

● 董克^{1,2}, 刘德洪², 江洪²

(1. 中国科学院 研究生院, 北京 100049; 2. 中国科学院 国家科学图书馆武汉分馆, 湖北 武汉 430071)

基于三方关系组的引用网络结构分析

摘要: 社会网络分析法可以分为不同的分析层次。本文首先论述三方关系组方法的原理, 并将其应用到引用网络结构的研究中, 分析各种三方关系类型在引用网络中的具体意义。以 SSCI 收录的 18 种图书情报学期刊近 10 年的文章为数据源, 运用三方关系组揭示了图书情报领域高产作者引用网络的内部结构, 通过计算不同类型聚类系数研究了网络整体结构倾向。

关键词: 社会网络分析; 引文分析法; 引用关系网络; 三方关系组

Abstract: Methods of social network analysis can be divided into different levels. This article first describes the principle of the triad census method, applies it in the research on the structure of citation network, and analyzes the specific significance of various triad census types in citation network. Taking the papers published by the 18 types of journals of library and information science in recent 10 years and included by SSCI as data sources, the article uses the triad census to reveal the internal structure of the citation network of the high-yield authors in the library and information science field, and by calculating the different types of clustering coefficients, studies the trend of the whole network structure.

Keywords: social network analysis; citation analysis method; citation network; triad census

社会网络分析方法和信息计量方法结合的相关研究由来已久, 1987 年就有学者应用网络分析方法研究科学交流中的知识流和“无形学院”, 并对信息计量学与社会网络的关系作了分析^[1]。随着分析技术的不断发展和研究条件的不断成熟, 社会网络分析在信息计量学中的应用越来越广泛, 并且克服了以往如数据量小等问题^[2]。现有的研究中, 合作网络被认为是研究社会网络的理想平台, 研究成果也较多^[3], 但利用社会网络分析结构类型方法对作者合作和引用网络进行的研究暂时还没有。结构类型方法在引用网络研究中更为适用, 本研究就是基于社会网络分析的三方关系组不同结构类型对作者引用网络进行分析。

1 社会网络分析的不同层次

社会网络分析有狭义与广义之分^[4], 狭义社会网络分析主要分为个体网络分析 (Ego-centric Network) 和整体网络分析 (Complete Network), 所涉及的指标如网络规模, 网络密度, 成分分析, 派系, 可达性, 区块模型。广义社会网络分析包括中心度, 弱连接, 桥问题等研究。

从具体分析的规模和对象上来看, 社会网络分析可以分为 5 个具体层次^[5], 相对应的度量系统测度不同的网络

属性。

1) 个体层次分析。分析主要针对网络中的节点, 即行动者本身的属性, 如点的出度、入度等。

2) 二方关系对分析。以两个行动者间的联结关系为研究主体, 如行动者之间关系的内容, 关系的互惠问题, 多元关系的交叉和分布等。

3) 三方关系组分析。以三方关系组的基本型为依据, 分析三个行动者间的可能发生的各种联结关系, 如关系与信息的传递和共享; 通过三方关系组计算相应聚类系数, 解释整个网络的属性。

4) 子图层次与多关系结构分析。主要是针对网络内点的归属, 包括点的聚类分类, 类间关系的相关性问题。

5) 整体网络分析。在某特定的范围内, 研究该范围内所有行动者的所有关系的状态, 还包括整体的统计分布规律, 如无标度性, 六度分离等。

在各种分析层次中, 对三方关系组理论及其应用的研究一直都比较欠缺^[6]。对引用网络中不同结构类型三方关系具有的传递和循环特性进行分析, 可以研究信息的传递和知识的共享; 通过计算不同类型的聚类系数, 能够判断信息和知识交流的程度, 对研究学科主题的发展具有重要的意义。

2 实证数据

ISI Web of Science 是信息计量领域的重要研究工具和数据来源,其中收录图书情报学期刊 60 余种,本研究参考 JCR-SSCI Edition,选取了 1999—2008 年均被 SSCI 收录排名前 18 的图书情报学期刊,以其刊载 1999—2008 年的 3 种主要文献类型 (Article, Proceedings Paper, Review) 为数据源,共获得记录 8 623 条,其中包括作者 13 313 位,第一作者 5 677 位,被引作者 93 267 位,选取发文量 10 篇以上的高产作者 105 人,构成 105 × 105 引用关系矩阵并进行二值化。

部分高产作者和矩阵片段如表 1 和表 2 所示。社会网络分析软件使用 Ucinet^[7] 和 Pajek^[8]。

表 1 部分高产作者

编号	作者姓名	发文数 (篇)
N1	Thelwall M	70
N2	Egghe L	58
N3	Oppenheim C	48
N4	Rousseau R	46
N5	Bates D W	43
N6	Glanzel W	42
N7	Spink A	39
N8	Leydesdorff L	38
N9	Cimino J J	34
N10	Benbasat I	29

表 2 作者引用关系矩阵片段

编号	N1	N2	N3	N4	N5	N6
N1	1	1	1	1	0	1
N2	0	1	1	1	0	1
N3	1	1	1	0	0	1
N4	1	1	1	1	0	1
N5	0	0	1	0	1	0
N6	0	1	0	1	0	1

3 作者引文网络的三方关系组

在研究中利用 Pajek 或 Netdraw 等软件进行可视化,往往会遇到节点数量比较大,从可视化结果上能够进行的分析受到限制的问题,可视化结果的意义也被削弱。在引用关系网络中,多个作者之间的引用关系,可以分解为若干个三人引用关系,这些三人构成的关系在社会网络分析中有专门的名称——三方关系组 (Triad Census), 这种分析方法对于研究作者引用关系具有重要的借鉴意义。

3.1 引文网络中的三方关系组

三方关系组可以有 16 种表现形式 (见图 1)^[9]。图形

下方 1~16 表示序号,其排序是按照“对称关系—不对称关系—无关系”模型来列出的,例如 021 指对称关系数量为 0,不对称关系数量为 2,无关系数量为 1。最后的字母 D (Down) 表示关系方向向下, U 表示关系方向向上 (Up), C 表示为循环 (Cycle), T 表示传递 (Transitive), 三方关系组分析方法主要应用于研究不同的社会交流结构,这些字母是基于一些社会学的假设。下面针对不同的三方关系组在作者引用关系中的意义进行说明。

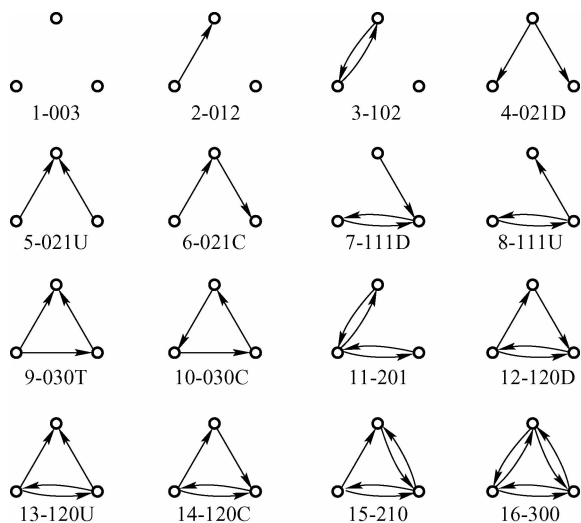


图 1 16 种可能出现的三方关系组

箭头表示作者之间的引用关系 (假定 A→B 表示 B 引用 A), 在作者引用网络可能出现全部 16 种三方关系。其中 003 表示三人之间完全无关系, 012 和 102 实质上是两人之间的关系, 这三种都是特殊的三方关系。剩下 13 种三方关系代表三人之间产生的关系, 从结构上可分为非封闭三方关系和封闭三方关系两组 5 个类型 (见表 3)。

表 3 13 种三方关系的不同类型

组系	类型	序号
非封闭三方关系组	基础型	021D
		021U
		021C
	发展型	111D
		111U
封闭三方关系组	传递型	201
		030T
		120D
	循环型	120U
		030C
	综合型	120C
210		
		300

非封闭三方关系表示彼此之间的交流存在一定的缺失, 其中的基础型是所有三方关系中最简单的形态。021D、021U 与 021C 表示两两之间不存在相互引用关系, 互相交流程度较低, 中间点在关系中处于核心地位, 其他两个点处于从属地位。021D 构成了最简单的文献耦合关系, 而 021U 则代表了共被引的最简单形态, 021C 是一个单向的信息流动过程。发展型是基础形态最简单的扩充, 111D、111U 和 201 这三种发展型三方关系与基础型相比均有一个相互引用的关系对, 表明作者之间的关系相对紧密, 但仍然存在彼此间无交流情况, 不能构成一个完整的引用关系环。

随着交流程度的进一步提高, 在非封闭三方关系基础上逐步出现了三人之间均存在直接交流的情况, 封闭三方关系由此产生。030T、120D 和 120U 三种类型体现了一种引用关系的传递。援引社会网络理论“禁现三方组”假设, 三人中如果 A 引用了 B, B 引用了 C, 则就会出现 A 引用 C 情况, 这个假设在引用网络中并不一定成立, 但特定的形态构成了一个类型, 即引用关系的传递性类型。030C 体现了引用关系之间的循环, 在这类结构中, 每个人既充当了信息的产生者, 同时也充当了信息的传递者, 是比较典型的知识交流团体。120C、210 和 300 三种三方关系既包含了关系的传递, 也包含了关系的循环, 属于具有综合特征的三方关系类型。210 和 300 所包含的关系数量是 13 种三方关系中最多的, 是理想的知识交流模型; 300 形态表示三人之间的关系最为紧密, 是“无形学院”产生的重要基础。

3.2 三方关系组统计分析

表 4 显示了网络中 13 种三方关系组的各项信息, 其中第一列为三方关系类型代号, 第二列为网络结构中实际数量, 第三列为随机模型计算出的数量。其卡方值 123010. 1025 *** (*** 表示在 0.001 的显著水平) 显示网络结构中各种三方关系数量分布明显不符合随机分布规律, 表明图书情报领域的高产作者引用网络是一个存在明显结构倾向的网络结构。在 13 种三方关系组结构中, 021D、021U 与 021C 等三种基础型三方关系实际数值均低于随机产生值, 而在此基础上发展型三方关系 111D、111U、201 的实际数值均高于随机产生值。这说明图书情报领域的高产作者相互引用的情况明显, 高产作者之间的学术交流和信息交流较为充分。

如果两个人同时引用第三人, 或者两个人同时被第三人所引用, 则这两人之间存在一定科学关系, 这是耦合和共被引研究的基本理论假设, 有学者认为这是在用结果的有效性说明方法的有效性^[10]。030T、120D 和 120U 这三种传递型三方关系类型在网络中的实际数量均超过了随机

表 4 引用网络三方关系组程序分析结果

类型	实际数量	随机模型数量
4 - 021D	2401	4247. 01
5 - 021U	2457	4247. 01
6 - 021C	4014	8494. 02
7 - 111D	2411	1045. 7
8 - 111U	1513	1045. 7
9 - 030T	1171	1045. 7
10 - 030C	133	348. 57
11 - 201	465	64. 37
12 - 120D	382	64. 37
13 - 120U	333	64. 37
14 - 120C	331	128. 74
15 - 210	499	15. 85
16 - 300	151	0. 33

模型产生数量; 120C、210 和 300 三种兼具传递型特征综合型三方关系在网络中的实际数量也高于随机模型产生数量。这表明在图书情报领域的高产作者中, 如果两人之间存在耦合关系或者共被引关系, 两者之间存在联系的可能性大大增加。这个结果在一定程度上证明了耦合和共被引方法理论理性。

循环型三方关系 030C 在网络中的实际数量小于随机模型产生数量, 但由于 120C、210 和 300 三种综合型三方关系同时也具备了循环型的特点, 且在社会学假设中, 120C 与 030C 属性更接近, 这个统计结果表明在图书情报领域高产作者引用网络中, 由于间接引用催生出直接引用情况是比较常见的现象。

3.3 聚类系数分析

集聚性是社会网络有别于其他网络结构的一个重要特征。集聚性体现在两个方面, 一是网络中节点之间发生关系的可能性; 二是不同类型的三方关系出现的比率。节点之间发生关系的可能性主要利用实际存在的边与网络中可能存在的边书目的比值确定^[11]; 利用三方关系组计算网络的聚类系数则可以分析网络中关系传递和关系循环的具体情况, 以此判断整体的结构倾向^[12]。合著网络关系紧密程度的聚类系数可以用如下方法进行计算^[3]:

$$C = \frac{3 * \text{网络中三角形的个数}}{\text{顶点关联的三点组个数}}$$

合著网络由于作者之间的关系是双向的, 因此完全的交流形态只有一种, 从结构形态能够计算的聚类系数也只有一个。作者引用关系网络和合著网络的不同在于其有向性, 三方关系类型的丰富使引用网络的具有多个聚类系数。令 N 表示各种三方关系组数量, 则可以定义传递性

聚类系数与循环性聚类系数计算公式如下:

$$C_{传} = \frac{N_{030T} + N_{120D} + N_{120U} + N_{120C} + N_{210} + N_{300}}{6N}$$

$$C_{循} = \frac{N_{030C} + N_{120C} + N_{210} + N_{300}}{4N}$$

根据表4的数据可以计算出105人引用关系网络的传递型聚类系数 $C_{传} = 0.02939$, 循环型聚类系数 $C_{循} = 0.014074$ 。 $C_{传} > C_{循}$, 从整体上进一步证明图书情报领域高产作者引用网络中各人彼此之间的相互交流存在不足, 但存在学术权威, 引用关系整体上存在方向性。

4 结束语

网络分析法与个体分析的优势在于, 在相互的联系中探讨研究各个对象以及行动者系统的属性。以上分析了三方关系的不同类型及在引用网络中的具体含义, 在此基础上计算了图书情报领域高产作者引用网络中不同类型三方关系数量, 对比分析了图书情报领域高产作者引用网络所具有的结构特征, 在一定程度上证明了耦合和共被引方法理论假设的合理性; 进一步通过计算聚类系数研究了网络整体的结构倾向。

在实证分析中发现, 作为一种网络基本结构类型研究方法, 三方关系组分析方法为引用网络的研究提供了一个新的视角, 如果对不同结构类型从结构层次或时间的角度进一步进行分析, 能够得出更为丰富的结论, 这将是以后研究的重要方向。□

参考文献

- [1] LIEVROUW L A, et al. T-riangulation as a research strategy for identifying invisible colleges among biomedical science [J]. Social Net-works, 1987 (9): 217-238.

(上接第56页)

事实也表明只有网络知识产权才能对智力创造活动产生足够的激励, 平衡好各方的利益要求, 促进更多创新性智力成果的产生, 持续推动网络文化事业的进步。□

参考文献

- [1] 卢山. 网络知识产权的特征及其保护 [J]. 钦州师范高等专科学校学报, 2005 (2): 74-77.
- [2] 钱锋. 网络环境下知识产权的矛盾冲突及对策研究 [D]. 南京: 东南大学, 2004.
- [3] 王金玲. 电子商务的法律与规范 [M]. 沈阳: 东北大学出版社, 2003: 200.
- [4] 郑笑寒. 论知识产权理论的哲学反思——以创新收益为视角 [D]. 南昌: 南昌大学, 2007.
- [5] 陶鑫良, 袁真富. 知识产权法总论 [M]. 北京: 知识产

- [2] NEWMAN M E J. Scientific collaboration networks. I. Network construction and fundamental results [J]. Physical Review E, 2001, 64 (1).
- [3] NEWMAN M E J. The structure of scientific collaboration networks [J]. PNAS, 2001, 98 (2): 404, 408.
- [4] 汪丹. 科学合作中的媒介角色 [D]. 北京: 中国科学院文献情报中心, 2009: 11-13.
- [5] WASSERMAN S, FAUST K. Social network analysis: methods and application [M]. New York: Cambridge University Press, 1994.
- [6] KIDUFF M, TSAI W. 社会网络与组织 [M]. 王凤彬, 等译. 北京: 中国人民大学出版社, 2006: 7.
- [7] BORGATTI S P, EVERETT M G, FREEMAN L C. Ucinet for windows: software for social network analysis [M]. Harvard, MA: Analytic Technologies, 2002.
- [8] <http://vlado.fmf.uni-lj.si/pub/networks/pajek>.
- [9] BATAGELJ V, MRVAR A. A subquadratic triad census algorithm for large sparse networks with small maximum degree [J]. Social Networks, 2001 (23): 237-243.
- [10] LEYDESDORFF L. Various methods for the mapping of science [J]. Scientometrics, 1987 (5/6): 295-324.
- [11] WATTS D J. Small worlds: the dynamics of networks between order and randomness [M]. Princeton: Princeton University Press, 2003: 33.
- [12] SCHANK T, WAGNER D. Approximating clustering coefficient and transitivity [J]. Journal of Graph Algorithms and Applications, 2005, 9 (2): 265-275.

作者简介: 董克, 男, 1986年生, 硕士生。

刘德洪, 男, 1962年生, 研究员。

江洪, 女, 1968年生, 副研究员。

收稿日期: 2010-08-02

权出版社, 2005: 17-18.

- [6] 冯晓青. 知识产权法利益平衡理论 [M]. 北京: 中国政法大学出版社, 2006: 511.
- [7] 蒋玉宏, 单晓光. 知识产权制度对城市竞争力的影响——基于创新激励的机理分析 [J]. 知识产权, 2007 (3): 26-31.
- [8] 周寄中, 徐倩云. 知识经济中的知识产权制度及其激励功能 [J]. 研究与发展管理, 2002 (2): 51-56.
- [9] 冯晓青. 论利益平衡原理及其在知识产权法中的适用 [J]. 江海学刊, 2007 (1): 141-146.
- [10] 费安玲. 论著作权的正当性——历史的透视与权利要素理论的思考 [J]. 知识产权研究, 2004 (4): 42-51.

作者简介: 王英, 女, 1985年生, 硕士生。

收稿日期: 2010-07-14