



Análise da rede de co-autoria de professores pesquisadores da Unipampa campus Uruguaiana

Amanda Fagundes Lopes 161151957

Sandro Matheus Vila Nova Marques 1701570595

Wesley Ferreira de Ferreira 161151193

Sumário

1. Introdução
2. Metodologia
3. Resultados obtidos
4. Conclusão





Introdução

Essa análise foi feita em dados obtidos a partir de arquivos “HTML” gerados pelos Lattes dos Docentes da Universidade Federal do Pampa(Unipampa), campus Uruguaiana do Rio Grande do Sul. Os objetivos deste trabalho são identificar os autores que mais colaboram e os termos mais recorrentes em suas pesquisas, utilizando ferramentas de análise de dados e geração de grafos para visualização.



Metodologia

Os materiais analisados neste trabalho foram os currículos encontrados no Lattes, disponíveis publicamente. O currículos foram baixados manualmente um a um, pois o sistema de “captcha” não permitirá métodos de automatização. Após a extração os currículos passaram pela ferramenta de análise scriptLattes, escrita na linguagem Python, para retirada das informações necessárias, para que então, pudéssemos gerar os grafos com a ferramenta Gephi, que serão vistos a seguir. Também foram necessários scripts desenvolvidos por nós em python para extração de HTML's gerados pelo scriptLattes





Resultados Obtidos

Após os currículos serem baixado e serem analisados automaticamente pelo ScriptLattes retornou dados úteis para o nosso contexto de análise de duas formas diferentes, um grafo no formato GDF e HTML's com uma compilação de informação dos autores e publicações.

O grafo obtido foi importado pelo software Gephi e a partir dele fizemos a representação gráfica do Grafo e tivemos acesso a uma série de métricas podem ser vistos na tabela 1.

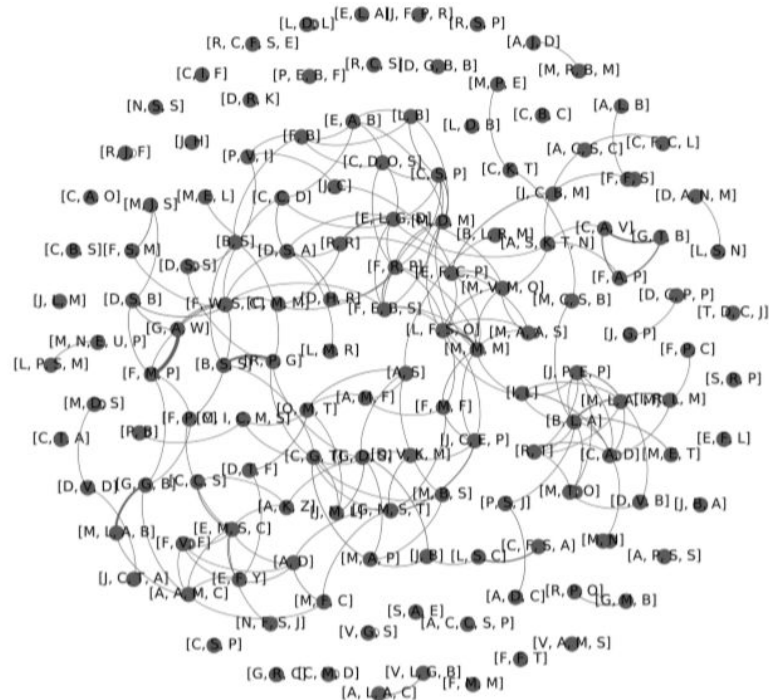


Figura 1. Grafo de Co-Autoria dos Professores da Unipampa Uruguiana



Métrica	Valor
Nós	138
Arestas	203
Componentes conectados	41
Diâmetro da rede	12
Densidade do grafo	0.02

Tabela 1. Métricas fornecidas pelo Gephi sobre o grafo



A partir do software Gephi é possível exportar uma tabela em formato CSV contendo métricas mais detalhadas por componente do grafo(Pesquisadores). Através desta tabela podem ser feitas análises mais profundas.


Na tabela gerada pelo Gephi foram aplicados alguns métodos da biblioteca “**Pandas**” da linguagem de programação Python 3 no ambiente Jupyter Notebook para manipular esses dados.

Os resultados obtidos sobre essas análises podem ser vistos nas tabelas 2, 3, 4.




ID	Label	Produção Bibliografica
71	[F, P, C]	426
31	[M, J, S]	242
73	[A, S]	239
118	[M, N]	230
74	[J, C, E, P]	190

Tabela 2. 5 professores com maior número de publicações.




ID	Label	Grau Ponderado
121	[F, F, S]	280
118	[M, N]	255
115	[L, D, L]	212
122	[M, T, O]	181
4	[G, A, W]	137

Tabela 3. 5 professores com maior grau de colaboração.



Total de produções bibliográficas	9448
Centralidade média de grau	3.057971
Centralidade média de proximidade	0.245964
Centralidade média de intermediação	0.011688

Tabela 4. Informações gerais extraídas da tabela.



Ainda sobre o ambiente anteriormente descrito e os HTML's gerados pelo ScriptLattes, foi feita uma extração de todos os títulos dos artigos por meio da biblioteca “re” da linguagem Python .

Essa biblioteca faz a extração de informações textuais com base em expressões regulares.

A expressão regular utilizada foi `(\w.*)`

Com base nas strings extraídas foi gerada uma wordcloud(figura 2) contendo as principais palavras chaves dos títulos(salvo palavras chaves de grande ocorrência no mundo científico) a fim de identificar as áreas de maior pesquisa pelos professores.



Figura 2. Wordcloud dos títulos das produções bibliográficas



Conclusão

Com base nas análises, percebeu-se que os professores têm poucas publicações em conjunto, como pode ser visto no número de componentes conectados (tabela 1), em comparação a quantidade de professores (Nós do Grafo), na densidade e no grafo da figura 1. Porém, percebe-se também, que os professores que co-autoram entre si, o fazem de maneira intensa, como pode ser visto na tabela 3.

Na análise dos títulos das publicações pode ser observado, que uma grande preocupação das pesquisas sobre a saúde física das pessoas, visto que termos como "Educação Física" e "Atividade física" aparecem com maior recorrência (em maior tamanho), nota-se também uma preocupação muito grande com idosos e pessoas em período escolar (crianças e adolescentes) e também saúde da família.