# A Vertex Similarity Probability Model for Finding Network Community Structure

Kan Li[1] and Yin Pang[1,2]

[1] Beijing Key Lab of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, 100081, China
likan@bit.edu.cn
[2] Beijing Institute of Tracking and Telecommunication Technology, Beijing, 100094, China
pangyin@bit.edu.cn

**Abstract.** Most methods for finding community structure are based on the prior knowledge of network structure type. These methods grouped the communities only when known network is unipartite or bipartite. This paper presents a vertex similarity probability (VSP) model which can find community structure without priori knowledge of network structure type. Vertex similarity, which assumes that, for any type of network structures, vertices in the same community have similar properties. In the VSP model, "Common neighbor index" is used to measure the vertex similarity probability, as it has been proved to be an effective index for vertex similarity. We apply the algorithm to real-world network data. The results show that the VSP model is uniform for both unipartite networks and bipartite networks, and it is able to find the community structure successfully without the use of the network structure type.

**Keywords:** community structure, type of the network structure, vertex similarity, common neighbor index.

## 1 Introduction

As part of the recent surge of research on large, complex networks, attention has been devoted to the computational analysis of complex networks [1-4]. Complex networks, such as social networks and biological networks, are all highly dynamic objects which grow and change quickly over time. These networks have a common feature, namely "community structure". Communities, also known as clusters or modules, are groups of vertices which could share common properties and/or have similar roles within the graph [5]. Finding community structure and clustering vertices in the complex network, is key to learning a complex network topology, to understanding complex network functions, to founding hidden mode, to link prediction, and to evolution detection. Through the analysis of community structure, researchers have achieved a lot results, such as in [6, 7], V. Spirin *et al*. revealed the relationship between protein function and interactions inherent; in [8, 9], Flake *et al.* found the internal relations of hyperlink and the main page; in [10, 11], Moody *et al*. identified the social organizations to evolve over time and so on.

P.-N. Tan et al. (Eds.): PAKDD 2012, Part I, LNAI 7301, pp. 456–467, 2012.

The most popular method for finding community structure is the modularity matrix method [12, 13] proposed by Newman *et al.* which is based on spectral clustering. The Modularity model proves that, if the type of the network structure is known, modularity optimization is able to find community structure in both unipartite and bipartite networks by the maximum or minimum eigenvalue separately. Then, some scientists have sought to detect the community in bipartite networks like Michael J. Barber [14]. BRIM proposed by Barber and his colleagues can determine the number of communities of a bipartite network. Furthermore, in [15], Barber and Clark use the label-propagation algorithm (LPA) for identifying network communities. However, [14, 15] can not be used without knowing the type of network.

There are other methods to find community structure. Hierarchical clustering is adopted frequently in finding community structures, in which vertices are grouped into communities that further are subdivided into smaller communities, and so forth, as in [12]. Clauset, Moore and Newman propose HRG [16] using the maximum likelihood estimation to forecast the probability of connections between vertices. Hierarchical methods perform remarkably in clear hierarchy network, but not so impressive under contrary circumstance. Moreover, a hierarchical method always has high computational complexity. In 2009, Roger Guimera and Marta Sales-Pardo proposed a stochastic block model [17] based on HRG. Different from traditional concept which divide network by principle of "inside connection dense outside sparse", in [17], the probability that two vertices are connected depends on the blocks to which they belong. However, the assumption that vertices in same blocks have same connection probability is not accurate. Recently, Karrer and Newman [18] also proposed a stochastic block model which considers the variation in vertex degree. This stochastic block model solves the heterogeneous vertex degrees problem and got a better result than other previous researches without degree correction. It can be used in both types of networks, but different types of networks should be dealt with separately none the less.

In some cases, researchers have no priori knowledge of the network structure. For example, when we know the interaction of vertex in the protein network, we may have no knowledge of the network structure type. Moreover, when we get a network which consists of people's relationships in schools, the type of network may not be sure. It is because that if links only exist between students, the network will be a unipartite network; or if links exist between students and teachers, the network will be a bipartite one. An effective method used for finding community structure in both unipartite and bipartite networks is needed.

It is discussed before that most methods deal with the unipartite network or bipartite network separately, because the properties of networks are different in different types of the network structure. Unipartite networks assume that connections between the vertices in same community are dense, and between the communities are sparse, such as Social network [19], biochemical network [20] and information network [21]. However, some real networks are bipartite with edges joining only vertices of different communities, such as shopping networks [22], protein-protein interaction networks [23], plant-animal mutualistic networks [24], scientific publication networks [25], etc. Although the properties of "edges" in the two types of networks are different, vertices in the same communities should be similar because vertices in same

communities have similar properties. In this paper, we develop a uniform VSP model which is based on the vertex similarity. Therefore, the VSP model can be used in any type of networks as long as we put similar vertices in same communities. The VSP model gets ideal result both by theoretical proof and experimental analysis.

The paper is organized as follows. In section 2, we prove vertex similarity theory is suitable for finding community structure. We present the VSP model and the method to group network into two communities in section 3. In section 4, we make the experiment in both unipartite and bipartite network. Compared with Newman's modularity, the VSP model is an accurate uniform model which can find community structure without prior knowledge of type of the network structure. Finally, we draw our conclusions.

## 2 Vertex Similarity in Finding Community Structure

The concept of community informs that vertices in the same community should share common properties no matter in unipartite or bipartite network. It means that vertices in the same community should be similar, although edges in different type of the network structures are connected in different ways. Therefore, we change our focus from "edges" to "vertices" for finding communities.

Vertex similarity is widely studied by researchers in complex network. It is sometimes called structural similarity, to distinguish it from social similarity, textual similarity, or other similarity types. It is a basic premise of research on networks that the structure of a network reflects real information about the vertices the network connects, so it is reasonable that meaningful structural similarity measures might exist [26]. In general, if two vertices have a number of common neighbors, we believe that these two vertices are similar. In community detection, we assume that two similar vertices have similar properties and should be grouped in the same community.

Let $\Gamma_x$ be the neighborhood of vertex $x$ in a network, i.e., the set of vertices that are directly connected to $x$ via an edge. Then $|\Gamma_x \cap \Gamma_y|$ is the number of common neighbors of $x$ and $y$. Common neighbor index, Salton index, Jaccard index, Sorenson index, LHN (Leicht-Holme-Newman) index, and Adamic-Adar index [27-31] are five famous methods for vertex similarity. Many researchers have analyzed and compared these methods. Liben-Nowell[32] and Zhou Tao[33] proved that the simplest measurement "common neighbor index" performs surprisingly well. We use "common neighbor index" to measure the vertex similarity in our VSP model.

Definition 1. For two vertices $x$ and $y$, if there is a vertex $z$ to be the neighbor of $x$ and $y$ at the same time, we call $x$ and $y$ a pair, denoted as pair($x$, $y$). $z$ is called the common neighbor of pair($x$, $y$).

Since vertices which are in the same community have similar properties, we assume vertices in the same community are similar vertices. The more similar the vertices inside a community are the more common neighbors they have. The number of common neighbors $N_{ij}$ of vertices $i$ and $j$ is given by,

$$N_{ij} = \left|\Gamma_i \cap \Gamma_j\right|, \text{ and } N_{ii} = 0 .$$

The sum of common neighbors with vertices in same communities $N_{in}$ is given by

$$N_{in} = \sum_{\substack{i,j \in \text{same} \\ \text{community}}} N_{ij} \cdot$$

And the sum of common neighbors with vertices in different communities $N_{out}$ is given by

$$N_{out} = \sum_{\substack{i,j \notin \text{same} \\ \text{community}}} N_{ij} \cdot$$

Therefore, the task of maximizing the number of common neighbors in the same community is to get $\max(N_{in})$ or to get $\min(N_{out})$. The sum of common neighbors in the network *R* is given by

$$R = \frac{1}{2} \sum_{i,j \in n} N_{ij} \cdot$$

We define the adjacency matrix *A* to be the symmetric matrix with elements $A_{ij}$. If there is an edge joining vertices *i* and *j*, $A_{ij} = 1$; if no, $A_{ij} = 0$. Define $\mathbf{a_i}$ as *i* th vector of *A*, so as *A* can be rewritten as $A = [\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_n}]$. If and only if $A_{ik} A_{kj} = 1$, the vertex k is a common neighbor of vertices *i* and *j*. Therefore $N_{ij}$ can be rewritten as

$$N_{ij} = \sum_k A_{ik} A_{kj} = \mathbf{a_i} \cdot \mathbf{a_j},$$

when *i* and *j* are two different vertices. As $\mathbf{a_i} \cdot \mathbf{a_i} = k_i$, matrix *N* is

$$N = A^T A - \Lambda_k,$$

where $\Lambda_k = diag(k_1, k_2, ..., k_n)$. It allows us to rewrite *R* as

$$\begin{aligned} R &= \frac{1}{2} \sum_{\substack{i,j \in n \\ i \neq j}} \mathbf{a_i} \cdot \mathbf{a_j} \\ &= \frac{1}{2} \left( \sum_{i,j \in n} \mathbf{a_i} \cdot \mathbf{a_j} - \sum_i k_i \right) \\ &= \frac{1}{2} \left( \sum_i \mathbf{a_i} \cdot (k_1, k_2, ..., k_n)^T - \sum_i k_i \right) \\ &= \frac{1}{2} \left( (k_1, k_2, ..., k_n) \cdot (k_1, k_2, ..., k_n)^T - \sum_i k_i \right) \\ &= \sum_i \frac{1}{2} k_i (k_i - 1) \end{aligned} \tag{1}$$

Definition 2. According to Eq.(1), *R* is only related to a function of vertex degree. To analyze the relationship between a vertex x and common neighbor index, we define the function as a common neighbor degree index, denoted as $c_x$. Let $c_x = k_x(k_x - 1)/2$. Therefore, $R = \sum_{x \in n} c_x$.

Total number of common neighbors in the network equals the number of common neighbors in same communities plus the number of common neighbors different communities, $R$ also can be written as $R = N_{in} + N_{out}$.

The following proves that using common neighbor index in finding community structure is suitable in both unipartite networks and bipartite networks.

### 2.1 Common Neighbor Index in Unipartite Network

For a unipartite network, the basic community detection principle is "edges inside communities are dense, outside are sparse". Let the sum of edges with vertices in different communities is $A_{out}$, where

$$A_{out} = \sum_{\substack{i,j \notin \text{same} \\ \text{community}}} A_{ij} .$$

The task is to minimize $A_{out}$, written as $\min(A_{out})$.

Suppose $i$ and $j$ are two vertices in different communities. If $i$ and $j$ are connected, there are $k_i - 1$ pairs (where $k_i$ is the degree of $i$) with a common neighbor $i$, each of which is formed by $j$ and a neighbor of $i$. In a unipartite network, neighbors of a vertex are almost in the same community. As a result, for $i$, most of its neighbor should be in the same community with $i$ except $j$ (if $j$ is a neighbor of $i$). As shown in Fig. 1.

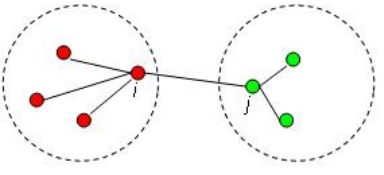


**Fig. 1.** An example of two vertices in different communities in a unipartite network

If $i$ and $j$ are not connected, no common neighbor is counted. Therefore the number of common neighbors with pairs of vertices in different communities is

$$N_{out} = \sum_{\substack{i,j \notin \text{same} \\ \text{community}}} A_{ij}(k_i - 1) \tag{2}$$

$A$ is symmetric which allows us to rewrite Eq.(2) as

$$N_{out} = \sum_{\substack{i,j \notin \text{same} \\ \text{community} \\ \text{and i<j}}} A_{ij}(k_i + k_j - 2) \tag{3}$$

For two vertices $i$ and $j$ in different communities, $N_{out}$ is related to $A_{ij}$, $k_i$ and $k_j$. If there is an edge between $i$ and $j$, $A_{out}$ will plus 1 and $N_{out}$ will plus $k_i + k_j - 2$. As $k_i + k_j - 2 \geq 0$, we consider $A_{out}$ and $N_{out}$ have the same growth trend. It means getting $\min(N_{out})$ is equivalent to getting $\min(A_{out})$. The conclusion is in line with the basic principles of the unipartite network community detection.

### 2.2 Common Neighbor Index in Bipartite Network

For a bipartite network, the basic community detection principle is "edges inside communities are sparse, outside are dense". The task is to maximize $A_{out}$, written as $\max(A_{out})$.

In a bipartite network, almost all adjacent vertices are in different communities. For a pair of vertices which are in the same community, the common neighbor should be in a different community, as shown in Fig. 2.

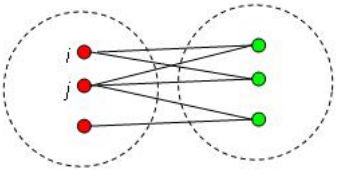


**Fig. 2.** An example of two vertices in same communities in a bipartite network

As a result, for any pair of vertices *i* and *j* which are in the same community, $N_{ij}$ have $2N_{ij}$ edges between different communities. In the overall network, each edge will be counted $(k_i + k_j - 2)$ times.

$$2N_{in} = \sum_{\substack{i,j \in \text{same} \\ \text{community}}} A_{ij}(k_i + k_j - 2) \cdot \tag{4}$$

*A* is symmetric which allows us to rewrite Eq.(4) as

$$N_{in} = \sum_{\substack{i,j \in \text{same} \\ \text{community} \\ \text{and i<j}}} A_{ij}(k_i + k_j - 2) \tag{5}$$

Similar as section 2.1, we consider $A_{out}$ and $N_{in}$ have same growth trend. It means getting $\max(N_{in})$ is equivalent to getting $\max(A_{out})$. The conclusion is in line with the basic principles of the bipartite network community detection.

In summary of section 2.1 and 2.2, the common neighbor index of vertex similarity is suitable for finding community structure in both unipartite and bipartite networks.

## 3 A VSP Model for Finding Community Structure

In this section, we propose our VSP model to find community structure. In [13], Newman *et al.* proved that a good division of a network in to communities "in which the number of edges inside groups is bigger than expected". It can get a better result than the measures based on pure numbers of edges between communities. Similarly, a good division of a network into communities should be one which the number of common neighbors within communities is bigger than expected. Let

$$\text{Q=(common neighbors within communities- expected number of such common neighbors)} \cdot \tag{6}$$

It is a function that divides the network into groups, with larger values indicating stronger community structure. We build a random network in which vertices have same common neighbor degrees as the vertices in the complex network, and assume the expected number of common neighbors as the number in the random network. In section 2, we have proved that common neighbor index can be used to find communities instead of edges. However, we can also find that $N_{out}$ in unipartite network and $N_{in}$ in bipartite network are both affected not only by edges but also by vertex degree. It is known that common neighbor degree $c_i$ is a function of $k_i$ and R is the sum of $c_i$. We use $c_i$ to calculate the common neighbors in the random network. The probability of a random vertex to be a common neighbor of a particular vertex $i$ depends only on the expected common neighbor degree $c_i$. The probabilities of a random vertex to be a common neighbor of two vertices are independent on each other. This implies that the expected number of common neighbors $P_{ij}$ between vertices $i$ and $j$ is the product $f(c_i)f(c_j)$ of separate functions of the two common neighbor degrees, where the functions must be the same since $P_{ij}$ is symmetric. Hence $f(c_i)=Cc_i$ for some constant C,

$$\sum_{i,j\in n} P_{ij} = \sum_i f(c_i) \sum_j f(c_j) = C^2 R^2 \tag{7}$$

Vertices in random network have the same common neighbor degree just like in complex network, $\sum_{i,j\in n} P_{ij} = \sum_{i,j\in n} N_{ij} = 2R$. So, $C = \sqrt{\frac{2}{R}}$ and

$$f(c_i) = \sqrt{\frac{2}{R}} c_i \tag{8}$$

We get the expected number of common neighbors of pair($x$, $y$) as follows,

$$P_{ij} = f(c_i) f(c_j) = \frac{2c_x c_y}{R} \cdot \tag{9}$$

The VSP model can be written,

$$Q = \frac{1}{2R} \sum_{\substack{i,j\in \text{same} \\ \text{community}}} [N_{ij} - \frac{2c_i c_j}{R}] \cdot \tag{10}$$

What we should notice is that,

$$\sum_{i,j\in n} \frac{2c_i c_j}{R} = 2\frac{\sum_{j\in n} c_j \sum_{i\in n} c_i}{R} = 2R = \sum_{i,j\in n} N_{ij} \cdot \tag{11}$$

Thus,

$$\frac{1}{2}\sum_{i,j\in n}[N_{ij}-\frac{2c_ic_j}{R}]=0 \tag{12}$$

Let

$$B_{ij}=N_{ij}-\frac{2c_ic_j}{R}\cdot \tag{13}$$

$B$ is the VSP matrix, and $\sum_{i,j\in n}B_{ij}=0$.

We use the VSP matrix instead of modularity matrix to find the community structure. In the VSP model, the higher value of $Q$, the more similar vertices are in the same community. It can be applied to both unipartite networks and bipartite networks without knowing the exact type of network structure in advance. It is more flexible than the previous methods which deal with the grouping separately according to the type of the network structure.

## 4 Experimental Results

In this section, we apply the VSP model to a unipartite network and two bipartite networks with Pajek [34]. The unipartite network shows the dolphin social network studied by Lusseau *et al*. [35]. The bipartite networks show the interactions of women in the American Deep South at various social events [36] and Scotland Corporate Interlock in early twentieth century [37].

Since we know the actual communities for the real networks, we measure the accuracy of the VSP model by directly comparing with the known communities. We take use of the normalized mutual information $I_{normn}$ [38] for the comparison. When the found communities match the real ones, we have $I_{norm}$=1, and when they are independent of the real ones, we have $I_{norm}$=0.

We compare the VSP model with the Modularity model in unipartite networks and bipartite networks by three properties: the edges outside communities; $Q$ of the Modularity model, where $Q$ is the edges within communities minus expected number of such edges, written as *Q-Modularity*; and $I_{norm}$.

### 4.1 Finding Community Structure in Unipartite Network

$Q$ in Eq. (6) is written as *Q-VSP* in this section. For a unipartite network, the VSP model maximizes *Q-VSP* to find the community structure, while the Modularity model maximizes *Q-Modularity*.

The dolphin social network is a classical unipartite social network. The vertices in this network represent 62 bottlenose dolphins living in Doubtful Sound, New Zealand, with social ties between dolphin pairs established by direct observation over a period of several years. It is used a lot in community detection because the dolphin group split into two smaller subgroups following the departure of the population. Fig.3 shows the clustering results using the VSP model and the Modularity model respectively.

Two red vertices are grouped into the green community in the VSP model, while three red vertices are grouped into the green community in the Modularity model.

In a unipartite network, edges inside the communities are dense, while outside are sparse. Edges outside communities should be small; *Q-Modularity* should be large; $I_{norm}$ close to 1. Properties of the dolphin social network are shown in Table.1. It shows that two properties of VSP model are better than the Modularity model in this unipartite network. *Q-Modularity* of the VSP model is 0.381 which is approximately equal to the one of the Modularity model. The VSP model performs well in unipartite networks.

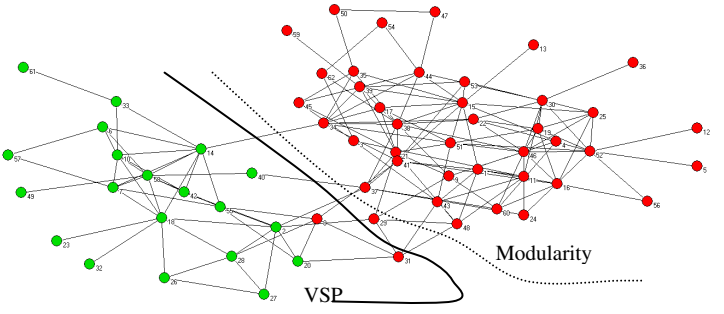


**Fig. 3.** Finding community structures of the dolphin social network. The red and green vertices represent the division of the network. The solid curve represents the division of the VSP model.. The dotted curve represents the division of the Modularity model.

**Table 1.** Properties of the dolphins social network

| | Edges outside communities | Q-Modularity | $I_{norm}$ |
|---|---|---|---|
| VSP | 8 | 0.381 | 0.813 |
| Modularity | 9 | 0.386 | 0.752 |

## 4.2 Finding Community Structure in Bipartite Network

In a bipartite network, the VSP model also maximizes *Q-VSP* to find the community structure, the Modularity model minimizes *Q-Modularity*, which is contrary to in the unipartite network.

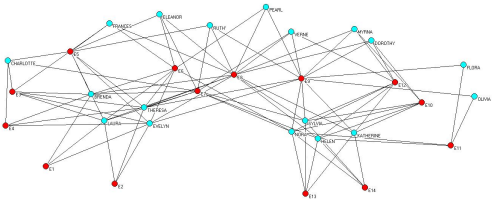

**Fig. 4.** Find community structures of Southern women network using the VSP model

The Southern women data set describes the grouping of 18 women in 14 social events constitute a bipartite network; and an edge exists between a woman and a social event if the woman was in attendance at the event. We use this network here to group it into two communities, shown in Fig.4. It shows that the VSP model groups the network accurately into two communities of "women" and "events".

Although using other finding community structure methods can also get the same result, they should know the type of the network in advance. For example, in modularity, it gets the smallest value of the Modularity model but not the biggest because the southern women network is a bipartite network.

As a second example of bipartite network, we consider a data set on Scotland Corporate Interlock in early twentieth century. The data set is characterized by 108 Scottish firms, detailing the corporate sector, capital, and board of directors for each firm. The data set includes only those board members who held multiple directorships, totaling 136 individuals. Unlike the Southern women network, the Scotland corporate interlock is not connected. We got the division of one community with 102 vertices and the other with 142 vertices when *Q-VSP* is the maximum value.

We also compare three properties of the VSP model and the Modularity model. In a bipartite network, edges inside the communities are sparse, while outside are dense. Edges outside communities should be large; *Q-Modularity* should be small; $I_{norm}$ close to 1. Properties of the Scotland corporate interlock are shown in Table 2. All the three properties of the VSP model are better than the Modularity model. It proves that the VSP model also finds community structure accurately in bipartite networks.

**Table 2.** Properties of Scotland corporate interlock network

| | Edges inside communities | Q-Modularity | $I_{norm}$ |
|---|---|---|---|
| VSP | 47 | -0.372 | 0.767 |
| Modularity | 150 | -0.169 | 0.377 |

In summary of section 4.1 and 4.2, the VSP model is a uniform model for finding community structure which can be used in both unipartite networks and bipartite networks. It is flexible and applicable to a wide range. For instance, in the protein network which people only knows its tip of iceberg, the VSP model can find the community structure only with the topology of the network, even when we have no idea of the type of the network structure.

## 5 Conclusion

In this paper, we define a VSP model for finding the community structure in complex networks. The VSP model is based on the vertex similarity using the common neighbor index. As common neighbor index is proved an effective measurement of the vertex similarity methods in complex network, it is applied to the VSP model to measure the vertex similarity. We prove that calculating the common neighbor inside communities of the network is equivalent to calculation the least edges outside communities in a unipartite network and the most edges outside communities in a bipartite network.

Therefore, it is suitable for finding community structure in both unipartite and bipartite network. Then we give the expectation of the common neighbor between any two vertices and gave the VSP model. At last, we apply our model in the dolphin social network, Southern women event network and Scotland corporate interlock network separately. Results showed that the VSP model is effective for finding community structure without the need of the network structure type.

**Acknowledgments.** The research was supported by Natural Science Foundation of China (No.60903071).

## References

1. Wasserman, S., Faust, K.: Social Networks Analysis. Cambridge University Pres, Cambridge (1994)
2. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U.: Complex networks: Structure and dynamics. Physics Reports-Review Section of Physics Letters 424, 175–308 (2006)
3. Lu, L.Y., Zhou, T.: Link prediction in complex networks: A survey. Physica a-Statistical Mechanics and Its Applications 390, 1150–1170 (2011)
4. Boguna, M., Krioukov, D., Claffy, K.C.: Navigability of complex networks. Nature Physics 5, 74–80 (2009)
5. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17, 734–749 (2005)
6. Spirin, V., Mirny, L.A.: Protein complexes and functional modules in molecular networks. Proceedings of the National Academy of Sciences of the United States of America 100, 12123–12128 (2003)
7. Chen, J.C., Yuan, B.: Detecting functional modules in the yeast protein-protein interaction network. Bioinformatics 22, 2283–2290 (2006)
8. Flake, G.W., Lawrence, S., Giles, C.L., Coetzee, F.M.: Self-organization and identification of web communities. Computer 35, 66–71 (2002)
9. Dourisboure, Y., Geraci, F., Pellegrini, M.: Extraction and classification of dense communities in the web. In: Proceedings of the 16th International Conference on the World Wide Web, pp. 461–470. ACM, New York (2007)
10. Moody, J., White, D.R.: Structural cohesion and embeddedness: A hierarchical concept of social groups. American Sociological Review 68, 103–127 (2003)
11. Wellman, B.: The development of social network analysis: A study in the sociology of science. Contemporary Sociology-a Journal of Reviews 37, 221–222 (2008)
12. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E 69 (2004)
13. Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. Physical Review E 74 (2006)
14. Barber, M.J.: Modularity and community detection in bipartite networks. Physical Review E 76 (2007)
15. Barber, M.J., Clark, J.W.: Detecting network communities by propagating labels under constraints. Physical Review E 80 (2009)
16. Clauset, A., Moore, C., Newman, M.E.J.: Hierarchical structure and the prediction of missing links in networks. Nature 453, 98–101 (2008)

17. Guimera, R., Sales-Pardo, M.: Missing and spurious interactions and the reconstruction of complex networks. Proceedings of the National Academy of Sciences of the United States of America 106, 22073–22078 (2009)
18. Karrer, B., Newman, M.E.J.: Stochastic blockmodels and community structure in networks. Physical Review E 83 (2011)
19. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America 99, 7821–7826 (2002)
20. Holme, P., Huss, M., Jeong, H.W.: Subnetwork hierarchies of biochemical pathways. Bioinformatics 19, 532–538 (2003)
21. Rosvall, M., Bergstrom, C.T.: An information-theoretic framework for resolving community structure in complex networks. Proceedings of the National Academy of Sciences of the United States of America 104, 7327–7331 (2007)
22. Chuma, J., Molyneux, C.: Coping with the costs of illness: The role of shops and shopkeepers as social networks in a low-income community in coastal kenya. Journal of International Development 21, 252–270 (2009)
23. Li, F., Long, T., Lu, Y., Quyang, Q., Tang, C.: The yeast cell-cycle network is robustly designed. PNAS 101(14), 4781–4786 (2004)
24. Bascompte, J., Jordano, P., Melian, C.J., Olesen, J.M.: The nested assembly of plant-animal mutualistic networks. Proceedings of the National Academy of Sciences of the United States of America 100, 9383–9387 (2003)
25. Guimera, R., Uzzi, B., Spiro, J., Amaral, L.A.N.: Team assembly mechanisms determine collaboration network structure and team performance. Science 308, 697–702 (2005)
26. Leicht, E.A., Holme, P., Newman, M.E.J.: Vertex similarity in networks. Physical Review E 73 (2006)
27. Salton, G., McGill, M.J.: Introduction to modern information retrieval. MuGraw-Hill, Auckland (1983)
28. Jaccard, P.: Nouvelles recherches sur la distribution florale. Bulletin de la Societe Vaudoise des Science Naturelles 44, 223–270 (1908)
29. Sφrensen, T.: A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. Det Kongelige Danske Videnskabernes Selskab. Biologiske Skrifter 5(4), 1–34 (1948)
30. Salton, G.: Automatic text processing: The transformation, analysis, and retrieval of information by computer. Addison-Wesley, Boston (1989)
31. Adamic, L.A., Adar, E.: Friends and neighbors on the Web. Social Networks 25, 211–230 (2003)
32. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. Journal of the American Society for Information Science and Technology 58, 1019–1031 (2007)
33. Zhou, T., Lu, L.Y., Zhang, Y.C.: Predicting missing links via local information. European Physical Journal B 71, 623–630 (2009)
34. Pajek: http://vlado.fmf.uni-lj.si/pub/networks/pajek
35. 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? Behavioral Ecology and Sociobiology 54, 396–405 (2003)
36. Davis, A., Gardner, B.B., Gardner, M.R.: Deep South. University of Chicago Press (1941)
37. Scott, J., Hughes, M.: The anatomy of Scottish capital: Scottish companies and Scottish capital. Croom Helm, London (1980)
38. Danon, L., Diaz-Guilera, A., Arenas, A.: The effect of size heterogeneity on community identification in complex networks. Journal of Statistical Mechanics-Theory and Experiment, P11010 (2006)