Rankings analysis with the Optimized Pareto method.

Carlos Andres Delgado-Saavedra[1], Angel García-Baños[2], Victor Andrés Bucheli-Guerrero[3]
[1,2,3]Universidad del Valle, Colombia

Received: March 20, 2020
Approved: April 20, 2020

Abstract— Rankings compare the performance of organizations. In many cases, rankings provide a good assessment of successful organizations. However, rankings often generate controversy and debate since they support the making decisions. A ranking is a weighted linear combination of indicators, and the weights assigned to each of the indicators can lead to different rank orders. In most cases, rankings are used as a tool to support making decisions, such as resource allocation; therefore, these decisions can be affected by the assignment of such weights. In this article, we analyze the behavior of a ranking and the weights; simulations are used to calculate the change in the order of the equally weighted ranking and of the randomly weighted ranking. In this regard, we present a discussion and ranking design alternatives.

Keywords: Multi-objective optimization, linearization, Pareto, ranking.

*Author of correspondence
Email address: cardel87@gmail.com (Carlos Andres Delgado Saavedra)
Peer review is the responsibility of the University of Santander.

How to cite: C. A. Delgado Saavedra, A. García-Baños y V. A. Buchel-Guerreroi, “Rankings analysis with the Optimized Pareto method”, Aibi journal of research, management and engineering, vol. 8, no. 2, pp. 92-97, 2020.

I. INTRODUCTION

Rankings are used in many fields. For example, the sports rank-ings, such as the FIFA rankings[1]; social rankings such as that of the richest people in the world[2]. Currently, in organizations such as universities, rankings are fundamental for their prestige at an international level, the most prominent are the Times Higher Education, Academic Ranking World Universities Shanghai and QS World Ranking from which the universities of the United States and Europe are better rated[3]. A ranking is composed of a set of indicators, which are collapsed into a single measure that generates a sorted list. More precisely, the goal is to maximize the performance indica-tors simultaneously. This can be achieved by calculating the average of all the indicators or assigning weights to each of them. This results in a single indicator that synthesizes the information and needs to be maximized[4]

Rankings make it possible to compare organizations and deter-mine which ones are best according to a list sorted by rank. The ranking order can be changed by both the chosen indicators and the weighting scheme used. Some decisions, such as allocating resources, investing, choosing a place for an event, or selecting a house to buy, can be affected by a ranking. However, what are the implications of the ranking method in the decision-making process? In this study, we provide a first look at this question, which is currently controversy[5], [6].

We analyze the issue of using different weights or indicators in the ranking methods. Simulations were performed to understand how a ranking works internally because indicators can significantly change the ranking order.

This paper is organized as follows: the next section describes the rankings. The third section shows the results of the simula-tion and Pareto optimization, in addition to the tests performed with random sets of indicators and random weights for each indicator, comparing the incorrect rankings against the rankings proposed.

II. RANKINGS

Rankings stem from the need of knowing which organization is best to make decisions based on the ranking order. Currently, there are rankings in almost any field, as is the case for universities and the quality of life of countries, among others. Rankings consist of a group of indicators; for example, the ranking indicators for quality of life in different countries include average lifespan and average income, among others. To develop a ranking, a linear weighting is often performed, namely, each indicator is assigned a weight, and each indicator is normalized[7]. Finally, the ranking is the result of the sum of the product of these indicators and their respective weights, which yields a synthetic indicator that incorporates all the characteristics of the measurement.

However, according to some studies, there are problems when generating rankings. Wilcoxon[8] argues that weighted-rankings could get very different results with a small variation in the weigths. Altbach indicated the rankings may be affected by previous measurements due to the prestigie of the institution. Also, some indicator have noise due to practices as article self-citing, because these cannot be performed properly[9]. For these reasons, other studies seek alternatives to build rankings of universities; the Leiden Ranking is one example[10]. Hence, different measuring methods have been proposed, including a model based on the partial sum of quadratic differences[11], multidimensional analysis[12] and models based on fuzzy logic[13], [14].

Other studies reveal the negative impact of rankings on the innovation of organizations. For example, business schools make the generation of new structures or lines of thought difficult because a new area can correspond to a low rating in the indicators[15] or get a unfair adventange with the publication of poor quality papers in predatory journals[16]. Other studies show the impact of rankings on the perception of tourism[17] , security[18], [19], life quality[20] and proposing measuring alternatives in countries in the worst measures. In addition, there are studies that show how rankings are affected by the quality of the data used because there have been some problems with transparency and indicators with unexpected behaviors that add noise to the measurement[21], [22].

Recent publications about rankings explored the problem of the grow of data about the indicators[23], Cantù Ortiz argues that data about publications doubles every 9-15 years, then it increases the complexity of analysis of the indicators[24], the increment of the noise in the rankings data[25] or quality of the data[26]. Authors as Cousijn[27] and Singh[28], had concluded is necessary to create tools to the correct data collection of the measurement factors.

Another question is how the organizations can respond to multiple rankings; Pollock argues the institutions can be surrounded to measures of different rankings. These rankings can use their own methodologies and indicators, this an important challenge to organizations transformation with the goal of improve their positioning[29]. The universities are the institutions with the major challenge because the rankings are the most common strategy to categorize their quality and some rankings ignore some factors about researching, teaching and pedagogy. It is a big problem because the universities must respond to multiple rankings [30].

III. MULTIOBJETIVE OPTIMIZATION

In some optimization cases, there are several objectives to maximize, which may be mutually exclusive. For example, in selecting the best organization by assessing two parameters, one parameter may be the size of the organization and the other its impact on society. Each objective is important, but they are not directly comparable since, for example, small organizations can have a significant impact on society, and vice versa.

In these cases, multi-objective optimization techniques are used[31], [32]. In the literature exists different techniques such as, weigthed sum[33], genetic algorithms[34] or fuzzy models[35].

Our approach is the Pareto Method, that are based on the concept of dominance and on sets of optimals[36], [37]. These concepts are defined as follows:

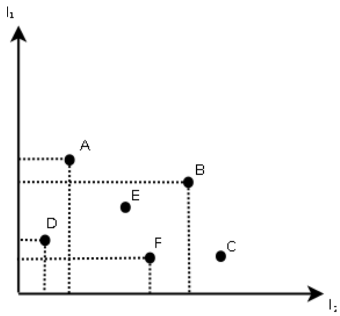
Dominance definition: There are two organizations x_a and x_b. Organization x_a dominates organization x_b if and only if $\exists i \in I \therefore F(x_a,i) > F(x_b,i)$ where P is the set of Pareto optimals, U is the set of organizations, I the set of indicators, NI is the number of indicators, NU is the number of organizations, NP is the number of organizations in the set of Pareto optimals, and $F(x_a,i_j)$ is the indicator value $i_j \in I, 1 \leq j \leq \ni$ for organization $x_i \in U, 1 \leq i \leq NU$

Pareto optimal definition: when one organization is not dominated by any other, then it is said that the organization belongs to P.

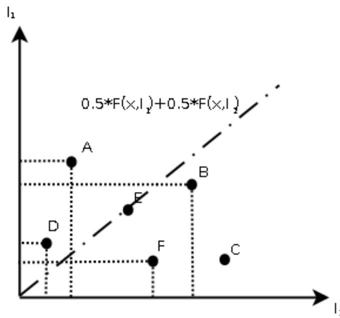
Pareto frontier definition: If an organization $x_k \in P, 1 \leq k \leq NP$, then it is part of the valid solutions of the multi-objective problem.

Figure 1 shows an example of multi-objective optimization. There are 6 organizations and two measurement indicators, F1 and F2. Each of the organizations of the example has different indicator values. figure 1a shows that A dominates D because it is better in both measurement indicators. B dominates D, E and F since it exceeds them in both measurement

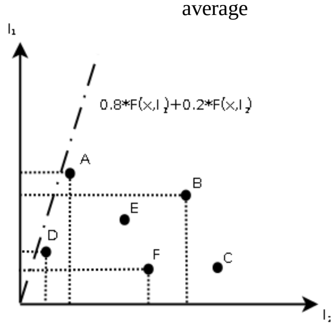
indicators. C does not dominate any other point. Consequently, points A, B, and C belong to the Pareto frontier because they are not dominated by any other point. A comparison of the ranking classification obtained by varying the weight of the indicators is shown in figure 1b and 1c. It is noted that the weighting can change the classification produced by the ranking. In figure 1b, points B, C and F are in the top places, whereas in figure 1c, the top places are occupied by A, E and B, which indicates that changing the weights of the indicators can produce different classifications.



a) A, B and C form a Pareto frontier



b) Ranking via a linear weighted average



c) Ranking via linear weighting with a higher weight given to one of the indicators

Figure 1: Example of multi-objective Pareto optimization and ranking by weighting.
Source: Own elaboration.

IV. EXPERIMENTS

In this section, we perform computer simulations of the behavior of rankings to understand the impact of changing the weights of the indicators. This means that the change in the order of the rankings classified linearly and by Pareto optimization is observed.

For the simulation, we assume that organizations are evaluated by indicators randomly that are generated and evenly distributed over a range of values (each indicator has a different range). Below are the experiments we performed:

The indicators are averaged to obtain a single summary indicator, which is sorted from highest to lowest (namely, a ranking is made). Subsequently, the differences between this ranking and that obtained via Pareto optimization are measured.

Random weights are generated under a uniform distribution, ensuring that the sum equals one. Then, a summary indicator of the sum of the weighted indicators is obtained. This summary indicator is sorted from highest to lowest; consequently, rankings are obtained for each of the possible weightings. Finally, the differences in the positions of each ranking obtained are analyzed.

a. Simulation parameters

The number of organizations (NU), number of indicators (NI) and weights (V) are parameterized in the experiments. The experiments are parameterized as follows:

Between 2 and 10 organizations (NU)
Between 2 and 20 indicators (NI)
Weights between 0 and 1.

b. *Experiment 1: Simulation of organizations ranking and Pareto frontier.*

Generating the set of Pareto optimals P and a sorted list R(m) of the linear weighted ranking.

A set of organizations U of size UN is generated, each one with NI indicators. The values taken by each indicator are floating numbers between 0 and U={X_1,X_2,...,X_NU}such that X_j={X_j1,X_j2,...,X_jNI}, where $1\leq j\leq \ni$ and $1\leq k\leq NU$

The set P of the Pareto optimals of U is calculated. For each organization U, the indicators are linearized using a set of weights α_k. The set of weights is the same for all the organizations. More specifically, for each organization, the following calculation is performed:

f(x_j)=∑_(i=1)^NU▒f(x_(j,) I_i) ,such that ∑α_k=1, for $1\leq j\leq \ni$ and $1\leq k\leq NU$ (1)

In the first experiment, all weights are equal, α_k=1/NI,∀k because this is how many rankings are performed. In the second experiment, the weights are random.

4All points U are ranked from highest to lowest using f(x_j). This produces the sorted list of organizations, also called the ranking R(m) of size NR with $0\leq NR\leq NU$.

If R(n)∈P∀n is $1\leq n\leq NR$, then indicator f(x_n) coincides with the first S cases. S is then added to the number of correct answers obtained.

The previous process is repeated, but iterating from the bottom of the ranking and counting the items M that do not belong to P. If R(n)∉P∀nis ≥n≥M with M≤1, then NU-M+1 will also be added to the number of correct answers. Since these elements exist at the bottom of the ranking in addition to in the Pareto set, these are also considered correct ranking answers. It should be noted that this method of measuring the quality of the ranking accepts organizations in any order if these are consecutive and at the top of the list. The same happens with the bottom of the list.

False negatives are calculated as all the points that are in P but are not correct answers in the ranking. This means that even though these points are on the Pareto frontier, the ranking does not place them correctly.

False positives are calculated as all other ranked elements. This means that the ranking says that these are good, but the Pareto frontier does not.

c. Experiment 2: Simulation of organization ranking with random weights.

Random weights are used in this experiment. The difference between the highest and lowest position that results from modifying the weights is calculated for each organization. Therefore, the greater the difference is, the greater the sensitivity to the weights.

The process is repeated by calculating the average number of correct answers, false positives, false negatives, and standard deviation of the correct answers.

V. TEST AND RESULTS

The implementation of the algorithms in the previous section was performed in Ruby. Unit tests were performed to verify the implementation using the Behavior Driven Development (BDD)[38] methodology implemented with the Cucumber tool[39]. This makes it possible to generate test scenarios in natural language.

Figure 2 shows an actual test for which a ranking of items sorted from highest to lowest, from 1 to 10, was created. Assuming the Pareto frontier is P={1,2,6,8}, this shows the number of correct answers of the ranking (items that are very high in the ranking and that are part of the Pareto frontier, and also items that are low in the ranking and are not part of the Pareto frontier); false positives (items that are high in the ranking but are not part of the Pareto frontier); and false negatives (items that appear low in the ranking but are part of the Pareto frontier). It can be observed that items 1, 2, 9 and 10 are correct answers because 1 and 2 belong to the Pareto frontier and are highly ranked. Similarly, items 9 and 10 are correct answers because they are not on the frontier. Items 6 and 8 are false negatives because they are on the Pareto frontier and are not in the top positions of the ranking. Finally, items 3, 4, 5 and 7 are false positives because they are in intermediate positions of the ranking but are not on the Pareto frontier.

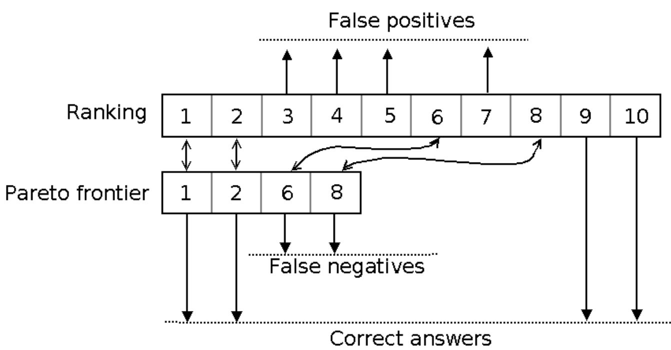


Figure 2. Example of false positives and false negatives in the theoretical measure.
Source: The authors.

The averages of 300 repetitions are taken for each experiment to determine the percent of correct answers, percent of false positives and percent of false negatives for all the possible combinations from 2 to 10 indicators and 2 to 100 organizations.

For simulation 1, a box plot is used to represent the results. The average of the data is in red, and the deviations are indicated in blue. The green circles indicate mild outliers, and the green crosses indicate extreme outliers that represent isolated cases.

Simulation 2 produces a graph of false positives versus correct answers to show the performance of the correct answers based on the variation of the indicators.

a. Results experiment 1

Regarding the correct answers, figure 3 shows that if the indicators increases, there is a tendency toward 100% of correct answers, i.e., perfect agreement between the Pareto solution and ranking with weights. Additionally, the percentage of correct answers is low (approximately 40% in the measurement) with fewer indicators (less than 10 in the experiments), regardless of the number of organizations

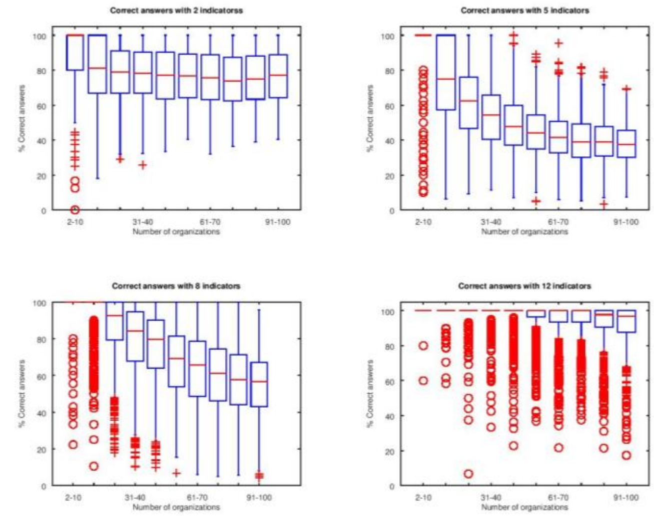


Figure 3: Correct answers per number of indicators.
Source: The authors.

b. Results experiment 2

An analysis of the false positives and false negatives found in the experiments is performed after comparing the Pareto optimal method to the average position obtained by varying the weights in a linear ranking.

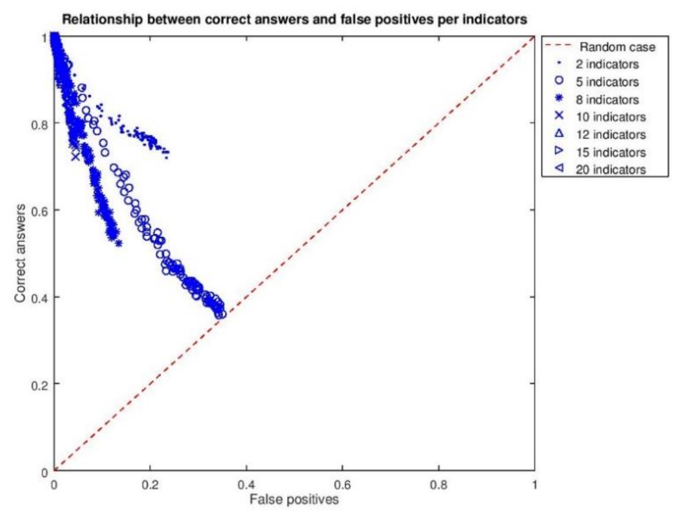


Figure 4. Relationship between correct answers and false positives.
Source: The authors.

Figure 4 shows that as the number of indicators increases, the system tends toward the optimal classification, which consists of 100% correct answers and 0% false positives. When there are few indicators (between 2 and 5), a different behavior is observed because if the number of indicators is increased, the performance of the system approaches the random case.

There are rankings in many fields, and not only in the classification of organizations. For example, in the admission processes of universities, there are several indicators (such as entrance exams in subject areas such as mathematics, physics, and English) that are added using weights that depend on the importance of the subject area for each specific degree. There is also a ranking for the indicators of the research groups of Colciencias[40], among others.

Also, the weights of the linear rankings can deliver completely different results compared with the Pareto method in cases in which there are few indicators. The most important characteristic to examine is the number of indicators because these vary substantially among different rankings. In our measurements, it is observed that the Pareto method and weighted ranking yield similar results in the cases with many measurement indicators because slight variations in the weights do not affect the ranking greatly. The trend that is generally found is shown in figure 5, which shows that as the number of indicators increases, the maximum percentage of correct answers is achieved more rapidly.

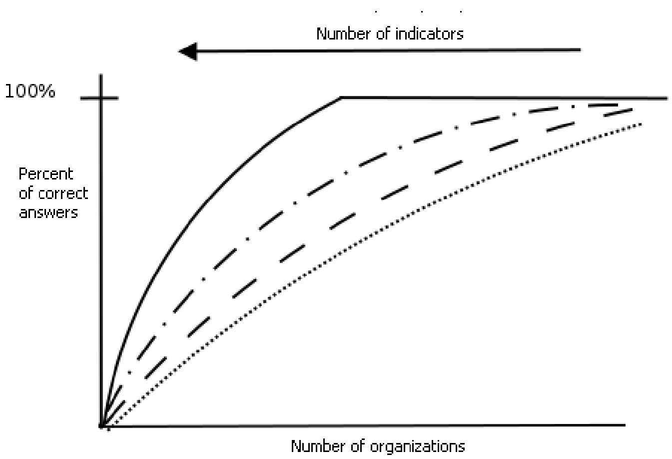


Figure 5. Trend of the percentage of correct answers in relation to the increase in the number of indicators.
Source: The authors.

The experiments show that the ranking method has a strong dependence on the number of indicators and the relationship between false positives and correct answers presented in figure 5 show that as the number of indicators increases, the Pareto method behaves as a linear ranking; consequently, Pareto optimization with few indicators is not a good strategy for classification because its behavior is like that of the random case.

VI. DISCUSSION

There are rankings in many fields, and not only in the classification of organizations. For example, in the admission processes of universities, there are several indicators (such as entrance exams in subject areas such as mathematics, physics, and English) that are added using weights that depend on the importance of the subject area for each specific degree. There is also a ranking for the indicators of the research groups of Ministerio de Ciencias[40], among others.

Figure 3 shows that the weights of the linear rankings can deliver completely different results compared with the Pareto method in cases in which there are few indicators. The most important characteristic to examine is the number of indicators because these vary substantially among different rankings. In our measurements, it is observed that the Pareto method and weighted ranking yield similar results in the cases with many measurement indicators because slight variations in the weights do not affect the ranking greatly. The trend that is generally found is shown in figure 5, which shows that as the number of indicators increases, the maximum percentage of correct answers is achieved more rapidly.

The experiments show that the ranking method has a strong dependence on the number and representative of the indicators. According to Agrawal[41] is necessary to identify the good regions of the desired performance, therefore, we argue the importance of the selection of the adequate indicators in the measure.

The results regarding the relationship between false positives and correct answers presented in figure 4 shows that as the number of indicators increases, the Pareto method behaves as a linear ranking; consequently, Pareto optimization with few indicators is not a good strategy for classification because its behavior is like that of the random case.

VII. CONCLUSIONS

The Pareto optimization method is presented in this article as an alternative for the generation of rankings. It is found that the ranking method yields good results in measurements involving many organizations (more than 10 in the experiments) because the organizations on the Pareto frontier can be found in the top positions in the rankings, regardless of the weights of the indicators. However, if there are many indicators, most organizations will be on the Pareto frontier because most organizations will stand out in terms of at least one indicator.

It is important to note that the classification obtained via linearly weighted rankings largely depends on the weights of the indicators. Therefore, an organization with the best rating in a group of indicators could be classified either incorrectly if the weights are unfavorable or correctly if the weights are favorable. In contrast, in the Pareto optimization method, such an organization will be on the frontier, which means that it will be classified correctly because it is the best in terms of one of the indicators and also because it will be in one of the top positions in a linear ranking with an appropriate weighting scheme.

VIII. REFERENCES

[1] P. Macmillan and I. Smith, "Explaining International Soccer Rankings," J. Sports Econom., vol. 8, no. 2, pp. 202–213, 2007.

[2] O. Kaltmeier and M. Breuer, "Social inequality," Routledge Handb. to Polit. Econ. Gov. Am., 2020.

[3] T. Luque-Martínez and N. Faraoni, "Meta-ranking to position world universities," Stud. High. Educ., vol. 45, no. 4, pp. 819–833, 2020.

[4] A. R. Bradley, Z. Manna, and H. B. Sipma, "Linear Ranking with Reachability," in Computer Aided Verification, 2005, pp. 491–504.

[5] B. P. G. Altbach, "Views The State of the Rankings," pp. 11–14, 2010.

[6] A. Rauhvargers, Global University Rankings and Their Impact. 2011.

[7] Philip G. Altbach, "The State of the Rankings," The inside highered, pp. 11–14, 11-Nov-2011.

[8] F. Wilcoxon, "Individual Comparisons by Ranking Methods," pp. 196–202, 1992.

[9] L. Waltman, "A review of the literature on citation impact indicators," J. Informetr., vol. 10, no. 2, pp. 365–391, 2016.

[10] A. F. J. Van Raan, T. N. Van Leeuwen, M. S. Visser, N. J. van Eck, and L. Waltman, "Rivals for the crown: Reply to Opthof and Leydesdorff," J. Informetr., vol. 4, no. 3, pp. 431–435, 2010.

[11] N. K. Jajo and J. Harrison, "World university ranking systems: an alternative approach using partial least squares path modelling," J. High. Educ. Policy Manag., vol. 36, no. 5, pp. 471–482, 2014.

[12] E. Triperina, C. Sgouropoulou, I. Xydas, O. Terraz, and G. Miaoulis, "Assessing the performance of educational institutions: A multidimensional approach," IEEE Glob. Eng. Educ. Conf. EDUCON, pp. 1337–1344, 2017.

[13] M. Erdoğan and İ. Kaya, "A type-2 fuzzy MCDM method for ranking private universities in İstanbul," Proc. World Congr. Eng. (Vol. 1, pp. 2-4)., vol. 1, pp. 2–4, 2014.

[14] E. Celik, M. Gul, N. Aydin, A. T. Gumus, and A. F. Guneri, "A comprehensive review of multi criteria decision making approaches based on interval type-2 fuzzy sets," Knowledge-Based Syst., vol. 85, pp. 329–341, 2015.

[15] G. Martin, S. Siebert, and I. Robson, "Conformist innovation: An institutional logics perspective on how HR executives construct business school reputations," Int. J. Hum. Resour. Manag., Oct. 2016.

[16] D. Pyne, "The rewards of predatory publications at a small business school," J. Sch. Publ., vol. 48, no. 3, pp. 137–160, 2017.

[17] J. I. Pulido-Fernández and B. Rodríguez-Díaz, "Reinterpreting

the World Economic Forum's global tourism competitiveness index," Tour. Manag. Perspect., vol. 20, pp. 131–140, 2016.
[18] R. P. Curiel and S. R. Bishop, "A metric of the difference between perception of security and victimisation rates," Crime Sci., vol. 5, no. 1, 2016.
[19] E. Lora, The distance between perception and reality in the social domains of life. 2016.
[20] J. Ratcliffe, B. Kaambwa, C. Hutchinson, and E. Lancsar, "Empirical Investigation of Ranking vs Best–Worst Scaling Generated Preferences for Attributes of Quality of Life: One and the Same or Differentiable?," Patient, 2020.
[21] A. Bonaccorsi and T. Cicero, "Nondeterministic ranking of university departments," J. Informetr., vol. 10, no. 1, pp. 224–237, 2016.
[22] S. Marginson, "University rankings and social science," Eur. J. Educ., vol. 49, no. 1, pp. 45–59, 2014.
[23] G. Doğan and U. Al, "Is it possible to rank universities using fewer indicators? A study on five international university rankings," Aslib J. Inf. Manag., vol. 71, no. 1, pp. 18–37, 2019.
[24] D. Ross, "Times higher education world university rankings," in Research Analytics: Boosting University Productivity and Competitiveness through Scientometrics, CRC Press, 2017, pp. 137–146.
[25] F. L. Bookstein, H. Seidler, M. Fieder, and G. Winckler, "Too much noise in the Times Higher Education rankings," Scientometrics, vol. 85, no. 1, pp. 295–299, 2010.
[26] P. Katerattanakul, B. Han, and S. Hong, "Objective quality ranking of computing journals," Commun. ACM, vol. 46, no. 10, pp. 111–114, 2003.
[27] H. Cousijn, P. Feeney, D. Lowenberg, E. Presani, and N. Simons, "Bringing citations and usage metrics together to make data count," Data Sci. J., vol. 18, no. 1, 2019.
[28] A. Singh, "Enabling Researchers to Make Their Data Count," SSRN Electron. J., 2019.
[29] N. Pollock, L. D'Adderio, R. Williams, and L. Leforestier, "Conforming or transforming? How organizations respond to multiple rankings," Accounting, Organizations and Society, vol. 64, pp. 55–68, 2018.
[30] C. Spence, "'Judgement' versus 'metrics' in higher education management," High. Educ., vol. 77, no. 5, pp. 761–775, 2019.
[31] Y. Sawaragi, H. Nakayama, and T. Tanino, Theory of Multiobjective Optimization. Academic Press, 1985.
[32] S. Lozano and L. Calzada-Infante, "Efficiency ranking using dominance network and multiobjective optimization indexes," Expert Syst. Appl., vol. 126, pp. 83–91, 2019.
[33] R. T. Marler and J. S. Arora, "The weighted sum method for multi-objective optimization: New insights," Struct. Multidiscip. Optim., vol. 41, no. 6, pp. 853–862, 2010.
[34] A. Konak, D. W. Coit, and A. E. Smith, "Multi-objective optimization using genetic algorithms: A tutorial," Reliab. Eng. Syst. Saf., vol. 91, no. 9, pp. 992–1007, 2006.
[35] G. S. Mahapatra and T. K. Roy, "Fuzzy multi-objective mathematical programming on reliability optimization model," Appl. Math. Comput., vol. 174, no. 1, pp. 643–659, 2006.
[36] M. Morishima and A. M. Pujana, Teoría económica de la sociedad moderna, 1st ed. Barcelona, España: Antoni Bosch, 1981.
[37] J. M. Lourenco and L. Lebensztajn, "Post-Pareto Optimality Analysis with Sum of Ranking Differences," IEEE Trans. Magn., vol. 54, no. 8, 2018.
[38] C. Solís and X. Wang, "A study of the characteristics of behaviour driven development," Proc. - 37th EUROMICRO Conf. Softw. Eng. Adv. Appl. SEAA 2011, pp. 383–387, 2011.
[39] I. Dees, M. Wynne, and A. Hellesoy, Cucumber Recipes: Automate Anything with BDD Tools and Techniques. Pragmatic Bookshelf, 2013.
[40] M. de Ciencias, "Convocatoria medición de grupos," 2020. [Online]. Available: https://minciencias.gov.co/convocatorias/fortalecimiento-capacidades-para-la-generacion-conocimiento/medicion-grupos. [Accessed: 12-May-2020].
[41] G. Agrawal, K. Lewis, K. Chugh, C. H. Huang, S. Parashar, and C. L. Bloebaum, "Intuitive visualization of Pareto Frontier for multi-objective optimization in n-dimensional performance space," Collect. Tech. Pap. - 10th AIAA/ISSMO Multidiscip. Anal. Optim. Conf., vol. 3, pp. 1523–1533, 2004.