

# 复杂网络中社团结构发现的多分辨率密度模块度\*

张聪<sup>1)†</sup> 沈惠璋<sup>1)</sup> 李峰<sup>1)</sup> 杨何群<sup>2)</sup>

1) (上海交通大学安泰经济与管理学院, 上海 200052)

2) (上海市卫星遥感与测量应用中心, 上海 201199)

(2011年9月7日收到; 2011年12月7日收到修改稿)

现实中的许多复杂网络呈现出明显的模块性或社团性. 模块度是衡量社团结构划分优劣的效益函数, 它也通常被用作社团结构探测的目标函数, 但最为广泛使用的 Newman-Girvan 模块度却存在着分辨率限制问题, 多分辨率模块度也不能克服误合并社团和误分裂社团同时存在的缺陷. 本文在网络密度的基础上提出了多分辨率的密度模块度函数, 通过实验和分析证实了该函数能够使社团结构的误划分率显著降低, 而且能够体现出网络社团结构是一个有机整体, 不是各个社团的简单相加.

**关键词:** 复杂网络, 社团结构, 模块度, 网络密度

**PACS:** 89.75.Fb, 89.75.Hc

## 1 引言

自然界和人类社会中的大量复杂系统可以被描述为网络或图, 系统的组成部分和它们之间的交互作用可视为网络中的节点和连边<sup>[1-6]</sup>, 例如: 互联网、社会网络、食物链网络、生物网络以及引文网络等<sup>[7-11]</sup>. 复杂系统通常呈现出群组或社团的组织形式, 不同的群组或社团表现出不同的功能或角色. 社团内节点间的连接比较紧密, 而社团间节点间的连接比较松散<sup>[12,13]</sup>.

社团结构发现是复杂网络研究领域中的一个重要课题. 由物理学、数学、计算机科学、生物学和社会学中的原理和方法发展出许多社团结构发现算法<sup>[14-25]</sup>. 根据这些算法可得到多种社团结构, 但是这么多种社团结构需要一个量化的标准来评估它们的优劣. 迄今为止, 最广泛使用的评价函数是 Newman 和 Girvan<sup>[26]</sup> 提出的模块度函数  $Q$  (下文称 NG 模块度). 该函数源于“随机网络不会具有

明显的社团结构”这样的思想, 通过衡量实际覆盖度 (覆盖度是社团内部边数占总边数的比例) 与随机连接情况下期望覆盖度的差异大小来评估社团结构<sup>[26]</sup>. 然而, Fortunato 和 Barthélemy<sup>[27]</sup> 研究证明模块度函数存在着分辨率限制问题, 以模块度最大化为目标, 可能无法发现一些规模较小的社团, 即当最大模块度值对应的社团结构中包含小于一定规模的社团时, 就不能确定这些社团是单独的社团或更小社团的弱连接合并. 还有一些相似的评论见文献<sup>[28, 29]</sup>. Li 等<sup>[30]</sup> 依据弱社团结构定义<sup>[31]</sup> 提出了一个模块化密度函数  $D$ , 该函数在一些网络的社团结构划分中没有表现出模块度  $Q$  的分辨率限制.

虽然模块度  $Q$  存在着分辨率限制问题, 但是由 Reichardt 和 Bornholdt<sup>[32]</sup> 以及 Arenas 等<sup>[33]</sup> 提出的多分辨率模块度被认为可以克服此限制, 这些多分辨率模块度试图通过对  $Q$  函数加入可调节参数来实现探测不同粒度社团结构的目的. 然而 Lancichinetti 和 Fortunato<sup>[34]</sup> 研究发现多分辨率

\* 国家自然科学基金 (批准号: 71071096, 71001068) 资助的课题.

† E-mail: z.c.cn@163.com

模块度无法准确地识别现实中大量存在的具有无标度特征的网络社团结构, 具体表现在多分辨率模块度既可能将一些相互连接的小社团合并识别为一个“大社团”, 又可能将大社团识别为其分裂而成的若干“小社团”, 而这两种情况基本上是无法同时避免的.

本文定义了与常用网络密度不同的自然密度, 进一步在网络密度矩阵的基础上提出了密度模块度函数、由不同系数矩阵形成的三种典型函数以及多分辨率密度模块度函数, 最后通过对函数的形式分析以及在随机网络和无标度网络上的实验, 说明本文的密度模块度相比 NG 模块度, 能够在不同的分辨率上得到更好的社团结构.

## 2 网络密度和密度模块度

设  $G = (V, E)$  是一个连通的无向网络,  $V$  是节点集,  $E$  是边集,  $k$  是节点的度,  $A = (a_{ij})$  是  $G$  的邻接矩阵. 节点个数  $n = |V|$ , 总边数  $L = |E|$ . 设节点集  $V_i, V_j \subseteq V$ , 则  $V_i$  到  $V_j$  的连接数  $l_{ij} = \sum_{u \in V_i, v \in V_j} a_{uv}$ . 给定一个划分  $P = \{G_1(V_1, E_1), \dots, G_m(V_m, E_m)\}$ , 其中  $V_i$  和  $E_i$  分别是  $G_i$  的节点集和边集 ( $i = 1, \dots, m$ ;  $n_i = |V_i|$ ;  $l_{ii} = 2|E_i|$ ), 且  $\bigcup_{i=1}^m V_i = V$ . NG 模块度  $Q$  在划分  $P$  的基础上定义为如下形式<sup>[17]</sup>:

$$Q(P) = \sum_{i=1}^m \left[ \frac{l_{ii}}{2L} - \left( \frac{d_i}{2L} \right)^2 \right], \quad (1)$$

其中  $d_i = \sum_{j=1}^m l_{ij}$  是  $G_i$  中所有节点的度之和,  $Q$  的取值范围为  $(-1, 1)$ . 模块度  $Q$  看似一个非常有效的评价函数, 但是当某些社团内部的连接数小于  $\sqrt{2L}$  时, 利用模块度  $Q$  将可能无法分辨这些社团, 这就使划分的社团结构处在一个粗劣的层次上.

本文提出的密度模块度函数是在网络密度的基础上定义的, 但用到的密度与常用的网络密度有所区别, 密度  $d(G) = \frac{2L}{n(n-1)}$  是网络中实际边数占最大可能边数的比率, 而本文的自然密度定义为网络中节点的平均度与节点数的比率, 如下式:

$$d_N(G) = \frac{\langle k \rangle}{n} = \frac{2L}{n^2}. \quad (2)$$

网络密度  $d$  与自然密度  $d_N$  在取值范围上是不同的, 网络密度的取值范围是  $[0, 1]$ , 而  $d_N$  的取

值范围为  $\left[0, \frac{n-1}{n}\right]$ . 假设两个网络  $G_1, G_2$  都是全连接网络 (派系), 且节点数  $n_1 > n_2 > 0$ , 则  $d(G_1) = d(G_2)$ ,  $d_N(G_1) > d_N(G_2)$ , 虽然都是派系, 但从每个节点来看,  $G_1$  比  $G_2$  的连接度更高. 显然密度  $d$  适合作为比较相同规模网络 (相同节点数的网络) 连接稠密程度的指标, 而自然密度  $d_N$  适合于相同规模网络 and 不同规模网络两种情况. 此外, 当网络只有一个节点时, 网络密度  $d$  不存在, 而自然密度  $d_N = 0$ , 这也使自然密度不需要做例外处理就可使用于各种网络或社团. 除了网络内部的密度, 还需要用到网络外部的密度. 两个子网络  $G_i, G_j$  之间的密度  $d(G_i, G_j) = \frac{l_{ij}}{n_i n_j}$  表示网络之间实际边数占最大可能边数的比率. 那么, 根据自然密度  $d_N$  和网络间密度  $d(G_i, G_j)$  可构造划分  $P$  的密度矩阵, 如下所示:

$$D(P) = (d_{ij})_{m \times m} = N^{-1} L N^{-1}, \quad (3)$$

其中  $d_{ij} = \frac{l_{ij}}{n_i n_j}$ , 对角线上的密度是自然密度, 非对角线的密度是网络间密度, 对角矩阵  $N = \text{diag}(n_1, \dots, n_m)$ , 矩阵  $L = (l_{ij})_{m \times m}$ . 分析划分  $P$  的密度矩阵可发现, 对网络进行划分得到社团结构, 其目标就是使社团 (子网络) 内部的密度大, 社团间的密度小. 那么当社团内部密度与外部密度的差距最大时, 即对应于网络社团结构的最佳划分. 由此, 给出本文的密度模块度函数:

$$Q_N(P) = \sum_i \left( w_{ii} d_{ii} - \sum_{j \neq i} w_{ij} d_{ij} \right), \quad (4)$$

$$Q_N(P) = 2\text{Tr}(\mathbf{W} \circ D(P)) - \|\mathbf{W} \circ D(P)\|. \quad (5)$$

两式分别为用元素和矩阵表达的形式, 其中  $\mathbf{W} = (w_{ij})_{m \times m}$  为系数矩阵,  $\circ$  代表矩阵的阿达马乘积 (Hadamard product),  $\text{Tr}$  代表矩阵的迹,  $\|\cdot\|$  代表矩阵的和范数. 密度模块度函数  $Q_N$  衡量社团内部的连接密度与社团之间的连接密度的差值, 从直观上看, 这个差值的大小反映了网络划分所得社团结构质量的优劣.

系数矩阵  $\mathbf{W}$  的取值对密度模块度函数的效果是有显著影响的. 本文讨论三种最典型的系数矩阵.

1) 等系数矩阵: 对每个元素取相等的系数  $w_{ij} = \frac{1}{m}$ , 此时明显不能反映出不同大小社

团对模块度贡献的差异.

$$Q_{N1}(P) = \frac{1}{m} \sum_i \left( d_{ii} - \sum_{j \neq i} d_{ij} \right). \quad (6)$$

2) 节点比例系数矩阵: 按照社团节点个数与网络总节点个数的比值取系数  $w_{ij} = \frac{n_j}{n}$ , 此时的  $Q_N$  与文献 [30] 中的模块度  $D$  是等价的, 虽然每个社团有了各自的系数, 但对于一个社团而言, 其内部密度和外部密度的系数却是相同的.

$$Q_{N2}(P) = \frac{1}{n} \sum_i \frac{l_{ii} - \sum_{j \neq i} l_{ij}}{n_i}. \quad (7)$$

3) 边比例系数矩阵: 以社团内或社团间连接数占总连接数的比例作为系数  $w_{ij} = \frac{l_{ij}}{2L}$ , 显然此时的系数更能体现出各个密度对模块度的不同贡献.

$$Q_{N3}(P) = \frac{1}{2L} \sum_i \left( \langle k_i \rangle^2 - \sum_{j \neq i} \frac{l_{ij}^2}{n_i n_j} \right), \quad (8)$$

其中  $\langle k_i \rangle$  是社团  $i$  的平均度.

从三个函数的形式来分析, 当网络中所有社团的节点数都相同时, (6) 式和 (7) 式是一致的, 而三个函数都可以根据社团内部与社团之间的密度差异, 得到社团内部连接紧密, 社团之间连接稀疏的社团结构; (8) 式与 (6) 式和 (7) 式的差异仅在于: 由于系数的作用使社团内部密度项  $Q_{N+} = \sum_i w_{ii} d_{ii}$  与社团之间密度项  $Q_{N-} = \sum_i \sum_{j \neq i} w_{ij} d_{ij}$  的差异增大.

当网络中社团规模大小不同时, 主要考虑是否误将大社团分裂为若干“小社团”或误将某些小社团合并为“大社团”; 因 (6) 式使用等系数矩阵, 只需考虑社团的密度大小, 当大社团本身的密度小于其内部某些局部的密度时, 存在分裂大社团的可能性, 当然也存在将一些密度较小的小社团合并的可能性, 那么, 得到的社团结构极有可能是大多数社团其规模小密度大, 极少数社团其规模大密度小; 对 (7) 式和 (8) 式而言, 因系数矩阵中的元素体现了社团规模大小对模块度值的不同贡献, 明显使上述“误分裂”或“误合并”的可能性大大降低.

最后, 对 (7) 式和 (8) 式进行比较, 前文提到它们形式上的差别在于单个社团其各个系数 (系数矩阵每列中的各个元素) 的异同, 此差异体现在网络中各个社团的相互关联和影响上. 不失一般性, 考虑某个社团  $x$  分裂为两个新社团  $x_1$  和  $x_2$ ,

在计算 (7) 式的模块度值时, 除社团  $x$  外的其他社团的模块度分量之和  $\sum_{i \neq x} \frac{l_{ii} - \sum_{j \neq i} l_{ij}}{n n_i}$  在分裂前后是不变的 ( (1) 式中的  $\sum_{i \neq x} \left[ \frac{l_{ii}}{2L} - \left( \frac{d_i}{2L} \right)^2 \right]$  也是如此 ), 而 (8) 式的其他社团的模块度分量之和  $\frac{1}{2L} \sum_{i \neq x} \left( \langle k_i \rangle^2 - \sum_{j \neq i} \frac{l_{ij}^2}{n_i n_j} \right)$  在分裂前后是不同的 ( 因  $\frac{l_{ix}^2}{n_i n_x}$  与  $\frac{l_{ix_1}^2}{n_i n_{x_1}} + \frac{l_{ix_2}^2}{n_i n_{x_2}}$  不同 ), 即某个社团的分裂影响到了其他社团的模块度分量; 同理, 社团合并时的情形也是如此. (8) 式所表现出的这种网络中各个社团的相互关联和影响, 体现了网络所描述的系统是一个有机的整体, 不是各个部分的机械组合或简单相加. 同时这种相互关联和影响无疑是其他模块度函数所不具备的优势. 从以上分析来看, (8) 式所表达的密度模块度函数比其他各式更好, 下文将通过一系列实验来进一步验证.

### 3 多分辨率密度模块度

在给出本文的多分辨率密度模块度之前, 先介绍 Reichardt 和 Bornholdt<sup>[32]</sup> 提出的多分辨率模块度  $Q_\lambda$ :

$$Q_\lambda(P) = \sum_{i=1}^m \left[ \frac{l_{ii}}{2L} - \lambda \left( \frac{d_i}{2L} \right)^2 \right], \quad (9)$$

其中  $\lambda$  是可调节参数,  $\lambda$  逐渐增加时, 分辨率提高, 可识别某些规模小的社团. 而 Lancichinetti 和 Fortunato 指出, 当  $\lambda$  较小时, 所识别的“大社团”可能是一些实际小社团的合并, 当  $\lambda$  较大时, “小社团”可能是由实际大社团分裂而成, 而且不存在两种可能同时避免的情况.

本文的多分辨率密度模块度函数与  $Q_\lambda$  的函数形式类似, 也是通过加入一个可调节参数  $\lambda$  来实现多分辨率识别社团结构的. 函数如下所示:

$$Q_{N\lambda}(P) = \sum_i \left( w_{ii} d_{ii} - \lambda \sum_{j \neq i} w_{ij} d_{ij} \right). \quad (10)$$

当参数  $\lambda$  逐渐增加时, 与  $Q_\lambda$  的变化趋势相反,  $Q_{N\lambda}$  的分辨率降低,  $\lambda$  逐渐减少时,  $Q_{N\lambda}$  的分辨率提高. 通过图例可解释这种变化趋势. 图 1 是  $\lambda = 1$  时依据  $Q_{N\lambda}$  和  $Q_\lambda$  两种模块度对具有社团结构性

质的网络 (形如图 2 所示) 划分后所得到的社团数目 - 模块度值示意图. 图 1(a) 中的三条曲线分别代表了  $Q_{N\lambda}$  及其社团内部密度项  $Q_{N+}$  和社团之间密度项  $Q_{N-}$  随社团数目的变化趋势, 当社团数目  $m = 1$  时, 由于  $Q_{N-} = 0$ , 因而  $Q_{N\lambda}$  与  $Q_{N+}$  相同, 当社团数目与节点数相同  $m = n$  时,  $Q_{N+} = 0$ , 而  $Q_{N-} = 1$ ,  $Q_{N\lambda} = -1$ . 分析图例可知, 当  $\lambda$  逐渐增加时,  $Q_{N-}$  项随之增加, 使  $Q_{N\lambda}$  取最大值时

所对应的社团数目变小; 反之, 当  $\lambda$  逐渐减小时,  $Q_{N-}$  项也减小, 使  $Q_{N\lambda}$  取最大值时所对应的社团数目变大. 图 1(b) 给出了  $Q_{\lambda}$  及其正负项的变化趋势, 当  $m = 1$  时,  $Q_+ = Q_- = 1$ , 那么  $Q_{\lambda} = 0$ ; 当  $m = n$  时,  $Q_+ = 0$ , 而  $Q_{\lambda} = -Q_-$ . 对比分析图 1(a) 和 (b) 两图, 明显可得出二者趋势相反的结论. 下文我们将通过在无标度网络上的实验, 验证密度模块度在参数变化时社团结构划分的准确性.

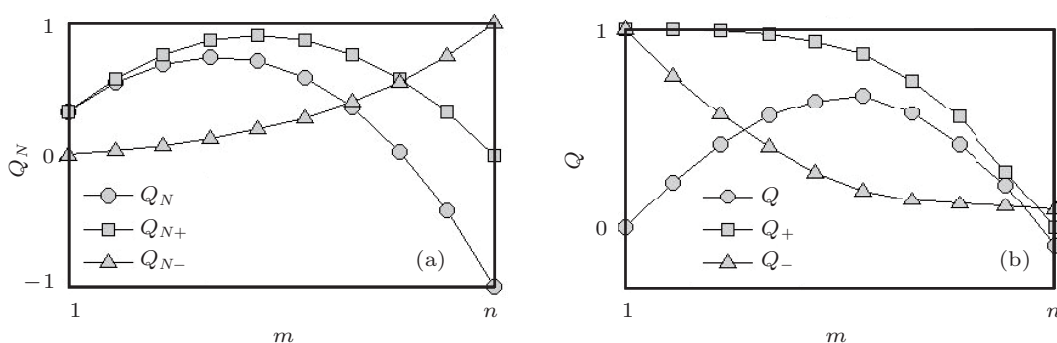


图 1 (a) 和 (b) 分别为  $Q_N$  和  $Q$  及其正负项的社团数目 - 模块度值示意图 (图中圆形、正方形和三角形分别代表模块度值及其正负项的值)

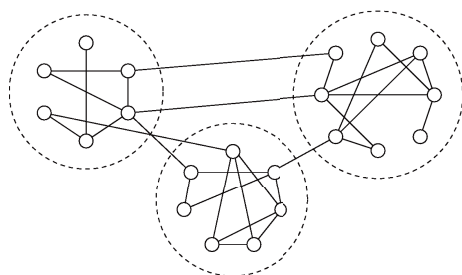


图 2 一个具有社团结构性质的网络示意图

## 4 数值实验及结果

我们分别在随机网络和无标度网络上进行实验. 由于此网络划分问题是 NP 完全问题, 本文以四种模块度为目标函数, 采用模拟退火算法对各网络进行社团结构划分.

### 4.1 随机网络

为了比较密度模块度和 NG 模块度发现社团结构的优劣, 我们进行了文献 [13, 28] 中描述的基础实验. 实验中 128 个节点被平均地分为 4 个组,

节点平均度为 16. 用适当的概率调节节点的组内平均连边数  $k_{in}$  和组间平均连边数  $k_{out}$ , 当节点与组外的平均连接数不断增长时, 发现正确的社团结构将变得越来越难.

本文实验的计算结果见表 1, 分别给出了不同  $k_{out}$  所对应的节点划分正确比率. 实验包含三组. 第一组: 网络包含 4 个社团, 每个社团的节点数相同, 随着  $k_{out}$  的增加, 划分的正确率降低, 四种模块度的正确率差异不大. 第二组: 网络包含两个社团, 一个有 96 个节点, 另一个有 32 个节点, 这种非对称网络使模块度  $Q$  和  $Q_{N1}$  难于得到正确的划分, 但是模块度  $Q_{N2}$  和  $Q_{N3}$  有很高的划分正确率, 这表明元素相等的系数矩阵和按社团大小比例取值的系数矩阵使密度模块度在非对称网络情况下的社团划分效果有显著差异. 第三组: 网络包含两个社团, 各有 64 个节点, 但是两个社团节点的平均度是不同的, 分别为 8 和 24, 当  $k_{out} = 2, 3, 4$  时, 密度模块度  $Q_{N2}$  和  $Q_{N3}$  明显能够以相当高的比率将节点正确划分, 而另两个模块度则越来越难得到正确的社团结构, 这与第二组实验表明的一致. 表 2 给出了四种方法识别的社团数量的正确率

和平均值, 从中同样可以发现在非对称情况下, 模块度  $Q_{N2}$  和  $Q_{N3}$  比  $Q$  和  $Q_{N1}$  有更高的准确性. 模块度  $Q$  的分辨率限制问题使其不能对非对称性强的网络做出准确的社团划分, 而  $Q_{N1}$  准确性低的原因主要在于其每个社团的系数是相等的, 一些局部连接紧密的节点更容易被识别为社团, 这样使得  $Q_{N1}$  倾向于将实际的社团分裂为若干小社团. 另外, 由于此实验社团数目较少, 使得  $Q_{N2}$  和  $Q_{N3}$  识别社团结构的差异不明显.

表 1 在 100 个网络上实验得到的节点划分正确率的

均值和标准差					
社团	$k_{out}$	$Q$	$Q_{N1}$	$Q_{N2}$	$Q_{N3}$
相同	6	0.99(0.01)	0.99(0.01)	0.99(0.01)	0.99(0.01)
	7	0.97(0.02)	0.97(0.02)	0.97(0.02)	0.98(0.02)
	8	0.89(0.05)	0.90(0.03)	0.91(0.03)	0.93(0.03)
节点数不同	6	0.85(0.04)	0.89(0.04)	0.99(0.01)	0.99(0.01)
	7	0.80(0.03)	0.87(0.05)	0.97(0.02)	0.98(0.02)
	8	0.74(0.05)	0.82(0.05)	0.93(0.04)	0.93(0.04)
边数不同	2	1.00(0.01)	0.98(0.01)	1.00(0.00)	1.00(0.00)
	3	0.96(0.03)	0.81(0.06)	1.00(0.01)	1.00(0.00)
	4	0.73(0.10)	0.68(0.15)	0.98(0.01)	1.00(0.00)

表 2 在 100 个网络上实验得到的社团数目的正确率和均值

社团	$k_{out}$	$Q$	$Q_{N1}$	$Q_{N2}$	$Q_{N3}$
相同	6	1.00(4.00)	0.78(4.91)	1.00(4.00)	1.00(4.00)
	7	1.00(4.00)	0.54(5.63)	1.00(4.00)	1.00(4.00)
	8	0.70(4.33)	0.00(8.31)	0.82(4.18)	0.95(4.05)
节点数不同	6	0.00(4.94)	0.00(7.75)	1.00(2.00)	1.00(2.00)
	7	0.00(4.99)	0.00(7.87)	1.00(2.00)	1.00(2.00)
	8	0.00(5.29)	0.00(8.12)	0.68(1.70)	0.93(1.93)
边数不同	2	0.00(3.12)	0.00(6.41)	1.00(2.00)	1.00(2.00)
	3	0.00(4.51)	0.00(6.98)	1.00(2.00)	1.00(2.00)
	4	0.00(5.56)	0.00(7.55)	1.00(2.00)	1.00(2.00)

## 4.2 无标度网络

为了验证多分辨率模块度和密度模块度在大规模网络上社团结构探测的准确性, 我们在 Lancichinetti-Fortunato-Radicchi (LFR) 基准网络<sup>[35]</sup>上进行实验. 此网络以典型的无标度网络为基础, 其节点的度值和社团规模(节点数)都服从

幂率分布, 参数分别为  $\gamma$  和  $\beta$ . 在网络连边上需做一定的调整, 使每个节点与社团外部的连接数占其度值的比例为  $\mu$ , 与社团内部连接的比例为  $1 - \mu$ . 对少量度值低的节点, 难于完全符合这个比例, 只需接近于此比例就可以了. 按照 LFR 基准构造的网络具有明显的社团结构, 而且社团数目较多, 具有验证模块度性能所需的确定性和复杂性.

本文实验在 10000 个节点的 LFR 网络上展开, 网络的平均度  $\langle k \rangle = 20$ , 最大度  $k_{max} = 100$ , 最大社团的节点数  $c_{max} = 1000$ , 最小社团节点数  $c_{min} = 10$ , 度分布参数  $\gamma = 2$ , 另两个参数分别选择了两组数据, 社团规模参数  $\beta$  选择了 2 和 3, 连接比例参数  $\mu$  分别是 0.1 和 0.3, 这两组参数划定的范围代表了绝大部分具有明显社团结构的无标度网络. 在测试网络上分别对  $Q_\lambda$  和本文给出的三种  $Q_{N\lambda}$  进行了实验, 实验结果见图 3—图 6, 图中的每个点都是在 100 个网络上计算得到的平均数. 各图给出了被错误划分节点所占的比例与  $\lambda$  之关系, 图中的圆形和方形分别代表误合并和误分裂所对应的错误划分节点之比率(下文简称为误合并率和误分裂率). 从总体上看, 第一种密度模块度  $Q_{N1\lambda}$  的误划分率(误合并率与误分裂率之和)最高, 均在 20% 以上, 且随着  $\lambda$  的减少, 其变化比较平缓. 而  $Q_\lambda$ ,  $Q_{N2\lambda}$  和  $Q_{N3\lambda}$  的变化比较剧烈, 当  $\lambda$  为 1 时, 这三种模块度误合并了许多实际社团, 但误分裂的社团极少; 当  $\lambda$  值增大(对于  $Q_\lambda$ ) 或减小(对于  $Q_{N2\lambda}$  和  $Q_{N3\lambda}$ ) 时, 误合并社团情况迅速减少, 误分裂社团对应的误分裂率越来越高. 需要指出, 这种误分裂率的增高是合理的, 因为提高分辨率必然使得社团划分越来越细, 将最佳划分所对应的各社团进一步划分成更小的社团, 那么误分裂是在所难免的. 而最小误划分率则是需要衡量的首要指标, 因最小误分辨率直接反映了各模块度函数所对应的最佳划分社团结构与 LFR 网络原始社团结构的差异. 首先, 从图中对比可知, 三者的最小误划分率有显著差异,  $Q_\lambda$  有较高的最小误划分率, 均在 10% 以上;  $Q_{N2\lambda}$  各图的最小误划分率在 3%—5% 之间;  $Q_{N3\lambda}$  的最小误划分率大大降低, 均在 6% 以下, 而且仅就误合并率来看,  $Q_{N3\lambda}$  的误合并率显然是所有模块度中最低的. 此外, 误合

并和误分裂两类误划分同时存在的重叠部分的大小, 是衡量多分辨率模块度性能优劣的另一个方面. 如果不存在重叠部分, 无疑是最理想的情况, 即得到了相互一致的多粒度社团结构. 从各图的对比中可观察到, 图 4 的重叠部分包括了所有

的  $\lambda$  值, 图 5 比图 3 的情况有所改善, 而图 6 的重叠部分又进一步显著减小. 由以上最小误划分率和其重叠大小的分析可表明, 在大规模网络上,  $Q_{N3\lambda}$  比其他模块度能得到更好的多分辨率社团结构.

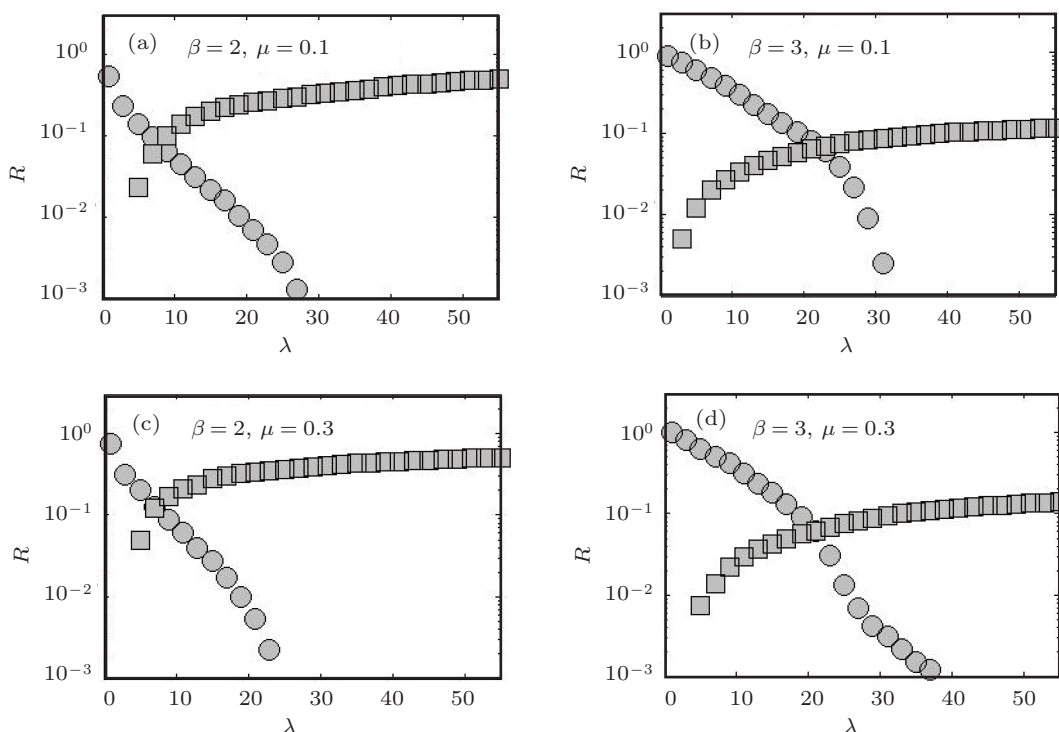


图 3 (a)—(d) 分别对应于  $\beta = 2$  或  $3, \mu = 0.1$  或  $0.3$  四种情况下  $Q_\lambda$  的误划分率 (图中的圆形和方形分别代表误合并率和误分裂率)

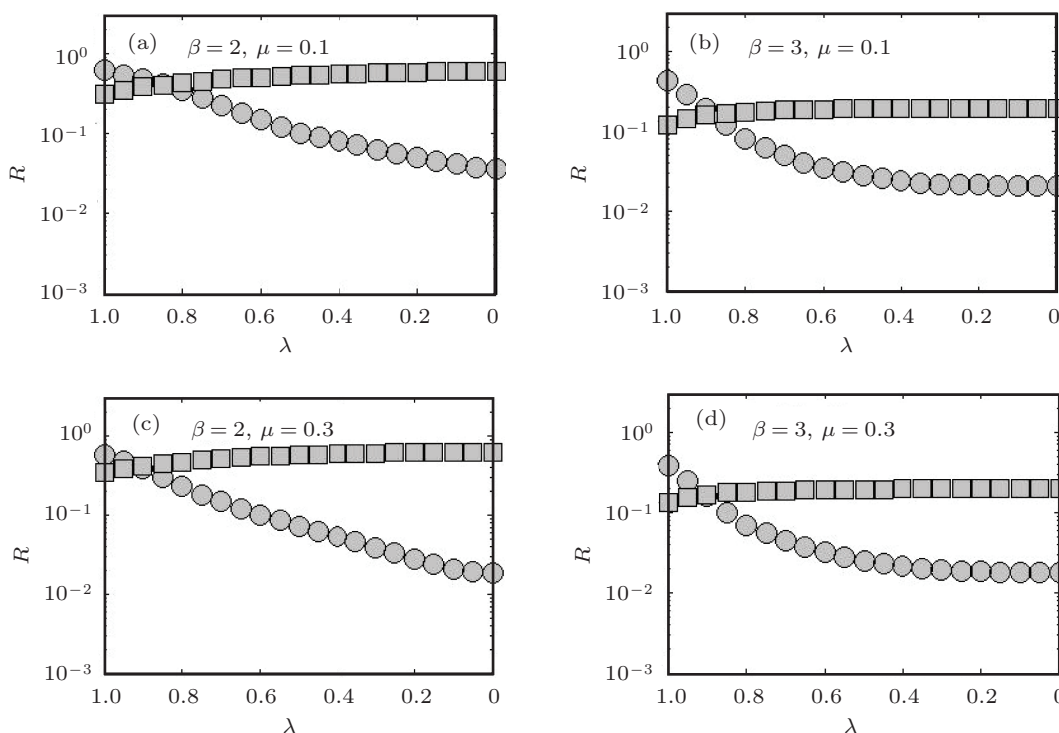


图 4 (a)—(d) 分别对应于  $\beta = 2$  或  $3, \mu = 0.1$  或  $0.3$  四种情况下  $Q_{N1\lambda}$  的误划分率 (图中的圆形和方形分别代表误合并率和误分裂率)

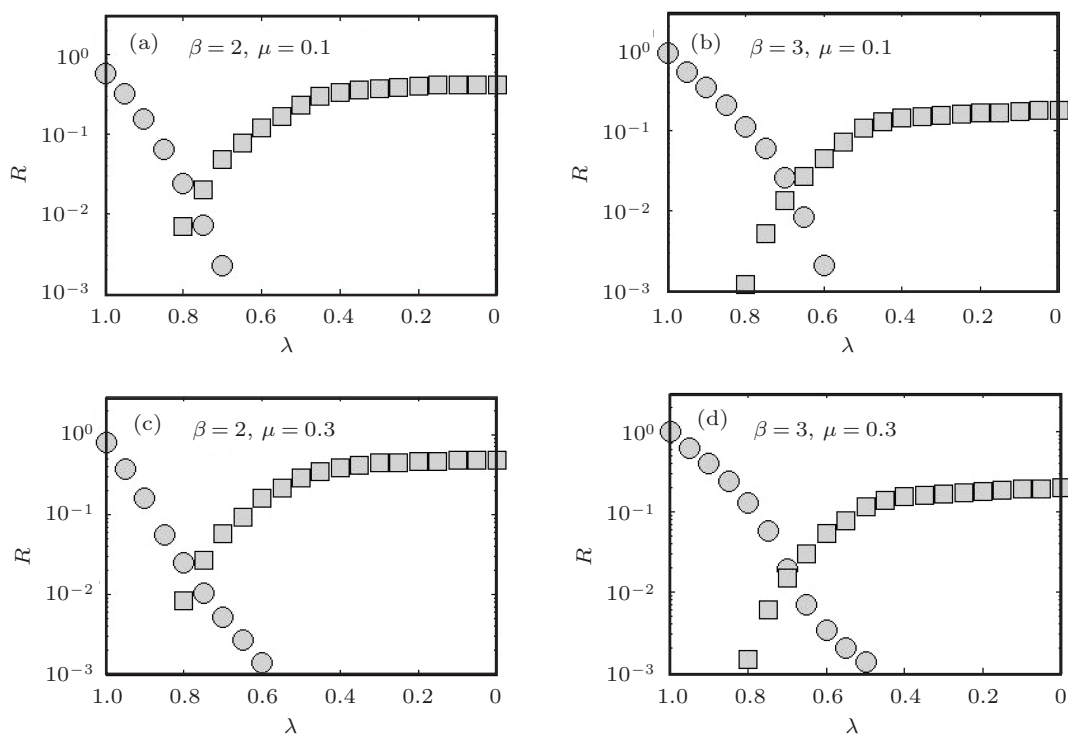


图5 (a)—(d) 分别对应于  $\beta = 2$  或  $3, \mu = 0.1$  或  $0.3$  四种情况下  $Q_{N2\lambda}$  的误划分率 (图中的圆形和方形分别代表误合并率和误分裂率)

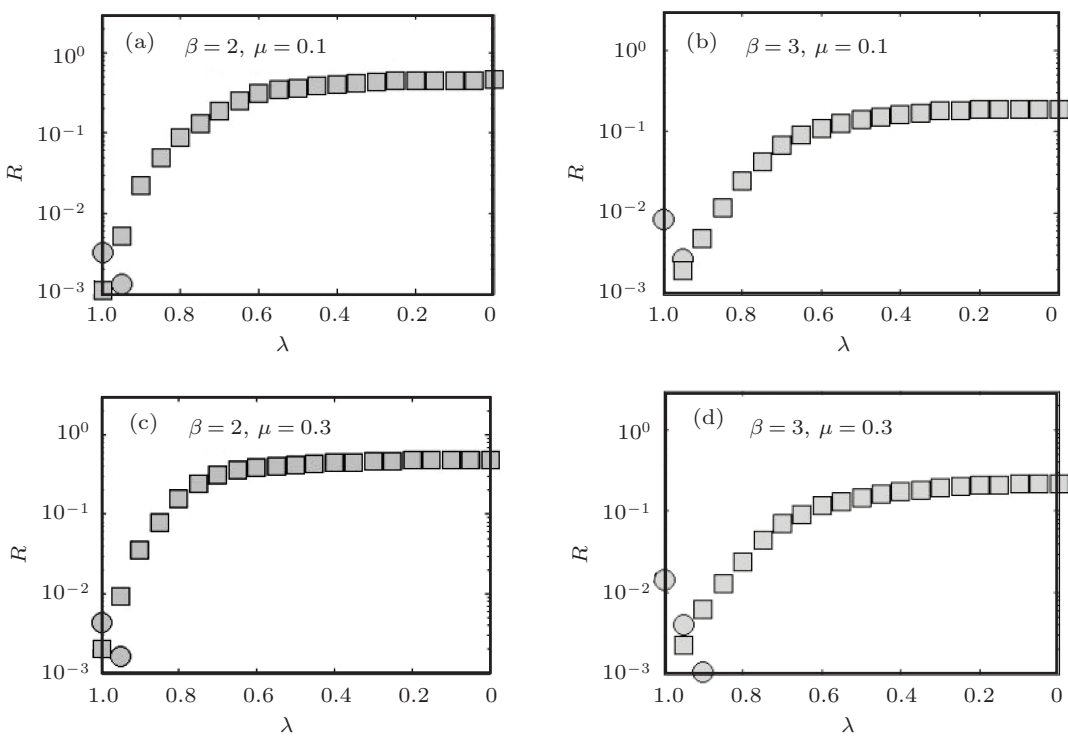


图6 (a)—(d) 分别对应于  $\beta = 2$  或  $3, \mu = 0.1$  或  $0.3$  四种情况下  $Q_{N3\lambda}$  的误划分率 (图中的圆形和方形分别代表误合并率和误分裂率)



## 5 结论

模块度是复杂网络中社团结构划分的目标函数,也是对社团结构评价的依据.最为广泛使用的 NG 模块度  $Q$  存在着分辨率限制问题,不能识别小于一定规模的社团;而  $Q$  的扩展版本多分辨率模块度  $Q_\lambda$  则存在着误合并社团和误分裂社团同时存在的现象,这使多分辨率社团结构成为空谈.本文在网络密度的基础上,以“明显的社团结构应存在着社团内部密度与社团之间密度的显著差异”为基本思想,提出了密度模块度函数和多分辨率的函

数,进一步对三种典型系数矩阵构造的密度模块度和 NG 模块度进行了对比分析,并在随机网络和无标度网络上开展了模拟实验,最终说明了以边比例系数矩阵形成的第三种密度模块度不仅能够表现出网络中各个社团的相互关联和影响,而且与其他模块度相比能够得到误合并率和最小误划分率显著降低的社团结构.此外,因多分辨率密度模块度是由网络密度构造的,而在有权网络上的网络密度只需用边权和代替无权网络的边数即可,那么本文提出的模块度显然也适用于有权网络.

- [1] Wang X F, Li X, Chen G R 2006 *Complex Networks Theory and its Application* (Beijing: Tsinghua University Press) (in Chinese) [汪小帆, 李翔, 陈关荣 2006 复杂网络理论及其应用 (北京: 清华大学出版社)]
- [2] He D R, Lin Z H, Wang B H 2008 *Complex System and Complex Networks* (Beijing: Higher Education Press) (in Chinese) [何大初, 刘宗华, 汪秉宏 2008 复杂系统与复杂网络 (北京: 高等教育出版社)]
- [3] Wang L, Dai G Z 2009 *Scale-free Feature, Phenomena and Control in Complex Networks* (Beijing: Science Press) (in Chinese) [王林, 戴贯中 2009 复杂网络的 Scale-free 性、Scale-free 现象及其控制 (北京: 科学出版社)]
- [4] Zhang P P, He Y, Zhou T, Su B B, Chang H, Zhou Y P, Wang B H, He D R 2006 *Acta Phys. Sin.* **55** 60 (in Chinese) [张培培, 何阅, 周涛, 苏蓓蓓, 常慧, 周月平, 汪秉宏, 何大初 2006 物理学报 **55** 60]
- [5] Liu H K, Zhou T 2007 *Acta Phys. Sin.* **56** 106 (in Chinese) [刘宏鲲, 周涛 2007 物理学报 **56** 106]
- [6] Xu D, Li X, Wang X F 2007 *Acta Phys. Sin.* **56** 1313 (in Chinese) [许丹, 李翔, 汪小帆 2007 物理学报 **56** 1313]
- [7] Albert R, Jeong H, Barabasi A L 1999 *Nature* **401** 130
- [8] Andrei B, Ravi K, Farzin M, Prabhakar R, Sridhar R, Raymie S, Andrew T, Janet W 2000 *Comput. Netw.* **33** 309
- [9] Williams R J, Martinez N D 2000 *Nature* **404** 180
- [10] Amaral L A N, Scala A, Barthélemy M, Stanley H E 2000 *Proc. Natl. Acad. Sci. USA* **97** 11149
- [11] Gleiser P, Danon L 2003 *Adv. Complex Syst.* **6** 565
- [12] Fortunato S 2010 *Phys. Rep.* **486** 75
- [13] Girvan M, Newman M E J 2002 *Proc. Natl. Acad. Sci. USA* **99** 7821
- [14] Newman M E J 2009 *Phys. Rev. Lett.* **103** 058701
- [15] Newman M E J 2004 *Eur. Phys. J. B* **38** 321
- [16] Sales-Pardo M, Guimera R, Moreira A A, Amaral L A N 2007 *Proc. Natl. Acad. Sci. USA* **104** 15224
- [17] Blondel V D, Guillaume J L, Lambiotte R, Lefebvre E 2008 *J. Stat. Mech.* **2008** 10008
- [18] Shen H, Cheng X, Cai K, Hu M B 2009 *Physica A* **388** 1706
- [19] Zhang S, Wang R S, Zhang X S 2007 *Physica A* **374** 483
- [20] Zhang S, Wang R S, Zhang X S 2007 *Phys. Rev. E* **76** 046103
- [21] Gao Z K, Jin N D 2008 *Acta Phys. Sin.* **57** 6909 (in Chinese) [高忠科, 金宁德 2008 物理学报 **57** 6909]
- [22] 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]
- [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] Wang G X, Shen Y 2010 *Acta Phys. Sin.* **59** 842 (in Chinese) [王高峡, 沈轶 2010 物理学报 **59** 842]
- [25] Shao F, Jiang G P 2011 *Acta Phys. Sin.* **60** 078902 (in Chinese) [邵斐, 蒋国平 2011 物理学报 **60** 078902]
- [26] Newman M E J, Girvan M 2004 *Phys. Rev. E* **69** 026113
- [27] Fortunato S, Barthélemy M 2007 *Proc. Natl. Acad. Sci. USA* **104** 36
- [28] Rosvall M, Bergstrom C T 2007 *Proc. Natl. Acad. Sci. USA* **104** 7327
- [29] Muff S, Rao F, Caflisch A 2005 *Phys. Rev. E* **72** 056107
- [30] Li Z, Zhang S, Wang R S 2008 *Phys. Rev. E* **77** 036109
- [31] Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D 2004 *Proc. Natl. Acad. Sci. USA* **101** 2658
- [32] Reichardt J, Bornholdt S 2006 *Phys. Rev. E* **74** 016110
- [33] Arenas A, Fernández A, Gómez S 2008 *New J. Phys.* **10** 053039
- [34] Lancichinetti A, Fortunato S 2011 arXiv: 1107.1155v1 [physics.soc-ph]
- [35] Lancichinetti A, Fortunato S, Radicchi F 2008 *Phys. Rev. E* **78** 046110



# Multi-resolution density modularity for finding community structure in complex networks\*

Zhang Cong<sup>1)†</sup> Shen Hui-Zhang<sup>1)</sup> Li Feng<sup>1)</sup> Yang He-Qun<sup>2)</sup>

1) (*Antai College of Economics and Management, Shanghai Jiaotong University, Shanghai 200052, China*)

2) (*Shanghai Center for Satellite Remote Sensing and Measurement Application, Shanghai 201199, China*)

(Received 7 September 2011; revised manuscript received 7 December 2011)

## Abstract

In reality many complex networks present modules or community structures obviously. Modularity is a benefit function used in quantifying the quality of a division of a network into communities. And it usually can be used as a basis for optimization methods of detecting community structure in networks. But the most popular modularity which is proposed by M. E. J. Newman and M. Girvan has the resolution limit in community detection. Multi-resolution modularity cannot overcome the misclassifications caused by merging and splitting the communities either. In this paper, we propose a multi-resolution density modularity based on the network density. The proposed function is tested on the artificial networks. Computational results show that it can reduce the rate of misclassification considerably. And the systematicness of the community structures can be demonstrated by the multi-resolution density modularity.

**Keywords:** complex networks, community structure, modularity, network density

**PACS:** 89.75.Fb, 89.75.Hc

---

\* Project supported by the National Natural Science Foundation of China (Grant Nos. 71071096, 71001068).

† E-mail: z.c.cn@163.com