

TAXONOMY AND SURVEY OF COMMUNITY DISCOVERY METHODS IN COMPLEX NETWORKS

Ahlem Drif¹ and Abdallah Boukerram²

¹Computer Science Department, University of Setif1, Algeria

²Computer Science Department, University of Abderrahmane Mira, Bejaia, Algeria

ABSTRACT

The community detection in complex networks has attracted a growing interest and is the subject of several researches that have been proposed to understand the network structure and analyze the network properties. In this paper, we give a thorough overview of different community discovery strategies, we propose taxonomy of these methods, and we specify the differences between the suggested classes which helping designers to compare and choose the most suitable strategy for the various types of network encountered in the real world.

KEYWORDS

Community detection methods, complex networks, quality function, graph, social networks.

1. INTRODUCTION

In complex networks, the communities are groups of nodes which share probably a common proprieties and/or similar functions. The communities may be correspond, for example, to groups of Web pages accessible over the Internet that have the same subject [19], functional modules as cycles and pathways in metabolic networks [25], [47], a set of people or groups of people with some pattern of contacts or interactions between them [24], [38], and subdivisions in the food webs [49], [32]. The main objective of community detection is to discover a pertinent community structure. See Fig. 1 for a toy example with this kind of a structure which shows the network structure of the Web site of Ferhat Abbas University (Algeria). As we can see from the figure, this network contains six communities that identify groups of users with similar behaviour for which personalized versions of the Web site may be created. The principle idea for community discovery domain proposed in [24], [46] was to focus on a different measure of the quality of a division other than the simple cut size or its variants [43].

The community discovery methods have attracted considerable interest and curiosity from the science community in recent decades. As the field of community identification has grown quite popular and the number of published proposals for community discovery algorithms as well as reported applications is high, we do not even pretend to be able to give an exhaustive survey of all the methods, but rather an explanation of the methodologies commonly applied and pointers to some of the essential publications related to each research branch in order to provide a taxonomy for community discovery methods. We begin by given basic definitions of community structure and the essential elements of this thematic in section 2. In section 3, we propose taxonomy of the detection community methods and provide a survey defining the basic idea of each one, the measure and/or process used to identify the community structure and to which class each method belongs. Conclusion is given in section 4.

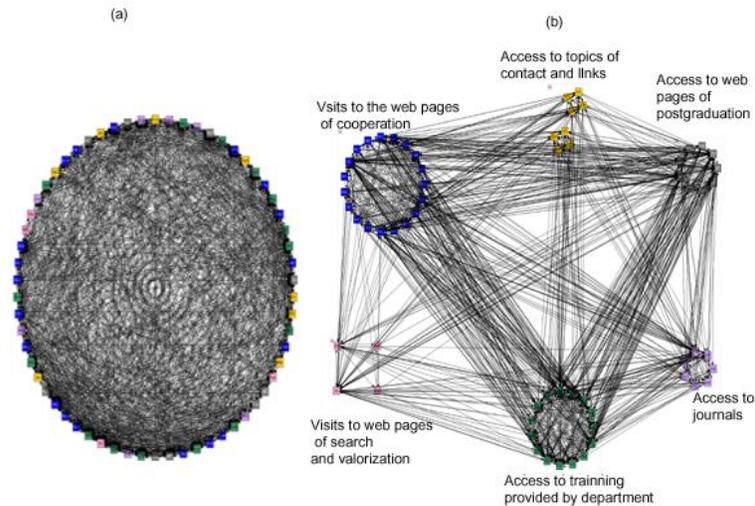


Figure 1. (a) The network structure of the Web site of Ferhat Abbas University (Algeria). The structure identifies the users' session and all the sessions (the nodes represent resources and edges represent the browsing sequences of users during each session).
 (b) We have obtained 6 communities according to the users' behavior. Communities, labeled by colors, were detected by applying Fast algorithm [45].

2. COMMUNITIES DESCRIPTION

Despite the large amount of study in this area, a consensus on what is the definition of community has not been reached. Conceptually, the definitions can be separated into two main categories, self-referring and comparative definitions. Central to all such definitions is the concept of sub graph.

2.1. Graph concepts

We consider a graph $G = (V, E)$, with V is the set of vertices and E is the set of edges. Vertices are also known as nodes, points and (in social networks) as actors, agents or players. Two vertices u and v are adjacent if there exist an edge (u, v) that connects them. In an undirected graph, each edge is an unordered pair v, w . In a directed graph, edges are ordered pairs. The number of nodes in a graph is usually denoted n while the number of edges is usually denoted m . Graph is used to represent all types of complex networks. In this survey, the presented algorithms divide the graph top down into communities, or also work bottom up merging singleton sets of nodes iteratively into communities.

2.2. Comparative definitions

The comparative definitions consist to compare the number of internal links to the number of external links. Authors sometimes use a criterion of similarities between nodes to discover the community structure. The intuitive notion of criterion of similarities derives both from the relative strength, frequency, density, or closeness of links within a subgroup, and the relative weakness, infrequency, sparseness, or distance of links from subgroup members to non-members. A community may be defined as LS set that is a subgraph definition that compares links within the subgroup to links outside the subgroup by focusing on the greater frequency of links among subgroup members compared to the links from subgroup members to outsiders [59]. Comparative

definitions include also that of strong community, and that of weak community that introduced by Radicchi et al [52].

2.3. Self-referring definitions

The self-referring definitions consider the subgraph alone. It identifies classes of subgraphs like cliques, n -cliques, k -plexes, etc... They are maximal subgraphs, which cannot be enlarged with the addition of new vertices and edges without losing the property which defines them. Self-referring definitions can include the definition introduced by Newman in [41] which defines a community as indivisible subgraph.

2.4. Quality Functions

How can we know if detected communities are good or no and how to value such partitions? What is the better partition for the network in question? When we break the dendrogram to obtain the level of adequate partition of the network or else a large number of appropriate communities? To answer to those questions, Newman and Girvan [46] introduce a measure of quality of a particular partition which they called modularity. The modularity is based on assortative mixing measure proposed by Newman [44]. The modularity is defined in the following way: Given a particular division of a network into k communities, let e denote a $k \times k$ matrix whose element is the fraction of all edges in the network which connects vertices in community i to those in community j . The trace of this matrix gives the fraction of edges in the network that connect vertices in the same community:

$$Tre = \sum_i e_{ii} \quad (1)$$

A good community partition should have a high value of the trace, but the trace on its own is not a good indicator of the quality of the division since, for example, placing all vertices in a single community would give the maximal value of $Tr = 1$ without giving any information about the community structure. The row (column) sum represents the fraction of edges that connect to vertices in community i , is then defined as

$$a_i = \sum_j e_{ij} \quad (2)$$

If the network does not exhibit community structure, or if the partitions are allocated without any regard to the underlying structure, the expected value of the fraction of links within partitions can be estimated. It is simply the probability that a link begins at a node in i , a_i , multiplied by the fraction of links that end at a node in i , a_i . So the expected number of intra-community links is just $a_i a_i$. On the other hand we know that the real fraction of links exclusively within a partition is e_{ii} . So, we can compare the two directly and sum over all the partitions in the graph.

$$Q = \sum_i (e_{ii} - a_i^2) = Tre - \|e^2\| \quad (3)$$

It measures the fraction of edges in a community, minus the expected value of the same quantity in a network with the same community divisions but random connections between the nodes. If a particular division gives no more with in community edges that would be expected by random chance the modularity is zero. Values other than 0 indicate deviations from randomness, and values above 0.3 indicate a modular structure [45]. In practice, values above 0.7 are rare, and

indicate a very clear structure. However, it is possible that the partition of high modularity don't suit to more pertinent community partitions [20].

3. PROPOSITION OF TAXONOMY OF COMMUNITY DISCOVERY METHODS

The problem of community detection goes to traditional approaches in computer science which were the graph partitioning and the data clustering approaches. However, these traditional approaches are not adapted to discover a structure community. In the traditional algorithm of partitioning graph [31] [51] [18], the number and the size of groups must be specified in input. In data clustering methods, the graphs considered have not the specific proprieties of complex networks. Moreover, in hierarchical clustering when many partitions recovered how can we know which one is the best? Therefore, several approaches were proposed in order to deal with the incapacity of classical methods and to provide a significant community structure.

In this section, we survey the research in the area of communities identification, suggest a classification, and we situate some of the research with respect to the proposed taxonomy. According to our analysis, we have proposed taxonomy for community discovery methods in order to well locate these different approaches (see Fig.2 and Fig.3). It seems that this taxonomy allows us to explicit the different choices of one of these community discovery approaches, decides which is the most suitable for a given type of complex networks and for which reasons are. These appropriate choices will give us a considerable flexibility in choice of the division quality functions. Our classification is based on the three following points of view:

A. Agglomeration versus divisive:

According to manner of grouping the nodes in groups, Jain and Dubes [28] defined two distinct approaches to do: agglomerative and divisive. As the same, we conclude that all the community discovery methods use an agglomerative or divisive approach, depending on whether the partition is refined or coarsened during each iteration:

- Bottom-up or agglomerative algorithms that start with each node in its own singleton community or another set of small initial communities, iteratively merging these communities into larger ones.
- Top-down or divisive algorithms that split the network iteratively or recursively into smaller and smaller communities.

B. Stochastic versus deterministic:

In the context of evaluation of performance of community discovery methods, we have noted that the deterministic and the stochastic methods differed on function of time execution and partition quality. Therefore, we think that is fort important to distinguish between the deterministic method and the stochastic ones.

C. Various computer implementations of community discovery methods:

In many approaches, the communities are characterised and detected, directly or indirectly, by some global proprieties of graph, such intermediary, betweenness, etc..., or by some process as random walks, synchronisation, etc..., communities also may be interpreted as a topological organization form. All of these remarks allow us to classify the existent methods according to the technical details or the process used during the discovery of communities.

3.1. Agglomerative Methods

Agglomerative methods start with a state in which each node (or a small set of nodes) is the sole member (or members) of one of n communities, they repeatedly join communities together in pairs until obtain one community which corresponding to all complex network. In this section, we present the agglomerative approaches, we give a classification based on which the method is deterministic or stochastic, and then we give a grouping of methods in function of the process used during the discovery of communities.

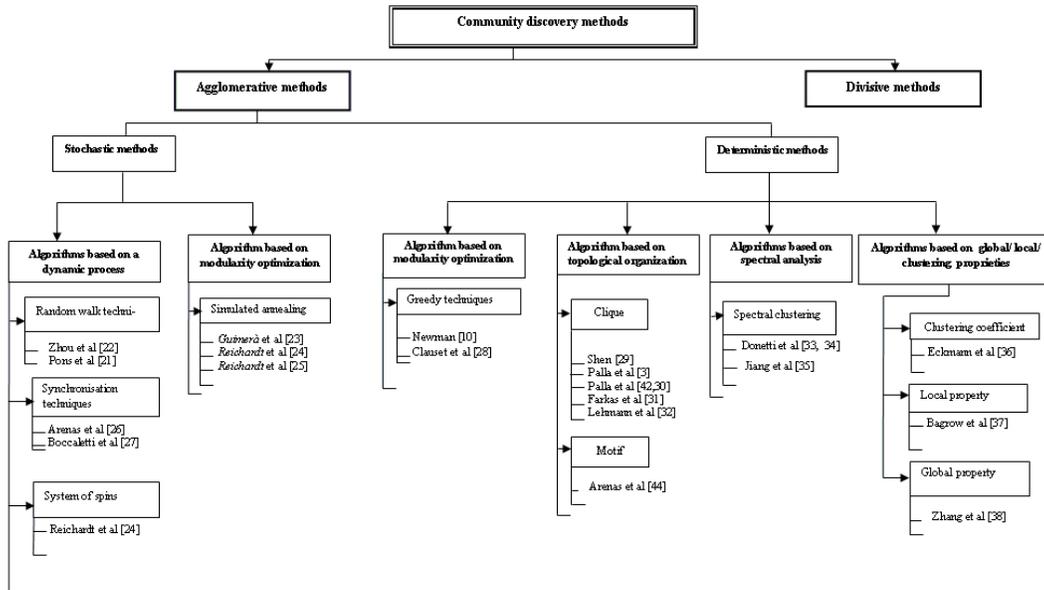


Figure 2. Community discovery methods taxonomy: Agglomerative methods.

3.1.1 Stochastic methods

In this section, we present the agglomerative approaches that use a stochastic process in which future states only depend on the current state, not the past, taking values from some countable state space.

3.1.1.1 Algorithms based on a dynamic process

This Section describes methods employing processes running on the graph and focusing on spin-spin interactions, random walk and synchronization.

3.1.1.1.1 Random walk techniques

The method suggested by Zhou and Lipowsky [66] is based on the average number of stages so that a Brownian particle (random movement of a particle) reaches a given node since another source node. The authors used the Brownian movement to introduce the Netwalk algorithm (NW) which is based on the proximity index. The proximity index for each nearest-neighboring pair of nodes i and j is defined

$$\Lambda(i, j) = \frac{\sqrt{\sum_{k \neq i, j}^N [d_{ik} - d_{jk}]^2}}{(N - 2)} \quad (4)$$

If two nearest-neighboring nodes i and j belong to the same community, then the mean-first-passage-time d_{ik} from i to any another node k will be approximately equal to that from j to k . So, it will be small if i and j belong to the same community and large if they belong to different communities. Initially, the Netwalk algorithm considers each node as a single community. Then it merge the two communities with the lowest proximity index into a single community and then update the proximity index between this new community and all the other remaining communities that are connected to it. This merging process is continued until all the nodes are merged into a single community corresponding to the whole network. This algorithm runs in time $O(n^3)$ which make the application of these approaches on large graphs inadmissible.

Pons et al. [50] proposed the Walktrap algorithm by using the intuitive report that the random walks will be made trap in zones dense, in other words when a walker is in a community it has a strong probability of remaining in the same community at the following stage. The authors define metric of distance which is related to the spectral approaches which are based on the fact that two close nodes belonging to the same community have similar components on the principal eigenvectors. The algorithm computes the connected components, and applies then an agglomerative algorithm which discovered communities on connected subgraphs. This algorithm gives better performance in $O(n.m.\log(n))$ but when the length of the steps becomes important, the quality of the results decreases.

3.1.1.1.2 Synchronization techniques

Using the synchronization process to discovery communities based on the idea that there is a relationship between topological scales and dynamic time scales in complex networks. Several studies show [4], [7] that high densely interconnected sets of oscillators - where oscillators are placed at the nodes - synchronize more easily that those with sparse connections. This scenario suggests that for a complex network with a non-trivial connectivity pattern, starting from random initial condition that highly interconnected oscillators forming local clusters will synchronize first, whereas a full synchronization requires a longer time. This process occurs at different time scales if a clear community structure exists. In [7], the algorithm scales in a time $O(n.m)$, or (n^2) on sparse graphs, and gives good results on practical examples. However, synchronization-based algorithms may not be reliable when communities are very different in size.

3.1.1.1.3 System of spins

System of spins is another promising dynamic process to reveal communities in complex networks. Reichardt et al. [53] proposed a discovered community algorithm which is based on the Potts models with Q states. The Potts model is one of the most models used in statistical physics in order to describe the behaviour of the magnetic bodies [30]. It corresponds to model these bodies like the spins with Q states located at the nodes of a network and which are in interaction between neighbours in order to align its for a ferromagnetic body, or with being well in opposition for an antiferromagnetic body, according to the sign of the constant of coupling. Fu and Anderson [23] showed by analogy that there is a relation between the energy of the physical systems, which is represented by the Hamiltonian and the cost function in an optimization combinative problem. The Hamiltonian of a spin is given by

$$H = -J \sum_{(i,j) \in E} \delta_{\sigma_i \sigma_j} + \gamma \sum_{s=1}^q \frac{n_s(n_s - 1)}{2} \quad (5)$$

Here E is the set of edges, $\sigma_i (i = 1..n)$ denotes the individual spins which are allowed to take q values. n_s indicates the number of spins that have s spin such that $\sum_{s=1}^q n_s = N$, j is the

ferromagnetic interaction strength, δ : is a positive parameter, and σ is the Kronecker delta. Each node is characterized by a spin which can have q possible values. The first sum is the standard ferromagnetic Potts term which represents a homogeneous distribution of the spins in the network, and is minimized by $H_{ferr} = -JM$. The second term sums up all the possible pairs of spins which have equal value. It represents the diversity of the configuration of spins or the existing classes of spins. We find the system fundamental state in order to define the community structure. The communities correspond to the classes of nodes having equal values of spin. The q number of possible spins corresponds to the maximum number of communities which we can detect and it must be selected so that it is higher than the communities effective number. The authors use Monte Carlo single spin flip heat-bath algorithm to determine the community structure. The energy optimization system, which is represented by the second term of the Hamiltonian, is found by simulated annealing algorithm [54]. This minimization of energy corresponds to support the edges intra-community and to optimize the edges inter-community. The algorithm has a non deterministic and non hierarchical nature, so it is able to detect the affiliation of the nodes which belong to the overlapped communities, and it allows the quantification of the communities' stability. However, simulated annealing is not a method of total optimization effective and the algorithm is not convenient to be applied to wide area networks.

3.1.1.2 Algorithms based on modularity optimization

Using the modularity optimization approaches to discover communities comes from the idea that a great value of modularity (Q) represents a good partition in communities. Several studies proposed to optimize this value of all the possible partitions in order to find the best modularity.

3.1.1.2.1 Simulated annealing

Guimer et al. [26] show that finding the modularity of a network is analogous to finding the ground-state energy of a spin system. They demonstrate that, due to fluctuations, stochastic network models give rise to modular networks. The authors were employed a simulated annealing procedure for modularity optimization. This optimization is based on local moves, where a single node is shifted from one cluster to another, taken at random, and a global moves, consisting of merges and splits of communities. Simulated annealing converges generally more closely towards the optimal solution but it can be used only for small networks.

3.1.2 Deterministic methods

In this section, we describe the agglomerative approaches that use deterministic methods in order to reveal correctly all the communities.

3.1.2.1 Algorithms based on modularity optimization

A great value of modularity represents a good division of a network into communities, and then one should be able to find such good divisions by searching through the possible candidates for ones with high modularity. While finding the global maximum modularity over all possible divisions seems hard in general, reasonably good solutions can be found with approximate optimization techniques. In this section, we survey the algorithms which are based on a greedy optimization.

3.1.2.1.1 Greedy techniques

The fast algorithm proposed by Newman [45] uses a greedy optimization in which, starting with each node being the sole member of each community, we repeatedly join together the two

communities whose amalgamation produces the largest increase in Q but don't join the pair of communities whose there are no edges between them. For a network of n nodes, after $(n - 1)$ such joins, the algorithm stops when the results of merging process is a single community. Thus the total running time is (mn) , or $O(n)^2$ on a sparse graph. The output of the algorithm can be represented in the form of a dendrogram and the optimal cross-section of the dendrogram found by looking for the optimal value of Q . Clauset et al [10] proposed an improvement of the fast algorithm which performs well in $O(n \log^2 n)$, the proposed method computes the change in modularity and finds a pair of communities $i; j$ with the largest ΔQ_{ij} . These algorithms based its decisions on local information of the various communities. However, the structure of community isn't a local quantity, whereas the algorithms that based on local information such the GN algorithm [24] finds this structure more correctly.

3.1.2.2 Algorithms based on topological organization

Several approaches of discovered communities are based on the observation that a community can be interpreted as the union of a set of sub graphs which share nodes between them. This observation based on such a topological organization allows expressing the tendency of the nodes to gather in communities. In this section, we describe the methods that belong, in the diagram of our classification, to the topological organization class which are by nature agglomerative methods.

3.1.2.2.1 Clique

Palla et al [47], [12] have defined a new method that employed a percolation process to detect the overlapped communities in the networks. The clique percolation method (CPM) uses the high density of the internal edges which seem to form cliques. So, two k -cliques are adjacent if they share $(k - 1)$ nodes, and a community is defined as a union of all the k -cliques that can be reached by chains of adjacent k -cliques. Such communities can be better visualized using a k -clique template which is an isomorphic object for a complete graph of k -nodes. This object can be placed on a k -clique of the network and be moved towards the adjacent k -clique by changing one of its nodes and keeping its others $(k - 1)$ nodes. Thus, the community (k -clique percolation cluster) is the largest connected subgraph which can be entirely explored while rolling the object on the k -clique.

An extension of CPM algorithm was proposed for the weighted networks [17] and the directed networks [48]. To detect communities, from k -cliques, the maximal cliques should first of all be computed. The algorithm complexity is exponential and directly proportional to the size of the network, nevertheless the authors proved that the algorithm can be carried out in an acceptable short time on real networks with up to 10^5 nodes. However, CPM method supposes that the graph has a large number of cliques, thus it can fail to detect significant partitions for graphs containing just a few cliques.

Lehmann et al [34] have addressed the problem of community detection on bipartite networks and it has been shown that if the bipartite information is available, the proposed biclique community detection algorithm retains all of the advantages of the k -clique algorithm [12], but avoids discarding important structural information when performing a one-mode projection of the network.

In [56], Shen et al. presented an algorithm to detect at the same time the hierarchy of the communities and their overlapping, using the whole of maximal cliques. The maximal cliques whose the nodes belong to others largest maximal cliques are called subordinates maximal cliques and the majority of them have small sizes. The subordinate maximal cliques may degrade the

performance of the algorithm and should be eliminated. Thus, subordinate maximal cliques that are lower than a threshold k are eliminated. In real-world networks, the threshold value k is typically between 3 and 6. After the elimination phase, some nodes that don't belong to any remaining maximal cliques are called subordinate nodes. Initially, the algorithm finds all the maximal cliques by using Bron-Kerbosch algorithm [9]. Then, the subordinate maximal cliques are neglected, and each of the maximal cliques and each of the subordinate nodes are determined as initial communities. Then, it repeatedly computes the similarity between each pair of communities and join together the two communities whose amalgamation produces a maximum similarity. The similarity between two communities C_1 and C_2 is defined as

$$M = \frac{1}{2m} \sum_{v \in C_1, w \in C_2, v \neq w} [A_{vw} - \frac{k_v k_w}{2m}] \quad (6)$$

Where A_{vw} is the adjacency matrix, k_v is the degree of the node v , and $m = \frac{1}{2} \sum_{vw} A_{vw}$ is the total number of the edges in the network. The authors of the same paper [56] introduced an extended modularity to determine the quality of division in communities

$$EQ = \frac{1}{2m} \sum_i \sum_{v \in C_i, w \in C_i} \frac{1}{O_v O_w} [A_{vw} - \frac{k_v k_w}{2m}] \quad (7)$$

Where O_v is the number of communities containing node v . The algorithm determines the suitable cut of the dendrogram according to the maximum value of EQ . EAGLE algorithm don't have only the capacity to detect sub communities until none of them can be divided (indivisible graph), but also it detects the communities which overlap without loss of their hierarchical details. However, the algorithm is time-consuming due to the research of all the maximum cliques in the network.

3.1.2.2 Motif

Arenas et al. [3] redefined the communities as set of classes which contains motifs. The motifs make it possible to represent the suitable structure of the network. The authors developed a modularity extensions such the motif modularity which was defined as the fraction of motifs inside the communities minus the fraction in a random network which preserves the nodes strengths. Then a generalization of modularity was proposed by relaxing the condition that all nodes of the motif should be fully inside the modules. The authors proposed path modularity and envisaged the length of paths in order to give the best partitions in communities according to the characteristics of the studied networks.

3.1.2.3 Algorithms based on spectral analysis

The spectral methods consist in embedding graph in an Euclidean space such that the nodes strongly connected are represented in the same part of space and the nodes without or with few connections are represented remotely.

3.1.2.3.1 Spectral clustering

Donetti et al. [13] proposed an approach based on the spectral properties of the Laplacien matrix of the graph. The coordinate i and j of the eigenvectors corresponding to the smallest non null eigenvalues are correlated when nodes i and j are in the same community. An Euclidean distance or an angular distance between nodes is computed starting from these eigenvectors, then this

distance being used in a hierarchical clustering algorithm. The number of eigenvectors that have to be taken into account is a priori not known. At every step of the clustering process the modularity is computed. Once the whole dendrogram is completed, the splitting with the maximum modularity is chosen as the output for the corresponding number of eigenvectors. An improvement of this approach has been proposed by using of a normalized Laplacien matrix version [14]. In [29], Jiang et al have reformulated the modularity and employed the spectral clustering in order to maximize the modularity and in consequence to correctly detect the community structure in the network. These spectral methods generate good results but the eigenvectors calculations are time-consuming which give a limited performance.

3.1.2.4 Algorithms based on global, local, or clustering proprieties

This section describes the algorithms that deal with some proprieties and adopt an agglomerative framework. Methods that use local proprieties are based on analysis of local connection patterns taking into consideration the internal connectivity within the clusters. The other methods that use global proprieties look not only for the inter-connectivity between clusters but also internal connectivity within the clusters and capture large-scale network structure.

3.1.2.4.1 Algorithm based on clustering proprieties

Eckmann et al. [16] defined a method based on the clustering proprieties. The key idea of the algorithm is that the high curvature region of a network will belong to the same community. In the first phase the method regroups nodes which are directly related and illuminated the connective links that arise from physical proximity in order to leaving the more relevant remote links. If two nodes A and B are congruent and B and C are congruent, then A and C will be congruent with high probability, forming a triangle. To quantify the aggregation of triangles with congruent edges, the method defines a local curvature at a node n . This curvature defines a World Wide Web landscape whose connected regions of high curvature characterize a common topic.

3.1.2.4.2 Algorithm based on local proprieties

Bagrow et al [5] have presented detection communities method which uses local information. They introduced the concept of shell that is defined as a set of nodes of geodesic distance from a starting node. The first shell includes the closest neighbors of the starting node and the second one includes the next neighbors of the closest neighbors of this starting node and so on until the l shell. Initially the l shell $l = 0$ starts from a node j which is then added to the list of community members. Repetitively, the algorithm extends the number of shell by adding nodes which are from l hop from node j in the list of community members. When the constraint of threshold is reached or the total connected component is obtained the process is stopped. The authors defined an adhesion matrix which gathers the vectors representing the communities of each starting node in order to obtain an idea about the total structure of the network. Then, according to the distance between these vectors, a process is carried out in order to permute the lines which have a short distance between them in order to gather the sub communities belonging to the same community. However, the founded communities depends closely on the localization of the starting node in particular if it is close from nodes that do not belong to its community. The algorithm generates a high cost in $O(n^3)$.

3.1.2.4.3 Algorithm based on global proprieties

An iterative community detection algorithm based on a measure of information discrepancy (MID) was proposed by Zhang et al. [62]. The authors defined the profile of any node based on the shortest path (SP) between it and all the other nodes in the network; if the profiles of a set of

nodes are similar, then they are possible in the same community, because the profile of each node characterizes its overall connection information in the network. The principle idea of the algorithm is that if two nodes i and j have similar SP profiles, they must have a very close link relationship. In this algorithm, the nodes which have a larger degree and have a proportion less than some predetermined value C to all the nodes in the network are called hubs. If node in a community structure is assembled around some hubs it is called the hub community, otherwise, it is called the non-hub community.

2.2. Divisive methods

The divisive methods split the graph into several communities by removing gradually the edges which connect two distinct communities. Thus, the network is divided into several components representing communities. This process of suppression of edges can be stopped at any stage according to constraint used in each divisive method. In this section, we go on to describe the various divisive methods using the main classification proposed in section (3).

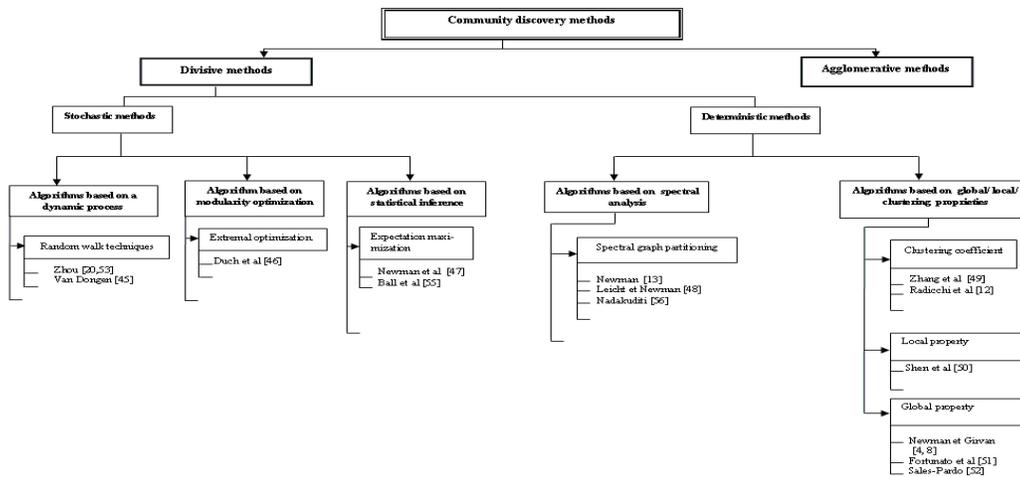


Figure 3. Community discovery methods taxonomy: Divisive methods

3.2.1 Stochastic Methods

Several community detection approaches have been developed dealing with stochastic methods and perform a divisive process in order to well reveal communities.

3.2.1.1 Algorithms Based on a Dynamic Process

3.2.1.1.1 Random Walk Techniques

Zhou et al. [64] [65] introduced the concept of network random walking and distance measure (as described in section 3.1.1.1.1). The Authors had proposed a divisive algorithm which based on dissimilarity index between nearest neighbors of a network. Initially, the approach [65] considers the whole network as just a single community. For each community, a threshold parameter is introduced and takes the initial value θ_{upp} (the upper dissimilarity threshold) of that community.

Repetitively, the algorithm carries out the necessary changes of the dissimilarity threshold and examines all the edges in the community to see whether two nearest neighbors are friends; if

$\Lambda(i, j) \leq \theta$ (see eq. 4) nodes i and j are marked as friends. Different friends' sets are then formed. A node which does not have any friends is moved to the friends set that has the strongest interaction with it. In each phase, the nodes of the community are distributed into a number of disjointed communities. After all the communities are processed, the dendrogram is drawn to show the relationship between the various communities as well as the upper and lower dissimilarity thresholds of each community.

3.2.1.2 Algorithms based on modularity optimization

The search for the optimal modularity value is an NP-hard problem that means [8] that the space of possible partitions grows faster than any power of the system size.

3.2.1.2.1 Extremal optimization

Duch et al. [46] proposed a heuristic search method to restrict the search space and find the optimal modularity value. They consider that the total modularity Q is the sum of the local modularity q_i on each node such as q_i is normalized in the interval $[-1,1]$ and defined

$$\lambda_i = \frac{q_i}{k_i} = \frac{k_r(i)}{k_i} - a_r(i) \quad (8)$$

Where λ_i : fitness of node is i , k_i : is the degree of node i and $k_r(i)$ is the number of edges between node i and the node which belongs to the same community r . Initially, the nodes of the network are divided into two random partitions containing the same number of nodes. In each iteration, the system self-organizes by moving the node that has the lower fitness to another partition and the fitness must be recalculated. Therefore the edges between the partitions are removed and we proceed recursively with each new resultant component. The process is repeated until the modularity Q cannot be improved. The algorithm gives an optimal modularity Q which is better than several existing algorithms of this kind and runs in $O(n^2 \log n)$ but the final partition depends closely on the initialization phase of the random partition in the network.

3.2.1.3 Algorithms based on statistic inference

3.2.1.3.1 Expectation maximization

Newman et al [47] have used the probabilistic mixture models and the expectation maximization algorithm to inferring module assignments and to identifying the optimal number of modules in a complex network by dividing its nodes into classes such that the members of each class have similar patterns of connection to other nodes. However, expectation maximization algorithms are known to converge to local maxima of the likelihood but not always to global maxima, and hence it is possible to get different solutions from different starting nodes. The used method almost converges to the same solution or to another similar one, whereas for others it is necessary to perform several runs with different initial conditions to find a good maximum of the likelihood.

Ball et al. [55] have proposed a probabilistic model of link communities for detecting overlapping and non overlapping communities. The type of links is a main factor in determining all communities, i.e., it is the edges that are partitioned, where the expected number of edges of color

z that lie between nodes i and j is $\theta_{iz}\theta_{jz}$ (or $\frac{1}{2\theta_{iz}\theta_{jz}}$ in case of self-edges). This approach defines a generative model and uses an expectation-Maximization (EM) algorithm to find maximum log likelihood, it solves the following equations

$$\theta_{iz} = \frac{\sum_j A_{ij} q_{ij}(z)}{\sqrt{\sum_{ij} A_{ij} q_{ij}(z)}} \quad (9)$$

The value of $q_{ij}(z)$ in Eq. (10) is the probability that an edge between i and j has color z , which is precisely the quantity needed in order to infer link communities in the network.

$$q_{ij}(z) = \frac{\theta_{iz} \theta_{jz}}{\sum_z \theta_{iz} \theta_{jz}} \quad (10)$$

The algorithm start with random initial conditions and iterate until $\theta_{iz}^{(i)}$ converge. The approach can be applying to networks of millions of nodes and gives results competitive with several existing algorithms. Nonetheless, it doesn't precise a criterion for determine the parameter k that represent the number of communities in the network.

3.2.2 Deterministic Methods

3.2.2.1 Algorithms Based On Spectral Analysis

3.2.2.1.1 Spectral Graph Partitioning

Newman and Leicht [48] have proposed an extension of the spectral optimization method and have adapted the modularity [13] to the oriented complex network. To rewrite the function of modularity for oriented networks, Newman and Leicht [48] proceeded as follows: Consider two nodes, i and j . Node i has high out-degree but low in-degree while node j has the reverse situation. This means that a given edge is more likely to run from i to j than vice versa. Then the probability that a link from a node i is directed to i is: $k_i^{in} k_j^{out} / m$ where k_i^{in} and k_j^{out} are the in- and out-degrees of the nodes. This proposal defines the equivalent of Eq.3 as follows

$$Q = \frac{1}{m} \sum_{ij} \left[A_{ij} - \frac{k_i^{in} k_j^{out}}{m} \right] \delta_{c_i, c_j} \quad (11)$$

The elements of modularity matrix are defined as

$$B_{ij} = A_{ij} - \frac{k_i^{in} k_j^{out}}{m} \quad (12)$$

To obtain the symmetric modularity matrix, Q is written as

$$Q = \frac{1}{4m} s^T (B + B^T) s = \beta_i (v_i^T s)^2 \quad (13)$$

Where β_i is the eigenvalues of $(B + B^T)$ corresponding to eigenvector v_i , and s_i is +1 if node i is assigned to community 1 and -1 if it is assigned to community 2. Then to divide the network into two communities, we calculate the eigenvector corresponding to the largest positive eigenvalue of the symmetric modularity matrix $(B + B^T)$ and the communities are identified based on the signs of the elements of the eigenvector. Thus to discover communities, a generalization of the modularity matrix is

$$B_{ij}^{(g)} = B_{ij} - \delta_{ij} \sum_{k \in g} B_{ik} \quad (14)$$

Where $B^{(g)}$ is the modularity matrix of the subgraph. These spectral optimization methods extract oriented information of links which result in identifying a significant community structure in time $O(n^2 \log n)$.

In [56], the authors focus on the spectral properties of the adjacency and modularity matrices using random matrix methods. This approach is built based on the stochastic block model and find the spectrum of eigenvalues of the modularity matrix. The spectrum illustrate the presence of a sharp transition between a regime in which there is a structure of community and a regime in which there is none. The method is a poor method of community detection of real world networks but it is an optimal method in the sense that no method can detect communities in the regime where the modularity method fails [57].

3.2.2.2 Algorithms Based on Global, Local, or Clustering Proprieties

3.2.2.2.1 Algorithm Based on Local Proprieties

Shen et al. [50] have proposed a filtration recursive method using a random model for networks in order to simultaneously carry out the suppression of several edges in each operation of filtration. To quantify the quality of division, the algorithm apply a recursive community coefficient (CRC) such as if the CRC of under network is smaller than that of its father network, then it consider under network as an indivisible built local community. The method of Shen et al. [50] offers complexity gain $O(m^2 + (c+1)m)$, for a network of m edges and c communities. Moreover, the method can detect the local communities according to densities' of their external edges in increase order in particular in the wide-area networks. Nevertheless, this method becomes slow and vague when the density of edges between the communities is proximate to the density of the edges within the communities.

The approaches proposed by Newman et al. [4] [8] are inspired from Freeman works [58]. The intuitive design of a central point in the communication, which based on the structural property betweenness, allows defining this point as being connecting between other points along their shortest paths of communication [58]. Then, the betweenness centrality of a node i is the number of the set of all geodesic paths that pass through i [9]. Newman et al. [8] defined three measures: shortest-path betweenness, current-flow betweenness and random walk betweenness. The algorithm can calculate the shortest paths between a particular pair of nodes using the breadth-first search in time $O(m^2)$ [59], [60]. Newman has proposed [61] a powerful algorithm which finds all edge betweenness in time $O(mn)$. After that, the detection of the communities is carried out by the removal of the edges with largest betweenness. The algorithm produces high quality partitions for networks which have small size. However, at each step, when removing an edge, both version of the shortest path betweenness algorithm and random walk betweenness algorithm update all computations which is very expensive in computation and need to be carried out in $O(n^3)$. The current-flow betweenness algorithm runs in $O(n^4)$ on a sparse graph which makes it not convenient for larger network.

Sales-Pardo et al [52] have observed that the hierarchical structure gives a very significant knowledge of the dynamics of several complex networks such biological networks [62] [63], determines the organization of complex systems, and extracts the relevant information at each level. The authors of this paper employ the proximity concept in the hierarchy between all pairs

of nodes, this measure, which called node affinity, based on Newman-Girvan modularity. First of all, the algorithm performs the search of partitions that are local maxima in the modularity landscape, next it find the affinity matrix using those partitions and their basin of attraction [64], and ,at last, the hierarchical tree obtained by using box clustering method. This method reveals the hierarchical organization of the network by the nested-box pattern along the diagonal of affinity matrix which is block-diagonal. The algorithm gives meaningful partitions for some social, technological and biological networks but it is quite slow.

Fortunato et al [51] have suggested a divisive algorithm that uses a centrality measure [65] which is based on the concept of efficient propagation of information over the network. The efficiency in the communication between two nodes i and j is equal to the inverse of the shortest path length, and the average efficiency of the graph G is defined as the average of the individual efficiencies over all $n(n-1)$ ordered pairs of distinct nodes. Information centrality is the relative drop in the network efficiency cause by removal of the edge from the graph G . The algorithm finds and removes iteratively the edge with the highest information centrality. The major drawback of this algorithm is the computational cost. Its running time is $O(n^4)$. In [66], Fortunato discussed in some detail the different centrality measure proposed in literature. Partitions obtained with these techniques are consistent, mainly because information centrality has a strong correlation with edge betweenness.

3.2.2.2.3 Algorithm based on clustering coefficient proprieties

Watts and Strogatz [67] have proposed a simple model which represents several proprieties in social networks such as the clustering coefficient and the degree distribution. The clustering coefficient quantifies how well connected are the neighbors of a node in a network. Radicchi et al [12] have proposed a divisive algorithm for detecting community using the concept of edge clustering coefficient. They define the edge clustering coefficient in analogy with the node clustering coefficient, it explain the fraction between the number of triangles to which a given edge belongs and the number of triangles that might potentially include it. Formally, for the edge-connecting node i to node j , the edge-clustering coefficient is defined as

$$C_{i,j}^{(3)} = \frac{z_{i,j}^{(3)} + 1}{\min[(k_i - 1), (k_j - 1)]} \quad (15)$$

Where $z_{i,j}^{(3)}$ the number of triangles is built on the edge (i, j) and $\min[(k_i - 1), (k_j - 1)]$ is the maximal possible number of these triangles. To address the problem of lacking of triangles in the network structure, Radicchi et al [12] have defined cycles of a higher order g as

$$C_{i,j}^{(g)} = \frac{z_{i,j}^{(g)} + 1}{s_{i,j}^{(g)}} \quad (16)$$

Where $z_{i,j}^{(g)}$ is the number of cyclic structures of order g that contain the edge (i, j) and $s_{i,j}^{(g)}$ is the number of all possible cyclic structures of order g . The algorithm removes at each step the smallest edge clustering coefficient and each removal operation requires only a local update of clustering coefficients, so the algorithm is much faster than several algorithms.

Lind et al [68] studied the clustering coefficient of bipartite networks in which there are no cycles of size three, and therefore, the standard definition of clustering coefficient given in [69] cannot be used. Thus, the coefficient is defined as the number of existing square over the total number of all possible squares. Zhang et al. [49] have proposed a community detection algorithm for the

bipartite networks whose principle idea is to remove at each step the edge which have smallest value of edge clustering coefficient. The authors define the coefficient LC_4 and LC_3 . The edge clustering coefficient LC_4 is defined as

$$LC_{4,iX} = \frac{q_{iX}}{(k_i - 1)(k_X - 1) + k_i^{(2)} + k_X^{(2)}} \quad (17)$$

Where q_{iX} is the number of squares to which a given edge l_{iX} belongs. k_i is the degree of node i and $k_i^{(2)}$ is the number of the second neighbours of node i except the nodes which are the first neighbors of node X . In bipartite networks, triples are the basic unit which includes two nodes from the same set. It is the basic unit which gives the relationship of two nodes in the same set. The clustering coefficient LC_3 of edge l_{iX} is the average of edges similarity of all the triples to which edge l_{iX} belongs. Here node i and X belong to different sets. It is given as

$$LC_{3,iX} = \frac{1}{k_i + k_X - 2} \left(\sum_{m=2}^{k_X} \frac{t_{mi}}{k_m + k_i - t_{mi}} + \sum_{N=2}^{k_i} \frac{t_{NX}}{K_N + k_X - t_{NX}} \right) \quad (18)$$

Where nodes m and i are from the same set and i and x are not in the same set. t_{mi} is the number of triples which contain node m and i . t_{NX} is the number of triples which contain node N and x . The divisive algorithm of bipartite structure produced a community structure more significant than that produced by applying existing algorithms for non-bipartite networks. However, the number of detected communities has to be determined in advance and there is not a stopping criterion during the partition in communities.

4. CONCLUSIONS

The discovery of a significant structure of communities is an important mechanism for understanding the structures and functions of complex networks. That's why several meaningful methods have been proposed in the literature. In this article, we have presented a survey on community discovery methods. We have described some of the essential definitions and techniques of community identification and have proposed taxonomy of them. For future work we should applied community detection methods on different type of complex networks according to their classification in the proposed taxonomy in order to conclude which class of methods is most appropriate for each type of complex networks.

REFERENCES

1. G.W.Flake, S.Lawrence, C.L.Giles, and F.M.Coetzee. Self-organization and identification of web communities. *Computer*, 35(3):66-70, 2002.
2. R.Guimera and L.A.N.Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895-900, 2005.
3. G.Palla, I.Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Arxiv preprint physics/0506133*, 435:814{818, 2005.
4. M.Girvan and M.E.J.Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821-7826, 2002.
5. D.Lusseau and M.E.J. Newman. Identifying the role that animals play in their social networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(Suppl 6):S477-S481, 2004.
6. S.L.Pimm. The structure of food webs. *Theoretical Population Biology*, 16(2):144-158, 1979.
7. AE Krause, KA Frank, DM Mason, RE Ulanowicz, and WW Taylor. Compartments exposed in food-web structure. *Nature*, 426:282-285, 2003.

8. M.E.J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.
9. Mark Newman. *Networks: An Introduction*. OUP Oxford, March 2010.
10. M.E.J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6):066133, 2004.
11. Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, November 1994.
12. Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2658-2663, March 2004.
13. M.E.J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577-8582, June 2006.
14. M.E.J. Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, 2003.
15. Santo Fortunato and Marc Barthlemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36-41, February 2007.
16. Lin S Kernighan BW. An efficient heuristic procedure for partitioning graphs. *The Bell System Technical Journal*, 49:291-307, 1970.
17. A.Pothen, H.D.Simon, K.P.Liou, et al. Partitioning sparse matrices with eigenvectors of graphs. *SIAM J. MATRIX ANAL. APPLIC*, 11(3):430-452, 1990.
18. M.Fiedler. Algebraic connectivity of graphs. *zechoslovak Mathematical Journal*, 23:298-305, 1973.
19. A.K.Jain and C. RICHARD. *Algorithms for clustering Data*. Prentice Hall Advance Reference Series, 1988.
20. Haijun Zhou. Network landscape from a brownian particles perspective. *Physical Review E*, 67(4):041908, April 2003.
21. Latapy M. Pons P. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10(2):191-218, 2006.
22. Lipowsky R. Zhou H. Network brownian motion: A new method to measure vertex-vertex proximity and to identify communities and subcommunities. *International Conference on Computational Science*, pages 1062-1069, 2004.
23. R.Guimera, M. Sales-Pardo, and L. A. N. Amaral. Modularity from uctuations in random graphs and complex networks. *Physical Review E*, 70(2):025101, 2004.
24. J.Reichardt and S. Bornholdt. Detecting fuzzy community structures in complex networks with a potts model. *Physical Review Letters*, 93(21):218701, 2004.
25. Jrg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Physical Review E*, 74(1):016110, July 2006.
26. Alex Arenas, Albert Daz-Guilera, and Conrad J. Prez-Vicente. Synchronization reveals topological scales in complex networks. *Physical Review Letters*, 96(11):114102, March 2006.
27. S.Boccaletti, M. Ivanchenko, V. Latora, A. Pluchino, and A. Rapisarda. Detecting complex network modularity by dynamical clustering. *Physical Review E*, 75(4):045102, April 2007.
28. Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical Review E*, 70(6):066111, December 2004.
29. Huawei Shen, Xueqi Cheng, Kai Cai, and Mao-Bin Hu. Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications*, 388(8):1706-1712, April 2009.
30. G.Palla, I.J.Farkas, P. Pollner, I. Derenyi, and T. Vicsek. Directed network modules. *New Journal of Physics*, 9:186, 2007.
31. I.Farkas, D.bel, G. Palla, and T. Vicsek. Weighted network modules. *New Journal of Physics*, 9:180, 2007.
32. Sune Lehmann, Martin Schwartz, and Lars Kai Hansen. Biclique communities. *Physical Review E*, 78(1):016108, July 2008.
33. L.Donetti and M.A.Munoz. Detecting network communities: a new systematic and efficient algorithm. *Journal of Statistical Mechanics: Theory and Experiment*, 2004:P10012, 2004.
34. L.Donetti and M. A. Munoz. Improved spectral algorithm for the detection of network communities. *arXiv:physics/0504059*, pages 104-107, April 2005.
35. J.Q.Jiang, A. W. M. Dress, and G. Yang. A spectral clustering-based framework for detecting community structures in complex networks. *Applied Mathematics Letters*, 2(9):14791482, 2009.
36. Jean-Pierre Eckmann and Elisha Moses. Curvature of co-links uncovers hidden thematic layers in the world wide web. *Proceedings of the National Academy of Sciences*, 99(9):5825-5829, April 2002.

37. J.P.Bagrow and E.M.Bollt. Local method for detecting communities. *Physical Review E*, 72(4):046108, 2005.
38. Junhua Zhang, Shihua Zhang, and Xiang-Sun Zhang. Detecting community structure in complex networks based on a measure of information discrepancy. *Physica A: Statistical Mechanics and its Applications*, 387(7):1675-1682, March 2008.
39. P.W. Kasteleyn and C.M. Fortuin. *J. Phys. Soc. Jap. Suppl.* 26, 11 (1969); C.M. Fortuin and P.W. Kasteleyn, *Physica*, 57:536, 1972.
40. Y.Fu and P. W. Anderson. Application of statistical mechanics to NP-complete problems in combinatorial optimisation. *Journal of Physics A: Mathematical and General*, 19:1605, 1986.
41. Jr.S.Kirkpatrick, C. D. Gillat and M. P.Vecchi. *Science* 220, 671, 1983.
42. Imre Deryni, Gergely Palla, and Tams Vicsek. Clique percolation in random networks. *Physical Review Letters*, 94(16):160202, April 2005.
43. C.Bron and J.Kerbosch. Finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575-577, 1973.
44. A Arenas, A Fernandez, S Fortunato, and S Gmez. Motif-based communities in complex networks. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224001, June 2008.
45. S.M.van Dongen. Graph clustering by flow simulation. university of utrecht, netherlands. 2000.
46. Jordi Duch and Alex Arenas. Community detection in complex networks using extremal optimization. *Physical Review E*, 72(2), August 2005.
47. M.E.J. Newman and E. A. Leicht. Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences*, 104(23):9564-9569, May 2007.
48. E.A.Leicht and M.E.J.Newman. Community structure in directed networks. *Physical Review Letters*, 100(11):118703, March 2008.
49. Peng Zhang, Jinliang Wang, Xiaojia Li, Menghui Li, Zengru Di, and Ying Fan. Clustering coefficient and community structure of bipartite networks. *Physica A: Statistical Mechanics and its Applications*, 387(27):6869-6875, December 2008.
50. Wenjiang Pei Yi Shen. Recursive _literation method for detecting community structure in networks. *Physica A: Statistical Mechanics and its Applications*, (26):6663-6670.
51. Santo Fortunato, Vito Latora, and Massimo Marchiori. A method to find community structures based on information centrality. *arXiv:cond-mat/0402522*, February 2004.
52. Marta Sales-Pardo, Roger Guimer, Andr A. Moreira, and Lus A. Nunes Amaral. Extracting the hierarchical organization of complex systems. *Proceedings of the National Academy of Sciences*, 104(39):15224-15229, September 2007.
53. H.Zhou. Distance, dissimilarity index, and network community structure. *Physical review e*, 67(6):061901, 2003.
54. U.Brandes, D.Delling, M. Gaertler, R.Goerke, M.Hoefer, Z.Nikoloski, and D.Wagner. Maximizing modularity is hard. *arXiv:physics/0608255*, August 2006.
55. Brian Ball, Brian Karrer, and M. E.J.Newman. An efficient and principled method for detecting communities in networks. *arXiv:1104.3590*, April 2011. *Phys. Rev. E* 84, 036103 (2011).
56. Raj Rao Nadakuditi and M.E.J.Newman. Graph spectra and the detectability of community structure in networks. *Physical Review Letters*, 108(18):188701, May 2012.
57. Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborov. Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107(6):065701, August 2011.
58. Linton Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35-41, March 1977.
59. Ravindra K.Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1 edition, February 1993. Published: Hardcover.
60. R.L.Rivest T.H.Cormen, C.E.Leiserson and C.Stein. *Introduction to Algorithms*. MIT Press, Cambridge, 2nd ed. edition, 2001.
61. M E Newman. Scientific collaboration networks. II. shortest paths, weighted networks, and centrality. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 64(1 Pt 2):016132, July 2001. PMID: 11461356.
62. Pennisi.E (2005) *Sciences*, page 309:94.
63. Kashtan N Milo R Itzkovitz M Alon U Itzkovitz S, Levitt R. Coarse-graining and self-dissimilarity of complex networks. *Phys Rev, E* 71:016127, 2005.
64. Weber TA Stillinger FH. *Phys Rev A*, 25:978-989, 1982.

65. Vito Latora and Massimo Marchiori. A measure of centrality based on the network efficiency. arXiv:cond-mat/0402050, February 2004.
66. Santo Fortunato, "Community Detection in Graphs," Physics Reports, vol. 486, Issues 3–5, pp. 75–174, February 2010.
67. Duncan Watts and Steven Strogatz. Collective dynamics of small-world networks. Nature, 393(6684):440{442, June 1998.
68. Pedro G. Lind, Marta C. Gonzalez, and Hans J. Herrmann. Cycles and clustering in bipartite networks. arXiv:cond-mat/0504241, April 2005.
69. R. Luce and Albert Perry. A method of matrix analysis of group structure. Psychometrika, 14(2):95-116, 1949.

Authors

Ahlem DRIF received an engineering degree in computer science from University of Sétif in Algeria (UFAS) in 2002 and magister degree in computer science in 2006. Currently, she is assistant professor and has 9 years teaching experience. She is a Ph.D. student and a member at the Laboratory of Network and Distributed System (LRSD) at the UFAS and has 10+ papers in international conferences. Her main research interests lie in the areas of social networks, community discovery methods, dynamic topology, ad hoc networks, power-aware protocols for wireless networks



Abdallah BOUKERRAM is professor of Computer Science at University of Bejaia in Algeria. He obtained the PhD degree in Computer Science from University of Louis Pasteur Strasbourg 1991 (France). He is director of Laboratory of Network and Distributed System (LRSD) at the UFAS. His current research domains are social networks, network security, distributed systems, and grid computing.

