University Rankings; what do they measure?

Friso Selten

March 12, 2019

Abstract

University rankings are an important tool for university employees, policy makers, journalist and students to asses and compare the quality of universities. Therefore these rankings have a big impact on the status of a university, the number of students they attract and their ability to raise funding. But these rankings receive mixed reactions. On the one hand people are content with them because they provide a clear and seemingly objective tool to assess a university's performance. On the other hand there has been criticism that the rankings are subjective and use wrong measurements. This research will investigate the methodology of the three major university rankings, by looking from a data science perspective at how they are constructed. Using Correlations, Principal Component Analysis and Clustering techniques on this data, it is going to be studied what and how the rankings measure.

1 Personal details

My email `Friso.Selten@gmail.com`

My supervisors email `P.T.Groth@uva.nl`

In collaboration with Centre for Culture and Technology; Curtin University

The wiki on my github account `https://github.com/ccat-lab/university-clustering`

2 Research questions

The aim of this research is getting a better understanding of what it is university rankings really measure. This is going to be done by analyzing the data from individual rankings, combining data from different rankings and add external data sources to the ranking data. Doing this we want to provide better insights in how the rankings are constructed. For this the following research questions are going to be answered;

- Are there latent factors underlying the ranking data, and if so what are these latent factors?
- What longitudinal trends can be observed in the rankings?
- How do the major university rankings compare to each other?

3 Related Literature

The public sector has been subject to significant administrative reforms over the past decades. Aim of these reforms was to increase efficiency, effectiveness and accountability [1]. This demand has sparked 'a flood of social measures' designed to evaluate the performance of organizations and make accountability more transparent [1, 2]. Universities, as part of this public sector, have also been subject to these reforms.

One of the measures taken in the higher education domain to satisfy this demand for accountability and transparency is the popularization of University Rankings (UR's). These rankings make it easy to quantify results and compare universities. The rankings were thus met with relief and greed by policy makers and journalists but also by students and universities themselves. The idea of comparing Universities dates to the 1880's, it however gained major attention in 2003 with the release of the Shanghai league table [3]. Many new UR's were established since then; the most notable ones: the THE-QS and the Webometrics Ranking of World Universities in 2004, the NTU Ranking and the CWTS Leiden Ranking in 2007. In 2009, the THES and QS rankings split up and published their separate rankings since 2010.

The impact of UR's is unmistakable. They affect judgments of university leaders and prospective students. But also, the decisions made by policy makers and investors. [4]. For certain politicians, it has even become a goal by itself to have high ranked universities in their country [5].

The rankings, however met with enthusiasm by some, also quickly became subject to criticism [3]. These criticisms can be divided into two categories; Fundamental criticism, i.e. rankings do not measure the right things, and methodical criticism, i.e. the right things are not measured well.

The fundamental criticism focuses on the fact that there are certain social choices made that will boost the scores of certain universities and decrease the scores of other universities. Certain researchers argue that the rankings do not measure university quality, but that they are slotted in a preordained global hierarchy [6]. One of the reasons to suggest this is that English speaking countries dominate the rankings. This might be caused by the fact that citations, which are an important factor in rankings, are often only measuring citations in a subset of mostly English journals [7]. Another criticism from a social perspective argues that university rankings are 'self-fulfilling prophecies'. The rankings create expectations about a university, it can therefore be argued that they are therefore no objective measure but influence influence the behavior of and towards a university. For example; prior rankings influence surveys that determine future ones, they influence funding decisions and universities conform their activities to the ranking criteria [2, 6].

Methodical criticism focuses on the fact the measurements of the UR's are wrong or at least arbitrary. This criticism centers around the different components the rankings measure and the weights that are assigned to these components. Marginson et al. and Saisana et al. [4, 5] show that making little changes to the weights can cause a major shift in ranking positions. The rankings are thus highly based on their underlying model and certain research even argues that, beneath the top 10, the rankings can therefore not be used to compare individual universities [5].

This research focuses mostly on studying this second criticism. By perform-

ing multiple analysis on the data from the rankings, combining the data from the different rankings data and adding external data sources we are going to study what it is the rankings really measure. It contributes to the existing literature by taking a data science approach; where more data is analyzed then in other study's, by analyzing more years and multiple rankings; and new techniques are used, like clustering and the combining with external data sources.

4 Methodology

4.1 Description of the data

The main data sources for this project are the three university rankings ARWU, THE and QS. These all collect data about universities on a scale from 0 to 100. They measure mostly overlapping concepts, but use different names and measurements for overlapping concepts (see table 1 where the broader concepts the rankings measure are mapped to the specific ranking variable and appendix A for a full overview of what the rankings exactly measure). For a full overview of the number of universities which are measured each year see table 3 in appendix B.

One of the research questions is finding if there are latent factors the rankings measure. For this, next to the data from the rankings themselves, external data is going to be added to the rankings. Although not yet entirely certain what external data sources are relevant to use, an interesting approach would be to add geographical data like in the research of Marginson [8]. This would help prove or disprove the point made by researchers like Pusser [6] who say that rankings do not measure university quality, but just a reflection of global hierarchy between countries.

The processing of this data consists of scraping the ranking websites and linking all universities together. This however has already been done, see appendix C.

Table 1: Main variables the rankings measure

Concept	ARWU	THES	QS
Education	Quality of education	Teaching Quality	Teaching
Citations	Research output	Citations	Citations
Reputation	Quality of Faculty	Employer reputation	Research
Internationalization	-	Internationalization	International outlook
Income	-	-	Industry Income

4.2 Exploratory analysis

Some explanatory analysis has already been performed. Here the most common techniques seen in previous research on university rankings are performed on the latest ranking data (year 2018 for ARWU and 2019 for QS and THE) to retrieve deeper insights into the data that is collected. The most important results of this analysis will be discussed here.

First, a general Spearmans correlation within and between the three total ranking has been performed to gain insights in the data. This approach is also

taken in the research of Aguillo et al. [9]. This analysis showed that the rankings are highly correlated with each other. Especially the top fifty ranked universities in the different rankings are very similar, after these initial fifty the correlations however a much weaker relationship between the three rankings is observed. Also, if we dive further in the ranking we see that very similar components of the rankings are not very correlated, for example the number of citations which all three rankings measure show a weak correlation.

Also, Principal Component Analysis (PCA) has been used. The first two principal components explained a high amount of the variance in the rankings. The loading's of the PCA can be viewed in table 4 in appendix E. These loading's are interesting because they describe how the Principal Components are constructed. Which is also used in the research of Docampo and Dehon et al. [10, 11].

Lastly, a start has been made to cluster the data, multiple clustering algorithms (K-means, hierarchical clustering, HDbscan) have been used to gain insights into the data. The results of one of one of these clustering techniques (K-means) combined with the PCA to visualize can be seen in figure 1 in appendix D. What exactly these clusters and PCA tell us, must be further investigated.

4.3 Evaluation

The methods used in this research are unsupervised. It is therefore impossible to evaluate the results using direct measures like precision or recall. Therefore the results from the analysis have to be critically evaluated using statistical tests and knowledge from previous study's.

Next to assessing how the results of this study fit in the outcomes of previous research, this research is performed in collaboration with the Centre for Culture and Technology at the Curtin University; who have a lot of knowledge on this topic. Therefore they are actively involved in deciding what analysis are suitable, if the results from the analysis make sense and how the results can be interpreted.

5 Risk assessment

Biggest risk in any data science project is not being able to collect, or have access to the necessary data sources. This risk is already addressed, because all data from the university ranking websites is already scraped and stored.

Another risk might be that the data is too noisy to gain useful insights from. This might however in this case not be so much of a risk, but rather an interesting observation which can be further investigated. Because if a ranking is based upon only noise, what does it tell us? Thereby, the exploratory analyses already shows interesting results to further investigate.

6 Project plan

Table 2: Project planning

Week	Deliverable
0	All data from university websites is scraped, universities linked to grid-id's.
1	Literature review and and analysis within ranking data finished, explore external data.
2	longitudinal analysis finished, start exploring external data.
3	External data collected, first analysis on external data performed.
4	Reflect on performed analyses to see what is missing, decide on what direction to proceed.
5	First analysis parts are written, continue with performing analyses.
6	Most analysis are performed, check if they help answering research questions
7	All analysis are performed, start writing results section, validate results in literature.
8	Result section is written, reflect and validate results.
9	Reflection points addressed, literature, method and result are a logically connected story.
10	Conclusion, Discussion and introduction are written.
11	Write last parts, putting everything together.
12	Thesis finished.

References

[1] Barbara S Romzek. "Dynamics of public sector accountability in an era of reform". In: *International review of administrative sciences* 66.1 (2000), pp. 21–44.

[2] Wendy Nelson Espeland and Michael Sauder. "Rankings and reactivity: How public measures recreate social worlds". In: *American journal of sociology* 113.1 (2007), pp. 1–40.

[3] Konstantinos I Stergiou and Stephan Lessenich. "On impact factors and university rankings: from birth to boycott". In: *Ethics in Science and Environmental Politics* 13.2 (2014), pp. 101–111.

[4] Simon Marginson. "University rankings and social science". In: *European Journal of Education* 49.1 (2014), pp. 45–59.

[5] Michaela Saisana, Béatrice d'Hombres, and Andrea Saltelli. "Rickety numbers: Volatility of university rankings and policy implications". In: *Research policy* 40.1 (2011), pp. 165–177.

[6] Brian Pusser and Simon Marginson. "University rankings in critical perspective". In: *The Journal of Higher Education* 84.4 (2013), pp. 544–568.

[7] Nancy J Adler and Anne-Wil Harzing. "When knowledge wins: Transcending the sense and nonsense of academic rankings". In: *Academy of Management Learning & Education* 8.1 (2009), pp. 72–95.

[8] Simon Marginson. "Global university rankings: Implications in general and for Australia". In: *Journal of Higher Education Policy and Management* 29.2 (2007), pp. 131–142.

[9] Isidro Aguillo et al. "Comparing university rankings". In: *Scientometrics* 85.1 (2010), pp. 243–256.

[10] Domingo Docampo. "On using the Shanghai ranking to assess the research performance of university systems". In: *Scientometrics* 86.1 (2011), pp. 77–92.

[11] Catherine Dehon, Alice McCathie, and Vincenzo Verardi. “Uncovering excellence in academic rankings: A closer look at the Shanghai ranking”. In: *Scientometrics* 83.2 (2009), pp. 515–524.

Appendices

A Detailed overview of ranking data

1. Academic Ranking of World Universities (ARWU)
    - Quality of education
        - Alumni of an institution winning Nobel Prizes and Fields Medals
    - Quality of faculty
        - Staff of an institution winning Nobel Prizes and Fields Medals
        - Highly cited researchers in 21 broad subject categories
    - Research output
        - Papers indexed in Science Citation Index-expanded and Social Science Citation Index
    - Per capita performance
        - Per capita academic performance of an institution
2. Times Higher Education Ranking(THE)
    - Teaching
        - Reputation survey
        - Staff-to-student ratio
        - Doctorate to bachelor's ratio
        - Doctorates awarded to academic staff
        - Institutional income
    - Research
        - Reputation survey
        - Research income
        - Research productivity
    - Citations
        - Bibliometric data provided by Elsevier
    - International outlook
        - Proportion of international students
        - Proportion of international staff
        - International collaboration
    - Industry Income
        - Industry income scaled against the number of academic staff it employs
3. Quacquarelli Symonds Ranking (QS)
    - Academic reputation
        - Academic Survey collect under 80,000 individuals
    - Employer reputation

  - 40,000 responses to our QS Employer Survey
- Teaching Quality
  - Teacher/student ratios
- Citations per faculty
  - Elsevier's Scopus database
- Internationalization
  - International faculty ratio
  - International student ratio

B Number of universities measured per year

Table 3: Number of universities measured per year

Year	ARWU	THES	QS
2005 - 2010	500	-	-
2011	500	200	-
2012	500	402	-
2013	500	400	-
2014	500	400	-
2015	500	401	-
2016	500	800	918
2017	500	981	936
2018	500	1103	980
2019	500	1258	1021

C Description of data processing

Main data sources for this research are of course the ranking data from the university rankings themselves. This is all numerical data which is in a tabular form. This means that the data is relatively clean and easy to process. The data from the rankings for all years available, is already scraped and stored and thus available to use.

The most difficult step for the analysis is linking universities from the different rankings together. In the first ranking a university might be named 'Universiteit of Amsterdam', in the second one 'Amsterdam University' and in the third one 'the University of Amsterdam'. Therefore, to make a comparison between the rankings all universities have to be linked to their id's in the Global Research Identifier Database (GRID). This linkage is done in four steps.

First all perfect string matches are linked together, if no perfect match is found then we search for the id in a database retrieved from Wikidata. This database contains a lot of universities and for a most of them also the GRID id's and ARWU-, QS- and THES id. This ensures that we can link the id from the rankings, which are scraped from the ranking websites, to the GRID id in the Wikidata database. If this also did not yield a connection we continue with

step three. Here we use Python's fuzzywuzzy package, this applies Levenshtein distance to match strings, using manual inspection this method proved to be accurate if the matching score was above 95 percent. The last step is manually match the university's which were not matched using one of the above procedures. After this all universities are linked to their corresponding GRID ID and the data is tidy stored is a CSV so that it can easily be analyzed.

D PCA & K-Means clustering visualization

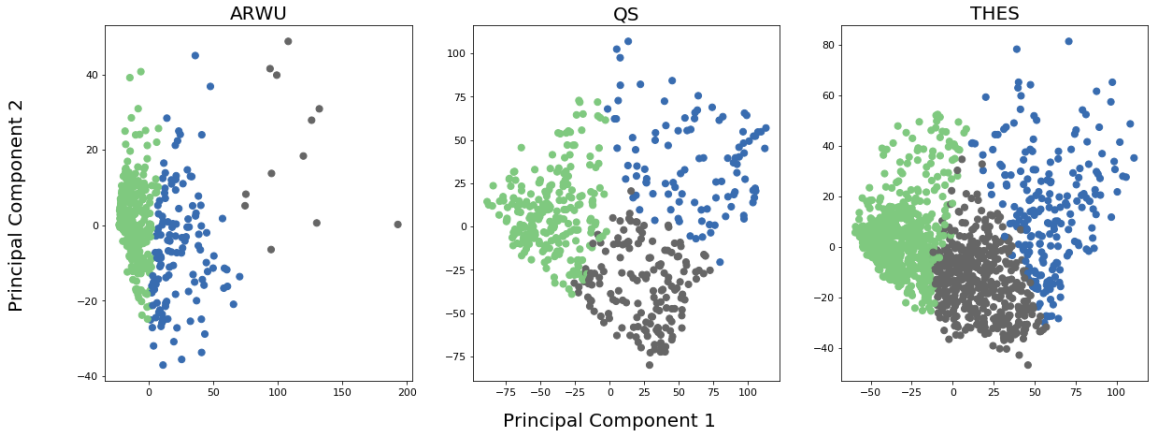

Figure 1: PCA and K-Means Clustering on individual rankings

E PCA Loadings

Table 4: PCA Loadings

ARWU	0.670	0.150
Variable	PCA 1	PCA 2
Alumni	0.422	0.341
Award	0.535	0.467
HiCi	0.392	-0.410
NS	0.427	-0.079
PUB	0.338	-0.690
PCP	0.289	0.119

QS	0.425	0.222
Variable	PCA 1	PCA 2
AcaRep	0.332	0.533
EmpRep	0.374	0.514
FacStu	0.007	0.448
IntFac	0.605	-0.397
IntStu	0.541	-0.290
Citations	0.298	0.082

THE	0.643	0.162
Variable	PCA 1	PCA 2
Teaching	0.290	0.396
Research	0.391	0.416
Citations	0.705	-0.198
Ind Income	0.161	0.624
Int Outlook	0.488	-0.489