

南京医科大学2016年度 英文论文发表的作者合作网络分析

江胜强¹, 赵 杨^{2,3}, 于 洁², 陈会会²

1.南京医科大学科技处,2.公共卫生学院,3.生物医学大数据重点实验室,江苏 南京 211166

摘要:对南京医科大学2016年英文论文的作者进行统计,使用R语言的igraph包,采用复杂网络的研究视角,探究论文发表合作网络情况。分析表明,作者所形成的科研合作关系网络是一个非连通网络,网络统计参数呈现出无尺度网络特性和小世界网络特性。通过筛取合作网络内部较大的连通子网络并对其进行研究,可了解学校学术研究中新思想的传播方式。

关键词: 社会网络分析;作者合著;科研合作网络;小世界模型

中图分类号: G644

文献标志码: A

文章编: 1671-0479(2018)01-077-003

doi: 10.7655/NYDXBSS20180119

随着世界一体化进程不断加速,人与人之间的协作变得越来越紧密,而这一点对于科研工作者而言尤其重要。体现科研工作者协作成果的方式之一是论文共同署名,不难想像,两位作者在发表的论文中同时出现的次数越多,说明两位作者协作越紧密,成果越丰硕。在作者科研协作这一网络中,我们能够运用社会网络分析(social network analysis, SNA)方法来详细地描述科研工作者之间的合作关系。SNA能对处于网络中的所有个体间的关系进行量化分析^[1],不同的个体或者群体所组成的关系当中的结构和属性也有所不同。运用具体的量化分析,SNA可以在中心性等角度对作者的论文合著关系进行网络分析^[2]。

本文使用SNA,从作者合著这一角度来研究科研论文中的合作关系,运用R软件构建作者科研合作网络,达到科研合作网络的可视化,分析科研合作网络的特点和特征^[3],结合现有数据呈现出的作者科研合作关系,对南京医科大学科研工作的发展进行评估,对知识传播产生积极影响的团体或个人进行发掘。

一、材料与方法

(一)数据来源及整理

本文所使用的数据为南京医科大学2016年英

文论文发表的数据,一共955篇文献,由南京医科大学科技处提供。两位研究人员独立提取了论文的有关信息,所提取的信息包括论文的作者、文献标题、文献类型、基本科学指标数据库、作者地址、团体作者、被引频次、出版年等。为了避免重名现象对分析结果产生干扰,两位研究人员独立复核了所有作者的姓名,对于可能重名的作者,由学校人事部门加以区别,最终共识别了4 383位独立作者。

将数据转换为 $4\ 383 \times 4\ 383$ 矩阵的形式,每个唯一的作者作为矩阵的一行(一列)。若两位作者从来没有一起合作过论文,则对应元素取值为0;否则,作者间每共同发表一篇论文,则矩阵相应位置数值增加1。这里暂不区分合作者的署名顺序,同一篇论文的各个作者间的合作关系用全连接方式表达。

(二)数据分析

可以把作者与作者间的联系看成一种社会网络(social network)。每个作者可以认为是社会网络中的一个节点(node),而若作者与作者间共同发表过论文,则可用边(edge)将他们联接。针对SNA, Watts等^[4]首次提出了“小世界”(small-world)网络模型。该模型指出,现实世界中的网络很多都具有很高的聚集性,但大多数节点属于分散状态,两节点之间的距离很短。真实网络几乎都具有小世界效应^[5-6]。此外,研究者还发现大量真实网络的节点度

基金项目:江苏高校哲学社会科学研究基金项目(2017SJB0278);江苏省青蓝工程;江苏省高校优势学科建设工程;江苏高校品牌专业建设工程资助项目(PZY2015A067)

收稿日期: 2017-09-15

作者简介: 江胜强(1979—),男,江苏溧阳人,助理研究员,研究方向为科技管理。

服从幂律分布^[7],即具有很强的异质性,下降趋势是指数型的。节点度服从幂律分布的网络被称为无标度网络(scale-free networks),即所谓“网络的无标度特性”^[8]。

我们利用R语言的igraph扩展包对本数据进行SNA。分析将从以下几个角度开展:首先对网络图的节点和边的特征、网络的凝聚性特征进行描述性分析,再利用不同算法和布局方式寻找最佳的网络图展示方式,最后再从复杂网络分析的角度,结合实际网络的特性提出具体可行的分析方法。

二、结果

(一)科研合作网络的整体特征

根据已有的论文信息,首先初步构建了一个复杂的科研合作网络图。这个科研合作网络是由4 383个节点组成的复杂网络,是一个非连通网络。各节点度的分布如图1。从4 383个节点,42 068条边构成的直方图可见分布的异质性很强。度值的大小与科研合作网络相当,网络中还有一定数量的节点度值很高的节点。使用双对数坐标将更直观体现这一特点(图1),取对数后的频率随着取对数后的度值变化,呈现较好的线性下降趋势。幂律递减的规律显示出复杂网络的度分布具有无标度网络特征。

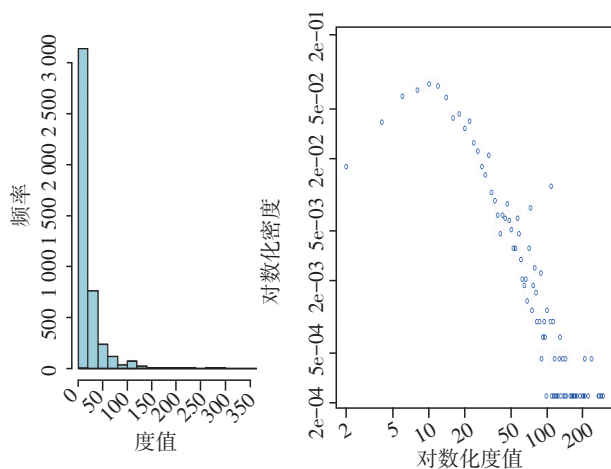


图1 科研合作网络分布直方图和双对数图

(二)科研合作网络的小世界性

利用弹簧模型,我们对可能的绘图空间进行高效的搜索,生成的图形较好权衡了局部和全局结构的细节。同时,考虑到越在外侧的节点在网络中越处于游离的位置,换言之,在科研网络中与他人的合作越少,对科研合作网络图的影响微小,故在后继分析中将其删除。为了使图形的显示更为清晰,我们将只合作发表过一次论文的两作者间的边隐藏作图,将图中心部分放大,如图2。可以看出科研合作有一定的小世界性。作者合作网络是一个非连通的网络,并且由许多相互独立的子网络组

成。图中间有两个比较大的子网络,周围分布着一圈星状的子网络,越在外围的作者意味着与他人的合作越少。处于图2中间的两个最大的连通子网络,这两个活跃的科研团队主要来自于该校肿瘤学、外科学及生殖医学等团队的研究者,发表论文数量较多,不同作者之间的协作次数也较多,作者之间的相互影响力较大,这极大地推动了学校的科研发展。

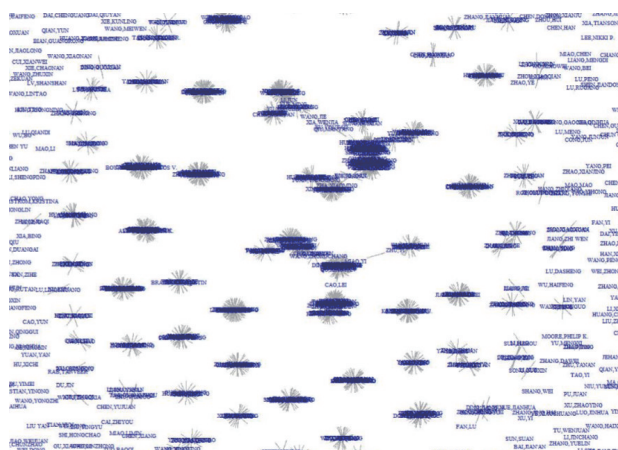


图2 科研合作关系图之边大于2(放大后)

(三)科研合作网络的细化分析

对该校英文论文合作网络取出部分节点做可视化分析,对其中较大的连通子网络进行分析,添加节点度大于110的节点条件后,从4 383个节点中选出了43个节点(图3)。这也印证了本文上述结果。提取图3每个节点周围(一阶)的邻居,可以得到邻居规模最大的4个节点,提取该节点所处的连通子网络图,如图4。

图4给出了包含16个节点的连通子网络图,它是科研合作网络中全部连通子网络图中节点数最多的,此网络图的边数为76,它的平均路径长度为2,最大的路径长度为4,网络图平均聚类系数为0.619,大约是 N^{-1} (N为节点个数)的10倍,网络同时具有小的平均距离和明显的聚集性,是一个小世界网络图。在这个子网络中作者间的合作最为紧密,合作交流的频率比较高,极大推动了该校的科研发展。节点度的平均值为4.75,也说明图中作者的合作比较紧密。观察此子网络图可发现,这个网络图表现出了一定的聚集性,即网络由较为明显的小群体组成,群体内部节点间相对紧密,而社团之间则比较稀疏。

从全部的科研合作网络图,我们提取的节点度最高的5名作者,观察到他们每个人都会隶属多个不同科研团队,或者他们研究范围是交叉学科,在很多科研领域都有需求。其中,节点度最高的前3位作者相对数分别为0.514,0.241和0.241。

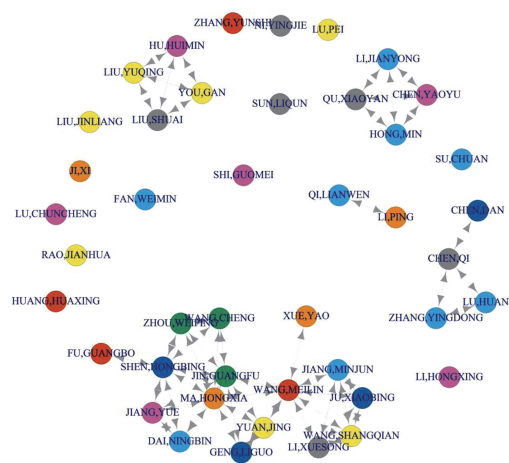


图3 大于110节点网络分布

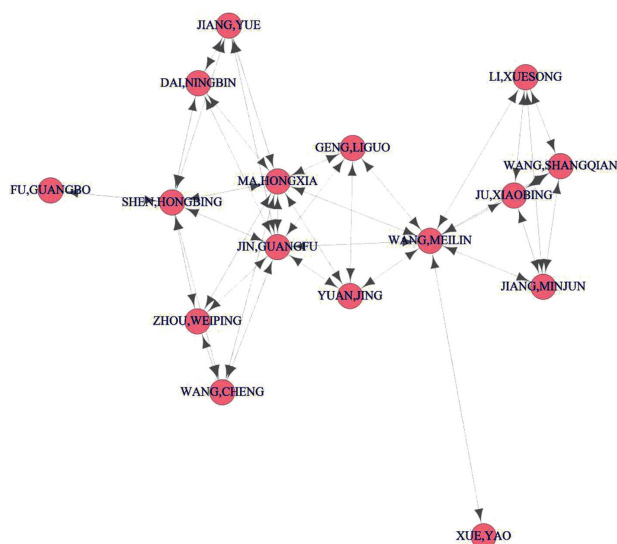


图4 连通子网络分布图

三、讨论

将科研合作视为一种复杂的社交关系,可以将社交网络的分析方法引入科学情报分析^[9]。社会网络分析是一种有效的网络分析方法,本文对南京医科大学2016年英文论文科研合作网络进行了分析,有助于了解南京医科大学的科研工作者的科研协作情况。研究显示该科研合作网络是复杂网络,且整体较大,但由于其节点度的分布符合幂律分布,因此该网络是无标度网络,具有小世界特性。在连通子网络中也有鲜明性,小团体由其中的科研骨干

力量领导,形成十分有活力的科研团队,成为科研中知识传播与创造的中坚力量。

分析中发现,中心节点度最高的作者相对数为0.514,其次2位作者相对数均为0.241。根据Freeman^[10]给出的介数中心性的定义,我们可以理解为这3位作者的中介性较强,其他2人合作最短距离都经过该“中介”,说明这个作者很重要,联系性很强,是这个连通的科研合作子网络的科研骨干。

本研究仍旧存在一定的局限性,由于数据的规模较大,在追求可视化的同时,不得不牺牲掉部分信息;研究时只保留了节点的合作信息,而将文献标题、文献类型、基本科学指标数据库、作者地址、团体作者、被引频次、出版年等信息人为删除了,损失了部分信息;同时,由于条件限制,我们只使用了2016年的出版文献,若能纳入5年以上的结果,将能更好地体现合作的变化情况。这些将在后期研究中加以解决。

参考文献

- [1] Barabasi AL, Albert R. Emergence of scaling in random network[J]. Science, 1999, 286(5439): 509-512
- [2] 刘军. 整体网分析讲义: UCINET软件实用指南[M]. 格致出版社, 2009
- [3] 李进, 刘瑞璟, 于伟, 等. 作者科研合作网络构建及影响分析——以《复杂系统与复杂性科学》期刊为例[J]. 复杂系统与复杂性科学, 2014, 11(3): 86-93
- [4] Watts DJ, Strogatz SH. Collective dynamics of “small-world” networks[J]. Nature, 1998, 393(6684): 440-442
- [5] 周涛, 柏文洁, 汪秉宏, 等. 复杂网络研究概述[J]. 物理, 2005, 34(1): 31-36
- [6] Lü L, Pan L, Zhou T, et al. Toward link predictability of complex networks, Proceedings of the National Academy of Sciences[J]. Proc Natl Acad Sci USA, 2015, 112(8): 2325-2330
- [7] Newman M. Pareto distributions and Zipf's law[J]. Contemporary Physics, 2005, 46(5): 323-351
- [8] Barabási A, Bonabeau E. Scale-free networks [J]. Sci Am, 2003, 288(5): 60
- [9] 侯剑华. 国际科学合作领域主流学术团体与代表人物分析[J]. 现代情报, 2012, 32(1): 84-91
- [10] Freeman L. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40(1): 35-41