A Rainbow at the Skyline after the Storm of Indicators for Ranking Scientists

Georgios Stoupas[1], Antonis Sidiropoulos[2], Antonia Gogoglou[3], Dimitrios Katsaros[4], Yannis Manolopoulos[3]

[1] *grgstoupas@gmail.com*, [2] *asidirop@it.teithe.gr*
Alexander Technological Educational Institute of Thessaloniki, Greece

[3] *{agogoglou, manolopo}@csd.auth.gr*
Aristotle University of Thessaloniki, Greece

[4] *dkatsar@inf.uth.gr*
University of Thessaly, Volos, Greece

Introduction

A plethora of bibliometric indices have been proposed to quantify scientific output. To deal with this storm of valuable indicators, the need arises for a classification scheme of scientists according to multiple evaluation metrics. In this work, we expand upon the concept of the skyline operator (Sidiropoulos, Gogoglou, Katsaros, & Manolopoulos, 2016), and introduce a new indicator, namely the Rainbow Ranking. For this study, we collected data from Microsoft Academic Search (MAS), and extracted full citation data starting from year 1950 up to 2015 for computer science scientists.

Rainbow-Ranking (RR-index)

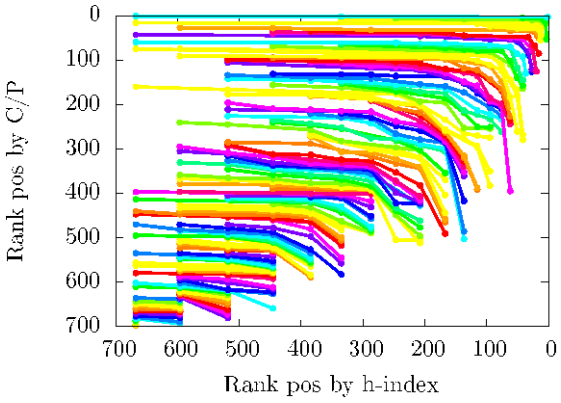

Figure 1. Rainbow Ranking graph

The *Rainbow-Ranking* applies the skyline operator iteratively until all scientists are classified into a skyline level. Figure 1 shows a graphical representation of the skyline levels with two dimensions: citations per publication and the *h*-index. Every point in Figure 1 corresponds to a scientist. Each line connecting the points corresponds to a different skyline level. The x-axis represents ranking positions of each scientist according to their *h*-index, whereas on the y-axis the respective ranking positions according to citations per publication. Since this iterative procedure results into a plot with grouped curves as shown, we have named in the *Rainbow Ranking*.

For the rest of our experiments we select as dimensions of the *skyline operator* the *h*-index (Hirsch, 2005), the *Perfectionism* Index (Sidiropoulos, Katsaros, & Manolopoulos, 2015), and the *A*-index (Jin, Liang, Rousseau, & Egghe, 2007). Given a set of scientists $A=X_1$, the first call of skyline produces the first skyline level. We denote this first set of scientists as set S_1. Next, we compute set $X_2=X_1-S_1$, which contains the scientists that were not classified in the first skyline set S_1. For the set X_2 the skyline operator is applied once more and the result is the second skyline level (S_2). The process continues until all the scientists of the dataset are assigned a value.

To summarize the ranking levels into a single number metric, given a set of scientists A and a set of dimensions *dims*, we define the RR-index of a scientist a based on *dims* as follows:

$$RR(dims) = 100 - 100 * \left(\frac{|A_{above}(a,dims)|}{|A|} + \frac{|A_{tie}(a,dims)|}{2*|A|}\right)$$

$|A|$ is the total number of scientists, $|A_{above}(a,dims)|$ is the number of scientists ranked at higher skyline levels than scientist a based on dimensions *dims*. Level 1 is considered higher than level 2. $|A_{tie}(a,dims)|$ is the number of scientists who are ranked at the same level with scientist a, excluding scientist a. Thus, the following holds for the RR-index: $0 < RR(dims) \leq 100$.

Results

The following table illustrates the *RR*-based ranking. The first skyline level is occupied by scientists who can be grouped into two groups; one group (in italics) are those who have worked in core computer science (e.g., networking, compilers, databases), and the second group are those who have worked in computational biology. In the second skyline level, we mainly encounter core computer scientists (in italics), but also a political scientist and economist (Simon Herbert); the others are computational biologists.

Table 1. First two skyline levels' members

Name	*C*	*P*	*C/P*	*h*-index	*A-index*	*PI*	*RR*	Skyline level
Shenker Scott	38557	473	81.52	90	361.02	12187	99.99	1
Foster Ian	39052	730	53.50	87	365.48	-9320	99.99	1
Ullman Jeffrey	38019	445	85.44	82	394.98	14977	99.99	1
Haussler David	27799	320	86.87	78	305.29	15007	99.99	1
Tibshirani Robert	47661	344	138.55	69	636.06	33447	99.99	1
Miller Webb	54262	532	102.00	42	1272.76	35446	99.99	1
Higgins Desmond	41527	190	218.56	21	1974.43	38419	99.99	1
Lipman David	48638	97	501.42	20	2425.05	47498	99.99	1
Altschul Stephen	46730	78	599.10	19	2453.42	45970	99.99	1
Gish Warren	26065	33	789.85	9	2894.11	25930	99.99	1
Thompson Julie	36441	450	80.98	8	4552.50	32969	99.99	1
Gibson Toby	36329	427	85.08	8	4538.63	33041	99.99	1
Zhang Jinghui	28638	94	304.66	5	5727.20	28218	99.99	1
Garcia-Molina Hector	25743	578	44.54	86	205.88	-9173	99.98	2
Estrin Deborah	34706	446	77.82	85	344.86	11246	99.98	2
Culler David	27360	363	75.37	77	296.17	11267	99.98	2
Simon Herbert	31620	1194	26.48	75	389.40	-46680	99.98	2
Lander Eric	42201	430	98.14	67	612.10	22369	99.98	2
Rivest Ronald	38336	294	130.39	58	615.40	28012	99.98	2
Vapnik Vladimir	31324	123	254.67	49	618.14	30099	99.98	2
Leiserson Charles	23147	155	149.34	36	627.36	20159	99.98	2
Myers Eugene	32210	286	112.62	33	954.42	24950	99.98	2
Cormen Thomas	16707	50	334.14	14	1189.57	16399	99.98	2
Shannon Claude	13554	32	423.56	7	1935.57	13428	99.98	2
Woods Richard	11642	41	283.95	6	1940.33	11468	99.98	2
Schaffer Alejandro	24096	42	573.71	5	4818.20	23936	99.98	2

Summary

This article addresses the following problem: "Given a set of bibliometric indicators, selected in any algorithmic way, can we successively rank scientists into layers based on these indicators, such that the scientists in each layer outperform those of the lower layers according to (at least one) indicator?" We employed the skyline and iteratively applied it to scientists, thus designing the *Rainbow Ranking* indicator. We evaluated it against computer scientists and showed intuitive results.

References

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572.

Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The r- and ar-indices: Complementing the h-index. *Chinese Science Bulletin*, 52(6), 855-863.

Sidiropoulos, A., Gogoglou, A., Katsaros, D., & Manolopoulos, Y. (2016). Gazing at the Skyline for Star Scientists. *Journal of Informetrics, 10*(3), 789 - 813.

Sidiropoulos, A., Katsaros, D., & Manolopoulos, Y. (2015). Ranking and identifying influential scientists versus mass producers by the Perfectionism Index. *Scientometrics, 103*(1), 1-31.