

一种具有指数截断和局部集聚特性的网络模型^{*}

袁韶谦[†] 赵海 李超 张昕

(东北大学信息科学与工程学院, 沈阳 110004)
(2007 年 11 月 12 日收到, 2008 年 3 月 25 日收到修改稿)

针对真实网络局域演化的特点, 提出了一种具有局部集聚特性的网络演化模型——局部集聚模型(LC 模型). 理论分析和模拟实验表明, LC 模型的节点度服从一种具有指数截断的幂律分布, 同时它的平均聚类系数要远大于局域世界模型, 接近真实网络. 模拟了 LC 模型对恶意攻击和随机错误的抵抗力, 发现高聚类系数的 LC 模型对恶意攻击更加脆弱.

关键词: 局部集聚, 指数截断, 脆弱性, 无标度网络

PACC: 0590, 0175

1. 引 言

人类对网络的研究由来已久, 最早可以追溯到 18 世纪伟大的数学家欧拉对著名“Konigsberg 七桥问题”的研究. 现代复杂网络的研究兴起于上个世纪末, 标志性的研究成果就是小世界网络模型(SW 模型)^[1]和无标度网络模型(BA 模型)^[2]的提出. 现实世界中的网络有着各自不同的特点, 网络建模研究一直都是复杂网络研究的一个很重要的方面^[3-11].

真实网络通常都很庞大, 单一节点一般无法获知网络的全局信息. 李翔和陈荣荣^[8]的研究表明, 在某些网络(比如世界贸易网)中, 优先连接机制是被限制在网络的一些局部区域起作用的, 他们因此提出了局域世界模型(LW 模型). LW 模型为每个新节点都生成一个局部集团, 而新节点只与局部集团中的节点建立连接, 用这种方式可以得到度分布介于幂律分布与指数分布之间的网络. LW 模型采用随机的方式来生成新节点要加入的局部集团, 而我们认为, 真实网络中的局部集团通常都是由具有密切联系的节点所组成的. 例如, 世界贸易网中的局部集团都是按地区划分的, 在新成员加入之前, 这些区域集团内的成员就已经建立了密切的贸易关系. 又比如学生关系网, 一个新学生最先接触的大多是同一

班级或同一实验室中的人, 这些人之间也是存在密切联系的. 随机选取的局部集团的内部联系通常都比较松散. LW 模型的平均聚类系数远小于大多数的真实网络^[8], 正好也说明了这一点.

对网络局域演化的深入探讨, 有助于人们更好地理解 and 描述真实世界中的网络. 本文设计了一种局部集聚的网络演化模型(LC 模型), 通过选取一组具有密切关系的节点作为新节点要加入的局部集团, LC 模型具有接近真实网络的平均聚类系数. 同时, 由于“钝化”效应的影响, LC 模型节点度近似服从幂律分布, 并具有指数截断效应.

2. 局部集聚模型

在真实网络中, 新节点通常无法获得网络的全局信息, 而 BA 模型却假定新节点可获知网络中全部节点的度值, 这就使其应用受到了极大的限制. 为了表现网络的局域演化, LC 模型采用了与 LW 模型一样的思路——先为新节点选择一个局部集团, 新节点只与这个局部集团中的节点建立连接, 同时 LC 模型采用了不同的选取方法以便能使局部集团内存在密切联系. LC 模型的构造分为两个步骤. 首先是初始设定, 构造一个具有 m_0 个节点和 e_0 条边的随机网络. 然后是网络增长. 每次加入一个新节点, 按照局部集团的选取方法, 从当前网络中选取

^{*} 国家自然科学基金(批准号: 69873007)资助的课题.

[†] E-mail: yuanshaoqian@gmail.com

具有 M ($M \leq m_0 + t$, 其中 t 为网络中的新节点数) 个节点的局部集团, 新节点从中随机选取 m ($m \leq M$) 个节点来建立连接.

在 LC 模型中, 局部集团的选取方法是一个关键. 为了选取一组具有密切关系的节点作为新节点的局部集团, LC 模型采用了一种树状的选取方法, 并设选取树的层数为 d , 每层的节点数 $M_d = M/d$.

首先, 从网络中随机选取一个节点, 作为初始的生成节点, 然后选择它的 M_d 个邻节点, 加入局部集团, 并随机地选择其中一个作为新的生成节点.

选择新生成节点的 M_d 个不属于当前局部集团的邻节点加入局部集团, 并从中随机选取一个节点作为新的生成节点. 重复这一步, 直至得到有 M 个节点的局部集团. 按这种方式选取的局部集团(如图 1 所示)不包含第一个生成节点.

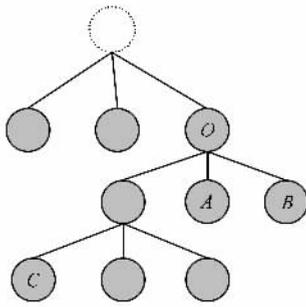


图 1 局部集团示意图 $M=9, d=3$

当生成节点的邻节点数小于 M_d 时, 选取全部满足条件的邻节点并选择其中一个作为新的生成节点. 当新生成节点没有满足条件的邻节点时, 可以在当前的局部集团中随机选择一个节点作为新的生成节点. 在网络连通的情况下, 上述选取方法能最终得到有 M 个节点的局部集团.

树形的选取方法意味着局部集团中的节点是通过一条边而被选中的, 节点通过这种方式被选中的概率与节点的度 k 有关^[7], 所以 LC 模型中新节点随机地与局部集团中的节点建立连接.

3. 理论分析与数值模拟

3.1. 度分布分析

节点的度 k 是指与节点相连的边的数量, 节点度的分布可以很直观地刻画网络的结构特性, 有必要对 LC 模型的度分布进行深入分析. 与 LW 模型

一样^[8], 讨论节点度分布时, LC 模型也有 $M = m_0 + t$ 和 $M = m$ 两种特殊情况.

当 $M = m_0 + t$ 时, 整个网络都被看成是新节点的局部集团, 这时新节点从整个网络中随机选取 m 个节点建立连接, 节点的度服从指数分布^[7].

当 $M = m$ 时, 新节点和局部集团内的每个节点都建立连接. 所以, 其与原网络中的节点建立连接的概率相当于通过树形选取方法得到一个节点的概率 Π ^[7], 即

$$\Pi(k_i) = \frac{k_i}{\sum_i k_i}. \quad (1)$$

式中 k_i 为第 i 个节点的度. 从(1)式可知, k 值越大的节点被选中的概率就越大. (1)式正是 BA 模型的优先连接公式, 此时 LC 模型等价于 BA 模型^[7], 节点度服从幂指数为 3 的幂律分布.

在一般情况下, 设 M 固定且 $M \ll m_0 + t$, $m < M$, 由(1)式可得, 网络中为 k_i 的节点被选入局部集团的概率 Π_1 为

$$\begin{aligned} \Pi_1(k_i) &= M\Pi(k_i) \\ &= M \frac{k_i}{\sum_i k_i}. \end{aligned} \quad (2)$$

由于概率最大只能取 1, 因而当 $\Pi_1 > 1$ 时, Π_1 取为 1, 这表示节点 i 一定会被选入局部集团. 要使 $\Pi_1 > 1$, 第 i 个节点的度 k_i 要满足

$$k_i > \frac{\sum_i k_i}{M}. \quad (3)$$

根据文献[12]可知

$$\sum_i k_i = \chi(e_0 + mt) \approx 2mt. \quad (4)$$

将(4)式代入(3)式可得

$$k_i > \frac{2mt}{M}. \quad (5)$$

满足(5)式的节点, 被选入局部集团的概率 $\Pi_1 > 1$, 说明这些节点一定会被选入局部集团.

当 $M \ll m_0 + t$ 时, k_i 必须非常大才能满足(5)式, 在不考虑(5)式的情况下, 可认为(2)式对所有节点都成立. 新节点随机选择局部集团中的节点建立连接, 局部集团中的节点被选中的概率 Π_s 可表示为

$$\Pi_s = \frac{m}{M}. \quad (6)$$

由(2)(6)式可以得到新节点与原网络中节点 i 建

立连接的概率 Π_c 为

$$\begin{aligned} \Pi_c(k_i) &= \Pi_l(k_i)\Pi_s \\ &= m \frac{k_i}{\sum_i k_i} \end{aligned} \quad (7)$$

假设节点度 k_i 的值连续,由(4)(7)式可得节点度的变化率

$$\frac{\partial k_i}{\partial t} \approx \frac{k_i}{2t} \quad (8)$$

根据文献 [12] 的推导可知,网络节点度分布

$$P(k) \propto 2m^2/k^3.$$

所以在一般情况下,LC 模型节点度也近似服从幂指数为 3 的幂律分布.

任何 k_i 满足条件(5)式的节点,其被选入局部集团的概率 $\Pi_l = 1$.这时新节点与其建立连接的概率 Π_c 将变成

$$\begin{aligned} \Pi'_c(k_i) &= \Pi_l(k_i)\Pi_s(k_i) \\ &= \frac{m}{M} \end{aligned} \quad (9)$$

即所有满足临界条件(5)式的节点,与新节点都以相同概率进行连接.这是因为这些节点都会被选入局部集团,而新节点可随机选择局部集团中的节点连接,这时度值差别将不再使它们之间有相对优势.

当 M 固定时,随着网络规模的增大(5)式所表示的临界值也将不断增大,其对网络的影响较小.而当 M 不固定时,设 $M = \rho(m_0 + t) \approx \rho t (0 < \rho \leq 1)$,此时条件(5)式变为

$$k'_i \geq \frac{2m}{\rho} \quad (10)$$

(10)式中的 $2m/\rho$ 是一个不随网络规模变化而变化的临界值.而随着 ρ 的增大,临界值会逐渐变小,当 ρ 趋近于 1 时,临界值变得很小,绝大多数的节点都以相同的概率与新节点建立连接,这时 LC 模型节点度服从指数分布.

所有满足(5)式的节点与新节点建立连接的概率要小于(7)式给出的概率,这就降低了网络中大度值节点出现的概率,本文称这种效应为“钝化”效应,“钝化”效应使得 k 超过临界值的节点按照指数网络的规律进行演化.所以 LC 模型的节点度实际上服从一种具有指数截断的幂律分布.

图 2、图 3 给出了 M 取不同值时 LC 模型的度分布.网络节点数 $N = 20000$.从图 2 可以看到,当 $M = 30, m = 3$ 时,LC 模型的节点度近似服从幂律分布.而从图 3 可以看到,当 $M = 0.5(m_0 + t)$ ($m_0 +$

t) 时,LC 模型的节点度逐渐趋近于指数分布.

图 4 给出了 $M = 0.5(m_0 + t), m = 3$ 时 LC 模型和 BA 模型的度分布.从图 4 可以看到,相对于 BA 模型,LC 模型的度分布明显地具有“钝化”效果, k 值较大的节点数量明显少于 BA 模型, $k > 60$ 的节点已经不存在.从图 4 还可以看到, k 在 20 左右的度分布有个截断变化.

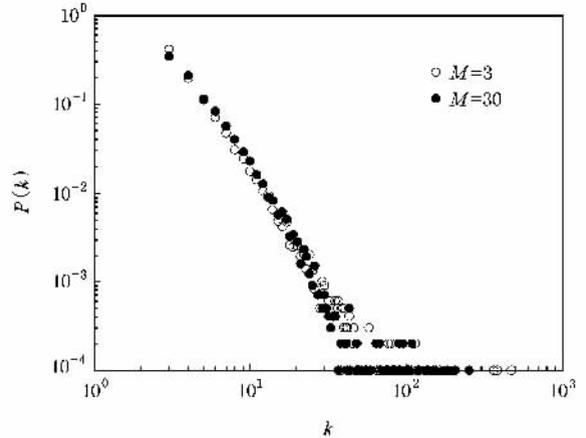


图 2 $m = 3, M = 3, 30$ 时 LC 模型的度分布

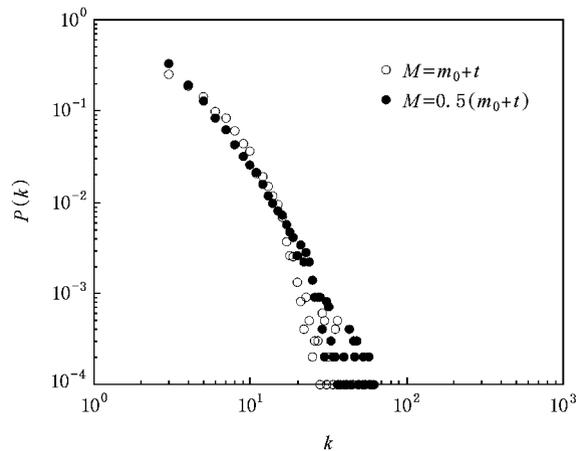


图 3 $m = 3, M = 0.5(m_0 + t), m_0 + t$ 时 LC 模型的度分布

由(5)式可知,随着 M 的增大截断临界值降低,“钝化”效应将得到加强,网络中 k 值较大的节点减少, k 值较小的节点增多,网络的异构性将会逐渐降低.网络的异构性可以用节点度的均方差 D_k 来刻画,即

$$D_k = \sqrt{\frac{\sum (k_i - \bar{k})^2}{N}} \quad (11)$$

其中 \bar{k}_i 为网络的平均度, D_k 表示网络节点度值偏

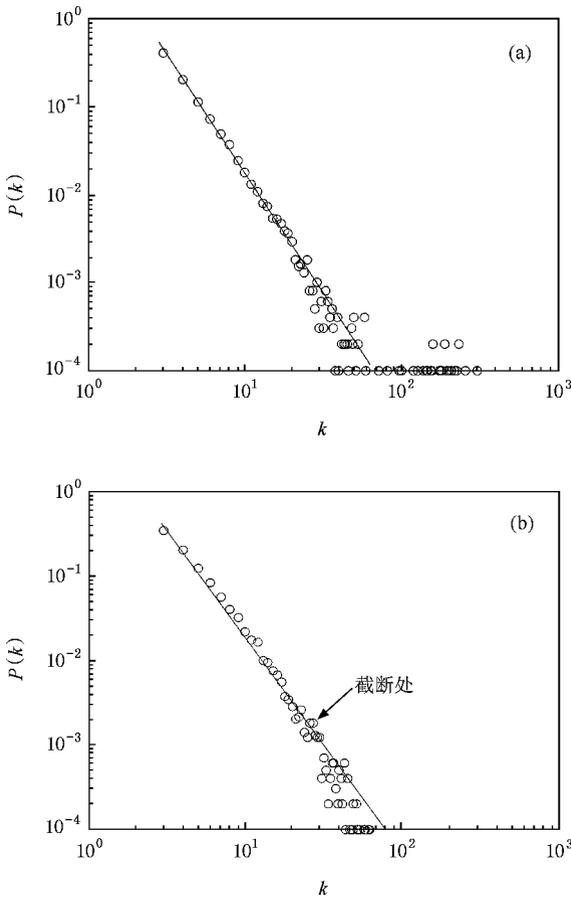


图4 BA模型($m=3$)的度分布和LC模型的度分布($M=0.3m_0 + t, m=3$) 网络规模 $N=20000$. (a)BA模型 (b)LC模型

离平均值的程度.

图5给出了LC模型和LW模型的节点度均方差 D_k 随 M 的变化情况($m=3$),图中也给出了BA模型和指数模型的对比.从图5可以看到,当 $M > 3$ 时,LC模型的节点度均方差 D_k 小于BA模型的 D_k ,这验证了“钝化”效应的存在.从图5还可以看到,随着 M 的增加,LC模型的节点度均方差有变小的趋势,而LW模型的节点度均方差则逐渐变大,与文献[8]对LW模型节点度异构性的分析相一致.

指数截断型幂律度分布已经在一些真实网络中被发现,如演员合作网^[13]和万维网^[14],LC模型可以为这种现象的出现提供另一种解释.

3.2. 网络最短路径

按照文献[1]对SW网络的定义,具有小世界特性的网络的最短路径 L 应该满足

$$L \propto \frac{\ln(N)}{\ln(\langle k \rangle)}$$

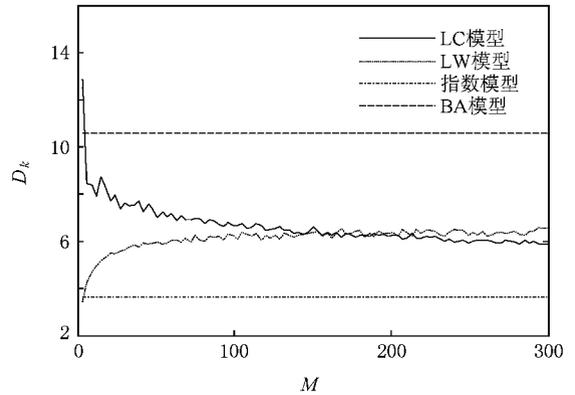


图5 节点度的均方差 D_k 随参数 M 的变化 $m=3$

我们定义网络的小世界系数 A_{sw} 为

$$A_{sw} = \frac{L \ln(\langle k \rangle)}{\ln(N)}. \quad (12)$$

对于具有小世界特性的网络, A_{sw} 不随网络规模 N 的变化而变化.图6给出了BA模型、LW模型和LC模型的小世界系数 A_{sw} 随 N 的变化.从图6可以看到,随着 N 的增长,三种网络的 A_{sw} 都近似保持不变,这说明LC模型具有小世界特性.从图6中还可以看到,LC模型和LW模型相对于BA模型,网络的平均最短路径要大一些,而LC和LW模型的网络平均最短路径则比较接近.

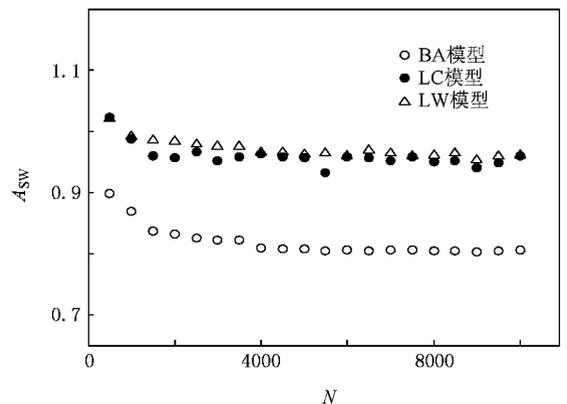


图6 BA模型、LW模型和LC模型的小世界系数随网络规模的变化

3.3. 聚类系数

节点 i 的聚类系数 c_i 反映了它的邻节点之间互相连接的程度^[15],网络平均聚类系数可表示为

$$C = N^{-1} \sum_i c_i.$$

C 反映了通过三边连接三点而形成的三角形子图

在全网络中的密度. 真实网络大多具有比较大的平均聚类系数, 而 BA 模型和 LW 模型的平均聚类系数却都很小, 近似等于随机网络的平均聚类系数, 因此并不能很好地应用于真实网络的建模. LC 模型因为选取一组具有密切关系的节点作为局部集团, 生成的网络具有大得多的平均聚类系数, 下面对此进行分析和验证.

目前对网络平均聚类系数的分析大都应用了平均场理论^[7,16,17], 应用该理论可以得到平均聚类系数的精确解析解, 但推导却很繁琐, 用于分析 LC 模型的聚类系数也较困难. 我们用一种简单的方法对 m 较小时 LC 模型的平均聚类系数做了估算.

对于节点度服从幂律分布的网络, 网络中大部分节点的 k 值都比较小, 这些节点的聚类系数决定着网络的平均聚类系数. 当 m 较小时, 新节点只对原网络中极少数节点有影响, 并且主要是 k 值较大的节点, 可以认为网络中大量 k 值较小节点的聚类系数近似不变, 与这些节点刚加入网络时的聚类系数相当, 因此我们用新节点加入网络时的聚类系数作为网络平均聚类系数的估计值.

新节点的聚类系数与局部集团内部的连接情况有关. 局部集团内有三类边, 下面以图 1 为例进行说明.

首先是生成节点和它的邻节点间的边, 如图 1 中的边 OA 和 OB . 这类边是确定存在的, 其总数为

$$N_d = (d-1) \frac{M}{d}. \quad (13)$$

其次是生成节点的邻节点间的边, 如节点 O 的邻节点 A 与 B 间的边. 这类边是不确定的, 与网络的平均聚类系数有关, 其总数为

$$N_u = \frac{1}{2} dC \frac{M}{d} \left(\frac{M}{d} - 1 \right). \quad (14)$$

最后是完全无关的两个节点之间的边, 如节点 A 与 C 间的边. 这类边完全是随机的, 存在的概率较小, 可以忽略不计.

因为新节点的 m 个邻节点是从局部集团的 M 个节点中随机选取的, 因此它的聚类系数 c_i —— m 个节点之间存在的三角形子图的密度等于局部集团内三角形子图的密度, 即

$$C = c_i = \frac{\chi(N_d + N_u)}{M(M-1)}. \quad (15)$$

将(13)(14)式代入(15)式, 经化简得到

$$C = \frac{2}{M} \quad (M \neq 1, d \neq 1). \quad (16)$$

当 m 较小的情况下, 由(16)式可知 LC 模型的平均

聚类系数反比于 M .

图 7 给出了不同参数下经数值模拟得到的 LC 模型的平均聚类系数. 从图 7 可以看到, 在 $m=3$ 时, LC 模型的平均聚类系数与(16)式的估计值一致, 而 $m=30$ 时, 实际值与估计值的偏差比较大. 从图 7 中还可以看到, 当 $M < 200$ 时, BA 模型和 LW 模型的平均聚类系数要远小于 LC 模型的平均聚类系数.

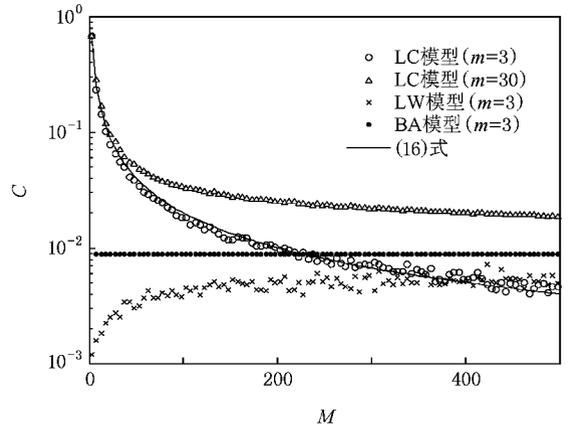


图 7 网络平均聚类系数 C 与 M 的关系

在进一步的模拟实验中, 我们发现 LC 模型的平均聚类系数不随网络规模的改变而改变, 而 LW 模型则与 BA 模型一样^[16], 随着网络规模的增大网络平均聚类系数会逐渐变小. 由此可见, LC 模型还具有保持平均聚类系数稳定的能力.

4. 容错性和抗攻击性

容错性和抗攻击性是指一个系统对随机错误和恶意攻击的抵抗能力. 对复杂网络而言, 它是指网络中随机的部分节点或者某些关键节点失效的情况下, 网络维持它本身一些特性(如连通性、最短路径长度)的能力.

无标度网络对于随机故障具有很强的鲁棒性, 而在恶意攻击下却非常脆弱^[6,18,19]. 我们用实验模拟了 LC 模型在发生随机故障和受到恶意攻击时的表现. 采用随机从网络中删除一部分节点的方法来模拟随机错误, 采用删除 k 值较大节点的方法来模拟恶意攻击.

由于计算最短路径的算法具有较高的时间复杂度 $O(N^3)$, 所以我们仅用 $N=2000$ 的网络来进行模拟实验. 为了对比, 我们也模拟了 LW 模型和 BA 模

型的表现. 实验参数如下:对于 LC 模型, $M = 6$, $m = 3$, $d = 2$, 对于 LW 模型, $M = 6$, $m = 3$, 对于 BA 模型, $m = 3$.

图 8、图 9 显示了 BA 模型、LW 模型和 LC 模型发生随机错误和受到恶意攻击时,网络最大连通区域和直径随失效节点比例的变化情况. 由此可以看到,三种模型在受到恶意攻击时都表现得非常脆弱,而发生随机错误时则都比较稳定.

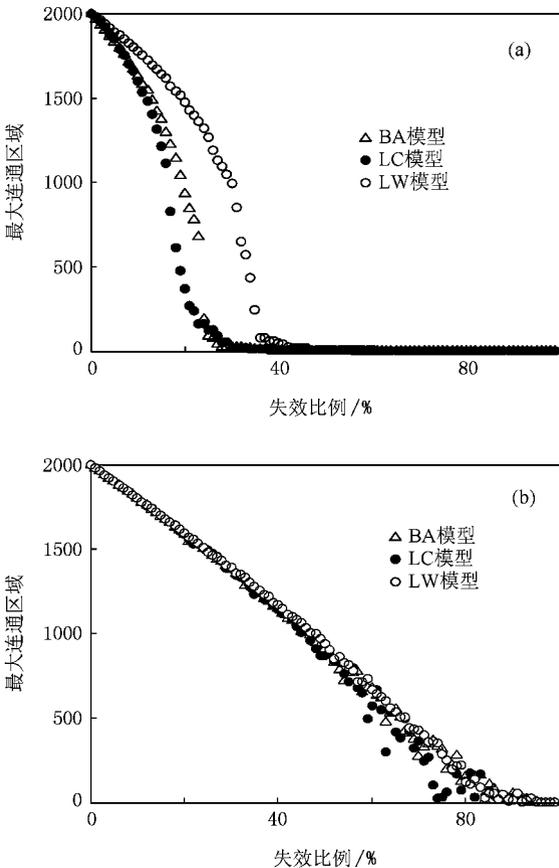


图 8 网络在受到恶意攻击和发生随机错误时,最大连通区域的变化与节点失效比例的关系 (a)受到恶意攻击时 (b)发生随机错误时

从图 8、图 9 中还可以看到,三种模型对随机错误的抵抗力没有太大的差异,但在恶意攻击下所表现的脆弱程度却不尽相同,按抵抗力强弱的顺序依次为 LW 模型→BA 模型→LC 模型. 对于恶意攻击的抵抗力与网络的异构程度有一定关系,网络异构性越大,网络中的 k 值较大节点的作用就越突出,删除这些节点后对网络的影响就越大. 在 $M = 6$, $m = 3$ 时, LW 模型节点度的均方差最小,即异构性最弱,因此当受到恶意攻击时, LW 模型的抵抗力相对更强.

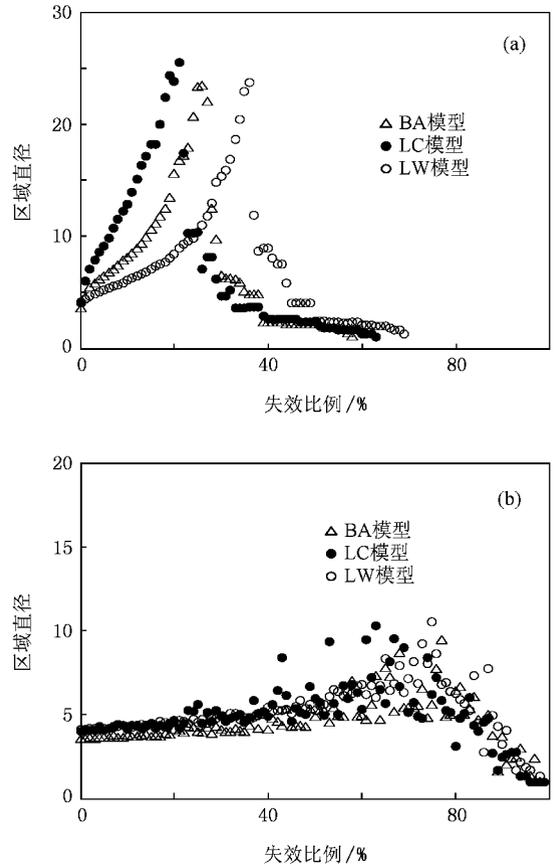


图 9 网络在受到恶意攻击和发生随机错误时,最大连通区域直径与节点失效比例的关系 (a)受到恶意攻击时 (b)发生随机错误时

抗攻击能力的强弱与网络的平均聚类系数也有关系^[19]. 在 $M = 6$, $m = 3$ 时, LC 模型的节点度均方差要小于 BA 模型(8.471 和 10.596),但 LC 模型对恶意攻击的抵抗力却更弱. 文献[19]指出,高的聚类系数降低了网络的效率,更多的最短路径将通过同样的关键节点,使得网络更易于被攻击. 在 $M = 6$, $m = 3$ 时, LC 模型的平均聚类系数为 0.33,远高于 BA 模型(0.0088). 高的聚类系数意味着网络中的很大一部分边都是作为一种冗余而存在,不利于维持网络的稳定. 这就是 LC 模型比 BA 模型和 LW 模型更加脆弱的原因.

5. 结 论

本文对网络的局域演化特性进行了探讨,提出了一种具有局部集聚特性的网络演化模型. 理论分析和模拟实验表明, LC 模型可以生成具有指数截断型幂律度分布的网络,并且由于选择一组具有密切

关系的节点作为新节点所加入的局部集团, LC 模型具有很大的平均聚类系数, 这使其更适于对真实网络进行建模. LC 模型和 LW 模型一样, 都是基于局域演化的思想而提出, 但 LC 模型又不同于 LW 模

型, 它不是 LW 模型的简单扩展. LC 模型抗攻击性的模拟实验表明, LC 模型对于恶意攻击比 BA 模型和 LW 模型更加脆弱. 对此, 需要作进一步深入研究.

-
- [1] Watts D J , Strogatz S H 1998 *Nature* **393** 440
- [2] Barabási A L , Albert R 1999 *Science* **286** 509
- [3] Dorogovtsev S N , Mendes J F F 2000 *Phys. Rev. E* **62** 1842
- [4] Bianconi G , Barabási A L 2001 *Europhys. Lett.* **54** 436
- [5] Jost J , Joy M P 2002 *Phys. Rev. E* **66** 036126
- [6] Albert R , Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
- [7] Saramäki J , Kaski K 2004 *Physica A* **341** 6
- [8] Li X , Chen G 2003 *Physica A* **328** 274
- [9] Chen F , Chen Z Q , Yuan Z Z 2007 *Chin. Phys.* **16** 287
- [10] Deng K , Tang Y 2004 *Chin. Phys. Lett.* **21** 3
- [11] Guo J L 2007 *Chin. Phys.* **16** 1239
- [12] Barabási A L , Albert R , Jeong H 1999 *Physica A* **272** 14
- [13] Amaral L A N , Scala A , Barthelemy M , Stanley H E 2000 *Proc. Natl. Acad. Sci. USA* **97** 11149
- [14] Mossa S , Barthélémy M , Stanley H E , Amaral L A N 2002 *Phys. Rev. Lett.* **88** 138701
- [15] Newman M E J 2003 *SIAM Rev.* **45** 2
- [16] Fronczak A , Fronczak P , Holyst J A 2003 *Phys. Rev. E* **68** 046126
- [17] Newman M E J , Moore C , Watts D J 2000 *Phys. Rev. Lett.* **84** 3201
- [18] Albert R , Jeong H , Barabási A L 2000 *Nature* **406** 5
- [19] Holme P , Kim B J , Yoon C N , Han S K 2002 *Phys. Rev. E* **65** 056109

A network model with exponential cut-off degree distribution and local-cluster property^{*}

Yuan Shao-Qian[†] Zhao Hai Li Chao Zhang Xin

(School of Information Science and Engineering , Northeastern University , Shenyang 110004 , China)

(Received 12 November 2007 ; revised manuscript received 25 March 2008)

Abstract

According to the character of local evolving mechanism in real network , we propose a new local-clustered network evolving model (LC model). Theoretical analysis and numerical simulations show that the LC model obeys a power-law degree distribution with an exponential cut-off , and its cluster coefficient is much bigger than the local-world model (LW model) proposed by Li and Chen and is more , close to real network . We simulate the performance of LC model against random errors and malice attacks , and discover that LC model is more fragile under attacks because of its large cluster coefficient .

Keywords : local clustered , exponential cut-off , frangibility , scale-free network

PACC : 0590 , 0175

* Project supported by the National Natural Science Foundation of China (Grant No. 69873007).

† E-mail : yuanshaoqian@gmail.com