

Relatório

Trabalho de Teoria dos Grafos: Análise das Redes de Colaboração Científica

Equipe: Lucas de Farias Teixeira, Gustavo Henrique Campestrini e Nicolas Andrei Ceruti

1. Introdução

Este trabalho tem como objetivo analisar as redes de colaboração científica entre professores do programa de pós-graduação em Desenvolvimento Regional da FURB, utilizando dados obtidos dos currículos Lattes.

2. Coleta de Dados

A coleta de dados foi realizada a partir de uma base em formato XML fornecida pelo professor, contendo informações extraídas dos currículos Lattes dos professores do programa de pós-graduação em Desenvolvimento Regional da FURB.

Foi necessário tratar esses dados para extrair informações específicas e relevantes, como os nomes dos pesquisadores, os autores das publicações de artigos completos ou em periódicos, e os orientandos de mestrado e doutorado. Além disso, realizamos a limpeza de dados irrelevantes e corrigimos inconsistências, como variações nos nomes dos autores e problemas de codificação.

3. Desafios Enfrentados

3.1. Coleta de Informações do Lattes

Problema:

Os nomes dos pesquisadores foram encontrados em formatos variados nos currículos Lattes. Algumas das inconsistências incluíam:

- Uso de iniciais seguidas de sobrenomes;
- Nomes completos com ordens diferentes;
- Abreviações inconsistentes;
- Diferenças na acentuação dos caracteres.

Solução Adotada:

O problema foi solucionado da seguinte maneira:

1. **Listas de formas de citação:** Para currículos que continham listas de possíveis formas de citação, utilizamos essas informações para padronizar os nomes.
2. **Expressões regulares (Regex):** Para currículos sem listas de citação, utilizamos expressões regulares para identificar e padronizar os nomes extraídos.
3. **Normalização dos nomes:** uma função em Python baseada na biblioteca unicodedata foi implementada para remover acentos e transformar caracteres acentuados em suas versões equivalentes no padrão ASCII.

Trecho do código utilizado para normalizar os nomes dos pesquisadores

```
def remove_acentos(texto: str) -> str:
    nfkd_form = unicodedata.normalize('NFKD', texto)
    only_ascii = nfkd_form.encode('ASCII', 'ignore')
    return only_ascii.decode(encoding="iso-8859-1")

def normaliza_nome(nome: str) -> str:
    nome = remove_acentos(nome).lower().replace("-", " ")
    if "," in nome:
        nomes = nome.split(",")
        nome = nomes[1].strip() + " " + nomes[0].strip()

    return nome
```

Impacto na Análise:

A padronização dos nomes foi essencial para evitar duplicidades e garantir a integridade do grafo. Erros nessa etapa poderiam resultar na representação de um mesmo pesquisador por múltiplos vértices, comprometendo a análise.

3.2. Problemas com Codificação de Caracteres nos Arquivos XML

Problema:

Os arquivos XML extraídos dos currículos Lattes apresentaram diferenças de codificação de caracteres (como UTF-8 e ISO-8859-1), causando erros ao abrir e processar os dados.

Solução Adotada:

Para resolver o problema, utilizamos o encode ISO-8859-1 no momento de abrir os arquivos. Isso garantiu que os caracteres fossem lidos corretamente.

Código utilizado para ler o arquivo na codificação correta

```
for filename in os.listdir(basepath):
    with open(basepath + filename, "r+", encoding="iso-8859-1") as file:
        curriculos.append(xmltodict.parse(file.read(), encoding="iso-8859-1"))
```

Impacto na Análise:

Essa solução foi fundamental para garantir a integridade dos dados, evitando duplicatas e perdas de informações causadas por diferenças de codificação. Sem esse tratamento, a construção do grafo poderia ser prejudicada devido à interpretação incorreta dos dados.

3.3. Desafio na Visualização do Grafo

Problema:

O grafo gerado possuía uma grande quantidade de vértices e arestas, o que resultava em grafo extremamente difícil de visualizar de forma clara. A grande quantidade de informações tornou a interpretação desafiadora, dificultando a identificação de padrões de colaboração e a análise da estrutura do grafo.

Solução Adotada:

Para resolver esse problema, utilizamos o Kamada-Kawai Layout, que é um algoritmo de layout para grafos que proporciona uma distribuição mais equilibrada dos vértices. Além disso, para melhorar a clareza da visualização, reduzimos o tamanho dos vértices e definimos cores específicas para diferentes tipos de colaboração. Por exemplo:

- **Orientados de mestrado** foram representados com a cor azul;
- **Orientados de doutorado** foram representados com a cor verde.

Essa estratégia de diferenciação visual facilitou a compreensão de como os vértices estavam organizados e o que cada um representava no grafo, tornando a visualização mais intuitiva.

Impacto na Análise:

Essa abordagem permitiu uma melhor compreensão da rede de colaboração, facilitando a interpretação das relações entre os pesquisadores e a visualização dos diferentes tipos de colaboração. A clareza na apresentação visual também contribuiu para a análise de clusters e comunidades.

4. Importância das colaborações e da formação acadêmica

Com base na nossa pesquisa, realizada a partir dos dados do Currículo Lattes, observamos que a probabilidade de colaboração entre pesquisadores com o mesmo orientador é de apenas 11,78%. Isso mostra que muitos pesquisadores acabam não colaborando nem mesmo com sua própria "equipe", ou seja, com colegas que compartilharam um vínculo acadêmico direto durante a formação.

Apesar da baixa taxa de cooperação entre pesquisadores com o mesmo orientador, a formação acadêmica oferece uma base inicial indispensável para criar conexões que podem ser expandidas ao longo da carreira. Incentivar e fortalecer essas colaborações é essencial para diversificar as redes de pesquisa, aumentar a troca de conhecimentos e acelerar avanços científicos de forma mais ampla e integrada.