

我国城市信息化发展现况研究——基于文献计量与社会网络分析

李万辉 张文德 陈振标

福州大学信息管理研究所 福州 350108

[摘要] 以城市信息化学术论文为研究对象,运用文献计量分析、共词分析、内容分析、社会网络分析方法,对1990-2009城市信息化部分学术论文进行统计分析,以期揭示研究热点和薄弱点。

[关键词] 城市信息化 文献计量分析 共词分析 社会网络分析

1 引言

在经济全球化和全球信息化的大背景下,城市作为全球化经济的网络节点和区域经济中心,在推动国家和区域全球化和信息化过程中面临新的功能转移。城市信息化概念的提出,源于前美国副总统戈尔于1998年提出的“数字地球”,但城市信息化这一概念最早出现于我国的信息规划中。本文采用文献计量学方法,对论文作者、单位、引文期刊、年代分布情况进行统计,对论文摘要自动分词与词频统计,在论文关键词共词分析的基础上,结合社会网络分析法,实现数据可视化,以期把握该领域的研究热点与发展趋势。

2 研究方法、数据来源和处理程序

2.1 研究方法

学术期刊作为各类科研项目成果发表的主要形式,对其进行统计分析,可以较为客观,全面地反应该领域的研究情况。本文主要采用文献计量分析、共词分析、内容分析、社会网络分析方法进行研究。文献计量分析(Bibliometrics Analysis)是情报学独有的分析方法,在科学学与科技管理、科学评价与预测等学科领域及其定量管理等领域得到了广泛应用,无疑也是学科分析的最好方法之一^[1]。共词分析法的思想来源于文献计量学的引文耦合与共被引概念,即当两个能够表达某一学科领域研究主题或研究方向的专业术语(一般为主题词或关键词)在同一篇文献中出现时,表明这两个词之间具有一定的内在关系,并且出现的次数越多,表明关系越密切、距离越近^[2]。内容分析法是一种对研究对象的内容进行深入分析,透过现象看本质的科学方法^[3]。社会网络分析(Social Network Analysis, SNS)发端于20世纪50年代,最初用手心理心研究后被用于社会学、人类学、科学学等其它领域。作为一种分析方法,它是对社会网络中行为者之间相互关系进行量化研究,将被分析对象视为社会行动者和他们关系的集合^[4]。研究的原理在于,关键词是网络中的一个节点,而它们的共现则体现为节点之间有直接的联系。在虚拟的关键词网络中,通过关键词共现的频次不同,可以定量地确定某个关键词在网络中的地位,它反映了词与词之间的关系^[5]。它对了解某一主题领域的研究热点、成熟度、知识结构、研究的规模等情况具有非常重要的意义。

2.2 数据来源

笔者以篇名或关键词为“城市信息化”；时间界定为 1990 年 1 月 1 日—2009 年 12 月 31 日；选择 CNKI 中国学术文献网络出版总库作为数据源，并以中国博士学位论文全文数据库和中国优秀硕士学位论文全文数据库为补充；匹配模式为“精确”；检索得到文献记录共 1007 条，其中核心期刊载文 118 篇，占论文总量的 12%，表明该领域的研究具有较高的研究质量。

2.3 数据处理

2.3.1 数据处理软件工具

本研究主要用到的软件工具有微软公司开发的 Microsoft Excel，中国医科大学医学信息系开发的 BICOMB 软件，IBM 公司开发的 SPSS 软件，武汉大学虚拟学习团队开发的 ROST CM 软件以及由加州大学欧文大学开发的 UCINET 软件。

2.3.2 数据处理流程

利用书目共现分析系统 BICOMB 和 Excel 对数据进行初步处理。

- 在将数据导入到 BICOMB 系统后，分别进行作者、作者单位、期刊、关键词、年代进行提取并做相应的统计。

- 将作者、作者单位、期刊、年代相关数据分别导入到 Excel 表中，进行相关统计分析后，制得相关统计图。

- 高频主题词统计：利用 ROST CM 软件对论文摘要进行自动分词与词频统计，并制得主题词高频表。

- 高频关键词统计：在步骤 1 的基础上，我们共得到 8 229 个关键词，去重后得到 4 532 个关键词，篇均关键词为 4.5 个。根据研究要求，取词频大于等于 10 的关键词（60 个）作为表现当前城市信息化领域研究热点的标志，利用 BICOMB 软件输出一个 60*60 的高频关键词共现矩阵。考虑到关键词的规范性问题，并在参考步骤 3 研究结果的基础上，对相近的同义词进行合并，如电子政务与电子政务建设合并为电子政务，数字城市与数字化城市合并为数字城市，以及去掉一些非高频主题词以及无意义词，如地名：上海，北京，广州，数字奥运等词。经处理后，得到一个 49*49 的关键词共现矩阵。

- 转化为 0-1 矩阵：将矩阵中同一个关键词的共现变为 1，矩阵中所有元素的平均值为 1.696，取近似值 2 为阈值。将矩阵中小于 2 的取值为 0，大于等于 2 的取 1，得到一个只有 0 或 1 的二值矩阵。利用 UCINET 软件将二值矩阵结果保存为 UCINET 的数据文件，其文件名为 city.###d 与 city.###d。

- 构建关键词共现网络图：启动 NewDraw>Open>Ucinet dataset>NetWork；再选择相应的数据文件，即可实现关键词共现网络图，并对该图做相应的调整。其详细流程如图 1 所示：

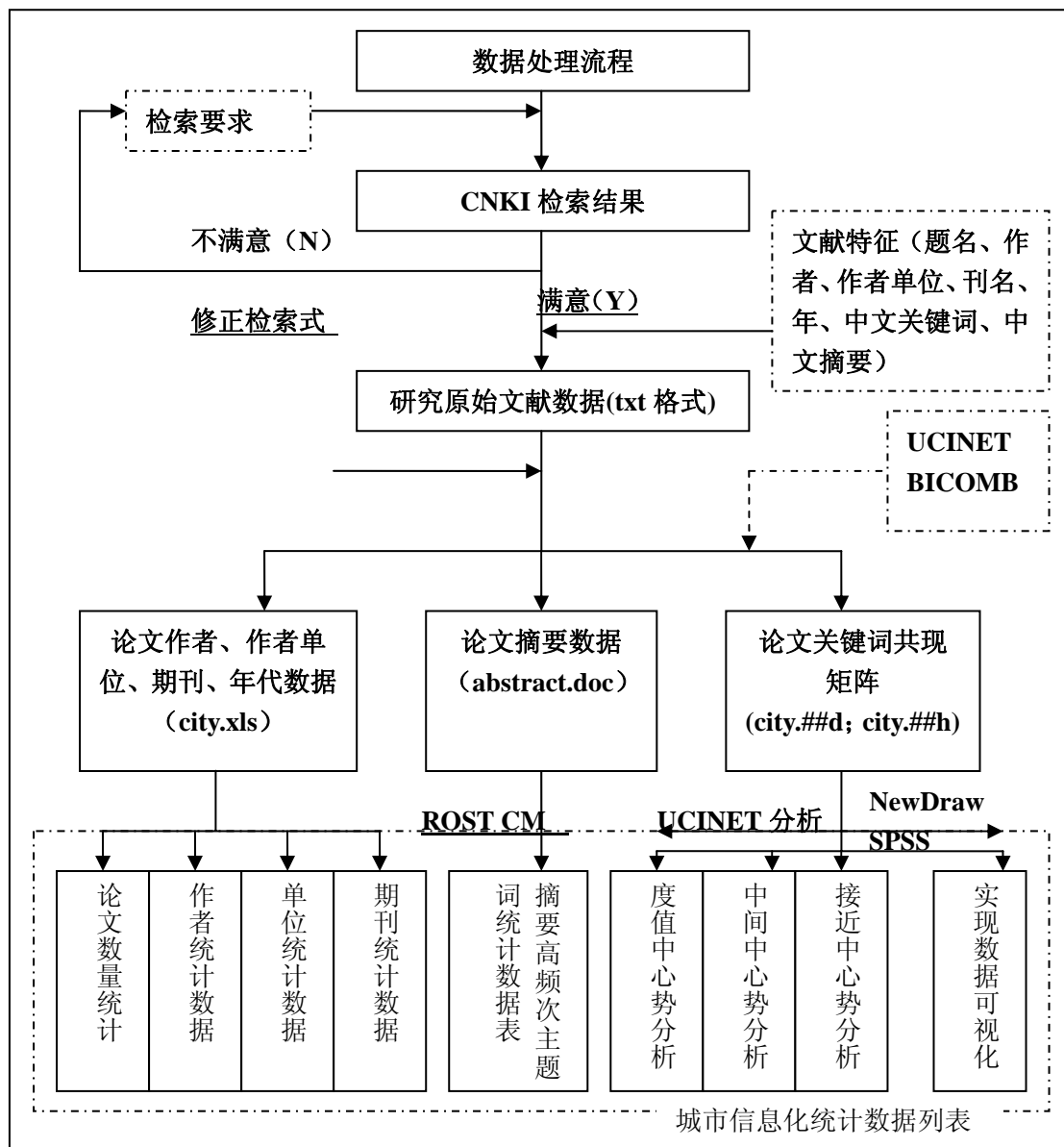


图1 城市信息化数据处理流程

3 城市信息化文献计量分析

3.1 论文年代分析

对城市信息化主题论文年代进行分析, 可以了解其研究历史, 提示其发展趋势。图 2 展示了 1990—2009 年城市信息化论文发表情况如图 2 所示:

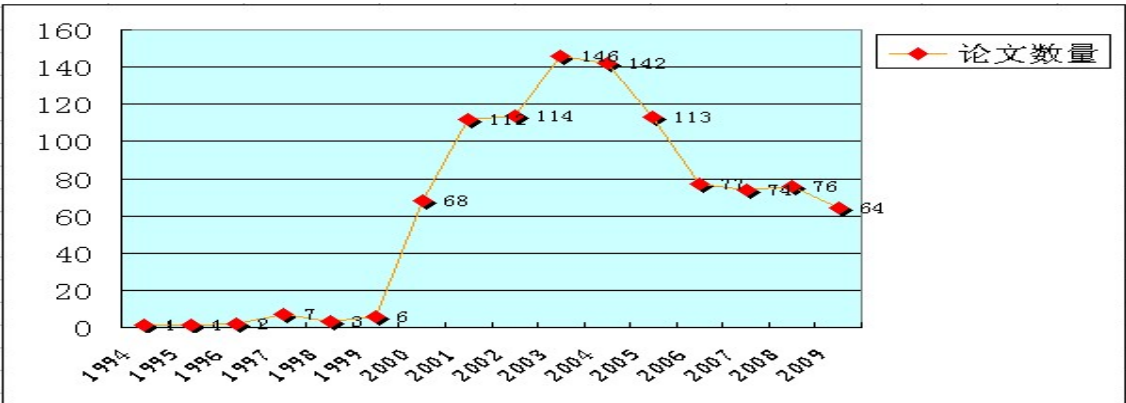


图2 城市信息化论文年化分布

图2可以分三个部分①从1994年到1999年的起始阶段，论文共20篇，仅点问题的1.99%；②2000年到2005年的快速发展阶段，随着全球一体化和信息化趋势的推动，人们对城市信息化的研究越来越热，到2005年末，开始出现大量的研究成果；③2006年到2009年有所减缓阶段，该阶段对城市信息化的研究趋势有所减弱，但仍然高于平均水平。

3.2 论文作者分布

在对论文作者统计中，本研究仅对第一责任人进行统计分析，因为，在论文发表中，署名为第一作者的著者，是该文献的主要贡献者，也是知识产权所有人。据统计，在1 007篇文献中，去掉无署名论文586篇，共有377位著者。表1给出了发表论文在2篇及以上的分布情况，其中，徐峰个人发表论文5篇，是城市信息化领域的高产作者，而发表3篇论文的作者有11位，发表2篇论文的作者有18位，发表1篇论文的作者有347位。

表1 城市信息化论文作者分布图

序号	作者名称	论文数量	序号	作者名称	论文数量
1	徐峰	5	16	张晓楠	2
2	李卓华	3	17	高荆民	2
3	张永乐	3	18	耿军	2
4	章彩云	3	19	郭理桥	2
5	杨雄	3	20	刘景明	2
6	金江军	3	21	贾莉	2
7	徐马陵	3	22	田凤宾	2
8	郑宏剑	3	23	寇有观	2
9	吴伟萍	3	24	马天蔚	2
10	郑大永	3	25	罗建萍	2
11	汪礼俊	3	26	谢世诚	2
12	石德万	3	27	王丹	2
13	杨潇	2	28	任宇子	2
14	谢学宁	2	29	程娟	2
15	盛柏桢	2	30	朱兵	2

3.3 论文期刊分布

据统计，1 007篇期刊论文共分布于427种期刊中，其中核心期刊**篇，表明对城市信息化的研究具有较高的质量。排名第一的《中国信息界》发文47篇，占12.2%；而排名第二、三名的《每周电脑报》、《信息化建设》载文篇数分别为39，31。表2为刊载8篇以上的期刊分布，共载论文385篇，占总论文数的38.2%，而载文只有1篇的期刊共有293种，说明该领域的论文期刊分布面较广。如表2所示：

表2 城市信息化刊载8篇以上的期刊分布

期刊名称	频次	所在比例 (%)
中国信息界	47	12.2
每周电脑报	39	10.1
信息化建设	31	8.1
中国计算机用户	26	6.8
电子政务	23	6
中国电子商务	21	5.4
信息系统工程	20	5.2
计算机周刊	20	5.2
上海信息化	19	4.9
通信世界	17	4.4
互联网周刊	16	4.1
上海微型计算机	15	3.9
软件世界	13	3.4
数码世界	13	3.4
情报杂志	10	2.6
邮电商情	10	2.6
建设科技	10	2.6
智能建筑与城市信息	10	2.6
中国城市经济	9	2.3
中国建设信息	8	2.1
情报科学	8	2.1

3.4 作者单位分布

在 1 006 篇文献中,作者单位(第一作者单位)约为 326 个,表 3 是发表论文数量前 12 的单位。其中,北京大学以 15 篇论文高居榜首,而仅随其后的分别是同济大学和信息产业部,论文数分别为 13、8,表中,不难发现,排名前 12 名的单位中,有 8 所高等院校,说明,我国高等院所在推动城市信息化起着举足轻重的作用。如表 3 所示:

表3 发表论文前12作者单位分布

作者单位 (第一作者)	北京大学	同济大学	信息 产业部	武汉 理工大学	武汉大 学	上海市 信息 化办	华中 科技大 学	华东 师范大 学	北京 工业大 学	中国 科学院	天津大 学	上海 社会科 学院
频次	15	13	8	6	6	6	6	6	6	5	5	5

在这 326 个机构中,有政府部门、科研院所、企业公司、也包括一些其部门。在对城市信息化的研究中,科研院所与政府部门投入的力度最大,分别占 41%, 34%,是该领域的核心研究者。机构类型分布如图 4 所示:

4 城市信息化主题分析

4.1 摘要主题内容分析

对文献摘要进行分析的原因在于作为科技论文的重要组成部分,它是以提供论文内容梗概为目的,不加评论和补充解释,简明、确切地概况论文的主要内容。通过对文献摘要的分析,可以在一定程度上把握该领域的研究热点和发展趋势。笔者首先将检索所得的数据按年代编号,并将摘要数据导入到 Word 文档

中，然后利用 ROST CM 中文词频统计软件自动统计 Word 文档各词词频，并利用过滤功能将一些停用的词去掉，如此我们得到城市信息化论文摘要词频出现高的主题词，将出现频次前 100 的主题词筛选出来，如表 4 所示：

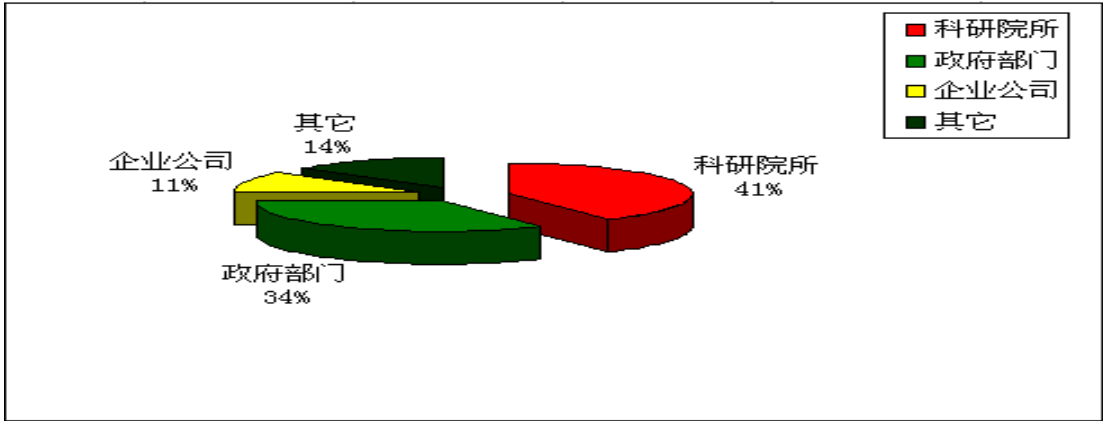


图4 作者单位类型

表4 1990-2009城市信息化学术论文摘要出现频次前75位主题词

序号	主题词	频次	所在比较 (%)	序号	主题词	频次	所在比较 (%)	序号	主题词	频次	所在比较 (%)
1	数字城市	210	15.49	26	智能建筑	12	0.885	51	信息平台	9	0.664
2	信息产业	91	6.715	27	中心城市	12	0.885	52	指标体系	9	0.664
3	电子政务	89	6.568	28	城镇化与信息化	12	0.885	53	信息港	8	0.590
4	信息技术	66	4.870	29	城市空间	12	0.885	54	信息产业发展	8	0.590
5	城市发展	47	3.468	30	区域信息化	12	0.885	55	解决方案	8	0.590
6	无线城市	42	3.099	31	信息化程度	12	0.885	56	领域信息化	7	0.516
7	信息化水平	31	2.287	32	中国信息	12	0.885	57	国家信息化	7	0.516
8	信息资源	29	2.140	33	空间信息	11	0.811	58	社区信息化	7	0.516
9	电子商务	27	1.992	34	城市功能	11	0.811	59	有线电视	7	0.516
10	企业信息化	25	1.845	35	宽带城域网	11	0.811	60	信息安全	7	0.516
11	社会信息化	25	1.845	36	中国信息化	11	0.811	61	经济社会	7	0.516
12	信息基础设施	24	1.771	37	西部地区	10	0.738	62	新型工业化	7	0.516
13	地理信息系统	22	1.623	38	政府信息化	10	0.738	63	可持续发展	7	0.516
14	信息共享	21	1.549	39	发展战略	10	0.738	64	城市群	7	0.516
15	城市现代化	18	1.328	40	信息化战略	10	0.738	65	宽带网	7	0.516
16	基础设施	18	1.328	41	产业结构	10	0.738	66	公共服务	7	0.516
17	信息网络	17	1.254	42	空间数据	10	0.738	67	科学发展观	7	0.516
18	信息服务	17	1.254	43	农村信息化	10	0.738	68	信息资源	7	0.516
19	信息系统	16	1.180	44	农村	10	0.738	69	总体规划	6	0.442
20	数字鸿沟	15	1.107	45	信息素养	10	0.738	70	信息管理	6	0.442
21	信息时代	15	1.107	46	系统工程	10	0.738	71	电子信息产业	6	0.442
22	发达国家	14	1.033	47	经济全球化	10	0.738	72	网络经济	5	0.369
23	电子政府	13	0.959	48	数字地球	10	0.738	73	管理信息系统	5	0.369
24	基础设施建设	13	0.959	49	宽带网络	10	0.738	74	支柱产业	5	0.369
25	关键技术	12	0.885	50	网络技术	9	0.664	75	行业信息化	5	0.369

从表4中可见,近20年内城市信息化持续研究的热点词有:数字城市、信息产业、电子政务、信息技术、无线城市、信息资源建设、电子商务、企业信息化、社会信息化、信息共享、信息服务等词。

4.2 关键词共现社会网络分析

文献关键词能鲜明而直观地表述文献论述或表达的主题,使读者在未看学术论文的文摘和正文之前便能一目了然地知道论文论述的主题,而各关键词之间具有某种内在的联系,某一主题的多篇论文,关键词之间,相互交叉,关系错综复杂。因此,笔者用社会网络分析方法,来开展对关键词的分析,能够更好地把握城市信息化发展的热点和趋势。利用UCINET的绘图功能,将二值矩阵的数据转化为一个关键词共现网络图,下面将从网络图的中心度,以及聚类分析展开探讨。网络图及中心度如图2,表5所示:

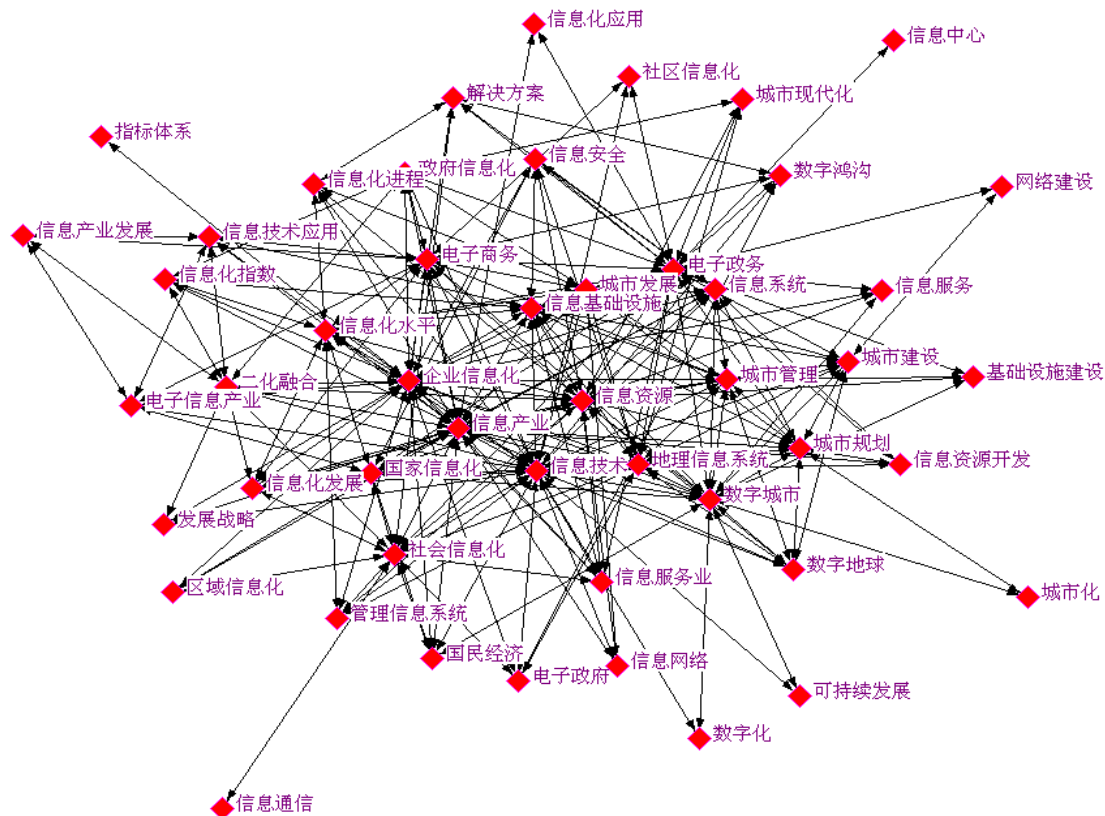


图2 城市信息化调频关键词共现网络

表5 城市信息化领域高频关键词中心度

序号	关键词	频次	度值中心度	中间中心度	接近中心度
1	电子政务	119	320	189.883	59
2	数字城市	110	272	148.823	60
3	信息技术	107	317	192.219	55
4	信息产业	99	276	171.935	53
5	企业信息化	91	276	171.447	54
6	信息资源	66	206	108.885	60
7	电子商务	42	118	82.026	68
8	城市管理	39	128	59.274	70
9	信息系统	38	105	49.732	69
10	城市发展	37	97	43.936	71
11	地理信息系统	35	108	65.992	70
12	国家信息化	33	98	60.574	68
13	城市规划	32	114	46.37	73

14	信息产业部	31	99	84.125	66
15	社会信息化	30	86	41.355	75
16	城市建设	28	55	18.153	83
17	信息化发展	27	82	43.145	68
18	信息化水平	26	87	53.095	67
19	数字鸿沟	26	56	46.197	80
20	信息基础设施	26	93	42.838	69
21	二化融合	24	69	57.047	71
22	城市化	22	43	31.259	79
23	电子政府	19	47	32.202	78
24	信息服务业	18	61	30.995	77
25	政府信息化	17	61	50.695	70
26	信息服务	16	38	15.741	81
27	管理信息系统	14	44	18.56	78
28	信息技术应用	14	42	27.016	76
29	信息化指数	14	45	37.656	79
30	信息化进程	14	42	25.173	80
31	数字化	13	30	2.755	84
32	区域信息化	13	38	28.058	79
33	发展战略	13	34	34.201	82
34	信息安全	13	42	15.275	79
35	电子信息产业	12	40	28.291	78
36	信息通信	12	21	8.23	91
37	信息中心	12	27	13.302	86
38	网络建设	12	28	13.361	84
39	社区信息化	12	25	8.745	86
40	信息网络	11	31	8.595	83
41	国民经济	11	35	12.516	81
42	信息产业发展	11	27	18.978	86
43	指标体系	11	21	5.657	87
44	可持续发展	11	22	7.433	87
45	基础设施建设	11	27	10.99	85
46	信息资源开发	10	36	21.7	81
47	数字地球	10	38	4.755	87
48	信息化应用	10	32	27.907	77
49	城市现代化	10	33	19.227	78

4.2.1 中心度与中心势分析

“中心性”是社会网络分析中的重点之一，反映出个人或组织在其社会网络中具有怎么样的权力，或居于什么样的中心地位^[6]。比较常用的中心性指标包括度值中心度（degree centrality）、中间中心度（betweenness centrality）、接近中心度（closeness centrality）。下面对城市信息化高频关键词社会网络中心性展开详细分析：

• 度值中心度是用来衡量一个网络中某个节点中心性的测量,可用该节点与网络图中直接连接的节点个数来度量。本文研究中,度值中心度反映了关键词之间的共现关系,关键词与其它关键词共现的频率越高,该值就越大,说明这些词是城市信息化领域内的研究热点并处于核心地位,备受研究者青睐。所以“电子政务”、“信息技术”、“信息产业”、“企业信息化”、“数字城市”、“信息资源”、“城市管理”、“电子商务”、“城市规划”、“地理信息系统”、“信息系统”等度值中心度最高,说明它们是城市信息化领域的研究热点。

• 中间中心度是衡量网络中节点对资源控制能力的测量,一个节点处于其他节点对最短路径的条数最多,其中间中心度越高,起到的“中介”就越重要,对资源的控制能力也就越强。在本文中,中间中心度反映的是某关键词对其它关键词出现在同一篇文章中影响能力的大小,如表 5 中,“信息技术”、“电子政务”、“信息产业”、“企业信息化”中间中心度最高,说明它们对关键词的共现影响最大。

• 接近中心度衡量网络中节点不受其他节点控制的测量,它反映的是某节点与其他所有节点的亲近程度。在本文中,反映的是某关键词与其他关键词共现可能性的大小。新近中心度越大,说明越不容易与其他关键词同时出现在某一篇文章中。如“信息通信”的接近中心度最高,说明它与其他关键词共现的几率最小。而“信息产业”、“企业信息化”、“电子政务”的接近中心度较小,则与其他关键词共现的几率最高。

• 三指标相互关系,通过分析,我们可以得知,接近中心度由度值中心度与中间中心度共同影响。将表 5 中三个中心度标准化后,用 SPSS 进行曲线拟合制得图 3,可见,三个中心度都反映出其核心地位,如“信息基础设施”,“二化融合”,当度值中心度与中间中心度都较高时,接近中心度反而较低,相反,当度值中心度与中间中心度都较低时,而接近中心度却较高,如图 3 所示:

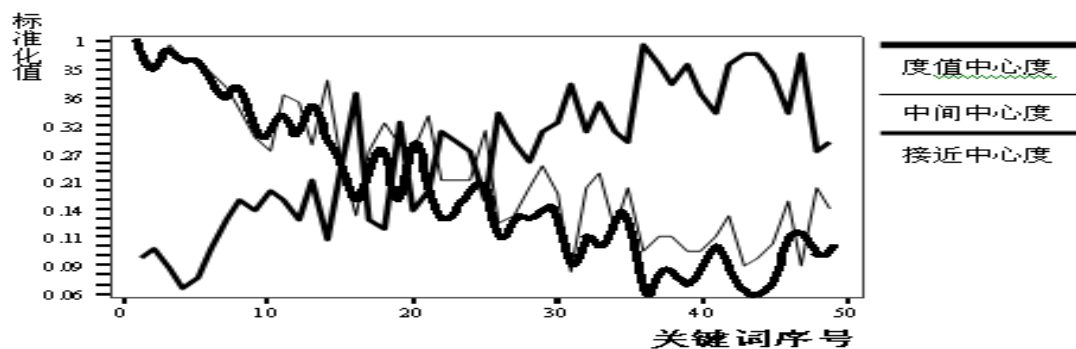


图3 三个中心指标折线

4.2.2 聚类分析

用 SPSS 13.0 软件对高频关键词进行聚类分析,得到关键词聚类图,如图 4 所示:

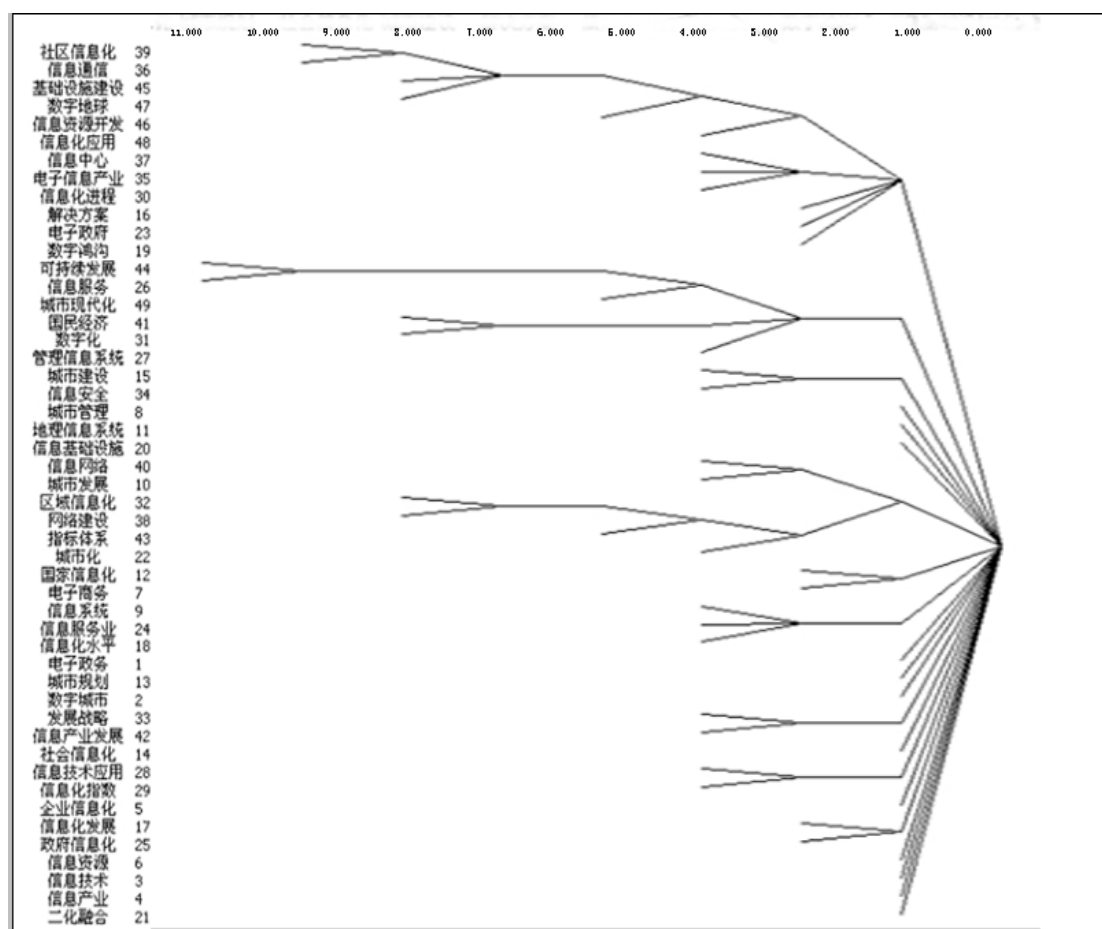


图4 高频关键词聚类

图4反映了城市信息化领域研究内容间错综复杂的关系以及各研究主题的等级关系。如，“电子政务”、“城市规划”、“信息化水平”以及“信息中心”、“电子信息产业”、“信息化进程”分别被聚在同一类，说明它们的关系紧密，研究成果也比较多。

5 结 语

经过近二十年的发展，我国城市信息化研究不断升温，越来越受到重视，地位不断提升。通过对城市信息化学术论文作者、作者单位、引文期刊、关键词以及摘要这几方面的分析，笔者发现：

- 城市信息化在近20年中发展较快，特别是2000-2005这一期间，随着党的十六大提出了我国信息化的建设方针，大力推动了城市信息化的发展，为此涌现了大量的研究者。从作者的情况来看，徐峰、李卓华、张永乐等是该领域的高产作者。从研究期刊来看，大多是来自科研院所与政府部门，总共占到了75%。通过分析，该领域的核心研究机构有北京大学、同济大学、信息产业部，以及武汉理工大学、武汉大学等单位，他们的研究成果丰富和推动了城市信息化的发展。

- 研究热点：通过分析，我们发现，对城市信息化的研究，形成了以电子政务、电子商务、企业信息化、社会信息化、数字城市、二化融合、地理信息系统、信息技术等词为核心的研究，并不断向信息资源建设、缩短数字鸿沟、建设统一标准方向发展。信息指标体系，信息化水平评价也是城市信息化研究的热点。随着研究的不断深入，城市信息化的研究理论越来越丰富，也推动着我国信息化进程。

- 研究薄弱点：上述研究表明，我国城市信息化的研究存在的主要问题：①基础研究较多，深入研究少，研究深度不够。大多数研究集中在电子政务建设、企业信息化、社会领域信息化以及电子商务。对

城市信息化程度评价的研究成果较少。②研究体系松散,城市信息化各环节连接不紧密,没有形成一个紧密的链条。③众多研究者,都在研究制定城市信息化规划,而对于如何制定城市信息化规划、规划制定过程中应当注意的问题、一般流程是什么的研究成果还较少,这些方面值得研究。

• 发展趋势:鉴于以上研究,笔者认为:城市信息化未来发展,数字城市是基础,无线城市是目标,即在实现数字城市的基础之上,近而实现无线城市。

参考文献:

- [1] 邱均平. 信息计量学. 武汉:武汉大学出版社. 2007:1-20.
- [2] 马费成. 国内知识管理研究结构探讨——以共词分析为方法. 情报学报, 2008(2):93-101.
- [3] 邱均平, 邹菲. 关于内容分析法的研究. 中国图书馆学报. 2004(2): 12- 17.
- [4] 朱庆华, 李亮. 社会网络分析法及其在情报学中的应用. 情报理论与实践, 2008 (2): 179-183.
- [5] 魏瑞斌. 社会网络分析在关键词网络分析中的实证研究. 情报杂志, 2009, 28 (9):46-49.
- [6] 刘军. 社会网络分析. 北京: 社会科学文献出版社, 2004:114, 121-122.

[作者简介]

李万辉, 男, 1984 年生, 硕士, 研究生, 已发表论文 3 篇

张文德, 男, 1962 年生, 研究馆员, 教授, 馆长, 博士后, 博士生导师, 发表论文 100 余篇。

陈振标, 男, 1982 年生, 助理馆员, 发表论文 9 篇。