

基于超图结构的科研合作网络演化模型*

胡枫¹⁾²⁾ 赵海兴^{2)†} 何佳倍³⁾ 李发旭¹⁾²⁾ 李淑玲²⁾ 张子柯^{4)‡}

1) (陕西师范大学计算机科学学院, 西安 710062)

2) (青海师范大学计算机学院, 西宁 810008)

3) (电子科技大学互联网科学中心, 成都 611731)

4) (杭州师范大学信息经济研究所, 杭州 310036)

(2013年4月20日收到; 2013年6月27日收到修改稿)

基于科研论文作者的合作方式, 用超图理论构建了一个科研合作超网络演化模型. 利用平均场理论分析了作者发表论文的演化规律, 发现作者的超度 (即发表论文数) 分布符合幂律分布. 进一步理论分析得到分布的幂指数 γ 与合作领域作者增长速度相关. γ 越大, 新作者增长速度越快, 且存在关系: $\gamma = 1 + \frac{L}{M} \left(\frac{L}{M} \right)$ ($\frac{L}{M}$ 为作者增长率). 并通过对《物理学报》与《中国科学》2003—2012年期间作者发表论文进行了数据分析, 实证结果与理论分析及模拟结果能很好地符合. 本文对科研合作网络的理论和实证研究有一定的借鉴意义.

关键词: 复杂网络, 超图, 科研合作网络, 演化模型

PACS: 89.75.Fb, 02.10.0x, 87.23.Ge, 05.65.+b

DOI: 10.7498/aps.62.198901

1 引言

随着 20 世纪 60 年代由著名的数学家 Erdős 和 Rényi 提出 ER 随机图模型^[1] 开创复杂网络的研究, 20 世纪末 Watts 和 Strogatz 揭示复杂网络的小世界特性^[2], Barabási 和 Albert 揭示复杂网络的无标度特性^[3], 掀起了复杂网络研究的热潮, 成为物理、数学、计算机、生物、化学、经济与社会学等多个学科共同关注的前沿热点. 如今, 复杂网络的研究领域已经拓展到了生命科学的各种网络, 如新陈代谢网、蛋白质网、神经网络等, 技术网络中的因特网、万维网、电力网等, 社会网络中的科研合作网、公司董事网、语言网等等.

在众多社会网络中, 合作网络由于现实世界的普遍性, 一直以来是复杂网络研究的热点之一, 它包含社会关系中的众多方面, 如企业合作网络^[4]、演员合作网络^[5]、科研合作网络^[6]、科研项目合

作网络^[7]、人类性关系网络^[8]等等. 在这些合作网络中人们往往为了各自的利益合作, 企业合作网为了企业本身的利益以及团队的作用等; 演员合作网络关注演员与其他演员之间合作的影片数、演员的影响力和合作人数等, 而科研合作网更多关注的是科学家之间的合作贡献、科研创新能力及合作团队的学术带头作用等等. 科研合作网络作为一种典型的社会网络受到了复杂网络领域众多学者的关注, 不少国内外研究人员对科研合作网络的模型构建、特征分析等做了大量的研究. 如 2001 年 Newman^[9,10] 在 *Physical Review E* 的同一期上连续发表了两篇论文, 报道了对科研合作网络所做的实证分析结果; Barabási 等^[11] 通过数学和神经科学领域内 8 年期间的相关杂志电子数据库的数据分析提出了合作网络的演化模型, 验证了科研合作网络是无标度网络, 网络演化由优先连接机制控制; 周涛等^[12] 提出了合作网络的一般模型, 在该模型

* 科技部 973 前期专项资助项目 (批准号: 2010CB334708)、教育部长江学者与创新团队支持计划 (批准号: IRT1068)、国家自然科学基金地区科学基金 (批准号: 60863006, 61164005)、国家自然科学基金青年科学基金 (批准号: 11105024, 11205040, 1147015)、青海省自然科学基金 (批准号: 2012-Z-943) 和杭州师范大学科研启动基金及“攀登工程”计划资助的课题.

† 通讯作者. E-mail: h.x.zhao@163.com

‡ 通讯作者. E-mail: zhangzike@gmail.com

中,通过调节择优连接参数,可以得到在指数分布与幂律分布之间转换的度分布;何阅等^[7]对项目负责人与助手之间的科研合作建立了双粒子自适应发展模型,对模型的各种不同投影单粒子图进行了数值模拟,其结果与Barabási等^[11]的统计结果可以符合得很好.

到目前科研合作网络的研究中,普遍采用的拓扑结构是Newman在文献[6]中提出的普通图结构,即图中的节点表示作者,合作者之间连边.假设一篇论文有1000个合作者,则每个作者的度大于等于1000.然而度大并不能代表这些作者都有强的合作能力,更不能充分说明这些科学家的科研水平.所以,在普通图中,虽然能很容易计算作者的合作次数(即节点的度),但无法区分一篇论文是两个或多个作者合作完成的关系.解决这个问题也可以采用二部图的方法,用一组不相交的点集表示作者,另一组不相交的点集来表示作者发表的论文.Ramasco等^[5]就提出了一个自组织的合作网络模型(RDP模型),该模型采用了二部图的拓扑结构,在每个时间步内添加一篇新论文,论文中新旧作者的结合采取边优先连接机制,逐步生成了增长的自组织的演化模型,研究了合作网络的度相关性及度簇相关性,但不足之处是RDP模型中每次合作的人数以及新加入合作的人数是固定值,这与实际科研合作网络的演化模型并不相符.另外,在二部图中,由于存在作者和论文两类节点,点的“同质性”失去了.此外,还存在一些网络很难用二部图或多部分图模型表示,如在文献[13]中讨论的含有用户、商品以及标签三类节点的推荐系统,用三部分图是很难完整地刻画该系统所包含的关系.为了解决这个问题,可采用超图方式,其中作者用“节点”代表,而若干个作者合作写的一篇文章用“边”表示,即图里面的“边”可包含若干个节点.关于超图理论的基础研究开始较早,Berge^[14,15]给出了超图理论的基本概念及基本性质,即在普通图中通过在两个节点之间的连接只能表示一对节点之间的关系,而在超图中的“超边”,则可以包含任意多个节点,用来表示这多个节点之间的关系.而基于超图拓扑结构构建的网络称为超网络.近年来,超网络的研究主要集中在基于现实世界的超网络特性的研究.Estrada等^[16]对于复杂超网络的子图中心度和聚集系数进行了研究;Ghoshal等^[17]针对随机三部超图及它们的应用进行了研究;Zlatic等^[18]定义和分析了基于三部超图模型的统计特性;张子柯

等^[19]建立了一种基于用户背景知识和对象、标签双重优先连接机制的超图增长模型;王建伟等^[20]采用超边不断增长和优先连接机制构造了一种超网络的动态演化模型;胡枫等^[21]也给出了一种超网络的动态演化模型,该模型中每次将新增加的一个节点和超网络中已有的若干个节点结合成超边,逐步生成均匀超网络,通过理论和仿真解析得出超网络的几种度分布和一般复杂网络一样,显示为无标度特性.以上给出的几种基于超图的超网络模型通常侧重于研究社会化标签网络,而对科研合作的超网络模型构建的研究迄今仍是空白.因此,本文构建了一个基于非均匀超图结构的科研合作超网络演化模型,在此模型上重点理论分析作者发表文章数量的演化规律,利用计算机仿真分析的方法,验证了理论分析的结果,并通过与《物理学报》与《中国科学》2003—2012十年期间作者发表文章进行了实证分析,其结果与模型结果符合得非常一致.

2 超图的定义^[14]

设 $V = \{v_1, v_2, \dots, v_n\}$ 是一个有限集.若 $E_i \neq \Phi (i = 1, 2, \dots, e)$ 和 $\bigcup_{i=1}^e E_i = V$,则称二元关系 $H = (V, E)$ 为超图.其中 V 的元素 v_1, v_2, \dots, v_n 称为超图的节点或顶点, $E = \{E_1, E_2, \dots, E_e\}$ 为超图的边集合, E 中的元素 $E_i = \{v_{i_1}, v_{i_2}, \dots, v_{i_j}\} (1 \leq j \leq n)$ 称为超图的超边.包含顶点 v_i 的超边的个数称为顶点 v_i 的超度,记为 $d_H(v_i)$.如果两个顶点属于同一条超边,则称这两个顶点邻接;如果两条超边的交集不为空,称为这两条超边邻接.如果每条超边中的顶点数都相等,则称为均匀超图或一致超图.

超图 H 用图形来表示^[14],即由点的集合表示 V 中的元素,超边则用一条包含该超边中所有顶点的简单闭合曲线表示.如图1所示的超图 H 中,顶点集 $V = \{v_1, v_2, v_3, v_4, v_5, v_6\}$,超边集 $E = \{E_1, E_2, E_3, E_4, E_5\}$,其中 $E_1 = \{v_1, v_2, v_6\}$, $E_2 = \{v_2, v_3, v_4\}$, $E_3 = \{v_4, v_5\}$, $E_4 = \{v_5, v_6\}$, $E_5 = \{v_6\}$.图1中顶点 $v_1, v_2, v_3, v_4, v_5, v_6$ 的超度分别是1, 2, 1, 2, 2, 3.

3 科研合作超网络演化模型

科研合作网络是描述作者科研合作关系的网络,顶点表示作者,作者之间通过合作写论文来建

立连接. 这里我们给出科研合作的超网络表示, 顶点仍然表示作者, 一篇论文可能有多个作者, 作者之间合作一篇论文就表示为超网络中的一条超边, 每条超边中的多个顶点即合作同一篇论文的作者之间采取全连接的方式, 表示这多个作者是相互合作的. 不同的论文中, 可能有相同的作者, 即不同的超边之间通过公共的顶点邻接. 因此许多篇科研论文根据相同的作者就会构成一个作者合作论文关系的超网络. 在此基础上, 本文的超网络演化模型的构建过程如下:

1) 初始化

假设初始时网络中有 n_0 个作者, 合作写的一篇论文组成一条超边.

2) 超边增长

按照时间 $t(t = 1, 2, \dots)$ 的变化, 在每个时间步内添加一篇论文, 添加规则如下:

(i) 以概率 p_i 表示新增论文中有 i 个作者的概率 (合作撰写了该篇论文), 其中 $\sum_{i=1}^m p_i = 1$ (假设论文中作者的人数最多为 m 个).

(ii) 以概率 q_j 表示新增论文中的作者中与原有网络中的 $j(j = 0, 1, 2, \dots, m)$ 个旧作者相同, 其中 $\sum_{j=0}^m q_j = 1$. 则 $j = 0$ 表示论文中的作者全是新作者, $j = m$ 表示论文中的作者全是旧作者.

(iii) 基于超度的择优连接

我们在选取合作作者时, 通常倾向于选取发表

论文数量多的作者, 因此每个时间步内添加一篇论文 (超边) 时, 该论文中选取与原有网络中的旧作者 i 连接, 采用超度优先连接方式, 每次选取连接的节点 i 的概率 $\Pi(i)$ 等于节点 i 的超度 $d_H(i)$ 与超网络中的已有节点 j 的超度 $d_H(j)$ 总和之比, 即

$$\Pi(i) = \frac{d_H(i)}{\sum_j d_H(j)}, \quad (1)$$

其中分子 $d_H(i)$ 表示节点 i 的超度即作者 i 发表的论文数量, 分母表示所有节点的超度之和, 即为网络中已有的所有作者发表的论文总数. 图 2 模拟了增加到 4 篇论文的演化过程, 其中实心圆点表示旧作者, 空心圆点表示新加入的作者, 虚曲线表示新添加的论文.

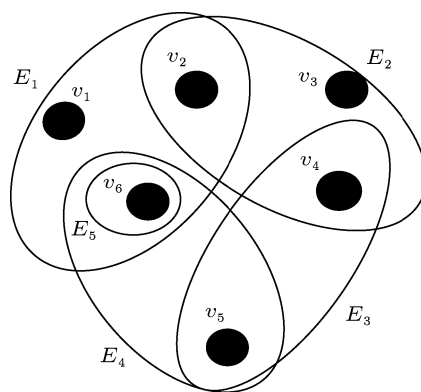


图 1 超图 H 的一般表示

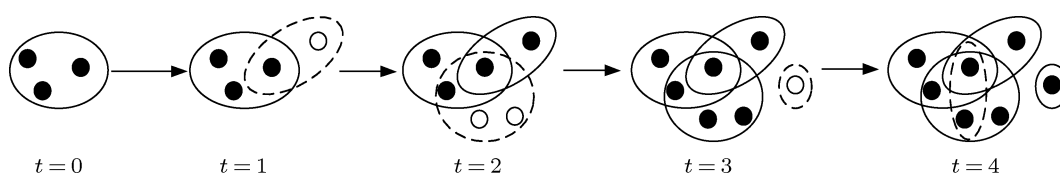


图 2 科研合作超网络的演化过程示意图

4 作者发表论文数量分析

4.1 数据分析

我们通过对《物理学报》与《中国科学》中文版 2003.1—2012.12 十年期间每种刊物发表论文中作者数量进行统计, 结果如表 1 所示.

从表中分析《中国科学》中 A 刊合作者为 2 人最多, E, F, G 刊的合作者是 3 人的最多, 而 B, C, D 合作者 4 人及以上的偏多, 《物理学报》也是合作者 4 人偏多. 根据表 1 数据绘制出的合作者人

数分布图如图 3 所示, 插图为累计分布图, 表现为明显的单峰指数衰减函数, 这与周涛等^[12]的分析是一致的. 分析在不同学科中合作者所占比例不同的原因, 数学属于理论研究, 需要的合作者较少, 而化学、生命科学和地球科学需要更多的实践和实验分析, 所以合作者偏多, 物理、信息科学和技术科学介于二者之间.

4.2 平均场理论分析

在科研合作超网络中, 我们通常会关注一个作

者发了多少篇论文(即节点的超度值),以此来判断作者的科研水平、能力以及他的知名度,超度也会影响其他作者以后是否选择与他的合作,我们当

然有意识或无意识的选择发表论文较多、知名度较高的作者合作.下面利用平均场理论对作者发表论文的演化规律进行理论分析.

表 1 << 物理学报 >>、<< 中国科学 >>2003.1—2012.12 期间发表论文中合作者的数量统计表

论文篇数/人	<< 物理学报 >>	<< 中国科学 >> 中文版刊物							统计总数	所占比例/%
		A 辑	B 辑	C 辑	D 辑	E 辑	F 辑	G 辑		
1	487	211	51	36	61	74	16	91	1027	5
2	1800	458	142	94	202	294	105	232	3327	16
3	2464	205	252	139	298	449	168	294	4269	21
4	2499	32	284	152	310	374	153	240	4044	20
5	1926	4	256	151	357	221	76	161	3152	15
6	1343	3	180	130	248	99	23	115	2141	10
7	687	0	119	109	154	59	15	54	1197	6
8	425	0	48	69	76	23	4	44	689	3
9	220	0	18	57	29	11	3	12	350	2
≥ 10	278	0	17	83	59	15	1	43	496	2
总论文数	12129				8563				20692	

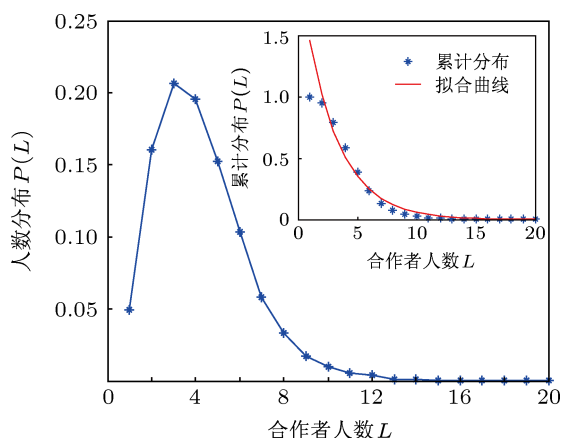


图 3 论文中合作者人数分布图

平均场理论在本文中体现的主要思想 [22] 是:首先建立节点 i 的超度 $d_H(i)$ 满足动力学方程,其次令 i 服从均匀分布,反映节点 i 随机选择,然后求出概率分布函数 $P_H(d_H, t)$, 并取极限得到最终的概率分布 $P_H(d_H)$.

根据以上超网络模型的构建算法,在每个时间步 t 内,添加一篇论文(即超边),论文中有 $l(l=1, 2, \dots, m)$ 个作者(由概率 p_l 决定),在 l 个作者中有 $j(j=0, 1, 2, \dots, m)$ 个旧作者(由概率 q_j 决定),引起了节点 i 的超度变化,节点 i 的超度 $d_H(t)$ 满足动力学方程

$$\frac{\partial d_H}{\partial t} = \left(\sum_{l=1}^m p_l \left(\sum_{j=0}^l q_j j \right) \right) \Pi(i), \quad (2)$$

其中 $\Pi(i) = \frac{d_H(i)}{\sum_j d_H(j)}$, 分母是 t 时刻所有节点的超度之和,即为 t 时刻所有作者发表的论文数之和,所以

$$\sum_j d_H(j) = n_0 + t \left(\sum_{l=1}^m p_l l \right), \text{ 令}$$

$$\sum_{j=0}^l q_j j = J, \quad \sum_{l=1}^m p_l J = M, \quad \sum_{l=1}^m p_l l = L,$$

则 (2) 式可化为

$$\frac{\partial d_H}{\partial t} = M \Pi(i) = M \frac{d_H}{n_0 + tL}, \quad (3)$$

当 t 充分大时,

$$\frac{\partial d_H}{\partial t} \approx \frac{M}{L} \cdot \frac{d_H}{t}. \quad (4)$$

每个节点加入超网络时节点的超度初始值 $d_H(t_i) = 1$, 则解方程 (4) 可得 $d_H(t) = \left(\frac{t}{t_i} \right)^{M/L}$, 由于超网络中加入超边的节点是随机选择的,因此节点具有超度 d_H 的概率为

$$P_H(d_H(t) < d_H) = P_H \left(t_i > \frac{t}{d_H^{L/M}} \right). \quad (5)$$

假设在相同的时间间隔,新添加的超边服从均匀分布,即 t_i 值具有常数的概率密度 $\rho(t_i) = \frac{1}{t}$, 从

而代入方程 (5) 得

$$P_H\left(t_i > \frac{t}{d_H^{L/M}}\right) = 1 - P_H\left(t_i \leq \frac{t}{d_H^{L/M}}\right) = 1 - \frac{t}{d_H^{L/M} \cdot t} \quad (6)$$

对上式求导, 则网络的瞬时超度分布 $P_H(d_H, t)$ 为

$$P_H(d_H, t) = \frac{\partial P(d_H(t) < d_H)}{\partial d_H} = \frac{L}{M} d_H^{-(1+L/M)},$$

即节点超度分布 $P_H(d_H) = \frac{L}{M} d_H^{-(1+L/M)}$.

这说明节点的超度分布和很多复杂网络模型的度分布一样符合幂律分布^[3-6], 具有无标度特性. 幂指数 γ 为 $1 + \frac{L}{M}$. γ 越大, 表示该期刊的新作者增加速率越快.

4.3 蒙特卡洛数值模拟分析

依据前面的模型构建算法, 用蒙特卡洛数值模拟方法生成论文中作者数量和连接的旧作者数量, 反复生成时间序列, 逐步构建科研合作超网络演化模型, 用计算机模拟随时间变化的作者发表论文数的分布特性.

假设初始时只有一篇论文, 作者数量为 4, 随着时间 t 的变化, 每个时间步内添加一篇论文 (假设论文的人数不超过 10), 根据表 1 中作者的合作人数统计, 论文中作者人数的概率取值为 $p_1 = 0.05, p_2 = 0.16, p_3 = 0.21, p_4 = 0.20, p_5 = 0.15, p_6 = 0.10, p_7 = 0.06, p_8 = 0.03, p_9 = 0.02, p_{10} = 0.02$. 选取连接的旧作者的人数概率取两组值. (a) 组的 $q_0 = 0.25, q_1 = 0.30, q_2 = 0.30, q_3 = 0.10, q_4 = 0.05$; (b) 组的概率取值为 $q_0 = 0.10, q_1 = 0.15, q_2 = 0.25, q_3 = 0.25, q_4 = 0.15, q_5 = 0.10$. 我们用计算机模拟了论文总数 N 为 20000, 在选取两组不同旧作者时的作者发表论文数双对数坐标下的分布图 (如图 4 对应 (a) 组, 图 4(b) 对应 (b) 组). 作者发表的论文数量分布遵循幂律分布, 显示了无标度特性, 其幂指数分别是 $\gamma \approx 3.2$ 和 $\gamma \approx 2.6$.

4.4 实证数据分析

为了应用实证数据验证本文的演化模型, 我们通过知网统计了 2003.1—2012.12 年十年期间发表在《物理学报》与《中国科学》A, B, C, D, E, F, G 辑上作者发表论文数量. 如表 2 所示.

根据表 2 的数据绘制出如图 5 所示双对数坐标下的《物理学报》及《中国科学》C, G 辑及

综合后的作者发表的论文数量分布图, 其中《物理学报》幂指数 $\gamma \approx 2.6$. 《中国科学》幂指数 $\gamma \approx 3.2$. 这一结果和图 4 的理论及模拟结果一致.

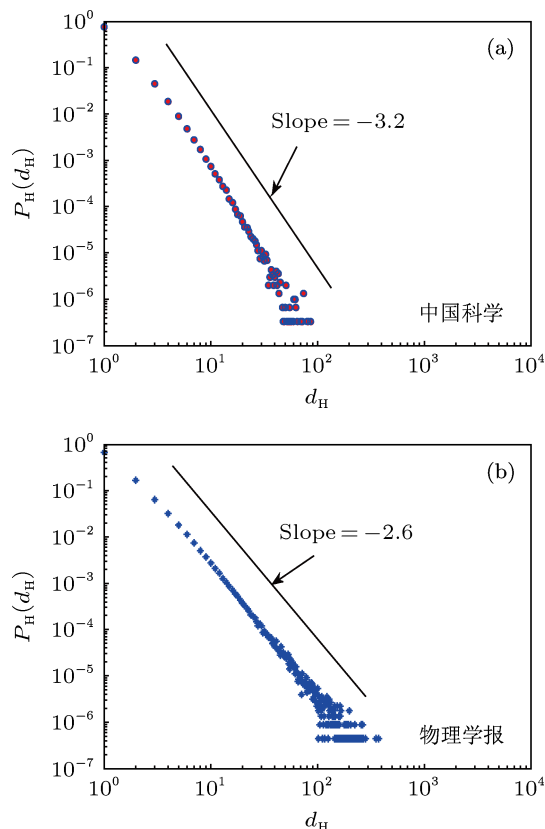


图 4 不同参数下模型演化结果与理论分析的超度分布图 (其中参数 p 取值均相同, 参数 q 取值分别为 (a) $q_0 = 0.25, q_1 = 0.30, q_2 = 0.30, q_3 = 0.10, q_4 = 0.05$; (b) $q_0 = 0.10, q_1 = 0.15, q_2 = 0.25, q_3 = 0.25, q_4 = 0.15, q_5 = 0.10$. 模拟结果均为 50 次仿真结果的平均值)

4.5 理论结果比较分析

理论分析得到作者发表论文数量 (超度) 的分布指数 $\gamma = 1 + \frac{L}{M}$. 由 4.2 节中的定义: $L = \sum_{l=1}^m p_l l$ 表示在 t 时刻添加的一篇论文中有 L 个作者, $M = \sum_{l=1}^m p_l \sum_{j=0}^l q_j$, j 表示在 t 时刻添加的一篇论文 L 个作者中有 M 个是已有合作网络中的作者数目. 显然有 $L \geq M$. 极值情况下, 当 $L = M$ 时, 表示论文中的合作者都是原有作者, 其作者超度分布的指数达到最小值 $\gamma=2$; 当 $L > M$ 时, M 所占比例越大, γ 值越小, 即合作者中原有作者数量越多, γ 越接近于 2 (但不低于 2); 相反, 如果演化过程中新作者数越多, 也即参与该领域的新作者数目增长速度越快, M 就越小, γ 就越大. 而且, 因为 $\frac{L}{M} \leq L \leq m$ (m 为所有论文中最多的合作者数量), 则 $2 < \gamma \leq 1 + m$.

表2 《物理学报》与《中国科学》2003.1—2012.12期间作者发表的论文数量统计表

作者数量/篇	《物理学报》	所占比例/%	《中国科学》中文版刊物									
			A辑	B辑	C辑	D辑	E辑	F辑	G辑	综合	比例/%	
1	12565	58.02	1091	3837	3801	3641	3690	1373	3027	20460	78.91	
2	4015	18.54	198	557	489	925	575	194	521	3459	13.34	
3	1780	8.22	67	188	130	308	168	35	156	1052	4.06	
4	962	4.44%	27	76	37	168	59	16	69	452	1.74	
5	658	3.04	13	25	13	96	34	7	34	222	0.86	
6	407	1.88	5	16	8	46	13	5	14	107	0.42	
7	264	1.22	1	7	9	24	9	2	21	73	0.28	
8	224	1.03	0	3	2	18	8	2	1	34	0.13	
9	164	0.76	1	2	1	10	5	1	2	22	0.09	
10	120	0.55	0	2	3	5	2	0	1	13	0.05	
11	80	0.37	0	0	0	9	1	0	1	11	0.04	
12	72	0.33	0	0	1	3	1	0	2	7	0.02	
≥13	347	1.60	0	4	2	8	1	0	2	27	0.06	
作者总数	21658									25929		

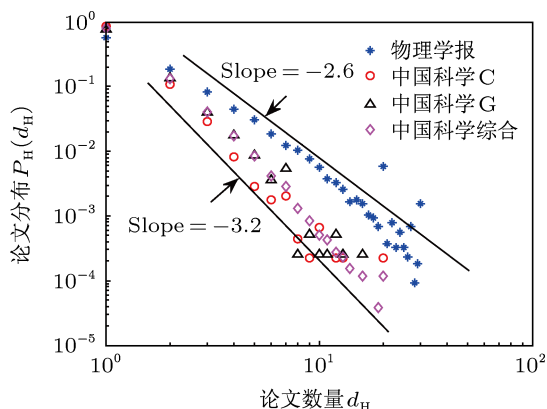


图5 作者超度(即发表论文数量)分布实证图

为了检验作者增长率与作者超度分布指数的关联关系, 本文进一步实证分析了《中国科学-综合》和《物理学报》的数据. 由于在模型的演化过程中, L 和 M 是动态变化的. 因此, 本文按照论文的发表顺序将作者排序. 发表在其中同一期的论文则按照发表页码排序, 同一篇论文中的作者按照作者序排序. 实证数据分析结果如图6所示. 图6(a)显示《中国科学-综合》的作者增长率(即 $\frac{L}{M}$)为2.24, 换算成作者的超度分布指数 $\gamma = \frac{L}{M} + 1 = 3.24$, 这和实际的指数 $\gamma = 3.2$ 非常一致; 类似地, 图6(b)显示《物理学报》的作者增长率(即 $\frac{L}{M}$)为1.62, 换算成作者的超度分布指数 $\gamma = \frac{L}{M} + 1 = 2.62$, 这和实际的指数 $\gamma = 2.6$ 也非常符合.

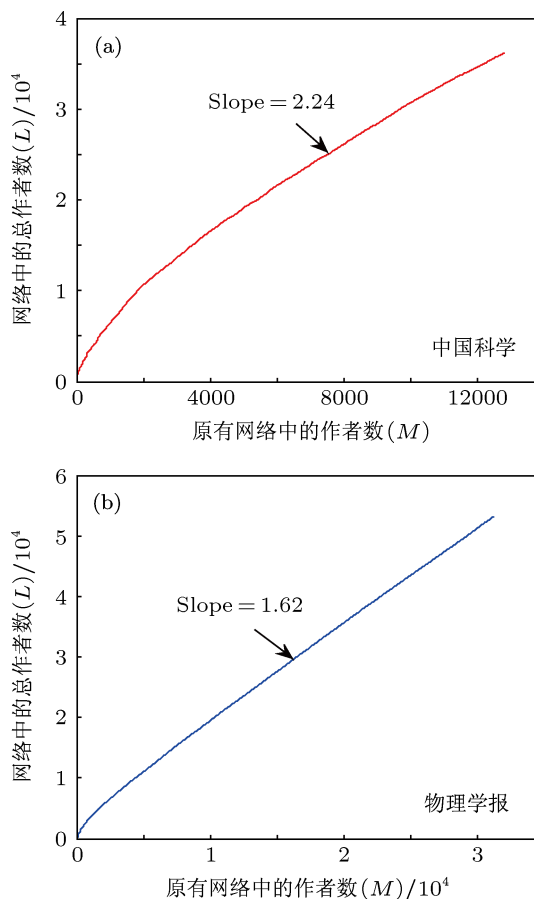


图6 原网络中作者(M)与累计作者总数(L)的实证关系 (a) 中国科学-综合; (b) 物理学报

5 结论

本文根据超图理论, 基于科研论文的作者合作

关系, 建立了一个科研合作超网络演化模型, 在此模型上通过理论分析作者发表论文的演化规律, 发现作者的超度分布 (即发表的论文数量) 遵循幂律分布, 此结果与蒙特卡洛仿真实验和实证分析的结果一致, 且理论分布指数 $\gamma = 1 + \frac{L}{M}$. 同时, 我们也得出 γ 值与领域的作者增长率相关, γ 越大, 则新发表的论文中新作者比例越大. 这一关系也得到了来自《物理学报》与《中国科学》2003—2012 年

实证数据的支持. 虽然本文模型是针对科研合作网络提出的, 实际上该模型可适用于各种的合作网络, 其表现的特性显示了合作网络的普适性, 即个体参与若干个小团体的行为表现为无标度特性, 因此, 对一类超网络的演化问题有一定的借鉴意义.

感谢青海师范大学数学系陈占寿教授和杭州师范大学信息经济研究所刘闯博士的讨论及分析.

- [1] Erdős P, Rényi A 1960 *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17
- [2] Watts D, Strogates S H 1998 *Nature* **393** 440
- [3] Barabási A L, Albert R 1999 *Science* **286** 509
- [4] Davis G F, Greve H R 1997 *Amer. J. Sociol.* **103** 1
- [5] Ramasco J J, Dorogovtsev S N, Pastor-Satorras R 2004 *Phys. Rev. E* **64** 016131
- [6] Newman M E J 2001 *Proc. Natl. Acad. Sci. U.S.A.* **98** 404
- [7] 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]
- [8] Liljeros F, Edling C R, Amaral L A N, Stanley H E, Aberg Y 2001 *Nature* **411** 907
- [9] Newman M E J 2001 *Phys. Rev. E* **64** 016131
- [10] Newman M E J 2001 *Phys. Rev. E* **64** 016132
- [11] Barabási A L, Jeong H, Néda Z, Ravasz E, Schubert A, Vicsek T 2001 *Physica A* **311** 590
- [12] Zhou T, Jin Y D, Wang B H, He D R, Zhang P P, He Y, Su B B, Chen K, Zhang Z Z 2007 *Int. J. Mod. Phys. C* **18** 297
- [13] Lü L, Medo M, Yeung C H, Zhang Y C, Zhang Z K, Zhou T 2012 *Phys. Rep.* **519** 1
- [14] Berge C 1973 *Graphs and Hypergraphs* (vol. 6) (New York: Elsevier) p389
- [15] Berge C 1989 *Hypergraphs: Combinatorics of Finite Sets* (Vol. 45) (Amsterdam: North-Holland Holl) p3
- [16] Estrada E, Juan A J, Rodríguez V 2006 *Physica A* **364** 581
- [17] Ghoshal G, Zlatic, Caldarelli G, Newman M E J 2009 *Phys. Rev. E* **79** 066118
- [18] Zlatic V, Ghoshal G, Caldarelli G 2009 *Phys. Rev. E* **80** 036118
- [19] Zhang Z K, Liu C 2010 *J. Stat. Mech.* P10005
- [20] Wang J W, Rong L L, Deng Q H, Zhang J Y 2010 *Eur. Phys. J. B* **77** 493
- [21] Hu F, Zhao H X, Ma X J 2013 *Sci. Sin. Phys. Mech. Astron.* **43** 16 (in Chinese) [胡枫, 赵海兴, 马秀娟 2013 中国科学: 物理学力学天文学 **43** 16]
- [22] Shi D H 2011 *Theory of network Degree Distributions* (Vol. 1) (Beijing: Higher Education Press) p60 (in Chinese) [史定华 2011 网络度分布理论 (第 1 版) (北京: 高等教育出版社) 第 60 页]

An evolving model for hypergraph-structure-based scientific collaboration networks*

Hu Feng¹⁾²⁾ Zhao Hai-Xing^{2)†} He Jia-Bei³⁾ Li Fa-Xu¹⁾²⁾
Li Shu-Ling²⁾ Zhang Zi-Ke^{4)‡}

1) (School of Computer Science, Shaanxi Normal University, Xi'an 710062, China)

2) (School of Computer, Qinghai Normal University, Xining 810008, China)

3) (Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, China)

4) (Institute of Information Economy, Hangzhou Normal University, Hangzhou 310036, China)

(Received 20 April 2013; revised manuscript received 27 June 2013)

Abstract

In this paper, by considering the collaborative scientific behaviors, a model for evolving hyper-network is proposed based on the hypergraph theory. We study analytically the dynamics of authors' hyper-degree using mean-field approach. Results of both theoretical and numerical studies indicate that such a hyper-degree distribution follows a power-law decay, with the exponent $\gamma = 1 + \frac{L}{M}$, where $\frac{L}{M}$ suggests the author growth rate of an evolving scientific network. In addition, empirical studies on ten-year journal data of *Science China* and *Acta Physica Sinica* from 2003 to 2012, shows that the present model can find good agreement with the empirical data. The proposed model may shed some light on the in-depth understanding of the structure and scientific collaboration networks.

Keywords: complex networks, hypergraphs, scientific collaboration, evolving model

PACS: 89.75.Fb, 02.10.0x, 87.23.Ge, 05.65.+b

DOI: 10.7498/aps.62.198901

* Project supported by the National Basic Research Program of China (No. 2010CB334708), the Program for Changjiang Scholars and Innovative Research Team in Universities (No. IRT1068), the Fund for Less Developed Regions of the National Natural Science Foundation of China (Grant Nos. 60863006, 61164005), the Yong Scientists Fund of the National Natural Science Foundation of China (Grant Nos. 11105024, 11205040, 1147015), the Nature Science Foundation from Qinghai Province (No. 2012-Z-943), and the start-up research funding and Pandeng Project of Hangzhou Normal University.

† Corresponding author. E-mail: h.x.zhao@163.com

‡ Corresponding author. E-mail: zhangzike@gmail.com