International Research Journal of Computer Science (IRJCS)
Issue 06, Volume 6 (June 2019)
SPECIAL ISSUE - 5th International Conference - "ACCE – 2019"
ISSN: 2393-9842
www.irjcs.com

IDENTIFY IMPORTANT USERS IN ONLINE SOCIAL NETWORK

Srikantaiah K C, Saikumar K, Shivaprasad K G, Ramya Shridhar Moger, Manu kumar R
Department of Computer Science and Engineering, SJB Institute of Technology, Bengaluru, Karnataka, India
srikantaiahkc@gmail.com, saikumark1237@gmail.com, shivaprasadkg1996@gmail.com,
ramyashridharmoger@gmail.com,manukumarr86@gmail.com

Manuscript History
Number: IRJCS/RS/Vol.06/Issue06/JNCS10109
Received: 29, May 2019
Final Correction: 30, May 2019
Final Accepted: 02, June 2019
Published: June 2019
doi://10.26562/IRJCS.2019.JNCS10109
Editor: Dr.A.Arul L.S, Chief Editor, IRJCS, AM Publications, India
Copyright:**©2019 This is an open access article distributed under the terms of the Creative Commons Attribution License, Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited**

Abstract—Nowaday's Online Social Networks (OSN) are Experiencing high rise in popularity due to increase in number of users in it. and it is necessary to know the users connections with other users present in these social structures where it can be done through forming communities based on their connections. In this we have formulated procedure of problem of finding the communities and high influencer or important user in aonline social network. The Communities of users are formed using Louvain algorithm in the social network. and by using Degree and Betweenness Centrality measures we have obtained the top influencer in all obtained communities.
Index Terms — Betweenness Centrality, Louvain algorithm, Social Network, Important Users.

I. INTRODUCTION

In the ModernWorld, due to the improvement of the internet and computer technology, many social networking sites have widely involved into our daily life. Where some of the Social networking sites are Facebook, Twitter, Snapchat and Google+. The universality of networks has made us sensitive to the new types of network risks. A Network is outlined as assortment of computers, servers, mainframes, other network devices, peripherals or alternative devices connected to one another to permitting the sharing of knowledge or information. Network can also be defined as a set of relationships (links) that connect people or objects (nodes).A Social network could be a website that permits individuals with similar interests to return along and share data, digital photos and digital videos. moreover Social networking sites enable users to share ideas, digital photos, videos and posts and to tell others concerning online or real-world activities and events with individuals in their network [1].

This may either used in good and bad ways i.e., sometimes it may be useful and sometimes it may be dangerous too. Since the information spreads faster in a social network, social network can be used for spreading or sharing information. This will be disadvantage when it come to the spreading fake news.

In this context of online social networks, The figure 1 will give us a clear concept that how the actual structure of a social network will appear. Though there are nothing strict rules that we must always extract the structure of the community. The community may be synonymous with clusters, groups in different contexts. The structure of the community is one of the great future in the social networks. The graphs that will generate randomly doesn't have such sort of feature. And the main aim is to divide the networks and to form the communities, wherever every node is assigned with separate community identity. Where It is related to the clustering algorithm. the concept of community detection is very important. If we consider practically, a community is a cluster of users who share similar interests, consume similar information or interact with each other in many other ways [2]. Often, in online social network a user may connect to other users who share content of their interest. A link between the two users will increases the possibilities of them sharing information or common interests to each other. Thus to determine

© 2014-19, IRJCS- All Rights Reserved

communities one needs to know both the link structure of the network and the information consumption behaviour of various users which makes it a challenging problem.

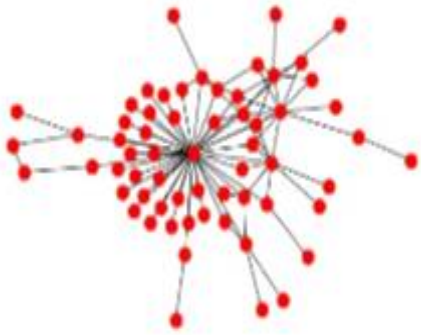

Figure 1: Online Social Network Structure.

Communities detection is important since it gives us a greater insight into the structure of the social network and can potentially has a number of practical applications such as predicting unobserved or future links between uses and for using marketing campaigns only for specific communities of users for greater effectiveness. and after forming communities influencer is identified based on the degree and between ness centrality of nodes or users which are connected to each other through links. Since the influencer nodeor important node has more possibility of sharing message fast in a social network.

II. LITERATURE SURVEY

Prior work on Communities identification, many methods are proposed. Some of them are Fuzzy community detection algorithm [7], Hierarchical clustering for community detection [8], Greedy optimization community detection algorithm, Girvan Newman community detection algorithm, Modularity maximization community detection algorithm and statistical inference community detection algorithm.Hierarchical clustering is used to group small clusters into large clusters at different levels according to the similarity of nodes. This technique failed to resolve the problem of fuzzy community structure.Greedy optimization community detection [9] uses optimize modularity of a partition of the network. It follows two steps: In the first step it looks for small communities by optimizing modularity regionally andIn the second step it aggregates nodes that belongs to the same community. This step is repeated until hierarchy of communities is formed.

This method will work with the variety of graph structures to detect task of communities. Few others planned ways area of sub-detection of community and detection of communities in on-line social network for the users.Statistical inference community detection is used signed networks i.e., positive links and negative links. The performance depends on predefined optimization (parameter estimation).

Community detection[3]for the weighted networks can use cluster method, the most motive is to maximise overall weight for all chosen clusters and maximize a similar options among the chosen clusters. Here overall weights of all chosen clusters square measure calculated with the similarity between the clusters. The community detection analysis is for characterizing sturdy grouping by the network property that supports in understanding the structure of social networks [4] also because the operate of the social network. It will conjointly serving to future social network growth. The methods used to discover the communities are all based on the agglomerative clustering. All these techniques can emphases size solely on the graph structure and their communities, however it doesn't think about the user interest and conjointly the impact the user's fame on the social networks. Sathiyakumari K and Vijaya M S proposed model dependent on Girvan Newman Algorithm procedure[5] for the location and investigation of community structure where it relies on the iterative disposal or elimination of edges, characterizing the edge betweenness of an edge as the number of shortest paths between pairs of nodes that run along it If there's quite one shortest path between a group of nodes in network structure,every path is allotted with equal weight specified the overall weight of all of the methods is adequate unity. If a network consists of the communities or teams that area unit loosely connected by a couple of inter-group edges.

Girvan and Newman outlined a new measure called as modularity. It had been introduced as a stopping condition for the community detection algorithm that uses edge betweenness to calculate modularity. Currently Modularity is employed wide for community detection and analysis of quality of partitions. This amount measures the fraction of edges within a community less the expected value of constant fraction. where The modularity of a partition value ranges between -1 and 1. Clauset et al [6] proposed the community approximation algorithm for optimization of modularity on large scale online social networks. However, this methodology has few drawbacks like it produces significally low values of modularity as compared to Louvain method and it additionally also as tendency to cluster nodes together on networks that don't have a major community structure.

© 2014-19, IRJCS- All Rights Reserved

III. SYSTEM ARCHITECTURE

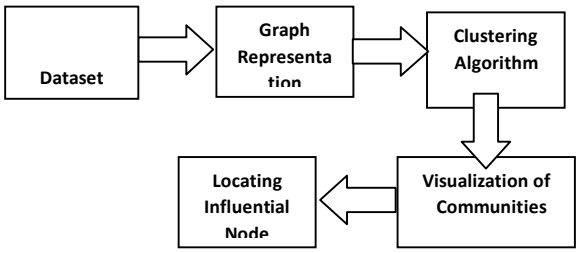

Figure 2: System Architecture

A. Problem Definition

The problem is to identify the Communities of users based on their connection that is how densely they are connected to each other in social network and to know the influencer users or important users in the obtained communities.

B. Data Collection

In this paper, we have made use of offline available Twitter Dataset. Where this twitter dataset consists of tweets data of users who have tweeted and retweeted for specific keywords in their timeline. We have extracted this dataset from figshare website. To download the dataset refer the link [4].

The Downloaded Twitter Dataset consists of four folders namely: annotations, threads, annotation_scheme.json, README. README file has all the description about the file downloaded and details of the data present in the file. Annotation file contains structure of the data present in the dataset. Threads file contains two folders: en and dn. We will be dealing only with file en i.e., English. File en contains eight tweets information of users for specific keywords. In that eight tweets we consider only one tweets information. Where this has the following document folders: images, reacts, source-tweets, urls-content, annotation.json, image.dat, reaction.json, structure.json, urls.json,who-follows-whom.dat. From dataset we extract the data of who follows whom connections of users and their followers and convert it into .txt file.

C. Graph Representation

A Network can be graphically represented using NetworkX package. Where it comes as a built in package in python. By making use of NetworkX the users and their followers present in the .txt file are represented in graphical format using this built in library in python. From this graph we can know the connections of users and followers.

D. Clustering

Clustering of users and their followers is done using the Louvain algorithm. Here Communities are formed based on modularity where it first identifies the small community and this small community is considered as single node in the network and this single node is made comparison with other nodes in the network this step continues till all the users in the network are in the communities.

E. Visualization of Communities

After applying Louvain Algorithm, the group or communities of users are obtained. These obtained communities are visualised using NetworkX and matplotlib packages in python.

F. Locate Influential Node or Important Users

The user or a node with high influence in a community of a network is identified using the betweenness centrality. Between ness centrality is a process of identifying the node with high information propagation in a network. Here the top ten nodes with high influence are fetched and influential Users or important users are visualized in communities.

Applications:

- By using the influencer nodes or users the marketing of new product can be made easy.
- It can be used for spreading both positive and negative message. The sensitive information can be circulated or spread easily by using this influencer nodes.
- It can be used to build a recommendation System for users.

© 2014-19, IRJCS- All Rights Reserved

IV. ALGORITHM

Algorithm Identifying _community

Purpose: Identifying Communities and Influencer or Important Users
Input: Twitter Dataset
Output: Communities and Important Node
Step 1: Inputing the data present in the dataset
Step 2: Preprocessing data present in dataset.
Step 3: Graphical representation of users present inthe Dataset.
Step 4: Apply Louvain algorithm to the graph, where it returns the communities based on modularity measures.
Step 5: Obtained communities are graphical represented or visualized.
Step 6: Apply Degree and Between ess centrality measures to the communities obtained to get the top influencer or important users in each community and Visualize it.
Step 7: Stop

Initially, we have considered offline available Twitter dataset where it consists of connections of users i.e., the users who follows whom based on the specific tweets tweeted by the users and there followers in there timeline inthe twitter online social network. The connections of users and there followers are present in .dat file format, where this .dat file is a binary data format so we convert it into .txt file in step 1.

In step 2 we will read the users and there followers who have tweeted for specific keywords, in step 3 these users are represented in graphical format by making of Networkx package with programming language python. In step 4 the converted Graph is provided as input to the Louvain algorithm and communities are formed based on maximisingthe modularity values of each obtained community. In step 5 obtained communities are graphical represented using network x package. In step 6 obtained communities are sent as input to between ness centrality to return the top influencers or important users in the network. Where between ness centrality is a centrality measure to fing the influence of a node or users over the flow of information between the every two pair of vertices, and this centrality measure assumes that information primarly flows only over through the shortest paths between them. Where this between ness centrality measure returns the important users in network. In step 7 algorithm terminates.

V. EXPERIMENTAL RESULT

A. Experimental Setup

Initially Data present in dataset will be in (.dat format) and from that file data nodes and their follower nodes are extracted and converted into (.txt file) using Python. From this file the nodes are represented in Graphical form using NetworkX Package (Figure 4), the Graphical form output is clustered (Figure 5) using Clustering Algorithm such as Louvain Method. From Cluster, the top influential node is identified by using Betweenness Centrality concept.
In Software Requirement we use a windows 10 as an Operating System, we are using a language as a python, we are using IDE as a Pycharm.
In Hardware Requirement we use a System Processor as Intel i5 core processor with 2.4 GHZ and 7th generation, we are using Hard Disk as a 500 GB and Ram as a 4 GB.

B. Results

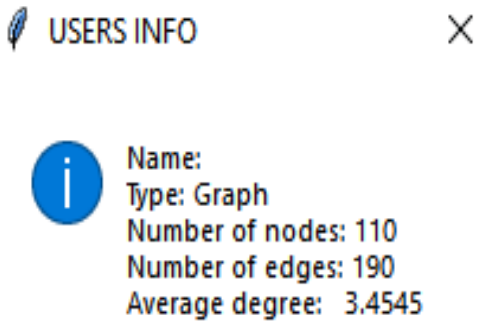

Figure 3: User Information

© 2014-19, IRJCS- All Rights Reserved

In Figure 3 describes the user information, where the dataset consists of 110 number of users or nodes, 190 number of edges where these nodes as average degree of 3.4545, the average degree describes number of edges connected to it.

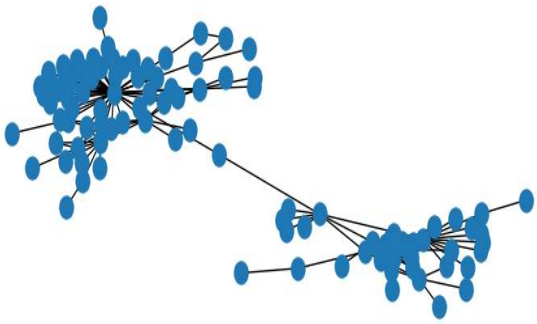

Figure 4: Graphical Representation of Users in Social Network.

In Figure 4, each node represents the users, where each user is connected with other users in the online social network. and nodes are interconnected with each other with the edges of undirected graph, In the blue colour we represent the nodes and in the black colour we represent the edges. From Figure 4, Graphical Representation of users before applying clustering.

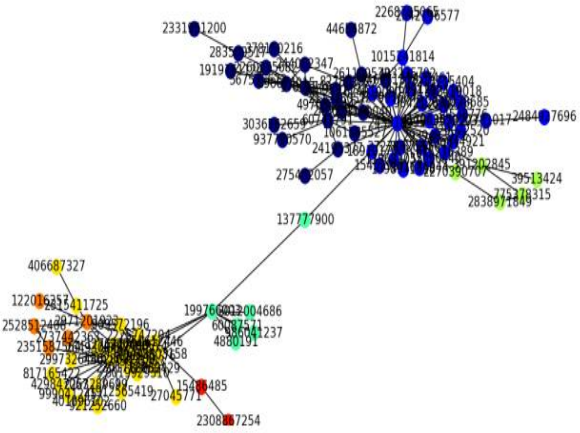

Figure 5: Obtained Communities

From Figure 5, we can see the graphical representation of users forming communities. Where communities are formed based on how well the users are connected with other users in social network.

UserIds of Users	Percentage of Influence
243318995	0.670283
199766003	0.424397
137777900	0.404179
331658004	0.236162
17801749	0.096603
435603158	0.093940
550612446	0.092730
60741791	0.037874
1015241814	0.033554
391202845	0.033129

Table 1: Top 10 influential users

The table 1 describes top 10 influential nodes, the name represents the user id, cent represents the percentage of influential node and the id 243318995 is the top influencer node in a network because it has a highest percentage of influential among all nodes.

© 2014-19, IRJCS- All Rights Reserved

CONCLUSION

In this paper, systematic study ofproblem of Identifying Communities in online social networks is modelled using Louvain algorithm where it works on maximising modularity values of the nodes. Louvain algorithm works on concept of how well the nodes are connected to the other users in the online social network. After Identifying Communities in a social network we have applied between ness Centrality concept to it, to obtain the top influencers or important users in the particular group of communities in online social network.These influencer or important users can be used for marketing of new or existing products or for building any recommendation system and for the future work we considered by making use of this important users we can know the way of information propagation in online social network.

REFERENCES

1. Social Networks, Information regarding Social Networking Services,[Online],Available: https://en.wikipedia.org/wiki/Social_networking_service
2. Natarajan, N., Sen, P andChaoji, V. "Community detection in content-sharing social networks," Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining ASONAM, 2013.
3. Shen, K., Song, L., Yang, X., and Zhang, W, "A hierarchical diffusion algorithm for community detection in social networks,"In Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 276-283, 2010
4. Ahajjam, S., El Haddad, M., and Badir, H. "Leaders Rank: Towards a new approach for community detection in social networks," In Computer Systems and Applications (AICCSA) in IEEE/ACS 12th International Conference , pp. 1-8, 2010
5. Sathiyakumari, K. and Vijaya, M. S. " Community Detection Based on Girvan Newman Algorithm and Link Analysis of Social Media," Communications in Computer and Information Science, 223–234.doi:10.1007/978-981-10-3274-5_1, 2016
6. Aaron Clauset, M. E. J. Newman and Cristopher Moore, "Finding Community structure in very large networks", Physical Review, E 70, 06611, 2014
7. Wang, Q and Fleury, E, "Community Detection with Fuzzy Community Structure," 2011 International Conference on Advances in Social Networks Analysis and Mining, doi:10.1109/asonam, 2011
8. Li, C., Bai, J., Wenjun, Z., and Xihao, Y, "Community detection using hierarchical clustering based on edge-weighted similarity in cloud environment," Information Processing & Management, 56(1), pp. 91–109.doi:10.1016, 2018
9. Jiang, Y., Jia, C and Yu, J, " An efficient community detection algorithm using greedy surprise maximization," Journal of Physics A: Mathematical and Theoretical,pp.47(16), doi:10.1088/1751-8113/47/16/165101, Vol. 23, 2014

© 2014-19, IRJCS- All Rights Reserved