

基于熵权TOPSIS的211院校网站影响力分析

陈超群^{1,2}, 邓三鸿^{1,2}, 王昊^{1,2}

(1.南京大学 信息管理学院, 江苏 南京 210023; 2.江苏省数据工程和知识服务重点实验室, 江苏 南京 210023)

摘要:【目的/意义】探讨利用大学网站间的引用结构网络来研究大学网站综合影响力的可行性,同时避免社会网络分析中使用单一中心度来评估节点影响力的片面性。【方法/过程】以112所211大学作为研究对象,通过英语版必应搜索引擎收集高校网站间的互链数据,然后运用传统社会网络分析中的中心性理论分析各高校在网站引用网络中的影响力,最后运用熵权TOPSIS方法得到了高校网站影响力的综合排名。【结果/结论】结果表明熵权TOPSIS作为一种多属性决策方法,在研究节点影响力方面可以发挥重要作用,同时从211高校网站综合影响力的排名结果可以看出其与大学科研实力并不是完全相关的,高校在提高科研实力的同时应该加强网站建设。

关键词: 网络计量; 社会网络分析; 熵权TOPSIS

中图分类号: G250.7 **文献标识码:** A **文章编号:** 1007-7634(2017)09-146-07

Analysis on the Influence of Project 211 Universities Websites Based on Entropy TOPSIS

CHEN Chao-qun^{1,2}, DENG San-hong^{1,2}, WANG Hao^{1,2}

1 (1.School of Information Management, Nanjing University, Nanjing 210023, China;

2.Jiangsu Key Lab of Data Engineering and Knowledge Service, Nanjing 210023, China)

Abstract: 【Purpose/significance】This paper explores the feasibility of using the reference structure between the university websites to study the comprehensive influence of the university website, and avoids the one-sidedness of using a single center degree to evaluate the influence of the node in the social network analysis. 【Method/process】The article takes the Project 211 universities as the object of study, through the English version of Bing search engine to collect the interlinked data between university websites, uses the Centrality theory of traditional social network analysis to analyze the websites' influence in the network. Then the entropy TOPSIS method is used to measure the influence of the website of 211-project-universities with multiple centrality as the metric. 【Result/conclusion】The results show that entropy TOPSIS can play an important role in studying the influence of nodes. At the same time, we can see that the website influence and university research capacity are not completely related from the universities websites' comprehensive influence ranking results, and universities should strengthen the website construction while improving their scientific research capacity.

Keywords: webometrics; social network analysis; entropy TOPSIS

1 引言

20世纪90年代中期,随着计算机网络技术的迅速发展,信息资源越来越呈现出数字化、网络化的特征,网络信息资源的交流迅速激增。原有的信息计量方法已不能够对电子

和网络信息进行测度和计量,这就促成了网络计量学(Webometrics)的诞生^[1]。

由于科学文献的引文与网络链接之间的相似性使得研究者开始将文献计量学中的引文分析法应用于网络计量研究中,由此产生了网络计量学的重要研究方法——链接分析法。1996年,美国爱荷华州立大学图书馆的理论馆员 Gerry

收稿日期:2017-06-06

基金项目:国家社科基金项目(12BTQ037)

作者简介:陈超群(1994-),女,湖南常德人,研究生,主要从事网络计量与数据挖掘研究。

McKiernan根据文献计量学中引文“Citation”的含义,首次提出了“Sitiation”(网络引文)的概念,来描述网站之间相互链接的行为。1997年,Ronald Rousseau发表了论文,“Sitiation”一词首次正式出现在文献题名当中。此后,有不少学者进行了网络链接的理论以及应用研究,目前,网络链接分析已经成为了网络计量学的研究热点与重点^[2],在分析政府网站^[3]、新闻网站^[4]、商业网站^[5]以及期刊网站^[6]等链接结构中发挥了重要作用,而分析高校网站的链接结构这一研究方向尤为突出。

随着信息技术、网络技术、计算机技术的快速发展,万维网在不断提高学术信息交流、加速学术合作以及促进国际学术合作网络的形成上发挥了重要作用。大学和专业院校不仅是人才培养的机构,而且也是重要的科研和学术交流组织。当今,几乎所有的世界大学都建立了自己的网站,成为介绍和宣传大学、树立大学形象、传播学术思想和开展学术交流的工具。在过去十几年间,已经有不少学者在研究大学的国际学术网络结构,学者们通过分析大学网页的可见性来形成世界大学之间的超链接网络结构,这已经被证明是衡量一个大学的学术水平、科研能力和教育能力的重要指标。一般来说,分析大学网站间的引用结构主要分为两种类型。一是研究大学网站间引用结构的形成受到哪些因素影响,例如Barnett等^[7]根据Cybermetrics实验室基于大学网站影响力做出的大学网站排名,选取了世界上58个国家排名前1000位的大学对其网络链接结构进行了描述,同时从大学层面和国家层面多个角度探讨了影响大学网络链接结构的原因。虽然Barnett在收集数据时包含了亚洲大学,然而在分析的过程中却没有体现。于是中国科学院国家科学图书馆成都分馆的几位学者^[8]同样根据Cybermetrics实验室的排名选取了中国排名前100的大学,通过社会网络分析法揭示了其主要网络特征,并从科研能力、师资能力、区域政治经济水平等方法研究了大学网络链接结构的影响因素。二是使用不同链接分析和社会网络分析角度去分析大学网站间引用结构的特征。国外学者,例如英国、美国、加拿大、西班牙等,关于大学网络链接的一些研究揭示了某个特定国家和地区某些学科间的联系^[7],同时也有学者通过分析国际上大学的超链接结构,发现亚洲的大学网站与欧洲大学的联系更紧密,而不是彼此联系,这是学术交流国际化的体现^[9]。在国内,傅余洋子等利用因子分析和主成分分析等方法对江苏省11所高校的网站链接结构进行了分析^[10]。邓三鸿提出用“Title Mention”替代链接分析,以我国985高校为研究对象,对这39所高校的名称共现和相互引荐数据进行了相关性分析^[11]。笔者发现虽然有不少学者利用网络计量方法来研究高校间的网站链接结构,但目前少有学者针对大学网站间的引用结构网络研究大学网站综合影响力。本文选取了中国教育部确定的112所211大学作为研究对象,以英语版必应搜索引擎作为工具收集211大学网站间的互链数据,通过社会网络分析方法研究高校网站的中心性排名,并首次将熵权TOPSIS方法运用到大学网站链接结构分析中,对高校网站

综合影响力进行排名。

2 数据收集与预处理

本文为了研究中国大学网站间的链接结构中具有影响力的节点,选取国家教育部确定的112家211院校作为研究对象^[12],这些大学的教育水平以及网站建设质量相对较好。在数据来源上,世界上最大的搜索引擎谷歌已经退出中国,而中国最大的搜索引擎百度其网页覆盖范围小、返回结果不稳定且不支持检索式“university A web domain”site: university B web domain来检索B入链到A大学的记录,所以本文使用英语版必应搜索引擎收集大学间互相引用的关系。注意的是,在笔者手工检索阶段,发现同样是必应搜索引擎,英语版的必应搜索引擎返回结果要好于中文版的必应搜索引擎返回结果,例如,在中文版必应搜索引擎中输入检索式“nju.edu.cn”site:tongji.edu.cn检索从同济大学网站链入到南京大学网站的网页数,结果是0,而在英语版必应搜索引擎中会返回多条结果,类似的情形多次出现使得使用英语版必应搜索引擎是更好的选择,可以获得更完整的数据。笔者编写了爬虫代码自动在英语版必应搜索引擎中通过检索式“university A web domain”site: university B web domain来检索大学网站间的互链数据,检索时间为2017年1月。本文没有使用大学名称作为检索字段,主要是考虑到大学的曾用名(如厦门理工学院曾用名鹭江职业大学)以及自然语言的歧义会对检索结果产生影响。当然并不是所有的返回结果都表示这两个大学间的网络引用关系,例如论文合作者的邮箱地址包含检索式域名的会被作为结果返回,但是这种现象本质上反映了大学之间的联系,与实际网络链接分析并无很大区别,因而不会排除这种返回结果。最终收集到的原始网站引用频次形成了一个112*112的矩阵,其中第x行第y列表示大学网站x对大学网站y的网站引用次数。因为本研究只考虑大学网站间的引用关系,所以在收集数据时排除了大学网站的自引用情况,即表示矩阵对角线上是0。由于所得的网站引用频次变异程度很大,(最高的有7510000,最低的为0次)不利于发现数据蕴含的意义。因此本研究借鉴了中国科学院国家科学图书馆成都分馆的几位学者的做法^[4],对原始数据进行了min-max标准化法,使得数据落在[0,1]区间,并保留五位小数。

将标准化的矩阵用R语言的igraph包进行分析,计算得到图的密度是0.1868,可见中国112所211高校网站之间交流不是很紧密。图1是使用R语言绘制出的211高校网站引用关系网络图,为了显示清楚,只有节点度数大于56的节点标签可见,节点度数的大小决定了节点大小,笔者使用了随机游走社群发现算法^[13]将网络图的大学进行聚类,在RStudio软件中相同的颜色表示属于同一个类。

由图1可以看出,北京大学、清华大学、浙江大学、上海交通大学、西安交通大学、复旦大学等高校彼此之间交流密切,形成一个大类,这些高校虽然属于中国不同的省份,但是

大部分都是处于中国东部经济发达的城市,同时也是中国高校中科研实力比较强的学校。除此之外,还有同济大学、中国药科大学和西北大学等联系也比较密切。由此可见,高校之间联系程度与地理位置无关。通过观察节点大小,可以发现东南大学、西南交通大学和中南大学等与其他较多高校的网站之间存在直接联系。



图1 211高校网站引用关系网络图

3 传统社会网络中心性分析

在社会网络分析中,中心性分析是主要研究方法之一,包括节点中心性和网络中心势。其中节点中心性的指标有四种:点度中心性(degree centrality)、接近中心性(closeness centrality)、中介中心性(betweenness centrality)、特征向量中心性(Eigenvector centrality)。

点度中心性是指某节点存在直接联系的其他节点的数目,是衡量节点重要性大小的一项指标。中心度越高的节

点越能影响网络中的其他节点,在网络中处于重要位置。反之,中心度越低的节点对网络中的其他节点影响力越小。可见,度中心性可以作为衡量各高校在社会网络中重要性的指标^[14]。

接近中心性是指网络中某点越接近其他点,则传递信息越容易,也越可能成为网络中的中心^[15]。接近中心性越高表明该大学网站可通过越短的路径与多个网站相连,相对来说该网站更加不受其他网站的控制。

中介中心性是以经过某个节点的最短距离数目来刻画节点重要性的指标。中介中心性可以作为结构洞的一个度量指标,具备高中介中心性的节点可以被认为是结构洞的占据者。结构洞是美国社会学家波特提出的概念,是指社会网络中的空隙,即社会网络中某个或某些个体和有些个体发生直接联系,但与其他个体不发生直接联系,即无直接关系或关系间断,从网络整体看好像网络结构中出现了洞穴^[16]。高中介中心性的节点显示了信息流被限制到通过这些节点的路径。在某种程度上,该指标代表了在社会网络中社会资本或者智力资本的集中区域,因此可以作为大学网站影响力的指标^[17]。

特征向量中心性是节点在网络中的整体中心性的度量指标^[18]。该指标认为一个节点的重要性既与它邻居节点的数量有关,而且也与邻居节点的重要性有关,也就是说一个节点连接到更多中心节点则其变得更加中心,在整个社会网络中的位置更加重要。

表1至表4是分别是运用R语言计算得到的点度中心性、接近中心性、中介中心性、特征向量中心性排在前20位的大学。从上述表格可以看出,不同的中心性指标得到的大学排名各不相同。

在211大学社会网络中度中心性处于领先地位的有:中

表1 大学网站引用网络——点度中心性排在前20位的大学

Total degree	Values	Out-degree	Values	In-degree	Values
中南大学	141	北京大学	74	南京师范大学	111
西安交通大学	138	吉林大学	62	中南大学	111
东南大学	134	上海交通大学	52	中国药科大学	111
兰州大学	133	厦门大学	51	兰州大学	111
哈尔滨工业大学	133	清华大学	50	华中农业大学	111
南京师范大学	132	浙江大学	41	华东理工大学	111
大连理工大学	129	北京航空航天大学	40	哈尔滨工业大学	111
中国药科大学	127	武汉大学	38	大连理工大学	111
电子科技大学	127	中国科学技术大学	37	第四军医大学	111
华中农业大学	126	中山大学	36	电子科技大学	111
华东理工大学	126	复旦大学	36	中国传媒大学	111
西北大学	125	华中科技大学	36	西安交通大学	111
第四军医大学	124	南京大学	31	东南大学	111
中国传媒大学	124	中南大学	30	西北大学	110
北京大学	116	大连海事大学	29	上海大学	81
上海大学	109	上海大学	28	北京大学	42
清华大学	87	中国人民大学	27	复旦大学	40
上海交通大学	77	西安交通大学	27	同济大学	39
复旦大学	76	北京师范大学	26	清华大学	37
吉林大学	72	同济大学	25	西藏大学	36

表2 大学网站引用网络—接近中心性排在前20位的大学

Closeness	Values	Out-closeness	Values	In-closeness	Values
浙江大学	24.34275	大连海事大学	0.00047	太原理工大学	10.78865
同济大学	24.32498	安徽大学	0.00045	南京理工大学	10.33805
清华大学	24.30724	上海外国语大学	0.00043	北京工业大学	10.18434
北京大学	24.28363	中国石油大学	0.00043	内蒙古大学	10.11634
上海交通大学	24.26007	北京交通大学	0.00043	西藏大学	10.09387
上海大学	24.26007	对外经济贸易大学	0.00043	北京邮电大学	10.08369
中国农业大学	24.2483	东北林业大学	0.00043	清华大学	9.75515
华中科技大学	24.23655	北京中医药大学	0.00043	上海大学	9.74849
武汉大学	24.23068	中国政法大学	0.00043	武汉大学	9.74754
吉林大学	24.1955	海南大学	0.00043	浙江大学	9.74469
中国科学技术大学	24.18965	北京林业大学	0.00043	中国矿业大学	9.74279
西藏大学	24.17795	中南财经政法大学	0.00043	中国科学技术大学	9.74184
西安电子科技大学	24.16626	延边大学	0.00043	华中科技大学	9.73994
厦门大学	24.16043	北京体育大学	0.00043	西安电子科技大学	9.73994
南京大学	24.15459	天津医科大学	0.00043	山东大学	9.73899
中山大学	24.15459	东北农业大学	0.00043	第二军医大学	9.73805
北京师范大学	24.15459	宁夏大学	0.00043	南昌大学	9.73615
北京航空航天大学	24.14876	石河子大学	0.00043	南京大学	9.73425
中国人民大学	24.14293	华北电力大学	0.00043	苏州大学	9.72857
复旦大学	24.14293	辽宁大学	0.00043	中央民族大学	9.71534

南大学、西安交通大学和东南大学。由于网站链接关系矩阵为非对称矩阵,点度中心性可以分为入度中心性和出度中心性。在中国211高校网站链接结构中,南京师范大学、中南

表3 大学网站引用网络—中介中心性排在前20位的大学

Betweenness	Values
北京大学	2598.21902
西安交通大学	962.33969
中南大学	946.19689
清华大学	650.28722
兰州大学	607.01065
哈尔滨工业大学	596.04406
上海交通大学	525.38966
南京师范大学	449.44753
上海大学	414.66483
复旦大学	386.57258
东南大学	345.49151
厦门大学	315.99654
吉林大学	278.37565
中国科学技术大学	216.33394
武汉大学	194.5599
浙江大学	176.99228
中国药科大学	160.80473
中国农业大学	129.23544
华中科技大学	122.49648
东北大学	112.90299

大学和中国药科大学的入度中心性处于前列,说明这三所高校的网站具有权威性,发布的信息具有较高价值;而出度中心性较高的是北京大学、吉林大学和上海交通大学,反映出其在信息传递以及与其他大学网站交流中的主动性和积极性较高。就特征向量中心性而言,在网络中处于重要地位的

是中南大学、西安交通大学和哈尔滨工业大学,它们都是度中心性比较高的大学。观察中介中心性,北京大学、西安交通大学和中南大学排名靠前,可以认为这三所学校在211高校网站结构图中处于结构洞的位置,是信息的集中区域。

表4 大学网站引用网络—特征向量中心性排在前20位的大学

Eigenvector	Values
中南大学	1
西安交通大学	0.99382
哈尔滨工业大学	0.95894
东南大学	0.957
兰州大学	0.94183
南京师范大学	0.93697
大连理工大学	0.91363
电子科技大学	0.89296
华东理工大学	0.88835
华中农业大学	0.88414
西北大学	0.88212
中国药科大学	0.87937
北京大学	0.86757
上海大学	0.86617
第四军医大学	0.8648
中国传媒大学	0.8648
清华大学	0.74568
复旦大学	0.71479
上海交通大学	0.70544
浙江大学	0.6704

从接近中心性的角度,浙江大学、同济大学和清华大学这三所高校处于前列。类似点度中心性,接近中心性也分为入接近中心性和出接近中心性,排在第一位的分别是太原理

工大学和大连海事大学,说明太原理工大学的网站知名度较高,被链接的捷径多,更容易被其他高校网站访问,而在大连海事大学网站寻找信息资源比较容易。不同的度量指标都代表了某一角度上的大学网站影响力,在实际使用中,只根据一个指标来衡量某个大学的网站影响力过于片面。而且对于接近中心性,我们可以发现无论是总体的中心性还是入接近中心性或者出接近中心性,其数值差别都很细微,因此以这些指标中的某一个来给中国211大学网站的影响力排名都是片面的。

4 基于熵权TOPSIS的大学网站综合影响力评估

单独考虑一种中心度来评估社会网络节点的影响力是很片面的,这些度量分别蕴含着节点的重要性、声望以及地位信息等,所以在评估节点影响力的过程中,比较周全的方法是将各种中心度量指标结合起来解决社会网络中节点影响力的评估问题,这种方法称为多属性决策(Multiple Attribute Decision Making, MADM)。TOPSIS作为一种经典的MADM法,已经吸引了众多学者和实践者的关注^[19-21]。

TOPSIS的全称是“逼近与理想值的排序方法”(Technique for Order Preference by Similarity to Ideal Solution),是Hwang和Yoon与1981年提出的一种适用于根据多项指标、多个方案进行比较分析的分析方法^[22]。这种方法的中心思想是选择同时具有距离正理想解的最短距离和距离负理想解的最远距离的解决方案。正理想解最大化利益标准并最小化成本标准,而理想解使成本标准最大化并使得利益标准最小化。TOPSIS充分利用属性信息,并且不要求属性偏好是独立的。图2给出了Hwang和Yoon实施TOPSIS的步骤过程。首先将初始判定矩阵进行归一化处理,然后建立加权归一化判定矩阵,步骤三是确定的正理想解和负理想解,并在步骤四中计算每个目标与正负理想解的距离。该过程以计算每个目标的相对接近系数而结束,根据相对接近系数进行降序排序从而得到各个目标的排序结果列表。具体步骤如下:

设有 m 个目标节点, n 个度量指标,用 y_{ij} 表示第 i 个目标节点在第 j 个指标上的得分,多属性决策问题的决策矩阵 $Y=\{y_{ij}\}$,则:

步骤一:将决策矩阵归一化处理: $Z_{ij}=\frac{y_{ij}}{\sum_{i=1}^m y_{ij}}$, $i=1, \dots, m, j=1, \dots, n$,得到规范化决策矩阵 $Z=\{z_{ij}\}$ 。

步骤二:构建权重决策矩阵:设各指标的权重为 $w=(w_1, w_2, \dots, w_n)^T$,则 $x_{ij}=w_{ij} \cdot z_{ij}$,得到决策规范矩阵 $X=\{x_{ij}\}$ 。

步骤三:计算理想解和负理想解:正理想解表示为 $x^*=(x_1^*, x_2^*, \dots, x_n^*)^T$,每个值对应该指标下的最优值;负理想解表示为 $x'=(x_1', x_2', \dots, x_n')^T$ 每个值对应该指标下的最劣值。

步骤四:计算各个目标到正负理想解的距离:备选目标 x_i 到理想解的距离为 $d_i^*=\sqrt{\sum_{j=1}^n (x_{ij}-x_j^*)^2}$, $i=1, \dots, m$;备选目标 x_i 到负理想解的距离为 $d_i'=\sqrt{\sum_{j=1}^n (x_{ij}-x_j')^2}$, $i=1, \dots, m$ 。

步骤五:计算各个目标的综合评价指数: $C_i=\frac{d_i'}{d_i'+d_i^*}$ 。

最后按照 C_i 由大到小排列各目标的次序,获得节点的排名。

本文将TOPSIS方法运用于大学网站影响力排名,选择的指标有:度中心性、出度中心性、入度中心性、特征中心性、中介中心性、接近中心性、出接近中心性和入接近中心性,分别用1~8表示。虽然TOPSIS法能够定量反映不同比较对象的优劣程度,方便、灵活,但是没有引入权重思想,通常在计算权重时采用专家打分法或者经验判断法,在一定程度上削弱了该方法的客观性^[23]。因此本文将熵权法应用到TOPSIS方法的权重确定中,信息熵是信息论中标度不确定的量,某项指标携带和传输的信息越多,不确定就越少,熵也越小,其表示该指标对决策的作用较其他携带和传输较少信息的指标要大。因此,本文利用信息熵的大小来反映不同评价指标在决策中所起作用的程度。信息熵权的具体公式是:

①首先求各个指标的熵值。

$$H_j=-k \sum_{i=1}^m z_{ij} \ln z_{ij}, j=1, 2, 3, \dots, n。$$

其中 z_{ij} 为规划化决策矩阵中第 i 个节点在第 j 个指标上的数值, k 为常规系数, $k=1/\ln(m)$,同时假定当 $z_{ij}=0$ 时, $z_{ij} \ln z_{ij}=0$ 。

$$\textcircled{2} \text{求各指标的权值: } w_j=\frac{1-H_j}{n-\sum_{j=1}^n H_j}, j=1, 2, 3, \dots, n。$$

经过计算,得到高校影响力各个度量指标的熵值分别是: $H_1=0.927$ 、 $H_2=0.980$ 、 0.772 、 0.608 、 0.966 、 0.941 、 0.999 、 0.984 ,最终得到的权重是: $w_1=0.0884$ 、 $w_2=0.0246$ 、 $w_3=0.2773$ 、 $w_4=0.4767$ 、 $w_5=0.0412$ 、 $w_6=0.0718$ 、 $w_7=0.0002$ 、 $w_8=0.0198$ 。可以发现在这8个度量指标中,权重排在前面的是特征向量中心性和入度中心性,说明这两个指标在大学网站影响力排名中贡献比较大,而接近中心性权重较低,尤其是出接近中心性,主要是因为在该指标上这112所高校的得分并没有很大差异,因而在网站影响力排名上其贡献不大。然后利用excel实现上文说明的TOPSIS过程,得到这112所211高校网站影响力的综合排名,结果如表5所示。网站综合影响力排在第一位的是北京大学,该校的特征向量中心性和入度中心性都是位于前列的,证明了这两个指标在高校网站影响力上的重要作用。可以发现这些排名与大学的科研实力并不是完全相关的,如教育水平和科研实力处于中国高校前列的清华大学,其排名要低于西安交通大学、中南大学和兰州大学。高校网站是一个学校向外宣传自己的途径,同时也是学校间进行信息交流与资源共享的一个平台,高校在提高科研水平的同时应该注重加强网站建设,将科研成果和学术信息及时网络化,增强网站吸引链接的能力,提高高校网站社会影响力。

5 结 语

本文运用英语版必应搜索引擎获取了112所中国211高校网站的互链数据。通过分析可以发现高校网站间的互链

表5 112所211高校网站影响力的综合排名

大学名称	排名	大学名称	排名	大学名称	排名	大学名称	排名
北京大学	1	北京工业大学	29	河海大学	57	北京外国语大学	85
西安交通大学	2	西藏大学	30	上海财经大学	58	四川大学	86
中南大学	3	南京大学	31	北京科技大学	59	大连海事大学	87
兰州大学	4	山东大学	32	东华大学	60	四川农业大学	88
清华大学	5	中山大学	33	中国海洋大学	61	湖南师范大学	89
哈尔滨工业大学	6	西安电子科技大学	34	哈尔滨工程大学	62	安徽大学	90
上海交通大学	7	天津大学	35	中国矿业大学	63	中国石油大学	91
南京师范大学	8	中国人民大学	36	苏州大学	64	中央音乐学院	92
上海大学	9	内蒙古大学	37	暨南大学	65	贵州大学	93
东南大学	10	北京师范大学	38	中央民族大学	66	上海外国语大学	94
复旦大学	11	南京航空航天大学	39	第二军医大学	67	对外经济贸易大学	95
中国药科大学	12	华南理工大学	40	华南师范大学	68	东北林业大学	96
厦门大学	13	江南大学	41	中央财经大学	69	北京林业大学	97
大连理工大学	14	重庆大学	42	北京化工大学	70	北京交通大学	98
电子科技大学	15	北京邮电大学	43	西南交通大学	71	中国政法大学	99
华东理工大学	16	北京航空航天大学	44	广西大学	72	海南大学	100
华中农业大学	17	南开大学	45	郑州大学	73	北京中医药大学	101
吉林大学	18	湖南大学	46	南京农业大学	74	中南财经政法大学	102
中国传媒大学	19	南昌大学	47	西南大学	75	青海大学	103
第四军医大学	20	国防科学技术大学	48	东北师范大学	76	延边大学	104
西北大学	21	华中师范大学	49	长安大学	77	北京体育大学	105
中国科学技术大学	22	福州大学	50	西北农林科技大学	78	天津医科大学	106
武汉大学	23	武汉理工大学	51	合肥工业大学	79	东北农业大学	107
浙江大学	24	南京理工大学	52	陕西师范大学	80	宁夏大学	108
同济大学	25	北京理工大学	53	中国地质大学	81	石河子大学	109
中国农业大学	26	西北工业大学	54	云南大学	82	华北电力大学	110
华中科技大学	27	太原理工大学	55	西南财经大学	83	辽宁大学	111
东北大学	28	华东师范大学	56	新疆大学	84	河北工业大学	112

数在绝对值上比较低,互链网络总体密度值是0.1868,远低于中国网站排名网2012年排名前50位的新闻类网站构成的网络图密度0.3127^[4],这说明211高校网站间联系不紧密,信息交流和资源共享的水平尚有待提高。同时以传统社会网络分析方法中中心性分析得到的高校影响力排名各不相同,使用单一的度量指标来衡量高校网络影响力是不全面的,因而本文使用熵权TOPSIS方法得到了211高校的网站影响力的综合排名,这是首次将多属性决策方法应用于高校网站影响力分析。但是由于数据完备性欠缺,得到的结果可能不完全正确。主要原因是,在链接分析中,使用的是国外搜索引擎,它会按照自己的取舍标准进行挑选,这与我们统计数据处理过程中要求覆盖所有结果的要求本身就有差距。同时由于相对链接的使用也会影响数据的完整性,绝对链接给出了完全的URL,可被搜索引擎提取,但是相对链接仅仅给出了一个与参考网页相关的向导和文件路径,搜索引擎不能提取。这些都将导致数据收集上的误差从而影响最终结果的准确性。当然随着技术不断进步和搜索引擎的不断发展,相信未来可以使用更加完备的数据得到更加准确的分析结果。

参考文献

- 1 龚立群,朱庆华.网络计量学的研究方法及应用[J].新世纪图书馆,2003,(6):6-9.
- 2 邱均平,张洋.网络信息计量学综述[J].高校图书馆工作,2005,25(1):1-12.
- 3 晏尔伽.中国省会城市政府网站链接分析[J].情报科学,2008,26(2):218-223.
- 4 肖建英,张辉.中国新闻网站的影响力与链接关系研究[J].情报杂志,2012,31(10):65-70.
- 5 Vaughan L, Wu G. Links to commercial websites as a source of business information[J]. Scientometrics, 2004, 60(3): 487-496.
- 6 Vaughan L, Hysen K. Relationship between links to journal Web sites and impact factors[J]. Aslib Proceedings, 2002, 54(6): 356-361.
- 7 Barnett G A, Han W P, Jiang K, et al. A multi-level network analysis of web-citations among the world's universities[J]. Scientometrics, 2014, 99(1):5-26.
- 8 唐川,刘春江,徐婧,等.中国大学网站URL引用网络结构特征与相关因素研究[J].情报杂志,2014,33(4):46-51.
- 9 Park H W, Thelwall M. Web-science communication in the age of globalization[J]. New Media & Society

- ety, 2006, 8(4):629-650.
- 10 傅余洋子, 邓三鸿, 陆宇飞. 基于链接分析的江苏省高校网站建设与交流研究[J]. 图书馆学研究, 2015, (14):43-50.
 - 11 邓三鸿, 王 昊. 基于 Title Mention 的 985 高校名称共现及引荐关系分析[J]. 情报学报, 2014, 33(12): 1322-1334.
 - 12 中华人民共和国教育部. “211 工程”学校名单[EB/OL]. http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_94/201002/82762.html, 2017-02-28.
 - 13 Pons P, Latapy M. Computing communities in large networks using random walks[J]. Journal of Graph Algorithms and Applications, 2006, 10(2):191-218.
 - 14 魏志皓, 施国良. URL 与网络标题共现数据的特征对比实证研究——以“985 高校”为例[J]. 图书馆学研究, 2015, (23):80-86.
 - 15 刘 军. 整体网分析讲义 UNINET 软件实用指南[M]. 上海: 格致出版社, 2009:105.
 - 16 Burt R S. Structural holes: The Social Structure of Competition[M]. Cambridge: Harvard University Press, 2009:18.
 - 17 Leydesdorff L. Betweenness centrality as an indicator of the interdisciplinarity of scientific journals[J]. Journal of the American Society for Information Science and Technology, 2007, 58(9): 1303-1319.
 - 18 Bonacich P. Factoring and weighting approaches to status scores and clique identification[J]. Journal of Mathematical Sociology, 1972, 2(1): 113-120.
 - 19 Chen C T. Extensions of the TOPSIS for group decision-making under fuzzy environment[J]. Fuzzy sets and systems, 2000, 114(1): 1-9.
 - 20 陈 雷, 王延章. 基于熵权系数与 TOPSIS 集成评价决策方法的研究[J]. 控制与决策, 2003, 18(4): 456-459.
 - 21 朱 乾, 徐淑琼. 基于熵权 TOPSIS 法的社会网络节点影响力评估研究[J]. 南京航空航天大学学报: 社会科学版, 2016, 18(1): 42-46.
 - 22 Hwang, C. L., & Yoon, K. P. Multiple attribute decision making: Methods and applications[M]. New York: Springer-Verlag, 1981:39.
 - 23 李 伟, 杨国才. 基于熵权 TOPSIS 法的城市竞争力综合比较与时间演化分析——以“中四角”城市群为例[J]. 暨南学报(哲学社会科学版), 2014, 36(10):77-86.

(责任编辑:毛秀梅)

(上接第 145 页)

- 2010, 26(Z1):114-119.
- 14 Kuang L L. Application of Association Rule in Data Mining of Students' borrowing Books from Library in University[J]. Computer Knowledge & Technology, 2010, (4):784-786, 813.
 - 15 邓桂龙, 刘智慧, 贾志东. 作战仿真实验数据关联规则挖掘研究[J]. 军事运筹与系统工程, 2008, 22(4):46-50.
 - 16 Wang C, Satuluri V, Parthasarathy S. Local probabilistic models for link prediction[C]//Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on. IEEE, 2007: 322-331.
 - 17 Badia A, Kantardzic M. Link analysis tools for intelligence and counterterrorism[C]//International Conference on intelligence and security informatics, 2005, 3495: 49-59.
 - 18 Placida T, N. Deepika. Expert System to Detect Suspicious Words in Online Messages for Intelligence Agency Using FP-growth Algorithm[J]. International Journal of Computer Science and Mobile Computing, 2015, 4(7): 103-108.
 - 19 孙细明, 龚成芳. 关联规则在购物篮分析中的应用[J]. 计算机与数字工程, 2008, 36(6):57-60.
 - 20 王小虎. 关联规则挖掘综述[J]. 计算机工程与应用, 2003, 39(33):190-193.
 - 21 Han J, Jian P, Yin Y, et al. Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach[J]. Data Mining and Knowledge Discovery, 2004, 8(1):53-87.
 - 22 Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation[C]// ACM SIGMOD International Conference on Management of Data. ACM, 2000:1-12.
 - 23 郎 胜, 王爱立. 中华人民共和国反恐怖主义法释义[M]. 北京: 法律出版社, 2016:47.
 - 24 凤凰资讯. 新疆部分地区学习识别 75 种宗教极端活动[EB/OL]. http://news.ifeng.com/a/20141122/42785382_0.shtml, 2017-06-08.
 - 25 刘明吉, 王秀峰, 黄亚楼. 数据挖掘中的数据预处理[J]. 计算机科学, 2000, 27(4):54-57.

(责任编辑:毛秀梅)