



University leagues alongside rankings

Zsuzsanna Banász^{1,2} · Zsolt T. Kosztyán^{2,3,4,5} · Vivien V. Csányi³ ·
András Telcs^{2,3,6,7}

Accepted: 8 March 2022
© The Author(s) 2022

Abstract

University rankings can both orient and disorient potential students. In rankings, universities with very different characteristics are compared, which makes interpretation difficult. We propose the application of a clustering method, which creates groups of universities that are close to each other with respect to a subset of indicators, but the indicators also show homogeneity with respect to the universities in that group. We call such groups leagues. These leagues are defined by the data themselves and are not based on subjective criteria. We demonstrate our proposition using one member of the family of the two-way clustering method, namely, biclustering. The case we present is based on the Round University Ranking (RUR) 2020 dataset. The use of leagues could provide better guidance not only for potential applicants but also for university funding organizations and policy-makers. Our case study led to a somewhat surprising observation. In the top league (based on the RUR data and indicators), the three most important indicators measure reputation, not scientific or educational performance.

Keywords Universities · Rankings · Leagues · Biclustering

1 Introduction

As Lawrence and Green (1980) state, higher education systems are too complex to rank on the basis of a couple of indicators. So in the past decades, several rankings were developed and launched with the objective of adequately measuring and giving insight into the quality of institutions to students, scholars, and decision-makers.

One can find a wide range of rankings depending on which combination of the following three aspects are relevant: (i) those at the global, regional (in a geographical or economic sense) or national level, (ii) universities or the whole higher education system of countries, and (iii) whether or not they are thematic. This study examines global nonthematic university rankings. Although university rankings are criticized in many ways (see, e.g., Soh (2017); Mussard and James (2018)), they are widely used by potential students before deciding which institution to apply to (see, e.g., Griffith and Rask (2007); Helbig and Ulbricht (2010); Horstschräer (2012); Kosztyán et al (2019a); Bowman and Bastedo

Extended author information available on the last page of the article

(2009)). As a result, university leaders cannot avoid taking these rankings into account when making strategic decisions about their institution, despite their drawbacks. University management needs the indicator composition of the ranking in order to be able to improve their university position.

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 who are thinking about where to apply presumably do not investigate the reasons for these large 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. We think that rankings would not be as popular among students if they knew that these rankings are based primarily on the research (and not educational) performance of faculty or on how quickly they could achieve their dream job. A counterexample is the Financial Times, whose ranking also takes into account the salaries of graduates.¹Bell and Brooks (2019) found in the UK that students are more satisfied with universities where the level of research is lower. Kosztyán et al (2019a) 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. The study aims to present a method defining university leagues on a neutral base.

“Leagues” (not in the sense we use the term) are already used to eliminate the heavily criticized deficiency of global rankings. Those leagues are based on the universities’ major field of activity (medicine, business) or other characteristic features (such as size or financial constitution). It is certain that such a definition of the scope reduces the incommensurability of the selected items but at the same time 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. In our study, we present a method of league creation that is free of ad-hock choices or a suspect of bias. We demonstrate 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. We found that universities earned their privileged position in the top league, having high scores only in their three reputation-based indicators.

In this study, leagues are specified by an unsupervised biclustering method. Leagues are defined simultaneously by a set of indicators and a set of universities. Based on a given threshold, the top-, mid-, and low-quality leagues are specified. 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. League membership has a double message for students. 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.

¹ <https://www.ft.com/content/5f04d8c4-1035-11ea-a225-db2f231cfeae>.

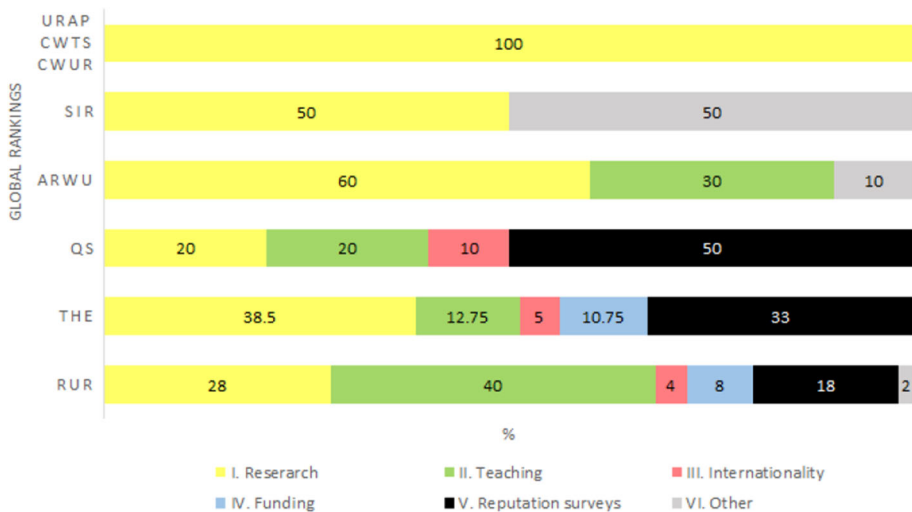


Fig. 1 Indicator categories for 8 global rankings and their weights. (colour figure online)

2 Literature review

2.1 Categorization of ranking indicators

The indicators vary with each ranking. Figure 1 shows the indicators of eight rankings² categorized by the authors.

Rankings are usually a mixture of indicators measuring the following four areas: I. research, II. teaching, III. internationalization, and IV. funding. Indicators related to publications and citations were classified into research categories. Publications made through international collaboration appear as separate variables in several rankings. We could have grouped them into the international category, but instead we categorized them by publications. Teaching is measured by different ratios of the number of academic staff to students at certain tertiary ISCED³ levels. Internationalization consists of the share of international academic staff and students. 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.). Other (rarely used) indicators were accounted for in category VI. In the cases of ARWU

² ARWU: Academic Ranking of World Universities, also known 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>.

³ International Standard Classification of Education.

and RUR, the average of certain indicators from categories I-IV were taken. SIR measures innovation and societal factors in the 'other' category.

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. Our 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 Fig. 1). The RUR also uses survey-based indicators to a lesser extent. In addition, the qualitative and quantitative indicators are much more balanced.

2.2 Leagues or rankings

At the beginning of the twentieth century, developed economies started to transition 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 experienced a sudden boom as well. This phenomenon has placed more emphasis on the measurement of the quality of institutions and has led to the need for classification systems that can differentiate between universities in terms of 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 UK, 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 in addition to the Russell Group: the Cathedrals Group, GuildHE, MillionPlus, Unaffiliated, and University Alliance⁴ Table 1 shows some rankings that use the term 'league'.

Among the rankings listed in Figure 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 ranking⁵. Their names are reminiscent of country ratings by Standard & Poor's or Fitch (Genc and Basar 2019). U-Multirank⁶ 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 multidimensional 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.

⁴ <https://ukstudyoptions.com/uk-university-groups-a-quick-guide/>.

⁵ For example, <https://urapcenter.org/cdn/storage/PDFs/vhKaHMN3xHTJiFjNT/original/vhKaHMN3xHTJiFjNT.pdf>.

⁶ <https://www.umultirank.org/> U-Multirank did not appear in Fig. 1 because they create 6 different types of readymade rankings.

Table 1 Leagues of RUR, former URAP and U-Multirank

RUR 2020		URAP 2017-8		U-multirank 2020	
n = 829		n = 2500		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

¹ “Universities of science and technology rankings” out of the 6 readymade rankings of the U-multirank

Furthermore, instead of publishing the rank of universities, U-Multirank rates them 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 categories⁷ can also be considered leagues.

Table 1 shows that the leagues in these rankings were developed based on the ranks or overall value or rank of the universities. We find it problematic to rank all universities according to all indicators and then classify them into leagues based on that ranking (as shown in Table 1). This is similar to awarding an Olympic gold medal based on the overall performance achieved in all sports. Just as it is worthwhile to award medals by sport at the Olympics, it would be worthwhile to determine the ranking of universities within leagues. In another analogy, Real Madrid does not play football with a county team. Interpreting this in the case of universities, Table 2 provides an example of very different operating conditions for universities. Table 2 compares the total operating revenue of Harvard University (A)⁸ and Hungary’s entire national higher education budget (B)⁹ for the past 5 years (2014–2018). (For conversion between currencies, the annual average of daily central exchange rates of Hungarian Central Bank was employed).¹⁰ Harvard’s data were approximately 20–60 times higher than the national data from Hungary (Banász 2019). With this example, we do not want to suggest that only money matters, but, obviously it matters greatly. 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’ (van der Wende (2008), p. 67). The author proposes the classification of universities based on their mission and characteristics.

⁷ <https://www.umultirank.org/export/sites/default/press-media/documents/Rank-Group-Calculation-2018.pdf>.

⁸ <https://finance.harvard.edu/annual-report>.

⁹ All items of The Closing Accounts Acts of Hungary, which included the phrase ‘higher education’ <http://kfib.hu/hu/torvenyek-zarszamasok>.

¹⁰ <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 Budget of Harvard University vs. Hungary

		Year				
		2014	2015	2016	2017	2018
Harvard University (A)	In millions USD	4409	4526	4777	4999	5215
Hungary	In millions HUF	48,121	37,251	41,161	23,196	51,457
Exchange rate	HUF ^a / USD	233	279	281	274	270
Hungary (B)	In millions USD	207	133	146	85	190
Harvard / Hungary (A/B)	Rate	21	34	33	59	27

^aHungarian Forint

As Kováts (2015) points out, institutions do not have to be good in all indicators, only in those that are in line with the institutions' strategies and policies. Kosztyán et al (2019b) also suggests that with the help of creating leagues, countries (or institutions) can determine which indicators they should improve to move up to a better league.

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". Other scholars suggest using different grouping algorithms to avoid comparability problems and to create homogeneous groups of universities. A short summary of these papers can be seen in Table 3.

The common point in these papers is that all of them 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 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 our opinion, it was not a useful research aim to compare the rankings and publications because, as shown in Fig. 1, we also considered teaching, internationalization 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)

Table 3 Leagues by clustering

Citation	The subject of the investigation			Clustering method used to form the leagues	Leagues found
	Ranking	Year	Cases		
Valadkhani and Worthington (2006)	-	1998–2002	36 Australian Universities	Hierarchical cluster analysis, factor analysis	Higher-performing, lower-performing
Nolle (2010)	Sunday Times University Guide	2010	144 UK Universities	Self-organizing map, k-means	Best-performing, middle-performing, old universities, worst-performing
Jarocka (2012)	ARWU	2011	101 Universities	k-means	Cluster 1-5, 1 is the best
Barnett and Moher (2019)	-	2016–2017	app. 750 Universities	Bayesian clustering model	Cluster 1-5, 1 is the best
Johnes (2018)	The Complete UniversityGuide	2018	129 UK Universities	Data envelopmentanalysis (DEA)	Tier 1-6, 1 is the best
Raponi et al (2016)	-	2009–2010	55 Italian Economics faculties	Biclustering	Public, private
Kosztyán et al (2019b)	Universitas21	2014	50 Countries	Biclustering / iBBiG, BiCARE	A: Upper league B: Middle league C: Lower league

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 abovementioned studies; instead of using a university ranking, he worked with the enriched version of the European Tertiary Education Register (ETER). He classified over 2000 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 rather to 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.

The papers of Raponi et al (2016) and Kosztyán et al (2019b) are different from the analyses mentioned thus far because the authors used a biclustering technique to create leagues. Raponi et al (2016) used productivity, teaching, research, and internationalization indicators of 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. On the other hand, Kosztyán et al (2019b) used the Universitas21 (U21) ranking of 50 countries' entire higher education systems to show that instead of rankings, the leagues of countries should be used. They proposed using 3 leagues, namely, the lower, middle, and upper leagues. Their method can be considered 'fair' because instead of using the whole predefined indicator set, the biclustering algorithm decides which indicators (and countries) belong to a certain bicluster, that is, league. In this way, the countries in the same league can be compared to each other across those indicators that characterize that league.

We do not agree with forming leagues based on overall rankings, such as those shown in Table 1; instead, we recommend that rankings be formed only within leagues containing similar universities. We recommend using league-based rankings rather than rank-based leagues. Therefore, we propose unsupervised biclustering methods to create leagues that specify indicators and universities simultaneously.

The method of biclustering is most widespread in bioinformatics. It also has much potential within the social sciences, as it can be used to define leagues, 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 3), Raponi et al (2016) applied this method to the data of 55 Italian faculty of economics concerning the academic years 2009–2010.

3 Methodology

To create leagues, biclustering was applied. Clustering shows which universities are similar in all ranking indicators. In contrast, biclustering shows which universities are similar in which ranking indicators. Therefore, against the clustering and two-step clustering, biclustering finds indicators and comparable universities simultaneously. Three kinds of leagues are specified. League A (C) includes indicators and universities whose values are significantly *higher (lower)* than those of the remainder. While universities within the type B league have the same characteristics, such as the same strengths and weaknesses, the variance in indicators is minimal. Since biclustering is an NP-hard problem (Tanay et al 2002), leagues are specified iteratively. Based on the specification of homogeneity, the type A league collects the top universities, the type B league collects the middle, and the type C league collects the lower-ranked universities. However, in addition, owing to the simultaneous indicator and university selections, the selected indicators of the leagues are also identified. The applied method allows the creation of leagues within a league. Creating leagues by just one level of biclustering has important properties. The submatrices corresponding to the league are more homogeneous than the remaining ones), *stable* for permutation of rows and columns, and *relevant*, which means the partial rankings of universities within a league are highly correlated by the original ranking; in other words, biclustering collects the relevant indicators (Koszytan et al 2019b).

Koszytan et al (2019b) used this method on Universitas21 data for the year 2014, which ranks 50 countries according to their entire higher education systems. Our paper fits in line with those papers, but our paper is based on global institutional rankings.

The research aims to present the results of the biclustering method applied to a global nonthematic university ranking that is as diverse as possible in terms of the number of ranking areas. Biclustering 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 Fig. 1). There are two rankings in Fig. 1 containing all four main areas (I IV): THE and RUR.

The latest data of RUR (World University Ranking 2020¹¹) on 829 institutions were selected for analysis because the weights of the reputation surveys in RUR are less than those in THE. Bowman and Bastedo (2011) that anchoring effects have an influence on reputational assessments. More precisely, being ranked highly in a particular ranking increases reputation, not the other way around. This means that rankings that also include reputation surveys are particularly troubling because reputation surveys are biased toward elite universities. Furthermore, as Berghoff and Federkeil (2006) and Bowman and Bastedo (2009) point out, reputation is stable and changes very slowly over time. The study of Lawrence and Green (1980) reviewed 5 reputational studies and concluded that the reputations of UC Berkeley, Chicago, Harvard, Michigan, Princeton, Wisconsin, and Yale were secured and changed almost nothing in the 1925-1970 period. Additionally, Berghoff and Federkeil (2006) state that reputation differs between social groups and varies in terms of subjects and geographical areas. These conditions make it even harder to measure reputation. Because of these factors, we chose not to use THE (as surveys count higher in their rankings than in the RUR). Despite concerns about rankings that include reputation surveys, we did not want to ignore them, as many rankings contain the results of such surveys. Table 4 shows the construction of the RUR. Only the 20 basic indicators were

¹¹ https://roundranking.com/ranking/indicator_ranking.html.

Table 4 The leagues formed on RUR 2020

			Leagues			
			A	B	C	
No. of institutions at a threshold of 0.50:			398	280	430	
			0.75:		174	192
			0.90:		78	81
No. of indicators at a threshold of 0.50:			17	10	17	
			0.75:		11	15
			0.90:		3	15
INDICATORS						
Teaching (T) 8-8% 40%	T1	Academic staff / students				
	T2	Academic staff / bachelors degrees awarded				
	T3	Doctoral Degrees awarded / academic staff	X		X	
	T4	Doctoral degrees awarded / bachelors 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 to determine Leagues A and C:

- X: threshold = 0.5
- X: threshold = 0.5 and 0.75
- X: threshold = 0.5, 0.75 and 0.9
- without X (F3 in League C): threshold = 0.9

employed; the four aggregated subindicators and the overall scores were ignored. The abbreviations in Table 4 are according to the original categorization of RUR. This does not completely overlap with the way we classified the indicators in Fig. 1.

Our detailed (step-by-step) analyses on the RUR 2020 ranking are the same as those described in the Kosztyán et al (2019b) article. Analyses were performed in R software¹². There are several methods of biclustering. Two of these methods were used by us to determine leagues: iterative biclustering of genes (iBBiG) and biclustering analysis and results exploration (BicARE). The criterion for the iBBiG method is to normalize the data

¹² R code available at <https://kmt.gtk.uni-pannon.hu/kutatas/RUR/EN/BIC/RUR.html>.

to 0-1 intervals. Within this 0-1 interval, a threshold can be specified (which is the median (Med) by default, i.e., 0.5). The iBBiG method gathers universities and ranking indicators into a bicluster that is above this threshold. This decision binarizes the original dataset. iBBiG is used to determine League A (best). If the same iBBiG method is used on the reverse (1-X) of the basic data (X), the C league (laggards) can be obtained.

The iBBiG algorithm balances the homogeneity (in this case entropy) of the selected submatrix with the size of the league. Formally, the iBBiG algorithm maximizes the following target function if the original binary dataset is denoted as B :

$$\max \leftarrow \text{score} := (1 - H_B)^\alpha \begin{cases} \sum_i \sum_j [B]_{ij}, & \text{if } Me(B) > \tau \\ 0, & \text{if } Me(B) \leq \tau \end{cases} \quad (1)$$

where $score$ is the score value of the submatrix (bicluster, league) $B \in B$. H_B is the entropy of submatrix B ; $Me(B)$ is the median of bicluster B ; $\alpha \in [0, 1]$ is the exponent; and τ is the threshold. If τ or α is increased, we obtain a smaller but more homogeneous submatrix. Former studies (see e.g. Gusenleitner et al 2012; Kosztyán et al 2019b,) suggest that the balance exponent (α) must be set to 0.3, while the threshold (τ) was 0.5, 0.75, and 0.9 to specify Med, Q1, D1, and leagues, respectively.

The B league (midfield) is defined using the BicARE method, which collects the submatrix (universities and ranking indicators) of basic data where the variance is lowest.

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

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

In addition to the 0.5 threshold (median) applied for the iBBiG method in the study of Kosztyán et al (2019b), the present research refines the results with other thresholds: 0.75 (upper and lower quartiles) and 0.9 (upper and lower deciles) for leagues A and C.

4 Results

Table 4 summarizes the results as follows: the number of universities and the indicators classified into each league (A, B, C) using the thresholds 0.5, 0.75, and 0.9, respectively. The threshold does not affect the indicators in League B. The higher the threshold, the fewer the number of universities and indicators entering the A and C leagues. Table 4 also indicates the specific indicators included in each league.

At a threshold of 0.5, the indicators marked with X were classified into Leagues.

Out of the 20 variables:

- (i) Both Leagues 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, there are 3 indicators that were missing from both Leagues A and C (these were not included in League B either)¹³: academic staff per student (T1), which measures the quality of education, as the more lecturers per student, the more effective the education; academic staff per bachelor's degree awarded (T2), which narrows the previous indicator of

¹³ <https://roundranking.com/methodology/methodology.html>.

undergraduate-level bachelor's programs as this level is the basis of higher education in the world; and papers per unit of research income (F3), which shows the financing level of publications.

The 10 indicators in finding (ii) are those with the lowest variance in the universities included in League B; however, they are decisive in that their high (low) value is required for League A (C), in addition to the 7 other indicators. These 10 variables played a role in the development of all three leagues: citations per academic and research staff (R1), normalized citation impact (R3), papers per academic and research staff (R4), share of international academic staff (I1), international level (I5), institutional income per academic staff (F1), and research income per academic and research staff (F4). Furthermore, an interesting finding is that all three reputation surveys were included here: world teaching reputation (T5), world research reputation (R5), and reputation outside region (I4).

In addition to the 10 indicators listed above, the high (low) value of the 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 by academic staff (T3), by bachelor's degrees awarded (T4), by admitted PhD students (R1), share of international students (I2), share of international coauthored papers (I3), institutional income per student (F2), and research income per unit of institutional income (F5).

To refine the results, Leagues A and C were also generated to higher thresholds by the iBBiG method. This modifies columns A and C in Table 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 remained in Leagues A and C.

In the following, we will focus only on League A, which contains the best results. At the threshold of 0.5, a high value of 17 indicators ensured the classification of an institution in the League A; at the threshold of 0.75, 11 indicators; and at the threshold of 0.9, only 3 indicators. The latter means that if we collect universities in a league whose 0-1 normalized data are above 0.9, only 3 indicators will determine the best institutions. These are the 3 international reputation surveys based on the annual data of the Academic Reputation Survey of Clarivate Analytics (which was implemented by Ipsos Media CT): world teaching reputation (T5), world research reputation (R5), and reputation outside region (I4). The latter also takes into account teaching and research, but only the opinions of respondents who live outside the university region matter. The regions considered are Asia, Europe, North America, Oceania, and South America.

These surveys included universities that received at least 1 vote. Participation in these surveys was by invitation only and did not operate on a self-report. It is not permitted to vote for your own university. Annually, 10,000 respondents give 60,000 votes for universities. Each respondent can select up to 15 universities that are the most powerful in teaching/research based on their opinion.

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 are like large corporations that remain stable while small companies go out of business or merge with other companies.

5 Summary and conclusions

In this work, we showed how to form university leagues with biclustering through the example of the Round University Ranking 2020. Two-way clustering algorithms are used to create a less controversial ranking of universities, pooling them into leagues. The leagues by intention contain universities with similar profiles and pooled indicators on which the similarity of the profiles of the universities are based. The method can be considered a dimension reduction or feature selection method that helps to interpret leagues of universities as well as the rankings within the leagues.

This method has already been used to establish leagues of countries based on their higher education system as a whole Kosztyán et al (2019b) and to group economics faculties into leagues within a country (Italy) Raponi et al (2016). Compared to these studies, the novelty of the present paper is twofold. The results were presented on the basis of an international university ranking and on using we investigated the effect of the change of threshold to the obtained leagues, in particular the change of university membership into a particular league and the entrance or dropout of indicators.

Partial rankings of indicators within leagues of comparable universities better satisfy the requirement of fair rankings. The paper suggests leagues with an unsupervised biclustering method. The proposed method ensures that the universities within the leagues are comparable. To specify leagues, the selection of universities and indicators must be done simultaneously, while the different thresholds to specify Q1, D1 leagues are considered as well. Finally, the overlap of biclustering highlights indicators that must be improved to increase the ranking and/or league positions. To the best of our knowledge, traditional clustering or two-step clustering methods do satisfy these requirements.

We examined 20 indicators of the 829 universities included in the RUR 2020. Two of the indicators turned out to be irrelevant, namely, T1 'Academic staff per student' and T2 'Academic staff per bachelor's degree awarded'. A university ensures its membership in League A, in the club of excellence, if it is significantly above the median value of the 17 indicators. 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 three qualitative and the most subjective indicators proved to be the most important. Directly influencing these reputation indicators considered unethical, but there are many indirect and innovative methods that enough universities can utilize.

Partial and regional rankings also try to increase the comparability of universities; nevertheless, the applied selection of universities and indicators is arbitrary. The proposed unsupervised biclustering methods ensure the comparability of universities within a league. Therefore, the main implication for scholars are to apply this tool to specify the set of leagues that are ranked. In addition, universities within the leagues are comparable and can also be ranked by the specified indicators. The main benefit of the proposed biclustering method for university management is to identify those indicators that must be improved to increase their ranking and league positions. The main benefit for students is to see not only the ranking positions and the indicator values of a given university but also similar universities around the world in a particular league.

As a future research direction, another dimension reduction method can be proposed. Applying a principal component analysis (PCA) or factor analysis (FA) on the subset of indicators assigned to a given league, new composite and meaningful indicators can be

obtained and used for later ranking of universities without ad hoc waiting of the indicators.

Author Contributions These authors contributed equally to this work.

Funding Open access funding provided by University of Pannonia. This work was supported by the TKP2020-NKA-10 project financed under the 2020-4.1.1-TKP2020 Thematic Excellence Programme by the National Research, Development and Innovation Fund of Hungary, the National Research, Development and Innovation Fund (TUDFO/51757/2019-ITM, Thematic Excellence Program), and by the Higher Education Excellence Program of the Ministry of Human Capacities in the frame of the Artificial Intelligence research area of Budapest University of Technology and Economics (BME FIKP-MI/SC).

Data availability and materials The data are available at: https://roundranking.com/ranking/indicator_ranking.html#academic-staff-per-students_SO-2020 (see footnote 11).

Code availability <https://kmt.gtk.uni-pannon.hu/kutatas/RUR/EN/BIC/RUR.html>.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Banász, Z.: R&D trends in Hungary. In: Pintér G, Csányi S, Zsiborács H (eds) *Innovation Challenges in the 21st Century*. Georgikon Faculty, University of Pannonia, pp. 14–22, (2019) https://napok.georgikon.hu/hu/cikkadatbazis/cikkek-2012/cat_view/3-cikkadatbazis/90-2019/98-vi-szekcio-kozgazdasagtan-i
- Barnett, A.G., Moher, D.: Turning the tables: A university league-table based on quality not quantity. *F1000Research* **8**(31316755):583–583 (2019). <https://doi.org/10.12688/f1000research.18453.2>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6611132>
- Barr, R.S., Durchholz, M., Seiford, L.M.: Peeling the DEA onion: Layering and rank-ordering DMUs using tiered DEA. Tech. rep., Southern Methodist University, (2000). <http://faculty.smu.edu/barr/pubs/layer.pdf>
- Bell, A.R., Brooks, C.: Is there a 'magic link' between research activity, professional teaching qualifications and student satisfaction?. *High. Educ. Policy* **32**(2), 227–248 (2019). <https://doi.org/10.1057/s41307-018-0081-0>, <https://link.springer.com/article/10.1057/s41307-018-0081-0>
- Bengoetxea, E., Buéla-Casal, G.: The new multidimensional and user-driven higher education ranking concept of the European union. *Int. J. Clin. and Health Psychol.* **13**(1), 67–73 (2013). [https://doi.org/10.1016/S1697-2600\(13\)70009-7](https://doi.org/10.1016/S1697-2600(13)70009-7), <http://www.sciencedirect.com/science/article/pii/S1697260013700097>
- Berghoff, S., Federkeil, G.: Reputation indicators and research performance. In: Conference presentation at the 2nd meeting of International Ranking Expert Group (IREG), Berlin, Germany, 18–21 May 2006 (2006)
- Borden, V.M.H., McCormick, A.C.: Accounting for diverse missions: can classification systems contribute to meaningful assessments of institutional performance? *Tert. Educ. Manag.* **26**(3), 255–264 (2020). <https://doi.org/10.1007/s11233-019-09045-w>
- Bowman, N., Bastedo, M.: Getting on the front page: organizational reputation, status signals, and the impact of U.S. News and World Report on student decisions. *Res. High. Educ.* **50**, 415–436 (2009)

- Bowman, N.A., Bastedo, M.N.: Anchoring effects in world university rankings: exploring biases in reputation scores. *High. Educ.* **61**(4), 431–444 (2011). <https://doi.org/10.1007/s10734-010-9339-1>
- Daraio, C., Bonaccorsi, A.: Beyond university rankings? Generating new indicators on universities by linking data in open platforms. *J. Am. Soc. Inf. Sci.* **68**(2), 508–529 (2017). <https://doi.org/10.1002/asi.23679>
- Dolnicar, S., Kaiser, S., Lazarevski, K., et al.: Biclustering: overcoming data dimensionality problems in market segmentation. *J. Travel Res.* **51**(1), 41–49 (2012). <https://doi.org/10.1177/0047287510394192>
- Genc, E.G., Basar, O.D.: Comparison of country ratings of credit rating agencies with moora method. *Bus. Econ. Res. J.* **10**(2), 391–404 (2019). <https://doi.org/10.20409/berj.2019.175>, <https://www.berjournal.com/comparison-of-country-ratings-of-credit-rating-agencies-with-moora-method>
- Griffith, A., Rask, K.: The influence of the us news and world report collegiate rankings on the matriculation decision of high-ability students: 1995–2004. *Econ. Educ. Rev.* **26**(2), 244–255 (2007). <https://EconPapers.repec.org/RePEc:eee:ecoeud:v:26:y:2007:i:2:p:244-255>
- Gusenleitner, D., Howe, E.A., Bentink, S., et al.: ibbig: iterative binary bi-clustering of gene sets. *Bioinformatics* **28**(19), 2484–2492 (2012). <https://doi.org/10.1093/bioinformatics/bts438>, https://arxiv.org/abs//oup/backfile/content_public/journal/bioinformatics/28/19/10.1093_bioinformatics_bts438/1/bts438.pdf
- Helbig, M., Ulbricht, L.: Perfekte Passung: Finden die besten Hochschulen die besten Studenten? In: Verbeet M, Trepte S (eds) *Allgemeinbildung in Deutschland*. VS Verlag für Sozialwissenschaften, pp. 107–118, (2010). https://link.springer.com/chapter/10.1007%2F978-3-531-92543-1_7
- Horstschräer, J.: University rankings in action? the importance of rankings and an excellence competition for university choice of high-ability students. *Econ. Educ. Rev.* **31**(6), 1162–1176 (2012). <https://doi.org/10.1016/j.econedurev.2012.07.018>, www.sciencedirect.com/science/article/pii/S0272775712001045
- Jarocka, M.: University ranking systems-from league table to homogeneous groups of universities. *Int. J. Soc. Behav. Educ. Econ. Bus. Ind. Eng.* **6**(6), 1377–1382 (2012). <https://pdfs.semanticscholar.org/b892/8175625ae9e2e5d0da60f3b512ec497579cb.pdf>
- Johnes, J.: University rankings: What do they really show? *Scientometrics* **115**, 585–606 (2018). <https://link.springer.com/article/10.1007%2Fs11192-018-2666-1>
- Kosztján, Z.T., Banász, Z., Csányi, V.V., et al.: Examining the mobility of higher education applicants by economic network models. *Statistikai Szemle* **97**(11), 1007–1049 (2019a). <https://doi.org/10.20311/stat2019.11.hu1007>, http://www.ksh.hu/statszemple_archive/all/2019/2019_11/2019_11_1007.pdf
- Kosztján, Z.T., Banász, Z., Csányi, V.V., et al.: Rankings or leagues or rankings on leagues? - ranking in fair reference groups. *Tert. Educ. Manag.* **25**(4), 289–310 (2019b). <https://doi.org/10.1007/s11233-019-09028-x>
- Kováts, G.: “New” Rankings on the Scene: The U21 Ranking of National Higher Education Systems and U-Multirank, pp. 293–311. Springer International Publishing, Cham (2015) https://doi.org/10.1007/978-3-319-20877-0_20
- Lawrence, J.K., Green, K.C.: A question of quality: the higher education ratings game. Tech. rep., Education Resources Information Center (ERIC), (1980). <https://eric.ed.gov/?id=ED192667>
- Lepori, B.: The heterogeneity of European higher education institutions: a configurational approach. *Studies in Higher Education*, pp. 1–17 (2021). <https://doi.org/10.1080/03075079.2021.1968368>
- Marginson, S.: Open source knowledge and university rankings. *Thesis Eleven* **96**(1), 9–39 (2009). <https://doi.org/10.1177/0725513608099118>
- Mussard, M., James, A.P.: Engineering the global university rankings: Gold standards, limitations and implications. *IEEE Access* **6**, 6765–6776 (2018). <https://doi.org/10.1109/ACCESS.2017.2789326>, <https://ieeexplore.ieee.org/document/8247171>
- Nolle, L.: Cluster-based benchmarking of universities as an alternative to league tables, vol. XXVI. In: *Research and Development in Intelligent Systems*, pp. 499–504. Springer, Oxford (2010) https://doi.org/10.1007/978-1-84882-983-1_40
- Petrarca, F., Terzi, S.: The global competitiveness index: an alternative measure with endogenously derived weights. *Qual. Quant.* **52**, 2197–2219 (2018). <https://doi.org/10.1007/s11135-017-0655-8>
- Raponi, V., Martella, F., Maruotti, A.: A biclustering approach to university performances: an Italian case study. *J. Appl. Stat.* **43**(1), 31–45 (2016). <https://doi.org/10.1080/02664763.2015.1009005>
- Soh, K.: The seven deadly sins of world university ranking: a summary from several papers. *J. High. Educ. Policy Manag.* **39**(1), 104–115 (2017). <https://doi.org/10.1080/1360080X.2016.1254431>
- Tanay, A., Sharan, R., Shamir, R.: Discovering statistically significant biclusters in gene expression data. *Bioinformatics* **18**(suppl_1):S136–S144 (2002). https://doi.org/10.1093/bioinformatics/18.suppl_1.S136, https://arxiv.org/abs//oup/backfile/content_public/journal/bioinformatics/18/suppl_1/10.1093/bioinformatics/18.suppl_1.s136/2/18s136.pdf

- Valadkhani, A., Worthington, A.: Ranking and clustering Australian university research performance, 1998–2002. *J. High. Educ. Policy Manag.* **28**(2), 189–210 (2006). <https://doi.org/10.1080/13600800600751101>
- Vogel, G.: Good grades for Germany's project to build an Ivy league. *Science* **351**(6273), 545–545 (2016). <https://doi.org/10.1126/science.351.6273.545>, <https://science.sciencemag.org/content/351/6273/545>, <https://arxiv.org/abs/https://science.sciencemag.org/content/351/6273/545.full.pdf>
- Wende, M.V.D.: Rankings and classifications in higher education: a European perspective. In: *Higher Education*, pp. 49–71. Springer, Dordrecht (2008) https://doi.org/10.1007/978-1-4020-6959-8_2

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Zsuzsanna Banász^{1,2}  · Zsolt T. Kosztyán^{2,3,4,5}  · Vivien V. Csányi³  ·
András Telcs^{2,3,6,7} 

✉ Zsuzsanna Banász
banasz.zsuzsanna@gtk.uni-pannon.hu

Zsolt T. Kosztyán
kosztyan.zsolt@gtk.uni-pannon.hu

Vivien V. Csányi
csanyi.vivien@gtk.uni-pannon.hu

András Telcs
telcs.andras@gtk.uni-pannon.hu

- ¹ Department of Economics, University of Pannonia, Egyetem str. 10, Veszprém, Hungary
- ² MTA-PE Budapest Ranking Research Group, Eötvös Loránd Research Network (ELKH), Egyetem str. 10, Veszprém, Hungary
- ³ Department of Quantitative Methods, University of Pannonia, Egyetem str. 10, Veszprém, Hungary
- ⁴ Institute of Advanced Studies, Koszeg, Koszeg, Hungary
- ⁵ Research Centre of Faculty of Business and Economics University of Pannonia, Egyetem str. 10., Veszprém, Hungary
- ⁶ Head of Department of Computational Sciences, MTA Wigner Research Center for Physics, Budapest, Hungary
- ⁷ Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Budapest, Hungary