Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-021-03508-y

ORIGINAL RESEARCH

An improved Girvan–Newman community detection algorithm using trust-based centrality

Mahsa Zahiri[1] · Javad Mohammadzadeh[1] · Sasan Harifi[1]

Received: 28 August 2020 / Accepted: 13 September 2021

Abstract
Accumulative structure or cluster-like shape is one of the important features of social networks. These structures and clusters are communities in a complex network and are fully detectable. Common group behaviors of different communities can be categorized using community detection methods. Categorize behavior allows the study of each part of the network to be done centrally. This paper uses trust-based centrality to detect the communities that make up the network. Centrality determines the relative importance of a node in the graph of social networks. Redefining the trust-based centrality makes it possible to change the position in the analysis of centrality and separates the local central nodes and global central nodes. Then, a trust-based algorithm is proposed to express the strength of trust penetration conceptually between nodes to extract communities in networks. This method has led to the achievement of a flexible and effective community detection method. The proposed algorithm is applied to four benchmark networks. The experiments consist of two independent parts. The first part is to use the proposed algorithm to detect clusters and communities. After that, the algorithm is compared with a Girvan–Newman inspired method. The second part is the implementation of the proposed algorithm with a large number of iterations with the aim of modularity maximization and comparing it with other community detection algorithms. Although, the modularity criterion has been used to validate and compare the solution quality in both independent parts of the experiments. The results show about 1.4–5.2% improvement in community detection.

Keywords Community detection algorithm · Trust-based centrality · Social network · Complex network · Clustering · Girvan–Newman

1 Introduction

Nowadays, community detection in large networks has become a very important issue and extensive research has been done on this issue. Along with the growth of social networks, there is a need to manage and categorize these communities (Gilbert et al. 2011). On the other hand, with the computer revolution, a platform has been provided for researchers to provide a huge amount of data and computational resources for processing and data analysis.

✉ Javad Mohammadzadeh
j.mohammadzadeh@kiau.ac.ir

Mahsa Zahiri
maahsa.zaahiri@gmail.com

Sasan Harifi
s.harifi@kiau.ac.ir

1 Department of Computer Engineering, Karaj Branch, Islamic Azad University, Karaj, Iran

An important feature of networks representing real systems is their social structure or cluster structure. This means that the vertices are placed in special groups so that the accumulation of edges within these groups is high and among the groups is low. Identifying these groups, called clusters or communities, is of particular value in social science, biology, computer science, and other disciplines that study complex systems. As mentioned, one of the most important features of complex networks is the presence of communities in them. With the help of community extraction methods, common group behaviors can be categorized and each part of the network can be studied. There are various methods for community detection (Kırer and Çırpıcı 2016). Traditional methods use clustering. In recent years, many solutions have been considered and various algorithms have been proposed.

Many algorithms have been introduced by researchers. However, these algorithms have been evaluated on a limited number of networks with a small number of nodes. These limitations on evaluation indicate that these methods are

Published online: 25 September 2021

not comprehensive and cannot be used in large or real networks. Therefore, there is no best algorithm for identifying a community appropriate to the type of network (Xiao et al. 2020). The main challenge is that the number of communities on social networks is usually unknown. Most community detection algorithms work well on small networks, but the performance of these algorithms on real-world networks with millions of nodes is severely reduced (Alghamdi and Greene 2019). In many cases, they are not even able to identify communities. High computations in determining paths for one class of algorithms and clustering constraints for another class of these algorithms affect the quality of the output results. Some methods are also not scalable for large networks (Yang et al. 2016). As mentioned, extracting community structure is an important topic in social network analysis, and identifying hidden communities in social networks helps to better understand the structural features of networks in the real world. Also, community detection is useful for monitoring public opinion, identifying leading opinions, and implementing suggestions to individuals (Nerurkar et al. 2019).

Compared to simple user relationships or content, considering the trust characteristics of multiple users provides a more conceptual description of the link between users. Therefore, the aim is to provide a non-overlapping community detection algorithm based on the trust mechanism to identify the structure of society (Nikolaev et al. 2015). First, the definition of trust between users should be examined to express the amount of trust, which includes direct and indirect trust, and then the method of calculating trust should be presented to quantify the amount of trust (Grabner-Kräuter and Bitter 2015). Centrality is an important element in social network analysis. In social networks, each member can communicate with any other member of the network. This has a significant effect on centrality in that it reduces the throughput of connections in the social graph because network members do not necessarily have to follow connections (Muller and Peres 2019). In this paper, a new criterion for centrality is presented, which is based on trust. In other words, in this study, the trust centrality is used to identify the communities that make up the network. The trust-based centrality allows for a change in the analysis of centrality and separates the local and global central nodes. Then, starting from the trust relationship, it combines edge compatibility with the detection of a non-overlapping community and proposes a trust-based algorithm to express the penetration power of trust conceptually between nodes to extract communities in networks. This approach achieves a flexible and effective circulation community detection method. Hence, the goal of this study is to achieve a flexible and efficient method of detecting communities. Therefore, the modularity criterion has been used to validate and compare the solution quality. Although the trust concept was used in graph theories, the main contribution of this paper is defining a community and creating a graph based on trust centrality, and calculating trust for the node instead of using iteratively removing high-betweenness edges in a hierarchical clustering procedure in the Girvan–Newman algorithm.

The rest of the paper is structured as follows: Sect. 2 introduces the literature review. Section 3 represents the trust-based algorithm. Section 4 includes experimental results. Section 5 provides conclusions.

2 Literature review

In the literature, social networks have been considered since the nineteenth century. Research in this area has accelerated since the 1940s with the definition of tools such as the social graph. In social networks, community detection methods can be used to categorize people and their desires (Freeman 2011). The purpose of community detection is to sort the samples into clusters that have a relatively strong degree of relationship between the members of the cluster and a relatively weak degree between the members of the different clusters. Communities provide valuable information about the type of communication between users, how information is transmitted between them and how users are distributed on social networks, and in fact, is considered as a key component of these networks (Barabási 2013).

The detection of communities can be generally divided into four categories. The first is the detection of communities as vertices, in which the vertices of a group are included in the calculations, and communities are determined based on certain criteria (Rosvall et al. 2019). Second is group-based community detection, in which case the similarity of groups at the specific level of a network is examined (Kumar and Carley 2018). The third is the detection of communities as network-based, in which case all the links in the network are examined and similarities are searched extensively at the network level (Mahajan and Raipurkar 2018). Finally, the fourth is the identification of communities in a hierarchical manner, which in this method, based on the network topology, is hierarchically determined communities (Zhang et al. 2014). Figure 1 shows the classification of community detection methods.

A community is a subgraph of a graph whose number of connections between the members of that subgraph is greater than the number of connections that connect it to the rest of the graph. The goal of most community detection algorithms is to divide the graph into several connected subgraphs, each node of which must belong to one community and not to other communities. In the simplest case, dividing the nodes into communities is obtained by dividing the graph so that the vertices between the two groups are minimal. In other words, the nodes of each part have the most connection with

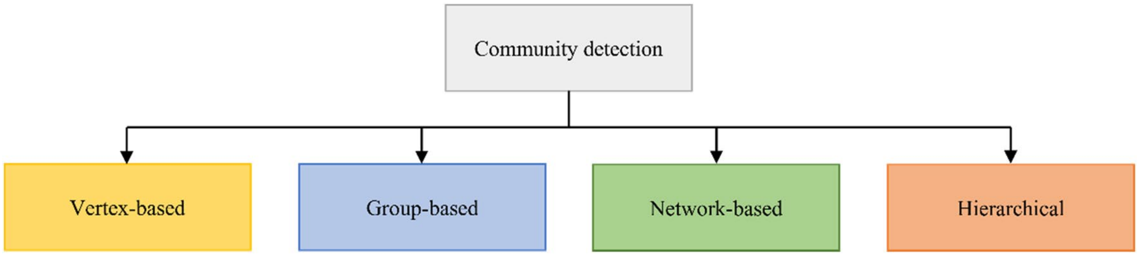

Fig. 1 Classification of community detection methods

each other but have the least connection with the nodes of the other part. One of the existing algorithms for dividing a graph is Kernighan-Lin algorithm (Kernighan and Lin 1970). In this algorithm, if the number and size of communities are not known, the concept of dividing a graph change to a new concept called community detection. Other algorithm can be mentioned is Girvan–Newman algorithm. The Girvan–Newman algorithm, introduced by Girvan and Newman (2002) and Du et al. (2008), is one of the clustering methods used to detection of communities.

Lancichinetti et al. (2008) introduced a new measure called Lancichinetti–Fortunato–Radicchi (LFR), which somewhat improved some of the problems of Girvan–Newman. In Girvan–Newman, all nodes have comparable degrees and all communities are the same size, but the degree of distribution in real networks is different, so the Girvan–Newman algorithms may not work in real networks. The LFR method considers both the degree of the node and the size of the community. Zhao and Zhang (2011) proposed a new clustering method to the structure detection of society in the network, which was suitable for use in the analysis of some social networks. In their method, individuals and their relationships were represented by a weighted graph and were based on the density analysis of weighted subgraphs. A method called Infomap was presented by Miyauchi and Kawase (2016). This method uses data compression to the detection of the community. It does this by optimizing a quality function for community detection in directional and weighted networks. Danowski (2012) stated the main methods of detection communities and examined the limitations and shortcomings of each method.

In other studies, Moosavi et al. (2017), proposed a method to explore the community, which in addition to communication information between nodes, used content information to improve the quality of community detection. Their method was a new approach based on an iterative pattern and based on users' operations on the network, and in particular, it was implemented on internet social networks where users choose their favorite operations. Jin et al. (2011) and Wang et al. (2013) used the betweenness degree measurement to find the communities of each group. A member with a high betweenness degree means a strong connection in the group. Another type of modularity was proposed by Nicosia et al. (2009) to evaluate the quality of the community structure. This modularity was proposed based on the clustering coefficient below the maximum graph considering that the network is weightless and directionless. Ahn et al. (2010) presented an algorithm that performs clustering the relationship between nodes instead of clustering nodes. In this algorithm, the similarity of a pair of links was determined by the neighbors of the nodes connected by them.

In more recent studies, Zhang et al. (2017) applied node importance and label influence concepts for using them in a Label Propagation Algorithm (LPA). The quality of community detection through their method was much improved. Their method also shortened the iteration period and had appropriate accuracy and stability in large-scale networks. Deng et al. (2019) applied fuzzy C-means membership vectors for labels of vertexes in each community. In their method, the use of fuzzy C-means stabilized the status of the communities. Carnivali et al. (2020) utilized a coarse-grained vertex clustering to deal with the costly Louvain method. In this method, they used the original graph preprocessing to forward a graph of reduced size. Jiang et al. (2020) used central node-based link prediction for finding missing information. Their method was able to deal with an ambiguous community structure. Hu et al. (2020) considered a framework for learning continuous feature representations in networks. Within the framework provided by them, their algorithm could learn a mapping of nodes to low-dimensional space of features. Guo et al. (2020) utilized an algorithm based on the internal force between nodes. Ji et al. (2020) viewed community detection as a multi-objective problem. They used Ant Colony Optimization (ACO) algorithm for their method. In their method, they simultaneously considered ratio cut and negative ratio association as two objective functions. Harifi et al. (2021) used metaheuristic methods to maximize the modularity value. They used a hybrid algorithm based on nature inspired Emperor Penguins Colony (EPC) algorithm. The advantage of using metaheuristic methods is the appropriate speed and the ability to solve problems with high dimensions. Therefore, it is very suitable for networks with many nodes. Table 1 lists the studies or methods along with their descriptions and strengths.

3 The proposed trust-based algorithm

3.1 Trust definitions

Trust is one of the main characteristics of social network users and is usually considered in two direct and indirect ways, which is expressed in the following mathematical relations (Chen et al. 2018).

Direct trust: in the field of sociology and psychology, trust is an individual and relative characteristic. That is, trust has different definitions according to different cultures. In general, user actions such as active collaboration, communication, and concern in the field of social network to some extent indicate the development of their trust. Direct trust can be based on the tie strength of the connection or based on the node similarity.

According to the definition by Chen et al. (2018) for a pair of nodes together, the direct trust obtained from their tie strength is obtained from the following equation,

$$d\_trust\_r(u,v) = \frac{w(u,v)}{w(u)} \tag{1}$$

where $d\_trust\_r(u,v)$ is the degree of direct trust between u and v. $d\_trust\_r(u,v)$ is a member of (0.1]. $w(u,v)$ is the strength between u and v. $w(u)$ is the total strength of the tie between u and the nodes in its neighborhood.

For two adjacent users, direct trust based on node similarity is obtained from the following equation,

$$d\_trust\_s(u,v) = \sum_{t \in N(u) \cap N(v)} \frac{1}{I(t)} \tag{2}$$

where $d\_trust\_s(u,v)$ is the degree of direct trust between u and v. $N(u)$ is neighboring sets of u and $N(v)$ is neighboring sets of v. $I(t)$ is penetration degree of t.

Therefore, according to the above relationships, direct trust is obtained from the sum of two relationships, direct trust based on tie strength and, direct trust based on node similarity (Chen et al. 2018). So we have,

$$d\_trust(u,v) = d\_trust\_r(u,v) + d\_trust\_s(u,v) \tag{3}$$

Indirect trust: in the social network, adjacent nodes make indirect connections through intermediate nodes, and the

Table 1 Different studies in the field of community detection

Study	Method	Descriptions and strengths
Lancichinetti et al. (2008)	LFR	Based on the Girvan–Newman method, considering the same size of communities, considering degree and size of society
Jin et al. (2011)	Betweenness degree measurement	Detect strong connection in the group based on checking high betweenness degree
Zhao and Zhang (2011)	Weighted graph analysis	Create much smaller hierarchical trees that clearly show meaningful clusters
Wang et al. (2013)	Betweenness degree measurement	Detect strong connection in the group based on checking high betweenness degree
Nikolaev et al. (2015)	Entropy-based centrality	Using entropy-based centrality and embedding it in the structure of the Girvan–Newman community detection algorithm
Zhang and Wang (2015)	Infomap	Data compression and use the quality function
Miyauchi and Kawase (2016)	Infomap	Data compression and use the quality function
Moosavi et al. (2017)	Content information	Using iterative pattern and based on users' operations
Zhang et al. (2017)	LPA-NI	Using node importance and label influence concepts for modifying label propagation
Deng et al. (2019)	LPA and fuzzy C-means	Using fuzzy C-means membership vectors for modify label propagation
Carnivali et al. (2020)	CoVeC	Suitable for sparse graphs and reduce the cost of the first iterations of the Louvain method
Hu et al. (2020)	Spectral clustering	An algorithmic framework for learning continuous feature representations for nodes in networks
Ji et al. (2020)	Decomposition-based ant colony	Considering community detection as a multi-objective problem
Jiang et al. (2020)	Central node based link prediction	Finding missing information and deal with the networks with an ambiguous community structure
Harifi et al. (2021)	Hybrid-EPC metaheuristic	Using a hybrid metaheuristic algorithm maximization of modularity with high speed in reaching the final community
The current work	Trust-based centrality	Using trust-based centrality and embedding it in the structure of the Girvan–Newman community detection algorithm

indirect trust of adjacent nodes is obtained based on the direct trust of these adjacent nodes. Based on the different paths between the source node and the destination node, the calculation of indirect trust is done in two methods, single-path and multi-path (Chen et al. 2018).

When the source node is not adjacent to the destination node and there is only one transmission path between them, the trust between the two nodes is a single-path indirect trust. The relationship of this type of trust is as follows,

$$i\_trust\_s(u \cdot v) = \begin{cases} mt \times \frac{d_{max} - d_{u,v} + 1}{d_{max}}, & \text{if} d_{u,v} \le d_{max} \\ 0, & \text{if} d_{u,v} > d_{max} \end{cases} \quad (4)$$

where mt is minimum of $d\_trust(u, u_1), d\_trust(u, u_2), \dots, d\_trust(u_n, v)$. In fact, mt is the route length between u and v. d_{max} is the maximum distance of trust transfer. In the trust transfer process, trust decreases as the distance between users increases.

If the source node is not adjacent to the destination node and there are at least two transmission paths between them, the trust between the two nodes is a multi-path indirect trust (Chen et al. 2018). The relationship of this type of trust is calculated as follows,

$$i\_trust\_m(u.v) = \max_{paths(u,v)} \{i\_trust\_s(u, v)\} \quad (5)$$

where $paths(u, v)$ is the path set of the node u and v.

Therefore, considering the direct and indirect trust relationships for the two nodes u and v, we have,

$$trust(u \cdot v) = \begin{cases} d\_trust(u, v), & \text{if} u \text{and} v \text{are adjacent} \\ i\_trust\_m(u, v), & \text{else} \end{cases} \quad (6)$$

3.2 The algorithm

The proposed algorithm for community detection is inspired by the algorithm proposed by Girvan and Newman (2002), which iteratively removes high-betweenness edges in a hierarchical clustering procedure. By defining a community and creating a graph based on trust and calculating trust for the node, the algorithm presented in this paper can be used to community detection in the network. Algorithm 1 shows the pseudo-code of the algorithm presented in this paper.

Algorithm 1: Pseudo-code of the trust-based algorithm.

```
STEP 1:
  Input number of nodes in network as (N);
  Input weighted matrix of trust as (T);
  Input iteration as (K);
  STEP 2: for k=1 to K do
    STEP 3: for i=1 to N and j=1 to N, i≠j do
      remove the link between nodes i and j, if it exists;
      revise the structure to calculate trust from node i to j;
      recalculate the weighted matrix of trust for nodes;
      remember/update the link for which the trust decrease is maximum;
    END STEP 3
    remove the link for which the trust decrease is maximum;
  END STEP 2
  sort to identify the obtained clusters;
END STEP 1
```

To implement, first, the adjacency matrix equivalent to the data graph is extracted. This matrix is square for each data and its nodes are individuals in the community and the edges show the relationship between individuals in the community. According to the definition of direct trust, which expresses the direct relationship between individuals, this type of trust is characterized by the separation of the edges of the data matrix. Therefore, the direct trust of each node is the sum of the direct connection of the edges of this node with the adjacent nodes, if available. The indirect trust of each node with the adjacent node according to the relevant equation, is the shortest path between this node and the adjacent node. Hence, the indirect trust of each node, if any, is equal to the sum of the shortest paths with the neighboring nodes. Total trust is equal to the sum of direct and indirect trust for each node.

The proposed community detection algorithm, based on the structure defined for each data, operates in such a way that in the first step, the algorithm starts from a selected node in the initial state, and the direct and indirect trust values are calculated for all nodes of the community. In the next step, again by selecting an initial node, the edge related to the path of this node with the neighboring nodes is removed and the trust values for the nodes are recalculated with an iterative loop. Finally, the edge is selected to be removed, which, if removed, reduces the amount of trust with a lower

Fig. 2 The flowchart of proposed trust-based algorithm

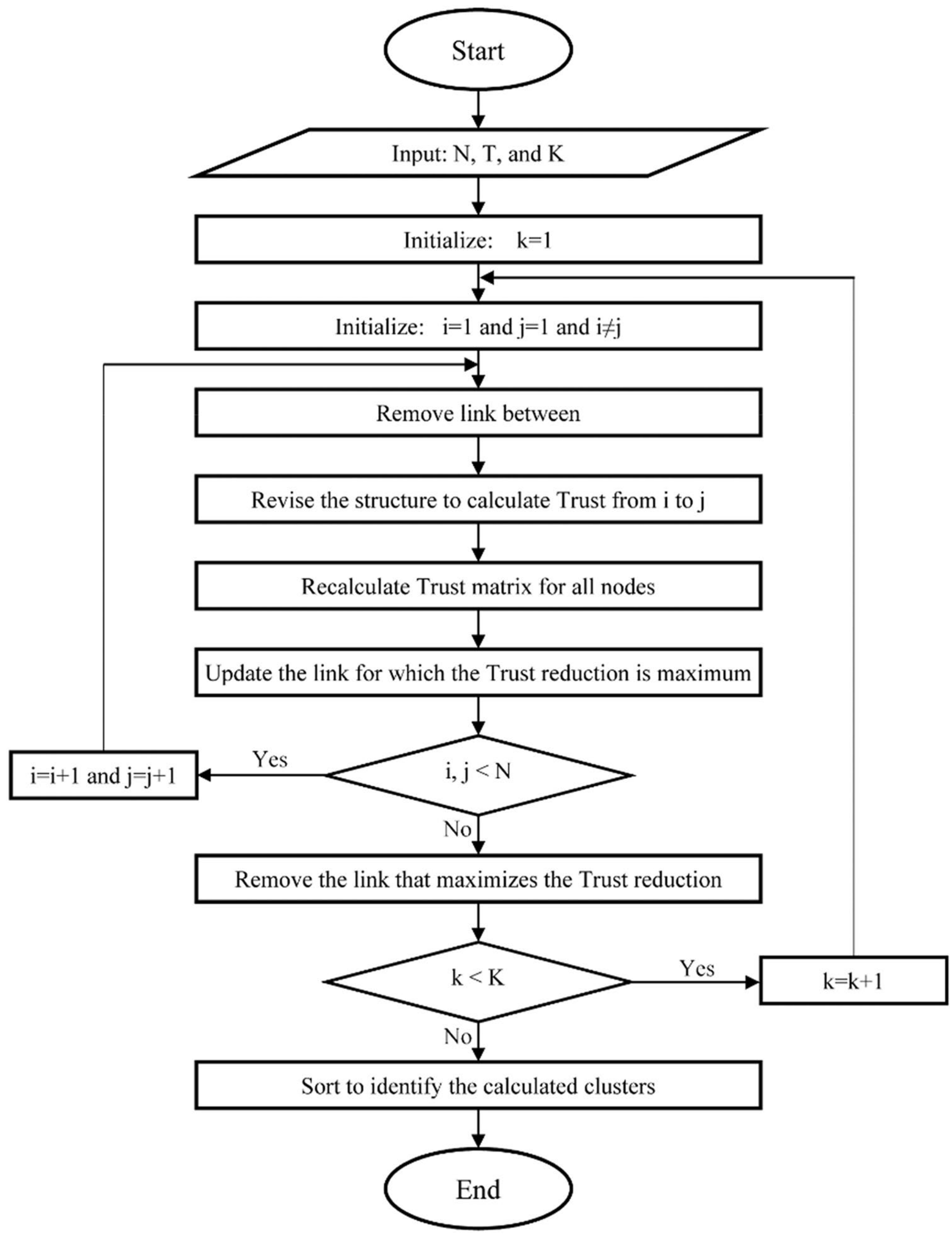

Table 2 The features of social network datasets

Network	Nodes	Edges	Avg of degree	Avg of cluster coefficient	Avg of path length
Zachary's karate club	34	78	4.588	0.588	2.408
Dolphin network	62	159	5.129	0.303	3.357
US football network	115	613	10.661	0.403	2.508
PolBooks network	105	441	8.400	0.488	3.079

slope. Then the algorithm goes to the next node and this process continues again. The removal of edges continues until there is no path left between the nodes of a community and the nodes of neighboring communities. In this case, the algorithm is stopped and independent communities are determined.

In our implementation, because the graph of the studied data is not weighted, the direct reliability coefficient for direct connection and the presence of an edge between two nodes is set equal to one. Finally, after stopping the community detection algorithm and determining the independent communities, the number of nodes of each community is determined and the criterion of modularity is calculated for the set. Figure 2 shows the flowchart of the proposed algorithm.

4 Experimental results

4.1 Benchmark networks

For the experiments, we used four benchmark networks, which are described below, and Table 2 shows their characteristics.

Zachary's karate club. This network is the study and collection of information about the members of a karate club that was done between 1970 and 1972. The club has 34 members and documents the relationship between club members outside the club. During the information gathering process, there were disagreements between the club manager and the club coach, which led to divisions among the members. Almost half of the club members formed another group with their coach. Other members either found a new coach or gave up karate.

The dolphin network. The network covers a particular species of dolphin in New Zealand between 1994 and 2001. There are 62 dolphins in this network, including males and females. If you see a pair of dolphins swimming, an edge is drawn between them as a sign of connection. Different pairs have formed over time. Sometimes no swimming is observed between two specific dolphins, but each can communicate indirectly through its partners.

The US football network. The network includes 115 American football teams competing in the United States throughout a season. Each node represents a team, and if two teams play during a football season, an edge is drawn between them. According to the zoning of the teams' game, which is called the conference, there are 11 conferences. Teams play between 7 and 13 games per conference. Most teams may not play together due to the zoning of games or conferences. If a team is in a conference, it will play most of its games in that conference. Conferences are connected through the games of the winners of each conference.

The PolBooks network. It is an information network in which each node represents a political book sold by Amazon during the 2004 US election (most of which were published and edited in the US). In this network, the connection between the two nodes is that if a book is sold along with another book on the same subject, these two books are connected by an edge. For example, if a reader buys a book, the same reader buys another book with a similar subject. His two purchased books are connected by an edge. In addition, each node in the network is marked as "liberal", "conservative" or "neutral" by reviews written by readers. The goal is to find books with a specific label that identifies specific societies, such as liberal or conservative communities, and so on.

4.2 Community quality evaluation criterion

In this paper, the modularity function Q is used to evaluate the coherence of the structure. The Q function is defined as follows,

$$Q = \frac{1}{2m} \sum_{u,v} \left[A_{u,v} - \frac{k_u k_v}{2m} \right] \delta(C_u, C_v) \tag{7}$$

where, in the above relation, m is the total number of edges. $A_{u,v}$ is the adjacent matrix element. k_u is the node degree of u and k_v is the node degree of v. C_u and C_v are communities that include node u and v. If u and v are in the same community, the $\delta(C_u, C_v)$ will be equal to one, otherwise it will be zero. The maximum modularity value is obtained when all vertices of each community are connected to each other and communities are not connected together. In other words, a higher value of the Q function means that the effect of dividing the community is more pronounced. Usually, in social

Table 3 Clusters and communities detected by trust-based algorithm after applying on the Zachary's karate club

Criterion	Nodes in clusters and communities
The clusters	Cluster_1: 1, 2, 3, 4, 8, 10, 12, 13, 14, 18, 20, 22 Cluster_2: 5, 6, 7, 11, 17 Cluster_3: 9, 15, 16, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34
The real communities	Community_1: 1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 17, 18, 20, 22 Community_2: 9, 10, 15, 16, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34

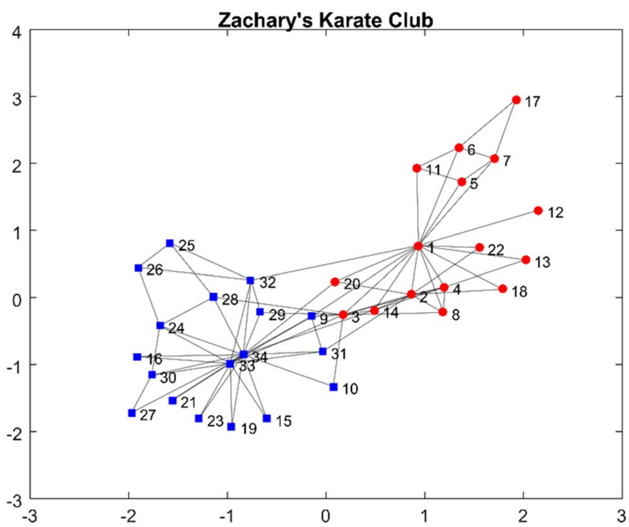

Fig. 3 Detected communities in the Zachary's karate club

network graphs, this value is between 0.3 and 0.7 (Harifi et al. 2021).

4.3 Results

In this subsection, the results of community detection and the value of obtained modularity are presented. It is better to point out that, our experiments consist of two independent parts with independent results. In the first part, the comparison of the proposed algorithm with the entropy-based algorithm (Nikolaev et al. 2015) has been performed. The mentioned algorithm is also Girvan–Newman inspired algorithm that uses entropy centrality for social network analysis and community detection. Therefore, the purpose of the first part is to compare the proposed method with a Girvan–Newman inspired algorithm. For this part, the number of iterations of the main loop of the compared algorithms is at most ten. This number of iterations is sufficient to detect the community, but the maximum amount of modularity is not obtained. In the second part of the experiments, the number of iterations is considered to the extent that the modularity value is maximized. The purpose of the second part is to reach the maximum value of modularity so that the proposed algorithm can be compared with other community detection algorithms. All evaluation experiments have been run on an Intel® Pentium® processor CPU G645 2.90 GHz with 2 GB RAM.

As the first part of the experiments, we have identified the results include clusters and communities. Table 3 shows the results obtained from applying the proposed algorithm to the Karate Club network. After applying the algorithm, three main clusters and two real communities were observed. The value of modularity obtained is 0.2354. The modularity value is also calculated based on the entropy-based method. The modularity value for this network is 0.1857 based on the entropy-based method. As can be seen, the modularity value based on the proposed method is more appropriate than the entropy-based method. Figure 3 shows the detected communities by applying the proposed algorithm on the Karate Club network. In this figure, each color represents an independent community.

Table 4 Clusters and communities detected by trust-based algorithm after applying on the Dolphin network

Criterion	Nodes in clusters and communities
The clusters	Cluster_1: 1, 3, 11, 29, 31, 43, 48 Cluster_2: 2, 8, 18, 20, 23, 26, 27, 28, 32 Cluster_3: 4, 5, 9, 12, 13, 15, 16, 17, 19, 21, 22, 24, 25, 30, 34, 35, 36, 37, 38, 39, 40, 41, 44, 45, 46, 51, 52, 53, 54, 56, 59, 60, 62 Cluster_4: 6, 7, 10, 14, 33, 42, 49, 55, 57, 58, 61 Cluster_5: 47, 50
The real communities	Community_1: 1, 3, 4, 5, 9, 11, 12, 13, 15, 16, 17, 19, 21, 22, 24, 25, 29, 30, 31, 34, 35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 50, 51, 52, 53, 54, 56, 59, 60, 62 Community_2: 2, 6, 7, 8, 10, 14, 18, 20, 23, 26, 27, 28, 32, 33, 42, 49, 55, 57, 58, 61

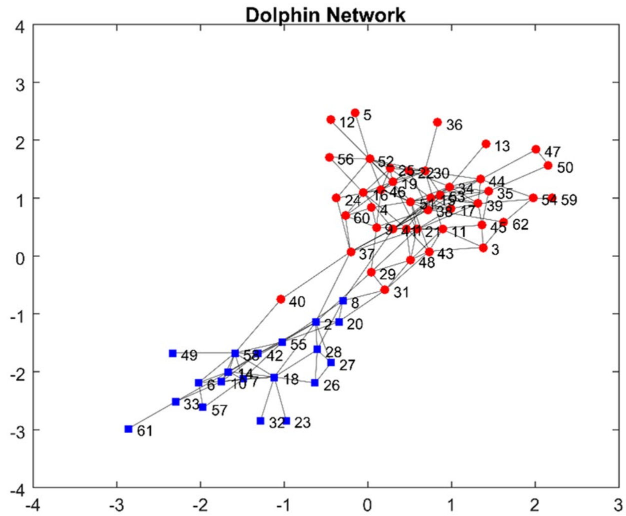

Fig. 4 Detected communities in the Dolphin network

Table 4 shows the results obtained from applying the proposed algorithm to the dolphin network. The results show five clusters and two real communities. The modularity value according to the proposed method is 0.4021. Also, modularity in the entropy-based method showed 0.3774. Comparing the modularity values, we conclude that the proposed algorithm performed better for this network. Figure 4 shows the detected communities by applying the proposed algorithm on the dolphin network. In this figure, each color represents an independent community.

The results obtained from applying the proposed algorithm to the US football network are shown in Table 5. For this data set, after applying the proposed algorithm, eleven clusters and twelve real communities are obtained. Modularity is also calculated for this network. The modularity obtained based on the proposed method is 0.4914. Also, the modularity for this network after applying the entropy-based method is 0.4397. In this network, the modularity criterion in the trust-based method, which is the method proposed in this paper, is more appropriate than the entropy-based method. Figure 5 shows the detected communities by applying the proposed algorithm on the US football network. In this figure, each color represents an independent community.

The results obtained from applying the proposed algorithm to the PolBooks network are also shown in Table 6. After applying the algorithm, six main clusters and three real communities are obtained. The modularity value for this network is 0.4665. Based on the entropy-based method, the modularity value is 0.4528. As can be seen, for this network, the modularity value based on the proposed method is more appropriate than the entropy-based method. Figure 6 shows the detected communities by applying the proposed algorithm on the PolBooks network. In this figure, each color represents an independent community.

Table 5 Clusters and communities detected by trust-based algorithm after applying on the US football network

Criterion	Nodes in clusters and communities
The clusters	Cluster_1: 1, 5, 10, 12, 17, 24, 25, 29, 42, 51, 70, 91, 94, 105
	Cluster_2: 2, 26, 34, 38, 46, 90, 104, 106, 110
	Cluster_3: 3, 7, 14, 16, 33, 40, 48, 61, 65, 101, 107
	Cluster_4: 4, 6, 11, 41, 53, 73, 75, 82, 85, 99, 103, 108
	Cluster_5: 8, 9, 22, 23, 52, 69, 78, 79, 109, 112
	Cluster_6: 13, 15, 19, 27, 32, 35, 39, 43, 44, 55, 62, 72, 86, 100
	Cluster_7: 18, 21, 28, 57, 63, 66, 71, 77, 88, 96, 97, 114
	Cluster_8: 20, 30, 31, 36, 56, 80, 81, 83, 95, 102
	Cluster_9: 37, 59, 60, 64, 98
	Cluster_10: 45, 49, 58, 67, 76, 87, 92, 93, 113
	Cluster_11: 47, 50, 54, 68, 74, 84, 89, 111, 115
The real communities	Community_1: 2, 26, 34, 38, 46, 90, 104, 106, 110
	Community_2: 20, 30, 31, 36, 56, 80, 95, 102
	Community_3: 3, 7, 14, 16, 33, 40, 48, 61, 65, 101, 107
	Community_4: 4, 6, 11, 41, 53, 73, 75, 82, 85, 99, 103, 108
	Community_5: 45, 49, 58, 67, 76, 87, 92, 93, 111, 113
	Community_6: 37, 43, 81, 83, 91
	Community_7: 13, 15, 19, 27, 32, 35, 39, 44, 55, 62, 72, 86, 100
	Community_8: 1, 5, 10, 17, 24, 42, 94, 105
	Community_9: 8, 9, 22, 23, 52, 69, 78, 79, 109, 112
	Community_10: 18, 21, 28, 57, 63, 66, 71, 77, 88, 96, 97, 114
	Community_11: 12, 25, 51, 60, 64, 70, 98
	Community_12: 29, 47, 50, 54, 59, 68, 74, 84, 89, 115

As mentioned, modularity is a measure of the quality of communities. In this criterion, the strength of the connection in the communities is measured in comparison with the base strength. Table 7 shows the obtained modularity values aggregated as a table for better comparison. The modularity value of the trust-based method for Zachary's karate club, Dolphin network, US football network, and PolBooks network is 4.9, 2.5, 5.2, and 1.4% better than the entropy-based method, respectively. These results indicate a better

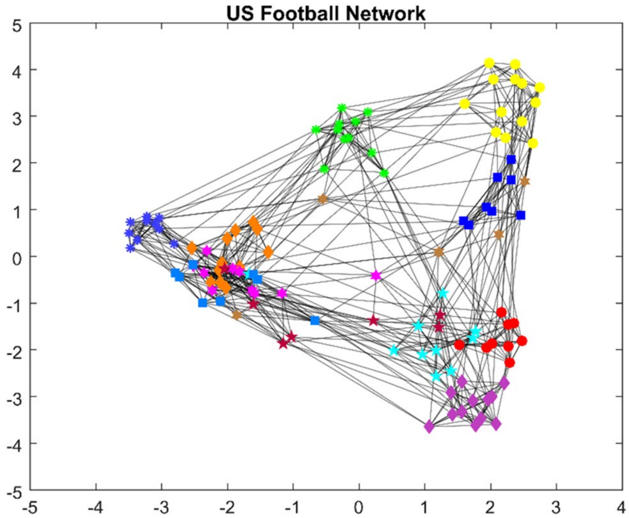

Fig. 5 Detected communities in the US football network

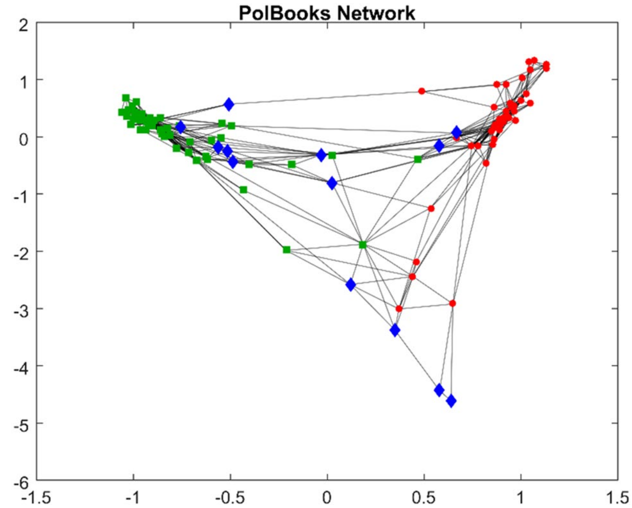

Fig. 6 Detected communities in the PolBooks network

Table 6 Clusters and communities detected by trust-based algorithm after applying on the PolBooks network

Criterion	Nodes in clusters and communities
The clusters	Cluster_1: 1, 2, 3, 5, 6, 7, 8, 30 Cluster_2: 4, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 54, 55, 56, 57, 58 Cluster_3: 29, 31, 32, 67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 87, 88, 89, 90, 91, 92, 93, 97, 98, 99, 101 Cluster_4: 52, 53, 59, 65, 66, 68, 69, 70, 86, 104, 105 Cluster_5: 60, 61, 63, 64, 100 Cluster_6: 62, 94, 95, 96, 102, 103
The real communities	Community_1: 31, 32, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 71, 72, 73, 74, 75, 76, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103 Community_2: 1, 5, 7, 8, 19, 29, 47, 49, 52, 70, 77, 104, 105 Community_3: 2, 3, 4, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48, 50, 51, 53, 54, 55, 56, 57, 58, 59, 78

Table 7 The obtained modularity values

Network	Entropy-based	Trust-based
Zachary's karate club	0.1867	0.2354
Dolphin network	0.3774	0.4021
US football network	0.4397	0.4914
PolBooks network	0.4528	0.4665

and more effective performance of the proposed method in detecting social network communities.

As the second part of the experiments, we once again performed the experiments with more iterations. This experiment was performed to maximize the value of modularity by increasing the iterations of the main loop of the algorithm. Where in each iteration changes in trusts cause a change in the amount of modularity. The goal here is to achieve the maximum modularity value so that the algorithm can be compared to some of the existing non Girvan–Newman inspired methods. We once again chose the entropy-based method for this part of the experiment. Also, we added the Hybrid-EPC (Harifi et al. 2021), LPA-FCM (Deng et al. 2019), and LPA-NI (Zhang et al. 2017) to our comparisons. The Hybrid-EPC is a hybrid metaheuristic method for community detection. The LPA-FCM method is a label propagation algorithm using fuzzy C-means. The LPA-NI is a label propagation algorithm based on node importance and labels influence. The results of the mentioned methods have been taken from its reference and we only run the entropy-based method and the proposed trust-based method once again to reach the maximum modularity value. Table 8 shows the maximum modularity value of the compared algorithms.

To achieve the maximum value of modularity, we considered a different number of iterations for each network. For example, for Zachary's karate club network, the highest modularity is obtained with several iterations between 25 and 30. We also considered that more iterations may result in more granular clusters. However, for a large network like US football, the number of iterations increased to 200. Table 8 shows that the proposed algorithm based on trust centrality, which is an improvement on the Girvan–Newman community detection algorithm, can provide acceptable results if the number of iterations increases. According to the table,

Table 8 Maximum modularity values obtained by algorithms

Network	Hybrid-EPC	LPA-FCM	LPA-NI	Entropy-based	Trust-based
Zachary's karate club	**0.4200**	0.4198	0.4151	0.3917	**0.4200**
Dolphin network	0.5280	0.5264	0.5143	0.4941	**0.5296**
US football network	**0.6080**	0.6046	0.5805	0.5952	0.6068
PolBooks network	0.5310	0.5257	N/A	0.5234	**0.5343**

The best results are marked in bold

in three networks it has better results than other methods, and only in one network, the Hybrid-EPC method is better. Hybrid-EPC uses a metaheuristic algorithm to achieve the maximum value of modularity, and given the benefits of soft computing (Harifi et al. 2020), it is clear that it provides acceptable results. However, the proposed method has also provided acceptable results, so that it is completely better than the entropy-based method, which is one of the improvements of the Girvan–Newman algorithm. It also has acceptable results in comparison with LPA-based methods.

By redefining trust-based centrality, different locations of centrality can be achieved. This makes local central and global central nodes more distinct. This increases the flexibility and efficiency of iteration-based community detection. These advantages are used to overcome some limitations of the Girvan–Newman community detection algorithm. As expected, the computational efficiency of centralization values based on trust over all network nodes is high. The maximum value of modularity also proves that the proposed trust-based community detection algorithm has been able to improve the Girvan–Newman algorithm well.

5 Conclusions

One of the most important issues that have been extensively studied is the issue of community detection in large networks. Social networks are growing so there is a need to manage and categorize communities. There are many methods and algorithms in the field of community detection, but solving the problem of community detection in a coherent framework has not been performed so far. Centrality determines the relative importance of a vertex in a network graph. In this paper, the centrality criterion based on trust was used to analyze the network. Trust as centrality is one of the main characteristics of social network users and is usually considered in both direct and indirect ways. Some of the actions of users, such as activity, communicating, and some concerns in the field of social networks, to some extent indicate the development of their trust. In this paper, the community graph was first created based on trust, which was done by calculating trust for each node. Then the community detection algorithm was applied to the graph in which node trust was used to identify the community. The ultimate goal of the algorithm has been to find the correct and effective communities that make up the network in a way that uses the trust of users to community detection. The algorithm was run on four networks as experiments. The experiments consisted of two independent parts. The first part was to apply the algorithm to detect the clusters and communities and compare it to a Girvan–Newman inspired method. The second part was the implementation of the algorithm with a large number of iterations to achieve the maximum value of modularity and compare it with other community detection algorithms. Modularity criteria were also used to assess the quality of communities in both independent parts of the experiments. The results showed that in the worst case 1.4% improvement and the best case 5.2% improvement of community detection was obtained through the proposed trust-based algorithm resulted from the first part experiment. Also, results were acceptable based on the second part of the experiments.

Detecting communities requires further investigation and analysis. This has been one of the challenges of this research, which means that it can never be enough to study and analyze a network based on various criteria and tests performed on it, and there is always a need for further study and analysis. The study also faced hardware limitations. Analyzing graphs with multiple nodes and edges requires a powerful processor and a strong computer system configuration. This limitation makes processing and analysis time-consuming. Hardware limitations also make it almost impossible to analyze the proposed algorithm on a larger and more complex dataset or network. Therefore, considering the challenges and limitations of this study, the use of datasets with much larger and more complex networks and the effectiveness of the proposed method for this data could be the subject of future researches. As future work, we will try to improve the proposed method using soft computing methods, including methods based on metaheuristic algorithms.

Acknowledgements The authors would like to thank the anonymous referees for their valuable comments and suggestions that improved the presentation of this paper.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Ahn YY, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. Nature 466(7307):761–764

Alghamdi E, Greene D (2019) Active semi-supervised overlapping community finding with pairwise constraints. Appl Netw Sci 4(1):1–27

Barabási AL (2013) Network science. Philos Trans R Soc A 371(1987):20120375

Carnivali GS, Vieira AB, Ziviani A, Esquef PA (2020) CoVeC: coarse-grained vertex clustering for efficient community detection in sparse complex networks. Inf Sci 522:180–192

Chen X, Xia C, Wang J (2018) A novel trust-based community detection algorithm used in social networks. Chaos Solitons Fract 108:57–65

Danowski JA (2012) Social media network size and semantic networks for collaboration in design. Int J Organ Des Eng 2(4):343–361

Deng ZH, Qiao HH, Song Q, Gao L (2019) A complex network community detection algorithm based on label propagation and fuzzy C-means. Physica A 519:217–226

Du N, Wang B, Wu B, Wang Y (2008) Overlapping community detection in bipartite networks. In 2008 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, IEEE, pp 176–179

Freeman LC (2011) The development of social network analysis—with an emphasis on recent events. SAGE Handb Soc Netw Anal 21(3):26–39

Gilbert F, Simonetto P, Zaidi F, Jourdan F, Bourqui R (2011) Communities and hierarchical structures in dynamic social networks: analysis and visualization. Soc Netw Anal Min 1(2):83–95

Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc Natl Acad Sci 99(12):7821–7826

Grabner-Kräuter S, Bitter S (2015) Trust in online social networks: a multifaceted perspective. Forum Soc Econ 44(1):48–68

Guo K, He L, Chen Y, Guo W, Zheng J (2020) A local community detection algorithm based on internal force between nodes. Appl Intell 50(2):328–340

Harifi S, Mohammadzadeh J, Khalilian M, Ebrahimnejad S (2020) Giza Pyramids Construction: an ancient-inspired metaheuristic algorithm for optimization. Evol Intel. https://doi.org/10.1007/s12065-020-00451-3

Harifi S, Mohammadzadeh J, Khalilian M, Ebrahimnejad S (2021) Hybrid-EPC: an Emperor Penguins Colony algorithm with crossover and mutation operators and its application in community detection. Prog Artif Intell 10(2):181–193

Hu F, Liu J, Li L, Liang J (2020) Community detection in complex networks using Node2vec with spectral clustering. Physica A 545:123633

Ji P, Zhang S, Zhou Z (2020) A decomposition-based ant colony optimization algorithm for the multi-objective community detection. J Ambient Intell Humaniz Comput 11(1):173–188

Jiang H, Liu Z, Liu C, Su Y, Zhang X (2020) Community detection in complex networks with an ambiguous structure using central node based link prediction. Knowl-Based Syst 195:105626

Jin JH, Park SC, Pyon CU (2011) Finding research trend of convergence technology based on Korean R&D network. Expert Syst Appl 38(12):15159–15171

Kernighan BW, Lin S (1970) An efficient heuristic procedure for partitioning graphs. Bell Syst Tech J 49(2):291–307

Kırer H, Çırpıcı YA (2016) A survey of agent-based approach of complex networks. Ekonomik Yaklasim 27(98):1–28

Kumar S, Carley KM (2018) Towards group-activities based community detection. In: Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers, ACM, pp 1178–1183

Lancichinetti A, Fortunato S, Radicchi F (2008) Benchmark graphs for testing community detection algorithms. Phys Rev E 78(4):046110

Mahajan SP, Raipurkar AR (2018) Network based community detection by using bisecting hierarchical clustering. HELIX 8(5):4077–4081

Miyauchi A, Kawase Y (2016) Z-score-based modularity for community detection in networks. PLoS ONE 11(1):e0147805

Moosavi SA, Jalali M, Misaghian N, Shamshirband S, Anisi MH (2017) Community detection in social networks using user frequent pattern mining. Knowl Inf Syst 51(1):159–186

Muller E, Peres R (2019) The effect of social networks structure on innovation performance: a review and directions for research. Int J Res Mark 36(1):3–19

Nerurkar P, Chandane M, Bhirud S (2019) A comparative analysis of community detection algorithms on social networks. In: Verma N, Ghosh A (eds) Computational intelligence: theories, applications and future directions—Volume I. Advances in intelligent systems and computing. Springer, pp 287–298

Nicosia V, Mangioni G, Carchiolo V, Malgeri M (2009) Extending the definition of modularity to directed graphs with overlapping communities. J Stat Mech Theory Exp 2009(03):P03024

Nikolaev AG, Razib R, Kucheriya A (2015) On efficient use of entropy centrality for social network analysis and community detection. Soc Netw 40:154–162

Rosvall M, Delvenne JC, Schaub MT, Lambiotte R (2019) Different approaches to community detection. In: Batagelj V, Ferligoj A (eds) Doreian P. Advances in network clustering and blockmodeling, Wiley, pp 105–119

Wang GA, Jiao J, Abrahams AS, Fan W, Zhang Z (2013) ExpertRank: a topic-aware expert finding algorithm for online knowledge communities. Decis Support Syst 54(3):1442–1451

Xiao J, Ren HF, Xu XK (2020) Constructing real-life benchmarks for community detection by rewiring edges. Complexity 2020:1–16

Yang Z, Algesheimer R, Tessone CJ (2016) A comparative analysis of community detection algorithms on artificial networks. Sci Rep 6:30750

Zhang Z, Wang Z (2015) Mining overlapping and hierarchical communities in complex networks. Phys A Stat Mech Appl 421:25–33

Zhang L, Ye Q, Shao Y, Li C, Gao H (2014) An efficient hierarchy algorithm for community detection in complex networks. Math Probl Eng 2014:1–12

Zhang XK, Ren J, Song C, Jia J, Zhang Q (2017) Label propagation algorithm for community detection based on node importance and label influence. Phys Lett A 381(33):2691–2698

Zhao P, Zhang CQ (2011) A new clustering method and its application in social networks. Pattern Recogn Lett 32(15):2109–2118