Analysis Framework for Initiating Dynamic Behaviors in Friendship Network

R.Venkatesan[#1], M.Selvakumar[*2], S.Nandhagopal[#3], R.Gopal[#4]

[#] *Assistant Professor & Department of Computer Science and Engineering & KSR Institute for Engg & Tech.*

[#] *Assistant Professor & Department of Information Technology & KSR Institute for Engg & Tech.*

[#] *Assistant Professor & Department of Information Technology & KSR Institute for Engg & Tech.*

 # Assistant Professor & Department of Computer Science and Engineering & KSR Institute for Engg & Tech.

Abstract *— In this digital world, social media can be a valuable addition to a department's communications strategy. Because most of the people's have expressed an interest in developing and maintaining a digital relationship among them in social web environments like Facebook, Linkedin, Twitter etc. Such kind of relationships might be in personal and professional capabilities. Social media is an internet-based form of communication. It allows users to have conversations, share information, community formations and create web content. Basically in social network the relationship among people can be expressed in 3 forms such as, one way communication, mutual communication and maintained relationship. Today conversations are going in the form of either mutual or one to many i.e., Group chat. While conversing in social molecule, cannot ensure that all the people in the community may not be close with each other. In general social molecules are formed with similar people by considering any common characteristics among them. So that some close friends, some acquaintances and some anonymous people's also can form molecule. The mode of conversations might be either in the form of text based communication or using built in emojis or Stickers. When conversing in public forum each and everyone must follow the policy of official communications. Here the system analyze whether people's are following such policies or violating. If violations occur in a social molecule then the system ensure the atom to follow the policy by analyzing their conversations. Text messages can be analyzed by using Text Categorization techniques. By considering the level of violation the system will initiate the dynamic change in molecule by excluding an atom and giving warning massage. This system also analyses the mutual communication and calculate the internal score of each atom to maintain the relationship or better to avoid.*

Keywords *- Text Categorization, Social Networks, Social Media, Sociogram, Sentiment Analysis, Homophily, Betweeness Centrality.*

I. INTRODUCTION

Social networking on social media websites involves the use of the internet to connect users with their friends, family and acquaintances. Social media websites are not necessarily about meeting new people online, although this does happen. Instead, they are primarily about connecting with friends, family and acquaintances you already have in real life. The most well-known social media sites are Facebook, Twitter, Instagram and LinkedIn. These sites allow you to share photos, videos and information, organise events, chat, and play online games.

Often, each of your "friends" (Facebook) or "followers" (Twitter) will be connected to each other. Just like in real life, the connections between people aren't just one-on-one, but a network of connections. This online social network is useful for spreading information, pictures and videos and generally staying in touch with people you wouldn't normally get to interact with all the time. For example, you can easily set up a Facebook page with details and pictures of an event you might be planning, such as a school fete. The page allows you to easily send out invitations to other users of the social media site.

Social media sites have a variety of privacy settings you can adjust. This means you can control who sees your profile page and other information you share on the site. Some people do not mind having their personal information available for anyone to view online. However, the system strongly recommend that you don't publish your home address and be mindful of posting other personal information about yourself (including your birthday), or others especially if you don't have their permission.

II. RELATED WORKS

Kyunglag Kwon, Yunwan Jeon et al. developed a system called Sentiment Trend Analysis in Social Web Environment[1]. They used the ACO algorithm and SentiWordNet. Their system compute the users sentiment scores for the computed pheromone values with respect to the sentiment words with SentiWordNet. They were developed the system by considering the online marketing to analyze and support many kinds of product selling companies. They didn't consider the formation of friendship network.

Kerstin Denecke developed multilingual sentiment analysis using SentiWordNet. He was leading the system to identify the polarity level by considering the positive score of a document. The positivity had been calculated with respect to the sentiment bearing words like adjectives. The system deals with the document classification.

Hao Ma, Tom Chao Zhou, et.al[2] developed a system for improving the recommender system by incorporating social contextual information. They designed a factor analysis approach based on the on probabilistic matrix factorization to alleviate the data sparsity problem and poor prediction accuracy problems by incorporating social contextual information, such as social networks and social tags. They deals with the recommender systems of academia and industry. They didn't apply their methods to analyze the dynamic behavior of friendship network.

Jinpeng Wang, Wayne Xin Zhao et.al[3] they are deal with the review of online customers in e-commerce applications. They gave higher importance to identify the demographic information of different kinds of product adopters. They used the weighted regularized matrix factorization techniques.

Alvaro Ortigosa , José M et.al[4] presented a method to analyze the facebook massages written by users, supports: i) to extract information about the user sentiment polarity level (positive, negative or netural) ii) to model the user's usual sentiment polarity level to detect the emotional changes. They implemented the method in SentBuk. SentBuk method retrieve the massage from user and classify them according to the polarity levels. By the level of polarity their system will initiate the activities to be tackled at each time by considering the user emotional state. System deals with the individual user massage or post polarity level in Facebook.

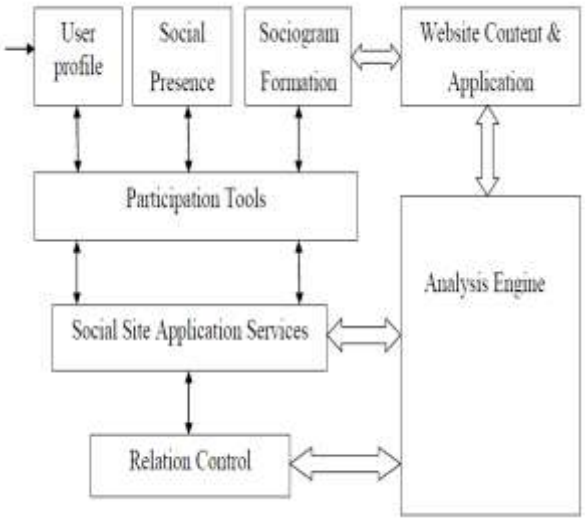
Alessia D'Andrea, Fernando Ferri et.al[5] presented a survey of Sentiment Analysis Applications. They conducted survey to tackle data and analyze based on the different kinds such as business data, politic, public actions and finance. They presented the different Sentiment classification approaches such as, machine learning, lexion based and hybrid approach. Tanvi Hardeniya1 D. A. Borikar[6] conducted a comparative study of sentiment analysis techniques which uses various algorithms. They developed a framework of sentiment analysis using dictionary based approach to calculate a polarity score of a sentence.

III. PROPOSED METHOD

The proposed system uses the ANN - Artificial Neural Network and Text Categorization techniques. The overall system has been divided into two layers such as, decision making layer and application layers. In this system all the internal functionalities has been considered for analyzing the bulling words. The internal functionalities covers posting, commenting and chatting in social network. Initially the system has been trained with the two kinds on training datasets namely bulling and polite and which are mapped with semantics using ANN technique. The training datasets are dynamic over the web.

After that system applies the text categorization technique on user post, comments and chatting in mutual case or in public forum. Once text categorization applied the system computes the values of 3 classes namely positive, negative and neutral class. With respect to the majority of the percentage of respective functionality, the dynamic behavior will be initiated in friendship network.

System Architecture



Architectural Description

Initially the user can sign up in the social network. The user need to provide all the personal and security information for the user profile. After creating an account, the users can login and search for new friends and they may chat among their friends. User profiles and conversations are maintained as individual profile. Data mart has two different dataset called polite words,

bullying words and friendship network information's are maintained. In analysis engine the system runs text categorization algorithm to monitor all the conversations between friends. If a person use bullying words in chat or posting messages in their timeline. The system will send warning message sent to the user depends on the usage of bullying words and update the weight of individual profile. If social molecule formed the conversation among the molecule also analyzed whether conversation goes polite or not. If a people found that they are using bullying words depend on the degree of usage they will be intimated and excluded from community.

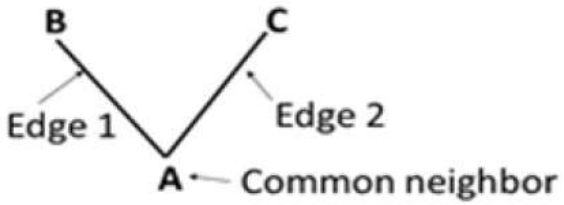
In case a person is found with highest degree of bullying words, the user may block. In another side weight age of individual profile updated depends on their comments and chats.

The entire system has been divided into five modules,

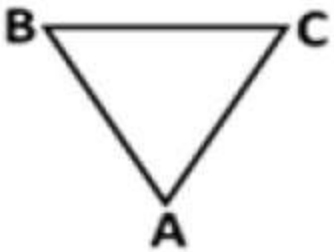
A. Friendship Network Formation
B. Analyzing Mutual Communication
C. Identify Betweenness Centrality
D. Analyze Homophily
E. Dynamics of FFB (Friendship Formation and Behaviour)

A. *Friendship Network Formation*

The formation of friendship network has been initiated by randomly inter connecting two peoples to form network. To create such a friendship network it should follow a property called triadic closure. It is all about defining how so far peoples are close to each other among them. Triad means a person knows two people's which called triad.



The triad becomes triadic closure if two people in social network have a friend in common, then there is an increased likelihood that they will become friends themselves at same point in the future.



To achieve such triad becomes triadic closure the network is initialized with N nodes, each node having one link to a randomly chosen node among N.

B. *Analyzing Mutual Communication*

In general communication in social network has been categorized into three ways such as, Mutual Communication, One way Communication, Maintained Relationship. A link represents mutual communication, if the user both sent messages to the friend at the other end of the link, and also received messages from them during the observation period. A link represents one-way communication if the user sent one or more messages to the friend at the other friend of the link (whether or not these messages were reciprocated). A link represents a maintained relationship if the user followed information about the friend at the other end of the link, whether or not actual communication took place; "following information" here means either clicking on content via Facebook's news feed service or visiting the friend's profile more than once.

The system analyses the mutual communication among the friendship network by applying text categorization techniques. Usually conversations are goes in text mode.

Text Categorization

Text categorization (TC)[7][8] –is a given set of categories (Bulling, Polite) and a collection of text documents, the process of finding the correct topic for each document.

The task of approximating an unknown category assignment function

$$F : D\ X\ C\ \rightarrow \{0, 1\}$$

Where,

D - Any chosen functionality of Facebook activity.(Post, Comment, Chat)

C - Training datasets

$$F(d,c) = \begin{cases} = 1\,; iff\ d \in c \\ = 0\ ;\ Otherwise \end{cases}$$

The approximating function,

$$M : D\ X\ C \rightarrow \{0,1\}$$

is called a classifier, and the task is to build a classifier that produces results as "close" as possible to the true category assignment function F.

The system takes a description of an instance, $\in X$, Where X is the instance language or instance space. A fixed set of categories: $C = \{c1, c2, c3, \ldots, cn\}$

To determine category of $x : c(x) \in C$, where $c(x)$ is a categorization function whose domain is X and whose range is C.

After the categorization process system follows text indexing. It follows Supervised Learning - process of applying the known true category assignment function on the training set. Indexing of Texts Using Controlled Vocabulary, The documents according to the user queries, which are based on the key terms. The key terms all belong to a finite set called controlled vocabulary. The task of assigning keywords from a controlled vocabulary to text documents in a alphabetical order is called text indexing.

A set of sufficient conditions for a document to be labeled with a given category bulling or polite is defined.

If DNF (Disjunction of Conjunctive Clauses) formula then category else ¬category, Such rule may look like the following,

If ($(journey$ & $trip)$ or $(journey$ & $tour)$ or $(journey$ & $ride)$ or $(journey$ & $safari)$ or $(journey$ & $outing$ ¬ walk)) $then$ journey else ¬ journey

At last system calculate the internal score of their activity in web environment. Based on the score dynamic behavior is initiated.

An internal score of an atom can be measured from the bulling term frequency. By following the Boolean frequency[9] measure the bulling term frequency is calculated as,

$$tf(t,d) = \begin{cases} = 1\,; if\ t\ occurs\ in\ d \\ = 0\,;\ Otherwise \end{cases}$$

If the frequency of a bulling term is greater than 5 times then the system will give the 1% to atom.

$$if \sum_{i=1}^{5} t_i = \begin{cases} t_i += 1; if\ t_i == 5 \\ 0 \quad ; \quad Otherwise \end{cases}$$

The summation of percentage reaches closer to the 60% then initiate the dynamic behavior over the web which is discussed in dynamics of FFB, remains the same otherwise.

C. *Identify Betweenness Centrality*

In a randomly formed friendship network there might be chance to form different communities or clusters within the larger network to identify such communities. We have to calculate betweenness the centrality of a larger network. This centrality can be measured with metrics like degree, closeness. Networks can vary in the way node-to-node transmission takes place, or in the way a course through the network is taken. The edge which has highest betweenness centrality will say that there is a community in the network. This shows that weak tie between communities, Strong communication between them.

$$BC(v) = \sum_{u,v \in V} \left(\frac{\sigma_{u\,w}(v)}{\sigma_{u\,w}} \right)$$

$\sigma_{u\,w}$ - Total number of shortest paths between node u and w.

$\sigma_{u\,w}(v)$ - Total number of shortest paths between node u and w that pass through v.

Betweenness centrality for all the nodes and the paths that it travels from the initial to final stage. If there exists highest degree of edge in a graph then we can conclude that there is a community occurs.

D. *Analyze Homophily*

Once betweenness centrality has been identified then the next step is to test the degree of homophily. In fact homophily is so powerful a principle that whole communities on Facebook by considering some common characteristics—genders, races, ethnicities, ages, class backgrounds, educational attainment, etc.—appear to have very different qualities. Also some of the friendships gets formed based on the life style similarities[10] between peoples and considering the recommended friends on their pages.

$$Homophily = 1 - \left(\frac{Actual\ No.of\ Edges}{Expected\ No.of\ Edges} \right)$$

Here a common characteristics ethnicities or city or educational institutions has been taken to identify the degree of homophily.

Homophily test can be defined as, if the fraction of cross-gender edges is significantly less than 2pq, then there is belief for homophily.

where,

p - probability of male gender

q - probability of female gender

p^2 - probability of both the end of edge will be male gender

q^2 - probability of both the end of edge will be female gender

$$if\ H\ is \begin{cases} = 0\ ; There\ is\ no\ homophily \\ < 0\ ; Hetrogeneity \\ > 0\ ; Homophily \end{cases}$$

E. *Dynamics of FFB (Friendship Formation And Behavior)*

Dynamic changes are initiated by computing the clustering coefficient of a person and average clustering coefficient of a community.

A clustering coefficient is a measure of the degree to which friendship relation in a graph tend to cluster together. Evidence suggests that in most real-world networks, and in particular social networks, friendship tend to create tights knit groups characterized by a relatively high density of ties; this likelihood tends to be greater than the average probability of a tie randomly established between two friends.

Clustering coefficient is a local measure. It has been calculated as clustering coefficient of a person by using following formula:

$$C_i = \frac{2L}{K_i(K_i - 1)}$$

Where,

Ki- is the degree of node i.

Li - is the number of edges between the Ki neighbours of node i.

The clustering coefficient of entire graph is average clustering coefficient of entire graph and can be calculated as:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

$$If\ ACC = \begin{cases} = 0; There\ is\ no\ strong\ connection\ between\ community\ people \\ = 1; Everyone\ of\ community\ members\ are\ connected\ with\ Everybody \end{cases}$$

In the above case the value of clustering coefficient = 1 or closure to 1 the community will not be destructed and will not be notified, Even though the usage of harness degree is higher and the individual score becomes greater than 60%. If a candidate doesn't have strong relation with all in the community and clustering coefficient also poor, then they will be notified or excluded from the public forum. Then implicitly candidate advised to have the personal converse with their friend.

IV. EXPERIMENTAL RESULT

A. *Chat Analysis using Text Categorization*

The system has been experimented with a conversation between two friends in real time. Then text categorization and Boolean frequency measure were applied on the chat to know whether their chats uses bullying or polite. In case the usage of bullying words in chat becomes higher than the threshold value i.e., 40% then the system will sent warning message to the candidate. If they use bullying words continuously i.e., 60% automatically they will be blocked.

Table 1 Chat Analysis

FRIEND 1	FRIEND 2	FRIEND 1		FRIEND 2	
		Bullying %	*Polite %*	*Bullying %*	*Polite %*
Hello	Hi	0%	80 %	0%	85%
How are you?	I am good, what about you?				
I am also good.	So what are you doing these days?				
I work with an MNC in Noida.	Wow! That's great.				
Bye	Bye				

B. *Clustering Analysis*

The system is experimented with 240 users and friendship between them. Depends on the post, comment, chat are analyzed with the training dataset. The analysis process can be carried out by using text categorization and Boolean frequency analysis techniques. By considering the degree of harness the system will intimate the friend to use soft and polite words. It has been analyzed in both personal chat between individuals and public chat in a forum i.e., community or clustering. In public forum, the usage of harness will be analyzed and notified depends on the clustering coefficient of individuals and averaging clustering coefficient of whole community.

This means that as per the system, the dynamic changes will be prohibited since their average is closure to 1. Even though the usage of harness word is higher.

Table 2 Cluster Analysis

Friends id	Cluster co-efficient	Average
A	0.93	0.95
B	1	
C	0.93	
D	0.93	
E	0.93	
F	0.93	
G	1	

If a candidate doesn't have strong relation with all in the communities, clustering coefficient also poor which means that they will be notified or excluded from the public forums. Then implicitly candidate advised to maintain personal converse with their friends.

V. CONCLUSION AND FUTURE ENHANCEMENT

The developed web application yield to create well-formed friendship network over the web. Friendship network is dynamic one since whenever new friends may be added or deleted. With this application friends can share the news and post in their timeline. Similar interested friends can form social molecule in web and by considering any common parameter can form either homogeneous or heterogeneous molecule. Friends can do chat with their friends and they can converse in molecule as well. The system analyzed and experimented with the chat between two friends and conversation in social molecule. To analyze the chat the system uses the text categorization technique to know degree of bullying and politeness word. It will update the internal score of each friend based on the comments, like, and chat. The social molecule has been measured with clustering coefficient of each atom. The dynamic change in friendship network has been initiated by considering both average and atom clustering coefficient. People notified or blocked by the system if the degree of bullying words in chat is higher than the threshold value. The whole system will give dynamic updation in friendship network by considering the internal score of each atom. In future the application can consider the various parameters like different form of liking, commenting and symbolic language such as emojis to calculate the internal score of an atom. Can rank the profiles based on score to recommend a friend to achieve the triadic closure property and can create the complete friendship network.

REFERENCES

[1] Kyunglag Kwon, Yunwan Jeon, Chanho Cho, Jongwoo Seo, and In-Jeong Chung., " Sentiment trend analysis in social web environments" in IEEE Transaction On Knowledge And Data Engineering, Vol 28, No 7, March 2017.

[2] Hao Ma, Tom Chao Zhou, " Improving Recommender Systems by Incorporating Social Contextual Information" in ACM Transactions on Information Systems, April 2011

[3] Jinpeng Wang, Wayne Xin Zhao, " Leveraging Product Adopter Information from Online Reviews for Product Recommendation" in Ninth International AAAI Conference on Web and Social Media, 2015

[4] Alvaro Ortigosa *, José M. Martín, Rosa M. Carro., " Sentiment analysis in Facebook and its application to e-learning" in Elsevier - Computers in Human Behavior, August 2013

[5] Alessia D'Andrea, Fernando Ferri, Patrizia Grifoni, Tiziana Guzzo, " Approaches, Tools and Applications for Sentiment Analysis Implementation" in International Journal of Computer Applications (0975 – 8887) Volume 125 – No.3, September 2015

[6] Tanvi Hardeniya1 D. A. Borikar, " An Approach To Sentiment Analysis Using Lexicons With Comparative Analysis of Different Techniques" in IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 3, Ver. I (May-Jun. 2016).

[7] S.Niharika, V.Sneha Latha, D.R.Lavanya, "A Survey on Text Categorization" in International Journal of Computer Trends and Technology- volume3Issue1- 2012

[8] Manasa M Tilakraj, Deepika D. Shetty et.al, "Emotion Finder: Detecting Emotions From Text, Tweets and Audio " in IJSEAS – Volume-2, Issue-5,May 2016

[9] Parul Kalra Bhatia, Tanya Mathur, Tanaya Gupta, " Survey Paper on Information Retrieval Algorithms and Personalized Information Retrieval Concept" in International Journal of Computer Applications (0975 – 8887) Volume 66– No.6, March 2013.

[10] Mayuri Suresh Pawar, Shraddha Popat, " Friendbook: A Scalable and Efficient Way to Recommend Friends on Social Networks through Life-Style" in International Journal of Engineering Science and Computing, April 2017.