



10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



ESTRUTURA DA PRODUÇÃO CIENTÍFICA DA PÓS-GRADUAÇÃO BRASILEIRA:

análise cientométrica de docentes a partir da Plataforma Lattes

Guilherme Mascarenhas Dias

<https://orcid.org/0009-0001-7760-4692>
Guilhermemdias2501@gmail.com
Instituto Federal de Minas Gerais (IFMG)

Thiago Magela Rodrigues Dias

<https://orcid.org/0000-0001-5057-9936>
thiagomagela@cefetmg.br
Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)

Patrícia Mascarenhas

<https://orcid.org/0000-0002-8448-6874>
patriciamdias@gmail.com
Universidade do Estado de Minas Gerais (UEMG)

Resumo:

Este estudo apresenta o desenvolvimento de um arcabouço computacional destinado à análise em larga escala da produção científica de docentes vinculados à pós-graduação brasileira. A metodologia baseia-se na extração e processamento de currículos da Plataforma Lattes, permitindo a construção de uma base analítica estruturada para geração de indicadores bibliométricos e cientométricos. A partir dessa base são analisados padrões de produtividade científica, distribuição institucional e redes de colaboração acadêmica. Os resultados evidenciam a presença de forte concentração institucional da produção científica e padrões estruturais de colaboração entre pesquisadores no sistema científico brasileiro.

Palavras-Chave: Plataforma Lattes. Cientometria. Pós-graduação brasileira. Redes de colaboração científica. Produção científica.

EIXO TEMÁTICO:

Produtividade e Colaboração Científica

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.1172

COMO CITAR ESTE ARTIGO:

DIAS, Guilherme Mascarenhas; DIAS, Thiago Magela Rodrigues; MASCARENHAS, Patrícia. Estrutura da produção científica da pós-graduação brasileira: análise cientométrica de docentes a partir da Plataforma Lattes. In: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. **Anais [...]**. Curitiba, 2026. DOI: 10.22477/x.ebbc.1172. Disponível em: <https://doi.org/10.22477/x.ebbc.1172>.



1 INTRODUÇÃO

De acordo com Ferreira (2010), a ciência produz conhecimento havendo a necessidade e o compromisso de tornar seus resultados públicos, a partir da divulgação junto à comunidade científica das atividades realizadas durante a pesquisa. Candoni (2002), por sua vez, destaca que disponibilizar novas descobertas e inovações originadas de pesquisas científicas é um dos principais objetivos da ciência, independentemente da área considerada.

Nesse cenário, a Bibliometria (Araújo, 2006), que possui grande relação com a Cientometria (Silva *et al.*, 2011), se destaca como uma das principais ciências métricas de análise de conteúdo, podendo ser aplicada a fontes de dados científicos com o intuito de se obter informações quantitativas sobre publicações (Pritchard, 1969).

No âmbito da pós-graduação brasileira, os programas reconhecidos e avaliados pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) representam o principal espaço institucional de produção científica e formação de pesquisadores no país. Os docentes vinculados a esses programas desempenham papel central na geração de conhecimento científico, na formação de recursos humanos qualificados e na consolidação de redes de colaboração acadêmica. Dessa forma, a análise das atividades desses pesquisadores permite compreender importantes aspectos da dinâmica científica nacional.

Diante desse cenário, este trabalho apresenta o desenvolvimento de um arcabouço computacional destinado à análise de currículos de docentes da pós-graduação brasileira. O objetivo principal consiste em estruturar um processo metodológico capaz de extrair informações relevantes dos currículos Lattes, organizar

esses dados em uma base analítica e gerar indicadores cientométricos e redes de colaboração científica que permitam caracterizar a estrutura da produção científica desses pesquisadores.

2 METODOLOGIA

A metodologia adotada neste estudo baseia-se na construção de um arcabouço computacional voltado ao processamento em larga escala de currículos da Plataforma Lattes. Esse *pipeline* foi projetado para lidar com grandes volumes de dados mantendo eficiência computacional e baixo consumo de memória, permitindo a análise sistemática de currículos de docentes vinculados a programas de pós-graduação reconhecidos pela CAPES.

O processo metodológico inicia-se com a identificação dos docentes vinculados a programas de pós-graduação reconhecidos pela CAPES, identificados a partir do Portal de Dados Abertos da CAPES no período de 2021 a 2024. Em seguida, foram identificados os currículos cadastrados na Plataforma Lattes referentes a esse conjunto de docentes, resultando em 98.692 docentes. Consequentemente, foi realizada a leitura incremental dos currículos identificados. Essa estratégia de processamento incremental permite que cada currículo seja lido, interpretado e processado individualmente, evitando o carregamento simultâneo de grandes volumes de dados na memória do sistema.

Além das informações institucionais e identificadoras, o arcabouço realiza a extração e quantificação de diferentes tipos de produção acadêmica registrados nos currículos. Entre os elementos analisados encontram-se artigos completos publicados em periódicos científicos, trabalhos apresentados em eventos acadêmicos, livros publicados ou organizados, capítulos de livros, orientações acadêmicas concluídas, orientações em andamento e projetos de pesquisa registrados pelos pes-

quisadores. Esses elementos são utilizados para caracterizar o perfil de produção científica dos docentes analisados.

Após a extração dos dados, as informações são armazenadas em uma base analítica estruturada utilizando o sistema de gerenciamento de banco de dados SQLite. A partir dessa base estruturada são gerados diversos indicadores bibliométricos e cientométricos que caracterizam o conjunto de docentes analisados. Inicialmente são realizadas análises sobre a distribuição temporal de atualização dos currículos, permitindo avaliar o grau de atualização das informações acadêmicas registradas na Plataforma Lattes.

Outra dimensão explorada refere-se à distribuição disciplinar dos docentes, obtida a partir da primeira grande área de atuação declarada nos currículos. Essa informação permite observar como os docentes se distribuem entre diferentes áreas do conhecimento e possibilita a identificação de padrões disciplinares na estrutura da pós-graduação brasileira.

O arcabouço também realiza análises sobre a estratificação da produção científica dos docentes, considerando diferentes tipos de produção acadêmica registrados nos currículos. Essas análises permitem observar padrões de produtividade científica e identificar variações entre pesquisadores em relação à quantidade e ao tipo de produção registrada.

Além disso, a distribuição da produtividade científica entre os docentes é analisada a partir de modelos de distribuição de potência, frequentemente observados em sistemas científicos. Nesse contexto é utilizada uma abordagem baseada na distribuição de Zipf/Pareto para examinar a relação entre o ranking de pesquisadores e o volume de sua produção científica, permitindo identificar padrões de concentração e desigualdade na produtividade acadêmica.

A etapa final da metodologia consiste na construção de redes de colaboração científica entre os docentes analisados. A identificação dessas colaborações é realizada a partir da comparação de títulos de publicações presentes em diferentes currículos. Para reduzir a ocorrência de falsos positivos, os títulos das publicações são previamente normalizados e concatenados ao ano da publicação, formando uma chave única de identificação da produção científica. Quando essa chave é encontrada em mais de um currículo, estabelece-se uma conexão entre os pesquisadores correspondentes.

A rede resultante é exportada em formatos compatíveis com ferramentas especializadas em análise de redes, como o *software* Gephi, permitindo a aplicação de métricas de centralidade, identificação de comunidades científicas e análise da estrutura de colaboração entre pesquisadores.

3 RESULTADOS

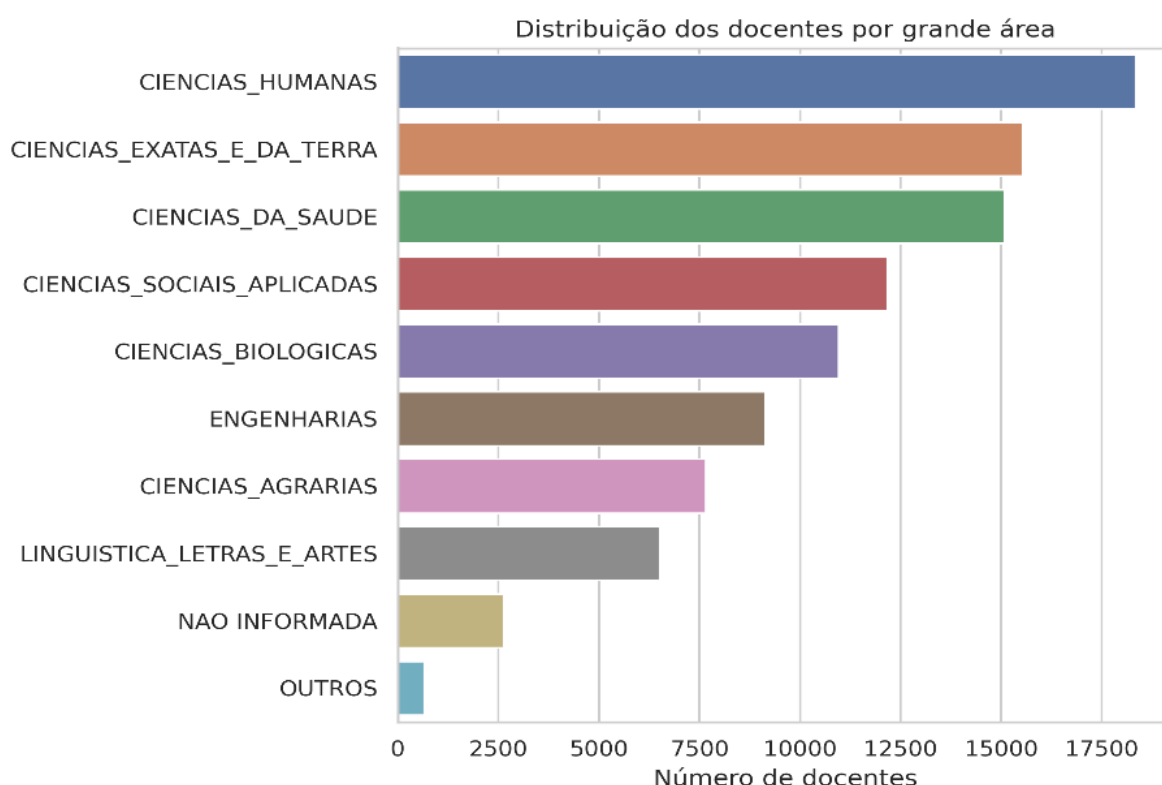
O processo de identificação dos docentes resultou em um conjunto inicial de 98.692 pesquisadores vinculados a programas de pós-graduação brasileiros. A partir desse conjunto foram recuperados os respectivos currículos da Plataforma Lattes, que constituem a principal fonte de dados utilizada neste estudo. Após o processo de leitura e extração das informações, foi possível estruturar uma base analítica contendo registros detalhados de produção científica, vínculos institucionais e histórico de atualização curricular.

A análise da distribuição temporal de atualização dos currículos revela padrões relevantes relacionados ao comportamento dos docentes da pós-graduação brasileira em relação à manutenção de seus registros acadêmicos na Plataforma Lattes. Observa-se uma tendência de concentração das atualizações em anos mais recentes, sugerindo uma forte relação entre a atualização curricular e ciclos insti-

tucionais de avaliação científica, especialmente aqueles associados aos processos de avaliação periódica conduzidos pela CAPES. Esse comportamento indica que a atualização dos currículos não ocorre de forma homogênea ao longo do tempo, mas tende a intensificar-se em períodos próximos a processos de avaliação institucional ou submissão de relatórios acadêmicos.

A distribuição dos docentes por grandes áreas do conhecimento (Figura 1) evidencia a diversidade disciplinar presente na pós-graduação brasileira. Entretanto, essa distribuição não é uniforme, revelando diferenças importantes na presença relativa de docentes entre áreas científicas. Algumas áreas concentram maior número de pesquisadores, refletindo tanto a expansão histórica de determinados campos científicos quanto as prioridades institucionais e políticas científicas estabelecidas ao longo do desenvolvimento do sistema de pós-graduação brasileiro.

Figura 01 – Distribuição dos docentes de pós-graduação por grande área de atuação.



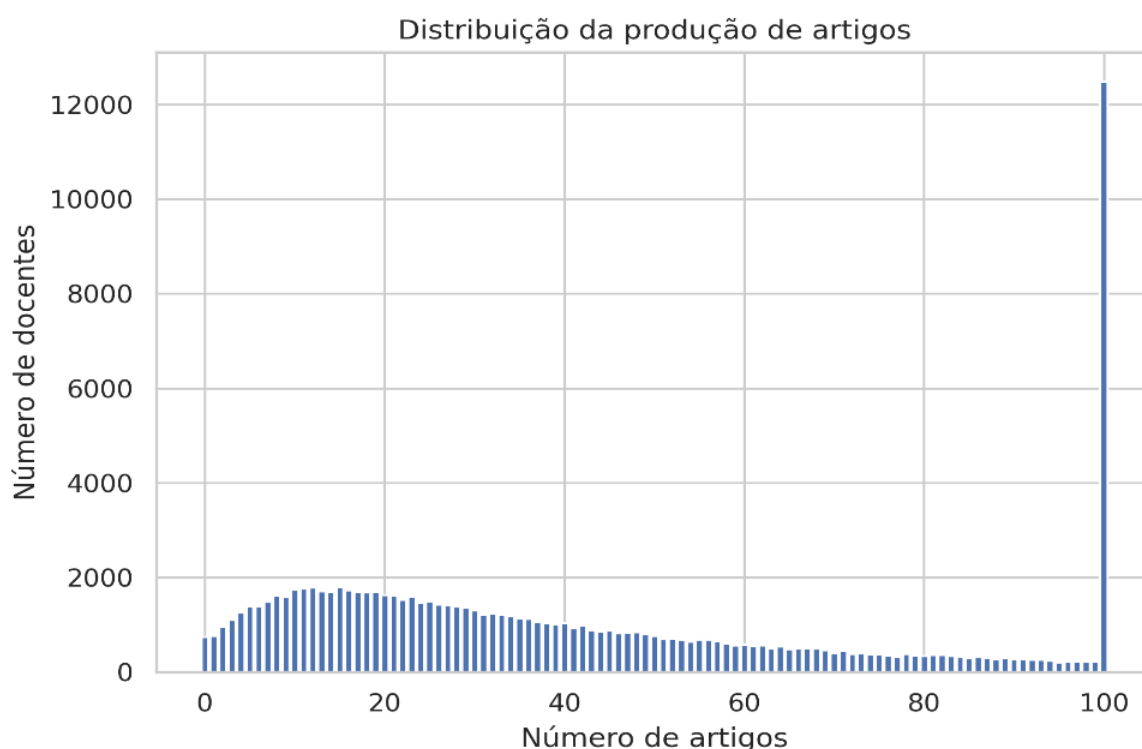
Fonte: Os autores (2026).

A análise da distribuição institucional dos docentes revela um padrão significativo de concentração em um conjunto relativamente reduzido de instituições de ensino superior. Universidades públicas tradicionais, particularmente universidades federais e estaduais consolidadas, concentram parcela significativa dos docentes vinculados à pós-graduação brasileira.

Esse resultado sugere que o sistema científico brasileiro apresenta uma estrutura fortemente hierarquizada, na qual algumas instituições atuam como polos centrais de produção científica e formação de pesquisadores, enquanto outras desempenham papéis mais periféricos. Esse padrão tem implicações importantes para políticas de ciência e tecnologia, especialmente no que se refere à distribuição de recursos e à expansão do sistema científico nacional.

A distribuição da produção de artigos científicos (Figura 2) revela a presença de uma cauda longa de produtividade. A maioria dos docentes apresenta produção científica moderada, enquanto um subconjunto relativamente pequeno apresenta níveis significativamente mais elevados de produção. Esse padrão reforça a presença de heterogeneidade na produtividade científica e sugere que a produção acadêmica não está distribuída de maneira uniforme entre pesquisadores.

Figura 02 – Distribuição da produção de artigos em periódicos dos docentes.



Fonte: Os autores (2026).

Esse comportamento é amplamente documentado em estudos bibliométricos e cientométricos e está associado a fenômenos como a vantagem cumulativa e o chamado efeito Mateus, nos quais pesquisadores mais produtivos tendem a ampliar progressivamente sua capacidade de produção científica ao longo do tempo. A presença desse padrão na análise dos docentes da pós-graduação brasileira indica que o sistema científico nacional apresenta características estruturais semelhantes às observadas em outros sistemas científicos internacionais.

A construção das redes de colaboração científica permitiu identificar padrões estruturais nas relações de coautoria entre docentes da pós-graduação brasileira. A análise das comunidades científicas identificadas por meio de algoritmos de detecção de comunidades (Figura 8) revela a presença de agrupamentos de pesquisadores que compartilham padrões recorrentes de colaboração acadêmica.

Esses agrupamentos frequentemente correspondem a comunidades científicas associadas a áreas específicas do conhecimento ou a redes institucionais de pesquisa. A presença de clusters relativamente coesos indica que a colaboração científica tende a ocorrer dentro de comunidades disciplinares ou institucionais relativamente bem definidas. Entretanto, a rede também apresenta conexões entre diferentes clusters, sugerindo a existência de colaborações interdisciplinares e interinstitucionais.

4 CONSIDERAÇÕES

Este estudo apresentou o desenvolvimento de um arcabouço computacional destinado à análise em larga escala de currículos da Plataforma Lattes de docentes vinculados a programas de pós-graduação reconhecidos pela CAPES. A metodologia proposta integra técnicas de processamento incremental de dados, construção de bases analíticas estruturadas e geração de indicadores bibliométricos, cientométricos e redes de colaboração científica.

Os resultados obtidos demonstram o potencial da Plataforma Lattes como fonte de dados para estudos métricos da informação e análises sobre a estrutura da ciência brasileira. A metodologia desenvolvida permite explorar de maneira sistemática informações presentes nos currículos acadêmicos, contribuindo para a compreensão de padrões de produtividade científica, concentração institucional e colaboração entre pesquisadores.

REFERÊNCIAS

- ARAÚJO, Carlos Alberto. A. Bibliometria: evolução histórica e questões atuais. **Em questão**, v. 12, n. 1, p. 11-32, 2006. Disponível em: <https://dialnet.unirioja.es/descarga/articulo/6134719.pdf>. Acesso em: 15 mai. 2026.
- CANDOTTI, Ennio. **Ciência na Educação Popular**. Ciência e público: caminhos da divulgação científica no Brasil. Rio de Janeiro: Casa da Ciência-Centro Cultural de



Ciência e Tecnologia da UFRJ. 2002. Disponível em: <https://juntosnacasadaciencia.wordpress.com/wp-content/uploads/2020/06/cic3aancia-e-pc3bablico-caminhos-da-divulgac3a7c3a3o-cientc3adfica-no-brasil.pdf>. Acesso em: 14 mar. 2026.

FERREIRA, Ana Gabriela Clipes. Bibliometria na avaliação de periódicos científicos. **DataGramaZero-Revista de Ciência da Informação**, v. 11, n. 3, p. 1-9, 2010. Disponível em: <https://cip.brapci.inf.br/download/7110>. Acesso em: 13 mai. 2026.
PRITCHARD, Alan. Statistical bibliography or bibliometrics. **Journal of documentation**, v. 25, p. 348, 1969.

SILVA, Márcia Regina; HAYASHI, Carlos Roberto Massao; HAYASHI, Maria Cristina Piumbato Innocentini. Análise bibliométrica e cientométrica: desafios para especialistas que atuam no campo. **InCID: revista de ciência da informação e documentação**, v. 2, n. 1, p. 110-129, 2011. DOI: <https://doi.org/10.11606/issn.2178-2075.v2i1p110-129>. Acesso em: 10 mar. 2026.