

信息管理系列丛书

SHEHUIWANGLUO FENXIFANGFA ZAI
TUSHUQINGBAOXUEKE DE YINGYONGYANJIU

姜鑫 王德庄 马海群 著

社会网络分析方法在 图书情报学科的应用研究



知识产权出版社

全国百佳图书出版单位

本书受 2015 年度黑龙江省哲学社会科学研究规划专项项目“我国图书情报学科知识结构及动态演化的可视化研究”(15TQD03)、2016 年度黑龙江省高校基本科研业务费专项资金项目“我国图书情报学科知识结构的动态演化研究：基于科学知识图谱视角”(HD-JDY201615) 资助

社会网络分析方法在图书 情报学科的应用研究

姜 鑫 王德庄 马海群 著



知识产权出版社

全国百佳图书出版单位

图书在版编目 (CIP) 数据

社会网络分析方法在图书情报学科的应用研究/姜鑫, 王德庄, 马海群著. —北京: 知识产权出版社, 2019. 2

ISBN 978 - 7 - 5130 - 6057 - 8

I. ①社… II. ①姜… ②王… ③马… III. ①社会网络—分析方法—应用—图书情报学—研究 IV. ①G250

中国版本图书馆 CIP 数据核字 (2019) 第 015926 号

内容提要

本书通过大量实例并结合 Ucinet 软件工具系统介绍社会网络分析方法在图书情报学科应用的原理、方法与案例, 具体内容涉及中心性分析、凝聚子群分析、结构洞分析、核心-边缘结构分析等社会网络分析方法在知识管理学、文献计量学、网络计量学等图书情报学科分支领域的应用研究。本书可作为图书情报学科及社会科学相关学科教师、学生及相关研究人员参考用书。

责任编辑: 许 波

责任印制: 孙婷婷

社会网络分析方法在图书情报学科的应用研究

姜 鑫 王德庄 马海群 著

出版发行: 知识产权出版社有限责任公司

网 址: <http://www.ipph.cn>

电 话: 010-82004826

<http://www.laichushu.com>

社 址: 北京市海淀区气象路 50 号院

邮 编: 100081

责编电话: 010-82000860 转 8380

责编邮箱: xubo@cnipr.com

发行电话: 010-82000860 转 8101

发行传真: 010-82000893

印 刷: 北京虎彩文化传播有限公司

经 销: 各大网上书店、新华书店及相关专业书店

开 本: 720mm×1000mm 1/16

印 张: 14.5

版 次: 2019 年 2 月第 1 版

印 次: 2019 年 2 月第 1 次印刷

字 数: 282 千字

定 价: 58.00 元

ISBN 978-7-5130-6057-8

出版版权专有 侵权必究

如有印装质量问题, 本社负责调换。

前 言

社会网络 (social network) 是一种基于“网络”(节点之间的相互连接) 而非“群体”(明确的边界和秩序) 的社会组织形式, 也是西方社会学从 20 世纪 60 年代兴起的一种分析视角。随着工业化、城市化的发展和新的通信技术的兴起, 社会越来越呈现网络化的趋势。在 2012 年出版的 *Networked: The New Social Operating System* 一书中, Lee Rainie 和 Barry Wellman 将社会网络革命 (social network revolution)、移动革命 (mobile revolution) 与互联网革命 (internet revolution) 并列为新时期影响人类社会的三大革命。

社会网络分析 (social network analysis, SNA) 以行动者及其相互之间的关系作为研究内容, 通过对行动者之间的关系模型进行描述, 分析这些模型所蕴含的结构及其对行动者和整个群体的影响。社会网络分析可以应用于群体、部门、组织和组织间等多个分析层次, 具有较宽泛的应用领域和较强的问题解释能力。20 世纪 80 年代起, 社会网络分析的思想和方法被介绍到国内, 主要应用于社会学领域。图书情报学界对社会网络分析的研究始于人际网络分析。2003 年, 包昌火、谢新洲和申宁在《情报学报》上发表《人际网络分析》, 这是国内图书情报学界应用社会网络分析研究的开始。目前, 图书情报学科应用社会网络分析法的研究成果不断涌现, 已成为国内社会网络分析应用最重要的学科领域。

本书通过大量实例并结合 Ucinet 软件工具系统地介绍了社会网络分析方法在图书情报学科应用的原理、方法与案例, 具体内容涉及中心性分析、凝聚子群分析、结构洞分析、核心-边缘结构分析等社会网络分析方法在知识管理学、文献计量学、网络计量学等图书情报学科分支领域的应用研究, 希

望能为图书情报学科及社会科学相关学科的研究人员进行有关研究提供参考。本书第一章、第三章、第七章由姜鑫撰写，第二章、第六章由姜鑫、马海群撰写，第四章、第五章、第八章由姜鑫、王德庄撰写，第九章、第十章由姜鑫、王德庄、马海群撰写。本书撰写过程中参考了大量中外学者的论著，在此一并表示衷心的感谢！

目 录

基础篇

社会网络分析 (SNA) ——理论、方法与软件

第一章 社会网络资料的收集	3
第一节 社会网络资料的相关概念	3
第二节 社会网络资料的收集方法	12
第二章 社会网络的数学表达	23
第一节 关系网络的图形表达	23
第二节 关系网络的矩阵表达	27
第三章 社会网络的基本概念	41
第一节 与“关联性”有关的概念	41
第二节 与“距离”有关的概念	49
第四章 中心性分析	60
第一节 点度中心度	62

第二节 中介中心度	66
第三节 接近中心度	70
第五章 凝聚子群分析	77
第一节 基于距离的凝聚子群	77
第二节 基于度数的凝聚子群	85
第六章 结构洞分析	93
第一节 结构洞的相关概念	94
第二节 结构洞的测度指标	96
第七章 核心-边缘结构分析	107
第一节 核心-边缘结构的特点	107
第二节 核心-边缘结构的类型	108

实践篇

社会网络分析 (SNA) 在图书情报学科的应用

第八章 SNA 在知识管理领域的应用	117
第一节 中心性分析在知识管理中的应用	120
第二节 结构洞分析在知识管理中的应用	121
第九章 SNA 在文献计量学中的应用	133
第一节 SNA 在共词分析中的应用	133
第二节 SNA 在共被引分析中的应用	153
第三节 SNA 在耦合分析中的应用	164
第四节 SNA 在互引分析中的应用	174

第十章 SNA 在网络计量学中的应用	187
第一节 SNA 在网络关联性研究中的应用	187
第二节 SNA 在网络信息交流研究中的应用	195
附 录	
附录 1 91 位用户的微博地址	207
附录 2 143 位用户的微博地址	211
参考文献	216

基础篇

社会网络分析(SNA)——理论、方法与软件



第一章 社会网络资料的收集

社会网络分析 (social network analysis, SNA) 是在人类学、心理学、社会学、数学及统计学等领域中发展起来的, 在国外是从 20 世纪 30 年代末出现并在最近 20 多年里得到重要发展。社会网络分析是研究社会结构的最新方法和技术, 目前已经形成了一系列专有术语和概念, 并已成为一种全新的社会科学研究范式。社会网络分析所用的数据资料有其自身的类型与特征, 本章主要介绍社会网络资料及其收集方法。

第一节 社会网络资料的相关概念

首先简单介绍社会网络的相关概念。“社会网络”(social network) 是指社会行动者 (social actor) 及他们之间关系的集合。也可以说, 社会网络是由多个节点 (社会行动者) 和各个节点之间的连线 (代表行动者之间的关系) 所组成的集合^①。社会网络分析中的节点代表社会行动者, 连线代表行动者之间的关系。“行动者”(actors) 可以是个体、公司、学校乃至组织、国家等。关于节点的信息既可以是静态的, 也可以是动态的。行动者之间的“关系”(ties) 是指具体的关联内容 (relational content) 或者现实中发生的实质性的关系。行动者之间的关系类型是多种多样的, 可以是个人之间的朋友关系、上下级关系, 也可以是城市之间的距离关系、国家之间的国际贸易关系等。

社会网络分析中的模 (mode) 是指社会行动者的集合, 模数是指社会行动者集合类型的数目。由一个行动者集合内部各个行动者之间的关系所构成

① 刘军. 整体网分析讲义——UCINET 软件实用指南 [M]. 上海: 上海人民出版社, 2009: 1-3.

的网络称为1-模网络 (one-mode network)。例如, 一个公司内 150 名员工之间的朋友关系网络就是 1-模网络。由一类行动者集合与另一类行动者集合之间的关系所构成的网络称为 2-模网络 (two-mode network)。还有一类特殊的 2-模网络被称为“隶属网络” (affiliation network)。隶属网络中的两个模态分别是行动者和事件, 即一个模态为行动者集合, 另外一个模态为这些行动者所“隶属”的“事件” (如俱乐部或自愿者组织等)。

社会网络资料是指一组反映社会行动者之间关系的数据, 它是进行社会网络分析的基础和材料。首先, 社会网络资料是关于社会关系的数据信息, 简称关系数据 (relational data)。其次, 关系数据有一定的表现形式, 即通过一定的变量或数据把关系特征、结构等反映出来^①。所以, 关系数据不同于属性数据 (attribute data), 不仅其本质内容不同, 其表现形式也不同。属性数据是关于行动者的自然状况、态度、观点以及行为等方面的数据, 一般被视为个人或者群体所具有的财产、性质、特点等属性。通过调查和访谈收集到的一些资料通常被看成是特定个体的一些属性, 需要运用现有的统计程序对它们进行定量分析。适用于分析属性数据的方法主要是“变量分析” (variable analysis), 如相关分析、回归分析、列联分析等。各种属性被看成是特定变量 (如年龄、性别、学历、收入等) 的取值。在进行变量分析时, 属性数据的矩阵表现形式由个案和变量所构成; 矩阵的行表示个案 (case), 列表示变量 (variable), 如图 1-1 所示。

		变 量			
		年龄	性别	收入	...
个 案	1	属性数据			
	2				
	3				
	4				
	5				
	⋮				

图 1-1 属性数据矩阵

① 林聚任. 社会网络分析: 理论、方法与应用 [M]. 北京: 北京师范大学出版社, 2009: 65 - 68.

这种个案 - 变量数据矩阵 (case-by-variable matrices) 不能用来表示关系数据。关系数据的矩阵表示形式为个案 - 隶属关系矩阵 (case-by-affiliation matrices): 矩阵的行表示个案 (case), 即一些特定的行动者, 他们构成了分析单元; 矩阵的列表示隶属项 (affiliation), 即这些行动者所卷入的组织、事件或活动等。用矩阵的各个列代表各个隶属项, 可以区分出哪些行动者参与了何种隶属项。从这种个案 - 隶属关系矩阵中, 可以引出关于行动者之间的直接关系和间接关系的信息。例如, 图 1-2 就是一个简单的个案 - 隶属关系矩阵, 它表示了 3 个人 (标记为 1、2、3) 参与 3 个事件 (标记为 A、B、C) 的情况。如果有某人参与了某一事件, 则将矩阵对应交叉处的格值记为 “1”, 否则记为 “0”。由图 1-2 可见, 3 个人都参与了事件 A, 都没有参与事件 B 和事件 C。与这个关系矩阵相对应的社群图表示了一个在个体之间存在相互联系的简单的三人结构。可以将这个社群图解读为: 每个人在某一特定事件过程中都与另外两个人存在联系^①。

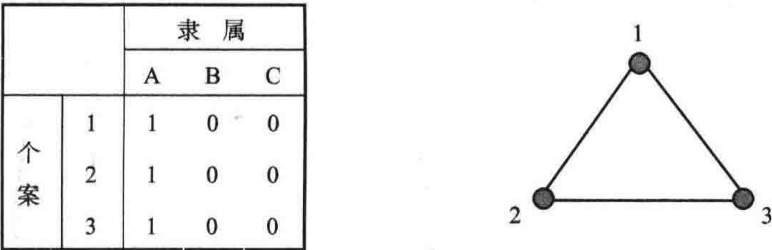


图 1-2 一个简单的关系矩阵及其社群图

图 1-3 表示了社会网络关系矩阵的三种一般形式, 最基本的形式就是前面提到的个案 - 隶属关系矩阵 [图 1-3 (a)], 其中各行表示行动者, 各列表示行动者的隶属项。因为行和列表达的是不同的数据集合, 因此这是一类 2- 模 (2-mode) 长方形矩阵, 矩阵中的行数和列数通常是不相等的。从这种基本的长方形矩阵中可以推导出两个 1- 模 (1-mode) 方阵, 其中一个是个案 - 个案方阵 [图 1-3 (b)], 其中行和列都代表个案 (case), 矩阵中的每个格值表示特定的一对行动者之间是否由于共同隶属于一个事件而关联在一

① 约翰·斯科特. 社会网络分析法 [M]. 刘军, 译. 2 版. 重庆: 重庆大学出版社, 2009: 33-51.

起。因此，个案－个案方阵表示了行动者之间实际存在的关系，并且该矩阵与其对应的社群图所表达的信息是等同的。另一个是隶属－隶属方阵[图 1-3 (c)]，其中行和列都代表隶属项 (affiliation)，矩阵中的每个格值表示特定的一对隶属项之间是否由于拥有共同的行动者而关联在一起。这种隶属－隶属方阵在社会网络分析中非常重要，因为它通常可以展现出个案－个案方阵所不能明显反映的有关社会结构的重要信息。

		隶 属				
		A	B	C	D	E
个 案	1					
	2					
	3					
	4					

(a) 个案－隶属关系矩阵

		个 案			
		1	2	3	4
个 案	1				
	2				
	3				
	4				

(b) 个案－个案方阵

		隶 属				
		A	B	C	D	E
隶 属	A					
	B					
	C					
	D					
	E					

(c) 隶属－隶属方阵

图 1-3 社会网络关系矩阵三种形式

一个简单的 2-模长方形数据矩阵可以转换成两个 1-模方阵，其中一个描述了初始阵的各行，另外一个描述了初始阵的各列。长方阵和两个方阵都是针对同一关系数据的相同表达方式，长方阵通常称为“发生阵”(incidence matrix)，而两个方阵又称为“邻接阵”(adjacency matrix)。研究者通常需要将所收集到的个案及其隶属关系的 2-模数据整理成一个发生阵，然后将这个发生阵转换为社会网络分析中使用的 1-模邻接阵。在某些情况下，研究者也可以直接收集个案－个案方阵形式的关系数据，如在一个公司中员工之间的

朋友关系就可以直接构建方阵。以下通过一个具体的实例来说明如何应用 Ucinet 软件工具分析关系矩阵。图 1-4 是一个虚构的 4 个公司之间的连锁董事数据。当某个人兼任两个或者多个公司的董事时，则可认为存在一个“连锁董事”（interlocking directorates），并使其兼任董事的两个或多个公司建立了联系。通常将公司看成是个案，将各个公司共享的董事看成是隶属项。

		董 事				
		A	B	C	D	E
公 司	1	1	1	1	1	0
	2	1	1	1	0	1
	3	0	1	1	1	0
	4	0	0	1	0	1

(a) 公司 - 董事发生阵

	1	2	3	4
1	—	3	3	1
2	3	—	2	2
3	3	2	—	2
4	1	2	2	—

(b) 公司 - 公司邻接阵

	A	B	C	D	E
A	—	2	2	1	1
B	2	—	3	2	1
C	2	3	—	2	2
D	1	2	2	—	0
E	1	1	2	0	—

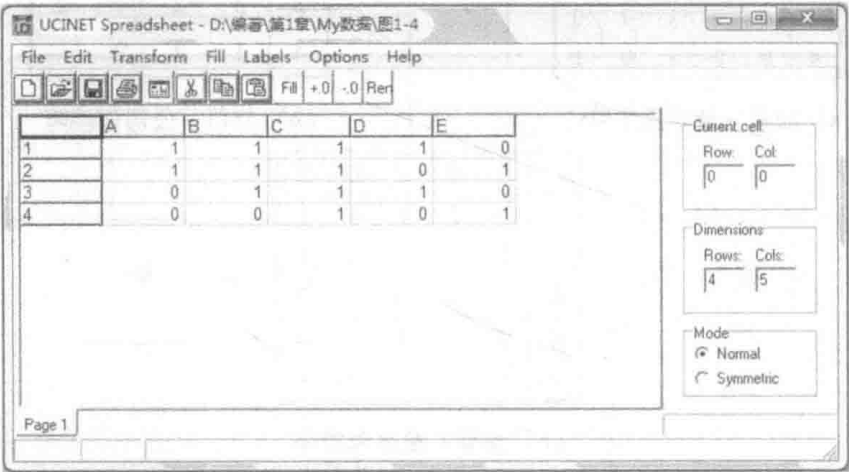
(c) 董事 - 董事邻接阵

图 1-4 4 个公司之间的连锁董事数据

由图 1-4 (a) 可见，公司 1 有 4 个董事（A、B、C、D），而董事 A 同时也是公司 2 的董事，这表明公司 1 和公司 2 之间有关联。图 1-4 (b) 表达了所有公司之间的连锁董事情况，矩阵中的每个格值不仅表示了连锁董事是否存在，还表示了每对公司之间共享董事的个数。每个格值并不是简单的二进制数，因为公司之间共享的董事可能多于一个。从初始发生阵的列中可以看出，公司 1 和公司 4 只有一个共同的董事，即 C；而公司 2 和公司 3 则有两个共同的董事，即 B 和 C。关系的强度可以通过关系本身包含的“连锁”数来测量。“最强的”关系存在于共享董事数目最多的公司之间，即公司 1 和公司 2 之间以及公司 1 和公司 3 之间，每组之间都共享 3 个董事。“最弱的”关系存在于只共享一个董事的公司之间，即公司 1 和公司 4 之间。

公司 - 董事发生阵（图 1-4 (a)）在 Ucinet 软件中可以通过以下两种方法输入。

(1) 利用 Ucinet 软件本身的数据表程序 (spreadsheet) 直接输入。单击 Ucinet 主界面中的数据编辑器快捷方式 (matrix spreadsheet) 或者单击 “Data → Data editors → Matrix editor”, 就可以打开 Ucinet 软件的数据编辑器, 可以在数据编辑器中直接键入数据, 也可以利用 “复制” “粘贴” 等方式将 Excel 文件中的数据复制到 Ucinet 软件中。矩阵的行数 (Rows) 和列数 (Cols) 的默认值是 30, 分别将行数和列数修改为 4 和 5, 输入完毕后命名为 “图 1-4”, Ucinet 软件将会自动生成两个文件: “图 1-4. ##d” 和 “图 1-4. #h” (图 1-5 (a))。



(a) 用 “数据编辑器” 输入初始发生阵



(b) 用 “文本编辑器” 输入初始发生阵

图 1-5 在 Ucinet 软件中输入初始发生阵的两种方法

(2) 利用文本编辑器输入。在 Ucinet 软件中, 常见的数据输入方法是在一个文本文件中输入数据, 可以利用任何一种字处理程序, 如 “写字板” 或者 Word; 或者在 Ucinet 软件中单击 “File → Text Editor”, 打开文本编辑器,

也可以输入数据文件。采用关联列表形式 (linked list formats) 输入数据, 只需指定数据中实际有关联的关系。关联列表形式可分为两类: 点列表 (nodelists) 和边列表 (edgelists)。点列表形式 (nodelist format) 又为: ①点列表形式-1 (nodelist1), 用于输入1-模方阵数据; ②点列表形式-2 (nodelist2), 用于输入长方形的2-模矩阵数据; 两者都只能用来输入1-0矩阵数据。由于公司-董事发生阵是2-模矩阵, 故采用点列表形式-2 (nodelist2), 在文本编辑器中输入相应语句 (图1-5 (b))。

关键词“dl” (代表 data language, 即数据语言) 是必需的, 它界定了该文件是一个数据语言文件 (DL filetype), 并且必须放在文件的最前端。长方形的2-模矩阵需指定矩阵的行数和列数, 语句“nr = 4”表示矩阵包含4行 (其中, nr 是“number of rows”的缩写), “nc = 5”表示矩阵包含5列 (其中, nc 是“number of columns”的缩写)。DL 文件也可以包含行动者的标签 (lables), 并可在标题行中加入标签信息。标签的长度最长为18个字符, 不能包含空格或者逗号, 建议用英文符号表示标签。可以针对行和列分别指定标签, 即行标签 (row labels) 和列标签 (column labels), 也可以用 embedded 命令将标签嵌入到数据中。关键词“data:”表示标题信息 (关于数据的信息) 结束, 接下来就是数据本身了。语句“format = nodelist2”指定一种数据形式, 即每行中的第一个数字指定一个“个案”, 其关系指向与之对应的“隶属项”, 例如, 公司1共享董事A、B、C、D, 公司2共享董事A、B、C、E, 公司3共享董事B、C、D, 公司4共享董事C、E。在输入上述数据之后, 在指定文件夹中保存该文件成为纯文本文件, 并且给该文件命名为“图1-4.txt”。如果将该文件中的信息转换为Ucinet数据, 需要在Ucinet软件中单击“Data → Import text files → DL”, 选择保存的文本文件“图1-4.txt”, 即可打开成为Ucinet形式的数据文件。

在Ucinet软件中单击“Data → Affiliations (convert 2-mode data to 1-mode)”, 可以将2-模发生阵转换为2个1-模邻接阵。其中, 模态 (which mode) 有两种选择: 行模式 (row) 生成以个案为计算标准的矩阵, 即个案-个案邻接阵 (公司-公司邻接阵); 列模式 (column) 生成以隶属项为计算标准的矩阵, 即隶属-隶属邻接阵 (董事-董事邻接阵)。转换的方法 (method) 有两种: ①对应乘积法 (Cross-Products (co-occurrence)), 这是默认的