International Journal of Advanced Computer Technology
ISSN: 2319-7900 www.ijact.org Volume-X, Issue-V, October 2020

Community Detection over Social Media: A Compressive Survey

Shraddha Sharma[1], Yogadhar Pandey[2]
Department of Computer science & Engineering, TIT, Bhopal
[1]sshraddha0903@gmail.com

Abstract:-*Social media mining is an emerging field with a lot of research areas such as, sentiment analysis, link prediction, spammer detection, and community detection. In today's scenario, researchers are working in the area of community detection and sentiment analysis because the main component of social media is user. Users create different types of community in social world. The ideas and discussions in the community may be negative or positive. To detect the communities and their behavior researcher have done a lot of work, but still two major issues are presents per survey, Scalability and Quality of the community. These issues of community detection motivate to work in this area of social media mining. This paper gives a bird eye view over social media and community detection.*

Keywords: - Social Network Analysis, cloud, spammer, communities, Network Influence.

I. INTRODUCTION

All modern information systems can view as digitally connected systems and models of connected groupings with patterns of interactions. For these models of organization to advance beyond the metaphorical phrase, a coherent framework and accompanying methods of analysis that can capture relationships and interactions are essential—Social Network Analysis (SNA) with its established approaches bridges this aspect. SNA is not just a theoretical platform in sociology but rather a strategy for understanding social relationships and entities [1] [2] [3] [4] [5]. Social media is becoming an integral part of daily life nowadays. Social media generate user-based content. Users want to create their profile on social media in which they have to mention their personal information, as shown in figure 1. Social media are facing increasing security threats from local and distributed elements.

Up till now, the researcher has done great work in this field of social media [6-9]. Stopping release of user information and attack in social networks in its initial stage may result in saving time energy and space. Keeping the social media secure and user friendly is of supreme importance to every user. Neighbourhood privacy theft and spammer detection are attacks on social networking sites. Therefore community detection is necessary to know the characteristics of the users. It is a process of detecting communities in the social network. Community detection is essential in social media due to the reason that users create groups based on their interest. There are two types of communities' real-world communities and virtual communities. Users have common social-economic or political Interests termed as Real-world communities and the communities which created one the social sites are called Virtual Communities. Social Networking Services Online is a popular web service which provides users to create profiles and share these profiles within other Peoples can share their daily life and opinions within their friends one Internet within this Service. The popular online social networking sites like MySpace Facebook Google and Bebo have attracted thousands and thousands of users [1], [2]. The connection between the users constitutes a Network for every site, and the network of these Social Networking sites is called "Social Networks Online".

Figure 1: Social Media Land Escape [24]

Network science is an interdisciplinary field of study that looks at different physical virtual and real-world systems as networked models. A network is a simple model of the connected structure and has as its core concept nodes that are connected by edges. SNA is a subgenre of network

science and focuses on one entity and relationships. It notated the intention of this papery to provide detailed coverage of Network or SNA research as both Network and SNA have growth within extensive publications and applications. Here, we assume that readers have basic familiarity with SNA. For those who wish to gain a foundational understanding of SNA may want to refer to [1] [2] [3] and many other books, journals and publications.

II. SOCIAL MEDIA MINING

A process of extracting and analyzing applicable patterns over the social network is known as Social Media Mining. Through Social Media Mining, social theories have integrated within the computational methods. Social Media Mining defines the basic principle and concepts for investigating a vast amount of social media data. There are various tools which are used in social media mining to measure and extract meaningful patterns over large Social Networks. Social media sites generate user data which is different from traditional attribute-values of data for Hellenic data mining. The Data which generated from social sites are noisy distributed not in proper structure and frequent. All the characteristics of social media data pose challenges for data mining task, and for that, new techniques and algorithmic have to be developed [11, 12].

III. SOCIAL MEDIA MINING CHALLENGES

It is known that every field of research or technology has some challenges. In the same manner, social media mining has some Size of Social Media Data challenges.

3.1 Size of Social Media Data

Data generated through the social media site is huge and significant in size. If they want to zoom into individual information, then there should be relevant recommendations about the individual. The Data which is available about the individual is littlie for each specific individual.

3.2 Extraction of Data

Social media data extracted through API (application programming interface); we can extract an only limited amount of data on daily purpose. There are specific tools available for extracting social media data, but, they also have a limit for extracting data.

3.3 Removal of Noise

The portion of noise in social media data is large. Therefore removal of noise from Data is a challenging task in mining. We cannot remove noise blindly, because in this case, valuable information can be eliminated. In this scenario definition of noise is dependent on our task at hand.

3.4 Evaluation Process

In data mining, there is a standard procedure of evaluating pattern from data, and this is done based on ground truth. Data set for Ground truth is divided into two parts, training dataset and testing dataset. In social media mining, often ground truth is not available. Therefore, evaluating patterns from social media data is a tough, challenging task.

IV. SOCIAL THEORIES

There are three types of social theories such as; Balance theory Status theory and Social correlation.

4.1 Balance Theory

Social balance theory represents the structural consistency in the friend/foe relationship among individuals. In graphs nodes represents individuals and edges represents the status of the individual with the sign positive and negative. Suppose there are two vertices v1 and v2, with an edge having a positive sign. It means v1 and v2 are friends, and if the sign of edge is negative, then they are not connected.

4.2 Status Theory

Social status theories represent how consistent an individual connection its neighbours. We can summarize this by an example, Suppose A has lower status then B and B has lower status then C. A should have a lower level than C. Positive directed edge from Y to X node show that Y has more excellent status than X in the same manner rest of the nodes and edges with sign represent all the connection.

4.3 Social Correlation

Social correlation is related to individual behaviour in the social network. It is divided into three parts: influence, homophily, confounding.

A. ***Influence:-*** it is an art or act by which an individual affects another user in the network. The influence user is more appropriate to influential figure, as shown in figure 2 that guided influence

propagation over the network. Influence is the process by which an individual (the influential) affects another individual such that the influenced individual becomes more similar to the influential figure. If most of one friend switch to a mobile company, he might be influenced by his friends and switch to the company as well as shown in figure 2.

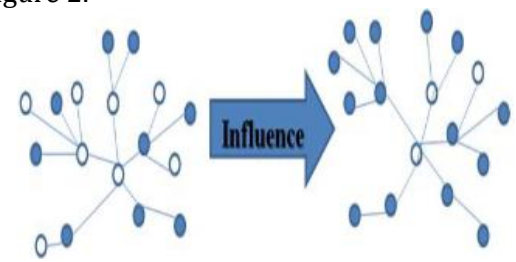

Figure 2: Influence Propagation over the Network.

B. ***Measures of Influence:-*** There are two types of measures for calculating influence between the individuals: prediction-based and observation-based.

I. ***Prediction-based:-*** In this measure, an individual is how much influence is predicted. Prediction made based on the attributes of the individual and where it situated in the network. For instance, it assumed that the number of friends of an individual could tell us about how influential the individual is. If we want to find out the significant user in the twitter network, then the in-degree attribute is used to calculate the influence of the user [3].

II. ***Observation-based:-*** In this type of measure, the influence of the user is quantified based on the influence parameter. An individual can influence differently in different settings. Here we have described three different stages in which individual influence differently. When an individual treated as a role model: This type of setting is done when we talk about fashion industry celebrities and teachers. Size of the audience depends on the influence rate. Several likes can act as an accurate measure of influence. When an individual communicates information: In this Scenario, influence is based on the number of hops traversed by the information. It also depends on the rate at which users get influenced. When the value of an item increases due to user participation: In this type of setting value of a product is increased according to the purchase of the product. Therefore the number of purchase/sells of the product decides the influence of that product [3].

C. ***Homophily:-***Homophily is the tendency of an individual to associate and bond with similar individuals. People interact more often with people who are "like them" than with people who are dissimilar as shown in figure 3[3].

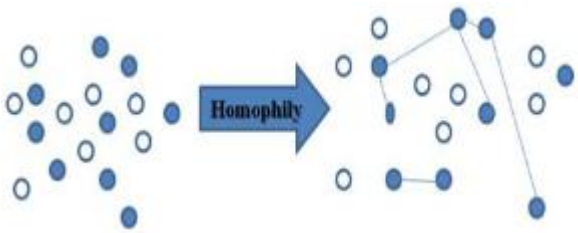

Figure 3: Homophily over the Network

It is realized when similar individuals become friends due to their high similarity. Two musicians are more likely to become friends, as shown in figure 3. Note that both influence and homophily social forces give rise to assortative networks. After either of them affects a network the network exhibits more similar nodes; however, when "friends become similar," we denote that as influence and when "similar individuals become friends," we call it homophily. Figure 3-4 depicts how both influence and homophily affect social networks.

V. RELATED WORK

Community Detection formalizes the strong social groups based on social network properties. Network interaction provides rich information about the relationship between users, but Communities Detection in Multi-mode Multi-Dimensional Networks is a challenging task. In this Section-VII, we have discussed the related work that uses various approaches we have described earlier on the Hierarchical Classification of Community Detection Algorithm.

Qiyao Wang et al. [2], Proposes an independent group-based dense community's cascade-based model for influence maximization termed as IMIC-O which is going to calculate positive influence and also considers about the Influence Maximization

that refers to the process of detecting influential users who make the most of product adoption. Assumption of the author is that influential users spread positive opinions. At the beginning user's initial opinions can be positive or negative. Users balanced their own opinions as well as of their neighbours when more users involved in the discussions. To calculate the maximum positive influence on the three real networks Facebook, HEP- PH and Epinions, various experiments were conducted, and it was based on the IMIC-OC model. The proposed model resulted in a more considerable positive influence and accordingly indicating better overall performance.

Sang Yup Lee et al. [3] examined homophily and social group-based balance communities, Two online casual game players having similar game genre preferences tend to be friends with each other Or over the periodic of time players gaming preferences become more similar to those of her Kongregate friends which is examined by influence processes. For this study, demographic attributes game genre preferences gaming frequencies and relational ties for 2488 game players for two time periods from Kongregate are collected by authors. SIENA (Simulation Investigation for Empirical Network Analysis), is a program which is used for social network analysis of panel data and was analyzed with the R version of SIENA. The results suggest that there might not be strong homophily.

Barbieri, N et al. [4] propose a group-based balance communities stochastic framework for the diffusion model, which is based on community-level influence. The framework assumes that items adoption governed by diffusion process over the unobserved social network. To learn the community membership and the level of influence of each user in each community, an analysis is done by putting the model parameters to the user activity log. This allows identification of "key" users for each community; key users are the leaders who are most likely to influence the rest of the community to adopt a certain item.

Razieh Hosseiniet. al. [5] proposes an improved label propagation algorithmic for finding community structure in social networks is known as memory-based label propagation algorithmic (MLPA). In this algorithmic for every node of the graph, a simple memory element is designed and iteratively the most frequent joint adoption of labels stored in this element. The algorithm MLPA holds several labels of each node, and finally, the most frequent standard label for each node forming communities of the networks are extracted by the algorithm.

Xiaokang Zhou et al. [13] Focus on mining and tracking the dynamic communities based on social networking analysis. Based on Member-based community detection, node similarity and generic framework for the dynamic community discovery, a computational approach is developed to extract user's static and dynamic features for the temporal trend detection. A dynamically socialized user networking model is then presented to describe users' various social relationships.

Zhou, X., Yen et.al.[14] present Member-based Community detection primarily Based on node similarity and a way to analyze and extract meaningful information according to the needs of current users and interests of social currents using two advanced algorithms and go further to integrate this current data organization which is described as Assembly undulations in the search system in order and improve the importance of the results obtained in the search engine and users feedback with a new perspective of the issues necessary to guide the search for more information this can benefit both enrich the search experience and facilitate the search process.

Adam Roughton et al. [15] has suggested a classification of interaction group-based models based on the multitudes and the classification of literature. According to this classification Crowds, the traditional definition should be reviewed in the mild of advances in communication technologies such as smartphones and cloud-based infrastructure. Authors have expanded the definition to include virtual scattered masses by having a look at the basic components of crowd-based activities. Authors suggest simplified interactive control for highly synchronized and cooperative economic activities. They argued that the equal, cooperative element and the identical economy could be obtained with a rich reflexive control.

Zhou X., iJin Q et al. [16] tries to discover the potential and dynamic links of users who use those social flows that have been reorganized according to the current interests and needs of the user to

help in the search for information. The author developed Node reachability Based Member-based community detection mechanism to construct the social networking model (DSSON), described fixed of measures (which includes a degree of interest degree of popularity) and concepts (such as complementary tideland strong tie) that may detect and represent current users and dynamic connections. The corresponding algorithms are advanced respectively. Based on these authors then discussed the scenario of the DSSON application with the result of the experiment.

VI. CONCLUSION

Community detection is the most important feature of social media. It is similar to the clustering feature of data mining. In member-Based community detection, a whole lot of work has been completed, but identifying community through Influence is a different way of detection in social media mining. Most of the community detection approaches do not consider the information about the formation of communities in a social network. Community members have individual social roles: leaders, members, the elite in real-world communities. Treasons to other communities, some are extra influential than others, and some are more susceptible to influence individual roles. The identity of influential users in a social network is a hassle that has received enormous interest in recent studies.

REFERENCES

[1]. Samadi, Mohammadreza, Alexander Nikolaev, and Rakesh Nagi. "A subjective evidence model for influence maximization in social networks." Omega 59 (2016): 263-278.

[2]. Wang, Qiyao, et al. "Influence maximization in social networks under an independent cascade-based model." Physica A: Statistical Mechanics and its Applications 444 (2016): 20-34.

[3]. Lee, Sang Yup. "Homophily and social influence among online casual game players." Telematics and Informatics 32.4 (2015): 656-666.

[4]. Barbieri, Nicola, Francesco Bonchi, and Giuseppe Manco. "Influence-based network-oblivious community detection." 2013 IEEE 13th International Conference on Data Mining. IEEE, 2013.

[5]. Hosseini, Razieh, and Reza Azmi. "Memory-based label propagation algorithm for community detection in social networks." 2015 The International Symposium on Artificial Intelligence and Signal Processing (AISP). IEEE, 2015.

[6]. Yang, Jaewon, Julian McAuley, and Jure Leskovec. "Community detection in networks with node attributes." 2013 IEEE 13th International Conference on Data Mining. IEEE, 2013.

[7]. Fortunato, Santo. "Community detection in graphs." Physics reports 486.3-5 (2010): 75-174.

[8]. Malliaros, Fragkiskos D., and Michalis Vazirgiannis. "Clustering and community detection in directed networks: A survey." Physics reports 533.4 (2013): 95-142.

[9]. Gong, Maoguo, et al. "Community detection in networks by using multiobjective evolutionary algorithm with decomposition." Physica A: Statistical Mechanics and its Applications 391.15 (2012): 4050-4060.

[10]. Shen, Huawei, et al. "Detect overlapping and hierarchical community structure in networks." Physica A: Statistical Mechanics and its Applications 388.8 (2009): 1706-1712.

[11]. Xing, Yan, et al. "A node influence based label propagation algorithm for community detection in networks." The Scientific World Journal 2014 (2014).

[12]. Bonchi, Francesco. "Influence Propagation in Social Networks: A Data Mining Perspective." IEEE Intell. Informatics Bull. 12.1 (2011): 8-16.

[13]. Zhou, Xiaokang, et al. "Dynamic community mining and tracking based on temporal social network analysis." 2016 IEEE International Conference on Computer and Information Technology (CIT). IEEE, 2016.

[14]. Roughton, Adam, et al. "The crowd in the cloud: moving beyond traditional boundaries for large scale experiences in the cloud." Proceedings of the Twelfth Australasian User Interface Conference-Volume 117. 2011.

[15]. Zhou, Xiaokang, and Qun Jin. "User correlation discovery and dynamical profiling based on social streams." International Conference on Active Media Technology. Springer, Berlin, Heidelberg, 2012.

[16]. Newman, Mark EJ, and Michelle Girvan. "Finding and evaluating community structure in networks." Physical review E 69.2 (2004): 026113.

[17]. Shen, Wei, Jiawei Han, and Jianyong Wang. "A probabilistic model for linking named entities in web text with heterogeneous information networks." Proceedings of the 2014 ACM SIGMOD international conference on Management of data. 2014.
[18]. Lancichinetti, Andrea, et al. "Finding statistically significant communities in networks." PloS one 6.4 (2011): e18961.
[19]. Long, Bo, et al. "Unsupervised learning on k-partite graphs." Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. 2006.
[20]. Long, Bo, Zhongfei Mark Zhang, and Philip S. Yu. "A probabilistic framework for relational clustering." Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. 2007.
[21]. Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu, "Social Media Mining", https://dmml.asu.edu/smm 2014
[22]. Newman, Mark EJ. "Fast algorithm for detecting community structure in networks." Physical review E 69.6 (2004): 066133.
[23]. Flake, Gary William, Steve Lawrence, and C. Lee Giles. "Efficient identification of web communities." Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. 2000.
[24]. Girvan, Michelle, and Mark EJ Newman. "Community structure in social and biological networks." Proceedings of the national academy of sciences 99.12 (2002): 7821-7826.
[25]. Newman, Mark EJ. "Modularity and community structure in networks." Proceedings of the national academy of sciences 103.23 (2006): 8577-8582.
[26]. Ferrara, Emilio. "Community structure discovery in facebook." International Journal of Social Network Mining 1.1 (2012): 67-90.
[27]. Roughton, Adam, et al. "The crowd in the cloud: moving beyond traditional boundaries for large scale experiences in the cloud." Proceedings of the Twelfth Australasian User Interface Conference-Volume 117. 2011.
[28]. Zhou, Xiaokang, et al. "Enriching user search experience by mining social streams with heuristic stones and associative ripples." Multimedia tools and applications 63.1 (2013): 129-144.
[29]. Wang, Chang-Dong, Jian-Huang Lai, and S. Yu Philip. "NEIWalk: Community discovery in dynamic content-based networks." IEEE transactions on knowledge and data engineering 26.7 (2013): 1734-1748.