# Agrupamento Hierárquico

**Prof. Eduardo Raul Hruschka**

# Créditos

- O material a seguir consiste de adaptações e extensões dos originais:
  - de Eduardo R. Hruschka e Ricardo J. G. B. Campello
  - de (Tan et al., 2006)
  - de E. Keogh (SBBD 2003)

# Agenda

- Algoritmos Hierárquicos
  - Conceitos e Definições
    - Dendrogramas
    - Grafos de Proximidade
  - Métodos Aglomerativos
    - *Single Linkage*
    - *Complete Linkage*

# Relembrando...

- **Matriz de Dados X:**
  - *N* linhas (objetos) e *n* colunas (atributos):

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nn} \end{bmatrix}$$

- Cada **objeto** (linha da matriz) é denotado por um vetor $\mathbf{x}_i$
  - Exemplo:

$$\mathbf{x}_1 = \begin{bmatrix} x_{11} & \cdots & x_{1n} \end{bmatrix}$$

# Relembrando...

- **Matriz de Proximidade (**Dissimilaridade ou Similaridade**):**
  - $N$ linhas e $N$ colunas:

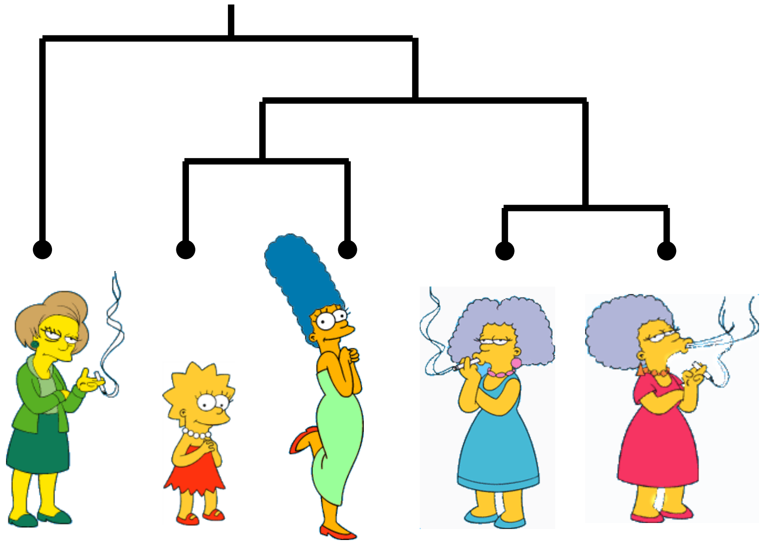
$$\mathbf{D} = \begin{bmatrix} d(\mathbf{x}_1, \mathbf{x}_1) & d(\mathbf{x}_1, \mathbf{x}_2) & \cdots & d(\mathbf{x}_1, \mathbf{x}_N) \\ d(\mathbf{x}_2, \mathbf{x}_1) & d(\mathbf{x}_2, \mathbf{x}_2) & \cdots & d(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & & \ddots & \vdots \\ d(\mathbf{x}_N, \mathbf{x}_1) & d(\mathbf{x}_N, \mathbf{x}_2) & \cdots & d(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}$$

- Simétrica se proximidade $d$ apresentar propriedade de simetria

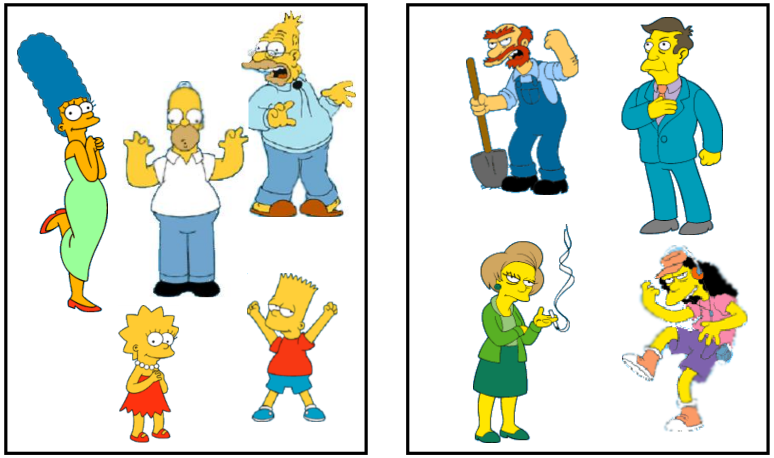
# Relembrando...

- **Agrupamento Particional:** constrói uma *partição* dos dados
- **Agrupamento Hierárquico:** constrói uma *hierarquia de partições*

Hierárquicos



Particionais



Keogh, E. A Gentle Introduction to Machine Learning and Data Mining for the Database Community, SBBD 2003, Manaus.

# Definição de Partição de Dados

- Consideremos um conjunto de $N$ objetos a serem agrupados: $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$
- **Partição** (rígida): coleção de $k$ grupos não sobrepostos $\mathbf{P} = \{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_k\}$ tal que:

$$\boxed{\begin{array}{c} \mathbf{C}_1 \cup \mathbf{C}_2 \cup \ldots \cup \mathbf{C}_k = \mathbf{X} \\[1em] \mathbf{C}_i \neq \varnothing \\[1em] \mathbf{C}_i \cap \mathbf{C}_j = \varnothing \ \text{ para } \ i \neq j \end{array}}$$

  - Exemplo: $\mathbf{P} = \{ (\mathbf{x}_1), (\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6), (\mathbf{x}_2, \mathbf{x}_5) \}$

# Definição de Hierarquia

- **Hierarquia** (de partições de dados):
  - **Sequência de partições aninhadas**
    - Uma partição $\mathbf{P}_1$ está *aninhada* em $\mathbf{P}_2$ se cada componente (grupo) de $\mathbf{P}_1$ é um subconjunto de um componente de $\mathbf{P}_2$
- **Exemplo:**

$\mathbf{P}_1 = \{ (\mathbf{x}_1), (\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6), (\mathbf{x}_2, \mathbf{x}_5) \}$

$\mathbf{P}_2 = \{ (\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6), (\mathbf{x}_2, \mathbf{x}_5) \}$

- **Contra-Exemplo:**

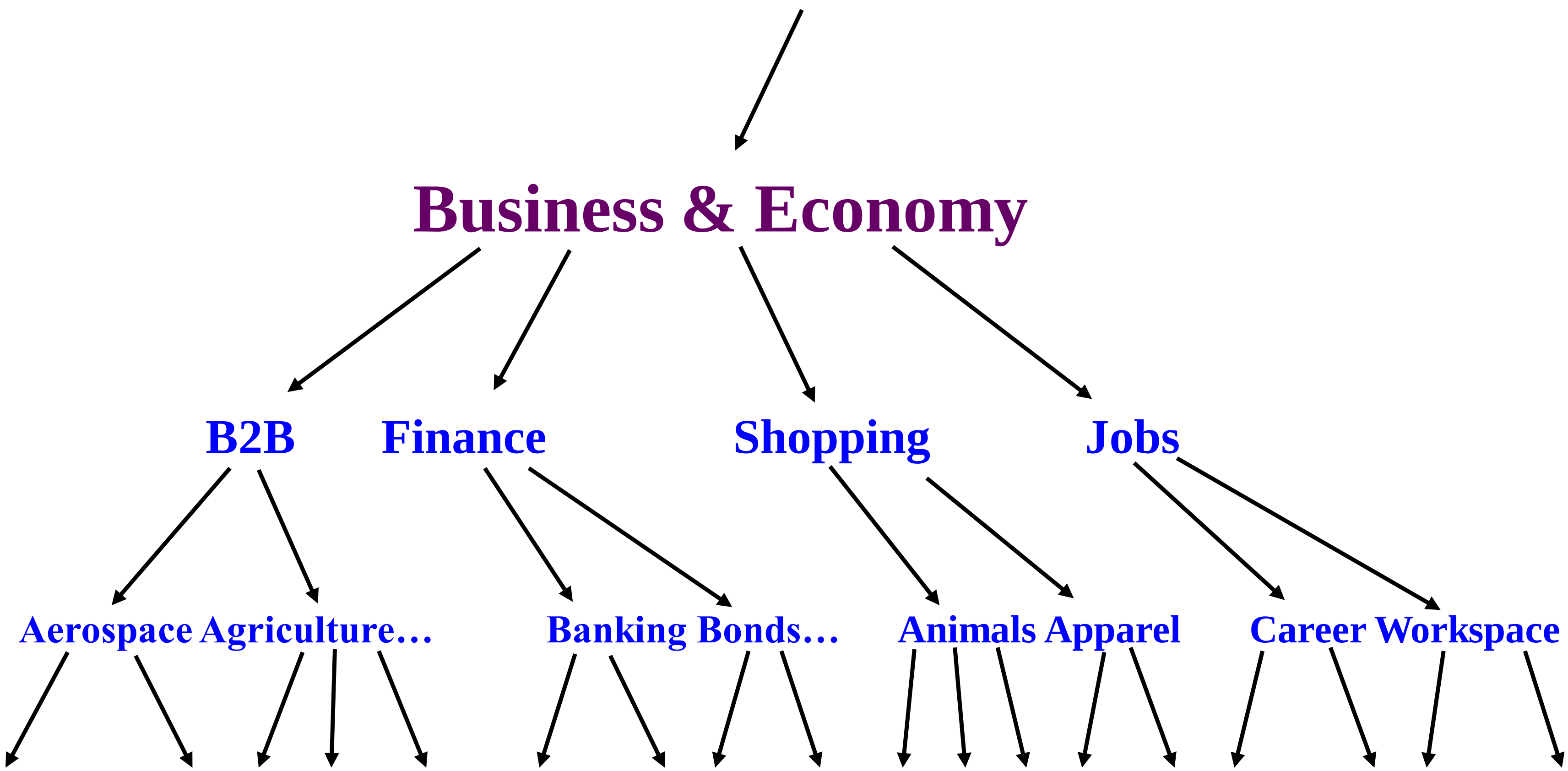
$\mathbf{P}_3 = \{ (\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6), (\mathbf{x}_2, \mathbf{x}_5) \}$

$\mathbf{P}_4 = \{ (\mathbf{x}_1, \mathbf{x}_2), (\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6), (\mathbf{x}_5) \}$

# Definição de Hierarquia

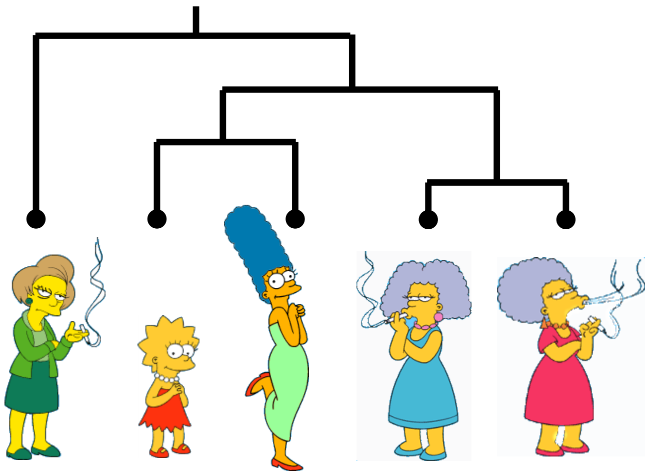
- Uma hierarquia completa:
  - Inicia ou termina com partição totalmente disjunta
    - ***Disjoint clustering***: apenas grupos **atômicos** (***singletons***)
    - Exemplo: $\mathbf{P} = \{ (\mathbf{x}_1), (\mathbf{x}_2), (\mathbf{x}_3), (\mathbf{x}_4), (\mathbf{x}_5), (\mathbf{x}_6) \}$
      - Também denominada "solução trivial"
  - Inicia ou termina com partição totalmente conjunta
    - ***Conjoint clustering***: grupo único com todos os objetos
    - Exemplo: $\mathbf{P} = \{ (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6) \}$

Hierarquias são comumente usadas para organizar informação, como, por exemplo, num portal


Keogh, E. A Gentle Introduction to Machine Learning and Data Mining for the Database Community, SBBD 2003, Manaus.

# Métodos Clássicos para Agrupamento Hierárquico



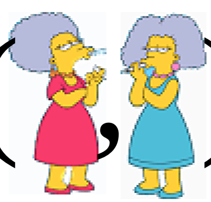
**Bottom-Up (aglomerativos):**
- Iniciar colocando cada objeto em um *cluster*
- Encontrar o melhor par de *clusters* para unir
- Unir o par de *clusters* escolhido
- Repetir até que todos os objetos estejam reunidos em um só *cluster*

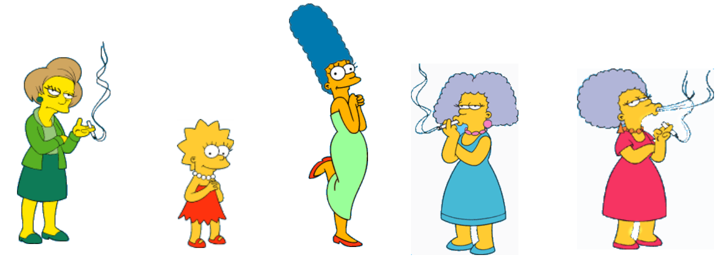
**Top-Down (divisivos):**
- Iniciar com todos objetos em um único *cluster*
- Sub-dividir o *cluster* em dois novos *clusters*
- Aplicar o algoritmo recursivamente em ambos, até que cada objeto forme um *cluster* por si só

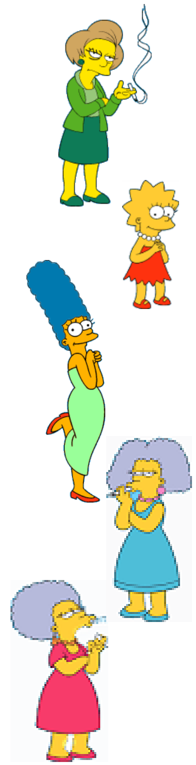
Keogh, E. A Gentle Introduction to Machine Learning and Data Mining for the Database Community, SBBD 2003, Manaus.

Algoritmos hierárquicos podem operar somente sobre uma matriz de distâncias: são (ou podem ser) **relacionais**.

D(  ) = 8

D(  ) = 1



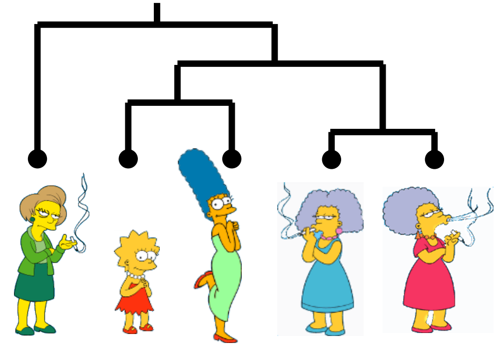


| | | | | |
|---|---|---|---|---|
| 0 | 8 | 8 | 7 | 7 |
| | 0 | 2 | 4 | 4 |
| | | 0 | 3 | 3 |
| | | | 0 | 1 |
| | | | | 0 |

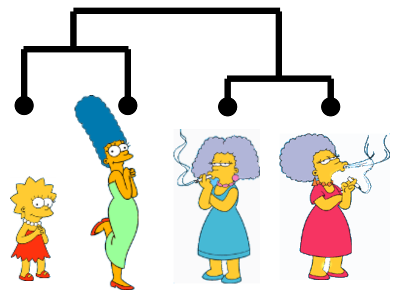
Keogh, E. A Gentle Introduction to Machine Learning and Data Mining for the Database Community, SBBD 2003, Manaus.

## Bottom-Up (aglomerativo):

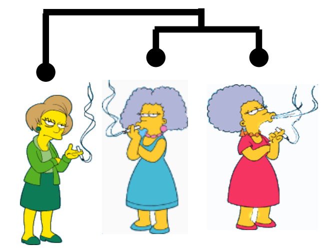
Iniciando com cada objeto em seu próprio cluster, encontrar o melhor par de *clusters* para unir em um novo *cluster.* Repetir até que todos os *clusters* sejam fundidos em um único *cluster*.



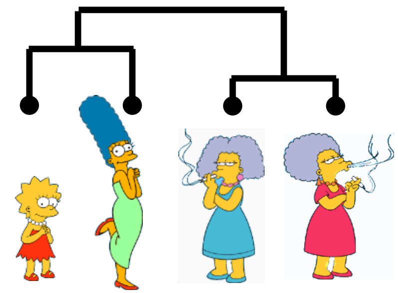
Considerar todas as uniões possíveis …



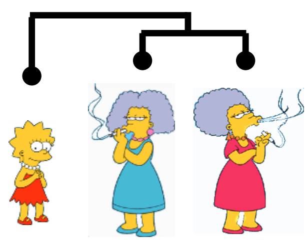
…

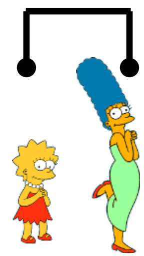


Escolher a melhor



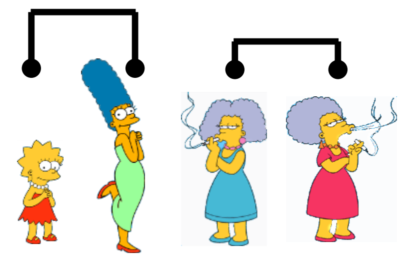
Considerar todas as uniões possíveis …



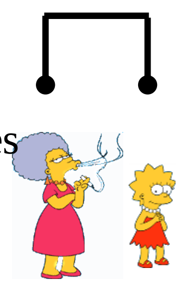


…

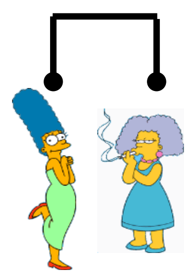
Escolher a melhor



Considerar todas as uniões possíveis …

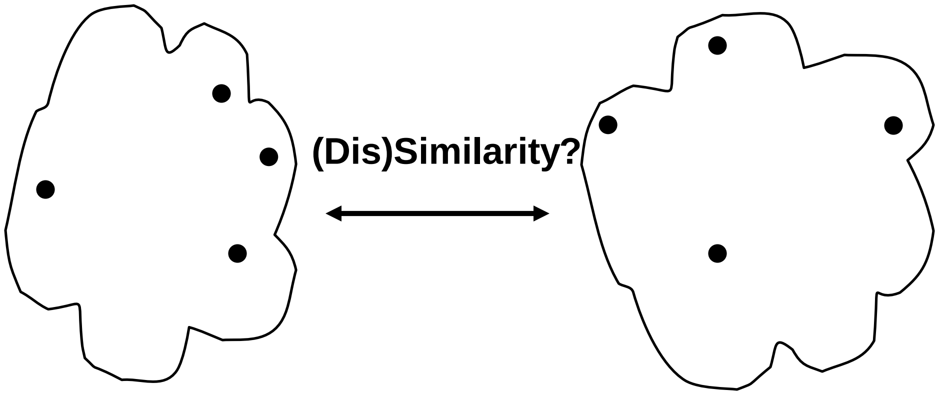


…



Escolher a melhor

# How to Define Inter-Cluster (Dis)Similarity

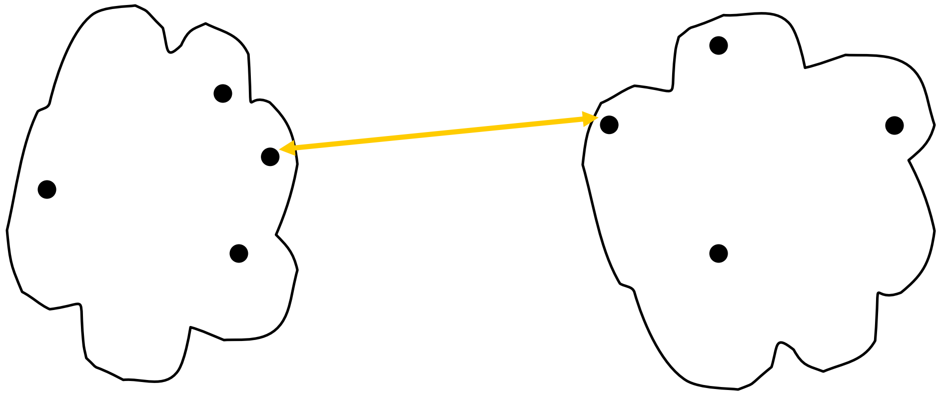



| | p1 | p2 | p3 | p4 | p5 | . . . |
|---|---|---|---|---|---|---|
| p1 | | | | | | |
| p2 | | | | | | |
| p3 | | | | | | |
| p4 | | | | | | |
| p5 | | | | | | |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |

**Proximity Matrix**

- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods
  - Ward's
  - …

# How to Define Inter-Cluster (Dis)Similarity

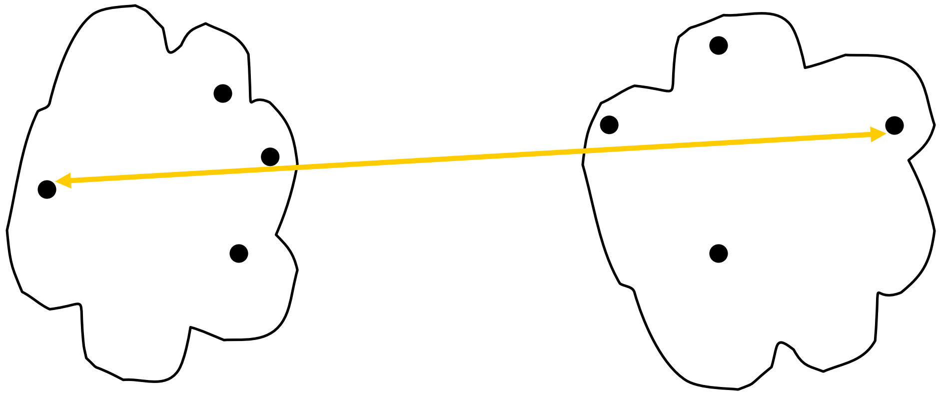


| | p1 | p2 | p3 | p4 | p5 | . . . |
|---|---|---|---|---|---|---|
| p1 | | | | | | |
| p2 | | | | | | |
| p3 | | | | | | |
| p4 | | | | | | |
| p5 | | | | | | |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |

**Proximity Matrix**

- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods
  – Ward’s
  – …

# How to Define Inter-Cluster (Dis)Similarity

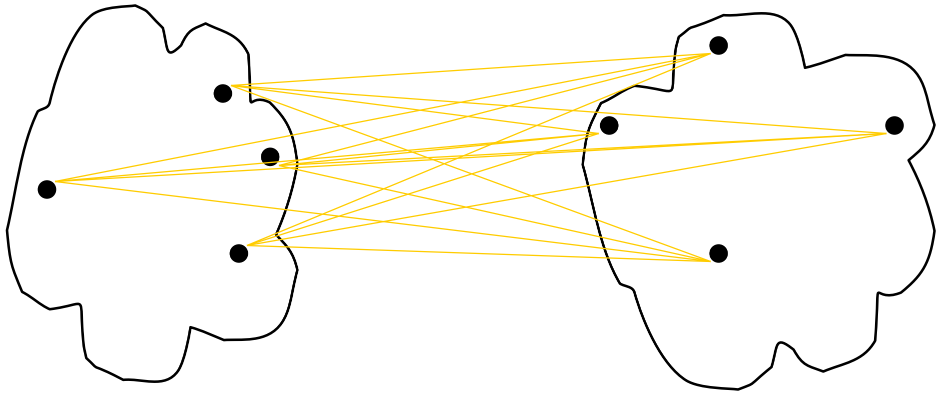


| | p1 | p2 | p3 | p4 | p5 | . . . |
|---|---|---|---|---|---|---|
| p1 | | | | | | |
| p2 | | | | | | |
| p3 | | | | | | |
| p4 | | | | | | |
| p5 | | | | | | |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |

**Proximity Matrix**

- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods
  - Ward's
  - …

# How to Define Inter-Cluster (Dis)Similarity

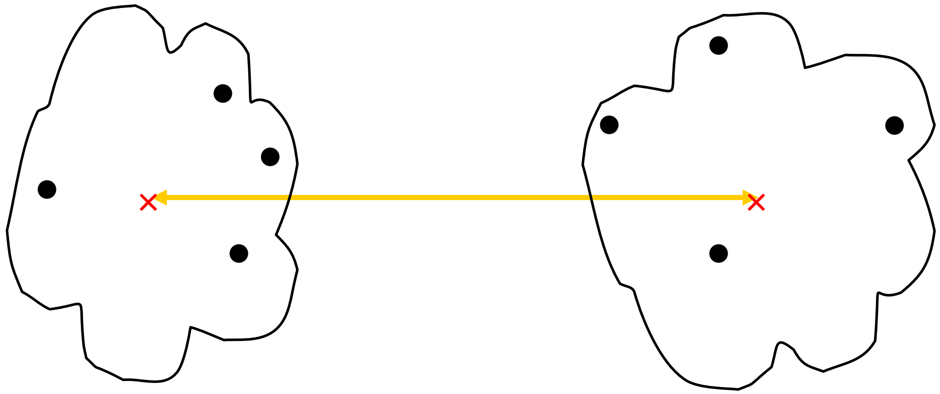


| | p1 | p2 | p3 | p4 | p5 | . . . |
|---|---|---|---|---|---|---|
| p1 | | | | | | |
| p2 | | | | | | |
| p3 | | | | | | |
| p4 | | | | | | |
| p5 | | | | | | |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |

**Proximity Matrix**

- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods
  - Ward's
  - …

# How to Define Inter-Cluster (Dis)Similarity



| | p1 | p2 | p3 | p4 | p5 | . . . |
|---|---|---|---|---|---|---|
| p1 | | | | | | |
| p2 | | | | | | |
| p3 | | | | | | |
| p4 | | | | | | |
| p5 | | | | | | |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |

**Proximity Matrix**

- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods
  - Ward’s
  - …

# Como Comparar os Clusters?

- ***Single Linkage*** , Min, ou Vizinho mais Próximo :
  - Dissimilaridade entre *clusters* é dada pela menor dissimilaridade entre 2 objetos (um de cada *cluster*)

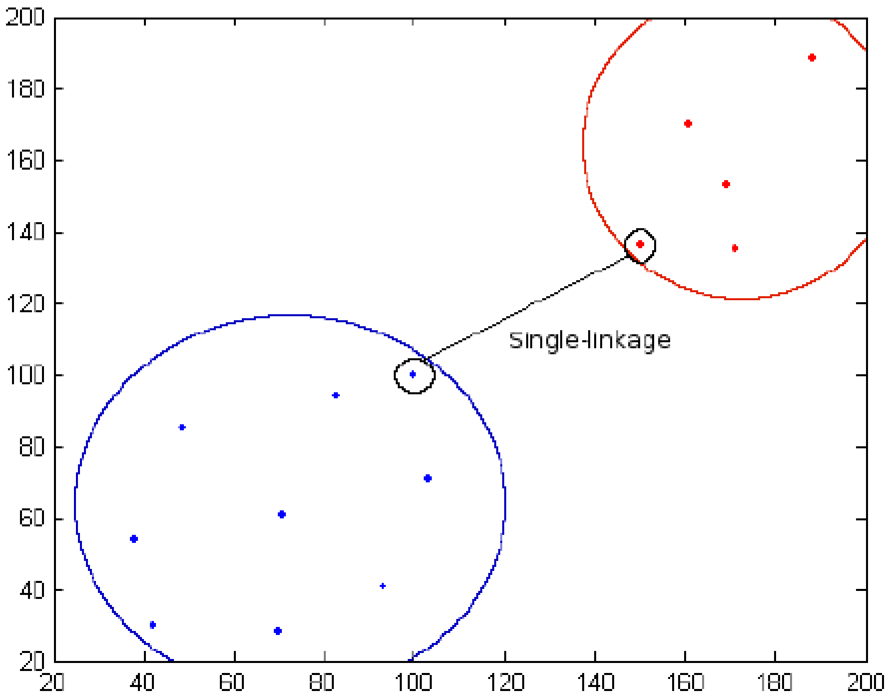


Figura por Lucas Vendramin

***single link*** (Florek, 1951; Sneath, 1957)

<u>Originalmente baseado em **Grafos**</u>:
menor aresta entre dois vértices
de subconjuntos distintos

# Propriedade Útil

- Propriedade da Função Mínimo (min):
  - min{$\mathbf{D}$} = min{ min{$\mathbf{D}_1$} , min{$\mathbf{D}_2$} }
    - $\mathbf{D}$, $\mathbf{D}_1$ e $\mathbf{D}_2$ são conjuntos de valores reais tais que $\mathbf{D}_1 \cup \mathbf{D}_2 = \mathbf{D}$
  - Exemplo:
    - min{10, -3, 0, 100} = min { min{10, -3}, min{0, 100} } = -3
  - Propriedade vale recursivamente (para min{$\mathbf{D}_1$} e min{$\mathbf{D}_2$})
- Utilidade para *Single-Linkage*
  - Dada a distância entre os grupos **A** e **B** e entre **A** e **C**
    - É trivial calcular a distância entre **A** e (**B** $\cup$ **C**).

## **Exemplo de Single Linkage**: Método de Johnson (1967)

- Consideremos a seguinte matriz de distâncias iniciais ($\mathbf{D}_1$) entre 5 objetos {1,2,3,4,5}. Qual par de objetos será escolhido para formar o 1º *cluster* ?

$$\mathbf{D}_1 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 0 & & & & \\ \boxed{2} & 0 & & & \\ 6 & 5 & 0 & & \\ 10 & 9 & 4 & 0 & \\ 9 & 8 & 5 & 3 & 0 \end{bmatrix}$$

- A menor distância entre objetos é $d_{12}=d_{21}=2$, indicando que estes dois objetos serão unidos em um *cluster*. Na seqüência, calcula-se:

$d_{(12)3}=min\{d_{13},d_{23}\}=d_{23}=5;$

$d_{(12)4}=min\{d_{14},d_{24}\}=d_{24}=9;$

$d_{(12)5}=min\{d_{15},d_{25}\}=d_{25}=8;$

- Desta forma, obtém-se uma nova matriz de distâncias ($\mathbf{D}_2$), que será usada na próxima etapa do agrupamento hierárquico:



Exemplo de (Everitt et al., 2001)

$$\mathbf{D}_2 = \begin{array}{c} 12 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 0 & & & \\ 5 & 0 & & \\ 9 & 4 & 0 & \\ 8 & 5 & \boxed{3} & 0 \end{bmatrix}$$

- Qual o novo *cluster* a ser formado?
- Unindo os objetos **4** e **5** obtemos três clusters: {1,2}, {4,5}, {3}
- Como $d_{(12)3}$ já está calculada, calculamos na sequência:

$d_{(12)(45)} = min\{d_{(12)(4)}, d_{(12)(5)}\} = d_{(12)(5)} = 8$

$d_{(45)3} = min\{d_{43}, d_{53}\} = d_{43} = 4$

obtendo a seguinte matriz:

$$\mathbf{D}_3 = \begin{array}{c} 12 \\ 3 \\ 45 \end{array} \begin{bmatrix} 0 & & \\ 5 & 0 & \\ 8 & \boxed{4} & 0 \end{bmatrix}$$

* Unir *cluster* {3} com {4,5};

* Finalmente, unir todos os *clusters* em um único *cluster*

- A sequência de partições obtidas neste exemplo é, portanto:

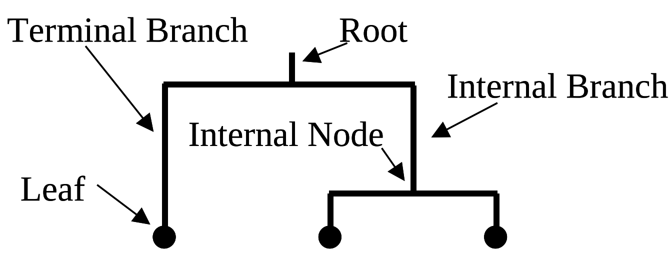
  { (1), (2), (3), (4), (5) } → { (1, 2), (3), (4), (5) } →

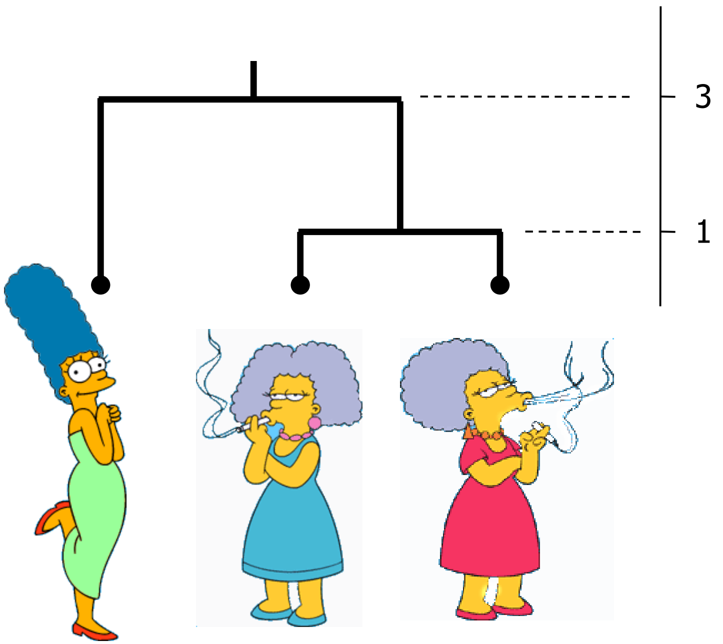
  { (1, 2), (3), (4, 5) } → { (1, 2), (3, 4, 5) } → { (1, 2, 3, 4, 5) }

- **Nota**: Para *single link*, a dissimilaridade entre 2 clusters pode ser computada naturalmente a partir da matriz atualizada na iteração anterior, sem necessidade da matriz original
  - Isso vale devido à propriedade da função *min* vista anteriormente
  - No nosso exemplo, simplificamos o cálculo de $d_{(12)(45)}$ como $\min\{d_{(12)(4)}, d_{(12)(5)}\}$ fazendo uso daquela propriedade:
    - $\min\{d_{(12)(4)}, d_{(12)(5)}\} = \min\{9, 8\} = \min\{d_{14}, d_{24}, d_{15}, d_{25}\}$

# Dendrograma

**Dendrograma**: Hierarquia + Dissimilaridades entre Clusters



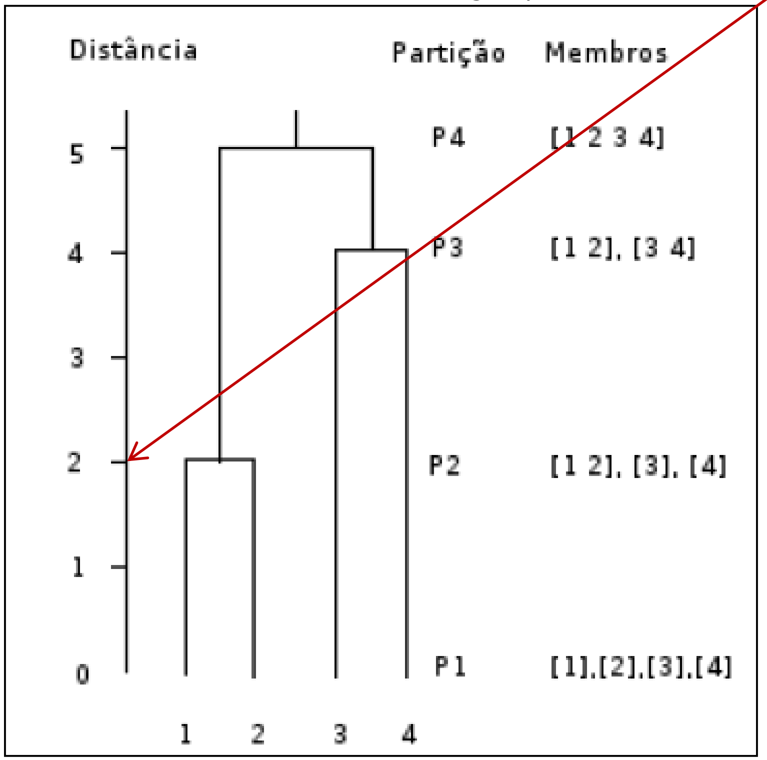

* A dissimilaridade entre dois clusters (possivelmente **singletons**) é representada como a altura do nó interno mais baixo compartilhado

Keogh, E. A Gentle Introduction to Machine Learning and Data Mining for the Database Community, SBBD 2003, Manaus.

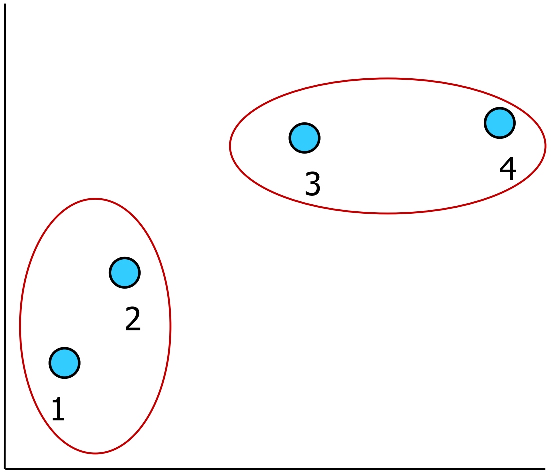
# Exemplo de Dendrograma

$$\mathbf{D} = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} 0 & 2 & 7 & 13 \\ 2 & 0 & 5 & 10 \\ 7 & 5 & 0 & 4 \\ 13 & 10 & 4 & 0 \end{bmatrix}$$

Figura por Lucas Vendramin

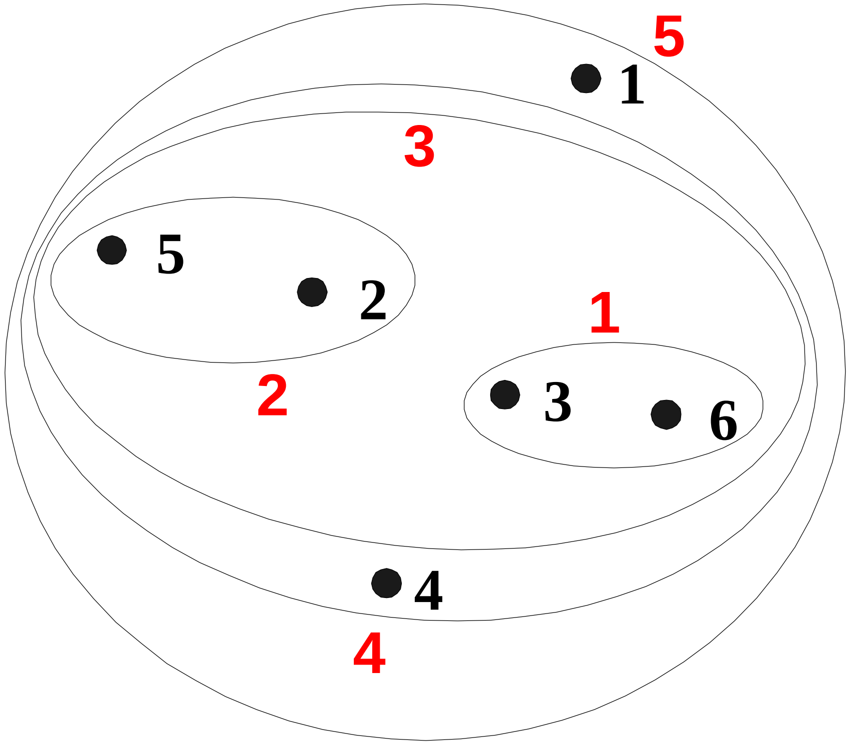


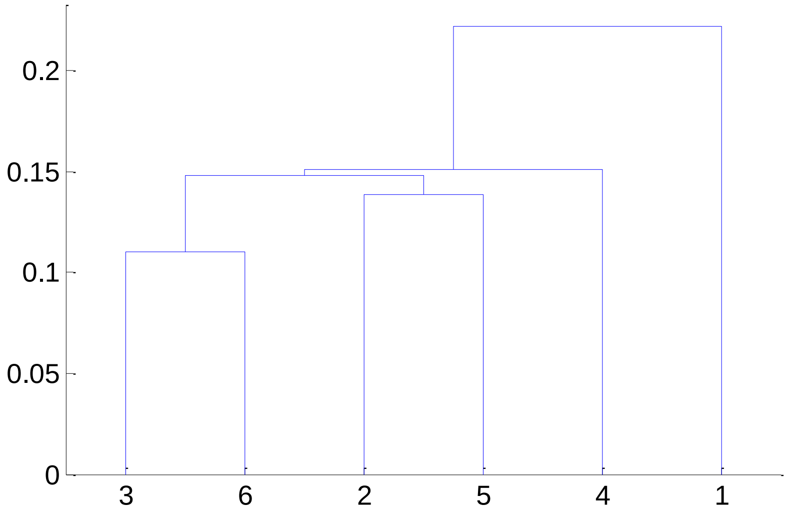

Dendrograma




uma das partições aninhadas

# Outro Exemplo de Dendrograma




***Nested Clusters*** ***Dendrogram***

# Exercício:

- Obtenha o dendrograma completo para o exemplo visto de execução do *single linkage* (matriz de distâncias abaixo)

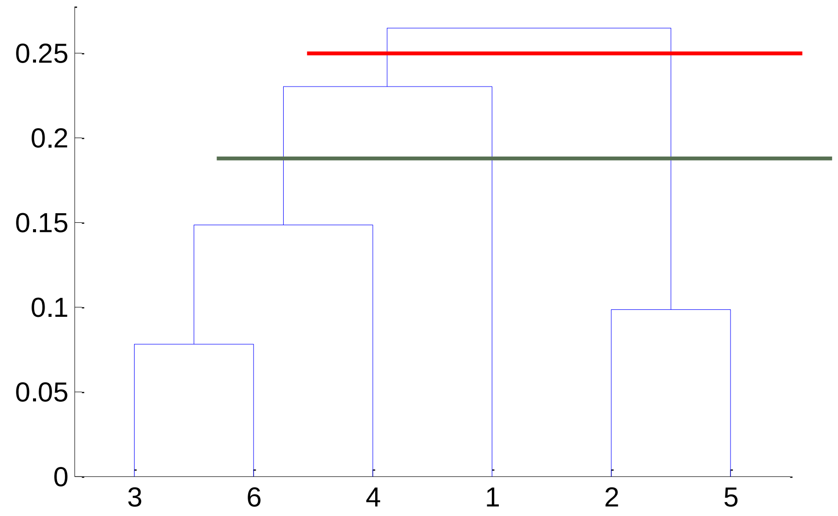
$$\mathbf{D}_1 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 0 & & & & \\ 2 & 0 & & & \\ 6 & 5 & 0 & & \\ 10 & 9 & 4 & 0 & \\ 9 & 8 & 5 & 3 & 0 \end{bmatrix}$$

# Dendrogramas e Partições

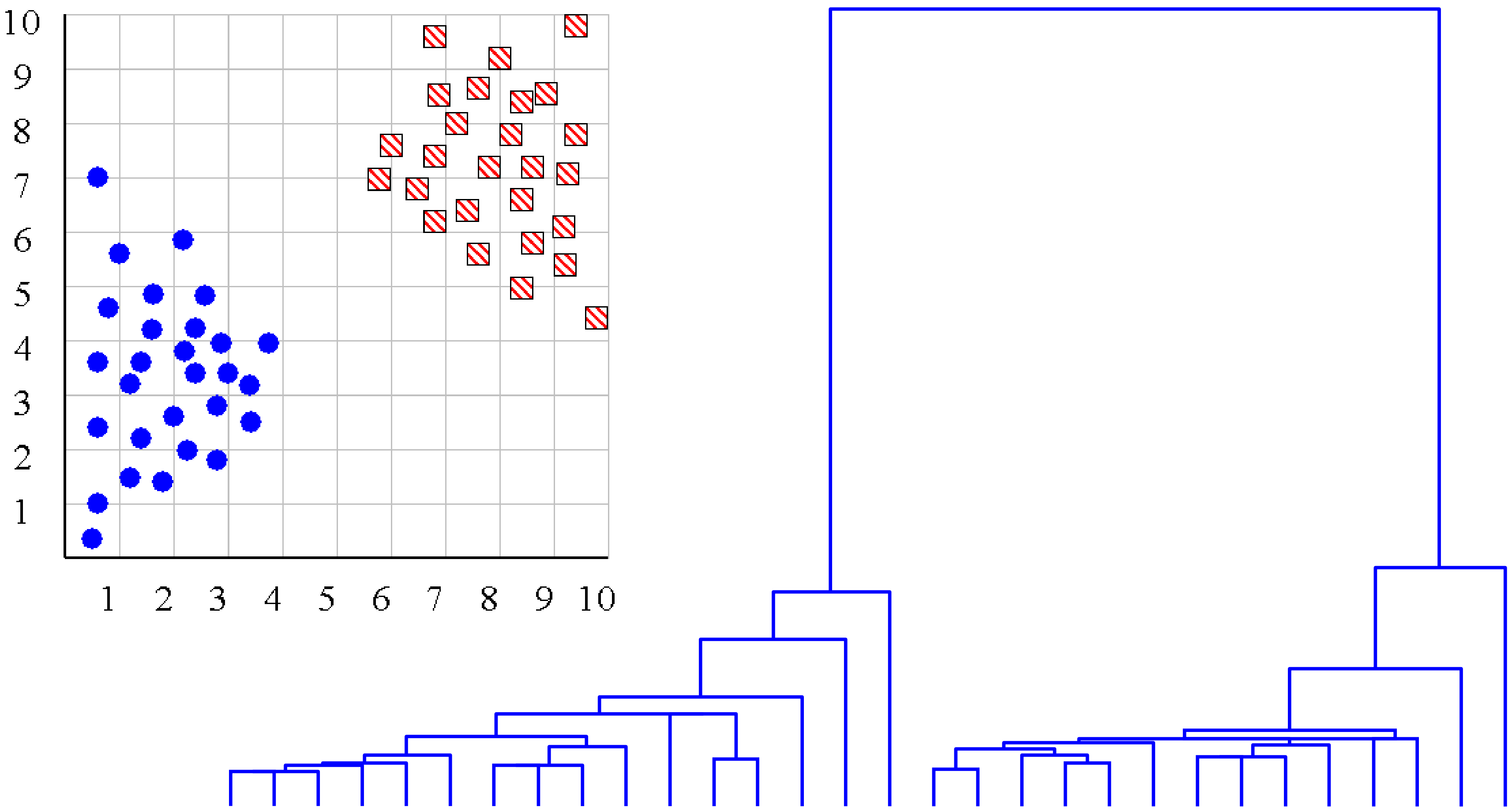
- Partições são obtidas via **cortes** no dendrograma
  - cortes horizontais
  - no. de grupos da partição = no. de interseções
- **Exemplos:**

$P_2 = \{ (\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6), (\mathbf{x}_2, \mathbf{x}_5) \}$

$P_1 = \{ (\mathbf{x}_1), (\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6), (\mathbf{x}_2, \mathbf{x}_5) \}$



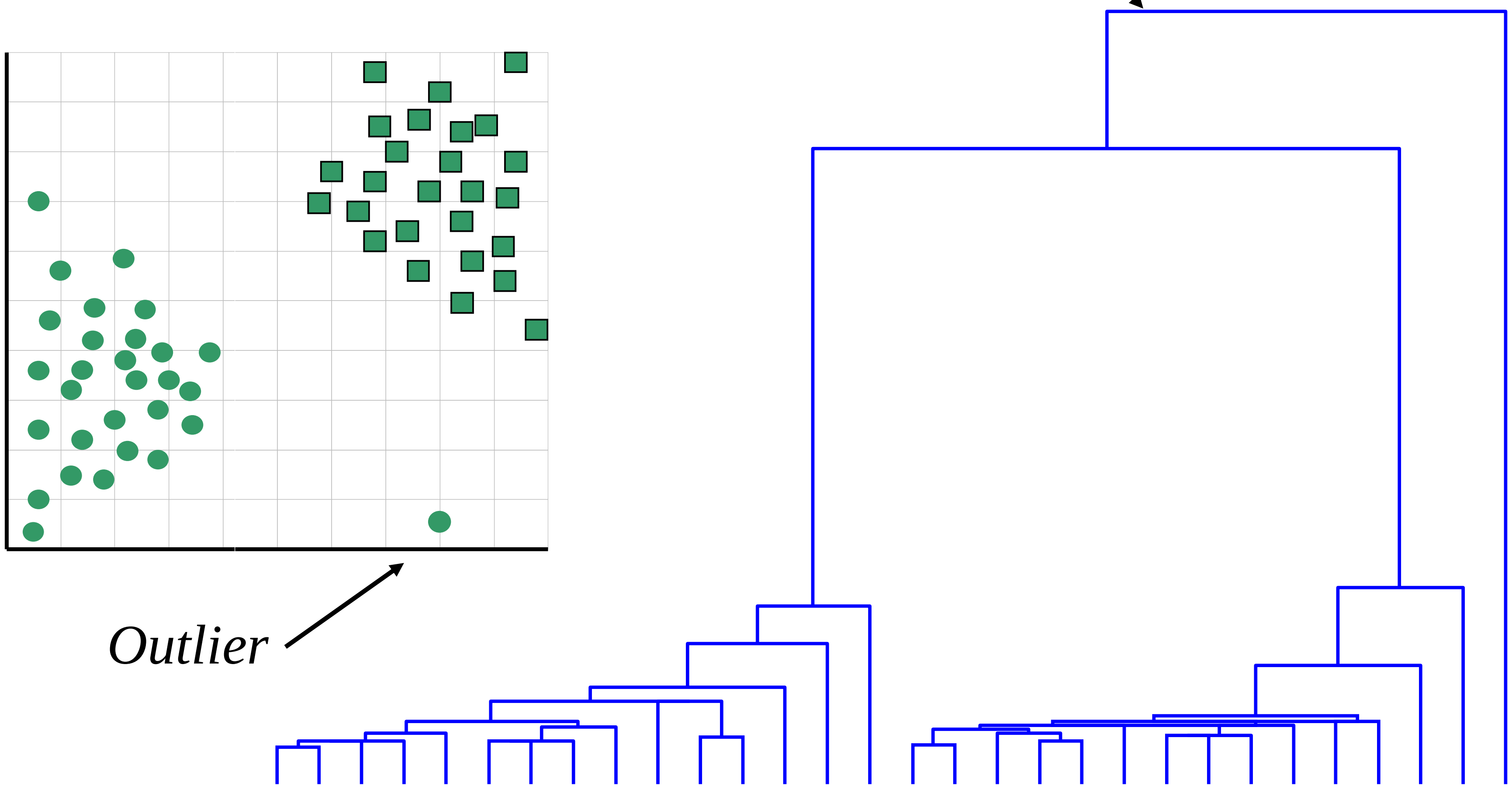
Pode-se examinar o dendrograma para tentar estimar o número *mais natural* de clusters. No caso abaixo, existem duas sub-árvores bem separadas, sugerindo dois grupos de dados. Infelizmente, na prática, as distinções não são tão simples…




Keogh, E. A Gentle Introduction to Machine Learning and Data Mining for the Database Community, SBBD 2003, Manaus.

# Pode-se usar o dendrograma para tentar detectar *outliers*:

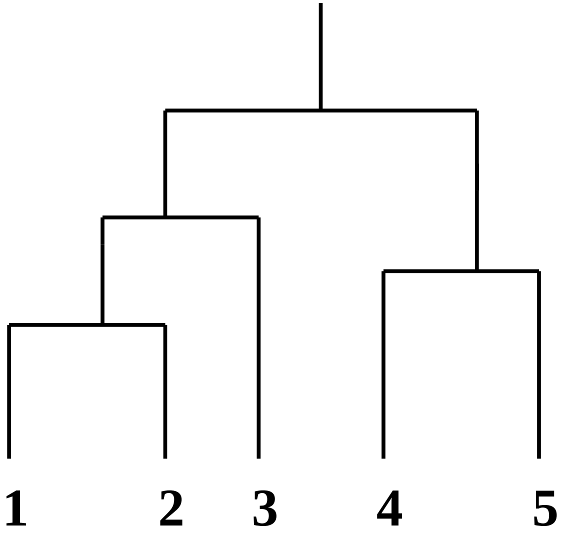
Ramo isolado sugere que o objeto é muito diferente dos demais.



# Voltando ao Single Linkage (Min)...

- Similarity of two clusters is based on the two most similar (closest) points in the clusters
  - Determined by **one pair of points**
  - i.e., by **one link** in the the **proximity graph**

| | I1 | I2 | I3 | I4 | I5 |
|---|---|---|---|---|---|
| I1 | 1.00 | 0.90 | 0.10 | 0.65 | 0.20 |
| I2 | 0.90 | 1.00 | 0.70 | 0.60 | 0.50 |
| I3 | 0.10 | 0.70 | 1.00 | 0.40 | 0.30 |
| I4 | 0.65 | 0.60 | 0.40 | 1.00 | 0.80 |
| I5 | 0.20 | 0.50 | 0.30 | 0.80 | 1.00 |



# Strength of MIN

- **Can handle non-elliptical shapes**

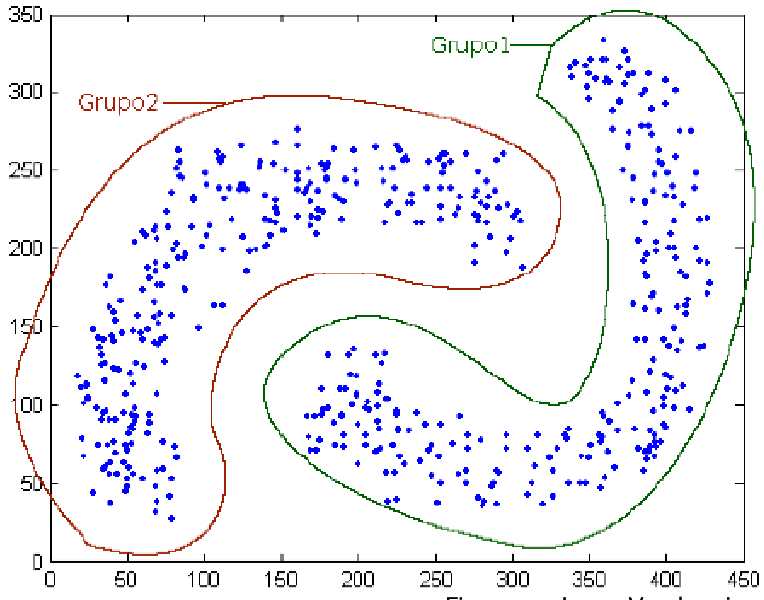
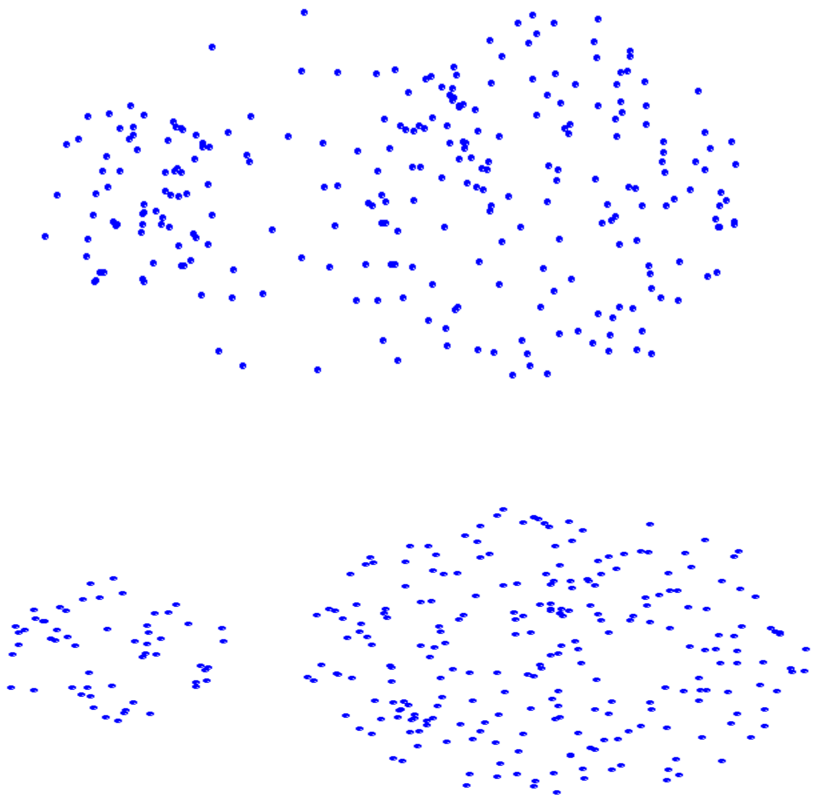


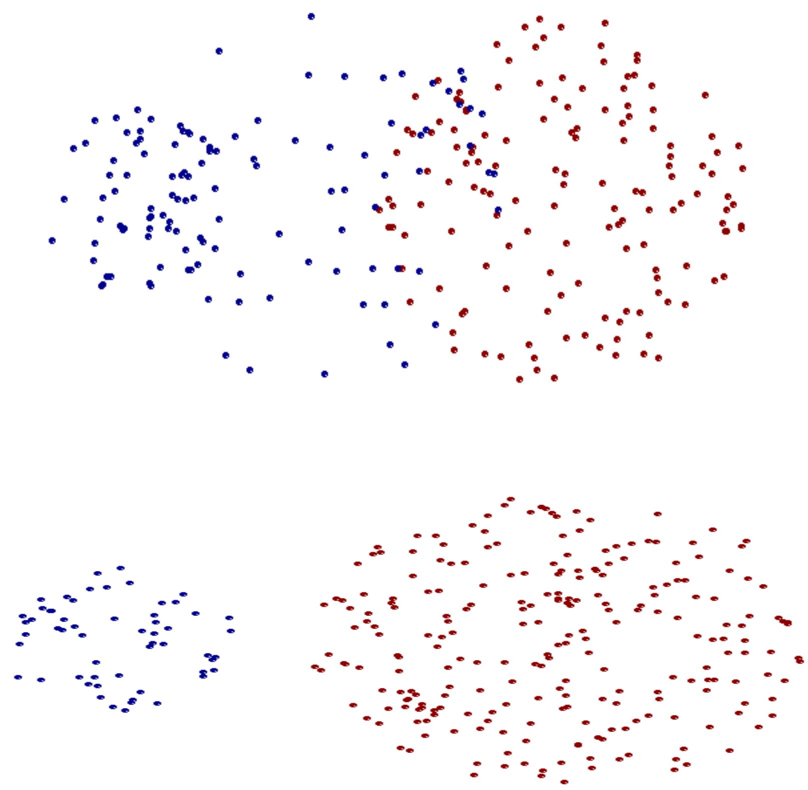

Figura por Lucas Vendramin

# Main Limitations of MIN

- **Sensitive to noise and outliers**



**Original Points**



**Two Clusters**

# Como Comparar os Clusters?

- ***Complete Linkage*** , Max, ou Vizinho mais Distante:
  - Dissimilaridade entre *clusters* é dada pela maior dissimilaridade entre dois objetos (um de cada *cluster*)

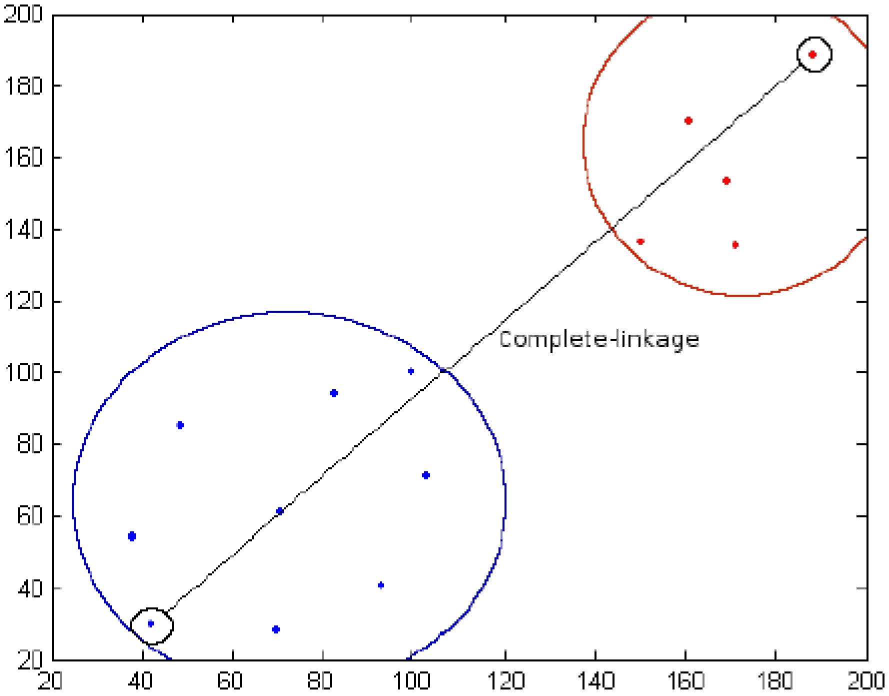



Figura por Lucas Vendramin

***complete link*** (Sorensen, 1948)

<u>Originalmente baseado em **Grafos**</u>: maior aresta entre dois vértices de subconjuntos distintos

# Propriedade Útil

- Propriedade da Função Máximo (max):
  - $\max\{\mathbf{D}\} = \max\{ \max\{\mathbf{D}_1\} , \max\{\mathbf{D}_2\} \}$
    - $\mathbf{D}$, $\mathbf{D}_1$ e $\mathbf{D}_2$ são conjuntos de valores reais tais que $\mathbf{D}_1 \cup \mathbf{D}_2 = \mathbf{D}$
  - Exemplo:
    - $\max\{10, -3, 0, 100\} = \max \{ \max\{10, -3\}, \max\{0, 100\} \} = 100$
  - Propriedade vale recursivamente (para $\max\{\mathbf{D}_1\}$ e $\max\{\mathbf{D}_2\}$)
- Utilidade para *Complete-Linkage*
  - Dada a distância entre os grupos **A** e **B** e entre **A** e **C**
    - É trivial calcular a distância entre **A** e (**B** $\cup$ **C**).

- Seja a seguinte matriz de distâncias iniciais ($\mathbf{D}_1$) entre 5 objetos :

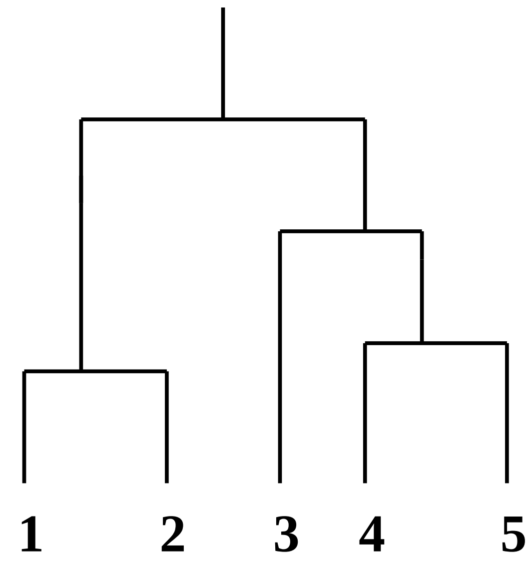
$$\mathbf{D}_1 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 0 & & & & \\ 2 & 0 & & & \\ 6 & 5 & 0 & & \\ 10 & 9 & 4 & 0 & \\ 9 & 8 & 5 & 3 & 0 \end{bmatrix}$$

- <u>Exercício</u>: executar o *complete linkage* através de sucessivas atualizações da matriz de distâncias (método de Johnson).

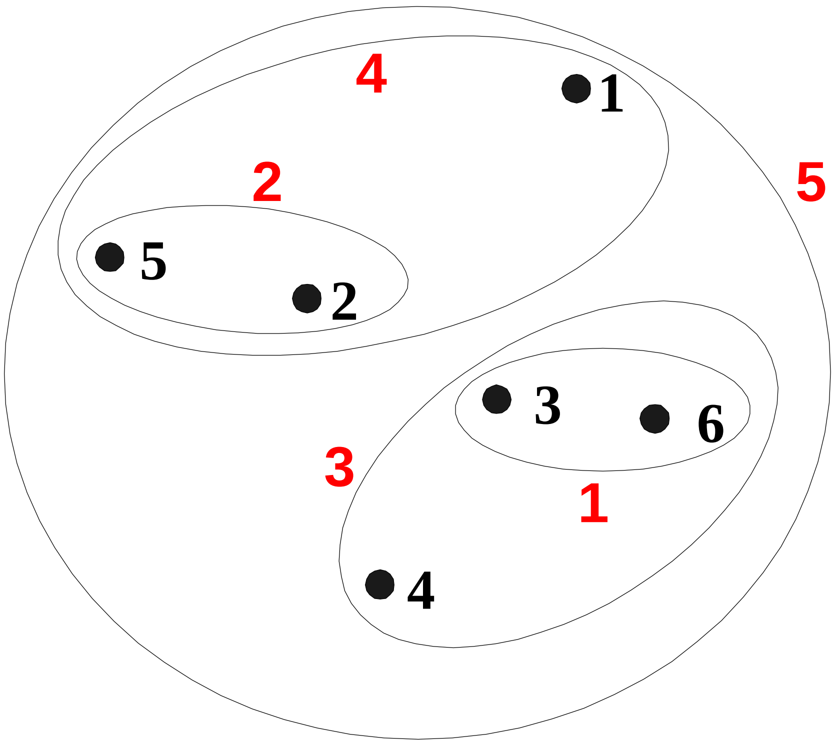
# Cluster Similarity: MAX or Complete Linkage

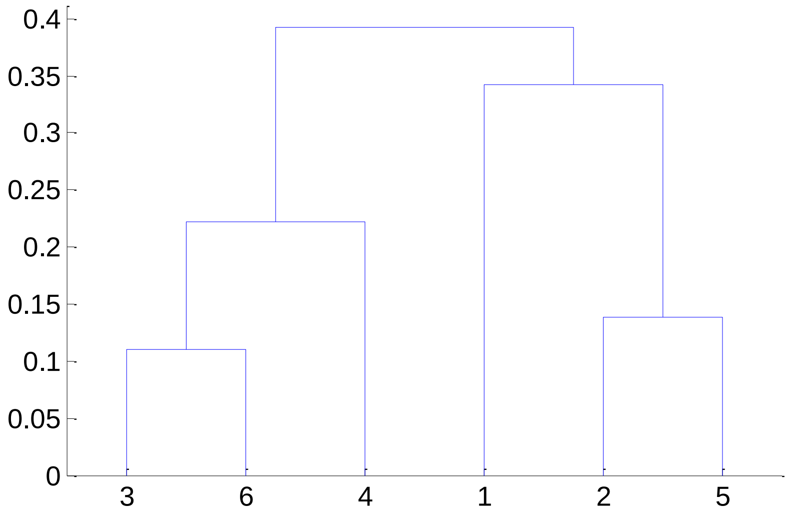
- Similarity of two clusters is based on the two least similar (most distant) points in the clusters
  - Determined by **one pair of points**

| | I1 | I2 | I3 | I4 | I5 |
|---|---|---|---|---|---|
| I1 | 1.00 | 0.90 | 0.10 | 0.65 | 0.20 |
| I2 | 0.90 | 1.00 | 0.70 | 0.60 | 0.50 |
| I3 | 0.10 | 0.70 | 1.00 | 0.40 | 0.30 |
| I4 | 0.65 | 0.60 | 0.40 | 1.00 | 0.80 |
| I5 | 0.20 | 0.50 | 0.30 | 0.80 | 1.00 |



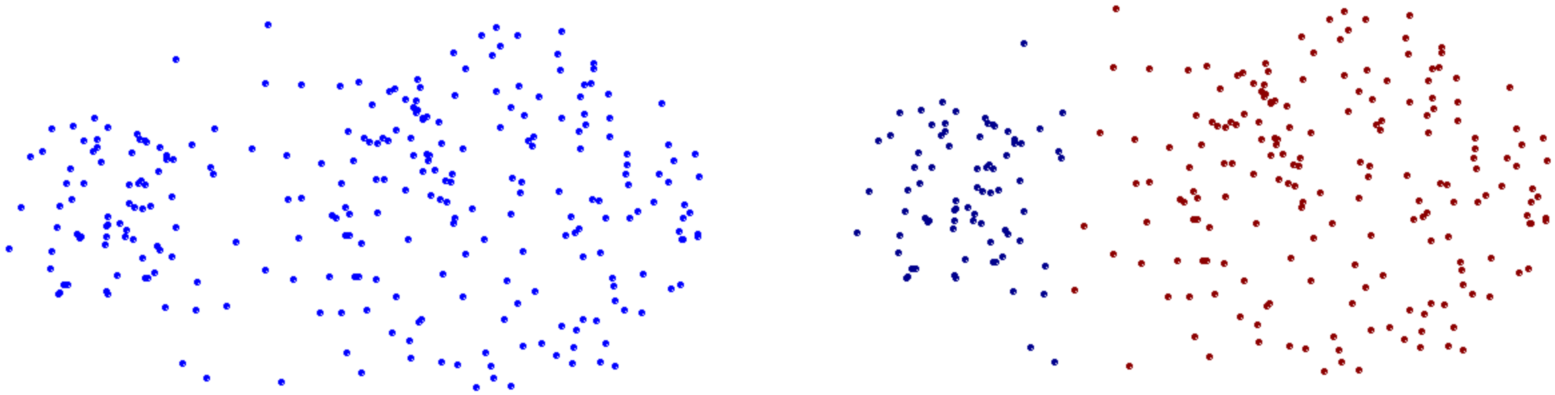
# Hierarchical Clustering: MAX




**Nested Clusters** **Dendrogram**

# Strength of MAX



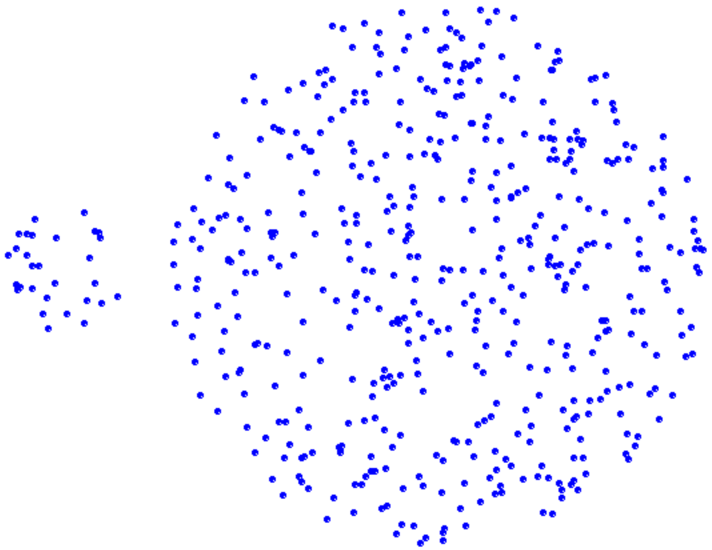
Original Points

Two Clusters

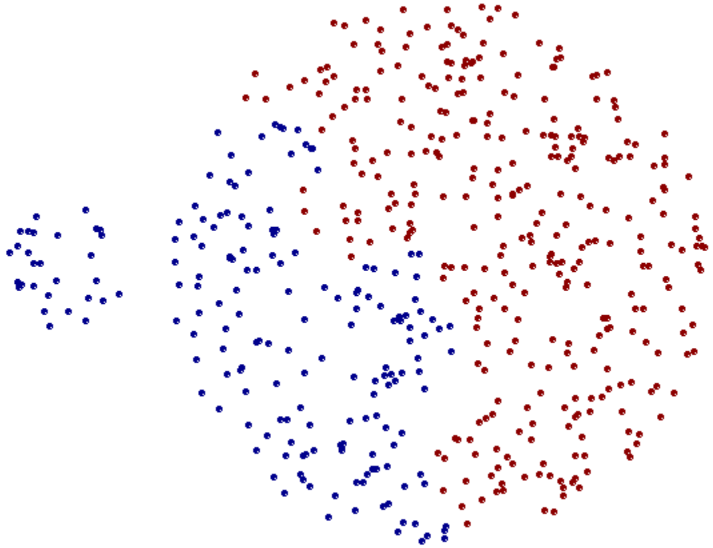
- **Less susceptible to noise and outliers**

# Main Limitations of MAX

**Original Points**



**Two Clusters**



- **Tends to break large clusters**
- **Biased towards globular clusters**

# Referências


- Jain, A. K. and Dubes, R. C., Algorithms for Clustering Data, Prentice Hall, 1988
- Everitt, B. S., Landau, S., and Leese, M., *Cluster Analysis*, Arnold, 4th Edition, 2001.
- Tan, P.-N., Steinbach, M., and Kumar, V., *Introduction to Data Mining*, Addison-Wesley, 2006