

A Literature Survey on Algorithms for Detecting Communities in Online Network

Anupama tyagi

M.Tech Student
Department of CSE & IT
Madhav institute of technology & science
Gwalior, India

Dr.sanjiv sharma

Assistant professor
Department of CSE & IT
Madhav institute of technology & science
Gwalior, India

Abstract—Social network analysis is emerging field of research. In this field various researches have been conducted by various researchers. Recently, Researchers are interested for detecting community so that they can analyze structure of social network for sugrouping or partitioning. This paper focuses on overview of available researches of community detection and analyzed their performance on the basis of various measurements of social network analysis. Community detection is also shows cohesiveness of nodes in any social network

Keywords—community detection, cohesion, partitioning, online social network subgroup.

1. INTRODUCTION

Social network analysis [1] views social relationships in terms of network theory consisting of nodes and ties (also called edges, links, or connections). Nodes are the individual Communities within the network, and ties are the relationships between the Communities, such as members of online communities sharing information using relationships among them. In this paper a state of the art survey of the works done on community detection in social network. The main goal is to provide a road map for researchers working on different measures for tracking communities in Social Network This popularity provides an opportunity to study the characteristics of online social network graphs at large scale. online social networks have gained significant popularity. People make investments greater electricity in social media network than whatever different internet activity wherein many humans stay interact with the Face book and Twitter than every other social media network. Social network analysis focuses on the analysis of patterns of relationships among pepole, organization states and such social entities online social network provide social networking service over internet.the members of online social network can interact to each other in online communities.communities are considered is groups of

densely connected members that are only lossely connected to the rest of the network.

produces huge amount of user generated data. Such social media data accommodates rich source of records that is able to offer tremendous opportunities for organizations to effectively attain out to a massive quantity of target audience. Tracking the users topic of interest from to be had such large supply of statistics is challenging and very critical mission for targeting right audience in various domains e.g. Marketing, Politics etc. Also it's far located that users that are related to each other through dating hyperlinks in such social media network do no longer necessarily percentage precise hobby.

Thus for spreading right statistics to right target market it will become critical to pick out group of users with common interest in such huge network. Detecting community of such likeminded people from large social media networks can offer benefits to applications of diverse domains and for suggesting likeminded people to user which are still obscure to him/her. An essential viable issue in social networks is to find communities of customers based on their shared content and relationship with other users community in network is a sample with dense links internally and sparse hyperlinks externally.

These hyperlinks can be characterized with the aid of the content similarity among users, friendship among them and also different similarities of their personal data which include their location, gender, age and so forth. These close systems can then be used for various functions consisting of focused advertising schemes, terrorist cells.

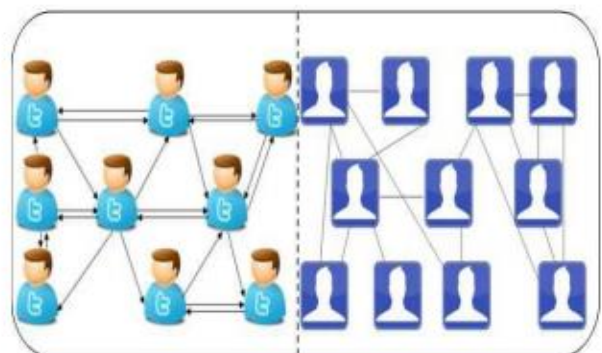


Fig. 1. First phase represents Follower-Followee courting between Twitter users at the same time as 2d segment represents Friend relationship among Facebook users

The social links of friendship is an important part of maximum social networks. These social links often provide upward push to communities in social media network. Communities in an social network may speak to genuine social groupings, maybe by intrigue or connection topology to recognize these communities. It will help to understand and exploit these networks more effectively. The ability to locate community structure in a network has sensible packages in various domain names. Most of the existing approaches for community detection depend on connection analysis and ignore the vast amount of other information available in social networks. Besides that greater part of the community detection algorithms separate whole network into disjoint arrangement of nodes. Which can be referred to as cohesively linked set of nodes or community.. This is unquestionably not valid in social networks. In social networks like Facebook and Twitter one client can be a piece of more than one cluster Also Twitter has different sorts of hyperlinks inside the form of follower following relationships, retweets, mentions and replies. Tweet incorporates rich source of textual statistics along with tweet tags e.g. Hashtag and mention. Additionally, Twitter provides loads of metadata within the shape of person area, age and gender which can be used for clustering.

II. DETECTION OF COMMUNITIES

In social networks, detection of communities is finished by observing the nodes which are like each other, i.e. keeping those nodes in one community. There is more number of nodes inside a single community, than in between two communities. When the nodes of a network may be arranged At the point when the hubs of a system might be orchestrated to shape a group such that the nodes are stated to have linked internally, then that network is said to have a community shape Community systems are pretty not unusual in real networks. Social networks include community companies (the starting place of the time period, in truth) based on not unusual location, pastimes, occupation, and so forth. Metabolic networks have communities in view of intentional groupings. Citation networks form communities by utilizing research topic Being able to become aware of these sub-structures within a community can provide insight into how community function and topology have an effect on each other. Such perception may be useful in enhancing some algorithms on graphs such as spectral clustering. The static properties are depicted by the structure of previews of charts in the specific system; while dynamic properties are portrayed by the network structure develops after some time. The above insinuated properties of social networks, can be for un-weighted or weighted charts, where weights may constitute multi-edges (e.g Multiple telephone calls starting with one man or lady then onto the next), or side weights (e.g. Money related sums among a benefactor and a beneficiary in a

political donation network) [2]. Meaning of community detection given by various creators is distinctive, and definitions are subjective. Every single strategy for recognizing communities has distinctive methodologies. Thus the result for detecting the communities by different method is different for the same network. So far, the researches achieved on community detection are in numerous approaches, and are based totally on an idea of various range [3]. The main advantage in the field of detecting communities is accessing the information from diverse sources and clusters. Detection of communities make possible to exchange Detection of communities make conceivable to trade and offer data since individuals in a group regularly have comparative tastes and longings. A community comprehends the structure of informal organizations since groups are considered as segments of social networks and determine the capacities notwithstanding elements of the network Communities' help in the visualization of big-scale social networks, and the relation of communities, clear the system of statistics sharing and dissemination of information Community detection also gives a vision on network future growth.

III. LITERATURE SURVEY

Chengling Zhao et al. [2016] this study aims to expand the current understandings on the behavioral patterns of learners' discussion within the Online Learning Space (OLS) at the Central China Normal University. This study also compares the differences in behavioral patterns between core participants and marginal participants. This study employed Social Network Analysis (SNA), Quantitative Content Analysis (QCA) and Sequential Analysis over a 44-days period of teaching activities joined by 40 college students as participants The results shows that the online discussion activity formed a relatively dense network of interaction, but the knowledge construction level contributed is relatively low, mostly at a low level stage of KC1, KC2, etc. Based on the findings, this study further implicates on teaching practices and future researches [5].

Mengnan Zhao et al. [2016] used social network analysis to characterize the extent of and relationship among users in a quit smoking intervention conducted through private groups on Facebook. Social media speaks to a promising system to deliver and engage participants in smoking cessation intervention. Previous analyses have highlighted the potential of this medium as a resource for requesting or offering information, discussing smoking, and exchanging social support in the context of a public quit smoking Facebook page [6].

Raji P S et al. [2016] different types of RDF representations and various SNA methods using RDF are studied. The first W3 Consortium standard to represent metadata with rich amount of information resource of the web is referred as Resource Description Framework (RDF). RDF semantic web

approaches are used on different areas of research today such as social network analysis, discovering semantic associations, semantic web engineering and so on. Social networking in modern society is a communication mechanism to generate large amount of user data in real time. Key capacities and soul of social network might be abused through Social Network Analysis (SNA) SNA used to extract network of interactions between organizations, peoples and transactions. Recent studies shows semantic web technologies are used to build up applications that can analyze data on various social networking sites [7].

Shufeng Ma et al. [2016] ciples from Communities of Practice inconvenience in the SNA train We then present the design of a social analysis network study that we are executing to study and analyze whether this change effort has been successful in achieving its goals. We present some preliminary data to demonstrate the promise of this approach for executing and studying institutional change in engineering education and STEM education more broadly. Creating systemic change in undergraduate engineering and STEM education is difficult to achieve and just as difficult to study. It has been proposed that organizational learning and change theories can be coupled with social network analysis to achieve both of these goals [8].

Mehmet KARAKOSE et al. [2016] proposes a new conceptual version for improvement of education lifeloop control method. The proposed model consists of three main stages and two supporting stages. This model has a natural loop between three main stages supported with two stages. The proposed overall approach has a loop that renew with a continuously loops and a structure that includes effects of all factors. The usability of comparable technology in the higher education is very important for supporting knowledge management at industry assisted technology. An Information Technology (IT) platform with administrators, instructors, and students is necessary to achieve a powerful learning content. This process requires an educational lifeloop management approach. This approach provides some features such as active and responsible students according to passive students, differentiated instructional methods according to standard instructional methods, student-focused learning experiences according to lecturer-based instructional methods, and analyzed, sorted, and provided student data in real time [9].

Chih-Chien Wang et al. [2016] attempt to adopt an actual case to research the social network of spammers by way of K-core and Clique analysis. Our research results show that the social connection among spammers is stronger than that among non-spammers. Consumer created item surveys are considered as more persuasive than commercial advertising, and are now an important message source to make purchase decision. Nevertheless, corporations may additionally deliberately procure spammers to make fake critiques to sell their merchandise and to demote merchandise in their competitors.

To create the opinion majority, firms may hire a group of spammers rather than just one or few individual spammers to write fake reviewers. These spammers may additionally act as a collection to assist different spammers to create a social consensus or majority of critiques [10].

Benjamin Cabrera et al. [2016] make a contribution in two methods to the usage of pattern distributions in networks evaluation. First, we introduce two one of a kind sampling-based totally algorithms for computing the referred to pattern distributions. Second we cartoon two unique thoughts which may be used to harness the records in pattern distributions for gaining insights on worldwide and neighborhood properties of Social Networks. Deciding the frequencies and the circulation of small sub graph designs in a colossal enter chart is a basic piece of many diagram based mining assignments, for example, Frequent Sub chart Mining (FSM) and Motif Detection. Due to the exponential wide variety of such graph patterns the interpretation of the mining outcomes is broadly speaking constrained to finding unexpectedly frequent patterns, and in standard identifying few in particular thrilling patterns to then understand their function in the network. However, the full distribution of patterns itself encodes much more information about the underlying graph. Looking at this sample distribution will be of specific interest in SNA due to the fact social networks appear to have extra random noise and much less hard structural constraints in comparison to other styles of graphs. This makes it unlikely to find meaning in only the most frequent patterns [11].

Madan Krishnamurthy, et al. [2016] Intends to discover developing tendencies in Mental Health and Behavioral Studies (MHB) which lays emphasis on co-current situations resulting in comorbidity. We present a unique method in which personality developments inferred from unstructured textual content of sufferers and well-known social users are compared via statistical evaluation. This is completed by using our Psychiatric Disorder Determination (PDD) algorithm. Further, Social media data of customers showing personality developments of patients is subjected to semantic primarily based textual content classification the usage of Natural Language Processing (NLP) and Ontology Based Information Extraction (OBIE) in our Addiction Category Determination (ACD) algorithm. This gives order of client diaries to regular themes of examination by alluding to ontologies DB pedia, Freebase and YAGO2s. The very last class therefore acquired may be expected to be a trending concern of subject for users with Psychiatric disorders developing Addictive behavioral personalities [12].

Alireza Farasat et al. [2016] study at the utility of SNA methodologies inside the data fusion region. Given fused data that combine a couple of intelligence reviews from the equal surroundings, social network extraction and high value individual (HVI) identification are of interest. The research at the feasibility of such sports can also help now not simplest in

methodological developments in community technological know-how but additionally in checking out and evaluation of fusion nice. This paper gives a parallel computing-based technique to extract a social network of people from fused facts, captured as a cumulative associated statistics graph (CDG). To gain the preferred social network, procedures which include a hop depend weighted and a direction salience technique are evolved and in comparison. A supervised mastering system is executed for parameterizing the extraction calculations [13].

IV. SOME IMPORTANT OBSERVATIONS IN THE EXISTING APPROACHES

As we've got already discussed a number of the prevailing strategies. So in this section we are going to have a look at some benefits and drawback in the existing techniques. The begin with algorithm was proposed for the detection of communities in social network is Newman-Girvan algorithm which is used for detecting communities' structure in both PC produced and true network data. [14]. For a calculation was proposed for the recognition of communities based on edge content and link content in the detection of communities in social media network. Edges offer a much richer characterization of network conduct, because the content material models the traits of pair wise interactions as opposed to individual actors'. The authors have examined the proposed approach for email and flicker based approaches only [15].

One of the ways for the formation of communities is clustering, so the author's have proposed a Boolean programming problem for the partition of clusters. In this approach it ensures that only a fixed number of clusters should be selected and when a data object is required to be assign then a cluster must be selected. The proposed Boolean programming problem approach is not cleared by author's that for which type of network will it work appropriately. The proposed community detection in a coordinated internet of things (IOT) and Social Network (SN) engineering approach have picked the graph mining method for settling the issue introduce in complex system of IOT and SN. The approach for detection of communities in an integrated environment is very useful, because in the case of any search operation performed by any node, the results obtained for intercommunity are more relevant than intercommunity. This integrated environment approach fails to give results for directed networks, such as Twitter, and is not generalized to all networks [16]. An empirical look at of community and sub-community detection displays the shape of communities as well as sub-communities happening in a social network by applying Newman-Girvan algorithm. The author's have implemented this network detection algorithm on real international networks. The results differ from those presented earlier in the sense that the author's have defined another idea of sub-communities. The main drawback of NewmanGirvan algorithm is that the computational cost is relatively high [17].

The detection of groups in distributed surroundings in Web-scale systems in view of a troupe learning plan which offers a way to become aware of high quality partitions from an ensemble of partitions with lower quality. Through the detection of communities in distributed environment it is possible to process a graph as large as 3.3 billion edges on small Hadoop cluster with 50 nodes in just a few hours. Density based absolutely spatial clustering of projects with noise (DBSCAN), is additionally utilized as a part of interpersonal organization investigation. The DBSCAN algorithm has a thought of noise, and is effective to anomalies. Through this approach is not an incremental approach, which means that it requires the whole records set earlier. Through this DBSCAN algorithm it is feasible to do away with outlier members which result in a correct clustering end result that enables with the network detection inconvenience in the SNA train [18]. Bayesian network (probabilistic graphical model) was used for showing the relationship between statistical model variables and statistical model is presented for the participations between casual social network's characters. The detection of communities was done based on the Expectation Maximization (EM) algorithm, which drive estimates for the model parameters. for the detection of groups. The algorithm works well for weighted, un-weighted networks and additionally for directed and un-directed networks [19]. Graph mining techniques also are very beneficial for extraction of understanding from the network graph, so a set of rules based totally on graph mining procedures got to be proposed for the detection of communities. The graph mining technique approach is very much useful for large number of nodes and is examined for finding communities in villages [20].

V. TABULATED VIEW OF DIFFERENT METHODS FOR DETECTION OF COMMUNITIES

Table 1. Different methods with its advantages and disadvantages [14]

SL. No.	Algorithms	Advantages	Disadvantages
1.	Community detection using DBSCAN algorithm	DBSCAN algorithm has belief of noise, and is robust to outliers.	DBSCAN is not so much deterministic i.e. fringe focuses that are reachable from more than one group can be a piece of either cluster
2.	Community detection method in an integrated	An actor or a node in the network can be part of multiple	This approach is not generalized to all the networks. It fails to give

	internet of things and social network architecture	communities, similar to real life situation. This approach can be used to suggest friends	results for directed networks, such as twitter
3.	Community detection strategy with edge content in online social media	Edge content gives better supervision to the community detection process	This approach is useful for email and flicker images based clustering only.
4.	Community detection methods in weighted network	Add up to weight of every single chose clusters in the network is calculated and also the similarity in between the clusters.	The proposed approach is not cleared for the type of network for it will work for.
5.	Community & sub-community detection method using new mangirvan algorithm	Author's have characterized another idea of detecting sub-communities	Computational cost is relatively high
6.	Community detection methods for distributed environment in web scale networks.	The proposed approach makes possible to process a graph as large as 3.3 billion edges on small Hadoop cluster with 50 nodes in just a few hours.	The runtime complexity is $O(3^{n/3}/m)$ (exponential) where n is the number of vertices and m is the number of nodes.
7.	Community detection method by using Bayesian Network and expectation Maximization technique.	It works well with directed and undirected networks and with weighted and un-weighted networks.	It requires indicating the quantity of communities ahead of time.
8.	Community detection technique in a social network the use of graph mining approach.	Useful for extremely large variety of nodes.	

9.	Community detection method based on spectral clustering	It not only find the real or near the real partition but also is able to find out solutions with underling hierarchical structures and fuzzy nodes that cal hardly be discovered by single objective optimization approaches	It may be sensitive to choice of parameters and computationally expensive for large datasets.
10.	Community detection method based on overlapping communities.	It detects the overlapping communities and also analyze the community structure	To detect the overlapping community only modularity method has been used as a fitness function.

CONCLUSION

Community detection algorithms provide a way for analyzing overall network and relations among components of organization. This paper shows the basic concept of social network analysis and provides scope for exploring the basic methods for detection of community in social network. Many researches reveal community detection mechanism using different formulation and algorithms. In spite of these, some limitations are found in previous research work in the field of community detection in social network.

REFERENCES

- [1] Kwan Hui Lim and Amitava Datta, Finding Twitter Communities with Common Interests using Following Links of Celebrities, MSM'12, Pages 25-32, ACM 2012.
- [2] Mary McGlohon, Leman Akoglu and Christos Faloutsos, "Statiscal properties of social networks," C. C. Aggarwal (ed.), Social Network Data Analytics, Springer Science + Business Media, LLC 2011.
- [3] Andry Alamsyah, Budi Rahardj, Kuspriyanto, "Community Detection Methods in Social Network Analysis", 2011 American Scientific Publishers.
- [4] Chengling Zhao, Yunzhen Liang, Qingtang Liu "Analysis of Social Network and Knowledge Construction Levels in Online Discussion" 2016 International Conference on Educational Innovation through Technology.
- [5] Mengnan Zhao, Christopher C. Yang , Johannes Thrul, Danielle Ramo "Patterns of interaction within smoking cessation groups on Facebook: A social network analysis" International Conference on Healthcare Informatics, 978-1-5090-6117-4/16 2016 IEEE.
- [6] Raji P S, Subu Surendran "RDF Approach on Social Network Analysis" International Conference on Research Advances in Integrated Navigation Systems (RAINS - 2016), April 06-07, 2016, R. L. Jalappa Institute of Technology, Doddaballapur.
- [7] Sushama Patil ,Mansi Bhonsle "Survey on Community Detection and Strength Analysis Methods" (IJETT) – Volume 36 Number 9 – June 2016.

- [8] Shufeng Ma, Geoffrey L. Herman, Matthew West, Jonathan Tomkin and Jose Mestre "Studying Faculty Communities of Practice Through Social Network Analysis" 978-1-5090-1790-4/16/\$31.00 ©2016 IEEE.
- [9] Mehmet KARAKOSE "Hardware, Software, and Human in the Loop Education Supported with Social Network Analysis and Mining" 978-1-5090-0778-3/16/\$31.00 ©2016 IEEE.
- [10] Chih-Chien Wang, Min-Yuh Day, Yu-Ruei Lin "Toward Understanding the Cliques of Opinion Spammers with Social Network Analysis." IEEE/ACM ASONAM 2016, August 18-21, 2016, San Francisco, CA, USA 978-1-5090-2846-7/16/\$31.00 ©2016 IEEE.
- [11] Benjamin Cabrera "Towards Using Subpattern Distributions in Social Network Analysis" 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).
- [12] Madan Krishnamurthy, Khalid Mahmood, Pawel Marcinek "A Hybrid Statistical and Semantic Model for Identification of Mental Health and Behavioral Disorders using Social Network Analysis" 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).
- [13] Alireza Farasat, Geoffrey Gross, Rakesh Nagi, and Alexander G. Nikolaev "Social Network Analysis With Data Fusion" 2329-924X © 2016 IEEE, Transactions on computational social systems.
- [14] M. E. J. Newman, M. Girvan, "Finding and evaluating community structure in networks", Phys. Rev. E, Columbia, vol. 69, pp. 1-15, August 2003.
- [15] Guo-Jun Qi¹, Charu C. Aggarwal² and Thomas Huang¹, "Community Detection with Edge Content in Social Media Networks", Data Engineering (ICDE), 2012 IEEE 28th International Conference.
- [16] Sudip Misra, Romil Barthwal, Mohammad S. Obaidat, "Community Detection in an Integrated Internet of Things and Social Network Architecture", Global Communications Conference (GLOBECOM), 2012 IEEE
- [17] Deepjyoti Choudhury, Saprativa Bhattacharjee, Anirban Das, "An Empirical Study of Community and Sub-Community Detection in Social Networks Applying Newman-Girvan Algorithm", Emerging Trends and Applications in Computer Science (ICETACS), 2013 1st International Conference.
- [18] Yomna M. ElBarawy, Ramadan F. Mohamedt and Neveen I. Ghali, "Improving Social Network Community Detection Using DBSCAN Algorithm", Computer Applications & Research (WSCAR), 2014 World Symposium, 2014 IEEE.
- [19] Ahmed Ibrahim Hafez, Abaul ella Hassanien, Aly A. Fahm and M.F.Talba, "Community Detection in Social Networks by using Bayesian network and Expectation Maximization technique", 2013 IEEE.
- [20] Bapuji Rao, Anirban Mitra, "A New Approach for Detection of Common Communities in a Social Network using Graph Mining Techniques", 2014 IEEE.
- [21] Mehjabin Khatoon a, W. Aisha Banu "A Survey on Community Detection Methods in Social Networks" I.J. Education and Management Engineering, 2015, 1, 8-18 Published Online May 2015 in MECS.
- [22] S. Yang, B. Wang, H. Zhao, and B. Wu. Efficient dense structure mining using mapreduce. In ICDMW '09. IEEE Conference, 2009.
- [23] Sanjiv Sharma, G.N. purohit, "A New Centrality Measure for Tracking Online Community in Social Network", IJITCS, vol.4, no.4, pp.47-53, 2012.
- [24] Sanjiv Sharma, G.N. purohit "Evaluation of Alternative Centrality Measure Algorithm For Tracking Online Community In Social Network", Volume No.1, Issue No.1, pp : 28-33 01 Nov. 2012.
- [25] Sanjiv Sharma, G.N. "Purohit Analysis of Spectral clustering approach for tracking community formation in social network", Vol.1, No.1, July 2012, pp. 31-37.