

基于聚合的世界大学综合排名方法研究

吴 俊, 张 洋, 吕 欣

(国防科技大学 系统工程学院, 湖南 长沙 410073)

摘 要: 随着社会对高等教育的需求日益提升, 形形色色的世界大学排行榜应运而生。然而, 各种榜单之间的显著差异使得决策者在运用榜单时困惑不已。在总结当前世界大学排名存在的问题基础上, 提出基于聚合的世界大学排名思想进而提出基于竞争图的世界大学综合排名方法。实证分析结果表明本方法提升了大学排名的准确性和客观性, 可以平滑大学排名中的误差, 而且对各地区的排名更加公正。

关键词: 大学排名; 排名聚合; 高等教育评价; 竞争图

中图分类号: G647 **文献标识码:** A **文章编号:** 1672-8874 (2019) 01-0083-08

On Comprehensive World University Ranking Based on Ranking Aggregation

WU Jun, ZHANG Yang, LV Xin

(College of Systems Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: With the increasing demand for high education, many university rankings have been proposed in recent decades. However, the remarkable divergence among various rankings leads to confusion for decision-makers. In this paper, we propose a comprehensive world university ranking by aggregating existing individual university rankings. We present a new graph-based rank aggregation method by defining a competition graph of universities. The results show that the aggregated ranking increases the accuracy and objectiveness of ranking and can efficiently eliminate the outliers and regional partiality as a “smoother”.

Key words: university ranking; ranking aggregation; higher education evaluation; competition graph

一、引言

随着社会对高等教育的需求日益提升, 形形色色的世界大学排行榜应运而生。特别是在我国政府大力推进“双一流”建设背景下, 大学排名作为一种有效的高等教育质量评价工具正受到社会界、学术界的广泛重视^[1-2]。虽然各国政府一般不做官方排名, 只履行公权对教育基本质量进行监控并促进质量持续改进, 但是各种社会团体和研究机构推出的大学排行榜层出不穷。到目前为止,

维基百科上一共列出了 20 多个不同组织提出的世界大学排行榜, 这些排行榜主体不同、用途各异, 分别从不同的评价理念出发、采用不同的评价方法得到了不同的排名结果, 在国际上影响力各异。

高等教育的质量直接决定了人才培养、教育质量、社会文明的发展水平, 对其进行评价也是高等教育自身的需要。大学排名是一种高等教育质量评价的必然表现, 这种表现是用一种显然甚至突显的方式来进行的, 目前也许在实践中有了缺失、偏差, 但其作为高等教育自身评价的本体意义尚存, 也该是其自身最具生命力之所在。大

收稿日期: 2018-10-22

作者简介: 吴 俊 (1980-), 男, 湖北荆门人。国防科技大学系统工程学院教授, 博士, 博士研究生导师, 主要从事复杂系统与复杂网络研究。

学排名的初衷是提供数据和信息,为各类主体决策提供依据^[3]。然而,随着大学排行榜越来越多,人们越来越困惑,到底哪一个排行榜好呢?到底该参考哪一个排行榜呢?本文将在总结当前世界大学排名存在的问题基础上,提出基于聚合的世界大学排名思想,进而建立基于竞争图的世界大学综合排名方法。

二、当前世界大学排名存在的问题

大学排名在近十年来发展迅猛,不同的机构和组织在提出新的大学排名时也在不断地完善大学排名的体系和方法。但是,大学是个包含学生、教师、教学、就业等诸多方面的复杂系统,以排序作为结果的大学排名从出现就不停地遭到质疑。存在的主要问题主要包括以下几方面:

1. 指标和权重的主观性。当前绝大多数大学排行榜都是首先建立一个指标体系,然后再将指标体系中的各项指标加权求和得出大学的整体得分,最后根据整体得分确定最终排名。然而在这个综合评价过程中,由于各排行榜的价值维度、关注重点、数据获取方式等存在差异,导致指标选取和权重设定不可避免的具有主观性,最终排名缺少必要的效度、信度和可比度。

2. 排名数据的不稳定性。大学是一个相对稳定的机构,提高质量也是一个恒久的过程,从这一视角出发,一般认为大学排名也应该是相对稳定的。但是由于单个大学排行榜是根据其选取的指标基于榜单获取的数据对大学进行排名,这种方式会导致小部分学校的排名变化会产生异常波动。这种不异常波动的产生有两个原因:一是数据获取和处理时产生的误差,二是大学根据排名指标特意造就出的虚假的数据,这种异常不利于排名的稳定性和客观性。

3. 大学排名中的地域偏见性。大学作为国际竞争的“代言人”和国际贸易的“产品”给大学排名赋予了考量国家利益乃至发挥政治作用的意义。某项排名在此种逻辑下揭示的绝不仅仅是大学间的评比结果,甚至大学本身的排名高低并不重要,重要的是大学作为一种国家形象的“代言”应该具备的政治意义^[4]。在这种政治意义下,不同机构产生的大学排名会从设计的评价体制和指标上对一些地区产生优待或者偏见,尽管在原始

数据上看是客观公正的,但是这种偏见是真实存在的并且隐藏在这种公正之下。

三、基于聚合的世界大学排名思想

所谓排名,就是将目标根据其重要性加以排列。现在看来,就研究的活跃程度和研究者的兴趣而言,排名问题的研究进展已经达到了一个顶峰,越来越多的排名方法被提出和完善。但在目前的形势下,我们的聚焦点从提出新的排名方法转变成利用现用的排名通过排名聚合方法来更好的解决问题。

“总体大于个部分之和”这句名言,正是排名聚合方法的基本思想,将单个排名聚成一个新的更好的排名并不是一个新想法,在政治投票领域、网络搜索领域上都已经取得了很好的应用效果^[5-6]。例如,元搜索引擎便是基于排名聚合的思想,将多个搜索引擎给出的排名结果综合成一个超级父列表,该列表保持了单个列表的良好性质^[7]。伴随着大学排行榜的出现,虽然没有哪个排行榜能够“十全十美”,但是每个大学排名都能在某些方面保证一定的客观性,我们可以通过排名聚合的方法产生一个新的排名,“取其精华、去其糟粕”,该排名在保持单个排名的优秀的性质的同时也能够对一些异常的情况进行优化。例如在一个榜单中,出于某些原因,某所大学的排名出现异常,与其他榜单相比有着明显的不一致,这会给榜单的使用者造成一定的影响。而通过聚合排名,这个异常所造成的影响就会被削弱。聚合排名可以起到“平滑器”器的作用,移除各个榜单中的任何异常的排名凸起,并使得排名结果更加平滑。

本文试图通过聚合现有世界大学排行榜获得一个更加科学合理的世界大学综合排名,其意义如下:

1. 聚合排名有助于增加大学排名的客观性和公正性。单个榜单中因为方法和指标以及外界的影响将会产生一些误差和偏见。由于排名使用权重把若干测量结合起来构成一个总体,所以这种误差和偏见仅仅分析单个榜单是很难察觉到的。如果我们分析多个排名,这种差异就会明显的暴露出来。基于聚合的方法如同鸡尾酒一般将不同种类的世界大学排名聚合在一起,将单个排名中因政治因素、数据因素、指标因素导致的偏差通过一个“过滤器”将其稀释到最低,最终获得一个更加客观和公正的榜单。

2. 单一的榜单有助于管理者和学生进行决策。虽然不同的榜单从不同的角度反映了大学的情况,提供了更丰富的信息,但是在大多数决策情况下,众多的参考榜单反而会给决策者带来困惑,尤其是每个榜单都有一定的社会影响力和公正客观性,各个榜单中的不一致性就会给决策者带来困难,而如果仅用一个排名作为标准又有失偏颇。因此,通过聚合的方法产生的唯一的榜单在保证更加客观公正的同时也能够满足决策者的需要。

在排名聚合方法的选择上,我们将创新提出一种基于竞争图的聚合排名方法,用一个有向图来刻画所有排名信息,再利用网络的结构属性获得最终的排名,这种方法较之前排名相加取平均(波达计数法)要精确,并且能够用于不完全信息排名的聚合。

四、基于竞争图的世界大学排名方法

(一) 榜单聚合方法

假设用来聚合的大学排名列表的数量为 N , 出现在列表内的大学数量为 M , 每个大学列表用排名列表集合 $R_i = [r_{i1} \ r_{i2} \ \dots \ r_{im}]$ 来表示, 其中 r_{ii} 为大学 u_i 在排名列表 R_i 中的位置, 如果大学 u_i 未出现在 R_i 中, 则令 $r_{ii} = 0$ 。对于每一个排名列表, 我们都能够基于大学的排名得到一个转移矩阵 $P^i = (p_{ij}^i)_{M \times M}$, 如果大学 u_i 排在 u_j 的前面或者并列, 我们就令 $p_{ij}^i = 1$, 如果其中一所大学未被 R_i 收录, 我们令 $p_{ij}^i = p_{ji}^i = 0$ 。 $P^i = (p_{ij}^i)_{M \times M}$ 表示如下:

$$P^i = (p_{ij}^i)_{M \times M} = \begin{cases} p_{ij}^i = 1 \ p_{ji}^i = 0 & r_{ij} > r_{ji} > 0 \\ p_{ij}^i = 1 \ p_{ji}^i = 1 & r_{ij} = r_{ji} > 0 \\ p_{ij}^i = 0 \ p_{ji}^i = 0 & r_{ij} \times r_{ji} = 0 \end{cases}$$

定义 $A = (a_{ij})_{M \times M}$ 为大学竞争图的邻接矩阵, 当

$\sum_{i=1}^N p_{ij}^i > \sum_{i=1}^N p_{ji}^i$ 时, 令 $a_{ij} = 1 \ a_{ji} = 0$; 当 $\sum_{i=1}^N p_{ij}^i = \sum_{i=1}^N p_{ji}^i > 0$ 时, 令 $a_{ij} = 1 \ a_{ji} = 1$; 当 $\sum_{i=1}^N p_{ij}^i = \sum_{i=1}^N p_{ji}^i = 0$ 时, 令 $a_{ij} = 0 \ a_{ji} = 0$; 根据邻接矩阵, 我们可以得到一个竞争图 $G_c = (V, E)$, 每个节点表示一所大学, 每条边表示两所大学之间的关系, 由相对好的大学指向差的大学。

根据竞争图 $G_c = (V, E)$ 我们可以得到每个节点 i 的入度 $d_i^- = \sum_{i=1}^N a_{ij}$ 和出度 $d_i^+ = \sum_{j=1}^N a_{ij}$ 。节点 i 的

出度越大, 代表着大学 u_i 超过的其他大学越多。在极端的情况下 $d_i^+ = 0$, 意味着大学 u_i 排在聚合列表的最后一位; 节点 i 的入度越大, 代表着该大学 u_i 被越多的其他大学所战胜, 当 $d_i^- = 0$ 时, 意味着大学 u_i 排在聚合列表的第一位。

我们定义出入度比 $\alpha_i = (d_i^+ + 1) / (d_i^- + 1)$ 来表示大学 u_i 的质量并且根据 α 的值对所有的大学进行排序。公式中的项“1”是为了避免出现 $\alpha \rightarrow \infty$ 的特殊情况。对于每所大学, α 的值越大, 大学的质量越高, 排名越靠前。

(二) 榜单评价指标

由于排名方法和指标的差异以及排名机构或者排名者的主观印象, 同一所大学 u_i 在不同的列表中的位置可能会出现很大的差异。在给定的排名列表集合 $R_1 \ R_2 \ \dots \ R_n$ 中, 大学 u_i 的排名分别为 $r_i^1, r_i^2, \dots, r_i^m$, 将该对象所有排名的均值记作 $\bar{r}$, 标准差记作 σ , 定义 z 得分 $\psi_i^t = (r_{ii} - \bar{r}) / \sigma$ 来衡量大学 u_i 在排名列表 R_j 中的异常情况。如果 $|\psi_i^t|$ 远远大于 0 (一般 $|\psi_i^t| > 3$) 我们就可以认定排名列表 R_j 对该大学的排名存在异常。我们将列表中每个对象的 $|\psi_i^t|$ 值相加, 记为 $\Psi = \sum_{u_i \in R_i} |\psi_i^t|$, 以此用来衡量排名列表中异常排名的程度。显而易见, Ψ 越大说明该排名列表中的异常排名越多。

一个排名对象集合内的对象会有许多属性, 根据某个属性可以将对象划分为不同的群体。比如大学排名中不同的大学属于不同的区域, 同一个区域的大学可以划分到一个群体; 在学校教师排名中, 可以按照院系(计算机院、光电学院、管理学院、人文社科院等)将老师划分到不同的群体中。我们以群体作为分析对象, 研究不同的群体在排名中的地位, 定量地分析同一个群体在不同的排名中的排名是否受到刻意的抬高或者拉低。我们用集合 C 表示区域, 如果大学 u_i 属于区域 C , 那么记为 $S \in C$ 。定义指标 Q 来表示区域的大学质量:

$$Q = \sum_{u_i \in C} \frac{1}{r_i}$$

Q 越大, 表明群体的质量越好, 群体中的对象数目相对较多, 排名相对靠前。不同区域在不同的排名列表中有不同的 Q 值, 我们定义 $\bar{Q}$ 为群体 C 所有 Q 值的均值, σ_c 为标准差。为了衡量排名列表 R_j 对群体 C_i 的排名的公正程度, 定义排名列表 R_j 对区域 C_i 的 Z 得分 $\phi_i^j = (Q_i^j - \bar{Q}_{C_i}) / \sigma_{C_i}$ 为了评价排名列表是否对所有的群

体能够公正的排名 我们定义 $\Phi = \sum_{i=1}^N |\varphi_i^j|$ 。显然 Φ 值越小 排名越能公正的评价群体。

五、世界大学综合排名实证分析

(一) 原始榜单数据

我们选取了2015—2016年度5个世界大学排名列表的前200名进行实例分析,一方面选取影响力大,社会认同度高的排名,另一方面也考虑排名机构所在地,尽量选取不同的地区产生的排名来避免地区偏见。以下是对这5个排名的介绍:

USNEWS: 2014年《美国新闻与世界报道》(U. S. News & World Report) 首次发布“US News 世界大学排名(US News Best Global Universities)”,为全球学生选择理想大学提供科学依据。

ARWU “世界大学学术排名”(Academic Ranking of World Universities) 是世界上最早的综合大学排名,于2003年由上海交通大学高等教育研究院首次发布。

QS: QS世界大学排名(QS World University Rankings),是由教育组织 Quacquarelli Symonds 所发表的年度世界大学排名。QS世界大学排名是历史第二悠久的全球大学排名,第一次发布于2004年,仅次于《世界大学学术排名》(2003年)。

THE: 泰晤士高等教育(Times Higher Education)

以教学、研究和国际展望作为标准,为全世界大学排列名次。

URAP: 世界大学学术表现排名(University Ranking by Academic Performance) 是由土耳其中东理工大学(Middle East Technical University) 发布的一项世界排名。

在本研究中,我们仅选取该5个榜单的前200所大学并分析榜单之间的差异。选取前200基于两个理由:一是ARWU仅提供了前200名的详细分数可以计算出精确的排名;二是200名之后的大学之间的差距很小,精确排名已经没有什么很大的意义。在这5个榜单中,一共有308所大学,但仅有118所大学同时出现在全部榜单中。在剩下的190所大学里,有75所是仅出现在一个榜单中的,这种信息不完全的榜单给普通的聚合方式带来极大地困难。如图1所示,对于这118所同时上榜的大学,其不同榜单的排名也有着显著的差异,并且这种差异性随着大学排名的降低变得更加显著,例如香港大学(在ARWU中排名186, QS中排名30)、Queen Mary University London(在ARWU中排名197, QS中排名109)。为了进一步研究榜单两两之间的差异性,我们在图2中展现了5个榜单之间的相关性。从图2中我们可以定性地观察到不同榜单之间的排名有着巨大的差异,某些大学在一个榜单中排在前50而在另一个榜单中排在了150名开外。

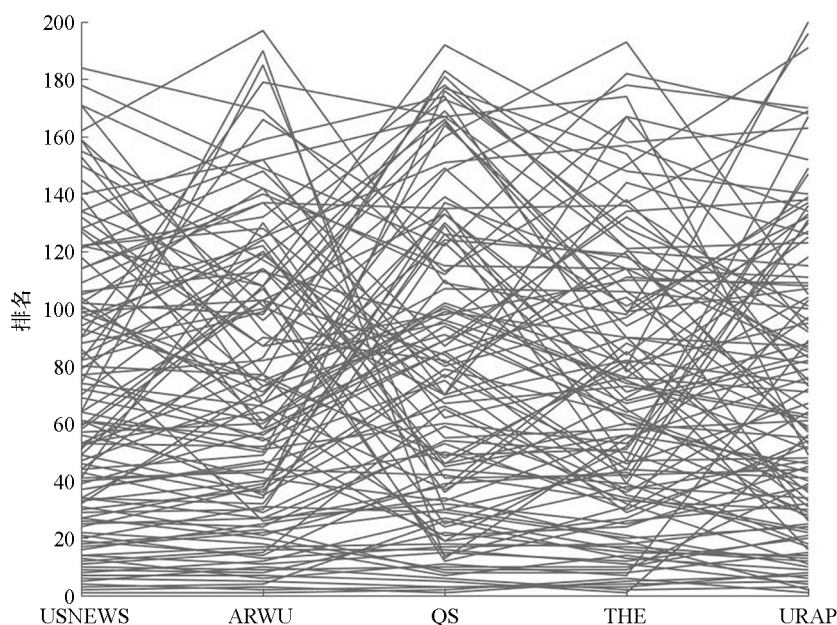


图1 五个榜单排名的差异性

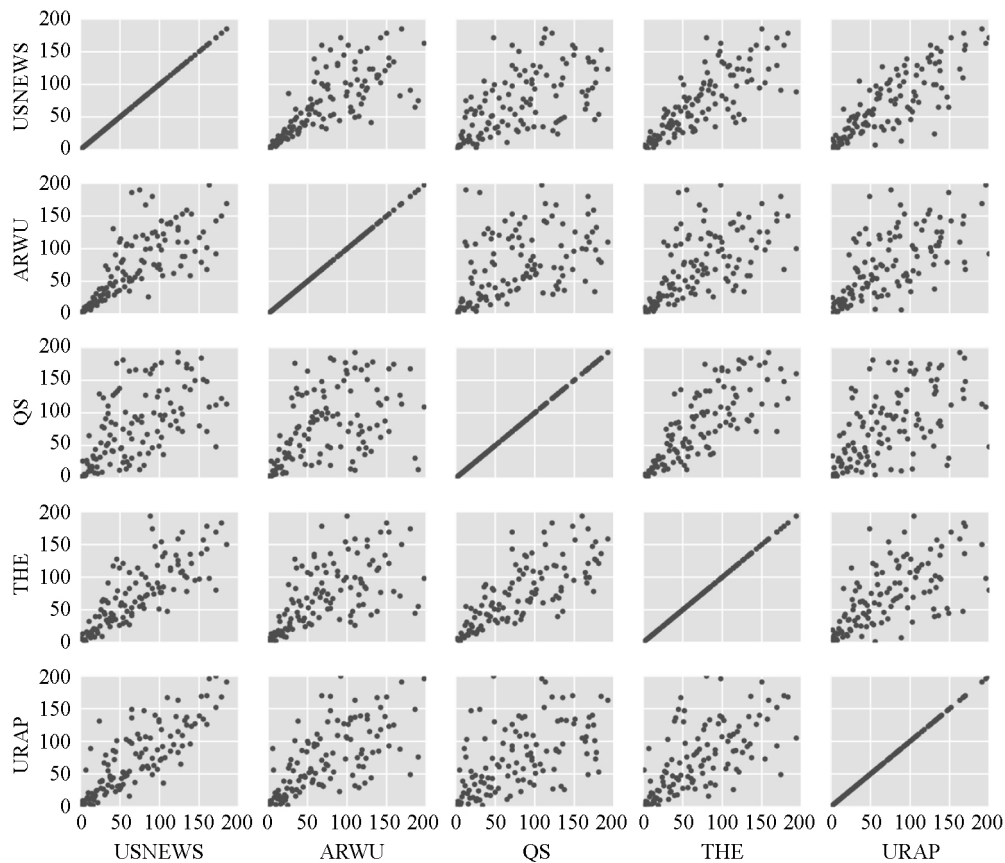


图2 五个榜单的相关性

(二) 综合排名结果分析

我们用基于竞争图的方法将上述的5个榜单进行聚合得到了一个新的排名（Comprehensive world university ranking based on ranking aggregation，简称CWUR）。在图3中我们将结果用竞争图的方式展

现出来，节点越大表示排名越靠前。在表1列出了排名前20的具体信息。在我们的排名中，哈佛大学位居第一，即使它在QS（2）和THE（6）中的排名轻微的偏后，麻省理工大学排名第二，并在QS中排名第一，在其他几个榜单中也排名靠前。

表1 聚合排名 top20

大学	国家	USNEWS	ARWU	QS	THE	URAP	CWUR	α_i
哈佛大学	美国	1	1	2	6	1	1	308
麻省理工大学	美国	2	3	1	5	7	2	153.5
斯坦福大学	美国	4	2	3	3	8	3	102
牛津大学	英国	5	10	6	2	3	4	60.8
剑桥大学	英国	6	5	3	4	5	4	60.8
加州理工学院	美国	7	7	5	1	56	6	50.5
加州大学伯克利分校	美国	3	4	26	13	9	7	43.14
约翰霍普金斯大学	美国	12	16	16	11	4	8	33.33
普林斯顿大学	美国	13	6	11	7	89	8	33.33
芝加哥大学	美国	10	9	10	10	21	10	29.9
哥伦比亚大学	美国	9	8	22	15	14	11	27.09
伦敦大学学院	英国	22	18	7	14	6	12	24.75

续表 1

大学	国家	USNEWS	ARWU	QS	THE	URAP	CWUR	α_i
耶鲁大学	美国	14	11	15	12	20	13	22.77
加利福尼亚大学，洛杉矶	美国	8	12	27	16	12	14	21.07
宾夕法尼亚大学	美国	14	17	18	17	13	15	18.31
伦敦帝国理工学院	英国	18	23	8	8	15	15	18.31
康奈尔大学	美国	21	13	17	18	25	17	17.18
多伦多大学	加拿大	16	25	34	19	2	18	15.26
密歇根大学安阿伯分校	美国	17	22	30	21	10	18	15.26
瑞士联邦理工学院	瑞士	27	20	9	9	39	20	13.71

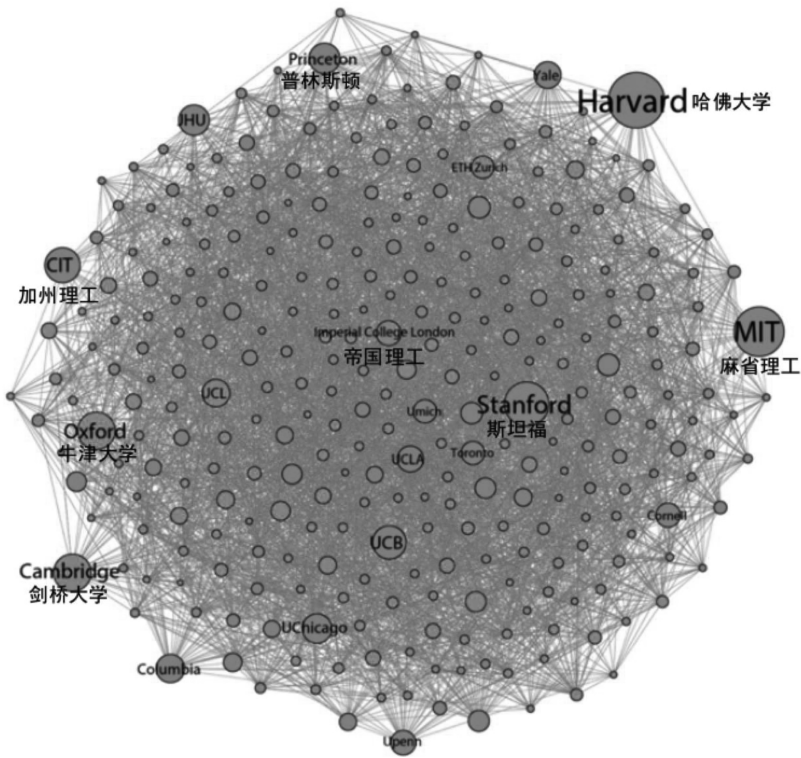


图 3 世界大学排名竞争图

在图 4 中，我们可以清楚的观测到世界大学地理上分布的严重不平衡。在前 100 中有 84 所大学集中在美国和西欧，尤其是在美国的东海岸。一小部分大学零星地分布在东亚和澳大利亚，并且排名在 51—100 之间。而在地域广袤的非洲和拉丁美洲，没有一所大学入围前 100。这反映出了世界各地高等教育上的严重不平衡。

图 5 展现了同时出现在 6 个榜单的 118 所大学在各个榜单的排名异常分布图。可以看出 QS 的排名异常程度最高，异常程度值达到 144.30，榜单中存在大量排名异常的学校。聚合榜单的异常程度最低，异常程度值只有 47.53。值得注意的是，

聚合榜单中也存在极少数的异常现象，这是因为聚合榜单的长度是 308，而其他榜单的长度为 200，这种长度的增加会导致部分异常。

为了衡量榜单的地区差异，我们将世界划分为四个地区：北美（美国和加拿大）、欧洲、亚澳、其他地区，并根据这四个地区将大学分为四类。图 6 展现了这四个地区的大学质量，气泡越大说明该学校的质量越好。从图 6 可以看出，北美的大学质量在每个榜单中都是最高的，接下来是欧洲和亚澳地区。通过横向观察，我们发现北美地区的大学在 USNEWS 和 ARWU 中的排名相对要更好，欧洲的大学在 THE 和 QS 中的排名较其他榜单

也相对更好。亚澳地区的大学在 QS 中也获得更好的排名。换句话讲,发行榜单的机构所在地会对当地的学校有一定的优待。

在表 2 中,我们利用 Z 得分计算各个榜单的地区公正性, $|\phi|$ 越接近 0,榜单对该地区的排名就

越公正,THE 中存在的地区不公正是最多的,这也和其指标中的主观评价有关。URAP 和我们的聚合排名的 $|\phi|$ 都很小,说明这两个排名对待不同地区的大学相对更客观公正。



图 4 聚合排名 Top100 的地理分布图

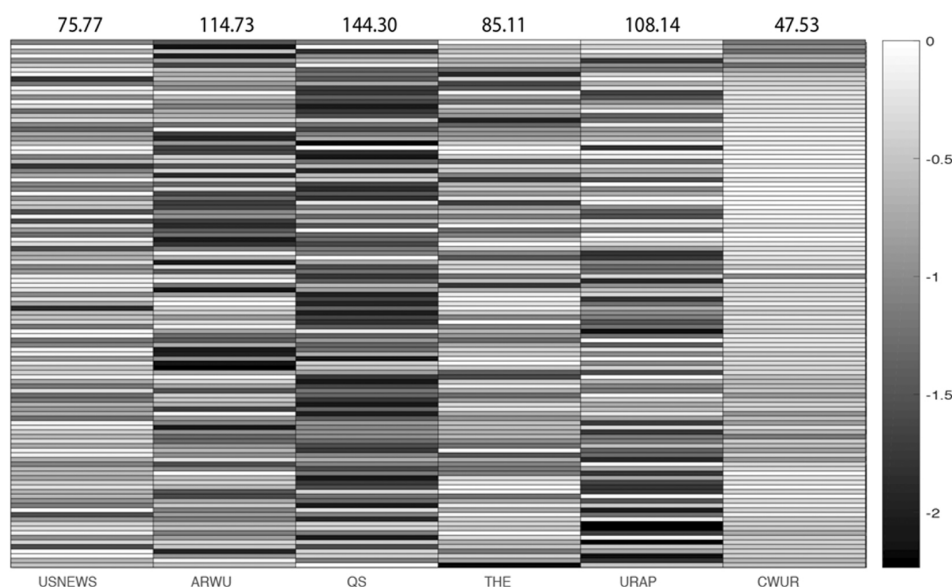


图 5 大学排名异常热图

表 2 榜单地区公正性

榜单	ϕ'_k				ϕ'
	北美地区	欧洲地区	亚澳地区	其他地区	
USNEWS	1.049	-1.027	-0.524	-0.493	3.093
ARWU	1.061	-1.008	-0.761	1.074	3.904
QS	-1.08	0.604	1.668	0.206	3.557
THE	-0.714	1.236	-0.573	-1.454	3.977
URAP	-0.316	0.195	0.19	0.668	1.369
CWUR	0.377	-0.252	-0.127	-0.868	1.624

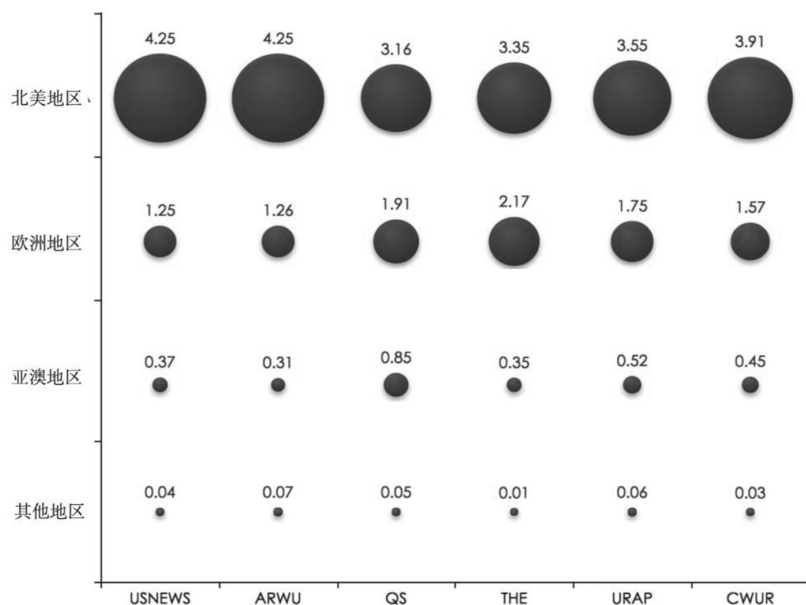


图6 地区大学质量气泡图

六、结语

作为高等教育质量评估保障体系的重要组成部分,大学排名已经在世界范围类广泛开展和运用。大学排名榜在“以评促建”提高大学建设水平、提供高校报考信息、教育教学改革等方面扮演着越来越重要的角色。然而,由于当前单个大学排名榜中存在着诸多问题,排名方法和结果一直存在争议。因此,本文提出了基于聚合的大学排名思想,采用转移矩阵生成竞争图,借助有向网络的出度入度比这一指标对大学进行综合排名。实证分析结果表明本方法提升了大学排名的准确性和客观性,可以平滑大学排名中的误差,而且对各地区的排名更公正。

参考文献:

- [1] 王洪才. 大学排名榜: 现状·困境·展望[J]. 复旦教育论坛 2007(6): 45-50.
- [2] 葛少卫, 杨晓江. 基于网络计量学的大学排名榜影响力研究[J]. 高教探索 2018(1): 28-37.
- [3] 张旺, 华晓玲, 邵艳霞. 大学排名对学生择校和大学招生的影响——以《美国新闻与世界报道》的大学排名为例[J]. 高教探索 2011(5): 44-47.
- [4] 赵锋, 白玫. 大学排名: 批判、存在与发展[J]. 河北师范大学学报: 教育科学版 2016(3): 102-107.
- [5] Xiao Y, Deng Y, Wu J, et al. Comparison of rank aggregation methods based on inherent ability [J]. Naval Research Logistics 2017(6): 556-565.
- [6] Langville A N, Meyer C D. Who's #1?: The Science of Rating and Ranking [M]. Princeton: Princeton University Press 2012: 159-181.
- [7] Dwork C, Kumar R, Naor M, et al. Rank aggregation methods for the Web [C]. Hong Kong: ACM, 2001: 613-622.

(责任编辑: 王新峰)