EXCELLENCE WITH LEADERSHIP: THE CROWN INDICATOR OF SCIMAGO INSTITUTIONS RANKINGS IBER REPORT

Veljko Jeremić, Marina Jovanović-Milenković, Zoran Radojičić, and Milan Martić

Abstract
Although there are many models for ranking higher education institutions, the SCImago Institutions Rankings methodology stands out for its ability to present quantitative and qualitative indicators of scientific output. Besides Total number of published papers, several indicators are concerned with quality dimensions of published papers, such as International collaboration, Scientific leadership or High quality publications. However, official rankings are provided solely on the basis of one indicator: Output (total number of published papers). This paper presents a statistical I-distance method that integrates all the indicators into one value, which therefore represent a rank and show which of the input indicators is the most important for the process of ranking. Our results clearly showed that Excellence with Leadership occupies the most significant spot.


Veljko Jeremić is an assistant professor at the Department of Operational Research and Statistics at the Faculty of Organizational Sciences, University of Belgrade. He was born in 1985, finished his PhD thesis in the field of computational statistics in the 2012. He has published over 60 scientific papers with emphasis on ISI indexed journals (over 20 papers). He is co-author of one university textbook. He has acted as a reviewer for many ISI indexed journals such as: Jasist, Scientometrics, Journal of applied statistics.

http://orcid.org/0000-0002-7761-1841
jeremic.veljko@fon.bg.ac.rs

Marina Jovanović-Milenković is an assistant professor at the Department of Information Technologies at the Faculty of Organizational Sciences, University of Belgrade. She was born in 1976, finished her PhD thesis in the field of information and communication technology in health care management in 2011. She published over 30 scientific papers with emphasis on ISI indexed journals (5 papers). She has participated in the organization of the doctoral degree study program Management of the health system. Her major interests are e-health, ICT implementations in health systems and business decision making.

http://orcid.org/0000-0003-3687-5282
marinaj@fon.bg.ac.rs

Zoran Radojičić is an associate professor at the Department of Operational Research and Statistics at the Faculty of Organizational Sciences, University of Belgrade. He defended his PhD thesis in the field of computational statistics in 2007. Main research interests include: statistics, applied statistics, social sciences and biostatistics. He has published over 40 papers at SCI Thomson Reuters database. He has participated in Tempus and numerous projects supported by the Ministry of Science of Republic of Serbia. He was included in international projects financed by World Bank and EBRD in the region.

http://orcid.org/0000-0002-1536-8591
radojicic.zoran@fon.bg.ac.rs

Milan Martić is a dean and full professor at the Department of Operational Research and Statistics at the Faculty of Organizational Sciences, University of Belgrade. He defended his PhD thesis in the field of operational research in 1999. He has published over 130 scientific papers with emphasis on ISI indexed journals (over 12 papers). He is co-author of seven university textbooks. He has acted as a reviewer for many ISI indexed journals. His major interests are data envelopment analysis, efficiency models and operational research.

http://orcid.org/0000-0002-0289-6678
martic.milan@fon.bg.ac.rs

University of Belgrade, Faculty of Organizational Sciences, Jove Ilica 154, 11000 Belgrade, Serbia
1. Introduction

The increasing number of methodologies for ranking higher education institutions (HEI) has attracted many different stakeholders, especially students. Consequently, those rankings are quite often used as an indicator of a university’s reputation and performance. Although there are many models to classify institutions of higher education, the methodology of SCImago Institutions Rankings (SIR) is one of the most influential.

Keywords
Ranking of universities, I-distance method, SCImago Institutions Rankings Iber reports, Statistical methods.

Título: Excelencia con liderazgo: el indicador rey del SCImago Institutions Rankings Iber report

Resumen
Aunque hay muchos modelos para clasificar instituciones de educación superior, la metodología de SCImago Institutions Rankings destaca por su capacidad de presentar indicadores cuantitativos y cualitativos de la producción científica. Además del Número total de artículos publicados, varios indicadores se refieren a aspectos de calidad de los trabajos, como Colaboración internacional, Liderazgo científico o Publicaciones de alta calidad. Sin embargo, las clasificaciones oficiales se ofrecen únicamente en base a un indicador: Resultados (número total de artículos publicados). Este trabajo presenta el método de I-distancia estadística que integra todos los indicadores en un valor, lo que facilita el establecimiento de una lista o ranking, y muestra cuál de los indicadores es el más importante para el proceso de clasificación. Los resultados muestran claramente que Excelencia con liderazgo es el más influyente.

Palabras clave
Rankings de universidades, Método I-distancia, SCImago Institutions Rankings Iber reports, SIR, Métodos estadísticos.

The SIR approach integrates one quantitative and various qualitative variables. The Output (O) indicator is a measure of the quantity of an institution’s publication output. It represents the total number of documents published in scholarly journals indexed in Scopus (Romero-Fernández et al., 2011).

Seven other variables represent the quality dimension of scientific output: International collaboration (IC), Normalized impact (NI), High quality publications (Q1), Specialization index (SI), Excellence rate (ER), Scientific lead (Lead) and Excellence with leadership (Ewl).

The IC variable represents an institution’s output ratio produced in collaboration with foreign institutions. The values are computed by analyzing output with affiliations including more than one country address (Guerrero-Bote; Olmeda-Gómez; De-Moya-Anegón, 2013; Lancho-Barrantes; Guerrero-Bote; De-Moya-Anegón, 2013). However, the question raised is whether this characteristic actually provides insight into scientific output, since cooperation of Serbian and Montenegrin universities, for instance, is international while these countries were until recently one state union. Further, NI compares the average scientific impact of the institution with the world average (taken as 1). Thus, a score of 0.8 implies a performance 20% below average, whereas a score of 1.3 means the institution is considered 30% above average (González-Pereira; Guerrero-Bote; De-Moya-Anegón, 2010). One can argue that since “citation is counted only if it is made to an item published in the three previous years”, some important citations have been excluded (Getchew-Dinku, 2011).

Also, Q1 is the ratio of publications that the institution publishes in what the SCImago team takes as the most influential scholarly journals of the world: those ranked in the first quartile (25%) in their categories as calculated by SCImago.
Journal Rank. Since this is reported as a percentage, the ratio (Q1/25) is a crude normalized proxy for quality of publication, with a value of 1 taken as the world average (Miguel; Chinchilla-Rodríguez; De-Moya-Anegón, 2011). It is essential to mention that SJR doesn’t include journals’ self-cites (as done by Thomson Reuters IF) and the weight of citations depends upon the “prestige” of the citing journal (Falagas et al., 2008).

The SI variable indicates the extent of thematic concentration/dispersion of an institution’s scientific output. The values of this indicator range between 0 and 1, indicating more generalist or specialized institutions, respectively (López-Illescas; De-Moya-Anegón; Moed, 2011). On the other hand, ER indicates the percentage of an institution’s scientific output that is included in the set formed by 10% of the most-cited papers in their respective scientific fields. This indicator serves as a measure of the high-quality output of research institutions. Again, the ratio ER/10 allows one to normalize this indicator so that the world average becomes 1 (Bornmann; De-Moya-Anegón; Leydesdorff, 2012). Lead indicates an institution’s “output as main contributor”, that is the number of papers in which the corresponding author belongs to the institution (De-Moya-Anegón, 2012). Finally, EwI indicates the amount of documents in the Excellence rate in which the institution is the main contributor (SIR, 2013).

Nonetheless, although the SIR Iber report presents all the valuable data, the official rankings are presented based only on the number of Total published papers (indicator Output - O). Having said this, it is essential to provide a potential upgrade of current framework and create a synthesised indicator that will incorporate both the quantitative and qualitative dimensions of SIR Iber report. Further, it is vital to establish which of these dimensions provides better insight into scientific excellence of a HEI. As a possible remedy to the issue, the statistical I-distance method is elaborated and applied.

2. I-distance method

Quite frequently, the score obtained in a specific league list can seriously affect the process of taking exams, entering competitions, UN projects participation, medicine selection and many other areas (Jeremić; Radojičić, 2010; Al-Lagilli et al., 2011). I-distance is a metric distance in an n-dimensional space. It was originally proposed and defined by Branislav Ivanović, and has appeared in various publications since 1963 (Ivanović, 1977). Ivanović devised this method to rank countries according to their level of development on the basis of several indicators; many socio-economic development indicators had been considered and the problem was how to use all of them in order to calculate a single synthetic indicator which would thereafter represent the rank.

For a selected set of variables $X_i$ (for $i=1,...,k$) and $j$ chosen to characterize the entities, the I-distance between the two entities $e_r=(X_{i1},X_{i2},...,X_{ik})$ and $e_s=(X_{s1},X_{s2},...,X_{sk})$ is defined as

$$D(r,s)=\frac{\sum_{i=1}^{k}d_i(r,s)}{\sigma_i}\prod_{j=1}^{i-1}(1-r_{ji}^{12...j-1})$$

where $d_i(r,s)$ is the distance between the values of variable $X_i$ for $e_r$ and $e_s$, e.g. the discriminate effect,

$$d_i(r,s)=x_{ir}-x_{is}, \ i \in \{1,...,k\}$$

$\sigma_i$ the standard deviation of $X_i$, and $r_{ji}^{12...j-1}$ is a partial coefficient of the correlation between $X_i$ and $X_j$ ($j<i$) (Bulajić et al., 2012; Dobrota; Jeremić; Marković, 2012).

The construction of the I-distance is iterative; it is calculated through the following steps:

- Calculate the value of the discriminate effect of the variable $X_1$ (the most significant variable, that which provides the largest amount of information on the phenomena that are to be ranked).
- Add the value of the discriminate effect of $X_1$ which is not covered by $X_1$
- Add the value of the discriminate effect of $X_2$ which is not covered by $X_1$ and $X_2$
- Repeat the procedure for all variables (Jeremić et al., 2012; Radojičić et al., 2012).

Sometimes, it is not possible to achieve the same sign mark for all variables in all sets, and, as a result, a negative correlation coefficient and a negative coefficient of partial correlation may occur (Jeremić et al., 2011; Maletić et al., 2012). This makes the use of the square I-distance even more desirable. The square I-distance is given as:

$$D^2(r,s)=\sum_{i=1}^{k}\frac{d_i^2(r,s)}{\sigma_i^2}\prod_{j=1}^{i-1}(1-r_{ji}^{212...j-1})$$

In order to rank the entities (in this case, universities), it is necessary to have one entity fixed as a referent in the observing set using the I-distance methodology (Jeremić et al., 2012; Jovanović et al., 2012). The entity with the minimal value for each indicator or a fictive minimal entity should be utilized as the referent entity, as the ranking of the entities in the set is based on the calculated distance from the referent entity (Seke et al., 2013).

3. Results of the I-distance method

For this study, the latest release of the SCImago Institutions Rankings (SIR) Iber reports (SIR, 2013) was analyzed. Out of the 1,600 leading research institutions ranked in the SIR 2013 Iber reports, 148 universities with the Output indicator larger than 1,000 papers were selected and the I-distance method was performed on that sub dataset. The results achieved by means of the squared I-distance method for the first 20 HEI are shown below in table 1 (full list is available upon request). The construction of the I-distance uses an iterative approach, and the crucial idea is to reduce the duplication of information. It is done by partially integrating the variables into the creation of one value which will therefore represent the rank. Particularly interesting is that variables with different types of measurements (percentages, GDP, student enrolment rate, etc.) could easily be integrated into one variable.
As can be seen from table 1, *Universidade de São Paulo* tops the I-distance method. This university has an impressive *Output* indicator, with 47,833 published papers. On the other hand, all the other indicators are solid but far from the best. One should note that the *Universitat Rovira i Virgili* is highly placed (2nd spot), although it has a rather small number of published papers – just 4,186 (less than 10% of *São Paulo’s* output). Precisely this information is crucial because it is essential to elaborate other variables in which the *Universitat Rovira i Virgili* impresses. For instance, in a quality indicator such as *Excellence with leadership* (~ *Ewl* (10.6) and *Excellence rate* ~ *ER* (17.17), the *Universitat Rovira i Virgili* is one of the best universities (number one concerning the indicator “*Excellence with leadership*”, third in the “*Excellence rate*”). A similar conclusion has been noted by *Radojičić & Jeremić* (2012), for *Rockefeller University*, a postgraduate-only institution with a small number of published papers but fully impressive in quality indicators such as *QI* (88.6%) and *ER* (48.8).

Consequently, it is essential to determine which of the eight input indicators is the most important for the process of ranking. Thus, this data set has been further examined and the correlation coefficients of each variable with the I-distance values have been determined. The results shown in table 2 demonstrate that the most significant variable for the calculated I-distance value is *Excellence with leadership* (*Ewl*), highly correlated with the I-distance value (r = 0.676, p < 0.01). Total number of published papers (*Output*), *Excellence rate* (*ER*), *Normalized impact* (*NI*) and *High quality publications* (*Q1*) are also very important indicators, with each correlation larger than 0.5 (p < 0.01). This finding clearly shows that the qualitative dimension of scientific output must not be neglected at any cost.

### 4. Concluding remarks

The increasing number of ranking methodologies used to identify the world’s best universities is pushing the academic world into becoming even more concerned with the assessment of higher education. With these rankings often used as a marketing tool for universities to show their educational or research excellence, the need to provide rankings as accurate as possible becomes exceptionally important (*Radojičić & Jeremić*, 2012). The analysis presented here stresses potential improvements in the *SCimago Ranking* methodology, with emphasis on the quality indicator of universities’ scientific output. The idea is consistent with the approach taken by *Torres-Salinas et al.* (2011), which presents the IFQ2A index as an integrated index of qualitative and quantitative scientific indicators. In addition, the *Leiden ranking* (Leiden, 2013) has some similarities with *SCimago* in terms of bibliometric data, and rankings focused on the research performance of institutions. However, in *Leiden ranking* the journals that do not have a strong international scope (i.e., not published in English or articles with authors concentrated in one or a few countries) or have a small number of references to other journals in the *Web of Science* database are being excluded from the analysis (*Waltman et al.*, 2012). Perhaps this could be a way to go for following SIR lists. Moreover, by default the *Leiden ranking* reports size-independent indicators (average statistics per publication, such as a university’s average number of citations per publication). The advantage of size-independent indicators is that they enable comparisons between smaller and larger universities (Leiden, 2013). As an alternative to size-independent indicators, the *Leiden ranking* can also report size-dependent indicators, which provide overall statistics of the publications of a university (the total number of citations of the publications of a university). Size-dependent indicators are strongly influenced by the size of a university and therefore tend to be less useful for comparison purposes (*Waltman et al.*, 2012). Furthermore, our approach could contribute to the emerging efforts to map regions of academic excellence and scientific output (*Bornmann; Leydesdorff*, 2011; *Bornmann; Waltman*, 2011; *Prathap*, 2011). Additionally, I-distance methodology could embrace the postulates of bootstrapping and we hope to encourage potential researchers in this area.
Acknowledgment

The authors would like to thank the editor and anonymous reviewers for their valuable comments and suggestions, which significantly improved the quality of the paper.

5. References


García, José A.; Rodríguez-Sánchez, Rosa; Fernández-Valdivia, Joaquín; Torres-Salinas, Daniel; Herrera, Francisco (2012). “Ranking of research output of universities on the
basis of the multidimensional prestige of influential fields: Spanish universities as a case of study”. Scientometrics, v. 93, n. 3, pp. 1081-1099. http://dx.doi.org/10.1007/s11192-012-0740-7


Jovanović, Milica; Jeremić, Veljko; Savić, Gordana; Bulajić, Milica; Martić, Milan (2012). “How does the normalization of data affects the ARWU ranking?”. Scientometrics, v. 93, n. 2, pp. 319-327. http://dx.doi.org/10.1007/s11192-012-0674-0


Radojičić, Zoran; Jeremić, Veljko (2012). “Quantity or qual-