

Análise comparativa de métricas topológicas em redes de coautoria para o problema de predição de *links*

*Mariana Magalhães de Mattos Coelho, Claudia Marcela Justel
Engenharia de Computação, Instituto Militar de Engenharia
Praça General Tibúrcio 80, 2229-270, Praia Vermelha, Rio de Janeiro, RJ, Brasil.
*mariana@ime.eb.br

RESUMO: O problema denominado predição de *links* consiste em estimar o surgimento de arestas entre nós de um grafo que representa uma rede de elementos interligados (e.g., uma rede de coautoria cujos vértices e arestas representam, respectivamente, os autores e as publicações entre eles). Diversas abordagens para resolver esse problema foram propostas nos últimos anos. Dentre as diferentes abordagens existentes, neste trabalho consideramos a abordagem topológica, para a qual foram definidas diferentes métricas. O objetivo deste trabalho é comparar quatro métricas topológicas por meio de experimentos em uma rede de coautoria. Apresentamos os resultados e as conclusões obtidas a partir dos experimentos executados em uma rede real desenvolvida por alunos participantes do projeto de pesquisa Algoritmos em Grafos.

PALAVRAS-CHAVE: análise de redes sociais, predição de *links*, aplicações de grafos.

ABSTRACT: The link prediction problem consists in estimating the appearance of edges between nodes of graphs representing a network of interconnected elements (e.g. a co-authorship network whose vertices and edges represent, respectively, the authors and the publications between them). In the last years, several approaches to solve this problem were proposed. Among them, one line of work is considering the topology of the network. This paper is about different metrics used to solve the topological approach of the link prediction problem. Our goal is to compare 4 different metrics by conducting experiments in a collaboration network. We present the results and conclusions obtained with the experiments for a real network developed by students, working in the research project Graph Algorithms.

KEYWORDS: social network analysis, link prediction, graph applications.

1. Introdução

Neste trabalho, abordamos o problema de predição de *links*, um problema fundamental da área Análise de Redes Sociais, com aplicações em diferentes domínios, como, por exemplo, predição da evolução em redes dinâmicas, indicação de novas amizades em redes sociais, recomendação de produtos e serviços [1], entre outros. Uma rede de coautoria é representada por vértices e arestas, em que os nós são os autores e as arestas são as publicações entre eles.

O problema de predição de *links* procura identificar ligação entre pares de nós para os quais essa conexão não existe. Para resolver esse problema, existem diferentes soluções. Algumas delas utilizam características ou atributos dos nós, e outras só utilizam informação estrutural do grafo. As primeiras são conhecidas

como abordagens baseadas em características, e as últimas como abordagens topológicas. Também existem abordagens que usam ambas, informação de características dos nós e informação estrutural, e são denominadas abordagens híbridas [2].

O artigo de Liben-Nowell e Kleinberg [3] introduziu diferentes métricas topológicas para resolver o problema de predição de *links* que associam a um par de nós não conectados x, y de um grafo G num tempo t , um coeficiente, denominado $score(x, y)$. Depois disso, uma lista ordenada pelos valores de $score(x, y)$ é produzida para gerar um preditor de novas conexões.

Em [4], Nassar *et al.* propõem uma nova abordagem topológica para predição de *links* denominada “predição aos pares” que, em vez de considerar um par de nós, determina qual nó tem mais chance de formar um triângulo com uma aresta existente. Dessa

forma, uma nova versão das métricas propostas por [3] é introduzida.

Ambos os tipos de métricas, que denominaremos “tradicional” e “aos pares”, são métricas topológicas de predição de *links* para redes homogêneas [5]. Mas, de acordo com as informações de que dispomos, não foram comparadas as versões tradicional e aos pares das métricas propostas em [3] e [4], respectivamente, isto é, não existem artigos fazendo a comparação sobre o desempenho das duas versões das métricas numa mesma rede real. Por esse motivo, consideramos importante desenvolver experimentos que permitam identificar vantagens e desvantagens das duas versões, tradicional e aos pares.

O objetivo deste trabalho é fazer uma análise comparativa do desempenho de diferentes métricas topológicas nas versões tradicional e aos pares em redes de coautoria. Para tanto, utilizamos um *dataset* de uma rede de coautoria gerada a partir de informações da Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) para a realização dos experimentos que permitem fazer a comparação e, posteriormente, analisamos os resultados obtidos pelas duas versões. Pelos experimentos realizados, podemos concluir que a versão aos pares obteve uma pequena vantagem em relação à versão tradicional.

O restante deste artigo está organizado como se segue. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 explicita a descrição do problema de predição de *links*, assim como as métricas introduzidas em [3] e [4] e as definições dos conceitos utilizados neste trabalho. A metodologia utilizada para viabilizar a comparação das métricas nas duas versões propostas pelos dois artigos mencionados anteriormente é descrita na Seção 4, e a Seção 5 apresenta os experimentos realizados com seus respectivos resultados. Finalmente, a Seção 6 destaca as conclusões do trabalho.

2. Trabalhos relacionados

No artigo de Liben-Nowell e Kleinberg [3], trabalho importante em predição de *links*, é analisada uma rede de

1 Redes homogêneas possuem um único tipo de nós e *links*.

coautoria acadêmica utilizando características topológicas da rede para prever a formação de arestas entre dois nós não conectados. Nesse caso, a rede de coautoria é homogênea.

Os autores de [6] propõem uma adaptação das métricas tradicionais em redes homogêneas descritas em [3], para serem utilizadas em redes heterogêneas bipartidas, isto é, uma rede na qual os nós são de dois tipos diferentes (uma bipartição do conjunto de nós) e todas as arestas têm extremidades em conjuntos diferentes da bipartição. Os autores propõem transformar o conjunto $\Gamma(u)$ de vizinhos de u , em $\Gamma'(u) = \bigcap_{v \in \Gamma(u)} \Gamma(v)$ (vizinhos dos vizinhos do nó u), ou seja, substituir $\Gamma(u)$ por $\Gamma'(u)$ no cálculo das métricas topológicas tradicionais.

Em 2010, [7] apresentou uma solução diferente da proposta em [6] para tratar redes heterogêneas bipartidas. Nesse caso, é utilizada uma projeção do grafo que representa a rede sobre um dos dois conjuntos da bipartição, e então são definidas as métricas de acordo com essa projeção. A Fig. 1 ilustra as duas projeções para um exemplo de grafo bipartido, segundo [7].

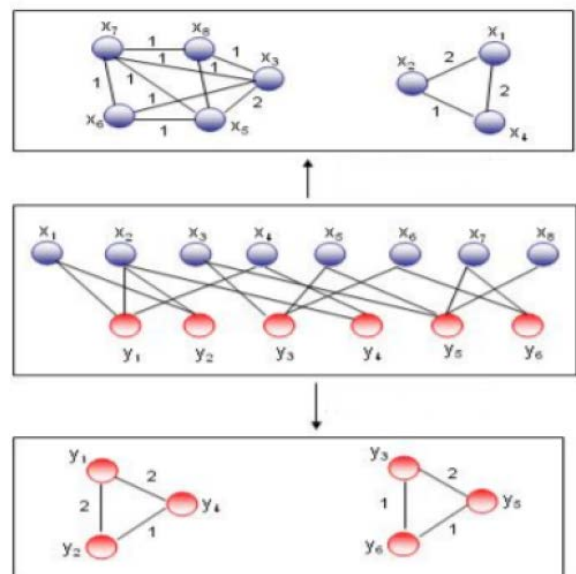


Fig. 1 – Duas projeções do grafo bipartido $G = (X \cup Y, E)$. Fonte: [7].

Em [8], foi desenvolvida uma solução para uma rede heterogênea multipartida. O método proposto, denominado MRLP (*multi-relational link prediction*), cujo componente principal utiliza um esquema de

pesos para diferentes tipos de combinações de arestas, a partir da contagem de subgrafos formados por três nós que existem na rede.

Já em 2019, Nassar *et al.* [4] propuseram prever a formação de arestas considerando um nó e uma aresta existente na rede. Os artigos [3] e [4] consideram de formas diferentes o fechamento de triângulos ao analisar subgrafos formados por três nós em redes homogêneas (Fig. 2 e Fig. 3).

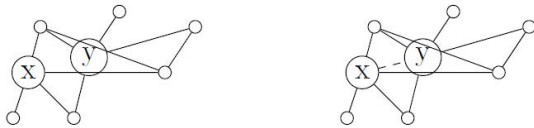


Fig. 2 – O desenho à esquerda representa G num tempo t e as ligações existentes até esse instante (x e y são vértices não adjacentes); o desenho à direita representa G num tempo t' e as ligações existentes até esse instante, para $t < t'$. A aresta tracejada seria uma possível ligação entre x e y no instante t' usando métrica para predição de links tradicional. **Fonte:** [4].



Fig. 3 – O desenho à esquerda representa G num tempo t e as ligações existentes até esse instante (u e v são adjacentes); o desenho à direita representa G num tempo t' e as ligações existentes até esse instante, para $t < t'$. As arestas tracejadas seriam possíveis ligações entre u e w e entre v e w usando métricas para predição de links aos pares. **Fonte:** [4].

3. Duas abordagens topológicas

Redes sociais são objetos altamente dinâmicos, que crescem e mudam rapidamente ao longo do tempo pela adição de novas arestas, de acordo com o surgimento de novas interações no grafo original. O problema de predição de *links* está relacionado com a evolução da rede social ao longo do tempo e pode ser definido da seguinte forma: dado um retrato instantâneo de uma rede social num tempo t , o problema de predição de *links* procura prever com certa precisão arestas que serão adicionadas nessa rede durante o intervalo de tempo entre t e um tempo futuro t' [3].

Todas as métricas consideradas por [3] para predição de *links* associam um coeficiente a pares de nós não adjacentes x, y , denominado *score*(x, y), a partir de um grafo de entrada e produzem uma lista ordenada na ordem não crescente desses coeficientes. Os coeficientes podem ser considerados como uma medida de proximidade ou similaridade entre um par de nós, e serão chamados de **métricas tradicionais**. Essas métricas foram adaptadas de algumas técnicas utilizadas em Teoria de Grafos e Análise de Redes Sociais. Elas, em geral, não foram criadas para computar a similaridade entre nós de um grafo, portanto foi necessário modificá-las para o novo propósito.

A notação utilizada é a seguinte: $G = (V, E)$ um grafo não direcionado, $x \in V$ um nó, $\Gamma(x) = \{y \in V : (x, y) \in E\}$ o conjunto de vizinhos do nó $x \in V$, e $|\Gamma(x)|$ a cardinalidade do conjunto $\Gamma(x)$. A Tabela 1 apresenta quatro métricas topológicas tradicionais utilizadas em [3].

Tabela 1 – Métricas na versão tradicional.

Tradicional	
$VC(x,y) =$	$ \Gamma(x) \cap \Gamma(y) $
$JS(x,y) =$	$\frac{ \Gamma(x) \cap \Gamma(y) }{ \Gamma(x) \cup \Gamma(y) }$
$AA(x,y) =$	$\sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log \Gamma(z) }$
$LP(x,y) =$	$ \Gamma(x) \Gamma(y) $

Fonte: [3]

Em [4], Nassar *et al.* observam que a predição de *links* tradicional, definida em [3], pode ser descrita em termos da seguinte pergunta: dado um nó x na rede, quais são os nós mais propensos a serem ligados a *ele*? Nassar *et al.* consideram uma nova versão das métricas, a partir da pergunta: dada uma aresta (u, v) na rede, quais são os nós mais propensos a se conectarem às extremidades da aresta, ou seja, aos vértices u e v ? O objetivo é definir nova versão da proximidade ou similaridade entre cada um dos vértices que são da extremidade da aresta com o nó, que os autores

denominam “aos pares” e denominaremos **métricas aos pares**. Será utilizada a notação a seguir:

$\Gamma^*((u, v)) = \{z \in V : \Gamma(u) \cap \Gamma(v) - \{u, v\}\}$ o conjunto de vizinhos da aresta $(u, v) \in E$; $|\Gamma^*((u, v))|$ a cardinalidade do conjunto $\Gamma^*((u, v))$. As métricas adaptadas utilizando a versão aos pares são denotadas Vizinhos Comuns (VC^*), Similaridade de Jaccard (JS^*), Adamic-Adar (AA^*) e Ligação Preferencial (LP^*). A Tabela 2 apresenta quatro métricas topológicas aos pares.

Tabela 2 – Métricas na versão aos pares.

Aos Pares	
$VC^*(w, (u, v)) =$	$ \Gamma(w) \cap \Gamma^*((u, v)) $
$JS^*(w, (u, v)) =$	$\frac{ \Gamma(w) \cap \Gamma^*((u, v)) }{ \Gamma(w) \cup \Gamma^*((u, v)) }$
$AA^*(w, (u, v)) =$	$\sum_{z \in \Gamma(w) \cap \Gamma^*((u, v))} \frac{1}{\log \Gamma(z) }$
$LP^*(w, (u, v)) =$	$ \Gamma(w) \cap \Gamma^*((u, v)) $

Fonte: [4]

4. Metodologia

Neste trabalho, consideramos um grafo $G=(V,E)$, no qual cada aresta $e \in E$ representa uma interação entre dois nós u e v num instante de tempo $t(e)$ numa rede social. Não serão consideradas múltiplas interações entre u e v . Para um dado instante de tempo t , notamos por G_t o subgrafo de G que contém todas as arestas e tal que $t(e) \leq t$. A formulação matemática do problema é dada a seguir. Escolhemos dois instantes de tempo, $t < t'$, e consideramos um algoritmo que acesse o grafo que representa a rede até o instante t , G_t , e que retorne uma lista de pares de elementos (dois nós não adjacentes, ou um nó e uma aresta em G_t), que são predições de arestas para $G_{t'}$. Os intervalos $(0, t]$ e $(t, t']$ são referidos como intervalo de treino e teste, respectivamente.

Cada preditor p considerado retorna uma lista ordenada L_p de pares em $V \times V$, formada pelas predições de novas interações em $G_{t'}$ em ordem não crescente de confiança.

A medida de desempenho para o preditor p é determinada da seguinte maneira: escolhendo os primeiros k pares de predições de novas interações da lista ordenada L_p (Top- k). Logo, para poder comparar cada métrica M_i na versão tradicional ($i = 1$) e aos pares ($i = 2$), é preciso construir a partir do preditor $p(M_i)$ o Top- k , ou seja, os primeiros k elementos da lista $L_p(M_i)$, para $i = 1, 2$. Finalmente, para cada lista $L_p(M_i)$, determinam-se as medidas de qualidade de classificação do preditor $p(M_i)$, $i = 1, 2$, usando as fórmulas das Eq. 1 a Eq. 4:

$$Precisão = \frac{VP}{VP + FP} \quad (1)$$

$$Acurácia = \frac{VP + VN}{VP + FP + FN + VN} \quad (2)$$

$$Revocação = \frac{VP}{VP + FN} \quad (3)$$

$$F1 = \frac{2 \text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (4)$$

onde,

- VP = verdadeiro positivo (um *link* é classificado, ou previsto, positivo e existe no grafo $G_{t'}$);
- FP = falso positivo (um *link* é classificado, ou previsto, positivo e não existe no grafo $G_{t'}$);
- FN = falso negativo (um *link* é classificado, ou previsto, negativo e existe em $G_{t'}$); e
- VN = verdadeiro negativo (um *link* é classificado, ou previsto, negativo e não existe em $G_{t'}$).

De acordo com a descrição do tipo de rede na qual é possível utilizar as métricas tradicionais e aos pares feita na Seção 3, foi escolhido um *dataset*, correspondente a uma rede de coautoria, para realizar os experimentos comparativos. A seguir, é apresentada a descrição do *dataset*.

Para determinar o *dataset* a ser utilizado nos experimentos, observamos que tanto as métricas tradicionais quanto aos pares não produzem bons resultados em redes heterogêneas bipartidas. Realizamos alguns experimentos preliminares, mostrando que a adaptação proposta em [6] não funciona com as métricas aos pares. Portanto, escolhemos como *dataset* uma rede de coautoria, que é uma rede homogênea. A rede de coautoria pode ser pensada como a projeção, proposta em [7], no conjunto dos autores da rede heterogênea bipartida que relaciona autores e artigos.

Em [9], foi criada uma rede de coautoria a partir de dados retirados em 11 de outubro de 2010 da Plataforma Lattes do CNPq. Os nós do grafo que representa a rede correspondem aos autores, e as arestas identificam pelo menos uma publicação conjunta entre dois autores (portanto o grafo é não direcionado, sem arestas múltiplas) e $t(e)$ corresponde ao ano da primeira publicação conjunta entre os vértices que são extremidades da aresta e .

Posteriormente, a rede criada em [9] foi atualizada com dados até 10 de outubro de 2014 [10]. Tanto a criação da rede quanto a primeira atualização foram desenvolvidas por alunos de graduação num trabalho de final de curso e em um projeto de iniciação científica inserido no projeto *Algoritmos em Grafos*. As Fig. 4 e 5 ilustram os grafos G_{2011} e G_{2014} que representam as duas redes de coautoria antes mencionadas.

No contexto da pesquisa deste artigo, foi realizada uma nova atualização dos dados em 5 de janeiro de 2021. A Tabela 3 apresenta os tamanhos dos conjuntos de nós e arestas dos grafos G_{2011} , G_{2014} e G_{2020} , com informações correspondentes aos períodos (2010, 2014] e (2014, 2020], respectivamente. A Fig. 6 representa o grafo G_{2020} .

Tabela 3 – Informações das redes de coautoria obtidas da plataforma Lattes-CNPq.

$G = (V, E)$	$ V $	$ E $
G_{2011}	207	520
G_{2014}	280	756
G_{2020}	425	1.350

Os experimentos foram realizados em um computador *Intel Core i7*, CPU 1.80GHz, 8GB de RAM. A linguagem de programação *Python* (versão 3.9.4 [11]) foi usada para implementar as métricas, determinar os preditores e obter os Top- k e a biblioteca *NetworkX* 2.5.1 para manipular grafos ([12]).

5. Descrição dos experimentos realizados

5.1 Experimento Lattes-CNPq 2011-2014

Nesse primeiro experimento, utilizamos a rede de coautoria Lattes e produzimos os grafos

G_t e $G_{t'}$ com $t < t'$ que representam a rede, para valores $t = 2011 < t' = 2014$. Observamos que nesse caso existem 32 arestas em G_{2014} entre pares de nós não adjacentes em G_{2011} , ou seja, 32 novas arestas. Houve pequenas diferenças dos valores de Top- k obtidos com os k considerados ($k = 3, 5, 7$), portanto mostramos unicamente os resultados para $k = 7$. O valor $k = 7$ foi escolhido porque a métrica tradicional Vizinhos Comuns apresentou unicamente nove valores de escores diferentes. Todos os valores das medidas da qualidade da classificação dos links preditos obtidos nesse experimento são apresentados na Tabela 4.

A partir da Tabela 4, podemos concluir que, para Vizinhos Comuns e Similaridade de Jaccard, os resultados obtidos pelas métricas na versão aos pares foram melhores ou iguais, exceto na *Revocação* para Vizinhos Comuns. Para Ligação Preferencial, as métricas na versão tradicional tiveram melhores resultados em todos os casos. E para Adamic-Adar os resultados são balanceados para as duas versões das métricas.

Tabela 4 – Resultados para G_{2011} e G_{2014} .

Top-7		
	VC	VC*
Precisão	0,01010101	0,025
Acurácia	0,965818951	0,994732053
F-1	0,019310345	0,035087719
Revocação	0,21875	0,058823529
JS		
	JS	JS*
Precisão	0	0
Acurácia	0,985241094	0,992497372
F-1	0	0
Revocação	0	0
LP		
	LP	LP*
Precisão	0,142857143	0
Acurácia	0,998221239	0,998029793
F-1	0,051282051	0
Revocação	0,03125	0
AA		
	AA	AA*
Precisão	0,142857143	0,058823529
Acurácia	0,998221239	0,996928246
F-1	0,051282051	0,058823529
Revocação	0,03125	0,058823529

5.2 Experimento Lattes-CNPq 2014-2020

Nesse experimento, utilizamos o grafo de coautoria Lattes e produzimos as informações necessárias para trabalhar com $t < t'$, para valores $t = 2014$ e $t' = 2020$. Os resultados foram avaliados para diferentes valores de Top- k ($k = 3, 5, 7$).

Observamos que nesse caso existem 105 arestas em G_{2020} entre pares de nós não adjacentes em G_{2014} , ou seja, 105 novas arestas. Houve pequenas diferenças dos resultados obtidos com os valores de Top- k considerados, portanto mostramos unicamente os resultados para $k = 7$. O valor $k = 7$ foi escolhido porque a métrica tradicional Vizinhos Comuns apresentou unicamente oito valores de escores diferentes. Todos os valores das medidas da qualidade da classificação dos links preditos obtidos nesse experimento são apresentados na Tabela 5.

Dos resultados mostrados na Tabela 5, observamos que, para as métricas Vizinhos Comuns e Similaridade de Jaccard, os resultados da versão aos pares foram melhores quando comparados com as da versão tradicional, exceto na *Revocação*. Para Ligação Preferencial, as duas versões, tradicional e aos pares, apresentaram valor zero na *Precisão*, *F1* e *Revocação*. A *Acurácia* para a métrica Ligação Preferencial tradicional foi um pouco melhor que aos pares. E para Adamic-Adar os resultados obtidos pela versão aos pares mostram uma pequena vantagem, exceto na *Acurácia*.

Tabela 5 – Resultados para G_{2014} e G_{2020}

Top-7		
	VC	VC*
Precisão	0,00791526	0,033950617
Acurácia	0,776524645	0,989178834
F-1	0,015639374	0,05
Revocação	0,647619048	0,094827586
JS		
	JS	JS*
Precisão	0,002932551	0,006024096
Acurácia	0,988408521	0,992981544
F-1	0,004484305	0,007352941
Revocação	0,00952381	0,009433962

continua

Top-7		
	LP	LP*
Precisão	0	0
Acurácia	0,99707602	0,997050532
F-1	0	0
Revocação	0	0
AA		
	AA	AA*
Precisão	0	0,023255814
Acurácia	0,99707602	0,996166584
F-1	0	0,013422819
Revocação	0	0,009433962

6. Conclusão

Neste trabalho, apresentamos os resultados de experimentos comparativos entre duas versões para as métricas de predição de *links*, tradicional e aos pares, em uma rede real de coautoria.

A partir dos resultados obtidos, podemos concluir que as duas versões para as quatro métricas consideradas, Vizinhos Comuns, Similaridade de Jaccard, Ligação Preferencial e Adamic-Adar, apresentam comportamentos parecidos, com uma pequena vantagem para a versão aos pares. Em todos os experimentos realizados, a versão aos pares para Vizinhos Comuns obteve discreta melhoria na *Precisão*, *Acurácia* e *F1*. Em Similaridade de Jaccard, os resultados foram iguais ou melhores em *Precisão*, *Acurácia* e *F1*, enquanto para Adamic-Adar a melhora foi em *F1* e *Revocação*.

Observamos que para Ligação Preferencial no Experimento 2014-2020, tanto a versão tradicional quanto a versão aos pares obtiveram valores iguais a zero para *Precisão*, *F1* e *Revocação*. Já no Experimento 2011-2014, os valores dessas três medidas de qualidade para Ligação Preferencial foram diferentes de zero para a versão tradicional.

A partir dos resultados preliminares, podemos considerar que a versão aos pares, que foi introduzida recentemente, também pode ser usada para resolver a abordagem topológica do problema de predição de *links*, no caso de três das métricas consideradas neste artigo.

Como trabalhos futuros, sugerimos continuar analisando os resultados para os *datasets* do repositório de *preprints ArXiv* em cinco campos de astrofísica ([13]), que já foram usados na literatura para avaliar os resultados de métodos de predição de *links*. Os *datasets* a serem considerados são: *General Relativity and Quantum*

Cosmology (gr-qc), *Astrophysics (astro-ph)*, *Condensed Matter (cond-mat)*, *High Energy Physics – Phenomenology (hep-ph)* e *High Energy Physics – Theory (hep-th)*. Nessa análise futura, pretendemos também incluir experimentos usando uma alteração da versão aos pares, originalmente proposta pelos mesmos autores de [4], publicada em [14].



Fig. 4 – Grafo Lattes-CNPq G_{2011} . Fonte: [9]

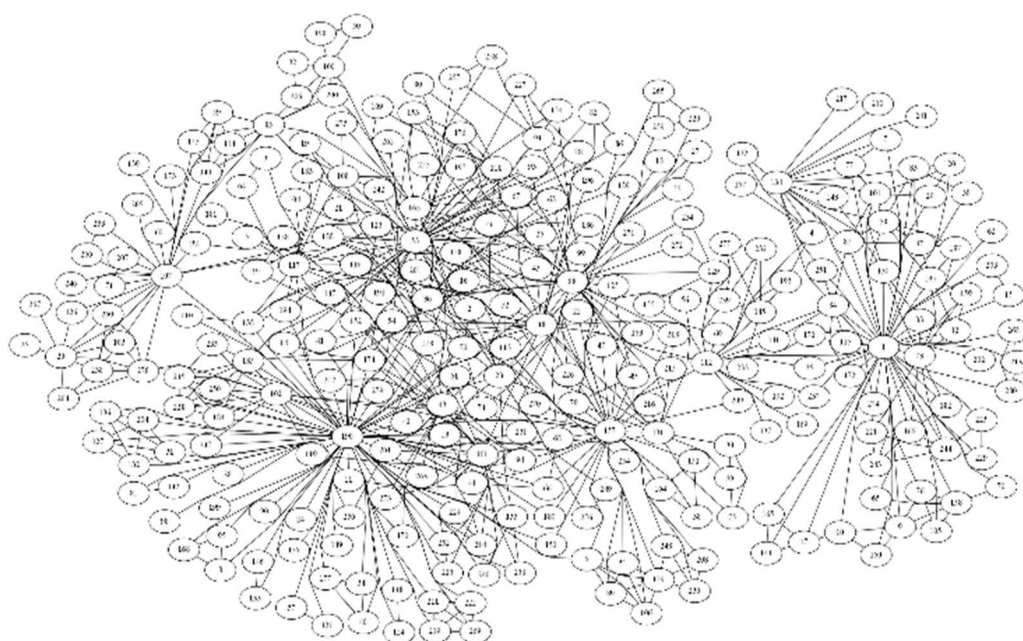


Fig. 5 – Grafo Lattes-CNPq G_{2014} . Fonte: [10]

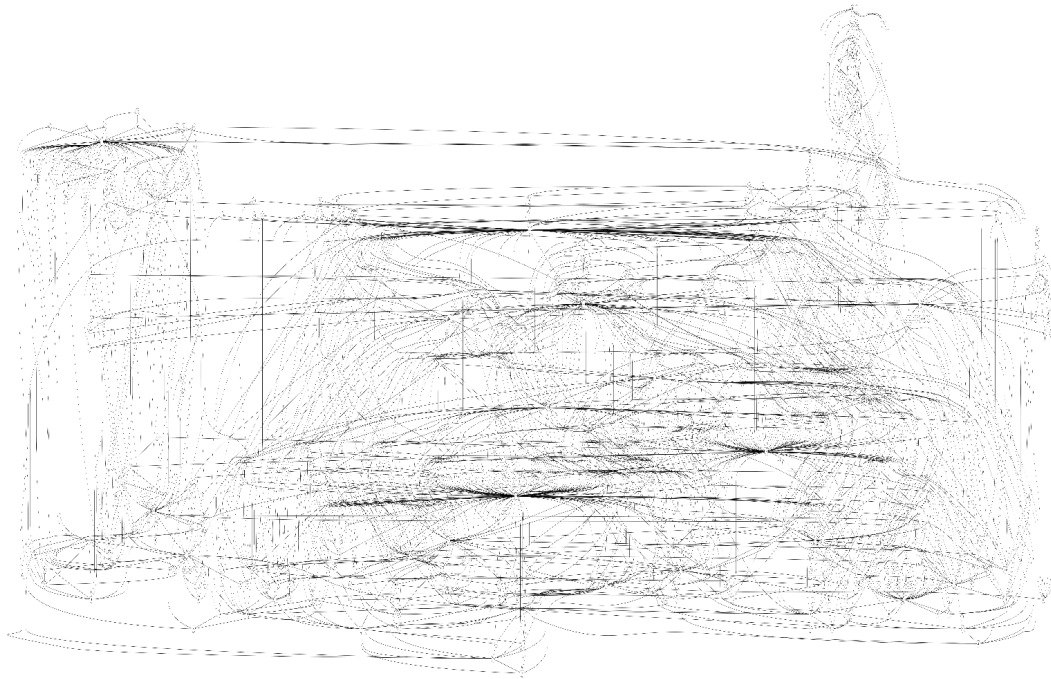


Fig. 6 – Grafo Lattes-CNPq G_{2020} .

Referências

- [1] ZAREIE, A.; SAKELLARIOU, R. Similarity based link prediction in social networks using latent relationships between the users. *Scientific Reports*, v. 10, n. 20137, p. 1–11, 2020.
- [2] PUJARI, M. *Link Prediction in Large-scale Complex Networks (Application to bibliographical Networks)*. Paris: University Sorbonne Paris Cité, 2015.
- [3] LIBEN-NOWELL, D.; KLEINBERG, J. The Link-Prediction Problem for Social Networks. *Journal of the American Society for Information Science and Technology*, v. 58, n. 7, p. 1019–1031, 2007.
- [4] NASSAR, H.; BENSON, A. R.; GLEICH, D. F. Pairwise Link Prediction. arXiv:1907.04503v1 [cs.SI]. 10 Jul 2019.
- [5] AL HASAN, M.; ZAKI, M. J. A survey of link prediction in social networks. In AGGARWAL, C. C. *Social Network Data Analytics*. New York: Springer, 2011. p. 243–275.
- [6] HUANG, Z.; LI, X.; CHEN, H. Link prediction approach to collaborative filtering. In *ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*. New York: ACM Digital Library, 2005. p. 141–142.
- [7] BENCHETTARA, N.; KANAWATI, R.; ROUVEIROL, C. Supervised machine learning applied to link prediction in bipartite social networks. In *International Conference on Advances in Social Networks Analysis and Mining*. Piscataway: IEEE, 2010.
- [8] DAVIS, D.; LICHTENWALTER, R.; CHAWLA, N. V. Supervised methods for multi-relational link prediction. *Social Network Analysis and Mining*, v. 3, p. 127–141, 2013.
- [9] BARBOSA, D. A. B. L.; AVELINO, L. B.; SOUZA, R. F.; OLIVEIRA, C. C. G. F.; JUSTEL, C. M. Medidas de centralidade e detecção de comunidades em rede de co-autoria. In *Anais do XVIII Simpósio Brasileiro de Pesquisa Operacional*. Rio de Janeiro: Sobrapo, 2011. p. 2574–2583.
- [10] MAGNANI, H. M. *Redes Sociais e Comunidades*. Relatório Final Projeto Institucional de Iniciação Científica CNPq-IME, 2014.
- [11] PYTHON. Disponível em: <https://www.python.org/>. Acesso em: 6 abr. 2021.
- [12] NETWORKX. Disponível em: <http://networkx.org>. Acesso em: 6 abr. 2021.
- [13] ARXIV. Disponível em: <https://arxiv.org/>. Acesso em: 6 abr. 2021.
- [14] NASSAR, H.; BENSON, A. R.; GLEICH, D. F. Pairwise Link Prediction. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. New York: ACM Digital Library, 2019. p. 386–393.