Multidisciplinary Review Journal

Journal of Critical Reviews

ISSN- 2394-5125 Vol 7, Issue 4, 2020

Review Article

AN OPTIMIZED OVERLAPPING AND DISJOINT COMMUNITY DETECTION TECHNIQUES USING IMPROVED COMMUNITY OVERLAP PROPAGATION ALGORITHM IN COMPLEX NETWORKS

C.S. Saradha[1], Dr.P. Arul[2]

[1]Research Scholar, Bharathiar University, Coimbatore. sara.kumarasamy@gmail.com
[2]Assistant Professor, Department of Computer Science, Govt Arts College, Trichy. phdarul2004@yahoo.co.in

Received: 14.12.2019 *Revised: 17.01.2020* *Accepted: 19.02.2020*

Abstract
The Internet and Social network has become the essential part of life which ease people for sharing information and media with friends. The formation of community structure is particularly important in network analysis practically. The community structure is one in which it merges the alike users by interacting with friends in social networks. Apart from this recognition of online communities users in social networks, there are various applications such as identifying a group of expert customers, a group of customers with shared endeavors, and a alike people's group for marketing objectives. The structural Properties of social networks are analyzed with the help of Community detection algorithms. Enhanced Nearest Neighbor-Based Clustering (ENNC) is best among the community detection algorithms in which modularity based community detection is achieved in the prevailing technique. Although there exist some hindrances of overlapping communities due to the probability of user joining in more than a group is high. Hence there requires a detecting overlapping communities while analyzing realistic network. This research concentrates on Optimized Overlapping and disjoint Community Detection (OODCD) technique by adopting a by Improved Vertex Imitation Co-efficient based Community Overlap Propagation Algorithm (IVIC-COPRA).In this, the Enhanced Nearest Neighbor-Based Clustering (ENNC) approach is utilized basically which is modularity based approach for partitioning the network into minor local communities. The Louvain method is then used for detecting the disjoint community in the given network. The belonging matrix plays a vital in this research which is regularly updated whose matrix elemental value decides the role of node belonging to a community and thus the overlapping communities is found with the support of Improved Vertex Imitation Co-efficient Community Overlap Propagation Algorithm (IVIC-COPRA).The Performance of the anticipated research is examined and contrasted with the numerous eminent other overlapping community detection algorithms by means of simulation.

Keywords: complex networks, Community structure, Enhanced Nearest Neighbor-Based Clustering, modified modularity, disjoint community, Louvain method, overlapping communities, Improved Vertex Imitation Co-efficient based Community Overlap Propagation Algorithm (IVIC-COPRA).

DOI: http://dx.doi.org/10.31838/jcr.07.04.146

INTRODUCTION

The various real-world complex systems such as social networks, scientist's cooperation networks, Web networks, protein interaction networks, etc. [1] show a significant part in defining complex systems. The nodes in entity represent by each node and the associations amid the entities are defined by links in case of complex networks. There are many communities (modules or clusters) in many real networks which are revealed by extensive researches which shows the nearby interconnections and scarce associations which is the peculiar property of significant complex networks topology structure [2]. The theoretical implication plays a vital role in the complex network community structure. The arrangement and utility of the network are the key aspect in finding the hidden laws and predicting their behavior which is achieved by means of detection of community structure. The most challenging research objective lies in the community detection.

Community Detection is demarcated as the method of recognizing the interrelated sets or clusters in the network. The interrelated communities are compactly connected to each other when compared to different communities [3] group which supports in identifying the node which are linked by common properties and thereby assessing relationship between them, Considering a Facebook example where the network is constructed by representing the user and exemplifying the associations amid the users and the community. Let us consider same football club or support the same presidential nominee in which individual users is assessed by identifying such groups, the interaction between these users and forecast the missing data [4]. Various researches are being carried out in finding the information in the real-world networks and detecting communities.

The data quantity to be stored and retrieved back resourcefully is the main significant difficult hitches in investigating these networks and it forms the important function of Social network analysis [5]. The community structure is one which is formed by the network vertices which is fragmented into either disjoint or overlapping vertices sets with the constraint that the number of edges within a group go beyond the number of edges between any two groups by certain evenhanded extent. This sort of network with community structure is said to possess hierarchical community structure. Though there are many worthy storing community structures, there requires cost-effective and productive methods for data retrieval [6].

The application for each algorithm is always restricted due to the disastrous irreconcilable difference between "disjoint" and "overlapping" Community Detection algorithms. The overlapping communities cannot use "disjoint" algorithm for a network and similarly "disjoint" algorithm [7] is used for disjoint communities. The appropriate algorithm in its proper use gives better efficiency. We can select many “disjoint” algorithms [8] for identifying overlapping communities which can also be used by few “overlapping” algorithms. Additionally, improved “disjoint” algorithms can be exploited for problem detection of overlapping communities which is the main focus of future research.

Community is not defined anywhere basically and there does not exist standard algorithm for ascertaining communities. Though there are plenty of algorithm generated for network community and each algorithm varies in effectiveness and speed [9].It is assumed that the networks are uni partite, does not possess direction and un weighted edges and understanding the communities relation in a network depicts the dissimilarity amid algorithms. In many algorithms, it is anticipated that the vertices are flat set of disjoint communities members [10].In some case, overlapping is possible if each individual possibly appears in multiple community.

The Concept of Optimized Overlapping and disjoint Community Detection (OODCD) technique by Improved Community Overlap Propagation Algorithm (ICOPRA) is introduced in this work for boosting the community detection algorithm efficacy. In this, the Enhanced Nearest Neighbor-Based Clustering (ENNC) approach is utilized basically which is modularity based approach for partitioning the network into minor local communities. The Louvain method is then used for detecting the disjoint community in the given network. The belonging matrix plays a vital in this research which is regularly updated where each matrix element .decides the role of node belonging to a community and thus the overlapping communities is found with the support of Improved Vertex Imitation Co-efficient Community Overlap Propagation Algorithm (IVIC-COPRA).

This research paper is systematized in the subsequent manner; section 1 designates the social networks basic models and communities. In section 2, the review of various methods for community detection is deliberated. In section 3 explains the proposed community detection algorithm trailed by standard datasets list utilized for investigation in social networks. In section 4 explains the benefits of exhausting a proposed system with respect to other community structure by means of simulation result. The section 5 ends up with conclusion and future work.

LITERATURE SURVEY

Chakraborty et al [11] projected a plan for identifying overlapping community structure in a two-stage framework by a technique namely PVOC (Permanence based Vertex-replication algorithm for Overlapping Community detection). This paper concentrates on non-overlapping community structure with the support of standard disjoint community detection algorithm which is said to possess greater resemblance by way of its definite overlapping community structure with the exception of overlapping part. The author used a novel post-processing technique in which prevailing disjoint community detection algorithm is merged with new vertex-based metric termed as permanence which supports in determining overlapping candidates with their community memberships. The PVOC technique provide better experimental results compared to the overlapping community detection algorithms with respect to high similarity output due to which it provides significant overlapping communities from the network.

Gregory et al [12] suggested a innovative two-phase detecting overlapping communities in which the vertices are divided by split concept and disjoint community detection algorithm is utilized for transformed network. Thereby achieving the transformation between disjoint community detection algorithm and overlapping community detection algorithm. It is demonstrated that experimental results shows that utilizing various “disjoint” algorithms offers better execution times than formed by focused “overlapping” algorithms.

Sheikholeslami et al [13] anticipated a robust community detection technique for sparse tensor-based representation which unveils more affluent structure in contrast with its matrix counterpart. The decomposition method is utilized in a constrained tensor approximation framework for constructing network structure. The alternating direction method of multipliers (ADMM) is also involved here for handling arising constrained trilinear optimization via alternating minimization for ensuring the convergence. The soft community memberships help in overlapping community assignments. The time-varying graphs in which edge set in addition to the primary communities are compared with interval which is assessed by tests on benchmark synthetic graphs in addition to the real-world networks.

Yang et al [14] suggested a novel method namely BIGCLAM (Cluster Affiliation Model for Big Networks) for overlapping community detection which supports in scaling huge networks of masses of nodes and edges. The concept of overlapping between communities is connected in a dense manner which is inverse to the other present community detection methods in which the connections of communities are sparsely done. Model-based community detection algorithm is deployed for compactly populated communities, systematically grouped as well as non overlapping communities massive networks. Experimental results are obtained for huge social, collaboration and information networks with ground-truth community information which exhibits improved performance in terms of the detected communities quality as well as rapidity and scalability of the algorithm.

Lee et al [15] proposed a concept called Greedy Clique Expansion (GCE) for community assignment issue. In this research, seeds are nothing but the distinct cliques and local fitness function is optimized by means of expansion of the seeds. The GCE’s performance is measured by exploiting wide-ranging benchmarks on synthetic data which is more robust with respect to diverse graph topologies. The synthetic graphs are well evaluated proficiently by means of Greedy Clique Expansion (GCE) which is well suited for multiple communities. GCE is an efficient method for recognizing protein interaction data functional modules and Facebook friendship data college dorm tasks etc.

Gregory et al [16] demonstrated a new idea for propagating between neighbouring vertices created upon label propagation technique of vertices. In this work, community members attain consent on their community membership by adopting label propagation technique of vertices in which labels are propagated amid neighbouring vertices. The extension of the label and propagation step is to take account of multiple communities information among which all vertex consist of v communities, where v represents parameter of the algorithm. The algorithm can manage weighted and bipartite networks. The algorithm is tested on real networks and independently considered set of benchmarks which offers high efficiency in recovering overlapping communities. This technique is rapid and processing of very large and dense networks can be accomplished in a diminutive period.

Ball et al [17] suggested new statistical methodology for generative network models for the purpose of overlapping communities. The execution time is minimized by means of fast, closed form expectation-maximization algorithm which investigates many nodes in network within stipulated running times. This new methodology delivers best results when executed on real-world networks and on synthetic benchmarks in contrast with existing method. The fastest running time is the biggest advantage achieved for extracting non overlapping community divisions via a relaxation method.

Gopalan et al [18] addressed a new scalable methodology for overlapping communities for resolving the community detection in case of massive real world networks. Bayesian model of networks is the basis for the nodes to take part in multiple communities and the updation of the communities estimate is accomplished subsequently by sub sampling from the network. The exact community structure is ascertained with the support of the algorithm precisely by demonstrating on large simulated networks. By which, massive networks are analyzed using sophisticated statistical models.

Evans et al [19] projected a strategy for community detection which utilizes line graph with weighted version for analyzing degree heterogeneity. The community structure is revealed by network link partitioning and it permits communities nodes overlapping by which nodes may perhaps be present in multiple community and thus nodes partitioning along the line graph of the original network. This in turn output the link partitioning.

Psorakis et al [20] utilized Bayesian nonnegative matrix factorization (NMF) model in community detection for abstraction of overlapping modules from a network. This method addresses variety of benchmarks problems and contrasted with various community detection algorithms. This novel approach benefits the soft-partitioning elucidations, task of node participation scores to modules and an instinctive foundation.

Zhang et al [21] suggested a method for overlapping community detection by means of symmetric binary matrix factorization model (SBMF).Every node is assigned a clear community membership and also differentiates the overlapping nodes. Additionally, the community structure quality is assessed by suggesting a modified partition density and governing the most appropriate number of communities. The work is carried out in synthetic benchmarks and real world networks which is used to demonstrate proposed method efficacy.

Nepusz et al [22] analyzes community detection by finding out optimal membership degrees with the support of fuzzy community structure algorithm. In this three vertices namely outlier vertices, bridge vertices and regular vertices are analyzed. The outlier vertices defines the vertices that is not considered in none of the communities, bridge vertices defines the vertices that are appropriate in multiple communities and consistent vertices are those necessarily confine their own community interactions. The uncertainty is also predicted in the dataset with the aid of fuzzified variant of the modularity function. The Fuzzy community structure of different real world networks is identified in this research nevertheless not narrowed to social networks, scientific collaboration networks and cortical networks with high confidence.

PROPOSED METHODOLOGY

Local network community detection is the key aspect of the suggested methodology. Here, in this research Optimized Overlapping and disjoint Community Detection (OODCD) are combined with the Improved Community Overlap Propagation Algorithm (ICOPRA). There are three basic steps involved in OODCD (1) Spotting local communities, (2) finding the disjoint communities and (3) Combining overlapping communities. In OODCD , social network is fed as input thereby executing the above three steps sequentially for ascertaining the network overlapping community structure. In this, the Enhanced Nearest Neighbor-Based Clustering (ENNC) approach is utilized basically which is modularity based approach for partitioning the network into minor local communities. The Louvain method is then used for detecting the disjoint community in the given network. The belonging matrix plays a vital in this research which is regularly updated where every matrix element decides the role of node belonging to a community and thus the overlapping communities is found with the support of Improved Vertex Imitation Co-efficient Community Overlap Propagation Algorithm (IVIC-COPRA).

Representation and Initialization

The representation of graph is done by means of a graph and the term community refers to a cluster of nodes (a sub graph) densely interconnected amid each other and when lying outside the community, it will be sparsely connected [23]. Community refers to different meaning both literally and technically, there is no proper definition with the concept of connection density. Each cluster corresponds to a node in a non-overlapping scenario. But here, the research concentrates on overlapping scenario where a node is shared by many communities or clusters.

Consider the adjacency matrix be A with size is N×N for a network with N nodes. The value aij of A at position (i, j) is 1 if there exist an edge from i to j and 0 otherwise.

The definition of degree Di of a node i is as follows:

$$D_i = \sum_{j=1}^{N} a_{ij} \tag{1}$$

Considering a sub graph C; there exist two contributions by the total degree of a node i belonging to C. First, the internal degree $D_i^{in}(c)$is defined as the number of edgesconnecting i to other nodes in C:

$$D_i^{in}(c) = \sum_{j \in C} a_{ij} \tag{2}$$

Second, the external degree $D_i^{out}(c)$ of a node i ∈ C is defined as the number of edges connecting i to all nodes outside community C:

$$D_i^{out}(c) = \sum_{j \notin C} a_{ij} \tag{3}$$

Considered sub graph C to be a community- Strongly, if each node has additional connections inside the community compared with remaining network given by:

$$D_i^{in}(c) > D_i^{out}(c), \forall_i \in C \tag{4}$$

Sub graph C is a community in the weak sense if the sum of all degrees within C is higher than the sum of all degrees towards other communities, that is:

$$\sum_{j \notin C} D_i^{in}(c) > \sum_{j \notin C} D_i^{out}(c) \tag{5}$$

A community basically refers to sub graph recognized by the maximization of the subsequent fitness:

$$f_c = \frac{D^{in}(c)}{[D^{in}(c)+D^{out}(c)]^{\alpha}} \tag{6}$$

Where:

- $D_i^{in}(c)$ represents the total internal degree of nodes in community C; and almost twice the number of internal links of that community.
- $D_i^{out}(c)$ represents the total external degree of nodes in community C; and can be computed by the number of links linking module each member with the remaining network.
- α is a positive real-valued parameter, which regulates the community size. The parameter α tunes the resolution of the method. The scale setting of focus in the network refers to α. Large values of α yield very small communities and small values provide large modules. If α is small enough, all the nodes end up in the same cluster, the network itself. In most cases, for α< 0:5, there is only one

community; for α> 2; one recovers the smallest communities. A natural excellent is α = 1; as it is the ratio of the external degree to the total degree of the community which refers to weakcommunity. In most cases, for α = 1 is relevant, whichcontributesvaluable information about the actual networkcommunity structure.

Enhanced Nearest Neighbor-based Clustering (ENNC) Algorithm

Enhanced Nearest Neighbor-Based Clustering (ENNC) Algorithm is utilized to assess the similarity parameter which is nothing but the closeness amid a pair of nodes. The different performance for detecting complex networks community structure [24] is achieved by node metrics based upon the local information. Let us consider set of data points $x_1, \dots\dots, x_n$ and some notion of similarity $s_{ij} \geq 0$ amongentire pairs of data points x_i andx_j, the objective of grouping are several groups of data point fragmenting. This indicates that the data points are alike if they are in same group and data points are dissimilar if they are in different group. The best approach for representation of data is nothing but the similarity graph $G = (V, E)$, when no more similarities happenbetween data points. Here v_i notates the vertex in the graph denoting data point x_i. Whenever there exist similarity s_{ij}among the conforming data points x_i and x is positive amid two vertices, in other words if the value exceeds the threshold , the weighted edge represented by s_{ij} and said to be connected. Similarity graph is a key conern in the clustering issue which can also be reformulated.The reformulation is done in a manner such that graph partitioning is achieved consequently different groups edges possess low weights and the edges within a group possess high weights.

Next lies the modeling of local neighborhood which lies between the data points, thereby creating similarity graphs. Gaussian kernel function plays a significant role in that and the structure process is established on the Gaussian kernel model and given in Equation (7).

$$S_{ij} = \begin{cases} \exp\,(-d\,(i,j)^2/\sigma^2 & i \neq j \\ 1 & i = j \end{cases} \tag{7}$$

Where $d(i, j)$ represents Euclidean distance concerningx_i and x_j

σ represents fixed kernel parameter and does not diverge with the surroundings change.

In this a local scale parameter σ_i for each point to substitute the fixed parameter σ, permitting the similarity self-tuning capability. Usually, $\sigma_i = d(x_i, x_m)$where x_m is the m-th closest neighbor of the point x_i, and the similarity function defined in Equation (8).

$$S_{ij} = \begin{cases} \exp\,(-d\,(i,j)^2/(\sigma_i\sigma_j)) & i \neq j \\ 1 & i = j \end{cases} \tag{8}$$

The local density of different vertices and edges is characterized by the formation of enhanced nearest neighbor. A set $N(x_i)$and point x_j are constructed by means of closest kd nearest neighbors of point x_i,then the shared neighbor vertexes between $x_i$and $x_j$are given in Equation (9).

$$E_{NN}(x_i, x_j) = |N(x_i) \cap N(x_j)| \tag{9}$$

Higher similarity is said to possess if the vertex lies in the same cluster and a higher local density region is obtained if occurs in different manifolds. The shared nearest neighbors is the key parameter for similarity assessment between vertex x_i and x_j. The construct similarity function is defined in Equation (10).

$$E_{NN_{ij}} = \begin{cases} \exp\,(-d\,(i,j)^2/(\sigma_i\sigma_j E_{NN}(x_i, x_j) + 1)) & i \neq j \\ 1 & i = j \end{cases} \tag{10}$$

If the common nearest neighbors is greater, which replicates higher similarity between two vertexes? The network topology non-homogeneity plays a significant role in the network which in turn reflects the importance of each node. The similarity of two vertices is associated not only to the number of neighbors shared, but also closely to the prominence of the shared neighbor vertices.

Louvain Algorithm

The Louvain algorithm plays a significant role in this research since it is one among the fastest algorithms in community detection along with the precise results. The fastest is termed since it can manipulate a network constructed up of more than six million users in lesser than a minute. It works by setting all graph vertices in individual communities, one per vertex and sweeps inner loop vertices sequentially. The algorithm has two main calculations 1) modularity gain ΔQ when placing vertex i in the community of any neighbor j calculation. 2) Selecting the neighbor j that produces the largest gain in ΔQ and juncture the corresponding community [25].This Process is continued till no gain is achieved. Finally, first level partitioning is achieved. These cod level of process involves the first level partitioning which turn into super vertices and thereby manipulating the weight of edges among all super vertices. There exists connection between two super vertices if the conforming partitioning of an edge between vertices corresponds to the edges sum of the weights and two super vertices edge weights at the lower level. The repetition of these two steps is accomplished for producing new hierarchical levels and super graphs. Whenever the communities are stable, the algorithm halts. The convergence of the algorithm is achieved rapidly and communities are identified in less iteration.

There are three phases involved in this algorithm. The first phase involves a disjoint community detection algorithm. It is mainly meant for applications involving huge scale social networks involving more billions of connection along with processing of one million nodes in minimum time namely ~45sec and two million nodes network takes two minutes time only. While there is an acceptable accuracy in detection of disjoint communities, it cannot be suitable for overlapping communities. Also the modularity factor increases in every step which implicates different results for each iterations. In each run, shared nodes are combined to other community. During First phase, after little iteration, belonging coefficient expectation of every single node to communities is obtained.

The community detection is achieved by measuring by the Modularity factor. However, modularity optimization is an NP-complete problem.

$$\Delta Q_{u\to c} = \left[\frac{\sum_{in}^{c} + w_{u\to c}}{2m} - \left(\frac{\sum_{t\,ot}^{c} + w(u)}{2m}\right)^2\right] \tag{11}$$

$$-\left[\frac{\sum_{in}^{c}}{2m} - \left(\frac{\sum_{t\,ot}^{c}}{2m}\right)^2 - \left(\frac{w(u)}{2m}\right)^2\right] \tag{12}$$

$$= \frac{w_{u|\to c}}{2m} - \frac{\sum_{t\,ot}^{c} * w(u)}{2m^2} \tag{13}$$

Equation 12 narrates the maximization of modularity gain $\Delta Q_{u\to c}$ when the isolated vertex u is progressed into community c in Louvain algorithm, where Σ_{in}^{c}, Σ_{tot}^{c}, and m are defined in Equation 13, w(u) is the sum of the edges weightsincident to vertex u, and $w_{u\to c} = \Sigma_{v\in c}\, w_{u,v}$ is the sum of theedges weights from vertex u to vertices in community c.

Algorithm 1: communities with the maximum modularity Q

```
Input: G = (V, E): graph representation
Output: Over  lapping community
Loop outer
                $C \leftarrow \{\{u\}\}, \forall u \in V;$
        $\sum_{in}^{c} \leftarrow \sum w_{u,v}, e(u,v) \in E, u \in c \; and \; v \in c;$
        $\sum_{t\,ot}^{c} \leftarrow \sum w_{u,v}, e(u,v) \in E, u \in c \; and \; v \in c;$
  // Phase 1
  Loop inner
for$u \in V \; and \; u \in c \; do$
      // Find the best community for vertex u.
              $\hat{c} \leftarrow argmax \; \Delta Q_{u \to c'}$
      If $\Delta Q_{u \to c'} > 0 \; then$
      // Update $\sum_{t\,ot} and \; \sum_{in}$
        $\sum_{t\,ot}^{\hat{c}} \leftarrow \sum_{t\,ot}^{\hat{c}} + w(u); \sum_{in}^{\hat{c}} \leftarrow \sum_{in}^{\hat{c}} + w_{u \to \hat{c}};$
        $\sum_{t\,ot}^{c} \leftarrow \sum_{t\,ot}^{c} - w(u); \sum_{in}^{c} \leftarrow \sum_{in}^{c} - w_{u \to c};$
      // Update the community information
              $\hat{c} \leftarrow \hat{c} \cup \{u\}; c \leftarrow c - \{u\};$
if No vertex moves to a new community then
exit inner loop;
// Calculate community set and modularity
$Q \leftarrow 0;$
for$c \in C \; do$
$Q \leftarrow Q + \frac{\sum_{in}^{c}}{2m} - \left(\frac{\sum_{t\,ot}^{c}}{2m}\right)^2$
$C' \leftarrow \{c\}, \forall c \in C; print \; C' and \; Q;$
// Phase 2: Rebuild Graph
$V' \leftarrow C';$
$E' \leftarrow \{e(c,c')\}, \exists e(u,v) \in E, u \in c, v \in c';$
        $w_{c,c'} \leftarrow \sum w_{u,v}, \forall e(u,v) \in E, u \in c, v \in c';$
If No community changes then
   Exit outer Loop;
                $V \leftarrow V'; E \leftarrow E'$
```

The algorithm 1 describes the community identification with the maximum modularity Q as shown. The modularity gain (ΔQ) is evaluated by taking each vertex u, each neighbor v of u which supports in eradicating u from its community c and by employing it in the neighbor's community c'.This employment is done for maximizing ΔQ in the community ĉ.The absence of modularity gain indicates u to stay back in the parent community. The Processing of all vertices is accomplished till no further improvement is required for attaining the community information $(C)'$ and modularity Q. The subsequentstep involves constructing a fresh graph in which the vertices are the communities instigate in the aforementioned iteration.The latest communities are involved in new set of vertices V', and the sum of the weight of the edges amid vertices in the conforming two communities is found from the weights of the edges amid the new vertices.Self-looping in the new graph may occur due to the edges between vertices of the same community. This iteration is carried out for the above phases for the subsequent level until the communities are steady.

Overlapping community detection using Improved Vertex Imitation Coefficient based Community Overlap Propagation Algorithm (IVIC-COPRA*)*

The extending version of LPA leads to COPRA which is purely prolonged for community detection which are overlapped in which there are *v* labels in every node, v representing parameter of the algorithm.

The main process of COPRA proceeds below.

1) At initial process, inimitable label is assigned for each vertex with the condition that belonging coefficient is made 1.
2) The labels are regularly updated by summing process followed by belonging coefficients of vertices normalization with respect to neighbor set of x. Next involves the COPRA , in which the parameter v is utilized to restrict the vertex possessing the maximum number of communities by which vertex owning all community identifiers is avoided. Once after the satisfaction of stop criteria, the propagation procedure halts after undergoing several iterations.
3) The communities possessed by other communities are completely eradicated.
4) The communities which are not continuous are fragmented.

The propagation procedure in COPRA involves synchronous updating strategy for enhanced results compared with asynchronous updating. Each label has the ability to store many communities and preserving activity is not required over there. Also the belonging coefficient is determined using COPRA for each label of each vertex. The belonging coefficients of the labels whose value is less than threshold are removed. The global threshold is represented as 1/v where v is the parameter of COPRA denoting maximum number of communities to which any vertex can belong. In this the label's belonging coefficients are summed to 1.The vertex-independent parameter v is the prime reason for assigning the belonging coefficients of labels to a vertex whose value is below than the threshold. The COPRA maintains the greatest belonging coefficient label and the remaining is eradicated. While the lower belonging coefficient label possessed by more than one label are said to access randomly and this random fashion mitigates the algorithm stability.

COPRA achieves improved computational performance in contrast with LPA and produces worthy results in various cases. In addition to it, it also embraces a global vertex-independent parameter due to the availability of few vertices with several community memberships in a network, hence there arises tough situation to select a suitable value of v.

Consider the case which contains plenty of non-overlapping vertices where the value of v plays a vital role in recognizing overlapping vertices. There arise two cases: value of v is small representing difficult in recognizing overlapping vertices and greater value will wrongly recognize non-overlapping vertices to be overlapping vertices. The subsequent chapters elucidate a novel update stratagem with a vertex-dependent parameter.

Vertex Imitation Coefficient update strategy (VIC)

There are v labels present mostly in every vertex in the COPRA algorithm. But this Vertex Imitation Coefficients (VIC) update strategy projected in this work does not hold any limitation how many communities it possess belonging to the vertex. The VIC update strategy consists of labels of a vertex that possess belonging coefficients.

Similar to COPR every vertex is labeled as x by means of a set of pairs (c, b), in which c indicating community identifier, imitation coefficient is represented by I which is nothing but the strength of x's membership of community c , whose sum value is equal to

1. Every propagation step x's label involves union of its neighbors' labels and sum the imitation coefficients of the communities over all neighbors to be grouped. The determination of label c max with maximum imitation coefficient i $_{max}$ is the next step. Equation (14) narrates the community identifier

$$\frac{i}{i_{max}} \geq p, \qquad (14)$$

Where pis the threshold parameter, and p∈(0,1).

The maximum imitation coefficient, equation (14) should be satisfied when the balancing of imitation coefficient of a community identifier is accomplished and normalization of imitation coefficients of the retained labels is done. The update procedure is explained previously .The central vertex consists of five neighbors and labels of this vertex are computed by adding neighboring label imitation coefficients. Once the target value 9/6 is obtained, then division is accomplished by the maximum value for obtaining the ratio. At last, final labels of the central vertex in this iteration are nothing but the labels with ratio greater than the threshold. The recalculation of all vertices' labels is performed in a random manner in each iteration. Like other algorithms initialize labels of all vertices i.e. every vertex is denoted by a unique label. If p is selected, the vertices will possess all neighboring community identifiers.

The updation is accomplished by prompting higher label influence for core nodes rather than further community structures nodes. Each node is measured for its label influence. Then ordering of nodes and its label influence mutually depends on each other in a manner such that sorted in descending order for selection of nodes.

The node with the highest label influences the effectiveness of the algorithm. VIC based methods utilizes detecting of initial nodes for updating their label which is the most important step. Initial cores are determined by the nodes with higher label influence. The estimation of the label influence is accomplished using Eq. (16) whose similarity value (u, v) is manipulated by Eq. (15). The optimization of convergence speed is achieved by this strategy and there attains significant stability while updating decision.

$$S(i,j) = \beta A + \beta^2 A^2 + \beta^3 A^3, \qquad (15)$$

where A denotes adjacency matrix, A^2refers to the number of different paths of length 2 connecting nodes i and j, A^3 stands for the path counts of length 3, and $\beta = 1/d_{mean}$. The d_{mean} refers to the average degree of the network.

$$k(u) = \sum v \in \Gamma_{1(u)} \setminus \{u\}\big(similarity(u,v)\big), \qquad (16)$$

Where$\Gamma_{1(u)}$ denotes the node uimmediate neighbors.

The subsequent section compares the existing Community detection techniques with the proposed IVIC-COPRA by which the effectiveness of the initialization procedure and the detection of disjoint and overlapping community is verified.

RESULTS AND DISCUSSION

Normalized Mutual Information

The community detection algorithm performance comprises instituting the segmentation conveyed by the algorithm is to segment the required data [23] utilizing criterion definition.

Normalized mutual information (NMI) is nothing but the similarity proportion in the information theory system. The normalized mutual information $N(X|Y)$is as follows:

$$N(X|Y) = \frac{H(X)+H(Y)-H(X,Y)}{(H(X)+H(Y))/2} \qquad (17)$$

Where H(X), H(Y) is the entropy of the random variable X(Y) associated with the partition $C'$and C'', and H(X,Y) is the joint entropy. This value ranges between 0 and 1 and 1 and the maximum value 1 represents the two partitions $C'$and C'' are exactly coincident**.**

$$N(X|Y) = 1 - \tfrac{1}{2}[H(X|Y)_{norm} + H(Y|X)_{norm}]$$

(18)

The above equation is calculated by using the subsequent equation

$$H(X|Y)_{norm} = \frac{1}{|C'|}\sum_k H(X_k|Y)_{norm} \qquad (19)$$

Where,

$$H(X|Y)_{norm} = \frac{H(X_k|Y)}{H(X_k)} \qquad (20)$$

and:

$$H(X|Y)_{norm} = \min_{l\in\{1,2,......|C''|\}} H(X_k|Y_l) \qquad (21)$$

$$H(X|Y)_{norm} = H(X_k|Y_l) - H(Y_l) \qquad (22)$$

Where$H(X_k|Y)$ is the conditional entropy of X$_k$ with respect to all the components of Y and $H(X|Y)_{norm}$ is the conditional entropy of X with respect to Y. The range of extended NMI lies between 0 and 1 and the maximum value 1 represents perfect match.

Table 1. Comparison of NMI between the community member of different techniques with different months of a year.

Table 1. NMI comparison between the community member of different techniques with different months of the year 2003 from Amazon product co-purchasing network.

Months	**March 2 2003**					**March 12 2003**					**May 2003**					**June 2003**				
No. of data	**No. of Community member**					**No. of Community member**					**No. of Community member**					**No. of Community member**				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
MIGA	0.96	0.93	0.9	0.87	0.85	0.93	0.90	0.86	0.84	0.82	0.93	0.91	0.87	0.86	0.84	0.92	0.89	0.87	0.85	0.82
NGTCDA	0.97	0.94	0.92	0.89	0.87	0.94	0.91	0.88	0.86	0.84	0.95	0.92	0.90	0.88	0.86	0.94	0.92	0.89	0.87	0.84
ENNC-EABC	0.98	0.97	0.95	0.93	0.91	0.96	0.94	0.92	0.89	0.87	0.98	0.96	0.94	0.91	0.90	0.95	0.93	0.90	0.89	0.87
IVIC-COPRA	0.98	0.98	0.96	0.94	0.92	0.97	0.96	0.94	0.92	0.90	0.98	0.97	0.95	0.93	0.92	0.97	0.95	0.92	0.90	0.89

The table .1 displays the Comparison of NMI among the community member of different techniques with different months of a year.

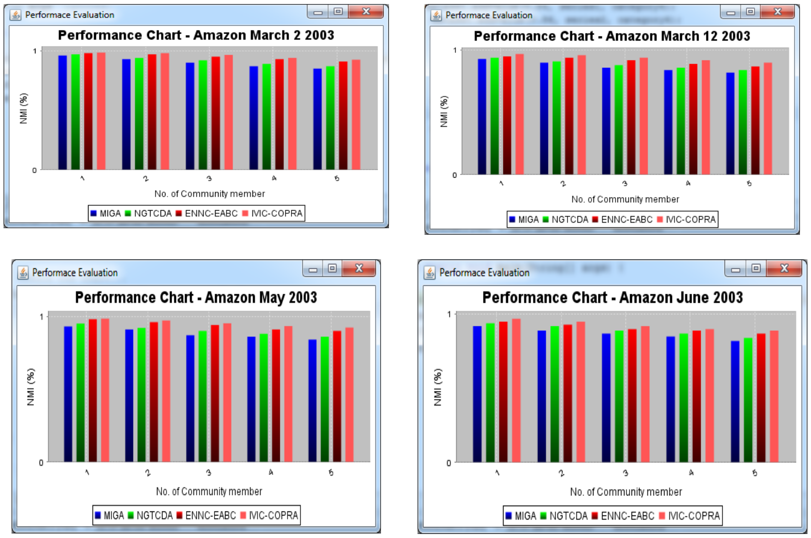


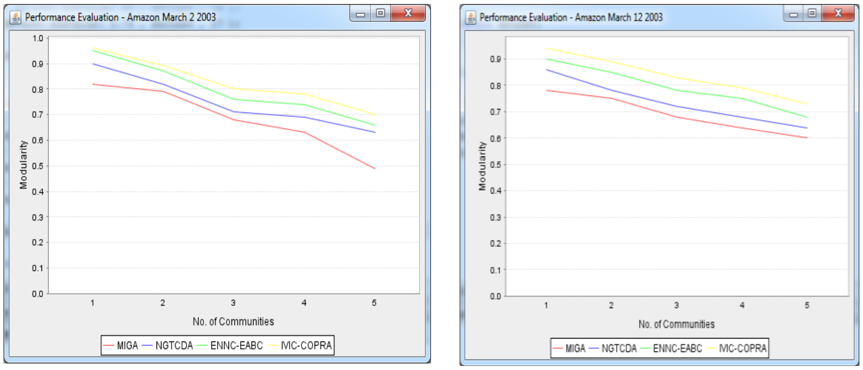
Fig.1. Comparison of NMI among the community members in a network of different techniques along with different months of a year

Fig.1. shows the comparison of NMI among the community members for detecting the community in a spatial network of different techniques along with 4 months in the year 2003. The simulation result reveals that the proposed IVIC-COPRA technique delivers better NMI results when related to the existing ENNC-EABC, NGTCDA and MIGA methods.

Table 2. Comparison of Modularity between the community member of different techniques with different months of a year. The Community size is defined based on the modularity value as an approximation for how users are related in it.

Months	March 2 2003					March 12 2003					May 2003					June 2003				
No. of data	No. of. community					No. of. community					No. of. community					No. of. community				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
MIGA	0.82	0.79	0.68	0.63	0.49	0.78	0.75	0.68	0.64	0.60	0.79	0.77	0.65	0.60	0.48	0.80	0.76	0.65	0.61	0.52
NGTCDA	0.9	0.82	0.71	0.69	0.63	0.86	0.78	0.72	0.68	0.64	0.89	0.80	0.68	0.66	0.62	0.88	0.81	0.73	0.68	0.61
ENNC-EABC	0.95	0.87	0.76	0.74	0.66	0.90	0.85	0.78	0.75	0.68	0.93	0.86	0.74	0.72	0.65	0.93	0.85	0.78	0.73	0.66
IVIC-COPRA	0.96	0.89	0.80	0.78	0.70	0.94	0.89	0.83	0.79	0.73	0.96	0.90	0.82	0.77	0.69	0.95	0.88	0.85	0.80	0.74

The table 2 Comparison of Modularity between the community member of different techniques with different months of a year.



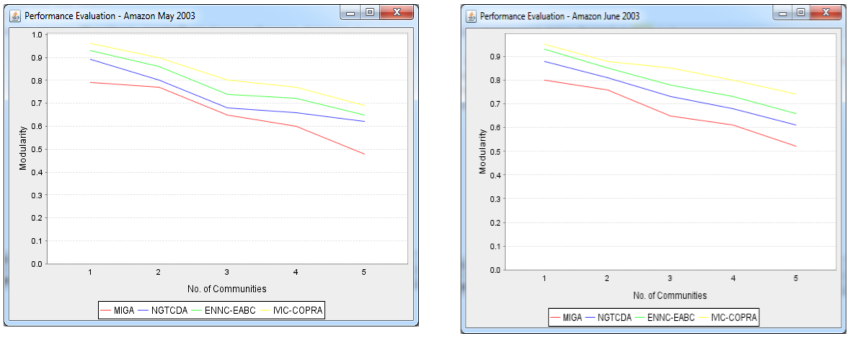


Fig.2. Modularity assessment comparisonbetween the proposed and existing methods

Fig.2. shows the Modularity assessment comparison between the proposed and existing methods for community detection with different months of a year. The simulation result reveals that the proposed IVIC-COPRA technique delivers higher modularity rate when related to the existing ENNC-EABC, NGTCDA and MIGA methods in the presence of more community.

Table 3. Comparison ofRunning time between the community member of different techniques with different months of a year.

Months	March 2 2003					March 12 2003					May 2003					June 2003				
No. of data	No. of. data					No. of. data					No. of. data					No. of. data				
	1000	2000	3000	4000	5000	1000	2000	3000	4000	5000	1000	2000	3000	4000	5000	1000	2000	3000	4000	5000
MIGA	66	74	82	87	93	70	73	78	85	92	67	75	83	88	95	64	71	78	85	89
NGTCDA	55	66	70	82	90	65	69	76	82	90	58	68	72	84	92	52	63	68	78	83
ENNC-EABC	47	56	62	74	87	60	64	71	80	87	50	59	65	78	88	45	52	59	71	79
IVIC-COPRA	44	52	58	69	83	52	59	67	76	84	46	54	61	73	84	40	49	56	67	75

The table .3 shows the Assessment of Running time between the community member of different techniques with different months of a year.

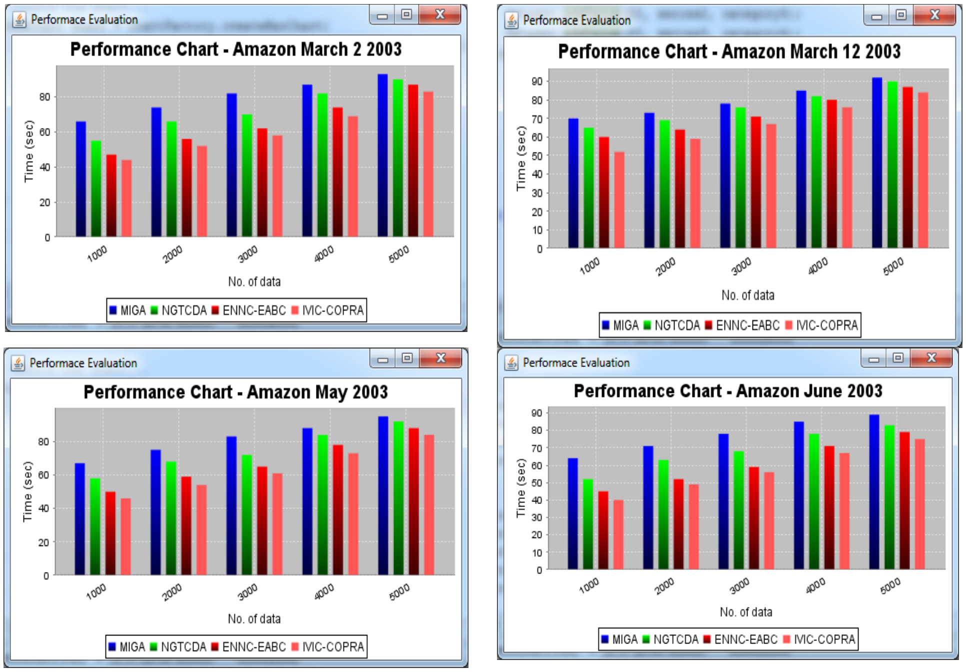


Fig.3. Running time comparison between the proposed and existing methods

Fig.3. shows the assessment of running time among the proposed and existing methods for community detection with different months of a year.

By applying community detection for Amazon co-purchase network, this work explores the way these communities behave for different products over months. The simulation result reveals that the proposed IVIC-COPRA technique delivers lesser running time when related to the existing ENNC-EABC, NGTCDA and MIGA methods in the presence of more community.

CONCLUSION

This research paper concentrates on investigating disjoint and overlapped communities with the support of modularity based community detection algorithms. The best available method is chosen by performing Quantitative analysis with the modularity score. Also Optimized Overlapping and disjoint Community Detection (OODCD) technique by Improved Vertex Imitation Co-efficient based Community Overlap Propagation Algorithm (IVIC-COPRA) is also introduced in this work. It has grabbed the attention since is segregates the “overlapping” and “community detection” issues, permitting the finest algorithm to be designated for each phase.

In this research, Enhanced Nearest Neighbor-Based Clustering (ENNC) approach is utilized basically which is modularity based approach for partitioning the network into minor local communities. The Louvain method is then used for detecting the disjoint community in the given network. The belonging matrix plays a vital in this research which is regularly updated where each matrix elemental value decides the role of node particular to that community and thus the overlapping communities is found with the support of Improved Vertex Imitation Co-efficient Community Overlap Propagation Algorithm (IVIC-COPRA). The outcomes acquired by the proposed technique is beneficial in investigating overlapping communities effectively even from the enormous real world systems as far as high similarity with the ground-truth network structure.In future the propose stratagem can be utilized in a dynamic social network with machine learning techniques. The strength of community node in the social network field can also be analyzed with the support of edge weights.

REFERENCES

1. Aggarwal, C. C. (2011). An introduction to social network data analytics. In *Social network data analytics* (pp. 1-15). Springer, Boston, MA.
2. Lancichinetti, A., Fortunato, S., &Radicchi, F. (2008). Benchmark graphs for testing community detection algorithms. *Physical review E*, *78*(4), 046110.
3. Leskovec, J., Lang, K. J., & Mahoney, M. (2010, April). Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th international conference on World wide web* (pp. 631-640). ACM.
4. Bródka, P., Filipowski, T., &Kazienko, P. (2011, September). An introduction to community detection in multi-layered social network. In *World Summit on Knowledge Society* (pp. 185-190). Springer, Berlin, Heidelberg.
5. Tasgin, M., Herdagdelen, A., &Bingol, H. (2007). Community detection in complex networks using genetic algorithms. *arXiv preprint arXiv:0711.0491*.
6. Xie, J., Kelley, S., & Szymanski, B. K. (2013). Overlapping community detection in networks: The state-of-the-art and comparative study. *Acm computing surveys (csur)*, *45*(4), 43.
7. Gregory, S. (2010). Finding overlapping communities in networks by label propagation. *New Journal of Physics*, *12*(10), 103018.
8. Danon, L., Diaz-Guilera, A., Duch, J., & Arenas, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, *2005*(09), P09008.
9. Lancichinetti, A., Radicchi, F., Ramasco, J. J., &Fortunato, S. (2011). Finding statistically significant communities in networks. *PloS one*, *6*(4), e18961.
10. Ahn, Y. Y., Bagrow, J. P., & Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. *nature*, *466*(7307), 761.
11. Chakraborty, T. (2015). Leveraging disjoint communities for detecting overlapping community structure. *Journal of Statistical Mechanics: Theory and Experiment*, *2015*(5), P05017.
12. Gregory, S. (2009). Finding overlapping communities using disjoint community detection algorithms. In *Complex networks* (pp. 47-61). Springer, Berlin, Heidelberg.
13. Saini, R., Saini, S., Sharma, S.Potential of probiotics in controlling cardiovascular diseases(2010) Journal of Cardiovascular Disease Research, 1 (4), pp. 213-214.
DOI: 10.4103/0975-3583.74267
14. Yang, J., &Leskovec, J. (2013, February). Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining* (pp. 587-596). ACM.
15. Lee, C., Reid, F., McDaid, A., & Hurley, N. (2010). Detecting highly overlapping community structure by greedy clique expansion. *arXiv preprint arXiv:1002.1827*.
16. Gregory, S. (2010). Finding overlapping communities in networks by label propagation. *New Journal of Physics*, *12*(10), 103018.
17. Nishi Gupta, Garima Vishnoi, Ankita Wal, Pranay Wal. "Medicinal Value of Euphorbia Tirucalli." Systematic Reviews in Pharmacy 4.1 (2013), 40-46. Print. doi:10.4103/0975-8453.135843
18. Gopalan, P. K., &Blei, D. M. (2013). Efficient discovery of overlapping communities in massive networks. *Proceedings of the National Academy of Sciences*, *110*(36), 14534-14539.
19. Evans, T. S., &Lambiotte, R. (2009). Line graphs, link partitions, and overlapping communities. *Physical Review E*, *80*(1), 016105.
20. Psorakis, I., Roberts, S., Ebden, M., & Sheldon, B. (2011). Overlapping community detection using bayesian non-negative matrix factorization. *Physical Review E*, *83*(6), 066114.
21. Zhang, Z. Y., Wang, Y., &Ahn, Y. Y. (2013). Overlapping community detection in complex networks using symmetric binary matrix factorization. *Physical Review E*, *87*(6), 062803.
22. Nepusz, T., Petróczi, A., Négyessy, L., &Bazsó, F. (2008). Fuzzy communities and the concept of bridgeness in complex networks. *Physical Review E*, *77*(1), 016107.
23. Chopade, P., & Zhan, J. (2017). A framework for community detection in large networks using game-theoretic modeling. *IEEE Transactions on Big Data*, *3*(3), 276-288.
24. He, X., Zhang, S., & Liu, Y. (2015). An Adaptive Spectral Clustering Algorithm Based on the Importance of Shared Nearest Neighbors. Algorithms, 8(2), 177-189.
25. Gach, O., &Hao, J. K. (2013, October). Improving the Louvain algorithm for community detection with modularity maximization. In *International Conference on Artificial Evolution (Evolution Artificielle)* (pp. 145-156). Springer, Cham.
26. Fu, X., Liu, L., & Wang, C. (2013). Detection of community overlap according to belief propagation and conflict. *Physica A: Statistical Mechanics and its Applications*, *392*(4), 941-952.