

一个描述合作网络顶点度分布的模型^{*}

张培培^{1)†} 何 阅²⁾ 周 涛³⁾ 苏蓓蓓⁴⁾ 常 慧⁴⁾ 周月平⁴⁾ 汪秉宏³⁾ 何大韧⁴⁾

1) 江苏大学理学院物理系, 镇江 212013)

2) 无锡市第一中学, 无锡 214031)

3) 中国科学技术大学近代物理系, 合肥 230026)

4) 扬州大学物理科学与技术学院, 扬州 225002)

(2005 年 3 月 4 日收到, 2005 年 6 月 16 日收到修改稿)

讨论一类社会合作网络以及一些与其拓扑结构相似的技术网络的度分布. 建议一个最简化模型, 通过解析的方法说明这些网络演化的共同动力学机理, 而且说明顶点的度分布和项目度分布之间具有密切的一致关系, 而项目所含的顶点数分布对度分布的影响较小, 对模型的更一般情况进行数值模拟, 说明上述结论具有一定的普遍性. 这个模型显示这类广义的合作网络一般具有处于幂函数和指数函数这两种极端情况之间的度分布. 简要介绍对一些实际合作网络做统计研究的结果, 说明本模型的合理性.

关键词: 合作网络, 度分布, 项目度分布, 项目含顶点数

PACC: 0590, 0175, 0210

1. 引 言

近年来复杂网络的研究引起了很大兴趣^[1,2]. 在许多实际或理论问题的研究中, 可能把复杂系统基本单元之间的相互作用简化地表示为图论中顶点(即上述单元)之间的赋权或不赋权的“边”. 这样, 这些复杂系统就可以在一定规定下表示为一个网络. 这是对复杂系统建模的一种新尝试.

网络中每个顶点连接其他顶点的边的数目称为它的顶点度, 用 k 表示. 顶点度为 k 的顶点出现的概率, 或称为顶点度分布, $P(k)$, 可能在相当的程度上说明网络的拓扑性质和演化机理的特征, 因此一直受到重视(为了方便, 本文采取文献[3]中提出的顶点度分布的定义, 即 $P(k)$ 表示顶点度为 k 的顶点数目, 以下其他分布的定义类似). 近来 Barabasi 和 Alber(BA) 举出大量的实际统计例子, 证明许多实际网络的顶点度分布精确地或近似地遵循幂律, 即 $P(k) \sim k^{-\gamma}$, 并且提出了可能揭示这种无标度分布律最主要产生机理的 BA 网络生长模型^[1,4]. 模型有两个要点: 第一, 网的顶点一个个地增加, 第二, 新出现顶点与已经存在的顶点之间遍及整个网络,

涉及绝大多数的边)都按照线性优选法则(即所谓“富者更富”法则)建立边的联系^[4]. 由此模型得到的解析和模拟结果显示理想的幂律. Barabasi 和 Albert 同时解析地说明了如果新出现顶点与已经存在的顶点之间完全随机地选择连接, 那么顶点度的分布满足一个指数分布 $P(k) \propto \exp(-k/m)$, 其中 m 是常数^[1,4]. 在此后的许多旨在改进或修正 BA 网络生长模型的提议中, 我们特别注意以下两个. 第一是 Liu 和 La(LL) 等人建议的介于无标度网和随机网之间的网络模型^[5]. 他们举出了科研合作网的数据统计结果介于指数函数和幂函数之间作为一个例证^[6], 说明网络的演化存在两种机理的竞争. 类似地, 他们假设网的顶点一个个地增加, 但是设新顶点以概率 $p(0 \leq p \leq 1)$ 随机连接旧顶点, 以概率 $(1-p)$ 按照线性优选法则连接旧顶点. 当 $p=0$ 时, LL 模型退化为 BA 模型, 而 $p=1$ 时它退化为随机模型. 在 $0 < p < 1$ 时, 模型具有介于无标度和指数之间的顶点度分布. 第二个模型是 Li 和 Cher(LC) 提出的, 表现了关于优选和随机选择这两种机理竞争的更具体、更深刻的思想^[7]. 他们以世界贸易网为例, 说明许多网络中存在“优先互相作用”的“局域世界”, 例如内部的经济合作和贸易关系优先的各种区域经济合作组织、互

^{*} 国家自然科学基金(批准号: 70371071, 70271070, 10472116, 70471033)资助的课题.

[†]E-mail: ppzh@ujs.edu.cn, xueren2001@sohu.com

联网中的域-路由器结构,以及蛋白质的族系或超族系.在 LC 模型中网的顶点一个个地增加,但是线性优选法则只适于随机选取一部分顶点构成的局域世界内部.当局域世界线度 M (它包含的顶点数)等于 m (每个新顶点在它的局域世界中连接的新边的数目)时,局域世界中每对顶点都连接,无优选可言,LC 模型退化为 BA 随机模型.当 $M = m_0 + t$ (m_0 是初始顶点数, $m_0 + t$ 是 t 时刻的顶点数),局域世界就是整个网络,连接法则则是完全的线性优选,LC 模型退化为 BA 模型.当 $m < M < m_0 + t$ 时,模型具有介于无标度和指数之间的顶点度分布.

大家早就注意到许多实际网络的度分布介于无标度分布(幂律分布)和指数分布之间.这种分布被称为“幂律尾分布(power law tail)”、“去头的幂律(truncated power law)分布”等等.它们可能采取从非常接近于幂函数到非常接近于指数函数这一个广阔范围内的任一种形式.我们倾向于使用 Laherrere 和 Somette 在 1998 年提出的所谓的“广延指数分布”(stretched exponential distribution SED 分布)对它们进行比较定量的描述^[8]. SED 分布的原始分布规律为 $P(x)dx = \mu(x^{m-1}/x_0^m)\exp(-(x/x_0)^\mu)dx$, Laherrere 和 Somette 证明了对应的“累计统计分布”规律为 $P(x) = \exp(-(x/x_0)^\mu)$.它更方便使用 $\ln P(x) \propto x^\mu$ 形式表示.其中的关键参数 $\mu = 1$ 时表示指数分布, $\mu \rightarrow 0$ 时表示幂律分布. μ 在 1 和 0 之间时,分布介于指数分布和幂律分布之间,数据点的中段在双对数平面上显示一个明显的线性部分(无标度区),然而数据点的头尾部分又可能显示遵循指数规律的弯曲. μ 越小于 1,无标度区越大,规律越近于幂律^[9,10].

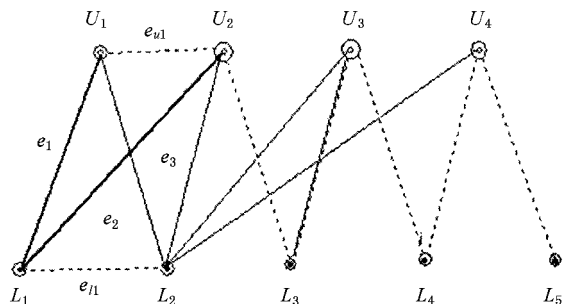


图 1 描述社会合作网的双粒子图

近年来,我们注意到社会网络中特殊的一类,即所谓“社会合作网络(Social Cooperation Networks)^[11].已经有许多讨论社会合作网络的论

文^[12],其中我们特别注意 Ramasco, Dorogavtsev 和 Pastor-Satorras (RDP)最近发表的论文^[13]以及他们提出的社会合作网络演化模型,其内容将在下一节中简单介绍.这类网络中的边通常表示它们的基本单元之间的某种合作关系,适于用图 1 所示的双粒子图来描述^[1,11-13].一类粒子可称为“项目”(act,例如电影影片或科研论文),在图 1 中用上一列顶点 ($U_i, i = 1, 2, \dots$)表示,另一类粒子可以称为“参与者”(actor,例如演员或科研人员),在图 1 中用下一列顶点 ($L_i, i = 1, 2, \dots$)表示.在这种图中一般只考虑两个不同类粒子之间的无向边,在图 1 中用粗实线 $e_i, i = 1, 2, \dots$ 表示.这样的一条边代表一个参与者参加一个合作项目.如果只考虑一类粒子(称为向一种单粒子图投影)就把两条共一个另一类单粒子的 e_i 边投影为这种单粒子之间的边.例如常常只讨论演员的合演关系,这时图 1 中两个粒子 L_1, L_2 的边 e_{11} 表示他们在同一部影片 U_1 中的合作关系.边 e_{11} 正是代表他们参加 U_1 合作项目的两条双粒子图的 e_1, e_3 边的向下投影.如果他们曾在两部影片合演,则他们之间具有两条“重复边”(multiple edges)表示合作的频繁程度或强度,具有重要的意义.类似地,也可以只讨论影片曾经聘用同样演员的关系,这时可以定义两部影片(两个 U_i 顶点)之间的边表示它们(曾经)聘用同一个演员.例如图 1 中两个粒子 U_1, U_2 之间的边 e_{21} 表示它们曾经共聘演员 L_1 的合聘关系.类似地,它是两条双粒子图的 e_1, e_2 边的向上投影.如果两部影片曾经聘用过许多相同的演员,它们之间就具有多条“重复边”.这两部影片的风格和特色一定有许多共同之处,同样具有重要的意义.如果特别注意一个演员 i 曾经演过几部影片,可以定义一个量 h_i 称为它的“顶点项目度”,等于这个 L 粒子在双粒子图中连接的 U 粒子数,如图 1 中由 L_2 粒子出发的四条细实线所示.显然,如果不计多重边,这四条双粒子图边连接的四个 U 粒子 (U_1, U_2, U_3, U_4)由于曾经聘用过同一个演员在影片单粒子图中构成一个完全图(complete graph,即每一对 U 粒子顶点之间都有一条 e_u 边连接),它们中的每一个的顶点度都等于 $h_i - 1$.类似地,如果特别注意一个影片 j 同时聘用几个演员,可以定义一个量 T_j 称为它的“项目含顶点数”,等于这个 U 粒子在双粒子图中连接的 L 粒子数.类似地这些 L 粒子也由于在同一部影片中出演而在演员单粒子图中构

成一个完全图,它们中的每一个的顶点度都等于 $T_j - 1$. 当然,两个这样的完全图很可能共用一条甚至好几条边,但是如果计算多重边,顶点度等于 $T_j - 1$ 的结论仍旧成立. 如果我们把完全图的概念推广到这种计算多重边的情况(即认为包含多重边,每两个顶点都连边的图仍叫做完全图),仍可以说网络是完全图的集合.

本文特别注意一些“非社会合作网络”或“技术合作网络”,例如城市内的公交线路网络、旅游线路网络、中药方剂网络以及中国菜肴网络等. 可以强调这些网络中的顶点(参与者,例如公交车站、旅游景点、中药药材、食品原料等)之间的合作,把它们之间的边定义为在项目(公交线路、旅游线路、中药方剂以及中国菜肴等)中的合作关系. 这样,在每个项目中,每一对顶点之间同样都具有表示合作相互作用的边连接,或者说在进行“完全”的合作,以至于每个项目成为一个完全图,网络成为完全图的集合^[9,10]. 我们建议把这些网络与上述的社会合作网络归于一类(也许可以称为“广义合作网络”),来研究它们的共同性质.

本文的目的在于用一个模型来描述广义合作网络的度分布及其可能表示的网络演化机理的特征. 在第二节中将介绍上述的 RDP 模型,并参照该模型建议一个最简化模型,试图通过解析的方法说明这一类网络的一种普遍演化机理,并且说明顶点的度分布 $P(k)$ 和项目度分布 $P(h)$ 之间具有密切的一致关系,而项目含顶点数的分布 $P(T)$ 对度分布的影响较小. 然后将对模型的更一般情况进行数值模拟,希望说明上述结论在定性的意义上具有一定的普遍性. 在第三节中简要地介绍一些实际系统的统计结果,希望说明本模型的正确性. 同时将进行本文的总结.

2. 模 型

2.1. RDP 模型

RDP 模型^[13]设想每步时间演化过程中组建一个包含 T 个参与者的项目(项目含顶点数为 T). T 可能是常数,也可能是遵从项目含顶点数概率分布 $P(T)$ 的随机数. 其中 m 个是新参与者,其余 $T - m$ 个从旧顶点(旧参与者)中按照正比于它的顶点项目度 h 的概率优选. 当仅考虑 T 和 m 均为常数,且不

计重复边的简单情况时,他们解析地证明了在参与者的单粒子投影图中,度分布 $P(k)$ 和顶点项目度分布 $P(h)$ 均为严格的幂函数,而且标度因子相同,都等于 $2 + m/(T - m)$. 这个模型产生的模拟数据与 RDP 对美国演员网和科研合著网的统计结果相当好地符合^[13].

我们在一些实际网络中得到的关于度分布 $P(k)$ 和顶点项目度分布 $P(h)$ 的统计结果倾向于说明 RDP 模型得出的结论是正确的,但是可能只代表一种极端情况. 为此我们借鉴上一节中叙述过的 BA 模型、LL 模型、LC 模型和 RDP 模型,建议一个可能产生幂律、指数函数以及它们之间的各种 SED 分布的度分布和顶点项目度分布的简化模型,这种结果与我们的广义合作网络实际统计结果更加符合. 以下分两步来叙述本文建议的模型.

2.2. 本文的最简化模型

类似地,设初始 $t = 0$ 时有 m_0 个顶点,已经联接成若干个完全图项目,它们的顶点项目度 h_{i0} 之和为 h_0 . 每步时间演化过程增加一个新顶点,在旧顶点中按照一定法则选取 $T - 1$ 个(T 为常数),把这 T 个顶点中两两之间尚未连接的也都连接一条边,构成一个新的完全图项目.

2.2.1. 最简化模型的优选连接情况

设上述选取 $T - 1$ 个旧顶点的法则是选取每个旧顶点 i 的概率正比例于它的顶点项目度 h_i (顶点项目度线性优选法则,即选取演员是看他已经演过多少部影片),则仿照 BA 模型的准连续近似解法^[14],我们列出顶点项目度 h_i 的准连续近似演化方程

$$\frac{\partial h_i}{\partial t} = (T - 1) \frac{h_i}{\sum_j h_j}, \quad (1)$$

由于在 t 够大时, $\sum_j h_j = T(t - 1) + h_0 \approx Tt$, 此方程可以简化为

$$\frac{\partial h_i}{\partial t} \cong (T - 1) \frac{h_i}{Tt}, \quad (2)$$

由此解得 $h_i(t) = Ct^\beta$, 其中 $\beta = (T - 1)/T$, $C = 1/t_i^\beta$ (注意当 $t = t_i$, $h_i(t_i) = 1$). 由此可得 $P(h_i < h) = P(t_i > t/h^{1/\beta})$. 考虑到 $P(t_i) = 1/(m_0 + t)$, 可解得 $t_i < t/h^{1/\beta}$ 的概率为 $t/h^{1/\beta} - 1/(m_0 + t)$. 因此,由 $P(h_i) = \partial P(t/t_i)^\beta < h) \partial h$, 可得到顶点项目度分布为

$$P(h) = \frac{t}{(m_0 + t)\beta} h^{-(\frac{1}{\beta} + 1)}, \quad (3)$$

即顶点项目度分布为幂律, 标度因子 $\gamma = 1/\beta + 1 = (2T - 1)(T - 1)$. 它随常数项目含顶点数 T 的变化而变化, 若 $T = 2, \gamma = 3; T = 3, \gamma = 2.5; T = 5, \gamma = 2.25; T = 7, \gamma = 2.17; T = 9, \gamma = 2.12$; T 趋于无穷时, γ 趋于 2. 如上所述, 当计算重复边时, 总有 $k_i = h_i(T - 1)$, 由此可得重复边的顶点度分布为

$$P(k) \propto k^{-\nu} \quad (4)$$

且 $\nu = \gamma$, 即与 RDP 模型的结论相同, 度分布 $P(k)$ 和顶点项目度分布 $P(h)$ 均为严格的幂函数, 而且标度因子相同. 图 2 和图 3 显示了我们对这个简化情况的数值模拟结果. 在顶点总数 N 不是很大的情况下, 数值模拟结果与上述解析结论仍旧比较好地符合.

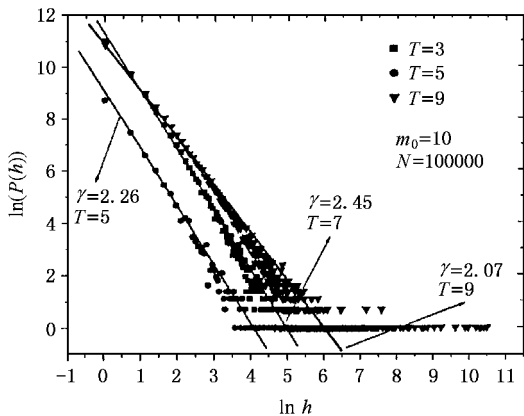


图 2 最简化模型线性优选情况不同项目含顶点数 T 对应的顶点项目度分布(模型计算参数为: $m_0 = 10$, 计算顶点数 $N = 100000$)

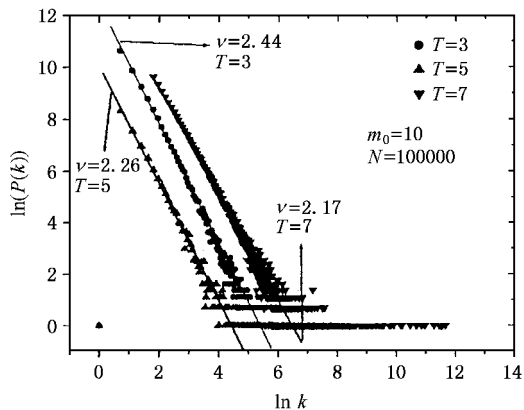


图 3 最简化模型线性优选情况不同项目含顶点数 T 对应的重复边顶点度分布(模型计算参数为: $m_0 = 10$, 计算顶点数 $N = 100000$)

2.2.2. 最简化模型的随机连接情况

若上述选取 $T - 1$ 个旧顶点的法则是随机选取, 则类似地仿照 BA 模型的准连续近似解法^[4], 有

$$\frac{\partial h_i}{\partial t} = (T - 1) \frac{1}{m_0 + t - 1}. \quad (5)$$

类似考虑当 $t = t_i, h_i = 1$ 解得

$$h_i(t) = (T - 1) \ln \frac{m_0 + t - 1}{m_0 + t_i - 1} + 1. \quad (6)$$

由此可得

$$P(h_i < h) = P(t_i > (m_0 + t - 1) \times \exp[(1 - h)(T - 1)] - m_0 - 1).$$

考虑到

$$P(t_i) = 1/(m_0 + t),$$

可解得

$$P(h) = B e^{\frac{1-h}{T-1}}, \quad (7)$$

或

$$\ln P(h) = (T - 1) \ln B + 1 - \frac{h}{T - 1}, \quad (8)$$

其中 $B = (t + m_0 - 1)[(T - 1)(m_0 + t)]$. 即顶点项目度分布为指数规律, 斜率因子为 $\gamma = -1/(T - 1)$. 它也随常数项目含顶点数 T 的变化而变化, T 趋于无穷时, γ 趋于 0. 类似地, 当计算重复边时, 总有 $k_i = h_i(T - 1)$, 由此可得重复边的顶点度分布为

$$\ln P(k) \propto -k/(T - 1). \quad (9)$$

这是 RDP 模型未讨论的另一个极端情况. 这时度分布 $P(k)$ 和顶点项目度分布 $P(h)$ 均为严格的指数函数, 但是斜率因子不同. $P(k)$ 分布的斜率因子为 $P(h)$ 分布的斜率因子的平方. 图 4 和图 5 显示了本文对这个简化情况的数值模拟结果. 在顶点总数

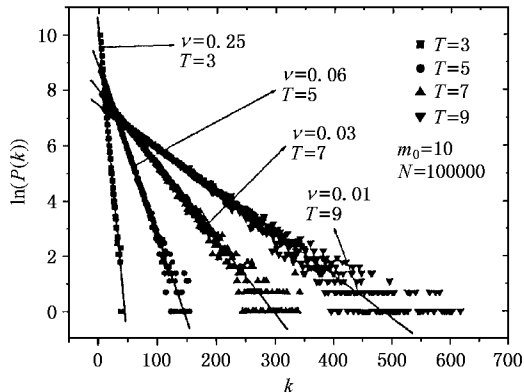


图 4 最简化模型随机情况的顶点项目度分布(模型计算参数为: $m_0 = 10$, 计算顶点数 $N = 100000$)

N 不是很大的情况下,数值模拟结果与上述解析结论也比较好地符合.

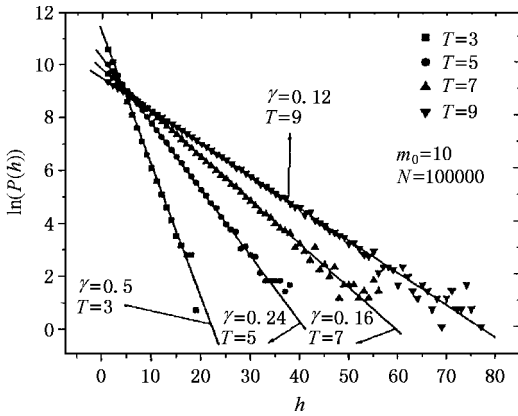


图5 最简化模型随机情况的重复边顶点度分布(模型计算参数为: $m_0=10$,计算顶点数 $N=100000$)

2.2.3. 最简化模型的部分优选连接、部分随机连接情况

如果上述选取 $T-1$ 个旧顶点的法则是以一定的概率 p 随机连接、以其余的概率 $1-p$ 优选连接,容易想到度分布 $P(k)$ 和顶点项目度分布 $P(h)$ 均应该在指数函数和幂函数之间,可以用一个 SED 分布来描述.图6显示了我们对于 $p=0.8$ 的情况的数值模拟结果.对其他情况的数值模拟结果同样显示度分布 $P(k)$ 和顶点项目度分布 $P(h)$ 均为很好的 SED 分布,而且两种分布的关键参数 μ 相同.

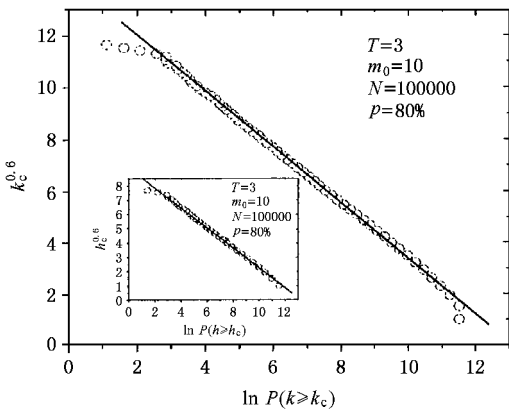


图6 最简化模型部分随机情况($p=0.8$)的顶点度分布(插图显示了相应的顶点项目度分布)

2.3. 本文模型的更近于实际情况

在下一节简短报道的实证统计结果将说明一般情况下 T 并不是常数,在我们调研的系统中(细节将另文报道)项目含顶点数 T 的分布函数 $P(T)$ 都

呈现一个不对称的单峰函数,但是具体的拟合函数形式用一个泊松分布的半支来近似拟合.我们对此情况将提出另外一个模型来讨论^[14].本文仅非常近似地取项目含顶点数 T 的分布函数为 $p_\lambda(T) = \frac{1}{A} \frac{\lambda^T}{T!} \cdot e^{-\lambda}$ 的形式(即泊松分布的半支,希望参照累计分布的实证结果,能够定性地代表实证统计情况.其中 λ 为函数的峰值, T 为大于或等于峰值 λ 的整数, $1/A$ 为归一化因子.在这里 λ 都取 8.5,以做到使泊松分布函数峰值的数量级与我们大多数的实际统计结果(包括本文未报导的一些结果)一致.我们认为参数 A 并不重要,所以在下面各图中也不再注明).本文仅对此情况做数值模拟的研究,说明在此复杂情况下顶点的度分布 $P(k)$ 和项目度分布 $P(h)$ 之间仍定性地具有密切的一致关系,而项目含顶点数的分布 $P(T)$ 对度分布的影响较小.

2.3.1. 优选连接情况

我们根据大多数实际统计结果,令 T 为以上文所述的 $P_\lambda(T)$ 概率分布取值的随机数,然后在旧顶点中线性优选地(即与其顶点项目度正比例地)选取 $T-1$ 个构成一个完全图项目.图7显示了数值模拟结果.结果显示这时顶点项目度分布 $P(h)$ 和度分布 $P(k)$ 都是幂律分布,而且两种分布的幂律因子相同.取 λ 为其他值时结果相同.

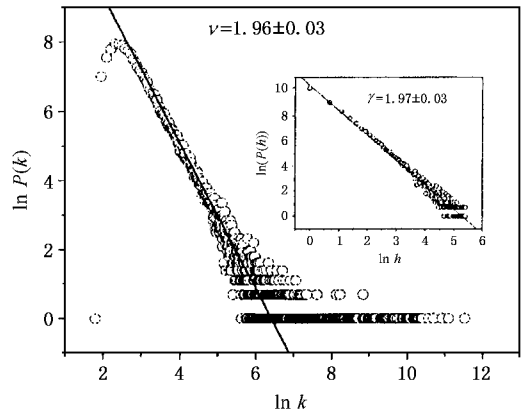


图7 项目含顶点数为 $P_\lambda(T)$ 概率分布($\lambda=8.5$),然后优选的重复边顶点度分布(插图显示了顶点项目度分布.模型计算参数为: $m_0=10$,计算顶点数 $N=100000$)

2.3.2. 随机连接情况

图8显示了我们对于 T 为上述半枝泊松分布,随机选取 $T-1$ 个旧顶点时的数值模拟结果.结果显示这时顶点项目度分布 $P(h)$ 和度分布 $P(k)$ 都是指数函数分布,但是两种分布的斜率因子不相同(仍

旧满足 $\nu \approx \gamma^2$).

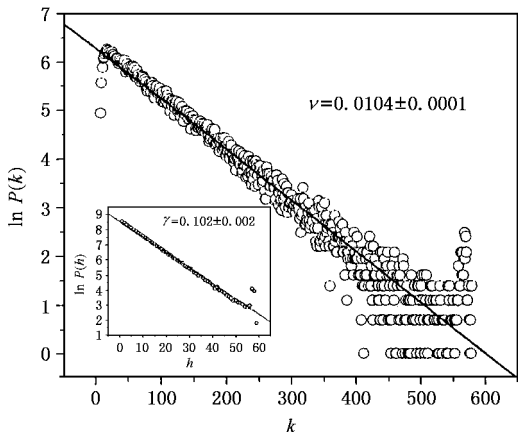


图 8 项目含顶点数为 $P_\lambda(T)$ 概率分布 ($\lambda = 8.5$), 然后随机选择的重复边顶点度分布(插入图显示了顶点项目度分布. 模型计算参数为: $m_0 = 10$, 计算顶点数 $N = 100000$)

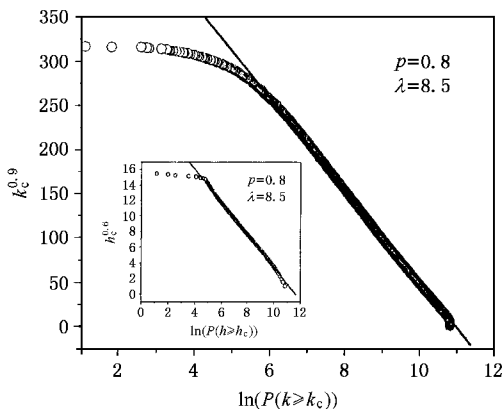


图 9 项目含顶点数为 $P_\lambda(T)$ 概率分布 ($\lambda = 8.5$), 然后部分随机、部分优选的重复边顶点度分布(插入图显示了顶点项目度分布. 模型计算参数为: $m_0 = 10$, 计算顶点数 $N = 100000$, $p = 0.8$)

所作计算的结果定性相同。

图 9 显示了我们对于 T 为上文所述的 $P_\lambda(T)$ 概率分布取值的随机数, 然后在旧顶点中以概率 $p = 0.8$ 随机选取 $1 - p = 0.2$ 线性优选 $T - 1$ 个旧顶点时的数值模拟结果. 结果显示这时顶点项目度分布 $P(h)$ 和度分布 $P(k)$ 都是很好的 SED 分布, 但是两种分布的关键参数 μ 不完全相同. 对概率 p 为其他值

3. 一些实证统计结果

我们对中国旅游线路系统、淮阳菜肴系统、北京和扬州的公交线路系统、以及好莱坞演员的合作系统进行了实证统计研究. 结果证明了上述模型的合理性. 限于篇幅, 本文仅能非常简短地列表报道这些

表 1 一些实证的统计结果

系统	顶点总数	项目总数	边总数	项目含顶点数 累计分布形式	顶点项目度 累计分布形式	μ 参数 (或斜率)	重复边顶点度的 累计分布形式	μ 参数 (或斜率)
2003 年中国旅游线路	171	240	719	泊松分布的半枝	SED 分布	0.5	SED 分布	0.5
淮阳菜肴	242	329	1713	泊松分布的半枝	SED 分布	0.2	SED 分布	0.3
北京 2003 年公交线路	460	65		泊松分布的半枝	近于指数函数分布	0.49	近于指数函数分布	0.03
扬州 2003 年公交线路	289	26		泊松分布的半枝	近于幂函数分布	2.2	近于幂函数分布	1.67
好莱坞演员合作	392304	181455		近似于泊松分布的半枝	SED 分布	0.45	SED 分布	0.45

研究的结论, 详细的情况将另文报道。

在表 1 所列的(从上至下)系统的网络描述中, 规定: 在系统(1)中, 一个旅游线路中的各旅游点为顶点(参与者), 处于同一旅游线路中(同一批游客同一次游览所经历)的两个旅游点之间的合作关系为边, 一个旅游线路对应一个完全图项目. 在系统(2)中, 把一道菜肴看作一个由若干参与者(顶点, 食品)构成的项目, 其中的两两顶点之间都有食品烹饪的相互作用, 用顶点之间的连线(边)来表示. 在系统(3, 4)中, 公共交通工具(多少路(次)车)看作合作项目, 每个合作项目中交通工具停靠的站点定义为顶点(参与者), 两个顶点之间的边定义为它们在同

一条公交线路中的运输合作关系. 在系统(5)中, 定义顶点为演员, 项目为影片. 我们认为这些统计结果与本文建议的模型的结论比较好地符合, 倾向于说明本文建议的模型可以描述广义合作网络的一部分演化机理与特征.

作为结论, 我们建议注意这样的一类系统, 其中许多参与者在参与许多合作项目. 在这些系统的研究中只考虑参与者在项目中的合作关系, 不讨论(或不存在)它们之间的竞争、对抗或其他类型的关系. 我们建议用一个网络来简化地描述这样的系统, 定义参与者在项目中的合作关系为它们之间的边, 这样, 一个项目就可以用一个它包含的参与者两

两连边所构成的完全图来表示. 这样的网络不一定是社会网络, 也可以包含其他许多类网络, 但是由于拓扑结构具有共同特征, 它们的统计性质也具有共同规律. 首先, 项目是这类网络中的主导因素, 项目如何组成是网络演化的关键问题, 参与者的项目度分布(即参与者参与多少个项目的数目的分布)很可能是这类网络最主要的统计性质, 在很大程度上左右顶点的度分布.

由于我们注重远离平衡的复杂系统的研究, 这类网络中各个参与者的信息往往是严重缺乏的. 通常它们只在各自的一个或几个局域世界中比较密切地互相关联, 知道其他参与者的一些信息, 因此也只有在这些范围中才可能按照其他参与者的影响力的大小, 以不同程度的积极性去参加它们所在的项目, 而越出这些局域世界, 就不再具备这样的信息, 只能随机地去参与项目. 这样, 只有在空间和时间的覆盖

范围都很狭小的超小数据网络中, 或者在虽然数据很多, 覆盖范围很大, 但信息流通非常便利、快捷的网络中才可能出现严格的幂律项目度分布和顶点度分布, 一般情况下这类合作网络都显示处在幂律和指数规律之间的, 可能被一个 SED 分布函数更好描述的项目度分布和顶点度分布. SED 分布函数的关键参数 μ 的大小也许有可能描述系统远离平衡的程度以及信息流通的程度. 这些思想需要更深入的研究来修正或支持.

汪秉宏感谢高教博士特别基金(20020358009)的资助. 周涛感谢中国科学技术大学研究生创新基金(KD200408)的支持. 全体作者深切地感谢上海理工大学系统科学研究院车宏安教授多次倡导和组织的讨论, 若没有这些讨论本文的思想不可能产生. 我们同时感谢大连理工大学系统工程研究所章忠志、荣莉莉教授提供他们关于类似题材的非常有启发性的预印文本.

- [1] Albert R, Barabasi A L 2002 *Reviews of Modern Physics* **74** 47
- [2] Newman M E J 2003 *SIAM Review* **45**(2) 167
- [3] Krapivsky P L, Redner S 2003 *Statistical Mechanics of Complex networks*, ed P-Satorras R, Rubi M and D-Guilera A (Berlin Heidelberg: Springer-Verlag) p4
- [4] Barabasi A L, Albert R 1999 *Science* **286** 509
- [5] Liu Z, Lai Y C, Ye N *et al* 2002 *Phys. Lett. A* **303** 337
- [6] Newman M E J 2001 *Phys. Rev. E* **64** 016132
- [7] Li X, Chen G 2003 *Physica A* **328** 274
- [8] Laherrere J, Sornette D 1998 *Eur. Phys. J. B* **2** 525
- [9] He D R, Zhang P P, Xu T, He Y, *Proceeding of the Nationwide Conference on Complex Networks*, Hangzhou, 2004/08, p103 (in Chinese)[何大韧、张培培、许田、何阅, 全国复杂网络会议文集, 杭州 2004/08, p103]
- [10] Sun A, Zhang P, He Y, Su B, He D R 2004 *Bulletin of APS* **49** (1) 1006
- He Y, Zhang P, Su B, He D R 2004 *Bulletin of APS* **49**(1) 1006
- Zhang Y, He Y, He D R 2004 *Bulletin of APS* **49**(1) 1007
- Hu J, Zhang P, Qiu R, Tang J, He D R 2004 *Bulletin of APS* **49** (1) 1008
- [11] Wasserman S, Faust K 1994 *Social Network Analysis: Methods and Applications* (Cambridge: Cambridge University Press)
- [12] He Y, Zhang P P, Xu T, Jiang Y M, He D R 2004 *Acta Phys. Sin.* **53** 1710 (in Chinese)[何阅、张培培、许田、姜玉梅、何大韧 2004 物理学报 **53** 1710]
- [13] Ramasco J J, Dorogavtsev S N, Pastor-Satorras R 2004 *Phys. Rev. E* **70** 036106
- [14] Zhou T, Jin Y D, Wang B H, He D R, Zhang P P, He Y, Su B, Chen K, Preprint

A model describing the degree distribution of collaboration networks^{*}

Zhang Pei-Pei^{1)†} He Yue²⁾ Zhou Tao³⁾ Su Bei-Bei⁴⁾ Chang Hui⁴⁾

Zhou Yue-Ping⁴⁾ Wang Bing-Hong³⁾ He Da-Ren⁴⁾

1) (*Department of Physics , College of Science , Jiangsu University , Zhenjiang 212013 , China*)

2) (*Wuxi No.1 Middle School , Wuxi 214031 , China*)

3) (*Department of Modern Physics , University of Science and Technology of China , Hefei 230026 , China*)

4) (*College of Physics Science and Technology , Yangzhou University , Yangzhou 225002 , China*)

(Received 4 March 2005 ; revised manuscript received 16 June 2005)

Abstract

This paper discusses the degree distribution of a type of social networks and some technological networks with similar topological structures. Firstly , we propose a most simplified model and try to show analytically that the degree distribution and the act degree distribution are well consistent , but the distribution of the number of the nodes inside an act have much smaller influence on the degree distribution. Secondly , we perform numerical simulations in a more general situation of the model , so as to show that the above mentioned conclusion is rather universal. This model leads to a conclusion that these generalized collaboration networks usually have a degree distribution between the power law and the exponential function. In the last part of the article , we introduce very briefly the results of the empirical statistical investigation on some practical collaboration networks to show that the model is valid for generalized collaboration networks.

Keywords : collaboration networks , degree distribution , act degree distribution , number of nodes inside an act

PACC : 0590 , 0175 , 0210

^{*} Project supported by the National Natural Sciences Foundation of China (Grant Nos. 70371071 , 70271070 , 10472116 and 70471033).

[†]E-mail : ppzh@ujs.edu.cn , xueren2001@sohu.com