STI 2018 Leiden

23rd International Conference on Science and Technology Indicators
"Science, Technology and Innovation Indicators in Transition"

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at https://hdl.handle.net/1887/64521

ISBN: 978-90-9031204-0

23rd International Conference on Science and Technology Indicators (STI 2018)
"Science, Technology and Innovation indicators in transition"
12 - 14 September 2018 | Leiden, The Netherlands
#STI18LDN

With or without h-index? Comparing aggregates of rankings based on seven popular bibliometric indicators[1]

Andrey Subochev[*] and Vladimir Pislyakov[**]

[*] *asubochev@hse.ru*
DeCAn Lab and Department of Mathematics, Faculty of Economic Sciences, National Research University Higher School of Economics, Myasnitskaya 20, Moscow, 101000 (Russia)

[**] *pislyakov@hse.ru*
Library, National Research University Higher School of Economics, Myasnitskaya 20, Moscow, 101000 (Russia)

Abstract

We apply five majority-rule-based ordinal ranking methods to data on economic, management and political science journals in order to produce aggregate journal rankings. First, we calculate aggregates for the set of rankings based on seven popular bibliometric indicators (impact factor, 5-year impact factor, immediacy index, article influence score, h-index, SNIP and SJR). Then, we exclude the Hirsch index and repeat the calculations. We perform the comparative correlation analysis of the aggregates and the initial rankings. We use two rank measures of correlation, Kendall's τ_b and the share of coinciding pairs r. The analysis demonstrates that aggregate rankings represent the set of single-indicator-based rankings better than any of the seven rankings themselves. Among the single-indicator-based rankings themselves, the best representations of their set are produced by the 5-year impact factor. The least representative are rankings based on the immediacy index. The exclusion of the Hirsch index from the set of indicators does not change these results.

Introduction

The emergence of the Scopus database and the invention of the h-index (Hirsch, 2005) revitalized the interest in developing various bibliometric measures. However, their growing multiplicity generates two questions.

(a) How do the rankings based on different measures correlate with each other?
(b) What a decision-maker can do if there are several rankings but he/she needs just one?

Thus, we began with analysis of correlations between the rankings based on seven popular indicators, which are impact factor (IF), 5-year impact factor (IF-5), immediacy index (II), article influence score (AI), h-index (Hirsch), SNIP and SJR. This had already been done in a number of comparative studies, which were focused either on indicators from different databases (Archambault et al., 2009; Delgado & Repiso, 2013; Leydesdorff, 2009), or on citation, network and usage metrics (Bollen et al., 2009). The reviews of Waltman (2016), Rousseau (2002) and Glänzel (2003) may serve as an introduction to the vast literature on citation indicators. In agreement with the previous results, we confirmed that all rankings are

---

[1] The study was financially supported through the Basic Research Program at the National Research University Higher School of Economics (HSE) and by the Russian Academic Excellence Project '5-100'.

positively correlated with each other. Nevertheless, it was also found that there was a non-negligible percentage of contradictions.

The multiplicity of contradicting evaluations is a problem for a decision-maker. To make decisions, there should be just one ranking. An obvious solution is to choose the best indicator. Unfortunately, the academic discussion concerning relative advantages of various indicators has been inconclusive so far. Since there is no compelling reason to presume that one indicator is somehow inferior to the others, it is problematic to make the choice rationally.

Instead of choosing the best indicator, a decision-maker may choose an appropriate aggregation procedure and use all rankings available. The theory of aggregation is a well-developed area, and, consequently, it allows one to make quite definite conclusions regarding the appropriateness of such a choice.

To construct an aggregate ranking is to rank on a basis of multiple criteria. There exists a formal analogy between the multicriteria decision-making and the social choice (Arrow & Raynaud, 1986). Therefore, a decision-maker may consider the whole panoply of extensively studied and well-behaved social choice procedures. We propose to use *ordinal* aggregation methods based on the *majority rule*. In our paper (Subochev, Aleskerov & Pislyakov, 2018), we presented an axiomatic analysis of the aggregation functionals and provided the theoretical arguments in favor of these methods. Here, we present some empirical evidence supporting our proposal. We perform the formal comparative correlation analysis of the aggregates and the initial rankings. In order to check the robustness of our conclusions, we use two measures of correlation, Kendall's τ_b and the share of coinciding pairs r. The rank correlation analysis confirms that the aggregates thus obtained reduce the number of contradictions and represent the sets of single-indicator-based rankings better than any member of a set does.

Data

We consider three sets of journals representing three academic disciplines: economics, management and political science. Rankings are computed for each set separately. Sets of journals were taken from Journal Citation Reports (JCR) database from Clarivate Analytics (then Thomson Reuters IP), along with their IF, 5-year IF, immediacy index and AI indicators (all for JCR-2011 edition). SNIP and SJR metrics for 2011 were taken from Journal Metrics website powered by Scopus database; h-index for each journal was calculated manually by searching Web of Science database. To make h-index more definite, the exact publication and citation windows have been applied. Only papers appeared from 2007 to 2011 have been considered, and citations to them made during the same period, 2007–2011. After exclusion of publications with missing values, the sets contain 212 economic journals, 93 management science journals and 99 political science journals.

The main selection criteria for indicators were their popularity and diversity of data sources and methodologies. The latter is particularly important, since it is senseless to aggregate rankings if they are based on identical indicators. In order to capture a relatively vague concept of the "journal influence", it seems better to use several measures, and these measures should be as independent and dissimilar as it is possible.

The set of selected indicators contains all kinds of metrics. There are un-weighted as well as weighted (AI, SJR) measures, size-dependent (h-index) as well as size-independent ones. The indicators use different publication windows, from one (immediacy index) to five (5-year IF, AI) years. Moreover, they are taken from different databases. A choice of a database may significantly change the values of indicators even when they are based on the same methodology (Pislyakov, 2009). Data sources and properties of metrics are summarized in Table 1.

Table 1. Indicators: sources and properties.

	Database	*Year*	*Publication window, years*	*Weighted*	*Size-dependent*
Impact factor (IF)	WoS/JCR	2011	2	No	No
5-year impact factor (IF-5)	WoS/JCR	2011	5	No	No
Immediacy index (II)	WoS/JCR	2011	1	No	No
Article influence (AI)	WoS/JCR	2011	5	Yes	No
h-index (Hirsch)	WoS	2007–2011 (papers and citations)	5	No	Yes
SNIP	Scopus	2011	3	No	No
SJR	Scopus	2011	3	Yes	No

Since there is a disagreement among scientometricians concerning desirability of aggregating rankings which are based on size-dependent indicators with rankings based on size-independent ones, we excluded h-index from the set of indicators at the second stage of the research and repeated all the calculations for the set of six size-independent indicators only. That is, we obtained two sets of results, with and without h-index.

Aggregation methods

We consider ranking of journals as a multicriteria decision problem. It is possible to frame any multicriteria decision problem as a social choice problem if one treats a ranking based on a certain criterion as a representation of preferences of a certain voter. In our case, the set of rankings based on corresponding bibliometric indicators is treated as a profile of opinions of either seven or six virtual experts.

Let A denote the set of feasible alternatives; let N denote a group of experts making a collective decision. *Preferences* of a voter i, $i \in N$, are revealed through pairwise comparisons of alternatives and are modeled by a binary relation P_i on A, $P_i \subseteq A \times A$: if voter i prefers x to y, then the ordered pair (x, y) belongs to the relation P_i. If a voter is unable to compare two alternatives or thinks they are of equal value, it will be presumed that he is indifferent regarding the choice between them. Probably, the best method to construct *social preferences* P of group N is to apply *the majority rule*: (x, y) belongs to P if the number of those who think x is better than y is greater than the number of those who think y is better than x: $xPy \Leftrightarrow |N_1| > |N_2|$, where $N_1 = \{i \in N | \ xP_iy\}$, $N_2 = \{i \in N | \ yP_ix\}$. In this case, P is called *the majority relation*. We present the arguments in favor of this particular rule of aggregation in (Subochev, Aleskerov & Pislyakov, 2018).

The majority relation quite often happens not to be a ranking itself since it is generally not transitive, either positively or negatively. For instance, the majority relation may contain cycles. This result is known as *the Condorcet paradox* (Condorcet, 1785). In order to check if the majority relation is transitive or not and to evaluate how nontransitive it is, we calculate the number of 3-step P-cycles, 4-step P-cycles and 5-step P-cycles for the set of seven indicators (Table 2) and for the set without h-index (Table 3).

Table 2. Numbers of 3-, 4- and 5-step *P*-cycles for the set of seven indicators.

	3-step cycles	*4-step cycles*	*5-step cycles*
Economics	2446	22427	226103
Management	203	787	3254
Political Science	149	430	1344

Table 3. Numbers of 3-, 4- and 5-step *P*-cycles for the set of six indicators (without h-index).

	3-step cycles	*4-step cycles*	*5-step cycles*
Economics	167	822	3140
Management	19	36	57
Political Science	21	58	142

As we see, the Condorcet paradox occurs in all six cases. When we exclude h-index, all numbers drop. This is because the number of aggregated indices becomes even. As a result, the number of ties in the majority relation significantly increases. In our case, it has increased sixfold. New ties break *P*-cycles; therefore, the numbers of cycles decrease.

In order to bypass the nontransitivity problem, various majority-rule-based ranking methods have been proposed. Effectively, all such methods are ways to "mend" the majority relation whenever it happens not to be a ranking itself. We consider two versions of the Copeland rule (Copeland, 1951), a version of the Markovian method (Daniels, 1969; Ushakov, 1971) and the sorting procedure based on two tournament solutions - the uncovered set (Miller, 1980) and the minimal externally stable set (von Neumann & Morgenstern, 1944; Aleskerov & Kurbanov 1999; Subochev, 2008; Aleskerov & Subochev, 2013). The detailed description of these five methods is given in (Subochev, Aleskerov & Pislyakov, 2018) and in (Aleskerov, Pislyakov & Subochev, 2014). The table with ranks of all journals in aggregates of the seven single-indicator-rankings and in seven rankings themselves can be found in (Aleskerov, Pislyakov and Subochev, 2014) as well.

Correlation analysis

To evaluate the (in)consistency of two rankings, we measure their correlation. In this paper, we use two related but not identical measures based on the Kendall distance, namely, the Kendall rank correlation index τ_b and the share of coinciding pairs r. The share of coinciding pairs r is a percentage of pairs ranked in the same way in both rankings. This measure has a simple probabilistic interpretation. If someone knows that alternative x is ranked above alternative y in ranking R_1 and guesses that in ranking R_2 they are placed in the same order, then r is the probability of her being correct. When r=50%, probability of being right equals probability of being wrong, which means two rankings do not correlate. The main difference between τ_b and r is that the latter "punishes" rankings containing too many ties, while the former does not.

The corresponding numerical values of τ_b and r can be found in (Aleskerov, Pislyakov & Subochev, 2014) and in (Subochev, Aleskerov & Pislyakov, 2018).

We employ the same idea of binary multicriteria comparisons to evaluate all rankings formally. The problem of aggregation can be reformulated as a choice of a single object representing a given group of objects. In our case, we need to choose a ranking that serves as the best representative for the set of rankings based on the selected bibliometric indicators. We have either twelve or eleven candidates: the five aggregates and the prime rankings themselves. If the prime rankings were the preferences of some votes, then we would expect that in a binary contest a voter would vote for a representative whose preferences are closer to his or her own. Let us again use the majority rule to determine the best representations. Let us say that ranking X_1 represents a given set of rankings $\{R_i\}$, $i=1\div n$, better than ranking X_2 if X_1 is better correlated with the majority of rankings from $\{R_i\}$ than X_2. In our case, $\{R_i\}$ is a set of single-indicator-based rankings, n equals either 7 or 6, and each ranking X is characterized by two n-tuples of values of τ_b and r. A component number i of an n-tuple is a value of a corresponding correlation measure for the ranking X and a corresponding single-indicator-based ranking R_i. For each correlation measure and for each of the three sets of journals, we compare these n-tuples and compute the corresponding voting matrix $\mathbf{V}$. Entry v_{xy} of the voting matrix $\mathbf{V}$ is a natural number; it is a number of rankings R_i, with which ranking X is better correlated than ranking Y. Then for each $\mathbf{V}$, we calculate the majority relation P on the set of the rankings compared, $(X, Y) \in P \Leftrightarrow v_{xy} > v_{yx}$. Finally, we compute the majority relation for the results of our previous study (Aleskerov, Pislyakov & Subochev, 2011), where we ranked management journals by values of the same seven bibliometric indicators measured for the earlier periods. The voting matrices and the matrix representations of the majority relations are given in (Aleskerov, Pislyakov & Subochev, 2014).

If we apply the Copeland rule (2nd version) to the majority relations obtained, we will get the four sets of rankings, denoted Q_k, of ranking methods. These rankings are presented in Tables 4a and 4b. The methods that produce rankings which are better representations are ranked higher. The aggregates are highlighted.

Table 4a. The Copeland ranking of rankings (with h-index)

compared by τ_b

rank	Economics	Management	Political Science	Management (old results)
	Q_1	Q_2	Q_3	Q_4
1	*MES*	*MES*	*MES*	*UC*
2	*UC*	*UC*	*UC*	*MES*
3	Copeland 3	Copeland 2	Copeland 3	Copeland 3
4	Copeland 2	Copeland 3	Copeland 2	Copeland 2
5	Markov	Markov	Markov	Markov
6	IF-5	IF-5	IF-5	IF
7	IF	SNIP	Hirsch	IF-5
8	SJR	Hirsch	AI / IF / SJR	SJR
9	AI	AI		AI / Hirsch / SNIP
10	SNIP	SJR		
11	Hirsch	IF	SNIP	
12	II	II	II	II

compared by *r*

	Q_5	Q_6	Q_7	Q_8
1	Copeland 3	Copeland 3	Copeland3 / Copeland2 / Markov	Copeland 3
2	Copeland 2	Copeland 2		Copeland 2
3	Markov	Markov		Markov

4	*UC*	*UC*	*UC*	*UC*
5	IF-5	IF-5	IF-5	*MES*
6	IF	*MES*	*MES*	IF
7	*MES*	SNIP	AI	IF-5
8	AI	AI	IF	SJR
9	SNIP	IF / Hirsch / SJR	SNIP	SNIP
10	SJR		SJR	AI
11	Hirsch		Hirsch	Hirsch
12	II	II	II	II

Table 4b. The Copeland ranking of rankings (without h-index)

compared by τ_b

rank	**Economics**	**Management**	**Political Science**
	Q_9	Q_{10}	Q_{11}
1	*UC*	*UC / MES*	*UC*
2	*MES*		*MES*
3	Copeland2 / Copeland3	Copeland 3	Copeland2 / Copeland3
4		Copeland 2	
5	Markov	Markov	Markov
6	IF-5	IF-5	IF-5
7	IF	SNIP	IF
8	SJR	AI	SJR
9	AI / SNIP	IF / SJR	AI / SNIP
10			
11	II	II	II

compared by *r*

	Q_{12}	Q_{13}	Q_{14}
1	Copeland2 / Copeland3 / Markov	Copeland3 / Markov	Copeland2 / Copeland3 / Markov
2			
3		Copeland 2	
4	IF-5 / *UC*	IF-5	IF-5 / *UC*
5		*UC*	
6	*MES*	*MES*	*MES*
7	IF	SNIP	IF
8	AI / SNIP	AI	AI / SNIP
9		IF / SJR	
10	SJR		SJR
11	II	II	II

In all fourteen cases, the ranking by values of the immediacy index demonstrates the lowest level of correlation with the single-indicator-based rankings. In all the cases except two related to the older data, Q_4 and Q_8, the rankings based on the 5-year impact factor demonstrate the highest level of correlation among the single-indicator-based rankings. In the previous study (Q_4 and Q_8), the most correlated ranking was one based on the classic impact factor, the 5-year impact being the second best. The rankings based on h-index and SJR contain far fewer ranks than there are journals. The numbers of ranks in all rankings are presented in Table 5. Other systematic differences between single-indicator-based rankings are not observed.

In all cases when rankings are compared by τ_b, i.e. when one compares only Q_1, Q_2, Q_3, Q_4, Q_9, Q_{10} and Q_{11}, all aggregate rankings are placed above all single-indicator-based ones.

Table 5. Total number of ranks.

	Economics	Management	Political Science	Management (older results)
Total number of journals	212	93	99	82
IF	200	90	95	81
IF-5	207	92	98	81
II	159	84	72	66
AI	204	91	95	80
Hirsch	30	30	19	22
SNIP	201	92	97	81
SJR	65	41	28	41
with h-index				
Copeland 2	135	68	69	58
Copeland 3	139	69	66	58
UC	59	42	42	40
MES	37	33	36	30
Markov	211	93	97	81
without h-index				
Copeland 2	136	61	63	
Copeland 3	139	64	62	
UC	44	29	35	
MES	46	30	33	
Markov	207	92	97	

When rankings are compared by r, Hirsch, SJR, *UC* and *MES* go down in all cases, while relative positions of all other rankings remain practically the same.[2] This is explained by the fact that rankings based on h-index and SJR and aggregate rankings based on *UC* and *MES* contain significantly fewer ranks and, consequently, more tied pairs than other rankings. As a result, the values of r for the pairs that include one of these four rankings are lower, since this measure, unlike τ_b, "punishes" rankings containing too many ties. Indeed, a pair of journals tied in a ranking with many ties most probably will not be a tie in a ranking which is more refined. Thus, this pair will not contribute to the numerator of r, while r's denominator remains constant across all pairs.

This difference between two correlation measures explains why sorting by *MES* in Q_5, Q_6, Q_7, Q_{12}, Q_{13}, Q_{14} is placed below IF-5 and even below IF in Q_5, and why sorting by *UC* is placed below IF-5 in Q_{13} or tied with it in Q_{12} and Q_{14}. Taking into account the nature of this exception, we may safely conclude that all aggregate rankings are better representations of a set of initial single-indicator-based rankings in all cases considered. This supports our assertion that the aggregation based on the majority rule produces rankings that represent a set of single-indicator-based rankings better than any ranking from the set.

The exclusion of the h-index from the set of indicators changes almost nothing. There are just 6 inconspicuous inversions. In Q_9 and Q_{11}, *UC* is placed above *MES*, while *MES* is above *UC* in, correspondingly, Q_1 and Q_3. IF is below *MES* in Q_{12} and above it in Q_5. Copeland 3 is

---

[2] If one excludes Hirsch, SJR, *UC* and *MES* and compares Q_1 with Q_5, Q_2 with Q_6, Q_3 with Q_7, Q_4 with Q_8, Q_9 with Q_{12}, Q_{10} with Q_{13} and Q_{11} with Q_{14}, there will be just two inversions and a number of broken ties. Copeland 2 is placed above Copeland 3 in Q_2, but their order is reversed in Q_6. Copeland 2 is placed above Markov in Q_{10}, but their order is reversed in Q_{13}.

placed below Markov and *UC* is below IF-5 in Q_{13}, while their order is reversed in Q_6. Finally, the order of IF and AI is different in Q_7 and Q_{14}.

It is interesting to note that the Copeland rankings are almost never[3] placed below the Markovian ones despite the latter contain on average 1.5 times more ranks than the former.

Conclusion

Replacing the set of single-indicator-based rankings with majority-relation-based aggregates is justified, at least for the datasets considered. Judging from Tables 4a and 4b, the best aggregation method seems to be some version of the Copeland rule when one is interested in obtaining a fine ranking. If a coarse filtration is needed then one may use the sorting by either *UC* or *MES*. The exclusion of the Hirsch index from the set of indicators does not change these results.

References

Aleskerov, F., Kurbanov, E. (1999). Degree of manipulability of social choice procedures. In A. Alkan, Ch.D. Aliprantis N.C. Yannelis (Eds.), *Current Trends in Economics: Theory and Applications* (pp. 13–27). N.Y.: Springer-Verlag.

Aleskerov, F.T., Pislyakov, V.V., Subochev, A.N. (2014). *Ranking Journals in Economics, Management and Political Science by Social Choice Theory Methods. WP BRP 27/STI/2014*. Moscow: HSE. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2437850

Aleskerov, F.T., Pislyakov, V.V., Subochev, A.N., Chistyakov, A.G. (2011). *Rankings of Management Science Journals Constructed by Methods from Social Choice Theory. Working paper WP7/2011/04.* Moscow: Higher School of Economics. (in Russian).

Aleskerov, F., Subochev, A. (2013). Modeling optimal social choice: Matrix-vector representation of various solution concepts based on majority rule. *Journal of Global Optimization*, 56(2), 737–756.

Archambault, É., Campbell, D., Gingras, Y., Larivière, V. (2009). Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of the American Society for Information Science and Technology*, 60(7), 1320–1326.

Arrow, K.J., Raynaud, H. (1986). *Social Choice and Multicriterion Decision-Making*. Cambridge (Mass.): MIT Press.

Bollen, J., Van de Sompel, H., Hagberg, A., Chute, R. (2009). A principal component analysis of 39 scientific impact measures. PLoS ONE, 4(6), Article number e6022.

Condorcet, Marquis de. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.* Paris: L'imprimerie royale.

Copeland, A.H. (1951). *A reasonable social welfare function, Seminar on Application of Mathematics to the Social Sciences, University of Michigan, Ann Arbor*. Mimeo.

Daniels, H.E. (1969). Round-robin tournament scores. *Biometrica*, 56(2), 295–299.

Delgado, E., Repiso, R. (2013). The impact of scientific journals of communication: Comparing Google Scholar metrics, Web of Science and Scopus. *Comunicar*, 21(41), 45–52.

---

[3] The only exception is Copeland 2 in Q_{13}.

Glänzel, W. (2003). Bibliometrics as a Research Field: A course on theory and application of bibliometric indicators. Course Handouts. Leuven.

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

Leydesdorff, L. (2009). How are new citation-based journal indicators adding to the bibliometric toolbox? *Journal of the American Society for Information Science and Technology*, 60(7), 1327–1336.

Miller, N.R. (1980). A new solution set for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting. *American Journal of Political Science*, 24(1), 68–96.

von Neumann, J., Morgenstern O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Pislyakov, V. (2009). Comparing two "thermometers": Impact factors of 20 leading economic journals according to Journal Citation Reports and Scopus. *Scientometrics*, 79(3), 541–550.

Rousseau R. (2002). Journal evaluation: Technical and practical issues. *Library Trends*, 50(3), 418-439.

Subochev, A. (2008). *Dominant, Weakly Stable, Uncovered Sets: Properties and Extensions. Working paper WP7/2008/03.* Moscow: SU – Higher School of Economics. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2681061

Subochev, A., Aleskerov, F., Pislyakov, V. (2018). Ranking journals using social choice theory methods: A novel approach in bibliometrics. *Journal of Informetrics*, 12(2), 416–429.

Ushakov, I.A. (1971). The problem of choosing the preferable object. *Izvestiya Akademii Nauk SSSR. Tekhnicheskaya Kibernetika*, 4, 3–7. (in Russian)

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics.* 10(2), 365–391.