

一种基于随机行走和策略连接的网络演化模型*

姜志宏[†] 王晖 高超

(国防科技大学信息系统与管理学院系统工程系, 长沙 410073)

(2010年5月28日收到; 2010年8月13日收到修改稿)

本文提出了一个基于随机行走和策略选择的复杂网络局域演化模型 RAPA. 新节点加入系统不需要全局知识, 而是通过随机行走构造局域世界; 然后依据概率采用随机连接, “扶贫”连接或“亲富”连接策略, 从局域世界中选择节点增加连接边; 最终自组织演化具有幂律特点的复杂网络. 初步的解析计算和仿真实验都表明, RAPA 模型不仅重现了具有小世界特性、整体上的无标度特性, 还可以演化出小变量饱和以及指数截断等现象, 同时也具有明显的聚类特性, 并能够构造出同配或异配等不同混合模式的网络.

关键词: 复杂网络, 模型, 随机行走, 策略连接

PACS: 89.75.-k, 89.75.Fb

1. 引言

复杂网络是复杂系统的高度抽象, 是对复杂系统元素之间关系的描述. 复杂网络的研究对象充满着自然界、工程界和社会界, 如细胞新陈代谢网络、蛋白质网络、互联网、万维网、社会关系网络、科研协作网络以及电力网络等等. 人们对描述真实系统拓扑结构的研究, 从最早仅采用规则结构表示, 到 20 世纪开始采用随机图的思想来描述, 称之为随机网络. 到了本世纪初, 研究人员们发现大量的真实网络并不是类似于随机网络的均质网络, 而是异质的. 复杂网络的度分布不是服从 Poisson 分布, 而是更多地是服从幂律分布, 这类度分布服从幂律分布的网络称为无标度网络^[1-3].

幂律的发现解释了复杂网络的许多涌现特征, 使人们对现实网络有了更进一步的认识, 但是节点度分布仍然仅仅是复杂网络特征属性的一个方面. 文献[4-6]的研究表明满足同样幂律分布的网络可以呈现出完全不同的拓扑结构. 除度分布特征以外, 真实网络往往都呈现很强的关联性, 而且不同网络也通常呈现出不同的关联特性. 文献[2, 7, 8]研究表明科研协作网络等社会网络具有正的关联系数, 而蛋白质网络, 互联网等生物网络和技术

网络具有负的关联系数. 而经典随机网络模型、WS 模型^[9]、BA 模型^[10]及大量的 BA 扩展模型^[2, 10]等并没有很好地模拟和解释出实际网络的高聚类、节点度关联性等特征. 研究和建立能够同时重现真实网络多种拓扑特征的网络局域演化模型, 有助于更好地理解 and 描述真实网络, 对于更进一步了解复杂网络的拓扑结构及其形成机理、研究复杂网络传播动力学和网络的鲁棒性都具有重要意义.

2. 相关研究及本文工作

近年来, 人们在各种各样的实际网络中观察到了幂律分布. 这些网络涵盖了从社会关系网络到人造信息网络的广大领域. 在这些网络中, 一个节点的度数为 k 的概率 $p(k) \propto ck^{-r}$, 其中 c 和 r 为常数. 文献[10]提出的 BA 模型, 是复杂网络研究上的一个里程碑. BA 模型采用网络增长和择优连接两个因素解释了幂律的成因, BA 模型中的节点度服从指数为 3 的幂次分布. BA 模型的生成算法利用了网络的全局信息, 即假设每个新加入节点都具备全局知识, 然后根据这些信息进行择优连接. 然而在真实网络, 例如社会网络和万维网等现实网络中, 新节点通常是在不完全信息情况下, 确定增长网络的连接机理, 最终自组织演化成为无尺度网络, 这

* 国家高技术研究发展计划(863 计划)(批准号:2008AA01Z407)和国家自然科学基金(批准号:60872053)资助的课题.

[†] E-mail: jonathan.jiang@263.net

就使得类 BA 模型的应用受到了很大限制. 文献 [11] 给出了通过简单随机行走 (random walk) 形成无尺度网络的方法, 并证明了该模型和 BA 模型等价. 文献 [12] 修改了 BA 模型的择优连接方式, 采用随机选择边的端节点与新增节点之间增加连接边的方式构造幂律网络. 这两个模型都是在不完全信息下, 通过新的连接机制, 构造新的无尺度演化网络模, 重现了和 BA 模型等价的度分布特征. 单数测量计算得到的现实中各种复杂网络并且常常具有一些非幂律特征, 如指数截断^[13] (exponential cutoffs) 或陡尾 (steep tail), 小变量饱和 (saturation for small variables) 或“平头” (flat head) 等. 文献 [14] 认为在现实网络中, 由于局域连接性的存在, 每个节点都有各自的局域世界, 因而也只占有和使用整个网络的局部连接信息, 因此需要在择优连接过程中引入了局域世界的概念. 即择优连接不是从全局节点中选择, 而是从随机选取的一个节点子集中择优, 并指出基于局域世界的演化模型, 其度分布可在指数分布和幂次分布之间演化. 文献 [15, 16] 等对局域世界的构造方法进行了扩展, 但是这些局域世界演化模型中使用的局域世界都是在演化步中随机选择若干节点构成的, 显然忽略了局域世界内部节点之间的关系和局域世界的形成原因, 因此模型的局域演化仍然缺少足够的实际意义.

很多网络中通常包含不只一种类型的节点, 例如在性关系网络中, 可以按照民族、人种等对节点进行分类, 节点间存在边相连接的概率常常依赖于节点的类型, 这种依赖关系称之为混合模式 (mixing pattern). 而以节点度作为节点分类的依据则具有更加普遍意义. Newman^[5, 7] 的研究表明, 在某些实际网络中, 高度节点倾向于与其他高度节点连接, 称之为异配 (disassortative mixing), 如信息网络和神经网络; 而在另外一些网络中, 高度节点则倾向于和低度节点连接, 称之为同配 (assortative mixing), 如众多的社会关系网络, 如论文合作网络和演员合作网络. 混合模式可通过即计算节点度-度的相关系数 (degree-degree correlation coefficient, DCC) 来进行量化描述, 目前已有研究表明大部分真实网络的度-度相关系数位于 $(-0.4, 0.4)$ 之间^[2]. 文献 [17] 提出了相互吸引模型, 对于每个节点都赋予一个初始的吸引力强度值, 并将该强度值引入择优过程, 从而获得的模型, 其度-度相关系数可以在 $(-0.6, 0.3)$ 之间转变; 文献 [18] 提出的大统一的

混合网络模型 (LUHNM), 认为现实网络中的连接选择方式是随机性和确定性的混合, 而确定性选择既可能“亲富”, 又可能“扶贫”, 也可能采取“折中”或者“平衡”, 因此扩充 BA 择优连接机理为混合方式连接, LUHNM 的度-度相关系数可以在 $(-1, 1)$ 之间大范围内转变. 但是这两个模型仍然需要新加入节点掌握全局信息, 用于计算新增边的概率.

本文提出了一种基于随机行走和策略连接 (politic attachment) 的网络局域演化模型 RAPA, 新增节点根据随机行走得到的局域世界, 并依据策略从中选择节点增加连接边. 模型演化的无向网络最终呈现出明显的小世界、幂律以及高聚类特征. 同时, 通过调节模型参数, 演化网络还可以重现出不同的混合模式, 其度-度相关系数在 $(-0.9, 0.5)$ 较大范围内变化的.

3. 基于随机游走的复杂网络模型 RAPA

在真实网络中, 新加入节点增加连接边不可能采用 BA 模型中全局择优的方法, 而更多是通过随机行走和选择的方式选择邻居节点. 例如万维网络中, 一个新节点可能以一些随机节点为起点, 通过页面超链进行浏览漫游, 并将漫游中遇到的一些节点其作为自己的新邻居, 建立自己到这些节点的超链. 人际关系网络中, 新加入的个体可能通过朋友的介绍认识新的朋友, 从而建立新的人际关系连接. 论文引用网络中, 论文作者也可能通过已有论文的引用关系进行检索, 并从中选择作为正在撰写的论文的引用文献. 因此可见, 随机行走的连接构造方式比全局择优方式更具现实意义.

简单随机游走模型^[11] 已经被证明和 BA 模型具有等价关系, 但是其度-度关联性仍然趋近于零, 即生成网络不具备明显的关联性. 真实网络中的随机行走节点, 并不是向行走过程中发现的所有节点都增加连接边, 而是有选择地建立连接关系. 选择可能是随机的, 或者是按照某种确定性策略. 文献 [18] 认为不论是社会网络, 还是物理网络、生物网络和技术网络, 都包含随机性和确定性两种混合过程, 只是混合的程度不同. 比如在真实网络中通常存在“扶危济贫”和“亲富疏贫”两种确定性选择的典型情况. 这些确定性选择的存在, 可能是导致真实网络复杂表现的重要原因.

本文提出的 RAPA 演化模型在随机选择的基础上,引入了“扶贫”和“亲富”两种确定性的节点连接策略. 该模型包括三个重要特征:

增长:与 BA 模型类似,即网络的规模是不断扩大的.

随机行走:新加入节点在网络中进行一轮随机行走构造局域世界.

策略连接:新节点在局域世界中选择节点相连接. 选择节点的策略分为随机、“扶贫”(选择度最低的节点)和“亲富”(选择度最高的节点)三种.

模型演化算法描述如下:

1) 初始时刻 $t=0$ 时,假定系统仅包含一个由每个节点组成的全连通网络.

2) 系统增加一个新节点 A .

3) 在系统中随机选择一个节点作为行走开始节点,进行一个长度为 S 的随机行走,行走路径上的全部节点构成局域世界 $V_{\text{localworld}}$.

4) 然后以概率 p 从 $V_{\text{localworld}}$ 中随机选择 $m(0 < m \ll S)$ 个不同节点,增加这些节点到节点 A 的连接边;以概率 q 从 $V_{\text{localworld}}$ 中选择度最小的 m 个不同节点,增加到节点 A 的连接边,即所谓“扶贫”;以概率 $1-p-q$ 从 $V_{\text{localworld}}$ 上选择度最大的 m 个不同节点,增加到节点 A 的连接边,即所谓“亲富”.

5) 返回 2),直到网络达到所需要的大小为止.

在 S 和 m 确定的情况下,模型中包含两个可调参数 p 和 $q, p(0 \leq p \leq 1)$ 表示随机性选择邻居节点的概率, $q(0 \leq q \leq 1-p)$ 表示采用“扶贫”策略的概率,则 $1-p-q$ 表示采用“亲富”策略的概率. 同时,为了便于算法分析,对参数 q 进行归一化处理,引入新参数 $q_n = q/(1-p)$.

4. 节点度分布

本文提出的 RAPA 模型,在 t 时刻,新节点加入网络 $G_t(E_t, V_t)$,并从 V_t 中随机选择节点启动一个步长为 S 的随机行走,得到行走路径上的无重复节点序列记为 Path_i^S . 为了简化计算分析,我们假设随机行走采用无重复的 self-void 方式,因此局域世界中 $|\text{Path}_i^S| = |V_{\text{localworld}}| = S$. 然后从 Path_i^S 中按照不同策略挑选 m 个节点作为新节点的邻居. 由文献 [11] 可知,从随机节点上经过 S 步随机行走,最终到达节点 i 的概率为 $\frac{k_i}{\sum k}$. 记 $P(A_i^S)$ 表示网络中任意

节点 i 出现在随机行走路径 Path_i^S 中的概率,有

$$P(A_i^S) = 1 - \left(1 - \frac{k_i}{\sum k}\right)^S$$

$$= 1 - \left(1 - \frac{k_i}{m(m-1) + 2mt}\right)^S. \quad (1)$$

当 $t \rightarrow \infty$ 时可得

$$P(A_i^S) \approx 1 - \left(1 - \frac{k_i}{2mt}\right)^S. \quad (2)$$

根据平均场方法 [19],可假设节点 i 的度 $k_i(t)$ 是连续变化的. 采用随机策略从 Path_i^S 中选择邻居节点时,节点 i 的度 k_i 随时间的变化率为

$$\left(\frac{\partial k_i}{\partial t}\right)_{\text{Ra}} = P(A_i^S) \times \frac{m}{S}$$

$$= \left(1 - \left(1 - \frac{k_i}{2mt}\right)^S\right) \times \frac{m}{S}. \quad (3)$$

记 $P(A_{ij}^L)$ 表示从节点 i 开始,经过一个步长为 L 的随机行走,到达节点 j 的概率,文献 [20] 已经证明,当 L 较大时,有 $P(A_{ij}^L) \approx \frac{k_j}{\sum k}$. 因而经过步长为 L 的随机行走,从节点 i 到达任意节点 j 且同时满足 $k_j \leq k_i$ 的概率为

$$P(A_{ij, k_j \leq k_i}^L) = \sum_{k_j=0}^{k_i} \left(\frac{k_j}{\sum k} \times P(k = k_j) \times |V_i|\right)$$

$$= \sum_{k_j=0}^{k_i} \left(\frac{k_j}{2mt + m(m-1)} \times P(k) \times (m+t)\right)$$

$$\approx \frac{1}{2m} \sum_{k=0}^{k_i} kP(k). \quad (4)$$

因此采用“亲富”策略选择邻居节点,当 $t \rightarrow \infty$ 且 S 充分大时, k_i 随时间的变化率为

$$\left(\frac{\partial k_i}{\partial t}\right)_{\text{Ma}} = P(A_i^S) \times \sum_{r=S-m}^{S-1} \binom{S-1}{r}$$

$$\times P(A_{ij, k_j \leq k_i}^L)^r \times (1 - P(A_{ij, k_j \leq k_i}^L))^{S-1-r}$$

$$= \left(1 - \left(1 - \frac{k_i}{2mt}\right)^S\right)$$

$$\times \left(1 - F_{Bi}\left(S-m; S-1, \sum_{k=0}^{k_i} \frac{kP(k)}{2m}\right)\right)$$

$$= \left(1 - \left(1 - \frac{k_i}{2mt}\right)^S\right)$$

$$\times I_{\sum_{k=0}^{k_i} \frac{kP(k)}{2m}}(S-m+1, m-1)$$

$$\approx \left(1 - e^{-\frac{k_i}{2mt} \times S}\right) \times I_{\sum_{k=0}^{k_i} \frac{kP(k)}{2m}}$$

$$\times (S-m+1, m-1), \quad (5)$$

其中 $P(k)$ 为节点度分布函数, $F_{Bi}(k; n, p)$ 表示参数为 (N, p) 的累积二项分布函数, $I_x(a, b)$ 表示正则化不完全 Beta 函数 (regularized incomplete beta function) [21].

同样可知采用扶贫策略选择邻居节点时, 节点 i 的度 k_i 随时间的变化率为

$$\left(\frac{\partial k_i}{\partial t}\right)_{Mi} \approx (1 - e^{-\frac{S}{2m^2 k_i}}) \times I_{1 - \sum_{k=0}^{k_i-1} \frac{kP(k)}{2m}} \times (S - m + 1, m - 1). \quad (6)$$

RAPA 模型综合考虑上述三种策略, 其演化模型中 k_i 满足动力学方程

$$\frac{\partial k_i}{\partial t} = p \left(\frac{\partial k_i}{\partial t}\right)_{Ra} + (1 - p) q_n \left(\frac{\partial k_i}{\partial t}\right)_{Mi} + (1 - q_n)(1 - p) \left(\frac{\partial k_i}{\partial t}\right)_{Ma}. \quad (7)$$

显然, RAPA 模型有几个特殊情形, 下面分别进行讨论.

特殊情形 A $p = 1$

此时, 仅仅采用随机策略添加新节点的连接边, 方程(7)退化成为方程(3)的形式. 对于大多数的中低度节点 ($k_i \sim m$) 有

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \sum_{j=0}^{s-1} \frac{1}{S} \left(1 - \frac{k_i}{2mt}\right)^j \approx \frac{k_i}{2t} \sum_{j=0}^{s-1} \frac{1}{S} \left(1 - \frac{1}{2t}\right)^j \approx \frac{k_i}{2t}. \quad (8)$$

此时 RAPA 模型等价于无标度网络模型 [19], 即在中低度节点上的度分布服从幂率分布 $P(k) \propto k^{-3}$.

对于高度节点 ($k_i \sim t$) 当 $t \rightarrow \infty$ 时, 有

$$\frac{\partial k_i}{\partial t} \approx \frac{m}{S} \times \left(1 - \left(1 - \frac{1}{2m}\right)\right)^S. \quad (9)$$

特别的如果 $S = m + t$ 有

$$\frac{\partial k_i}{\partial t} \approx \frac{m}{S} = \frac{m}{m + t}. \quad (10)$$

此时随机行走路径几乎覆盖整个网络, 随机行走经过节点 i 的概率接近 1, 这意味着为新加入节点选择邻居时近似于采用全网随机选择的方式, 根据文献 [19] 可以证明此时的 RAPA 模型的度分布趋近于指数分布 $P(k) \propto e^{-\frac{k}{m}}$.

可知 $p = 1, q = 0$ 时, RAPA 模型的度分布显示出由幂次分布向指数分布演化的特点. $S < m + t$ 时, 在大多数的中低度节点上表现出幂次分布的无标度性, 而在尾部的高度节点上表现出指数分布曲

线的特征, 形成指数截断 (exponential cutoffs) 现象 [13]. 类似的度分布特征可见于蛋白质网络 [22]、电子邮件网络 [23] 以及演员合作网络 [24] 等真实网络中. RAPA 演化模型的特殊情形 A 仿真结果实验如图 1 所示, 其中实线表示对度分布函数进行幂次拟合结果, 拟合的幂次系数为 2.87—2.93, 与 BA 无标度模型一致. 从图 1 插图的互余累积度分布函数还可以看出, 对于高度节点, 即在曲线尾部出现了向指数分布偏离的倾向, 证实了解析计算的分析结果. 同时, 还看出参数 m 和 S 都与标度指数无关, 而参数 S 对指数截断的程度产生影响. 当随机行走步长 S 越大, 指数截断现象越为明显.

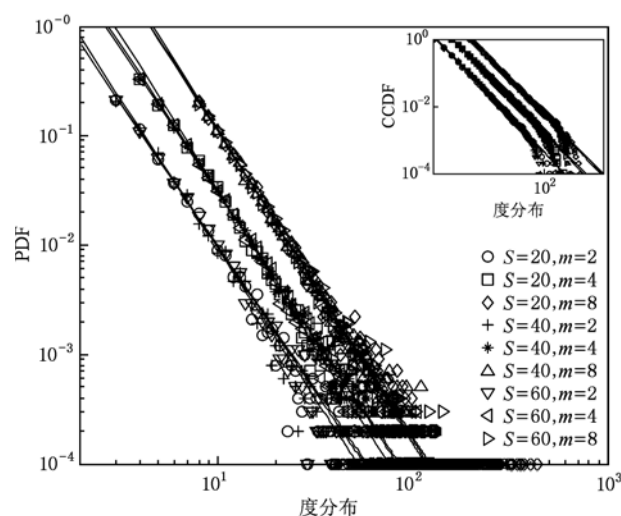


图 1 RAPA 演化模型的度分布函数特殊情形 A ($p = 0, N = 10000, S = 20, 40, 60, m = 2, 4, 8$, 插图为互余累积度分布曲线)

特殊情形 B $p = 0, q_n = 0$

此时, 仅仅采用亲富策略添加新节点的连接边, 方程(7)退化成为方程(5)的形式:

$$\begin{aligned} \frac{\partial k_i}{\partial t} &= \left(\frac{\partial k_i}{\partial t}\right)_{Ma} \\ &= (1 - e^{-\frac{k_i}{2m^2 \times S}}) \times I_{\sum_{k=0}^{k_i} \frac{kP(k)}{2m}} (S - m + 1, m - 1). \end{aligned} \quad (11)$$

虽然通过 (12) 式解析计算度分布函数比较困难, 但是我们仍然可以通过定性分析推导度分布函数的性质. 易知 $0 < (1 - e^{-\frac{k_i}{2m^2 \times S}}) < 1$, 当 k_i 较大即 $k_i \sim m + t$ 时 $(1 - e^{-\frac{k_i}{2m^2 \times S}}) \rightarrow 1$. 同时 $0 \leq \sum_{k=0}^{k_i} \frac{kP(k)}{2m} \leq 1$ 且 $\sum_{k=0}^{k_i} \frac{kP(k)}{2m}$ 是关于 k_i 的增函数, 因为 $S - m$

+1 ≫ m - 1, 由正则化不完整 Beta 函数^[21]的性质可知在任意 t 时刻, k_i 的变化率关于 k_i 的曲线呈现“J”形, 即当取值于最大节点度附近时有 $\frac{\partial k_i}{\partial t} \rightarrow 1$, 而在小于最大节点度时的变化率迅速降低为 0. 显然此时的演化网络表现出明显的星型网络特点, 即极少数节点的度趋近于 m + t, 同时大部分节点的度保持为 m. RAPA 演化模型特殊情形 B 仿真结果实验如图 2 所示, 说明在 N = 10000 的网络规模下, 有极少数节点度接近 10000, 而其余所有节点的度都保持在 m 附近. 参数 m 的变化仅改变平均度而不会对分布曲线的形状产生大的影响. 当参数 S 越大, 即 S ≫ m 越明显时, 度分布曲线上甚至仅包含最高度和最低度两个点, 此时 RAPA 网络几乎完全退化成星型网络.

特殊情形 C p = 0, q_n = 1

此时, 仅仅采用扶贫策略添加新节点的连接边, 方程(7)退化成为方程(6)的形式:

$$\begin{aligned} \frac{\partial k_i}{\partial t} &= \left(\frac{\partial k_i}{\partial t} \right)_{Mi} \\ &= (1 - e^{-\frac{S}{2m}k_i}) \times I_{1-\sum_{k=0}^{k_i} \frac{kP(k)}{2m}} (S - m + 1, m - 1). \end{aligned} \quad (12)$$

同样进行定性分析, 由函数(1 - e^{- $\frac{S}{2m}k_i$})以及正则化不完整 Beta 函数^[21]的性质可知在任意 t 时刻, 在上述方程, k_i 的变化率在最小节点度附近时取得最大值, 而在节点度 k_i 偏离 m 时其变化率迅速降低为 0. 当 t → ∞ 时, 相当部分节点的度趋近于网络平均度, 同时高度节点的概率急速下降. 显然可知此时的演化网络表现出均匀网络倾向, 网络度分布具有一定程度的正态性. 特殊情形 C 下 RAPA 模型的度分布函数仿真结果如图 3 所示, 其中插图为 RAPA 互余累积度分布函数. 仿真结果与我们的解析分析是一致的, 而且还可以看出参数 m 仅决定了度分布概率密度曲线的峰值位置, 即网络的平均度, 同时当参数 S 越大, 即 S ≫ m 越明显时, 峰值位置越接近平均度 2m.

从 RAPA 模型的几种特殊情形分析来看, 当满足 m ≪ S ≪ m + t 时, 参数 S 和 m 并不对模型的度分布特征产生大的影响, 可见此时行走路径长度 S 和节点初始度 m 不是模型的演化的关键因素. 为了简化仿真过程同时又不失典型性, 我们在后续实验中取经验值 S = 40, m = 4, 重点讨论参数 p, q 对模型演化的影响. 当 0 < p < 1, 0 < q < 1 时, RAPA 模型表现

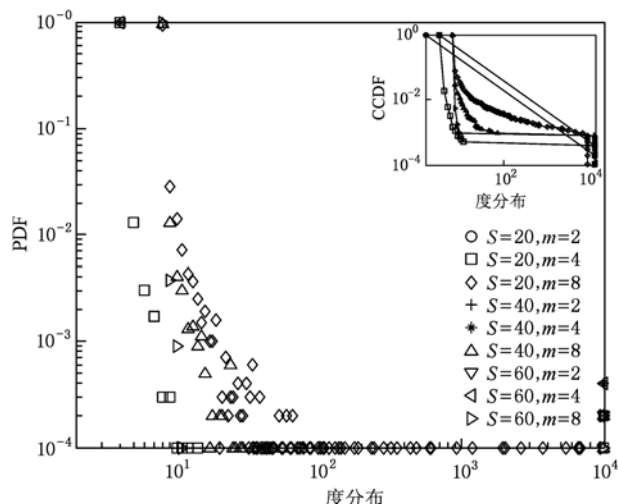


图2 RAPA 演化模型的度分布函数特殊情形 B (p = 0, q_n = 0, N = 10000, S = 20, 40, 60, m = 2, 4, 8, 插图为互余累积度分布曲线)

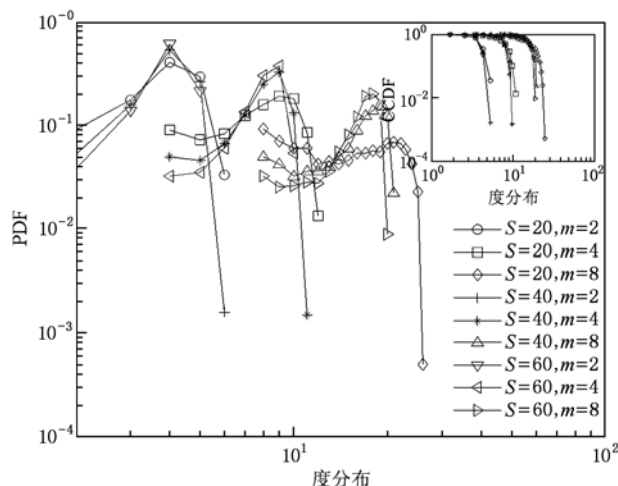


图3 RAPA 演化模型的度分布函数特殊情形 C (p = 0, q_n = 1, N = 10000, S = 20, 40, 60, m = 2, 4, 8, 插图为互余累积度分布曲线)

为上述三种特殊情形的混合. 图 4 是此时 RAPA 模型的度分布函数仿真结果. 当 p → 1 时, 随机选择策略作为优势策略, 度分布曲线表现为幂次曲线, 幂指数 λ ≈ 3, 但从插图的互余累积度分布曲线看, 在高度节点范围存在一定的指数截断现象. 当 0 ≤ p < 1 时, 度分布在整体上仍然呈现出幂次特点, 而且 p 值越小幂指数越大. 当 q_n → 1 时, 扶贫策略逐渐成为确定性优势策略, 使得度分布在平均度 2m 附近呈现峰值 (在 p 较大时峰值不明显, 度分布形成“平头”), 而在更大的度上接近幂次分布, 其幂指数 λ 随 p 减小而增大, 说明此时网络度分布较幂次网络更为均匀, 而且在一定范围内具有正态性特点. 当 p < 1 且 q_n → 0 时, 亲富策略逐渐成为确定性优势策

略,加重了了度分布曲线的长尾特征,使得极少数节点拥有较大的节点度,而在中低度节点范围内仍然接近幂次分布,其幂指数 λ 仍然随 p 减小而增大,说明此时网络演化具有星型网络的特点.

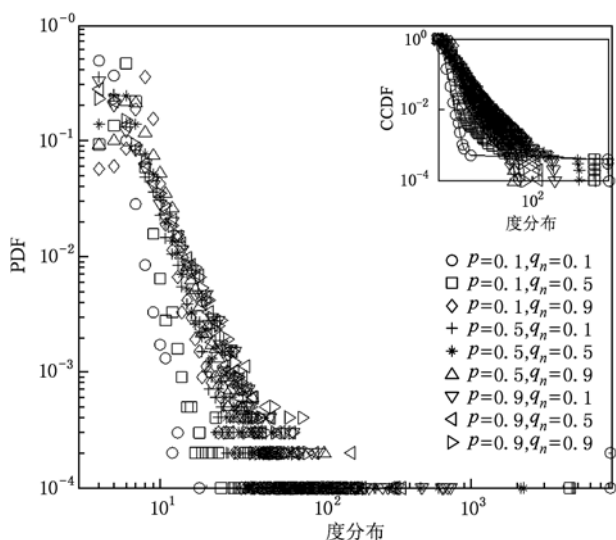


图4 RAPA 模型在不同 p, q 参数下的度分布 ($N = 10000, S = 40, m = 4$, 插图为互余累积度分布曲线)

仿真结果和初步理论解析都说明,参数 p 决定了度分布整体上的幂次指数,而参数 q_n 则影响度分布曲线的头部和尾部形状,当 q_n 较大时,扶贫策略使得度分布曲线出现“平头”现象,尤其是当 $q_n \rightarrow 1$ 时,度分布曲线更在平均度附近出现单峰. 反映在互余累积度分布曲线中,即形成小变量饱和现象.

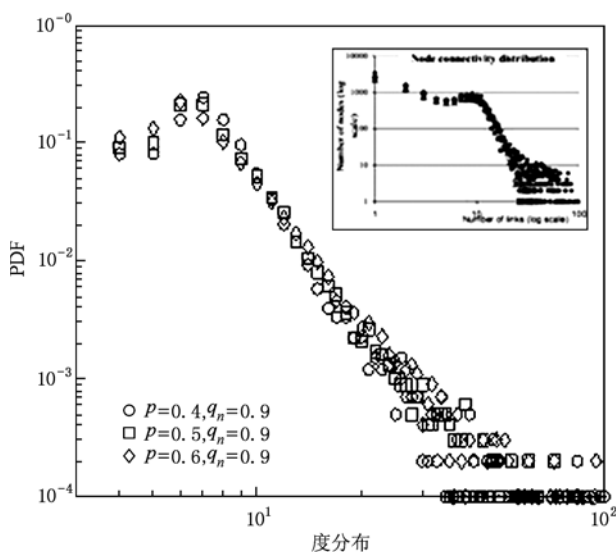


图5 RAPA 模型在 $q_n \rightarrow 1$ 时的度分布(插图表示 Gnutella 的节点连接度分布)

而类似的分布也可见于 Gnutella 网络^[25] 中(见图 5). 说明在这些实际系统中,由于负载均衡实现或存在节点最大连接数限制等原因,节点连接策略中引入了“扶贫”因素,即在一定程度上优先连接低度节点. 当 q_n 较小时,亲富加剧了度分布曲线的“长尾”现象,尤其是当 $q_n \rightarrow 0$ 时少数节点成为“垄断”节点,即所谓的“赢者通吃(winner-takes-all)”.

5. 其他拓扑特征

5.1. 平均路径长度

平均路径长度是指网络中两节点间最短路径的平均值,通常以节点间的跳数作为度量来计算. 平均路径长度与网络的连通性、可达性以及传输延迟等特征密切相关. 平均路径的计算采用的公式^[2] 为

$$d = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}, \quad (13)$$

d_{ij} 为网络中节点 i 和节点 j 之间的最短路径长度. 图 6 是在不同模型参数下的演化模型的平均路径长度,其中“+”标注的曲线是 BA 模型平均路径长度的理论值 $l \approx \ln N / \ln \ln N$ ^[26]. 显然,本文提出的 RAPA 模型具有明显的小世界特性. 而且当 q_n 参数较小时,即邻居节点选择更多考虑“亲富”,平均路径长度有轻微下降. 一个原因是确定性“亲富”情形的增加,导致新加入节点大量连接到确定性的少数高度节点上,从而降低了整个网络的平均路径长度.

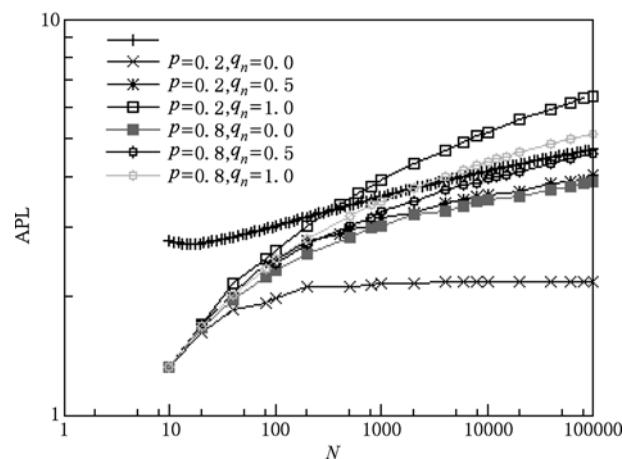


图6 RAPA 模型在不同 p, q_n 取值下的平均路径长度

5.2. 聚类系数

聚类系数 (clustering coefficient) 用于描述一个

节点邻居之间的相互连接的紧密程度,即网络的集团化程度,是网络拓扑的另一个重要参数. 对于网络的平均聚类系数的量化定义的公式^[2]为

$$\bar{C} = \sum_k P(k)C(k), \quad (14)$$

其中, $P(k)$ 表示节点的度分布, $C(k) = \langle M_{nn}(k) \rangle / (k(k-1)/2)$ 为局部聚类系数, $\langle M_{nn}(k) \rangle$ 表示节点度为 k 的节点的邻居之间平均存在的连接数.

当网络规模充分大时, BA 模型的聚类系数趋向于零, 而许多大规模的实际网络都具有明显的聚类特性, 他们的聚类系数尽管远小于 1, 但是却比 $O(N^{-1})$ 要大得多^[2]. 我们的仿真实验表明, RAPA 模型具有较高的聚类系数, 当网络规模较小时, 聚类系数随着网络规模增加而降低, 而随着网络规模进一步增加时, 聚类系数不再发生明显变化(图 7). 尤其在 p 较小的情况下, 聚类系数随网络规模增加而降低的速度更加缓慢. 在网络规模达到 100,000 时, 仍然具有非常明显的聚类特征.

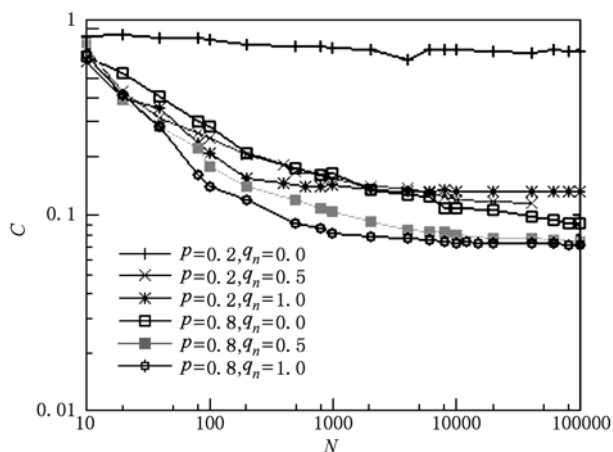


图 7 RAPA 模型在不同 p, q_n 取值下的平均聚类系数

5.3. 度相关性

在实际网络中, 节点之间的连接通常会表现出某种倾向性, 例如高度节点倾向于与其他高度节点连接, 称之为同配网络; 或者反之, 高度节点则倾向于和低度节点连接, 称之为异配网络. 这种节点之间连接的相关性, 对网络可靠性、传播动力学等都有着不同于度分布特征的重要影响, 越来越受到研究者的重视. Newman^[5, 7] 提出了一种用关联系数来定量描述混合模式的量化方法, 即计算节点度-度的相关系数 DCC(见(15)式).

$$\gamma = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}, \quad (15)$$

其中, M 是总边数, j_i 和 k_i 是第 i 条边关联的两个节点的度, $i = 1, 2, \dots, M$.

度-度的相关系数小于零表明网络具有异配特性, 大于零则具有同配特性. 当节点数量足够多时, BA^[2] 模型的度-度相关系数都趋向于零, 即没有任何倾向性. 在 RAPA 模型中, 根据 p, q_n 参数的取值, 度-度相关系数可以在 $(-0.9, 0.5)$ 之间变化(图 8). 当 $p = 1.0$ 时, 网络性质与 q_n 无关, DCC 趋近于 0, 节点间连接没有明显的倾向性; 当 $p < 0$ 且保持固定值时, q_n 参数值越大, DCC 越大. 当 q_n 保持固定值时, p 参数越大, DCC 的值越趋近于 0. 当 $p = 0, q_n = 0$ 时, 即完全采用确定性“亲富”策略时, DCC 取得最小值 -0.9 ; 当 $p = 0, q_n = 1.0$ 时, 即完全采用确定性“扶贫”策略时, DCC 取得最大值 0.5. 当 $q_n > 0.8$ 时, 演化网络呈现同配性, 当 $q_n < 0.6$ 时, 演化网络呈现明显异配性质; 由此可见, RAPA 模型能够在保持度分布符合幂律的情况下, 产生从异配到同配的具有不同混合模式的多种网络拓扑.

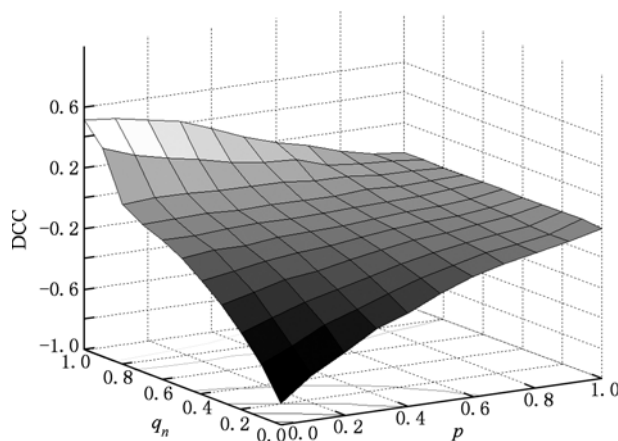


图 8 在不同 p, q_n 参数值下的度-度相关系数(网络规模 $N = 10000$)

6. 结 论

在复杂网络研究领域, 通过建立各种网络模型, 进行理论研究和仿真分析, 是一种重要的研究方法. 本文提出了一个基于随机行走和策略选择的

网络局域演化模型 RAPA. 与 BA 模型相较, RAPA 模型更加具有社会学和生物学意义. 新节点加入系统时, 不需要拥有全局知识, 仅仅通过随机选择引导节点进行随机行走, 以行走路径上的全体节点构造局域世界; 然后混合采用随机连接, 以及确定性“扶贫”连接和“亲富”连接等策略, 从局域世界中选择节点增加连接边; 最终自组织成为具有幂律结构的复杂网络. 初步的解析计算和仿真实验都表明, 与 BA 及其扩展模型不同, RAPA 模型不仅具有小世界特性、整体上的无标度特性, 还可以演化出指数截断(或“陡

尾”)、小变量饱和(或“平头”)等现象, 同时也具有明显的聚类特性, 能够构造出同配、异配等不同混合模式的网络. 依据不同的模型参数, RAPA 模型的度-度相关系数可在(-0.9, 0.5)内变化.

综上所述, RAPA 模型不仅更加符合真实网络演化的特点, 能够重现出更多的真实网络拓扑特征, 而且模型的实现和演化复杂度也比生成相同规模的 BA 网络要低. RAPA 模型的提出为进一步探索复杂网络拓扑特征的形成机理提供了一种有价值的思路.

- [1] Barabási A L, Albert R 1999 *Science* **286** 509
- [2] Newman M E J 2003 *SIAM Rev.* **45** 167
- [3] Barabási A L, Dezső Z, Ravasz E, Yook S H, Oltvai Z 2003 *Proceedings of Modeling of Complex Systems: Seventh Granada Lectures (AIP Conference Proceedings)* (Garrido P L, Marro J, eds.) (New York) American Institute of Physics 1.
- [4] Dorogovtsev S N 2004 *Phys. Rev. E* **69** 027104
- [5] Newman M 2002 *Phys. Rev. Lett.* **89** 208701
- [6] Zhou S, Mondragon R J 2004 *Phys. Rev. E* **70** 066108
- [7] Newman M E J 2003 *Phys. Rev. E* **67** 026126
- [8] Newman M E J, Park J 2003 *Phys. Rev. E* **68** 036122
- [9] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [10] Albert R, Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
- [11] Saramäki J, Kaski K 2004 *Phys. A* **341** 80
- [12] Zhang Z Z, Rong L L 2005 *Systems Engineering* **23** 1 (in Chinese) [章忠志、荣莉莉 2005 系统工程 **23** 1]
- [13] Fenner T, Levene M, Loizou G 2005 *Phys. A* **335** 641
- [14] Li X, Chen G 2003 *Phys. A* **328** 274
- [15] 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]
- [16] 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、韩祯祥 2008 物理学报 **58** 3597]
- [17] Wang W X, Hu B, Wang B H, Yan G 2006 *Phys. Rev. E* **73** 016133
- [18] Li Y, Fang J Q, Liu Q 2007 *Science & Technology Review* **25** 23 (in Chinese) [李永、方锦清、刘强 2007 科学导报 **25** 23]
- [19] Barabási A L, Albert R, Jeong H 1999 *Phys. A* **272** 173
- [20] Pons P, Latapy M 2005 *Proceedings of the 20th International Symposium on Computer and Information Sciences* 284
- [21] NIST/SEMATECH *e-Handbook of Statistical Methods*. <http://www.itl.nist.gov/div898/handbook/index.htm>; NIST/SEMATECH 2003
- [22] Jeong H, Mason S P, Barabási A L, Oltvai Z N 2001 *Nature* **411** 41
- [23] Ebel H, Mielsch L I, Bornholdt S 2002 *Phys. Rev. E* **66** 035103
- [24] Amaral L A N, Scala A, Barthélémy M, Stanley H E 2000 *Proceedings of the National Academy of Sciences* **97** 11149
- [25] Ripeanu M, Foster I, Iamnitchi A 2002 *IEEE Internet Computing Journal (special issue on peer-to-peer networking)* **6** 50
- [26] Reuven C, Shlomo H F 2003 *Phys. Rev. Lett.* **90** 058701

A evolving network model generated by random walk and policy attachment*

Jiang Zhi-Hong[†] Wang Hui Gao Chao

(College of Information Systems and Management, National University of Defense Technology, Changsha 410073, China)

(Received 28 May 2010; revised manuscript received 13 August 2010)

Abstract

Real-world networks always present some complex network properties simultaneously, such as small-world, scale-free, high clustering and assortative/disassortative mixing, etc., but only part of these properties can be reproduced in most of complex network models. In this paper, a new complex network model generated by random walk and policy attachment (RAPA) is proposed. A new peer constructs a local world by random walking, and attaches itself to peers in the local world following the policy of “random selection”, “poverty alleviation” or “favoring the rich”. The results of analysis computing and simulation demonstrate that RAPA model can reproduce not only small-world and scale-free features, but some non-power-law features such as exponential cutoff and saturation for small variables. In addition to these, RAPA model also constructs some networks with evident clustering structure and assortative/disassortative mixing pattern.

Keywords: complex network, model, random walk, policy attachment

PACS: 89.75.-k, 89.75.Fb

* Project supported by the National High Technology Research and Development Program of China (Grant No. 2008AA01Z407) and the National Natural Science Foundation of China (Grant No. 60872053).

[†] E-mail: jonathan.jiang@263.net