



# **VIÉS DE CATEGORIZAÇÃO DA WEB OF SCIENCE NA DETERMINAÇÃO DAS PUBLICAÇÕES HISTORICAMENTE MAIS INFLUENTES DAS ÁREAS DE PESQUISA: uma crítica do resultado obtido para a Ciência da Informação & Biblioteconomia**

Gerson Pech<sup>1</sup>  
Catarina Delgado<sup>2</sup>

**Resumo:** A categorização dos artigos científicos em campos de pesquisa pela Web of Science têm sido criticada pelos autores da Ciência da Informação e Biblioteconomia (CI&B). Este estudo tem como objetivo mostrar que esta categorização pode levar a resultados incompatíveis na determinação das publicações historicamente mais relevantes dos campos de pesquisa. Para CI&B, revelamos que a determinação das publicações mais influentes é fortemente influenciada por periódicos que além de pertencerem à CI&B, também pertencem a área de Gestão. Isso mostra que a categorização do WoS pode estar distorcendo os resultados de estudos que determinam as publicações historicamente mais influentes das áreas científicas.

**Palavras-Chave:** Web of Science. Categorização. Publicações historicamente influentes. CRExplorer.

## **1 INTRODUÇÃO**

A identificação dos artigos mais influentes de um campo de pesquisa é um processo essencial em estudos de revisão e prospectivos. Um propósito fundamental é que esse processo de identificação não privilegie nenhum período de publicação ou assunto específico do campo de pesquisa para não comprometer os resultados do estudo. Entretanto, várias tendências em desacordo com esse propósito têm sido apontadas nos métodos usuais de seleção de artigos influentes. Por exemplo, Pech e Delgado (2021) mostraram que as métricas de citações mais utilizadas para selecionar artigos para revisões de literatura privilegiam determinados períodos enquanto negligenciam outros. De fato, tanto o período de publicação quanto o ano em que o artigo é citado devem ser considerados para normalizar as métricas de seleção de artigos mais influentes de um campo de pesquisa. Considerando esses dois elementos, Thor *et al.* (2018) introduziram um novo indicador para descobrir as publicações historicamente mais

---

<sup>1</sup> Universidade do Estado do Rio de Janeiro (UERJ)

<sup>2</sup> Universidade do Porto

influentes de um campo. Os autores utilizaram o software CRExplorer. Esse indicador, também difere dos usuais, pois em vez de contar o número total de citações dos artigos, conta apenas as citações dentro de uma mesma área. Ampliando o estudo, Thor *et al.* (2021) definiram o indicador N\_TOPO\_1+ como sendo o número de anos em que o artigo esteve entre os 0,1% mais citados do ano, considerando citações na mesma área. As publicações com N\_TOPO\_1+ mais altos são as historicamente mais influentes (*landmarks*). Na categoria Ciência da Informação & Biblioteconomia (CI&B) da Web of Science (WoS), os *landmarks* são:

Porter ME. *Estratégia competitiva: técnicas para analisar indústrias e concorrentes*. New York: Free Press, 1980 (*Porter-landmark*)

Giddens A. *A constituição da sociedade: Esboço da teoria da estruturação*. Cambridge: Polity Press, 1984 (*Giddens-landmark*)

Por pertencerem à área de CI&B, Thor *et al.* (2021) consideraram seu resultado como sendo “contra-intuitivo (no contexto de nosso conhecimento especializado)”. Ou seja, um resultado completamente fora do esperado por pesquisadores dessa área. De fato, essas não são as publicações historicamente mais relevantes da área da Ciência da Informação, principalmente no que se refere aos temas voltados ao estudo da Bibliometria, Cienciometria, Informetria ou Webometria (denominados como campo de pesquisa “Métricas de Informação”). Por que, então, o resultado obtido foi esse? Esse é um questionamento relevante pois pode ajudar a entender as limitações existentes na categorização das áreas de pesquisa pelo WoS.

Assim, este estudo tem como objetivo mostrar que a categorização usada pelo WoS pode levar a resultados incompatíveis na determinação das publicações historicamente mais relevantes dos campos de pesquisa. Mais especificamente, nosso estudo tem o propósito de analisar a razão pela qual as duas publicações acima lideram o ranking das mais influentes da nossa área de pesquisa. Traduzimos esse objetivo nas seguintes questões de pesquisa:

RQ1. Quais periódicos foram responsáveis pela identificação de *Porter/Giddens-landmarks*? Ou seja, os artigos que citam esses dois *landmarks* foram publicados em quais periódicos? Esses periódicos são populares no campo “Métricas de Informação”? Além de serem classificados em CI&B, em qual outra categoria do WoS esses periódicos estão classificados?

RQ2. Quais são os *landmarks* do campo “Métricas de Informação”? Esses trabalhos são reconhecidos na área como historicamente relevantes? Ou eles também são “contra-intuitivos”?

## 2 DADOS E MÉTODOS

### 2.1 RQ1: PERIÓDICOS DE CI&B

Dados de todos os artigos e revisões dos jornais classificados em CI&B (1980-2017) foram recuperados do WoS em 14 de novembro de 2021. Este período corresponde ao mesmo do estudo anterior Thor *et al.* (2021). As referências citadas de cada publicação, também foram incluídas. Um total de 93.330 *papers* foram obtidos. O método consistiu em identificar os títulos dos periódicos dos *papers* que citaram *Porter/Giddens-landmarks*. Para isso, realizamos uma busca detalhada nas referências citadas dos 93.330 artigos recuperados. Cada periódico possui um padrão diferente para registro das referências citadas (por exemplo, alguns periódicos usam o primeiro nome do autor por extenso, outros usam apenas a primeira letra). Essas diferentes formas de registro de referência podem dificultar algumas buscas automáticas desenvolvidas para *Big Data Analysis*, pois estão sempre pressionadas a reduzir o tempo de processamento. Assim, para fazer uma busca rigorosa, testamos e usamos diferentes formatos de escrita para os *Porter/Giddens-landmarks*. Para esta análise, usamos o MS Excel 365.

### 2.2 RQ2: PERIÓDICOS DO CAMPO “MÉTRICAS DE INFORMAÇÃO”

Recuperamos os dados de todos os artigos e revisões (1980-2017) do WoS dos seguintes periódicos: *Scientometrics*; *Journal of Informetrics*; *Journal of Information Science*; *Journal of the American Society for Information Science*; *Journal of the American Society for Information Science and Technology*; and *Journal of the Association for Information Science and Technology*. Esses três últimos periódicos são, na verdade, nomes diferentes para o mesmo periódico, que teve seu nome alterado em três períodos diferentes. Essa lista de periódicos é baseada na seleção realizada por Leydesdorff *et al.* (2014) para tratar os periódicos do campo “Métricas de Informação”. Dados de 9.818 *papers* foram exportados em 14/11/2021, contendo 184.037 referências diferentes. No ambiente do WoS, os conjuntos de dados foram coletados usando a opção "Salvar em outros formatos de arquivo"; em Conteúdo de Gravação, selecionamos "Registro Completo e Referências Citadas" e, como Formato de Arquivo, selecionamos "Outro Software de Referência". Para identificar as publicações que são os *landmarks* desse conjunto de dados, ou seja, do campo “Métricas de Informação”, usamos o mesmo método de Thor *et al.* (2021). Assim, os artigos foram classificados de acordo com o indicador  $N\_TOP0\_1+$ , e a publicação com maior valor de  $N\_TOP0\_1+$  foi considerada a historicamente mais influente neste campo de pesquisa. Conforme já

mencionado (seção 1), o indicador  $N\_TOP0\_1+$  de uma publicação é o número de vezes em que a publicação aparece no grupo dos 0,1% mais citados do ano. Portanto, considerando o conjunto de dados de “Métricas de Informação”, para cada ano do período estudado (1980-2017) listamos todas as referências citadas no ano e colocamos essa lista em ordem decrescente do número de citações no ano. Considere  $i=1$  a posição da primeira publicação da lista e  $N$  (número total de publicações no ano) a posição da última da lista. Portanto, as publicações que possuem  $i < N/1000$  ganham uma pontuação igual a 1 para  $N\_TOP0\_1+$ . Se isso acontece para uma determinada publicação  $x$  vezes ( $x$  anos), então  $N\_TOP0\_1+ = x$  para essa publicação. Para esta análise usamos o CRExplorer e o MS Excel 365.

### 3 RESULTADOS E DISCUSSÕES

#### 3.1 RQ1: PORTER/GIDDENS-LANDMARKS NOS PERIÓDICOS CI&B

Encontramos 319 *papers* em 51 periódicos citando *Porter-landmark*. Em 31 desses periódicos, essa referência aparece apenas uma ou duas vezes. No caso de *Giddens-landmark*, a encontramos em 421 *papers* de 61 periódicos diferentes; em 24 desses periódicos, a referência aparece apenas uma ou duas vezes. As Tabelas 1 e 2 mostram os dez periódicos que mais citam *Porter-landmark* e *Giddens-landmark*, respectivamente. A primeira coluna apresenta o título da revista e sua categoria no WoS. Como todos esses periódicos pertencem à CI&B, essa categoria não foi exibida na tabela. A segunda coluna mostra o número de vezes em que *Porter/Giddens-landmarks* foram citados; no terceiro, o percentual relativo ao total (319 e 421); e por último, a posição do periódico no ranking do número de artigos considerando a CI&B (R). Nenhum desses periódicos são periódicos do campo “Métricas de Informação” (seção 2.2). Eles estão principalmente relacionados à área de gestão; de negócios; de sistemas de informação e de gestão de recursos de TI<sup>3</sup>. Nas Tabelas 1 e 2, vemos que nove das 14 revistas pertencem à área de Gestão. Esse resultado mostra o peso substancial que a área de Gestão teve na determinação dos *landmarks* na pesquisa de Thor *et al.* (2021). Observe que a última coluna de cada tabela traz informações adicionais, mostrando que esses periódicos não se enquadram nessas tabelas por possuírem o maior número de artigos na área.

---

<sup>3</sup> Nas páginas de alguns desses periódicos ([journals.elsevier.com/information-and-management](http://journals.elsevier.com/information-and-management); [tandfonline.com/toc/tjis20/current](http://tandfonline.com/toc/tjis20/current) e [misq.org](http://misq.org)) fica claro que os assuntos principais são voltados à área de Gestão.

Tabela 1 - Os *top10* periódicos da CI&B que citam *Porter-landmark*. Na terceira coluna mostramos a porcentagem de artigos de cada periódico que citam esta referência em relação ao total de artigos que a citam e, na última, o ranking relativo ao total de artigos CI&B (R)

Título do periódico (categorias do WoS em adição a CI&B)	<i>Porter-landmark</i>	%	R
Information & Management (Gestão; Ciência da Computação, Sistemas de Informação)	44	13.8	8
Mis Quarterly (Gestão; Ciência da Computação, Sistemas de Informação)	32	10.0	21
International Journal of Information Management	30	9.4	18
Journal of Strategic Information Systems (Gestão; Ciência da Computação, Sistemas de Informação)	25	7.8	77
European Journal of Information Systems (Gestão; Ciência da Computação, Sistemas de Informação)	21	6.6	36
Journal of Information Technology (Gestão; Ciência da Computação, Sistemas de Informação)	21	6.6	54
Information Systems Research (Gestão)	19	6.0	30
Journal of Management Information Systems (Gestão; Ciência da Computação, Sistemas de Informação)	16	5.0	45
Information Systems Journal	12	3.8	61
Telecommunications Policy (Telecomunicações; Comunicação)	10	3.1	11

Tabela 2 - O mesmo do que na Tabela 1, mas para *papers* do CI&B que citam *Giddens-landmark*

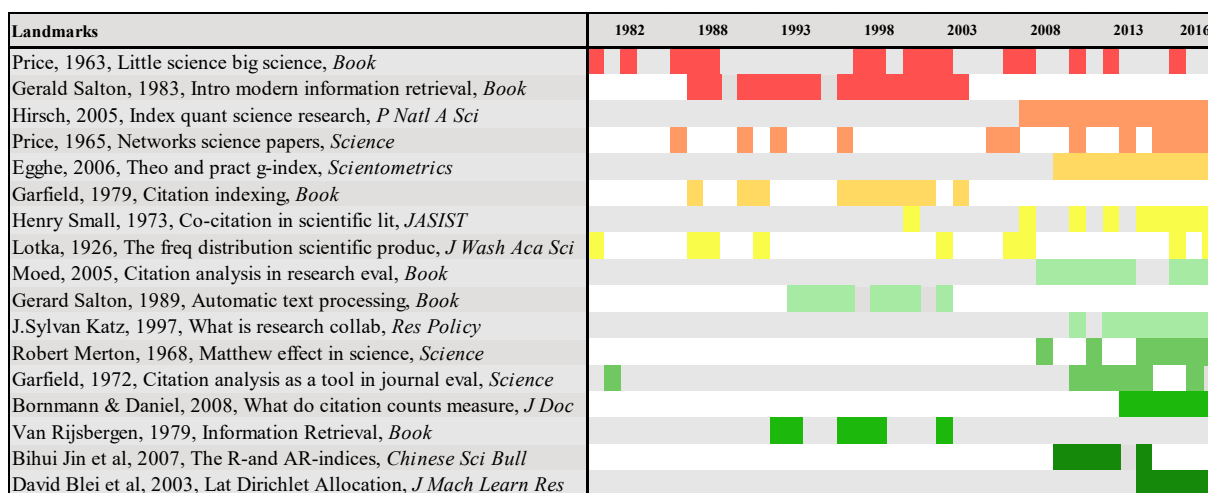
Título do periódico (categorias do WoS em adição a CI&B)	<i>Giddens-landmark</i>	%	R
Mis Quarterly (Gestão; Ciência da Computação, Sistemas de Informação)	49	11.6	21
European Journal of Information Systems (Gestão; Ciência da Computação, Sistemas de Informação)	34	8.1	36
Information Systems Journal	32	7.6	61
Journal of Information Technology (Gestão; Ciência da Computação, Sistemas de Informação)	23	5.5	54
Information and Organization (Gestão)	18	4.3	127
Information Society (Comunicação)	18	4.3	63
Information Systems Research (Gestão)	18	4.3	30
Social Science Information Sur les Sciences Sociales (Ciências Sociais, Interdisciplinaridade)	17	4.0	38
Journal of Strategic Information Systems (Gestão; Ciência da Computação, Sistemas de Informação)	15	3.6	77
Journal of the Association for Information Systems (Gestão; Ciência da Computação, Sistemas de Informação)	15	3.6	82

### 3.2 RQ2: LANDMARKS DO CAMPO “MÉTRICAS DE INFORMAÇÃO”

A Figura 1 mostra os *landmarks* considerando os periódicos mencionados na seção 2.2. Ou seja, periódicos com escopo voltado para pesquisas de métricas de informação, como bibliometria, informetria e cientometria. As publicações são classificadas pelo valor de N\_TOP0\_1+, e a figura mostra os artigos com N\_TOP0\_1+ maior do que quatro. O quadrado colorido revela os anos em que a publicação ficou no grupo dos 0,1% mais citados. A figura também revela claramente a linha do tempo da influência de cada publicação.

O primeiro *landmark* é o renomado livro “*Little Science Big Science*” de Derek de Solla Price, “pai” do estudo quantitativo da ciência da ciência e de suas aplicações em política científica. O segundo é outro livro marcante da área. “*Introduction to Modern Information Retrieval*” foi escrito por Gerard Salton, “pai” da recuperação da informação. A terceira publicação do ranking é o artigo de Hirsch em que ele propõe o índice H. O seguinte é o artigo de Price que antecipou os estudos das redes de citações, largamente utilizadas em estudos atuais. De fato, as publicações dessa lista são compatíveis com a expectativa deste campo de pesquisa, pois revelam publicações reconhecidas na literatura (LEYDESDORFF *et al.*, 2014; HOU, 2017).

Figura 1 - *Landmarks* do campo “Métricas de Informação” com N\_TOP0\_1+ maior do que cinco



Os quadrados coloridos indicam os anos em que a publicação esteve no Top0.1% dos mais citados dentro na área. Na figura, N\_TOP0\_1+ está na faixa de 15 (vermelho) a 5 (verde escuro).

## 4 CONCLUSÕES

Recentemente, Thor *et al.* (2021) identificaram livros de Michael Porter e Anthony Giddens como sendo as publicações historicamente mais influentes no campo de CI&B. No entanto, esses resultados não são os esperados pelos pesquisadores dessa área. Neste artigo,

investigamos as razões que motivaram esses resultados. Mostramos que ele tem origem no fato de que vários periódicos que publicam majoritariamente nas áreas de Gestão, Negócios e Gestão de Sistemas de Informação estão classificados pelo WoS também na área de CI&B. Ou seja, o resultado tem origem em *papers* que, de fato, pertencem à CI&B, mas cujo vínculo mais forte é com outro campo de pesquisa. Com esse resultado, respondemos completamente à RQ1 estabelecida para este estudo. Em relação à RQ2, mostramos que os *landmarks* do campo “Métricas de Informação”, obtidos pelo nosso estudo, são, de fato, os trabalhos reconhecidos na área como historicamente relevantes. Sendo assim, este paper mostrou que a categorização de *papers* definida pelo WoS pode estar introduzindo um viés nos estudos que buscam identificar as publicações mais influentes de cada área. Adicionalmente, esse viés também deve estar presente em estudos que utilizam essa categorização como fonte para a construção de seu conjunto de dados de pesquisa.

Os resultados apresentados por este estudo confirmam e ampliam trabalhos recentes que mostraram, através de outros métodos, que a classificação dos papers em áreas de pesquisa através dos periódicos (usado por Scopus e WoS) são pouco úteis para a pesquisa científica (SHU *et al.*, 2019). Portanto, esta pesquisa pode contribuir com estudos que (i) utilizam a classificação das áreas do WoS; (ii) que planejam essa utilização; ou (iii) que estejam desenvolvendo outros métodos para a classificação. Analisar as falhas da classificação do WoS é relevante para o desenvolvimento de outros modelos, pois a classificação adequada dos artigos em seus campos de pesquisa é fundamental para avaliar corretamente seu impacto – uma vez que as métricas de quantidade e qualidade usadas na avaliação acadêmica precisam estar normalizadas por campo (BORNMANN, 2018; LEYDESDORFF; BORNMANN, 2016).

## REFERÊNCIAS

- BORNMANN, L. Field classification of publications in Dimensions: a first case study testing its reliability and validity. *Scientometrics*, Budapest, v. 117, n. 1, p. 637–640, 2018.
- HOU, J. Exploration into the evolution and historical roots of citation analysis by referenced publication year spectroscopy. *Scientometrics*, Budapest, v. 110, p. 1437–1452, 2017.
- LEYDESDORFF, L.; BORNMANN, L.; MARX, W.; MILOJEVIĆ, S. Referenced Publication Years Spectroscopy applied to iMetrics: *Scientometrics*, *Journal of Informetrics*, and a relevant subset of JASIST. *Journal of Informetrics*, Amsterdam, v. 8, n. 1, p. 162-174, 2014.
- LEYDESDORFF, L.; BORNMANN, L. The operationalization of “fields” as WoS subject categories (WCs) in evaluative bibliometrics: the cases of “library and information science”

and “science & technology studies”. **Journal of the Association for Information Science and Technology**, Hoboken, v. 67, n. 3, p. 707–714, 2016.

PECH, G.; DELGADO, C. Screening the most highly cited papers in longitudinal bibliometric studies and systematic literature reviews of a research field or journal: Widespread used metrics vs a percentile citation-based approach. **Journal of Informetrics**, Amsterdam, v. 15, n. 3, p. 101161, 2021.

SHU, F.; JULIEN, C. A.; ZHANG, L.; QIU, J.; ZHANG, J.; LARIVIÈRE, V. Comparing journal and paper level classifications of science. **Journal of Informetrics**, Amsterdam, v. 13, n. 1, p. 202–225, 2019.

THOR, A. *et al.* Identifying single influential publications in a research field: new analysis opportunities of the CRExplorer. **Scientometrics**, Budapest, v. 116, n.1, p. 591–608, 2018.

THOR, A.; BORNMANN, L.; HAUNSCHILD, R.; LEYDESDORFF, L. Which are the influential publications in the Web of Science subject categories over a long period of time? CRExplorer software used for big-data analyses in bibliometrics. **Journal of Information Science**, London, v. 47, n. 3, p. 419-428, 2021.