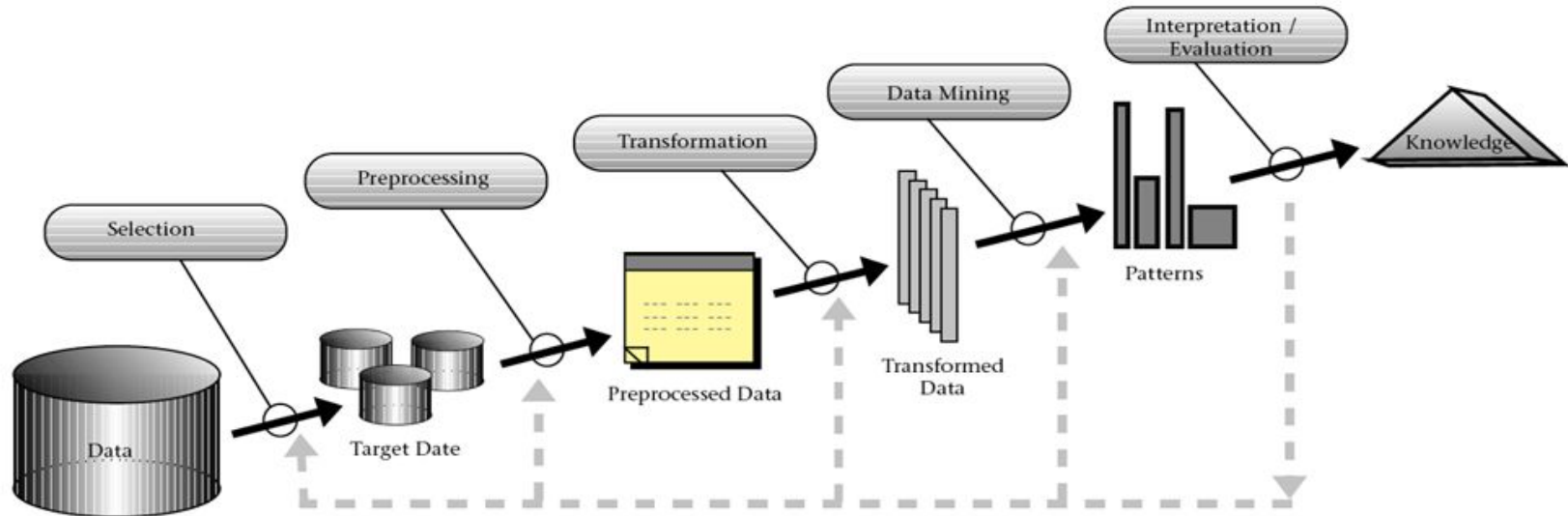


IA e Ciência de dados nas relações de coautoria nos PPGCC no Brasil

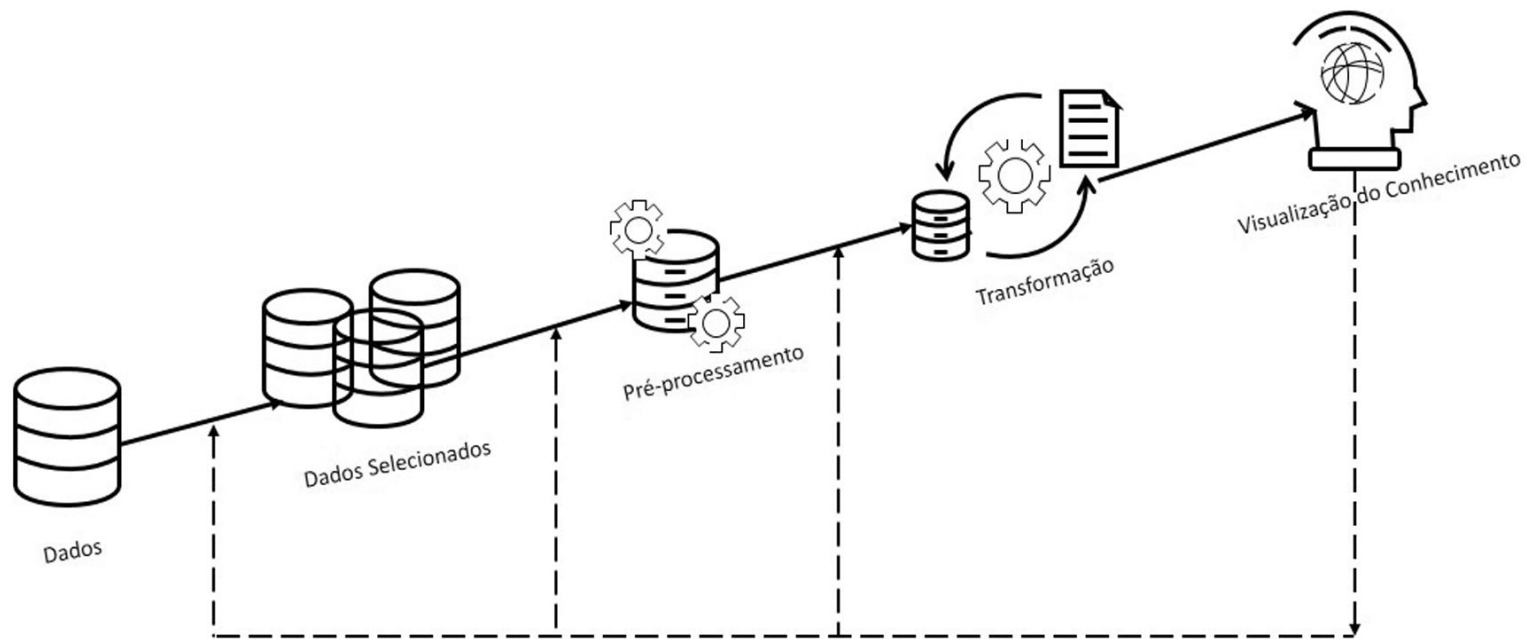
Jeoildo Pereira
Prof. Dr. Reginaldo Cordeiro dos Santos Filho

- Analisar relações de coautoria nos PPGCC no Brasil utilizando dados do currículo lattes.
- Visualizar as relações de coautoria entre os professores
- Mapear as relações entre PPGCC com base nas colaborações acadêmicas.

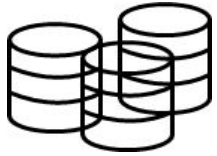
Knowledge Discovery in Databases - KDD (geral)



KDD nas relações de coautoria nos PPGCC



Knowledge Discovery in Databases - KDD ...cont



→ **Seleção dos dados**



Universidades e Instituto: 65
PPGCC no Brasil: 67
Docentes PPGCC: 1810



Currículo
Lattes



Currículo lattes dos docentes PPGCC: 1810
(*feramenta de automatização*)

Knowledge Discovery in Databases - KDD ...cont



→ **Pre-processamento dos dados**



Currículo
Lattes

Curriculo.xml // exemplo:

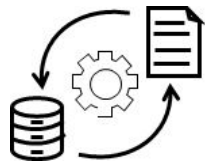
<http://lattes.cnpq.br/915742238690032> → Prof. Reginaldo

<http://lattes.cnpq.br/4742268936279649> → Prof. Claudomiro

Dados-gerais, produção-bibliográfica, produção-técnica, outra-produção e dados-complementares

Dados-gerais,
Produção-bibliográfica → 2014 a 2023

Knowledge Discovery in Databases - KDD ...cont



→ **Transformação dos dados**

- 1. Normalização dos nomes dos professores**
(padronizar, resolver inconsistências, eliminar dados duplicado)
- 2. Criação da relações de coautoria**
(Identificar, criar matriz de coautoria, construir um grafo)
- 3. Filtragem de dados relevantes**
(Remover artigos sem coautoria, publicações dentro de período 2014 a 2023)
- 4. Conversão para um formato de análise de redes**
(Lista de Arestas e matriz de adjacência)
- 5. Cálculo de métricas da rede de coautoria**
(Grau de cada autor, grau de coautoria entre ins/prog., densidade da rede, e centralidade)

Knowledge Discovery in Databases - KDD ...cont



→ **Visualização dos conhecimento**

1. Diagrama de cordas
2. Diagrama de sankey
3. Diagrama de barra
4. Diagrama de linha
5. Diagrama de rede (grafo)

Diagrama de cordas (chord diagram)

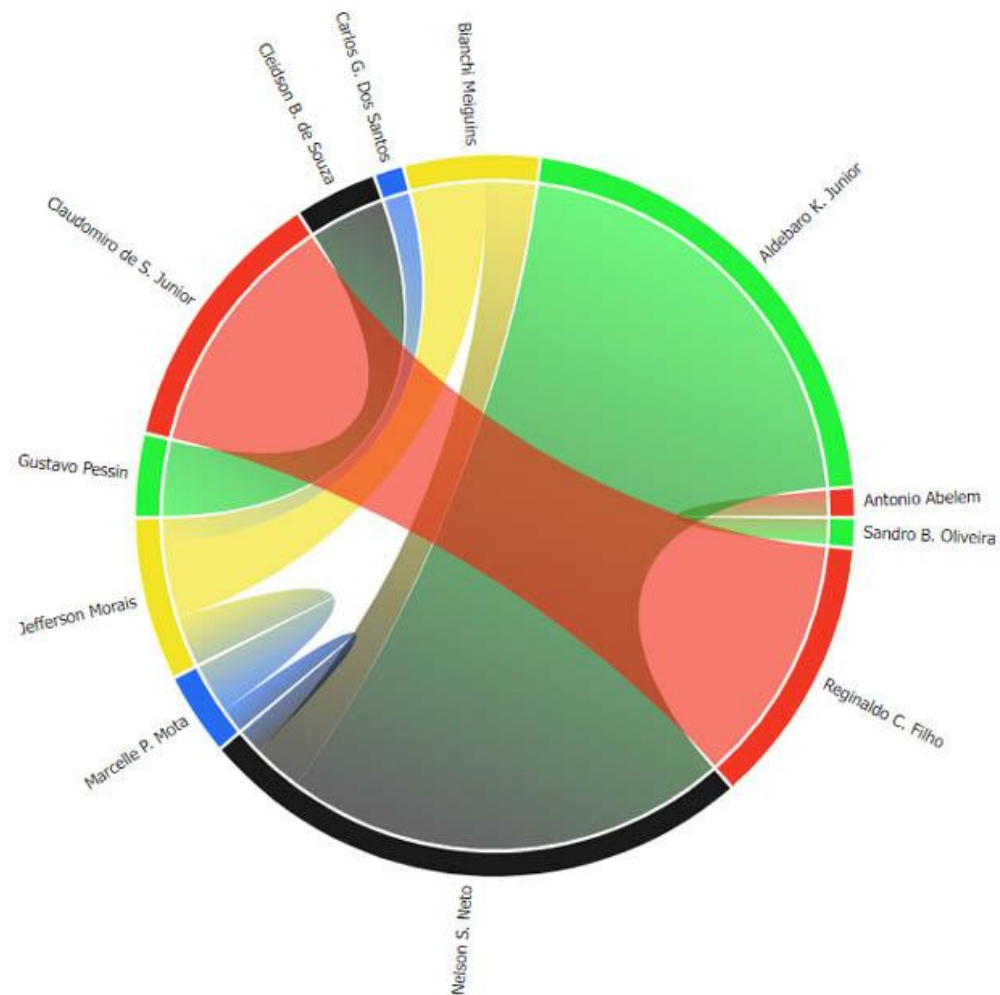


Diagrama de sankey

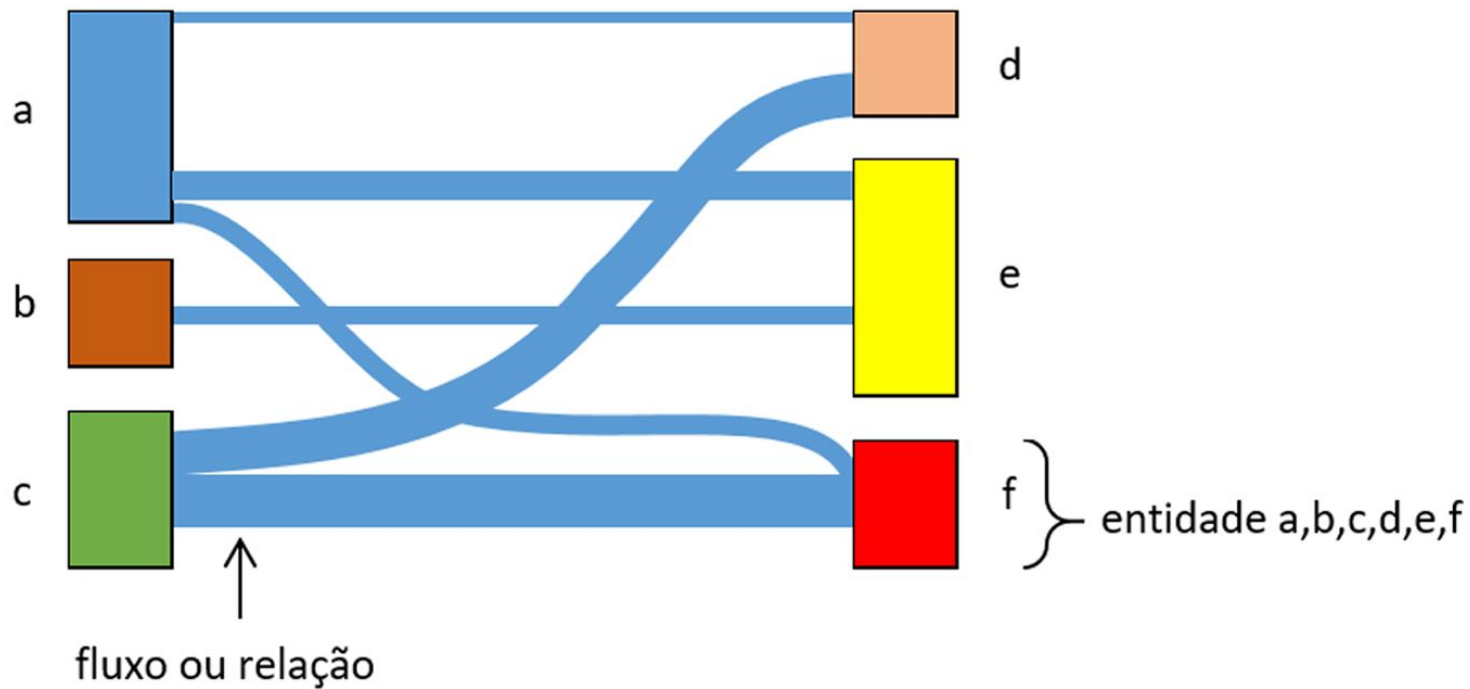


Diagrama de barra

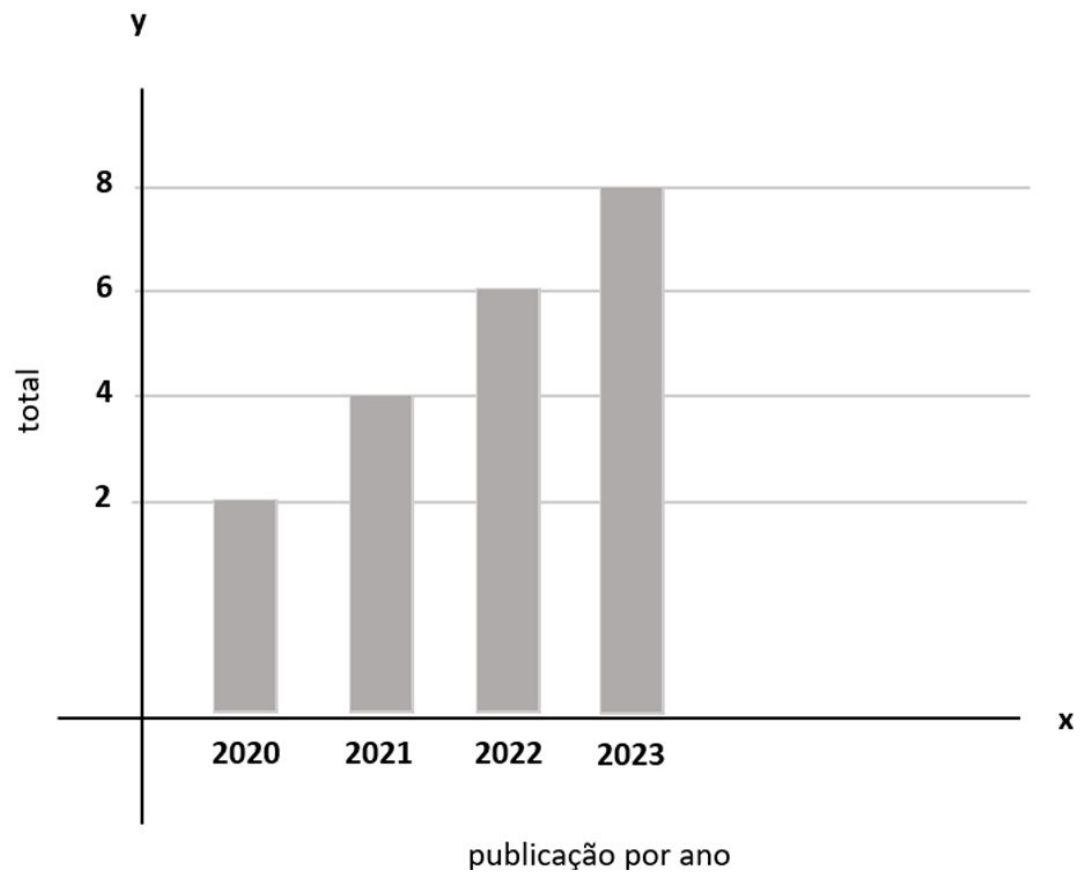


Diagrama de linha

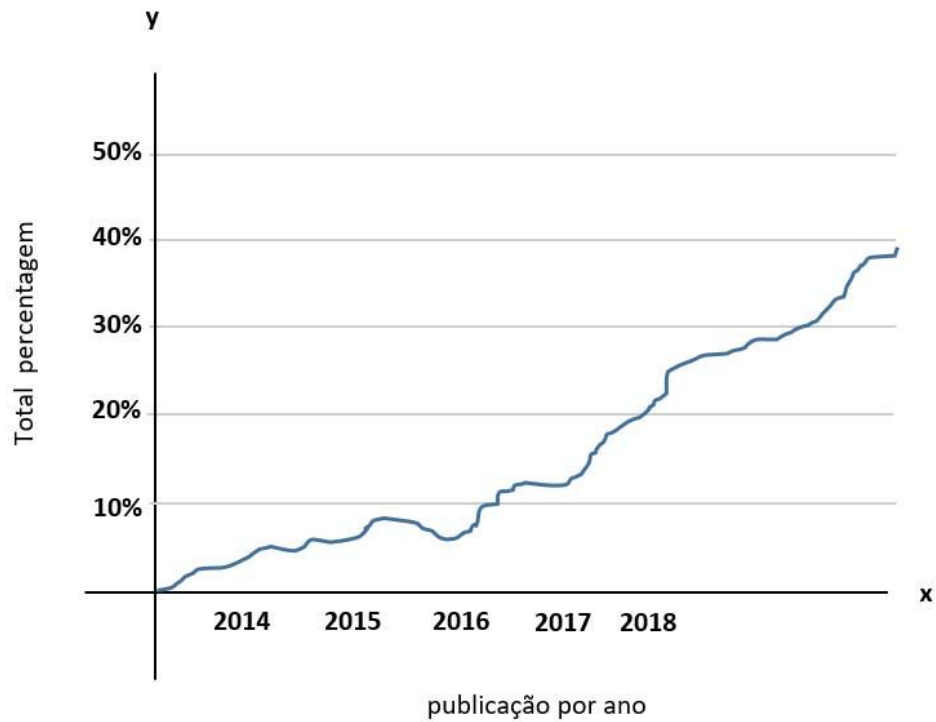
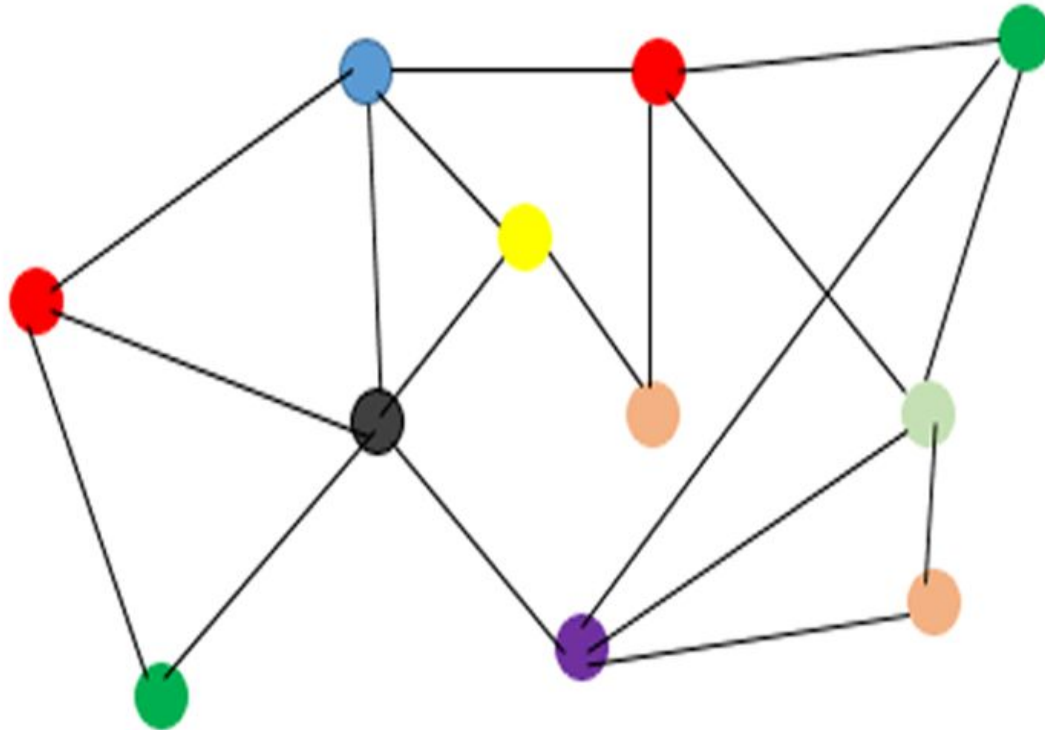


Diagrama de rede (grafo)



IA nas relações de coautoria nos PPGCC no Brasil

Como IA pode ser usada :

1. Machine Learning

- Agrupamento de autores com base em padrões de colaboração
- Previsão de colaborações futuras

2. Graph Neural Networks (GNNs)

- Análise profunda da estrutura e relações na rede de coautoria

3. NLP (Processamento de Linguagem Natural)

- Análise dos títulos e palavras-chave das publicações

4. Detecção de Anomalias

- Identificação de padrões de colaboração incomuns

Conclusão

1. Contribui significativamente para o entendimento das dinâmicas de colaboração científica entre pesquisadores/professores nos programas de pós-graduação em ciência da computação (PPGCC) no Brasil.
2. Contribuir no desenvolvimento de uma pipeline de dados robusta para análise de coautoria acadêmica.
3. Essa abordagem facilita a construção de redes de colaboração e pode ser adaptada para outras áreas do conhecimento.
4. Visualização como diagrama de cordas e diagrama de sankey que tornam mais clara a compreensão das relações de coautoria e instituições, universidades ou programas.

Obrigado barak / Muito obrigado