

Análise de rede de co-autoria dos professores pesquisadores da Universidade Federal do Pampa campus Uruguaiiana

Amanda Lopes¹, Sandro Matheus¹, Wesley Ferreira¹

¹ Universidade Federal do Pampa (Unipampa) - Campus Alegrete
Av. Tiarajú – 810 – Ibirapuitã – Alegrete – RS – Brazil

(amandagobus, wferreira531)@gmail.com, sandro-matheus@hotmail.com

Resumo. *Esse artigo foi um estudo sobre a rede de co-autoria dos professores pesquisadores da Universidade Federal do Pampa campus Uruguaiiana aonde foram usados ferramentas e métodos de análise de dados e de redes aplicados sobre os currículos da Plataforma Lattes cujos links estavam disponíveis no portal da universidade. Foi possível identificar os pesquisadores que mais publicam, os que mais co-autoram e as áreas de maior interesse.*

1. Introdução

Esse artigo apresenta uma análise da rede de colaboração científica dos pesquisadores da Universidade Federal do Pampa (Unipampa), campus de Uruguaiiana. Os objetivos deste trabalho são identificar os autores que mais colaboram e os termos mais recorrentes em suas pesquisas. Para isso foram usadas os currículos disponíveis publicamente na Plataforma Lattes e ferramentas de extração desses currículos e de análise de rede. Com isso foi possível traçar um perfil colaborativo e perfil das áreas de maior incidência de pesquisas desses profissionais.

2. Referencial Teórico

A ciência por trás da análise de redes sociais se apóia em um conceito central: nossas relações interpessoais, definidas por como somos e como agimos com base nas nossas personalidade, na nossa educação, história, raça e etnia [Maksim Tsvetovat 2011].

A análise de rede social é um assunto que tem atraído o interesse de pesquisadores. Um dos principais usos dela é a identificação dos atores mais importantes em uma rede [Wasserman and Faust 1993].

Uma rede social pode ser representada por grafos, artifícios matemáticos representados por um conjunto $G(V,E)$, sendo V os vértices(atores) e E as arestas(relações). Tendo varias representações diferentes e varias métricas aplicáveis a tal modelo, tais como as medidas de centralidade, aplicadas nesse artigo.

De acordo com Wasserman(1993), a centralidade de grau mede a quantidade de relações que um ator tem com outros dentro da rede, podendo ou não levar em consideração o peso das conexões; a centralidade de proximidade mede o quão perto um ator está de todos os outros dentro de uma rede. A centralidade de intermediação no entanto, como descrito por Freeman(1977), quantifica o número de vezes que um ator age como parte de um menor caminho entre dois outros atores.

3. Materiais e Métodos

O material usado para o desenvolvimento deste artigo foi os currículos dos professores pesquisadores da Universidade Federal do Pampa campus Uruguaiana, disponível publicamente na Plataforma Lattes e cujos *links* estão disponíveis no portal da universidade referente ao campus analisado.

Os currículos foram baixados manualmente, pois, de forma simples não é possível a automação do processo devido ao uso de *captcha* pela Plataforma Lattes, após eles foram analisados pela ferramenta ScriptLattes, desenvolvida na linguagem python, que faz o *parsing* e a extração das informações mais importantes, gerando um grafo de co-autoria e uma lista das publicações feitas por todos eles.

O grafo gerado foi importado para o *software* de análise de redes Gephi da onde foi possível extrair a imagem do gráfico e uma tabela de várias estatísticas que puderam ser analisadas, juntamente com a lista de publicações por meio da linguagem python, usando as bibliotecas "pandas" e "re".

4. Análise dos dados

Após os currículos serem baixados manualmente e serem analisados automaticamente pelo ScriptLattes foi retornado dados úteis para o nosso contexto de análise de duas formas diferentes, um grafo no formato GDF e HTML's com uma compilação de várias informações sobre os autores e publicações.

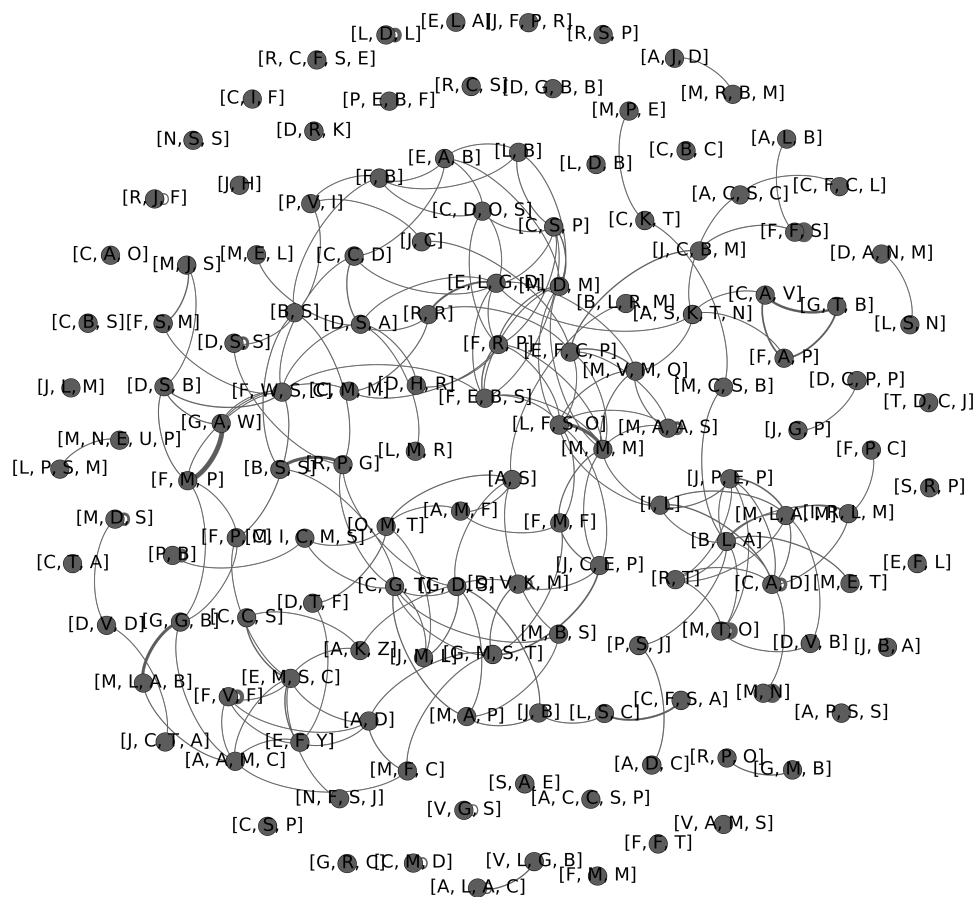
O grafo obtido, como descrito anteriormente, foi importado pelo *software* Gephi e a partir dele tivemos acesso a representação gráfica do grafo, visível na figura 1 e de uma série de métricas aplicáveis, visíveis na tabela 2.

A partir do *software* Gephi também é possível exportar em formato CSV uma tabela que contém métricas mais detalhadas individualmente por componente do grafo (Pesquisadores) onde podem ser feitas análises mais profundas. Sobre essa tabela foi aplicadas alguns métodos da biblioteca *pandas* da linguagem de programação Python 3 no ambiente Jupyter Notebook para operar nessa tabela. Os resultados obtidos sobre essas análises podem ser vistos nas tabelas 3, 4, 1.

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 que faz extração de informações em textos com base em expressões regulares. A expressão regular escrita para a extração dos títulos dos arquivos que continham todos artigos sintetizados foi ` (\w. *) `, e 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.

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

Tabela 1. Informações gerais extraídas da tabela.



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

Tabela 2. Métricas fornecidas pelo Gephi sobre o grafo

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 3. 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 4. 5 professores com maior grau de colaboração.

5. Resultados e Discussões

Percebe-se com base nas análises feitas, que dentre os professores pesquisadores da Unipampa campus Uruguaiana, temos poucos deles que publicam em conjunto, como poder ser visto nos número de componentes conectados na tabela 2 em comparação a quantidade de professores(Nós do grafo), na densidade do grafo e no grafo presente na 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 4.

Na análise dos títulos das publicações pode ser observado, com a ajuda da *wordcloud*(nuvem de termos) na figura 2, 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. Outrossim, é perceptível com base na *wordcloud* a preocupação de testes em ambientes controlados(in vitro e em animais) para diminuição no risco de vida dos pacientes.

Referências

- Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks.
- CNPq, C. N. d. D. C. e. T. Plataforma lattes.
- Freeman, L. (1977). *A Set of Measures of Centrality Based on Betweenness*, volume 40.
- Maksim Tsvetovat, A. K. (2011). *Social Network Analysis for Startups*.
- McKinney, W. (2017). *Python for Data Analysis 2nd Edition*. oreilly.
- Mena-Chalco, J. P. and Junior, R. M. C. (2009). scriptLattes: an open-source knowledge extraction system from the lattes platform. *Journal of the Brazilian Computer Society*, 15(4):31–39.
- Wasserman, S. and Faust, K. (1993). *Social Network Analysis Methods And Applications*, volume 8.