

ANÁLISE DA GENEALOGIA ACADÊMICA DO CORPO DOCENTE DE INSTITUIÇÕES DE ENSINO SUPERIOR E SEUS RESPECTIVOS IMPACTOS

Paulo Vinicius Araujo Feitosa¹, Luciano Rossi²
^{1,2} Ciência da Computação, Centro Universitário FEI
¹ unifpfeitosa@fei.edu.br
² lrossi@fei.edu.br

Resumo: O presente trabalho apresenta um método para a análise da genealogia acadêmica do corpo docente de instituições de ensino superior, centrado em quatro etapas principais: extração e pré-processamento de dados de Currículos Lattes por meio de técnicas de *web scraping*; estruturação dos registros em um modelo de dados orientado a grafos; construção de grafos acadêmicos que representam as relações de orientação entre docentes e discentes; e geração de metagrafos capazes de revelar padrões coletivos de influência e interdependência entre áreas de conhecimento. Essa abordagem metodológica visa oferecer uma ferramenta robusta para identificar dinâmicas de transmissão de saberes e subsidiar estratégias institucionais voltadas à promoção da diversidade epistemológica e à renovação intelectual.

1. Introdução

A genealogia acadêmica consiste em um mapeamento sistemático das relações de orientação entre docentes e discentes, permitindo visualizar a transmissão de teorias, métodos e práticas de pesquisa ao longo das gerações. Esse levantamento possibilita identificar a consolidação de escolas de pensamento, o estabelecimento de redes de colaboração que transcendem instituições e fronteiras nacionais, bem como os perfis de investigação que moldam a formação intelectual de novos pesquisadores.

O método proposto fundamenta-se na extração e no processamento de dados provenientes de um repositório estruturado que reúne informações sobre pesquisadores e suas orientações concluídas. A abordagem metodológica inclui quatro etapas principais: obtenção e pré-processamento dos registros disponíveis; modelagem das relações de orientação entre pesquisadores e seus orientandos de doutorado; construção de grafos acadêmicos para representar esses vínculos; e geração de metagrafos capazes de revelar padrões coletivos de orientação e áreas de concentração epistemológica.

Como estudo de caso, utiliza-se a própria Plataforma Acácia [1] aplicada ao Centro Universitário FEI, na qual serão selecionados e analisados os currículos do corpo docente da instituição privada de ensino superior. Essa aplicação ilustrará os procedimentos de coleta, organização e visualização dos dados.

1.1. Objetivo

Diante deste contexto, o presente estudo tem como objetivo principal analisar as relações acadêmicas a nível de doutorado entre os docentes, seus ascendentes

(orientadores de doutorado) e seus descendentes (orientados de doutorado), por meio da estruturação e análise de dados acadêmicos. O estudo será desenvolvido por meio de um método composto por quatro etapas: extração dos dados, estruturação dos dados, construção dos grafos dos pesquisadores e, por fim, geração do metagrafo das áreas de doutorado. A coleta dos dados será realizada por meio de técnicas de *web scraping*, com a extração automática das relações acadêmicas a partir de repositórios que disponibilizam este tipo de informação. A partir desses dados, será possível identificar e analisar as conexões entre os pesquisadores, tendo como principal atributo investigativo a grande área do conhecimento à qual pertencem seus trabalhos de doutorado.

2. Estudo de Caso

A Plataforma Acácia foi concebida pelo Grupo de Pesquisa em Cientometria da UFABC com o propósito de documentar as relações formais de orientação em programas de pós-graduação brasileiros, utilizando dados públicos disponíveis na Plataforma Acácia. Além da visualização de redes de ascendência e descendência acadêmica, a plataforma oferece métricas como fecundidade, descendência, índice genealógico e geração [4].

Neste estudo de caso, a Plataforma Acácia será utilizada como banco de dados para alimentar o método de construção da genealogia acadêmica. A coleta de registros do corpo docente será realizada por meio de técnicas automatizadas de *web scraping*, extraindo sistematicamente dados de orientação (doutorado) disponíveis na plataforma. Tais dados serão estruturados para inserção em um banco de grafos Neo4j, com representação fiel das conexões entre orientadores e orientandos. A Acácia serve, portanto, como fonte primária e já organizada de vínculos acadêmicos, viabilizando a posterior geração de grafos e metagrafos para análise de influência e interdependência entre áreas do conhecimento.

3. Metodologia

A metodologia adotada para o desenvolvimento deste projeto, compreende duas etapas principais, primeiro, foi feita a extração dos dados acadêmicos dos docentes com doutorado do Centro Universitário FEI, além de seus orientadores do doutorado, e, se houver, também será extraído os seus orientados de doutorado, e, por último, foi feita o estudo das áreas de conhecimento de cada pesquisador, analisando quais áreas de conhecimento tem mais

influência, ou seja, o estudo de quais áreas de conhecimento mais orientam e quais são mais orientadas.

Para a realização deste projeto, foram utilizadas os seguintes componentes:

a) Plataforma Acácia: Utilizada como banco de dados para a extração das informações dos pesquisadores com doutorado; b) *Python*: Linguagem de programação utilizada para a extração e análise dos dados; c) *Playwright*: Biblioteca principal utilizada para o *web scraping*; d) Neo4j: Banco de dados utilizado para armazenar os pesquisadores, as relações de orientações e a influência das áreas. e) *Gephi*: Para uma melhor visualização dos grafos e metagrafos, foi utilizado este programa para o estudo deste trabalho.

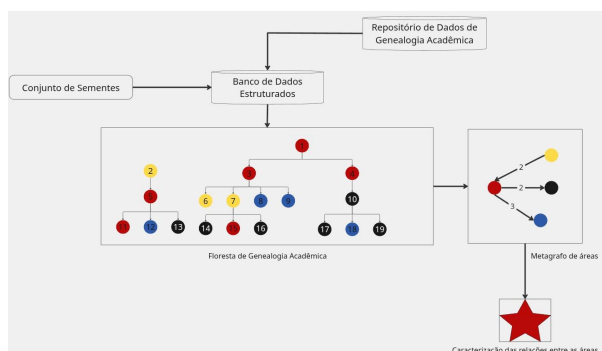


Figura 1: Fluxograma da Iniciação Científica

Com esses componentes, o projeto inicia-se pela extração dos pesquisadores disponíveis na Plataforma Acácia, abrangendo tanto os pesquisadores quanto seus orientadores, por meio do *framework Playwright*; finalmente, procede-se à construção de um metagrafo que evidencia as áreas de conhecimento mais influentes e aquelas mais influenciadas, possibilitando uma análise estruturada das relações acadêmicas. Para melhor visualização do caminho utilizado neste trabalho, o fluxograma abaixo mostra o passo desta iniciação científica de forma genérica, para que possa ser utilizado em outros trabalhos.

A floresta da genealogia acadêmica é a visualização dos pesquisadores doutores, seus ascendentes, descendentes e área de doutorado. Para resumo, será explicado a família do nó 2. Este nó representa um pesquisador com doutorado na área do conhecimento amarela e este pesquisador foi orientador de doutorado do pesquisador 5, que teve seu trabalho na área vermelha. O pesquisador participou como orientador de três doutorados, do pesquisador 11, 12 e 13, sendo respectivamente da área do conhecimento vermelha, azul e preta.

Com a floresta destes pseudos-pesquisadores feita, podemos entender o metagrafo das áreas, que tem as quatro áreas de conhecimento da floresta, em que

tivemos dois pesquisadores da área amarela orientando pesquisadores com o doutorado da área vermelha. Também é possível analisar que tivemos pesquisadores da área vermelha que orientaram um doutorado da área preta duas vezes, e da área azul três vezes.

Por conseguinte, é feita a caracterização das relações entre as áreas, ou seja, a análise deste metagrafo criado a partir da floresta da genealogia acadêmica.

3.1. Web Scraping

Para automatizar a coleta de dados, adotou-se a técnica de *web scraping*, que consiste na extração programática de informações a partir da interface web de um site. Através de *frameworks*, como *Playwright*, variáveis críticas como orientadores, orientações e ano do doutorado são extraídas após o carregamento do site. O processo de coleta e inserção no banco de dados foi separada em dois diferentes códigos, um para cada ação.

Primeiramente para a coleta dos pesquisadores, foi extraído do próprio site da instituição do estudo de caso [2] o nome completo de todos os doutores participantes do corpo docente.

Com a lista dos doutores pronta, foi feita uma investigação dentro da plataforma para uma facilitação da automatização, e, assim, foi descoberto que é possível fazer a pesquisa automática dos docentes pelo próprio URL. A plataforma utiliza o nome dos pesquisadores dentro da URL (por exemplo: <https://plataforma-acacia.org/profile/luciano-rossi/>), com esta informação, somente foi necessário formatar a lista com o nome dos pesquisadores (Luciano Rossi para luciano-rossi) e usar a URL base (<https://plataforma-acacia.org/profile/>) para a pesquisa automatizada.

```
def normalizar_nome(nome: str) -> str:
    nome = nome.lower()
    nome = nome.replace(" ", "-")
    return nome

def processar_arquivo(entrada: str, saida: str):
    with open(entrada, "r", encoding="utf-8") as f:
        nomes = f.readlines()
        nomes_formatados=[normalizar_nome(n.strip())
        ]) for n in nomes if n.strip()]
    with open(saida, "w", encoding="utf-8") as f:
        f.write("\n".join(nomes_formatados))

processar_arquivo("docentes-fei.list", "docentes-fei-formatados.list")
```

Este código foi feito para formatar os nomes dos pesquisadores para ser utilizado na URL da plataforma. Com a lista de nomes formatados para a URL, o trabalho pode ir ao passo de automatização.

```
def extrai_pesquisador(nome, page, grupo):
    nome_chave = nome_formatado(nome)
    if nome_chave in pesquisadores_dict:
        return
    url = base_url + nome
    tentativas = 0
```

```

max_tentativas = 5
while tentativas < max_tentativas:
    try:
        page.goto(url, timeout=20000)
        html = page.content()
        if "<h1>Not Found</h1>" in html or "The
requested resource was not found on this server." in
html:
            logger.warning(f'Pesquisador '{nome_chave}'
não encontrado na plataforma. Pulando.')
            return
        if "503" in html or "Service Temporarily
Unavailable" in html:
            raise Exception("Erro 503 detectado no
HTML")
        area, universidade, ascendentes, descendentes =
extraí_info(html)
        if area or universidade or ascendentes or
descendentes:
            pesquisadores_dict[nome_chave] = {
                "grupo": grupo,
                "grande-area": area,
                "universidade": universidade,
                "ascendentes":
                    nome_formatado(ascendentes[0]) if ascendentes else "",
                "descendentes": [nome_formatado(d) for d
in descendentes]
            }
            logger.info(f'Extraído: {nome_chave}
({grupo}) - universidade: {universidade} - ascendentes:
{ascendentes} - descendentes: {descendentes}')
            break
    except Exception as e:
        tentativas += 1
        logger.warning(f'Tentativa {tentativas} falhou
para {nome_chave}: {e}')
        if tentativas < max_tentativas:
            logger.info(f'Recarregando página para
{nome_chave} em 5 segundos...')
            time.sleep(5)
        time.sleep(6)

```

A função `extraí_pesquisador` foi o centro da extração. Esta função teve a responsabilidade de: verificar se a página do docente existe; verificar se teve um erro no site e, caso tiver, fazer até mais quatro tentativas com pausas; se a página do docente existir e não tivesse falhas, invoca a função `extraí_info`, que extraí as informações do pesquisador; armazena as informações extraídas (área, universidade, ascendentes e descendentes) num dicionário; e fazer logs detalhados de cada etapa. Nas tentativas manuais de entrada ao site, foi visto que o servidor retornava *Service Temporarily Unavailable*, e, após um intervalo de tempo, a página carregava normalmente, por esta razão foi feito os intervalos de seis segundos para cada docente extraído e cinco segundos caso houvesse o retorno de erro na página.

Como tido anteriormente, a função `extraí_info` tem a responsabilidade de extrair os dados dos doutores (área, universidade, ascendentes e descendentes) e retornar uma tupla com estes dados já tratados para evitar quebras de linha ou espaços extras.

```

def extraí_info(html):
    soup = BeautifulSoup(html, "html.parser")
    # Extraí área
    h3 = soup.find("h3", class_="subtitle is-size-6 mb-0")
    if h3:
        area = " ".join(h3.stripped_strings)
        area = area.replace("\n", " ").replace("\r", " ")
        area = " / ".join([part.strip() for part in
area.split("/")])
        area = " ".join(area.split())
    else:
        area = ""
    # Universidade
    universidade = ""
    span_uni = soup.find("span", itemprop="name",
class_="is-size-6")
    if span_uni:
        universidade = span_uni.text.strip()
    # Ascendentes
    ascendentes = []
    asc_table = soup.find("table",
id="table-profile-ascendants")
    if asc_table:
        for tr in asc_table.find("tbody").find_all("tr"):
            tags = tr.find_all("span", class_="tag is-d")
            if tags:
                nome_asc = tr.find("span", itemprop="name")
                if nome_asc:
                    ascendentes.append(nome_asc.text.strip())
    # Descendentes
    descendentes = []
    desc_table = soup.find("table",
id="table-profile-descendants")
    if desc_table:
        for tr in desc_table.find("tbody").find_all("tr"):
            tags = tr.find_all("span", class_="tag is-d")
            if tags:
                nome_desc = tr.find("span",
itemprop="name")
                if nome_desc:

```

```

descendentes.append(nome_desc.text.strip())

```

```

return area, universidade, ascendentes, descendentes

```

A função `extraí_pesquisador` também invoca a função `nome_formatado`, que foi feita para formatar o nome dos pesquisadores para facilitar a análise posterior (luciano-rossi para Luciano Rossi). Tendo consciência deste processo, podemos verificar o fluxo principal do código.

```

with sync_playwright() as p:
    browser = p.chromium.launch(headless=True)
    context = browser.new_context()
    page = context.new_page()

```

```

for nome in nomes:

```

```

    # Extraí pivô
    extraí_pesquisador(nome, page, grupo="pivo")
    nome_chave = nome_formatado(nome)
    if nome_chave not in pesquisadores_dict:
        continue # Não achou o pivô, pula
    # Extraí pai (ascendente)

```

```

                                asc      =
pesquisadores_dict[nome_chave]["ascendentes"]
    if asc:
        asc_url = asc.replace(" ", "-").lower()
        extrai_pesquisador(asc_url, page, grupo="pai")
        # Se o pai existir, extrai seus descendentes como
filhos
        pai_chave = nome_formatado(asc)
        if pai_chave in pesquisadores_dict:
            for desc in
pesquisadores_dict[pai_chave]["descendentes"]:
                if desc:
                    desc_url = desc.replace(" ", "-").lower()
                    extrai_pesquisador(desc_url, page,
grupo="filho")

            # Extrai filhos (descendentes) do pivô
normalmente
            for desc in
pesquisadores_dict[nome_chave]["descendentes"]:
                if desc:
                    desc_url = desc.replace(" ", "-").lower()
                    extrai_pesquisador(desc_url, page,
grupo="filho")

    browser.close()

with open(saida_json, "w", encoding="utf-8") as f:
    json.dump(pesquisadores_dict, f, ensure_ascii=False,
indent=2)

```

Este fluxo utiliza o *framework Playwright* para inicializar o *Chromium* em modo *headless*, para que o processo funcione sem que abra as páginas do navegador visualmente. Este fluxo invoca a função `extrai_pesquisador` para fazer o processo explicado anteriormente, e, caso encontre a informação de ascendente e/ou descendente no docente, eles também são extraídos.

Estes dados coletados por estas funções foram armazenados num arquivo `.json` com a seguinte formatação:

```

{
  "Nome do pesquisador": {
    "grupo": "pivo/filho/pai",
    "grande-area": "grande area / area",
    "universidade": "Nome da universidade de seu
doutorado",
    "ascendentes": "Nome do orientador (pai)",
    "descendentes": [
      "Nome do orientado (filho)"
    ]
  }
}

```

3.3. Neo4j

Com todas as informações dos pesquisadores e suas relações coletadas e armazenadas no arquivo `.json`, o próximo passo é colocá-las em um banco de dados. O banco de dados escolhido foi o Neo4j por ter um armazenamento dos dados em grafos, o que facilita a

análise das relações, em que cada área é um nó, e as arestas são as suas relações, que, neste estudo, será a influência.

Neste passo, foi utilizado o arquivo JSON com os dados dos pesquisadores para inseri-los no Neo4j com o segundo código escrito em Python. Este código foi responsável para inserir os pesquisadores, as suas relações (ascendente e descendente), e também pelas áreas e suas relações (influência). No próximo capítulo será discutido sobre como foi feita a visualização destes pesquisadores e áreas no Neo4j, e como analisar estes dados.

Para que cada pesquisador seja criado no Neo4j com as suas relações corretas foi feito o seguinte código:

```

with driver.session() as session:
    # Cria vértices
    for nome, dados in pesquisadores_dict.items():
        grande_area = dados.get("grande-area", "")
        universidade = dados.get("universidade", "")
        session.write_transaction(create_pesquisador,
nome, grande_area, universidade)

    # Cria relações ascendentes
    for nome, dados in pesquisadores_dict.items():
        asc = dados.get("ascendentes", "")
        if asc and asc in pesquisadores_dict:
            session.write_transaction(cria_relacao, nome,
asc, "ASCENDENTE")
        # Cria relações descendentes
        for desc in dados.get("descendentes", []):
            if desc and desc in pesquisadores_dict:
                session.write_transaction(cria_relacao, nome,
desc, "DESCENDENTE")

```

Este código lê o dicionário criado anteriormente pela função `extrai_pesquisador` e cria o nó do pesquisador com os seus respectivos atributos (grande área, nome, universidade), depois cria as arestas, que representa o vínculo de ascendência e descendência deste pesquisador com outros pesquisadores. Segue abaixo a *query* para criar o nó do pesquisador e a *query* para criar as arestas de relações:

```

MERGE (p:Pesquisador {nome: $nome})
SET p.grandeArea = $grande_area,
p.universidade = $universidade;
MATCH (a:Pesquisador {nome: $nome1})
MATCH (b:Pesquisador {nome: $nome2})
MERGE (a)-[:relacao.upper()]->(b);

```

Desse modo, foi concluído a extração dos doutores do corpo docente do Centro Universitário FEI e a inserção dos mesmos no banco de dados Neo4j. No próximo capítulo será feito o estudo do resultado deste trabalho.

4. Resultados e Discussão

Neste capítulo será abordado sobre os resultados do caso de uso do Centro Universitário FEI, e também para uma verificação e comparação de casos de uso, será também abordado as áreas de doutorado do

corpo docente do Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP).

4.1. Centro Universitário FEI

Após oito horas do código em execução, foram processados todos os docentes doutores do Centro Universitário FEI, seus orientadores e orientandos registrados na Plataforma Acácia, resultando um total de 1472 pesquisadores, como entrada, tivemos 214 docentes doutores. Com estes dados, foi criado vários grafos destes pesquisadores com relação de ascendência e/ou descendência, sendo 1370 vértices de pesquisadores e 2112 arestas de ascendentes e descendentes. Todos os grafos tem como pivô um pesquisador com doutorado que faz parte do corpo docente do Centro Universitário FEI. Também foi feito os grafos das áreas e suas influências, com um total de 65 áreas e 179 relações de influência. Para visualizar os pesquisadores e suas relações no Neo4j foi feito o seguinte código abaixo com a linguagem Cypher.

```
MATCH p=()-[:DESCENDENTE]->()
MATCH f=()-[:ASCENDENTE]->()
RETURN p, f;
```

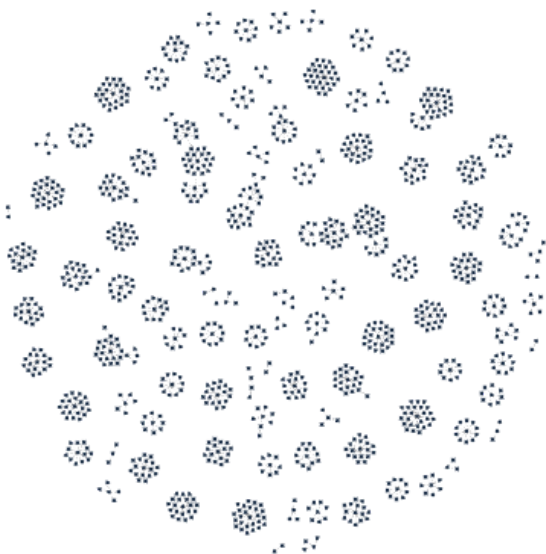


Figura 2: Grafos de relação de ascendente e descendente com os docentes doutores do Centro Universitário FEI.

Para fins de melhor compreensão e aprofundamento no estudo deste trabalho adiante, utilizaremos como exemplo o pesquisador pivô Cássio Leandro Dal Ri Barbosa, um docente do setor de Física dentro do Centro Universitário FEI.

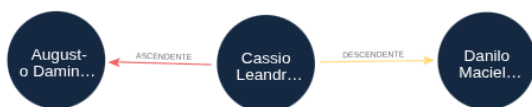


Figura 3: Gráfico com o pivô Cássio Leandro Dal Ri Barbosa e sua família acadêmica de primeira geração.

Neste grafo de exemplo, podemos analisar que o docente pivô tem somente um descendente, o pesquisador Danilo Maciel Lopes Gusmão, e seu ascendente, Augusto Damineli Neto, não tem nenhum outro descendente.

Alguns pesquisadores não tem o registro da grande área de seu doutorado e em qual instituição foi feita dentro da Plataforma Acácia, podendo ser uma consequência do pesquisador não informar estes dados em seu currículo Lattes, onde foi coletado as informações da Plataforma Acácia.

A análise do grafo em Neo4j mostrou que alguns poucos orientadores apresentam alto grau de saídas (orientandos), configurando-se como “hubs” de orientação, enquanto a maioria possui poucos ou nenhum descendente imediato. Ao agrupar as relações por grande área, verificou-se que as Engenharias orientam majoritariamente dentro da própria área, com pouca transferência de doutores para áreas distintas, refletindo a estrutura hierárquica rígida das áreas de conhecimento. Entretanto, surgiram casos isolados de docentes formados em Ciências Exatas que orientam doutorandos em áreas de Engenharia, indicando pontes interdisciplinares.

Com os pesquisadores e suas relações analisados, podemos passar para o estudo das áreas destes pesquisadores e suas influências. Assim, podemos visualizar todas as áreas e suas relações no Gephi, assim, temos um metagrafo para análise.

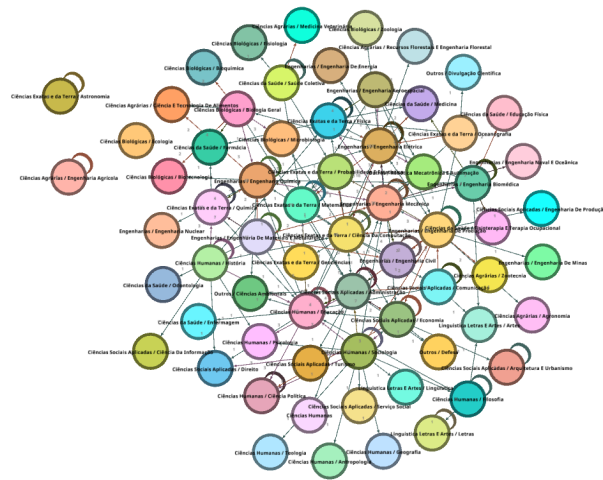


Figura 4: Metagrafo de todas as áreas e suas relações do Centro Universitário FEI.

Na análise do metagrafo das áreas, podemos visualizar com mais precisão a dominância da grande área de Engenharias em que, das 65 áreas analisadas, 14 tem a grande área como Engenharias. Aprofundando mais nesta grande área, podemos ver que dos 1472 pesquisadores processados pelo sistema, 577 deles fizeram o seu doutorado na grande área de Engenharias, que pode ser uma consequência da instituição em análise ter um grande foco nos cursos superiores de Engenharia. Como exemplo, vamos analisar a área com mais influência destes grafos, a Engenharia Elétrica.

Para analisar somente a Engenharia Elétrica e suas influências, foi feita a visualização pelo próprio Neo4j e foi utilizado o seguinte código no Neo4j:

```
MATCH p = (e:Area {nome: 'Engenharias / Engenharia Elétrica'})-[:INFLUENCIA*1..1]->(a:Area)
RETURN p;
```

Assim, o Neo4j nos retorna o metagrafo com as áreas que são influenciadas e/ou influenciam a área de Engenharia Elétrica, além da própria.

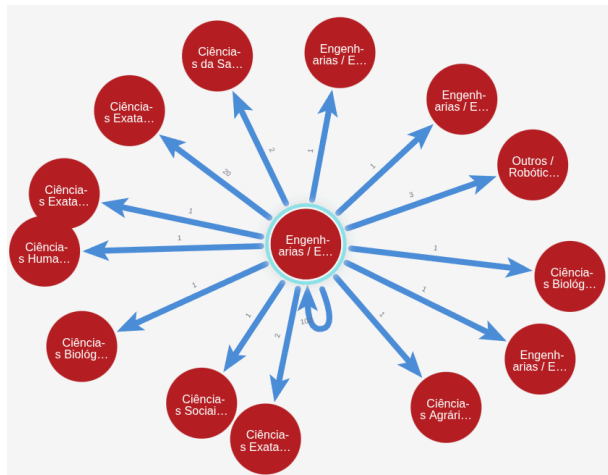


Figura 5: Metagrafo de Engenharia Elétrica e as áreas influenciadas

Neste metagrafo, podemos ver que a Engenharia Elétrica influencia principalmente as áreas da própria Engenharia, mas também influencia outras grandes áreas, como Ciências Biológicas, Ciências Humanas, Ciências Exatas e da Terra, Ciências Sociais Aplicadas e Ciências Agrárias, sendo a área de Ciência da Computação a mais influenciada, com 20 pesquisadores influenciados, além da sua própria área, com 102 doutorados cujo tem o orientador e orientado de Engenharia Elétrica.

Na figura 4 também é possível ver que existem duas áreas isoladas, ou seja, não tem relação com outro nó, somente recebe influência e influência dentro de sua área, sendo as áreas de Engenharia Agrícola e Astronomia. Aprofundando mais na área de Astronomia, percebemos que o pesquisador docente Cassio Leandro Dal Ri Barbosa é o "culpado" desta área aparecer neste metagrafo.

A análise das relações entre grandes áreas de conhecimento evidencia a forte predominância das Engenharias como núcleo central de influência no corpo docente do Centro Universitário FEI. O maior número de ocorrências corresponde à relação de orientação intra-área dentro das Engenharias, revelando que os vínculos acadêmicos são majoritariamente endógenos, ou seja, docentes e orientandos são desenvolvidos e tendem a permanecer dentro da mesma grande área. Esse resultado corrobora a tradição institucional da FEI voltada para cursos de Engenharia, ao mesmo tempo em que aponta para uma relativa baixa mobilidade

interdisciplinar de doutores em comparação a Engenharia.

Outro dado expressivo é o comportamento das Ciências Exatas e da Terra, que aparecem tanto como influenciadoras de si mesmas (sete ocorrências) quanto como fonte de formação para as Engenharias (quatro ocorrências). Essa interação mostra a importância das Exatas como base epistemológica para áreas aplicadas, funcionando como elo estruturante do conhecimento científico. Ainda que em menor escala, há também um fluxo inverso, com Engenharias influenciando as Ciências Exatas e da Terra (2 ocorrências), sugerindo uma troca mútua, ainda que assimétrica, entre teoria fundamental e aplicação tecnológica.

As relações interdisciplinares mais periféricas, mas não menos relevantes, incluem a influência das Ciências Humanas sobre as Ciências Sociais Aplicadas (3 ocorrências) e sobre as Ciências Exatas e da Terra (1 ocorrência), indicando que há conexões conceituais e metodológicas que transbordam para além das áreas tecnológicas.

Por fim, observa-se a presença de áreas específicas, como Engenharia de Materiais e Metalúrgica, que aparece como área influenciada em apenas uma ocorrência, evidenciando sua menor centralidade no conjunto do corpo docente analisado. A influência das Ciências da Saúde sobre as Exatas também surge de maneira isolada, sugerindo potenciais, ainda que incipientes, de integração interdisciplinar.

Em termos acadêmicos, esse panorama aponta para uma estrutura de formação e orientação altamente concentrada nas Engenharias e nas Ciências Exatas, que se retroalimentam como pólos dominantes do corpo docente da FEI. As demais áreas, embora presentes, exercem influência restrita ou marginal. Tal concentração contribui para a consolidação de uma identidade institucional fortemente tecnológica, mas, ao mesmo tempo, coloca em evidência a necessidade de ampliar conexões interdisciplinares que favoreçam maior diversidade epistemológica e inovação científica.

4.2. Universidade de São Paulo

Neste estudo de caso adicional foi examinado a Instituição de Matemática e Estatística da Universidade de São Paulo (IME-USP). Foi escolhida esta faculdade por também abranger a área de exatas, porém não ter o foco na grande área de Engenharias, como o estudo de caso anterior. Este outro estudo de caso seguiu o mesmo fluxo que o estudo anterior. Como entrada, foi extraído do site do próprio instituto [3] e tivemos somente 40 pesquisadores com doutorado.

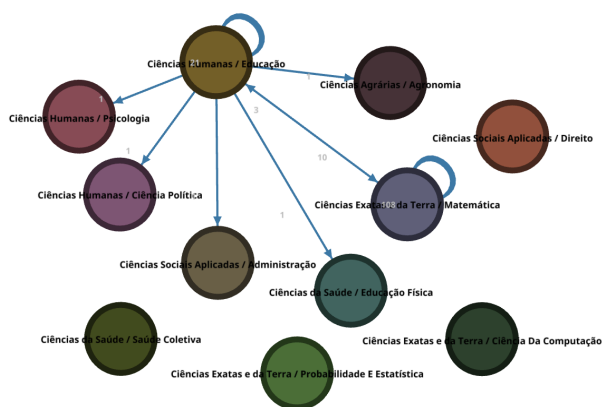


Figura 6: Metagrafo de todas as áreas e suas relações do IME-USP.

A análise dos pesquisadores extraídos da USP revela um perfil disciplinar fortemente concentrado nas Ciências Exatas e da Terra — em particular em Matemática —, a partir dos registros presentes no arquivo de origem, que contabiliza 292 pesquisadores no conjunto analisado, dos quais 185 ($\approx 63,4\%$) estão classificados em Matemática.

Esse predomínio configura a Matemática como o nó central da topologia relacional observada: há elevado número de orientações intra-área (108 casos em que orientador e orientando pertencem à mesma grande área Matemática), o que indica uma forte endogeneidade formativa — isto é, vínculos acadêmicos que tendem a permanecer dentro da mesma grande área.

Adicionalmente, o conjunto registra 62 entradas sem grande-área preenchida (aproximadamente 21,2% do total), hipótese explicativa para fluxos aparentes entre categorias não identificadas e a Matemática (por exemplo, 69 ocorrências classificadas como “*Unknown* → Ciências Exatas e da Terra / Matemática”), sugerindo que existem informações não preenchidas nas páginas dos pesquisadores na plataforma Acácia.

Em termos de fluxos inter-área, embora a maior parte da influência seja intra-Matemática, verificam-se trocas relevantes com Ciências Humanas (especialmente Educação): há registro de dez ocorrências em que orientadores classificados em Educação aparecem como formadores de doutorados em Matemática, além de 3 ocorrências no sentido inverso (Matemática para Educação), o que pode indicar coorientações, trajetórias de formação interdisciplinar ou ainda inconsistências de classificação.

No plano dos atores individuais, o recorte apresenta alguns orientadores-hubs que moldam a topologia relacional: destacam-se, entre outros, Luiz Antonio Barrera San Martin (37 descendentes, da área Matemática), Lisete Regina Gomes Arelaro (26 descendentes, da área Educação), Alexandre Nolasco de Carvalho (20 descendentes, da área Matemática), Ivan Chestakov (20 descendentes, da área Matemática) e Sonia Barbosa Camargo Igliori (18 descendentes, da área Educação), perfis que apontam para polos de

difusão de orientação com impacto marcado sobre a estrutura observada.

Comparativamente ao perfil descrito para o corpo docente da FEI — onde as Engenharias constituem o núcleo central de influência e há ocorrência de áreas relativamente isoladas — o recorte IME-USP aqui analisado revela uma centralidade distinta (Matemática em lugar de Engenharias) e a ausência de grandes áreas completamente isoladas; por outro lado, a presença expressiva de registros sem classificação (“Unknown”) impõe cautela à comparação direta entre as duas instituições.

5. Conclusão

O presente trabalho demonstrou a viabilidade e utilidade de um fluxo metodológico que integra extração automatizada de currículos (via Plataforma Acácia), modelagem em grafos (Neo4j) e visualização (*Gephi*) para mapear a genealogia acadêmica de docentes e suas áreas de doutorado. A execução prática do método permitiu não apenas a construção de grafos individuais de ascendência/descendência, mas também a síntese em metagrafos de áreas que revelaram padrões de influência e estruturas disciplinares institucionais.

No estudo de caso do Centro Universitário FEI, a aplicação do método sobre 1472 pesquisadores resultou em uma rede densa (1370 vértices e 2112 arestas) e na identificação de 65 áreas com 179 relações de influência, evidenciando a forte centralidade das Engenharias — em especial a Engenharia Elétrica — tanto como fonte quanto como destino de orientações. Observou-se predominância de orientações intra-área na grande área de Engenharias. Esses resultados apontam para uma identidade institucional tecnicamente concentrada e para uma mobilidade interdisciplinar relativamente limitada no corpo docente analisado.

A análise complementar do IME-USP, ainda que em escopo menor (40 pesquisadores pivôs extraídos para o caso), confirmou que perfis institucionais distintos produzem topologias de rede distintas, o que sublinha a pertinência de análises comparativas entre unidades e universidades para compreender variações disciplinares e trajetórias formativas.

Contudo, a investigação também revelou limitações metodológicas e de qualidade de dados que impactam a interpretação dos resultados. Parte significativa dos registros apresentou campos de grande-área ou universidade não preenchidos, e a presença de “ascendentes” oriundos de fora do recorte gerou fluxos classificados como “*Unknown*”, o que pode enviesar estimativas de influência interárea e a centralidade observada. Assim, inferências robustas exigem etapas adicionais de normalização, preenchimento de metadados e validação manual quando necessário.

A partir dos achados e das limitações identificadas, recomenda-se para trabalhos futuros: (i) a utilização direta da base da Plataforma Lattes (quando viável) para reduzir a perda de metadados e ampliar a cobertura temporal e institucional; (ii) procedimentos de limpeza e harmonização das categorias de grande-área

(mapeamento de rótulos divergentes e reconciliação de entradas “*Unknown*”); (iii) a inclusão de métricas topológicas quantitativas (centralidade de grau, betweenness, modularidade, medidas de assortatividade) para complementar a leitura qualitativa dos metagrafos; (iv) a expansão amostral para outras instituições e para um horizonte temporal maior, de modo a permitir análises longitudinais da transmissão de saberes; e (v) a exportação sistemática e tabelas agregadas para facilitar reuso em ferramentas como Gephi e análises reprodutíveis.

Finalmente, do ponto de vista institucional e científico, o método proposto constitui uma ferramenta útil para diagnosticar concentrações disciplinares, identificar orientadores-hubs e mapear potenciais pontes interdisciplinares. Esses insumos podem apoiar decisões estratégicas de formação, cooperação e inovação, por exemplo, iniciativas para fomentar coorientações inter-área, diversificar a origem formadora do corpo docente e promover políticas que estimulem maior mobilidade epistemológica. Em suma, o estudo oferece um protótipo metodológico replicável e extensível, cuja aplicação ampliada pode contribuir de maneira relevante para políticas institucionais de formação e para o entendimento das dinâmicas de reprodução e transformação do capital científico.

O método implementado no trabalho mostra-se eficaz para automatizar a coleta de dados do Lattes através da Acácia e estruturar grafos de relações de orientação, incorporando informações de grande área e universidade que permitem análises mais refinadas. Porém é importante dizer que, para estudos futuros com escopos de instituições ou áreas diferentes, a melhor base de dados para utilizar como estudo de caso é a plataforma Lattes, que foi a base de dados utilizada para a Plataforma Acácia.

Ademais, este trabalho pode ser utilizado em centros de educação que buscam analisar e/ou aumentar a diversidade epistemológica de seu corpo docente com doutorado, podendo ter uma maior gama de áreas de conhecimento. Este trabalho também pode ser utilizado por estudantes que buscam uma universidade com grande foco em sua área.

5. Referências

- [1] DAMACENO, R.; ROSSI, L.; MENA-CHALCO, J. Identificação do grafo de genealogia acadêmica de pesquisadores: Uma abordagem baseada na plataforma lattes. In: . [S.l.: s.n.], 2017.
- [2] Centro Universitário FEI. Corpo Docente da FEI. 2025. Acesso em 20 de Agosto de 2025. Disponível em: <<https://portal.fei.edu.br/corpo-docente-da-fei>>.
- [3] USP. Corpo Docente IME-USP. 2025. Acesso em 5 de Setembro de 2025. Disponível em: <<https://www.ime.usp.br/mat/docentes/>>.
- [4] DAMACENO, R.; ROSSI, L.; MENA-CHALCO, J. Identificação do grafo de genealogia acadêmica de

pesquisadores: Uma abordagem baseada na plataforma lattes. In: . [S.l.: s.n.], 2017. Acesso em 7 de Julho de 2025

Agradecimentos

À instituição FEI pela realização da Iniciação Científica e ao Luciano Rossi pela confiança e apoio para a realização deste trabalho.

¹ Aluno de IC do Centro Universitário FEI (ou FAPESP, CNPq ou outra). Projeto com vigência de 02/2025 a 01/2026.