

Mineração de Dados em Biologia Molecular



Visualização

Docente: André C. P. L. F. de Carvalho
PAE: Victor Hugo Barella



Tópicos

- Introdução
- Objetivos
- Motivações
- Representação visual
 - Objetos
 - Atributos
 - Relacionamentos

© André de Carvalho - ICMC/USP

2



Introdução

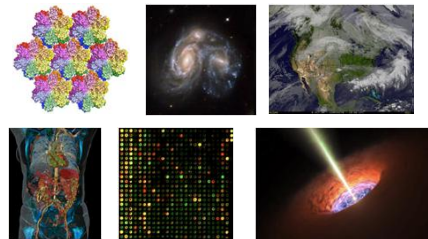
- Figuras e imagens são frequentemente utilizadas para ilustrar:
 - Condições do tempo
 - Perfis populacionais
 - Resultados de eleições
 - Estruturas de proteínas
 - Níveis de expressão de genes
- Facilitam a compreensão da informação transmitida

© André de Carvalho - ICMC/USP

3



Exemplos



© André de Carvalho - ICMC/USP

4



Introdução

- Visualização tem um papel importante em análise de dados
 - Uma das técnicas mais poderosas para exploração dos dados
 - Facilita visualização de dados e resultados
 - *Visual data mining*
 - Uso de técnicas de visualização em mineração de dados
 - Importante área de Mineração de Dados

© André de Carvalho - ICMC/USP

5



Um dos primeiros usos



Napoleão

Exército francês, comandado por Napoleão
Invade a Rússia, em 1812



© André de Carvalho - ICMC/USP

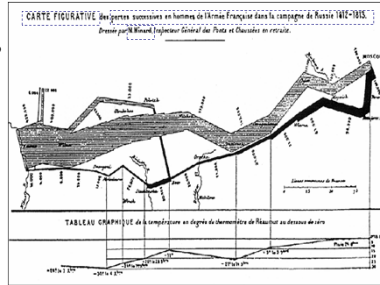
6



Um dos primeiros usos

Perdas humanas do exército francês, sob o comando de Napoleão, na invasão da Rússia em 1812

Charles Joseph Minard (1869)



© André de Carvalho - ICMC/USP

7

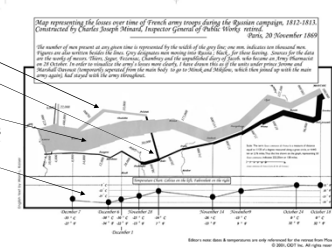


O que mostra

Cinza: entra na Rússia
Preto: sai da Rússia

Cada mm de largura
Equivale a 10.000 homens

Variação de temperatura



© www.odt.org - from <http://www.odt.org/Papers/minard.jpg>, used by permission

© André de Carvalho - ICMC/USP

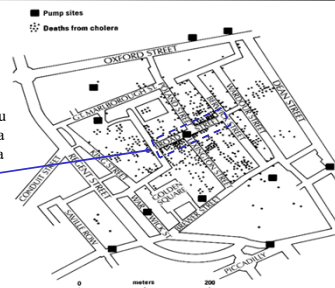
8



Outro exemplo

Mapa da Cólera em Londres
(Snow) 1855

Distribuição da doença permitiu identificar que a fonte da cólera era uma bomba de água pública na Broad Street



© André de Carvalho - ICMC/USP

9



Exemplo mais recentes

Ásia durante a noite



© André de Carvalho - ICMC/USP

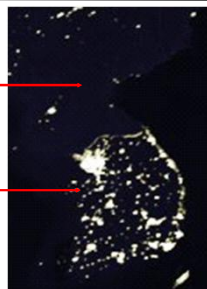
10



Exemplo mais recentes

Coreia do Norte

Seul,
Coreia do Sul



© André de Carvalho - ICMC/USP

11



Motivação

- Pessoas se saem bem na detecção de estruturas em imagens
 - Têm mais facilidade para entender informação representada visualmente
 - E encontrar padrões
 - Conseguem usar conhecimento do domínio que elas têm, mesmo sem estar cientes
 - Difícil usar em ferramentas estatísticas ou computacionais

© André de Carvalho - ICMC/USP

12



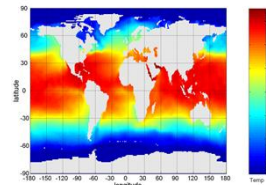
Benefícios

- Resume informação presente em dados e resultados experimentais
 - Torna mais claros
 - Padrões e tendências gerais
 - Anomalias e outliers
- Facilita análise de grandes conjuntos de dados



Exemplo

- Temperatura da superfície marítima em graus Celsius para julho de 1992
 - Figura resume ≈250.000 valores numéricos



Que padrões é possível observar?



Visualização de dados

- Utiliza representações gráficas para ilustrar os principais aspectos de um conjunto de dados
- Na mineração de dados
 - Visualização facilita compreensão dos dados por:
 - Usuários
 - Cientistas de dados
 - Para uma boa visualização, é essencial:
 - A escolha adequada da representação visual de objetos e atributos



Visualização de dados

- Mostra a informação em um formato gráfico ou tabular
 - Técnicas convertem informação em um conjunto de dados para um formato visual
 - Características dos objetos e atributos e relacionamentos podem ser:
 - Analisadas
 - Exploradas



Visualização de dados

- Exibição de um conjunto de dados em um formato visual
 - Recebe:
 - Valores de objetos, atributos e relacionamentos
 - Transforma em:
 - Objetos, atributos e relacionamentos visuais
 - Para isso, utiliza elementos gráficos
 - Pontos, linhas, formatos e cores



Visualização de dados

- Principais objetivos:
 - Permitir a interpretação por seres humanos da informação apresentada visualmente
 - Formação de um modelo mental da informação
- Aspectos importantes
 - Representação dos dados
 - Arranjos



Visualização

- Importantes para uma boa visualização
 - Escolha da representação visual para objetos, atributos e relacionamentos
 - Arranjo de objetos e atributos para salientar relacionamentos
 - Seleção de objetos e atributos



Representação de dados

- Utiliza forma gráfica para representar
 - Objetos
 - Atributos
 - Relacionamentos entre objetos
 - Explícitos
 - Implícitos



Representação de objetos

- Considerando um único atributo categórico
 - Atributo é exibido em entrada de tabela ou área da tela e cada valor assume uma cor ou tom diferente
- Considerando mais de um atributo
 - Por uma linha ou coluna de uma tabela ou uma linha em um grafo
- Mais de um objeto
 - Cada objeto pode ser por interpretado como um ponto (figura) em um espaço de 2 ou 3 dimensões



Representação de atributos

- Depende do tipo de atributo
- Ordinais e numéricos
 - Podem ser mapeados por características gráficas contínuas
 - Tamanho, cor ou localização em um eixo
- Nominais
 - Cada valor pode ser mapeado para uma:
 - Posição, cor, formato orientação ou coluna de uma tabela
 - Cuidado para não inserir relação de ordem



Representação de relacionamentos

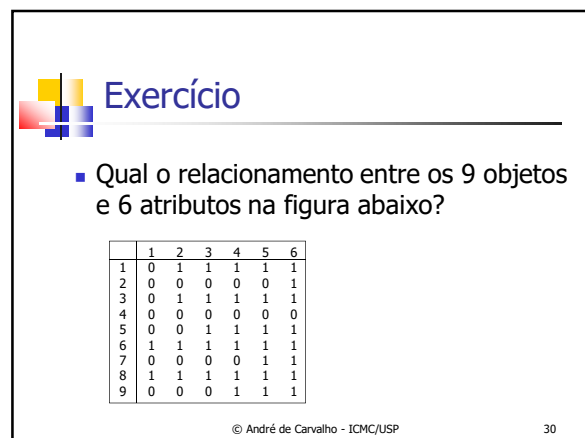
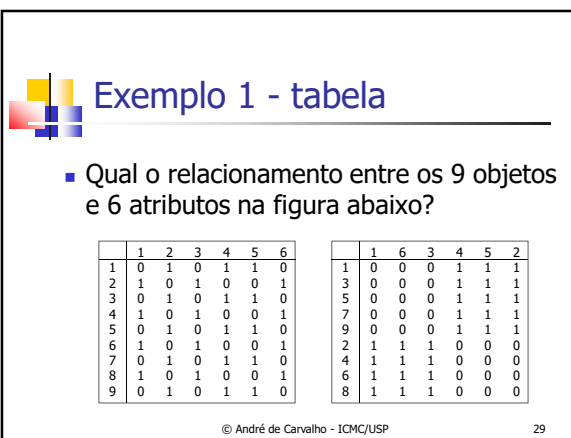
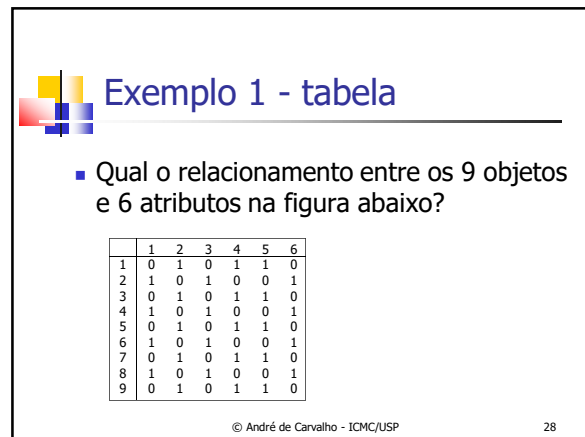
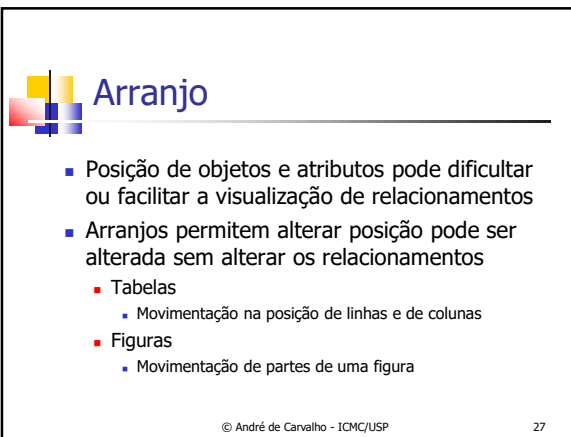
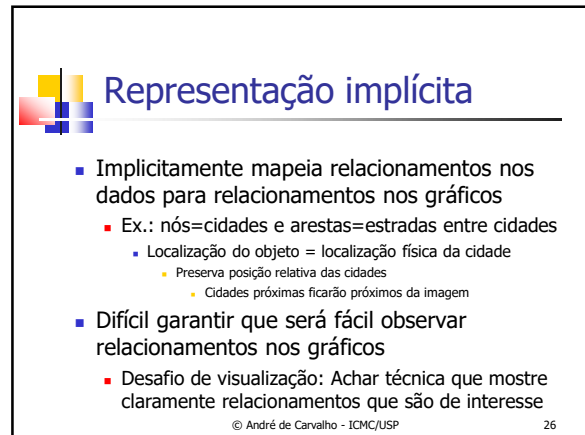
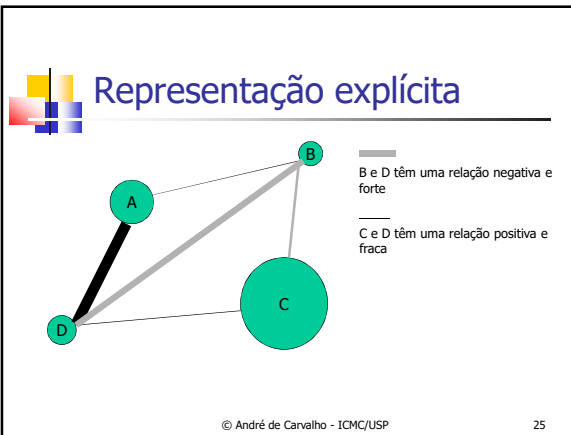
- Mostra como objetos de um conjunto de dados se relacionam
- Representação pode ocorrer de duas formas
 - Explícita
 - Implícita



Representação explícita

- Para dados representados por grafos
 - Usar representação para grafos
 - Com nós (objetos) e arestas (relacionamentos) entre nós
- Se objetos e relacionamentos têm atributos ou características próprias
 - Isso é representado graficamente
 - Ex.: nós = cidades e arestas = estradas entre cidades
 - Tamanho dos nós = população e largura da aresta = fluxo na estrada







Exercício

- Qual o relacionamento entre os 9 objetos e 6 atributos na figura abaixo?

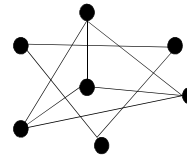
	1	2	3	4	5	6
1	0	1	1	1	1	1
2	0	0	0	0	0	1
3	0	1	1	1	1	1
4	0	0	0	0	0	0
5	0	0	1	1	1	1
6	1	1	1	1	1	1
7	0	0	1	0	0	1
8	1	1	1	1	1	1
9	0	0	1	1	0	1

	1	2	5	4	3	6
4	0	0	0	0	0	0
2	0	0	0	0	0	1
7	0	0	0	0	1	1
9	0	0	0	1	1	1
5	0	0	1	1	1	1
1	0	1	1	1	1	1
3	0	1	1	1	1	1
8	1	1	1	1	1	1
6	1	1	1	1	1	1



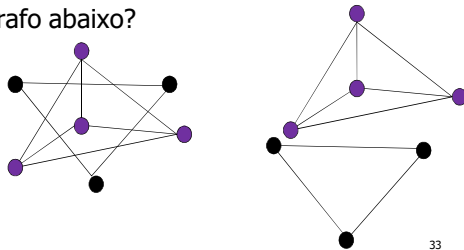
Exemplo 2 - grafo

- Qual o relacionamento entre os nós do grafo abaixo?



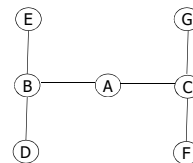
Exemplo 2 - grafo

- Qual o relacionamento entre os nós do grafo abaixo?



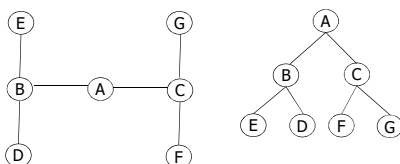
Exercício

- Qual o relacionamento entre os nós do grafo abaixo?



Exercício

- Qual o relacionamento entre os nós do grafo abaixo?



Seleção

- Objetos são selecionados quando é difícil representar todos graficamente
 - Presença de muitos objetos e/ou atributos (dimensões) dificulta a visualização de objetos
 - Alternativas para seleção:
 - Eliminar ou destacar alguns objetos
 - Utilizar uma amostra dos dados
 - Focar em uma região de dados que seja de maior interesse
 - Eliminar ou selecionar atributos



Seleção de atributos

- Fácil mapear poucos atributos para 2 ou 3, mas é difícil mapear muitos atributos
 - Para muitos atributos, selecionar um subconjunto deles (geralmente 2)
 - Se número de atributos não é grande, pode ser feito para todos os pares de atributos
 - Alternativamente, pode ser mostrada uma sequência de figuras com 2 dimensões



37



Seleção de atributos

- Alternativas
 - Agregação
 - Análise de componentes principais (PCA)
 - Ranking
 - Filtro
 - Subconjunto
 - Wrapper

38



Conjunto de dados iris

- Iris (lírio): planta com flor
 - Atributos de entrada numéricos
 - Tamanho sépala (cm)
 - Largura sépala (cm)
 - Tamanho pétala (cm)
 - Largura pétala (cm)
 - Classes
 - Iris Setosa
 - Iris Versicolour
 - Iris Virginica
- 150 exemplos, com distribuição 50/50/50



© André de Carvalho - ICMC/USP

39



Conclusão

- Objetivos
- Motivações
- Representação visual
 - Objetos
 - Atributos
 - Relacionamentos

40



Perguntas



© André de Carvalho - ICMC/USP

41