文章编号:1674-7070(2009)04-0360-05

科研合作网络的描述与分析

冯郁[1]

摘要

科研合作网络是描述作者之间合作关系的复杂网络.以某杂志为例,以其中作者之间的合作关系构建复杂网络,研究了该网络的基本特性,发现其具有比较明显的无标度网络特征;分析了其中最大的3个连通子网,发现其呈现小世界网络特征.

关键词

科研合作关系;复杂网络;无标度;小世界

中图分类号 G350;N94
文献标志码 A

收稿日期 2009-08-31

作者简介

冯郁,女,硕士,讲师,主要从事动力系统的研究. dgy@ nau. edu. cn

1 南京审计学院 数学与统计学院,南京,210029

0 引言

Introduction

复杂网络(Complex Network)的理论研究始于20世纪60年代由著名数学家Erdos和Renyi提出的ER随机图模型.1998年Watts和Strogatz在《Nature》杂志上发表文章,引入了小世界(Small-World)网络模型,以描述从完全规则网络到完全随机网络的转变.小世界网络既具有与规则网络类似的聚类特性,又具有与随机网络类似的较小的平均路径长度.1999年Barabasi和Albert在《Science》上发表文章指出:许多实际的复杂网络的连接度分布具有幂律形式,该类网络被称为无标度(Scale-Free)网络.无标度网络的节点度分布服从幂律(Power-Law)分布:$p(k) \sim k^{-\lambda}$,因此,无标度网络的连接分布极不均匀,网络中大量节点拥有少量的连接,而少量节点却拥有网络的大多数连接.现实世界中许许多多的复杂网络,如WWW、Internet、邮件系统、科研合作网络、新陈代谢系统、食物链、社会关系网等,都是无标度或小世界的网络.这2种网络都同时具有2个基本特征:高平均集聚程度、小的最短路径,而无标度网络的度分布又具有幂律分布特征[1-3].

对复杂网络的研究主要包括它的几何性质、形成机制、演化的统计规律、结构稳定性,以及演化动力学机制等问题.它为我们提供了一种复杂性研究的新视角、新方法,同时,也为在社会科学领域研究人与人之间的关系比如科研合作关系提供了新途径.

科研合作网络是描述作者之间科研合作关系的网络.通常把每个作者定义为1个顶点,如果2个作者合作发表过1篇论文,就把这2个顶点用1条边连接,对于一定范围内的科研合作情况进行考量,最终会生成1个描述科研合作关系的复杂网络.科研合作网络的研究最早起源是对数学家Pául Erdös的研究,他一生与超过500位科学家合作发表了1 600多篇论文.后来用Erdös数表示从某人到Pául Erdös的"路径",Erdös数为1表示直接与Pául Erdös有合作,Erdös数为2表示与Pául Erdös的1个合作者有合作,依此类推.这样,以Pául Erdös为中心形成了1个巨大的科研合作复杂网络[4].

国内外不少研究人员对科研合作网络的模型构建、特性分性等做了大量的研究.Newman对美国Medline、Los Alamos e-print Archive、Spires和NCSTRL等4个大型公用文献数据库1995—1999年的数据

进行分析发现:1)Medline 有1 520 251个节点,网络直径为24,平均路径长度4.6,聚集系数0.066;2)Los Alamos e-Print Archive 中4个子库分别有8 361、16 726、16 706和52 909个节点,网络直径分别为19、18、14和20,平均路径长度分别为6.91、6.4、4.66和5.9,平均聚类系数分别为0.327、0.348、0.414和0.43;3)NCSTRL 有11 994个节点,平均聚类系数0.496,网络直径为31,平均路径长度9.7;4)Spires有56 627个节点,网络直径为19,平均路径长度4,平均聚类系数0.726[4-5]. 刘杰等[6]对1998年1月至2004年6月间发表于《物理学报》和《Chinese Physics》上混沌科学方面的科学研究论文作者合作研究所形成的小型网络进行了分析,共涉及266篇论文360位作者,其中最大连通组群含有41位作者(节点),网络直径为11,平均路径长度2.175 37,聚类系数为0.703. 李凯旋等[7]对2001年1月至2006年12月期间发表于《情报科学》上的科学研究论文作者进行统计,研究了6年内作者合作所形成的非连通网络,涉及2 400篇论文2 629位作者,其最大连通组群有221位作者,网络直径为19,平均路径长度8.545,聚类系数为0.000 085. 王福生等[8]构建了1种科研合作网络模型,该模型能够很好地描述作者合作网络的演化过程.

1 科研合作网络的特性分析

Analysis of characteristics of the scientific collaboration network

要构建科研合作网络,首先是要选取合适的数据源. 一般而言,在同一学科领域,作者之间的合作关系相对较多. 选定合适的数据源后,还需要对其中的数据进行筛选,删除一些无用或冗余的数据. 最后,利用网络分析工具及可视化工具构建科研合作网络,并对该网络的网络连通性、节点度的分布等基本特性进行分析.

本文从中国知网(http://www.cnki.net)数据库中选取《审计研究》杂志1985年发刊起至2007年第3期上发表的所有论文,去除其中无作者信息或作者信息不明确的论文,共计1 354篇论文,1 288位作者. 根据作者之间的合作关系,建立1个科研合作网络,每1位作者用网络图中的1个节点表示,不同作者间只要有合作(即共同署名)就把他们之间的这种关系用1条连线(边)连接起来,不区分署名顺序,多次合作仅记录1次.

笔者发现,该科研合作网络是1个由712个互相独立的连通子网络组成的复杂网络,可见该网络是1个非连通网络. 具体数据见表1.

表1 连通子网络统计

Table 1 Statistics of the connected subnets

连通子网节点/个	1	2	3	4	5	6	7
网络/个	456	154	64	18	5	2	6
连通子网节点/个	9	10	13	17	20	49	63
网络/个	1	1	1	1	1	1	1

由表1可以看出,35%的作者(456人)从未与其他作者有过合作,24%的作者(308人)仅与1位作者有过合作且从未与第3人合作. 同时,也存在着6个节点数大于10的连通子网,其中最大的连通子网含有63个节点.

在科研合作网络中,节点的度就意味着该节点(作者)与其他节点(作者)的直接合作关系的多少,本文计算了《审计研究》科研合作网络中的各节点(作者)的度. 具体数据见表2.

表2 节点度的分布

Table 2 Distribution of the nodes and their degrees

度	0	1	2	3	4	5	6
节点/个	456	442	270	50	24	10	8
度	7	8	9	10	11	12	13
节点/个	5	1	3	12	4	1	2

在双对数坐标中,通过绘制度和节点数的关系图(见图1),并进行直线拟合,发现度的分布符合幂

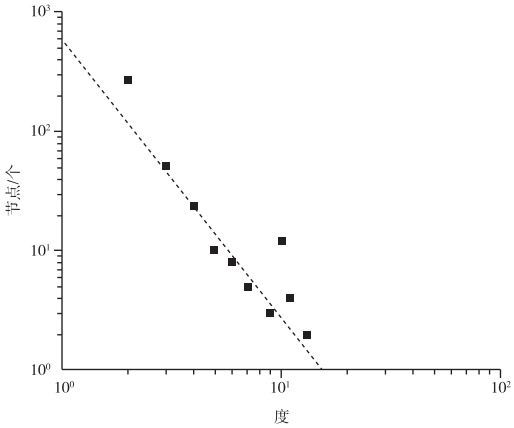

图1 节点度的分布,虚线为拟合直线

Fig. 1 Distribution of the nodes and their degrees with the dotted line as the fitting straight line

律特征:$p(k) \sim k^{-\lambda}$,λ 值约为 2.33,即该科研合作网络具有的明显的无标度网络的特性.

2 连通子网的分析

Analysis of connected subnet

对连通子网通过平均路径长度(Average Path Length)、直径(Diameter)、平均聚类系数(Clustering Coefficient)和节点的度及分布几个指标进行分析. 2 个节点间的距离为连接 2 者最短路径的边的数目,网络的直径为任意两点间最短路径的最大值;网络的平均路径长度则是所有节点之间距离的算术平均值;聚类系数 C 是用来描述网络中节点的聚集情况,假设节点 i 与 k 个节点相连接,如果这 k 个节点都相互连接,它们之间应该存在 $k(k-1)/2$ 条边,E 为这 k 个节点之间实际存在的边的条数,则节点 i 的聚类系数定义为$C_i = \frac{2E}{k(k-1)}$;平均聚类系数为整个网络中所有节点的聚类系数的平均值,在完全随机网络中,$C \approx N^{-1}$(N 为节点个数),在全连通网络(每个节点都与其他所有的节点相连)中,聚类系数等于 1[9].

仍然以《审计研究》杂志科研合作网络为例,分别计算了节点大于 10 个的 6 个连通子网的各项指标,数据见表 3.

由表 3 可以看出:这 6 个网络的平均路径长度最大为4.926 8,最小为2.021 5;平均聚类系数从0.409 5到0.821 3不等,均远大于 N^{-1}. 所以,这些网络都具有明显的小世界网络特征(较小的平均路径长度,较大的平均聚类系数).

进一步对最大的 3 个连通子网的拓扑结构进行

表 3 6 个较大连通子网的主要特征

Table 3 Chief characteristics of the six major connected subnets

节点/个	边/条	平均路径长度	直径	平均聚类系数	节点平均度	节点最大度
63	161	4.926 8	11	0.427 8	3.396 2	12
49	119	4.133 5	8	0.566 4	2.775 5	13
20	89	2.021 5	4	0.821 3	6.7	13
17	46	2.375	4	0.471 8	3.176 4	10
13	34	2.410 2	5	0.629 6	2.923 0	7
10	23	2.177 8	4	0.409 5	2.2	7

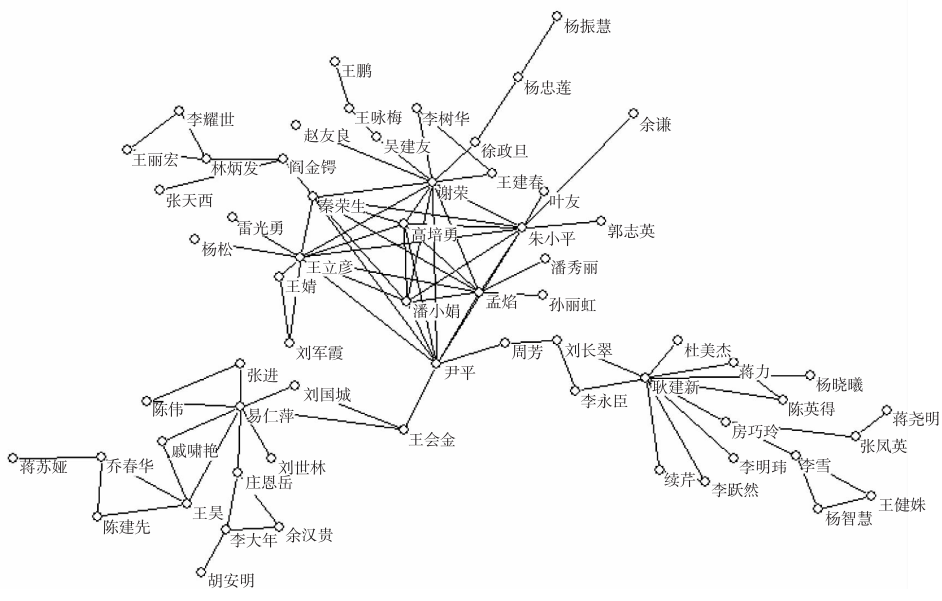

图 2 含有 63 个节点的网络图

Fig. 2 Diagram of the network containing 63 nodes

分析.

1）含有 63 个节点的子网络为整个《审计研究》杂志科研合作网络中的最大的连通子网(见图 2),其直径为 11,为 6 个子网中最大,它的平均路径长度为4.926 8,也是 6 个子网中的最大值;平均聚类系数为0.426 8,是 N^{-1}的约 27 倍,显示了比较明显的聚类特征;节点的最大度为 12,拥有者是国家会计学院(上海)的谢荣教授,其他度较大的节点都是北京大学、中央财经大学、南京审计学院等院校的知名学者.另外,以中国人民大学耿建新教授(度为 10)为中心节点,形成了 1 个类似星型网络.

2）含有 49 个节点的网络为第 2 大连通子网络(图 3),其直径为 8,平均路径长度为4.133 5,平均聚类系数为0.566 4,是 N^{-1}的约 28 倍.由拓扑图中可以发现,该网络中形成了 3 个比较明显的类星型网络,其中心节点分别为复旦大学李若山教授、厦门大学陈汉文教授和中山大学张立民教授,其中,节点“陈汉文”的度为 13,是《审计研究》杂志科研合作网

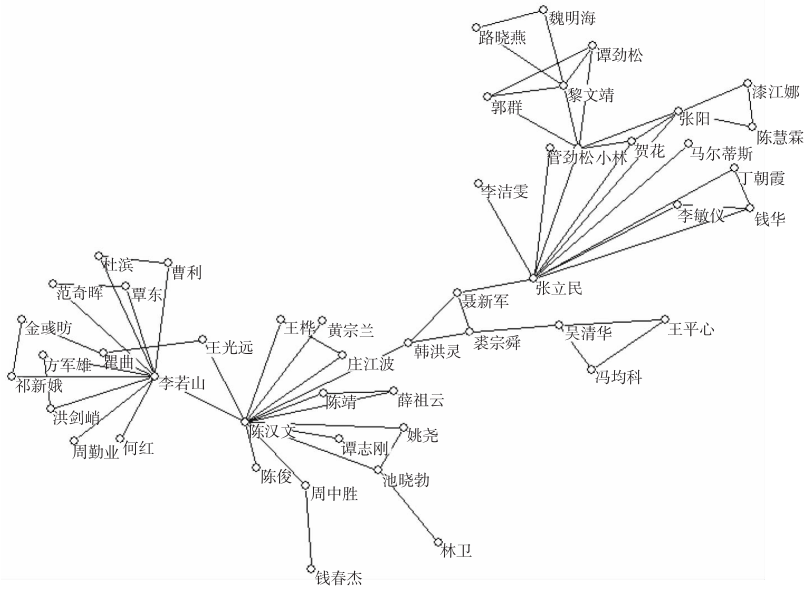

图 3　含有 49 个节点的网络

Fig. 3　Diagram of the network containing 49 nodes

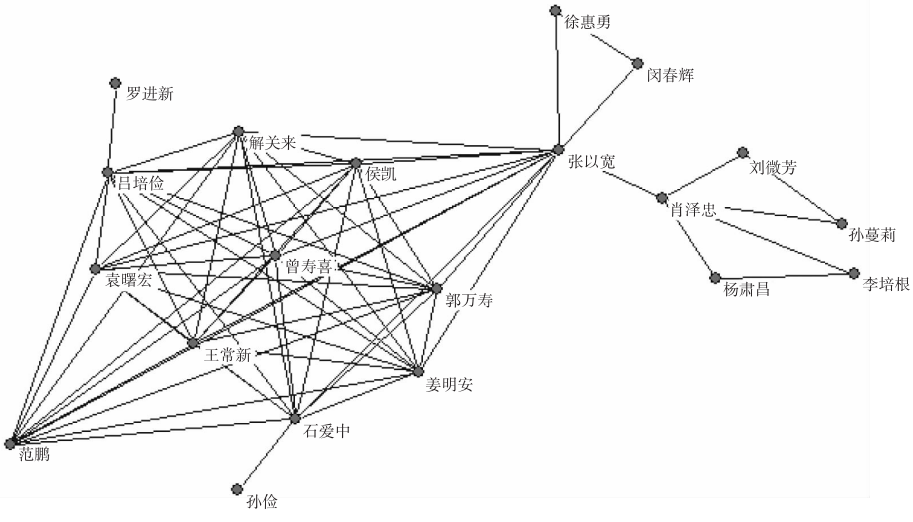

图 4　含有 20 个节点的网络

Fig. 4　Diagram of the network containing 20 nodes

络中的最大值. 这 3 位教授分别为各自学术团体的学术核心.

3）含有 20 个节点的子网为第 3 大连通子网络（图 4），其直径为 4，平均路径长度为2.021 5，平均聚类系数为0.821 3，该子网拥有 6 个连通子网中的最小的平均路径长度和最大的平均聚类系数，小世界特征最为明显.

3 小结

Conclusion

本文以《审计研究》杂志作者间的合作关系为例构建了科研合作网络，并对其网络基本特征进行了初步的分析，发现该复杂网络具有无标度特征，其内部连通子网络呈现了小世界特征. 利用复杂网络可视化方法绘制了较大连通子网的拓扑结构并进行了初步的分析，对未来其他类型复杂网络的描述与分析有一定的指导价值.

本文考虑作者合作关系时，未区分作者次序、未考虑合作次数，这对科研合作网络的精确性有一定的影响. 另外，本文仅考虑了 1 本刊物一段时间内的作者科研合作关系，很明显，这是一个静态的复杂网络，如何构造 1 种动态的随时间增长的复杂网络模型并对其实现可视化和特性分析是将来要研究的课题.

参考文献

References

[1] Newman M J. The structure and function of complex network[J]. SIAM Review, 2003, 45(2): 167-256

[2] Erdos P, Renyi A. On the evolution of random graphs[C]//Publications of the Mathematical Institute of the Hungarian Academy of Science, 1960, 5(1): 17-60

[3] Watts D J, Strogatz S H. Collective dynamics of "small world" networks[J]. Nature, 1998, 393: 440-442

[4] Newman M E J. Scientific collaboration networks. I. Network construction and fundamental results[J]. Physical Review E, DOI: 10.1103/PhysRevE.64.016131

[5] Newman M E J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality[J]. Physical Review E, DOI: 10.1103/PhysRevE.64.016132

[6] 刘杰，陆君安. 一个小型科研合作复杂网络及其分析[J]. 复杂系统与复杂性科学，2004(3): 56-61
LIU Jie, LU Junan. A small scientific collaboration complex network and its analysis[J]. Complex Systems and Complexity Science, 2004(3): 56-61

[7] 李凯旋，林娜，杨洪勇. 基于《情报科学》期刊作者合作网络模型分析[J]. 现代情报，2007(9): 57-61
LI Kaixuan, LIN Na, YANG Hongyong. The analysis of journal author collaboration network based on Information Science[J]. Modern Information, 2007(9): 57-61

[8] 王福生，杨洪勇. 作者科研合作网络模型与实证研究[J]. 图书情报工作，2007(10): 68-71
WANG Fusheng, YANG Hongyong. Author scientific collaboration network model and its demonstration study[J]. Library and Information Service, 2007(10): 68-71

[9] Albert R, Barabbsi A L. Statistical mechanics of complex networks[J]. Review of Modern Physics, 2002, 74(1): 47-97

Description and analysis of scientific collaboration network

FENG Yu[1]

1 School of Mathematics and Statistics, Nanjing Audit University, Nanjing 210029

Abstract The scientific collaboration network is the complex one describing the relationship between the authors who have coauthored a paper together. This paper constructs a scientific collaboration network based on the authors' collaborative relationship contained in the papers published in audit research, and then analyses the scale-free characteristic of the network and small-world characteristic of the three biggest connected subnets.

Key words scientific collaboration; complex networks; scale-free; small world