

A base ORCID e a representação da carreira docente no Brasil: potencial, cobertura e desafios

The ORCID database and the representation of the academic career in Brazil: potential, coverage, and challenges

La base ORCID y la representación de la carrera docente en Brasil: potencial, cobertura y desafíos

Ana Maria Carneiro

Universidade Estadual de Campinas, Núcleo de Estudos de Políticas Públicas, Campinas, São Paulo, Brasil

Daniela Maciel Pinto

Universidade Estadual de Campinas, Departamento de Política Científica e Tecnológica, Campinas, São Paulo, Brasil

Embrapa Territorial, Transferência de Tecnologias, Campinas, Brasil

João Gabriel Pedreira de Moura Gomes

Universidade Estadual de Campinas, Faculdade de Engenharia Civil, Arquitetura, Campinas, São Paulo, Brasil

ORIGINAL

Resumo

Objetivo. A profissão acadêmica tem sido explorada usando uma diversidade de métodos e fontes de dados, tais como currículos, bases de publicações e, mais recentemente, a base ORCID (Open Researcher and Contributor ID). O objetivo deste trabalho é explorar a cobertura da base ORCID no Brasil em relação à comunidade acadêmica, focando no caso de professores em instituições de ensino superior (IES) públicas e privadas em 2020. **Método.** Todos os registros de indivíduos afiliados a organizações no Brasil do ORCID foram extraídos, limpos e organizados em uma base de dados, que passou por procedimentos de normalização das variáveis (cidades, unidades da federação, regiões, nome das IES). As ocupações foram filtradas para selecionar apenas os docentes, sendo também categorizadas segundo atuação e nível na carreira. As IES foram classificadas segundo a tipologia de Schwartzman et al. (2021). Os dados da base ORCID foram comparados com o Censo da Educação Superior do Ministério da Educação. **Resultados.** Os registros de docentes na base ORCID representam cerca de 7,5% do total de docentes, com uma maioria de professores do sexo masculino, um único vínculo, distribuídos em 1834 IES. Em termos de nível na carreira, em primeiro lugar aparecem os registros nos quais não foi possível identificar a posição. Em segundo lugar, as categorias identificadas mais frequentes reproduzem os níveis da carreira magistério superior de forma invertida (da posição mais alta para as posições de ingresso). Há uma concentração maior na Região Sudeste, especialmente em São Paulo. Há uma sobrerepresentação de perfis de docentes em IES mais intensivas em pesquisa. **Conclusões.** Apesar de se tratar de uma amostra pequena do conjunto de docentes brasileiros, os perfis na base ORCID apresentam distribuições semelhantes ao universo em termos de gênero e distribuição geográfica, repetindo as desigualdades e assimetrias existentes. Entretanto, a base ORCID apresenta vieses importantes em termos da posição na carreira e na vinculação aos tipos de IES, o que pode estar ligado aos incentivos recebidos para o registro no ORCID e o formato da base de campos não padronizados. Tendo em vista o potencial de crescimento da cobertura, abre-se uma agenda de pesquisa para estudos sobre a carreira acadêmica no Brasil, incluindo a análise de mobilidade entre instituições, a progressão na carreira e diferenças entre áreas, com a complementação de dados de publicações. E também a comparação com outras bases nacionais, como a Plataforma Lattes, e a cobertura em outros países.

Palavras-chave: ORCID, carreira acadêmica, instituições de ensino superior, cientometria, Brasil

Abstract

Objective. The academic profession has been explored using various methods and data sources, such as curricula vitae, publication databases, and more recently, the ORCID (Open Researcher and Contributor ID) database. This study aims to explore the coverage of the ORCID database in Brazil concerning the academic community, focusing on the case of professors in public and private higher education institutions (HEIs) in 2020. **Method.** All records of individuals affiliated with organizations in Brazil from the ORCID database were extracted, cleaned, and organized into a database that underwent normalization procedures for variables (cities, states, regions, and the names of HEIs). Occupations were filtered to select only faculty members, who were further categorized by their area of work and career level. The HEIs were classified according to the typology of Schwartzman et al. (2021). The data from the ORCID database were compared with the Higher Education Census from the Ministry of Education. **Results.** The records of faculty members in the ORCID database represent about 7.5% of the total number of faculty, with a majority of male professors, with single affiliation, distributed across 1834 HEIs. Regarding career level, first, there are the records in which it was not possible to identify the position. Second, the most frequently identified categories reflect the higher education teaching career levels in reverse order (from the highest position to entry-level positions). There is a higher concentration in the Southeast region, especially São Paulo. There is an overrepresentation of faculty profiles in research-intensive HEIs. **Conclusions.** Although this is a small sample of the total Brazilian faculty, the profiles in the ORCID database show similar distributions to the overall population in terms of gender and geographical distribution, mirroring existing inequalities and asymmetries. However, the ORCID database presents significant biases regarding career position and affiliation with types of HEIs, which may be related to the incentives for registering with ORCID and the non-standardized fields in the database. Given the potential for increased coverage, a research agenda opens up for studies on academic careers in Brazil. This includes analyzing mobility between institutions, career progression, and field differences, complemented by publication data. Additionally, comparisons with other Brazilian databases, such as the Lattes Platform, and other countries could be made.

Keywords: ORCID, academic career, higher education institutions, scientometrics, Brazil

Resumen

Objetivo. La profesión académica ha sido explorada utilizando una diversidad de métodos y fuentes de datos, tales como currículos, bases de publicaciones y, más recientemente, la base ORCID (Open Researcher and Contributor ID). El objetivo de este trabajo es explorar la cobertura de la base ORCID en Brasil en relación con la comunidad académica, enfocándose en el caso de los profesores en instituciones de educación superior (IES) públicas y privadas en 2020. **Método.** Se extrajeron, limpian y organizaron todos los registros de individuos afiliados a organizaciones en Brasil de la base ORCID, creando una base de datos que pasó por procedimientos de normalización de variables (ciudades, unidades federativas, regiones, nombre de las IES). Las ocupaciones fueron filtradas para seleccionar solo a los docentes, siendo también categorizadas según su actuación y nivel en la carrera. Las IES fueron clasificadas según la tipología de Schwartzman et al. (2021). Los datos de la base ORCID fueron comparados con el Censo de Educación Superior del Ministerio de Educación. **Resultados.** Los registros de docentes en la base ORCID representan cerca del 7,5% del total de docentes, con una mayoría de profesores de sexo masculino, un único vínculo, distribuidos en 1834 IES. En términos de nivel en la carrera, en primer lugar, aparecen los registros en los que no fue posible identificar la posición. En segundo lugar, las categorías identificadas más frecuentes reproducen los niveles de la carrera del magisterio superior de forma invertida (de la posición más alta hacia las posiciones de ingreso). Hay una mayor concentración en la Región Sudeste y en São Paulo. Existe una sobrerepresentación de perfiles de docentes en IES más intensivas en investigación. **Conclusiones.** A pesar de tratarse de una muestra pequeña del conjunto de docentes brasileños, los perfiles en la base ORCID presentan distribuciones semejantes al universo en términos de género y distribución geográfica, repitiendo las desigualdades y asimetrías existentes. Sin embargo, la base ORCID presenta sesgos importantes en cuanto a la posición en la carrera y la vinculación a los tipos de IES, lo que puede estar relacionado con los incentivos recibidos para el registro en ORCID y el formato de la base con campos no estandarizados. Teniendo en cuenta el potencial de crecimiento de la cobertura, se abre una agenda de investigación para estudios sobre la carrera académica en Brasil, incluyendo el análisis de la movilidad entre instituciones, la progresión en la carrera y las diferencias entre áreas, con la complementación de datos de publicaciones. También la comparación con otras bases nacionales, como la Plataforma Lattes, y la cobertura en otros países.

Palabras clave: ORCID, carrera académica, instituciones de educación superior, cienciometría, Brasil

1 Introdução

Os assuntos relacionados com a profissão acadêmica têm sido explorados usando uma diversidade de métodos e bases de dados. Os estudos buscam responder questões sobre o que ocorre com um/a acadêmico/a quando muda de instituição e os fatores determinantes da progressão na carreira e produtividade. Diante da miríade de fontes de dados disponíveis (e de outros tantos indisponíveis), interessa explorar as oportunidades e desafios da

utilização da base ORCID, que permite, por exemplo, analisar os currículos, carreiras e produções científicas de forma integrada (Costas et al., 2022).

A base de dados Open Researcher and Contributor ID (ORCID) é uma base global de registros persistentes para pesquisadores com estrutura de acesso aberto que baseia-se

em dados biográficos auto-reportados para mapear carreiras acadêmicas em múltiplas perspectivas, incluindo a formação dos pesquisadores, dados de publicações e emprego e as conquistas profissionais (ex. projetos de pesquisa, filiações acadêmicas etc.), junto com o conhecimento sobre os relacionamentos dos pesquisadores com diferentes organizações (Costas et al., 2022, p. 2-3).

A iniciativa do ORCID começou em 2009 (Sobrido Prieto et al., 2016) e foi consolidada em 2012, quando passou a ser promovida por editoras, agências de financiamento e organizações de pesquisa como um identificador universal para pesquisadores (Youtie et al., 2017). O ORCID surgiu no esteio de outras iniciativas de criar identificadores persistentes, para organizações (Research Organization Registry - ROR), pessoas (ORCID), produtos de pesquisas (Digital Object Identifier - DOI) e automação de repositórios institucionais (Institutional Current Research Information Systems - CRIS) (Schnieders et al., 2022). Parece um passo incremental, mas representou uma etapa relevante em um área em que a “validade da autoria é uma moeda crítica” (Porter, 2022, p. 1), pois marca a identidade do pesquisador frente a editoras, financiadores e instituições. O ORCID foi criado a partir de um mix de abordagens bottom up (instituições) e mid-level (consórcios nacionais, agências de fomento, editoras) e top down (países que têm obrigado os pesquisadores a utilizar o identificador em relatórios e processos acadêmicos) (Porter, 2022). Ao longo do tempo, tem sido integrado a estas outras iniciativas de forma a aumentar a interoperabilidade de sistemas (Fernández-Marcial et al., 2023).

Desta forma, o ORCID envolve agências de fomento, universidades e institutos de pesquisa, editoras e gestores públicos. Estes atores tanto atuam como incentivadores a sua adoção, quanto se beneficiam do uso de seus metadados para fins de avaliação, accountability e planejamento. A plataforma vem sendo adotada crescentemente. Segundo estatísticas da base em janeiro de 2025, há 9,08 milhões pesquisadores ativos, sendo que o Brasil ocupa o quarto lugar entre os países com mais perfis, com mais de 423 mil registros (ORCID Statistics, 2025).

O uso no Brasil iniciou-se em 2016 a partir da adesão das três universidades estaduais paulistas (Universidade de São Paulo - USP, Universidade Estadual de Campinas - UNICAMP e Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP) (RNP, 2018), o que foi reforçado em 2017 pela decisão da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) de começar a solicitar este identificador em seu sistema de financiamento e pelo Scientific Electronic Library Online (SciELO), nas publicações da base (Meijas, Gabriela, 2018).

Atualmente, há 13 instituições brasileiras membros do ORCID, incluindo quatro agências de fomento (Conselho Nacional das Fundações Estaduais de Amparo à Pesquisa - CONFAP, Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, CAPES e Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP), cinco instituições de pesquisa (Fundação Getúlio Vargas, Hospital Israelita Albert Einstein, UNICAMP, UNESP e Universidade Federal de Uberlândia), dois repositórios (Instituto Brasileiro de Informação em Ciência e Tecnologia - Ibict e SciELO) e dois órgãos governamentais (Serviço Geológico do Brasil e Rede Nacional de Ensino e Pesquisa - RNP) (ORCID Statistics, 2025).

Apesar do potencial da base ORCID para a pesquisa sobre a profissão acadêmica e a avaliação da pesquisa (Haak et al., 2018), a prevalência da adoção deste identificador por professores e pesquisadores não ocorre de forma aleatória, sendo adotado mais por determinados perfis de pesquisadores, disciplinas e regiões geográficas. Desta forma, é necessário analisar as variações sistemáticas em sua adoção de forma a levá-la em conta nos estudos que a utilizam (Youtie et al., 2017). A base ORCID vem sendo estudada desde pelo menos 2012. A maioria das pesquisas limitam-se a explicar o que é a plataforma e como pode apoiar a comunidade acadêmica, sem explorar todo seu potencial (Costas et al., 2022). Além disso, há relativamente poucos estudos que fazem a análise da adoção deste identificador (Costas et al., 2022; Fernández-Marcial et al., 2023; Porter, 2022; Youtie et al., 2017).

O objetivo deste trabalho é investigar a cobertura da base ORCID no Brasil em relação à comunidade acadêmica, focando no caso de professores em instituições de ensino superior (IES) públicas e privadas em 2020. A motivação surgiu no âmbito de um projeto maior, na área da pesquisa da pesquisa, que busca analisar os efeitos da mobilidade na carreira acadêmica no Brasil. Este projeto tem como objetivo desenvolver e empregar abordagens que gerem evidências sobre como a pesquisa nas principais áreas do conhecimento pode ser melhor planejada,

financiada, avaliada e comunicada, de forma a gerar mais e maiores impactos positivos para a sociedade.

Neste sentido, buscou-se explorar a distribuição dos perfis da carreira docente em termos geográficos e por tipos de instituição, comparando-se com dados do Censo da Educação Superior 2020 do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira do Ministério da Educação que indicavam a existência de 314.384 docentes em exercício no ensino superior. A questão que se coloca é até que ponto a base ORCID representa a distribuição geográfica e institucional dos professores em Instituições de Ensino Superior no Brasil, e como essa representatividade se compara aos dados do Censo da Educação Superior de 2020.

Trata-se de um trabalho pioneiro no Brasil, que mapeou a cobertura da base ORCID no Brasil por unidade da federação, bem como a diversidade de perfis presentes em termos de gênero, ocupação e tipos de IES.

2 Revisão de literatura

Segundo revisão de literatura sobre mobilidade acadêmica de Autor et al. (2023), utilizando 22 artigos, a maioria dos estudos utilizou desenhos qualitativos de pesquisa, principalmente entrevistas (Nachatar Singh, 2021; Nikunen & Lempiäinen, 2020; Schittenhelm, 2022; Tzanakou, 2021; Vohlídalová, 2017), mas também métodos antropológicos como auto-ethnografia e outros baseados em diários (Gibb, 2022; Henderson, 2021; Jenkins, 2020).

O estudo da profissão acadêmica por meio de pesquisas quantitativas tem utilizado dados primários advindos de surveys (Aarrevaara et al., 2021; Franzoni et al., 2012; Malakhov et al., 2020; Uhly et al., 2017) e variados tipos de bases de dados secundários. Nestes últimos podem ser encontrados pelo menos quatro tipos de fontes de informação: a) bases de currículos (Cañibano et al., 2008; Mascarenhas & Dias, 2022; Reis et al., 2021; Robles-Belmont, 2021; Siekierski et al., 2019; Yan et al., 2020), b) bases de dados de recursos humanos (Aksnes et al., 2019; Wendt et al., 2022); c) processos de seleção e recrutamento (Moratti, 2021); e d) bases de publicações acadêmicas (Abel et al., 2019; Aman, 2018; Malakhov et al., 2020; Miranda-González et al., 2020; Paraskevopoulos et al., 2020; Zhao et al., 2022). No caso das bases de dados baseadas em currículos, há trabalhos que utilizam fontes de agências de fomento (por ex. Cañibano et al., 2008) e bases nacionais, como os trabalhos baseados na Plataforma Lattes brasileira (ex. Reis et al., 2021; Mascarenhas & Dias, 2022; Siekierski et al., 2019).

Cada desenho de pesquisa possui vantagens e desvantagens em termos de esforço, tipo de questão que consegue responder, representatividade e possibilidade de generalização dos resultados. Por exemplo, os estudos qualitativos possibilitam analisar, de maneira aprofundada, como questões relacionadas à desigualdade de gênero na carreira acadêmica afetam as decisões de mobilidade, progressão e ocupação de posições de liderança, entre outros fatores (Autor et al., 2023). Entretanto, tanto os surveys quanto os estudos qualitativos demandam um grande esforço para serem desenvolvidos e conseguem alcançar, em geral, apenas pequenas amostras. Por outro lado, os estudos quantitativos, como os baseados na bibliometria, têm a vantagem da abrangência do escopo (Autor et al., 2023), mas faltam outras informações como formação, nacionalidade, início da carreira em termos de ano de admissão ou término do doutorado, entre outros.

Alguns estudos recentes têm avançado para o uso da base de dados ORCID (Open Researcher and Contributor ID) (Gomez et al., 2020; Yan et al., 2020; Zhang, Qian, et al., 2023; Zhao et al., 2022). O preenchimento da base é feito pelos pesquisadores de forma auto-declarada, mas há também mecanismos que podem alimentar determinados campos automaticamente. O indivíduo pode configurar seu perfil para receber atualizações automáticas do Crossref (Porter, 2022). Suas tabelas incluem informações sobre formação (etapas, onde e quando se formaram), publicações (título, nome do periódico ou da conferência, ano da publicação), emprego (nome da organização, cidade e país, ocupação e período do emprego) e realizações profissionais como projetos de pesquisa, filiação a sociedades científicas e prêmios (Costas et al., 2022; Gomez et al., 2020).

O foco no indivíduo permite análises empíricas “diretas”, sem a necessidade de proxies baseadas em produtos que muitas vezes trazem um retrato no qual as comunidades de pesquisa do Sul Global são sub representadas (Costas et al., 2022). A inclusão do período de início e fim dos vínculos permite o estudo longitudinal, fundamental para compreender a mobilidade acadêmica (Yan et al., 2020).

Três estudos ilustram a potencialidade dos estudos longitudinais (Gomez et al., 2020; Yan et al., 2020; Zhao et al., 2022). Zhao et al. (2022) exploraram a persistência das colaborações em pesquisa dos cientistas que se movem entre instituições usando uma amostra de 4.343 cientistas norte-americanos e suas publicações na Web of Science. O uso da base Orcid permitiu detectar a mobilidade entre instituições, a direção da mobilidade e os

países nos quais o cientista trabalhou. Na junção com a Web of Science permitiu, então, analisar a produção bibliográfica em co-autoria, as áreas de pesquisa e a temporalidade da colaboração.

Yan et al. (2020) analisaram a mobilidade de professores de universidades nos Estados Unidos usando a base ORCID. As informações disponíveis na base de dados permitiram, então, identificar os professores filiados a instituições de ensino superior e classificar cada profissional segundo o estágio na carreira (assistant, associate ou full professorship). Com a visão da trajetória longitudinal dos indivíduos, foi possível analisar as situações de mobilidade interinstitucional dentro dos EUA e explorar padrões de mudança em relação ao cargo (progressão, descenso ou permanência), localização (rural ou urbana) e intensidade em pesquisa da instituição.

Gomez et al (2020) destacam que o uso de bases bibliométricas para analisar mobilidade dos cientistas não é totalmente adequado, pois faltam informações que apenas os currículos completos trazem, tais como país de doutorado (considerado o ponto de início da carreira), posições em diferentes instituições e quando houve mobilidade. Neste sentido, usaram uma amostra de 116.400 doutores da base ORCID para classificar cada mudança em movimento doméstico, regional (no mesmo continente) e global. A análise mostrou que a mobilidade entre países está estagnada desde os anos 1980, quando comparada com a mobilidade doméstica. Entretanto, os cientistas que se movem internacionalmente têm percorrido distâncias menores dentro dos continentes. Neste sentido, o estudo mostrou os cientistas brasileiros com uma crescente tendência de mobilidade regional, especialmente dentro da América Latina.

Como apresentado, alguns estudos têm integrado as informações do ORCID com outras fontes. O estudo de Valentine e Williams (2023) agregou informações disponíveis dos metadados de publicações, perfis no ORCID e no Scopus e bases de currículos da Austrália para analisar o treinamento acadêmico formal na área de engenharia. Já a pesquisa de Hou, Wu e Xie (2022) buscou identificar fatores preditivos de performance acadêmica nos estágios iniciais da carreira na área de administração combinando a base Scopus e as informações do ORCID.

A base ORCID apresenta uma série de vantagens para o estudo da profissão acadêmica. A primeira, já mencionada, é a possibilidade de realizar estudos longitudinais centrados no pesquisador desde o período de formação até o vínculo empregatício mais recente. Esta é uma característica que supera limitações dos estudos centrométricos. Estes tomam a primeira publicação como marco do início da carreira. Entretanto, o pesquisador pode eventualmente ter realizado mobilidade antes de publicar, como apresentam Costas et al. (2022, p. 6–7):

O ORCID oferece algumas soluções para estas limitações ao incluir informações sobre educação e emprego. Assim, o ORCID permite analisar as transições de afiliação desde o local onde os investigadores realizaram a sua educação e formação (por exemplo, graduação, mestrado ou mesmo doutorado) até ao local onde estiveram empregados (por exemplo, como pós-doc, professor assistente, professor titular ou mesmo numa empresa). Este tipo de informação não está obviamente disponível nas bases de dados centrométricas, uma vez que não está incluído nas publicações científicas.

A segunda vantagem é a de poder realizar estudos em escala global, de forma a lidar com limitações de estudos baseados em surveys ou entrevistas, que em geral referem-se a países específicos ou pequenas amostras (Costas et al., 2022). A terceira vantagem é que o formato da base permite incluir mais tipos de produtos da pesquisa que as bases tradicionais, como trabalhos em eventos, livros, preprints, patentes etc. (Costas et al., 2022). A quarta vantagem, também já mencionada, é a sua estrutura aberta e inclusiva (Costas et al., 2022; Youtie et al., 2017). Desta forma, tornou-se um instrumento importante de meta-pesquisa:

O ORCID oferece uma base de dados inclusiva, interdisciplinar, sem fins lucrativos, aberta e de fácil utilização para aumentar a visibilidade de todos os tipos de atividades científicas. Proporciona também à comunidade científica, aos financiadores e aos responsáveis pela política científica uma base de dados construída pela própria comunidade científica para saber como a ciência funciona e se desenvolve. (Costas et al., 2022, p. 12).

A quinta vantagem é aquela mais diretamente ligada ao identificador único, que apoia iniciativas e metodologias de desambiguação de nomes (Carley et al., 2019; Zhang, Lu, et al., 2023).

Apesar destas vantagens, é importante apontar as limitações do ORCID. A primeira refere-se ao grau de disponibilidade de informações. Como os dados são na sua maioria fornecidos pelos pesquisadores, o nível de preenchimento varia bastante entre os campos. A maior parte das iniciativas das editoras, organizações de pesquisa e agências de fomento que incentivam, ou tornam obrigatório o uso do ORCID demandam que o pesquisador apenas crie o perfil com, no mínimo, algumas poucas informações, como nome e filiação institucional. Desta forma, os indivíduos não necessariamente se engajam em completar o preenchimento dos perfis ou atualizá-los (Porter, 2022; Fernández-Marcial et al., 2023). Além disso, muitos pesquisadores escolhem manter

as informações privadas devido a preocupações com a “existência digital como um todo” (Porter, 2022, p. 5). Segundo levantamento de Costas et al. (2022), em 2020 99,8% dos registros no ORCID possuíam pelo menos o nome do pesquisador e cerca de 40,7% tinham pelo menos uma informação em alguma das outras seções. Segundo as estatísticas do ORCID de janeiro de 2025, 63% dos perfis ativos possuem informação de filiação e 50% de publicações, sendo que apenas 33% possuem os dois tipos de informação (ORCID, 2025). Segundo pesquisa de Porter (2022), a partir da integração de informações do ORCID, Crossref e Dimensions, com pesquisadores que tinham publicações entre 2015 e 2019, o Brasil teria cerca de 42% de adoção (utilização ativa) e 42% de engajamento (completude das informações).

A segunda limitação seria um caso extremo da primeira, pelo fato de a maioria dos dados não serem verificados (Costas et al., 2022; Porter, 2022). Os “perfis fantasmas” ou “perfis vazios” têm levantado preocupações com relação às fábricas de artigos (*paper mills*). Teixeira da Silva (2021) aponta que o uso do ORCID não necessariamente garante a autenticidade dos autores devido a utilização em artigos produzidos por fábrica de artigos, por não garantir em alguns casos a desambiguação dos nomes autores, especialmente com sobrenomes chineses. Segundo o autor, operadores dessas fábricas poderiam estar por trás da criação de perfis fantasmas não verificados e seria urgente o desenvolvimento de ferramentas de código aberto para auxiliar acadêmicos, editores e editoras a detectar estes problemas. Desta forma, a incompletude do preenchimento limitar a função da desambiguação (Fernández-Marcial et al., 2023).

A terceira limitação refere-se à falta de padronização das informações, pois a grande maioria dos campos são abertos, aceitando qualquer tipo de dado, o que leva à perda de dados e inconsistências (Porter, 2022). Neste sentido, repositórios de registros institucionais públicos poderiam ter um papel relevante no aperfeiçoamento, curadoria e validação dos metadados do ORCID (Costas et al., 2022), bem como iniciativas de interoperabilidade entre sistemas e mudanças na cultura organizacional (Fernández-Marcial et al., 2023).

As demais limitações referem-se à cobertura desigual da base ORCID de acordo com determinados fatores, como posição na carreira, localização geográfica, área do conhecimento e financiamento. Segundo Gomez et al (2020), há uma maior presença de pesquisadores mais jovens e que se formaram mais recentemente. Este aspecto é corroborado por Porter (2022), que aponta que pesquisadores em início de carreira, ou ainda durante as etapas de formação, têm sido mais incentivados a se registrar no ORCID.

Sobre as áreas do conhecimento, o estudo de Porter (2022) indica que as áreas de maior adoção são ciências da terra e química e as de menor adoção são humanidades. Nesta última, uma possível explicação seria o menor número de autores por artigo.

Em relação ao financiamento, segundo Porter (2022), os autores com projetos financiados apresentam taxas de adoção significativamente mais altas que as médias dos seus países, o que refletiria a influência de editoras e agências de fomento.

A cobertura geográfica do ORCID varia entre as regiões e países e tem se expandido ao longo do tempo. Os estudos sobre este tema são escassos, mas os primeiros estudos mostraram uma cobertura maior na Europa do que na Ásia e Estados Unidos (Gomez et al., 2020; Youtie et al., 2017). Segundo Porter (2022, p. 12):

Em contraste com a internacionalização da pesquisa, os padrões de adoção e engajamento no ORCID são regionais, sendo que países como Portugal, Polônia, Dinamarca e Austrália assumiram a liderança e gigantes da pesquisa, como os Estados Unidos, a China e o Japão, ficaram para trás. Os pesquisadores de países com baixas taxas de adoção do ORCID têm também maior probabilidade de não se engajarem com a atualização dos seus perfis.

Entretanto, este quadro mudou ao longo do tempo devido à iniciativas que têm promovidoativamente a adoção e interoperabilidade do ORCID com sistemas nacionais, incluindo Reino Unido, Brasil, Espanha, Austrália e Canadá (Sixto-Costoya et al., 2021; Marín-Arraiza & Meijas, 2020; Dasler, 2017). Além disso, desde 2018 a União Européia tem recomendado o uso do ORCID para acesso à financiamentos dos programas-quadro (Fernández-Marcial et al., 2023). Segundo as estatísticas do ORCID de janeiro de 2025, 13% dos registros são dos Estados Unidos, seguidos por China (11%), Índia e Brasil (6% cada) (ORCID, 2025). Uma análise mais precisa da cobertura geográfica é dificultada em parte pela incompletude do preenchimento dos dados. Analisando manualmente os perfis dos professores da Faculdade de Artes e Humanidades da Universidade do Porto, Fernández-Marcial et al. (2023) verificaram que mais de 90% tinham um registro na base ORCID, mas menos da metade tinham completado o campo país.

Não há estudos que analisem possíveis disparidades na cobertura dos registros do ORCID em relação à gênero. Entretanto, pesquisas mais amplas indicam desigualdades persistentes em termos de representação, reconhecimento e financiamento de pesquisadoras (Cruz-Castro et al., 2022; Hipólito et al., 2022; Atães, 2023), o que pode refletir em uma presença menor no ORCID, especialmente em níveis mais sêniores da carreira acadêmica (Arêas et al., 2020; Valentova et al., 2017).

Além disso, há preocupações éticas com o *status* de uso compulsório por parte de periódicos e organizações de pesquisa (Costas et al., 2022). Como visto, a riqueza de informações e possibilidades do uso da base ORCID para pesquisas sobre a profissão acadêmica justificam a necessidade de estudos que explorem suas limitações em contextos específicos, como o objeto deste trabalho.

3 Metodologia

Trata-se de uma pesquisa quantitativa, de caráter descritivo-exploratório, com procedimentos majoritariamente documentais. O objetivo foi explorar a cobertura da base ORCID no Brasil em relação à comunidade acadêmica, com foco específico em professores de IES. A escolha por essa abordagem é justificada pela possibilidade de analisar padrões em um grande volume de dados, permitindo uma caracterização da comunidade acadêmica vinculada à base ORCID. O caráter exploratório da pesquisa é essencial para investigar um tema relativamente novo e ainda pouco documentado na literatura, enquanto a dimensão descritiva busca mapear as características e especificidades dos registros disponíveis. Os procedimentos adotados envolveram, primeiramente, a coleta de dados provenientes da base ORCID, complementados por bases institucionais e públicas. Em seguida, os dados foram tratados e normalizados, a fim de padronizá-los para as etapas analíticas. A análise e visualização dos dados foi conduzida com o uso de bibliotecas disponíveis para a linguagem R. As próximas subseções detalham os tipos de dados utilizados, suas respectivas fontes e os procedimentos adotados para coleta, tratamento, normalização e análise.

3.1 Dados

A análise foi realizada a partir da extração, em setembro de 2023, de todos os registros da base ORCID referentes a indivíduos filiados a organizações de qualquer tipo no Brasil. Essa extração ocorreu por meio da conta oficial do ORCID (<https://orcid.figshare.com/>) no repositório Figshare, reunindo dados desde o início da atividade de registro na plataforma ORCID até o final do ano de 2022. Nesse aspecto, foram extraídas informações de três tabelas principais: educação, qualificação e emprego. Este estudo priorizou as tabelas de emprego e qualificação devido ao foco nas ocupações profissionais. É importante ressaltar que o número de perfis de professores brasileiros na base pode ser maior que o extraído, dado que como visto na revisão de literatura, muitos indivíduos não preenchem todas as informações de todas as tabelas.

A tabela “emprego”, originalmente, continha 449.124 registros (ou seja, o número de linhas da base de dados), e 17 variáveis, (ou seja, campos), sendo uma delas correspondente à chave única de identificação (número do ORCID). Já a tabela qualificação possuía um total de 190.594 registros e as mesmas 17 variáveis, servindo de complemento à tabela “emprego”. Com exceção da variável chave, todas as variáveis apresentaram uma significativa falta de padronização. Por exemplo, havia 115.861 tipos de carreiras distintas, assim como 4.811 regiões brasileiras e 12.622 municípios diferentes. A lista das 17 variáveis e sua descrição está disponível na tabela 1A do Apêndice deste trabalho.

Além das tabelas relativas aos dados do ORCID, utilizou-se a base de dados organizada no âmbito do projeto maior, a partir de microdados do Censo da Educação Superior de 2001 a 2022 e de dados da Plataforma Sucupira da CAPES (Schwartzman & Bueno, 2023). Essa base, denominada “Base de tipologias de IES”, continha informações relativas às Instituições de Ensino Superior (IES) e foi utilizada para classificar os tipos de IES existentes no país, com base na proposta de Schwartzman (2021).

Ressalta-se que os dados extraídos da base ORCID incluíram todos os registros disponíveis, independentemente do preenchimento completo das informações. Para a investigação deste trabalho, foi necessário realizar um filtro, selecionando apenas os registros que continham informações preenchidas na variável correspondente à identificação da organização/instituição (IES). Esse processo de filtragem foi realizado utilizando a linguagem R, com o pacote dplyr e a função filter().

Após o filtro, os dados resultantes foram submetidos a tratamentos de normalização. Para isso, foram utilizados dados complementares de fontes oficiais, como o dicionário "Área Territorial - Brasil, Grandes Regiões, Unidades da Federação e Municípios", do Instituto Brasileiro de Geografia e Estatística (IBGE), que serviu para normalizar os nomes de municípios e regiões brasileiras. Dados da "Lista de Todas as Universidades do Brasil", proveniente do Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), utilizado para normalizar o nome das IES; os dados de códigos de IES do Cadastro e-MEC; e a API da OpenAI, com o modelo GPT-4 (OpenAI, 2024). Em síntese, os dados utilizados nesta pesquisa estão organizados no Quadro 1.

Quadro 1

Tipos de fontes de dados e sua utilização no estudo

Fonte de Dados	Fonte	Finalidade
Base ORCID (tabelas emprego e qualificação)	ORCID	Analise de ocupações profissionais e identificação das organizações/instituições (IES), considerando-se abordagem geográfica
Base de tipologias de IES	Schwartzman et al. (2021)	Classificação e identificação de tipos de IES existentes no Brasil, por nome e código da IES
Área territorial - Brasil, Grandes Regiões, Unidades da Federação e Municípios"	Instituto Brasileiro de Geografia e Estatística (IBGE, 2022)	Processamento automático para normalização do nome de municípios e regiões do Brasil
Lista de Todas as Universidades do Brasil	Instituto Nacional de Estudos e Pesquisas Educacionais (INEP)	Processamento automático para normalização do nome da IES
Código de IES	Cadastro e-MEC	Processamento manual para normalização do código das IES
GPT-4	Open AI	Processamento e tratamento automático para normalização do nome e código das IES

Nota. Fonte: Elaboração própria (2025). *[Início da Descrição]* O Quadro apresenta três colunas, Fonte de Dados, Fonte e Finalidade, organizando as bases utilizadas no estudo, suas origens e o objetivo de cada uma. A linha Base ORCID (tabelas emprego e qualificação) indica como fonte o ORCID, com a finalidade de análise de ocupações profissionais e identificação das organizações e instituições considerando abordagem geográfica. A Base de tipologias de IES tem como fonte Schwartzman et al. (2021) e é usada para classificação e identificação de tipos de IES no Brasil, por nome e código. A linha Área territorial – Brasil, Grandes Regiões, Unidades da Federação e Municípios apresenta como fonte o Instituto Brasileiro de Geografia e Estatística, IBGE 2022, sendo utilizada para processamento automático na normalização do nome de municípios e regiões. A Lista de Todas as Universidades do Brasil tem como fonte o Instituto Nacional de Estudos e Pesquisas Educacionais, INEP, e serve ao processamento automático para normalização do nome das IES. A linha Código de IES indica como fonte o Cadastro e-MEC, com a finalidade de processamento manual para normalização do código das IES. A última linha traz GPT-4 como fonte OpenAI, utilizado para processamento e tratamento automático na normalização do nome e código das IES *[Fim da descrição]*.

3.2 Tratamento e Normalização dos dados

Durante o tratamento dos dados, foi necessário corrigir problemas de padronização e inconsistência das informações que poderiam comprometer a qualidade da análise que, como visto, é uma das limitações da base. Para alcançar um nível adequado de uniformidade e precisão, foram adotadas técnicas de limpeza e transformação de dados, conforme descrito a seguir.

3.2.1 Identificação e Remoção de Duplicidades

Os primeiros procedimentos de identificação de duplicidades foram realizados a partir da tabela "emprego". Para garantir a precisão dos dados, foram identificados e removidos registros duplicados, considerando como duplicados aqueles em que todas as 17 variáveis apresentavam exatamente o mesmo conteúdo. Assim, verificou-

se que 172.880 registros eram duplicados, restando 276.244 e 17 variáveis. Durante esse processo, foram mantidos os registros de uma mesma pessoa que apresentavam dados distintos, particularmente para permitir a análise de números de ocupações e vínculos institucionais diferentes. Isso foi alcançado removendo linhas com informações completamente idênticas, mas preservando registros únicos que refletiam múltiplas ocupações ou variações de dados associados ao mesmo indivíduo.

3.2.2 Normalização de Regiões e Estados

A normalização dos dados geográficos, como nomes de regiões, municípios e unidades da federação, foi realizada utilizando dicionários baseados nas classificações oficiais do IBGE (2022). Dado que, em algumas regiões e estados, existem nomes de locais que podem ser análogos, o município foi utilizado como uma proxy para determinar a região correta, levando em consideração o estado associado.

3.2.3 Adequação dos Nomes das Instituições de Ensino Superior (IES)

Os primeiros procedimentos de identificação de duplicidades foram realizados a partir da tabela “emprego”. Para garantir a precisão dos dados, foram identificados e removidos registros duplicados, considerando como duplicados aqueles em que todas as 17 variáveis apresentavam exatamente o mesmo conteúdo. Assim, verificou-se que 172.880 registros eram duplicados, restando 276.244 e 17 variáveis. Durante esse processo, foram mantidos os registros de uma mesma pessoa que apresentavam dados distintos, particularmente para permitir a análise de números de ocupações e vínculos institucionais diferentes. Isso foi alcançado removendo linhas com informações completamente idênticas, mas preservando registros únicos que refletiam múltiplas ocupações ou variações de dados associados ao mesmo indivíduo.

3.2.4 Construção de Novas Variáveis Corretivas

Com base nas correções e normalizações mencionadas, novas variáveis foram criadas para substituir as versões inconsistentes. Essas novas variáveis foram construídas para refletir as informações corrigidas, como os nomes normalizados de regiões, estados e IES, bem como a classificação adequada de docentes e pesquisadores e somaram outras 11 variáveis (ver tabela 2A do Apêndice). Esse processo permitiu a criação de um conjunto de dados refinado e padronizado, adequado para análises subsequentes e comparações com outras bases de dados, como as fornecidas pelo Censo da Educação Superior e outros registros oficiais.

3.2.5 Construção de Novas Variáveis Corretivas

Um dos maiores desafios no tratamento dos dados foi a identificação de docentes vinculados a IES. Esse procedimento foi realizado de duas formas:

- a) **Automática:** neste caso foram utilizadas técnicas de processamento de texto e a aplicação de expressões regulares para a recuperação de dados, em conjunto com dois dicionários de termos. Esses dicionários foram desenvolvidos pela equipe do projeto e continham termos considerados sinônimos e diretamente relacionados às atividades de docência no Brasil. Neste procedimento, foram identificados 76.705 registros considerados diretamente relacionados às atividades de interesse da pesquisa.
- b) **Manual:** na sequência foi realizada a verificação das ocupações manualmente, com dois focos: 1. identificar a ocupação associada a “professores” de IES; e 2. identificar as instituições vinculadas às ocupações, como forma de garantir que esses vínculos fossem restritos às IES. Neste caso, o trabalho foi realizado sob supervisão da coordenadora do trabalho e resultou em 40.677 registros.

Como foi realizado o filtro por instituições, caso o indivíduo não tenha preenchido esta informação, ele/a não consta na base analisada. A partir do conjunto de 40.677 registros, procedeu-se à identificação dos docentes únicos em exercício no ano de 2020, com o objetivo de compará-los com os dados do Censo da Educação Superior de 2020 e a tipologia das instituições de ensino superior.

Paralelamente, buscou-se padronizar manualmente as ocupações dos docentes, utilizando-se os termos disponíveis nos dados, segundo tipo de atuação (por exemplo, modalidade de ensino, tipo de carreira, estabilidade do vínculo) e nível na carreira, classificando-os em 27 categorias. A base traz termos em português e inglês. Segundo Rodrigues Filho (2015), tomando-se o conjunto total de IES do Brasil, há relativamente poucas com carreiras padronizadas, como a carreira Magistério Superior no sistema público federal (Lei Federal 12.772/2012). A mesma lei rege a carreira de docentes do Ensino Básico, Técnico e Tecnológico. Em São Paulo em 2010, o Conselho de Reitores das Universidades Estaduais Paulistas reestruturou a carreira do Magistério Superior, harmonizando os cargos entre as três universidades para fins de reajuste salarial (Resolução Cruesp 10/2010). As demais IES possuem planos de carreira “locais”, especialmente as privadas. Em muitos casos, os níveis da carreira estão relacionados à titulação. Mas é importante ter em conta que segundo a Lei de Diretrizes de Base (Brasil. Lei de Diretrizes e Bases da Educação Nacional, 1996), apenas no caso das universidades e centros universitários é necessário ter no mínimo 1/3 do corpo docente com mestrado ou doutorado.

Procurou-se manter os termos originais, agregando-os nos casos de equivalência, especialmente na carreira Magistério Superior federal e das universidades estaduais paulistas, como indicado no Quadro 2.

Quadro 2

Equivalência entre as carreiras Magistério Superior

Carreira Magistério Superior do sistema federal	Carreira Magistério superior das Universidades Estaduais Paulistas	Outros termos agregados a partir da base ORCID	Titulação exigida
Professor Auxiliar	Auxiliar de Ensino ou Instrutor (MS-1)		Graduação ou Especialização
Professor Assistente	Assistente ou Professor Assistente (MS-2)		Mestrado
Professor Adjunto	Professor Doutor ou Professor Assistente Doutor (MS-3)	Senior lecturer; Assistant Professor doctor (ou Ph.D)	Doutorado
Professor Associado	Professor Livre Docente (MS-4), Professor Adjunto ou Professor Associado (MS-5)		Doutorado
Professor Titular	Professor Titular (MS-6)	Senior professor	Doutorado

Nota. Fonte: Baseado em Brasil (2012) e CRUESP (2010). [Início da Descrição] O Quadro apresenta quatro colunas que comparam categorias da carreira do magistério superior: Carreira Magistério Superior do sistema federal, Carreira Magistério superior das Universidades Estaduais Paulistas, Outros termos agregados a partir da base ORCID e Titulação exigida. Na linha Professor Auxiliar, o sistema federal apresenta Professor Auxiliar, enquanto as universidades paulistas utilizam Auxiliar de Ensino ou Instrutor (MS-1); não há termos adicionais da base ORCID nessa linha, e a titulação exigida é Graduação ou Especialização. Na linha Professor Assistente, o sistema federal registra Professor Assistente e as universidades paulistas utilizam Assistente ou Professor Assistente (MS-2); na coluna de termos da base ORCID aparecem Senior lecturer e Assistant Professor doctor (ou Ph.D), e a titulação exigida é Mestrado. Na linha Professor Adjunto, o sistema federal apresenta Professor Adjunto, enquanto as universidades paulistas utilizam Professor Doutor ou Professor Assistente Doutor (MS-3), Professor Livre Docente (MS-4), Professor Adjunto ou Professor Associado (MS-5); não há termos adicionais da base ORCID nessa linha e a titulação exigida é Doutorado. Na linha Professor Associado, o sistema federal registra Professor Associado e as universidades paulistas não apresentam correspondência específica nesta tabela; não há termos adicionais e a titulação exigida é Doutorado. Na linha Professor Titular, o sistema federal traz Professor Titular e as universidades paulistas utilizam Professor Titular (MS-6); na coluna de termos da base ORCID aparece Senior professor, e a titulação exigida é Doutorado [Fim da Descrição].

Em alguns casos, a categorização levou em conta a titulação para adequar a classificação. Por exemplo, “professor titular mestre” foi classificado como “professor (sem indicação de nível)”, pois o nível titular demanda doutorado e é comum que “titular” seja associado ao contrato permanente, com estabilidade. Os termos “tutor interno”, “tutor externo” foram associados à ocupação “tutor EAD”. Os professores temporários em qualquer nível da carreira foram categorizados como “professor substituto”. Os professores aposentados foram classificados

como “professor colaborador”, acrescidos do nível indicado. Os casos de estágio docente no ensino superior foram classificados como “auxiliar de docente”.

Ao mesmo tempo, realizou-se um procedimento automatizado, seguido de uma verificação manual para obtenção da informação sobre o gênero dos docentes, uma vez que o ORCID não apresenta essa informação.

Neste sentido, após todos os procedimentos, foram identificados 23.793 docentes, vinculados a IES, e em exercício no ano de 2020, incluindo professores que iniciaram o vínculo antes deste marco ou mesmo no ano de 2020. No caso de docentes com mais de um vínculo ativo em 2020, foi considerado aquele mais antigo. Este conjunto foi cruzado com a tabela “qualificação”, a fim de se obter as ocupações distintas de um mesmo docente. Para tanto, após esse cruzamento, o conjunto de dados somou 25.063 registros. A Tabela 1 resume as ações realizadas.

3.2.6 Classificação das IES

Para a classificação das IES, utilizou-se a proposta de classificação de IES desenvolvida por Schwartzman et al. (2021), que estabelece nove categorias distintas de IES, sendo elas: 1 - Privadas de grande porte; 2 - Públlicas de grande porte com pós-graduação alta; 3 - Públlicas especializadas com pós-graduação alta; 4 - Privadas comunitárias e diferenciadas; 5 - Universidades e Centros universitários privados; 6 - Outras instituições públicas; 7 - Faculdades privadas isoladas; 8 - Instituições de educação tecnológica e 9 - Instituições de Pesquisa e Pós-Graduação. Assim, cruzou-se os dados finais de ORCID (25.063) com a “Base de tipologias de IES”, a partir da sigla das IES, registrando-se o código de cada IES. Importa ressaltar que os dados disponíveis na “Base de tipologias de IES” referem-se a vínculos (com repetição devido a contratos múltiplos dos docentes com mais de uma IES), enquanto a base ORCID possui os dois dados.

Tabela 1

Relação das principais ações e quantitativo de dados decorrentes da tabela emprego

Etapa	Dados (Início)	Dados (pós)	Variáveis
Exclusão de duplicatas	449.124	276.244	17
Normalização dos dados e informações e expansão de variáveis	276.244	276.244	27
Identificação de docentes de forma automática	276.244	76.705	28
Identificação manual de docentes e IES	76.705	40.677	28
Identificação única de docentes vinculados a IES, em exercício até o ano de 2020	40.677	23.793	28
Vínculos empregatícios dos docentes em exercício	23.793	25.063	28
Docentes únicos	23.793	23.793	28

Nota. Fonte: Elaboração própria (2025). **[Início da Descrição]** A tabela apresenta três colunas intituladas Etapa, Dados (início), Dados (pós) e Variáveis, descrevendo o fluxo de processamento dos dados. Na linha Exclusão de duplicatas, os dados passam de 449.124 registros para 276.244 e utilizam 17 variáveis. Na etapa Normalização dos dados e informações e expansão de variáveis, os dados permanecem em 276.244 registros, com aumento para 27 variáveis. Na Identificação de docentes de forma automática, os registros caem de 276.244 para 76.705 e totalizam 28 variáveis. Na Identificação manual de docentes e IES, os registros passam de 76.705 para 40.677, mantendo 28 variáveis. Na Identificação única de docentes vinculados a IES em exercício até o ano de 2020, os registros passam de 40.677 para 23.793, novamente com 28 variáveis. Na etapa Vínculos empregatícios dos docentes em exercício, os registros aumentam de 23.793 para 25.063, mantendo 28 variáveis. Por fim, em Docentes únicos, os valores permanecem em 23.793 registros, totalizando 28 variáveis **[Fim da Descrição]**.

3.3 Análise e ferramentas utilizadas

A análise foi realizada para identificar padrões nas ocupações, explorar a distribuição geográfica das Instituições de Ensino Superior (IES), entender as características dos docentes nas IES brasileiras, bem como identificar as

tipologias das IES existentes nos registros de dados do ORCID. Para tanto, utilizou-se a linguagem de programação R (versão 4.3.1; 2023-06-16 ucrt) com o ambiente de desenvolvimento RStudio (versão 2023.6.1.524).

Diversos pacotes R foram empregados para facilitar a manipulação de dados textuais, análise e visualização, incluindo pacotes para análise espacial e processamento de texto. Assim, para a manipulação e preparação dos dados, foram utilizados os pacotes dplyr, tidyverse, tidy e janitor. Esses pacotes foram empregados para a limpeza, transformação e reorganização dos dados, incluindo a normalização de nomes de colunas, padronização das informações geográficas (regiões, estados e municípios), dos nomes das IES, e manipulação de tabelas para adequação às necessidades analíticas. O pacote janitor, em particular, foi utilizado para ajustar os nomes das colunas e garantir que fossem consistentes e padronizados ao longo de todas as tabelas. As funções dos pacotes utilizados foram empregadas com parâmetros definidos de acordo com as necessidades analíticas. Por exemplo, para filtrar registros de docentes vinculados a IES, foi utilizada a função filter() do pacote dplyr, com condições como !is.na(nome_instituicao); para renomear variáveis, utilizou-se clean_names() do pacote janitor; e, na padronização de regiões e IES, mutate() e case_when() foram aplicadas. O mapeamento geográfico foi viabilizado com st_as_sf() e geom_sf() dos pacotes sf e geobr. Todos os scripts completos, com os parâmetros detalhados, estão disponíveis no repositório GitHub¹.

Para o processamento de texto, os pacotes typer e tidytext foram utilizados com o objetivo de explorar a frequência de termos associados às atividades de pesquisa e docência. Esses pacotes possibilitaram a tokenização de textos e a análise de frequências sobre as palavras e termos mais utilizados nos contextos de pesquisa e ensino.

Análises espaciais foram conduzidas para mapear a distribuição geográfica das IES e das ocupações usando os pacotes sf e geobr, os quais possibilitaram a manipulação de dados espaciais e a criação de mapas da distribuição regional das IES, conforme os docentes identificados.

A identificação do gênero dos docentes foi feita a partir dos pacotes genderBR, gender e genderizer que foram aplicados ao campo "Name" dos dados do ORCID, em conjunto com um algoritmo para desagregação e identificação do primeiro nome do autor. Desta forma, os pacotes utilizados permitiram a inferência do gênero pelo nome do docente, uma vez que possuem, em sua estrutura, dicionários de nomes em diferentes línguas. Adicionalmente, utilizou-se uma base de dados de nomes de pesquisadores brasileiros registrados na Plataforma Lattes, ambos desenvolvidos por Jesus Mena-Chalco. Para aproximadamente 4% dos registros, que não tiveram o gênero identificado automaticamente, foi necessário um trabalho manual para complementar a base de dados e garantir a completude das informações.

Complementarmente, utilizou-se a API da OpenAI, os softwares Tableau (versão 2023.2.24.0115.0353), LibreOffice Calc (versão 7.5.5.2), OpenOffice Calc (versão 4.1.14) e Google Sheets foram adotados para consolidação, análises e apresentação dos dados.

4 Resultados

Considerando o objetivo deste trabalho - explorar a cobertura da base ORCID no Brasil em relação à comunidade acadêmica, com foco nos professores de Instituições de Ensino Superior (IES) - e conforme descrito na metodologia, utilizou-se um conjunto de dados que incluiu 23.793 docentes únicos registrados no ORCID e em exercício no ano de 2020, totalizando 25.063 vínculos. Inicialmente, realiza-se uma análise dos dados sobre os docentes em relação à gênero, quantidade de vínculos e posição na carreira. Em seguida, apresenta-se a análise da distribuição geográfica dos docentes, destacando-se a presença e concentração desses indivíduos em diferentes regiões, estados e cidades do Brasil para, então, explorar sua distribuição segundo tipo das IES.

4.1 Gênero e vínculos institucionais

Com base nos dados do Censo da Educação Superior, em 2020, o Brasil possuía 314.384 docentes em exercício no ensino superior (Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), 2024), o que equivale a 378.492 vínculos (Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), 2020). Assim, os registros na base ORCID correspondem a aproximadamente 7,5% do contingente total de docentes e 7% dos vínculos do país.

¹ Disponível em: <https://bit.ly/3ME1E8g>

Foi possível identificar o gênero de 99,8% dos docentes. A Tabela 2 apresenta a relação entre o número de IES ao qual um docente esteve vinculado pela quantidade de docentes identificados para o número de vínculos com IES, considerando-se o gênero. Considerando os nomes identificados, há uma maioria de homens (13.006 docentes ou 54,6%), enquanto 10.756 (45,2%) são mulheres. Em comparação com os dados do Censo da Educação Superior, há uma leve sobrerepresentação de homens que ali correspondem a 53,2%. Como apontado pela literatura, a sobrerepresentação masculina na academia é refletida na base ORCID.

Tabela 2**Número de vínculos institucionais por número de docentes**

Vínculo	Mulher	Homem	% Mulher	% Homem	Total	%
Vínculo Único	10.323	12.550	43,39	52,75	22.873	96,13
Vínculo Duplo	408	420	1,71	1,77	828	3,48
Vínculo Triplo	24	28	0,10	0,12	52	0,22
Vínculo Quádruplo	1	8	0,004	0,03	9	0,04
Não identificado					31	0,13
Total geral					23.793	100

Nota. Fonte: Elaboração própria (2025). *[Início da Descrição]* A tabela apresenta seis colunas relacionadas aos tipos de vínculo dos docentes: Vínculo, Mulher, Homem, porcentagem de mulheres, porcentagem de homens e Total, acompanhado do percentual total. Em Vínculo Único, há 10.323 mulheres e 12.550 homens, representando 43,39 por cento de mulheres, 52,75 por cento de homens e total de 22.873 docentes, equivalente a 96,13 por cento. Em Vínculo Duplo, são 408 mulheres e 420 homens, com 1,71 por cento de mulheres, 1,77 por cento de homens e total de 828 docentes, 3,48 por cento. Em Vínculo Triplo, são 24 mulheres e 28 homens, representando 0,10 por cento de mulheres, 0,12 por cento de homens e total de 52 docentes, 0,22 por cento. Em Vínculo Quádruplo, há 1 mulher e 8 homens, representando 0,004 por cento de mulheres, 0,03 por cento de homens e total de 9 docentes, 0,04 por cento. A categoria Não identificado registra 31 docentes, equivalente a 0,13 por cento. O Total Geral reúne 23.793 docentes, correspondendo a 100 por cento do conjunto analisado *[Fim da Descrição]*.

Os vínculos podem ser simultâneos ou sequenciais. Observa-se que a maioria dos docentes, tanto mulheres quanto homens, informou apenas um vínculo com uma IES, o que foi denominado de "Vínculo Único". Especificamente, 22.873 professores (95,6%), representando a vasta maioria, têm esse tipo de condição. Além do vínculo único, há também professores que possuem vínculos com diferentes IES. O "Vínculo Duplo", que envolve 828 professores (4%), diz respeito àqueles que informaram afiliações com duas IES. Representando apenas 1% da amostra, as categorias "Vínculo Triplo", que inclui 52 professores, e "Vínculo Quádruplo", com apenas nove professores inseridos, são menos comuns. Eles indicam que um mesmo professor mantém ou manteve vínculo ativo com até quatro instituições. A distribuição percentual entre homens e mulheres é muito semelhante, conforme mostrado nos dados.

A amostra estudada apresentou um total de 1835 IES diferentes, o que representa 62,3% do total de IES em 2020 segundo o Censo da Educação Superior (Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), 2020). A Universidade de São Paulo (USP), possui o maior número de docentes registrados no ORCID, com 1.691. A Universidade Estadual de Campinas (Unicamp) aparece em segundo lugar com 673 docentes, seguida da Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), com 590 docentes registrados. Cabe lembrar que foram as três instituições que primeiro aderiram à plataforma. A tabela 3 apresenta as dez instituições com o maior número de docentes registrados no ORCID.

Tabela 3

IES com maior número de docentes registrados no ORCID

Sigla	IES	Quantidade
USP	Universidade de São Paulo	1691
UNICAMP	Universidade Estadual de Campinas	673
UNESP	Universidade Estadual Paulista Júlio de Mesquita Filho	590
UFRJ	Universidade Federal do Rio de Janeiro	511
UFMG	Universidade Federal de Minas Gerais	397
UFRGS	Universidade Federal do Rio Grande do Sul	349
UFSC	Universidade Federal de Santa Catarina	307
UFF	Universidade Federal Fluminense	288
UFPE	Universidade Federal de Pernambuco	238
UNB	Universidade de Brasília	237

Nota. Fonte: Elaboração própria (2025). *[Início da Descrição]* A tabela apresenta três colunas intituladas Sigla, IES e Quantidade, listando instituições de ensino superior e o número de docentes vinculados a cada uma. Os valores são: USP, Universidade de São Paulo, com 1.691 docentes; UNICAMP, Universidade Estadual de Campinas, com 673; UNESP, Universidade Estadual Paulista Júlio de Mesquita Filho, com 590; UFRJ, Universidade Federal do Rio de Janeiro, com 511; UFMG, Universidade Federal de Minas Gerais, com 397; UFRGS, Universidade Federal do Rio Grande do Sul, com 349; UFSC, Universidade Federal de Santa Catarina, com 301; UFF, Universidade Federal Fluminense, com 288; UFPE, Universidade Federal de Pernambuco, com 238; e UNB, Universidade de Brasília, com 237 docentes *[Fim da Descrição]*.

4.2 Tipo de atuação e posição na carreira

Os diferentes títulos acadêmicos adotados na base foram agrupados em 27 categorias segundo o tipo de atuação e a posição na carreira. A tabela 4 mostra a distribuição dos docentes por categoria, gênero e percentuais de gênero sobre a amostra total, considerando-se o total de vínculos (N=25.063). A categoria mais frequente foi “professor” sem especificação de posição na carreira, que representa 25,58% das mulheres e 14,63% dos homens, sendo o grupo mais numeroso entre os docentes (40,2%). Essa categoria diz respeito aos docentes que não declararam seu nível específico em relação às IES às quais estão vinculados. Os termos associados mais comuns foram: docente (com ou sem indicação da titulação), teacher, lecturer, docente do ensino superior, professor especialista, docente do magistério superior, college professor, entre outros. Em vários casos havia a indicação da disciplina ministrada, da atuação concomitante em cargos administrativos e também do nível (graduação e/ou pós-graduação).

Em seguida, as categorias mais frequentes refletem a hierarquia nos níveis da carreira magistério superior da posição mais alta (titular – 16,49% do total), passando pela posição intermediária (professor associado – 15,11%) para as de ingresso (adjunto – 11,14%, assistente – 8,22%, auxiliar - 1,25%). Outras categorias que se destacaram foram às relacionadas à Educação a Distância (tutor e docente), à carreira de Docente do ensino básico, técnico e tecnológico (EBTT), que atuam nas Instituições Federais de Ensino Básico, Técnico e Tecnológico vinculadas ao Ministério da Educação, e posições temporárias (professor visitante e professor substituto) ou de colaborador.

Tabela 4

Categorias de docentes por gênero

Perfil	Mulher	Homem	% Mulher	% Homem	Total	% Total
Professor (sem informação de nível)	6.410	3.666	25,58	14,63	10.076	40,20
Professor titular	1.129	3.005	4,50	11,99	4.134	16,49
Professor associado	1.255	2.532	5,01	10,10	3.787	15,11
Professor adjunto	849	1.943	3,39	7,75	2.792	11,14
Professor assistente	657	1.403	2,62	5,60	2.060	8,22
Tutor EaD	393	266	1,57	1,06	659	2,63
Professor auxiliar	120	194	0,48	0,77	314	1,25
Professor visitante	61	143	0,24	0,57	204	0,81
Docente EaD	125	128	0,50	0,51	253	1,01
Docente do ensino básico, técnico e tecnológico	129	127	0,51	0,51	256	1,02
Professor colaborador (titular)	11	65	0,04	0,26	76	0,30
Professor substituto	62	58	0,25	0,23	120	0,48
Auxiliar de docente	45	44	0,18	0,18	89	0,36
Professor colaborador (sem informação de nível)	45	26	0,18	0,10	71	0,28
Outro tipo de tutor	24	12	0,10	0,05	36	0,14
Outras categorias	34	51	0,14	0,20	85	0,34
Total	11.349	13.663	45,28	54,51	25.012	99,80
Não identificado		51		0,20	51	0,20
Total Geral		25.063			100	25.063 100,00

Nota. Fonte: Elaboração própria (2025). [Início da Descrição] A tabela intitulada Categorias de docentes por gênero apresenta a distribuição de docentes por perfil, com os números de mulheres e homens, seus percentuais e os totais por categoria. Os valores são: Professor sem informação de nível, 6.410 mulheres e 3.666 homens, com 25,58 por cento de mulheres, 14,63 por cento de homens e total de 10.076 docentes, correspondendo a 40,20 por cento; Professor titular, 1.129 mulheres e 3.005 homens, 4,50 por cento de mulheres, 11,99 por cento de homens e total de 4.134 docentes, 16,49 por cento; Professor associado, 1.255 mulheres e 2.532 homens, 5,01 por cento de mulheres, 10,10 por cento de homens e total de 3.787 docentes, 15,11 por cento; Professor adjunto, 849 mulheres e 1.943 homens, 3,39 por cento de mulheres, 7,75 por cento de homens e total de 2.792 docentes, 11,14 por cento; Professor assistente, 657 mulheres e 1.403 homens, 2,62 por cento de mulheres, 5,60 por cento de homens e total de 2.060 docentes, 8,22 por cento; Tutor EaD, 393 mulheres e 266 homens, 1,57 por cento de mulheres, 1,06 por cento de homens e total de 659 docentes, 2,63 por cento; Professor auxiliar, 120 mulheres e 194 homens, 0,48 por cento de mulheres, 0,77 por cento de homens e total de 314 docentes, 1,25 por cento; Professor visitante, 61 mulheres e 143 homens, 0,24 por cento de mulheres, 0,57 por cento de homens e total de 204 docentes, 0,81 por cento; Docente EaD, 125 mulheres e 128 homens, 0,50 por cento de mulheres, 0,51 por cento de homens e total de 253 docentes, 1,01 por cento; Docente do ensino básico, técnico e tecnológico, 129 mulheres e 127 homens, 0,51 por cento de mulheres, 0,51 por cento de homens e total de 256 docentes, 1,02 por cento; Professor colaborador titular, 11 mulheres e 65 homens, 0,04 por cento de mulheres, 0,26 por cento de homens e total de 76 docentes, 0,30 por cento; Professor substituto, 62 mulheres e 58 homens, 0,25 por cento de mulheres, 0,23 por cento de homens e total de 120 docentes, 0,48 por cento; Auxiliar de docente, 45 mulheres e 44 homens, 0,18 por cento de mulheres, 0,18 por cento de homens e total de 89 docentes, 0,36 por cento; Professor colaborador sem informação de nível, 45 mulheres e 26 homens, 0,18 por cento de mulheres, 0,10 por cento de homens e total de 71 docentes, 0,28 por cento.

cento de homens e total de 71 docentes, 0,28 por cento; Outro tipo de tutor, 24 mulheres e 12 homens, 0,10 por cento de mulheres, 0,05 por cento de homens e total de 36 docentes, 0,14 por cento; Outras categorias, 34 mulheres e 51 homens, 0,14 por cento de mulheres, 0,20 por cento de homens e total de 85 docentes, 0,34 por cento. O total identificado soma 11.349 mulheres e 13.663 homens, representando 45,28 por cento de mulheres, 54,51 por cento de homens e totalizando 25.012 docentes, equivalente a 99,80 por cento. A linha Não identificado registra 51 docentes sem informação de gênero, equivalente a 0,20 por cento. O Total Geral apresenta 25.063 docentes, correspondendo a 100 por cento do conjunto analisado *[Fim da Descrição]*.

A tabela apresenta a distribuição de docentes por categoria profissional, discriminada por gênero e percentuais relativos. As colunas mostram o perfil do docente, o número e percentual de mulheres e homens, e os totais gerais. Observa-se predominância de docentes classificados como “Professor (sem informação de nível)”, representando 40,2% do total, seguidos por professores titulares (16,49%) e associados (15,11%). O conjunto de docentes é composto por 45,3% de mulheres e 54,5% de homens, indicando leve predominância masculina nas posições de maior titulação.

4.3 Localização Geográfica dos docentes

A análise da distribuição geográfica dos professores foi realizada levando-se em conta o número total de vínculos, com repetição. Do total de 25.063 vínculos, 333 não possuíam informação sobre a localização da instituição. A região Sudeste se destaca com quase a metade dos vínculos de professores. Em seguida, a maior concentração está nas regiões Nordeste e Sul.

Conforme se vê na tabela 5, em relação à distribuição nas unidades da federação, há vínculos em todas as unidades da federação. São Paulo concentra ¼ dos perfis dos docentes, seguido por Rio de Janeiro (9,9%), Minas Gerais (8,5%), Paraná (7,5%) e Rio Grande do Sul (6,2%).

Tabela 5

Distribuição dos vínculos dos docentes ativos em 2020 segundo Região e Unidade da Federação

Região	UF	Quantidade de Vínculos	%
Sudeste	São Paulo	6.327	25,50
	Rio de Janeiro	2.434	9,90
	Minas Gerais	2.089	8,50
	Espírito Santo	261	1,10
	Total (Sudeste)	11.111	44,30
	Bahia	1.207	4,90
Nordeste	Rio Grande do Norte	898	3,60
	Pernambuco	831	3,40
	Ceará	772	3,10
	Paraíba	685	2,80
	Alagoas	543	2,20
	Maranhão	321	1,30
Sul	Piauí	265	1,10
	Total (Nordeste)	5.724	22,80
	Paraná	1.850	7,50
	Rio Grande do Sul	1.540	6,20
Centro-Oeste	Santa Catarina	999	4,00
	Total (Sul)	4.389	17,50
	Goiás	772	3,10
	Distrito Federal	622	2,50
	Mato Grosso	369	1,50

Região	UF	Quantidade de Vínculos	%
Norte	Mato Grosso do Sul	200	0,80
	Total (Centro-Oeste)	1.963	7,80
	Pará	652	2,60
	Amazonas	278	1,10
	Tocantins	219	0,90
	Rondônia	160	0,60
	Acre	116	0,50
	Amapá	72	0,30
	Roraima	46	0,20
	Total (Norte)	1.543	6,20
Não Identificado	Total (Não Identificado)	333	1,30
Total Geral		25.063	100%

Nota. Fonte: Elaboração própria (2025). *[Início da Descrição]* A tabela apresenta quatro colunas intituladas Região, UF, Quantidade de vínculos e porcentagem, organizando a distribuição dos vínculos docentes por unidade da federação, somados por região. Na região Sudeste, São Paulo possui 6.327 vínculos, equivalente a 25,50 por cento; Rio de Janeiro tem 2.434 vínculos, 9,90 por cento; Minas Gerais tem 2.089 vínculos, 8,50 por cento; Espírito Santo tem 261 vínculos, 1,10 por cento; totalizando 11.111 vínculos, correspondentes a 44,30 por cento. Na região Nordeste, Bahia possui 1.207 vínculos, 4,90 por cento; Rio Grande do Norte tem 898 vínculos, 3,60 por cento; Pernambuco tem 831 vínculos, 3,40 por cento; Ceará tem 772 vínculos, 3,10 por cento; Paraíba tem 685 vínculos, 2,80 por cento; Alagoas tem 543 vínculos, 2,20 por cento; Maranhão tem 321 vínculos, 1,30 por cento; Piauí tem 265 vínculos, 1,10 por cento; totalizando 5.724 vínculos, 22,80 por cento. Na região Sul, Paraná possui 1.850 vínculos, 7,50 por cento; Rio Grande do Sul tem 1.540 vínculos, 6,20 por cento; Santa Catarina tem 999 vínculos, 4,00 por cento; totalizando 4.389 vínculos, 17,50 por cento. Na região Centro-Oeste, Goiás apresenta 772 vínculos, 3,10 por cento; Distrito Federal tem 622 vínculos, 2,50 por cento; Mato Grosso tem 369 vínculos, 1,50 por cento; Mato Grosso do Sul tem 200 vínculos, 0,80 por cento; totalizando 1.963 vínculos, 7,80 por cento. Na região Norte, Pará tem 652 vínculos, 2,60 por cento; Amazonas registra 278 vínculos, 1,10 por cento; Tocantins tem 219 vínculos, 0,90 por cento; Rondônia tem 160 vínculos, 0,60 por cento; Acre tem 116 vínculos, 0,50 por cento; Amapá tem 72 vínculos, 0,30 por cento; Roraima tem 46 vínculos, 0,20 por cento; totalizando 1.543 vínculos, 6,20 por cento. A categoria Não Identificado registra 333 vínculos, equivalentes a 1,30 por cento. O total geral apresenta 25.063 vínculos, correspondendo a 100 por cento do conjunto analisado *[Fim da Descrição]*.

Em relação às cidades, há vínculos em 908 municípios diferentes, o que equivale a 16,3% dos municípios brasileiros. Os municípios com mais vínculos são 12 capitais (São Paulo - 10,9%, Rio de Janeiro - 6,4%, Belo Horizonte - 2,7%, Curitiba - 2,7%, Porto Alegre - 2,6%, Brasília - 2,5%, Recife - 2,4%, Salvador - 2%, Fortaleza - 1,7%, Goiânia - 1,6%, Natal - 1,4%, João Pessoa - 1%), que juntamente com cidades universitárias tradicionais (Campinas - 3%, Ribeirão Preto - 1,4%, Londrina - 1,4%, São Carlos - 1,3% e Niterói - 1,3%), concentram 50% dos vínculos.

A distribuição de docentes pela quantidade de cidades em que atuam revela que a vasta maioria trabalha em apenas uma cidade. Especificamente, 23.427 docentes (ou cerca de 99% do total - N = 23720) têm atuação em uma única cidade, conforme indica a tabela 7. Um caso a se destacar é o de um único docente da região Sul, que atua em quatro cidades diferentes da mesma região, sendo este o único exemplo de um docente com múltiplos vínculos em quatro cidades.

Para analisar a distribuição dos perfis dos professores no ORCID em relação ao universo de vínculos docentes do país, foi realizada a comparação com os dados do Censo da Educação Superior, base 2020, a partir da consulta aos microdados. Os dados são apresentados na Figura 1. É interessante notar que, apesar de ser uma amostra pequena do universo, a distribuição mostra-se semelhante entre as unidades federativas.

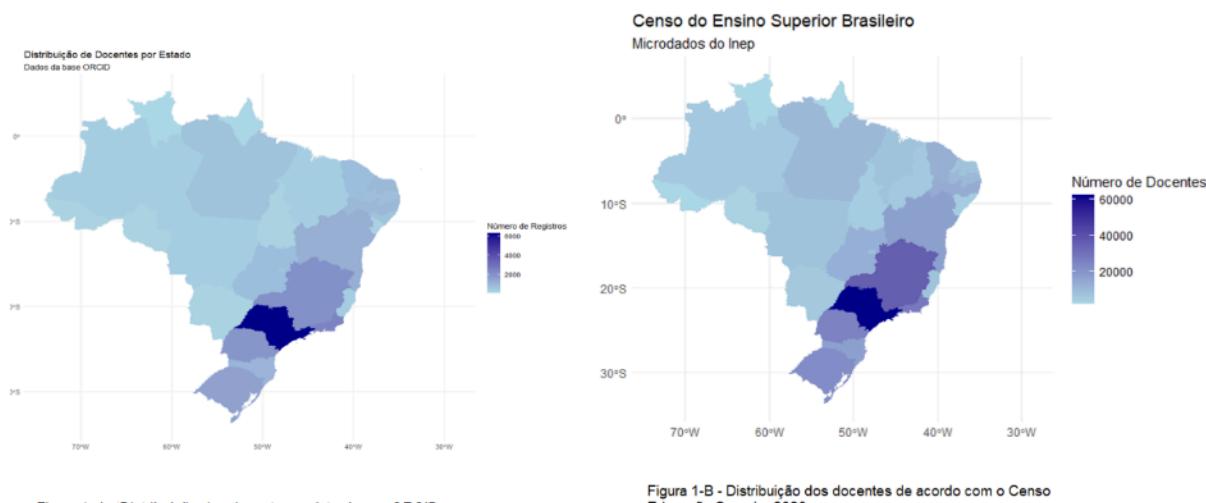
4.4 Distribuição segundo tipologia de IES

Outra forma de analisar possíveis vieses da base ORCID em relação ao universo de docentes é a comparação dos tipos de IES utilizando a classificação de Schwartzman et al. (2021), que agrupa as instituições a partir de vários critérios como intensidade em pesquisa (traduzida como pós-graduação) e ensino, categoria administrativa, organização, entre outras categorias. Durante a análise comparativa entre os dados do ORCID e a base de tipologias de IES proposta por Simon Schwartzman, foi identificado um significativo mismatch entre as IES, especialmente no que se refere às siglas utilizadas. No ORCID, fomos capazes de identificar um total de 1.835 IES, sendo que apenas 51 registros ficaram sem identificação (0,002%). Deste conjunto, foi possível encontrar correspondência para apenas 814 IES, quando realizamos o cruzamento com a base de Schwartzman, o que representou 45%.

Na comparação com os microdados do Censo da Educação Superior de 2020, nota-se uma sobrerepresentação na base ORCID de vínculos com IES mais intensivas em pesquisa (tipos 2, 3 e 9) e uma sub-representação nos demais tipos de IES, especialmente naquelas instituições públicas e privadas mais dedicadas ao ensino de graduação (tipos 1, 4, 5 e 7) e a cursos vocacionais (tecnológicos) (tipo 8).

Figura 1

Mapa da distribuição dos vínculos docentes por Unidade da Federação, ORCID e Censo da Educação Superior 2020



Nota. Fonte: Elaboração própria (2025). **[Início da Descrição]** A imagem contém duas figuras lado a lado. A figura à esquerda se chama Figura 1-A – Distribuição dos docentes registrados no ORCID. Ela apresenta um mapa do Brasil com os estados coloridos em tons de azul, variando do azul-claro ao azul-escuro conforme o número de docentes registrados na base ORCID. Os estados do Sudeste exibem as tonalidades mais escuras, com destaque para São Paulo, seguido por Rio de Janeiro e Minas Gerais. Regiões do Sul, como Rio Grande do Sul e Paraná, também apresentam intensidade elevada. Estados das regiões Norte e Nordeste aparecem em tons mais claros, indicando menor quantidade de registros. Uma legenda à direita define o gradiente de cores, variando aproximadamente de 2.000 a 6.000 registros. A figura à direita se chama Figura 1-B – Distribuição dos docentes de acordo com o Censo Educação Superior 2020. Ela também apresenta o mapa do Brasil, estruturado da mesma forma, com variação de azul-claro a azul-escuro indicando o número de docentes registrados no Censo do Ensino Superior. Nesta figura, São Paulo aparece novamente com a maior intensidade, seguido por Minas Gerais, Paraná e Rio Grande do Sul. Estados com menor número de docentes surgem em tons mais claros, a exemplo de estados do Norte como Roraima, Acre e Amapá. A legenda indica valores aproximados entre 20.000 e 60.000 docentes. As duas figuras utilizam a mesma base territorial, mas representam bases de dados distintas, destacando diferenças de concentração regional **[Fim da Descrição]**.

Tabela 6

Distribuição de docentes ativos e número de IES, segundo dados do Censo da Educação Superior 2020 e da base ORCID

Tipologia	% Docentes Censo	% Docentes ORCID	% IES Censo	% IES ORCID
1 - Privadas de grande porte	5,7	3,6	1	2,3
2 - Públicas de grande porte com pós-graduação alta	10,7	25,3	0,4	1,6
3 - Públicas especializadas com pós-graduação alta	9,2	13,9	1	2,8
4 - Privadas comunitárias e diferenciadas	8,6	6,3	3	8,4
5 - Universidades e Centros universitários privados	18,5	8,5	10	22,6
6 - Outras instituições públicas	23,5	23,9	5	15,8
7 - Faculdades privadas isoladas	20,4	4,1	53	37,7
8 - Instituições de educação tecnológica	3,3	1,5	9	4,3
9 - Instituições de Pesquisa e Pós-Graduação	0,1	0,6	3	1,5
IES Sem informação na Base ORCID	0	17,6	14	2,9
% Total	100	100	100	100
IES não identificadas no cruzamento	-	-	-	1021
IES identificadas no cruzamento	-	-	-	814
Total	N= 314.383	N=23.793	N=2.945	N=1835

Nota. Fonte: Elaboração própria (2025). *[Início da Descrição]* A tabela apresenta cinco colunas: Tipologia, porcentagem de docentes segundo o Censo, porcentagem de docentes segundo o ORCID, porcentagem de IES segundo o Censo e porcentagem de IES segundo o ORCID. As tipologias e seus valores são: 1 – Privadas de grande porte, com 5,7 por cento dos docentes no Censo, 3,6 por cento no ORCID, 1 por cento das IES no Censo e 2,3 por cento no ORCID; 2 – Públicas de grande porte com pós-graduação alta, 10,7 por cento dos docentes no Censo, 25,3 por cento no ORCID, 0,4 por cento das IES no Censo e 1,6 por cento no ORCID; 3 – Públicas especializadas com pós-graduação alta, 9,2 por cento dos docentes no Censo, 13,9 por cento no ORCID, 1 por cento das IES no Censo e 2,8 por cento no ORCID; 4 – Privadas comunitárias e diferenciadas, 8,6 por cento dos docentes no Censo, 6,3 por cento no ORCID, 3 por cento das IES no Censo e 8,4 por cento no ORCID; 5 – Universidades e Centros universitários privados, 18,5 por cento dos docentes no Censo, 8,5 por cento no ORCID, 10 por cento das IES no Censo e 22,6 por cento no ORCID; 6 – Outras instituições públicas, 23,5 por cento dos docentes no Censo, 23,9 por cento no ORCID, 5 por cento das IES no Censo e 15,8 por cento no ORCID; 7 – Faculdades privadas isoladas, 20,4 por cento dos docentes no Censo, 4,1 por cento no ORCID, 53 por cento das IES no Censo e 37,7 por cento no ORCID; 8 – Instituições de educação tecnológica, 3,3 por cento dos docentes no Censo, 1,5 por cento no ORCID, 9 por cento das IES no Censo e 4,3 por cento no ORCID; 9 – Instituições de Pesquisa e Pós-Graduação, 0,1 por cento dos docentes no Censo, 0,6 por cento no ORCID, 3 por cento das IES no Censo e 1,5 por cento no ORCID. A linha IES Sem informação na Base ORCID registra zero por cento no Censo, 17,6 por cento no ORCID, 14 por cento das IES no Censo e 2,9 por cento no ORCID. A linha % Total indica 100 por cento nas quatro colunas de distribuição. As linhas finais indicam IES não identificadas no cruzamento, totalizando 1.021 instituições, e IES identificadas no cruzamento, totalizando 814 instituições. Os valores totais apresentados ao final são N = 314.383 para docentes no Censo, N = 23.793 para docentes no ORCID, N = 2.945 para IES no Censo e N = 1.835 para IES na base ORCID *[Fim da Descrição]*.

5 Discussão

Os registros do ORCID representam uma amostra pequena do conjunto de docentes brasileiros. Apesar de pequena, os perfis na base ORCID apresentam distribuições semelhantes ao universo em termos de gênero e distribuição geográfica, repetindo as desigualdades de gênero e assimetrias regionais existentes, com uma concentração maior no Sul e Sudeste.

Em linha com a hipótese levantada a partir dos estudos mais amplos sobre desigualdade de gênero na profissão acadêmica (Cruz-Castro et al., 2022; Hipólito et al., 2022; Atães, 2023), a sub representação feminina na academia é refletida na base ORCID.

Entretanto, a base ORCID apresenta vieses importantes em termos da posição na carreira (mais frequente nos níveis mais seniores, dentre os identificados) e na vinculação aos tipos de IES (mais frequentes em IES mais intensivas em pesquisa). Em relação à distribuição dos vínculos, segundo os tipos de IES, os resultados vão na mesma linha que o estudo de Porter (2022, p.6), que verificou que “altos níveis de engajamento em pesquisa implicam um alto nível de manutenção de registros ORCID”.

Em relação ao nível na carreira, os achados apontam numa direção diferente dos estudos de Porter (2022) e Gomez et al. (2020), que reportaram o uso mais frequente do ORCID por pesquisadores em início de carreira, pois a cultura organizacional e as pressões das editoras e agências de financiamento têm sido maior sobre estes no sentido da adoção do ORCID. Também têm sido mais comum a utilização desde as etapas de formação (graduação, mestrado e doutorado), o que não ocorreu com os professores de gerações mais velhas, que tendem a estar em posições mais seniores da carreira acadêmica. Além disso, os docentes nas posições mais seniores tendem a ser menos numerosos na carreira acadêmica. Por exemplo, a Unicamp estabeleceu em 2007 um teto máximo de 35% de professores titulares em cada unidade (Unicamp, 2014).

A diferença de gênero entre os professores nos níveis da carreira também chamou a atenção, pois como apresentado na Tabela 4, há uma proporção maior de professores homens em relação às professoras em todas as posições. Essa discrepância é especialmente notável na posição mais elevada, na qual os homens (11,99%) representam quase três vezes mais que as mulheres (4,5%). É importante destacar que não há dados sobre a posição na carreira disponíveis no Censo da Educação Superior, o que limita a possibilidade de contextualizar esses resultados no universo de docentes de IES brasileiras. Neste sentido, pode ser que as mulheres preencham menos a informação sobre a posição na carreira ou que os dados mostrem, de fato, a desigualdade de gênero, refletindo o “efeito tesoura” na ciência brasileira (maior presença nas posições de ingresso e menores nas posições mais altas quando comparadas aos homens), como já constatado por outros estudos (Arêas et al., 2020; Valentova et al., 2017).

A base parece refletir melhor os professores que atuam mais intensamente na pesquisa e pós-graduação, além de estarem vinculados a IES que provavelmente têm tornado compulsório o registro no ORCID. Pode indicar, ainda, a maior cobertura de profissionais com maior circulação internacional, o que demandaria a inclusão do CV (Curriculum Vitae) no ORCID, que dá uma maior visibilidade neste contexto. Entretanto, é necessário levar em consideração que a classificação manual do nível da carreira feita neste estudo pode apresentar vieses.

Um resultado interessante foi a alta frequência de vínculos únicos, o que pode indicar alta estabilidade e baixa mobilidade, em linha com o estudo de Furtado et al. (2015), que identificou baixas taxas de mobilidade acadêmica, seja nacional ou internacional. Mas também pode indicar a incompletude dos dados, pois como apontado por Porter (2022), os pesquisadores podem adotar o ORCID e criar um registro com as informações atuais, mas não se engajarem em preencher dados passados ou continuar atualizando-os.

Como mencionado na revisão de literatura, ao redor do mundo existem iniciativas institucionais que incentivam a adesão ao ORCID, incluindo o Consórcio formado no Brasil envolvendo CAPES, CNPq, CONFAP, SciELO, Ibict, RNP, Unesp, USP, Unicamp e Fiocruz. A Unesp e a USP, por exemplo, foram pioneiras no Brasil ao estimular seus professores a se registrarem, o que pode ter gerado uma maior concentração de docentes dessas instituições na base. Além disso, a adesão pioneira dessas instituições ajuda a explicar a concentração geográfica em São Paulo e no Sudeste, além da maior concentração de IES mais intensivas em pesquisa no Sudeste (Sidone, Haddad & Mena-Chalco, 2016).

Desta forma, é importante reconhecer que, como a adesão ao ORCID não é obrigatória para todos os docentes no Brasil, isso pode introduzir um viés de autosseleção. Docentes mais envolvidos em pesquisa ou com maior visibilidade internacional tendem a ser mais propensos a manter perfis ativos, enquanto outros podem não estar representados na base.

6 Conclusões

A base ORCID coloca-se como uma grande promessa para estudos cientométricos, mas há poucos trabalhos que exploram este potencial (Costas et al., 2022). O quadro tem mudado rapidamente e se nota cobertura crescente no Sul Global e no Brasil (Costas et al., 2022; Porter, 2022; Youtie et al., 2017).

Este estudo pioneiro atende ao gap identificado na literatura de estudos locais sobre a cobertura devido às diferenças na adoção ao redor do mundo e por características dos profissionais, como posição na carreira e gênero. Neste sentido, o estudo traz informações relevantes sobre a qualidade dos dados no Brasil para contextualizá-los em próximos estudos internacionais.

O primeiro resultado de destaque é que se trata de uma amostra pequena do universo de professores brasileiros. Na comparação com dados censitários do Ministério da Educação, esta amostra parece representativa do universo em termos de gênero e distribuição geográfica, refletindo distribuições desiguais existentes. Também parece refletir a baixa mobilidade e alta estabilidade na profissão acadêmica no país. Entretanto, possui vieses relevantes em termos de posição na carreira e distribuição segundo tipos de instituições de ensino superior.

É importante situar os limites desta pesquisa, alguns deles referentes às próprias limitações da base ORCID. Em primeiro lugar, a adoção dos filtros sucessivos (organização localizada no Brasil e ocupação de professor), excluiu perfis que não preencheram estas informações no ORCID.

Outro limite diz respeito à classificação das posições na carreira. O campo aberto permite o detalhamento da ocupação e nível na carreira, mas a falta de padronização ou de instruções para o preenchimento dificulta a categorização. A classificação pode ter privilegiado posições dos planos das carreiras das instituições públicas federais e estaduais paulistas, dado que os planos de carreiras no setor privado são locais (Rodrigues Filho, 2015).

O ORCID depende de informações auto-declaradas, e muitos perfis não estão completos (Fernández-Marcial et al., 2023; Porter, 2022). Isso exige o desenvolvimento de estratégias para sanar as incompletudes ou tratar os dados, o que pode envolver tarefas complexas. Como já mencionado, a falta de padronização se estende a todos os campos da base. No caso específico da identificação das IES e suas siglas correspondentes, isso pode gerar dificuldades na correspondência entre a base ORCID e outras bases, como a do Censo da Educação Superior. Uma recomendação seria que o ORCID adotasse padrões globais de padronização de instituições e siglas, utilizando identificadores reconhecidos, como o Research Organization Registry (ROR).

O mismatch identificado no cruzamento das IES da base ORCID com a tipologia de IES sugere a necessidade de uma análise mais detalhada e rigorosa sobre os dados de ambas as bases. A padronização das siglas e nomenclaturas de instituições é essencial para uma análise mais consistente dos dados.

Tendo em vista o potencial de crescimento da cobertura, abre-se uma agenda de pesquisa para estudos sobre a carreira acadêmica no Brasil usando a base ORCID, incluindo a análise da endogamia, mobilidade entre instituições, a progressão na carreira, caracterização dos docentes em relação à posição na carreira e diferenças entre áreas, com a complementação de dados de publicações.

Outra frente de estudos é a comparação com outras bases nacionais, como a Plataforma Lattes e a Base RAIS (Relação Anual de Informações Sociais). A Plataforma Lattes é reconhecida com uma base de dados de cobertura praticamente censitária em relação à academia, diferente de outros países. Já a base RAIS é censitária em relação aos professores com vínculo empregatício formal. A base ORCID, apesar da menor cobertura, apresenta uma oportunidade de realizar comparações internacionais, bem como acompanhar os cientistas que saem do Brasil e deixam de atualizar o Lattes.

Por fim, a terceira agenda para estudos futuros inclui a comparação da cobertura da base ORCID com outros países da América Latina e Sul Global.

Referências

- Aarrevaara, T., Finkelstein, M., Jones, G. A., & Jung, J. (2021). The Academic Profession in the Knowledge-Based Society (APIKS): Evolution of a Major Comparative Research Project. Em T. Aarrevaara, M. Finkelstein, G. A. Jones, & J. Jung (Orgs.), *Universities in the Knowledge Society* (Vol. 22, p. 49–64). Springer International Publishing. https://doi.org/10.1007/978-3-030-76579-8_4
- Abel, G. J., Muttarak, R., Bordone, V., & Zagheni, E. (2019). Bowling Together: Scientific Collaboration Networks of Demographers at European Population Conferences. *European Journal of Population*, 35(3), 543–562. <https://doi.org/10.1007/s10680-018-9493-1>
- Aksnes, D. W., Piro, F. N., & Rørstad, K. (2019). Gender gaps in international research collaboration: A bibliometric approach. *Scientometrics*, 120(2), 747–774. <https://doi.org/10.1007/S11192-019-03155-3>
- Aman, V. (2018). A new bibliometric approach to measure knowledge transfer of internationally mobile scientists. *Scientometrics*, 117(1), 227–247. <https://doi.org/10.1007/s11192-018-2864-x>
- Arêas, R., Santana, A. E., & Barbosa, Marcia C. (2020). A discriminação da intelligentsia no campo científico brasileiro. *Revista Educação, Cultura e Sociedade*, 10(3), 02-18. <https://doi.org/10.30681/ecs.v10i3.3943>
- Atães, D. (2023). Gênero, raça e profissão acadêmica: desigualdades em instituições de ensino superior brasileiras. *Anais do Congresso Brasileiro de Sociologia*, Belém. <https://www.sbs2023.sbsociologia.com.br/arquivo/downloadpublic?q=YTo4OntzOjY6InBhomFtcyl7czozNToiYToxOntzOjEwOjJJRF9BUlFVSVZPljtzOjQ6ljUzNzEiO30iO3M6MTaiaCl7czozMjoiYWlzMjZlYzFIMzlxDhY2l5NTgzOGUxZDI3ZmlxMDAiO30%3D>
- Brasil. (2012). *Lei No 12.772, de 28 de dezembro de 2012*. Dispõe sobre a estruturação do Plano de Carreiras e Cargos de Magistério Federal; sobre a Carreira do Magistério Superior, de que trata a Lei no 7.596, de 10 de abril de 1987; sobre o Plano de Carreira e Cargos de Magistério do Ensino Básico, Técnico e Tecnológico e sobre o Plano de Carreiras de Magistério do Ensino Básico Federal, de que trata a Lei no 11.784, de 22 de setembro de 2008; altera remuneração do Plano de Cargos Técnico-Administrativos em Educação e dá outras providências. https://www.planalto.gov.br/ccivil_03/ato2011-2014/2012/lei/L12772compilado.htm
- Brasil. *Lei de Diretrizes e Bases da Educação Nacional, Lei no 9394 (1996)*. http://www.planalto.gov.br/ccivil_03/leis/l9394.htm
- Brasil. Ministério da Educação. (2024). *Cadastro Nacional de Cursos e Instituições de Educação Superior – Cadastro e-MEC* [Dataset]. <https://emecc.mec.gov.br/>
- Cañibano, C., Otamendi, J., & Andújar, I. (2008). Measuring and assessing researcher mobility from CV analysis: The case of the Ramón y Cajal programme in Spain. *Research Evaluation*, 17(1), 17–31. <https://doi.org/10.3152/095820208X292797>
- Carley, S. F., Porter, A. L., & Youtie, J. L. (2019). A Multi-match Approach to the Author Uncertainty Problem. *Journal of Data and Information Science*, 4(2), 1–18. <https://doi.org/10.2478/jdis-2019-0006>
- Costa, Alexandre (2015). Lista de Todas as Universidades do Brasil (retirado do portal de indicadores do INEP/ENADE). [Dataset] <https://gist.github.com/alexandremcosta/c9361cc23722a5aa1133>
- Costas, R., Corona, C., & Robinson-Garcia, N. (2022). *Could ORCID play a key role in meta-research? Discussing new analytical possibilities to study the dynamics of science and scientists*. <https://doi.org/10.31235/osf.io/sjck6>
- CRUESP. (2010). *Resolução CRUESP no 10/2010 de 21/12/2010*. Altera a Resolução CRUESP no 01/2010, de 01/03/2010, que dispõe sobre a reestruturação da Carreira Magistério Superior dos docentes da Universidade de São Paulo, Universidade Estadual de Campinas e Universidade Estadual Paulista “Júlio de Mesquita Filho” e dá outras providências. CRUESP.
- Cruz-Castro, L., Ginther, D., & Sanz-Menendez, L. (2022). *Gender and Underrepresented Minority Differences in Research Funding*. NBER Working Paper Series 30107. National Bureau of Economic Research. <https://doi.org/10.3386/w30107>

- Dasler, R., Deane-Pratt, A., Lavasa, A., Rueda, L., & Dallmeier-Tiessen, S. (2017). *Study of ORCID adoption across disciplines and locations*. Zenodo. <https://doi.org/10.5281/zenodo.841777>
- Fernández-Marcial, V., González-Solar, L., & Vale, A. (2023). Is ORCID your ID? A case study at the Faculty of Arts and Humanities of the University of Porto. *Learned Publishing*, 36(4), 564–576. <https://doi.org/10.1002/leap.1562>
- Franzoni, C., Scellato, G., & Stephan, P. E. (2012). *Foreign-born scientists*: Mobility patterns for 16 countries. *Nature Biotechnology*. <https://doi.org/10.1038/nbt.2449>
- Gibb, C. (2022). The home-based postdoctoral mother in the neoliberal university. *Gender, Place & Culture*, 29(2), 248–272. <https://doi.org/10.1080/0966369X.2021.1975102>
- Gomez, C. J., Herman, A. C., & Parigi, P. (2020). Moving more, but closer: Mapping the growing regionalization of global scientific mobility using ORCID. *Journal of Informetrics*, 14(3). <https://doi.org/10.1016/j.joi.2020.101044>
- Haak, L. L., Meadows, A., & Brown, J. (2018). Using ORCID, DOI, and Other Open Identifiers in Research Evaluation. *Frontiers in Research Metrics and Analytics*, 3(28). <https://doi.org/10.3389/frma.2018.00028>
- Henderson, E. F. (2021). Sticky care and conference travel: Unpacking care as an explanatory factor for gendered academic immobility. *Higher Education*, 82(4), 715–730. <https://doi.org/10.1007/s10734-020-00550-1>
- Hipólito, J., Shirai, L. T., Halinski, R., Guidolin, A. S., Querino, R. B., Quintela, E. D., Pini, N. D. S. D., Pires, C. S. S., & Fontes, E. M. G. (2022). Brazilian female researchers do not publish less despite an academic structure that deepens sex gap. *PLOS ONE*, 17(8). <https://doi.org/10.1371/journal.pone.0273291>
- Hou, L., Wu, Q., & Xie, Y. (2022). Does early publishing in top journals really predict long-term scientific success in the business field? *Scientometrics*, 127(11), 6083–6107. <https://doi.org/10.1007/s11192-022-04509-0>
- Instituto Brasileiro de Geografia e Estatística (IBGE). (2022). FTP do IBGE. 2022. *Lista de municípios com limites atualizados na malha municipal, segundo Censo 2022* [Dataset]. https://ftp.ibge.gov.br/Censos/Censo_Demografico_2022/Populacao_e_domicilios_Primeiros_resultados/Anexos/CD2022_Municipios_com_limites_atualizados_na_malha_municipal_2022.xlsx
- Instituto Nacional de Estudos e Pesquisas Educacionais (INEP). (2020). *Sinopse Estatística da Educação Superior 2020*. INEP. <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/sinopses-estatisticas/educacao-superior-graduacao>
- Instituto Nacional de Estudos e Pesquisas Educacionais (INEP). (2024). *Estatísticas do Censo da Educação Superior* [Dataset]. INEP. <https://app.powerbi.com/view?r=eyJrIjoiMGJiMmNiNTAtOTY1OC00ZjUzLTg2OGUtMjAzYzNiYTA5Yjliwi dCI6ljI2ZjczODk3LWM4YWMtNGIxZS05NzhmLWVhNGMwNzc0MzRiZiJ9&pageName=ReportSection40 36c90b8a27b5f58f54>
- Jenkins, K. (2020). Academic motherhood and fieldwork: Juggling time, emotions, and competing demands. *Transactions of the Institute of British Geographers*, 45(3), 693–704. <https://doi.org/10.1111/tran.12376>
- Malakhov, V., Vasil'eva, I., & Belov, F. (2020). Structure of International Migration of Researchers. Case Study of Russia. *Economic and Social Changes: Facts, Trends, Forecast / Экономические и социальные перемены: факты, тенденции, прогноз*, 5(71). <https://doi.org/10.15838/esc.2020.5.71.14>
- Mascarenhas, H. A., & Dias, T. M. (2022). Análise da mobilidade acadêmica brasileira com ênfase na internacionalização e intensidade de relações. *Em Questão*, 25(4). <https://doi.org/10.19132/1808-5245284.120243>
- Marí-n-Arraiza, P., & Mejias, G. (2020). *Identificadores persistentes: la adopción del orcid iD en España*. Anuario ThinkEPI, 14. <https://doi.org/10.3145/thinkepi.2020.e14e06>
- Mejias, G. (2018). *ORCID na América Latina*: An update. <https://www.contentmind.com.br/orcid-na-america-latina>

- Miranda-González, A., Aref, S., Theile, T., & Zagheni, E. (2020). Scholarly migration within Mexico: Analyzing internal migration among researchers using Scopus longitudinal bibliometric data. *EPJ Data Science*, 9(1), 34. <https://doi.org/10.1140/epjds/s13688-020-00252-9>
- Moratti, S. (2021). A woman's place is in the 'home'? Gender-specific hiring patterns in academia in gender-equal Norway. *Journal of Sociology*, 57(4), 916–934. <https://doi.org/10.1177/1440783320960530>
- Nachatar Singh, J. K. (2021). International academics' lived experiences in gaining leadership positions at Australian universities. *International Journal of Leadership in Education*, 28(1), 1–14. <https://doi.org/10.1080/13603124.2021.1988717>
- Nikunen, M., & Lempäinen, K. (2020). Gendered strategies of mobility and academic career. *Gender and Education*, 32(4), 554–571. <https://doi.org/10.1080/09540253.2018.1533917>
- OpenAI. (2024). OpenAI API: GPT-4 model (September 2024) [Large language model]. Retrieved from <https://platform.openai.com/docs>
- ORCID (2025). Discover research from ORCID. [Datasets] <https://orcid.figshare.com/>
- ORCID Statistics. (2025). *ORCID*. <https://info.orcid.org/resources/orcid-statistics/>
- Paraskevopoulos, P., Boldrini, C., Passarella, A., & Conti, M. (2020). Dynamics of Scientific Collaboration Networks Due to Academic Migrations. In S. Aref, K. Bontcheva, M. Braghieri, F. Dignum, F. Giannotti, F. Grisolia, & D. Pedreschi (Orgs.), *Social Informatics*. 283–296. Springer International Publishing. https://doi.org/10.1007/978-3-030-60975-7_21
- Porter, S. J. (2022). Measuring Research Information Citizenship Across ORCID Practice. *Frontiers in Research Metrics and Analytics*, 7, 779097. <https://doi.org/10.3389/frma.2022.779097>
- Reis, J. E. D., Sarvo, D. D. O., Faria, L. I. L. D., & Amaral, R. M. D. (2021). Impact of teacher education abroad in international co-authorship: A study of Federal University of São Carlos's scientifics production indexed in the Web of Science. *Transinformação*, 33. <https://doi.org/10.1590/2318-0889202133e200061>
- Robles-Belmont, E. (2021). Mobility of the Members of the National System of Researchers of Mexico in their Academic Training in Latin America. *Journal of Scientometric Research*, 10(1s), 71–77. <https://doi.org/10.5530/jscires.10.1s.23>
- Rodrigues Filho, A. F. (2015). *Trabalho docente no ensino superior: Uma análise das políticas retributórias e seus efeitos* [Tese de doutorado, Universidade Estadual de Campinas]. <https://doi.org/10.47749/T/UNICAMP.2015.963455>
- RNP (2018). *Lançado consórcio brasileiro para assinatura de identificador digital de pesquisadores*. <https://www.rnp.br/lanado-consorcio-brasileiro-para-assinatura-de-identificador-digital-de-pesquisadores/>
- Schittenhelm, K. (2022). Negotiating unequal mobility opportunities: Young academics' mobility in the EU and their partner-related living arrangements. *Geoforum*, 133, 117–127. <https://doi.org/10.1016/j.geoforum.2022.03.020>
- Schnieders, K., Mierz, S., Boccalini, S., Meyer Zu Westerhausen, W., Hauschke, C., Hagemann-Wilholt, S., & Schulze, S. (2022). ORCID coverage in research institutions—Readiness for partially automated research reporting. *Frontiers in Research Metrics and Analytics*, 7, 1010504. <https://doi.org/10.3389/frma.2022.1010504>
- Schwartzman, S., & Bueno, A. C. (2023). A New Typology of Higher Education Institutions in Brazil. *Proceedings* of the 35th annual conference of the Consortium of Higher Education Researchers (CHER). 35th annual conference of the Consortium of Higher Education Researchers (CHER), Viena. https://www.researchgate.net/publication/374170892_A_New_Typology_of_Higher_Education_Institutions_in_Brazil
- Schwartzman, S., Filho, R. L. S., & Coelho, R. R. A. (2021). Por uma tipologia do ensino superior brasileiro: Teste de conceito. *Estudos Avançados*, 35(101), 153–186. <https://doi.org/10.1590/s0103-4014.2021.35101.011>

- Sidone, O. J. G., Haddad, E. A., & Mena-Chalco, J. P. (2016). A ciência nas regiões brasileiras: evolução da produção e das redes de colaboração científica. *Transinformação*, 28(1), 15–32. <https://doi.org/10.1590/2318-08892016002800002>
- Siekierski, P., Lima, M. C., & Borini, F. M. (2019). International academic mobility and patent filing in the country of origin. *Revista de Administracão Pública*, 53(3), 560–574. <https://doi.org/10.1590/0034-761220170378>
- Sixto-Costoya, A., Robinson-Garcia, N., Leeuwen, V. T., & Costas, R. (2021). Exploring the relevance of ORCID as a source of study of data sharing activities at the individual-level: A methodological discussion. *Scientometrics*, 126(8), 7149–7165. <https://doi.org/10.1007/s11192-021-04043-5>
- Sobrido Prieto, M., Gutiérrez Couto, U., & González Gutián, C. (2016). From the standardization of scientific signatures to identification with digital signatures. *Index de Enfermería*, 25(1–2), 56–59. <https://www.index-f.com/index-enfermeria/v25n1-2/sumario.php>
- Teixeira da Silva, J. A. (2021). Abuse of ORCID's weaknesses by authors who use paper mills. *Scientometrics*, 126(7), 6119–6125. <https://doi.org/10.1007/s11192-021-03996-x>
- Tzanakou, C. (2021). Stickiness in academic career (im)mobilities of STEM early career researchers: An insight from Greece. *Higher Education*, 82(4), 695–713. <https://doi.org/10.1007/s10734-020-00596-1>
- Uhly, K. M., Visser, L. M., & Zippel, K. S. (2017). Gendered patterns in international research collaborations in academia. *Studies in Higher Education*, 42(4), 760–782. <https://doi.org/10.1080/03075079.2015.1072151>
- Unicamp (2014). *Define procedimentos para atribuição de vagas para concursos públicos para provimento de cargo de professor titular e para promoção por mérito para o nível MS-6*, Deliberação CONSU-A-026/2014. <https://www.pg.unicamp.br/norma/3729/0>
- Valentine, A., & Williams, B. (2023). *Identifying career pathways to engineering education research in Australia*. 2023 5th International Conference of the Portuguese Society for Engineering Education (CISPEE), 1–5. <https://doi.org/10.1109/CISPEE58593.2023.10227653>
- Valentova, J. V., Otta, E., Silva, M. L., & McElligott, A. G. (2017). *Underrepresentation of women in the senior levels of Brazilian science*. PeerJ, 5, e4000. <https://doi.org/10.7717/peerj.4000>
- Vohlídalová, M. (2017). Academic couples, parenthood and women's research careers. *European Educational Research Journal*, 16(2–3), 166–182. <https://doi.org/10.1177/1474904116668883>
- Wendt, K., Gunnes, H., & Aksnes, D. W. (2022). International migration of researchers and gender imbalance in academia—The case of Norway. *Scientometrics*, 127(12), 7575–7591. <https://doi.org/10.1007/s11192-022-04365-y>
- Yan, E., Zhu, Y., & He, J. (2020). Analyzing academic mobility of U.S. professors based on ORCID data and the Carnegie Classification. *Quantitative Science Studies*, 1(4), 1451–1467. https://doi.org/10.1162/qss_a_00088
- Youtie, J., Carley, S., Porter, A. L., & Shapira, P. (2017). Tracking researchers and their outputs: New insights from ORCIDs. *Scientometrics*, 113(1), 437–453. <https://doi.org/10.1007/s11192-017-2473-0>
- Zhang, L., Lu, W., & Yang, J. (2023). LAGOS-AND: A large gold standard dataset for scholarly author name disambiguation. *Journal of the Association for Information Science and Technology*, 74(2), 168–185. <https://doi.org/10.1002/asi.24720>
- Zhang, L., Qian, Y., Ma, C., & Li, J. (2023). Continued collaboration shortens the transition period of scientists who move to another institution. *Scientometrics*, 128(3), 1765–1784. <https://doi.org/10.1007/s11192-022-04617-x>
- Zhao, Z., Bu, Y., & Li, J. (2022). Does the mobility of scientists disrupt their collaboration stability? *Journal of Information Science*, 48(2), 199–209. <https://doi.org/10.1177/0165551520948744>

Apêndice A – Seção 3 Metodologia

Tabela 1A.

Relação das variáveis da base Orcid

Nome	Descrição
path	Identificador único (ID ORCID) associado a um indivíduo
name	Nome completo do indivíduo
department	Departamento ou programa/área específica dentro da instituição
role	Função ou grau buscado pelo indivíduo (por exemplo, Bacharel, Mestre, Doutor)
start_date_year	Ano de início da função docente
start_date_month	Mês de início da função docente
start_date_day	Dia de início da função docente
end_date_year	Ano de término da função docente
end_date_month	Mês de término da função docente
end_date_day	Dia de término da função docente
organization_name	Nome da organização/instituição (IES)
organization_city	Cidade onde a organização/instituição (IES) está localizada
organization_region	Região/estado onde a organização/instituição (IES) está localizada
organization_country	País onde a organização/instituição (IES) está localizada
organization_url	Site ou URL associado à organização/instituição (IES)
disambiguated_organization_identifier	Identificador único da organização (por exemplo, RINGGOLD, GRID)
disambiguated_source	Fonte do identificador único da organização (por exemplo, RINGGOLD, GRID, FUNDREF)

Nota. Fonte: Elaborado pelo(s) autor(es) (2025). *[Início da Descrição]* A tabela apresenta duas colunas, Nome e Descrição, que organizam variáveis relacionadas a indivíduos e instituições. Ela inclui: “path”, identificador único ORCID associado a um indivíduo; “name”, nome completo do indivíduo; “department”, departamento ou área específica dentro da instituição; “role”, função ou grau buscado pelo indivíduo, como bacharel, mestre ou doutor; “start_date_year”, ano de início da função docente; “start_date_month”, mês de início da função docente; “start_date_day”, dia de início da função docente; “end_date_year”, ano de término da função docente; “end_date_month”, mês de término da função docente; “end_date_day”, dia de término da função docente; “organization_name”, nome da instituição; “organization_city”, cidade onde a instituição está localizada; “organization_region”, região ou estado da instituição; “organization_country”, país da instituição; “organization_url”, site ou URL da instituição; “disambiguated_organization_identifier”, identificador único da organização, como RINGGOLD ou GRID; e “disambiguated_source”, fonte do identificador da organização, como RINGGOLD, GRID ou FUNDREF *[Fim da Descrição]*.

Tabela 2A**Variáveis criadas no âmbito do estudo**

Nome	Descrição
é professor?.x	Indica se o indivíduo é professor (sim ou não)
organization_city_padrao.x	Cidade padronizada da organização/instituição de ensino
UF.x	Unidade Federativa (estado) onde a instituição está localizada
Região.x	Região geográfica brasileira da instituição
IES.x	Nome da Instituição de Ensino Superior (IES)
SIGLA.x	Sigla da Instituição de Ensino Superior (IES)
é ies?.x	Indica se a instituição é uma IES (sim ou não)
COIES.x	Código identificador único da IES
tipologia_2023.x	Classificação tipológica da IES em 2023
Rev AMC.x	Descrição do nível ou título do professor revisado (ex.: professor titular, associado)
gender_final.x	Gênero do indivíduo (male, female ou não informado)

Nota. Fonte: Elaborado pelo(s) autor(es) (2025). *[Início da Descrição]* A tabela intitulada Variáveis criadas no âmbito do estudo apresenta duas colunas, Nome e Descrição, organizando os identificadores das variáveis e seus significados. Ela inclui: “é professor?.x”, que indica se o indivíduo é professor (sim ou não); “organization_city_padrao.x”, que registra a cidade padronizada da organização ou instituição de ensino; “UF.x”, que informa a unidade federativa onde a instituição está localizada; “Região.x”, que aponta a região geográfica brasileira da instituição; “IES.x”, que apresenta o nome da Instituição de Ensino Superior; “SIGLA.x”, que registra a sigla da IES; “é ies?.x”, que identifica se a instituição é uma IES (sim ou não); “COIES.x”, que traz o código identificador único da instituição; “tipologia_2023.x”, que descreve a classificação tipológica da IES em 2023; “Rev AMC.x”, que apresenta o nível ou título do professor revisado, como professor titular ou associado; e “gender_final.x”, que informa o gênero do indivíduo (male, female ou não informado) *[Fim da Descrição]*.

Dados de publicação**Ana Maria Carneiro**

Doutora

Universidade Estadual de Campinas, Núcleo de Estudos de Políticas Públicas, Campinas, São Paulo, Brasil

anamacs@unicamp.br<https://orcid.org/0000-0001-6688-1881>

Doutora em Política Científica e Tecnológica pela Universidade Estadual de Campinas (Unicamp - 2007), mestre em Sociologia pela Unicamp (2000) e graduada em Ciências Sociais pela Universidade Federal de Goiás (1997). Possui Pós Doutorado em Avaliação em CTI (Unicamp, 2009) e em Educação Superior (University of California, Berkeley, 2016). Atualmente é Pesquisadora da Carreira Pq no Núcleo de Estudos de Políticas Públicas

(NEPP/Unicamp), Professora Permanente do Programa de Pós-Graduação em Política Científica e Tecnológica (DPCT/IG/Unicamp); Coordenadora Associada do Grupo de Estudos Sobre Organização da Pesquisa e da Inovação (GEOPI/DPCT/IG/Unicamp). Entre 2018 e 2021 foi assessora para avaliação institucional na Coordenadoria Geral da Unicamp. Tem experiência na pesquisa interdisciplinar, atuando principalmente nos seguintes temas: avaliação de resultado e impactos, educação superior, Ciência Tecnologia e Inovação, equidade, diversidade e inclusão (EDI) na carreira acadêmica e diáspora científica

Daniela Maciel Pinto

Mestre

Universidade Estadual de Campinas, Departamento de Política Científica e Tecnológica, Campinas, São Paulo, Brasil

Embrapa Territorial, Transferência de Tecnologias, Campinas, Brasil

daniela.maciel@embrapa.br

<https://orcid.org/0000-0003-0539-622X>

É analista de dados com especialização em Banco de Dados (Ulbra) e MBA em Data Science and Analytics pela Universidade de São Paulo (USP). Possui um mestrado também pela USP, focado em Ciência da Informação, e é formada em Biblioteconomia e Documentação pela ECA/USP. Atualmente, é doutoranda do Programa de Política Científica e Tecnológica da Universidade Estadual de Campinas (Unicamp - 2026). Além disso, trabalha na Embrapa Territorial, onde está envolvida em atividades relacionadas à transferência de tecnologia e inovação. Foi coordenadora de Infraestrutura de Dados Espaciais da Embrapa (GeoInfo) e supervisora do Setor de Prospecção e Avaliação de Tecnologia (SPAT) na área de Transferência de Tecnologia da Embrapa Territorial. Liderou o desenvolvimento da metodologia de avaliação das Tecnologias de Informação e Comunicação (TICs), Ambitec-TICs, que identifica o impacto (ex post) das soluções tecnológicas em base territorial. Também idealizou e liderou o desenvolvimento do aplicativo “TecAmazônia”, financiado pelo Fundo Amazônia, voltado para a transferência de soluções tecnológicas para o Bioma Amazônia.

João Gabriel Pedreira de Moura Gomes

Bacharel

Universidade Estadual de Campinas, Faculdade de Engenharia Civil, Arquitetura, Campinas, São Paulo, Brasil

j241749@dac.unicamp.br

<https://orcid.org/0009-0005-1887-9281>

Graduando em Arquitetura e Urbanismo pela Faculdade de Engenharia Civil, Arquitetura e Urbanismo da Universidade Estadual de Campinas (FECFAU/UNICAMP). Membro do Laboratório de Estudos sobre Organização da Pesquisa e da Inovação (Lab-GEOPI) do Instituto de Geociências da UNICAMP.

Originalidade

Declaramos que o texto é original e não está sendo revisado por nenhuma outra publicação.

Preprint

O manuscrito não foi submetido a nenhuma plataforma preprint.

Informações sobre o trabalho

O manuscrito é resultado do projeto de pesquisa denominado “Pesquisa da pesquisa e da inovação: indicadores, métodos e evidências de impactos”, financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Processo n.: 21/15091-8), coordenado pelo prof. Sergio Luiz Monteiro Salles Filho, no âmbito do Laboratório de Estudos sobre Organização da Pesquisa e da Inovação (LabGeopi) da Universidade Estadual de Campinas (Unicamp).

Agradecimentos

Os autores agradecem a equipe do projeto “Pesquisa da pesquisa e da inovação: indicadores, métodos e evidências de impactos”, que apoiaram a realização deste estudo.

Contribuição dos autores

Concepção e elaboração do manuscrito: AM Carneiro, DM Pinto, JGPM. Gomes

Coleta de dados: AM Carneiro, JGPM. Gomes

Análise de dados: DM Pinto

Discussão dos resultados: AM Carneiro

Redação – rascunho original: AM Carneiro, DM Pinto, JGPM. Gomes

Visualização: DM Pinto

Revisão e aprovação: AM Carneiro

Supervisão: AM Carneiro

Uso de inteligência artificial

Os autores informam que utilizaram a API da OpenAI, especificamente o modelo GPT- 4, para o processamento e tratamento automático de normalização dos nomes e códigos das IES, conforme descrito na seção de Metodologia do estudo.

Financiamento

O estudo foi financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) - Processo n.: 21/15091-8, a partir do projeto “Pesquisa da pesquisa e da inovação: indicadores, métodos e evidências de impactos”.

Permissão para usar imagens

Não aplicável.

Aprovação do Comitê de Ética em Pesquisa

Não aplicável.

Conflito de interesses

Os autores declaram que não possuem qualquer conflito de interesse relacionado à publicação deste artigo.

Declaração de disponibilidade de dados

Este estudo reutiliza dados provenientes de três fontes públicas. A base ORCID (2025), Discover research from ORCID, foi utilizada para extrair registros de docentes e informações institucionais. O conjunto ORCID Statistics (2025) forneceu indicadores complementares sobre distribuição e características dos registros. O dataset de Costa (2015), Lista de Toda as Universidades do Brasil, derivado do portal de indicadores do INEP/ENADE, foi empregado nos processos de normalização e identificação institucional. O reúso respeita as condições de acesso e as licenças definidas por cada fonte.

Costa, Alexandre (2015). Lista de Todas as Universidades do Brasil (retirado do portal de indicadores do INEP/ENADE). [Dataset] <https://gist.github.com/alexandremcosta/c9361cc23722a5aa1133>

ORCID (2025). Discover research from ORCID. [Datasets] <https://orcid.figshare.com/>

ORCID Statistics. (2025). ORCID. <https://info.orcid.org/resources/orcid-statistics/>

Licença de uso

Os autores concedem à BiblioS direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob uma Licença Creative Commons Atribuição (CC BY) 4.0 Internacional. Esta licença permite que terceiros remixem, adaptem e desenvolvam o trabalho publicado, dando os devidos créditos pela autoria e publicação inicial neste periódico. Os autores estão autorizados a firmar acordos adicionais separados para distribuição não exclusiva da versão publicada do trabalho no periódico (por exemplo, publicação em um

repositório institucional, em um site pessoal, publicação de uma tradução ou como um capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

Editor

Publicado pelo Sistema de Bibliotecas Universitárias da Universidade de Pittsburgh. Responsabilidade compartilhada com universidades parceiras. As ideias expressas neste artigo são dos autores e não representam necessariamente as opiniões dos editores ou da universidade.

Editores

João de Melo Maricato, Janicy Aparecida Pereira Rocha e Lúcia da Silveira

Histórico

Recebido: 12-09-2024 - Aprovado: 29-01-2025 - Publicado em: 08-12-2025



Os artigos neste periódico estão licenciados sob uma Licença Creative Commons Atribuição 4.0 Estados Unidos.



This journal is published by [Pitt Open Library Publishing](http://biblios.pitt.edu/).