

Correlation Analysis Of Community Detection In Social Network Of Big Data Methodical Using Set Theorem

^[1] K.M. Monica, ^[2] Dr. R. Parvathi

^{[1][2]} Department of computer science, Vel's University, Chennai, Tamil Nadu, India

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, the 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. Partitioning the network into dense regions of the graph is the ultimate aim of the community detection method. But, in general, those regions will correlate with close related entities which can be 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.

I. INTRODUCTION

A. Big Data: Definition

A broad 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 analysis, capture, data duration,

searching the terms, sharing the data, storage they require information and transfer, visualization, information with privacy and security, 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 particular 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 detect trends, measure things, and etc., the descriptive concept data with high information density has been utilized by the Business Intelligence. To disclose the relationships between the dependencies and perform predictions of outcomes of behaviors, several laws like regressions, non-linear relationships of data, and causal effects have been inferred from the large couple sets of data with low information density and big data analytic to makes use of inductive statistics analysis report based on the concepts from the non-linear classification identification of data set.

B. 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, and 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 is 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 the people for similar kind of likings and tastes, as on today social media sites were permitting them to extend their life with the society in unprecedented ways as it is harder for people to meet their friends and relatives in the physical humanity. Though the social media sites, they find themselves very easy and faster between their friends with similar interests' online, interesting patterns and properties through these real-world social networks and exhibit a community structure, social networks were having a characteristic property. While partitioning of the network into either disjoint the connectivity or overlapping with their sets of vertices, if the number of edges within a set exceeds to the number of edges between any two sets or more number of set by rational amount, then it can be said that the network displays a community structure. Community detection is known as the process of discovering the coupling and cohesive groups or it may be called as clusters in the network or social network analysis. In many applications, the communities detection of social networks may be useful for group decisions or joining or together for communication any action taken, e.g., multicasting a message of interest to a community instead of sending it to each one in the group or recommending a set of products to a community.

C. Social Media Pattern

Possibly, the most discussed Big Data usage pattern is social media and customer sentiment. Big Data can be used to figure out what customers are saying about your competition. Moreover, this new found insight can be used to figure out how the decisions taken by you were getting impacted by this sentiment and the way your company engages. More specifically, to determine the sales, the effectiveness or receptiveness of marketing campaigns, the accuracy of marketing mix like the product, price and so on were being impacted by the sentiment.

D. Community Structure

People within a group structure and it can be interact with each other as well as with the outside of group more frequently by creating a community for them. If in case, any non-overlapping community has been found, the network will obviously get divided into groups of nodes with internal denser connections and the sparser connection between the groups. 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 four type of community criteria; they are (i) Node-Centric Community: Each of the nodes in a group must satisfy certain properties. (ii) Group-Centric Community: connections within the group were considered as a whole and the group should convince different properties without taking a deeper look at the node-level. (iii) Network-Centric Community: The entire network is partitioned into various disjoint sets. (iv) Hierarchy-Centric Community: A group with hierarchical structure of different communities was constructed.

II. DETERMINATION OF SAMPLE ANALYSIS OF COMMUNITIES DETECTION

For example, on the origin of different languages, India is divided into various regions and beyond the scrutiny of the social behaviors e.g. in social networks, image processing and cyber networks etc. The community detection problem is having 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 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 easy access of social media, agree to people to extend their social life in unparalleled ways and when it is not easy to meet friends in the physically, but much better way to find the friend online with similar interests/group and interactions between nodes can help determine communities.

III. FUNDAMENTAL LITERATURE SURVEY OF COMMUNITIES

A. Community detection method based on overlapping communities

In addition to the structure of the community, the overlapping is also best example for social network attributes and these overlapping communities will be detected to which group is belonging to more than one social community/group. In general, a person may belong to more than community if he/she wishes as they might have connections to several social groups such as family, friends, and colleagues. Today, a large number of communities have been proposed which can be then divided into two categories as node based and link based overlapping community detection algorithms.

The network nodes will be categorized by the node based overlapping community detection algorithms whereas the edges of the network in clusters will be classified by the link based overlapping community detection algorithm. In order to detect the overlapping communities, many of the researchers have proposed the parliamentary optimization algorithm that helps in analyzing the structure of the community [1].

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

In a social network, the graph mining techniques and tools have been used to detect the communities' identification in a social network. Based on the complete mutuality, nodes reach ability, and nodal degrees of path, either the community or a group will be detected in this proposed approach. If a sub-graph is formed by considering more than two nodes, a group can be shaped completely which are neighboring to each other termed as a clique/group. Inside a community, the reach ability is possible between the two actors or nodes only if there is a path between those nodes. For the nodal degree condition, it should be checked whether the nodes within the group are relatively adjacent to a large number of group members or not [2].

C. 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 particular node of a network is belonging to various communities simultaneously. The author to unified framework: HOC Tracker, for tracking the development of hierarchical system and overlapping with the communities in social networks has been proposed and community detection methods will have a common limitation while studying the community identification and community evolution problems independently.

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 most important part of HOC Tracker involves examining the local neighborhood of each node in the network considering the time complexity. For each of the nodes, the cost is proportional to its out of the degree [3].

D. Node-Centric Detection of Overlapping Communities in Social Networks

During the recent years, a great attention has been got in detecting the overlapping communities while node is belonging to several communities as the researchers were focused only on detecting disjoint communities. NECTAR, a node-centric overlapping community detection algorithm is presented and it generalizes node-centric local search heuristic so that it can be applied also for network processing overlapping community structure. A distinctive 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 organization 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 from network which objective of the function is to use based on the graph. This detection algorithm is performed the average detection quality and was best among or second best for almost all networks in the communication [4][5].

E. 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 or social media requires simple and effective fast technique, analysis the detection and analysis of communities in graphs. Currently, existing community detection methods were suffering of 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].

F. A Novel Algorithm for Community Detection and Influence Ranking in Social Networks

An inherent knowledge based connectivity and closeness encoded in the network topology has been developed and proposed by a novel algorithm in both the community detection and influence ranking. A new influence diffusion model has been proposed that embeds with influence into a node and passes it around in the network communities. Moreover, total influence a node spreads is found out, but also finding correlative distributed is kept tracking in order to build its influence vector for community detection. This algorithm is tested through the validation and verification methods on real world datasets and these set of simulated networks using the LFR model with the performance analysis was evaluated by comparing with the ground truth and executed of the algorithms [7].

IV. PROPOSAL METHODOLOGY FOR COMMUNITIES' DETECTION USING 'SET THEORY'

A. 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 introduced 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 of every node network, and it is related 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.

B. Proposal methodology: Approaches to probability using set theory for finding the Communities' detection

A primary concept of set theory is that a membership between 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 a 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 : An 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. Symbolically, this relation is denoted by $A \subseteq B$, or equivalently $B \supseteq A$. Alternatively, $A \subseteq B \Leftrightarrow (x) (x \in A \rightarrow x \in B) \Leftrightarrow B \supseteq A$. When compare with probability of communities is (A, B) or (B, A).

Probability 2: An 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)

By 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 mention the Venn diagram clearly about the union and intersection of the communities.

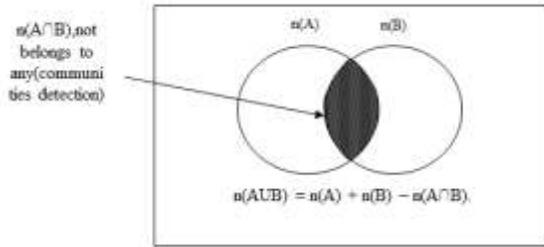


Figure 1.1 Union and intersection of communities

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 found 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.

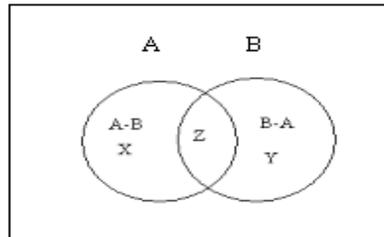


Figure 1.2 Inclusions and Equality Set of 'SET A' and 'SET B'

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 Set Theory communities' detection.

| S. No | Properties | Syntax | Justification |
|-------|-------------------|--|--|
| 1. | Inclusion of Sets | $A \subseteq B \Leftrightarrow (x) (x \in A \rightarrow x \in B)$ $\Leftrightarrow B \supseteq A$ | 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 | $A \cap B = \{x \mid (x \in A) \wedge (x \in B)\}$ | The intersection of any two sets of 'A' and 'B', then written as $A \cap B$ is the set consisting to detection of communities or not belongs to any . |
| 4. | Disjointed Sets | $A \cap B = \phi$ | A and B have not in common |

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

V. CONCLUSION

A way to discover the network structure is the community detection where the interconnections between the nodes, that are found to be denser than the intra-connection between the communities group. In social networks, a solid problem is identify the perfect community structure as the size of the network and sparseness of the network that makes the task as very difficult. Existing methods are limited by their computational works and depend heavily on the network topology, which fails into the 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 Altunbey and Bilal Alatas, "Overlapping Community Detection in Social Networks Using Parliamentary Optimization Algorithm", International Journal of Computer Networks and Applications Volume 2, Issue 1, January - February (2015).
- [2] Bapuji Rao, Anirban Mitra , "A New Approach for Detection of Common Communities in a Social Network using Graph Mining Techniques", 2014 IEEE.
- [3] Sajid Yousuf Bhat and Muhammad Abulaish, Senior Member, HOCTracker: Tracking the Evolution of Hierarchical and Overlapping Communities in Dynamic Social Networks. IEEE Transactions on Knowledge and Data Engineering, VOL. 27, NO. 4, APRIL 2015
- [4] Yehonatan Cohen, Danny Hendler, Amir Rubin "Node-Centric Detection of Overlapping Communities in Social Networks" Computer Science Department, Ben-Gurion University of the Negev, Be'er-Sheva, Israel fyehonatc,hendlerd,amirrubg@cs.bgu.ac.il 2016
- [5] W.Zhan, J.Guan & H.Chen "Identifying overlapping communities in networks using evolutionary method Of Journal Physics A: 2016
- [6] Proceedings of 2016 IEEE 13th International Conference on Networking, Sensing, and Control Mexico City, Mexico, April 28-30, 2016 Ranking Features in Facebook to Detect Overlapping Communities
- [7] Wenjun Wang and W. Nick 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, 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)
- [8] Various Approaches of Community Detection in Complex Networks: A Glance I.J. Information Technology and Computer Science, 2016, 4, 35-41 published Online April 2016 in MECS (<http://www.mecspress.org/>),DOI: 10.5815/ijitcs.2016.04.05.