RANKINGS REVISITED

Ignacy Kaliszewski*
Polish Academy of Sciences
Warsaw School of Information Technology

Abstract: *We argue that rankings, as they are commonly used, can be, and perhaps are, misleading and potentially harmful.*

With little extra effort, however, one can gain much more insight into relations among the objects ranked and, in the consequence, gain a better understanding of the ranking. The fundamental notion used to compare and evaluate rankings in our analysis is that of Pareto optymality. General claims are illustrated with the ranking of Polish universities published by Perspektywy monthly in 2016.

This note is based on results that are well known in the areas of multiobjective optimization and multiple-criteria decision analysis. The objective of the note is to point to the shortcomings and potential pitfalls behind the common use and understanding of rankings.

Key words: *rankings, dominance, incomparability, subjectivity.*

O RANKINGACH PONOWNIE

Streszczenie: *W pracy przedstawiam argumenty za tym, że rankingi funkcjonujące obecnie w gospodarce i społeczeństwie mogą być, i prawdopodobnie są, mylące i potencjalnie szkodliwe.*

Stwierdzam, że przy niewielkim dodatkowym wysiłku możemy uzyskać znacznie głębszy wgląd we wzajemne relacje pomiędzy obiektami podlegającymi rankingom i w konsekwencji mieć bogatszy ogląd rzeczywistości. Centralnym pojęciem w przedstawionych tu rozważaniach jest optymalność w sensie Pareto. Dla zilustrowania istotności tak ogólnego stwierdzenia posłużę się danymi po-

* Ignacy Kaliszewski, Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warszawa, e-mail: ignacy.kaliszewski@ibspan.waw.pl oraz Warsaw School of Information Technology, ul. Newelska 6, 01-447 Warszawa.

Ministry of Science and Higher Education Republic of Poland

English-language editing of that article was financed under Agreement No. 645/P-DUN/2018 with funds from the Ministry of Science and Higher Education allocated to the popularization of science.

 DOI: 10.7206/DEC.1733-0092.115

chodzącymi z rankingu polskich szkół wyższych za rok 2016, opublikowanego przez miesięcznik „Perspektywy".

Praca bazuje na znanych od dawna wynikach, w szczególności w obszarze optymalizacji wielokryterialnej i wielokryterialnego podejmowania decyzji. Jest także wyrazem moich refleksji i jednocześnie niepokoju odnośnie sposobów wykorzystywania rankingów w życiu codziennym, wskazując przy tym ich ułomności i potencjalne pułapki.

Słowa kluczowe: *rankingi, dominacja, nieporównywalność, subiektywność.*

1. Introduction

Rankings are ubiquitous. Their purpose is to turn the multi-aspect universe we live in, where objects are related by dominance and incomparability relations, into a purely linear one.

The notion of a ranking has two common connotations. Possibly the most popular and widespread meaning is that of a ranking as a given sequence of objects arranged according to values on a number of different criteria. The second meaning, a less common one, understands a ranking as arranging objects into a sequence. If the arranging objects is done according to one criterion – for instance, time in a track and field sprint or distance in a discus throw – the problem is trivial. The problem becomes nontrivial when we have two or more criteria. In the latter case, a ranking cannot be produced without subjective decisions of its author.

If we accept rankings without understanding the process used to produce them, we lose control of the objective and the subjective aspects in the underlying algorithm generating a ranking. Without such understanding one has no option but to believe that the resulting ranking adequately represents relations between objects which are, in general, complex and difficult to see. The second meaning still raises questions about the mechanism used to produce the ranking.

In this note, we attempt to distinguish between objective and subjective aspects of a ranking. Considering the objective aspect of a ranking to be a common denominator, we show how using a simple extra effort each individual can exercise control over ranking's subjective aspects. We also address the issue of ranking creation mechanisms.

2. The problem

Suppose we have a set of objects and each object is described on several numerical criteria. Assuming that all criteria are of the more, the better type, rankings attempt to order objects from the best to the worst using numerical values of objects on the criteria. It is often done by summing up values across all the criteria and ordering objects according to the decreasing values of the scores (below we discuss an alternative way of doing this). If criteria have different importance, they can be weighted by numbers, called weights, that reflect their relative importance. This is, for instance, how *Perspektywy 2016 rankings* of the Polish institutions of higher education were constructed (*Perspektywy 2016* ranking 2016).

We claim that there is no such thing as an objective ranking. Even if an influential institution, like a grant giving agency, adopts a ranking as its official instrument, this still reflects just that institution's specific perspective and nothing more. The only objective element in rankings are data. (Naturally, we assume that the data are accurate, so nobody can contest them). Subjective in any ranking are the selection of criteria and the selection of weights. Therefore, rankings published by ranking providers, who always try to market themselves as objective bodies, are nothing but subjective[1]. Moreover, as they follow the principle "one ranking for all", they fall into a group of analytics recently referred to as the Procrustean method (Taleb 2016). In brief, the purpose of a ranking is to compress or stretch reality to fit the preconceived ideas.

If rankings are, indeed, subjective, what would be wrong about it? As we will argue below, the popular understanding of rankings as a universal evaluation device, skews the reality and provides a perspective that is too narrow for potential users.

3. The outline

We will illustrate our argument with a running example (the corresponding fragments, as e.g. this one, are set with an alternative font) of a data set for 50 Polish universities – they are the objects to be ranked – taken from the 2016 edition of the annual rankings published by the *Perspektywy* monthly. Criteria in that edition were: prestige, innovation, scientific potential, scientific merit, education infrastructure, international cooperation. All these criteria are of the more, the better type.

[1] A fitting metaphor would be a restaurant in which the cook offers just one best (he believes) dish, even though from the ingredients available in the kitchen he could prepare a range of courses that would meet individual preferences of his guests.

DOI: 10.7206/DEC.1733-0092.115

For the sake of notational simplicity, we will identify universities by the numbers assigned to them in the *Perspektywy 2016* ranking.

Using an electronic companion to this work (Rankings 2018) built with the *Perspektywy 2016* data, one may check all computations presented in this note, run other analyses, and derive alternative rankings.

Data for a ranking consists of a set of objects, each represented by a vector of k values on the corresponding k criteria.

An object $\overline{x}$, an element of set of objects X, is said to be *Pareto optimal*, if there is no other object in this set with all criteria values equal or greater and at least one criterion value strictly greater. If such an object exists, let's call it x, then $\overline{x}$ is said to be *dominated* by x and x is said to be *dominating* $\overline{x}$. Two objects such that none of them dominates the other, are said to be *incomparable*.

In *Perspektywy 2016* ranking, there are ten Pareto optimal universities, namely #1 to #6, #23, #24, #33, #50.

It is interesting to note that the object ranked in the very last, the 50th position, is, in fact, Pareto optimal.

Any function defined over k arguments ranks objects with respect to k criteria. In practice, all rankings make use of the linear weighted scoring function

$$w_1 y_1(x) + \ldots + w_k y_k(x); \tag{1}$$

where $y_l(x)$ is the value of l-th criterion for object x, and w_l is the (positive) value of l-th weight, $l = 1, \ldots, k$.

Since a Pareto optimal object is not dominated, it seems to be a natural leader in the set of all objects. It is an element in the Pareto equlibrium – any other Pareto optimal object has the value of at least one criterion smaller and the value of at least one criterion larger. It shares this distinctive position with other leaders - other Pareto optimal objects. For any leader, we can ask what set of weights in scoring function (1) would put it in the first position in the ranking[2]. A set of weights will satisfy that condition if the value of the scoring function for this leader is greater or equal than for any other Pareto optimal object. Such weights, if they exist, can be found by solving a linear programming problem.

Given the scoring function (1) it can happen that for some leaders there is no ranking, i.e. there is no set of weights, in which that leader is ranked highest.

[2] If there were no such ranking and in consequence the leader could be placed at most, say, at the second position in any ranking, this would contradict its leadership.

 DOI: 10.7206/DEC.1733-0092.115

In *Perspektywy 2016* data set there are five leaders like that. For leaders: #3, #4, #6, #24 there is no scoring function (1) that will put them in the first position. This can be proved in the following way. Suppose leader #i is checked for the existence of weights that would put him in the first place under the scoringx function (1). To have it happen, the following system of conditions has to be consistent:

$$w_1\, y_1(\#i) + \ldots + w_k\, y_k(\#i) \geq$$
$$w_1\, y_1(\#j) + \ldots + w_k\, y_k(\#j) \text{ for all leaders } j;\ j \neq i; \qquad (2)$$
$$w_l > 0;\ l = 1, \ldots, k.$$

In the above test, all dominated (i.e. not a leader) objects are neglected since they produce redundant inequalities. And for leaders: #3, #4, #6, #24 and #50 systems of inequalities (2) are inconsistent.

For leaders #1, #2, #5, #23 and #33, weights that put them in the first place in the ranking are any weights for which the respective systems of inequalities (2) are consistent. Theoretically, such weights can be unique but typically such set of weights is infinite.

For example, for leader #1 of *Perspektywy 2016* ranking, an example of a ranking in which it takes the first position is defined by weights w_l: prestige – 27, innovation – 33, scientific potential – 9, scientific merit – 9, education infrastructure – 9, international cooperation – 13.

For leader #2, one ranking in which it takes the first position is defined by weights wl: prestige – 30, innovation – 3, scientific potential – 18, scientific merit – 25, education infrastructure – 15, international cooperation – 9.

Since a Pareto dominance relation induces a partial order which is typically not linear, a useful way to represent relations between objects are Hasse diagrams, where dominated objects are placed below dominating ones. Hasse diagrams are a convenient way of depicting dominance relations when the number of criteria is greater or equal to 4, i.e. when standard graphical representations of objects by their criteria values, that may be useful for $k = 2$ or $k = 3$, are of little, if any, use.

Just by looking at the Hasse diagram of objects and the corresponding ranking it is very clear that rankings alone, when provided as the only solution, conceal and distort the complex structure of dominance relations between objects. But such an information, to the Author's best knowledge, never accompanies rankings. If present, it could cause at least a reflection on the merits on rankings, purported to be fair representations of such structures.

 DOI: 10.7206/DEC.1733-0092.115

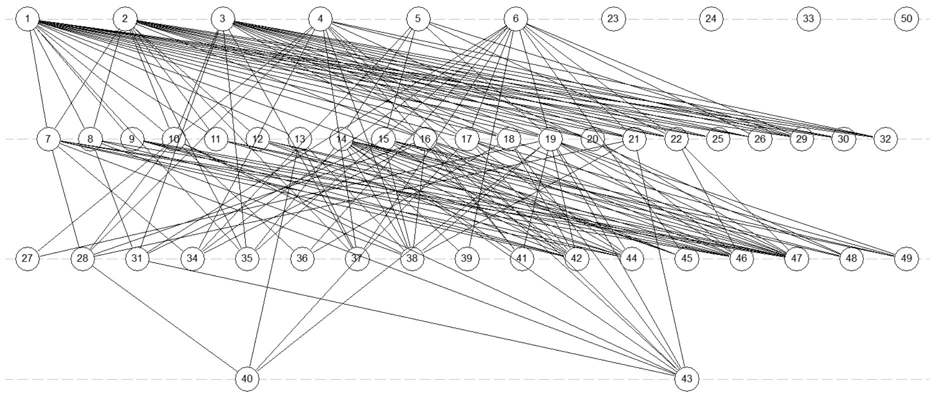

***Figure 1*. The Hasse diagram for the Perspektywy 2016 data set**

Figure 1 shows Hasse diagram[3] for the *Perspektywy 2016* data set – a true and undistorted picture of the 50 Polish university structure under six criteria adopted for the *Perspektywy 2016* ranking.

It is natural to ask whether there is a simple way to construct a ranking that would put each leader in the first position? The answer is affirmative. Moreover, as shown below, such a construction is simple and intuitive.

Denote the highest value over all objects in criterion l by $\hat{y}_l$ with l ranging from 1 to k, and $\hat{y} = \{\hat{y}_1, ..., \hat{y}_k\}$. If there were an object with all criteria values equal to the corresponding components of $\hat{y}$, this object would be *ideal*. It would dominate all other objects. In general, this is rather uncommon.

Given object x, $\hat{y}_l - y_l(x)$ measures the distance from $\hat{y}$ to $y(x)$ along the l-th criterion. For all criteria, this distance can be thought of "the smaller, the better" type of a criterion. The function

$$\max_l \lambda_l \, (\hat{y}_l - y_l(x)), \tag{3}$$

where λ_l is the (positive) value of l-th weight, yields the biggest of those k distances. Here we use λ_l to denote weights, to avoid mixing them up with weights w_l in scoring function (1). Scoring function (3) is an option to scoring function (1). Table 2 lists the most significant properties of the two scoring functions. The most important is property 2, which states that for EVERY Pareto optimal object there are weights λ_l, $l = 1, ..., k$, for which this object is the minimizer of scoring function (3). In case of multiple minimizers of (3), it is easy to check which one of them is Pareto optimal.

[3] The diagram has been produced with freeware software DART (Decision Analysis via Ranking Techniques).

 DOI: 10.7206/DEC.1733-0092.115

In terms of rankings, property 2 of scoring function (3) means that NO leader is a priori excluded from the chance to be ranked first (with some weights). Scoring function (1) does not have this property. As we have noted before, some leaders will not be ranked first under any instance of scoring function (1).

Moreover, the weights for which a selected Pareto optimal object $\overline{x}$ minimizes scoring function (3) are easily calculable[4]:

$$\lambda_l = \frac{1}{\hat{y}_l - y_l(\overline{x})}, \quad l = 1, ..., \mathrm{k} \tag{4}$$

To avoid zeros in the denominator, which will appear when $y_l(\overline{x})$ is equal to $\hat{y}_l$, we have to use values y_l^* that are slightly bigger than $\hat{y}_l$, i.e. $y_l^* = \hat{y}_l + \varepsilon$, $\varepsilon > 0$, $l = 1, ..., k$. With ε small enough, this does not affect the relations between objects and thus does not affect their rankings.

Leader #3 minimizes scoring function (3) with weights λ_l (we arbitrarily assume $\varepsilon = 1$):

$$\frac{1}{101-62.95}, \frac{1}{101-46.84}, \frac{1}{101-92.36}, \frac{1}{101-81.70}, \frac{1}{101-79.53}, \frac{1}{101-58.98}.$$

Likewise, leader #50 minimizes scoring function (3) for weights λ_l:

$$\frac{1}{101-18.59}, \frac{1}{101-1.08}, \frac{1}{101-54.06}, \frac{1}{101-41.78}, \frac{1}{101-35.19}, \frac{1}{101-63.78}.$$

Formula (4) yields a ranking with scoring function (3) for any object.

Both scoring functions have the property that a dominated object is never placed in a ranking above the dominating one.

With weights λ_l for leader #50 defined above, the resulting ranking is given in Table 1. The second row in the Table identifies universities by their positions in the *Perspektywy 2016* ranking.

4. Have your own rankings!

Ranking agencies provide a great service to the society by collecting data and making them available to everyone interested. This is good. However, in the next step they present THEIR rankings for the collected data. Given pervasive social inertia and widespread lack of proper understanding, aggressively marketed and well-publicized rankings tend to be regarded and worse yet, used as objective, true

[4] This formula follows from the constructive proof of the property 2 of scoring function (3) (Kaliszewski et al. 2016).

 DOI: 10.7206/DEC.1733-0092.115

and unique representations of the relations between objects ranked. Yet, as we have shown above, such a ranking is merely an instance of many others.

Of a true service to society would be an interactive ranking system, where one could input his or her proprietary weights and generate HIS or HER subjective ranking. As long as such services are not commonly available, they can be easily forged with a spreadsheet like the one provided by the electronic companion to this note. Using the spreadsheet, we can do more comprehensive analyses like the one that was presented in the note.

Table 1
The ranking with scoring function (3) and weights defined by leader #50

1	2	3	4	5	6	7	8	9	10	11
#50	#4	#23	#3	#5	#24	#1	#2	#7	#9	#12
12	13	14	15	16	17	18	19	20	21	22
#11	#14	#32	#18	#16	#15	#6	#8	#17	#19	#44
23	24	25	26	27	28	29	30	31	32	33
#42	#22	#34	#20	#46	#33	#47	#30	#41	#31	#27
34	35	36	37	38	39	40	41	42	43	44
#25	#49	#21	#29	#10	#43	#26	#35	#13	#28	#39
45	46	47	48	49	50					
#37	#38	#45	#40	#48	#36					

Rankings with scoring function (3) are more fair than that with scoring function (1); with (3) every Pareto optimal object has a ranking in which it takes the first position. Since every Pareto optimal object is a top element under the dominance relation, it seems intuitively right that each such element may end up as the highest ranked.

Moreover, the minimizers of scoring function (3) tend to have more balanced criteria values (since the minimizers minimize $\max_l \lambda_l (y_l^* - y_l(x))$ over all objects). This function admits no substitution between criteria values. In contrast, criteria values in maximizers of scoring function (1) can be much more scattered, since this function admits substitutions between criteria. We should ask, especially with the university rankings, what sense does it make to admit substitutions between, say, the number of sold licenses and a measure of academic performance, as was the case with the *Perspektywy 2016* ranking.

With respect to all issues raised here, the decision about which scoring function and which weights to use should be left to an individual. Everyone should be able to judge according to their individual, subjective preferences, convictions and tastes.

 DOI: 10.7206/DEC.1733-0092.115

Table 2
Properties of two scoring functions considered

	Function	
	(1) maximized	(2) minimized
1. Is the function extremizer Pareto optimal?	Yes	Yes, if it is unique; otherwise, one of them is Pareto optimal.
2. Is every Pareto optimal object an extremizer?	No	Yes
3. Calculations of weights for which a Pareto optimal object is the function extremizer.	Requires: – Knowledge of all Pareto optimal objects, – Solving a linear programming problem.	– Independent of other Pareto optimal objects, – Involves only elementary arithmetic calculations.

5. Concluding remarks

This note is the result of the Author's concern that some simple facts about rankings remain unknown to broad audiences. Scoring function (3), called in the domain of multiobjective optimization and multiple criteria decision making after the famous Russian mathematician, the Chebyshev function (or the Chebyshev distance)[5], and its properties have been known for a long time (Kaliszewski 1994,2006, Miettinen 1999, Ehrgott 2005, Kaliszewski et al. 2016, just to mention general monographs, relevant papers are numerous). However, to the best knowledge of the Author, outside the academic world Chebyshev functions are never used to rank objects. But why should we eliminate, a priori, any Pareto optimal objects? Why should we disqualify them as winners in all possible rankings? As long as rankings are regarded as non-harmful curiosities, no damage is done. Once, however, rankings are used to make decisions that have social and economic consequences, one should be aware of the foregone opportunities. And it is not so that missed opportunities are costless. Everyone, both individuals and societies, pay possibly heavy tolls and we don't realize that we do (Kaliszewski, Samuelson 2016).

A good illustration of this claim is the case of the Via Baltica expressway. As shown in the paper by Jastrzębski and Kaliszewski (2011), a consulting company that was contracted to analyze variants of possible expressway routes on the territory of Poland, failed to determine all Pareto optimal variants just because it used the scoring function (1). It is difficult, or perhaps impossible, to know to what extent this flaw contributed to economic and environmental damages (the expressway passes nearby the famous Rospuda river valley, a protected natural preserve). One

[5] However, in the number theory the term „Chebyshev function" denotes a different construct.

conclusion we can make with certainty, however, is that the analysis behind the final recommendation was incomplete. We will now have to live with the consequences of a deficient analysis no matter what these consequences will turn out to be.

The electronic companion

An electronic companion to this note (http://www.ibspan.waw.pl/~kaliszew) contains an Excel sheet in which one can generate his or her own rankings with the *Perspektywy 2016* data set, using scoring function (1) or (3) and using their own preferences expressed as a choice of weights over the criteria used in the ranking.

References

[1] *Perspektywy 2016 ranking* (2016). http://www.perspektywy.pl/RSW2016/ranking-uczelni-akademickich, as by October 2018 .

[2] Ehrgott, M. (2005). *Multicriteria Optimization*. Springer.

[3] Jastrzębski, P., Kaliszewski, I. (2011). Computer-aided decision making {the multiple criteria approach; the case of planning Via Baltica expressway} (in Polish). *Contemporary Management Problems, 1*, 63-70. Warsaw School of Information Technology.

[4] Kaliszewski, I. (1994). *Quantitative Pareto Analysis by Cone Separation Technique*. Kluwer Academic Publishers, Boston.

[5] Kaliszewski, I. (2006). *Soft Computing for Complex Multiple Criteria Decision Making*. Springer, New York.

[6] Kaliszewski, I., Miroforidis, J., Podkopaev, D. (2016). *Multiple Criteria Decision Making by Multiobjective Optimization - A Toolbox*. Springer.

[7] Kaliszewski, I., Samuelson, D.A. (2016). Models that are never wrong. *OR/MS Today, June 2017*, 24-27.

[8] Miettinen, K.M. (1999). *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Boston.

[9] *Rankings* (2018). An electronic companion. http://www.ibspan.waw.pl/_kaliszew/Rankingi.

[10] Simon, H. (1977). *The New Science of Management Decision*. Prentice Hall.

 DOI: 10.7206/DEC.1733-0092.115