



10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



A INFLUÊNCIA DA BASE DE DADOS EM RANKINGS UNIVERSITÁRIOS: comparação entre *Web of Science* e *OpenAlex* no *Leiden Ranking*

Marília Augusta de Freitas

<https://orcid.org/0000-0003-0737-4542> | mariliaaf@unb.br
Universidade de Brasília (UnB)

Raquel Viana

<https://orcid.org/0009-0003-6517-0174> | raquelviana@unb.br
Universidade de Brasília (UnB)

Resumo:

Este estudo analisa os efeitos da substituição da base bibliográfica utilizada em rankings universitários sobre os resultados institucionais. A pesquisa compara os indicadores de universidades brasileiras em duas edições de um ranking internacional baseada em diferentes fontes de dados bibliográficas. Foram examinados indicadores de produção científica, impacto, colaboração e acesso aberto por meio de análise comparativa e correlações estatísticas. Os resultados indicam ampliação da cobertura da produção científica e elevada consistência nos indicadores absolutos, embora métricas percentuais apresentem maior sensibilidade às diferenças de cobertura das bases.

Palavras-Chave: rankings universitários. Avaliação científica. Bases bibliográficas. Produção científica. Bibliometria.

EIXO TEMÁTICO:

Bases e Fontes de Dados

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.1015

COMO CITAR ESTE ARTIGO:

FREITAS, Marília Augusta de; VIANA, Raquel. A influência da base de dados em rankings universitários: comparação entre Web of Science e OpenAlex no Leiden Ranking. In: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. **Anais [...]**. Curitiba, 2026. DOI: 10.22477/x.ebbc.1015. Disponível em: <https://doi.org/10.22477/x.ebbc.1015>.



1 INTRODUÇÃO

Os *rankings* universitários consolidaram-se nas últimas décadas como instrumentos relevantes de comparação e posicionamento institucional no ensino superior, influenciando políticas de internacionalização, estratégias de gestão acadêmica e diagnósticos de desempenho científico. Apesar das críticas quanto à sua capacidade limitada de representar a diversidade das missões universitárias, esses *rankings* continuam exercendo forte influência simbólica e política, funcionando como parâmetros de visibilidade e excelência no sistema global de ciência e tecnologia (Woicolesco, Cassol-Silva e Morosini, 2022; Goulart, Liboni e Cezarino, 2021).

Uma dimensão central desse debate refere-se às bases bibliográficas que sustentam os indicadores utilizados nos *rankings*. A maioria dos *rankings* globais utiliza bases proprietárias, como *Web of Science* e *Scopus*, cujos critérios de indexação tendem a privilegiar periódicos em língua inglesa, áreas STEM e países do Norte Global. Como consequência, parte significativa da produção científica de países do Sul Global, bem como de áreas como as ciências humanas e sociais, permanece sub-representada nesses sistemas (Miguel, Tannuri de Oliveira e Grácio, 2021; Digiampietri et al., 2025). Evidências empíricas indicam que apenas uma fração da produção científica registrada em sistemas nacionais é efetivamente indexada nessas bases, o que pode afetar diretamente indicadores de produção e impacto utilizados em *rankings* universitários (Digiampietri et al., 2025).

Nos últimos anos, as bases de dados abertas de dados científicos têm ganhado destaque como alternativas às bases proprietárias. Nesse contexto, o *OpenAlex* foi desenvolvido para oferecer acesso aberto a dados bibliográficos e documentação metodológica, alinhando-se aos princípios da ciência aberta. Embora apresente vantagens em termos de transparência e acesso, estudos apontam desafios re-

lacionados à padronização, consistência de metadados e estabilidade de cobertura quando comparado a bases com maior controle curatorial, como a *Web of Science* (Mongeon et al., 2024; Zheng et al., 2025). Em contextos avaliativos, essas diferenças podem produzir efeitos relevantes sobre indicadores bibliométricos e interpretações de desempenho institucional.

O *Leiden Ranking* ocupa posição singular nesse cenário por adotar uma abordagem estritamente bibliométrica, baseada exclusivamente em publicações e citações, sem incorporar indicadores reputacionais ou dados autodeclarados pelas instituições. Desenvolvido pelo *Centre for Science and Technology Studies* da Universidade de Leiden, o ranking destaca-se pela transparência metodológica e pelo uso de indicadores normalizados de impacto, colaboração e acesso aberto (Kochetkov, 2024).

A partir de 2024, o *Leiden Ranking* passou a disponibilizar uma edição baseada em dados do *OpenAlex*, mantendo inalterados os indicadores e procedimentos bibliométricos. Essa mudança cria uma oportunidade analítica para avaliar empiricamente os efeitos da substituição da base bibliográfica sobre os resultados institucionais.

Diante disso, este estudo investiga em que medida a substituição da base bibliográfica, da *Web of Science* para o *OpenAlex*, afeta a mensuração do desempenho das universidades brasileiras no *Leiden Ranking*. O objetivo é analisar a robustez e a sensibilidade dos resultados institucionais a essa substituição, comparando os resultados das universidades brasileiras nas duas edições do ranking e discutindo as implicações metodológicas do uso de bases de dados abertas na avaliação científica.

2 PROCEDIMENTOS METODOLÓGICOS

Este estudo adota uma abordagem quantitativa e comparativa, com base em indicadores bibliométricos do CWTS *Leiden Ranking*, com o objetivo de analisar a robustez e a sensibilidade dos resultados das universidades brasileiras à substituição da base bibliográfica subjacente ao *ranking*.

Foram analisadas duas edições do *Leiden Ranking*: a *Traditional Edition*, baseada na *Web of Science*, e a *Open Edition*, baseada no *OpenAlex*. Em ambas, manteve-se constante o arcabouço metodológico do ranking, considerando-se exclusivamente o campo *All Sciences*, o período 2020–2023, o método de contagem fracionada (*fractional counting*) e, no caso da *Open Edition*, apenas as *core publications*, de modo a garantir a comparabilidade conceitual entre as edições.

Os dados foram extraídos a partir das tabelas oficiais do ranking, organizadas nas dimensões de impacto científico, colaboração e acesso aberto. O conjunto analítico foi definido por junção interna (*inner join*) entre as edições, resultando em 39 universidades brasileiras presentes em ambas. A *Traditional Edition* (WoS) conta com $k_1 = 39$ instituições brasileiras, enquanto a *Open Edition* (OpenAlex) registra $k_2 = 75$, das quais 39 coincidem com a edição tradicional e 36 são exclusivas da edição aberta.

Foram considerados indicadores absolutos e percentuais amplamente utilizados em *rankings* universitários, incluindo: número total de publicações (P), número e proporção de publicações entre as 10% mais citadas (P(top 10%) e PP(top 10%)), indicadores de colaboração (P(collab) e PP(collab)) e de acesso aberto (P(OA) e PP(OA)). Não foram realizadas agregações adicionais, imputações de valores ausentes ou transformações nos indicadores originais.

A comparação entre as edições foi realizada por meio de correlações de Spearman (ρ), método apropriado para a análise de posicionamento relativo e distribuições assimétricas típicas de indicadores bibliométricos. Também foram calculados, de forma complementar, as outras três formas de análise para a comparação de *rankings* universitários propostas por Bar-Ilan et al. (2007), Overlap (O), Footrule norm (F), Fagin (G) e a Medida de rank inverso (M). Em conjunto, essas quatro medidas permitem uma melhor análise sobre a estabilidade ordinal, a sobreposição institucional e a sensibilidade das posições de topo do ranking à mudança de base bibliográfica.

3 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A análise concentrou-se na comparação dos resultados das universidades brasileiras entre as edições *Traditional* e *Open do CWTS Leiden Ranking*. Essa estratégia permitiu avaliar a consistência dos indicadores diante da substituição da base bibliográfica subjacente ao ranking, mantendo-se constantes os demais parâmetros metodológicos.

3.1 COBERTURA E VOLUME DE PUBLICAÇÕES

A sobreposição institucional (O) entre as duas edições é de $O = 0,5200$, indicando que 39 das 75 universidades presentes na *Open Edition* (52%) também figuram na *Traditional Edition*. Esse nível de sobreposição é estruturalmente condicionado pelo próprio escopo das edições: a *WoS* indexa um conjunto menor de periódicos, naturalmente resultando em um universo institucional mais restrito. Observou-se também uma diferença expressiva no volume total de publicações atribuídas às universidades brasileiras entre as duas edições analisadas. A *Open Edition* registrou 139.030 publicações, enquanto a *Traditional Edition* contabilizou 116.896 no mesmo período (2020–2023), representando um acréscimo aproximado de 18,9%

na cobertura da produção científica.

Esse resultado indica que o *OpenAlex* captura um universo mais amplo de publicações em comparação à *Web of Science*, convergindo com evidências da literatura que apontam que bases abertas ampliam a visibilidade da produção científica de países do Sul Global, frequentemente sub-representada em bases proprietárias (Digiampietri et al., 2025). Essa diferença de cobertura constitui, portanto, um elemento contextual relevante para a interpretação dos resultados comparativos entre as duas edições do ranking.

3.2 RELAÇÕES ENTRE INDICADORES ABSOLUTOS

Para o indicador P (total de publicações), os resultados são consistentemente altos: $F = 0,9066$, $M = 0,9063$ e $p = 0,9754$ ($p < 0,001$). Esses valores indicam que a hierarquia das universidades brasileiras por volume de produção científica é praticamente idêntica nas duas edições. A medida M, por sua vez, confirma que essa estabilidade se mantém inclusive nas primeiras posições do ranking, as três maiores universidades brasileiras (USP, Unesp e Unicamp) ocupam exatamente as mesmas posições nas duas edições.

Para o par $P(\text{top } 10\%)$ vs $P(\text{collab})$, que compara indicadores absolutos de natureza distintas entre as edições, os valores permanecem elevados: $F = 0,8711$, $M = 0,8447$ e $p = 0,9540$ ($p < 0,001$). Mesmo comparando indicadores diferentes (impacto científico na Traditional e colaboração na Open) a ordenação das universidades apresenta forte consistência, sugerindo que o tamanho institucional é o principal determinante do posicionamento relativo em ambas as edições, independentemente do indicador absoluto considerado.

3.3 CORRELAÇÃO ENTRE INDICADORES PERCENTUAIS

Em contraste com os indicadores absolutos, os indicadores percentuais revelam padrão de maior sensibilidade à substituição da base bibliográfica, com graus de variação que diferem substancialmente entre os pares analisados. Para os indicadores de colaboração, a proporção de publicações colaborativas (PP(collab)) apresentou $p = 0,8749$ ($p < 0,001$) na comparação com PP(top 10%), indicando consistência moderada-alta na ordenação das universidades. O caso mais extremo, porém, é o do par PP(top 10%) vs PP(collab), cujos resultados revelam padrão radicalmente distinto dos demais: a medida F cai para 0,2934, o valor de M torna-se negativo (-0,0335) e a correlação de Spearman é de $p = -0,1296$, não significativa e próxima de zero. Em termos práticos, isso significa que os *rankings* ordenados pelas proporções percentuais das duas edições são essencialmente sem relação entre si como ordenações. Esse resultado não indica inconsistência das bases, mas sim que indicadores percentuais de natureza diferente, impacto e colaboração, produzidos por infraestruturas distintas, não devem ser diretamente comparados como ordenações. De modo mais amplo, esses achados convergem com estudos que apontam que métricas percentuais são mais suscetíveis a diferenças de cobertura, critérios de indexação e padronização de metadados entre bases de dados científicas (Mongeon et al., 2024; Zheng et al., 2025).

3.3 DIFERENÇAS MÉDIAS ENTRE AS EDIÇÕES

A análise das diferenças médias entre os indicadores percentuais reforça a distinção entre estabilidade e sensibilidade observada entre as edições. A proporção média de publicações entre as 10% mais citadas apresentou variação mínima (+0,11 ponto percentual), enquanto a proporção de publicações colaborativas foi inferior na *Open Edition* (-3,35 pontos percentuais). Em contraste, a proporção

média de publicações em acesso aberto foi substancialmente superior na edição baseada no *OpenAlex*, com acréscimo aproximado de 10,10 pontos percentuais. Esse resultado converge com estudos que apontam maior sensibilidade das métricas de acesso aberto às bases de dados e aos esquemas de classificação adotados por bases abertas, sugerindo que o *OpenAlex* tende a capturar de forma mais abrangente a diversidade de modelos de publicação em acesso aberto (Zheng et al., 2025).

3.4 ESTABILIDADE E MUDANÇAS NO POSICIONAMENTO INSTITUCIONAL

A elevada correlação observada nos indicadores absolutos reflete-se também na estabilidade do posicionamento das universidades de maior porte. As instituições brasileiras com maior volume de publicações mantiveram posições praticamente idênticas nas duas edições do ranking, indicando que a substituição da base bibliográfica não altera a hierarquia das universidades líderes em produção científica no país. Em contraste, universidades posicionadas fora do núcleo superior do ranking apresentaram variações mais perceptíveis, ainda que geralmente limitadas a poucas posições.

A análise das variações de posição para o ranking por P (total de publicações) entre as duas edições revela que a média do deslocamento absoluto é de apenas 2,15 posições, com máximo de 7 posições. Esse padrão reforça a análise de Kochetkov (2024), segundo a qual *rankings* baseados em indicadores bibliométricos absolutos tendem a reproduzir hierarquias institucionais consolidadas, enquanto ajustes metodológicos ou de cobertura afetam de forma mais pronunciada instituições situadas em faixas intermediárias da distribuição.

4 CONSIDERAÇÕES FINAIS

A temática central deste artigo discutiu o papel dos dados bibliográficos que sustentam os indicadores utilizados em *rankings* universitários, tradicionalmente baseados em bases de dados proprietárias. A partir da comparação entre uma base aberta (*OpenAlex*) e uma base proprietária (*Web of Science*), o estudo analisou empiricamente os efeitos da substituição da base bibliográfica nos resultados institucionais. Os resultados evidenciaram ampliação da cobertura da produção científica das universidades brasileiras na base aberta, com aumento aproximado de 18,9% no volume de publicações, indicando maior capacidade de captar parcelas da produção sub-representadas em bases proprietárias. Apesar dessa diferença de cobertura, as medidas de similaridade de *rankings* e as correlações de Spearman indicaram elevada consistência entre as edições para os indicadores absolutos, sugerindo que a substituição da base não altera de forma substantiva o posicionamento relativo das universidades brasileiras quando consideradas contagens agregadas de produção científica. Essa consistência, contudo, não se estende aos indicadores percentuais: as métricas relativas apresentam sensibilidade significativa à mudança de base bibliográfica, chegando, no caso de PP(top 10%) vs PP(collab), à incompatibilidade ordinal completa, evidência de que indicadores percentuais de natureza distinta, produzidos por bases diferentes, não devem ser diretamente comparados como ordenações institucionais. De modo geral, os resultados indicam que fontes de dados bibliográficas abertas podem ampliar a visibilidade da produção científica sem comprometer a estabilidade estrutural dos *rankings*, embora diferenças de cobertura influenciem de forma pronunciada a composição da produção e os indicadores relativos, reforçando a importância de considerar explicitamente a base de dados na construção de indicadores avaliativos e no debate sobre transparência e democratização da avaliação científica.



REFERÊNCIAS

BAR-ILAN, Judit; LEVENE, Mark; LIN, Ayelet. Some measures for comparing citation databases. **Journal of Informetrics**, v. 1, n. 1, p. 26-34, 2007. DOI: <https://doi.org/10.1016/j.joi.2006.08.001>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1751157706000058?via%3Dihub>. Acesso em: 06 maio 2026.

DIGIAMPIETRI, Luciano; GALLARDO, Osvaldo; BARANGER, Denis; BEIGEL, Fernanda. Approaching bibliometrics and prosopography: the comprehensive publishing landscape of CNPq (Brazil) and CONICET (Argentina) and its coverage in global databases. **Encontros Bibli**, Florianópolis, v. 30, e103528, 2025. DOI: <https://doi.org/10.5007/1518-2924.2025.e103528>. Disponível em: <https://www.scielo.br/j/eb/a/Nd4hT7CCJCnRqsSVvn39Tnr/?lang=en>. Acesso em: 10 fev. 2026.

GOULART, Vera; LIBONI, Lara; CEZARINO, Luciana. Balancing skills in the digital transformation era: The future of jobs and the role of higher education. **Industry and Higher Education**, 2021. DOI: <https://doi.org/10.1177/09504222211029796>. Disponível em: <https://journals.sagepub.com/doi/10.1177/09504222211029796>. Acesso em: 10 fev. 2026.

KOCHETKOV, Dmitry. University rankings in the context of research evaluation: a state-of-the-art review. **Quantitative Science Studies**, Cambridge, v. 5, n. 1, p. 1-23, 2024. DOI: https://doi.org/10.1162/qss_a_00317. Disponível em: https://direct.mit.edu/qss/article-pdf/5/3/533/2472669/qss_a_00317.pdf. Acesso em: 10 fev. 2026.

MIGUEL, Sandra; TANNURI DE OLIVEIRA, Ely Francina; CABRINI GRÁCIO, Maria Claudia. Scientific production on open access: A worldwide bibliometric analysis in the academic and scientific context. **Publications**, v. 9, n. 1, p. 12, 2021. DOI: <https://doi.org/10.3390/publications4010001>. Disponível em: <https://www.mdpi.com/2304-6775/4/1/1>. Acesso em: 10 fev. 2026.

MONGEON, Philippe et al. **Investigating Document Type, Language, Publication Year, and Author Count Discrepancies Between OpenAlex and Web of Science**. Dalhousie University, 2024. DOI: <https://doi.org/10.48550/arXiv.2508.18620>. Disponível em: <https://arxiv.org/abs/2508.18620>. Acesso em: 10 fev. 2026.

WOICOLESCO, Vanessa Gabrielle; CASSOL-SILVA, Carla Camargo; MOROSINI, Marília. Internationalization at Home and Virtual: a Sustainable Model for Brazilian Higher Education. **Journal of Studies in International Education**, v. 26, n. 2, p. 222-239, 2022. DOI: 10.1177/10283153221076898. Disponível em: <https://journals.sagepub.com/doi/10.1177/10283153221076898>. Acesso em: 10 fev. 2026.

ZHENG, Mengxue; MIAO, Lili; BU, Yi; LARIVIÈRE, Vincent. Understanding discrep-



10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



ancies in the coverage of OpenAlex: The case of China. **Journal of the Association for Information Science and Technology**, v. 76, n. 11, p. 1591–1601, 2025. DOI: <https://doi.org/10.1002/asi.70013>. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70013>. Acesso em: 10 fev. 2026.