Chameleon

Algoritmo de agrupamento hierárquico baseado em conectividade e proximidade

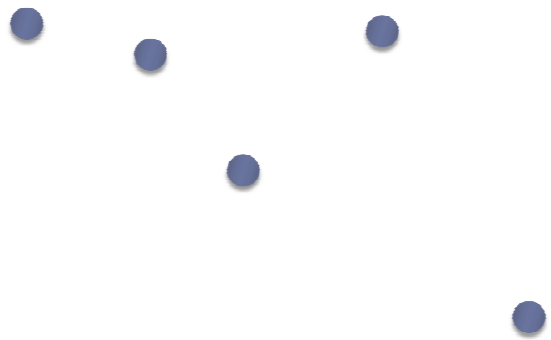
Chameleon

- Hierárquico aglomerativo.
- Encontrar grupos de formato arbitrário.
  - Sem certas desvantagens do Single-link.
- Modela os dados como um grafo de vizinhos mais próximos.
- Modelo dinâmico de similaridade entre grupo.
- Uso de proximidade e conectividade no modelo de similaridade entre grupo.

Algoritmo Chameleon

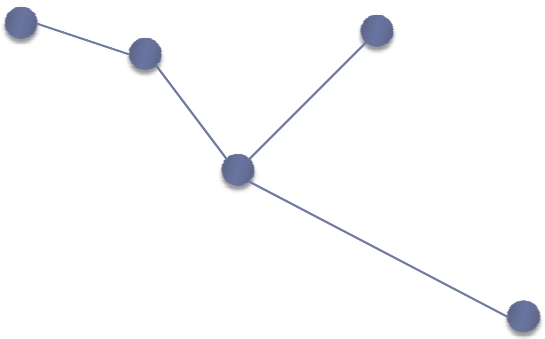
- **INPUT**: R, **k**
- Construção do grafo dos **k** vizinhos mais próximos

Grafo dos **k** vizinhos mais próximos



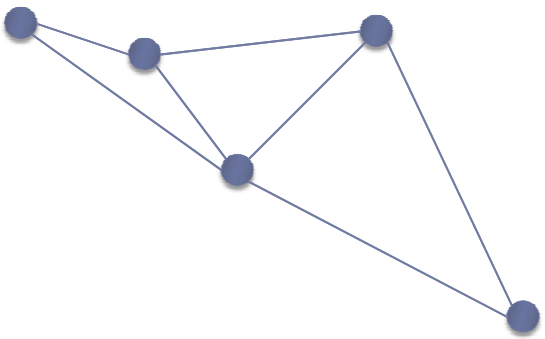
k=0

Grafo dos **k** vizinhos mais próximos



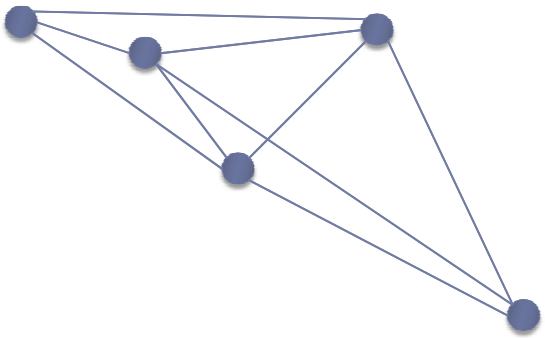
k=1

Grafo dos **k** vizinhos mais próximos



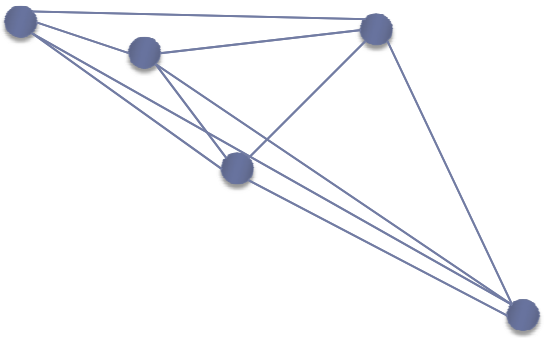
k=2

Grafo dos **k** vizinhos mais próximos



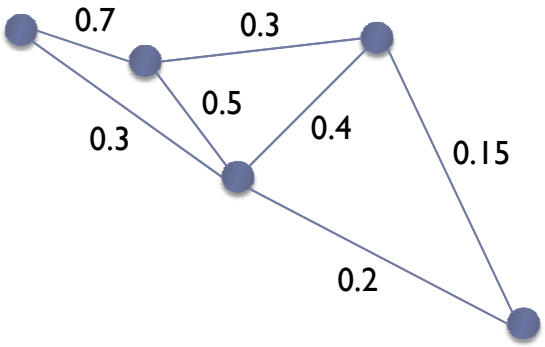
k=3

Grafo dos **k** vizinhos mais próximos



k=4

Grafo dos **k** vizinhos mais próximos

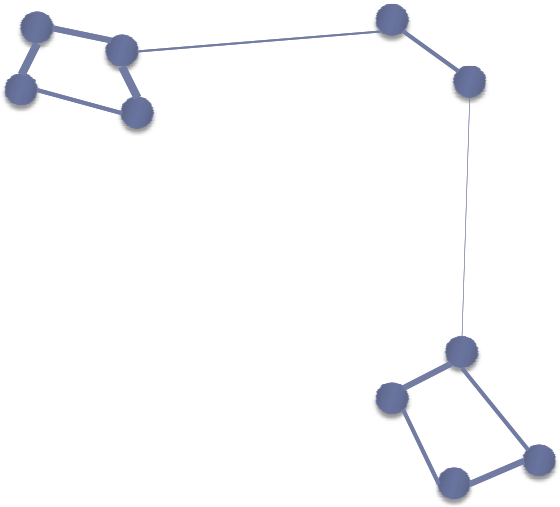


k=2

Algoritmo Chameleon

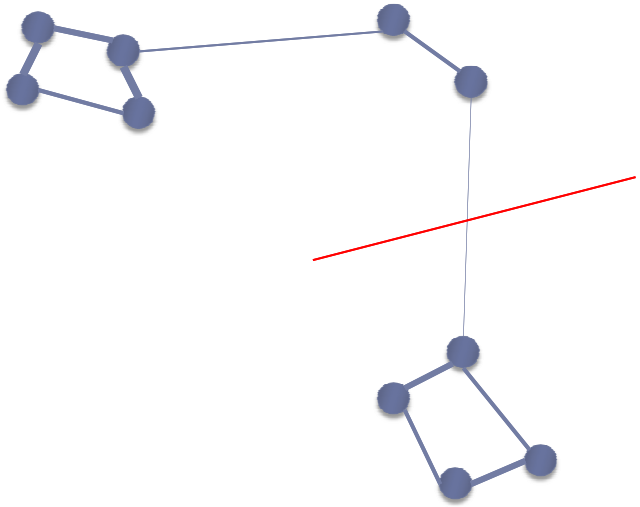
- **INPUT**: R, **k**, minpts
- Construção do grafo dos **k** vizinhos mais próximos
- Particionamento do grafo
  - Consiste na divisão balanceada dos vértices em grupos de forma a minimizar a soma dos pesos das arestas entre os grupos

Particionamento Inicial

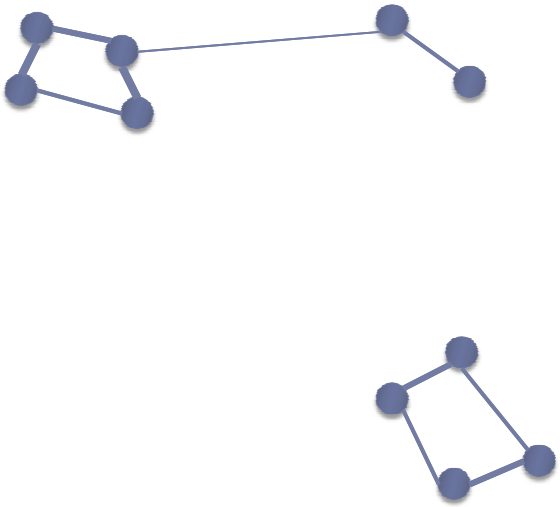


k=2

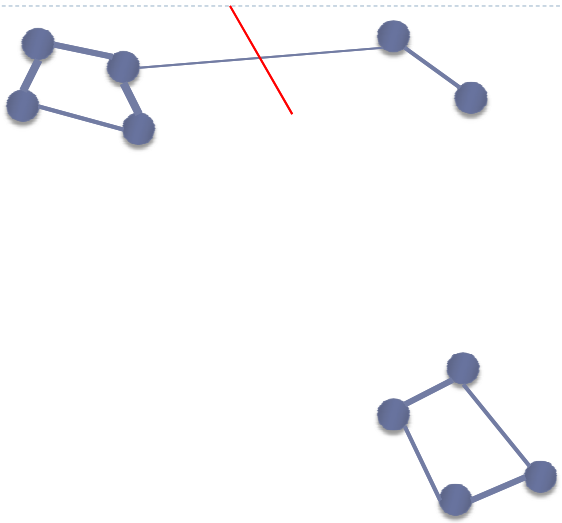
Particionamento Inicial



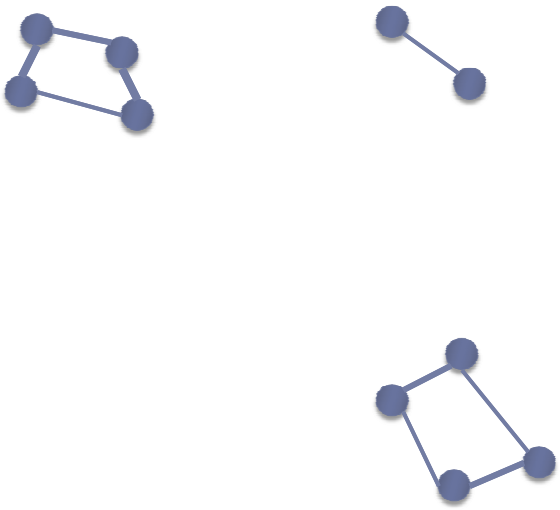
Particionamento Inicial



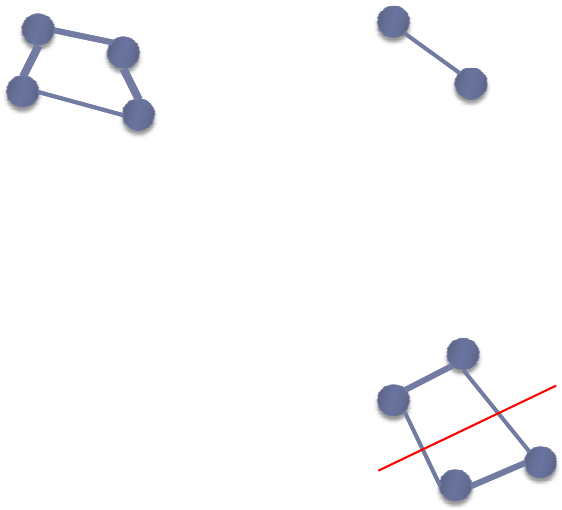
Particionamento Inicial



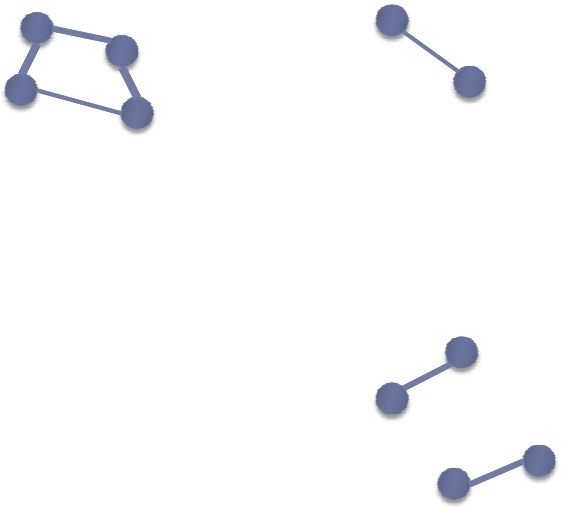
Particionamento Inicial



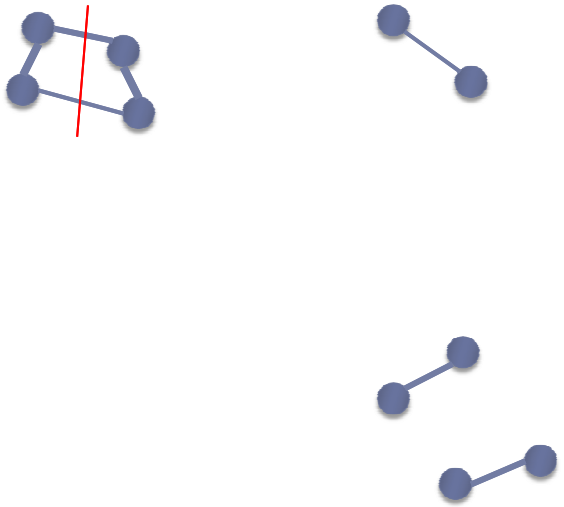
Particionamento Inicial



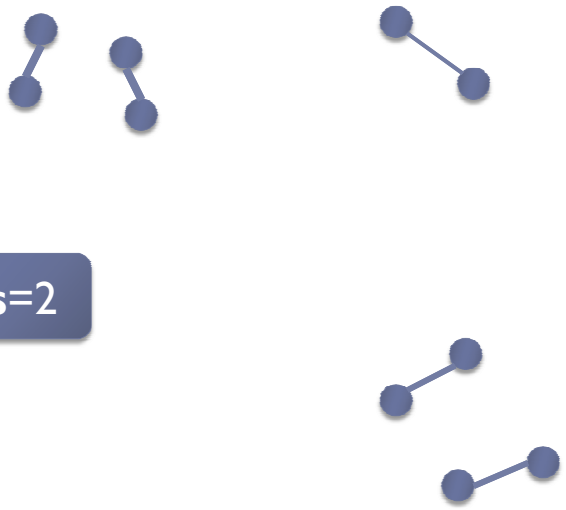
Particionamento Inicial



Particionamento Inicial

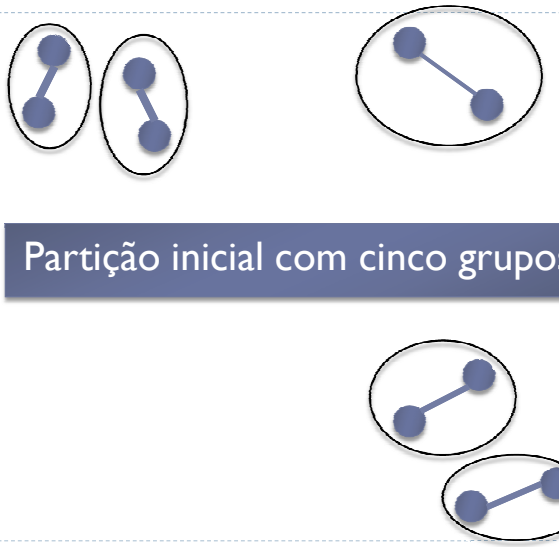


Particionamento Inicial



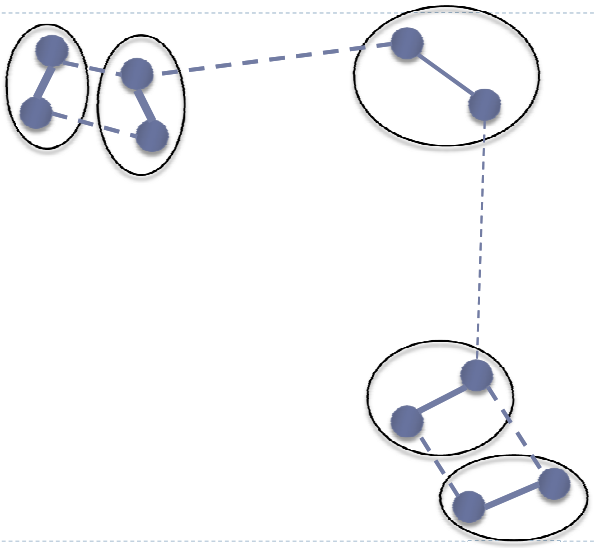
minpts=2

Particionamento Inicial



Partição inicial com cinco grupos

Particionamento Inicial



Algoritmo Chameleon

- **INPUT**: R, **k**, minpts
- Construção do grafo dos **k** vizinhos mais próximos
- Particionamento do grafo
- **Faça**
  - Una os dois grupos mais similares
- **Enquanto** puder unir grupos

Algoritmo Chameleon

- **INPUT**: R, **k**, minpts, α
- Construção do grafo dos **k** vizinhos mais próximos
- Particionamento do grafo
- **Faça**
  - Una os dois grupos mais similares
- **Enquanto** puder unir grupos

$$S(G_i, G_j) = CR(G_i, G_j) \cdot PR(G_i, G_j)$$

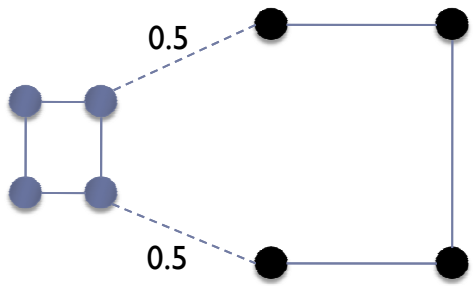
Conectividade Relativa

- Conectividade relativa é estabelecida a partir de conectividades absolutas

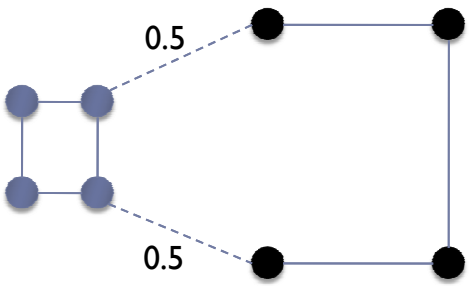
$$CA(G_i, G_j)$$

$$CA(G_i) \quad CA(G_j)$$

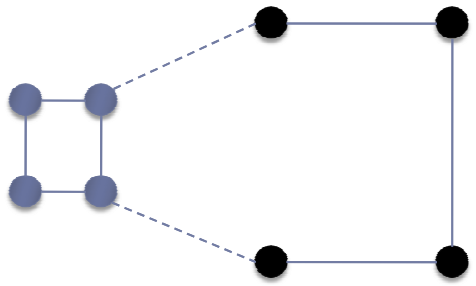
$$CA(G_i, G_j) = ?$$



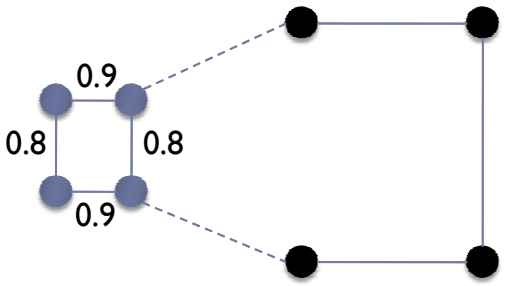
$$CA(G_i, G_j) = 1$$



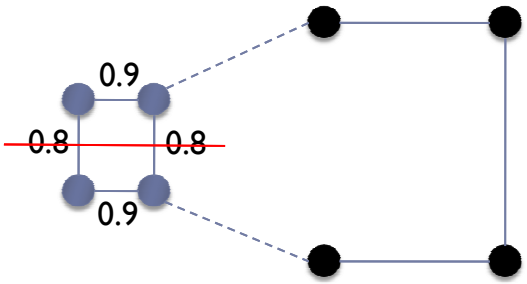
$$CA(G_i) = ?$$



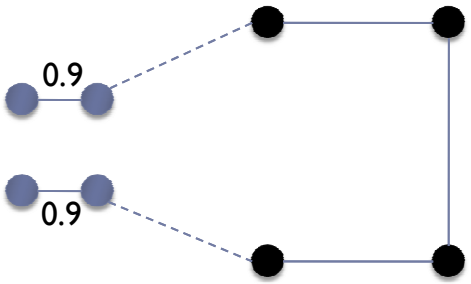
$$CA(G_i) = ?$$



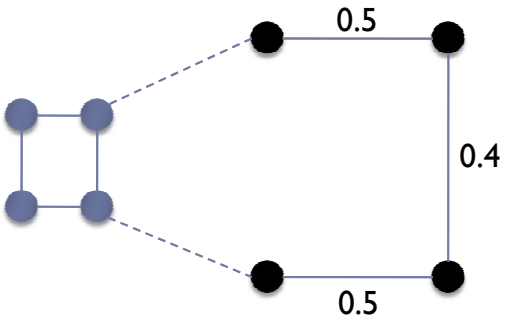
$CA(G_i) = ?$



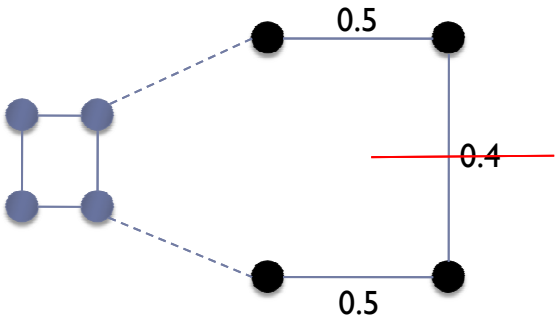
$CA(G_i) = 1.6$



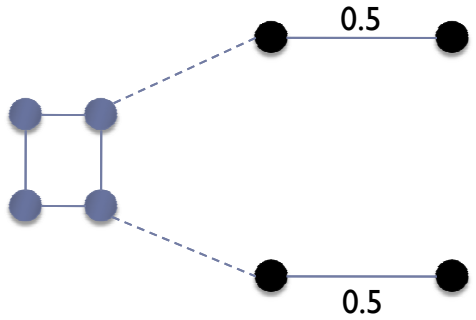
$CA(G_j) = ?$



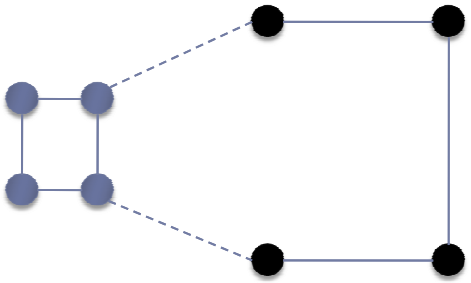
$CA(G_j) = ?$



$$CA(G_j) = 0.4$$



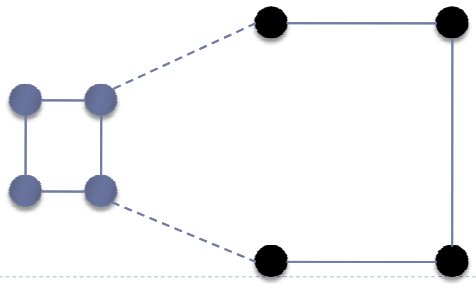
$$CA(G_i, G_j) = 1$$



$$CA(G_i) = 1.6 \qquad CA(G_j) = 0.4$$

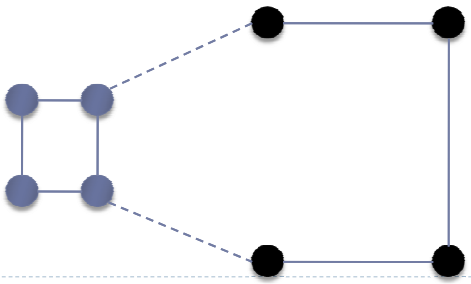
Conectividade Relativa

$$CR(G_i, G_j) = \frac{CA(G_i, G_j)}{\frac{CA(G_i) + CA(G_j)}{2}}$$



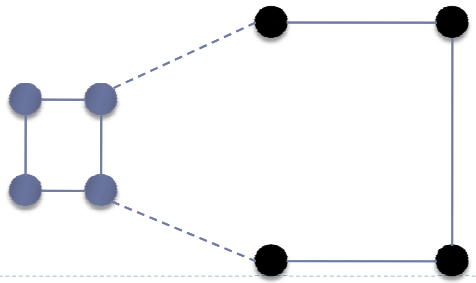
Conectividade Relativa

$$CR(G_i, G_j) = \frac{1}{\frac{1.6 + 0.4}{2}}$$



Conectividade Relativa

$$CR(G_i, G_j) = 1$$



Similaridade entre G_i e G_j

$$S(G_i, G_j) = 1 \cdot PR(G_i, G_j)$$

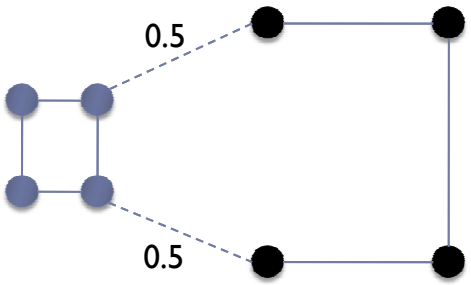
Proximidade Relativa

- Proximidade relativa é estabelecida a partir de proximidades absolutas

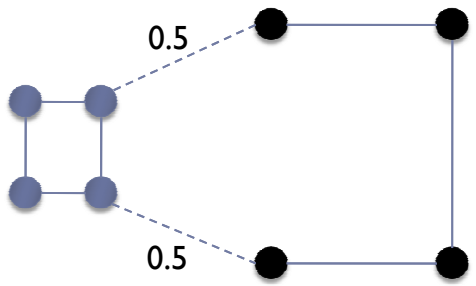
$$PA(G_i, G_j)$$

$$PA(G_i) \quad PA(G_j)$$

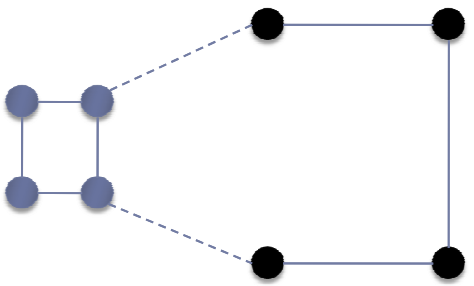
$$PA(G_i, G_j) = ?$$



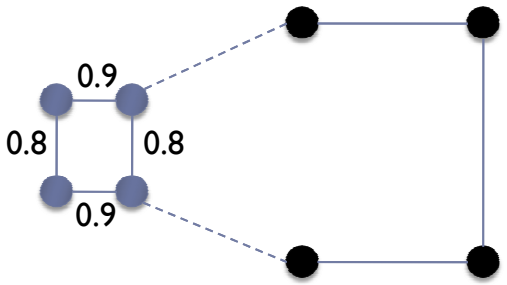
$$PA(G_i, G_j) = 0.5$$



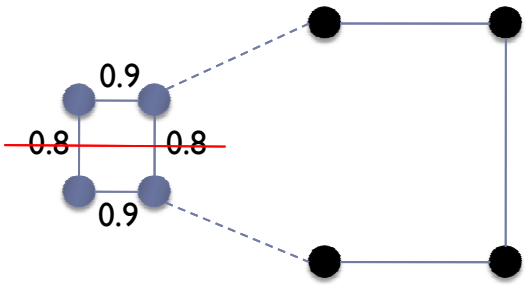
$$PA(G_i) = ?$$



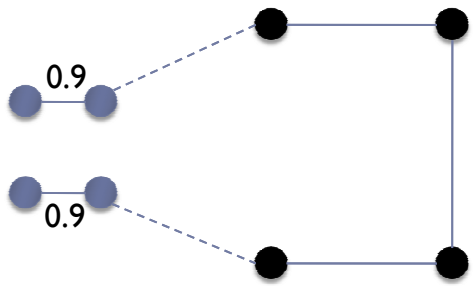
$$PA(G_i) = ?$$



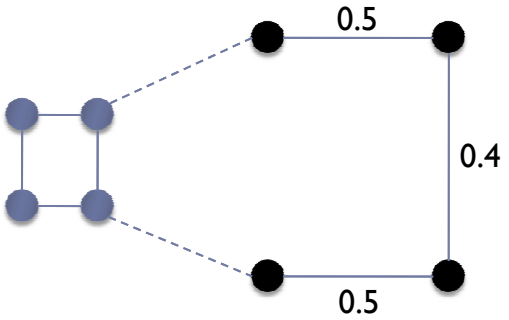
$$PA(G_i) = ?$$



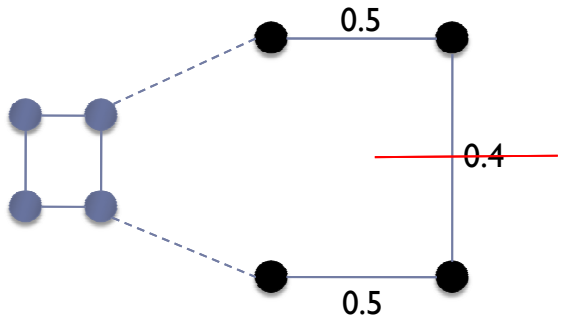
$$PA(G_i) = 0.8$$



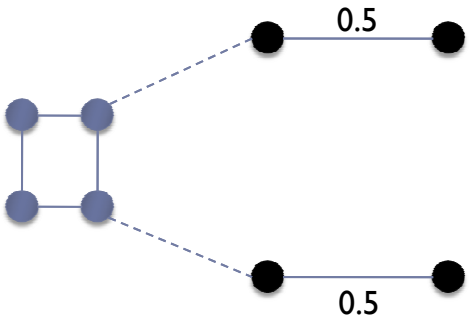
$$PA(G_j) = ?$$



$$PA(G_j) = ?$$

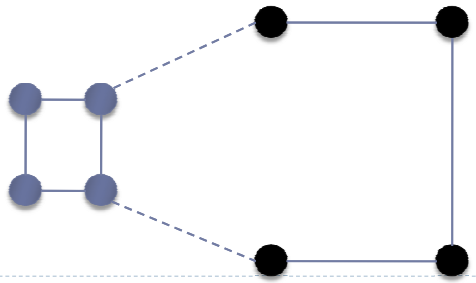


$$PA(G_j) = 0.4$$



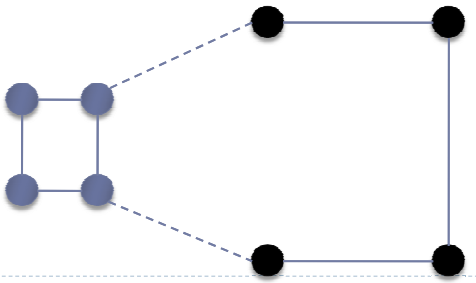
Proximidade Relativa

$$PR(G_i,G_j)=\frac{PA(G_i,G_j)}{\frac{|G_i|PA(G_i)+|G_j|PA(G_j)}{|G_i|+|G_j|}}$$



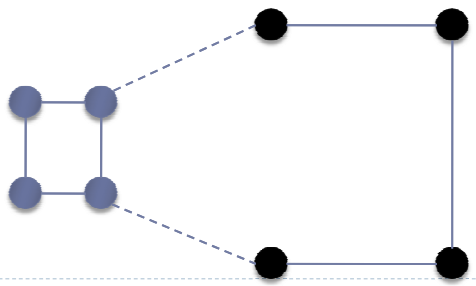
Proximidade Relativa

$$PR(G_i,G_j)=\frac{0.5}{\frac{4\cdot 0.8+4\cdot 0.4}{4+4}}$$



Proximidade Relativa

$$PR(G_i,G_j)=\frac{5}{6}$$



Similaridade entre G_i e G_j

$$S(G_i,G_j)=1\cdot\frac{5}{6}=0.83$$

Algoritmo Chameleon

- **INPUT**: R, **k**, minpts, α
- Construção do grafo dos **k** vizinhos mais próximos
- Particionamento do grafo
- **Faça**
  - Una os dois grupos mais similares
- **Enquanto** puder unir grupos

$$S(G_i, G_j) = CR(G_i, G_j) \cdot PR(G_i, G_j)$$

Conectividade

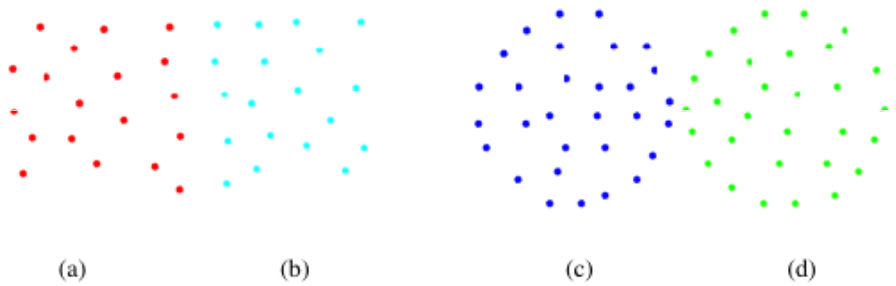


 Figura retirada inescrupulosamente do artigo *CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling*, 1999.

Conectividade Relativa

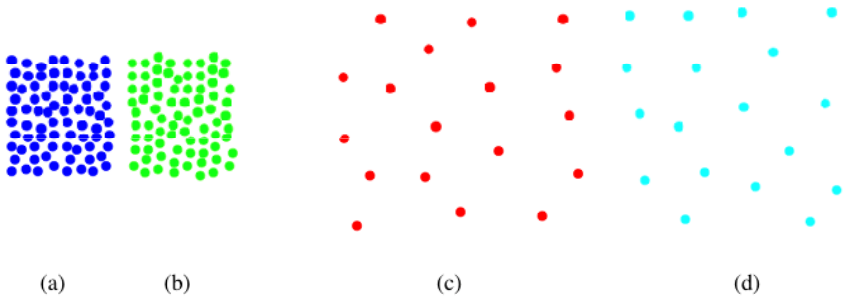


 Figura retirada inescrupulosamente do artigo *CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling*, 1999.

Proximidade (local)

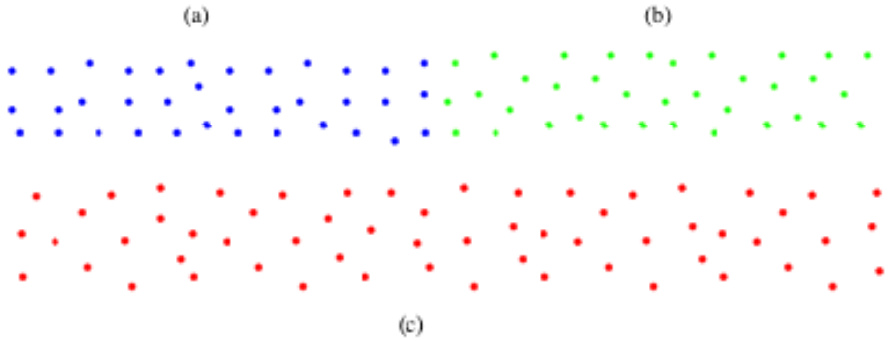


 Figura retirada inescrupulosamente do artigo *CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling*, 1999.

This is the end