Journal of Theoretical and Applied Information Technology
30th June 2024. Vol.102. No. 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A SURVEY ON COMMUNITY DETECTION: APPLICATIONS, ALGORITHMS, AND CHALLENGES

LAASSEM BRAHIM [1], MAMMASS MOUAD [2], CHERKAOUI CHIHAB-EDDINE [3] AND IDARROU ALI [4]

[1,2,3,4] Laboratory IRF-SIC, Faculty of Sciences, University Ibn Zohr Agadir, Morocco

E-mail: [1]brahim.laassem@edu.uiz.ac.ma, [2]m.mammass@uiz.ac.ma, [3]ccherkaoui@gmail.com, [4]a.idarrou@uiz.ac.ma

ABSTRACT

Community detection in networks is the process of identifying groups of nodes with more connections within the group than with the rest of the network. Community detection plays an important role in the field of complex network analysis. It can be used to understand the relationships and dynamics within a network, design better recommendation systems, and create powerful network visualizations. This paper surveys the most important approaches in this field and classifies them into categories based on their fundamental working principle. We have briefly discussed real-world community detection applications. Moreover, the paper discusses the strengths and limitations of each approach and the challenges and future research directions in this area. Overall, this article provides a comprehensive overview of the current state of the art in community detection and serves as a valuable resource for researchers in this field.

Keywords: *Social Network, Complex Network, Community Detection, Graph Clustering*

1. INTRODUCTION

In many disciplines, it is possible to represent real systems in the form of networks. These networks are modeled by graphs whose nodes constitute the system's objects, and the links represent their interactions. One of the essential problems on networks is the detection of communities. Communities are mesoscopic structures of the network, known to be present in most field graphs, especially in small-world networks [1]. We can define them as sets of nodes that have more connections within the group than with the rest of the network.

The terms cluster, module, and group are often used interchangeably in literature and refer to community. The major component in the community detection (CD) algorithm is to have the right definition of what's a community. Different definitions have been made in the literature, and they are implemented in various approaches [2]. Therefore, community division results can differ within the same network depending on the definition set out by the analyst.

In recent years, community detection (CD) has attracted the attention of researchers from many fields, such as biology, economy, social science, epidemic propagation, marketing, and so on. The application of community detection is discussed in section 2. Community detection has been and remains a highly active research area in social network analysis. For example, given the big data generated and the growth of social network services, community detection algorithms have been used in many social network analysis tasks, such as user classification, friend recommendation, and behavior prediction.

Generally, studying the community structure in networks is closely related to graph partitioning in graph theory. Since 2002, where the introduction of the well-known algorithm by Girvan and Newman [3], the amount of work published on community detection is substantial, and it is impossible to discuss all of it in one single work. In this paper, we summarize several key contributions in the field, with a particular emphasis on highly cited articles. Additionally, we devote attention to newer and innovative approaches, recognizing their potential to exert significant influence in the future. Moreover, we include simple examples to help explain the functioning of some of the approaches discussed.

Thus, we present a new research survey on community detection techniques. Our survey goes beyond simply listing CD techniques by offering new critiques and insights into existing CD approaches while proposing a novel taxonomy that categorizes them based on their fundamental

Journal of Theoretical and Applied Information Technology
30th June 2024. Vol.102. No. 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

working principles, including hierarchical methods, spectral methods, label propagation methods, modularity optimization, and metaheuristic algorithms. In addition, by outlining existing challenges and proposing several promising directions for future research, our work not only consolidates current knowledge but also motivate further development in the field. The main contributions of our work are the following:

- ✓ We describe and analyze the most prominent strategies for community detection.
- ✓ We introduce a new taxonomy that categorizes existing methods depending on the technique used.
- ✓ We outline several points of interest based on the study of each family of community detection papers.
- ✓ We provide useful guidance for future research in the field of community detection.

The remainder of this paper is organized as follows. Section 2 briefly introduces the community detection task and its application. It introduces basic definitions which are required for this study. Section 3 is divided into five major sub-sections, each studying a different category of community detection methods. Section 4 describes issues and challenges and provides recommendations for future work on this topic. Section 5 presents conclusions.

2. PRELIMINARIES

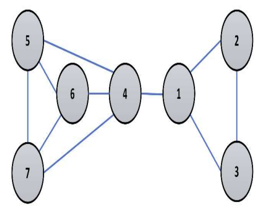
2.1 Terminology Definition

Complex networks are at the core of the natural and human sciences because they make it possible to represent the interactions between the different elements of a system. These networks are often modeled by graphs, a mathematical structure allowing the encoding of relational data. For example, the World Wide Web (WWW) is a complex network with nodes representing web pages and edges representing the hyperlinks between them. Analyzing this network can help to understand and evaluate relationships between web pages, how they are grouped into different themes or communities by identifying similar web pages, and which are the most central or important.

Formally, a graph G is defined as a couple $G = (V, E)$ with $n = |V|$ the number of vertices known as the size of the graph, and $m = |E|$ the number of edges in the graph.

The graph can be directed, i.e., links between the vertices are oriented (the relations are unidirectional). In contrast, an undirected graph is composed of undirected links (relationships are bidirectional). Also, the graph can be weighted, i.e., a weight (or cost) is associated with each link, or unweighted. Figure 1 illustrates a small example of a simple undirected unweighted graph and its adjacency matrix.

Graph $G\ (V, E)$ can be represented by an adjacency matrix $A \in R^{n\times n}$. $A_{ij} = 1$ if there is a link between v_i and v_j where $i, j \in \{1, 2, ..., n\}$, otherwise $A_{ij} = 0$. $N(i)$ represents the neighborhood set of vertex v_i (all adjacent nodes) with $N(i) = \{j, v_j \in V \text{ and } \{ v_i, v_j \} \in E\}$. The degree of node v_i is defined as the number of edges between node v_i and other nodes, noted $k_i = |N(i)|$.



$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

(a) (b)

Figure 1. An Undirected Graph $G = (V, E)$, with $V=\{1,2,3,4,5,6,7\}$ and $E=\{(1,2),(1,3), (1,4), (2,3), (4,5), (4,6),(4,7),(5,6),(5,7),(6,7)\}$. (a)The Graphical Representation and (b) the Associated Adjacency Matrix. $N(v_4) = \{v_5, v_6, v_7, v_1\}$ and $k_4=4$.

2.2 Problem Statement

2.2.1 Community detection

Let $G = (V, E)$ be a network, the task of community detection is to identify k sub set of nodes which are represented by: $C= \{C_1, C_2,..., C_k\}$ and $V = \cup_{i=1}^{k} C_i$. This partition needs to satisfy the used definition for the community. Thus, a community can be viewed as a set of similar nodes sharing certain common characteristics, or as a densely induced subgraph with a greater number of internal links than external ones.

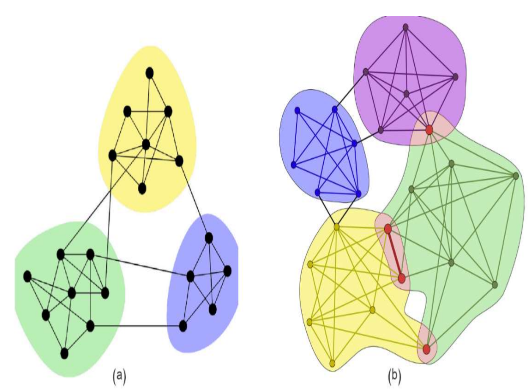


Figure 2. (a) Non-overlapping Communities. (b) Overlapping Communities

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The existing algorithms can be divided into two categories - disjoint and overlapping (see Figure 2). In disjoint mode, $C_i \cap C_j = \emptyset \; \forall (i,j) \epsilon [1,k]$. In overlapping mode, one can find a pair of communities C_i and C_j such that $C_i \cap C_j \neq \emptyset$.

2.2.2 Community detection application

Community discovery provides a way to organize a large-scale network into small components to reduce analysis complexity and better understand network structure and topology. It has a wide range of applications in several areas:

- ✓ **Social network analysis:** Community detection makes several social network analysis (SNA) applications more effective and efficient: the diffusion of a new idea or new technologies can be maximized by identifying groups of people interested in a given subject, the suggestion of friends or groups can be improved by also taking into account the user profiles and the behaviors of people in the same communities [4], [5], the search tasks of experts could be more precise if the users are previously subdivided into thematic groups. Additionally, community detection can be used to identify criminals based on online social behaviors [6] and identify people (groups) who support and spread criminal ideas or support terrorist acts [7].
- ✓ **Smart Advertising and Target Marketing:** Businesses can provide a better service solution if they know their customer groups (i.e., customer segmentation). Then they can advertise and market to specific detected groups [8], [9]. Community detection is also used in financial market modeling [10] as an effective decision-support tool.
- ✓ **Biology and biochemistry:** proteins with a similar function in protein-protein interaction networks can form a community [3], [11], [12]. Also, community detection is used to detect diseases such as cancers and tumors [13].
- ✓ **Link prediction:** Link prediction evaluates the possibility of future links between entities by analyzing the topological structure of the network [14], [15]. Link prediction is basically used to detect missing and fake links and predict the future existence of links in the evolution of the network. For example, one might predict future friendships when analyzing social networks or future co-authors in a collaborative network.
- ✓ **The spread of an epidemic:** Community structure is essential to understanding the spread of infectious diseases in populations represented by networks [16], [17]. A common way to study the spread of an epidemic in complex networks is to simulate it as a discrete process. Initially, all network nodes are healthy and susceptible to infection. Then a randomly chosen vertex gets infected and then the outbreak begins to spread. Most of the time, epidemic spread occurs between overlapping communities whose members are strongly bonded [18].
- ✓ **Intrusion detection and malware analysis:** Community detection algorithms have been of focus recently in identifying anomalies across various domains, including security, finance, healthcare, and law enforcement [19], [20]. By leveraging the principles of CD within the context of network analysis, anti-malware systems can enhance their capabilities to identify and mitigate malicious activities [21], [22], [23]. These approaches often involve the analysis of network structures, relationships, and communication patterns to uncover potential threats. Thus, understanding the community structure within a network is very important to enhance the accuracy and effectiveness of intrusion detection mechanisms [24].
- ✓ **Partitioning and visualization of large graphs:** Today, the size of the networks to be analyzed is increasing rapidly and exponentially, hence the need to decompose the network or partition the vertices into clusters for a better understanding and exploration of the system [25], [26].

2.3 Related Surveys

A plethora of community detection methods exists, each with its own characteristics, advantages, and disadvantages. Several surveys and comparative studies of the area have been proposed in the last decade in particular [2], [27], [28], [29], [30], [31], [32], [33], [34], Table 1. gives a summary of the previous surveys in the field of community detection and lists, for each contribution, the proposed classification for all cited algorithms.

Instead of focusing on a single research direction, such as deep learning methods or local community finding, this paper comprehensively reviews, summarized the state-of-the-art community detection approaches. We categorize these techniques for better understanding and provide in-depth discussions for each algorithm,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

including their features, advantages, disadvantages, and technical details. Additionally, the survey explores the practical applications of community detection algorithms, offering a broad perspective on their real-world relevance. Finally, it concludes with an exploration of open issues and the identification of promising future directions in the field of community detection.

3. ALGORITHMS FOR COMMUNITY DETECTION

In the past decade, there has been a lot of research in the field of community detection on networks, with attention to designing powerful new algorithms that can be applied to large graphs. In the following, we propose a taxonomy of community detection techniques and provide technical details for each cited approach.

3.1 Hierarchical Methods

The detection of the community structure in a network using hierarchical clustering is generally intended as a procedure for mapping the network into a tree. By taking the order of construction of the tree (called dendrogram) into consideration by adding links to or removing links from the network, these algorithms are divided into two classes, agglomerative (merging) and divisive (splitting). A summary of the discussed methods is given in Table 2.

- ✓ **Agglomerative:** Starting from the set of all nodes and no edges, links are iteratively added between pairs of nodes based on the similarity measure. At each step, nodes are grouped into larger and larger communities, and the tree is built up to the root, representing the whole network.
- ✓ **Divisive:** This is a top-down clustering approach; initially, all the nodes in the network are assigned into one cluster. Starting from the root, the division is performed recursively as one moves down (removing edges) the hierarchy until the desired number of clusters is obtained.

The results of hierarchical clustering are usually presented in a dendrogram (Figure 3(c)), which can be used to identify the multilevel community structure of the network. Many approaches calculate the modularity [35] for each partition of a network into communities moving down in the dendrogram and look for its maximal value, which indicates the best network partition.

3.1.1 Divisive algorithms

The algorithm proposed by Newman and Girvan [36] is still the reference for community identification. The approach is based on the method of "finding edges with the highest betweenness", a generalization of the centrality betweenness [37] [38]. The edge betweenness of an edge is the number of shortest paths between pairs of vertices that pass through it. This measure is high for edges between communities compared to those inside communities. The global idea is that if two communities are joined by only a few inter-community edges, then all the shortest paths from one community to another must traverse one of those few edges. In the following, the major steps of the algorithm, an example of execution is given in Figure 3:

1. Find and remove the edge with the highest betweenness centrality (BC) score in the network using one of these measures:
   - ✔ <u>shortest-path betweenness:</u> Find the shortest paths (geodesic distance) between all pairs of vertices and count how many runs along each edge.
   - ✔ <u>Random-walk Betweenness:</u> The number of times a random walk between all vertex pairs in the graph will pass down a particular edge.
   - ✔ <u>current-flow betweenness:</u> The network is imagined as a resistor network; edges represent resistors, and nodes are junctions between resistors. The current-flow betweenness measures the fraction of current that flows through a vertex (or an edge) along all the possible pairs of vertices in the network.
2. Recalculate betweenness for all remaining edges.

Repeating the two steps mentioned above until the whole graph is divided into a set of isolated nodes, then calculate the modularity for each partition of a network into communities

according to dendrogram levels, and look for its maximal value, which indicates the best division.

The method proposed by Fortunato, Latora, and Marchiori [40] for finding community structure is a modification of the method by Girvan and Newman [36] based on removing the edge having the highest edge betweenness. The proposed algorithm makes use of a centrality information measure that is based on the concept of efficient propagation of information over the network. The approach consists of the iterative removal of the edges with the highest information centrality until the graph breaks into isolated nodes.

Table 1. Summary of Previous Survey Articles

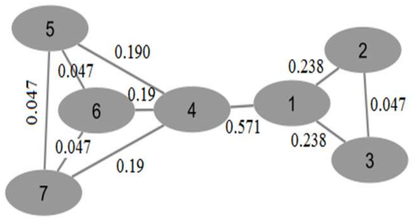
Ref	Authors	Year	Type	Prime contribution	Proposed classification
[2]	Coscia, Giannotti, and Pedreschi	2011	Review	Reviewed techniques for Community detection. Techniques investigated are classified based on the definition of community.	✓ Feature Distance ✓ Internal Density ✓ Bridge Detection ✓ Diffusion ✓ Closeness ✓ Structure ✓ Link Clustering ✓ Meta Clustering
[32]	Plantié and Crampes	2013	Survey	Presented novel classification of community detection algorithms according to their input (Graph, Hypergraph, Galois lattices) and output (disjoint or overlap) data schemes.	Classification rules are proposed based on the mixing of input data and output data type
[31]	Amelio and Pizzuti	2014	Survey	Reviewed state-of-the-art approaches for the detection of overlapped communities in static and dynamic networks.	✓ Node seeds and local expansion ✓ Link clustering ✓ Label propagation ✓ Clique expansion ✓ Dynamic networks ✓ Other approaches
[33]	Drif and Boukerram	2014	Literature Survey	A survey and taxonomy on community discovery methods.	Classification into two main categories (agglomerative, divisive) each divided into subcategories (Stochastic Methods and Deterministic Methods)
[29]	Bedi and Sharma	2016	Review	Presents a review of the existing algorithms, datasets, and metrics for the detection of communities in social networks.	✓ Clustering-based ✓ Genetic Algorithms ✓ Label Propagation-based ✓ Clique-based Methods for Overlapping ✓ Non-Clique Methods for Overlapping ✓ Algorithms for Dynamic Networks
[27]	Yang, Algesheimer, and Tessone	2016	Review	A comparative analysis of the most widely used community detection algorithms on artificial networks.	This study is limited to 8 algorithms implemented within the igraph library and tested on LFR benchmarks
[34]	Fortunato and Hric	2016	User guide	The authors give a complete survey of many aspects of the complex network and discuss community detection concepts, techniques, and performances. This study expands upon and provides additional insights into the previous survey [39] conducted by Fortunato in 2010.	✓ Consensus clustering ✓ Spectral methods ✓ Overlapping communities ✓ Methods based on statistical inference ✓ Methods based on optimization ✓ Methods based on dynamics ✓ Clustering algorithms for dynamic networks
[30]	Khan and Niazi	2017	Review and Visual Survey	Analysis of all relevant articles from the Web of Science to identify the most influential, central author, article, and journal. The authors use CiteSpace for visualizing and analyzing the literature on the community detection field.	✓ Traditional Techniques ✓ Overlapping techniques ✓ Algorithms for dynamic networks
[28]	Jin et al.	2021	Survey	Many important clustering methods are discussed which utilize deep learning and probabilistic graphical models of the network.	✓ Probabilistic graphical model-based methods ✓ Deep learning-based methods

Because the Girvan-Newman [36] algorithm is computationally expensive, it requires the repeated evaluation of edge betweenness for each edge in the whole graph, Radicchi et al. [41] attempted to further explore the local measure of nodes and introduce a new algorithm based on removing the

Journal of Theoretical and Applied Information Technology
30th June 2024. Vol.102. No. 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

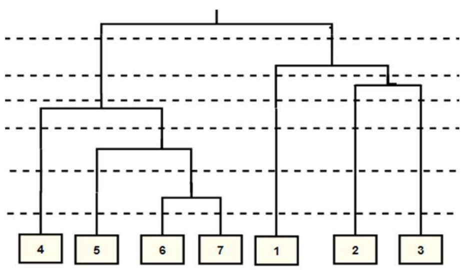
edge having smallest edge clustering defined in the following :



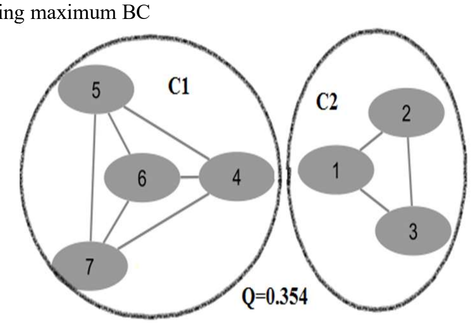
(a) The Graph with betweenness centralities of all edges, in next we remove the edge (1,4) having maximum BC and recalculate the BC for all remaining edges

Removed edge	Modularity
(1,4)	0.354
(1,2)	0.354
(1,3)	0.214
(2,3)	0.134
(4,5)	0.134
(4,6)	0.134
(4,7)	0.0149
(5,6)	0.0149
(5,7)	-0.095
(6,7)	-0,15

(b) Modularity obtained after removing one by one the edge having maximum BC



(c) dendrogram of community division, cutting at a specific level gives a set of clusters.



(d) Community structure with maximum modularity obtained when we remove edge (1,4)

Figure 3. Illustration of the Girvan Newman Algorithm : (a) Network Employed as an Example with Betweenness Centralities of all Edges, (b) Modularity Obtained After Successive Removing of Edges, (c) Result Dendrogram, Each Level Indicates a Aommunity Structure (d) Community Structure Having Maximum Modularity

Edge clustering: the number of triangles to which a given edge belongs, divided by the number of triangles that might potentially include it. The edge-clustering coefficient for the edge connecting *i* to *j* is:

$$C_{i,j}^{3} = \frac{z_{i,j}^{3} + 1}{\min\left[(k_i - 1), (k_j - 1)\right]} \quad (1)$$

$z_{i,j}^{3}$ *the number of triangles of order 3 built on edge* e_{ij}

Based on the degree of a node, the author gives two definitions of what is a community:

- ✓ **Strong community:** strong interactions between nodes inside the community than other nodes, i.e., few inter-community edges.
- ✓ **Weak community:** the sum of all degrees within a community is larger than the sum of all degrees toward the rest of the network.

The proposed algorithm starts by choosing a definition of community (Weak or Strong), then repeats the following steps until the whole graph is divided into a set of isolated nodes:

1. Compute the edge clustering for all edges and remove those with the smallest value, repeat this step if the removal does not split the (sub-) graph.

2. If the removal divides the (sub) graph, test if the two subgraphs resulting meet the community definition chosen. If so, keep the corresponding part on the dendrogram.

Many optimization algorithms use modularity as the optimized objective equation so that a good community division result is obtained when the objective function reaches the maximum. Duch and Arenas [42] proposed a divisive algorithm to find communities in complex networks based on an extremal optimization algorithm [43], the modularity is the global variable to optimize and λ_i ,the fitness of node i, is the local variable involved in the extremal optimization process defined as:

$$\lambda_i = \frac{k_{r(i)}}{k_i} - a_{r(i)} \quad (2)$$

- ✓ $k_{r(i)}$ The number of links that node *i*, belonging to a community *r*, has with nodes in the same community.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- ✓ $a_{r\ (i)}$ the fraction of links from i to vertices inside the community r

The proposed algorithm follows these steps:

- ✓ Split the whole graph into two random-equal partitions.
- ✓ Repeatedly moving the node with the lower fitness from one partition to the other until an "optimal state" with a maximum value of modularity is reached.
- ✓ Proceed recursively with every partition until the modularity cannot be improved.

3.1.2 Agglomerative algorithms

Modularity optimization is often combined with hierarchical clustering in many works. The algorithm proposed by Newman [44] starts with a state of n communities. It repeatedly merges communities in pairs, choosing at each step the join that results in the greatest increase (or smallest decrease) of modularity. In other work, Clauset, Newman, and Moore [45] investigated the performance of such an algorithm by using more sophisticated data structures (balanced binary trees and max heaps) combined with greed optimization to analyze an extremely large network, the proposed method is well known as FastGreedy. The greedy optimization used by [45] does not necessarily lead to a fully optimized result, so Wakita and Tsurumi [46] propose three heuristic methods to improve this algorithm in a large-scale social network.

An *ℓ-shell* with radius ℓ centered at node i is defined as the subgraph containing all nodes connected to i through paths no longer than ℓ. Starting from this definition, the proposed algorithm by [47] performs an ℓ-shell spreading over neighbors to detect the local community by comparing the change in the total emerging degree to some threshold when this measure ceases to grow, all vertices covered by shells of a $depth \leq \ell$ formed a community around started vertex. A global application of this method is also introduced to detect all communities using a membership matrix obtained by performing a local community detection to all nodes. However, the drawback of matrix operations complexity has limited its functionality in practice.

Another well-known agglomerative algorithm is the Louvain algorithm proposed in [48], this approach is similar to [44], [45], where the modularity value is to be optimized in an iterative process. The Louvain algorithm starts from n partitions and tries to maximize modularity locally by considering the nodes' merges which increase modularity. Then, the algorithm builds a new network by considering the found communities as nodes. The process of aggregating nodes is applied repeatedly until no further improvement of modularity, and the complexity of this algorithm is estimated to run in time $O(n \log n)$. The Leiden [49] algorithm is introduced as an improvement of the Louvain algorithm. While the Louvain algorithm iteratively visits all nodes in a network to detect movements that enhance the quality function, the Leiden algorithm employs a fast local move procedure, focusing solely on nodes whose neighborhoods have changed. As a result of its additional refinement steps, the Leiden algorithm is more complex than the Louvain algorithm but faster and produces better results.

Studies of node similarity measures have attracted considerable attention in the field of agglomerative community detection methods. Starting from the individual nodes, larger communities are formed by merging groups of nodes based on their similarity. Based on such local information and node properties, several interesting node similarity metrics have been proposed and applied to community detection, some were discussed in [50]. Most of the proposed clustering algorithms using node similarities depend on the features of node neighborhoods. Based on the Zhou_Lü_Zhang index [50], Pan et al. [51] proposed a new metric using only common neighbors of connected nodes. H. Yang et al. [84] found the initial community structure using the Jaccard index combined with the adjacency matrix. Then, some small communities were merged with a measure called link strength to get the final result. In [56], Saoud and Moussaoui combine node similarity with modularity to detect predefined k clusters.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2. Summary of Hierarchical Community Detection Methods. All Methods are for Undirected Unweighted Graph

Autor and Ref	Publication Year	Method class	Based on	Complexity
M. E. J. Newman and M. Girvan **[36]**	2004	Hierarchical divisive	Betweenness centrality	$O(n^3)$
Newman **[44]**		Hierarchical Agglomerative	greedy optimization	$O(n^2)$
Radicchi et al **[41]**		Hierarchical divisive	edge clustering	$O(n^2)$
Clauset et al **[45]**		Hierarchical Agglomerative	greedy optimization	$O(n\log^2 n)$
Donetti and Mûnoz **[53]**		Hierarchical Agglomerative	Spectral representation	$O(n^3)$
Fortunato, Santo, et al. **[40]**		Hierarchical divisive	Information Centrality	$O(n^4)$
Zhou and Lipowsky **[54]**		Hierarchical Agglomerative	Random walks	$O(n^3)$
Pascal Pons and Matthieu Latapy **[55]**	2005	Hierarchical Agglomerative	Random walk	$O(n^2 \log n)$
Duch J and Arenas **[42]**		Hierarchical divisive	the extremal optimization	$O(n^2 \ln n)$
Bagrow et al. **[47]**		Hierarchical Agglomerative	l-shell spreading	$O(n^3)$
Wakita and Tsurumi **[46]**	2007	Hierarchical Agglomerative	Greedy optimization with heuristic methods	----
Blondel et al. **[48]**	2008	Hierarchical Agglomerative	Modularity optimization	$O(n \log n)$
Ying et al. **[51]**	2010	Hierarchical Agglomerative	Node similarity	$O(n*k)$ *k mean degree of network*
Traag et al. **[49]**	2019	Hierarchical Agglomerative	Modularity optimization	----

3.2 Methods Based On Modularity Optimization

An effective way to detect the community structure of a network is to start from an objective function. By optimizing this objective function, we can detect the community structure of a given network. The most used quality function is modularity. It measures the quality of a particular network clustering by summing the difference between the actual fraction of edges that connect nodes from the same community and such fraction expected in an equivalent randomized graph with the same number of nodes and the same degree distribution. The problem of Modularity maximization is NP-hard [57] in general; therefore, many approximation algorithms or heuristics are used to solve such a problem. Those approaches discussed in [58], include extremal optimization [42], greedy algorithms [44], [48], simulated annealing [59], [60], spectral methods [61], [62], sampling technique [63], and mathematical programming [64]. Some methods are previously discussed in detail.

Modularity depends on the number of links in the network, it suffers from a problem known as the resolution limit established by Fortunato and Barthemely [65]. Optimizing modularity leads, under certain conditions, to the disappearance of small communities and incorrect estimation of the number of clusters, i.e., we cannot affirm that a cluster associated with a peak of modularity is a single cluster or can be divided to more smaller clusters.

3.3 Label Propagation Methods

The label propagation approach is similar to message passing and epidemic spreading, as it involves the exchange of information or influence between vertices in the network. It is based on the idea that a vertex is likely to belong to the same community as most of its neighbors. In other words, a signal or message emitted by a node and passed along to its neighbors is more likely to stay within the same community as the source node, rather than spread to other communities. This is because the communities are relatively dense and connected within themselves, and the connections between communities are weaker.

The first attempt to use label propagation for community detection was made by Raghavan et al. [66]. In this work, the authors proposed a label propagation algorithm (LPA) for detecting communities in large-scale networks. The algorithm starts with an initial partition where each vertex is initiated with a different label denoting the community to which they belong. Then, the label of each node is updated to the most common label among its neighbors. This process is repeated until the labels of all nodes have converged or until a maximum number of iterations is reached. In the following, the major steps of the Raghavan algorithm (LPA), the method is detailed in Algorithm 1, and an example of execution is given in Figure 4:

Algorithm 1: The LPA Algorithm

Input: undirected connected network $G = (V, E)$, the maximum number of iterations t_{max}

Initialization: for each node $v \in V$, set label $\ell(v)=v$

Step 1: Set $t = 1$.

Step 2: Create a set X as a random permutation of V.

Step 3: For each $x \in X$, x chooses the label with the highest frequency among neighbors. Choose one randomly if more than one label has the maximum frequency

Step 4: Stop the iterations if the label of each node does not change or $t = t_{max}$, otherwise, set $t = t+1$ and go to step (2).

Output: Assign all nodes with the same label into a community then return the final partition.

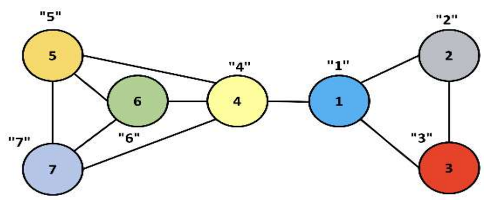
LPA is an efficient and fast algorithm that uses only the network structure to uncover significant communities within different large networks without needing pre-defined objective functions or previous community information. Due to a random choice during its iterations, LPA is non-deterministic, meaning it does not always produce the same results when applied to the same input. Thus, various methods have been suggested to enhance the stability and efficiency of the standard LPA algorithm. The popular approaches using LPA are briefly summarized in Table 3. (A more comprehensive study can be found in [67]).

Table 3. Summary Of Label Propagation-based Community Detection Methods. This Table is Adapted from [68]

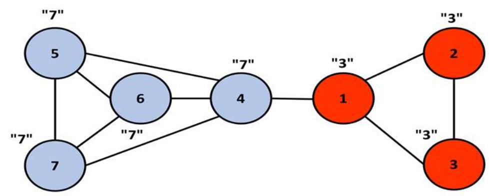
Year and Authors	Method name	Idea	Overlapping	Network type
(Barber and Clark 2009) [69]	LPAm	Find the maxima of an objective function.	✗	simple
(Liu and Murata 2010) [70]	LPAm+	Multistep greedy agglomerative algorithm	✗	
(Lou, Li, and Zhao 2013) [71]	LPA-CNP	Weighted coherent neighborhood propinquity	✗	
(X.-K. Zhang et al. 2014) [72]	LPAc	Edge clustering coefficient	✗	
(Xing et al. 2014) [73]	NIBLPA	Node influence +k-shell	✗	
(X.-K. Zhang et al. 2015) [74]	LPALC	The nearest neighbor with a local cycle	✓	
(X.-K. Zhang et al. 2016) [75]	RWLPA	Similarity matrix using random walks	✗	
(X.-K. Zhang et al. 2017) [76]	LPA_NI	Node importance and label influence	✗	
(Francisquini, Rosset, and Nascimento 2017) [77]	GA-LP	Genetic-based algorithm	✗	Simple and directed
(Chen et al. 2017) [78]	LPA-E	Mutual information of direct and indirect neighbors	✓	simple
(Berahmand and Bouyer 2018) [79]	LP-LPA	Link influence	✗	
(Kong et al. 2018) [80]	LPA-INTIM	The intimacy between nodes	✗	
(Joghan, Bagheri, and Azad 2019) [81]	WLPA-LEB	Local Edge-Betweenness	✗	
(Song et al. 2019) [82]	{SAL, SOR, JAC, SOR, HDI, HPI}-LPA	Node similarity based on local information	✗	
(Wang et al. 2020) [83]	NI-LPA	Signal propagation of nodes with Jaccard distance	✗	

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

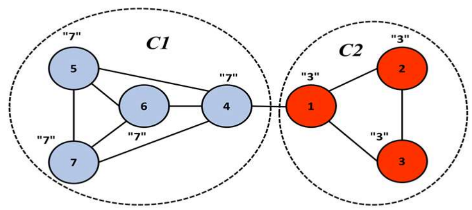
(Xu et al. 2021) [84]	TNS-LPA	Two-level neighborhood similarity	✗	
(Laassem et al. 2022) [85]	LPA_CL	Similarity matrix using Coulomb's law	✗	



(a) The graph with label initialization



(b) Graph after the first iteration



(c) Graph after the final iteration, nodes with the same label are grouped to form a community

Figure 4. A Toy Example Of The Standard Label Propagation Algorithm. Initialization: Each Node Gets A Unique Label Same As The Node Id And Different Labels And Colors Denote The Community Of Nodes. For This Run, The Random Order Chosen Was ⟨6,4,7,2,1,3,5⟩. (b) Iterative Process: It Begins With Node 6 Receiving The Random Label "7", Followed By Node 4 Selecting The Maximum Frequency Label "7", Repeating The Process For The Remaining Nodes. LPA Ends After This First Iteration When All Nodes' Labels Match Those Of Their Neighbors. Finally, In (c), Nodes With The Same Labels Are Grouped To Form Two Communities: C1 {4,5,6,7) And C2 {1,2,3}.

3.4 Spectral Methods

Spectral clustering [86], [87] has gained widespread popularity in recent years due to its simplicity, efficiency, and effectiveness. It is easy to implement and can be solved using standard linear algebra algorithms, and it often performs better than traditional clustering algorithms like k-means. Spectral clustering techniques use the eigenvectors of a similarity matrix (i.e., Adjacency, Laplacian) derived from the graph to perform dimensionality reduction, projecting nodes into a lower-dimensional space in which it can be more easily clustered. The general procedure for spectral clustering has four key steps, as shown in Figure 5.

- ✓ **Pre-processing:** Compute a similarity matrix $S \in \mathbb{R}^{nxn}$ between nodes. This matrix represents the relationships between the nodes in the graph and can be calculated using various measures of similarity or affinity. For example, one way to calculate a similarity matrix is to use the graph's adjacency matrix. Constructing a good affinity matrix is an important step in spectral clustering, as the quality of the affinity matrix can significantly affect the quality of the resulting clusters.
- ✓ **Decomposition:** Compute eigenvalues and corresponding eigenvectors of the matrix representation throughout the Laplacian matrix. Several forms of the Laplacian matrix can be used to represent a graph. The most common forms are [88]: The unnormalized Laplacian matrix $L = D–A$ and the normalized Laplacian matrix $L_n = D^{\frac{-1}{2}}\, L\, D^{\frac{-1}{2}}$. D is the degree matrix, and A is the adjacency matrix of the graph which can be replaced by the computed affinity matrix S. When the number of clusters k is given in advance, we consider only the k largest or smallest eigenvalues. Alternatively, several heuristic methods can be used to determine the optimal number of clusters.
- ✓ **Spectral embedding:** Represent a graph in a low-dimensional space using the selected eigenvectors of the used matrix. The number of eigenvectors used for the projection determines the dimensionality of the final embedding.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- ✓ **Grouping:** Apply traditional clustering algorithms, such as *k-means* or hierarchical clustering, to the embedded data.

Figure 5. Four Key Steps For The Spectral Clustering Process

The choice of a similarity function significantly impacts the effectiveness of spectral clustering. Thus, instead of using the network affinity matrix as the node similarity representation, researchers often use other similarity indexes to compute a new similarity matrix that encodes rich topological and structural information.

Spectral clustering is a powerful technique for community detection, but it also has several disadvantages that should be considered. One of the main limitations of spectral clustering is that it is sensitive to noise and outliers, which can lead to poor performance and inaccurate results. Additionally, spectral clustering requires a large amount of computational resources, especially when computing eigenvalues and eigenvectors, making it challenging to apply to larger networks. Furthermore, spectral clustering requires a good initialization of the number of clusters which can be difficult to achieve. Therefore, various authors have suggested heuristic approaches for determining the optimal number of clusters.

Finally, a summary of different spectral approaches and their key characteristics are presented in Table 4. , including the main idea, the similarity matrix employed, and the network type used in the experimentation.

3.5 Metaheuristic Algorithms

Metaheuristic algorithms are approximate approaches that can find satisfactory solutions over an acceptable time instead of finding the optimal solution. Many community detection algorithms use metaheuristic algorithms to minimize computational complexity. These algorithms will be able to optimize an objective function, often utilizing metrics like modularity, with appropriate accuracy within an acceptable time. Population-based metaheuristic optimization methods, such as Genetic Algorithm (GA) [89], Ant Colony Optimization (ACO) [90], Simulated Annealing (SA)[91], Tabu Search (TS) [92], and Particle Swarm Optimization (PSO) [93], have been recently applied in the context of community detection. Among these, GA stand out as the most developed and widely utilized algorithms in this category.

Identifying communities can be viewed as searching for the approximately optimum solutions to combinatorial optimization problems (COP). A genetic algorithm often used as a search technique in COP is a computer simulation of the natural evolution of the biological system, including the concepts of mutation, natural selection, inheritance, and crossover. A comprehensive review, of definitions, and concepts can be found in [94].

Generally, a genetic algorithm starts from a set of candidate solutions to a problem, which are described as chromosomes, and an objective function to be optimized on search space (or solution space). The objective function (called fitness function) plays an important role in the evolution process of genetic algorithms, and many optimization objectives have been proposed. In [95], Tasgin, Herdagdelen, and Bingol used modularity as a fitness function, and the initial generation is constituted by n genes and the value assigned to the i^{th} gene is the community identifier of node i. They chose to accelerate the convergence of *GA* by transferring a community identifier of random nodes (limited at 40% of n) to its neighbors and then using a non-standard one-way crossover operation in which, given two individuals A and B, a community identifier i is chosen at random from A, and i is assigned to the same nodes of B. Firat, Chatterjee, and Yilmaz in [96] integrate the random walk in a genetic algorithm using the k-medoids representation in which each cluster center is represented with one of the network nodes. The number of communities is initially given, and the fitness function tries to minimize the sum of all pair-wise distances between nodes. Pizzuti presented GA-Net [97] a well-known effective algorithm coded in MATLAB for detecting communities on small networks. This approach introduces two new concepts. The first is the "community score", which measures the quality of partitioning the network into communities. The second is the notion of a "safe individual" used to reduce calculation time. Instead of modularity, Guoqiang and Xiaofang [98] utilized the modularity density in a genetic algorithm. Furthermore, Jin et al. [99] proposed a local search-based genetic algorithm (*GALS*) that uses modularity as an objective function and locus-based adjacency (LAR) presentation. Similarly, in [100], Shi et al. proposed another *GA* that employs modularity as a fitness function and LAR as a genetic representation.

PSO is inspired by the social behavior of bird flocking or fish schooling, where individuals adjust

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

their movements based on personal experience and the experiences of their peers. In the context of community detection, PSO approaches [116], [117], [118], [119], [120], [121], conceptualize potential solutions (communities) as particles in a multidimensional space. These particles iteratively adjust their positions influenced by both individual and group insights, aiming to optimize a quality function.

Table 4. Summary Of Spectral Community Detection Methods.

Reference	Approach	Community number	Network Type used in the experiments	Used matrix	Overlapping
[53] (Donetti and Muñoz 2004)	Hierarchical clustering with modularity optimization in the eigenvector space.	Input	Simple	*L=D-A*	Yes
[101] (Zhang and You 2011)	Constructing a random walk matrix for the graph, which captures the probability of transitioning between nodes in the graph.	Input	Simple	A modified version of the Laplacian matrix called the stochastic matrix: *P = D - W*	No
[102]	Study the spectrum of the adjacency matrix	Obtained from the spectrum of the adjacency matrix	Direct and undirected graphs	Adjacency matrix	No
[61] (Newman 2006)	Spectral optimization of modularity	Input	Simple	The modularity matrix: $B_{ij} = A_{ij} - \frac{k_i k_j}{2m}$ Where the strength of node *i* is: $k_i = \sum_j A_{ij}$	No
[103] (Shen et al. 2010)	Based on the modularity matrix, authors proposed a method for analyzing the multiscale community structure of networks using the concept of covariance and the correlation matrix	Input	Simple	Correlation matrix: $C_{ij} = \frac{B_{ij}}{\sqrt{k_i - \frac{k_i^2}{2m}} - \sqrt{k_j - \frac{k_j^2}{2m}}}$	No
[104] (Capocci et al. 2005)	Spectral method to find communities in a weighted directed graph	Input	Weighted directed	$Q^{-1}ww^T$ Where *W* is the weight matrix and $q_{ij} = \delta_{ij}\sum_{l j=1}^{n} w_{il}w_{jl}$	No
[105] (Xie et al. 2009)	Replaced the adjacency in the normalized Laplacian with the proposed similarity matrix called shared nearest neighbor (SNN).	Modularity maximization	Simple	$K^{-1}S'$ where: $S'_{ij} \begin{cases} S_{ij}A_{ij}/k_i, i \neq j \\ 0, i = j \end{cases}$ S_{ij} shared neighbor numbers of vertices *i* and *j*, *K* the diagonal matrix of S'	No
[106] (Chaudhuri et al. 2012)	The authors introduced a modified version of the graph Laplacian called the degree-corrected graph Laplacian, which	Input	Simple	The degree-corrected Random-walk Laplacian is defined as: $\mathcal{L} = I - (D+\tau I)^{-1/2} A (D\tau I)^{-1/2}$ Where D is a diagonal matrix of degrees and τ is a tuning parameter.	No

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	incorporates a regularization term into the diagonal matrix of degrees.				
[107], [108] (Newman 2006; Leicht and Newman 2008):	The method involves repeatedly bisecting the network by calculating the eigenvector corresponding to the largest positive eigenvalue of the used matrix. The communities are then assigned based on the signs of the elements of this eigenvector.	Modularity maximization	Directed	$B + B^T$ B: the modularity matrix	No
[109] (Krzakala et al. 2013)	investigated the spectrum of the non-backtracking matrix of a sparse network.	Obtained from the spectrum	Directed	$B\big((u,v),(x,y)\big) = \begin{cases} 1 \; if \; v = x \; and \; u \neq y \\ 0 \; otherwise \end{cases}$ Where (u, v) and (x, y) are edges set between two nodes	No
[110] (Li et al. 2015)	A local spectral analysis of the subgraphs formed by the random walk diffusion starting from the known seed nodes.	Input	Simple	The normalized Laplacian matrix using subgraphs around the interest nodes.	Yes
[111] (Fanuel et al. 2017)	The proposed method is inspired by quantum physics	Estimated using the magnetic eigenmaps	Directed	Normalized magnetic Laplacian	No
[112] (Xu 2020)	Defined the communicability modularity and showed that it can be maximized by finding the leading eigenvector of the communicability modularity matrix.	Input	Simple	Communicability modularity matrix	No
[113] (Xu et al. 2018)	Particle swarm optimization and Simulated Annealing are applied to obtain the global optimal solution using an eigenvector space constructed from a similarity matrix merging attribute and relationship information of nodes.	Input	Attributed network	Laplacian matrix using a similarity matrix merging attribute and relationship information of nodes.	No
[114] (Hu et al. 2020)	Used the learned node embeddings to perform spectral clustering	input	Simple	A normalized Laplacian with a similarity matrix generated using node2vec embeddings	No
[115] (Laassem et al. 2022)	By considering the network as a set of charged particles, the authors	Estimated using the second largest	Simple	A normalized Laplacian with a similarity matrix M was generated using Coulomb's law.	No

	computed similarities between nodes using Coulomb's law and then performed spectral clustering.	eigengap		$M_{i,j} = \begin{cases} \frac{k_i * k_j}{r^2}, & if\ i,j\ are\ connected \\ 0, & otherwise \end{cases}$ Where r is the shortest path between i and j	

Ant Colony Optimization (ACO) draws its inspiration from the foraging behavior of ants, where they use pheromone trails to communicate and navigate towards food sources. In the context of community detection, numerous ACO-based methods have been developed [122], [123], [124], [125], [126], [127], [128]. These approaches simulate artificial ants traversing the network iteratively to identify communities within the network structure.

The simulated annealing approach is inspired by the process of cooling molten metals. The gradual cooling process helps converge to a globally optimal solution by adjusting the temperature and exploring the search space. SA strategy is used in some CD methods [129], [130], [131], [132] to maximize the modularity of a network and escape from local minima.

In community detection, Tabu Search methods iteratively explore neighboring solutions, employing a memory structure called the tabu list to direct the search process and ensure avoidance of revisiting previously examined solutions. Currently, there are fewer Tabu search-based methods developed for CD [133], [134], [135], indicating a gap in research within this area. Further efforts and research are needed to enhance and refine this approach.

3.6 Other Approaches

3.6.1 Clique-based methods:

In graph theory, a clique is a complete subgraph, i.e., every two distinct nodes are adjacent. Moreover, a k-clique [136] community is defined as a union of all k-cliques (cliques containing k vertices) that can be reached from each other through a series of adjacent k-cliques (where adjacency means sharing k-1 nodes). By this, Palla et al. [137] proposed an overlapping community detection algorithm called: The clique percolation method (CPM). The numerical determination of a set of all *k-cliques* of an undirected graph is a polynomial problem [138], but CPM uses an exponential algorithm *O (exp (n))*. The method is based on first locating all cliques (maximal complete subgraphs) of the network and then identifying the communities by analyzing the clique–clique overlap matrix. The main advantage of this approach is the possibility to "zoom" in on a single unit in a network and uncover small communities and the communities connected to these by changing the value of k related to finding the *k-clique*. Based on this approach, Lehmann et al. in [139], developed a method for detecting communities in bipartite networks. They showed that incorporating the bipartite information in the process of community detection reveals a good community structure.

Shen et al. used maximal clique and modularity in an agglomerative approach and developed a method called EAGLE [140] to detect the overlapping and hierarchical community structure.

3.6.2 Random walk - based methods:

To solve the problem of finding communities in a large network, Pons and Latapy [55] proposed a new measure of similarity between nodes based on random walks and developed a popular method known as Walktrap. A distance between nodes is computed from the information given by random walks in the graph, and it is related to the probability that a random walker moves from one node to another in a fixed number of steps t. the main idea of this approach is that if nodes are in the same community, this distance tends to be smaller and large otherwise. The result of the clustering process after the *n-1* step is represented by a dendrogram, and the modularity of GN is chosen to evaluate the best partition. This method has a high computational complexity; therefore, a prior parameter is introduced to deal with large networks. Similarly, Zhou and Lipowsky introduced a new proximity index and developed an agglomerative clustering procedure called Netwalk [54] based on random walks.

Rosvall and Bergstrom proposed a method known as Infomap [141] based on the concept of information flow on networks, such that a group of nodes among which information flows quickly and easily can be aggregated and form a single well-connected module or community. This is realized using random walks in the entire network as a core for the information flow; to minimize the description length, these random walks are described using a two-level Huffman compression code.

In 2016, Hurley and Duriakova [142] proposed a method for community detection that combines the Infomap method with block modeling. The proposed method used a model of random walks between groups or blocks to identify a particular mesoscale structure. Toth et al. proposed *Synwalk* [143], in comparison to *Infomap* and *Walktrap*, *Synwalk* predicts communities in undirected networks by using candidate random walks.

Self-avoiding random walks (SAW) is a walk in which a walker does not visit the same node more than once. Based on SAW, Guzzi Bagnato et al. in [144] proposed a method that reveals community structure with high modularity scores.

4. OPEN ISSUES AND CHALLENGES

The concept of community discovery is relatively mature. However, is still of continuing interest to researchers and practitioners due to the number of challenges that have yet to be overcome. Some examples are briefly discussed below:

- ✓ **Scalability:** Due to the rapid development and evolution of social networks, classical algorithms of CD can become computationally expensive to run on these networks. For instance, Facebook has billions of users who log in every month. Therefore, developing a highly efficient community detection algorithm is one of the grand challenges in large networks. Using parallel processing techniques with efficient memory management for these new algorithms is important.
- ✓ **Overlapping communities:** Many real-world networks have overlapping communities, where nodes belong to multiple communities at once. For example, one user may contribute to various Facebook groups, and users with many common interests can form multiple overlapping communities. Furthermore, there are many applications whose data does not always cluster in disjoint ways, such as document classification or medical diagnosis [145]. Developing an overlapping CD method is an even more challenging problem since, beyond the used definition for a community, the structure of the overlapping region must also be considered.
- ✓ **Validation and evaluation:** In the context of evaluating and comparing the performance and accuracy of community detection methods, several datasets (real and artificial) and measures have been used. The most used artificial benchmarks are GN (Girvan and Newman) [3] and LFR (Lancichinetti, Fortunato, Radicchi) [146]. However, recognizing the constraints inherent in the LFR benchmark, particularly its scalability issues and lack of robust theoretical foundations, researchers have introduced the Artificial Benchmark for Community Detection graph (ABCD) [147]. The model and its variants are presented as a random graph with a power-law distribution for the degree distribution and community size distributions. There is a real need for designing new benchmarks that accurately approximate a real-life network with options to generate directed and weighted networks as well as overlapping community structures. Additionally, real datasets suffer from several issues, such as the lack of ground truth and its correlation with metadata, and the absence of standardization related to the size or types of networks. On the other hand, there are two types of measures used to evaluate the quality of detected community structures [148]: internal ones (Using only network topology: the most used one is Modularity) and external ones (compare results with a ground truth: The most used is NMI). Finally, several problems with evaluation metrics have been identified [149]. There is a need to design new metrics combining internal and external measures with the ability to evaluate overlapping and hierarchical community structures.
- ✓ **Resolution limit:** Some algorithms have a resolution limit, meaning that they may not be able to detect small communities within a large network.
- ✓ **Definition of community:** The definition of a community can vary depending on the context, making it difficult to design algorithms that are applicable to a wide range of networks.
- ✓ **Community Evolution:** Many real-world networks evolve over time as nodes and edges are added or removed, and it is challenging to design algorithms that can track the changes in community structure.
- ✓ **The optimal number of communities**: A fundamental problem in cluster analysis is determining the optimal number of clusters a priori. In many existing clustering algorithms, the number of clusters is given as input or estimated using statistics or metaheuristics methods. The a priori determination of this parameter is known as automatic clustering, which has continued to attract more attention from researchers.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Although many algorithms are developed each year to tackle the challenges of community detection, it is still necessary to improve their accuracy, efficiency, and adapt them for different graph structures.

5. CONCLUSIONS

In conclusion, community detection is a vital area of research for understanding the structure and dynamics of networks. In this survey paper, we have summarized many of the important works in the field with a focus on highly cited papers. We have discussed various community detection methods, including traditional and recent approaches. Each method has its own advantages and limitations, which are suitable for different types of networks. Moreover, we introduce the applications of CD and present the current problems faced in this field to explore possible opportunities for future work.

While this survey offers a solid foundation, the vast amount of new research necessitates a more in-depth analysis of recent advancements. Additionally, with the emergence of big data, advanced algorithms, powerful hardware accelerators, and parallel processing, further research is needed to develop more accurate and efficient community detection algorithms capable of addressing the complexities of dynamic networks and large-scale big data networks. Overall, this survey will help researchers to better understand the community detection filed and the state-of-the-art algorithms.

REFERENCES:

[1] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” vol. 393, p. 3, 1998.

[2] M. Coscia, F. Giannotti, and D. Pedreschi, “A classification for community discovery methods in complex networks,” *Statistical Analy Data Mining*, vol. 4, no. 5, pp. 512–546, Oct. 2011, doi: 10.1002/sam.10133.

[3] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002, doi: 10.1073/pnas.122653799.

[4] C.-Y. Liu, C. Zhou, J. Wu, Y. Hu, and L. Guo, “Social recommendation with an essential preference space,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

[5] D. Deshmuk and D. Ingle, “A community detection and recommendation system,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 7, pp. 752–757, 2017.

[6] T. Sangkaran, A. Abdullah, and N. Jhanjhi, “Criminal community detection based on isomorphic subgraph analytics,” *Open Computer Science*, vol. 10, no. 1, pp. 164–174, 2020.

[7] T. Waskiewicz, “Friend of a Friend Influence in Terrorist Social Networks,” p. 5.

[8] K. K. Mkhitaryan, “Realization of Recommender Framework Based on Community Detection,” *Mathematical Problems of Computer Science*, vol. 51, pp. 57–65, 2019, doi: 10.51408/1963-0033.

[9] D. Lalwani, D. V. L. N. Somayajulu, and P. R. Krishna, “A community driven social recommendation system,” in *Proceedings of the 2015 IEEE International Conference on Big Data (Big Data)*, in BIG DATA ’15. USA: IEEE Computer Society, Oct. 2015, pp. 821–826. doi: 10.1109/BigData.2015.7363828.

[10] R. N. Mantegna, “Hierarchical structure in financial markets,” *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 11, no. 1, pp. 193–197, Sep. 1999, doi: 10.1007/s100510050929.

[11] V. Satuluri, S. Parthasarathy, and D. Ucar, “Markov clustering of protein interaction networks with improved balance and scalability,” in *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology*, in BCB ’10. New York, NY, USA: Association for Computing Machinery, Aug. 2010, pp. 247–256. doi: 10.1145/1854776.1854812.

[12] S. Brohée and J. van Helden, “Evaluation of clustering algorithms for protein-protein interaction networks,” *BMC Bioinformatics*, vol. 7, no. 1, p. 488, Nov. 2006, doi: 10.1186/1471-2105-7-488.

[13] N. Haq and Z. J. Wang, “Community detection from genomic datasets across human cancers,” in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2016, pp. 1147–1150. doi: 10.1109/GlobalSIP.2016.7906021.

[14] L. Lü and T. Zhou, “Link prediction in complex networks: A survey,” *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 6, pp. 1150–1170, 2011, doi: https://doi.org/10.1016/j.physa.2010.11.027.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[15] T. Yi, S. Zhang, Z. Bu, J. Du, and C. Fang, "Link prediction based on higher-order structure extraction and autoencoder learning in directed networks," *Knowledge-Based Systems*, vol. 241, p. 108241, 2022, doi: https://doi.org/10.1016/j.knosys.2022.108241 .

[16] M. Salathé and J. H. Jones, "Dynamics and Control of Diseases in Networks with Community Structure," *PLOS Computational Biology*, vol. 6, no. 4, p. e1000736, Apr. 2010, doi: 10.1371/journal.pcbi.1000736.

[17] C. Stegehuis, R. van der Hofstad, and J. S. H. van Leeuwaarden, "Epidemic spreading on complex networks with community structures," *Sci Rep*, vol. 6, no. 1, Art. no. 1, Jul. 2016, doi: 10.1038/srep29748.

[18] J. Shang, L. Liu, F. Xie, and C. Wu, "How Overlapping Community Structure Affects Epidemic Spreading in Complex Networks," in *2014 IEEE 38th International Computer Software and Applications Conference Workshops*, Vasteras, Sweden: IEEE, Jul. 2014, pp. 240–245. doi: 10.1109/COMPSACW.2014.43.

[19] L. Akoglu, H. Tong, and D. Koutra, "Graph-based Anomaly Detection and Description: A Survey." arXiv, Apr. 28, 2014. Accessed: Feb. 03, 2024. [Online]. Available: http://arxiv.org/abs/1404.4679

[20] K. Anand, J. Kumar, and K. Anand, "Anomaly Detection in Online Social Network: A Survey," 2017.

[21] H. Safdari and C. De Bacco, "Anomaly detection and community detection in networks," *J Big Data*, vol. 9, no. 1, p. 122, Dec. 2022, doi: 10.1186/s40537-022-00669-1.

[22] M. A. Prado-Romero and A. Gago-Alonso, "Community Feature Selection for Anomaly Detection in Attributed Graphs," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, vol. 10125, C. Beltrán-Castañón, I. Nyström, and F. Famili, Eds., in Lecture Notes in Computer Science, vol. 10125. , Cham: Springer International Publishing, 2017, pp. 109–116. doi: 10.1007/978-3-319-52277-7_14.

[23] Z. Chen, W. Hendrix, and N. F. Samatova, "Community-based anomaly detection in evolutionary networks," *J Intell Inf Syst*, vol. 39, no. 1, pp. 59–85, Aug. 2012, doi: 10.1007/s10844-011-0183-2.

[24] Q. Ding, N. Katenka, P. Barford, E. Kolaczyk, and M. Crovella, "Intrusion as (anti)social communication: characterization and detection".

[25] R. Santamaría and R. Therón, "Overlapping Clustered Graphs: Co-authorship Networks Visualization," in *Smart Graphics*, A. Butz, B. Fisher, A. Krüger, P. Olivier, and M. Christie, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 190–199.

[26] A. Lambert, F. Queyroi, and R. Bourqui, "Visualizing patterns in node-link diagrams," in *2012 16th International Conference on Information Visualisation*, IEEE, 2012, pp. 48–53.

[27] Z. Yang, R. Algesheimer, and C. J. Tessone, "A Comparative Analysis of Community Detection Algorithms on Artificial Networks," *Sci Rep*, vol. 6, no. 1, p. 30750, Aug. 2016, doi: 10.1038/srep30750.

[28] D. Jin, Z. Yu, P. Jiao, S. Pan, P. S. Yu, and W. Zhang, "A Survey of Community Detection Approaches: From Statistical Modeling to Deep Learning," *arXiv:2101.01669 [physics]*, Jan. 2021, Accessed: May 20, 2021. [Online]. Available: http://arxiv.org/abs/2101.01669

[29] P. Bedi and C. Sharma, "Community detection in social networks," vol. 6, p. 21, 2016.

[30] B. S. Khan and M. A. Niazi, "Network Community Detection: A Review and Visual Survey," p. 39, 2017.

[31] A. Amelio and C. Pizzuti, "Overlapping Community Discovery Methods: A Survey," *arXiv:1411.3935 [physics]*, 2014, doi: 10.1007/978-3-7091-1797-2.

[32] M. Plantié and M. Crampes, "Survey on Social Community Detection," in *Social Media Retrieval*, N. Ramzan, R. van Zwol, J.-S. Lee, K. Clüver, and X.-S. Hua, Eds., in Computer Communications and Networks. , London: Springer London, 2013, pp. 65–85. doi: 10.1007/978-1-4471-4555-4_4.

[33] A. Drif and A. Boukerram, "Taxonomy and Survey Of Community Discovery Methods in Complex Networks," *IJCSES*, vol. 5, no. 4, pp. 1–19, Aug. 2014, doi: 10.5121/ijcses.2014.5401.

[34] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics Reports*, vol. 659, pp. 1–44, Nov. 2016, doi: 10.1016/j.physrep.2016.09.002.

[35] M. E. J. Newman, "Mixing patterns in networks," *Phys. Rev. E*, vol. 67, no. 2, p. 026126, Feb. 2003, doi: 10.1103/PhysRevE.67.026126.

[36] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in

networks," *Phys. Rev. E*, vol. 69, no. 2, p. 026113, Feb. 2004, doi: 10.1103/PhysRevE.69.026113.
[37] J.M. Anthonisse, "The rush in a directed graph," *Stichting Mathematisch Centrum. Mathematische Besliskunde*, no. BN 9/71. Stichting Mathematisch Centrum, Jan. 01, 1971. Accessed: Jun. 05, 2021. [Online]. Available: https://ir.cwi.nl/pub/9791
[38] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978, doi: https://doi.org/10.1016/0378-8733(78)90021-7.
[39] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, Feb. 2010, doi: 10.1016/j.physrep.2009.11.002.
[40] S. Fortunato, V. Latora, and M. Marchiori, "Method to find community structures based on information centrality," *Phys. Rev. E*, vol. 70, no. 5, p. 056104, Nov. 2004, doi: 10.1103/PhysRevE.70.056104.
[41] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, "Defining and identifying communities in networks," *Proceedings of the National Academy of Sciences*, vol. 101, no. 9, pp. 2658–2663, Mar. 2004, doi: 10.1073/pnas.0400054101.
[42] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization," *Phys. Rev. E*, vol. 72, no. 2, p. 027104, Aug. 2005, doi: 10.1103/PhysRevE.72.027104.
[43] S. Boettcher and A. G. Percus, "Optimization with Extremal Dynamics," *Phys. Rev. Lett.*, vol. 86, no. 23, pp. 5211–5214, Jun. 2001, doi: 10.1103/PhysRevLett.86.5211.
[44] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E*, vol. 69, no. 6, p. 066133, Jun. 2004, doi: 10.1103/PhysRevE.69.066133.
[45] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *PHYSICAL REVIEW E*, p. 6, 2004.
[46] K. Wakita and T. Tsurumi, "Finding Community Structure in Mega-scale Social Networks," *arXiv:cs/0702048*, Feb. 2007, Accessed: Jun. 06, 2021. [Online]. Available: http://arxiv.org/abs/cs/0702048
[47] J. Bagrow and E. Bollt, "A Local Method for Detecting Communities," *arXiv:cond-mat/0412482*, Mar. 2005, doi: 10.1103/PhysRevE.72.046108.
[48] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech.*, vol. 2008, no. 10, p. P10008, Oct. 2008, doi: 10.1088/1742-5468/2008/10/P10008.
[49] V. A. Traag, L. Waltman, and N. J. Van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," *Sci Rep*, vol. 9, no. 1, p. 5233, Mar. 2019, doi: 10.1038/s41598-019-41695-z.
[50] T. Zhou, L. Lu, and Y.-C. Zhang, "Predicting Missing Links via Local Information," *Eur. Phys. J. B*, vol. 71, no. 4, pp. 623–630, Oct. 2009, doi: 10.1140/EPJB/E2009-00335-8.
[51] Y. Pan, D.-H. Li, J.-G. Liu, and J.-Z. Liang, "Detecting community structure in complex networks via node similarity," *Physica A*, p. 9, 2010.
[52] H. Yang *et al.*, "A Node Similarity and Community Link Strength-Based Community Discovery Algorithm," *Complexity*, vol. 2021, pp. 1–17, Mar. 2021, doi: 10.1155/2021/8848566.
[53] L. Donetti and M. A. Muñoz, "Detecting network communities: a new systematic and efficient algorithm," *J. Stat. Mech.: Theor. Exp.*, vol. 2004, no. 10, p. P10012, Oct. 2004, doi: 10.1088/1742-5468/2004/10/P10012.
[54] H. Zhou and R. Lipowsky, "Network Brownian Motion: A New Method to Measure Vertex-Vertex Proximity and to Identify Communities and Subcommunities," in *Computational Science - ICCS 2004*, vol. 3038, M. Bubak, G. D. van Albada, P. M. A. Sloot, and J. Dongarra, Eds., in Lecture Notes in Computer Science, vol. 3038. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 1062–1069. doi: 10.1007/978-3-540-24688-6_137.
[55] P. Pons and M. Latapy, "Computing Communities in Large Networks Using Random Walks," p. 10.
[56] B. Saoud and A. Moussaoui, "Node similarity and modularity for finding communities in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 492, pp. 1958–1966, Feb. 2018, doi: 10.1016/j.physa.2017.11.110.
[57] U. Brandes *et al.*, "On Modularity Clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 2, pp. 172–188, Feb. 2008, doi: 10.1109/TKDE.2007.190689.
[58] M. Chen, K. Kuzmin, and B. K. Szymanski, "Community Detection via Maximization of

Modularity and Its Variants," *IEEE Trans. Comput. Soc. Syst.*, vol. 1, no. 1, pp. 46–65, Mar. 2014, doi: 10.1109/TCSS.2014.2307458.

[59] A. Medus, G. Acuña, and C. O. Dorso, "Detection of community structures in networks via global optimization," *Physica A: Statistical Mechanics and its Applications*, vol. 358, no. 2–4, pp. 593–604, Dec. 2005, doi: 10.1016/j.physa.2005.04.022.

[60] R. Guimerà and L. A. Nunes Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, no. 7028, pp. 895–900, Feb. 2005, doi: 10.1038/nature03288.

[61] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E*, vol. 74, no. 3, p. 036104, Sep. 2006, doi: 10.1103/PhysRevE.74.036104.

[62] M. E. J. Newman, "Spectral methods for network community detection and graph partitioning," *Phys. Rev. E*, vol. 88, no. 4, p. 042822, Oct. 2013, doi: 10.1103/PhysRevE.88.042822.

[63] M. Sales-Pardo, R. Guimera, A. A. Moreira, and L. A. N. Amaral, "Extracting the hierarchical organization of complex systems," *Proceedings of the National Academy of Sciences*, vol. 104, no. 39, pp. 15224–15229, Sep. 2007, doi: 10.1073/pnas.0703740104.

[64] G. Agarwal and D. Kempe, "Modularity-maximizing graph communities via mathematical programming," *Eur. Phys. J. B*, vol. 66, no. 3, pp. 409–418, Dec. 2008, doi: 10.1140/epjb/e2008-00425-1.

[65] S. Fortunato and M. Barthelemy, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences*, vol. 104, no. 1, pp. 36–41, Jan. 2007, doi: 10.1073/pnas.0605965104.

[66] 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. 036106, Sep. 2007, doi: 10.1103/PhysRevE.76.036106.

[67] S. E. Garza and S. E. Schaeffer, "Community detection with the Label Propagation Algorithm: A survey," *Physica A: Statistical Mechanics and its Applications*, vol. 534, p. 122058, Nov. 2019, doi: 10.1016/j.physa.2019.122058.

[68] L. Brahim, B. Loubna, and I. Ali, "A Literature Survey on Label Propagation for Community Detection," in *2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS)*, Fez, Morocco: IEEE, Oct. 2021, pp. 1–7. doi: 10.1109/ICDS53782.2021.9626716.

[69] M. J. Barber and J. W. Clark, "Detecting network communities by propagating labels under constraints," *Phys. Rev. E*, vol. 80, no. 2, p. 026129, Aug. 2009, doi: 10.1103/PhysRevE.80.026129.

[70] X. Liu and T. Murata, "Advanced modularity-specialized label propagation algorithm for detecting communities in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 7, pp. 1493–1500, Apr. 2010, doi: 10.1016/j.physa.2009.12.019.

[71] H. Lou, S. Li, and Y. Zhao, "Detecting community structure using label propagation with weighted coherent neighborhood propinquity," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 14, pp. 3095–3105, Jul. 2013, doi: 10.1016/j.physa.2013.03.014.

[72] X.-K. Zhang, X. Tian, Y.-N. Li, and C. Song, "Label propagation algorithm based on edge clustering coefficient for community detection in complex networks," *Int. J. Mod. Phys. B*, vol. 28, no. 30, p. 1450216, Dec. 2014, doi: 10.1142/S0217979214502166.

[73] Y. Xing, F. Meng, Y. Zhou, M. Zhu, M. Shi, and G. Sun, "A Node Influence Based Label Propagation Algorithm for Community Detection in Networks," *The Scientific World Journal*, vol. 2014, pp. 1–13, 2014, doi: 10.1155/2014/627581.

[74] X.-K. Zhang, S. Fei, C. Song, X. Tian, and Y.-Y. Ao, "Label propagation algorithm based on local cycles for community detection," *Int. J. Mod. Phys. B*, vol. 29, no. 05, p. 1550029, Feb. 2015, doi: 10.1142/S0217979215500290.

[75] X.-K. Zhang, C. Song, J. Jia, Z.-L. Lu, and Q. Zhang, "An improved label propagation algorithm based on the similarity matrix using random walk," *Int. J. Mod. Phys. B*, vol. 30, no. 16, p. 1650093, Jun. 2016, doi: 10.1142/S0217979216500934.

[76] X.-K. Zhang, J. Ren, C. Song, J. Jia, and Q. Zhang, "Label propagation algorithm for community detection based on node importance and label influence," *Physics Letters A*, vol. 381, no. 33, pp. 2691–2698,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Sep. 2017, doi: 10.1016/j.physleta.2017.06.018.

[77] 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 Systems with Applications*, vol. 74, pp. 127–138, May 2017, doi: 10.1016/j.eswa.2016.12.039.

[78] N. Chen, Y. Liu, H. Chen, and J. Cheng, "Detecting communities in social networks using label propagation with information entropy," *Physica A: Statistical Mechanics and its Applications*, vol. 471, pp. 788–798, Apr. 2017, doi: 10.1016/j.physa.2016.12.047.

[79] K. Berahmand and A. Bouyer, "LP-LPA: A link influence-based label propagation algorithm for discovering community structures in networks," *Int. J. Mod. Phys. B*, vol. 32, no. 06, p. 1850062, Mar. 2018, doi: 10.1142/S0217979218500625.

[80] H. Kong, Q. Kang, C. Liu, W. Li, H. He, and Y. Kang, "An improved label propagation algorithm based on node intimacy for community detection in networks," *Int. J. Mod. Phys. B*, vol. 32, no. 25, p. 1850279, Oct. 2018, doi: 10.1142/S021797921850279X.

[81] H. S. Joghan, A. Bagheri, and M. Azad, "Weighted Label Propagation Algorithm based on Local Edge Betweenness," p. 29, 2019.

[82] C. Song, G. Huang, B. Yin, B. Zhang, and X. Liu, "Label propagation algorithm based on node similarity driven by local information," *Int. J. Mod. Phys. B*, vol. 33, no. 30, p. 1950363, Dec. 2019, doi: 10.1142/S0217979219503636.

[83] T. Wang, S. Chen, X. Wang, and J. Wang, "Label propagation algorithm based on node importance," *Physica A: Statistical Mechanics and its Applications*, vol. 551, p. 124137, Aug. 2020, doi: 10.1016/j.physa.2020.124137.

[84] G. Xu, J. Guo, and P. Yang, "TNS-LPA: An Improved Label Propagation Algorithm for Community Detection Based on Two-Level Neighbourhood Similarity," *IEEE Access*, vol. 9, pp. 23526–23536, 2021, doi: 10.1109/ACCESS.2020.3045085.

[85] B. Laassem, A. Idarrou, L. Boujlaleb, and M. Iggane, "Label propagation algorithm for community detection based on Coulomb's law," *Physica A: Statistical Mechanics and its Applications*, vol. 593, p. 126881, 2022, doi: https://doi.org/10.1016/j.physa.2022.126881.

[86] H.-W. Shen and X.-Q. Cheng, "Spectral methods for the detection of network community structure: a comparative analysis," p. 14, 2010.

[87] U. von Luxburg, "A tutorial on spectral clustering," *Stat Comput*, vol. 17, no. 4, pp. 395–416, Dec. 2007, doi: 10.1007/s11222-007-9033-z.

[88] F. R. K. Chung, *Spectral graph theory*. in Regional conference series in mathematics, no. no. 92. Providence, R.I: Published for the Conference Board of the mathematical sciences by the American Mathematical Society, 1997.

[89] J. H. Holland, *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. in Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. Oxford, England: U Michigan Press, 1975, pp. viii, 183.

[90] M. Dorigo and T. Stützle, *Ant colony optimization*. in A Bradford book. Cambridge, Mass.: MIT Press, 2004.

[91] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983, doi: 10.1126/science.220.4598.671.

[92] F. Glover, "Tabu Search—Part I," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, Aug. 1989, doi: 10.1287/ijoc.1.3.190.

[93] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, Nov. 1995, pp. 1942–1948 vol.4. doi: 10.1109/ICNN.1995.488968.

[94] C. Pizzuti, "Evolutionary Computation for Community Detection in Networks: a Review," *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, p. 21, 2017.

[95] M. Tasgin, A. Herdagdelen, and H. Bingol, "Community Detection in Complex Networks Using Genetic Algorithms," p. 6, 2007.

[96] A. Firat, S. Chatterjee, and M. Yilmaz, "Genetic clustering of social networks using random walks," *Computational Statistics*, p. 10, 2007.

[97] C. Pizzuti, "GA-Net: A Genetic Algorithm for Community Detection in Social Networks," p. 10, 2008.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[98] C. Guoqiang and G. Xiaofang, "A Genetic Algorithm Based on Modularity Density for Detecting Community Structure in Complex Networks," p. 4, 2010.

[99] D. Jin, D. He, D. Liu, and C. Baquero, "Genetic Algorithm with Local Search for Community Mining in Complex Networks," p. 8, 2010.

[100] C. Shi, Z. Yan, Y. Wang, Y. Cai, and B. Wu, "A GENETIC ALGORITHM FOR DETECTING COMMUNITIES IN LARGE-SCALE COMPLEX NETWORKS," p. 15, 2010.

[101] X. Zhang and Q. You, "An improved spectral clustering algorithm based on random walk," *Front. Comput. Sci. China*, vol. 5, no. 3, pp. 268–278, Sep. 2011, doi: 10.1007/s11704-011-0023-0.

[102] S. Chauhan, M. Girvan, and E. Ott, "Spectral properties of networks with community structure," *Phys. Rev. E*, vol. 80, no. 5, p. 056114, Nov. 2009, doi: 10.1103/PhysRevE.80.056114.

[103] H.-W. Shen, X.-Q. Cheng, and B.-X. Fang, "Covariance, correlation matrix, and the multiscale community structure of networks," *Phys. Rev. E*, vol. 82, no. 1, p. 016114, Jul. 2010, doi: 10.1103/PhysRevE.82.016114.

[104] A. Capocci, V. D. P. Servedio, G. Caldarelli, and F. Colaiori, "Detecting communities in large networks," *Physica A: Statistical Mechanics and its Applications*, vol. 352, no. 2–4, pp. 669–676, Jul. 2005, doi: 10.1016/j.physa.2004.12.050.

[105] F. Xie, M. Ji, Y. Zhang, and D. Huang, "The detection of community structure in network via an improved spectral method," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 15–16, pp. 3268–3272, Aug. 2009, doi: 10.1016/j.physa.2009.04.036.

[106] K. Chaudhuri, F. Chung, and A. Tsiatas, "Spectral Clustering of Graphs with General Degrees in the Extended Planted Partition Model," in *Proceedings of the 25th Annual Conference on Learning Theory*, JMLR Workshop and Conference Proceedings, Jun. 2012, p. 35.1-35.23. Accessed: Apr. 26, 2022. [Online]. Available: https://proceedings.mlr.press/v23/chaudhuri12.html

[107] E. A. Leicht and M. E. J. Newman, "Community Structure in Directed Networks," *Phys. Rev. Lett.*, vol. 100, no. 11, p. 118703, Mar. 2008, doi: 10.1103/PhysRevLett.100.118703.

[108] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103, no. 23, pp. 8577–8582, Jun. 2006, doi: 10.1073/pnas.0601602103.

[109] F. Krzakala *et al.*, "Spectral redemption in clustering sparse networks," *Proceedings of the National Academy of Sciences*, vol. 110, no. 52, pp. 20935–20940, Dec. 2013, doi: 10.1073/pnas.1312486110.

[110] Y. Li, K. He, D. Bindel, and J. Hopcroft, "Uncovering the Small Community Structure in Large Networks: A Local Spectral Approach," *arXiv:1509.07715 [physics]*, Sep. 2015, Accessed: May 26, 2021. [Online]. Available: http://arxiv.org/abs/1509.07715

[111] M. Fanuel, C. M. Alaíz, and J. A. K. Suykens, "Magnetic eigenmaps for community detection in directed networks," *Phys. Rev. E*, vol. 95, no. 2, p. 022302, Feb. 2017, doi: 10.1103/PhysRevE.95.022302.

[112] Y. Xu, "A spectral method to detect community structure based on the communicability modularity," *Physica A: Statistical Mechanics and its Applications*, vol. 537, p. 122751, Jan. 2020, doi: 10.1016/j.physa.2019.122751.

[113] Y. Xu, Z. Zhuang, W. Li, and X. Zhou, "Effective community division based on improved spectral clustering," *Neurocomputing*, vol. 279, pp. 54–62, Mar. 2018, doi: 10.1016/j.neucom.2017.06.085.

[114] F. Hu, J. Liu, L. Li, and J. Liang, "Community detection in complex networks using Node2vec with spectral clustering," *Physica A: Statistical Mechanics and its Applications*, vol. 545, p. 123633, May 2020, doi: 10.1016/j.physa.2019.123633.

[115] B. Laassem, A. Idarrou, L. Boujlaleb, and M. Iggane, "A spectral method to detect community structure based on Coulomb's matrix," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, p. 3, 2023, doi: 10.1007/s13278-022-01010-7.

[116] B. S. Rees and K. B. Gallagher, "Overlapping community detection using a community optimized graph swarm," *Social Network Analysis and Mining*, vol. 2, no. 4, pp. 405–417, Dec. 2012, doi: 10.1007/s13278-012-0050-3.

[117] X. Li, X. Wu, S. Xu, S. Qing, and P.-C. Chang, "A novel complex network community detection approach using discrete particle swarm optimization with particle diversity and mutation," *Applied Soft*

Computing, vol. 81, p. 105476, Aug. 2019, doi: 10.1016/j.asoc.2019.05.003.

[118] S. Rahimi, A. Abdollahpouri, and P. Moradi, "A multi-objective particle swarm optimization algorithm for community detection in complex networks," *Swarm and Evolutionary Computation*, vol. 39, pp. 297–309, Apr. 2018, doi: 10.1016/j.swevo.2017.10.009.

[119] C. Zhang, X. Hei, D. Yang, and L. Wang, "A Memetic Particle Swarm Optimization Algorithm for Community Detection in Complex Networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 30, no. 2, 2016, doi: 10.1142/S0218001416590035.

[120] Q. Cai, M. Gong, L. Ma, S. Ruan, F. Yuan, and L. Jiao, "Greedy discrete particle swarm optimization for large-scale social network clustering," *Information Sciences*, vol. 316, pp. 503–516, Sep. 2015, doi: 10.1016/j.ins.2014.09.041.

[121] Y. Chen and X. Qiu, "Detecting Community Structures in Social Networks with Particle Swarm Optimization," in *Frontiers in Internet Technologies*, J. Su, B. Zhao, Z. Sun, X. Wang, F. Wang, and K. Xu, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 266–275.

[122] J. Zhao, X. You, Q. Duan, and S. Liu, "Multiple Ant Colony Algorithm Combining Community Relationship Network," *Arabian Journal for Science and Engineering*, vol. 47, no. 8, pp. 10531–10546, Feb. 2022, doi: 10.1007/s13369-022-06579-x.

[123] C. Mu, J. Zhang, Y. Liu, R. Qu, and T. Huang, "Multi-objective ant colony optimization algorithm based on decomposition for community detection in complex networks," *Soft Computing*, vol. 23, no. 23, pp. 12683–12709, Feb. 2019, doi: 10.1007/s00500-019-03820-y.

[124] C. J. Hou and M. Z. B. M. Kamali, "An ant colony approach in the detection of communities in complex networks," in *AIP Conference Proceedings*, AIP Publishing, 2019. doi: 10.1063/1.5136488.

[125] X. Zhou, Y. Liu, J. Zhang, T. Liu, and D. Zhang, "An ant colony based algorithm for overlapping community detection in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 427, pp. 289–301, Jun. 2015, doi: 10.1016/j.physa.2015.02.020.

[126] Chang Honghao, Feng Zuren, and Ren Zhigang, "Community detection using Ant Colony Optimization," in *2013 IEEE Congress on Evolutionary Computation*, Cancun, Mexico: IEEE, Jun. 2013, pp. 3072–3078. doi: 10.1109/CEC.2013.6557944.

[127] D. He, J. Liu, D. Liu, D. Jin, and Z. Jia, "Ant colony optimization for community detection in large-scale complex networks," in *2011 Seventh International Conference on Natural Computation*, IEEE, 2011, pp. 1151–1155.

[128] S. Sadi, S. Etaner-Uyar, and Ş. Gündüz-Öğüdücü, "Community detection using ant colony optimization techniques," in *15th International conference on soft computing*, Citeseer, 2009, pp. 206–213.

[129] J. Liu and T. Liu, "Detecting community structure in complex networks using simulated annealing with k<math><mi is="true">k</mi></math>-means algorithms," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 11, pp. 2300–2309, Jun. 2010, doi: 10.1016/j.physa.2010.01.042.

[130] J. Yang *et al.*, "A Memetic Algorithm Based on Adaptive Simulated Annealing for Community Detection," in *Intelligence Science IV*, Springer International Publishing, 2022, pp. 20–28. doi: 10.1007/978-3-031-14903-0_3.

[131] J. He, D. Chen, and C. Sun, "A fast simulated annealing strategy for community detection in complex networks," in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, IEEE, 2016, pp. 2380–2384.

[132] C.-H. Mu, J. Xie, Y. Liu, F. Chen, Y. Liu, and L.-C. Jiao, "Memetic algorithm with simulated annealing strategy and tightness greedy optimization for community detection in networks," *Applied Soft Computing*, vol. 34, pp. 485–501, Sep. 2015, doi: 10.1016/j.asoc.2015.05.034.

[133] B. Saoud, "Community structure detection in networks based on Tabu search," *Journal of Control and Decision*, vol. 0, no. 0, pp. 1–11, 2022, doi: 10.1080/23307706.2022.2146010.

[134] O. Gach and J.-K. Hao, "Combined neighborhood tabu search for community detection in complex networks," *RAIRO-Oper. Res.*, vol. 50, no. 2, pp. 269–283, Apr. 2016, doi: 10.1051/ro/2015046.

[135] Z. Lü and W. Huang, "Iterated tabu search for identifying community structure in complex networks," *Phys. Rev. E*, vol. 80, no. 2, p. 026130, Aug. 2009, doi: 10.1103/PhysRevE.80.026130.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[136] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an undirected graph," *Commun. ACM*, vol. 16, no. 9, pp. 575–577, Sep. 1973, doi: 10.1145/362342.362367.

[137] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, Jun. 2005, doi: 10.1038/nature03607.

[138] B. Rossman, "On the constant-depth complexity of k-clique," in *Proceedings of the fortieth annual ACM symposium on Theory of computing*, Victoria British Columbia Canada: ACM, May 2008, pp. 721–730. doi: 10.1145/1374376.1374480.

[139] S. Lehmann, M. Schwartz, and L. K. Hansen, "Biclique communities," *Phys. Rev. E*, vol. 78, no. 1, p. 016108, Jul. 2008, doi: 10.1103/PhysRevE.78.016108.

[140] H. Shen, X. Cheng, K. Cai, and M.-B. Hu, "Detect overlapping and hierarchical community structure in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 8, pp. 1706–1712, Apr. 2009, doi: 10.1016/j.physa.2008.12.021.

[141] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118–1123, Jan. 2008, doi: 10.1073/pnas.0706851105.

[142] N. Hurley and E. Duriakova, "An information theoretic approach to generalised blockmodelling for the identification of meso-scale structure in networks," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, San Francisco, CA, USA: IEEE, Aug. 2016, pp. 319–322. doi: 10.1109/ASONAM.2016.7752252.

[143] C. Toth, D. Helic, and B. C. Geiger, "Synwalk -- Community Detection via Random Walk Modelling," *arXiv:2101.08623 [physics]*, Jan. 2021, Accessed: Jul. 11, 2021. [Online]. Available: http://arxiv.org/abs/2101.08623

[144] G. de Guzzi Bagnato, J. R. F. Ronqui, and G. Travieso, "Community detection in networks using self-avoiding random walks," *Physica A: Statistical Mechanics and its Applications*, vol. 505, pp. 1046–1055, Sep. 2018, doi: 10.1016/j.physa.2018.04.006.

[145] S. Khanmohammadi, N. Adibeig, and S. Shanehbandy, "An improved overlapping k-means clustering method for medical applications," *Expert Systems with Applications*, vol. 67, pp. 12–18, Jan. 2017, doi: 10.1016/j.eswa.2016.09.025.

[146] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *PHYSICAL REVIEW E*, p. 5, 2008.

[147] B. Kamiński, P. Prałat, and F. Théberge, "Artificial Benchmark for Community Detection (ABCD)—Fast random graph model with community structure," *Net Sci*, vol. 9, no. 2, pp. 153–178, Jun. 2021, doi: 10.1017/nws.2020.45.

[148] T. Chakraborty, A. Dalmia, A. Mukherjee, and N. Ganguly, "Metrics for Community Analysis: A Survey," *ACM Comput. Surv.*, vol. 50, no. 4, pp. 1–37, Nov. 2017, doi: 10.1145/3091106.

[149] X. Liu, H.-M. Cheng, and Z.-Y. Zhang, "Evaluation of Community Detection Methods," *arXiv:1807.01130 [physics]*, Feb. 2019, Accessed: Jun. 12, 2021. [Online]. Available: http://arxiv.org/abs/1807.01130