”Who Mentions Whom?”- Understanding the Psycho-Sociological Aspects of Twitter Mention Network

R.Sudhesh Solomon **Srinivas P Y K L** **Abhay Narayan** **Amitava Das**
Indian Institute of Information Technology, Sri City
{sudheshsolomon.r,srinivas.p,abhay.n,amitava.das}@iiits.in

Abstract

Users in social network 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 - *whether the at-mention or the uni-casting pattern in social media is purely random in nature or there is any user specific selectional preference*? To answer the question we present an empirical analysis to understand the psycho-sociological aspects of Twitter mentions network within a social network community. To understand the psychological pattern we have analyzed personality (Big5 model: *Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism*) of users and to understand the the sociological behavior we analyze values (Schwartz model:*Achievement, Benevolence, Conformity, Hedonism, Power, Security, Self-Direction, Stimulation, Traditional, and Universalism*) of all the users inside a community. Empirical results suggest that personality and values traits are indeed salient cues to understand how the mention-based communication network functions. For example, we notice that achievement-oriented communities talk to each other more often than other people. We also observe that neurotic people are more involved in communication within their community.

1 Introduction

Information diffusion is a 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.

To understand the propagation process we need to understand who is connected with whom and in what manner. At mention on Twitter is the way of one-to-one conversation. The question we raise here is whether the at-mention pattern is purely random in nature or is there any user specific selectional preference? Selectional preference implies to the choice that certain kind of people make for direct communication but, while they are interested in broadcasting they behave differently. To understand the notion this paper presents an empirical analysis to understand the psycho-sociological aspects of Twitter mentions network within a social network community. To this end, we analyze personality and values of all the users in a social network community. First, we categorize social network communities based on values types and analyze mention network within each type (i.e. power, hedonic, etc.) of communities. We notice that achievement-type communities talk to each other more often than other people. Then we analyze how people with certain personality type (i.e. open, extrovert, etc.) interact with other types of people inside a community. We observe that conscientious people are more involved in communication within their network.

Empirical results suggest that personality and values traits are indeed a salient cue to understand how the mention-based communication network functions. In our analysis, we found that the members from stimulation, achievement, and benevolent oriented communities are closely con-

S Bandyopadhyay, D S Sharma and R Sangal. Proc. of the 14th Intl. Conference on Natural Language Processing, pages 418–426, Kolkata, India. December 2017.

nected among themselves while the members in other communities do not show significant connectivity among themselves. Thus in our analysis, we were able to find that universal, extrovert, and open people prefer broadcasting over unicasting messages (via @ mention).

Communication dynamics in human society is a complex phenomenon. Here, in this paper, we present empirical results to establish correlations of the user's unicasting behavior vs his/her psycho-sociological traits. We believe that there are many properties that are not considered (content of the message, age, gender) which affect the at-mention dynamics in the social network. However, we are not considering those aspects of this paper. Our future research is motivated towards that direction.

2 Related Work

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, atmention 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 pieces of information or topics are popular and diffuse the most*, (ii) *how, why and through which paths is the information diffusing, and will be diffused in the future*, (iii) *which members of the network play important roles in the spreading process?*

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 and etc. In the following paragraphs we are describing such related works.

Research endeavors by (Kimura et al., 2010) and (Wani and Ahmad, 2014) 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. (Kimura et al., 2010) explains about combinatorial optimization problem, which is a way to find out the most influential nodes in a social network. In (Wani and Ahmad, 2014) 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 (Gomez Rodriguez et al., 2013), 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 (Romero et al., 2011), 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 (Apolloni et al., 2009) 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 (Zaman et al., 2010) moduled the information diffusion task as a predicting 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 (Yang and Counts, 2010) 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 (Cox and Oakes, 1984) 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 hypothesis 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.

3 Computational Psycho-Sociological Models

In recent years, there have been significant efforts on determining the opinion or sentiment or emotion about a specific topic held by the author of a piece of text, and on automatic sentiment strength analysis of text, classifying it into either one of the classes – positive, negative or neutral, or into Ekman's classes – happy, sad, anger, fear, surprise, and disgust. However, grouping people based on positive, negative, or neutral comments and then understanding their behavior would be spurious. Therefore we propose, psycholinguistic and sociolinguistic models in order to capture user's intrinsic selectional preferences.

The Big 5 personality (Goldberg, 1990) model and Schwartz values (Schwartz, 2012) model can be considered as a person level sentiment model (depicted in table 1) and a societal sentiment model, respectively. Traditional sentiment analysis systems detect sentiment at text-level, whereas the personality model aims at understanding the sentiment/personality of each individual whereas the Schwartz model describes the societal sentiment of groups of individuals forming communities in social networks.

Personality- User Level Sentiment: There has been a growing interest in the scientific community on doing automatic personality recognition from their language usage and behaviour in social media. A milestone in this area was the 2013 workshop and shared task on Computational Personality Recognition (WCPR) (Celli et al., 2013), repeated in 2014 (Celli et al., 2014). Two corpora were released for the 2013 task. One is a Facebook corpus, consisting of about 10,000 Facebook status updates from 250 users, plus their Facebook network properties, labelled with personality traits. The other corpus comprises 2,400 essays written by several participants labelled with personality traits. The best performing system (F-score = 0.73) was developed by (Verhoeven et al., 2013). The various features and methods used by all the participant groups can be viewed as either linguistic and non-linguistic. Another relevant research work on developing computational models for personality on Twitter corpus of 335 users was the (Quercia et al., 2011). They showed that a user's personality traits can be predicted only using three features : following, followers and listed counts.

Personality	Description
Openness	Imaginative, insightful and have wide interest
Conscientiousness	Organised, thorough and planned
Extroversion	Talkative, energetic and assertive
Agreeableness	Sympathetic, kind and affectionate
Neuroticism	Tense, moody and anxious

Table 1: Description for OCEAN Model

Schwartz Values - Societal Sentiment: The societal sentiment model introduced by Schwartz and Bilsky (1990) and modified by Schwartz (1992). The model defines ten basic and distinct personal ethical values (henceforth only values), that respectively are given in the table 2:

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

Table 2: Description for Schwartz Values

The computational Schwartz model has been first proposed by (Maheshwari et al., 2017). 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.

3.1 Psycholinguistic and Network Features

Several different types of features are used depending upon classifiers. An exhaustive set of features include – (**f1**) Word N-grams; (**f2**) POS tags; (**f3**) Lingustic Features (LIWC[1]; Harvard General Inquirer, MRC psycholinguistic feature; Sensicon[2]); (**f4**) Network properties (network size, betweenness centrality, density and transitivity); (**f5**) speech-act classes; (**f6**) sentiment amplifiers (exclamation marks, quotes, ellipses, interjections, emoticons, word/sentence length); (**f7**) misspelt words (SMS slang, stressed words, capitalized words, wrong spellings); (**f8**) presence of *umm/bint* or *abu* in username (a common suffix for women and men respectively in Arabic); (**f9**) Sentiment/Emotion lexica (NRC emotion Lexicon (Mohammad et al., 2013), Sentiwordnet (Baccianella et al., 2010)); (**f10**) Topics words obtained from topic model. A brief overview about the Sociological models and features used are illustrated below.

3.2 Building Classifiers and Performance

We collected data from several sources to build five classification models. Here, for each model,

[1] http://www.liwc.net/

[2] https://hlt-nlp.fbk.eu/technologies/sensicon

Features	Model	F-Score (SVM)	F-Score (LR)	F-Score(RF)
Lexicon	Personality	0.78	0.62	0.65
	Values	0.74	0.59	0.62
+Non-Linguistic	Personality	0.79	0.66	0.68
	Values	0.76	0.61	0.65
+Speech-Act	Personality	**0.80**	0.70	0.71
	Value	**0.81**	0.63	0.67

Table 3: Performance of Personality and Values Models.

we report the best classifier. All the results reported in Table 3 are based on 10-fold cross validation on the respective dataset. **Personality**: A SVM-based model outperforms the state-of-the-art (Verhoeven et al., 2013) by 10%, achieving average F-Score of 79.35%. **Values and Ethics**: A SVM-based values classifiers achieves an average F-Score of 81%. Features used in this model (both personality and values) are reported in Table 3.

4 Semantic Interpretation of Communities

A community in a social network is considered to be a group of nodes densely connected internally and sparsely connected externally. In this paper, we attempt to understand whether individuals in a community possess similar personalities, values and ethical background.

In order to analyze the behaviour of optimists/pessimists at societal level, the egocentric twitter network released by SNAP is used. The Twitter network, released by SNAP (Leskovec and Krevl, 2015) (nodes: 81,306, edges: 1,768,149) has been used to study community structure. We considered 1,562 ground-truth communities (after discarding communities having size less than 5 and with tweets less than 100).

In order to analyze whether people within the same community tend to be homogeneous with respect to their background values/ethics, we measure Shannon's Entropy (measure of the uncertainty) (Lin, 1991) for each dimension separately.

Higher entropy scores suggest lower similarity. To calculate the entropy score vector $X_{(i)}$ for a community $C_{(i)}$ consisting of n users as $u_{(1)}, u_{(2)}, u_{(3)}...u_{(n)}$, a matrix $A_{(i)}$ is created where $A_{(i,j)}$ row vector represents the estimated scores of each of the ten values for a user $u_{(j)}$ and $A_{(i,:,k)}$ column vector represents the estimated scores of k^{th} class for all n users. The $A_{(i,:,k)}$ column vector was transformed to a probability distribution vector $N_{(i,:,k)}$ using softmax-normalization:

$$N_{(i,j,k)} = \frac{exp(A_{(i,j,k)})}{||exp(A_{(i,:,k)})||_1} \quad (1)$$

The entropy score $X_{(i,k)}$ for $N_{(i,:,k)}$ can be calculated using the following formulation:

$$X_{(i,k)} = -\sum_{j=1}^{n} N_{(i,j,k)} * \log N_{(i,j,k)} \quad (2)$$

	AC	BE	CO	HE	PO	SE	SD	ST	TR	UN
$u_{(1)}$	0.91	0.47	0.02	0.07	0.32	0.24	0.65	0.78	0.94	0.10
$u_{(2)}$	0.97	0.40	0.49	0.50	0.56	0.83	0.62	0.73	0.04	0.08
$u_{(3)}$	0.99	0.75	0.50	0.72	0.38	0.60	0.75	0.02	0.57	0.62
$u_{(4)}$	0.77	0.44	0.40	0.16	0.19	0.55	0.73	0.08	0.53	0.25
$u_{(5)}$	0.29	0.02	0.26	0.56	0.41	0.23	0.95	0.02	0.79	0.86
$X_{(i)}$	1.54	1.40	1.40	1.39	1.55	1.50	1.59	0.99	1.42	1.28
$S_{(i)}$	0.87	-0.12	-0.12	-0.19	0.95	0.57	1.26	-2.35	0.00	-0.87
$T_{(i)}$	1.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0

Table 4: Illustrates entropy calculation for values model. Here $T_{(i)}$ represents the binary estimate of fuzzy distribution of values and $S_{(i)}$ represents the zero-mean unit-variance scaled values of $X_{(i)}$ for a community $C_{(i)}$. Similarly, binary estimates for five personality traits $P_{(i)}$ of user $u_{(i)}$ are calculated.

After normalization, $N_{(i,:,k)}$ vector represents the probability distribution of k^{th} value class across n users where entropy score $X_{(i,k)}$ represents the randomness in community along k^{th} value class. The lower the randomness, higher the k^{th} class is dominant in the $C_{(i)}$ community. Now, in order to obtain binary estimates $T_{(i)}$ for each of the ten values and classes in $C_{(i)}$ community, the entropy score vector $X_{(i)}$ is scaled using zero-mean unit-variance method and for numerical values greater than 0, 1 was assigned and for numerical values less than 0, 0 was assigned as class label for $C_{(i)}$ community. Instead of labelling a community $C_{(i)}$ with a class having minimum entropy, the scaling approach is used for the purpose of preserving the fuzzy distribution of values at community level. The obtained $T_{(i)}$ vector represents the fuzzy distribution of values and is thus a representation to capture the semantic information about the community. Having built the model and the classifiers we now try to understand the Psycho-Sociological aspects of the mention network in Twitter.

5 Understanding Psycho-Sociological Aspects of Twitter Mention Network

Network Creation: From the obtained tweets of the SNAP dataset, community level communication networks were created by looking at the @ mention in users tweets. For the network creation, Gephi API[3] is used. In the networks each node represents a user in the network and the edge represents mention link. The users who are never mentioned by somebody and never mentioned someone else were discarded at this stage as they will not contribute anything in understanding the dynamics of intra-community mention network. Once networks were formed, we analyzed their detailed characteristics. For example, let us take into consideration the following network in Figure 1 in order to analyze the following parameters. The nodes in the network represent the users are in the network and the edges represents the connection 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 Figure 1 into consideration we are able to find that there are a total of 10 users labeled from A to J and 6 nodes(users) are connected within the network whereas 4 nodes (users) are disconnected from the network. After we created the network we tried to understand the level of connectivity, after which we analyzed the community type and the personality type along with the mention pattern.

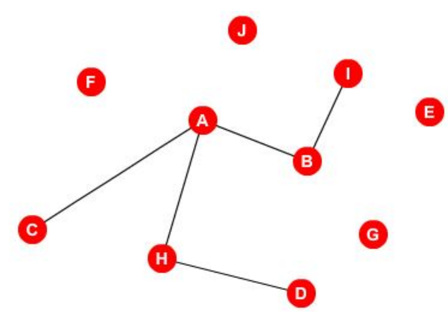


Figure 1: The network of users

Understanding the Level of Connectivity: To understand the dynamics of intra-community mention network we calculate user specific eigenvector centrality (Newman, 2008). Consider the centrality of vertex *i* could be denoted by x_i, could be calculated by making x_i proportional to the average of the centralities of i's network neighbors which is given in the below formula:

$$1/\lambda \sum_{j=1}^{n} A_{ij} x_j \quad (3)$$

, where λ is a constant. Let us assume the exam-

[3] https://gephi.org/

ple of Figure 1 here. In this case the eigen-vector centrality for each node [A, B, C,....I, J] is calculated. The centrality measure is one of the most fundamental measure in the network structure. It is used to determine the central node, and helps in identifying the centralized person who other people are connected with in the social network (Newman, 2008).

Community Type and Mention Patterns: After calculating the eigen-vector centrality at the user level i.e for each nodes, we calculated the eigen-vector centrality for each community. This analysis was done in order to determine how the users behave with each others within their own community. Now, lets say user u_i is connected with *n* number of users $[u_j, u_k, ..., u_n,]$ via mention links within a community. From Figure 1,we are able to notice that A is connected to C,H,B and I. D is connected indirectly to A via node H and hence we consider D only while calculating the eigen-vector centrality for node H. Now for user A, we calculate pair-wise eigen-vector centrality between each pair of users [(A, B), (A, C), (A, H), (A, I)] and obtain a vector for the user A. To get the final average intra-community connectivity score for user A, all the *n* scores are further averaged by dividing their sum by *n*. The score that is obtained now is the average connectivity of the user A within the community. Following this method we obtain connectivity score of all the users within the community and those scores are then further averaged by the total number of users (excluding users who never got mentioned or never mention someone else) within the community. This score obtained is the average score of the community for that particular user. Similarly, average connectivity scores are calculated for each community in the SNAP network and further those obtained scores are averaged based on community type (i.e., power, hedonic, etc.). These category-wise average connectivity scores are finally reported to understand the intra-community psycho-sociological aspects of the Twitter mention network.

Personality Type and Mention Patterns: To understand who is mentioning whom we consider user specific average eigen-vector centrality. For example, in the first iteration we take all the open type people and find out their average eigen-vector centrality over all communities. Then we further divide them into 5 classes (i.e., open, conscientious, etc.). Thus class-to-class average connectivity (i.e., open-open, open-conscientious, open-extrovert, open-agreeable, and open-neurotic) was obtained. This process was repeated for all other personality types.

Finally, for more granular understanding the eigen-vector centrality scores were divided into three bands - high, mid, and low. These values were then scaled between 0-100 by looking at the overall connectivity score (high, low) distribution at corpus level as shown in Figure 3.

6 Obtained Mention Network Patterns

We present our findings in three parts. First part details psycho-sociological patterns of the mention network, whereas the second part tries to answer the question – *who-mentions-whom*? Finally we try to analyze the relationship between closeness and reciprocity of the community with different community sizes.

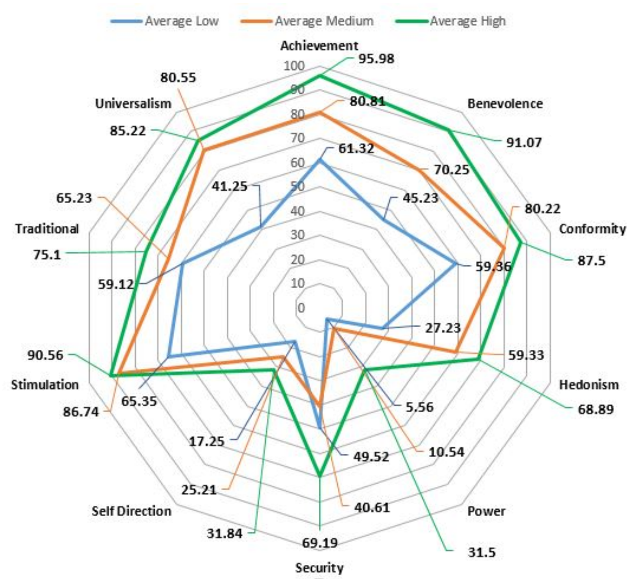
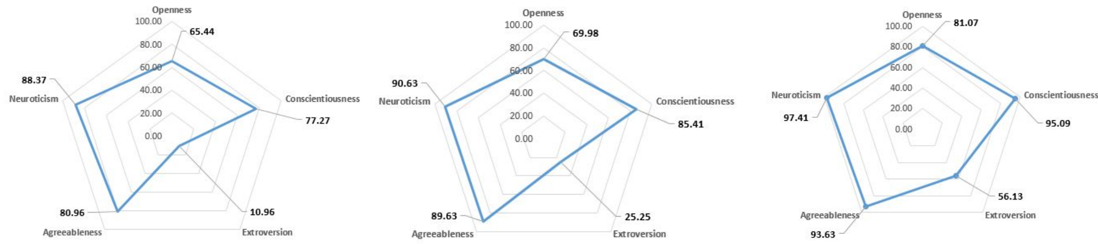


Figure 2: Communication within each community of the values model

Psycho-Sociological Patterns of Mention Network: Figure 2 reports obtained results of our empirical analysis. Results indicate that members from stimulation, achievement, and benevolent communities are closely connected among themselves while the members in other communities do not show significant connectivity among themselves.

We also observe that people who are independent, i.e, self-directed do not involve much in connecting themselves with other members in their



(a) Low connectivity (b) Medium connectivity (c) High connectivity

Figure 3: Psycho-sociological patterns of Twitter mentions network

community. Security oriented people i.e., those who follow strict rules and regulations are found to have reasonably balanced connectivity. People belonging to the traditional groups who follow rituals blindly, are loosely connected to the world. One significant observation was that the universal people, who are the people tending to be more inclined towards social justice and tolerance show high connectivity with other people in the community. Further analysis reveal that in general universal people tend to unicast (i.e., @mention someone specific) messages rather than broadcast , which is justifiable to their nature. The members of power-oriented communities are those who seek to dominate other people in their community and hence the communication is low among these people.

Node	Closeness
A	0.40
B	0.31
C	0.25
D	0.21
E	0.0
F	0.0
G	0.0
H	0.31
I	0.21
J	0.0

Table 5: Closeness Centrality for Figure 1

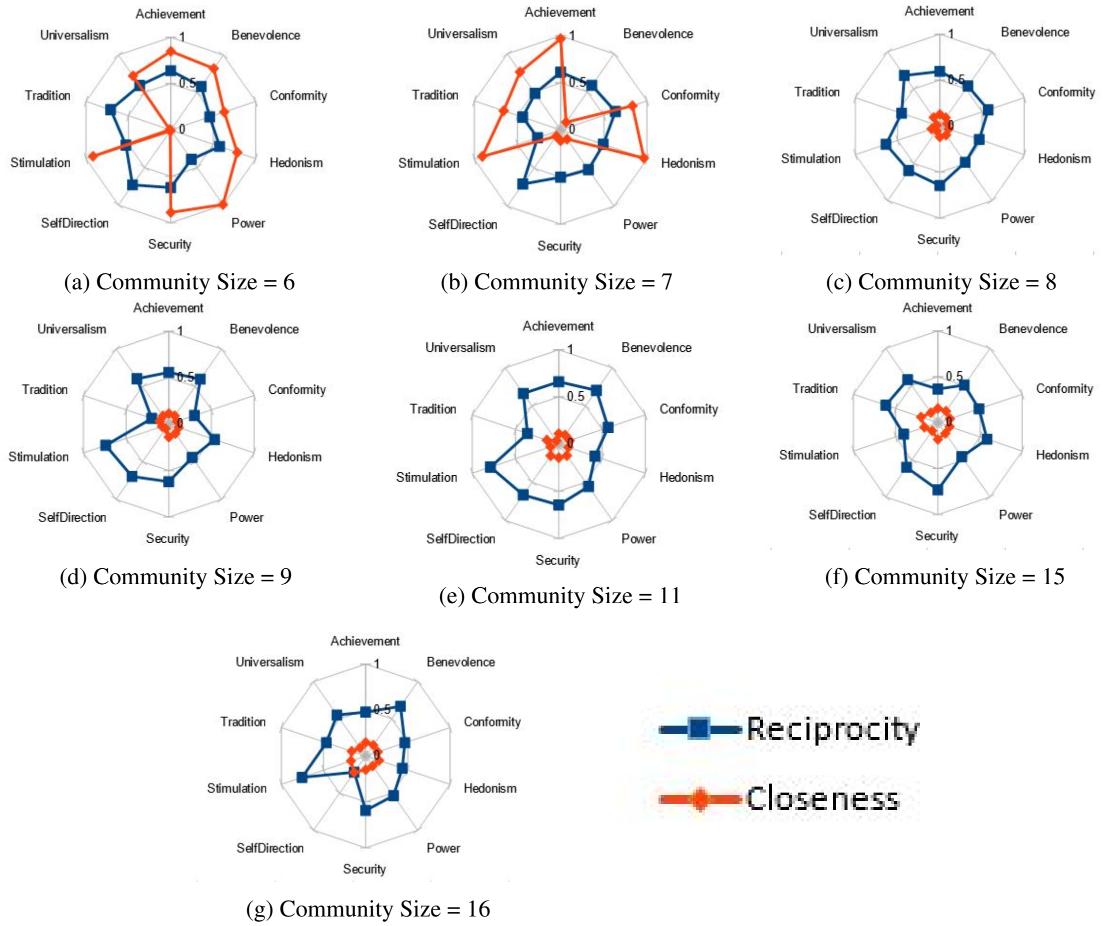
Who-Mentions-Whom: Results of this analysis is reported in Figures 3a, 3b and 3c. The result is presented in three sets i.e., low, medium, high connectivity. We notice that in the highly connected communities, the neurotic people who are mostly tense, moody, anxious are more connected to other people in the network. Agreeable and conscientious people also maintain a good relationship with others. In the case of medium and low connected communities , the neurotic people tend to maintain good connectivity with others than people possessing other values. We also infer that extroverts (high,mid and low) also tend to broadcast their messages rather than sending it to someone specific, therefore their connectivity is low in the mention network.

Similar practice has also been noticed by open people in low connected group. Therefore, we can conclude universal, extrovert, and open people prefer broadcasting over unicasting messages (via @ mention)

Closeness vs Reciprocity: In a connected graph,the closeness or the closeness centrality of a node is used to measure how close a particular node is with respect to other nodes in the 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. It can be seen that, the more central the node is, the more closer it is to the other nodes.

For example, in Figure 1 we are able to find that node A is more closely associated with other nodes and has the highest closeness centrality among the other nodes. It is also observed from Table 5 that those nodes which are not connected in the network shown in Figure 1 are having 0 as their closeness centrality measure.

The social norm of reciprocity is the expectation that people will respond to each other in similar ways. Therefore in a mention network, if a particular user mentions another user in his/her tweet and the other user on the other hand mentions him/her back then we can find that reciprocity can be achieved between those two users. The result is provided as an analysis of closeness vs reciprocity for various sizes of the community in Figure 4. From the analysis, we observe that as the size of the community increases the reciprocity de-



(a) Community Size = 6 (b) Community Size = 7 (c) Community Size = 8

(d) Community Size = 9 (e) Community Size = 11 (f) Community Size = 15

(g) Community Size = 16

Figure 4: Closeness vs. Reciprocity for various community sizes

creases and the closeness increases. This is because the closeness centrality is calculated for an entire network and hence as the size of the network increases the closeness tend to increase.

Reciprocity on the other hand is calculated for the mentions network and as the size of the network increases there is a high chance that two users do not mention each other in their tweets. Let us consider one example from our analysis, here we consider the self-directed community of people of community size 15 and 16 respectively. From Figure 4 we are able to find that as the size of the community increases from 15 to 16 the reciprocity decreases and the closeness increases.

7 Conclusion and Future Work

This paper presents an empirical analysis to understand the psycho-sociological aspects of Twitter mentions network within a social network community. Here, we take the explanatory approach; however, we strongly believe that the obtained empirical results could be further used to predict individual/group communication behavior. We would be woking on finding similar patterns in (i) Twitter favorite network i.e., *who has liked whom*, and (ii) Retweets network i.e., *who retweets whose tweet*. We would also like to understand inter-community mention network pattern – i.e., is there any selectional preference when someone chooses to communicate with someone outside his/her community? We believe that this kind of models may become extremely useful in the future for various purposes like Internet advertising (specifically social media advertising), community detection, computational psychology, recommendation systems, sociological analysis over social media.Now that we have calculated various measures and obtained analytical results. In the future, based on these results given a community we try to predict who will mention whom?

References

Andrea Apolloni, Karthik Channakeshava, Lisa Durbeck, Maleq Khan, Chris Kuhlman, Bryan

Lewis, and Samarth Swarup. 2009. A study of information diffusion over a realistic social network model. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, pages 675–682. IEEE.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204.

Fabio Celli, Fabio Pianesi, David Stillwell, and Michal Kosinski. 2013. Workshop on computational personality recognition (shared task). In *Proceedings of the Workshop on Computational Personality Recognition*.

Fabio Celli, Bruno Lepri, Joan-Isaac Biel, Daniel Gatica-Perez, Giuseppe Riccardi, and Fabio Pianesi. 2014. The workshop on computational personality recognition 2014. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 1245–1246. ACM.

David Roxbee Cox and David Oakes. 1984. *Analysis of survival data*, volume 21. CRC Press.

Lewis R Goldberg. 1990. An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216.

Manuel Gomez Rodriguez, Jure Leskovec, and Bernhard Schölkopf. 2013. Structure and dynamics of information pathways in online media. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 23–32. ACM.

Masahiro Kimura, Kazumi Saito, Ryohei Nakano, and Hiroshi Motoda. 2010. Extracting influential nodes on a social network for information diffusion. *Data Mining and Knowledge Discovery*, 20(1):70–97.

Jure Leskovec and Andrej Krevl. 2015. {SNAP Datasets}:{Stanford} large network dataset collection.

Jianhua Lin. 1991. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151.

Tushar Maheshwari, Aishwarya N Reganti, Upendra Kumar, Tanmoy Chakraborty, and Amitava Das. 2017. Semantic interpretation of social network communities. In *AAAI*, pages 4967–4968.

Saif M Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242*.

Mark EJ Newman. 2008. The mathematics of networks. *The new palgrave encyclopedia of economics*, 2(2008):1–12.

Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. 2011. Our twitter profiles, our selves: Predicting personality with twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 180–185. IEEE.

Daniel M Romero, Brendan Meeder, and Jon Kleinberg. 2011. 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*, pages 695–704. ACM.

Shalom H Schwartz and Wolfgang Bilsky. 1990. Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of personality and social psychology*, 58(5):878.

Shalom H Schwartz. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in experimental social psychology*, 25:1–65.

Shalom H Schwartz. 2012. An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, 2(1):11.

Ben Verhoeven, Walter Daelemans, and Tom De Smedt. 2013. Ensemble methods for personality recognition. In *Proceedings of the Workshop on Computational Personality Recognition*, pages 35–38.

Mudasir Wani and Manzoor Ahmad. 2014. Survey of information diffusion over interaction networks of twitter. *International Journal of Computer Application*, 3(4):310–313.

Jiang Yang and Scott Counts. 2010. Predicting the speed, scale, and range of information diffusion in twitter. *ICWSM*, 10:355–358.

Tauhid R Zaman, Ralf Herbrich, Jurgen Van Gael, and David Stern. 2010. Predicting information spreading in twitter. In *Workshop on computational social science and the wisdom of crowds, nips*, volume 104, pages 17599–601. Citeseer.