



10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



DIAGNÓSTICO DA ENDOGAMIA ACADÊMICA DOS PESQUISADORES DA UNIVERSIDADE FEDERAL DE SÃO CARLOS – CAMPUS SÃO CARLOS

Vinícius Rafael Micali Soares

<https://orcid.org/0000-0002-0421-866X>
viniciusmicali@ufscar.br
Universidade Federal de São Carlos
(UFSCar)

Leandro Innocentini Lopes de Faria

<https://orcid.org/0000-0002-8369-1315>
leandro@ufscar.br
Universidade Federal de São Carlos
(UFSCar)

Dênis Leonardo Zaniro

<https://orcid.org/0000-0003-2638-9264>
deniszaniro@estudante.ufscar.br
Universidade Federal de São Carlos
(UFSCar)

Resumo:

Este estudo analisa a endogamia acadêmica na Universidade Federal de São Carlos – Campus São Carlos por meio de uma arquitetura computacional multi-linguagem. O fenômeno é investigado diante da necessidade de compreender características de colaboração científica em bases densas. A amostra abrange 86.341 vínculos de coautoria em 24.612 artigos, envolvendo 664 pesquisadores institucionais e 7.322 colaboradores externos. A metodologia fundamenta-se na extração automatizada de dados da Plataforma Lattes e no subsequente tratamento de metadados para análise georreferenciada. Os resultados revelam a inexistência de correlação linear entre endogamia e distância geográfica, validando a eficácia da abordagem proposta para a análise de métricas bibliométricas complexas.

Palavras-Chave: Coautoria científica. Endogamia acadêmica. Visualização da Informação. Geolocalização. Plataforma Lattes.

EIXO TEMÁTICO:

Produtividade e Colaboração Científica

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.1008

COMO CITAR ESTE ARTIGO:

SOARES, Vinícius Rafael Micali; FARIA, Leandro Innocentini Lopes de; ZANIRO, Dênis Leonardo. Diagnóstico da endogamia acadêmica dos pesquisadores da Universidade Federal de São Carlos – campus São Carlos. In: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. **Anais [...]**. Curitiba, 2026. DOI: 10.22477/x.ebbc.1008. Disponível em: <https://doi.org/10.22477/x.ebbc.1008>.



1 INTRODUÇÃO

Uma investigação crítica sobre os efeitos da endogenia acadêmica é fundamental para que formuladores de políticas compreendam seus impactos — tanto positivos quanto negativos — no sistema científico brasileiro (Borenstein; Perlin; Imasato, 2022). Tal debate torna-se urgente diante da crescente pressão social por maior transparência e eficiência acadêmica, demandando que o fenômeno seja explorado com maior profundidade.

Tradicionalmente, a análise da produção científica recorre a indexadores internacionais como Web of Science (WoS) e Scopus, ou iniciativas como ORCID e Google Scholar. Contudo, tais ferramentas impõem desafios à operacionalização de análises relacionais complexas, uma vez que não disponibilizam registros nativamente estruturados sobre vínculos de orientação *stricto sensu* e geolocalização institucional. Sob outra perspectiva, a Plataforma Lattes (PL) constitui um repositório nacional consolidado, que integra registros detalhados sobre trajetórias profissionais, formação acadêmica e produção científica.

Nesse contexto, o problema central que motiva este trabalho não reside na escassez de informações na PL, mas na complexidade de operacionalizar analiticamente os vínculos de orientação e coautoria presentes na plataforma. Embora a base seja exaustiva, há uma lacuna metodológica na conversão desses registros brutos em indicadores de distribuição geográfica e, especificamente, em métricas de endogamia acadêmica. O desafio, portanto, é superar a opacidade das grandes bases de dados para identificar características de colaboração que moldam a produção científica brasileira.

O conceito de endogenia acadêmica, derivado da biologia e definido por Altbach,

Yudkevich e Rumbley (2015) como a contratação de docentes pela mesma instituição em que se formaram, possui variações terminológicas importantes. No Brasil, o termo “endogenia” é predominante, e em Portugal utiliza-se “endogamia”, correspondente ao inglês “*academic inbreeding*”.

Considerando essas nuances, este trabalho adota o termo endogamia. Operacionalmente, define-se aqui a relação endogâmica como uma coautoria em artigo científico que satisfaça, simultaneamente, três condições: 1) a presença do orientador como autor; 2) a presença do orientado como coautor; e 3) a publicação ter ocorrido após a conclusão do primeiro doutorado do orientado.

A relevância do estudo parte da premissa de que a endogamia acadêmica pode acarretar riscos à produção do conhecimento, como a redução da diversidade de pensamento, a manutenção de determinados vieses investigativos e o declínio do potencial de inovação.

Partindo da hipótese de que métodos de visualização de informação podem tornar explícitas as relações de colaboração que estão ocultas na estrutura tabular da PL, o objetivo deste estudo foi desenvolver uma abordagem automatizada de extração e tratamento de dados que, por meio de visualização, permite identificar padrões de endogamia e de dispersão geográfica. O objeto de pesquisa é formado pelos pesquisadores da Universidade Federal de São Carlos (UFSCar) – Campus São Carlos, credenciados em Programas de Pós-Graduação (PPGs) da universidade.

2 METODOLOGIA

Esta pesquisa é experimental e exploratória, com uma abordagem quantitativa (Gil, 2017), pautada na mensuração de indicadores de endogamia para a compre-

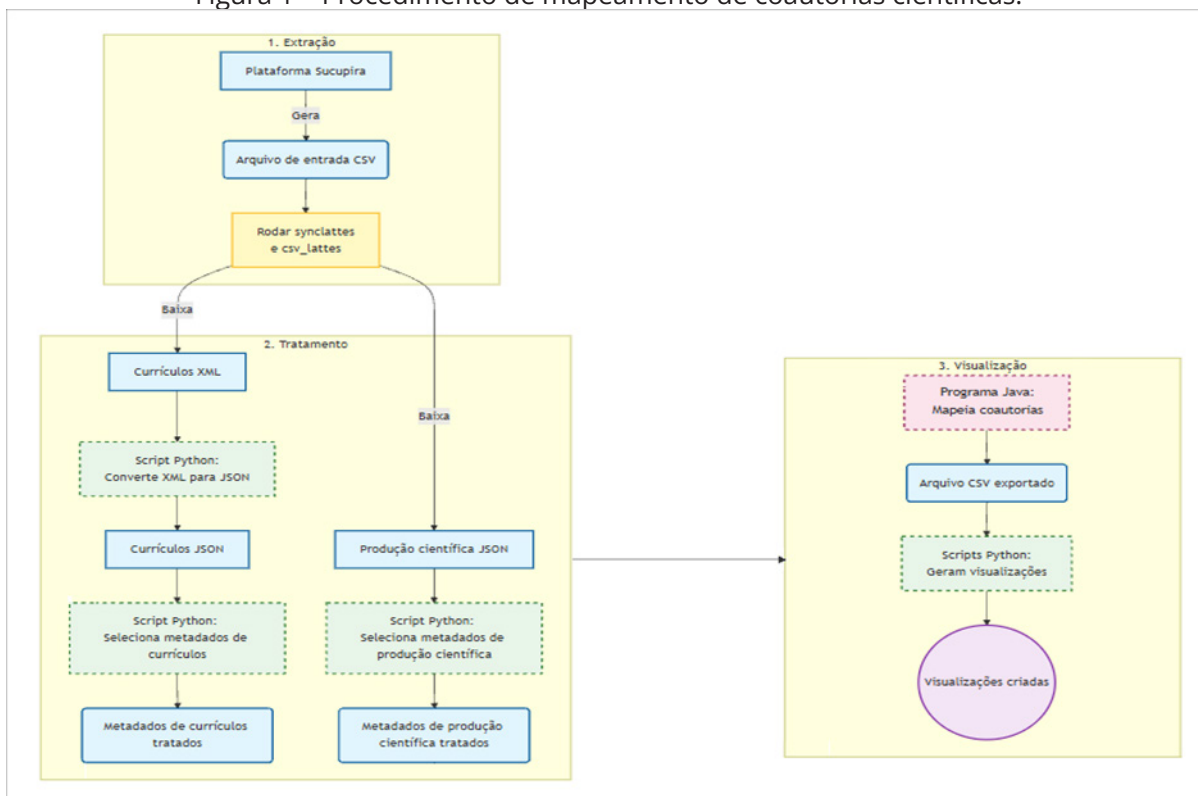
ensão do fenômeno no cenário analisado.

Alinhada aos princípios da Ciência Aberta, a *Application Programming Interface* (API) livre Nominatim (OpenStreetMap)¹ foi adotada para a geocodificação dos dados, convertendo endereços institucionais em coordenadas geográficas (Soares, 2026, Apêndice G). Foi implementado um *script* para contagem de ocorrências de latitude e longitude (Soares, 2026, Apêndice I), onde, para cada par origem-destino — variáveis “latitude_origem”, “longitude_origem”, “latitude_destino” e “longitude_destino” —, é calculado o número de ocorrências, representando a frequência com que aquela relação geográfica se manifesta. A fórmula de Haversine foi empregada para o cálculo de distâncias entre coordenadas (Soares, 2026, Apêndice F).

Na Figura 1, é apresentado o procedimento computacional desenvolvido neste trabalho.

¹ Nominatim. Disponível em: <https://nominatim.org>. Acesso em: 16 fev. 2026.

Figura 1 – Procedimento de mapeamento de coautorias científicas.



Fonte: Soares, 2026.

Conforme esquematizado na Figura 1, a arquitetura metodológica deste trabalho estrutura-se em três etapas sequenciais e interdependentes: Extração, Tratamento e Visualização.

A etapa de Extração inicia-se com a obtenção de dados primários da Plataforma Sucupira, gerando um arquivo de entrada no formato *Comma-Separated Values* (CSV) que lista os IDs Lattes dos pesquisadores². Este arquivo atua como insumo para as ferramentas synclattes (Matias, 2015) e csv_lattes (Matias; Amaral; Matias, 2017), que executam a coleta massiva dos registros de produção científica em *JavaScript Object Notation* (JSON)³ e dos Currículos Lattes (CVs Lattes) em *Extensible Markup Language* (XML)⁴.

² IDs Lattes dos pesquisadores credenciados em PPGs da UFSCar. Disponível em: <https://zenodo.org/records/18371308>. Acesso em: 28 abr. 2026.

³ Produção científica dos pesquisadores credenciados em PPGs da UFSCar. Disponível em: <https://zenodo.org/records/18371506>. Acesso em: 28 abr. 2026.

⁴ CVs Lattes dos pesquisadores credenciados em PPGs da UFSCar. Disponível em: <https://zenodo.org/records/18371406>. Acesso em: 28 abr. 2026.

Na etapa intermediária de Tratamento, emprega-se um *script* para converter os currículos do formato XML para JSON (Soares, 2026, Apêndice A), assegurando a interoperabilidade e padronização dos dados. Subsequentemente, rotinas adicionais são executadas, para filtrar e selecionar os metadados pertinentes tanto da produção científica (Soares, 2026, Apêndice C) quanto dos CVs Lattes (Soares, 2026, Apêndice E), preparando o conjunto de dados para a análise relacional.

A etapa final, designada Visualização, utiliza programas para processar os dados tratados (Soares, 2026, Apêndice D) e mapear as relações de coautoria (Soares, 2026, Apêndice H). Para garantir a pertinência do escopo, o algoritmo aplica filtros institucionais rigorosos: considera-se como ponto de origem apenas os pesquisadores que possuam vínculo formal com a UFSCar (código institucional 033500000006) e cujo endereço profissional esteja declarado na cidade de São Carlos/SP. Para cada pesquisador elegível, são examinadas as publicações indexadas, verificando a existência de relações formais de orientação em nível de doutorado entre os coautores. Em consonância com a definição de endogamia previamente adotada, são consideradas endogâmicas apenas as coautorias ocorridas em anos posteriores à titulação do orientado, assegurando que a análise reflita parcerias científicas consolidadas após a formação acadêmica. Adicionalmente, é implementada uma estratégia iterativa de enriquecimento de dados: durante a varredura das publicações, coautores cujos currículos não constam no conjunto são registrados em um arquivo CSV auxiliar para uma nova extração na PL, possibilitando expandir progressivamente a cobertura da rede. Após, é realizada a mesclagem (Soares, 2026, Apêndice B) dos CVs Lattes dos pesquisadores da UFSCar com os CVs Lattes dos seus coautores⁵. Por fim, o fluxo encerra-se com a geração de dois artefatos principais⁶: um arquivo estruturado em CSV, contendo as

⁵ Dados tratados: CVs Lattes dos pesquisadores credenciados em PPGs da UFSCar e dos coautores dos pesquisadores da UFSCar – Campus São Carlos credenciados em PPGs da universidade. Disponível em: <https://zenodo.org/records/18446038>. Acesso em: 28 abr. 2026.

⁶ Arquivos de saída da execução do programa de mapeamento de coautorias científicas dos pesquisadores da UFSCar – Campus São Carlos credenciados em PPGs da universidade. Disponível em: <https://zenodo.org/records/18446156>. Acesso em: 28 abr. 2026.

coordenadas geográficas necessárias para a plotagem dos mapas de colaboração, e um registro de log detalhado. Esses documentos permitem a rastreabilidade das operações, contribuindo para a transparência, auditabilidade e reprodutibilidade de todo o processo analítico.

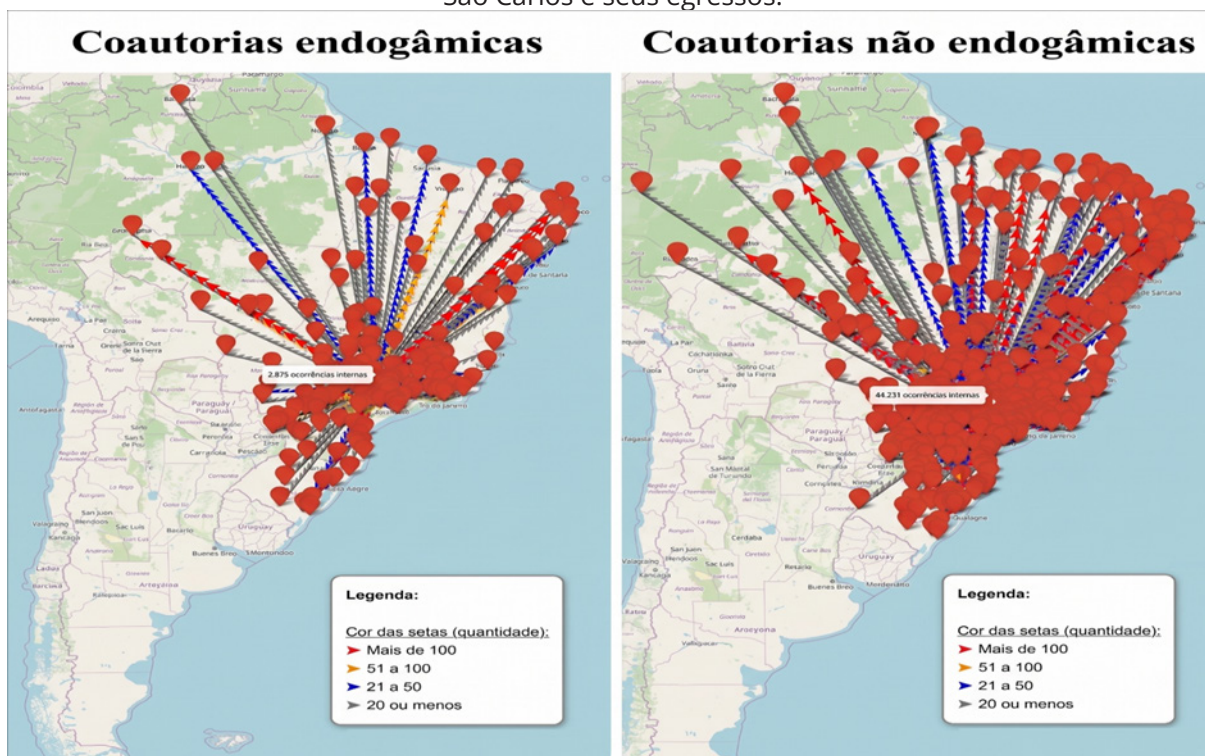
3 RESULTADOS

Foram processados 86.341 registros de coautorias em artigos completos de revistas até o ano de 2024, referentes a 24.612 produções únicas. Esses dados englobam 664 pesquisadores da UFSCar – Campus São Carlos credenciados em PPGs da universidade e 7.322 coautores com endereço profissional cadastrado.

Verificou-se que a endogamia é um fenômeno minoritário na amostra, representando 7% das coautorias e 22% das publicações. Embora reduzida nos vínculos de colaboração, sua presença triplica quando analisada a partir da produção bibliográfica final, evidenciando maior expressividade no volume de trabalhos do que no perfil das parcerias.

Para analisar o fluxo espacial de coautorias foi desenvolvido um *script* que gera um mapa interativo (Soares, 2026, Apêndice J) — referente a etapa de Visualização do procedimento de mapeamento de coautorias —, podendo ser visto na Figura 2.

Figura 2 – Distribuição espacial das coautorias entre pesquisadores de PPGs da UFSCar – Campus São Carlos e seus egressos.



Fonte: Adaptada de Soares (2026).

Como pode ser observado na Figura 2, a topologia de rede dos mapas é similar, não havendo um padrão linear evidente que correlacione a distância geográfica com o nível de endogamia. A espessura das conexões está padronizada para manter a clareza dos fluxos, reservando o detalhamento da frequência de coautorias para a interação via recurso de *tooltip*, disponível no formato HTML⁷.

4 CONCLUSÕES

Este estudo mostrou a viabilidade técnica de utilizar a PL como fonte primária para a extração, o tratamento e a visualização de dados, superando as limitações de recursos nativos dessa plataforma.

Constatou-se que a proximidade geográfica não atua como um fator determinante

⁷ Visualizações dinâmicas para analisar a endogamia das coautorias científicas dos pesquisadores da UFSCar – Campus São Carlos credenciados em PPGs da universidade. Disponível em: <https://zenodo.org/records/18446461>. Acesso em: 28 abr. 2026.

para os níveis de endogamia, evidenciando a coexistência de diferentes frequências de coautorias endogâmicas ao longo de todo o espectro geográfico analisado.

Conclui-se, que a endogamia acadêmica é um fenômeno emergente de dinâmicas relacionais e institucionais complexas, que não podem ser explicadas por variáveis isoladas. Os resultados contribuem para consolidar a visualização da informação como uma ferramenta indispensável para explicitar a estrutura oculta das redes de colaboração científica e subsidiar políticas institucionais.

Por fim, ressalta-se que o recorte metodológico adotado não exaure o fenômeno em sua totalidade. Essa delimitação, contudo, sinaliza horizontes para pesquisas futuras, que poderão expandir o escopo para comparações interinstitucionais ou adotar perspectivas qualitativas sobre os fatores que sustentam a endogamia no ambiente acadêmico.

REFERÊNCIAS

ALTBACH, P. G.; YUDKEVICH, M.; RUMBLEY, L. E. Academic inbreeding: local challenge, global problem. **Asia Pacific Education Review**, p. 317-330, 2015. DOI: 10.1007/s12564-015-9391-8. Disponível em: https://www.researchgate.net/publication/282753989_Academic_inbreeding_local_challenge_global_problem. Acesso em: 09 mar. 2026.

BORENSTEIN, D.; PERLIN, M. S.; IMASATO, T. The Academic Inbreeding Controversy: Analysis and Evidence from Brazil. **Journal of Informetrics** 16, 2022. <https://doi.org/10.1016/j.joi.2022.101287>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1751157722000396>. Acesso em: 28 abr. 2026.

GIL, A. C. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2017. MATIAS, M. S. O. **Base referencial para o povoamento de repositórios institucionais**: coleta automatizada de metadados da Plataforma Lattes. 2015. Dissertação (Mestrado em Gestão de Organizações e Sistemas Públicos) - Universidade Federal de São Carlos, São Carlos, 2015. Disponível em: <https://repositorio.ufscar.br/handle/20.500.14289/6932>. Acesso em: 28 abr. 2026.

MATIAS, M. S. O.; AMARAL, R. M.; MATIAS, P. **Proxy customizado para acesso ao web service da Plataforma Lattes**: apresentação. São Carlos: Universidade



Federal de São Carlos, 2017. Disponível em: https://www.researchgate.net/publication/317348485_Proxy_customizado_para_acesso_ao_web_service_da_Plataforma_Lattes_Presentation. Acesso em: 29 abr. 2026.

SOARES, V. R. M. **Análise da endogamia das coautorias científicas de pesquisadores da UFSCar credenciados em programas de pós-graduação.** 2026. Tese (Doutorado em Ciência, Tecnologia e Sociedade) - Universidade Federal de São Carlos, São Carlos, 2026. Disponível em: <https://repositorio.ufscar.br/handle/20.500.14289/23969>. Acesso em: 28 abr. 2026.