

科研合作网络的重要作者发现*

□ 朱天 吴斌 王柏 / 北京邮电大学北京市智能通信软件与多媒体重点实验室 北京 100876

摘要:近年来,使用复杂网络理论对文献的科研合作网进行分析得到了广泛的研究。文章对DBLP数据库中1998年至2007年的作者合作数据构造科研合作网络,并且通过复杂网络的基本统计度量,如度、聚集系数等对网络的总体面貌进行了宏观上的描述。在微观层面,文章提出了高效的重要作者发现算法,能够从作者的合作数量以及合作范围对重要作者进行排名。通过分析科研合作数据作者的影响力,从而为科研人才评价提供参考。

关键词: 社会网络分析, 科研合作网, 核心节点发现

DOI: 10.3772/j.issn.1673-2286.2010.08.005

1 引言

现在信息技术发展日趋全球化,科研合作成为科学研究的主流方式,科研合作的数量与日剧增。至今为止,已经有很多理论和技术被引入到科研文献分析中,复杂网络是其中应用比较广泛的一种。复杂网络研究的兴起时间并不长,但人们对它的研究方兴未艾。其应用领域涉及工程技术、社会、政治、医药、经济、管理等不同方面。广泛的应用前景使得复杂网络的研究倍受国内外的密切关注,引起了不同学科的高度重视。从国际范围看,有关复杂网络的各种形式的学术研讨会越来越多,近几年的*Nature*、*Science*杂志上都有关于复杂网络结构发现及网络演化等方面的文章,NetSci、KDD等重要会议论文集也涵盖了大量有关复杂网络的文章。美国NSF等机构近几年也对网络科学相关项目给予了积极的

支持。在国内,包括中国科学院、中国科学技术大学、上海交通大学、武汉大学、北京邮电大学等许多院校都专门成立了复杂网络研究中心和研究小组。

使用复杂网络理论研究社会网络结构时,一般从网络的静态特征入手。一些经典的统计方法有度分布特征(Degree Distribution)、聚集系数(Cluster Coefficient)等。这些静态统计能够反映网络的某些重要特性,如“小世界特性”(Small World)、“富人俱乐部”(Rich Club)等^[1]。本文以计算机领域的研究成果为基础,分析了计算机领域DBLP数据库收录的从1998年至2007年的论文,根据作者之间的合作关系,建立科研合作网络。每一位作者表示为科研合作网络中的节点,不同作者之间的合作(即公开发表共同署名的学术论文)关系表示为网络中的边,作者间的合作次数为网络中边的权值,建立了一个

无向加权科研合作网络^[2-5]。文章对该网络进行了初步研究,将它的静态结构特征与过去对文献科研合作网络的研究成果进行了比较,虽然一些特性与其他领域的网络特征不同,但仍然具有比较明显的无尺度网络特性。

本文还进一步对该科研合作网络进行了核心节点分析,也称节点的中心性分析。节点中心性也是网络分析的最基本概念之一。这些年来,研究人员依据不同的标准提出了许多中心性指标来判定网络中哪些节点比其他节点更重要。这些指标已经被广泛应用于各种领域,帮助研究人员分析和理解节点在各种类型网络中的重要性。根据节点的重要性,研究人员能针对不同类型的节点制定出不同的方案进行研究。在科研合作网络中,人们利用中心性分析方法可以检测出合作数量最多、最广泛和最具有影响力的作者。然而,仅依赖于节点的度,

* 本文得到国家自然科学基金项目(90924029)和国家“十一五”科技支撑计划项目(2006BAH03B05)资助。

只能得到合作数量最多的作者，不足以判断作者的影响力。因此，本文针对重要作者满足合作次数多、范围广的特性提出了新的重要作者发现算法。该算法复杂度低，能够适用于大型文献网络。

2 相关工作

科研合作网是一个社会网络，Barabási等人研究了数学领域与神经科学领域的科研合作网络，他们在参考文献[6]中指出，该网络具有以下特点：（1）节点的度分布符合power-law分布。（2）节点间的聚集系数随时间降低，节点间的平均最短路径随时间减小。（3）网络是不断增长的。随着时间的推移，新的作者加入到这个网络中，网络中的节点在不断增加，一个作者的合作者也会越来越多。（4）优先连接。因为一个作者的度越大，他的科研能力和影响力会越来越大，其他节点更趋向于与这个作者合作（相连接）。（5）网络的最大聚类大小逐渐增大。Newman在参考文献[2]中从统计学的几个基本特征对生物学领域、数学领域和生物医学领域的科研合作网进行了对比分析。Huang等人在参考文献[7]中则对计算机领域的科研合作网络的演化进行了分析和建模，他们对科研合作网进行了网络层次、社团层次、单个节点层次等三层分析，结果指出科研合作网符合复杂网络的小世界特性；对社团的演化研究结果显示了6个不同研究主题的社团存在不同的作者合作模式；在对节点的研究中，作者还提出了一个新的模型SPOT，根据作者当前的邻居结构预测作者间将来的合作关系。

刘则渊^[8-10]等人发表论文，论述了复杂网络分析方法及其在合作研究中的应用及前景，构建了《科学计量学》的引文网络，使用UCINET对其进行了分析，发现这个网络是一个小世界无标度网络。李鹏翔等^[11]发表《一个师生学术群体的网络研究案例》，对咨询、沟通、科研合作和友谊网络进行了核心和外围结构的分析，考察了群体成员获得帮助的主要途径。王福生等人^[12]对《情报学报》从2001年1月到2006年12月期间发表的论文进行科研合作网分析，从节点的最大度以及最大介数的角度对作者的影响力进行分析，得到了一些研究结果。杨胜琦等人^[13]对生命科学领域近五年的论文引文网络进行分析，使用两种不同的网络构造方法形成了有向引文网络，以及无向同引网络，对这两种网络进行了社团划分以及社团演化研究。Wang Chi等人^[14]对科研合作网络进行建模，以发现合作者之间的导师与学生的关系。Erjia Yan等人^[15]通过研究中国的图书馆和信息科学（LIS）中18个核心期刊六年的科研合作数据，分析了科研合作网络的作者合作模式和网络结构，通过分析发现，构建的科研合作网络符合小世界特征。他们还比较了用以发现核心作者的几种基本的中心性发现方法，以及这些方法的缺陷。唐杰等人^[16-18]开发了科技文献数据挖掘工具ArnetMiner，并且在文献[16]中介绍了使用ArnetMiner分析和挖掘文献网络的方法和结果。ArnetMiner也提供了重要作者发现的功能，主要考虑了作者间的合作次数。也就是说，他们假设一个作者与其他作者的合作次数越多就越重要。虽然Arnetminer列出了一些作

者重要性评价指标用于领域专家排名，例如：引用次数（citation）、发文数量（publication）、H-指数（H-index）、科研持续度（Longevity）、活跃度（Activity）、多元性（Diversity）等，但多数都是统计指标，并且都没有从科研合作这个角度进行评估。尽管有社会合作度量（Sociability）涉及合作的指标，但对合作网分析相对较浅，主要考虑了作者间的合作次数。本文主要是从合作网的角度出发，提出基于社会网络的新的作者重要性度量。

3 计算机领域科研合作网分析

3.1 科研合作网基本结构的分析

DBLP是数字书目索引与图书馆项目（Digital Bibliography & Library Project）的缩写。DBLP数据库提供了计算机领域科学文献的搜索服务，它储存这些文献的相关元数据，如标题、作者、发表日期等。本文使用了1998年至2007年DBLP收录的论文形成的作者科研合作网。

表1中将DBLP科研合作网与其他科研合作网进行了比较。从表中可知，DBLP科研合作网符合小世界特性。它的静态结构特性与同是计算机领域的CiteSeer、NCSTRL合作网络比较接近。作者平均发表论文数与论文平均作者数都较小，低于物理与生物医学领域；网络的平均度大于CiteSeer、NCSTRL合作网以及数学领域合作网，小于物理学、生物医学与神经科学领域。聚集系数与CiteSeer相近，小于物理

表1 合作网络静态结构特性比较

合作网	DBLP	CiteSeer	NCSTRL	Maths	SPIRES	MEDLINE	NeuroSci
参考文献	本文	[7]	[5]	[6]	[5]	[5]	[6]
#论文	502,947	451,305	13,169	70,901	66,652	2,163,923	210,750
#作者 (N)	398,455	283,174	11,994	70,975	56,627	1,520,251	209,293
作者平均发表论文数	3.545	4.06	2.55	-	11.6	6.4	-
论文平均作者数	2.808	2.55	2.22	-	8.96	3.75	-
平均距离	7.0	7.1	9.7	9.5	4.0	4.6	6
平均度(<k>)	6.277	5.56	3.59	3.9	173	18.1	11.54
最大度(k)	460	291	10.7	120	1200	5800	400
聚集系数(C)	0.636	0.634	0.496	0.59	0.726	0.066	0.76

与神经科学领域，略大于NCSTRL以及数学领域，远大于医学领域。与Erdős-Rényi随机网络相比，它的聚类系数较高，这就意味着与相同作者有过科研合作的作者之间倾向于将来有更多的合作。此外，根据表中显示，DBLP合作网络的作者之间的平均距离为7.0，与复杂网络著名的六度分离论吻合，证明六度分离在文献的科研合作网络依然有效。进一步地，网络中作者间平均距离的对数比也约为7.0，进一步验证了该合作网络的小世界特性：

$$\frac{\log(N)}{\log(\langle k \rangle)} = 7.02 \approx l_{real} \quad (1)$$

图1展示了每一年的作者数、不考虑重复合作的作者合作数量、考虑重复合作的作者合作数量、科研合作网络的作者最大度，以及作者平均度。由图中可知，随着时间的推移，作者数、作者合作总数、网络的聚集系数都在不断增长，说明投身计算机领域科研工作的人员在不断增多，科研成果也在不断增加，人们倾向于更紧密地合作。

图2中以2000年为例，展示了

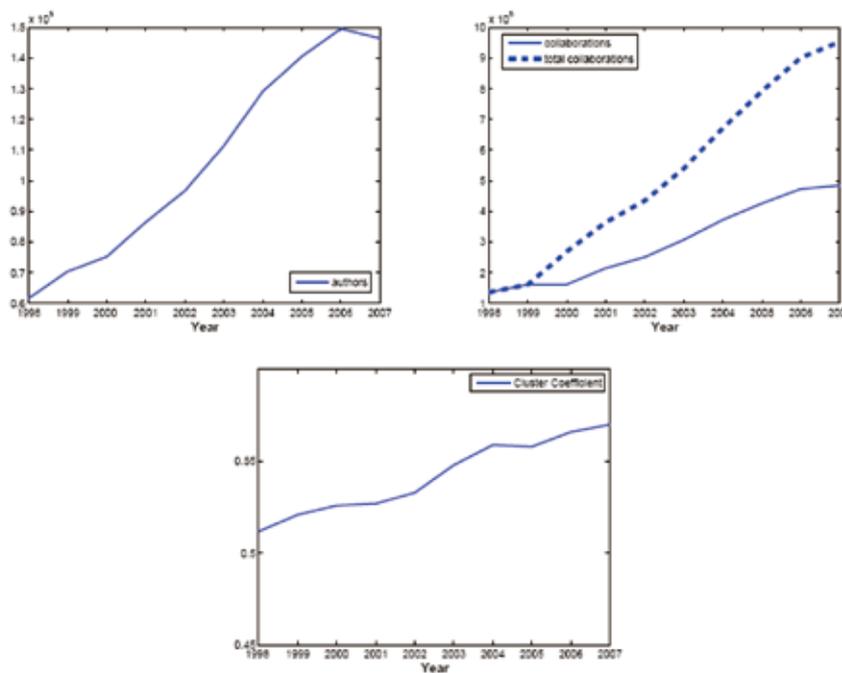


图1 1998年至2007年作者数、合作数以及网络聚集系数发展趋势

科研合作网络的度分布。定性来看，它符合幂率 (power-law) 分布的特点，满足复杂网络无标度特性。

3.2 科研合作网重要作者的分析

3.2.1 核心节点发现方法介绍

在复杂网络中，节点或边的中心度量常用的有四种^[1]。(1)最简单的是结点度数 (Degree)，即连接该点的边的数目，结点i的度 $K_i = \sum_{j=1}^n A_{ij}$ 。在一些社会背景下，结点

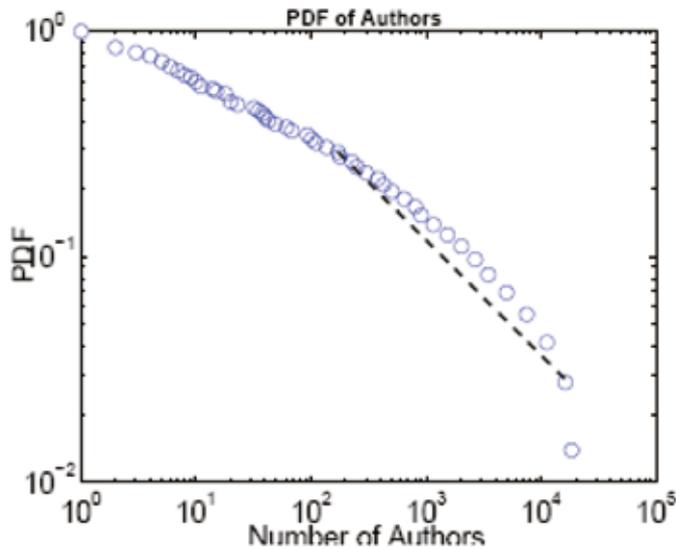


图2 2000年科研合作网络度分布

度数能够高效地度量结点的影响力及重要性；(2) 结点*i*的接近度中心度 (Closeness Centrality) 是指从*i*到其余各结点的最短距离 d_{ij} 的平均值，为 $CC_i = \frac{1}{\sum_j d_{ij}}$ ，容易证明越核心的结点接近度中心度的值越小；(3) 结点或边u的中介度中心度 (Betweenness Centrality) 定义为 $B_u = \sum_{ij} \frac{\sigma_{(i,u,j)}}{\sigma_{(i,j)}}$ ，其中 $\sigma_{(i,u,j)}$ 表示结点*i*与*j*间经过结点或边u的最短路径的数量， $\sigma_{(i,j)}$ 表示结点*i*与*j*间最短路径的总数量。其中，接近度中心度与中介度中心度的计算复杂度很高，无法直接应用于大型网络结构的重要节点发现。对于本文中的数十万个点，近百万条边的科研合作网络无法适用，而节点度未考虑到合作网络的链接信息。

基于链接结构分析的排序算法有两种，一种是斯坦福大学Sergey Brin和Lawrence Page提出的PageRank算法^[19]，他们提出一种用户行为的模型：假设有一个随机的网络游走者，任意给定一个网页，持续点击链接，最终厌倦并开始访问另一个随机页面。这样，一

个随机的网络游走者访问一个页面的可能性就是该网页的PageRank值。另一种算法是康奈尔大学Jon Kleinberg提出的HITS (Hypertext-Induced Topic Search) 算法^[20]。KleinBerg将网页分为两个类别，即权威页 (authorities) 和中心页 (hubs)。权威页是表达某一主题的页面，对这一主题它的价值很高，依赖于指向它的页面。中心页是指向多个权威页的页面，它把这些权威页链接在一起，依赖于它所指向的页面。HITS算法为每个页面引入两个权值，即authority权值和hub权值，通过一定的计算可得到针对某个检索提问的最具价值的网页，即排名最高的权威页。

3.2.2 改进的重要作者发现算法

接近度中心度与中介度中心度复杂度高，无法于应用大型网络，节点度度量没有考虑网络的链接信息，PageRank与HITS算法这两种著名的排序算法又都是应用于有向网

络的。本文构建的科研合作网络为无向加权网络，故不能直接适用。

事实上，PageRank算法的基本思想是借鉴传统的科研文献的引文分析方法，并把这一思想应用到了Web页面中，即一篇文献的重要性可以通过其他文献对其引用的数量来衡量。如果页面A通过超链接指向了页面B，相当于页面A给页面B投了一票，页面A需要把自己的一部分PageRank值分给页面B。最后，根据每个页面的PageRank值来判断页面的重要性，重要的页面会在搜索引擎的搜索结果中位于前列。如果一个网页有许多网页都指向它，那么它可能获得很高的PageRank值；如果一个网页被一个本身PageRank值很高的页面所指向，那么它同样可能具有很高的PageRank值。本文希望将PageRank的思想重新引入科研文献分析中，通过改进应用于科研合作网络的分析中。故本节主要讨论将PageRank算法思想引申至无向加权网络，应用于科研合作网络的重要作者发现。

我们认为，一个作者能够成为该领域中的核心作者，需要满足两个条件：其一，他与其他作者合作发表了大量的论文；其二，他与广泛作者合作发表过论文。为第一个条件，我们将PageRank度量进行改进，令其加权无向化，判断他的核心程度；为第二个条件，我们使用桥接系数^[21]的概念，判断他与其他作者合作的广泛程度。

由参考文献[19]可知，传统的PageRank值的计算公式为：

$$PR(A) = d(PR(T_1)/C(T_1) + PR(T_2)/C(T_2) + \dots + PR(T_n)/C(T_n)) + (1-d) \quad (2)$$

假设页面 $T_1 \sim T_n$ 都有超链接指向页面A。其中PR(A)表示页面

A的PageRank值；参数d是一个衰减因子，根据不同情况可以设定d在0~1之间，通常设定为0.85。C(A)表示页面A指向其他页面的链接个数。

令其加权无向化后，改进的PageRank值的计算公式可以定义为：

$$PR(A)=d \sum_{N \in pa(A)} \frac{W_{NA}}{W_N} PR(N)+(1-d) \quad (3)$$

其中， $N \in pa(A)$ 表示作者N在作者A的邻居集合中， W_{NA} 为作者N与作者A的合作次数， W_N 为作者N与其他作者的合作总次数。

作者A的桥接系数 $\Psi(A)$ 定义如下：

$$\Psi(A)=\frac{1}{d(A)} \sum_{N \in N(A)} \frac{\delta(N)}{d(N)-1} \quad (4)$$

其中， $d(A)$ 为作者A的度，表示作者N在作者A的邻居集合中，表示为作者N的邻居且不为作者A的邻居的作者数量。

结合以上两个度量，本文使用的判断重要作者发现度量 A_{score} 为：

$$A_{score}=PR(A)*\Psi(A) \quad (5)$$

3.2.3 结果分析

本文的实验环境是建立在单机上，处理器为双核2.66GHz，内存为2G。由于DBLP十年的数据规模很大，中介度以及中心度度量对其分析的计算效率较低，文章采用了另一个真实的科研合作网络，复杂网络分析领域的研究人员的合作网络^[22]，来比较度量的结果差异。该网络数据是由Newman等人收集的，共有1589个节点，2742条边。实验中对度、中介度、接近性中心度以及 A_{score} 对节点进行了中心值计算，表2展示了这四种度量方

法得到的结果的Kendal关联系数。从表中可以看出，度、中介度以及 A_{score} 三种中心性度量有较强的关联，他们与接近度都是负相关。在时间效率上，度、接近度、中介度以及 A_{score} 的时间复杂度分别为 $O(n)$ 、 $O(mn)$ 、 $O(n^2 * \log(n))$ 、 $O(k(m$

$+ n)$ 。其中，n为网络的节点数，m为网络的边数，k为迭代次数。一般情况下，它们的运行效率为中介度<接近度< A_{score} <度，随着数据量的加大以及网络连接越紧密， A_{score} 相比于中介度，接近度的速度优势更加显著。

表2 复杂网络分析领域科研合作网中各中心值的关联系数

	Degree	Closeness	Betweenness	A_{score}
Degree	1	-0.530	0.416	0.620
Closeness	-0.530	1	-0.274	-0.657
Betweenness	0.416	-0.274	1	0.380
A_{score}	0.620	-0.657	0.380	1

对于DBLP科研合作网络，我们使用新的作者影响力度量 A_{score} ，也得到了作者影响力排名列表。而中介度与接近度的效率很难在单机上得到该网络中心性的运算结果，故文章只比较了度与 A_{score} 的排名结

果。由于排名第一的作者2005年前一直是Alberto L. Sangiovanni-Vincentelli，故文章中只列出了以2000年以及2005年为例的作者影响力排名的比较数据。表3和表4列出了排名前十的作者名字，图3中

表3 2000年排名前十的作者列表比较

排名	度	A_{score}
1	Alberto L. Sangiovanni-Vincentelli	Alberto L. Sangiovanni-Vincentelli
2	Frank Thomson Leighton	Robert K. Brayton
3	Gio Wiederhold	Alok N. Choudhary
4	Leonidas J. Guibas	Philip S. Yu
5	Moti Yung	Leonidas J. Guibas
6	Hector Garcia-Molina	Thomas S. Huang
7	Ian T. Foster	Prithviraj Banerjee
8	Masahiro Fujita	Hector Garcia-Molina
9	Hiroaki Kitano	Toshio Fukuda
10	Avi Wigderson	Moti Yung

展示了两种不同度量得出的排名分别为前10、20、30、50、100、200的作者相似率的比较。根据与Arnetminer对计算机领域作者重要性排名的结果比较^①，尽管使用的数据集不尽相同，我们的算法发现出的重要作者在其数个度量得到的结果中都有所体现。Alberto L. Sangiovanni - Vincentelli, Wen

Gao, Wei Wang等作者都出现在其社会合作度量（Sociability）的Top 10列表中。值得一提的是，从我们的2005年的数据中通过 A_{score} 度量排名第8的作者，根据“度”度量计算未被列入Top 10的作者Philip S. Yu，在其作者列表中排名第9。而根据“度”度量被列入Top10的作者Li Zhang，在其排名列表中未列

入Top100。由此，本文提出的算法能够作为传统的排名算法的一个补充。

4 结论与展望

本文对DBLP数据库中1998年至2007年十年来的文献数据作者信息构建了科研合作网，对该网络的静态特征，例如度、聚集系数等进行了分析，并且与其他研究成果中不同领域的科研合作网进行了比较，得出DBLP科研合作网与其他合作网一样，满足小世界特性以及无标度特征。由于传统网络的核心度量，例如中介度、中心性度量等计算复杂度较高，不能适用于大型网络。文章还进一步提出了适用于大型科研合作网络的重要作者发现算法，分析了网络中作者的影响力，用以高效地发现合作数量多、范围广泛的核心作者。

本文只是对大型科研合作网络作了初步的分析，利用科研合作网还可以做很多工作。例如，科研网络中作者的合作模式是多样的，如何发现不同的合作模式，以及不同领域作者间的合作模式有什么不同是一个可以深入分析的研究方向。相信随着对复杂网络的研究，利用科研合作网的分析可以挖掘出更多的信息，为文献服务提供新的分析角度。

表4 2005年排名前十的作者列表比较

排名	度	A_{score}
1	Wei Wang	Wen Gao
2	Wei Li	Wei Wang
3	HongJiang Zhang	HongJiang Zhang
4	Wei-Ying Ma	Wei-Ying Ma
5	Alberto L. Sangiovanni-Vincentelli	Alberto L. Sangiovanni-Vincentelli
6	Wen Gao	Mahmut T. Kandemir
7	Ian T. Foster	Thomas S. Huang
8	Francky Catthoor	Philip S. Yu
9	Li Zhang	Francky Catthoor
10	Thomas S. Huang	Wei Li

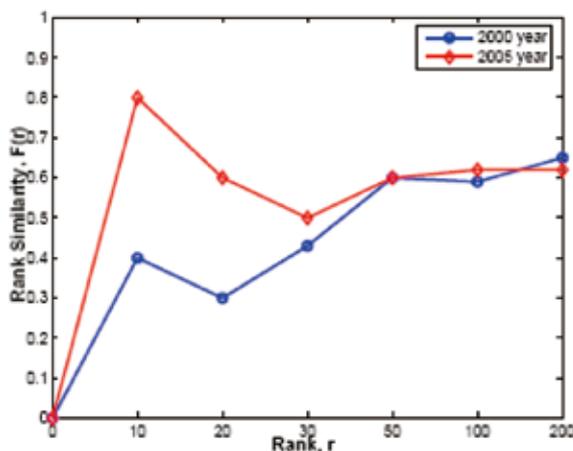


图3 作者排名列表相似率

^① <http://www.arnetminer.org/expertrank/list/sociability/-1>.

参考文献

- [1] ALBERT R, BARABÁSI A L. Statistical mechanics of complex networks [J]. Rev. Mod. Phys., 2002(74):47-97.
- [2] NEWMAN M E J. Coauthorship networks and patterns of scientific collaboration [J]. PNAS, 2004(101):5200-5205.
- [3] NEWMAN M E J. Who is the best connected scientist? A study of scientific Coauthorship [J]. Phys.Rev, E 64,016132,2001.
- [4] BARABÁSI A L. Linked: The New Science of Networks [M]. Massachusetts: Persus Publishing, 2002.
- [5] NEWMAN M E J. The structure of scientific collaboration networks [J]. PNAS, 2001,98(2):404-409.
- [6] BARABÁSI A L, JEONG H, NEDA Z, et al. Evolution of the social network of scientific collaborations [J]. Physica A: Statistical Mechanics and its Applications, 2001(311):3-4.
- [7] HUANG J, ZHUANG Z M, LI J, et al. Collaboration over Time: Characterizing and Modeling Network Evolution [C]// The First ACM International Conference on Web Search and Data Mining (WSDM).
- [8] 刘则渊,尹丽春,徐大伟. 论述复杂网络分析方法在合作研究中的应用[J]. 科学管理研究,2005(12):267-269.
- [9] 谢彩霞,刘则渊. 科研合作及其科研生产力功能[J]. 科学技术与辩证法,2006,23(1):99-103.
- [10] 尹丽春,刘则渊. 《科学计量学》引文网络演化研究[J]. 中国科技期刊研究,2006,17(5):668-722.
- [11] 李鹏翔,姚小涛,席酉民. 一个师生学术期刊的网络研究案例[J]. 网络与信息化,2005(7):175-177.
- [12] 王福生,杨洪勇. 《情报学报》作者科研合作网络及其分析[J]. 情报学报,2007,26(5):659-663.
- [13] 杨胜琦,吴斌. 生命科学近五年论文引文情况分析[J]. 数字图书馆论坛,2008(6):12-20.
- [14] WANG C, HAN J, TANG J, et al. Mining Advisor-Advisee Relationships from Research Publication Networks [C]. Sixteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD' 2010).
- [15] YAN E, DING Y, ZHU Q. Mapping library and information science in China: a coauthorship network analysis [J]. Scientometrics,2010(83):115-131.
- [16] TANG J, ZHANG J, YAO L, LI J, ZHANG L, SU Z. ArnetMiner: Extraction and Mining of Academic Social Networks [C]// Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD' 2008), 990-998.
- [17] TANG J, JIN R, ZHANG J. A Topic Modeling Approach and its Integration into the Random Walk Framework for Academic Search [C]// Proceedings of 2008 IEEE International Conference on Data Mining (ICDM' 2008), 1055-1060.
- [18] TANG J, JIN R, ZHANG J. A Topic Modeling Approach and its Integration into the Random Walk Framework for Academic Search [C]// Proceedings of 2008 IEEE International Conference on Data Mining (ICDM' 2008), 1055-1060.
- [19] BRIN S, PAGE L. The anatomy of a large-scale hypertextual Web search engine [C]// Proceedings of the 7th World Wide Web Conference (WWW7). 1998.
- [20] KLEINBERG J M. Authoritative Sources in a Hyperlinked Environment [J]. Journal of the ACM, 1999.
- [21] HWANG W, KIM T, RAMANATHAN M, et al. Bridging Centrality: Graph Mining from Element Level to Group Level [C]// Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2008.
- [22] NEWMAN M E J. Finding community structure in networks using the eigenvectors of matrices [J]. Preprint physics/0605087, 2006.

作者简介

朱天, 博士研究生, 研究方向为数据挖掘、图挖掘。通讯地址: 北京邮电大学179信箱 100876。E-mail: zhutian.bupt@gmail.com
 吴斌, 副教授, 主要研究领域为数据挖掘、复杂网络及智能信息处理。通讯地址: 同上。E-mail: wubin@bupt.edu.cn
 王柏, 教授, 主要研究领域为电信系统软件、分布计算技术、数据挖掘。通讯地址: 同上。

Core Author Discovery in Science Collaboration Network

Zhu Tian, Wu Bin, Wang Bai / Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing, 100876

Abstract: Analyzing the science collaboration data using complex network theory has received extensive attention in recent years. The author collaboration network built from DBLP database during 1998-2007 is studied in this paper. The paper gives a macroscopic characterization of the collaboration network by analyzing the statistic measures, such as degree, cluster coefficient, et al. It also presents a microscopic characterization of individual level by proposing an efficient core author ranking metric. The new ranking metric gives authors score through the number and the scope of cooperation of the author. Based on the studying of the collaboration network, the authors wish to provide a method to find the important researchers for science literature analysis.

Keywords: Social network analysis, Author collaboration network, Core node discovery

(收稿日期: 2010-05-31)