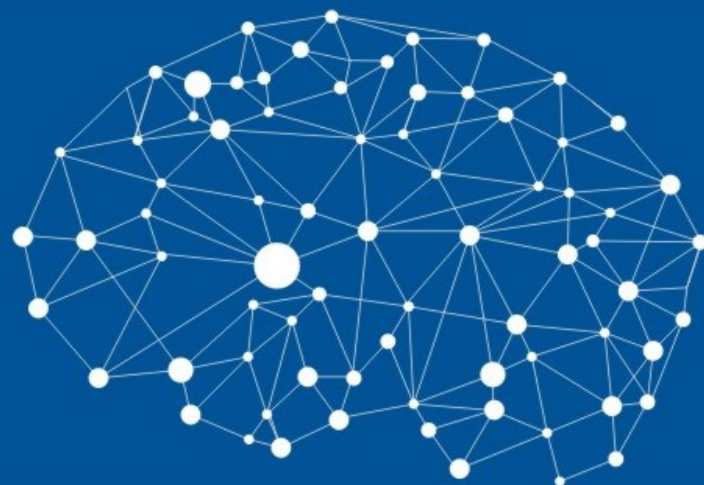




# COSSENO DE SALTON, ÍNDICE DE JACCARD E CORRELAÇÃO DE PEARSON: comparando índices normalizados e absolutos em análise de cocitação de autores

Bruno Henrique Alves<sup>1</sup>; Ely Francina Tannuri Oliveira<sup>2</sup>

ALVES, B. H.; OLIVEIRA, E. F. T.. COSSENO DE SALTON, ÍNDICE DE JACCARD E CORRELAÇÃO DE PEARSON: comparando índices normalizados e absolutos em análise de cocitação de autores In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 5., 2016, São Paulo. **Anais...** São Paulo: USP, 2016. p. A14

<sup>1, 2</sup> Universidade Estadual Paulista - UNESP- FFC - Marília

# **COSSENO DE SALTON, ÍNDICE DE JACCARD E CORRELAÇÃO DE PEARSON: comparando índices normalizados e absolutos em análise de cocitação de autores**

Eixo temático: Métodos, Técnicas e Ferramentas para Estudos Bibliométricos e Cientométricos

Modalidade: Apresentação oral

## **1 INTRODUÇÃO**

Na contribuição dos estudos de citação e cocitação, destacam-se os estudos de Small (1973), White e Griffith (1981), entre outros. Segundo estes autores, as citações indicam os paradigmas das comunidades formadas, seus procedimentos metodológicos, os grupos de cientistas, suas publicações e evidenciam os pesquisadores de impacto de uma área.

Associada à análise de citações, a Análise de Cocitações de Autores (ACA) foi aprofundada por White e Griffith, em 1981. Cocitação é definida por Small (1973, p. 265) “como a frequência com que dois documentos citados estão juntos em um artigo”. O número de vezes que os autores ou documentos foram citados juntos determina a força da cocitação entre eles, bem como o núcleo da literatura da área. A frequência de cocitação entre dois autores determina como a estrutura de conhecimento da área é percebida pelos pesquisadores.

Os estudos de coautoria e cocitação são, em geral, realizados a partir da construção de matrizes de valores absolutos. No entanto, quando se pretende realizar estudos comparativos, dadas as especificidades e peculiaridades de cada área, destaca-se a relevância dos indicadores normalizados, que possibilitam avaliações comparativas, uma vez que padronizam as unidades de medida, além de revelar aspectos não explicitados nos dados brutos, obtidos nas matrizes com valores absolutos (GLÄNZEL et al., 2009), tais como a intensidade e proximidade de relações entre os autores cocitados.

Segundo os estudos de Luukkonen et al. (1993), os índices absolutos e normalizados trazem tipos diferentes de informações: os absolutos mostram os autores centrais ou os mais periféricos nas redes, enquanto os índices normalizados mostram a intensidade das relações de colaboração ou de cocitação entre os pares de pesquisadores cocitados.

Apresentam-se vários procedimentos para normalização dessas medidas, tais como: Cosseno de Salton ( $C_s$ ), Índice de Jaccard (IJ) e Coeficiente de Correlação de Pearson, denotado por ( $r$ ) de Pearson, indicadores estes que são o foco dessa pesquisa, utilizados aqui para análise de cocitações de autores. Há outros procedimentos matemáticos e estatísticos para normalização, que não são abordados no âmbito deste trabalho.

Considerando o exposto, objetiva-se nesta pesquisa realizar um estudo comparativo entre os indicadores absolutos e os normalizados, a saber: Cosseno de Salton ( $C_s$ ), Índice de Jaccard (IJ) e Correlação de Pearson ( $r$ ) e para Análise de Cocitação de Autores. De forma mais específica, propõe-se apresentar os três índices normalizados, apontar as diferenças no uso entre os indicadores absolutos e os normalizados, relativos ao  $C_s$ , IJ e  $r$  de Pearson, analisar algumas questões sobre a escolha desses diferentes procedimentos e apresentar um estudo teórico-aplicado desses indicadores, comparando resultados extremados.

Esta pesquisa se justifica, considerando as especificidades das diferentes áreas, relativas aos diferentes comportamentos no que se refere às citações. Ela se propõe a analisar alguns procedimentos e técnicas para estudos relativos à normalização de cocitações, uma vez que esta padroniza as unidades de medida nas matrizes de cocitação de autores. Ainda, por evidenciar algumas questões existentes no âmbito da comunidade científica relativas ao procedimento mais adequado a ser utilizado nas diferentes situações.

## **2 NORMALIZAÇÃO DOS INDICADORES ABSOLUTOS E DISCUSSÃO**

Os estudos de cocitação de autores utilizam matrizes de proximidade, na medida em que deixam explícita a intensidade da proximidade ou o maior distanciamento entre os pares de autores cocitados, sob a ótica dos autores citantes. Nas matrizes de proximidade, quanto maior for o valor do indicador de cocitação, mais similares e próximos são os dois autores cocitados, nos aspectos teóricos e/ou metodológico (LEYDERSDOFF; VAUGHAM, 2006). Assim, todas as matrizes de frequência absoluta ou normalizada de cocitação são chamadas de similaridade.

O  $C_s$  é um índice normalizado, calculado a partir da matriz de coocorrência dos dados absolutos, tanto para a coautoria como para cocitações, conforme Luukkonen et al. (1993). É expresso pela frequência de coautorias ou cocitações de dois autores  $i, j$ . A fórmula do  $C_s$

presente nos estudos de Hamers et al. (1989), na qual as coocorrências representam cocitações, pode ser assim expressa:

$$C_s(i,j) = \frac{coc(i,j)}{(cit_{(i)}, cit_{(j)})^{1/2}}$$

Onde,

$coc(i,j)$  = total de cocitações entre o autor i e j

$cit_{(i)}$  = total de citações recebidas pelo autor i

$cit_{(j)}$  = total de citações recebidas pelo autor j

Small (1973) foi o primeiro a apontar o IJ como índice de mormalização, utilizado somente para cocitações. Mede a intensidade da relação entre dois autores, revistas ou documentos, e é definido como a frequência de cocitação de dois autores A e B, chamados aqui de Cocit (A, B). É compreendido como o número de vezes que os autores A e B são citados conjuntamente, representados por Cocit (A, B), dividido pelo número total de citações recebidas pelos dois autores A e B, representadas conceitualmente por  $Cit(A) + Cit(B) - Cit(A \cap B)$ . A fórmula de IJ, segundo Luukkonen et al. (1993), pode ser expressa por:

$$IJ = \frac{Cocit(A, B)}{Cit(A) + Cit(B) - Cit(A \cap B)}$$

Tanto o  $C_s$  como IJ apresentam valores variando entre zero e um: quanto mais próximo de um estiver o valor desses indicadores, mais similares são os dois autores; quanto mais próximo de zero, mais distante (menos intensa) é a associação dos dois autores, quer seja como coautores ou na percepção do conjunto de autores citantes. Valor zero para  $C_s$  ou IJ indica ausência de coautorias ou cocitações entre os dois autores.

A correlação mede a associação de pares de autores cocitados, considerando o conjunto de valores de cocitações de cada um deles, com os demais em estudo. O  $r$  de Pearson foi a medida padrão na ACA ( $-1 \leq r \leq +1$ ) e, nesse intervalo, a correlação varia de sentido (de negativo para positivo) e de força (fraca, moderada ou forte).

Considerando  $x_i$  e  $y_i$ , as frequências de cocitações de um autor X e de um autor Y com os demais autores, para i variando de 1 a n, com n igual a quantidade de autores em estudo, o  $r$  de Pearson é definido por:

$$r = \frac{\sum x_i \cdot y_i - \frac{\sum x_i \cdot \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right) \left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}}$$

A escolha do uso do índice de similaridade para a normalização  $C_s$  ou  $IJ$  ou  $r$  de Pearson, no caso da Análise de Cocitação de autores (ACA), é controversa. Em 2003, Ahlgren et al., criticaram o uso de  $r$  de Pearson, mostrando que o mesmo não satisfaz isoladamente como medida de similaridade. Os autores citados fornecem argumentos para o uso do  $C_s$ , em vez de  $r$  de Pearson, especialmente quando se pretende a visualização da estrutura das redes sociais ou MDS. Porém Bensman (2004) e White (2003) consideram que, ao propor uma análise estatística (multivariada, por exemplo), deve-se usar a correlação de Pearson.

Quando o foco de análise é a frequência de cocitação entre dois autores, independente do total do volume de cocitações, quer seja frequência absoluta ou a normalizada pelo  $C_s$  ou  $IJ$ , tem-se uma medida local, também chamada bilateral de similaridade. Quando a análise de cocitação entre dois autores está relacionada com todos os outros autores do conjunto de dados analisados, tem-se uma medida global ou multilateral, como é o caso de  $r$  de Pearson. Ahlgren et al. (2003) apontam que o uso de um ou outro método (local ou global) leva a resultados totalmente diferentes.

O  $C_s$  é sempre maior que o  $IJ$ . O resultado do índice para o “Cosseno de Salton é, em geral, o dobro do valor para o Índice de Jaccard” (HAMERS, 1989, p.315), mas tem-se com o  $IJ$  uma medida tão boa como  $C_s$ , porém Egghe e Leydesdorff (2009) preferem utilizar o  $C_s$ , considerando que as diferenças entre os resultados são mínimas.

### 3 METODOLOGIA

Conceituaram-se os três índices abordados, com questões sobre seu uso. Utilizaram-se os artigos publicados no periódico *Scientometrics*, sobre a temática “Estudos Métricos”, na base de dados *SCOPUS*, com os termos de busca: “*Bibliometr\**” OR “*Cientometr\**”, recuperando-se 234 artigos, publicados no período de 2013 e 2014, em junho de 2015. Identificaram-se 9327 pesquisadores citados, em um total de 8610 referências.

Construiu-se uma tabela em ordem decrescente de citações, desconsiderando-se as autocitações. O corte na tabela foi feito para os pesquisadores que foram citados em pelo menos 28 artigos, totalizando o grupo-alvo de 21 pesquisadores. Construiu-se a matriz de dados absolutos (21x21). Para o cálculo do  $C_s$  e  $IJ$ , utilizou-se o Microsoft *Excel*, e para o cálculo de  $r$  de Pearson, o software *SPSS*. Destacaram-se os índices que apresentaram valores

notáveis, apontando algumas contribuições destes três indicadores para a visualização das proximidades entre os autores cocitados, tendo como referência o valor absoluto.

## 4 APRESENTAÇÃO E ANÁLISE DOS DADOS

Apresentam-se a Figura 1 matriz absoluta e Figuras 2, 3 e 4 com as matrizes normalizadas pelo  $C_s$ ,  $IJ$  e  $r$  de Pearson.

Figura 1-Matriz de valores absolutos

|                   | Leidesdorff, L. | Van Raan, A. | Moed, H. | Schubert, A. | Bormann, L. | Hirsch, J. | Egghe, L. | Rousseau, R. | Waltman, L. | Van Eck, N. | Braun, T. | Ho, Y. | Van Leeuwen, T.N. | Moya-Angelín, F. | Bordons, M. | Martin, B. | Persson, O. | Ding, Y. | Wagner, C. | Akman, D. | Total de citações |
|-------------------|-----------------|--------------|----------|--------------|-------------|------------|-----------|--------------|-------------|-------------|-----------|--------|-------------------|------------------|-------------|------------|-------------|----------|------------|-----------|-------------------|
| Leidesdorff, L.   | 77              | 25           | 21       | 23           | 24          | 9          | 10        | 13           | 12          | 10          | 8         | 15     | 14                | 8                | 11          | 11         | 7           | 1        | 1          | 1         | 1                 |
| Van Raan, A.      | 25              | 70           | 14       | 23           | 26          | 18         | 17        | 12           | 21          | 20          | 12        | 5      | 20                | 12               | 13          | 14         | 6           | 10       | 10         | 12        | 79                |
| Moed, H.          | 21              | 14           | 65       | 21           | 29          | 19         | 18        | 16           | 12          | 13          | 11        | 18     | 15                | 11               | 10          | 9          | 7           | 5        | 15         | 65        | 65                |
| Schubert, A.      | 23              | 26           | 21       | 67           | 25          | 23         | 21        | 23           | 18          | 16          | 15        | 25     | 8                 | 12               | 11          | 9          | 10          | 16       | 13         | 12        | 9                 |
| Bormann, L.       | 9               | 18           | 16       | 65           | 22          | 19         | 16        | 10           | 6           | 14          | 6         | 8      | 13                | 14               | 7           | 7          | 7           | 13       | 65         | 65        |                   |
| Hirsch, J.        | 10              | 20           | 12       | 19           | 12          | 50         | 35        | 21           | 11          | 10          | 14        | 4      | 9                 | 6                | 9           | 7          | 4           | 8        | 2          | 8         | 50                |
| Egghe, L.         | 12              | 21           | 13       | 18           | 16          | 12         | 4         | 13           | 11          | 13          | 5         | 5      | 5                 | 5                | 5           | 5          | 5           | 5        | 5          | 5         | 12                |
| Rousseau, R.      | 13              | 14           | 16       | 15           | 10          | 14         | 4         | 9            | 6           | 9           | 7         | 4      | 8                 | 2                | 8           | 5          | 5           | 5        | 5          | 5         | 13                |
| Waltman, L.       | 10              | 10           | 7        | 13           | 6           | 14         | 6         | 8            | 5           | 3           | 8         | 9      | 8                 | 5                | 3           | 8          | 9           | 8        | 5          | 3         | 42                |
| Van Eck, N.       | 15              | 20           | 8        | 25           | 14          | 6          | 8         | 5            | 3           | 8           | 9         | 8      | 5                 | 3                | 8           | 9          | 8           | 5        | 3          | 8         | 16                |
| Braun, T.         | 8               | 6            | 5        | 15           | 5           | 5          | 5         | 5            | 5           | 5           | 5         | 5      | 5                 | 5                | 5           | 5          | 5           | 5        | 5          | 5         | 8                 |
| Ho, Y.            | 14              | 10           | 7        | 13           | 6           | 14         | 6         | 8            | 5           | 3           | 8         | 9      | 8                 | 5                | 3           | 8          | 9           | 8        | 5          | 3         | 27                |
| Van Leeuwen, T.N. | 8               | 10           | 5        | 15           | 5           | 5          | 5         | 5            | 5           | 5           | 5         | 5      | 5                 | 5                | 5           | 5          | 5           | 5        | 5          | 5         | 28                |
| Moya-Angelín, F.  | 11              | 14           | 9        | 10           | 16          | 12         | 4         | 13           | 11          | 13          | 5         | 5      | 5                 | 5                | 5           | 5          | 5           | 5        | 5          | 5         | 29                |
| Bordons, M.       | 11              | 14           | 9        | 10           | 16          | 12         | 4         | 13           | 11          | 13          | 5         | 5      | 5                 | 5                | 5           | 5          | 5           | 5        | 5          | 5         | 30                |
| Martin, B.        | 7               | 10           | 5        | 15           | 5           | 5          | 5         | 5            | 5           | 5           | 5         | 5      | 5                 | 5                | 5           | 5          | 5           | 5        | 5          | 5         | 31                |
| Persson, O.       | 1               | 10           | 5        | 15           | 5           | 5          | 5         | 5            | 5           | 5           | 5         | 5      | 5                 | 5                | 5           | 5          | 5           | 5        | 5          | 5         | 31                |
| Ding, Y.          | 1               | 10           | 5        | 15           | 5           | 5          | 5         | 5            | 5           | 5           | 5         | 5      | 5                 | 5                | 5           | 5          | 5           | 5        | 5          | 5         | 31                |
| Wagner, C.        | 1               | 10           | 5        | 15           | 5           | 5          | 5         | 5            | 5           | 5           | 5         | 5      | 5                 | 5                | 5           | 5          | 5           | 5        | 5          | 5         | 31                |
| Akman, D.         | 1               | 10           | 5        | 15           | 5           | 5          | 5         | 5            | 5           | 5           | 5         | 5      | 5                 | 5                | 5           | 5          | 5           | 5        | 5          | 5         | 31                |
| Total de citações | 77              | 70           | 65       | 67           | 65          | 50         | 50        | 47           | 42          | 36          | 33        | 35     | 27                | 28               | 29          | 30         | 31          | 31       | 29         | 28        | 190               |

Fonte: Elaborado pelos autores.

Figura 2-Matriz normalizada pelo Cosseno de Salton

|                   | Leidesdorff, L. | Van Raan, A. | Moed, H. | Schubert, A. | Garfield, E. | Bormann, L. | Hirsch, J. | Egghe, L. | Rousseau, R. | Waltman, L. | Van Eck, N. | Braun, T. | Ho, Y. | Van Leeuwen, T.N. | Moya-Angelín, F. | Bordons, M. | Martin, B. | Persson, O. | Ding, Y. | Wagner, C. | Akman, D. |
|-------------------|-----------------|--------------|----------|--------------|--------------|-------------|------------|-----------|--------------|-------------|-------------|-----------|--------|-------------------|------------------|-------------|------------|-------------|----------|------------|-----------|
| Leidesdorff, L.   | 1               | 0,34         | 0,30     | 0,35         | 0,31         | 0,29        | 0,15       | 0,17      | 0,23         | 0,23        | 0,32        | 0,19      | 0,18   | 0,32              | 0,30             | 0,17        | 0,31       | 0,29        | 0,28     | 0,38       | 0,24      |
| Van Raan, A.      | 0,34            | 1            | 0,59     | 0,34         | 0,39         | 0,41        | 0,30       | 0,30      | 0,22         | 0,42        | 0,42        | 0,24      | 0,13   | 0,45              | 0,27             | 0,28        | 0,30       | 0,13        | 0,23     | 0,22       | 0,27      |
| Moed, H.          | 0,30            | 0,59         | 1        | 0,32         | 0,45         | 0,33        | 0,32       | 0,29      | 0,23         | 0,37        | 0,28        | 0,23      | 0,15   | 0,42              | 0,35             | 0,35        | 0,22       | 0,20        | 0,16     | 0,12       | 0,35      |
| Schubert, A.      | 0,35            | 0,34         | 0,33     | 1            | 0,38         | 0,40        | 0,36       | 0,41      | 0,34         | 0,33        | 0,32        | 0,72      | 0,19   | 0,28              | 0,25             | 0,20        | 0,22       | 0,35        | 0,29     | 0,27       | 0,21      |
| Garfield, E.      | 0,31            | 0,39         | 0,33     | 0,41         | 1            | 0,39        | 0,33       | 0,34      | 0,31         | 0,27        | 0,29        | 0,14      | 0,19   | 0,30              | 0,32             | 0,16        | 0,16       | 0,16        | 0,30     | 0,27       | 0,27      |
| Bormann, L.       | 0,29            | 0,41         | 0,33     | 0,36         | 0,30         | 1           | 0,36       | 0,31      | 0,28         | 0,45        | 0,39        | 0,29      | 0,11   | 0,35              | 0,29             | 0,24        | 0,13       | 0,13        | 0,11     | 0,18       | 0,24      |
| Hirsch, J.        | 0,15            | 0,30         | 0,32     | 0,40         | 0,34         | 0,36        | 1          | 0,37      | 0,46         | 0,26        | 0,25        | 0,33      | 0,11   | 0,24              | 0,19             | 0,23        | 0,18       | 0,10        | 0,21     | 0,05       | 0,21      |
| Egghe, L.         | 0,17            | 0,42         | 0,32     | 0,41         | 0,31         | 0,29        | 0,37       | 1         | 0,61         | 0,19        | 0,23        | 0,32      | 0,08   | 0,19              | 0,24             | 0,27        | 0,13       | 0,10        | 0,30     | 0,14       | 0,19      |
| Rousseau, R.      | 0,23            | 0,22         | 0,29     | 0,41         | 0,31         | 0,29        | 0,45       | 0,39      | 1            | 0,03        | 0,23        | 0,16      | 0,44   | 0,28              | 0,24             | 0,18        | 0,09       | 0,15        | 0,22     | 0,09       | 0,09      |
| Waltman, L.       | 0,22            | 0,42         | 0,37     | 0,34         | 0,31         | 0,28        | 0,45       | 0,39      | 0,03         | 1           | 0,04        | 0,24      | 0,13   | 0,43              | 0,29             | 0,12        | 0,16       | 0,09        | 0,16     | 0,19       | 0,10      |
| Van Eck, N.       | 0,32            | 0,42         | 0,37     | 0,34         | 0,31         | 0,29        | 0,37       | 0,46      | 0,26         | 0,24        | 1           | 0,04      | 0,24   | 0,13              | 0,43             | 0,29        | 0,12       | 0,16        | 0,09     | 0,16       | 0,10      |
| Braun, T.         | 0,23            | 0,28         | 0,23     | 0,72         | 0,19         | 0,29        | 0,14       | 0,19      | 0,30         | 0,32        | 0,04        | 1         | 0,00   | 0,23              | 0,24             | 0,16        | 0,06       | 0,12        | 0,23     | 0,14       | 0,08      |
| Ho, Y.            | 0,18            | 0,13         | 0,15     | 0,19         | 0,14         | 0,19        | 0,30       | 0,32      | 0,16         | 0,16        | 0,13        | 0,00      | 1      | 0,00              | 0,03             | 0,23        | 0,16       | 0,44        | 0,28     | 0,24       | 0,18      |
| Van Leeuwen, T.N. | 0,32            | 0,45         | 0,42     | 0,28         | 0,30         | 0,33        | 0,45       | 0,27      | 0,28         | 0,30        | 0,32        | 0,23      | 0,24   | 1                 | 0,00             | 0,24        | 0,13       | 0,43        | 0,29     | 0,12       | 0,16      |
| Moya-Angelín, F.  | 0,30            | 0,27         | 0,35     | 0,25         | 0,16         | 0,16        | 0,06       | 0,12      | 0,23         | 0,14        | 0,08        | 0,23      | 0,24   | 0,24              | 1                | 0,00        | 0,14       | 0,10        | 0,12     | 0,07       | 0,11      |
| Bordons, M.       | 0,38            | 0,23         | 0,35     | 0,29         | 0,16         | 0,16        | 0,06       | 0,12      | 0,23         | 0,14        | 0,08        | 0,23      | 0,24   | 0,14              | 0,10             | 1           | 0,00       | 0,14        | 0,10     | 0,12       | 0,07      |
| Martin, B.        | 0,29            | 0,16         | 0,16     | 0,35         | 0,16         | 0,16        | 0,06       | 0,12      | 0,23         | 0,14        | 0,08        | 0,23      | 0,24   | 0,14              | 0,10             | 0,12        | 1          | 0,00        | 0,14     | 0,10       | 0,12      |
| Persson, O.       | 0,27            | 0,21         | 0,12     | 0,21         | 0,30         | 0,13        | 0,11       | 0,09      | 0,09         | 0,16        | 0,09        | 0,16      | 0,09   | 0,09              | 0,09             | 0,09        | 0,09       | 1           | 0,00     | 0,19       | 0,17      |
| Ding, Y.          | 0,28            | 0,22         | 0,35     | 0,27         | 0,16         | 0,16        | 0,06       | 0,12      | 0,23         | 0,14        | 0,08        | 0,23      | 0,24   | 0,14              | 0,10             | 0,12        | 0,07       | 0,30        | 1        | 0,00       | 0,17      |
| Wagner, C.        | 0,38            | 0,23         | 0,35     | 0,29         | 0,16         | 0,16        | 0,06       | 0,12      | 0,23         | 0,14        | 0,08        | 0,23      | 0,24   | 0,14              | 0,10             | 0,12        | 0,07       | 0,19        | 0,00     | 1          | 0,00      |
| Akman, D.         | 0,24            | 0,27         | 0,35     | 0,21         | 0,30         | 0,13        | 0,11       | 0,09      | 0,09         | 0,16        | 0,09        | 0,16      | 0,09   | 0,09              | 0,09             | 0,09        | 0,09       | 0,27        | 0,17     | 0,14       | 0,00      |

Fonte: Elaborado pelos autores

Figura 3-Matriz normalizada pelo Índice Jaccard

|                   | Leidesdorff, L. | Van Raan, A. | Moed, H. | Schubert, A. | Bormann, L. | Hirsch, J. | Egghe, L. | Rousseau, R. | Waltman, L. | Van Eck, N. | Braun, T. | Ho, Y. | Van Leeuwen, T.N. | Moya-Angelín, F. | Bordons, M. | Martin, B. | Persson, O. | Ding, Y. | Wagner, C. | Akman, D. |
|-------------------|-----------------|--------------|----------|--------------|-------------|------------|-----------|--------------|-------------|-------------|-----------|--------|-------------------|------------------|-------------|------------|-------------|----------|------------|-----------|
| Leidesdorff, L.   | 1,00            | 0,20         | 0,17     | 0,21         | 0,18        | 0,13       | 0,08      | 0,09         | 0,12        | 0,18        | 0,17      | 0,10   | 0,08              | 0,17             | 0,15        | 0,14       | 0,15        | 0,14     | 0,20       | 0,12      |
| Van Raan, A.      | 0,20            | 1,00         | 0,24     | 0,20         | 0,24        | 0,25       | 0,18      | 0,17         | 0,12        | 0,22        | 0,24      | 0,11   | 0,07              | 0,26             | 0,14        | 0,15       | 0,16        | 0,11     | 0,11       | 0,14      |
| Moed, H.          | 0,17            | 0,24         | 1,00     | 0,19         | 0,20        | 0,19       | 0,17      | 0,13         | 0,12        | 0,15        | 0,13      | 0,12   | 0,11              | 0,19             | 0,10        | 0,09       | 0,08        | 0,08     | 0,12       | 0,12      |
| Schubert, A.      | 0,21            | 0,20         | 0,19     | 1,00         | 0,21        | 0,21       | 0,17      | 0,13         | 0,12        | 0,15        | 0,13      | 0,12   | 0,11              | 0,19             | 0,10        | 0,09       | 0,08        | 0,08     | 0,12       | 0,12      |
| Bormann, L.       | 0,18            | 0,25         | 0,18     | 0,21         | 1,00        | 0,24       | 0,20      | 0,18         | 0,11        | 0,09        | 0,14      | 0,07   | 0,09              | 0,14             | 0,11        | 0,10       | 0,08        | 0,08     | 0,14       | 0,10      |
| Hirsch, J.        | 0,13            | 0,12         | 0,13     | 0,12         | 0,11        | 1,00       | 0,22      | 0,20         | 0,18        | 0,15        | 0,08      | 0,12   | 0,11              | 0,19             | 0,10        | 0,09       | 0,08        | 0,08     | 0,12       | 0,12      |
| Egghe, L.         | 0,08            | 0,12         | 0,12     | 0,12         | 0,11        | 0,22       | 1,00      | 0,22         | 0,20        | 0,18        | 0,15      | 0,08   | 0,12              | 0,19             | 0,10        | 0,09       | 0,08        | 0,08     | 0,12       | 0,12      |
| Rousseau, R.      | 0,09            | 0,14         | 0,14     | 0,13         | 0,11        | 0,08       | 0,22      | 1,00         | 0,24        | 0,20        | 0,18      | 0,15   | 0,08              | 0,19             | 0,10        | 0,09       | 0,08        | 0,08     | 0,12       | 0,12      |
| Waltman, L.       | 0,15            | 0,16         | 0,15     | 0,12         | 0,10        | 0,14       | 0,18      | 0,20         | 1,00        | 0,24        | 0,20      | 0,18   | 0,15              | 0,19             | 0,10        | 0,09       | 0,08        | 0,08     | 0,12       | 0,12      |
| Van Eck, N.       | 0,17            | 0,11         | 0,07     | 0,07         | 0,12        | 0,12       | 0,12      | 0,12         | 0,20        | 1,00        | 0,24      | 0,20   | 0,18              | 0,19             | 0,10        | 0,09       | 0,08        | 0,08     | 0,12       | 0,12      |
| Braun, T.         | 0,10            | 0,11         | 0,11     | 0,11         | 0,08        | 0,08       | 0,08      | 0,08         | 0,10        | 0,20        | 1,00      | 0,24   | 0,20              | 0,19             | 0,10        | 0,09       | 0,08        | 0,08     | 0,12       | 0,12      |
| Ho, Y.            | 0,14            | 0,15         | 0,15     | 0,15         | 0,14        | 0,14       | 0,14      | 0,14         | 0,16        | 0,18        | 0,24      | 1,00   | 0,24              | 0,20             | 0,19        | 0,10       | 0,09        | 0,08     | 0,08       | 0,12      |
| Van Leeuwen, T.N. | 0,17            | 0,15         | 0,15     | 0,15         | 0,14        | 0,14       | 0,14      | 0,14         | 0,16        | 0,18        | 0,24      | 0,24   | 1,00              | 0,24             | 0,20        | 0,19       | 0,10        | 0,09     | 0,08       | 0,12      |
| Moya-Angelín, F.  | 0,15            | 0,14         | 0,14     | 0,14         | 0,14        | 0,14       | 0,14      | 0,14         | 0,16        | 0,18        | 0,24      | 0,24   | 0,24              | 1,00             | 0,24        | 0,20       | 0,19        | 0,10     | 0,09       | 0,12      |
| Bordons, M.       | 0,15            | 0,14         | 0,14     | 0,14         | 0,14        | 0,14       | 0,14      | 0,14         | 0,16        | 0,18        | 0,24      | 0,24   | 0,24              | 0,24             | 1,00        | 0,24       | 0,20        | 0,19     | 0,10       | 0,12      |
| Martin, B.        | 0,11            | 0,11         | 0,11     | 0,11         | 0,10        | 0,10       | 0,10      | 0,10         | 0,12        | 0,12        | 0,12      | 0,12   | 0,12              | 0,12             | 0,12        | 1,00       | 0,24        | 0,20     | 0,19       | 0,12      |
| Persson, O.       | 0,12            | 0,12         | 0,12     | 0,12         | 0,10        | 0,10       | 0,10      | 0,10         | 0,12        | 0,12        | 0,12      | 0,12   | 0,12              | 0,12             | 0,12        | 0,24       | 1,00        | 0,24     | 0,20       | 0,12      |
| Ding, Y.          | 0,20            | 0,11         | 0,09     | 0,09         | 0,08        | 0,08       | 0,08      | 0,08         | 0,10        | 0,10        | 0,10      | 0,10   | 0,10              | 0,10             | 0,10        | 0,10       | 0,24        | 1,00     | 0,24       | 0,12      |
| Wagner, C.        | 0,12            | 0,12         | 0,12     | 0,12         | 0,10        | 0,10       | 0,10      | 0,10         | 0,12        | 0,12        | 0,12      | 0,12   | 0,12              | 0,12             | 0,12        | 0,12       | 0,24        | 0,24     | 1,00       | 0,12      |
| Akman, D.         | 0,12            | 0,12         | 0,12     | 0,12         | 0,10        | 0,10       | 0,10      | 0,10         | 0,12        | 0,12        | 0,12      | 0,12   | 0,12              | 0,12             | 0,12        | 0,12       | 0,12        | 0,12     | 0,12       | 1,00      |

Fonte:

Analisando-se e comparando-se os valores de  $C_s$  e  $IJ$ , nos cinco pares de cocitações destacados, observam-se que os valores de  $C_s$  são maiores que  $IJ$  (HAMERS, 1989).

Nas normalizações, analisando-se as Figuras 1, 2, 3 e 4 e a Tabela 1, quer seja por  $C_s$ ,  $IJ$  e  $r$  de Pearson, os maiores valores são para Waltman e Van Eck. Embora não tenham o maior valor absoluto para cocitação, eles possuem os maiores valores normalizados, isto porque eles diferenciam a intensidade de proximidade, não discriminada pelas frequências absolutas, com o maior valor igual a 0,99 para  $r$  de Pearson, significando correlação quase perfeita entre os dois autores. Esse resultado advém de 32 cocitações absolutas relativizadas, para 36 citações do Waltman e 33 citações do Van Eck. As demais correlações seguem igual raciocínio, mas com menor intensidade de cocitação.

Tabela 1 - Autores e valores extremados dos indicadores absolutos e normalizados

| Autores cocitados         | Índic. absoluto | Indicadores Normalizados |      |                |
|---------------------------|-----------------|--------------------------|------|----------------|
|                           |                 | $C_s$                    | $IJ$ | $r$ de Pearson |
| Vaan Ran, A. x Moed, H.   | 34 cocitações   | 0,50                     | 0,34 | 0,86           |
| Schuber, A. x Braun, T.   | 35 cocitações   | 0,72                     | 0,52 | 0,84           |
| Hirsch, J. x Egghe, L.    | 35 cocitações   | 0,72                     | 0,56 | 0,93           |
| Egghe, L. x Rousseau, R.  | 27 cocitações   | 0,61                     | 0,44 | 0,88           |
| Waltman, L. x Van Eck, N. | 32 cocitações   | 0,93                     | 0,76 | 0,99           |
| Moed, H. x Ho, Y.         | 2 cocitações    | 0,05                     | 0,02 | 0,28           |
| Ho, Y. x Van Leeuwen, T.  | 1 cocitação     | 0,04                     | 0,02 | 0,19           |
| Ho, Y. x Bordons, M.      | 1 cocitação     | 0,04                     | 0,02 | 0,06           |
| Ho, Y. x Martin, B.       | 1 cocitação     | 0,03                     | 0,02 | 0,41           |

**Fonte:** Elaborado pelos autores.

Por outro lado, compararam-se os quatro menores valores absolutos com suas respectivas normalizações,  $C_s$ ,  $IJ$  e  $r$  de Pearson (em cor verde), tendo os três últimos pares apenas uma cocitação. Observa-se que  $C_s$  e  $IJ$  estão próximos para os nove valores e distantes dos valores de  $r$  de Pearson, explicado pelo fato de serem índices de correlação local e o  $r$  de Pearson ser um índice global, influenciado por todo o conjunto (ALHGREN et al., 2003).

Em continuação a esta pesquisa, no sentido de aprofundá-la e obter conclusões mais consistentes, propõe-se calcular a Correlação de Postos (Correlação de Spearman) entre os três diferentes índices normalizados -  $C_s$ ,  $IJ$  e  $r$  de Pearson - a partir dos dados de cada uma das matrizes, ordenados em postos. Comparando os três resultados obtidos pela Correlação de Postos, podem-se obter indicadores que contribuam para destacar se algum dos índices normalizados deve ser utilizado de forma mais discriminativa ou se eles podem ou não ser usados indiscriminadamente. Pode ocorrer que o recorte da amostra tenha excluído casos mais destacados, para os quais os índices apresentam comportamentos bastante distintos.

## 5 CONSIDERAÇÕES FINAIS

Apesar de não existir uma indicação unívoca para o uso dos índices de normalização, as pesquisas na temática sugerem que a escolha do procedimento está articulada com os objetivos da pesquisa e natureza dos dados, para a qual se propõe a normalização. Assim, quando se pretende visualizar uma rede de cocitações, tanto o  $C_s$  como o  $IJ$  são índices bem aceitos. Quando o objetivo do estudo se encaminha para uma análise estatística mais avançada, a exemplo da multivariada, o  $r$  de Pearson é mais aconselhado. Observe-se que as informações utilizadas para  $C_s$  e  $IJ$  diferem daquelas usadas para o cálculo do  $r$  de Pearson.

Indica-se a continuidade e o aprofundamento de estudos nas questões relativas às normalizações de índices, especialmente de cocitações ou autorias.

**AGRADECIMENTOS:** Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

## REFERÊNCIAS

- AHLGREN, P. et al. Requirements for a Cocitation Similarity Measure, with Special Reference to Pearson's Correlation Coefficient. **Journal of the American Society for Information Science and Technology**, v. 54, n. 6, p.550-560, 2003.
- BENSMAN, S. J. Person's  $r$  and author cocitation analysis: a commentary on the controversy. **Journal of the American Society of Information Science & Technology**, New York, v. 55, n. 10, p. 935-936, 2004.
- EGGHE, L.; LEYDESDORFF, L. The relation between Pearson's correlation coefficient  $r$  and Salton's cosine measure. **Journal of the American Society for Information Science and Technology**, New York, v. 60, n. 5, p. 1027-1036, 2009.
- GLÄNZEL, W. et al. Subfield-specific normalized relative indicators and a new generation of relational charts: methodological foundations illustrated on the assessment of institutional research performance. **Scientometrics**, Amsterdam, v. 78, n. 1, p. 165–188, 2009.
- HAMERS, L. et al. Similarity measures in scientometric research: the Jaccard Index versus Salton's Cosine formula. **Information Processing de Management**, New York, v. 25, n. 3, p. 315-318, 1989.
- LEYDESDORFF, L.; VAUGHAN, L. Co-occurrence Matrices and their applications in Information Science: Extending ACA to the Web environment. **Journal of the American Society for Information Science and Technology**, v. 57, n. 12, p.1616-1628, 2006.
- LUUKKONEN, T. et al. The measurement of international scientific collaboration. **Scientometrics**, Amsterdam, v. 28, n. 1, p. 15-36, 1993.
- SMALL, H. Co-citation in the scientific literature: a new measure of the relationship between two documents. **Journal of the American Society for Information Science**, Washington, v. 24, n. 4, p. 265-269, 1973.
- WHITE, H.D.; GRIFFITH, B. Author co-citation: a literature measure of intellectual structure. **Journal of the American Society for Information Science & Technology**, v. 32, n. 2., p. 163-171, 1981.
- WHITE, H.D. Author cocitacion analysis and Pearson's  $r$ . **Journal of the American Society of Information Science & Technology**, New York, v.54, n.13, p.1250-1259, 2003.