

# 网络自然密度社团结构模块度函数

张 聪, 沈惠璋

(上海交通大学安泰经济与管理学院 上海 长宁区 200052)

**【摘要】**针对常用的NG模块度存在分辨率限制,不能识别小于一定规模的社团的问题,提出了网络自然密度的定义,用以衡量网络的连接紧密程度,在此基础上构造了密度模块度函数来评估社团结构的划分;进一步分3种情况证明了密度模块度函数,克服了NG模块度函数的分辨率限制问题;最后通过人工网络和经典现实网络验证了密度模块度函数的有效性。

**关键词** 社团探测; 复杂网络; 模块度; 自然密度

中图分类号 N941

文献标识码 A

doi:10.3969/j.issn.1001-0548.2012.02.003

## Modularity Function for Community Structure Based on Natural Density of Networks

ZHANG Cong and SHEN Hui-zhang

(Antai College of Economics & Management, Shanghai JiaoTong University Changning Shanghai 200052)

**Abstract** The most popular modularity optimization may fail to identify communities smaller than a scale. A natural density of networks is proposed for describing the degree of interconnectedness of modules. The density modularity function is constructed to evaluate the community structure partitioning based on the natural density. Three cases study proves that the density modularity function can overcome the resolution limit of NG's modularity. The density modularity has been tested on both artificial networks and classical real-world networks. Computational results demonstrate the effectiveness of the density modularity.

**Key words** community detection; complex networks; modularity; natural density

自然界和社会中的许多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>[1-2]</sup>。这些系统包括了互联网、社会网络、食物链网络、生物网络以及引文网络等<sup>[3-7]</sup>。复杂系统通常由群组或社团组织而成,不同的群组或社团表现出不同的功能和角色。群组内节点间的连接比较紧密,而群组间节点的连接比较松散<sup>[8]</sup>。

社团结构探测或群组发现是复杂网络领域中的一个重要课题。由物理学、应用数学、计算机科学、生物学及社会学原理、方法发展出许多社团结构探测算法<sup>[9-16]</sup>。根据这些算法可得到很多社团结构,为了辨别它们的优劣,需要一个量化的标准来评估各种社团结构。迄今为止,被最广泛使用的评价函数就是文献[17]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”的思想,那么通过比较实际覆盖度(覆盖度是社

团内部连接数占总连接数的比例)与随机连接情况下覆盖度的差异来评估社团结构<sup>[17]</sup>。模块度可用来比较不同社团结构划分的质量,如,发现最佳的层次结构分裂图、验证不同的社团结构发现算法以及作为优化过程的适应度函数,其中 $Q_{\max}$ 对应于网络的最佳社团结构。实际应用中, $Q$ 的最大值一般在0.3~0.7的范围内。然而,模块度函数存在分辨率限制问题,文献[18]研究发现模块度函数不能探测到一些较小的社团(相对整体网络而言),甚至当这些社团是派系(即完全图)时。那么,当最大模块度对应的社团结构划分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组合。还有一些相似的评论见文献[19-22],其中文献[19]提出的信息论基础上的社团结构发现方法、文献[21]的利用极值优化(extremal optimization)计算社团结构方法从不同程度上克服了模块度函数 $Q$ 的分辨率缺陷。文献[23]提出了一个模块化密度函数 $D$ ,

收稿日期: 2011-07-18; 修回日期: 2012-01-14

基金项目: 国家自然科学基金(71071096)

作者简介: 张聪(1975-),男,博士生,主要从事复杂网络与数据挖掘方面的研究。

模块化密度的大小与网络节点的平均度是相关的,使得不同网络社团结构的模块化密度值的比较没有意义,这与不同网络的模块度 $Q$ 能够相互比较不同。

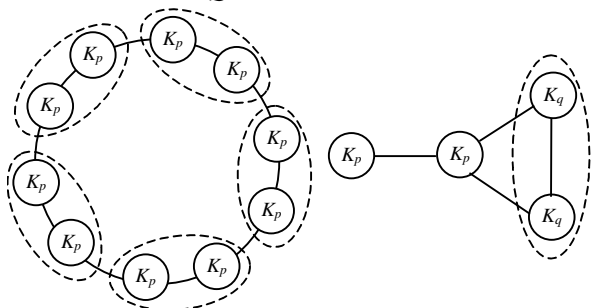
本文定义了网络自然密度来衡量社团内部的连接紧密程度,进一步构造了一个量化的函数来评估社团结构的划分,称为密度模块度函数 $U$ 。而且证明了密度模块度函数改善了NG模块度函数的分辨率限制问题。在实验部分,通过人工网络和经典现实网络验证了密度模块度函数的有效性。

## 1 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, $A$ 是 $G$ 的邻接矩阵。节点个数 $n=|V|$ ,边的个数 $L=|E|$ 。设节点集 $V'$ , $V''\subseteq V$ ,则 $V'$ 到 $V''$ 的连接数 $L(V',V'')=\sum_{i\in V',j\in V''}A_{ij}$ 。给定一个划分 $P=\{G_1(V_1,E_1),G_2(V_2,E_2),\dots,G_m(V_m,E_m)\}$ ,其中, $V_i$ 和 $E_i$ 分别是 $G_i$ 的节点集和边集( $i=1,2,\dots,m$ ;  $n_i=|V_i|$ ;  $L_i=|E_i|$ ),且 $\bigcup_{i=1}^m V_i=V$ 。模块度 $Q$ 在划分 $P$ 的基础上定义为<sup>[17]</sup>:

$$Q = \sum_{i=1}^m q_i = \sum_{i=1}^m \left[ \frac{L(V_i, V_i)}{L(V, V)} - \left( \frac{L(V_i, V)}{L(V, V)} \right)^2 \right] \quad (1)$$

式中, $q_i$ 代表社团 $i$ 的实际覆盖度与期望覆盖度的差; $Q$ 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边数小于 $\sqrt{2L}$ 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。通过文献[18]给出的两个典型网络如图1所示,可形象地表明模块度 $Q$ 的分辨率限制问题<sup>[18]</sup>。



a. 多个相同派系构成的环状网络 b. 4个不同派系构成的非对称网络

图1 说明模块度 $Q$ 分辨率限制问题的示意图

图1中的 $K_p$ 和 $K_q$ 表示派系, $p$ 和 $q$ 分别是派系中节点的个数。图1a中相邻的两个派系之间有一条边连接,组成了一个环形的网络,当派系的个数大于 $\sqrt{L}$ 时,模块度 $Q$ 评价的最佳社团结构将把两个或两个

以上相邻的派系做为社团。图1b由4个派系和4条边连接组成,其中 $p>q$ ,当 $p$ 足够大时(例如 $p=20,q=5$ ),模块度 $Q$ 将把虚线圈内两个派系合并为一个社团。

若对模块度中的每项都乘以一个因子,此因子如能反映社团的连接稠密程度,则上述分辨率限制问题可能得到解决。借鉴数论中的自然密度定义了网络的自然密度,并将其作为模块度的密度因子。自然密度定义为网络中节点的平均度与节点总数的商,公式如下:

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

自然密度与常用的网络密度 $\text{density} = \frac{2L}{n(n-1)}$ 是不同的,网络密度的取值范围是 $[0,1]$ ,而 $ND$ 的取值范围为 $[0,1)$ ,如果两个网络 $G_1, G_2$ 都是派系,节点数 $n_1 > n_2$ ,那么 $ND(G_1) > ND(G_2)$ 。显然自然密度反映了网络真实的连接稠密程度。此外,文献[23]提出的模块化密度定义为:

$$D = \sum_{i=1}^m d_i = \sum_{i=1}^m \frac{L(V_i, V_i) - L(V_i, \bar{V}_i)}{n_i} \quad (3)$$

式中, $L(V_i, \bar{V}_i)$ 表示社团 $i$ 与其外部的所有连接数,可见,模块化密度的取值与节点的平均度是相关的,而且其取值范围为 $(-\infty, +\infty)$ ,这使得不同网络的最大模块化密度值失去相互比较的意义。

本文定义的密度模块度为:

$$U = \sum_{i=1}^m u_i = \sum_{i=1}^m \left[ \begin{array}{l} ND(G_i) \frac{L(V_i, V_i)}{L(V, V)} \\ - ND(R(G_i)) \left( \frac{L(V_i, V)}{L(V, V)} \right)^2 \end{array} \right] \quad (4)$$

式中, $ND(G_i) = \frac{L(V_i, V_i)}{n_i^2}$ 是社团 $i$ 的自然密度;

$ND(R(G_i)) = \frac{(L(V_i, V))^2}{n_i^2 L(V, V)}$ 表示当随机连接时社团 $i$ 的

自然密度 $\left( \frac{L(V_i, V)}{L(V, V)} \right)^2$ 是随机连接时社团 $i$ 的期望连

接数);两种密度分别对应于社团 $i$ 的实际社团结构和期望社团结构。那么,上式可化简为:

$$U = \sum_{i=1}^m \left[ \frac{(L(V_i, V_i))^2}{n_i^2 L(V, V)} - \frac{(L(V_i, V))^4}{n_i^2 (L(V, V))^3} \right] \quad (5)$$

密度模块度 $U$ 与模块度 $Q$ 的取值范围相同,都是 $(-1,1)$ 。直观上看,当出现如图1a所示的情况时,每个派系的自然密度显然大于若干个派系构成社团的自然密度,这使得密度模块度不出现分辨率限制成

为可能。

为何选择使用网络自然密度而不是网络密度呢? 随着网络社团结构的逐渐分裂, 划分得到的社团数目不断增加, 社团的自然密度呈现先增大再减小, 其最大值出现在派系首次成为社团时; 而对于网络密度则是一个非递减的变化过程(派系的密度都是1), 这明显不利于找到最佳社团结构。此外, 当单个节点作为一个社团出现时, 其网络自然密度是0, 而网络密度是不存在的, 那么如果不做额外处理, 由网络密度构造的密度模块度函数则无法计算。

## 2 分辨率限制问题的改善

模块度  $Q$  虽然是最广泛使用的评价社团结构的函数, 但文献[18]还是指出了其的分辨率限制。本文提出的密度模块度克服了这些分辨率限制。

### 2.1 密度模块度不会把派系分割为两部分

对于一个含有  $n$  个节点的派系, 密度模块度  $U$  不会将其划分为两部分或多部分。当把派系视为一个社团时,  $U_1 = 0$ ; 当把派系划分为两个部分时, 两部分分别含有  $n_1$  和  $n_2$  个节点, 则有:

$$U_2 = \left[ \frac{n_1^2(n_1-1)^2}{n_1^2 n(n-1)} - \frac{n_1^4(n-1)^4}{n_1^2 n^3(n-1)^3} \right] + \left[ \frac{n_2^2(n_2-1)^2}{n_2^2 n(n-1)} - \frac{n_2^4(n-1)^4}{n_2^2 n^3(n-1)^3} \right] = \frac{n^2((n_1-1)^2 + (n_2-1)^2) - (n-1)^2(n_1^2 + n_2^2)}{n^3(n-1)}$$

当  $n \geq 2$ ,  $n_1 \geq 1$ ,  $n_2 \geq 1$  时, 显然  $U_2 < 0$ , 即  $U_1 > U_2$ , 那么, 派系不会被密度模块度划分成两部分。

### 2.2 密度模块度可以分辨大部分网络中的模块

利用文献[18]中给出的示意图(图1a)来测试密度模块度的有效性。这个示例是由若干个相同的派系通过单边连接所组成的环形网络。每个派系有  $p$  ( $p \geq 3$ ) 个节点, 网络共有  $m$  ( $m \geq 2$ ) 个派系, 假设  $m$  可被  $h$  ( $h \geq 2$ ,  $h$  是整数) 整除, 总节点数  $n = mp$ , 总边数  $L = mp(p-1)/2 + m$ 。

该网络有显而易见的社团结构, 其中的每个派系都是一个社团, 但是模块度  $Q$  却将相邻的两个或多个派系分辨为一个社团。下面给出用密度模块度得到的结果。当每个派系都是一个社团时, 有:

$$U_{\text{single}} = \frac{m}{p^2} \left[ \frac{p^2(p-1)^2}{mp(p-1)+2m} - \frac{(p(p-1)+2)^4}{(mp(p-1)+2m)^3} \right] = \frac{(p-1)^2}{p(p-1)+2} - \frac{p(p-1)+2}{m^2 p^2}$$

当连续的  $h$  个派系(虚线内的部分)组成一个社团时, 有:

$$U_h = \frac{m}{h} \frac{1}{h^2 p^2} \left[ \frac{(hp(p-1)+2(h-1))^2}{mp(p-1)+2m} - \frac{(hp(p-1)+2h)^4}{(mp(p-1)+2m)^3} \right] =$$

$$\frac{(hp(p-1)+2(h-1))^2}{h^3 p^2(p(p-1)+2)} - \frac{h(p(p-1)+2)}{m^2 p^2}$$

假如  $p \geq 3$ ,  $m \geq 2$ ,  $h \geq 2$ , 则有:

$$U_{\text{single}} - U_h = \frac{h^3 p^2(p-1)^2 - (hp(p-1)+2(h-1))^2}{h^3 p^2(p(p-1)+2)} +$$

$$\frac{(h-1)(p(p-1)+2)}{m^2 p^2} >$$

$$\frac{h^3 p^2(p-1)^2 - (hp(p-1)+2(h-1))^2}{h^3 p^2(p(p-1)+2)} =$$

$$\frac{h^2(h-1)p^2(p-1)^2 - 4(h(h-1)p(p-1) + (h-1)^2)}{h^3 p^2(p(p-1)+2)} >$$

$$\frac{h^2(h-1)p^2(p-1)^2 - 4(h(h-1)p^2)}{h^3 p^2(p(p-1)+2)} \geq$$

$$\frac{h^2(h-1)p^2(p-1)^2 - h(h-1)p^2(p-1)^2}{h^3 p^2(p(p-1)+2)} =$$

$$\frac{h(h-1)^2 p^2(p-1)^2}{h^3 p^2(p(p-1)+2)} > 0$$

虽然上述结论是针对这样一个特例网络得到的, 但是通过相似的过程, 显然可以证明上述结论对于任何派系构成的社团结构同样有效。而且最优化的  $U$  不会将派系划分为两部分, 因此, 密度模块度可得到每个派系都被视为社团的最佳社团结构。

### 2.3 密度模块度可分辨不同规模社团组成的网络

对于网络含有不同大小模块(形如图1b)的情况, 两种派系中分别含有  $p$  和  $q$  个节点,  $p \geq q \geq 3$ , 模块度  $Q$  优化可能将若干小的模块合并为社团。而密度模块度  $U$  排除了这种情况产生的可能性。

用  $U_{\text{separate}}$  表示两个小的派系都是社团时的密度模块度值,  $U_{\text{merge}}$  表示两个小的派系合并为一个社团时的密度模块度值。

$$U_{\text{separate}} = u_{p_1} + u_{p_2} + \frac{2}{q^2} \left[ \frac{q^2(q-1)^2}{2p(p-1)+2q(q-1)+8} - \frac{(q(q-1)+2)^4}{(2p(p-1)+2q(q-1)+8)^3} \right]$$

$$\begin{aligned}
U_{\text{merge}} &= u_{p_1} + u_{p_2} + \frac{1}{4q^2} \left[ \frac{(2q(q-1)+2)^2}{2p(p-1)+2q(q-1)+8} - \frac{(2q(q-1)+4)^4}{(2p(p-1)+2q(q-1)+8)^3} \right] \\
U_{\text{separate}} - U_{\text{merge}} &= \frac{2}{q^2} \left[ \frac{q^2(q-1)^2}{2p(p-1)+2q(q-1)+8} - \frac{(q(q-1)+2)^4}{(2p(p-1)+2q(q-1)+8)^3} \right] - \\
&\frac{1}{4q^2} \left[ \frac{(2q(q-1)+2)^2}{2p(p-1)+2q(q-1)+8} - \frac{(2q(q-1)+4)^4}{(2p(p-1)+2q(q-1)+8)^3} \right] = \\
&\frac{2q^2(q-1)^2 - (q(q-1)+1)^2}{q^2(2p(p-1)+2q(q-1)+8)} + \\
&\frac{4(q(q-1)+2)^2 - 2(q(q-1)+2)^2}{q^2(2p(p-1)+2q(q-1)+8)^3} > \\
&\frac{q^2(q-1)^2 - 2q(q-1) - 1}{q^2(2p(p-1)+2q(q-1)+8)} = \\
&\frac{(q(q-1)-1)^2 - 2}{q^2(2p(p-1)+2q(q-1)+8)} > 0
\end{aligned}$$

通过上述证明可知, 密度模块度可将大小不等的模块都正确的划分为社团。显然, 这是因为加入自然密度, 使相对小的模块的  $u_i$  之和大于它们合并形成社团的  $u$  值。

从上面的3种情况可知, 正确的网络划分对应于最大的密度模块度值。由此说明, 利用密度模块度评价社团结构是有效的。

### 3 数值实验

分别在人工模拟网络和经典的现实网络上进行了实验。由于此优化是NP完全问题, 本文采用了模拟退火算法对问题进行求解。

#### 3.1 人工模拟网络

为了比较密度模块度  $U$  与模块度  $Q$  的性能, 本文进行了文献[8,19]中描述的基准实验。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, 探测正确的社团结构会变得越来越难。

本文实验的计算结果如表1所示, 分别给出了不同  $k_{out}$  所对应的节点划分正确比率。实验包含了3个部分: 1) 网络包含4个组, 每组的节点数相同, 随着  $k_{out}$  的增加, 划分的正确率降低, 而两种方法的正确率差异很小。2) 网络包含两个组, 一组有96个节

点, 另一组有32个节点, 这种非对称网络使模块度  $Q$  难于得到正确的划分, 但是密度模块度  $U$  还是有高的划分正确率。3) 网络包含两个组, 每组有64个节点, 但是两组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  $U$  明显能够以接近100%的比率发现正确的社团结构, 而模块度  $Q$  则越来越难得到正确的社团结构。表2给出了两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, 密度模块度  $U$  比模块度  $Q$  有更好的性能。另外, 密度模块度与信息论方法得到的结果相比, 节点的划分正确率基本相当, 社团数目正确率则略优于后者。

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	$k_{out}$	$U$	$Q$	$C$
相同	6	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)
	8	0.88(0.08)	0.89(0.05)	0.87 (0.08)
节点数不同	6	0.99(0.01)	0.85(0.04)	0.99 (0.01)
	7	0.98(0.02)	0.80(0.03)	0.96 (0.04)
	8	0.85(0.08)	0.74(0.05)	0.82 (0.10)
边数不同	2	1.00(0.00)	1.00(0.01)	1.00 (0.00)
	3	1.00(0.00)	0.96(0.03)	1.00 (0.00)
	4	1.00(0.00)	0.74(0.10)	1.00 (0.01)

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	$k_{out}$	$U$	$Q$	$C$
相同	6	1.00(4.00)	1.00(4.00)	1.00 (4.00)
	7	1.00(4.00)	1.00(4.00)	1.00 (4.00)
	8	0.85 (4.04)	0.70(4.33)	0.14 (1.93)
节点数不同	6	1.00(2.00)	0.00(4.95)	1.00 (2.00)
	7	0.92(1.93)	0.00(4.97)	0.80 (1.80)
	8	0.71(1.27)	0.00(5.29)	0.06 (1.06)
边数不同	2	1.00(2.00)	0.00(3.10)	1.00 (2.00)
	3	1.00(2.00)	0.00(4.48)	1.00 (2.00)
	4	1.00(2.00)	0.00(5.55)	1.00 (2.00)

#### 3.2 经典网络

空手道俱乐部网络是社会网络分析的一个经典案例。该网络由34个成员作为节点, 而78条边代表了成员之间的朋友关系。该文作者用了两年的时间来观察美国一所大学中的空手道俱乐部成员间的相互社会关系, 从而构造了该网络。由于该俱乐部的主管与校长之间因某些问题而产生了争执, 结果该俱乐部分裂成了两个分别以主管和校长为核心的小俱乐部(图2中粗线划分而成的左右两部分)。通过本文密度模块度函数可将空手道俱乐部网络划分为与实际划分相同的两部分, 而在  $U$  最大化时, 网络被

划分为如图2所示的4个社团, 从网络的拓扑结构看, 这样的划分是合理的, 且符合文献[24]提出的社团结构的弱定义, 此时的密度模块度值  $U_{\max} = 0.2131$ 。

模块度  $Q$  的最大值也出现在图2的4个社团时,  $Q_{\max} = 0.4198$ , 由于社团结构比较简单, 二者的结果是一致的。

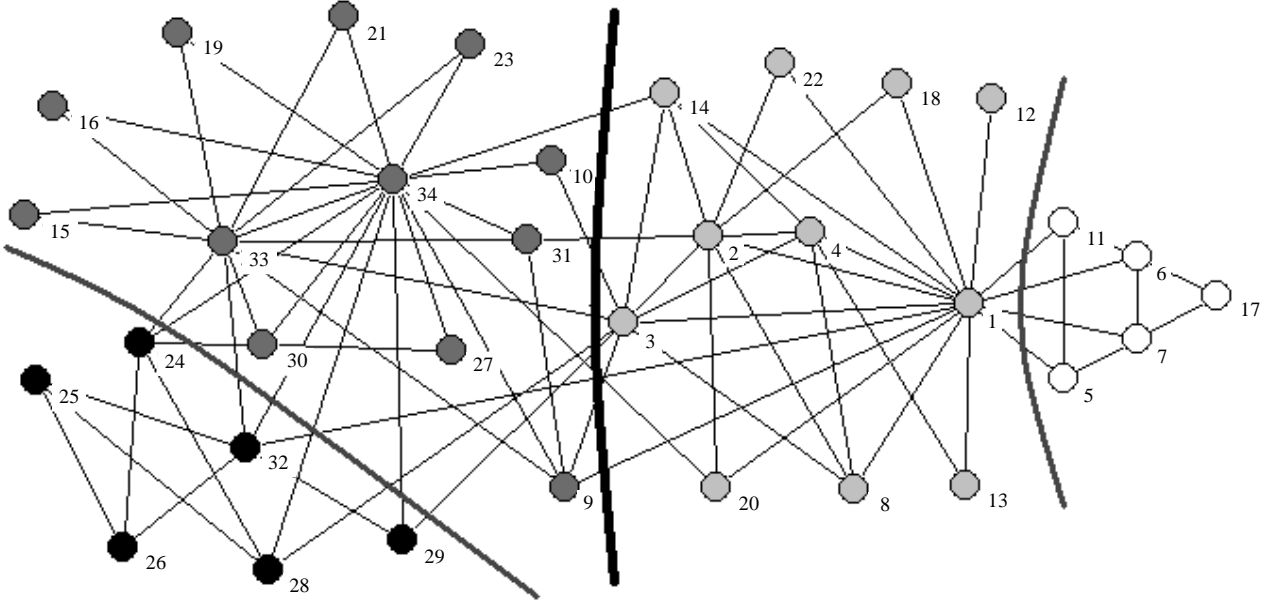


图2 空手道俱乐部网络的划分结果

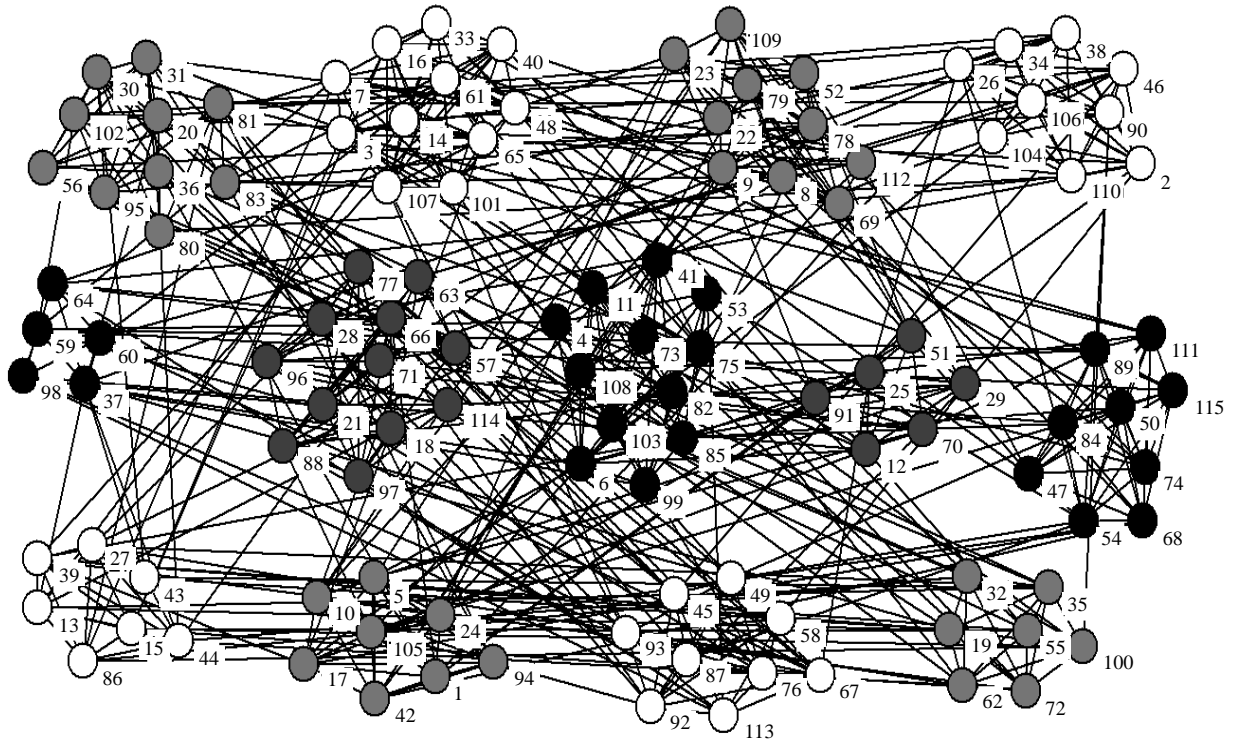


图3 大学橄榄球联盟网络按照密度模块度划分的社团结构

本文实验研究的第二个真实网络是美国大学橄榄球联盟, 网络中115个节点表示115支球队, 而613条边代表了613场赛事。这些球队被分成若干小组, 每个小组由8~12支球队组成, 同组内球队之间的比赛要比组间球队之间的比赛多。根据模块度  $Q$  得到

的最佳社团结构含有11个社团。按照密度模块度函数的评估将橄榄球联盟网络划分为13个社团, 比前者多两个社团, 如图3所示。两种模块度在不同社团个数时的取值如图4所示, 密度模块度的曲线形状是山峰状的, 当社团个数比较少时, 其  $U$  值是较低的,

这种现象是合理的,使得密度模块度能更好地克服分辨率限制问题。从图中的第3条曲线可知,在达到最佳社团结构之前,两种模块度取值的差异是明显的,随着社团个数的增加,二者的差异越来越小,慢慢趋于稳定。这显然是自然密度所引起的,未达到最佳社团结构时,各社团的自然密度较小,随着社团个数的增加,自然密度逐渐增加,使两种模块度的差异减小;而当各个社团逐渐成为派系时,自然密度开始缓慢下降,两种模块度的差异缓慢增加。

当然,由于经典网络的社团划分无标准答案,而且在以上二网络上两种模块度划分结果的差异不大,以上经典网络的分析可以看作是对人工模拟网络实验的补充和佐证。

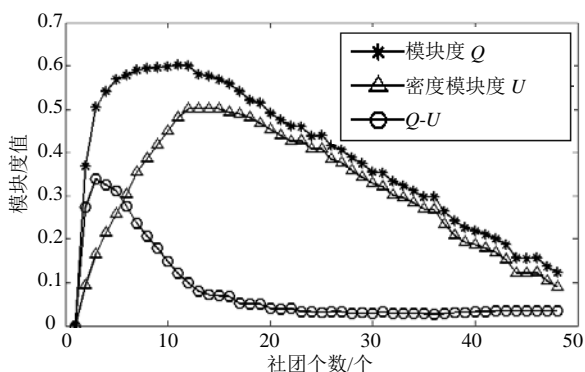


图4 两种模块度及其差值的社团个数-模块度值曲线图

## 4 结论与讨论

通过以上证明和实验,表明本文提出的密度模块度函数能够分辨出NG模块度所无法分辨的规模小的社团,包括了对称的和不对称的网络。这显然是由于自然密度因子的加入所带来的。自然密度不仅反映了社团内部的连接紧密程度,同时使密度模块度考虑了社团中的节点个数因素,从而克服了NG模块度中无节点个数因素所带来的片面性。

按照密度模块度进行优化,可以得到最佳的社团结构,而且对于有权网络同样适用。密度模块度函数计算的时间复杂度与NG模块度函数是一致的,都是 $O(n^2)$ ,当整个网络为稀疏连接时,可使用以边为存储对象的数据结构,计算所需的时间复杂度为 $O(L)$ 。

由于本文主要研究密度模块度克服分辨率限制问题,在此仅讨论密度模块度用于重叠社团结构划分的问题。笔者认为只需对密度模块度的函数形式做适当变形即可使用,譬如按照文献[15]中给出的基于模糊聚类的社团划分方法将密度模块度变形为:

$$\tilde{U} = \sum_{c=1}^m \left[ \frac{(B(\tilde{V}_c, \tilde{V}_c))^2}{\tilde{n}_c^2 B(V, V)} - \frac{(B(\tilde{V}_c, V))^4}{\tilde{n}_c^2 (B(V, V))^3} \right] \quad (6)$$

式中,  $\tilde{n}_c = \sum_i \mu_{ic}$  为社团  $c$  中节点的模糊隶属度之和;  $B(\tilde{V}_c, \tilde{V}_c) = \sum_{i,j \in V_c} ((\mu_{ic} + \mu_{jc})/2) A_{ij}$ ;  $B(\tilde{V}_c, V) = B(\tilde{V}_c, \tilde{V}_c) + \sum_{i \in V_c, j \in V \setminus \tilde{V}_c} ((\mu_{ic} + (1 - \mu_{jc}))/2) A_{ij}$ ;  $B(V, V) = \sum_{i,j} A_{ij}$ ; 隶属度矩阵  $M_{n \times m} = (\mu_{ic})$ 。

## 参考文献

- [1] FORTUNATO S. Community detection in graphs[J]. Physics Reports, 2010, 486: 75-174.
- [2] PORTER M A, ONNELA J P, MUCHA P J. Communities in networks[J]. Notices of the AMS, 2009, 56(9): 1082-1097.
- [3] ALBERT R, JEONG H, BARABSI A L. Diameter of the world-wide web[J]. Nature, 1999, 401(6749): 130-131.
- [4] ANDREI B, RAVI K, FARZIN M, et al. Graph structure in the Web[J]. Computer Networks, 2000, 33: 309-320.
- [5] WILLIAMS R J, MARTINEZ N D. Simple rules yield complex food webs[J]. Nature, 2000, 404(6774): 180-183.
- [6] AMARAL L A N, SCALA A, BARTHELEMY M, et al. Classes of small-world networks[J]. Proc Natl Acad Sci USA, 2000, 97(21): 11149-11152.
- [7] GLEISER P, DANON L. Community structure in jazz[J]. Advances in Complex System, 2003, 6(4): 565-573.
- [8] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks[J]. Proc Natl Acad Sci USA, 2002, 99: 7821-7826.
- [9] 汪小帆, 刘亚冰. 复杂网络中的社团结构算法综述[J]. 电子科技大学学报, 2009, 38(5): 537-543.  
WANG Xiao-fan, LIU Ya-bing. Overview of algorithms for detecting community structure in complex networks[J]. Journal of University of Electronic Science and Technology of China, 2009, 38(5): 537-543.
- [10] NEWMAN M E J. Random graphs with clustering[J]. Phys Rev Lett, 2009, 103: 058701.
- [11] NEWMAN M E J. Detecting community structure in networks[J]. Eur Phys J B, 2004, 38(2): 321.
- [12] SALES-PARDO M R, GUIMERA A, MOREIRA A, et al. Extracting the hierarchical organization of complex systems[J]. Proc Natl Acad Sci USA, 2007, 104: 15224.
- [13] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of community hierarchies in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 10: 10008.
- [14] SHEN H, CHENG X, CAI K, et al. Detect overlapping and hierarchical community structure in networks[J]. Physica A, 2009, 388: 1706-1712.
- [15] ZHANG S, WANG R S, ZHANG X S. Identification of overlapping community structure in complex networks using fuzzy c-means clustering[J]. Physica A, 2007, 374: 483-490.
- [16] ZHANG S, WANG R S, ZHANG X S. Uncovering fuzzy

- community structure in complex networks[J]. Phys Rev E, 2007, 76(4): 046103.
- [17] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. Phys Rev E, 2004, 69(2): 026113.
- [18] FORTUNATO S, BARTHÉLEMY M. Resolution limit in community detection[J]. Proc Natl Acad Sci USA, 2007, 104(1): 36-41.
- [19] ROSVALL M, BERGSTROM C T. An information-theoretic framework for resolving community structure in complex networks[J]. Proc Natl Acad Sci USA, 2007, 104(18): 7327-7331.
- [20] MUFF S, RAO F, CAFLISCH A. Local modularity measure for network clusterizations[J]. Phys Rev E, 2005, 72(5): 056107.
- [21] DUCH J, ARENAS A. Community detection in complex networks using extremal optimization[J]. Phys Rev E, 2005, 72(2): 027104.
- [22] LANCICHINETTI A, FORTUNATO S. Limits of modularity maximization in community detection[J/OL]. Eprint arXiv, 2011, 1107: 1155[2011-07-06]. <http://www.arXiv.org>.
- [23] LI Z, ZHANG S, WANG R S. Quantitative function for community detection[J]. Phys Rev E, 2008, 77(3): 036109.
- [24] RADICCHI F, CASTELLANO C, CECCONI F, et al. Defining and identifying communities in networks[J]. Proc Natl Acad Sci USA, 2004, 101(9): 2658-2663.

编辑 蒋晓