



Algoritmo para análise de presença *online* de pesquisadores da base ORCID

Algorithm for analyzing the online presence of researchers in the ORCID database

Rafael Gutierrez Castanha¹



<http://lattes.cnpq.br/4834832439175113>



<https://ORCID.org/0000-0002-3117-1780>

Francielle Franco²



<http://lattes.cnpq.br/0132426572975221>



<https://ORCID.org/0000-0002-3686-595X>

Resumo

Este estudo teve como objetivo desenvolver um algoritmo capaz de identificar automaticamente a presença *online* de pesquisadores registrados na ORCID, analisando websites, mídias sociais e outros identificadores acadêmicos vinculados aos perfis. Para isso, foi desenvolvido um código em R que acessa a API pública da ORCID, extrai informações dos perfis cadastrados e gera um relatório com quatro tabelas (dados gerais, dados sumarizados, País x Plataforma, Coocorrência entre Plataformas) e um grafo de coocorrência entre as plataformas cadastradas. A prova de conceito foi realizada com 997 pesquisadores de universidades brasileiras, resultando na geração de quatro tabelas e uma rede coocorrência entre as plataformas. Os resultados demonstraram que 37,2% dos pesquisadores analisados possuíam informações sobre presença *online*, ou seja, este percentual de pesquisadores inseriu os links de seus perfis em outras plataformas nos campos determinados dentro da ORCID. As plataformas mais frequentes foram o *Currículo Lattes*, *Linkedin* e *Google Scholar*. A análise evidencia o potencial da ORCID como agregador de presença *online*, mas ressalta a importância da atualização dos perfis pelos pesquisadores. Como desenvolvimento futuro, propõe-se a criação de uma interface gráfica para facilitar o uso do algoritmo.

Palavras-chave: dados científicos; mídias sociais; identificadores únicos; ORCID; presença *online*.

Abstract

This study aimed to develop an algorithm capable of automatically identifying the online presence of researchers registered on ORCID, by analyzing websites, social media platforms, and other academic identifiers linked to their profiles. To achieve this, an R script was developed to access ORCID's public API, extract information from registered profiles, and generate a report with four tables (general data, summarized data, Country x Platform, and Platform Co-occurrence) as well as a co-occurrence network graph of the registered platforms. The proof of concept was carried out with 997 researchers from Brazilian universities, resulting in the generation of the four tables and a co-occurrence network among platforms. The results showed that 37.2% of the analyzed researchers had information regarding their online presence, meaning this percentage had inserted links to their profiles on other platforms in the designated fields within ORCID. The most frequently mentioned platforms were the Lattes Curriculum, LinkedIn, and Google Scholar. The analysis highlights the potential of ORCID as an aggregator of

¹ Doutor em Ciência da Informação, Universidade Estadual Paulista (Unesp), Marília, São Paulo, Brasil. Docente, Universidade de Marília (Unimar), Marília, São Paulo, Brasil. r.castanha@gmail.com.

² Mestre em Ciência da Informação, Universidade Federal do Rio Grande do Sul (UFRGS), Rio Grande do Sul, Brasil. Doutoranda em Ciência da Informação, Universidade Estadual Paulista (Unesp), Marília, São Paulo, Brasil. francielle.franco@unesp.br.

online presence, while also emphasizing the importance of researchers keeping their profiles up to date. As a future development, the creation of a graphical user interface is proposed to facilitate the use of the algorithm.

Keywords: scientific data; social media; unique identifiers; ORCID; *online* presence.

1 INTRODUÇÃO

A presença *online* de pesquisadores refere-se à forma como os pesquisadores são percebidos no espaço digital, podendo influenciar significativamente sua reputação profissional e impacto acadêmico (Araújo; Pedri, 2023). Essa presença é crucial pois permite que os pesquisadores compartilhem seus trabalhos, se conectem com outras pessoas e participem de conversas científicas mais amplas: A participação ativa em discussões *online* e o compartilhamento de pesquisas podem levar a um maior reconhecimento e oportunidades de colaboração (Bik; Goldstein, 2013). Uma das formas de medir a eficácia dessa presença é por meio da autoridade social, que reflete sua influência e alcance em plataformas como o X por exemplo (Araújo; Pedri, 2023). Portanto, uma presença on-line robusta é essencial para que os pesquisadores estabeleçam credibilidade e promovam a colaboração em suas áreas (Bik; Goldstein, 2013; Araújo; Pedri, 2023).

A importância de pesquisadores terem perfis em plataformas da web é cada vez mais reconhecida na comunidade acadêmica, pois podem aumentar a visibilidade, o engajamento e a reputação profissional. Ter perfis ativos e atualizados em plataformas como o *ResearchGate* ou *Google Scholar* é importante para que um pesquisador possa mostrar o seu trabalho, o que pode levar a uma maior visibilidade dentro da comunidade acadêmica (Orduña-Malea; Martín-Martín; Delgado-López-Cózar, 2016). Entretanto, pensando em termos de divulgação científica, destaca-se que a integração com as mídias sociais é fundamental porque contribui para a reputação *online* de um pesquisador e facilita o engajamento com um público mais amplo (Araújo, 2014). Portanto, manter perfis em plataformas da web é crucial para que pesquisadores aumentem sua visibilidade, promovam conexões profissionais e meçam seu impacto no cenário acadêmico.

Dentre as plataformas que disponibilizam a criação de perfis de pesquisadores destacam-se aquelas de cunho acadêmico (*ResearchGate* e *Google Scholar*), as de uso pessoal e profissional (*Facebook*, *X*, *LinkedIn*) e os perfis que disponibilizam identificadores únicos a parte do registro em base de dados (ORCID, Researcher ID e Scopus Author ID). No que tange

ao aprimoramento e estabelecimento da presença *online* de pesquisadores em distintos ambientes *online* observa-se que a plataforma ORCID (*Open Researcher and Contributor ID*) tem papel fundamental (Orduña-Malea; Martín-Martín; Delgado-López-Cózar, 2016).

Nesse sentido, a gestão de identidade acadêmica refere-se ao conjunto de práticas e estratégias adotadas por pesquisadores para organizar, manter e divulgar suas informações profissionais, produção científica e conexões institucionais em ambientes digitais. Essa gestão envolve a atualização constante de perfis em plataformas acadêmicas (como ORCID, *ResearchGate*, *Google Scholar*, entre outras), bem como o uso consciente das mídias sociais como ferramentas de visibilidade e reputação profissional (Barata *et al.*, 2018). Costas, Corona e Robinson-García (2024) destacam que a manutenção de perfis atualizados em plataformas como o ORCID tem papel central no ecossistema da pesquisa, por integrar informações dispersas e promover maior transparência e interoperabilidade na comunicação científica. Assim, a gestão da identidade acadêmica vai além da curadoria pessoal de dados, sendo uma prática essencial para o fortalecimento da presença digital dos pesquisadores e para a integridade dos sistemas de avaliação e análise da ciência.

Para além da função de garantir a correta atribuição de autoria, a ORCID ainda pode ajudar os pesquisadores a manterem uma identidade *online* consistente, vinculando seu trabalho em várias plataformas (Costas; Corona; Robinson-García, 2024). O fornecimento de um identificador único facilita o gerenciamento centralizado de perfis, integrando-se a outras plataformas, aumentando a visibilidade e incentivando atualizações regulares (Wang *et al.*, 2024). Assim, ter um perfil na ORCID não apenas ajuda os pesquisadores individuais a gerenciar suas atividades acadêmicas, mas também melhora a qualidade geral e a confiabilidade dos dados acadêmicos.

A ORCID foi projetada para atender as necessidades de pesquisadores, agências de fomento e instituições e esse aspecto comunitário solidifica ainda mais seu papel como agregador, pois conecta várias partes interessadas e facilita o compartilhamento de informações (Costas; Corona; Robinson-García, 2024). Embora a ORCID tenha o potencial de atuar como um agregador de presença *online* para pesquisadores, sua eficácia é influenciada pela integridade dos perfis e pelas motivações por trás de sua adoção. Assim, o aprimoramento da atualização dos perfis ORCID poderia solidificar ainda mais sua posição

como uma ferramenta fundamental para a manutenção da presença *online* de pesquisadores (Arroyo-Machado *et al.*, 2025; Wang *et al.*, 2024).

Os perfis na ORCID viabilizam um repositório estruturado de informações de pesquisadores, facilitando a integração de dados e ressaltando sua importância como fonte de dados para várias atividades relacionadas à pesquisa (Costas; Corona; Robinson-García, 2024). A ORCID conta com mais de 9 milhões de cadastros, sendo o Brasil o terceiro país com maior número de perfis (447.181 mil). Quanto à integridade dos dados, destaca-se que 66% dos perfis apresentam a informação de filiação, 50% apresentam informações de publicações. Quando esses dados são analisados em conjunto, identifica-se que apenas 34% dos perfis têm ambas as informações registradas. Existem diversas possibilidades de integração com a plataforma, atualmente a ORCID conta com 5941 integrações ativas (ORCID, 2025).

Arroyo-Machado *et al.* (2025) coletaram os metadados dos perfis de pesquisadores espanhóis para estudar a frequência das atualizações, bem como investigar seções de metadados preenchidas e características comuns entre os perfis e por fim aumentar a confiabilidade da ORCID para rastrear a mobilidade do pesquisador. Costas, Corona, Robinson-García (2024) também relacionaram os dados da ORCID como um facilitador na análise das relações de financiamento e da mobilidade de pesquisadores. Enquanto base de dados, a ORCID é explorada em função de seu potencial para analisar aspectos relacionados à carreira acadêmica no Brasil, incluindo a mobilidade entre instituições (Carneiro; Pinto; Gomes, 2024). Barata *et al.* (2018) discute que a integração da ORCID às práticas acadêmicas no Brasil faz parte de um movimento maior para melhorar a presença *online* dos pesquisadores, impulsionado pelas políticas institucionais e pelo cenário em evolução da comunicação científica.

Portanto, a crescente importância da presença *online* dos pesquisadores na web, evidenciada pelo impacto das mídias sociais, destaca a necessidade de mecanismos automatizados para identificar e analisar essa presença. Assim, define-se como questão norteadora: de que maneira um algoritmo pode contribuir para a extração e visualização de informações sobre a presença *online* de pesquisadores registrados na ORCID, considerando websites, mídias sociais e outros identificadores cadastrados? O objetivo deste estudo é desenvolver um algoritmo capaz de identificar automaticamente a presença *online* de

pesquisadores registrados na ORCID, analisando websites, mídias sociais e outros identificadores acadêmicos vinculados aos perfis.

2 PROCEDIMENTOS METODOLÓGICOS

A fim de construir um algoritmo capaz de identificar a presença *online* de pesquisadores registrados na ORCID, desenvolveu-se um algoritmo, em linguagem de programação R, capaz de acessar os dados cadastrais informados pelos pesquisadores em seus respectivos perfis na ORCID. O código é aberto e foi disponibilizado via repositório digital Zenodo em Castanha e Franco (2025). Para acessar estes dados, utilizou-se a API (*Application Programming Interface*) pública da base de dados da ORCID, que permite a consulta aos dados de um perfil de pesquisador que tenha tornado as suas informações públicas. As respostas das consultas via API são em formato *Extensible Markup Language* (XML).

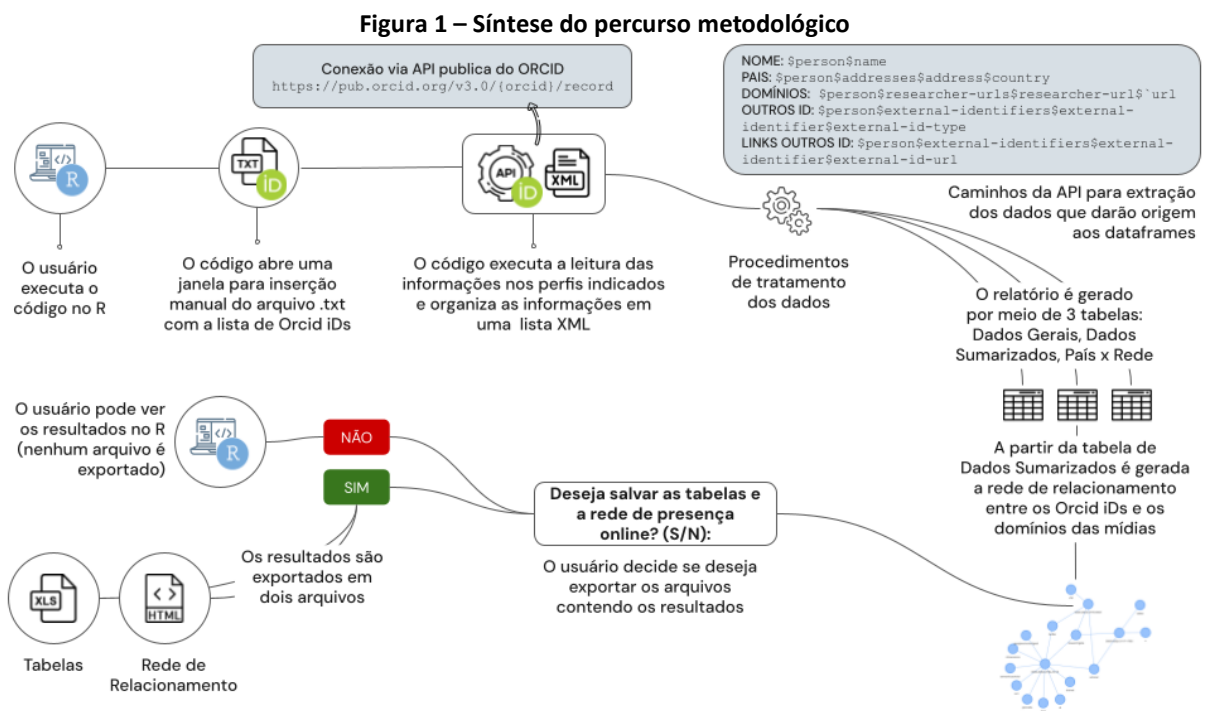
Tendo foco na presença *online* dos pesquisadores registrados na base, o algoritmo em questão deveria ser capaz de coletar os dados referentes a: nome de cada pesquisador, país de origem, nomes e *urls* dos websites e mídias sociais (*Websites e social links*), nomes e *urls* de outras plataformas fornecedoras de identificadores únicos (*Other IDs*). Para que essas informações fossem coletadas, definiu-se a necessidade de submeter uma lista de números de ORCID IDs ao algoritmo para que o código fosse capaz de identificar e extrair todas as informações requeridas. Assim, o usuário necessita enviar ao algoritmo uma lista contendo uma única coluna com os ORCID IDs de interesse em seu formato alfanumérico padrão, contendo 16 caracteres (0000-0000-0000-000X), sem a parte da *url* que precede a parte numérica (<https://ORCID.org/>). Portanto, é pré-requisito para uso deste código que o usuário possua uma lista de ORCID IDs inseridos em um arquivo .txt, organizados conforme descrito.

Tendo a lista de ORCID IDs, identificou-se na API os campos e os caminhos das informações (supracitadas) a serem extraídas. Estas informações encontram-se nos seguintes caminhos: ORCID IDs:

- a) ORCID IDs: `$`ORCID-identifier`$path`
- b) Primeiro Nome: `$person$name$`given-names`;`
- c) Sobrenome: `$person$name$`family-name``
- d) País de origem: `$person$addresses$address$country`

- e) Websites & social links: `$person$`researcher-urls`$`researcher-url`$`url-name``
- f) Links dos websites e mídias sociais: `$person$`researcher-urls`$`researcher-url`$`url``
- g) Outros Identificadores: `$person$`external-identifiers`$`external-identifier`$`external-id-type``
- h) Links dos outros Identificadores: `$person$`external-identifiers`$`external-identifier`$`external-id-url``

Na etapa seguinte realizou-se a extração de todas as informações. Para fins de organização foi gerado um relatório contendo todos os dados obtidos a partir de cada perfil da ORCID. Este relatório é composto por quatro tabelas e uma rede (grafo) que relaciona as plataformas cadastradas. Todo processo é ilustrado na Figura 1.



Fonte: Elaboração própria.

A primeira tabela refere-se aos dados gerais de cada pesquisador contendo oito colunas: Número de ORCID (ORCID IDs), Nome, país de origem, websites e mídias sociais (*Websites and social links*), *urls* dos respectivos websites e mídias sociais, Domínio (Plataforma), outros identificadores (*Other IDs*) e as respectivas *urls* dos outros identificadores. O item descrito como “Domínio (Plataforma)” foi gerado a partir das *urls* de

cada website e mídia social referindo-se aos seus respectivos domínios, por exemplo, em casos que o pesquisador registrou *Google Scholar Profile*, *Google Scholar*, *Google Acadêmico*, ou similares, estes rótulos foram armazenados na coluna “Web sites e social links”. A partir das *url* destes rótulos (presente na quarta coluna), extraiu-se somente o domínio de cada *url*. Por exemplo, todos rótulos exemplificados possuem o mesmo domínio que é *https://scholar.google.com/{...}*. Com isso, o código exclui todos os elementos antes e depois de “scholar”, sendo esta a informação registrada na coluna “Domínio (Plataforma)”. Este procedimento é realizado a partir da criação de um Regex. Esta operação foi primordial para padronização das mídias sociais de cada pesquisador, visto que, diversos rótulos de mídias sociais podem corresponder a um mesmo domínio.

A segunda tabela gerada pelo algoritmo é uma tabela sumarizada contendo somente cinco colunas: Nome, número de ORCID (ORCID IDs), país de origem, domínio e Outros ID. Esta tabela possui linhas com valores únicos, aglutinando nas colunas “Domínio (Plataforma)” e “Outros ID”, em uma única célula, estas informações utilizando o separador ponto e vírgula (;). Esta tabela resume o panorama da presença *online* do conjunto de pesquisadores a serem analisados.

A terceira aba apresenta uma tabela conjugada de frequência, refere-se ao cruzamento de dados entre cada domínio do website ou mídia social e o país de origem dos pesquisadores. Por fim a tabela 4 apresenta os dados de coocorrência entre Plataformas, ou seja, com qual frequência as plataformas (expressas pelos domínios) são mencionadas simultaneamente em cada cadastro. A partir da geração destas quatro tabelas, o algoritmo fornece a opção de salvá-las em arquivo externo de extensão *.xlsx (Excel)* gerando quatro abas distintas: Aba 1: Dados Gerais; Aba 2: Dados Sumarizados; Aba 3: País x Plataforma; Aba 4: Coocorrência entre Plataformas.

O algoritmo ainda gerou uma rede (grafo) de relacionamento entre as plataformas mais mencionadas de forma simultânea em cada cadastro. Para isso, utilizou-se a biblioteca *visnetwork3D*. A rede gerada fornece uma visão geral da presença *online* de pesquisadores em websites e mídias sociais, de modo que cada aresta da rede denota um relacionamento de presença *online* de um pesquisador específico em diferentes plataformas. A opção pelo uso da rede de coocorrência se dá pela possibilidade de visualizar como websites e mídias sociais aparecem conjuntamente nos perfis de pesquisadores registrados na ORCID. A partir de um

conjunto de ORCID IDs analisados, essa abordagem permite identificar padrões de conexão, similaridades na presença *online* e a distribuição das atividades dos pesquisadores em diferentes plataformas.

Cabe ressaltar que tanto o processo de inserção, quanto a exportação dos dados ocorre a partir da interação do usuário com o código. Quando o código é executado, o console no R apresenta a frase “Selecione seu conjunto de ORCID IDs”, abre-se uma janela para a inserção manual de arquivo .txt. Após essa ação, o console mostra a mensagem “Aguarde...”. Passado o processamento, o console exibe a mensagem: “Deseja salvar as tabelas e a rede de presença *online*? (S/N):”, caso o usuário digite “S” ele confirma a exportação do relatório (tabelas) e a rede de presença *online* no formato .html. Desse modo os arquivos são exportados e salvos no diretório padrão do usuário.

A fim de testar o algoritmo proposto, realizou-se uma prova de conceito para demonstração das funcionalidades do programa. Realizou-se a coleta de um conjunto de 1000 ORCID IDs composto pelos 200 primeiros resultados de busca no site da ORCID (ORCID.org) das cinco universidades brasileiras mais bem ranqueadas no *Scimago Institution Ranking 2025*: Universidade de São Paulo (USP), Universidade de Campinas (UNICAMP), Universidade Estadual Paulista (UNESP), Universidade Federal de Minas Gerais (UFMG) e Universidade Federal do Rio de Janeiro (UFRJ). Buscou-se pela sigla de cada uma das universidades para gerar o conjunto com os 200 primeiros resultados de cada busca, gerando uma lista de 1000 ORCID IDs. Este conjunto de dados, bem como os resultados gerados pelo algoritmo também estão no repositório digital Zenodo em Castanha e Franco (2025).

3 RESULTADOS

A partir da prova de conceito, foi possível testar as funcionalidades propostas pelo algoritmo. Sendo assim, foram coletadas todas as informações de 997 pesquisadores únicos (3 registros estavam duplicados) vinculados às referidas universidades. O algoritmo coletou as informações gerais de cada pesquisador, bem como, foi capaz de identificar os perfis de pesquisadores que não cadastraram qualquer informação nos campos consultados. Desse modo, pode-se analisar a presença *online* de 371 pesquisadores, isso porque 626 foram identificados como “Sem presença *online*”, ou seja, não apresentavam qualquer informação

nos campos correspondentes aos websites, mídias sociais ou outros identificadores únicos. Esta ausência pode indicar limitações no preenchimento dos perfis pelos próprios usuários. Tal constatação reforça as discussões sobre a importância da motivação pessoal dos pesquisadores para manterem seus perfis atualizados visando maior eficácia da ORCID como agregador de presença *online* (Arroyo-Machado *et al.*, 2025; Sixto-Costoya *et al.*, 2021).

Ao observar as tabelas geradas pelo algoritmo, verificou-se que o campo “Nome” apresentou algumas inconsistências sendo elas: um pesquisador que não preencheu o nome, oito perfis que fizeram uso dos prefixos “Professor” ou “Doutor”, e ainda oito perfis que não correspondiam à pesquisadores individuais mas sim à autorias institucionais, como revistas, fóruns ou portais de periódicos, como por exemplo: Fórum de Pesquisa Discente FORPED/PPGGOC/UFMG (0000-0003-3627-6482), UFMG Revista Docência do Ensino Superior (0000-0002-2372-1914), Portal de Periódicos UFMG (0000-0003-4190-1430). Sixto-Costoya *et al.* (2021) evidenciaram que a falta de uniformidade nos dados do ORCID dificulta a extração e interpretação automática das informações, além de representar um obstáculo para estudos que dependem de dados sobre a atividade científica e a identidade dos pesquisadores. Portanto, a falta de padronização no preenchimento dos dados de identificação dos pesquisadores pode ter impacto nas análises subsequentes.

Referente à distribuição geográfica dos pesquisadores observou-se que, para além do Brasil (879 pesquisadores únicos), outros países foram identificados em menores proporções como Ucrânia (15 pesquisadores), Estados Unidos (12 pesquisadores) e Espanha (10 pesquisadores). O cruzamento das informações de País X Plataforma (na tabela de dados sumarizados) permitiu a análise da frequência das plataformas em cada país. Destaca-se que no Brasil foram identificadas 93 plataformas diferentes (perfis em mídias sociais, sites institucionais, plataformas fornecedoras de IDs, páginas pessoais, etc.), sendo o *Currículo Lattes* (151 ocorrências), o *LinkedIn* (32 ocorrências), o *Instagram* (12 ocorrências) e o *ResearchGate* (9 ocorrências) as plataformas mais mencionadas. Portanto, o algoritmo proposto pode apresentar dados de presença *online* que estimulem as discussões acerca da presença *online* de pesquisadores de acordo com as suas particularidades geográficas, evidenciado pela tabela “País x Plataforma”.

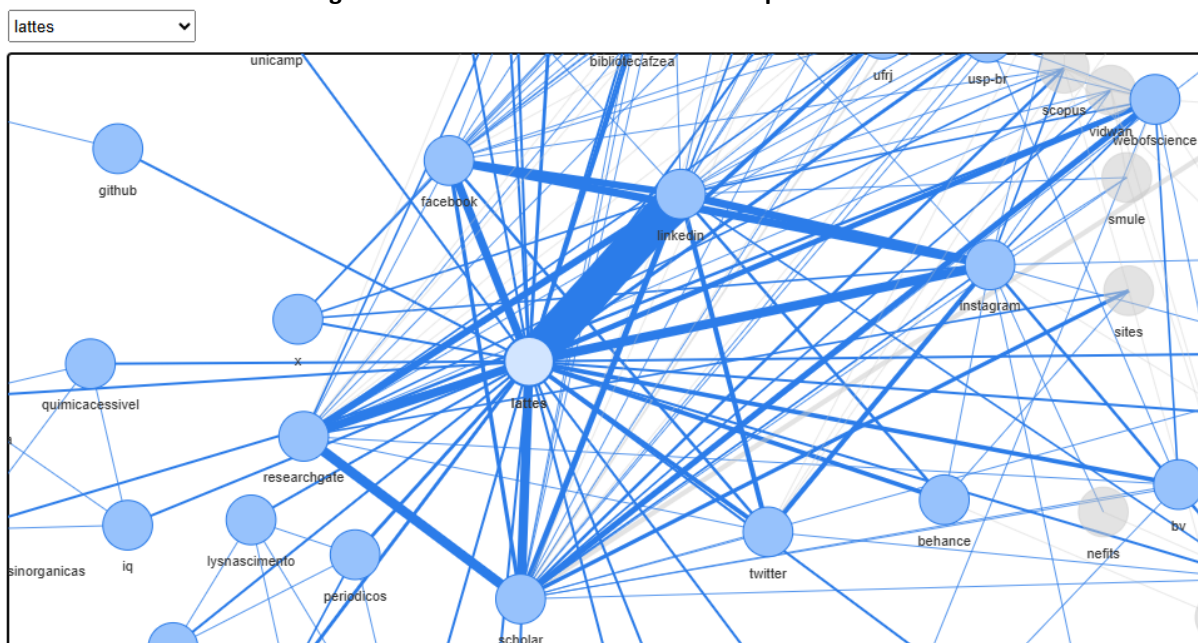
Países como Espanha e Estados Unidos apresentam resultados que indicam uma preferência equilibrada entre mídias sociais acadêmicas (*ResearchGate*, *Google Scholar*) e

mídias sociais gerais (*Instagram, X*). Barata *et al.* (2018) menciona que no Brasil os pesquisadores utilizam predominantemente o *Currículo Lattes* como plataforma de identidade acadêmica, ao passo que o uso de redes sociais como *Facebook, X* e *ResearchGate* varia conforme as áreas e intenções de visibilidade. Dessa forma a localização influencia a escolha das plataformas, muitas vezes impulsionada por políticas institucionais e culturais específicas do país.

Referente aos dados de frequência de menção das plataformas, destaca-se que foram utilizadas as informações coletadas a partir do campo “Domínio (Plataforma)”, assim pode-se garantir a maior precisão da contagem. As plataformas mais frequentes foram o *Currículo Lattes* (93), *Linkedin* (53), *Google Scholar* (27), *ResearchGate* (24) e *Instagram* (22). Conforme outros estudos, esses resultados destacam a importância das mídias sociais acadêmicas (*ResearchGate, Google Scholar*) para aumentar a visibilidade e das mídias sociais gerais (*Linkedin, Instagram*) para ampliar o engajamento dos pesquisadores com outros públicos mais amplos (Araújo, 2014; Orduña-Malea, Martín-Martín e Delgado-López-Cózar, 2016).

O campo “*Other IDs*”, onde são inseridas as informações de perfis em outras plataformas que disponibilizam identificadores únicos, apresentou informações em apenas 91 perfis. Destacam-se as plataformas *ResearcherID* (53) e *Scopus Author ID* (38), além das plataformas *Ciência ID* (9), *Sciprofiles* (5) e *Loop profile* (4). Os identificadores *Researcher ID* e *Scopus Author ID* estão vinculados às bases de dados *Web of Science* e *Scopus* respectivamente. Portanto, a existência dessa informação depende da indexação da produção do pesquisador nas referidas bases de dados, assim, observam-se indícios da baixa representação da produção acadêmica destes pesquisadores na *Scopus* e na *Web of Science*. A presença desses identificadores adicionais reforça a ideia apresentada por Costas, Corona e Robinson-García (2024) sobre o papel agregador da ORCID ao vincular diferentes perfis acadêmicos em um único ambiente. Desse modo, a funcionalidade de apresentar tais informações também se mostra igualmente útil para análises de presença *online*. Por fim, apresenta-se a rede de coocorrência das plataformas cadastradas pelos pesquisadores. A Figura 2 apresenta a rede onde o nó do *Currículo Lattes* (*lattes*) está selecionado.

Figura 2 – Rede de coocorrência entre as plataformas.



Fonte: Elaboração própria.

Analisando a rede acima observa-se que muitos pesquisadores deste universo utilizam principalmente o *Currículo Lattes* como plataforma de presença *online*, seguido por outras plataformas como *ResearchGate*, *Google Scholar* e *Linkedin*. As combinações mais frequentes incluem Lattes com *Instagram* e *Linkedin*; *ResearchGate* com *Google Scholar*; *Facebook* com *Instagram*. A rede de coocorrência confirma a relevância das plataformas destacadas anteriormente e apresenta a possibilidade de analisar visualmente a presença *online* dos pesquisadores.

A partir dessa análise visual observam-se evidências de que o uso simultâneo dessas plataformas pode indicar uma tentativa de maximizar o impacto acadêmico e social. A diversificação da presença em diferentes plataformas é um componente essencial da chamada *Ciência 2.0*, onde visibilidade, acessibilidade e engajamento com públicos diversos se tornam objetivos relevantes (Araújo, 2014). O uso combinado de plataformas acadêmicas (como *ORCID*, *ResearchGate* e *Google Scholar*) e redes sociais gerais (como *X*, *Facebook*, *Linkedin*) não é aleatório, mas sim uma estratégia deliberada de muitos pesquisadores para aumentar sua visibilidade, alcançar diferentes públicos e potencializar seu impacto, tanto na esfera acadêmica quanto na esfera pública e social (Araújo, 2014; Costas; Corona; Robinson-García, 2024).

A análise do grafo pode ser combinada com as informações apresentadas na coluna “Domínio (Plataforma)” na tabela de dados sumarizados para obter uma imagem mais ampla da coocorrência das menções. Observou-se que *Linkedin* e Lattes foram mencionados simultaneamente 36 vezes, *Instagram* e *Linkedin* 6 vezes e *Google Scholar* e *Web of Science* 5 vezes. Esses resultados sugerem que existe uma forte tendência dos pesquisadores em utilizar conjuntamente uma plataforma acadêmica (Lattes) e uma mídia social profissional (*Linkedin*). Barata *et al.* (2018) corrobora ao mencionar que as redes sociais profissionais são empregadas para fins distintos das redes acadêmicas, e, nesse contexto, o *Linkedin* aparece como uma ferramenta relevante para a construção da reputação profissional, colaboração e divulgação de conquistas e atividades acadêmicas em ambientes mais amplos.

4 CONSIDERAÇÕES FINAIS

A partir da prova de conceito pode-se demonstrar que o algoritmo desenvolvido é eficaz na identificação automática da presença *online* de pesquisadores registrados na ORCID. A análise dos dados permitiu a caracterização da presença digital dos pesquisadores, identificando padrões geográficos, preferências por plataformas e a integração entre diferentes identificadores acadêmicos. Dessa forma, o estudo responde diretamente para a questão norteadora de como identificar automaticamente a presença *online* de pesquisadores em diferentes ambientes web.

Os resultados reforçam a importância da gestão da presença *online* dos pesquisadores. O estudo evidenciou que a presença *online* dos pesquisadores varia de acordo com o país, sendo o *Currículo Lattes* predominante no universo analisado, enquanto *ResearchGate* e *Google Scholar* têm maior relevância em outros países. A análise da rede de coocorrência evidenciou padrões de uso combinados entre mídias sociais acadêmicas e de uso pessoal ou profissional, demonstrando a relevância da diversificação da presença *online* para ampliar a visibilidade dos pesquisadores.

Entre as limitações do estudo, destaca-se a dependência da atualização das informações nos perfis da ORCID, uma vez que a metodologia utilizada se baseia nos dados fornecidos voluntariamente pelos pesquisadores. Outra possível limitação refere-se à própria definição de presença *online*, vale ressaltar que a simples menção de perfis de outras

plataformas na ORCID não garante que essa presença tenha qualidade, ou seja, são necessários outros estudos que possam verificar se os perfis indicados estão atualizados e se os dados são válidos. Assim, definiu-se que, para o objetivo proposto, a menção de uma url ou ID nos campos indicados na ORCID representa um indício de presença *online* daquele pesquisador.

Enquanto trabalhos futuros, pretende-se realizar a validação dos dados a fim de analisar a qualidade da presença *online* desses pesquisadores verificando se os perfis estão ativos e atualizados. Além disso, pretende-se desenvolver uma interface gráfica que permita o uso do algoritmo sem a necessidade de interação direta com o *script* do algoritmo, tornando a ferramenta mais acessível. Os resultados reforçam a relevância da ORCID como um agregador de perfis acadêmicos e demonstram o potencial de abordagens automatizadas para mapear a presença *online* de pesquisadores. Por fim, ao fornecer um método estruturado para coletar e analisar esses dados, este estudo contribui para a compreensão das dinâmicas de visibilidade acadêmica na web e para o aprimoramento de estratégias voltadas à presença *online* de pesquisadores.

REFERÊNCIAS

ARAÚJO, R. F. Ciência 2.0 e a presença *online* de pesquisadores: visibilidade e impacto. **Ciência da Informação em Revista**, Maceió, v. 1, n. 3, p. 32-40, 2014. DOI; <http://doi.org/10.28998/cirev.%y132-40>. Disponível em: <https://www.seer.ufal.br/index.php/cir/article/view/1608>. Acesso em: 28 fev. 2025.

ARAÚJO, R. F.; PEDRI, P. Cientistas da informação no Twitter: presença, visibilidade e autoridade social. **P2P e Inovação**, Rio de Janeiro, v. 10, n. 1, p. 371-385, 2023. DOI; <http://doi.org/10.21728/p2p.2023v10n1.p371-385>. Disponível em: <https://cip.brapci.inf.br/download/228365>. Acesso em: 9 abr. 2025.

ARROYO-MACHADO, W. *et al.* Exploring ORCID adoption and metadata presence in Spain's research landscape. **Scientometrics**, [S. l.], v. 130, n. 1, p. 123-145, 2025. DOI: <http://doi.org/10.1007/s11192-025-05248-8>. Disponível em: <https://link.springer.com/article/10.1007/s11192-025-05248-8>. Acesso em: 7 mar. 2025.

CASTANHA, R. G.; FRANCO, F. Conjunto de dados utilizados no estudo "algoritmo para análise de presença online de pesquisadores da base ORCID". **Zenodo**. Version v1. 2025. DOI: <https://doi.org/10.5281/zenodo.15001663>. Acesso em: 11 maio 2025.

BARATA, G. *et al.* O uso de mídias sociais por acadêmicos brasileiros. *In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA*, 6., 2018, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: EBBC, 2018. p. 209-217. DOI: <http://doi.org/10.22477.047989>. Disponível em: <https://ebbc.inf.br/ojs/index.php/ebbc/article/view/715>. Acesso em: 7 mar. 2025.

BIK, Holly M.; GOLDSTEIN, Miriam C. An introduction to social media for scientists. **PLOS Biology**, [S. l.], v. 11, n. 4, e1001535, 2013. Disponível em: <https://doi.org/10.1371/journal.pbio.1001535>. Acesso em: 9 abr. 2025.

CARNEIRO, A. M.; PINTO, D. M.; GOMES, J. G. P. M. ORCID como fonte de dados para análise da carreira acadêmica no Brasil: oportunidades e desafios. *In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA*, 9., 2024, Brasília, DF. **Anais [...]**. Brasília, DF: EBBC, 2024. Disponível em: <https://doi.org/10.22477/ix.ebbc.360>. Acesso em: 6 mar. 2025.

COSTAS, R.; CORONA, C.; ROBINSON-GARCÍA, N. Could ORCID play a key role in meta-research? discussing new analytical possibilities to study the dynamics of science and scientists. *In: OANCEA, A.; DERRICK, G. E.; NUSEIBEH, N.; XU, X. (org.). Handbook of meta-research*. Cheltenham: Edward Elgar Publishing, 2024. cap. 18, p.215-232.

ORCID. **ORCID Statistics**. 2025. Disponível em: <https://info.ORCID.org/pt/resources/ORCID-statistics/>. Acesso em: 6 mar. 2025.

ORDUÑA-MALEA, E.; MARTÍN-MARTÍN, A.; DELGADO-LÓPEZ-CÓZAR, E. The next bibliometrics: ALMetrics (author level metrics) and the multiple faces of author impact. **Profesional de la información**, [S. l.], v. 25, n. 3, p. 485-496, 2016. Disponível em: <https://doi.org/10.3145/epi.2016.may.18>. Acesso em: 28 fev. 2025.

SIXTO-COSTOYA, A. *et al.* Exploring the relevance of ORCID as a source of study of data sharing activities at the individual-level: a methodological discussion. **Scientometrics**, [S. l.], v. 126, n. 8, p. 7149-7165, 2021. Disponível em: <https://arxiv.org/abs/2105.11825>. Acesso em: 8 mar. 2025.

WANG, T. *et al.* Is ORCID a reliable source for CV analysis? exploring the data availability of ORCID academic profiles. **Scientometrics**, [S. l.], v. 129, n. 3, p. 1637-1662, 2024. Disponível em: <https://doi.org/10.1007/s11192-024-04944-1>. Acesso em: 28 fev. 2025.



4.0 Internacional

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

Como citar este trabalho:

CASTANHA, Rafael Gutierrez; FRANCO, Francielle. Algoritmo para análise de presença *online* de pesquisadores da base ORCID. *In: WORKSHOP DE INFORMAÇÃO DADOS E TECNOLOGIA*, 8., 2025, Marília, SP. **Anais [...]**. Marília, SP: Universidade de Marília, 2025. DOI: <https://doi.org/10.22477/viii.widat.265>.