·复杂性科学·

演化网络的Mandelbrot律

任学藻[1]，杨紫陌[2]，汪秉宏[3,4]

(1. 西南科技大学理学院　四川 绵阳　621010;　2. 电子科技大学互联网科学中心　成都　610054;
3. 中国科技大学近代物理系　合肥　230026;　4. 上海理工大学复杂系统科学研究中心　上海 杨浦区　200093)

【摘要】本文研究了满足线性择优模式的生长网络，解析得到了度分布的递推表达式，在此基础上证明了在一定条件下其近似解析解符合Mandelbrot分布 $p(k)\propto(k+c)^{-\gamma}$ 分布，并得到了幂律指数 γ 和漂移量 c 的具体表达式。文章指出了传统解析中差分近似中存在着对漂移量 c 的估计偏差，通过数值对比实验，验证了本文所提出的方法的准确性。

关　键　词　复杂网络;　演化网络;　Mandelbrot分布
中图分类号　N94　　**文献标识码**　A　　doi:10.3969/j.issn.1001-0548.2011.02.001

Mandelbrot Law of Evolution Networks

REN Xue-zao[1], YANG Zi-mo[2], and WANG Bing-hong[3,4]

(1. School of Science, Southwest University of Science and Technology　Mianyang Sichuan　621010;
2. Web Science Center, University of Electronic Science and Technology of China　Chengdu　610054;
3. Department of Modern Physics, University of Science and Technology of China　Hefei　230026;
4. Research Center for Complex System Science, University of Shanghai for Science and Technology　Yangpu Shanghai　200093)

Abstract　This paper studies the growing networks with linear preferential attachment The recursive formula of degree distribution is presented. we prove that the approximate analytical solution of the recursive formula is in accordance with Mandelbrot law distribution $p(k)\propto(k+c)^{-\gamma}$ in certain conditions and then we obtain the concrete expressions of the scaling exponent γ and the shifting coefficient c. It is concluded that there is a deviation of shifting coefficient c in the approximation of traditional difference analysis. The accuracy of this method is confirmed by numerical experiments.

Key words　complex networks;　evolutionary networks;　Mandelbrot law distribution

现实世界中许多系统都可以用复杂网络进行描述[1-7]，网络节点可以广义地理解为系统元素和边为元素间的相互作用或联系。复杂网络将统计物理、随机数学与系统科学融为一体，不仅可以体现出生物、技术及社会等各种复杂系统的结构，而且是研究它们特征和功能的有效工具。其中，网络的演化模型及其背后隐藏的演化机制，是复杂网络研究的基本问题。近年来，文献[8-9]提出的小世界网络和无标度网络，为实际网络分析中观察到的随网络规模对数增长的平均距离、高于随机网络的局部集聚性和近似幂律的度分布等特征提供了较为合理的解释，成为推动复杂网络研究热潮的原创力量。

文献[8-9]认为，增长和择优是网络自组织演化过程中涌现出标度律的两个基本机制。它们从“富者更富”理念出发建立网络模型(BA网络)，引入择优连接 $\prod(k_i)=k_i\Big/\sum_j k_j$ 演化机制，通过平均场近似方法，解析出网络的度分布为 $p(k)=2m^2k^{-3}$，其中，$p(k)$表示度分布，$2m$为网络的平均度[9-10]。

事实上，真实网络的度分布往往不严格满足幂函数律，而仅仅是接近于幂律分布。如科学家合作网络的度分布服从带有指数截断的幂函数律[11]，用户−对象二部分图中，用户的度分布服从广延指数分布[12]；航空网络的度分布服从双段幂律分布[13]等。本文特别关注 $p(k)\propto(k+c)^{-\gamma}$，该分布最早被文献[14]提出，称为Mandelbrot分布律，也被称做漂移幂

收稿日期：2011－01－28
基金项目：国家自然科学基金(10635040)；国家自然基金重大研究计划(90924011)；四川省科技厅资助项目(2010HH0002)
作者简介：任学藻(1963－)，男，教授，主要从事凝聚态物理方面的研究.

律[15]，其中，γ 是幂律指数，c 是漂移量。最近的实证研究显示，很多网络的度分布可以用Mandelbrot分布律更好地进行刻画[16-17]，即使是BA网络，利用主方程严格计算的度分布为 $p(k)=\frac{2m(m+1)}{(k+2)(k+1)k}$ [3]，近似可化为 $p(k)\propto(k+1)^{-3}$，也满足Mandelbrot分布律。

可采用多种方法对演化网络的度分布进行解析，如平均场、主方程及率方程等[18-20]。以前的研究过分关注幂指数，忽略了对于漂移量的解析和数值研究，并且由于对微分方程近似处理不当，产生了较大的误差。本文利用率方程的方法[21-23]，引入与度 k 线性择优的连接方式，严格导出度分布的递推关系，并得到了更为精确的Mandelbrot律指数 γ 和漂移量 c 的具体解析表达式，指出了传统解析中差分近似存在的对漂移量估计的偏差。数值实验显示，漂移量c会对网络度分布的形态产生不可忽略的影响，并且本文得到的解析结果更加精确。

1 线性择优模型

本文的模型不同于经典的BA模型[9]，用线性择优的方式，实现了在网络增长时择优连接和随机连接的过渡。构建网络算法如下：1) 初始化。初始给定 m_0 个节点的全连通网络，其中相比于网络演化的最终节点数 N，m_0 要远小于 N；2) 演化。每个时间步，在原网络中增加一个新节点，并且从该新加入的节点引出 m 条边，与原网络中的节点相连，与每个点相连的概率与该节点的度线性相关，即：新加入的节点与原网络中的节点 i 连接的概率满足：

$$\prod(k_i)=\frac{\alpha k_i+\beta}{N}=\frac{2m(\alpha k_i+\beta)}{2mN}=\frac{2m(\alpha k_i+\beta)}{\sum_i k_i} \tag{1}$$

式中，k_i 为节点 i 的度数；α 和 β 为可调参数。本文将证明 α 和 β 中只有一个是独立变量。此外，需要注意排除自身到自身的连线和重复连线。

2 率方程解析

率方程的解析基于当 N 相对于 m_0 较大时，每一次新加入节点和连边，对于度的分布几乎没有影响的假设。

在某一时间步，当网络中的节点为 N 并且 N 相对于 m_0 较大时，网络增长处于稳态，记度为 k 的节点在整个网络中所占的比率为 $p(k)$，则度为 k 的节点数为 $Np(k)$，其中度分布满足加归一化条件 $\sum_k p(k)=1$。下一个时间步，加入一个新的节点后，度为 k 的节点数为 $(N+1)p(k)$。参考式(1)对于 $\pi(k)=\alpha k+\beta$ 的线性择优连边模式，其中k为节点的度。每个新加入的节点度为 m，用 $m\pi(k)p(k)$ 表示度为 k 的节点与新加入的节点连边的数目，因此有 $\sum_k m\pi(k)p(k)=m$。根据每一次新加入节点和连边对于度的分布几乎没有影响的假设，由率方程可得：

$$(N+1)p(k)=Np(k)+m\pi(k-1)p(k-1)-m\pi(k)p(k)+\delta_{km} \tag{2}$$

式中，$m\pi(k-1)p(k-1)$ 和 $m\pi(k)p(k)$ 分别表示加入新的节点后，新连边产生的由原来度为 $k-1$ 的节点变成度为 k 的节点和原来度为 k 的节点变成度为 $k+1$ 的节点的数目。当 $k>m$ 时，式(2)变成：

$$p(k)(1+m\pi(k))=m\pi(k-1)p(k-1) \tag{3}$$

线性择优连边模式 $\pi(k)=\alpha k+\beta$ 中的参数 $\alpha\geqslant 0$ 和 $-1\leqslant\beta\leqslant 1$ 需满足归一化条件 $\sum_k \pi(k)p(k)=1$，即 α 和 β 两个参数中只有一个是独立的，即：

$$\sum_k \pi(k)p(k)=\sum_k(\alpha k+\beta)p(k)=\alpha<k>+\beta=2m\alpha+\beta=1 \tag{4}$$

解得：

$$\alpha=\frac{1-\beta}{2m} \tag{5}$$

将式(5)代入式(3)和式(2)，得到：

$$p(k)\left[k+\frac{2(1+m\beta)}{(1-\beta)}\right]=\left[k+\frac{2m\beta}{(1-\beta)}-1\right]p(k-1) \tag{6}$$

令：

$$\begin{cases} a=\dfrac{2m\beta}{(1-\beta)}-1 \\ b=\dfrac{2(1+m\beta)}{(1-\beta)} \end{cases} \tag{7}$$

得到度分布满足的递推关系为：

$$p(k)=\frac{k+a}{k+b}p(k-1) \tag{8}$$

由于 a 和 b 不是整数，所以对式(8)一般不能写出简单形式的解析表达式。但当 $m\beta$ 较小时，a 和 b 都是有限大小的数，当 k 较大时，可以得到近似的解析解，将式(8)两边取对数：

$$\lg\frac{p(k)}{p(k-1)}=\lg\frac{k+a}{k+b} \tag{9}$$

式(9)即可用Mandelbrot分布律 $p(k)\propto(k+c)^{-\gamma}$ 作很好的近似，此时将横坐标改为 $\lg(k+c)$，在新的双对数坐标下，幂律指数 γ 为：

$$\gamma=\frac{\lg\dfrac{k+a}{k+b}}{\lg\dfrac{k-1+c}{k+c}} \tag{10}$$

变形得到：

$$\lg\frac{1+a\dfrac{1}{k}}{1+b\dfrac{1}{k}}=\gamma\lg\frac{1+(c-1)\dfrac{1}{k}}{1+c\dfrac{1}{k}} \tag{11}$$

将式(11)两边按 $\frac{1}{k}$ 进行泰勒展开，当展开到2阶项时，可得到幂律指数为：

$$\gamma=b-a=1+\frac{2}{1-\beta} \tag{12}$$

漂移量为：

$$c=\frac{b+a+1}{2}=\frac{1+2m\beta}{1-\beta} \tag{13}$$

由式(12)和式(13)可知，幂律指数 γ 虽然只取决于独立变量 β，漂移量 c 与 β 和 m 都有关系。当 $m\beta$ 较大或 $\beta\to1$ 时，a 和 b 都是很大的数，式(11)的两边不能按 $\frac{1}{k}$ 展开，网络的度分布趋于随机网络的指数分布。当取 $\beta=0$ 时，$\alpha=\frac{1}{2m}$ 是与度成正比的择优连接。可计算得到 $a=-1$，$b=2$，$\gamma=3$，$c=1$，网络是标准的BA网络，Mandelbrot分布律近似可化为 $p(k)=-\frac{2}{\psi(2,m+1)}(k+1)^{-3}$，其中 $\psi(x)$ 为digamma function，$\psi(n,x)=\frac{d^n\psi(x)}{dx^n}$。

对于式(8)，如采用传统的差分方法求解，会对漂移量的计算带来很大误差，这是因为若采用向后差分方法，由于：

$$\frac{\mathrm{d}p}{\mathrm{d}k}=p(k)-p(k-1) \tag{14}$$

其中，$p(k)=\frac{k+a}{k+b}\left[p(k)-\frac{\mathrm{d}p}{\mathrm{d}k}\right]$ 移项得到：

$$\frac{\mathrm{d}p}{\mathrm{d}k}=(1-\frac{k+b}{k+a})p(k)=\frac{a-b}{k+a}p(k)$$

又得到：

$$p(k)\propto(k+a)^{-(b-a)} \tag{15}$$

若采用向前差分方法，由于：

$$\frac{\mathrm{d}p}{\mathrm{d}k}=p(k+1)-p(k) \tag{16}$$

首先将式(8)化为：

$$p(k+1)=\frac{k+1+a}{k+1+b}p(k) \tag{17}$$

再将式(17)代入式(16)得到：

$$p(k)+\frac{dp}{dk}=\frac{k+1+a}{k+1+b}p(k)$$

移项得到：

$$\frac{\mathrm{d}p}{\mathrm{d}k}=\left(\frac{k+1+a}{k+1+b}-1\right)p(k)=\frac{a-b}{k+1+b}p(k)$$

又可得到：

$$p(k)\propto(k+b+1)^{-(b-a)} \tag{18}$$

与式(13)比较，在 k 较小时，显然式(15)和式(18)会带来很大误差。

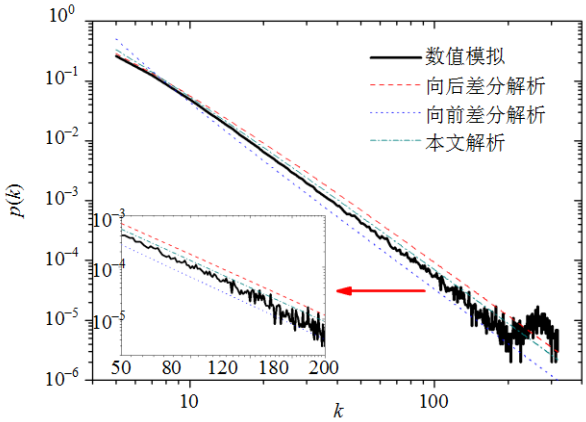

图1　线性择优演化网络度分布的数值模拟与解析结果比较

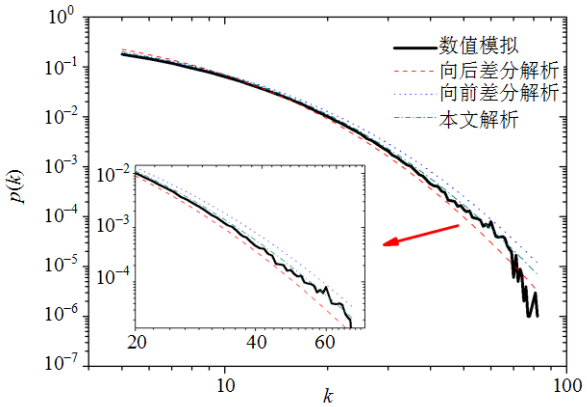

图2　线性择优演化网络度分布的数值模拟与解析结果比较

3　数值模拟

通过数值模拟，比较了本文的解析结果式(13)与向后差分式(15)和向前差分式(18)的漂移量解析结果。如图1和图2所示，本文的结果能够更好刻画线性择优演化网络的度分布。图1中粗黑实线是数值

模拟结果，虚线、点线和点划线分别代表向后差分、向前差分和本文的解析结果。小图是度在50～200一段的放大图。网络参数为 $m_0=11$， $m=5$， $N=20\ 000$， $\beta=0$， $\alpha=\frac{1}{2m}=0.1$；数值模拟结果是100次独立试验的平均值。

图2中粗黑实线是数值模拟结果，虚线、点线和点划线分别代表向后差分、向前差分和本文的解析结果。小图是度在50～200一段的放大图。网络参数 $m_0=11$，$m=5$，$N=20\ 000$，$\beta=0.2$，$\alpha=\frac{1-\beta}{2m}=0.08$；数值模拟结果是100次独立试验的平均值。

特别地，由上面的推导可知，一般网络的度分布均遵从“漂移幂律”，根据式(13)，要求漂移量 $c=0$，此时，$\beta=-\frac{1}{2m}$，$\alpha=\frac{2m+1}{4m^2}$，$a=-\frac{4m+1}{2m+1}$，$b=\frac{2m}{2m+1}$，$\gamma=3-\frac{2}{2m+1}$，实现这样的网络只要将BA网络中的正比择优连接，修正为如下的线性择优连接：

$$\prod(k_i)=\frac{\left(m+\frac{1}{2}\right)k_i-m}{m\sum_i k_i}=\frac{\left(m+\frac{1}{2}\right)k_i-m}{2m^2N}$$

即：

$$\pi(k)=\alpha k+\beta=\frac{\left(m+\frac{1}{2}\right)k-m}{2m^2} \tag{19}$$

显然：

$$\sum_i\prod(k_i)=\frac{m\sum_i k_i+\frac{1}{2}2mN-mN}{m\sum_i k_i}=1 \tag{20}$$

率方程为：

$$(N+1)p(k)=\frac{(m+1/2)(k-1)-m}{2mN}Np(k-1)+Np(k)\left(1-\frac{(m+1/2)k-m}{2mN}\right) \tag{21}$$

即：

$$p(k)=\frac{(m+1/2)(k-1)-m}{2mN}Np(k-1)-\frac{(m+1/2)k-m}{2mN}Np(k)$$

得到：

$$p(k)=\frac{k-\frac{4m+1}{2m+1}}{k+\frac{2m}{2m+1}}p(k-1) \tag{22}$$

采用与前文相同的方法，两边取对数后展开近似，可以解得度分布为：

$$p(k)\propto k^{-3+\frac{2}{2m+1}} \tag{23}$$

对应的漂移量 $c=0$，可称为无偏差的无标度网络。

图3给出了对漂移量为0和漂移量为45的两个模拟结果的对比，数值模拟结果是100次独立试验的平均值。可以看出，具有较大漂移量的网络，存在低头现象，且尾部下降很快，更接近个指数函数的形式。

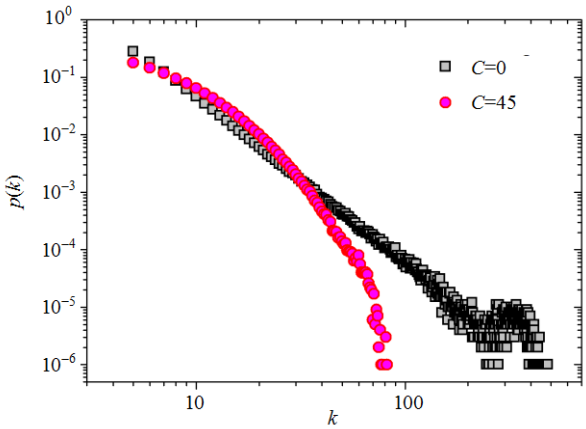

图3 漂移量对网络度分布的影响

在线性择优模式下，如果新加入的节点的度为 m，则：

$$\pi(k)=\alpha k+\beta=\frac{1-\beta}{2m}k+\beta$$

本文严格证明了度分布的递推关系满足 $p(k)=\frac{k+a}{k+b}p(k-1)$；其中 $a=\frac{2m\beta}{(1-\beta)}-1$；$b=\frac{2(1+m\beta)}{(1-\beta)}$。当 $m\beta$ 较小时，网络的度分布可以表示为漂移幂律分布，$p(k)\sim(k+c)^{-\gamma}$，其中，漂移幂律指数 $\gamma=1+\frac{2}{1-\beta}$，漂移量 $c=\frac{b+a+1}{2}=\frac{1+2m\beta}{1-\beta}$。当 $m\beta$ 较大或 $\beta\to1$ 时，网络趋于随机网络。

取特殊线性择优模式 $\pi(k)=\frac{\left(m+\frac{1}{2}\right)k-m}{2m^2}$ 时，所得到的网络是没有偏差的无标度网络 $p(k)\propto k^{-3+\frac{2}{2m+1}}$。

4 结 论

本文指出了传统解析中差分近似中存在的对漂

移量估计的偏差，通过数值对比实验，验证了所提出方法的准确性；揭示了率方程解析线性择优演化网络中对漂移量估计存在的问题；理清了解析近似中存在的错误认识；提出了更精确的近似方法。本文对深入理解并正确运用率方程有重要的借鉴意义。

参 考 文 献

[1] STROGATZ S H. Exploring complex networks[J]. Nature, 2001, 410: 268-276.

[2] DOROGOVTSEV S N, MENDES J F F. Evolution of networks[J]. Adv Phys, 2002, 51: 1079-1187.

[3] ALBERT R, BARABASI A L. Statistical mechanics of complex networks[J]. Rev Mod Phys, 2002, 74: 47.

[4] NEWMAN M E J. The structure and function of complex networks[J]. SIAM Review, 2003, 167: 45.

[5] 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): 18-46.
WU Jin-shan, DI Zeng-ru. Complex networks in statistical physics. Progress in Physics[J]. 物理学进展, 2004, 24(1): 18-46.

[6] 周涛, 柏文洁, 汪秉宏, 等. 复杂网络研究概述[J]. 物理, 2005, 34(1): 31-36.
ZHOU Tao, BAI Wen-Jie, WANG Bing-Hong, et al. A brief review of complex networks[J]. Physics, 2005, 34(1): 31-36.

[7] 吕琳媛, 陆君安, 张子柯, 等. 复杂网络观察[J]. 复杂系统与复杂性科学, 2010, 7(2): 173-186.
LÜ Lin-yuan, LU Jun-an, ZHANG Zi-ke, et al. Looking into complex networks[J]. Complex Systems and Complexity Science, 2010, 7(2): 173-186.

[8] WATTS D J, STROGATZ S H. Collective dynamics of small-world networks[J]. Nature, 1998, 393: 440.

[9] BARABáSI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286: 509.

[10] BARABáSI A L, ALBERT R, JEONG H. Mean-field theory for scale-free random networks[J]. Physica A, 1999, 272: 173.

[11] NEWMAN M E J, The structure of scientific collaboration networks[J]. PNAS, 2001, 98: 404.

[12] SHANG M S, LÜ L, ZHANG Y C, et al. Empirical analysis of web-based user-object bipartite networks[J]. EPL, 2010, 90: 48006.

[13] 刘宏鲲, 周涛. 中国城市航空网络的实证研究与分析[J]. 物理学报, 2007, 56(1): 106-112.
LIU Hong-Kun, ZHOU Tao. Empirical study of Chinese city airline network[J]. Acta Physics Sinica, 2007, 56(1): 106-112.

[14] MANDELBROT B. Information theory and psycholinguistics[M]. NewYork: Basic Books Publishing Co, 1965.

[15] 何大韧, 刘宗华, 汪秉宏. 复杂系统与复杂网络[M]. 北京: 高等教育出版社, 2009.
HE Da-ren, LIU Zong-hua, WANG Bing-hong. Complex systems and complex networks[M]. Beijing: Higher Education Press, 2009.

[16] CHANG H, SU B B, ZHOU Y P, et al. Assortativity and act degree distribution of some collaboration networks[J]. Physica A, 2007, 383: 687.

[17] WANG Y L, ZHOU T, SHI J J, Empirical analysis of dependence between stations in Chinese railway network[J]. Physica A, 2009, 388: 2949.

[18] 史定华. 复杂网络的随机刻画和演化规律[J]. 力学进展, 2008, 38(6): 679.
SHI Ding-hua. Stochastic features of complex networks and their patterns in dynamic evolution[J]. Advances in Mechanics, 2008, 38(6): 679.

[19] 何大韧, 刘宗华, 汪秉宏. 复杂网络研究的一些统计物理学方法及其背景[J]. 力学进展, 2008, 38(6): 692.
HE Da-ren, LIU Zong-hua, WANG Bing-hong. Some statistical physics methods for investigations of complex networks and their background[J]. Advances in Mechanics, 2008, 38(6): 692.

[20] 史定华. 无标度网络: 基础理论和应用研究[J]. 电子科技大学学报, 2010, 39(5): 644-650.
SHI Ding-hua. Scale-free networks: Basic theory and applied research[J]. Journal of University of Electronic Science and Technology of China, 2010, 39(5): 644-650.

[21] KRAPIVAKY P L, REDNER S, LEYVRAZ F. Connectivity of growing random networks[J]. Phys Rev Lett, 2000, 85: 4629.

[22] ZHOU T, YAN G, WANG B H. Maximal planar networks with large clustering coefficient and power-law degree distribution[J]. Phys Rev E, 2005, 71: 046141.

[23] MOORE C, GHOSHAL G, NEWMAN M E J. Exact solutions for models of evolving networks with addition and deletion of nodes[J]. Phys Rev E, 2006, 74: 036121.

编 辑 蒋 晓