

高校科研团队的识别与网络分析研究

李远明,谭世明

(湖北民族学院 图书馆,湖北 恩施 445000)

摘要:基于文献计量的频次统计原理,分析了某高校“十一五”期间科研人员发文概况,并从作者共现视角入手,对科研人员合作发文的网络结构进行了深度分析。运用社会网络分析的组件、密度及中心性等整体网分析指标对其校科研人员合作的紧密程度、团队合作的基本结构、学术科研团队的数量及团队成员的构成进行了揭示,并对分析结果进行了可视化展示。

关键词:高校科研团队;文献计量;作者共现;社会网络分析;可视化分析;频次统计

DOI:10.6049/kjbydc.2011050094

中图分类号:G644

文献标识码:A

文章编号:1001-7348(2012)11-0147-04

0 引言

以合作研究为目的的高校科研团队已成为我国知识创新和科技开发的重要群体,高校通过组建团队进行科技创新已成为相当普遍的一种科研形式,并得到各级科研管理部门的大力提倡^[1]。传统意义上的科学家个人奋斗情形已经不再适合当今世界科研工作的要求,科学的发展需要同一学科相同领域或不同领域甚至不同学科间的科研人员进行交流与合作^[2]。国内关于高校科研团队的研究主要集中在:学术研究合作的缘由、团队带头人的培养、团队的形成和管理机制等方面^[3],而如何准确科学地识别高校科研团队并了解团队成员间知识的交流与合作情况、如何发现团队的核心成员、如何揭示团队的结构特征等问题的实证研究正日益受到关注。本文拟根据文献计量学原理,采用社会网络分析法和数据挖掘工具,以科研人员间的合作关系为研究对象,对某大学(以下称H大学)“十一五”期间科研团队的状况、合作的基本模式、团队的基本结构进行研究,以便为高校科研团队的分析和研究拓展出新的方法和空间,为高校科研规划及科研管理提供借鉴。

1 研究方法与研究流程

1.1 研究方法

本研究采用了文献计量和社会网络分析两种方

法。

文献计量是集数学、统计学、文献学于一体,注重量化的综合性知识计量方法。通过对文献量(各种出版物,尤以期刊论文和引文居多)、作者数(个人、集体或团体)、词汇数(各种文献标识,其中以叙词居多)进行计量^[4]。采用频次统计技术,能从一定程度上定量地反映出某研究者、某机构、某学术团体或某学科的学术研究概况。基于此方法,本文提取了H大学十一五期间发表的学术论文,然后抽取出每篇文献的作者,得到了该校十一五期间发文的作者集,再对作者集进行频次统计,得到了频次统计表,依据频次统计结果,对作者的合作情况进行共现分析,最后形成作者共现矩阵。

共现分析方法最早在20世纪70年代中后期由法国文献计量学家提出。1986年,法国国家科学研究中心CNRS(Centre National de la Recherche Scientifique)的M. Callon、J. Law和A. Rip出版了《Mapping the Dynamics of Science and Technology》^[5]。作者共现指的是 $n(n \geq 2)$ 个作者合作发表科技论文,并同时在同一篇科技论文中署名的情况。共现关系可反映为数众多的作者基于合作而聚集成一个个作者群体,从而及时发现学科专业人员之间的联系和结构特点,进而反映出他们所从事的学科专业之间的联系及发展变化趋势。共现频率越高,即说明他们之间的关系越密切。

基于共现关系形成的矩阵是共现频次数值矩阵,其数据是多值的。为便于社会网络分析,需要将多值数据矩阵转化为邻接矩阵,即二值矩阵。数值“1”表示

收稿日期:2011-01-14

基金项目:湖北省高等学校教研项目(20070324);湖北民族学院教研项目(2010JY01);湖北民族学院图书馆科研项目(2011GLB004)

作者简介:李远明(1969—),男,湖北鹤峰人,硕士,湖北民族学院图书馆副研究馆员,研究方向为信息计量、情报分析与信息检索;谭世明(1969—),男,湖北利川人,博士,湖北民族学院图书馆教授,研究方向为区域经济可持续发展与高等教育管理。

二者之间有合作关系,数值“0”表示二者之间没有合作关系。本文构建转化算法公式如下:

$$a_{ij} = \begin{cases} 1_+ & \text{若 } a_{ij} \geq \frac{\sum_{i=1, j=1}^n a_{ij}}{n \times n - k} \\ 0_+ & \text{若 } a_{ij} < \frac{\sum_{i=1, j=1}^n a_{ij}}{n \times n - k}; \text{若 } i = j \end{cases} \quad (1)$$

上式中 n 为矩阵中数据的行列数, k 为矩阵中元素为 0 的个数。即矩阵元素大于等于阈值时赋值为 1, 否则赋值为 0, 矩阵中对角线上的值赋为 0。

社会网络分析(Social Network Analysis, SNA)是人、集团、组织、计算机或者其它信息与知识处理实体的关系和流(flows)的映射及测量^[6]。它是一种以关系为基本单位进行分析的实证研究方法,来自于数学之图形理论,最早可追溯到 20 世纪 30 年代的心理学和人类学研究。20 世纪 70 年代后,随着“新哈佛学派”的出现,社会网络分析逐渐成熟发展起来。20 世纪 90 年代以来,社会网络分析进入快速发展时期。

社会网络指的是社会行动者及其关系的集合,集合由多个节点和节点之间的连线组成,连线即节点之间的关系,如:朋友关系、亲属关系、合作关系、借贷关系等,而节点可以是个体、单位、公司、城市或国家等。个体之间的交往行为、家族之间的联姻、学者论文的相互引用、学术论文的合著等均可构成各自的社会网络。社会网络分析以“在互动的单位之间存在的关系非常重要”^[7]这一独特视角,分析社会网络内部成员之间的关系和交互,力图发现它们的组织结构、组织特点、行为方式、个性特征等^[8]。其独有的方法和技术为我们定量地分析和研究实体间的关系提供了可能性和便利性。通过社会网络中的密度、规模、关系、距离、中心性和图的度数中心势指数等概念,一方面能形象直观地反映社会网络的概况,同时又能挖掘出网络中具有重要地位的论著、作者或是它们之间的关系,进而发现他们的结构和特征等。

社会网络通常采用社群图法和矩阵代数方法来加以描述。本文采用作者共现矩阵来描述作者间的合作关系,并通过社会网络分析工具 UNINET 对这一合作关系进行研究。

1.2 研究流程

本研究流程主要有以下几个步骤:

第一步:数据收集。根据研究主题确定数据源,并明确数据的收集范围。

第二步:数据清洗。将收集到的数据进行规范化处理,并去掉不符合要求的“脏”数据。

第三步:数据计算。对清洗后的数据进行频次统计和共现分析。

第四步:制图。将计算结果运用合适的可视化工

具绘制成图形,便于研究者和用户分析和理解。

第五步:形成研究论文或报告。根据研究目的,在计算结果和可视化图形的基础上形成研究论文或报告。

2 实证研究:以 H 大学为例

2.1 数据来源及样本的提取

根据研究需要,本文选择中国知网(CNKI)的中国学术期刊网络出版总库为数据来源,该库收录了 7 579 种学术期刊,文献来源覆盖率达到 99%。检索时间限定为 2006. 1. 1—2010. 12. 31,作者单位以“H 大学”(隐去真实名称)为检索词,以模糊匹配方式进行检索,共得到记录 4 149 条,去掉 1 条无用数据,共得到数据 4 148 条。经过数据清洗后,作为样本数据进行分析。

统计发现,样本数据中总署名次数为 7 829 人次,该校共有 2 328 名作者撰写了学术论文,人均署名 3. 4 次。有 1 187 名作者在样本数据中只出现过一次,约占 51%。

2.2 作者频次统计

对样本数据的论文作者进行频次统计,可以得到该校每位作者发表的论文篇数(含非第一作者身份发表的论文)。表 1 是该校发表论文较多的作者(作者名用姓名首字母代替)。

表 1 发表论文在 20 篇及以上的作者

序号	作者名	篇数	序号	作者名	篇数	序号	作者名	篇数
1	XDS	52	17	QEH	27	33	ZGL	22
2	ZXJ	52	18	JL	27	34	YH	22
3	YAN	43	19	JJL	27	35	HQ	22
4	ZK	41	20	ZC	27	36	CS	21
5	ZDZ	39	21	DL	27	37	CPZ	21
6	HWB	36	22	LHH	25	38	ZJM	21
7	LXP	36	23	LXP	25	39	CXB	21
8	ZYF	33	24	XBJ	25	40	LLW	20
9	AXR	32	25	WBQ	24	41	LW	20
10	TQY	29	26	YYM	24	42	YH	20
11	LJL	29	27	JYW	24	43	XYP	20
12	CC	29	28	WY	24	44	HF	20
13	ZYC	29	29	MYP	24	45	PC	20
14	LL	29	30	DL	24	46	HLY	20
15	LYS	28	31	LH	23	47	TZX	20
16	MKJ	27	32	TXC	22			

根据本文研究需要,假定“十一五”期间该校作者年均能发一篇文章,我们选定作者频次 5 以上(含 5 次)的作者共 439 名,占发文作者总数的 18. 9%,统计生成作者共现矩阵。矩阵片断见图 1。

2.3 H 大学学术科研团队分析

2.3.1 H 大学学术科研团队的可视化描述

将图 1 中的共现矩阵转换成邻接矩阵后,借助 Ucinet 软件进行可视化描述,可视化图形见图 2。

*	XDS	ZXJ	YAN	ZK	ZDZ	HVB	LXP	ZYF	AXR	TQY	LJL	CC	ZYC	LL	LYS	MKJ	QEH	JL
XDS	52	14	0	30	10	0	17	0	0	4	1	0	3	0	0	0	1	9
ZXJ	14	52	0	7	11	0	14	0	0	5	19	0	5	0	0	0	0	7
YAN	0	0	43	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0
ZK	30	7	0	41	3	0	7	0	0	0	0	0	0	0	0	0	0	2
ZDZ	10	11	0	3	39	0	11	2	0	12	1	0	19	0	0	0	0	4
HVB	0	0	15	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0
LXP	17	14	0	7	11	0	36	0	0	6	0	0	6	0	0	0	2	27
ZYF	0	0	0	0	2	0	0	33	2	15	1	0	3	0	0	0	7	0
AXR	0	0	0	0	0	0	0	2	32	0	0	0	0	0	0	0	0	0
TQY	4	5	0	0	12	0	6	15	0	29	0	0	11	0	0	0	5	2
LJL	1	19	0	0	1	0	0	1	0	0	29	0	0	0	0	0	1	0
CC	0	0	0	0	0	0	0	0	0	0	0	29	2	0	0	16	1	0
ZYC	3	5	0	0	19	0	6	3	0	11	0	2	29	0	0	2	1	2
LL	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	1	0
LYS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0
MKJ	0	0	0	0	0	0	0	0	0	0	0	16	2	0	0	27	2	0
QEH	1	0	0	0	0	0	2	7	0	5	1	1	1	1	0	2	27	2
JL	9	7	0	2	4	0	27	0	0	2	0	0	2	0	0	0	2	27

图 1 作者频次 5 次(含 5 次)以上的共现矩阵片断

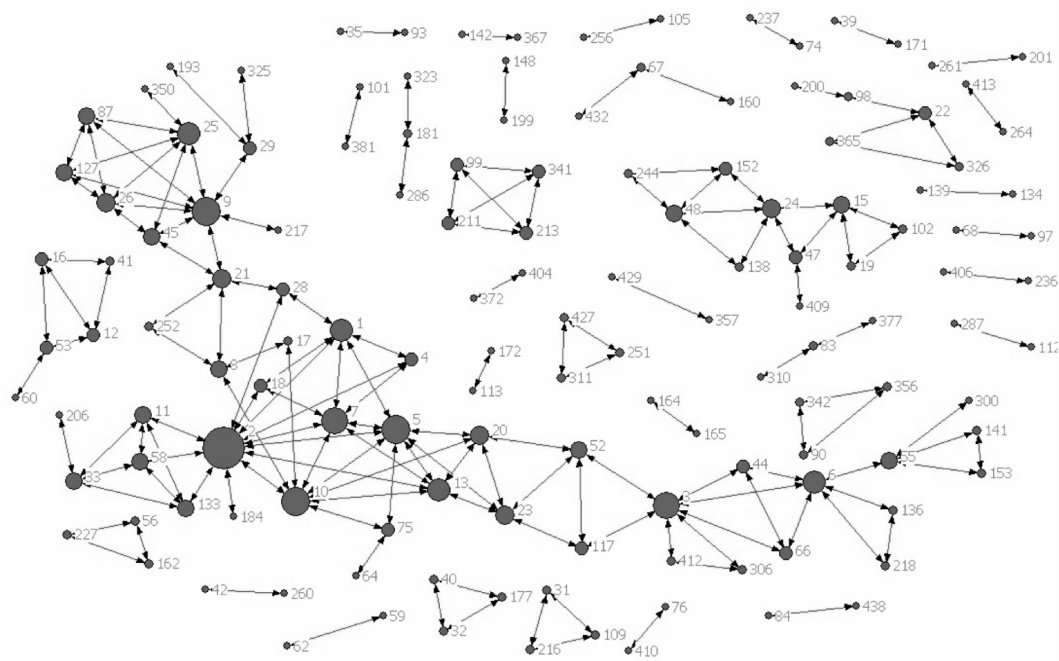


图 2 H 大学科研团队网络

该网络用点和点之间的关系连线组成,网络中的每个节点代表一名科研人员,本文隐去其真实姓名而用数字序号表示。节点的大小表示科研人员的发文数量,大的表示发文多,小的表示发文数量相对较少。节点之间的连线表示科研人员间的合作关系。无连线的表示他们是独立的个体,没有或基本没有合作关系,并将这部分人员放置在图的左侧。

由图 2 可见,该校科研团队整体网络呈现如下特点:单点型和双核型网络占绝大多数,尤其以单点型网络居多。单点型指的是独立发表文章的子网络,图 2 中单点型网络有 301 个。双核型指的是仅两人合作发文形成的子网络,图 2 中双核型网络有 21 个。以科研人员 1、2、10 等为核心的人员组成了一个共 48 人的较复

杂的核心型网络。该图形显示出该校部分学科的团队建设很强势,而其它学科的团队建设明显不足。

2.3.2 H 大学学术科研团队整体分析

本文将样本数据中提取出的 439 位作者视为一个整体网,分析研究该网络的组件、密度及中心性等方面的指标。

一个图形的组件是最大的相连的子图形^[9],一个组件就是最大的一块相连的节点。单点型网络是组件中的特例。图 2 中显示出来的组件即是 H 大学的科研团队概况。分析图 2 中的组件构成,旨在了解 H 大学学术科研团队的个数、团队结构及科研人员构成情况。H 大学学术科研团队整体网的组建构成如表 2 所示。

表2 H大学学术科研团队构成

团队类型	数量	团队成员	团队占团队总数(%)	成员占样本总数(%)
单点型	301	301个单独成员	89.9	68.6
双核型	21	(35 93);(39 171);(42 260);(59 62);(68 97);(74 237);(76 410);(84 438);(101 381);(105 256);(112 287);(113 172);(134 139);(142 367);(148 199);(164 165);(201 261);(236 406);(264 413);(357 429);(372 404)	6.3	9.6
三角型	8	(31 109 216);(32 40 177);(56 162 227);(67 160 432);(83 310 377);(90 342 356);(181 286 323);(251 311 427)	2.4	5.5
多点核心型	5	(1 2 3 4 5 6 7 8 9 10 11 13 17 18 20 21 23 25 26 28 29 33 44 45 52 55 58 64 66 75 87 117 127 133 136 141 153 184 193 206 217 218 252 300 306 325 350 412);(15 19 24 47 48 102 138 152 244 409);(22 98 200 326 365);(12 16 41 53 60);(99 211 213 341);	1.5	16.4
合计	335	439	100	100

在社会关系网中,密度是一个用来表征社会群体结构形态的重要指标,一个团体可以有紧密关系,也可以有疏离关系。关系紧密的团体合作行为较多,信息流通容易,团体工作绩效也会比较好;而关系疏离的团体,其情况恰好相反。对社会关系图而言,其整体密度即是图形中实际存在的线与可能数量的线的比例,密度越大,该网络对其中的成员产生的影响越大,其取值范围在0和1之间,用公式(2)表示为:

$$\Delta = \frac{2L}{g(g-1)} \quad \begin{matrix} L = \text{图中线的数目} \\ g = \text{图中节点的数目} \end{matrix} \quad (2)$$

经计算,图2中线的数目共有328条,其密度为0.0034,密度偏低。

中心性是社会网络分析中衡量个人在网络结构中的位置的重要指标,用以评价一个人重要与否,或者说分析这个人居于什么样的中心地位。研究发现,中心性指标与群体效率有关,也与参与群体的个人满意度有关^[10]。本文用程度中心性来测量处于团队中最主要位置的中心人物。拥有高中心性的个人,在团体中的学术地位也高。使用群体程度中心性(也称图的度数中心势指数)来反映团队的集中趋势,取值范围在0和1之间,数值越大,说明团队的集中趋势越明显。

经对样本数据进行分析,中心性均值为0.747,程度中心性在1(含1)以上的共138人。群体程度中心性为0.0258,说明该校学术科研团队建设明显不足。表3列出了中心性4(含4)以上的成员,共27人。

表3 H大学科研团队程度中心性

成员	程度中心性	成员	程度中心性	成员	程度中心性
2	12	21	5	8	4
10	8	24	5	55	4
9	8	20	5	133	4
5	8	23	5	11	4
7	7	26	5	87	4
3	7	58	4	33	4
1	6	15	4	52	4
6	6	48	4		
13	6	127	4		
25	6	45	4		

3 结语

在文献计量的基础上,运用社会网络分析技术,使用社会网络中的组件指标对学术科研团队的关系结构进行量化分析,可以更加直观高效地反映学术科研团队的全局状态,判断并分析学术科研团队的多层次结构及成员状况,发现学术科研团队中隐藏的各种关系。使用密度和中心性指标能对团队的合作程度及团队成员的学术地位进行揭示。本文认为,将文献计量学和社会网络分析结合起来在信息分析、信息服务领域有广阔的应用前景。在本研究的基础上可进一步分析团队的演化趋势、学科结构及知识交流情况。管理部门可依据本研究成果制定科研团队的发展规划并培植优秀团队。

参考文献:

- [1] 王怡然,陈士俊,张海燕,等.高校创新团队建设的若干理论问题研究[J].科技进步与对策,2007,24(8):194-197.
- [2] 王业军,刘艳艳.基于社会网络分析法的学术创新团队的研究[J].技术与创新管理,2009(3):276-279.
- [3] 汤建民.学术研究团队的可视化识别及评估方法研究:以科学学研究领域为例[J].情报学报,2010,29(2):323-330.
- [4] 文献计量学[EB/OL].http://baike.baidu.com/view/40533.htm.2011-04-02.
- [5] 钟伟金,李佳.共词分析法研究(一)——共词分析的过程与方式[J].情报杂志,2008,27(5):70-72.
- [6] GETOOR L, CHRISTOPHER P. Link mining: a survey [G]. SIGKDD Explorations, 2005, 7(2): 3-12.
- [7] 刘军.社会网络分析导论[M].北京:社会科学文献出版社,2004:12-13.
- [8] 刘蓓,袁毅,BOUTIN E.社会网络分析法在论文合作网中的应用研究[J].情报学报,2008,27(3):407-417.
- [9] 罗家德.社会网分析讲义[M].北京:社会科学文献出版社,2005:98-99.
- [10] FREEMAN L C. Centrality in social networks conceptual clarification[J]. Social Networks, 1978, 1(3): 215-239.

(责任编辑:陈晓峰)