



Universiteit
Leiden
The Netherlands

A refined approach to the dynamics of rankings: taking ranking differences into account

Garcia-Zorita, C.; Rousseau, R.; Marugan-Lazaro, S.; Sanz-Casado, E.

Citation

Garcia-Zorita, C., Rousseau, R., Marugan-Lazaro, S., & Sanz-Casado, E. (2018). A refined approach to the dynamics of rankings: taking ranking differences into account. *Sti 2018 Conference Proceedings*, 1058-1064. Retrieved from <https://hdl.handle.net/1887/65251>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/65251>

Note: To cite this publication please use the final published version (if applicable).



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

© of the text: the authors

© 2018 Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands



This ARTICLE is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

A refined approach to the dynamics of rankings: taking ranking differences into account

Carlos Garcia-Zorita ^{*a,b}, Ronald Rousseau ^{**c,d}, Sergio Marugan-Lazaro ^{***a} and Elias Sanz-Casado ^{****a,b}

* czorita@bib.uc3m.es

^aLaboratory of Metric Studies on Information (LEMI). Department of Library and Information Science. Carlos III University of Madrid. C/Madrid 126, Getafe, 28903 Madrid, Spain.

^bResearch Institute for Higher Education and Science (INAECU). Carlos III University of Madrid-Autonomous University of Madrid. C/Madrid 126, Getafe, 28903 Madrid, Spain.

** ronald.rousseau@uantwerpen.be; ronald.rousseau@kuleuven.be

^cUniversity of Antwerp, Faculty of Social Sciences, B-2020 Antwerpen, Belgium.

^dKU Leuven, Facultair Onderzoekscentrum ECOOM, Naamssestraat 61, Leuven B-3000, Belgium.

*** smarugan@pa.uc3m.es

^aLaboratory of Metric Studies on Information (LEMI). Department of Library and Information Science. Carlos III University of Madrid. C/Madrid 126, Getafe, 28903 Madrid, Spain.

**** elias.sanz@uc3m.es

^aLaboratory of Metric Studies on Information (LEMI). Department of Library and Information Science. Carlos III University of Madrid. C/Madrid 126, Getafe, 28903 Madrid, Spain.

^bResearch Institute for Higher Education and Science (INAECU). Carlos III University of Madrid-Autonomous University of Madrid. C/Madrid 126, Getafe, 28903 Madrid, Spain.

Abstract

The notion of ‘ranking dynamics’ was established in a previous work (García-Zorita et al. Rankings: competitiveness versus stability, *STI International conference*. Paris, 2017) where we include the study of rankings with ties, entrants and leavers. In this contribution, we generalize our method by taking the absolute difference in position between consecutive rankings into account. We introduced the concept of Relative Ranking Volatility (RRV) to normalizing the absolute volatility of a ranking using a factor that is defined using an absolute score normalization. Finally, we show a real-example taken from the *Web of Science / Journal Impact Factor* for each IF-quartile distribution in the *Information Sciences & Library Sciences* category. In this real example, the lowest volatility (0.19) is obtained in Q1 journals.

Keywords

Ranking dynamics, relative ranking volatility, normalizing factor, ranking stability.

1. Introduction

Following Criado, García, Pedroche and Romance (2013) and Pedroche, Criado, García, Romance & Sánchez (2015) we introduced a method to study the dynamics of rankings (García-Zorita et al., 2017). Besides taking ties into account, which was also done in (Pedroche et al., 2015) we included leavers and entrants in the rankings. At any moment, some elements in the rankings are active, i.e. included in the ranking at that moment, or inactive, i.e. not included in the ranking at that moment. The inactive elements have been or will be active, or both.

Following Criado et al. (2013) our measure to study dynamics is based on pair-wise changes in position between consecutive rankings. In this contribution, we generalize our method by taking the absolute difference in position between consecutive rankings into account.

2. Notation

Similar to the Criado et al. (2013) framework we consider a set S of n elements or nodes (when described in a network context), denoted as $\{s_1, s_2, \dots, s_n\}$; hence $\#S = n$. Next we consider a strictly ordered column of r rankings, referred to as instances. Each instance is a ranked set of elements, where ties are allowed. By definition, tied elements have the same rank. We add to each instance those elements in S which are absent in this particular instance, ranking them as ties in the last position. The original elements in a given instance are called the active elements; the added ones are called the inactive elements. Being active or inactive is referred to as the state of an element in a given instance. In the notation for instances active and inactive elements are separated by the symbol “;”. Denoting a tie between elements e and f as $\dots, \underbrace{e, f}, \dots$ and assuming that elements x, y and z are missing in instance c , this means that

$c = (s_1, s_2, \dots, s_{n-3})$ is rewritten as $c = (s_1, s_2, \dots, s_{n-3}; \underbrace{x, y, z})$. The rank of an element s in instance c is denoted as $r_c(s)$. Tied elements are given the same – average – rank (hence ranks do not have to be natural numbers). Although this may rule may also be applied to the inactive elements we will not need it further on. We denote by $n_a(c)$ the number of active elements at instance c .

We say that element s_i changes position with element s_j if they exchange their relative positions between two consecutive rankings. Roughly speaking the more position shifts the more dynamic a ranking system,

3. Counting volatility

We obtain a volatility score for each element, e , by comparing with each other element. This score is obtained by a comparison of the positions of e and the other element in consecutive instances, leading to partial volatility scores. The sum of all these partial volatility scores resulting of a comparison of e with each other element, is the volatility score of element e . Partial volatility scores are symmetric: the partial score between elements e and f is the same as the one between f and e . In each comparison between instances the score either stays unchanged or increases by at least one and at most two (see further). The initial score is zero. When the state of (at least) one of the two elements changes from active (A) to inactive (I) or vice versa then the score increases by one; if both elements stay inactive then the score stays the same. If both elements stay active and their relative position stays the same, also then the score does not change. If the relation $<$ (between $r_c(e)$ and $r_c(f)$) changes to $>$ or vice versa, then the volatility score increases. This increase is equal to one plus $\frac{|r_c(e) - r_c(f)| + |r_{c+1}(e) - r_{c+1}(f)|}{MAX}$, where MAX is equal to $(n_a(c) - 1) + (n_a(c + 1) - 1)$. MAX is obtained if e and f are the unique first and last active elements in instances c and $c+1$, but with

their roles reversed. This means that the partial volatility score of two elements changes at most by 2.

If the two elements become tied then the score does not change, but the previous relative position is kept in memory. If the two elements were tied and are still tied, then nothing changes; if they were tied (instance c) and are not tied anymore and are both active (instance $c+1$) then the last time they were not tied (at instance c_0) determines if there is a position change and hence if the volatility score increases or not. If the score increases it increases by $1 + \frac{|r_{c_0}(e)-r_{c_0}(f)|+|r_{c+1}(e)-r_{c+1}(f)|}{MAX}$ where MAX is equal to $(n_a(c_0) - 1) + (n_a(c + 1) - 1)$.

Finally, if the two elements have always been tied and active since the first instance and they are not tied anymore (instance $c+1$) then this counts as a position shift and the volatility score increases by $1 + \frac{|r_{c+1}(e)-r_{c+1}(f)|}{MAX}$ where MAX is equal $(n_a(c + 1) - 1)$.

In the very special case that two elements are inactive since the beginning (hence tied), then become both active and tied (instance $c+1$), and then at a later instance ($c+k$, $k>1$) one is ranked before the other the volatility score is increased by $1 + \frac{|r_{c+k}(e)-r_{c+k}(f)|}{MAX}$ where MAX is equal to $(n_a(c + k) - 1)$. Finally, if they stay tied till the last instance, then their score is increased by 1 (as the result of becoming active).

If two elements were active and become inactive this implies that they are tied and their previous position is kept in memory (see further for a formal description of this). If one of the two becomes active, their score is raised by one and the memory is cleared as they are not tied anymore.

This scoring method has the following properties:

- 1) leaving or entering is taken into account
- 2) changing positions receives a strictly higher score than leaving or entering;
- 3) in case of a position change the relative position change is taken into account

Once the volatility of each element is determined we add all values leading to an absolute ranking volatility score. Finally we divide by a normalizing factor (see further) yielding the final relative ranking volatility score, (in short: RRV-score) which is a value between zero and one. It corresponds to the NS-value in our previous investigations.

4. A formal way of denoting changes and counting volatility

To describe formally what happens step by step we use the following 9-tuple, referred to as a dynamic ranking tuple. These tuples are the basis of a computer program to obtain volatility scores.

This 9-tuple consists of the following code elements: (element1, state of element1, element2, state of element2, instance, number of active element in this instance, relational symbol, counter, memory containing an additional information symbol).

We want to determine the volatility score of element1 and compare with element2; these elements are either active (A) or not (inactive, indicated by I); the fifth position of the tuple denotes the second instance used in the comparison between two instances, the sixth gives the number of active elements in this instance, the seventh shows the relation between element1 and element2 in that order: the relational symbol is either $<$, $>$ or $=$; the seventh position is the partial volatility score up to that instance and the ninth position is a memory symbol that helps to deal with equalities: it is either $*$ (meaning that element1 and element2 are not tied); $<(c_j)$ (meaning that element1 and element2 are tied and the last time they were not tied we had element1 $<$ element2; this happened at instance c_j ; or $>(c_j)$ referring to the opposite relation; finally $\#$ means that element1 and element2 have always (since c_1) been tied.

The value of the score at the last instance is the partial volatility score for the pair element1-element2.

5. An artificial example

Consider the four instances

$$c_1 = (s, t, u)$$

$$c_2 = (s, u, w, x)$$

$$c_3 = (\underbrace{s, u}, w, y, z)$$

$$c_4 = (z, y, u, \underbrace{s, t})$$

The set S is here $\{s,t,u,w,x,y,z\}$, with $\#S=7$. Consequently, the four instances are rewritten as:

$$c_1 = (s, t, u; \underbrace{w, x, y, z}); n_a(c_1) = 3$$

$$c_2 = (s, u, w, x; \underbrace{t, y, z}); n_a(c_2) = 4$$

$$c_3 = (\underbrace{s, u}, w, y, z; \underbrace{t, x}); n_a(c_3) = 5$$

$$c_4 = (z, y, u, \underbrace{s, t}; \underbrace{w, x}); n_a(c_4) = 5$$

We provide some examples of calculations in which we show each step.

Case $\{s,t\}$.

CODE 0: (s,A,t,A,c1,3,<,0,*) start

CODE 1: (s,A,t,I,c2,4,<,1,*) an increase by 1 because t becomes inactive

CODE 2: (s,A,t,I,c3,5,<,1,*) no volatility changes

CODE 3: (s,A,t, A, c4,5,=,2,*) an increase by 1 because t becomes active and tied with s (which has no influence here). The final score is 2.

Case $\{s,u\}$.

CODE 0: (s,A,u,A,c1,3,<,0,*) start

CODE 1: (s,A,u,A,c2,4,<,0,*) no relative changes

CODE 2: (s,A,u,A, c3,5,=,0, <(c2)) s and u are tied and we had, at instance c2, s < t

CODE 3: (s,A,u, A, c4, 5, >, 1+(1+1.5)/(3+4)=19/14, *) s and u have changed positions with respect to the situation in c2; their differences in position were 1 in c2 and 1.5 in c4; there were 4 active elements in c2 and 5 in c4. This yields a final score of 19/14.

Case $\{t,y\}$.

CODE 0: (t, A, y, I, c1, 3, <, 0, *) start

CODE 1: (t, I, y, I, c2, 4, =, 1, *) t becomes inactive

CODE 2: (t, I, y, A, c3, 5, >, 2, *) y becomes active

CODE 3: (t, A, y, A, c4, 5, >, 3, *) t becomes active. The final score is 3.

Table 1. Details of the volatility calculations for the artificial example

elements	c1-c2	c2-c3	c3-c4	total	elements	c1-c2	c2-c3	c3-c4	total
s-t	1	0	1	2	u-w	1	0	1	2
s-u	0	0	19/14	19/14	u-x	1	1	0	2
s-w	1	0	1	2	u-y	0	1	23/16	39/16
s-x	1	1	0	2	u-z	0	1	27/16	43/16
s-y	0	1	13/8	21/8	w-x	1	1	1	3
s-z	0	1	15/8	23/8	w-y	1	1	1	3

t-u	1	0	1	2		w-z	1	1	1	3
t-w	1	0	1	2		x-y	1	1	0	2
t-x	1	1	1	3		x-z	1	1	0	2
t-y	1	1	1	3		y-z	0	1	5/4	9/4
t-z	1	1	1	3						

The results shown in Table 1 lead to the volatility scores shown in Table 2. Element u has the lowest volatility as a result of being always among the first three. Element z has the highest volatility: it became active, was once last and once first. Element y has a similar history but moved up less than z, explaining the small difference in volatility.

Table 2. Volatility of the elements in S and the absolute ranking volatility (sum)

Elements	s	t	u	w	x	y	z	sum
Volatility	12.86	15	12.48	15	14	15.31	15.81	100.46

In Figure 1, we show the heatmap of the partial scores matrix between the elements in the theoretical example. The colour of the map ranges from dark blue (less volatility) to red (more volatility).

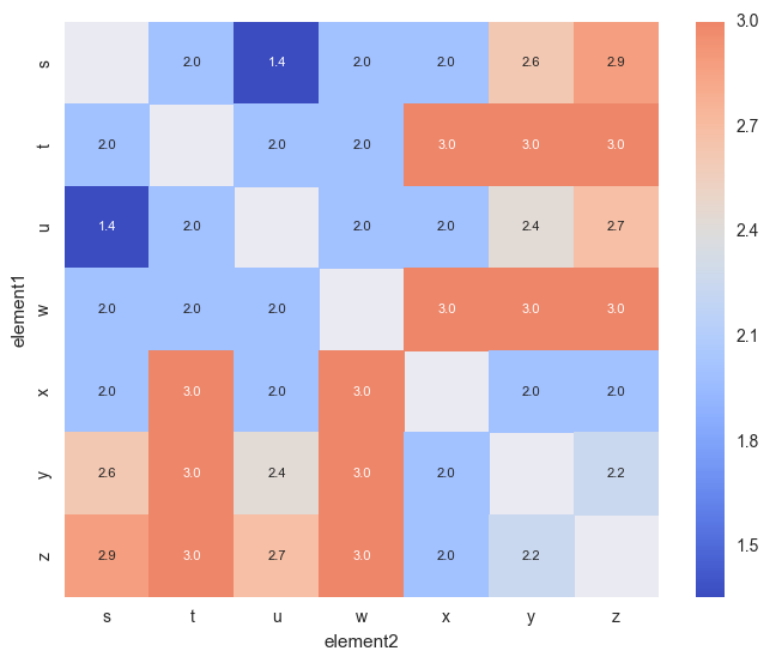


Fig. 1. Heatmap of the partial scores between elements in S

5. Absolute score normalization

How to normalize the results shown in Table 2? In the previous article we took the sum and divided by $n(n-1)(r-1)$. Here the situation is more complicated.

Assume that $c = (e_1, e_2, e_3, \dots, e_n)$ and $c+1 = (e_n, e_{n-1}, \dots, e_2, e_1)$. This comparison yields the largest possible sum of volatilities for one comparison. Next we calculate its absolute score.

First we have is $n(n-1)$, corresponding to the number of pairs of elements, each counted twice. To this we add twice the following scores:

$$\left[\frac{1}{n-1}(n-1) + \frac{2}{n-1}(n-2) + \frac{3}{n-1}(n-3) + \dots + \frac{n-1}{n-1}(n-(n-1)) \right].$$

This yields:

$$\frac{1}{n-1} [\sum_i^{n-1} i(n-i)] = \frac{1}{n-1} [n \sum_i^{n-1} i - \sum_i^{n-1} i^2].$$

The part between square brackets is:

$$n \frac{(n-1)n}{2} - \frac{(n-1) \cdot n \cdot (2n-1)}{6} = \frac{(n-1) \cdot n \cdot (n+1)}{6}.$$

Bringing all this together yields:

$$n(n-1) + \frac{2}{n-1} \cdot \frac{(n-1) \cdot n \cdot (n+1)}{6} = \frac{2n(2n-1)}{3}.$$

This maximum may be attained for each comparison, namely if all elements are always active and rankings are reversed all the time. Hence the sum of all volatilities must be divided by

$$\frac{2n(2n-1)}{3} \cdot (r-1). \tag{1}$$

Normalizing using the normalizing factor (1) leads to a relative ranking volatility score (RRV-score) of $3 \cdot (100.46) / (14 \cdot 13 \cdot 3) = 0.552$

6. A real-world example

As an illustrative real-world example, we calculated the ranking dynamics of the WOS-JIF (Web of Science - Journal Impact Factor) for the SSCI/JCR (Social Sciences Citation Index/Journal Citation Reports) class of *Information Sciences & Library Sciences*. We consider the period 1997-2016 ($r=20$). We have calculated the volatility of all journals in the IS&LS category ($n=117$) leading to an absolute ranking volatility of 40,110.48 (The normalizing factor (1) was 336,640.69). Hence, the relative ranking volatility is 0.1191 or 11.91%.

We have moreover calculated the volatility of each of the four quartiles of the JIF ranking of the IS&LS category. The results are shown in Table 3.

Table 3: IS&LS category: Volatility for each quartile

r=20	Q1	Q2	Q3	Q4
n(*)	50	71	64	66
Absolute Volatility	11,915.84	35,515.40	33,876.13	26,956.52
Normalizing factor	62,700.00	126,806.00	102,954.67	109,516.00
Relative Ranking Volatility (RRV)	0.1900	0.2801	0.3290	0.2461

(*) $\sum_{n \neq 117}$ (Same journal can be in several categories and several quartiles)

The dynamics in the fourth quartile is less than in the third quartile that has the greater volatility, while the least volatility occurs in Q1 where the position shifts between journals are less than in the others quartiles.

7. Conclusions

Previous investigations about the dynamics of rankings, based on pairwise changes of elements, are refined by taking the actual difference in rankings into account, leading to a refined RRV-score. We recall that one minus the relative ranking volatility score (1-RRV) is a measure for stability.

8. References

- Criado, R., Garcia, E., Pedroche, F. & Romance, M. (2013). A new method for comparing rankings through complex networks: Model and analysis of competitiveness of major European soccer leagues. *Chaos*, 23, 043114.
- García-Zorita, C., Rousseau, R., Marugan-Lazaro, S. & Sanz-Casado, E. (2017). Rankings: competitiveness versus stability. In: *Open indicators: innovation, participation and actor-based STI indicators*. STI 2017, Paris.
- Pedroche, F., Criado, R., García, E.H., Romance, M. & Sánchez, V.E. (2015). Comparing series of rankings with ties by using complex networks: An analysis of the Spanish stock market (IBEX-35 index). *Networks and Heterogeneous Media*, 10(1), 101-125. doi:10.3934/nhm.2015.10.101.