基于组合赋权法的高校网络影响力综合评价

——以“双一流”高校为例

邓三鸿　陈超群　王　昊　张　越

[提要]随着互联网的快速发展，各个高校越来越重视网络信息资源的建设，塑造自身在网络环境中的形象和提升知名度。本文综合现有的高校评价和网络影响力评价研究，从官方平台、社交媒体和第三方平台三个网络影响力传播渠道构建了一个多维度的全面高校网络影响力评价指标体系，包含“量”和“质”两个角度的3个一级指标、6个二级指标和45个三级指标。然后以39个双一流高校为例，运用基于加法的综合归一化方法将层次分析法得到的主观权重和熵权法得到客观权重进行融合，得到更为科学合理的组合权重。最后运用TOPSIS方法进行高校网络影响力的排名分析，并运用谱聚类方法将39所高校依据网络影响力划分为5个类别，结合各个类别的特点对高校提升网络影响力提供有参考性的建议。

[关键词]双一流高校；组合赋权法；网络影响力；聚类分析

中图分类号：G250　　**文献标识码**：A　　**文章编号**：1004—3926(2019)09—0227—09

基金项目：国家社科基金“大数据环境下学术成果真实价值与影响的实时预测及长期评价研究”(19BTQ062)、中央高校基本科研业务费专项资金项目(010814370301)阶段性成果。

作者简介：**邓三鸿**，南京大学信息管理学院、江苏省数据工程与知识服务重点实验室教授，研究方向：科学计量；**陈超群**，南京大学信息管理学院硕士研究生，研究方向：信息检索；**王昊**，南京大学信息管理学院教授，研究方向：信息处理；**张越**，南京大学信息管理学院博士研究生，研究方向：信息处理。江苏 南京　210023

引言

2018年9月在全国教育大会上，习近平总书记提出“教育是国之大计，党之大计”[1]。一直以来，教育进步都是促进社会发展、民族繁荣昌盛的核心。高等教育的根本使命是培养人才，目的在于帮助每一个人认识自己、认识社会和改变社会，从而为国家科技、经济和文化的向前发展源源不断地输送优秀人才，推动中华民族伟大复兴的进程。在此背景下，对我国高等教育学校进行评价，可以从侧面反映出目前中国高等教育的发展状况，并为相关部门的决策提出建议，促进我国教育的全面发展。

随着互联网的普及以及智能手机的广泛应用，网络日益成为人们获取信息、提出建议以及制造舆论的重要渠道，有时甚至影响了人们的决策行为。高校的网络形象建设和信息化平台的搭建成为了高校发展新的要求，如何利用互联网来提升自己的网络影响力成为了一个亟待解决的课题。目前，几乎所有高校都搭建了自己的官方网站、借助官方微博账号和微信公众号等多种渠道来丰富自身的网络信息资源，突破时空的限制，通过网络来加强其与社会公众之间的联系，给大家提供更加人性化的服务，从而扩大高校在网络虚拟社区中的影响力。网络信息资源是对高校在网络上的传播力和影响力的一种反映，与高校在信息时代的全面发展息息相关，例如，某个学校在网络信息资源越丰富，那么它在网络上的知名度和传播力可能就会越高，可以帮助相关政府组织和社会个人在充分认识高校的发展状况和影响力的基础上做出正确决策。鉴于以上原因，为了更准确、更有效地评估高校网络影响力的现状，需要找出能够反映高校网络信息资源“质”和“量”两方面的具有代表性的指标，建立一套科学完善的评价指标体系，然后选择合适的评价方法对高校的网络影响力进行评估，为社会提供一个观察高校网络信息资源建设的窗口。

一、相关研究综述

事实上，目前对高校网络影响力进行评价的相关研究已经不少，前期主要是基于链接分析方法，选择网页数量、内部链接数、网络影响因子、链

接效率等指标对高校网站内部结构和外部辐射能力进行评价,发现在网站影响力评价中贡献度比较高的指标,为高校的网站建设与优化提出指导性的意见。21世纪初,从web链接中提取网站信息引起了很多研究人员的兴趣,Thelwall改进了Ingwerson在1998年提出的WIF(Web impact factor)计算公式,对比分析英国大学不同类型网站的网络影响因子,结果表明网络影响因子与大学研究水平排名密切相关[2]。国内学者邱均平通过对中国100所大学的网站进行分析,发现高校网站的总链接数、外部链接数和网络影响因子与中国大学排名以及科研得分之间具有显著关系,其中外部链接数相关性最高,这说明一个高校声誉越高、实力越强,指向其网站的链接也会越多[3]。与此类似的是,吴茵茵在对中国和美国的26所大学进行网络影响力进行测定时,发现外部网络影响因子与链接效率的线性正相关度要高于总体的网络影响因子,也说明了外部链接数的重要性[4]。

如今,新兴媒体形式,如微博、微信等作为一种快速及时且全民参与的信息传播渠道,逐渐成为了组织机构的主要信息发布方式,让各个组织机构实现了与外界的双向沟通,对提升其网络影响力也有一定的作用。基于社会媒体的高校网络影响力分析主要是通过采集新媒体平台上的数据对高校在社会大众中产生的影响程度进行分析。其中尤以微博影响力的研究居多,如赵扬等以高校图书馆官方微博为研究对象,从微博内容层面和用户层次对影响高校图书馆微博影响力的主要因素进行了定量分析,发现了对提升图书馆微博信息传播具有重要影响的信息类型、展现形式等因素[5]。此外,于洋洋等通过分析用户的行为特征构建了一个微博影响力综合评价模型,并以我国25所"985"高校为例进行了实证分析,用主成分分析和聚类分析对高校微博网络影响力进行了比较研究[6]。随着近年来微信的活跃用户越来越多,微信公众号在移动互联网世界的资讯传播上扮演了重要角色,少数学者也开始以微信公众号为例进行实例研究,提出了一种新的网络影响力评价方式[7]。

但是上述研究基本上都是围绕着高校官方网站和高校图书馆等,只是从单一的角度出发,缺乏系统的多视角对高校在整个网络环境中的影响力的研究。网络影响力的评价不能局限于单一对象,网络信息资源的形式多种多样,需要建立一个多维度的综合评价方法。本文通过对过去关于高校评价和网络影响力相关研究的多维度综合,采用科学计量学研究方法和工具,构建了一套由3个一层评价指标、6个二层评价指标、45个三层评价指标组成的高校网络影响力评估体系,以对现有的高校网络影响力评价体系进行进一步扩充与完善。同时,本文以39所双一流高校为研究对象,抓取了互联网真实客观存在的数据,在此基础进行定量的数据分析,融合熵权客观赋权法和主成分分析主观赋权法确定组合权重,较好地克服了这两种方法各自的不足之处,最后运用TOPSIS和聚类方法对高校网络影响力进行定量分析,根据评价结果对高校网络影响力建设提出了有价值的建议。

二、高校网络影响力评价指标体系构建

根据高校网络影响力的产生渠道,结合相关文献关于网络影响力的现有研究,高校主要借助官方网站、教育类政府网站、官方微博、官方微信公众号、百科词条和搜索引擎等不同渠道来展示高校相关信息和资讯、与社会进行交流,其在网络上留下的痕迹已经成为学生选择学校、用人单位选择学生、政府进行教育相关资源分配决策的辅助性工具。

高校网络影响力的传播渠道可以分为三种:官方平台、社交媒体以及第三方平台。官方平台的信息传播是最具有公信力、信息传播最准确的渠道,通过网页主动展示信息可以使大众快速获取高校资讯,进而树立自身形象;社交媒体主要包括微博和微信,是目前信息传播速度最为迅速,而且大众参与最广泛的渠道,利用频繁更新的博文和文章来跟大众进行互动交流,发布大众可能感兴趣的话题来提升高校网络影响力;第三方平台本文主要考虑的是百度百科以及百度新闻,可以说是目前发展最为成熟的信息传播渠道,同时也是大众获取高校简介信息和实时信息的主要途径。借鉴吕嘉构建的广东省国家森林公园网络影响力评价指标体系[8]和李敏谦的古村落网络信息资源评价指标体系[9],本文构建了一个三级评价指标体系,一级指标包括官方平台影响力、社交媒体影响力、第三方平台影响力。二级指标由官方网站网络影响力、政府网站网络影响力、官方微博网络影响力、微信公众号网络影响力、百科网络影响力和新闻网站网络影响力构成。最后结合客观性、相关性、全面性、必要性和可操作性的原则[10],基于链接分析和情感分析方法,从"质"和"量"两个角度选取三级指标,最终得到的评价指标体系以及对应指标的解释说明如下表所示:

表1 高校网络影响力评价指标体系

一级指标	二级指标	三级指标	解释说明
官方平台影响力 A1	官网网络影响力 B1	网页数量 C1	该学校域名网站下的网页数量
		网站总链接数 C2	链接指向学校网址的网页数量
		内部链接数 C3	该校网站网页中指向本网站其他网页的网页数量
		外部链接数 C4	在该校网站范围以外检索得到的指向该网站的网页数
		网络影响因子 C5	总链接数/网页数
		外部影响因子 C6	外链接数/网页数
		教育类网站反链数 C7	除本校外其他大学的网页中指向该学校网址的网页数量
		PR 值 C8	Google 用于标识网页的等级、重要性、网站的好坏的重要标准之一。级别从 0 到 10 级为满分。PR 值越高说明该网页越受欢迎。
		DPV 值 C9	人均页面访问次数
	政府网站网络影响力 B2	教育部提及数 C10	教育部门户网站提及到该学校的网页数
		科学技术部提及数 C11	科学技术部网站提及到该学校的网页数
社交媒体影响力 A2	官方微博网络影响力 B3	粉丝数量 C12	关注官方微博账号的粉丝数量
		日均微博数 C13	官方微博账号平均每日发布的微博数，包括转发微博与原创微博，可以反映出账号的活跃度
		平均每条微博转发数量 C14	平均每条微博被他人转发的数量，反映出微博传播的广度
		最高转发数 C15	评估期内所发微博的最高转发数
		平均每条微博评论数量 C16	平均每条微博被他人评论的数量，反映出微博传播的深度
		最高评论数 C17	评估期内所发微博的最高评论数
		平均每条微博点赞数量 C18	平均每条微博被他人点赞的数量，反映出公众对微博的认可度
		最高点赞数 C19	评估期内所发微博的最高点赞数
		平均每条微博中性评论数量 C20	平均每条微博被他人评论数中中性评论的数量
		平均每条微博正面评论数量 C21	平均每条微博被他人评论数中正面评论的数量，体现公众的积极情感
		平均每条微博负面评论数量 C22	平均每条微博被他人评论数中负面评论的数量，体现公众的消极情感
		微博等级 C23	用户使用微博时间长短及活跃情况的体现
	微信公众号网络影响力 B4	日均发文数 C24	官方微信公众号平均每日发表的文章数，包括原创文章与转发文章
		原创文章比例 C25	评估期所有发布的文章中原创文章的占比
		活跃粉丝数 C26	参与度比较高的粉丝数，包括阅读、点赞、评论、回台回复等行为
		平均每篇阅读数 C27	平均每篇文章被他人阅读的数量，反映出文章传播的广度
		最高阅读数 C28	评估期内所发文章的最高阅读数
		平均每篇点赞数 C29	平均每篇文章被他人点赞的数量，体现公众对文章内容的认可度
		最高点赞数 C30	评估期内所发文章的最高点赞数
		平均每篇文章的评论数量 C31	平均每篇文章被他人评论的数量，反映出文章传播的深度
		最高评论数 C32	评估期内所发文章的最高点赞数
		平均每篇文章的中性评论数量 C33	平均每篇文章被他人评论中中性评论的数量
		平均每篇文章的正面评论数量 C34	平均每篇文章被他人评论中正面评论的数量，体现公众的积极情感
		平均每篇文章的负面评论数量 C35	平均每篇文章被他人评论中正面评论的数量，体现公众的消极情感
		WCI(微信传播指数)C36	通过微信公众号推送文章的传播度、覆盖度、账号的成熟度和影响力来反映微信整体热度和公众号的发展走势。

一级指标	二级指标	三级指标	解释说明
第三方平台影响力 A3	百科网络影响力 B5	词条的编辑次数 C37	该学校的词条被编辑的次数,反映公众的关注度
		词条的浏览次数 C38	该学校的词条被浏览的次数,反映公众的关注度
		词条的长度 C39	该学校的词条总字数,反映信息的完善程度
		词条的转发次数 C40	该学校的词条被转发的次数,反映信息传播的广度
		词条的点赞次数 C41	该学校的词条被点赞的次数,反映公众对词条的认可度
	新闻网站网络影响力 B6	百度新闻报道的数量 C42	评估期内有关该学校的所有新闻数
		百度新闻报道的正面新闻比例 C43	评估期内有关该学校的所有新闻中正面新闻所占比例,体现公众的积极情感
		百度新闻报道的负面新闻比例 C44	评估期内有关该学校的所有新闻中负面新闻所占比例,体现公众的消极情感
		百度新闻报道的中性新闻比例 C45	评估期内有关该学校的所有新闻中中性新闻所占比例

三、数据收集与预处理

限于篇幅,本文以双一流高校为研究对象。在教育部公布的首批42 所双一流高校名单中[11],国防科技大学至今还没有开通微博,北京理工大学2018 年9 月30 日才开通微博,新疆大学虽然开通微博较早,但是最新的一条微博发布在2018 年4 月14 日,微博数据比较少,因此在本文研究中排除了这三所学校,最终以39 所高校作为研究对象在互联网上收集数据,包括搜索引擎、微博、微信、百度百科等,数据收集时间2018 年11 月1 日至12 月10 日。

对于采集到的数据,需要进行异常数据的清洗。一般来说,高校的一条微博的评论数普遍是几十条,少部分是几百条,但存在一些热门事件会引起社会群体的广泛关注,出现微博的评论数、点赞数、转发数突然激增的情况。例如,2018 年8 月22 日湖南大学发布了一条查询四六级成绩的微博,参与这个话题的人数远超平常,有5827 位用户点赞,3025 位用户转发,7513 位用户评论了这条微博。本文将评论数超过1000 的微博视为热门事件的微博,需要注意的是,这些热门事件只是偶然发生的,每个高校在评估期内并不一定都会发生热门事件,而且持续周期以及引起社会关注和讨论程度也是不一致的,这些微博数据可以称之为微博网络影响力评估体系中的极端值,在计算平均每条微博转发数量、平均每条微博评论数量等均值类指标时,如果将极端值考虑进去得到的结果往往难以真正代表"平均水平",因此为了得到更加客观的结果,在进行微博网络影响力指标数值计算时删除了热门事件的微博。

清洗后的数据可以分为两类,一是数值型数据,如网站网页数量、新闻报道数量和微博关注数量等;另一类为文本数据,如微博评论、微信评论和新闻标题等,对于这类数据需要将其转化为可量化的指标,本文采用中文文本分析工具 Snownlp 进行情感分析,对于每条文本,Snownlp 会输出一个[0,1]的值,表示该文本为积极情感的概率,文章在情感分类处理时将输出概率在[0,0.4]之间的划分为负面类文本,(0.4,0.6)之间的为中性类文本,[0.6,1]之间的为正面类文本,以此得到平均每条微博中性评论数量、平均每条微博中性评论数量等"量"指标的具体数值。

由于网络影响力评价指标的来源和性质不同,收集到的原始数据具有不同的数量级和量纲,例如,三级指标网页数量和 PR 值,PR 值这一指标的最大值是9,最小值是0,而网页数量的数值往往都是十万级,甚至是百万级,这两个指标数值的水平相别很大,为了不同量级的指标具有可比性,需要对原始数据进行标准化的处理,以消除单位和量纲的不一致性。本文选择归一化方式进行数据标准化,设 x_{ij} 表示第 j 个样本的第 i 个指标数值标准化后的结果,v_{ij} 表示第 i 个样本的第 j 个指标原始数值,n 表示样本的数量,转换公式为:$x_{ij}=\frac{v_{ij}}{\sum_{i=1}^{n} v_{ij}}$。

在对原始数据进行观察时,发现除了原创文章比例、百度新闻报道的负面新闻比例、百度新闻报道的中性新闻比例这三个指标的最小值为0,其他指标数值的最小值都大于0。但是指标为0 的值归一化处理后结果还是0,这样无法进行取对数的运算,因而本文对于数值为0 的指标用0.00001

替换后再进行归一化处理，这样做的好处在于在几乎不会给原始数据带来的影响的情况下，为后续用熵权法求客观权重做准备。

四、指标权重确立

在对高校的网络影响力进行定量评价之前，需要对评价指标体系中各个指标的权重进行计算，考虑到主观赋权法存在的主观随意性以及客观赋权法存在的与实际情况不相符的缺陷，本文采用组合赋权法，运用基于加法的线性归一化方法将主观权重和客观权重进行融合。

主观权重的确定采用的是层次分析法，在构建判断矩阵的的过程中，笔者借鉴了德尔菲法的思想，用问卷的形式邀请6位专家，由3位研究生、2位博士生、1位教授组成，分别对不同层次的指标的重要性做出判断，他们对这个领域的基本知识均有一定的了解。接下来运用层次分析法，基于得到的指标重要性的综合判断构建判断矩阵，采用几何平均法对层次单排序和层次总排序，然后进行一致性检验，计算每层各指标的主观权重。

按照从上到下的原则，首先对一级指标进行权重分配，其判断矩阵为 $\begin{bmatrix} 1 & 1/2 & 2 \\ 2 & 1 & 4 \\ 1/2 & 1/4 & 1 \end{bmatrix}$，采用几何平均法得到的一级指标权重为 P_1 = (0.29,0.57,0.14)，计算得到一致性指标 CI = 0.0，参考龚木森、许树柏得出的1－15阶重复计算1000次的平均随机一致性指标[15]查询对应的随机一致性指标 CR = 0.0，其值小于0.1，矩阵通过了一致性检验。同理，对二级指标进行同样的操作来确定各指标的权重。A1对应的判断矩阵为 $\begin{bmatrix} 1 & 1/2 \\ 2 & 1 \end{bmatrix}$，A2对应的判断矩阵为 $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$，A3对应的判断矩阵为 $\begin{bmatrix} 1 & 1/2 \\ 2 & 1 \end{bmatrix}$，对于A1计算得到的权重为 P_2 = (0.67,0.33)，A2得到的权重为 P_3 = (0.5,0.5)，A3得到权重为 P_4 = (0.67,0.33)，因为二阶判断矩阵本身就具有完全一致性，RI = 0，因此二级指标的判断矩阵均符合一致性要求。

最后，对三级指标的权重进行计算。B1对应的判断矩阵为

$$\begin{bmatrix} 1 & 1/2 & 2 & 1/4 & 1/2 & 1/4 & 1/6 & 1/6 & 1/2 \\ 2 & 1 & 4 & 1/2 & 1 & 1/2 & 1/3 & 1/3 & 1 \\ 1/2 & 1/4 & 1 & 1/8 & 1/4 & 1/8 & 1/9 & 1/9 & 1/4 \\ 4 & 2 & 8 & 1 & 2 & 1 & 1/2 & 1/2 & 2 \\ 2 & 1 & 4 & 1/2 & 1 & 1/2 & 1/3 & 1/3 & 1 \\ 4 & 2 & 8 & 1 & 2 & 1 & 2 & 2 & 1/2 \\ 6 & 3 & 9 & 2 & 3 & 1/2 & 1 & 1 & 3 \\ 6 & 3 & 9 & 2 & 3 & 1/2 & 1 & 1 & 3 \\ 2 & 1 & 4 & 1/2 & 1 & 2 & 1/3 & 1/3 & 1 \end{bmatrix}$$

，求的权重为 P_5 = (0.04,0.08,0.02,0.14,0.08,0.16,0.20,0.20,0.09)，其中 CI = 0.06136，得到 CR = 0.042 < 0.1，故判断矩阵符合一致性要求。B2对应的判断矩阵为 $\begin{bmatrix} 1 & 1/2 \\ 2 & 1 \end{bmatrix}$，求得的指标权重为 P_6 = (0.33,0.67)，RI = 0，判断矩阵符合一致性要求。B3对应的判断矩阵为

$$\begin{bmatrix} 1 & 3 & 1/2 & 1 & 1/4 & 1/3 & 1/2 & 1 & 4 & 1/3 & 1/5 & 1 \\ 1/3 & 1 & 1/5 & 1/3 & 1/6 & 1/5 & 1/5 & 1/3 & 1 & 1/6 & 1/7 & 1/3 \\ 2 & 5 & 1 & 2 & 1/3 & 1/3 & 1 & 2 & 3 & 1/2 & 1/3 & 2 \\ 1 & 3 & 1/2 & 1 & 1/3 & 1/2 & 1/2 & 1 & 2 & 1/2 & 1/3 & 1 \\ 4 & 6 & 3 & 3 & 1 & 2 & 2 & 4 & 8 & 2 & 1/2 & 4 \\ 3 & 5 & 3 & 2 & 1/2 & 1 & 1/2 & 1 & 2 & 1/2 & 1/3 & 3 \\ 2 & 5 & 1 & 2 & 1/2 & 2 & 1 & 2 & 3 & 1/2 & 1/3 & 2 \\ 1 & 3 & 1/2 & 1 & 1/4 & 1 & 1/2 & 1 & 2 & 1/2 & 1/3 & 1 \\ 1/4 & 1 & 1/3 & 1/2 & 1/8 & 1/2 & 1/3 & 1/2 & 1 & 1/2 & 1/3 & 1/4 \\ 3 & 6 & 2 & 2 & 1/2 & 2 & 2 & 2 & 2 & 1 & 1/2 & 3 \\ 5 & 7 & 3 & 3 & 2 & 3 & 3 & 3 & 3 & 2 & 1 & 4 \\ 1 & 3 & 1/2 & 1 & 1/4 & 1/3 & 1/2 & 1 & 4 & 1/3 & 1/4 & 1 \end{bmatrix}$$

求得的指标权重为 P_7 = (0.05,0.02,0.08,0.05,0.18,0.09,0.09,0.05,0.03,0.12,0.20,0.05)，其中 CI = 0.05975，得到 CR = 0.0388 < 0.1，判断矩阵符合一致性要求。B4对应的判断矩阵为

$$\begin{bmatrix} 1 & 2 & 1/3 & 1/2 & 1 & 1/2 & 1 & 1/3 & 1/2 & 1 & 1/4 & 1/5 \\ 1/2 & 1 & 1/4 & 1/3 & 1 & 1/3 & 1 & 1/3 & 1/2 & 1/3 & 1/6 & 1/7 \\ 3 & 4 & 1 & 1/2 & 2 & 1/2 & 2 & 1/3 & 1/2 & 3 & 1/2 & 1/3 \\ 2 & 3 & 2 & 1 & 2 & 1 & 2 & 1/3 & 1/2 & 1/2 & 1/3 & 1/4 \\ 1 & 1 & 1/2 & 1/2 & 1 & 1/2 & 1 & 1/3 & 1/2 & 1 & 1/4 & 1/5 \\ 2 & 3 & 2 & 1 & 2 & 1 & 2 & 1/2 & 1/2 & 1/2 & 1/3 & 1/4 \\ 1 & 1 & 1/2 & 1/2 & 1 & 1/2 & 1 & 1/3 & 1/2 & 1 & 1/4 & 1/5 \\ 3 & 3 & 3 & 3 & 3 & 2 & 3 & 1 & 1/2 & 3 & 5 & 6 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 2 & 3 & 4 \\ 1 & 3 & 1/3 & 2 & 1 & 2 & 1 & 1/3 & 1/2 & 1 & 1/4 & 1/5 \\ 4 & 6 & 2 & 3 & 4 & 3 & 4 & 1/5 & 1/3 & 4 & 1 & 1/2 \\ 5 & 7 & 3 & 4 & 5 & 4 & 5 & 1/6 & 1/4 & 5 & 2 & 1 \end{bmatrix}$$

指标权重为 P_8 = (0.04,0.02,0.06,0.06,0.04,0.06,0.04,0.15,0.13,0.05,0.12,0.15,0.09)。其中 CI = 0.150,得到 CR = 0.09 < 0.1,判断矩阵符合一致性要求。B5 对应的判断矩阵为

$$\begin{bmatrix} 1 & 1/2 & 2 & 1/4 & 1/3 \\ 2 & 1 & 3 & 1/3 & 1/2 \\ 1/2 & 1/3 & 1 & 1/6 & 1/5 \\ 4 & 3 & 6 & 1 & 2 \\ 3 & 2 & 5 & 1/2 & 1 \end{bmatrix}$$

,求得的指标权重为 P_9 = (0.10,0.16,0.06,0.43,0.27),其中 CI = 0.0116,得到 CR = 0.010 < 0.1,判断矩阵符合一致性要求。B6 对应的判断矩阵为

$$\begin{bmatrix} 1 & 1/2 & 1/3 & 2 \\ 2 & 1 & 1/2 & 3 \\ 3 & 2 & 1 & 5 \\ 1/2 & 1/3 & 1/5 & 1 \end{bmatrix}$$

,求得的指标权重为 P_{10} = (0.16,0.27,0.48,0.09) 其中 CI = 0.00484,得到 CR = 0.00931 < 0.1,判断矩阵符合一致性要求。

接下来需要对整体一致性进行检验,由于第一层指标均为二阶判断矩阵,完全符合一致性检验要求,因此这里仅对第二层指标做整体一致性检验。第二层对目标层的组合一致性指标为 $CI^2 = \sum_{i=1}^{m^2} w_i^2 \ CI_j^2$ = 0.19 * 0.0614 + 0.09 * 0 + 0.29 * 0. . 0598 + 0.29 * 0. 1500 + 0.05 * 0.0116 + 0.09 * 0.0048 = 0.0735,第二层对目标层的随机一致性指标为 $RI^2 = \sum_{i=1}^{m^2} w_i^2 \ RI_j^2$ = 0.19 * 1.46 + 0.09 * 0 + 0.29 * 1.54 + 0.29 * 1.56 + 0.05 * 1.12 + 0.09 * 0.89 = 1.3125,则 $CR^2 = \frac{CI^2}{RI^2}$ = 0.056 < 0.1,整体一致性检验通过。由上述结果可知,各重要性判断矩阵均可以满足一致性要求,因此不需要对判断矩阵进行调整。

客观权重的计算本文采用的是熵权法。在上一章节,已经对在互联网上收集到的 39 所高校第三层指标的具体数据进行了标准化处理,根据公式 $H_j = -\frac{1}{\ln n}\sum_{i=1}^{n} z_{ij} \ln z_{ij}$ 可以直接计算各个指标的熵值,其中 z_{ij} 表示第 j 所高校第 i 个指标数值标准化后的结果,接着利用 $w_j = \frac{1 - H_j}{1 - \sum_{i=1}^{m} H_k}$ 公式计算得到各个指标的客观权重。

由表 2 中第三层指标的客观权重和主观权重结果可以看到,按照层次分析法得到的主观权重值最大的是的平均每条微博负面评论数量,由于微博具有的强大的号召力和影响力,一些负面的评论会给网络环境下高校的形象带来不好的影响。由熵权法计算得到的网页数量客观权重值是很大的,可见在网页数量这一指标上高校之间的变异程度很大,对于高校网络影响力评价体系来说,即包含的信息量比较大的。因此可以发现,指标的主观权重与客观权重有比较大的差距,为了实现主观权重与客观权重的有机结合,得到更加科学合理的权重指标值,依据基于加法的综合归一化组合赋权法进行组合赋权,计算公式为:

$$w_i = \lambda \ w_i^1 + (1 - \lambda) \ w_i^2 \quad (1)$$

其中 w_i^1 是第 i 个指标的主观权重,w_i^2 是第 i 个指标的客观权重,λ 是主观权重的偏好系数,本文将主观权重与客观权重同等看待,设置 λ = 0.5,得到的各指标的组合赋权值结果如下表所示,通过观察,对于前文提到的网页数量这一指标,虽然客观权重值比较高,但由于根据专家的判断,这个指标对于高校网络影响力评价的贡献不是很大,经过二者的加权求和计算,得到一个较为合理的权重值,说明该组合赋权法是有效的。

表 2 第三层指标的组合权重结果

指标	主观权重	客观权重	组合权重	指标	主观权重	客观权重	组合权重
网页数量	0.0071	0.0493	0.0151	日均发文数	0.0102	0.0026	0.0011
网站总链接数	0.0143	0.0321	0.0198	原创文章比例	0.0070	0.0493	0.0149
内部链接数	0.0038	0.0389	0.0064	活跃粉丝数	0.0184	0.0199	0.0158
外部链接数	0.0269	0.0397	0.0461	平均每篇阅读数	0.0161	0.0132	0.0092
网络影响因子	0.0143	0.0149	0.0092	微信最高阅读数	0.0100	0.0079	0.0034
外部影响因子	0.0314	0.0591	0.0801	平均每篇点赞数	0.0166	0.0186	0.0133

指标	主观权重	客观权重	组合权重	指标	主观权重	客观权重	组合权重
反链接数	0.0381	0.0318	0.0522	微信最高点赞数	0.0100	0.0261	0.0112
PR 值	0.0381	0.0274	0.0451	平均每篇文章的评论数量	0.0433	0.0179	0.0334
DPV 值	0.0167	0.0005	0.0004	微信最高评论数	0.0381	0.0111	0.0183
教育部提及数	0.0390	0.0101	0.0169	平均每篇文章的中性评论数量	0.0130	0.0199	0.0112
科学技术部提及数	0.0552	0.0194	0.0463	平均每篇文章的正面评论数量	0.0329	0.0173	0.0245
粉丝数量	0.0137	0.0281	0.0166	平均每篇文章的负面评论数量	0.0426	0.0246	0.0452
日均微博数	0.0056	0.0167	0.0040	微信传播指数(wci)	0.0269	0.0017	0.0020
平均每条微博转发数量	0.0218	0.0254	0.0239	词条的编辑次数	0.0044	0.0048	0.0009
微博最高转发数	0.0148	0.0403	0.0257	词条的浏览次数	0.0073	0.0075	0.0023
平均每条微博评论数量	0.0508	0.0244	0.0534	词条的长度	0.0026	0.0020	0.0002
微博最高评论数	0.0250	0.0100	0.0108	词条的转发次数	0.0199	0.0469	0.0403
平均每条微博点赞数量	0.0262	0.0273	0.0309	词条的点赞次数	0.0126	0.0271	0.0148
微博最高点赞数	0.0153	0.0498	0.0329	百度新闻报道的数量	0.0257	0.0096	0.0107
平均每条微博中性评论数量	0.0077	0.0201	0.0067	百度新闻报道的正面新闻比例	0.0150	0.0004	0.0003
平均每条微博正面评论数量	0.0339	0.0230	0.0337	百度新闻报道的负面新闻比例	0.0260	0.0309	0.0347
平均每条微博负面评论数量	0.0566	0.0272	0.0666	百度新闻报道的中性新闻比例	0.0462	0.0245	0.0488
微博等级	0.0139	0.0009	0.0006				

五、网络影响力聚类分析

为了更加直观地观察高校网络影响力的差异，接下来结合聚类方法进行分析。聚类分析是一种典型的数据挖掘算法，它可以根据成员之间特征的相似程度将一组样本数据（或变量）分组，而无需任何先验知识[12]。聚类分析的优点是可以综合利用多个变量对样本进行聚类，聚类分析的结果比传统的分类方法更加细致、全面和合理，而且它可以有效地消除由于数据的微小差异而导致的区别，避免了直接排名的限制。传统的聚类算法，k－means 和 ME 算法等前提假设就是样本分布在凸球形的空间里，不能解决非凸（non－convex）数据，而近年来提出的谱聚类就是一种可以在任意形状的样本空间上进行聚类、并且能够收敛到全局最优解的一种新型聚类方法[13]，因此本文采用谱聚类进行高校网络影响力的聚类分析，这种方法的其他优势在于在聚类过程中只需要样本的相似度矩阵，而且使用了降维，在处理类似本文中的高维数据时效果会比传统聚类方法好，其具体步骤为：

输入：样本集 D＝（s_1，s_2，…，s_n），相似矩阵的生成方式，降维后的维度 k_1，聚类方法，聚类后的维度 k_2

输出：簇划分 C（c_1，c_2，…，c_{k_2}）

1）根据输入的相似矩阵的生成方式构建样本的相似矩阵 S

2）根据相似矩阵 S 构建邻接矩阵 W，构建度矩阵 D，D 矩阵为对角矩阵

3）计算拉普拉斯矩阵 L

4）构建归一化后的拉普拉斯矩阵 $L = D^{-\frac{1}{2}} LD^{-\frac{1}{2}}$

5）计算归一化后 L 矩阵的 k_1 个最小特征值及对应的特征向量 f

6）将 k_1 个特征向量 f 组成的矩阵按行标准化，最终组成 $n \times k_1$ 维的特征矩阵 F

7）对特征矩阵 F 按照输入的聚类方法进行聚类，类簇数量为 k_2

8）得到簇划分 $C(c_1, c_2, \cdots, c_{k_2})$

相似矩阵是有样本的相似度计算得到，一般都是采用的欧几里得距离，假设有 m 个属性，属性 j 的权重为 w_j，第 i 个样本属性 j 的值为 v_{ij}，样本 s_{n_i} 和 s_{n_j} 距离计算公式为：

$$d(s_{n_i}, s_{n_j}) = \sqrt{\sum_{j=1}^{m} (v_{n_i j} - v_{n_j})^2} \quad (2)$$

这种方法将所有属性同等看待，会导致计算

得到的两个样本点的距离失真,在重要属性上比较相似但是在无关属性上差别较大的两个点在欧几里得空间上的距离往往会比较远,相当于放大了无关属性的影响。因此,基于属性权重的距离计算方法时很有必要的,可以体现出不同的属性在相似度计算上的贡献,降低聚类过程的模糊性,得到更加准确的聚类结果,本文采用公式(3)来计算距离。

$$d(s_{n_i}, s_{n_j}) = \sqrt{\sum_{j=1}^{m} w_j (v_{nj-} v_{nj})^2} \quad (3)$$

在谱聚类算法在执行之前,需要指定输入到数据集中产生的分类个数,即类簇数量,选择合适的类簇数量是很重要的,它影响到了最终分类结果的准确性与合理性。一般通过各个簇内的样本点到所在簇质心的距离平方和(Sum of Squared Error,SSE)来进行判断,SSE 越小说明各个类簇越收敛。当然 SSE 不是越小越好,因为一种极端情况时将所有的样本点视为簇,此时 SSE 为 0,但是这样显然达不到分类的效果,因此需要在类簇数量与 SSE 之间找一个平衡,肘部法则提供了这样的方法。肘部法则[14]是一种在聚类分析中一致性解释与检验的方法,它将 SSE 解释为类簇数量的函数。首先,指定一个 i 值,即可能的最大类簇数,然后将类簇数从 1 开始递增,一直到 i,计算出 i 个 SSE。根据数据的潜在模式,当设定的类簇数不断逼近理想的类簇数时,SSE 呈现快速下降态势,而当设定类簇数超过理想类簇数时,SSE 也会继续下降,但是下降趋势会迅速趋于缓慢。通过画出 K - SSE 曲线,找出下降途中的拐点,即可较好的确定 K 值。图 1 展示了随着类簇数量增大时 SSE 的变化,可以看到类簇数量取 5 时比较合适,因此本文将 39 所高校的网络影响力数据划分为 5 类,最终得到的分类结果见表 3。

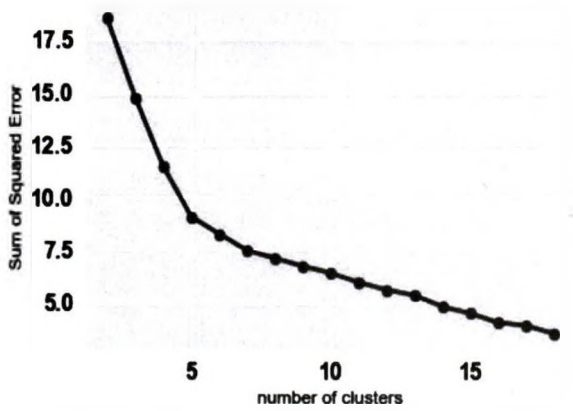

图 1 不同类簇数量对应的 SSE 值

表 3 谱聚类分类结果

类别	成员名称
1	南京大学、华南理工大学、兰州大学、云南大学
2	中国人民大学、北京师范大学、中央民族大学、电子科技大学、西北工业大学、东北大学、西北农林科技大学、华东师范大学
3	北京大学、清华大学、上海交通大学、复旦大学、武汉大学、浙江大学、厦门大学、山东大学
4	北京航天航空大学、南开大学、天津大学、大连理工大学、中国海洋大学、中南大学、中山大学、四川大学、西安交通大学、郑州大学、湖南大学
5	中国科学技术大学、中国农业大学、吉林大学、哈尔滨工业大学、同济大学、东南大学、华中科技大学、重庆大学

第 1 个类别是发展均衡型,所属的 4 个高校综合排名处于中间部分,从表 3 来看,这些学校网络影响力各方面的发展大致是处于相同水平,虽然个别指标上排名比较突出或者落后,但是对整体结果影响不大。南京大学在新闻网络影响力上的排名处于第 31 位,但在其他指标上得分都相对比较高,同时新闻网络影响力指标对整体得分的贡献比较低,因而此高校可以划分为均衡发展类型。发展均衡型的学校需要在稳步发展时有所创新,促进网络影响力的全面提升。

第 2 个类别是全面落后型,包括东北大学、中国人民大学、电子科技大学在内的 8 所高校,综合网络影响力排名位置相对集中,总体均处于整体样本的中下游水平,整体网络影响力不高。除了个别指标外,在多数二级指标上的网络影响力排名都比较靠后,在互联网时代,树立正面的网络形象、扩大网络号召力和影响力有利于帮助高校招生、引进优秀人才和获取好的资源,这些高校亟需从各个方面加强网络信息资源建设,提高网络影响力,发挥其在信息传播和舆论引导上的重要作用。

对于第 3 个类别,属于强势领先型,共有 8 所高校,它们在官方网站影响力、教育类网站上的影响力等 6 个二级指标上大部分具有领先优势,树立了良好的网络形象,尤其是在权重比较大的二级指标上,虽然其中的上海交通大学在微博网络影响力上排名稍微落后,但是由于在一级指标官方平台影响力下的官方网站和政府网站影响力具有明显优势,所以综合来说也是处于领先地位的。

第 4 类是综合落后型,此类高校成员包括北京航天航空大学、南开大学等 11 所高校,数量多。

这几所高校大部分都在新闻网络影响力上具有优势，但是在其他指标上处于劣势，同时新闻网络影响力对高校综合网络影响力的贡献度不高，所以整体网络影响力排名并不理想。这类高校应该将网络信息资源的建设放在重要地位，从各个方面提升网络影响力。

第5类是综合领先型，对比强势领先类别的高校，这8所高校在多数二级指标上的网络影响力都不具有优势，例如华中科技大学在微博网络影响力、百科网络影响力和新闻网络影响力上排名略低，但在权重较大的官网和微信网络影响力上得分具有比较大的优势，因此它的综合网络影响力的排名上是处于领先地位的。这类高校应该注重全面发展，在排名较低的指标上投入更多的时间和精力。

六、小结

高校网络影响力评价有很大的研究空间，但鲜有采用本文所用的加权方法。[16] 本文从官方平台、社交媒体和第三方平台三个方面出发，构建的高校网络影响力三级评价指标体系，不仅涵盖了多种渠道的网络信息资源，而且从“量”和“质”来两个角度对衡量高校的网络影响力，其中“量”体现在官方网站的网页数量、微博关注量、新闻提及量等指标，“质”主要从微博或者微信的文章的评论内容、新闻报道的情感倾向进行体现。希望研究结果年能有助于各高校加强Web信息资源管理工作，提升自身的舆论引导能力和塑造良好的网络形象，保证信息的有效传播，为公众提供信息服务。同时，本文研究尚有许多细节方面值得继续完善，如指标的覆盖性、更大跨度的数据范围等，在后续的研究中，我们将多方深入思考。

参考文献：

[1]中国纪检监察报. 教育是国之大计党之大计[EB/OL]. [2018-12-15]. http://csr.mos.gov.cn/content/2018-09/12/content_67522.htm.

[2]Thelwall, Mike. A comparison of sources of links for academic Web impact factor calculations[J]. Journal of Documentation, 2002, 58(1): 66-78.

[3]邱均平，陈敬全，段宇锋. 中国大学网站链接分析及网络影响因子探讨[J]. 中国软科学，2003(6).

[4]吴茵茵. 中美大学网络影响因子研究[J]. 情报科学，2008(7).

[5]赵杨，宋倩，高婷. 高校图书馆微博信息传播影响因素研究——基于新浪微博平台[J]. 图书馆论坛，2015(1).

[6]于洋洋，袁珈琳，朱周熠，et al. 基于PCA和CA的高校官方微博影响力评价与比较研究[J]. 中国商论，2018(25).

[7]石建华. 我国一流高校图书馆微信公众号影响力评估研究[J]. 河南图书馆学刊，2018，38(5).

[8]吕嘉. 广东省国家级森林公园网络影响力评价研究[D]. 中南林业科技大学，2017.

[9]李敏谦. 古村落网络信息资源评价与村落画像分析[D]. 大连理工大学，2017.

[10]李宇啸. 基于网络分析法的高校微信公众号传播影响力评价研究[D]. 合肥工业大学，2018.

[11]中华人民共和国教育部. 教育部、财政部、国家发展改革委公布世界一流大学和一流学科建设高校及建设学科名单[EB/OL]. [2018-11-25]. http://www.moe.gov.cn/srcsite/A22/moe_843/201709/t20170921_314942.html.

[12]Lucheng H, Yanhua Y. Evaluation on the Industrialization Potential of Emerging Technologies Based on Principal Component and Cluster Analysis[C]// International Conference on Computer Modelling & Simulation. IEEE, 2010: 317-322.

[13]蔡晓妍，戴冠中，杨黎斌. 谱聚类算法综述[J]. 计算机科学，2008(7).

[14]维基百科 Elbow method[EB/OL]. [2019-03-05]. https://en.wikipedia.org/wiki/Elbow_method_(clustering).

[15]许树柏. 实用决策方法：层次分析法原理[M]. 天津：天津大学出版社，1988.

[16]汤建民. 建构和发展我国的评价科学[J]. 西南民族大学学报(人文社科版)，2019(1).

收稿日期 2019-03-20 **责任编辑** 吴 俊