

A Survey on Social Network Analysis and their Issues

Amit Aylani

Smt. S.R. Patel Engineering College
Computer Engineering Department
Unjha, Gujarat

Nikita Goyal

Shri G.S Institute of Tech. & Sci
Computer Eng. Department
Indore (M.P.)

ABSTRACT

Social network has become an important part of every individual's daily routine. Social Networking Sites (SNS) like Twitter, Facebook, LinkedIn and Google+ are user interfaces for social network which can be accessed through internet and web 2.0 technologies. It provides a platform for its users, where they connect, communicate, share, update etc. People are developing more interest for social networks as there is dependency for information, news, daily updates and psychology of other humans on a generalized scenario is easily available with a ease of access. With this increasing use and interest of any SNS a lot more enhancement and more utility can be provided for its users which is a crucial and a tedious area of interest for researchers with growing trend and demand. In this paper a attempt is made to summarize basic ideas behind a Social Network and working of a SNS with some existing work done and areas where there is scope of improvement by categorizing the research topics for a online social network.

General Terms

Analysis, Relationship

Keywords

Social Network, Social Networking Sites, Community, Social Connection.

1. INTRODUCTION

A Social network can be viewed as a graph containing nodes and edges. Where the nodes represent users and edges shows the relationships or interactions between users. A social network can be generalized as a network of interactions and social relationships with or within the users. This relationship which is represented by edges makes graph densely or sparsely connected i.e. density of any social network is directly proportional to the social relationship among the users. A Social Networking Site (SNS) is an online service or platform that creates an interface for users. It is the way in which they are connected through different social familiarity ranging from casual acquaintances to close bonds [1]. SNS builds and reflects social networks or social relations among people, who share their interests or perform activities. A SNS consists of a representation of each user's social links and a variety of services. Most social network services are web-based and provide means for users to interact over the internet. Social networking sites allow users to share their ideas, activities, events and interests within their individual networks.

Online social networks in the last few years have grown drastically. These online social networks are rich in data which makes it more challenging and a large amount of opportunities are available from the perspective of knowledge discovery and data mining. A Social network can mainly be categorized into following three types. First a personalized

social network developed for one to one connectivity of users and also for content based sharing i.e. Facebook. It is largest and most popular social network now a day's having billions of users and it also provides a provision for one to one connectivity and also for selection of users whom one wants to connect by using the concept of friend request which make it a personalized network. Facebook is used for sharing of contents like photos, videos, tagging, status, information, news etc. This huge amount of data and sharing interests, attract many user and which in-turn attracts many researcher to perform more analysis and develop new optimized algorithms for this growing network.

Second type of social network is based on public opinion and revolves around public figures or influential bodies of our surrounding i.e. twitter. It works with tweets and twitter handlers. By using tweet, users try to express their opinions for a particular topic or even by using tweet important messages are conveyed to one who relates to it irrespective of their high position in society. A public network contains user's sentiments, their public opinion, govt. policy issues, about natural calamities etc. In this social network, it contains millions of tweet data which is common to all users and hence can be used to analysis users sentiments which is again a critical research area. Third type of social network is mobile based social network which provide connectivity and media sharing i.e. Instagram which was designed with a motivation of providing a platform for sharing of images and videos through a mobile application. It provides users an instantaneous way to capture and share their life moments with friends through a series of (filter manipulated) pictures and videos [2]. It also has a concept similar to Twitter of sending and receiving follow requests but in a different style. It has features for content based searching for which data needs to be handled properly for providing best results in optimum time. Searching based on face recognition as it is a network of images and selfies can be of interest.

In this paper we summarize a survey on various social networks and present some areas of research and issues on social network. With growing social networking as dependency is more on internet and mobile a deep analysis and requirement gathering is very important to develop any new algorithm which works for these SNS or social network. Rest of the paper is organized as follows where seven sections (2-8) describes various research areas and issues for a social network which includes data extraction, analysis of social network, community detection, hash tags, social network connections, Sybil attack and sentiments analysis. Above mentions areas of research are then summarized and concluded in section 9.

2. DATA EXTRACTION

With the growing social network, size of data is also increasing proportionally. Handling this data accurately and

efficiently with algorithms with optimized memory and time complexity constraints is very essential. Along with this, extracting this data from any social network requires several techniques as these networks also have restriction because of their privacy policies as it also contains personal details of user's along with public information, which may be sensitive and needs to be kept confidential from users who are not friends or are connected to each other in any ways. In any social network data extraction is the most important and difficult which makes it a typical part for research area. For data extraction various SNS provide some techniques to extract public attribute information of users. Facebook API or graph API is used to extract user's information; this information is called as attribute. Attributes can be user's id, name, current affiliation, High school detail, home town etc. This API helps to extract various attributes from Facebook. Another data extraction technique used for twitter is REST API or Twitter API using which one can extract user's large amount of information i.e. tweets data, which can be used to analyze reactions of other users in order to track events and public issues, which identifies the human behavior and can also be used to detect the current trends and demands.

3. ANALYSIS OF SOCIAL NETWORK

Social network can be analyzed as a static network and as a dynamic network [3]. The networks which change with slow pace of time comes under static network and other which changes rapidly are the ones known as dynamic network. The main differences which arise are in the context of social network algorithms. Static networks on the whole are analyzed in batch mode over particular snapshots taken. For networks like, bibliographic network where new events occur slowly over time is analyzed using static analysis. On the other hand, for networks like instant messaging networks, continuous interactions occur at a very large rate which leads to network streams and therefore such networks are analyzed under the heading of dynamic network and are more challenging and tedious to analyze and is a subject of recent research.

Dynamic networks are also created by social networks which are mobile application, interacting users are continuously mobile or are changing their location over time. Berger-Wolf and Saia [4] proposed a framework for identifying communities in dynamic social networks, making explicit use of temporal changes. These dynamic social networks are depicted as dynamic graphs where the edges change constantly over time and these dynamic graphs lead to massive challenges in processing because of the extremely large number of connections between the entities which may need to be tracked simultaneously. For such dynamic graphs a application known as graph stream mining is required to perform efficient online analysis. Such applications summarize the network structure of the data in real time and use it for a variety of mining applications.

4. COMMUNITY DETECTION

In social network, community detection is a wide research area now a day. Community can be considered as a set of users where each user interacts more frequently within the group of users than outside the group [5]. Community is structure which is focused on sharing of different types of information such as messages, photos, and videos etc. within the group of user who interact more frequently. It play important role in providing personalized services, product marketing and increasing privacy from public users. Various community detection algorithms have been proposed by various researchers. A method proposed by Liaoruo et al. [6]

is based on determining common followers to find the community of users. This algorithm is based on maximum cardinality search; the problem is formalized in an optimization framework. By using common followers technique it may possible user have more than one followers. It is used to find user's behavior but it is not possible to find accurate community. Another approach by Xutao et al. [7] proposed community discovery scheme in multidimensional network. A group of users is found based on attributes such as tags, photos, comments and stories. This algorithm calculates the probability of visiting contents and compares their values to generate communities. It cannot find exact possibility of users contents because it is depend on user's choice. Newman et al. [8] proposed an algorithm to find communities in which all the user profiles are treated as nodes of a graph and the edges represent relation between the users. This algorithm works by iteratively removing edges until the social network graph becomes partitioned. At this point, different partitions are considered as communities. In order to determine the edge to be removed at each step, a metric known as "Betweenness Centrality" for each edge was proposed. To compute this metric the shortest path between each pair of vertices needs to be determined. This method is highly computation intensive. Vasavi et al. [9] proposed measurement of similarity between users on the basis of social, geographic, educational, professional attributes, shared interests, pages liked, common interested groups or communities and mutual friends. These attributes were assigned weights manually then string and semantic similarity metrics were used to predict the most similar profiles. Manual weights to every attribute are not possible to detect meaningful community. All the above discussed approaches are based on user's attribute, but to find members who interact with each other, there is a requirement of strong social relationship within a community. For this, to identify user's relationship with other users more analysis is required.

5. HASH TAGS

A hash tag is a string of characters preceded by the hash (#) character and it is used on platforms like Twitter as a descriptive label or to build communities around particular topics [10]. Hash tags, originally introduced in Twitter are now becoming the most used or a popular way to tag short messages on social networks. Hash tags in Twitter are considered very important keywords since they add valuable meta-knowledge to a particular piece of text that is limited to 140 characters. In order to track certain events and to annotate them properly users indirectly agree to hash tag Tweets with a predefined keyword. Zangerle et al. [11] assumed that the primary purpose of the hashtags is to categorize the tweets and facilitate the search. It is used to share user public opinion and their views. Researcher focus on hash tag classification, hash tags similarity to identify and user can follow same topic which is very useful for data mining purpose. The hash tags have also been studied in order to determine the feelings attached or hidden within the tweets. Rodrigues et al [12] described hash tags, which can be categorized into positive, negative or neutral. In this way, they count the number of tweets that are included to these categories to analyze, if the hash tags can be used in order to determine feelings depicted in a tweet. The above consideration lacks for, as hash tags do not always represent opinions, sometimes it also contains references for a event or to a person Using these hash tags, patterns of a user can be recognized which can further be used for prediction of his behavior and opinions and a hash tag actually works as a content classifier, using these hash tags tweets can be classified into categories which can further be

used for analysis and classifying these hash tags into accurate classes is a major area of research for social networks using these values.

6. SOCIAL NETWORK CONNECTIONS

For the analysis of social connections in a social network graph theory is used, probably the most important method from early times. Social network is determined by its key features such as nodes and links or edges. For a social network node defines users and links are representation for relation between nodes or users. Graph theory influence to identify and analyze user's social connections. There is various graph property and tools that are used to analyze the node, links and the relationship between them. Density, clustering coefficient, Degree centrality, Clique, Reachability

Matrix, Adjust matrix, Eccentricity Centrality, Betweenness Centrality, In-degree, out-degree etc. are the important graph properties which are used in social network to help in analysis of the user's behavior and interaction between them. These Graph properties used are available in various tools as inbuilt functions. These tools help to analyze the user's data and create a graphical view of a social network. Social Network Visualizer (SocNetV)[13] is a cross-platform, user-friendly software application for social network analysis and visualization of the network. Social Networks are created or imported in the form of mathematical graphs, where vertexes depict actors/agents and edges represent relations between these users. NodeXL [14] is open source tool that provides easy access to social media network data streams, advanced network metrics, text and sentiment analysis and powerful report generation. By using these tools and graph properties which are used to analyze the node, edges and which will identify their social connection or relationship, interaction and behavior. This analysis is the main research area to be focused.

7. SENTIMENTS ANALYSIS

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic [15]. It is important to identify the user's emoticons and sentiments in social network. In twitter social network, user expresses their opinion by using tweet. In the tweet data it is required to analyze according to user's view and opinion. There are several challenges to be faced while performing sentiment analysis. First is analyzing words of opinion which can be treated as ambiguous i.e. positive in one situation and negative in another situation. Second challenge in sentiment analysis is the way of expressing opinions, every individual has its own way of expression and this makes it difficult to justify opinion. In most of the traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. Asso et al. [16] discussed a new technique for sentiment analysis named as SSSA (Semantic Scoring Sentiment Analysis); this technique assigns score to the words of tweet which are used for classification and analysis of user opinion. This technique is less effective or can become more tedious as for words which express sentiments there can be several synonyms, due to which multiple classes for classification end up having same sentiment but with a different name. In Sentiment analysis, however, "the lunch was delicious" is very different from "the lunch was not delicious" which conflicts with the traditional text processing beliefs. People can be contradictory in their statements. Most reviews will have a combination of both positive and negative comments, which can be managed by analyzing sentences one at a time. However, in the more informal medium like twitter or blogs, the more likely people

are to combine different opinions and a negative and a positive comment together in the same sentence which is easy for a human to understand, but more difficult for a computer to parse it because some comment are in form of taunt or a dual meaning sentence which related to human sentiments and cannot be detected using a parser. Several research work is done in this sentiment analysis area for social network but still there is requirement of more techniques to perform this sentiment analysis which is text mining, Polarity classification [17], Tokenization, sentiment classification, data mining technique etc.

8. SYBIL ATTACK

Online social networks like Facebook, LinkedIn, Twitter and other social network contains a lot of sensitive data which makes it more vulnerable to attacks and networking threats. This attack includes a type of most common attack which is a sybil attack. In a sybil attack, attackers create and maintain several fake identities or fake accounts on a social network which are known as sybils and these sybils inject malwares into the social network. Sybil has a capability to develop several attacks simultaneously specially to those real users who tend to be more careless and do not verify the person before accepting its request of becoming a friend or even some allow strangers to connect with them. In social network, because of sybil attack it is very difficult to identify trusted node or fake node (sybil). Krishna et al. [18] proposed security against sybil attack in which sybil node is identified and a community is created which contains trusted users by identify their edges. Various researcher proposed several algorithms to detect Sybil node and protect from sybil attack. H. Yu et al. proposed SybilGuard [19] and SybilLimit [20] which are a decentralized approach to detect sybil node. W. Wei et al. proposed SybilDefender [21] to prevent sybil attack. To prevent Sybil attack it is required to determine the behavior of a sybil node and mainly identify its pattern. Sybil attack is a vast area in social network which requires a lot of research work and evolution of new techniques to be done.

9. CONCLUSION

Social network is a strong emerging network who have entered in every human's daily life has now become a necessity and a important part as a network. With growing demand and users more requirement of developing an efficient social network is increasing for this new networks can also be introduced along with enhancement in existing social networks. Increasing users, also the amount of data is growing which is a major area of concern as to store such large amount of data and use it in a optimized manner is also a necessity. Time and space complexities with growing users and amount of data needs techniques, which can help in mining data accurately, which is also responsible for the working of search engines within the social network. Search results should be accurate and within minimum amount of time. Another major issue related to these social networks are security and privacy. As the dependencies on online networks are increasing it is becoming more prone to attacks like sybil attack and other vulnerability threats. Apart from threats to security, privacy is also a concern with personalized details, images, videos which a user want to be shared with friends or with close acquaintances only along with privacy policies of social network. In this paper an attempt was made to summarize important research areas, issues and a brief introduction to working of a social network and for overall improvement and development of any network all the loop holes should vanish with enhancement using new approaches and techniques.

10. REFERENCES

- [1] A.Hanneman and M.Riddle, 2005 Introduction to social network methods, online at <http://faculty.ucr.edu/hanneman/nettext>.
- [2] Yuheng Hu, Lydia Manikonda and Subbarao Kambhampati, 2014. What We Instagram: A First Analysis of Instagram Photo Content and User Types. Weblogs and Social Media. In 8th International AAAI Conference.
- [3] Charu C. Aggarwal, 2011. Social Network Data Analytics. Springer New York Dordrecht Heidelberg. Library of Congress Control Number.
- [4] T. Y. Berger-Wolf and J. Saia, 2006. A framework for analysis of dynamic social networks. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining.
- [5] Jaewon Yang, Julian McAuley and Jure Leskovec, 2013. Community Detection in Networks with Node Attributes In IEEE 13th International Conference on Data Mining.
- [6] Liaoruo Wang, Tiancheng Lou, Jie Tang and John E. Hopcroft, 2011. Detecting Community Kernels in Large Social Networks. In IEEE 11th International Conference on Data Mining.
- [7] Xutao Li, Michael K. Ng, and Yunming Ye, 2014. MultiComm: Finding Community Structure in Multi-Dimensional Networks. In IEEE Transactions on Knowledge and Data Engineering.
- [8] M. E. J. Newman and M. Girvan, 2004. Finding and Evaluating Community Structure in Networks. In Physical Review E, 69:026113.
- [9] Vasavi Akhila Dabeeru, 2014. User Profile Relationships Using String Similarity Metrics in Social Networks. Social and Information Networks (cs.SI).
- [10] O. Tsur and A. Rappoport., 2012. What's in a hashtag?: content based prediction of the spread of ideas in micro blogging communities. In Proceedings of the 5th ACM international conference on Web search and data mining.
- [11] Zangerle, E., Gassler, W., 2011. Recommending #-tags in twitter. In Proceedings of the CEUR Workshop.
- [12] Rodrigues B., G. A., Silva, I. S., Zaki, M., Meira Jr, W., Prates, R. O., & Veloso, A., 2012. Characterizing the effectiveness of twitter hash tags to detect and track online population sentiment. Proceedings of ACM annual conference extended abstracts on Human Factors.
- [13] SOcNETV: Social Network Analysis and Visualization Software <http://socnetv.sourceforge.net>.
- [14] NodeXL: Network Overview, Discovery and Exploration for Excel <https://nodexl.codeplex.com>
- [15] G.Vinodhini and RM.Chandrasekaran, 2012. Sentiment Analysis and Opinion Mining: A Survey, In International Journal of Advanced Research in Computer Science and Software Engineering.
- [16] Asso Hamzehei, Mohammad Ebrahimi, Elahe Shafiei Bavani, Raymond K. Wong, and Fang Chen, 2015. Scalable Sentiment Analysis for Microblogs based on Semantic Scoring. IEEE International Conference on Services Computing.
- [17] Theresa Wilson, Janyce Wiebe and Paul Hoffmann, 2005. Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of the conference on Human Language Technology and Empirical Methods. In Natural Language Processing.
- [18] Krishna B. Kansara Narendra M. Shekokar, 2016. Security against sybil attack in social network. International Conference on Information Communication and Embedded System.
- [19] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman, 2006. SybilGuard: Defending against Sybil Attacks via Social Networks. Proceeding ACM SIGCOMM.
- [20] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, 2008. SybilLimit: A Near-Optimal Social Network Defense against Sybil Attacks. IEEE Symposium Security and Privacy.
- [21] W. Wei, F. Xu, C.C. Tan, and Q. Li, 2012. SybilDefender: Defend against Sybil Attacks in Large Social Networks. Proceeding IEEE INFOCOM.