

SURVEY

Community Detection Algorithms in Healthcare Applications: A Systematic Review

MEHRDAD ROSTAMI¹, MOURAD OUSSALAH^{1,2}, (Senior Member, IEEE),
KAMAL BERAHMAND³, AND VAHID FARRAHI^{1,2}

¹Center for Machine Vision and Signal Analysis (CMVS), Faculty of Information Technology and Electrical Engineering, University of Oulu, 90570 Oulu, Finland

²Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, 90570 Oulu, Finland

³School of Computer Sciences, Science and Engineering Faculty, Queensland University of Technology (QUT), Brisbane, QLD 4000, Australia

Corresponding author: Mehrdad Rostami (Mehrdad.Rostami@oulu.fi)

This work was supported in part by the Academy of Finland Profi5 DigiHealth-Project, a Strategic Profiling Program at the University of Oulu under Project 326291; and in part by the Ministry of Education and Culture under Grant OKM/20/626/2022 and Grant OKM/76/626/2022.

ABSTRACT Over the past few years, the number and volume of data sources in healthcare databases has grown exponentially. Analyzing these voluminous medical data is both opportunity and challenge for knowledge discovery in health informatics. In the last decade, social network analysis techniques and community detection algorithms are being used more and more in scientific fields, including healthcare and medicine. While community detection algorithms have been widely used for social network analysis, a comprehensive review of its applications for healthcare in a way to benefit both health practitioners and the health informatics community is still overwhelmingly missing. This paper contributes to fill in this gap and provide a comprehensive and up-to-date literature research. Especially, categorizations of existing community detection algorithms are presented and discussed. Moreover, most applications of social network analysis and community detection algorithms in healthcare are reviewed and categorized. Finally, publicly available healthcare datasets, key challenges, and knowledge gaps in the field are studied and reviewed.

INDEX TERMS Social network analysis, community detection, graph theory, healthcare application, medical data analysis.

I. INTRODUCTION

A. SOCIAL NETWORKS AND HEALTHCARE

The amount and number of accumulated data in healthcare databases have increased substantially in the last decade [1], [2]. This is mainly due to the increased digitalization trend and the growing enforcement policies for maintaining patient data either in dynamic databases or offline repositories for a limited time period. Typically, large-scale healthcare data arise from at least four primary sources: Electronic Health Record [3], Medical Imaging Data [4], Unstructured Clinical Notes [5], and Genetic Data [6]. Intuitively, analyzing such a huge amount of data would generate new opportunities for knowledge discovery enables us to gain new insights, and generate new hypotheses that can eventually lead to improvement in health services. The use of e-Health has also

The associate editor coordinating the review of this manuscript and approving it for publication was Kostas Kolomvatsos¹.

been promoted by World Health Organization (WHO) as an efficient way to achieve the 3rd United Nation Sustainable Development Goal of Good Health and Well-being [7].

Electronic Health Records, Patient Health Records, and Mobile Health are all subdomains of e-Health. The goal of e-Health is to provide affordable, secure, and efficient health services through digital and smart technologies. The use of e-Health ultimately benefits both individuals and organizations. For example, e-Health data analysis facilitates the delivery of health care services in developing countries [8], [9].

Although the massive amount of healthcare data provides valuable opportunities to improve patient health experience, it also prompts various sources of uncertainties (e.g., missing data, spelling errors, inefficient updates, anomalous patterns), which may lead to abusive activities and access restriction [17], [18], [19], [20]. The discrepancy and variety of the sources of data used in healthcare create additional challenges

associated with formatting and protection issues. This calls for innovative handling of healthcare data in a way to ease knowledge discovery and overcome such inherent challenges.

Grounded in theoretical physics and complex system theories [10], where the interaction between subsystems, network topology, and behavior of individual units are emphasized, social network analysis (SNA) emerged as a promising methodological framework that enables understanding the nature of relationship between these units within a system and how these relationships influence the system as a whole [11]. This offers a mean for mapping and exposing the hidden channels of communication and information flow among various units, exhibiting collaboration opportunities, and any potential gaps of this interaction [12]. In this context, individual units can refer to actors, organizations, group of individuals or events, depending on the context of study. This empowered SNA to emerge in several fields related to healthcare and health sciences such as neuroscience [13], epidemiology [14], social-support [15], cancerology [16], pharmacovigilance [17], rehabilitation [18], among others.

The link between healthcare and SNA can be viewed from different standpoints [19]. First, according to WHO, healthcare manifests through interactions between various organizations, actors, and service providers whose primary intention is to promote, restore, or maintain health [20]. This fully accommodates the structure and purpose of SNA systems. Second, SNA theories were historically born under the umbrella of health domain when Moreno [21] attempted to understand why an epidemic in a school has spread so quickly among pupils in 1934 in New York. Third, a multiplicity of third-party providers in healthcare systems, including public and private sectors, has led to the emergence of informational and technological innovations that benefit health services and citizens [22]. Therefore, SNA-based methods are becoming increasingly popular. For instance, SNA has been extensively used to comprehend the dynamics of the spread of infectious diseases [23]. An illustrative example is *contact tracing* apps that have been recently employed to model the dynamics of COVID-19 outbreak [24]. In [25], SNA was used to explore the health professional relationships in a way to identify potential gaps in these relationships, which would trigger appropriate action by health authorities. In [26], the authors investigated the application of SNA in obesity, hypothesizing that individuals with similar body weight tend to form a community. In [27], the authors used SNA to gain insights for detecting drug adverse reaction by mapping large patient communities. Social support is another domain where SNA has reportedly achieved exemplary success in uncovering the relationship among various care providers [28]. The preceding testifies to the recognition of both the research and health communities on the merits of SNA technology to tackle the increasing challenges associated with massive healthcare datasets.

Indeed, social network analysis exhibits a distinguished set of methodological frameworks that enable mapping, evaluating, and analyzing social relationships between individuals,

organizations, or any real entities, exploiting the power of network and graph theory. In this respect, the social structures are defined in terms of nodes, corresponding to individual actors, organizations, or entities, and the edges (or ties or links) representing the relationships or interaction among those nodes [29], [30]. Nodes can also be used to represent events or actions, depending on the focus of the studies. In the health sector, the structure of social networks (e.g., size, information diffusion, member relationships) has been hypothesized to model a member's health behaviors and orientations by influencing access to skills, opportunities, and limitations.

B. SNA AND COMMUNITY DETECTION

One of the most attractive techniques in SNA used for exploration and mining relevant insights from the raw data is Community Detection [31]. In network science, a community is a group of nodes having a stronger interaction compared to other nodes [32]. Community detection is an important SNA method developed for detecting a subset (or subgraph) of nodes that are similar to each other with respect to some similarity criteria [33], [34]. Community detection has emerged as a vital graph mining approach that has been applied to many real-world health applications. Some examples of these applications are mining complex molecular structures as in protein-to-protein interaction networks, mapping the interconnections among patients with respect to a predefined treatment in order to assess the impact of interventions, exploring the interaction between healthcare actors in a way to assess a set of predefined communication effectiveness, and identifying a cohesive and collaborative network of healthcare professionals to enhance the quality of service delivery [26].

Intuitively, targeting community may be more interesting than individuals, because many health interventions are rather targeting groups instead of individuals. Implicitly linked to SNA is the construction of the network itself. Although this is often viewed as part of SNA framework, it bears important ramifications in terms of the expected results that can be generated. In general, social network graph construction encompasses the specificity of research questions addressed in the underlined study. In this context, we distinguish methods that use automatic information retrieval techniques to extract useful data according to some predefined criteria or protocol, e.g., querying electronic health records for a given set of queries, scrapping data from web resources, or through performing some field observation or questionnaire and/or interview with stakeholders in relation to a question of interest. Although there are already a few interesting review papers associated with the application of SNA in health systems (i.e., [35] and [36]), the constant evolution of the technology, availability of new large-scale medical datasets, and sparsity of social network research call for new reviews in a way to benefit both health practitioners and researchers. This paper seeks to provide a comprehensive overview of community detection methods and their categorization in healthcare applications, highlighting their merits, pitfalls, and

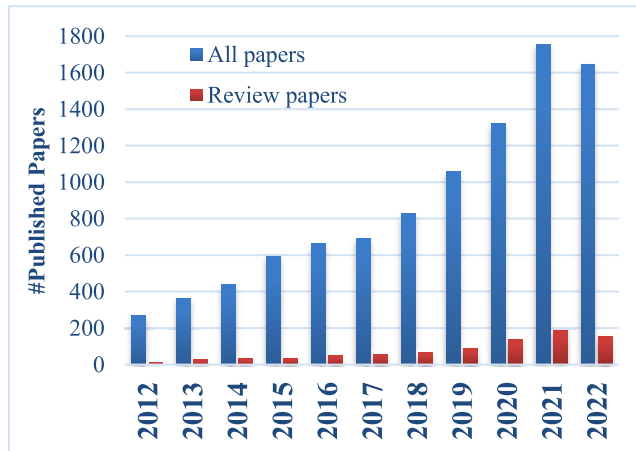


FIGURE 1. “All published papers” as well as “review papers” in Web of Science since 2012 with keywords “Social Network” and “Healthcare.”

future directions as well as surveying existing related healthcare datasets to highlight both theoretical and technological developments.

The application of SNA and community detection in healthcare are gaining increased interest year by year as illustrated in FIGURE 1, which highlights the trend of all published papers as well as *review* papers in the topics of “Social Network” and “Healthcare” since 2012.¹ This analysis shows that more than 60% of all papers were published during the period 2018-2022. Given the evolving of healthcare technologies, researchers have increasingly applied advanced SNA, for example, community detection and graph clustering approaches, to deal with complex healthcare problems.

C. OBJECTIVES AND RESEARCH QUESTIONS

The overall objective of this review is to investigate the scope and nature of the applications of community detection methods in the healthcare field, highlighting the key-methodological frameworks, state-of-the-art achievements, multi-disciplinary aspects, gaps and complementary aspects.

More specifically, this review addresses three research questions:

- RQ1: What are the main methodological frameworks underpinning the development of community detection-based approaches in the healthcare domain?
- RQ2: How multidisciplinary aspect impacted the development of community detection methods in healthcare domain?
- RQ3: What are the challenges and knowledge gaps in this domain?

D. COMPARISON WITH PREVIOUSLY REVIEW STUDIES

Despite its importance, as far as our research is concerned, there is no comprehensive research on the applications of community detection in healthcare. The surveyed existing

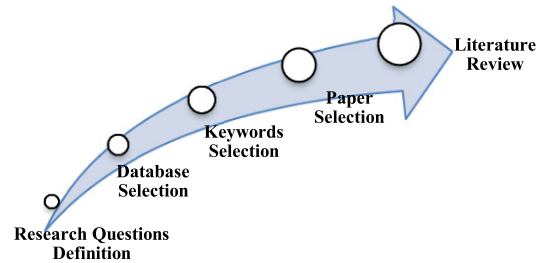


FIGURE 2. Search methodology for the paper selection.

review papers have focused on a specific thematic area and no systematic review has been conducted yet.

In Table 1, existing review papers is listed, and compared with our study. To the best of our knowledge, our study is the first to categorize the applications of SNA in healthcare, and discusses their advantages and challenges, and identifies existing datasets in the field.

E. REVIEW PROTOCOL

We followed the Systematic Literature Reviews protocol presented by Kitchenham and Charters [37]. This protocol is divided into three phases: plan phase, conduct phase, and report phase. In FIGURE 2, these phases and the steps involved in all phases are shown.

In the planning phase, research questions, related keywords, and publication databases were chosen. The main goal of this systematic literature review is to feature the works that have been performed in the healthcare application domain using SNA techniques. After defining the research questions, the search protocol was determined, including which databases to select, and what search strings to employ. We restricted our search to studies published in the years 2010-2021. The following scientific databases were explored:

- ACM Portal
- Springer
- IEEE Xplore
- Science Direct
- PubMed

These databases were used as the main scientific databases due to their comprehensiveness in “Community Detection”. Alternative searches used for Community Detection were “graph clustering”, “social network analysis”, “graph theory”, “community structure”, “clustering” and “modularity”. Additional keywords were used to limit the search to healthcare-related areas only, including “healthcare”, “health”, “cure”, “medicine”, or “treatment”. In the initial query 234, 315, 237, 315 and 137 papers (in total 1238) were identified from ACM Portal, Springer, IEEE Xplore, Science Direct and PubMed scientific databases respectively.

Each of these papers was screened separately two of the authors, based on the abstract and the title content. If the abstract or the title involved SNA in healthcare application, then it was marked for further investigation, otherwise, the article was discarded. Then, the papers that either focused

¹<https://github.com/Meh1988/SystematicReview->

TABLE 1. Comparison of our article with the previous review or survey articles.

Paper	Year	Brief description	Limitations
Chambers et al. [38]	2012	The authors reviewed ten bibliographic databases, reference lists, hand-searched selected journals and websites, and contacted experts in the healthcare domain. They included 52 completed research, studied in 62 publications. They found less evidence to use SNA in healthcare.	No categorization of SNA applications. Restricted topics. Available datasets were not discussed.
Sabot et al. [35]	2017	They studied professional advice and performance among healthcare providers. Two independent reviewers evaluated 5970 articles identified from ten databases between 1990 and 2016.	Focused on professional advice and performance among healthcare providers. Patient-Patient communities discarded.
Aqib Javed et al. [39]	2018	The authors sought a simultaneous survey on the methods of community detection and its applications in the different fields. The advantages and disadvantages of previously community detection methods are discussed. Open issues and future trends are also considered.	Healthcare applications were not considered. Publicly available datasets were not discussed.
Shelton et al. [36]	2019	This review sought to investigate the different applications of social network analysis in the development, dissemination, performance, and stability of health behavior for adults.	Review limited to SNA informed intervention. This leaves other SNA studies unaccounted.
Chunaev [40]	2020	The authors described and reviewed community detection in attribute social network from data fusion perspective (structure and node attribute), distinguishing early fusion methods, simultaneous fusion and late fusion methods according to clustering phase.	No specific focus on healthcare applications.
Moscato and Sperli [41]	2021	New classification based on SNA methods was provided (topology, game theory, AI model, fuzzy and greedy methods). Special review of SNA using neural networks. The problem of the applicability of community detection methods on the different online social networks is discussed. Open issues and challenges of SNA discussed.	No focus on healthcare applications. No review protocol employed.
AL-ANDOLI et al [42]		Community detection methods in social networks from classical to deep learning methods were reviewed. The related papers on deep learning with parallel and Meta-Heuristic techniques were studied and highlighted.	Healthcare applications were not considered.
Attea et al. [43]	2021	Heuristics and metaheuristics-based community detection methods in social networks were reviewed and categorized. Current applications and future progress and trends were reviewed.	Focused on non-heuristics community detection. Publicly available datasets were not discussed.
Bhambere [19]	2022	Light discussion of the role of social media in healthcare applications, together with the evolution of the National Institute for Health Research (NIHR). This is followed by a light review of Data mining techniques, and associated challenges.	No categorization of community detection. Very light review without known protocol.
Alamri et al [9]	2022	This study was conducted on the Blockchain-based Identity and Access Management systems in health IoT applications to investigate the security aspect. The authors summarized the main components and techniques in general Blockchain-based identity and access management systems.	Healthcare applications were not considered. Publicly available datasets were not discussed.
Our Paper	2022	A comprehensive categorization of community detection algorithms is presented. The strengths and weaknesses of previous community detection methods are discussed. The applications of SNA techniques, especially community detection algorithms, in healthcare domains are reviewed and categorized. The advantages and disadvantages of the previous SNA-based methods in healthcare applications were discussed. Publicly available social network datasets in healthcare applications are discussed. Open issues and challenges in the community detection methods are identified.	The experimental comparison of previously community detection methods is outside the scope of our study, and it has not been considered.

on community detection in the non-healthcare application domain or used non-social network analysis-based technique in healthcare application field were also excluded. In other words, only relevant works that used only community detection in healthcare applications were included. This led to a total of 45 papers.

This paper is organized as follows. In Section II, previously community detection methods are briefly reviewed and categorized. In Section III, previous applications of community detection in healthcare domain are reviewed and categorized. Section IV reviews previous community detection-based algorithms employed in healthcare

applications. In Section V, publicly available social network healthcare datasets are reviewed. Section VI discusses the open issues and challenges. Section VII presents the conclusion and finally, in Section VIII, future opportunities and works are pointed out.

II. COMMUNITY DETECTION ALGORITHMS

In recent years, researchers from different disciplines were attracted to the SNA field. Complex networks have become an effective way to represent the relationships between objects in many real-life systems. Typically, a network is modeled as a graph, where nodes represent objects, and edges

represent the relationships between these objects [44], [45], [46]. The community structure of a social network, which corresponds to some division of this network into groups of nodes that share dense interconnections and sparse interconnections, is a fundamental feature of social networks that has been intensively studied in recent years. In many real-world complex systems, such as sociology, biology, and transportation community structures can be defined as groups of nodes that are densely connected with each other according to some similarity metric.

Community detection has emerged as a primary method for understanding how network structure affects behavioral patterns in many systems [47], [48]. In this respect, communities are recognized as an effective way of identifying the underlying structures inferred from network topology and any additional information about the content of these nodes. This includes, for instance, finding potential friends from social media graph, identifying products that can be recommended to users from some user-product purchase network, and identifying social opinions from user-discussion network [49], [50]. The detection of a community group in the graph captures the tendency of nodes to create the groups, according to the employed similarity metric. Another terminology used for community detection is cluster. Although, clusters and communities are used interchangeably in graph theory literature, clustering sometimes stand for a single modality approach of community detection [51], [52]. The ability to detect communities of a given network contributes to the understanding of the inherent properties and structure of this network. Additionally, it provides an overview of communication among nodes of the same community as well as an analysis of the utility of these nodes [53], [54]. Typically, a community is defined as a collection of nodes that have a high degree of connection among themselves, but a low degree of connection with nodes outside the community set. Community detection presents a challenging field of study because there is no clear-cut model that describes the precise characteristics of a “community”. Especially, community detection involves the use of some optimization strategy for transforming a large-scale complex network into a set of disjoint and compact subgroups, without prior knowledge about the number of subgroups and their sizes.

Several algorithms for community detection have been developed, combining techniques and tools from different disciplines such as biology [55], physics [56], statistics [57], social sciences [58], [59], [60], cognitive sciences [61], mathematics [62], economics [63], and computer sciences [64]. It is commonly acknowledged that there is no unique community detection algorithm that can universally accommodate all kinds of social networks with high accuracy because of the discrepancy in network types and purposes. For example, in one type of network, algorithmic biases can increase the efficiency, while in another type, they may decrease the efficiency [65], [66], [67], [68]. A distinguished class of traditional community detection methods only considers the

structure of the network (i.e., relationships between nodes) and ignores node attributes, if any. For instance, in the case, where nodes represent patients, the latter can include attributes like age, gender, education, which leads to what is referred as a node-attributed social network [69]. In general, a complex network is subdivided into two dimensions where the first one corresponds to the network structure and the second one corresponds to the node attributes [70], [71]. Therefore, accounting for both network structure and node attributes, when available, is necessarily to yield a global and useful community [72]. In FIGURE 3, a categorization of previous community detection algorithms is shown and in the remainder of this section, these methods are reviewed.

A. NON-ATTRIBUTED COMMUNITY DETECTION METHODS

[73] In non-attributed community detection, only the linkage network structures are considered, while the node attributes are fully ignored [73]. We distinguish traditional graph-based methods and dimensional reduction-based methods for structural community detection. Dimensional reduction-based techniques are motivated by the need of a preprocessing stage that involves dimension reduction to ease the complexity of the data, prior to any clustering-like method to discover various communities. This technology is already available in the form of spectral clustering, non-negative matrix factorization, and more recently, deep clustering approaches.

1) TRADITIONAL METHODS

A conventional approach to community detection uses a graph structure to implement clustering algorithms that enable community categorization. Traditional community detection algorithms are typically classified into five categories: Hierarchical clustering [74], Modularity [75], Random Walk [76], Label Propagation [77], and Density-based [78]. These traditional community detection methods have the advantage of fast and easy implementation. Typical disadvantages of these methods include dependence on the initial solutions and non-scalability for large social networks.

a: HIERARCHICAL CLUSTERING

Hierarchical clustering is one of the most prominent and oldest approaches for finding communities in complex network analysis. The definition of a similarity measure is the starting point for any hierarchical clustering algorithm. Hierarchical clustering algorithms can be divided into two groups: agglomerative methods and divisive methods for community detection. Divisive algorithms use a top-down approach for finding clusters, in which a cluster is iteratively split into clusters based on edges connecting nodes with low similarity. Examples of divisive clustering methods include Girvan-Newman (GN) [79], edge clustering coefficient [80], and Information centrality [81]. Each of these algorithms uses a different approach for detecting suitable edges for the splitting tasks. Agglomerative algorithms are bottom-up

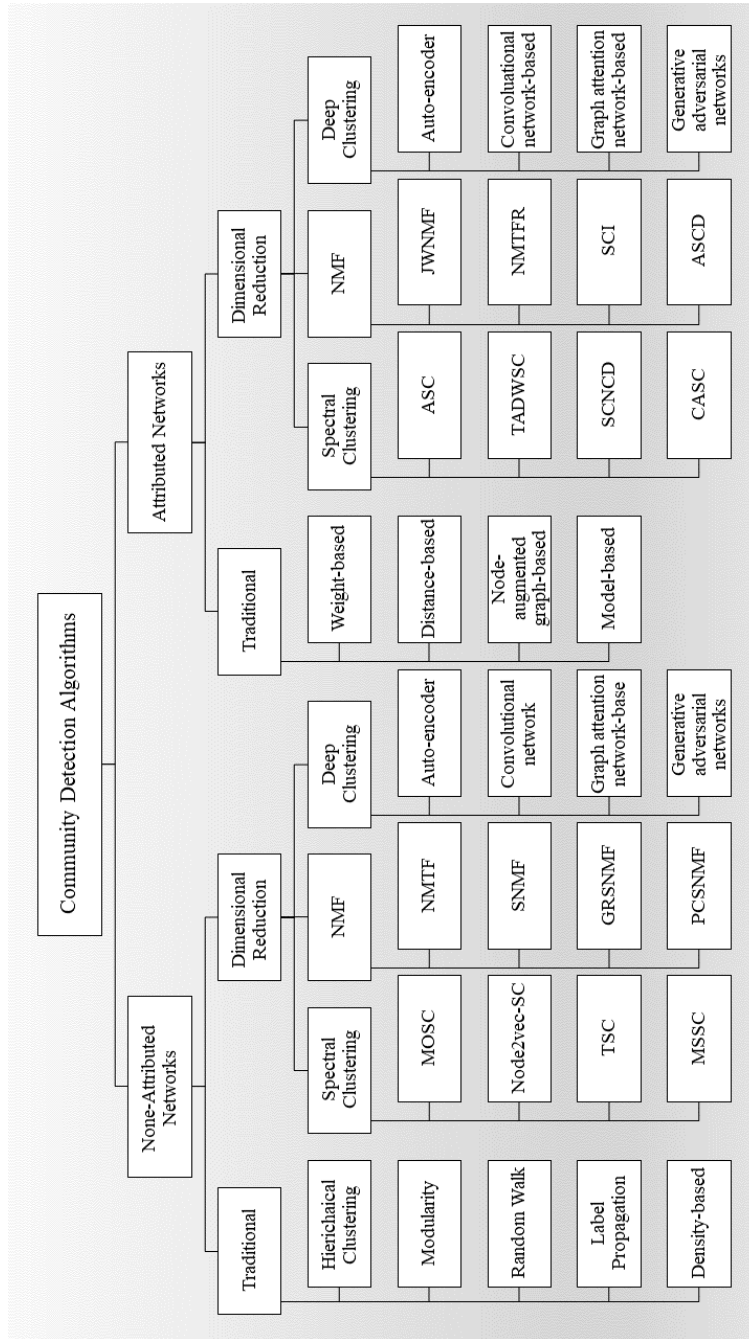


FIGURE 3. Categorization of methods for community detection.

methods, in which clusters are iteratively merged if their similarity is sufficiently high. Agglomerative algorithm uses node similarity while devise algorithm uses edge similarity.

b: MODULARITY

Modularity based approaches attempt to detect communities using some modularity measure. For a well-separated community, these strategies assume a high modularity value. The Sterling number of the second kind $S_n(k)$ gives the number of ways to partition n nodes into k non-empty groups, and hence the number of distinct community divisions is the Bell

number. As a result, modularity optimization is shown to be *NP*-complete. The purpose of all modularity-based approaches is to find a network division that maximizes the modularity value. The proposed modularity maximization strategies can be divided into two categories: greedy-based [82] and heuristic methods [83].

c: RANDOM WALK

In random walk methods, each node starts with a walker. After that, each walker will choose a neighbor of the node they are currently standing on at random to localize. Random

walk is based on the premise that walks tend to become stuck in dense parts of a network connecting communities [76]. Many community discovery approaches based on random walks have been proposed recently. In this area, popular methods include Markov Clustering algorithm (MCL) [84], Walktrap [76], and Infomap [85].

d: LABEL PROPAGATION

Raghavan et al. [77] proposed the label propagation algorithm (LPA) as a popular and rapid method for community discovery. A unique label is assigned to each node in the network at the start. In the next stage, each node updates its label with the label that is most frequently used by its neighbors. When several neighboring labels are equally weighted, the algorithm chooses a label at random from the more frequent labels. The procedure of label propagation is repeated until all nodes with the same label are grouped into a single community. The key advantages of LPA are its near-linear time complexity, ability to use local knowledge, independence from free parameters and objective function, and ease of implementation [86]. Despite its advantages, this method has some weaknesses, including instability, low quality, and the formation of huge communities due to its random behavior in initial node selection and node label update. Various modifications to LPA have been suggested in order to increase its stability and robustness [87].

e: DENSITY-BASED

Density-based clustering for community detection has been addressed in [78], [88], [89], and [90]. DENGRAPH is a density-based method that has been used for community detection in social networks [88]. It finds communities based on the DBSCAN clustering algorithm by incorporating an appropriate interaction-based distance function in the clustering process. The distance function is defined in terms of the interactions between the nodes of the network where the closeness of nodes is assessed by the number of interactions between them. Based on the incremental version of the DBSCAN algorithm, DENGRAPH provides an adaptable community. It calculates the aggregated number of interactions between two users in a social network. Nevertheless, this algorithm is one of the oldest density-based methods that has many problems as well, such as difficulty to accommodate diverse datasets, low stability, and inability to often detect evident and clear communities.

2) DIMENSIONAL REDUCTION

Clustering, as a process for discovering hidden groups in data, is an essential approach in statistical data analysis with applications ranging from community detection to image segmentation and advanced signal processing. One important notion to consider before getting into clustering is dimensionality reduction, which is often viewed as part of data preprocessing. This is especially important for large-scale and high-dimensional data, such as big graphs or streams

issued from a large collection of medical sensors. Large dimensionality often leads to high computational complexity, which in turn causes longer training and inference times. Furthermore, one may face a specific set of mathematical challenges, referred to as the “curse of dimensionality” [91]. Finally, higher dimensionality can result in more complex geometrical structures, potentially making the specification of a good distance function difficult. The exploration of structural features of high dimensional data represented by the low dimensional manifold helps in achieving good density structures. In general, depending on which method is used to reduce the dimension, the dimensional reduction community approaches can be divided into three categories: spectral clustering, which uses eigenvalue and eigenvector to reduce dimensionality; nonnegative matrix factorization method, which uses matrix factorization to reduce dimensionality; and the deep clustering method, which uses the graph embedding method to reduce dimensionality.

a: SPECTRAL CLUSTERING

Spectral Clustering (SC) is one of the most fundamental clustering approaches in dimensional reduction, and it has proven to be effective at grouping nodes by analyzing complex data structural information. The spectral clustering technique reduces the number of dimensions of the data matrix by using its spectrum [92].

The aim behind SC is to reduce dimensionality by using the eigenvalues of the data affinity matrix before performing clustering in a lower dimension space. Specifically, spectral clustering begins by constructing an affinity matrix from data, then derives the graph Laplacian matrix from the affinity matrix. Next, it, computes the first k eigenvectors corresponding to the k -largest eigenvalues of the graph Laplacian matrix, and, finally, clusters the k dimensional row using the k -means algorithm. The quality of the affinity graph has a significant impact on the performance of spectral clustering. Different techniques have been proposed to improve the spectral clustering algorithm, typically focusing on constructing an accurate affinity matrix that can faithfully reflect the similarity information between each pair of nodes. Popular SC algorithms include Mixed-Order Spectral Clustering (MOSC) [93], Node2vec-SC [94], multi-similarity spectral (MSSC) [95], and Tensor Spectral Clustering (TSC) [96].

b: NONNEGATIVE MATRIX FACTORIZATION (NMF)

Nonnegative matrix factorization (NMF) is a classical low-rank matrix factorization model. It is very useful for analyzing matrices with all non-negative elements [97]. Mathematically, given a non-negative matrix $X = [X_1, X_2, \dots, X_N] \in R_+^{m \times n}$ composed of n by m -dimensional data vectors, and the desired reduced dimension d ($d \ll \min(m, n)$), NMF aims to find two non-negative matrices $W = [W_{ip}]^{m \times d} \in R_+^{m \times d}$ and $H = [H_{jp}]^{n \times d} \in R_+^{n \times d}$, which can well approximate to the original matrix X in the form of their product:

$$X \cong WH^T \quad (1)$$

where \mathbf{W} and \mathbf{H} are respectively called the basis matrix and the coefficient matrix. In order to estimate the factorization matrices, the following objective function is minimized:

$$\left\| X - WH^T \right\|_F^2, s.t. W \geq 0, H \geq 0 \quad (2)$$

Since the non-negativity constraints on both factor matrices lead to part-based representation, which is consistent with the concept of learning the parts to build the whole. *NMF* has been widely used in various fields, particularly in community detection [98]. For directed or undirected networks, the conventional *NMF* model: $X \cong WH^T$ can be directly used to detect communities by replacing \mathbf{X} with \mathbf{A} . Each column in the matrix \mathbf{W} corresponds to a representation of the community. Each column of matrix \mathbf{H} denotes the membership between overall pairs of communities and nodes. Therefore, it can assign a community label to a given node by finding the index of the largest element in the coefficient matrix, and the detection rule for *NMF* is denoted as:

$$\text{community}(v_j) = \underset{i}{\text{argmax}} \quad i = 1, 2, \dots, V_{j1} \quad (3)$$

Popular *NMF* approaches for community detection include NMTF (Non-negative Matrix Tri-Factorization) [99], SNMF (Symmetric Non-negative Matrix Factorization) [100], GRSNMF (Graph Regularized Non-negative Matrix Factorization) [101], and PCSNMF (Pairwisely Constrained Symmetric Non-negative Matrix Factorization) [102]. Existing *NMF*-based community discovery approaches are generally effective, but they all have the same disadvantages [103]. To begin, they frequently construct the feature matrix \mathbf{X} by using the neighboring matrix \mathbf{A} , which is insufficient to capture the complex network global features. This in turn degrades the overall performance of the community detection. Second, they frequently use the Frobenius norm to create the objective function. This is insufficiently resilient because a few noises with large errors can easily dominate the objection function due to the form of the squared errors. Finally, their solutions to the *NMF*-based model for community finding include the multiplication operations of multiple $n \times n$ matrices, yielding a quadratic time complexity of $O(n^2)$ for each. Although we can lower the time complexity by using sparse matrix multiplication or other optimization tactics, the computational cost is still high when dealing with large-scale complex networks.

c: DEEP CLUSTERING

Deep clustering techniques try to combine data transformation and clustering into a single architecture that can then be tuned simultaneously by minimizing a loss function [104]. The architectures utilized for the feature learning step are drawn from the field of deep learning. A deep clustering method consists of two main parts, dimension reduction and data clustering. Depending on which deep learning method is used to reduce the dimensions, deep clustering can be classified into four major groups [105]: Auto encoder-based [106], [107], [108], [109], Generative

Adversarial Network-based [110], [111], [112], [113], Graph Convolutional Network-based [114], [115], [116], [117], and Graph Attention Network-based [118], [119].

B. ATTRIBUTED COMMUNITY DETECTION METHODS

Real-world graphs where nodes have multiple attributes are becoming more popular. Social networks, for example, feature both friendship relationships and user data such as hobbies and demographic attributes. As another example, the expression levels associated with each protein may be included in a Protein-Protein Interaction (PPI) network, in addition to the interaction links. A graph can be used to represent such information, with nodes representing items, edges representing relationships between them, and feature vectors linked with nodes representing properties. An attributed graph is a term used to describe this type of graph data. Unlike non-attributed community detection, attributed graph methods focus on both structural and attribute knowledge [120], [121], [122]. For this purpose, attributed community detection methods combine information from both networks and attributes and detect clusters of densely linked vertices that show high similarities in their attributes [123], [124], [125].

1) TRADITIONAL APPROACHES

Let $G = (V, E, A)$ be an attribute network, defined over a set of vertices \mathbf{V} , edges \mathbf{E} , and attributes \mathbf{A} . Let K be the number of clusters, the attributed network community is consisted of sets of vertices V_1, V_2, \dots, V_K where $V = \bigcup_{i=1}^K V_i$ and $V_i \cap V_j = \emptyset$ for any $i \neq j$. As a result, nodes within clusters are tightly connected in terms of structure, whereas nodes in other (non-community) clusters are sparsely connected; and nodes within the same cluster have limited diversity in their attribute values, whereas nodes in separate clusters may have diverse attribute values. These approaches integrate structural information with attribute characteristics. Nevertheless, the major problem of such solutions lies in their inefficiency to tackle large amounts of data due to the time complexity of various search operations. Based on previous research, there are four types of attributed community detection methods [128], including (1) Edge weighting Methods, (2) Distance-based Methods, (3) Node-augmented graph-based methods, and (4) Model-based Methods.

a: EDGE WEIGHTING METHODS

The edge weighting method category of community detection attempts to convert the original attributed network into a weighted graph. The methods of this category construct a weighted graph based on the attribute information of the nodes in the network. Then, node attributes are eliminated from the vertices by acquiring their information from the edges of the network, which is done by assigning to each attribute a similarity value between two nodes as the weight of it. After the graph is reshaped as a weighted graph, different graph clustering algorithms can be used, considering the weight of edges. Clusters of nodes with similar attribute

values can be obtained if edges are maintained at high weight values during clustering [126], [127]. FocusCO [128] is one of the most important community detection methods in this category. To use this method, a set of user-provided exemplars are employed to infer user preferences by analyzing the attributes. Then, the communities and outliers are combined based on this user preference in this new scenario. The FocusCO method detects outliers, extracts the communities of focus, and detects the final communities.

b: DISTANCE-BASED METHODS

In distance-based methods, the graph structure information is stored as a distance measure between vertices, and then integrated with the attribute similarity measure. This category of community detection utilizes distance-based techniques and measures such as the SI-Cluster [129], SA-Clustering [130], and CODICIL [131]. Several similarity (distance) criteria are developed for this aim, and the classic distance-based clustering algorithms can be also employed.

c: NODE-AUGMENTED GRAPH-BASED METHODS

Through an augmented graph, this class of methods attempts to merge topological structure with attribute information. Additional vertices called attribute vertices and new edges called attribute edges are added to the original graph topological structure. The attribute similarity is expressed in the augmented graph as the vertex neighborhood: two vertices sharing the same attribute value are connected by a shared attribute vertex. Because each vertex v_i has T attribute values, the original graph will have $|V| \times T$ attribute edges. Two vertices in the augmented graph are close if they are connected through a large number of other original vertices, or if they have a large number of shared attribute vertices as neighbors or both. Following the creation of the augmented graph, a distance measure that estimates pairwise vertex closeness or a community discovery technique can be used to identify the clusters [132].

d: MODEL-BASED METHODS

The model-based category utilizes a probabilistic model to avoid the design of an unrealistic distance criterion. The model-based attributed community detection methods often advocate the inclusion of some probabilistic optimization criteria in the mechanism of community generation process. CESNA [133], Bayesian probabilistic model [134], and PCL-DC [135] are among the most popular model-based methods in this category.

2) DIMENSIONAL REDUCTION

In this approach, a graph with attributes that includes structural information utilizes a preprocessing step of dimensional reduction on structural and attribute data. We distinguish three main approaches for this purpose: spectral clustering, non-negative matrix factorization and deep clustering.

a: SPECTRAL CLUSTERING

Spectral clustering creates an affinity matrix from original data, then it derives the graph Laplacian matrix from the affinity matrix and computes the first k eigenvectors corresponding to the k largest eigenvalues of the graph Laplacian matrix. This method finally clusters the k dimensional row vectors of the matrix formed by the first k eigenvectors of the Laplacian as columns using the k -means algorithm [136]. The quality of the affinity graph has a significant impact on the performance of spectral clustering. As a result, creating an affinity matrix that accurately displays the similarity information between each pair of nodes is crucial. Spectral Clustering based on Node Convergence Degree (SCNCD) [137], Attributed Spectral Clustering (ASC) [138], Text-Associated DeepWalk-Spectral Clustering (TADW-SC) [139], and covariate-assisted spectral clustering (CASC) [140] are famous algorithms in this category.

b: NON-NEGATIVE MATRIX FACTORIZATION (NMF)

Clustering in attributed networks simultaneously considers the connectivity and attributes of the vertices. Non-negative Matrix Factorization (NMF) is acknowledged as a powerful technique that can combine structure and information for community detection in attributed networks. In these methods, the primary challenge was finding a way to effectively fuse two types of information [98]. Suppose we have the adjacency matrix $A = [A_{ij}]^{n \times n} \in \mathbb{R}_+^{n \times n}$ and attribute matrix $Y = [Y_{ij}]^{m \times n} \in \mathbb{R}_+^{m \times n}$ where n is the number of nodes and m is the number of attributes. Then one common approach, which was applied in Nonnegative Matrix Tri-Factorization (NMTFR) [141], and Joint Weighted Nonnegative Matrix Factorization (JWNMF) [142], with the same factor $H = [H_{ij}]^{n \times k} \in \mathbb{R}_+^{n \times k}$, referred to as community indicator matrix in the following optimization problem:

$$\begin{aligned} \min \mathcal{L}(W, H) &= \|A - HH^T\|_F^2 + \alpha \| \Lambda Y - WH^T \|_F^2, \\ \text{s.t. } H &\geq 0, \quad W \geq 0 \end{aligned} \quad (4)$$

where Λ diagonal matrix is to assign a weight for each attribute and $W = [W_{ij}]^{m \times k} \in \mathbb{R}_+^{m \times k}$ is the community attribute matrix and the contribution of attributes of nodes can be determined through α .

Another approach used to benefit from the node attributes in community detection consists of applying the chain factorization model, as proposed in SCI [143], ASCD [144], using the following objective function:

$$\begin{aligned} \min \mathcal{L}(W, H) &= \|A - HH^T\|_F^2 + \alpha \|H - Y^T W\|_F^2, \\ \text{s.t. } H &\geq 0, \quad W \geq 0 \end{aligned} \quad (5)$$

c: DEEP CLUSTERING

Deep clustering has been trending in the attributed graph as a popular approach in recent years. Deep clustering is divided into two stages: dimensional reduction using a deep learning algorithm and data clustering in Euclidean space. Deep

learning approaches make it easier to capture the structure and content information in attributed graphs through feature aggregation. This had a lot of success with attributed graph clustering as well. It entails converting each node n of an attributed graph into a latent space, low-dimensional feature vector $F(n) \in R^d$, where d is the latent dimension. In this approach, $F(n)$ encodes both node properties and graph structure information of node n in such a way that nodes in the same cluster have similar feature vectors. This methodology, like structural methods, is separated into four primary groups based on the deep learning strategy chosen to minimize the dimension: auto Encoder-based [145], [146], [147], [148], Generative Adversarial Network-based [111], [149], [150], Graph Convolutional Network-based [149], [151], [152], [153], and Graph Attention Network-based [118], [154].

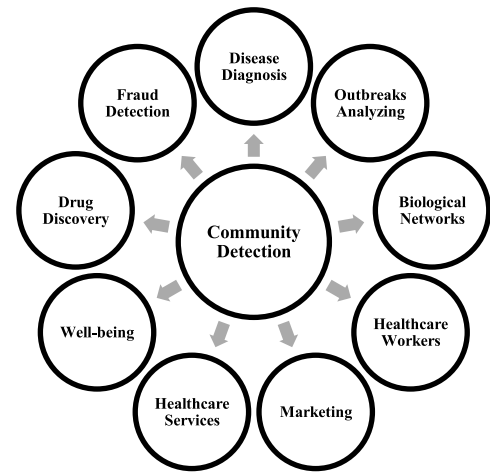


FIGURE 4. Applications of community detection in healthcare.

III. COMMUNITY DETECTION IN HEALTHCARE DOMAIN

Research in healthcare systems examines how people, institutions, and resources work together to meet the health needs of target populations [155], [156], [157]. A graph of health systems can reveal, for instance, the inherent complexity of the overall healthcare system, shed light on the interdependence among its various parts, and identify the corresponding challenges associated with these interconnections. On the other hand, various systems have accumulated a vast amount of medical data, namely big data, due to the extensive application of technologies such as mobile Internet, cloud computing, and the Internet of Things in the healthcare domain. Medical care, medical emergency communication, and various medical and healthcare research have extensively utilized such big data to deliver insights or decision aid. Especially, utilizing medical big data effectively is seen as a tool to reduce medical costs, improve medical quality, and increase patient prognosis, and, subsequently, will have a significant effect on society, the economy, and public health. In order to describe and evaluate such dynamic and open systems, new and innovative methods are needed. As a result, researchers in this field have extensively used artificial intelligence tools to improve the quality of medical care.

Indeed, nowadays, with the boom in digitalization and new regulation that enforces patient data preservation, the healthcare system is rapidly being dominated by artificial intelligence and machine learning [158], [159], [160]. This transformed the manual health system into automatic systems, in which clinicians and health operators perform a routine patient record search and update to manage various patient and system resources in a way to ensure transparency and efficiency. This provided a substantial boost to healthcare services in terms of resource optimization, healthcare personalization, and enhanced diagnosis [44], [161].

Artificial intelligence, SNA and more recently community detection, are key techniques that have demonstrated the ability to translate these medical big data into actionable knowledge resources [50], [162], [163], [164]. Therefore, social network analysis and community detection can contribute to

improved patient diagnosis [165]. Especially, a community detection is a specific approach for mapping, measuring, and analyzing social relationships between individuals, teams, and organizations.

We can broadly categorize most applications of community detection in healthcare into Healthcare Event Monitoring and Disease Diagnosis, Analyzing and Tracking Outbreaks, Analyzing Healthcare Workers Network, Network Analysis of Healthcare Services, Analyzing Marketing in Healthcare Agencies, Characterizing Lifestyle, Behavioral Health and Well-being, Drug Discovery and Fraud Detection. These applications of SNA for healthcare are shown in Fig. 4. In the remainder of this section, the previous applications of SNA in these subfields are briefly detailed.

A. HEALTHCARE EVENT MONITORING AND DISEASE DIAGNOSIS

In the modern era, more and more healthcare data are being collected by computer systems and storage technologies [166]. Massive data growth in healthcare fields has posed a major challenge to healthcare data analysis techniques [167], [168], [169]. Medical doctors and specialists cannot analyze such an overwhelming amount of medical data in a short period of time to make medical diagnoses, predictions, or treatment schedules. Therefore, it is important to propose an intelligent algorithm that can handle large quantities of medical data accurately. Due to this, artificial intelligence and machine learning-based models have become essential for the healthcare industry. Community detection can be utilized in the medical diagnosis field by capitalizing on the interconnections among various diseases, and/or diagnoses, and/or treatments. In [170], a multi-objective gene selection model for microarray data classification is developed. In this paper, using a graph representation and employing node centrality and community detection, a set of relevant and non-redundant genes is selected for the final disease diagnosis phase. The experimental results on colon,

leukemia, prostate tumor, and lung cancer datasets shows that this model can improve the accuracy of disease diagnosis as well as decrease its computational complexity. Ali et al. [171] integrated advanced models, including community detection, wireless sensor data analysis, ontologies, and deep learning, to develop a novel health monitoring system for chronic patients. In [172], the authors focused on high-risk patients that obtain common care in an initial care outpatient practice where they build a community-based lung cancer diagnosis model. In [173], two community detection methods were utilized to group a lymph node tissue section that had been diagnosed with metastatic infiltration. The results indicated that this developed clustering technique can improve the accuracy of cancer detection. Moreover, the authors of [174] developed a community detection-based feature selection method in the gene selection problem. In their work, the features of the dataset are represented as a graph model, and then, grouped using community detection algorithms into several communities. Next, features are selected using genetic algorithm and considering the communities of each feature whose outcomes are inputted to an ensemble-classifiers (i.e., KNN, SVM and AdaBoost) that enable disease prediction evaluation. In e-healthcare systems, wireless body area networks (WBANs) are successfully used to monitor the condition of remote patients through wearables or implantable biomedical devices. A novel clustering and routing scheme for multi-WBANs has been developed in [8] that clusters similar sensors and aggregates data.

B. ANALYZING AND TRACKING OUTBREAKS

Epidemic or widespread infections, such as COVID-19, Influenza, Swine flu, Ebola, and Cholera need systematic surveillance of clinical syndromes that affect the public's health where early warning of a rising trend in disease prevalence can help to prepare patients and increase efforts to distribute the appropriate vaccine or treatment. On the other hand, the knowledge of a declining trend can be a metric of success [175], [176]. Classical public health monitoring relies on surveying and compiling primary data from healthcare providers and pharmacists. Online platforms, including social media platforms such as Twitter and Facebook, have created new opportunities for real-time monitoring of disease outbreaks, yielding new epidemiology research areas at the intersection of computer science and public health for infection outbreaks prediction [177]. Community detection has been employed to help draw overall statistics regarding pandemic spread, isolate people at risk, among others [178]. In [178], the authors developed a model for detecting targeted immunization nodes using community detection. Three steps are included in the developed method, including community detection, narrowing of potential candidates, and a meta-heuristic method for selecting immunization nodes. To detect communities of related cases of an infectious disease, the authors of [179] proposed a new model for integrating multiple data streams. They incorporated pre-existing disease

information into their model and explicitly accounted for under-reporting. Moreover, the authors of [176] used machine learning techniques and social network analysis to raise awareness of COVID-19. In this study, node grouping and community detection on social networks were performed using some edge similarity, edge degree, and node clustering measures techniques.

C. ANALYZING BIOLOGICAL NETWORKS

Due to the rapid development of the data storage technologies, large scale PPI data as well as unannotated protein sequences are now available, which offers golden opportunities to apply various SNA techniques to mine protein structures. Especially, in biological network representation, vertices denote certain biological entities and edges indicate the links between these biological entities, which can vary according to the context of the study. In [180], a novel community detection method is developed for large-scale PPI networks. As a result of the developed modularity measure and functional cohesion, protein communities are identified in this model. In [181] protein-protein interaction network of non-alcoholic fatty liver disease is analyzed. Using experimentally validated interactors as well as the interaction network of human proteins associated with Non-Alcoholic Fatty Liver Disease (NAFLD), the authors identified novel associations between human proteins that may be implicated in non-alcoholic fatty liver disease. In [182], a new method analyzes the disease network using PCA algorithm based on disease-gene network detecting algorithm, which is useful for examining the relationships between nodes within a single cluster. In [183], an automated method for finding communities of interconnected proteins in protein interaction networks and retrieving annotations associated with these clusters is described. Based on Girvan and Newman's Edge-Betweenness community detection method, proteins' interaction graphs were divided into subgraphs of interconnected proteins. The finding demonstrates the ability of the method for fast screening of small to medium size protein interaction networks.

D. ANALYZING HEALTHCARE WORKERS NETWORK

Understanding how connections are organized in healthcare worker networks can reveal useful insights to comprehend how these networks affect healthcare workers' life, behaviors, productivities, and communication failures [184]. A better understanding of this issue would improve communication strategies for the diffusion of healthcare information. A physician social network, for instance, is widely credited with influencing physician preferences, attitudes, and beliefs about the effectiveness of new technologies in the clinical setting. In [185], a new multi-scale community detection is introduced to assess the characteristics of healthcare providers and their impact on patient outcomes. This approach can be used by health service researchers to assess provider collaboration and its impact on patient outcomes. In [186], the effects of

interruptions on different unique roles of work in an ICU are investigated using a sociotechnical system. The study used SNA such as node centrality and community detection to identify systems-based interventions that can decrease unnecessary interruptions and avoid unintentional outcomes, which would impose supplementary burdens on staff working in the ICU. In [187], the authors developed a novel algorithm to detect physicians' real medical specialties by analyzing their prescription history. Scientists could use this knowledge to analyze physicians' drug prescriptions more accurately. In systems where patients have the freedom to choose their physician, identifying informal networks of ambulatory physicians is particularly challenging. Using routine data, authors of [184] showed that SNA can identify informal networks of ambulatory physicians.

E. NETWORK ANALYSIS OF HEALTHCARE SERVICES

Healthcare services can be represented as a graph model and this model can be investigated using SNA. Understanding the organization of this graph of healthcare services not only help in analyzing existing healthcare services at wide but also aid in evaluating and discovering unknown patterns that harm the development of healthcare service. For instance, in [188], different techniques and models for creating and analyzing the network of healthcare services are reviewed. To detect communities of services, a model based on PCA algorithm and SNA is developed and then a graphical representation for demonstrating and recognizing patterns of care is proposed. In [189], essential details about the professional advice networks in Ethiopia's primary healthcare units are examined and assessed using a set of SNA-based methodology and qualitative interviews. De Brún and McAuliffe [190] described used community structure-based techniques for describing, diagnosing, and evaluating healthcare systems and networks in a case study of senior managers in a newly established hospital group. In [191], disease regional areas are detected and monitored using a combination of machine learning and Louvain community detection algorithm on drug prescriptions. The first step in the process was to design the network of prescriptions and then apply the Louvain community detection algorithm for the community discovery of 50000 Iranian prescriptions collected in 2014. Validation of the experiments is based on the modularity measure and specialist opinions. Six communities are extracted using Louvain community detection algorithm with adequate performance. In [192], a novel community detection algorithm is proposed for transforming heterogeneous knowledge into operational information on Information Systems Medicalization Program data and its results are compared with Agglomerative Hierarchical Classification and the diagnosis related groups classification.

F. ANALYZING MARKETING IN HEALTHCARE AGENCIES

Marketing is one of the most vital skills for many healthcare agencies to handle fierce competition in the field [193], [194].

Using online social networks, healthcare service providers, recipients of healthcare services or potentially new customers can interact with each other in a way to increase visibility, boost collaboration, raise brand awareness and resolve any concerns raised by the community [195]. The use of social networks can therefore be beneficial for promoting a company's profile, or to increase patients' loyalty. In this respect, community structure analysis can help healthcare agencies to adopt a data-driven marketing strategy involving users' behavioral analysis and accommodating community needs to ultimately build a healthier world [196].

G. CHARACTERIZING LIFESTYLE, BEHAVIORAL HEALTH AND WELL-BEING

The concept of well-being emerged as an important health indicator whose evaluation goes beyond common economic indicators as it affects the overall health, productivity, and quality of life of the underlined country. Studies of life satisfaction and well-being differ conceptually from flu or allergy forecasts because they aim to provide insights into factors and mechanisms contributing to happiness, rather than just determine regional variations in it. SNA can therefore be called upon to understand the influence that various social networks may have on health outcomes and behaviors. In [197], the authors examined the language of well-being in order to estimate the cognitive quality of life satisfaction where a billion tweets were collected over a year's time span across various US cities. Then using a questionnaire inferred from semantic textual analysis, linear regression is employed to predict the concept of life satisfaction scores derived from such questionnaires. In [36], the current applications of community structure analysis algorithms for implementing, disseminating, and sustaining health behavior interventions for adults are reviewed. Zong et al. [198] developed a framework for categorizing user-health levels that aggregates results derived from a variety of medical social networks. In the first step, they construct a multiple affinity graph to characterize a user's relationship with each attribute. Thereafter, the massive number of users are clustered into several medically related social communities using community detection methods. By using these communities, the health status of each user is distinguished by their classified health level based on their companions. Finally, a multi-view medical attribute learning framework is presented for combining the health level of users from various attributes.

H. DRUG DISCOVERY

Introducing a new drug to the market is a long and costly procedure for the pharmaceutical industry. Hence, a comprehensive analysis based on existing drugs can be used to improve the positioning of drugs that have been identified with new clinical indications, reduce development risks and costs, and accelerate drug discovery to the market [199]. Community detection can reveal effective in analyzing the relationships between exiting drugs and/or target indicators.

For instance, Galan-Vasquez and Perez-Rueda [200] examined the relationships between drugs and their targets using the DrugBank dataset and annotations from the United States Food and Drug Administration. CDDI (Core Drug Discovery for Indications) is currently a popular research topic in Traditional Chinese Medicine. Moreover, in [201] an improved label propagation community detection is developed for analyzing a drug network to discover drug communities and core drugs for treating different indications in Traditional Chinese Medicine. In this method, communities are detected by adopting fixed label update order based on the ascending order of node importance. The results on different dataset networks indicate that the developed community detection method acquires better performance and stability than previous methods. Moreover, this method can find efficient core drugs in drug networks. Using modularization to shape metabolic networks is one way to understand how communities in networks influence network stability and evolution. In this respect, authors in [202] used different kinds of thermophiles and mesophiles as the study subjects and built their metabolic graph utilizing the Pajek algorithm. Then, the “component” description and community detection method were employed to detect the clusters in the metabolic graph.

I. FRAUD DETECTION

Data mining and social network analysis play an essential role in early fraud detection and surveillance of fraud. The relationship between physicians and patients forms sophisticated communities, which can potentially lead to fraud detection as well. Scientists typically study clusters of physicians and patients to identify suspicious communities. In [203], different types of communications are investigated and examined, where communications that appear suspicious and may reveal healthcare fraud are highlighted. The study proposed two methods to find these small and exclusive groups. The authors proposed high scalability methods that can be applied to large medical data. These developed methods were evaluated with different synthesized data. The finding of this study showed that these algorithms are very effective and can measure the community structures of 50,000 providers in about a minute.

In Table 2, a summary of previously mentioned community detection-based studies employed in healthcare applications, a subfield of their application, utilized techniques, purpose, advantages, and disadvantages, are outlined.

IV. DEVELOPED COMMUNITY DETECTION ALGORITHMS IN HEALTHCARE APPLICATIONS

Community detection plays an important role in many healthcare applications, including metabolic network analysis, protein-protein interactions, food web construction, food-drug interaction network analysis, gene regulation, and pollination. In this section, previous community detection algorithms in healthcare applications are reviewed and their strengths and weaknesses are discussed.

A. LOUVAIN ALGORITHM

Louvain community detection is an algorithm for detecting communities in social networks using optimization of a modularity measure. Louvain is an unsupervised algorithm, and it does not require the input of the number of communities nor their sizes before the clustering process. This algorithm is simple, powerful, and easy-to-implement to identify network communities. The Louvain algorithm can be employed to detect communities in very large networks with low execution times. It consists of two simple steps: the first step assigns each node to a community that maximizes network modularity Q ; the second step calculates the gain from moving a node i to a community C as follows:

$$\Delta Q = \frac{\sum_C + k_i^C}{2m} - \left(\frac{\sum_{\hat{C}} + k_i}{2m} \right)^2 - \left[\frac{\sum_C}{2m} - \left(\frac{\sum_{\hat{C}}}{2m} \right) - \left(\frac{k_i}{2m} \right) \right] \quad (6)$$

where \sum_C indicates the sum of the weights of the edges inside community C , $\sum_{\hat{C}}$ is the sum of the weights of the edges incident to nodes in community C , k_i is the sum of the weights of the edges incident to node i , k_i^C is the sum of the weights of the edges from i to nodes in C , m denotes the sum of the weights of all the edges in the network.

As a second step, a new network composed of nodes that represent previously discovered communities is simply created. Iterations are carried out until the modularity of the network has been significantly improved. Several network analysis software packages, such as NetworkX, [204] and Gephi [205] implemented the Louvain community detection algorithm.

The computational complexity of the Louvain algorithm is $O(n \log n)$, where n is the number of nodes in the social network. This algorithm can therefore be utilized to identify communities in very large networks within relatively short computing times. Due to its simplicity and high performance, the Louvain community detection algorithm has been extensively used for social network analysis in healthcare applications [174], [187], [191]. Despite its good performance, this method also has drawbacks. Traag et al. [206] showed that the Louvain community detection method has a significant problem, for both modularity and the Constant Potts Model. As a result of the resolution limit, modularity may cause smaller communities to be clustered into larger ones. In other words, modularity may mask smaller communities and can produce communities with significant substructures. In addition to the well-known issue of the resolution limit of modularity, this method may produce arbitrarily inadequate connected communities. In some cases, communities may even be disconnected [206].

B. TRADITIONAL CLUSTERING ALGORITHMS

In traditional algorithms, the problem of community detection is represented as a graph, and then using a clustering algorithm the community of the network will be detected.

TABLE 2. Outlining the reviewed community detection-based method in healthcare applications.

Authors	Subfield of application	Purpose	Advantages	Challenges
Rostami et al. [170]	Healthcare monitoring and disease diagnosis	Feature selection for disease diagnosis	Reducing the time complexity of microarray data while increasing classification accuracy.	Lack of explainability
Ali et al. [171]		Intelligent healthcare monitoring	Healthcare data can be stored and analyzed precisely, and classification accuracy can be improved.	It is not possible to accurately predict patient mental health and drug side effects using the proposed model.
Bechtel et al. [172]		Lung cancer diagnosis	High-risk patients can be identified in early treatment	Lack of explainability
Wang et al. [173]		Cancer diagnosis	Improve final diagnosis accuracy	Traditional clustering methods have limited the model's performance
Rostami et al. [174]		Feature selection in medical dataset	A smaller set of features is selected for final classification	The usability in very high dimensional medical datasets is limited
Javaheri et al [8]		Routing protocol for wireless body area networks	Enhancing the effectiveness of wireless body area networks	The performance of the model has not been evaluated on real medical datasets.
Wang et al. [178]	Analyzing and tracking outbreaks	Preventing epidemic spreading	Selecting targeted immunization nodes by employing community detection technique	Limited performance in large-scale networks
Choudhury et al. [176]		COVID-19 community detection	Enhancing COVID-19 public health care links using spatial information and community structure	Using traditional clustering models in real-world large-scale data has been limited the performance of the method
Cori et al. [179]		Outbreak detection	An early assessment of outbreaks of infectious diseases	Low performance in partially missing data
Chen et al. [180]	Analyzing biological networks	Large-scale PPI networks analysis	Understanding cellular organization and exploring disease causes through a structural approach	Limitations to using this model in dynamic PPI networks
Amanatidou and Dedoussis [181]		Analyzing PPI network of non-alcoholic fatty liver disease	Analyze human protein interaction networks more effectively	Limitations for investigation of various aspects of PPI network
Liu and Chen [182]		Analyzes the disease network	Modeling community structure to better understand the disease network	Limitation of use in very large disease network
Dunn et al. [183]		Investigation of protein interaction networks	Create a way to screen protein interactions using community analysis in a fast manner	Suitable only for research on particular areas, and not for research that involve a larger proteome
Ostovari and Yu [185]	Analyzing healthcare workers network	Healthcare provider network analysis	Analyzing healthcare provider network attributes in the treatment of the chronic situations	The datasets used in this study are limited to claims and offer narrow data about providers
McCurdie et al. [186]		An investigation of interruptions in healthcare based on social network analysis	Utilization of the sociotechnical systems perspective to uncover the interdependencies between the healthcare work roles that result in interruptions in ICU.	This study does not capture the communications for each role.
Shirazi et al. [187]		Detection of the real specialty of physicians	Improve the accuracy of identifying physicians' real specialties	The use of only a few informative attributes
Flemming et al. [184]		Identify networks of physicians	Enhancing the identification of informal networks of ambulatory physicians by analyzing community structure	There is no analysis of how informal networks differ based on the patient networks

TABLE 2. (Continued.) Outlining the reviewed community detection-based method in healthcare applications.

Niyirora et al. [188]	Network analysis of healthcare services	Analyze the medical care services	An analysis of the network of medical care services and a representation and assessment of treatment patterns	The research was narrowed to a single network of medical care services
Sabot et al. [189],		Analyzing attributes of professional advice networks in Ethiopia's primary healthcare units	Helping to understand professional advice networks and their potential in healthcare	Protecting healthcare data privacy is an important challenge
De Brún et al. [190]		Better understanding and analyzing the structures of health systems	Demonstrating, diagnosing, and evaluating the healthcare services and networks	Actors outside the network and their influences are not considered
Shirazi et al. [191]		Identifying disease fields	Detecting disease areas efficiently with big data and SNA	Limitation for analyzing the large scale networks
Elghazel et al. [192]		Transforming heterogeneous healthcare data into operational information	Finding a new type of hospital stay by adopting community detection	Limitation for analyzing the large scale networks
Baktha et al et al. [196]	Analyzing marketing in healthcare agencies	Health service recommendation using users' behavioral analysis	Healthcare agencies can utilize community detection techniques to raise brand awareness, satisfaction, attract potential candidates, and ultimately build a healthier world	Ethical issues; lack of explainability
Schwartz et al. [197]	Characterizing lifestyle, behavioral health and well-being	Prediction of life satisfaction	Helping the healthcare providers for measuring and evaluating life satisfaction and well-being factors	Semantic and conceptual models are not used in well-being analysis
Zong et al. [198]		Health level classification	An effective method of predicting a user's health level using SNA	Developing a black box model for health level classification
Galan-Vasquez and Perez-Rueda [200]	Drug discovery	Analysis of drug-target interactions	Generate an improved understanding of how drugs interact with their targets.	Inability to model multilayer networks
Zhang et al. [201]		Core drug discovery	Detecting drug clusters and core drugs more efficiently with community detection algorithms	Due to the random nature of the algorithm used to detect communities, it may cause an instability issue.
Yanrui et al. [202]		Metabolic network analysis	Detection of metabolic communities efficiently	Lack of usability in real large networks
Gangopadhyay, and Chen [203]	Fraud Detection	Healthcare fraud detection	Accurate fraud discovery in the healthcare network; early detection of fraud	Ethical issues, permit access to healthcare data

Community detection algorithms based on traditional methods are typically classified into four categories: Partitional clustering [207], Hierarchical clustering [74], Spectral clustering [75], and Graph partitioning [208]. These traditional community detection methods have the advantage of being fast and easy to implement. The disadvantages of these methods include dependence on initial solutions and non-scalability for large social networks. Additionally, in many traditional community detection algorithms, it is needed to provide the number of communities in advance.

Traditional hierarchical clustering (e.g., k-means and fuzzy c-means) has been widely used in healthcare applications because of its simplicity and acceptable efficiency [173], [209], [210].

C. LABEL PROPAGATION ALGORITHM (LPA)

In Label Propagation-based and dynamical algorithm-based community detection methods, nodes with the same label tend to generate a group or a community [77], [211]. LPA is semi-supervised machine learning algorithm where each node carries a label of the community it belongs to, and then it uses existing node-label allocation to assign labels to unlabeled nodes using network structure until convergence occurs. LPA [77] is one of the most important agent-based community detection methods. LPA is a simple, efficient, and easy-to-implement method for detecting communities in large social networks. Their scheme utilizes only the graph structure as a guideline rather than a predetermined evaluation criterion and it needs neither prior community

knowledge nor optimization. Due to its low time (linear order) and spatial complexity (using local information), this community detection method is very popular. On the other hand, because of its random behavior, LPA can be unstable and produce less accurate community detection. Due to their simplicity and low computational complexity, LPA has been widely used for social network analysis in healthcare applications [201].

D. GIRVAN-NEWMAN ALGORITHM

In Modularity-based algorithms, optimizing techniques for analyzing network structure often use modularity function to detect communities. This modularity can be measured by the strength of its division into communities or clusters. The nodes in high modularity networks are densely connected within their respective community but sparsely connected between communities. It has been demonstrated that modularity suffers a resolution limit, so it is unable to distinguish small communities. Popular community detection algorithms based on optimization of such a modularity measure are Extremal optimization [212], [213], Spectral optimization [50], [163], Greedy optimization [82], [214], and Heuristics and Metaheuristics algorithms [215], [216], [217], [218].

Girvan–Newman algorithm, introduced by Girvan and Newman [79], is one of the greedy optimization-based algorithms used for the detection of communities. This community detection method detects and eliminates modules' connections and then measures the importance of edges using the edge betweenness criterion. Girvan–Newman algorithm includes four phases:

1. Compute betweenness values of all edges
2. Eliminate edge with the largest betweenness.
3. Recompute the betweenness of the other edges.
4. Iterate phases of 2) and 3)

Repeated calculations of the four previous steps are then used to construct a hierarchical clustering tree.

Girvan–Newman algorithm is more efficient as compared to traditional community detection methods. The time complexity of the calculation of modularity function in Girvan–Newman algorithm is $O(mn)$, where n is the number of nodes in the network and m the number of edges, respectively. Despite being understandable and simple, Girvan–Newman algorithm has its own challenges. With a social network containing many nodes and edges, this method is not very efficient. In large and complex networks, communities are hard to detect. Previous studies have used Girvan–Newman algorithm for social network analysis in healthcare applications due to its simplicity and acceptable performance [183], [202].

E. PCA-BASED ALGORITHM

The PCA-based community detection model utilizes the Principal Component Analysis (PCA) to select the best number of eigenvectors and then the Laplacian matrix to map the

nodes in a low-dimensional space [219], [220]. In [219], a novel PCA-based community detection algorithm is developed to identify the overlapping communities. In this method, first, the PCA technique is utilized for the extraction of the key features of network structure, and then the membership index is employed for node classification. This algorithm can detect communities accurately and obtain approximate measure. In some of the healthcare applications, a PCA-based community detection algorithm is employed for the analysis of disease-gene networks [182].

F. GRAPH COLORING ALGORITHM

The graph coloring algorithm is one of the most interesting graph-theoretic algorithms that can be used in many different domains [221], [222]. B-coloring is the procedure of assigning colors to nodes in each graph so that no two adjacent nodes have the same color, and each color class has at least one dominating node that is adjacent (dissimilar) to all the other color classes.

Graph coloring techniques are already widely used to identify communities in social networks [223], [224]. Assume that $G = (V, E)$ is a graph without loops that has a node set V and an edge set E . B-coloring for G is the coloring of nodes in such a way that:

Each pair of adjacent nodes $(x, y) \in G$ has different colors (proper coloring), and each color class has at least one node that has neighbors in all other color classes. These nodes are known as dominating nodes. A color that has a dominant node is called a dominant color.

$\varphi(G)$ is the largest number k in a graph G which has a b-coloring with k colors. A b-coloring is a maximum coloring of the vertices of graph G , as opposed to a minimal coloring. Therefore, $\varphi(G)$ is the degree over all the nodes of a graph G , and the degree of a node is the number of its neighbors. It has been shown that:

$$\chi(G) \leq \varphi(G) \leq \Delta(G) + 1 \quad (7)$$

where chromatic number $\chi(G)$ denotes to the minimum number of colors utilized to have a proper coloring of G . In some healthcare applications [192], a graph b-coloring is utilized for community detection. This algorithm provides a real representation of clusters based on a dominant object which is employed to evaluate cluster quality.

In Table 3, the strengths and weaknesses and computational complexity of previously implemented community detection algorithms in healthcare applications are discussed.

V. PUBLICLY AVAILABLE SOCIAL NETWORK HEALTHCARE DATASET

To evaluate and visualize the performance of most social network analysis techniques, the datasets are crucial. In this section, some of the publicly available social network healthcare datasets are described and the characteristics of these datasets are summarized in Table 4.

TABLE 3. Summary of implemented community detection algorithms in healthcare applications.

Algorithm	References	Advantages	Weaknesses	Complexity	Applications
Louvain Algorithm	[174, 187, 191]	Simple, easy to implement, and high performance.	Eliminate smaller communities and generate communities containing significant sub-communities; Detect arbitrarily badly connected communities	$O(n \log n)$	Healthcare event monitoring and disease diagnosis, network analysis of healthcare services and analyzing healthcare workers network
Traditional Hierarchical clustering (k-means and c-means)	[173, 209, 210]	Fast and easy to implement.	Need to determine the number of communities; Dependence on initial solutions and non-scalability for large social networks	$O(n^2)$	Healthcare event monitoring and disease diagnosis, network analysis of healthcare services
Label Propagation Algorithm	[201]	Simplicity and low computational complexity	Random behavior, unstable and low performance	$O(n)$	Drug discovery
Girvan-Newman Algorithm	[183, 202]	Simplicity and acceptable performance	Not suitable for the very large social network, High computational complexity in complex networks	$O(mn)$	Analyzing biological networks and drug discovery
PCA-based Algorithm	[182]	High performance	High computational complexity in complex networks	$O(m^2n)$	Analyzing biological networks
Graph Coloring Algorithm	[192]	Real representation of clusters; determine the optimal number of clusters automatically	High execution time	$O(m^2 + n)$	Network analysis of healthcare services

TABLE 4. Summary of publicly available social network healthcare dataset.

Dataset	Details	Size	Application	Reference
ASUS/CNES	Historical records of Brazilian healthcare professionals	628 healthcare units	Analyzing Healthcare Workers Network	[225]
Drug Reviews	Patient reviews on specific drugs	215063 reviews from Drugs.com and 3551 reviews from Druglib.com.	Network Analysis of Healthcare Services	[226]
HCOS	Social network of physicians prescribing pain medications	500,000 organization and 1.1 million distinct professionals	Analyzing Healthcare Workers Network	[227]
National Neighborhood Dataset	Diet and physical activity dataset	Collected tweets related to diet and physical activity	Healthcare Monitoring and Disease Diagnosis, Network Analysis of Healthcare Services,	[228]
PMSI	Patient Medical records	26.5 million acute care hospital stays	Healthcare Monitoring and Disease Diagnosis, Network Analysis of Healthcare Services,	[229]
MIMIC-III	Patients admitted to a hospital's critical care units.	49,785 admitted patients	Healthcare Monitoring and Disease Diagnosis, Network Analysis of Healthcare Services, Analyzing Healthcare Workers Network	[230]

A. DATASUS/CNES

DATASUS/CNES [225] is a Brazilian public health dataset that includes the historical records of Brazilian healthcare professionals, available with a delay of up to 45 days, per month of competence. The city of Bauru has 628 health-

care units, containing seven public hospitals and three specialized hospitals. The samples of this dataset include all the physicians that work in these public healthcare units and have employment contracts with other public and private healthcare units. As a result, the graph representation

of the physicians with employment contracts resulted in these public and private healthcare units in the city of Bauru and other regions. Based on the DATASUS/CNES information, a database named DDB (Detailed Data Bank) was created, which contained the following information: healthcare units, physician's code identification, Brazilian classification of occupation (BCO), public and private spheres identification, workload, type of employment contract within the healthcare unit, and physician's occupation and location.

B. DRUG REVIEWS

Drug Reviews [226] dataset prepares patient reviews on particular drugs along with associated characteristics. In this dataset, the 10-star rating represents the overall satisfaction of the patient. Data was collected by crawling online pharmaceutical review sites. The intention was to study (1) Sentiment analysis of drug experience over a variety of aspects (i.e., sentiments learned on peculiar aspects such as efficacy and side effects), (2) Model transferability across domains, and (3) The ability to transfer models between different data resources. This dataset for drug side effect prediction was acquired from the UCI repository.

C. HEALTHCARE ORGANIZATION SERVICES (HCOS)

Medical-prescriptions dataset and the Healthcare Organization Services (HCOS) [227] is a public physician-affiliation dataset in healthcare applications that can be used to generate a social network of physicians prescribing pain medications. The prescription dataset contained billions of physicians in the United States prescribing certain pain medications between the years 1996 and 2017. The physician has an affiliation to a plurality of hospitals, and each hospital has an affiliation with an integrated delivery network (IDN). An IDN is an official system of providers and hospitals that offers both healthcare services and a health insurance plan to patients.

D. NATIONAL NEIGHBORHOOD DATASET

National Neighborhood Dataset [228] is happiness, diet, and physical activity dataset in healthcare applications. Between February 2015 and March 2016, a random 1% sample of publicly available tweets with latitude and longitude coordinates is continuously collected using Twitter's streaming application programming interface (API). The authors of this study constructed neighborhood indicators at both census tract and zip code levels in order to address the potential data needs of neighborhood researchers. These researchers tend to use and be interested in data at the census tract level differently than at the zip code level.

E. PMSI

French national hospital promoted a database called PMSI (Programme de Médicalisation des Systèmes d'Information).

It contains information related to all patients hospitalized for metastatic melanoma. PMSI is a publicly available social network healthcare dataset that contains records of all hospitalizations, including detailed medico-administrative information in France in acute care (medicine, surgery, or obstetrics), rehabilitation centers providing short-term care, and home care, including public and private units. Approximately 26.5 million acute care hospital stays were documented in the PMSI database in 2013. An anonymous patient identifier can be used to follow a patient across multiple hospitalizations. The diagnostic codes of the 10th Revision of the International Classification of Diseases (ICD-10) were used to identify the reasons for hospitalization for each diagnosis related groups. Researchers can access this database free of charge on request, upon contacting the relevant agency and additional considerations [229].

F. MIMIC-III

The Medical Information Mart for Intensive Care (MIMIC) is a publicly available large-scale dataset that includes information related to patients admitted to intensive care units and contains data such as vital signs, medications, observations and notes, diagnostic codes, etc. The data in each case contain signals and periodic measurements acquired from a bedside monitor as well as clinical data acquired from the patient's health record. For example MIMIC-III [230] data was collected between 2001 and 2012 at Beth Deaconess Medical Center in Boston, Massachusetts. The MIMIC-III includes more data regarding patients who are seriously ill or injured than, for instance, data from routine checkups for those admitted to ICU.

VI. OPEN ISSUES AND CHALLENGES

Community detection is still an active research field with numerous open issues. In this section, some open issues and challenges in community detection in healthcare applications are discussed. These challenges, their fields, the health disciplines involved, and potential approaches to overcome them are summarized in Table 5.

A. COMMUNITY DETECTION IN MULTILAYER BIOLOGICAL NETWORKS

The structure of various biological systems has been modeled using networks. By manipulating signals, biological network research aims to steer biological system structures toward desired states. The linear control of single-layer networks has advanced significantly over the past few decades, but many complex biological systems exhibit highly complicated nonlinear processes and a complex multilayer network structure. Genetic, biochemical, and environmental factors can be linked to human diseases through a multilayer network. In recent years, the healthcare and biological applications of modeling based on multilayer networks have received more attention than before. These networks have been studied because different layers can demonstrate users' communica-

TABLE 5. Summary of community detection challenges in healthcare applications.

Challenges	Challenges	Reason of Challenge	Health disciplines involved	Potential approach
Community Detection in Multilayer Biological Networks	Efficiency	Most traditional algorithms cannot perform well on a multilayer network without modification.	Biological complex network analysis (e.g., protein-protein interaction network, human brain networks analysis, network neuroscience analysis), Food-Drug interaction analysis	Formulation of the multilayer social network as multi-objective and many-objective optimization problems.
Number of Communities	Efficiency	A key limitation of many of the previous community detection algorithms is that the number of communities should be determined as an input parameter.	Analyzing healthcare workers network, healthcare monitoring and cancer diagnosis/prediction	Modularity maximization, statistical inference
Deep Learning-based Methods	Computational Complexity	Due to prohibitive demands on memory and computation, many existing deep learning-based community detection algorithms are ineffective when dealing with large scale social network.	Healthcare monitoring and disease diagnosis, analyzing biological networks, network analysis of healthcare services, drug discovery	Parallel computing using GPUs, Metaheuristic algorithms
Ethical Issues of SNA Usage in Healthcare	Ethical challenges	Many guidelines for using social network analysis in healthcare applications are still incomplete and often conflicting	Healthcare monitoring and disease diagnosis, analyzing and tracking outbreaks, analyzing healthcare workers network, analyzing marketing in healthcare agencies	Redetermine the boundaries of the investigation

tions from different point of views. However, their analysis raised a new challenge concerning the application of different layers to detect different communities. Community detection in multilayer networks is especially important in representing healthcare networks where each node has multiple tasks and relationships. Complex network systems prepare powerful tools and techniques to study biological fields from interactions between genes and proteins, to the investigation of disease and organ functions, and even brain activity research areas. Studying community detection in multilayer networks has more issues to target. Especially, one of the main issues in multilayer networks concerns how to merge the different layers, which model node-node communications from different perspectives. Furthermore, there are no standards for multi-layer networks that can be used for evaluation. To overcome this challenge, the problem of community detection in a multilayer network can effectively be addressed by multi-objective and many-objective optimization models. For this purpose, multi/many-objective mathematical functions can be developed which consider the different aspects of a community structure, simultaneously.

B. NUMBER OF COMMUNITIES

In many of the previous community detection algorithms, the number of communities should be determined as an input parameter, reducing the efficiency of community detection as an intelligent mechanism to analyze the network structure.

In practice, there are two techniques to estimate the optimal number of communities. The first technique uses modularity maximization, which itself can also be employed for community detection. The second one uses a statistical inference-based approach. Due to the difficulty of estimating the optimal number of communities in healthcare applications, future research needs to pay more attention to this challenge.

C. DEEP LEARNING-BASED METHODS

The emergence of large social networks is observed in the last recent years. Deep learning is a vital technique to handle these large social networks and the extraction of useful knowledge from them. However, because of high computational complexity and storage requirements, the complexity of deep learning-based approaches is increased considerably. Deep learning models require high-performance computing machines to be developed at large social networks. However, designing precise and efficient deep learning-based community detection models, especially when dealing with large social network, is always a continuous study effort to pursue. Metaheuristic methods have shown their potential to increase deep learning performance in terms of efficiency (i.e., computational complexity) and effectiveness (i.e., accuracy). Moreover, parallel computing is used to improve the efficiency of deep learning approaches trained on the large social network by splitting a task into subtasks and performing them simultaneously. Deep learning approaches are more efficient through

parallel computing by using computers with graphical processing units (GPUs). Furthermore, metaheuristic methods have shown their potentials in improving deep learning models in terms of solution quality and computational cost.

D. ETHICAL ISSUES OF SNA USAGE IN HEALTHCARE

Today, SNA research in healthcare applications faces unprecedented ethical challenges due to a combination of technological development (big data) and increasing attention towards institutionalizing ethic governance. Ethics is the study of right and wrong, as well as moral duty and obligation. Public health ethics focuses on the moral aspects of public health interventions for disease prevention, life extension, or psychological and physical well-being. Medical ethics, on the other hand, focuses on the relationship between doctors and patients. Health professionals and patients alike face the challenge of how ethical principles may be applied to SNA and community detection algorithms in healthcare studies. Low relationship obstacles in social networks, limited privacy, and security challenges lead to issues from an ethical aspect. Therefore, for ethical reasons, the boundaries of the investigation should be redetermined and only cooperation among those stakeholders who consented to collaborate in the research should finally be investigated.

VII. CONCLUSION

In the past few years, routine data sources in healthcare databases have become increasingly fast. Health informatics generates huge amounts of data, and analyzing these voluminous data creates new opportunities for knowledge discovery. In various scientific fields including healthcare, there has been an increase in the use of SNA. Social networks can be utilized in healthcare applications because they offer huge data volumes and continuous monitoring, which enable rapid response time. Community detection is an essential SNA technique widely used in the healthcare applications. The community detection approach seeks to identify clusters of nodes that are similar according to some similarity criteria. From a general perspective, existing algorithms for community detection can be divided into two groups of non-attributed and attributed networks, which include traditional models and dimensionality reduction algorithms. Here, a categorization of different community detection algorithms is presented, while the recent and state-of-the-art methods are reviewed and compared. We categorized most applications of community detection in healthcare into nine groups as: Healthcare Event Monitoring and Disease Diagnosis, Analyzing and Tracking Outbreaks, Analyzing Healthcare Workers Network, Network Analysis of Healthcare Services, Analyzing Marketing in Healthcare Agencies, Characterizing Lifestyle, Behavioral Health and Well-being, Drug Discovery, and Fraud Detection. We also discussed four important open issues and challenges related to community detection in healthcare applications, such as Community Detection in Multilayer Networks, the Number of Communities, Deep Learning-based Methods, and Ethical Issues related to

SNA usage in healthcare. Our study also explores potential approaches to overcome such challenges.

VIII. FUTURE WORKS

Social network analysis plays a major role in crisis management, since it can potentially identify areas in danger, investigate the outbreak events, generate summarized reports, and anticipate future requirements, contributing greatly to situational awareness and disaster response. When the crisis is an outbreak of infection or a virus pandemic, SNA can be employed to anticipate the outbreak, monitor the spread of the disease, and check the state of the infection. Due to the extensive application of community detection methods for precise analyzing of online social network, there is a prospect that researchers can use community detection methods to identify high-risk infectious disease communities and monitor infection spread. For example, high-level knowledge can be found with online SNA using community detection methods. Moreover, using SNA and community detection, various interactions and health issues that are observed in society/localities can be modeled accordingly to better comprehend the spread of diseases and effects of policy.

While novel and more especially deep learning-based community detection algorithms can be used to easily identify communities in healthcare applications, there is still one significant barrier to the use of these techniques in the healthcare sectors. This relates to the explainability problem, which occurs when healthcare authorities are reluctant to trust results that are hardly transparent or explainable. The lack of explainability and transparency is a major limitation of community detection and SNA techniques for healthcare applications. There are many healthcare applications that require understanding how a community detection algorithm made its analysis to build trust among stakeholders (e.g., physicians, specialists, patients, researchers, and the general public). In short, eXplainable Artificial Intelligence (XAI) can provide a human-understandable explanation of its behavior. Despite limited success, a variety of techniques have been proposed adding transparency to artificial intelligence in a way that earns clinician trust and yields positive clinical outcomes. The incorporation of interpretable features in a post-hoc explanation is an open area for research for improving community detection methods in future works. The goal is to create a set of easily understandable features for end-users and healthcare stakeholders.

REFERENCES

- [1] F. Schwendicke and J. Krois, "Data dentistry: How data are changing clinical care and research," *J. Dental Res.*, vol. 101, no. 1, pp. 21–29, 2022.
- [2] S. Kalkman, J. van Delden, A. Banerjee, B. Tyl, M. Mostert, and G. van Thiel, "Patients' and public views and attitudes towards the sharing of health data for research: A narrative review of the empirical evidence," *J. Med. Ethics*, vol. 48, no. 1, pp. 3–13, 2022.
- [3] C. A. Kronk, A. R. Everhart, F. Ashley, H. M. Thompson, T. E. Schall, T. G. Goetz, and L. Hiatt, "Transgender data collection in the electronic health record: Current concepts and issues," *J. Amer. Med. Inform. Assoc.*, vol. 29, no. 2, pp. 271–284, 2022.

- [4] M. E. Tschuchnig and M. Gadermayr, "Anomaly detection in medical imaging—A mini review," in *Data Science-Analytics and Applications*. Wiesbaden, Germany: Springer, Mar. 2022, pp. 33–38.
- [5] M. Mahbub, S. Srinivasan, I. Danciu, A. Peluso, E. Begoli, S. Tamang, and G. D. Peterson, "Unstructured clinical notes within the 24 hours since admission predict short, mid & long-term mortality in adult ICU patients," *PLoS ONE*, vol. 17, no. 1, 2022, Art. no. e0262182.
- [6] A. C. Langedijk, E. R. Harding, B. Konya, B. Vrancken, R. J. Lebbink, A. Evers, J. Willemsen, P. Lemey, and L. J. Bont, "A systematic review on global RSV genetic data: Identification of knowledge gaps," *Rev. Med. Virol.*, vol. 32, no. 3, pp. 1–10, May 2022.
- [7] W. Guideline, "Recommendations on digital interventions for health system strengthening," World Health Org., Geneva, Switzerland, Tech. Rep., 10-2020, 2019.
- [8] D. Javaheri, P. Lalbakhsh, S. Gorgin, J.-A. Lee, and M. Masdari, "A new energy-efficient and temperature-aware routing protocol based on fuzzy logic for multi-WBANs," *Ad Hoc Netw.*, vol. 139, Feb. 2023, Art. no. 103042, doi: [10.1016/j.adhoc.2022.103042](https://doi.org/10.1016/j.adhoc.2022.103042).
- [9] B. Alamri, K. Crowley, and I. Richardson, "Blockchain-based identity management systems in health IoT: A systematic review," *IEEE Access*, vol. 10, pp. 59612–59629, 2022.
- [10] A. Zeng, Z. Shen, J. Zhou, J. Wu, Y. Fan, Y. Wang, and H. E. Stanley, "The science of science: From the perspective of complex systems," *Phys. Rep.*, vols. 714–715, pp. 1–73, Nov. 2017.
- [11] C. Webb and O. Bodin, "A network perspective on modularity and control of flow in robust systems," in *Complexity Theory for a Sustainable Future*, J. Norberg and G. Cumming, Eds. New York, NY, USA: Columbia Press, 2008.
- [12] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1994.
- [13] A. C. Krendl and R. F. Betzel, "Social cognitive network neuroscience," *Social Cognit. Affect. Neurosci.*, vol. 17, no. 5, pp. 510–529, May 2022.
- [14] L. E. Young, J. L. Tang, and Y. Nan, "Social media communication and network correlates of HIV infection and transmission risks among black sexual minority men: Cross-sectional digital epidemiology study," *JMIR Formative Res.*, vol. 6, no. 10, Oct. 2022, Art. no. e37982.
- [15] G. Huang, M. Sun, and L. C. Jiang, "Core social network size is associated with physical activity participation for fitness app users: The role of social comparison and social support," *Comput. Hum. Behav.*, vol. 129, Apr. 2022, Art. no. 107169.
- [16] S. A. Rains and J. Meng, "Social enhancement and compensation in online social support among cancer patients: The role of social network properties," *Health Commun.*, vol. 37, no. 4, pp. 490–497, Mar. 2022.
- [17] M. Salas, J. Petracek, P. Yalamanchili, O. Aimer, D. Kasthuril, S. Dhingra, T. Junaid, and T. Bostic, "The use of artificial intelligence in pharmacovigilance: A systematic review of the literature," *Pharmaceutical Med.*, vol. 36, pp. 1–12, Jul. 2022.
- [18] S. Jellema, K. Bakker, M. W. G. Nijhuis-van der Sanden, R. van der Sande, and E. M. Steultjens, "The role of the social network during inpatient rehabilitation: A qualitative study exploring the views of older stroke survivors and their informal caregivers," *Topics Stroke Rehabil.*, vol. 29, no. 1, pp. 30–39, Jan. 2022.
- [19] S. Bhambere, "Social network structure and data analysis in healthcare," in *Social Network Analysis: Theory and Applications*. Hoboken, NJ, USA: Wiley, 2022, pp. 63–82.
- [20] *The World Health Report 2000: Health Systems: Improving Performance*, World Health Organization, Geneva, Switzerland, 2000.
- [21] J. L. Moreno, *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Socio-Drama*, 2nd ed. Washington, DC, USA: Beacon House, American Psychological Association, 1953.
- [22] G. Bloom and H. Standing, "Future health systems: Why future? Why now?" *Social Sci. Med.*, vol. 66, no. 10, pp. 2067–2075, May 2008.
- [23] W. Jo, D. Chang, M. You, and G.-H. Ghim, "A social network analysis of the spread of COVID-19 in South Korea and policy implications," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, Apr. 2021.
- [24] T. Kuchler, D. Russel, and J. Stroebel, "JUE insight: The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook," *J. Urban Econ.*, vol. 127, Jan. 2022, Art. no. 103314.
- [25] H. Hu, Y. Yang, C. Zhang, C. Huang, X. Guan, and L. Shi, "Review of social networks of professionals in healthcare settings—where are we and what else is needed?" *Globalization Health*, vol. 17, no. 1, pp. 1–17, 2021.
- [26] S. Zhang, K. de la Haye, M. Ji, and R. An, "Applications of social network analysis to obesity: A systematic review," *Obesity Rev.*, vol. 19, no. 7, pp. 976–988, Jul. 2018.
- [27] E. Sukkar, "Searching social networks to detect adverse reactions," *Pharmaceutical J.*, vol. 294, pp. 75–78, Jan. 2015.
- [28] A. D. Faber and S. Wasserman, "Social support and social networks: Synthesis and review," in *Social Networks and Health (Advances in Medical Sociology)*, vol. 8, J. A. Levy and B. A. Pescosolido, Eds. Bingley, U.K.: Emerald Group Publishing Limited, 2002, pp. 29–72.
- [29] Z. Ghalmane, C. Cherifi, H. Cherifi, and M. E. Hassouni, "Extracting modular-based backbones in weighted networks," *Inf. Sci.*, vol. 576, pp. 454–474, Oct. 2021, doi: [10.1016/j.ins.2021.06.087](https://doi.org/10.1016/j.ins.2021.06.087).
- [30] D. Camacho, Á. Panizo-Lledot, G. Bello-Organ, A. Gonzalez-Pardo, and E. Cambria, "The four dimensions of social network analysis: An overview of research methods, applications, and software tools," *Inf. Fusion*, vol. 63, pp. 88–120, Nov. 2020, doi: [10.1016/j.inffus.2020.05.009](https://doi.org/10.1016/j.inffus.2020.05.009).
- [31] G. Bello-Organ, S. Salcedo-Sanz, and D. Camacho, "A multi-objective genetic algorithm for overlapping community detection based on edge encoding," *Inf. Sci.* vol. 462, pp. 290–314, Sep. 2018, doi: [10.1016/j.ins.2018.06.015](https://doi.org/10.1016/j.ins.2018.06.015).
- [32] V. Moscato and G. Sperli, "Community detection over feature-rich information networks: An eHealth case study," *Inf. Syst.*, vol. 109, Nov. 2022, Art. no. 102092.
- [33] I. Inuwa-Dutse, M. Liptrott, and I. Korkontzelos, "A multilevel clustering technique for community detection," *Neurocomputing*, vol. 441, pp. 64–78, Jun. 2021, doi: [10.1016/j.neucom.2021.01.059](https://doi.org/10.1016/j.neucom.2021.01.059).
- [34] W. Zhao, J. Luo, T. Fan, Y. Ren, and Y. Xia, "Analyzing and visualizing scientific research collaboration network with core node evaluation and community detection based on network embedding," *Pattern Recognit. Lett.*, vol. 144, pp. 54–60, Apr. 2021, doi: [10.1016/j.patrec.2021.01.007](https://doi.org/10.1016/j.patrec.2021.01.007).
- [35] K. Sabot, D. Wickremasinghe, K. Blanchet, B. Avan, and J. Schellenberg, "Use of social network analysis methods to study professional advice and performance among healthcare providers: A systematic review," *Systematic Rev.*, vol. 6, no. 1, p. 208, Oct. 2017, doi: [10.1186/s13643-017-0597-1](https://doi.org/10.1186/s13643-017-0597-1).
- [36] R. C. Shelton, M. Lee, L. E. Brotzman, D. M. Crookes, L. Jandorf, D. Erwin, and E. A. Gage-Bouchard, "Use of social network analysis in the development, dissemination, implementation, and sustainability of health behavior interventions for adults: A systematic review," *Social Sci. Med.*, vol. 220, pp. 81–101, Jan. 2019, doi: [10.1016/j.socscimed.2018.10.013](https://doi.org/10.1016/j.socscimed.2018.10.013).
- [37] O. Kaiwartya, A. H. Abdullah, Y. Cao, A. Altameem, M. Prasad, C. T. Lin, and X. Liu, "Guidelines for performing systematic literature reviews in software engineering," Keele Univ., Keele, U.K., Durham Univ., Durham, U.K., Joint Rep., 2016.
- [38] D. Chambers, P. Wilson, C. Thompson, and M. Harden, "Social network analysis in healthcare settings: A systematic scoping review," *PLoS ONE*, vol. 7, no. 8, Aug. 2012, Art. no. e41911.
- [39] M. A. Javed, M. S. Younis, S. Latif, J. Qadir, and A. Baig, "Community detection in networks: A multidisciplinary review," *J. Netw. Comput. Appl.*, vol. 108, pp. 87–111, Apr. 2018, doi: [10.1016/j.jnca.2018.02.011](https://doi.org/10.1016/j.jnca.2018.02.011).
- [40] P. Chunaev, "Community detection in node-attributed social networks: A survey," *Comput. Sci. Rev.*, vol. 37, Aug. 2020, Art. no. 100286, doi: [10.1016/j.cosrev.2020.100286](https://doi.org/10.1016/j.cosrev.2020.100286).
- [41] V. Moscato and G. Sperli, "A survey about community detection over on-line social and heterogeneous information networks," *Knowl.-Based Syst.*, vol. 224, Jul. 2021, Art. no. 107112, doi: [10.1016/j.knosys.2021.107112](https://doi.org/10.1016/j.knosys.2021.107112).
- [42] M. N. Al-Andoli, S. C. Tan, W. P. Cheah, and S. Y. Tan, "A review on community detection in large complex networks from conventional to deep learning methods: A call for the use of parallel meta-heuristic algorithms," *IEEE Access*, vol. 9, pp. 96501–96527, 2021, doi: [10.1109/ACCESS.2021.3095335](https://doi.org/10.1109/ACCESS.2021.3095335).
- [43] B. A. Attea, A. D. Abbood, A. A. Hasan, C. Pizzuti, M. Al-Ani, S. Özdemir, and R. D. Al-Dabbagh, "A review of heuristics and meta-heuristics for community detection in complex networks: Current usage, emerging development and future directions," *Swarm Evol. Comput.*, vol. 63, Jun. 2021, Art. no. 100885, doi: [10.1016/j.swevo.2021.100885](https://doi.org/10.1016/j.swevo.2021.100885).
- [44] X. Zhao, J. Liang, and J. Wang, "A community detection algorithm based on graph compression for large-scale social networks," *Inf. Sci.*, vol. 551, pp. 358–372, Apr. 2021, doi: [10.1016/j.ins.2020.10.057](https://doi.org/10.1016/j.ins.2020.10.057).

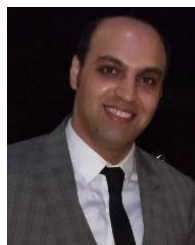
- [45] Z. Wu, X. Wang, W. Fang, L. Liu, S. Tang, H. Zheng, and Z. Zheng, "Community detection based on first passage probabilities," *Phys. Lett. A*, vol. 390, Feb. 2021, Art. no. 127099, doi: 10.1016/j.physleta.2020.127099.
- [46] Y. Wang, J. Cao, Z. Bu, J. Jiang, and H. Chen, "Proximity-based group formation game model for community detection in social network," *Knowl.-Based Syst.*, vol. 214, Feb. 2021, Art. no. 106670, doi: 10.1016/j.knsys.2020.106670.
- [47] H. Zardi, B. Alharbi, W. Karamti, H. Karamti, and E. Alabdulkreem, "Detection of community structures in dynamic social networks based on message distribution and structural/attribute similarities," *IEEE Access*, vol. 9, pp. 67028–67041, 2021, doi: 10.1109/ACCESS.2021.3076594.
- [48] B. Sun, R. Al-Bayaty, Q. Huang, and D. Wu, "Game theoretical approach for non-overlapping community detection," *Tsinghua Sci. Technol.*, vol. 26, no. 5, pp. 706–723, Oct. 2021, doi: 10.26599/TST.2020.9010017.
- [49] D. He, D. Jin, C. Baquero, and D. Liu, "Link community detection using generative model and nonnegative matrix factorization," *PLoS ONE*, vol. 9, no. 1, Jan. 2014, Art. no. e86899.
- [50] M. Chen, K. Kuzmin, and B. K. Szymanski, "Community detection via maximization of modularity and its variants," *IEEE Trans. Computat. Social Syst.*, vol. 1, no. 1, pp. 46–65, Mar. 2014, doi: 10.1109/TCSS.2014.2307458.
- [51] S. X. Wu, Z. Wu, S. Chen, G. Li, and S. Zhang, "Community detection in blockchain social networks," *J. Commun. Inf. Netw.*, vol. 6, no. 1, pp. 59–71, Mar. 2021, doi: 10.23919/JCIN.2021.9387705.
- [52] S. Mittal, D. Sengupta, and T. Chakraborty, "Hide and seek: Outwitting community detection algorithms," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 4, pp. 799–808, Aug. 2021, doi: 10.1109/TCSS.2021.3062711.
- [53] J. Reichardt and S. Bornholdt, "Detecting fuzzy community structures in complex networks with a Potts model," *Phys. Rev. Lett.*, vol. 93, no. 21, Nov. 2004, Art. no. 218701, doi: 10.1103/PhysRevLett.93.218701.
- [54] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 1, 2006, Art. no. 016110, doi: 10.1103/PhysRevE.74.016110.
- [55] S. Azadifar, M. Rostami, K. Berahmand, P. Moradi, and M. Oussalah, "Graph-based relevancy-redundancy gene selection method for cancer diagnosis," *Comput. Biol. Med.*, vol. 147, Aug. 2022, Art. no. 105766, doi: 10.1016/j.combiomed.2022.105766.
- [56] D. Hric, R. K. Darst, and S. Fortunato, "Community detection in networks: Structural communities versus ground truth," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 90, no. 6, Dec. 2014, Art. no. 062805.
- [57] J. Banks, S. Mohanty, and P. Raghavendra, "Local statistics, semidefinite programming, and community detection," in *Proc. ACM-SIAM Symp. Discrete Algorithms (SODA)*, 2021, pp. 1298–1316.
- [58] J. Fumanal-Idocin, A. Alonso-Betanzos, O. Cordon, H. Bustince, and M. Minárová, "Community detection and social network analysis based on the Italian wars of the 15th century," *Future Gener. Comput. Syst.*, vol. 113, pp. 25–40, Dec. 2020.
- [59] J. Sheng, J. Hu, Z. Sun, B. Wang, A. Ullah, K. Wang, and J. Zhang, "Community detection based on human social behavior," *Phys. A, Stat. Mech. Appl.*, vol. 531, Oct. 2019, Art. no. 121765.
- [60] D. Liu, G. Yang, Y. Wang, H. Jin, and E. Chen, "How to protect ourselves from overlapping community detection in social networks," *IEEE Trans. Big Data*, vol. 8, no. 4, pp. 894–904, Aug. 2022.
- [61] Q. Zhu, J. Yang, S. Wang, D. Zhang, and Z. Zhang, "Multi-modal non-Euclidean brain network analysis with community detection and convolutional autoencoder," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, May 12, 2022, doi: 10.1109/TETCI.2022.3171855.
- [62] Z. Zhong, X. Wang, C. Qu, and G. Wang, "Efficient algorithm based on non-backtracking matrix for community detection in signed networks," *IEEE Trans. New. Sci. Eng.*, vol. 9, no. 4, pp. 2200–2211, Jul. 2022.
- [63] R. E. Mansano, L. E. Allem, R. R. Del-Vecchio, and C. Hoppen, "Balanced portfolio via signed graphs and spectral clustering in the Brazilian stock market," *Quality Quantity*, vol. 56, no. 4, pp. 2325–2340, Aug. 2022.
- [64] Y. Zhen and J. Wang, "Community detection in general hypergraph via graph embedding," *J. Amer. Stat. Assoc.*, vol. 2022, pp. 1–10, Jan. 2022.
- [65] M. Li, S. Lu, L. Zhang, Y. Zhang, and B. Zhang, "A community detection method for social network based on community embedding," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 2, pp. 308–318, Apr. 2021, doi: 10.1109/TCSS.2021.3050397.
- [66] S. Yazdanparast, T. C. Havens, and M. Jamalabdollahi, "Soft overlapping community detection in large-scale networks via fast fuzzy modularity maximization," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 6, pp. 1533–1543, Jun. 2021, doi: 10.1109/TFUZZ.2020.2980502.
- [67] D. Zhuang, M. J. Chang, and M. Li, "DynaMo: Dynamic community detection by incrementally maximizing modularity," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 5, pp. 1934–1945, May 2021, doi: 10.1109/TKDE.2019.2951419.
- [68] H. Zare, M. Hajiabadi, and M. Jalili, "Detection of community structures in networks with nodal features based on generative probabilistic approach," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 7, pp. 2863–2874, Jul. 2021, doi: 10.1109/TKDE.2019.2960222.
- [69] X. Wang, J. Li, L. Yang, and H. Mi, "Unsupervised learning for community detection in attributed networks based on graph convolutional network," *Neurocomputing*, vol. 456, pp. 147–155, Oct. 2021, doi: 10.1016/j.neucom.2021.05.058.
- [70] M. Qin and K. Lei, "Dual-channel hybrid community detection in attributed networks," *Inf. Sci.*, vol. 551, pp. 146–167, Apr. 2021, doi: 10.1016/j.ins.2020.11.010.
- [71] X. Xueshuo, W. Jiming, Y. Junyi, F. Yaozheng, L. Ye, L. Tao, and W. Guiling, "AWAP: Adaptive weighted attribute propagation enhanced community detection model for Bitcoin de-anonymization," *Appl. Soft Comput.*, vol. 109, Sep. 2021, Art. no. 107507, doi: 10.1016/j.asoc.2021.107507.
- [72] H. Gmati, A. Mouakher, A. Gonzalez-Pardo, and D. Camacho, "A new algorithm for communities detection in social networks with node attributes," *J. Ambient Intell. Humanized Comput.*, vol. 2018, pp. 1–13, Oct. 2018, doi: 10.1007/s12652-018-1108-5.
- [73] E. Akbas and P. Zhao, "Attributed graph clustering: An attribute-aware graph embedding approach," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Jul. 2017, pp. 305–308.
- [74] L. Zhang, Q. Ye, Y. Shao, C. Li, and H. Gao, "An efficient hierarchy algorithm for community detection in complex networks," *Math. Problems Eng.*, vol. 2014, pp. 1–12, Jan. 2014.
- [75] M. Meilä and J. Shi, "A random walks view of spectral segmentation," in *Proc. Int. Workshop Artif. Intell. Statist.*, 2021, pp. 203–208.
- [76] P. Pons and M. Latapy, "Computing communities in large networks using random walks," in *Proc. Int. Symp. Comput. Inf. Sci.* Cham, Switzerland: Springer, 2005, pp. 284–293.
- [77] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 76, no. 3, 2007, Art. no. 036106.
- [78] K. Subramani, A. Velkov, I. Ntoutsis, P. Kroger, and H.-P. Kriegel, "Density-based community detection in social networks," in *Proc. IEEE 5th Int. Conf. Internet Multimedia Syst. Archit. Appl.*, Dec. 2011, pp. 1–8.
- [79] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, 2001.
- [80] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, "Defining and identifying communities in networks," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 9, pp. 2658–2663, 2004.
- [81] S. Fortunato, V. Latora, and M. Marchiori, "Method to find community structures based on information centrality," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 70, no. 5, Nov. 2004, Art. no. 056104.
- [82] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 6, 2004, Art. no. 066133, doi: 10.1103/PhysRevE.69.066133.
- [83] R. Guimera, M. Sales-Pardo, and L. A. N. Amaral, "Modularity from fluctuations in random graphs and complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 70, no. 2, Aug. 2004, Art. no. 025101.
- [84] H. S. Shon, S. Kim, C. S. Rhee, and K. H. Ryu, "Clustering DNA microarray data by MCL algorithm," in *Proc. Intell. Syst. Mol. Biol. (ISMB)*, 2007.
- [85] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 2, pp. 1118–1123, Jan. 2008.
- [86] K. Berahmand and A. Bouyer, "A link-based similarity for improving community detection based on label propagation algorithm," *J. Syst. Sci. Complex.*, vol. 32, no. 3, pp. 737–758, Jun. 2019.
- [87] 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. 6, Mar. 2018, Art. no. 1850062.

- [88] T. Falkowski, A. Barth, and M. Spiliopoulou, "DENGRAPH: A density-based community detection algorithm," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Nov. 2007, pp. 112–115.
- [89] M. Gong, J. Liu, L. Ma, Q. Cai, and L. Jiao, "Novel heuristic density-based method for community detection in networks," *Phys. A, Stat. Mech. Appl.*, vol. 403, pp. 71–84, Jun. 2014.
- [90] H. Sun, J. Huang, J. Han, H. Deng, P. Zhao, and B. Feng, "GSkeletonClu: Density-based network clustering via structure-connected tree division or agglomeration," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2010, pp. 481–490.
- [91] R. A. Rossi, D. Jin, S. Kim, N. K. Ahmed, D. Koutra, and J. B. Lee, "On proximity and structural role-based embeddings in networks: Misconceptions, techniques, and applications," 2019, *arXiv:1908.08572*.
- [92] K. Taşdemir, "Vector quantization based approximate spectral clustering of large datasets," *Pattern Recognit.*, vol. 45, no. 8, pp. 3034–3044, 2012, doi: [10.1016/j.patcog.2012.02.012](https://doi.org/10.1016/j.patcog.2012.02.012).
- [93] Y. Ge, P. Peng, and H. Lu, "Mixed-order spectral clustering for complex networks," *Pattern Recognit.*, vol. 117, Jan. 2021, Art. no. 107964.
- [94] F. Hu, J. Liu, L. Li, and J. Liang, "Community detection in complex networks using Node2Vec with spectral clustering," *Phys. A, Stat. Mech. Appl.*, vol. 545, Jan. 2020, Art. no. 123633.
- [95] X. Qin, W. Dai, P. Jiao, W. Wang, and N. Yuan, "A multi-similarity spectral clustering method for community detection in dynamic networks," *Sci. Rep.*, vol. 6, no. 1, pp. 1–11, 2016.
- [96] A. R. Benson, D. F. Gleich, and J. Leskovec, "Tensor spectral clustering for partitioning higher-order network structures," in *Proc. SIAM Int. Conf. Data Mining*, 2015, pp. 118–126.
- [97] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
- [98] C. He, X. Fei, Q. Cheng, H. Li, Z. Hu, and Y. Tang, "A survey of community detection in complex networks using nonnegative matrix factorization," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 2, pp. 440–457, Apr. 2022.
- [99] F. Wang, T. Li, X. Wang, S. Zhu, and C. Ding, "Community discovery using nonnegative matrix factorization," *Data Mining Knowl. Discovery*, vol. 22, no. 3, pp. 493–521, 2011, doi: [10.1007/s10618-010-0181-y](https://doi.org/10.1007/s10618-010-0181-y).
- [100] D. Kuang, C. Ding, and H. Park, "Symmetric nonnegative matrix factorization for graph clustering," in *Proc. SIAM Int. Conf. Data Mining*, 2012, pp. 106–117.
- [101] Z. Gao, N. Guan, and L. Su, "Graph regularized symmetric non-negative matrix factorization for graph clustering," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Apr. 2018, pp. 379–384.
- [102] X. Shi, H. Lu, Y. He, and S. He, "Community detection in social network with pairwise constrained symmetric non-negative matrix factorization," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2015, pp. 541–546.
- [103] C. He, X. Fei, H. Li, Y. Tang, H. Liu, and S. Liu, "Improving NMF-based community discovery using distributed robust nonnegative matrix factorization with SimRank similarity measure," *J. Supercomput.*, vol. 74, no. 10, pp. 5601–5624, Oct. 2018.
- [104] X. Su, S. Xue, F. Liu, J. Wu, J. Yang, C. Zhou, W. Hu, C. Paris, S. Nepal, D. Jin, Q. Z. Sheng, and P. S. Yu, "A comprehensive survey on community detection with deep learning," 2021, *arXiv:2105.12584*.
- [105] D. Jin, Z. Yu, P. Jiao, S. Pan, D. He, J. Wu, P. S. Yu, and W. Zhang, "A survey of community detection approaches: From statistical modeling to deep learning," 2021, *arXiv:2101.01669*.
- [106] L. Yang, X. Cao, D. He, C. Wang, X. Wang, and W. Zhang, "Modularity based community detection with deep learning," in *Proc. IJCAI*, vol. 16, 2016, pp. 2252–2258.
- [107] X. Shen and F.-L. Chung, "Deep network embedding for graph representation learning in signed networks," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1556–1568, Aug. 2018.
- [108] Y. Xie, X. Wang, D. Jiang, and R. Xu, "High-performance community detection in social networks using a deep transitive autoencoder," *Inf. Sci.*, vol. 493, pp. 75–90, Jan. 2019.
- [109] J. Zhang, C. Xia, C. Zhang, L. Cui, Y. Fu, and S. Y. Philip, "BL-MNE: Emerging heterogeneous social network embedding through broad learning with aligned autoencoder," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 605–614.
- [110] Y. Zhang, Y. Xiong, Y. Ye, T. Liu, W. Wang, Y. Zhu, and P. S. Yu, "SEAL: Learning heuristics for community detection with generative adversarial networks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2020, pp. 1103–1113.
- [111] L. Yang, Y. Wang, J. Gu, C. Wang, X. Cao, and Y. Guo, "JANE: Jointly adversarial network embedding," in *Proc. IJCAI*, 2020, pp. 1381–1387.
- [112] Y. Jia, Q. Zhang, W. Zhang, and X. Wang, "CommunityGAN: Community detection with generative adversarial nets," in *Proc. World Wide Web Conf.*, 2019, pp. 784–794.
- [113] S. Mukherjee, H. Asnani, E. Lin, and S. Kannan, "ClusterGAN: Latent space clustering in generative adversarial networks," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 4610–4617.
- [114] X. Xin, C. Wang, X. Ying, and B. Wang, "Deep community detection in topologically incomplete networks," *Phys. A, Stat. Mech. Appl.*, vol. 469, pp. 342–352, Jan. 2017.
- [115] D. Jin, Z. Liu, W. Li, D. He, and W. Zhang, "Graph convolutional networks meet Markov random fields: Semi-supervised community detection in attribute networks," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 152–159.
- [116] T. Zhang, Y. Xiong, J. Zhang, Y. Zhang, Y. Jiao, and Y. Zhu, "CommDGI: Community detection oriented deep graph infomax," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 1843–1852.
- [117] O. Shchur and S. Günnemann, "Overlapping community detection with graph neural networks," 2019, *arXiv:1909.12201*.
- [118] C. Park, D. Kim, J. Han, and H. Yu, "Unsupervised attributed multiplex network embedding," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 4, pp. 5371–5378.
- [119] X. Fu, J. Zhang, Z. Meng, and I. King, "MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding," in *Proc. Web Conf.*, 2020, pp. 2331–2341.
- [120] M. P. Boobalan, D. Lopez, and X. Z. Gao, "Graph clustering using k -neighbourhood attribute structural similarity," *Appl. Soft Comput.*, vol. 47, pp. 216–223, Oct. 2016.
- [121] P. Wu and L. Pan, "Multi-objective community detection method by integrating users' behavior attributes," *Neurocomputing*, vol. 210, pp. 13–25, Oct. 2016, doi: [10.1016/j.neucom.2015.11.128](https://doi.org/10.1016/j.neucom.2015.11.128).
- [122] E. Alinezhad, B. Teimourpour, M. M. Sepehri, and M. Kargari, "Community detection in attributed networks considering both structural and attribute similarities: Two mathematical programming approaches," *Neural Comput. Appl.*, vol. 32, no. 8, pp. 3203–3220, Apr. 2020, doi: [10.1007/s00521-019-04064-5](https://doi.org/10.1007/s00521-019-04064-5).
- [123] G. P. Guedes, E. Ogasawara, E. Bezerra, and G. Xexeo, "Discovering top- k non-redundant clusterings in attributed graphs," *Neurocomputing*, vol. 210, pp. 45–54, Oct. 2016.
- [124] Z. Huang, X. Zhong, Q. Wang, M. Gong, and X. Ma, "Detecting community in attributed networks by dynamically exploring node attributes and topological structure," *Knowl.-Based Syst.*, vol. 196, May 2020, Art. no. 105760, doi: [10.1016/j.knosys.2020.105760](https://doi.org/10.1016/j.knosys.2020.105760).
- [125] K. Berahmand, M. Mohammadi, F. Saberi-Movahed, Y. Li, and Y. Xu, "Graph regularized nonnegative matrix factorization for community detection in attributed networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 1, pp. 372–385, Jan. 2023.
- [126] C. Pizzuti and A. Socievole, "Multiobjective optimization and local merge for clustering attributed graphs," *IEEE Trans. Cybern.*, vol. 50, no. 12, pp. 4997–5009, Dec. 2020.
- [127] A. Baroni, A. Conte, M. Patrignani, and S. Ruggieri, "Efficiently clustering very large attributed graphs," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Jul. 2017, pp. 369–376.
- [128] B. Perozzi, L. Akoglu, P. Iglesias Sánchez, and E. Müller, "Focused clustering and outlier detection in large attributed graphs," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2014, pp. 1346–1355.
- [129] Y. Zhou and L. Liu, "Social influence based clustering of heterogeneous information networks," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2013, pp. 338–346.
- [130] Y. Zhou, H. Cheng, and J. X. Yu, "Graph clustering based on structural/attribute similarities," *Proc. VLDB Endowment*, vol. 2, no. 1, pp. 718–729, Aug. 2009.
- [131] Y. Ruan, D. Fuhrly, and S. Parthasarathy, "Efficient community detection in large networks using content and links," in *Proc. 22nd Int. Conf. World Wide Web*, May 2013, pp. 1089–1098.
- [132] H. Tong, C. Faloutsos, and J.-Y. Pan, "Fast random walk with restart and its applications," in *Proc. 6th Int. Conf. Data Mining (ICDM)*, Dec. 2006, pp. 613–622.
- [133] J. Yang, J. McAuley, and J. Leskovec, "Community detection in networks with node attributes," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dec. 2013, pp. 1151–1156.

- [134] Z. Xu, Y. Ke, Y. Wang, H. Cheng, and J. Cheng, "A model-based approach to attributed graph clustering," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, May 2012, pp. 505–516.
- [135] T. Yang, R. Jin, Y. Chi, and S. Zhu, "Combining link and content for community detection: A discriminative approach," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jun. 2009, pp. 927–936.
- [136] M. E. J. Newman, "Spectral methods for community detection and graph partitioning," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 88, no. 4, 2013, Art. no. 042822.
- [137] W. Li, S. Jiang, and Q. Jin, "Overlap community detection using spectral algorithm based on node convergence degree," *Future Gener. Comput. Syst.*, vol. 79, no. 1, pp. 408–416, 2018.
- [138] K. Berahmand, M. Mohammadi, A. Faroughi, and R. P. Mohammadiani, "A novel method of spectral clustering in attributed networks by constructing parameter-free affinity matrix," *Cluster Comput.*, vol. 25, no. 2, pp. 869–888, Apr. 2022, doi: [10.1007/s10586-021-03430-0](https://doi.org/10.1007/s10586-021-03430-0).
- [139] K. Berahmand, E. Nasiri, R. P. Mohammadiani, and Y. Li, "Spectral clustering on protein-protein interaction networks via constructing affinity matrix using attributed graph embedding," *Comput. Biol. Med.*, vol. 138, Nov. 2021, Art. no. 104933.
- [140] F. Tang, C. Wang, J. Su, and Y. Wang, "Spectral clustering-based community detection using graph distance and node attributes," *Comput. Statist.*, vol. 35, no. 1, pp. 69–94, Mar. 2020.
- [141] Y. Pei, N. Chakraborty, and K. Sycara, "Nonnegative matrix trifactorization with graph regularization for community detection in social networks," in *Proc. 24th International Joint Conf. Artif. Intell.*, 2015, pp. 1–10.
- [142] Z. Huang, Y. Ye, X. Li, F. Liu, and H. Chen, "Joint weighted nonnegative matrix factorization for mining attributed graphs," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*, Cham, Switzerland: Springer, 2017, pp. 368–380.
- [143] X. Wang, D. Jin, X. Cao, L. Yang, and W. Zhang, "Semantic community identification in large attribute networks," in *Proc. AAAI Conf. Artif. Intell.*, 2016, vol. 30, no. 1, pp. 1–4.
- [144] M. Qin, D. Jin, K. Lei, B. Gabrys, and K. Musial-Gabrys, "Adaptive community detection incorporating topology and content in social networks?" *Knowl.-Based Syst.*, vol. 161, pp. 342–356, Jan. 2018.
- [145] C. Wang, S. Pan, C. P. Yu, R. Hu, G. Long, and C. Zhang, "Deep neighbor-aware embedding for node clustering in attributed graphs," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108230.
- [146] H. Sun, F. He, J. Huang, Y. Sun, Y. Li, C. Wang, L. He, Z. Sun, and X. Jia, "Network embedding for community detection in attributed networks," *ACM Trans. Knowl. Discovery Data*, vol. 14, no. 3, pp. 1–25, Jun. 2020.
- [147] J. Cao, D. Jin, L. Yang, and J. Dang, "Incorporating network structure with node contents for community detection on large networks using deep learning," *Neurocomputing*, vol. 297, pp. 71–81, Jul. 2018.
- [148] D. Jin, B. Li, P. Jiao, D. He, and W. Zhang, "Network-specific variational auto-encoder for embedding in attribute networks," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 2663–2669.
- [149] H. Gao, J. Pei, and H. Huang, "ProGAN: Network embedding via proximity generative adversarial network," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 1308–1316.
- [150] L. Sun, X. Liu, M. Zhao, and B. Yang, "Interpretable variational graph autoencoder with noninformative prior," *Future Internet*, vol. 13, no. 2, p. 51, Feb. 2021.
- [151] J. Ji, Y. Liang, and M. Lei, "Deep attributed graph clustering with self-separation regularization and parameter-free cluster estimation," *Neural Netw.*, vol. 142, pp. 522–533, Oct. 2021.
- [152] X. Zhang, H. Liu, Q. Li, and X.-M. Wu, "Attributed graph clustering via adaptive graph convolution," 2019, *arXiv:1906.01210*.
- [153] L. Xue, W. Yao, Y. Xia, and X. Li, "Deep attributed network embedding with community information," in *Proc. Int. Conf. Multimedia Modeling*, Cham, Switzerland: Springer, 2021, pp. 653–665.
- [154] C. Wang, S. Pan, R. Hu, G. Long, J. Jiang, and C. Zhang, "Attributed graph clustering: A deep attentional embedding approach," 2019, *arXiv:1906.06532*.
- [155] H. Fan, W. Du, A. Dahou, A. A. Ewees, D. Yousri, M. A. Elaziz, A. H. Elsheikh, L. Abugaligh, and M. A. A. Al-Qaness, "Social media toxicity classification using deep learning: Real-world application UK Brexit," *Electronics*, vol. 10, no. 11, p. 1332, Jun. 2021.
- [156] A. Namtirtha, A. Dutta, and B. Dutta, "Weighted kshell degree neighborhood: A new method for identifying the influential spreaders from a variety of complex network connectivity structures," *Expert Syst. Appl.*, vol. 139, Jan. 2020, Art. no. 112859.
- [157] F. Piccialli, F. Giampaolo, E. Prezioso, D. Camacho, and G. Acampora, "Artificial intelligence and healthcare: Forecasting of medical bookings through multi-source time-series fusion," *Inf. Fusion*, vol. 74, pp. 1–16, Oct. 2021, doi: [10.1016/j.inffus.2021.03.004](https://doi.org/10.1016/j.inffus.2021.03.004).
- [158] T. E. Tavolara, M. K. K. Niazi, A. C. Gower, M. Ginese, G. Beamer, and M. N. Gurcan, "Deep learning predicts gene expression as an intermediate data modality to identify susceptibility patterns in mycobacterium tuberculosis infected diversity outbred mice," *EBioMedicine*, vol. 67, May 2021, Art. no. 103388, doi: [10.1016/j.ebiom.2021.103388](https://doi.org/10.1016/j.ebiom.2021.103388).
- [159] J. W. Hughes, N. Yuan, B. He, J. Ouyang, J. Ebinger, P. Botting, J. Lee, J. Theurer, J. E. Tooley, K. Nieman, M. P. Lungren, D. H. Liang, I. Schnittger, J. H. Chen, E. A. Ashley, S. Cheng, D. Ouyang, and J. Y. Zou, "Deep learning evaluation of biomarkers from echocardiogram videos," *EBioMedicine*, vol. 73, Nov. 2021, Art. no. 103613, doi: [10.1016/j.ebiom.2021.103613](https://doi.org/10.1016/j.ebiom.2021.103613).
- [160] A. Brahim, R. Jennane, R. Riad, T. Janvier, L. Khedher, H. Toumi, and E. Lespessailles, "A decision support tool for early detection of knee OsteoArthritis using X-ray imaging and machine learning: Data from the OsteoArthritis initiative," *Computerized Med. Imag. Graph.*, vol. 73, pp. 11–18, Apr. 2019, doi: [10.1016/j.compmedimag.2019.01.007](https://doi.org/10.1016/j.compmedimag.2019.01.007).
- [161] S. Kumar and B. S. Panda, "Identifying influential nodes in social networks: Neighborhood coreness based voting approach," *Phys. A, Stat. Mech. Appl.*, vol. 553, Sep. 2020, Art. no. 124215, doi: [10.1016/j.physa.2020.124215](https://doi.org/10.1016/j.physa.2020.124215).
- [162] L. Li, M. Du, G. Liu, X. Hu, and G. Wu, "Extremal optimization-based semi-supervised algorithm with conflict pairwise constraints for community detection," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2014, pp. 180–187, doi: [10.1109/ASONAM.2014.6921580](https://doi.org/10.1109/ASONAM.2014.6921580).
- [163] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 23, pp. 8577–8582, Feb. 2006, doi: [10.1073/pnas.0601602103](https://doi.org/10.1073/pnas.0601602103).
- [164] L. C. Ribas, R. Riad, R. Jennane, and O. M. Bruno, "A complex network based approach for knee osteoarthritis detection: Data from the osteoarthritis initiative," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103133, doi: [10.1016/j.bspc.2021.103133](https://doi.org/10.1016/j.bspc.2021.103133).
- [165] Y. Nasser, R. Jennane, A. Chetouani, E. Lespessailles, and M. E. Hassouni, "Discriminative regularized auto-encoder for early detection of knee OsteoArthritis: Data from the osteoarthritis initiative," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2976–2984, Sep. 2020.
- [166] L. Yousfi, L. Houam, A. Boukrouche, E. Lespessailles, F. Ros, and R. Jennane, "Texture analysis and genetic algorithms for osteoporosis diagnosis," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 34, no. 05, May 2020, Art. no. 2057002.
- [167] S. M. J. Jalali, M. Karimi, A. Khosravi, and S. Nahavandi, "An efficient neuroevolution approach for heart disease detection," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 3771–3776, doi: [10.1109/SMC.2019.8913997](https://doi.org/10.1109/SMC.2019.8913997).
- [168] S. Ahmadian, S. M. J. Jalali, S. M. S. Islam, A. Khosravi, E. Fazli, and S. Nahavandi, "A novel deep neuroevolution-based image classification method to diagnose coronavirus disease (COVID-19)," *Comput. Biol. Med.*, vol. 139, Dec. 2021, Art. no. 104994, doi: [10.1016/j.combiomed.2021.104994](https://doi.org/10.1016/j.combiomed.2021.104994).
- [169] S. M. J. Jalali, M. Ahmadian, S. Ahmadian, A. Khosravi, M. Alazab, and S. Nahavandi, "An oppositional-cauchy based GSK evolutionary algorithm with a novel deep ensemble reinforcement learning strategy for COVID-19 diagnosis," *Appl. Soft Comput.*, vol. 111, Nov. 2021, Art. no. 107675, doi: [10.1016/j.asoc.2021.107675](https://doi.org/10.1016/j.asoc.2021.107675).
- [170] M. Rostami, S. Forouzandeh, K. Berahmand, M. Soltani, M. Shahsavari, and M. Oussalah, "Gene selection for microarray data classification via multi-objective graph theoretic-based method," *Artif. Intell. Med.*, vol. 123, Jan. 2022, Art. no. 102228, doi: [10.1016/j.artmed.2021.102228](https://doi.org/10.1016/j.artmed.2021.102228).
- [171] F. Ali, S. El-Sappagh, S. M. R. Islam, A. Ali, M. Attique, M. Imran, and K.-S. Kwak, "An intelligent healthcare monitoring framework using wearable sensors and social networking data," *Future Gener. Comput. Syst.*, vol. 114, pp. 23–43, Jan. 2021, doi: [10.1016/j.future.2020.07.047](https://doi.org/10.1016/j.future.2020.07.047).

- [172] J. J. Bechtel, W. A. Kelley, T. A. Coons, M. G. Klein, D. D. Slagel, and T. L. Petty, "Lung cancer detection in patients with airflow obstruction identified in a primary care outpatient practice," *Chest*, vol. 127, no. 4, pp. 1140–1145, Apr. 2005, doi: [10.1016/S0012-3692\(15\)34459-7](https://doi.org/10.1016/S0012-3692(15)34459-7).
- [173] X. Wang and J. M. Garibaldi, "A comparison of fuzzy and non-fuzzy clustering techniques in cancer diagnosis," in *Proc. 2nd Int. Conf. Comput. Intell. Med. Healthcare*, vol. 28, 2005, pp. 1–25.
- [174] M. Rostami, K. Berahmand, and S. Forouzandeh, "A novel community detection based genetic algorithm for feature selection," *J. Big Data*, vol. 8, no. 1, pp. 1–10, Dec. 2021, doi: [10.1186/s40537-020-00398-3](https://doi.org/10.1186/s40537-020-00398-3).
- [175] Z. Wang, M. A. Andrews, Z.-X. Wu, L. Wang, and C. T. Bauch, "Coupled disease-behavior dynamics on complex networks: A review," *Phys. Life Rev.*, vol. 15, pp. 1–29, Dec. 2015, doi: [10.1016/j.plrev.2015.07.006](https://doi.org/10.1016/j.plrev.2015.07.006).
- [176] T. Choudhury, R. Arunachalam, A. Khanna, E. Jasinska, V. Bolshev, V. Panchenko, and Z. Leonowicz, "A social network analysis approach to COVID-19 community detection techniques," *Int. J. Environ. Res. Public Health*, vol. 19, no. 7, p. 3791, Mar. 2022.
- [177] G. Chowell, L. Sattenspiel, S. Bansal, and C. Viboud, "Mathematical models to characterize early epidemic growth: A review," *Phys. Life Rev.*, vol. 18, pp. 66–97, Sep. 2016, doi: [10.1016/j.plrev.2016.07.005](https://doi.org/10.1016/j.plrev.2016.07.005).
- [178] S. Wang, M. Gong, W. Liu, and Y. Wu, "Preventing epidemic spreading in networks by community detection and memetic algorithm," *Appl. Soft Comput.*, vol. 89, Apr. 2020, Art. no. 106118, doi: [10.1016/j.asoc.2020.106118](https://doi.org/10.1016/j.asoc.2020.106118).
- [179] A. Cori, P. Nouvellet, T. Garske, H. Bourhy, E. Nakoué, and T. Jombart, "A graph-based evidence synthesis approach to detecting outbreak clusters: An application to dog rabies," *PLOS Comput. Biol.*, vol. 14, no. 12, Dec. 2018, Art. no. e1006554.
- [180] J. Chen, K. Li, K. Bilal, A. A. Metwally, K. Li, and P. Yu, "Parallel protein community detection in large-scale PPI networks based on multi-source learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Aug. 31, 2020, doi: [10.1109/TCBB.2018.2868088](https://doi.org/10.1109/TCBB.2018.2868088).
- [181] A. I. Amanatidou and G. V. Dedoussis, "Construction and analysis of protein-protein interaction network of non-alcoholic fatty liver disease," *Comput. Biol. Med.*, vol. 131, Apr. 2021, Art. no. 104243, doi: [10.1016/j.cmpbiomed.2021.104243](https://doi.org/10.1016/j.cmpbiomed.2021.104243).
- [182] W. Liu and L. Chen, "Community detection in disease-gene network based on principal component analysis," *Tsinghua Sci. Technol.*, vol. 18, no. 5, pp. 454–461, Oct. 2013, doi: [10.1109/TST.2013.6616519](https://doi.org/10.1109/TST.2013.6616519).
- [183] R. Dunn, F. Dudbridge, and C. M. Sanderson, "The use of edge-betweenness clustering to investigate biological function in protein interaction networks," *BMC Bioinf.*, vol. 6, no. 1, p. 39, 2005, doi: [10.1186/1471-2105-6-39](https://doi.org/10.1186/1471-2105-6-39).
- [184] R. Flemming, W. Schüttig, F. Ng, V. Leve, and L. Sundmacher, "Using social network analysis methods to identify networks of physicians responsible for the care of specific patient populations," *BMC Health Services Res.*, vol. 22, no. 1, pp. 1–17, Dec. 2022.
- [185] M. Ostovari and D. Yu, "Impact of care provider network characteristics on patient outcomes: Usage of social network analysis and a multi-scale community detection," *PLoS ONE*, vol. 14, no. 9, Sep. 2019, Art. no. e0222016.
- [186] T. McCurdie, P. Sanderson, and L. M. Aitken, "Applying social network analysis to the examination of interruptions in healthcare," *Appl. Ergonom.*, vol. 67, pp. 50–60, Feb. 2018, doi: [10.1016/j.apergo.2017.08.014](https://doi.org/10.1016/j.apergo.2017.08.014).
- [187] S. Shirazi, A. Albadvi, E. Akhondzadeh, F. Farzadfar, and B. Teimourpour, "A new application of community detection for identifying the real specialty of physicians," *Int. J. Med. Informat.*, vol. 140, Aug. 2020, Art. no. 104161.
- [188] J. Niyirora and O. Aragones, "Network analysis of medical care services," *Health Informat. J.*, vol. 26, no. 3, pp. 1631–1658, Sep. 2020.
- [189] K. Sabot, K. Blanchet, D. Berhanu, N. Spicer, and J. Schellenberg, "Professional advice for primary healthcare workers in Ethiopia: A social network analysis," *BMC Health Services Res.*, vol. 20, no. 1, p. 551, Dec. 2020, doi: [10.1186/s12913-020-05367-3](https://doi.org/10.1186/s12913-020-05367-3).
- [190] A. De Brún and E. McAuliffe, "Social network analysis as a methodological approach to explore health systems: A case study exploring support among senior managers/executives in a hospital network," *Int. J. Environ. Res. Public Health*, vol. 15, no. 3, p. 511, Mar. 2018, doi: [10.3390/ijerph15030511](https://doi.org/10.3390/ijerph15030511).
- [191] S. Shirazi, H. Baziyad, N. Ahmadi, and A. Albadvi, "A new application of Louvain algorithm for identifying disease fields using big data techniques," *J. Biostatist. Epidemiol.*, vol. 5, no. 3, pp. 183–193, Jul. 2020.
- [192] H. Elghazel, V. Deslandres, M.-S. Hacid, A. Dussauchoy, and H. Kheddouci, "A new clustering approach for symbolic data and its validation: Application to the healthcare data," in *Proc. Int. Symp. Methodologies Intell. Syst.* Cham, Switzerland: Springer, 2006, pp. 473–482.
- [193] K. Alkhatib, M. Almahmood, O. Elayan, and L. Abualigah, "Regional analytics and forecasting for most affected stock markets: The case of GCC stock markets during COVID-19 pandemic," *Int. J. Syst. Assurance Eng. Manag.*, vol. 13, pp. 1–11, Oct. 2021.
- [194] P. Sumari, S. J. Syed, and L. Abualigah, "A novel deep learning pipeline architecture based on CNN to detect COVID-19 in chest X-ray images," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 2001–2011, Apr. 2021.
- [195] K. A. Qureshi, R. A. S. Malick, M. Sabih, and H. Cherifi, "Complex network and source inspired COVID-19 fake news classification on Twitter," *IEEE Access*, vol. 9, pp. 139636–139656, 2021, doi: [10.1109/ACCESS.2021.3119404](https://doi.org/10.1109/ACCESS.2021.3119404).
- [196] K. Baktha, M. Dev, H. Gupta, A. Agarwal, and B. Balamurugan, "Social network analysis in healthcare," in *Internet of Things and Big Data Technologies for Next Generation Healthcare*. Cham, Switzerland: Springer, 2017, pp. 309–334.
- [197] H. A. Schwartz, J. Eichstaedt, M. Kern, L. Dziurzynski, R. Lucas, M. Agrawal, and G. Park, "Characterizing geographic variation in well-being using tweets," in *Proc. 7th Int. AAAI Conf. Weblogs Social Media*, 2013, pp. 583–591.
- [198] X. Zong, X. Ding, and Z. Zhou, "Health level classification by fusing medical evaluation from multiple social networks," *Future Gener. Comput. Syst.*, vol. 114, pp. 574–580, Jan. 2021, doi: [10.1016/j.future.2020.08.008](https://doi.org/10.1016/j.future.2020.08.008).
- [199] K. Abbas, A. Abbasi, S. Dong, L. Niu, L. Yu, B. Chen, S.-M. Cai, and Q. Hasan, "Application of network link prediction in drug discovery," *BMC Bioinf.*, vol. 22, no. 1, pp. 1–21, Apr. 2021.
- [200] E. Galan-Vasquez and E. Perez-Rueda, "A landscape for drug-target interactions based on network analysis," *PLoS ONE*, vol. 16, no. 3, Mar. 2021, Art. no. e0247018.
- [201] Y. Zhang, Y. Liu, Q. Li, R. Jin, and C. Wen, "LILPA: A label importance based label propagation algorithm for community detection with application to core drug discovery," *Neurocomputing*, vol. 413, pp. 107–133, Nov. 2020, doi: [10.1016/j.neucom.2020.06.088](https://doi.org/10.1016/j.neucom.2020.06.088).
- [202] D. Yanrui, Z. Zhen, W. Wenchao, and C. Yujie, "Identifying the communities in the metabolic network using 'component' definition and Girvan-Newman algorithm," in *Proc. 14th Int. Symp. Distrib. Comput. Appl. Bus. Eng. Sci. (DCABES)*, Aug. 2015, pp. 42–45.
- [203] A. Gangopadhyay and S. Chen, "Health care fraud detection with community detection algorithms," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, Apr. 2016, pp. 1–5.
- [204] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX," in *Proc. 7th Python Sci. Conf.*, Pasadena, CA, USA, 2008, pp. 11–15.
- [205] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," in *Proc. Int. AAAI Conf. Weblogs Social Media*, 2009, pp. 1–15.
- [206] 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](https://doi.org/10.1038/s41598-019-41695-z).
- [207] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. Cham, Switzerland: Springer, 2013.
- [208] A. Pothén, "Graph partitioning algorithms with applications to scientific computing," in *Parallel Numerical Algorithms*, D. E. Keyes, A. Sameh, and V. Venkatakrishnan, Eds. Dordrecht, The Netherlands: Springer, 1997, pp. 323–368.
- [209] R. A. Haraty, M. Dimishkieh, and M. Masud, "An enhanced k-means clustering algorithm for pattern discovery in healthcare data," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 6, Jun. 2015, Art. no. 615740.
- [210] S. Rallapalli, R. Gondkar, and G. V. M. Rao, "Cloud based k-means clustering running as a Mapreduce job for big data healthcare analytics using Apache Mahout," in *Information Systems Design and Intelligent Applications*. Cham, Switzerland: Springer, 2016, pp. 127–135.
- [211] S. Gregory, "Finding overlapping communities in networks by label propagation," *New J. Phys.*, vol. 12, no. 10, Oct. 2010, Art. no. 103018.
- [212] S. Boettcher and A. G. Percus, "Extremal optimization for graph partitioning," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 2, Jul. 2001, Art. no. 026114, doi: [10.1103/PhysRevE.64.026114](https://doi.org/10.1103/PhysRevE.64.026114).

- [213] 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](https://doi.org/10.1103/PhysRevLett.86.5211).
- [214] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 70, no. 6, 2004, Art. no. 066111, doi: [10.1103/PhysRevE.70.066111](https://doi.org/10.1103/PhysRevE.70.066111).
- [215] J. Liu and T. Liu, "Detecting community structure in complex networks using simulated annealing with k -means algorithms," *Phys. A, Stat. Mech. Appl.*, vol. 389, no. 11, pp. 2300–2309, Jun. 2010, doi: [10.1016/j.physa.2010.01.042](https://doi.org/10.1016/j.physa.2010.01.042).
- [216] M. Gong, B. Fu, L. Jiao, and H. Du, "Memetic algorithm for community detection in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 84, no. 5, Nov. 2011, Art. no. 056101, doi: [10.1103/PhysRevE.84.056101](https://doi.org/10.1103/PhysRevE.84.056101).
- [217] M. Gong, L. Ma, Q. Zhang, and L. Jiao, "Community detection in networks by using multiobjective evolutionary algorithm with decomposition," *Phys. A, Stat. Mech. Appl.*, vol. 391, no. 15, pp. 4050–4060, 2012, doi: [10.1016/j.physa.2012.03.021](https://doi.org/10.1016/j.physa.2012.03.021).
- [218] C. Liu, J. Liu, and Z. Jiang, "A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2274–2287, Dec. 2014, doi: [10.1109/TCYB.2014.2305974](https://doi.org/10.1109/TCYB.2014.2305974).
- [219] P. Yuan, W. Wang, and M. Song, "Detecting overlapping community structures with PCA technology and member index," in *Proc. 9th EAI Int. Conf. Mobile Multimedia Commun.*, 2016, pp. 121–125.
- [220] L. Lin, X. Zheng-Min, S.-H. Li, and S.-N. L. Zhi-Huang, "Detecting overlapping community structure via an improved spread algorithm based on PCA," in *Proc. Int. Conf. Comput. Sci. Softw. Eng. (CSSE)*, 2014, pp. 115–121.
- [221] K. Shimizu and R. Mori, "Exponential-time quantum algorithms for graph coloring problems," in *Proc. Latin Amer. Symp. Theor. Informat.* Cham, Switzerland: Springer, 2021, pp. 387–398.
- [222] O. Goudet, B. Duval, and J.-K. Hao, "Population-based gradient descent weight learning for graph coloring problems," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106581.
- [223] T. Dokeroglu and E. Sevinc, "Memetic teaching-learning-based optimization algorithms for large graph coloring problems," *Eng. Appl. Artif. Intell.*, vol. 102, Jun. 2021, Art. no. 104282.
- [224] O. Şeker, T. Ekim, and Z. C. Taşkın, "An exact cutting plane algorithm to solve the selective graph coloring problem in perfect graphs," *Eur. J. Oper. Res.*, vol. 291, no. 1, pp. 67–83, May 2021.
- [225] DATASUS. Accessed: Oct. 1, 2022. [Online]. Available: <http://cnes.datasus.gov.br/>
- [226] F. Gräßer, S. Kallumadi, H. Malberg, and S. Zaunseder, "Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning," in *Proc. Int. Conf. Digit. Health*, Apr. 2018, pp. 121–125.
- [227] *Healthcare Organization Services: Professional and Organization Affiliations Maintenance Process*, IM S Health, Bedford, U.K., 2012.
- [228] Q. C. Nguyen, D. Li, H.-W. Meng, S. Kath, E. Nsoesie, F. Li, and M. Wen, "Building a national neighborhood dataset from geotagged Twitter data for indicators of happiness, diet, and physical activity," *JMIR Public Health Surveill.*, vol. 2, no. 2, p. e158, Oct. 2016.
- [229] J. Fernandes, B. Bregman, P. Combemale, C. Amaz, L. de Léotoing, A. Vainchtock, and A.-F. Gaudin, "Hospitalisation costs of metastatic melanoma in France; the MELISSA study (melanoma in hospital costs assessment)," *BMC Health Services Res.*, vol. 17, no. 1, pp. 1–13, Dec. 2017.
- [230] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-Wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 1, pp. 1–9, 2016.



MEHRDAD ROSTAMI received the B.Sc. degree (Hons.) in computer engineering from Razi University, Kermanshah, Iran, in 2012, and the M.Sc. degree (Hons.) in artificial intelligence from the University of Kurdistan, Sanandaj, Iran, in 2014. He is currently pursuing the Ph.D. degree with the University of Oulu, Oulu, Finland. He is also a Researcher with the University of Oulu. His research interests include pattern recognition, medical diagnosis, machine learning, recommender systems, big data mining, and bioinformatics.



MOURAD OUSSALAH (Senior Member, IEEE) received the Ph.D. degree in robotics and artificial intelligence from the University of Paris XII, in 1998. He held research and academic positions with KU Leuven, the City, University of London, and Birmingham University, U.K. He is currently a Research Professor with the Faculty of Information Technology and Electrical Engineering, University of Oulu, Finland. His research interests include information processing, data mining, and text mining. He has published more than 200 technical articles and led several projects in his research areas. He is a fellow of the Royal Statistical Society.



KAMAL BERAHMAND received the M.S. degree in information technology engineering from Azarbaijan Shahid Madani University, Tabriz, Iran, in 2017. He is currently pursuing the Ph.D. degree with the Faculty of Computer Engineering and Information Technology, Queensland University of Technology, Brisbane, Australia. His current research interests include machine learning, network science, graph neural networks, and knowledge graph.



VAHID FARRAHI received the Ph.D. degree in medical physics and technology from the Faculty of Medicine, University of Oulu, Finland, in 2021. He received funding from the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie Grant (713645) for his doctoral research. He is currently a Postdoctoral Fellow with the University of Oulu, where he conducts multidisciplinary problem-oriented research with a focus on machine-learning and data-driven solutions. His Ph.D. thesis received multiple recognition and was selected as the Best Doctoral Thesis from the Faculty of Medicine, University of Oulu, in 2021.

...