RESEARCHERID
THOMSON REUTERS

[ICRTCET-2018]

ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

OVERLAPPING AND DISJOINT COMMUNITY DETECTION IN SOCIAL NETWORKS- A REVIEW

Manoj Kumar Gottimukkla[*1] & Dr Hari Seetha[2]

[*1]Research Scholar, Computer Science and Engineering, VIT-AP,Amaravati,Andhra Pradesh, India
[2]Professor, Computer Science and Engineering, VIT-AP,Amaravati,Andhra Pradesh, India

ABSTRACT

Of late, social networks have become very popular and have attained the maximum consideration because of its utility in various sectors such as healthcare, management, computer science and the sociology domain. Moreover, identification of communities within the networks was a significant task because, in real-time, weighted networks are more general that has the overlapping communities such that the node could depend on many communities. Moreover, it provides response to various queries based on the person's interaction patterns and their behavior. However, identification of the community is a major concern in the social network system for exploring and understanding the network frame. Moreover, the community denotes individuals getting grouped into sub-groups based on the provided framework which is specific to the social interactions. This study an overall review of the various techniques utilized to identify communities within the network. Moreover, it also analyses the conventional community detection overlapping, non-overlapping algorithms and the existing quality metrics. The major goal of this study is to investigate the varied overlapping and disjoint community detection techniques in the network. Even though there are numerous studies relating to this community detection, there are drawbacks such as larger processing time. Hence, a need for an efficient algorithm is proposed for detecting communities with larger network within a limited process time.

Keywords: - *Community detection, social network, community structure, cluster, Disjoint Community Detection*

I. INTRODUCTION

Social networks are general platforms utilized for interacting with different people located in different locations [1]. Specifically, a group of people interact amongst themselves through various ways which then creates social networks [2], [3]. Moreover, it is based on web service through which personal profiles are created and it is shared with another person through mediums like Facebook, MySpace, Google+ [4], [5]. Therefore, it developed as a significant medium for people to be in communication with each other throughout the world [3], [6], [7]. Social network analysis method is utilized to define the network framework and features by creating the relationships within the entities and also the matches within them [1]. Based on the similarity amongst individuals, a network is divided into communities. Moreover, it delivers mathematical and optical investigation of the relationship of the individual [3]. The network is separated into communities based on the similarities amongst individuals. The community is the nodal group that is identical within itself however non-identical to the remaining network.

Currently, wide researches have been undertaken to describe and identify the community existing in the network. The major goal of this community detection techniques is to detect the cluster of nodes within the network [1]. However, it exposes the internal structure and the specific relationships within the nodes which are not accessed in an easier way through appropriate empirical tests [8].

Additionally, it delivers visions regarding the general features or theories amid persons who separate them from certain other communities [9]. The community identification is identical to the problems related to the network partition where the division of network is based on the n groups of nearly identical size and the edges numbering within the groups are reduced. It is a serious concern [9], and a larger amount of heuristic techniques are established for resolving the problem occurring in the networks. The social network individuals are generally categorized using multiple community memberships [10].The individual normally has certain connections with the groups such as

family, friends and colleagues. Moreover, there are no limitations and hence the individual could be active in many of the social groups depending on his desire. This may also exist in the several other complicated networks like biological networks in which the node may have multiple function tasks. Generally, Conventional community detection algorithms focused specifically on the disjoint communities in which the consumer is dependent on tone group. However, most of the real-time organizations could not be exactly denoted using the disjoint network subset [8].

Within the overlapping community networks, disjoint community identification approach was applied to categorize the overlapped nodes into several communities rather than real-network frame structure. Hence, from the past years, larger consideration was specified to overlapping community detection in which the client could depend on two or more groups. The overlapping community detection problem is fascinating and demanding [6], [11] yet it delivers an improved understanding of the structural topology and the principles of every community. In Kelley et al. [12] and Reid et al. [13], the researcher has demonstrated that the overlap is certainly an important characteristic of the most of the real-time social networks. Based on this motive, the overlapping community detection techniques that detect the group of clusters which are not essentially disjoint and nodes which are depended to two or more clusters could also exist. In this paper, a review is presented with regards to detecting communities in social networks. Following the introduction, the paper presents a brief overview of concepts, outlines the diverse approaches used to detect communities in social network and elucidates the methodology which is then followed by a discussion and conclusion.

II. OVERVIEW

A. Social Network

A social network indicates the graphical demonstration of people, whereas people are denoted as nodes, and the edges within the nodal pair denoted certain types of communication within them [14]. It is the general habit of people to communicate with like-minded persons. Moreover, these selective communications create the classic network

News Channels and Politics

Maddow Michelle framework called the community. Within the communities, individuals communicate regularly with community members and communicate less with people outside of this community [11].

B. Community

The online social networks community structure was essential in assisting scholars for exploring the person's behaviour, particularly the overlapping structure, that could capture the features of the clients depended on certain communities [15]. The simple community structure is shown in figure 1.

ISSN 2348 – 8034
Impact Factor- 5.070

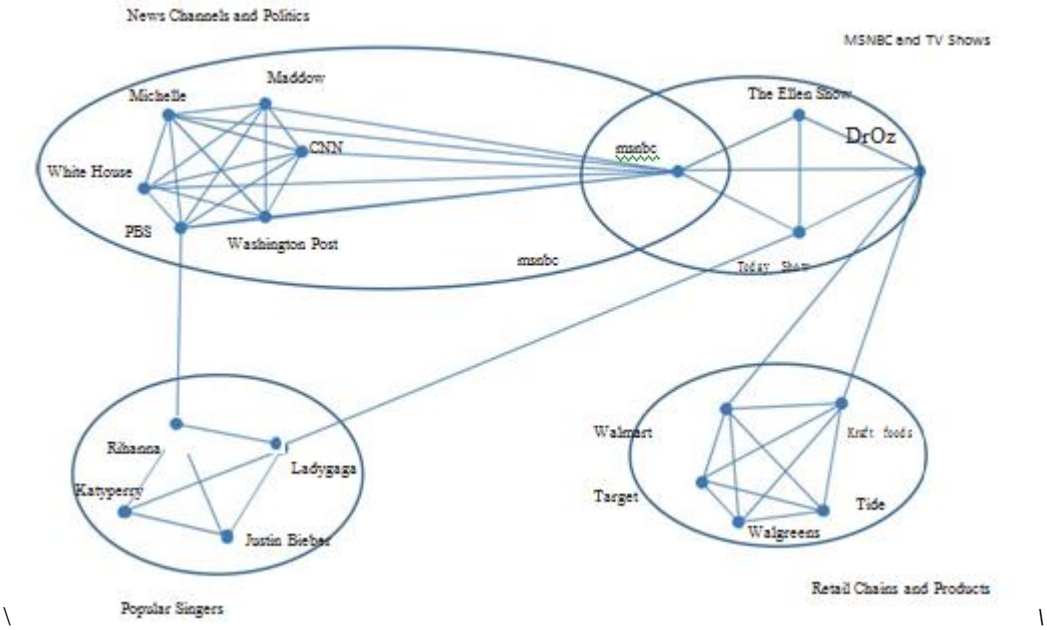

Figure 1. Network with Four Communities.

Source: Adopted from Vinay et al. [1]

C. Community Detection

The term community detection is utilized towards grouping nodes based on the relationships amongst them to form strongly linked subgroup from the whole group.[16]. Yet there is no standard definition, if we predict the network in a graphical format, then the community must be the connected subgraph having added edges in its inner part more than the outer edges of the remaining graph [17], because the community members maintain an inner relationship within it other than people from the outer community. Various data including the network framework, operators profile and their locations data's etc. are utilized in the community detection process.

D. Types of Community Detection

The communities are broadly divided into disjoint community and overlapping community. In the disjoint community, the node will only belong to a single community. There will be no overlapping between the nodes. This type of community is termed as a CRISP assignment for the nodes in which only binary relationship is allowed between any node and community. In an overlapping community, a node may belong to more than one community at the same time. This is sometimes called as the FUZZY assignment of the nodes in which a node may be present in different communities simultaneously [18].

Almost every algorithm creates disjoint communities by allocating each node to a particular community. However, in real time, nodes could spread among many communities in social networks which are identified as overlapping communities. Moreover, community detection algorithm could have the capability to generate disjoint communities whereas actual networks consist of the overlapping community frame structure.

E. Application

The community detection technique was largely utilized in various applications including metabolic network analysis, criminal network mining, genetic and protein interaction network analysis, web search engine, analysis of operator's activity in online, social marketing etc. Moreover, it was also utilized for detecting the financial crime [19], and also the outliers on the basis of the community framework of the network[20].

ISSN 2348 – 8034
Impact Factor- 5.070

III. METHODOLOGY

The data for identifying the Community exists in the socio-networks are gathered from Google Scholar, IEEE transaction, and also from Springer, ACM and Elsevier. Moreover, there is specific time duration for studies that have been reviewed. Only studies conducted between 2002 to 2016 were reviewed. Mostly, this research considered review papers and perused papers from references included in different articles and only studies related to community detection were taken into account. Moreover, 60 appropriate studies were selected that affords the related outcomes related to the community detection. It may also comprise actual researchers who evaluated social networks and excluded the studies which discussed the overlap detection not related to this social network and those that were not in the English language.

IV. RESULTS

A. Approaches to Detect Communities

Generally, more approaches are framed to identify the social network communities. However, such designed approaches utilize various techniques for performing the detection process. Especially, few of them have focused towards genetic algorithms such as [6], [21], [22] and [23] and some of them have focused towards a clustering approach as [24]–[27] and [28]. Many of them have focused towards overlapping community detection such as [2], [29]–[31] and

Hence, few of them have focused towards a fuzzy partitioning approach. The studies related to these are discussed as follows.

B. Studies related to the Genetic Approach

The genetic approach based community detection in the socio-networks is reviewed as follows:

Shang et al. [21] have proposed a community detection technique based on the “Modularity and Improved Genetic Algorithm (MIGA)”. MIGA has considered the modularity as an objective function, that made the simplification of the algorithm and utilized prior information (i.e., the number of community structures), which causes the algorithm to be more objective and increased the accuracy and stability of the community detection. Moreover, MIGA also considers the simulated annealing technique as the local search technique by enhancing the local search capability by regulating the parameters. These simulation outcomes on the on the network obtained from the computer and four physical networks revealed MIGA’s efficiency while comparing with the conventional algorithms.

Su and Havens [22] framed the O(n2) “Genetic Algorithm for Fuzzy Community Detection (GAFCD)” algorithm which identifies the fuzzy communities existing in the network information by exploiting the generalized modularity in a direct manner. GAFCD was the one and only community detection algorithm that is based on the modularity which returns the community in the fuzzy partitions crisp subset if it is suitable. The author stated that the future work should be focused on the additional works which are necessary to allow this GAFCD feasible for the bigger social networks. It is assumed that bigger social networks are generally sparse graphs, which tries to lessen the money and time consumed for computing the Qg for the purpose of the fuzzy partition. Hence, it is necessary to develop more enhanced algorithm as a substitute for the existing mutation operator.

Meena and Devi [6] developed a Genetic Algorithm (GA) which contains two stages to identify overlapping communities by representing the node. Initially, disjoint communities were identified to detect overlapping communities. Moreover, modularity is utilized as the primary optimization function. Experiments were done in actual and artificial networks for validating the efficiency and its scalability. On the whole, non-overlapping community identification delivered enhanced overlapping communities. Examining the belongings of boundary nodes in more community than the single community is the better approach of handling the overlapping communities.

Ebrahimi et al. [33] developed the genetic-based approach to identify overlapping communities by utilizing pareto fronts. This technique involves two major processes such as framing a mathematical model having two objective functions, and the optimization is done using "Non-dominated Sorting Genetic Algorithm II" (NSGA-II) and the next stage includes the community identification by utilizing Pareto optimal solutions in Pareto front. The developed approach extracting the suitable communities from overall Pareto optimal solutions rather than selecting an individual optimal solution and has two major benefits including the scalability and the detecting capability of the overlapping communities. The results obtained from real and synthetic networks stated that the proposed technique outperformed the conventional approach with its overlapping ratio, interwining and overlapping membership's functions and produced stable scalability results.

Francisquini et al. [23] developed a GA to identify communities in socio- networks on the basis of local data to create offspring. Moreover, it utilizes the target ROI existing in solution space. However, this process gets affected with crossover operator. The proposed technique stores the individual locally as labels in the vertices and this creates better flexibility and can be used in various applications like dynamic networks. The proposed technique has been proved experimentally to be faster and to provide better results compared to other examined networks. This technique would be further developed to be used in WSN routers and hence the speed of calculating the fitness would be improved by measuring the modularity locally. To reduce the danger of uncontrolled community growth, a size controlling technique will be used.

Overall, the genetic approach can be useful in community detection Another genetic based algorithm brought in increased flexibility. Hence, the genetic approach is deemed advantageous in community detection.

C. Studies related to clustering approach

Previous studies related to clustering approach for detecting the community are discussed as follows:

Cai et al. [24] framed the clustering approach on the basis of a new random walk with the implementation of the clustering tuning aspects to make the appropriate clustering structure. However, the performance was investigated based on two actual world socio-networks with enhanced structure framework, a network of "Zachary's karate club" and "American college football team" by utilizing the knowledge and statistical based metrics for the performance evaluation of the proposed technique and also the other conventional algorithm. Moreover, the simulated outcomes indicated that this proposed technique outperformed another conventional algorithm which provides the enhanced quality in the clustering technique based on the overlapping proportion stated by the users by utilizing the network topology and also the attained communities' knowledge. Though, the enhancement of the clustering quality based on the application of the fitness detection on clusters should be focused in future [34].The running time would be raised based on the raise in the network size. Moreover, the alteration of this technique needed to match this approach to extensive larger networks which are to be explored in the future. Gu et al. [25] developed the dual algorithm for community detection overlapping on the basis of the Neighborhood Vector Propagation Algorithm (NVPA), which could identify the disjoint communities with an increased accuracy rate. The Link Partition of Overlapping Communities (LPOC) which is the initial algorithm based on the link graph clustering rectifies the link clustering disadvantage of producing many overlapping vertices. Even though it maintains the higher accuracy of the NVPA; it has the high-level complexity of computing. Hence to rectify this issue, the researcher developed another algorithm called "Candidate Overlapping Nodes Screening (CONS)" algorithm that utilizes the NVPA on the node graph to identify the non-overlapping communities, by framing the quality function and the screening technique for detecting the nodes which actually linked multiple communities. Moreover, the researcher evaluated this algorithm on LFR benchmarks and actual world networks, and the comparison is made with conventional techniques. The proposed results stated that it had more accuracy rate and outperformed the other algorithm regarding the modularity, WAC value and partition density. However, the major restraint found in this two algorithm was that the time taken for the processing of NVPA is more and hence the algorithms are comparatively less.

Ferreira and Zhao [26] developed the approach for clustering of the time series through the community detection in the more complex networks. Initially, they framed the network in which each vertex denotes the time series which

are linked with more identical ones. By implementing different time series distance functions, the similarity was calculated.

The researcher applied the community detection algorithm for identifying the vertices groups which are connected in a strong manner for generating the time series clusters. Moreover, they checked the enhanced operation of the particular distance function with each clustering technique and compared it to the proposed technique. Moreover, from the results, it is found that this technique performed better than the conventional clustering techniques in a statistical manner. Moreover, to improvise the performance regarding the selection of neighbour's best number, these work needed to be attentive on additional distance functions, construction techniques, rival approaches and community detection techniques.

A study by Zhou et al., [35] framed the link clustering algorithm on the basis of density to enhance the accuracy rate in detecting overlapping communities. A novel similarity formula is computed in this study to allocate the similarity within the links and clusters including the core edges on the basis of core density reachable model in expansion stage are generated. Additionally, the enhanced approach used for unclassified edges are framed to allocate them to nearest cluster. Moreover, similarity metric to evaluate the similarity between two edges is demonstrated. The results attained indicated that the proposed technique outperformed conventional techniques in terms of community and overlapping nodes detection such that it obtained 15% more NMI value compared with the conventional techniques. Future work has to be done on concentrating the management of weighted and directed networks.

Ferreira and Zhao [27] framed the approach for the process of clustering the time series by utilizing community detection in complicated networks. Initially, they developed the technique for the transformation of the time series group to the network by utilizing the various distance functions, in which every time series is signified using the vertex and more identical series are getting linked. Moreover, community detection algorithms were applied to detect the vertices groups which are connected in a strong manner (known as the community) and accordingly detect the clusters of time series. The results found indicated enhanced performance than conventional techniques. Though, the present study focused on the univariate times series it is necessary to consider solving certain time series problems such as the prediction of the time series.

Zhou [28] developed the AR –cluster which is based on graph clustering for detecting communities existing in complex networks. The framed technique embraces two theories which are the attracting degree and recommending degree for supporting the structural resemblances with the vertices. Additionally, the path that has the extreme recommending degree has been considered for evaluating the pairs that are having the relationship in an indirect way. Even though the obtained results indicated enhanced performance of the proposed system, there is a difficulty in utilizing the larger graphs.

Overall, this section looked into studies done by several authors into approaches related to clustering. One author proposed a clustering algorithm based on a random walk. It was found that the random walk clustering algorithm was more effective as compared to other algorithms. The quality of the clustering is also substantially improved. However, this technique should concentrate on establishing fitness detection clusters to improve the outcome quality of clustering.

Another paper proposed two algorithms viz., the LPOC and CONS algorithm. The said two algorithms were found to have high accuracy levels as compared to other algorithms with regards to modularity, the density of partition and WAC value. Hence, this algorithm has drawback pertaining to a larger processing time .Moreover, several researchers framed time series clustering approaches. However, It was found that the said approach excelled from a statistical standpoint as compared to other conventional algorithms for clustering. Nonetheless, this approach would further be enhanced if it takes into account more distance functions, community detection algorithms, rival methods and network construction methods. An efficient graph clustering algorithm termed AR-Cluster was proposed which adopted two techniques recommending degree and attracting degree. Hence, the findings indicated that this approach was very effective nonetheless, using it in large graphs can be quite complex.

D. Studies Related to Overlapping Community Detection
The previous studies related to community detection are discussed as follows:

Dai et al. [29] developed the "Multi-Label Propagation Algorithm (MLPA)" for identifying overlapping communities within the networks. In MLPA, Propagating Intensity (PI) that associates the vertices nearness and the propagated label data developed for the propagation guidance. Moreover, the results within the real and synthetic networks indicated that MLPA performs better than the various conventional methods, such as the traditional label propagation techniques in the overlapping communities' detection. However, it has the disadvantage that identified communities are based on the algorithm parameters in which the parameter values are difficult to be determined which are to be enhanced in future.

Boudebza et al. [36] has developed a framework to detect overlapping communities in social networking sites called online label clique percolation method (OLCPM).In this technique, community frame structures are updated with the node involved in the event rather than evaluating entire k-cliques for the overall network in every event happened in 530 networks. Hence, within this updation, computing time gets reduced. Moreover, 535 label propagation based on post-process is developed to identify the peripheral nodes. The results attained indicated enhanced performance in terms of both computational time and quality detection. However, this technique has certain drawbacks associated with CPM including the k parameter (clique size) such as it increases the cost while execution for every step. Hence in future, this framed OLCPM could be expanded in case of enhancing a post-process online version.

Xu et al. [30] proposed detection algorithm based on network's dynamic evolution technique to analyze the impact of nodes inclusion or removal function and edges in community detection to determine the group's removal or inclusion in a dynamic network in order to simulate process including splitting, merger, creation, community retrial and finally to detect the community structure containing overlapping characteristics. The validated results demonstrated the algorithm's effective performance. However, this technique is appropriate for the network having the local alteration; this incremental algorithm could not assure higher quality within the division of community if the changes happen in the larger range.

Dhouioui and Akaichi [2] developed a novel algorithm with the aim of overlapping community detection, particularly within social networks. Initially, they found the probable central nodes for extending the communities and improve this partition in the final stage. Subsequently they verified the performance of the OCDA in real-time networks and determined a standard accurate rate. Moreover, this work required to enhance the algorithm for handling the numerous networks that have a larger scale and must satisfy the demand of the nodes which could depend on two or more community having numerous degrees.

Elyasi et al. [31] developed the rapid algorithm for the detection of overlapping community . Initially, Louvain technique was implemented to the provider network. Later, updation of the belonging matrix was done in which every component defines the level of nodes belonging to its community. In the end, some of the identified communities are combined on the basis of the modularity metrics. From the performance analysis, it was found that the proposed technique shows the enhanced performance than conventional algorithms. Though the clusters numbers are known, it is used to enhance the clustering process speed, and accordingly, the final phase of the algorithm was not considered. Therefore, it is necessary to identify a more rapid method for the final phase of the algorithm.

Huang et al. [37] developed a new technique known as "Overlapping Community Detection in a Heterogeneous Social Network (OCDHSN)" to perform the detection process with improved accuracy and reduced power consumption. This proposed technique involves seed selection using vertex set conductance and community initializing process using seed- neighbourhood. The experiment results indicated that the proposed approach performed better than conventional techniques in terms of evaluation metrics such as precision-recall, omega index, modularity, normalised mutual index. It also has the enhanced efficiency that is needed to perform high scale SNSs.

Moreover, the author stated the need to focus on extending the multiple seeds selection approach to attain enhanced effective community detection and to evaluate the performance in actual heterogeneous social networks.

Chakraborty et al. [32] framed two common structure for the detection of ensemble community which is EnDisCo that detects disjoint community frame structure and the MeDOC that identifies the overlapping and disjoint community frame structure. By using this LFR benchmark, with an actual dataset of ground-truth community structure indicated that two approaches operated with maximized accuracy than the traditional CD techniques. However, in future, it is needed to enhance the justification for the supremacy of the ensemble techniques than the discrete prototypes. Moreover, it is also needed to parallelize the ensemble structure.

Overall, the approaches used in this section pertained to (MLPA) for detecting overlapping communities. This approach is known to outperform several other approaches. However, this particular approach is not without flaws. One of the primary shortcomings is that the detected communities largely rely on the algorithm parameter. Moreover, the determination of the parameter's value was complex. Even though several existing techniques are available, there are certain drawbacks within these approaches such as the processing time, quality, speed and accuracy.

E. Studies Related to Fuzzy Partition

Previous studies related to fuzzy partitioning approach are discussed as follows:

Zhang et al. [38] developed clustering approach with fuzzy based community detection by providing relative membership within network vertices by framing into n-dimensional space with spectral mapping. By providing the communities number K, the upper k-1 eigenvectors were evaluated, and the network is again getting mapped in to the Euclidean space with d-dimensions. Fuzzy c-means (FCM) clustering technique was utilized to attain the softer assignment. However, the computational efficiency and the partition accuracy depended on the value K, which was generally complex to predict in future.

Sun [39] presented the identical metrics on the basis of the edge centrality and made the network modeling as the fuzzy relation for clustering the fuzzy-based community. By the implementation of the threshold parameter, the fuzzy assignment could be transformed into a crisp assignment. Moreover, it is the edge centrality which is the global measure and the parameter selection mostly based on the networks frame structure.

Sun [39] recommended the fuzzy transitive rules that expose the community frame structure within complex networks. The suppleness in the selection of the thresholds of this technique leads to the separation of the network into numerous communities that are having multiple resolutions. The experimental results indicate that this technique which is on the basis of Rule I attained improved performance than conventional techniques while it is nearer to 0.8 and hence from these studies, clear knowledge about the partitioning of the network and community detection is obtained. Additionally, the author stated that it is necessary to concentrate on distortion problem existing in fuzzy clustering in community detection process.

Al-Ayyoub et al. [40] developed the study to focus on the community detection algorithm to enhance the operational performance of fuzzy C-mean and K-mean approach by using the Graphics Processing Units. Various parallel applications are considered including the Hybrid, Dynamic Parallel and Hybrid Nested Parallel. The results obtained by Fuzzy C-Mean -based algorithm indicated that the parallel implementations attained the better speedups of about 4.45X (dynamic Parallel), 8.3X (Hybrid) and 12.58X (Hybrid Nested Parallel) implementations. At the same time for the K-Mean based algorithm attained the speedups of about 5.33X (dynamic Parallel), 1.88X (Hybrid) and 8.37 X (Hybrid Nested Parallel) implementations which indicated that the greater speedups are attained for lesser dimensional quantity. The developed hybrid nested parallel implementation outperforms the conventional techniques. Hence, the author stated that the effects of dimensions decrease with the raising the abilities of Graphics Processing Units cards.

However certain fuzzy community detection techniques were presented in current years, most of the studies need the basic understanding regarding the community frame structure including the quantity of the communities etc [41]–[43] and the necessity to adjust the parameters such as the probability threshold [44], [45] that could reduce its operational performance within complex networks.

Overall, this section outlined approaches that used fuzzy partitions wherein several reviewers presented diverse approaches such as the fuzzy clustering method, fuzzy transitive rules, fuzzy community detection methods and similarity measures based on edge centrality. While all of the said approaches can be beneficial, it is implied that these approaches need to be explored further to fine tune the approaches in order to achieve enhanced results.

V. DISCUSSION

The community detection process was ae type of exploration that mostly concentrated on the socio-network, partitioned the network nodes into the various communities, and classical non-overlapping community detection techniques are analyzing of ideological spectrum techniques, clustering coefficient method, GN algorithm, hierarchical clustering algorithm, Newman greedy algorithm, Random walk algorithm, Potts model algorithm etc. [30]. Nauyen and Dinh presented the efficient algorithms to detect the non-overlapping network community frame structure such as identifying the pairwise disjoint communities, in static [46], and dynamic networks [46], [47].

Additionally, overlapping community detection technique offers an alternate for denoting the variety and changeability in the client's attention. Previous works regarding community detection anticipated that the communities are either disjoint or non-overlapping. Hence to improve the identification speed and the quality, Newman developed the Fast Newman algorithm [48] in which the modularity functions are embraced for the overall enhancement which is being largely utilized for the process of community detection. Moreover, Palla et al. [49] introduced the “Clique Percolation Method (CPM)” which is the first overlapping community detection technique that provided the enhanced performance in identifying the overlapping structure within the networks, Similarly, Lancichinetti and Fortunato [50] demonstrated that Modularity function could not be appropriate to the overlapping community context in a better manner due to the problem existing with the resolution limitation and in greater degradation in the overlapping cases. Overlapping community identification technique on the basis of the local optimization was currently developed. Moreover , in the study by Lancichinetti et al. [51] developed the Local Fitness Maximization (LFM) technique which made the benefits of overlapping and the community hierarchies. De Meo et al. [52] developed a work-related with pre-processing stage among which links gets weighed based on centrality with an objective to improvise the conventional community detection process.

Moreover, Lee et al. [53] established pervasively overlapping communities in numerous networks, that consist of larger amount of exterior links than that of the interior links that interrupt the general network community properties. They described that most of the nodes depended on multi-communities hence in future, an extended LFM algorithm was presented to manage the issues with pervasively overlapping community [53], [54].

Mostly, larger techniques are presented for identifying communities such as the linkage clustering [55], betweenness-based techniques [14], [56],local approaches [51], [57], [58] etc. Typically this approach does not allow overlapping communities. However certain networks are implementing the statistical statement of overlapping and nested communities. On the basis of this concept, Palla et al. [49]presented clique percolation for investigating major statistic characteristics of interwoven communities set and initiating the coverage of the overlapping modular framework of the complex structures [49], [59]. The most probable wider approach utilized in the benefit function (Q) maximization in the communities and the predicted quantity of certain edges based on the random format. However, the modularity maximization was the NP-complete complexity problem [60].Hence, in general ,certain applications of the modularity technique are generally utilized including the greedy algorithm [48], [61] simulated annealing algorithm [62], [63], spectral techniques [64],each having varied accuracy and complexity [65].

Conventional community detection generally depends on the global objective function optimization [16]. Certain research attainments including the GN Algorithm [14], [66], Fast Newman (FN) Algorithm [67], Radicchi fast split

Algorithm [68], Extreme-value optimization Approach [69], Simulated Annealing based GA Algorithm [70] and more modularity based optimization techniques that converted community detection problem into optimization problem and made an attempt to detect an optimal solution for the specific objective function which should also be enhanced [71], [72].

Though the existing schemes are focused mainly on non-overlapping communities in various real networks whereas a non-overlapping community does not have any connection within every community and nodes exist in the same community. Palla established overlapping communities with global existence in several actual socio-networks and in the genetic networks [49].Moreover, on the basis of Palla's clique percolation concept, certain overlapping community identification techniques were developed which are Clique Graph [73], EAGLE algorithm [74],CPM algorithm [49],

The next technique uses fuzzy clustering based technique

It increases the closeness between all the available communities and nodes and finds whether it has a beneficial adhesion to a group depending upon its degree. The most well-known algorithm in this class is "Fuzzy C-Means (FCM)" which mitigates the objective function thereby reducing the intra-cluster variance [75]. FCM only considers the interval between the nodes and hence cannot plot the structure [76] has modelled the detection of overlapping community nonlinearly in a constrained optimization problem that may be resolved by methods like simulated annealing. NMF [45] is a model which is based on Bayesian non-negative matrix factorization which has also been used in OSBM [77], [78], etc. The mentioned fuzzy techniques calculate communities which are in a spherical shapes mainly due to the constraints imposed on the membership degrees [45], [75], [77]. This is a huge disadvantage since, communities are of arbitrary shapes in real networks. Table 1 discusses the summary of previous studies and table 2 has discussed the pros and cons of community detection approach

Table I. summary of previous studies

Author	**Algorithm / Technique**	**Results**	**Advantage**	**Limitation/Future scope**
Xu et al. [30]	Community detection algorithm	Suitable for a network with local change	High efficiency	Not applicable to the high quality of community division
Elyasi et.al [31]	Louvain method	The measure of the modularity is used to mergre the communities	High Accuracy Obtained the obtain the maximum value of modularity	Needs to improve the clustering speed
Su and Havens [22]	Genetic algorithm	Effectively explored the search space for all possible community numbers	Better performance in finding the best partition of a network in terms of modularity Finds the best partition in a network	A better algorithm must be replaced with.

			effectively on the basis of modularity	
Gu et al. [25]	Link Partition of Overlapping Communities, and neighbourhood vector propagation algorithm	Detected communities more efficientlyy as compared to other techniques	High accuracy. Better partition density WAC value and modularity	Processing task time is slow
Dai et al. [29]	Multi-label propagation algorithm	Effectively improved the detection technique of overlapping communities	Better than other detection techniques.	The determination of the parameter value complicated issue
Cai et al. [24]	Random walk based clustering algorithm	The structure of the community and quality of clustering has been has been studied.	Clustering quality has been improved to a great extent by using network topologies	Fitness detection can be applied to boost the performance of clustering results
Ferreira and Zhao [27]	Time series clustering approach	The proposed technique achives better result than existing techniques.	Highly efficient in detecting patterns of the shape	Considered only for univariate time series
Francisquini et al. [23]	Genetic-based algorithm	Exploited the targetted areas of interest in the solution space	Obtained good results in the tested network.	To solve the routing issues in WSN, a distributed version of this technique can be built.
Shang et al.	Modularity and an	The ability of the local search has been	High efficiency	Performance needs to be

[21]	improved genetic algorithm	improved varying the parameters.		improved
Shang et al. [21]	Genetic algorithm	Evolution of network topology is captured	High efficiency	Needs to focus robustness issue
Zhou et al. [28]	Graph based AR cluster approach	Adapted similarity measure to calculate node similarities	Distance value is less when they are nearby Clustering quality is high	It is not easy to use the data to plot large graphs.
Ferreira and Zhao [26]	Time series clustering approach	Calculated different time series distance functions Also produced the time series clusters	Found accurate distance function Better to detect groups of series that present similar patterns	Need to focus towards enhance the performance when choosing the best number of neighbours
Sun [39]	Fuzzy transitive rules	The data has been transmitted between two nodes until it attains stability.	Nodes which are present in the same group give a better result compared to those of different group.	The distortions that occur in clustering of fuzzy would be addressed.

Table II. summary of overview

CDM Algorithms	Advantages	Disadvantages
Edge Content in Social Media Networks.	It supervises the community detection better	This approach is useful for email and flicker images based clustering only.
weighted network	The total weight of the clusters in the network and the	The proposed technique does not work well for
	affinity between the different clusters has been	some types of network.
	calculated.	
Bayesian network and Expectation Maximization	It performs well in all types of network	The quantity of the communities must be known
technique.		in the first place.
Overlapping communities.	It identifies the communities that are overlapping and	Only modularity method has been used to to

	analyzes the structure of the community	identify the communities that are overlapping.

From the analysis, drawbacks of conventional techniques have been identified such as they are cost –expensive, and most of them are based on the network –topology [16].Moreover, the clustering efficiency relies on the quantity of required computational stages and also on the memory units quantity which is required to be assigned consecutively for executing the computation process [79].Therefore , the larger amount of socio-network's data has influence over the efficiency range. However, conventional techniques could not handle this conditions such as the creation of various community structures within various socio-network by a certain group of people [15].Hence, most of the conventional techniques could not manage this larger dimensionality of data within the socio-networks.

VI. CONCLUSION

In this present study, analysis of overlapping community detection process in social networks are done by reviewing conventional overlapping community detection algorithms and significance of algorithm implementations in these community detection techniques are examined. Moreover, the concepts of social networks, community detection and its types and applications are presented and also community detection approaches such as genetic, fuzzy based overlapping community detection algorithms are reviewed.

Hence from this review, it is known that conventional techniques are easier to implement, but it is complex to determine the clustering size for community detection process. Moreover, with overlapping detection techniques, the drawback is that they are inappropriate for the networks with higher uncertainty. However, local approaches are effective in detection process even with reduced complexity yet there are certain flaws with the techniques such that they could identify individual community alone and also clustering size should be known beforehand.

Moreover, with the advancement in social networking technologies, the researchers in this social networking field have to perform with large volume of data. Therefore, an efficient algorithm which could operate on huge data within limited time should be proposed to attain maximized accuracy and speed rate in the community detection process. Also, evaluation metrics are to be considered to analyze the performance of the proposed algorithm such as community coverage, modularity, conductance and transitivity.

Hence within this performance evaluation, community coverage metric used to calculate the user's quantity allocated to appropriate communities, modularity to detect the integrity of overlapping communities, conductance metric to concentrate on both the exterior and interior community connections, and transitivity to measure the link among the two individuals. Therefore the proposed algorithm has the advantages that it performs overlapping community detection with enhanced accuracy rate and reduced overhead by concentrating on both social and dynamic networks.

In addition, it is established that conventional algorithms are mostly focused on the weighted, un-weighted, directed and undirected networks. Conversely, in case of heterogeneous and in bipartite networks, community detection process attained higher accuracy even with complex data but there are only a few works on community detection based on this heterogeneous network .Therefore, the proposed algorithm should focus on bipartite and heterogeneous networks in future.

REFERENCES

1. *Vinay, Sumit, and J. Parkash, "A Review paper for Detection of Overlapping Communities in Complex Networks," Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 3, pp. 3339–3341, 2014.*
2. *Z. Dhouioui and J. Akaichi, "Overlapping community detection in social networks," in 2013 IEEE International Conference on Bioinformatics and Biomedicine, 2013, pp. 17–23.*
3. *Richa, "A Review on Overlapping Community detection in Social Network," Int. Res. J. Eng. Technol., vol. 3, no. 7, pp. 2947–2957, 2016.*

4. *danah m. Boyd and N. B. Ellison, "Social Network Sites: Definition, History, and Scholarship," J. Comput. Commun., vol. 13, no. 1, pp. 210–230, Oct. 2007.*
5. *M. Gjoka, M. Sirivianos, A. Markopoulou, and X. Yang, "Poking facebook: characterization of osn applications," in Proceedings of the first workshop on Online social networks - WOSP '08, 2008, p. 31.*
6. *J. Meena and V. S. Devi, "Overlapping Community Detection in Social Network Using Disjoint Community Detection," in 2015 IEEE Symposium Series on Computational Intelligence, 2015, pp. 764–771.*
7. *Y. Gu, B. Zhang, G. Zou, M. Huang, and K. Jiang, "Overlapping Community Detection in social network based on Microblog User Model," in 2014 International Conference on Data Science and Advanced Analytics (DSAA), 2014, pp. 333–339.*
8. *G. T. Prabavathi and V. Thiagarasu, "A Review on Overlapping Community Detection Algorithms," Int. J. Appl. Res. Stud., vol. 2, no. 3, p. 347, 2013.*
9. *U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," Phys. Rev. E, vol. 76, no. 3, p. 36106, Sep. 2007.*
10. *J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks," ACM Comput. Surv., vol. 45, no. 4, pp. 1– 35, Aug. 2013.*
11. *A. K. Singh and S. Gambhir, "Greedy local algorithm for overlapping community detection in online social networks," in 2014 5th International Conference - Confluence The Next Generation Information Technology Summit (Confluence), 2014, pp. 155–162.*
12. *S. Kelley, M. Goldberg, M. Magdon-Ismail, K. Mertsalov, and A. Wallace, "Defining and discovering communities in social networks," in Handbook of Optimization in Complex Networks, Springer, Berlin: Springer, 2011, pp. 139–168.*
13. *F. Reid, A. McDaid, and N. Hurley, "Partitioning Breaks Communities," in Proceedings of the International Conference on Advances in Social Networks Analysis and Mining, 2013, pp. 79– 105.*
14. *M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," Proc. Natl. Acad. Sci., vol. 99, no. 12, pp. 7821–7826, Jun. 2002.*
15. *W. Fan, K.-H. Yeung, and W. Fan, "Overlapping community structure detection in multi-online social networks," in 2015 18th International Conference on Intelligence in Next Generation Networks, 2015, pp. 239–234.*
16. *V. . Sumithra and S. Surendran, "A computational geometric approach for overlapping community (cover) detection in social network," in 2015 International Conference on Computing and Network Communications (CoCoNet), 2015, pp. 98–105.*
17. *T. N. D. T. M. Nguyen and N. . Dinh, "Overlapping community structures and their detection on social networks," in IEEE Third International Conference on Social Computing (SocialCom) and Privacy, Security, Risk and Trust (PASSAT), 2011.*
18. *W. Zhi-Xiao, L. Ze-chao, D. Xiao-fang, and T. Jin-hui, "Overlapping community detection based on node location analysis," Knowledge-Based Syst., vol. 105, pp. 225–235, Aug. 2016.*
19. *L. Tang, G. Barbier, H. Liu, and J. Zhang, "A social network analysis approach to detecting suspicious online financial activities," in Advances in Social Computing, ser. Lecture Notes in Computer Science, S.-K. Chai, J. Salerno, and P. Mabr, Eds. Berlin, Heidelberg: Springer Berlin / Heidelberg, 2010, pp. 390–397.*
20. *J. Gao, F. Liang, W. Fan, C. Wang, Y. Sun, and J. Han, "On community outliers and their efficient detection in information networks," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10, 2010, p. 813.*
21. *R. Shang, J. Bai, L. Jiao, and C. Jin, "Community detection based on modularity and an improved genetic algorithm," Phys. A Stat. Mech. its Appl., vol. 392, no. 5, pp. 1215–1231, Mar. 2013.*
22. *J. Su and T. C. Havens, "Fuzzy community detection in social networks using a genetic algortihm," in 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2014, pp. 2039–2046.*
23. *R. Francisquini, V. Rosset, and M. C. V. Nascimento, "GA-LP: A genetic algorithm based on Label Propagation to detect communities in directed networks," Expert Syst. Appl., vol. 74, pp. 127–138, May 2017.*

24. *B. Cai, H. Wang, H. Zheng, and H. Wang, "An improved random walk based clustering algorithm for community detection in complex networks," in 2011 IEEE International Conference on Systems, Man, and Cybernetics, 2011, pp. 2162–2167.*
25. *K. Gu, J. Tang, L. Pan, and J. Li, "Overlapping Community Detection Using NVPA," in 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), 2015, pp. 197– 202.*
26. *L. N. Ferreira and L. Zhao, "A Time Series Clustering Technique based on Community Detection in Networks," Procedia Comput. Sci., vol. 53, pp. 183–190, 2015.*
27. *L. N. Ferreira and L. Zhao, "Time series clustering via community detection in networks," Inf. Sci. (Ny)., vol. 326, pp. 227–242, Jan. 2016.*
28. *H. Zhou, J. Li, J. Li, F. Zhang, and Y. Cui, "A graph clustering method for community detection in complex networks," Phys. A Stat. Mech. its Appl., vol. 469, pp. 551–562, Mar. 2017.*
29. *Q. Dai, M. Guo, Y. Liu, X. Liu, and L. Chen, "MLPA: Detecting overlapping communities by multi-label propagation approach," in 2013 IEEE Congress on Evolutionary Computation, 2013, pp. 681– 688.*
30. *B. Xu, L. Deng, Y. Jia, B. Zhou, and Y. Han, "Overlapping Community Detection on Dynamic Social Network," in 2013 Sixth International Symposium on Computational Intelligence and Design, 2013, pp. 321–326.*
31. *M. Elyasi, M. Meybodi, A. Rezvanian, and M. A. Haeri, "A fast algorithm for overlapping community detection," in 2016 Eighth International Conference on Information and Knowledge Technology (IKT), 2016, pp. 221–226.*
32. *S. Chakraborty et al., "Programmed Molecular Engineering: Stepwise, Multicomponent Assembly of a Dimetallic Metallotriangulane," European J. Org. Chem., vol. 2016, no. 30, pp. 5091–5095, Oct. 2016.*
33. *M. Ebrahimi, M. R. Shahmoradi, Z. Heshmati, and M. Salehi, "A novel method for overlapping community detection using Multi-objective optimization," Phys. A Stat. Mech. its Appl., vol. 505, pp. 825–835, Sep. 2018.*
34. *D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-*