

12/05/21



Defesa Projeto de Graduação em Computação

Aluna: Amanda Cruz Francesconi

Orientador: Jesus Mena

Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- Introdução
- Conceitos básicos
- Método
- Resultados
- Conclusão

Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- **Introdução**
- Conceitos básicos
- Método
- Resultados
- Conclusão

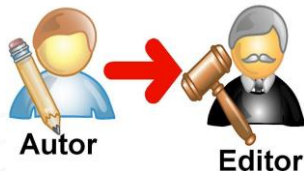
Introdução

Contexto



Pesquisa científica

Revisão por pares



Somente a FAPESP faz a análise de **19 mil projetos** de pesquisa por ano

38 mil horas gastas
R\$3.8M*

Esse projeto busca construir uma ferramenta que auxilie em uma etapa do processo da pesquisa científica: **a atribuição de avaliadores para que seja realizada a revisão por pares.**

*considerando o tempo de 2h para decisão de avaliadores e R\$100 reais/hora

Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- Introdução
- **Conceitos básicos**
- Método
- Resultados
- Conclusão

Conceitos básicos

Revisão por pares

PLN

Grafos

Emparelhamento

A revisão por pares é o procedimento utilizado no meio científico para **avaliação de trabalhos**, sejam eles artigos, projetos de pesquisa, etc.

Conceitos básicos

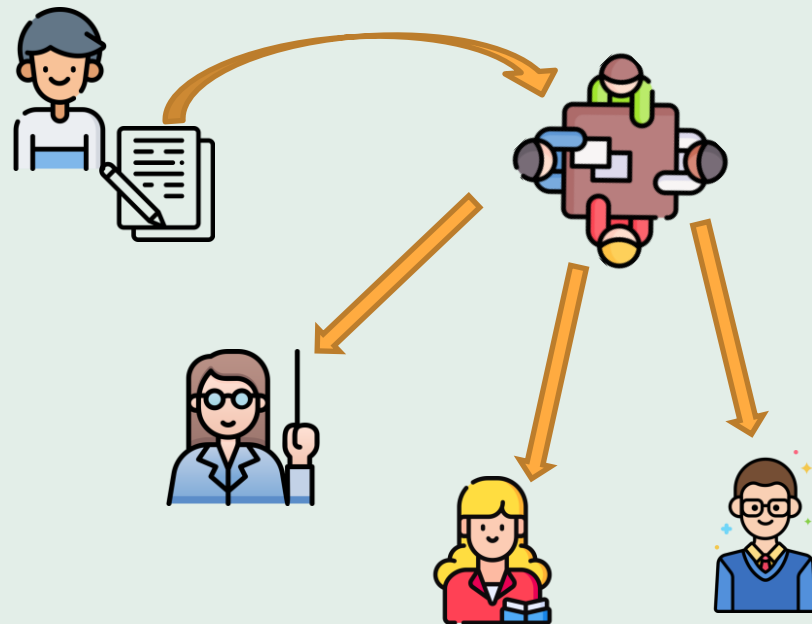
Revisão por pares

PLN

Grafos

Emparelhamento

A revisão por pares é o procedimento utilizado no meio científico para **avaliação de trabalhos**, sejam eles artigos, projetos de pesquisa, etc.



Conceitos básicos

Revisão por pares

PLN

Grafos

Emparelhamento

Radicalização (stemming)

Técnica para extrair informação principal da palavra
reduzindo-a ao seu radical

Estudo » Estud
Estudante » Estud
Estudar » Estud

Saco-de-palavras (bag-of-words)

Técnica para simplificação de textos onde somente o conteúdos das palavras importa, sendo desconsiderados estrutura gramatical e ordem das palavras

“Como avaliar
esse
trabalho?”



Bigramas

São sequências de 2 palavras consecutivas em uma frase, utilizado para facilitar análise textuais

“Como avaliar esse trabalho?”

- o como avaliar
- o avaliar esse
- o esse trabalho

Conceitos básicos

Revisão por pares

PLN

Grafos

Emparelhamento

Grafos são estruturas

Conceitos básicos

Revisão por pares

PLN

Grafos

Emparelhamento

Grafos são estruturas
formados por vértices

Diretor
A

Diretor
B

Diretor
C

Ator
A

Ator
B

Ator
C

Ator
D

Ator
E

Conceitos básicos

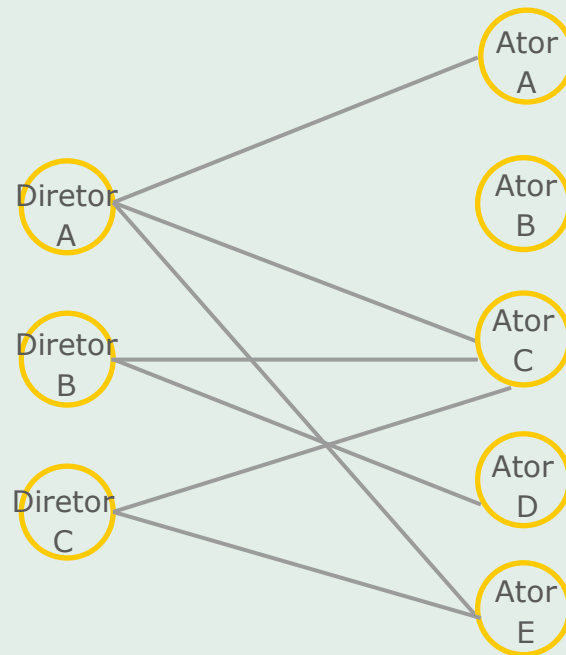
Revisão por pares

PLN

Grafos

Emparelhamento

Grafos são estruturas formados por vértices e arestas.



Conceitos básicos

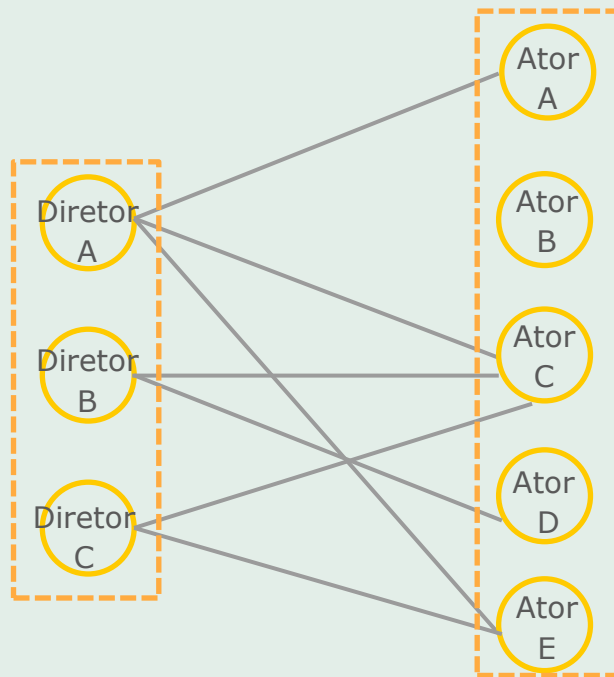
Revisão por pares

PLN

Grafos

Emparelhamento

Grafos são estruturas formados por vértices e arestas.
Pode ser um grafo bipartido



Conceitos básicos

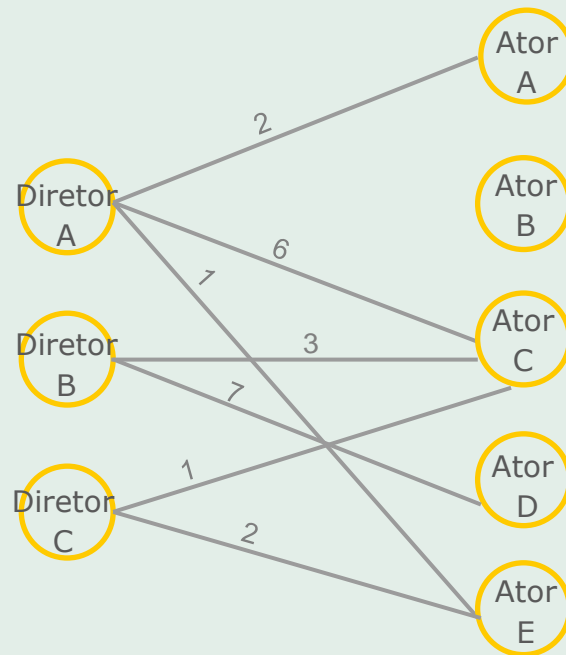
Revisão por pares

PLN

Grafos

Emparelhamento

Grafos são estruturas formados por vértices e arestas.
Pode ser um grafo bipartido e ponderado.



Conceitos básicos

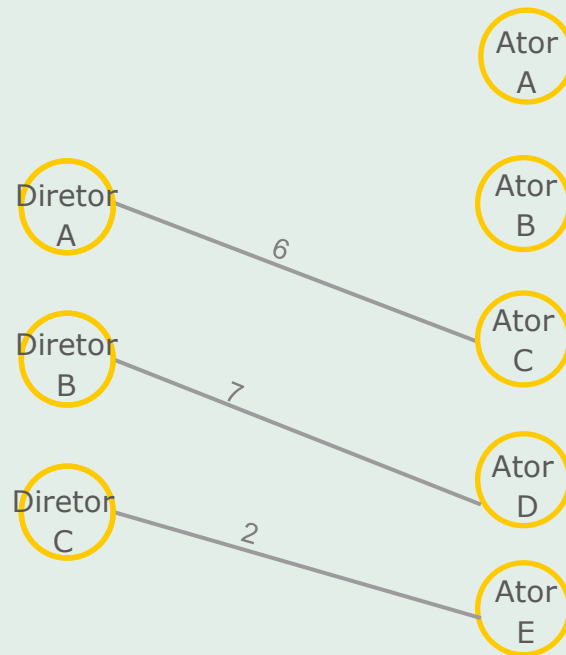
Revisão por pares

PLN

Grafos

Emparelhamento

Emparelhamento é um conjunto de arestas de um grafo em que todos os vértices estão conectados 2 a 2 e nenhuma aresta é adjacente.



Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- Introdução
- Conceitos básicos
- **Método**
- Resultados
- Conclusão

Método

Avaliadores

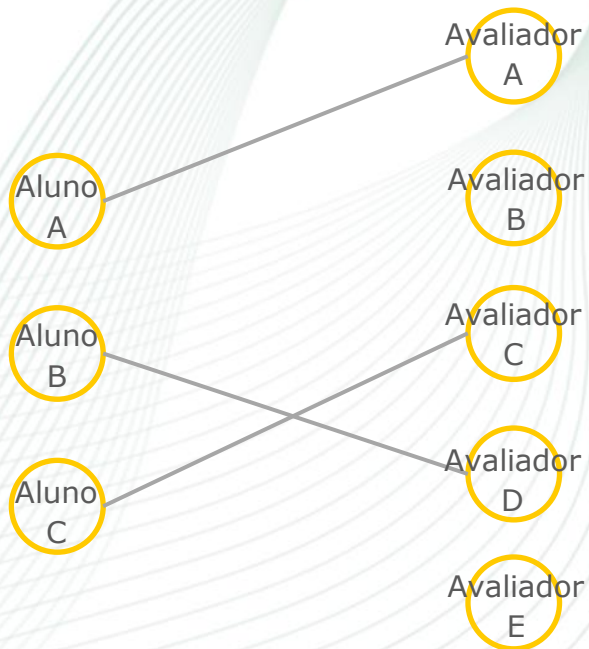
- Nome
- Artigos publicados
- Livros e capítulos de livros

Projetos

- Nome aluno
- Título
- Palavras-chave
- Resumo



Input



Output

Método - PLN

Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,
cognição



Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,
cognição

Avaliador: Bruno Oliveira

Método - PLN

Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,
cognição



Remoção de stop words

Livro: Ensino
computação ensino
médio
Artigo:
aprendizado
algoritmos auxilia
desenvolvimento cognitivo
adolescentes
Palavras-chave: ensino,
cognição

Avaliador: Bruno Oliveira

Método - PLN

Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,
cognição



Remoção de palavras frequentes

Livro:
computação
médio
Artigo:
aprendizado
auxilia desenvolvimento
cognitivo adolescentes
Palavras-chave:
Cognição

Avaliador: Bruno Oliveira

Método - PLN

Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,
cognição



Remoção de acentos e pontuação

Livro:
computacao
medio
Artigo:
aprendizado
auxilia desenvolvimento
cognitivo adolescentes
Palavras-chave:
Cognicao

Avaliador: Bruno Oliveira

Método - PLN

Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,
cognição



Stemming

Livro:
computac
med
Artigo:
Aprend
auxil desenvolvim
cogn adolescent
Palavras-chave:
Cogn

Avaliador: Bruno Oliveira

Método - PLN

Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,
cognição

Tratamentos textuais

Livro:
computac
med
Artigo:
Aprend
auxil desenvolv
cogn adolescent
Palavras-chave:
Cogn

Avaliador: Bruno Oliveira

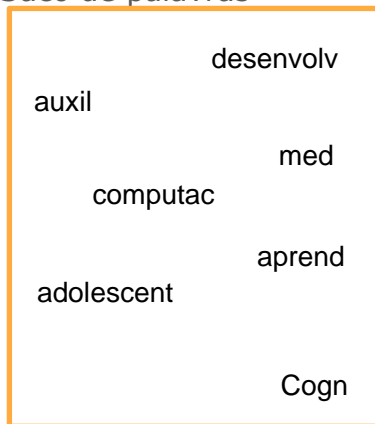
Saco de palavras

desenvolv
auxil
computac
adolescent
med
aprend
Cogn

Avaliador: Bruno Oliveira

Método – Criação do grafo

Saco de palavras



Avaliador: Bruno Oliveira

Avaliador
A

ID nó: Bruno Oliveira

Bipartição: avaliador

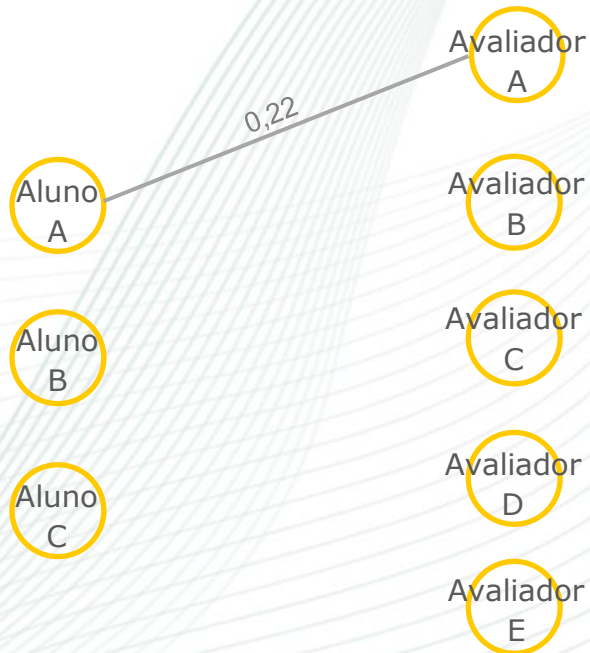
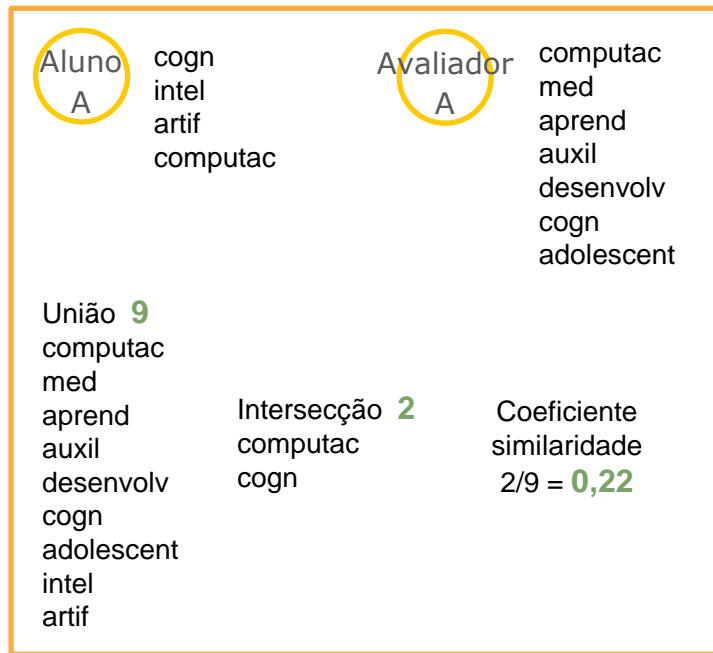
Saco-de-palavras:

computac
med
Aprend
auxil
desenvolv
cogn
adolescent

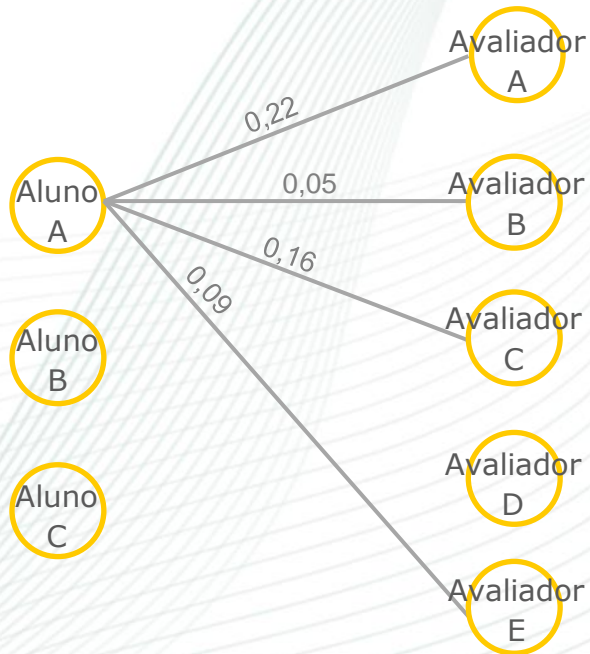
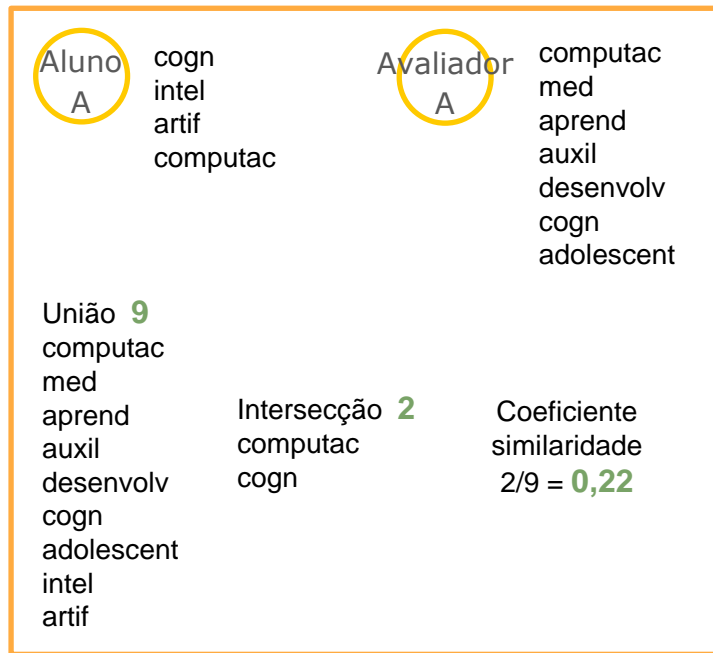
Método – Criação do grafo



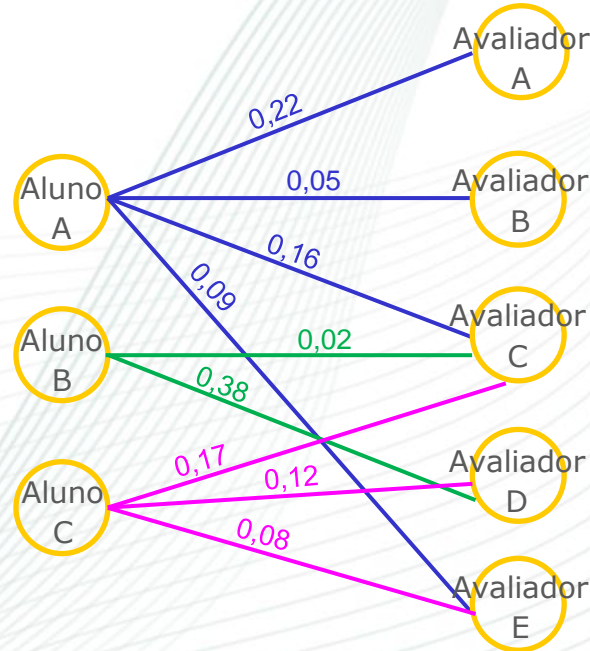
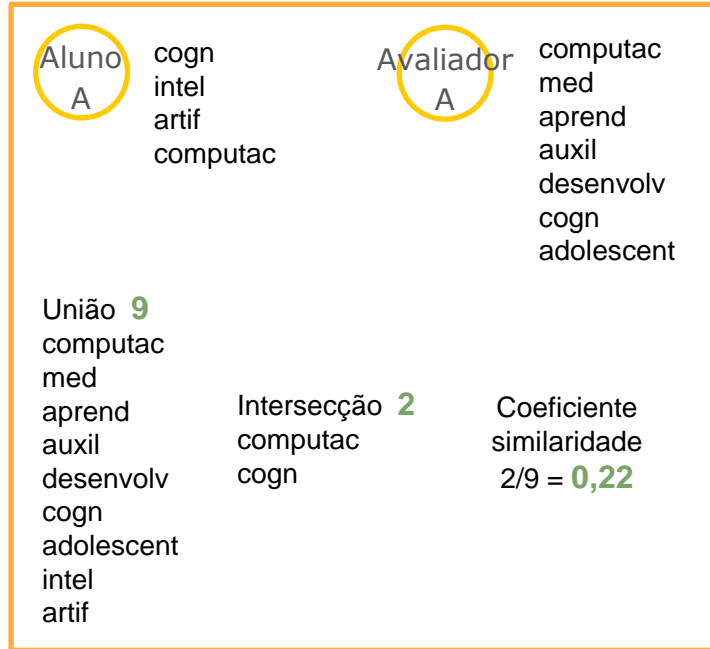
Método – Criação do grafo



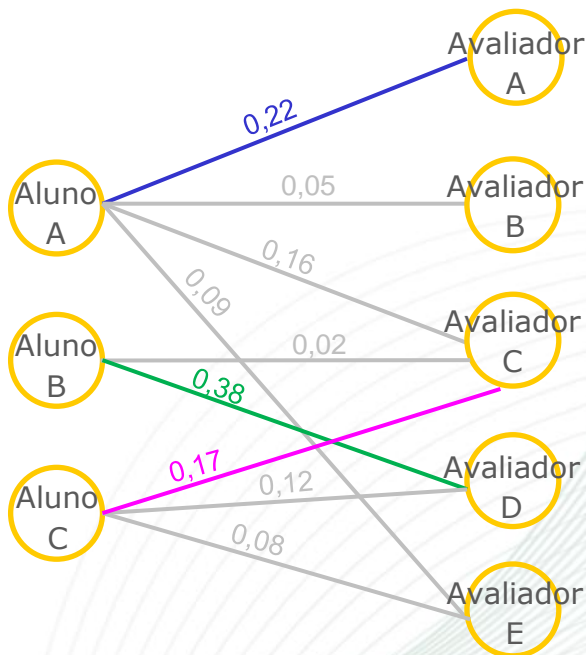
Método – Criação do grafo



Método – Criação do grafo



Método – Emparelhamento



Algoritmo de Hopcroft–Karp

$M=0$

enquanto houver caminho extensor para M

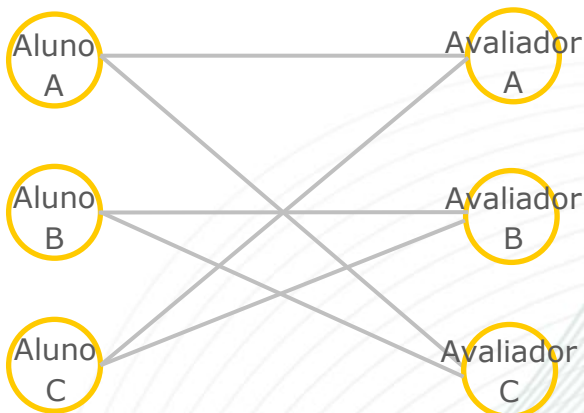
Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M

Método – Emparelhamento

1ª iteração



Algoritmo de Hopcroft-Karp

$M=0$

enquanto houver caminho extensor para M

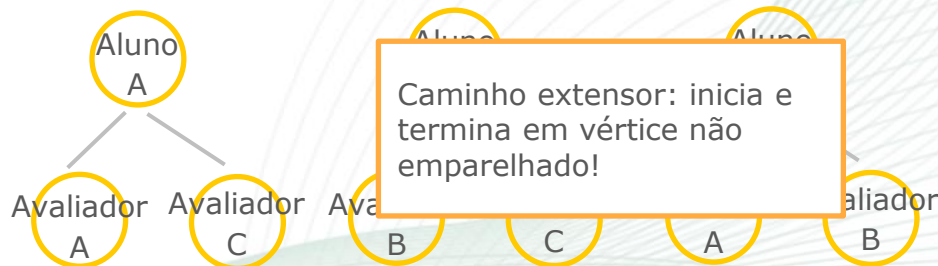
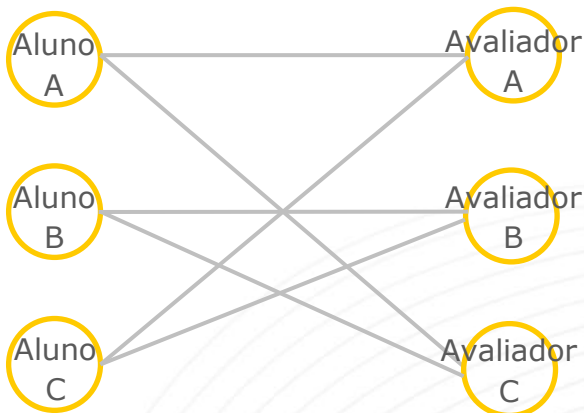
Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M

Método – Emparelhamento

1ª iteração



Caminho extensor: inicia e termina em vértice não emparelhado!

Algoritmo de Hopcroft–Karp

$M=0$

enquanto houver caminho extensor para M

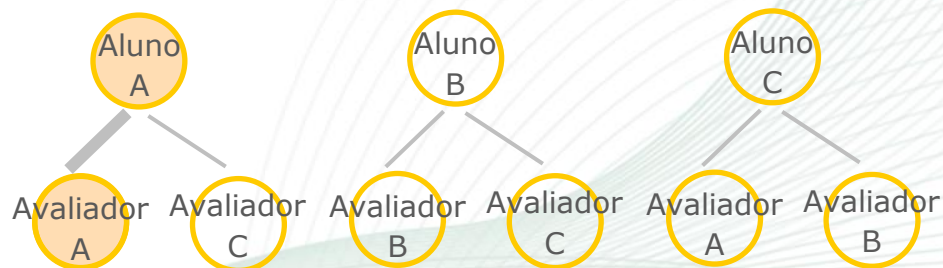
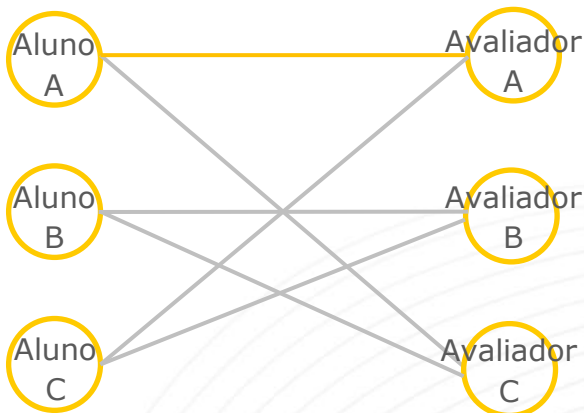
Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M

Método – Emparelhamento

1ª iteração



Algoritmo de Hopcroft-Karp

$M=0$

enquanto houver caminho extensor para M

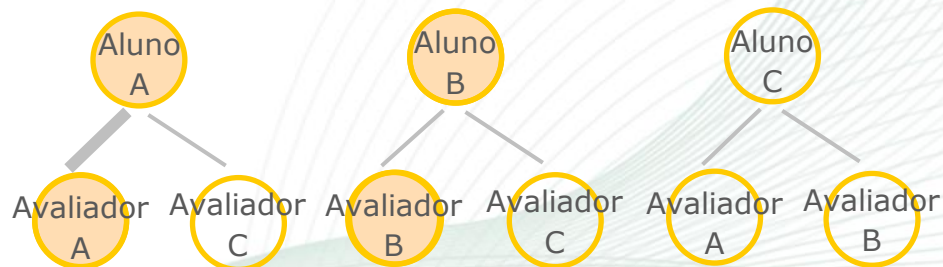
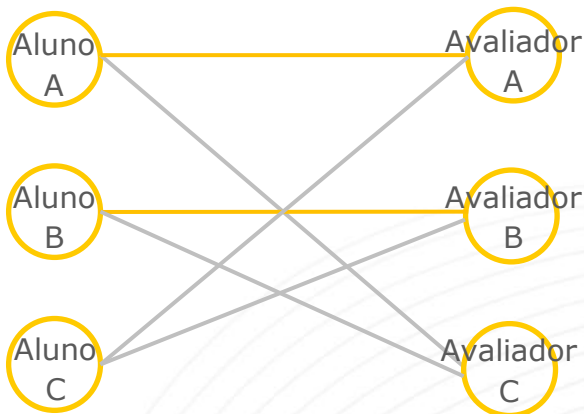
Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M

Método – Emparelhamento

1ª iteração



Algoritmo de Hopcroft-Karp

$M=0$

enquanto houver caminho extensor para M

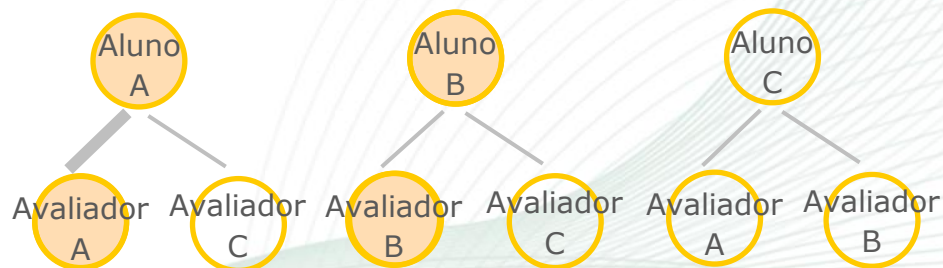
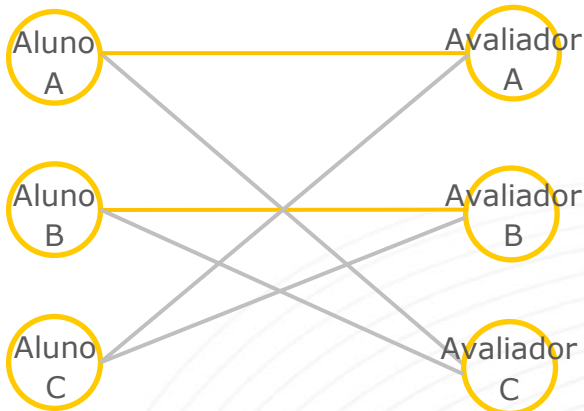
Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M

Método – Emparelhamento

1ª iteração



Algoritmo de Hopcroft-Karp

$M=0$

enquanto houver caminho extensor para M

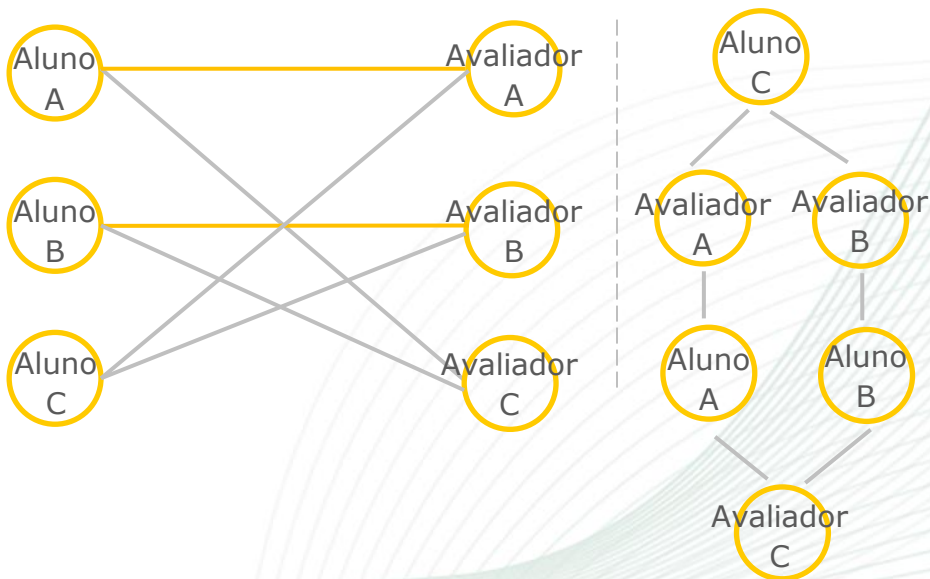
Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos **disjuntos** em G que aumentem o emparelhamento atual de M

retornar M

Método – Emparelhamento

2ª iteração



Algoritmo de Hopcroft-Karp

$M=0$

enquanto houver caminho extensor para M

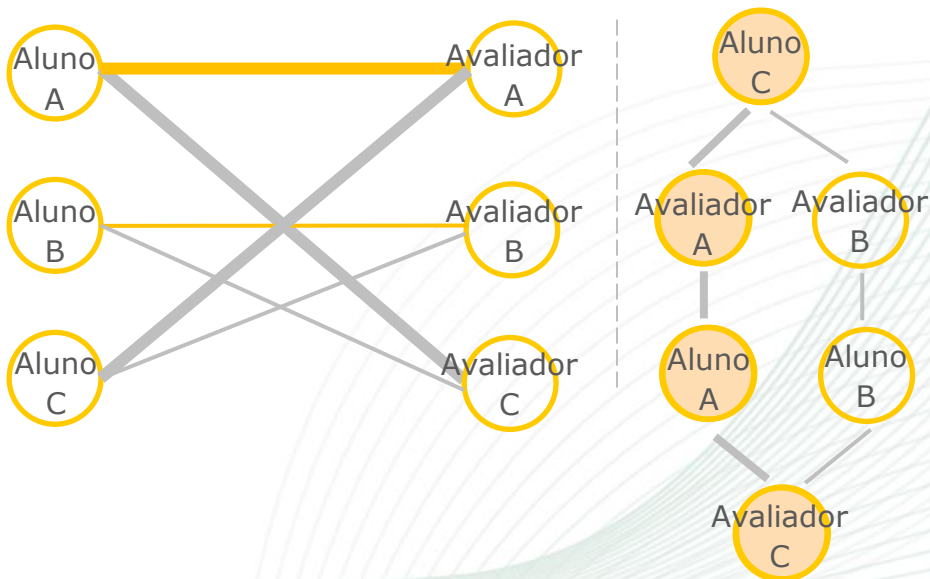
Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M

Método – Emparelhamento

2ª iteração



Algoritmo de Hopcroft-Karp

$M=0$

enquanto houver caminho extensor para M

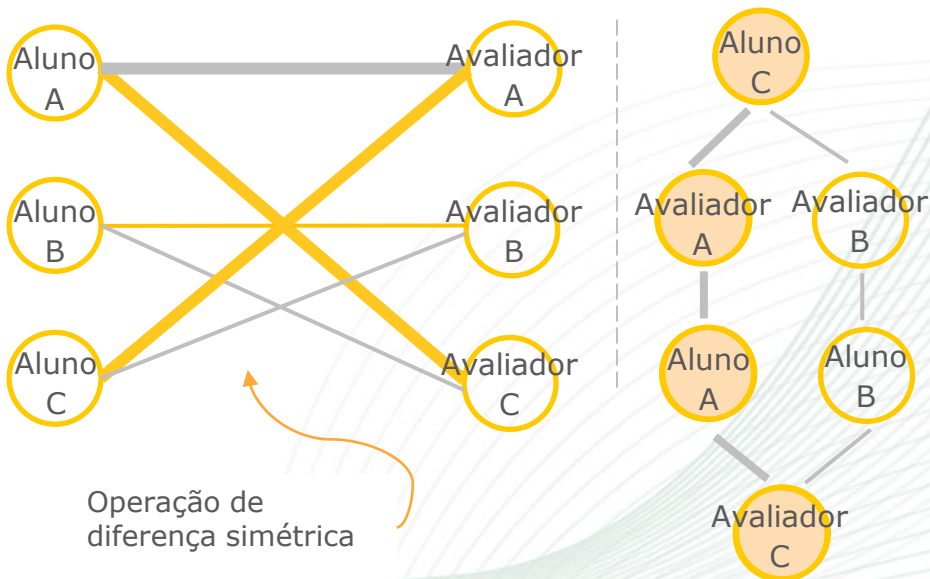
Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M

Método – Emparelhamento

2ª iteração



Algoritmo de Hopcroft-Karp

$M=0$

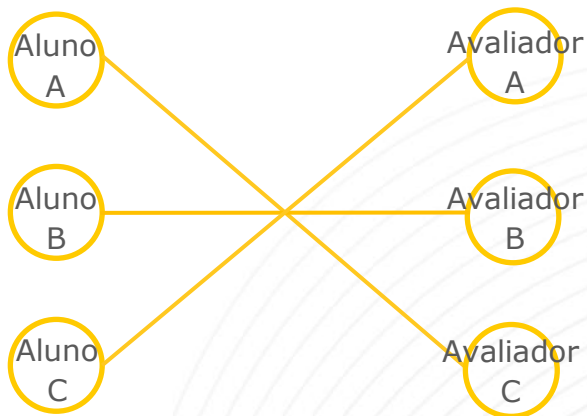
enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M

Método – Emparelhamento



Algoritmo de Hopcroft-Karp

$M=0$

enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M

Método – Emparelhamento

Análise complexidade

Complexidade total:

$$O(|E| * \sqrt{|V|})$$

$$O(\sqrt{|V|})$$

$$O(|E|)$$

Algoritmo de Hopcroft–Karp

$M=0$

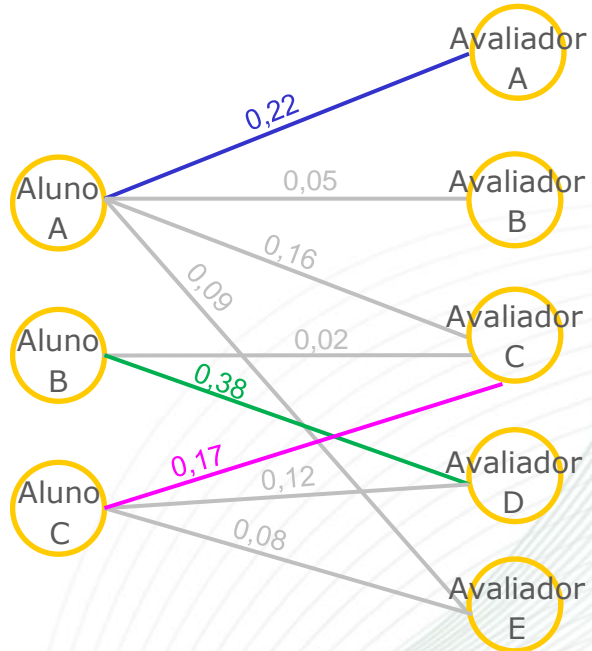
enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M

Método – Emparelhamento



SUGESTÃO DE AVALIADORES PARA PROJETOS

Aluno A	Avaliador A
Aluno B	Avaliador D
Aluno C	Avaliador C

Método – Complexidade do código

Imports e definições:

$O(1)$

Captura de dados:

$O(\text{palavras_proj} + \text{palavras_aval})$

Banco palavras:

$O(\text{retirar})$

Tratamento PLN:

$O((\text{palavras_proj} + \text{palavras_aval}) * \text{retirar})$

Criação dos vértices:

$O(\text{palavras_proj} + \text{palavras_aval})$

Criação das arestas:

$O(\text{palavras_proj} * \text{palavras_aval})$

Emparelhamento:

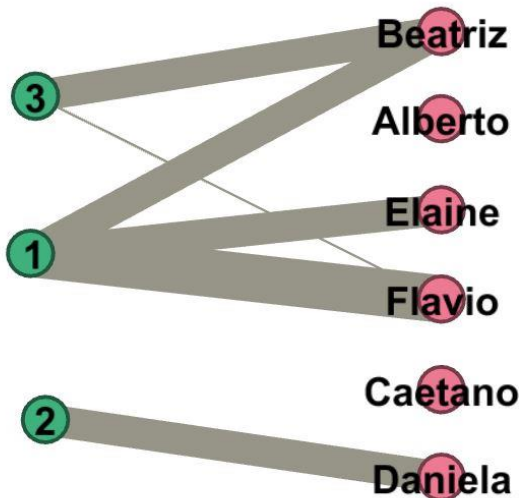
$O(\text{arestas} * \sqrt{(\text{vértices})})$

Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- Introdução
- Conceitos básicos
- Método
- **Resultados**
- Conclusão

Resultados – teste emparelhamento

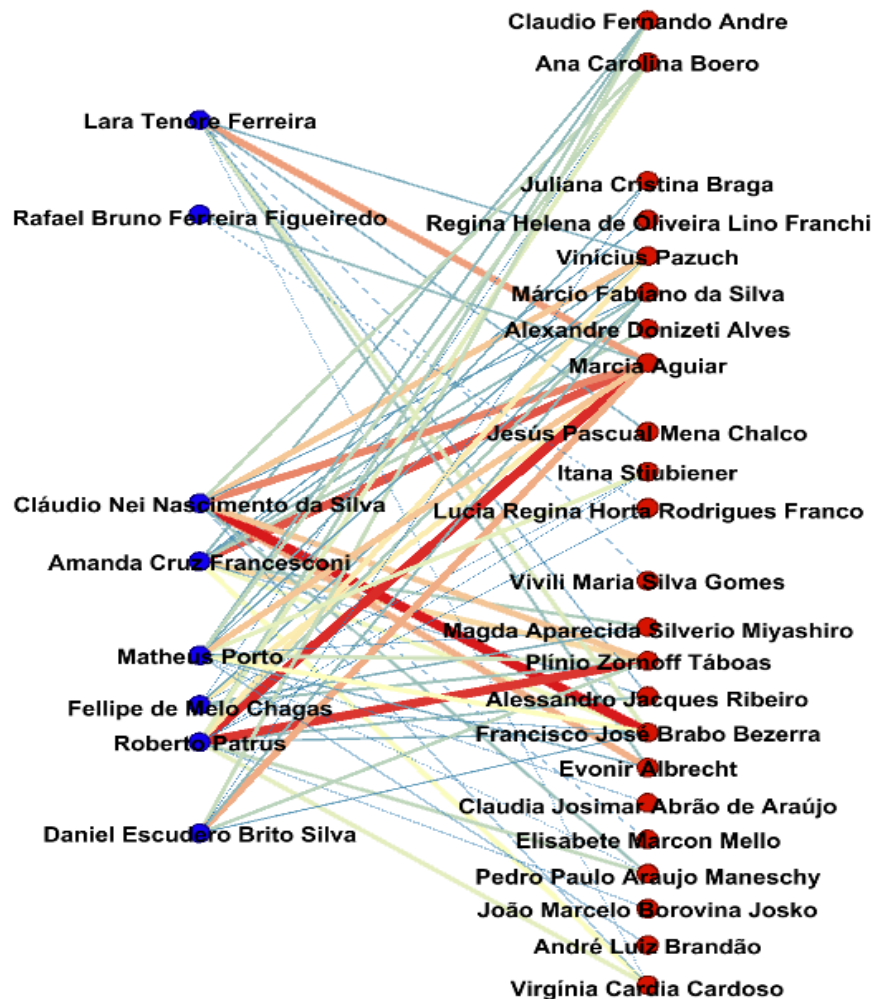


Bipartição 1	Bipartição 2
1	Flavio
2	Daniela
3	Beatriz

O teste realizado apresentou o resultado esperado

Resultados – teste sem PLN

O teste realizado sem nenhum tratamento apresentou um número grande de conexões e altos coeficientes de similaridade



Resultados – teste sem PLN

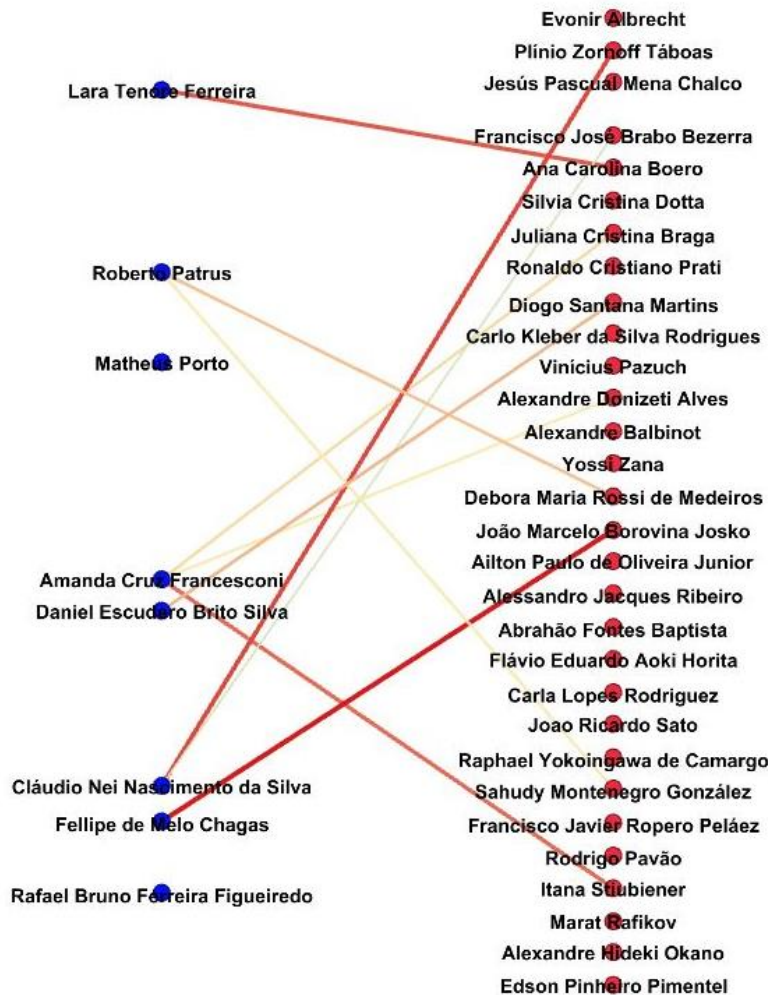


Autor projeto	Avaliador sugerido	Peso	Termos em comum
Amanda Cruz Francesconi	Virgínia Cardia Cardoso	0.076	uma, que, vez, a, estudo, em, é, um, para, com, o, no, da, de, projeto, ensino, como, na, processo, e, abordagem, número
Cláudio Nei Nascimento da Silva	Francisco José Brabo Bezerra	0.093	uma, dificuldades, que, A, a, estudo, os, da, partir, em, um, dos, com, o, no, sobre, as, de, onde, na, processo, do, e
Daniel Escudero Brito Silva	Claudio Fernando Andre	0.072	em, de, como, estado, na, para, com, o, computacional, A, O, e, a, da, conhecimento
Fellipe de Melo Chagas	Magda Aparecida Miyashiro	0.067	de, um, utilizando, para, o, a, da
Lara Tenore Ferreira	Marcia Aguiar	0.084	uma, desenvolvido, que, história, desenvolvimento, A, a, em, partir, tarefa, O, para, com, o, no, da, de, como, contexto, na, das, do, e, por
Matheus Porto	Itana Stiubiener	0.075	em, uma, de, do, e, a, da
Rafael Bruno Ferreira Figueiredo	Lucia Regina Rodrigues Franco	0.061	uma, sistema, a, em, controle, O, um, dos, para, com, o, 3, da, as, rede, de, na, problema, do, e, aplicações
Roberto Patrus	Plínio Zornoff Táboas	0.091	de, dos, para, na, História, o, no, e, sobre, da, as,



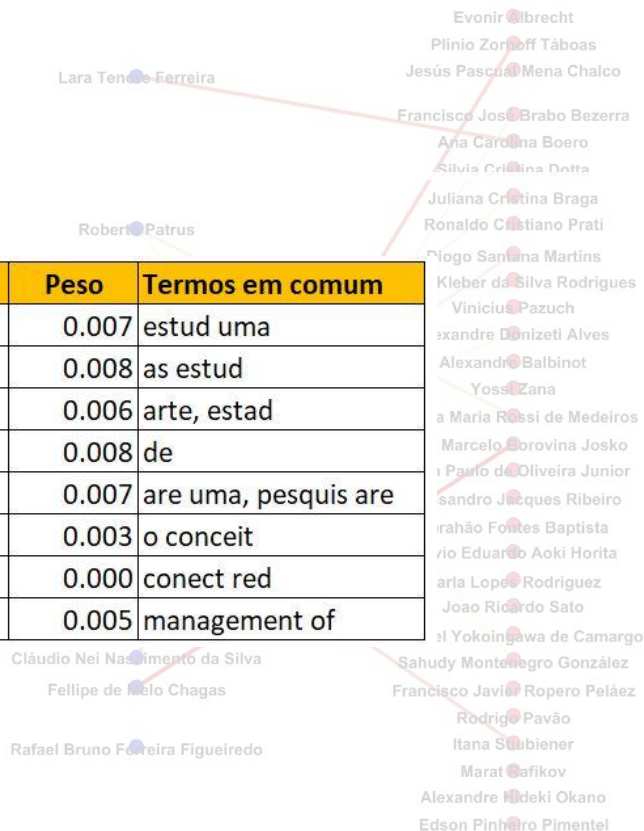
Resultados – teste bigramas

O teste realizado com tratamento de PLN e utilizando bigramas apresentou um número baixo de conexões



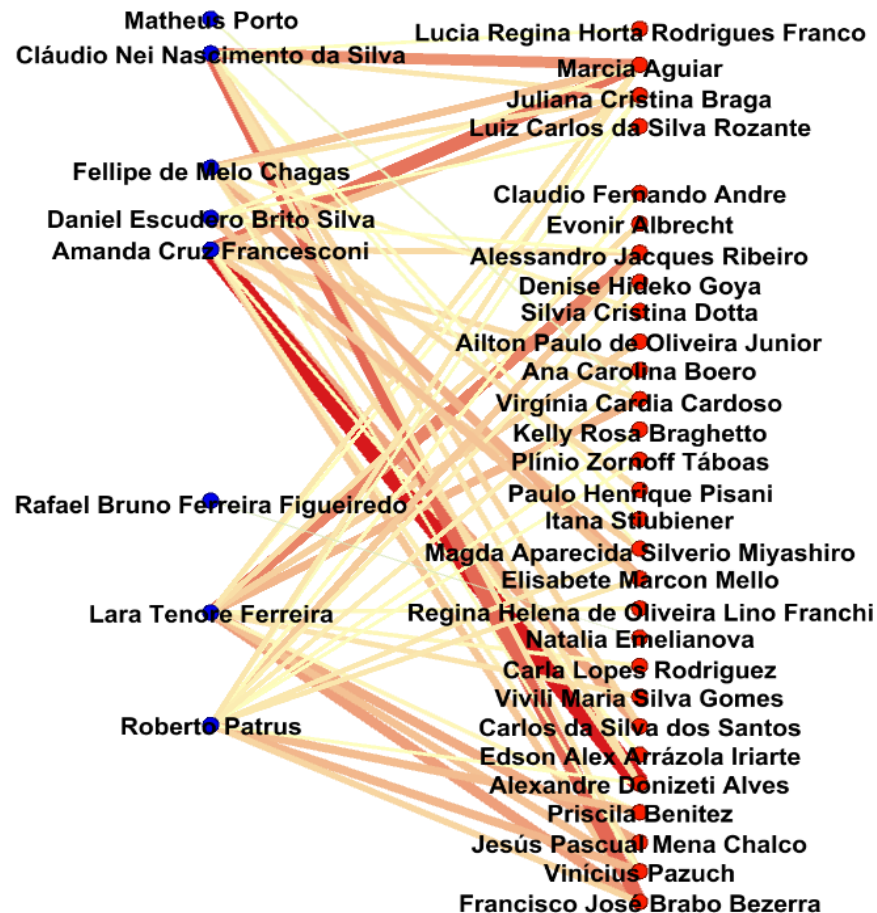
Resultados – teste bigramas

Autor projeto	Avaliador sugerido	Peso	Termos em comum
Amanda Cruz Francesconi	Itana Stiubiener	0.007	estud uma
Cláudio Nei Nascimento da Silva	Plínio Zornoff Táboas	0.008	as estud
Daniel Escudero Brito Silva	Diogo Santana Martins	0.006	arte, estad
Fellipe de Melo Chagas	João Marcelo Borovina Josko	0.008	de
Lara Tenore Ferreira	Ana Carolina Boero	0.007	are uma, pesquis are
Matheus Porto	Evonir Albrecht	0.003	o conceit
Rafael Bruno Ferreira Figueiredo	Joao Ricardo Sato	0.000	conect red
Roberto Patrus	Debora Maria Rossi de Medeiros	0.005	management of



Resultados – teste com PLN e palavras

O teste realizado com tratamento de PLN e utilizando somente palavras apresentou um número intermediário de conexões e coeficientes de similaridade



Resultados – teste com PLN e palavras



Autor projeto	Avaliador sugerido	Peso	Termos em comum
Amanda Cruz Francesconi	Alexandre Donizeti Alves	0.079	especif, bas, linguag, estud, are, ferrament, natural, process
Cláudio Nei Nascimento da Silva	Francisco José Brabo Bezerra	0.065	exist, conheç, part, estud, signific, onde, compar, dificultad, process
Daniel Escudero Brito Silva	Elisabete Marcon Mello	0.050	especial, constru, propost, registr
Fellipe de Melo Chagas	Marcia Aguiar	0.050	curricul, bas, pesquis, conheç, process
Lara Tenore Ferreira	Alessandro Jacques Ribeiro	0.061	calcul, compreensa, histor, divers, futur, busc, conjunt, signific, context, projet, impact, pesquis, consider, part, estud, taref, ii, relaco, process
Matheus Porto	Virgínia Cardia Cardoso	0.037	filosof, profund, leitur, critic, construca
Rafael Bruno Ferreira Figueiredo	Natalia Emelianova	0.034	encontr, pont, simpl, font, red, uso, ambient
Roberto Patrus	Vinícius Pazuch	0.052	pesquis, of, histor, review, cientif, uso, teoric, artig, process

Edson Alex Arrázola Iriarte
 Alexandre Donizeti Alves
 Priscila Benitez
 Jesús Pascual Mena Chalco
 Vinícius Pazuch
 Francisco José Brabo Bezerra

Resultados – teste com dados reais

O teste realizado para uma aplicação real de atribuição considerou todo o tratamento de PLN e o uso de somente palavras. Os ajustes feitos em relação aos resultados anteriores:

- Três avaliadores por projeto
- Máximo de dois projetos por avaliador

Candidato	Avaliador 1	Similaridade	Termos em comum	Avaliador 2	Similaridade	Termos em comum	Avaliador 3	Similaridade	Termos em comum
1	A	0.1267	118	E	0.1128	94	A C	0.1116	125
2	B	0.0859	70	A L	0.054	65	Y	0.0988	131
3	C	0.0682	121	A M	0.0997	158	G	0.1236	203
4	D	0.095	62	L	0.0853	49	A B	0.0647	29
5	E	0.0947	76	A	0.0965	88	O	0.0787	143
6	F	0.0702	71	R	0.1239	163	A Q	0.0918	147
7	G	0.1208	186	P	0.0518	69	I	0.0939	131
8	H	0.0947	82	A H	0.1117	79	A H	0.1119	79
9	I	0.0908	114	X	0.0642	76	-	-	-
10	J	0.0516	132	C	0.0595	115	AA	0.0634	95

Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- Introdução
- Conceitos básicos
- Método
- Resultados
- **Conclusão**

Conclusão

Melhor resultado

Grafo gerado com **todas as técnicas de PLN** e palavras ao invés de bigramas apresentou resultado satisfatório para o objetivo proposto com **boa correlação** entre avaliadores e projetos de pesquisa.

Ao ser testado em uma **aplicação real** de atribuição de avaliadores conseguiu **auxiliar na tomada de decisão** do órgão responsável.

Possíveis melhorias

- Input manuais para ajustes finos
- Software para facilitar o acesso a ferramenta

Contribuição

Método inusitado na literatura para resolução de problema ainda em aberto e que ocorre em abundância na atualidade.

Referências

- Bond, J. A., & Murty, U.S. 1976. Graph theory with applications. The Macmillan Press Ltd.
- Digiampietri, Luciano Antonio, Mena-Chalco, Jesus Pascual, Perez-Alcazar, Jose de Jesus, Fernandez Tuesta, Esteban, Delgado, Karina Valdivia, Mugnaini, Rogerio, & Silva, Gabriela S. 2012. Minerando e caracterizando dados de curriculos lattes. Pages 1–12 of: Brazilian workshop on social network analysis and mining – brasnam
- Erciyes, Kayhan. 2015. Complex networks: An algorithmic perspective. CRC Press.
- Fernando A. Ribeiro Serra, Manuel Portugal Ferreira, Gabriela Gonçalves Fiates Lenise Costa. 2007. Publicar e difícil ou faltam competências? o desafio de pesquisar e publicar em revistas científicas na visão de editores e revisores internacionais. Encontro de ensino e pesquisa em administração e contabilidade, 9(4).
- Jurafsky, Daniel, & Martin, James. 2000. Speech and language processing. Prentice Hall.



Obrigada!

12/05/21



Defesa Projeto de Graduação em Computação

Aluna: Amanda Cruz Francesconi

Orientador: Jesus Mena

Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo