

小世界网络与无标度网络的社区结构研究^{*}

杜海峰^{1,2,3)†} 李树茁¹⁾ W. F. Marcus²⁾ 悦中山⁴⁾ 杨绪松⁴⁾

1) 西安交通大学人口与发展研究所, 西安 710049)

2) 斯坦福大学人口与资源研究所, 美国斯坦福 94305)

3) 西安交通大学机械工程学院, 西安 710049)

4) 西安交通大学管理学院, 西安 710049)

(2007 年 3 月 4 日收到, 2007 年 5 月 25 日收到修改稿)

模块性(modularity)是度量网络社区结构(community structure)的主要参数. 探讨了 Watts 和 Strogatz 的小世界网络(简称 W-S 模型)以及 Barabási 等的 B-A 无标度网络(简称 B-A 模型)两类型复杂网络模块性特点. 结果显示, 网络模块性受到网络连接稀疏的影响, W-S 模型具有显著的社区结构, 而 B-A 模型的社区结构特征不明显. 因此, 应用中应该分别讨论网络的小世界现象和无标度特性. 社区结构不同于小世界现象和无标度特性, 并可以利用模块性区别网络类型. 因此网络复杂性指标应该包括社区结构. 现实社会网络可能同时具有小世界现象、无标度特性和清晰的社区结构, 而目前的复杂网络模型只揭示了其部分特征, 为了更全面地认识社会复杂性, 需要建立新的复杂网络模型.

关键词: 模块性, 社区结构, 小世界网络, 无标度网络

PACC: 0590, 0175

1. 引言

自然界中存在的大量系统都可以通过网络加以描述^[1,2], 复杂网络已经成为研究这些系统的新工具. 同时, 也是复杂性科学研究的新热点^[3-7]. 小世界现象和无标度特性是目前两类最典型的复杂网络特征^[8,9]. 小世界现象揭示了客观世界许多复杂网络运动中最为有效的信息传递方式之一, 即一个高度聚集的包含了“局部连接”节点的子网, 连同一些有助于产生短路径的长距离随机连接^[10]. 小世界现象目前还没有精确的定义. 一般认为, 如果网络的聚类系数远远大于相应随机网络, 而平均路径长度相当, 则称该网络具有小世界现象. 聚类系数和平均路径长度是考察网络小世界现象的两个重要指标^[11]. 节点度满足幂律分布的网络具有无标度特性, 相应网络被称为无标度网络^[5,12]. 与泊松分布和高斯分布尾部呈指数下降不同, 幂律分布的尾部下降很慢, 因

此可能会出现网络中心节点. 有研究者认为, 无标度网络是小世界网络的一种^[13].

进一步的研究发现, 多数现实网络是不均匀的, 即由许多子网络构成. 这些子网络内个体之间的关系比较紧密, 而子网络之间个体的关系比较稀疏^[14-18]. 这一现象在社会网络中尤为常见^[1,2,19]. Newman^[20]将异构网络中由不同性质、类型的节点组成的关系丰富的结构称为“社区”(子网络). 社区内节点关系稠密, 而不同社区节点之间关系稀疏的结构——社区结构(community structure)是复杂网络的特征之一^[20,21].

一方面, 以无标度特性、小世界现象以及社区结构为内容的复杂网络研究才展开不久. 对无标度特性和小世界现象的研究多集中在有关网络模型的讨论以及现实网络复杂性特征的揭示^[22-29], 对社区结构的研究也多集中在探测方法的研究^[20,30-39]. 虽然社区结构是网络的特征之一^[1,2,20,21], 但是对小世界网络和无标度网络等复杂网络的社区结构特点的研

^{*} 国家自然科学基金(批准号: 70671083, 50505034), 国家社会科学基金(批准号: 05ARK001, 05CRK002), 教育部新世纪优秀人才支持计划(批准号: NCET-04-0931), 国家教育振兴行动计划(批准号: 07200701), 美国 Santa Fe 研究所国际合作计划和斯坦福大学健康与老龄化人口经济统计中心种子基金(批准号: AG017253-06)资助的课题.

[†] E-mail: haifengdu@mail.xjtu.edu.cn

究还很少见.另一方面,对小世界现象和无标度特性的关系研究还不够深入,尤其缺乏基于社区结构的对比研究.

本文通过仿真和有关数学分析,研究 Watts 和 Strogatz 的小世界网络(简称 W-S 模型)以及 Barabási 等提出的 B-A 无标度网络(简称 B-A 模型)的社区结构及主要影响因素,进而从社区结构的角度比较这两类典型复杂网络的异同,探讨社区结构与小世界现象、无标度特性等复杂网络特征之间的关系.这不仅有利于揭示这两类典型复杂网络的形成特点、组织结构特征,也可以为认识复杂网络提供新视角,还可以为实践中判断网络复杂性特征提供新指标.另外,本研究获得的复杂网络社区结构特征,可为建立新的复杂网络模型提供借鉴.

2. 模型与方法

2.1. 模型

网络可以表示为

$$G = (V, E),$$

其中 V 表示给定节点,即网络成员的集合, E 表示边,即网络成员关系的集合. $n = |V|$ 是网络成员数目, $A = (a_{ij})$ 是相应的邻接矩阵,对于 $0, 1$ 矩阵 A ,

如果节点 i 和节点 j 相连,且 $i \neq j$,有 $a_{ij} = 1$,否则 $a_{ij} = 0$.

本文考察的复杂网络包括 W-S 模型和 B-A 模型.传统社会网络分析中的网络参数很多,包括度、中心性以及中心势等^[1, 2].复杂网络强调的特征通常有 3 个:平均聚类系数、平均路径长度以及度分布^[9, 24].作为判断网络小世界现象的主要参数,平均聚类系数和平均路径长度可以利用公式计算.度分布是否符合幂律是无标度特性判断的主要指标,然而实际网络度分布的判定因无法利用理论计算实现而相对困难,多依据对数图形或采用拟合的方法判断^[40, 41].模块性是度量网络社区结构的指标,也可以通过计算求得.

2.1.1. W-S 模型

W-S 模型是从规则网络向随机网络过渡的中间网络形态^[11].图 1 以一维环状点阵为例说明这一网络的构造方法.从一个有 n 个节点环状网络中,如果每个结点向与它最近邻的 k 个结点连出 k 条边就形成如图 1(a)所示的规则网络,对图 1(a)中每条边,以概率 p 改变它的目的节点来重新连接此边,并保证没有重复的边出现.当 $p = 0.1$ 时,可以获得如图 1(b)所示的 W-S 模型.而当 $p = 1$ 时,上述网络就变成了图 1(c)所示的随机网络.图 1(a)中 $n = 20, k = 4$.

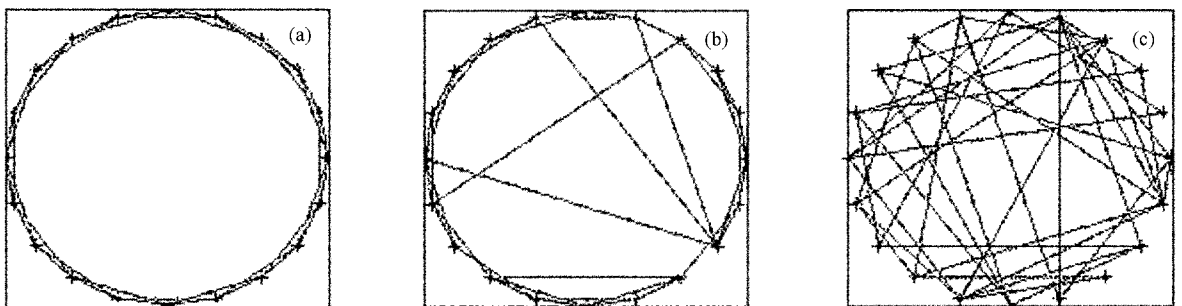


图 1 W-S 模型的构造过程 (a) 规则网络, $p = 0$ (b) Watts 小世界网络, $p = 0.1$ (c) 随机网络, $p = 1$

2.1.2. B-A 模型

节点 i 的度一般指的是与节点直接相连边的数目,即

$$k_i = \sum_{j=1}^n a_{ij}. \quad (1)$$

节点的度反映网络成员个体的交际能力,进而反映网络成员占有网络资源的情况.一般而言,节点的度越大,说明其交际能力越强,占有的网络资源也越多.无标度特性关注网络节点度的分布情况.

Barabási 等提出的 B-A 模型构造了具有幂律度分布的网络^[12],具体算法如下:

第一步,节点增长.开始于较少的节点数量 (m_0),在每个时间间隔增添一个具有 m ($m \leq m_0$) 条边的新节点,连接这个新节点到 m 个不同的已存在于系统的节点上.

第二步,偏好依附连接.在选择新节点的连接点时,假设新节点连接到节点 i 的概率取决于节点 i 的度 k_i ,即

$$P(k_i) = \frac{k_i}{\sum_j k_j}. \quad (2)$$

理论分析表明,由此构造的网络度分布满足幂律分布,即

$$P(k) = \frac{2m^2 t}{m_0 + t} \frac{1}{k^3}. \quad (3)$$

Barabási 等^[12]指出,节点增长和偏好依附连接是 B-A 模型的关键要素,而将节点度分布满足幂律分布的网络称为无标度网络^[40].

2.1.3. 平均聚类系数 C

反映网络集团化的程度,即考察连接在一起的集团各自的近邻之中有多少是共同的近邻.聚类系数有不同的定义,彼此之间有差异^[42],比较常用的一种定义如下:对于一个节点 i ,其直接相连的节点集合为 V_i ,记 V_i 的节点数量为 $n_i = |V_i|$,则顶点 i 的聚集系数可以定义为^[11]

$$c_i = \frac{\sum_{i,j \in V_i} a_{ij}}{\binom{n_i}{2}}. \quad (4)$$

对于所有的节点,平均聚类系数为

$$C = \frac{1}{n} \sum_{i=1}^n c_i. \quad (5)$$

2.1.4. 平均路径长度 L

平均路径长度用来度量整体网络节点之间通信的有效性,即一个个体平均经过多少步才能达到另一个个体.定义节点 i 与节点 j 间的最短距离为 $l(i,j)$,平均路径长度为^[11]

$$L = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n l(i,j). \quad (6)$$

2.1.5. 模块性指标 Q

设网络节点集合 V_p, V_q 是 V 的真子集,即 $V_p \neq \emptyset, V_q \neq \emptyset$,且 $V_p \subset V, V_q \subset V$.若 $V_p \cap V_q = \emptyset$,则有 $A_{pq} = \{a_{ij}\}, i \in V_p, j \in V_q$,且 $A_{pq} \subset A$.记

$$\|A\| = \sum_{i=1}^n \sum_{j=1}^n a_{ij},$$

因此当 $p \neq q$ 时,

$$\|A_{pq}\| = \sum_{i \in V_p} \sum_{j \in V_q} a_{ij} \quad (a_{ij} \in A)$$

为子集 V_p, V_q 之间的关系数量.而当 $p = q$ 时,

$$\|A_{pp}\| = \sum_{i \in V_p} \sum_{j \in V_p} a_{ij} \quad (a_{ij} \in A)$$

为子集内部的关系数量.对于一个有 t 个社区的网络,若定义

$$e_{pq} = \frac{\|A_{pq}\|}{\|A\|}, \quad (7)$$

则模块性指标可定义为^[15]

$$Q = \sum_{p=1}^t [e_{pp} - (\sum_{q=1}^t e_{pq})^2]. \quad (8)$$

Newman 进一步指出,如果 Q 大于 0.3 就表明网络存在明显的社区结构. Q 越大,网络的社区结构就越明显^[20,37].

2.2. 分析方法

对于 W-S 模型,本文在不同的重新连接概率 p 、邻近节点数 k 以及网络规模 n 时,依据 W-S 模型产生相应网络,然后计算平均聚类系数、平均路径长度以及模块性,考察其统计平均值随网络主要参数 p, k 和 n 的变化情况,探讨影响 W-S 模型社区结构的变化规律和主要因素. B-A 模型社区结构的分析思路与 W-S 模型基本一致,所不同的是网络参数主要集中在边增加数 m 和网络规模 n .由于初始节点数 m_0 要远远小于 n ^[12],其对网络结构的影响较小,因此没有考虑.本文采用的标准 W-S 模型和 B-A 模型,其度分布的研究成果很多^[12,24],所以不再深入讨论.多数实际网络比较稀疏,孤立节点较多,因此,通常平均聚类系数和平均路径长度的计算无法严格按照公式进行.对于平均聚类系数,本文只计算 $k_i > 1$ 的节点,对于平均路径长度,只考虑 $l(i,j)$ 值存在的节点,即 $l(i,j) \leq n-1$,相关结果与 Ucinet 软件的一致.本文也基于社区结构讨论 B-A 模型和 W-S 模型的关系、社区结构和小世界现象以及无标度特性等复杂网络特征之间的关系.

本文的网络产生和相应参数计算利用 Matlab 软件完成,每一个参数都是多次随机独立试验的统计平均,其中,模块性值依据 Aaron 算法^[38]迭代获得(为了克服相关社区结构算法计算量大的问题,我们已经提出一种改进的 Aaron 算法^[43]).需要说明的是,一方面,由于随着网络规模的增加,邻接矩阵的规模扩大且 Aaron 算法的计算量增加迅速,另一方面,虽然网络规模对模块性、平均路径长度、聚类系数等有一定影响,但是已有的结果已表明网络规模对本文的基本结论影响不大,所以本文的仿真网络规模最大为 600.

3. 试验结果

3.1. W-S 模型的模块性

对于 W-S 模型的不同连接概率 p ,平均聚类系

数 C 、平均路径长度 L 及模块性 Q 分别记为 $C(p)$ 、 $L(p)$ 和 $Q(p)$ 因此 $C(0)$ 、 $L(0)$ 和 $Q(0)$ 是规则网络的参数,而 $C(1)$ 、 $L(1)$ 和 $Q(1)$ 则是相应随机网络的参数.当 $n = 160$, $k = 8$ 时,图 2 给出 W-S 模型的相关参数 $L(p)$ 、 $C(p)$ 和 $Q(p)$ 随连接概率 p 变化的情况,图中每一 p 值对应的参数值都是 100 次独立试验的统计结果.同时,为了方便作图,以规则网络的相应参数为基准,对 $C(p)$ 、 $L(p)$ 和 $Q(p)$ 做了归一化处理.

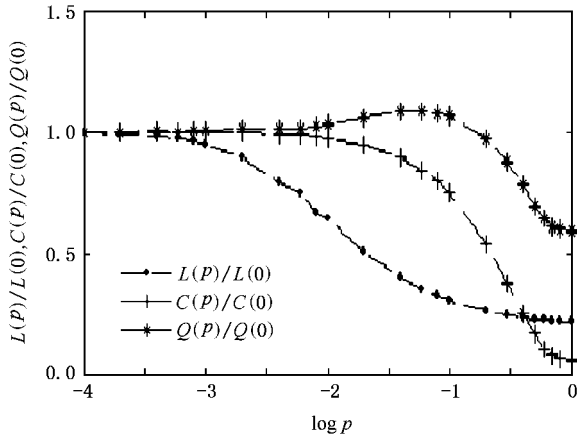


图 2 L 、 C 和 Q 随 p 的变化情况

从图 2 可以看出,随着重新连接概率 p 的增加,平均路径长度和聚类系数逐渐减少,网络表现出从规则到随机变化的小世界现象^[11].模块性却并不是随着连接概率 p 单调增减,而是出现了一个明显的峰值.当 $0.01 \leq p \leq 0.1$ 时,W-S 模型的模块性大于规则网络.

当 $k = 6$ 时,不同规模 n 的 W-S 模型模块性随连接概率 p 的变化规律如图 3 所示,图中每一 p 值对应的参数值都是 100 次独立试验的统计结果.需要说明的是,当 k 取不同值时,虽然模块性 Q 值略有变化,但是其变化规律仍如图 3 所示,因此本文没有再列出对应的 W-S 模型网络模块性随连接概率 p 的变化图.

试验表明,对于不同的 k 和 n ,W-S 模型的模块性值随 p 变化均出现了峰值,而且规律一致,即当 $0.01 \leq p \leq 0.1$ 时,W-S 模型的模块性大于规则网络.这说明当连接概率 p 取值合适时,W-S 模型可以具有比规则网络更好的社区结构.网络的稀疏程度是影响 W-S 模型网络社区结构的主要因素之一,网络连接越稀疏,模块性值越大,社区结构也就越清晰.

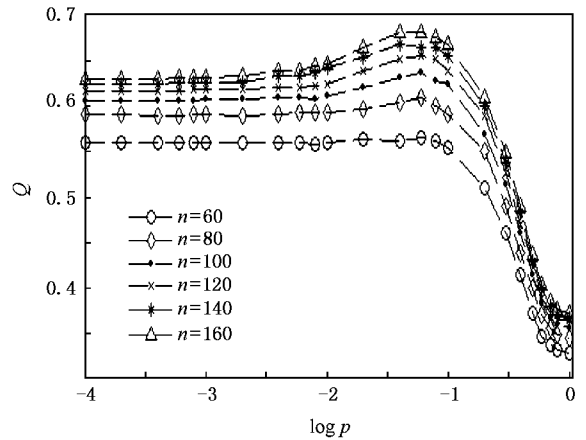


图 3 模块性随重新连接概率 p 的变化情况

3.2. B-A 模型的模块性

图 4 所示为在网络规模 $n = 600$, m 取不同值时,B-A 模型平均聚类系数 C 、平均路径长度 L 及模块性 Q 变化的情况,并分别记为 $C(m)$ 、 $L(m)$ 和 $Q(m)$.为了作图方便,以 $m = 4$ 时的参数为标准,将 $C(m)$ 、 $L(m)$ 和 $Q(m)$ 做了形式上的标准化.为了进一步比较 B-A 模型的特点,图 4 也给出了对应随机网络的参数,分别用 $C_r(m)$ 、 $L_r(m)$ 和 $Q_r(m)$ 表示.这些随机网络和相应的 B-A 模型具有相同的规模和连接关系数.由于网络规模不同时相应的 B-A 模型依然具有如图 4 所示的规律,故本文没有一一列出.图 5 所示为 B-A 模型的模块性随 m 和网络规模 n 变化的情况.图 4 和图 5 中所有参数值均为 20 次独立随机试验的统计结果.

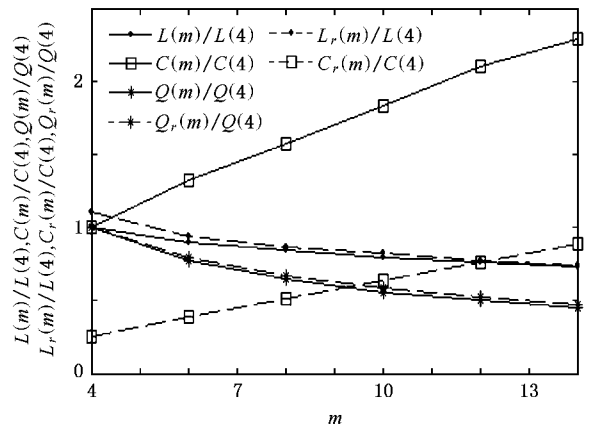


图 4 三个统计量随 m 的变化

由图 4 可以看出,B-A 模型的平均路径长度与相应的随机网络相当,而聚类系数要明显大于相应

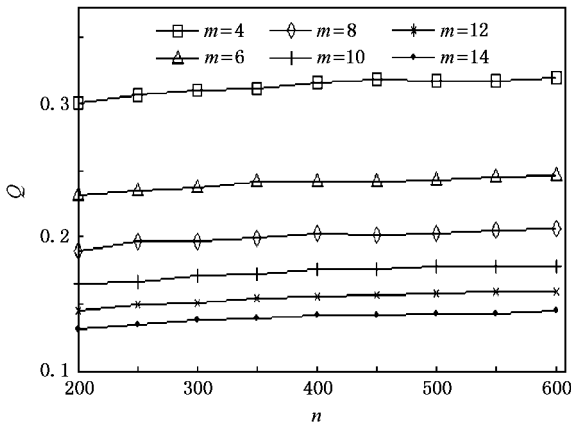


图5 B-A模型网络的模块性随 m 和 n 的变化

的随机网络,一般在2倍左右.由此可见,B-A模型所形成的无标度网络的确存在小世界现象,但是相对较弱.B-A模型的模块性值与随机网络基本一致,而且基本都小于0.3,即B-A模型所形成的无标度网络没有明显的社区结构.随着 m 的增加,B-A模型的平均路径长度和模块性逐渐减小,而聚类系数逐渐增加.不同的网络规模,上述规律不仅一致,而且其标准化之后的数值差异也不大.由此可见, m 是影响网络性质的主要参数.

图5进一步表明,B-A模型所形成的无标度网络的模块性同样受到网络密度的影响.当 m 增加时,网络连接稠密性也随之增加,其模块性值逐渐变小,网络社区结构就不清晰.对于同一 m 值,不同网络规模时,也可以发现相同的规律,但是网络规模的影响相对 m 要小得多.

4. 讨 论

4.1. 社区结构是网络的基本结构特征之一

与随机网络相比,规则网络和W-S模型的社区结构明显,而B-A模型则没有明显的社区结构特征.W-S模型的社区结构受到重新连接概率的影响,随着连接概率的增加,W-S模型模块性不是单调增减,而是出现明显的波峰.大量仿真结果显示,当 $0.01 \leq p \leq 0.1$ 时,W-S模型的社区结构最为清晰.文献[44,45]对于网络结构与知识传播进行的研究发现,在W-S模型的传播过程中,知识的平均水平不是重新连接概率 p 的单调函数,而是在 $0.01 \leq p \leq 0.1$ 区间存在明显的峰值,且峰值一般出现在 $p =$

0.06附近,本文的结果与这一研究结果相符.Watts和Strogatz关于小世界现象的判断标准只是一个定性的描述^[11].结合社区结构特征的研究,本文认为可以进一步明确该判断标准,即W-S模型具有与相应随机网络相当的平均路径长度和较大的聚类系数以及比相应规则网络大的模块性.对于图3所示的多数W-S模型而言,当 $0.01 \leq p \leq 0.1$ 时,小世界现象最为显著.

从模块性角度可以清楚地发现B-A模型与W-S模型的不同,正如Newman^[20,37]指出,模块性应该作为复杂网络指标之一.表1列出了不同网络特性之间的比较.

表1 不同网络的参数比较

网络类型	平均路径长度	聚类系数	模块性	度分布
规则网络	大	大	大	—
随机网络	小	小	小	泊松分布
W-S模型	小	大	大	不确定
B-A模型	小	较大	小	幂律分布

4.2. 模块性受到网络稀疏程度的影响

规则网络、随机网络、W-S模型以及B-A模型的社区结构划分结果都表明,网络越稀疏,其模块性值越大.由于网络结构差异较大,很难从理论上分析不同性质网络模块性的变化规律.特别地,对于如图1(a)所示的规则网络而言,其模块性 Q 为

$$Q = \frac{nk - tk'(k' + 1)}{nk} - \sum_{i=1}^t \left(\sum_{q=1}^t m_{iq} \right)^2, \quad (9)$$

式中

$$k' = \frac{k}{2},$$

$$\sum_{i=1}^t \left(\sum_{q=1}^t m_{iq} \right) = 1.$$

当

$$\sum_{q=1}^t m_{1q} = \sum_{q=1}^t m_{2q} = \dots = \sum_{q=1}^t m_{tq}$$

时 $\sum_{i=1}^t \left(\sum_{q=1}^t m_{iq} \right)^2$ 最小,所以有

$$Q = 1 - \left(\frac{k+2}{4n}t + \frac{1}{t} \right). \quad (10)$$

如果 $\frac{k}{n}$ 很小,即网络的连接稀疏,模块性 Q 将变大.社区数 t 是影响模块性值的另外一个因素.如果

$$t = \sqrt{\frac{4n}{k+2}},$$

且 t 是整数,就有

$$Q_{\max} = 1 - \sqrt{\frac{k+2}{n}}. \quad (11)$$

对于规则网络,如果 $k \leq \ln(0.49n - 2)$, 则 $Q_{\max} \geq 0.3$. 显然这个条件是很容易满足的,因此规则网络的模块性值一般都比较大,具有明显的社区结构. W-S 模型在重新连接概率 p 很小的时候,基本没有破坏规则网络的社区结构,而且因为长随机连接的出现,可能使得社区结构变得更加清晰.但是随着 p 的不断增大,规则网络的结构向随机网络方向变化很大,使得网络的社区结构破坏严重,这可能是 W-S 模型的模块性随 p 值非单调变化的主要原因.对于 B-A 模型,随着 m 的增加,一方面使得网络更加稠密.另一方面,由于偏好依附连接的作用,使得节点连接本质体现出有选择的随机性,从而导致其社区结构特征不明显.

4.3. 无标度网络与小世界网络的差别

与对应的随机网络相比, B-A 模型无标度网络具有相似的平均距离和较大的聚类系数,所以有学者认为无标度网络是一类小世界网络^[13].事实上,上述大量仿真结果显示,虽然无标度网络的聚类系数大于相应的随机网络,但一般只有 2 倍左右,与 Watts 提出的标准^[11]相差较远.另外, W-S 模型与 B-A 模型还有其他一些不同之处.首先,两者的构造过程不一样, W-S 模型是基于规则网络的随机变化,而 B-A 模型则强调网络构造的动态(节点增加)和节点之间联系(偏好依附连接).其次,两者的表现不同, W-S 模型强调局部连接丰富,而随机长连接稀少,其度分布基本对称,不满足幂律分布,而 B-A 模

型强调“核心节点”的出现,这些核心节点占有了绝大部分网络资源,其度分布满足幂律分布.另外,如果从模块性来判断, W-S 模型的社区结构特征明显,而 B-A 模型几乎不具备社区结构特征.总之,本文认为 B-A 模型与 W-S 模型是两类复杂网络, B-A 模型形成的无标度网络仅具有小世界现象.

4.4. 社区结构与复杂网络特征间的关系

小世界现象揭示的稀疏随机长连接伴随丰富的局部连接的网络结构特征,本质上就是社区结构的一种阐释,而清晰的社区结构可能是小世界现象产生的条件之一.为了进一步加以说明,考虑按下述模型生成网络,简称为 R-R 网络:规模为 $\{n_1, n_2, \dots, n_s\}$ 的 s 个网络,各网络内部节点间的连接概率为 $\{p_1, p_2, \dots, p_s\}$,不同网络间的节点按照概率 p_c 连接,就构成一个新网络.

表 2 给出了相应网络规模和网络内连接概率取值规律不同时的试验结果,具体包括三种情况.第一种情况是 $n_1 = n_2 = \dots = n_{15} = 10$, $p_1 = p_2 = \dots = p_{15} = 0.5$, $p_c = 0.01$. 第二种情况是 $n_1 = n_2 = \dots = n_{15} = 10$, $p_1 = p_2 = \dots = p_7 = 0.6$, $p_8 = p_9 = \dots = p_{15} = 0.3$, $p_c = 0.01$. 第三种情况是 $n_1 = n_2 = n_3 = 10$, $n_4 = n_5 = \dots = n_{10} = 6$, $n_{11} = n_{12} = \dots = n_{15} = 15$, $p_1 = p_2 = p_3 = 0.4$, $p_4 = p_5 = \dots = p_{10} = 0.5$, $p_{11} = p_{12} = \dots = p_{15} = 0.3$, $p_c = 0.01$.

每一种情况下产生的 R-R 网络与对应的随机网络的网络规模和连接总数相同.

表 2 R-R 网络与随机网络的参数对比

		平均度	平均路径长度	聚类系数	模块性
情况 1	R-R 网络	4.6267	2.7523	0.4015	0.6150
	随机网络	4.6267	2.4815	0.0659	0.2700
情况 2	R-R 网络	4.3067	2.8194	0.3675	0.6280
	随机网络	4.3067	2.5502	0.0635	0.2896
情况 3	R-R 网络	4.1905	2.8288	0.2907	0.5696
	随机网络	4.1905	2.5681	0.0550	0.2841

注:由于 R-R 网络是随机产生的,因此每次试验网络的平均度并不相同,为了便于和所对应的随机网络更好比较,所以表中仅给出了一次随机试验的结果,而非统计结果.多次试验的结果显示,表中的规律依然成立.

很显然,如 $p_i \gg p_c$, 则该网络具有明显的社区结构特征,与表中的结果一致.表 2 的试验结果也显示相关网络具有明显的小世界现象. R-R 模型与

W-S 模型差异很大,仅依据社区结构随机构造网络,却也表现出小世界现象,从而说明小世界现象与社区结构揭示网络的共同结构特征.清晰的社区结构

可以强化小世界现象,而小世界现象反过来又可能使得社区结构更加显著.社区结构与无标度特性间的关系相对复杂,无法用因果关系简单概括.一方面,无标度特性强调网络成员对网络资源的不平等占有,无标度特性揭示的节点度所遵循的幂律分布要求少数节点和其他多数节点直接相连,从而可能导致社区结构不够清晰.另一方面,社区结构是网络节点层次和整体层次的桥梁,它将网络分块,要求块与块之间连接稀疏,从而大大降低了少数节点控制整个网络的可能,因此清晰的社区结构可能使得无标度特性减弱.由于网络的复杂性以及相关参数计算的困难,故难以从理论上探讨上述关系.

5. 结 论

本文分析了 W-S 模型和 B-A 模型的社区结构特征,通过模块性指标 Q 的计算,探讨了 W-S 模型

和 B-A 模型的异同.研究表明,W-S 模型具有显著的社区结构,而 B-A 模型的社区结构特征不明显.网络清晰的社区结构是小世界现象产生的重要原因之一,但是社区结构不同于小世界现象和无标度特性,并可以利用模块性区别网络结构类型,因此网络复杂性特征应该包括社区结构.同时,W-S 模型的模块性随连接概率变化的特点表明,当 $0.01 \leq p \leq 0.1$ 时,其模块性大于规则网络,小世界现象最为显著.另外,本文也发现网络密度对社区结构模块性指标影响显著.

实际社会网络一般比较稀疏且结构复杂,可能同时具有小世界现象、无标度特性以及明显的社区结构^[43,46].目前的网络模型,不论是随机网络、规则网络,还是 W-S 模型和 B-A 模型,都不会同时具有上述复杂特征.因此,需要发展新的网络模型,以充分反应真实网络的特征,而在新网络模型中,社区结构特征是必需要考虑的因素之一.

- [1] Wasserman S, Faust K 1994 *Social Network Analysis: Methods and Applications* (New York, Cambridge: Cambridge University Press)
- [2] Scott J 2002 *Social Network Analysis: A Handbook* (London: Sage)
- [3] Dorogovtsev S N, Mendes J F 2001 *Phys. Rev. E* **63** 056125
- [4] Dorogovtsev S N, Mendes J F 2002 *Adv. Phys.* **51** 1079
- [5] Albert R, Barabási A L 2002 *Rev. Mod. Phys.* **74** 48
- [6] Watts D J, Dodds P S, Newman M E J 2002 *Science* **296** 1302
- [7] Li J, Wang B H, Jiang P Q, Zhou T, Wang W X 2006 *Acta Phys. Sin.* **55** 4051 (in Chinese) [李季、汪秉宏、蒋品群、周涛、王文旭 2006 物理学报 **55** 4051]
- [8] Zhou T, Fu Z Q, Niu Y W 2005 *Prog. Nat. Sci.* **15** 513 (in Chinese) [周涛、傅忠谦、牛永伟 2005 自然科学进展 **15** 513]
- [9] Xiao F W, Chen G R 2003 *IEEE Circ. Sys. Mag.* **3** 6
- [10] Kleinberg J 2000 *Nature* **406** 845
- [11] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [12] Barabási A L, Albert R R, Jeong H 1999 *Physic A* **272** 173
- [13] Amaral L A N, Scala A, Barthélémy M, Stanley H E 2000 *Proc. Natl. Acad. Sci. USA* **97** 11149
- [14] Flake G W, Lawrence S R, Giles C L, Coetzee F 2002 *IEEE Computer* **35** 66
- [15] Girvan M, Newman M E J 2002 *Proc. Natl. Acad. Sci. USA* **99** 7821
- [16] Holme P, Huss M, Jeong H 2003 *Bioinformatics* **19** 532
- [17] Guimer'a R, Amaral L A N 2005 *Nature* **433** 895
- [18] Mercedes P, Jennifer A D 2006 *Ecological Networks: Linking Structure to Dynamics in Food Webs* (Oxford: Oxford University Press)
- [19] Luo J D 2005 *Social Network Analysis* (Beijing: Social Science Academic Press) (in Chinese) [罗家德 2005 社会网分析讲义 (北京: 社会科学文献出版)]
- [20] Newman M E J 2004 *Eur. Phys. J. B* **38** 321
- [21] Wang L, Dai G Z 2005 *Sci. Techn. Rev.* **23** 62 (in Chinese) [王林、戴冠中 2005 科技导报 **23** 62]
- [22] Barabási A L, Albert R 1999 *Science* **286** 509
- [23] Davidsen J, Holger E, Stefan B 2002 *Phys. Rev. Lett.* **88** 128701
- [24] Konstantin K, Victor M 2002 *Phys. Rev. E* **65** 057102
- [25] Latoral V, Marchiori M 2003 *Eur. Phys. J. B* **32** 249
- [26] Guo J L 2006 *Acta Phys. Sin.* **55** 3916 (in Chinese) [郭进利 2006 物理学报 **55** 3916]
- [27] 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、何大韧 2006 物理学报 **55** 60]
- [28] Pan Z F, Wang X F 2006 *Acta Phys. Sin.* **55** 4058 (in Chinese) [潘灶烽、汪小帆 2006 物理学报 **55** 4058]
- [29] 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]
- [30] Newman M E J 2002 *Phys. Rev. Lett.* **89** 208701
- [31] Radicchi F, Castellano C, Ceconi F, Loreto V, Parisi D 2004 *Proc. Natl. Acad. Sci. USA* **101** 2658
- [32] Wu F, Huberman B A 2004 *Eur. Phys. J. B* **38** 331
- [33] Newman M E J 2003 *Phys. Rev. E* **67** 026126
- [34] Newman M E J 2003 *SIAM Rev.* **45** 167
- [35] Newman M E J, Girvan M 2003 *Statistical Mechanics of Complex Networks* (Berlin: Springer)

- [36] Newman M E J , Girvan M 2004 *Phys. Rev. E* **69** 026113
- [37] Newman M E J 2004 *Phys. Rev. E* **69** 066133
- [38] Aaron C , Newman M E J 2004 *Phys. Rev. E* **70** 066111
- [39] Scott W , Smyth P 2005 *International Conference on Data Mining* (Newport Beach : Society for Industrial and Applied Mathematics)
- [40] Barabási A L 2002 *Rev. Mod. Phys.* **74** 48
- [41] Newman M E J 2005 *Contem. Phys.* **46** 323
- [42] Newman M E J , Juyong P 2003 *Phys. Rev. E* **68** 036122
- [43] Du H F , Marcus W F , Li S Z , Jin X Y 2007 *Complexity J.* **13** (3) 53
- [44] Cowan R , Jonard N , Ozman M 2004 *J. Econ. Dynam. Contr.* **28** 1557
- [45] Morone P , Taylor R 2004 *J. Evolut. Econ.* **14** 327
- [46] Li S Z , Yang X S , Jin X Y , Marcus W F , Du H F 2006 *Mark. Demog. Anal.* **13** (5) 13 (in Chinese) [李树茁、杨绪松、靳小怡、Marcus W F、杜海峰 2006 市场与人口分析 **13** (5) 13]

Community structure in small-world and scale-free networks^{*}

Du Hai-Feng^{1 2 3 †} Li Shu-Zhuo¹ W. F. Marcus²

Yue Zhong-Shan⁴ Yang Xu-Song⁴

1 *Institute for Population and Development Studies , Xi'an Jiaotong University , Xi'an 710049 , China*

2 *Morrison Institute for Population and Resource Studies , Stanford University , Stanford 94305 , USA*

3 *School of Mechanical Engineering , Xi'an Jiaotong University , Xi'an 710049 , China*

4 *School of Management , Xi'an Jiaotong University , Xi'an 710049 , China*

(Received 4 March 2007 ; revised manuscript received 25 May 2007)

Abstract

Modularity is one of the important parameters to evaluate the community structure of network. The community structure of small-world networks (Watts-Strogatz model) and scale-free networks (Barabási-Albert model) were explored in the framework of the modularity using numerical simulations. We found that , firstly , the community structure in Watts ' small-world networks is clear , while in Barabási ' s scale-free networks it is unclear ; secondly , community structure is affected by the density of the network , namely , the sparser the network connections , the clearer is the community structure ; thirdly , modularity in Watts ' small-world networks is a dynamic variable which changes as a nonmonotonic function of the probability of rewiring connections. Hence , community structure should be one of the complex network properties. The real social networks may exhibit clear community structure , small-world phenomena and scale-free properties simultaneously. None of the current network models , such as the random network , Watts-Strogatz model , or Barabási-Albert model , fits these social networks very well. A new social network structure is pending construction.

Keywords : modularity , community structure , small-world network , scale-free network

PACC : 0590 , 0175

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 70671083 , 50505034) , the National Social Science Foundation of China (Grant Nos. 05ARK001 , 05CRK002) , the Program for the New Century Excellent Talents in University of Ministry of Education , China (NCET-04-0931) , the National Education Promotion Program of China (Grant No. 07200701) , the International Cooperation Program of Santa Fe Institute , USA and the Seed Foundation from the Center for Demography and Economics of Health and Aging at Stanford University , USA (Grant No. AG017253-06).

[†] E-mail : haifengdu@mail.xjtu.edu.cn