UNIVERSITY OF PANNONIA

DOCTORAL THESIS

Beyond Tradition: A New Approach to Constructing University Leagues

---

Author:
Vivien Valéria CSÁNYI

Supervisors:
Dr. Péter MIHÁLYI
Dr. Zsolt Tibor KOSZTYÁN

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Doctoral School in Management Sciences and Business Administration
Department of Quantitative Methods

January 30, 2024

Beyond Tradition: A New Approach to Constructing University Leagues

Thesis for obtaining a Ph.D. degree in the **Doctoral School in Management Sciences and Business Administration** of the University of Pannonia

in the field of **Social Sciences**
in the subject of **Management and Business Studies**

Written by: Vivien Valéria CSÁNYI

Supervisor(s): Dr. Péter MIHÁLYI
Dr. Zsolt Tibor KOSZTYÁN

propose acceptance (yes / no) ……………….............
(supervisor/s)

As reviewer, I propose acceptance of the thesis:

Name of Reviewer: …...................... …................ yes / no
……………….............
(reviewer)

Name of Reviewer: …...................... …................ yes / no
……………….............
(reviewer)

The PhD-candidate has achieved …..........% at the public discussion.

Veszprém, ……………………….....
(Chairman of the Committee)

The grade of the Ph.D. Diploma …..................................... (…….. %)

Veszprém,
……………………….
(Chairman of UDHC)

UNIVERSITY OF PANNONIA

Abstract

Doctoral School in Management Sciences and Business Administration
Department of Quantitative Methods

Doctor of Philosophy

Beyond Tradition: A New Approach to Constructing University Leagues

by Vivien Valéria CSÁNYI

Universities and Higher Education Systems (HESs) are often ranked by various well-known organizations. However, recent studies have raised questions as to whether it is fair to compare institutions and countries with different structures. These criticisms have merit, as it is challenging to create a one-dimensional ranking system that can accurately compare complex systems like universities or HESs. This paper introduces university "Leagues," which differ from the existing global rankings leagues.

The main challenge is to define leagues by simultaneously selecting criteria and countries/universities that meet the criteria. In this work, leagues are defined by an unsupervised bi-clustering method, using a set of indicators and a set of countries/universities. The bi-clustering methods are demonstrated on two different data sets: the ranking of Higher Education Systems and the global Round University Ranking of institutions. The top-, mid-, and lower-performing leagues are established based on a given threshold. The proposed set of leagues allows overlapping both on indicators and on universities. Overlapping can help university management determine how to improve their institution's ranking or move up to a higher league.

Resumen

Las universidades y los sistemas de enseñanza superior suelen ser objeto de clasificación por parte de diversas organizaciones de renombre. Sin embargo, estudios recientes han puesto en duda que sea justo comparar instituciones y países con estructuras diferentes. Estas críticas tienen fundamento, ya que es difícil crear un sistema de clasificación unidimensional que pueda comparar con precisión sistemas complejos como las universidades o los sistemas de enseñanza superior. Este documento introduce las „ligas" universitarias, que difieren de las actuales ligas de clasificación mundial.

El principal reto consiste en definir ligas seleccionando simultáneamente criterios y países/universidades que cumplan los criterios. En este trabajo, las ligas se definen mediante un método de biagrupación no supervisado, utilizando un conjunto de indicadores y un conjunto de países/universidades. Los métodos de bi-clustering se demuestran en dos conjuntos de datos diferentes: el ranking de Sistemas de Educación Superior y el Round University Ranking global de instituciones. A partir de un umbral determinado, se establecen las ligas de rendimiento superior, medio e inferior. El conjunto de ligas propuesto permite el solapamiento tanto de indicadores como de universidades. El solapamiento puede ayudar a la dirección de la universidad a determinar cómo mejorar la clasificación de su institución o ascender a una liga superior.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation of the Thesis

Since the first appearance of American universities' ranking in 1983 by the U.S. News and World Report and the first world university ranking by Shanghai Jiao Tong University, several university rankings are published yearly.

Rankings are a widely used tool to simplify complex systems into easily understandable ordered lists, typically classed from best to lowest performance. They enable users to compare and contrast entities that have been ranked and are frequently used to allocate resources according to the achieved rankings (Iñiguez et al., 2022). In their extensive research, Iñiguez et al. (2022) examined 30 ranking lists' evolvement over time from natural, social, economic, and infrastructural systems. As an example, they analyzed the Academic Ranking of World Universities (ARWU) and found that top performers, like Harvard and Stanford, maintain high scores over time, and others, at the middle and bottom of the ranking list, change their rank frequently.

The universities at the top of the rankings have concreted places, mainly because a large portion of the ranking score comes from reputation surveys that hardly change over time (Dill and Soo, 2005; Safón and Docampo, 2020).

Choosing a university solely based on its ranking can be challenging for students that not looking for elite universities. It is worth noting that smaller regional institutions can excel in certain areas, given that the comparison criteria are appropriate. Moreover, a majority of international rankings fail to consider factors such as tuition fees, living expenses, and other associated costs of attending a specific university. These metrics are only implicitly reflected in the rankings through other indicators (e.g. student - staff ratio). Universities positioned at the top of global rankings tend to be more expensive for students in comparison to those situated in the middle or at the lower end of the ranking.

University rankings are heavily criticized from several angles (see, for example, Liu and Cheng, 2005; Daraio and Bonaccorsi, 2017; Soh, 2017; Moed, 2017; Safón and Docampo, 2020; Chirikov, 2022). The author grouped the problems into three main categories that are described in Subsection 2.1.3. One common point of the criticisms is that rankings can not be considered "fair" because they compare entities with highly different input-output structures, sizes, and funding (Lawrence and Green, 1980; Bengoetxea and Buela-Casal, 2013; Daraio and Bonaccorsi, 2017). To address this, the author proposes that only similar institutions (or countries' Higher Education Systems) should be compared to achieve a fairer ranking.

The author considers a ranking "fair" if the compared entities are similar in some nature following the work of Lawrence and Green (1980), Bengoetxea and Buela-Casal (2013), and Daraio and Bonaccorsi (2017). A key aspect of fairness is that not all entities can be compared using the same indicators. Some entities excel in certain indicators, while others perform below average.

This study aims to present a method that can simultaneously choose a subset of indicators and a group of entities for comparison based on these selected indicators. This approach ensures the formed groups are consistent and include entities that share similarities.

There is an ongoing effort (e.g., Downing, 2013; Salmi, 2013) to define different and well-tailored leagues for benchmarking universities or countries instead of ranking them in one group. However, there is no generally accepted method for identifying such leagues.

The author agrees with Benneworth (2010) and Liu (2013) that universities that belong to similar Higher Education Systems (HESs) should be compared according to a given set of criteria that is also in accordance with the common features of the HESs.

The research was motivated by the assumption that university leagues are more useful for potential students than rankings. The reason for this is that the indicators of the rankings and their weights vary from ranking to ranking. This causes problems because students thinking about where to apply presumably do not investigate the reasons for these significant differences in the positions in the rankings. If a university is in the top 100 of one list but the same university is around the 500th place in another ranking, it can discourage students from applying. The author thinks that rankings would not be as popular among students if they knew that they were based primarily on the faculty's research (and not the educational) performance or on how quickly they could achieve their dream job. A counterexample is the Financial Times, whose ranking also considers graduates' salaries.[1] Bell and Brooks (2019) found in the UK that students are more satisfied with universities where the level of research is lower. (Kosztyán et al., 2019) showed on the application data of Hungarian students that the excellence of the faculty (measured by their research performance) played less of a role during the higher education institution selection process during the 2011-2017 period.

This study aims to present a method defining university leagues on a neutral base. "Leagues" (not in the sense the author uses the term) are already used to eliminate the heavily criticized deficiency of global rankings. Those "leagues" are based on the universities' major fields of activity (medicine, business) or other characteristic features (such as size or financial constitution). Such a definition of the scope certainly reduces the incommensurability of the selected items but simultaneously incorporates an ad-hock preselection or uses a specific indicator or feature. Such a choice may be considered an unsolicited preference toward the selected items and a dispreference against those that were omitted.

[1] https://rankings.ft.com/home/masters-in-business-administration

The term "league" in this work is borrowed from English football and signifies a group of teams of similar performance engaged in competitive sports, participating in contests against one another. "Group of teams" in this case are Higher Education Institutions competing for students, resources, funds, and talents, not just on their national field, but on an international level as well to achieve higher and better rankings.

This work presents a method of league creation that is free of ad-hock choices or a suspect of bias. The author demonstrates the usefulness of league creation, finding that the top league is a result of self-reinforcing dynamics. The dynamics resemble Matthew's "the rich get richer" effect.[2] The results show that universities earned their privileged position in the top league, having high scores only in their three reputation-based indicators.

In this work, leagues are specified by an unsupervised bi-clustering method. Leagues are defined simultaneously by a set of indicators and a set of countries/universities. The top-, mid-, and lower-performing leagues are specified based on a given threshold. The proposed set of leagues allows overlapping both on indicators and on universities. The overlaps show university management which indicators should improve the position of their institution in the ranking or permit entering a higher league. The member universities are similar with respect to a number of indicators. Membership in a particular league indicates a set of similar universities to students, i.e., they have comparable conditions and similar strengths and weaknesses.

In the following, for the case of Higher Education Systems (HESs) in countries and a global ranking of Higher Education Institutions (HEIs), the work shows how leagues can be developed as a new basis for comparing HEIs and HESs. The author utilizes existing indicators, recognizing that they may not be entirely bias-free but also acknowledging the significant effort that has gone into

[2]Matthew's "the rich get richer" principle is originally coming from the Bible. "For whoever has will be given more, and they will have an abundance. Whoever does not have, even what they have will be taken from them." (Matthew 25:29).

acquiring and cleansing the data. The author proposes a new method for grouping objects for comparison but does not introduce any new indicators. This approach illustrates the benefits of forming leagues without the confounding effect of new indicators.

It is important to note that creating these leagues demands significant effort and complex calculations. Given that this work aims to form groups where the entities (HEIs or HESs) are comparable, the direct outcome of the bi-clustering method does not yield an exact institutional ranking but leagues with varying numbers of entities. The ranks of entities within these groups are inexact.

Studying countries' entire higher education systems is essential, as higher education is a public good that directly affects a country's economic prosperity (Marginson, 2011). Improving a nation's higher education system can lead to increased labor productivity (Mankiw et al., 1992) and innovation capabilities (Romer, 1990). However, it is also important to recognize differences in the quality of education, as neglecting this can distort the relationship between education and economic growth (Hanushek and Wößmann, 2010). Assessing the quality of education can be done through university rankings, making the analysis of country-level education particularly valuable to the existing literature in this field.

Part of the results presented in this work have already been published in the following international and Hungarian scientific papers:

1. Zsolt T. Kosztyán, Zsuzsanna Banász, Vivien V. Csányi, and András Telcs (2019). "Rankings or Leagues or rankings on Leagues? - Ranking in fair reference groups". In: *Tertiary Education and Management* 25.4, pp. 289–310. DOI: 10.1007/s11233-019-09028-x. URL: `https://link.springer.com/article/10.1007/s11233-019-09028-x`

2. Zsolt T. Kosztyán, Zsuzsanna Banász, Vivien V. Csányi, and András Telcs (2019). "Felsőoktatási ligák, parciális rangsorok képzése biklaszterezési eljárásokkal". In: *Közgazdasági Szemle* 9, pp. 905–931. DOI: 10.18414/KSZ.2019.9.905. URL: `https://ideas.repec.org/a/ksa/szemle/1861.html`

3. Zsuzsanna Banász, Zsolt T. Kosztyán, Vivien V. Csányi, and András Telcs (2022). "University Leagues alongside Rankings". In: *Quality & Quantity* 57.1, pp. 721–736. DOI: 10.1007/s11135-022-01374-0. URL: `https://link.springer.com/article/10.1007/s11135-022-01374-0`

4. Zsolt T. Kosztyán, Zsuzsanna Banász, Vivien V. Csányi, László Gadár, and András Telcs (2020). "Egyetemi rangsorok tudománymetriai és statisztikai megalapozással". In: *Statisztikai Szemle* 98.8, pp. 930–957. DOI: 10.20311/stat2020.8.hu0930. URL: `https://www.ksh.hu/statszemle_archive/all/2020/2020_08/2020_08_930.pdf`

This work further extends the above-mentioned papers by discussing the problems of university rankings in more depth, focusing more on the results related to Hungary. Additionally, it introduces new insights derived from biclustering analysis applied to the latest ranking datasets.

In the following, the research questions and research proposals are described.

Research Questions

RQ1: Are universities comparable "fairly" based on an arbitrarily predefined set of ranking indicators?

RQ2: Is it possible to create homogeneous groups that contain entities (countries or institutions) that has above-average, below-average, and similar performance?

RQ3: Is it possible to determine a distinct set of indicators that specifies the entities' (countries or institutions) potential for development, leading them towards an above-average performing group?

RQ4: Are there any indicators that clearly identify entities (countries or institutions) belonging to the above-average performing group?

Research Proposals

P1: To make a fair comparison of universities, it is important to ensure that the entities being compared are similar in nature. This means that not all entities can be evaluated using the same set of indicators.

P2: The clustering method of bi-clustering can be used to create university leagues that simultaneously select the countries/universities and the set of indicators.

P2.1: The iterative Binary Bi-clustering of Genes (iBBiG) method can be used to determine the above-average performing group of entities (countries or institutions) and their common set of indicators, and the below-average performing group and their shared set of indicators.

P2.2: The Bi-clustering Analysis and Results Exploration (BicARE) can be used to determine those entities (countries and institutions) that have the same performance regarding the set of indicators selected by the method.

The dissertation is structured in the following manner. The literature review, which centers around university rankings, is presented in Chapter 2. Chapter 3 describes the ranking datasets used during the analyses, then Chapter 4 explains the methodology that was employed to establish university leagues. The findings are outlined in Chapter 5. Chapter 6 presents the contribution to the existing literature, and also gives insights on how the leagues evolved over time. Chapter 7 provides a summary of the research, and Chapter 8 outlines the implications of the research for other researchers and the private sector.

The next chapter reviews the scientific literature related to the research topic.

Chapter 2

Related Studies

2.1 University Rankings

University rankings serve several purposes, including assisting students in choosing universities, helping universities improve their performance, promoting competition, and increasing the visibility of institutions. For students, rankings can be a useful tool for comparing universities and making informed decisions about where to study. Employers can also benefit from rankings by assessing the quality of graduates from different universities, which can be helpful when making hiring decisions (Hazelkorn, 2009a).

These rankings can incentivize universities to strive for excellence and compete with one another, leading to an overall improvement in the quality of education (Marginson, 2007; Hazelkorn and Ryan, 2016). High rankings can also help universities attract top students, faculty, and funding opportunities, as well as increase their reputation and prestige. Rankings can also provide feedback on their strengths and weaknesses, helping them identify areas for improvement and adjust their strategies accordingly (Hazelkorn, 2009a).

Rankings can be a valuable tool for policymakers to assess the performance of their country's higher education system and make informed policy decisions (Hazelkorn, 2009a; Hazelkorn and Ryan, 2016). By reviewing them, they can gain insights into how their country's colleges compare to those in other countries and identify areas where improvement is needed. Policymakers can use this information to set goals and allocate resources to improve the quality and reputation of their country's universities (Marginson, 2007).

Monitoring the progress over time and benchmarking their country's performance against other countries can help to identify trends and track changes in the higher education system (Marginson, 2007). This information can be useful in setting targets, evaluating policies, and making data-driven decisions.

Different types of university rankings are published by various organizations and use different methodologies, criteria, and weightings to assess universities and colleges worldwide. Subsection 2.1.1 summarizes all these different types of rankings, and Subsection 2.1.2 presents the various indicators that they use.

University rankings have become increasingly influential in shaping the higher education landscape globally. However, there are several critiques and problems associated with these rankings that question their validity and usefulness. The shortcomings and critiques are summarized in Subsection 2.1.3.

2.1.1 Types of University Rankings

An increasing number of countries publish national rankings of their educational institutions at the secondary and tertiary levels. These rankings are prepared by research institutions based on orders of the governments or commercial actors (e.g., newspapers, non-governmental organizations) (Dill and Soo, 2005). On the national level, universities are more comparable because the national field has common characteristics for every university. One can find numerous national ranking tables, such as The Complete University Guide in the UK, the Center for Higher Education (CHE) University Ranking in Germany, the Maclean's University Rankings in Canada, Perspektywy University Ranking in Poland, or the CYD in Spain. However, rankings on the national level alone do not allow for comparing universities in a global space.

In addition to these national rankings, there are global rankings of higher education institutions (see Table 2.1). The best-known - and probably the most influential - systems are the Academic Ranking of World Universities (ARWU), often called the Shanghai Ranking; the Times Higher Education World University Ranking (THE); the World University Rankings by Quacquarelli Symonds (QS); the Leiden Ranking by the research institute Centre for Science and Technology Studies (CWTS) at Leiden University; and the U-Multirank.

Besides these rankings, several other rankings exist. One of them is CWUR which is published by the Center for World University Rankings located in the United Arab Emirates. Round University Ranking (RUR) publishes Academic World University Rankings and Reputation World University Rankings as well. The first one reckons the level of research performance of leading world universities, meanwhile, the second one assesses the teaching and research reputation of the institutions. Another example is the U.S. News & World Report (USNWR) which publishes not just the Best Global University Ranking, but also the Best Colleges rankings which only consider American HEIs.

One of the main issues with university rankings is that they use different sets of indicators which makes the comparison of the results more challenging (see Subsection 2.1.3 for more details). This phenomenon inspired not just this work, but other independent organizations as well, to create a more fair, objective ranking method.

Academic Influence claims that they "engineered an innovative and unbiased ranking technology that employs machine learning to measure the impact of work produced by the world's top institutions and academics" (Academic Influence, 2023a). Their engine is called "InfluenceRanking" which evaluates the influence of academics and HEIs. To get the institutional ranking, the influence of academics at a particular institution is calculated and cumulated. These scores then are normalized and ranked in order to get the university ranking (Academic Influence, 2023b).

Another example of the ranking world whose aim is close to this study is the Eduniversal ranking. Although they focus on the best 1000 Business Schools, they aim to not generate a vertical, but a horizontal ranking within 9 geographical regions (Eduniversal, 2023b). The schools are selected based on both quantitative and qualitative criteria, and then, for each country, the chosen schools are divided into five levels of excellence they call the "Palmes of Excellence" (Eduniversal, 2023c). The five levels are the following:

- 100 schools in the 5 Palmes League - Universal Business Schools with strong global influence;
- 200 schools in the 4 Palmes League - Top Business Schools with significant international influence;
- 400 schools in the 3 Palmes League - Excellent Business Schools with reinforcing international influence;
- 200 schools in the 2 Palmes League - Good Business Schools with strong regional influence;
- 100 schools in the 1 Palmes League - Business Schools with considerable local influence.

The third and last step is called "The Deans' Vote" (Eduniversal, 2023a). This step contradicts one of the aims of the Eduiversal ranking, which is not to generate a vertical ranking of institutions."The Deans' Vote" is a kind of assessment made by the peer HEIs where Deans and Directors are allowed to vote to create the final ranking within each League (Eduniversal, 2023c). This step is similar to reputation surveys as it involves experts giving feedback on institutions' performance.

QS has developed a system to distinguish between the quality of institutions, known as the QS Star rating system. This system assigns a rating of 0 to 5+ Stars to institutions based on various criteria. This rating system can assist students in making more informed decisions, while institutions can utilize it as a marketing tool to enhance their visibility (QS, 2023c).

By employing the rating system, institutions can identify their areas of weakness and receive a comprehensive evaluation for each criterion. The five core criteria used by QS include Research, Academic Development, Teaching, Employability, and Internationalisation (QS, 2023e). A leaflet has been published by QS, summarizing the requirements that universities must fulfill to attain a 5 Star rating (QS, 2023d). According to (QS, 2023d), the following criteria must be met by an institution to achieve this rating:

- Overall score of 700 out of 1000 points;
- 5% of faculty should be international;
- 5% of students should be international;
- At least 70 points in the Learning Environment category;
- At least 85 points in the Employability category;
- 150 academic references of three citations per faculty member - if assessed in the Research category;
- 105 points in the teaching category - if assessed in the Academic Development category.

Universities can hit the 5+ Star rating if they hit everything in the 5 Star category and have an overall point of 900 out of 1000.

According to QS website, they claim: "we rate universities rather than rank them. To make the rating process as straightforward and fair as possible, QS Stars methodology is based on several critical categories that assess universal, core strengths" (QS, 2023e). However, it is worth noting that the website does not explicitly specify how the thresholds mentioned earlier are determined.

Both the QS Star system and the Eduniversal Palmes of Excellence employ a subgroup approach to categorize institutions, enabling a more equitable comparison among universities compared to a global university ranking.

The idea of creating leagues (subgroups) is in line with the aim of this work, but instead of choosing the institutions and countries' higher education systems arbitrarily, an objective method is used to determine the leagues.

TABLE 2.1: Groups and Examples of Ranking Systems

		Ranking level	
		University	**HES**
Territorial coverage	***Global***	ARWU; THE; QS; CWTS; U-Multirank; ARTU; Academic Influence; CWUR; RUR; USNWR; Eduniversal	U21 (2012-2020); QS (2016, 2018); Lisbon Council (2008)
	Regional	QS & THE: Latin-America; QS: EECA; QS: BRICS; USNWR: US Regions	The scope of this study
	National	UK: The Complete University Guide; US: USNWR, Forbes, CCAP's Rankings; JP: THES; DE & AT: CHE; PL: Perspektywy University Ranking; ES: CYD; CA: Maclean	Cannot be interpreted

Most of these systems focus on universities. However, several initiatives (see, e.g., Salmi, 2013; Hazelkorn, 2015) have suggested that the excellence of tertiary educational institutions should be improved on the national level instead of the institutional level. The researchers who are of this opinion seek to measure the indicators of the HES as a whole. Hazelkorn (2015) sought to develop a "world-class system" rather than "world-class universities". These proposals are only theoretical. However, three practical efforts have been mentioned and developed by the following organizations: Lisbon Council, QS, and Universitas21 (U21). The first was a one-off venture. The Lisbon Council ranked 17 European OECD countries in 2008 based on six fields (inclusiveness, access, effectiveness, attractiveness, age range, and responsiveness), the use of which could measure the ability of their HESs to help citizens and society meet the genuine challenges of a 21st-century knowledge economy (Ederer et al., 2009). Additionally, in 2016, the QS published the "National System Strength Rankings", for which THE data were used in addition to their QS dataset. Their overall rank was determined using four fields: system strength, access, flagship institution, and economic context (Hazelkorn, 2015). In 2016, a similar country-based ranking on their HES was published as "QS Higher Education System Strength Rankings" (QS, 2016) which was repeated in 2018 as well (QS, 2018).

The ranking of U21 is the most ambitious of the initiatives mentioned in the previous paragraph, according to Hazelkorn (2015)'s statements. U21 was considered a novelty for the year 2012 in a report on rankings by the European University Association, although the positions of some countries were considered arguable. A methodological modification was recommended by Rauhvargers (2013) to refine the U21 ranking because the use of ARWU scores “strengthens the positions of big and rich countries whose universities are strong in medicine and natural sciences." (Rauhvargers, 2013, pp.14).

In this work, the author uses the U21 ranking to create leagues of higher education systems. One of the main reasons for applying U21 data is that the indicators are available; therefore, leagues of countries can be specified. The U21 rankings have been published annually since 2012 until 2020, and U21's methodology is one of the most transparent. For the U21 ranking, 2014 was the last year with the relatively lowest number of missing values in which every indicator was available. Therefore, this work focuses on the 2014 country ranking.

In addition to national and global rankings, regional rankings can be specified (see Table 2.1). To date, regional rankings are referred to as rankings within geographic regions, e.g., continents. Excellent examples include the Latin America and Asia University Rankings of QS[1] (Sowter et al., 2017) and THE,[2] and the

[1] https://www.topuniversities.com/university-rankings

[2] https://www.timeshighereducation.com/world-university-rankings

Arab Region University Rankings of QS. Similarly, U.S. News classifies their regional US rankings into four regions: North, South, Midwest and West.[3] Exceptions exist, which rank universities not only by geography but by economic factors: EECA (Emerging Europe and Central Asia) and BRICS (five major emerging national economies: Brazil, Russia, India, China, and South Africa) University Rankings of QS, the Young University Rankings, and BRICS & Emerging Economies University Ranking of THE.

In addition to geographic- and economic-based regional rankings, scholars (e.g., Jarocka, 2012; Abankina et al., 2016) recommend clustering universities to identify similar groups of similar universities and thereby determine the profiles of institutions and identify the directions of development. Nevertheless, those papers did not rank universities after clustering them.

Besides arbitrary classification, clustering methods are used to separate clusters (see, e.g., Rad et al., 2011; Ibáñez et al., 2013). Ibáñez et al. (2013) clustered public universities in the area of computer sciences into four groups based on their productivity, visibility, quality, prestige, and internalization. However, clustering alone cannot be used to specify regional or other rankings because, beforehand or in parallel, clustering indicators should be selected for ranking similar universities or countries (Poole et al., 2017).

The rankings as mentioned above are published by independent organizations (see overview in Table 2.2), such as the Shanghai Ranking Consultancy which publishes the Shanghai Ranking, or the Center for Higher Education Policy Studies which is one of the leaders of the consortium which makes the U-Multirank (Moed, 2017). There is also an international organization called the International Ranking Expert Group (IREG) dealing with - inter alia - approving global university rankings.

IREG aims to shed light on the range of issues related to rankings and help the public to better understand them. Besides being a guide, as (Brankovic et al., 2022) refers to it, it is also a "watchdog" because of its auditing procedure

[3] https://www.usnews.com/best-colleges/rankings/regional-universities

TABLE 2.2: Publishing Organizations of Global Rankings

Publishing Organization		Their Global Ranking	
Name	**Headquarter**	**Name**	**Abbr.**
ShanghaiRanking Consultancy	Shanghai, China	Academic Ranking of World Universities (ARWU)	ARWU (Sanghai)
U.S.News World Report	Washington, US	Best Global Universities	USNWR
SRG S.L./ Scimago Lab	Granada, Spain	Scimago Institutions Rankings / Higher Education	SIR
Times Higher Education	London, UK	World University Rankings	THE
Quacquarelli Symonds (QS)	London, UK	World University Rankings	QS
Round University Ranking Agency	Moscow, Russia	World University Rankings	RUR
Informatics Institute of Middle East Technical University	Ankara, Turkey	University Ranking by Academic Performance	URAP
Center for World University Rankings	Ras el-Kheima, United Arab Emirates	World University Rankings	CWUR
Centre for Higher Education (CHE)	Gütersloh, Germany	World University Rankings	U-Multirank
Eduniversal Evaluation Agency	Paris, France	Best Business Schools	Eduniversal
Education Access LLC's partner	Denton, US	Global University Ranking	AcademicInfluence

for ranking organizations. Rankers can go through the "IREG Seal of Approval Process" (Hägg and Wedlin, 2013; Brankovic et al., 2022; IREG, 2023) to get certified which means that their ranking method is in line with the Berlin Principles. By today, only five rankings earned the "IREG Approved" status (IREG, 2023). Three national ones: Perspektywy University Ranking (Poland), CHE University Ranking (Germany), and Russian University Ranking (Russian Federation); and two international: QS World University Rankings, National Ranking of Higher Educational Institutions of the Republic of Kazakhstan. One can wonder, why only a handful of rankings have undergone the audit so far. As Brankovic et al. (2022) points out, in the early years of IREG the audit process was advertised at their events, and as years went by, and IREG became a fully independent organization, this ambition gradually vanished.

2.1.2 Indicators Used in University Rankings

In order to create rankings, a range of indicators are used to assess different aspects of universities, and these indicators vary with each ranking. From an average user's point of view, it is challenging to fully understand the distinctions between what rankings capture with their various set of indicators.

This subsection aims to highlight the common and dissimilar areas and to give an overview of the most widely used indicators that appear in most of the rankings through the eight most popular global university rankings. Figure 2.1 shows the indicators of the selected eight rankings[4] categorized by the author.

Rankings are usually a mixture of indicators measuring the following five

[4]ARWU: Academic Ranking of World Universities, also knowns as the Shanghai Ranking `http://www.shanghairanking.com/ARWU-Methodology-2019.html`, CWTS: Centre for Science and Technology Studies, also called the Leiden Ranking `https://www.leidenranking.com/information/indicators`, CWUR: Center for World University Rankings `https://cwur.org/methodology/world-university-rankings.php`, SIR: Scimago Institutions Rankings `https://www.scimagoir.com/methodology.php`, THE: Times Higher Education `https://www.timeshighereducation.com/world-university-rankings/world-university-rankings-2020-methodology`, QS: Quacquarelli Symonds `https://www.topuniversities.com/qs-world-university-rankings/methodology`, RUR: Round University Ranking `https://roundranking.com/methodology/methodology.html`, and URAP: University Ranking by Academic Performance `https://www.urapcenter.org/Methodology`

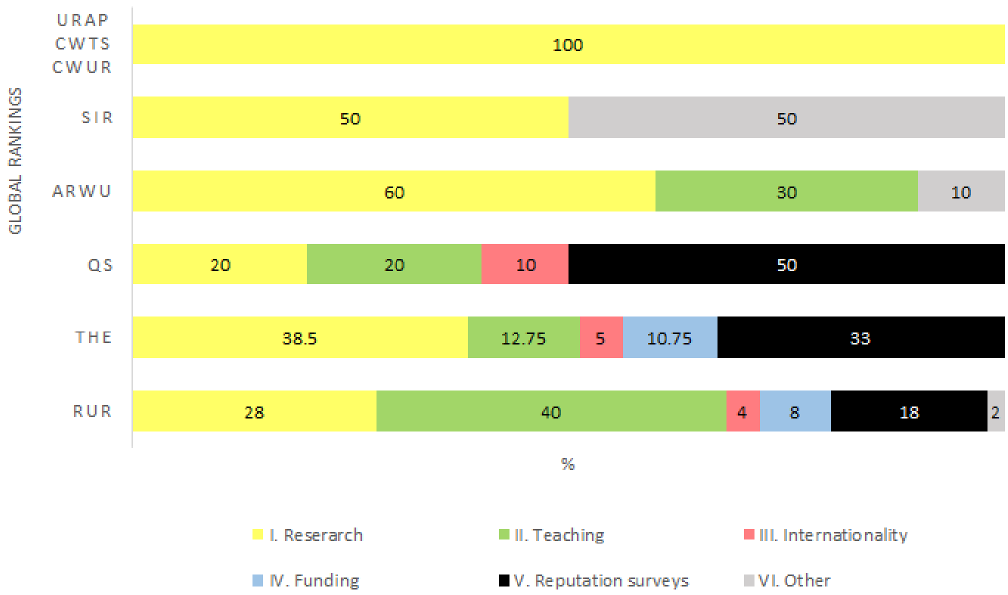


FIGURE 2.1: Indicator Categories for 8 Global Rankings and their Weights

areas: I. research, II. education/teaching, III. internationality/international outlook, IV. funding, and V. reputation. As Figure 2.1 shows, the proportion of the five categories varies between the examined rankings. URAP, CWTS, and CWUR purely measure research activities, while QS, and RUR use indicators from all five areas. Another point that is worth mentioning is the different proportions of the reputation measures amongst the global rankings. While half of the total score of QS is related to reputation, THE uses 33%, and RUR only employs 18%. The flaws of these reputation surveys-based indicators are discussed in Subsection 2.1.3.

Table 2.1.2 describes the indicators and weights of the five main areas. In the cases of ARWU and RUR, the average of certain indicators from categories I-IV was taken. Indicators related to publications and citations were classified into the research category. Research activity is usually measured by the number of research papers and by the number of citations in proportion to academic and research staff. ARWU uses the Highly Cited (HiCi) Researchers indicator to measure the quality of faculty. Every year, Clarivate publishes a list of the most influential researchers based on the highly cited papers (means the top 1%) over ten years.[5]

Publications made through international collaboration appear as separate variables in several rankings. These could have been grouped into the research category, but instead, the author decided to categorize them into the internationalization group. This main group also contains the proportion of international students and faculty members.

Quality of teaching and education is measured by different ratios of the number of academic staff to students at certain tertiary ISCED[6] levels. ARWU and CWUR also employ the number of alumni and staff of an institution winning Nobel Prizes and Fields Medals. Furthermore, CWUR measures the number of alumni who have held top positions in major companies. The size of a "major company" is relatively measured against the size of a given institution. CWUR uses this indicator to measure employability.

Funding is defined as any indicator that examines income, such as institutional or research income or industry earnings.

In addition to these hard indicators, some rankings consider the results of questionnaires (reputation surveys) as soft indicators. Therefore, these surveys were classified into a separate category (V.). The reputation survey of QS measures two types of prestige: academic reputation (40%) and employer reputation (10%). Respondents of the survey can be previous respondents from earlier surveys, from contact lists submitted by the institutions, sign-ups on QS' sign-up facility, or from the International Book Information Service (IBIS)[7] Worldwide Academic and Library database which is one of the leading sources of academic marketing data (QS, 2023a; QS, 2023b).

Respondents are asked to nominate up to ten institutions from their territory (country and knowledge) that they think are producing leading research in the scope of their faculty. They are also asked to provide a list of up to thirty HEIs

[5] In 2022 the list consists of 6938 researchers. https://clarivate.com/highly-cited-researchers/. See methodology details: https://clarivate.com/highly-cited-researchers/methodology/#methodology

[6] International Standard Classification of Education

[7] https://www.ibisacademic.com/about-us/

outside of their country within their knowledge area - in this case, their own institution is excluded (QS, 2023a). In the case of employer reputation, the method is almost the same. They are asked to provide a list of ten HEIs that they think are the best for producing graduates, and thirty international HEIs (QS, 2023b).

TABLE 2.3: Indicators and their Weights in Selected Global University Rankings

	Global University Ranking					
Indicators	**ARWU**	**THE**	**QS**	**RUR**	**URAP**	**CWUR**
I. RESEARCH	**60**	**36**	**20**	**26**	**85**	**40**
Number of research papers; Number of publications by academic and research staff	20	6		10	49	30
Number of citations by academic and research staff	20	30	20	16	36	10
Highly Cited (HiCi) Researchers	20					
II. EDUCATION/TEACHING	**30**	**12.75**	**20**	**40**		**60**
Academic staff to student ratio		4.5	20	8		
Academic staff to Bachelor's ratio				8		
Doctorate to Bachelor's ratio; Academic staff to Bachelor degrees awarded		2.25		8		
Doctoral degrees awarded; Doctoral degrees to admitted PhD				8		
Doctorates awarded to academic staff ratio		6		8		
Number of staff or alumni winning Nobel Prizes	30					35
Number of alumni who have held top positions at major companies						25
III. INTERNATIONAL OUTLOOK		**7.5**	**10**	**6**	**15**	
Proportion of international student or faculty		5	10	4		
Proportion of publications written in international collaboration		2.5		2	15	
IV. FUNDING		**10.75**		**8**		
Institutional income by staff; Institutional income by students		2.25		4		
Research income by academic staff; Research income by students		6		4		
Industry income (knowledge transfer)		2.5				
V. REPUTATION SURVEY		**33**	**50**	**18**		
OTHER	**10**			**2**		

The scientific merit of survey-based indicators is always questionable. Nevertheless, there are several highly prestigious rankings, such as QS or THE rankings, that mainly apply survey-based indicators, while others, such as Leiden's Rankings, do not use survey-based indicators. This study highlights that these indicators have an important role in ranking, which explains, among other things, why some rankings differ so much (see e.g. Bowman and Bastedo, 2011, and summarized in Figure 2.1). The RUR also uses survey-based indicators to a lesser extent. In addition, the qualitative and quantitative indicators are much more balanced in the case of RUR versus other widely used ones.

2.1.3 Critiques Raised against Rankings

The higher-education-related rankings suffer from numerous "deadly sins" as Soh (2017) calls them. Following the work of Daraio and Bonaccorsi (2017), the author grouped these issues into the following three main categories:

- Data and indicator-related problems;
- Methodology-related issues;
- Impact and implication of university rankings.

Each of the above-mentioned categories is further broken down into subcategories and summarized in Table 2.4, Table 2.5, and Table 2.6 with a brief description. Data and indicators-related and methodology-related issues are strongly tied together, and there is no fine line between some categories.

One problem is derived from the fact that global university rankings do not consider the different disciplinary/field compositions of institutions. Most universities are internally diverse, with different missions and staff compositions (Liu and Cheng, 2005; Charon and Wauters, 2007; Bengoetxea and Buela-Casal, 2013), which makes the institutional-level comparison problematic (Daraio and Bonaccorsi, 2017; Bengoetxea and Buela-Casal, 2013). In recent years, subject rankings have appeared next to global rankings - see, for example, the QS World

University Rankings by Subject or the THE World University Ranking by Subject.

The HEIs provide part of the data used in rankings to the ranking agencies. Ishikawa (2009) shows the struggles of a Japanese research university to become a "world-class" institution. During the data-providing process for QS, they often faced definitions-related problems. Ishikawa (2009) mentions "International students" as an example where the outcome number heavily depends on the definition. In Japan, international students are defined based on visa status. Still, after asking for clarification from QS, they were instructed to include all non-Japanese nationals in the international category - such a measure was not available for them then.

This problem arises not only when an institution needs to provide data but also for students when checking different rankings to get a picture of an institution's quality. Moed (2017) compares five popular rankings and shows that even though indicators in different rankings have very similar names, there is no correlation between them. His results suggest that QS Faculty-Student Ratio compared with THE Student-Staff Ratio, and QS International Faculty, compared to U-Multirank International Academic Staff, have very little in common. The way rankings define "staff" and "academic staff", whether they include professors or all researchers, can lead to substantial differences in the ranking results (Charon and Wauters, 2007).

A common approach for rankers is only to consider the extreme top data quantiles, such as Nobel prize winners, papers in Nature and Science, or highly cited researchers (HiCi). This approach leads to not measuring quality but HEIs capability to attract top scientists (Bonaccorsi and Daraio, 2008).

Bibliometric indicators also account for a large part of the overall score of institutions. Missing institutional names, affiliations, mergers, and splits can cause identification problems (Charon and Wauters, 2007; Liu and Cheng, 2005; Frey and Rost, 2010). Hospitals or research units without institutional mention present a thorny challenge. Some do not wish to include their names or do not allow hospital staff to publish papers (Liu and Cheng, 2005). University-owned hospitals - also called academic hospitals - fall into a different category as they are heavily pushed to publish papers in high impact factor journals, and it is also common that one paper has thousands of authors (Abed et al., 2022). One further critique usually raised when dealing with rankings is that rankers emphasize hard sciences more than Humanities. Publications in Human and Social Sciences are underrepresented, partly because there is an imbalance in the production of articles (Liu and Cheng, 2005; Saisana et al., 2011). Citation indicators have several other problems that are not examined further since they are out of this work's scope. See an excellent overview of Frey and Rost (2010) that lists six significant shortcomings.

World university rankings are biased towards a small group of institutions. They favor old research-intensive universities with long ranking histories that use English language (Dill and Soo, 2005; Charon and Wauters, 2007; Bengoetxea and Buela-Casal, 2013; Boyadjieva, 2017). Rankers claim that their rankings are global. However, Moed (2017) shows that ARWU is biased towards North America, THE towards Anglo-Saxon countries, and Leiden towards emerging Asian nations. As Bengoetxea and Buela-Casal (2013) points out, only 2-3% of HEIs are listed; smaller, lesser-known, more diverse institutions are left out.

This leads to the problem that they do not consider the institutions' embeddedness into their unique systems and fail to account for the input-output relations as well (Bonaccorsi and Daraio, 2008; Daraio and Bonaccorsi, 2017). As Daraio and Bonaccorsi (2017) argue, rankers should consider that institutions act as strategic units, and they combine their available resources to produce their output (e.g., teaching, research). Different output mixes can be produced; if they produce more from one, they may produce less from another.

The ranking organization arbitrarily chooses the indicators and weights used in rankings (Hrubos, 2014). The weight values can greatly impact the outcome, and this fact often remains unnoticed (Becker et al., 2017). Furthermore, the chosen weights lack any theoretical foundation, and users assume that weights are maintained as specified (Dill and Soo, 2005; Lukman et al., 2010; Soh, 2011a; Soh, 2014). Soh (2011a) uses the example of the 2010 ARWU ranking to examine whether the effects of weights. The original ARWU ranking's methodology states that "Staff winning Nobel Prizes and Field Medals" is worth twice as much (20% of overall score) than "Alumni winning Nobel Prizes and Field Medals" (10% of overall score). Regression analysis' standardized coefficients (beta-weights) show that Staff's contribution to the overall score is about 24 times than Alumni's. Both Soh (2011a) and Soh (2014) conclude that assigned (nominal) weights and actual (attained) weights differ, thus leading users to misinterpret the ranking results.

Another stream of criticism is related to the deterministic settings of the rankings. They create an ordinal ranking by solely considering the mean of the distribution of indicators (Bonaccorsi and Cicero, 2016; Daraio and Bonaccorsi, 2017). The indicators are aggregated into composite metrics ignoring the underlying attributes, just averages taken from distributions (Bonaccorsi and Cicero, 2016). This leads to a crucial point that the difference between universities in rankings might be statistically indistinguishable from zero (Saisana et al., 2011; Bonaccorsi and Cicero, 2016). This "spurious precision", as Soh (2011b) and Soh (2017) refer to it, biases the ranking users to believe that two adjacent universities have significant differences, even if there is only a minute difference in the second decimal of their overall score.

Further problems arise because rankings are robust only at the top and less reliable at the bottom (Dill and Soo, 2005; Saisana et al., 2011). Saisana et al. (2011)'s uncertainty analysis shows that ARWU ranking is more robust than THE. Still, none of them should be used to compare the performance of individual entities because the assigned weights are very sensitive to the underlying methodology.

The aggregation formulas that rankers use are also questionable. Tofallis (2012) walks the reader through different normalization and aggregation techniques and shows that the chosen method significantly impacts the outcomes.

Tofallis (2012) suggests that rankings should consider multiplicative models instead of additive aggregation methods because that would make weight interpretation easier. Bonaccorsi and Daraio (2008) argues that since universities have three different missions, their indicators can not be summed since these three terms are not additive.

Rankings claim to measure the institution's quality, but there is a debate on which indicators can be used. Van Dyke (2005) and Marginson and van der Wende (2007) point out that different ranking systems use different indicators for measuring quality. Van Dyke (2005) examined ten rankings containing 72 different indicators, and she concluded that no single indicator is used by all rankings. Marginson and van der Wende (2007, pp. 319) argues, "there is no necessary connection whatsoever between the quality of teaching and learning and the quantity and quality of research".

Last but not least, one main problem with university rankings is the heterogeneity of institutions which is also the scope of this work. Several authors argue that entities should not be compared if they have differences in size, funding, and budgets (Dill and Soo, 2005; Guarino et al., 2005; Charon and Wauters, 2007; Marginson and van der Wende, 2007; Saisana et al., 2011; Bengoetxea and Buela-Casal, 2013; Daraio and Bonaccorsi, 2017). For example, in 2021, Harvard University's annual budget was approximately \$5.2 billion, whereas Hungary's annual budget for tertiary education was approximately \$2.9 billion in 2021 (Eurostat, 2021). Rankings compare institutions such as Harvard with significantly smaller HEIs. Moreover, in 2006, the 16 Berlin Principles on Ranking of Higher Education Institutions stated that rankings must specify the linguistic, cultural, and economic contexts of the institutions (IREG, 2006) so users can better understand and interpret the results.

Daraio and Bonaccorsi (2017) also defines the principles of "fair" comparison. First of all, the compared entities should have similar input structures. Secondly, the trade-off between outputs should be explicitly recognized. Thirdly, a higher ranking should be associated with higher performance.

One possible solution to eliminate the biases stemming from the heterogeneity of HEIs is to create regional and subject rankings. Regional rankings consider the ranked entities' socio- and economic embeddedness, while subject rankings narrow down the leagues where universities compete with each other. Subsection 2.2 deals with potential resolutions to eliminate the unfairness of ranking systems. This work's primary aim is to propose a method that can ensure a more fair comparison of the entities. The method is called bi-clustering and is introduced in Chapter 4.

University rankings have had noticeable effects on institutions' policy and behavior since Shanghai Jiao Tong University (SJTU) published the first global ranking in 2003 (Marginson and van der Wende, 2007).

Several authors claim rankings are ideological and neoliberal (Daraio and Bonaccorsi, 2017; Charon and Wauters, 2007). They transform higher education into a market sector, and new managerial techniques emerge, putting more importance on market results than social outcomes (Amsler and Bolsmann, 2012; Lynch, 2014; Daraio and Bonaccorsi, 2017). Hazelkorn (2009a)'s and Hazelkorn (2009b)'s research shows, based on an international survey (2006) and interviews of German, Australian, and Japanese higher education leaders and faculty (2008), that 63% of respondents took strategic, managerial, and organizational actions in response to rankings. With the appearance of rankings, competition between universities started to arise. Their first and foremost aim is to become "world-class" and to attract top scientific talents to secure their high positions in global rankings (Altbach, 2003; Altbach, 2006; Hazelkorn, 2009a; Hazelkorn, 2009b; Shin and Harman, 2009). National policies are tailored to improve their HEIs' places in the most prestigious rankings. For example, China's *Project 985* is aimed at enhancing the performance of its leading institutions to attain a "world-class" status, as highlighted by Török and Nagy (2020) and Török and Nagy (2021). This global competition enhances vertical differentiation between HEIs that may lead to flattening national typologies, resulting in more unitary national systems (Marginson and van der Wende, 2007).

TABLE 2.4: Shortcomings of University Rankings - Data and Indicators

Data/indicators-related critiques		
Main issues	**Brief description**	**References**
Heterogenity of data	*Mixed disciplinary composition of universities*	Johnes (1989), Liu and Cheng (2005), Charon and Wauters (2007), Bengoetxea and Buela-Casal (2013), Daraio and Bonaccorsi (2017), and Török and Konka (2020)
Definitions of indicators	*Definitions are varying between rankings, also different at HEIs*	Johnes (1989), Ishikawa (2009), Moed (2017), Török and Konka (2020), and Chirikov (2022)
Extreme top quantiles of data used only	*Nobel prize winners, Nature and Science publications used to measure quality*	Marginson and van der Wende (2007), Bonaccorsi and Daraio (2008), and Saisana et al. (2011)
Bibliometric shortcomings	*Not exact measures (e.g. Hospitals); research departments with missing affiliations*	Johnes (1989), Charon and Wauters (2007), Frey and Rost (2010), and Shin and Harman (2009)
More weight on hard sciences vs. Humanities	*Publications in Human and Social Sciences are under-represented*	Liu and Cheng (2005), Charon and Wauters (2007), Ishikawa (2009), Bengoetxea and Buela-Casal (2013), and Boyadjieva (2017)
Fail to account for input-output relations	*Rankings are mostly based on output data with only limited indication of inputs; HEIs embeddedness into institutional system failed to taken into account*	Johnes (1989), Bonaccorsi and Daraio (2008), Bengoetxea and Buela-Casal (2013), Boyadjieva (2017), Daraio and Bonaccorsi (2017), and Török and Konka (2020)
Biased toward a small group of universities	*Favors large old universities, most of the indicators are measures of past performance*	Johnes (1989), Charon and Wauters (2007), Marginson and van der Wende (2007), Marginson (2009), Bengoetxea and Buela-Casal (2013), Olcay and Bulu (2017), and Daraio and Bonaccorsi (2017)
Biased for English language	*Citation measures are based on publication in English language; native papers are excluded resulting in a bias against non-English speaking world*	Liu and Cheng (2005), Charon and Wauters (2007), Marginson and van der Wende (2007), Ishikawa (2009), Shin and Harman (2009), Bengoetxea and Buela-Casal (2013), and Olcay and Bulu (2017)

TABLE 2.5: Shortcomings of University Rankings - Methodology

Methodology-related critiques		
Main issues	**Brief description**	**References**
Arbitrary weights	*Lack of theoretical foundation*	Guarino et al. (2005), Lukman et al. (2010), Soh (2011a), Soh (2014), Hrubos (2014), and Becker et al. (2017)
Deterministic settings	*Only considers the mean of the distribution of indicators; small statistically not significant variation in data causes differences in rank outcomes*	Dill and Soo (2005), Guarino et al. (2005), Marginson and van der Wende (2007), Saisana et al. (2011), Soh (2014), and Bonaccorsi and Cicero (2016)
Only robust in the top positions	*Large uncertainty intervals; less reliable in the middle and bottom of the rankings*	Dill and Soo (2005) and Saisana et al. (2011)
Heterogeneity of institutions	*Differences in size, funding, budgets are not taken into account*	Dill and Soo (2005), Altbach (2006), Marginson and van der Wende (2007), Bonaccorsi and Daraio (2008), Ishikawa (2009), Bengoetxea and Buela-Casal (2013), Olcay and Bulu (2017), and Boyadjieva (2017)
Aggregation formulas	*Various formulas and procedures; the three mission of the universities are not additive terms*	Bonaccorsi and Daraio (2008), Frey and Rost (2010), Saisana et al. (2011), Tofallis (2012), Soh (2014), and Becker et al. (2017)
Measuring quality	*Quality can not be measured solely based on the used indicators*	Guarino et al. (2005), Charon and Wauters (2007), Marginson and van der Wende (2007), Shin and Harman (2009), Saisana et al. (2011), and Bengoetxea and Buela-Casal (2013)

TABLE 2.6: Shortcomings of University Rankings - Impact

Impact-related critiques		
Main issues	**Brief description**	**References**
Rankings are ideological and neoliberal	*Transforms HE into a market sector; rankings are used as marketing and public relation tools*	Charon and Wauters (2007), Lynch (2014), and Daraio and Bonaccorsi (2017)
More importance on market results	*Social outcomes are ignored*	Dill and Soo (2005), Amsler and Bolsmann (2012), and Daraio and Bonaccorsi (2017)
Reinforces social hierarchies	*Becoming a "world-class university" is the main purpose*	Altbach (2003), Altbach (2006), Liu and Cheng (2005), Marginson and van der Wende (2007), Hazelkorn (2009a), Hazelkorn (2009b), Shin and Harman (2009), Török and Nagy (2020), and Török and Nagy (2021)
Affects the perception of performance	*Reputation is not based on prior performance but on achieved ranking*	Liu and Cheng (2005), Guarino et al. (2005), and Altbach (2006)
Distorts the institutional policy, enhances vertical differentiation	*Institutions are losing their mission diversity, thus distorting the national hierarchy as well*	Marginson and van der Wende (2007), Teichler (2008), Marginson (2009), Ishikawa (2009), Shin and Harman (2009), Bengoetxea and Buela-Casal (2013), and Boyadjieva (2017)
Halo-effect	*Well-known institutions generate Halo-effect*	Dill and Soo (2005), Altbach (2006), Marginson and van der Wende (2007), Saisana et al. (2011), and Safón and Docampo (2020)
Conflict of interest	*Institutions are paying rankers for guidance, branding, and marketing tools*	Jacqmin (2021) and Chirikov (2022)

HEIs are losing their mission diversities as they shape their strategies to keep up with other institutions in the global space. Merging departments and institutions, and establishing new programs, especially in English languages, to attract more international students and faculty. Since most of the global rankings measure research, HE leaders realized "research matters more now, not more than teaching necessarily, but it matters more right now" Hazelkorn (2009b, pp. 8).

Besides research quality measures, university rankings use reputational surveys to assess institutions' quality. As Altbach (2006, pp. 2) points out, it is a "popularity contest", asking academic groups' opinions about peer institutions. The problem is that well-known HEIs easily generate Halo-effect. Raters assign their positive perception of high reputation to HEIs, departments, and faculties (Safón and Docampo, 2020). At the same time, usually, these people do not know the quality of all university programs accurately; their judgments are based on the existing reputation of an institution. (Dill and Soo, 2005). This "reputation-ranking-reputation" - as Safón and Docampo (2020, pp. 2202) refer to it - circle heavily influences rankings, such as ARWU, affecting students' and decision-makers' opinions.

If one knew about the biases mentioned earlier, one still thinks that university rankings provide impartial information for the users. In reality, rankers do not just rank universities but also provide them with consulting, analytics, and advertising services (Chirikov, 2022). Examining the effect of advertising on university rankings is not a widely researched area yet; it only has a few empirical pieces of evidence, primarily due to the lack of high-granularity data on rankings (Jacqmin, 2021; Chirikov, 2022). Jacqmin (2021) examined the THE ranking with a fixed-effect identification strategy and found that advertising in the printed version of THE magazine is associated with better rankings — the rank of the universities that paid advanced fifteen ranks.

Chirikov (2022) analyzed the QS ranking and concluded the same. QS offers universities a fee-based rating system that evaluates them from 0 to 5+ stars indicating the institution's quality. The number of stars appears on the website next to the university's name, and students can also search and compare institutions based on the number of stars (QS, 2023c). The results of Chirikov (2022) show that the paid star system is associated with 140 positions advancement.

Both Jacqmin (2021) and Chirikov (2022) note that their findings have limitations. It is hard to quantify the advertisement's effect since ranking agencies also provide consultancy that can help institutions change and adapt their strategy, which can also lead to better-ranking positions.

In summary, while university rankings face substantial criticism for their limitations, various studies emphasize that their absence would result in a lack of meaningful comparison, transparency, and accountability in the academic landscape (Mihályi, 2020; Fábri, 2020).

The following section provides an overview of studies suggesting alternative methodologies, arguing that the current hierarchical rankings should be reconsidered to ensure a more impartial comparison of institutions.

2.2 Leagues or Rankings

At the beginning of the twenty-first century, developed economies transitioned from manufacturing and mass production-based economic systems to knowledge-based economies. In response to this process, the number of higher education institutions also started to increase, and students' participation in higher education also experienced a sudden boom. This phenomenon has placed more emphasis on measuring the quality of institutions. It has led to the need for classification systems that can differentiate between universities regarding their missions and specializations (Borden and McCormick, 2020). In recent decades, several classification systems have been developed, such as the Carnegie Classification in the U.S. or the U-Map/U-Multirank in Europe (see an excellent overview in Borden and McCormick (2020, Table 1.).

Initially, elite (top-tier, world-class) universities were given separate league names, such as the Ivy League in the U.S., the Russell Group in the U.K., or the Group of Eight (Go8) in Australia. In addition to these historically established elite universities, several countries have attempted to form elite groups of universities using "do it yourself" systems, which, in Germany, for example, seems to be successful so far (Vogel, 2016). By now, in addition to the elite league, the leagues of other institutions have also been given various notable names. The Complete University Guide uses the following leagues besides the Russell Group: the Cathedrals Group, GuildHE, MillionPlus, Unaffiliated, and University Alliance.[8] Table 2.7 shows some rankings that use the term "league".

TABLE 2.7: Leagues of RUR, former URAP and U-Multirank

RUR 2020 n=829		URAP 2017-8 n=2500		U-multirank 2020 n = 1528[1]	
League	Rank	League	Rank	Group	Value
Diamond	1-100	A++	1-108	A	Me+25% <value
Golden	101-200	A+	109-258	B	Me <value ≤ Me+25%
Silver	201-300	A	259-517	C	Me-25% <value ≤ Me
Bronze	301-400	B++	518-1015	D	0 <value ≤ Me-25%
Cooper	401-500	B+	1016-1501	E	value = 0
World	501+	B	1502-2261		
		B-	2262-2500		

n: number of institutions ranked
[a]"Universities of science and technology rankings" out of the 6 readymade rankings of the U-multirank

Among the rankings listed in Figure 2.1, only RUR and URAP introduce this classification. The names of the RUR leagues are similar to those of the Olympic medal system. URAP displayed the leagues on their website until their 2017-2018 ranking.[9] Their names are reminiscent of country ratings by Standard & Poor's or Fitch (Genc and Basar, 2019).

U-Multirank[10] uses the term group instead of league and labels universities from A to E. U-Multirank is different from other rankings since it is a multi-dimensional user-driven ranking. Multidimensionality means that it does not create composite indicators. It is also user-driven since users can select the indicators for ranking certain institutions. Furthermore, instead of publishing the

[8]https://ukstudyoptions.com/uk-university-groups-a-quick-guide/
[9]For example, https://urapcenter.org/cdn/storage/PDFs/vhKaHMN3xHTJiFjNT/original/vhKaHMN3xHTJiFjNT.pdf
[10]https://www.umultirank.org/ U-Multirank did not appear in Figure 2.1 because they create six different types of readymade rankings.

rank of universities, U-Multirank rates institutions by grouping them into five categories. These five broad categories are from A to E, where A is very good and E is weak (Kováts, 2015). These categories[11] can also be considered leagues.

Table 2.7 shows that the leagues in these rankings were developed based on the universities' ranks or overall value. The author finds it problematic to rank all universities according to all indicators and then classify them into leagues based on that ranking (as shown in Table 2.7). This is similar to awarding an Olympic gold medal based on the overall performance in all sports. Just as it is worthwhile to award medals by sport at the Olympics, it would be advantageous to determine the ranking of universities within leagues. To put it another way, Real Madrid does not engage in football matches against a country team.

Interpreting this in the case of universities, Table 2.8 provides an example of very different operating conditions for universities. Table 2.8 compares the total operating revenue of Harvard University (A)[12] and Hungary's entire national higher education budget (B)[13] for the past years (2014-2022). For conversion between currencies, the annual average of daily central exchange rates of Hungarian Central Bank was employed.[14] Harvard's data were approximately 12-60 times higher than the national data from Hungary (Banász, 2019). With this example, the author does not want to suggest that only money matters, but it matters greatly. Török and Konka (2020) further observes that while comparing Hungarian universities to Serbian or Austrian counterparts may be relevant, such comparisons lose significance when comparison is extended to institutions in the United Kingdom or Ethiopia.

According to van der Wende (2008), global rankings favor research-intensive universities, but there are other types of universities. "If it is absolutely necessary to rank institutions, care must be exercised to compare similar institutions"

[11] `https://www.umultirank.org/export/sites/default/press-media/documents/Rank-Group-Calculation-2018.pdf`

[12] `https://finance.harvard.edu/annual-report`

[13] All items of The Closing Accounts Acts of Hungary, which included the phrase "higher education" `http://kfib.hu/hu/torvenyek-zarszamadasok`

[14] `https://www.mnb.hu/arfolyam-tablazat?devizaSelected=USD&deviza=rbCustom&datefrom=2014.01.01.&datetill=2018.12.31.&order=1&customdeviza%5B%5D=USD`

TABLE 2.8: Budget of Harvard University vs. Hungary

		year								
		2014	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
Harvard University (A)	*in million USD*	4,409	4,526	4,777	4,999	5,215	5,213	5,373	5,249	5,836
Hungary	*in million HUF*	48,121	37,251	41,161	23,196	51,457	54,990	34,534	64,115	187,179
Exchange rate	*HUF[1]/USD*	233	279	281	274	270	291	308	303	373
Hungary (B)	*in million USD*	207	134	146	85	191	189	112	212	502
Harvard / Hungary (A/B)	*rate*	21	34	33	59	27	28	48	25	12

[a]Hungarian Forint

(van der Wende, 2008, p. 67). The author proposes the classification of universities based on their mission and characteristics.

As Kováts (2015) points out, institutions do not have to be good in all indicators, only in those that align with the institutions' strategies and policies.

The fairness of university rankings has been questioned since they first appeared (Marginson, 2009). According to Daraio and Bonaccorsi (2017), the comparison of institutions can be considered fair if (1) the universities have a similar input structure; (2) during the measurement, the trade-off between the outputs is fully and clearly expressed; and (3) a higher ranking means a better performance. This means that universities or higher education systems with different sizes or funding shall not be compared (Bengoetxea and Buela-Casal, 2013). Lawrence and Green (1980, p. 3) also notes that "if comparisons must be made, they should be made between similar types of institutions". Hrubos (2010) and Hrubos (2014) also argue that the currently existing and used rankings, which rely on a hierarchical approach, should be reconsidered, or at least it should not be treated as the only solution. Other scholars suggest using different grouping algorithms to avoid comparability problems and create homogeneous university groups. A summary of these papers can be seen in Table 2.9.

The common point in these papers is that they all emphasize the need for fairer contrast. Valadkhani and Worthington (2006) examined thirty-six Australian universities over the period 1998-2002. Using hierarchical clustering on research-related indicators (PhD completions, publications, and grants), they showed that two clusters are optimal. One of the clusters contains universities from the Go8, while the other cluster has institutions with a lower level of performance. Nolle (2010) and Jarocka (2012) used existing rankings as a basis of

TABLE 2.9: Leagues by clustering

Citation	The subject of the investigation				Clustering method used to form the leagues	Leagues found
	ranking	year	cases	indicators		
Valadkhani and Worthington, 2006	-	1998-2002	36 Australian universities	7	hierarchical cluster analysis, factor analysis	higher-performing, lower-performing
Nolle, 2010	Sunday Times University Guide	2010	144 UK universities	8	self-organizing map, k-means	best-performing, middle-performing, old universities, worst-performing
Jarocka, 2012	ARWU	2011	101 universities	6	k-means	cluster 1-5, 1 is the best
Barnett and Moher, 2019	-	2016-2017	app. 750 universities	1	Bayesian clustering model	cluster 1-5, 1 is the best
Johnes, 2018	The Complete University Guide	2018	129 UK universities	10	data envelopment analysis (DEA)	tier 1-6, 1 is the best category
Raponi et al., 2016	-	2009-2010	55 Italian Economics faculties	24	bi-clustering	public, private

clustering. Nolle (2010) analyzed the 8 indicators of the Sunday Times University Guide, which contains 144 universities from the UK. The author identified four groups of institutions. The 101 universities of ARWU were examined by Jarocka (2012). Using k-means, five clusters were found, and in the first cluster, there was just one university (Harvard University). In the other clusters (No. 2-5), 5, 5, 27, and 63 institutions were assigned, respectively. Barnett and Moher (2019) examined approximately 750 universities from around the world based on the number of publications. The authors assigned 4,408 papers with 47,876 author affiliations to the institutions and compared the results with the World University Rankings. (In the author's opinion, it was not a useful research aim to compare the rankings and publications because, as shown in Figure 2.1, THE also considers teaching, internationalism, and funding, not only publications.) They set the number of clusters a priori to five. The first cluster contains universities with top performance, while the fifth cluster has institutions with lower performance. Johnes (2018) proposed an alternative to rankings by using data envelopment analysis (DEA) to create tiers of universities. The author identified six tiers of institutions using the "peeling the DEA onion" method based on the paper of Barr et al. (2000). The findings show that in the first two tiers, universities have a very similar performance across the examined indicators; they have the highest average values. In contrast, in the fifth and sixth tiers, the original rank of the institutions varies between 76 and 129 (out of 129).

The paper of Lepori (2021) differs from the above-mentioned studies; instead of using a university ranking, he worked with the enriched version of the European Tertiary Education Register (ETER). He classified over 2,000 institutions into 6 classes using latent class clustering among two main dimensions: research vs. educational orientation and subject specialization. He did not intend to rank the universities but categorize them into meaningful classes. The results help to differentiate and distinguish the European HEIs into several categories instead of the two most commonly used categories (research-oriented vs. educational-oriented).

The first class of the six classes contains HEIs that are top-ranked in international rankings, such as Cambridge and Oxford. Class 2 has science and technology-oriented institutions, while in Class 3, most HEIs focus on the applied sciences. Class 4 has generalist HEIs that are middle-sized, multidisciplinary universities. Institutions that specialize in the social sciences and humanities are in Class 5, and Class 6 contains purely educational institutions with no research or technology output.

Raponi et al. (2016) used productivity, teaching, research, and internationalization indicators of 55 Italian economic faculties to create biclusters. They found two different clusters based on the nature of the institutions. One of the clusters contains public universities, while the other cluster has private universities. Their results help to better understand the strengths and weaknesses of the faculties, and they suggest that their approach could potentially serve as the foundation for a multidimensional framework, offering institutions guidance on areas requiring improvement.

The author does not agree with forming leagues based on overall rankings, such as those shown in Table 2.7; instead, recommends that rankings should be formed only within leagues containing similar universities. It is also recommended to use league-based rankings rather than rank-based leagues. Therefore, this work proposes unsupervised bi-clustering methods to simultaneously create leagues that specify indicators and universities.

The method of bi-clustering is most widespread in bioinformatics. It also has much potential within the social sciences, as it can be used to define leagues (see for example, for countries based on their competitiveness indicators Petrarca and Terzi (2018) or Dolnicar et al. (2012)). Within the subject of university rankings (as previously mentioned concerning the last two lines of Table 2.9), Raponi et al. (2016) applied this method to the data of 55 Italian faculty of economics concerning the academic years 2009-2010.

In the upcoming chapter, the data sources are described that were used as a base of the analyses. After Chapter 3, the bi-clustering method is explained in detail, along with the steps of analysis that are required to form leagues.

Chapter 3

Data sources

In this section, a detailed description of the two datasets utilized for the analysis to create leagues is provided. The first dataset used is the renowned Universitas21 (U21) ranking, which assesses and ranks higher education systems of countries based on various factors. The second dataset used is the global ranking of Round University Ranking (RUR), which ranks universities based on their performance across four key areas, including teaching, research, international diversity, and financial sustainability.

3.1 U21 - The Ranking of Countries' Higher Education System

The U21 rankings of countries by their HESs (Williams et al., 2012; Williams et al., 2013; Williams et al., 2014; Williams et al., 2015; Williams et al., 2016; Williams et al., 2017; Williams and Leahy, 2020) are developed at the University of Melbourne. In what follows, the paper presents the evaluation of the U21 rankings and their indicators in details. The U21 rankings cover 9 years (2012-2020) and 50 countries. The rankings for a given year are published in May of that year. Forty-eight countries were examined in 2012, and Saudi Arabia and Serbia were added in 2013. Table 3.1 summarizes countries in order of ranking for the year 2014.

The overall U21 rank scores are calculated from 4 groups based on resources (R), environment (E), connectivity (C) and output (O). Each (sub)indicator is a

TABLE 3.1: Countries of U21 (Williams et al., 2014)

Countries (1-10)		Countries (11-20)		Countries (21-30)		Countries (31-40)		Countries (41-50)	
1	United States	11	Norway	21	Korea, Rep. (South)	31	Poland	41	Argentina
2	Sweden	12	Austria	22	Taiwan	32	Greece	42	Thailand
3	Canada	13	Belgium	23	Spain	33	Chile	42	Ukraine
3	Denmark	14	Germany	24	Portugal	34	Serbia	44	Croatia
5	Finland	15	Hong Kong	25	Slovenia	35	China	45	South Africa
6	Switzerland	16	New Zealand	26	Czech Republic	35	Russia	46	Mexico
7	Netherlands	17	Ireland	27	Italy	37	Slovakia	47	Turkey
8	United Kingdom	18	France	28	Malaysia	38	Brazil	48	Indonesia
9	Australia	19	Israel	29	Hungary	39	Romania	49	Iran
10	Singapore	20	Japan	30	Saudi Arabia	40	Bulgaria	50	India

weighted average of multiple variables. Table 3.2 lists names and weights of the indicators. The resource, environment, and connectivity groups have a 20% weight, and output contributes 40% to the final rank.

The overall scores U21 ranking are available for each year, but the (sub)indicators are available only for the years 2012-2014 and 2019-2020 period. For the appropriate application of bi-clustering, only the (sub)indicators must be considered. As a starting point, since (sub)indicators of the U21 rankings are not available from 2015, the year 2014 was selected. Then, the bi-clustering method was run on the U21:2020 ranking, and the results were compared to the U21:2014 results.

Several sources, such as the OECD's Education at a Glance 2013 report, UNESCO Statistics, The World Economic Forum's Global Competitiveness Report 2013-2014, and the U21 Scopus data bank and survey, were used to collect indicators for U21. These values were then scaled to a 0-100 interval for each variable. The overall scores for countries were calculated by weighted sums of these indicators. Although U21 published the score values of the indicators, these values cannot be verified completely. Firstly, although most sources of raw data are published, a few are not available (e.g., the qualitative measure of the policy environment (E4), which based on surveys, the results of which are not available). Secondly, various series are derived from previously obtained on-line search results. For example, the variable connectivity webometrics visibility index (C4) measures the external links that university web domains received from third parties. Webometrics does not contain archived data, so it is impossible to re-calculate the indicator of C4. Thirdly, there are several missing values ($42/1200 = 3.42\%$), and the methodology used to treat missing data is

TABLE 3.2: List of indicators (Williams et al., 2014.

w	**Abbr.**		**Variables**
5.0%	**R**esources 20%	R1	Government exp. on tertiary education institutions as a % of GDP
5.0%		R2	Total exp. on tertiary education institutions as a % of GDP
5.0%		R3	Annual exp. per student (full-time equivalent) by tertiary education institutions in USD, PPP
2.5%		R4	Exp. in tertiary education institutions for R&D as a % of GDP
2.5%		R5	Exp. in tertiary education institutions for R&D per head of population at USD, PPP
2.0%	Environ-ment 20%	E1	% of female students in tertiary education
2.0%		E2	% of female academic staff in tertiary institutions
2.0%		E3	A rating of data quality.
14.0%		E4	Qualitative measure of the policy environment.
4.0%	Connectivity 20%	C1	% of international students in tertiary education
4.0%		C2	% of articles that are co-authored with international collaborators (coverage is all institutions that publish at least 100 papers).
2.0%		C3	Webometrics web transparency measure: sum of values from 4,200 universities divided by the country's population.
2.0%		C4	Webometrics visibility index (external links that university web domains receive from third parties). Sum of data for 10,000 tertiary institutions divided by the country;s population.
4.0%		C5	Responses to question "Knowledge transfer is highly developed between companies and universities", which was asked of business executives in the annual survey by IMD World Development Centre, Switzerland.
4.0%		C6	% of university research publications that are co-authored with industry researchers
13.3%	**O**utput 40%	O1	Total number of journal articles that are produced by higher education institutions
3.3%		O2	Total number of articles that are produced by higher education institutions per capita
3.3%		O3	Average impact of articles, as measured by citations in 2014 of articles that were published in previous years using the Karolinska Institute normalized impact factor.
3.3%		O4	Depth of world-class universities in a country. This is calculated as an average of the institutions' score of a country that is listed in the top 500 of the Shanghai ranking, divided by the country's population
3.3%		O5	Excellence of a nation's best universities, which is calculated by summing the Shanghai Jiao Tong scores for the nation's three best universities
3.3%		O6	Enrollment in tertiary education as a % of the eligible population, which is defined as the 5-year age group after secondary education
3.3%		O7	% of the population aged 25-64 with a tertiary qualification
3.3%		O8	Number of researchers (full-time equivalent) in the nation per population
3.3%		O9	Unemployment rates among tertiary-educated aged 25-64 years compared with unemployment rates for those with only upper-secondary or post-secondary non-tertiary education

)

unpublished.

Table 3.3 summarizes various descriptive statistics of the 24 indicators, which are scaled to scores of 0-100. Most of the cells (11) are missing for the proportion of female academic staff (E2). The least varied data (the indicator in which the countries are the most similar) is the proportion of female students (E1). Its scores ranged in a 20.7 score interval, and the relative standard deviation of countries' data is the smallest, at only 4.1%. The countries are the most different in terms of the number of journal articles (O1); its mean is extremely small (8.3), and its relative standard deviation is the highest (196.2%).

TABLE 3.3: U21 Descriptive Statistics

Variable	Missing Values	Min	Range	Mean	Relative SD
R1	0	22.1	77.9	47.5	34.9%
R2	0	25.0	75.0	55.2	30.9%
R3	0	3.8	96.2	44.7	54.1%
R4	6	2.8	97.2	40.5	57.7%
R5	6	0.3	99.7	36.4	82.1%
E1	0	79.3	20.7	98.6	4.1%
E2	11	35.8	64.2	82.6	16.4%
E3	0	68.2	31.8	94.1	7.9%
E4	0	53.4	46.6	80.6	13.9%
C1	3	0.5	99.5	27.1	96.6%
C2	0	22.5	77.5	64.2	30.6%
C3	0	4.0	96.0	34.7	73.4%
C4	0	2.8	97.2	34.1	72.2%
C5	3	27.1	72.9	63.5	31.5%
C6	0	0.1	99.9	43.7	58.8%
O1	0	0.1	99.9	8.3	196.2%
O2	0	0.1	99.9	42.6	74.7%
O3	0	23.4	76.6	61.2	32.7%
O4	0	0.0	100.0	25.2	109.9%
O5	0	0.0	100.0	19.5	95.4%
O6	3	23.1	76.9	64.2	28.3%
O7	1	6.7	93.3	50.9	45.2%
O8	2	1.2	98.8	38.7	65.2%
O9	6	33.5	66.5	62.1	26.5%

Note: SD = standard deviation, N = 50, Max = 100

The countries of U21 have varied geographic locations, with different income levels and histories. Most of the countries (27) are from Europe. There are 14 countries from Asia, 6 from America, 2 from Oceania (Australia and New Zealand) and 1 from Africa (South Africa). The countries are also varied in their income levels. They are grouped into high, higher-middle, lower-middle and low-income categories by the World Bank.[1] Most of the countries (36) were classified as high-income countries by the World Bank in 2014. The remaining 14 countries were given middle ratings. Three of the 14 countries were categorized into the lower-middle income category (India, Indonesia, and Ukraine); the other 11 countries were placed in the upper-middle income category. The countries also have different historical pasts. There are 38 developed market economies and 12 post-socialist countries.

The indicators with the largest and smallest relative standard deviation (SD) are examined in more detail. There are only 4 indicators in Table 3.3 that have relative SDs less than 20%. These cover all environmental indicators. Two of them are extremely low, i.e., under 10%. The proportion of female students in tertiary education (E1) has the smallest relative SD, only 4.1%. Thirty-nine of the 50 countries obtained the maximum score of 100 for this indicator, and 9 countries' scores are between 90 and 100. The remaining 2 countries also have high values of approximately 80: India's score is 83.5, and South Korea's is 79.3. The rating for data quality (E3) has the second lowest relative SD, 7.9%. This indicator was derived from each quantitative series by U21 as a categorical variable: 1 indicates available data, 0.5 indicates some available data with adjustments needed, and 0 indicates any other case. Among the 50 countries, 21 have the maximum score of 100, 17 countries' scores are between 90 and 100, 9 countries' scores are between 80 and 90, and the 3 remaining countries have lower scores (Saudi Arabia and South Africa 77.3, India 68.2).

Considering the highest relative SDs, the largest (196.2%) can be observed on the total number of journal articles produced by higher education institutions (O1). The United States had the maximum score of 100, which was extremely high compared to those of other countries. The second highest score of 58.9 was

[1] `http://databank.worldbank.org/data/download/site-content/OGHIST.xls`

obtained by China, followed by three countries with scores of 20-30 (UK 25.2, Japan 24.4, and Germany 20.7). There are 7 countries with scores of 10-20. The remaining 38 countries had scores under 10. In detail, the scores of 5 countries are in the interval [5,8], the scores of 16 countries are in [2,5], 17 countries' scores are less than 2, and the scores of 8 out of 17 countries are lower than 1. The average scores of the top 500 Shanghai institutes (O4) have the second largest relative SD (109.9%). In the case of this indicator, there is only one country (Switzerland) with a score of 100. It is followed by Sweden (94.5). The next 10 countries' scores are between 50 and 80, 15 are between 10 and 40 and 23 are less than 10 (including 7 countries with scores of 0.0).

3.2 RUR - The Global Ranking of Universities

This work aims to present the results of the bi-clustering method applied to a global non-thematic university ranking as well, which is as diverse as possible in terms of the number of ranking areas. Bi-clustering is less useful for thematic rankings (because these rankings are sufficiently delimited to a narrow area) or if a ranking examines only one field of science, e.g., research (see URAP, CWTS and CWUR in Figure 2.1). There are two rankings in Figure 2.1 containing all four main areas (I-IV): THE and RUR.

The data of RUR (World University Ranking 2020[2]) on 828 institutions were selected for analysis because the weights of reputation surveys in RUR are less than those in THE.

Bowman and Bastedo (2010) showed that anchoring effects have an influence on reputational assessments. More precisely, being ranked highly in a ranking increases reputation, not the other way around. This means that reputation surveys are biased towards elite universities, and because of this, the author chose not to use THE (as surveys count higher in their rankings than in the RUR). The

[2]`https://roundranking.com/ranking/indicator_ranking.html#academic-staff-per-students_SO-2020`

author also conducted analyses on RUR's 2022 and 2023 rankings and compared the results to those of 2020.

Table 5.4 shows the construction of RUR. Only the 20 basic indicators were employed; the four aggregated subindicators and the overall scores were ignored. The abbreviations and description of indicators in Table 3.4 are according to the original categorization of RUR. This does not entirely overlap with how the author classified the indicators in Figure 2.1.

The RUR framework comprises four essential elements: teaching (T), research (R), international diversity (I), and financial sustainability (F). Each component is further divided into five subcategories with equal weights assigned to them. Refer to Table 3.4 for more details.

The number of universities ranked by RUR varies across continents and countries. Europe has the highest number of ranked institutions, with 323, followed by Asia, with 278. North America ranks third with 165, while Oceania has 34 institutions. Africa and South America have ten listed HEIs in the 2020 global ranking.

In Europe, the Middle-East, and Africa region, the distribution of ranked institutions can be seen in Appendix A in Figure A.1. Russia holds the highest number of HEIs, 82, followed by the United Kingdom with 60 universities. Germany, Spain, Italy, and France each have more than ten entities listed in the global rankings. On the other hand, Bulgaria (University of Sofia), Estonia (University of Tartu), Iceland (University of Iceland), Latvia (Riga Technical University), Lithuania (Vilnius University), Slovakia (Comenius University in Bratislava), and Nigeria (University of Ibadan) only have one university included in the global ranking.

In terms of higher education institutions, the Americas region has a number of esteemed universities that can be seen in Appendix A in Figure A.2. The leading country is the United States with an impressive 137 ranked universities, followed by Canada with 22 and Mexico with Brazil with 6. Additionally, there

TABLE 3.4: RUR Indicators and Weights

Variable		Description	Weights
T		**TEACHING**	**40%**
	T1	Academic staff per students	8%
	T2	Academic staff per bachelor degrees awarded	8%
	T3	Doctoral Degrees awarded per academic staff	8%
	T4	Doctoral degrees awarded per bachelor degrees awarded	8%
	T5	World teaching reputation	8%
R		**RESEARCH**	**40%**
	R1	Citations per academic and research staff	8%
	R2	Doctoral degrees awarded per admitted PhD	8%
	R3	Normalized citation impact	8%
	R4	Papers per academic and research staff	8%
	R5	World research reputation	8%
I		**INTERNATIONAL DIVERSITY**	**10%**
	I1	Share of international academic staff	2%
	I2	Share of international students	2%
	I3	Share of international co-authored papers	2%
	I4	Reputation outside region	2%
	I5	International level	2%
F		**FINANCIAL SUSTAINABILITY**	**10%**
	F1	Institutional income per academic staff	2%
	F2	Institutional income per students	2%
	F3	Papers per research income	2%
	F4	Research income per academic and research staff	2%
	F5	Research income per institutional income	2%

are a few notable institutions in the region, including Austral University in Argentina, Universidad Pontificia Bolivariana de Medellin-Colombia in Colombia, and the University of Puerto Rico in Puerto Rico that solely represent their own country in the 2020 RUR ranking.

The Asia-Pacific region's number of ranked universities can be seen in Appendix A in Figure A.3. China takes the lead with 61 ranked institutions, followed by Iraq with 37. Japan and Taiwan both have 34 universities listed in the global ranking. Kuwait University, University of Macau, University of Qatar, Ivane Javakhishvili Tbilisi State University (Republic of Georgia), and Can Tho University (Vietnam) are the only universities from their respective countries to be ranked.

Taking a closer look at RUR's methodology and how they determine final scores for universities is worthwhile. The process begins with RUR utilizing the initial data sets submitted by universities. Next, universities are ranked from largest to smallest based on these initial values. Each institution is then assigned a percentile based on the 20 sub-indicator values, resulting in a value on a 100-point scale that considers both rank and sample size. (RUR, 2023).

Looking at the statistics of the 20 subindicators of RUR, due to the 100-point scaling, it is difficult to draw any conclusions from the figures. The means of the indicators are around 50, indicating even distribution around the scale's midpoint. Nevertheless, the relative SDs are quite high for all indicators, ranging from 45.2% to 57.8%, suggesting significant variation in the data. The published data set is complete, with no missing values. In the process of calculating rankings, universities that fail to submit a value are assigned 25% of the average value of their country. In cases where there is only one university in a country, the institution receives 25% of the world average (RUR, 2023).

The upcoming chapter provides a comprehensive introduction to the method of bi-clustering, which is a powerful technique used in data analysis to simultaneously cluster both rows and columns of a dataset.

Chapter 4

Methodology

4.1 The Method of Bi-clustering

In addition to arbitrary classification, clustering methods are used to separate clusters (see, e.g., Rad et al., 2011; Ibáñez et al., 2013). Ibáñez et al. (2013) clustered public universities in the area of computer sciences into four groups based on their productivity, visibility, quality, prestige, and internalization. However, clustering alone cannot be used to specify regional or other rankings because, beforehand or in parallel, clustering indicators should be selected for ranking similar universities or countries (Poole et al., 2017).

Bi-clustering methods are relatively new, almost entirely unknown, and unused in the social sciences. The author demonstrates the capabilities of these methods in clustering and ranking Higher Education Systems (countries) and Higher Education institutions. One can find meaningful but far-from-evident leagues of both countries and indicators using well-chosen elements of the family of bi-clustering methods. The selected indicators shed light on HEIs' and countries' strengths, weaknesses, and positions in the rankings. Last but not least, the proposal opens a new direction of multivariate analysis free of subjective or ad-hock weights and does not require indicator selection over non-comparable indicators.

A fair comparison of HEIs can be performed within leagues. In the present paper, the author creates three leagues within HESs and HEIs, which are denoted as A, B, and C and have simple characteristics to make the methods and results as transparent as possible while still being able to make nontrivial observations.

League A: Upper league,

League B: Middle league,

League C: Lower league.

Bi-clustering is a data mining technique that enables the simultaneous clustering of the rows and columns of a matrix. The term was first introduced by Mirkin (1998) to name a technique that was introduced many years previously, in 1972, by J. A. Hartigan (1972). This clustering method was not generalized until 2000 when Cheng and Church (2000) proposed a bi-clustering algorithm based on the variance and applied it to biological gene expression data. Many bi-clustering algorithms have been developed for bioinformatics; see an excellent review in Pontes et al. (2015). Until recently, these methods were rarely used in other fields of science.

Despite the very few publications that use bi-clustering algorithms in the social, business and economic sciences (see, e.g., Liu et al., 2009; Huang, 2011 for an exception), there is already a publication (see Raponi et al., 2016) on the bi-clustering of university performances. This study clearly demonstrates how to select indicators and universities simultaneously.

The term *bi-clustering* or *biclustering* was coined by Cheng and Church (2000) that refers to a distinct class of clustering algorithms that perform simultaneous clustering on rows and columns (Madeira and Oliveira, 2004). Several other proposed and used names exist in the literature such as coclustering, bidimensional clustering, subspace clustering, and block clustering (Hartigan, 1972; Madeira and Oliveira, 2004).

A *bicluster* refers to a subset of rows that display similar behavior across a subset of columns, and vice versa (Madeira and Oliveira, 2004).

There are different types of bi-clusters (Madeira and Oliveira, 2004):

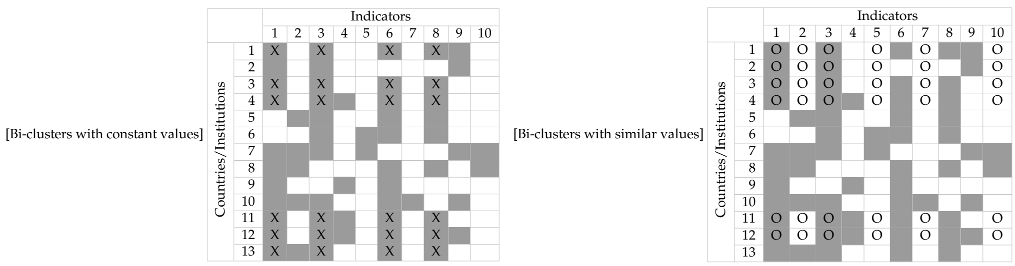
BIC1 Bi-clusters with constant values (in rows and/or columns) (see Table 4.1(a));

BIC2 Bi-clusters with similar values (on rows and/or columns) (see Table 4.1(b)).

The BIC1-type bi-clustering algorithms re-order the rows and columns of the matrix in an attempt to bring similar rows and columns as close together as possible at the same time and then to find bi-clusters with similar (constant) values (see, e.g., Table 4.1(a)). In contrast, BIC2-type algorithms seek bi-clusters with similar values in rows and columns. Similarity can be measured in many ways; the simplest way is by analyzing the variance between groups using the co-variance between rows and columns. In Cheng and Church (2000)'s theorem, a bi-cluster is defined as *a subset of rows and columns with almost the same score*. The score is the measure of the similarity of the rows and columns. Typical clustering algorithms are based on global similarities of rows or columns of the expression (or feature) matrix.

Cheng and Church (2000) developed a function called the Mean Squared Residue Score to score sub-matrices and locate those with high row and column correlation (bi-clusters). The exhaustive search for and scoring of all sub-matrices is NP-hard, and they employed a Greedy Search Heuristic in their approach. Tanay et al. (2002) proved that bi-clustering is an NP-hard problem, meaning no known algorithm can solve it in polynomial time (Garey and Johnson, 1979). In the original paper of Cheng and Church (2000), the rows corresponded to genes and the columns to conditions. In this analysis, the rows correspond to the countries or to the institutions, and the indicators of the ranking (U21 and RUR) are the columns (refer to Table 3.2 and Table 3.4).

TABLE 4.1: Cell Selection Results



Note: X, O denote the selected cells; ■: upper/ ❏: lower than a specified threshold.

Grey cells represent those cells that are above a given threshold, here, the median of the total matrix. The “X"-s in Table 4.1(a) indicate a possible choice for a subset of cells that form a similar subset as well as a bi-cluster with respect to rows and columns.

It is imperative to discuss the effect of the choice of threshold. The lower the threshold is, the larger and less similar the cluster (see, e.g. Gusenleitner et al., 2012). The balance between the similarity and the size of the bi-cluster can be set by parameter selection for a target function (Gusenleitner et al., 2012).

Table 4.1(b) shows another possible selection, where “O" indicates the maximal entries of the selected columns and the correlations among rows are maximal. The method seeks to find a balance between the size of the bi-cluster and the similarity, which, in this case, is measured by the row correlation. The measure of similarity, i.e., the distance between the indicators, is a freely chosen parameter of the method, as in classical clustering methods. This choice requires particular care because the results inherently depend on it. Proper interpretation can become challenging in the application of classical clustering methods, and this applies to bi-clustering even more.

This work first demonstrates the method on a relatively small number of objects, namely, the U21 countries’ HESs, then performs the analysis on a larger data set of institutions to show that well-selected bi-clustering methods can identify leagues (countries/institutions and indicators simultaneously). For simplicity, the paper identifies only three leagues: upper league A, middle league B, and lower league C. For that purpose, two bi-clustering methods are used.

The first one is the **i**terative **B**inary **Bi**-clustering of **G**enes (iBBiG) (Gestraud et al., 2014) method.

This algorithm is a BIC1-type method that produces bi-clusters, where the cells exceed the threshold (i.e., median) (see Table 4.1(a)). The procedure starts with the normalization of the indicators, as defined in Equation (4.1).

iBBiG does not require all unique cells within a bi-cluster to be above or below a threshold (i.e., the median). However, the medians for the selected cells must be above/below both the row/column median and the medians of the excluded rows and columns.

$$x'_{ij} := \frac{x_i - m(j)}{M(j) - m(j)}, \tag{4.1}$$

where $m(j) = \min_i x_{ij}, M(j) = \max_i x_{ij}$. The next step in iBBiG involves determining a threshold based on the median of the matrix. A new binary matrix is then created, where cells with values above the threshold are assigned a value of one, while all other cells are assigned a value of zero. The key step of iBBiG is thus to find the cells that form similar rows and columns.

As a result, we obtain League A. The binary reversed data and the same procedure yield League C. The iBBiG method can produce more than one bi-cluster (i.e., leagues), which can overlap if the above procedures are applied with different thresholds.

Let the author note here that when using different thresholds to develop several alternative clusters, a quality test is needed to evaluate the results. For simplicity, the author does not apply multiple thresholds; instead, to identify the middle league, another bi-clustering method, namely, **Bi**-**c**lustering **A**nalysis and **R**esults **E**xploration (BicARE), is used. Through the implementation of the BicARE technique, we are able to produce a bi-cluster that effectively defines a middle league (League B) of nations/institutions that intersect with both (A) and (C) (see Figure 4.1), thereby yielding a more comprehensive understanding of their respective accomplishments. The position of the countries with respect to the created leagues is depicted in Figure 5.4.

BicARE is a BIC2-type method, where the similarity measure is the correlation (see Table 4.1(b)). BicARE (Gestraud et al., 2014) is the improved and enhanced version of the **FL**exible **O**verlapped bi**C**lustering (FLOC) algorithm proposed by Yang et al. (2003). This method is based on the notion of residue, which is a measure of the similarity of the elements in a bi-cluster (see Yang et al., 2005 for a definition of the residue). The smaller the residue is, the more similar the elements of the bi-cluster are. Similarly to the interpretation of the upper and lower leagues, when interpreting the middle league (see the cells of Table 4.1(b) that are marked by "O"), the BicARE method specifies a group (submatrix) of countries/institutions and indicators whose values are similar (their variances are as small as possible) for both countries and indicators.

Figure 4.1 depicts the three different leagues that will be determined by the two different bi-clustering methods. Using two different bi-clustering techniques ensures finding all the members of the three different leagues. The iBBiG algorithm is capable of pinpointing the best-performing entities (League A), and its application to the reversed dataset effectively identifies the lower-performing entities (League C). In contrast, the BicARE method can identify the most homogeneous group of entities, encompassing those that exhibit strong performance in certain indicators while lagging in others. Incorporating the BicARE method is essential for capturing the full spectrum of countries/institutions and ensuring a thorough exploration of diverse performance patterns.

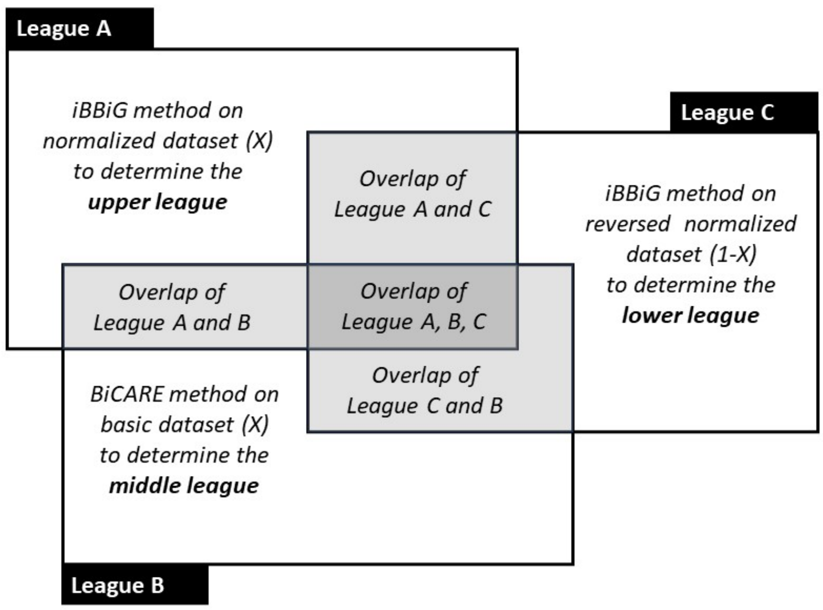


FIGURE 4.1: The Leagues and its Overlaps Determined by the Bi-clustering Methods

To obtain a preliminary picture of the possible bi-clusters and to later compare these potential bi-clusters with the obtained bi-clusters, a visualization method, i.e., a seriation method can be used. Seriation is an exploratory combinatorial data analysis technique for reordering objects into a sequence (Liiv, 2010). Typically, finding an optimal seriated matrix is also an NP-hard problem (similar to finding bi-clusters). Therefore, heuristic methods are usually applied. In this study, the hierarchical cluster-based matrix seriation (Hahsler et al., 2008) is used.

In the upcoming section (4.2), the reader will find a concise explanation of the necessary analysis steps to identify the bi-clusters.

4.2 Steps of Analysis

The analysis consists of 5 steps, both in case of countries and institutions:

Step 1: Replacing missing values;

Step 2: Normalization;

Step 3: Data binarization and reversal of binary entries;

Step 4: 100 iterations of bi-clustering and selection of bi-clusters with the largest significant score values; and

Step 5: Calculation of partial rankings for the significant bi-clusters.

As a result, the following three bi-clustering can be defined:

- League A (the bests): iBBiG on normalized basic data (X)
- League B (the midfield): BicARE on basic data (X)
- League C (the laggards): iBBiG on the reverse (1-X) of normalized basic data (X)

Overlaps can also be found between these leagues for the indicators and countries/institutions.

The applied iBBiG algorithm is robust to missing values (Gestraud et al., 2014), but the BicARE algorithm requires a complete database. Choosing the appropriate method for replacing of the missing values is important because data can be missing completely at random (MCAR), at random (MAR) and not missing at random (NMAR). In our case, the missing data is MAR-type because the values could be calculated based on other indicators (Little and Rubin, 2002).

There are several methods of replacing missing values, but their application is recommended if missing values data does not exceed 5% (Scheffer, 2002). Since the ratio of missing values was low $(41/1200 = 3.42\%)$, in the first step (Step 1), missing values were replaced. In order to minimize the potential methodological dangers that can be caused by replacing the missing values, the missing scores were calculated based on the original rank of the given country. The author did not use the software solutions offered to replace missing data (e.g., mean or median), but replaced the missing data in such a way that the original rankings could be reproduced. The original scores of groups R, E, C, and O were decompiled. In those cases in which there was only one unknown value, the missing score could be calculated easily. If there was more than one unknown score, their sum could be calculated and divided equally among the missing cells. Then (in Step 2), the cell values were normalized via a min-max normalization formula (see Eq. 4.1).

Since the original iBBiG method finds only the League(s) A, in the next step (Step 3), the reversed normalized data (1-nomalized data) are also calculated to specify League(s) C. This step is not used when specifying League(s) B because the applied BicARE algorithm treats variances instead of binarized values. The binarization is also ignored when applying the BicARE algorithm.

Before bi-clustering, the data matrix was ordered using a seriate algorithm. A hierarchical clustering algorithm was used to classify both rows and columns. To use this ordered matrix as the initial matrix for both the iBBiG and BicARE algorithms, the distance function for rows (countries) was the Euclidean distance, and the distance function for columns (indicators) was the Spearman correlation.

Since both iBBiG and BicARE are heuristic methods, in step four (Step 4), every calculation was iterated 100 times, and the bi-clusters with the highest score values were selected for further analysis.

F-statistics were calculated from the two-way ANOVA model with row and column effects. A bi-cluster was considered a *significant bi-cluster* if both the row and column effects were significant.

In the last step (Step 5), partial rankings were calculated and compared to the corresponding part of the U21 and RUR rankings. When calculating partial rankings for countries and institutions in the specified bi-cluster(s), the original weights of U21's and RUR's indicators were used, and the total scores for the countries were calculated using the selected indicators in the given bi-cluster.

Step 1: The main components are calculated as weighted averages of the scores of indicators.

Step 2: The highest score for each of the four components is increased to 100, and the component score values of every country are re-scaled proportionally.

Step 3: The overall score values are similarly calculated as a weighted mean. The highest score values are transformed into 100, and the remaining overall scores are re-scaled proportionally.

Step 4: In the final step, the entities are ordered by their overall scores.

In the below a detailed step-by-step pseudo-code can be seen for calculating the partial rankings for U21 and RUR.

Pseudo Code 1: Calculating U21 Partial Rankings

```
# Input: orig_mtx (matrix), weights (of the original
   ↪ ranking)

# Perform bi-clustering
C <- biclust::bicluster(orig_mtx)

# Convert the bicluster to a data frame
B <- as.data.frame(C[[1]])

# Select weights corresponding to the columns of B
selectedweight <- weights[, colnames(B)]

# Group columns based on specific patterns
in their names
colsC <- find_columns_with_pattern(B, "C")
colsO <- find_columns_with_pattern(B, "O")
colsE <- find_columns_with_pattern(B, "E")

# Calculate weighted sums and normalize to a
   ↪ percentage scale for each group
swC <- weights[, colsC]
swO <- weights[, colsO]
swE <- weights[, colsE]

BC <- rowSums(B[, colsC] * swC) * 100 / max(rowSums(B
   ↪ [, colsC] * swC))
BO <- rowSums(B[, colsO] * swO) * 100 / max(rowSums(B
   ↪ [, colsO] * swO))
BE <- rowSums(B[, colsE] * swE) * 100 / max(rowSums(B
   ↪ [, colsE] * swE))
```

Pseudo Code 2: Calculating U21 Partial Rankings (contd.)

```
# Add individual scores to the data frame
B$E_Score <- BE
B$C_Score <- BC
B$O_Score <- BO

# Calculate an overall score based on a weighted
   ↪ combination of individual scores
B$Overall_Score <- calculate_overall_score(B$E_Score,
   ↪  B$C_Score, B\$O_Score)

# Assign ranks based on the overall score
B$Rank <- rank(-B$Overall_Score)

# Output: Data frame B containing the entities of the
   ↪  selected league with Overall_Score and Rank
   ↪ columns
```

Pseudo Code 3: Calculating RUR Partial Rankings

```
# Input: mtx (matrix), weights (of the original
   ↪ ranking)

# Perform bi-clustering
C <- biclust::bicluster(mtx)

# Extract the bicluster and convert it to a data
   ↪ frame
B <- as.data.frame(C[[1]])

# Extract selected columns from the "weights" data
   ↪ frame based on the columns of "B"
selectedweight <- weights[, colnames(B)]

# Store column names of "B"
cols <- colnames(B)

# Extract corresponding columns from "weights"
sw <- weights[, cols]

# Calculate the column sums across the selected rows
   ↪ and multiply by the selected weights and
   ↪ normalize
BR <- rowSums(B[, cols] * sw) / max(rowSums(B[, cols]
   ↪  * sw))
```

Pseudo Code 4: Calculating RUR Partial Rankings (contd.)

```
# Add Overall_Score column to "B" with the calculated
   ↪  values
B$Overall_Score <- BR

# Add a Rank column to "B" with ranks based on
   ↪ Overall_Score in descending order
B$Rank <- rank(-B$Overall_Score)

# Output: "B" data frame containing the entities of
   ↪ the selected league with Overall_Score and Rank
   ↪  columns
```

Chapter 5

Results

This chapter provides a detailed account of the results obtained from bi-clustering. In Subsection 5.1, a comprehensive discussion is given on the leagues of countries. The section provides a detailed analysis of the bi-clustering outcomes, including the composition of each league. In Section 5.2, the results of the leagues created on the Round University Ranking are presented with particular interest in the effects of different thresholds applied during the analyses. Finally, in Section 5.3, the author presents the results related to Hungary. The section offers a detailed analysis of the results specific to Hungary, comparing them to similar countries. The source files and code that were used during the bi-clustering are available on the below links.

- Results of the U21:2014 league creation.[1]
- Results of the U21:2020 league creation.[2]
- Results of the RUR:2020 league creation.[3]
- Results of the RUR:2022 league creation.[4]
- Results of the RUR:2023 league creation.[5]

[1] https://kmt.gtk.uni-pannon.hu/kutatas/u21/EN/BIC/
[2] https://bit.ly/bicluster
[3] https://kmt.gtk.uni-pannon.hu/kutatas/u21/EN/BIC/
[4] https://bit.ly/bicluster
[5] https://bit.ly/bicluster

5.1 The Leagues of Countries

It is important to note that, unlike classical clustering, bi-clusters can overlap, depending on the method applied. Moving forward, the author will highlight scenarios where belonging to a single cluster or multiple clusters holds particular significance. In both cases, it is essential to consider both the country and indicator positions simultaneously.

After seriation, two bigger homogeneous blocks can be identified based on Figure 5.1. The block of the darker cells on the top left corner of Figure 5.1 indicates the top league, while the bigger lighter block, which indicates the remaining (lower) league, can be discovered at the bottom of the figure. The dendrogram of two-way clustering also shows that regarding rows and columns two main blocks can be specified. Even though the heat map of the normalized data suggests two bi-clusters, only the bi-clustering algorithm, and F-tests will help to determine the significant bi-clusters.

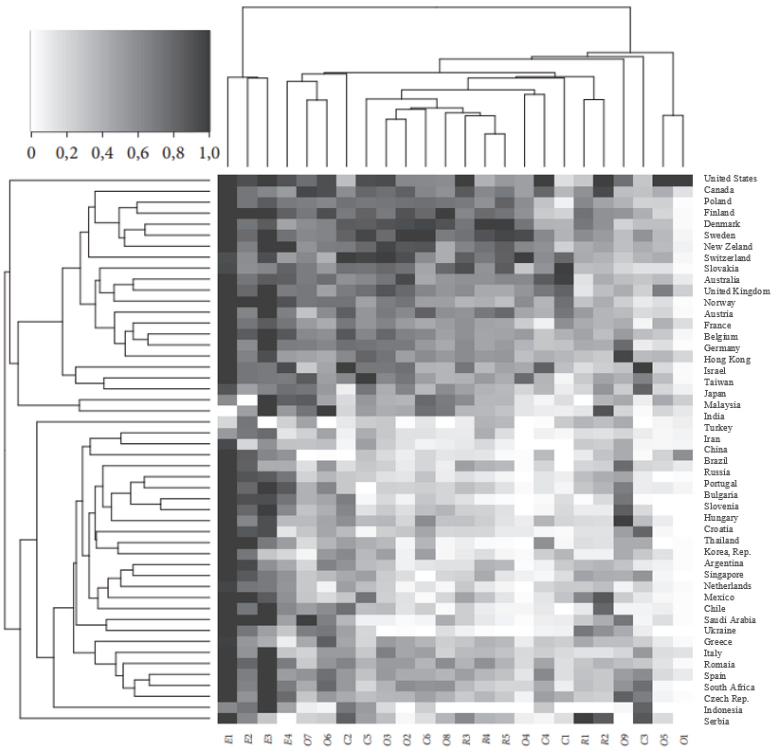


FIGURE 5.1: Heat Map of the Normalized and Seriated Matrix

After the 100 runs, two bi-clusters with higher frequencies appeared; the others had negligible hits. Table 5.1 shows that only cluster number 1 has acceptable significance for both dimensions.

TABLE 5.1: Results of scores and significances for iBBiG bi-clustering algorithms

Dataset	№	Score	Rows	Cols	F-tests for Row Effects	F-tests for Col (p-values)
U21	1	287.3794	23	19	0.0000	0.0000
U21	2	78.8670	22	5	0.1866	0.0000
Reverse U21	1	535.1661	38	19	0.0000	0.0000
Reverse U21	2	52.1089	11	7	0.9350	0.0000

The iBBiG algorithm on normalized data specifies League A because the cell values from the bi-cluster are significantly higher than those of the excluded data. The iBBiG algorithm on the reversed data identifies League C.

When selecting League(s) A and C, it is important to also specify League(s) B in a similar manner. To determine the middle league, a unique concept of similarity is utilized. The author aims to identify a middle league where the differences between countries and indicators are minimal. This is achieved using the BicARE method, which generates bi-clusters that meet these criteria. Then, one can identify a significant bi-cluster by conducting an F-test to compare variances for both countries and indicators between included and excluded cells.

To illustrate the disparities between the countries included in each league and those excluded, it is beneficial to visually depict the median and variance across both the rows (representing countries) and columns (representing indicators) (refer to Figure 5.2 and Figure 5.3).

The median of the countries and indicators included in League A is higher but the variance is lower and they exhibit significantly less variability than the countries that are left out. The countries and indicators in League C have lower median and variance. The same is true for the countries and indicators classified in League B. Figure 5.2 and Figure 5.3 illustrate to what extent the medians and variances of the cell values of the variables/countries included in and excluded

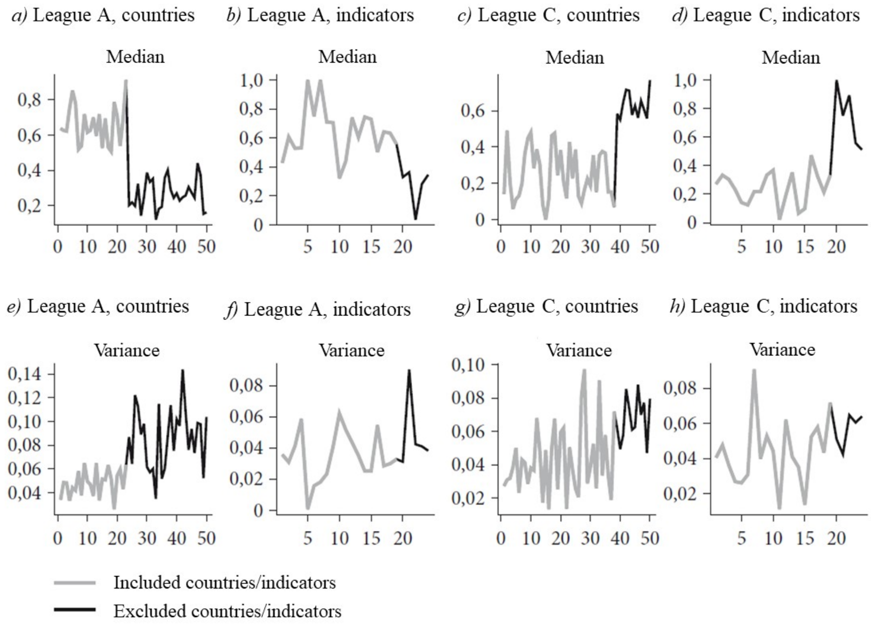


FIGURE 5.2: Medians [*a)-d)*] and Variances [*e)-h)*] of League A and League C

from the significant leagues differ from each other.

Since a country can have several high and low values simultaneously, it can be a member of more than one league. Similarly, if an indicator has a high relative variance (see Table 3.3), its high-value cells can be included in League A, and lower-value cells can be included in League C (see the overlaps of columns of cells that are labeled X or O in Table 4.1). Therefore, the results of bi-clusters can specify overlaps (see Figure 5.4). An in-depth analysis can highlight which countries are separated, and the analysis of the overlaps can provide a detailed picture of the countries and indicators.

As mentioned at the beginning of this section, bi-clusters might (or might not) have overlaps (see Figure 5.4), which is worth analyzing case by case.

League A: League A contains 23/50 countries and 19/24 indicators. The remaining variables are journal articles (O1), the score of the nation's best three universities by Shanghai (O5), unemployment rates (O9), government expenditure (R1), and international students (C1). These are the indicators for which

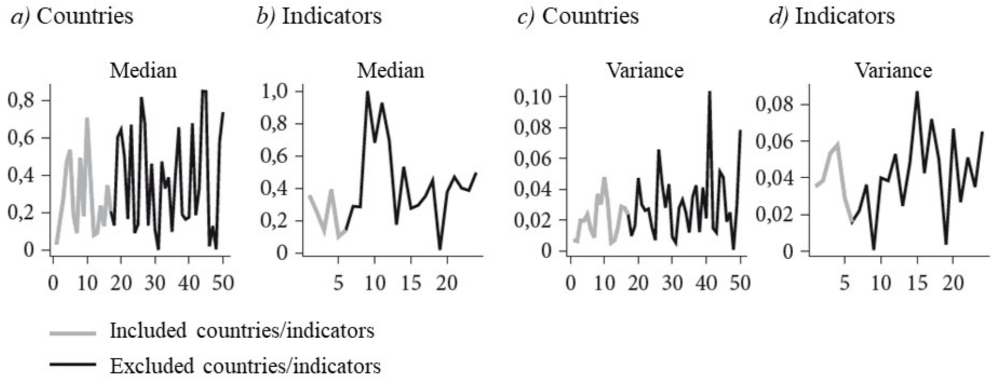


FIGURE 5.3: Medians [*a)-b)*] and Variances [*c)-d)*] of League B

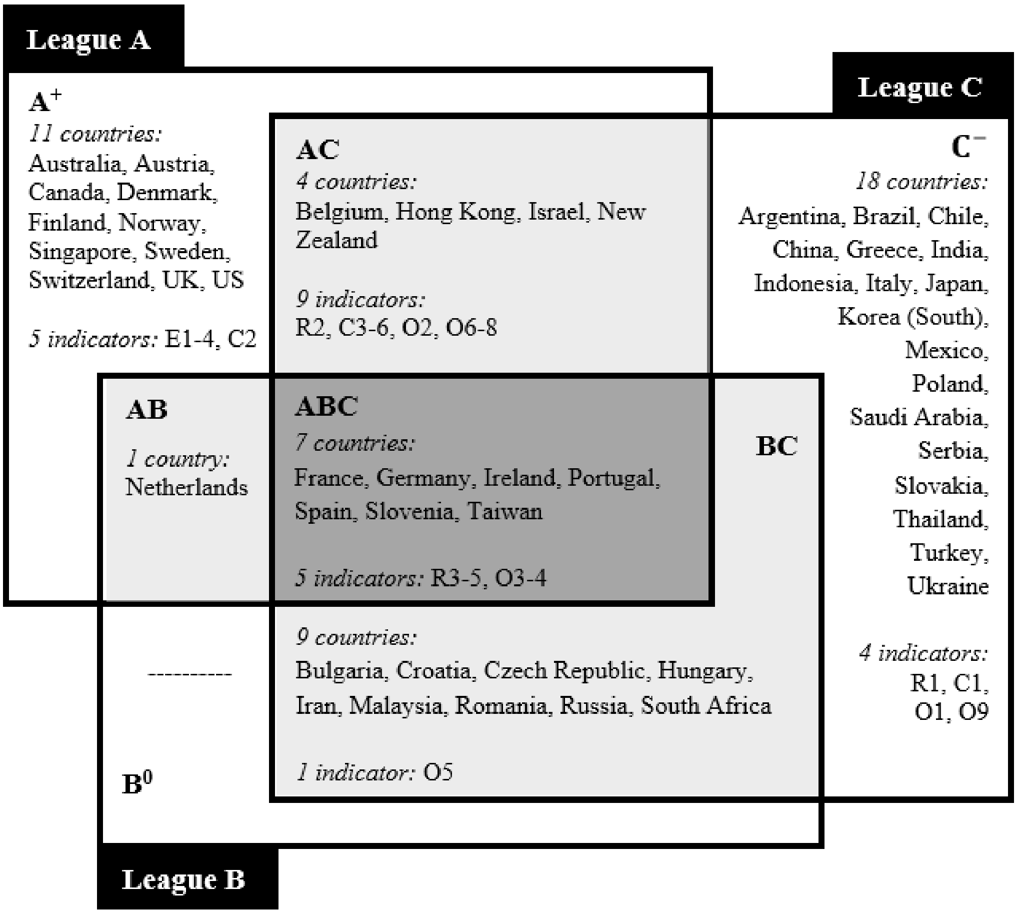


FIGURE 5.4: Results of Bi-clustering Algorithms Specifying Leagues

countries of League A do not perform equally well. The absence of indicator O1 in League A is not surprising because, among all the indicators, this one has the highest relative standard deviation. Table 3.3 shows that std.dev=196.2% for all 50 countries. Because there are 23 countries in League A, it is still very high (185.1%). Slovenia has the lowest score (0.6), and the US has the highest score (100.0). Of the 23 countries, only 7 have O1 scores above 10: Spain (10.1), Australia (11.0), Canada (14.8), France (16.6), Germany (20.7), the UK (25.2) and the US (100.0)

League A^+***:*** League A comprises 11 countries and five distinct indicators that are not shared with any other leagues (which is denoted as A^+). These countries are among the top 12 countries in the original U21 ranking. The method employed is capable of identifying the indicators that differentiate the countries in League A. These indicators primarily pertain to the environment (E1-4) and one that is associated with connectivity (C2).

Even though some countries may have a higher GDP per capita[6], the author does not believe that differences in resources are the main factor causing the separation. Additionally, the indicators related to the environment (E1-4) are only indirectly connected to HESs. Based on these findings, the author concludes that the only indicator directly impacting the separation of the top group is the articles co-authored with international collaborators (C2).

League C: League C includes most of the countries (38) and those 19 indicators which were not in League A^+. This means there are more less-well-performing countries (38 in League C) than well-performing ones (23 in League A). Nevertheless, the number of indicators in League A and League C are equal (19), and 14 of them are common. In addition to these 14 common indicators, the countries of League A perform well in the environmental indicators (E1-4) and in the articles with international collaborators (C2). The countries of League C usually perform worse in government expenditure (R1), international students (C1), journal articles (O1), the nation's best three universities by the Shanghai

[6] `http://databank.worldbank.org/`

ranking (O5) and unemployment rate (O9).

League C$^-$: The part of League C that does not overlap with other leagues (which is denoted as C$^-$) contains 18 countries and 4 indicators, of which one belongs to resources (R1), one to connectivity (C1) and two to output (O1, O9). There was no indicator from the environment category because all of the 18 countries have relatively high scores in these indicators. These 18 countries are in the middle (20^{th}, 21^{st} and 27^{th} place) and in the last 20 places of the original U21 ranking. Comparing League C$^-$ and A$^+$, only League C$^-$ contains a resource indicator (R1).

League AC and League ABC: There are 14 indicators that correspond to countries in both League A and League C. These indicators are from the resources, connectivity and output categories. Four resource (R2-5) indicators exist in the intersection of League A and League C. These indicators are significant for higher education and for specifying both League A and League C and can compare countries within these two leagues. League A requires high values on R2-5 regardless of government expenditure (R1). A low rate of government expenditure (R1) is associated with few international students (C1) and a high unemployment rate among tertiary-educated people (O9), which pull countries toward League C$^-$.

In League ABC, there are 7 countries and 5 indicators that appear in all three leagues. Most of the resource indicators (3/5) are in this league: expenditure per student (R3), R&D expenditure as a % of GDP (R4) and per capita (R5), and two output indicators (O3, O4).

League B: League B includes 17 countries and 6 indicators from the resources (R3-5) and output (O3-5) categories. The 17 countries of League B are from the middle and lower segments (14-49) of the original U21 ranking, except the Netherlands (which can be found in the 7th place of the original U21 ranking). This result shows that League A is better separated from League B than League C. The applied method (BicARE) assigned those countries and indicators to this league which became more similar after bi-clustering. Environmental indicators belong to A^+ because of their higher means and lower variances. The absence of connectivity indicators could be caused by their large variance.

League B^0 and League AB: There is no common country or indicator of League B^0. Additional evidence of the better separation of the top league (League A^+) is that there is only one country (Netherlands) in League AB but there are 9 in League BC. This reflects the big break between the top league and other leagues.

League BC: The overlap of Leagues B and C includes 9 countries and 1 indicator from the output category: the nation's best three universities by the Shanghai ranking (O5). Thus, if a country performs well on this indicator, the country could move to a higher league (from League C to League B).

Table 5.2 summarizes the results of the original U21 rankings and the results of partial rankings within leagues. All 50 countries are listed in Table 5.2 in order of their original U21 ranks, and the countries of Leagues A, B and C are specified. The 23 countries of League A can be found in the top 25 places in the U21 ranking. The 38 countries of League C can be found in the bottom 38 places of the U21 list. The 17 countries of League B are more scattered, with original U21 ranks between 18 and 49.

Table 5.3 shows the correlation for leagues between rankings by U21 and by the authors. Each correlation is significant at the 0.001 level (2-tailed). The positive nature of all of the correlation coefficients indicates that the order within each league is consistent with the original U21 ranking. Two measures of rank correlations are calculated: Kendall's τ_B and Spearman's ρ.

For the upper and lower leagues (A and C), the ranking within each league is strongly correlated with the original U21 ranking. The correlation in the middle class (League B) is slightly weaker but moderately strong. Although the countries in League B are more similar for the selected indicators, compared to the U21 ranking, the selected countries' U21 rank positions are more scattered.

Bi-clustering methods offer more than just partial rankings; they provide a more comprehensive understanding and the possibility of fair comparison. One advantage is the ability to analyze overlaps. Creating leagues can be difficult as

TABLE 5.2: Partial Ranking on U21 Leagues

	rank by U21	rank within League A			rank within League B			rank within League C		
		by U21	by author	diff.	by U21	by author	diff.	by U21	by author	diff.
US	1	1	3	-2						
Sweden	2	2	1	1						
Denmark	**3**	**3**	2	1						
Canada	**3**	**3**	6	-3						
Finland	5	5	4	1						
Switzerland	6	6	7	-1						
Netherlands	7	7	5	2	1	1	0			
UK	8	8	11	-3						
Australia	9	9	9	0						
Singapore	10	10	10	0						
Norway	11	11	8	3						
Austria	12	12	14	-2						
Belgium	13	13	13	0				1	2	-1
Germany	14	14	18	-4	2	2	0	2	1	1
Hong Kong	15	15	12	3				3	8	-5
New Zealand	16	16	16	0				4	5	-1
Ireland	17	17	17	0	3	4	-1	5	3	2
France	18	18	19	-1	4	3	1	6	6	0
Israel	19	19	15	4				7	4	3
Japan	20							8	7	1
South Korea	21							9	9	0
Taiwan	22	20	20	2	5	7	-2	10	10	0
Spain	23	21	23	0	6	6	0	11	11	0
Portugal	24	22	21	3	7	5	2	12	13	-1
Slovenia	25	23	22	3	8	8	0	13	12	1
Czech Republic	26				9	10	-1	14	14	0
Italy	27							15	15	0
Malaysia	28				10	12	-2	16	20	-4
Hungary	29				11	9	2	17	17	0
Saudi Arabia	30							18	18	0
Poland	31							19	22	-3
Greece	32							20	16	4
Chile	33							21	**26**	-5
Serbia	34							22	24	-2
China	**35**							**23**	19	4
Russia	**35**				12	14	-2	**23**	21	2
Slovakia	37							25	**26**	-1
Brazil	38							26	23	3
Romania	39				13	16	-3	27	31	-4
Bulgaria	40				14	17	-3	28	34	-6
Argentina	41							29	30	-1
Thailand	**42**							**30**	33	-3
Ukraine	**42**							**30**	28	2
Croatia	44				15	13	2	32	25	7
South Africa	45				16	**11**	5	33	29	4
Mexico	46							34	36	-2
Turkey	47							35	32	3
Indonesia	48							36	38	-2
Iran	49				17	15	2	37	35	2
India	50							38	37	1

Notes: **ties** fall back moving forward

TABLE 5.3: Results of Partial Correlations

	League A			League B			League C		
	Correlation Coefficient	p-value	N	Correlation Coefficient	p-value	N	Correlation Coefficient	p-value	N
Kendall's τ_B	.824	.000	23	.583	.001	17	.855	.000	38
Spearman's ρ	.956	.000	23	.785	.000	17	.966	.000	38

it is challenging to strictly divide countries and indicators. However, overlaps show that certain countries can belong to multiple leagues. In fact, countries in overlapped regions may outperform others when considering the elite league's indicators. Still, improvements in several indicators are necessary to differentiate these countries from those in lower leagues (refer to Figure 5.5). Overlap analysis also reveals shared indicators, allowing for comparisons between countries in different leagues.

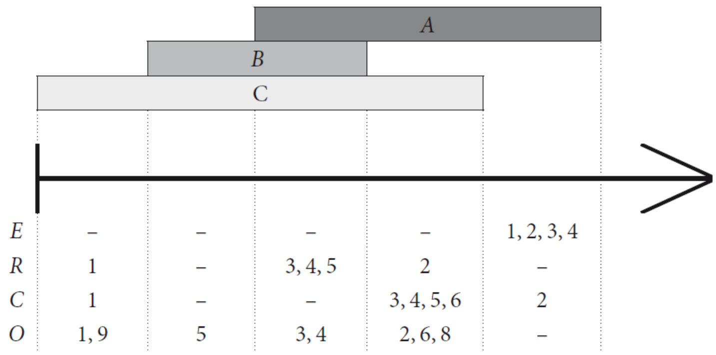


FIGURE 5.5: Opportunities for Development Across Leagues

Figure 5.5 displays the indicators for each league. Policymakers can use this figure to identify the areas they need to improve in order to move up a league.

As shown in Figure 5.4, Hungary falls into the intersection of League B and League C. This suggests that it can be classified at most into the middle range of the countries. Specifically, Hungary shows similarity (homogeneity) in the indicators belonging to League B, but performs worse than the median in the indicators belonging to League C.

In order for Hungary to advance to League A, it is crucial to focus on improving the R3-5 and O3-4 indicators. This entails increasing the amount spent on higher education per student, which may or may not be from government sources (R3). The country should also aim to improve the ratio of higher education R&D expenditures in the share of GDP (R4) as well as per capita expenditure (R5), which will require significant investment in research and development activities in the higher education sector.

Moreover, Hungary needs to increase the number of citations (O3) and it is also important to achieve a higher ranking in the Shanghai Ranking (O4). This will require the country to focus on improving the quality of higher education, research and innovation, and increase investment in these areas. These observations are also true for other countries in League BC, like Bulgaria, Croatia, Romania, and the Czech Republic.

In the following section, the analysis is expanded to include a global university ranking. This demonstrates that the bi-clustering techniques presented can also be used to create leagues of higher education institutions.

5.2 The Leagues of HEIs

In addition to the 0.5 threshold (median) applied for the iBBiG method for bi-clustering the countries, the present section refines the results with other thresholds: 0.75 (upper and lower quartiles) and 0.9 (upper and lower deciles) for leagues A and C.

Table 5.4 summarizes the results as follows: the number of universities and the indicators classified into each league (A, B, C) using different thresholds (0.5, 0.75, 0.9). The higher the threshold, the fewer the universities and indicators entering the leagues. Table 5.4 also indicates the specific indicators included in each league.

At a threshold of 0.5, the indicators marked with "light gray background X" were classified into League A and C. The threshold does not affect the indicators in league B, denoted by X.

Out of the 20 variables:

- i) both League A and C included the same 17 indicators,
- ii) 10 of them are in League B, too.

Finding i) is interesting in two respects. On the one hand, the best institutions are the best in the same indicators as those in which the lagging universities are the worst. On the other hand, 3 indicators were missed from both League A and C (these were not included in League B either[7]):

- T1 "Academic staff per students", which measures the quality of education, as the more lecturers per student, the more effective the education is.
- T2 "Academic staff per bachelor's degrees awarded", which narrows the previous indicator undergraduate-level bachelor's programs because this level is the basis of higher education in the world.
- F3 "Papers per research income", which shows the financing level of the publications.

The 10 indicators in finding ii) are the ones with the lowest variance in the universities included in League B; however, they are decisive in the fact that their high (low) value is required for League A (C) - in addition to 7 other indicators. These 10 variables played a role in the development of all three leagues:

- an interesting finding is that all three reputation surveys were included here:
  - T5 World teaching reputation
  - R5 World research reputation
  - I4 Reputation outside region
- R1 Citations per academic and research staff
- R3 Normalized citation impact
- R4 Papers per academic and research staff
- I1 Share of international academic staff
- I5 International level
- F1 Institutional income per academic staff
- F4 Research income per academic and research staff

TABLE 5.4: The Leagues Formed on RUR 2020

			Leagues		
			A	B	C
		No. of institutions at a threshold of 0.50:	398		430
		0.75:	174	280	192
		0.90:	78		81
		No. of indicators at a threshold of 0.50:	17		17
		0.75:	11	10	15
		0.90:	3		15
INDICATORS					
Teaching (T) 8-8% 40%	T1	Academic staff / students			
	T2	Academic staff / bachelor degrees awarded			
	T3	Doctoral Degrees awarded / academic staff	X		X
	T4	Doctoral degrees awarded / bachelor degrees awarded	X		X
	T5	World teaching reputation	X	X	X
Research (R) 8-8% 40%	R1	Citations / academic and research staff	X	X	X
	R2	Doctoral degrees awarded / admitted PhD	X		X
	R3	Normalized citation impact	X	X	X
	R4	Papers / academic and research staff	X	X	X
	R5	World research reputation	X	X	X
International diversity (I) 2-2% 10%	I1	Share of international academic staff	X	X	X
	I2	Share of international students	X		X
	I3	Share of international co-authored papers	X		X
	I4	Reputation outside region	X	X	X
	I5	International level	X	X	X
Financial sustainability (F) 2-2% 10%	F1	Institutional income / academic staff	X	X	X
	F2	Institutional income / students	X		X
	F3	Papers / research income			
	F4	Research income / academic and research staff	X	X	X
	F5	Research income / institutional income	X		X

Notations of the results of the different thresholds applied in the iBBiG method for determine league A and C:

- X: threshold = 0.5
- threshold = 0.5 and 0.75
- threshold = 0.5, 0.75 and 0.9

In addition to the ten indicators listed above, the high (low) value of 7 indicators determines whether an institution will be placed in League A (or C), i.e., the most important indicators are as follows:

- Doctoral degrees awarded per
  - T3 academic staff
  - T4 bachelor degrees awarded
  - R2 admitted PhD

[7] https://roundranking.com/methodology/methodology.html

- I2 Share of international students
- I3 Share of international co-authored papers
- F2 Institutional income per students
- F5 Research income per institutional income

To RUR's ***League A*** (at the 0.5 threshold), the algorithm assigned 398 institutions. In this League, we can find the Anglo-Saxon countries' most prestigious universities, like Cambridge, Imperial College London, Oxford, and Harvard. These institutions exhibit high scores in all the 17 indicators selected by the method. They have an average score of 0.7 (out of 1.0) on the Teaching (T) indicators with a high score of "World teaching reputation" (T5). They perform well in the Research (R) category as well, and have a high score of the "World research reputation" (R5).

In RUR's ***League C*** (at the threshold of 0.5) there are 430 institutions. Russia is represented by the largest number of universities, accumulating 17% of the institutions. It is followed by China and Iraq with 37 universities. Hungary has three institutions in League C: the Eotvos Lorand University, the University of Szeged, and the University of Debrecen. Eotvos Lorand University secures an impressive 43rd place, positioning it in the upper-middle range of the partial ranking. The University of Szeged holds the 121st place, and the University of Debrecen is at the 154th place. Gwangju Institute of Science and Technology (Republic of Korea) holds the 1st place, Ason University (UK) has the 2nd, and Istanbul Technical University (Turkey) has the 3rd spot.

These institutions show lower average scores for the 17 indicators. They have an average score of 0.26 (out of 1.0) for the "World teaching reputation" (T5), a similar value of the "World research reputation" (R5), and 0.3 for the "Reputation outside the region" (I4).

The ***League B*** of RUR has 280 institutions and 10 indicators. More than 20% of the institutions are from the USA, but notably, Russia also has 40 universities listed in this League. Only one Hungarian university can be found here: the University of Szeged.

California Institute of Technology (USA), Stanford University (USA), Harvard (USA), and Princeton (USA) hold the first four places in this League, and the University of Szeged secures the 75th position. Due to the method, the variances of these institutions and indicators are minimal.

To refine the results, League A and C were also generated to higher thresholds by the iBBiG method. This modifies columns A and C in Table 5.4. League B is not affected by changing the threshold, as it is determined differently (by the BicARE method). At a threshold of 0.75/0.9, the indicators marked with medium/dark gray background X remained in League A, B, and C.

The following focuses only on League A, which contains the best. At the threshold of 0.5, the high value of 17 indicators ensured the classification of an institution in League A, at the threshold of 0.75, 11 of them, and at the threshold of 0.9 only 3. The latter means that if we collect universities in a league with 0-1 normalized data above 0.9, only three indicators will determine the best institutions. These are the three international reputation surveys based on the annual data of the Academic Reputation Survey of Clarivate Analytics (which was implemented by Ipsos Media CT):

- T5 World teaching reputation
- R5 World research reputation
- I4 Reputation outside region: both teaching and research are taken into account, but only respondents' opinions who live outside the university region. The regions considered are as follows: Asia, Europe, North America, Oceania, and South America.

The universities that received at least one vote were included in this survey. Participation in the survey was by invitation only and did not rely on self-reporting. It was not allowed to vote for one's own university. Every year, 10,000 respondents cast 60,000 votes for universities. Each respondent could select up to 15 universities that they deemed the most effective in teaching and research.

The three indicators remaining in the top 10% of League A confirm the dominant role of reputation surveys in the RUR ranking. The reputation of universities is historically very strongly defined and changes very slowly. The well-known elite (the larger Western universities) are like large corporations that remain stable while small companies go out of business or merge with other companies.

5.3 The Case of Hungary

This subsection takes a deeper look at the results associated with Hungary.

In the 2014 U21 ranking Hungary has the 29th place while in 2019 it fell to the 35th place. According to U21's overall 2014-2020 ranking, Hungary is ranked 35th. Regarding the Resources category, its overall rank is 44, for Environment it is 46, for Connectivity it is 18, and 32 for Output. Government expenditure on higher education, as a percentage of GDP, is placed 40th, while total expenditure per student is 27th. Research expenditure as a percentage of GDP earns a rank of 36th. In the Connectivity category, the country achieves the fifth rank in joint publications with industry, although knowledge transfer in business is ranked lower at 32nd. Joint publications with international authors secure a 19th position. In the Output category, Hungary stands third for the tertiary qualifications of the workforce compared to school leavers, ranks 31st for publications per head, and holds the 24th spot for their impact (U21, 2020).

The bi-clustering results position Hungary within League C, as illustrated in Figure 5.6. In this depiction, blue cells denote lower indicator values, while red cells signify higher indicator values. Lower League C encompasses a total of 38 countries. The corresponding Resource score (R_Score), Connectivity score (C_Score), Output score (O_Score), Overall score, and the country's rank are detailed in Table 5.5.

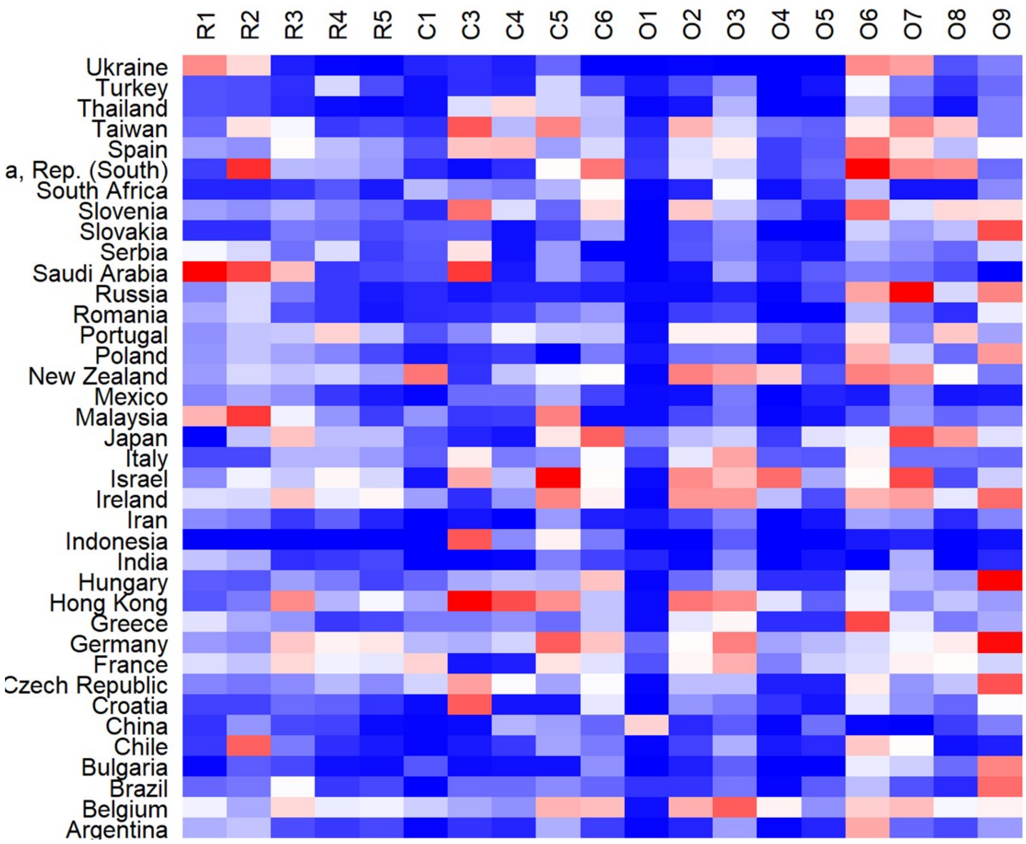


FIGURE 5.6: The Lower League C of U21

Hungary attains a Resource score of 51.32, a Connectivity score of 62.85, Output score of 49.83, and an Overall score of 63.86, securing the 17th position, placing it in the upper-middle range of the League C.

In comparison to neighboring countries based on the Overall score, Hungary is outperformed by the Czech Republic across all scores, while Slovakia, Serbia, and Romania lag behind. Slovakia surpasses Hungary in Output indicators, whereas Serbia excels in Resource indicators.

In the Resources category, Saudi Arabia claims the top spot, followed by Ireland in second place. Belgium, Germany, and France secure the third to fifth positions. Hungary ranks 24th, trailing behind Poland (21st), Greece (22nd), and Brazil (23rd). Serbia, Ukraine, and Slovenia beat Hungary, securing the 15th, 16th, and 18th positions.

For Connectivity, Hong Kong leads, with the Czech Republic in second place. Hungary secures a commendable 11th position. Among neighboring countries, only Slovenia performs better (7th). Croatia ranks 13th, Slovakia 27th, Romania 28th, Serbia 32nd, and Ukraine 37th.

Regarding the Output score, Ireland has the first place, Israel the second. Hungary is ranked in the 24th position. Slovakia is at the 12th place, Slovenia is in the 16th place. Serbia, Romania, and Ukraine underperform Hungary in the Output category.

Hungary's high rank in connectivity and relatively good output suggest a good position in international collaboration, web visibility, and the production of quality research outputs. While Hungary performs well overall, the Environment category, particularly the qualitative policy environment, might be an area for improvement to enhance the overall ecosystem for tertiary education. Continued emphasis on international collaboration and visibility may further boost Hungary's performance.

These conclusions align with the research conducted by Bögel and Mátyás (2020), revealing that Hungary's research performance is commensurate with its size. The study indicates that Hungary allocates a comparatively lower budget

TABLE 5.5: The Partial Ranking of League C of U21

Country	R Score	C Score	O Score	Overall Score	Rank
Ireland	90.66	64.58	100	100	1
Hong Kong	79.01	100	81.94	95.7	2
Israel	81.99	76.11	96.73	94.81	3
Belgium	88.27	75.66	81.37	91.96	4
Germany	87.12	78.1	77.07	90.19	5
New Zealand	75.05	62.88	91.18	90.16	6
Taiwan	64.61	68.23	89.37	87.71	7
France	84.7	54.55	87.08	87.56	8
Korea, Rep. (South)	80.74	47.67	87.44	83.49	9
Japan	75.77	49.71	84.35	82.81	10
Spain	68.56	64.55	82.85	78.96	11
Czech Republic	62.62	78.28	63.92	75.65	12
Malaysia	80.61	48.83	52.7	73.96	13
Slovenia	58.72	65.34	66.24	71.96	14
Portugal	76.36	56.16	66.36	71.8	15
Saudi Arabia	100	52.58	28.4	65.75	16
Hungary	**51.32**	**62.85**	**49.83**	**63.86**	**17**
Italy	57.32	52.98	53.21	60.94	18
Brazil	52.82	33.75	75.11	60.39	19
Croatia	42.02	59.02	53.02	59.98	20
Greece	55.45	46.78	53.53	58.92	21
Russia	47.53	20.35	84.15	56.17	22
Poland	56.38	23.47	61.25	55.59	23
Slovakia	41.51	32.74	81	55.39	24
Serbia	64.57	27.92	35.68	46.12	25
Thailand	28.32	56.04	50.69	45.99	26
Chile	57.69	29.92	36.24	45.07	27
South Africa	32.41	59.83	31.74	43.83	28
Ukraine	64.33	16.88	33.97	41.99	29
Turkey	48.59	33.52	33.37	41.95	30
China	33.72	31.3	48.74	41.52	31
Romania	44.51	32.35	34.51	41.06	32
Mexico	47.44	37.31	22.6	40.73	33
Argentina	47.65	28.62	33.61	38.99	34
Iran	40.19	22.77	45.72	37	35
Indonesia	17.07	54.03	33.58	34.27	36
Bulgaria	27.99	25.5	35.39	32.2	37
India	45.92	15.5	23.43	30.49	38

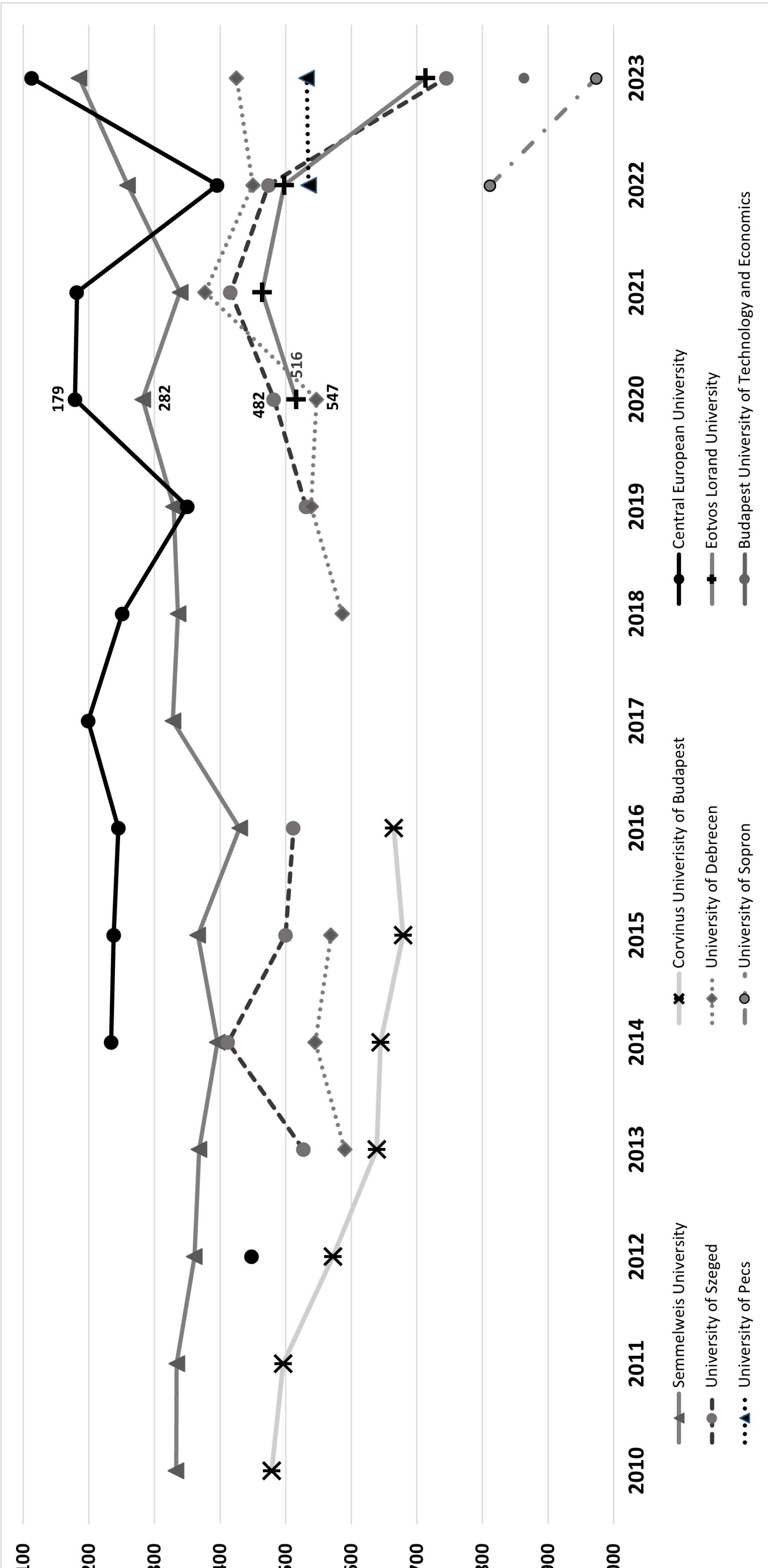
to R&D in comparison to countries of similar size and historical context. To enhance the nation's research capabilities, it is recommended to increase R&D investments and provide greater financial support to researchers.

The RUR ranking has been documenting Hungarian universities since 2010. Figure 5.7 illustrates the trajectory of their rankings from 2010 to 2023. Notably, Semmelweis University has maintained a continuous presence since 2010, progressively achieving higher rankings during this period. Starting in 2020, more Hungarian institutions made their appearance in the ranking. The University of Debrecen managed to advance to the 425th place, and the Central European University, though technically situated in Vienna, climbed to the 113th place in 2023. However, the ranks of the University of Szeged, Eötvös Loránd University, and the University of Sopron declined from 2022 to 2023.

This study demonstrated the league creation on the 2020 RUR ranking, which ranked five Hungarian universities. The individual rankings of each institution in 2020 are depicted in Figure 5.7. Among them, the Central European University performed the best, securing the 179th position, while the University of Debrecen held the lowest rank at 547th.

Regarding the indicators, 17 fall under League A. Notably, "Academic Staff/Students" (T1), "Academic Staff/Bachelor degrees awarded" (T2), and "Papers/Research income" (F3) are not included in League A. A substantial number of universities in Upper League A (with a threshold of 0.5) belong to the United States of America, with the United Kingdom, Germany, and Australia also contributing significantly. Hungary is represented by two institutions: the Central European University and Semmelweis University.

In the partial ranking, which includes only the indicators and universities selected by the bi-clustering method, the Central European University secures the 176th place, while Semmelweis is positioned at 322nd. Comparatively, among Hungary's neighboring countries, Slovenia has one institution (University of Nova Gorica) at the 299th place, and Austria has one university (Medical University of Vienna) at the 156th place. The University of Cambridge (UK) claims



Source: www.roundranking.com

FIGURE 5.7: Ranks of Hungarian Universities in the RUR Ranking, 2010-2023

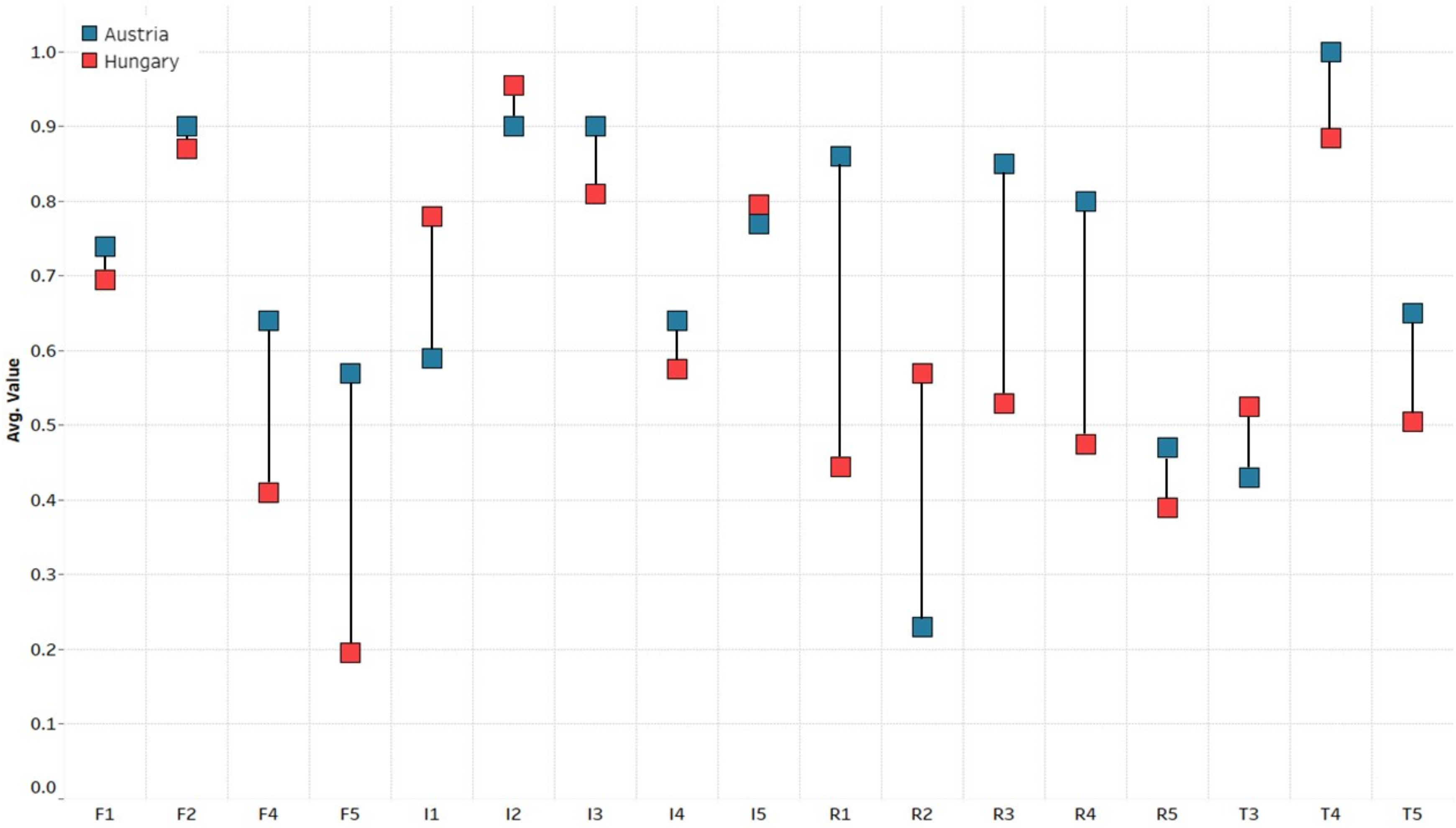
the top spot, followed by Imperial College London (UK) in second place, and Caltech (USA) in third.

Figure 5.8 compares the average scores of the two Hungarian universities and one Austrian university in League A. The red rectangles represent the average scores of the Hungarian universities, and the blue ones are the Austrian ones. In each category, the red and blue rectangles are connected by a black line to show the difference between the average scores. The longer the black line is the bigger the difference.

In terms of the International diversity category (I), the institutions' scores are very similar, and the largest differences can be observed in the Research (R) category. The "International level" (I5) is almost identical which is the average of I1-I4 indicators measuring the share of international staff, international students, and the share of the international co-authored papers. The "Institutional income/students" (F2) is also very similar, meaning that the per capita income for these universities is quite similar.

Turning to the larger differences, the "Citations/academic and research staff" (R1) shows the biggest difference, in favor of Austria. "Research income/institutional income" (F5) has the second largest difference which measures the proportion of the income coming from research to the overall institutional income. Hungary underperforms in this category. Even though it has a high value of the "Institutional income/students" (F2), only a small income is related to research. The score of these two indicators (F2 and F5) indicate that the three examined universities has a very similar number of students and income level, but the Austrian university's income is rather coming from research.

Further notable distinctions exist in the Research category. The "Doctoral degrees awarded/admitted PhD" (R2), the "Normalized citation impact" (R3), and the "Papers/academic and research staff" (R4) also show large differences. These indicators measure the publication performance, taking into account the citations as well. The Austrian university has a higher number of papers compared to the staff, and citation impact.



Where F denotes Financial sustainability indicators, I denotes International diversity indicators, R covers Research indicators, and T stands for Teaching indicators. The smallest and largest differences can be observed in the following indicators:

- F2: Institutional income/students
- I5: International level
- R1: Citations/academic and research staff
- F5: Research income/institutional income

FIGURE 5.8: Average scores of Hungarian and Austrian universities in RUR's League A

Hungary's representation by two institutions in the League shows that these universities exhibit strengths comparable to institutions from more established educational systems, like the UK, or Germany. The top positions in the partial ranking are occupied by globally recognized institutions, which is not a surprise, underlining the competitiveness of the academic landscape. Comparing its performance to Austria, the results show that the universities have larger differences in the research-related areas, whilst they are very similar in the international diversity category.

In the middle League B, there are 280 institutions, and only one Hungarian university can be found here: the University of Szeged. More than 20% of the institutions are from the USA, but notably, Russia also has 40 universities listed in this league. Indicator-wise, the method selected 10 indicators. Only one indicator from the Teaching category which is the reputation metric ("World teaching reputation" - T5). All Research indicators are presented except the "Doctoral degrees awarded/Admitted PhDs" (R2). In the International diversity category, the "Share of international academic staff" (I1), the "Reputation outside of the region" (I4), and the "International level" (I5) are presented. From the Financial sustainability category, the "Institutional income/Academic staff" (F1), and the "Research income/Academic and research staff" (F4) are listed.

In the partial ranking, the University of Szeged secures the 75th position. Neighboring countries such as Ukraine, have four institutions, while Romania and Croatia each have two. Slovenia and the Czech Republic are represented by a single institution. Nevertheless, all universities from neighboring countries lag behind, occupying positions ranging from 86th to 274th. The top four positions are dominated by famous institutions, namely Caltech, Stanford, Harvard, and Princeton.

University of Szeged's representation in this league, in comparison to a diverse set of global peers, demonstrates its solid and resilient reputation, coupled with a significant degree of internationalization.

League C (with a threshold of 0.5) contains most of the institutions, 430. Russia is represented by the largest number of universities, accumulating 17% of the institutions. It is followed by China and Iraq with 37 universities. China's presence in League C is notable, while Iraq has a nearly equivalent number of institutions in League B. Among the indicators, only three metrics are absent in League C. Two pertain to the Teaching category: "Academic staff/Students" (T1) and "Academic Staff/Bachelor degrees awarded" (T2). One originates from the Financial sustainability category: "Papers/Research income" (F3).

Hungary has three institutions in League C: the Eötvös Loránd University, the University of Szeged, and the University of Debrecen. In comparison to surrounding countries, Ukraine has 9 institutions, Romania has 6, while Slovenia, the Czech Republic, and Croatia each have 2, and Slovakia has 1 institution in this league.

Eötvös Loránd University secures an impressive 43rd place, positioning it in the upper-middle range of the partial ranking. The University of Szeged holds the 121st place, and the University of Debrecen is at the 154th place. While they exhibit lower scores in reputation indicators (T5 and R5), they excel in "Share of international co-authored papers" (I3), with Eötvös Loránd University achieving a notable score for "Normalized citation impact" (R3) as well. Compared to institutions in neighboring countries, only Romania's West University of Timisoara has a higher score for R3. On the other hand, in the I3 metric, Kyiv National Economic University from Ukraine has the highest score, followed by another Ukrainian university: South Ukrainian National Pedagogical University.

Overall, Hungary's universities show competitive performances across the three Leagues, with individual strengths in teaching, research, and internationalization. The diverse representation in different leagues reflects a nuanced landscape of academic excellence and areas for improvement, especially in terms of reputation indicators.

Chapter 6

Discussion

One of the main messages of this study is that it is worth comparing universities of those countries, which have similar higher education systems. The results on upper/lower leagues are consistent with the U21 ranking because the countries of the upper league (League A) come from the top of the U21's 50 countries. In the same way, the lower league (League C) covers the bottom of the original U21 ranking.

However, this method points out that League A countries' performance is not uniform in the indicators outside of League A. For example, Norway as a member of the upper league (League A) has only 7.4 score in the field of "proportion of international students (C1)", which is under the median (17.8) of all 50 countries. Similarly, the UK's government expenditure as a percentage of GDP (R1) examples that an upper-league member country is not necessarily excellent in indicators outside its league. UK has 32.0 score under the median of all 50 countries (44.8 score). Saudi Arabia and Indonesia from League C are another good examples that the countries are good (87.5, 78.2) at (upper the median, 67.8) "The proportion of articles co-author with international collaborators (C2)", which is an indicator outside of League C.

In order to confirm the accuracy of the bi-clustering results, it is advisable to compare them with another established ranking. This comparative approach helps to validate the results presented in this work by aligning them with a widely accepted and respected benchmark.

The comparison of U21:2016 and QS:2016 rankings[1] shows that the rank correlation is moderately strong: ρ=0.622, τ_B=0.435. League A is comprised of 23 countries, but Slovenia is not included in the QS ranking. Out of the remaining 22 countries, 17 are included in both League A and the QS 2016 ranking. The QS 2016 ranking does not include 8 countries (Bulgaria, Croatia, Hungary, Iran, Romania, Serbia, Slovakia, Slovenia) from the 38 countries in League C. However, 19 out of 30 of these countries are included in both rankings.

The reason for these differences can be derived from the different raking calculation methods and the partially different considered indicators. QS's overall rank came up from a series of four fields with equal weighting: system strength, access, flagship institution and economic context.[2] Each of them is based on their own QS World University Rankings. The fields of QS ranking can only match two U21 (O1, O4) indicators, which are related to the Shanghai ranking.

League B is the smallest league containing homogeneous countries on the included indicators. It consists of only 18 countries and only 6 indicators, which indicates most countries are inhomogeneous within most indicators.

Taking into account the nations within the specified leagues, the question is how these leagues relate to geographical or economic regions. Figure A.4 in the Appendix shows the geographical distribution of the leagues. The findings lead to a novel categorization that exclusively concentrates on the higher education systems of these nations. The process of bi-clustering results in an alternative classification of these nations, diverging from conventional groupings like economic or geographical ones. From a geographic perspective, each league and a substantial portion of their intersections encompass countries at least three different continents.

A similar assertion can be drawn from an economic standpoint. As an example, BRICS countries are ranked only by QS, despite the outcomes indicating that they belong to a lower league (League C). Nevertheless, two of these countries

[1]QS:2014 are not available. The U21 rankings are very stable. Comparing U21:2014 and U21:2016 rankings, the Spearman's ρ is 0.982, and the Kendall's τ_B is 0.903.

[2]https://www.topuniversities.com/system-strength-rankings/methodology

are already part of the middle league (League B), which aligns with the author's findings. Furthermore, countries like South Africa and Russia resemble middle league (League B) nations due to the similarity in the patterns of indicators that hold significance in that league.

The separation of the countries between the top of League A (League A^+) and the remaining countries (i.e. countries in League C) is caused mostly by the proportion of articles co-authored with international collaborators (C2). It means that if this rate is high - besides the high value of other indicators - it could determine that a country could be a member of the top of the best league (League A^+).

It is interesting, that "other indicators" do not contain the total number of journal articles (O1), the sum of Shanghai scores of the best 3 universities of a country (O5), the proportion of international students (C1), unemployment rates (O9), and one of the most interesting findings: the government expenditure on tertiary education institutions as a percentage of GDP (R1) is excluded from the best league (League A). Whether the value of these five indicators is high or low, it does not matter: if all of the other 19 indicators are high, the country probably would belong to the top (League A). Except for government expenditures (R1), all other resources (R2-5) indicators matter in League A. If the government spends a large percentage of GDP on higher education (R1), it does not pull this country to the best league (League A) instantly, because this rate is high vainly, if the GDP is low. Ukraine from the lower league (League C^-) is a good example of this, because its government spending on higher education in the percentage of GDP is the third highest (78.3 score), but its GDP is low enough. The counterexample is the UK, because UK is at the top of the best league (League A^+), despite its governmental spending measured in the rate of GDP being the smallest (32.0 score).

Three resources (R3-5) and 1 output (O4) indicator are common in all leagues

(League ABC), therefore according to these indicators all countries can be compared. Since both input and output indicators are common in all leagues, countries can also be benchmarked by the effectiveness of their higher education system. A higher value of these four indicators pulls countries toward the best league. The lower value of these four indicators pulls countries towards the lower league. Small variances of these indicators which variances are similar to other indicator's variances that matter in the middle league (O1 and O5) pull countries towards the middle league.

The results of bi-clustering suggested which countries should be compared and ranked. However, this method also shows which countries should be evaluated separately. Two given countries from two different leagues should be compared by only the common indicators. For example, countries like Argentina, Brazil, China or India from the League C^- and the USA or UK from the League A^+ should not be compared or ranked by the "proportion of articles co-authored with international collaborators" (C2). However, countries from League A can be compared to each other. Similarly, when considering the indicator R1: "government expenditure on tertiary education as a % of GDP" is not involved in League A, which can mean that upon a level this indicator does not determine the higher education position, while under a level it is one of the most important of the indicators. The result suggests that the ranking should only include indicators of resource R1 when considering countries in League C.

Analyzing overlaps shows that there are 3 inputs (R3-R5) and 1 output (O4) which are common for all leagues. Therefore, based on these indicators countries can be compared and a global ranking or global benchmarking can be specified. However, the results of this work conform with the suggestion (see Benneworth, 2010; Liu, 2013) that partial rankings should be used instead of global rankings. If more (than 4) indicators should be involved in the comparison, partial rankings should be specified instead of global rankings. However, the bound of partial rankings was an open question. The proposed method specifies bounds and also the set of indicators where the first n or the last m (i.e.

$n = 23$, USA to Slovenia; $m = 38$, Belgium to India) countries can be compared, and partial rankings can be calculated based on the involved indicators.

An important aspect to consider is whether the results obtained are consistent over time. To examine this, the author conducted further analyses on the 2020 dataset of U21 (U21:2020), to compare the changes observed with those in the year 2014 (U21:2014). It is worth noting that 2020 was the final year when the organization responsible for U21 published its country ranking.

In the U21:2020 ranking the number of countries is the same as in the U21:2014 ranking. However, between the 2014-2020 period, several changes were made in the indicators and their weights. New indicators were introduced, for example, E5 which comes from the responses to a survey: „How well does the educational system in your country meet the needs of a competitive economy?" (see more details in Williams and Leahy, 2020).

The bi-clustering method was run on the U21:2020 dataset, searching for League A, League B, and League C. Figure 6.1 juxtaposes the outcomes of the 2014 and 2020 datasets through a Sankey chart. The left side illustrates the 2014 standings, while the right side shows the composition of leagues in 2020. Each side displays the count of countries within each league. The ribbons in the chart signify the transition of countries between leagues, and their width is proportional to the number of countries involved, as indicated by the numerical values on the ribbons. Figure A.5 in the Appendix depicts the results of the bi-clustering and shows the countries and indicators in each league.

A notable contrast in the outcomes of the two years lies in the increased representation of countries in League A^+ in 2020, coupled with a decrease in the count of countries in League C^-. Specifically, the Netherlands, Belgium, and New Zealand successfully climbed up to League A^+ by 2020. In 2014, the Netherlands stood as the sole representative in League AB, while the remaining two countries were situated at the intersection of League A and League C (League AC). The other 11 countries that comprised League A in 2014 maintained their position within the same league in 2020.

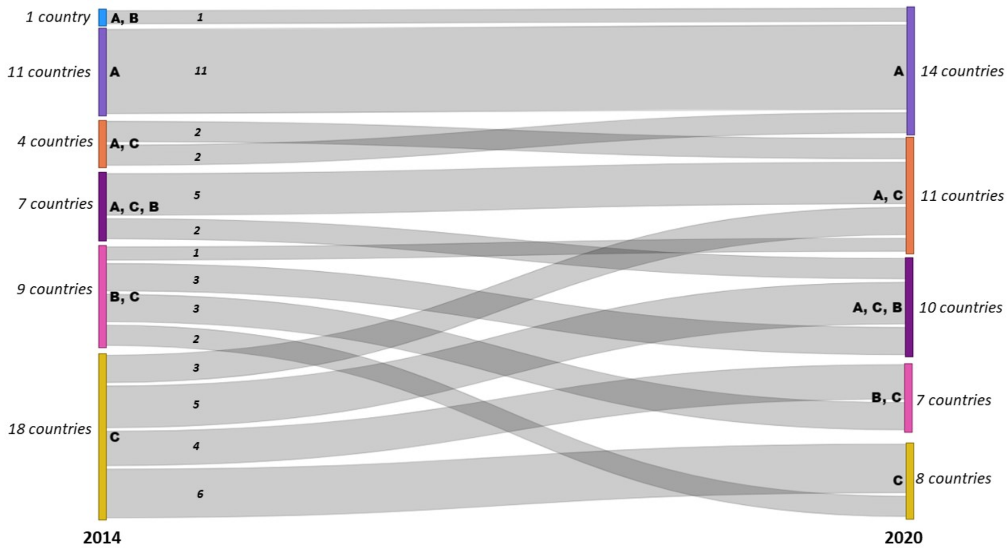


FIGURE 6.1: Comparison of the Results of the Leagues of U21:2014 to U21:2020

Of the 18 countries in League C^- in 2014, 6 countries remained in the same league: Brazil, China, India, Korea, Slovakia, and Turkey. This means they were not successful in developing the governmental expenditure on tertiary education institutions as a share of GDP (R1), in the percentage of international students (C1), in the number of journal articles (O1), nor the unemployment rates (O9).

Out of the 18 countries initially placed in League C^- in 2014, only six nations - Brazil, China, India, Korea, Slovakia, and Turkey - remained in the same league by 2020. This implies their lack of success in advancing across multiple metrics, including governmental expenditure on tertiary education institutions as a share of GDP (R1), the percentage of international students (C1), the number of journal articles (O1), and unemployment rates (O9).

The remaining 12 countries made upward movements to higher leagues. Greece, Japan, and Saudi Arabia moved to League AC due to an increase in the number of articles written collaboratively with international counterparts (C2), engagement with industry researchers (C6), and success in knowledge transfer (C5). Additionally, these countries demonstrated better performance in two other indicators: the average impact of articles (O3) and the share of the population with tertiary qualifications (O7).

Argentina, Chile, Indonesia, Italy, and Ukraine managed to move up to the intersection of the three leagues (League ABC). Mexico, Poland, Serbia, and Thailand moved to League BC, indicating they moved closer to the middle league by 2020.

Romania and South Africa were initially placed in League BC in 2014 and by 2020 they fell back to League C^{-}. This decline is attributed to their diminished scores in various key metrics. In 2020, both countries exhibited lower values for governmental and total expenditure on tertiary education as a share of GDP (R1, R2), a reduced percentage of international students (C1), a decrease in the number of journal articles (O1, O2), a decline in the number of researchers (O8), and higher unemployment rates (O9) when compared to other countries.

The changes that have occurred in the structure of countries' leagues between 2014 and 2020 offer valuable information about the progression of higher education. These changes also demonstrate the usefulness and interpretability of the bi-clustering method. The method can identify significant patterns in the data, which allows for a detailed comprehension of a country's performance. This, in turn, enables researchers to draw conclusions about the factors that influence changes over time.

Looking at the results of the leagues of universities, the findings reveal insights into the underlying reasons that shape the classification of institutions into leagues A, B, and C. The shared indicators between Leagues A and C, along with the stark differences in their performance, point to the critical role these indicators play in distinguishing between top-tier and lower-ranked universities. It suggests that excelling in these specific areas can elevate an institution to League A, while under-performance may lead to placement in League C.

The exclusion of three indicators, namely T1 "Academic staff per students", T2 "Academic staff per bachelor's degrees awarded", and F3 "Papers per research income", from all leagues raises questions about their significance in the ranking process. These indicators are pivotal in assessing the quality of education. Their omission could be attributed to the challenges in accurately quantifying these aspects or the need for further refinement to account for their complexities.

The 10 indicators with low variance in League B but decisive in separating League A and C demonstrate the critical factors influencing institutional ranking. Reputation surveys, international diversity, research impact, and financial performance emerge as key determinants.

The inclusion of reputation surveys underscores the recognition and perception of institutions in the academic world. The fact that all three reputation indicators are part of League A at different cutoffs highlights how important these surveys are. At the highest cutoff (0.9), only these three indicators remain in the top league. This makes it clear that if an institution wants to be in the top league, it really needs to do well in reputation surveys. As the results of reputation surveys hardly change over time (Dill and Soo, 2005; Safón and Docampo, 2020), it is very hard for smaller universities to become competitive in this area.

Meanwhile, internationalization and research indicators contribute significantly to an institution's standing, reflecting the global influence of universities in League A.

The additional 7 indicators further shed light on the factors driving performance differences between League A and C. Doctoral degrees awarded per academic staff, international student and faculty presence, and institutional financial health are revealed as crucial elements in determining an institution's league placement. These factors emphasize the importance of research output, global engagement, and financial sustainability in achieving higher rankings.

Besides creating leagues, the uniqueness of this study is that it also illustrates how various thresholds can influence the outcomes of bi-clustering. Modifying these thresholds allowed for a more intricate refinement of the results and a clearer insight into the indicators that remain or are excluded.

Refining the results with higher thresholds (0.75 and 0.9) for League A and

C leads to fewer indicators entering the leagues, emphasizing the sensitivity of the ranking system to different performance levels.

In order to see the changes in the results, the bi-clustering methods were also performed on the latest available rankings of RUR. The author ran the analyses on both the ranking of 2022 (RUR:2022) and the ranking of 2023 (RUR:2023) datasets. The RUR:2022 ranking still uses the same indicators as in RUR:2020. On the other hand, in 2023, RUR decided to change its three major survey-based reputation indicators to be more objective in their rankings. They "firmly believe that this data is far more valuable and will enable the evaluation of universities' reputation and their influence on society in a more balanced manner" RUR (2023).

The RUR:2020 results revealed that three indicators – namely, academic staff per student (T1), academic staff per bachelor degree awarded (T2), and papers per research income (F3) – remained unassigned to any leagues. This suggested that the RUR:2020 ranking might have been formulated without considering these metrics. However, in the subsequent RUR:2022 results, all indicators found placement within one of the leagues or intersections.

In RUR:2022, T1, T2, and F3 were assigned to League AC alongside 6 other indicators. Out of the 9 metrics within League AC for RUR:2022, 5 were consistent with League AC placements in the RUR:2020 ranking. The values of these 9 metrics within League AC can significantly influence institutions, either elevating or demoting them to upper or lower leagues.

The reason for the inclusion of the initially unassigned T1, T2, and F3 indicators in League AC, as opposed to remaining unassigned, lies in the altered landscape of the RUR:2022 ranking. While the RUR:2020 ranking evaluated 828 institutions, the RUR:2022 dataset expanded to include 1021 HEIs. This increase in both the data values for institutions and the number of institutions in the ranking over the two years likely accounts for the incorporation of the three indicators into League AC.

A significant contrast between the RUR:2020 and RUR:2022 results is the exclusive assignment of one key indicator, institutional income per academic staff (F1), to League A in 2022. In the 2020 result, this was assigned to League ABC. The shift to League A can mean that per academic staff institutional income must be high for institutions to be able to pull into League A, indicating the increasing emphasis on the financial dimension.

The bi-clustering methods were run at different thresholds for the RUR:2022 ranking as well. The RUR:2020 results showed the dominance of the three reputation indicators at the 0.9 threshold, indicating that to become a world-class institution, the high value of these three indicators is required.

The outcome of the RUR:2022 further emphasized the dominance of the reputation surveys. At the 0.9 threshold of League A, four indicators remained. These included the three reputation survey indicators (T5, R5, I4), underscoring the sustained influence of reputation-related metrics. Additionally, a financial metric measuring institutional income per student (F2) found a place among the influential indicators in this league.

Since the RUR:2020 and RUR:2022 bi-clustering results showed the dominance of the reputation-based indicators, the author was curious whether the new three indicators that were introduced to replace them would dominate again or not. The "World teaching reputation" (T5) was changed to "Online visibility" which measures the university's prominence and the frequency with which users access its resources via the Google search engine (RUR 2023). The "World research reputation" (R5) was changed to "Social media visibility" which assesses the university's level of engagement with its audience across key social media platforms like Facebook, Twitter, Instagram, LinkedIn, or YouTube. The "Reputation outside region" (I4) was changed to "New media impact" representing the average number of subscribers to a university's social media resources.

In order to see the connection between the old (RUR:2020) and new indicators (RUR:2023), the author ran a correlation analysis (see Table 6.1). There is a positive moderate correlation between the research reputation (2022:R5) and social media visibility (2023:R5), and also between reputation outside region (2022:I4) and new media impact (2023:I4). Furthermore, there is a strong positive correlation between teaching reputation (2022:T5) and online visibility (2023:T5). These results suggest that institutions with strong reputations tend to have higher social media visibility.

TABLE 6.1: The Correlation between RUR 2022 and 2023 Indicators

	2023:T5	***2023:R5***	***2023:I4***	***2022:T5***	***2022:R5***	***2022:I4***
2023:T5	1					
2023:R5	0.58	1				
2023:I4	0.62	0.92	1			
2022:T5	0.72	0.60	0.61	1		
2022:R5	0.72	0.59	0.60	0.95	1	
2022:I4	0.74881	0.62	0.64	0.92	0.92	1

The bi-clustering results of RUR:2023 for League A showed that these three new indicators remained again in the top league at the threshold of 0.9. This result underscores not just the robustness of the bi-clustering method, but also the importance of these indicators. Even though the measurements were changed to a "more object" method, the bi-clustering, and also the correlation analysis confirm that to become a world-class institution, it is a must to maintain a high level of media visibility. Studies showed that social media has a significant impact on students' decisions when selecting an institution (Constantinides and Stagno, 2012; Gautam and Bahl, 2020). As Generation Z, known for its high reliance on social media (Mude and Undale, 2023), shapes perceptions about institutional reputation based on content observed on platforms such as Facebook, Instagram, YouTube, and others, the significance of media visibility in the academic landscape becomes increasingly evident.

These findings suggest that maintaining a strong presence in the media, particularly on social platforms, is a crucial factor in shaping the perceptions of prospective students. The results emphasize the need for institutions to strategically manage their media visibility to enhance their standing in the competitive landscape of higher education.

Overall, the inclusion and exclusion of indicators in the various leagues based on threshold values offer valuable insights into the crucial factors that shape institutional rankings. It underscores the significance of reputation, internationalization, research impact, and financial aspects in determining an institution's standing within the academic landscape. However, careful consideration of the performance thresholds and the potential limitations of certain indicators is essential to refine the ranking system and provide a more comprehensive evaluation of universities worldwide.

Despite the rankers' intention to introduce more objective indicators, an intriguing trend emerges in the new media visibility indicators – they continue to exhibit a bias towards established, prestigious institutions from the past. Even with attempts to shift the focus from reputation indicators to visibility indicators, the ranking persists in reflecting a preference for historically esteemed institutions. This suggests a persistence of reputation influence in the visibility metrics. Interestingly, this observation prompts the realization that in the eyes of Generation Z, visibility and reputation appear to converge as nearly synonymous metrics. Regardless of the specific indicators employed, the alignment of visibility and reputation underscores their interconnected nature and their joint significance in shaping institutional standing in the evolving landscape of ranking assessments.

As the study pointed out, after the creation of Leagues by the bi-clustering method, partial rankings can be formed. These partial rankings fulfill the fairness criteria because the entities in the Leagues are similar in the nature of the method-selected indicators. In terms of the selected indicators, they perform better than the average *(League A)*, below the average *(League C)*, or share the same characteristics *(League B)*. HEIs and countries belonging to the same League can be objectively compared across the selected indicators.

In discussing the methodologies employed, it is crucial to acknowledge certain limitations inherent in the heuristic nature of the presented methods. While iterative processes can help approach optimal results, ensuring the absolute best outcome (highest homogeneity) is not guaranteed, especially as the original database size increases.

It is crucial to note that the proposed method works best when comprehensive data is available with all indicators for all entities (countries or institutions). Even though the iBBiG method can handle missing values, determining regional rankings requires estimating or specifying all values. Additionally, the BicARE method is sensitive to missing values, so it is necessary to replace them before using the algorithm.

When applying these methods to different datasets representing various rankings, different results are obtained. However, despite these variations, the outcomes remain robust, providing a stable basis for analysis. This consistency across datasets is valuable for examining year-over-year changes and gaining insights into trends.

Chapter 7

Summary

The results obtained for the HESs upper/lower leagues are consistent with the U21 ranking. As stated in Chapter 1, the research questions were focused on determining how to define comparable leagues across countries and institutions. The work has demonstrated that the bi-clustering approach is an effective method for defining these leagues. Moving forward, the interpretation of the bi-clustering results is discussed.

Based on the research conducted, it has been shown that by utilizing suitable bi-clustering techniques, it is possible to identify different leagues of countries' HESs and HEIs effectively. The author has employed the BicARE method to recognize the middle league, i.e., League B, while the upper and lower leagues, namely Leagues A and C, have been identified using the iBBiG method.

Using effective bi-clustering techniques, the author has demonstrated the possibility of a fresh classification system for countries. This classification system differs somewhat from traditional economic groups and significantly from geographic regions (refer to Figure A.4 and Figure A.6). Instead, it highlights new groups that align well with the U21 ranking and illuminates how indicators determine a country's position.

A notable discovery using the U21:2014 dataset is that in League A, all resource indicators (R2-5) are significant, except for government expenditures (R1). This suggests that investing a large portion of GDP in higher education (R1) does not necessarily improve a country's standing in the upper league. However, dedicating more resources (either as a percentage of GDP or per capita) towards higher education and research and development from any source (R2-5) can boost a country's position.

The study showed that by using carefully chosen bi-clustering methods, countries and indicators can be categorized together. This approach generates an unbiased, "fair" ranking of HESs, as it eliminates any intentional pre-selection of indicators. The resulting leagues can be useful in providing a clear understanding of the roles of the obtained indicators.

An extensive examination was carried out on the outcomes of bi-clustering. This process involves assessing countries grouped and identifying those that are distinguished to determine the strengths and weaknesses of Higher Education Systems. By doing so, a critical area that requires intervention could also be discovered (refer to Figure 5.5).

This research also demonstrated a method for creating university leagues, specifically by using Round University Ranking 2020 as an example, and also on the 2022 and 2023 datasets. Bi-clustering algorithms were utilized to generate a more agreeable ranking of universities, grouping them into leagues. The leagues were intentionally composed of universities with comparable profiles and combined indicators based on the similarity of their profiles. This technique functions as a feature selection or dimension reduction method, making it easier to comprehend the university leagues and rankings within them.

This method has already been used to group selected economic faculties into leagues within one country (Italy) (Raponi et al., 2016). They found two different clusters based on the nature of the institutions. One of the clusters contains public universities, while the other cluster has private universities. Compared to this study, the present work's novelty lies in applying this method in two respects. The results were presented, on the one hand, based on international university rankings and, on the other hand, with different thresholds.

Compared to Raponi et al. (2016), this work covers a broader range of countries and institutions as it uses international university rankings that encompass a larger extent of the world's universities. Additionally, the analyses were conducted on rankings from different years to demonstrate the stability of the results and the changes that may have occurred over time.

Based on the 20 indicators of the 828 universities included in the RUR 2020, determining excellence above the median, the high value of 17 indicators is required to enter League A. The most surprising result of the study is that, on a stricter interpretation of excellence (pulling the threshold at the upper decile), the high values of only 3 indicators are enough to enter League A, namely, reputation surveys (T5, R5, I4). As a result of the quantitative analyses, these three qualitative (the three most subjective) indicators proved to be the most important.

The significance of reputation surveys was also revealed when the algorithms were run on the 2022 data, as only three indicators (plus one financial) remained at the upper decile threshold. These indicators were further confirmed to be crucial in the 2023 dataset, even though RUR changed them to "more subjective" metrics measuring media visibility. Despite this change, only these three new metrics remained at the 0.9 threshold, further proving the dominance and importance of reputation surveys.

It is considered unethical for universities to directly influence reputation surveys. Therefore, they should focus on improving the quality of their teaching and research activities to indirectly raise their reputation. Additionally, universities should implement effective media visibility strategies since studies have shown that social media plays a significant role in students' decision-making process when selecting an institution (Constantinides and Stagno, 2012; Gautam and Bahl, 2020; Mude and Undale, 2023).

As the outcome of the dissertation, four theses were defined:

Thesis 1. **The proposed method can simultaneously find homogenous Leagues, containing the maximum possible number of indicators and entities (countries or institutions). The proposed method is capable of identifying three primary types of Leagues.**

Thesis 1. 1. **The Top League (A) includes the maximum number of indicators and entities (countries or institutions) that exhibit performance above a predefined threshold in terms of the selected indicators determined by the method. The Lower League (C), in contrast, contains entities that demonstrate performance below a specific threshold with respect to the method-selected indicators.**

Thesis 1. 2. **The Middle League (B) includes the highest possible number of entities (countries or institutions) that have the same performance level in terms of the indicators selected by the method.**

Thesis 2. **The proposed method is capable of defining overlaps of the Leagues. These intersections contain entities and indicators that are part of multiple Leagues, indicating the strength of these entities across multiple academic domains.**

Thesis 3. **The overlap results assist in establishing a developmental trajectory for entities. As these entities demonstrate strength across various academic domains, focusing on refining appropriate indicators can promote them into higher Leagues.**

Thesis 4. **The partial rankings made on the different Leagues can be considered fair as the entities in the Leagues are similar in nature.**

The research results were published in the following international scientific papers:

1. Zsolt T. Kosztyán, Zsuzsanna Banász, Vivien V. Csányi, and András Telcs (2019). "Rankings or Leagues or rankings on Leagues? - Ranking in fair reference groups". In: *Tertiary Education and Management* 25.4, pp. 289–310. DOI: 10.1007/s11233-019-09028-x. URL: `https://link.springer.com/article/10.1007/s11233-019-09028-x`

2. Zsolt T. Kosztyán, Zsuzsanna Banász, Vivien V. Csányi, and András Telcs (2019). "Felsőoktatási ligák, parciális rangsorok képzése biklaszterezési

eljárásokkal". In: *Közgazdasági Szemle* 9, pp. 905–931. DOI: 10.18414/KSZ.2019.9.905. URL: https://ideas.repec.org/a/ksa/szemle/1861.html

3. Zsuzsanna Banász, Zsolt T. Kosztyán, Vivien V. Csányi, and András Telcs (2022). "University Leagues alongside Rankings". In: *Quality & Quantity* 57.1, pp. 721–736. DOI: 10.1007/s11135-022-01374-0. URL: https://link.springer.com/article/10.1007/s11135-022-01374-0

4. Zsolt T. Kosztyán, Zsuzsanna Banász, Vivien V. Csányi, László Gadár, and András Telcs (2020). "Egyetemi rangsorok tudománymetriai és statisztikai megalapozással". In: *Statisztikai Szemle* 98.8, pp. 930–957. DOI: 10.20311/stat2020.8.hu0930. URL: https://www.ksh.hu/statszemle_archive/all/2020/2020_08/2020_08_930.pdf

Chapter 8

Implications

When comparing countries or universities, the first and most fundamental question is which subjects can be compared and which indicators can be used in the comparison. In this regard, the author believes that the bi-clustering method can play an important role in ranking and benchmarking. Although interpreting bi-clustering is more challenging than explaining the results of traditional clustering, analyzing overlaps and separations provides an opportunity to understand why top countries/institutions are separated from others and why some of the entities belong to more than one league.

The proposed bi-clustering methods can identify common indicators that can be used for global rankings or benchmarks. Even if there is no common indicator, bi-clusters can be specified to define regional or partial rankings. This approach ensures that entities are evaluated based on comparable indicators rather than arbitrarily determined ones from a selected region. By analyzing the results of bi-clustering, one can gain a detailed understanding of countries belonging to the same league or those that are separated. This analysis can help to identify the strengths and weaknesses of a given HES. Additionally, one may uncover a point of necessary intervention (refer to Figure 5.5).

The implications of using the bi-clustering method for ranking and benchmarking countries or universities are significant and offer valuable insights for scholars in the field of higher education and global rankings. The adoption of bi-clustering allows for a more nuanced and sophisticated approach to understanding the factors that contribute to the success or differentiation of institutions.

The analysis of overlaps provides valuable details about why certain top-performing countries or institutions stand out from others. Scholars can delve into the distinctive characteristics and strengths that distinguish these high-performing entities, shedding light on best practices and successful strategies in higher education.

The identification of shared characteristics among entities in lower-performing bi-clusters can highlight areas that need improvement. Scholars can pinpoint weaknesses and challenges faced by specific countries or institutions, leading to informed interventions and targeted efforts to enhance their performance.

The adoption of bi-clustering methods in ranking and benchmarking has the potential to deepen the understanding of higher education systems. It provides scholars with a detailed and comprehensive view of the landscape, enabling them to make evidence-based decisions and recommendations for enhancing the quality and effectiveness of higher education institutions. It also has the potential to shape policies and also can be used in strategic planning.

The results of bi-clustering offer benefits not only to scholars but also to students. Rather than relying on pre-selected indicators that rank all entities uniformly, students can use bi-clustering to compare institutions within the same League. This allows them to identify a group of universities that share their preferred fields of study or research areas and allows for a fair comparison of the institutions.

They can leverage the bi-clustering results for benchmarking purposes as well. Given the high cost of higher education, the ability to check the Leagues enables students to identify institutions that offer the best price-quality balance while meeting all their specific criteria. This empowers students to make well-informed decisions when choosing an educational institution that best suits their requirements.

Bi-clustering and university Leagues offer unique advantages that go beyond traditional ranking methods, providing decision-makers with valuable insights and tools to improve their decision-making processes. The method allows decision-makers to identify the subjects and indicators that can be meaningfully compared across countries or universities. This ensures a more accurate and relevant evaluation of entities, as it focuses on comparable factors rather than arbitrary criteria.

The analysis of overlaps and separations in bi-clusters provides decision-makers with a deeper understanding of the factors that differentiate top-performing countries or institutions from others. This knowledge allows them to recognize the specific strengths and successful strategies employed by high-performing entities, offering valuable insights that can be emulated or adopted to enhance the performance of other institutions or systems.

The approach of creating Leagues with the bi-clustering method recognizes the uniqueness of each institution while still allowing for relevant comparisons and evaluations. Decision-makers can tailor their strategies and interventions based on the results and can apply targeted and effective improvements.

A multitude of opportunities arises when considering the implementation of the suggested bi-clustering techniques. Further research could examine long-run year-over-year changes to gain a better understanding of how the higher education landscape and league formation have evolved over time uncovering trends that contribute to a more nuanced understanding of the academic landscape.

The usage of these techniques is not limited to universities or countries' higher education systems. Another potential area for future research could involve applying these methods to assess the competitiveness of countries. This would help identify groups with similar performance characteristics among different entities.

The techniques employed in this work are not limited to academic purposes alone. The private sector can also derive substantial benefits from their application. One potential area of application involves the identification of customer segments that exhibit similar patterns of interest and behavior. The insights gleaned from such analysis could enable businesses to prepare targeted marketing campaigns that are tailored to the specific needs of these segments. By doing so, businesses can more effectively reach and engage their target audiences, thereby increasing the efficacy of their marketing efforts.

Publications

Most of the introduced methodologies and figures are previously appeared in the scientific articles listed below:

1. Zsolt T. Kosztyán, Zsuzsanna Banász, Vivien V. Csányi, and András Telcs (2019). "Rankings or Leagues or rankings on Leagues? - Ranking in fair reference groups". In: *Tertiary Education and Management* 25.4, pp. 289–310. DOI: 10.1007/s11233-019-09028-x. URL: `https://link.springer.com/article/10.1007/s11233-019-09028-x`

2. Zsolt T. Kosztyán, Zsuzsanna Banász, Vivien V. Csányi, and András Telcs (2019). "Felsőoktatási ligák, parciális rangsorok képzése biklaszterezési eljárásokkal". In: *Közgazdasági Szemle* 9, pp. 905–931. DOI: 10.18414/KSZ.2019.9.905. URL: `https://ideas.repec.org/a/ksa/szemle/1861.html`

3. Zsuzsanna Banász, Zsolt T. Kosztyán, Vivien V. Csányi, and András Telcs (2022). "University Leagues alongside Rankings". In: *Quality & Quantity* 57.1, pp. 721–736. DOI: 10.1007/s11135-022-01374-0. URL: `https://link.springer.com/article/10.1007/s11135-022-01374-0`

4. Zsolt T. Kosztyán, Zsuzsanna Banász, Vivien V. Csányi, László Gadár, and András Telcs (2020). "Egyetemi rangsorok tudománymetriai és statisztikai megalapozással". In: *Statisztikai Szemle* 98.8, pp. 930–957. DOI: 10.20311/stat2020.8.hu0930. URL: `https://www.ksh.hu/statszemle_archive/all/2020/2020_08/2020_08_930.pdf`

Appendix A

Figures

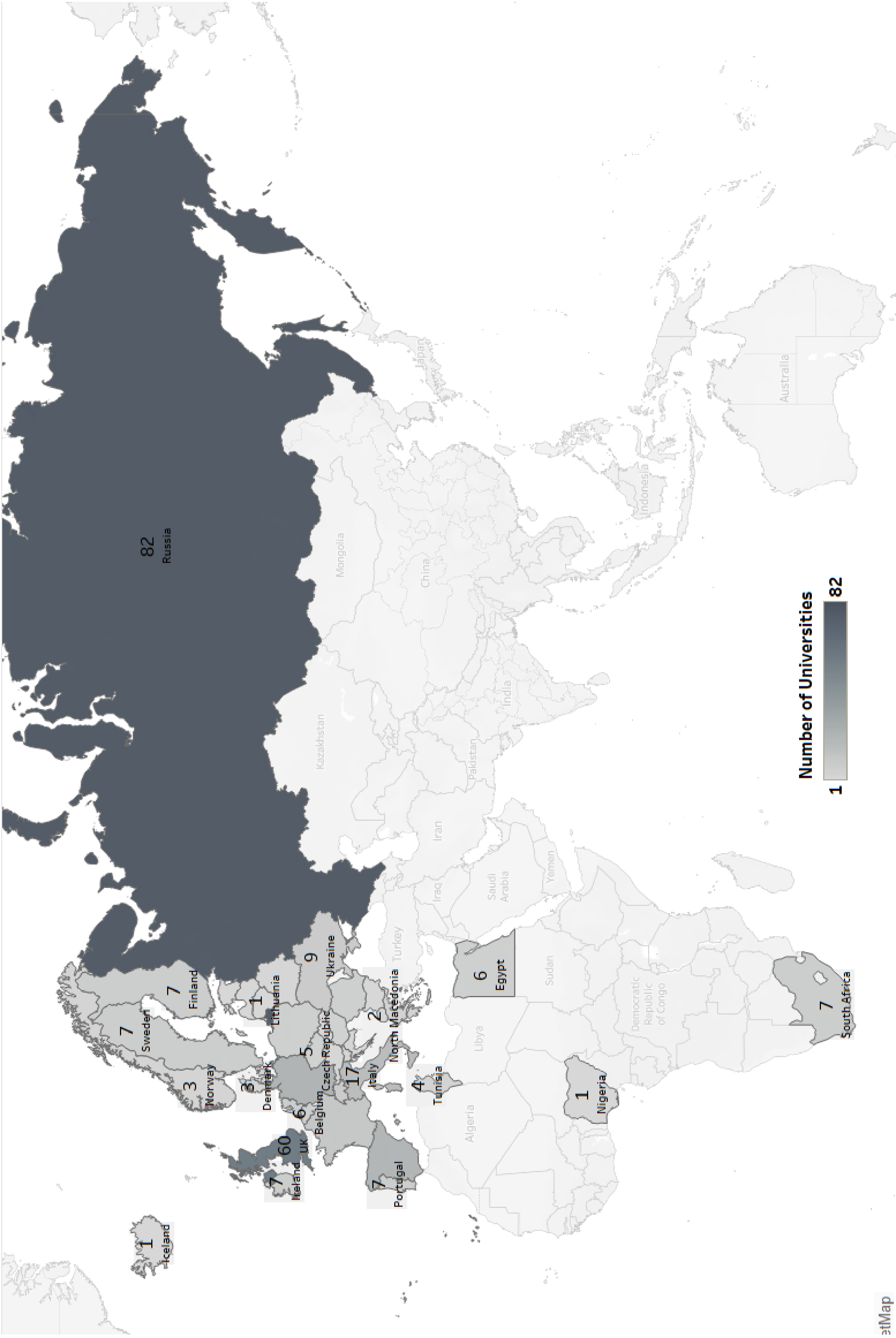


FIGURE A.1: Number of Ranked Universities in RUR Ranking - Europe, the Middle-East, and Africa Region

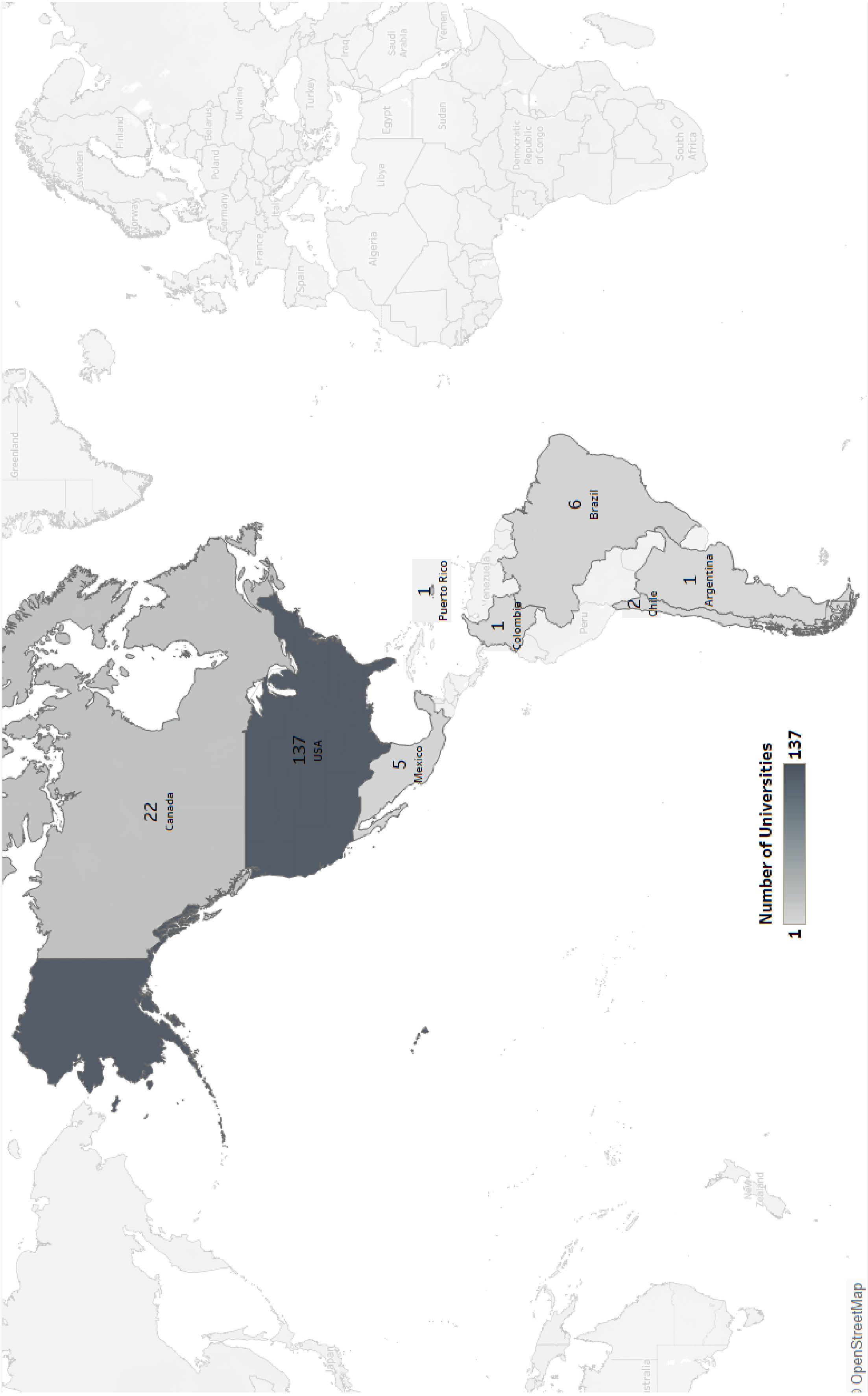


FIGURE A.2: Number of Ranked Universities in RUR Ranking - Americas

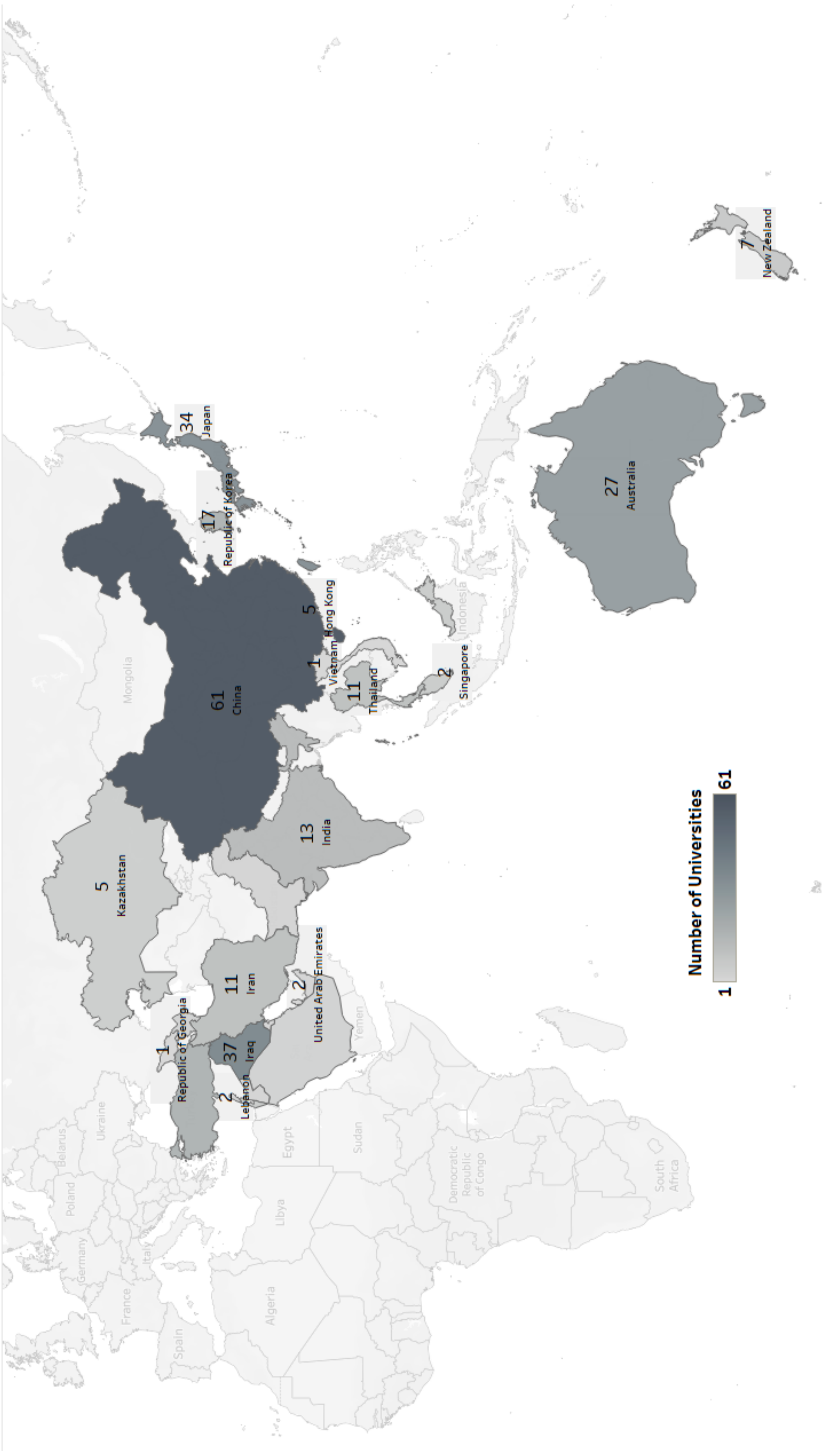


FIGURE A.3: Number of Ranked Universities in RUR Ranking - Asia-Pacific Region

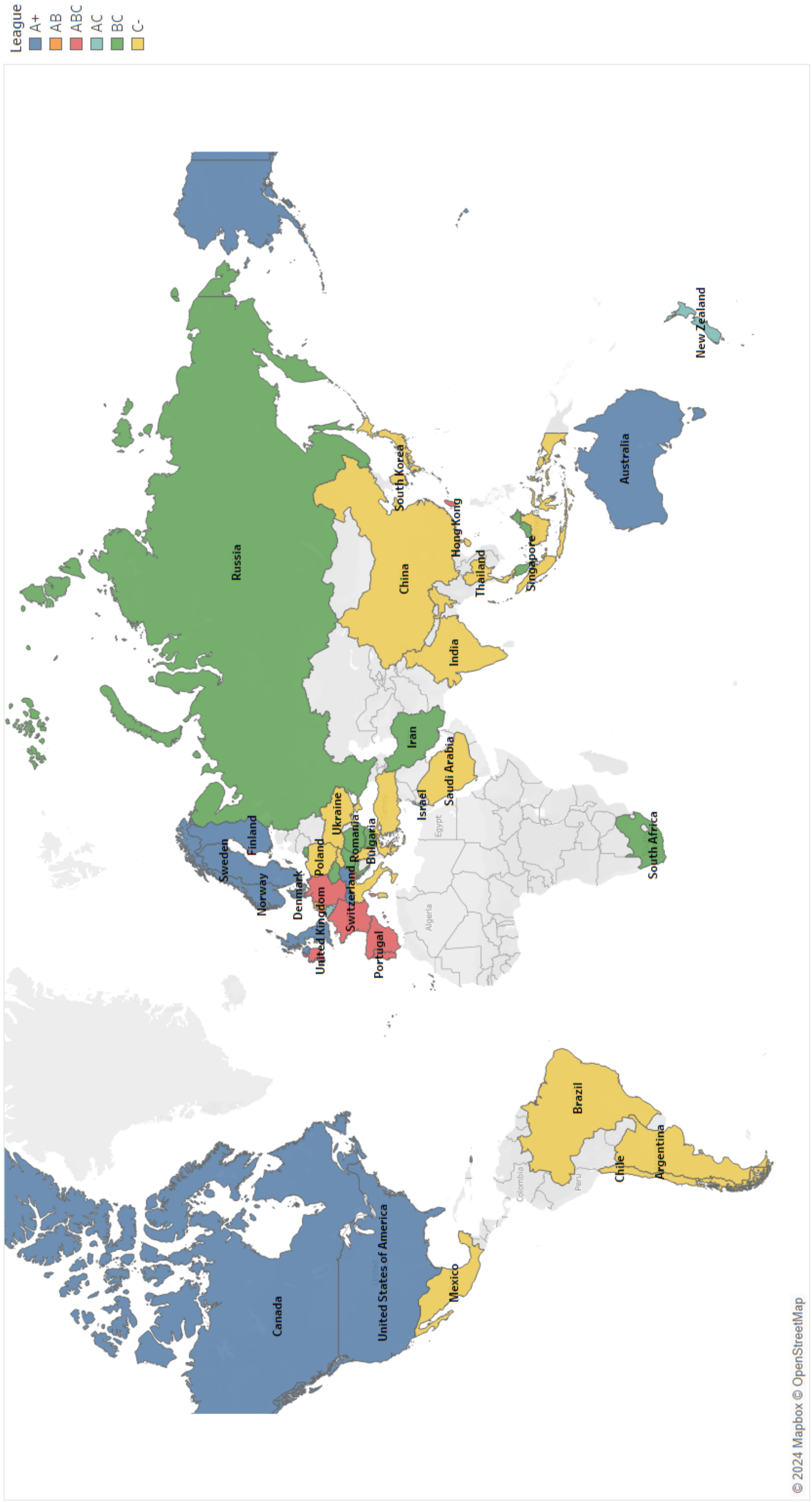


FIGURE A.4: The Geographical Distribution of Leagues of U21:2014

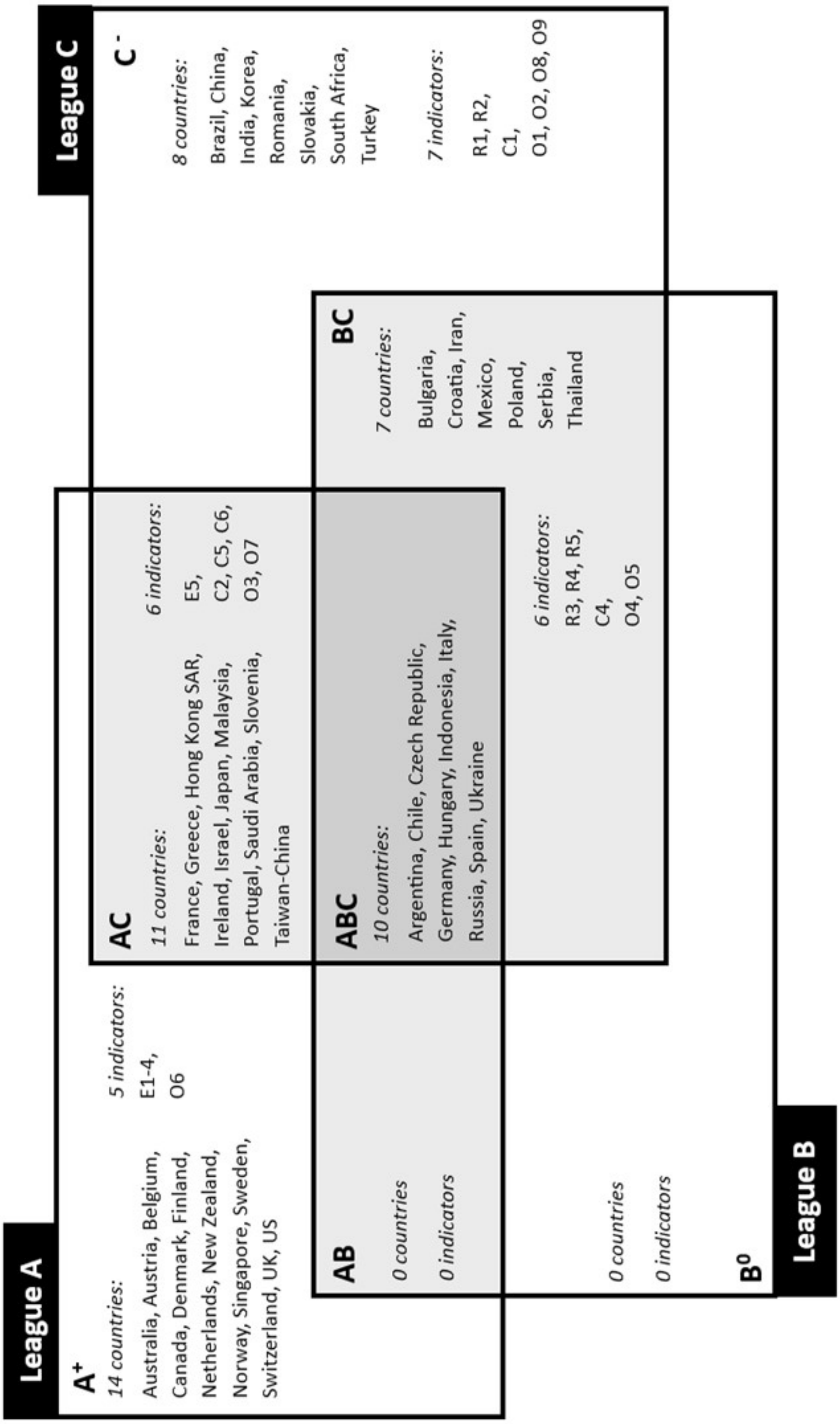


FIGURE A.5: The Leagues Created on U21:2020

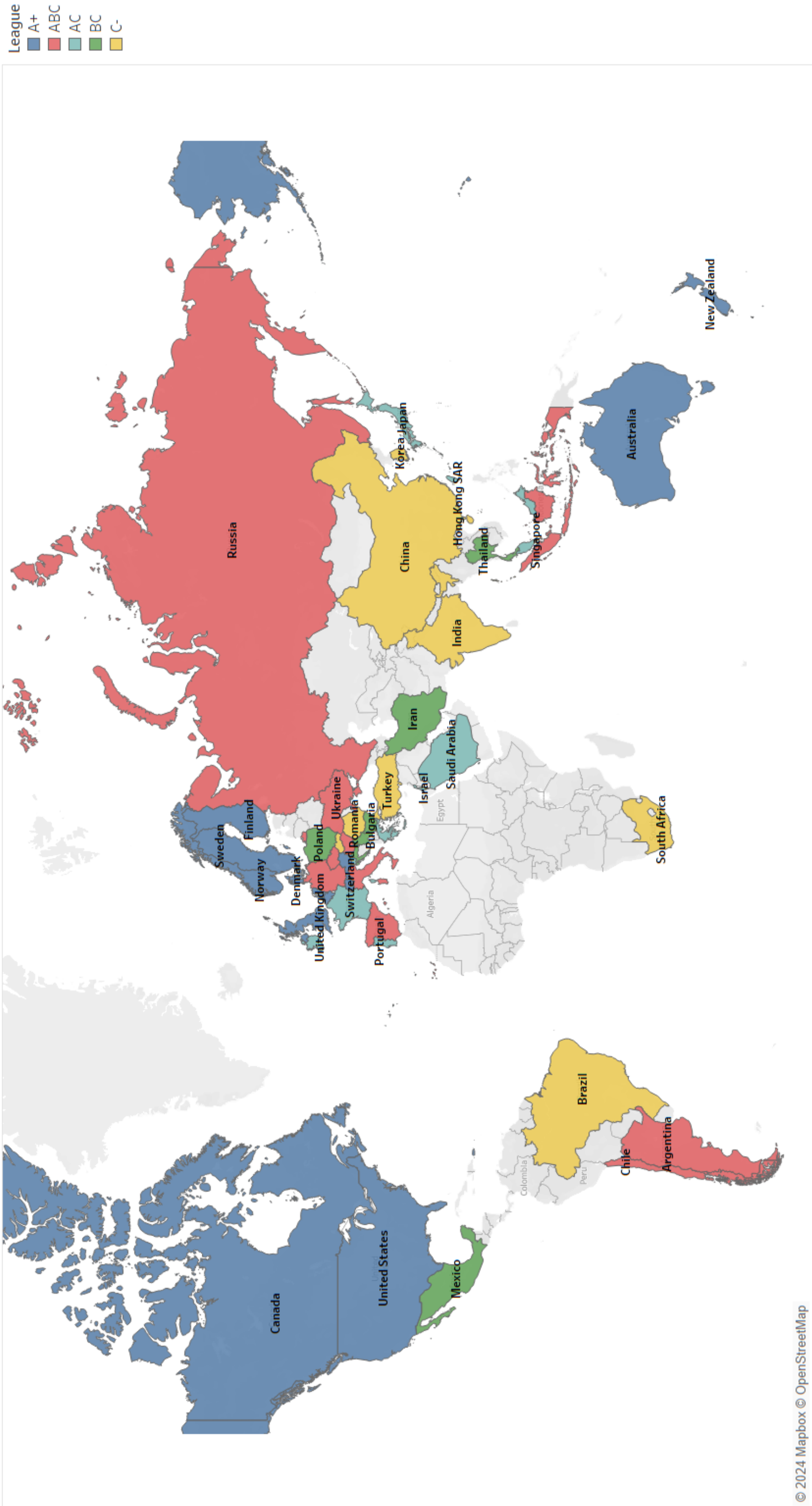


FIGURE A.6: The Geographical Distribution of Leagues of U21:2020

Bibliography

Abankina, I., Aleskerov, F., Belousova, V., Gokhberg, L., Kiselgof, S., Petrushchenko, V., Shvydun, S., and Zinkovsky, K. (2016). "From equality to diversity: Classifying Russian universities in a performance oriented system". In: *Technological Forecasting and Social Change* 103.Supplement C, pp. 228–239. ISSN: 0040-1625. DOI: https://doi.org/10.1016/j.techfore.2015.10.007. URL: http://www.sciencedirect.com/science/article/pii/S0040162515003029.

Abed, V., Koolmees, D. S., Elhage, K., Hessburg, L., and Makhni, E. C. (2022). "Institution Origin and Medical School Rank Impact the Citation Frequency and Publication Rate in Orthopaedic Sports Medicine Journals". In: *Arthroscopy, Sports Medicine, and Rehabilitation* 4.2, e295–e300. ISSN: 2666-061X. DOI: https://doi.org/10.1016/j.asmr.2021.09.005. URL: https://www.sciencedirect.com/science/article/pii/S2666061X21001693.

Academic Influence (2023a). *About AcademicInfluence.com*. URL: https://academicinfluence.com/about#mission.

Academic Influence (2023b). *Influence Ranking Engine*. URL: https://academicinfluence.com/about/influence-ranking-engine.

Altbach, P. (2003). "The Costs and Benefits of World-Class Universities". In: *International Higher Education* 33. DOI: 10.6017/ihe.2003.33.7381. URL: https://ejournals.bc.edu/index.php/ihe/article/view/7381.

Altbach, P. (2006). "The Dilemmas of Ranking". In: *International Higher Education* 42. DOI: 10.6017/ihe.2006.42.7878. URL: https://ejournals.bc.edu/index.php/ihe/article/view/7878.

Amsler, S. S. and Bolsmann, C. (2012). "University ranking as social exclusion". In: *British Journal of Sociology of Education* 33.2, pp. 283–301. DOI: 10.1080/

01425692.2011.649835. eprint: https://doi.org/10.1080/01425692.2011.649835. URL: https://doi.org/10.1080/01425692.2011.649835.

Banász, Zs. (2019). "R&D trends in Hungary". In: *Innovation Challenges in the 21st Century*. Ed. by G. Pintér, Sz. Csányi, and H. Zsiborács. Georgikon Faculty, University of Pannonia, pp. 14–22. URL: https://napok.georgikon.hu/hu/cikkadatbazis/cikkek-2012/cat_view/3-cikkadatbazis/90-2019/98-vi-szekcio-kozgazdasagtan-i.

Barnett, A. G and Moher, D. (2019). "Turning the tables: A university league-table based on quality not quantity". In: *F1000Research* 8.31316755, pp. 583–583. DOI: 10.12688/f1000research.18453.2. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6611132.

Barr, R. S., Durchholz, M., and Seiford, L. M. (2000). *Peeling the DEA Onion: Layering and Rank-Ordering DMUs Using Tiered DEA*. Tech. rep. Southern Methodist University. URL: http://faculty.smu.edu/barr/pubs/layer.pdf.

Becker, W., Saisana, M., Paruolo, P., and Vandecasteele, I. (2017). "Weights and importance in composite indicators: Closing the gap". In: *Ecological Indicators* 80, pp. 12–22. ISSN: 1470-160X. DOI: https://doi.org/10.1016/j.ecolind.2017.03.056. URL: https://www.sciencedirect.com/science/article/pii/S1470160X17301759.

Bell, A. R. and Brooks, C. (2019). "Is There a 'Magic Link' Between Research Activity, Professional Teaching Qualifications and Student Satisfaction?" In: *Higher Education Policy* 32.2, pp. 227–248. DOI: 10.1057/s41307-018-0081-0. URL: https://link.springer.com/article/10.1057/s41307-018-0081-0.

Bengoetxea, E. and Buela-Casal, G. (2013). "The new multidimensional and user-driven higher education ranking concept of the European Union". In: *International Journal of Clinical and Health Psychology* 13.1, pp. 67–73. ISSN: 1697-2600. DOI: https://doi.org/10.1016/S1697-2600(13)70009-7. URL: http://www.sciencedirect.com/science/article/pii/S1697260013700097.

Benneworth, P. S. (2010). *A University Benchmarking Handbook. Benchmarking in European Higher Education*. Ed. by Centre for Higher Education Policy Studies

CHEPS. ESMU. URL: https://research.utwente.nl/en/publications/a-university-benchmarking-handbook-benchmarking-in-european-higher-education(d9fde861-31e3-4177-96bb-5a3dfcd2ebed).html.

Bonaccorsi, A. and Cicero, T. (2016). “Nondeterministic ranking of university departments”. In: *Journal of Informetrics* 10.1, pp. 224–237. ISSN: 1751-1577. DOI: https://doi.org/10.1016/j.joi.2016.01.007. URL: https://www.sciencedirect.com/science/article/pii/S1751157716000031.

Bonaccorsi, A. and Daraio, C. (2008). “The differentiation of the strategic profile of higher education institutions. New positioning indicators based on microdata”. In: *Scientometrics* 74, 15–37. DOI: https://doi.org/10.1007/s11192-008-0101-8.

Borden, V. M. H. and McCormick, A. C. (2020). “Accounting for diverse missions: can classification systems contribute to meaningful assessments of institutional performance?” In: *Tertiary Education and Management* 26.3, pp. 255–264. ISSN: 1573-1936. DOI: 10.1007/s11233-019-09045-w. URL: https://doi.org/10.1007/s11233-019-09045-w.

Bowman, N. A. and Bastedo, M. N. (2011). “Anchoring effects in world university rankings: exploring biases in reputation scores”. In: *Higher Education* 61.4, pp. 431–444. ISSN: 1573-174X. DOI: 10.1007/s10734-010-9339-1. URL: https://doi.org/10.1007/s10734-010-9339-1.

Bowman, N.A. and Bastedo, M.N. (2010). “Anchoring effects in world university rankings: exploring biases in reputation scores”. In: *Higher Education* 61, pp. 431–444.

Boyadjieva, P. (2017). “Invisible higher education: Higher education institutions from Central and Eastern Europe in global rankings”. In: *European Educational Research Journal* 16.5, pp. 529–546. DOI: 10.1177/1474904116681016. eprint: https://doi.org/10.1177/1474904116681016. URL: https://doi.org/10.1177/1474904116681016.

Brankovic, J., Ringel, L., and Werron, T. (Oct. 2022). “Spreading the gospel: Legitimating university rankings as boundary work”. In: *Research Evaluation* 31.4, pp. 463–474. ISSN: 0958-2029. DOI: 10.1093/reseval/rvac035. eprint: https://academic.oup.com/rev/article-pdf/31/4/463/49227625/rvac035.pdf. URL: https://doi.org/10.1093/reseval/rvac035.

Bögel, Gy. and Mátyás, L. (2020). "Kis pénz, nagy teljesítmény? A hazai tudományos kutatás versenyképességéről". In: *Közgazdasági Szemle* 67, pp. 88–102.

Charon, A. and Wauters, J-P. (Oct. 2007). "University ranking: a new tool for the evaluation of higher education in Europe". In: *Nephrology Dialysis Transplantation* 23.1, pp. 62–64. ISSN: 0931-0509. DOI: 10.1093/ndt/gfm279. eprint: https://academic.oup.com/ndt/article-pdf/23/1/62/5420340/gfm279.pdf. URL: https://doi.org/10.1093/ndt/gfm279.

Cheng, Y. and Church, G. M. (2000). "Biclustering of Expression Data". In: *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, pp. 93–103. ISBN: 1-57735-115-0. URL: http://dl.acm.org/citation.cfm?id=645635.660833.

Chirikov, I. (2022). "Does conflict of interest distort global university rankings?" In: *Higher Education*. DOI: https://doi.org/10.1007/s10734-022-00942-5.

Constantinides, E. and Stagno, M. C. Z. (2012). "Higher Education Marketing: A Study on the Impact of Social Media on Study Selection and University Choice". In: *International Journal of Technology and Educational Marketing* 2, pp. 41–58. DOI: https://doi.org/10.4018/ijtem.2012010104.

Daraio, C. and Bonaccorsi, A. (2017). "Beyond university rankings? Generating new indicators on universities by linking data in open platforms". In: *Journal of the Association for Information Science and Technology* 68.2, pp. 508–529. ISSN: 2330-1643. DOI: 10.1002/asi.23679. URL: http://dx.doi.org/10.1002/asi.23679.

Dill, D. D. and Soo, M. (2005). "Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems". In: *Higher Education* 49.4, pp. 495–533. ISSN: 1573-174X. DOI: 10.1007/s10734-004-1746-8. URL: https://doi.org/10.1007/s10734-004-1746-8.

Dolnicar, S., Kaiser, S., Lazarevski, K., and Leisch, F. (2012). “Biclustering: Overcoming data dimensionality problems in market segmentation”. In: *Journal of Travel Research* 51.1, pp. 41–49. DOI: https://doi.org/10.1177/0047287510394192.

Downing, K. (2013). “What’s the use of rankings?” In: ed. by P.T.M. Marope, P.J. Wells, and E. Hazelkorn. Rankings and Accountability in Higher Education: Uses and Misuses. UNESCO. Chap. 11, pp. 197–208.

Ederer, P., Schuller, P., and Willms, S. (2009). “University Systems ranking: citizens and Society in the Age of the Knowledge”. In: *Educational Studies* 3, pp. 169–202.

Eduniversal (2023a). *The Deans’ Vote*. URL: https://www.eduniversal-deans-vote.com/voting-process/prepare-your-voting-session.

Eduniversal (2023b). *The Eduniversal Evaluation System*. URL: https://www.eduniversal-ranking.com/methodology/eduniversal-evaluation-system.html.

Eduniversal (2023c). *The Eduniversal Ranking Methodology*. URL: https://www.eduniversal-ranking.com/methodology/official-selection-of-the-best-business-schools.html.

Eurostat (2021). *Government Expenditure on Education*. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Government_expenditure_on_education#Evolution_of_.27education.27_expenditure_over_1995-2021.

Frey, B. S. and Rost, K. (2010). “Do rankings reflect research quality?” In: *Journal of Applied Economics* 13.1, pp. 1–38. ISSN: 1514-0326. DOI: https://doi.org/10.1016/S1514-0326(10)60002-5. URL: https://www.sciencedirect.com/science/article/pii/S1514032610600025.

Fábri, Gy. (2020). “Van-e kiút az egyetemi rangsorok bűvöletéből?” In: *Magyar Tudomány* 181, pp. 1309–1317.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability; A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman & Co. ISBN: 0716710447.

Gautam, R. D. and Bahl, S. K. (2020). “Measuring the impact of brand reputation through social media on choice of Higher Education”. In: *International Journal of Advanced Research in Engineering and Technology* 11, pp. 105–110. DOI: https://doi.org/10.34218/IJARET.11.12.2020.11.

Genc, E. G. and Basar, O. D. (2019). “Comparison of Country Ratings of Credit Rating Agencies with MOORA Method”. In: *Business and Economics Research Journal* 10.2, pp. 391–404. DOI: 10.20409/berj.2019.175. URL: https://www.berjournal.com/comparison-of-country-ratings-of-credit-rating-agencies-with-moora-method.

Gestraud, P., Brito, I., and Barillot, E. (2014). *BicARE: Biclustering Analysis and Results Exploration*. URL: http://www.bioconductor.org/packages/release/bioc/vignettes/BicARE/inst/doc/BicARE.pdf.

Guarino, C., Ridgeway, G., Chun, M., and Buddin, R. (2005). “Latent Variable Analysis: A New Approach to University Ranking”. In: *Higher Education in Europe* 30.2, pp. 147–165. DOI: 10.1080/03797720500260033. eprint: https://doi.org/10.1080/03797720500260033. URL: https://doi.org/10.1080/03797720500260033.

Gusenleitner, D., Howe, E. A., Bentink, S., Quackenbush, J., and Culhane, A. C. (2012). “iBBiG: iterative binary bi-clustering of gene sets”. In: *Bioinformatics* 28.19, pp. 2484–2492. DOI: 10.1093/bioinformatics/bts438. eprint: /oup/backfile/content_public/journal/bioinformatics/28/19/10.1093_bioinformatics_bts438/1/bts438.pdf. URL: +http://dx.doi.org/10.1093/bioinformatics/bts438.

Hahsler, M., Hornik, K., and Buchta, C. (2008). “Getting things in order: an introduction to the R package seriation”. In: *Journal of Statistical Software* 25.3, pp. 1–34.

Hanushek, E. A. and Wößmann, L. (2010). “Education and Economic Growth”. In: *International Encyclopedia of Education*. Ed. by Penelope Peterson, Eva Baker, and Barry McGaw. Vol. 2. Oxford: Elsevier, pp. 245–252.

Hartigan, J. A. (1972). "Direct Clustering of a Data Matrix". In: *Journal of the American Statistical Association* 67.337, pp. 123–129. DOI: 10.1080/01621459.1972.10481214. eprint: http://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1972.10481214. URL: http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1972.10481214.

Hazelkorn, E. (2009a). "Impact of Global Rankings on Higher Education Research and the Production of Knowledge". In: *UNESCO Forum on Higher Education, Research and Knowledge* Occasional Paper No. 15.

Hazelkorn, E. (Jan. 2009b). "Rankings and the battle for world-class excellence: Institutional strategies and policy choices". In: *Higher Education Management and Policy* 21, pp. 4–4.

Hazelkorn, E. (2015). *Rankings and the reshaping of higher education: The battle for world-class excellence*. Springer.

Hazelkorn, E. and Ryan, M. (2016). "The Impact of University Rankings on Higher Education Policy in Europe: a Challenge to Perceived Wisdom and a Stimulus for Change". In: *The Globalisation Challenge for European Higher Education*. Ed. by Pavel Zgaga, Ulrich Teichler, and John Brennan. Peter Lang Verlag.

Hrubos, I. (2010). "Régi-új megközelítések a felsőoktatási intézmények sokféleségének és teljesítményének értelmezésében". In: *Felsőoktatási Műhely* 4, pp. 15–24.

Hrubos, I. (2014). "Verseny - értékelés - rangsorok". In: *Educatio* 4, pp. 541–549.

Huang, Q. (2011). "Discovery of time-inconsecutive co-movement patterns of foreign currencies using an evolutionary biclustering method". In: *Applied Mathematics and Computation* 218.8, pp. 4353–4364. ISSN: 0096-3003. DOI: https://doi.org/10.1016/j.amc.2011.10.011. URL: http://www.sciencedirect.com/science/article/pii/S0096300311012410.

Hägg, I. and Wedlin, L. (2013). "Standards for quality? A critical appraisal of the Berlin Principles for international rankings of universities". In: *Quality in Higher Education* 19.3, pp. 326–342. DOI: 10.1080/13538322.2013.852708. eprint: https://doi.org/10.1080/13538322.2013.852708. URL: https://doi.org/10.1080/13538322.2013.852708.

Ibáñez, A., Larrañaga, P., and Bielza, C. (2013). "Cluster methods for assessing research performance: exploring Spanish computer science". In: *Scientometrics* 97.3, pp. 571–600. ISSN: 1588-2861. DOI: 10.1007/s11192-013-0985-9. URL: https://doi.org/10.1007/s11192-013-0985-9.

IREG (2006). *Berlin Principles on Ranking of Higher Education Institutions. International Ranking Expert Group*. International Ranking Expert Group. URL: https://www.che.de/downloads/Berlin_Principles_IREG_534.pdf.

IREG (2023). *IREG Seal of Approval*. URL: https://ireg-observatory.org/en/initiatives/ranking-seal-of-approval/ (visited on 03/12/2023).

Ishikawa, M. (2009). "University Rankings, Global Models, and Emerging Hegemony: Critical Analysis from Japan". In: *Journal of Studies in International Education* 13.2, pp. 159–173. DOI: 10.1177/1028315308330853. eprint: https://doi.org/10.1177/1028315308330853. URL: https://doi.org/10.1177/1028315308330853.

Iñiguez, G., Pineda, C., Gershenson, C., and Barabási, A. (2022). "Dynamics of ranking". In: *Nature Communications* 13.1646. DOI: https://doi.org/10.1038/s41467-022-29256-x.

Jacqmin, J. (2021). "Do ads influence rankings? Evidence from the higher education sector". In: *Education Economics* 29.5, pp. 509–526. DOI: 10.1080/09645292.2021.1918642. eprint: https://doi.org/10.1080/09645292.2021.1918642. URL: https://doi.org/10.1080/09645292.2021.1918642.

Jarocka, M. (2012). "University ranking systems-from league table to homogeneous groups of universities". In: *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering* 6.6, pp. 1377–1382. URL: https://pdfs.semanticscholar.org/b892/8175625ae9e2e5d0da60f3b512ec497579cb.pdf.

Johnes, G. (1989). "Ranking University Departments: Problems and Opportunities". In: *Politics* 9.2, pp. 16–22. DOI: 10.1111/j.1467-9256.1989.tb00253.x. eprint: https://doi.org/10.1111/j.1467-9256.1989.tb00253.x. URL: https://doi.org/10.1111/j.1467-9256.1989.tb00253.x.

Johnes, J. (2018). "University rankings: What do they really show?" In: *Scientometrics* 115, pp. 585–606. URL: https://link.springer.com/article/10.1007%2Fs11192-018-2666-1.

Kosztyán, Zs. T., Banász, Zs., Csányi, V. V., Neumanné, V. I., and Telcs, A. (2019). "Examining the mobility of higher education applicants by economic network models". In: *Statisztikai Szemle* 97.11, pp. 1007–1049. DOI: https://doi.org/10.20311/stat2019.11.hu1007. URL: http://www.ksh.hu/statszemle_archive/all/2019/2019_11/2019_11_1007.pdf.

Kováts, G. (2015). ""New" Rankings on the Scene: The U21 Ranking of National Higher Education Systems and U-Multirank". In: *The European Higher Education Area: Between Critical Reflections and Future Policies*. Ed. by Adrian Curaj, Liviu Matei, Remus Pricopie, Jamil Salmi, and Peter Scott. Cham: Springer International Publishing, pp. 293–311. ISBN: 978-3-319-20877-0. DOI: 10.1007/978-3-319-20877-0_20. URL: https://doi.org/10.1007/978-3-319-20877-0_20.

Lawrence, J. K. and Green, K. C. (1980). *A Question of Quality: The Higher Education Ratings Game*. Tech. rep. Education Resources Information Center (ERIC). URL: https://eric.ed.gov/?id=ED192667.

Lepori, B. (2021). "The heterogeneity of European Higher Education Institutions: a configurational approach". In: *Studies in Higher Education* 0.0, pp. 1–17. DOI: 10.1080/03075079.2021.1968368. eprint: https://doi.org/10.1080/03075079.2021.1968368. URL: https://doi.org/10.1080/03075079.2021.1968368.

Liiv, I. (2010). "Seriation and matrix reordering methods: An historical overview". In: *Statistical Analysis and Data Mining* 3.2, pp. 70–91. ISSN: 1932-1872. DOI: 10.1002/sam.10071. URL: http://dx.doi.org/10.1002/sam.10071.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. second. Wiley Series in Probability and Statistics. Hoboken, New Jersey: John Wiley & Sons, Inc. ISBN: 0471183865.

Liu, N. C. (2013). "The Academic Ranking of World Universities and its future direction". In: ed. by P.T.M. Marope, P.J. Wells, and E. Hazelkorn. Rankings and Accountability in Higher Education: Uses and Misuses. UNESCO. Chap. 1, pp. 23–39.

Liu, N. C. and Cheng, Y. (2005). "The Academic Ranking of World Universities". In: *Higher Education in Europe* 30.2, pp. 127–136. DOI: 10.1080/03797720500260116. eprint: https://doi.org/10.1080/03797720500260116. URL: https://doi.org/10.1080/03797720500260116.

Liu, S., Chen, Y., Yang, M., and Ding, R. (2009). "Bicluster Algorithm and Used in Market Analysis". In: *2009 Second International Workshop on Knowledge Discovery and Data Mining*, pp. 504–507. DOI: 10.1109/WKDD.2009.224.

Lukman, R., Krajnc, D., and Glavič, P. (2010). "University ranking using research, educational and environmental indicators". In: *Journal of Cleaner Production* 18.7. Going beyond the rhetoric: system-wide changes in universities for sustainable societies, pp. 619–628. ISSN: 0959-6526. DOI: https://doi.org/10.1016/j.jclepro.2009.09.015. URL: https://www.sciencedirect.com/science/article/pii/S0959652609003047.

Lynch, K. (2014). "New managerialism, neoliberalism and ranking". In: *Ethics in Science and Environmental Politics* 13.2, pp. 141–153. DOI: https://doi.org/10.3354/esep00137.

Madeira, S. C. and Oliveira, A. L. (2004). "Biclustering Algorithms for Biological Data Analysis: A Survey". In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 1.1, pp. 24–45. ISSN: 1545-5963. DOI: 10.1109/TCBB.2004.2. URL: http://dx.doi.org/10.1109/TCBB.2004.2.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). "A contribution to the empirics of economic growth". In: *Quarterly Journal of Economics* 107.2, pp. 407–437.

Marginson, S. (2009). "Open Source Knowledge and University Rankings". In: *Thesis Eleven* 96.1, pp. 9–39. DOI: 10.1177/0725513608099118. eprint: https://doi.org/10.1177/0725513608099118. URL: https://doi.org/10.1177/0725513608099118.

Marginson, S. (2011). "Higher Education and Public Good". In: *Higher Education Quarterly* 65, pp. 411–433.

Marginson, S. and van der Wende, M. (2007). "To Rank or To Be Ranked: The Impact of Global Rankings in Higher Education". In: *Journal of Studies in International Education* 11.3-4, pp. 306–329. DOI: 10.1177/1028315307303544. eprint: https://doi.org/10.1177/1028315307303544. URL: https://doi.org/10.1177/1028315307303544.

Marginson, Simon (2007). "Global University Rankings: Implications in general and for Australia". In: *Journal of Higher Education Policy and Management* 29. DOI: 10.1080/13600800701351660.

Mihályi, P. (2020). "Hol tart a világ az egyetemi rangsorokról való gondolkodásban?" In: *Magyar Tudomány* 181, pp. 1298–1308.

Mirkin, B. (1998). "Mathematical classification and clustering: From how to what and why". In: *Classification, data analysis, and data highways*. Springer, pp. 172–181.

Moed, H. F. (2017). "A critical comparative analysis of five world university rankings". In: *Scientometrics* 110, pp. 967–990.

Mude, G. and Undale, S. (2023). "Social Media Usage: A Comparison Between Generation Y and Generation Z in India". In: *International Journal of E-Business Research* 19, pp. 1–20. DOI: https://doi.org/10.4018/IJEBR.317889.

Nolle, L. (2010). "Cluster-Based Benchmarking of Universities as an Alternative to League Tables". In: *Research and Development in Intelligent Systems*. Ed. by Bramer M., Ellis R., and Petridis M. Vol. XXVI. Oxford: Springer, pp. 499–504. URL: https://link.springer.com/chapter/10.1007/978-1-84882-983-1_40.

Olcay, G. A. and Bulu, M. (2017). "Is measuring the knowledge creation of universities possible?: A review of university rankings". In: *Technological Forecasting and Social Change* 123.Supplement C, pp. 153–160. ISSN: 0040-1625. DOI: https://doi.org/10.1016/j.techfore.2016.03.029. URL: http://www.sciencedirect.com/science/article/pii/S004016251630021X.

Petrarca, F. and Terzi, S. (2018). "The Global Competitiveness Index: an alternative measure with endogenously derived weights". In: *Quality & Quantity* 52, pp. 2197–2219. DOI: https://doi.org/10.1007/s11135-017-0655-8.

Pontes, B., Giráldez, R., and Aguilar-Ruiz, J. S. (2015). "Biclustering on expression data: A review". In: *Journal of Biomedical Informatics* 57.Supplement C, pp. 163–180. ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2015.06.028. URL: http://www.sciencedirect.com/science/article/pii/S1532046415001380.

Poole, S. M., Levin, M. A., and Elam, K. (2017). "Getting out of the rankings game: a better way to evaluate higher education institutions for best fit". In: *Journal of Marketing for Higher Education* 0.0, pp. 1–20. DOI: 10.1080/08841241.2017.1311981. eprint: https://doi.org/10.1080/08841241.2017.1311981. URL: https://doi.org/10.1080/08841241.2017.1311981.

QS (2016). *Higher Education System Strength Rankings - a ranking of national higher education system*. International Ranking Expert Group. URL: https://www.topuniversities.com/system-strength-rankings/2016.

QS (2018). *Higher Education System Strength Rankings - a ranking of national higher education system*. URL: https://www.topuniversities.com/system-strength-rankings/2018.

QS (2023a). *Academic Reputation*. URL: https://support.qs.com/hc/en-gb/articles/4405952675346.

QS (2023b). *Employer Reputation*. URL: https://support.qs.com/hc/en-gb/articles/4407794203410.

QS (2023c). *QS Stars University Ratings*. URL: https://www.topuniversities.com/qs-stars#sorting=overall+country=+rating=+order=desc+orderby=uni+search=.

QS (2023d). *QS Stars University Ratings - How to achieve a 5 Star Rating?* URL: https://content.qs.com/qsiu/FAQ_leaflet.pdf.

QS (2023e). *QS Stars University Ratings Methodology*. URL: https://www.topuniversities.com/qs-stars/qs-stars-methodology.

Rad, A., Naderi, B., and Soltani, M. (2011). "Clustering and ranking university majors using data mining and AHP algorithms: A case study in Iran". In: *Expert Systems with Applications* 38.1, pp. 755–763. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2010.07.029. URL: http://www.sciencedirect.com/science/article/pii/S0957417410006469.

Raponi, V., Martella, F., and Maruotti, A. (2016). "A Biclustering Approach to University Performances: An Italian Case Study". In: *Journal of Applied Statistics* 43.1, pp. 31–45. DOI: 10.1080/02664763.2015.1009005. eprint: https://doi.org/10.1080/02664763.2015.1009005. URL: https://doi.org/10.1080/02664763.2015.1009005.

Rauhvargers, A. (2013). *Global university rankings and their impact: Report II*. European University Association Brussels.

Romer, P. M. (1990). "Endogenous Technological Change". In: *Journal of Political Economy* 98.5, S71–S102. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/2937632.

RUR (2023). *RUR Methodology*. URL: https://roundranking.com/methodology/methodology.html.

Safón, V. and Docampo, D. (2020). "Analyzing the impact of reputational bias on global university rankings based on objective research performance data: the case of the Shanghai Ranking (ARWU)". In: *Scientometrics* 125, 2199–2227. DOI: https://doi.org/10.1007/s11192-020-03722-z.

Saisana, M., d'Hombres, B., and Saltelli, A. (2011). "Rickety numbers: Volatility of university rankings and policy implications". In: *Research Policy* 40.1. Special Section on Heterogeneity and University-Industry Relations, pp. 165–177. ISSN: 0048-7333. DOI: https://doi.org/10.1016/j.respol.2010.09.003. URL: http://www.sciencedirect.com/science/article/pii/S0048733310001812.

Salmi, J. (2013). "If ranking is the disease, is benchmarking the cure?" In: ed. by P.T.M. Marope, P.J. Wells, and E. Hazelkorn. Rankings and Accountability in Higher Education: Uses and Misuses. UNESCO. Chap. 13, pp. 235–255.

Scheffer, J. (2002). "Dealing with missing data". In: *Research Letters in the Information and Mathematical Sciences* 3, pp. 153–160.

Shin, J. C. and Harman, G. (2009). "New challenges for higher education: global and Asia-Pacific perspectives". In: *Asia Pacific Education Review* 10, 1–13. DOI: https://doi.org/10.1007/s12564-009-9011-6.

Soh, K. (2011a). "Don't Read University Rankings Like Reading Football League Tables: Taking a Close Look at the Indicators". In: *Higher Education Review* 44, pp. 15–29.

Soh, K. (2011b). "Mirror, mirror on the wall: a closer look at the top ten in university rankings". In: *European Journal of Higher Education* 1.1, pp. 77–83. DOI: 10.1080/21568235.2011.577179. eprint: https://doi.org/10.1080/21568235.2011.577179. URL: https://doi.org/10.1080/21568235.2011.577179.

Soh, K. (2014). "Nominal versus attained weights in Universitas 21 Ranking". In: *Studies in Higher Education* 39.6, pp. 944–951. DOI: 10.1080/03075079.2012.754866. eprint: https://doi.org/10.1080/03075079.2012.754866. URL: https://doi.org/10.1080/03075079.2012.754866.

Soh, K. (2017). "The seven deadly sins of world university ranking: a summary from several papers". In: *Journal of Higher Education Policy and Management* 39.1, pp. 104–115. DOI: 10.1080/1360080X.2016.1254431. eprint: https://doi.org/10.1080/1360080X.2016.1254431. URL: https://doi.org/10.1080/1360080X.2016.1254431.

Sowter, B., Hijazi, S., and Reggio, D. (2017). "Ranking World Universities". In: ed. by Kevin Downing, Fraide A., and Jr. Ganotice. IGI Global. Chap. A Decade of Refinement, and the Road Ahead, pp. 1–24. ISBN: 9781522508199. DOI: 10.4018/978-1-5225-0819-9.ch001. URL: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-5225-0819-9.ch001.

Tanay, A., Sharan, R., and Shamir, R. (2002). "Discovering statistically significant biclusters in gene expression data". In: *Bioinformatics* 18.suppl_1, S136–S144. DOI: 10.1093/bioinformatics/18.suppl_1.S136. eprint: /oup/backfile/

content_public/journal/bioinformatics/18/suppl_1/10.1093/bioinformatics/18.suppl_1.s136/2/18s136.pdf. URL: http://dx.doi.org/10.1093/bioinformatics/18.suppl_1.S136.

Teichler, U. (2008). "Diversification? Trends and Explanations of the Shape and Size of Higher Education". In: *Higher Education* 56.3, pp. 349–379. ISSN: 00181560, 1573174X. URL: http://www.jstor.org/stable/40269081 (visited on 04/12/2023).

Tofallis, C. (2012). "A different approach to university rankings". In: *Higher Education* 63, 1–18. DOI: https://doi.org/10.1007/s10734-011-9417-z.

Török, Á. and Konka, B. (2020). "A felsőoktatási rangsorkészítés tíz szakmai problémája - Módszertankritikai megjegyzések". In: *Statisztikai Szemle* 98, pp. 909–929.

Török, Á. and Nagy, A. M. (2020). "China: A candidate for winner in the international game of Higher Education?" In: *Acta Oeconomica* 70, pp. 127–152.

Török, Á. and Nagy, A. M. (2021). "A "nagy ugrás" a felsőoktatásban - Kína a nemzetközi verseny élmezőnyében?" In: *Közgazdasági Szemle* 68, pp. 375–398.

U21 (2020). *U21 Rankings*. URL: https://universitas21.com/what-we-do/u21-rankings/u21-ranking-national-higher-education-systems-2020/comparison-table.

Valadkhani, A. and Worthington, A. (2006). "Ranking and Clustering Australian University Research Performance, 1998–2002". In: *Journal of Higher Education Policy and Management* 28.2, pp. 189–210. DOI: 10.1080/13600800600751101. eprint: https://doi.org/10.1080/13600800600751101. URL: https://doi.org/10.1080/13600800600751101.

van der Wende, Marijk (2008). "Rankings and Classifications in Higher Education: A European Perspective". In: *Higher Education*. Ed. by John C. Smart. Dordrecht: Springer Netherlands, pp. 49–71. ISBN: 978-1-4020-6959-8. DOI: 10.1007/978-1-4020-6959-8_2. URL: https://doi.org/10.1007/978-1-4020-6959-8_2.

Van Dyke, N. (2005). "Twenty Years of University Report Cards". In: *Higher Education in Europe* 30.2, pp. 103–125. DOI: 10.1080/03797720500260173. eprint:

https://doi.org/10.1080/03797720500260173. URL: https://doi.org/10.1080/03797720500260173.

Vogel, G. (2016). "Good grades for Germany's project to build an Ivy League". In: *Science* 351.6273, pp. 545–545. ISSN: 0036-8075. DOI: 10.1126/science.351.6273.545. eprint: https://science.sciencemag.org/content/351/6273/545.full.pdf. URL: https://science.sciencemag.org/content/351/6273/545.

Williams, R. and Leahy, A. (2020). *U21 Ranking of National Higher Education Systems 2020*. Research rep. University of Melbourne. eprint: https://universitas21.com/sites/default/files/2020-04/U21_Rankings%20Report_0320_Final_LR%20Single.pdf. URL: https://universitas21.com/sites/default/files/2020-04/U21_Rankings%20Report_0320_Final_LR%20Single.pdf.

Williams, R., Rassenfosse, G. de, Jensen, P., and Marginson, S. (2012). *U21 ranking of national higher education systems 2012*. Research rep. University of Melbourne. eprint: http://www.universitas21.com/article/projects/details/190/2012-rankings. URL: http://www.universitas21.com/RelatedFile/Download/422.

Williams, R., Rassenfosse, G. de, Jensen, P., and Marginson, S. (2013). *U21 ranking of national higher education systems 2013*. Research rep. University of Melbourne. eprint: http://www.universitas21.com/article/projects/details/243/2013-rankings. URL: http://www.universitas21.com/RelatedFile/Download/561.

Williams, R., Rassenfosse, G. de, Jensen, P., and Marginson, S. (2014). *U21 ranking of national higher education systems 2014*. Research rep. University of Melbourne. eprint: http://www.universitas21.com/article/projects/details/288/2014-rankings. URL: http://www.universitas21.com/RelatedFile/Download/659.

Williams, R., Rassenfosse, G. de, Jensen, P., and Marginson, S. (2015). *U21 ranking of national higher education systems 2015*. Research rep. University of Melbourne. eprint: http://www.universitas21.com/article/projects/details/

303/2015-rankings. URL: http://www.universitas21.com/RelatedFile/Download/772.

Williams, R., Rassenfosse, G. de, Jensen, P., and Marginson, S. (2016). *U21 ranking of national higher education systems 2016*. Research rep. University of Melbourne. eprint: http://www.universitas21.com/article/projects/details/312/2016-rankings. URL: http://www.universitas21.com/RelatedFile/Download/857.

Williams, R., Rassenfosse, G. de, Jensen, P., and Marginson, S. (2017). *U21 ranking of national higher education systems 2017*. Research rep. University of Melbourne. eprint: http://www.universitas21.com/article/projects/details/312/2016-rankings. URL: http://www.universitas21.com/RelatedFile/Download/857.

Yang, J., Wang, H., Wang, W., and Yu, P. S. (2003). "Enhanced biclustering on expression data". In: *Third IEEE Symposium on Bioinformatics and Bioengineering, 2003. Proceedings.* pp. 321–327. DOI: 10.1109/BIBE.2003.1188969.

Yang, J., Wang, H., Wang, W., and Yu, P. S. (2005). "An improved biclustering method for analyzing gene expression profiles". In: *International Journal on Artificial Intelligence Tools* 14.05, pp. 771–789. DOI: 10.1142/S0218213005002387. eprint: http://www.worldscientific.com/doi/pdf/10.1142/S0218213005002387. URL: http://www.worldscientific.com/doi/abs/10.1142/S0218213005002387.