic through Facebook. The reactions about a particular post is mixed, thus it is difficult to look into the problems since there are too many suggestions. Identifying the sentiment analysis components of Facebook posts in order to identify that whether the particular post has positive or negative responses. This process makes use of the Natural Language Processing concepts for sentiment analysis.

Analysis of social networks has always been a sought after area of research. With online socializing on the rise day by day, a lot of data and information is being generated. The volume

of data is just not small and because of big data, there have also been several problems generated. In addition, such data often opens up an avenue for researchers in different fields. Analysis of social networks is one such area [1]. One can gain immense insights into how users communicate and socialize with each other with social network research. With such social data, users' actions can be recorded. Liking a post, sharing a post, following an individual, adding a friend to a network are all the ways in which users' interact with each other on social networks. The data from June 2012 to October 2012 on Twitter is collected to assess the success model of tweets. To create a learning model that predicts viral tweets with high accuracy, the most impactful features are chosen. All experiments are done on a real-world dataset, extracted from the TRE-C 2011 microblog corpus through a public Twitter API based on user IDs. Thus in this paper at first the comments of a particular post or the page is being extracted and then making use of the nltk tool in python which makes use of the natural language processing concept the sentiments are being analyzed for that particular post based on the comments.

2. Related Work

The authors have taken into account the posts or the comments written in English or Malay in order to do sentiment analysis and opinion mining of the comments. The authors have used Json library in order to extract the comments and have given the result based on the percentage of happy, unhappy and neutral comments [2]. The paper proposes a system which makes use of sentence level classification in order to categorize the post as positive, negative or neutral. The aims of the proposed system are to form their own corpus based on the data retrieved from Facebook API, then to accept the data which can contain both words and emoticons. The method is used for

categorizing the status update as positive or negative. This paper makes use of Bayes probability in order to determine that whether the status is positive or negative. The paper aims to perform sentiment analysis on the comments of the people on the particular post. The proposed system makes use of the hybrid system which is a combination of Machine Learning technique and Lexicon based technique. The sentiment analysis is done with the help of Facebook App. This paper makes use of Sentimental Orientation. In order to predict the sentiments of the words or the emoticons they are compared with pre tagged lexicons and the emoticons respectively. In order to know about the context of the words Named Entity Recognition is used with the help of Part of Speech Tagging. Another such study [3] explored social media interactions within and between extreme right groups. The work involved choosing established community groups on Twitter and examining their social media behavior habits and answering questions such as are they different or is there anything special that distinguishes these individuals from other regular social media users. A paper entitled "Surfacing contextual hate speech words in social media"[4] came with an innovative way of detecting hate speech by trying to solve the challenge of detecting hate speech as users try to use clear patterns to prevent automated detection and hate speech filters with continuous research in the same area of hate speech.

3. Proposed Work

This paper aims to analyses, study and draw conclusions of a learning model that predicts viral tweets with high accuracy, the most impactful features are chosen. To do so, an online-available dataset specific to this event was downloaded and used. The dataset was mainly composed of tweet objects.

3.1 Dataset Collection

1. Twitter API was used to gather the tweets
2. Submit a review to Twitter, permission was obtained for the using Twitter API
3. A python program using Tweepy library was
4. CSV file is maintained with the tweets

3.2 Graph Modelling

Member – Party graph and User – User follow graph and User – User retweet graph was computed for identifying the relationship between the tweets.

3.3 Social Network Analysis

3.3.1 Centrality Measures:

Closeness centrality detects nodes that, through a graph, can very efficiently spread information. The closer or shorter the distance a node has to with every other node, then the greater the centrality of closeness would be. Similarly the betweenness calculates how much a node has to travel along the shortest path to reach every other node. The amount of control a node has over the flow of information in a graph is detected. Betweenness centrality can give an insight into which users are better at connecting two different communities or groups.

3.3.2 Community Detection

With a distinctive sticker, each node is initialized (an identifier). Via the network, these marks spread. At each propagation iteration, each node changes its label to the one to which its neighbors' maximum numbers belong.

3.3.3 Page rank

To detect key players in the network who can quickly spread messages, page rank can be used. Higher page rank would normally mean that there are more supporters for the given tweets, and thus the information communicated would easily reach a wider audience and can also have a greater effect.

4. Results and Discussion

The results obtained by analyzing through social network analysis are detailed in this section.

To extract the comments the software called Facebook pager has been used. In this software the access token number has to be given. The access token is different for different users of Facebook. The user needs to have an account on Facebook. This token number can be accessed by the Facebook Graph API in Facebook's Developer page. Then the id of the page or the post can be identified by find my id site. Now Dataset can be extracted.

Sentiment Analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral. In the proposed methodology we have used sentiment analysis for identifying the main focus of the people on the particular post. The process proposed takes into account the comments of the particular post and do sentiment analysis on these comments based on the percentage of positive and the negative words. Thus in the given dataset we have removed unnecessary columns and taken only sentiment and text columns for sentiment analysis.

The degree centrality obtained for the tweets are shown in the figure 1.

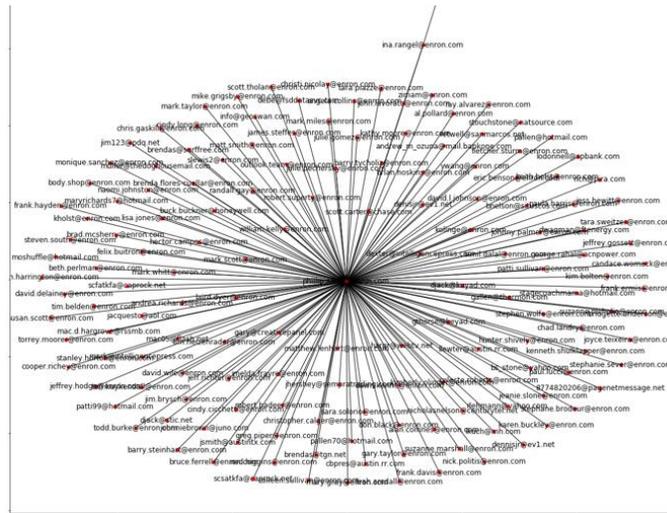


Figure 1. Degree Centrality of a tweet

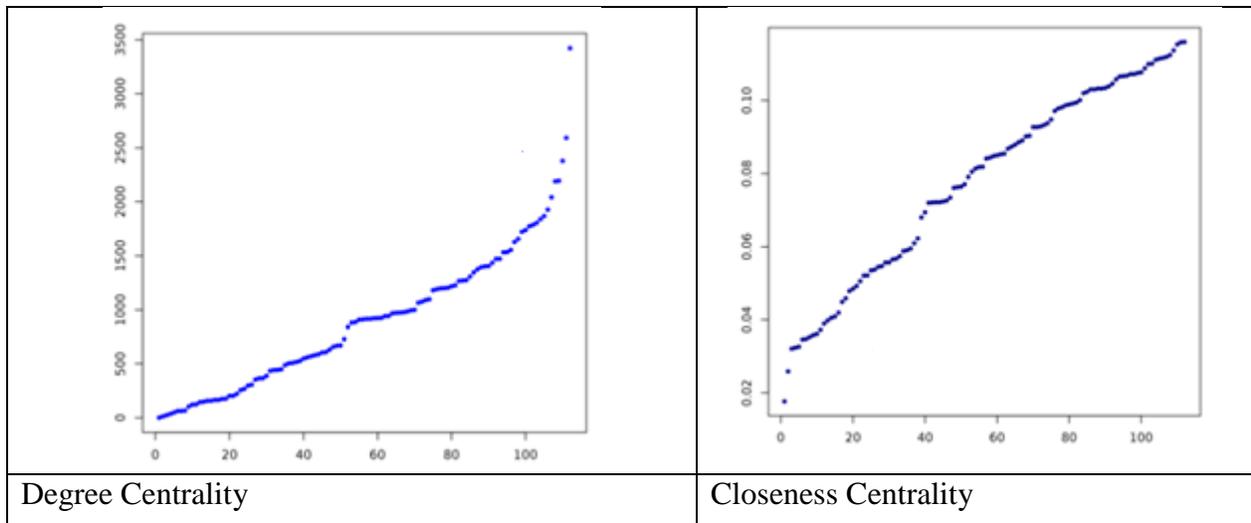
For predicting the viral tweets the model is trained on a collection of tweets specified by the $\varphi(x) = (x_1, \dots, x_n)$ and the tweets that have earned more retweets than a threshold. The model learns weights "wi"

for each feature based on this data and, given the feature vector of a tweet, it is measured by a "virality" score $v(x) = \sum_i^n w_i x_i$, where $i=1$. and is shown in the table 1.

Table 1. Viral Tweets predicted

PREDICTING VIRAL TWEETS		
0	449	52
Positive Sentiment is 0%		
Negative Sentiment is 89.62%		
Neutral Sentiment 10%		
7	116	2
Retweet Positive Sentiment is 5.6%		
Retweet Negative Sentiment is 92.62%		
Retweet Neutral Sentiment 1.6%		
Positive Neutral Negative		
Retweet 5.6% 1.6% 92.8%		
Pure 0.0% 10% 89.62%		
0.9		
p(r n): 0.20589279375221867		
p(r neu): 0.03993610223642173		
p(r p): 1.1182108626198082		
Probability of popularity of statement in our data which is most popular : 0.25		

Similarly the degree, closeness, eigen value and clustering coefficient obtained for the analyzed dataset is shown in the figure 2.



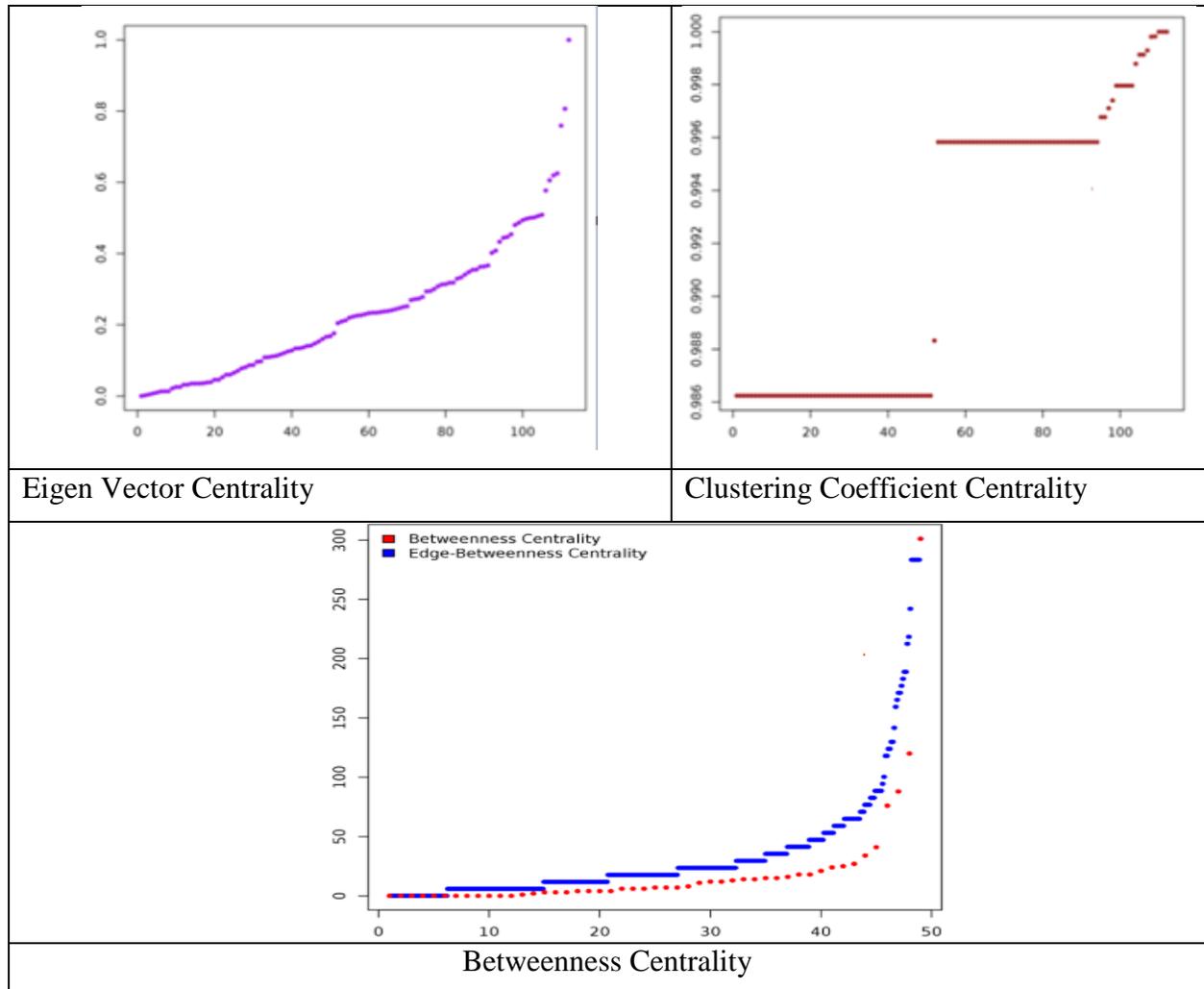


Figure 2. Centrality Measures

5. Conclusion

In conclusion, while Twitter data might not be able to predict retweets with a high degree of precision, it may still help to draw general conclusions about the effects of different users on their messages. Nevertheless, more sophisticated data techniques such as natural language processing may help to provide a better understanding of the path of retweets. Analyzing this re-tweet data allows businesses to satisfy the rising trend in order to achieve those business values, such as increasing customer numbers, boosting customer loyalty, customer satisfaction and business credibility, and achieving higher profits and overall

revenue. Social organizations will also seek the opinion of citizens on current debates or on topics such as the next leaders etc.

References

- [1].D Arshad,L.Jani Anbarasi, Modigari Narendra, Dhanya D, "Investigating the effect of social media campaign on German 2017 Elections", International journal of Recent technology and Engineering, vol8, Iss.9, pp. 2122-2129
- [2].N. Azmina m. Zamani, siti z. Z. Abidin, nasiroh omar, m. Z. Z.abiden Sentiment

- Analysis: Determining People's Emotions in Facebook
- [3].Derek O'Callaghan, Derek Greene, Maura Conway, Joe Carthy, P'adraig Cunningham, "Uncovering the Wider Structure of Extreme Right Communities Spanning Popular Online Networks", Proceedings of the 5th Annual ACM Web Science Conference, 2014
- [4].Cherries Taylor, Melvyn Peignon, Yi-Shin Chen, "Surfacing contextual hate speech words within social media",CoRR,2017.
- [5].Y. Han and K. K. Kim, "Sentiment analysis on social media using morphological sentence pattern model," *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, London, 2017, pp. 79-84, doi: 10.1109/SERA.2017.7965710.
- [6].C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno and J. Caro, "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning," *IISA 2013*, Piraeus, 2013, pp. 1-6, doi: 10.1109/IISA.2013.6623713.
- [7].S. M. Junaid, S. W. Jaffry, M. M. Yousaf , L. Aslam, S. Sarwar, Sentiment Analysis and Opinion Mining – A Facebook Posts and Comments Analyzer, *International Journal of Engineering Research & Technology (IJERT)*, 2018
- [8].P. Andr'e, M. S. Bernstein, and K. Luther. Who Gives A Tweet? Evaluating microblog Content Value. In Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW, pages 471–474, 2012.