

加权网络权重自相似评判函数 及其社团结构检测*

沈 毅[†] 徐焕良

(南京农业大学信息科学技术学院, 南京 210095)

(2010 年 12 月 17 日收到; 2010 年 1 月 24 日收到修改稿)

提出了权重自相似性加权网络社团结构评判函数,并基于该函数提出一种谱分析算法检测社团结构,结果表明算法能将加权网络划分为同一社团内边权值分布均匀,而社团间边权值分布随机的社团结构.通过建立具有社团结构的加权随机网络分析了该算法的准确性,与 WEO 和 WGN 算法相比,在评判权重自相似的阈值系数取较小时,该算法具有较高的准确性.对于一个具有 n 个节点和 c 个社团的加权网络,社团结构检测的复杂度为 $O(cn^2/2)$.通过设置评判权重自相似的阈值系数,可检测出能反映节点联系稳定性的层化性社团结构.这与传统意义上只将加权网络划分为社团中边权值较大而社团间边权值较小的标准不同,从另一个角度更好地提取了加权网络的结构信息.

关键词: 权重自相似, 加权网络, 社团结构, 谱分析算法

PACC: 0590, 0547

1. 引 言

许多大规模的分布式复杂系统都可以用复杂网络的形式描述,例如 Internet^[1,2], WWW^[2,3], 电力网络^[3], 社会网络^[4,5] 和生物网络^[5] 等. 可以观察到这些网络都具有模块特性或称为社团特性,即基于不同概念如节点相似性等形成的一些节点组. 社团结构这一特征使得架起了从微观的单个节点联系整个宏观网络系统的桥梁^[6,7]. 因此,社团结构检测成为了复杂网络研究领域的一个热点.

近年来,广大研究者提出了很多社团结构检测算法,见综述文献[8]. 然而几乎所有的算法都有一个共同点,它们都针对无权网络的社团结构检测. 无权网络中的社团结构检测旨在如何将网络划分为若干个社团,使得社团内部边密度相对社团间最大. 但是,除了对网络拓扑结构本身的讨论外,不可忽视的一点就是很多现实网络,例如 WWW^[2,3], 食物链网络^[5], 新陈代谢网络^[9] 等在本质上是加权网络,即不仅要观察两个节点之间是否有关联并且还要观察其关联程度的大小. 例如在社会网络中,个

体间就有强的和弱的联系;在新陈代谢网络^[5]中,不同的反应路径上有不同的物质流量. 不仅如此,在很多情况下,边权重将影响整个网络的功能和特性,例如疾病传播^[10,11], 振子的动力学同步^[12-16], 相继故障^[17,18] 等. 显然,边权重在网络中是一个非常重要的变量,如果忽略边的权重,那就会丢掉网络中很多有用的信息^[19,20].

近年来,报道了一些关于加权网络中社团结构检测的算法,包括基于状态转移矩阵的随机游走算法^[21], WGN (weighted GN algorithm) 算法^[22] 和 WEO (weighted extremal optimization algorithm) 算法^[23]. 基于状态转移矩阵的随机游走算法首先将网络看成一个连通的导电实体,计算节点之间的电抗性,然后根据电抗性的大小建立状态转移矩阵,游走质子根据状态矩阵在网络上行走,形成的轨迹和网络社团联系在一起,从而揭示网络社团结构. 由于计算网络电抗矩阵的计算复杂度较高,因此检测速度较慢^[24]. WGN 算法^[22] 是由检测无权网络的 GN (Newman and Girvan) 算法^[25] 演变而来,演变方法是将 GN 算法中的边介数替换为关于权重的边介数^[22],通过每次删除边介数最大的边分离网络,得

* 中央高校基本科研业务费专项资金(批准号:KYZ200916)和南京农业大学青年科创基金资助的课题.

[†] E-mail: shen_yi1979@126.com

到社团结构,研究表明对于高密度网络该算法的准确度不高^[24,26]. WEO 算法^[23,26]是由检测无权网络社团结构的 EO 算法^[23]演变而来,演变方法是将极值优化函数转化为 Newman 提出的 Q^w ^[22],研究表明该算法的准确度比 WGN 和随机游走高^[24,26]. 通过优化 Newman 提出的 Q^w ^[22],上述这些算法将网络划分为若干个社团,使社团中权值总和相对社团间最大,即从整体联系强度上体现社团结构. Q^w 表示为^[22]

$$Q^w = \frac{1}{2T} \sum_{ij} \left[w_{ij} - \frac{T_i T_j}{2T} \right] \delta(c_i, c_j), \quad (1)$$

其中 w_{ij} 代表节点 i 和 j 之间边的权值, T_i 为节点 i 的权重, $T_i = \sum_j w_{ij}$ 为连接节点 i 的所有边的权值, c_i 表示节点 i 所属的社团,当 c_i 和 c_j 为同一社团是时, $\delta(c_i, c_j) = 1$,反之则为 0.

然而,值得一提的是边的权值不仅表明了节点之间联系的强度,还表明了节点功能的相似性,权值分布特征包含了很多有待我们发现的网络结构信息. 因此,加权网络的社团结构检测不应仅仅局限于如何将网络划分为若干个内部边权值较大的社团. 在某些情况下,例如在社会网络中,具有相同兴趣爱好的团体,个体之间联系强度往往较稳定,表现为边权值分布较均匀,体现其节点功能的相似性.

基于此,本文根据加权网络边权重自相似原则,研究如何将加权网络划分为若干个在社团内部边权值分布均匀而社团间分布较随机的社团. 为此,本文提出了一个加权模块度函数,并基于该模块度函数提出了一种谱分析的算法. 为了验证该算法的正确性,构建了一种特殊的加权随机网络分析,取得了令人满意的结果. 最后对一个现实网络进行分析,并通过调整边权重自相似阈值系数揭示了其基于权重相似性的层次化结构,这能获得更多关于加权网络结构的信息.

2. 模块度函数构造

如何衡量社团划分的优劣是正确检测社团结构的基础,采用模块度函数是一种行之有效的办法,在广泛的无权网络社团结构检测研究中,Newman 提出的模块度 Q 函数被很多研究者所采用^[5],并在此基础上产生的很多算法都具有很好的检测效果. 因此,在阐述本文提出的权重自相似模

块度函数前,先分析无权网络的模块度函数,然后引出权重自相似模块度函数.

2.1. 无权网络的模块度函数

Newman 提出的模块度 Q 函数定义为社团中的边减去社团中因该有的边数. 众所周知,优化 Q 函数可以得到正确的社团结构^[8]. 基于随机网络无社团结构的启发性思想^[27],用配置模型描述节点 i 和 j 之间因该有的边数为 $k_i k_j / 2m$ ^[27],其中 k_i 为节点 i 的度, m 是网络中总边数. 模块度函数表示为^[27]

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(g_i, g_j), \quad (2)$$

其中 A_{ij} 为网络的邻接矩阵元素, g_i 是节点 i 属于的网络,如果节点 i 和节点 j 属于同一个社团,则 $\delta(g_i, g_j) = 1$,反之则为 0.

在无权网络配置模型中,节点对之间应该有的边数只取决于它们的度数. 在一般简单情况下,即网络中相连节点对间只有一条边,节点本身无自连边时, $k_i k_j / 2m$ 越大,对应节点 i 和 j 之间应该有边的概率越大. 显然,社团内的边密度越大,对模块度值的贡献也越大. 因此,对于各种可能的划分取最大的 Q 值就对应正确的社团结构.

2.2. 权重自相似模块度函数

从上述分析可知,优化模块度函数所得的最终划分结果依赖于配置模型的形式. 对于无权网络,基于完全随机网络无社团结构的启发性思想,因该有的边应连接节点度乘积较大的节点对,可通过无权网络模块度函数中的配置模型进行描述. 对于加权网络,考虑两个节点 A 和 B ,如果它们具有相同权值的边很多,那么它们之间应该有边连接概率较大,因为只有权值相同的边才可能相连. 反之,如果它们没有权值相同的边,那么它们之间不可能有边连接. 因此,如果 A 和 B 没有权值相同的边,而当它们之间假设有边时,自然形成一种不协调,因对模块度函数值有所贡献,因为设计模块度函数本身的思想是让不协调的成分对其值有所贡献^[23]. 根据模块度函数的设计理念,提出如下的配置模型函数:

$$P_{ij}^w = \frac{\sum_{ij} \sum_{d \in k_i} \sum_{e \in k_j} \Phi(w_{id}, w_{je})}{2m}, \quad (3)$$

其中, k_i 和 k_j 为节点 i 和 j 的度, w_{id} 和 w_{je} 分别是节点 i 的第 d 条边和节点 j 的第 e 条边. 定义 $\Phi(a, b)$

$= I_{(|a-b| \leq \zeta)}$ 为一个事件示性函数, 当 $||a| - |b|| \leq \zeta$ 时, $\Phi(a, b) = 1$, 反之为 0. ζ 是一个阈值系数用来衡量边权值之间的相似性. 与(2)式中无权网络的配置模型不同, (3)式描述的配置模型融合了边权重和节点度数的信息. 这里提出的加权模块度函数表示为

$$Q^W = \frac{1}{2m} \sum_{ij} [A_{ij} - P_{ij}^W] \delta(g_i, g_j). \quad (4)$$

优化(4)式即可找出具有边权值近似的社团. 通过(4)式可将无权网络, 有向网络和加权网络模块度函数统一在相同框架体系下. 对于无权网络, 只要将边的权值看成是相等的. 对于有向网络, 只需在 w_{id} 加上方向信息, 改写成 w_{id}^{in} 或 w_{id}^{out} 即可.

3. 模块度谱分析算法

由于(4)式与(2)式形式相似, 因此许多通过优化模块度函数分析无权网络社团结构的算法经过稍加改动或不改动便能用到加权网络分析中. 例如模拟退火算法^[28]和贪婪算法^[25]等, 这里提出一种基于权重自相似模块度函数的谱分析算法.

定义 s_i 等于 +1 时节点 i 属于组 1, -1 时节点 i 属于组 2, 这样 $\sum_i s_i^2 = n$ 且 $\delta(g_i, g_j) = \frac{1}{2}(s_i s_j + 1)$, 那么模块度函数可写为

$$Q^W = \frac{1}{4m} \sum_{ij} [A_{ij} - P_{ij}^W] (s_i s_j + 1) = \frac{1}{4m} \mathbf{s}^T \mathbf{B}^W \mathbf{s}, \quad (5)$$

向量 \mathbf{s} 的元素为 s_i , \mathbf{B}^W 为模式矩阵, 其元素为 $B_{ij} = A_{ij} - P_{ij}^W$.

优化的目的是找出向量 \mathbf{s} , 对于给定的 \mathbf{B}^W 使得 Q^W 达到最大. \mathbf{s} 可以写成 \mathbf{B}^W 归一化特征向量 \mathbf{u}_i 的线性组合形式 $\mathbf{s} = \sum_{i=1}^n a_i \mathbf{u}_i$. 由于 $a_i = \mathbf{u}_i^T \mathbf{s}$, 那么 $Q^W = \frac{1}{4m} \sum_i a_i^2 \beta_i$, 其中 β_i 为 \mathbf{B}^W 的特征向量 \mathbf{u}_i 对应的特征值. 优化 Q^W 取最大值, 则应该让 \mathbf{s} 平行于主特征向量 $\mathbf{u}^{(\max)}$, 即最大特征值 β_i 对应的特征向量. 但 s_i 只能为 $s_i = \pm 1$, 且 s_i 不能取小数使 Q^W 达到理论最大值. 只能将 s_i 尽量接近于 $\mathbf{u}^{(\max)}$. 这样, 当 $u_i^{(\max)} \geq 0$ 时, $s_i = +1$; 反之 $u_i^{(\max)} \leq 0$ 时, $s_i = -1$, 其中 $u_i^{(\max)}$ 是向量 $\mathbf{u}^{(\max)}$ 第 i 个元素. 该算法可将网络每次分裂成两部分, 多次分裂可对分裂所得的子网络反复使用该算法. 当分裂至不能增大 Q^W 时, 分裂得到的子网络即为社团.

对于一个具有 n 个节点的网络, 计算所有特征值和特征向量的复杂度为 $O(n^3)$. \mathbf{B}^W 矩阵为对称矩阵且网络通常为稀疏网络, 因此可采用快速 Lanczos^[8]算法. 由于只需关注主特征向量, 因此计算主特征向量的复杂度为 $O(n^2)$. 假设加权网络有 c 个社团, 则最多需要分裂 $c/2$ 次, 因此社团结构检测的复杂度为 $O(cn^2/2)$.

为了清楚地说明算法效果, 这里专门设计一种检测该算法的 32 节点随机加权网络. 在这个网络中, 设节点对之间有边的概率为 p , 边一开始无权值, 因此这个初始无权网络就是一个无社团结构的随机网络, 将这个网络分为 16 个节点一组, 组中节点对之间的边权值设置均为 1, 组间的边权值设置为从 0.0 到 1.0 带有一位小数随机权值. 设置阈值系数 ζ 为 0, 即只有当边权值相等时, 配置模型中的事件函数 $\Phi = 1$. 对该网络分析后, 结果如图 1 所示.

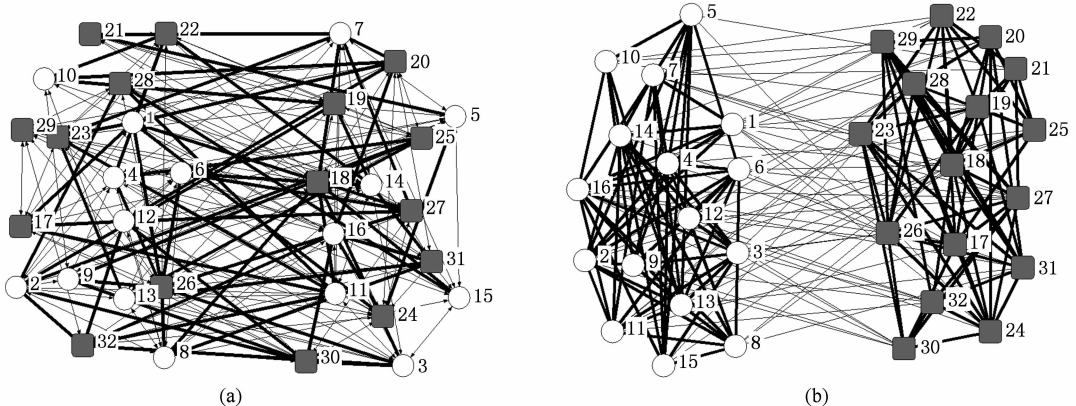


图 1 32 节点随机加权网络分析结果(粗边表示具有权值为 1) (a)忽略边权值的谱分析算法结果;(b)基于权重自相似模块度函数的谱分析算法结果

如图 1 所示,当忽略边权值,只考虑节点之间有无边连接时,网络等同于无社团结构的完全随机网络,因此算法不能找出社团结构.当考虑边权值时,基于权重自相似模块度函数的谱分析算法精确地将网络划分两个社团,社团内部边权值相等,而社团之间边权值分布随机.

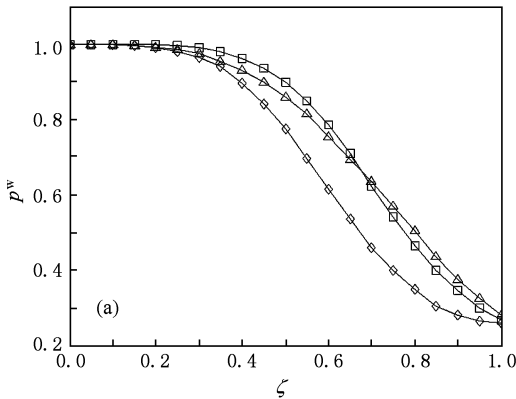
4. 模拟实验分析

在这节中,为检测本文所提出社团结构,构造了一种特殊随机加权网络.将 Q^w 作为目标优化函数,对基于 Q^w 的谱分析算法进行模拟实验分析,并和 WNG, WEO 算法进行了比较.最后分析一个具有社团结构的现实加权网络^[30].

4.1. 具有社团结构的加权随机网络

先构造一种随机网络检测算法效果,该网络具有 128 个节点分为 4 个社团,每个社团有 32 个节点,设置每个节点的度为 16,且让每个节点有 8 条边连接本社团内部的节点,8 条边连接社团外部的节点,这意味着当忽略边的权值时,即作为无权网络检测时,是一个无社团结构的完全随机网络.这里设置两种情况进行分析,一是社团中的边权值为 1,二是社团 1, 2, 3, 4 中边的权值分别为 0.25, 0.50, 0.75, 1,这两种情况的社团间边权值都为 0—1 的随机值,且最小间隔为 0.05.为了清楚地分析算法检测正确性,对于上述两种情况各建立 200 个随机网络分析.设置 ζ 从 0 增加到 1 检测算法的准确度,分析结果如图 2 所示,纵坐标 p^w 为检测正确的节点数占总节点数的百分比,横坐标为 ζ .

如图 2 所示,对于上述两种情况,从 $\zeta = 0$ 到 $\zeta =$



0.6, 基于 Q^w 的谱分析算法能正确检测超过 80% 的节点,并且准确度高于 WEO 和 WGN. 由于第一种情况社团中的边权值相似性要比第二种大,因此对于基于 Q^w 的谱分析算法,随着 ζ 的增大,第一种情况的检测结果比第二种好.当 $\zeta \geq 0.6$ 时,由于 ζ 过大而不能提供衡量边相似性的精确尺度,使得算法检测性能随 ζ 的增大快速下降.对于第一种情况,当 $\zeta > 0.7$ 时,基于 Q^w 的谱分析算法准确度低于 WEO; 对于第二种情况,当 $\zeta > 0.6$ 时,基于 Q^w 的谱分析算法准确度低于 WEO.当 $\zeta = 1$ 时,在基于权重自相似的尺度下,网络成完全随机,算法不能检测出社团结构,因此各种算法对两种情况检测的准确度收敛于 0.25.

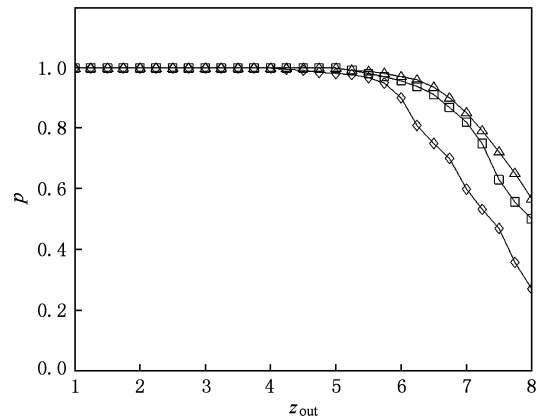


图 3 带有矩形的线为本文提出的谱分析算法结果(带有三角形的线为 EO 算法结果,带有菱形的线为 GN 算法结果)

当 ζ 取较大值时,衡量加权网络的尺度转化为无权网络,即对边权进行了二值化,其目标优化函数从 Q^w 转换为 Q .为了进一步分析算法检测无权网络社团结构的准确性和普遍适用性,构造具有社

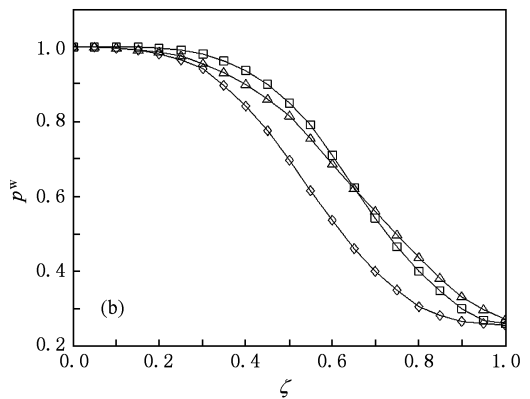


图 2 具有社团结构的加权随机网络检测结果 (a)和(b)分别为对第一种情况和第二种情况的检测结果(带有矩形的线为本文提出的基于 Q^w 的谱分析算法结果,带有三角形的线为 WEO 算法结果,带有菱形的线为 WGN 算法结果)

团结构的无权随机网络^[26]对算法进行了分析,构造的网络共 128 节点,分为 4 个社团,每个社团节点数均为 32,节点之间的边随机独立分布,平均每个节点有 z_{in} 条边与其同一社团内的节点连接, z_{out} 条边与其社团外部的节点连接,并且满足 $z_{in} + z_{out} = 16$,这样可以方便地通过改变 z_{out} 控制社团外与社团内的相对边密度,依此用来检测算法随网络结构变化的准确性. 取不同的 z_{out} 各建立 200 个随机网络,对本文提出的算法进行分析,并且与 CN 算法和 EO 作了比较,分析结果如图 3 所示,纵坐标 p 为检测正确的节点数占总节点数的百分比,横坐标为 z_{out} .

随 z_{out} 增大,社团内的边密度与社团间的边密度开始接近,网络的社团结构出现模糊,这时如果检测正确的节点数占总节点数的百分比越高,则说明算法检测准确性越好. 该无权随机网络已被很多研究者用来检验算法的准确性^[26],其中模拟退火算法准确度最高,当 $z_{out} = 8$ 时,平均检测正确节点数超

过 80%,但由于该算法需要选取参数和高的计算复杂度,因此不适用于规模较大的网络. 本文算法为 52%,略低于 EO 算法,但是高于 GN 算法.

4.2. 现实网络

用该算法分析一个科研合作网络^[30],该网络共有 152 个节点表示研究者,如果两个研究者在 1992—1993 年间有论文合作发表,则在他们之间添加一条边,边上的整数表示他们合作发表论文的篇数,这样就形成了一个加权网络,边的权重清楚地表示了科研者之间的科研合作情况. 采用基于权重自相似模块度函数的谱分析算法检测社团结构,为了清楚地表示检测过程,这里先图示该网络,如图 4 (a) 所示,然后以课题组为单位分别命名,如图 4 (b) 所示. 图 5 为算法随 ζ 从 5 减小到 2 的分析结果,由于社团大小不同会影响谱分析法的精确性^[31],因此在算法运行过程中采用一种“精确协

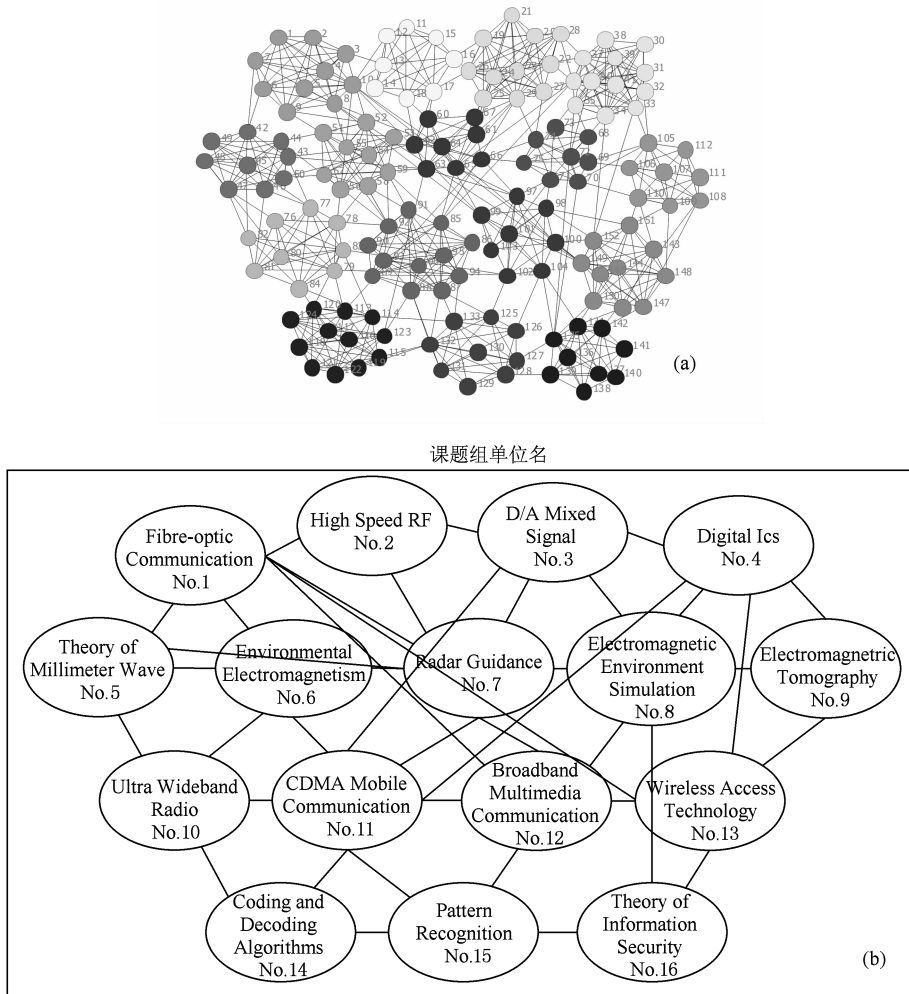


图 4 (a) 科研协作网络; (b) 与 (a) 对应的 16 个研究小组 (分别用序号表示)

调^[31] (在不同社团间移动一些节点使得模块度尽量增大)”。

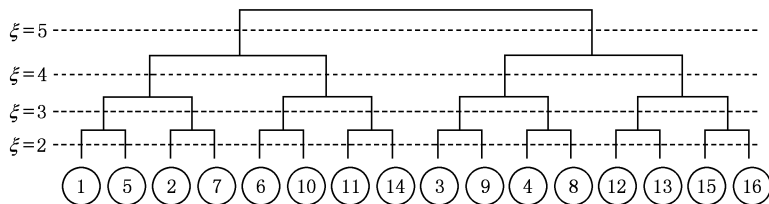


图5 随 ζ 从 5 到 2 的划分结果

很显然,对于同一课题感兴趣的研究者往往共同研究并合作撰写论文,因此他们之间合作的论文数量较多且数量近似.这种课题与研究之间复杂的协作关系使网络产生了社团结构.随 ζ 的减小形成的划分顺序揭示了网络内节点联系稳定性的一种层次化结构.当 ζ 取较大值时,划分的社团结构体现了一个较广的研究领域,在这个较广的研究领域内,包括的研究者自然较多,且分布于各研究小组内.有较稳定合作关系的研究者们将随着 ζ 的减少,被划分在同一个社团内.这里当 $\zeta = 2$ 时,除了少数几个节点被划分到其他社团外,该算法检测的 16 个社团正好对应 16 个研究小组(为了图示层次化划分的清楚性,图 5 忽略了这些节点).当 $\zeta = 1$ 时,网络被划分为 43 部分,其中最大部分包含 5 个节点,其余大都只包含 3 或 4 个节点,虽然算法揭示了网络中某些节点间最稳定的连接关系,但由于包含的节点数量较少,它们与一般传统意义下的社团^[32]有所区别.从以上分析结果可以看出,基于权重自相似模块度函数的谱分析算法能准确地揭示加权网

络中描述强度连接关系的层次化社团结构.

5. 结 论

传统的加权网络社团结构检测研究如何将网络划分了若干个社团,使社团内边权值总和相对于社团间较大.与之不同,为了进一步了解网络内节点之间联系的稳定关系,本文致力于将加权网络如何划分同一社团内部边权值分布相对均匀而社团之间的权值分布相对随机的社团结构.为此,本文提出一个新的加权模块度函数,并基于该模块度函数提出了一种谱分析算法.通过建立具有社团结构的加权随机网络分析了该算法的准确性,与 WEO 和 WGN 算法相比,在评判权重自相似的阈值系数 ζ 取较小时,该算法具有较高的准确性.对于一个具有 n 个节点和 c 个社团的加权网络,社团结构检测的复杂度为 $O(cn^2/2)$.通过设置不同的 ζ ,该算法能准确地揭示加权网络中描述节点连接稳定性的层次化社团结构.

[1] Watts D J, Strogatz S H 1998 *Nature* **393** 440
 [2] Albert R, Jeong H, Barabási A L 1999 *Nature* **401** 130
 [3] Wang G Z, Cao Y J, Bao Z J, Han Z X 2009 *Acta Phys. Sin.* **58** 3597 (in Chinese) [王光增、曹一家、包哲静、韩祯祥 2009 物理学报 **58** 3597]
 [4] Zhang L, Liu Y 2008 *Acta Phys. Sin.* **57** 5419 (in Chinese) [张立、刘云 2008 物理学报 **57** 5419]
 [5] Newman M E J 2002 *Proc. Natl. Acad. Sci. U. S. A.* **99** 7821
 [6] Newman M E J, Barabási A L, Watts D J 2006 *The Structure and Dynamics of Networks* (Princeton: Princeton University Press)
 [7] Shen Y, Pei W J, Wang K, Li T, Wang S P 2008 *Physica A* **387** 6663
 [8] Danon L, Duch J, Guilera A D, Arenas A 2005 *J. Stat. Mech.* p09008
 [9] Holme P, Huss M, Jeong H 2003 *Bioinformatics* **19** 532

[10] Guimerà R, Amaral L A N 2005 *Nature* **433** 895
 [11] Colizza V, Barrat A M, Vespignani A 2006 *Proc. Natl. Acad. Sci. U. S. A.* **103** 2015
 [12] Ni S J, Weng W G, Fan W C 2009 *Acta Phys. Sin.* **58** 3707 (in Chinese) [倪顺江、翁文国、范维澄 2009 物理学报 **58** 3707]
 [13] Ma X J, Wang Y, Zheng Z G 2009 *Acta Phys. Sin.* **58** 4426 (in Chinese) [马晓娟、王延、郑志刚 2009 物理学报 **58** 4426]
 [14] Lü L, Zhang C 2009 *Acta Phys. Sin.* **58** 1462 (in Chinese) [吕翎、张超 2009 物理学报 **58** 1462]
 [15] Song Q S, Feng Z R, Li R H 2009 *Acta Phys. Sin.* **58** 5057 (in Chinese) [宋青松、冯祖仁、李人厚 2009 物理学报 **58** 5057]
 [16] Shen Y, Pei W J, Wang K, Wang S P 2009 *Chin. Phys. B* **18** 3783
 [17] Zhang Z, Fu Z Q, Yan G 2009 *Chin. Phys. B* **18** 2209
 [18] Wang J W, Rong L L 2009 *Acta Phys. Sin.* **58** 3714 (in Chinese)

- [王建伟、荣莉莉 2009 物理学报 **58** 3714]
- [19] Ouyang M, Fei Q, Yu M H 2008 *Acta Phys. Sin.* **57** 6763 (in Chinese) [欧阳敏、费 奇、余明晖 2008 物理学报 **57** 6763]
- [20] Xu Q X, Xu X J 2009 *Chin. Phys. B* **18** 933
- [21] Nelson A A 2007 *Phys. Rev. E* **76** 036101
- [22] Newman M E J 2004 *Phys. Rev. E* **70** 056131
- [23] Duch J, Arenas A 2005 *Phys. Rev. E* **72** 027104
- [24] Fan Y, Li M, Zhang P, Wu J S, Di Z R 2006 *Physica A* **370** 869
- [25] Newman M E J, Girvan M 2004 *Phys. Rev. E* **69** 026113
- [26] Lancichinetti A, Fortunato S, 2009 *Phys. Rev. E* **80** 056117
- [27] Newman M E J 2006 *Phys. Rev. E* **74** 036104
- [28] Roger G, Marta S P, Luis A, Nunes A 2004 *Phys. Rev. E* **70** 025101
- [29] Newman M E J 2004 *Phys. Rev. E* **69** 066133
- [30] Frank K A 1996 *Soc. Networks* **18** 93
- [31] Danon L, Díaz-Guilera A, Duch J, Arenas A 2006 *J. Stat. Mech.* p11010
- [32] Newman M E J 2006 *Proc. Natl. Acad. Sci. U. S. A.* **103** 8577

The evaluation function of weight similarity and its application in community detection in weighted networks^{*}

Shen Yi[†] Xu Huan-Liang

(College of Information Science and Technology, Nanjing Agricultural University, Nanjing 210095, China)

(Received 17 December 2009; revised manuscript received 24 January 2010)

Abstract

An evaluation function of weight similarity in weighted network is proposed, and a spectral algorithm for detecting community structure based on the function is presented. The results show that the algorithm can divide the weighted network into several groups within each of them the edges' weights distribute uniformly but at random between them. The algorithm is analyzed by constructing random weighted networks with known community structure. Compared with WEO and WGN, the algorithm has high accuracy when the threshold coefficient takes small values. For a network with n nodes and c communities, the computation complexity of the algorithm is $O(cn^2/2)$. By setting different threshold coefficients, a special hierarchical organization which describes the various steady connections between nodes in groups can be discovered by the algorithm. It is different from the conventional concept of community detection in weighted networks which divides the weighted network into several groups in which the edges' weights are relatively larger than those in-between them, such that it extracts the information about the structure of weighted networks from another perspective.

Keywords: weight similarity, weighted networks, community structure, spectral algorithm

PACC: 0590, 0547

^{*} Project supported by the Fundamental Research Funds for the Central Universities (Grant No. KYZ200916) and the Excellent Young Scientists Fund of Nanjing Agricultural University.

[†] E-mail: shen_yi1979@126.com