

## **Perfil institucional da produção científica de docentes da pós-graduação stricto sensu da UFRJ (2017-2020) em diferentes bancos de dados**

### Introdução

A cientometria foi a princípio definida como a “aplicação de métodos quantitativos que lidam com a análise da ciência enquanto processo informacional” (Nalimov and Mul'chenko 1969). Atualmente, ela se ocupa dos mais diversos temas, incluindo a mensuração de qualidade/impacto da pesquisa; análise e entendimento dos processos de citações; e identificação de tópicos relevantes/emergentes ou da estrutura conceitual de um campo de estudo (Mingers and Leydesdorff 2015; Aria and Cuccurullo 2017).

Com o aumento da produção científica, o processo de peer-review, oneroso e sujeito a vieses e subjetividade (Rose 1989), começa a ser usado mais estrategicamente. Alternativas econômicas e objetivas *para a avaliação da performance acadêmica nos mais diversos níveis se tornam então uma necessidade para gestores da área de Ciência e Tecnologia (CT). Nesse contexto, surge a aplicação da ciëntometria de maior destaque: o uso de suas métricas de avaliação de produção científica na tomada de decisão e definição de políticas nos mais diversos âmbitos da gestão em CT* (Mingers and Leydesdorff 2015).

O uso de métricas na avaliação da produção científica influencia diretamente a atuação e carreira dos pesquisadores. A obtenção de um cargo como professor universitário, financiamentos para projetos e bolsas de estudos para alunos dependem cada vez mais de avaliações quantitativas a nível de indivíduo, programa de pós-graduação ou mesmo instituição. No Brasil, por exemplo, temos que dentre as 49 áreas definidas pelo Qualis Periódicos, 29 utilizam o Fator de Impacto (JCR Impact Factor) como principal definidor de classificação para seus periódicos (Oliveira and Amaral 2017). Por sua vez, a CAPES utiliza o Qualis Periódicos para avaliar programas de pós-graduação.

Entretanto, não apenas as métricas podem ser utilizadas de forma inadequada (Pendlebury 2009; Haustein and Larivière 2015), como seu cálculo é altamente influenciado pelo banco de dados utilizado (Garner et al. 2018). Assim sendo, é muito importante se atentar não só à métrica utilizada, como também a características da plataforma usada para o cálculo da mesma (e.g. a cobertura de publicações por campo). A escolha da fonte informacional usada é um passo essencial para que a aplicação da métrica calculada seja o mais adequada quanto possível.

Algumas das maiores fontes informacionais para análise da comunicação científica englobam bancos de dados acadêmicos multidisciplinares internacionais, como Scopus e Web of Science, amplamente usados em trabalhos cientométricos. Entretanto, novos bancos de dados, como o Dimensions () e Microsoft Academic () surgiram recentemente e vêm ganhando

projeção. Em especial, a plataforma Dimensions é cada vez mais utilizada em pesquisas científicas, dada sua enorme cobertura em termos de Journals (Singh et al. 2021) e por também indexar preprints (Fraser et al. 2021).

Assim sendo, iremos realizar um estudo de caso. Recuperaremos a produção científica (limitadas a artigos de pesquisa originais e revisões) dos docentes cadastrados em programas de pós-graduação *stricto sensu* da UFRJ durante os anos 2017-2020 e iremos checar a cobertura da mesma nas quatro plataformas supracitadas. Para além disso, compararemos métricas calculadas com base nas publicações recuperadas em cada uma das plataformas. Nossas análises discriminarão também os resultados específicos para cada grande área do conhecimento, permitindo comparativos da cobertura diferencial dos bancos de dados e das métricas calculadas por disciplina. Desta forma, teremos um panorama geral de quais bancos de dados e métricas seriam os mais adequados para a avaliação de disciplinas específicas da instituição, recurso valioso para órgãos de fomento e gestão.

### Objetivos

- Obter a produção científica dos docentes cadastrados em programas de pós-graduação *stricto sensu* da UFRJ;
- Averiguar o quanto da produção científica está contemplada por quatro diferentes bancos de dados bibliométricos (Scopus, Web of Science, Dimensions e Microsoft Academic);
- Calcular métricas com base no acervo de cada um dos bancos de dados e compará-las entre si
- Analisar os dados por grande área para averiguar diferenças disciplinares
- Disponibilizar todos os dados e análises publicamente (Github/Zenodo)

### Abordagem experimental

Uma tabela contendo os docentes vinculados à pós-graduação *stricto sensu* no Brasil entre 2017 e 2020 será obtida através da plataforma “Dados Abertos CAPES” (<https://dadosabertos.capes.gov.br/dataset/2017-a-2020-docentes-da-pos-graduacao-stricto-sensu-no-brasil>). Esse arquivo .csv será filtrado de forma a manter apenas docentes da UFRJ, gerando uma lista que permitirá a recuperação da produção de cada docente no período de interesse através da Plataforma Lattes (<https://lattes.cnpq.br/>). Esse processo será semi-automático, realizado através do ScriptLattes (Mena-Chalco and Cesar Junior 2009). Somente produções que contenham um identificador único (doi) serão mantidas para etapas posteriores.

Em seguida, as publicações serão procuradas pelo seu identificador em cada uma das grandes plataformas de informação bibliográfica, obtendo-se assim uma estimativa da cobertura de cada área. Isso será feito automaticamente através de código escrito em Python,

que irá se comunicar com as plataformas via suas respectivas APIs. Para as plataformas que disponibilizam dados de citação, métricas como os índices h, g e m serão calculados para a instituição, grandes áreas e programas de pós-graduação específicos usando o pacote do R Bibliometrix (Aria and Cuccurullo 2017). Para as análises comparativas entre disciplinas/PPGs, a produção de professores atuantes em mais de uma grande área serão excluídos.

Fora isso, o ambiente de programação R e o programa VosViewer (Eck and Waltman 2009) serão utilizados para mapeamentos/análises adicionais da produção científica dos docentes da instituição no recorte temporal estipulado, como acoplamento bibliográfico e análises de co-autoria. Esses resultados proverão um rico panorama da produção da UFRJ, índices bibliométricos associados a elas e o efeito dos bancos de dados na obtenção das mesmas, um recurso valioso para a tomada de decisão/gestão institucional de forma mais informada e crítica.

### Referências Bibliográficas

- Aria, Massimo, and Corrado Cuccurullo. 2017. "Bibliometrix : An R-Tool for Comprehensive Science Mapping Analysis." *Journal of Informetrics* 11 (4): 959–75. <https://doi.org/10.1016/j.joi.2017.08.007>.
- Bornmann, Lutz, and Rüdiger Mutz. 2011. "Further Steps towards an Ideal Method of Measuring Citation Performance: The Avoidance of Citation (Ratio) Averages in Field-Normalization." *Journal of Informetrics* 5 (1): 228–30. <https://doi.org/10.1016/j.joi.2010.10.009>.
- Eck, Nees van, and Ludo Waltman. 2009. "Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping." *Scientometrics* 84 (2): 523–38. <https://doi.org/10.1007/s11192-009-0146-3>.
- Fraser, Nicholas, Liam Brierley, Gautam Dey, Jessica K. Polka, Máté Pálfi, Federico Nanni, and Jonathon Alexis Coates. 2021. "The Evolving Role of Preprints in the Dissemination of COVID-19 Research and Their Impact on the Science Communication Landscape." *PLOS Biology* 19 (4): e3000959. <https://doi.org/10.1371/journal.pbio.3000959>.
- Garner, Rebecca M., Joshua A. Hirsch, Felipe C. Albuquerque, and Kyle M. Fargen. 2018. "Bibliometric Indices: Defining Academic Productivity and Citation Rates of Researchers, Departments and Journals." *Journal of Neurointerventional Surgery* 10 (2): 102–6. <https://doi.org/10.1136/neurintsurg-2017-013265>.
- Gingras, Yves. 2016. *Bibliometrics and Research Evaluation: Uses and Abuses*. MIT Press.
- Haustein, Stefanie, and Vincent Larivière. 2015. "The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects." In *Incentives and Performance*, edited by Isabell M. Welp, Jutta Wollersheim, Stefanie Ringelhan, and Margit Osterloh, 121–39. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-09785-5\\_8](https://doi.org/10.1007/978-3-319-09785-5_8).
- Mena-Chalco, Jesús Pascual, and Roberto Marcondes Cesar Junior. 2009. "ScriptLattes: An Open-Source Knowledge Extraction System from the Lattes Platform." *Journal of the Brazilian Computer Society* 15 (December): 31–39. <https://doi.org/10.1007/BF03194511>.
- Mingers, John, and Loet Leydesdorff. 2015. "A Review of Theory and Practice in Scientometrics." *European Journal of Operational Research* 246 (1): 1–19. <https://doi.org/10.1016/j.ejor.2015.04.002>.
- Nalimov, V., and Z. M. Mul'chenko. 1969. "Naukometriya, the Study of the Development of Science as an Information Process in Russian." Nauka, Moscow.

- Oliveira, TM de, and Livio Amaral. 2017. "Public Policies in Science and Technology in Brazil: Challenges and Proposals for the Use of Indicators in Evaluation." *En: Mugnaini, R*, 189–217.
- Pendlebury, David A. 2009. "The Use and Misuse of Journal Metrics and Other Citation Indicators." *Archivum Immunologiae et Therapiae Experimentalis* 57 (1): 1–11.  
<https://doi.org/10.1007/s00005-009-0008-y>.
- Rose, Suzanna. 1989. "Women Biologists and the 'Old Boy' Network." *Women's Studies International Forum*, Feminism and science: In memory of Ruth Bleier, 12 (3): 349–54.  
[https://doi.org/10.1016/S0277-5395\(89\)80011-1](https://doi.org/10.1016/S0277-5395(89)80011-1).
- Singh, Vivek Kumar, Prashasti Singh, Mousumi Karmakar, Jacqueline Leta, and Philipp Mayr. 2021. "The Journal Coverage of Web of Science, Scopus and Dimensions: A Comparative Analysis." *Scientometrics* 126 (6): 5113–42. <https://doi.org/10.1007/s11192-021-03948-5>.