**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p. 1414-1420
https://publishoa.com
ISSN: 1309-3452

# A Survey and Analysis on Online Social Networks

**R.Sudha,**
Assistant professor, Dept.of IT, PSNA College of Engineering and Technology, Dindigul,Tamilnadu
**Dr.N.Uma Maheswari,**
Professor, Dept.of CSE, PSNA College of Engineering and Technology,Dindigul, Tamilnadu
**Dr.R.Venkatesh,**
Professor, Dept.of IT, PSNA College of Engineering and Technology, Dindigul, Tamilnadu
Mail-id: :sudha@psnacet.edu.in, numamahi@psnacet.edu.in, venkatit@psnacet.edu.in

**Received** 2022 March 15; **Revised** 2022 April 20; **Accepted** 2022 May 10.

**Abstract**

Social network analysis is a multidisciplinary research covering informatics, mathematics, sociology, management, psychology, etc. In the last decade, the development of online social media has provided individuals with a fascinating platform of sharing knowledge and interests. With hundreds of millions of users worldwide, online social networks (OSNs) offer incredible opportunities for social connection, learning, political and social change, and individual entertainment and enhancement in multiple contexts. Though OSNs are highly beneficial, it suffers from different kinds of malicious bots (e.g., spammers, fake followers, social bots, and content polluters). Malicious bots have abused the power of OSNs such as Twitter, Face book, and Weibo continuously caused significant disturbance to the overall online social environment and shaped unhealthy trends, bias, and misbelieve in societies. These social networks have many challenges like community detection based on the user behavior, influential nodes etc., In this paper, these two challenges have been taken and compared against various parameters.

## Introduction

The internet was not concerned with data trade alone: it was a modern multidisciplinary tool empowering people to create content, communicate with one another, and indeed elude reality. The rise of the internet has started a talk about approximately how online communication influences social connections. The internet is the tool we utilize to connect with one another, and postures unused challenges to privacy and security.

The term Social Network is utilized to depict web-based administrations that permit people to make a public/semi-public profile inside a space such that they can communicatively interface with other users inside the network. In network hypothesis, a social network is commonly modeled by a graph which comprises of users or groups called nodes associated by patterns of contacts or interactions called edges or links [1]. The one of a kind component of social organized information is that they bring unused openings to understand people and society, given the acknowledgment and believe, people have shown towards them.

Security has risen near the top of the agenda in step with an expanding awareness of the suggestions of utilizing social media. Much of the time, individuals begun to utilize social media with no genuine thought of the perils, and have wised up as it were through trial and error. Unseemly utilize of social media appears to hit the headlines each day. Celebrities posting inappropriate comments to their profiles, private pictures and tapes spilled to the Internet at large, companies showing presumption toward clients, and indeed criminal exercises include private-data abuse or social media abuse.

Social media permits people to assemble and express themselves in a much more basic and prompt mold. Interdisciplinary research on social systems is encountering phenomenal development, fuelled by the combination of the field of social network analysis and the increasing accessibility of information from digitalized platforms. Extricating insights from such information has getting to be a rapidly extending multidisciplinary area that requests the cooperation of scientific tools and ability [2]. Key analysis practices include social network analysis, sentiment analysis, trend analysis and collaborative recommendation. Social network information also has lately incorporated in recommendation systems. The latter are capable of dealing with the problems of information overload and information filtering.

The amazing development of SNS can be considered as a start that burst the Enormous growth of big data era. It makes accessible an uncommon scale of individual information, information about events and social relationships, public sentiments. The greatest distinction between conventional researches and social network analysis is that, focusing on the connections of actors in network analysis while focusing on actors and qualities in conventional researches. In network analysis, people and their qualification are not taken in to consideration [3][4]. We have taken only two challenges community detection and influential nodes to compare the various findings.

**Community detection based on the User behavior**

**Definition of Community**

A community is a group of people who share something in common. You can define a community by the shared attributes of the people in it and/or by the strength of the connections among them. You need a bunch of people who are alike in some way, who feel some sense of belonging or interpersonal connection.

The process of discovering the cohesive groups or clusters in the network is known as community detection. It forms one of the key tasks of social network analysis. The community has five functions: production-distribution-consumption, socialization, social control, social participation, and mutual support. A community is made up of different people with different interests, experiences and backgrounds in a social network. The built and natural environments together make up the physical structure of a community. By changing that structure, you may be able to change community members' attitudes, behaviors, prospects for health and well-being, economic opportunity, social interactions, and quality of life.

**Algorithms**

**Louvain Algorithm:**

The Louvain method for community detection is an algorithm for detecting communities in networks. It maximizes a modularity score for each community, where the modularity quantifies the quality of an assignment of nodes to communities. This means evaluating how much more densely connected the nodes within a community are, compared to how connected they would be in a random network.

The Louvain algorithm is a hierarchical clustering algorithm, that recursively merges communities into a single node and executes the modularity clustering on the condensed graphs. Time complexity of this is O(m).

Different methods have emerged over the years to efficiently uncover communities in complex networks. The most famous principle is maximizing a measure called modularity in the network, which is approximately equivalent to maximizing the number of edges (relationships) inside the communities and minimizing the number of edges between the communities. The first greedy (in terms of computation) algorithm based on modularity was introduced by Newman in 2004.

Another method to detect communities is by simulating random walks inside the network. It is based on the principle that a random walker will tend to stay inside densely connected areas of the graph. That is the idea behind the walktrap algorithm, introduced by Pascal Pons and Matthieu Latapy in 2005. These algorithms are working based on the one community per node, but in real networks, a node belongs to more than one community. So considering the overlapping communities is important for many real networks.

Jaewon Yang and Jure Leskovec also published a paper in 2013 for their own overlapping community detection algorithm called BigClam. The particularity of this algorithm is that it is very scalable (unlike the algorithms mentioned before) thanks to non-negative matrix factorization. The authors claim that contrary to common hypothesis, community overlaps are denser than the communities themselves, and that the more communities two nodes share, the more likely they are to be connected. We have to establish the criteria we want to compare the algorithms on. The most efficient way of evaluating a partition is to compare it to the communities we know exist in our network, they are called "ground-truth" communities. Several measures enable us to make that comparison, notably the normalized mutual information (NMI) and the F-Score.

**Kernighan Lin Technique:**

Partitioning based techniques have used in the past by the researchers to divide the network into groups in such a way that these groups have less connections. It divides the nodes of a network into different subsets of provided sizes to minimize the cost on edge cuts. A limitation of this technique is before execution number of groups needs to define. The technique is good as worst case time complexity is O(n2 ).

There are two famous benchmark networks generators for community detection algorithms: Girvan-Newman (GN) and Lancichinetti-Fortunato-Radicchi (LFR).

**Related Work**

Community detection has many applications like advertising, marketing, recommender systems, and many more medical applications. There are two approaches for community detection in social networks such as stochastic block model and modularity maximization. These models are only used for low dimensional space which is linear. Since the real world networks looks like non-linear structure, these traditional methods become less-practical for real world networks. Deep learning, neural networks, artificial neural networks, Deep neural networks can be used for the non-linear structure of real world networks. In [5], deep learning neural networks have been used based on multiple auto-encoders and it has been proposed to classify and link prediction by training the auto-encoders. The user profiles are also used for community detection since they include the topologies and node contents. The user profiles are very much useful to predict the community so it becomes one of the hot research topics. However, they often suffer from two drawbacks: 1) they cannot extract a potential deep representation of the network; 2) they cannot automatically weight different information sources with adequate balance parameters. A deep integration representation (DIR) algorithm [6] via deep joint reconstruction was proposed by having deep feed-forward auto-encoders and spectral clustering in terms of matrix reconstruction. It also learns the most suitable balance between different components automatically. Graph clustering algorithms are mainly used to identify the communities in real or artificial networks. They use the structural characteristics of the network. It is very difficult to compute the efficiency of such algorithms in real world social networks and its analysis. In [7], the performance of eight state-of-the-art graph clustering algorithms is demonstrated on small egocentric graphs, obtained from Facebook. In [8], a novel community detection method overrules the traditional k-means approach in terms of precision and stability while adding extra computational costs. Result of extensive experiments undertaken on computer-generated networks and real-world datasets illustrate acceptable performances of the introduced algorithm in comparison with other typical community detection algorithms.

The original adjacency matrix in social network is reconstructed based on the opinion leader and nearer neighbors for obtaining spatial proximity matrix. The spatial eigenvector of reconstructed adjacency matrix can be extracted by an auto-encoder based on convolution neural network for the improvement of modularity. Four open datasets of practical social networks were selected to evaluate the deep community detection method [9] and it obtained higher modularity than other methods. Analyzing the dynamics of social networks using machine learning techniques to locate maximal cliques and to find clusters for the purpose of identifying a target demographic is a daunting task. There are supervised and unsupervised learning in machine learning. Unsupervised machine learning techniques are designed and implemented to analyze a dataset from YouTube to discover communities in the social network and find central nodes. Different clustering algorithms are implemented and applied to the YouTube dataset. The well-known Bron-Kerbosch algorithm is used to find maximal cliques[10]. In recent years, many researchers have concentrated on feature selection and network embedding methods for node clustering. Detecting the community structure whether it is linear or non-linear is important task in any social network analysis. In [11], the researchers proposed a model for learning graph representation using deep neural networks. It concentrated to feed to stack auto-encoders for learning the model. The overlapping clustering algorithm is deployed to extract the overlapping communities.

YouTube has 2 billion users worldwide (Statista, 2019). Statista gives us information about the most popular social networks worldwide as of April 2019 and the number of active users they have on a monthly basis. The only social network that has more monthly active users than YouTube is Facebook. In fact, nearly eight out of ten (78.8 percent) of marketers consider it to be the most effective platform for video marketing (GO Globe, 2019). In comparison, at 58.5 percent, far fewer marketers consider Facebook, the king of social media, to be the most effective video marketing platform. Everyday people watch one billion hours of videos on YouTube and generate billions of views (YouTube, 2019). Let's do the math. If every single person on earth watched a video, that's around 8.4 minutes per day per person. That's a mindblowing number that just adds to the credibility of video as a content source for people. Facebook reached 2.45 billion monthly users, up 1.65%, from 2.41 billion in Q2 2019 when it grew 1.6%, and it now has 1.62 billion daily active users, up 2% from 1.587 billion last quarter when it grew 1.6%. In social network, nodes are called as a vertices and link between any two nodes vertices is called as an edge. The edges can be directed or undirected, weighted or un weighted. The social network can be represented using the graphs.

Figure 1: A social Network Example

In figure 1,the circles represent the vertices and links between the vertices represent edges. If the edge density is inhomogeneous, then the group of nodes with high concentration of edges within them and low concentration of edges between different groups is called as grouping of nodes or clustering or community structures in social network analysis. Community detection and clustering can be applied in Biology- communities are likely to group proteins having the same specific function within the cell and especially in world wide web(WWW)- group of pages dealing with the same or related topics, and so on. Members within a community are more similar among each other. Communities within the networks form the densely connected nodes. The connections among the vertices inside the community might be better and stronger than the rest of the network. The notion of communities is based on the number of edges within a group (density) compared to the number of edges between different groups. A community corresponds to a group of nodes with more intra-cluster edges than inter-clusters edges like in figure 2.

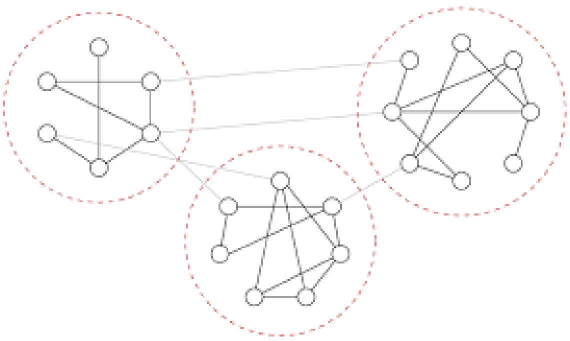

**Figure 2: An intra-clustering and Inter-clustering Community Structure**

There are more number of metrics or measurements used such as Hub/Authorities, Modularity, Density, Centrality/ Betweenness, and Clustering coefficient etc.,

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p. 1414-1420
https://publishoa.com
ISSN: 1309-3452

**Detecting the Influential Nodes**

The study of Influence Maximization (IM) has become a heated topic concerning the social network, especially in recent years, because it plays a key role in a wide range of fields, such as rumor control, network monitoring, information recommendation, and so on. Formally, the goal of IM is to identify the most influential network seed nodes as soon as possible and maximize the influence through the interaction between the seed nodes and other nodes simultaneously. Unfortunately, with the rapid growth of network scale, how to effectively solve the IM problem in reality, especially for the large-scale social networks, is very challenging. The most influential nodes can be identified by the time interval they communicate and vary across different topics. Detecting the influential nodes in dynamic social networks presents a very recent field that has gained considerable interest from researchers. One interesting approach is to update influential nodes incrementally taking into consideration the structural evolution of the social networks. Influence maximization in a social network refers to the selection of node sets that support the fastest and broadest propagation of information under a chosen transmission model. Nowadays, individuals interact with each other in more complicated patterns than ever. It is a challenging task to identify influencers in social networks for the various kinds of interactions.

**Related work**

This paper aims to effectively solve the problem of the influence maximization [13] in social networks. For this purpose, an influence maximization method that can identify influential nodes via the community structure and the influence distribution difference is proposed. Firstly, the network embedding-based community detection approach is developed, by which the social network is divided into several high-quality communities. Secondly, the solution of influence maximization is composed of the candidate stage and the greedy stage. The candidate stage is to select candidate nodes from the interior and the boundary of each community using a heuristic algorithm, and the greedy stage is to determine seed nodes with the largest marginal influence increment from the candidate set through the sub-modular property-based Greedy algorithm. Finally, experimental results demonstrate the superiority of the proposed method compared with existing methods, from which one can further find that our work can achieve a good tradeoff between the influence spread and the running time.

An effective influence assessment model based both on the total valuation and variance in valuation of neighbor nodes, motivated by the possibility of unreliable communication channels was proposed in [14]. We then develop a discrete moth-flame optimization method to search for influence-maximizing node sets, using local crossover and mutation evolution scheme atop the canonical moth position updates. To accelerate convergence, a search area selection scheme derived from a degree-based heuristic is used.

[15 ] proposed an incremental approach for detecting influential nodes by inspecting social networks evolution. First, we identify the influential nodes in the original network. Then, we propose a method for finding the changed elements. Finally, we present our algorithm for updating influential nodes in dynamic social networks.

A general multilayer network model [16] was developed to represent the multiple social networks, and then proposes the node influence indicator merely based on the local neighboring information. Extensive experiments on 21 real-world datasets are conducted to verify the performance of the proposed method, which shows superiority to the competitors. It is of remarkable significance in revealing the evolutions in social networks.

**Comparison**

| Parameters | User Behavior Analysis(%) | Influential Node Prediction (%) |
|---|---|---|
| Accuracy | 82.71% | 87.7% |
| TPR | 79.4% | 80.4% |
| TNR | 83.6% | 89.1% |

| | | |
|---|---|---|
| Computational time | t | 6.8t |

Conclusion:

In this paper, we have taken two challenges such as user behavior and influential node prediction and compared against the various parameters like accuracy, true positive rate and true negative rate and computational time.

References


1. Anni Sapountzi , Kostas E. Psannis " Social Networking Data Analysis Tools & Challenges"
2. Wei Deng, Rajvardhan Patil, Lotfollah Najjar, Yong Shi, Zhengxin Chen, "Incorporating Community Detection and Clustering Techniques into Collaborative Filtering Model", Information Technology and Quantitative Management (ITQM2014), ScienceDirect Procedia Computer Science 31 ( 2014 ) 66 – 74
3. Nursen Aydin," Social Network Analysis: Literature Review", AJIT-e: Online Academic Journal of Information Technology 2018- Cilt/Vol: 9-Sayı/Num: 34 DOI: 10.5824/1309-1581.2018.4.005.x
4. Shazia Tabassum, Sofia Fernandes, Fabiola Souza Fernandes Pereira, João Gama, "Social network analysis: An overview", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery · April 2018, DOI: 10.1002/widm.1256
5. M. Dhilber, S. Durga Bhavani, "Community Detection in Social Networks Using Deep Learning". Proceedings of International Conference on Distributed Computing and Internet Technology(ICDCIT) 2020, pp 241-250
6. Di Jin, Meng Ge, Zhixuan Li, Wenhuan Lu, Dongxiao He, Françoise Fogelman-Soulié, "Using Deep Learning for Community Discovery in Social Networks". Proceedings of IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), pp 160-167.
7. Pranav Nerurkar, Madhav Chandane, Sunil Bhirud," A Comparative Analysis of Community Detection Algorithms on Social Networks". Proceedings of Computational Intelligence: Theories, Applications and Future Directions - Volume I pp 287-298.
8. Feifan Wang, Baihai Zhang, Senchun Chai , and Yuanqing Xia, "An Extreme Learning Machine-Based Community Detection Algorithm in Complex Networks". Complexity, special Issue on Bio-Inspired Learning and Adaptation for Optimization and Control of Complex Systems Hindawi, Volume 2018, Article ID 8098325
9. Ling Wu, Qishan Zhang , and Chi-Hua Chen," Deep Community Detection Method for Social Networks". doi:10.20944/preprints201911.0019.v1
10. Craigory Coppola and Heba Elgazzar, "Novel Machine Learning Algorithms for Centrality and Cliques Detection in YouTube Social Networks". International Journal of Artificial Intelligence and Applications (IJAIA), Vol.11, No.1, January 2020 pp 65-77.
11. S. M. M. Salehi, A. A. Pouyan, "Detecting Overlapping Communities in Social Networks using Deep Learning". IJE Transactions C: Aspects Vol. 33, No. 3, (March 2020) pp 366-376
12. https://www.kernix.com/article/community-detection-in-social-networks/
13. Zufan Zhang, Xieliang Li, and Chenquan Gan, "Identifying influential nodes in social networks via community structure and influence distribution difference". Digital Communications and Networks 7 (2021) 131–139
14. Lu Wang, Lei Ma, Chao Wang, Neng-gang Xie, Jin Ming Koh, Kang Hao Cheong," Identifying Influential Spreaders in Social Networks through Discrete Moth-Flame Optimization". IEEE Transactions on Evolutionary Computation, DOI 10.1109/TEVC.2021.3081478.
15. Nesrine Hafiene, Wafa Karoui, and Lotfi Ben Romdhane," An incremental approach to update influential nodes in dynamic social networks". International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Procedia Computer Science 176 (2020) 781–790
16. Xinyu Huang, Dongming Chen, Dongqi Wang and Tao Ren, "Identifying Influencers in Social Networks". Entropy 2020, MDPI ; DOI:10.3390/e22040450

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p. 1414-1420
https://publishoa.com
ISSN: 1309-3452