Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/090138

Discovering Mesoscopic-level Structural Patterns on Social Networks: A Node-similarity Perspective

*Qing Cheng**, *Zhong Liu*, *Jincai Huang and Guangquan Cheng*

Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology,Changsha, Hunan, 410073, P. R. China

Received: 7 Apr. 2014, Revised: 8 Jul. 2014, Accepted: 9 Jul. 2014
Published online: 1 Jan. 2015

Abstract: Structural pattern analysis is of fundamental importance as it provides a novel perspective on illustration of the relationship between structure and function, as well as to understand the dynamics, of social networks. So far, scientists have uncovered a multitude of structural patterns ubiquitously existing in social networks in different levels, they may be microscopic, mesoscopic or macroscopic. Our work mainly characterizes the mesoscopic-level structural patterns on social networks from the node-similarity viewpoint and reviews some latest representative methods, focusing on the improved methods of community measure and community structure detection, role discovery methods, as well as the structural group discovery approaches used to reveal hidden but unambiguous structures. Finally, we also outline some important open problems, which may be valuable for related research domains.

Keywords: Structural pattern, community, role, structural group

1 Introduction

A network is, in its simplest form, a collection of nodes joined together in pairs by edges, and social networks [1, 2], in which the nodes are people or group and the edges represent any of a variety of different types of social interaction including friendship, collaboration, business relationships or others. Examples include the network of scientific collaboration (Figure 1(a)), network of Enron communication (Figure 1(b)), Co-authorship network (Figure 1(c)) and Facebook friend network (Figure 1(d)) [3].

In the past decade there has been a surge of interest in both empirical studies of networks and development of mathematical and computational tools for extracting insight from network data. However, a difficult problem when studying networks is that of global values of statistical measures can be misleading, and cannot clearly unveil insights into their functional organization [4], and there is no standard network visualization and quantitative description method show clear large-scale network. In order to synthesize realistic social networks, we usually start with the studies of the structural patterns. So far, scientists have uncovered a multitude of structural patterns ubiquitously existing in social networks [9]. They may be microscopic, such as motifs [5]; mesoscopic, such as communities [6]; or macroscopic, such as small worlds [7] and scale-fee phenomena [8]. See Figure 2(a), one can observe structural patterns in different levels, and hierarchy describes how the various structural patterns are combined. In the spite of the great efforts of pattern analysis having been made, we will focus our discussion in this paper on mesoscopic level. Because based on Mesoscopic-level structural patterns, such as community, role and structural group, one can make a step towards the uncovering of the modular structure of social networks and unveil insights into their functional organization, which would greatly benefit both theoretical studies and practical applications. For example, in a metabolic network, the network of chemical reactions within a cell?a community might correspond to a circuit or pathway that carries out a certain function, such as synthesizing or regulating a vital chemical product [10], and mesoscopic-level structural patterns can also be use to compress a huge network, resulting in a smaller network. In other words, problem solving is accomplished at group level, instead of node level. In the same spirit, a huge network can be visualized at different resolutions, offering an intuitive solution for network analysis and navigation [11]. However, to the best of our knowledge,

* Corresponding author e-mail: sgggps@163.com

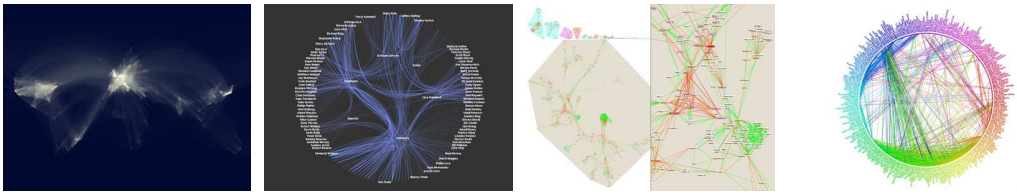


Fig. 1: (a)Map of scientific collaboration(b) Enron communication graph (c) Co-authorship network-LRI Lab (d) Facebook friend wheel

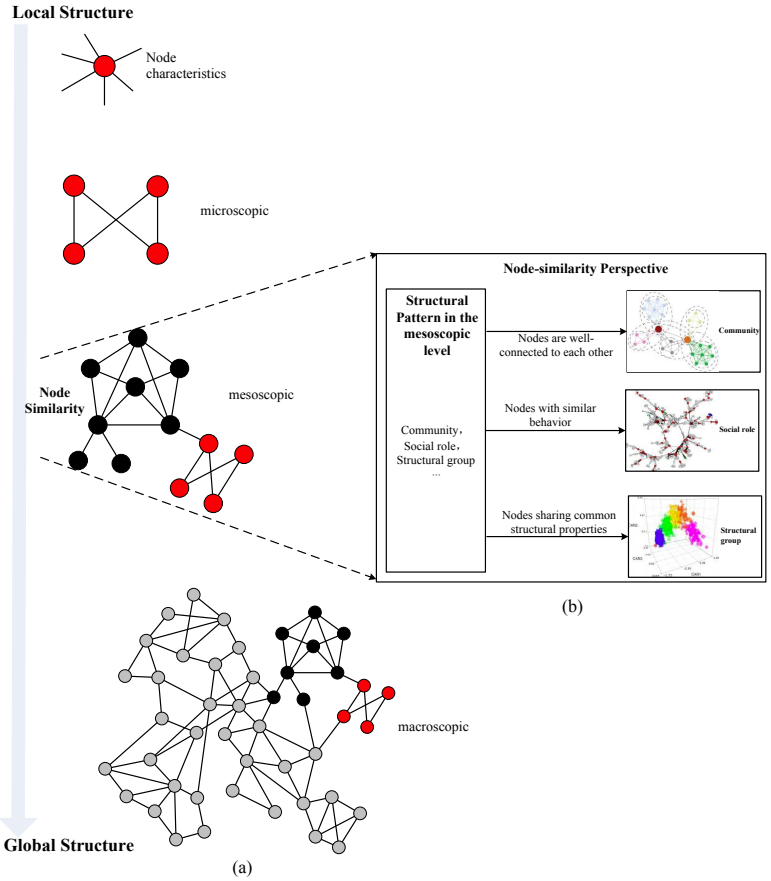


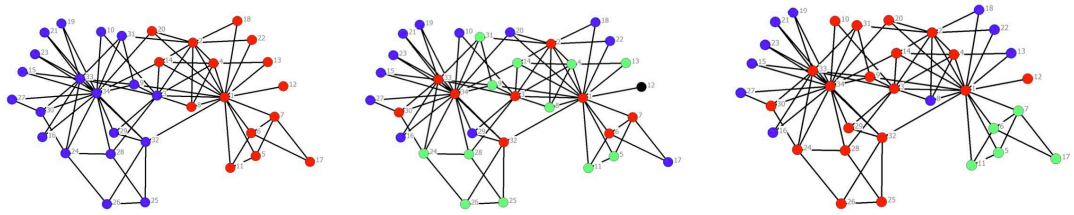
Fig. 2: Schematic illustrations of the scales of organization of social networks. (a)Observing structural patterns in different levels, hierarchy describes how the various structural patterns are joined into the entire network.(b) characterize mesoscopic-level structural pattern in the from a node similarity viewpoint

there have been no studies in the literature that explicitly and adequately characterize mesoscopic-level structural patterns, thus, this article aims to characterize such structures in a simple way and review the studies of mesoscopic-level structural pattern (For the sake of presentation, without loss of rigor, we will use the term structure to denote mesoscopic-level structural pattern in the rest of this paper).

This article is organized as follows. In the next section, we will characterize structure from a node-similarity viewpoint and mainly introduce three kinds of structure: community, role and structural group. In Section 3, we will review the community discovery from two aspects: the measure for community and structure of community, emphasizing on the optimization method and hierarchical clustering, due to they are widely used for discovering simultaneously both the hierarchical

Table 1: The definitions of community, social role and structural group

Structure	Node-similarity	Definition
Community	nodes densely-connect	Community, is a densely connected subset of nodes that is only sparsely linked to the remaining network. [6]
Social role	nodes with similar behavior	Social role, groups nodes of similar structural behavior (or function) [1]. In social network analysis position refers to a collection of individuals who are similarly embedded in networks of relation. While role refers to the patterns of relations which obtain between actors or between positions. In this paper we cannot distinguish two conceptions, and both of them are called as social role in the rest paper.
Structural group	nodes sharing common properties	Structural groups, defined as subsets of nodes sharing common structural properties that set them apart from other nodes in the network. [14]



(a) (b) (c)

Fig. 3: Comparison of community discovery, role discovery and structural group discovery(a)The 2 communities that GN algorithm [6] finds on the karate network.(b)The 4 roles discovered by optimization method based on regular equivalence (by UCINET tool): "bridge" nodes(as red), "periphery" nodes(as blue). (c) The 3 structural groups discovered by visual analytics method using the selection of 28 node properties [14]

and overlapping community structures. In Section 4, we introduce the social role discovery from perspectives of sociology and data mining. The structural group discovery methods are presented in Section 5, including the maximum likelihood methods and visual analytics methods. Finally, we outline some future challenges of structure discovery.

2 Mesoscopic-level structural patterns

The structure detection problem is challenging that a precise definition of what a "structure" really is does not exist at the current stage. However, it is widely agreed that structure groups nodes have similar property, function or behavior, such as community, social role and structural group. Based on this, we characterize structures by using a node-similarity viewpoint, seeking to identify and classify the structure and gasp its topological properties. There are three major types of structures: community, social role, and structural group (Figure 2(b)), which by far the most studied and best known structures in the literature, and they definitions as table 1.

More specifically, from the perspective of density-based similarity measure, it is obvious that two nodes are considered similarity if they are well-connected, such as they share many of the same network neighbors. Therefore, the community structure can be defined as a densely connected subset of nodes that is only sparsely linked to the remaining network [6].

There are, however, many cases in which nodes occupy similar structural position in networks without having well-connected, for instance, two store clerks in different towns occupy similar social positions by virtue of their numerous professional interactions with customers, although it is quite likely that they have none of those customers in common and they are not well-connected. In this case, nodes are referred as similar if they have similar behavior or their pattern of relationships is equivalent, and we call this structure type as social role.

Additionally, node similarity can be defined by using the essential attributes of nodes: two nodes are considered to be similarity if they have many common features, not only restricting to structural attributes, but also including quantifiable properties, such as age, income and level of education. Therefore, the structure is called structural

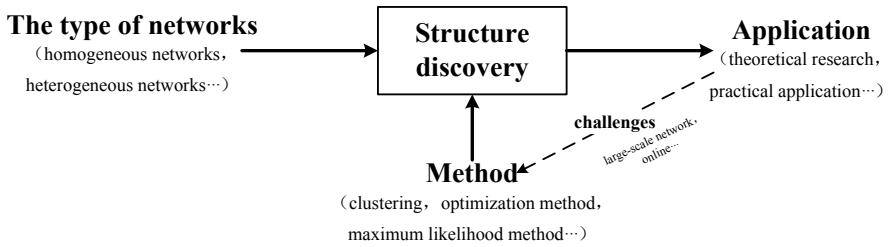


Fig. 4: the process of structure discovery

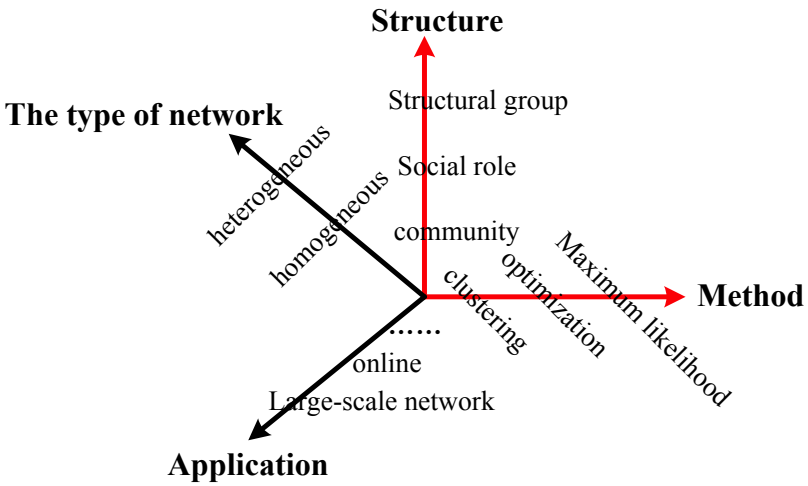


Fig. 5: Diagram showing the focus of this paper using four dimensions

group which can be defined as subsets of nodes sharing common properties that set them apart from other nodes in the network [14].

Furthermore, we want to emphasize that community is fundamentally different from (and complementary to) the social role: the former groups nodes are well-connected to each other, but the latter groups nodes have similar behavior [12]. Figure 3 depicts the difference between role discovery and community discovery for the karate network. The GN algorithm [6] discovers 2 communities (Figure 3(a)) vs. the 4 roles (Figure 3(b)) that the optimization method finds [13]. And structural group discovery tend to reveal a hidden but unambiguous structures beyond communities and social roles. For example, in Figure 3(c), the 3 structural groups discovered by visual analytics method [14] using the selection of 28 node properties. Group 1(blue) is characterized by low degree, clustering coefficient of one, and being connected to high-degree nodes with high betweenness and subgraph centrality. Group 2(red) forms the core of the network and includes all the high-degree, high-centrality nodes [14]. In fact, structural group includes but not limited to community and role, because the densely-connected and behavior may be regarded as a kind of node property. Within the wide range of possible structures expressible though different properties, the structural group discovery method can help discover a specific structure of interest and interpret it using a ranking of the node properties.

Based on community, social role and structural group, various structure discovery methods have been proposed. Generally, what methods are used to discovery structures relies on both what type of network one wish to answer and what practical application one confronted with, and various practical application give challenges to the existing methods, meanwhile, these challenges facilitate the method improvement and technical innovation. Figure 4 presents the process of structure discovery. Obviously, one discovers structures in social network should start with four aspects: type of networks, practical application, methods and challenges, as Figure 5. We aim to give detailed discussion on the methods as follow (as red directions in Figure 5). Figure 6 presents a synoptic picture of the works that will be reviewed, organized according to the community, role and structural group discovered by each method and the solution technique.

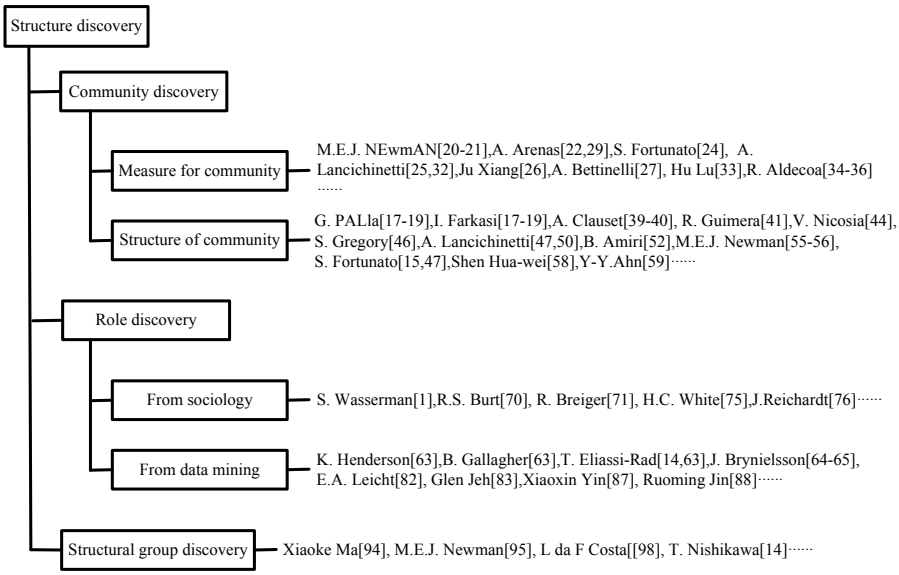


Fig. 6: The taxonomy of the reviewed structure discovery method

3 Community discovery

Community detection has become one of the most important topics in social network analysis. A huge variety of different methods of community detection have been designed by a truly interdisciplinary community of scholars, including physicists, computer scientists, mathematicians and social scientists (for earlier reviews see Refs. [15,16]). Generally speaking, a "good" community is taken to refer to a subset of nodes that are (1) well connected among themselves, and (2) well separated from the rest of the graph. Hence, algorithms for network clustering are often based on a subjective quality measure that can be applied on a potential cluster, meanwhile, communities are usually overlapping and hierarchical [17,18,19], and recently many researchers started to focus on the problem of identifying such realistic structures. Therefore, in general, algorithms for network clustering differ in (1) how the quality of the proposed clusters is measured, and (2) what kind of technique is used to obtain this desire quality, especially finding hierarchical and overlapping community. Moving from considerations about the meaning and structure of the community, we mainly review some latest representative researches from the aspects of measure for community and structure of community.

3.1 Measure of community

There is no consensus criterion for measuring the community structure, which is a main drawback in many algorithms. To tackle this difficulty, most methods are based on modularity function. Which is introduced by Newman et al [20]. The modularity function Q which measures the quality of a given partition of a network,

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(C_i, C_j) \qquad (1)$$

where k_i is the degree of node i and m is the total number of edges in the network. A_{ij} is an element of the adjacency matrix, $\delta(C_i, C_j)$ is the Kronecker delta symbol, and C_i is the label of the community to which node i is assigned, and $\delta(C_i, C_j) = 1$ if $C_i = C_j$ otherwise 0. Then one maximizes Q over possible divisions of the network into communities, the maximum being taken as the best estimate of the true communities in the network. That is to say, high values of modularity indicate stronger community structure, corresponding to more dense connections within communities. And the modularity function is extended to weight networks [21] and directed networks [22,23] for detecting community structure. Modularity is by far the most used and best known quality function for the measure.

However, the modularity maximization suffers from a resolution limit [24,25]: small communities may be undetectable in the presence of larger ones even if they are very dense. S. Fortunato et al. recently claimed that modularity optimization may fail to identify network in which the number of communities is larger than about $\sqrt{L}$, where L is the number of edges in entire network [24], moreover, the theoretical analysis and the experimental tests in several network examples indicated that the limitation depends on the degree of interconnectedness of small communities and the difference between the sizes of small communities and large communities, while independent of the size of the

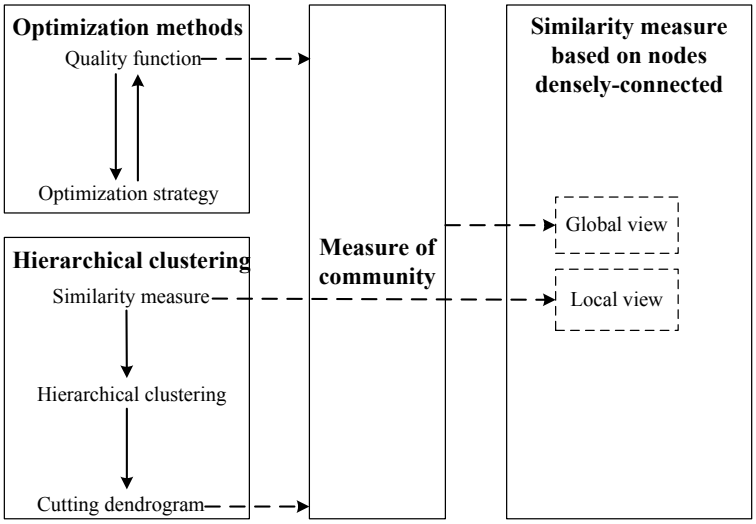


Fig. 7: The measure of community and the main method for overlapping and hierarchical communities discovery

whole network [26]. To solve the resolution problem, various methods have been proposed. These methods can be roughly classified into two categories in terms of their ideas:

The first kind of methods modifies the modularity function through tunable parameters [27]. In particular, a type of multi-resolution methods in community detection was introduced, which can adjust the resolution of modularity by modifying the modularity function with tunable resolution parameters, for example, Reichardt and Bornholdt (RB) [28] discussed a modified version of the modularity function which introduces a parameter to tune the contribution of the null model in the modularity; Arenas, Fernandez and Gomez (AFG) [29] also proposed a multi-resolution method by providing each node with a self-loop of the same magnitude r, which is equivalent to modifying the modularity function by the parameter r, and Andrea Bettinelli et al. extended modularity to parametric modularity by using a single parameter that balances the fraction of within community edges and the expected fraction of edges according to the configuration model [27], and so on. These multi-resolution methods indeed can help us find the communities of networks at different scales. To some extent, by varying their parameters to adjust the resolution of the modularity, but before being applied to the real problem of community detection, the methods based on modularity optimization all should have been thoroughly understood to ensure that the substructures found in networks are reasonable [30].

The second kind of methods aims to solve the resolution problem by introducing new quality function [31,32]. For example, the network community coefficient C is defined as the average community coefficient over all nodes in the network. While it depends on the correctness of partitioning methods. Only when the community structure is correctly divided can optimizing the value of C correctly identify the optimal number of communities [33]. Rodrigo Aldecoa et al introduced a new global measure, called Surprise(S), which has an excellent behavior in all networks tested [34,35,36], Surprise measures the probability of finding a particular number of intracommunity links in a given partition of a network, assuming that those links appeared randomly, and so on. Ideally, we would like a more reliable method to solve resolution problem.

3.2 Structure of community

Communities may be in complicated shapes. Palla et al. [17,18,19] revealed that complex network models exhibit an overlapping community structure, and Ravasz et al. [37] proved the existence of the hierarchical organization of modularity in metabolic networks. These overlapping and hierarchical communities are more realistic than average ones. Now, only a few efficient algorithms can uncover such realistic structure [38]. We focus on the well-known technique: optimization methods and hierarchical clustering, which based on the measure of community, corresponding to communities are characterized by groups of densely connected nodes, such as Figure 7.

(1) Optimization methods, are those that view the community-detection problem as an optimization task. The basic idea is to define a quality function that is high for “good” divisions of a network and low for “bad” ones, and then to search through possible divisions for the one with the highest score. Most classic methods treat modularity function as the quality function, then, the community detection is switch to modularity

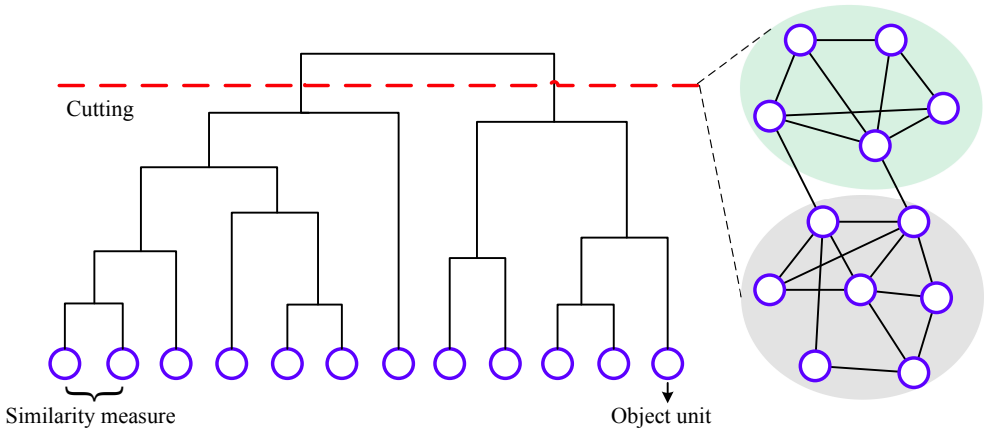


Fig. 8: The process of hierarchical clustering for community detection

maximization. Although finding the optimal Q value is an NP-hard problem, there are currently several methods able to find fairly good approximations of the modularity maximum in a reasonable time, such as greedy techniques [39,40], simulated annealing [41], external optimization [42], and genetic algorithms [43]. We are able to combine node overlap with hierarchical structure in a united framework and have converted the task of finding overlapping and hierarchical communities into a optimization problem, using the quality (objective) function that reflecting the community structure of overlapping or(and) hierarchy. Some previous researchers proposed many efficient methods for extending the modularity function or improving the optimization strategy to meet the requirement for discovering hierarchical and overlapping of community [44,45,46]. Moreover, one could choose a different expression for the quality functions, another criterion to define the most meaningful cover, or a different optimization procedure of the quality functions for a single cluster from different perspectives. For example, A.Lancihinetti et al. [47,48] think a community is subgraph identified by the maximization of a property or fitness of its nodes, the method based on the local optimization of a fitness function was presented to find both overlapping communities and the hierarchical structure. In addition, they also presented OSLOM(Order Statistics Local Optimization Method) [50] explores the clusters in networks accounting for overlapping communities and hierarchies, based on the local optimization of a fitness function expressing the statistical significance of clusters with respect to random fluctuations, which is estimated with tools of Extreme and Order Statistics. And Bo Yang et al explores the nature of community structure for a probabilistic perspective and introduces a novel community detection algorithm named as PMC(probabilistically mining communities), to meet a good trade-off between effectiveness and efficiency. In PMC, community detection is modeled as a constrained quadratic optimization problem that can be efficiently solved by a random walk based heuristic [49], etc.

Unfortunately, these methods employ single optimization criteria, which may not be adequate to represent the structures in social networks. Many researchers [51,52,53] suggest community detection process as a multi-objective optimization problem (MOP) for investigating the community structures in social networks. That is, the community detection corresponds to discovering community structures that are optimal on multiple objective functions, instead of one single-objective function in the single-objective community detection, such as multi-objective community detection algorithm (MOCD) [53], MOGA-Net [54] and EFA(enhanced firefly algorithm) [52]. The experimental results on synthetic and real world complex networks suggest that the methods based multi-objective optimization can discover more the accurate and comprehensive community structure compared to those well-established community detection algorithms, as well as provides useful paradigm for discovering overlapping or(and) hierarchical community structures robustly.

However, most of optimization method can effectively achieve overlapping community detection, to some extent the hierarchical community is relatively weak.

(2) Hierarchical clustering. An early, and still widely used, method for detecting communities in social networks is hierarchical clustering [55,56]. Strategies for hierarchical clustering generally fall into two types: agglomerative and divisive. Considering divisive hierarchical clustering was rarely applied in community detection and hard to detection overlapping community, we focus on agglomerative hierarchical clustering (we will use the hierarchical clustering to denote agglomerative hierarchical clustering in the rest of this paper). Hierarchical clustering is in fact not a single

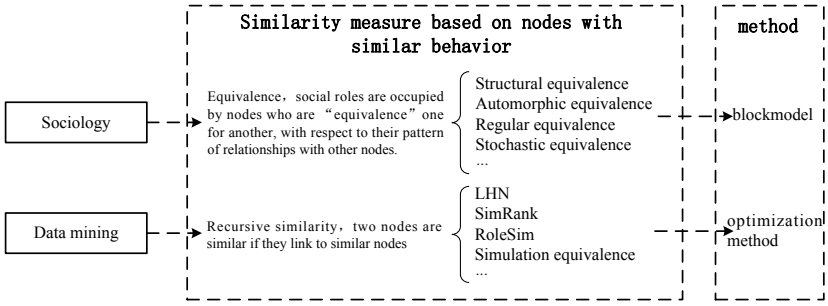


Fig. 9: the sketch map of role discovery from perspectives of sociology, data mining and computer science

technique but an entire family of techniques, with a single central principle:

The object units are taken as the initial communities if we calculate the similarity between each pair of communities(generally, node is often regarded as the object [39,40,46,53,57] due to it is the basic elements of the network, while the maximal cliques [58], links [59] and hyper-edge [60] also regard as the object in some researches.); then 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. This process is repeated until all nodes belong in a single cluster, the order in which nodes clustered together is then stored in a hierarchical tree of object unit known as a dendrogram, such as Figure 8. Finally, choosing the cut through the dendrogram that corresponds most closely to the known division of the communities, as indicated by the dotted line in the Figure 8.

Simply, the hierarchical clustering 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. Here, the key problems are how to define similarity between objects, and where to cut the dendrogram.

There are various ways to define a similarity between two objects, the node-dependent similarity and path-dependent similarity are by far the most used and best known in community detection. The node-dependent similarity index is common neighbors where the similarity of two objects is directly given by the number of common neighbors, such as Jaccard Index and Cosine Index [59,60], and the path-dependent similarity index assume that two objects are similar if they are connected by many paths, such as GENs(Generalized Erdos Numbers) based similarity [57] and shortest path based similarity [61]. In addition, there are many other ways to define a similarity between two objects we can refer to [62]. But no matter how to define similarity, it is dependent on nodes densely connected.

To find meaningful communities rather than just the hierarchical organization pattern of communities, it is crucial to know where to partition the dendrogram. In fact, that is to say, using a quality measure that helps us evaluate the goodness of communities generated by cutting the dendrogram at a particular threshold. What quality measure we could adopt relies on the special application we confronted with. The modified modularity has been widely used for this purpose, and can apply to overlapping communities [58]. And one could choose a different criterion to define quality measure, such as partition density, that measures the quality of a link partition [59].

Hierarchical clustering is a method that used widely to find overlapping or(and) hierarchical community, and it is straightforward to understand and to implement. But it has a tendency to group together those nodes with the strongest connections but leave out those with weaker connections, so that the divisions it generates may not be clean divisions into groups, but rather consist of a few dense cores surrounded by a periphery of unattached nodes [16].

We remark that the quality function in optimization method and the quality measure in hierarchical clustering are problems of measures of community, corresponding to communities are characterized by groups of densely connected nodes. Both optimization method and hierarchical clustering are techniques to obtain this desire quality, as shown in Figure 7. Thus, combining optimization method with hierarchical clustering in a united framework may provide useful hints for discovering hierarchical and overlapping of community.

4 Social role discovery

Social role discovery was first introduced in sociology, and recent studies have found not only do roles appear in social networks, but also in other types of networks, including food webs, world trade networks, and even software systems. A key question in studying the roles in a network is how to define role similarity. In terms of different definitions of role similarity, we review some latest representative researches about role discovery from

perspectives of sociology and data mining [63], as Figure 9.

4.1 Sociology viewpoint

In sociology, role analysts seek to define categories and variables in terms of similarities of the patterns of relations among actors (nodes), rather than attributes of actors. That is, the definition of a category, or a "social role" or "social position" depends upon its relationship to another category [66]. In an intuitive way, we would say social roles are occupied by nodes who are "equivalence" one for another, with respect to their pattern of relationships with other nodes (relational ties). This "equivalence" can be classified into two categories: deterministic equivalences and probabilistic equivalences [63], such as Figure 10 where the deterministic

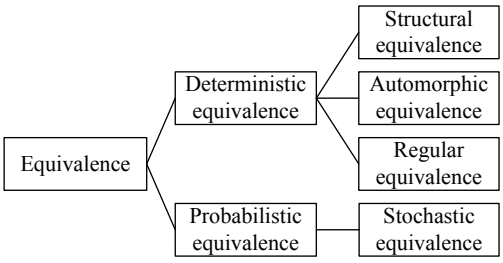


Fig. 10: The various categories of equivalence

equivalences fall into one of three categories: structural equivalence [67], automorphic equivalence [68] and regular equivalence [69].

Generally speaking, two nodes are said to be exactly structural equivalence if they have the same relationships to all other nodes, in Figure 11 there are seven structural equivalence classes: $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5,6\}$, $\{7\}$, $\{8,9\}$. Because exact structural equivalence is likely to be rare, we often are interested in examining the degree of structural equivalence, rather than the simple presence or absence of exact equivalence. Any measure of structural equivalence quantifies the extent to which pairs of actors meet the definition of structural equivalence. Euclidean distance and correlation are the most commonly used measures (Euclidean distance in STRUCTURE [70], and correlation in CONCOR [71], and both are widely available in network analysis computer programs as well as in standard statistical analysis packages). Besides, the researcher could consider alternative similarity measures, such as dichotomous relation and an ordered scale.

The idea of automorphic equivalence is that sets of nodes can be equivalent by being embedded in local structures that have the same patterns of ties¿'parallel" structures. In Figure 11, there are actually five

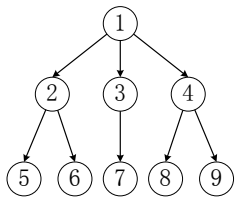


Fig. 11: Wasserman-Faust network to illustrate equivalence classes [1]. There are seven structural equivalence classes: $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5,6\}$, $\{7\}$, $\{8,9\}$, five automorphic equivalence classes: $\{1\}$, $\{2,4\}$, $\{3\}$, $\{5,6,8,9\}$, $\{7\}$, and three regular equivalence classes: $\{1\}$, $\{2,3,4\}$, $\{5,6,7,8,9\}$.

automorphic equivalence classes: $\{1\}$, $\{2,4\}$, $\{3\}$, $\{5,6,8,9\}$, $\{7\}$. Simply, these classes are groupings who's members would remain at the same distance from all other nodes if they were swapped, and, members of other classes were also swapped [66]. Compare with structural equivalence, automorphic equivalence is a bit more relaxed.

Two nodes are said to be regularly equivalent if they have the same profile of ties with members of other sets of actors that are also regularly equivalent. More generally, if node i and j are regularly equivalent, and node i has a tie to/from some node, k, then node j must have the same kind of tie to/from some node, l, and actor k and l must be regularly equivalent. In Figure 11 there are three regular equivalence classes: $\{1\}$, $\{2,3,4\}$, $\{5,6,7,8,9\}$. One of the earliest and most widely used measures of regular equivalence is embodied in the algorithm REGE proposed by White and Reitz [72]. More recently, authors have focused on methods for assigning actors to subsets such that the partition of actors is optimal in the sense that nodes in the same subset are nearly regularly equivalent [73,74].

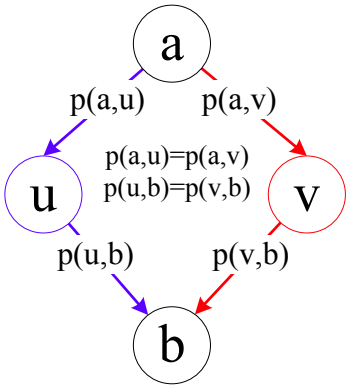


Fig. 12: A example network to illustrate stochastic equivalence [63], node u and node v are stochastically equivalent.

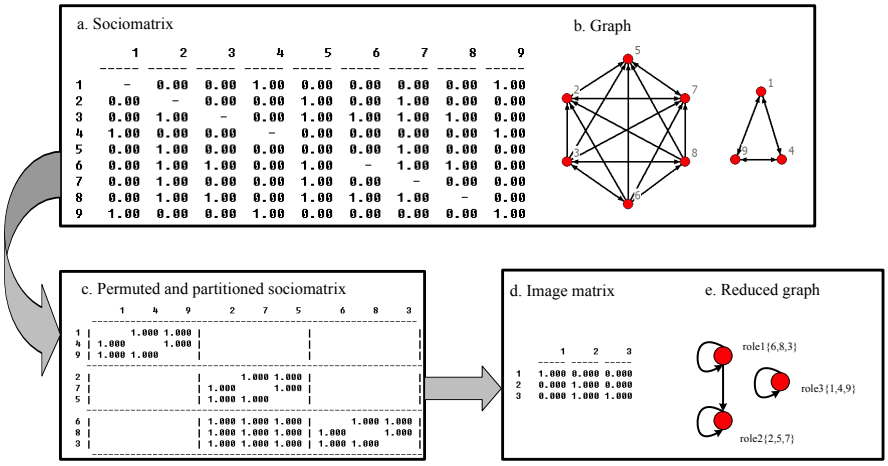


Fig. 13: [77] Example simplifying a network using blockmodel

Stochastic equivalence is similar to structural equivalence but probabilistic. Formally, two nodes are stochastically equivalent if they are "exchangeable" w.r.t a probability distribution, such as Figure 12.

By using "equivalence" as the measure of similarity among nodes, the social role discovery in sociology aims to group nodes with equivalence relation into a class, called a role. Blockmodels were widely used model for the discovery and analysis of social roles [1,75,76]. For instance, CONCOR is a kind of blockmodel based on the structural equivalence. The process of creating a blockmodel contains following steps:

1. Identify which type of equivalence is applied to measure the similarity among nodes, for example, structural equivalence, regular equivalence, etc.
2. According to the equivalence, partition nodes in the network into discrete subgroups positions, i.e. permute and partition sociomatrix, as Figure 13(c).
3. For each pair of positions, presence or absence of relational ties, create the image matrix, as Figure 13(d).
4. Finally, create the reduced graph according to image matrix that is social role, as Figure 13(e).

Blockmodel was proposed to discover social role as well as relationship among the roles. Since then there have been many articles describing blockmodels from a methodological standpoint, comparing blockmodels with alternative data analytic methods, discussing alternative methods for constructing blockmodels, and proposing some generalized blockmodels. Specially, recently several authors have generalized blockmodels by describing stochastic blockmodels [78,79,80,81], there have also been many applications of blockmodels and generalized blockmodels to social role discovery, you can find more details in [1].

4.2 Data mining viewpoint

From data mining viewpoint, role similarity is based on the following principles:"two nodes are similar if they link to similar nodes", based on it, the nodes can be partitioned into classes using a ranking of the node similarity, the widely used methods as LHN(was proposed by Leicht, Holme and Newman) [82], SimRank [83], RoleSim [88] and simulation relation [64,65] etc.

The fundamental principle behind LHN similarity is that i is similar to j if i's neithbor is similar to j, as Figure 14:

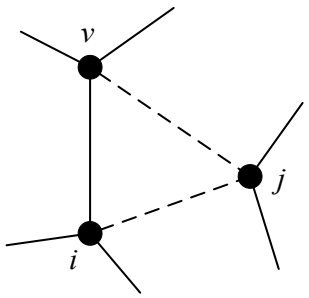


Fig. 14: A node j is similar to node i (dashed line) if I has a network neighbor v (solid line)that is itself similar to j [82]

Thus, the LHN similarity can written as

$$S_{ij} = \phi \sum_{v} A_{iv} S_{vj} + \psi \delta_{ij} \qquad (2)$$

where S_{ij} is similarity of node i to node j, which consist of two components: The direct similarity of node i to

node j, denoted by δ_{ij}, and the indirect similarity based on the local path between two nodes, denoted by $\sum_v A_{iv}S_{vj}$, and ϕ, ψ are free parameters whose values control the balance between the two components of the similarity. It is obvious that LHN similarity is a kind of recursive structural similarity, and its precise mainly depends on the choice of parameters. The generalization of LHN and the method of parameter tuning can be found in [82].

SimRank [83] was used to match text across documents. Recently, several researchers have tried to apply it to role modeling [84]. The SimRank similarity between nodes a and b is the average similarity between a's neithbors and b's neithbors:

$$s(a,b) = \frac{C}{|I(a)||I(b)|}\sum_{i=1}^{|I(a)|}\sum_{j=1}^{|I(b)|} s(I_i(a),I_j(b)) \qquad (3)$$

where C is a constant between 0 and 1. $I(a)$ is the set of in-neighbors of a. Mathematically, for any two different nodes u and v, SimRank computes their similarity recursively according to the average similarity of all the neighbor pairs. A fixed-point algorithm was presented for computing SimRank scores, as well as methods to reduce its time and space requirements. Inspired from SimRank, the number of variants of simrank soared [84,85,86,87], such as SimRank++ and P-SimRank.

SimRank has a problem when there is an odd distance between two nodes. Nodes u and v are automorphically equivalent, but because there are no nodes that are an equal distance from both u and v, $s(u,v) = 0$. And other variants of SimRank also do not meet the automorphic equivalence property [88]. According to this problem, the first real-valued similarity measure RoleSim, confirming automorphic equivalence, was proposed in [88]. Given two nodes u and v, where $N(u)$ and $N(v)$ denote their respective neighborhoods and N_u and N_v denote their respective degrees. The RoleSim measure realizes the recursive node structural similarity principle "two nodes are similar if they relate to similar objects" as follows

$$\mathrm{RoleSim}(u,v) = (1-\beta)\max_{M(u,v)}\frac{\sum_{x,y\in M(u,v)}\mathrm{RoleSim}(x,y)}{N_u+N_v-|M(u,v)|}+\beta \qquad (4)$$

Where $x \in N(u)$, $y \in N(v)$, $M(u,v) = \{x \in N(u), y \in N(v),$ and no other $(x',y') \in M(u,v),$ s.t.,$x = x'$ or $y = y'\}$, the parameter β is a decay factor, $0 < \beta < 1$.

RoleSim values can be computed iteratively and are guaranteed to converge, just as in SimRank. But unlike SimRank, which considers the average similarity among all possible pairings of neighbors, RoleSim counts only those pairs in the matching of the two neighbor sets which maximizes the targeted similarity function. The experiments in [88] shown that the iterative RoleSim computation generates a real-valued, admissible role similarity measure.

The simulation relation [89] creates a partial order on the set of nodes in a network and we can use this order to identify nodes that have characteristic properties. And the simulation relation can also be used to compute simulation equivalence. We use simulation equivalence to create equivalence classes that form roles in the social network.

Definition 1*Simulation preorder [64,65], The relation $\preceq$ is a preorder if it is reflexive and transitive. And $u \preceq v$ denotes node u simulate v if the fact that:*

1. *u and v have the same label.*
2. *For each $a \in \Sigma$ and $(v,v') \in E_a$, there is an edge $(u,u') \in E_a$ such that $u' \preceq v'$.*

where Σ is set of labels of nodes and edges, and if network G has edge labels in Σ, the E is a Σ-indexed family E_a of sets of edges, such that $E_a \subseteq V \times V$, for each $a \in \Sigma$. According to the definition of simulation preorder, for each $u,v \in V$, u and v is simulation equivalent if $u \preceq v$ and $v \preceq u$.

Therefore, we tentatively term the equivalence classes determined by simulation equivalence social roles [89]. In addition, in computer science, the regular equivalence is often referred to as the bisimulation, which is widely used in automata and modal logic [90]. After the simulation relation was applied in social role analysis, many authors tried to generalize and revise simulation relation to effectively and efficiently identify social roles or groups, such as strong simulation [91] and bounded simulation [92].

According to discussed above, we conclude that these methods from data mining viewpoint are all based on recursive structural similarity measure and comply with the "equivalence" requirement, the LHN and simulation relationship confirm regular equivalence, the Simrank confirms structural equivalence and the RoleSim confirms automorphic equivalence. Moreover, these methods use optimization algorithms for computing the maximal similarity scores on the network to obtain the social role that is available in heterogeneous and is still computable in polynomial time. This means that these methods can be computed in reasonable time for very large networks.

However these methods fail to achieve automatically extracting roles. Recently, Keith Henderson et al. [12] proposed a new method RolX(Role eXtraction), a scalable, unsupervised learning approach for automatically extracting structural roles from general network data, and demonstrated the effectiveness of RolX on several network mining tasks. More precisely, RolX consists of three components: feature extraction, feature grouping, and model selection, and achieves the following two objectives. First, with no prior knowledge of the kinds of roles that may exist, it automatically determines the underlying roles in a network. Second, it appropriately assigns a mixed-membership of these roles to each node in the networks. The Figure 15 illustrates the process of RolX.

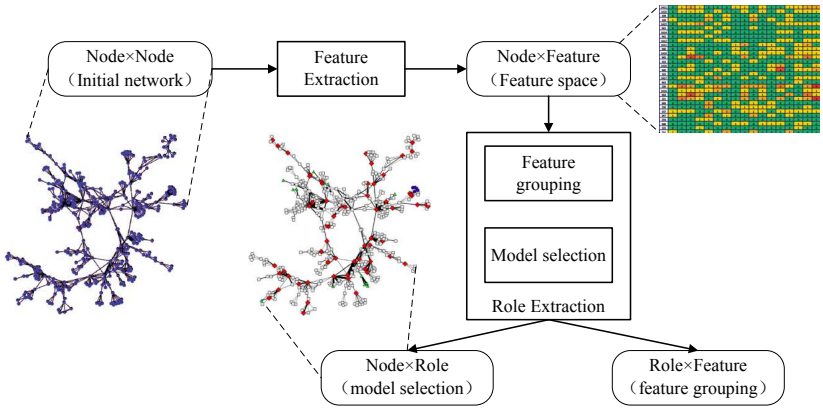


Fig. 15: RolX for role extraction

It is important to note that we do not argue that a method, e.g. simulation equivalence is a "better" way to study the social network than others, e.g., regular equivalence. All of these methods are useful, depending on the question one wishes to answer about the network. And other reason is that researchers have proposed a wide variety of definitions, from sociology and data mining, etc. With this array of definitions comes a corresponding array of algorithms that seek find the roles so defined. Unfortunately, it is no easy matter to determine which of these algorithms are the best, because the perception of good performance itself depends on how one defines a role and each algorithm is necessarily good at finding roles according to its own definition. So a unit/standard criterion to evaluate the methods of role discovery is key research direction in future.

5 Structural group discovery

Structural group discovery aims at seeking to capture more general structures characterized by other networks properties, tending to mine some hidden but unambiguous structures without knowing the groups a priori.

The previous work on structural group discovery has focused mainly on cluttering method. The choice of an appropriate similarity measure in clustering method is very important, which include not only density-based similarity, behavior-based similarity discussed above, but also many other similarity measure, such as feature-based similarity measures, distance-based similarity measures and probabilistic similarity measures, even nodes have quantifiable properties [93]. Usually, topological characteristics cannot be captured by one or two measure indexes. Thus, some methods simultaneously make use of several similarity measures, and grasp topological properties from different perspectives, for example, SNMF [94] make use of various similarity measures to detect structural groups via semi-supervised strategy. However, the remarkable feature of most clustering methods depending on what kind of structures that are of interest and we must know in advance which properties define the groups we seek to identify, for example, community discovery relies on community groups nodes that are well-connected to each other in advance, social role discovery relies on social role groups nodes of similar behavior in advance. This is difficult to mine hidden structures. In fact, in the case of purely structural features, two exploratory methods have been devised, which can identify patterns not anticipated by pre-conceptions. One based on maximum likelihood techniques [89] and the other base on data project techniques, they both aim resolving the internal structure of complex networks by organizing the nodes into groups that share something in common, even if we do not know a priori what the thing is.

5.1 The maximum likelihood method

The maximum likelihood method understand the structure of social networks from the statistical inference perspective and able to detect a wide variety of structural groups and, crucially, does so without requiring us to specify in advance which particular structure we are looking for. Its basic idea is that gaining understanding of the structure of networks by fitting them to a statistical network model. A very related study has been proposed recently by M.E.J. Newman et al [95], they show that it is possible to detect, without prior knowledge of what we are looking for, a very broad range of types of structure in networks, using the machinery of probabilistic mixture models and the expectation-maximization algorithm, whose objective is to groups nodes with common connection features into a predefined number of groups. The idea of maximum likelihood method is similar to the blockmodel, although the realization and the mathematical techniques employed are different, or, more

precisely, is kind of variant of blockmodel. In principle, any type of structure that can be detection by maximum likelihood, including community structures, bipartite or disassortative structures, structural groups and many others. However, in real world application, an obvious drawback of the maximum likelihood techniques is that it is very time consuming, which will definitely fail to deal with the huge networks.

5.2 Node project method

Recently, a novel clustering method viewed close data points as linked networks nodes was proposed [96]. Inspired by this idea, the method of "node project" was proposed which viewed tightly interacted network nodes as close data points in a feature space, and then one can study networks or discover hidden structural groups form a data analysis perspective, and systematic and sophisticated data analysis tools will be a great convenience. More specifically, as for structural group discovery, using a given set of node features (such as centrality and degree) as the coordinates for each node in the multi-dimensional feature space, we identify structural groups as clusters of points in this feature space [14,97,98]. The Figure 16 illustrates the schematic of the node projection method.

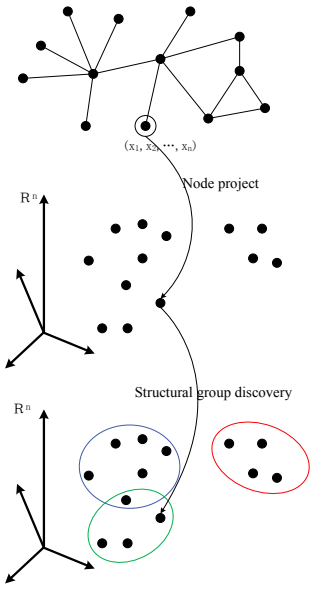


Fig. 16: Schematic illustration of the node projection method

According to Figure 16, the node project method has two key components: Node project and the structural group discovery in feature space.

A key question of node project is how to get the feature vector of nodes. The choice of node feature strictly depends on the application, since there is no defined rule to perform such a task. For instance, in the classification of highway networks into different models, one should consider features related to space and models with geographical constraints [99]. There are exists related work which exploits feature extraction from graphs for several data mining tasks. For example, W.Li et al extracted node features based on a signal spreading model [97], L da F Costa et al developed a classification approach which involves multivariate statistical methods and pattern recognition techniques to analyze complex networks based on local and global features extraction [98], and Keith Henderson et al proposed ReFeX(Recursive Feature eXtraction) [100], a novel algorithm, that recursively combines local features with neighborhood features, and outputs regional features-capturing "behavioral" information, and so on. Finding effective node features is the key step in all graph mining tasks, and is to continue to improve.

Many various methods were used to discover structural groups in multi-dimensional feature space by considering the principle: the closer the two nodes are in feature space, the more common properties they share. However, it is not easy to use high-dimensional clustering method for structural group detection, due to the correlation between features and the difficulty of their visualization. To overcome these limitations, it is necessary to use statistical methods for dimensionality reduction, such as component analysis (PCA) method and canonical variable analysis. These methods allow not only the elimination or, at least, reduction of the correlations between features but also the visualization of the observations into a reduced number of dimensions. In this way, although with a little loss of information, they still find effective structural groups [97,98]. In addition, to overcome the difficult of visualization of the observations into a high dimension feature space. Takashi Nishikawa et al proposed an approach based on visual analytics (called visual analytics method), which is conceptualized as exploratory statistics in which analytical reasoning is facilitated by a visual interactive interface [14]. The integration of the visual interactive interface allows the user not only to supervise the process, but also to learn and create intuition from raking parting in the process, thus facilitating the search for unanticipated network structures. And the results of applying this method to real networks suggest that it is capable of discovering not only group structures defined by link density, but also more general group structures, even when different types of structures coexist in the same network. For example, in Figure 17, although the teams are organized into 12 conferences (indicate 12 communities), the visual analytics method identifies 7 structural groups. The structural groups capture a higher-level organization of the conferences which is determined by the geographic proximity of the teams, which cannot be characterized by the community.

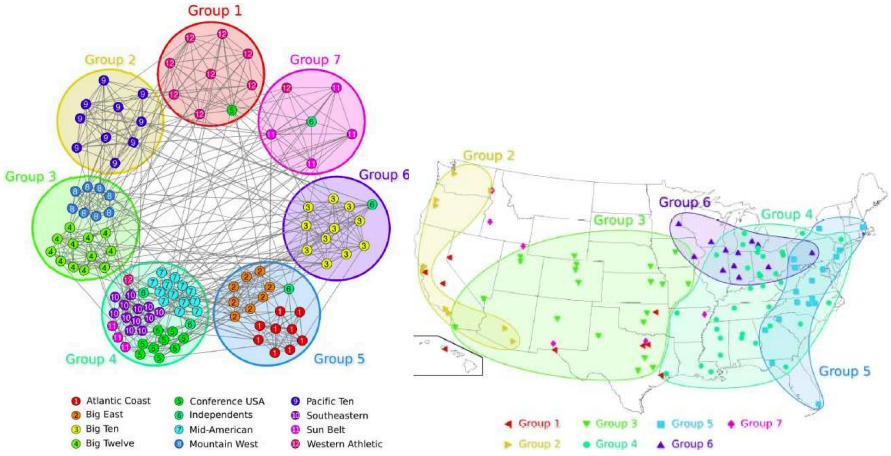


Fig. 17: [14] Characterizing seven structural groups discovered in the football network.(a)Layout of the network with the structural groups indicated by circles, color-coded as in the other panels. The number and color on a node indicate the college football conferences to which the corresponding team belongs. (b)Geographic distribution of nodes over the US, color-coded by the structural groups in panel (a).

In fact, structural group includes but not limited to community and role, because the densely-connected and behavior may be regarded as a kind of node property. But the application of structural group detection methods can detect hidden structures, which is an importantly scientific signification and potentially benefits, and has been attracting more and more attentions. Introducing the techniques from other fields to the study of structural groups may be helpful to further study, such as visual analytics method is a successful case.

6 Outlook

In this article, we briefly summarized the progress of studies on structure discovery, including community discovery, social role discovery and structural group discovery. And structure discovery are not limited to the social networks, but also widely used into the other real-world networks, such as web page clustering, protein function prediction and gene analysis, the field has appealed more and more attention from researchers across multiple domains and became even more flourished. In our opinion, Despite many new methods and tool have been presented, in others existing methods have been improved, becoming more accurate and faster, what motivation facilitate structure discovery development and technical innovation the most are the concrete applications. With the development of social media and the requirement of practical application, structure discovery encounters great challenges as well as opportunities. We will discuss a number of important open issues as follow.

Social media networks are often heterogeneous, having heterogeneous interactions or heterogeneous nodes. Heterogeneous networks are thus categorized into multi-dimensional networks, where it has multiple types of links between the same set of nodes, and multi-mode networks, where it involves heterogeneous nodes. However, the most existing structure discovery methods are constrained to homogeneous networks. Some work has been done to identify communities in a network of heterogeneous entities or relations [101,102,103] in terms of multitype relational clustering. Moreover, some methods are extend to handle this heterogeneity, for example, Lei Tang, et al. discuss potential extensions of community detection in one-dimensional networks to multi-dimensional networks, present a unified process, involving four components: network integration, utility integration, feature integration, and partition integration, to detect community in multi-dimensional networks [104], besides, they present an efficient and effective approach MROC to extract overlapping communities with different resolutions [105], and use an iterative latent semantic analysis process to capture evolving structures in multimode networks [106]. However, these methods only concentrate on multi-dimensional networks or multi-mode network and fail to reveal community of overlap and hierarchical. The most social role discovery methods can handle heterogeneity, such as RoleSim, RolX and simulation relation, as well as node project method can be applied in heterogeneous networks for discovering structural group, but their performance and effectiveness in heterogeneous networks require to be deeply discussed, especial for large-scale heterogeneous

networks. We are also expecting extended methods or new methods can also contribute to this domain.

As information technology has advanced, people are turning more frequently to electronic media for communication, and social relationships are increasingly found in online channels. The network presented in social media can be huge, often in a scale of millions of actors and hundreds of millions of connections, and it has exploded at a rapid pace. For example, Facebook claims to have more than 500 million active users as of August, 2010. While traditional structure discovery methods normally deals with hundreds of subjects or fewer, existing structure discovery methods might fail when applied directly to networks of this astronomical size, even if its runtime complexity is linear on the number of edges or nodes. Aim at this problem, one need to combine techniques for discovering structures, such as incremental algorithms and distributed algorithms. On the other hand, as the complete structure of the network is often unavailable since the entire network is too large and dynamic. We should try to explore structure from limited accessible region of a graph, for instance, many researchers have proposed several local methods that use partial knowledge of the network to discover the local community with a certain source vertex [107,108,109].

Up to now, it is no standard way or criterion to evaluate "good" or "bad" of structures. Although there are many typical approaches to determine which of community discovery method are the best, such as algorithms are tested against real-world networks and are tested against synthesized networks, for example, Mika Gustafsson et al compared and validated various classical community algorithms by using a class of computer generated networks and three well-studied real networks [110]. Unfortunately, it is still nothing reported about the evaluation criterion of the overlap and hierarchical quality, as well as no method or criterion to evaluate the role and structural group. In summary, the issue of evaluating the accuracy of structure methods discovery method in social networks is an important part.

Although social media present novel challenges for discovering structure, it also propose the amount of prior information (external information) of social networks, thus facilitating the research for discovering structure, such as the structure discovery methods, performance can be effectively enhanced by considering some prior information, like the attributes of nodes. Then, how to utilize the prior information to improve the structure methods have draw considerable attention in recent years. For example, based on semantic information extracted from user-comment content, ZhengYou Xia et al, proposed a useful method of discovering the latent communities, which can handle large-scale networks [111], Wenjun Zhou et al design a latent community model, called COCOMP(Collaborator COMmunity Profiling) [112], to uncover the communities of each user as well as their associated topics and communities by taking into account both the contacts and the topics, the experiments results demonstrate that the model can discover users, communities effectively, and provide concrete semantics, Besides, like PCB(belief propagation and conflict)method [113], RolX and visual analytics method are also give excellent results by using priori information, and so on. However, to design effective algorithms to discover structures by combing priori information and practical application, we need in-depth and comprehensive understanding of our application and priori information extraction.

There are various methods for structure discovery, but how to choose an appropriate method to discover structures in specific application is a key problem. We take friend recommendation as an example, it is obvious that individuals who share the same structure might be expected to share the same taste, interests, and so on. But what types of structure we can use to recommend friends, due to it has various patterns, such as community, role, group and so on. The individuals may have same interest when they are well-connected, i.e. in the same community, or have the same work when they have same behavior, i.e. in the same role, or they have the same taste when they are in the same city, just as in the same structural group. Thus, choosing an appropriate method for structure discovery could be of huge practical value, accounting for the specific application.

Furthermore, the study of structure discovery is a large and active field of endeavor, with new results appearing daily and an energetic community of researchers working on both methods and applications. Some of developments of structure discovery are of great importance not only in social network research, but also in biology, computer science, chemistry and so on.

Acknowledgement

This research work was supported by the National Natural Science Foundation of China under Grant No.61273322, 61201328 and Hunan Provincial Innovation Foundation For Postgraduate under Grant No. CX2013B024.

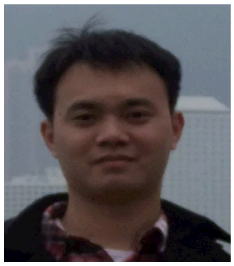
References

[1] S. Wasserman, K. Faust, Social Network Analysis, Cambridge Univ. Press, 1994.
[2] J. Scot, Social Network Analysis: A Handbook, 2nd edn, Sage, 2000.
[3] http://www.visualcomplexity.com/vc/.
[4] J. P. Bagrow, E. M. Bollt, J. D. Skufca, D. Ben-Avraham, Portraits of Complex Networks, Europhysics Letters, **81**, 68004 (2008).
[5] R. Milo, S. S. Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: Simple building blocks of complex networks, Science, **298**, 5594, 824-827 (2002).
[6] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, Proc. Natl Acad. Sci. USA, **99**, 7821-7826 (2002).

[7] D. J. Watts, S. H. Strogatz, Collective dynamics of small-world networks, Nature, **393**, 6684, 440-442 (1998).
[8] A. L. Barabasi, R. Albert, Emergence of scaling in random networks, Science, **286**, 5439, 509-512 (1999).
[9] Y. Bo, L. Jiming, L. Dayou, Characterizing and Extracting Multiplex Patterns in Complex Networks, IEEE Transaction on systems, man, and cybernetics-part B:cybernetics, **42(2)**, 469-481, (2012).
[10] R. Guimera, L. A. N. Amaral, Functional cartography of complex metabolic networks, Nature, **433**, 859-900 (2005).
[11] T. Lei, L. Huan. Community Detection and Mining in Social Media, MOGRAN&CLAYPOOL PUBLISHERS.
[12] H. Keith, G. Brian, E. R. Tina, T. Hanghang, RoIX: Structural Role Extraction & Mining in Large Graphs, The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Ming, KDD?12, Beijing, China, **2012**, 1231-1239 (2012).
[13] S. P. Borgatti, M. G. Everett, L. C. Freeman, Ucinet for Windows: Software for Social Network Analysis, 2002.
[14] N. Takashi, E. M. Adilson, Discovering Network Structure Beyond Communities Scientific reports, **1**, 151, 1-7 (2011).
[15] F. Santo, Community detection in graphs Physics Reports, **486**, 75-174 (2010).
[16] M. E. J. Newman, Communities, modules and large-scale structure in networks, nature physics, **8**, 25-31 (2012).
[17] G. Palla, I. Derenyi, I. Faraks, et al, Uncovering the overlapping community structure of complex networks in nature and society, Nature, **435**, 814-818 (2005).
[18] I. Farkas, D. Ábel, G. Palla, et al. Weighted network modules, New J Phys, **9**, 6, 180 (2007).
[19] I. Farkas, D. Ábel, G. Palla, et al. Directed network modules, New J Phys, **9**, 6, 186 (2007).
[20] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, Physical Review E, **69**, 026113 (2004).
[21] M. E. J. Newman, Analysis of weighted networks, PhysRev E, **70**, 056131 (2004).
[22] A. Arenas, J. Duch, A. Fernandez, et al, Community structure in directed networks, New J Phys, **9**, 176 (2007).
[23] M. E. J. Newman, E. A. Leicht, Community structure in directed networks, Proc Natl Acad Sci USA, **104**, 9564 (2007).
[24] S. Fortunato, M. Barthélemy, Resolution limit in community detection, Proc. Natl. Acad. Sci. USA, **104**, 36 (2007).
[25] L. Andrea and F. Santo, Limits of modularity maximization in community detection, Physical review E, **84**, 066122 (2011).
[26] X. Ju, H. Ke, Limitation of multi-resolution methods in community detection, Physica A, **391**, 4995-5003 (2012).
[27] B. Andrea, H. Pierre, L. Leo, Algorithm for parametric community detection in networks, Physical Review E, **86**, 016107 (2012).
[28] J. Reichardt, S. Bornholdt, Statistical mechanics of community detection, Physical Review E, **74**, 016110 (2006).
[29] A. Arenas, A. Fernández, S. Gómez, Analysis of the structure of complex networks at different resolution levels, New J. Phys, **10**, 053039 (2008).
[30] B. H. Good, Y. A. de Montjoye, A. Clauset, The performance of modularity maximization in practical contexts, Physical Review E, **81**, 046106 (2010).
[31] Z. Li, S. Zhang, R.S. Wang, X.S. Zhang, L. Chen, Quantitative Function for Community Detection, Physical Review E, **77**, 3, 036109 (2008).
[32] A. Lancichinetti, S. Fortunato, J. Kertesz, Detecting the overlapping and hierarchical community structure in complex networks, New J. Phys, **11**, 3, 033015 (2009).
[33] L. Hu, W. Hui, Detection of community structure in networks based on community coefficients, Physica A, **391**, 6156-6164 (2012).
[34] R. Aldecoa, I. Marin, Deciphering Network Community Structure by Surprise, PLoS ONE, **6**, e24195 (2011).
[35] R. Aldecoa, I. Marin, Surprise maximization reveals the community structure of complex networks, Sci.Rep., **3**, 1060 (2013).
[36] R. Aldecoa, I. Marin, Closed benchmarks for networks community structure characterization, Phys. Rev., **E85**, 026109 (2012).
[37] E. Ravasz, A. L. Somera, D. A. Mongru, et al., Hierarchical organization of modularity in metabolic networks, Science, **297**, 1551-1555 (2002).
[38] M. Xiaoke, G. Lin, Y. Xuerong, F. Lidong, Semi-supervised clustering algorithm for community structure detection in complex networks, Physica A, **389**, 187-197 (2010).
[39] A. Clauset, M. E. J. Newman, C. Moore, Fingding community structure in very large networks, Physical Review E, **70**, 6, 06611 (2004).
[40] A. Clauset, M. E. J. Newman, R. Lambiotte, et al, Fast unfolding of community hierarchies in large networks, J. of Statistical Mechanics, **10**, 10008 (2008).
[41] R. Guimerà, M. Sales-Pardo, L. A. N. Amaral, Modularity from fluctuations in random graphs and complex networks, Phys. Rev. E, **70**, 025101 (2004).
[42] J. Duch, A. Arenas, Community detection in complex networks using extremal optimization, Phys. Rev. E, **72**, 2, 027104 (2005).
[43] S. Ronghua, B. Jing, J. Licheng, J. Chao, Community detection based on modularity and an improved genetic algorithm, Physica A, **392**, 1215-1231 (2013).
[44] V. Nicosia, G. Mangioni, V. Carchiolo, et al, Extending the definition of modularity to directed graphs with overlapping communities, J Stat Mech, **3**, 03024 (2009).
[45] V. D. Blondel, J. L. Guillaume, R. Lambiotte, et al, Fast unfolding of community hierarchies in large networks, Journal of Statistical Mechanics: Theory and Experiment, **10**, 10008 (2008).
[46] S. Gregory, A fast algorithm to find overlapping communities in networks, Machine Learning and Knowledge Discovery in Databases, 408-423 (2008).
[47] A. Lancichinetti, S. Fortunto, J. Kertész, Detecting the overlapping and hierarchical community structure of complex networks, New J. Phys, **11**, 033015 (2009).
[48] W. Xiaohua, J. Licheng, W. Jianshe, Adjusting from disjoint to overlapping community detection of complex networks, Physica A, **388**, 5045-5056 (2009).
[49] Y. Bo, J. Di, L. Jiming, L. Dayou, Hierarchical community detection with application to real-world network analysis, Data&Knowledge Engineering, **83**, 20-38 (2013).
[50] A. Lancichinetti, F. Radicchi, J. J. Ramasco, S. Fortunato, Finding statistically significant communities in networks, PLoS ONE, **6**, e18961 (2011).

[51] J. Q. Jonathan, M. J. Lisa, Modularity functions maximization with nonnegative relaxation facilitates community detection in networks, Physica A, **391**, 854-865 (2012).
[52] A. Babak, H. Liaquat, C. W. John, W. T. Rolf, Multi-objective enhanced firefly algorithm for community detection in complex networks, Knowledge-Based Systems, **46**, 1-11 (2013).
[53] S. Chuan, Y. Zhenyu, C. Yanan, W. Bin, Multi-objective community detection in complex networks, Applied Soft Computing, **12**, 850-859 (2012).
[54] C. Pizzuti, A multi-objective genetic algorithm for community detection in networks, Proceedings of IEEE International Conference on Tools with Artificial Intelligence, (ICTA 2009), Newark, New Jersey, USA, 379-386 (2009).
[55] M. E. J. Newman, Netwroks: An Introduction, Oxford Univ. Press, 2010.
[56] M. E. J. Newman, Detecting community structure in networks, Eur. Phys. J. B, **38**, 2, 321 (2004).
[57] M. Greg, L. Mahadevan, Discovering Communities through Friendship, PLoS ONE, **7**, 7, e38704 (2012).
[58] S. Huawei, C Xueqi, C. Kai, et al, Detect overlapping and hierarchical community structure in networks, Physica A, **388**, 1706-1712 (2009).
[59] Y. Y. Ahn, J. P. Bagrow, S. Lehmann, Link communities reveal multiscale complexity in networks, Nature, **466**, 761-764 (2010).
[60] C. Qing, L. Zhong, H. Jincai, Z. Cheng, L. Yanjun, Hierarchical Clustering based on Hyper-edge Similarity for Community Detection, 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, Macau, China, 238-242 (2012).
[61] G. Mika, H. Michael, L. Anna, Comparison and validation of community structures in complex networks, Physica A, **367**, 559-576 (2006).
[62] W. Ruisheng, Z. Shihua, W. Yong, Z. Xiangsun, C. Luonan, Clustering complex networks and biological networks by nonnegative matrix factorization with various similarity measures, Neurocomputing, **72**, 134-141 (2008).
[63] E. R. Tina, F. Christos, Discovering Roles and Anomalies in Graphs: Theory and Applications, Proceedings of the 12th SIAM International Conference on Data Mining (SDM?12), Anaheim, California, USA, Tutorial (2012).
[64] J. Brynielsson, J. Hogberg, L. Kaati, C. Martenson, P. Svenson, Detecting social positions using simulation, 2010 International Conference on Advances in Social Networks Analysis and Mining International Conference on Advances in Social Networks Analysis and Mining, (ASONAM 2010), Odense, Denmark, 48-55 (2010).
[65] B. Joel, K. Lisa, S. Pontus, Social positions and simulation relations, Soc. Netw. Anal. Min, **2**, 39-52 (2012).
[66] H. A. Robert, R. Mark, Introduction to social network methods[EB/OL], http://faculty.ucr.edu/ hanneman/nettext.
[67] F. P. Lorrain, H.C. White, Structural equivalence of individuals in networks, J. Math. Sociology, **1**, 49-80 (1971).
[68] B. P. Stephen, E. G. Martin, Notions of position in social networks analysis, Sociological Methodology, **22**, 1-35 (1992).
[69] W. R. Douglas, R. P. Karl, Graph and semigroup homomorphisms on networks of relations, Social Networks, **5**, 193-234 (1983).
[70] R. S. Burt, W. Bittner, A note on inferences regarding networks subgroups, Social Networks, **3**, 1, 71-88 (1981).
[71] R. Breiger, S. Booman, P. Arabie, An algorithm for clustering relational data with applications to social network analysis and comparison with multi-dimensional scaling, Journal of Mathematical Psychology, **12**, 328-383 (1975).
[72] D. R. White, K. P. Reitz, Measuring role distance: Structural, regular and relational equivalence, Unpublished manuscript, University of California, Irvine (1985).
[73] V. Batagelj, P. Doreian, A. Ferligoj, An optimizational approach to regular equivalence, Social Networks, **14**, 121-135 (1992).
[74] E. Martin, B. Steve, Computing Regular Equivalence: Practical and Theoretical Issues. Metodoloski zvezki, **17**, 31-42 (2002).
[75] H. C. White, S. A. Boorman, R. L. Breiger, Social structure from multiple networks. I. Blockmodels of roles and positions, American Journal of Sociology, **81**, 730-778 (1976).
[76] J. Reichardt, D. R. White, Role models for complex networks, European Physical Journal B, **60**, 217-224 (2007).
[77] Z. Elena, N. Galileo, Stochastic Blockmodels: A Survey[EB/OL], http://www.cs.umd.edu/class/spring2008/cmsc828g/Slides/block-models.pdf.
[78] P. W. Holland, K. B. Laskey, S. Leinhardt, Stochastic Blockmodels: Some First Steps, Social Networks, **5**, 109-137 (1983).
[79] E. M. Airoldi, D. M. Blei, S. E. Fienberg, E. P. Xing, Mixed Membership Stochastic Blockmodels, Journal of Machine Learning Research, **9**, 1981-2014 (2008).
[80] K. Brian, M. E. J. Newman, Stochastic blockmodels and community structure in networks. Phys. Rev. E, **83**, 016107 (2011).
[81] S. Huawei, C. Xueqi, G. Jiafeng, Exploring the structural regularities in networks, Physical Review E, **84**, 056111 (2011).
[82] E. A. Leicht, Petter Holme, M. E. J. Newman, Vertex similarity in networks, Phys. Rev. E, **73**, 026120 (2005).
[83] J. Glen, W. Jennifer, Simrank: a measure of structural-context similarity, Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'02), Edmonton, Canada, 538-543 (2002).
[84] Z. Peixiang, H. Jiawei, S. Yizhou, P-rank: a comprehensive structural similarity measure over information networks, Proceedings of the 18th ACM conference on Information and knowledge management(CIKM'09), Hong Kong, China, 553-562 (2009).
[85] A. Ioannis, G. M. Hector, C. Chichao, Simrank++: query rewriting through link analysis of the click graph, PVLDB, **1**, 1, 408-421 (2008).
[86] L. Pei, C. Yuanzhe, L. Hongyan, H. Jun, D. Xiaoyong, Exploiting the block structure of link graph for efficient similarity computation, The 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2009), Bangkok, Thiland, 389-400 (2009).
[87] Y. Xiaoxin, H. Jiawei, Y. S. Philip, Linkclus: efficient clustering via heterogeneous semantic links, Proceedings of the 32nd international conference on Very large databases (VLDB'06), Seoul, Korea, 427-438 (2006).

[88] J. Ruoming, L. E. Victor, H. Hui, Axiomatic Ranking of Network Role Similarity, Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining(KDD'11), San Diego, California, USA, 21-24 (2011).
[89] R. Milner, Communication and Concurrency, Prentice Hall, 1989.
[90] M. Maarten, M. Michael, Regular equivalence and dynamic logic, Social Networks, **25**, 1, 51-65 (2003).
[91] S. Ma, Y. Cao, W. Fan, J. Huai, T. Wo, Capturing topology in graph pattern matching, PVLDB, **5**, 4, 310-321 (2011).
[92] W. Fan, J. Li, S. Ma, N. Tang, Y. Wu, Y. Wu, Graph pattern matching: From intractable to polynomial time, PVLDB, **3**, 1, 264-275 (2010).
[93] F. Gregory Ashby, E. M. Daniel, Similarity measures, Scholarpedia, **2**, 4116 (2007).
[94] M. Xiaoke, G. Lin, Y. Xuerong, F. Lidong, Semi-supervised clustering algorithm for community structure detection in complex networks, Physica A, **389**, 187-197 (2010).
[95] M. E. J. Newman, E. A. Leicht, Mixture models and exploratory analysis in networks, Proc. Natl. Acad. Sci., USA, **104**, 9564 (2007).
[96] F. J. Brendan, D. Delbert, Clustering by Passing Messages Between Data Points, Science, **315**, 972-976 (2007).
[97] W. Li, J. Y. Yang, W. C. Hadden, Analyzing complex networks from a data analysis viewpoint, Europhysics Letters, **88**, 6, 68007 (2009).
[98] L. da F. Costa, P. R. Villas Boas, F. N. Silva, F. A. Rodrigues, A pattern recognition approach to complex networks. Journal of Statistical Mechanics: Theory and Experiment, p11015 (2010).
[99] P. R. Villas Boas, et al , Modeling worldwide highway networks, Phys. Lett. A, **374**, 22, (2009).
[100] K. Henderson, B. Gallagher, L. Li, L. Akoglu, T. Eliassi-Rad, H. Tong, and C. Faloutsos, It's who you know: Graph mining using recursive structural features, Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'11), San Diego, California, USA, 663-671 (2011).
[101] J. Wang, H. Zeng, Z. Chen, H. Lu, L. Tao, W. Y. Ma, Recom:Reinforcement Clustering of Multi-Type Interrelated Data Objects, Proc. 26th Ann. Int'l ACM SIGIR Conf. Research andDevelopment in Informaion Retrieval (SIGIR '03), 274-281 (2003).
[102] B. Long, Z. M. Zhang, X. Wu, P. S. Yu, Spectral Clustering for Multi-Type Relational Data, Proc. 23rd Int'l Conf, Machine Learning (ICML '06), 585-592 (2006).
[103] B. Long, Z. M. Zhang, P. S. Yu, A Probabilistic Framework for Relational Clustering, Proc. 13th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '07), 470-479 (2007).
[104] T. Lei, W. Xufei, L. Huan, Community detection via heterogeneous interaction analysis, Data Min. Knowl. Disc., **25**, 1-33 (2012).
[105] W. Xufei, T. Lei, L. Huan, W. Lei, Learning with multi-resolution overlapping communities, Knowledge and Information Systems (KAIS), 1-19 (2012).
[106] T. Lei, L. Huang, Z. Jianping, Identifying Evolving Groups in Dynamic Multimode Networks, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, **24**, 72-85 (2012).
[107] J. Bagrow, E. Bollt, Local method for detecting communities, Physical Review E, **72**, 046108 (2005).
[108] A. Clauset, Finding local community structure in networks, Physical Review E, **72**, 026132 (2005).
[109] F. Luo, J. Wang, E. Promislow, Exploring local community structures in large networks, Web Intelligence and Agent Systems, **6**, 387-400 (2008).
[110] G. Mika, H. Michael, L. Anna, Comparison and validation of community structures in complex networks, Physica A, **367**, 559-576 (2006).
[111] X. Zhengyou, B. Zhan, Community detection based on a semantic network, Knowledge-Based Systems, **26**, 30-39 (2012).
[112] Z. Wenjun, J. Hongxia, L. Yan, Community Discovery and Profiling with Social Messages, Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'12), Beijing, China, 388-396 (2012).
[113] F. Xianghua, L. Liandong, W. Chao, Detection of community overlap according to belief propagation and conflict, Physica A, **392**, 941-952 (2013).



Qing Cheng Ph. D. student in Science and Technology on Information Systems Engineering Laboratory of the National University of Defense Technology. His research interests include information management, data mining and complex network.

Zhong Liu is a professor of National University of Defense Technology. His research interests are in information management and decision making support technology. He has published research articles in IEEE Intelligent Systems, IEEE Transactions on Intelligent Transportation Systems and IEEE Transaction on Systems, Man, and Cybernetics.

Jincai Huang is a professor of National University of Defense Technology. He is a visiting professor of the University of Edinburgh and his current research interests are in image processing, data mining and information management. He has published research articles in International Journal of Electronics and Communications.

Guangquan Cheng is a Lecturer of National University of Defense Technology. His current research interests are in image processing and data mining.