



# Visual Data Mining

Banco de Dados II

Prof. Dr. Carlos Roberto Valêncio

Murilo de Sales Menezes

RA: 151044414

Paulo Vitor de Queiroz Zanele

RA: 151044244



# Índice

- Objetivo
- Introdução
- Visual Data Mining
- Trabalhos recentes
- Conclusões
- Refêrendias bibliográficas



# Objetivo

- Apresentar conceitos introdutórios referentes a Visual Data Mining
- Apresentar diferentes naturezas de dados assim como, suas técnicas de visualização e interação
- Passar uma visão sobre os artigos selecionados



# Introdução



Homem

Computação

Psicologia  
Cognitiva

Tecnologia de  
Banco de Dados

Busca de  
Sentido

Tomada de  
Decisão

Visual Data Mining

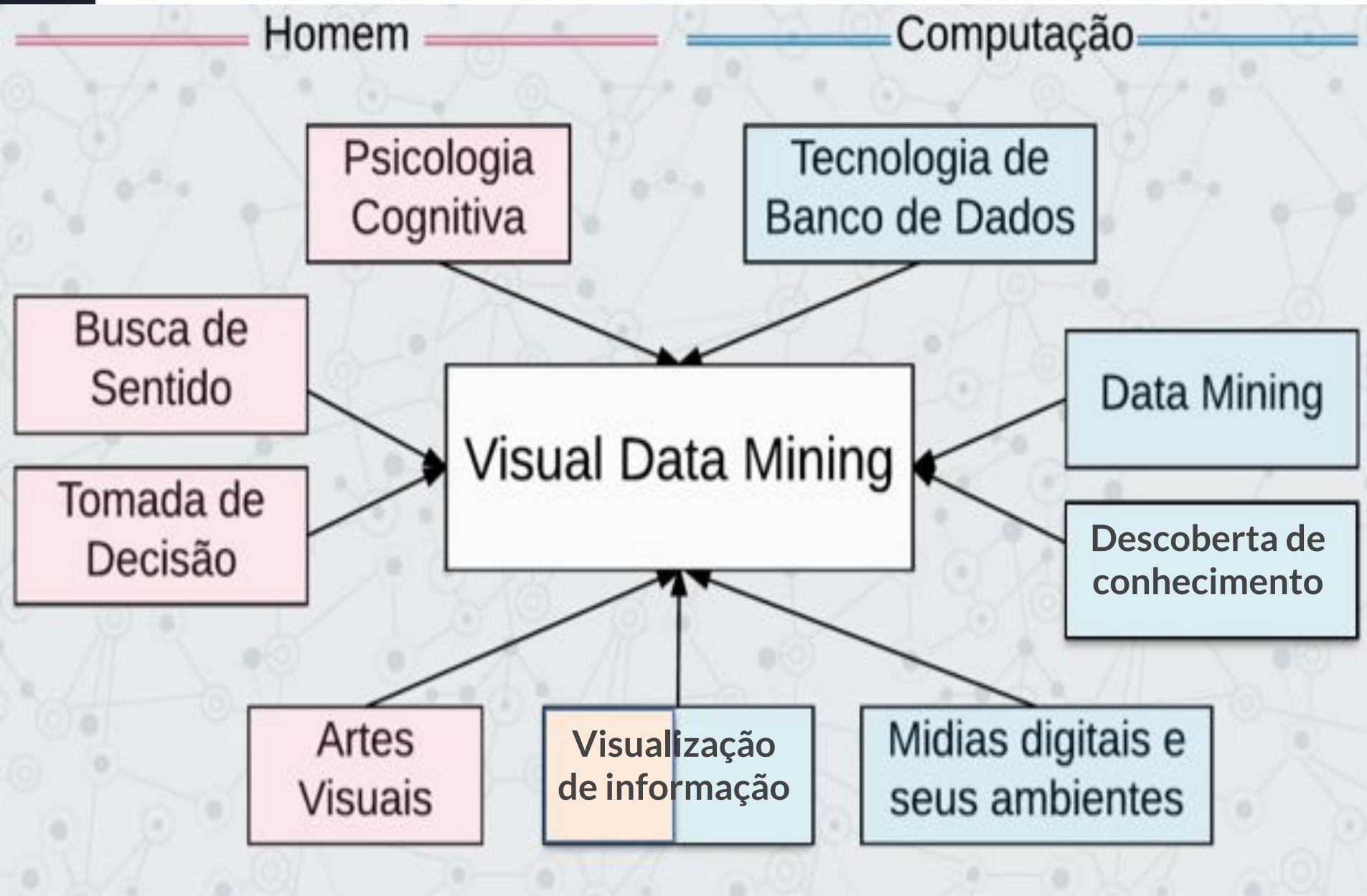
Data Mining


Descoberta de  
conhecimento

Artes  
Visuais

Visualização  
de informação

Mídias digitais e  
seus ambientes




- 
- Data Mining começa a ser muito utilizado
  - Necessidade de técnicas de visualização
    - Início da década de 90
  - Fim dos anos 90
    - Surge Visual Data Mining





# Visual Data Mining

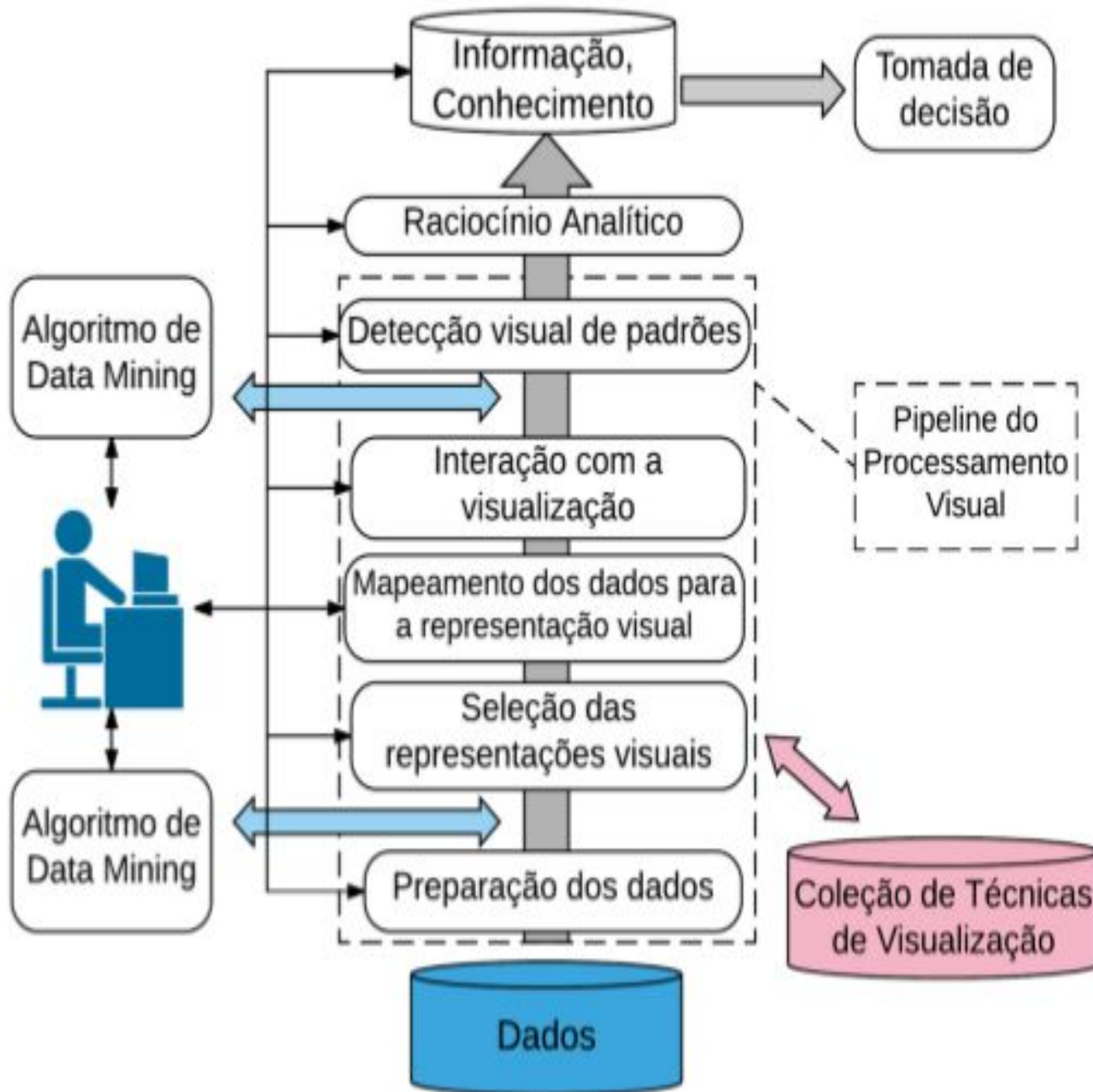


- 
- Mas o que é Visual Data Mining?
    - Interação e descobrimento de padrões
    - Visualizações abstratas



- 
- Visualização aplicada sobre resultados de algoritmos de Data Mining
  - Depende da capacidade de reconhecer padrões
  - Processa e mapeia os dados
    - Demanda muito poder computacional


- 
- Visual Data Mining depende de um processo pipeline visual interativo
  - Processo é focado no ser humano
    - Faz a visualização
    - Interage com os passos do pipeline







# Áreas de Aplicação

- Usada em áreas com grande quantidade de dados que no início são incompreensíveis
- Exemplos:
  - Ciência
  - Medicina - Análise da conectividade cerebral
  - Geografia - Mineração de dados na barragem de Itaipu (2010)

- 
- Visualização de informações usando dados reais
    - Obtidos por monitoramento na barragem
  - Os dados foram obtidos por extensômetros
    - Período de janeiro de 1995 a dezembro de 2004

- 
- Utilização de KDD no processo
  - KDD é o processo geral
  - Data mining foca na extração de padrões no meio dos dados
  - Visualização de informação já é indispensável no Data Mining

- 
- A VI permite ao usuário adquirir percepções sobre os dados, podendo surgir novas hipóteses do que está sendo analisado.
  - Visual Data Mining é um processo KDD usando VI como um canal de comunicação entre o computador e o usuário



# Benefícios

- Inclusão do usuário na análise
- Análise intuitiva
- Trata dados com ruídos e heterogêneas
- Substituição de algoritmos falhos
- Proporciona maior segurança





# Natureza dos dados

- Dado unidimensional
- Dado bidimensional
- Dado multidimensional
- Hierarquias e grafos
- Texto e Hipertexto
- Algoritmos e software



# Unidimensional

- Dados temporais
- Cores claras para altos valores
- Cores escuras para valores menores



# Bidimensional

- Sistemas de dados geográficos (SIG)
  - Latitude e Longitude
- Desenhos e Fotografias
  - Altura e Largura
- Deve se ter cuidado com a saturação



# Multidimensional

- Dados com mais de três atributos, visualização mais complexa
- Dimensões com o mesmo peso
- Exemplo:
  - Tomografia computadorizada
  - Tabelas de banco de dados



# Texto e hipertexto

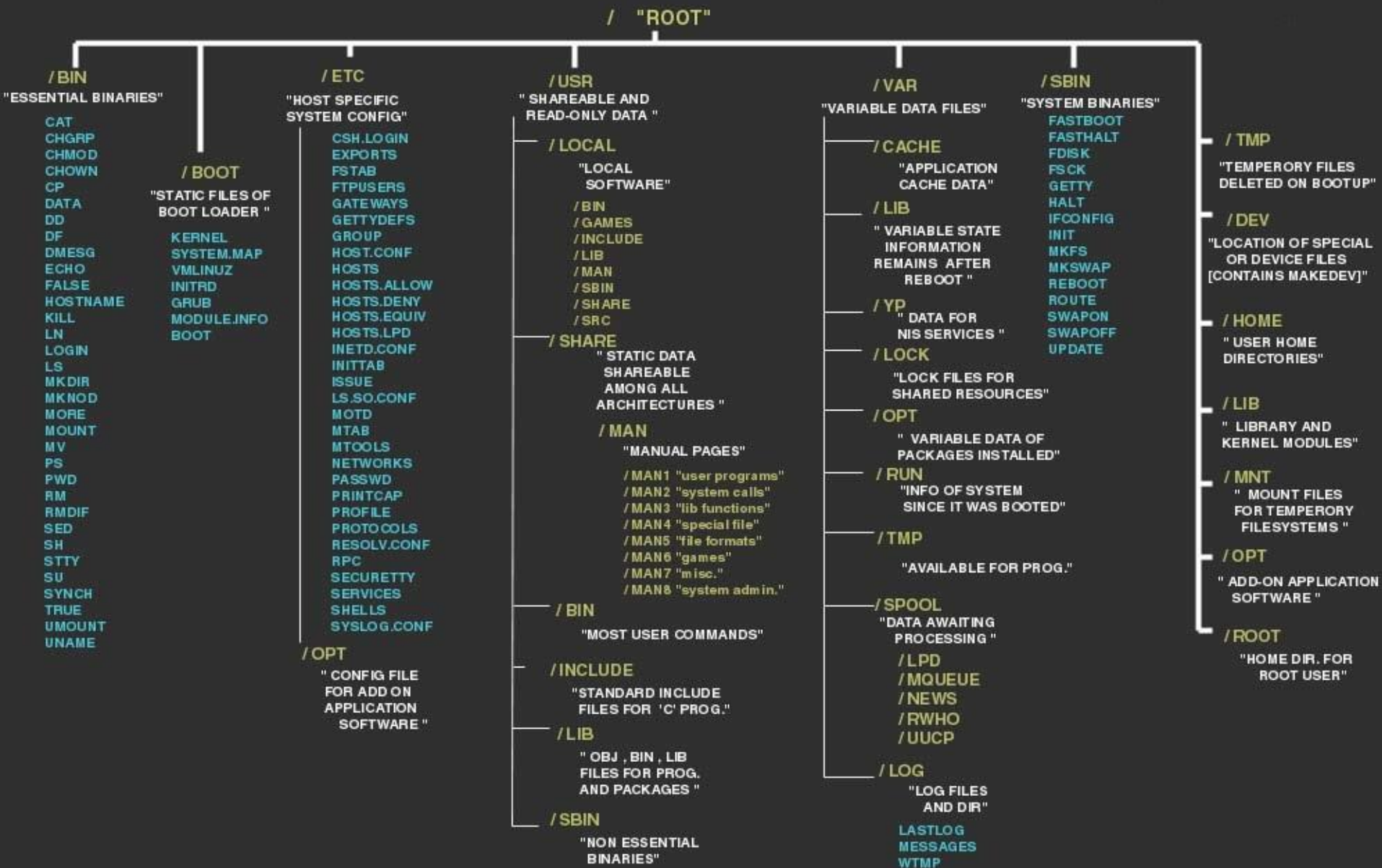
- Filtragem de dados
- Contar palavras
- Visualização de uma lista ordenada
- Diferenciação por meio de escala e cores



# Hierarquías e grafos

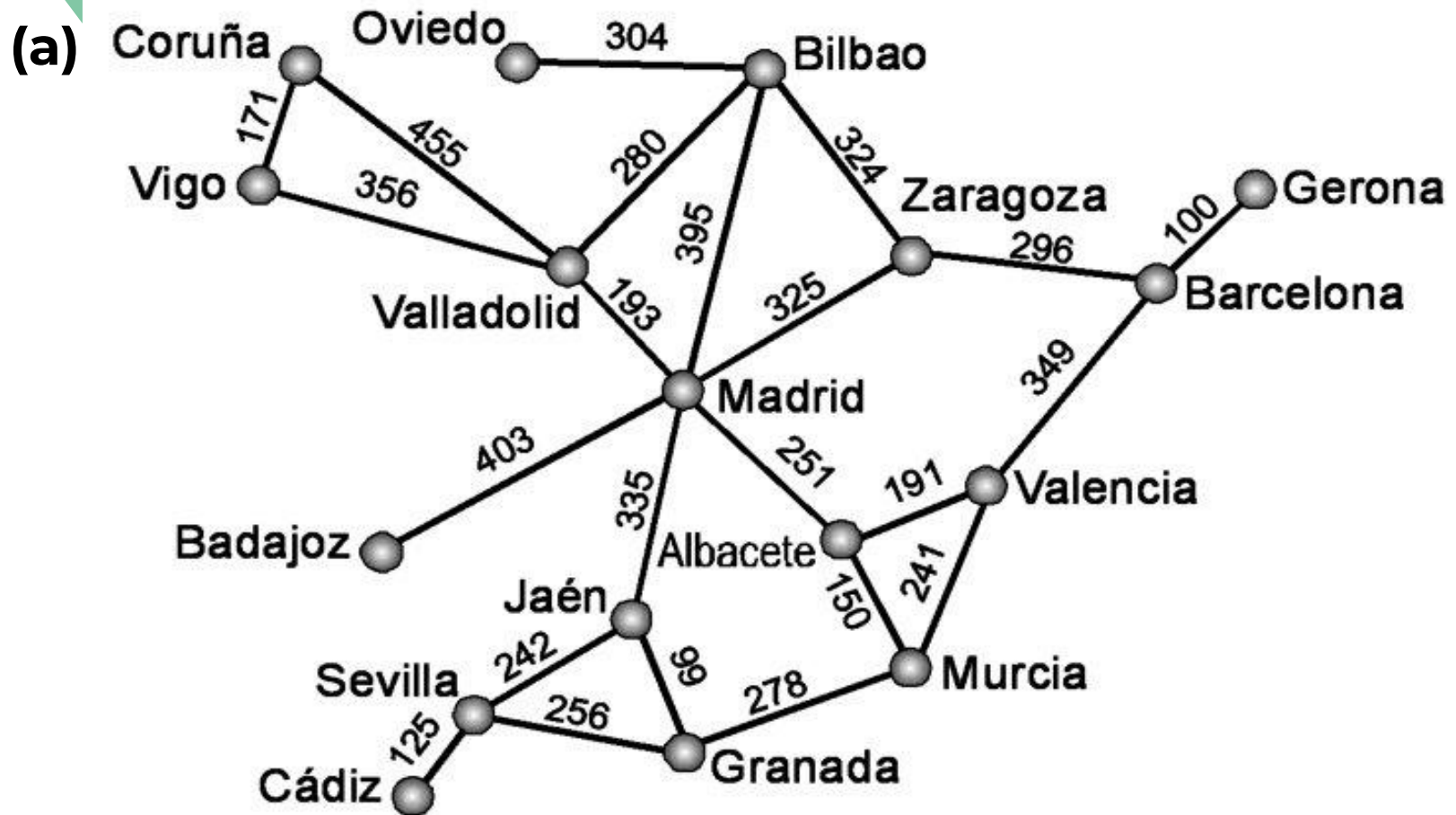
- Técnicas hierárquicas se baseiam em uma subdivisão de uma dimensão em subdimensões
  - Relacionamento entre pastas
- Grafos são utilizados para mostrar as interdependências

(a)



(a) Árvore de diretório linux

(a) Grafo representando a distância entre grandes cidades da Espanha







# Algoritmos e software

- Grandes projetos de software
- Conversão de milhares linhas de software em gráficos
- Facilitar a compreensão e depuração do código

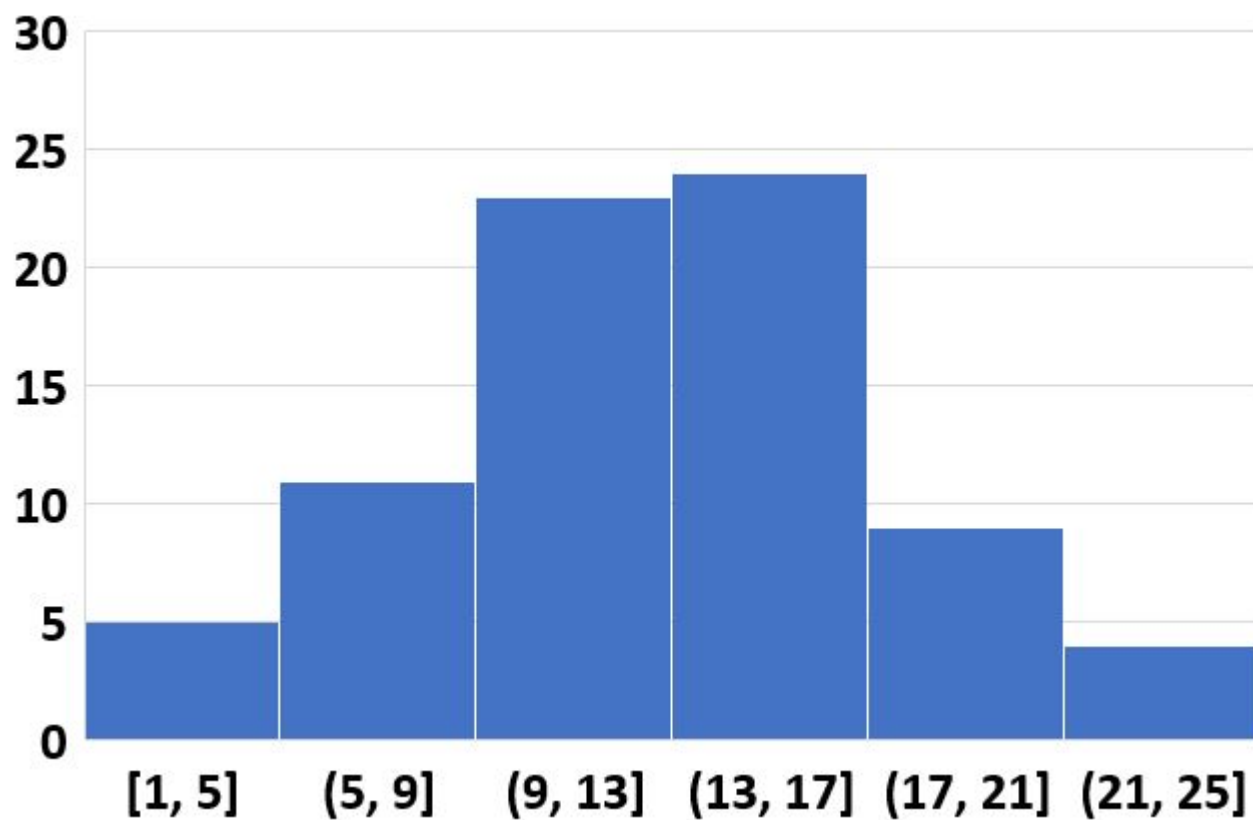


# Técnicas de visualização

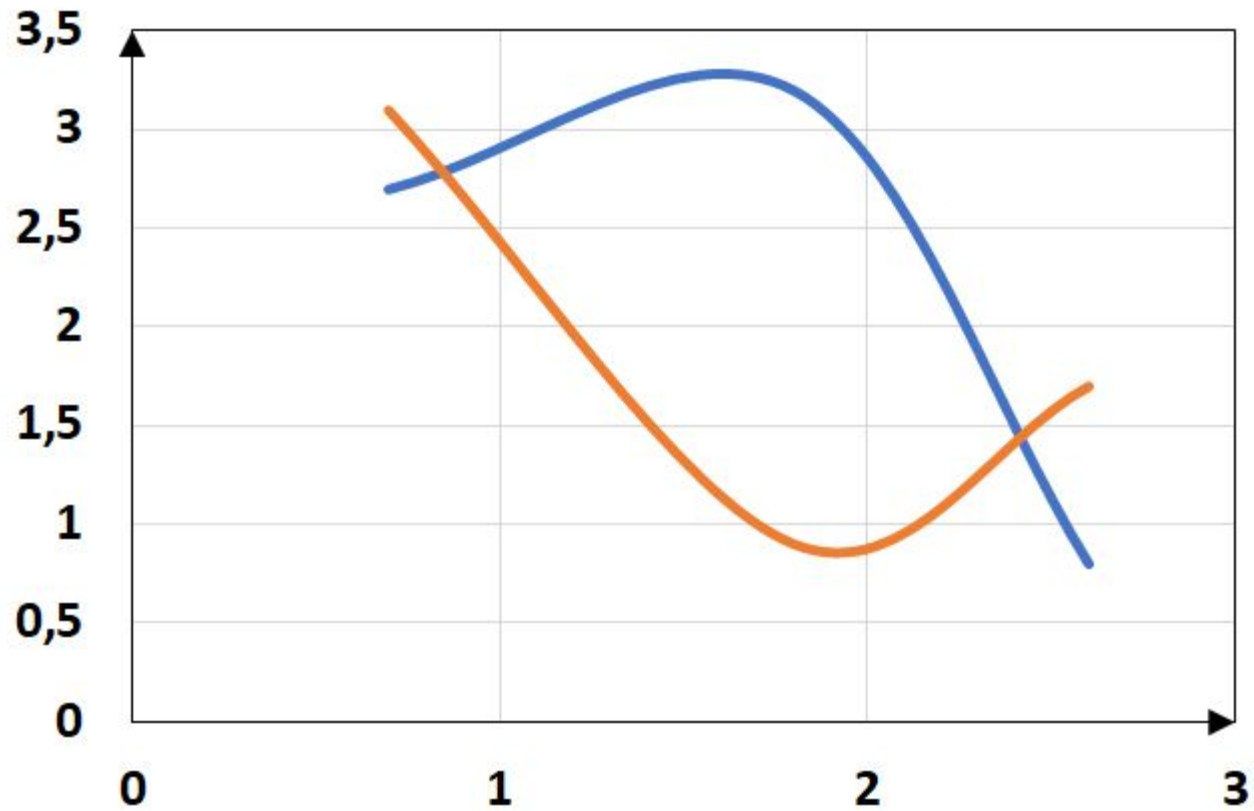
- Exibição padrões
- Exibição transformada geometricamente
- Exibição icônica
- Exibição por pixel denso
- Exibição empilhada

# Exibições Padrões

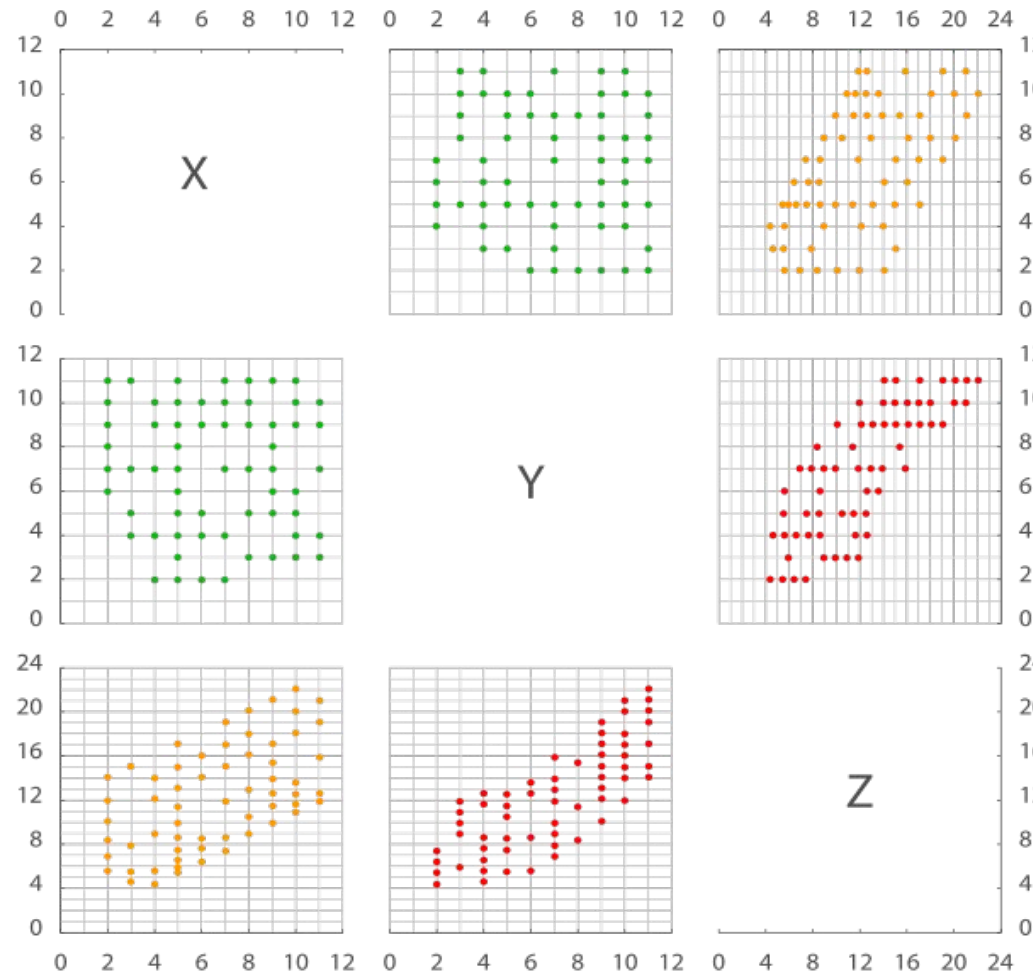
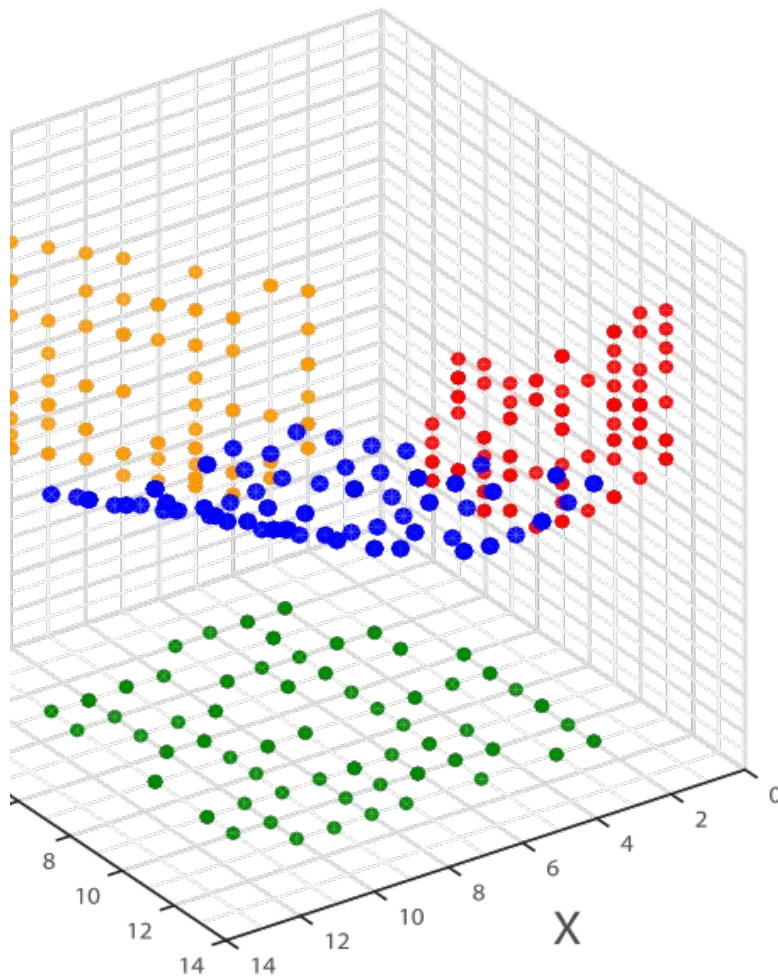
- Histograma



- Gráficos xy



- Gráfico de dispersão xyz

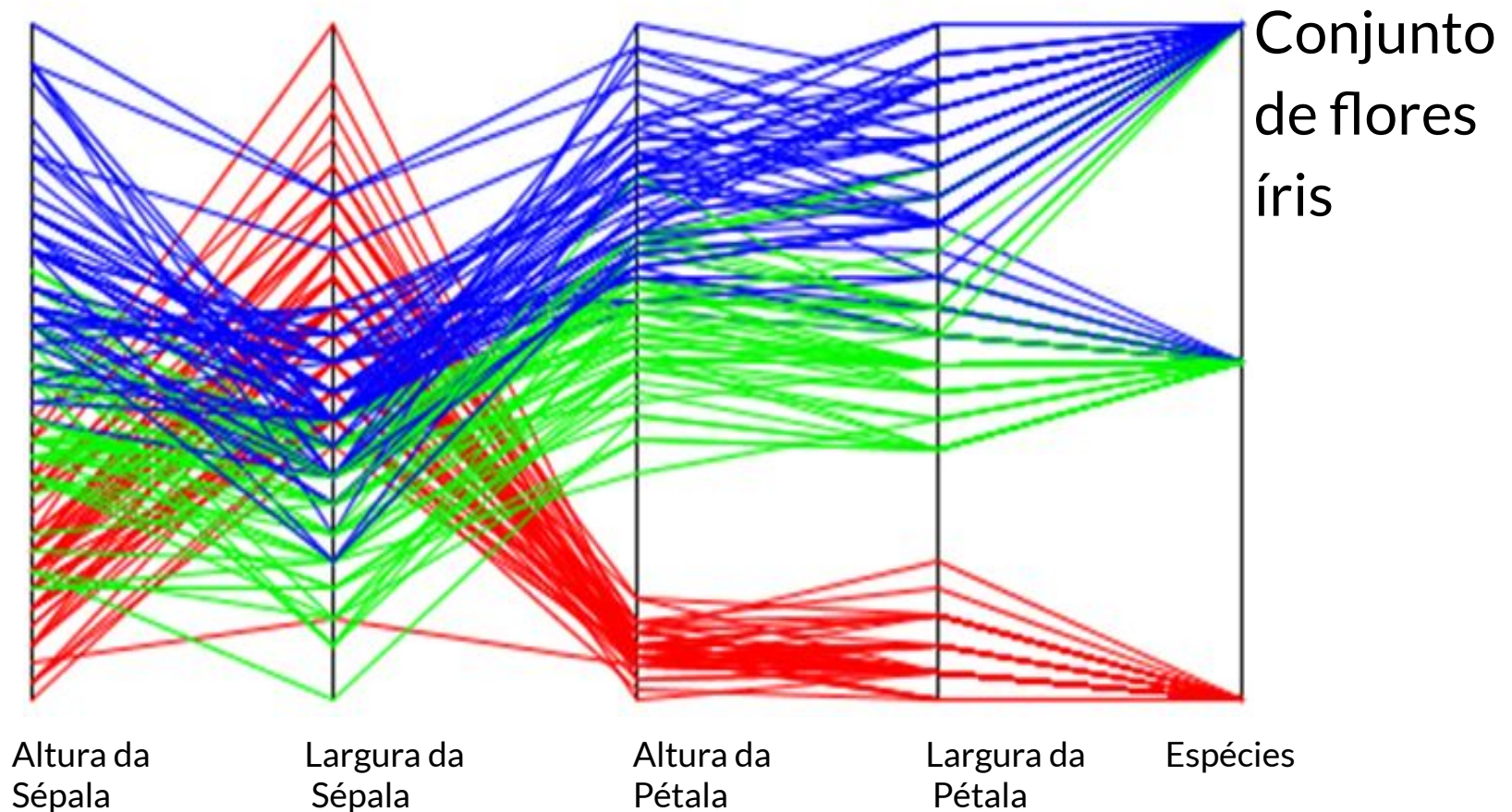




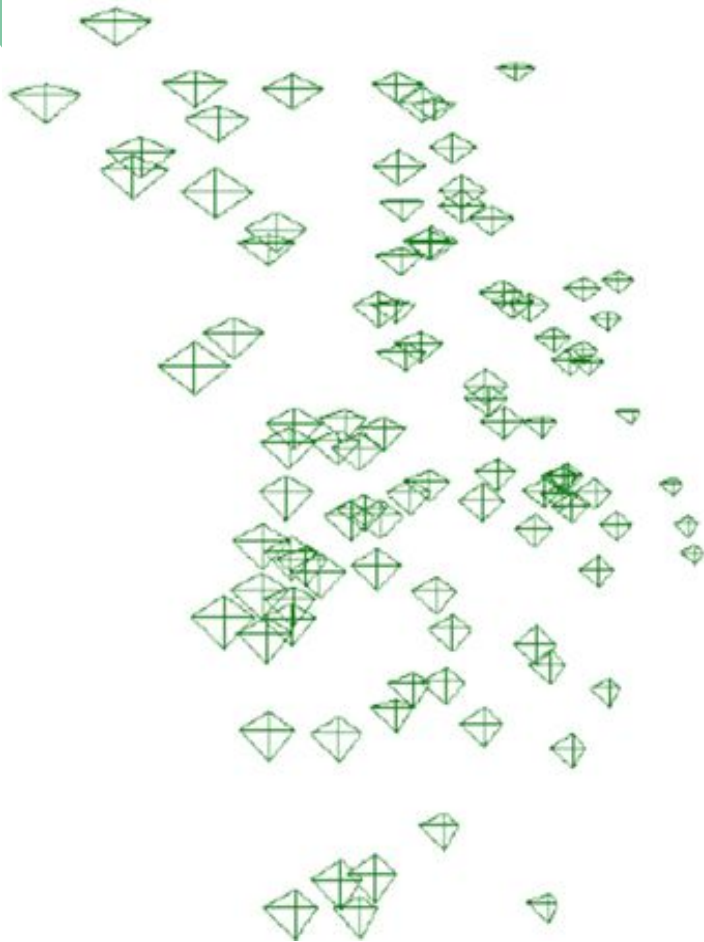
# Exibição transformada geometricamente

- Analisar características em conjunto de dados multidimensionais
- Tipos de representações:
  - Vista de prospecto
  - Coordenadas paralelas

# Exibição transformada geometricamente



# Exibição icônica



Conjunto de  
flores íris

- Variação de tamanho conforme a dimensão da amostra





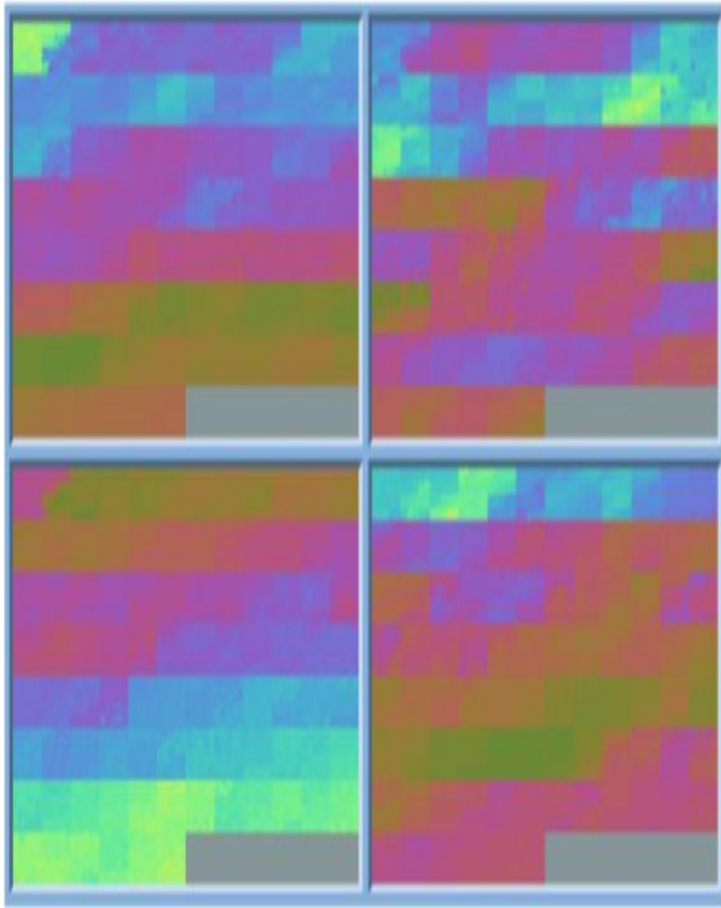


# Exibição por pixel denso

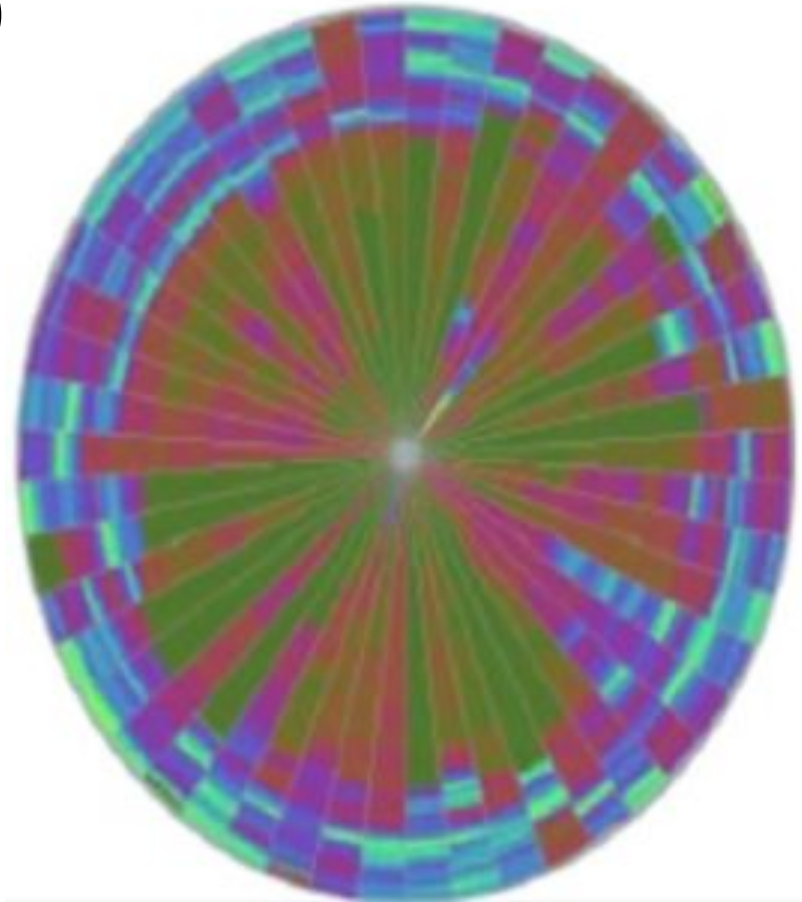
- Cada pixel representa um valor do dado
- Desafio na organização
- Tipos de representações:
  - Padrão recursiva
  - Segmentos circulares

# Exibição por pixel denso

(a)



(b)



Técnicas de exibição. (A) Padrão recursiva (B) Segmento de círculo

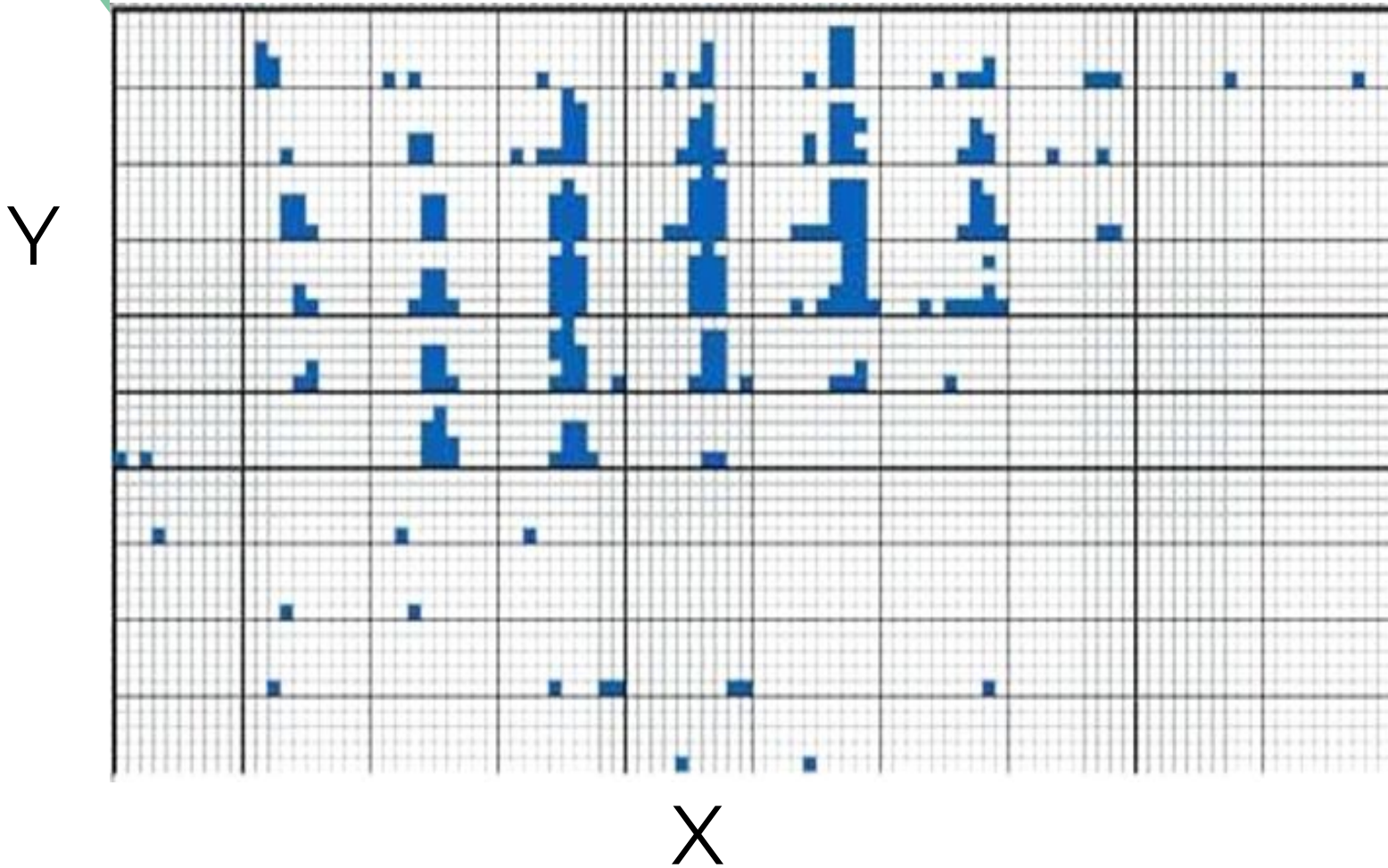


# Exibição Empilhada

- Particionamento de dados multidimensionais
- Representação por empilhamento dimensional
  - Criação de sistemas de coordenadas
  - Dimensões mais importantes primeiro
  - Ideia de hierarquia

# Exibição Empilhada

- Empresa de minério





# Técnicas de interação

- Padrão
- Projeção dinâmica
- Interação por filtragem
- Interação por zoom
- Interação por distorção
- Interação liga e escova

# Classificação



# Visual Data Mining em dados de produção acadêmica com características Big Data



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"  
Câmpus de São José do Rio Preto



SOFIA PAZZOTI

## **Visual Data Mining em dados de produção acadêmica com características Big Data**

Orientador: Prof. Dr. Carlos Roberto Valêncio

São José do Rio Preto  
2018



# Resumo

- Estudo de redes de coautoria
- Bibliometria e Cienciometria
- Limitações no universo de *Big Data*
- Criação de um visor interativo





# Bibliometria e Cienciometria

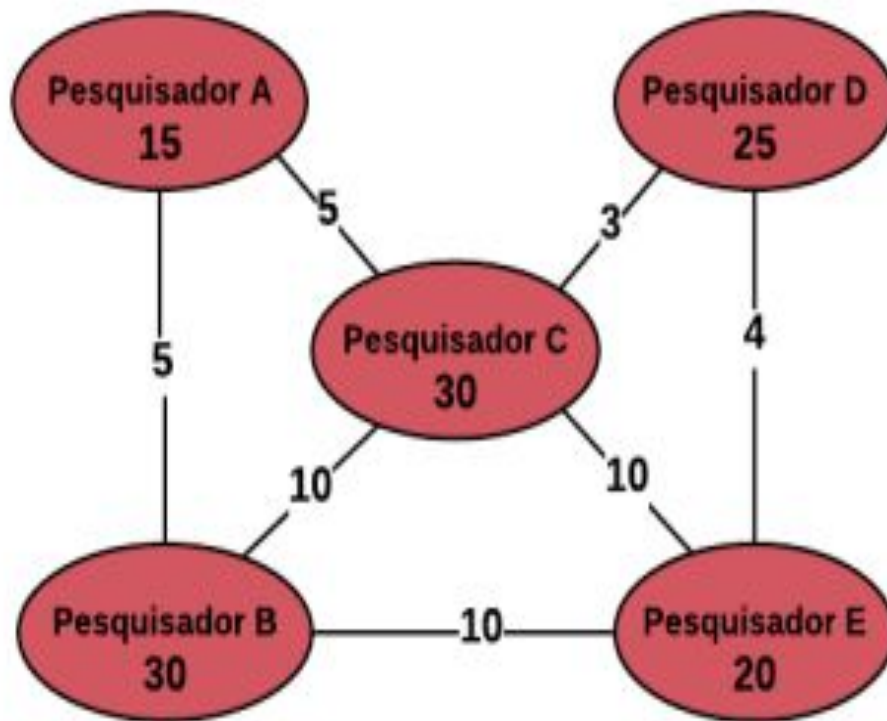
- Métodos matemáticos e estatísticos para verificar evolução da informação científica
- Estudo dos aspectos quantitativos da ciência
- Busca produzir modelos matemáticos e métricas



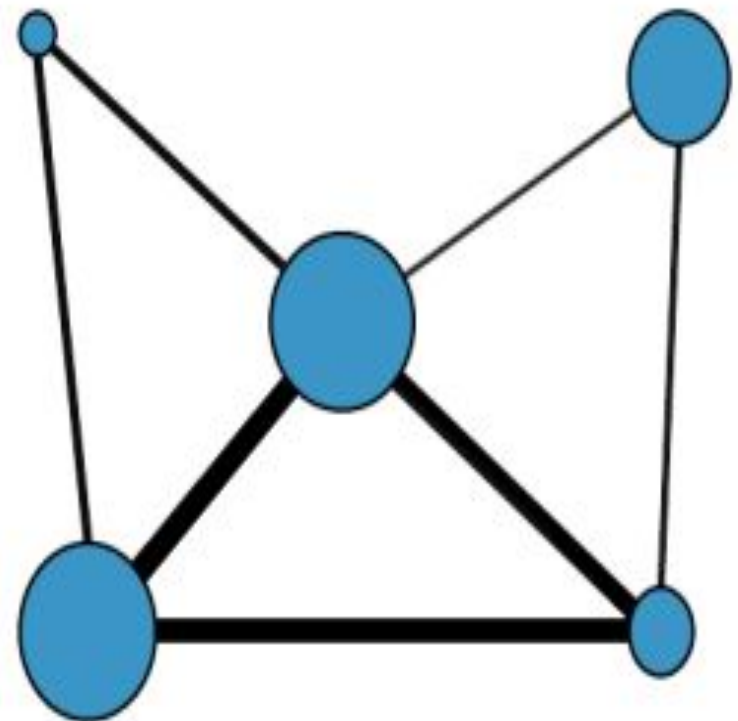
# Redes de colaboração científica

- Trabalho conjunto para atingir objetivos comuns
- Traz benefícios
- Podem ser expressas de diversas maneiras
- São estruturadas por grafos

Grafos que representam redes de coautoria. (a) Pesos representados por meio de números  
(b) Pesos representados por meio da espessura e tamanho



(a)



(b)



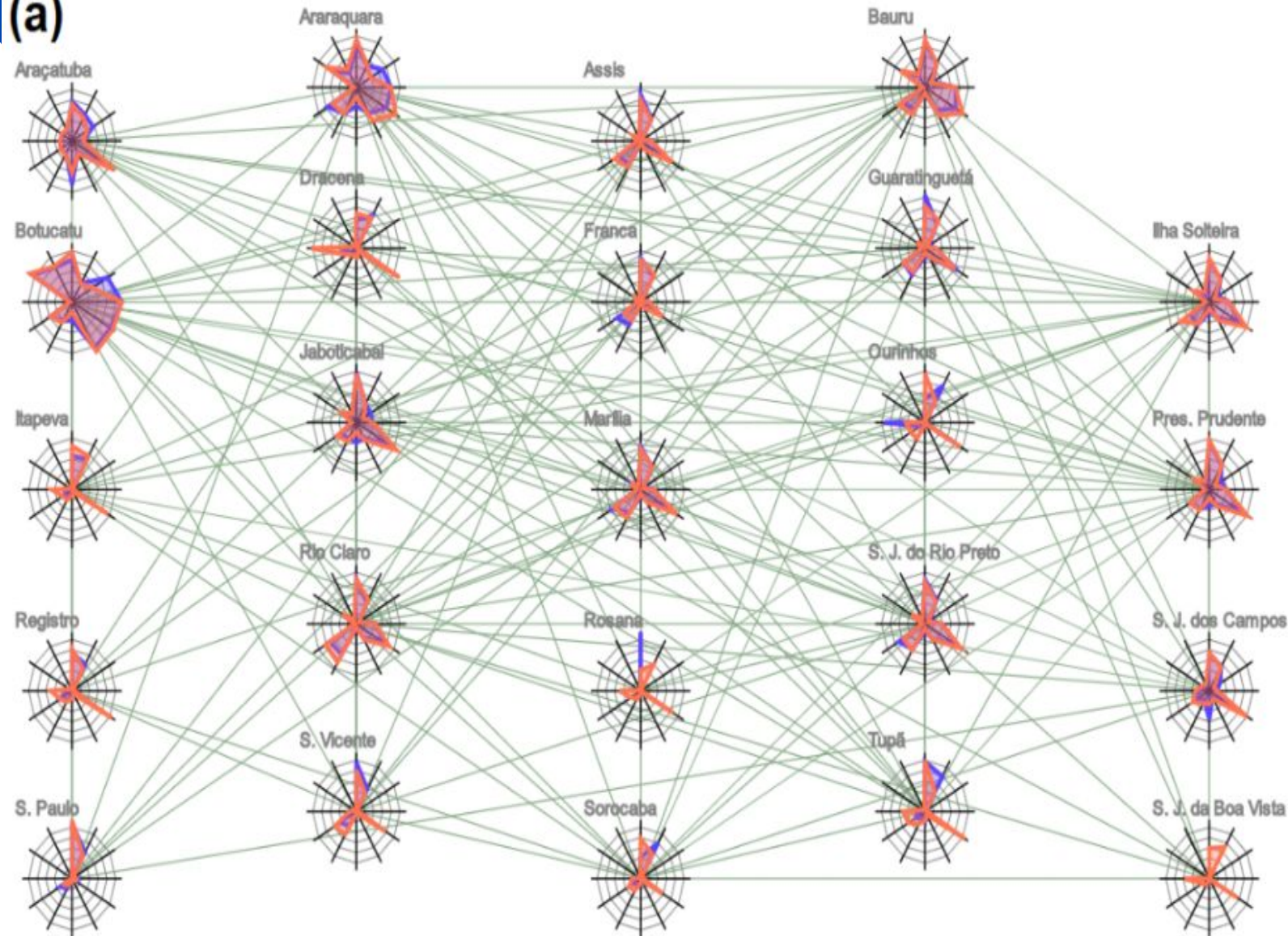
# Métricas Topológicas

- Conjunto de métricas podem ser aplicadas
  - Padrões e tendências
- Métricas como:
  - Total de coautores
  - Total de trabalhos em coautoria
  - Total de coautorias entre os pesquisadores
  - Transitividade

# Interface Principal do Visualizador com os Gráficos Radiais I selecionados. (a)

## Visualização (b) Filtros disponíveis

(a)



(b)

Selecione o Gráfico desejado ⓘ \*

Gráfico Radial 1

Selecione os períodos de anos consecutivos ⓘ \*

2006 a 2010

2015 a 2017

Intervalo de coautorias ⚙

De a coautorias

☐ Períodos ⓘ

Posicionamento ⓘ

Melhor Distr. para Visualiz

☐ Gráfico das citações ⓘ

Filtrar

Métricas ⓘ

Elementos Visualizados ⓘ

☒ Exibir gráficos radiais

☒ Exibir arestas

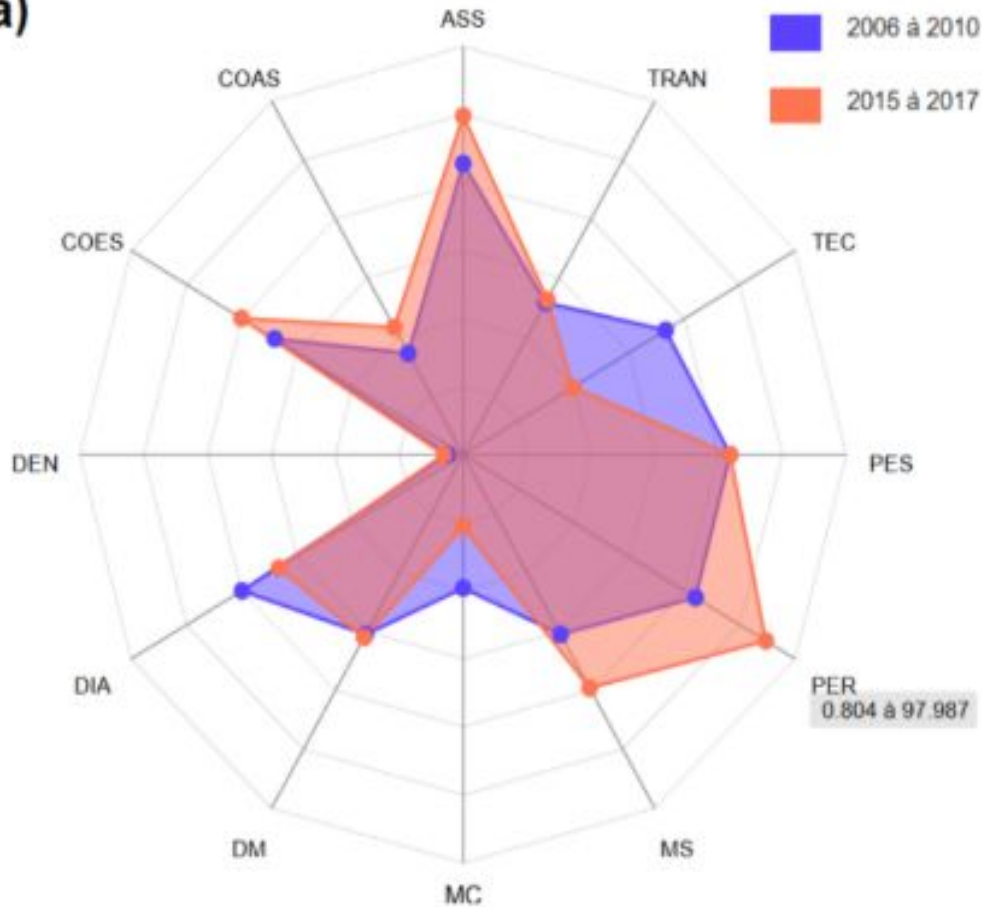
☐ Exibir peso das arestas



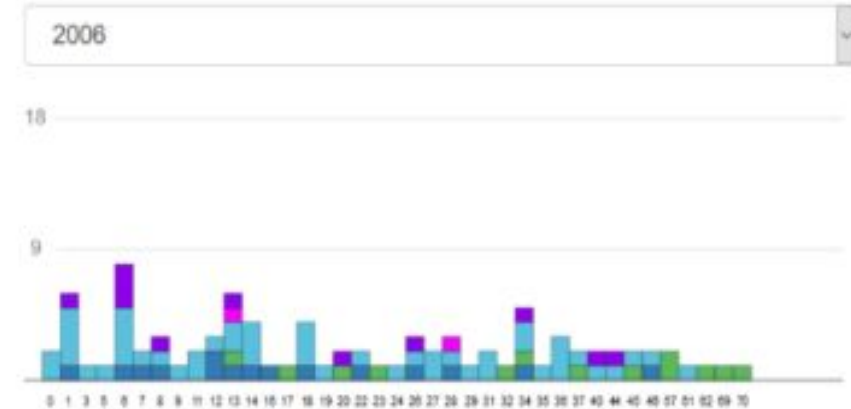
# Visualização das informações do câmpus específico. (a) Gráfico Radial I (b) Gráficos de barras

## Araraquara ⓘ

(a)



(b) Citações: ⓘ



ASS: Assortatividade  
COAS: Coautorias  
COES: Coautores  
DEN: Densidade  
DIA: Diâmetro  
DM: Distância Média  
MC: Média de Coautorias  
MS: Maior Subgrafo  
PER: Percentual do Maior Subgrafo  
PES: Pesquisadores  
TEC: Trabalhos em Coautoria  
TRAN: Translidade

Eixo Horizontal: Número de citações  
Eixo Vertical: Quantidade de publicações

Legenda de Categorias:

- Ciências Agrárias
- Ciências Exatas e da Terra
- Ciências Biológicas
- Ciências Sociais Aplicadas
- Ciências da Saúde
- Linguística Letras e Artes
- Ciências Humanas
- Engenharias
- Outros



# Avaliação da ferramenta

- Metodologia de experimentação
  - Base de dados SASD/Lattes
  - Tarefas de extração de conhecimento
  - Teste de uso



# Extração de conhecimentos

- Coautorias internas cada vez maiores
- Coautorias externas em câmpus recentes
- Câmpus com publicações mais relevantes
- Observação de um comportamento natural em citações de trabalhos





# Teste de uso

- Visualizador implementado em um servidor
- Questionário *online*
  - Critérios relacionados ao *Visual data mining*
  - E referentes ao desempenho e usabilidade



# Critérios relacionados ao *Visual data mining*

- Compreensibilidade
- Adequação
- Suficiência
- Suporte à análise




# Conclusão

- Os recursos de *Visual Data Mining*
  - Bom desempenho
  - Suportaram as análises dos dados
  - Adequados para apresentar tais dados
- Aprovação dos avaliadores
- Contribuições científicas



# Trabalhos Recentes






# **Implantação de mineração de dados visuais para campo de fluxo sanguíneo instável em um aneurisma da aorta**

Ano: 2011

Determinar relações entre cisalhamento da parede e ruptura do aneurisma na aorta.

- 
- Simulação de fluxo sanguíneo instável
  - Deslocamento da parede sanguínea é perigoso
  - VDM sugeriu características que causam o aneurisma
  - Identificou-se regiões críticas para a ruptura do aneurisma



**(a)** Whole shape layout



**(b)** Simulation domain in the descending aorta section with an aneurysm





# **Tractome: Uma Ferramenta de Visual Data Mining para análise da conectividade cerebral**

Ano: 2015

Uso de VDM para estudo da matéria branca  
do cérebro para detectar doenças  
neurológicas




- 
- Processo criado para análise de tractografias
  - Criação da ferramenta Tractome
    - Exploração rápida e fácil de dados de tractografia
  - Necessidade na área
    - Não havia método efetivo
  - Se mostrou eficiente

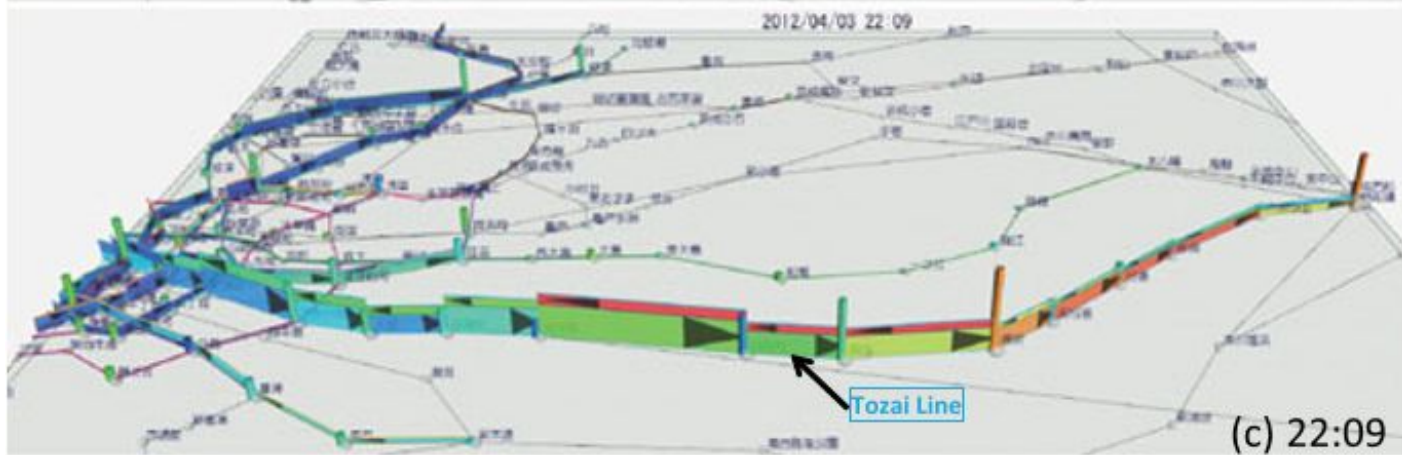
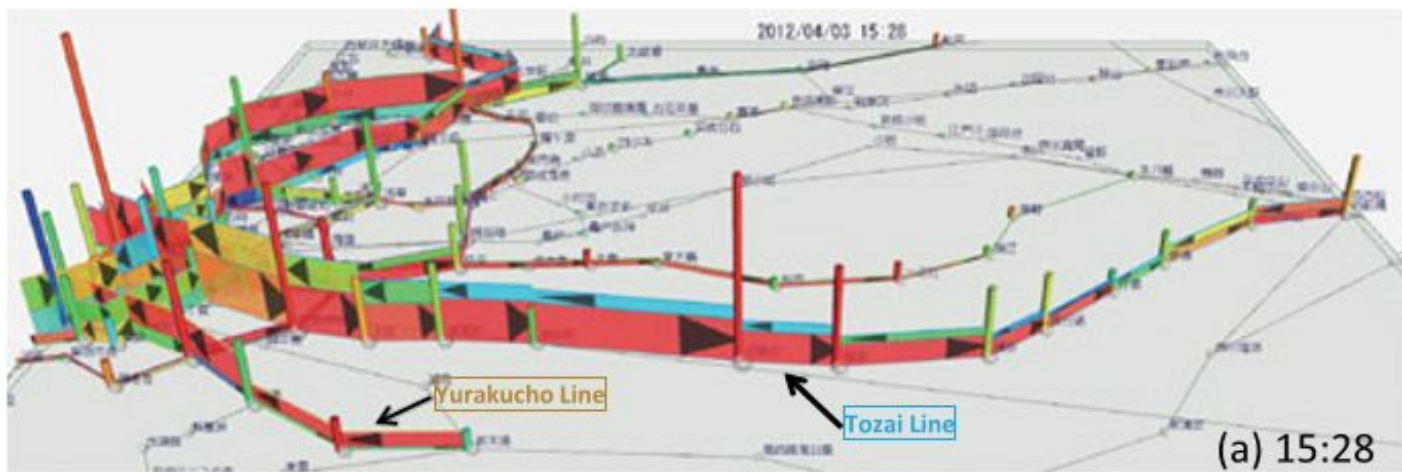


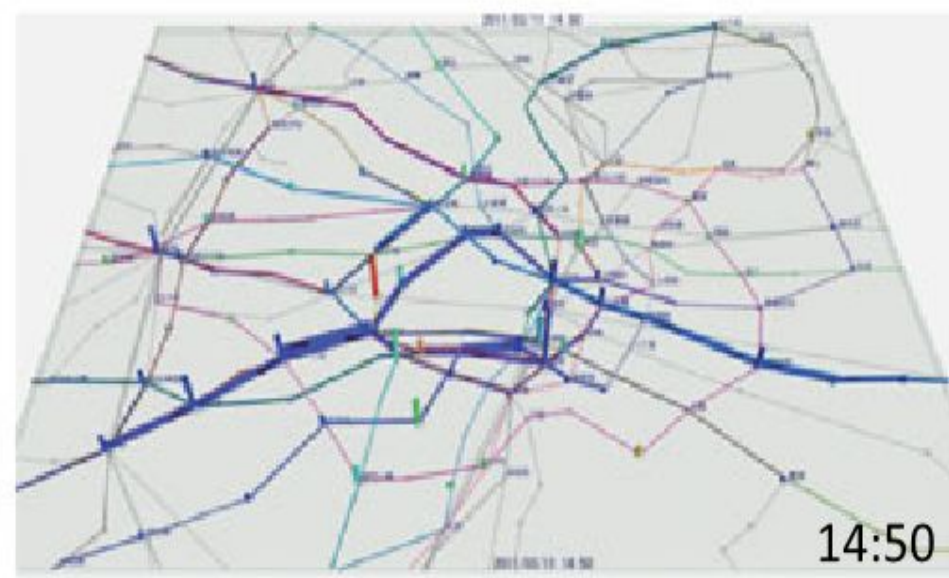
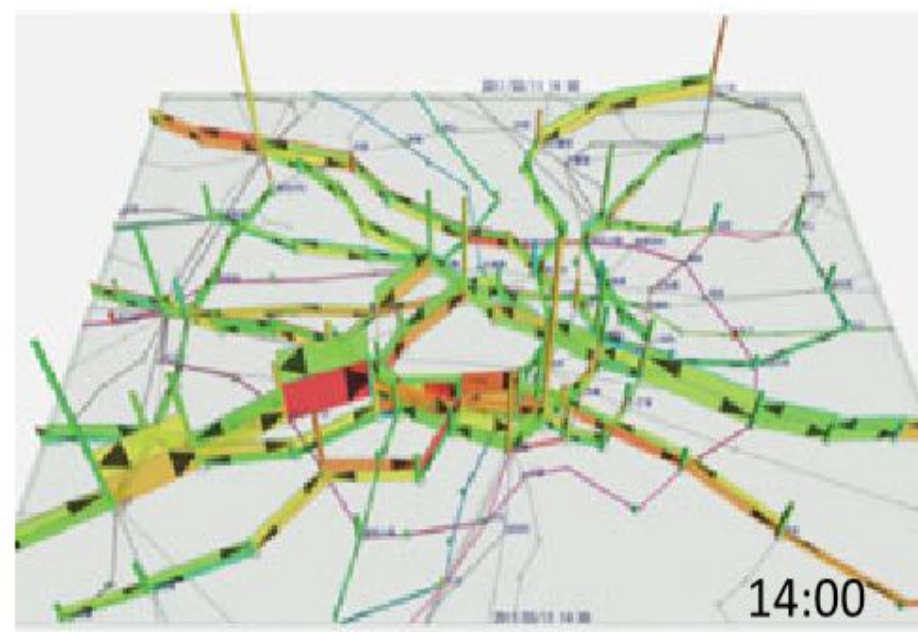
# **Exploração Visual de Mudanças nos Fluxos e Tweets de Passageiros na Rede de Metrô de Mega-cidades**

Ano: 2015

Análise de tráfego e da mídia social usando  
Big Data

- 
- Explora mudança no comportamento dos passageiros, além dos tweets feitos
  - A partir da nova ferramenta proposta, pode-se visualizar:
    - Fluxo dos passageiros no metrô
    - Fenômenos incomuns e a propagação
    - Duas formas de big-data (*smart card* e *Twitter*)
  - Todos utilizados para explorar causas e/ou efeitos de fenômenos
  - Planejam melhorar, fornecendo:
    - Previsão de eventos e fluxos de passageiros








# **Visão geral da mineração de dados e da análise visual para big data na smart grid**

Ano: 2016

Descreve a tecnologia de mineração de dados em redes inteligentes e a visualização em big data


- 
- Sistema mais eficiente, confiável e sustentável
  - Pesquisa muito recente
  - Alcanço de conquista preliminares
  - Alguns defeitos
    - Definição difusa
    - Modelo idealizado
  - Gargalo da tecnologia de visualização
    - Falta metodologia simples de desenvolvimento para o usuário
  - Possui grande potencial de desenvolvimento



# Conclusão






- 
- Mesmo tendo sido criado no fim da década de 90
    - Trabalhos recentes mostram que há muito o que se desenvolver ainda
  - Como a quantidade de dados não para de crescer
    - VDM ganha cada vez mais importância para se conseguir informações por meio de dados “perdidos”



# Referências



- 
- Keim, D.A.: Designing pixel-oriented visualization techniques: Theory and applications. IEEE Transactions on Visualization and Computer Graphics 6, 2000.
  - SOUKUP, Tom; DAVIDSON, Ian. Visual data mining: Techniques and tools for data visualization and mining. John Wiley & Sons, 2002.
  - Keim, D.A.: Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics 7, 2002.
  - RUIZ, Aguilar; TROYANO, Ferrer: Visual data mining. The Journal of Universal Computer Science, 2005.
  - CARVALHO, Elizabeth Simao; MARCOS, Adérito Fernandes. Visualização da informação.
  - PAZZOTI, SOFIA. Visual Data Mining em dados de produção acadêmica com características Big Data
  - NETO, Marco Aurélio Silva et al. Técnicas de mineração visual de dados aplicadas aos dados de instrumentação da barragem de Itaipu. **Visual data mining techniques applied for the analysis of data collected at Itaipu power plant gest. prod**, v. 17, n. 4, p. 721-734, 2010.
  - MORIZAWA, Seiichiro et al. Implementation of visual data mining for unsteady blood flow field in an aortic aneurysm. **Journal of visualization**, v. 14, n. 4, p. 393-398, 2011.
  - PORRO-MUÑOZ, Diana et al. Tractome: a visual data mining tool for brain connectivity analysis. **Data mining and knowledge discovery**, v. 29, n. 5, p. 1258-1279, 2015.
  - ITOH, Masahiko et al. Visual Exploration of Changes in Passenger Flows and Tweets on Mega-City Metro Network. **IEEE Trans. Big Data**, v. 2, n. 1, p. 85-99, 2016.
  - HOU, Lin et al. Overview of data mining and visual analytics towards big data in smart grid. In: **Identification, Information and Knowledge in the Internet of Things (IIKI)**, 2016 International Conference on. IEEE, 2016. p. 453-456.



**Obrigado!**