НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Популярные библиометрические индикаторы: сходства и различия

Ф.Т.Алескеров alesk@hse.ru

А.Н.Субочев asubochev@hse.ru

В.В.Писляков pislyakov@hse.ru

Национальный исследовательский университет
«Высшая школа экономики»

Библиометрические базы данных

В 1927 г. Гросс и Гросс опубликовали статью, в которой журналы по химии ранжировались с помощью матрицы цитирований.

В 1963 г. Юджин Гарфилд предложил формулу импакт-фактора и с 1964 г. начал публиковать The Science Citation Index.
Это привело к созданию базы *Web of Science*

Владелец WoS:	до 2016	Thomson Reuters Corporation
	с 2016	Clarivate Analytics

База *Scopus* создана в 2004
Владелец Scopus: Elsevier

Российская база научных публикаций - *Научная электронная библиотека*.
Создана в 1999 г. по инициативе РФФИ. Владелец: ООО НЭБ.
С 2005 г. по заказу Министерства образования и науки НЭБ создала и поддерживает систему Российского индекса научного цитирования (РИНЦ)

Импакт факторы (IF и IF5)

Классический IF (IF2) (Garfield, Sher, 1963).

IF5 публикуется Thomson Reuters (в н.в. Clarivate Analytics) с 2007.

$PUB(T)$ – число статей, опубликованных в данном журнале в году T.

$CIT(T_1, T_2)$ - число всех цитирований, полученных в году T_1 всеми статьями данного журнала, опубликованными в году T_2.

n – окно цитирования (число лет).

$$\mathrm{IF}n(T) = \frac{\sum_{t=1}^{n} CIT(T,\ T-t)}{\sum_{t=1}^{n} PUB(T-t)}$$

Индекс оперативности (Immediacy index)

PUB(*T*) – число статей, опубликованных в данном журнале в году *T*.

CIT(*T*, *T*) - число всех цитирований, полученных в году *T* всеми статьями данного журнала, опубликованными в году *T*.

$$II(T) = \frac{CIT(T,T)}{PUB(T)}$$

SNIP (source normalized impact per paper)

SNIP создан Хенком Мёдом (Moed, 2010), рассчитывается по базе Scopus.

Для каждого журнала определяется «индивидуальное предметное поле» - все публикации данного года, в которых цитируется хотя бы одна статья, опубликованная в данном журнале в предыдущие 10 лет.

ANR – среднее число ссылок (средняя длина списков литературы) в публикациях из индивидуального предметного поля.

$$\mathrm{SNIP} = \frac{IF3}{ANR}$$

Eigenfactor и индекс влиятельности статьи (article influence)

Eigenfactor предложен в 2007 г. исследователями Bergstrom Laboratory (University of Washington).

Для расчёта используется одна из версий метода собственного вектора. Айгенфактор журнала *i* равен значению *i*-той компоненты собственного вектора, соответствующего корню Фробениуса (максимальному собственному значению) матрицы цитирований.

Рассчитывается по базе WoS. Окно цитирования – 5 лет. Самоцитирования не учитываются.

Индекс влиятельности статьи журнала (AI) получается нормализацией айгенфактора делением на число статей в журнале.

SJR (SCImago Journal Rank)

Создан испанской исследовательской группой SCImago

Также используется одна из версий метода собственного вектора для оценки цитирований полученных из разных источников. Процедура вычислений - итеративная. Цитирования учитываются с весами, зависящими от цитируемости самих цитирующих источников, определённой на предыдущем этапе. Вычисления останавливаются тогда, когда изменения весов становятся меньше заданной погрешности. Итоговая сумма числа цитат нормализуется делением на число статей в журнале. Подробное описание методологии в (Gonzalez-Pereira et al., 2010)

Окно цитирования – 3 года. Рассчитывается по базе Scopus. Самоцитирования учитываются, но для объёма учитываемых самоцитирований установлена верхняя граница – 33% от общего числа цитат.

Индекс Хирша (h-index)

Создан Хорхе Хиршем (Hirsch, 2005). Оценивает цитируемость данного множества публикаций (а не отдельной публикации).

Изначально введён для оценки продуктивности исследователей.

Значение индекса данного множества публикаций равно *h* если ровно *h* публикаций из данного множества получили больше *h* цитирований, тогда как все остальные получили не больше *h* цитат.

Зависит от определения цитируемых и цитирующих источников, а также от библиометрической базы, по которой рассчитывается.

В нашем исследовании цитируемые и цитирующие источники – статьи журналов, индексируемых WoS, вышедшие с 2007 по 2011.

Свойства индикаторов

	Database	*Year*	*Publication window, years*	*Weighted*	*Field-normalized*
2-year IF	WoS/JCR	2011	2	No	No
5-year IF	WoS/JCR	2011	5	No	No
immediacy index	WoS/JCR	2011	1	No	No
article influence	WoS/JCR	2011	5	Yes	No
h-index	WoS	2007–2011 (papers and citations)	5	No	No
SNIP	Scopus	2011	3	No	Yes
SJR	Scopus	2011	3	Yes	No

- Economics: 212 журналов
- Management: 93
- Political Science: 99

Ранговые корреляции

Доля инверсий, % (экономические журналы)

	impact factor	5-year impact factor	immediacy index	article influence	Hirsch index	SNIP	SJR
impact factor		8,46	24,59	18,13	15,45	15,09	14,23
5-year impact factor	8,46		24,25	13,72	13,15	13,66	12,20
immediacy index	24,59	24,25		26,00	25,57	27,01	25,25
article influence	18,13	13,72	26,00		17,15	16,31	15,50
Hirsch index	15,45	13,15	25,57	17,15		18,47	15,05
SNIP	15,09	13,66	27,01	16,31	18,47		17,28
SJR	14,23	12,20	25,25	15,50	15,05	17,28	

Коэффициент Кендалла τ_b (экономические журналы)

	impact factor	5-year impact factor	immediacy index	article influence	Hirsch index	SNIP	SJR
impact factor		0,830	0,503	0,637	0,654	0,698	0,700
5-year impact factor	0,830		0,510	0,725	0,702	0,726	0,741
immediacy index	0,503	0,510		0,475	0,442	0,454	0,472
article influence	0,637	0,725	0,475		0,620	0,673	0,674
Hirsch index	0,654	0,702	0,442	0,620		0,592	0,650
SNIP	0,698	0,726	0,454	0,673	0,592		0,638
SJR	0,700	0,741	0,472	0,674	0,650	0,638	

Ранговые корреляции

Доля инверсий, % (Российские экономические журналы)

	Муравьёв	НИУ ВШЭ	Балацкий	IF РИНЦ	Science Index	IF5 РИНЦ
Муравьёв (2012)		10,8	26,1	29,2	27,8	25,3
НИУ ВШЭ (2015)	10,8		12,3	17,7	16,8	15,6
Балацкий (2015)	26,1	12,3		35,6	29,0	32,1
IF РИНЦ	29,2	17,7	35,6		28,2	11,4
Science Index	27,8	16,8	29,0	28,2		23,1
IF5 РИНЦ	25,3	15,6	32,1	11,4	23,1	

Ранговые корреляции

Коэффициент Кендалла τ_b (Российские экономические журналы)

	Муравьёв	НИУ ВШЭ	Балацкий	IF РИНЦ	Science Index	IF5 РИНЦ
Муравьёв (2012)		0,270	0,169	0,157	0,193	0,249
НИУ ВШЭ (2015)	0,270		0,308	0,185	0,212	0,244
Балацкий (2015)	0,169	0,308		0,191	0,334	0,265
IF РИНЦ	0,157	0,185	0,191		0,431	0,770
Science Index	0,193	0,212	0,334	0,431		0,533
IF5 РИНЦ	0,249	0,244	0,265	0,770	0,533	

Коллективный выбор

X – the *general set* of alternatives

A – the *feasible set* of alternatives: $A \subseteq X \wedge A \neq \varnothing$. The feasible set is a variable.

N – the *society* (a group of voters or a panel of experts)

$u_i(x)$ – the *utility* of alternative $x \in X$ for voter $i \in N$, $u_i(x): X \rightarrow \mathbb{R}$

$$u_i(y) > u_i(x) \Leftrightarrow \text{voter } i \text{ strictly prefers } y \text{ to } x$$

$U = \{ u_i(x) \mid i \in N \}$ – the profile of utility functions

R – (*weak*) *social preferences*, $R \subseteq X \times X$

R is presumed to be complete: $\forall\, x \in X, \forall\, y \in X, (x, y) \in R \vee (y, x) \in R$

P – *strict social preferences*, $P \subseteq R$: $(x, y) \in P \Leftrightarrow ((x, y) \in R \wedge (y, x) \notin R)$

It is presumed that $R = R(P)$ and $P = P(U)$.

Правило большинства

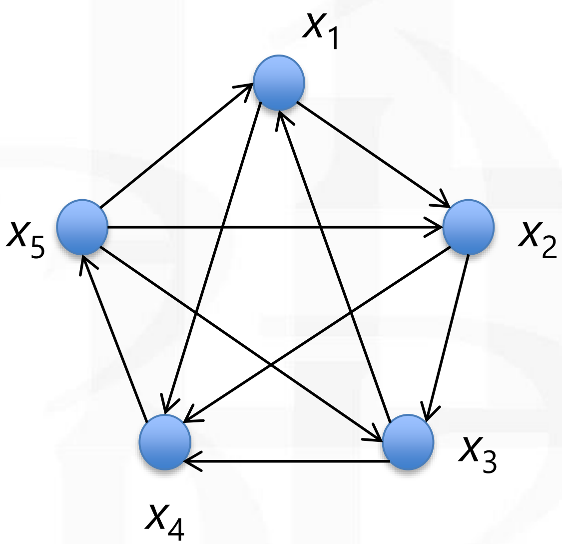
P – majority preference relation

$$(x, y) \in P \Leftrightarrow \#\{ i \in N \mid u_i(x) > u_i(y) \} > \#\{ i \in N \mid u_i(y) > u_i(x) \}$$

$\mathbf{M}=[m_{ij}]$ – matrix representing P $\quad m_{xy}=1 \Leftrightarrow (x, y) \in P,\ m_{xy}=0 \Leftrightarrow (x, y) \notin P$

	x_1	x_2	x_3	x_4	x_5
x_1	0	1	0	1	0
x_2	0	0	1	1	0
x_3	1	0	0	1	0
x_4	0	0	0	0	1
x_5	1	1	1	0	0

Majority matrix **M**



Majority digraph

Почему именно правило большинства? Аксиоматическое обоснование

Majority rule $R(P)$ uniquely satisfies the set of natural conditions (May 1952).

- ***Full domain***: the rule can be applied in all cases, that is, to any utility profile U
- ***Neutrality***: the rule treats all alternatives equally
- ***Anonymity***: the rule treats all voters (in our case, indicators) equally
- ***Pareto principle***: if x Pareto-dominates y, then xPy
- ***Monotonicity***: if utility profiles U and U' are such that

  $\forall i \in N, u'_i(x) \geq u_i(x) \wedge u'_i(y) = u_i(y)$, then $xP(U)y \Rightarrow xP(U')y$ and $xR(U)y \Rightarrow xR(U')y$
- ***Positive responsiveness***: if utility profiles U and U' are such that

  $\exists j \in N$: $(u_j(x) < u_i(y) \wedge u'_j(x) \geq u'_j(y)) \vee (u_j(x) = u_j(y) \wedge u'_j(x) > u'_j(y))$ and

  $\forall i \in N \setminus \{j\}, u'_i(x) = u_i(x) \wedge u'_i(y) = u_i(y)$ and $xR(U)y$ and $yR(U)x$ then $xP(U')y$
- ***Independence of irrelevant utilities***: $\forall A \subseteq X, P(U)|_A = P(U|_A)$
- ***Ordinality***

Ординальность или кардинальность?

Ordinality

If utility profiles U and U' are such that $\forall x \in X$, $\forall i \in N$, $u'_i(x) = g_i(u_i(x))$, where all g_i are strictly increasing real-valued functions of a real variable, then $P(U)=P(U')$.

Why ordinality? It removes the problem of non-comparability of individual utilities.

Individual utilities can be incomparable. Roughly speaking, we may not know the utility substitution rates; consequently, if the utility of person *i* decreases, we are unable to keep the social welfare constant by increasing the utility of person *j*.

Cardinal procedures are over-demanding from the informational point of view.

Their application may lead to meaningless results.

Почему именно правило большинства? Эпистемическое обоснование

If individual preferences are not subjective tastes but rather objective judgments concerning the state of affairs Q, the Condorcet Jury Theorem applies.

The Condorcet jury theorem (Condorcet 1785)

- If a binary judgment of each voter is more likely to be correct than otherwise, that is, if the conditional probability $p(xQy \mid u_i(x) > u_i(y))$ is greater than 0.5,
- and if judgments of different individuals are statistically independent,

then the judgment xPy obtained by the majority rule is likely to be true with the probability higher than that of any individual judgment:

$$\forall\ i \in N,\ p(xQy \mid xPy) > p(xQy \mid u_i(x) > u_i(y)).$$

Moreover, the probability $p(xQy \mid xPy)$ tends to 1 with number of voters $|N|$ increasing.

Парадокс Кондорсе

But the majority rule violates the axiom *Transitivity*, since the majority relation P may contain cycles. This result is known as the Condorcet paradox (Condorcet 1785).

In order to evaluate how nontransitive the majority relation is in our case, we calculate the number of 3-step, 4-step and 5-step P-cycles for three sets of journals.

Numbers of 3-, 4- and 5-step *P*-cycles for three sets of journals

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

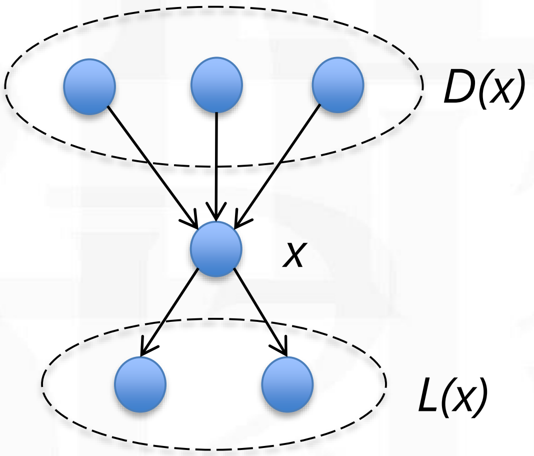
Правило Коупланда (Copeland 1951)

In order to get transitivity, which we need since we need a ranking, we sacrifice the independence of irrelevant utilities but keep the ordinality.

<u>The idea is to mend the majority relation, when it is nontransitive.</u>

The Copeland rule: when $|X| < \infty$, rank the alternatives by their score $s(x)$, determined in either of the following ways:

- Version 1. $s_1(x) = |\{y \in X \mid xPy\}| - |\{y \in X \mid yPx\}|$
- Version 2. $s_2(x) = |\{y \in X \mid xPy\}|$
- Version 3. $s_3(x) = |X| - |\{y \in X \mid yPx\}|$



Турнирные решения

A *social choice rule* is a correspondence S with arguments A and P and values in the set of subsets of A.

It is presumed that S depends on A and P only through restriction of P on A: $S=S(A, P)=S(P|_A) \subseteq A$, i.e. social choices are dependent on social preferences for available alternatives only.

A *tournament solution* is a social choice rule S that has the following properties:

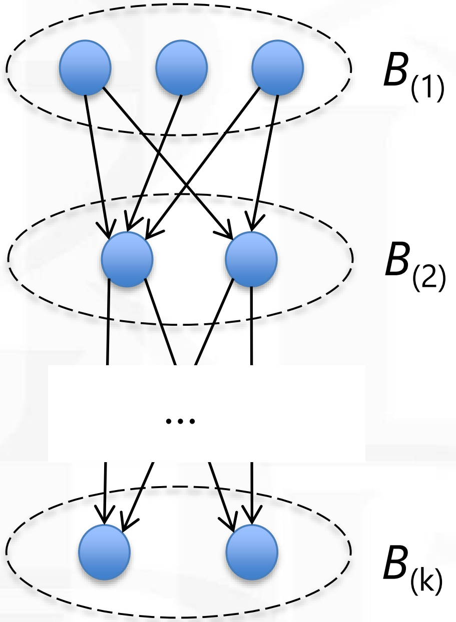
1. *Nonemptiness*: $\forall A, \forall P, \ S(P|_A) \neq \varnothing$;
2. *Neutrality*: permutation of alternatives' names and social choice commute;
3. *Condorcet consistency*:

if there is the Condorcet winner (P-maximal element) w for $P|_A$ then $S(P|_A) =\{w\}$.

Сортировка с помощью турнирного решения

Let us consider the following sorting procedure:

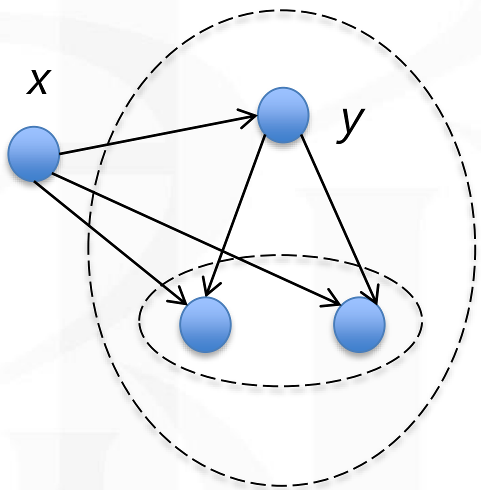
- Tournament solution S determines the set $B_{(1)}$ of social optima in A, $B_{(1)}=S(A)$.
- Let us exclude them and repeat the procedure for the set $A \setminus B_{(1)}$. The set $B_{(2)}=S(A \setminus B_{(1)})=S(A \setminus S(A))$ contains second best choices.
- After a finite number of selections and exclusions, all alternatives from A will be separated by classes $B_{(k)}=S(A \setminus (B_{(k-1)} \cup B_{(k-2)} \cup \ldots \cup B_{(2)} \cup B_{(1)}))$ according to their "quality", and these classes constitute a ranking.



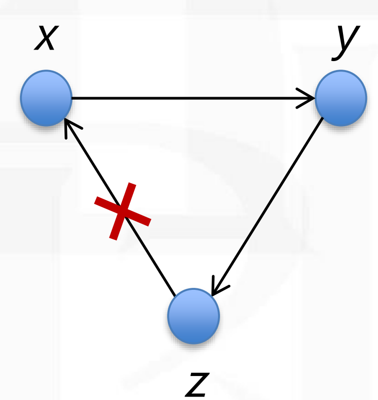
The uncovered set (Fishburn 1977, Miller 1980)

An alternative x **covers** an alternative y, if x is strictly more preferable (socially) than y, and all the alternatives, which are strictly less preferable than y, are also strictly less preferable than x:

$$xPy \wedge \forall z \in X, yPz \Rightarrow xPz.$$



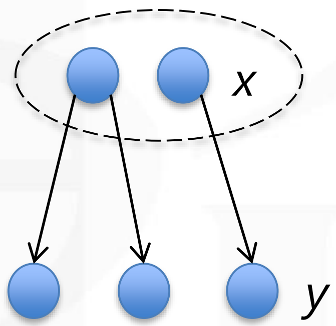
The best alternatives according to this solution concept are those that are not covered by any other alternative. The set of such alternatives is called the *uncovered set UC* (Fishburn 1977, Miller 1980).



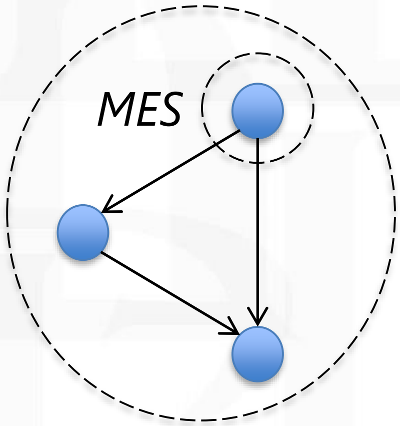
The minimal externally stable set

A nonempty subset *B* of a feasible set *A* is *externally stable* (von Neumann & Morgenstern 1944), if for any alternative *y* from *A* and outside *B*, there is an alternative *x* in *B*, that is strictly more preferable (socially) than *y*:

$$\forall y \in A \backslash B, \exists\, x \in B: xPy.$$

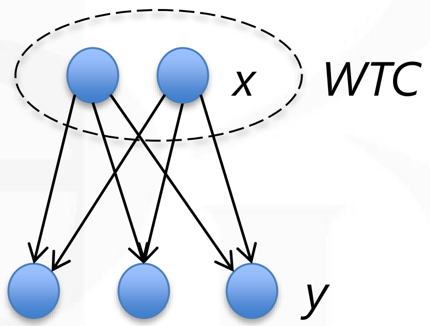


Externally stable set is called *minimal*, if none of its proper nonempty subsets is externally stable. The alternative is considered as "good" if it belongs at least to one minimal externally stable set. Thus the solution concept is a union of all such sets *MES* (Wuffl et al. 1989, Aleskerov & Kurbanov 1999, Subochev 2008, Aleskerov & Subochev 2013).



The weak top cycle (Ward 1961, Schwartz 1970, 1972, Smith 1973)

A set D, $D \subseteq A$, is called dominant in A, if every alternative from D is strictly more preferable (socially) than every alternative from $A \setminus D$: $\forall x \in D$, $\forall y \in A \setminus D$, xPy.



The weak top cycle WTC is a minimal dominant set.

Марковское ранжирование (Daniels 1969, Ушаков 1971)

First, sort the alternatives by *WTC*. Then, consider a set B of all the alternatives of a given sort. Imagine that a digraph representing $P|_B$ is a labyrinth: vertices are rooms, arcs are one-way passages between the rooms. Time is discrete. A visitor in a certain moment of time k is in a certain room x. Then at random with equal probability another vertex $y \in B\backslash\{x\}$ is chosen. If y is at least as preferable as x (yRx), then the visitor moves to y.

Let us denote alternatives in B by numbers. Let $\mathbf{p}^{(k)}$ – the vector, its component $p_x^{(k)}$ is probability that the visitor is in a room number x at a time moment k. Let us consider the vector $\mathbf{p}=\lim_{k\to\infty} \mathbf{p}^{(k)}$. Its value does not depend on initial conditions (i.e. on the value of $\mathbf{p}^{(0)}$). The greater is the number of visits to a room number x, the better is the corresponding alternative x. The relative number of visits to a room number x over infinite time period is proportional p_x, so we rank alternatives by this value.

Коэффициент Кендалла τ_b (экономические журналы)

	impact factor	5-year impact factor	immediacy index	article influence	Hirsch index	SNIP	SJR	Copeland (2)	Copeland (3)	*UC*	*MES*	Marcovian
impact factor		0,830	0,503	0,637	0,654	0,698	0,700	0,834	0,831	0,834	0,835	0,819
5-year impact factor	0,830		0,510	0,725	0,702	0,726	0,741	0,903	0,904	0,906	0,896	0,891
immediacy index	0,503	0,510		0,475	0,442	0,454	0,472	0,550	0,551	0,556	0,578	0,560
article influence	0,637	0,725	0,475		0,620	0,673	0,674	0,766	0,769	0,777	0,785	0,769
Hirsch index	0,654	0,702	0,442	0,620		0,592	0,650	0,738	0,737	0,737	0,747	0,729
SNIP	0,698	0,726	0,454	0,673	0,592		0,638	0,759	0,759	0,767	0,775	0,750
SJR	0,700	0,741	0,472	0,674	0,650	0,638		0,792	0,790	0,800	0,797	0,775
Copeland (2)	0,834	0,903	0,550	0,766	0,738	0,759	0,792		0,990	0,970	0,950	0,956
Copeland (3)	0,831	0,904	0,551	0,769	0,737	0,759	0,790	0,990		0,969	0,950	0,959
UC	0,834	0,906	0,556	0,777	0,737	0,767	0,800	0,970	0,969		0,955	0,954
MES	0,835	0,896	0,578	0,785	0,747	0,775	0,797	0,950	0,950	0,955		0,949
Markovian	0,819	0,891	0,560	0,769	0,729	0,750	0,775	0,956	0,959	0,954	0,949	

Ранговые корреляции (продолжение)

Коэффициент Кендалла τ_b (Российские экономические журналы)

	Муравьёв	НИУ ВШЭ	Балацкий	IF РИНЦ	Science Index	IF5 РИНЦ	Pareto	Core	*UC*	*MES*	Copeland (1)	Copeland (2)	Copeland (3)
Муравьёв		0,270	0,169	0,157	0,193	0,249	0,590	0,471	0,602	0,629	0,477	0,432	0,529
НИУ ВШЭ	0,270		0,308	0,185	0,212	0,244	0,596	0,545	0,568	0,554	0,528	0,474	0,547
Балацкий	0,169	0,308		0,191	0,334	0,265	0,465	0,637	0,500	0,489	0,628	0,731	0,571
IF РИНЦ	0,157	0,185	0,191		0,431	0,770	0,162	0,211	0,161	0,169	0,217	0,241	0,207
Science Index	0,193	0,212	0,334	0,431		0,533	0,222	0,291	0,246	0,250	0,302	0,354	0,275
IF5 РИНЦ	0,249	0,244	0,265	0,770	0,533		0,234	0,271	0,238	0,247	0,286	0,323	0,273
Pareto	0,590	0,596	0,465	0,162	0,222	0,234		0,810	0,950	0,954	0,813	0,710	0,887
Core	0,471	0,545	0,637	0,211	0,291	0,271	0,810		0,830	0,822	0,881	0,786	0,925
UC	0,602	0,568	0,500	0,161	0,246	0,238	0,950	0,830		0,978	0,829	0,751	0,892
MES	0,629	0,554	0,489	0,169	0,250	0,247	0,954	0,822	0,978		0,825	0,743	0,891
Copeland (1)	0,477	0,528	0,628	0,217	0,302	0,286	0,813	0,881	0,829	0,825		0,899	0,918
Copeland (2)	0,432	0,474	0,731	0,241	0,354	0,323	0,710	0,786	0,751	0,743	0,899		0,812
Copeland (3)	0,529	0,547	0,571	0,207	0,275	0,273	0,887	0,925	0,892	0,891	0,918	0,812	

Ранжирование ранжирований

<table>
<tr><th>ранг</th><th>Economics</th><th>Management</th><th>Political Science</th><th>Management (2008)</th></tr>
<tr><td>1</td><td>MES</td><td>MES</td><td>MES</td><td>UC</td></tr>
<tr><td>2</td><td>UC</td><td>UC</td><td>UC</td><td>MES</td></tr>
<tr><td>3</td><td>Copeland 3</td><td>Copeland 2</td><td>Copeland 3</td><td>Copeland 3</td></tr>
<tr><td>4</td><td>Copeland 2</td><td>Copeland 3</td><td>Copeland 2</td><td>Copeland 2</td></tr>
<tr><td>5</td><td>Markovian</td><td>Markovian</td><td>Markovian</td><td>Markovian</td></tr>
<tr><td>6</td><td>5-y.impact</td><td>5-y.impact</td><td>5-y.impact</td><td>impact</td></tr>
<tr><td>7</td><td>impact</td><td>SNIP</td><td>Hirsch</td><td>5-y.impact</td></tr>
<tr><td>8</td><td>SJR</td><td>Hirsch</td><td rowspan="3">AI/
impact/
SJR</td><td>SJR</td></tr>
<tr><td>9</td><td>AI</td><td>AI</td><td rowspan="3">AI/
Hirsch/
SNIP</td></tr>
<tr><td>10</td><td>SNIP</td><td>SJR</td></tr>
<tr><td>11</td><td>Hirsch</td><td>impact</td><td>SNIP</td></tr>
<tr><td>12</td><td>immediacy</td><td>immediacy</td><td>immediacy</td><td>immediacy</td></tr>
</table>

Публикации

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

Субочев А.Н. Насколько различны существующие рейтинги российских журналов по экономике и менеджменту и как их объединить. *Журнал НЭА*. 2016. №2 (30). С. 181-192.

Aleskerov, F., Pislyakov, V., Subochev, A. 2014. Ranking Journals in Economics, Management and Political Science by Social Choice Theory Methods. WP BRP 27/STI/2014. Moscow: HSE.

Aleskerov, F., Subochev, A. 2013. Modeling optimal social choice: matrix-vector representation of various solution concepts based on majority rule. *Journal of Global Optimization*, V. 56, Iss. 2. P. 737-756.

Алескеров Ф.Т., Писляков В.В., Субочев А.Н., Чистяков А.Г. 2011. Построение рейтингов журналов по менеджменту с помощью методов теории коллективного выбора**:** препринт WP7/2011/04. НИУ ВШЭ. – М.: Издательский дом Высшей школы экономики.

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Спасибо за внимание!

ul. Myasnitskaya, 20
Moscow, Russia, 101000
Phone: (495) 621-7983, Fax: (495) 628-7931
www.hse.ru