International Conference on "Recent Trends in Technology and its Impact on Economy of India"
Guru Nanak College for Girls, Sri Mukstar Sahib, Punjab (India) (ICRTTIEI-17)
24th October 2017, www.conferenceworld.in ISBN: 978-93-86171-74-0

Social Media Mining: Major Concerns in Social Media Mining

Amandeep kaur[1], Jaspreet kaur[2]

Guru Nanak College for girls, Sri muktsar sahib

ABSTRACT

***S**ocial Media has gained remarkable attention. This is due to the affordability of accessing social network sites such as Twitter, Google+, Facebook and other social network sites. Organizations and people relying on the Social Media for information and opinion of other users on diverse subject matters Online Social Network (OSN) mining has been a active area of research in the current years. The huge amount of Data is available in Online Social Networks (OSNs) poses a great challenge for researchers to analyze such networks. The data generated from OSNs is dynamic and need the intelligent mining of dynamic OSN data. . In this paper we take into consideration important current topics related to OSN mining*

***Keywords*: Online Social Networks, Data Mining, Influence Propagation, Community Detection, Sentiment Analysis , Recommendation algorithms.**

I. INTRODUCTION

We live in the an age of internet. Where millions of people spending hours on social media to communicate, connect and interact. Social media has become great source of big data. With the rise of social media, the web has become vibrant and lively where Information is collected, shared or consumed by thousands of individuals, who give spontaneous responses. Almost all the organizations and even government of countries follow the activities on social network. The network enables big organizations, celebrities, government official and government bodies to obtain knowledge on how their audience reacts to postings that concerns them out of the enormous data generated on social network .This increasing tendency of people to use Online Social Networks (OSNs) such as Facebook, Twitter have resulted in making different kinds of interactions which have lead to the generation and availability of a huge amount of valuable data that has never been available before. Figure below show different sources of information on social media.

ISBN: 978-93-86171-74-0

Fig.1 Common Social Media

Huge amount of data is collected from all these resources and this data can be used in some new, varied, eye-catching, and useful way. There are three dominant disputes with social network data : size, noise and dynamism. To resolve these disputes OSNs require appropriate data mining techniques. These data mining techniques are called social media mining. Social Media Mining is the process of representing, analyzing, and extracting actionable patterns from social media data.. The study and development of social media mining, an emerging discipline under the umbrella of data mining. Research in Social media mining is being carried out in various issues such as, influence propagation, expert finding, recommender systems, link prediction, community detection, opinion mining, mood analysis, prediction of trust and distrust among individuals, etc which is depicted in Fig. 2

ISBN: 978-93-86171-74-0

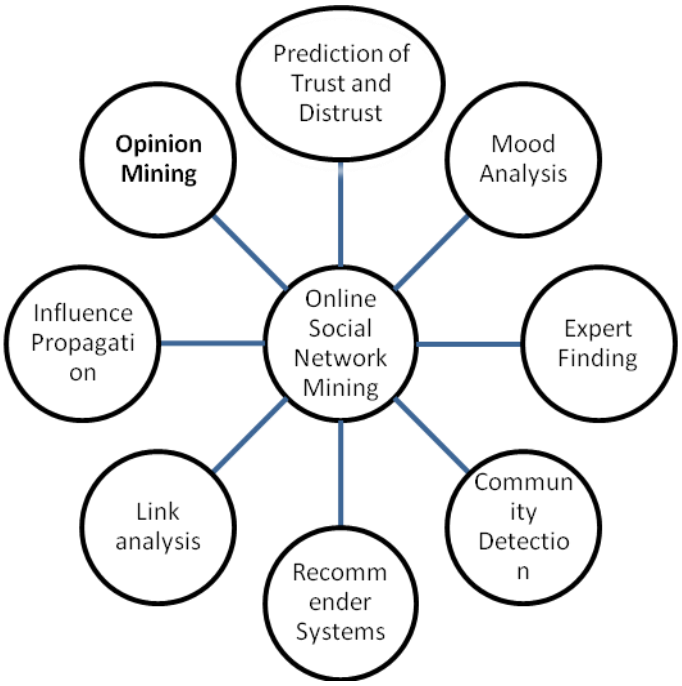


Figure 2:-various research issues in social media mining.

1.1 Influence propagation

In today's age of internet OSNs are attracting millions of people. Thus, influence propagation has become most important for effective viral marketing. Decisions taken by OSNs mainly rely on the influence of social sites . For example, influence propagation can help make a decision on which product to purchase, which audio/video to watch, which community to join, and so on. Online companies can convince customers to purchase products through the help of those active people in OSNs who can play a key role in influencing others. Hence study of influence propagation in OSNs is currently a hot topic related to OSN mining functionalities for researchers.

1.2 Community or Group Detection

Community or group detection in OSNs very important aspect. It is based on studying the OSN structure to club its users into groups by finding which individuals correlate with each other. Such detection can help to make an assessment about what products, services and activities an individual might be interested in. Hence, detection of such OSN community is of prime importance which attracts the researchers in data mining to make an in depth study of community detection in OSNs. Communities can be best explained with the help of graphs. In fact, one of the most applicable features of graphs for representing real systems is community structure. An example of such community graph is shown in figure 3.

ISBN: 978-93-86171-74-0

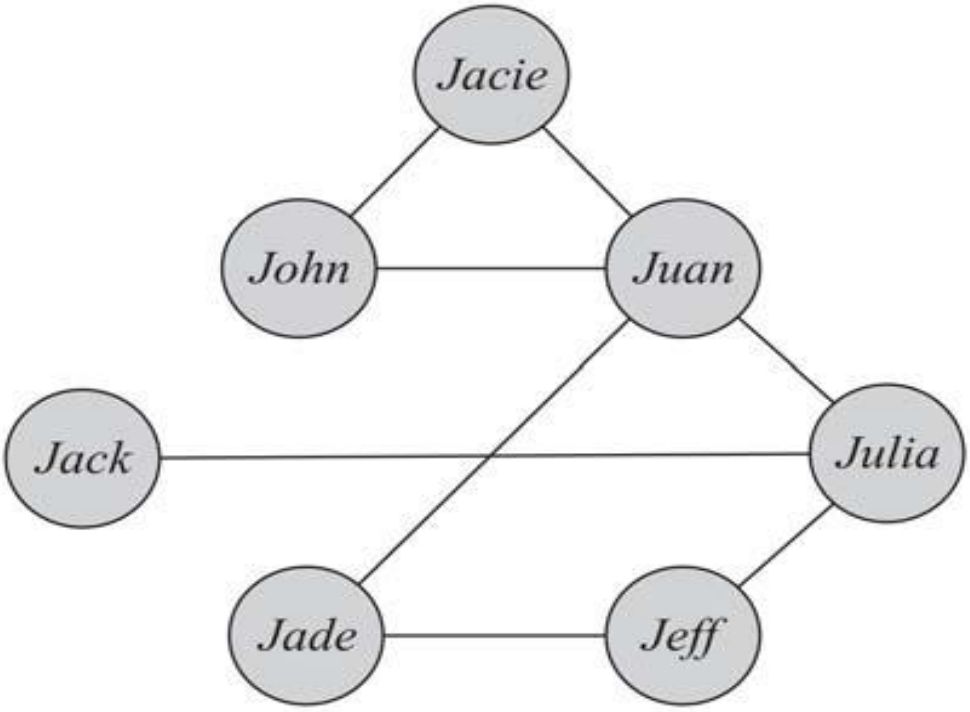


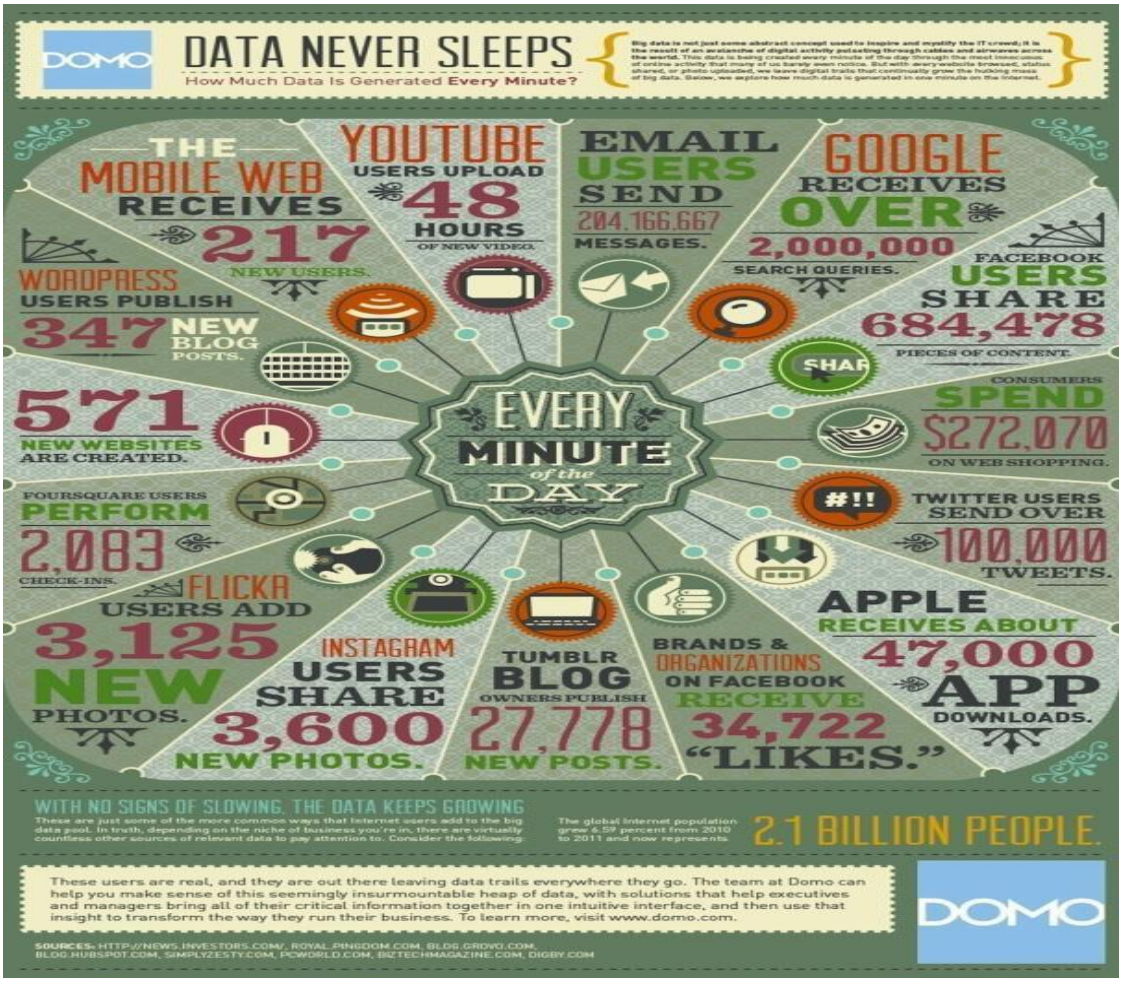
Figure 3:- A Sample Graph. In this graph, individuals are represented with nodes (circles), and individuals who know each other are connected with edges (lines).

1.3 Behavior analysis of individuals

What inspires individuals to join an online group? Individuals exhibit different behaviors in social media: as individuals or as part of a broader collective behavior. Collective behavior is that when a population of individuals behave in a similar way with or without any planning. Collective behavior of individuals on OSN major topic for analysis.

II. LITERATURE REVIEW

According to Technorati, about 75,000 new blogs and 1.2 million new posts giving veiws on products and services are generated every day. Large amount of data is generated every minute on different common *social* sites as shown in Figure below. The data collected on these sites make it difficult for traditional methods l to handle data.Iit is necessary to employ tools capable of analysing *Data* especially the expression of opinions/sentiments which are main characteristics of social media data.. Data mining techniques has shown to be capable of mining big data generated on *Social* sites. Massive data that is collected by social sites is shown figure.



The researches on OSN analysis reveals that intelligent data mining techniques can help effectively deal with the several challenges related to OSN data. First, OSN data sets are very large. For example, if we consider the millions Facebook users, analyzing such a network requires complex graph mining techniques. Second, OSN data sets can be noisy. For example, the unwanted or insignificant tweets in Twitter that need not be taken into consideration during OSN mining. Third, data from OSNs is dynamic, i.e., changes and updates occur frequently in a flicker of a second which poses a great challenge for researchers in dealing with such OSN data. Considering these challenges being faced in research related to OSN mining.

III. MAJOR CONCERNS IN OSN MINING

3.1Community analysis

While mining social media , analyzing communities is very important for many reasons. First, individuals often form groups based on their interests, and when studying individuals, we are interested in identifying these groups. A virtual community over social media comes into existence when like-minded users form a link and start interacting with each other. Communities on social media can be classified into types: **Explicit**: Where interactions done with respect to that interest. e.g., Yahoo! Groups **Implicit**: individuals who write blogs on the same or similar topics. In contrast to explicit communities, implicit communities and their members are unknown to many people. Community detection finds these implicit communities .There are a variety of

ISBN: 978-93-86171-74-0

community detection algorithms. While detecting communities, we are interested in detecting communities with either (1) specific members or (2) specific forms of communities. We denote the former as **member-based community** detection and the latter as **group-based community detection.** Figure below shows the various detection algorithms.

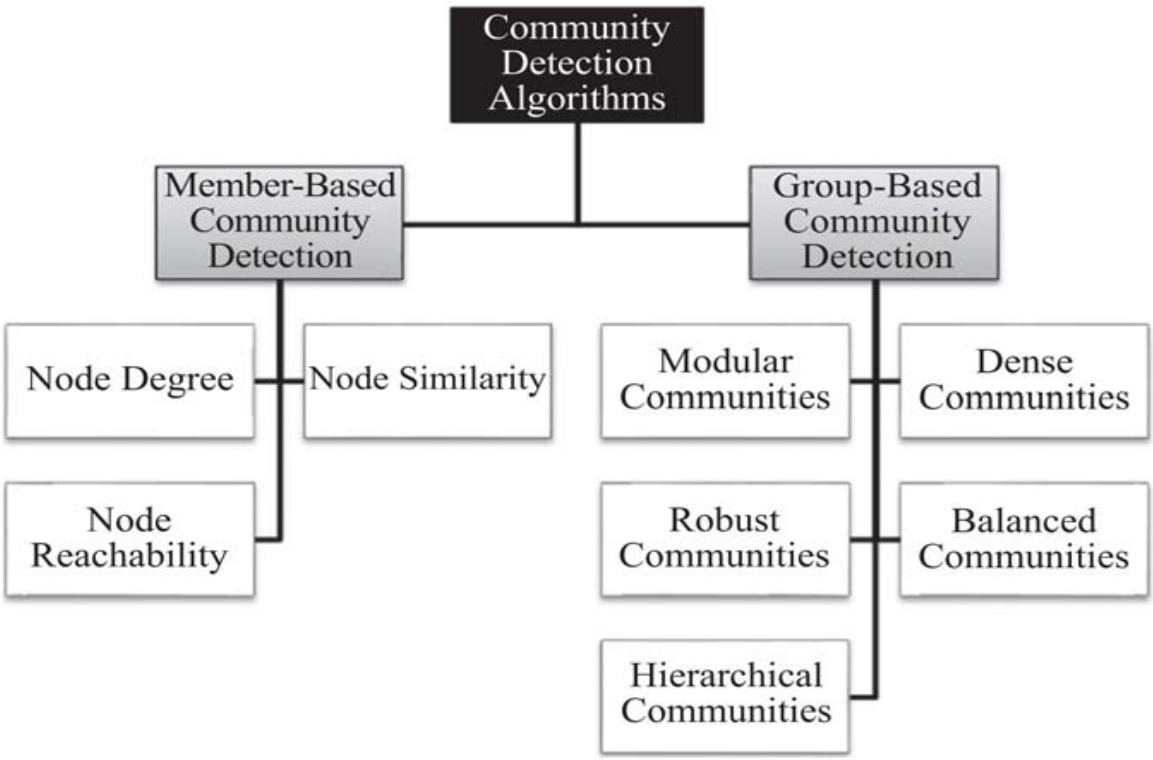


Figure 4: Community detection algorithms

3.2 Sentiment Analysis in social media:

The idea behind sentiment analysis is basically to recognize potential drift in the society as it concerns the attitudes, observations, sentiment and the expectations of stakeholder or the people and to make necessary decisions. It is more important to translate sentiment expressed to useful knowledge by way of mining and analysis. Widespread products are likely to attract thousands of reviews and this may make it difficult for prospective buyers to track usable reviews that may assist in making decision. So proper sentiment analysis in social data is important.

3.3 Recommendations Algorithms in social media

Users in social media make many decisions on a daily basis. These decisions are about buying a product, adding a friend in profile, and renting a movie, among others. The users often faces many options to choose from. Applications and algorithms are developed to help individuals decide easily, rapidly, and more accurately which option to choose and also set of choices that user can make. These algorithms are tailored to users' tastes such that customized recommendations are available for them. These algorithms are called recommendation

ISBN: 978-93-86171-74-0

algorithms or systems. Recommender systems are commonly used for product recommendation. Their goal is to recommend products that would be interesting to users.

3.3.1Challenges for Recommendation algorithms

Recommendation systems face many challenges, some of which are presented next:

Cold-Start Problem. Many recommendation systems use historical data or information provided by the user to recommend items. However, when users first join sites, they have not yet bought any product: they have no history. This makes it hard to decide what they are going to like when they start on a site. The problem is referred to as the cold-start problem. To address this issue, sites often ask users to rate a couple of items before they begin recommend others to them. some sites ask users to fill in profile
information, such as interests. This information serves as an input to the recommendation algorithm.

Data Sparsity. Unlike the cold-start problem, data sparsity relates to the system as a whole and is not specific to an user. In general, data sparsity occurs when a few individuals rate many items while many other individuals rate only a few items. When the information is not reasonably available, then it is said that a data sparsity problem exists. The problem is more prominent in sites that are not popular.

Attacks. The recommender system may be attacked to recommend items that otherwise not recommended. For example, a system recommends items based on similarity between ratings (e.g. item A is recommended for some product because they both have rating . Now, an attacker that has knowledge of the recommendation algorithm can create many fake user accounts and rate similar Item C (which is not as good as item A) highly such that it can get good rating This way the recommendation system will recommend C with product as well as A. This attack is called a push attack, because it pushes the ratings up.

Privacy. The more information a system has about the users, the better the recommendations it provides. but users often avoid revealing information about themselves due to privacy issue. Recommender systems has to face this challenge .

Explanation. Recommendation systems often recommend items without having an explanation why they did so. The system does not know why these items are bought together. Users prefer some reasons for buying items; therefore, recommendation algorithms should provide explanation when possible.

IV.CONCLUSION

Analysing data on social sites especially opinions/sentiments expressed users with data mining techniques has proved effective and useful . This is so because of the capacity data mining possess in handling noisy, large and dynamic data. OSN mining give insights into the social nature of users of OSNs As the number of social media users continues to grow, we will likely continue to see significant changes in the way we communicate and share information with each other. In this regard, the research community needs to continue to look to data mining approaches to provide users with the empowering ability to look deeper into these large data sets generated from OSNs in a more meaningful way. Also, researchers still need to continuously focus on several critical issues of OSN mining for attaining a better solution for the same.

ISBN: 978-93-86171-74-0

REFERENCES

[1] G. Nandi and A. Das, "A Survey on Using Data Mining Techniques for Online Social Network Analysis", in the International Journal of Computer Science Issues, November 2013, vol. 10, issue 6, pp. 162–167.

[2] SJ. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Briesen and N. S. Glance, "Cost-effective outbreak detection in networks", in Proceedings of the 13th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 420–429, 2007.

[3] J. Chen, W. Geyer, C. Dugan, M. Muller and I. Guy, "Make new friends, but keep the old: recommending people on social networking sites", in Proceedings of the 27th International Conference on Human factors in Computing Systems, pp. 201–210, 2009.

[4] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks", in Proceedings of the 12th International Conference on Information and Knowledge Management, pp. 556–559, 2003.

[5] J. Pan, H. Yang, C. Faloutsos and P. Duygulu, "Automatic multimedia cross-modal correlation discovery", in Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 653–658, 2004.

[6] G. Jeh and J. Widom, "SimRank: a measure of structural-context similarity", in Proceedings of the 8thACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 538–543, 2002.

[7] G. Palla, I. Derényi, I. Farkas and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society", Nature, http://dx.doi.org/10.1038/nature 03607, vol. 435, no. 7043, pp. 814-818, 2005.

[8] P. Domingos and M. Richardson, "Mining the network value of customers", in Proceedings of the 7th ACM SIGKDD International Conference on Knowledge discovery and data mining, pp. 57–66, 2001.

[9] B. Yang, W. Cheung and J. Liu, "Community mining from signed social networks", In IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 10, pp. 1333–1348, 2007.

[10] J. Kleinberg, "Authoritative sources in a hyperlinked environment", In Journal of the ACM, vol. 46, no. 5, pp. 604–632, 1999.

[11] L. Katz, "A new status index derived from sociometric analysis", In the Journal of Psychometrika, vol. 18, no. 1, pp. 39–43, 1953.

[12] G. Flake, S. Lawrence, C. Giles and F. Coetzee, "Self-organization of the web and identification of communities", In IEEE Computer, vol. 35, no. 3, pp. 66–71, 2002.

[13] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 61-70, 2002.

[14] D. Blei, A. Ng and M. Jordan, "Latent dirichlet allocation", in the Journal of Machine Learning Research, pp. 993–1022, 2003.

[15] A. Goyal, F. Bonchi and L. V. S. Lakshmanan, "A data-based approach to social influence maximization," in Proceedings of the VLDB Endowment, vol. 5, no. 1, pp. 73-84, 2011.

[16] K. Saito, R. Nakano and M. Kimura, "Prediction of information diffusion probabilities for independent cascade model," in Proceedings of the 12th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES), pp. 67-75, 2008.

[17] W. Chen et al, "Scalable influence maximization in social networks under the linear threshold model", in Proceedings of the 2010 IEEE International Conference on Data mining (ICDM), pp. 88–97, 2010.

[18] D. Kempe, J. M. Kleinberg and ´E. Tardos, "Maximizing the spread of influence through a social network", in Proceedings of the 9thACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD‟ 03), pp. 137-146, 2003.

[19] W. Chen, C. Wang and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data mining (KDD), pp. 1029-1038, 2010.

[20] M. Sachan, D. Contractor, T. A. Faruquie and L. V. Subramaniam, "Using Content and Interactions for Discovering Communities in Social Networks", in Proceedings of the 21st International Conference on World Wide Web, pp. 331-341, 2012.

[21] L. Adamic and E. Adar, "How to search a social network", In the Journal of Social Networks, vol. 27, no. 3, pp. 187–203, 2005.

[22] A. Papadimitriou, P. Symeonidis and Y. Manolopoulos, "Fast and accurate link prediction in social networking systems", In the Journal of Systems and Software, vol. 85, pp. 2119-2132, 2012.

[23] L. Liu, J. Tang, J. Han, M. Jiang and S. Yang, "Mining topic-level influence in heterogeneous networks," in Proceedings of the 19thACM International Conference on Information and knowledge management (CIKM), pp. 199-208, 2010.

[24] Y. Liu, A. Niculescu-Mizil and W. Gryc, "Topic-link lda: joint models of topic and author community", in Proceedings of the 26thAnnual International Conference on Machine Learning, pp. 665-672, 2009.

[25] N. Barbieri, F. Bonchi and G.Manco, "Topic-aware Social Influence Propagation Models", in Proceedings of the 2012 IEEE 12thInternational Conference on Data Mining (ICDM), pp. 81-90, 2012.

[26] T. Hofmann, "Probabilistic latent semantic indexing", in Proceedings of the22nd Annual International Conference on Research and Development in Information Retrieval (ACM SIGIR), pp. 50–57, 1999.

[27] M. Steyvers, P. Smyth, M. Rosen-Zvi and T. Griffiths, "Probabilistic author-topic models for information discovery", in Proceedings of the 10thACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 22–25, 2004.

[28] N. Pathak, C. DeLong, A. Banerjee, and K. Erickson, "Social topics models for community extraction", in Proceedings of the 2nd International Workshop on Advances in Social Network Mining and Analysis (SNAKDD), pp, 77-96, 2008.

[29] Jie Tang , Jimeng Sun , Chi Wang , and Zi Yang, "Social influence analysis in large-scale networks", in Proceedings of the 15th ACM SIGKDD internsational conference on Knowledge discovery and data mining, 2009.

[30] Jianshu Weng, Ee-Peng Lim, Jing Jiang, Qi He, "TwitterRank: finding topic-sensitive influential twitterers", in Proceedings of the 3rd ACM international conference on Web search and data mining, 2010.

[31] Cindy Xide Lin , Qiaozhu Mei , Jiawei Han , Yunliang Jiang , and Marina Danilevsky, "The Joint Inference of Topic Diffusion and Evolution in Social Communities", in Proceedings of the 2011 IEEE 11th International Conference on Data Mining", pp.378-387, 2011.