

一种可大范围调节聚类系数的加权无标度网络模型^{*}

潘灶烽 汪小帆[†]

(上海交通大学自动化系, 上海 200240)

(2005 年 10 月 17 日收到, 2006 年 4 月 14 日收到修改稿)

在 Barrat, Barthélemy 和 Vespignani (BBV) 加权无标度网络模型的基础上, 提出了一种可大范围调节聚类系数的加权无标度网络模型——广义 BBV 模型 (GBBV 模型). 理论分析和仿真实验表明, GBBV 模型保留了 BBV 模型的许多特征, 节点度、节点权重和边权值等都服从幂律分布. 但是, GBBV 模型克服了 BBV 模型只能小范围调节聚类系数的缺陷, 从而可以用于具有大聚类系数网络的建模.

关键词: 无标度网络, 加权网络, 聚类系数

PACC: 0590, 0250

1. 引 言

现实生活中, 复杂网络的例子比比皆是, 如 Internet^[1]、交通网 (traffic networks)、科学合作网^[2,3] (scientific collaboration networks)、社会网^[4,5] (social networks)、生物网 (biological networks)^[6,7] 等等. 人们也从不同的角度出发, 提出了各种各样的复杂网络模型, 其中最为著名的是 Erdős 和 Rényi 研究的随机图 (random graph)^[8]、Watts 和 Strogatz 提出的小世界网络 (small-world network) 模型^[9] 以及 Barabási 和 Albert 提出的无标度网络 (scale-free networks) 模型^[10]. 现有的大多数针对复杂网络的研究^[11,12] 主要属于无权网络 (unweighted networks) 的范畴, 即网络的节点与节点之间只用是否有连接表示, 而不管这种连接关系的强弱. 但是, 现实世界的许多网络都不是无权网络, 例如, 在交通网中连接重大城市之间的高速公路在交通网络里的重要性要远远大于连接一般乡村之间的低级公路. 因此, 有必要通过给节点与节点之间赋予一定的权值来形容它们之间的关系, 这样就形成加权网络 (weighted networks). 最近, 在无权网络的基础上研究加权网络已开始受到重视^[13-15].

研究表明, 不少实际的加权网络的度和节点权重都满足幂律分布^[13,16]. 依据这一特性, Barrat,

Barthélemy 和 Vespignani 提出了一个加权无标度网络模型, 称为 BBV 模型^[14]. 该模型综合考虑了网络结构和节点的权重等因素来研究网络的动态演化情况. BBV 模型具有结构简单且易于作理论分析的特点. 随着模型规模的增大, BBV 模型网络的度、边权值和节点的权重都呈现无标度特性.

BBV 模型给加权网络的研究奠定了良好的基础. 但现实的许多网络, 特别是社会网络, 除了满足上述规律之外, 还具有大聚类系数的特点. 也就是在社会网络里相邻的节点之间是非常紧密地连接在一起. 大聚类系数反映的是社会网络里“物以类聚, 人以群分”的特性. 尽管 BBV 模型自身可以通过原有参数的调节来控制聚类系数的大小, 但是调节的能力有限, 尤其是不能产生较大的聚类系数. 因此, 构造一个既符合 BBV 模型的特点, 同时又可以产生大聚类系数的加权网络模型是十分必要的. 最近, 胡波等^[17] 在 BBV 模型基础上提出了一种邻居相连的模型, 以通过参数调节达到了控制聚类系数的目的. 尽管该模型保留了 BBV 模型的无标度特性, 但是网络的演化过程、最终得到的网络结构都与 BBV 模型差异较大.

本文借鉴聚类系数可调的无权网络模型的思想^[18], 通过在 BBV 模型的基础上引入一个新的参数, 构造了一个扩展的 BBV 模型 (GBBV 模型), 我们发现 GBBV 模型既能保持原来 BBV 模型特性, 又能

^{*} 国家杰出青年科学基金 (批准号: 60225013) 和国家自然科学基金 (批准号: 70431002, 90412004) 资助的课题.

[†] 通讯联系人. E-mail: xfwang@sjtu.edu.cn

使聚类系数大范围可调.

2. BBV 模型

一个加权网络可以用一个加权的 $N \times N$ 的邻接矩阵表示. 矩阵的每个元素 w_{ij} 表示节点 i 与节点 j 之间边的权值大小. 本文只考虑无向边的情况, 即 $w_{ij} = w_{ji}$. 另外, 我们认为网络里不会有自环产生, 即 $w_{ii} = 0$. 显然, 作为对无权网络中节点 i 的度 (degree) k_i 的推广, 节点的权重 (strength) 定义为

$$s_i = \sum_{j \in \Gamma(i)} w_{ij},$$

其中 $\Gamma(i)$ 表示节点 i 的所有邻居节点所组成的集合. 直观上, 它表示一个节点在整个网络里的重要程度. 例如, 在交通网中站点的权重反映了途经这个站点的车流量的总和, 自然反映了真实的交通情况, 同时也反映了这个站点的规模和在整个交通网络中的重要性.

实际的统计分析研究^[13]表明, 许多网络的度分布和权重分布呈现无标度性质. 具体而言, 度分布满足

$$P(k) \propto k^{-\gamma},$$

其中 $2 \leq \gamma \leq 3$ ^[13], 权重分布满足

$$s \propto k^\beta,$$

其中 $\beta \geq 1$ ^[16]. 鉴于此, Barrat 等提出了 BBV 模型^[14], 综合了结构和权重, 来研究加权网络的动态演化情况.

BBV 模型的构造分为三个步骤.

(1) 初始设定. 给定 N_0 个节点, 它们组成一个全耦合网络, 其中每条边都赋予权值 ω_0 .

(2) 增长. 每次加入一个新节点 n , 让这个节点与之前的 m 个节点相连, 因此每次新加入了 m 条边. 连接节点的选择按照权重优先选择 (strength driven attachment) 进行, 即一个老节点 i 被选择的概率为

$$\Pi_{n \rightarrow i} = \frac{s_i}{\sum_j s_j}. \quad (1)$$

也就权重越大的节点越有被选择的可能.

(3) 边权值的动态演化. 每次新加入的边 (n, i) 都赋予一个权值 ω_0 . 另外, 为了简单起见, 认为新加入的边 (n, i) 只会局部地引发连接节点 i 与它的邻居节点 $j \in \Gamma(i)$ 边的权值的重新调整. 调整按照

$$w_{ij} \rightarrow w_{ij} + \Delta w_{ij}, \quad (2)$$

$$\Delta w_{ij} = \delta_i \frac{w_{ij}}{s_i} \quad (3)$$

规则进行. 也就是每次新引入一条边 (n, i) , 会给节点 i 带来额外的 δ_i 的流量负担, 而与之相连的边会按它们自身的权值 w_{ij} 的大小分担一定的流量. 因此, 总的节点 i 的权重调整为

$$s_i \rightarrow s_i + \delta_i + \omega_0. \quad (4)$$

3. GBBV 模型及理论分析

为了使模型满足可大范围调节聚类的特性, 我们借鉴 Holme 和 Kim^[18]研究无权网络时的方法, 引入一个新的边的连接选择方式——三角构成 (triad formation). 如果有一条边 (n, i) 在前面的权重优先选择的过程中被选中, 那么以概率 w_{ij}/s_i 选择一个节点 i 的邻居节点 $j \in \Gamma(i)$, 让它与新节点 n 相连, 即

$$\Pi_{n \rightarrow (j \in \Gamma(i))} = \frac{w_{ij}}{s_i}. \quad (5)$$

如果 i 的所有邻居节点都已经和节点 n 连接, 那么就做一次权重优先选择连接. 当然, 每次在节点与节点的边连接之后, 我们要像 BBV 模型那样, 引入新的流量负担, 并对权值进行再分配.

当一个新节点 n 加入网络时, 首先做一次权重优先选择连接. 此后, 以概率 p ($0 \leq p \leq 1$) 做三角构成连接, 而以概率 $1 - p$ 做权重优先选择连接. 因此, 每引入一个新的节点, 一共要做 $m_{\text{triad}} = (m - 1)p$ 次三角构成连接. 所得到的模型称为 GBBV 模型. 当 $p = 0$ 的时候, GBBV 模型即为 BBV 模型.

以上的构建机制有着简明而直观的物理意义和现实意义. 一般而言, 社会网络中一个新成员 n 总是倾向于先与网络中权重很大的成员 i 交往, 这对应于权重优先选择连接. 此后随着交际的增多, 新节点所连接的节点的邻居节点也被介绍、认识, 从而形成了三角构成连接. 值得注意的是, 在三角构成连接的过程中, 三角形子网的个数增加, 从而提高了网络的聚类系数. 另外, 一个新成员的加入会给网络带来新的负担, 边权值要得到重新分配以承担新到来的负担. 例如一条新的高速公路的竣工会刺激公路周边城市的汽车销售, 自然而然会给本地区范围内的公路带来局部新的车流负担, 但是由于地域关系, 在较短的一段时间内这种影响又不会波及交通网络里其他的遥远区域. 因此, 可以认为基于局部网络的权

值再分配原则即反映了边权值增长在较小范围内的剧烈性,同时又反映其在较大范围内影响的有限性.

下面具体分析 GBBV 模型 特别是网络的度、节点的权值、边的权值以及网络的聚类系数等重要指标.为简单起见,我们假定 $\delta = \text{const} = 1.0$ $\omega_0 = 1.0$.

当一个新的节点 n 加入网络时,对于网络中的节点 i ,易得 s_i 受四个因素的影响.(i)它被权重优先选择与节点 n 相连 (ii)它被三角构成连接与节点 n 相连 (iii)它的一个邻居 $j \in I(i)$ 被权重优先选择与节点 n 相连 (iv)它的一个邻居 $j \in I(i)$ 被三角构成连接与节点 n 相连.因此,一个时刻 s_i 的变化为

$$\begin{aligned} \frac{ds_i}{dt} &= (m - m_{\text{triad}}) \frac{s_i}{\sum_w s_w} (1 + \delta) \\ &+ \sum_{l \in I(i)} \frac{s_l}{\sum_w s_w} m_{\text{triad}} \frac{w_{il}}{s_l} (1 + \delta) \\ &+ \sum_{j \in I(i)} (m - m_{\text{triad}}) \frac{s_j}{\sum_w s_w} \delta \frac{w_{ij}}{s_j} \\ &+ \sum_{j \in I(i)} \left[\left[\sum_{l \in I(j)} \frac{s_l}{\sum_w s_w} m_{\text{triad}} \frac{w_{jl}}{s_l} \right] \delta \frac{w_{ij}}{s_j} \right] \\ &= (m - m_{\text{triad}}) \frac{s_i}{\sum_w s_w} (1 + \delta) \\ &+ m_{\text{triad}} \frac{s_i}{\sum_w s_w} (1 + \delta) \\ &+ (m - m_{\text{triad}}) \frac{s_i}{\sum_w s_w} \delta + m_{\text{triad}} \frac{s_i}{\sum_w s_w} \delta \\ &= m \frac{s_i}{\sum_w s_w} (1 + 2\delta). \end{aligned} \quad (6)$$

k_i 的变化为

$$\begin{aligned} \frac{dk_i}{dt} &= (m - m_{\text{triad}}) \frac{s_i}{\sum_w s_w} + \sum_{l \in I(i)} \frac{s_l}{\sum_w s_w} m_{\text{triad}} \frac{w_{il}}{s_l} \\ &= m \frac{s_i}{\sum_w s_w}. \end{aligned} \quad (7)$$

此外,每增加一条边,系统总的权重就要增加 $2 + 2\delta$,因此易得

$$\sum_w s_w \approx 2m(1 + \delta)t.$$

由微分方程的初始条件 $k_i(t=i) = s_i(t=i) = m$, 我们得到

$$s_i(t) = m \left(\frac{t}{i} \right)^{(2\delta+1)(2\delta+2)}, \quad (8)$$

以及

$$k_i(t) = \frac{s_i(t) + 2m\delta}{2\delta + 1}. \quad (9)$$

这样就得到了节点权重与度之间的关系

$$s_i = (2\delta + 1)k_i - 2m\delta. \quad (10)$$

(8)(9)两式的结果和 BBV 模型所得到的分析结果完全相同^[14],最终得出相同的结论,

$$P(s) \propto s^{-\gamma},$$

$$P(k) \propto k^{-\gamma},$$

即节点的权重 s 和节点的度 k 服从相同指数的幂率分布,其中

$$\gamma = \frac{4\delta + 3}{2\delta + 1}.$$

它们之间符合简单的线性关系.由于得到的结果完全与 m_{triad} 无关,说明 m_{triad} 的引入并没有破坏原来 BBV 模型的无标度特性.

类似地,我们也可以研究 GBBV 模型的边权值的演化情况.当一个新的节点连接节点 i 或者节点 j 时,都会引起 w_{ij} 的变化,因此容易得到

$$\begin{aligned} \frac{dw_{ij}}{dt} &= (m - m_{\text{triad}}) \frac{s_i}{\sum_w s_w} \delta \frac{w_{ij}}{s_i} \\ &+ (m - m_{\text{triad}}) \frac{s_j}{\sum_w s_w} \delta \frac{w_{ij}}{s_j} \\ &+ m_{\text{triad}} \sum_{l \in I(i)} \frac{s_l}{\sum_w s_w} \frac{w_{il}}{s_l} \delta \frac{w_{ij}}{s_i} \\ &+ m_{\text{triad}} \sum_{l \in I(j)} \frac{s_l}{\sum_w s_w} \frac{w_{il}}{s_l} \delta \frac{w_{ij}}{s_j} \\ &= m \frac{s_i}{\sum_w s_w} \delta \frac{w_{ij}}{s_i} + m \frac{s_j}{\sum_w s_w} \delta \frac{w_{ij}}{s_j} \\ &= \frac{\delta}{1 + \delta} \frac{w_{ij}}{t}. \end{aligned} \quad (11)$$

令 $t_{ij} = \max(i, j)$, 并且初始条件为 $w_{ij}(t_{ij}) = 1$, 我们得到了与 BBV 模型相同的结论,

$$w_{ij}(t) = \left(\frac{t}{t_{ij}} \right)^{\delta(\delta+1)} \quad (12)$$

最终得出

$$P(w) \propto w^{-\alpha},$$

即边权重分布也服从指数为 α 的幂律分布,其中 $\alpha = 2 + 1/\delta$.以上理论推导说明,GBBV 模型完全继承了 BBV 模型的无标度特性.

4. 仿真实验

在仿真实验中,我们取常数 $\delta = 1.0$,并取 $\omega_0 =$

1.0, $N = 3000, m = 3$. 图 1(a)和图 2(a)验证了关于节点权重和边权值的理论推导,而图 3(a)则证实了权重与度之间的线性关系.图 1(b)、图 2(b)和图 3(b)说明了三角构成连接概率 p 的引入没有破坏网

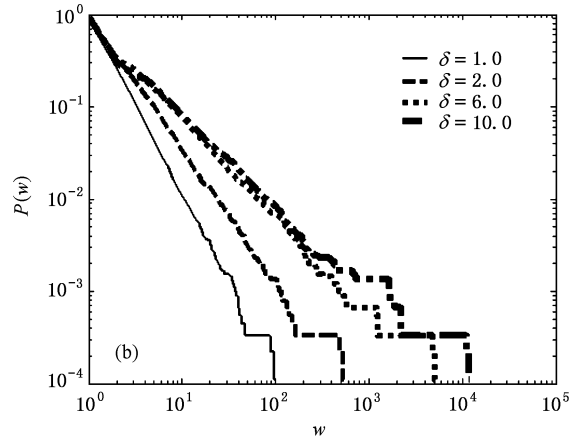
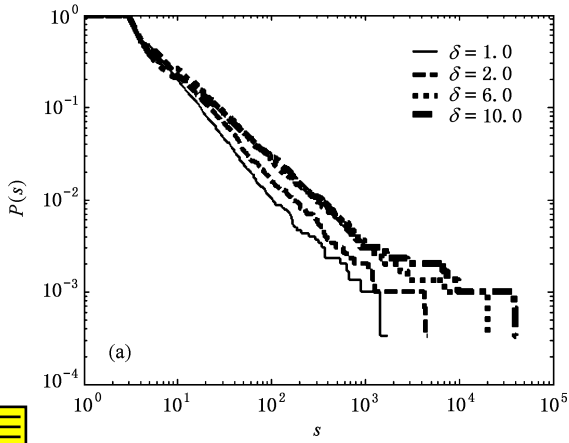


图 1 当 $p = 0.5$ 时,节点权重的分布情况(a)和边权值的分布情况(b) 网络规模 $N = 3000$

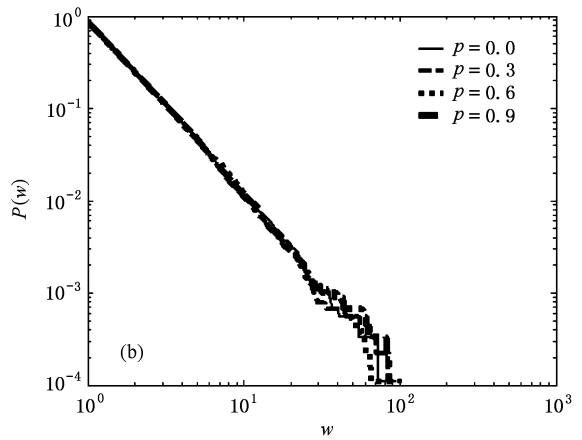
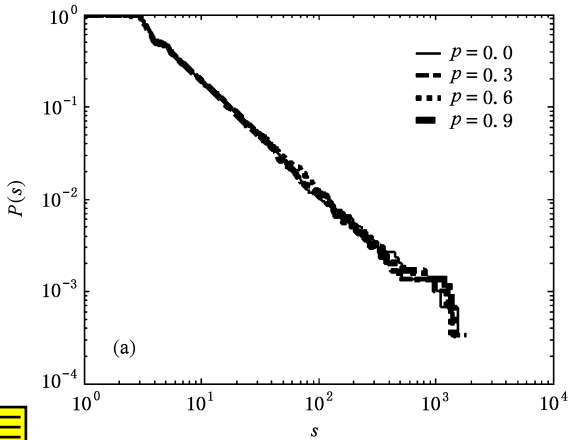


图 2 当 p 取不同值时,节点权重的分布情况(a)和边权值的分布情况(b) 网络规模 $N = 3000, \delta = 1.0$

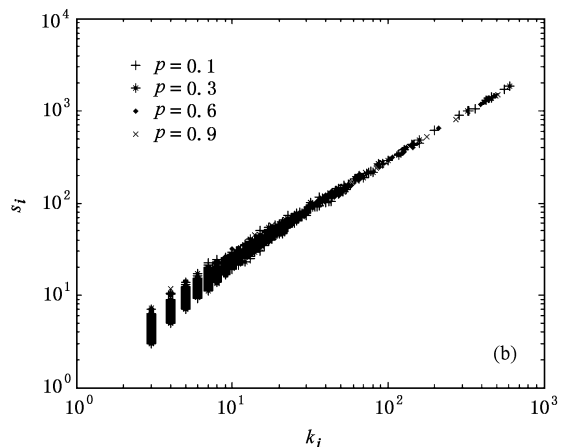
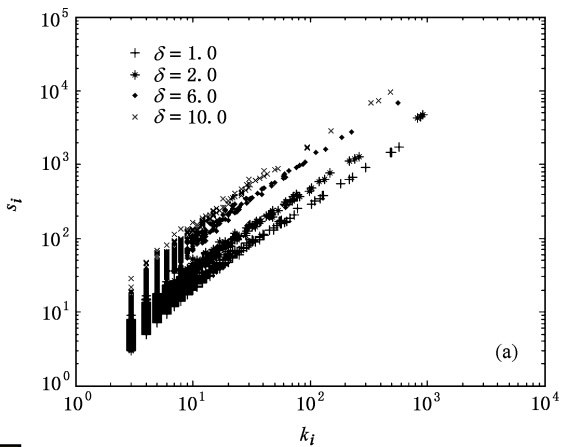


图 3 随着 δ 的变化(a)和 p 的变化(b)节点的权重 s_i 与节点的度 k_i 之间的变化关系 网络规模 $N = 3000, \delta = 1.0$

络的无标度性质,随着 p 的变化,网络的权重、权值分布基本没有改变,而权重与度之间的线性关系也得以保留.

5. 聚类系数

一个节点 i 的聚类系数 c_i 反映了它的邻居之间互相连接的程度^[19]. 网络的平均聚类系数

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

反映了通过三边连接三点而成的三角形子图在全网络里的密度. 更深入的研究则要涉及到度为 k 的节点的平均聚类系数

$$\alpha(k) = \frac{1}{NP(k)} \sum_{i|k_i=k} c_i. \quad (13)$$

在很多情况下, $\alpha(k)$ 是一个 k 的幂律下降函数,说明度小的节点属于平均聚类系数大的连通社区,反之,度大的节点属于平均聚类系数小的连通社区. 这对应于现实网络里的中心节点,在周围的一定区域,中心节点几乎单独地承担局部网络的负担任务.

现在考察 BBV 模型和 GBBV 模型的聚类调节特性. 在 BBV 模型中,节点的权重、边权值和度分布等都完全取决于唯一可调的参数 δ ,它决定了一个新节点的加入对网络结构与局部的权值变化所产生的影响. 由图 4 的内插图可知,随着 δ 的增大网络的平均聚类系数 C 也不断变大. 但是, δ 的增大又意味着网络将承担更大的负担,在网络承受能力有限的情况下(现实生活中大多数网络都有最大承受极限),显然通过增大 δ 的值来调节聚类系数是不完善的、低效率的.

现在,在 GBBV 模型中引入三角连接概率 p ,从而由参数 δ 和 p 共同控制网络结构的变化和权值的分配. 我们固定 $\delta = 1.0$ 而让 p 变化,发现随着 p 的增大,网络的平均聚类系数也不断提高(见图 5 的内插图),说明参数 p 可以控制平均聚类系数的变化. 由于 p 的存在,使网络不再只按 BBV 模型的情况演化,而是通过引入三角子网使网络的聚类程度提高,但是不影响网络的其他性质. 而且当 $p = 0$ 时,GBBV 模型即为 BBV 模型;当 $p = 1$ 时,网络的聚类系数很大.

图 4 和图 5 进一步反映了 GBBV 模型和 BBV 模型的区别. 从图 4 可以看出 BBV 模型聚类的一些特点,首先 $\alpha(k)$ 是一个 k 的幂律下降函数. 其次,随

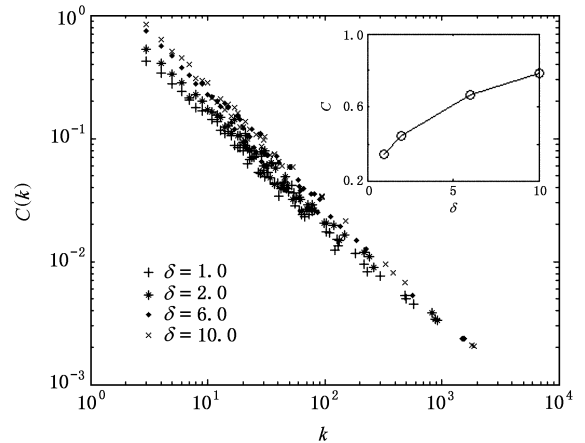


图 4 对于不同的 δ , $\alpha(k)$ 与 k 的变化图. 内插图是平均聚类系数 C 随着 δ 增大而增大的情况. 网络规模 $N = 3000$, $p = 0.5$

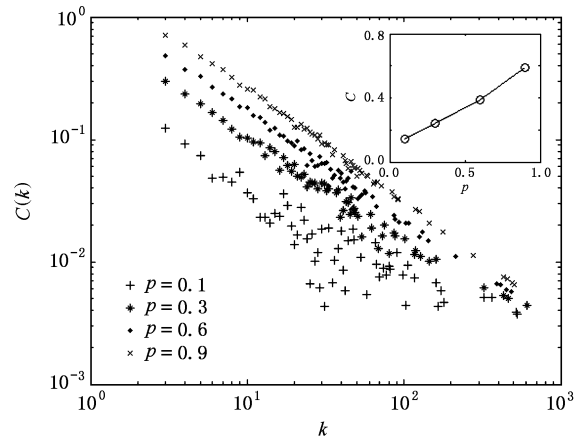


图 5 对于不同的 p , $\alpha(k)$ 与 k 的变化图. 内插图是平均聚类系数 C 随着 p 增大而增大的情况. 网络规模 $N = 3000$, $\delta = 1.0$

着 δ 的增大而引起 $\alpha(k)$ 变化的节点几乎都为度很小的节点,而度较大的节点聚类系数几乎不变. 另外,可以看出 $\alpha(k)$ 随着 δ 的增大而变化的情况并不敏感,对于不同的 δ 得出的 $\alpha(k)$ 的曲线几乎重合在一起. 图 5 则反映了 GBBV 模型聚类的另外一些特点. 首先,在边总数和 BBV 模型相同的情况下,三角连接概率 p 引起了聚类系数 $\alpha(k)$ 的较大波动,同时引起这些波动的范围扩大到了所有节点. 其次,无论是度大的节点还是度小的节点,随着 p 增大到一定的值后(例如 $p \geq 0.3$), $\alpha(k)$ 值也增大,且增幅基本相同,双对数曲线几乎成平行线. 比较图 4 和图 5 可以发现,虽然 δ 和 p 两个参数都可以调节聚类系数,但是它们所引起聚类系数变化的结果是不同的, δ 侧重于调节局部度小节点的聚类系数,对 $\alpha(k)$ 的影响较小,而 p 则将这种调节扩大到了全局从而引

起 α_k 较大的波动.

这样两种调节方法的存在,使得我们想到通过它们的组合来高效地调节聚类系数,以适应社会网络的研究需要.图 6 为平均聚类系数 C 随着 δ 和 p 的变化情况.从图 6 可以看出,在 δ 取值不是很大(例如 $\delta < 5$)又没有引入概率 p 的情况下,模型调节聚类系数的能力有限,范围大约在 0.1 到 0.5 以内.

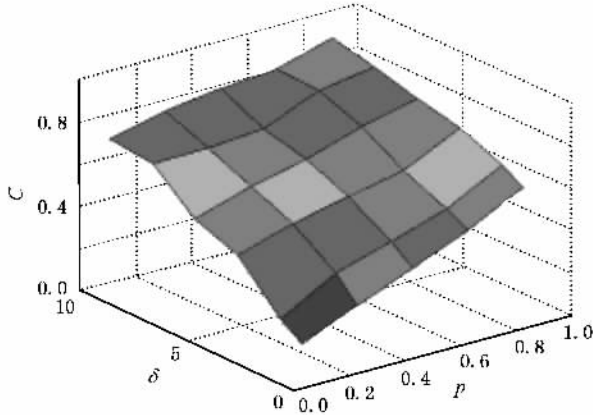


图 6 平均聚类系数 C 随着 δ 和 p 的变化情况 得到的数据都是 50 次试验后的平均结果,网络规模 $N = 3000$

而引入 p 之后,通过调节 p 值能使得聚类系数有较大幅度的提高,调节的最大范围扩大到了 0.1 到 0.9.鉴于此,我们认为通过 δ 和 p 两个参数的共同作用,聚类系数的调节范围大幅度提高,调节的效率也显著增强.

6. 结 论

本文借鉴聚类系数可调的无权网络模型的思想,通过在 BBV 模型的基础上引入一个新的参数,构造了一个新的加权无标度网络模型.理论分析和仿真表明,GBBV 模型既保持原来 BBV 模型的许多特性,又能使聚类系数在较大范围内可调.当然,GBBV 模型还有许多不完善的地方.首先,社会网络有一个显著的特征——相配性(assortativity),即网络中度大的节点倾向于与度大的节点相连.经验证,发现 GBBV 模型和原始的 BBV 模型一样,都不具备这样的性质.其次,与 BBV 模型一样,GBBV 模型考虑的重点是如何搭建加权无标度网络,从而忽略了一些在真实网络里影响网络演化扩张的因素.这些都是值得进一步研究的课题.

[1] Pastor-Satorras R, Vespignani A 2004 *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge: Cambridge University Press)

[2] Newman M E J 2001 *Phys. Rev. E* **64** 016132

[3] Barabási A L, Jeong H, Néda Z *et al* 2002 *Physica A* **590**

[4] Newman M E J 2001 *Proc. Natl. Acad. Sci. USA* **98** 404

[5] Amaral L A N, Scala A, Barthélemy M *et al* 2000 *Proc. Natl. Acad. Sci. USA* **97** 11149

[6] Uetz P, Giot L, Cangney G *et al* 2000 *Nature* **403** 623

[7] Lee T I, Rinaldi N J, Robert F *et al* 2002 *Science* **298** 799

[8] Erdős P, Rényi A 1959 *Publ. Math.* **6** 290

[9] Watts D J, Strogatz S H 1999 *Nature* **393**

[10] Barabási A L, Albert R, Jeong H 1999 *Physica A* **272** 173

[11] Huang Z X, Wang X R, Zhu H 2004 *Chin. Phys.* **13** 273

[12] Li Y, Liu Y, Shan X M *et al* 2005 *Chin. Phys.* **14** 2153

[13] Barrat A, Barthélemy M, Pastor-Satorras R *et al* 2004 *Proc. Natl. Acad. Sci. USA* **101** 3747

[14] Barrat A, Barthélemy M, Vespignani A 2004 *Phys. Rev. Lett.* **92** 228701

[15] Li W, Cai X 2004 *Phys. Rev. E* **69** 046106

[16] Pastor-Satorras R, Vázquez A, Vespignani A 2001 *Phys. Rev. Lett.* **87** 258701

[17] Hu B, Yan G, Wang W X *et al* 2005 *Nonlinear Sci.* A00505041

[18] Holme P, Kim B J 2002 *Phys. Rev. E* **65** 026107

[19] Watts D J, Strogatz S H 1998 *Nature* **393** 440

A weighted scale-free network model with large-scale tunable clustering^{*}

Pan Zao-Feng Wang Xiao-Fan[†]

(*Department of Automation, Shanghai Jiaotong University, Shanghai 200240, China*)

(Received 17 October 2005 ; revised manuscript received 14 April 2006)

Abstract

Based on the weighted scale-free network model proposed by Barrat, Barthélemy and Vespignani (BBV), we propose a generalized BBV (GBBV) model with large-scale tunable clustering coefficient. Theoretical analysis and numerical simulations show that the GBBV model retains many properties of the BBV model, such as power-law distributions of node degree, node strength and edge weight. However, the GBBV model overcomes the drawback of the BBV model that the clustering coefficient can only be tuned in a small interval, and therefore the GBBV model can be used for modeling networks with large clustering coefficients.

Keywords : scale-free network, weighted network, clustering coefficient

PACC : 0590, 0250

^{*} Project supported by the National Natural Science Foundation for Distinguished Young Scholars of China (Grant No. 60225013) and the National Natural Science Foundation of China (Grant Nos. 70431002, 90412004).

[†] Corresponding author. E-mail: xfwang@sjtu.edu.cn