

基于节点间依赖度的社团结构划分方法*

王兴元[†] 赵仲祥

(大连理工大学电子信息与电气工程学部, 大连 116024)

(2014年3月13日收到; 2014年5月13日收到修改稿)

本文提出了一种基于节点间依赖度的在复杂网络中划分社团结构的算法, 定义了节点对其邻居的依赖度以及节点对社团的依赖度和条件依赖度. 算法的基本要点是优先将最大依赖度不小于其他节点且有惟一依赖节点的节点划分到社团, 并将对社团的依赖度或条件依赖度达到一定值的节点吸收进社团, 直到所有节点都得到准确的社团划分. 本算法在几个实际网络的测试上, 都成功地划分出了满足条件的社团, 并且对社团结构已知的网络的划分结果符合实际情况.

关键词: 复杂网络, 社团结构, 划分, 依赖度

PACS: 89.75.Fb, 89.75.Hc

DOI: 10.7498/aps.63.178901

1 引言

18世纪30年代, Euler通过对七桥问题的抽象和论证, 开创了图论的研究. 而20世纪60年代, Erdős和Rényi建立的随机图理论开创了复杂网络理论的系统性研究^[1]. 在对复杂网络性质的物理意义和数学性质的研究中, 人们发现许多的实际网络都具有这样的性质: 整个网络由若干个群或者团组成, 每个群内部节点间的连接相对紧密, 但是各个群之间的连接相对稀疏, 人们称这种性质为社团结构. 图1显示了一个小型网络社团结构图的示例.

复杂网络的社团结构与计算机科学中的图形分割和社会学中的分级聚类都有着密切的关系^[2,3], 在理论和实际中都有重要的作用, 已经被应用于未知蛋白质功能预测^[4]、社会网络分析^[5], Web社区挖掘等众多领域^[6]. 复杂网络的社团结构对于网络上的许多动力学过程如随机行走等都有非常重要的影响^[7,8]. 人们提出过一些试探性算法来获得近似的社团结构划分, 其中著名的算法有Kernighan-Lin算法^[9], 这种方法是基于贪婪

算法原理将网络划分为两个大小已知社团的二分法; 还有从网络的Laplace矩阵特征值着手的谱平分法^[10,11]. 社会学中的分级聚类也可以用来解决复杂网络中的社团结构检测问题. 分级聚类方法又分为凝聚算法和分裂算法. 凝聚算法从相似性最高的节点对开始向节点数为 n 、边数为的网络中添加边, 这个过程可以随时终止, 终止时的网络就认为是由若干个社团组成的. 有代表性的凝聚算法有

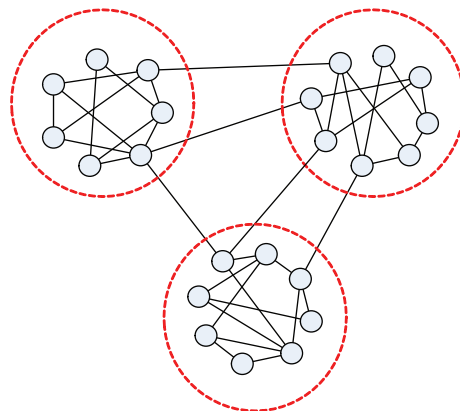


图1 复杂网络的社团结构示例

* 国家自然科学基金(批准号: 61370145, 61173183, 60973152)、高等学校博士学科点专项科研基金(批准号: 20070141014)、辽宁省高等学校优秀人才支持计划(批准号: LR2012003)和中央高校基本科研基金(批准号: DUT12JB06)资助的课题.

[†] 通讯作者. E-mail: wangxy@dlut.edu.cn

Newman 快速算法^[12], CNM 算法^[13]等. 分裂算法的过程与凝聚算法相反, 它从相似性最低的节点对开始从节点数为 n 、边数为 m 的网络中移除边, 分裂过程也可以随时终止, 把终止时的网络组成看作是若干社团的集合. 代表性的分裂算法有 GN 算法^[14]、快速分裂算法^[15]等. 分级聚类算法的过程可以用树状图(图 2)来表示, 从树状图的不同部位截断, 可实现复杂网络的满足一定要求的社团划分. 还有一类算法将物理学中的一些概念用于社团检测, 例如 Reichardt 和 Bornholdt 将网络中的节点与物理学模型中的粒子相对应, 通过修正模型的参数来检测网络中的社团结构^[16,17]; Wu 和 Huberman 提出的一种基于电阻网络电压的社团结构检测算法^[18]. 人们还从网络中的一些特殊节点入手, 通过各种判断标准确定节点间的关系紧密与否、依赖程度强弱, 然后将关系紧密的节点归于同一个社团, 进而将复杂网络划分为几个社团^[19-22]. 近些年来还有研究者们考察节点或节点簇间的相似性^[23], 由此而提出一些社团结构的探测算法^[24,25]. 还有很多的创新算法^[26-28]被提出来以划分网络中的社团结构. 然而, 这些算法或者效率较高, 或者精度较高, 但都难以同时达到高精度和高效率.

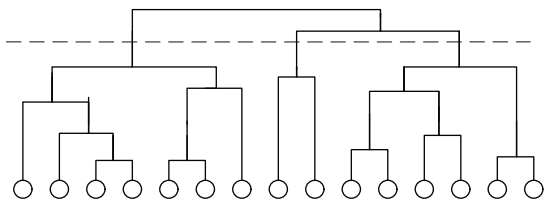


图 2 用树状图记录算法的结果

为此本文提出这种较为准确快速的社团划分方法. 经过实验验证, 本文算法具有较强的准确性和较高的时间效率.

2 方法的描述

2.1 主要思想

在大量的实验和观察中发现, 在划分社团结构的过程中, 一个节点的划分往往依赖于和它有最多公共邻居的邻居节点, 或者依赖于它的大多数邻居节点的划分.

2.2 基本概念

1) 节点对节点的依赖度

节点 a 对它的邻居节点 b 的依赖度 $D_{a,b}$ 定义为

$$D_{a,b} = (n_{a,b} + 1) / k_a, \quad (1)$$

其中 $n_{a,b}$ 表示节点 a 和节点 b 的共同邻居的数目, k_a 表示节点 a 的度. 称节点 a 为起始节点, 节点 b 为终止节点.

2) 节点对社团的依赖度

节点 a 对社团 C 的依赖度 $D_{a,C}$ 定义为

$$D_{a,C} = (n_{a,C}) / k_a, \quad (2)$$

其中 $n_{a,C}$ 表示社团 C 中节点 a 的邻居的数目.

3) 依赖节点

如果节点 b 是节点 a 的一个邻居节点, 并且节点 a 对节点 b 的依赖度不小于节点 a 对它的其他邻居节点的依赖度, 则称节点 b 是节点 a 的一个依赖节点, 称 a 对 b 的依赖度为节点 a 的最大依赖度.

4) 节点对社团的条件依赖度

节点 a 对社团 C 的条件依赖度 $D_{am,C}$ 定义为

$$D_{am,C} = (n_{am,C}) / k_{am}, \quad (3)$$

其中 $n_{am,C}$ 表示社团 C 中节点 a 的依赖节点的数目, k_{am} 表示节点 a 的依赖节点的总数目.

5) 社团的公共节点

如果一个节点对几个社团有着相同的依赖度和条件依赖度, 那么称这个节点是这几个社团的公共节点.

2.3 实现步骤

以下结合对著名的 Zachary 空手道俱乐部人际关系网络^[29](图 3 显示了该网络的原始状态)的社团划分过程来说明本文的社团划分方法.

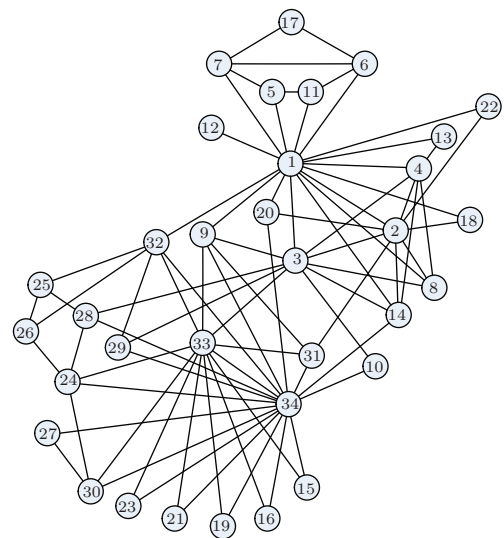


图 3 Zachary 空手道俱乐部网络原始状态

1) 去除复杂网络中度为1的节点, 并将它们邻居的度减1, 直到该复杂网络中不再存在度为1的节点. 在复杂网络中划分社团结构时, 度为1的节点必然和它的惟一邻居归于同一社团, 而它的存在对它的邻居的归属没有任何影响, 因此, 在划分社团之前可先将这类节点去除. 在空手道俱乐部网络中, 去掉度为1的节点12, 并将节点1的度减1.

表1 空手道俱乐部网络中各节点的最大依赖度

节点	度	依赖节点数	依赖节点	最大依赖度
1	15	1	2	0.533
2	9	1	1	0.889
3	10	1	1	0.600
4	6	1	1	1.000
5	3	1	1	1.000
6	4	2	1, 7	0.750
7	4	2	1, 6	0.750
8	4	4	1, 2, 3, 4	1.000
9	5	1	33	0.800
10	2	2	3, 34	0.500
11	3	1	1	1.000
13	2	2	1, 4	1.000
14	5	4	1, 2, 3, 4	0.800
15	2	2	33, 34	1.000
16	2	2	33, 34	1.000
17	2	2	6, 7	1.000
18	2	2	1, 2	1.000
19	2	2	33, 34	1.000
20	3	2	1, 2	0.667
21	2	2	33, 34	1.000
22	2	2	1, 2	1.000
23	2	2	33, 34	1.000
24	5	1	34	0.800
25	3	2	26, 32	0.667
26	3	2	25, 23	0.667
27	2	2	30, 34	1.000
28	4	2	24, 34	0.500
29	3	2	32, 34	0.667
30	4	1	34	1.000
31	4	3	9, 33, 34	0.750
32	6	1	34	0.500
33	12	1	34	0.917
34	17	1	33	0.647

2) 计算复杂网络中所有节点的最大依赖度, 并在这些最大依赖度中找到最大值 D_m . 找到这个最大值 D_m 所对应的所有起始节点(这些节点须具有惟一的依赖节点), 将它们分别和各自的依赖节点组成小的社团. 这些依赖节点称为社团的初始节点. 表1给出了空手道俱乐部网络中各节点的最大

依赖度, 在本步中 D_m 的值为1, 表2显示了这步得到的2个社团.

3) 对于社团的初始节点或者未划分到社团的节点 s , 如果 s 对现存的某个社团的依赖度大于0.5, 则将节点 s 吸收进这个社团; 如果节点 s 对它的邻居的最大依赖度不小于上步中的 D_m , 且 s 对于某个社团的条件依赖度大于0.5, 则将节点 s 吸收进这个社团. 如果节点 s 是某个社团的初始节点, 则将节点 s 所在社团中的所有节点吸收进这个社团中. 表3显示了经过本步骤得到的社团的结果, 13和27两个节点对两个社团的依赖度分别大于0.5, 被吸收进对应社团.

表2 本文算法第2)步得到的2个社团

社团	初始节点	社团中的节点
1	1	4, 5, 11
2	34	30

4) 重复步骤3), 直到没有节点被吸收进社团. 此时, 从未划分到社团的节点和社团的初始节点的依赖度中找到最大值, 将对应于该最大依赖度的具有惟一依赖节点的起始节点 s 吸收进社团(如果终止节点属于这个社团), 或者将节点 s 和终止节点组成一个小的社团(终止节点不在一个社团中). 在对空手道俱乐部网络的划分中, 本步骤的最大依赖度为0.917, 对应于节点33对节点34的依赖度, 将节点33吸收进社团2.

表3 本文算法第3)步后的社团结果

社团	初始节点	社团中的节点
1	1	4, 5, 11, 13
2	34	27, 30

5) 重复步骤4), 直到没有节点被吸收进社团并且没有新的社团成立.

6) 此时社团外的节点与它们的邻居关系不够紧密或者具有一定的歧义性. 在空手道俱乐部网络中, 此时社团外的节点为6, 7, 10, 17. 计算社团外节点对现存社团的依赖度和条件依赖度, 并从这些依赖度和条件依赖度中找到最大值, 将对应的节点(这些节点须具有对应于该最大值的惟一的社团)吸收进相应的社团中. 如果有节点满足步骤3)的条件, 则执行步骤3). 直到没有节点被吸收进社团为止. 表4显示了本步骤后空手道俱乐部社团的结果.

表4 本文算法第6)步后的社团结果

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

7) 对于现存的社团, 如果社团规模过小(例如社团中节点数目少于网络中节点总数目的5%), 则将该社团中的节点依照上述步骤重新划分入现存的满足条件的其他社团. 在对空手道俱乐部网络的划分过程中, 没有出现规模过小的社团.

8) 对于未划分到社团的节点, 根据步骤5)可知它们满足公共节点的定义, 属于与它们相连的社团的公共节点. 在空手道俱乐部网络中, 节点10作为社团1和社团2的公共节点而存在. 重新计算所有节点对现存社团的依赖度和条件依赖度, 确保所有节点对它所处的社团的依赖度大于对其他社团的依赖度, 或者在依赖度相等的情况下有最大的条件依赖度, 否则, 将不满足条件的节点以及受它归属影响的节点重新划分到合适的社团.

经过以上步骤, 再将步骤1)中的度为1的节点划归它们邻居所在的社团中, 即可得到该复杂网络的一种社团划分. 对空手道俱乐部网络, 度为1的节点12属于社团1, 其他节点的划分与表4保持一致, 节点10是两个社团的公共节点.

社团, 如图4所示. 而由于节点10和两个社团都分别有一条边相连, 符合关于社团公共节点的定义, 将它划分为这两个社团的公共节点.

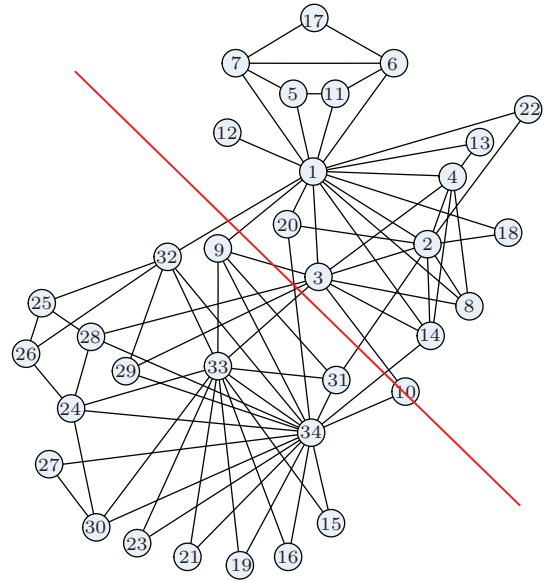


图4 Zachary空手道俱乐部网络中的两个社团

2.4 复杂度分析

对于一个包含 n 个节点和 m 条边的复杂网络, 本文算法在实现过程中, 步骤1)所需的时间复杂度为 $O(n)$. 步骤2)计算节点的最大依赖度所需的时间复杂度为 $O(mn)$, 为便于步骤2)和步骤4)从大到小取用, 将网络中所有节点的最大依赖度排序, 所需时间复杂度为 $O(n \log_2 n)$. 将一个度为 k 的节点吸收进社团最多需要计算 k 次节点对社团的依赖度, 每次计算的时间复杂度为 $O(k)$, 故总的时间复杂度为 $O(k^2)$, 将所有节点吸收进社团的时间复杂度为 $O(nk^2)$. 网络中节点的平均度为 $2m/n$, 而对于稀疏网络, $O(m)$ 等价于 $O(n)$, 故本文算法的时间复杂度为 $O(n^2)$.

3 实验结果与分析

在对 Zachary 空手道俱乐部人际关系网络的实验中, 本算法得到了和实际情况完全一致的两个

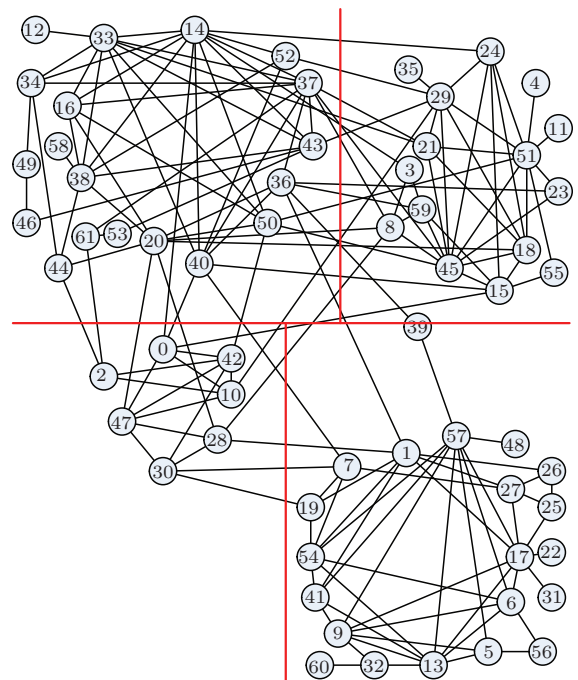


图5 海豚社会网络中的四个社团

Lusseau 等在新西兰对 62 只宽吻海豚的生活习性进行了长时间的观察, 他们研究发现这些海豚的交往呈现出特定的模式, 并构造了包含有 62 个结点的社会网络^[30]. 如果某两只海豚经常一起频繁活动, 那么网络中相应的两个结点之间就会有一条边存在. 本文算法将海豚社会网络划分为 4 个社团, 如图 5 所示, 节点 39 为它所连接的两个社团的公共节点.

美国足球队网络(文献^[5]中的一个例子)中, 每个节点代表了参加美国 2000 年橄榄球赛季的高校代表队, 连接两个节点的边表示对应的两支球队之间至少曾有过一场比赛. 美国足球队网络包含了 115 个节点和 614 条边. 本文算法将该网络划分为 7 个社团, 如图 6 所示, 红色线条将网络划分为 7 个部分.

在对空手道俱乐部网络的划分中, 前文提到的谱平分法^[10,11], GN 算法^[14], Tyler 等在 GN 算法基础上提出的新算法^[31]等, 都将节点 3 作为争议节点划分到了错误的社团中. 表面上看, 节点 3 和两边的社团都有同样数目的边相连, 但是实际上, 以节点 1 为中心的社团中节点 3 的邻居间的联系更为紧密, 节点 3 对节点 1 的依赖度更高, 故将节点 3 划分到节点 1 的社团中. 节点 10 和两个社团分别有一条边相连而被划分为公共节点, 本文所得的结果更加符合实际情况和社团结构的定义. 在文献^[14]

中提出了一种得到普遍认同的、对社团结构划分质量的评价标准: 模块度 Q 值, Q 值较大, 表明该算法所做划分较好. 表 5 是本算法和其他一些算法对海豚社会网络和美国足球队网络所做划分的 Q 值的比较. 从表 5 可以看出, 本文算法在这些复杂网络尤其是节点层次明显的网络的社团划分上有相当不错的表现.

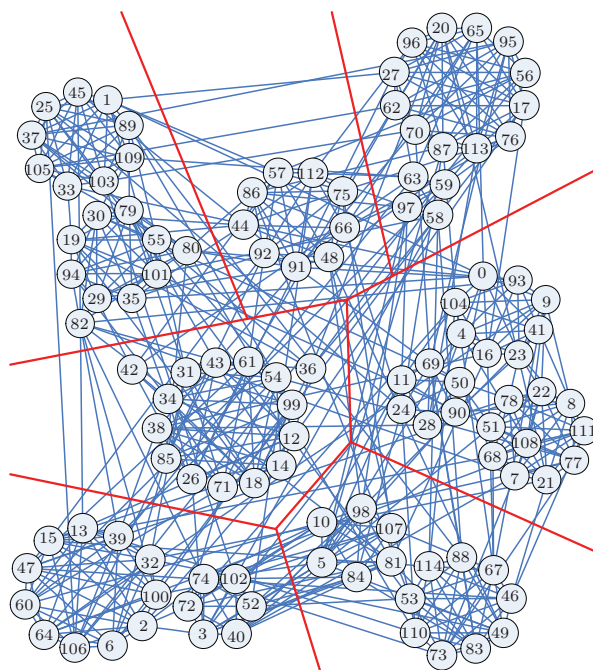


图 6 美国足球队网络中的七个社团

表 5 不同算法社团划分的 Q 值比较

算法	GN 算法	Newman 快速算法	文献 ^[22] 的算法	本文算法
海豚社会网络 ^{a)}	0.380	0.454	0.508	0.550
美国足球队网络 ^{a)}	0.427	0.546	0.581	0.569

a) <http://www-personal.umich.edu/~mejn/netdata/>.

4 结 论

本文提出了一种新的基于节点间依赖度的在复杂网络中划分社团结构的方法, 定义了节点对它的邻居的依赖度以及节点对社团的依赖度和条件依赖度. 首先找到具有惟一依赖节点且最大依赖度不小于其他节点的节点, 将它和它的依赖节点组成社团, 然后对社团的依赖度和条件依赖度满足要求的节点吸收进社团或者继续组成新的社团, 直到所有节点都被划分到社团. 算法的设计使得划分得到的社团中, 每个节点对它所在社团的依赖度都不会小于它对其他社团的依赖度, 也就是说, 任何节点

都和它的尽可能多的邻居划分到同一个社团, 和任何节点相连的边都尽可能多的成为社团内部的边, 使得结果更符合社团结构的定义. 通过对 3 个经典实际网络的测试, 本文算法都取得了不错的结果, 而算法的时间复杂度为 $O(n^2)$, 达到了较高的水准.

参考文献

- [1] Erdős P, Rényi A 1960 *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17
- [2] Garey M R, Johnson D S 1979 *Computers and Intractability: A Guide to the Theory of NP-Completeness* (vol. 1) (San Francisco: Freeman Publishers) p1

- [3] Scott J 2002 *Social Network Analysis: A Handbook* (vol. 2) (London: Sage Publications) p1
- [4] Wang Z, Zhang J Z 2007 *PLoS Computational Biology* **3** e107
- [5] Givan M, Newman M E J 2002 *Proc. Natl. Acad. Sci. USA* **99** 7821
- [6] Ino H, Kudo M, Nakamura A 2005 *Partitioning of web graphs by community topology Chiba*, May 10–14, **2005** 661
- [7] Zhang Z Z, Lin Y, Gao S Y, Zhou S G, Guan J H, Li M 2009 *Phys. Rev. E* **80** 051120
- [8] Zhang Z Z, Yang Y H, Lin Y 2012 *Phys. Rev. E* **85** 011106
- [9] Kernighan B W, Lin S 1970 *Bell System Technical Journal* **49** 291
- [10] Fiedler M 1973 *Czechoslovak Mathematical Journal* **23** 298
- [11] Pothén A, Simon H D, Liou K P 1990 *SIAM J. Matrix Anal. Appl.* **11** 430
- [12] Newman M E J 2004 *Phys. Rev. E* **69** 066133
- [13] Clauset A, Newman M E J, Moore C 2004 *Phys. Rev. E* **70** 066111
- [14] Newman M E J, Girvan M 2004 *Phys. Rev. E* **69** 026113
- [15] Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D 2004 *Proc. Natl. Acad. Sci. USA* **101** 2658
- [16] Reichardt J, Bornholdt S 2004 *Phys. Rev. Lett.* **93** 218701
- [17] Reichardt J, Bornholdt S 2006 *Phys. Rev. E* **93** 016110
- [18] Wu F, Huberman B A 2004 *Eur. Phys. J. B* **38** 331
- [19] Zanjani A A H, Darooneh A H 2011 *Phys. Rev. E* **84** 036109
- [20] Chen D B, Fu Y, Shang M S 2009 *Phys. A* **388** 2741
- [21] Pan Y, Li D H, Liu J G, Liang J Z 2010 *Phys. A* **389** 2849
- [22] Wang X Y, Li J Q 2013 *Phys. A* **392** 2555
- [23] Shen Y, Xu H L 2010 *Acta Phys. Sin.* **59** 6022 (in Chinese) [沈毅, 徐焕良 2010 物理学报 **59** 6022]
- [24] Newman M E J 2004 *Eur. Phys. J. B* **38** 321
- [25] Yuan C, Chai Y 2012 *Acta Phys. Sin.* **61** 218901 (in Chinese) [袁超, 柴毅 2012 物理学报 **61** 218901]
- [26] Zhan W H, Zhang Z Z, Guan J H, Zhou S G 2011 *Phys. Rev. E* **83** 066120
- [27] Shen Y 2011 *Chin. Phys. B* **20** 040511
- [28] Shen Y 2013 *Chin. Phys. B* **22** 058903
- [29] Zachary W W 1977 *J. Anthropol. Res.* **33** 452
- [30] Lusseau D 2003 *Proc R Soc. Lond. B* **270** 186
- [31] Tyler J, Wilkinson D, Huberman B 2005 *The Information Society: An International Journal* **21** 81

Partitioning community structure in complex networks based on node dependent degree^{*}

Wang Xing-Yuan[†] Zhao Zhong-Xiang

(Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, china)

(Received 13 March 2014; revised manuscript received 13 May 2014)

Abstract

In this paper, we present a new approach to partitioning communities in a complex network via degree of dependence of nodes . We define the dependence degree of a node on its neighbors, the dependencetce degree and the conditional dependence degree of a node on a cluster. The main point of the approach is to partition the nodes, which have the biggest dependence degree and are only dependent on nodes, firstly to clusters, then to absorb nodes whose dependence degree or conditional dependence degree on cluster gets the right value, until all the nodes are partitioned to the right communities. The partition of our approach in some real-world network satisfies the definition of communities, and in the network whose communities are already known, our partition method fits the physical truth.

Keywords: complex network, community structure, partitioning, dependence degree

PACS: 89.75.Fb, 89.75.Hc

DOI: [10.7498/aps.63.178901](https://doi.org/10.7498/aps.63.178901)

* Project supported by the National Natural Science Foundation of China (Grant Nos. 61370145, 61173183, 60973152), the Doctoral Program Foundation of Institution of Higher Education of China (Grant No. 20070141014), Program for Liaoning Excellent Talents in University, china (Grant No. LR2012003), and the Fundamental Research Funds for the Central Universities, china (Grant No. DUT12JB06).

† Corresponding author. E-mail: wangxy@dlut.edu.cn