# Skyline-Based University Rankings

Georgios Stoupas[1](✉), Antonis Sidiropoulos[2], Dimitrios Katsaros[3], and Yannis Manolopoulos[4]

[1] Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
grgstoupas@csd.auth.gr
[2] International Hellenic University, 57400 Thessaloniki, Greece
asidirop@gmail.com
[3] University of Thessaly, 38221 Volos, Greece
dkatsar@e-ce.uth.gr
[4] Open University of Cyprus, 2220 Nicosia, Cyprus
yannis.manolopoulos@ouc.ac.cy

**Abstract.** University rankings comprise a significant tool in making decisions in our modern educational process. In this paper, we propose a novel university ranking method based on the *skyline* operator, which is used on multi-dimensional objects to extract the non-dominated (i.e. "prevailing") ones. Our method is characterized by several advantages, such as: it is transparent, reproducible, without any arbitrarily selected parameters, based on the research output of universities only and not on publicly not traceable or random questionnaires. Our method does not provide meaningless absolute rankings but rather it ranks universities categorized in equivalence classes. We evaluate our method experimentally with data extracted from Microsoft Academic.

**Keywords:** University rankings · Skyline · Rainbow ranking index

## 1 Introduction

University rankings are of major importance for decision making by prospective students, by academic staff, and by funding agencies. The placement of universities in these lists is a crucial factor and academic institutions adapt their strategy according to the particular criteria of each evaluation system. The reason for this inclination can be understood by considering the "Thomas theorem" from sociology, which states "if men define situations as real, they are real in their consequences" [8]. As mentioned in [2]: "if rank positions between two universities define performance differences as real, they are real in their consequences (although the university ranking shows only slight differences between the universities' scores)".

Probably, the 3 most popular global rankings are: ARWU, QS and THE. Another quite well-known ranking list is Webometrics, whereas there also exist a few ranking lists developed by universities, e.g., CWTS ranking of Leiden


L. Bellatreche et al. (Eds.): ADBIS/TPDL/EDA 2020 Workshops and Doctoral Consortium, CCIS 1260, pp. 347–352, 2020.
https://doi.org/10.1007/978-3-030-55814-7_30

University, École Nationale Supérieure des Mines de Paris, Middle East Technical University, Wuhan University, and Shanghai Jiao Tong University, which founded the ARWU organization. All these lists base their respective ranking on some set of indicators, which differ from one organization to the other. The reader can retrieve these indicators from the respective sites[1, 2, 3, 4, 5]. Despite their popularity, these rankings are heavily criticized for their reproducibility, statistical soundness, etc. [1,4,5,9].

Here, we focus on an academic-based (teaching and research performance) ranking, i.e., in a similar approach to CWTS ranking of the Leiden University, and we propose an orthogonal method to rank academic institutions based on the *skyline* operator, which is applied on multi-dimensional objects to extract the non-dominated (i.e. "prevailing") ones. The contribution of our method and its advantages over the popular university rankings are the following:

- it focuses on the research output of the universities, and does not rely on questionnaires,
- it uses a set of indicators well-known to the whole academic community from Google Scholar metrics,
- it does not use arbitrary weights for each indicator, but treats all indicators equally in a symmetric manner,
- it avoids the nonsense absolute rankings, where there is no serious meaning in claiming that the $i$-th university is better the $(i+1)$-th one.
- it provides a list with a single structure, contrary to the popular rankings, where paradoxically the first few hundreds of universities are ranked in absolute order, whereas the rest follow in groups.
- it is not prone to inconsistent fluctuations from year to year,
- it is fully customizable in the sense that it can use any set of research key-performance indicators.

The structure of the remaining part of this paper is as follows. Section 2 explains the Skyline operator and its derivative, namely Rainbow Ranking. Section 3 gives the results of the application of the Rainbow Ranking to university ranking. Finally, Sect. 4 concludes the article.

## 2 Skyline and Rainbow Ranking

The Skyline operator is used as a database query to filter only those 'objects' that are not worse than any other (they are not dominated) [3]. A useful application of Skylines in scientometrics is reported in [6] where 3-d Skyline sets of 'dominating' researchers for each year of the period 1992–2013 were produced. An extension

[1] https://en.wikipedia.org/wiki/Academic_Ranking_of_World_Universities.
[2] https://en.wikipedia.org/wiki/QS_World_University_Rankings.
[3] https://en.wikipedia.org/wiki/Times_Higher_Education_World_University_Rankings.
[4] https://en.wikipedia.org/wiki/Webometrics_Ranking_of_World_Universities.
[5] https://en.wikipedia.org/wiki/CWTS_Leiden_Ranking.

of the Skyline operator, namely the Rainbow Ranking [7], applies iteratively the Skyline operator until all entities (i.e., scientists) of a dataset have been classified into a Skyline level. More specifically, given a set of scientists $X1$, the first call of the Skyline operator produces the first Skyline level, which is denoted as $S1$. Next, the Skyline operator is applied on the dataset $X1 - S1$, to derive the second Skyline layer, denoted as $S2$. This process continues until all the scientists of the dataset have been assigned to a particular Skyline level $Si$. To give more semantics to the method, a particular value should characterize the Skyline levels. Should this value be the iteration number, then this would convey limited interpretability since the relativeness would be lost. It is crucial to designate the position of scientist among their peers. Therefore, a normalization of this value is necessary. Thus, the $RR$-index of a researcher $a$ is defined as:

$$RR(a) = 100 - 100 \times \frac{|A_{above}(a)| + |A_{tie}(a)|/2}{|A|}$$

where $A$ is the set of scientists, $A_{above}(a)$ is the number of scientists at higher Skyline levels than scientist $a$, and $A_{tie}(a)$ is the number of scientists at the same Skyline level with scientist $a$, excluding scientist $a$. Apparently, it holds that: $0 < RR(a) \leq 100$. A key component for the $RR$-index concept is the number of the Skyline dimensions. By selecting different bibliometric indices as Skyline dimensions, $RR$-index can be fully customizable.

## 3 Ranking Universities with the *RR*-index

Here, the $RR$-index is generalized to higher conceptual levels. We present the dataset used and the Skyline dimensions. Then, we present the experimental results at three levels: at author, faculty and institutional level.

**Dataset**. For our experiments we have used the Microsoft Academic Search (MAS[6]) database. We have downloaded the Microsoft Academic Graph from the Open Academic Graph work-group (AMiner[7]). The initial dataset consisted of 253,144,301 authors with 208,915,369 publications. Out of this initial dataset we kept only the publications having a Document Object Identifier (DOI[8]) as well as the publication year. This cleaning led to selecting 77,080,039 publications authored by 84,818,728 distinct researchers. For our experiments, the authors of the Greek Universities were identified and two data sets were created:

1. the first dataset consists of the academic staff of 19 CS faculties of 17 major Greek universities, i.e., 539 persons.
2. the second dataset consists of all authors with affiliation in the aforementioned 17 Greek universities. This dataset consists of the academic staff of the universities plus every researcher affiliated to any of these universities.

[6] https://academic.microsoft.com.
[7] https://www.aminer.cn/oag2019.
[8] https://doi.org.

**Skyline Dimensions**. The dimensions of $RR$-index are the indicators used by Google Scholar: (a) $Cit$: number of citations to all publications, (b) $Cit$-5: number of citations during the last 5 years to all publications, (c) $h$-index, (d) $h$-index-5: largest number $h$ such that $h$ publications have at least $h$ new citations during the last 5 years, (e) $i10$: number of publications with at least 10 citations, (f) $i10$-5: number of publications that have received at least 10 new citations during the last 5 years.

### 3.1 *RR*-index for Faculty Members

Initially the $RR$-index was calculated at the level of individuals for the 539 members of the Greek CS faculties. The $RR$-index clusters the individuals into 47 groups. Table 1 presents the top-3 ranking levels as derived by the $RR$-index.

**Table 1.** Rainbow Ranking for authors, the 3 top $RR$-levels.

| Author | $RR$-level | $RR$-index | $Cit$ | $h$ | $i10$ | $Cit$-5 | $h$-5 | $i10$-5 |
|---|---|---|---|---|---|---|---|---|
| Nikos Hatziargyriou | 1 | 99.81 | 11009 | 42 | 115 | 5501 | 26 | 73 |
| George Karagiannidis | 1 | 99.81 | 8079 | 49 | 155 | 3856 | 34 | 98 |
| Ioannis Pitas | 1 | 99.81 | 12568 | 55 | 226 | 2774 | 25 | 77 |
| K.A. Antonopoulos | 2 | 98.77 | 1744 | 23 | 38 | 948 | 20 | 26 |
| Minos Garofalakis | 2 | 98.77 | 4824 | 40 | 80 | 871 | 18 | 29 |
| Yannis Manolopoulos | 2 | 98.77 | 5651 | 33 | 104 | 1569 | 17 | 40 |
| Petros Maragos | 2 | 98.77 | 7499 | 42 | 122 | 1500 | 17 | 47 |
| Konstantina Nikita | 2 | 98.77 | 3405 | 29 | 98 | 1246 | 18 | 34 |
| John Psarras | 2 | 98.77 | 3212 | 30 | 89 | 1342 | 19 | 39 |
| Grigorios Tsoumakas | 2 | 98.77 | 3383 | 24 | 38 | 1785 | 18 | 28 |
| Ioannis Vlahavas | 2 | 98.77 | 3468 | 28 | 64 | 1560 | 18 | 32 |
| Aggelos Bletsas | 3 | 96.98 | 4285 | 19 | 34 | 1205 | 13 | 21 |
| Pavlos Georgilakis | 3 | 96.98 | 2367 | 25 | 57 | 1439 | 14 | 29 |
| Aggelos Kiayias | 3 | 96.98 | 5862 | 25 | 33 | 554 | 14 | 16 |
| Stefanos Kollias | 3 | 96.98 | 4783 | 31 | 95 | 1007 | 14 | 19 |
| Aristidis Likas | 3 | 96.98 | 4068 | 34 | 64 | 1350 | 17 | 32 |
| Sotiris Nikoletseas | 3 | 96.98 | 2331 | 26 | 75 | 679 | 13 | 21 |
| Stavros Papathanassiou | 3 | 96.98 | 2620 | 25 | 41 | 1286 | 19 | 27 |
| Ioannis Pratikakis | 3 | 96.98 | 2729 | 28 | 55 | 1153 | 19 | 38 |
| Anastasios Tefas | 3 | 96.98 | 2675 | 26 | 64 | 1178 | 17 | 37 |
| Sergios Theodoridis | 3 | 96.98 | 3535 | 27 | 72 | 1128 | 16 | 30 |
| Yannis Theodoridis | 3 | 96.98 | 3920 | 31 | 65 | 1095 | 17 | 32 |

### 3.2 *RR*-index for CS Faculties

Stepping now to a higher conceptual level and generalizing the previous approach, we compute the $RR$-index of the 19 largest CS faculties, where the previous 539 individuals belong. This generalization is achieved by accumulating all the values of the adopted 6 features of all the faculty members belonging to each faculty. For example, the *Cit* value expresses the total number of citations received by all faculty members of each department. Table 2 shows the top-3 *RR*-index of these 19 CS faculties, which are grouped into 9 Skyline levels.

**Table 2.** Rainbow Ranking for CS faculties

| Fac-Univ | #Staff | *RR*-level | *RR*-index | *Cit* | *h* | *i*10 | *Cit*-5 | *h*-5 | *i*10-5 |
|---|---|---|---|---|---|---|---|---|---|
| ece-ntua | 71 | 1 | 100 | 91077 | 105 | 1822 | 30979 | 60 | 421 |
| di-uoa | 39 | 2 | 92.11 | 46897 | 88 | 899 | 12066 | 40 | 184 |
| inf-auth | 29 | 2 | 92.11 | 53740 | 98 | 877 | 17588 | 50 | 197 |
| csd-uoc | 24 | 3 | 78.95 | 29925 | 77 | 552 | 8188 | 36 | 120 |
| ece-tuc | 24 | 3 | 78.95 | 28858 | 75 | 432 | 8902 | 38 | 115 |
| ee-auth | 28 | 3 | 78.95 | 28842 | 72 | 612 | 10542 | 44 | 191 |

The first ranking level consists of 1 faculty only: the School of Electrical and Computer Engineering of the National Technical University of Athens. On the other hand, we notice that the second level consists of 2 CS faculties, whereas the third level consists of 3 CS faculties. This fact is a proof of concept, i.e. these faculties have the same *RR*-index and belong to the same equivalence class, without any of them dominating the others.

**Table 3.** Rainbow Ranking for Greek Universities

| University | *RR*-level | *RR*-index | *Cit* | *h* | *i*10 | *Cit*-5 | *h*-5 | *i*10-5 |
|---|---|---|---|---|---|---|---|---|
| uoa | 1 | 100 | 7078897 | 841 | 61172 | 3016544 | 552 | 24711 |
| auth | 2 | 91.18 | 3356467 | 548 | 32761 | 1578225 | 368 | 11249 |
| ntua | 2 | 91.18 | 2663388 | 498 | 19926 | 1416386 | 378 | 6886 |
| uoi | 3 | 79.41 | 2156665 | 513 | 20445 | 952141 | 344 | 8387 |
| uoc | 3 | 79.41 | 2125246 | 466 | 24320 | 805642 | 264 | 8769 |

### 3.3 *RR*-index for 17 Greek Universities

Finally, the *RR*-index values for the above 17 Greek universities were calculated using the second dataset. Again, note that each feature value was accumulated over the total number of the academic staff in each university. Notably, these 17 universities are grouped in 12 ranking levels. Table 3 shows the top 3 *RR*-index of these accumulated results for the 6 Skyline features. Table 4 shows the full names of the universities. The grouping created by applying our Rainbow Ranking method was relatively limited. This is due to the fact that the number of universities is small and the feature values vary widely. In turn, the latter fact is due to the different sizes of the universities both in terms of the number of faculties as well as the number of academic staff.

**Table 4.** Greek Universities acronyms and full names

| Acronym | University name | Acronym | University name |
|---|---|---|---|
| auth | Aristotle University of Thessaloniki | tuc | Technical University of Crete |
| ntua | National Technical University of Athens | uoc | University of Crete |
| uoa | National & Kapodistrian University of Athens | uoi | University of Ioannina |

## 4 Conclusions

This article proposes an alternative approach to rank universities by elaborating on the multidimensional Skyline operation, and the Rainbow Ranking methodology. In particular, our method provides ranked sets, in terms of equivalence classes, instead of ranked lists as provided by the traditional university rankings. The method alleviates many of the shortcomings of previous university rankings methods. The obtained results prove the validity of our approach. The proposed methodology can be further elaborated and tested towards richer multidimensional data representing other sets of key-performance indicators, such as more academic or non-academic ones. It can also be expanded across universities around the world and compared to existing rankings.

## References

1. Angelis, L., Bassiliades, N., Manolopoulos, Y.: On the necessity of multiple university rankings. COLLNET J. Scientometrics Inf. Manage. **13**(1), 11–36 (2019). https://doi.org/10.1080/09737766.2018.1550043
2. Bornmann, L., Marx, W.: Thomas theorem in research evaluation. Scientometrics **123**(1), 553–555 (2020). https://doi.org/10.1007/s11192-020-03389-6
3. Börzsönyi, S., Kossmann, D., Stocker, K.: The skyline operator. In: Proceedings 17th IEEE International Conference on Data Engineering (ICDE), pp. 1–20 (2001). https://doi.org/10.1109/ICDE.2001.914855
4. Johnes, J.: University rankings: what do they really show? Scientometrics **115**(1), 585–606 (2018). https://doi.org/10.1007/s11192-018-2666-1
5. Manolopoulos, Y., Katsaros, D.: Metrics and rankings: Myths and fallacies. In: Revised Selected Papers, 18th International Conference on Data Analytics & Management in Data Intensive Domains (DAMDID/RCDL), pp. 265–280. Moscow, Russia (2017). https://doi.org/10.1007/978-3-319-57135-5_19
6. Sidiropoulos, A., Gogoglou, A., Katsaros, D., Manolopoulos, Y.: Gazing at the skyline for star scientists. J. Inform. **10**(3), 789–813 (2016). https://doi.org/10.1016/j.joi.2016.04.009
7. Stoupas, G., Sidiropoulos, A., Gogoglou, A., Katsaros, D., Manolopoulos, Y.: Rainbow ranking: an adaptable, multidimensional ranking method for publication sets. Scientometrics **116**(1), 147–160 (2018). https://doi.org/10.1007/s11192-018-2731-9
8. Thomas, W., Thomas, D.: The Child in America. Knopf, Oxford (1928)
9. Van Raan, A.F.: Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods. Scientometrics **62**(1), 133–143 (2005). https://doi.org/10.1007/s11192-005-0008-6