

Investigating into the Emerging Research Areas of Social Network Analysis

Deepti Gupta

Department of Information Technology, Institute of Innovation in Technology and Management
Guru Gobind Singh Indraprastha University, New Delhi, India

Available online at: www.ijcseonline.org

Accepted: 10/Jun/2018, Published: 30/Jun/2018

Abstract— Social Network Analysis is one of the extensively analysed and researched areas in the current era. Due to the advancements in technology and hence the incremented count of Internet users, this field has gained remarkable attention in the past few years. The massive, noisy and unstructured nature of online datasets that are generating every minute from social networking profiles, blogs, comments, reviews in form of thumbs up and thumbs down, star ratings has posed great challenges for researchers and analysts in order to mine quality facts. The outcomes from the mining of this data expedite and improve the decision forming process. This paper outlines the different issues and the subfields that can be considered as part of Social Network Analysis and need to be studied .It also presents the related algorithms and techniques defined earlier.

Keywords— Data Mining, Influence Propagation, Recommender System, Opinion Mining, Link Analysis, Community Detection, Sentiment Analysis

I. INTRODUCTION

The communication, interactions or relationships among individuals through profiles, comments or other User Generated content has been made possible with the improvements in technology from Web 1.0 to 2.0. These web based services connect users from all over the world and this branch of research was formulated as Social Networking. The vast, distributed, noisy, unstructured and dynamic data generated from social network needs appropriate mining techniques which could result into extraction of effective and good quality information in the form of patterns and can be analysed to facilitate strategic decision making by the organizations.

As social network analysis involves study of interactions, it is nothing but a portion of learning analytics and has been included in various categories of learning like e-learning and m-learning. The opinions, sentiments, assessments, blogs, discussions are the different forms of data on social networks. They provide platforms for sharing of relationships and exchange of information among users from all over the world irrespective of their locations. The results of analysis of social networks facilitate individuals like celebrities, community groups and government to know views of people about them. Social networks are the information dissemination sources which are commonly known for profile creation and posting photos like on Facebook, Instagram, comments and tweets on Twitter, blogs and discussions on Edublogs, government policies and various other types of news through various Mobile Apps and websites. These are monitored and analysed using data mining or in-fact text mining techniques due to their unstructured nature, semantic and syntactic

ambiguities as people are not using proper English sentences. Pre-processing of data is very crucial before actually starting the mining process to clean the noisy data and prepare it for analysis. There might be some part of text that is neutral and does not going to impact the resultant patterns. These words can make the classification of text more difficult and hence needs to be removed.

A social network can be visualized as a graph where the network users can be treated as nodes and relationships or interactions between them as links.



Figure 1: Social Network as Graph [10]

Different applications or research areas of social network analysis includes Influence propagation, Recommender System, Opinion Mining ,Link Analysis, Community Detection, Mood Analysis. These techniques discover the relationships between nodes in the network according to a set of centrality measures.

The various sections of paper are set up as follows. Section 2 illustrates the literature review by the various authors in this area. The research areas that can be held down under Social Network Analysis (SNA) are explained in Section 3 of the paper. Section 4 finally concludes the paper.

II. RELATED WORK

Researchers have undergone a survey of various data mining techniques that can be employed for analysis of social networks. They have summarized all the past work that has been consummated regarding the various issues and areas of social network study like the linkage based structural analysis. They have listed out the various unsupervised and semi-supervised approaches and the corresponding tools required for mining the networks. They coined their proposed methodology as TRCM that is Transaction based Rule Change Mining that discovers the patterns related to original life events and news. Time frame windows were used by them as part of TRCM to compute the lifespan of hash tags on Twitter and to describe how the rules evolved over time by setting a Rule Matching threshold to match rules in different time intervals[1].

The authors have discussed the ongoing trends and research issues in online social network mining. They basically emphasized on the three sub-areas of social network mining that is influence propagation, community detection and link prediction. They have discussed the role and use of some models and algorithms related to these subfields that have been discovered and worked upon including Independent Cascade and Linear threshold Model, Cost Effective Lazy forward optimization for Influence propagation, Maximum Flow Community algorithm, HITS algorithm, CART model for community detection and SimRank for Link prediction[2].

Researchers have suggested the approaches and use of social network analysis to guide social learning. They discussed how results from analysis can be integrated with current learning systems and these tools can be used to encourage and facilitate learning process for both teachers and students. On one hand, they described how social networks can be related conceptually using some measures like degree, closeness, centrality to explain interactions between individuals, brokers and central connectors to analyse communities. On the other hands, they used graph theory to visualise networks mathematically. Last but not the least; they concluded how social learning environments can be optimized by using the potential of social network analysis in education evaluation[3].

The authors have performed an inspection of text mining algorithms that are discovered and employed for mining Facebook posts, photos, comments and Twitter tweets. They summarized the literature review done by the various authors regarding Facebook and Twitter analytics. They explained

that the unstructured and irregular type of social network data is forcing the mining towards text mining from data mining which is more based on structured datasets. They discussed the case study of drug addiction abuse where a classifier was designed to filter out the tweets or posts signalling the use of medication abuse. A case study which identifies the issues of engineering students that they are facing in academics was also analysed through their Twitter posts. Finally, they concluded that Text Mining in Arabic text needs further research especially in case of Sentiment Analysis [4].

III. RESEARCH AREAS OF SNA

The main objective behind Social Network Analysis (SNA) is exchange of information among users from all around the world regardless of their location. The network helps the organisations, government and celebrities to know the reactions of the targeted audience regarding their posts related to political debates or products launched or photos through comments and reviews given. This data on network is analysed to perform further research supporting the various sub-areas of SNA like Influence propagation, Opinion Mining, Sentiment Analysis and many more. All the research sub-areas that are covered under SNA are listed in the figure below:

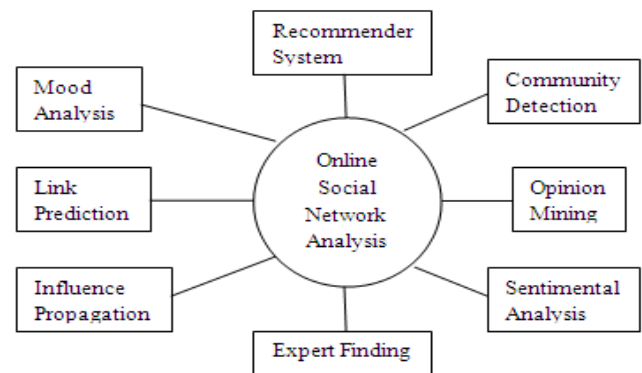


Figure 2: Classification of Research Areas of SNA

A. Influence Propagation

It is one of the key research areas under online social network analysis. As the growth in technology has attracted millions of people for the use of social networks, it has provided a great opportunity as well as challenge for computer scientists, researchers and analysts to study the phenomena of spread of influence and identifying the influential nodes or users in a network. With the explosion of blogging platforms like Twitter, the analysis of influence propagation on social networks has gained popularity. It has found its applications in personalized recommendations especially used by companies for viral marketing where they convince their customers to buy a particular product through the help of

influencers in a network [5]. Also known as the Word of Mouth, the main challenge in this case is to identify the influencers or the seed users that can maximize the influence. The influence maximization problem can be solved using one of the two basic diffusion models: Independent Cascade Model and Linear Threshold Model.

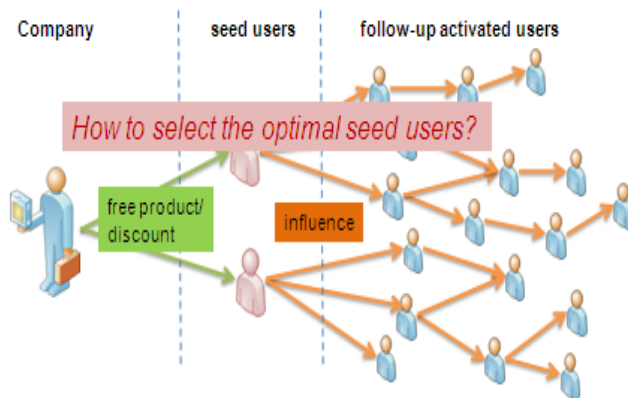


Figure 3: Word of Mouth Marketing [11]

Influence propagation is somehow similar to the way television and newspaper advertisements are used where celebrities endorse a product to influence and convince their fans.

B. Community Detection

The social networks can be considered as a graph where users are nodes and relationships between them as edges of graph. A community is a small cluster of nodes or users in social network that share some common characteristics. The communities might be formed on the basis of education, religion or location. It is very crucial to apply certain tools to detect the communities and understand its behaviour which can be used to model the dynamism of domain to which they belong. [1] Hierarchical Clustering is one of the algorithms that are used to find a community structure based on certain similarity measures like Hamming Distance, cosine similarity which group the nodes on the basis of some similarity between node pairs. Girvan Newman algorithm is another method that was designed for the same purpose. It uses Betweenness Centrality measure of Graph theory to detect community which is based on shortest path. A node having the highest value for this measure is the one through which major information will move and thus would have the maximum control over the network.

C. Recommender System

It is an important research area which is based on mutuality between nodes in a social network graph. Users that belong to

same community recommend products to other users of same community using their past experiences. It is just like the mutual friend feature of Facebook where a person who is a mutual friend of two persons can recommend one to other for Friend request. Recommender Systems can be implemented by multiple methods. One of the techniques is Collaborative Filtering which discovers associations among users on the basis of their similar preferences and tastes. Another method is content based which analyses the structure of data for making recommendations rather than on the basis of other user's opinions. Thus they have the advantage over CF on data scarcity. The author in [5] has used a hybrid approach where he merged the content based method with the Collaborative Filtering and gave an approach, Entrée defined to recommend restaurants initially.

D. Link Prediction

It is the prediction of probable future associations or link between two nodes on the basis of Similarity Index by mining the large amount of data available on social networks. It is found on Friend of a Friend (FOAF) Feature. The authors in [6] have compared the performance of different classification algorithms like the Decision tree, K nearest Neighbour, Neural Networks on two different datasets. One of the dataset is BIOBASE which is a collection of research publications in filed of biology. They have used certain proximity parameters like keyword count, sum of papers, and sum of neighbours for their link prediction. Researcher in [12] had proposed two improved algorithms for link prediction: CNGF that is based on local information in a network and KatzGF which uses global information of a network. Link Prediction has found many applications like in bioinformatics to discover interactions between proteins, in recommender systems like the 'people who bought this also bought this' feature of Amazon and identification of group of criminals or terrorists.

E. Opinion Mining

Millions of posts and comments on products on various shopping sites, thousands of tweets on Twitter generated massive amount of data every minute on social network. It is significant to mine this data as opinions are indicators of user's views and help other users in decision making either it is while purchasing a product online or selecting a right candidate for voting during elections. Analysing the text based opinions and filtering them as positive or negative is difficult problem for the analysts. Homophily Clustering is one of the techniques that are used in Opinion Mining where clusters are formed based on the opinion of influencers and people following it. At the initial stage of opinion mining, there are no clusters which begin to evolve with the introduction of cogent information. By this time, at the alert

stage, the influencer attracts follower ship of other minority users. A percolate stage is reached when minority starts forming an opinion different from other users.

F. Sentiment Analysis

Sentiment Analysis is an evolving research field in Social Network Analysis that uses NLP (Natural Language Processing) for extraction and classification of sentiments as positive or negative from User Generated Data. Due to the large volumes of data on Social network in the form of discussions, reviews, opinions, recommendations, comments and feedback, it becomes essential to mine and analyse sentiments from this data to support decision making tools. The main objective of Sentiment Analysis is to recognize the attitude and expectations of stakeholders that they have expressed through their sentiments. The opinion expressed by users can be overwhelming or just the assertions and the opinion word can vary from being positive to negative according to situation, so it becomes more challenging and important to filter out the false and true sentiments out of it. Sentiments can be analysed at three different levels: document level, sentence level and aspect level. The whole opinion document is classified as positive or negative at document level. At sentence level, each sentence is analysed whereas aspect level directly looks upon the opinions rather than the language constructs. Thus an accurate method is required for predicting sentiments which can be used for research in economics and marketing. There are multiple related fields with Sentiment Analysis that has attracted researchers and various articles like Emotion Detection that aims at extracting and analysing emotions, Transfer Learning and Building Resources. [7] Different approaches that can be used for Sentiment Analysis are Machine Learning, Rule based approach and Lexicon Based approach. Different Machine learning techniques are Support Vector Machine that uses decision plane to define decision boundaries and separates objects using decision plane that belongs to different classes, Naïve Bayes, a classifier based on conditional probability, K Nearest Neighbour and Maximum Entropy Classifier. [8] Lexicon based approach uses a dictionary called Sentiment Lexicon having a list of general words to improve data mining. It can be enhanced by removing the neutral words that depict neutral expressions. The dependence of users on Internet for information to facilitate selection of product from online shopping sites is mainly related to product ratings and reviews. The authors in [9] have proposed a matrix factorization method to increase predictions of rating of products and evaluate how strong is the trust between associations.

IV. CONCLUSION AND FUTURE SCOPE

The accelerated growth of Social network users has fascinated the researchers to divert their attention towards analysis of

social networks. This paper has presented the overview of the various research issues or the sub fields that are interrelated and are covered under Social Network Analysis. The review work performed by different authors in these areas is included in this paper. The relevance of the enormous volume of unregulated and noisy data from analyst's perspective to back decision making has made study of Social networks a very emerging and crucial area of research. Finally to conclude, further techniques and algorithms need to be worked upon to overcome the drawbacks of current discovered approaches.

REFERENCES

- [1] Mariam Adedoyin-Olowe, Mohamed Medhat Gaber, Frederic Stahl, 'A Survey of Data Mining Techniques for Social Network Analysis', School of Computing Science and Digital Media, Robert Gordon University Aberdeen, AB10 7QB, UK, School of Systems Engineering, University of Reading PO Box 225, Whiteknights, Reading, RG6 6AY, UK
- [2] G Nandi, A Das, 'Online Social Network Mining: Current Trends and Research Issues', IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308
- [3] Daniel Amo Filva, Francisco J. Garcia Penalvo, Marc Alier Forment, 'A Social Network Analysis Approaches for Social Learning Support', Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2014)(Salamanca, Spain, October 1-3, 2014) (pp. 269-274). New York, NY, USA: ACM. doi:10.1145/2669711.2669910
- [4] Said A. Salloum, Mostafa, Al-Emran, Azza Abdel Monem, Khaled Shaalan, 'A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives', Advances in Science, Technology and Engineering Systems Journal Vol. 2, No. 1, 127-133 (2017)
- [5] Burke, R, 'Hybrid recommender systems: Survey and experiments', User Modelling and User-Adapted Interaction, 12(4):331-370, 2002.
- [6] Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki, 'Link Prediction using Supervised Learning', Rensselaer Polytechnic Institute, Troy, New York 12180
- [7] Walaa Medhat, Ahmed Hassan, Hoda Korashy, 'Sentimental Analysis algorithms and application: A survey', Ain Shams Engineering Journal, 2090-4479 _ 2014 Production and hosting by Elsevier B.V.
- [8] Devika M D, Sunitha C, Amal Ganesh, 'Sentiment Analysis: A Comparative Study On Different Approaches', Science Direct, Fourth International Conference on Recent Trends in Computer Science & Engineering, Chennai, Tamil Nadu, India, Procedia Computer Science 87 (2016) 44 – 49
- [9] Au Yeung, C. M., and Iwata, 'Strength of social influence in trust networks in product review sites', In Proceedings of the fourth ACM international conference on Web search and data mining (pp.495-504). ACM, 2011.
- [10] Social-Network-Illustration in Concepts & Ideas|People|Technology,Preview-124730,Vexels.com
- [11] Suqi Cheng, Huawei Shen, Junming Huang, Guoqing Zhang, Xueqi Cheng, 'StaticGreedy: Solving the Scalability-Accuracy Dilemma in Influence Maximization', Research Centre of Web Data Sciences & Engineering Institute of Computing Technology, Chinese Academy of Sciences
- [12] Liyan Dong,1,2 Yongli Li,3 Han Yin,1,2 Huang Le,1,2 and Mao Rui, 'The Algorithm of Link Prediction on Social Network', Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2013, Article ID 125123

- [13] C.Amali Pushpam, J.Gnana Jayanthi,' *Overview on Data Mining in Social Media*', © 2017, IJCSE All Rights Reserved
147International Journal of Computer Sciences and Engineering,
Volume-5 Issue-11, E-ISSN: 2347-2693

Authors Profile



Ms. Deepthi Gupta pursued Bachelors in Technology, Information Technology from Guru Gobind Singh Indraprastha University, India in 2012 and Masters in Technology, Computer Science from Guru Gobind Singh Indraprastha University in year 2015. She is currently working as Assistant Professor in

Department of Information Technology, Institute of Innovation in Technology and Management, Guru Gobind Singh Indraprastha University, India since 2016. She has published many research papers in reputed international journals and conferences including IEEE and it are also available online. Her main research work focuses on Big Data Analytics and Data Mining. She has 2 years of teaching experience.
