

BA指数网络结构特性精确计算

章忠志*, 荣莉莉

(大连理工大学 系统工程研究所, 辽宁 大连 116024)

摘要: 度分布和平均路径长度是复杂网络的两个重要结构特性. 对于随机增长的网络, 其平均路径长度一般主要通过计算机模拟给出数值结果, 尚无一个普适的解析计算方法. 为此首先利用主方程的方法对 BA 随机指数网络的度分布进行解析推导, 与 Barabási 等的结果相比所得度分布与计算机模拟值更接近. 然后对 BA 随机指数网络和 BA 确定性指数网络的平均路径长度进行解析计算, 所得结果与经典的 ER 随机图相似, 即平均路径长度以网络大小的对数形式增长. 此外, 对 BA 随机指数网络平均路径长度的模拟值与解析计算结果相吻合. 最后, 对 BA 确定性指数网络的度分布与直径进行了解析计算, 并对两个网络的结构特性作了比较. 比较结果表明, BA 随机指数网络和 BA 确定性指数网络的结构性质虽然存在量上的差异, 但从定性角度来说, 其拓扑结构是相同的.

关键词: 复杂网络; BA 网络; 度分布; 平均路径长度; 指数网络
中图分类号: N94 **文献标识码:** A

0 引言

现实世界中许多系统都可以用复杂网络来描述^[1-4], 网络节点为系统元素, 边为元素间的相互作用和相互联系 (即关系). 复杂网络作为研究复杂性科学和复杂系统的有力工具, 近年来倍受国内外学者的广泛关注^[1-4]. 最经典的复杂网络有: Erdős 与 Rényi 的随机图^[5] (ER 随机图)、Watts 和 Strogatz 的小世界网络^[6] (WS 模型)、Barabási 和 Albert 的无尺度网络^[7-8] (BA 模型或 BA 网络). WS 模型和 BA 网络提出后, 在不同学科和领域内掀起了复杂网络^[1]、WS 模型^[9] 与小世界网络^[10-11]、BA 模型^[12] 与无尺度网络^[13] 的研究热潮. 在复杂网络的众多结构特性中, 度分布、平均路径长度 (或平均距离) 是两个重要的特征度量^[1]. 度分布表示节点度的概率分布函数 $P(k)$, 它指的是节点有 k 条边连接的概率. 平均路径长度是衡量网络平均传输性能与效率的参数. 到目前为止, 随机网络的平均路径长度主要通过模拟得到, 还没有一般性的解析计算方法. 复杂网络可以按照不同的方式来分类, 如根据节点度分布的不同, 可以将复杂网络分为指数网络和无尺度网络两大类. 指数网络的度分布 $P(k)$

随度 k 呈指数衰减, ER 随机图和 WS 模型属于指数网络; 无尺度网络的节点度 k 服从幂律分布, 即 $P(k) \propto k^{-\gamma}$, BA 模型属于此类网络. 从生成方式上又可将复杂网络分成随机网络和确定性网络.

本文研究 BA 网络的两个变体模型: BA 随机指数网络和 BA 确定性指数网络. 首先解析计算这两个网络的平均路径长度, 对随机情形, 利用主方程的方法, 解析并模拟计算度分布; 最后计算 BA 确定性指数网络度分布与直径的精确值, 并比较研究两个网络的主要结构特性.

1 BA随机指数网络

BA 模型对网络科学的发展起着重要的作用, 该模型捕捉到了无尺度网络形成的两个必不可少的机制——增长和择优连接^[7-8]. 为了验证这两个机制的必要性, Barabási 等研究了 BA 网络的两个变体模型^[8]. 第一个变体模型保留了 BA 网络的增长特性, 但新节点以等概率的形式与原系统的老节点进行连接. 具体生成过程如下:

(1) 增长

初始时刻 ($t = m$), 网络为 $m + 1$ 个节点连接成的完全图, 在以后的每一个时间间隔中, 新增一

收稿日期: 2005-07-10; 修回日期: 2006-12-23.
基金项目: 国家自然科学基金资助项目 (重点项目 70431001; 70571011).
作者简介: 章忠志* (1974-), 男, 博士; 荣莉莉 (1964-), 女, 教授, 博士生导师.
©1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

个度为 m 的点, 连接到网络中已经存在的 m 个不同的节点上. 需要指出的是, 网络初始状况对其最终演化成的拓扑结构影响甚微^[1-4], 这里为方便起见, 假设初始网络为完全图.

(2) 均匀连接

假定新节点与系统中原有节点的连接概率是相同的, 即与原有节点的度无关.

这样, 在时间步等于 t 时, 网络的节点总数 $N = t + 1$. Barabási 等^[8]提出并应用平均场近似的方法计算出其度分布为

$$P(k) = \frac{e}{m} \exp\left(-\frac{k}{m}\right) \tag{1}$$

该网络的度分布呈现出指数特性, 而且新节点与原有节点的连接是完全随机的, 故称之为 BA 随机指数网络.

1.1 度分布

由于 Barabási 等利用平均场近似的方法^[8]推导式 (1) 的过程不够严密, 导致式 (1) 表达的度分布不够精确, 误差较大^[14-15]. 为了得到关于 BA 随机指数网络度分布的精确表达式, 本文利用 Dorogovtsev 等提出的主方程的方法^[16-17]计算了该网络的度分布, 得到与式 (1) 不同的如下表达式:

$$P(k) = \frac{1}{m+1} \left[\frac{m+1}{m} \right]^m \left[\frac{m}{m+1} \right]^k; k \geq m \tag{2}$$

通过比较 (1)、(2) 两式不难发现, 对于较大的 k , 式 (1) 的结果要小于式 (2). 下面给出式 (2) 的具体推导过程.

根据节点的生成时间, 对节点进行标号, 即 $s = 0, 1, 2, \dots, t$. 设 $p(k, s, t)$ 表示 t 时刻, 节点 s 的度为 k 的概率, 则单个节点的度分布满足下面的主方程^[14-15]:

$$p(k, s, t+1) = \frac{m}{t+1} p(k-1, s, t) + \left[1 - \frac{m}{t+1} \right] p(k, s, t) \tag{3}$$

该方程的初始条件和边界条件分别为

$$p(k, s = \{0, 1, 2, \dots, m\}, t = m) = W_{k,m},$$
$$p(k, t, t) = W_{k,m}$$

其中 $W_{k,m} = \begin{cases} 1, k = m \\ 0, \text{其他} \end{cases}$.

整个网络的度分布可表示为

$$P(k, t) = \frac{1}{t+1} \sum_{s=0}^t p(k, s, t) \tag{4}$$

在方程 (3) 的两边应用求和公式 $\sum_{s=0}^t$ 对 s 求和, 可得如下关于整个网络度分布的方程:

$$(t+2)P(k, t+1) - (t+1)P(k, t) = mP(k-1, t) - mP(k, t) + W_{k,m} \tag{5}$$

当 $t \rightarrow \infty$ 时, 可以证明 $\lim_{t \rightarrow \infty} P(k, t)$ 存在^[18], 即 $\lim_{t \rightarrow \infty} P(k, t+1) = P(k, t)$, 因此方程 (5) 对应的稳态方程为

$$(m+1)P(k) - mP(k-1) = W_{k,m} \tag{6}$$

利用方程 (6), 可求得网络的度分布式 (2).

式 (2) 对所有 $k \geq m$ 都成立; 当 $k < m$, 显然有 $P(k) = 0$. 由式 (2) 知, BA 随机指数网络的度分布随着度的增加以指数形式下降. 特别地, 当 $m = 1$ 时, 度分布为 $P(k) = 2^{-k}$.

图 1 给出了度分布模拟结果的半对数图形, 图中符号 $\diamond, \triangle, \square$ 和 $\circ$ 分别表示规模为 30 万个节点的网络在 $m = 1, 3, 5, 7$ 时运行 10 次的平均模拟值. 从图 1 可以看出, 本文给出的解析表达式与模拟值更吻合.

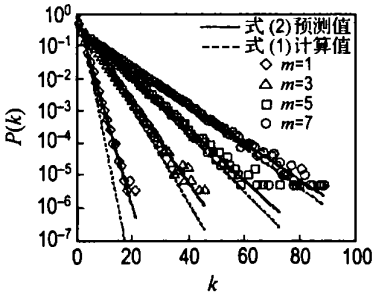


图 1 BA 随机指数网络的度分布

Fig. 1 Degree distributions of exponential BA networks

1.2 平均路径长度

对大多数现实网络而言, 不管规模多大, 其平均路径长度一般都很小. 平均路径长度指网络所有节点对之间的平均最短距离, 而节点间的距离指从一节点到另一节点所要经历边的最小数目, 其中所有节点对之间的最大距离称为网络的直径.

目前, 对于随机复杂网络的平均路径长度还没有一个普遍适用的解析计算方法, 人们一般通过计算机模拟得到其值. 本文利用类似于文献 [19] 中的方法, 解析求出了 BA 随机指数网络的平均路径长度. 先考虑 $m = 1$ 的情况, 对这一特殊情况, 有下面的引理.

引理 1 对任意两个节点 i 和 j , 每条从 i 到 j 的最短路径 SP_{ij} 都不经过任何节点 k , 其中 k 满足条件 $k > \max(i, j)$.

证明 用节点序列 $i \rightarrow v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_n \rightarrow$

j 表示从 i 到 j 长度为 $n+1$ 的一条最短路径 SP_{ij} . 显然, 当 $n=0$ 时, 引理成立. 当 $n>0$ 时, 假设 $v_k = \max(v_1, v_2, \dots, v_n)$, 如果 $v_k < \max(i, j)$, 则命题成立. 接下来用反证法证明 $v_k > \max(i, j)$ 不可能发生. 根据该网络的构造算法, 当节点 v_k 进入系统时, 只与一个节点进行连接. 而当 $v_k > \max(i, j)$ 时, 在路径序列 $i \rightarrow v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_n \rightarrow j$ 中, 有两个比节点 v_k 年轻的节点与 v_k 相连, 与构造算法矛盾. 因此, 如果 SP_{ij} 是从 i 到 j 的一条最短路径, 则路径序列中“最老”的节点是 i 或 j . $\square$

设 $d(i, j)$ 表示 i 到 j 的最短距离, $e(N)$ 表示顶点数为 N 的网络的所有节点对的距离和, 即

$$e(N) = \sum_{0 \leq i, j \leq N-1} d(i, j) \tag{7}$$

用 $L(N)$ 表示网络的平均路径长度, 则

$$L(N) = \frac{2e(N)}{N(N-1)} \tag{8}$$

根据引理, 新增节点不影响已经存在的节点之间的距离, 因此, 下面的式子成立:

$$e(N+1) = e(N) + \sum_{i=0}^{N-1} d(i, N) \tag{9}$$

假设节点 N 进入系统时与节点 w 相连, 则式 (9) 可以写成

$$e(N+1) = e(N) + \sum_{i=0}^{N-1} (d(i, w) + 1) = e(N) + N + \sum_{i=0}^{N-1} d(i, w) \tag{10}$$

方程 (10) 中 $\sum_{i=0}^{N-1} d(i, w)$ 表示节点集 $J = \{0, 1, 2, \dots, N-1\} \setminus \{w\}$ 中的所有 $N-1$ 个节点到一个特定节点 w 的距离和. 由于所讨论的模型属于指数网络, 考虑到指数网络的同质性^[1] 并利用式 (7), $\sum_{i=0}^{N-1} d(i, w)$ 可以用下式替代:

$$\sum_{i=0}^{N-1} d(i, w) = (N-1)L(N) = \frac{2e(N)}{N} \tag{11}$$

将式 (11) 代入方程 (10), 得

$$e(N+1) = e(N) + N + \frac{2e(N)}{N} \tag{12}$$

根据式 (12), 可得 $e(N)$ 关于 N 的变化方程

$$\frac{de(N)}{dN} = N + \frac{2e(N)}{N} \tag{13}$$

由方程 (13) 可得

$$e(N) = N^2(\ln N + _) \tag{14}$$

其中 $_$ 为常数. 当 $N \rightarrow \infty$ 时, 将式 (14) 代入式 (8), 得

$$L(N) = 2(\ln N + _) \tag{15}$$

因此, 当 $m=1$ 时, BA 随机指数网络的平均路径长与网络阶的对数成正比, 这一结果与 ER 随机图相似^[1, 5]. 图 2 给出了平均路径长度与网络规模关系的半对数图, 图中每个圆圈表示同一个网络运行 10 次的平均模拟结果, 直线为式 (15) 的解析值. 从图 2 容易看出, 解析结果与模拟值一致.

当 $m>1$ 时, 随着 m 的增大, 网络变得越来越稠密, 当顶点数相同时, 网络的平均路径长度随着 m 的增加而减小. 因此, BA 指数网络的平均路径长度最多以网络阶的对数形式增长.

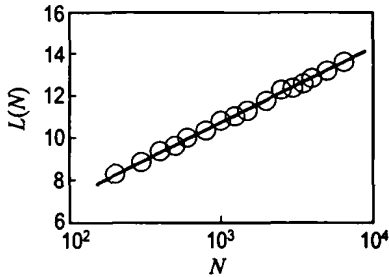


图 2 BA 随机指数网络的平均路径长度

Fig. 2 Average path lengths of exponential BA networks

2 BA 确定性指数网络

当 $m=1$ 时, BA 随机指数网络对应着一个确定性模型^[20], 由于该确定性网络的度分布随着度的增加以指数形式下降, 本文称之为 BA 确定性指数网络. 文献 [20] 给出了该确定性网络的构造方法, 描述如下: 初始时刻 ($t=0$), 网络由一条边连接起来的两个节点组成, 在随后的每一个时间步, 网络的每一个节点都生成一个新节点, 并与对应的新节点进行连接. 于是, 在时间步为 t 时, 网络的节点总数目为 $N_t = 2^{t+1}$. 文献 [20] 只给出了 BA 确定性指数网络的构造方法, 下面对该网络的主要性质 (度分布、平均路径长度和直径) 进行精确求解.

2.1 度分布

该确定性网络节点度的谱分布是离散的, 经过 t 个时间步后, 网络中度 $k=1, 2, 3, \dots, t-1, t, t+1$ 的节点数目 $m(k, t)$ 分别为 $2^t, 2^{t-1}, 2^{t-2}, \dots, 2^2, 2, 2$. 于是, 整个网络的度分布为

$$P(k) = \begin{cases} \frac{m(k, t)}{N_t} = 2^{-k}, & k = 1, 2, \dots, t \\ \frac{m(k, t)}{N_t} = 2^{t-k}, & k = t+1 \end{cases} \tag{16}$$

由式 (16) 易知, $P(k)$ 随着 k 的增加以指数形式下降. 由式 (16) 及文献 [4] 中的结论, 可知该确定性网络的累积度分布也以相同的指数幂下降.

2.2 平均路径长度

由于该网络是确定性的, 其平均路径长度可以通过计算得到精确值. 设 $l_t(k)$ 表示在时间步 t 时, 度为 k 的节点到网络中其他节点最短距离的平均值, 由文献 [20], $l_t(k)$ 的表达式为

$$l_t(k) = \frac{2}{2^{t+1} - 1} (2t + 2 - k) \quad (17)$$

于是当网络的顶点数 N_t 很大时, 有下式成立:

$$l(k, N_t) \cong \frac{\ln N_t}{\ln 2} - \frac{k}{2} \quad (18)$$

因此, 整个网络的平均路径长度

$$L(N_t) = l(k, N_t) P(k) \cong \frac{\ln N_t}{\ln 2} - 1 \quad (19)$$

可见, 该网络的平均路径长度是网络大小 N_t 的对数的线性函数, 这一性质与 ER 随机图一致.

2.3 直径

除了平均路径长度外, BA 确定性指数网络的直径也可以通过计算得到精确值. 设 $G(t)$ 表示时间步等于 t 时的 BA 确定性指数网络, $D(t)$ 表示 $G(t)$ 的直径, 则 $D(t)$ 一定是在时间 t 生成的节点对之间达到. 可以证明在时间步等于 t 时, $D(t) = 2t + 1$. 当 $t = 0, 1$ 时, 显然成立. 假设在时间步等于 $t - 1$ 时, 结论成立, 即 $D(t - 1) = 2(t - 1) + 1$. 由于在时间 t 生成的节点到 $G(t - 1)$ 的距离为 1, $D(t) = D(t - 1) + 2 = 2t + 1$, 所以命题成立. 由于网络阶的对数为 $\ln N_t = (t + 1) \ln 2$, $D(t)$ 又可用 $\ln N_t$ 表示成

$$D(t) = \frac{2}{\ln 2} \ln N_t - 1 \quad (20)$$

这样就证明了同平均路径长度一样, BA 确定性指数网络的直径也以网络大小的对数形式增长.

3 两种网络结构特性的比较

前文研究的 BA 随机指数网络和 BA 确定性指数网络同为指数网络, 它们的结构特性有何异同呢? 下面比较研究两者之间的关系. 为方便比较, 对 BA 随机指数网络, 这里仅考虑 $m = 1$ 的情况. 由 (2)、(16) 两式可以看出, BA 随机指数网络在 $m = 1$ 时的度分布与 BA 确定性指数网络相同. 而对于平均路径长度, 通过比较式 (15) 与式 (19) 可知, 尽管两者的度分布相同, 但当网络的规模很大时, BA 确定性指数网络的平均路径要比相同规模的 BA 随机指数网络的小, 而后的直径也

以网络阶的对数形式增长. 虽然两者的平均路径长度存在具体数值上的差异, 但都遵循相同的增长规律, 即以网络规模的对数形式增长. 下面, 顺便介绍一下网络的另外一个重要结构特性——集聚系数^[6]. 节点的集聚系数定义为与该节点直接相邻的节点间实际存在的边数目占最大可能存在的边数的比例; 而网络的集聚系数为所有节点集聚系数的平均值. 从定义知, 这两个网络的所有节点的集聚系数和整个网络的集聚系数均为零. 综上所述, 不难得出结论: 两个网络的结构特性虽然存在量上的差异, 但从定性的角度来说, 它们的性质是相同.

4 结 语

本文利用主方程方法对 BA 随机指数网络的度分布进行了解析计算, 与 Barabási 等的结果比较表明, 所得表达式的结果与数值模拟更吻合; 然后对 $m = 1$ 这一情况的平均路径长度进行了解析计算. 计算机模拟说明了解析方法及结果的有效性和准确性. 文章还对 BA 确定性指数网络的主要结构特性进行精确求解, 并与 BA 随机指数网络的性质进行一番比较, 结果表明两个指数网络的结构特性从定性的角度来说是一致的.

参考文献:

- [1] ALBERT R, BARABÁSI A L. Statistical mechanics of complex networks [J]. *Rev Modern Phys*, 2002, 74(1): 47-97
- [2] WANG X F. Complex networks topology, dynamical and synchronization [J]. *Int J Bifurcation and Chaos*, 2002, 22(5): 885-916
- [3] 吴金闪, 狄增如. 从统计物理学看复杂网络研究 [J]. *物理学进展*, 2004, 24(1): 19-46
- [4] NEWMAN M E J. The structure and function of complex networks [J]. *SIAM Rev*, 2003, 45(2): 167-256
- [5] ERDŐS P, RÉNYI A. On the evolution of random graphs [J]. *Publ Math Inst Hungarian Acad Sci*, 1960, 5: 17-61
- [6] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks [J]. *Nature*, 1998, 393: 440-442
- [7] BARABÁSI A L, ALBERT R. Emergence of scaling in random networks [J]. *Science*, 1999, 286: 509-512
- [8] BARABÁSI A L, ALBERT R, JEONG H. Mean-field theory for scale-free random networks.

- [J]. **Physica A**, 1999, **272**: 173–187
- [9] NEWMAN M E J. Models of the small world [J]. **J Stat Phys**, 2000, **101**: 819–841
- [10] ZHANG Z Z, RONG L L, COMELLAS F. Evolving small-world networks with geographical attachment preference [J]. **J Phys A Math and Gen**, 2006, **39**(13): 3253–3261
- [11] ZHANG Z Z, RONG L L, GUO C H. A deterministic small-world network created by edge iterations [J]. **Physica A**, 2006, **363**: 567–572.
- [12] 章忠志, 荣莉莉. BA 网络的一个等价演化模型 [J]. **系统工程**, 2005, **23**(2): 1–5
- [13] ZHANG Z Z, RONG L L, COMELLAS F. High dimensional random Apollonian networks [J]. **Physica A**, 2006, **364**: 610–618
- [14] BOLLOBAS B, RIORDAN O M, SPENCER J, *et al*. The degree sequence of a scale-free random graph process [J]. **Random Struct and Algorithms**, 2001, **18**: 279–290
- [15] GUO J L, BAI Y Q. A note on mean-field theory for scale-free random networks [J]. **Dyn of Continuous Discrete and Impulsive Syst Series B-Appl & Algorithms**, 2006, **13**(3–4): 523–531
- [16] DOROGOV TSEV S N, MENDES J F F, SAMUKHIN A N. Structure of growing networks with preferential linking [J]. **Phys Rev Lett**, 2000, **85**(21): 4633–4636
- [17] DOROGOV TSEV S N, MENDES J F F, SAMUKHIN A N. Size-dependent degree distribution of a scale-free network [J]. **Phys Rev E**, 2001, **63**(6): 062101
- [18] OZIK J, HUNT B R, OTT E. Growing networks with geographical attachment preference Emergence of small worlds [J]. **Phys Rev E**, 2004, **69**(2): 026108
- [19] ZHANG Z Z, RONG L L, ZHOU S G. Evolving Apollonian networks with small-world scale-free topologies [J]. **Phys Rev E**, 2005, **74**(4): 046105
- [20] DOROGOV TSEV S N, MENDES J F F, OLIVEIRA J G. Degree-dependent intervertex separation in complex networks [J]. **Phys Rev E**, 2006, **73**(5): 056122

Exact computations for structural characteristics of BA exponential networks

ZHANG Zhong zhi^{*}, RONG Li li

(Inst. of Syst. Eng., Dalian Univ. of Technol., Dalian 116024, China)

Abstract Degree distribution and average path length are two important structural characteristics of complex networks. For randomly growing networks, there has not been a generic analytical computation technique for their average path length, which is obtained mainly by computer simulations. To solve this problem, firstly, the expression of degree distribution for the random exponential BA network is derived analytically by master-equation approach, which is more consistent with the simulations than that given by Barabási, *et al*. Then the emphasis switches to the computation of the average path length for both random and deterministic exponential BA networks. The obtained average path length is similar to the counterpart of ER random graph, which increases logarithmically with the network size. Also, simulations are done for the properties of the random situation, which are in good agreement with the analytical results. Furthermore, the degree distribution and diameter of the deterministic exponential BA network are found exactly. Finally, comparative research is done on the structural characteristics of the two networks. The obtained results show that the two networks have quantitatively different topological properties, while they are qualitatively similar.

Key words complex networks; BA networks; degree distribution; average path length; exponential networks