World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:12, No:8, 2018

Understanding the Selectional Preferences of the Twitter Mentions Network

R. Sudhesh Solomon, P. Y. K. L. Srinivas, Abhay Narayan, Amitava Das

Open Science Index, Computer and Information Engineering Vol:12, No:8, 2018 publications.waset.org/10009415.pdf

***Abstract*—Users in social networks either unicast or broadcast their messages. At mention is the popular way of unicasting for Twitter whereas general tweeting could be considered as broadcasting method. Understanding the information flow and dynamics within a Social Network and modeling the same is a promising and an open research area called Information Diffusion. This paper seeks an answer to a fundamental question - *understanding if the at-mention network or the unicasting pattern in social media is purely random in nature or is there any user specific selectional preference*? To answer the question we present an empirical analysis to understand the sociological aspects of Twitter mentions network within a social network community. To understand the sociological behavior we analyze the values (Schwartz model: *Achievement, Benevolence, Conformity, Hedonism, Power, Security, Self-Direction, Stimulation, Traditional and Universalism*) of all the users. Empirical results suggest that values traits are indeed salient cue to understand how the mention-based communication network functions. For example, we notice that individuals possessing similar values unicast among themselves more often than with other value type people. We also observe that traditional and self-directed people do not maintain very close relationship in the network with the people of different values traits.

***Keywords*—Social network analysis, information diffusion, personality and values, Twitter Mentions Network.

I. Introduction

INFORMATION diffusion is the process of spreading information or content within a network via a particular path or pattern. A significant amount of research has been done in this area in the past few years. However, most of the previous efforts considered only network topology for the diffusion process. Here we bring into picture the human societal sentiment (values) factor to understand the at-mention communication behavior in Twitter at large scale.

To understand the societal sentiment we borrow the Schwartz model: *Achievement, Benevolence, Conformity, Hedonism, Power, Security, Self-Direction, Stimulation, Traditional and Universalism* and built an automatic NLP based model to categorize people in these values types by analyzing their language usage in social media and their social network behavior. Then to understand the propagation process we analyzed who (which values type) is connected with whom (vs. which values type) and in what manner at individual level. To understand the user level neighbouring preferences we have analyzed values vs. values closeness preferences on the at-mention network. Closeness centrality measures is the mean distance from a vertex to other vertices in a network. We are reporting values vs. values (10 X 10 matrix) preferences for the at-mention behaviour in terms of closeness. Our analysis reveals several interesting outcomes. For example, universal people maintain an average closeness with all other values types people and also on contrary power oriented people are more connected with conformity and security oriented people. Then to understand the communication preferences between a pair of users we have used the concept of reciprocity. Reciprocity is most commonly defined as the ratio between how many times a user pair communicate to each other directly. Reciprocity according to at-mention network can be expressed as, if a pair of users i.e. user A and user B at-mention each other frequently, then they are said to be reciprocate. Now, we analyze values vs. values (10 X 10 matrix) types preferences for the at-mention behaviour in terms of reciprocity. The results obtained, suggest that the value traits act as an important feature to understand the functionality of the mentions network.

II. Related Works

The research paradigm called information diffusion seeks to answer how information spreads in a social network and model how a given piece of information will propagate through a social network - more precisely what a user will do with a particular tweet (lets say), will he/she either retweet it, at-mention somebody or broadcast it again to spread it over to a wider audience within his/her reachability in the network. Essentially researchers seek to answer to the following questions :(i) *which snippets of data/information or subjects are very popular (familiar) and diffuse the most*, (ii) *how, why and through which ways is the data diffusing, and will be diffused later on*, (iii) *which members of the network play critical role in the spreading of information?* [1]

A considerable amount of work has been done in modeling the process of information diffusion in online social networks. Previous works on information diffusion have considered several influencing factors such as speed, scale, range, influential nodes, network topology, topics etc. In the following paragraphs we are describing such related works.

Research endeavors by [2], [3] discussed diffusion process based on network topology and they explain about the concept of influential nodes or in simple terms, which node/s will influence the other nodes in the diffusion process. Reference [2] explains about combinatorial optimization problem, which is a way to find out the most influential nodes in a social

R. Sudhesh Solomon, P. Y. K. L. Srinivas, Abhay Narayan and Amitava Das are with the Department of Computer Science and Engineering, Indian Institute of Information Technology, Sri City, Andhra Pradesh, 517541 India (e-mail: sudheshsolomon.r@iiits.in, srinivas.p@iiits.in, abhay.n@iiits.in, amitava.das@iiits.in).

Open Science Index, Computer and Information Engineering Vol:12, No:8, 2018 publications.waset.org/10009415.pdf

network. In [3] the authors explain about understanding the dynamics of social networks and modeling the same, dynamics here refer to the topological structure of the network. The authors also explained about various information diffusion parameters (diffusion rate, who influenced whom etc.) in this work. Research by [4], tried to capture time dimension of the diffusion pattern. The main motivation of the authors in this work was to infer the edges and the dynamics of the underlying network.

Some of the other works discussed about the topic based diffusion pattern. Work by [5], analyzed diffusion pattern based on hashtags categorizations such as celebrity, games, idioms, movies, tv, music, politics, sports, and technology. To describe the diffusion patterns the authors took two measures - Stickiness: *The measure of the contingency of an information. The peak value of the curve*. Persistence: *The time for which an information stays on a particular diffusion rate. The measure of rate of decay after the peak*. Then they empirically show how topical variations affect stickiness and persistence of information diffusion patterns. The other interesting work by [6] proposed a probabilistic model to understand how two people will converse about a particular topic based on their similarity: *based on demographic information. The popular idea of homophily and heterophily* and familiarity: *based on time that they spend together in same topic*.

Retweeting is the famous way of information cascading in Twitter. There are research endeavors to predict how retweeting diffusion pattern will be. The work by [7] moduled the information diffusion task as a predictive modeling. Using a large scale data on who has retweeted and what was retweeted a probabilistic collaborative filtering model was built to predict the future retweeting pattern. The model learnt on parameters like the tweet source (the tweeter), the user who was retweeting and the retweet content. Works by [8] discussed about several influencing factors such as speed, scale and range of retweeting behavior. The first factor analyzed was Speed – *whether and when the first diffusion instance will take place*. To perform the analysis on speed, two models were used. The first model answers when a tweet containing a particular topic is likely to be mentioned by another tweet containing the same topic. For example, when user A posts a tweet related to a topic XYZ, how quickly another user (say user B), responds to the tweet consisting XYZ mentioning user A. Secondly, the Cox proportional hazards model [9] was used to quantify the degree to which a number of features of both users and tweets themselves to predict the speed of diffusion to the first degree offspring. The second factor explained and analyzed in this work is Scale – *the number of affected instances at the first degree*. In this work, the number of times a person is mentioned in the retweet trail relating to a topic was analyzed and a probabilistic diffusion model has been proposed. The last factor considered in this work is Range – *how far the diffusion chain can continue on in depth.* The analysis on range was done by tracing a topic from a given start node to its second and third degree of offspring nodes, and so on.

A few works have discussed about behavior of group of individuals - **Herd Behavior**: a social behavior occurring when a group of individuals make an identical action, not necessarily ignoring their private information signals. However, user level sentimental preference is being ignored so far. Therefore, our current work is on understanding user societal sentiment behavior. Our theoretical point of departure is in psycho-socio-linguistic models, the Schwartz model *Achievement, Benevolence, Conformity, Hedonism, Power, Security, Self-Direction, Stimulation, Traditional and Universalism.*. We hypothesize that people have natural preferences for direct communications. That means certain type of people who possess one value type have preference over other kind of people of different value within their range. For example, we observe that the traditional people are less likely involved in communication (i.e, unicasting) compared to other communities of people of different value types.

III. Schwartz Values - The Societal Sentiment

The values model was introduced by Schwartz in [10] and modified in [11]. The model defines ten basic and distinct personal ethical values, that are are given in the Table I respectively:

TABLE I
Description for Schwartz Values

Values	Description
Achievement	sets goals and aims at achieving them
Benevolence	seeks to help others and provide general welfare
Conformity	obeys clear rules, laws and stuctures
Hedonism	seeks pleasure and enjoyment
Power	controls and dominates others, control resources
Security	seeks health and safety
Self-direction	wants to be free and independant
Stimulation	seeks excitement and thrills
Tradition	does things blindly because they are customary
Universalism	seeks peace, social justice and tolerance for all

The Schwartz' values model supports fuzzy membership. Schwartz' theory explains how the values are interconnected and influence each other. This is because of the fact that an individual is not constrained to one particular value rather he/she may possess several value traits. Fig. 1 represents similar fuzzy membership of schwartz values classes. For example, from Fig. 1 the fuzzy membership of the ACY oriented people is represented by outgoing red bands. The width of each outgoing band from ACY represents the degree of membership of ACY with other Values classes. Similarly, we can observe that there are 10 incoming bands of 10 different colours towards ACY, indicating the membership of each class in ACY. In each class there is a self-arc which represents membership of each class with itself (i.e., 100%). The intricate structure of the Circos figure rightly signifies how values are strongly connected with each other at societal level.

The computational Schwartz model has been first proposed by [12], [13]. The authors released a corpus of 367 unique users having 1,608 average tweets per user labelled with values traits. The highest number of tweets for one user was 15K, while the lowest number of tweets for a user was a mere 100. For building the automatic classifier, the authors proposed a comprehensive set of features such

Open Science Index, Computer and Information Engineering Vol:12, No:8, 2018 publications.waset.org/10009415.pdf

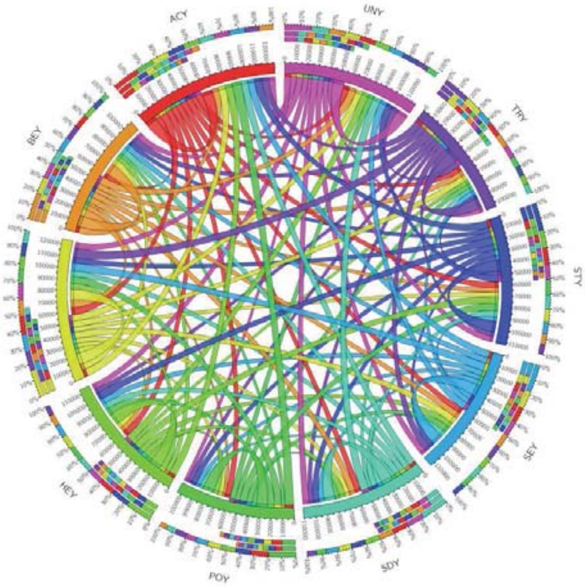


Fig. 1 Schwartz Values fuzziness for Twitter Values corpus using Circos Visualisation.

as For the automatic categorization of Personalities and Values, several psycholinguistic features were tested including Linguistic features (LIWC [14], Harvard General Inquirer, MRC psycholinguistic feature, and Sensicon [15]), network properties (Network size, betweenness centrality, density and transitivity), and Speech-Act classes. Their SVM-based model achieved an average F-Score of 0.81. In this research work, we have replicated the same classifier.

IV. Twitter Communication Corpus

TABLE II
Statistics about the Twitter Corpus

Total No. of users	15,496
Users considered	6,739
Total number of tweets	67,63,255
Highest No. of tweets (a user)	3,641
Lowest No. of tweets (a user)	100
Average tweets (per user)	2,406
Number of at-mentions	27,26,657

The Twitter data, released by SNAP [16] (nodes: 81,306, edges: 1,768,149), was used as a communication data for our work. The original dataset had 15,496 users. Further, all the non-existent accounts were discarded as well as those users who had less than 100 tweets. The tweet data of the remaining 6,739 users was downloaded. The highest (*resp.* lowest) number of tweets for one user was 3,641 (*resp.* 100) with the average number of tweets per user being 2,406. We had crawled nearly 67,63,255 tweets in total from the 6k users. Table II delineates the statistics about the dataset that was used for this analysis.

V. Understanding the Sociological Aspects of the Twitter Mention Network

The ten basic values relate to various outcomes and effects of a person's role in a society [17]-[20]. The values have also proved to provide an important and powerful explanation of the behaviour of the individual and how they influence it [21], [22]. Moreover, there are results that indicate how values of workforce and ethical practices in organizations are directly related to transformational and transactional leadership [23].

This paper seeks an answer to a fundamental question - *whether there is any preference in the choice in terms of personality or values type to establish a direct communication in Twitter?* Our theoretical point of departure is in Schwartz models. To understand the various sociological aspects of the Twitter Mention Network, we first created a Network comprising of source and target nodes in terms of at-mention. Once the network was created we performed an analysis on closeness and reciprocity between users of different values types pairs (i.e. Achievement-Achievement, Achievement-Benevolence and so on). as a result of this analysis we obtained a 10x10 matrix for both closeness and reciprocity separately. Based on these results we were able to come up with some very interesting observations (discussed in Section VI).

A. Network Creation

A directed graph is created from the source to the target of each at-mention at tweet level. The source corresponds to the individual who is tweeting and the target corresponds to the person being mentioned in the tweet. The generated network has 6738 nodes and 27, 26, 657 edges. In this process we have excluded nodes (read users) who have never mentioned someone or was never mentioned by someone else.

Fig. 2 portrays a toy example. Let us consider user 6 has tweeted the following **Looking forward to the next @user_10. It's been a little while since the last one. :)** and at-mentioned user 10. Thus from this tweet we were able to create a network where user 6 is the source and the user 10 is the target. For example user 9 is excluded from the network as it is never being mentioned someone or has never mentioned somebody else.

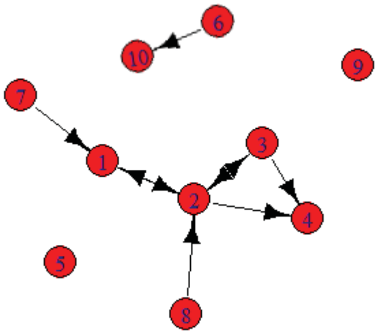


Fig. 2 A sample representation of 10 users in the network and their relationship

Fig. 2 shows a sample representation of a network of 10 users. The nodes represent the users in the network and the edges represents the connection or the relationship between the users. It is also important to note that not all the users in the network might not be connected. For example, when we take Fig. 2 into consideration we are able to find that there are

Open Science Index, Computer and Information Engineering Vol:12, No:8, 2018 publications.waset.org/10009415.pdf

a total of 10 users labeled from 1 to 10 and 6 nodes(users) are connected within a network, 2 nodes are connected separately (user 6 and user 10), and on the other hand, we are also able to find that there are some nodes that are not connected in the network (user 5 and user 9). After we created the network we tried to understand the sociological aspects of the network (at-mention network), for which we analyzed the closeness and reciprocity, both of which are formulated in Section V-B.

In the at-mention network we have considered the edges to be directed. From Fig. 2 we are able to find that user 6 has a directed edge towards user 10. This means that user 6 has mentioned user 10 in his/her tweet. This factor of directed graphs plays a vital role in analyzing who mentions whom? On the other hand, closeness does not require the graph to be directed. All that we need to calculate the closeness of the network are nodes and edges (either directed/undirected). The edges are used to determine how close is a user with other users in the network. For example, from Fig. 2 we are able to observe that user 2 is very close to others in the network than the other users.

B. Sociologial Aspects of the Twitter Mention Network

Once the network is created, we tried to understand various sociological aspects that influence the twitter mention network. To do this, we made an analysis on the mentions pattern based on the value types. Here, we considered factors like reciprocity and closeness. Reciprocity was calculated between each pair of values types users. For example (achievement-achievement, achievement-benevolence and so on) between each value types. Similarly closeness was also calculated between each value pairs.

Closeness: In a connected graph, the closeness or the closeness centrality of a node is a measure to understand how close a particular node is with respect to other nodes in terms network. It is calculated as the sum of the lengths of the shortest paths between the particular node and the other nodes in the graph. If the closeness measure of a node is higher - that means the node is closer to the other nodes. This measure is used to identify how close a particular node is with the other nodes in the network. For example in Fig. 2 we are able to observe that user 2 is comparatively more closer with other nodes in the graph and hence the closeness of the user 2 is higher than others. On the other hand, the disconnected users (users 5 and 9) will have a lower closeness centrality score i.e. 0.

$$Closeness = 1/sum(d(v, i), i! = v) \tag{1}$$

Reciprocity: The measure of reciprocity defines the proportion of the number of at-mentions in the network. It is most commonly defined as the probability that the opposite counterpart of a directed edge is also included in the graph. For example, from Fig. 2, let us consider users 2 and 3, we are able to infer that both the users mentions each other in their tweets and hence will have a high reciprocity score whereas, the connection between user 6 and 10 and the disconnected users (user 5 and user 9) will have a very low reciprocity score i.e. 0. In adjacency matrix notation [24]:

$$Reciprocity = sum(i, j, (A. * A')_{ij})/sum(i, j, A_{ij}) \tag{2}$$

where A.*A' is the element-wise product of matrix A and its transpose

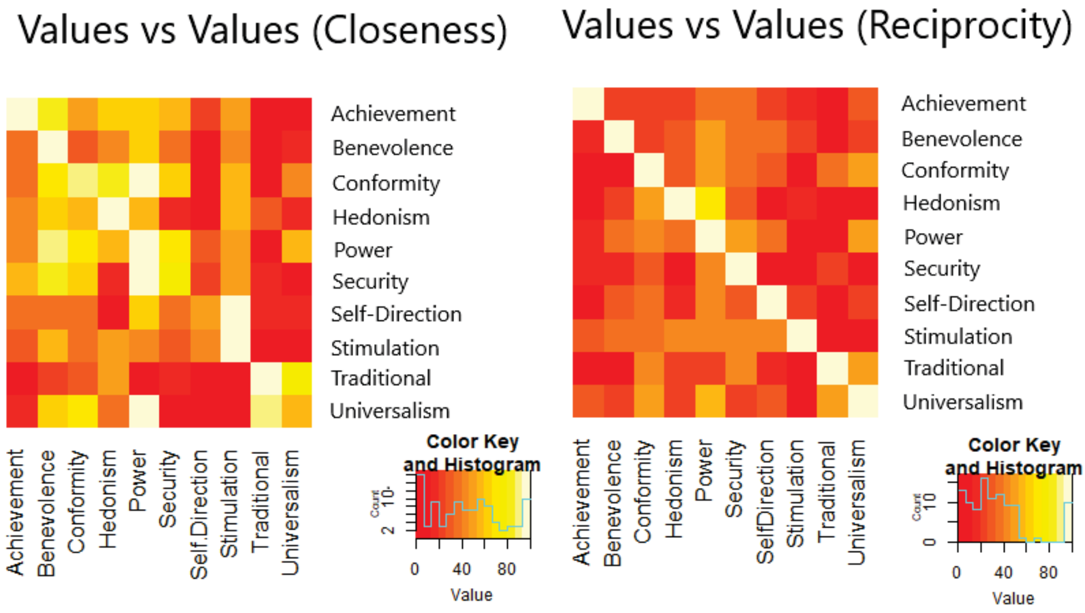
VI. Who Mentions Whom? - The Findings

To understand the notion of closeness in an at-mention network we have calculated values pair (achievement-achievement, achievement-benevolence, and so on) wise average closeness centrality, resulted in a 10 x 10 matrix. This is an analysis to understand who are close with whom. The result of the analysis is plotted in the form of a heatmap in the Fig. 3a.

We can observe from the analysis that people of same kinds prefer to remain closer among themselves. This possibly supports the well-defined homophily phenomena. But there are more to it. From the analysis we observe that conformity, security, self-directed, and universal people do maintain an average closeness with all the other values types of people. On the other hand we have also noticed that the power and stimulation oriented people are very close to their own type of people as well as conformity, security, and universal people. The power oriented people are those type of people who are dominant over the other types and they are close with the conformity-oriented people who are bound to follow rules and structures. The universal people on the other hand, who strive for social justice which can be achieved when someone with power is by their side are close with the power-oriented people and hence the closeness between power-oriented and universal people is high.On the contrary, there are certain types of people who are not very close to any particular type of people, other than their own kind. Traditional, self-directed and security seeking people exhibit such characteristics. When we relate them to Schwartz theory of human values we can possibly justify such behavior: 1) Security-oriented people are determined only about health and safety, hence they do not come into close relationship with various types of people like hedonic, stimulant people who are focused on excitement and enjoyment at that moment. 2) The self-directed people according to Schwartz theory like to be independent and free. We are clear from the definition that they want to be separate from groups and constraints imposed on themselves. 3) Lastly, the traditional people are those who believe in customs and traditions blindly and follow them. Hence they do not come into close relationship with other type of people who the feel would contradict their beliefs.

We did similar analysis in terms of reciprocity. Here we seek answer to the research question - *whether the at-mention or the unicasting pattern in social media is purely random in nature or is there any user specific selectional preference*? i.e to identify the choice of a person for unicasting his/her tweets. The reciprocity between each values pairs are calculated (achievement-achievement, achievement-benevolence, and so

Open Science Index, Computer and Information Engineering Vol:12, No:8, 2018 publications.waset.org/10009415.pdf



(a) Closeness Analysis

(b) Reciprocity Analysis

Fig. 3 Heatmap Visualization on the Analysis of various Sociological Aspects of Twitter Mention Network (Values)

on). Thus we obtained a 10 X 10 matrix which was then used to create the heatmap represented in Fig. 3b.

In this analysis we observed that the users having same values reciprocated highly among themselves. Similar homophily behavior in general. Moreover we observe that achievement-oriented people have high reciprocity with the power-oriented, stimulant and universal people, as both the power and achievement values people focus on social esteem. We also notice that traditional and conformity-oriented people preferred unicasting, i.e. had a high reciprocity score among themselves. The stimulant and self-directed people also had high reciprocity scores among themselves. This showed their intrinsic interests in novelty and mastery. On the other hand, the individuals who possessed the value of self-direction, i.e. people who prefer to be free and independent did not show much higher reciprocity with the other values types apart from stimulant people, thus preferring to broadcast the tweets more often.

VII. Discussion

This is an ongoing work. We are interested in understanding user level selectional preferences in terms of at-mention. Selectional preferences could be established once we have the complete picture on how a user is surrounded (in terms on follower and following network) i.e. by whom and then whom s/he choose to at-mention for a particular topic or message. The closeness analysis reported here is on at-mention network and not on the real network i.e. follower and following network. The limitation here arises on the fact that, when we consider the mentions network, we are not considering several users. For example, if user A never mentions user B and user B never mentions user C, although they are connected through some relationship, just because they did not mention someone or they never got mentioned by someone else they are being excluded. For example, in the Fig. 2, users 5 and 9 are being excluded from our current analysis. We are now working on closeness analysis based on the follower and following network. Secondly, the analysis which we have performed here in this work emphasize only on the sociological aspects of the twitter mentions network. We would like to extend this work in terms of psychological (personality) aspects as well, and finally obtain a psycho-sociological analysis on closeness and reciprocity on the Twitter mentions network. Finally, we would like to consider several other factors like age, gender, content type of the message which can possibly have some role to play in understanding the dynamics of the at-mention network.

VIII. Conclusion

In this paper our contributions are three folds - 1) We present an empirical analysis to understand the sociological aspects of the Twitter mentions network. 2) To establish that notion we present our analysis on user sociological trait vs. closeness and reciprocity. 3) Empirical suggests that there are strong correlations between the user's unicasting/broadcasting behaviour vs. his/her sociological traits.

Communication dynamics in human society is a complex phenomenon. The current paper is explanatory in nature, but we strongly believe such findings could be successfully used to solve several practical problems. We are now working on *link prediction*, where we are using all the analytical results obtained from the present study. We believe that this kind of models may become extremely useful in the future for various purposes like Internet advertising (specifically social media advertising), computational psychology, recommendation systems, psycho-sociological analysis about users over social media.

References

[1] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM SIGMOD Record*, vol. 42, no. 2, pp. 17–28, 2013.

[2] M. Kimura, K. Saito, R. Nakano, and H. Motoda, "Extracting influential nodes on a social network for information diffusion," *Data Mining and Knowledge Discovery*, vol. 20, no. 1, pp. 70–97, 2010.

[3] M. Wani and M. Ahmad, "Survey of information diffusion over interaction networks of twitter," *International Journal of Computer Application*, vol. 3, no. 4, pp. 310–313, 2014.

[4] M. Gomez Rodriguez, J. Leskovec, and B. Schölkopf, "Structure and dynamics of information pathways in online media," in *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 2013, pp. 23–32.

[5] D. M. Romero, B. Meeder, and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 695–704.

[6] A. Apolloni, K. Channakeshava, L. Durbeck, M. Khan, C. Kuhlman, B. Lewis, and S. Swarup, "A study of information diffusion over a realistic social network model," in *Computational Science and Engineering, 2009. CSE'09. International Conference on*, vol. 4. IEEE, 2009, pp. 675–682.

[7] T. R. Zaman, R. Herbrich, J. Van Gael, and D. Stern, "Predicting information spreading in twitter," in *Workshop on computational social science and the wisdom of crowds, nips*, vol. 104, no. 45. Citeseer, 2010, pp. 17 599–601.

[8] J. Yang and S. Counts, "Predicting the speed, scale, and range of information diffusion in twitter." *ICWSM*, vol. 10, pp. 355–358, 2010.

[9] D. R. Cox and D. Oakes, *Analysis of survival data*. CRC Press, 1984, vol. 21.

[10] S. H. Schwartz and W. Bilsky, "Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications." *Journal of personality and social psychology*, vol. 58, no. 5, p. 878, 1990.

[11] S. H. Schwartz, "Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries," *Advances in experimental social psychology*, vol. 25, pp. 1–65, 1992.

[12] T. Maheshwari, A. N. Reganti, U. Kumar, T. Chakraborty, and A. Das, "Semantic interpretation of social network communities." in *AAAI*, 2017, pp. 4967–4968.

[13] T. Maheshwari, A. N. Reganti, S. G. A. Jamatia, U. Kumar, B. Gambäck, A. Das, and S. S. I. AB, "A societal sentiment analysis: Predicting the values and ethics of individuals by analysing social media content."

[14] L. Inquiry and W. Count, http://liwc.wpengine.com/.

[15] Sensicon, https://hlt-nlp.fbk.eu/technologies/sensicon.

[16] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," http://snap.stanford.edu/data, Jun. 2014.

[17] A. Argandoña, "Fostering values in organizations," *Journal of Business Ethics*, vol. 45, no. 1, pp. 15–28, 2003.

[18] B. R. Agle and C. B. Caldwell, "Understanding research on values in business: A level of analysis framework," *Business & Society*, vol. 38, no. 3, pp. 326–387, 1999.

[19] G. Hofstede, "Cultures and organizations: software of the mind london," *UK: McGraw-Hill*, 1991.

[20] M. Rokeach, *The nature of human values*. Free press, 1973.

[21] L. R. Kahle, S. E. Beatty, and P. Homer, "Alternative measurement approaches to consumer values: the list of values (lov) and values and life style (vals)," *Journal of consumer research*, vol. 13, no. 3, pp. 405–409, 1986.

[22] C. J. Clawson and D. E. Vinson, "Human values: a historical and interdisciplinary analysis," *NA-Advances in Consumer Research Volume 05*, 1978.

[23] J. N. Hood, "The relationship of leadership style and ceo values to ethical practices in organizations," *Journal of Business Ethics*, vol. 43, no. 4, pp. 263–273, 2003.

[24] I. Documentation, http://igraph.org/r/doc/reciprocity.html.

Open Science Index, Computer and Information Engineering Vol:12, No:8, 2018 publications.waset.org/10009415.pdf