

doi:10.3772/j.issn.2095-915x.2015.05.006

基于社会网络分析的部委专业情报机构 链接关系分析

郭燕燕, 杨朝峰

(中国科学技术信息研究所 北京 100038)

摘要: 采用社会网络分析与链接分析的方法, 从万方机构数据库中检索出部委级专业情报机构, 剔除无官网机构, 统计这些网站的网页数量、出站链接数量、反向链接数量、PR 值, 并采用 Gephi 软件对情报机构网站间的链接关系进行可视化展示, 进而对情报机构链接网络结构进行分析。发现中国科学技术信息研究所、机械工业信息研究所、中科院文献情报中心等五所情报机构网站构成的链接网络网络密度、聚类系数比较大, 但是网群中存在大量孤立节点群以及链接子网络, 网络的整体密度与聚类系数比较小。因此, 在专业情报机构间有必要加强交流协作, 进而促进科技信息资源的共享, 加快科研成果的增值利用。

关键词: 社会网络分析, 链接分析, Gephi, 情报机构

中图分类号: G351

作者简介: 郭燕燕 (1990-), 中国科学技术信息研究所情报学硕士研究生; 研究方向: 科技战略与科技政策, E-mail: guoycbx2010@163.com; 杨朝峰 (1975-), 男, 中国科学技术信息研究所副研究员, 同济大学技术经济及管理专业博士, 研究方向: 科技政策、宏观经济政策。

The Link Relations Analysis among Professional Intelligence Agencies of all Ministries Based on Social Network Analysis

GUO Yanyan, YANG Chaofeng

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: We selected data from the institution database of wanfang, and then removed the professional intelligence agencies having no official website. Based on social network analysis and link analysis, we counted and analysed the number of webpage, outbound links, inbound links and the value of PR. With the Gephi, we visualized and analysed the link relations and the structure of the link network. We found that the network had high density and clustering coefficient, which constituted of five institution such as ISTIC and Machinery industry information institute and Chinese Academy of Science etc. But the whole network's density and clustering coefficient are low. As a result, to accelerate the sharing of science resources and the utilize of science achievements, it is necessary to strengthen communication and cooperation among information institutions.

Keywords: social network analysis, link analysis, Gephi, information Agencies

1 研究背景与意义

在日本所处的经济高速发展时期,日本科学技术会议就提出构筑“科学技术情报全国流通体系”(NIST)的构想。其基本内容是,以独立的信息服务机构为基础,国家从中加以调整,使各机构间分担协作并有机结合起来,实现科技信息在全国范围内的流通^[1]。美国情报机构管理部门也非常强调各情报部门间的配合,并开展竞争,一方面获取相互关联的情报,另一方面防止情报的虚假性,增强情报的可靠性^[2]。作为情报强国的美国与日本都强调情报机构间的沟通与协调工作,正是这种工作模式使得其情报机构在促进经济发展与保卫国家安全等多方面都发挥了极其重要的作用。作为一个正在实施科技创新驱动的发展大国,科技情报机构在情报搜集与情报分析方面理应首当其冲。而中国的科技情报机构现状如何呢?由于网络的快速发展,情报机构大多有自

己的网站,且网站的链接关系反映了机构之间的交流与合作关系,因此,我们希望通过分析情报机构网站的规模以及链接关系了解中国情报机构的现状。

本文主要采用链接分析法以及社会网络分析的方法,从机构网站链接角度分析机构间关系。链接在很大程度上可以看作作为一种推荐或认同^[3],它主要通过网络站点间的链接关系对网站自身信息组织、网站的外部互动性、开放性以及网站的综合影响力进行间接评价。链接分析与社会网络分析相结合,对网站之间构成的链接网络进行网络密度和度中心度的分析,有助于挖掘链接网络中的核心网站。链接分析和社会网络分析方法能够很好地揭示问题,被广泛应用,如2011年邱均平教授采用链接分析和社会网络分析的方法,对科学网博客基于好友链接的知识交流状况进行分析^[4];2012年门伟莉、邓尚民主要采用社会网络的分析方法对农业网

群影响力进行研究^[5]；吴凤慧等在2013年对链接分析研究进展进行综述^[6]；赵宇翔、彭希羨、孙建军采用链接分析方法以及灰色关联方法对非政府组织网站的综合影响力进行评价^[7]。王新才、丁家友在2013年采用复杂网络，对政府信息资源进行研究^[8]。

情报机构是我国主要的信息服务机构，尤其是部委专业情报机构（以下简称“情报机构”），他们直接为国家战略决策提供支持。在当前创新驱动发展战略的推动与实施下，科技情报机构的地位和作用更加突出，他们担负着促进科技信息界研究资源的共享、加快科研成果的增值利用的重任。因此，就必须加强对我国情报机构的现状研究，保证其健康发展，促进其重要作用发挥。另一方面，由于部委专业的情报机构有其特定的隶属部委政府机构，且分布在不同的行业，难以直观地了解科研情报机构的关系以及比较核心的科研情报机构，因此对我国情报机构的研究就更为迫切。采用链接分析和社会网络分析的方法对情报机构网站群的分析暂时还没有。因此，本文采用链接分析和社会网络分析相结合的方法，试图揭示部委专业情报机构网站的网络结构和链接关系，并将其可视化展示，发现情报机构网站间的问题，进而提出借鉴性的意见。

2 数据收集

部委专业情报（信息）机构直接参加国家战略决策，其网站建设相对比较完善，而部委属科研院所情报机构以及地方的情报机构大多隶属政府机构的一个职能部门，没有官方网站。同时考虑到专业情报机构之间的链接关系数量相对比较少，且可能存在多个专业情报机构共同链接一个或多个非专业情报机构的情况，为展示更全面的

链接关系，本文以专业情报机构为中心，将链接到部委属情报机构的网站也考虑在内。

在万方机构数据库中检索出部委专业情报机构，共47所，剔除无官网的部委专业情报机构得到有效网站36个；然后通过站长工具分别查询36所部委专业情报机构的网页总数、出站链接数、反向链接数、PR值。数据采集时间为2015年6月27日。

3 数据分析

Gephi是一款基于JVM的开源复杂网络分析软件，其主要用于各种网络和复杂系统，动态和分层图的交互可视化，其功能强大、可视化效果好且操作相对简单。因此，本次网站间的链接网络采用Gephi工具进行可视化处理和分析。

3.1 网页总数及PR值分析

网站的网页总数能反映网站的规模，网页数越多，在一定程度上就说明网站的规模越大。出站链接是指自身网站链接到其他网站的数量，反链数指其他网站链接到自身网站的链接数，反链数越大，说明网站的影响力越大。PR值全称为PageRank（网页级别），PR值越高说明该网页越受欢迎（越重要）。

由于篇幅有限，表1按照网页总数排名前10位的网站列出网站的出站链接数量、反链数量和PR值。

从表中数据可以发现，中国地质图书馆的网页数量最多，其网站规模最大，其次是中国物流信息中心、中国标准化研究院国家标准馆、中国科学技术信息研究所。水利部水利信息中心出站链接数量明显比其他情报机构网站多，一定程度上说明这两家机构与其他网站的交流合作比较多。从反链数量角度来看，水利部水利信息中心的反

表1 情报机构网站的基本情况

机构名称	网页总数	出站链接	反向链接	PR 值
中国地质图书馆	145000	63	8	6
中国物流信息中心	106000	16	3	4
中国标准化研究院国家标准馆	42600	40	13	6
中国科学技术信息研究所	42200	30	6	7
中国国家图书馆	37000	11	2	8
机械工业部科技信息院	32700	22	1	6
水利部水利信息中心	22700	70	23	6
中船综合技术经济研究所	19300	4	3	6
煤炭信息研究院	16300	13	3	5
国家知识产权局中国专利信息中心	13500	3	2	7

链数最多，有更多的情报机构通过链接指向他，说明其在本行业内具有较高的权威。

3.2 情报机构网站链接网络结构分析

机构网站间的链接关系是一种有向关系，链接关系网络是一种有向关系图。用 Gephi 构建链接关系网络图时，数据表格需要有 Source、Target 两列数据，如果 A 网站是 Source，B 网站是 Target，即表示 A 网站链接了 B 网站。将源网站放在 Source 列下，目标网站放在 Target 列下，将数据表格导入 Gephi 软件，构造网站链接关系网络。网络中节点表示情报机构的网站，两点间的连线表示链接关系，箭头指向表示链接方向。本文将从网络结构、网络中心性、网络的关联性(网络的平均路径长度和平均聚类系数)三个角度对比分析情报机构网站建设情况，旨在发现目前情报机构网站建设中存在的不足，从宏观上指导情报机构网站建设。

3.2.1 情报机构链接网络结构初步分析

采用 Gephi 可视化软件的模块化功能对构造的情报机构链接网络进行模块化处理。模块化

(modularity) 的概念最早由 Newman 和 Girvan 在 2004 年提出。模块化处理就是将整个网络划分为不同的子网络或者说是社区模块，模块化处理要求，同一模块内的节点连接更紧密，具有更大的密度，同时，模块化分类要求具有更好的稳定性，其分类的质量用 Q 表示。^[9]

Gephi 进行模块化处理时采用 Fast Unfolding 算法，通过模块化处理后，整个网络的 Q 值最大，模块化聚类的质量最好，且各个子模块间的联系最弱，子模块内的联系更紧密，密度更大。

部分情报机构与本行业内网站有链接关系，但与情报机构间没有链接关系，因此产生大量独立的节点，通过模块化处理后，呈现出情报机构网站链接网络图，如图 1 所示。同一个颜色代表同一类，节点的大小代表与该网站链接的情报机构网站的数量，链接数量越多，节点越大。为显示效果更加清晰，图中只显示度在 2 以上(包含度为 2)的节点标签，模块化后的结果将链接网络分为 31 个子模块。但是，图中存在大量的孤立节点以及两个单位之间的一个链接(如模块 2、模块 3、模块 6)，为展示更加清楚，进一步将链

基于社会网络分析的部委专业情报机构链接关系分析

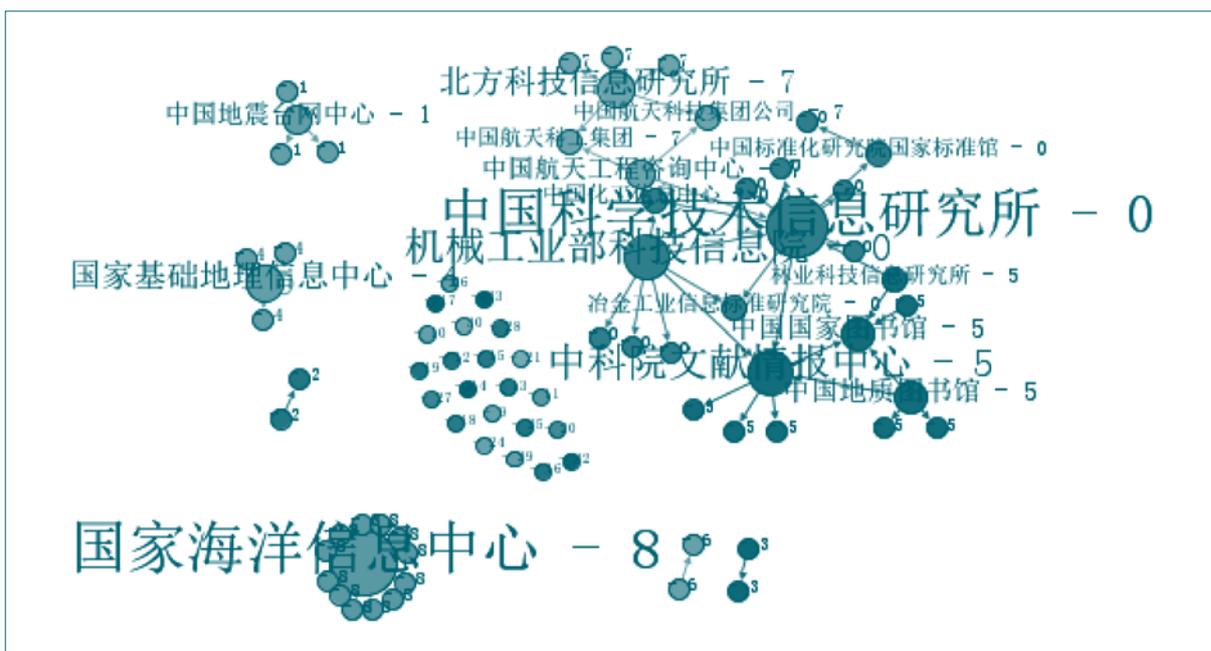


图 1 情报机构网站链接网络图

表 2 模块化分割的八个类别

模块化分类	网络描述	所占百分比
模块 0	以中国科学技术信息研究所、机械工业信息研究所为中心相链接构成的子网络。	16.46%
模块 1	以国家地震局地震数据信息中心（中国地震台网中心）为中心相链接构成的子网络。	5.06%
模块 2、 模块 3、 模块 6	中国核科技信息与经济研究院与核工业计算机应用研究所链接的子网络；国家食品药品监督管理局信息中心与国家医药监督管理局南方医药经济研究所链接的子网络；军事医学科学院卫生勤务和医学情报研究所与国家生物医学分析中心链接的子网络。	7.59%
模块 4	以国家基础地理信息中心为中心相链接构成的子网络。	5.06%
模块 5	以中科院文献情报中心、中国国家图书馆、中国地质图书馆、林业科技信息研究所为中心相链接构成的子网络。	12.66%
模块 7	以北方科技信息研究所、中国航天工程咨询中心中国航天科工集团为中心相链接构成的子网络。	8.86%
模块 8	以国家海洋信息中心为中心相链接构成的子网络	16.46%
模块 9	没有链接关系的孤立的节点群。	27.85%

接网络分为八大类，如表 2 所示。

密度用来描述整个网络联系的紧密度，是整体网络结构分析的重要指标之一。图中节点间的连线越多，说明该网络的密度越大，网络成员间的关系越密切。通过密度计算，可以分析网络节点间的关系是否密切，密度的取值范围为 [0 ~ 1]。

该情报机构网站链接网络的网络密度为 0.009，除去无链接关系的孤立节点后的网络密度为 0.018。

管理学派强调组织内部的资源能力是竞争优势的主要来源^[10]。网络密度反映了网络中联结的疏密情况以及网群中网站对资源（特别是稀缺性的关键资源）的可获得性及质量。在网群中，网

络密度高的网群内网站利用自身的社会资本，与相关资源拥有者建立关系，可以具有优先获得关键资源的可能^[11]。

从图1可以看到情报机构网站整理链接网络密度较低，存在大量无链接关系的孤立点，且模块1、模块2、模块3、模块4、模块6、模块8子网络间没有链接，同时与其他类子网络也没有链接关系，因此，情报机构之间的链接关系整体比较松散。

3.2.2 情报机构链接网络的度分析

一个网群中，各节点间的联系程度不一样，在网络中发挥的作用也不相同。中心度主要用来分析网络中各节点所处的位置。其中，点度中心度是测量网络中节点重要程度的关键指标之一。网络图中某节点的绝对点度是指与该点直接相连的其他点的数目。与该点直接相连的点的数量越大，该点的绝对点度中心度就越大，该点在网络中就处于中心位置^[12]。

出度是指指向其他节点的边的数量，在情报机构链接网络中，一个节点的出度代表该节点所代表的网站链接其他网站的数量，也即该网站的出站链接数量。入度是指从其他节点指向自身的边的数量，在情报机构链接网络中，一个节点的入度代表其他情报机构网站链接该节点代表网站的链接数，也即一个网站被链接的链接数量。在情报机构链接网络中，节点的入度越大，在一定程度上说明该机构拥有较高的权威，越值得借鉴学习。在 Gephi 软件中分别以度和入度来排序，其结果分别如图2、图3所示。

从度排序情况可知与中国科学技术信息研究所、机械工业部科技信息院、中科院文献情报中心、国家海洋信息中心相链接的情报机构网站比较多，但是从入度情况来看，中国国家图书馆、中国文献情报中心、中国科学技术信息研究所的网站入链相对其他都比较大，说明这几家情报机构的综合实力比较强。

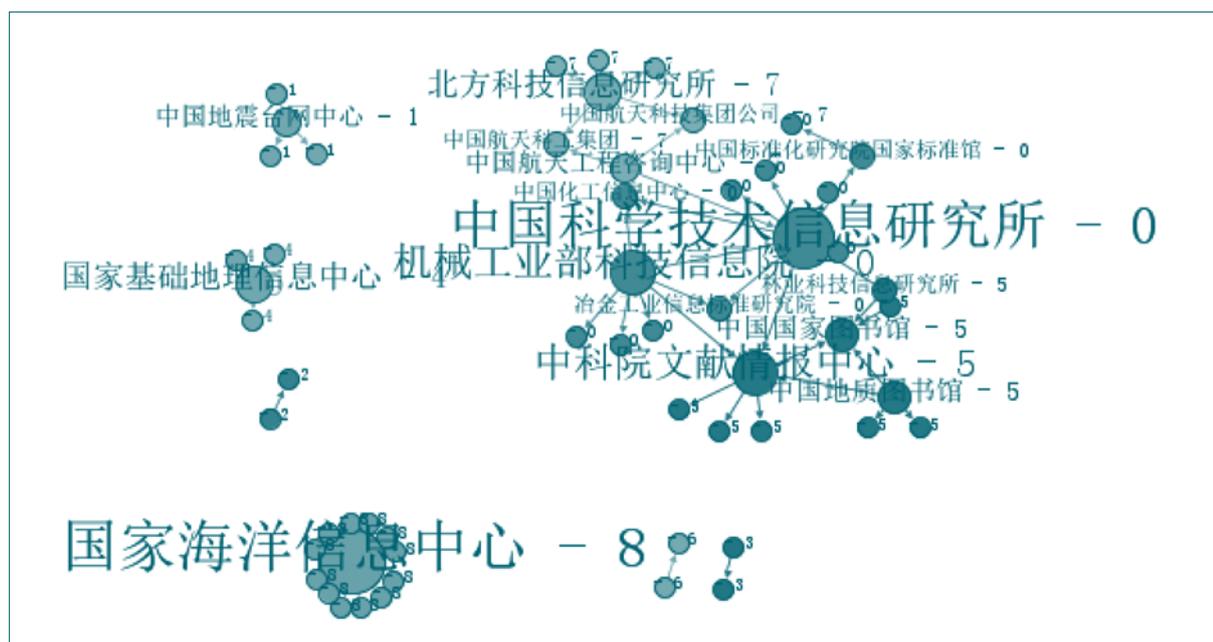


图2 以度排序生成的网络图



图3 以入度排序生成的网络图

3.2.3 情报机构网站链接网络关联性分析

网络的关联性分析采用平均路径长度和平均聚类系数两项指标^[5]。平均路径长度是指网络中所有节点对之间的平均最短距离，这里节点间的距离指的是从一个节点要经历的边的最小数目，平均路径长度和直径衡量的是网络的传输性能与

效率。聚集系数是表示一个图形中节点聚集程度的系数。平均路径长度越短，聚类系数越大，网络的传输效率越高，网络的关联度就越大。

情报机构网站链接网络整体比较松散，在此对比三组子网络的平均路径长度和平均聚类系数。如表3所示。

表3 部分子网络的平均路径长度与平均聚类系数

类别	模块 0、5 组成的子网络	模块 7、0、5 组成的子网络	模块 1、4、7、0、5 组成的子网络	除去孤立节点群构成的网络	整个网络
平均路径长度	1.614	1.679	1.631	1.535	1.55
平均聚类系数	0.057	0.044	0.035	0.023	0.017

从表3可以看出，由中国科学技术信息研究所、冶金工业信息标准研究院、中科院文献情报中心、中国国家图书馆为中心构成的网络（模块0、5）的平均聚类系数最大。由中国科学技术信息研究所、冶金工业信息标准研究院、中科院文献情报中心、中国国家图书馆、中国航天科工集团为中心构成的网络（模块7、0、5）的聚类系数次之。

4 情报机构网站链接网络综合分析

通过对情报机构网站链接网络整体的分析以及各子网络的对比分析，揭示了整体情报机构网络和子网络的结构特点，挖掘出了“权威”情报机构；另一方面，通过对情报机构网站间链接关系的可视化展示，发现情报机构间交流合作存在

的问题。

4.1 网络整体关系松散

从表2可看出与情报机构网站没有链接关系的孤立节点占到总结点的27.85%，且模块1、模块2、模块3、模块4、模块6、模块8子网络间没有链接，同时与其他类子网络也没有链接关系，这些节点代表的情报机构与其他情报机构关系疏远，信息交流较少。我们从表1也可以看出，中国物流信息中心、水利部水利信息中心等几个情报机构网站的规模以及网页排名都比较靠前，但在与情报机构的交流较少。比如，煤炭工业采掘机械化信息中心、水利部水利信息中心与其他情报机构没有链接关系。以中国地震台网信息中心为中心的子网络、以国家基础地理信息中心为中心的子网络、以国家海洋信息中心为中心的子网络与其他子网络之间没有链接关系，在情报交流方面形成小团体。整体情报机构网站链接网络的网络密度仅为0.009，平均聚类系数仅为0.017，机构间获取信息资源的效率以及科技信息的共享效率都极低，阻碍了国家科研成果的增值利用，降低了国家情报质量以及全面性。

4.2 网络局部关联紧密

从表3、图2以及图3可以发现中国科学技术信息研究所、机械工业信息研究所、中科院文献情报中心、冶金工业信息标准研究院、中国化工信息中心五所情报机构网站构成的链接网络网络密度、聚类系数都比较大，节点的度以及入度在整个网络中明显比其他节点大，同时，有其他几个部委属科研院所与之相链接，这说明这几所部委专业级情报机构在整个情报机构群中综合实力比较强，影响力大，具有一定的权威地位。该

现象与国家科技图书文献中心(NSTL)^[13]有一定关系，他联合了拥有理工农医科技资源的多所情报机构，这些情报机构之间交流相对比较频繁。

5 结论

本篇文章从万方机构数据库中选择47所部委级专业情报机构，经过筛除无官网的机构得到36所部委级专业情报机构的网站。对各个机构网站的总网页数、PR值、出站链接、反向链接进行分析，并采用Gephi软件将网站之间的链接网络进行可视化展示，并进一步对网络结构特征进行分析。发现中国科学技术信息研究所、中科院文献情报中心、机械工业信息研究所、冶金工业信息标准研究院在情报机构间沟通交流比较频繁，且占有权威地位；同时也发现情报机构网站整体链接网络关联松散。

当然，本文主要考虑的是情报机构之间的链接关系，作为不同部委专业的情报机构，从属于不同的行业，他们的链接关系可能主要集中在自己本行业内，另一方面，没有链接关系的部分机构间可能存在竞争关系，为避免有利信息泄露而减少交流。作为军工背景的国防科技信息中心链接数量也比较少，是考虑到军事安全问题。但从美国、日本情报机构的成功经验来看，我国不同行业、不同部委的专业情报机构，如能积极构建自己的官方网站，将会方便用户快速找到本行业的资讯，加快科技信息的流通。不同部委专业情报机构如果通过网络友情链接加强合作与交流，将对本行业以及行业外的用户使用最合适的数据资源提供极大的帮助。另一方面，增加向其他权威情报机构的链接，能够加快科技信息资源的共享，更有效地实现科技研究成果的增值利用。

参考文献

- [1] 乌云其其格,张新民.面向产业与创新的日本科技信息机构[J].中国信息导报,2006(4):42.
- [2] 刘天宇,李泽普.美国情报机构的变革及启示[J].法制与社会,2013(20):149.
- [3] 廖小琴,刘虹,孙建军.链接网络与核心节点评价指标研究综述[J].情报杂志,2012,31(5):166-171.
- [4] 邱均平,王菲菲.基于博客社区好友链接的知识交流状况分析——以科学网为例[J].图书情报知识,2011(6):25.
- [5] 门伟莉,邓尚民.基于社会网络分析的农业网群影响力研究[J].情报杂志,2012,6(31):59-60.
- [6] 吴夙慧,孙建军,郑彦宁等.链接分析研究进展[J].情报科学,2013,31(4):155-156.
- [7] 赵宇翔,彭希羨,孙建军.链接视角下国内NGO网站综合影响力评价研究[J].情报学报,2014,33(5):549-550.
- [8] 王新才,丁家友.复杂网络视角的国外政府信息资源研究[J].图书情报杂志,2013,57(21),119-126.
- [9] BLONDEL VD,GUILLAUMINE J,LAMBIOTTE R, et al. Fast unfolding of communities in large networks[EB/OL] [2015-4-27]. <http://iopscience.iop.org/1742-5468/2008/10/P10008>.
- [10] 季莹,于光,王铁成.中国作者在(Nature)杂志上发表论文的共词图谱分析[J].情报杂志,2011(5):28-32.
- [11] 门伟莉,邓尚民.组织危机决策影响因素与决策有效性实证研究——基于山东省企业组织危机决策的反思[J].情报杂志,2010,29(5):102-106.
- [12] 约翰·斯科特,著.《社会网络分析法》[M].重庆:重庆大学出版社,2007:200.
- [13] 张晓琳,刘细文,孙坦,等.国家科技图书文献中心的效用形式及其评价[J].图书情报工作,2008,52(3):62.