

◎网络、通信、安全◎

无尺度、小世界 Sierpinski 网络统一模型

吴月文¹,严华云^{1,2}WU Yue-wen¹,YAN Hua-yun^{1,2}

1.同济大学 电子与信息工程学院,上海 201804

2.湖州师范学院 信息工程学院,浙江 湖州 313000

1.College of Electronics & Information Engineering,Tongji University,Shanghai 201804,China

2.School of Information & Engineering,Huzhou Teachers College,Huzhou,Zhejiang 313000,China

E-mail:yuewen1983@hotmail.com

WU Yue-wen,YAN Hua-yun.Unifying model for Sierpinski networks with scale-free small world properties.Computer Engineering and Applications,2009,45(1):99-102.

Abstract: This paper introduces a single-parameter model of evolutionary Sierpinski network (ESN) that integrates the deterministic Sierpinski networks, and the random Sierpinski networks. Some of the properties of the evolutionary models on the basis of the complex network could be applied to the study of networks. This article also presents the form of the degree distribution which results from the algorithm proposed, and emulation consequence of the average path length of ESN reveals the small world feature of the new type of network.

Key words: complex networks; unified model; scale-free; small world

摘要: 引入了单一参数的演化 Sierpinski 网络模型,统一了确定性和随机性两类 Sierpinski 网络模型。基于复杂网络理论提出的演化模型的一些特征能应用于对现实网络的研究。给出了推导度分布的算法和平均路径长度的仿真结果,数值仿真验证了理论分析的正确性,揭示了新网络模型的“小世界”特征。

关键词: 复杂网络;统一模型;无尺度;小世界

DOI:10.3778/j.issn.1002-8331.2009.01.029 文章编号:1002-8331(2009)01-0099-04 文献标识码:A 中图分类号:TP393

1 引言

复杂网络已成为学术界研究的一个热点,其中最主要的原因之一就是复杂网络能够灵活地并且一般化地表示现实世界的网络。研究人员做了大量的实证研究,揭示了许多现实网络具有的普遍性质,如小世界效应,度的幂律分布特性,前者表现为较小的平均最短距离(APL)和高簇系数的特性,后者一般称为无尺度网络。

为了刻画现实世界中的小世界网络,科学家们提出了许多小世界网络模型,其中最著名的是1998年Watts和Strogatz提出的小世界网络模型-WS模型^[1],这一开创性工作的发表掀起了小世界网络、WS模型乃至复杂网络的研究热潮,随后,研究人员又提出了WS模型的替代模型和变体模型。WS模型及其变体模型让人们对于现实系统的形成有了较为直观的感觉,它们从不同的角度揭示了小世界效应的形成机制。

无尺度网络的研究最早起源于1999年的BA模型^[2],它是第一个无尺度网络的演化模型,对整个复杂网络的发展起着不可估量的作用。随着BA模型这一开创性工作的发表,国内外掀起了研究无尺度网络模型的高潮^[3-5],在短短的几年时间内,

涌现出了一大批无尺度网络的新模型,而且这些模型的性质与现实系统越来越接近^[3-5]。然而,各种真实网络的形成机制大不相同,存在着多样性、丰富性与复杂性。现有的模型远远不能囊括所有现实系统的生成机理与方式。目前,许多现实系统尚未在复杂网络的研究涉足到,应用复杂网络的观点对这些系统进行研究,捕捉其生成机制,建立相应的演化模型,并在此基础上研究模型的性质,是具有重大的理论意义和实用价值的。

在现实世界中,有一类系统可以用不相容图(或不相容网络)进行描述,因为在这类系统中存在接触或不相容关系。例如,城市道路系统就可以映射成一个不相容网络^[6],其中的每条道路被映射成网络节点,道路间的相互交叉关系可以定义为节点间的边,这一映射为研究人员分析城市导航的复杂性提供了方便。又如学校的排课系统也可以用不相容图来描述,把网络节点映射为不同的课程,如果两门课程在同一时间授课,则它们之间有一条边相连。另外,不相容网络曾被频繁应用到RNA的折叠研究^[7]以及聚合物物理的研究^[8]之中,并且取得了巨大的成功。遗憾的是,虽然不相容网络是普遍存在的,但是对相关网络模型的研究却很少。

基金项目:国家自然科学基金(the National Natural Science Foundation of China under Grant No.60573056);湖州市科技计划资助项目(No.2007YZ08)。

作者简介:吴月文(1983-),男,研究生,主研领域:复杂网络,演化博弈论;严华云(1972-),男,讲师,博士研究生,主研领域:信息检索、对等网。

收稿日期:2008-07-22 修回日期:2008-11-04

尽管现实世界中的许多系统可以用不相容网络来描述,但这类系统有各自不同的演化模式。为了对这类网络有个一般的了解,需要使用统一的方法来表示这些不相容网络的共性,同时避免单个系统的具体细节。在之前的论文中,将著名的 Sierpinski 分形^[9]映射成一类不相容网络,这类网络包括两个网络:一个是确定性的,另一个是随机的,它们都是极大平面图,同时具有无尺度、小世界、大的簇系数等性质。确定性 Sierpinski 网络^[10]和随机 Sierpinski 网络分别从不同角度较好地描述了现实网络的主要结构特性,但其拓扑结构存在差异,如前者的度指数小于后者。为了研究这些差异存在的原因,根据 Sierpinski 分形垫的构造过程,引入了一个参数 q ,提出了一类演化 Sierpinski 网络模型(ESN),将确定性 Sierpinski 网络和随机 Sierpinski 网络进行统一,并且将其纳入到同一个模型中。ESN 的度指数在区间 $[2, 3]$ 内连续可调,相对于确定性 Sierpinski 网络和随机 Sierpinski 网络而言,ESN 与实际系统的性质更加符合。

2 确定性和随机 Sierpinski 网络简介

为了构造网络,先介绍经典的 Sierpinski 分形垫(也称 Sierpinski 三角形)。Sierpinski 分形垫可以使用迭代的方法进行构造。设 S_t 表示经过 t 步演化后得到的 Sierpinski 分形垫,则其具体的构造过程可以描述如下:初始状态 S_0 是一个等边三角形;然后在第一个迭代步,连接等边三角形各边的中点,从而将原等边三角形分成四个小三角形,接着移走中间的一个小三角形,便得到 S_1 ;在第二个迭代步,对剩下的 3 个小三角形按上面同样的方法继续分割,并舍弃中间的三角形,从而得到 S_2 (如图 1 左图)。经过 t 步迭代后,便得到著名的 Sierpinski 模型 S_t 。

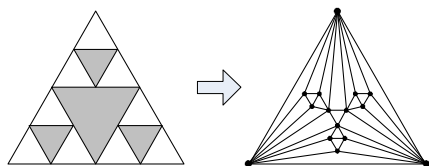


图1 Sierpinski 分形垫的前两步迭代构造(左图),确定性 Sierpinski 不相容网络的前两步构造(右图)

根据 Sierpinski 分形垫,通过如下映射方法可以很容易构造网络:将 Sierpinski 分形垫中每个被删除的三角形的边映射为网络节点,网络的边表示 Sierpinski 分形垫的所有被删除的三角形中边的接触关系。为了保持一致,分形垫中初始等边三角形 S_0 的三条边也分别对应于 3 个不同的节点。这样,便得到了一个网络(如图 1 右图)。由于该网络描述了 Sierpinski 分形垫中被删除三角形边的接触关系,所以称之为确定性 Sierpinski 不相容网络。

在介绍随机 Sierpinski 不相容网络的构造之前,先引入一种随机分形,该分形是确定 Sierpinski 分形垫的一个随机变体,称为随机 Sierpinski 分形垫。随机不相容网络正是基于随机 Sierpinski 分形垫而建立的,其构造方法如图 2 左图。与确定性 Sierpinski 分形垫类似,随机 Sierpinski 垫的初始情形也是一个等边三角形,随后进行如下操作:在第一步,连接初始三角形的中点形成 4 个小三角形,然后移去新生成的向下的小三角形,这样就得到初始三角形的 3 个副本;在此后的每一个时间步,随机选择一个等边三角形,再对所选择等边三角形的边进行等

分,并删除新形成的向下的小三角形,从而得到被选三角形的 3 个副本。重复执行就可以得到随机的 Sierpinski 分形垫。

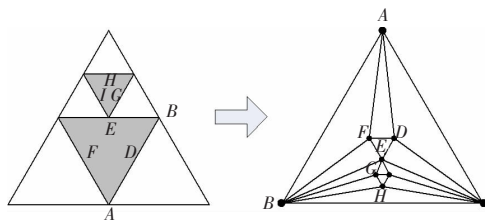


图2 随机 Sierpinski 分形垫的构造(图左),随机 Sierpinski 不相容网络示意图(图右)

与确定性的 Sierpinski 不相容网络所采用的映射方式相同,也可以将随机 Sierpinski 分形垫映射成网络——随机 Sierpinski 不相容网络。将被移除三角形的三条边映射成网络中的 3 个节点,网络的边表示所有被移除的三角形边之间的接触关系。为了保持一致,随机 Sierpinski 分形垫中初始三角形的 3 条边分别映射到 3 个节点(如图 2 右图)。

3 演化 Sierpinski 网络的迭代算法

下面介绍演化 Sierpinski 网络。首先,给出一种新的方法构造演化 Sierpinski 分形垫。初始结构和确定 Sierpinski 分形垫一致,此后每一迭代步,对每个三角形以概率 q 进行填充。特殊情况当 $q=1$ 时,它便弱化为确定的 Sierpinski 分形垫;当 $q \rightarrow 0$ 时,它便符合随机 Sierpinski 分形垫。ESN 正是由演化 Sierpinski 分形垫衍生而来;边映射为节点、边的连接映射为边。图 3 显示了网络增长过程。

在演化 Sierpinski 分形垫的构建过程中,在任意迭代步,对每个未被删除的等边三角形,连接各边中点形成 4 个小三角形,再删除其中间新生成的小三角形,从而得到 3 个原三角形的副本。当创建网络时,这与下列操作等价:对每组新加入的 3 个节点,将得到新生成 3 个小三角形,每个新生成的小三角形在下一迭代步又将生成 3 个新节点。根据以上规则,可以引入网络生成的一般算法。

用 $E(t)$ 表示经过 t 步迭代后得到的网络,则网络的构造算法如下:当 $t=0$ 时, $E(0)$ 为 3 个节点组成的一个三角形。接着在初始三角形中新增加 3 个节点,此新加入的节点两两相互连接构成一个新三角形,同时将此新三角形每条边的端点分别连向初始三角形的一个顶点,这样便得到了 $E(1)$ (如图 3 所示)。当 $t \geq 1$ 时, $E(t)$ 由 $E(t-1)$ 得到。为方便描述,先给出一个定义:即对任一 $E(t-1)$ 中在 $t-1$ 步生成的三角形,如果其三个顶点中仅有一个是 $t-1$ 步生成的节点,则称此三角形为活三角形。将 $E(t-1)$ 的每个活三角形以概率 q 用图 3 右边所示的簇替代,即可得到 $E(t)$ 。重复这一增长过程,一直持续到网络的节点数满足理想的要求为止。当 $q=1$ 时,该网络即为确定性 Sierpinski 网络;当 $q < 1$,此网络随机增长。特殊地,当 $q \rightarrow 0$ 时,该网络弱化为随机 Sierpinski 网络。

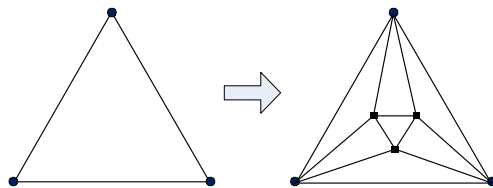


图3 网络的迭代构造方法

接下来计算网络 $E(t)$ 的节点和边数。设 $L_v(t), L_e(t)$ 分别表示第 t 步生成的节点数、边数。根据网络的构造算法(参见图 3), 每增加 3 个新节点就会产生 3 个新的活三角形, 同时增加 9 条边。在步骤 1, 网络中分别增加 $L_v(1)=3q$ 新节点和 $L_e(1)=9q$ 新边。经过简单计算, 可以得到在时间步 $t_i(t_i > 1)$ 新生成的网络节点和边分别为 $L_v(t_i)=3q(1+2q)^{t_i-1}$ 和 $L_e(t_i)=9q(1+2q)^{t_i-1}$ 。因此在时间步 t , 网络总的节点数和边数分别为

$$N_t = 3 + \sum_{i=1}^t L_v(t_i) = 3 + \frac{3}{2}((1+2q)^t - 1) = \frac{3}{2}((1+2q)^t + 1) \quad (1)$$

和

$$E_t = 3 + \sum_{i=1}^t L_e(t_i) = 3 + \frac{9}{2}((1+2q)^t - 1) = \frac{3}{2}(3(1+2q)^t - 1) \quad (2)$$

当 t 取足够大时, 平均度 $\bar{k}_t = \frac{2E_t}{N_t}$ 近似等于 6。

从式(1)和式(2)可以得到 $E_t = 3N_t - 6$ 。根据网络的构造方法, 任意两条边都不相交, 因此演化的 Sierpinski 不相容网络是最大平面图^[1]。

4 网络的拓扑性质

下文对 $E(t)$ 的度分布、簇系数、平均路径长度等拓扑性质进行解析计算。

4.1 度分布

当一个新的节点 i 在 $t_i(t_i \geq 1)$ 步进入网络时, 它的度为 4。设 $L_\Delta(i, t)$ 表示网络在第 t 步产生的部分活三角形数目, 这些活三角形在第 $t+1$ 步将生成新节点与节点 i 相连。显然, 在迭代步 $t_i, L_\Delta(i, t_i) = 1$ 。根据网络的迭代过程易知, 对于 t_i 后的每一个时间步, i 点的每两个新生成的邻居节点产生两个包含节点 i 的新的活三角形, 同时包含 i 的一个已存在的活三角形将失去“活性”而不再是活三角形。定义 $k_i(t)$ 表示节点 i 在时间 t 时的度数, 那么 $k_i(t)$ 和 $L_\Delta(i, t)$ 之间有如下关系:

$$L_\Delta(i, t) = \frac{k_i(t) - 2}{2} \quad (3)$$

下面计算 $L_\Delta(i, t)$ 。根据网络的构造有 $L_\Delta(i, t) = (1+q)L_\Delta(i, t-1)$, 再使用初始条件 $L_\Delta(i, t_i) = 1$, 可以推导出 $L_\Delta(i, t) = (1+q)^{t-t_i}$ 。因此, 在时间为 t 时节点 i 的度数为

$$k_i = 2(1+q)^{t-t_i} + 2 \quad (4)$$

因此网络中每个节点的度数都可以通过公式(4)获得, 从而能由公式 $P_{cum}(k) = \sum_{k' \geq k} N(k', t) / N_t \sim k^{-1-\gamma}$ 描述网络的累积度

分布, 式中 $N(k', t)$ 表示度为 k' 的节点的个数。详细的解析过程如下, 度为 $k=2(1+q)^{t-m}+2$ 的节点数为 $L_v(m)=3q(1+2q)^{m-1}$ 。所有拥有该度数的节点都产生于时间步 m 。因此得到

$$\sum_{k' \geq k} N(k', t) = 4 + \sum_{s=1}^m L_v(s) = 4 + \frac{3}{2}((1+2q)^m - 1)$$

节点的总数已经由式(1)给出:

$$[2(1+q)^{t-m} + 2]^{-\gamma} = \frac{4 + \frac{3}{2}((1+2q)^m - 1)}{\frac{3}{2}(1 + (1+2q)^t)}$$

因此, 当 t 取很大值时

$$2^{1-\gamma}[(1+q)^{t-m}]^{-\gamma} \approx (1+2q)^{m\gamma}$$

和

$$\gamma \approx 1 + \frac{\ln(1+2q)}{\ln(1+q)} \quad (5)$$

因此度指数 γ 是 q 的连续函数, 并且属于可调区间 $(2, 3]$, 当 q 从 0 增长到 1 时, γ 由 3 减小到 $1 + \frac{\ln 3}{\ln 2}$ 。在两种极限情况下, 当取 $q=1$ 和 $q \rightarrow 0 (q \neq 0)$ 时, 演化的 Sierpinski 网络弱化为确定性 Sierpinski 网络^[10]和随机 Sierpinski 网络。图 4 显示了在对数坐标下, 对应不同值的累积度分布 $P_{cum}(k)$ 的数值模拟结果。从图中可以看出, 随着 q 值的增大, γ 值逐渐减小, 因此, 模拟结果与理论分析是一致的。

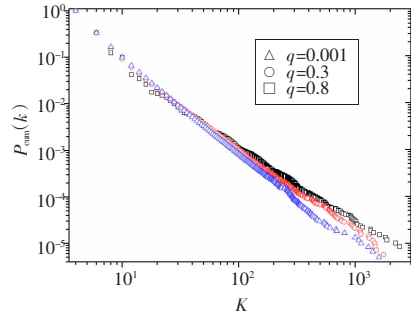


图 4 规模分别为 $t=20\ 000 (q=0.001), t=26 (q=0.3), t=12 (q=0.8)$ 网络累积度分布的双对数图

4.2 簇系数

根据定义, 度为 k_i 的节点 i 的簇系数^[11]为 $C_i = 2e_i / [k_i(k_i - 1)]$, 其中 e_i 表示 k_i 个邻居之间实际存在的边数, $k_i(k_i - 1) / 2$ 为 k_i 个邻居之间可能存在的最大边数。利用网络的生成规则, 能够直接分析计算度数为 k 的节点的簇系数 $C(k)$ 。当一个新节点加入网络时, 其 k_i 和 e_i 都等于 4。在以后的每一时间步, 如果包含该节点的一个活三角形被选择, 其和分别增加 2 和 3。因此, 在任一时间步, 对所有的节点均有下式成立: $e_i = 4 + \frac{3}{2}(k_i - 4)$ 。于是, 节点的簇系数和度数之间存在一个一一对应关系:

$$C(k) = \frac{2[4 + \frac{3}{2}(k-4)]}{k(k-1)} = \frac{4}{k} - \frac{1}{k-1} \quad (6)$$

当 k 很大时, 簇系数与成 k 反比。

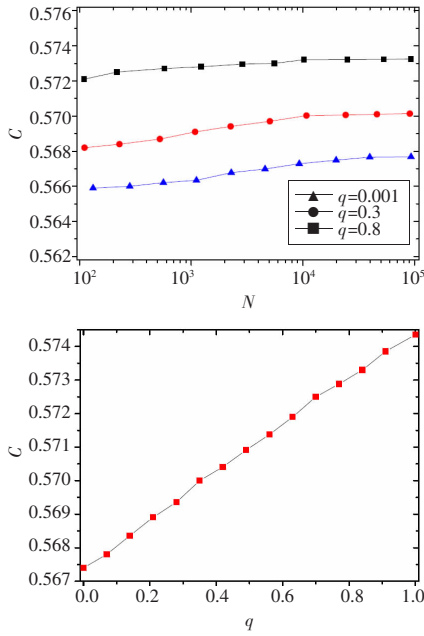
经过 t 步演化以后, 网络的平均簇系数 C_t 定义为网络中所有节点簇系数的平均值 C'_t , 由下式给出:

$$C'_t = \frac{1}{N} \sum_{r=0}^t [(\frac{4}{K_r} - \frac{1}{K_r-1}) L_v(r)] \quad (7)$$

其中连加符号所表示的和项为所有节点的簇系数之和, K_r 表示在第 r 步生成的节点在第 t 步时的度, 其度数 K_r 由式(4)给出。在 ESN 网络中可以很容易地证明 C'_t 随着 q 的增长而增长。当网络节点数 N_t 趋于无穷大时, q 从 0 增长到 1, 式(7)收敛值的范围从 0.567 4 增长到 0.574 5。因此该网络是个高集聚的网络。对于不同规模 ESN 的平均簇系数进行了数值模拟, 模拟结果如图 5 所示。从图 5 左图可以看出, 随着网络规模的增长, 不同 q 值 ESN 的簇系数均会慢慢收敛于一个稳定值。图 5 右图的方块对应不同 q 值 ESN 的数字模拟结果, 由图示数值模拟结果与理论分析结果是相吻合的。

4.3 平均路径长度

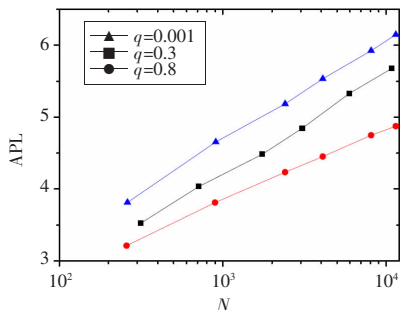
从以上讨论可以发现, ESN 同时具有无尺度特性和较大的

图5 不同参数 q 网络的平均簇系数 C

簇系数。事实上,ESN 模型还具有小世界性质。

利用平均场的方法^[12],可以从理论上预测 ESN 的平均路径长度(APL)。根据网络的构造方法,在每一个时间步,有 3 个新节点加入网络。为了区别不同的节点,可以采用如下的方式构建一个节点序列:在某一个给定的时间步,当 3 个新节点产生时,分别标记这 3 个点为 $M+1, M+2, M+3$,其中 M 表示新节点加入之前的节点总数。这样每一个节点都有唯一的整数标号,时间步 t 的总节点数目为 $N_t = \frac{3}{2}(1+(1+2q)^t)$ 。用 $L(N)$ 表示节点数为 N 时网络的 APL,于是有 $L(N) = \frac{2D(N)}{N(N-1)}$,其中 $D(N) =$

$\sum_{1 \leq i < j \leq N} d_{i,j}$ 表示不同节点对之间的最短路径长度。与文献[12]中的分析类似,可以很容易地推导出:当 N 足够大时, $D(N) \sim N^2 \ln N$,从而有 $L(N) \sim \ln N$ 。因此 ESN 的平均路径长度 $L(N)$ 随着网络规模 N 呈现对数增长,这说明该网络具有小世界效应。图 6 描绘了平均路径长度 $L(N)$ 和网络规模 N 之间的相互关系,从图中可以很明显地看到 $L(N)$ 随 N 的增长呈现对数增长。

图6 平均路径长度 L 关于网络规模 N 的半对数图

5 结论

综上所述,基于 Sierpinski 分形垫,提出并研究了一类演化网络:演化 Sierpinski 网络(ESNs)。根据网络的构造过程引入了生成网络的一般算法,并基于此对 ESN 网络的特征进行了解析计算和数字模拟,网络的度分布、簇系数、以及平均路径长度的解析结果与实际观察值较好地吻合。度指数能在 2 和 3 之间进行调整,具有较小的平均路径长度和较大的簇系数。

本文所揭示的 Sierpinski 分形垫与不相容网络之间的联系,对于研究相关的重要网络有一定的借鉴意义,所使用的映射方式为通过网络理论研究现实系统的复杂性提供了很好的范例。例如,用类似的方法将城市道路映射成不相容网络,可以量化道路系统的导航信息^[6]。另外,这一映射方式也可以用于旅行商问题^[13]。需要特别指出的是,提出的一类不相容网络为最大平面图,该特征对于设计印刷电路板有一定的帮助和借鉴价值。

参考文献:

- [1] Watts D J, Strogatz S H. Collective dynamics of "small-world" networks[J]. Nature, 1998, 393: 440-442.
- [2] Barabási A-L, Albert R. Emergence of scaling in random networks[J]. Science, 1999, 286: 509-512.
- [3] Dorogovtsev S N, Mendes J F F, Samukhin A N. Structure of growing networks with preferential linking[J]. Physical Review Letters, 2000, 85(21): 4633-4636.
- [4] Krapivsky P L, Redner S, Leyvraz F. Connectivity of growing random networks[J]. Physical Review Letters, 2000, 85: 4629-4632.
- [5] Fronczak, Fronczak P, Holyst J A. Mean-field theory for clustering coefficients in Barabási-Albert networks[J]. Physical Review E, 2003, 68.
- [6] Rosvall M, Trusina A, Minnhagen P, et al. Networks and cities: An information perspective[J]. Physical Review Letters, 2005, 94.
- [7] Kabacıoğlu A, Stella A L. Pseudoknots in a homopolymer[J]. Physical Review E, 2004, 70.
- [8] Kabacıoğlu A, Stella A L. Scale-free network hidden in a collapsing polymer[J]. Physical Review E, 2005, 72.
- [9] Majewski M. A tutorial of the realistic visualization of 3D Sierpinski fractals[J]. Computers and Graphics, 1998, 22(1): 129-142.
- [10] Zhang Zhongzhi, Zhou Shuigeng, Zou Tao, et al. Incompatibility networks as models of scale-free small-world graphs[J]. The European Physical Journal B, 2007, 60(2): 259-264.
- [11] West D B. Introduction to graph theory[M]. Upper Saddle River, NJ: Prentice Hall, 2001.
- [12] Zhou Tao, Yan Gang, Wang Binghong. Maximal planar networks with large clustering coefficient and power-law degree distribution[J]. Physical Review E, 2005, 71.
- [13] Jacobsen J L, Read N, Saleur H, et al. Traveling salesman problem, conformal invariance, and dense polymers[J]. Physical Review Letters, 2004, 93.