

Solução tecnológica para produção e análise de uma rede de colaboração a partir de dados da Plataforma Lattes

Technological solution for the production and analysis of a collaboration network based on data from the Lattes Platform

Richard William Valdivia
<https://orcid.org/0000-0002-9027-2944>

Mestre em Tecnologia, Gestão e Saúde Ocular. Universidade Federal de São Paulo (UNIFESP) – Brasil.
richard.valdivia@unifesp.br

Rosemberg Francisco Rangel
<https://orcid.org/0009-0005-3508-1300>

Mestre em Administração Pública – Universidade Federal de Alfenas (UNIFAL) - Brasil. rosemberg.rangel@gmail.com

Amanda Marta Jardim Souza
<https://orcid.org/0000-0002-9829-6035>

Mestre em Inovação Tecnológica. Universidade Federal de São Paulo (UNIFESP) – Brasil. amandamjardims@gmail.com.

Maria Elisabete Salvador
<https://orcid.org/0000-0001-8692-5968>

Doutora em Ciências de Saúde. Escola Paulista de Medicina. Universidade Federal de São Paulo (UNIFESP) - Brasil.
elisabete.salvador@unifesp.br

Fábio Luís Falchi de Magalhães
<https://orcid.org/0000-0003-3307-8659>

Doutor em Informática e Gestão do Conhecimento. Universidade Federal de São Paulo (UNIFESP) - Brasil. fabiosimp@gmail.com

RESUMO

Programas de Pós-Graduação brasileiros devem atender às diretrizes da CAPES para assegurar a qualidade acadêmica, o que demanda instrumentos robustos de acompanhamento da produção científica. A Plataforma Lattes constitui uma fonte estruturada de dados curriculares, porém a ausência de padronização ampla na web dificulta a recuperação eletrônica e automatizada de informações em larga escala. Neste contexto, este artigo tem como objetivo apresentar e avaliar a ferramenta automatizada LattesXtractor, desenvolvida para extrair, organizar e analisar dados da Plataforma Lattes, com foco na área de Medicina III da Universidade Federal de São Paulo (UNIFESP). Trata-se de uma pesquisa descritiva e exploratória, com abordagem qualitativa e apoio quantitativo, na qual foram incluídos docentes-orientadores de dez Programas de Pós-Graduação *Stricto Sensu* em Medicina III. A aplicação, desenvolvida em Java®, processa arquivos XML do Currículo Lattes e gera estruturas adequadas para análise em software de redes, especialmente o Gephi, possibilitando a construção de grafos de colaboração. Os resultados demonstraram que a maior parte dos pesquisadores – 38,26% – possui entre duas e cinco colaborações internas, enquanto nos extremos opostos concentram-se os casos mais críticos: 10,74% não registram qualquer parceria científica dentro de seus programas e apenas 4,03% atingem o grau mais elevado de conectividade, evidenciando assimetrias relevantes na distribuição das parcerias científicas. A análise visual das redes permitiu

identificar pesquisadores centrais, comunidades de colaboração, grupos fragmentados e pesquisadores isolados, oferecendo subsídios concretos para a gestão dos programas e para a reflexão sobre as implicações desses padrões nas avaliações periódicas da CAPES.

Palavras-chave: Plataforma Lattes; Pós-graduação; Dados Cientométricos; Redes de Colaboração.

ABSTRACT

Brazilian graduate programs must comply with CAPES guidelines to ensure academic quality, which requires robust instruments to monitor scientific output. In this context, the Lattes Platform constitutes a structured source of curricular data, but the lack of broader standardization on the web hinders large-scale electronic and automated information retrieval. This article aims to present and evaluate the automated LattesXtractor tool, developed to extract, organize, and analyze data from the Lattes Platform, focusing on the Medicine III area at the Federal University of São Paulo (UNIFESP). This is a descriptive and exploratory study, with a qualitative approach supported by quantitative analyses, including faculty advisers from ten Medicine III graduate programs (*stricto sensu*). The application, developed in Java®, processes XML files from the Lattes Curriculum and generates structures suitable for analysis in network software, especially Gephi, enabling the construction of collaboration graphs. The results showed that most researchers – 38.26% – have between two and five internal collaborations, while at opposite extremes lie the most critical cases: 10.74% show no scientific partnership within their programs and only 4.03% reach the highest level of connectivity, revealing relevant asymmetries in the distribution of scientific partnerships. The visual analysis of the networks made it possible to identify central researchers, collaboration communities, fragmented groups and isolated researchers, providing concrete inputs for program management and for reflection on the implications of these patterns in periodic CAPES evaluations.

Keywords: Lattes Platform; Graduate studies; Scientometric data; Collaboration Networks.

Recebido em 15/01/2026. Aprovado em 12/05/2026. Avaliado pelo sistema *double blind peer review*. Publicado conforme normas da ABNT.

<https://doi.org/10.22279/navus.v18.2286>

1 INTRODUÇÃO

Os programas de pós-graduação enfrentam o desafio de cumprir as normativas da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) para manter a excelência acadêmica, onde a Plataforma Lattes se destaca como um recurso valioso para acompanhar o desempenho científico (CAPES, 2020). Com o avanço da Tecnologia da Informação, técnicas computacionais facilitam a análise de grandes volumes de dados cientométricos, preenchendo uma lacuna na criação de ferramentas automatizadas para extrair informações relevantes da web (Mena-Chalco, 2009).

Sob essa perspectiva, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) desempenha um papel fundamental na manutenção da base de dados de currículos acadêmicos da já mencionada Plataforma Lattes, como principal repositório de informações sobre pesquisadores brasileiros. Esta plataforma possibilita realizar análises amplas e detalhadas da produção científica nacional, oferecendo um panorama consistente das atividades de pesquisa desenvolvidas no país (Digiampietri, Mena-Chalco, 2012).

Duas tendências emergem desses esforços: a valorização de pesquisadores e a análise tanto quantitativa como qualitativa dos dados acadêmicos. A agregação e o processamento de informações de diversos pesquisadores permitem a formulação de políticas que atendam às exigências da CAPES, ao mesmo tempo que fomentam com informação relevante o currículo Lattes dos pesquisadores (Digiampietri, Mena-Chalco, 2012).

Por outro lado, a ausência de padrões dos dados públicos na internet torna complexa a coleta automatizada de informações. Esse problema se intensifica diante do crescimento exponencial de dados digitais (fenômeno conhecido como Big Data). Para superar essa dificuldade, foi proposta a utilização de metadados, informações que descrevem e organizam outros dados, como estratégia para facilitar a recuperação de informações. Nesse contexto, a Plataforma Lattes apresentou-se como uma fonte de dados particularmente adequada para pesquisas e análises, devido à sua estrutura bem-organizada e padronizada (Lima, 2005).

Propõe-se então apresentar uma ferramenta automatizada para extrair e analisar dados da Plataforma Lattes (Mena-Chalco, 2009) inicialmente focada na área da saúde. O referencial teórico, dividido em aspectos de Redes de Colaboração, Pós-Graduação no Brasil e Engenharia de Software, serve como base para a argumentação e metodologia do trabalho.

Essa abordagem permite a organização e apresentação de resultados relevantes, com destaque para a construção de uma ferramenta de software capaz de extrair e gerar conhecimento a partir dos dados do Currículo Lattes, essencial para uma análise mais aprofundada tanto quantitativa quanto qualitativamente.

2 REVISÃO DA LITERATURA

No contexto atual em que tecnologias da informação norteiam e tornam mais robustas as tomadas de decisão, apresentamos aqui os fundamentos teóricos que embasam o desenvolvimento e a aplicação do *LattesXtractor*, ferramenta proposta neste trabalho para recuperação e organização de informações provenientes da Plataforma Lattes com objetivo de dar apoio à gestão da pós-graduação. Para tanto, a revisão está organizada em três

eixos temáticos. O primeiro aborda as redes de colaboração científica, evidenciando sua relevância para a compreensão das relações entre pesquisadores e grupos de pesquisa. O segundo contextualiza o sistema de pós-graduação brasileiro e seus mecanismos de avaliação, coordenados pela CAPES, que constituem o principal cenário de aplicação da ferramenta. O terceiro discute conceitos de Engenharia de Software e linguagens de programação que fundamentam as escolhas tecnológicas adotadas na implementação do *LattesXtractor*.

2.1 Redes de Colaboração

As redes de colaboração representam instrumentos de intercâmbio de informações e apoio mútuo entre pesquisadores, grupos ou instituições. Esses arranjos colaborativos desempenham um papel fundamental no desenvolvimento da pesquisa científica e facilitam a construção coletiva de conhecimento, além de acelerar o progresso tecnológico por meio da união de diferentes competências e recursos (Maia; Zanotto; Caregnato, 2011).

Tanto instituições públicas quanto programas de pós-graduação realizam pesquisas de redes como estratégia para aprimorar seus processos internos, o que resulta em maior competitividade no cenário acadêmico (FRANCISCO, 2018). No ambiente universitário, as redes de colaboração oferecem benefícios como melhor compreensão das atribuições e desafios enfrentados pelos diferentes grupos de pesquisa, além de promover o reconhecimento compartilhado das conquistas alcançadas (Balancieri, 2005).

Redes de colaboração não apenas conectam pesquisadores, mas também unem competências para criar novos conhecimentos e inovações tecnológicas. Essa abordagem resulta em diversas perspectivas sobre problemas científicos e aproveita novas direções quando pesquisas são realizadas em conjunto. A colaboração entre pesquisadores aumenta significativamente a qualidade das publicações científicas e economiza tempo e recursos, beneficiando as agências financiadoras de pesquisas e a sociedade (Maia; Zanotto; Caregnato, 2011).

No contexto da pós-graduação, a análise dessas redes por meio de métricas de colaboração como grau de conectividade, centralidade e densidade permite identificar padrões que subsidiam a gestão acadêmica e o planejamento estratégico dos programas. A identificação de pesquisadores isolados, nós centrais e comunidades de colaboração oferece insumos concretos para coordenadores de programas avaliarem a coesão interna de seus grupos e alinharem suas estratégias às diretrizes de avaliação da CAPES (Francisco, 2018; Balancieri, 2004). É nesse cenário que ferramentas como o *LattesXtractor* se inserem, automatizando, compilando e organizando dados para viabilizar análises que, de outra forma, demandariam esforço manual considerável.

2.2 Pós-Graduação no Brasil

O crescimento e a consolidação da pós-graduação no Brasil é resultado da convergência de múltiplos fatores importantes. Entre eles, destaca-se a atuação estratégica das agências de fomento, com expressiva presença nas diversas áreas de pesquisa, a implementação de processos regulares de avaliação dos programas de pós-graduação e o envolvimento ativo da comunidade acadêmica (CAPES, 2020).

A pós-graduação brasileira representa um exemplo notável de política pública que conquistou reconhecimento e relevância pelos diferentes governos ao longo das décadas. Essa estabilidade política permitiu a elaboração e execução de planos estratégicos consistentes, direcionados tanto para a expansão quantitativa do sistema quanto para o fortalecimento do apoio financeiro aos programas e pesquisadores (CAPES, 2020).

A implementação de um sistema de avaliação periódica dos programas de pós-graduação pela CAPES constituiu uma experiência pioneira no país, representando um desafio significativo tanto para a comunidade acadêmica quanto para a própria instituição. Este processo avaliativo revelou-se uma ferramenta educativa valiosa, contribuindo para o aprimoramento contínuo dos programas.

Além disso, conquistou ampla aceitação e credibilidade no meio acadêmico, especialmente por ter sido desenvolvido com a participação ativa e colaborativa da própria comunidade de pesquisadores e coordenadores de programas de pós-graduação (Neves, 2020).

O sistema de avaliação da CAPES, constantemente aprimorado, funciona como uma ferramenta na busca de um padrão de excelência acadêmica para os programas de mestrado e doutorado no país. Nesse contexto, a capacidade de monitorar e analisar as redes de colaboração entre docentes torna-se um elemento relevante para a gestão dos programas, uma vez que a produção científica conjunta e a interdisciplinaridade figuram entre os critérios considerados nas avaliações quadrienais da CAPES (CAPES, 2020; Neves, 2020).

3 MÉTODOS E INSTRUMENTOS

Este trabalho caracteriza-se como um estudo descritivo de pesquisa aplicada, voltado ao desenvolvimento e avaliação de uma ferramenta de software para extração e análise de dados da Plataforma Lattes. O caráter descritivo orienta o registro sistemático dos padrões de colaboração identificados na área de Medicina III.

A Engenharia de Software é constituída por um conjunto de conhecimentos sistematizados e desenvolvidos com o propósito de descrever os instrumentos técnicos empregados na construção de programas de computador. Esse campo abrange desde a avaliação econômica, que verifica se há viabilidade financeira para construção do código, até a organização de controles e métricas que quantificam o esforço necessário para a elaboração e conclusão do projeto (Lima et al., 2020).

No contexto desta pesquisa, a escolha da linguagem Java® para o desenvolvimento do *LattesXtractor* justifica-se por suas características de portabilidade, robustez e amplo suporte ao processamento de arquivos XML (eXtensible Markup Language, ou Linguagem de Marcação Extensível), formato adotado pela Plataforma Lattes para exportação de dados curriculares. A capacidade do Java de manipular estruturas hierárquicas de dados de forma eficiente tornou-o adequado para a construção de uma arquitetura modular capaz de processar simultaneamente os currículos de múltiplos pesquisadores (Deitel, 2016).

A XML difere das linguagens de programação convencionais, funcionando como uma linguagem descritiva que organiza e atribui significado às informações de forma independente. Essa tecnologia baseia-se em um sistema de etiquetas (*tags*) que estruturam e contextualizam o conteúdo textual,

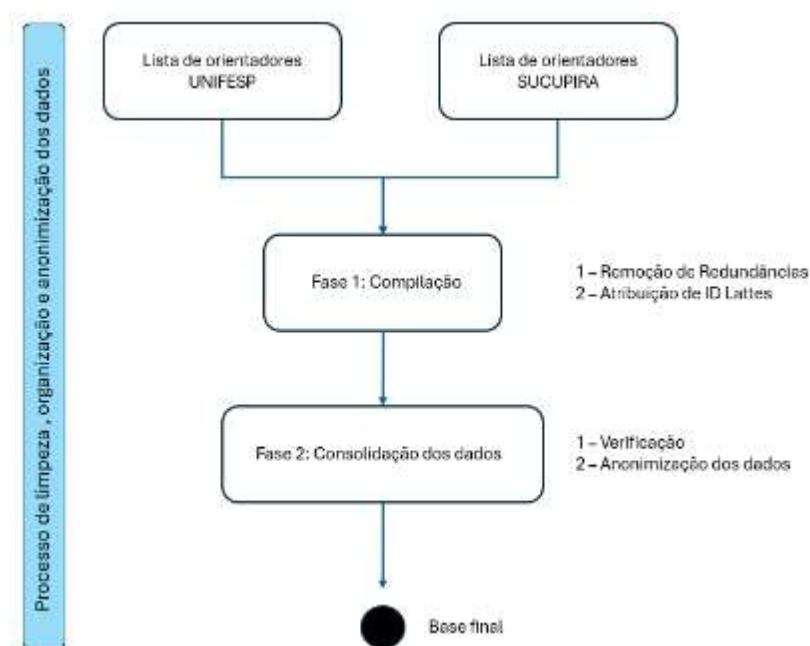
permitindo que diferentes sistemas computacionais possam interpretar e trocar informações de maneira eficiente. Essa capacidade de comunicação entre sistemas distintos - conhecida como interoperabilidade - torna a XML uma ferramenta valiosa para organizar e compartilhar dados em diversos contextos tecnológicos (Meddage et al., 2022).

No caso do *LattesXtractor*, a interoperabilidade proporcionada pela XML é central para o funcionamento da ferramenta: os arquivos gerados pelo *scriptLattes* no formato XML são lidos, interpretados e reorganizados pelo *LattesXtractor*, que por sua vez exporta os dados no formato .GEFX (*Gephi Files Extension*) também baseado em XML - para consumo pelo *Gephi*. Essa cadeia de interoperabilidade entre ferramentas distintas exemplifica o potencial da XML.

Para formar a base de dados utilizou-se a relação de docentes-orientadores dos Programas de Pós-Graduação stricto sensu da área de Medicina III da Universidade Federal de São Paulo (UNIFESP), considerando professores credenciados como orientadores de mestrado e/ou doutorado à época da extração dos dados, realizada em 2025. Inicialmente, foram obtidas informações na base institucional da UNIFESP, contemplando código CAPES do programa, nome do programa, nome do orientador e código Lattes de 16 dígitos.

Em seguida, realizou-se uma segunda busca na Plataforma Sucupira, a fim de conferir a lista de dados primários de orientadores, além de ajustar o corte dos dados, ainda que essa base não contivesse o ID Lattes. As duas bases foram integradas em planilha eletrônica, por meio de busca por nome do pesquisador; nos casos sem correspondência automática, o ID Lattes foi obtido por meio de uma busca nominal direta na Plataforma Lattes. Dessa forma, constituiu-se a base final de docentes utilizada nas etapas seguintes de extração e análise, conforme mostra a Figura 1.

Figura 1 - Consolidação dos dados



Os procedimentos metodológicos seguiram quatro etapas principais:

1. Revisão da literatura sobre redes de colaboração, pós-graduação no Brasil e engenharia de software, que fundamentou a proposta da ferramenta e da análise de redes;
2. Desenvolvimento de uma aplicação computacional em linguagem Java®, denominada *LattesXtractor*, responsável pela leitura e processamento dos arquivos XML do Currículo Lattes;
3. Exportação dos dados estruturados para posterior análise em software especializado em redes, em especial o Gephi, mediante geração de arquivos em formato adequado, como .CSV (*Comma Separated Values*) e .GEFX (*Gephi File Extension*).
4. Realização de um estudo de caso com os dez Programas de Pós-Graduação em Medicina III da UNIFESP, para validação prática da ferramenta e das visualizações de colaboração.

A confiabilidade dos dados processados pelo *LattesXtractor* foi verificada por meio de validação cruzada manual entre duas fontes independentes: a base institucional da UNIFESP, com 161 registros, e a Plataforma Sucupira, com 160 registros. A conciliação entre as bases foi realizada por correspondência nominal: cada registro da base UNIFESP foi comparado com a base Sucupira pelo nome do pesquisador. O critério de exclusão adotado foi rigoroso: foram removidos os registros cujos dados não eram exatamente coincidentes entre as duas bases, bem como aqueles presentes em apenas uma delas. Os registros sem correspondência automática tiveram o ID Lattes obtido por busca nominal direta na Plataforma Lattes. Após a remoção de 12 registros que não atenderam ao critério de coincidência exata entre as fontes, obteve-se uma base final de 149 pesquisadores, cuja consistência foi atestada pela convergência entre as fontes – sendo a Sucupira uma base oficial validada pela CAPES nas avaliações quadrienais. Esse processo de validação, ainda que manual, assegurou a integridade dos dados de entrada utilizados nas análises subsequentes.

O *LattesXtractor* foi projetado segundo uma arquitetura modular e orientada a objetos, organizando as informações curriculares em classes e estruturas de dados específicas, como *CurrículoLattes*, *Pesquisador* e *ArtigoVO*, entre outras. Esse modelo permite transformar os dados complexos do Currículo Lattes em objetos mais simples e manipuláveis, favorecendo sua reorganização sistemática para diferentes tipos de análises acadêmicas e científicas. A aplicação realiza a leitura dos arquivos XML gerados pelo *scriptLattes*, armazena os dados em memória em uma estrutura hierárquica (currículo-pesquisador-produção) e, a partir disso, gera arquivos tabulares e de rede que podem ser interpretados por ferramentas externas, como o Gephi.

A análise quantitativa contemplou a contagem de colaborações entre pesquisadores e a classificação dos graus de relacionamento por meio da função $G(N,M)$, em que N e M representam faixas de número de relações estabelecidas com outros docentes (por exemplo, $G(0,0)$, $G(1,1)$, $G(2,5)$, $G(6,10)$, $G(11,15)$ e $G(16,20)$). Esses resultados foram organizados em tabelas e gráficos, permitindo estimar percentuais de pesquisadores em cada faixa de colaboração e identificar, por exemplo, aqueles sem relações internas ou com alta conectividade.

Complementarmente, a análise qualitativa e visual foi realizada a partir dos grafos de colaboração gerados no *Gephi*, nos quais cada pesquisador é representado como vértice, enquanto cada publicação conjunta é representada como uma aresta. O tamanho dos vértices e a estrutura das

arestas possibilitaram identificar pesquisadores centrais, nós estratégicos, comunidades de colaboração, grupos fragmentados e pesquisadores isolados, bem como comparar padrões de organização interna entre programas específicos, como Medicina (Urologia), Oftalmologia e Ciências Visuais e Ciência, Tecnologia e Gestão Aplicadas à Regeneração Tecidual. Essas visualizações ofereceram suporte para interpretações sobre a coesão das redes, a distribuição das colaborações e possíveis implicações para a gestão e avaliação dos Programas de Pós-Graduação da área de Medicina III.

3.1 Especificação da ferramenta LattesXtractor

Para esta pesquisa, desenvolveu-se um módulo de software específico, capaz de extrair e processar informações da Plataforma Lattes. O programa opera a partir da leitura de arquivos XML gerados pelo *scriptLattes*, ferramenta complementar ao *LattesXtractor*, responsável pela coleta automatizada dos currículos na Plataforma Lattes que contêm dados curriculares de um conjunto de pesquisadores, reorganizando essas informações para posterior análise acadêmica.

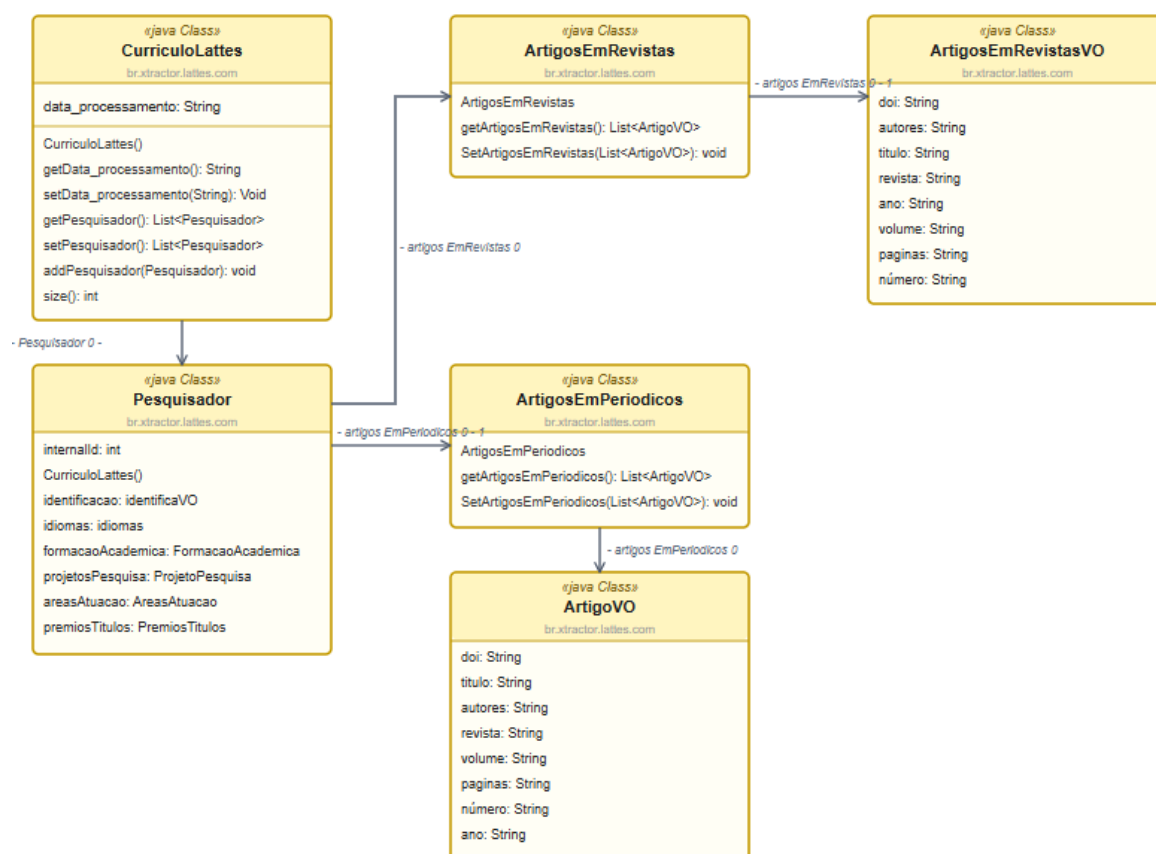
Após o processamento dos arquivos XML, os dados são organizados em memória em uma estrutura de dados própria, o que permite sua manipulação e adaptação para diferentes tipos de análises científicas. Essa estruturação viabiliza consultas flexíveis e a geração de arquivos em formatos adequados às etapas seguintes de tratamento e visualização.

O software foi desenvolvido em linguagem Java®, seguindo uma abordagem modular e orientada a objetos. Essa arquitetura permite organizar as informações curriculares de forma hierárquica - currículo, pesquisador e produção -, facilitando a extração seletiva de dados e a geração de arquivos nos formatos .CSV (*Comma Separated Values*) e .GEFX (*Gephi File Extension*), compatíveis com ferramentas de análise de redes como o *Gephi*.

A Figura 2 apresenta a estrutura simplificada da ferramenta, ilustrando como as informações curriculares são organizadas em níveis hierárquicos que permitem navegar do conjunto de pesquisadores até cada item de produção científica de forma estruturada.

Embora o *scriptLattes* já disponha de recursos para coleta e organização de dados da Plataforma Lattes, o *LattesXtractor* foi desenvolvido com um propósito distinto e complementar: atuar na etapa seguinte ao processo de extração, transformando os arquivos XML gerados pelo *scriptLattes* em estruturas de dados organizadas e exportáveis para análise de redes. Diferentemente da visualização nativa do Lattes - restrita à consulta individual de currículos e sem possibilidade de exportação, o *LattesXtractor* consolida simultaneamente os dados de um conjunto de pesquisadores, identifica relações de coautoria entre eles e gera arquivos no formato compatível com o *Gephi*. Isso viabiliza a aplicação de algoritmos especializados de análise de redes, como detecção de comunidades e cálculo de centralidade, que não estão disponíveis nas ferramentas existentes para consulta ao Lattes.

Figura 2 – Diagrama de Classes simplificado



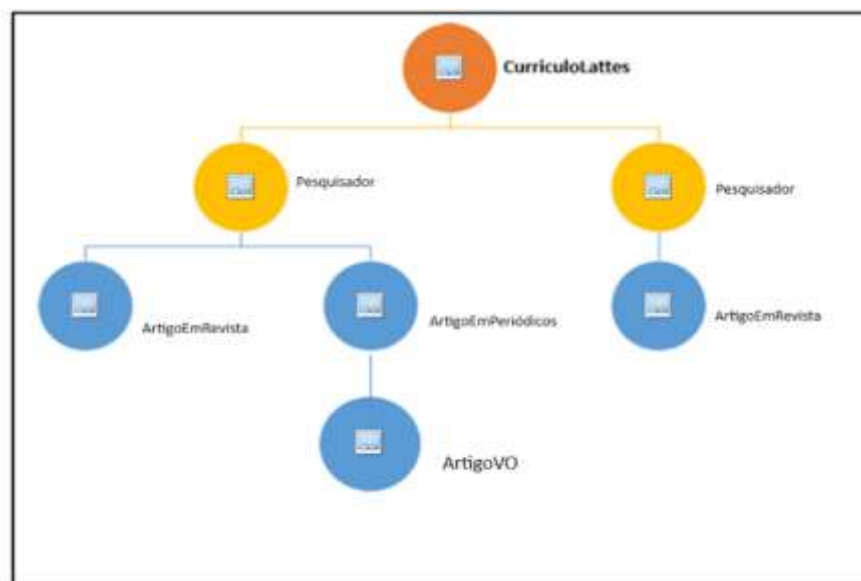
A classe **CurriculoLattes**, apresentada no diagrama, representa uma versão simplificada e organizada do currículo da Plataforma Lattes, reunindo uma coleção estruturada de pesquisadores. Essa classe dispõe de uma interface bem definida, que permite acessar as informações necessárias por meio de métodos específicos de consulta.

Para cada solicitação realizada, a ferramenta retorna uma lista correspondente aos dados requisitados, mantendo um funcionamento hierárquico. Parte-se de um conjunto de pesquisadores organizados em lista e, para cada pesquisador, é possível obter listas detalhadas de informações relacionadas à sua produção acadêmica.

Por exemplo, para um pesquisador, pode-se recuperar uma lista de artigos em periódicos publicados por ele, por meio de chamadas que retornam objetos do tipo **ArtigoVO**. Esse mecanismo permite navegar da entidade pesquisador até cada item de produção científica de forma estruturada.

Da mesma forma, é possível acessar outros tipos de produção, como artigos em revistas, por meio de objetos específicos, como **ArtigosEmRevistas** e **ArtigosEmRevistasVO**. Primeiro obtém-se o pesquisador, em seguida a lista de ArtigosEmRevistas e, por fim, o dado detalhado armazenado no objeto correspondente; a Figura 3 apresenta esse modelo em uma estrutura em árvore, facilitando sua visualização.

Figura 3 – Estrutura de classes em uma visualização em árvore



A estrutura é montada a partir da leitura de um arquivo XML que é criado após a execução do *scriptLattes* e que será detalhada mais adiante. A classe *LattesXtractorGUI* é o ponto de entrada da ferramenta conforme o diagrama de classes da Figura 2. Pode-se observar que ela contém o motor de leitura de dados do XML representado pela classe *ScriptLattesXmlParser*.

Para a seleção dos dados, o critério de organização para obtenção da base inicial foi que os pesquisadores deveriam ser docentes da pós-graduação *Stricto Sensu* e orientadores credenciados de mestrado e/ou doutorado nos programas de Pós-Graduação oferecidos pela UNIFESP da área de Medicina III.

Entretanto, entendendo que esse universo de dados é um universo de fluxo contínuo, isto é, docentes iniciam atividades de orientação e se desvinculam dos programas com frequência, realizou-se uma segunda busca de informações na Plataforma Sucupira para obter dados oficiais de orientadores cadastrados para o corte correto de dados. Essa segunda busca, entretanto, traz uma desvantagem: a base extraída possui nome do programa e o nome do pesquisador sem o ID Lattes.

Para contornar esse problema, colocaram-se as duas bases em uma planilha Excel, na qual foi criada uma busca que coincidissem o nome do pesquisador nas duas bases. Caso encontrasse a semelhança, aplicava-se na célula logo após o nome do pesquisador o ID Lattes obtido da base da UNIFESP. Para os nomes em que não foi encontrada correspondência, obteve-se o ID Lattes por meio da busca simples na Plataforma Lattes baseada no nome do pesquisador. Dessa forma foi possível criar uma base de dados final que serviu como fonte para a extração dos dados. A Tabela 1 mostra um exemplo dos quantitativos de dados obtidos.

Para a obtenção e organização de informações, a extração dos dados de orientadores da UNIFESP tomou-se o cuidado de se obter os seguintes campos para compor as informações:

- Código CAPES do Programa
- Nome do Programa
- Nome do Orientador
- Código Lattes de 16 dígitos

Dessa forma foi possível categorizar orientadores por grupos onde o Programa é a chave para o corte, por meio de uma query SQL (*Structured Query Language*) aplicada na base de dados Oracle da UNIFESP. O resultado é um arquivo texto com extensão CSV, facilmente lido pelo Excel e apresentado na Tabela 2. Na Plataforma Sucupira a extração dos dados foi realizada pelo endereço internet <https://sucupira.capes.gov.br> (pesquisa realizada em novembro de 2025). Os dados obtidos estão apresentados na Tabela 1.

Tabela 1 - Extração de dados

Pesquisadores	Total
Base de dados da UNIFESP	161
Plataforma Sucupira	160
Total sem duplicidade	149
Pesquisadores sem Lattes	2

Observando a Tabela 1, a quantidade de pesquisadores encontrados nas bases foi praticamente idêntica, tornando assim a base de dados confiável, uma vez que a base do Sucupira é base de dados oficial validada pela CAPES nas avaliações quadrienais. Após a extração, o *LattesXtractor* relacionou os registros dos docentes para cada programa verificando se o docente já havia publicado com outro pesquisador do mesmo grupo. Se essa condição fosse satisfeita, então se criava-se um relacionamento entre esses pares.

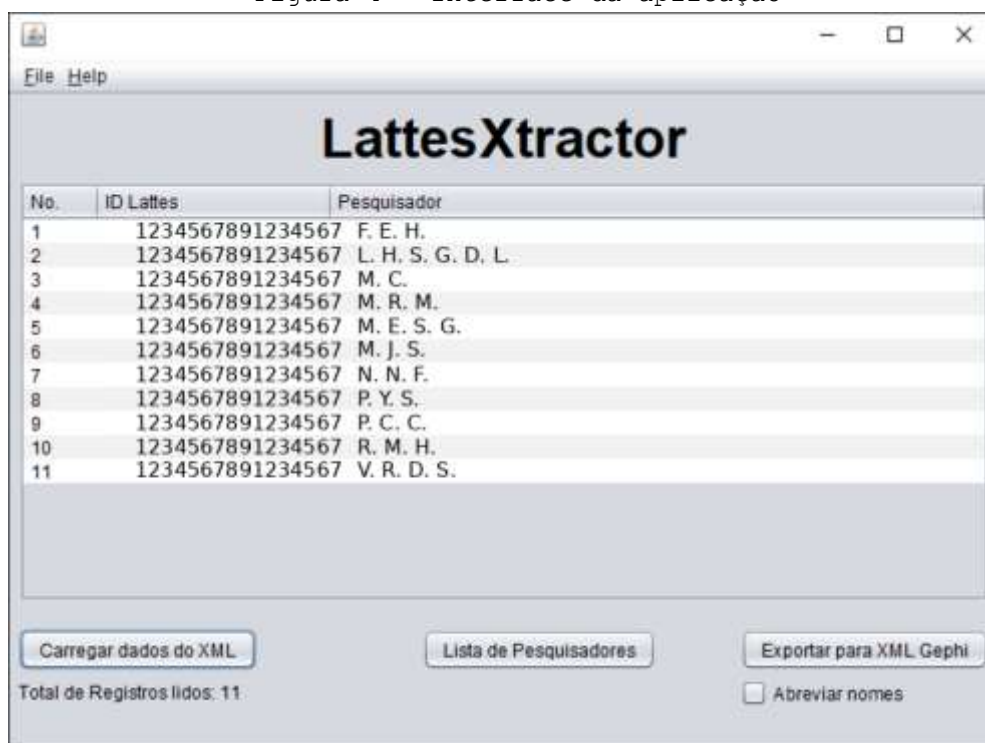
4 RESULTADOS

Nessa seção, são apresentados os resultados das análises realizadas com a base de dados coletada e seus principais produtos. Os nomes dos pesquisadores foram pseudo-anonimizados, de modo que aparecem apenas pelas iniciais. Para cada Programa de Pós-Graduação *Stricto Sensu* ofertado pela Universidade Federal de São Paulo (UNIFESP), foi construído um grafo de colaboração, representando visualmente as relações entre os docentes. Essas relações se configuram a partir de publicações conjuntas registradas no Currículo Lattes, conforme salientado por Balancieri (2004), que destaca a importância das redes de colaboração como instrumentos para compreender a dinâmica da produção científica.

A Figura 4 apresenta a tela principal do *LattesXtractor*, interface por meio da qual é realizada a carga dos dados provenientes de um arquivo XML exportado pelo *scriptLattes*. Ao acionar o botão Carregar dados do XML, a ferramenta processa o arquivo e organiza as informações em memória segundo uma estrutura hierárquica composta por três níveis – currículo, pesquisador e produção –, exibindo a lista de pesquisadores, acompanhada de seus identificadores na Plataforma Lattes, conforme ilustrado na Figura 4, em que os nomes foram substituídos por identificadores genéricos (Pesquisador_01 a Pesquisador_11) e os IDs Lattes ocultados (XXXXXXXXXXXXXXXXXX) em atendimento aos princípios da Lei Geral de Proteção de Dados Pessoais (LGPD – Lei nº 13.709/2018). A partir dessa estrutura, o usuário pode acionar o botão Lista de Pesquisadores para visualizar os dados organizados ou utilizar o botão Exportar para XML *Gephi*, que gera arquivos tabulares e de rede compatíveis com ferramentas externas de análise e visualização, como o *Gephi*. A opção abreviar nomes oferece ainda um recurso de pseudonimização dos nomes durante o processamento, substituindo-os por abreviações que dificultam a identificação direta dos pesquisadores, em consonância com os princípios de minimização e proteção

de dados estabelecidos pela Lei Geral de Proteção de Dados Pessoais (LGPD – Lei nº 13.709/2018).

Figura 4 – Interface da aplicação



A representação gráfica em forma de grafos permite visualizar cada pesquisador como um vértice, e cada publicação conjunta como uma aresta que conecta dois ou mais pesquisadores. Essa metodologia segue a abordagem de análise de redes sociais aplicada ao contexto acadêmico (Maia; Zanotto; Caregnato, 2011), possibilitando compreender padrões de centralidade, comunidades internas e lacunas de interação.

No caso específico da Medicina III da UNIFESP, foram extraídos os currículos Lattes de docentes-orientadores vinculados a dez programas de pós-graduação stricto sensu, cujas características estão descritas na Tabela 2.

Observa-se que os programas possuem diferentes níveis e notas atribuídas pela CAPES, variando de conceitos mais consolidados, como Cirurgia Translacional e Oftalmologia e Ciências Visuais (nota 6), até programas em consolidação, como Ciência, Tecnologia e Gestão Aplicadas à Regeneração Tecidual (nota 3). Essa diversidade é relevante, pois a literatura (NEVES, 2020; CAPES, 2020) destaca que a avaliação periódica da pós-graduação não apenas mede desempenho, mas também orienta políticas de incentivo à colaboração e à interdisciplinaridade.

Dessa forma, a análise dos grafos de colaboração permite não só compreender como os pesquisadores se organizam internamente em seus programas, mas também verificar como os níveis de excelência acadêmica, refletidos nas notas CAPES, podem estar associados a redes mais densas, equilibradas e integradas. Como aponta Francisco (2018), redes bem estruturadas promovem maior compartilhamento de competências e favorecem a inovação científica, o que justifica a importância de relacionar os resultados observados à política de avaliação e fomento da CAPES.

Foram extraídos os Currículos Lattes dos docentes-orientadores de 10 Programas de Pós-Graduação que fazem parte da área de Medicina III na UNIFESP cujos dados de identificação e avaliação pela CAPES estão apresentados na Tabela 2.

Tabela 2 - Programas pertencentes à área de Medicina III.

PROGRAMA DE PÓS-GRADUAÇÃO		NÍVEIS E NOTAS CAPES		
Código do Programa	Nome do Programa	ME	DO	MP
33009015009P1	Ciência Cirúrgica Interdisciplinar	4	4	-
33009015173P6	Ciências de Saúde Aplicada ao Esporte e a Atividade Física	-	-	3
33009015093P2	Ciência, Tecnologia e Gestão Aplicadas à Regeneração Tecidual	-	-	3
33009015038P1	Cirurgia Translacional	6	6	-
33009015014P5	Medicina (Ginecologia)	4	4	-
33009015013P9	Medicina (Obstetrícia)	4	4	-
33009015018P0	Medicina (Otorrinolaringologia)	4	4	-
33009015021P1	Medicina (Urologia)	4	4	-
33009015024P0	Oftalmologia e Ciências Visuais	6	6	-
33009015082P0	Tecnologia, Gestão e Saúde ocular	4	-	3

Os dados obtidos por meio do processo de extração das informações, em sua forma original, não proporcionam uma visualização clara que facilite a identificação de padrões ou a geração de novos conhecimentos, exigindo análises mais complexas e demoradas.

Para melhor organizar essas informações, os dados dos pesquisadores foram estruturados em formato tabular, apresentando o nome de cada pesquisador e o número de colaborações que ele estabelece dentro do grupo analisado. A Tabela 3 apresenta um exemplo desses resultados, mostrando os 10 pesquisadores com maior número de colaborações, organizados em ordem decrescente conforme seu grau de conectividade na rede.

Tabela 3 - Grau de colaboração aplicado para cada pesquisador.

Graus de Colaboração entre pesquisadores								
Pesquisador		Grau de Relacionamento	G(0,0)	G(1,1)	G(2,5)	G(6,10)	G(11,15)	G(16,20)
1	L.M.F	20	0	0	0	0	0	1
2	M.D.J.S	19	0	0	0	0	0	1
3	M.E.F.N	17	0	0	0	0	0	1
4	M.G.F.S	16	0	0	0	0	0	1
5	R.B.M.J	16	0	0	0	0	0	1
6	A.L.H.D.L.F	16	0	0	0	0	0	1
7	A.F.M	15	0	0	0	1	0	0
8	O.S	15	0	0	0	1	0	0
9	I.D.C.G.D.S	14	0	0	0	1	0	0
10	E.A.J	14	0	0	0	1	0	0

Para aprimorar a visualização dos dados, foram criadas colunas adicionais contendo as relações de colaboração representadas pela função $G(N,M)$, em que N representa o número mínimo de relações da faixa e M indica o número máximo de relações da faixa, conforme definido no Quadro 1.

Quadro 1 - Relações de Colaboração

G (N,M)	N-> N° mínimo de relações
	M-> N° máximo de relações

Assim M e N denotam a quantidade de relacionamento com outros pesquisadores:

- G(0,0) a Pesquisadores que não têm relacionamentos com outros pesquisadores;
- G(1,1) a Pesquisadores que têm relacionamentos com somente um pesquisador;
- G(2,5) a Com relacionamento entre dois e cinco pesquisadores;

Com os dados tabulados obtemos, a Tabela 4 com o percentual de relações encontradas o total de pesquisadores que possuem aquela relação.

Tabela 4 - Percentual de colaboração e quantitativo de pesquisadores com grau.

	G (0,0)	G (1,1)	G (2,5)	G (6,10)	G (11,15)	G (16,20)
%	10,74	10,07	38,26	8,72	28,12	4,03
Grau (total)	16	15	57	13	42	6

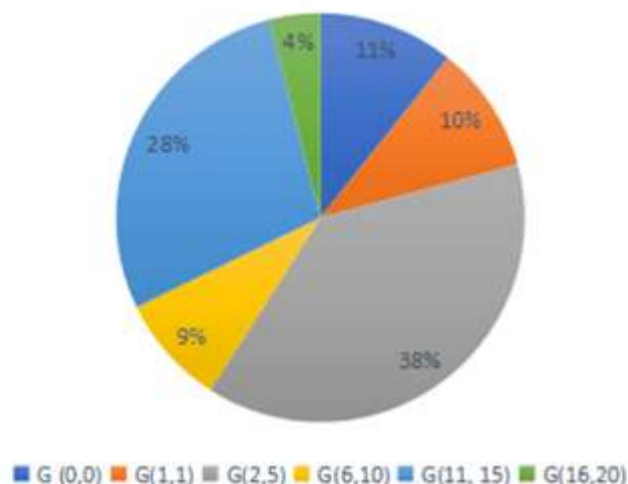
Observamos que o maior percentual de pesquisadores da Medicina III da UNIFESP é constituído por aqueles que possuem entre 2 e 5 relações com outros pesquisadores com 38,26% e no lado oposto, apenas 4,03% dos pesquisadores da UNIFESP possuem entre 10 e 20 relações. Outro dado interessante é que do universo de Programas da Medicina III da UNIFESP, 10,74% dos pesquisadores estão totalmente fora dos relacionamentos esperados. São aqueles que não publicam com nenhum outro ente do contexto dos Programas selecionados.

Porém observamos que essa organização resumida dos dados ainda não consegue demonstrar de maneira clara e intuitiva como ocorrem as interações entre os pesquisadores. Embora os percentuais apresentados sejam coerentes e forneçam uma base sólida para avaliações analíticas, ainda existe a necessidade de uma representação mais detalhada que permita compreender melhor a natureza e a dinâmica das colaborações acadêmicas estabelecidas entre os membros do grupo.

O Gráfico 1 apresenta a distribuição percentual do grau de colaboração entre os pesquisadores do Programa de Medicina III da UNIFESP. Embora esta visualização em formato de gráfico de setores (pizza) facilite a compreensão dos dados organizados na Tabela 3, ela ainda apresenta limitações importantes: não permite identificar com precisão as conexões específicas estabelecidas entre pesquisadores individuais nem a força dessas interações em detalhes.

Conforme destacam Maia, Zanutto e Caregnato (2011), análises baseadas exclusivamente em números e percentuais não conseguem capturar toda a complexidade das redes de colaboração científica. Torna-se necessário, portanto, utilizar métodos complementares que revelem tanto as colaborações oficiais e documentadas (como coautorias em publicações) quanto as interações mais informais que também influenciam o desenvolvimento da pesquisa acadêmica.

Gráfico 1 - Distribuição dos graus de colaboração.



Para superar essa limitação, foram utilizadas técnicas para aprofundar a análise qualitativa das relações. A partir do *LattesXtractor*, os dados extraídos em formato XML foram convertidos para um arquivo intermediário no padrão “.GEFX (*Gephi File Extension*)”, o que possibilitou o processamento no *software Gephi*. Esse recurso, conforme defendido por Lima (2005), exemplifica a importância da organização de metadados estruturados para viabilizar análises robustas em ambientes de big data. No *Gephi*, as relações são representadas em forma de grafos, permitindo observar interconexões, comunidades de maior densidade e discrepâncias entre pesquisadores.

Na rede de colaboração, cada pesquisador é representado como um vértice (nó) e cada publicação conjunta como uma aresta (linha). O tamanho do vértice indica o grau de centralidade (*degree centrality* – *métrica formal de Análise de Redes Sociais (ARS/SNA)*) do pesquisador: quanto maior o nó, maior a quantidade de conexões estabelecidas. Esse recurso gráfico possibilita distinguir redes normalizadas – com nós de tamanhos semelhantes – de redes concentradas em indivíduos específicos, confirmando os apontamentos de Francisco (2018) sobre o papel dos chamados “nós estratégicos” na dinâmica da produção científica.

Ao analisarmos especificamente o Programa de Medicina (Urologia), observou-se a presença de um pesquisador-chave, identificado como R.P.B., que publicou com oito dos 14 integrantes do grupo. Três pesquisadores, entretanto, não apresentaram nenhuma colaboração interna. Essa configuração é consistente com o que Balancieri (2004) descreve sobre redes de colaboração acadêmica: a presença de pesquisadores centrais pode favorecer a integração entre comunidades – como se observa entre os grupos lilás e verde da Figura 5 –, enquanto a ausência de conexões internas em alguns membros pode indicar diferentes perfis de inserção na rede, sem que isso implique necessariamente uma fragilidade estrutural do programa.

A análise detalhada do Programa de Pós-Graduação em Medicina (Urologia) evidencia, de forma clara, a existência de um pesquisador central (R.P.B.), que se conecta com nove dos catorze integrantes do grupo. Essa configuração é confirmada tanto pela rede de colaboração (Figura 5) quanto pelo grau de relacionamento em barras (Gráfico 2), onde R.P.B. aparece como o nó mais influente. Como destacam Balancieri (2004) e Maia, Zanotto e Caregnato (2011), a centralidade excessiva pode gerar dependência

em um único ator, enfraquecendo a resiliência da rede, ainda que também funcione como elo estratégico de integração entre comunidades.

Figura 5 – Rede de Colaboração de Medicina (Urologia)

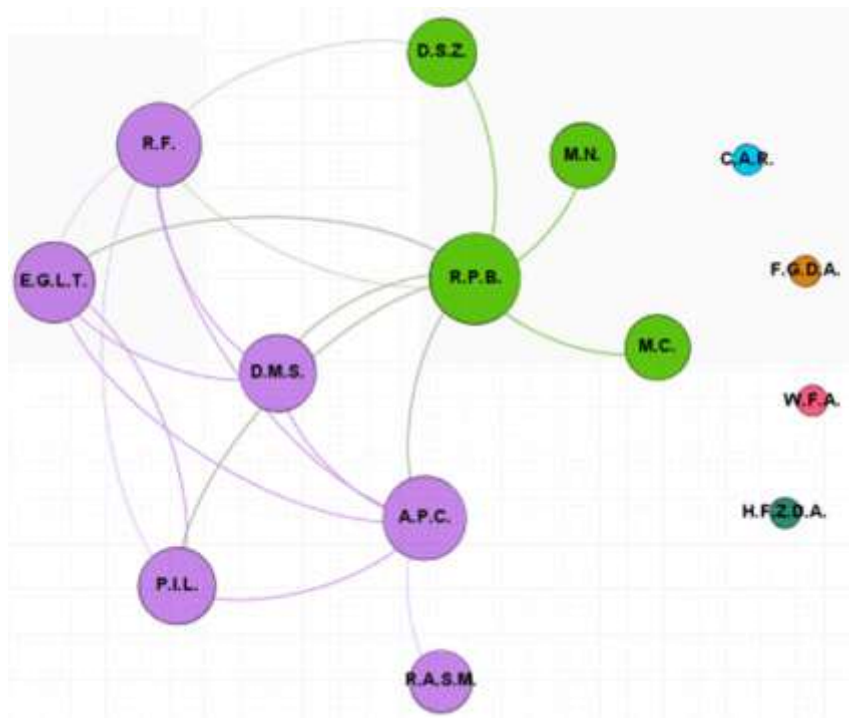
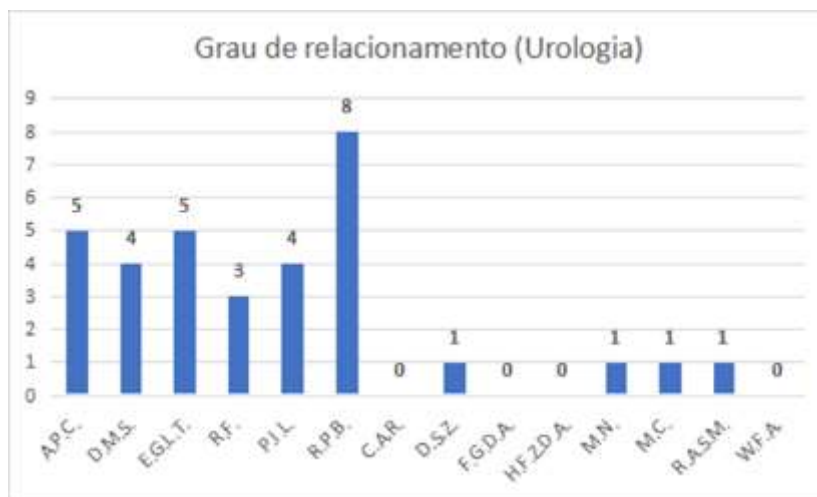


Gráfico 2 – Grau de relacionamento Medicina (Urologia)



A visualização das comunidades em cores permite aprofundar essa interpretação. Na rede, conforme a Figura 5, observam-se três grupos principais: um lilás, um verde e um conjunto mais fragmentado. O pesquisador R.P.B., ao lado de D.S.Z., atua como elo entre os grupos verde e lilás, enquanto M.N. e M.C. mantêm relações restritas apenas com R.P.B., sem sinergia com os demais pares. Essa fragmentação reforça a análise de Francisco (2018), segundo a qual redes equilibradas – caracterizadas por maior distribuição de colaborações entre os membros – tendem a ser mais robustas e sustentáveis do ponto de vista científico.

Outro ponto relevante é a presença de pesquisadores que aparecem totalmente isolados ou com apenas uma colaboração (como C.A.R., F.G.D.A., W.F.A. e H.F.Z.D.A.), o que pode refletir diferentes perfis de inserção na rede. Esses casos podem refletir desde questões técnicas, como atualização incompleta do Currículo Lattes, até fatores estruturais, como baixa inserção em grupos de pesquisa consolidados. A identificação desses padrões pode oferecer subsídios para reflexões sobre a organização interna dos programas, cabendo aos coordenadores avaliar, em cada contexto específico, se tais configurações demandam atenção (NEVES, 2020; CAPES, 2020).

A representação em grafo mostrou-se mais informativa do que as abordagens tabulares e em gráfico de barras: enquanto estas permitem identificar o pesquisador com maior número de conexões, não revelam com quem ele se relaciona nem a intensidade dessas relações. A visualização em rede, por sua vez, permite perceber de forma imediata a posição relativa de cada pesquisador, suas conexões específicas e o papel que desempenha na estrutura do grupo.

A Rede de Colaboração permite apresentar os resultados em comunidades que, nesse caso, são representadas por diferentes cores. Na Rede de Colaboração da Urologia observamos três comunidades distribuídas das cores lilás, verde e laranja. Os elementos que não se relacionam (não publicam com outros elementos do grupo) formam comunidades independentes.

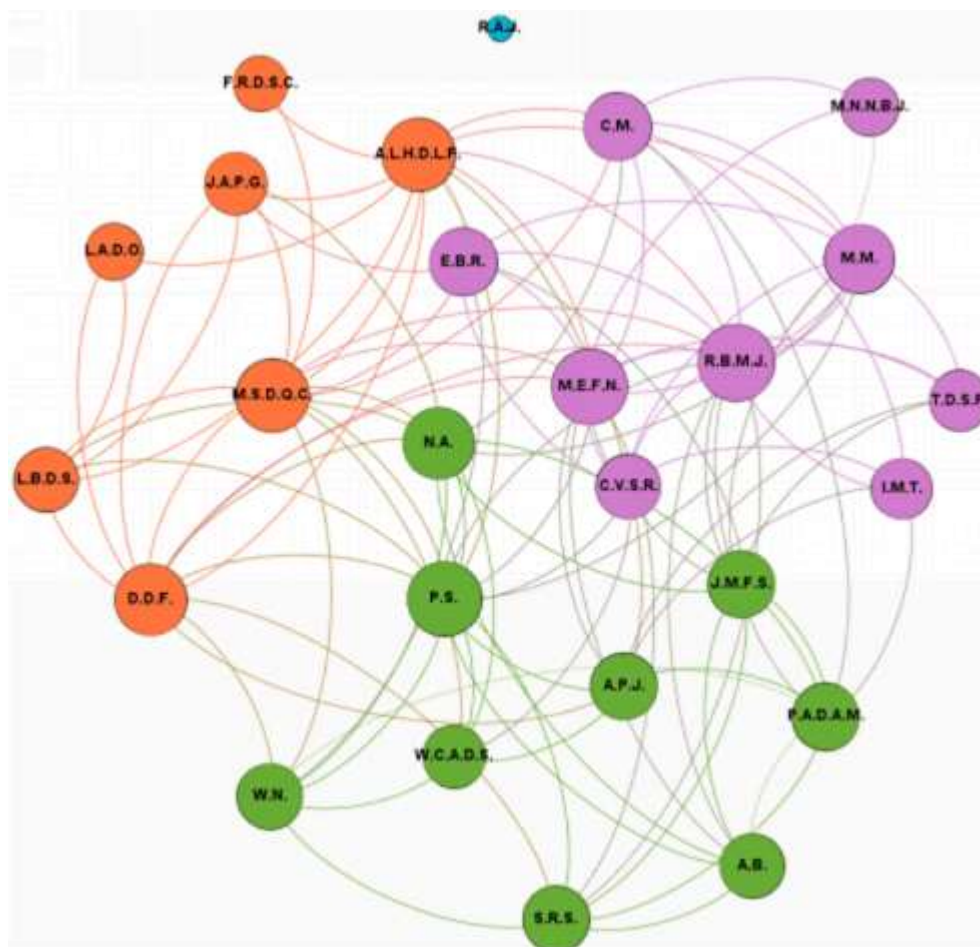
Essa relação de cores denota que tais pessoas costumam publicar entre si, então a relação é muito forte entre eles. Observa-se no gráfico que entre cada grupo existem muitas arestas em comum. Quando um grupo se liga a outro de forma mais fraca (poucas arestas de ligação), então essa passa a ter uma cor diferente.

No exemplo da Urologia, os pesquisadores R.P.B. e D.S.Z, ambos da comunidade verde, relacionam-se com outros elementos da comunidade lilás. Mas M.N. e M.C. relacionam-se somente com R.P.B. que é do grupo verde.

Essa diferenciação de cores por grupo ajuda a distinguir situações importantes em uma análise mais profunda. Nesse caso, a análise mostrou que há pesquisadores que não interagem com as pessoas do mesmo grupo e que a criação de grupos maiores seria um aspecto desejado no programa. Caberia, portanto, a um coordenador de área de um programa de pós-graduação verificar e intervir a fim de corrigir essa situação caso ela pudesse resolver algum problema relacionado ao alinhamento dos objetivos do Programa de Pós-Graduação em questão.

A rede de colaboração da Oftalmologia e Ciências Visuais, apresentada na Figura 6 mostra um resultado interessante de observação. É possível analisar agora um universo de colaboração que apresenta um grau de relacionamento maior que o exemplo anterior e a análise permite um grau de hipótese maior.

Figura 6 - Rede de Colaboração Oftalmologia e Ciências Visuais.



As três comunidades identificadas pelo algoritmo de detecção de comunidades e representadas nas cores lilás, verde e laranja evidenciam padrões distintos de densidade de arestas intracomunidade, indicando coesão interna diferenciada entre os subgrupos. Há aqui um fenômeno importante a ser observado: somente um pesquisador apresenta-se como nó periférico não comunitário, sem arestas que o vinculem a nenhuma das três comunidades identificadas.

Isso pode ser ocasionado por alguns motivos:

- R.A.J. não está com o Lattes atualizado;
- R.A.J. publica com outros pesquisadores fora do grupo ao qual pertence;
- R.A.J. ainda é novo no grupo;

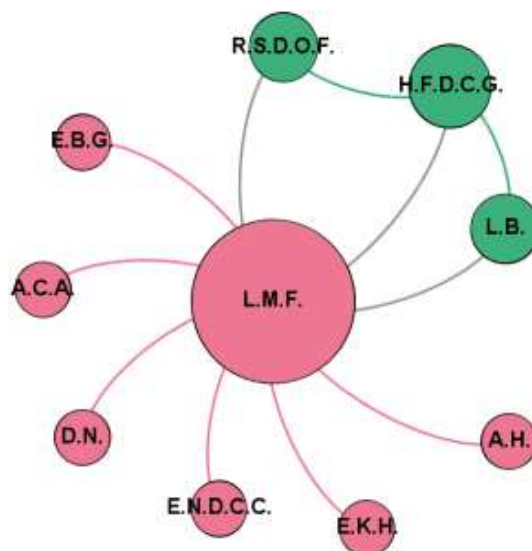
Essas hipóteses ilustram como a análise visual da rede pode levantar questões relevantes para o coordenador do programa, que dispõe de contexto institucional para interpretar adequadamente cada situação.

Por outro lado, observamos que o vértice que representa os pesquisadores possui uma normalização. Isso é interessante pois, como existe muita interação entre esse grupo, é natural que não exista um pesquisador chave. Essa observação pode ser também ratificada pela quantidade de arestas que são apresentadas mostrando um grupo muito equilibrado.

Na Figura 7 evidencia-se também um fenômeno relevante obtido pela análise da estrutura topológica da rede. Em uma análise comparativa dos

elementos que compõem o gráfico em relação ao grau nodal (node degree) representado pelo tamanho dos vértices, no Programa Ciências Tecnologia e Gestão Aplicada à Regeneração Tecidual, identifica-se um pesquisador com elevada centralidade de grau (*degree centrality*), caracterizando uma topologia na qual esse nó central concentra a maior parte das conexões enquanto os demais vértices periféricos apresentam baixo grau nodal e reduzida conectividade entre si.

Figura 7 - Ciências Tecnologia e Gestão Aplicada a Regeneração Tecidual

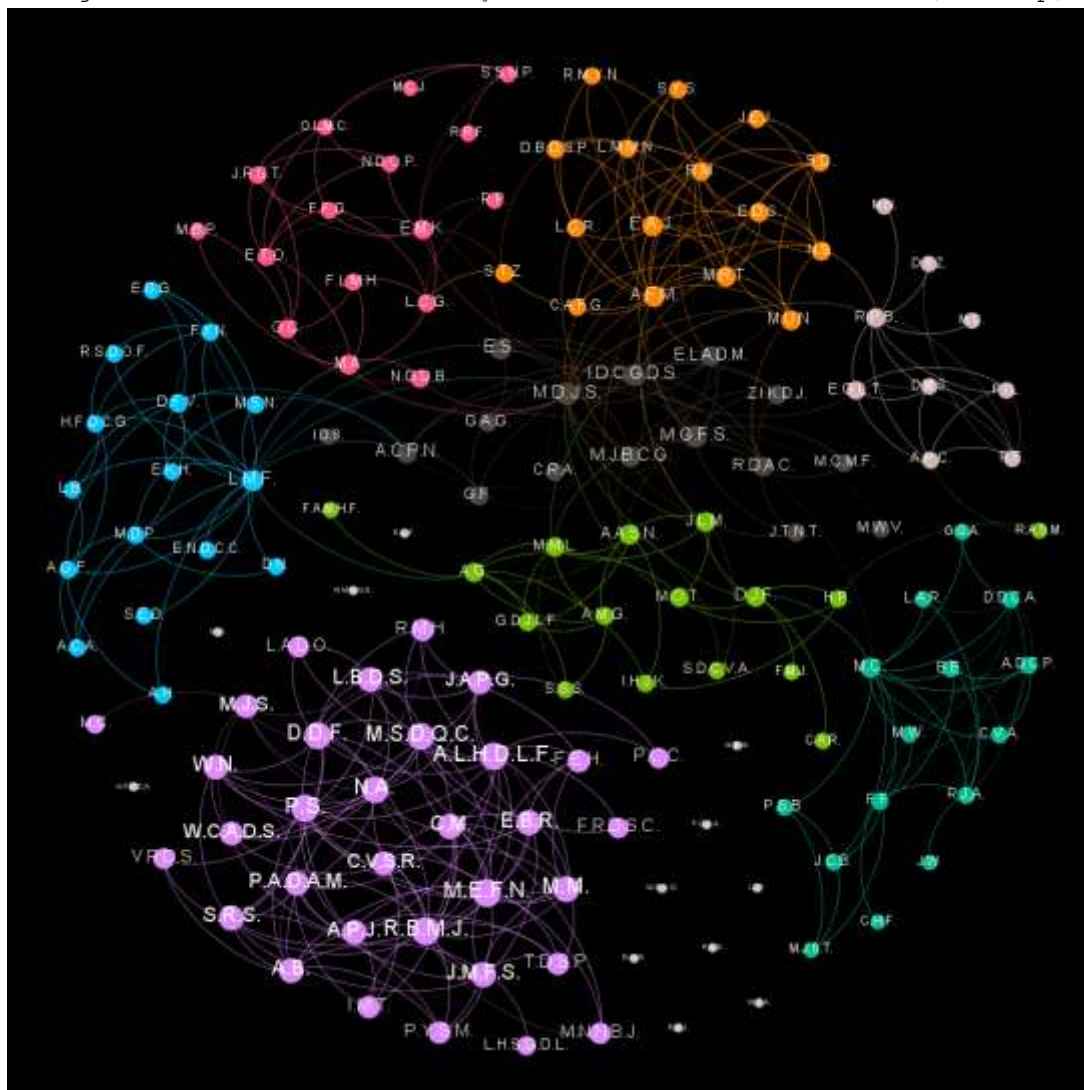


Por outro lado, L.M.F. pode evidenciar uma assimetria de centralidade na rede de coautoria do Programa. A concentração das colaborações em um único pesquisador, que é característica de redes com elevada heterogeneidade de grau (*degree heterogeneity*), fragiliza a capilaridade da produção científica coletiva, o que pode comprometer a avaliação do Programa frente aos critérios como consolidação dos grupos de pesquisa e diversidade da produção colaborativa. Considerando que Programas de Pós-Graduação são estruturados em múltiplas linhas de pesquisa com distintas especialidades temáticas, verifica-se que a centralização das coautorias em um único pesquisador pode indicar endogenia colaborativa, situação em que as parcerias científicas não emergem da convergência temática entre as linhas, mas da influência gravitacional do nó hub sobre a rede, comprometendo a autonomia produtiva dos demais pesquisadores e a sustentabilidade da colaboração científica no longo prazo.

Entretanto, talvez dentro desse Programa essa questão seja justa. Os questionamentos acima devem ser estudados dentro do universo de pesquisadores, mas a representação gráfica proposta dá poder a uma análise aprofundada baseada na inter-relação visual da rede.

Como última análise, iremos observar como estão organizados os pesquisadores e seus pares de uma forma global. A Figura 8 apresenta o relacionamento de toda a comunidade de Medicina III da Universidade Federal de São Paulo - UNIFESP, com os dados coletados.

Figura 8 - Rede de Colaboração da Área de Medicina III (Unifesp).



O primeiro ponto a ser observado na Figura 8 é que a rede de coautoria evidencia elevada densidade de conexões intercomunidade, indicando forte acoplamento colaborativo entre as comunidades identificadas. Apenas uma minoria de pesquisadores apresenta baixo grau nodal (*node degree*), caracterizando-os como nós periféricos na rede intrainstitucional de coautoria da UNIFESP. Entretanto, as oito comunidades identificadas, conforme visualizado na Figura 8 e descritas nominalmente na Figura 9, demandam análise criteriosa pelos Programas de Pós-Graduação, uma vez que a fragmentação da rede em múltiplas comunidades pode comprometer critérios avaliativos da CAPES relativos à consolidação e à capilaridade da produção colaborativa.

Atenção especial ao fato de que a rede representada na Figura 8 não apresenta componentes desconexos, evidenciando que as comunidades identificadas estão estruturalmente integradas por arestas intercomunidade. Uma vez que a interdisciplinaridade constitui diretriz central das políticas de avaliação da pós-graduação no Brasil (CAPES, 2020), esse padrão estrutural da rede está alinhado às diretrizes institucionais de fomento à colaboração interárea. Há expressiva densidade de arestas intercomunidade, o que sugere a interdisciplinaridade dos pesquisadores e aponta para o

potencial de geração de resultados de pesquisas com características translacionais

Figura 9 - Medicina III organizada em comunidades de relacionamento.

Verde	Ciência Cirúrgica Interdisciplinar
Verde Musgo	Ciências da Saúde Aplicada ao Esporte e à atividade física
Cian	Ciências Tecnologia, Gestão Aplicada à Regeneração Tecidual
Cinza	Medicina (Ginecologia)
Laranja	Medicina (Obstetria)
Vermelho	Medicina (Otorrino)
Grená	Medicina Urologia
Rosa	Oftalmologia e Ciências Visuais

A análise da rede global revelou que os Programas de Cirurgia Translacional e Tecnologia, Gestão e Saúde Ocular foram atraídos para outras comunidades, evidenciando um elevado grau de relacionamento de seus integrantes com pesquisadores externos a seus próprios grupos. Esse fator de atratividade interprogramas é relevante, pois, como apontam Maia, Zanotto e Caregnato (2011), redes científicas dinâmicas tendem a romper fronteiras disciplinares e gerar novos conhecimentos a partir da interação entre áreas distintas. Essa inferência não seria possível a partir da análise isolada dos programas, já que, quando observados individualmente, eles se apresentam como estruturas fechadas. A integração em rede, portanto, confirma o que Balancieri (2004) descreve como potencial de inovação pela colaboração.

No caso da Oftalmologia e Ciências Visuais (cor rosa), observou-se um raio do vértice ligeiramente maior em relação ao grupo de Medicina Obstetria (cor laranja). Embora ambos apresentem elevado grau de relacionamento e equilíbrio interno, o maior número de integrantes no programa de Oftalmologia explica sua maior projeção gráfica. Esse resultado está em consonância com Francisco (2018), ao destacar que redes com maior diversidade de atores possuem maior capacidade de articulação e resiliência.

De forma geral, os demais grupos da Medicina III demonstraram homogeneidade nas inter-relações, com poucos pesquisadores em situação de isolamento. Essa característica é positiva, pois reforça a ideia de redes normalizadas, nas quais não há concentração excessiva em poucos atores, mas sim uma distribuição mais equilibrada de colaborações (FRANCISCO, 2018).

Do ponto de vista prático, os resultados obtidos oferecem subsídios estratégicos aos Programas de Pós-Graduação. A coordenação de cursos pode utilizar essas análises para validar e estimular colaborações, identificar pesquisadores-chave e propor ações específicas para integrar aqueles que permanecem à margem da rede. Como defendem Neves (2020) e a CAPES (2020), a avaliação da pós-graduação deve ir além da mensuração de publicações, incorporando métricas de colaboração e interdisciplinaridade como indicadores de qualidade e sustentabilidade acadêmica.

Assim, a análise das redes de colaboração permite não apenas descrever o panorama atual da Medicina III, mas também apontar caminhos de fortalecimento. Estimular pesquisadores com maior receptividade à cooperação, promover o engajamento dos que permanecem isolados e compreender as razões do destaque de determinados atores são ações que

podem ampliar a produção científica e alinhar os programas às exigências contemporâneas de interdisciplinaridade e impacto social.

5. DISCUSSÃO

A predominância do grupo $G(2,5)$, com 38,26% dos pesquisadores, é consistente com o padrão descrito na literatura como típico de redes científicas: a maioria dos pesquisadores mantém um número moderado de colaborações, enquanto poucos atores concentram conexões mais amplas. Essa distribuição assimétrica, observada também nos programas de Medicina (Urologia) e Ciências, Tecnologia e Gestão Aplicadas à Regeneração Tecidual, é característica de redes com escala livre, nas quais nós centrais exercem papel estrutural relevante.

Por outro lado, o percentual de 10,74% de pesquisadores sem qualquer colaboração interna merece atenção analítica. O isolamento em redes acadêmicas pode ter origens diversas, por exemplo, a especialização temática, e não deve ser interpretado de forma determinística como indicador de desempenho. A ferramenta permite identificar esses casos, mas a interpretação de suas causas e implicações requer análise contextual por parte dos gestores dos programas.

A análise global da área de Medicina III revelou um padrão de interdisciplinaridade interprogramas, com pesquisadores de diferentes programas compondo comunidades comuns. Esse resultado é descrito como uma característica de redes científicas dinâmicas, nas quais a colaboração tende a romper fronteiras disciplinares. A visibilidade desse padrão é possível apenas pela análise conjunta dos programas, e não pela consulta individual ao Lattes exemplifica o valor analítico da abordagem proposta pelo *LattesXtractor*.

6. CONCLUSÃO

A aplicação do método adotado nesta pesquisa resultou no desenvolvimento do *LattesXtractor*, ferramenta capaz de transformar dados do Currículo Lattes em representações estruturadas e exportáveis para análise de redes de colaboração acadêmica. Ao automatizar a extração e organização dessas informações, a ferramenta amplia as possibilidades analíticas em relação às abordagens baseadas exclusivamente em tabelas e indicadores numéricos, sem, contudo, substituir a avaliação qualitativa mais ampla que o contexto de cada programa exige.

A visualização nativa de redes do Lattes é estática e não exportável para análise. O *LattesXtractor* gera arquivos XML compatíveis com o *Gephi* (Bastian; Heymann; Jacomy, 2009), abrindo a possibilidade de aplicar algoritmos de análise de redes entre eles: (i) detecção de comunidades, cujo resultado é visualizado pelos grupos coloridos na figura, obtidos por algoritmos como Louvain ou Girvan-Newman; (ii) métricas de centralidade (*degree*, *betweenness* e *closeness centrality*); e (iii) análise estrutural de redes de coautoria e colaboração científica. Além disso, o *LattesXtractor* constitui ponto de partida para a geração de esquemas de visualização cientométrica personalizados, produzidos por meio de algoritmos de layout especializados, como Force Atlas 2 e Fruchterman-Reingold, disponíveis no próprio *Gephi*.

A nova forma de visualização proporcionada pelo *LattesXtractor*, em integração com o *Gephi*, permitiu observar de maneira imediata a posição

relativa dos pesquisadores nas redes, destacando centralidade, presença de “nós estratégicos”, comunidades bem definidas e grupos fragmentados. Isso tornou mais evidente quais pesquisadores atuam como articuladores entre diferentes comunidades científicas, quais permanecem isolados e como se distribuem os graus de colaboração dentro e entre os programas da área de Medicina III. Elementos que, em quadros numéricos, exigiriam leituras demoradas e comparações manuais, passam a ser percebidos rapidamente por meio da análise visual.

No estudo de caso, a visualização revelou programas com perfis distintos de organização interna: alguns com redes mais equilibradas, sem concentração excessiva em poucos atores, e outros com pesquisadores centrais de alta conectividade. Esses padrões oferecem insumos para reflexão dos coordenadores de programas, embora sua interpretação deva considerar o contexto específico de cada área e não seja possível, com base apenas na estrutura visual das redes, estabelecer relações causais com o desempenho nas avaliações da CAPES.

Cabe destacar as principais limitações do presente estudo. A análise está restrita aos docentes-orientadores da área de Medicina III da UNIFESP, o que limita a generalização dos resultados para outros contextos institucionais ou áreas do conhecimento. Além disso, a qualidade dos dados depende diretamente da atualização dos Currículos Lattes pelos próprios pesquisadores, sendo que inconsistências cadastrais ou currículos desatualizados podem influenciar os padrões identificados nas redes. Por fim, as relações de coautoria registradas no Lattes representam apenas uma dimensão da colaboração científica, não contemplando outras formas de parceria acadêmica não formalizadas em publicações conjuntas.

Dessa forma, a principal contribuição não reside apenas na automatização da extração, mas na qualificação da leitura dos dados, ao oferecer uma perspectiva visual que facilita diagnósticos, comparações entre programas e identificação de oportunidades de fortalecimento das redes de colaboração. Os resultados indicam que a combinação entre extração estruturada de dados do Lattes e análise de redes constitui uma abordagem promissora para o apoio à gestão acadêmica, com potencial para refinamentos futuros que ampliem sua aplicabilidade a outras áreas e instituições.

REFERÊNCIAS

BALANCIERI, R. **Análise de redes de pesquisa em uma plataforma de gestão em ciência e tecnologia: uma aplicação à Plataforma Lattes**. 2004. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2004.

BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: an open source software for exploring and manipulating networks. In: **INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA (ICWSM)**, 3., 2009. Proceedings... [S. l.]: AAAI, 2009. v. 3, n. 1, p. 361-362. Disponível em: <https://gephi.org/publications/gephi-bastian-feb09.pdf>. Acesso em: 19 ago. 2025.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR. **Portal da CAPES**. Brasília, DF, 2020. Disponível em: <http://www.capes.gov.br>. Acesso em: 19 ago. 2025.

DEITEL, P.; DEITEL, H. **Java: como programar**. 10. ed. São Paulo: Pearson Education do Brasil, 2016.

DIGIAMPIETRI, L. A.; MENA-CHALCO, J. P.; PÉREZ-ALCÁZAR, J. J.; TUESTA, E. F.; DELGADO, K. V.; MUGNAINI, R.; SILVA, G. S. Minerando e caracterizando dados de Currículos Lattes. In: **BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING (BRASNAM)**, 1., 2012, Curitiba. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2012. p. 117-128. ISSN 2595-6094. Disponível em: <https://sol.sbc.org.br/index.php/brasnam/article/view/6868>. Acesso em: 19 ago. 2025.

FRANCISCO, R. P. **Gestão de redes de colaboração: conceitos e aplicações**. Campinas: Facunicamps, 2018.

LIMA, J. C.; CARVALHO, C. L. **Ontologias - OWL (Web Ontology Language)**. Goiânia: Universidade Federal de Goiás (UFG), 2005.

LIMA, J. V. V. et al. Metodologias ativas como forma de reduzir os desafios do ensino em Engenharia de Software: diagnóstico de um survey. In: **SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE)**, 31., 2020, Porto Alegre. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 172-181.

MAIA, F. S.; ZANOTTO, S. R.; CAREGNATO, S. E. Colaboração científica e análise de redes sociais. **Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 30, n. 1, p. 43-55, 2011.

MEDDAGE, D. P. P. et al. Explainable machine learning (XML) to predict external wind pressure of a low-rise building in urban-like settings. **Journal of Wind Engineering and Industrial Aerodynamics**, v. 226, p. 105027, 2022. DOI: <https://doi.org/10.1016/j.jweia.2022.105027>.

MENA-CHALCO, J. P.; CESAR JUNIOR, R. M. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, v. 15, n. 4, p. 31-39, dez. 2009. DOI: <https://doi.org/10.1007/BF03194511>.

NEVES, A. A. B. A pós-graduação no Brasil. **International Journal of Business and Marketing**, v. 5, n. 2, p. 23-29, 2020.