Data Mining Technique for Tracking of Information Diffusion in Online Social Network

Prakash Chandra Behera

Assistant Professor, Department of Computer Science, St.Claret College, Bangalore, India

***Abstract*: - Information diffusion is common in online social network .There are two different ways of how information reaches a person in a network. Information reaches us through connections in our social networks, as well as through the influence external out-of-network sources, like the mainstream media. Social media provides the means of interactions among people in which they create, share, and exchange information and ideas in virtual communities and networks. Online social networks play a major role in the spread of information at very large scale.**

Information diffusion is the communication of knowledge over time among members of a social system. Online social networks allow hundreds of millions of Internet users worldwide to produce and consume content. They provide access to a very vast source of information on an unprecedented scale. They also play a major role in the diffusion of information and have proven to be very powerful in many situations, like Facebook during the 2010 Arab spring or Twitter during the 2008 U.S. presidential elections. Still, the raw data produced by users of these networks. Tracking of information diffusion is a Very difficult task but by observing various factor of information diffusion, we can able to track the propagation information and we can predict about information propagation.

In this research paper we discuss some issue of data mining in social network and also we go through various model given by different researcher, we analysis these model and short out few weakness for each model. This will help the other researcher for extending their research work.

***Keywords*:-online social network (OSN), Information diffusion, Data mining, Information tracking.**

I. INTRODUCTION

Information diffusion is the communication of knowledge over time among members of a social system. The information diffusion field investigates how information (news, rumors etc) propagates among people. The diffusion process involves the following three elements:

1. Sender(s). A sender or a small set of senders initiate the information diffusion process.
2. Receiver(s). A receiver or a set of receivers receive diffused information. Commonly, the set of receivers is much larger than the set of senders and can overlap with the set of senders.
3. Medium. This is the medium through which the diffusion takes place. For example, when a rumor is spreading, the medium can be the personal communication between individuals.

Online Social Networks (OSNs), a revolutionary change has occurred in the social interactions of people of this decade. Many popular OSNs such as Facebook, Twitter, and LinkedIn have become increasingly popular. Nowadays, these OSNs allow many easy-to-learn online activities including chatting, online shopping, gaming, tweeting, etc. According to the site *henextweb.com*, Indian citizens spend one in four minutes online using social networking sites, more than any other Internet activity. In fact, social networking is considered to be the second-fastest growing activity, behind only entertainment. However, social media sites provide data which are vast, noisy, distributed and dynamic. Hence, data mining techniques provide researchers the tools needed to analyze such large, complex, and frequently changing social media data.

II. MINING DIFFICULTIES IN SOCIAL NETWORK:

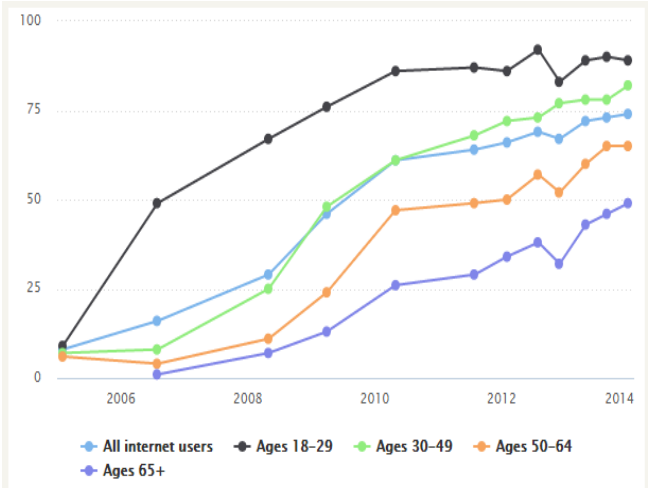
There are number of issue for mining in social network, which are

1. Age group mining in Social Network
2. Central control System in Network
3. Mental and psychological status of user in network
4. Influence spreading in network
5. Predicting Trust and Distrust among Individuals
6. Link Prediction
7. Lack of expert in network
8. Privacy in network
9. Sentimental community network
10. Lack of knowledge about information propagation in Social network

A. Age Group Mining in Social Network

In social network different age user are accessing their information, but the age at which a user are accessing social media is in between 15 to 60. But after that they do not have that mush use of social media. For example people of age more than 15 year they start using social media application and again people after age 60 year they have not that much interest to use social media. So whatever the activity in social networking is going on like hacking information, various social media crime, wrong data diffusion and etc

The percentage of online adults who use social networking sites has steadily risen. As of January 2014, 74% of all online adults use social networking sites. For adults ages 18-29, 89% of them use social networking sites. For adults ages 30-49, 82% of them do. For adults ages 50-64, 65% of them do, and for adults ages 65+, 49% of them use social networking sites. As per study report of Pew Research Center use of Social network is increasing



B. Central Control System in Network

Online social networks have witnessed phenomenal growth in popularity to an extent that today two thirds of the world's Internet population participates in some form of online social networking, Current social networks require users to place absolute faith in their operators, and the inability of operators to protect users from malicious agents has led to sensitive private information being made public. Many may not think about it, but using a social network ought to be a sobering activity, as it requires the user to place absolute faith in the Social Network Operator (SNO).In social network there is no such central control over all social network activity.

C. Mental and Psychological Status of User in Network

Mental and psychological status of user in social network is an important things in information diffusion, Discovering human mental and human interaction based on data mining techniques is also an interesting research field that is gaining huge attention in research. Here, human behavior may indicate any human-generated actions such as clicking on a specific advertisement, accepting a friend's request, joining a group or discussion forum, commenting on an image, music, etc, or dating with a person, mental status also play an important role for information diffusion in social network.

D. Influence Spreading in Network

Nowadays, as OSNs are attracting millions of people, the latter rely on making decisions based on the influence of such sites. For example, influence propagation can help decide which

movie to watch, which product to purchase, and so on. Thus, influence propagation has become an important mechanism for effective viral marketing, where companies try to promote their products and services through the word-of-mouth propagations in OSNs. This further motivates the research community to carry out extensive studies on various aspects of the influence propagation problem.

E. Predicting Trust and Distrust among Individuals

Due to the continuous expansion of communities in OSNs, the question of trust and distrust among individuals in a community has become a matter of great concern. Past assessments reveal that some users try to either disturb or take undue advantage of the normal atmosphere of such online communities. As a result, there arises a need of assessing each user of an OSN community to predict the level of trust or distrust that can be computed for them.

F. Link Prediction

The bulk amount of data available in OSNs can be mined to make predictions about 'who is a friend of whom' as an individual might be only a few steps away from a desirable social friend but may not realize it. By gathering useful information about an individual, OSNs can infer new interactions among members of an OSN that are likely to occur in the near future.

G. Lack of Expert in Network

OSNs consist of several experts in a specific domain and other people who join the network to receive help from these experts. These OSNs can be used to search for such experts within a group of people. For example, a topic related expert can be searched based on the study of the link between authors and receivers of emails.

H. Privacy in Network

The security of private information of users online is a critical topic, particularly since social networking applications became popular, Many people increasingly utilize social Networks, such as Facebook, MySpace, Twitter, Orkut, Linked in and etc. These networks allow users to publish details about themselves and their lives and also connect to their friends and colleagues. However some of the information revealed in these networks should remain private and not published at all. Companies that operate social networks are actually themselves collecting a range of data about their users (FB, Google, Twitter, MySpace etc.), both to personalize the services for the users, but more relevant in terms of privacy issues to sell this data to advertisers.

The availability of personal information online is also an opportunity for identity thieves, scam artists, debt collectors and, stalkers to use the information that people themselves have voluntarily provided in a ways harmful for the owner of the information.

I. Sentimental Community Network

Sentiment in social network is a new research issue in data mining, where sentimental factor play a great role for information diffusion in social network, sentiment analysis attempts to determine how individuals are feeling. Understanding the opinions of large groups of people is invaluable in many disciplines. Businesses need to gauge interest in their products and politicians determine their campaign platforms based on popular opinion. Tradition-ally, polling has been the standard method for gathering this information, but this is costly in terms of time, mo- ney, and manpower, and is difficult to distribute to large groups of people, Using social media platforms such as Facebook and Twitter, individuals can publish their ideas and distribute them very quickly throughout the network. By gathering information published in this way, researchers can acquire more accurate and widely representative data without concerns of survey bias. This computational method of gathering opinion from existing data is called sentimental analysis.

J. Lack of Knowledge about Information Propagation in Social Network

Some user in social network they keep updating and uploading their private information in social media, but they don't know that their private information may be propagate in some wrong community which will be so harmful to them. There are number of reason for information diffusion these are user don't have that much knowledge about social network and maintain privacy own information in social media.

III. RESEARCH ISSUE IN EXISTING MODEL

1. Adrien Guille (June 23,2013),New York,USA have proposed another model SONDY (Social Network Dynamics) an open source platform that helps understanding social network, user's interests and activity, but the model does not consider behavior and mental status of user's. The main weakness of these models are its only global variable not in local variable like human behaviors and mental and social conditions.
2. Ceren Budak ,Divyakant Agrawal,Amr EI Abbadi(2012),Santa Bharbara,USA have proposed a model GLCM(Gaussian Logit curve model) on diffusion of information in social network with user behavior with respect to the entire population and captures the innovativeness of a given user based on its actions. But in this model they didn't consider user's age factor which is a biggest weakness of this model.
3. Adrin Guille, Hakim hacid, Cecile Favre, Djamel A.Zighed (June 2013) have conducted a survey on information diffusion in online social network and they represent the issue on information diffusion in social network. They focus on the particular case of information diffusion in online social network which

are (i) which pieces of information or topics are popular and diffuse the most, (ii) how, why and through which paths information is diffusing, and will be diffused in the future, (iii) which members of the network play important roles in the spreading process? But this model didn't consider about who is diffusing information most and at what age.

4. Adrien Guille (June 23,2013),New York,USA have proposed a model T-BaSIC (Time- Based Asynchronous Independent cascades) a graph based model for detection to information diffusion .but this model depend on time and which parameter are function of time not on people thought or attitude .

IV. CONCLUSION

Information propagation in social network may be cause good impact or bad impact in network as well as on user. Tracking of all information diffusion is difficult task, but if we can think on above discuss point and weakness of already existing model, information diffusion or information propagation can be handle. The researcher can take consideration of all factor of user like age, social and mental status for information propagation in social network.by taking consideration of discussed ten points definitely we can tack information propagation.

REFERENCES

[1]. P. N. Howard and A. Duffy. Opening closed regimes, what was the role of social media during the arabspring? Project on Information Technology and Political Islam, pages 1{30, 2011.

[2]. A. Hughes and L. Palen. Twitter adoption and use in mass convergence and emergency events. International Journal of Emergency Management, 6(3):248{260, 2009.

[3]. M. Cataldi, L. Di Caro, and C. Schifanella. Emerging topic detection on twitter based on temporal and social terms evaluation. In MDMKDD '10, pages 5{14, 2010.

[4]. E. David and K. Jon. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. 2010.

[5]. W. Galuba, K. Aberer, D. Chakraborty,Z. Despotovic, and W. Kellerer. Out tweeting the twitterers - predicting information cascades in microblogs. In WOSN '10, pages 3{11, 2010.

[6]. J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters, 2001.

[7]. M. Gomez-Rodriguez, D. Balduzzi, and B. Sch□olkopf.Uncovering the temporal dynamics of diffusion networks. In ICML '11, pages 561{568, 2011.

[8]. A. Guille, C. Favre, H. Hacid, and D. Zighed.Sondy: An open source platform for social dynamics mining and analysis. In SIGMOD '13, (demonstration) 2013.

[9]. A. Guille and H. Hacid. A predictive model for the temporal dynamics of information diffusion in online social networks. In WWW '12 Companion, pages 1145–1152, 2012.

[10]. M. A. Hasan and M. J. Zaki. A survey of link prediction in social networks. In Social Network Data Analytics, pages 243–275.Springer, 2011.

[11]. H. W. Hethcote. The mathematics of infectious diseases. SIAM REVIEW, 42(4):599–653, 2000.

[12]. J. Ugander et al. The anatomy of the facebook social graph.In CoRR, 2011.

[13]. S. Wu et al. Who says what to whom on twitter? In WWW,pages 705–714, 2011.

[14]. Z. Yang et al. Understanding retweeting behaviors in social networks. In CIKM, pages 1633–1636, 2010.

[15]. Z. Zhou et al. Information resonance on twitter: watching iran. In I Social Media Analytics, SOMA '10, pages 123–131, 2010.

[16]. Josh Ong (August 2012), article retrieved from http://thenextweb.com/in/2012/08/20/social-networkingites-now-occupy-25-online-time-india/

[17]. Papadimitriou, P. Symeonidis, and Y. Manolopoulos, "Fast and accurate link prediction in social networking systems", The Journal of Systems and Software 85, 2012, pp. 2119–2132

[18]. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. S. Glance, "Cost-effective outbreak detection in networks," in Proc. of the 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'07), 2007, pp. 420–429

[19]. Kempe, J. M. Kleinberg, and ´E. Tardos, "Maximizing the spread of influence through a social network," in Proc. Of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'03), 2003, pp.137-146

[20]. http://www.pewinternet.org/data-trend/social-media/social-media-use-by-age-group/25/07/2015

[21]. http://www.telegraph.co.uk/technology/news/10619007/Children-using-social-networks-underage-exposes-them-to-danger.html