

Research on Community Division Method Based on Node Importance

Fang Liu, Caixia Gao*

School of Mathematical Sciences, Inner Mongolia University, Hohhot Inner Mongolia
Email: 1271037374@qq.com, *gaocx0471@163.com

Received: May 13th, 2019; accepted: Jun. 3rd, 2019; published: Jun. 10th, 2019

Abstract

By dividing the community structure existing in the complex network, it is helpful to discover the function, structure, hidden law and influence of the network. In order to get a better classification result of network community structure, and more ways of community division, this paper defines the node importance evaluation function of the network, and proposes a community partitioning algorithm based on node importance degree. The example shows that the algorithm can well classify the community structure in complex networks.

Keywords

Complex Network, Community Structure, Spectral Average Method, Community Division

基于节点重要度的社团划分方法研究

刘 芳, 高彩霞*

内蒙古大学数学科学学院, 内蒙古 呼和浩特
Email: 1271037374@qq.com, *gaocx0471@163.com

收稿日期: 2019年5月13日; 录用日期: 2019年6月3日; 发布日期: 2019年6月10日

摘 要

通过对复杂网络中存在的社团结构进行划分, 有助于发现整个网络的功能、结构、隐藏的规律及网络具有的影响力。为了得到较好的网络社团结构的划分结果, 多一种社团划分的途径, 本文定义了网络的节点重要度评价函数, 结合谱平分法提出了一种基于节点重要度的社团划分算法。实例表明, 该算法能很好地划分复杂网络中的社团结构。

*通讯作者。

关键词

复杂网络, 社团结构, 谱平分法, 社团划分

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来, 复杂网络正处于蓬勃发展的阶段, 成为一个能很好地描述社会科学, 自然科学, 数理学科, 生命学科和工程技术等领域相互关联的复杂模型。一个典型的网络由许多节点与连接节点之间的边组成, 节点代表系统中的个体, 边则表示节点间的作用关系。数学、计算机以及各个研究领域的基础内容等都是复杂网络的分析工具, 自组织, 自相似, 小世界, 无标度等是复杂网络的特性, 复杂系统则是复杂网络的研究目标。随着对网络性质的一些相关研究, 人们发现了很多实际网络都拥有一个共同特性, 即社团结构。复杂网络中社团结构的特性为: 社团内部各个节点的连边相对比较紧密, 而社团之间相互连接的边比较稀疏。现如今, 研究人员已经提出了很多有关复杂网络划分的方法, 这些方法为各个领域的研究人员提供了强有力的工具。无论从实际意义还是从理论研究的角度看, 研究网络的社团结构都很重要, 理由如下: 可以从网络结构上更好地理解节点与边的联系和功能; 可以以社团为基本单位从中去探究整个网络的动态性; 可以揭示网络的层次结构, 并有助于理解网络; 可以为真实世界中规模庞大的稀疏数据提供一种研究方法。通过对于复杂网络的划分从而揭示了网络中社团结构的特性, 不仅有助于了解网络的拓扑结构, 而且具有很强的实际应用价值[1] [2]。

谱平分法[1] [3] [4] [5]是复杂网络社团结构划分的一种有效算法, 在部分网络社团划分中存在错误识别的现象, 针对谱平分法错误识别的问题, 本文在谱平分法的基础上, 提出了基于节点重要度评价函数和社团划分相结合的一种算法。通过对实际网络和人工网络的划分证实了算法的可行性和有效性, 得到了较好地划分结果, 对复杂网络中社团划分的研究具有重要的参考意义。

2. 基础理论及相关概念

定义 1 [1] 一个图是由点集 $V = \{v_i\}$ 以及 V 中元素无序对的一个集合 $E = \{e_k\}$ 所构成的二元组, 记为 $G = (V, E)$, V 中的元素 v_i 称为节点, E 中元素 v_i 称为边。

定义 2 [1] 对于非赋权图 $G = (V, E)$, $|V| = n$, 构造一个矩阵 $A = (a_{ij})_{n \times n}$, 其中

$$a_{ij} = \begin{cases} 1, & (v_i, v_j) \in E \\ 0, & (v_i, v_j) \notin E \end{cases},$$

称矩阵 A 为非赋权图 G 的邻接矩阵。

定义 3 [1] 设网络具有 N 节点。则节点 v_i 的度指标定义为

$$C_d(v_i) = k_i,$$

式中: k_i 为与节点 v_i 直接相连的节点的数目, 即节点 v_i 的度。

定义 4 [1] 在具有 N 个节点的无向网络中一个节点的度不会超过 $N-1$, 故归一化的度指标定义为

$$C_D(v_i) = \frac{k_i}{(N-1)}.$$

定义 5 [1] 设 G 是无向连通网络, 网络节点数为 N , $A = (a_{ij})_{N \times N}$ 为网络 G 的邻接矩阵, D 表示对角线为 $d_{ii} = \sum_{j=1}^N a_{ij}$ 的对角矩阵, 定义网络 G 的 Laplace [6] (拉普拉斯) 矩阵(也称为 Kirchhoff 矩阵)为

$$\tilde{L} = D - A.$$

谱平分法的理论基础是在一个有 N 个节点的无向网络 G 中, 网络的 Laplace 矩阵是对称矩阵 $\tilde{L} = (l_{ij})_{N \times N}$, \tilde{L} 的对角线上的元素 l_{ii} 是节点 v_i 的度, 其他非对角线上的元素 l_{ij} 则表示节点 v_i 和节点 v_j 的连接关系: 如果这两个节点之间有边的连接, 则 $l_{ij} = -1$, 否则为 0。矩阵 \tilde{L} 有一个最小的特征值为 0, 且其对应的特征向量为 $l = (1, 1, \dots, 1)^T$ 。可以从理论上证明, 非零特征值所对应特征向量的各个元素中, 同一个社团内节点的对应元素是近似相等的[1] [2]。

假设矩阵 \tilde{L} 是一个对称矩阵, 矩阵 \tilde{L} 的 n 个特征值为 $0 = \lambda_1, \lambda_2, \dots, \lambda_n$ 且 $\lambda_1 < \lambda_2 < \dots < \lambda_n$, 其对应的 n 个特征向量为 v_1, v_2, \dots, v_n 且 $v_1 < v_2 < \dots < v_n$, 其中 λ_1 为 0, v_1 为全等向量, 第二小特征值 λ_2 对应的特征向量为 v_2 , 又被称作 Fiedler 向量[7]。根据网络 Laplace 矩阵第二小的特征值 λ_2 将其分为两个社团, 这是谱平分法的基本思想。当网络近似地分成两个社团时, 用谱平分法可以得到非常好的划分结果, 但是, 当网络不满足这个条件时, 谱平分法的优点就不能充分的得以体现[8]。

3. 基于节点重要度的社团划分方法

度是网络单个节点属性中简单且很重要的内容。从图中或邻接矩阵中, 可以明显看出一个节点的度越大, 即在该网络中就越“重要” [9]。仅仅用度指标去衡量网络中节点的重要性显然比较片面, 因此本文考虑邻居节点对所研究节点重要性的影响。这样网络中节点自身的重要程度和邻居对其重要性影响程度共同衡量该节点的重要度。本文定义节点重要度评价函数, 如下所示:

设在一个无权无向无自环的网络中, 定义一个节点重要度评价函数为

$$F = C_D(v_i) - e_{ij};$$

其中: 节点 v_i 的度为 k_i , 那么节点 v_i 自身的度指标为 $C_D(v_i) = \frac{k_i}{(N-1)}$,

$$e_{ij} = \begin{cases} \frac{1}{d_{ij}}, & i \neq j \\ 0, & i = j \end{cases},$$

e_{ij} 表示节点 v_i 对节点 v_j 的重要度影响力; d_{ij} 表示网络中节点 v_i 到节点 v_j 的最短距离。

空间自相关理论提出这样一个观点, 两个事物所处的距离越远, 那么彼此间依赖性就越弱, 鉴于此, 本文认为, 如果网络中任意两个节点间的距离越远, 那么就说明这两个节点彼此之间的影响力越低, 反之, 彼此间的影响力越高。现在大部分研究人员也认为有连边的节点之间会具有一些关系, 即节点在复杂网络中的重要性不能只用单一的指标去刻画, 需要从不同的角度, 利用多种属性进行较好的评价。

重要节点在一个网络中占有很显著的地位, 找到重要节点不仅有助于分析网络的结构, 而且可以更好地发现整个网络中隐藏的一些规律。利用节点的重要度对复杂网络中社团的划分也同样是复杂网络研究领域的重点。在这样的一个思路下, 故本文所提出的节点重要度评价函数是节点自身的度值与其邻居节点对其重要度影响力的一个差值, 本文算法便是基于此函数引出的。这个函数不仅包含节点自身的属

性, 也包含邻居节点对节点重要度的影响力, 从而进一步对节点的重要度进行评价。随后结合谱平分法的思想, 计算节点重要度评价函数的最小特征值及所对应的特征向量, 根据第二小特征值所对应特征向量的正负可将复杂网络中的社团进行划分。算法的具体描述如下:

对于一个无向无权含有 n 个节点的网络 $G = \langle V, E \rangle$, 基于谱平分法划分复杂网络社团结构的算法的基本步骤如下:

1) 输入具有 n 个节点的复杂网络 $G = \langle V, E \rangle$ 。写出网络的邻接矩阵并将邻接矩阵初始化, 输入邻接矩阵的上角矩阵;

2) 计算节点 v_i 的自身的度指标 $C_D(v_i) = \frac{k_i}{(N-1)}$; 随后计算节点 v_j 对节点 v_i 的重要度影响值

$$e_{ij} = \begin{cases} \frac{1}{d_{ij}}, i \neq j \\ 0, i = j \end{cases};$$

最后输出节点重要度评价函数 $F = C_D(v_i) - e_{ij}$ 所形成的矩阵;

3) 求此矩阵的最小的两个特征值及特征值所对应的特征向量, 根据元素的正负便可将网络划分成两个社团。特征向量中所有的正的元素对应的网络节点所形成的属于同一个社团, 其他负的元素对应的节点属于第二个社团;

4) 根据以上步骤, 可以将网络进行划分, 刷新图形便可得出划分后的图形。

为了更清楚的表达本文算法的思想和有效性, 选取一个构建好的小网络(见图 1)进行验证。首先输入网络的邻接矩阵, 然后计算出每个节点的度指标(如表 1)及其邻居节点对它的重要度影响值, 随后输出节点重要度评价函数矩阵, 最后计算此矩阵中最小特征值所对应的特征向量(如表 2), 根据特征向量的正负进行对网络的划分。

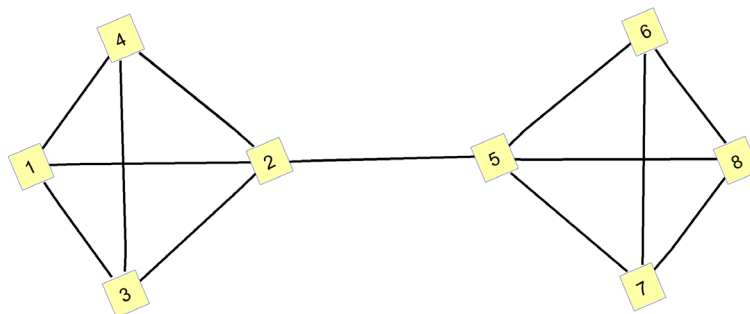


Figure 1. 8 points of undirected network
图 1. 8 个点的无向网络

Table 1. Degree indicators of the corresponding nodes
表 1. 对应节点的度指标

节点	1	2	3	4	5	6	7	8
度指标值	0.4286	0.5714	0.4286	0.4286	0.5714	0.4286	0.4286	0.4286

Table 2. Feature vectors for corresponding nodes
表 2. 对应节点的特征向量

节点	1	2	3	4	5	6	7	8
Vector1	-0.3393	-0.3393	-0.3393	-0.3393	-0.3393	-0.3393	-0.3393	-0.3393
Vector2	0.3849	0.2357	0.3849	0.3849	-0.2357	-0.3849	-0.3849	-0.3849

根据本文定义的节点重要度评价函数, 算出节点重要度评价函数矩阵对应的值(如表 3), 进一步求得节点重要度评价函数矩阵的特征值及特征向量, 那么特征向量 2 的正负便可将网络划分为两个社团, 得到最后的划分结果(如图 2)。

Table 3. Importance evaluation function values for corresponding nodes

表 3. 对应节点的重要度评价函数值

节点	1	2	3	4	5	6	7	8
1	0.4286	-1.0000	-1.0000	-1.0000	-0.5000	-0.3333	-0.3333	-0.3333
2	-1.0000	0.5714	-1.0000	-1.0000	-1.0000	-0.5000	-0.5000	-0.5000
3	-1.0000	-1.0000	0.4286	-1.0000	-0.5000	-0.3333	-0.3333	-0.3333
4	-1.0000	-1.0000	-1.0000	0.4286	-0.5000	-0.3333	-0.3333	-0.3333
5	-0.5000	-1.0000	-0.5000	-0.5000	0.5714	-1.0000	-1.0000	-1.0000
6	-0.3333	-0.5000	-0.3333	-0.3333	-1.0000	0.4286	-1.0000	-1.0000
7	-0.3333	-0.5000	-0.3333	-0.3333	-1.0000	-1.0000	0.4286	-1.0000
8	-0.3333	-0.5000	-0.3333	-0.3333	-1.0000	-1.0000	-1.0000	0.4286

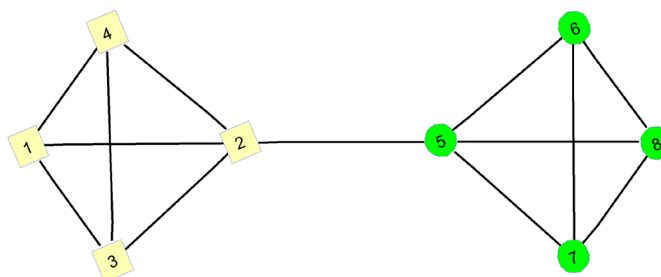


Figure 2. Results of the division of 8 points of undirected networks

图 2. 8 个点的无向网络的划分结果

针对上述步骤, 便可将给出的示例网络正确划分。从表 1 的数据得知节点 2 和 5 的重要度值相等且远远大于剩余节点 1, 3, 4, 6, 7, 8。而节点 1, 3, 4, 6, 7, 8 的重要度值是相等的。如果去掉节点 2 和 5, 会使网络的结构发生较大变化, 对整个网络产生一定的影响。由表 2 和图 2 给出的结果可以看出, 节点 1, 2, 3, 4 从节点重要度评价函数矩阵得出的特征向量 2 的元素都为正值, 即节点 1, 2, 3, 4 是属于同一个社团的, 图 2 的划分结果表明节点 1, 2, 3, 4 的形状, 颜色分别是正方形和米色。而节点 5, 6, 7, 8 从节点重要度评价函数矩阵得出的特征向量 2 的元素为负值, 即节点 5, 6, 7, 8 是属于另外一个社团, 图 2 的划分结果节点 5, 6, 7, 8 的形状, 颜色分别是圆形和绿色。即图 2 是 8 个节点无向网络的正确划分结果。

4. 实验结果与分析

Zachary 空手道俱乐部网络是复杂网络[10]等相关领域的典型测试网络的数据集。在 20 世纪 70 年代初期, 社会学家 Zachary 耗时两年的时间来研究美国一所大学中空手道俱乐部里 34 名成员之间的社会关系的网络, 基于这些成员在俱乐部内外部的交往情况, 构造了成员之间的社会网络体系。此网络包含 34 个节点和 78 条边, 节点代表俱乐部的成员, 两个节点之间有一条边则意味着相应的两个成员间至少是有交往关系的朋友。他通过观察发现, 是因为俱乐部中的主管 JohnA (node1)和教练员 Mr. Hi (node34)之间因为是否提高俱乐部收费问题产生了一些矛盾和分歧, 因此俱乐部便被分裂成分别以主管和教练为代表的两个团体。

图 3 中, 节点 1 和节点 34 分别代表了俱乐部中的主管和教练员, 而正方形和圆形的节点分别代表了分裂后的小俱乐部的各个成员。

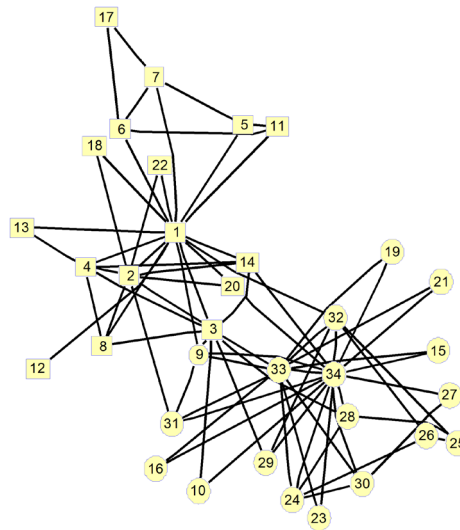


Figure 3. Karate club network membership network
图 3. 空手道俱乐部网络成员关系网络

本文算法首先输入空手道俱乐部网络的邻接矩阵, 通过计算对应节点的度指标(表 4)和邻居节点对其的重要度影响值, 然后输出重要度评价函数矩阵, 最后算出两个最小特征值及对应的特征向量(表 5), 根据特征向量的正负对空手道俱乐部网络进行划分。由划分的结果可以看出本文算法划分的结果(表 6)与现实中的划分结果是一致的。图 4 为本文算法分析空手道俱乐部网络得到的结果, 共同正方形节点和圆形节点代表了原空手道俱乐部网络的实际划分分裂的结果, 而米色和绿色则代表了利用本文算法得到的结果, 与实际结果完全一致, 达到了百分之百的正确率。图 5 是谱平分算法对空手道俱乐部网络得到的划分结果, 从图中得知节点 3 被错误归类, 正确率只达到了 97%。所以本文算法进一步提高了谱平分法的准确率, 克服了节点 3 错误识别的缺陷, 因此本文算法更加适合对空手道俱乐部网络进行划分。

Table 4. Degree indicators of each node of the karate club network
表 4. 空手道俱乐部网络各节点的度指标

节点	度指标值	节点	度指标值	节点	度指标值
1	0.4848	13	0.0606	24	0.1515
2	0.2727	14	0.1515	25	0.0909
3	0.3030	15	0.0606	26	0.0909
4	0.1818	16	0.0606	27	0.0606
5	0.0909	17	0.0606	28	0.1212
6	0.1212	18	0.0606	29	0.0909
7	0.1212	19	0.0606	30	0.1212
8	0.1212	20	0.0909	31	0.1212
9	0.1515	21	0.0606	32	0.1818
10	0.0606	22	0.0606	33	0.3636
11	0.0909	23	0.0606	34	0.5151
12	0.0303				

Table 5. Feature vectors of corresponding nodes in karate club network
表 5. 空手道俱乐部网络对应节点的特征向量

节点	vector1	vector2	节点	vector1	vector2	节点	vector1	vector2
1	-0.2311	0.2867	13	-0.1475	0.1802	24	-0.1674	-0.224
2	-0.1987	0.1721	14	-0.1946	0.0519	25	-0.1452	-0.0526
3	-0.2172	0.0428	15	-0.1514	-0.1769	26	-0.1444	-0.0725
4	-0.1852	0.163	16	-0.1514	-0.1769	27	-0.1469	-0.1795
5	-0.1519	0.2189	17	-0.1141	0.1791	28	-0.1767	-0.1172
6	-0.155	0.2376	18	-0.1497	0.1781	29	-0.1737	-0.0918
7	-0.155	0.2376	19	-0.1514	-0.1769	30	-0.1614	-0.2225
8	-0.1732	0.1539	20	-0.1836	0.0293	31	-0.179	-0.0829
9	-0.1949	-0.0215	21	-0.1514	-0.1769	32	-0.1993	-0.0456
10	-0.1654	-0.0825	22	-0.1497	0.1781	33	-0.2129	-0.2395
11	-0.1519	0.2189	23	-0.1514	-0.1769	34	-0.2324	-0.278
12	-0.1424	0.1627						

Table 6. Division of each node of the karate club network
表 6. 空手道俱乐部网络节点划分情况

社团 1	1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 17, 18, 20, 22
社团 2	9, 10, 15, 16, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34

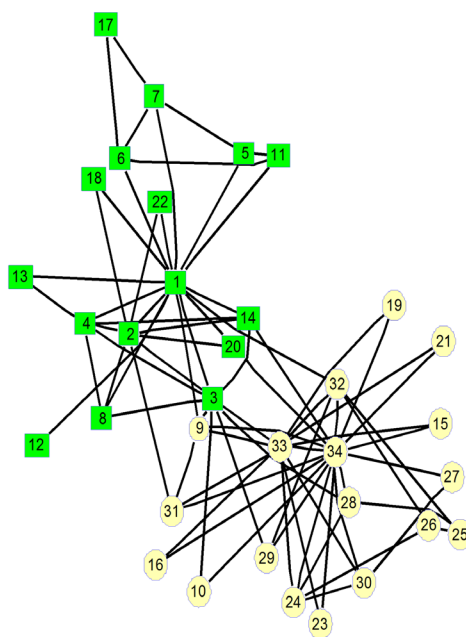


Figure 4. The results of the algorithm's division of the karate club network
图 4. 本文算法对空手道俱乐部网络的划分结果

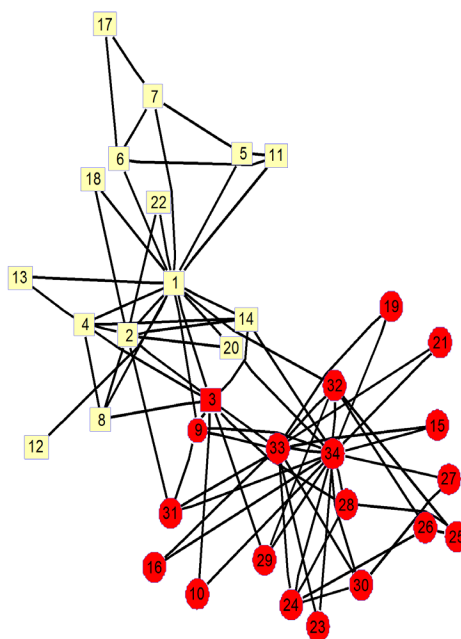


Figure 5. Results of the division of the karate club network by spectral equalization
图 5. 谱平分法对空手道俱乐部网络的划分结果

5. 结束语

本文结合了谱平分法的思想, 计算出了复杂网络各节点自身的度指标和邻居节点对其的重要度影响, 从而构造出一个节点重要度评价函数, 然后利用此评价函数的矩阵算得两个最小的特征值及其对应的特征向量, 提出了一种新的复杂网络社团划分的算法。在构造的 8 个节点的无向网络和经典的 Zachary 空手道俱乐部网络上得以正确的验证, 证明了该算法在复杂网络社团划分中具有可行性和有效性, 通过与经典的谱平分法进行对比后, 本文算法能较好的克服空手道俱乐部网络中个别节点错误识别的现象。本文算法最大的缺陷就是此算法每次只能将网络平分为两个小社团, 对于将一个网络划分为两个以上社团的情况, 就必须多次的进行重复运算。此算法对社团结构明显的网络划分很有效果, 对社团结构不是很明显的网络, 就不能得到很好地效果, 所以这也是本人日后学习研究需要改进的方面。

基金项目

国家自然科学基金(11261033), 内蒙古自治区高等学校科学研究项目 NJZY19005。

参考文献

- [1] 孙玺菁, 司守奎. 复杂网络算法与应用[M]. 北京: 国防工业出版社, 2015.
- [2] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006.
- [3] Pothen, A., Simon, H. and Liou, K.-P. (1990) Partitioning Sparse Matrices with Eigenvectors of Graphs. *SIAM Journal on Matrix Analysis and Applications*, **11**, 430. <https://doi.org/10.1137/0611030>
- [4] Scott, J. (2001) *Social Network Analysis: A Handbook*. 2nd Edition, Sage Publications, London.
- [5] Fiedler, M. (1973) Algebraic Connectivity of Graphs. *Czechoslovak Mathematical Journal*, **23**, 298-305.
- [6] Unser, M., Aldroubi, A. and Eden, M. (1993) B-Spline Signal Processing: Part II Efficient Design and Applications. *IEEE Transactions on Signal Processing*, **41**, 834-848. <https://doi.org/10.1109/78.193221>
- [7] 吴建平, 宋君强, 张卫民, 等. 计算 Fiedler 向量的一种高效准确方法[J]. 计算机学报, 2013(11): 2266-2273.

-
- [8] 吴卫江, 周静, 李国和. 一种基于节点重要度的社团划分算法[J]. 中南民族大学学报, 2016(1):0119-04.
- [9] 周璇, 张凤鸣, 李克武, 等. 利用重要度评价矩阵确定复杂网络关键节点[J]. 物理学报, 2012(5): 1-7.
- [10] Zachary, W.W. (1977) An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, **33**, 452-473. <https://doi.org/10.1086/jar.33.4.3629752>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2324-7991, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: aam@hanspub.org