

## 一种信息传播促进网络增长的网络演化模型\*

刘树新<sup>1)†</sup> 季新生<sup>1)</sup> 刘彩霞<sup>1)</sup> 郭虹<sup>2)</sup>

1)(国家数字交换系统工程技术研究中心, 郑州 450002)

2)(信息工程大学信息工程学院, 郑州 450002)

(2014年1月23日收到; 2014年4月9日收到修改稿)

为了研究信息传播过程对复杂网络结构演化的影响, 提出了一种信息传播促进网络增长的网络演化模型, 模型包括信息传播促进网内增边、新节点通过局域世界建立第一条边和信息传播促进新节点连边三个阶段, 通过多次自回避随机游走模拟信息传播过程, 节点根据路径节点的节点度和距离与其选择性建立连接。理论分析和仿真实验表明, 模型不仅具有小世界和无标度特性, 而且不同参数下具有漂移幂律分布、广延指数分布等分布特性, 呈现小变量饱和、指数截断等非幂律现象, 同时, 模型可在不改变度分布的情况下调节集聚系数, 并能够产生从同配到异配具有不同匹配模式的网络。

**关键词:** 复杂网络, 信息传播, 演化模型, 自回避随机游走

**PACS:** 89.75.-k, 89.75.Fb

**DOI:** 10.7498/aps.63.158902

## 1 引言

复杂网络是一个多学科交叉的研究领域, 自上世纪60年代ER随机图模型<sup>[1]</sup>的提出开启了复杂网络研究的大门以来, 便备受瞩目, 尤其是复杂网络的小世界特性<sup>[2]</sup>和无标度特性<sup>[3]</sup>的发现之后, 在丰富和改变人们对复杂网络认识的同时, 也掀起了各领域研究复杂网络的热潮<sup>[4]</sup>。近年来, 越来越多的现实网络也逐渐成为复杂网络的研究对象, 涉及互联网、社会网络、生物网络、经济网络、运输网络等多种复杂性系统<sup>[5-9]</sup>。对于复杂网络研究的范围也不断拓展, 从复杂网络拓扑建模<sup>[2,3]</sup>到网络动力学性质<sup>[10,11]</sup>, 从网络信息挖掘<sup>[12,13]</sup>到网络的干预和控制<sup>[14-16]</sup>, 然而对复杂网络演化机理的探索一直是各领域学者研究的热点。

Barabási和Albert提出的BA模型<sup>[3]</sup>, 揭示了无标度网络演化的基本机理为增长和优先连接, 其网络演化过程假定新节点具有网络的全局信息, 每增加一个新节点的同时, 新节点根据节点度择优建立 $m$ 条边。后续很多学者在BA模型的度优先连接

的连接概率和添加本地事件(local events)上进行了改进<sup>[17,18]</sup>。由于现实网络呈现高的集聚性, 文献<sup>[19]</sup>在BA模型优先连接阶段以概率加入了一个TF(Triad Formation)机理, 即新加入节点在优先连接的节点的周围邻居节点中随机选择节点建立连接, 但新节点在加入网络阶段仍然需要网络的全局信息。

在世界贸易网、互联网等网络研究中发现择优连接机理仅仅存在于局部范围内, 针对这一现象, Li等提出了一种局域世界演化模型<sup>[20]</sup>, 该模型不同于BA模型的全局节点择优连接, 而是在网络中随机选择 $M$ 个节点构成的局域世界中进行择优连接。文献<sup>[21]</sup>等在局域世界模型的基础上对其进行拓展, 但局域世界的形成多为随机节点的组合, 忽略了局域世界内部节点间的紧密联系<sup>[22,23]</sup>和局域世界的形成原因<sup>[24]</sup>。Saramäki和Kaski提出了一种通过随机游走(random walk)构造无标度网络的方法, 新节点和随机游走中标记的 $m$ 个节点建立连接, 产生了与BA模型相同的效果<sup>[25]</sup>。文献<sup>[24]</sup>进一步利用随机游走构造局域世界, 并依概率采

\* 国家高技术研究发展计划(批准号: 2011AA010605, 2011AA010604)资助的课题。

† 通讯作者。E-mail: liushuxin11@gmail.com

用随机连接、“扶贫”连接和“亲富”连接三种连接策略,然而其中的随机行走只是作为构造局域世界的方法,仅仅研究了新节点和其产生的连接,并没有研究局域信息在网络中随机行走对路径节点之间边的演化的影响,而且其连接选择过程中涉及的局域信息有限.

当前,充分考虑局域信息、节点相互作用且能够再现网络实证特点的耦合网络演化模型相对较少,朱陈平等提出了一种利用竞争排斥原理产生无标度网络的演化模型<sup>[26]</sup>,文献<sup>[5]</sup>则从网络的局部结构出发研究了共同邻居驱动网络演化的机理,而文献<sup>[6]</sup>则针对典型的社交媒体网络提出了基于优先选择三元闭包(triadic closure)的耦合网络演化模型.研究表明许多实际网络的度分布介于幂律分布和指数分布之间,如幂律尾分布(power law tail)、去头的幂律分布(truncated power law)等<sup>[27]</sup>,呈现出指数截断、小变量饱和等非幂律现象<sup>[28]</sup>,文献<sup>[27]</sup>和<sup>[29]</sup>分别利用广延指数分布(stretched exponential distribution)、漂移幂律分布(shift power-law distribution)对相关现象进行了实证研究,并取得了较好的效果.

现实世界中,局域信息的传播往往是新的连接产生的直接原因,例如:在人际关系网中,热点信息在朋友之间随机传播,个体通过朋友的介绍与传播路径上的个体建立新的人际连接;同样,对于新加入的个体,在接触到第一个朋友之后,其人际关系的建立也是如此.在社会网络中,各种话题、新闻通过转发在网络中传播,用户通过朋友的转发与信息路径上的用户建立新的朋友关系.可以看出,信息传播是为网络内部演化提供局域信息的重要源头,网络的演化过程不仅是单纯物理结构上的增删,也是网络内部信息交互的结果<sup>[5,6,26]</sup>.互联网等网络实证显示,网络边的生长速度远大于点的增加速度<sup>[30]</sup>,而网络内部边的急剧增长与网络内部信息的爆炸性传播是密不可分的.

基于信息传播对节点局部信息传播的促进作用,本文提出了一种信息传播促进网络增长的网络演化模型.该模型以局域信息的来源作为切入点,研究信息传播路径上连边对网络演化的影响,建立连接阶段同时考虑了路径节点的节点度和距离,产生包括随机连接、度优先连接、扶贫连接、近邻连接、远交连接以及可相互组合的连接方式.模型演化的网络具有明显的幂律特性、小世界特性和高集聚性,且可以在不改变度分布的情况下较大范

围内调节网络的集聚系数,其度度关联系数可以在(-0.6, 0.5)较大范围内变化.

## 2 信息传播促进网络增长的网络演化模型

### 2.1 信息传播及其路径上的连边机理

节点间的信息传播是一个随机的过程,理论上最短路径传播是一种最佳的传输方式,但最短路径传输的前提是每个节点均得知全局的连接信息<sup>[31]</sup>.由于随机游走仅仅需要局部的连接信息,所以它已经成为许多学者刻画实际网络中单个信息传播过程的重要方法之一.

在无限制随机游动过程中,游走节点(walker)存在重复游走路径节点的现象<sup>[25]</sup>,为了距离的计算方便和路径上节点数目的可调,又不影响信息传播的随机性,本文采用自回避随机游走模拟信息从节点 $s$ 经过 $L$ 跳传播到达随机节点 $j$ 这一过程.所谓自回避随机游走,就是一种每次从邻居节点中随机选择下一跳节点时,能够回避已选择节点的随机游走方法<sup>[32]</sup>.

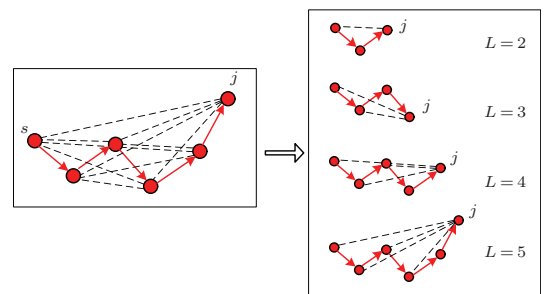


图1 传播路径上连边简化示意图

传播路径上任意两点都有可能产生连接,本文采用源点或终点和路径上的任意节点建立连接,通过调节不同的传输长度 $L$ 来达到任意两点建立连接的可能(如图1所示).节点间建立连接方式有很多种,与节点度相关的有BA模型的择优连接<sup>[3]</sup>,GLP模型的线性度优先连接<sup>[19]</sup>,Tang模型的超线性度优先连接<sup>[33]</sup>以及大一统混合网络模型(LUHNM)中的扶贫连接<sup>[34]</sup>;与距离相关的有文献<sup>[35]</sup>提出的距离优先连接、远交连接,TF机理<sup>[19]</sup>以及邻居优先连接<sup>[36]</sup>;当然也有随机连接机理<sup>[37]</sup>.上述连接机理表明,复杂网络中连接机理涉及不同的局域信息且对局域信息的选择处理方式各异,由于信息传播路径中传播的重要局域信息主要包括节

点度和距离, 结合两种局域信息, 本文定义了信息路径上节点间吸引力  $F_{ij}$  用于表征路径上节点  $i$  和节点  $j$  之间可能建立连接的程度:

$$F_{ij} = \frac{k_i^\alpha \cdot k_j^\alpha}{d_{ij}^\beta}, \quad d_{ij} \geq 1, \quad (1)$$

式中  $k_i, k_j$  为节点的度,  $\alpha$  为调节参数,  $d_{ij}$  为信息路径上的节点  $i$  到节点  $j$  所经过的跳数,  $\beta$  为  $d_{ij}$  的调节参数. 则节点  $j$  选择路径上的节点  $i$  建立连接的概率为

$$\Pi(k_i) = \frac{F_{ij}}{\sum_{r \in I} F_{rj}} = \frac{k_i^\alpha \cdot d_{ij}^{-\beta}}{\sum_{r \in I} k_r^\alpha \cdot d_{rj}^{-\beta}}, \quad (2)$$

其中  $I$  是路径上节点  $j$  的非邻居节点的集合. 可以通过调节参数的取值分析传播路径上的连接方式对网络演化的影响. 但从网络总体趋势上分析, 在传播路径上任意连接机理, 都具以下特点:

- 1) 在传播路径上以任意连接方式建立连接, 都是对信息传播效率的提高;
- 2) 网络局域信息的总体考虑是对不同网络中节点偏好的满足.

## 2.2 新节点的加入方式

BA 模型采用全局度优先的方式加入新节点, 后续产生了局域世界范围内择优连接的加入方式, 最近文献 [37] 提出了一种新节点与随机节点的邻居节点建立连接的无标度网络模型. 考虑到现实网络中, 新节点在建立第一条边时的知识范围有限, 本文采用局域世界作为新节点产生第一条边的局域信息来源.

通过随机选取任意节点构造局域世界的方法忽略了现实局域世界节点间的紧密联系 [23,24], 文献 [23] 提出了一种随机点和所有邻居节点组成局域世界的方法, 而文献 [24] 提出的随机游走构建局域世界也一定程度上避免了此种情况, 但现实网络中局域世界的产生并不仅仅是随机游走这一种方式, 而从局域信息传递的角度, 任何一种连通的网络形式都有可能成为局域世界. 本文采用随机生成连通网络的方法构建大小为  $M$  的局域世界, 具体如下:

- 1) 从网络中随机选取一个节点加入到局域世界中;
- 2) 随机选择一个与局域世界相连的点加入局域世界, 此过程重复  $M - 1$  次.

局域世界建立后, 新节点依 (2) 式选择第一个接入点时, 由于新节点对局域世界的距离信息获取有限, 假定此时新节点与任意网络节点的距离相同, 则 (2) 式简化为

$$\Pi(k_i) = \frac{k_i^\alpha \cdot d_{ij}^{-\beta}}{\sum_{r \in \text{local}} k_r^\alpha \cdot d_{rj}^{-\beta}} = \frac{k_i^\alpha}{\sum_{r \in \text{local}} k_r^\alpha}. \quad (3)$$

## 2.3 网络演化模型

在本文的模型中, 在新节点通过局域世界建立第一条连接加入到网络以后, 后续连边均属于网内信息传播促进内部连边这一过程, 其主要为两类: 一是新节点建立后续连接, 另一类便是网络内部所有节点间信息交互产生新连接, 即文献 [17] 中提到的本地事件, 鉴于内部边的重连可能分割网络 [19], 模型仅仅研究信息传播对网络内部连边产生的影响, 且作为初步的出发点, 在模型研究的当前阶段我们暂时忽略了信息传播的方向性. 在选择网络信息传播源时, 以节点度作为大规模网络中影响力的基本判断标准 [4], 即度大的节点在某一时间点拥有信息的可能更大.

基于上述考虑, 以信息传播促进网络边增长的角度, 建立模型如下:

- 1) 网络初始为  $m_0$  个节点,  $e_0$  条边, 且保证网络的连通性.
- 2) 信息传播促进网内增边: 以概率  $p(0 \leq p < 1)$  增加  $m$  条边. 在网络所有节点中, 依度优先选择节点  $s_0$  作为传播源, 通过信息传播  $L(L \geq 2)$  跳到达节点  $j$ , 节点  $j$  依概率  $\Pi(k_i)$  选择路径上的节点  $i$  建立一条连接 ((2) 式), 此过程重复  $m$  次.
- 3) 局域世界: 生成  $M(M \geq 1)$  个节点的连通网络作为局域世界 (如 2.2 所述).
- 4) 增长: 以概率  $1 - p$  加入一个新节点和  $m$  条边. 新节点依概率  $\Pi(k_i)$  选择局域世界中的节点  $i$  建立第一条边 ((3) 式), 后续  $m - 1$  条连边如 5) 所述.
- 5) 信息传播促进新节点连边: 以新节点为传播源, 传播  $L$  跳, 新节点和传播路径上的节点  $i$  建立一条, 选择概率为  $\Pi(k_i)$  ((2) 式), 此过程重复  $m - 1$  次.
- 6) 返回 2), 直至网络达到指定大小.

模型整体过程分为三阶段: 信息传播促进网内增边、新节点通过局域世界加入网络和信息传播促进新节点连边, 分别如图 2(a), (b), (c) 所示. 模型以自回避随机游走模拟信息传播过程, 在未完成  $L$

跳传播前,若当前节点没有可游走的邻居节点时,重新开始新一轮游走.

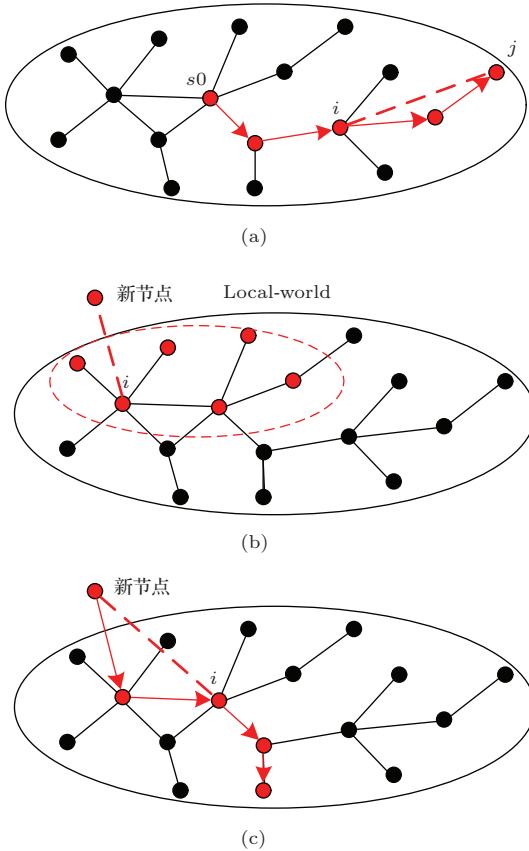


图2 模型分步骤演化示意图 (a) 信息传播促进网内增边; (b) 新节点加入网络; (c) 信息传播促进新节点连边

### 3 模型度分布分析

模型在  $t$  时刻,从网络  $G(V_t, E_t)$  中依概率选取一个节点  $s_0$  作为传播源,源点经过  $L$  跳信息传播到节点  $j$  形成一个  $L - 1$  的待连接路径序列  $I_t^{L-1}$ . 源节点  $s_0$  被选择的概率为  $P(s_0) = k_{s_0} / \sum k$ ,则在  $k_{s_0}$  个邻居节点中游走至下一跳节点  $s_1$  的概率为 [25]

$$P(s_1) = \frac{P(s_1/s_0)P(s_0)}{P(s_0/s_1)}, \quad (4)$$

其中  $P(s_1/s_0)$  表示在选择  $s_0$  后,游走到节点的  $s_1$  概率,即  $P(s_1/s_0) = 1/k_{s_0}$ ; 同理,  $P(s_0/s_1) = 1/k_{s_1}$ . 则 (4) 式化简为

$$\begin{aligned} P(s_1) &= \frac{k_{s_1}}{k_{s_0}} P(s_0) \\ &= \frac{k_{s_1}}{\sum k}. \end{aligned} \quad (5)$$

以同样的方式分析,在不考虑自回避情况下,节点游走过程中到达节点  $i$  的概率为  $P(i) =$

$k_i / \sum k$ , 则任意一个节点度为  $k_i$  的节点出现在  $I_t^{L-1}$  中的概率为

$$\begin{aligned} P(S_i^{L-1}) &= \sum_{r=0}^{L-2} \left(1 - \frac{k_i}{\sum k}\right)^r \cdot \frac{k_i}{\sum k} \\ &= 1 - \left(1 - \frac{k_i}{\sum k}\right)^{L-1}. \end{aligned} \quad (6)$$

采用准连续近似方法,假设节点度随时间连续变化,模型分析如下:

1) 以概率  $p$  在网络内部产生  $m$  条边. 此过程中,节点  $i$  的度随时间的变化为

$$\begin{aligned} \left(\frac{\partial k_i}{\partial t}\right)_{\text{intra}} &= mp \left[ \frac{k_i}{\sum k} + P(S_i^{L-1}) \frac{k_i^\alpha \cdot d_i^{-\beta}}{\sum_{r \in I_t^{L-1}} (k_r^\alpha \cdot d_r^{-\beta})} \right]. \end{aligned} \quad (7)$$

(7) 式中,第一项为信息传播的终点(图2中的节点  $j$ ) 的度的变化,第二项为在传播路径  $I_t^{L-1}$  上连边节点的度变化情况.

2) 以概率  $1 - p$  添加一个新节点和  $m$  条边. 在此阶段中,节点  $i$  的度随时间的变化为

$$\begin{aligned} \left(\frac{\partial k_i}{\partial t}\right)_{\text{inter}} &= (1 - p) \left[ \Pi'(i \in \text{local}) \frac{k_i^\alpha}{\sum_{\text{local}} k^\alpha} \right. \\ &\quad \left. + (m - 1) \cdot P(S_i^{L-1}) \frac{k_i^\alpha \cdot d_i^{-\beta}}{\sum_{r \in I_t^{L-1}} (k_r^\alpha \cdot d_r^{-\beta})} \right]. \end{aligned} \quad (8)$$

式中两项分别为建立第一条边和后续  $m - 1$  条边的结果. 对于  $\Pi'(i \in \text{local})$ , 在本文的局域世界构造方法中,通过一条边选择节点加入到局域世界的概率与节点的度  $k_i$  有关 [22], 即  $P(i) = k_i / \sum k$ , 则任意节点  $i$  出现在局域世界中的概率为

$$\begin{aligned} \Pi'(i \in \text{local}) &= 1 - \left(1 - \frac{1}{N_t}\right) \left(1 - \frac{k_i}{\sum k}\right)^{M-1}, \\ M &\geq 1. \end{aligned} \quad (9)$$

综合上述两个过程,节点度随时间的变化为

$$\begin{aligned} \frac{\partial k_i}{\partial t} &= \left(\frac{\partial k_i}{\partial t}\right)_{\text{intra}} + \left(\frac{\partial k_i}{\partial t}\right)_{\text{inter}} \\ &= \frac{mpk_i}{\sum k} + (1 - p) \Pi'(i \in \text{local}) \frac{k_i^\alpha}{\sum_{\text{local}} k^\alpha} \\ &\quad + (m + p - 1) \cdot P(S_i^{L-1}) \end{aligned}$$

$$\times \frac{k_i^\alpha \cdot d_i^{-\beta}}{\sum_{r \in I_t^{L-1}} (k_r^\alpha \cdot d_r^{-\beta})}. \quad (10)$$

根据模型参数不同, 模型存在几个特殊情况, 分别讨论如下:

**情形 A**  $\alpha = 0, \beta = 0$

此时模型为随机连接, 对于大多数中低度节点,  $t \rightarrow \infty$  时, 有  $(1 - \frac{k_i}{2mt}) \sim 1$ , 故节点出现在传播路径上或局域世界中且被选择的概率近似为  $\frac{(L-1)k_i}{\sum k}$  和  $\frac{1}{N_t} + \frac{(M-1)k_i}{\sum k}$ , 节点度的变化为

$$\begin{aligned} \frac{\partial k_i}{\partial t} &\approx \frac{mpk_i}{\sum k} + (1-p) \cdot \frac{1}{M} \left( \frac{1}{N_t} + \frac{(M-1)k_i}{\sum k} \right) \\ &\quad + \frac{(m+p-1)}{L-1} \cdot \frac{(L-1)k_i}{\sum k} \\ &\approx \left( \frac{p}{2} + \frac{(1-p)(M-1)}{2mM} + \frac{m+p-1}{2m} \right) \cdot \frac{k_i}{t} \\ &\quad + \frac{1-p}{M} \cdot \frac{1}{t} \\ &= A(p, m, M) \cdot \frac{k_i}{t} + B(p, M) \cdot \frac{1}{t}. \end{aligned} \quad (11)$$

根据上式, 模型的度分布推导结果为

$$P(k) = \frac{1}{A} (m + \theta)^{\frac{1}{\lambda}} \cdot (k + \theta)^{-\frac{1}{\lambda}-1}, \quad (12)$$

其中  $\theta = B/A$ , 模型服从  $\lambda = 1/A + 1$  的漂移幂律分布. 当  $\theta < 1$  时, 漂移幂律分布在双对数坐标下呈现一条近似幂律分布的斜线; 当  $\theta \geq 100$  时, 近似为指数分布;  $1 \leq \theta < 100$  时, 则呈现前端指数分布, 后续幂律分布的形式<sup>[29]</sup>. 由于漂移幂律分布同时具有指数和幂律特性, 常慧、何大韧等曾利用其刻画运输网络、技术协作网和社会网络等实际网络的度分布情况, 并取得了较好的实证效果<sup>[29]</sup>. 模型中当  $M \leq 2$  且  $p < \frac{2m+1-Mm}{2m+1+Mm}$  时, 存在  $\theta > 1$  的情况, 其他情况均服从幂律分布, 其中参数  $\lambda$  随  $p, M, m$  取值的不同而变化,  $M$  较大时,  $\lambda = 2/(p+1) + 1$ . 同理, 对于高度节点,  $t \rightarrow \infty$  时, 参数  $A(p, m, M)$  受节点度影响:

$$\begin{aligned} A(p, m, M) &= \frac{p}{2} + \frac{(1-p)}{2mM} \sum_{r=0}^{M-2} \left(1 - \frac{k_i}{2mt}\right)^r \\ &\quad + \frac{m+p-1}{2m(L-1)} \sum_{r=0}^{L-2} \left(1 - \frac{k_i}{2mt}\right)^r. \end{aligned} \quad (13)$$

随着节点度  $k_i$  的增大,  $A$  逐渐减小, 高度节点区域将会出现  $\theta > 1$  的情形, 即模型尾部呈现一定

程度的指数分布倾向, 且随着  $M, L$  值越大, 倾向越明显.

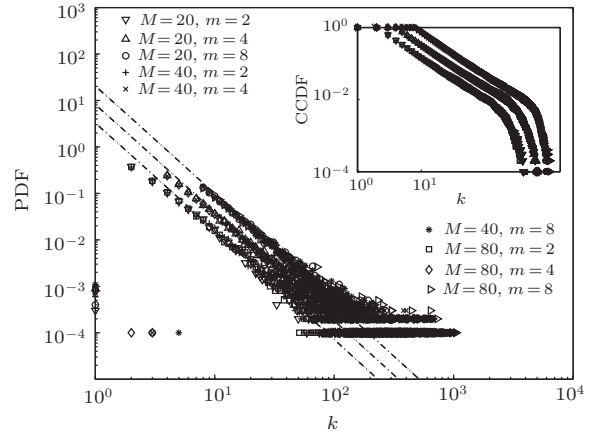


图3 模型特殊情形 A 中  $M, m$  对度分布的影响 ( $N = 10000, a = 0, b = 0, p = 0.5, L = 20$ , 插图为互余累积度分布曲线)

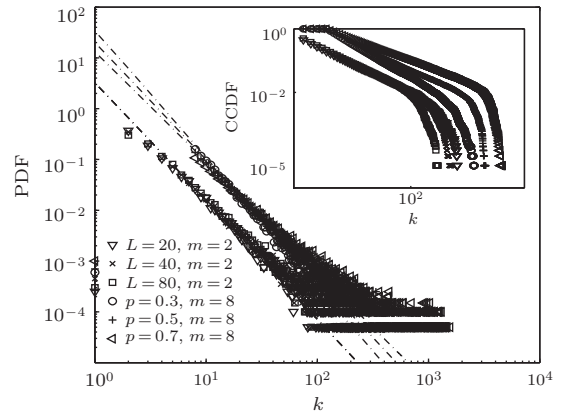


图4 模型特殊情形 A 中  $p, L$  对度分布的影响 ( $N = 10000, \alpha = 0, \beta = 0, p = 0.5, L = 20, M = 20$ , 插图为互余累积度分布曲线)

图3和图4显示了不同  $p, m, M, L$  值下的度分布情况, 其中虚线为(12)式理论分析结果, 在误差允许的范围, 可以较好的反映模型的度分布情况; 图3中可以看出, 在  $M$  较大时, 指数  $\lambda$  与参数  $m$  无关, 而图4中显示了  $\lambda$  随着概率  $p$  的增大减小的趋势, 这进一步证实了理论计算的结果, 且不同于 BA 模型中固定的参数  $\lambda$ , 其范围可在 (2, 3) 内变化<sup>[38]</sup>, 这与实际网络统计结果较为符合, 说明了信息传播促进网络增长具有一定的合理性和真实性; 插图的互余累积度分布函数曲线中均显示了高度节点呈现指数截断现象, 且随着信息传播的跳数  $L$  的增大而更加明显, 与理论推导结果一致, 但并未随着  $M$  增大而更加明显, 这与模型在局域世界中仅建立一条边有关. 指数截断现象广泛存在于蛋白

质网络 [7]、演员合作网 [8] 等复杂系统中, 模型可为其出现提供另一种解释. 由于  $M$  的影响较小, 在后续研究中取经验值  $M = 20$ .

**情形 B  $\beta = 0$**

此时, 模型为度优先、扶贫连接.  $t \rightarrow \infty$  时, 对大多数中低度节点有

$$\begin{aligned} \frac{\partial k_i}{\partial t} &\approx \frac{pk_i}{2t} + (1-p) \cdot \left( \frac{1}{t} + \frac{(M-1)k_i}{2mt} \right) \cdot \frac{k_i^\alpha}{\sum_{\text{local}} k^\alpha} \\ &\quad + (m+p-1) \frac{(L-1)k_i}{2mt} \frac{k_i^\alpha}{\sum_{I_t^{L-1}} k_r^\alpha} \\ &\approx \left( \frac{p}{2} + \frac{(1-p)(M-1) \cdot \sigma_{\text{local}}}{2m} \right. \\ &\quad \left. + (m+p-1) \frac{(L-1) \cdot \sigma_{I_t^{L-1}}}{2m} \right) \cdot \frac{k_i}{t} \\ &\quad + \frac{(1-p) \cdot \sigma_{\text{local}}}{t} \\ &= A'(p, m, M, k_i) \cdot \frac{k_i}{t} + B'(p, M, k_i) \cdot \frac{1}{t}, \quad (14) \end{aligned}$$

其中,  $\sigma_{\text{local}} = \frac{k_i^\alpha}{\sum_{\text{local}} k^\alpha}$ ,  $\sigma_{I_t^{L-1}} = \frac{k_i^\alpha}{\sum_{I_t^{L-1}} k_r^\alpha}$ , 令  $\theta' =$

$B'/A'$ , 通过漂移幂律分布对其做定性分析如下:

1) 度优先连接 ( $\alpha > 0$ ). 由于  $k_i \sim m$  附近的节点出现在局域世界或传播路径上的数目大于其他低度节点, 导致在低度节点时,  $\sigma_{\text{local}}, \sigma_{I_t^{L-1}}$  随着  $k_i$  的增大呈现减小的趋势, 因此, 模型头部呈现  $\lambda$  由小变大的幂律分布形式, 即存在指数分布的趋向, 且参数  $\alpha, L$  越大越明显; 随着  $k_i$  的逐渐增大, 此时  $\sigma_{\text{local}}, \sigma_{I_t^{L-1}}$  逐渐增大,  $\lambda$  逐渐减小, 呈现一种幂指数轻微变小的幂律分布, 且参数  $m, p$  的值越大, 参数  $A'$  的值越稳定,  $\lambda$  变化越小.

2) 扶贫连接 ( $\alpha < 0$ ). 随着  $k_i$  的增大,  $\sigma_{\text{local}}, \sigma_{I_t^{L-1}}$  一直持续的变小, 且参数  $\alpha$  越小变化越快, 此时  $\theta' < 1$ , 度分布为  $\lambda$  逐渐变大幂律分布, 变化缓慢时为近幂律分布, 而较快时为指数分布, 其中  $L$  越小, 变化越缓慢.

对于高度节点,  $A'(p, m, M, k_i)$  则为

$$\begin{aligned} &A'(p, m, M, k_i) \\ &= \frac{p}{2} + \frac{(1-p) \cdot \sigma_{\text{local}}}{2m} \sum_{r=0}^{M-2} \left( 1 - \frac{k_i}{2mt} \right)^r \\ &\quad + \frac{(m+p-1) \cdot \sigma_{I_t^{L-1}}}{2m} \sum_{r=0}^{L-2} \left( 1 - \frac{k_i}{2mt} \right)^r. \quad (15) \end{aligned}$$

度优先连接中, 随着  $k_i$  的增大参数  $A'$  会逐渐减小,  $\lambda$  逐渐增大, 导致模型尾部存在指数截断现

象, 且  $M, L$  越大越明显; 对于扶贫连接, 高度节点被选中的概率远小于 1, 其尾部也呈现急剧下降的趋势.

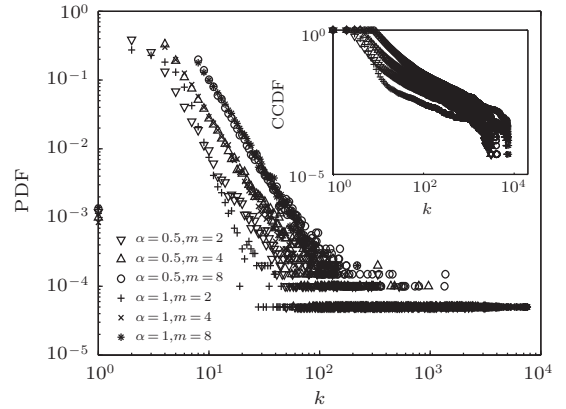


图5 模型特殊情形 B 中  $\alpha > 0$  时的度分布情况 ( $N = 10000, b = 0, p = 0.5, L = 20, M = 20$ , 插图为互余累积度分布曲线)

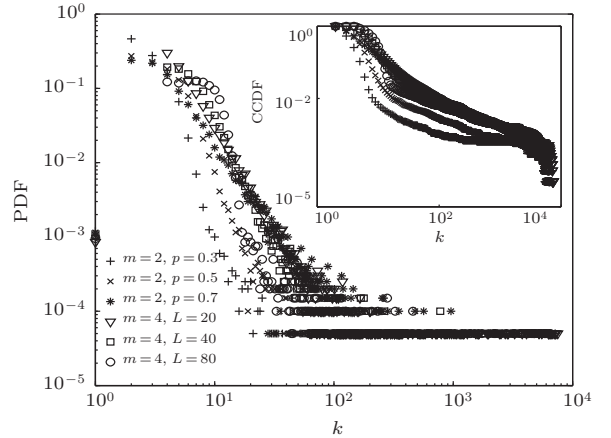


图6 模型特殊情形 B 中  $\alpha > 0$  时  $p, L$  对度分布的影响 ( $N = 10000, a = 1, b = 0, p = 0.5, L = 20, M = 20$ , 插图为互余累积度分布曲线)

模型情形 B 中度优先连接时的度分布情况如图 5 所示, 其中插图为互余累积度分布曲线, 模型头部和尾部均存在指数分布倾向, 表现为小变量饱和、指数截断现象, 中段遵循幂律分布, 总体上具有广延指数分布的特点, 且随着参数  $\alpha$  的减小, 幂律分布的范围逐渐增大, 与理论分析结果一致; 一些合作系统如中国旅游线路系统、淮阳菜肴系统以及好莱坞演员网均具有广延指数分布的特点, 且相关学者进行了一定的实证研究 [27], 模型可为其演化机理的研究提供一定的思路. 图 6 显示了在参数  $p$  增大或  $L$  减小时, 首、尾部分逐渐向幂律分布发展的趋势, 这说明信息传播的跳数也影响着幂律分布的范围, 互余累积度分布曲线显示模型存在指数截

断现象, 但受参数  $L$  影响较小. 此外, 模型存在长尾现象, 在  $\alpha > 1$  时可能存在节点度接近网络规模  $N$  的节点, 出现赢者通吃的现象.

图 7 显示了特殊情形 B 中扶贫连接时的度分布情况, 此时模型整体呈现从幂律分布向指数分布演变的过程, 且  $\alpha$  越大幂律分布的趋势越大. 参数  $p, L$  对此度分布的影响如图 8 所示, 信息传播的长度  $L$  不仅影响尾部的下降趋势, 也影响模型从幂律向指数演化的快慢.  $L$  越小, 模型幂律分布越明显, 图中  $L = 5$  时, 便呈现为幂律分布; 此外, 参数  $p$  影响了幂律分布的倾斜程度或指数分布的弯曲程度, 随着  $p$  的减小其呈现更快的下降趋势.

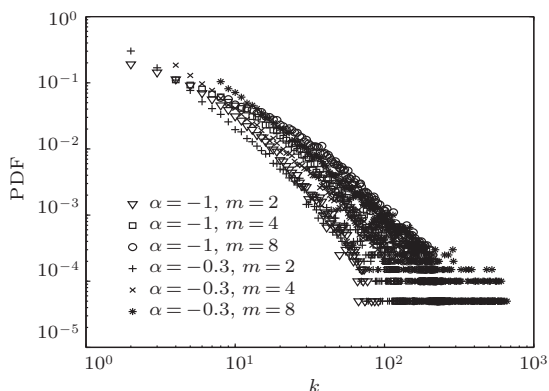


图 7 模型特殊情形 B 中  $\alpha < 0$  时的度分布情况 ( $N = 10000, \beta = 0, p = 0.5, L = 20, M = 20$ )

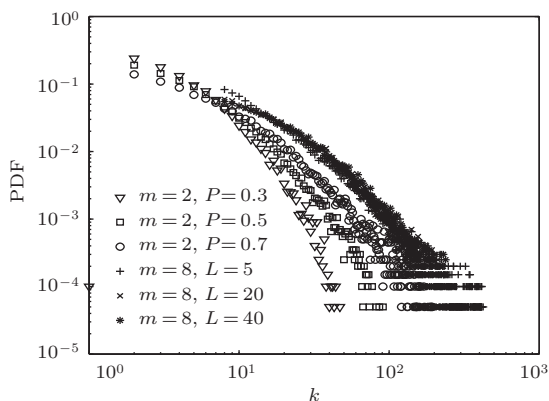


图 8 模型特殊情形 B 中  $\alpha < 0$  时  $p, L$  对度分布的影响 ( $N = 10000, \alpha = -1, \beta = 0, p = 0.5, L = 20, M = 20$ )

**情形 C  $\alpha = 0$**

此时, 模型在信息传播路径上进行近邻和远交连接. 由于节点出现在传播路径上的位置是随机的, 其被选中概率均为  $\frac{k_i}{(L-1)\sum k}$ , 理论上等同于情形 A 的情况. 图 9 显示了此情形下的度分布情况, 插图为互余累积度分布曲线, 其中虚线为情形 A 中 (12) 式的推导结果, 理论推导较好的反映了实

际分布情况; 同时也说明了距离的选择对度分布特性的影响较小.

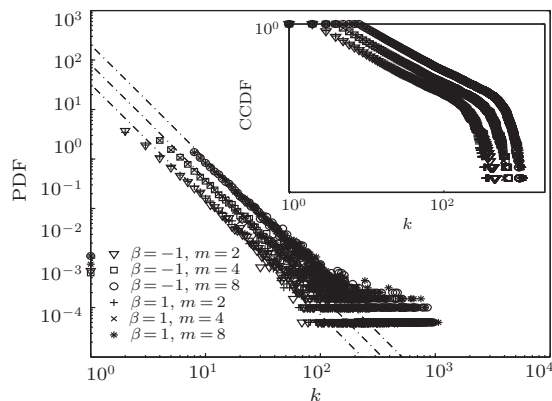


图 9 模型特殊情形 C 中度分布情况 ( $N = 10000, \alpha = 0, p = 0.5, L = 20, M = 20$ , 插图为互余累积度分布曲线)

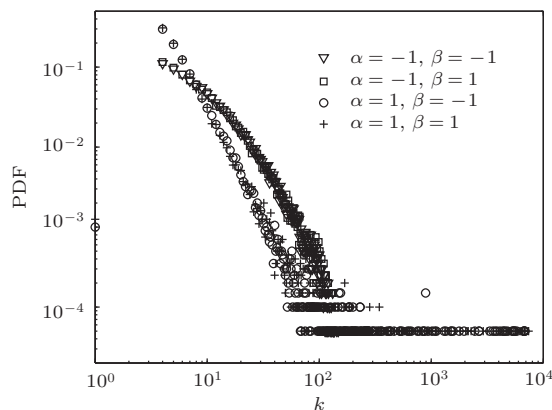


图 10 其他情形中度分布情况 ( $N = 10000, p = 0.5, L = 20, M = 20, m = 4$ )

其他情形的度分布如图 10 所示, 可以看出, 距离参数  $\beta$  对模型度分布并没有趋势性的影响; 模型在不同的度参数  $\alpha$  下呈现从指数分布到幂律分布的变化. 总之, 模型在  $\alpha = 0$  时, 可用漂移幂律分布刻画, 大多呈现幂律特性, 存在指数截断现象, 且  $L$  越大越明显; 在  $\alpha < 0$  时, 整体呈现幂律分布向指数分布变化的过程,  $\alpha$  的值较大时, 为幂律分布, 较小则呈现指数分布; 而  $\alpha > 0$  时, 模型具有广延指数分布的特点, 存在小变量饱和、指数截断现象, 且  $\alpha, L$  值越大, 越明显.

平均度也是与节点度密切相关的量化参数, 表 1 显示了  $m = 2$  时不同参数下模型的平均度 (不同网络规模基本相同, 此时为 10000), 可以看出连接方式对其影响较小, 网内连边概率越大, 平均度越高. 模型在一般参数下,  $\bar{k} \ll N$ , 明显具有实际网络的稀疏性, 也说明了稀疏网络的成因与网内节点相互作用存在一定的关联性.

表1 不同参数下模型的平均度 ( $m = 2$ )

$P$	$\alpha = 0; \beta = 0$	$\alpha = -1; \beta = 0$	$\alpha = 1; \beta = 0$	$\alpha = 0; \beta = -1$	$\alpha = 0; \beta = 1$	均值
0.3	5.68	5.69	5.66	5.65	5.68	5.67
0.5	7.96	8.01	8.00	7.98	7.99	7.99
0.7	13.09	13.22	13.20	13.11	13.17	13.16

## 4 其他重要参数分析

### 4.1 集聚系数

集聚系数 (clustering coefficient) 是用来刻画网络节点的邻居节点间相互连接的稠密程度, 反映了复杂网络内部集团化的程度. 网络平均集聚系数是所有节点集聚系数的平均值, 具体公式如下<sup>[39]</sup>:

$$C = \sum P(k) \cdot C(k), \quad (16)$$

式中,  $C(k)$  为局部集聚系数,  $C(k) = \langle E_k \rangle / (k(k-1)/2)$ , 其中  $\langle E_k \rangle$  为节点度为  $k$  的邻居节点中平均连接数. 统计研究表明, 大规模网络普遍具有明显的集聚特性, 多数网络的平均集聚系数远大于  $O(N^{-1})$ , 部分网络高达 0.8—0.9, 理论推测, 对于很多网络来说, 当网络规模  $N \rightarrow \infty$  时,  $C \rightarrow O(1)$ <sup>[38]</sup>. 本模型中, 不同的参数下, 平均集聚系数可以在较大范围内变化 (图 11), 对于节点度参数  $\alpha$ , 当  $\alpha > 0$  时, 随着网络规模的增大, 集聚系数逐渐增大, 在  $\alpha \geq 1$  时,  $n \rightarrow \infty$ , 存在  $C \rightarrow 1$  的趋势; 在  $\alpha = 0$  时, 呈现持续下降的趋势, 但依然远大于  $O(N^{-1})$ , 经过参数  $\beta = 1$  的调节可稳定在 0.2 左右; 而在  $\alpha < 0$  扶贫连接中, 随着网络规模的增大, 集聚系数便稳定在一定的范围内. 对于距离参数  $\beta$ , 在  $\beta > 0$  近邻连接时, 可以提高模型的整体集聚系

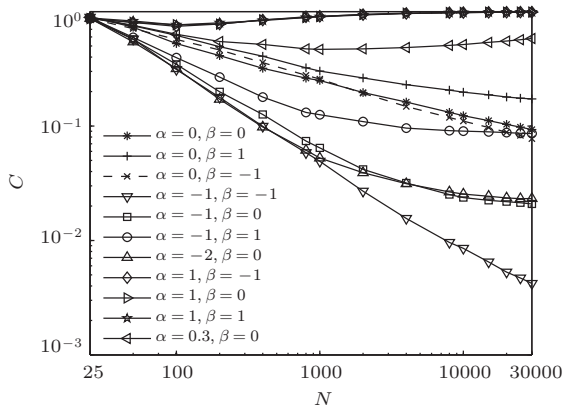


图 11 模型在不同  $\alpha, \beta$  下的平均集聚系数 ( $m_0 = 25$ ,  $e_0 = 250$ ,  $p = 0.5$ ,  $L = 20$ ,  $M = 20$ ,  $m = 4$ )

数, 极大的延缓集聚系数的下降速度; 在  $\beta < 0$  时, 则加快了集聚系数的下降速度. 不同于 BA 模型, 模型在大多参数设置下, 呈现较高的群集性, 且在规模较大时可保持相对稳定; 相比集聚系数取值范围有限的 RAPA 模型<sup>[24]</sup> 和一些集聚系数可调的演化模型<sup>[19]</sup>, 模型不仅可以通过度连接方式大幅度改变网络的集聚性, 而且可以在不改变度分布的情况下, 利用距离参数  $\beta$  对其进行一定程度的微调.

### 4.2 平均最短路径

现实网络普遍具有明显的小世界现象, 即网络具有较小的平均最短路径. 平均最短路径是任意两点间最短路径的平均值, 其计算公式为<sup>[39]</sup>

$$d = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}, \quad (17)$$

其中  $d_{ij}$  为节点  $i$  通过最短路径到达节点  $j$  经过的跳数. 图 12 是模型在不同  $\alpha, \beta$  下的平均最短路径, 而标注 “.” 的曲线为 BA 模型的理论值. 图中可以看出, 模型具有明显的小世界特性, 距离参数  $\beta$  对平均最短路径的影响较小, 而随着参数  $\alpha$  的增大, 平均最短路径变化越缓慢, 在  $\alpha \geq 1$  度优先连接时, 存在轻微减小的趋势, 这是由于部分节点出现一定程度的赢者通吃现象降低了网络的平均最短路径.

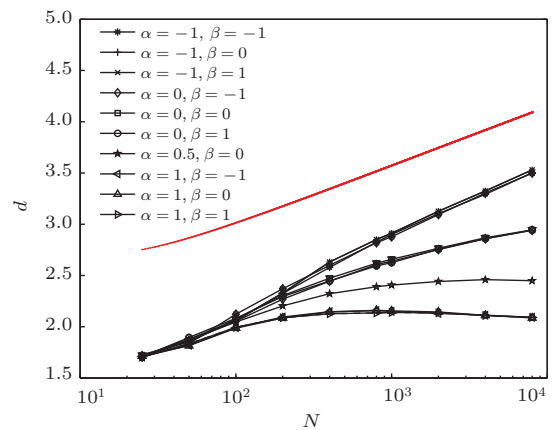


图 12 模型在不同  $\alpha, \beta$  下的平均最短路径 ( $p = 0.5$ ,  $L = 20$ ,  $M = 20$ ,  $m = 4$ )



### 4.3 度相关性

度相关性是描述网络内部不同节点间相互关系的一个重要指标, Newman 提出了一种通过计算关联系数  $r$  来量化网络的度相关性的方法<sup>[40]</sup>, 具体公式如下:

$$r = \left\{ M^{-1} \sum_i j_i k_i - \left[ M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2 \right\} \times \left\{ M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[ M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2 \right\}^{-1}, \quad (18)$$

式中,  $j_i, k_i$  分别表示与第  $i$  条边相连的节点  $j, k$  的度,  $M$  为网络的总边数.

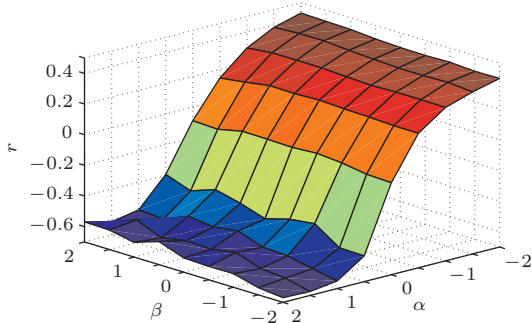


图 13 模型在不同  $\alpha, \beta$  下的度-度关联系数 ( $N = 10000, L = 20, M = 20, m = 4, p = 0.5$ )

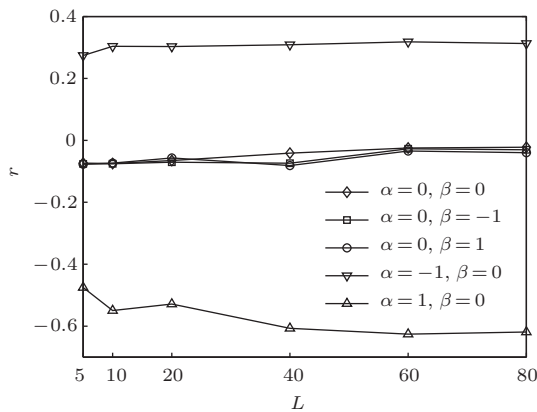


图 14 模型在不同  $L$  下的度-度关联系数 ( $N = 10000, M = 20, m = 4, p = 0.5$ )

图 13 中显示了不同  $\alpha, \beta$  下模型关联系数的变化情况, 其可以随着参数  $\alpha$  在  $(-0.6, 0.5)$  范围内变化, 而距离参数  $\beta$  对其影响较小. 当  $\alpha < 0$  扶贫连接时, 网络表现为同配特性; 相反, 在  $\alpha \geq 0$  度优先连接时, 表现为异配特性. 现实网络的度分布大多

具有幂律特性, 同时关联系数在  $(-0.4, 0.4)$  范围内变化<sup>[38]</sup>, 而模型  $\alpha < 0$  且  $L = 20$  时, 度分布多呈现指数分布的倾向, 与实际网络存在偏差, 针对这一现象, 本文进一步研究了信息传播长度  $L$  对关联系数的影响如图 14 所示,  $\alpha < 0$  时随着  $L$  的增大关联系数变化不大, 由度分布分析知在  $\alpha < 0$  且  $L$  较小时, 模型呈现幂律分布. 因此, 通过参数  $\alpha, L$  的调节, 模型可以在维持幂律分布特性的情况下, 产生从同配到异配具有不同关联模式的网络拓扑.

### 5 结 论

信息传播过程是现实网络中时时刻刻都在发生的行为, 网络的结构能够对信息传播过程造成巨大的影响, 同样, 信息传播过程也为网络内部结构的演化提供了重要的局域信息. 本文为了研究信息传播过程对网络演化的影响, 提出了一种信息传播促进网络增长的复杂网络演化模型. 新节点通过局域世界建立第一条边, 而后利用自回避随机游走模拟网络信息传播过程, 通过节点度和距离的综合考虑, 新节点以及内部节点分别与信息路径上的点选择性建立连接, 其中包括随机连接、度优先连接、扶贫连接、近邻连接、远交连接以及互相结合的连接方式. 理论计算和仿真实验表明, 模型不同于 BA 及其扩展模型, 不仅具有小世界特性、无标度特性, 不同情况下度分布还具有漂移幂律分布、广延指数分布的特点, 呈现出现实网络中的非幂律特性如: 指数截断、小变量饱和等现象, 同时模型演化的网络具有较高的集聚性, 且可以在不改变度分布的情况下, 较大范围内调节网络的集聚系数, 模型根据参数变化, 在保持幂律特性的基础上, 度度关联系数可在  $(-0.6, 0.5)$  范围内变化. 模型在完全随机选择的情况下, 产生了  $\lambda$  在  $(2, 3)$  范围内变化的幂律分布, 这说明了信息传播路径上的连边机理也可能是幂律分布产生的另一种机理, 对于幂律分布产生机理的研究具有一定的意义. 此外, 本文提出的模型从局域信息来源的角度出发, 不仅兼顾了多种网络连接的特点, 而且能够展现现实网络中重要的拓扑特性, 可以为复杂网络演化中信息传播对网络结构的影响提供一种研究思路, 其基本机理同样适用于现实中的含权、有向网络, 尤其有利于一些社会网络、信息传输网络 (如, Internet、通信网) 演化机理的研究.

参考文献

- [1] Erdős P, Rényi A 1960 *Publ. Math. Inst. Hungar. Acad. Sci.* **5** 17
- [2] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [3] Barabási A L, Albert R 1999 *Science* **286** 509
- [4] Fang J Q, Wang X F, Zheng Z G, Bi Q, Di Z R, Li X 2007 *Prog. Phys.* **27** 239 (in Chinese) [方锦清, 汪小帆, 郑志刚, 毕桥, 狄增如, 李翔 2007 物理学进展 **27** 239]
- [5] Cui A X, Fu Y, Shang M S, Chen D B, Zhou T 2011 *Acta Phys. Sin.* **60** 38901 (in Chinese) [崔爱香, 傅彦, 尚明生, 陈端兵, 周涛 2011 物理学报 **60** 38901]
- [6] Li M, Zou H, Guan S, Gong X, Li K, Di Z, Lai C H 2013 *Sci. Rep.* **3** 2512
- [7] Jeong H, Mason S P, Barabási A L, Oltvai Z N 2001 *Nature* **1** 41
- [8] Amaral L A N, Scala A, Barthelemy M, Stanley H E 2000 *Proc. Natl. Acad. Sci. USA* **97** 11149
- [9] Li S M, Xu X H, Meng L H 2012 *Chin. Phys. B* **21** 088901
- [10] Zhou T, Han X P, Yan X Y, Yang Z M, Zhao Z D, Wang B H 2013 *Journal of University of Electronic Science and Technology of China* **42** 482 (in Chinese) [周涛, 韩筱璞, 闫小勇, 杨紫陌, 赵志丹, 汪秉宏 2013 电子科技大学学报 **42** 482]
- [11] Zhao Z D, Yang Z, Zhang Z, Zhou T, Huang Z G, Lai Y C 2013 *Sci. Rep.* **3** 3472
- [12] Lü L, Zhou T 2011 *Physica A* **390** 1150
- [13] Wang B H, Zhou T, Zhou C S 2012 *Journal of University of Shanghai for Science and Technology* **34** 103 (in Chinese) [汪秉宏, 周涛, 周昌松 2012 上海理工大学学报 **34** 103]
- [14] Yang C L, Tang K S 2011 *Chin. Phys. B* **20** 128901
- [15] Chen G R 2013 *Acta Autom. Sin.* **39** 312 (in Chinese) [陈关荣 2013 自动化学报 **39** 312]
- [16] Liu Y Y, Slotine J J, Barabási A L 2013 *Proc. Natl. Acad. Sci. USA* **110** 2460
- [17] Albert R, Barabási A L 2000 *Phys. Rev. Lett.* **85** 5234
- [18] Bianconi G, Barabási A L 2001 *Phys. Rev. Lett.* **86** 5632
- [19] Holme P, Kim B J 2002 *Phys. Rev.* **65** 026107
- [20] Li X, Chen G 2003 *Physica A* **328** 274
- [21] Wang G Z, Cao Y J, Bao Z J, Han Z X 2009 *Acta Phys. Sin.* **58** 3597 (in Chinese) [王光增, 曹一家, 包哲静, 韩祯祯 2009 物理学报 **58** 3597]
- [22] Yuan S Q, Zhao H, Li C, Zhang X 2008 *Acta Phys. Sin.* **57** 4805 (in Chinese) [袁韶谦, 赵海, 李超, 张昕 2008 物理学报 **57** 4805]
- [23] Wang L N, Guo J L, Yang H X, Zhou T 2009 *Physica A* **388** 1713
- [24] Jiang Z H, Wang H, Gao C 2011 *Acta Phys. Sin.* **60** 58903 (in Chinese) [姜志宏, 王晖, 高超 2011 物理学报 **60** 58903]
- [25] Saramäki J, Kaski K 2004 *Physica A* **341** 80
- [26] Zhu C P, Zhou T, Yang H J, Xiong S J, Gu Z M, Shi D N, He D R, Wang B H 2008 *New J. Phys.* **10** 023006
- [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, 何大初 2006 物理学报 **55** 60]
- [28] Wang X F, Chen G R 2006 *Complex Networks Theory and its Applications* (Beijing: Tsinghua University Press) (in Chinese) [汪小帆, 李翔, 陈关荣 2006 复杂网络理论及其应用 (北京: 清华大学出版社)]
- [29] Chang H, Su B B, Zhou Y P, He D R 2007 *Physica A* **383** 687
- [30] Bornholdt S, Schuster H G 2003 *Handbook of graphs and networks* (New York: Wiley-Vch) pp318
- [31] Noh J D, Rieger H 2004 *Phys. Rev. Lett.* **92** 118701
- [32] Herrero C P, Saboyá M 2003 *Phys. Rev. E* **68** 026106
- [33] Bar S, Gonen M, Wool A 2004 *An incremental super-linear preferential Internet topology model* (Berlin: Springer-Heidelberg) pp53–62
- [34] Li Y, Fang J Q, Liu Q 2007 *Science & Technology Review* **25** 23 (in Chinese) [李永, 方锦清, 刘强 2007 科技导报 **25** 23]
- [35] Jost J, Joy M P 2002 *Phys. Rev. E* **66** 036126
- [36] Cao Y J, Wang G Z, Jiang Q Y, Han Z X 2006 *Phys. Lett. A* **349** 462
- [37] Yang X H, Lou S L, Chen G, Chen S Y, Huang W 2013 *Physica A* **392** 3531
- [38] Newman M E J 2003 *SIAM review* **45** 167
- [39] Wang X F, Chen G 2003 *IEEE Circuits and Systems Magazine* **3** 6
- [40] Newman M E J 2003 *Phys. Rev. Lett.* **89** 208701

# A complex network evolution model for network growth promoted by information transmission\*

Liu Shu-Xin<sup>1)†</sup> Ji Xin-Sheng<sup>1)</sup> Liu Cai-Xia<sup>1)</sup> Guo Hong<sup>2)</sup>

1) (National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China)

2) (Institute of Information System Engineering, Information Engineering University of PLA, Zhengzhou 450002, China)

( Received 23 January 2014; revised manuscript received 9 April 2014 )

## Abstract

In many real complex networks, information transmission occurs all the time. To study the effects of information transmission on the complex network evolution, we propose a new model for network growth promoted by the information transmission. The model includes three major steps: (i) New links attached to the nodes on the information transmission path, whose source point is chosen preferentially; (ii) the first link of the new node attached to the nodes in the local-world; (iii) other links of the new node attached to the nodes on the information transmission path, whose source point is the new node. The process of information transmission is simulated by self-avoiding random walk, and by considering the local information including its degree and distance; selective connection is established between the nodes on the information transmission path. Theoretical analysis and numerical simulation results show that the proposed model can not only reproduce small-world and scale-free network characteristics, but also indicate that “shift power-law distribution” and “truncated power law” function may form for different parameters which have some non-power-law features, such as exponential cutoff, and saturation for small variables. Moreover, in our model, the clustering coefficient is tunable without changing the degree distribution, and the model can also construct a network with assortative or disassortative mixed pattern.

**Keywords:** complex network, information transmission, evolving model, self-avoid random walk

**PACS:** 89.75.-k, 89.75.Fb

**DOI:** [10.7498/aps.63.158902](https://doi.org/10.7498/aps.63.158902)

---

\* Project supported by the National High Technology Research and Development Program of China (Grant Nos. 2011AA010605, 2011AA010604).

† Corresponding author. E-mail: [liushuxin11@gmail.com](mailto:liushuxin11@gmail.com)