Degree Centrality, Eigenvector Centrality and the relation between them in Twitter

Prantik Howlader, Sudeep K.S

***Abstract*—In Social Media the directed links formed between the users, are used for the transfer of information. Based on previous research, the rate of information transfer in a social network depends on the strength of connections of the user in the network, which is measured by the centrality value. In this paper, based on data collected from Twitter, we perform an analysis of eigenvector centrality approach of finding the influential users. We investigate the variation in indegree and eigenvector centrality of users participating in a hashtag in Twitter, with respect to change in the amount of interactions. Here interactions are: tweets, mentions and replies. We also investigate the relationship between indegree and eigenvector centrality in a given hashtag. We make the following interesting observations. First, in Twitter, users with high eigenvector centrality need not be influential users. Second, in a given hashtag, there is an increase in users with both high indegree and eigenvector centrality when there are more user interactions. Here interactions are: tweets, mentions and replies, indicating both indegree and eigenvector centrality should be considered when finding influential users. Third, there is a positive correlation between indegree and eigenvector centrality.**

Keywords—Twitter, Eigenvector Centrality, Degree Centrality.

I. INTRODUCTION

In social media, a group of users transfer information among each other. This transfer of information among a group of users can be represented by a social graph, which can be used to analyse the interactions in social network. The analysis of user interactions in social network is called Social Network Analysis(SNA) [1], [2].

In a social network, the centrality value of a node represents how many connections are there from that node to other nodes [1]. There are many ways to measure centrality of a node to measure its effect on the social network: betweenness centrality, closeness centrality, degree centrality and eigenvector centrality.

Centrality metrics have been used to identify influential nodes in dynamic processes, such as opinion competition, epidemic spreading and rumor propagation [3], [4], [5], [6]. In epidemic spreading, the identified influential nodes are the source nodes from which virus spreads and nodes with high spreading capacity. Centrality has also been used to select nodes which are to be immunized when a virus is prevalent [3], [7].

Degree centrality of a node is defined as the measurement of number of links the node has in a network [8], [9], [10]. Eigenvector centrality measure of a node depends on the centrality score of its adjacent nodes. The rationale behind this is that a student's popularity increases if he is voted as popular by other popular students. Thus eigenvector centrality of a node depends on the centrality of the adjacent nodes, not on the number of adjacent nodes [11], [12]. Previous research shows high eigenvector centrality value of a user in Twitter corresponds to higher influencing capability of the node without taking into consideration the effect of interactions. Here interactions are: *follows*, *replies* and *mentions* [2].

This paper investigates the eigenvector centrality approach to find the influential users in Twitter, taking into consideration the interactions. Here interactions are: follows, replies and mentions of the users. Further we investigate the variation in indegree and eigenvector centrality of users participating in a hashtag in Twitter, with respect to change in amount of interactions. Here interactions are: tweets, mentions and replies. We also investigate the relation between indegree and eigenvector centrality by finding the Pearson Correlation Coefficient.

II. METHODOLOGY

A. Degree Centrality

Degree Centrality of a graph is defined as[11]: for a graph $G = (V, E)$, where $|V|$ is the number of vertices in the graph. Degree Centrality of node v_i is defined as:

$$C_{D(v_i)} = deg(v_i) \qquad (1)$$

Where v_i in V. It is based on the idea that the number of "direct relations" of a user gives an indication of the structural importance of the user.

Finally, $\deg^{+(v_i)}$ gives the outdegree and $\deg^{-(v_i)}$ gives the indegree of the node v_i. Two users with same indegree, suggests that they have equal social status. While if one user X has higher indegree than user Y, then it means that user X is a celebrity compared to user Y [13].

B. Eigenvector Centrality

Eigenvector Centrality [12] gives a measure of the influence of the node based on the connections of the nodes to which it is connected. Same as degree centrality, eigenvector centrality favors nodes that have highest number of links. But unlike degree centrality, it also factors in the centrality of the adjacent nodes.It is defined as [2]:

For a given graph $G = (V, E)$, where $|V|$ is the number of vertices in the graph. Let $A = (a_{\{v,t\}})$ be the adjacency matrix, with $a_{\{v,t\}} = 1$ if vertex v and vertex t are adjacent in the graph otherwise $a_{\{v,t\}} = 0$. Eigenvector Centrality x_i of node i is defined as:

$$x_i = \mu \sum_{j=1}^{n} a_{ij} x_j \qquad (2)$$

Where a_{ij} is (i, j)th element of adjacencymatrix A, proportionality factor $\mu = \frac{1}{\lambda}$, λ a constant, and n is the number of nodes in graph G. So x_i is proportional to the sum of centrality scores of all nodes connected to it. Eigenvector centrality x_i of node i is the i-th entry in the normalized

Prantik Howlader, Department of Computer Science, National Institute of Technology, Calicut, Kerala, India (pratikbubun@gmail.com)

Sudeep K.S., Department of Computer Science, National Institute of Technology, Calicut, Kerala, India (sudeep@nitc.ac.in)

eigenvector belonging to the largest eigenvalue of adjacency matrix A.

III. TWITTER RELATIONSHIPS

In this paper we consider three types of relations that a user can have with the other users in Twitter. They are:

- **Follows:**Follows is a directed relationship. A user X is a follower of another user Y. Then all tweets by user Y can be seen by user X, but not vice versa.
- **Mentions and Replies:**This is directed communication. They involve the user making the reply and mention, and the user being replied or mentioned to. They are preceded by "@". Moreover, if a user X makes a reply or mention to a user Y, then followers of both user X and user Y can see that.

The aggregation of mentions, replies and the ratio of favorite tweets to the total number of tweets gives a measure. This measure indicates interpersonal activities between the user and other users [14]. Ratio of favorites to total number of tweets, provides an indication of acceptability of the user's information by other users.

IV. EXPERIMENTS

For performing these experiments, twitter data has been crawled using NodeXL. A total of 8 hashtags has been crawled having 11,324 users and 216,302 interactions which includes mentions, replies and follows.

In this paper, we conduct two experiments to analyse the eigenvector centrality approach to find the influential users in Twitter. Firstly, we find how users with high eigenvector centrality contribute in a hashtag. The contribution of a user in a hashtag is the total number of tweets, replies and mentions, the user has contributed in that hashtag. To measure this contribution of a user, we introduce a termActivity.

Activity Ac of a user i in hashtag H can be defined as:

$$Ac = T_i + M_i + R_i \tag{3}$$

where T_i is number of tweets by user i in hashtag H,M_i is number of mentions by user i in hashtag H, R_i is number of replies by user i in hashtag H.

Secondly we find, how information from users with different eigenvector centrality values are accepted by other users. Ratio of favorites to total number of tweets, provides an indication of acceptability of the user's information by other users. So, we did an analysis of how ratio of favorite tweets to tweets varies with eigenvector centrality of the users.

We also investigated the variation in indegree and eigenvector centrality of users participating in a hashtag in Twitter, with respect to the amount of interactions on days when the hashtag was trending in Twitter. Here interactions are: tweets, mentions and replies.

Further, we have investigated the relationship between indegree and eigenvector centrality by finding the Pearson Correlation Coefficient.

V. RESULTS AND DISCUSSION

First, we analyse results of the analysis of eigenvector centrality with respect to activity of the users. Fig 1 – Fig 2 shows the results on different hashtags in Twitter. We find that in 83.33% of the hashtags considered, users with high eigenvector centrality perform less activity in a hashtag than users with low eigenvector centrality.

Based on the above results, we observe that people with high eigenvector centrality always do not contribute many tweets, replies and mentions in a given hashtag.

Secondly, we analyse results of the analysis of ratio of favorites to tweets, with respect to the eigenvector centrality of the users participating in the hashtag. Fig 3 – Fig 4 shows the results on different hashtags in Twitter. We find that in all the cases, users with a high eigenvector centrality, have low ratio of favorites to tweets. In most cases, users with low eigenvector centrality have a much higher ratio of favorites to tweets than users with high eigenvector centrality.

Based on the above results, we observe that even if the user has a high eigenvector centrality. It does not mean information passed by him is readily accepted by the other users in the social network, as is indicated by low ratio of favorites to tweets.

Based on results of the above two experiments, we make our first observation, that even though in social networks, a high eigenvector centrality of a user indicates a high influential position in a network from a structural point of view. But when we consider the interactions like replies, mentions and follows. High eigenvector centrality does not indicate a corresponding higher influential position of the user among the other users. As can be observed from our experiments, users with high eigenvector centrality can provide very less number of tweets in a hashtag. In addition to that, users with high eigenvector centrality can have a low ratio of favorites to tweets. So of the few tweets propagated by a user of high eigenvector centrality, the chance of them being further propagated by the other users is less, as is indicated by low ratio of favorites to tweets.

Thirdly, we analyse results of the analysis of variation in indegree and eigenvector centrality of users participating in a hashtag in Twitter, with respect to interactions on days when the hashtag was trending in Twitter. Here interactions are tweets, mentions and replies. Fig 5– Fig 12 shows the results on different hastags. We find that there is an increase in users with both high indegree and eigenvector centrality participating in hashtags on days when the amount of interactions in that hashtag is high.

We observe, as we go through the days a hashtag was trending in Twitter. With increase in the amount of interactions in a given hashtag, there is an increase in the number of users participating in the hashtag, who have both high eigenvector centrality and indegree. Based on results of our analysis, we can observe that the amount of interactions in a hashtag depends on the users with both high eigenvector centrality and indegree participating in a hashtag. We suggest that indegree and eigenvector centrality should both be considered when finding users who are influential, because we find that there are more users with high values on both these factors on days when the hashtag is seeing more interactions, rather than on days when it is not.

Finally, we analyse results of the analysis of relationship between indegree and eigenvector centrality. Fig 13 shows the Pearson Correlation Coefficient between indegree and eigenvector centrality with respect to the days on which the different hashtags were trending in Twitter. The average Pearson Correlation Coefficient is 0.6. Thus indegree and eigenvector centrality are strongly correlated.

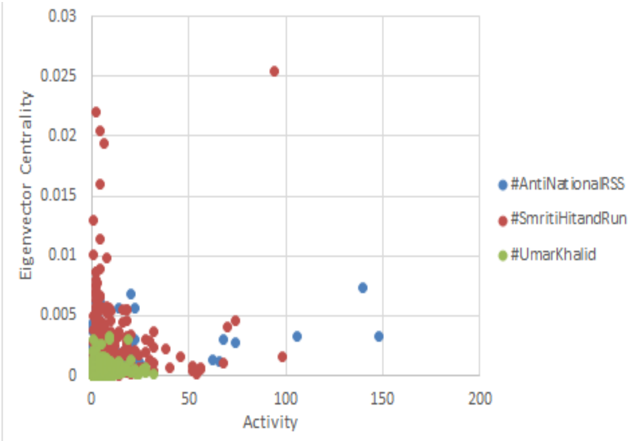

Fig. 1: Eigenvector Centrality with respect to Activity of #AntinationalRSS, #SmritiHitandRun and #UmarKhalid

VI. CONCLUSION

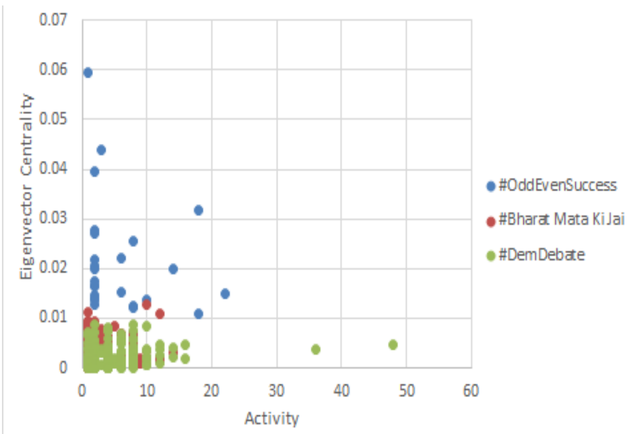

Fig. 1: Eigenvector Centrality with respect to Activity of #OddEvenSuccess, #Bharat Mata Ki Jai and #DemDebate

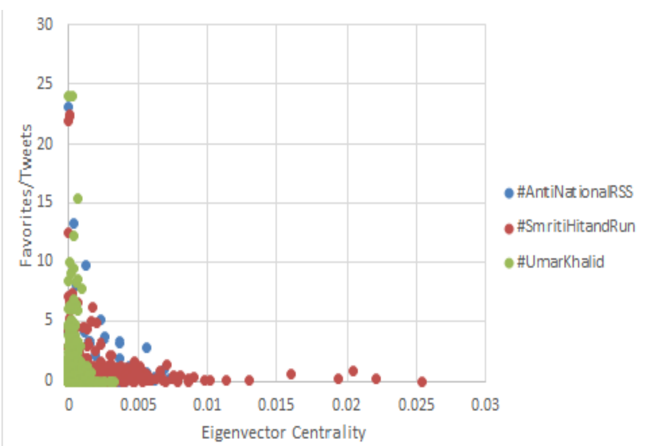

Fig. 2: Favorites/Tweets values with respect to Eigenvector Centrality of #AntiNationalRSS, #SmritiHitandRun and #UmarKhalid

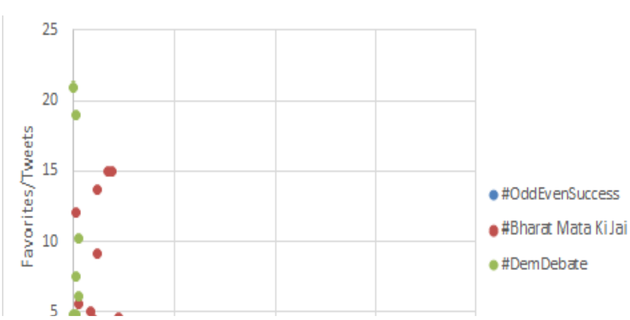

Fig. 3: Favorites/Tweets values with respect to Eigenvector Centrality of #OddEvenSuccess, #Bharat Mata Ki Jai and #DemDebate

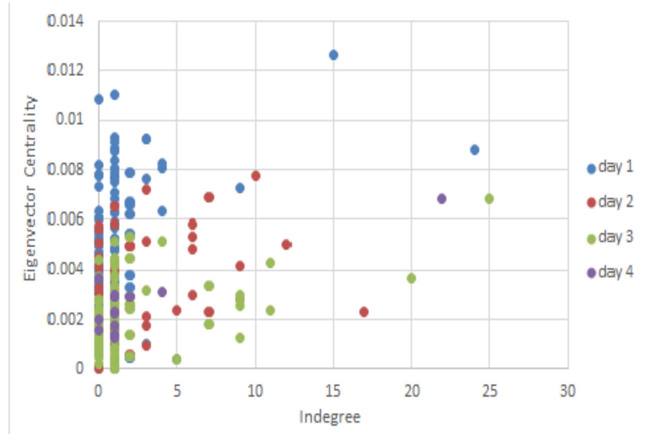

Fig. 4: Eigenvector Centrality with respect to Indegree of #Bharat Mata Ki Jai, on day 1, day 2, day 3 and day 4, the corresponding interactions on these days were 122, 141, 236 and 38

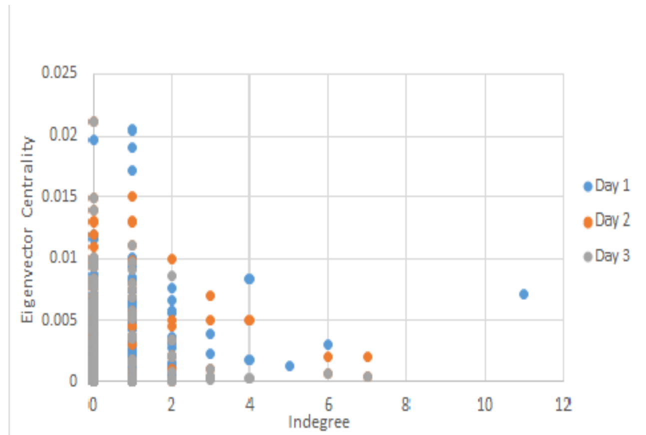

Fig. 5: Eigenvector Centrality with respect to Indegree of #WomensDay, on day 1, day 2 and day 3,the corresponding interactions on these days were 227, 213 and 213

We investigated the eigenvector centrality approach of finding influential users in Twitter. Our experimental results suggest that eigenvector centrality of users does not provide the influential users in Twitter, when we consider interactions. Here interactions are: follows, mentions and replies. We also investigated the variation in indegree and eigenvector centrality of users participating in a hashtag in Twitter, with respect to change in amount of interactions in that hashtag. Here interactions are: tweets, mentions and replies. Our experimental results show that for a given hashtag, there is an increase in users with both high indegree and eigenvector centrality, when there are more interactions. Based on the above results we suggest, indegree and eigenvector centrality should both be considered when finding influential users in Twitter. We further investigated the relation between indegree and eigenvector centrality, and found that there is is positive correlation between indegree and eigenvector centrality.

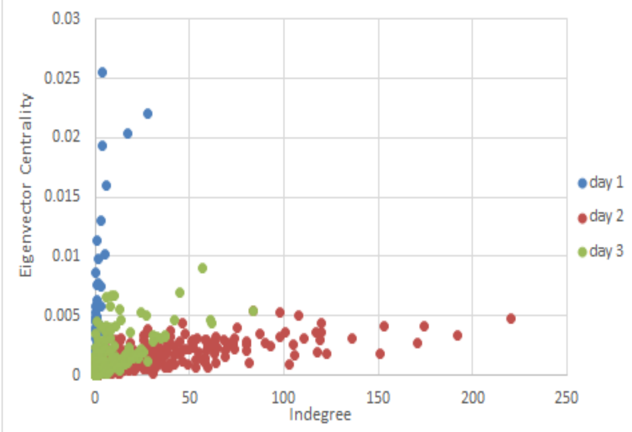

Fig. 6: Eigenvector Centrality with respect to Indegree of #SmritiHitandRun, on day 1, day 2 and day 3, the corresponding interactions on these days were 108, 1748 and 884

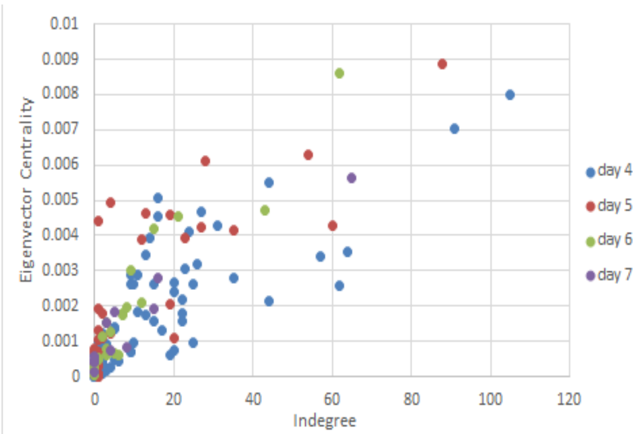

Fig. 7: Eigenvector Centrality with respect to Indegree of #SmritiHitandRun, on day 4, day 5, day 6 and day 7, the corresponding interactions on these days were 140, 60, 80 and 33

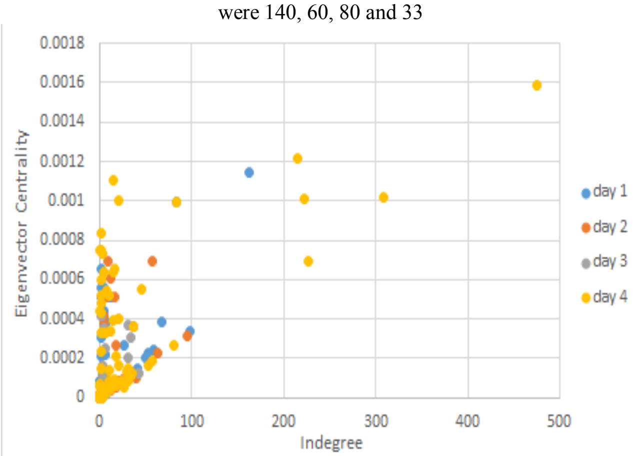

Fig. 8: Eigenvector Centrality with respect to Indegree of #UmarKhalid, on day 1, day 2, day 3 and day 4, the corresponding interactions on these days were 127, 110, 75 and 250

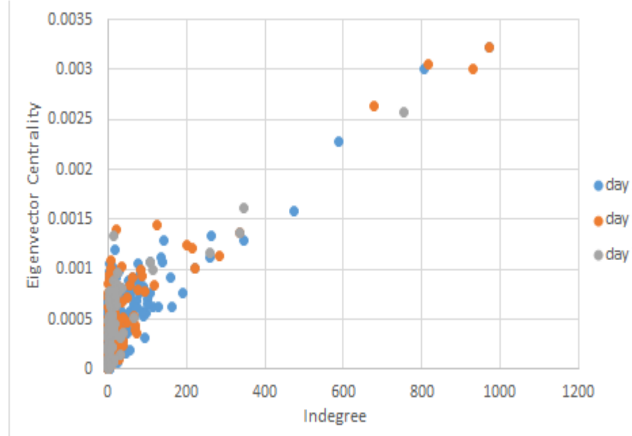

Fig. 9: Eigenvector Centrality with respect to Indegree of #UmarKhalid, on day 5, day 6 and day 7, the corresponding interactions on these days were 802, 701 and 214

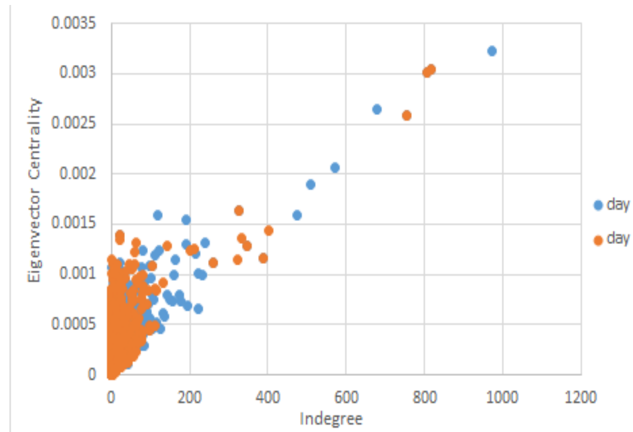

Fig. 10: Eigenvector Centrality with respect to Indegree of #UmarKhalid, on day 8 and day 9, the corresponding interactions on these days were 2048 and 1552

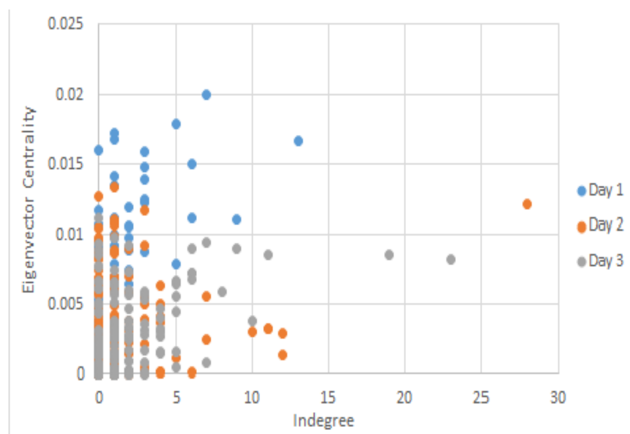

Fig. 11: Eigenvector Centrality with respect to Indegree of #Brussels, on day 1, day 2 and day 3, the corresponding interactions on these days were 120, 215 and 236

This research does not analyse the effect of retweets, links in tweets, betweenness centrality and closeness centrality on influential users in Twitter. For future research, these will be analysed to find, how different factors affect the influencers in Twitter.

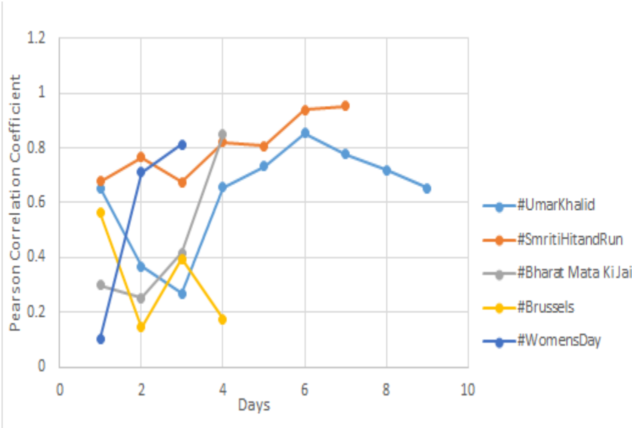

Fig. 12: Pearson Correlation Coefficient between Indegree and Eigenvector Centrality of 5 hashtags, with respect to the days the hashtags were trending in Twitter

REFERENCES

[1] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications. Cambridge University Press, 1994.

[2] Warih Maharani, Alfian Akbar Gozali, et al.Degree centrality and eigenvector centrality intwitter. In Telecommunication Systems Services and Applications (TSSA), 2014 8th International Conference on, pages 1-5. IEEE, 2014.

[3] Cong Li, Qian Li, Piet Mieghem, H. EugeneStanley, and Huijuan Wang. Correlation between centrality metrics and their application tothe opinion model, 88(3): 1-13, 2015.

[4] C. H. Comin and L. da Fontoura Costa. “Identifying the starting point of a spreading process in complex networks,” Phys. Rev. E, vol. 84, p. 056105, Nov 2011.

[5] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. “Identification of influential spreaders in complex networks,” Nature physics, vol. 6, no. 11, pp. 888–893, 2010.

[6] J. Borge-Holthoefer and Y. Moreno. “Absence of influential spreaders in rumor dynamics,” Phys. Rev. E, vol. 85, p. 026116, Feb 2012.

[7] R. Pastor-Satorras and A. Vespignani. “Immunization of complex networks,” Phys. Rev. E, vol. 65, p. 036104, Feb 2002.

[8] J. Scott, P. J. Carrington and S. Wasserman.Models and methods in social network analysis, volume 28. Cambridge University Press, 2005.

[9] Jon M Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S Tomkins. The web as a graph: measurements, models, and methods. In Computing and combinatorics, pages 1-17, Springer Berlin Heidelberg, 1999.

[10] Thomas W Valente, Kathryn Coronges, Cynthia Lakon, and Elizabeth Costenbader. How correlated are network centrality measures? Connections (Toronto, Ont.), 28(1):16, 2008.

[11] B. Ruhnau. Eigenvector-centrality a nodecentrality? Social networks, 22:357-365, 2000.

[12] P. Bonacich and P. Lloyd. Eigenvector-like measures of centrality for asymmetric relations. Social networks, 22:357-365, 2000

[13] Justin Cheng, Daniel M Romero, Brendan Meeder, and Jon Kleinberg. Predicting reciprocity in social networks. In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on, pages 49-56. IEEE, 2011.

[14] Yuheng Hu, Shelly Farnham, and Kartik Talamadupula. Predicting user engagement on twitter with real world events. In Ninth International AAAI Conference on Web and Social Media, 2015.