带偏置的信号传播的随机游走的社团检测算法

尹欣红 赵世燕 陈晓云
(兰州大学信息科学与工程学院 兰州 730000)

摘 要 复杂网络是从大量现实存在的复杂系统中抽象得到的,网络的整体功能体现在网络中节点间的相互作用上,社团结构是其关键性结构特征。社团对应于系统的功能模块,提取网络的功能模块有助于深层探究复杂网络的内部规律,从复杂网络中检测社团结构具有重要的理论研究意义和实用价值。因此,很多研究者对社团检测进行了研究,进而提出了很多社团检测算法,如基于模块度优化的社团检测算法、基于标签传播的社团检测算法、基于随机游走的社团检测算法等。在对这些算法进行充分研究的基础上,通过模拟随机游走的过程,结合信号传播过程中随着传播距离的增大,信号量会缓慢衰减的思想,提出了一种带偏置的信号传播机制的随机游走的社团检测算法。该算法从网络中选取一个节点作为信号源,随机选择与其相邻的节点作为下一跳节点,将衰减后的信号量传递到该节点,依次迭代并传递信号。考虑到信号的衰减,为每条边设置偏置,对信号传播过程进行限定。通过模拟信号的传播,将网络的每个顶点作为信号源来重复这一过程,得到传播矩阵。然后,为每个顶点添加自环,并结合邻接矩阵以及顶点间的相似性,形成具有新属性的相似性矩阵。根据新属性矩阵和传播矩阵为每个顶点构造属性。最后,使用 k-means 算法进行聚类,得到高质量的社团结构。为了验证该方法的性能,在 10 个实际网络数据集以及不同规模的人工合成网络上进行实验。实验结果充分证明,所提算法能够从网络中提取出高质量的社团结构,从而有效地为社团检测领域提供依据。

关键词 社团检测,偏置,随机游走,信号传播,社团结构
中图法分类号 TP311 **文献标识码** A **DOI** 10.11896/jsjkx.190700051

Community Detection Algorithm Based on Random Walk of Signal Propagation with Bias

YIN Xin-hong ZHAO Shi-yan CHEN Xiao-yun
(School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China)

Abstract Complex networks are abstracted from various complex systems. The overall function is reflected in the interaction among nodes, and community structure is one of the most significant structural properties presented in many networks. Generally, the community corresponds to the functional modules of the system. Therefore, extracting these communities of the network helps us to explore the internal rules, and it has important theoretical research significance and practical value for community detection of complex networks. As a result, it is paid attention widely by many researchers, and many community-detection algorithms are proposed, such as the algorithms based on modularity optimization, label propagation, and random walk. In the process of signal propagation, as the propagation distance increases, the signal quantity will decay slowly. On the basis of the full study of these algorithms, by simulating the process of random walk, a community detection algorithm with random walk based on the signal propagation mechanism with bias was proposed. The algorithm selects a node from the network as the signal source, chooses the neighbor node randomly as the next hop node, transmits the attenuated semaphore to the node, and iteratively selects the next hop node and transmits the signal randomly. Considering the attenuation of the signal, an attenuation factor is attached to each edge to constrain the signal propagation process. Through the propagation of the analog signal, each process of the network is repeated as a signal source to obtain a propagation matrix. Then, the self-loop is added for each vertex. By considering the similarity matrix with new attributes between the adjacency matrix and the similarity among vertices, attributes for each vertex are constructed based on the new attribute matrix and propagation matrix. Finally, k-means algorithm is used for clustering to obtain high-quality community structure. In the end, k-means algorithm is used for clustering to obtain high-quality community structure with the minimum cost. In order to verify the performance of this method, this paper conducted experiments on 10 actual network data sets and artificial synthetic networks of different sizes. The experimental results

到稿日期:2019-05-05 返修日期:2019-07-28

尹欣红(1993－),女,硕士生,主要研究方向为复杂网络社团检测;**赵世燕**(1994－),女,硕士生,主要研究方向为复杂网络社团检测;**陈晓云**(1954－),女,硕士,教授,博士生导师,CCF 高级会员,主要研究方向为数据挖掘、大数据分析、网络信息挖掘等,E-mail:chenxy@lzu.edu.cn(通信作者)。

fully prove that this algorithm can extract high-quality community structure from the network, thus effectively providing a basis for community detection field.

Keywords Community detection, Bias, Random walk, Signal propagation, Community structure

1 引言

现实世界中许多复杂系统都可被抽象地表示为复杂网络[1]，系统中的实体或对象抽象为顶点，而实体或对象之间的联系抽象为边。对复杂网络进行深入研究可以发掘顶点分布的情况，进而发现不同实体的分布情况。如在蛋白质交互作用的网络中[2-3]，每个顶点都表示一个蛋白质组织，而边则代表蛋白质之间的交互作用关系，研究该网络中顶点与边的关系，可以探究蛋白质组织在发生交互作用时所产生的关系。随着对复杂网络研究的不断深入，研究人员逐渐发现复杂网络表现出了一个显著的结构特征——社团结构[4]，该结构将网络中的顶点划分为若干个分组，其中分组之间的连接相对稀疏，分组内部的连接比较稠密。对于从现实世界系统中抽象得到的复杂网络，它的组成结构即社团，对应于系统中特定的功能模块。例如，在科研合作网络中[5]，具有相似研究方向的人员构成的集合往往对应该网络的社团[6-7]。

从复杂网络中提取社团结构就是社团检测[7]，随着社团检测研究的兴起，不同领域的研究人员逐渐意识到社团检测的作用。例如，在生物信息网络中，通过社团检测可以根据相似关系探索食物链中的营养、共生以及共栖关系[8]；在社会网络中[9]，通过社团检测能够帮助人们明显认识到他们与其他社团成员间信仰、爱好等方面的不同。

综上所述，社团检测对不同研究领域都有深远的影响，社团检测在理论研究和实际应用中都具有重要的作用和价值，从而引起了研究人员的广泛关注。因此，本文以降低现有算法的时间复杂度为目的，同时结合实际物理现象中信号传递过程中的信号衰减问题，提出了 RSPB(Community Detection algorithm based on Random Walk of Signal Propagation with Bias)算法。该算法通过基于偏置限制信号传播过程的思想，模拟随机游走的过程，使得信号能够不同程度地分布在网络中的所有节点上，从而划分得到最终的社团结构。

2 社团检测的相关理论基础与研究现状

2.1 相关符号的定义

复杂网络的概念来源于图，图可以形式化地表达网络的相关特征，网络可以表示为 $G=(V,E)$，其中 V 代表顶点集，E 代表边集，顶点数记作 $n=|V|$，边数记作 $m=|E|$。

对于网络 $G(V,E)$，存在任意 $u\in E$，若边 $(u,v)\in E$，则顶点 u 和 v 互为邻居。通常，网络中任意 v 的所有邻居所构成的集合是 $N(v)$，顶点 v 的邻居数目则称为该顶点的度，记作 d_v。

社团发现是指在 G 中确定 n_c 个社团：$C=\{C_1,C_2,\cdots,Cn_c\}$，使得各个社团中的顶点能够构成相应的集合，从而使得边 E 能被合理划分。

复杂网络的另一种表示方式是邻接矩阵(Adjacency Matrix)，该矩阵可以表示顶点之间的相邻关系，它的数学表达方式可定义为：在无向网络中，邻接矩阵是一个 $n*n$ 的方阵，用 $\boldsymbol{A}_{ij}$ 表示，定义为：

$$A_{ij}=\begin{cases}1, & \text{if}(u,v)\in E\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

2.2 常见的社团检测算法

Girvan 等对社团检测的研究引起了研究人员的极大兴趣[10]。目前，研究者已经提出了大量的社团检测方法，按照算法的工作模式可分为图论方法、模块度优化算法、层次算法、基于标签传播的算法、基于随机游走的算法等类型。

2.2.1 图论方法

与社团检测相关的图论方法即计算机中的图分割方法，其中 K-L 算法[11]和谱分析法[12-13]是典型的代表算法。K-L 是一种贪婪优化算法，该算法为每个顶点定义连接代价(cost)，从而确定交换的子网络顶点集，并将交换前后的顶点的连接代价的差作为增益值(gain)，以此为基础，重复迭代找到交换过程中增益值最大的一对顶点。它通过贪婪搜索并定义增益函数来获得增益函数值最大的网络划分。

谱分析法最早针对图的剖分问题而提出。最典型的谱分析法是由 Donath 等提出的谱聚类[14]。该方法的思想是，假设网络有 n 个顶点，首先计算拉普拉斯矩阵最小的前 k 个特征值及对应的特征向量，形成 $n*k$ 的矩阵，再利用 k-means 算法划分 k 个社团。Xu 等[15]提出了一种新的基于谱聚类的方法 SCBSP(基于模拟退火和粒子群优化的谱聚类方法)，该方法改善了光谱聚类(SC)中使用 k-均值聚类只能得到局部最优的问题。Gui 等[16]为了从全局的顶点层次结构中获取重叠社团，提出了新的分层组织算法(Spectral Analysis of Line Graph, SAoLG)。

2.2.2 模块度优化算法

模块度(Q)[17]是用于评判社团结构质量好坏的标准。不同的社团划分在给定网络时会得到不同的模块度，Q 值越大，社团质量就越高，由此出现了很多以获得更高 Q 值为目标的优化社团检测算法，主要有极值优化算法[18]、模拟退火算法[19]和遗传算法[20-21]等，其中文献[18]使用极值优化将模块度分解为一组局部变量的组合。首先将网络中的顶点随机划分到大小相同的两个子集，并不断移动以适应度最小的顶点，直到模块度值达到一个极大值，然后移除两个子集之间的边，递归执行，直到模块度值不再增大，从而得到最优的社团结构。

基于模块度方法的主要目的是探索能够使获得的模块度值最大[22]的所有可能性。FastQ[23]算法就是不断选择模块度增量最大的两个社团进行合并，当模块度值最大时，社团划分达到最佳。Blondel 等提出了 Louvain[24]算法，该算法第一步移动顶点，将其移动到使得模块度增量最大且为正的社团中，直到模块度增量不再变化；第二步将第一步形成的社团看作新顶点，构建一个新网络，其中顶点之间的边上带有权，其值为对应的两个社团之间边的权值之和，社团内部的边构成

顶点到其自身的环，其权值为社团内部的边的权值之和。重复迭代直到社团不再改变，从而得到最终的社团结构。

模块度优化算法还常与其他算法相结合来解决社团检测问题，例如 PPC[25] 算法结合了随机游走与模块度来精确有效地揭示社团结构。Aktunc 等[26] 为了快速检测大规模动态网络中的社团，提出了动态模块化优化框架(DMO)，该框架是通过修改已知的基于静态模块化的社团检测算法所构建的，即结合 Louvain 算法的演化算法，动态检测大规模网络的社团结构。

2.2.3　层次算法

层次聚类算法主要分为分裂算法和凝聚算法，其中最著名的分裂算法是 GN[6,10] 算法。GN 算法首先计算网络中每条边的边介数值，迭代删除与最大边介数值对应的边，将其分割成社团，并重新计算网络中剩余边的边介数值，重复上述步骤直至网络中所有边都删除完毕。GN 算法检测到的社团结构的输出形式是有层次结构的树状图(dendrogram)。凝聚算法中的代表算法是 FastQ[23] 算法，它的时间复杂度达到了 $O(m(n+m))$。其主要思想是：首先，将网络中每个顶点看作独立的社团；其次，将使得模块度增量最大的两个社团进行合并，重复迭代合并社团，直到该网络中所有顶点都被划分到相应社团，并且模块度增量不再变化，此时的社团结构就是最终的社团结构。

为了更好地衡量社团划分结果的好坏，Newman 等提出了模块度标准。基于该标准，Clauset 等[27] 发现，FastQ 算法在合并社团时需要计算具体合并哪两个社团，增加了算法的复杂度，因此他们改进了 FastQ 算法，采用平衡二叉树和大顶堆等数据结构的思想，提出了 CNM 算法。Danon 等[28] 改进了社团合并过程，将合并两个社团带来的模块度增量用社团关联的边在网络中所占的比例进行归一化，进而优化 FsatQ 算法。Wakita 等[29] 则采用合并比的策略来改进模块度增量，使得合并过程均衡化。

2.2.4　标签传播算法及其变体

Raghavan 等[30] 基于网络中的信息传播机制提出了 LPA 算法，其主要思想是：在初始阶段，给网络中每个节点一个唯一的标签属性，之后在每次迭代中，每个节点根据与其相邻的邻居节点的标签属性更新自身的标签，更新原则是在每次迭代过程中，网络中的每个顶点将其自身的标签更新为其邻居顶点中最频繁出现的标签，直到标签都与其邻居节点出现次数最多的标签一致，终止算法。LPA 在更新过程中，可采用同步更新与异步更新两种策略，在二分网络中同步更新会导致标签震荡，从而无法继续算法，使其无法继续运行。与此相反，异步更新策略有着更好的适用性。

由于 LPA 算法时间复杂度低，具有简单、高效的特点，研究者提出了一系列改进算法及变体[31]。Barbar 等在标签更新过程中，将该过程转换为一个形式化的优化问题，获取目标函数并增加模块度约束，从而优化了 LPA 算法，提出了 LPAm[32] 算法。Hu 等提出了 roLPA[33] 算法，将不同的角色分配给节点，如中心点、外围节点，以确定节点偏好的新度量，然后将节点偏好信息嵌入到均衡传播和收敛阶段中，以确保其稳定性和效率。Gui 等在 roLPA 算法的基础上，通过节点最大度确定初始社团，然后再确定社团归属度来标记剩余节点，形成了 LPA-CBD[34] 算法。Thakare 等提出 SkipLPA[35] 算法，将度大于相关阈值的节点赋予相应初始标签并使其参与传播阶段，如果节点的度小于阈值，则跳过该阶段。SS-CLPA[36] 算法增加了新约束及更新，减少了标签更新过程中的冗余。

2.2.5　随机游走算法

随机游走[25,37-38] 是最常见的动力学过程，其思想是：游走者从网络中某一顶点出发，根据一定的概率随机游走到与其有边的顶点，并且该顶点为游走者下次游走的初始点，重复迭代上述过程 n 次，停止游走。从该过程可知，游走者倾向于游走在它所属的社团内部，只有很小的概率会从社团中游走出去，亦即，游走者从某一顶点出发，在一定的步骤内进行随机游走，它从初始点所在社团游走到其他社团的概率很小。由此，它的这种特性可应用于社团检测，如 Zhou[39] 模拟随机游走的过程，通过计算顶点间的平均到达距离来定义基于局部吸引子和全局吸引子的社团扩充标准。Ponts 等提出了 WalkTrap[40] 算法，该算法初始都将每个节点视作小社团，从单节点出发，反复合并两个相邻社团并获得与该节点所属社团间最小的距离。Rosvall 等利用编码处理及优化随机游走的过程，用最短编码的方式得到最优的社团划分结果，形成 Infomap[41] 算法。之后他们继续对其进行改进，为了识别多层次的社团结构，提出 Infohiermap[42] 算法。Tabrizi 等将随机游走的过程与模块度自顶向下地结合，采取自顶向下的方式提取网络多层次的社团结构，形成 PPC(Personal Pagerank Clustering)[25] 算法。

基于随机游走的思想，研究人员提出了很多信号传递算法。Hu 等[43] 将网络视为一个可兴奋的系统，随机选择初始节点并赋于该节点一个非空信号值，剩余节点所带信号值为空。初始点将信号传递给自身以及与其相邻的节点并记录此时节点所带的信号量，按照此种方式依次迭代传递，形成带信号的向量，采取不同的方法对该向量进行处理，得到社团划分。Esmailian 等[44] 在信号传播的过程中引入不同规模的消极信息，加入映射方程和波茨模型，提出了 CPM 算法。Bahadori 等[45] 利用随机游走找到网络的种子节点形成特征向量，再合并具有较高共同特征集合的节点，提出了 LRW 算法。

近年来，复杂网络社团检测问题已成为各学术领域的研究热点，为了解决常见算法未涉及的问题，本文设计带偏置的信号传播机制的随机游走的社团检测方法，并对其进行深入研究。

3　基于带偏置的信号传播机制的随机游走算法

在实际存在的物理现象中，电流可能会受到电阻等的影响，不能百分之百地传递到下一个节点并启发之后的传递。考虑到信号传播的过程也会随着传播距离的增大而导致信号衰减，因此本文设计了 RSPB (A algorithm based on Random walk of Signal Propagation with Bias) 算法。为了寻找最优的社团结构，本文算法结合信号量衰减的思想，模拟随机游走

的过程，通过偏置的限定使得社团划分结果趋于稳定。RSPB 算法可分为 3 个步骤，具体如下。

3.1 基于带偏置的随机游走得到信号传播矩阵

在网络 $G=(V,E)$ 中，各个节点以及边不带任何信息。为了结合信号传播过程来模拟随机游走，将网络中的每个节点都视为一个可兴奋的系统，其中每个节点都可发送、接收、记录信号，并且每条边都有衰减因子，即存在偏置 μ_1。随着信号的不断传递，信号量会遍布整个网络，在传递的过程中每个节点只能影响与之有边相连的邻居节点，并且随着传播距离的不断增大，信号量会逐渐衰减。根据信号传播模拟随机游走得到传播矩阵。信号传播的步骤如下：

(1)首先输入网络 G，所有节点在初始状态下不带有任何信号，即该网络中每个节点所带的信号量值为 0。

(2)选择网络中的一个节点，赋予它一个非空信号值，本算法在初始状态下赋予单位信号量。赋值之后该节点就会拥有一定属性值，该属性即为该节点所带的信号量，而网络中其他节点不带任何信号量，并且每条边都带有偏置 μ_1。在之后的信号传递过程中要考虑该值带来的影响(该值整体对网络中节点所带信号量的衰减影响，详见下一步骤)。

(3)进行信号传递。当选择的源节点带有相应的信号量后，随机选择与其相连的邻居节点，并把信号量传递给它，将它作为下一跳节点。此时，每个被传递的节点都会记录所收到的信号量，并继续向下一跳邻居节点发送信号，将其带有的信号量依次进行传递。在信号传递的过程中，要考虑每条边所带的偏置 μ_1，因此下一跳邻居节点所带的信号量为当前节点所带的信号量的值减去偏置 μ_1 的值。基于这样的运算，就能知道偏置 μ_1 的作用，其主要体现在模拟随机游走时，信号量会因为一些客观因素而无法全部传递，设置偏置即代表某些客观因素，为每条边附带衰减因子，使得本文实验过程更加接近于实际网络中的物理现象。

(4)经过一次游走过程之后，由于源节点的影响，信号量会随机分布在网络中的某些节点上，此时将网络中所有节点所带相应的信号量记为向量 $\boldsymbol{N}_i$，不带信号量的节点在向量 $\boldsymbol{N}$ 中的值为 0，带信号的节点根据相关计算后，对应于向量 $\boldsymbol{N}$ 中的特定位置。记录之后，为了使下次游走过程中信号量的分布不受影响，将网络中相关节点所带的信号量都归 0。

(5)重复迭代步骤(1)－步骤(4)，经过 T 步传递之后，信号量会遍布整个网络，从而带有相似信号量的节点会趋于分布在同一个社团，此时把每次得到的向量集中表示，并且用 N 矩阵表示，该矩阵即为随机游走过程得到的传播矩阵。得到传播矩阵的算法如算法 1 所示。

算法 1 基于随机游走得到传播矩阵

```
Input:网络 G=(V,E),衰减因子 μ1,随机游走步长 T
Output:N,Signal matrix
1. Initialize the network,record the number of nodes and define N;
2. Define function:Biase_Random_Walker;
3. for (node:nodes) do
4.    signal_source←node;
5.    signal_index←node of index;
6.    signal←1;#initial a unit signal
7. for j ∈ range(T) do
8.       signal_index=nodes.index(signal_source);
   # get the index of signal_source
9.          next_node←randomly choose node from neighbors;#
            choice one neighbor of the signal souce
10.      next_node_index←next_node;#get the index of the neighbor
11.      signal_next_node_index+=signal_signal_index-μ1
12.      signal_source←next_node;
13.      end
14.   end
15. return signal;
16. Record the value of signal returned for every time,after T iera-
    tions,matrix N is formed;
17. return N.
```

算法 1 给出了 RSPB 算法通过传递信号模拟随机游走过程得到传播矩阵的伪代码。算法 1 中第 1 行定义了信号量矩阵为空。第 3－14 行为一重循环，代表每次随机游走后都将网络节点的信号量归 0 再迭代。第 7－13 行为随机选择邻居节点并计算节点所带信号量的过程，其中偏置的计算如第 11 行所示。对于偏置 μ_1 的设定，不同数据集的设置不同，具体见 4.2 节中的参数设置。第 15－16 行为记录每次输出的信号量，通过上述迭代，输出算法需要的信号传播矩阵 $\boldsymbol{N}$。

3.2 构造新属性矩阵

由于随机游走具有随机性，有些节点所带的信号量尽管会趋于接近并同处于同一社团，但不同社团之间的节点所带的信号量也不一定没有关系，因此只对 $\boldsymbol{N}$ 矩阵进行聚类以得到社团结构是不够的，它不能去除冗余信号的影响，从而使得该路径上的节点所带的信号量不稳定。因此，RSPB 算法结合邻接矩阵的特性，并利用单位矩阵对每个节点增加自环，再结合相似性矩阵与传播矩阵一起构造新的属性矩阵，以得到效果更好的社团结构。其步骤具体可描述为：

(1)首先，得到网络的邻接矩阵 $\boldsymbol{A}$，邻接矩阵的计算公式如式(1)所示。其次，利用网络的单位矩阵 $\boldsymbol{I}$，构造一个新属性矩阵 $\boldsymbol{B}$。此时引入单位矩阵 $\boldsymbol{I}$ 是因为该矩阵的特殊性可以增加对节点自环的限制，使得信号传播过程符合实际，从而更好地得到算法所需要的矩阵。相似性计算公式如下：

$$Sim(u,v)=\frac{|N(u)\bigcap N(v)|}{\sqrt{|N(u)||N(v)|}} \tag{2}$$

其中，分子代表两个顶点的公共邻居数量，分母代表两个顶点的公共邻居数量与总邻居数量之间的关系。

(2)定义一个新的 $n*n$ 的矩阵：$\boldsymbol{C}=\boldsymbol{B}+\boldsymbol{SIM}()$，其中 $\boldsymbol{SIM}$ 为相似性矩阵。此时，计算网络中节点的相似性，形成相似性矩阵，结合邻接矩阵和单位矩阵的特点，扩展原始的相似性，形成新相似性矩阵 $\boldsymbol{C}$。

(3)结合信号传播矩阵 $\boldsymbol{N}$ 和扩展后的相似性矩阵 $\boldsymbol{C}$，得到新的属性矩阵 $\boldsymbol{Y}$，即 $\boldsymbol{Y}=\boldsymbol{N}+\boldsymbol{C}$，从而使得在信号传递过程中，能够去除模拟随机游走后所得到的信号矩阵中一些网络的不相关节点，使其更加切合实际，以提高网络的划分效果。构造属性矩阵 $\boldsymbol{Y}$ 的伪代码如算法 2 所示。算法 2 中代码第 1－6 行为定义并获取具有新属性的相似性矩阵，在分别得到邻接矩阵、单位矩阵后，对 $\boldsymbol{C}$ 矩阵进行运算；第 7－12 行为得到网

络中每个节点的相似性矩阵，并迭代执行更新 $\boldsymbol{C}$ 矩阵；第 13 行为根据得到的 $\boldsymbol{C}$ 矩阵，结合传播矩阵 $\boldsymbol{N}$，得到最终要对其进行计算的矩阵 $\boldsymbol{Y}$。

算法 2　构造新属性矩阵

Input：网络 G=(V,E)，矩阵 **N**，信号量传播矩阵

Output：矩阵 **Y**，新属性矩阵

1. For the N matrix obtained by algorithm 1, new properties are constructed to form a new matrix;
2. num=G. number_of_nodes();
3. **A**←adj_matrix(G);
4. **I**=np. eye(num);
5. **B**←**A**+**I**;
6. **C**←num * num matrix;
7. for i:range(num) do
8.　　for j:range(num) do
9.　　**SIM**←calculate the simliarity matrix using equation (2);
10.　　**C**[i,j]=**B**[i,j]+**SIM**;
11.　　end
12.　end
13. **Y**=**N**+**C**;
14. return **Y**.

3.3　利用 *k*-means 聚类算法划分得到最终的社团结构

新属性矩阵可充分体现网络中节点的拓扑信息，这将为算法的后续步骤提供更好的依据。由于 *k*-means 算法[46]原理简单、容易实现，能够使得网络划分的社团结构更加清晰，因此 RSPB 算法利用 *k*-means 对社团进行聚类。具体过程如算法 3 所示。

算法 3　利用 *k*-means 算法划分得到最终社团结构

Input：网络 G=(V,E)，矩阵 **Y**，新属性矩阵

Output：网络最终社团结构 C_list

1. kmeans=KMeans(n_clusters,random_state). fit(Y);
2. label←kmeans. labels;
3. Define an empty dictionary:cluster, and process the label;
4. for i:range(label) do
5.　　cluster[nodes[i]]=label[i];
6.　　end
7. C_list ← ∅;
8. for v ∈ nodes do
9.　　if cluster[v] not in C_dict then
10.　　C_dict[cluster[v]]=[v];
11.　　else
12.　　C_dict[cluster[v]]+=[v];
13.　　end
14. Convert C_list to a list and print it out, and get the final community structure;
15. return C_list.

算法 3 中，第 1 行为直接调用 Python 工具中 *k*-means 算法的包，输入合适的 *n_clusters* 的个数。第 2－3 行将 *k*-means 算法中的标签存放在 label 中，以便处理，并且定义空字典 cluster 以便对标签进行处理。第 4－6 行遍历所有的标签，将网络中节点的编号与所得到簇内节点的编号进行一一对应。第 7 行定义一个空字典 *C_dict*，将相关节点的编号存入正确的字典中。第 8－13 行遍历所有的节点，对相关节点的标签进行判断：如果该节点的标签不在 *C_dict* 中，则将该节点加入字典中；否则选取下一个节点进行判断。为了便于计算模块度值与 *NMI* 值，将 *C_dict* 字典转换成列表打印输出，从而得到最终的社团。

4　实验

4.1　社团结构质量的评价指标

算法的目标是划分得到相应的社团结构，由于不同的检测算法采用不同的策略，因此得到的结果不一样。目前，被广泛认可的社团结构质量评价指标分别是模块度(Modularity，Q)和归一化互信息量(Normalized Mutual Information，NMI)，它们分别对应社团的内部评价标准和外部评价标准。

4.1.1　模块度(Modularity)

模块度是一种衡量网络社团结构强度的指标，最早由 Mark[6]提出。在社团结构中，社团内部的节点联系紧密且具有较强的模块度，则与之对应的模块度值也较大；反之节点之间联系稀疏则模块度值较小。模块度的计算公式为：

$$Q=\frac{1}{2m}\sum_{ij}\left(A_{ij}-\frac{k_ik_j}{2m}\right)\delta(C_i,C_j) \tag{3}$$

其中，C_i 是节点 i 所属的社团，m 为网络的总边数，k_i 和 k_j 是节点的度，$\delta(x,y)=1$，当且仅当 $x=y$，否则 $\delta(x,y)=0$；C_i 和 C_j 分别代表节点 i 和 j 所属的社团编号。Q 值的范围为[0，1]，其值越小，网络中边的分布越接近于随机网络中边的分布，即网络中社团结构的强度越低；反之，模块度值越高，其对应社团的强度越高。虽然模块度是最常见的检测社团结构质量好坏的标准，但由于分辨率限制问题[47]和自身限制，其最大值所对应的社团结构往往与真实的社团结构存在一定的偏差。因此，本文同时采用另一个评价标准来评价社团检测结果。

4.1.2　归一化互信息量(NMI)

归一化互信息量[48](Normalized Mutual Information，NMI)是一种基于信息论的社团结构质量评价指标，同时也是社团检测过程中需要以真实的社团结构作为基准进行计算进而衡量社团划分好坏的指标。与模块度标准不同的是，它需要一个基准才能对社团检测结果进行评价，即它是一种外部评价标准。*NMI* 用于衡量预测社团结构和真实社团结构之间的相似性，取值范围为[0，1]。该值越接近于 1，表示与其对应的社团结构越接近于真实社团结构，社团结构质量越高。特别地，当 *NMI* 值为 1 时，代表预测社团结构与真实社团结构完全一致。

给定网络 $G=(V,E)$，其顶点数为 n，标准社团结构为 $C_a=\{C_i^A=1,2,\cdots,C_a\}$，由社团检测算法从该网络中提取的社团结构为 $C_b=\{C_i^B=1,2,\cdots,C_b\}$，其中，$C_a$ 和 C_b 分别是标准社团结构和算法提取出的社团结构中社团的数目。社团结构 C_a 和社团结构 C_b 之间的归一化互信息量为：

$$NMI(C_a,C_b)=\frac{-2\times\sum_{i=1}^{|C_a|}\sum_{j=1}^{|C_b|}\frac{n_{ij}}{n}\times\log\left(\frac{\left(\frac{n_{ij}}{n}\right)}{\frac{n_i}{n}\times\frac{n_j}{n}}\right)}{\sum_{i=1}^{|C_a|}\frac{n_i}{n}\log\left(\frac{n_i}{n}\right)+\sum_{j=1}^{|C_b|}\frac{n_j}{n}\log\left(\frac{n_j}{n}\right)} \tag{4}$$

其中，$n_i=|C_i^A|$，$n_j=|C_j^B|$，$n_{ij}=|C_i^A\cap C_j^B|$。

4.2 参数设置

RSPB 算法的整个流程涉及 3 个参数：T，μ_1 和 $n\_clusters$。其中 T 为随机游走中游走过程的步长；μ_1 为设置在网络中每条边上的偏置，偏置的值越大，信号传递的衰减就越大，在检测社团的过程中，有些节点就很可能接收不到信号。这两个参数都为动态参数，当数据集的规模不同时，其值也会有所不同。对所用到的数据集进行多次实验验证，取在设置合适的 μ_1 值时，判断模块度值最大的值，它即为该网络的衰减因子。实验表明，T 的值一般取(3,6)，μ_1 的取值范围一般为(0.05,0.1)。

$n\_clusters$ 为聚类个数，它在真实网络数据集上的值是人为指定的社团数目，对于没有真实结构的数据集，该值为动态值，需实验验证。一般来讲，网络规模越大，划分的社团就越多。本实验对选取的所有数据集进行了多次验证，对不同的网络，设置不同的参数，并进行多次实验，取模块度值最高的社团划分即为最优的社团划分。

4.3 RSPB 算法在实际数据集上的实验结果及分析

4.3.1 实际网络数据集信息

本节实验使用存在真实结构的空手道俱乐部网络(Karate)、海豚网络(Dolphins)、游戏地图网络(RiskMap)、科研合作网络(Santafe)、足球网络(Football)和美国政治书籍网络(Polbooks)数据集，以及社团结构未知的 Jazz 协作网络、Lesmis 网络、YeastL 网络和 PGP 网络数据集，探索本文算法在这些数据集上呈现的结果以及实验对比结果。

Zachary 空手道俱乐部网络[10,49]是对某空手道俱乐部 34 名成员为期 2 年的观察数据。网络中每个顶点代表俱乐部中的一个成员，每条边代表成员之间的关系。

海豚社交网络是由 62 个顶点和 159 条边组成的网络，它分为 4 个社团，该网络由 Lusseas 等[50]创建。网络中的每个顶点代表海豚，顶点之间的边代表这两只海豚经常被观察到在一起。

Risk 游戏地图网络[51]是关于棋盘游戏 Risk 的一张世界地图。游戏是由法国电影导演 Albert Lamorisse 发明的，由 2 到 6 名玩家在一张世界地图上战斗。Risk 网络包含 42 个节点和 83 条边，并且整张游戏地图网络划分为 6 个大洲，每个小社团代表 1 个大洲。

科学家合作网络[10,52]是由 118 个顶点和 197 条边组成的社会关系网络，分为 6 个社团。网络的每个节点都表示科学家，顶点之间的边表示科学家之间存在科学合作关系。

足球网络[10]来源于 2000 年举办的美式足球比赛，由 115 个代表球队的顶点、613 条代表两支足球队之间对阵情况的边组成。在网络中，每对顶点之间连接的边表示两支球队之间有一场比赛要进行，并且根据对阵情况将所有球队划分成 12 个社团。

美国政治书籍网络[17]来源于 2004 年美国总统大选前后出版的有关美国政治的书籍，由 105 个节点和 441 条边组成，其中每个节点表示在电子商务网站亚马逊上出售的书籍，边表示同一买家经常共同购买的图书。

“悲惨世界”(Lesmis)网络[53]描述了维克多·雨果小说《悲惨世界》中主要人物之间的相互作用。节点代表标签所指的角色，而边表示在同一章中出现的任何一对角色。

Jazz 协作网络[54]是对爵士音乐家合作模式研究的网络，它由 198 个节点和 2742 条边组成。网络中每个节点代表每个爵士音乐家，边代表他们之间的合作关系模式。

YeastL 网络[55]记录了负责调控细胞中基因表达的过程，是由 2361 个顶点和 7182 条边组成。其中，顶点代表基因，边代表从编码转录蛋白的基因定向到由该转录因子转录到可调控的基因。

PGP 网络[56]是基于加密程序的信任网络，其中顶点代表证书，边代表将证书的所有者信息授权给另一个证书的所有者。

表 1 列出了本文算法所用到的真实网络数据集的相关信息。

表 1 真实网络数据集信息

Table 1 Information of real network datasets

网络名称	边数/条	平均度	社团数目
Karate 网络	78	4.59	2
RiskMap 网络	83	3.95	6
Dolphins 网络	159	5.13	4
Santafe 网络	197	3.34	6
Football 网络	613	10.67	12
Polbooks 网络	441	8.4	3
Lesmis 网络	253	6.57	—
Jazz 网络	2742	27.7	—
YeastL 网络	7182	6.08	—
PGP 网络	24316	4.55	—

4.3.2 实际网络结果及分析

图 1 给出了 RSPB 算法在空手道俱乐部(Karate)网络数据集上的实验结果。该算法将 Karate 网络划分为 2 个社团，得到了与真实社团结构完全一致的结果。这说明在信号游走的过程中，相似信号都分布在同一网络中。同时，在表 2 中 RSPB 算法的 NMI 值为 1，进一步说明了本文算法在 Karate 网络上的实验结果。

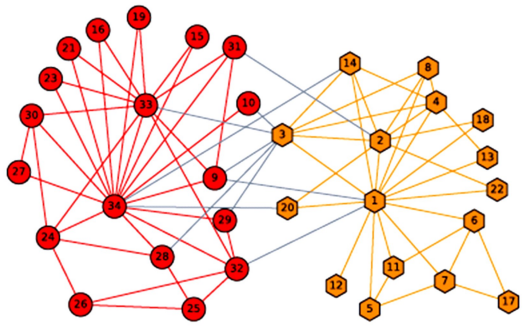
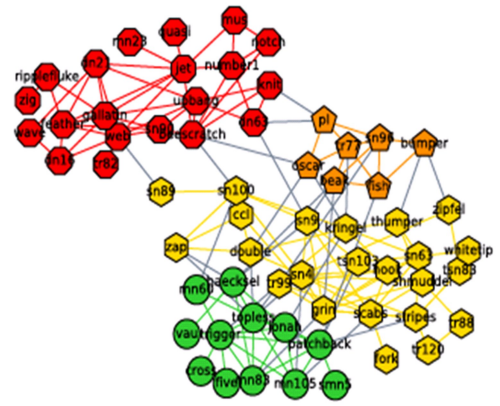


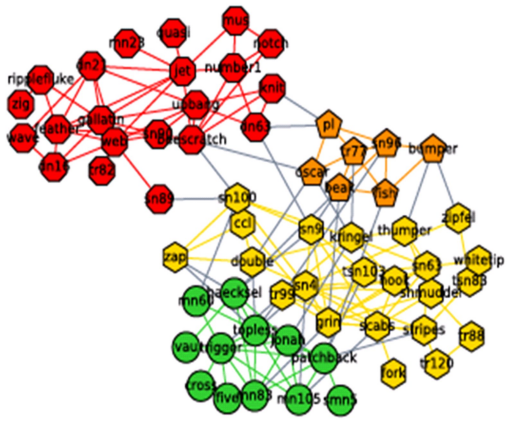
图 1 RSPB 算法在空手道俱乐部网络上的实验结果

Fig. 1 Experimental results of RSPB algorithm on Karate club network

在海豚(Dolphins)网络上，RSPB 算法得到的实验结果与真实网络社团结构的结果对比如图 2 所示。从图中可以看到，RSPB 算法与真实结构都将网络划分为 4 个社团，但稍有区别。节点“sn89”被错误地划分到别的社团，这是因为该节点处于社团边界，容易被误分，并且在信号传播的游走过程中，该节点所带的信号量趋向于与错误社团相似，从而出现误分情况。



(a)已知的社团结构

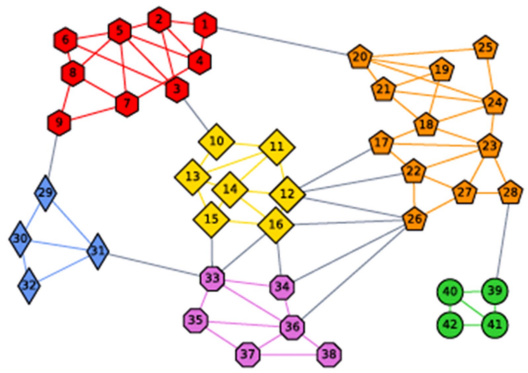


(b)RSPB 算法的检测结果

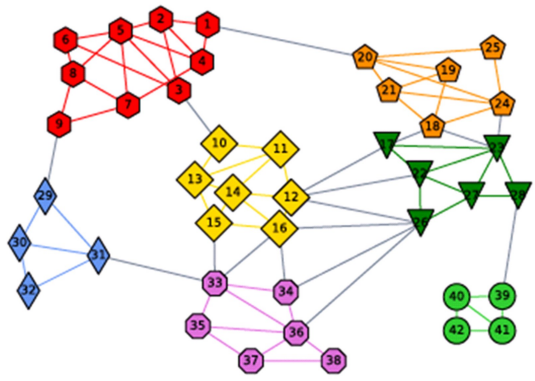
图 2　RSPB 算法在海豚网络上的实验结果

Fig. 2　Experimental results of RSPB algorithm on Dolphins network

图 3 给出了 Risk 游戏地图网络的真实结构与 RSPB 算法在该网络上的实验结果。对比可知，本文算法的结果基本与真实结构一致。节点“17”“18”“19”“20”等连续的 11 个节点所在的社团，被 RSPB 算法划分为两个小社团，节点“18”“19”“20”“21”“24”“25”处于一个社团，而节点“17”“22”“23”“26”“27”“28”处于一个社团，这是由于小社团内部节点信号量的分布结合了扩展的相似性矩阵，使它们更趋于分成两个社团，使得网络的模块度值更高，如表 2 所列。



(a)已知的社团结构



(b)RSPB 算法的检测结果

图 3　RSPB 算法在 Risk 游戏地图网络上的实验结果

Fig. 3　Experimental results of RSPB algorithm on Risk map network

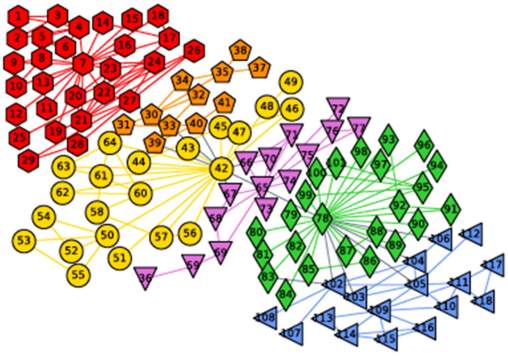
表 2　RSPB 算法的对比结果

Table 2　Comparison results of RSPB algorithm

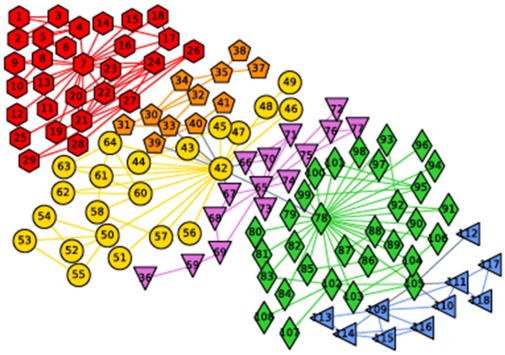
Network	Index	Walktrap	FastQ	LPA	IsoFdp	Attractor	PPC	RSPB
Karate	*Q*	0.3532	0.3807	0.3529	0.3715	0.3715	**0.4107**	0.3715
	NMI	0.5042	0.6925	0.6103	**1**	0.9241	0.5930	**1**
RiskMap	*Q*	0.6240	0.6248	0.5931	0.5186	0.5726	0.6302	**0.6337**
	NMI	0.9085	0.8547	0.8797	0.7142	0.7781	0.8342	**0.9453**
Dolphins	*Q*	0.4888	0.4915	0.4555	0.5049	0.4431	0.5126	**0.5203**
	NMI	0.6324	0.7328	0.7142	0.7444	0.6987	0.8729	**0.9495**
Santafe	*Q*	0.7329	**0.7489**	0.6358	0.6684	0.6938	0.7454	0.7340
	NMI	0.8180	0.8674	0.7428	0.8250	0.8354	0.8830	**0.9310**
Football	*Q*	0.6029	0.5497	0.5879	0.5989	0.6004	0.5997	**0.601**
	NMI	0.9543	0.7514	0.9404	0.9823	0.9165	0.9308	1
Polbooks	*Q*	0.5070	0.5020	0.4855	0.5179	0.5006	0.5150	**0.5266**
	NMI	0.5427	0.5308	0.5308	0.4849	0.4782	0.5476	**0.5695**
Lesmis	*Q*	0.5192	0.4994	0.5098	0.5099	0.4782	0.5212	**0.5350**
Jazz	*Q*	0.4384	0.4389	0.3770	0.4354	0.2738	0.4300	**0.4414**
YeastL	*Q*	0.5185	0.5637	0.3757	—	0.5107	**0.5653**	0.5556
PGP	*Q*	0.7894	0.8543	0.7673	0.7256	0.7681	**0.8712**	0.7967

科研合作网络在 RSPB 算法下得到的实验结果与真实结构的对比如图 4 所示。与真实社团结构相比，RSPB 算法在该网络上得到了与其基本一致的结果，都将网络划分为 6 个社团，其中 5 个社团的结构完全一致，只有一个社团的结构与已知社团结构稍有区别。节点“102”“103”“104”“105”“106”“107”和“108”被划分到错误社团。其中，节点“104”与节点“78”所在的社团有更大的相似性，并且由于经过一定步长的随机游走，在信号传递过程中，信号量趋于与该节点所在的社团相似，而节点“106”只与节点“78”和“104”有边相连，故而也被误分；节点“102”“103”和“106”都与节点“78”有一条边相

连，它们受到节点“78”影响的概率大于原本社团的影响而被误分；节点“107”和“108”处于社团边界，并且都与节点“102”相连，很容易被误分；由于扩展相似矩阵和信号矩阵共同的限制，导致节点“105”与节点“78”所在的社团有更强的相关属性，从而被误分。



(a)已知的社团结构



(b)RSPB 算法的检测结果

图 4 RSPB 算法在科研合作网络上的实验结果

Fig. 4 Experimental results of RSPB algorithm on Santafe network

Football 网络的实验结果如图 5 所示。很明显，RSPB 算法得到了与真实社团结构完全一致的结果，并且在表 2 中，Football 网络的 *NMI* 值为 1，更进一步地证明了本文算法在 Football 网络上能够取得最好的实验结果。

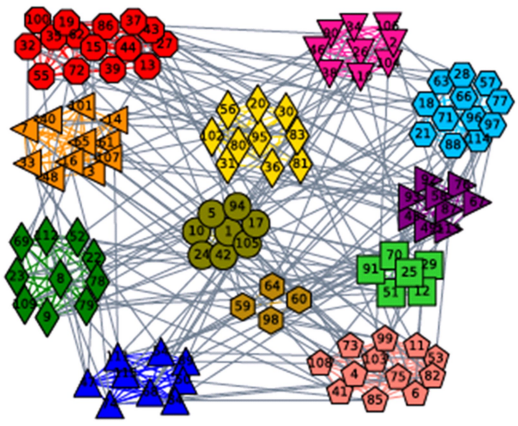
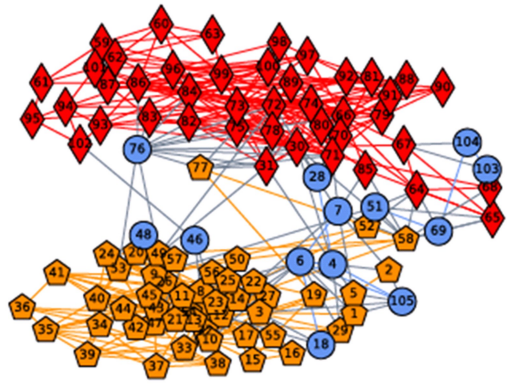


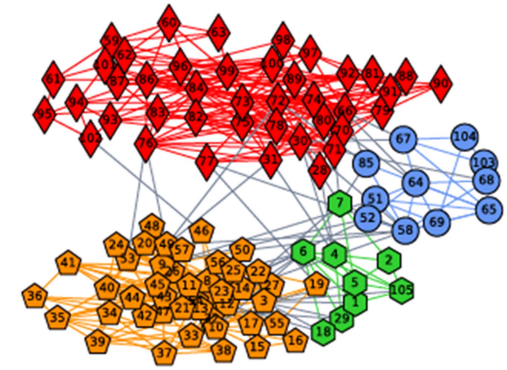
图 5 RSPB 算法在足球网络上的实验结果

Fig. 5 Experimental results of RSPB algorithm on Football network

图 6 给出了美国政治书籍(Polbooks)网络的真实结构与实验结果的对比。真实结构包含 3 个社团，而本文算法却将该网络划分为 4 个社团，使得社团结构更加明显。被误分的节点都划分到了别的社团，形成了新的社团结构，其中节点“1”“2”“4”“5”“6”“7”“18”“29”“105”形成了一个新社团，它们之间的紧密程度比起与原社团连接的紧密程度相对较大；节点“51”“52”“69”“85”“104”等形成了一个社团，使得该网络的模块度更高，如表 2 所列。



(a)已知的社团结构



(b)RSPB 算法的检测结果

图 6 RSPB 算法在美国政治书籍网络上的实验结果

Fig. 6 Experimental results of RSPB algorithm on Polbooks network

Jazz 协作网络、Lesmis 网络是两个社团规模较小的实际数据集，由于它们的社团结构未知，因此不能用可视化的图形来展现本文算法在该网络数据集上的可行性。因此，只选择 Q 来对这两个数据集的结果进行质量评价，评价结果如表 2 所列。

YeastL 网络和 PGP 网络也是实际数据集，但相对来说社团的规模较大，它们也没有公认的已知的社团结构。通过实验验证，RSPB 算法在 YeastL 网络和 PGP 网络上的模块度 *Q* 值分别为 0.5556 和 0.7967，与相关对比算法的结果相比，所提算法取得了较好的结果，并且该值所对应的社团结构都是较为合理的。

将本文算法与 Walktrap，FastQ，LPA，IsoFdp，Attractor 和 PPC 这 6 种算法进行对比。整体来看，RSPB 算法可以得到较好的社团结构划分结果，其在各个数据集上的实验结果对比如表 2 所列。

4.4 合成网络的实验结果及分析

4.4.1 人工合成网络信息

由于本实验所用到的已知社团结构的真实数据集的规模较小，因此 RSPB 算法还需要在 LFR 测试网络模型上生成不同规模的网络，以进一步验证 RSPB 算法的性能。人工合成网络是使用一定的基准网络模型工具生成的[57-58]，在生成数

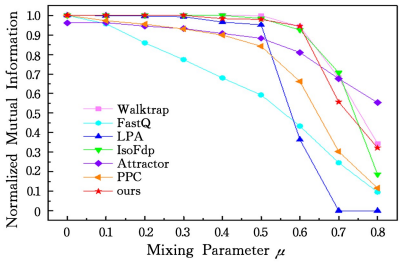
据集时需要设置一些参数，包括网络的平均度、最大度、数据集规模大小、顶点的度，以及社团规模分布指数、度分布指数、混合系数 μ(mixing parameter)、社团规模的范围。

对于人工合成网络，混合系数 μ 指社团之间的边占网络总边数的比例，是影响社团具体结构的重要参数。一般认为，μ 值越小，社团结构越清晰。实验中 μ 的取值范围为 0.1～0.8，生成了社团结构清晰与模糊时的网络，对 μ 的每一个取值生成 10 个相同的网络，并以平均值作为最终结果。实验中用 *NMI* 值对算法的性能进行量化。

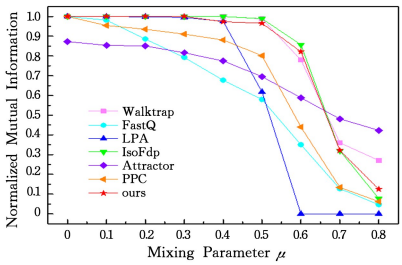
具体而言，在生成 LFR 网络时，将 LFR 网络的平均度设为 20，最大度设置为 50，度分布指数为 −2.0，社团规模分布指数为 −1.0。基于这些参数，生成以下 4 种规模的网络：包含 1000 个节点的网络，社团规模大小从 10 个节点变化到 50 个节点，称为 1000S；社团规模大小从 20 个节点变化到 100 个节点的网络，称为 1000B；另外两种社团大小的参数设置范围与 1000 节点的网络相同，但节点规模却为 5000，分别命名为 5000S 和 5000B。

4.4.2　合成网络的结果及分析

图 7 给出了不同算法在 LFR 上生成的 $N=1000$ 网络上的实验结果。图 7(a)是 1000S 的实验结果，随着 μ 的不断增大，*NMI* 值也会有所变化。当 $\mu\in[0,0.3]$时，RSPB 算法取得了最好的结果，即 $NMI=1$；当 $\mu\in(0.3,0.6)$时，μ 值越大，社团结构越不清晰，此时 RSPB 算法检测得到的 *NMI* 值有所下降，但下降幅度不大，仍处于 0.95 之上，仅次于 IsoFdp 算法和 Attractor 算法；当 $\mu>0.6$ 时，*NMI* 的下降幅度逐渐增大，但该值仍然不低于 0.3，且依旧仅次于 Walktrap 算法、IsoFdp 算法和 Attractor 算法。



(a)1000S



(b)1000B

图 7　在 $N=1000$ 节点的 LFR 网络上不同社团检测算法的结果对比

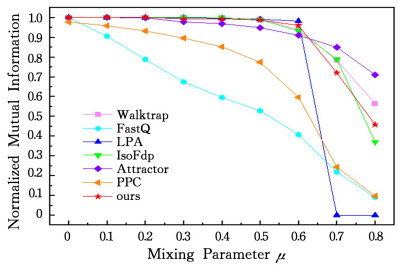
Fig. 7　Comparison of results of different community detection algorithms on LFR networks with $N=1000$ nodes

图 7(b)是 1000B 的实验结果，随着 μ 值的逐渐增大，其 *NMI* 值的整体变化趋势类似于 1000S，但有些许不同。当 $\mu=0.6$ 时，*NMI* 值仍不低于 0.8，且仅次于 IsoFdp 算法；当 $\mu>0.6$ 时，*NMI* 的下降趋势明显，并且比 1000S 的下降趋势更加强烈，随着 μ 值增大，其结果逐渐超过 IsoFdp 算法，但同样次于 Walktrap 算法和 Attractor 算法。

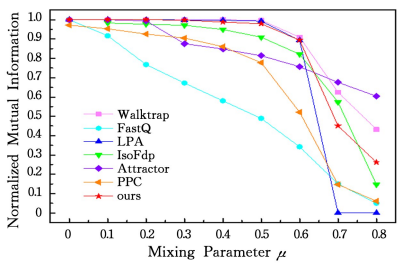
综上分析，当节点数为 1000，社团结构随着 μ 逐渐发生变化时，RSPB 算法在两种社团规模上得到 *NMI* 值的变化趋势基本一致，且 *NMI* 的取值较其他算法高。

LFR-5000 的实验结果对比如图 8 所示。图 8(a)显示的是 5000S 的结果，当 $\mu\in[0,0.2]$时，RSPB 算法得到的社团结构与网络的标准社团结构完全一致；当 μ 逐渐增大到 0.6 时，*NMI* 的下降趋势比较平缓，并且该值仍大于 0.95，因此本文算法得到的社团结构仍接近于实际社团结构；当 μ 的取值逐渐大于 0.6 后，*NMI* 的变化明显下降，且 *NMI* 值不低于 0.45，检测得到中等偏上的结果。

5000B 的变化趋势与 5000S 的变化趋势大体一致，但当社团结构越来越模糊时却有明显的不同，如图 8(b)所示。RSPB 算法在社团结构越清晰即 μ 越小时，得到的社团结构越接近标准社团结构，其 *NMI* 值越趋向于 1；当 μ 越大时，其结果与标准社团结构的差距越大，但与其他算法相比，该算法的表现并不是最差的，处于中等水平，仍可证明 RSPB 算法得到的结果是有效的。



(a)5000S



(b)5000B

图 8　在 $N=5000$ 节点的 LFR 网络上不同社团检测算法的结果对比

Fig. 8　Comparison of results of different community detection algorithms on LFR networks with $N=5000$ nodes

综上分析，RSPB 算法可以在较大规模数据集上进行实验，并且能够得到较好的结果。

结束语　为了能够降低算法的时间复杂度，同时结合实际物理现象中信号传递过程存在信号衰减的问题，本文提出了 RSPB 算法，基于信号传播的思想，模拟随机游走的游走过程，使得信号能够不同程度地分布在网络的所有节点上，并设置偏置来限制传播过程，结合构造新属性矩阵的过程，划分得到最终的社团结构。最后，在真实网络数据集和人工合成网

络上应用该算法，并将其与 6 种已有的算法进行比较，并使用 2 个评价指标进行评价。实验结果证实了 RSPB 算法能从网络中提取高质量的社团结构。

参 考 文 献

[1] ZACHARY W W. An information flow model for conflict and fission in small groups [J]. Journal of Anthropological Research, 1977, 33(4): 452-473.

[2] LEI X J, WANG F, WU F X, et al. Protein complex identification through markov clustering with firefly algorithm on dynamic proteinprotein interaction networks [J]. Information Sciences, 2016, 329(6): 303-316.

[3] ATAY Y, KOC I, BABAOGLU I, et al. Community detection from biological and social networks: A comparative analysis of metaheuristic algorithms [J]. Applied Soft Computing, 2017, 50: 194-211.

[4] FORTUNATO S, HRIC D. Community detection in networks: a user guide [J]. Physics Reports, 2016, 659: 1-44.

[5] NEWMAN M E. The structure of scientific collaboration networks [J]. Proceedings of the National Academy of Sciences of the United States of America, 2001, 98(2): 404-409.

[6] MARK E J N, GIRVAN M. Finding and evaluating community structure in networks [J]. Physical Review E, 2004, 69(2): 026113.

[7] SINGH A, HUMPHRIES M D. Finding communities in sparse networks [J]. Scientific Reports, 2015, 5(1): 8828.

[8] LEWIS A C F, JONES N S, PORTER M A, et al. The function of communities in protein interaction networks at multiple scales [J]. BMC Systems Biology, 2010, 4(1): 100.

[9] BEDI P, SHARMA C. Community detection in social networks [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2016, 6(3): 115-135.

[10] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences, 2002, 99(12): 7821-7826.

[11] KERNIGHAN B W, LIN S. An efficient heuristic procedure for partitioning graphs [J]. The Bell System Technical Journal, 1970, 49(2): 291-307.

[12] FIEDLER M. Laplacian of graphs and algebraic connectivity [J]. Banach Center Publications, 1989, 25(1): 57-70.

[13] URSCHEL J C, ZIKATANOV L T. Spectral bisection of graphs and connectedness [J]. Linear Algebra and its Applications, 2014, 449: 1-16.

[14] DONATH W E, HOFFMAN A J. Lower bounds for the partitioning of graphs [C]// Selected Papers of Alan J Hoffman: With Commentary. World Scientific, 2003: 437-442.

[15] XU Y, ZHUANG Z, LI W M, et al. Effective community division based on improved spectral clustering [J]. Neurocomputing, 2018, 279: 54-62.

[16] GUI C, ZHANG R S, HU R J. Guoming Huang, Jiaxuan Wei. Overlapping communities detection based on spectral analysis of line graphs [J]. Physica A: Statistical Mechanics and its Applications, 2018, 498: 50-65.

[17] NEWMAN M E J. Modularity and community structure in networks [J]. Proceedings of the National Academy of Sciences, 2006, 103(23): 8577-8582.

[18] DUCH J, ARENAS A. Community detection in complex networks using extremal optimization [J]. Physical Review E, 2005, 72(2): 027104.

[19] GUIMERA R, AMARAL L A N. Functional cartography of complex metabolic networks [J]. Nature, 2005, 433(7028): 895.

[20] GUERRERO M, MONTOYA F G, BAÑOS R, et al. Adaptive community detection in complex networks using genetic algorithms [J]. Neurocomputing, 2017, 266: 101-113.

[21] SAID A, ABBASI R A, MAQBOOL O, et al. Cc-ga: a clustering coefficient based genetic algorithm for detecting communities in social networks[J]. Applied Soft Computing, 2018, 63: 59-70.

[22] DŽAMIĆ D, ALOISE D, MLADENOVIĆ N. Ascent-descent variable neighborhood decomposition search for community detection by modularity maximization [J]. Annals of Operations Research, 2019, 272(1/2): 273-287.

[23] NEWMAN M E. Fast algorithm for detecting community structure in networks [J]. Phys Rev E Stat Nonlin Soft Matter Phys, 2004, 69(6 Pt 2): 066133.

[24] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks [J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 2008(10): P10008.

[25] TABRIZI S A, SHAKERY A, Asadpour M, et al. Personalized pagerank clustering: A graph clustering algorithm based on random walks [J]. Physica A: Statistical Mechanics and its Applications, 2013, 392(22): 5772-5785.

[26] AKTUNC R, TOROSLU I H, OZER M, et al. A dynamic modularity based community detection algorithm for large-scale networks: Dslm [C]// 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2015: 1177-1183.

[27] CLAUSET A, NEWMAN M E J, MOORE C. Finding community structure in very large networks [J]. Physical Review E, 2004, 70(6): 066111.

[28] DANON L, DÍAZ-GUILERA A, ARENAS A. The effect of size heterogeneity on community identification in complex networks [J]. Journal of Statistical Mechanics: Theory and Experiment, 2006, 2006(11): P11010.

[29] WAKITA K, TSURUMI T. Finding community structure in mega-scale social networks[C]// Proceedings of the 16th International Conference on World Wide Web. ACM, 2007: 1275-1276.

[30] RAGHAVAN U N, ALBERT R, KUMARA S. Near linear time algorithm to detect community structures in large-scale networks [J]. Physical Review E, 2007, 76(3): 036106.

[31] DENG Z H, QIAO H H, SONG Q, et al. A complex network community detection algorithm based on label propagation and fuzzy c-means [J]. Physica A: Statistical Mechanics and its Applications, 2019, 519: 217-226.

[32] BARBER M J, CLARK J W. Detecting network communities by propagating labels under constraints [J]. Physical Review E,

2009，80(2)：026129.

[33] HU X G，HE W，LI H Z，et al. Role-based label propagation algorithm for community detection [J]. arXiv：1601.06307，2016.

[34] GUI C，ZHANG R S，ZHAO Z L，et al. Lpa-cbd an improved label propagation algorithm based on community belonging degree for community detection [J]. International Journal of Modern Physics C，2018，29(02)：1850011.

[35] THAKARE S B，KIWELEKAR A W. Skiplpa：an efficient label propagation algorithm for community detection in sparse network [C]// Proceedings of the 9th Annual ACM India Conference. ACM，2016：97-106.

[36] CHIN J H，RATNAVELU K. A semi-synchronous label propagation algorithm with constraints for community detection in complex networks [J]. Scientific Reports，2017，7：45836.

[37] SU Y S，WANG B J，ZHANG X Y. A seed-expanding method based on random walks for community detection in networks with ambiguous community structures [J]. Scientific Reports，2017，7：41830.

[38] SUN H L，CHŸNG E，YONG X，et al. A fast community detection method in bipartite networks by distance dynamics [J]. Physica A：Statistical Mechanics and its Applications，2018，96：108-120.

[39] ZHOU H J. Network landscape from a brownian particle's perspective [J]. Physical Review E，2003，67(4)：041908.

[40] PONS P，LATAPY M. Computing communities in large networks using random walks [C]// International Symposium on Computer and Information Sciences. Springer，2005：284-293.

[41] ROSVALL M，BERGSTROM C T. Maps of random walks on complex networks reveal community structure [J]. Proceedings of the National Academy of Sciences，2008，105(4)：1118-1123.

[42] ROSVALL M，BERGSTROM C T. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems [J]. PloS One，2011，6(4)：e18209.

[43] HU Y Q，LI M H，ZHANG P，et al. Community detection by signaling on complex networks [J]. Physical Review E，2008，78(1)：016115.

[44] ESMAILIAN P，JALILI M. Community detection in signed networks：the role of negative ties in different scales [J]. Scientific Reports，2015，5(1)：14339.

[45] BAHADORI S，MORADI P. A local random walk method for identifying communities in social networks [C]// Artificial Intelligence and Robotics(IRANOPEN). IEEE，2017：177-181.

[46] HARTIGAN J A，WONG M A. Algorithm as 136：A k-means clustering algorithm [J]. Journal of the Royal Statistical Society. Series C (Applied Statistics)，1979，28(1)：100-108.

[47] FORTUNATO S，BARTHELEMY M. Resolution limit in community detection [J]. Proceedings of the National Academy of Sciences，2007，104(1)：36-41.

[48] ANA L N F，JAIN A K. Robust data clustering [C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE，2003.

[49] ZACHARY W W. An information flow model for conflict and fission in small groups [J]. Journal of Anthropological Research，1977，33(4)：452-473.

[50] LUSSEAU D，SCHNEIDER K，BOISSEAU O J，et al. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations [J]. Behavioral Ecology and Sociobiology，2003，54(4)：396-405.

[51] STEINHAEUSER K，CHAWLA N V. Identifying and evaluating community structure in complex networks [J]. Pattern Recognition Letters，2010，31(5)：413-421.

[52] NEWMAN M E J. Scientific collaboration networks. i. network construction and fundamental results [J]. Physical Review E，2001，64(1)：016131.

[53] KNUTH D E. The Stanford GraphBase：a platform for combinatorial computing [M]. New York：ACM，1993.

[54] GLEISER P M，DANON L. Community structure in jazz [J]. Advances in Complex Systems，2003，6(4)：565-573.

[55] MILO R，SHEN-ORR S，ITZKOVITZ S，et al. Network motifs：simple building blocks of complex networks [J]. Science，2002，298(5594)：824-827.

[56] BOGUÑÁ M，PASTOR-SATORRAS R，DÍAZ-GUILERA A，et al. Models of social networks based on social distance attachment [J]. Physical Review E，2004，70(5)：056122.

[57] LANCICHINETTI A，FORTUNATO S，RADICCHI F. Benchmark graphs for testing community detection algorithms [J]. Physical Review E，2008，78(4)：046110.

[58] FAGNAN J，ABNAR A，RABBANY R，et al. Modular networks for validating community detection algorithms [J]. arXiv：1801.01229，2018.