

网络的模块矩阵及其社团结构指标^{*}

王高峡^{1)2)†} 沈 轶²⁾

1)(三峡大学理学院, 宜昌 443002)

2)(华中科技大学系统工程研究所, 武汉 430074)

(2008 年 6 月 20 日收到; 2009 年 6 月 3 日收到修改稿)

探讨了复杂网络的模块矩阵的正(负)特征谱与网络的社团结构(反社团结构)的关系, 给出了反映网络社团结构性质的相关定义. 利用模块矩阵的多个特征值与特征向量, 引入反映个体对所处社团的依附程度一种结构中心化指标. 利用人工网络与实际网络数据, 将这种指标与几种经典的中心化指标进行了比较. 结果表明该指标具有较好的分辨率并与度指标具有一定程度的相关性.

关键词: 网络, 模块矩阵, 社团结构, 中心化

PACC: 0590, 0175

1. 引 言

网络的社团结构 (community structure) 及社团结构的发现是复杂网络的一个热点研究领域^[1-10]. 一般而言, 社团可包含模块 (module)、类 (clustering)、群或组 (group) 等各种含义. 网络的社团结构, 是指整个网络是由若干个群或组构成的, 每个群内部的节点之间的连接非常紧密, 而各个群之间的连接相对来说比较稀疏^[1].

网络可以用矩阵的形式加以描述, 邻接矩阵 (adjacent matrix) 刻画了网络中节点之间的连接关系. 在邻接矩阵的基础上, 可以定义其他形式的矩阵, 如拉普拉斯矩阵 (Laplacian matrix)、正规矩阵 (normal matrix)、归一化矩阵 (normalized matrix) 等, 这些矩阵都是邻接矩阵的变形, 因而也都包含着网络结构的有用信息. 通常称矩阵特征值的集合为矩阵的特征谱. 基于上述矩阵的特征谱, 已提出了不少社团划分的方法, 具体可见文献[11—14]等.

Newman 等^[1,2] 引入模块函数 (modularity function) 作为衡量社团结构划分效果的一种标准. 模块函数着重于社团内部的边的连接情况, 反映了社团内部的边相对于空模型 (null model) 的连接紧

密程度. 在文献[5,6]中, Newman 在前期工作的基础上, 重新演绎了模块函数并引入了模块矩阵 (modularity matrix), 从而将社团结构的识别问题转化成模块函数及模块矩阵的优化问题^[15,16]. 基于模块矩阵的社团结构的识别也很快得到研究, 并对实际网络具有较好的应用效果^[5-17].

与社团结构相对应, 网络也会呈现出“反社团”结构. 对于一种特定的划分, 如果网络的外部边数相对于空模型大于集合的内部边数, 此时对网络的这种划分就表现出一种“反社团”性质. 网络划分为两部分的情形, 若内部的节点相互之间没有连接, 所有的连边均为外部连边, 这种类型的网络称为二部图或二模图, 若内部仅有少量连边, 则为近似二部图. 在网络划分成多个部分的情形, 则有多部图或近似多部图^[5].

无论网络是具有社团结构还是“反社团”结构, 本文均称网络具有社团结构性质. 模块矩阵的特征值及其特征向量较好地反映了网络的社团结构信息^[5,6,16], 因而可以用依附于节点的特征向量的值来测量该节点在整个网络中相对于社团的重要性. 文献[18]中我们利用模块矩阵最大模度特征值所对应的特征向量, 引入了一种描述网络社团结构性质的模块中心化指标. 本文进一步讨论了正特征

^{*} 国家自然科学基金 (批准号: 60574025, 60740430664) 和高等学校博士学科点专项科研基金 (批准号: 20070487052) 资助的课题.

[†] E-mail: gaoxiawang@163.com

谱、负特征谱与网络社团结构性质的关系，基于模块矩阵给出了反映社团结构性质的相关定义，在此基础上定义了基于模块矩阵的多个特征值及其特征向量的社团结构化指标。这种指标反映了节点对所属社团能够具有的贡献值，同时也反映了对相应社团的依附强度。利用人工网络与大量实际网络，将社团结构化指标与几种经典的中心化指标就节点的分辨率、指标的相关性、指标对节点的排序结果等进行了比较与分析。

2. 模块函数、模块矩阵及其分解

用 $G(V, E)$ 表示一个有 n 个节点， m 条边的无向无权无自连边的网络。 V 为顶点集， E 为边集，网络的结构用 $n \times n$ 邻接矩阵 A 表示，当节点 i 与 j 之间有边相连时，矩阵对应的项 A_{ij} 等于 1，否则，其值为 0。定义 G 的一个划分为 V 的非空子集构成的集合 $T = \{G_1, G_2, \dots, G_c\}$ ，对每个节点 $i \in V$ ，存在唯一的 G_k ， $1 \leq k \leq h$ ，使得 $i \in G_k$ 。同时，在划分 T 下，各节点之间保持原有的连接关系不变。节点 i 与子集合 G_k 之间的关系可用一个 $n \times c$ 维的 0—1 指标矩阵 $X = (x_1, x_2, \dots, x_c)$ 表示。列向量 x_k 为对应第 k 个子集合的指标向量。 $x_{ik} = 1$ 表明 $i \in G_k$ ，也即 $G_k = \{i | x_{ik} = 1\}$ 。由于每个节点只能划分到一个社团，矩阵 X 的每一行只有一个元素为 1，同时，各指标向量是相互正交的。

模块函数 $Q(T)$ 是相对于网络的一种划分 T 的函数，其值是划分 T 下所对应的子集合内部的边数的比例减去在同样的划分下这些边的比例的期望值^[5,6]。其中边的比例的期望值可以通过选取一个空模型 (null model)，对这个模型按它与所讨论的网络具有同样的结构求出。对划分 $T = \{G_1, G_2, \dots, G_c\}$ ， $Q(T)$ 可以表示如下^[5]：

$$Q(T) = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(g_i, g_j), \quad (1)$$

这里， A_{ij} 是邻接矩阵的元素， P_{ij} 表示的是空模型中节点 i 和节点 j 之间有边相连的概率。 g_i 表示第 i 个节点所属的集合， δ 是 Kronecker 记号，如果 i, j 属于同一个集合，则 $\delta(g_i, g_j) = 1$ ；否则， $\delta(g_i, g_j) = 0$ 。

$Q(T)$ 的值可用以衡量相应划分下网络是否具有社团结构属性。通常认为 $Q(T)$ 取值大于 0.3 时，网络具有社团结构，此时，划分所得到的集合称为网络的一个社团。并且 $Q(T)$ 的值越高则社团

化程度越高^[5,6]。已有文献只考虑 $Q(T)$ 取正值的情形，当 $Q(T)$ 取负值时，本文认为网络对该划分呈现出反社团结构，为方便起见，仍将划分所得到的集合称为网络的一个社团。而 $Q(T)$ 取值为小于 0.3 的正数时，相应划分下网络不具有社团结构。

模块函数表达式中，空模型的选取可以有不同的形式，最简单也是最常用的一种形式是设概率矩阵为 $P = \frac{kk^T}{2m}$ 。这里， k 是以各节点的度 k_i 为元素构成的 n 元向量。模型矩阵的这种选取方式能够保证模型网络与所研究的网络的各个顶点有相同的顶点度。记 $M = A - P$ ， M 即为模块矩阵。易知 M 是一个实对称矩阵。利用指标矩阵及其指标向量，应有

$$\delta(g_i, g_j) = \sum_{k=1}^c x_{ik} x_{jk}, \text{ 从而 } Q \text{ 可以进一步表示为}$$
$$Q(T) = \frac{1}{2m} \text{Tr}(X^T M X). \quad (2)$$

将 M 的 n 个特征值按由大到小的次序排列，记为 $\beta_1, \beta_2, \dots, \beta_n$ ，与此相应的特征向量记为 $u_1, u_2, \dots, u_n$ ，把 M 表示成 $M = U D U^T$ ，这里 $U = (u_1, u_2, \dots, u_n)$ 是由 M 的特征向量构成的特征矩阵， D 是由特征值构成的对角矩阵 $D_{ii} = \beta_i$ 。方程 (2) 可以转换为

$$Q(T) = \sum_{j=1}^n \sum_{k=1}^c \beta_j \langle u_j, x_k \rangle, \quad (3)$$

式中 $\langle \cdot \rangle$ 表示向量的数量积。此式将 $Q(T)$ 表示成模块矩阵的特征值 $\beta_j (j = 1, \dots, n)$ 与相应特征向量 u_j 到对应指标向量 $x_k (k = 1, \dots, c)$ 的投影的各种可能的和的形式。此式也表明，通过选取恰当的指标向量，或者说，对网络进行恰当的划分，可以使第一正特征值 β_1 在所有正特征值中对 Q 的正值的贡献最大，同样，对网络的恰当划分也能使第一负特征值 β_n 对 Q 的负值贡献最大。反之，对应某种划分，当 Q 取到较大的正值或较小的负值时，在所有的特征值中，第一正特征值或负特征值所做出的正的或负的贡献也较大。 β_1, β_n 的绝对值的大小与网络 Q 值的变化范围有密切关系。

3. 模块矩阵的特征谱与网络社团结构

3.1. 网络的社团结构占优与反社团结构占优

对给定的网络 G ，可以有不同的划分，因而网络的社团结构与反社团结构可能同时存在。即使是

直观上具有很强社团结构的网络,也可以找出反社团结构.考虑由 100 个节点构成的网络,其中顺次每 10 个节点构成一个完全连接图,除此之外没有其他连接关系.对该网络,分别考虑划分

$$T1: \{ \{1, 2, \dots, 10\}, \{11, 12, \dots, 20\}, \dots, \{91, 92, \dots, 100\} \},$$

$$T2: \{ \{1, 11, 21, \dots, 91\}, \{2, 12, 22, \dots, 92\}, \dots, \{10, 20, \dots, 100\} \}.$$

相对于 $T1$, 网络表现出很强的社团结构特征,而相对于 $T2$, 则表现出很强的反社团结构.然而,尽管 $T1$ 的内部连接边数与 $T2$ 的外部连接边数相同,但相对于 $T1$ 的各社团内部的完全连接, $T2$ 的外部连接显然要稀疏一些.计算两种划分下的 Q 值, $Q(T1) = 0.9, Q(T2) = -0.1$, 因此,尽管网络社团结构与反社团结构共存,利用 Q 值这一衡量标准,可以认为划分 $T1$ 所具有的社团结构优于划分 $T2$ 所具有反社团结构.计算该网络的模块矩阵的特征值,第一正特征值为 9, 第一负特征值为 -1, 第一正特征值要大于第一负特征值的绝对值.

对同一网络,通过计算各种划分所对应的 Q 值来进行比较,这是一项很繁琐的工作,但网络给定之后,不管如何划分,网络模块矩阵的特征值不变.由(3)式所反映出的 Q 值与第一正负特征值的联系,给出网络的社团占优与反社团占优的定义.

定义 1 对上述网络 $G(V, E)$ 的模块矩阵 M 及 n 个按由大到小的次序排列的特征值 $\beta_1, \beta_2, \dots, \beta_n$, 若对某个 $\varepsilon > 0, \beta_1 - |\beta_n| > \varepsilon$, 则称网络社团结构占优;若 $|\beta_n| - \beta_1 > \varepsilon$, 则称网络反社团结构占优;当 $|\beta_1 - |\beta_n|| \leq \varepsilon$ 时,网络社团结构、反社团结构(如果存在的话)均不占优.定义中 ε 是一个可以充分小的正数,可根据其值的大小设定一种判断的严格尺度.

考虑社团结构识别问题文献中普遍采用的由计算机生成的 128 个节点的人工网络,将网络人为划分为 4 个社团,每个社团内的节点数均为 32 个.选取合适的概率,使各个社团保持同样的内部节点平均度 k_{in} 和外部节点平均度 k_{out} , 且使每个节点的度的期望值都为 $k_{in} + k_{out} = 16$. 图 1 为设定的 4 个社团的节点的内部边的平均值从 0 变化到 16 时,各个网络对应的第一正负特征值,图中每个数据是 30 次仿真结果得到的平均值.为便于比较,图中画出了第一负特征值的绝对值.对于该社团结构已知的网络,在社团结构、反社团结构均不占优的情形,

网络实际上是随机网络.在 k_{in} 的值从 0 变化到 16 的过程中,网络结构表现出反社团占优 $\rightarrow$ 随机 $\rightarrow$ 社团结构占优的变化. k_{in} 很小时,此时网络的反社团结构占优; k_{in} 的值在 5, 6 附近时,网络可认为是随机网络;当 $k_{in} \geq 8$ 时,网络有明显的 4 个社团构成.由图 1 中通过第一正负特征值得到的上述结论与通过网络构造原理得到的结论是相符的.此外,注意到随着 k_{in} 的值的增大,社团结构越来越明显, $\beta_1 - |\beta_{128}|$ 的值也在增大.

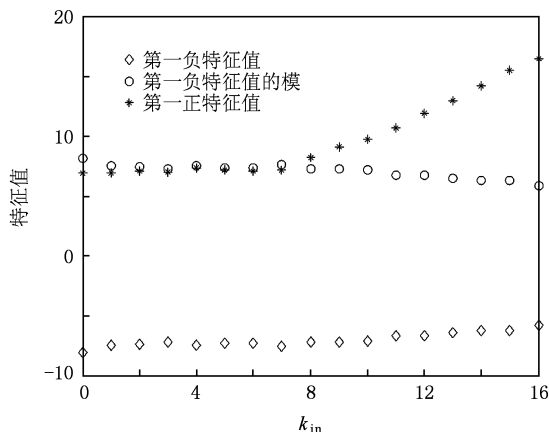


图 1 计算机生成的网络,当 k_{in} 从 0 变化到 16 时,第一正负特征值的变化情况

3.2. 网络特征值的领先集

对于一个具有明显的 c 个社团的网络采用基于 Laplacian 矩阵的谱方法时,相应的 Laplacian 矩阵有一个零特征值,在零特征值的附近存在 $c - 1$ 比 0 稍大一点的特征值,且节点的指标向量 $\mathbf{x}_1, \dots, \mathbf{x}_c$ 能够近似地由这 $c - 1$ 个特征值相应的特征向量线性表出^[14,19,20].

考虑网络的模块矩阵,仿真结果表明,模块矩阵的特征值也有类似的特征.图 2 为上述 128 节点的网络的特征值示意图(为节省篇幅,我们仅选取有代表性的 4 幅图, k_{in} 取其他值的情形,也都表现出相似的特征). $k_{in} = 0$ 时,3 个最小的负特征值与其他特征值之间存在一定的间距.当 $k_{in} = 4$ 时,网络的反社团结构微弱占优,其最小的负特征值在数值上区别于其他特征值.对社团结构明显占优的网络,如图 2(c),(d) 所示,存在着 3 个较大的正特征值,其数值上明显高于其他特征值.随着 k_{in} 增加,这种趋势愈明显.由此,定义模块矩阵的特征值的领先集.

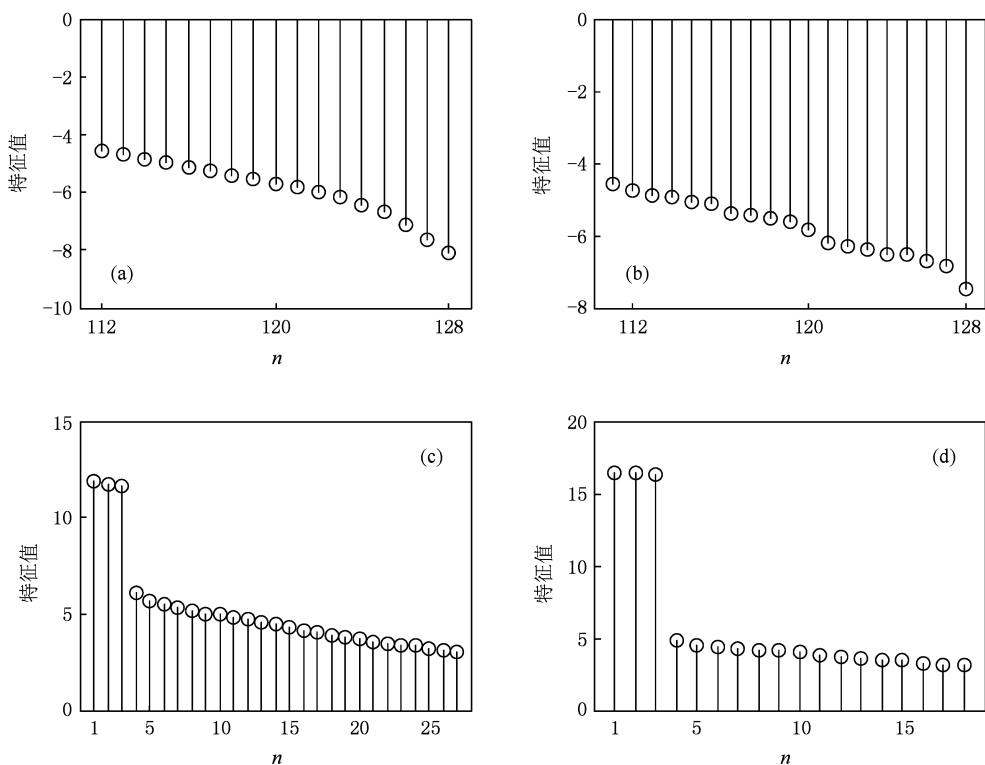


图2 计算机生成的网络, 对应于不同 k_{in} 时, 网络模块矩阵特征值的领先集示意图 (a) $k_{in} = 0$, (b) $k_{in} = 4$, (c) $k_{in} = 12$, (d) $k_{in} = 16$

定义 2 对上述网络的模块矩阵 M 及其特征值 $\beta_1, \beta_2, \dots, \beta_n$. 1) 若网络为社团占优的网络, 通过定距聚类, 对正特征值进行分组, 称处于第一集团, 数值上明显大于其他特征值的正特征值构成的集合为社团占优网络的领先集, 用符号 L^+ 表示, 其集合的个数记为 $|L^+|$. 2) 若网络为反社团占优的网络, 通过定距聚类, 对负特征值的模进行分组, 称处于第一集团, 其绝对值明显大于其他特征值的绝对值的负特征值构成的集合为反社团网络的领先集, 用符号 L^- 表示, 其集合的个数记为 $|L^-|$.

在图 2 所示的计算机生成的网络中, $k_{in} = 0, 4, 12, 16$, 领先集分别为 $\{\beta_{128}, \beta_{127}, \beta_{126}\}, \{\beta_{128}\}, \{\beta_1, \beta_2, \beta_3\}, \{\beta_1, \beta_2, \beta_3\}$. 注意到 $k_{in} = 4$ 的情形比较特殊. 通过对 Q 的计算表明, 当网络划分为 2 个社团时, Q 取得最小值 -0.012 , 与上述结果是符合的. 此例表明, 模块矩阵在反映网络的社团结构方面表现出与 Laplacian 矩阵类似的性质, 对社团占优的网络, 有 $|L^+| + 1$ 个明显的社团, 同样, 对反社团占优的网络, 有 $|L^-| + 1$ 个明显的反社团.

4. 基于模块矩阵的社团结构指标

4.1. 特征值与节点向量对模块函数的贡献

定义 2 涉及到多个特征值, 不妨假设模块矩阵有 p 个正特征值, q 个负特征值. 对节点 i , 分别利用正负特征值及其对应的特征向量, 可构造如下两个分别为 p 维和 q 维的节点行向量 s_i 和 t_i ^[5]:

$$[s_i]_j = \sqrt{\beta_j} U_{ij} \quad (j = 1, \dots, p), \tag{4}$$

$$[t_i]_j = \sqrt{-\beta_{n+1-j}} U_{i, n+1-j} \quad (j = 1, \dots, q). \tag{5}$$

再根据节点所属的社团, 引入社团向量 S_k 和 T_k 为

$$S_k = \sum_{i \in G_k} s_i, \quad T_k = \sum_{i \in G_k} t_i. \tag{6}$$

模 $|S_k|$ 可以表示为

$$\begin{aligned} |S_k| &= \frac{\langle S_k, S_k \rangle}{|S_k|} \\ &= \left\langle \sum_{i \in G_k} s_i, \frac{S_k}{|S_k|} \right\rangle \\ &= \sum_{i \in G_k} \langle s_i, \frac{S_k}{|S_k|} \rangle = \sum_{i \in G_k} \text{Pr}_j^\wedge S_k s_i, \end{aligned} \tag{7}$$

这里, $\hat{S}_k$ 是与 S_k 同向的单位向量. 这样, 每一个节点行向量对 $|S_k|$ 所作的贡献等于它到 $\hat{S}_k$ 的投影.

又由 $\text{Prj}_{\hat{S}_k} s_i = |s_i| \cos \theta_{ik}$, 这里 θ_{ik} 是 s_i 与 S_k 之间的夹角. $|s_i|$ 反映了节点 i 能对社团作出的正的最大贡献值, $\cos \theta_{ik}$ 是对节点在社团中所处的位置, 即节点是位于社团的“中心”还是“外围”的一种精确测量. 节点处于社团的“中心”, 也即节点向量与社团向量平行的时候, $|s_i|$ 是节点向量所能作出的最大的贡献值. 如果节点处于社团的“外围”, 则模 $|s_i|$ 的值即使很大实际上也有可能对模块仅作出较小的贡献. 行向量 t_i 与社团向量 T_k 之间也有类似的关系与结论.

4.2. 社团结构中心化指标

在引入节点行向量与社团向量之后, 模块方程 (2) 可以写为^[5]

$$Q = \sum_{k=1}^c |S_k|^2 - \sum_{k=1}^c |T_k|^2. \quad (8)$$

此式将 Q 的值表示成了由正特征值构成的社团向量与由负特征值构成的社团向量的模的差. 从而将特征值与特征向量对模块的正的贡献与负的贡献区分开来. 同时, 该式也清楚地表明, Q 的正值是由模块矩阵的正特征值及其对应的特征向量取得的, 而这种正的贡献来自于那些正特征值对应的特征向量中较大分量元素所对应的节点, 负的贡献则刚好相反.

对网络中的每个节点 i , 都有对应的节点行向量 s_i 和 t_i , 因而可以取适当维数的节点行向量, 定义一种基于模块矩阵的社团结构指标. 对社团结构占优的网络, 主要考虑由正特征值及其对应特征向量构成的节点向量 s_i , 而对反社团结构占优的网络, 则选取节点向量 t_i . 无论是 s_i 还是 t_i , 其模值都反映了节点所具有的对 Q 值的可能的最大贡献能力.

在近似处理中, 忽略除模值最大的特征值对应的最大特征向量之外的其他特征向量, 将节点向量对相应社团向量的投射缩减到最大特征向量的元素的模^[18]. 对社团结构占优的网络, 如果这个值较大, 若将对应的节点放在其他的社团, 将会损失对模块的贡献值, 这也说明该节点是相应社团的主要成员. 该值实际上从社团结构意义上明确了社团内部的中心节点. 这种近似实际上是非常有效的, 尤其是对大型的网络^[18]. 更进一步地, 结合前面领先

集的定义, 可以考虑更多的特征值及其相应的特征向量, 以得到更精确的刻画.

定义 3 对上述网络 $G(V, E)$ 的模块矩阵 M , 特征值 $\beta_1, \beta_2, \dots, \beta_n$, 以及节点行向量 $[s_i]_j (j = 1, \dots, p)$ 和 $[t_i]_j (j = 1, \dots, q)$. 1) 对社团结构占优, 领先集为 L^+ 的网络, 取 s_i 中前 $|L^+| + 1$ 个分量构成节点行向量, 定义节点的社团结构指标 (用 C^M 表示) 为归一化之后的节点行向量 s_i 的模, 对节点 i , 有 $C_i^M = \frac{|s_i|}{\max_j |s_j|}$. 2) 对反社团结构占优, 领先集为 L^- 的网络, 取 t_i 中前 $|L^-| + 1$ 个分量构成节点行向量, 定义节点的社团结构指标 (仍然用 C^M 表示) 为归一化之后的节点行向量 t_i 的模, 即对节点 i , 有 $C_i^M = \frac{|t_i|}{\max_j |t_j|}$.

定义 3 实际上定义了网络的一种中心化测度, 其基本意义是网络中的节点, 由于其在网络中所处的位置, 有能力对网络的整个模块作出正的或负的贡献. 这种贡献包含在与节点相应的元素的大小上.

5. 计算实例与讨论

5.1. 与人工网络的比较

本节利用典型的人工网络, 将本文所定义的社团中心化指标 C^M 与其他指标就对网络中节点的分辨能力进行比较. 符号 C^D , C^C , C^B , C^E 分别代表归一化的节点度、紧密度、介数中心化指标及特征向量中心化指标. 其中特征向量中心化指标指的是网络邻接矩阵的第一正特征向量, 也称之为特征向量构成的中心化指标^[21,22]. 其他中心化指标的定义见文献[23,24].

作为一种结构性指标, 本文所定义的社团中心化指标 C^M 具有较好的分辨能力. 考虑文献[25]中所讨论的两个典型例子 (见图 3). 图 3 所示网络为两个规则图, 图 3(a) 中各节点度均为 8, 图 3(b) 中各节点度均为 9. 通过计算, 在图 3(a) 中, 所有节点都有相同的 C^D (0.429), C^C (0.636) 以及 C^E (0.354), 意味着以这三种指标衡量, 各节点在图中的地位或重要性相同. 然而, 根据它们所参与的基本子图, 如边、三角形、四边形的数目上的差异, 图中存在三个节点可区分的组 $G_1 = \{1, 2, 8\}$, $G_2 = \{4, 6\}$, $G_3 = \{3, 5, 7\}$ ^[25]. 介数指标 C^B 能将三

个组区分开来，各组节点指标值分别为 0.095, 0.111, 0.071. C^M 指标同样能够将这三个组区分开来，指标值分别为 0.187, 0.854, 1.000. 但指标值

的排序与 C^B 不同,在 C^M 指标表示的是节点对所处的社团的依附强度这一意义下, G_2 中的节点具有最高的中心化值.

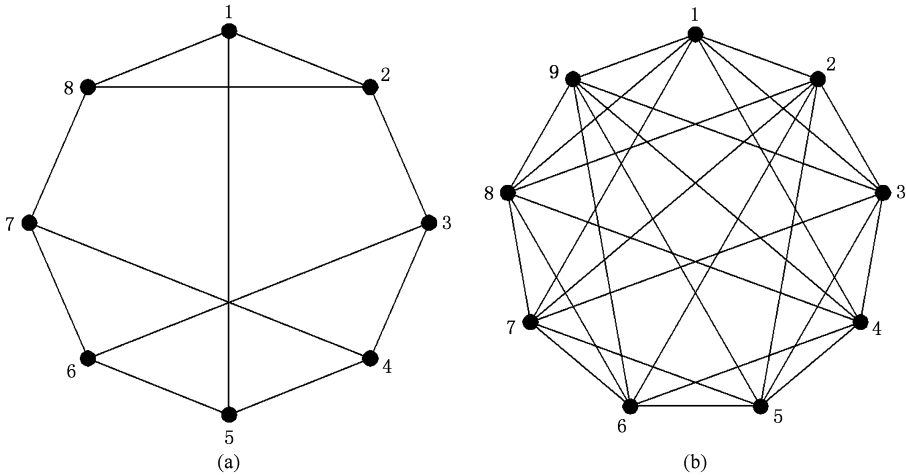


图 3 两个规则网络 (a)各节点度均为 8,(b)各节点度均为 9

图 3(b)是更为典型的例子. 图中所有的节点都有相同 C^D (0.750), C^B (0.036), C^C (0.800) 以及 C^E (0.333). 从节点参与构成子图的情况来看, 每个节点参与构成的三角形的数目也都相同. 然而 $G_1 = \{1, 3, 5, 6, 8\}$ 中各节点参与构成了 100 个四边形中的 44 个, $G_2 = \{2, 4, 7, 9\}$ 中各节点参与了 45 个. 因此, 这两个节点组仍然存在着细微的差别^[25]. 而本文定义的 C^M 指标能将这一差别显现出来. 计算结果表明, 对 G_1 中各节点, 有 $C^M = 0.800$, 而对 G_2 中各节点, 则有 $C^M = 1.000$. 这一结果与各节点参与构成子图的个数的排序结果相同, 表明了 C^M 具有较好的节点区分能力.

5.2. 实际网络的例子

考察 7 个实际网络的社团结构中心化测度. 1) 美国离婚人群生活状态网络 (divorce in US): 该网络为二模图, 59 个节点中 50 个节点代表美国的州名, 其余 9 个节点代表各州离婚人群的基本状况. 2) 海豚网 (dolphin): 社会关系网络, 表示的是居住在新西兰疑惑峡湾同一群体的 62 头海豚的交往情况^[26]. 3) 美国政治书籍网 (polbooks): 节点代表亚马逊网络在线书商 (Amazon. com) 售出的政治书籍, 边代表经常被同一顾客同时购买的书籍^[6]. 4) 美国航空网络 (US Air lines): 表示的是美国 1997 年的航空网络数据, 节点代表机场, 边则代表

航班. 5) “Graph products” 文献网 (Graphproducts): 该网络为作者-文章二模图, 根据 Imrich W, Klavžar S 1999 年出版的“Graph products”一书的参考文献编撰而成. 6) Erdos 合作网 (Erdos): 代表以数学家 Erdos 为核心的科学家合作网络, 节点代表作者, 有边相连代表彼此有过合作关系. 7) 美国西部电力网络 (USPower Grid): 节点代表发电机、变压器、变电站等电力设施, 边则代表设施之间高压传输线^[27].

上述网络中, 对于有向图, 本文忽略了方向并且在现有的分析中将其做为无向网络来处理. 另处, 对于 Erdos 网络作了简化, 只研究了网络的主要部分. 本文使用的数据来自 Pajek 程序数据库 (<http://vlado.fmf.uni-lj.si/pub/networks/data/>).

5.2.1. 网络的社团结构属性与领先集

表 1 中显示了七个实际网络的基本参数, 网络模块矩阵的第一正负特征值及其领先集. 根据定义 1, 美国离婚网络、美国航空网络、Graph products 文献网络以及 Erdos 合作网络为反社团占优的网络, 其中, 美国航空网络的第一正负特征值的绝对值的差非常小, 所以其反社团的属性并不是很明显. 海豚网络、美国政治书籍网络以及美国西部电力网络则为社团结构占优的网络. 根据定义 2 得到的各个网络的领先集在表 1 中最后一列列出.

表 1 七种实际网络的领先集

网络	节点数 n	边数 m	β_1	β_n	领先集
美国离婚网	59	225	5. 041	- 16. 106	$L^- = \{\beta_n\}$
海豚网	62	159	6. 258	- 3. 889	$L^+ = \{\beta_1\}$
美国政治书籍网	105	441	11. 695	- 5. 427	$L^+ = \{\beta_1\}$
美国航空网	332	2126	17. 882	- 18. 893	$L^- = \{\beta_{n-1}, \beta_n\}$
Graphproducts 文献网	674	613	5. 198	- 6. 149	$L^- = \{\beta_n\}$
Erdos 合作网	2155	7078	17. 819	- 29. 300	$L^- = \{\beta_n\}$
美国西部电力网	4941	6594	7. 390	- 4. 499	$L^+ = \{\beta_1, \beta_2\}$

5. 2. 2. 与其他中心化测度的比较

不同的中心化测度可能呈现出很强的相关性^[28], 考虑到各种中心化测度是对节点在网络中的侧重于某一方面的地位、能力或声望的一种量化, 这也是一个自然的结论. 例如, 高度值的节点通常参与构成了节点之间的最短路径, 从而使节点的度与几种中心化测度之间具有高度相关性. 从网络所具有社团属性来讲, 高度值节点在所处的社团内具有更大的直接影响力, 对所处的社团的贡献度也相应较大, 因而也往往具有较大的社团结构中心化值.

表 2 列出了七种实际网络的五种中心化指标的平均值. 表中 R^2 是相应的中心化测度与 $\overline{C}^M$ 的线性回归的平方相关系数. 如表 2 中所示, 与 $\overline{C}^M$ 相关程度最高的是 $\overline{C}^D$, 其次是 $\overline{C}^E$, 表明平均来看, 网络中度值较大的节点同样是相应社团的核心. 此外, 由于特征向量中心化基于邻接矩阵的主特征向量, 而模块矩阵是在邻接矩阵的基础上构造的, 所以彼此之间的联系也十分紧密.

表 2 七种实际网络的中心化值

网络	$\overline{C}^D$	$\overline{C}^B$	$\overline{C}^C$	$\overline{C}^E$	$\overline{C}^M$
美国离婚网	0. 132	0. 021	0. 463	0. 108	0. 361
海豚网	0. 084	0. 039	0. 307	0. 091	0. 247
美国政治书籍网	0. 081	0. 020	0. 330	0. 070	0. 267
美国航空网	0. 039	0. 005	0. 376	0. 031	0. 178
Graphproducts 文献网	0. 003	0. 002	0. 020	0. 010	0. 032
Erdos 合作网	0. 003	0. 001	0. 306	0. 011	0. 038
美国西部电力网	0. 001	0. 004	0. 054	0. 001	0. 009
R^2	0. 970	0. 563	0. 657	0. 924	

5. 2. 3. 对节点的排序

中心化测度的一个最重要的作用是根据所考虑的网络的拓扑性质来对网络中的节点进行排序, 从而给出对节点的重要性的一种评价. C^D 表示的是

节点对其邻居节点所施加的直接影响力. C^E 由于基于邻接矩阵, 实际上考虑了所有网络路径而避免了对路径的穷举搜索^[22], 因而能够表示节点传输信息的能力, 在社会网络分析中, 也表示了个体受欢迎的程度. C^M 基于网络的模块矩阵, 进一步排除了随机因素产生的影响, 反映了节点对社团潜在的最大贡献能力. 尽管不同的指标是有差异的, 但两种指标可能在网络中的重要节点的认定上表现出一定的相似性. Erdos 合作网中, 根据 C^D 计算的前 10 个重要节点中, 有 9 个与 C^M 的排序结果是相同的. 对于美国航空网络和政治书籍网络, 相同节点的数目均为 7. 在政治书籍网络中, 连接数目最多的是“A National Party No More”与“Off with Their Heads”, 两者的连接度相同. 这也是由 C^M 指标所显示的右翼色彩最浓的两本书, 但前者具有更大的指标值. 而由这一指标所确定的左翼色彩最浓的书为“Bushwhacked”. 而指标 C^E 则将“Bushwhacked”排在第一, 排在第二的是“American Dynasty”. 可见, 各种指标在有相似性的同时也有不同的地方.

6. 结 论

本文结合复杂网络社团结构的研究, 讨论了模块矩阵的特征谱中, 正负特征值及其相应特征向量对模块函数的贡献值与网络的社团结构属性的关系, 据此定义了网络的社团占优与反社团占优、模块矩阵的领先集等概念. 在此基础上提出了一种网络节点的中心化指标. 选取领先集中的矩阵特征值及其对应特征向量构成节点行向量, 取各行向量的模并进行归一化处理得到相应节点的中心化指标值, 这种中心化指标反映了节点对所处的社团所可能具有的最大贡献能力, 因而称之为社团中心化指标.

在对典型的人工网络进行的分析测度中, 相对

于度指标、介数指标、紧密度指标和特征向量指标这几种经典的中心化指标，社团中心化指标显示出较强的分辨能力。对七种实际网络的中心化指标值的分析结果表明，社团中心化指标与度指标和特征向量指标表现出较强的相关性，与介数指标和紧密

度指标的相关性不明显，对个体节点的排序表现出不同的特点。

感谢审稿人对本文提出的宝贵意见和建议，文中采用了审稿人举的一个例子，特此申明并表示感谢。

[1] Girvan M, Newman M E J 2002 *Proc. Natl. Acad. Sci. USA* **99** 7821

[2] Newman M E J 2004 *Phys. Rev. E* **69** 066133

[3] Newman M E J 2004 *Eur. Phys. J. B* **38** 321

[4] Clauset A, Newman M E J, Moore C 2004 *Phys. Rev. E* **70** 066111

[5] Newman M E J 2006 *Phys. Rev. E* **74** 036104

[6] Newman M E J 2006 *Proc. Natl. Acad. Sci. USA* **103** 8577

[7] Du H F, Li S Z, Marcus W F, Yu Z S, Yang X S 2007 *Acta Phys. Sin.* **56** 6886 (in Chinese) [杜海峰、李树苗、Marcus W F、悦中山、杨绪松 2007 物理学报 **56** 6886]

[8] Zhang G W, Kang J C, Xia C L, Li H S 2006 *Comput. Sci.* **33** 1 (in Chinese) [张光卫、康建初、夏传良、李鹤松 2006 计算机科学 **33** 1]

[9] Liu Y, Zhang L 2008 *Acta Phys. Sin.* **57** 5419 (in Chinese) [刘 云、张 立 2008 物理学报 **57** 5419]

[10] Gao Z K, Jin N D 2008 *Acta Phys. Sin.* **57** 6909 (in Chinese) [高忠科、金宁德 2008 **57** 6909]

[11] Pothen A, Simon H D, Liou K P 1990 *SIAM J. Matrix Anal. Appl.* **11** 430

[12] Seary A J, Richards W D 1995 *Proceedings of the International Conference on Social Networks London* **1** 47

[13] Capocci A, Servedio V D P, Caldarelli G, Colaioni F 2005 *Phys. A* **352** 669

[14] White S, Smyth P 2005 *Proceedings of the 5th SIAM International Conference on Data Mining Philadelphia* p274

[15] Barber M J 2007 *Phys. Rev. E* **76** 066102

[16] Wang G X, Shen Y, Ouyang M 2008 *Computers and Mathematics with Applications* **55** 2746

[17] Pujol J M, Bejar J, Delgado J 2006 *Phys. Rev. E* **74** 016107

[18] Wang G X, Shen Y, Luan E J 2008 *Prog. Natl. Sci.* **18** 1043

[19] Hagen L, Kahng A B 1992 *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **11** 1074

[20] Wang X F, Li X, Cheng G R 2005 *Theory and Application of Complex Networks* (Beijing: Tsinghua University Press) (in Chinese) [汪小帆、李 翔、陈关荣 2005 复杂网络理论及其应用 (北京:清华大学出版社)]

[21] Bonacich P 1972 *J. Math. Sociol.* **2** 113

[22] Bonacich P 1987 *Am. J. Sociol.* **92** 1170

[23] Wasserman S, Faust K 1994 *Social Networks Analysis: Methods and Applications* (Cambridge: Cambridge University Press)

[24] Scott J 2000 *Social Network Analysis: A Handbook* (London: Sage Publications)

[25] Estrada E, Rodriguez-Velazquez J A 2005 *Phys. Rev. E* **71** 056103

[26] Lusseau D, Schneider K, Boisseau O J, Haase P, Slooten E, Dawson S M 2003 *Behavioral Ecology and Sociobiology* **54** 396

[27] Watts D J, Strogatz S H 1998 *Nature* **393** 440

[28] Wuchty S, Stadler P F 2003 *J. Theor. Biol.* **223** 45

Modularity matrix of networks and the measure of community structure^{*}

Wang Gao-Xia^{1)2)†} Shen Yi²⁾

1) (*Science College, China Three Gorges of University, Yichang 443002, China*)

2) (*Institute of Systems Engineering, Huazhong University of Science and Technology, Wuhan 430074, China*)

(Received 20 June 2008; revised manuscript received 3 June 2009)

Abstract

The relationships between positive (negative) eigenspectrums and the structure properties of community (anti-community) of complex networks are investigated, and some corresponding definitions are given. By using the multieigenspectrums of modularity matrix of networks, a kind of structural centrality measure called the community centrality, is introduced. This individual centrality measure describes the strength that the individual adheres to the corresponding community. The measure is illustrated and compared with the standard centrality measures using several artificial networks and real world networks data. The results show that the community centrality has better discrimination, and it has positive correlation with the degree centrality.

Keywords: networks, modularity matrix, community structure, centrality

PACC: 0590, 0175

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 60574025, 60740430664) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20070487052).

[†] E-mail: gaoxiawang@163.com