Correlation analysis of community detection in social network of big data methodical using set theorem

K.M. Monica[1]*, R. Parvathi[2]

[1]Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India.
[2]Department of Computer Science and Engineering, VIT University, Chennai, India.
*Corresponding author E-mail:monica.km2016@vitstudent.ac.in

Abstract

A trending issue in the network system that aids in learning and understanding the overall network structure is the community detection in the social network. Actually, they are the dividing wall which divides the node of the network into several subgroups. While dividing, the nodes within the subgroups will get connected densely but, their connections will be sparser between the subgroups. The ultimate objective of the community detection method is to divide the network into dense regions of the graph. But, in general, those regions will correlate with close related entities which can be then said that it is belonging to a community. It is defined based on the principle that the pair of nodes will be connected only if they belong to the same community and if they don't share the communities, they are less likely to be connected. The vital problems across various research fields like the detection of minute and scattered communities have been necessitated with the ever growing variety of the social networks. The problem of community detection over the time has been recognized with the literature survey and the proposal methodology of set theorem to find the communities detection where the group belongs to activities. In addition to this, several basic concepts are stated in an exhaustive way where the research fields arise from social networks.

Keywords: Community detection, social network, big data, set theorem, probability.

1. Introduction

1.1. Big Data: Definition

A wide term for data sets, big data is either large or complex such that the applications available for the data process are inadequate. This involves several challenges such as study, capture, data duration, search, sharing, storage, transmit, image, information privacy, and querying. Often, this term refers to the extraction of the value from the data by using either the predictive analytics or some other advanced methods and seldom them to the exacting size of the data set. But, the decision making will be made confident with the accuracy of big data. Better decisions will end up in greater operational efficiency, reduction in cost, and risk reduction. There is a slight difference between the big data and business intelligence due to the growing maturity of the concept regarding their date and mode of usage. In order to measure the trends, measure things, and etc., the descriptive concept data with high information density has been utilized by the Business Intelligence. To disclose associations, its dependency and to execute the forecasting of outcomes and behaviors, several laws like regressions, non-linear associations, and its underlying effects have been inferred from the huge deposits of data with reasonably available substance and big data which makes use of inductive statistics and concepts from the non-linear system identification.

1.2. Social network

For an individual, a social network will be created by the way they interact with other people in the society and their relationships, social ties amongst the individuals through the rapid development of the internet facilities and the tremendous growth in the online interaction of the users. Apart from this, today there are lots of social networking sites have been arisen like the Facebook, Twitter etc. to facilitate the user interaction. With the tremendous increase in the number of interactions, tracking those communication means has turned out quite difficult. Currently, people are tending to interact with people of alike liking and tastes as the today social media sites were permitting them to extend their life with the society through unique means as it is rigid for people to join their associates and relatives in the tangible world. Though the social media sites, they find themselves very easy to discover their associates with same fascinations, interesting patterns and properties through existing communal network and exhibits a community structure, social networks were having a characteristic property. While partitioning, if the vertices within the network becomes disjoint or overlaps within the vertices, also when the number of edges within a set go beyond the number of edges among available two sets by rational amount, then it can be said that the system exhibits a community structure. Community detection is known as the method of ascertaining the organized grouping or clustering in the network and Social network analysis. In many applications, the recognition of clustering of people in network may be functional where grouping judgment are being taken, e.g., Distribution of messages through multicasting to a group of people instead of communicating to each one in the group or recommend a set of products to a community.

1.3. Social media pattern

Perhaps, the discussion about Big Data usage pattern is social media and client sentiment. However, Big Data may be used for

outlining what client's likings about the competition. Moreover, this new found insight can be used to figure out how decisions taken by you were getting impacted by this sentiment and a viable means for business connectivity. In precise, we could derive out the sales efficiency and receptivity of marketing campaigns, mix, about the product, price, promotion, placement and also the sentiment impacts.

1.4. Community structure

People inside a Group / Cluster can interact with every one outside the group very often by creating a community for them. If in case, any non-overlapping community has been found, the network will obviously get divided into group of nodes with domestic denser connectivity and the sparser connection among the clusters. Analyzing the social network also includes the learning and analysis of graphs for providing and identifying the patterns and the structures of the social network. Both the visual and mathematical analysis of human relationship will be provided by the social network. They are different type of community criteria; they are **Node-Centric Community**: Every nodes in the group must satisfy definite stuffs. **Group- Centric Community**: Connections within the group were considered as total and the group should convince different properties without taking a deeper look at the surface of the node. **Network-Centric Community:** The entire network set is partitioned and positioned with no elements in common. **Hierarchy- Centric Community:** Arranged entirely based on ranking structure of different communities and constructed.

2. Determination of sample analysis of communities detection

For example, on the origin of different languages, India is divided into a mixture of various regions inspite of the analysis of the communal activities such as metabolic networks, image processing and criminal networks, etc. community detection problem is have numerous applications. The different students are (nodes) from the same family and each of the students has taken different groups in colleges where it is divided into four departments and the groups are as follows. SET A :{ 1, 2, 3, 5,}, SET B: {4, 8, 10, 12, 15}, SET C: {6, 7, 11, 14} and SET D: {9, 13, 16}.In order to differentiate, the communities are highlighted with different colors according to the SET groups. 'SET A' highlighted with yellow, 'SET B' is highlighted with violet, 'SET C' is highlighted with blue and 'SET D' is highlighted with red. The colors mentions for to identifying the duplicate value are repeated in any SET groups and also detect the communities' detection of repetition in the network.

Communities in social networks are used since they are user-friendly pubic medium, permits users to enlarge their communal existence in unparalleled means and however it is not easy to meet up links in the physical world, and finds convenient to hunt friends online via alike opinions and interactions between nodes can help determine communities.

3. Fundamental literature survey of communities

1). Community detection method based on overlapping communities

In addition to the structure of the community, the overlapping also stand as one of the social network characteristics and the community overlaps will be examined if an individual belonging to additional social community. In general, an individual may fit in to more communities based on their wishes to have connections in several social groups such as family, friends, and colleagues. Today, a large number of groups which are available and planned in future might be divided into two categories as node based and link based overlapping community detection algorithms.

Network nodes categorization solely based on the node overlapping quality and community detection algorithms whereas the edges of network within the clusters may be classified by the link based overlapping community detection algorithm. Many of the research have been anticipated in order to detect the communities that overlap and the parliamentary optimization algorithm which aid in analyzing the structure of the community [1].

2). Community detection method in a Social Network using graph mining technique

In a social network, graph mining techniques may be used to identify the group in a social network on the basis of mutual understanding, node reach, and the degree of node , either the community or a group will be detected in this proposed approach. Also sub-graph is formed based on the consideration of more than two nodes, a group can be shaped completely which are adjacent to each other termed as a clique. Inside a community, the reach ability is possible between the two actors or nodes only when a path lies among those nodes. The nodal degree condition should be checked whether the nodes within the group are relatively next to a big number of cluster members or not [2]

3). HOC tracker: tracking the evolution of hierarchical and overlapping communities in dynamic social networks

Detecting the overlapping communities is the only challenging part in the community detection and this condition occurs when a specific node of a network is belonging to different communities simultaneously. The author to unified framework: HOC Tracker, for tracking the evolution of hierarchical and overlap communities in social network has been proposed and community detection methods will have a common constraint while studying the community detection and community evolution problems separately.

The novel density based framework called as HOC tracker for tracking all the community evolution in the social networks. Unlike another existing HOC tracker, an efficient log based approach has been used for mapping all the evolutionary relationships between the communities that occur at two successive times of a dynamic network. The main element of HOC Tracker involves examining the limited neighborhood of each node in the network considering the time complexity. For each of the nodes, the expenditure is proportional to its out degree [3].

4). Node-centric detection of overlapping communities in social networks

In recent years, a great attention has been found to detect the overlapping communities while a node belong to numerous communities as the researchers focused only on detecting disjoint communities. NECTAR, a node-centric overlapping community detection algorithm is presented in this paper which generalizes node-centric local search heuristic so that it can be applied for network processing overlapping community structure. A unique feature of NECTAR is that it decides where to use the weighted overlapping community clustering algorithm which is applied for overlapping community detection, depending on the structure of the graph. This algorithm will definitely provide the good results on graphs both at the low level and high level community overlaps. It is the first community detection algorithm that selects dynamically which objective function is to use based on the graph on which it is invoked. This detection algorithm outperformed all other algorithms in terms of average detection value and was best or second best for almost all networks [4][5].

5). Ranking features in facebook to detect overlapping communities

The risk of fake discoveries has been increased in the social networks with the large data and identifying the communities in Facebook which requires simple and effective fast technique to do the analysis of communities in graphs. Currently, existing community detection methods were suffering from the high computational cost which is caused by the huge structure. A new approach called as RELNA has been proposed to rank attributes of the social network based on the homophile property. By comparing the two algorithms namely RELNA and LINKREC (ranking algorithm) with Facebook datasets, the results were almost same. CESNA (overlapping detection community algorithm) were used to detect communities in order to make it precise, and these results show that this approach is much faster when comparing with other algorithms [6].

6). A novel algorithm for community detection and influence ranking in social networks

An inherent fact of influence-based connectivity and proximity encoded in the network topology has been developed and planned by a narrative algorithm for both the community detection and influence ranking. Here, the idea is drawn from the Page Rank algorithm with the sense that we cannot rely solely on the node degree. A new influence diffusion model has been proposed that embed influence into a node and passes it around in the network. Moreover, not only the total influence a node spreads is found out, but also where and how much the influence was distributed is kept tracking in order to build its influence vector for community detection. Here the algorithm is tested on valid world datasets and a set of virtual networks using the LFR model and the performance was evaluated by comparing with the ground truth and a set of delegate algorithms [7].

4. Proposal methodology for communities' detection using set theory'

4.1. Problem identification

Judgment of communities detection is a network within a limitation chosen of path in randomly, so the computationally is very difficult task due to the unequal size and density is varying depends upon the communities of the group. In spite of these difficulties, several methods are introduce for community to finding the various method, developed, algorithm are making with changeable levels of technical in receipt of the success. Few method of algorithm are discuss below

Minimum-cut method: The network is separated into a predetermined number of parts or events, and equal size can be distributed approximately, then it chosen that the number of edges between parts or groups is to be minimized.

Hierarchical clustering: Likeness measure calculation is compare with similarity node pairs.

Girvan–Newman algorithm: Identifies edges of the network are lie down to their communities and stick together for services, if any requirement to remove from the node and just leaving or behind communities themselves.

Modularity maximization: The modularity maximization method used to detection communities by pointed over their possible divisions in the network, if possible added one or more for have high modularity for strength to remove the detection.

Clique-based methods: Cliques are the sub-graphs in every node, and it is connected to every other node of the clique to strength the bound in a network.

From above observation of the various algorithms to be focus on the better method over other methods, not specific to finding exact detecting communities or not framing the group communities for better networking. The main objective of research is, not articulate the possibility or probability of detection communities of group activities. Through the set theory concepts the problem of these detecting communities can be articulate as solution of the research.

4.2. Proposal methodology: approaches to probability using set theory for finding the Communities' detection

A primary concept of set theory is that a membership belonging to two events or to a set or any object, enables. The member of an element of that set, and an object in sets may be a numbers, people, cars etc. They are two destination process (i) $X \in A$, which may be understood as "x belongs to A", or "X is an element of the set A". (ii) In contrast, $X \notin A$; it means "x does not belongs to A", or "X is not a member of the set A".

The fundamental concept in 'Set theory' is an Inclusion. Let 'A' and 'B' are the two sets of probabilities 1 and 'A', 'B', 'C' are the three set of probabilities 2. Let us discuss about the probabilities of inclusion set.

Probability 1: If every element of A is in element B, then A is called the subset of B, or A is said to be included in B, or B includes A. This relation is denoted by $A \subseteq B$, or equivalently $B \subseteq A$. Alternatively, $A \subseteq B \Leftrightarrow (x) (x \in A \rightarrow x \in B) \Leftrightarrow B \subseteq A$. When compare with probability of communities is (A, B) or (B, A).

Probability 2: If every element of A is in element B, then A is called the subset of B, or A is said to be included in B, or B includes A. A is said to be included in C, or C includes A. Symbolically, this relation is denoted by $A \subseteq B$, $B \subseteq A$ and $A \subseteq C$. Alternatively, $A \subseteq B \subseteq C \Leftrightarrow (x) (x \in A \rightarrow x \in B \rightarrow x \in C) \Leftrightarrow B \subseteq A \subseteq C$. When compare with probability of communities is (A,B) or (B,A) or (A,C) or (B,C).

Using Venn diagram of inclusion set theory, it is represent $n(A \cup B) = n(A) + n(B) - n(A \cap B)$. Whereas n $(A \cap B)$, may be called as communities detected i.e not belongs to any communities. The below figure 1.1 shows the Venn diagram clearly about the union and intersection of the communities.
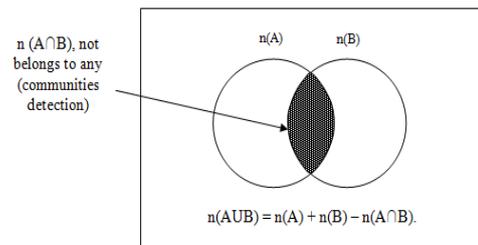
Figure 1.1: Union and intersection of communities

The detection level can be significantly found by using inclusion set theory of communities $n(A \cap B)$. For example, to more than three sets communities detection can be find out through $n(A \cup B \cup C)=n(A)+n(B)+n(C)-n(A \cap B)-n(B \cap C)-n(A \cap C)+n(A \cap B \cap C)$ and same for find the detection more than three number of communities is increase in union of inclusion set. The figure 1.2 shows the result of each community belongs with other, for example X is the result of A-B, Y is the result of B-A, Z is result of 'not belongs to any one' i.e communities detection from a set or group or events.
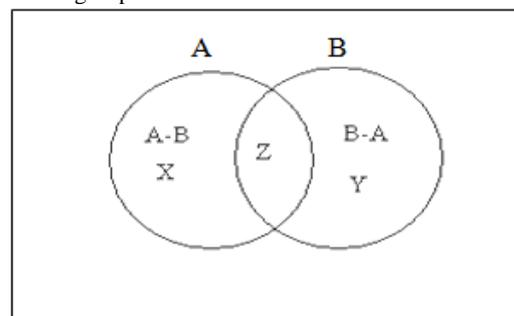
Figure 1.2: Inclusions and equality Set of 'SET A' and 'SET B'

The certain mathematical notations are representations are provided is in Table 1.1. These notations that use set operations have been derived for these properties for various combination of communities' detection of social network or any network. The suggested notations have been validated is depending upon the particular nature of properties and status, an suitable mathematical notation of set theory is suggested in table.

Table 1.1: Usage of Set Theory in Various Combination of Communities' Detection for Union and Intersection

S.No	Properties	Syntax	Justification
1.	Inclusion of Sets	$A \subseteq B \Leftrightarrow$ (x) (x∈ A→ x ∈B) $\Leftrightarrow B \subseteq A$	A is included in B, or B is included in A.
2.	Equality of Sets	$A = B \Leftrightarrow (A \subseteq B \wedge B \subseteq A)$	A and B are equal in number.
3.	Intersection of Sets	$A \cap B = \{x \mid (x E A) \wedge (x E B)\}$	Intersection of any two sets of A and B, written as $A \cap B$ is the set consisting of detection of communities or not belongs to any community
4.	Disjointed Sets	$A \cap B = \phi$	A and B have not in common

It satisfies the set theory link based on the SET A and SET B where the behaviors of communities in same group can find the communities detection (not belong to any group).The community detection is used to reduce the risk which would be effective during growth time, on effective memory utilization of the network and on-time delivery of the information to the network.

5. Conclusion

A way to discover the network structure is the community detection where the interconnections among the nodes are found to be denser than the intra-connection among the communities. In social networks, a hard problem is to discover the perfect community structure as the size of the network and sparseness of the network makes the task very difficult. Existing methods are restricted by their computational workings and depend heavily on the network topology, which fails in scale-free networks. Surveying numerous papers gave an idea on how the community detection was made with various algorithms and what are the pitfalls for each algorithm. Even though the algorithms were showing the different accuracy point in the result still, there are some drawbacks in each survey. An overlapping detection remains to be challenging till now. Hence, several graph-based algorithms are tried out to make the accuracy more perfect when compared with the other methods. Therefore, so the emphasis should be given to set theorem and Venn diagram of probabilities of given group which may find out communities detection in effective analytic of data set and it will be able to detect communities in a huge social network in allowable time.

References

[1] Feyza A & Bilal A, "Overlapping Community Detection in Social Networks Using Parliamentary Optimization Algorithm", *International Journal of Computer Networks and Applications*, Vol.2, No.1, (2015).

[2] Rao B & Mitra A, "A new approach for detection of common communities in a social network using graph mining techniques", *IEEE International Conference on High Performance Computing and Applications (ICHPCA)*, 2014, pp. 1-6.

[3] Bhat SY & Abulaish M, "HOC Tracker: Tracking the evolution of hierarchical and overlapping communities in dynamic social networks", *IEEE Transactions on Knowledge and Data engineering*, Vol.27, No.4, (2015), pp.1019-1013.

[4] Cohen Y, Hendler D & Rubin, A, "Node-centric detection of overlapping communities in social networks", *International Conference and School on Network Science*, Springer, Cham. (2017), pp.1-10.

[5] Zhan W, Guan J, Chen H, Niu J & Jin G, "Identifying overlapping communities in networks using evolutionary method", *Physica A: Statistical Mechanics and its Applications*, Vol.442, (2016), pp.182-192.

[6] Martinez-Seis B & Li X, "Ranking features in Facebook to detect overlapping communities", *IEEE 13th International Conference on Networking, Sensing, and Control (ICNSC)*, (2016), pp.1-6.

[7] Wenjun W & Nick W, "Street Department of Management Sciences University of Iowa, Iowa City, IA 52242, USA A Novel Algorithm for Community Detection and Influence Ranking in Social Networks", *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, (2014).

[8] Mahajan A & Kaur M, "Various approaches of community detection in complex networks: a glance", *International Journal of Information Technology and Computer Science (IJITCS)*, Vol.8, No.35, (2016).