

Chapter 2

Detecting the Overlapping and Hierarchical Community Structure in Networks

2.1 Introduction

As described in the previous chapter, community structure is a common and important topological characteristic of many real world complex networks. Examples include the World Wide Web, citations networks, various kinds of social and biological networks, and many others [1–3]. In the past decade, community structure has attracted much research attention from various scientific fields since it is crucial to understand the structural and functional properties of networks [4–6]. Many methods have been proposed to identify the community structure of complex networks [7–13]. The reader can refer to Ref. [14] for reviews.

These existing methods can be roughly classified into two categories in terms of the form of their results, i.e., to form a partition or a cover of the network. The first kind of methods produce a partition, i.e., each node belongs to one and only one community and is regarded as equally important. Different from classical graph partition problem, the number of communities and the size of each community are prior unknown. Among this kind of methods, the most successful ones are the methods based on the optimization of modularity [11, 15, 16], which is proposed by Newman et al. as a quality function to measure the goodness of a network partition [9]. A high value of modularity indicates a significant community structure. Generally, this kind of methods is suitable to understand the community structure of the whole networks, especially for the networks with small sizes. However, the modularity optimization methods also suffer several problems, e.g., the resolution limit problem [17, 18]. These problems pose concerns about the reliability of the community structure detected by directly optimizing the modularity.

The second kind of methods aim to discover the node sets i.e., *communities* with a high density of edges. In this case, overlapping is allowed, that is, some nodes may belong to more than one community. Meanwhile, some nodes may be neglected as subordinate nodes. Therefore, these methods result in an incomplete cover of the network. Compared to the partition methods, this kind of methods are appropriate to find the cohesive regions in the large scale networks. Ever since the problem of detecting overlapping community structure is proposed by Palla et al., many meth-

ods have been proposed [8, 19–26]. In [8], the community structure is uncovered by k -clique percolation and the overlaps between communities are guaranteed by the fact that one node can participate in more than one clique. However, the k -clique method gives rise to an incomplete cover of network, i.e., some nodes may not belong to any community. In addition, the hierarchical structure cannot be revealed for a given k . In [24], by introducing the concept of the belonging coefficients of each node to its communities, the authors proposed a general framework for extending the traditional modularity to quantify overlapping community structure. The method provides a new idea to find overlapping community structure. However, the physical meaning of the belonging coefficient lacks a clear explanation. Furthermore, the framework is hard to extend to large scale networks since it is difficult to find an efficient algorithm to search the huge solution space. Recently, Evans et al. [25] proposed a method to identify the overlapping community structure by partitioning a line graph constructed from the original network. This method only allows the communities to overlap at nodes. More important, there is no commonly accepted standard to evaluate the goodness of a cover up to now.

In real networks, communities are usually overlapping and hierarchical [8, 26–29]. Overlapping means that some nodes may belong to more than one community. Hierarchical means that communities may be further divided into sub-communities. The two kinds of existing methods, as mentioned above, investigate these two phenomena separately. The first kind of methods can be used to explore the hierarchical community structure. However, they are unable to deal with overlaps between communities. The second kind of methods can uncover overlapping community structure of networks, but they are incapable of finding the hierarchy of communities. Recently, Lancichinetti et al. make a pioneering attempt on the detection of both overlapping and hierarchical community structure in complex networks [26]. They try to detect the overlapped communities in the network based on the local optimization of a fitness function. Their method can uncover the hierarchical relation between these overlapped communities around a particular node. The remained problem lies in that the detection of the hierarchy of all overlapped communities in the network is not guaranteed due to the random choice of seed nodes.

In this chapter, we focus on the problem of detecting the overlapping and hierarchical community structure simultaneously. By taking maximal cliques as basic building blocks of communities, we propose an algorithm EAGLE (agglomerative hierarchical clusterinG based on maximalL cliquE) to detect community structure of networks. The overlaps among different maximal cliques guarantee the overlaps between communities and the hierarchy of these overlapped communities is uncovered by the process of agglomerative hierarchical clustering. Then, by extending the measurement for network partition, i.e., modularity, we propose a measurement Q_c to evaluate the quality of a cover of network. Then, we can find the overlapping community structure by directly optimizing the proposed measurement. Furthermore, we propose a method to construct a maximal clique network for a given network. With the maximal clique network at hand, finding the overlapping community structure by optimizing the new measurement Q_c on the original networks is equivalent to optimizing the standard modularity on the maximal clique network. In this way,

any method based on modularity optimization can be directly used to uncover the overlapping community structure.

2.2 EAGLE: Detecting the Overlapping and Hierarchical Community Structure

In this chapter, the algorithm EAGLE is presented to uncover both the overlapping and hierarchical community structure of networks. This algorithm deals with the set of maximal cliques and adopts an agglomerative framework. The effectiveness is then demonstrated by applications to two real world networks, namely the word association network and the scientific collaboration network.

2.2.1 The Algorithm

Before we introduce the details of the algorithm EAGLE, we use a schematic network to illustrate what EAGLE can do and compare it with the representative algorithms of the two kinds of existing methods introduced in the previous section. Figure 2.1(a) depicts the schematic network. This network is constructed according to the schematic network in [8], which has overlapping community structure. To construct the hierarchy of the overlapped communities, we remove the edge connecting nodes 9 and 13 and add two edges, one connecting 10 and 15 and the other one connecting 10 and 13. Figure 2.1(b) shows the community structure found by Newman's fast algorithm [11], which is the representative algorithm of partitioning network based on modularity optimization. Three communities are found when applying the algorithm to the schematic network. The hierarchy of communities can be revealed by applying the algorithm to each community further. For example, one of the three communities is divided into two sub-communities. However, overlaps between communities are not allowed. Figure 2.1(c) demonstrates the overlapping community structure found by the k -clique algorithm [8], which is the representative algorithm producing a cover of network. Unfortunately, this algorithm cannot reveal the hierarchy of community. Figure 2.1(d) shows the hierarchical and overlapping community structure found by the algorithm EAGLE. We can see that the algorithm EAGLE provides a possible way to investigate a more complete picture of the community structure.

Now we turn to the basic ideas behind the algorithm EAGLE. Generally speaking, a community can be regarded as a node set within which the nodes are more likely connected to each other than to the rest of the network. This indicates that a community usually has relatively high link-density. We know that the link-density of a clique is highest among all kinds of node subsets of a network. Furthermore, a densely-linked community usually contains a large clique, which could be regarded as the core of the community. Based on this observation, the algorithm EAGLE is proposed as an agglomerative hierarchical clustering algorithm to investigate

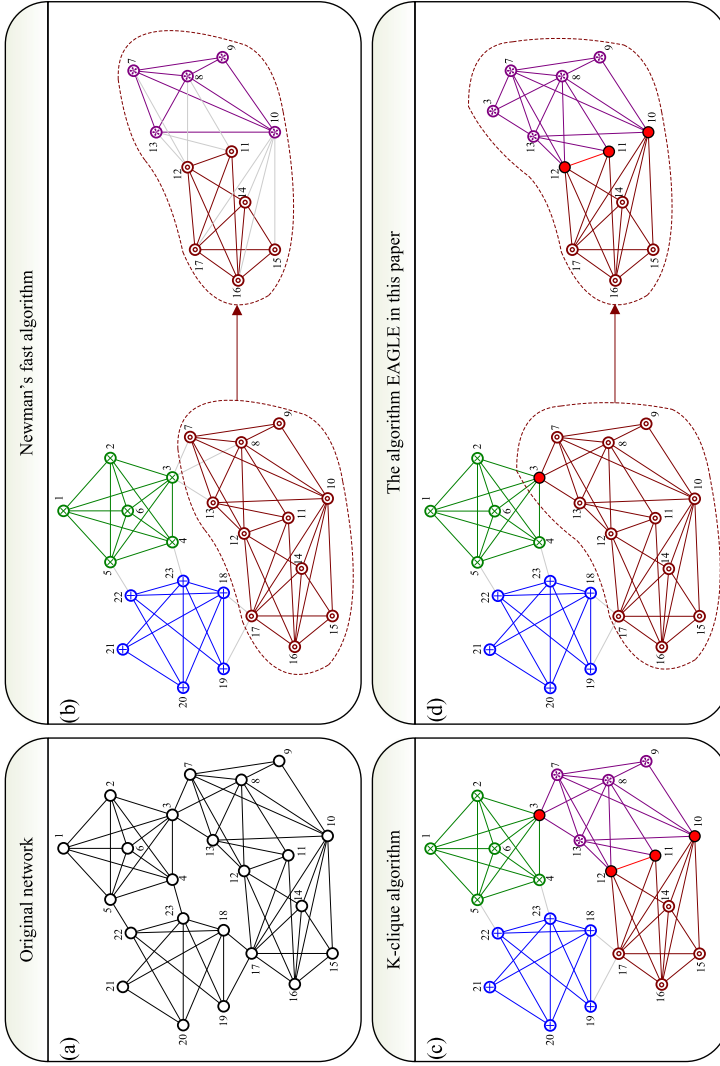


Fig. 2.1 Comparison of community structure found by different algorithms. Different communities are rendered in different colors (or markers for print). Edges between communities are colored in *light gray*. Overlapping regions between communities are emphasized in *red*. (a) The schematic network. (b) The hierarchical community structure found by Newman's fast algorithm. This algorithm is chosen as a representative of the first kind of algorithms. (c) The overlapping community structure found by the k -clique algorithm as a representative of the second kind of algorithms. (d) The hierarchical and overlapping community structure found by the algorithm EAGLE. Reprinted from Ref. [30], Copyright 2009, with permission from Elsevier

the community structure. Different from traditional agglomerative algorithms [11], the algorithm deals with the set of maximal cliques rather than the set of nodes.

A *maximal clique* is a clique which is not a subset of any other cliques. In the algorithm EAGLE, we need to first find out all the maximal cliques in the network. This can be done by many efficient parallel algorithms. Here we choose the well-known *Bron-Kerbosch* algorithm [31] for its efficiency and its simplicity in implementation. Note that not all maximal cliques are taken into account. The maximal cliques, whose nodes are from some other larger maximal cliques, are called *subordinate maximal cliques*. For example, in Fig. 2.1, nodes 4 and 23 form a subordinate maximal clique. Because node 4 is from another larger maximal clique $\{1, 2, 3, 4, 5, 6\}$ and node 23 is also from other larger maximal cliques, including $\{18, 20, 21, 23\}$, $\{18, 20, 22, 23\}$ and $\{18, 19, 22, 23\}$. Subordinate maximal cliques may mislead our algorithm and thus are discarded. Most subordinate maximal cliques have small sizes. Thus, we can discard them by setting a threshold k and neglecting all the maximal cliques with the size smaller than k . This simple tactic may also discard some non-subordinate maximal cliques. The higher the value of k is, the more non-subordinate maximal cliques are discarded by mistake. On the other hand, the smaller the value of k is, the more subordinate maximal cliques are remained. In real world networks, the threshold k typically takes value between 3 and 6. As to the network in Fig. 2.1, both 3 and 4 are appropriate threshold values. As to the networks used in Sect. 2.2.2, 4 is demonstrated to be an appropriate threshold [8]. After neglecting the maximal clique with the size smaller than the threshold k , some nodes do not belong to any remaining maximal cliques. We call these nodes as *subordinate nodes*.

The algorithm EAGLE has two stages. In the first stage, a dendrogram is generated. In the second stage, we choose an appropriate cut which breaks the dendrogram into communities. The first stage of the algorithm EAGLE can be described as follows:

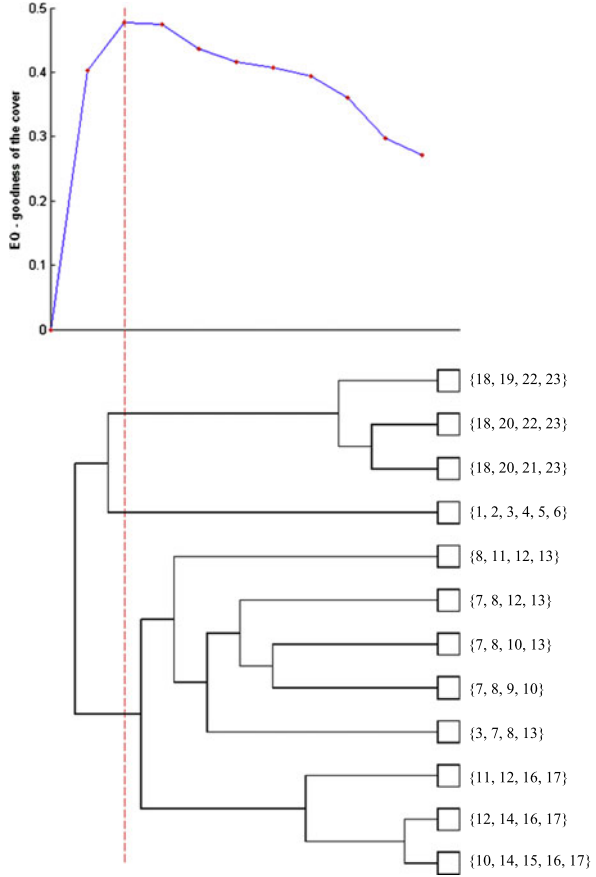
1. Find out all maximal cliques in the network. Neglect subordinate maximal cliques. The remainders are taken as the initial communities. Each subordinate node is also taken as an initial community comprising the sole node. Calculate the similarity between each pair of communities.
2. Select the pair of communities with the maximum similarity, incorporate them into a new one and calculate the similarity between the new community and other communities.
3. Repeat step 2 until only one community remains.

In the algorithm, the similarity S between two communities C_1 and C_2 is defined as

$$S = \frac{1}{2m} \sum_{v \in C_1, w \in C_2, v \neq w} \left[A_{vw} - \frac{k_v k_w}{2m} \right], \quad (2.1)$$

where A_{vw} is the element of adjacency matrix of the network (here, we only consider undirected, unweighted networks). It takes value 1 if there is an edge between node v and node w and 0 otherwise. The quantity $m = \frac{1}{2} \sum_{vw} A_{vw}$ is the total number of edges in the network and k_v is the degree of node v .

Fig. 2.2 Illustration for the process of the algorithm EAGLE. This illustration is according to the schematic network in Fig. 2.1. The *bottom part* is a dendrogram. The leaf nodes correspond to the non-subordinate maximal cliques. The label of each leaf node shows the nodes belonging to it. The (red) vertical dashed line is a cut through the dendrogram and it gives the best cover of the network. The *top part* of the figure is a graph which illustrates the curve of EQ corresponding to each cover of the network. The threshold k is set to be 4. Reprinted from Ref. [30], Copyright 2009, with permission from Elsevier



Similar to the fast algorithm in [11], the process of our algorithm corresponds to a dendrogram, which shows the order of the amalgamations of communities. Any cut through the dendrogram produces a cover of the network. As an illustration, Fig. 2.2 shows the dendrogram generated by our algorithm when applied to the network in Fig. 2.1.

The task of the second stage of the algorithm EAGLE is to cut the dendrogram. To determine the place of the cut, a measurement is required to judge the quality of a cover. In [24], an extension of modularity is proposed to evaluate the goodness of overlapped community decomposition. Here, for simplicity, we propose another extension of modularity, namely EQ . In Fig. 2.2, the cut gives the best cover with the maximum value of EQ . The extended modularity is defined as

$$EQ = \frac{1}{2m} \sum_i \sum_{v \in C_i, w \in C_i} \frac{1}{O_v O_w} \left[A_{vw} - \frac{k_v k_w}{2m} \right], \quad (2.2)$$

where O_v is the number of communities to which node v belongs.

Note that EQ reduces to Q in [9] when each node belongs to only one community (readers can refer to Ref. [15] for details), and EQ is equal to 0 when all nodes belong to the same community. In addition, it will be shown later in Sect. 2.2.2, a high value of EQ indicates a significant overlapping community structure.

Alike to modularity, the extended modularity suffers a resolution limit beyond which no modular structure can be detected even though these modules might have their own identities. For the algorithm EAGLE, however, these modules can be still detected by further applying the algorithm to each community found until none of them can be divided into smaller ones. Thus, we obtain a hierarchy of overlapping communities which reveals the community structure of network more completely.

Now we analyze the time complexity of the algorithm. Let n be the number of nodes, s be the number of maximal cliques in the initial state of the algorithm, and h be the number of pair of maximal cliques which are neighbors (connected by edges or overlap with each other). We firstly consider the first stage of the algorithm. In step 1, $O(n^2)$ operations are needed to calculate the similarity between each pair of initial communities. In step 2, we only consider the pairs of communities which are neighbors. Each selection costs h operations and each time of join costs $O(n)$ operations at most. Totally, we carry on a maximum of $s - 1$ join operations. Thus the first stage of the algorithm takes at most $O(n^2 + (h + n)s)$ operations. As to the second stage, we need to calculate the value of EQ corresponding to each cover. In our implementation, we calculate the value of EQ for the initial cover and update it after each join of two selected communities into a new one. Each time of update costs at most n^2 operations. Hence, the second state of the algorithm takes at most $O(n^2s)$ operations. In addition, we need to find out all the maximal cliques in the network. It is widely believed to be a non-polynomial problem. However, for real world networks, finding all the maximal cliques is easy due to the sparseness of these networks. Compared to the Newman's fast algorithm and the k -clique algorithm, the algorithm EAGLE is time-consuming. We leave it as a future work that how to improve the speed of EAGLE.

2.2.2 Applications

In this subsection, we apply the algorithm EAGLE to two real world complex networks, the word association network and the scientific collaboration network. The results show that EAGLE can discover new knowledge and insights underlying these networks.

The test data of the two networks are from the demo of the *CFinder* [32].¹ The two networks comprise 7207, 16662 nodes and 31784, 22446 edges, respectively. The average clustering coefficients [33] are approximately 0.15 and 0.19, which indicate that these networks have significant community structures in general.

¹CFinder is a free software for finding overlapping dense groups of nodes in networks, based on the clique percolation method. URL: www.cfinder.org.

The word association network is constructed from the South Florida Free Association norms list. The original network is directed and weighted. The weight of a directed link from one word to another indicates the frequency that the people in the survey associated the end point of the link with its start point. The directed links are replaced by undirected ones with a weight equal to the sum of the weights of the corresponding two oppositely directed links. Furthermore, the links with the weight less than 0.025 are deleted. The scientific collaboration network is from the co-authorship network of *Los Alamos e-print* archives. Each article in the archive between April 1998 and February 2004 contributes the value $1/(n-1)$ to the weight of the link between every pair of its n authors. The link with the weight less than 1.0 is omitted.

In the word association network, totally 17 communities are found by our algorithm—see Fig. 2.3(a), left panel. Among these communities, 63 of 136 possible pairs of communities overlap with each other. To investigate what is correlated to the community structure, we apply our algorithm to each of these communities again. The sub-community structure of one community is given in Fig. 2.3(a), middle panel. Each of these sub-communities has certain correlation with the semantic meaning of words. For example, most of the words in the community with size 112 are related to the family of animals in Africa. This community is explored further and four communities are found, shown in Fig. 2.3(a), right panel. Each community is associated with animals from the same family, namely rodentia, felidae & primates, cervidae & caprinae, and equidae respectively. The details of one community are also illustrated in Fig. 2.3(a), right panel. Two large communities correspond to words associated with animals from cervidae and caprinae respectively. The overlapped word *Animal* acts as a bridge between the two communities. Three small communities comprise peripheral words.

Applying our algorithm to the scientific collaboration network, we obtain totally 1754 communities—see Fig. 2.3(b), left panel, with the corresponding high value of $EQ \approx 0.86$. Three large communities contain 23.4 % of all the nodes, while the others are relatively small. The three large communities correspond closely to subject subareas: the biggest one mainly to *mes-hall* and *str-el*, the second biggest one to *str-el* and *supr-con*, and the other to *stat-mech*, *dis-nn* and *soft*. We further apply the algorithm to one community and it is broken down into 26 sub-communities—depicted in Fig. 2.3(b), middle panel. There appears to be a correlation between the sub-community structure and the regional divisions of the scientific researchers. For example, most of the members of the community with size 166 work in Europe. More specific regional information can be obtained when applying the algorithm to this community. The biggest one and its sub-community structure are given in Fig. 2.3(b), right panel. We can see that the author *G. Parisi* (who is well known for having made significant contributions in different fields of physics) acts as a hub in the community. Different communities can be associated with his different fields of interest.

Now, we compare the algorithm EAGLE with Newman’s fast algorithm and the k -clique algorithm by applying them to the scientific collaboration network. Figure 2.4 shows that the hierarchical community structure found by Newman’s fast

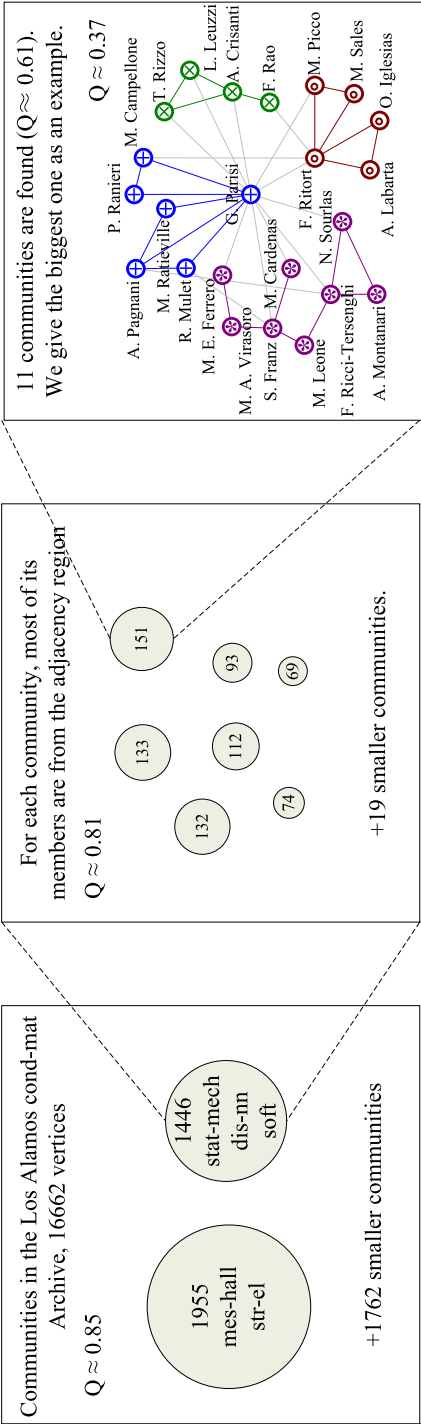
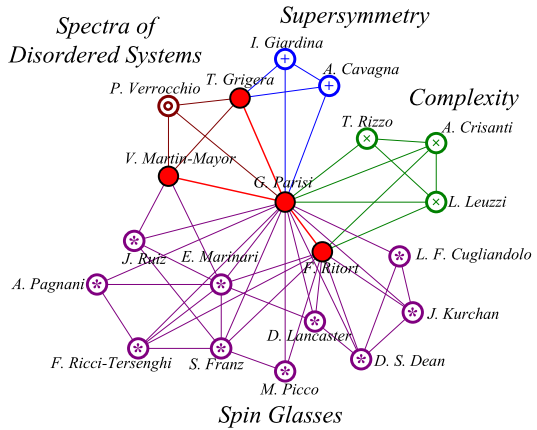


Fig. 2.4 The hierarchical community structure found by Newman's fast algorithm in the scientific collaboration network. Each *numbered circle* denotes a community and the *number in the circle* denotes its size. Communities connected by a link overlap with each other. Different communities are rendered in different colors (or markers for print). The overlapping nodes and edges between communities are colored in *red*. In addition, the values of the corresponding Q are also given when breaking networks (communities) down into communities (sub-communities). Reprinted from Ref. [30], Copyright 2009, with permission from Elsevier

Fig. 2.5 The overlapping community structure around the node *G. Parisi* in the scientific collaboration network. Different communities are rendered in different colors (or markers for print). The overlapping nodes and edges between communities are colored in red. Here, k is set to be 4. Reprinted from Ref. [30], Copyright 2009, with permission from Elsevier



algorithm. The number of communities at each level of the hierarchy is almost identical to that found by the algorithm EAGLE although the size of each community is somewhat different. Compare the left panel of Fig. 2.4 with that of Fig. 2.3(b), one community disappears. Actually, it is divided into several other smaller communities, which are not depicted. As to the right panels, the details of communities were given. The node *G. Parisi*, acting as a hub in Fig. 2.3, only appears in one community in Fig. 2.4. The reason is that Newman's algorithm gives rise to partitions of network, while the algorithm EAGLE allows overlaps between communities. Note that overlap between communities is a very common phenomenon in real networks and may contribute to the evolution of communities and the dynamics of networks.

Figure 2.5 shows the overlapping community structure around the node *G. Parisi* in the scientific collaboration network. Compare to Fig. 2.3, both the algorithm EAGLE and the k -clique algorithm can find the overlapping community structure, although the overlapped communities found by the two algorithms are somewhat different. However, the algorithm EAGLE can give the hierarchy of these overlapped communities compared to the k -clique algorithm. The hierarchy of communities is useful to understand the community structure of real world networks.

2.3 Extending Modularity to Quantify the Overlapping Community Structure

In this section, a measure for the quality of a cover is proposed to quantify the overlapping community structure referred as Q_c (quality of a cover). With the measure Q_c , the overlapping community structure can be identified by finding an optimal cover, i.e., the one with the maximum Q_c . The Q_c is based on a maximal clique view of the original network. A maximal clique is a clique (i.e. a complete sub-graph) which is not a subset of any other clique in a graph. The maximal clique view is according to a reasonable assumption that a maximal clique cannot be shared by

two communities due to that it is highly connective. To find an optimal cover, we construct a maximal clique network from the original network. We then prove that the optimization of Q_c on the original network is equivalent to the optimization of the modularity on the maximal clique network. Thus the overlapping community structure can be identified through partitioning the maximal clique network with an efficient modularity optimization algorithm, e.g., the fast unfolding algorithm in [34]. The effectiveness of the measure Q_c is demonstrated by extensive tests on both the artificial networks and the real world networks with known community structure and the application to the word association network.

2.3.1 Quantifying the Overlapping Community Structure

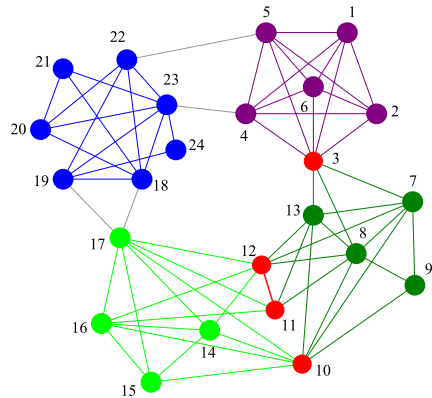
Before introducing how to quantify the overlapping community, we first illustrate the representation of overlapping community. Figure 2.6 shows an example network with overlapping community structure. The overlapping community structure can be represented by a cover of network. A cover of network is defined as a set of clusters such that each node is assigned to one or more clusters and no cluster is a proper subset of any other cluster. As to the network in Fig. 2.6, the overlapping community structure can be represented by the cover $\{\{1, 2, 3, 4, 5, 6\}, \{3, 7, 8, 9, 10, 11, 12, 13\}, \{10, 11, 12, 14, 15, 16, 17\}, \{18, 19, 20, 21, 22, 23, 24\}\}$.

We have known that the overlapping community structure can be represented as a cover of network instead of a partition of network. Therefore, the overlapping community structure can be quantified through a measure of a cover of network.

As well known, the modularity was used to measure the goodness of a partition of network. Given an un-weighted, undirected network $G(E, V)$ and a partition P of the network G , the modularity can be formalized as

$$Q = \frac{1}{L} \sum_{c \in P} \sum_{vw} \delta_{vc} \delta_{wc} \left(A_{vw} - \frac{k_v k_w}{L} \right), \quad (2.3)$$

Fig. 2.6 A schematic network with overlapping community structure. Communities are differentiated by colors and the overlapping regions are emphasized in red. The edges between communities are colored in gray. Reprinted from Ref. [35], Copyright 2009, with permission from IOP Publishing and SISSA



where A is the adjacency matrix of the network G , $L = \sum_{vw} A_{vw}$ is the total weight of all the edges, and $k_v = \sum_w A_{vw}$ is the degree of the node v .

In Eq. 2.3, δ_{vc} denotes whether the node v belongs to the community c . The value of δ_{vc} is 1 when the node v belongs to the community c and 0 otherwise. For a cover of network, however, a node may belong to more than one community. Thus δ_{vc} needs to be extended to a belonging coefficient α_{vc} , which reflects how much the node v belongs to the community c .

With the belonging coefficient α_{vc} , the goodness of a cover C can be measured by

$$Q_c = \frac{1}{L} \sum_{c \in C} \sum_{vw} \alpha_{vc} \alpha_{wc} \left(A_{vw} - \frac{k_v k_w}{L} \right). \quad (2.4)$$

The idea of the belonging coefficient was proposed in [24]. Its authors also pointed out that the belonging coefficient should satisfy a normalization property. This property is formally written as

$$0 \leq \alpha_{vc} \leq 1, \quad \forall v \in V, \forall c \in C \quad (2.5)$$

and

$$\sum_{c \in C} \alpha_{vc} = 1. \quad (2.6)$$

Equations 2.5 and 2.6 only give the general constraints on α_{vc} , which lead to such a huge solution space that the enumeration of all the solutions is impractical. To reduce the solution space and make the problem tractable, we introduce an additivity property for the belonging coefficient: the belonging coefficient of a node to a community c is the sum of the belonging coefficients of the node to all of c 's sub-communities.

For example, we assume that $C = \{c_1, c_2, \dots, c_{r-1}, c_r, \dots, c_s, c_{s+1}, \dots, c_n\}$ is a cover of the network G and $C' = \{c_1, c_2, \dots, c_{r-1}, c_u, c_{s+1}, \dots, c_n\}$ is another cover of G . The difference between C' and C is that the community c_u is the union of the communities $c_r, \dots, c_s$. The additivity property of belonging coefficient can then be formally denoted as

$$\alpha_{vc_u} = \sum_{i=r}^s \alpha_{vc_i}. \quad (2.7)$$

The belonging coefficient α_{vc} reflects how much a node v belongs to a community c . Intuitively, it is proportional to the total weight of the edges connecting the node v to the nodes in the community c , i.e.,

$$\alpha_{vc} \propto \sum_{w \in V(c)} A_{vw}, \quad (2.8)$$

where $V(c)$ denotes the set of nodes belonging to community c . Note that the additivity property of belonging coefficient requires that communities are disjoint from

a proper view of the network. Therefore, we introduce the maximal clique view to achieve this purpose. We define α_{vc} as the form

$$\alpha_{vc} = \frac{1}{\alpha_v} \sum_{w \in V(c)} \frac{O_{vw}^c}{O_{vw}} A_{vw}, \quad (2.9)$$

where O_{vw} denotes the number of maximal cliques containing the edge (v, w) in the whole network, O_{vw}^c denotes the number of maximal cliques containing the edge (v, w) in the community c , and α_v is a normalization term denoted as

$$\alpha_v = \sum_{c \in C} \sum_{w \in V(c)} \frac{O_{vw}^c}{O_{vw}} A_{vw}. \quad (2.10)$$

Obviously, the definition of α_{vc} in Eq. 2.9 satisfies the normalization property. It also satisfies the additivity property if we assume that each maximal clique only belongs to one community. This assumption is reasonable since a maximal clique is highly connective that any two communities sharing a maximal clique should be combined into a single one.

With Eqs. 2.4 and 2.9, we obtain the detailed form of Q_c as a measure to the quality of a cover of network. Note that when a cover degrades to a partition, Q_c becomes the modularity Q in [15] accordingly. In addition, $Q_c = 0$ when all nodes belong to the same community, and it will be shown later in Sect. 2.3.4 that a high value of Q_c indicates a significant overlapping community structure.

2.3.2 Identifying the Overlapping Community Structure

With the measure Q_c , the overlapping community structure of network can be identified by finding the optimal cover with maximum Q_c . To find the optimal cover, we construct a maximal clique network from the original network. Then the overlapping community structure can be identified through partitioning the maximal clique network.

2.3.2.1 Construction of the Maximal Clique Network

Given an un-weighted, undirected network G , a corresponding maximal clique network G' can be constructed through the following method.

The maximal clique network G' is constructed by defining its nodes and edges. We first find out all the maximal cliques in G . We can simply take all these maximal cliques as nodes of G' . In practice, however, we observe that some maximal cliques would not be so highly connective if their sizes are too small. Such a maximal clique either lies between different communities (e.g., the maximal cliques $\{4, 23\}$ and $\{5, 22\}$ in the network shown in Fig. 2.6) or connects a node to the whole network

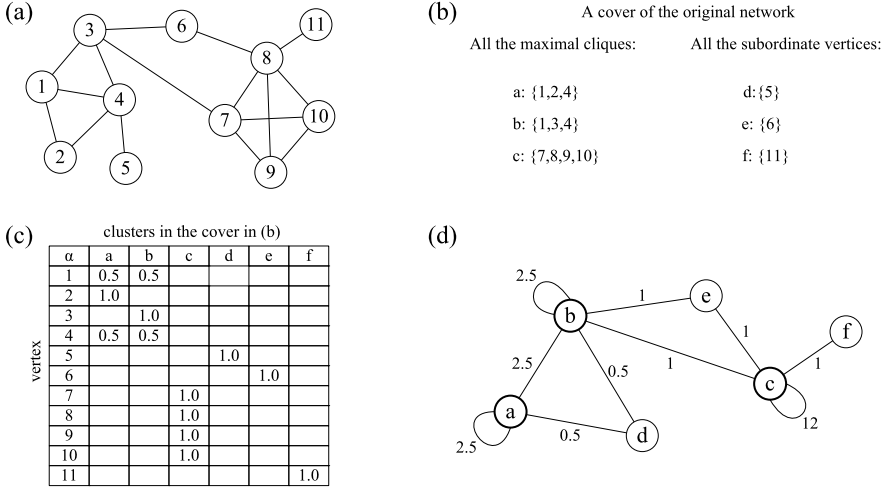


Fig. 2.7 Illustration for the construction process of the maximal clique network. Here, (a) The original example network. (b) A cover of the original network. In this cover, each maximal clique is a cluster and each subordinate node forms a cluster consisting of only one node. (c) The belonging coefficient of each node to its corresponding clusters in the cover. (d) The maximal clique network constructed from the example network. Here the parameter $k = 3$. Reprinted from Ref. [35], Copyright 2009, with permission from IOP Publishing and SISSA

(e.g., the maximal clique $\{8, 11\}$ in the network shown in Fig. 2.7(a)). To deal with these small maximal cliques, we introduce a threshold k . Specifically, given the parameter k , we only refer to those maximal cliques with the size no smaller than k as the maximal cliques, and refer to those with the size smaller than k as subordinate maximal cliques. We then denote the nodes only belonging to subordinate maximal cliques as subordinate nodes. In this way, each maximal clique or subordinate node in the original network G is taken as one node of G' .

Note that all the subordinate nodes and the maximal cliques form a cover C of the original network G . For a subordinate node v and a cluster c in the cover C , the value of α_{vc} is defined to be 1.0 when v belongs to the cluster c and 0.0 otherwise. As to other nodes, α_{vc} can be obtained according to Eq. 2.9.

Now we can define the edge of the maximal clique network G' by defining its adjacency matrix B . Let m_x denote the set of the original network's nodes corresponding to the x th node in G' . The element of B is defined as

$$B_{xy} = \sum_{vw} \alpha_{vm_x} \alpha_{wm_y} A_{vw} \quad (2.11)$$

and the strength (degree) of the x th node

$$s_x = \sum_y B_{xy} = \sum_v \alpha_{vm_x} k_v. \quad (2.12)$$

For clarity, Fig. 2.7 illustrates the construction process of the maximal clique network from an example network with the parameter $k = 3$. Figure 2.7(b) shows the subordinate nodes and the maximal cliques. Each of them becomes a node in the resulting maximal clique network. For example, the maximal clique $\{1, 2, 4\}$ corresponds to the node a and the subordinate node $\{5\}$ corresponds to the node d . Each of these maximal cliques or subordinate nodes is a cluster in a cover C of the original network. Their belonging coefficients corresponding to the cover C are shown in Fig. 2.7(c). According to these belonging coefficients and Eq. 2.11, the weight of each edge of the maximal clique network is obtained. Take the edge connecting the nodes a and b as an example. As known, the node a corresponds to the maximal clique $\{1, 2, 4\}$ and the node b corresponds to the maximal clique $\{1, 3, 4\}$. Using Eq. 2.11, the weight of this edge is $\alpha_{1a}\alpha_{3b} + \alpha_{1a}\alpha_{4b} + \alpha_{2a}\alpha_{1b} + \alpha_{2a}\alpha_{4b} + \alpha_{4a}\alpha_{1b} + \alpha_{4a}\alpha_{3b} = 0.5 + 0.25 + 0.5 + 0.5 + 0.25 + 0.5 = 2.5$.

The constructed maximal clique network is a weighted network though the original network is un-weighted. The total weight L' of all the edges in the maximal clique network is equal to the total weight (number) L of edges in the original network. The proof is

$$\begin{aligned}
 L' &= \sum_{xy} B_{xy} \\
 &= \sum_{xy} \sum_{vw} \alpha_{vm_x} \alpha_{wm_y} A_{vw} \\
 &= \sum_{vw} A_{vw} \sum_x \alpha_{vm_x} \sum_y \alpha_{wm_y} \\
 &= \sum_{vw} A_{vw} \\
 &= L.
 \end{aligned} \tag{2.13}$$

Each node in the original network corresponds to more than one node in the maximal clique network. For example, in Fig. 2.7, the node 1 corresponds to two nodes a and b in the maximal clique network. Thus, a partition of the maximal clique network can be mapped to a cover of the original network, which holds the information about the overlapping community structure of the original network.

2.3.2.2 Finding the Overlapping Community Structure

Now we investigate the overlapping community structure of the original network through partitioning its corresponding maximal clique network. To find the natural partition of a network, the optimization of modularity is the widely used technique. The partition with the maximum modularity is regarded as the optimal partition of network. We employ the algorithm proposed in [34] to partition our maximal clique network. As an example, Fig. 2.8 shows the partition of a maximal clique network. Different parts of the partition are differentiated by shapes or colors.

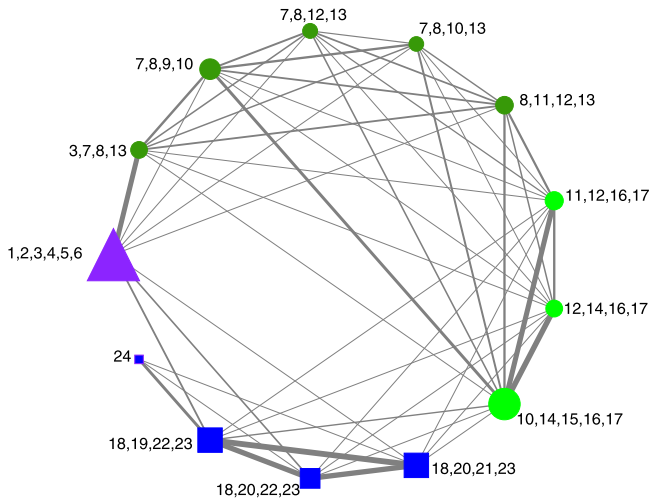


Fig. 2.8 The maximal clique network constructed from the schematic network in Fig. 2.6. The label near each node shows its corresponding nodes in the original network. The width of line indicates the weight of the corresponding edge. The self-loop edge of each node is omitted and its width is reflected by the volume of the associated circles, squares or triangles. In addition, the optimal partition of the maximal clique network is also depicted. The communities in this partition are differentiated by shapes. Furthermore, the circle-coded community can be partitioned into two sub-communities. The four communities are shown in different colors, which are identical to the communities depicted in Fig. 2.6. Here k is 4. Reprinted from Ref. [35], Copyright 2009, with permission from IOP Publishing and SISSA

As mentioned above, each partition of the maximal clique network corresponds to a cover of the original network and the cover tells us the overlapping community structure. The key problem lies in that whether the optimal partition of the maximal clique network corresponds to the optimal cover of the original network. To answer this question, we analyze the relation between the modularity of the maximal clique network and the Q_c of the original network.

Let $\mathcal{P} = \{p_1, p_2, \dots, p_l\}$ be a partition of the maximal clique network and $\mathcal{C} = \{c_1, c_2, \dots, c_l\}$ be the corresponding cover of the original network. Here, l is the size of $\mathcal{P}$ or $\mathcal{C}$, i.e., the number of communities. Using modularity, the quality of the partition $\mathcal{P}$ can be measured by

$$Q = \frac{1}{L'} \sum_i \sum_{x,y \in p_i} \left(B_{xy} - \frac{s_x s_y}{L'} \right). \quad (2.14)$$

Using Eqs. 2.11 and 2.12, we have

$$Q = \frac{1}{L'} \sum_i \sum_{x,y \in p_i} \left(\sum_{vw} \alpha_{vm_x} \alpha_{wm_y} A_{vw} - \frac{1}{L'} \sum_v \alpha_{vm_x} k_v \sum_w \alpha_{wm_y} k_w \right)$$

$$\begin{aligned}
&= \frac{1}{L'} \sum_i \sum_{x,y \in p_i} \sum_{vw} \alpha_{vm_x} \alpha_{wm_y} \left(A_{vw} - \frac{k_v k_w}{L'} \right) \\
&= \frac{1}{L} \sum_i \sum_{vw} \alpha_{vc_i} \alpha_{wc_i} \left(A_{vw} - \frac{k_v k_w}{L} \right) \\
&= Q_c.
\end{aligned} \tag{2.15}$$

Equation 2.15 tells us that the optimization of the Q_c on the original network is equivalent to the optimization of the modularity on the maximal clique network. Thus we can find the optimal cover of the original network by finding the optimal partition of the corresponding maximal clique network. The optimal cover reflects the overlapping community structure of the original network.

2.3.3 Discussions

As to our method, it is important to select an appropriate parameter k . On one hand, the parameter k affects the constituent of the overlapping regions between communities. According to the definition to subordinate nodes, they are excluded from the overlapping regions. Thus the larger the parameter k , the less the number of nodes which can occur in the overlapping regions. When $k \rightarrow \infty$, the maximal clique network is identical to the original network and no overlap is identified. On the other hand, since the subordinate maximal cliques are not so highly connective, the parameter k should not be too small in practice. The choice of the parameter k depends on the specific networks. Observed from many real world networks, the typical value of k is often between 3 and 6. Additionally, as to the networks where larger cliques are rare, our method is close to the traditional modularity-based partition methods. In this case, rare overlaps will be found.

Both the traditional modularity and the Q_c are based on the significance of link density in communities compared to a null-model reference network, e.g., the configuration model network. However, differently from the traditional modularity which requires that each node can only belong to one community, Q_c requires that each maximal clique can only belong to one community. In this way, Q_c takes advantage of both the local topological structure (i.e., the maximal clique) and the global statistical significance of link density.

The same to the traditional modularity, however, the measure Q_c also suffers the resolution limit problem [17], especially when applied to large scale complex networks. Recently, some methods [36] have been proposed to address the resolution limit problem of modularity. These methods are also appropriate to the measure Q_c .

Now we turn to the efficiency of our method. It is difficult to give an analytical form of the computational complexity of our method. Here we only discuss what influences the efficiency of our method. Our method consists of three stages, finding out the maximal cliques, constructing the maximal clique network and partitioning

the maximal clique network. As to the first stage, we need to find out all the maximal cliques in the network. It is widely believed to be a non-polynomial problem. However, for real world networks, finding all the maximal cliques is easy due to the sparseness of these networks. The computational complexity of the second stage depends on the number of edges in the original networks. Finally, the partition stage rests with the number of the maximal cliques and subordinate nodes. Taken together, our method is very efficient on real world networks.

In addition, as mentioned above, the overlapping community structure can be identified by the optimization of Q_c . Similarly, iteratively applying this method to each community, we can investigate the sub-communities correspondingly. In this way, a rigid hierarchical relation of overlapping communities can be identified from the whole network.

2.3.4 Results

In this section, we extensively test our method on the artificial networks and the real world networks with known community structure. Then we apply our method to a large real world complex network, which has been shown to possess overlapping community structure.

2.3.4.1 Tests on Artificial Networks

To test our method, we utilize the benchmark proposed in [37]. It provides benchmark networks with heterogeneous distributions of node degree and community size. In addition, it allows for the overlaps between communities. This benchmark poses a much more severe test to community detection algorithms than Newman's standard benchmark [9]. There are many parameters to control the generated networks in this benchmark, the number of nodes N , the average node degree $\langle k \rangle$, the maximum node degree $\max_k$, the mixing ratio μ , the exponent of the power-law node degree distribution $t1$, the exponent of the power-law distribution of community size $t2$, the minimum community size $\min_c$, the maximum community size $\max_c$, the number of overlapped nodes on , and the number of memberships of each overlapped node om . In our tests, we use the default parameter configuration where $N = 1000$, $\langle k \rangle = 15$, $\max_k = 50$, $t1 = 2$, $t2 = 1$, $\min_c = 20$, $\max_c = 50$, $on = 50$ and $om = 2$. By tuning the parameter μ , we test the effectiveness of our method on the networks with different fuzziness of communities. The larger the parameter μ , the fuzzier the community structure of the generated networks is.

To evaluate the effectiveness of an algorithm for the identification of overlapping community structure, a measure is needed to compare the cover found by the algorithm with the ground truth. In [26], a measure is proposed to compare two covers, which is an extension form of *variation of information*. The more similar two covers are, the higher the value of the measure is. Here, we adopt it to compare the overlapping community structure found by our method and the known overlapping community structure in the benchmark networks.

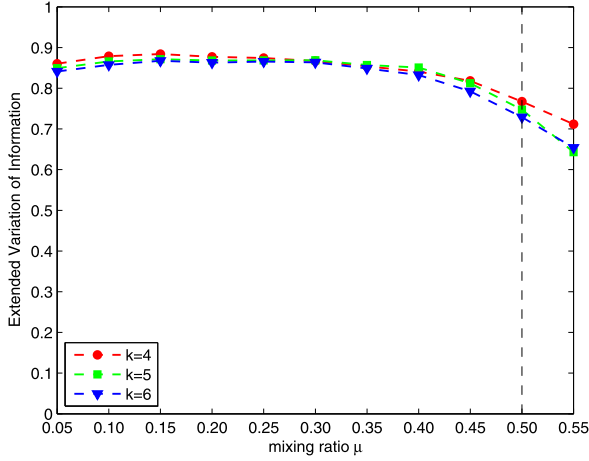


Fig. 2.9 Test of our method on the benchmark networks. The parameter k in the legend corresponds to the parameter k in our method. The threshold $\mu = 0.5$ (dashed vertical line in the figure) marks the border beyond which communities are no longer defined in the strong sense [10], i.e., such that each node has more neighbors in its own community than in the others. Each point corresponds to an average over 100 graph realization. Reprinted from Ref. [35], Copyright 2009, with permission from IOP Publishing and SISSA

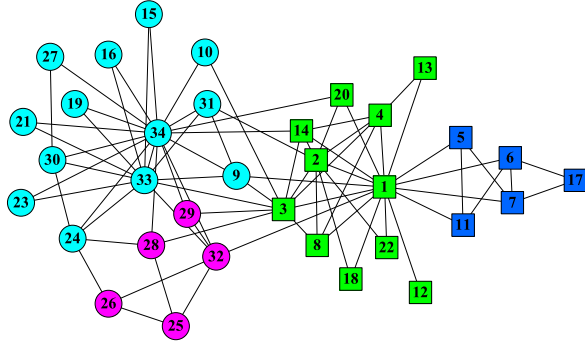
Figure 2.9 shows the results of our method with $k = 4, 5, 6$ on the benchmark networks. Our method gives rather good results when the μ is smaller than 0.5. All of the values of the variation of information are above 0.8. Note that in these cases, communities are defined in the strong sense [10], i.e., each node has more neighbors in its own community than in the others. We also test other settings of k which are larger than 6, and find similar results.

2.3.4.2 Tests on Real World Networks

Our first real world network for test is Zachary’s karate club network [38], which is widely used as a benchmark for the methods of community identification. This network characterizes the social interactions between the individuals in a karate club at an American university. A dispute arose between the club’s administrator and its principal karate teacher and as a result the club eventually split into two smaller clubs, centered around the administrator and the teacher respectively. The network and its fission is depicted in Fig. 2.10. The administrator and the teacher are represented by nodes 1 and 33 respectively.

Feeding this network into our method with the parameter $k = 4$, we obtain the result shown in Fig. 2.10. Similar to many existing community detection methods, our method partitions the network into four communities. This partition corresponds to the modularity with the value 0.417, while the real partition into two sub-networks has a modularity 0.371. Actually, no node is misclassified by our method. The real

Fig. 2.10 The network of the karate club studied by Zachary [38]. The real social fission of this network is represented by two different shapes, circle and square. The different colors show the partition obtained by our method with the parameter $k = 4$. Reprinted from Ref. [35], Copyright 2009, with permission from IOP Publishing and SISSA



split of the network can be obtained exactly by pair-wise merge of the four communities found by our method.

We also note that no overlaps are found when $k = 4$. Actually, no overlaps can be found when k is no smaller than 4 as to this network. Overlaps between communities emerge when the parameter k is set to 3. The value of Q_c corresponding to the resulting cover is 0.385 and in total three overlapped communities are found by our method. They are $\{1, 5, 6, 7, 11, 17\}$, $\{1, 2, 3, 4, 8, 9, 12, 13, 14, 18, 20, 22\}$ and $\{3, 9, 10, 15, 16, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34\}$. The overlapping regions consist of three nodes, being 1, 3 and 9. Each of them is shared by two communities. Such nodes are often misclassified by traditional partition-based community detection methods. Except the nodes occurring in the overlapping regions, other nodes reflects the real split of the network.

We also test our method on another real world network, a social network of 62 bottlenose dolphins living in Doubtful Sound, New Zealand. The network was constructed by Lusseau [39] with ties between dolphin pairs being established by observation of statistically significant frequent association. The network splits naturally into two groups, represented by the squares and circles in Fig. 2.11.

By applying our method with $k = 4$ to this network, four communities are obtained, denoted by different colors in Fig. 2.11. The green community is connected loosely to the other three ones. Regarding the three circle-denoted communities as a sole community, it and the green community correspond to the known division observed by Lusseau [39]. Furthermore, the three circle-denoted communities also correspond to a real division among these dolphins. The further division appears to have some correlation with the gender of these animals. The blue one consists mainly of females and the other two almost entirely of males.

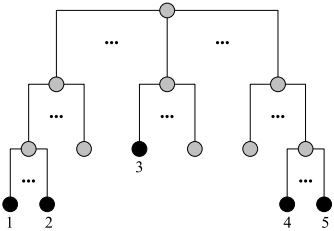
Alike to the Zarchay's karate network, the overlaps between communities cannot be detected when the parameter k is not less than 4. When $k = 3$, overlaps between the circle-denoted communities emerge while the green community keeps almost intact. The Q_c is 0.490 as to the resulting cover. The nodes occurring in overlapping regions are *Beak*, *Kringel*, *MN105*, *Oscar*, *PL*, *SN4*, *SN9* and *TR99* among which the nodes *Beak* and *Kringel* are shared by all the three circle-denoted communities. Again these overlapping nodes are often misclassified by traditional partition-based methods.

Table 2.1 The overlapping communities around the word *play*. Reprinted from Ref. [35], Copyright 2009, with permission from IOP Publishing and SISSA

No.	Description	Words in each community
1	theater	act actor actress bow character cinema curtsey dance director do drama entertain entertainment film guide involve juggler lead movie participate perform performance <i>play</i> portray producer production program scene screen show sing stage television theater
2	musical instrument	alto band banjo bass beep blues brass bugle cello clarinet clef compose concert conductor country drum fiddle fiddle flute guitar harp honk horn instrument jazz keyboard loud music oboe orchestra piano <i>play</i> rock saxophone symphony tenor treble trombone trumpet tuba tune viola violin woodwind
3	children	adults balls children family friends <i>fun</i> grown-ups guardians kids love mischief nursery parents <i>play</i> playground play_dough prank putty <i>toy</i> toys tricycle
4	sports	active arena athlete athletic baseball basketball black_and_white field football <i>fun game</i> illustrated inactive jock pigskin <i>play</i> recreation referee soccer sports stadium umpire
5	toys	board boardwalk checkers chess <i>fun game</i> games monopoly nintendo <i>play</i> plaything strategy <i>toy</i> toys vcr video winning yo-yo

Note: For each community, a short description is also given. The overlapped words are emphasized in italic type

Fig. 2.12 Part of the hierarchy of communities extracted from the word association network. The dark-filled circles correspond to the five communities shown in Table 2.1. Reprinted from Ref. [35], Copyright 2009, with permission from IOP Publishing and SISSA



ments, *children*, *sports* and *toys*. Except the common-shared word *play*, four other words are shared by some of these communities. They are *fun*, *game*, *toy* and *toys*. The overlap between these communities characterizes the direct, local relationship between them through sharing members. However, the extent of closeness between communities is sometimes reflected by the indirect, global relationship between them. One of this kind of relationship is the “genealogical” relationship between communities, which can be illustrated by the hierarchy of overlapping communities. Figure 2.12 is an example for hierarchy of communities. As shown in Fig. 2.12, the communities 1 and 2 are in the same branch of the hierarchy, indicating that the meanings represented by them are closer. This can be validated by examining the words contained in these two communities. Similarly, the communities 4 and 5 are also closely related. However, the distance between the communities 3 and 5 is larger although they share as many as 4 words. The overlaps between communities and the hierarchy of these communities provide us a more complete understanding to the relationship between communities.

2.4 Conclusions and Discussions

In this chapter, we have studied the problem of detecting both the overlapping and hierarchical community structure in networks. The distinct contribution is that we view a community as consisted of maximal cliques, instead of taking nodes as the building blocks of communities. In this way, the overlapping community structure of networks can be detected under the framework of traditional community detection methods.

Furthermore, representing the overlapping community structure as a cover of network, we propose two kinds of measurements to quantify the quality of a cover of network. The first of is a simple extension of modularity with the consideration that one node can simultaneously belong to more than one community with the same belonging coefficients. For the second one, we proposed a more general extension of modularity (namely Q_c) by using a relaxed belonging coefficients. With the Q_c at hand, the overlapping community structure can be detected by optimizing the Q_c to find the optimal cover of network. Then, a maximal clique network is constructed from the original network, and the overlapping community structure can be identified using any modularity optimization method on the maximal clique network.

In addition, Q_c takes advantage of both the local topological structure (i.e., the maximal clique) and the global statistical significance of link density compared with a null-model reference network. In addition, Q_c can be naturally used to simultaneously identify the overlapping and hierarchical community structure of networks. Such a method is helpful to more completely understand the functional and structural properties of networks. The effectiveness of the proposed methods are demonstrated by applications to several real world networks, including the word association network and the scientific collaboration network.

As the further work, we will consider the generalization to the weighted and/or directed networks. It is also an interesting problem about the selection of the parameter k in our method. We will further investigate how to determine an appropriate k for a given network later.

Finally, we give a brief introduction of further readings about overlapping community structure, which has been studied widely in the recent years.

References

1. Strogatz, S.H.: Exploring complex networks. *Nature* **410**, 268–276 (2001)
2. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002)
3. Newman, M.E.J.: The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003)
4. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99**, 7821–7826 (2002)
5. Guimerà, R., Amaral, L.A.N.: Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005)
6. Flake, G.W., Lawrence, S.R., Giles, C.L., Coetzee, F.M.: Self-organization and identification of Web communities. *IEEE Comput.* **35**, 66–71 (2002)

7. Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **74**, 036104 (2006)
8. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818 (2005)
9. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004)
10. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D.: Defining and identifying communities in networks. *Proc. Natl. Acad. Sci. USA* **101**, 2658–2663 (2004)
11. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. *Phys. Rev. E* **69**, 066133 (2004)
12. Newman, M.E.J.: Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **103**, 8577–8582 (2006)
13. Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* **76**, 036106 (2007)
14. Danon, L., Duch, J., Diaz-Guilera, A., Arenas, A.: Comparing community structure identification. *J. Stat. Mech.* P09008 (2005)
15. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. *Phys. Rev. E* **70**, 066111 (2004)
16. Duch, J., Arenas, A.: Community detection in complex networks using extremal optimization. *Phys. Rev. E* **72**, 027104 (2005)
17. Fortunato, S., Barthélemy, M.: Resolution limit in community detection. *Proc. Natl. Acad. Sci. USA* **104**, 36–41 (2007)
18. Kumpula, J.M., Saramaki, J., Kaski, K., Kertesz, J.: Resolution limit in complex network community detection with Potts model approach. *Eur. Phys. J. B* **56**, 41–45 (2007)
19. Baumes, J., Krishnamoorthy, M., Magdon-Ismail, M., Preston, N.: Finding communities by clustering a graph into overlapping subgraphs. In: *Proceedings of IADIS International Conference Applied Computing*, pp. 97–104 (2005)
20. Saito, K., Yamada, T., Kazama, K.: Extracting communities from complex networks by the k-dense method. In: *Proceedings of the 6th IEEE International Conference on Data Mining*, pp. 300–304 (2008)
21. Zhang, S., Wang, R.S., Zhang, X.S.: Identification of overlapping community structure in complex networks using fuzzy c-means clustering. *Physica A* **374**, 483–490 (2007)
22. Palla, G., Farkas, I.J., Pollner, P., Derényi, I., Vicsek, T.: Directed network modules. *New J. Phys.* **9**, 186 (2007)
23. Farkas, I.J., Ábel, D., Palla, G., Vicsek, T.: Weighted network modules. *New J. Phys.* **9**, 180 (2007)
24. Nicosia, V., Mangioni, G., Carchiolo, V., Malgeri, M.: Extending modularity definition for directed graphs with overlapping communities. *J. Stat. Mech.* P03024 (2009)
25. Evans, T.S., Lambiotte, R.: Line graphs, link partitions, and overlapping communities. *Phys. Rev. E* **80**, 016105 (2009)
26. Lancichinetti, A., Fortunato, S., Kertesz, J.: Detecting the overlapping and hierarchical community structure of complex networks. *New J. Phys.* **11**, 033015 (2009)
27. Sales-Pardo, M., Guimerà, R., Moreira, A.A., Amaral, L.A.N.: Extracting the hierarchical organization of complex systems. *Proc. Natl. Acad. Sci. USA* **104**, 15224–15229 (2007)
28. Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.L.: Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002)
29. Pons, P., Latapy, M.: Post-processing hierarchical community structures: Quality improvements and multi-scale view. *Theoret. Comput. Sci.* **412**, 892–900 (2011)
30. Shen, H.W., Cheng, X.Q., Cai, K., Hu, M.B.: Detect overlapping and hierarchical community structure in networks. *Physica A* **388**, 1706–1712 (2009)
31. Bron, C., Kerbosch, J.: Finding all cliques in an undirected graph. *Commun. ACM* 575–577 (1973)
32. Adamcsek, B., Palla, G., Farkas, I.J., Derényi, I., Vicsek, T.: CFinder: Locating cliques and overlapping modules in biological networks. *Bioinformatics* **22**, 1021–1023 (2006)

33. Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998)
34. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *J. Stat. Mech.* P10008 (2008)
35. Shen, H.W., Cheng, X.Q., Guo, J.F.: Quantifying and identifying the overlapping community structure in networks. *J. Stat. Mech.* P07042 (2009)
36. Arenas, A., Fernández, A., Gómez, S.: Analysis of the structure of complex networks at different resolution levels. *New J. Phys.* **10**, 053039 (2008)
37. Lancichinetti, A., Fortunato, S.: Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys. Rev. E* **80**, 016118 (2009)
38. Zachary, W.W.: An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **33**, 452–473 (1977)
39. Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E., Dawson, S.M.: The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Can geographic isolation explain this unique trait? *Behav. Ecol. Sociobiol.* **54**, 396–405 (2003)
40. Nelson, D.L., McEvoy, C.L., Schreiber, T.A.: The University of South Florida word association, rhyme, and word fragment norms (1998). <http://www.usf.edu/FreeAssociation/>

Community Structure of Complex Networks

Shen, H.-W.

2013, XIV, 117 p., Hardcover

ISBN: 978-3-642-31820-7