

一类作为网络模型的可平面无标度图

马 飞, 姚 兵*

(西北师范大学 数学与统计学院, 甘肃 兰州 730070)

摘要: 无标度特性普遍存在于大量的实际网络和人造网络中. 为了更好地研究这类无标度网络模型的拓扑性质和内在动力学, 大量的模型被建立, 如随机网络模型和确定性网络模型. 鉴于以往确定性模型中的无标度指数都是唯一不变的常数, 定义了一类具有广义自相似性的增长网络模型, 分析了它的一些拓扑性质: 平均度、聚集系数、直径、度分布、最多叶子生成树. 得出该模型具有无标度特性和小世界效应, 并且可以通过调整相应的参数来获得丰富的无标度指数.

关键词: 无标度图; 网络模型; 小世界效应

中图分类号: O157.5

文献标识码: A

doi: 10.7511/dllgxb201704016

0 引言

为了认识自然、改造自然, 学者们在面对大到天体, 小到人类及其他动物界, 甚至微生物界之间存在的种种联系时, 形象生动地提出了复杂网络这一概念, 随之产生了大量的复杂网络模型. Watts 等^[1]提出了小世界网络模型, Barabási 等^[2]首次提出了无标度网络模型. 在此之后的十几年, 学术界对复杂网络的研究热潮不衰反增, 吸引了众多学者的不断涌入, 使得复杂网络的研究不论纵向深入还是横向拓展都取得了丰硕的成果. 现今复杂网络研究涵盖了诸多学科领域, 如物理、数学、计算机科学、化学、生物学、经济学等, 在研究各领域中不同的研究对象时, 产生了大量的网络模型, 如物联网、地震网、食物链网、电力网、细胞新陈代谢网、神经信号传送网、股市等^[3-10].

对真实网络进行了海量的研究之后, 学者们不约而同地阐释了众多网络具有的拓扑共性: (1) 小的直径和平均距离; (2) 高的聚集系数; (3) 服从幂率分布(无标度性); (4) 自相似性; (5) 层次结构. 从而为复杂网络的进一步研究指明了方向. 为了认识、理解复杂网络, 学者们通过循环迭代的方法

构造新的随机模型, 讨论它的拓扑性质.

本文模型的建立亦采用当前比较热门的随时间循环迭代的生长方法. 不同于先前研究的是, 网络的增长是通过实施平行加边运算进行的, 并且引入另一个增长参数 d (时间 t 是一个增长参数), 这样使得文中的模型更具有一般性, 拓扑结构更加丰富多样. 此外, 新建立的模型是可平面的. 模型的可平面性有很多方面的应用与研究, 平面结构不仅是印刷电路中的结构, 它在数学中也有较多的研究, 如著名的平面图的四色问题. 在文献[11]中提到, 所有的无标度网络都是稀疏的, 并且在幂率分布式 $P_{\text{cum}}(k) \sim k^{-\gamma}$ 中 $\gamma \notin [0, 2]$, 然而本文中的模型 $M(t)$ 具有无标度特性, 但 $\gamma = \frac{\ln 3d}{\ln 2d} \in [0, 2]$, Barabási 等^[2]同样类似地认为无标度网络的幂率指数 $\gamma \in [2, 3]$, 本文认为可以拓展无标度的数学定义.

1 模型的迭代及其拓扑性质

1.1 模型的迭代

在本文的网络模型 $M(t)$ 中, 用顶点刻画研究中的每个实体对象, 顶点间的连边表示实体对象

间的相互联系, $L(t)$ 表示 $M(t)$ 的最多叶子生成树的叶子数, 文中未定义的术语和符号均采用图论中标准的术语和符号^[12]. 模型 $M(t)$ 可以通过迭代的方法得到, 如下:

(1) 当 $t=0$ 时, 模型 $M(0)$ 由一条活动边和与之关联的两个顶点组成.

(2) 当 $t>0$ 时, 模型 $M(t)$ 可以通过对模型 $M(t-1)$ 中的活动边实施平行加边运算获得. 这里的活动边是指模型 $M(t-1)$ 在 $t-1$ 时刻新产生的边, 每条活动边至少含有 1 个二度顶点, 不妨称在 $t-1$ 时刻之前的边为休眠边.

为了便于叙述, 本文采用符号 $V_g(t)$ 、 $E_g(t)$ 分别表示模型 $M(t)$ 在 t 时刻所产生的新顶点集合和活动边集合. 类似地, 采用符号 $|V_g(t)|$ 与 $|E_g(t)|$ 分别表示模型 $M(t)$ 在 t 时刻所具有的新顶点集中顶点的数目和活动边的数目, 符号 $|V(t)|$ 与 $|E(t)|$ 分别表示模型 $M(t)$ 的总顶点数目和总边数目. 显而易见, 每一步迭代过程中, 给每条活动边实施平行加 d 条边, 然后将这 d 条边的左右顶点分别与该活动边的左右顶点相连, 连接完毕后该活动边进入休眠. 总共涌现出 $3d$ 条

新的活动边, 即 $|E_g(t+1)| = 3d |E_g(t)|$. 根据 $|E_g(0)| = 1$, 有 $|E_g(t)| = (3d)^t$. 每迭代一步, 将有 $2d$ 个新的顶点与一条活动边对应, 即 $|V_g(t+1)| = 2d |E_g(t)| = 2d \cdot (3d)^t$, 其中 $|V_g(0)| = 2$. 易得模型 $M(t)$ 在 t 时刻所具有的总顶点和总边的数目分别如下:

$$|V(t)| = \sum_{i=0}^t |V_g(i)| = 2 + 2d \frac{(3d)^t - 1}{3d - 1} = \frac{2d \cdot (3d)^t + 4d - 2}{3d - 1} \quad (1)$$

$$|E(t)| = \sum_{i=0}^t |E_g(i)| = 1 + 3d \frac{(3d)^t - 1}{3d - 1} = \frac{(3d)^{t+1} - 1}{3d - 1} \quad (2)$$

通过上述迭代构造出的模型 $M(t)$ 是可平面的, 如果加边参数 d 同时满足 $d = m + n$, m 表示一条活动边上方加边的条数, n 表示一条活动边下方加边的条数, 这样便可以得到一族可平面的广义自相似网络模型. 随着时间的推移, 这一族网络模型将是相当复杂的^[8]. 图 1 为模型 $M(t)$ 在 $t=0, 1, 2, d=m+n=2(m=n=1)$ 的拓扑结构.

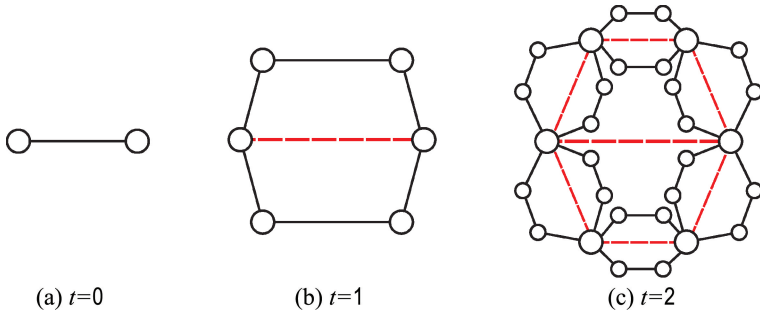


图 1 模型 $M(t)$ 拓扑结构

Fig. 1 Model $M(t)$ topological structures

1.2 模型的拓扑性质

1.2.1 平均度 网络的平均度 $\langle k \rangle$ 为边个数的 2 倍与顶点个数之比, 即

$$\langle k \rangle = \frac{2|E|}{|V|} = \frac{(3d)^{t+1} - 1}{d \cdot (3d)^t + 2d - 1} \quad (3)$$

随着时间的推移, 模型 $M(t)$ 的顶点和边数目越来越大, 不难发现, 当 $t \rightarrow \infty$ 时, 整个模型 $M(t)$ 的平均度趋于一固定常数 3, 即

$$\langle k \rangle = \frac{2|E|}{|V|} \rightarrow 3; t \rightarrow \infty \quad (4)$$

位于活动边 xy 上下不同位置的新顶点和新边会影响以后进入网络的新顶点和新边的位置, 从而导致模型 $M(t)$ 的拓扑结构的复杂性(图 2).

1.2.2 聚集系数 一般复杂网络中顶点 i 的聚集系数定义为 $\langle c_i \rangle = \frac{n_i}{k_i(k_i - 1)}$, 其中 k_i 表示与顶点 i 相邻的顶点的个数, n_i 表示这些相邻的 k_i 个顶点之间实际存在的边数. 在构造模型 $M(t)$ 的过程中, 与任何一个顶点相邻的顶点集合中任

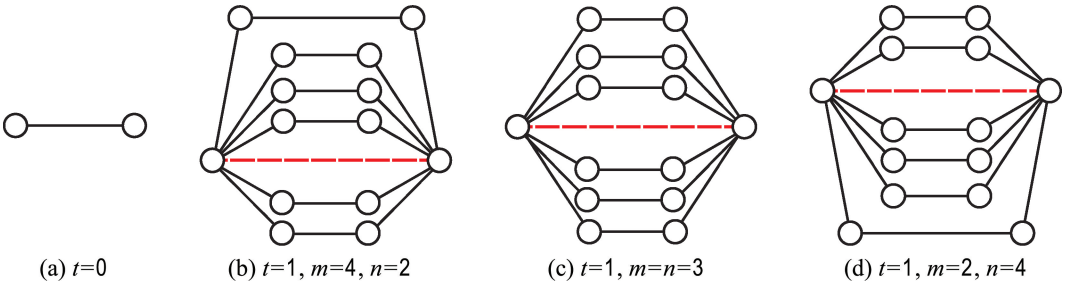


图 2 模型 $M(t)$ 在 $t=0,1,d=m+n=6$ 的拓扑结构

Fig.2 Model $M(t)$ topological structures at $t=0,1,d=m+n=6$

意两顶点与该顶点不能构成一个三角形,所以 $n_i=0$,故 $\langle c_i \rangle=0$. 对整个模型而言,它的聚集系数

$$\langle c \rangle = \frac{1}{|V(t)|} \sum_{i=1}^{|V(t)|} \langle c_i \rangle = 0.$$

1.2.3 直径 直径作为研究复杂网络的另一个重要参数,往往对网络的随机游走研究起到一定的作用,然而很多复杂网络的直径并不是轻松获得的,一般来说,获得一些网络直径的精确解是相当困难的. 模型 $M(t)$ 基于对称迭代构造,便于获得直径的精确解. 分析得知,模型 $M(t)$ 中任意两顶点间的距离中只有在 t 时刻新加入的对称顶点间的距离最长,根据直径的定义 $D(t)=\max\{d_{ij} \mid i,j \in V(t)\}$,其中 d_{ij} 表示顶点 i 与顶点 j 的距离,进一步分析得知, $M(t)$ 中的直径 $D(t)$ 可以表示为:顶点 i 先走到与之相邻的在 t 时刻休眠边的顶点 i_t ,顶点 j 先走到与之相邻的在 t 时刻休眠边的顶点 j_t ,直径 $D(t)=1+d_{i_t j_t}+1$. 根据模型的对称性知,顶点 i_t 与顶点 j_t 的距离 $d_{i_t j_t}$ 是模型 $M(t-1)$ 的直径,故 $D(t)=D(t-1)+2, t \geq 2$,因 $D(1)=1, M(t)$ 的直径为

$$D(t)=3+2(t-1) \tag{5}$$

从式(5)可以看出,当 t 很大时, $t \sim \ln |V(t)| \approx \ln(3d)^t = t \cdot \ln 3d$,也就是 $D(t) \sim \ln |V(t)|$. 说明模型 $M(t)$ 是稀疏的.

1.2.4 度分布 度分布是一个刻画复杂网络是否具有无标度性质的拓扑参数. 对模型 $M(t)$ 中顶点的度构成的度谱分析可知,这是一些离散的数据,为了获得模型 $M(t)$ 的度分布,采取累积度分布的计算方法^[4]:

$$P_{\text{cum}}(k) = \frac{\sum_{k' \geq k} |V(t, k')|}{|V(t)|} \tag{6}$$

其中 $V(t, k')$ 表示在 t 时刻 $M(t)$ 中 degree 为 k' 的顶点集合,下面做详细的论证. 在 $t=0$ 时, $M(0)$ 有两个度为 1 的顶点,之后的每一次迭代中新加入的顶点的度都是 2. 为了简化叙述,采用符号 $k(i, t), k_g(i, t), k_p(i, t)$ 表示在 $t \geq t_i$ 时与顶点 i 相关联的边的数目、活动边的数目、休眠边的数目. 显然有 $k(i, t) = k_g(i, t) + k_p(i, t)$. 鉴于模型 $M(t)$ 的构造,可以得到:当 $i=0$ 时,模型 $M(t)$ 中的两个属于 $M(0)$ 的初始顶点的度为

$$k(i, t) = \frac{d^{t+1}-1}{d-1} \tag{7}$$

在 $t_i \geq 1$ 时,

$$\begin{aligned} k_p(i, t+1) &= k_g(i, t) + k_p(i, t), \quad k_g(i, t_i) = 2 \\ k_g(i, t+1) &= 2dk_g(i, t), \quad k_p(i, t_i) = 0 \end{aligned} \tag{8}$$

由式(8)计算得

$$\begin{aligned} k_g(i, t) &= 2 \cdot (2d)^{t-t_i} \\ k_p(i, t) &= 2 \frac{(2d)^{t-t_i} - 2d}{2d-1} \end{aligned} \tag{9}$$

注意到在 t_i 时刻,有 $|V_g(t_i)| = 2d \cdot (3d)^{t_i-1}$ 个顶点进入网络,这些顶点在 t 时刻有着相同的度. 将式(9)代入 $k(i, t) = k_g(i, t) + k_p(i, t)$,得

$$k(i, t) = k_g(i, t) + k_p(i, t) = \frac{4d \cdot (2d)^{t-t_i} - 4d}{2d-1} \tag{10}$$

在 t_i 时刻,考察度为 k 的顶点的度分布,由式(10)可得

$$t-t_i = \frac{\ln \frac{k(2d-1)+4d}{4d}}{\ln 2d} \tag{11}$$

联立式(6)、(7)、(11),得到网络模型 $M(t)$ 的累积度分布为

$$P_{\text{cum}}(k) = \frac{1}{|V(t)|} \sum_{k' \geq k} |V(t, k')| =$$

$$\frac{1}{\frac{2d \cdot (3d)^t + 4d - 2}{3d - 1}} \times$$
$$\left[2 + \sum_{t_i=1}^{t_i} 2d \cdot (3d)^{t_i} \right] =$$
$$\frac{2d \cdot (3d)^{t_i} + 4d - 2}{2d \cdot (3d)^t + 4d - 2} =$$
$$\frac{2 \cdot (3d)^{\frac{\ln \frac{k(2d-1)+4d}{4d}}{\ln 2d} + 1} + 12d - 6}{2 \cdot (3d)^{t+1} + 12d - 6} \quad (12)$$

当 t 很大时,由式(11)知,式(12)可简化为

$$P_{\text{cum}}(k) \approx (3d)^{-\frac{\ln \frac{k(2d-1)+4d}{4d}}{\ln 2d}} = (3d)^{-\frac{\ln \left(\frac{k}{2} - \frac{k}{4d} + 1 \right)}{\ln 2d}} =$$
$$\left(\frac{k}{2} - \frac{k}{4d} + 1 \right)^{-\frac{\ln 3d}{\ln 2d}} =$$
$$k^{\frac{\ln 3d}{\ln 2d} \left(\frac{1}{2} - \frac{1}{4d} + \frac{1}{k} \right)^{-\frac{\ln 3d}{\ln 2d}}} \quad (13)$$

当 $k \gg 1$ 时,将获得如下表达式:

$$P_{\text{cum}}(k) \approx k^{-\frac{\ln 3d}{\ln 2d} \left(\frac{1}{2} - \frac{1}{4d} \right)^{-\frac{\ln 3d}{\ln 2d}}} \quad (14)$$

即度分布满足幂率分布 $P_{\text{cum}}(k) \approx k^{-\gamma}$, 其中 $\gamma = \frac{\ln 3d}{\ln 2d}$. 因此,模型 $M(t)$ 是一个无标度网络模型. 在文献[11]中提到,所有的无标度网络都是稀疏的,并且在幂率分布式 $P_{\text{cum}}(k) \sim k^{-\gamma}$ 中 $\gamma \notin [0, 2]$, 然而本文的模型 $M(t)$ 具有无标度特性,但 $\gamma = \frac{\ln 3d}{\ln 2d} \in [0, 2]$. 注意到,文献[8]也得到类似的结论,对无标度网络的认识又深入了一步.

1.2.5 最多叶子生成树 网络的生成树作为一个研究网络的工具目前正越来越受到诸多学者的青睐,很多文献对此做了不同程度的分析,提出了一个研究网络的重要参数——熵 $\left(h = \lim_{t \rightarrow \infty} \frac{\ln s(t)}{V(t)} \right)$, 其中 $s(t)$ 表示给定网络的生成树的总数, $V(t)$ 表示给定网络的顶点数). 熵刻画了给定网络的同步性和抗干扰能力,很受网络安全学者们的关注. 最多叶子生成树是一类重要的生成树,对研究网络的最优问题及随机游走问题起到至关重要的作用. 鉴于构造模型 $M(t)$ 的迭代方法,获得了 $M(t)$ 的最多叶子生成树的叶子数目 $L(t)$ 满足一个递推公式,具体分析如下: 在 $t = 0$ 时, $M(0)$ 中 $L(0) = 2$; 在 $t = 1$ 时, $M(1)$ 的最多叶子生成树可

通过删掉 $M(1)$ 中的新加入的 d 条活动边,得 $L(1) = 2d$; 在 $t = 2$ 时, $M(2)$ 的最多叶子生成树可通过依次删掉 $M(2)$ 和 $M(1)$ 中新加入的 $\frac{(3d)^2}{3}$ 和 d 条活动边,得 $L(2) = 6d^2 = 2d \cdot (3d)^{2-1}$; 在 $t = 3$ 时, $M(3)$ 的最多叶子生成树可通过依次删掉 $M(3)$ 、 $M(2)$ 和 $M(1)$ 中的新加入的 $\frac{(3d)^3}{3}$ 、 $\frac{(3d)^2}{3}$ 和 d 条活动边,得 $L(3) = 18d^3 = 2d \cdot (3d)^{3-1}$. 运用数学归纳法,假设在 $t - 1$ 时刻, $\frac{L(t-1)}{L(t-2)} = 3d$ 成立,那么在 t 时刻, $M(t)$ 的最多叶子生成树可通过依次删掉 $M(t)$ 、 $M(t-1)$ 、 $M(t-2)$ 、 $\dots$ 、 $M(1)$ 中的新加入的 $\frac{(3d)^t}{3}$ 、 $\frac{(3d)^{t-1}}{3}$ 、 $\frac{(3d)^{t-2}}{3}$ 、 $\dots$ 、 d 条活动边得到. 这样删边得到所有的叶子就是在 t 时刻进入网络的顶点,即

$$L(t) = |V_g(t)| = 2d \cdot (3d)^{t-1} = 3d \cdot L(t-1) \quad (15)$$

2 结 语

理论分析表明,本文建立的模型具有无标度特性和小世界效应. 在最近的生物网络研究中,有学者发现了一类特殊的现象:网络中每个顶点的邻居顶点集中任何顶点彼此之间没有边相连(聚集系数 $\langle c \rangle = 0$)^[8,13], 而且这样的网络具有较强的鲁棒性. 在实际生活中,可以提高对蓄意攻击的预防性,加强网络的安全性. 本文的模型正好吻合这一现象,有力推动对该类特殊网络的进一步研究,这也提醒人们在今后的研究中更加关注此类复杂网络. 复杂网络的研究正在如火如荼地进行着,研究成果层出不穷,研究方法也在日趋完善,日后将关注网络研究中的熵与随机游走等问题^[8,10].

参考文献:

[1] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks [J]. *Nature*, 1998, **393**(6684):440-442.

[2] BARABÁSI A L, ALBERT R, JEONG H. Mean-field theory for scale-free random networks [J]. *Physica A: Statistical Mechanics and its Applications*, 1999, **272**(1):173-187.

- [3] WATTS D J. **Small Worlds; The Dynamics of Networks between Order and Randomness** [M]. New Jersey: Princeton University Press, 1999.
- [4] ZHANG Zhongzhi, ZHOU Shuigeng, FANG Lujun, *et al.* Maximal planar scale-free Sierpinski networks with small-world effect and power-law strength-degree correlation [J]. **EPL**, 2007, **79**(3):417-429.
- [5] YAO Bing, YAO Ming, CHEN Xiangen, *et al.* Research on edge-growing models related with scale-free small-world networks [J]. **Applied Mechanics and Materials**, 2014 (513-517): 2444-2448.
- [6] ZHANG Z Z, WU B, COMELLAS F. The number of spanning trees in Apollonian networks [J]. **Discrete Applied Mathematics**, 2014, **169**:206-213.
- [7] TEUFL E, WAGNER S. Resistance scaling and the number of spanning trees in self-similar lattices [J]. **Journal of Statistical Physics**, 2011, **142**(4):879-897.
- [8] MIRALLES A, COMELLAS F, CHEN L C, *et al.* Planar unclustered scale-free graphs as models for technological and biological networks [J]. **Physica A: Statistical Mechanics and its Applications**, 2010, **389**(9):1955-1964.
- [9] SONG C M, KOREN T, WANG P, *et al.* Modelling the scaling properties of human mobility [J]. **Nature Physics**, 2010, **6**(10):818-823.
- [10] YAN G, TSEKENIS G, BARZEL B, *et al.* Spectrum of controlling and observing complex networks [J]. **Nature Physics**, 2015, **11**(9):779-786.
- [11] DEL GENIO C I, GROSS T, BASSLER K E. All scale-free networks are sparse [J]. **Physical Review Letters**, 2011, **107**(17):178701.
- [12] BONDY J A, MURTY U S R. **Graph Theory with Applications** [M]. London: The Macmillan Press Ltd. , 1976.
- [13] COMELLAS F, ZHANG Zhongzhi, CHEN Lichao. Self-similar non-clustered planar graphs as models for complex networks [J]. **Journal of Physics A Mathematical & Theoretical**, 2009, **42**(4):461-471.

A planar scale-free graphs as network models

MA Fei, YAO Bing*

(College of Mathematics and Statistics, Northwest Normal University, Lanzhou 730070, China)

Abstract: The scale-free feature is popular in amounts of real-life and artificial complex networks. In order to study the topological properties and intrinsic dynamics of this kind of scale-free network model, lots of models are presented, such as random network models and deterministic ones. Considering that the scale-free exponent is an unique constant in those previous deterministic models, a generalized self-similarity growing network model is proposed. Its topological properties, including average degree, clustering coefficient, diameter, degree distribution, maximum-leaf-spanning-tree are analyzed. It shows that this model has scale-free characteristics and small-world effects, and the scale-free exponent can be enriched by adjusting corresponding parameters.

Key words: scale-free graph; network model; small-world effect