



Rankings or leagues or rankings on leagues? - Ranking in fair reference groups

Zsolt T. Kosztyán^{1,2} · Zsuzsanna Banász^{1,2} · Vivien V. Csányi^{1,2} · András Telcs^{1,2,3,4}

Published online: 17 October 2019
© The Author(s) 2019

Abstract

There are several well-known rankings of universities and higher education systems. Numerous recent studies question whether it is possible to compare universities and countries of different constitutions. These criticisms stand on solid ground. It is impossible to create a one-dimensional ordering that faithfully compares complex systems such as universities or even higher education systems. We would like to convince the reader that using well-chosen elements of a family of state-of-the-art data mining methods, namely, bi-clustering methods, can provide an informative picture of the relative positions of universities/higher education systems. Bi-clustering methods produce leagues of comparable entities alongside the indicators, which produce a similar grouping of them. Within leagues, partial rankings could be specified and furthermore can serve as a proper basis for benchmarking.

Keywords University ranking · Leagues of countries · Bi-clustering · Partial rankings · Country rankings

Introduction

Global ranking systems have several shortcomings and face many challenges (Hendel and Stolz 2008; Olcay and Bulu 2017; Soh 2017). The most crucial problem likely stems from the

Andras Telcs is affiliated to VIAS, Virtual Institute of Advanced Studies

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11233-019-09028-x>) contains supplementary material, which is available to authorized users.

✉ Zsolt T. Kosztyán
kzst@gtk.uni-pannon.hu

¹ Department of Quantitative Methods, University of Pannonia, Veszprém, Hungary

² MTA-PE Budapest Ranking Research Group, Veszprém, Hungary

³ Department of Computational Sciences, MTA Wigner Research Center for Physics, Budapest, Hungary

⁴ Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Budapest, Hungary

comparison of universities or countries with very different financial backgrounds, scope and social environments. There is an ongoing effort (Downing 2013; Salmi 2013) to define different and well-tailored leagues for benchmarking universities or countries instead of ranking them in a single group. However, there is no generally accepted method for identifying such leagues.

The authors agree with Benneworth (2010) and Liu (2013) that universities that belong to similar higher education systems should be compared according to a given set of criteria that is also in accordance with the common features of the higher education systems. The present work hinges on a fundamental principle: *leagues (of countries) constitute both the set of countries and the set of indicators*. Specifically, the set of criteria might differ from league to league; however, some of the criteria may be common. This may also apply to universities. The primary challenge in specifying leagues is to simultaneously define a set of criteria and countries/universities that are similar according to the criteria.

The purpose of this study is to answer the following research question: How can leagues of comparable higher education systems be defined? The study does not aim at expanding the criticisms of the indicators of rankings, nor replacing them with other (better) indicators, nor developing new indicators.

In the following, for the case of higher education systems, we show how leagues, as a new basis for comparing higher education systems, can be developed. We use the available indicators as they are, acknowledging that some of them may impose some type of bias, while they are the result of enormous efforts of data acquisition and cleansing, which we surely cannot reproduce. In addition, the usage of well-known indicators follows the *ceteris paribus* principle; we introduce a new method forming groups of objects to compare, but we do not introduce new indicators at the same time. As a result, the gain of forming leagues can be demonstrated without the effect of new indicators.

In this paper, we focus on creating leagues of countries, while in a forthcoming paper we plan to use the method to create university leagues, study the rankings within them, and compare the results with global rankings. We consider this article as a starting point for further analyzes since bi-clustering universities is more complex and since more clusters can be established, i.e., more leagues can occur on larger data tables.

The relevance of examining countries (i.e., their whole higher education system) is reinforced by higher education being a public good (Marginson 2011). Developing a country's higher education system can affect the country's economic well-being. Education increases labor productivity (Mankiw et al. 1992) and promotes the innovation capability of a country (Romer 1990). On the other hand, Hanushek and Wößmann (2010) state that neglecting differences in quality of education can distort the relationship between education and economic growth. University rankings can be a proxy for the quality of education; this also supports our country-level analysis as a contribution to the existing literature in this field.

Related works

University rankings and their critiques

An increasing number of countries are publishing national rankings of their educational institutions at the secondary and tertiary levels. These rankings are prepared by research institutions based on order of the government or commercial actors (e.g., newspapers and non-governmental organizations) (Dill and Soo 2005). On the national level, universities are

Table 1 Groups and examples of ranking systems

		Ranking level	
		University	Higher Education System
Territorial coverage	Global	ARWU, THE, QS, CWTS, U-Multirank	U21 (2012–2017), QS (2008, 2016), Ederer et al. (2009)
	Regional	QS, THE: Latin America/Asia, QS: EECA, QS: BRICS	The scope of this paper.
	National	UK: The Complete University Guide in UK, US: Forbes and CCAP's Rankings	Cannot be interpreted.

Authors' compilation

more comparable because the national field presents common characteristics for every university. One can find numerous national league tables such as The Complete University Guide in the UK. However, rankings on the national level alone do not allow us to compare universities in a global space.

In addition to these national rankings, there are global rankings of higher education institutions (see Table 1). The best-known 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. These rankings are published by independent organizations, such as the Shanghai Ranking Consultancy, which publishes the Shanghai Ranking, and the Center for Higher Education Policy Studies, which is one of the leaders of the consortium that produces U-Multirank (Moed 2017). There is also an international organization called International Ranking Expert Group (IREG) addressing - inter alia - and approving global university rankings.

Most of these systems focus on universities. However, several initiatives (Hazelkorn 2015; Salmi 2013) 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 higher education system 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 Lisbon Council, QS and Universitas21 (U21). The first effort 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 higher education systems to help citizens and society meet the genuine challenges of a twenty-first-century knowledge economy (Ederer et al. 2009). Additionally, in 2008, 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 metrics: system strength, access, flagship institution and economic context (Hazelkorn 2015). In 2016, a similar country-based ranking on their higher education system was published as “QS Higher Education System Strength Rankings” (QS 2016).

The ranking of U21 is the most ambitious of the initiatives mentioned. 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 in this report 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). 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, and U21’s methodology is one of the most transparent. For the U21 ranking, 2014 was the last year in which every indicator was available. Therefore, this paper focuses on the 2014 country ranking.

The higher education rankings suffer from numerous shortcomings. First, they have a distorting effect on universities because they heavily rely on the reputations of institutions (Guarino et al. 2005; Marginson and van der Wende 2007). Second, the applied data, indicators and weights are chosen arbitrarily by each university ranking (Saisana et al. 2011; Bengoetxea and Buela-Casal 2013; Boyadjieva 2017; Pinar et al. 2019). Third, their “fairness” is questionable because institutions have differences in size and funding (Daraio and Bonaccorsi 2017). For example, in 2015, Harvard University’s annual budget was approximately \$4.5 billion, whereas Hungary’s annual budget for all levels of education was approximately \$5 billion in 2015 (European Commission 2015). 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). Rankings at the national and global levels are described in this subsection. The regional rankings (see Table 1) are detailed in Subsection 2.2.

Regional rankings - new directions and limitations

In addition to national and global rankings, regional rankings can be specified (see Table 1). To date, regional rankings refer to rankings within geographic regions, e.g., continents. Excellent examples include the Latin America and Asia University Rankings of QS (<https://www.topuniversities.com/university-rankings>) (Sowter et al. 2017) and THE (<https://www.timeshighereducation.com/world-university-rankings>), and the Arab Region University Rankings of QS. Similarly, U.S. News classifies their regional US rankings into four regions: North, South, Midwest and West (<https://www.usnews.com/best-colleges/rankings/regional-universities>). Exceptions exist that 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.

Although these regional rankings most likely better fulfill the fairness criterion of the International Ranking Expert Group (IREG), because the elements (universities/countries) are comparable, both geographical and economic rankings can be considered arbitrary classifications of universities or countries.

In addition to geographic and economic-based regional rankings, scholars (Abankina et al. 2016; Jarocka 2012) 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 the universities after clustering them.

In addition to arbitrary classification, clustering methods are used to separate clusters (Rad et al. 2011; Ibáñez et al. 2013). Ibáñez et al. (2013) clustered public universities in the area of computer science into four groups based on their productivity, visibility,

quality, prestige and internationalization. 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. 2018).

Bi-clustering methods are relatively new and are almost entirely unknown and unused in the social sciences. We demonstrate the capabilities of these methods in the clustering and ranking of higher education systems. Using well-chosen elements of the family of bi-clustering methods, we can find meaningful but far-from-evident leagues of both countries and indicators. The selected indicators shed light on higher education systems' strengths and weaknesses and their proper positions in the rankings. Finally, the proposal opens a new direction of multivariate analysis that is free of subjective or ad-hoc weights and does not require indicator selection over non-comparable indicators.

A fair comparison of higher education systems (universities) can be performed within leagues. In the present paper, we create three leagues, which are denoted 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.

Materials and methods

Introduction to the bi-clustering method

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 Hartigan (1972). This clustering method was not generalized until 2000 when Cheng and Church (2000) proposed a bi-clustering algorithm based on variance and applied it to biological gene expression data. Many bi-clustering algorithms have been developed for bioinformatics (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 (Huang 2011; Liu et al. 2009), there is already a publication (Raponi et al. 2016) on the bi-clustering of university performances. This study clearly demonstrates how to select indicators and universities simultaneously.

There are different definitions of bi-clusters and multiple bi-clustering algorithms (Madeira and Oliveira 2004):

BIC1: Bi-clusters with constant values (in rows and/or columns) (see Table 2(a));
BIC2: Bi-clusters with similar values (on rows and/or columns) (see Table 2(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 simultaneously and then to find bi-clusters with similar (constant) values. In contrast, BIC2-type algorithms seek

Table 2 Cell selection results

		Indicators									
		1	2	3	4	5	6	7	8	9	10
Countries	1	X		X			X		X		
	2										
	3	X		X			X		X		
	4	X		X			X		X		
	5										
	6										
	7										
	8										
	9										
	10										
	11	X		X			X		X		
	12	X		X			X		X		
	13	X		X			X		X		

		Indicators									
		1	2	3	4	5	6	7	8	9	10
Countries	1	O	O	O		O		O		O	
	2	O	O	O		O		O		O	
	3	O	O	O		O		O		O	
	4	O	O	O		O		O		O	
	5										
	6										
	7										
	8										
	9										
	10										
	11	O	O	O		O		O		O	
	12	O	O	O		O		O		O	
	13										

X, O: selected cells, ■: upper/ □: lower than a specified threshold
 Authors' compilation

bi-clusters with similar values in the rows and columns. Similarity can be measured in many ways, the simplest of which is by analyzing the variance between groups using the co-variance between rows and columns. In Cheng and Church's (2000) 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 (NP: Nondeterministic Polynomial time), and they employed a Greedy Search Heuristic in their approach. An NP-hard problem means that there is no known algorithm that 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 our analysis, the rows correspond to the 50 countries of U21, and the indicators of U21 are the columns. Gray cells are those that are above a given threshold, here, the median of the total matrix. The "X"s in Table 2(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.

The lower the threshold is, the larger and less similar the cluster (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. Table 2(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 so.

Bi-clustering of HESs

In the present paper, we demonstrate on a relatively small number of objects, namely the U21 countries' higher education systems, how well-selected bi-clustering methods can identify leagues (countries and indicators simultaneously). For simplicity, we identify only three leagues: upper league A, middle league B and lower league C. For this purpose, we use two bi-clustering methods.

The first method is the **iterative Bi-clustering of Genes (iBBiG)** (Gestraud 2008; 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 2(a)). The procedure starts with the normalization of the indicators, as defined in (1).

$$x'_{ij} = \frac{x_i - m(j)}{M(j) - m(j)}, \quad (1)$$

where $m(j) = \min_i x_{ij}$, $M(j) = \max_i x_{ij}$.

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. The next step in iBBiG is based on a threshold, which, in our case, is the median of the whole matrix. A new binary matrix is created in which a value of one is used for those cells whose original value is above that threshold. All other cells have 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 the upper league A. The binary reversed data and the same procedure yield the lower League C. The iBBiG method can produce more than one bi-cluster (i.e., leagues), which can overlap if the procedures are applied with different thresholds. When using different thresholds to develop several alternative clusters, a quality test is needed to evaluate the results.

For simplicity, we do not apply multiple thresholds; instead, to identify the middle league, another bi-clustering method, namely, **Bi-clustering Analysis and Results Exploration (BicARE)**, is used (Gusenleitner and Culhane 2011). With the BicARE method, we create a bi-cluster that identifies a middle league of countries overlapping with (A) and (C), thus providing a more detailed picture of the performances. The positions of the countries with respect to the created leagues are depicted in Figure 1.

BicARE is a BIC2-type method, where the similarity measure is the correlation (see Table 2 (b)). BicARE (Gestraud et al. 2014) is the improved and enhanced version of the **FLexible Overlapped biClustering (FLOC)** algorithm proposed by Yang et al. (2003). This method is based on the concept of residues, which are a measure of the similarity of the elements in a bi-cluster (Yang et al. 2005). The smaller the residue is, the more similar the elements of the bi-cluster. Similarly to the interpretation of the upper and lower leagues, when interpreting the middle league (see the cells of Table 2(b) that are marked by 'O'), the BicARE method specifies a group (submatrix) of countries and indicators whose values are similar (their variances are minimized) for both countries and indicators.

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 following sections, the research design and the results are introduced. All of the analysis was made in R. The detailed description of the used packages, the related datasets, the steps of the analysis, and the results can be found in this link: https://kmt.gtk.uni-pannon.hu/kutatas/u21/EN/U_21_EN.html.

Research design

Data

The U21 rankings of countries by their higher education systems (Williams et al. 2012, 2013, 2014, 2015, 2016, 2017) are developed at the University of Melbourne. In what follows, we present the evaluation of the U21 rankings and their indicators in detail. The U21 rankings cover 7 years (2012–2017) 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 summarizes the countries in order of ranking for the year 2014.

The overall U21 rank scores are calculated from four groups based on resources (R), environment (E), connectivity (C) and output (O). Each (sub) indicator is a weighted average of multiple variables. Table 4 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 for the U21 ranking are available for each year; however, the (sub) indicators are only available for the years 2012–2014. For the appropriate application of bi-clustering, only the (sub) indicators must be considered. Since (sub) indicators of the U21 rankings are not available from 2015, the year 2014 was selected.

Table 3 Countries of U21

Countries									
(1–10)		(11–20)		(21–30)		(31–40)		(41–50)	
1	United States	11	Norway	21	Korea (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

Williams et al. (2014)

Table 4 List of U21's indicators

<i>w (%)</i>	Abbr.	Variables
5.0	20 R(esources)	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	20 E(nvironment)	E1 % of female students in tertiary education
2.0		E2 % of academic staff who are female in tertiary institutions
2.0		E3 A rating for data quality.
14.0		E4 Qualitative measure of the policy environment
4.0	20 C(onnectivity)	C1 % of international students in tertiary education
4.0		C2 % of articles 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 4200 universities divided by 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 country's population
4.0		C5 Responses to question 'Knowledge transfer is highly developed between companies and universities', 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	40 O(utput)	O1 Total number of journal articles produced by higher education institutions
3.3		O2 Total articles produced by higher education institutions per head of population
3.3		O3 Average impact of articles as measured by citations in 2014 to articles published in previous years using the Karolinska Institute normalized impact factor.
3.3		O4 The depth of world-class universities in a country. This is calculated as a weighted average of the number of institutions listed in the top 500 according to the 2013 Shanghai Jiao Tong scores, divided by country population.
3.3		O5 The excellence of a nation's best universities calculated by totalling the 2013 Shanghai Jiao Tong scores for the nation's three best universities.
3.3		O6 Enrolments in tertiary education as a % of the eligible population, defined as the 5-year age group following on from 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 million of 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

w (%): weights, exp.: expenditure, PPP: purchasing power price

Williams et al. (2014)

The indicators of U21 were collected from several sources (e.g., Education at a Glance 2013 report of OECD, UNESCO Statistics, The Global Competitiveness Report 2013–2014 of World Economic Forum, and the Scopus data bank and survey of U21). For each variable, the original values were scaled to 0–100 intervals. The countries' overall scores were calculated as weighted sums of indicators.

Although U21 published the score values of the indicators, these values cannot be verified completely. First, although most sources of raw data are published, a few are not available (e.g., the qualitative measure of the policy environment (E4)). Second, various series are

Table 5 Descriptive statistics

Variables	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%

SD Standard deviation; *N* = 50, Max = 100.0

Authors' compilation based on data of Williams et al. (2014)

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; therefore, it is impossible to re-calculate the indicator of C4. Third, there are several missing values (42/1200 = 3.42%), and the methodology used to treat missing data is unpublished.

Table 5 summarizes various descriptive statistics of the 24 indicators, which are scaled to scores of 0–100. The greatest number (11) of cells are missing for the proportion of female academic staff (E2). The least varied data (the indicator in which the countries are the most similar) are the proportion of female students (E1). Those 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%).

Steps of the analysis

The analysis consisted of 5 steps:

- Step 1: Replacing missing vales;
- 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.

The applied iBBiG algorithm is robust to missing values (Gestraud et al. 2014); however, the BicARE algorithm requires a complete database. Choosing the appropriate method for replacing 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 are MAR-type data 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 data with missing values do not exceed 5% (Scheffer 2002). Since the ratio of missing values was low, in the first step (Step 1), the missing values were replaced. 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. We did not use the software solutions offered to replace missing data (e.g., mean or median); however, we replaced the missing data in such a way that the original U21 ranking 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 easily calculated. 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. 1).

Since the original iBBiG method finds only the upper league(s) A, in the next step (Step 3), the reversed normalized data (1-normalized 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 rankings. When calculating partial rankings for countries in the specified bi-cluster(s), the original weights of U21's indicators were used, and the total scores for the countries were calculated using the selected indicators in the given bi-cluster.

Calculation of partial rankings

Within leagues, partial rankings and partial scores are specified. The 4-step calculation method uses the original U21 scoring and ranking method, which can also be used if indicators are excluded from a league.

- Step 1: The four main components (R, E, C and O) are calculated as weighted averages of the scores of indicators (weights are listed in Table 4).
- 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 to 100, and the remaining overall scores are re-scaled proportionally.
- Step 4: In the final step, the countries are ordered by their overall scores.

Results

Descriptive analysis

The countries in U21 have varied geographic locations, different income levels and histories. Most of the countries (27) are in 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. The countries are grouped into high, higher middle, lower middle and low income categories by the World Bank (<http://databank.worldbank.org/data/download/site-content/OGHIST.xls>). 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 histories. 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 5 that have relative SDs of less than 20%. These cover all environmental indicators. Two indicators are extremely low, i.e., under 10%. The proportion of female students in tertiary education (E1) has the smallest relative SD, only 4.1%. A total of 39 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 score 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 achieve 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 for the total number of journal articles produced by higher education institutions (O1). The United States achieved the maximum score of 100, which was extremely high compared to other countries. The second highest score of 58.9 was 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 the interval [2,5], 17 countries' scores are less than 2, and the scores of 8/17 countries are lower than 1. The average scores of the top 500 Shanghai institutes (O4) have the second largest relative SD (109.9%). For this indicator,

Table 6 Results of scores and significances for iBBiG bi-clustering algorithms

Dataset	№	Score	Rows	Cols	F-tests for	
					Row Effects (<i>p</i> -values)	Col Effects (<i>p</i> -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

Authors' compilation based on data of Williams et al. (2014)

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).

Results of bi-clustering

In contrast to classical clustering (depending on the applied method), bi-clusters may overlap. In what follows, we pinpoint cases in which membership in a single cluster or membership in multiple clusters has a particular meaning. In both situations, countries and indicator positions must be considered simultaneously.

After the 100 runs, two bi-clusters with higher frequencies appeared; the others have negligible hits. Table 6 shows that only cluster number 1 has acceptable significance for both dimensions.

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.

After specifying League(s) A and C, League(s) B must be specified similarly. To find the middle league, we use a different concept of similarity. We seek a middle league in which the variances between countries and between indicators are low simultaneously. The BicARE method provides such bi-clusters. After performing the F-test to compare the variances for both countries and indicators between included and excluded cells, one significant bi-cluster can be specified.

Leagues

Since a country can have several high and low values simultaneously, it can be a member of more than one league simultaneously. Similarly, if an indicator has a high relative variance (see Table 5), its high-value cells can be included in League A, and low-value cells can be included in League C (see the overlaps of columns of cells that are labelled X or O in Table 2). Therefore, the results of bi-clusters can specify overlaps (see Fig. 1). 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 Fig. 1), which is worth analyzing case by case.

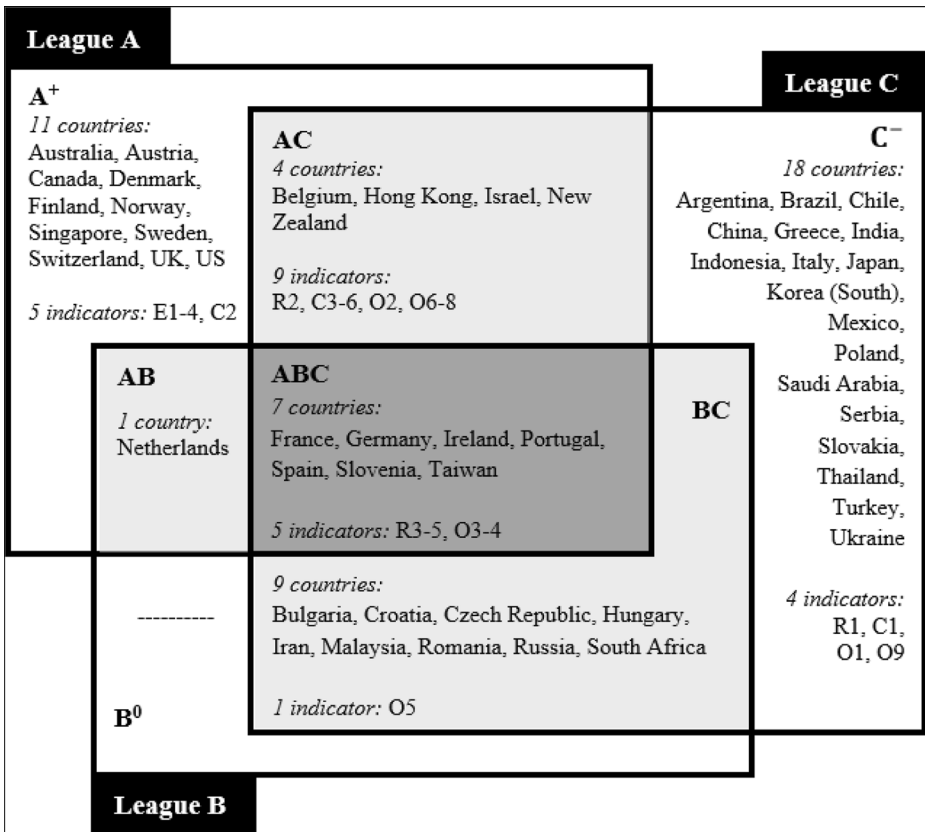


Fig. 1 Leagues specified by bi-clustering algorithms – results. Source: Authors’ compilation

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 three best universities by Shanghai (O5), unemployment rates (O9), government expenditure (R1) and international students (C1). These are the indicators for which 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 5 shows that standard deviation is 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⁺ The part of League A that does not overlap with other leagues (which is denoted as A⁺) contains 11 countries and 5 indicators. The countries are in the top 12 countries of the original U21 ranking. However, our method can highlight the indicators that clearly separate the countries of League A. These indicators are all indicators about the environment (E1–4) or connectivity (C2).

While these countries have higher GDP per capita (<http://databank.worldbank.org/>), we suspect that differences in resources do not directly cause the separation. Similarly, indicators about the environment (E1–4) are indirectly linked to higher education systems. We conclude that the only indicator that has a direct impact on the separation of that top group is that of the articles that were co-authored with international collaborators (C2).

League C League C includes most of the countries (38) and those 19 indicators that were not in League A⁺. This means that there are more less-well-performing countries (38 in League C) than well-performing countries (23 in League A). Nevertheless, the numbers 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 indicator (C2). The countries in League C usually perform worse in government expenditure (R1), international students (C1), journal articles (O1), the nation's three best universities by the Shanghai 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 18 countries have relatively high scores in these indicators. These 18 countries are in the middle (20th, 21st and 27th 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 resource, connectivity and output categories. There are four resource (R2–5) indicators in the intersection of League A and League C. These indicators are very important for both higher education and specifying both League A and League C and are capable of comparing 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 high unemployment rates among tertiary-educated people (O9), which pull countries toward League C⁻.

In League ABC, there are 7 countries and 5 indicators, which appear in all three leagues. Most of the resource indicators (3/5 indicators) 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 resource (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 7th place in the original U21 ranking). This result shows that League A is better separated from the midfield league 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

Table 7 Partial rankings on leagues

	rank by U21		rank within League A		rank within League B		rank within League C	
	by U21	diff.	by U21	by authors	by U21	by authors	by U21	by authors
US	1		1	3				
Sweden	2	-2	2	1				
Denmark	3	1	3	2				
Canada	3	-3	3	6				
Finland	5	1	5	4				
Switzerland	6	-1	6	7				
Netherlands	7	2	7	5	1	1		
UK	8	-3	8	11				
Australia	9	0	9	9				
Singapore	10	0	10	10				
Norway	11	3	11	8				
Austria	12	-2	12	14				
Belgium	13	0	13	13			1	-1
Germany	14	-4	14	18	2	2	2	1
Hong Kong	15	3	15	12			3	8
New Zealand	16	0	16	16			4	5
Ireland	17	0	17	17	3	4	5	3
France	18	-1	18	19	4	3	6	6
Israel	19	4	19	15			7	4
Japan	20						8	7
South Korea	21						9	9
Taiwan	22	2	20	20	5	7	10	10
Spain	23	0	21	23	6	6	11	11
Portugal	24	3	22	21	7	5	12	13
Slovenia	25	3	23	22	8	8	13	12
Czech Republic	26				9	10	14	14
Italy	27						15	15
Malaysia	28				10	12	16	20
Hungary	29				11	9	17	17
Saudi Arabia	30						18	18
Poland	31						19	22
Greece	32						20	16

Table 7 (continued)

	rank by U21	rank within League A	rank within League B	rank within League C	
Chile	33			21	26
Serbia	34			22	24
China	35			23	19
Russia	35	12	14	23	21
Slovakia	37			25	26
Brazil	38			26	23
Romania	39	13	16	27	31
Bulgaria	40	14	17	28	34
Argentina	41			29	30
Thailand	42			30	33
Ukraine	42			30	28
Croatia	44	15	13	32	25
South Africa	45	16	11	33	29
Mexico	46			34	36
Turkey	47			35	32
Indonesia	48			36	38
Iran	49	17	15	37	35
India	50			38	37

Notes: **ties** if diff. < 0: fall back if diff. > 0: moving forward
 Authors' compilation based on data of Williams et al. (2014)

lower variances. The absence of connectivity indicators could be caused by their large variance.

League B⁰ and League AB There is no common country or indicator of League B⁰. Additional evidence of the better separation of the top league (League A⁺) is that there is only one country (The Netherlands) in League AB, whereas there are 9 countries in League BC. This reflects the large gap 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 three best 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).

Results of partial rankings

Table 7 summarizes the results of the original U21 rankings and the results of partial rankings within leagues. All 50 countries are listed in Table 7 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 7 and 49.

Table 8 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 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.

Table 8 Results of partial correlations

	League A			League B			League C		
	Correlation Coefficient	<i>p</i> -value	N	Correlation Coefficient	<i>p</i> -value	N	Correlation Coefficient	<i>p</i> -value	N
Kendall's τ_B	0.824	0.000	23	0.765	0.000	17	0.855	0.000	38
Spearman's ρ	0.956	0.000	23	0.909	0.000	17	0.966	0.000	38

Authors' compilation based on data of Williams et al. (2014)

Discussion and summary

The results obtained for the upper/lower leagues are consistent with the U21 ranking. The research question was how the leagues of comparable higher education systems of countries can be defined. We have shown how the method of bi-clustering can help define leagues of comparable higher education systems of countries. In the following, we interpret the bi-clustering results.

This study showed that properly chosen bi-clustering methods are suitable for specifying leagues of countries' higher education systems. We specified the middle league (League B) by the BicARE method, while the upper and lower leagues (Leagues A and C) were specified by the iBBiG method. We have shown that using proper bi-clustering methods we can obtain a new classification of countries. The classification deviates somewhat from the usually defined economic groups and substantially from the geographic regions. Instead, it provides new groups that are in good accordance with the U21 ranking and reveals the roles of the indicators in positioning the countries.

One of the most interesting findings is that except for government expenditures (R1), all resource (R2–5) indicators are significant in League A. This means that spending a high percentage of GDP on higher education by the government (R1) does not pull a country into the upper league, but other resources do. Spending more on higher education and R&D (as a percentage of GDP or per capita amount) derived from any sources can pull up a country into a better position (R2–5).

In League A+, what matters the most (if other variables also have a high value) is the percentage of articles that are written with international co-authorship (C2) because the other indicators of League A+ (E1–4) are relatively high in all countries (see Table 5).

An analysis of overlaps shows that there are 3 input (R3–R5) and 2 output (O3, O4) indicators that are common among all leagues (League ABC). Therefore, all countries can be compared using this subset of indicators. Since both input and output indicators are common in all leagues, countries can also be benchmarked by the effectiveness of their higher education systems. Higher values of these five indicators pull countries toward the upper league. Lower values of these five indicators pull countries toward the lower league.

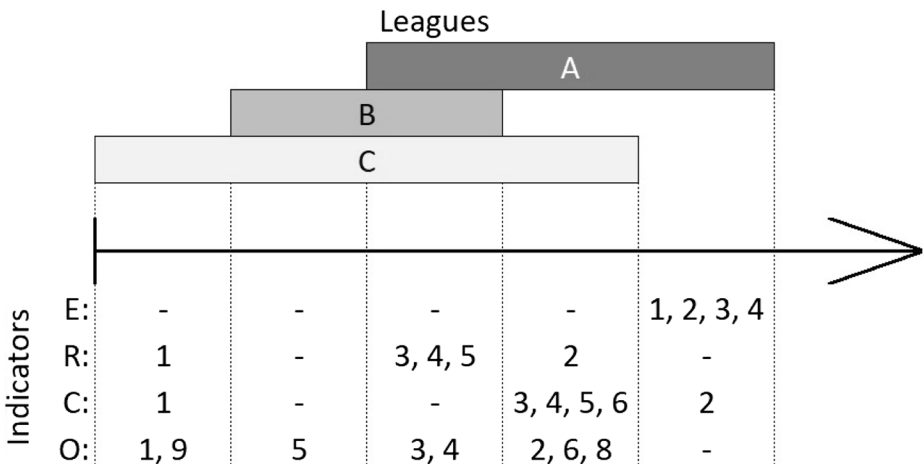


Fig. 2 Development opportunities among leagues. Source: Authors' compilation

In our paper, we demonstrated that leagues of countries and indicators can be defined together by well-selected bi-clustering methods. As a result, we can obtain a “fair” ranking of higher education systems that is not biased by the intentional pre-selection of indicators. The resulting leagues may be helpful in clarifying the roles of the obtained indicators.

Beyond specifying partial ranks, bi-clustering methods provide a much richer picture. The first opportunity is to analyze overlaps. The results of specifying leagues showed that it is very difficult to specify a strict partition of countries and indicators. Nevertheless, overlaps indicate that several countries can be members of more than one league. For example, countries in the overlapped regions can outperform other countries when considering the indicators of the elite league. However, improvements in several indicators must be achieved to separate these countries from lower leagues (see Fig. 2). Overlap analysis also identifies the common indicators. Based on these common indicators, countries from different leagues can also be compared.

The results of bi-clustering are analyzed in detail. By comparing countries that belong to the same league or observing which countries are separated (League A⁺ and League C⁻), the strengths and weaknesses of a higher education system can be identified. Consequently, a point of necessary intervention might be revealed as well (see Fig. 2).

Conclusion

When comparing countries (universities), the first and one of the most fundamental questions is which subjects can be compared and which indicators can be used in the comparison. The authors believe that the bi-clustering method can play an important role in both ranking and benchmarking. Despite the interpretation of bi-clustering being more challenging than explaining the results of conventional clustering, analyzing overlaps and the separations provide an opportunity to show why the top countries are separated from others, while other countries became members of more than one league. The proposed bi-clustering methods identify the common indicators that can be used for global rankings or as benchmarks. Nevertheless, even if there is no common indicator, after specifying bi-clusters, regional or more partial rankings can be defined within leagues. The countries in these leagues are not evaluated based on arbitrarily determined indicators of an arbitrarily selected region (see regional rankings in Table 1); however, similar countries can be compared by comparable indicators. Consequently, a point of necessary intervention may also be revealed in order to get into a better league (see Fig. 2).

Acknowledgements This publication/research has been supported by the European Union and Hungary and co-financed by the European Social Fund through the project EFOP-3.6.2-16-2017-00017, titled “Sustainable, intelligent and inclusive regional and city models.”

Funding Information Open access funding provided by University of Pannonia (PE).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abankina, I., Aleskerov, F., Belousova, V., Gokhberg, L., Kiselgof, S., Petrushchenko, V., Shvydun, S., & Zinkovsky, K. (2016). From equality to diversity: Classifying russian universities in a performance oriented system. *Technological Forecasting and Social Change*, *103*, 228–239.
- Bengoetxea, E., & Buela-Casal, G. (2013). The new multidimensional and user-driven higher education ranking concept of the European Union. *International Journal of Clinical and Health Psychology*, *13*(1), 67–73.
- Benneworth, P. S. (2010). *A University benchmarking handbook*. Benchmarking in European Higher Education. ESMU.
- Boyardjieva, P. (2017). Invisible higher education: Higher education institutions from central and Eastern Europe in global rankings. *European Educational Research Journal*, *16*(5), 529–546.
- Cheng, Y., & Church, G. M. (2000). Biclustering of expression data. In *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology* (pp. 93–103). AAAI Press.
- Daraio, C., & Bonaccorsi, A. (2017). Beyond university rankings? Generating new indicators on universities by linking data in open platforms. *Journal of the Association for Information Science and Technology*, *68*(2), 508–529.
- Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education*, *49*(4), 495–533.
- Downing, K. (2013). *What's the use of rankings?*, chapter 11, pages 197–208. Rankings and accountability in higher education: Uses and misuses. UNESCO.
- Ederer, P., Schuller, P., Willms, S., et al. (2009). University systems ranking: Citizens and society in the age of the knowledge. *Educational Studies*, *3*, 169–202.
- Comission, E. (2015). *National Sheets on education budget in Europe: Eurydice facts and figures*. Luxembourg: Publications Office of the European Union.
- Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability; A Guide to the Theory of NP-Completeness*. New York: W. H. Freeman & Co..
- Gestraud, P. (2008). BicARE: Biclustering analysis and results exploration. R package version 1.38.0.
- Gestraud, P., Brito, I., and Barillot, E. (2014). Bicare: Biclustering analysis and results exploration.
- Guarino, C., Ridgeway, G., Chun, M., & Buddin, R. (2005). Latent variable analysis: A new approach to university ranking. *Higher Education in Europe*, *30*(2), 147–165.
- Gusenleitner, D. and Culhane, A. (2011). iBBiG: Iterative binary Biclustering of Genesets. R package version 1.24.0.
- Gusenleitner, D., Howe, E. A., Bentink, S., Quackenbush, J., & Culhane, A. C. (2012). Ibbig: Iterative binary bi-clustering of gene sets. *Bioinformatics*, *28*(19), 2484–2492.
- Hahsler, M., Hornik, K., & Buchta, C. (2008). Getting things in order: An introduction to the R package seriation. *Journal of Statistical Software*, *25*(3), 1–34.
- Hanushek, E. A. and Woßmann, L. (2010). Education and economic growth. In Peterson, P., Baker, E., and McGaw, B., editors, *International encyclopedia of education*, volume 2, pages 245–252. Elsevier, Oxford.
- Hartigan, J. A. (1972). Direct clustering of a data matrix. *Journal of the American Statistical Association*, *67*(337), 123–129.
- Hazelkorn, E. (2015). *Rankings and the reshaping of higher education: The battle for world-class excellence*. Springer.
- Hendel, D. D., & Stolz, I. (2008). A comparative analysis of higher education ranking systems in europe. *Tertiary Education and Management*, *14*(3), 173–189.
- Huang, Q.-H. (2011). Discovery of time-inconsecutive co-movement patterns of foreign currencies using an evolutionary biclustering method. *Applied Mathematics and Computation*, *218*(8), 4353–4364.
- Ibañez, A., Larrañaga, P., & Bielza, C. (2013). Cluster methods for assessing research performance: Exploring Spanish computer science. *Scientometrics*, *97*(3), 571–600.
- IREG (2006). Berlin principles on ranking of higher education institutions. International Ranking Expert Group.
- Jaročka, M. (2012). University ranking systems—from league table to homogeneous groups of universities. *International Journal of Educational and Pedagogical Sciences. World Academy of Science, Engineering and Technology*, *6*(6), 1377–1382.
- Liiv, I. (2010). Seriation and matrix reordering methods: An historical overview. *Statistical Analysis and Data Mining*, *3*(2), 70–91.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley series in probability and statistics (2nd ed.). Hoboken: John Wiley & Sons, Inc..
- Liu, N. C. (2013). *The Academic Ranking of World Universities and its future direction*, chapter 1, pages 23–39. Rankings and accountability in higher education: Uses and misuses. UNESCO.
- Liu, S., Chen, Y., Yang, M., & Ding, R. (2009). Bicluster algorithm and used in market analysis. In *2009 Second International Workshop on Knowledge Discovery and Data Mining* (pp. 504–507).

- Madeira, S. C., & Oliveira, A. L. (2004). Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1(1), 24–45.
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(2), 407–437.
- Marginson, S. (2011). Higher education and public good. *Higher Education Quarterly*, 65, 411–433.
- Marginson, S., & van der Wende, M. (2007). To rank or to be ranked: The impact of global rankings in higher education. *Journal of Studies in International Education*, 11(3–4), 306–329.
- Mirkin, B. (1998). Mathematical classification and clustering: From how to what and why. In *Classification, data analysis, and data highways* (pp. 172–181). Springer.
- Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics*, 110, 967–990.
- Olcay, G. A., & Bulu, M. (2017). Is measuring the knowledge creation of universities possible?: A review of university rankings. *Technological Forecasting and Social Change*, 123, 153–160.
- Pinar, M., Milla, J., & Stengos, T. (2019). Sensitivity of university rankings: Implications of stochastic dominance efficiency analysis. *Education Economics*, 27(1), 75–92.
- Pontes, B., Girdlez, R., & Aguilar-Ruiz, J. S. (2015). Biclustering on expression data: A review. *Journal of Biomedical Informatics*, 57, 163–180.
- Poole, S. M., Levin, M. A., & Elam, K. (2018). Getting out of the rankings game: A better way to evaluate higher education institutions for best fit. *Journal of Marketing for Higher Education*, 28(1), 12–31.
- QS (2016). Higher Education System Strength Rankings - a ranking of national higher education system.
- Rad, A., Naderi, B., & Soltani, M. (2011). Clustering and ranking university majors using data mining and AHP algorithms: A case study in Iran. *Expert Systems with Applications*, 38(1), 755–763.
- Raponi, V., Martella, F., & Maruotti, A. (2016). A biclustering approach to university performances: An Italian case study. *Journal of Applied Statistics*, 43(1), 31–45.
- Rauhvargers, A. (2013). *Global university rankings and their impact: Report II*. European University Association Brussels.
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5), S71–S102.
- Saisana, M., d’Hombres, B., & Saltelli, A. (2011). Rickety numbers: Volatility of university rankings and policy implications. *Research Policy*, 40(1), 165–177 Special Section on Heterogeneity and University-Industry Relations.
- Salmi, J. (2013). *If ranking is the disease, is benchmarking the cure?*, chapter 13, pages 235–255. Rankings and accountability in higher education: Uses and misuses. UNESCO.
- Scheffer, J. (2002). Dealing with missing data. *Research Letters in the Information and Mathematical Sciences*, 3, 153–160.
- Soh, K. (2017). The seven deadly sins of world university ranking: A summary from several papers. *Journal of Higher Education Policy and Management*, 39(1), 104–115.
- Sowter, B., Hijazi, S., and Reggio, D. (2017). *Ranking World Universities*, chapter A decade of refinement, and the road ahead, pages 1–24. IGI Global.
- Williams, R., de Rassenfosse, G., Jensen, P., and Marginson, S. (2012). U21 ranking of national higher education systems 2012.
- Williams, R., de Rassenfosse, G., Jensen, P., and Marginson, S. (2013). U21 ranking of national higher education systems 2013.
- Williams, R., de Rassenfosse, G., Jensen, P., and Marginson, S. (2014). U21 ranking of national higher education systems 2014.
- Williams, R., Leahy, A., de Rassenfosse, G., and Jensen, P. (2015). U21 ranking of national higher education systems 2015.
- Williams, R., Leahy, A., de Rassenfosse, G., and Jensen, P. (2016). U21 ranking of national higher education systems 2016.
- Williams, R., Leahy, A., and Jensen, P. (2017). U21 ranking of national higher education systems 2017.
- Yang, J., Wang, H., Wang, W., & Yu, P. S. (2003). Enhanced biclustering on expression data. In *Third IEEE Symposium on Bioinformatics and Bioengineering, 2003. Proceedings* (pp. 321–327).
- Yang, J., Wang, H., Wang, W., & Yu, P. S. (2005). An improved biclustering method for analyzing gene expression profiles. *International Journal on Artificial Intelligence Tools*, 14(05), 771–789.