

MINERAÇÃO DE DADOS

Algoritmos de agrupamento



TIPOS DE ALGORITMOS DE AGRUPAMENTO

Particionais

Hierárquicos

Baseados em
densidade

Baseados em
grafos

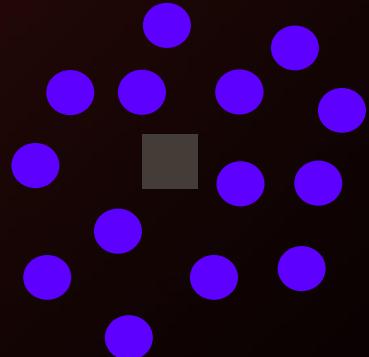
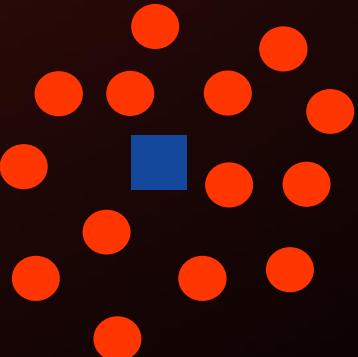
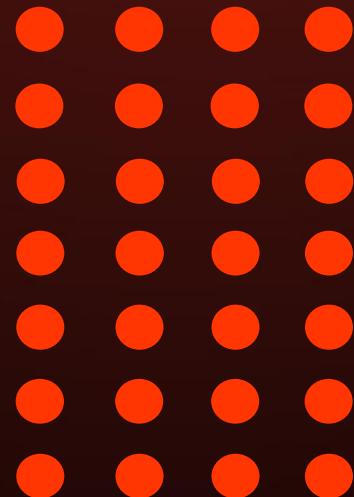
TIPOS DE ALGORITMOS DE AGRUPAMENTO

Particionais

Hierárquicos

Baseados em
densidade

Baseados em
grafos



k-médias

TIPOS DE ALGORITMOS DE AGRUPAMENTO

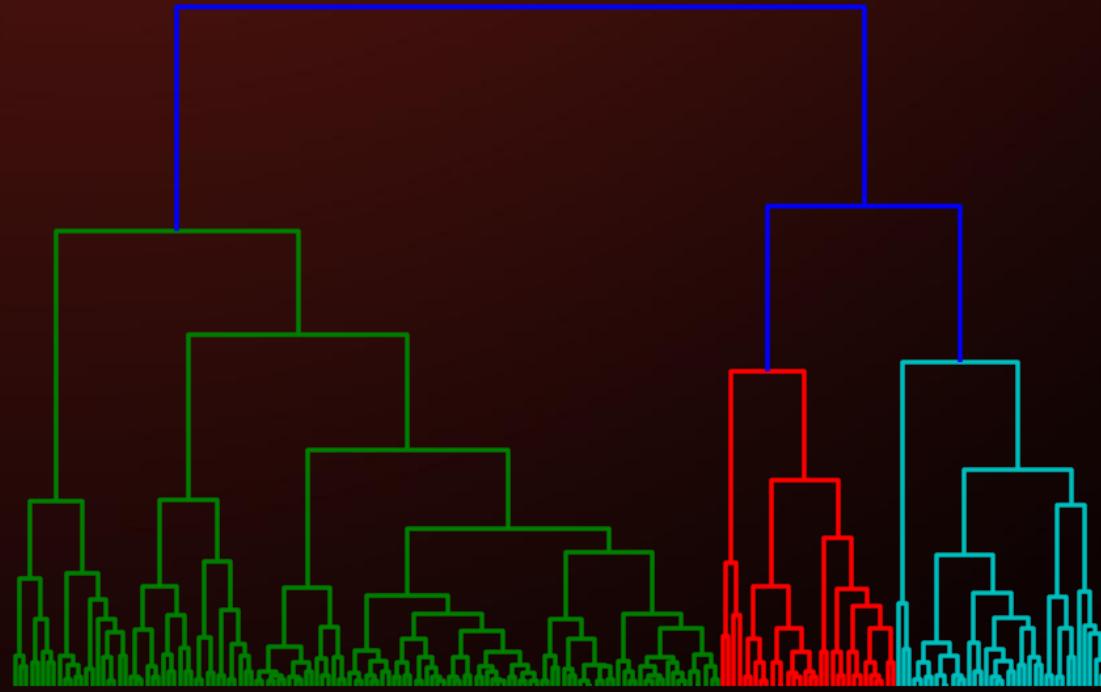
Particionais

Hierárquicos

Baseados em
densidade

Baseados em
grafos

Single-linkage



TIPOS DE ALGORITMOS DE AGRUPAMENTO

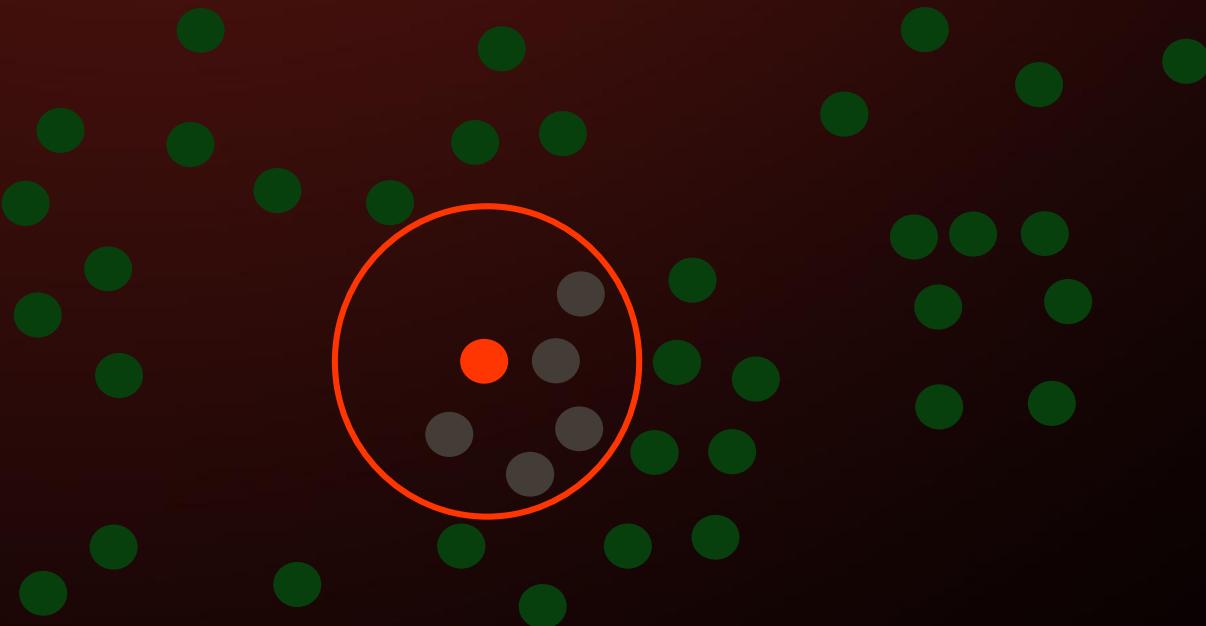
Particionais

Hierárquicos

Baseados em
densidade

Baseados em
grafos

DBSCAN



TIPOS DE ALGORITMOS DE AGRUPAMENTO

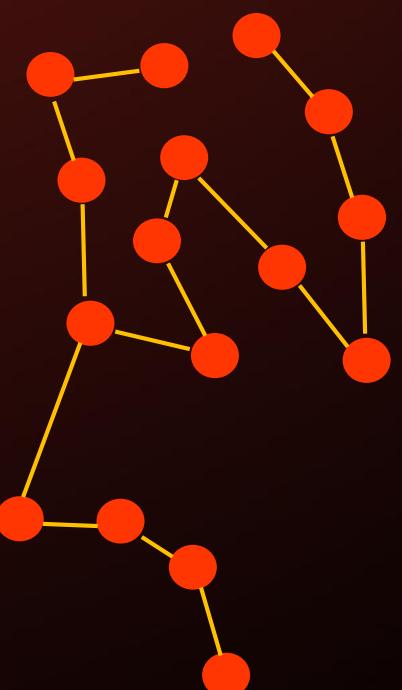
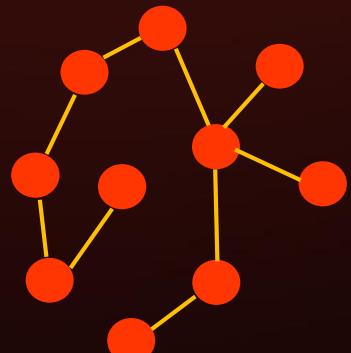
Particionais

Hierárquicos

Baseados em
densidade

Baseados em
grafos

MST (Árvore geradora mínima)



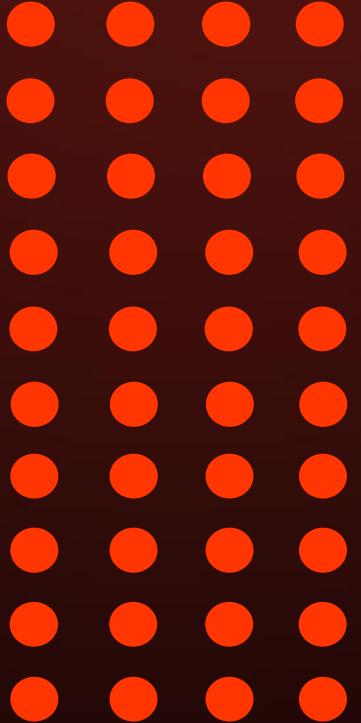
k-MÉDIAS

- Particiona n objetos em k grupos
- Alta similaridade intragrupos
 - centroide = média dos objetos do grupo
- Baixa similaridade intergrupos
- Refinamento iterativo
- Definição de k:
 - Heurística / tentativa e erro

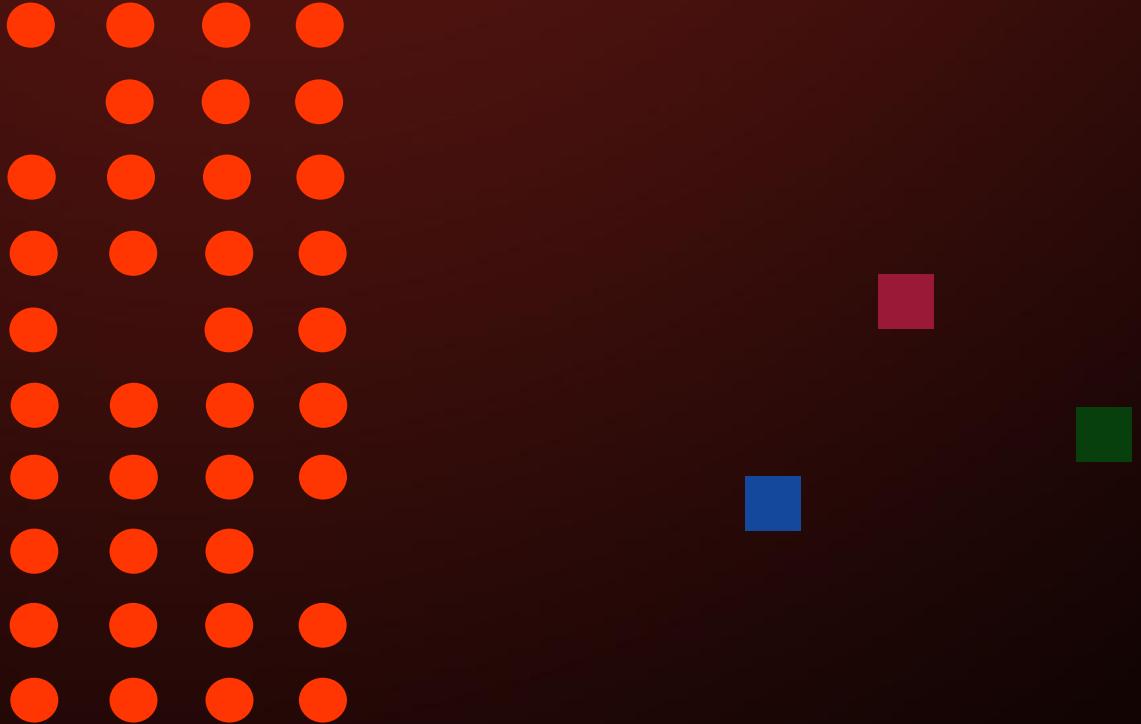
k-MÉDIAS

- Escolhe k objetos aleatoriamente (centroides)
- Calcula a distância entre cada objeto e os centroides
- Calcula o novo centroide
- Converge quando não há mais mudanças
 - Função de custo: erros quadráticos intragrupos

k-MÉDIAS

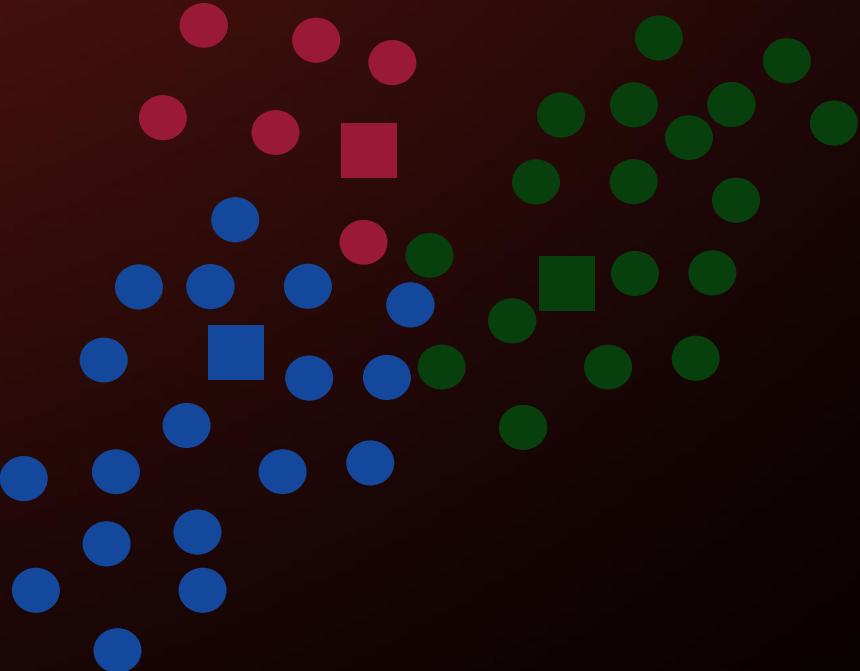


k-MÉDIAS

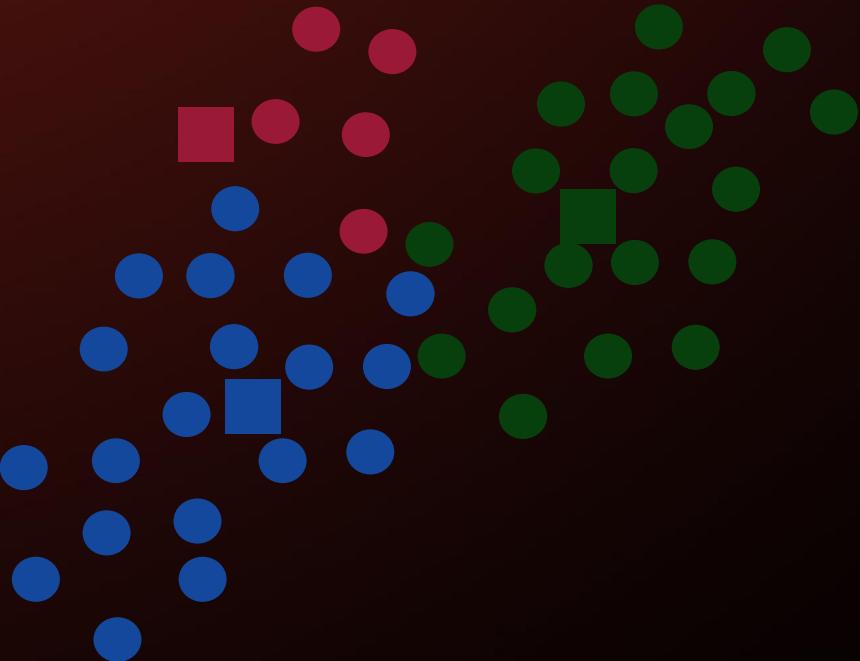


$k = 3$

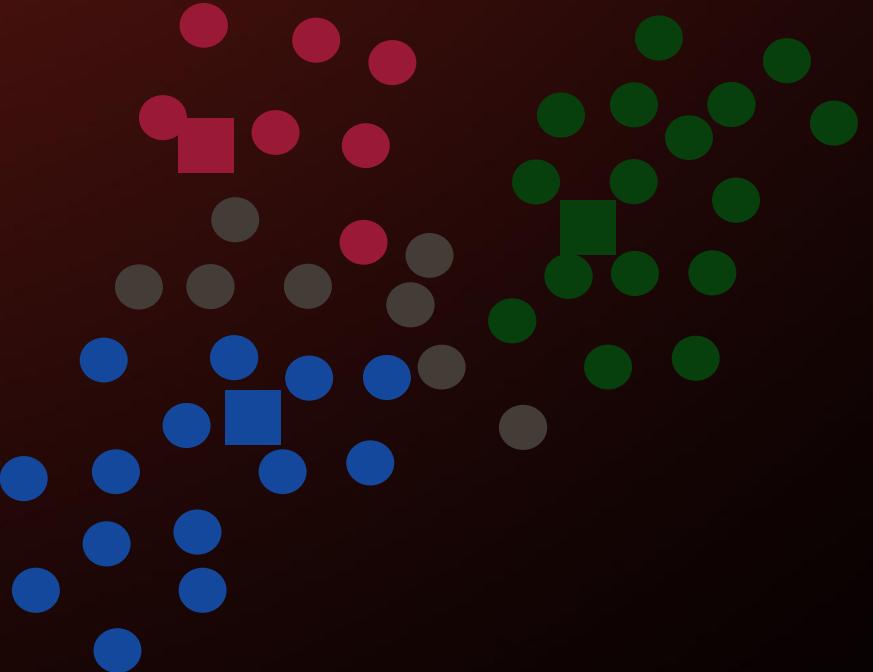
k-MÉDIAS



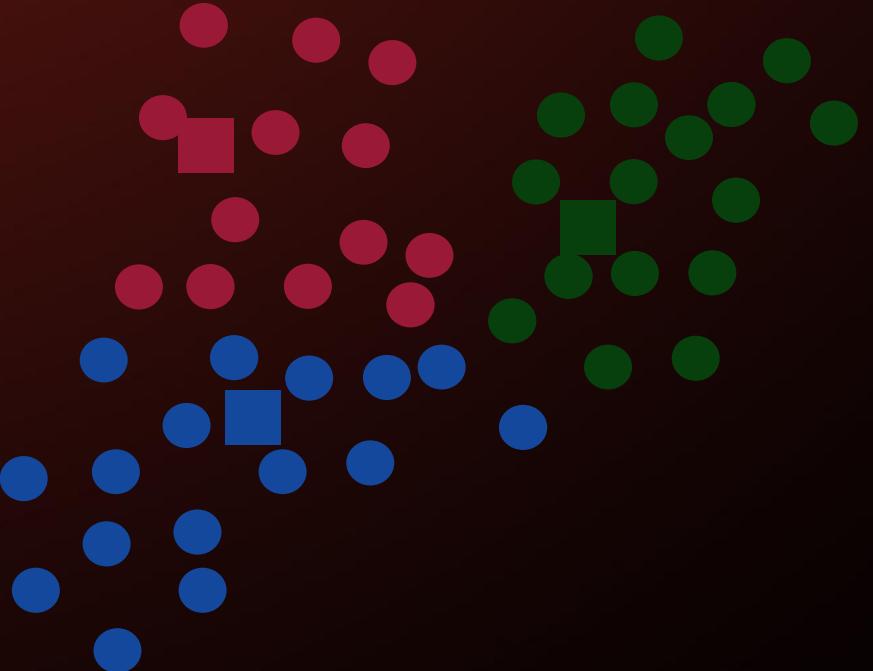
k-MÉDIAS



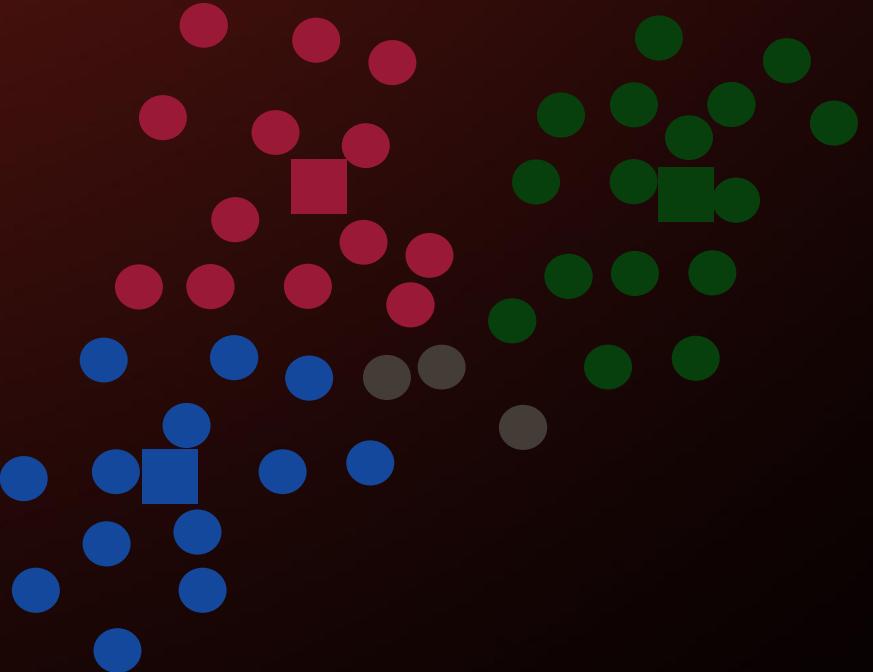
k-MÉDIAS



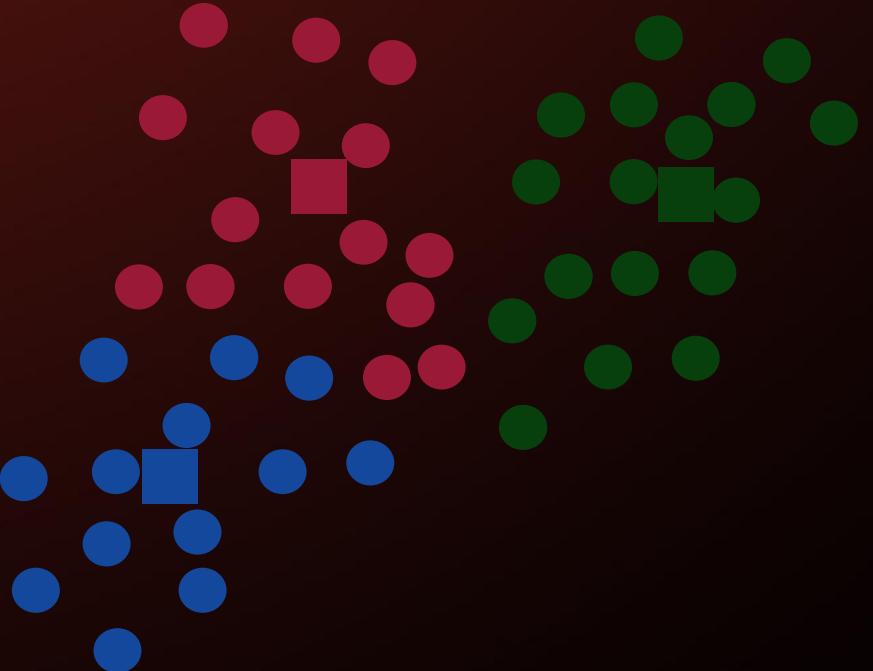
k-MÉDIAS



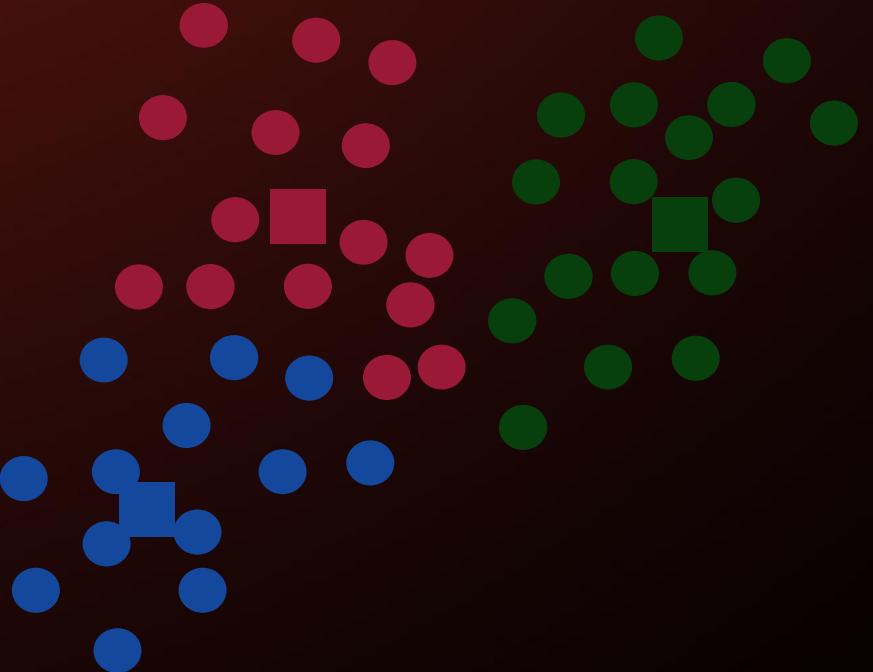
k-MÉDIAS



k-MÉDIAS



k-MÉDIAS



k-MEDOIDES

- Similar ao k-médias
- Medoide
 - Objeto mais ao centro do grupo
 - Protótipo real
- Mais robusto a ruídos

DBSCAN

Density Based Spatial Clustering of Applications with Noise

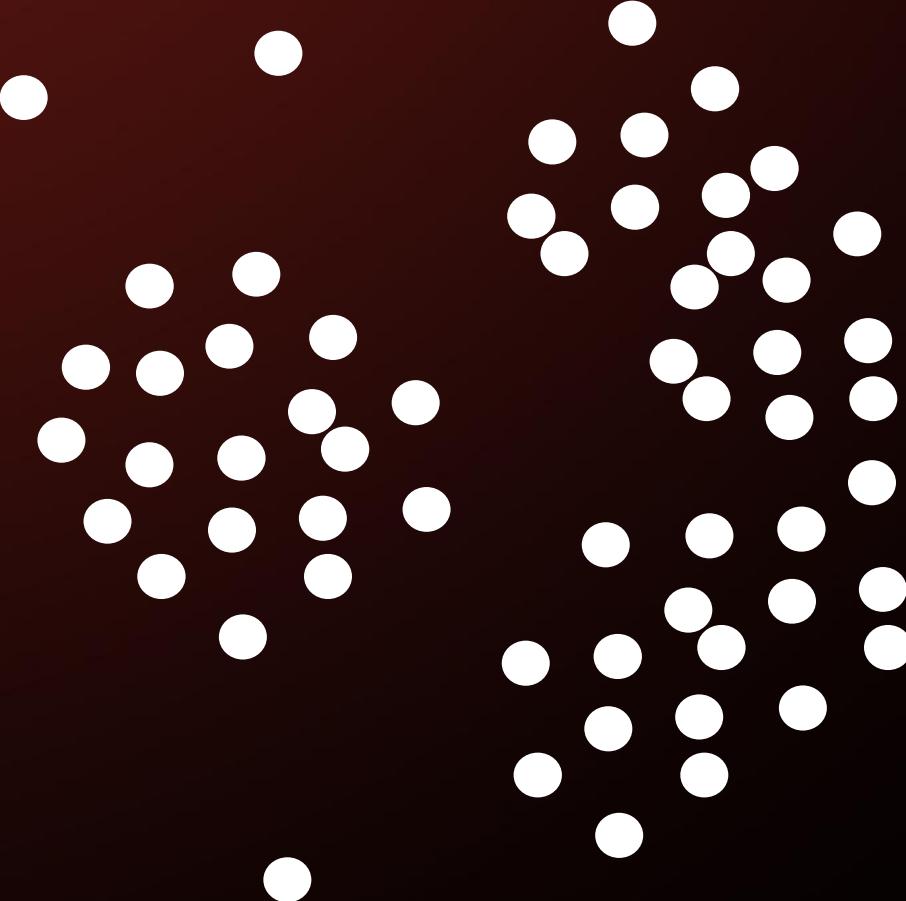
Agrupamento espacial baseado em densidade de aplicações com ruído

- Desenvolvido para agrupamentos em diferentes formatos
 - Uso: diagnóstico de imagens
- Densidade:
 - Objetos em um raio de vizinhança
- Define o número de grupos

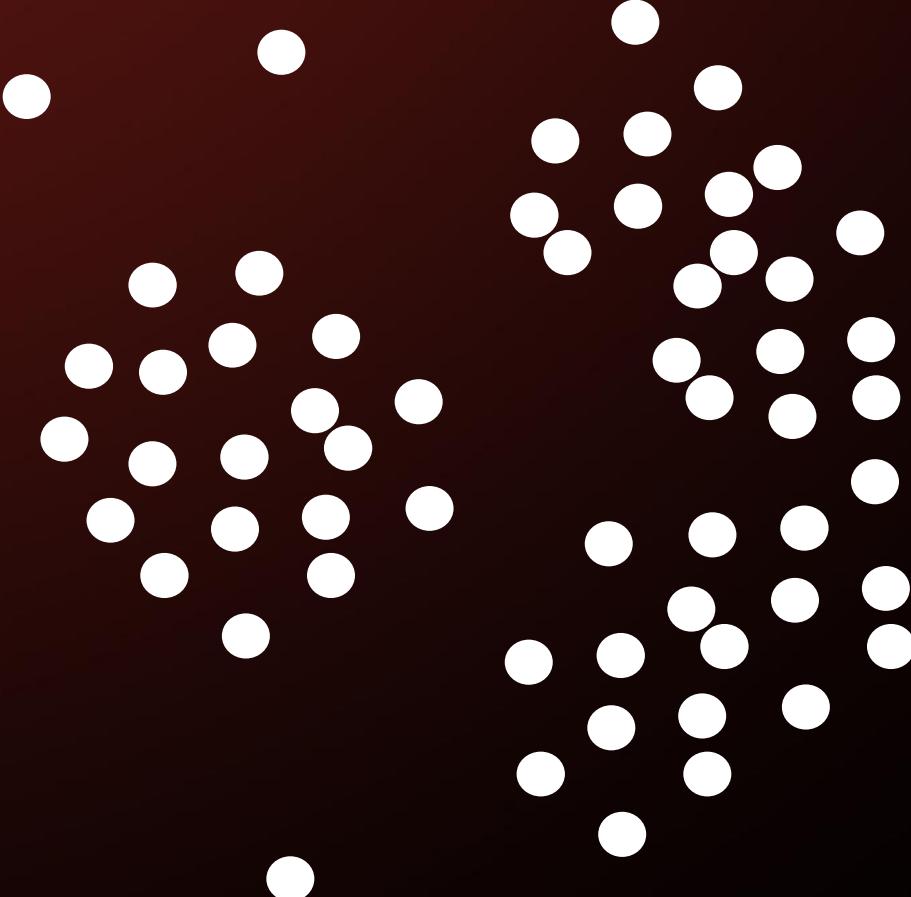
DBSCAN

- Recebe o raio e a quantidade mínima de vizinhos (`minPts`)
- Verifica se um objeto é núcleo de um novo grupo:
 - Verdadeiro se # vizinhos $\geq \text{minPts}$
- Verifica o próximo objeto não visitado
- Repete para todos os objetos não visitados
- Objetos sem vizinhos são ruídos:
 - Regiões de baixa densidade

DBSCAN



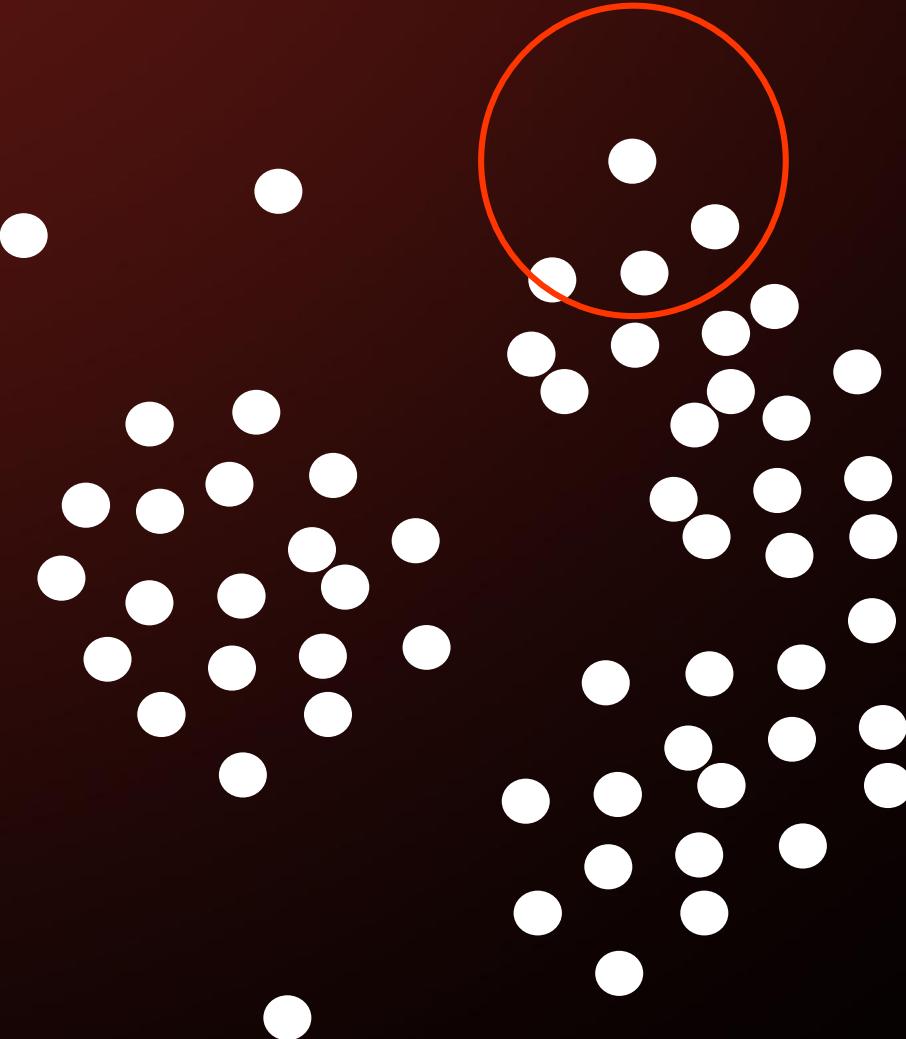
DBSCAN



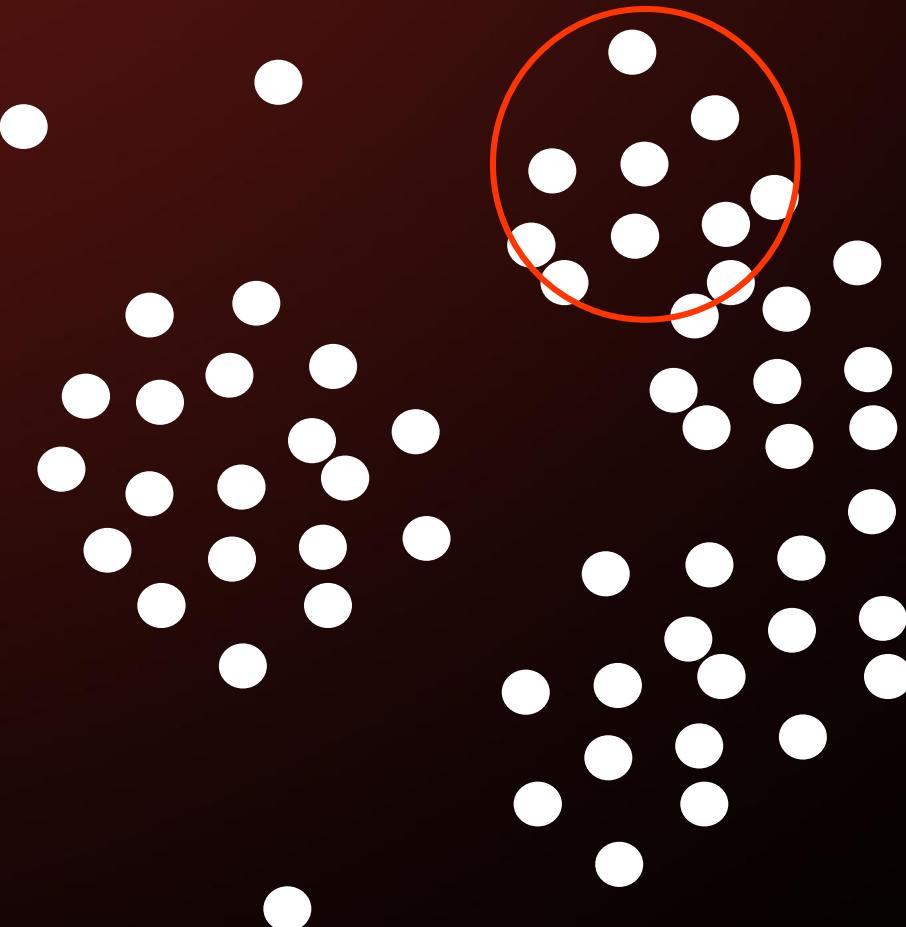
raio = 1

minPts = 5

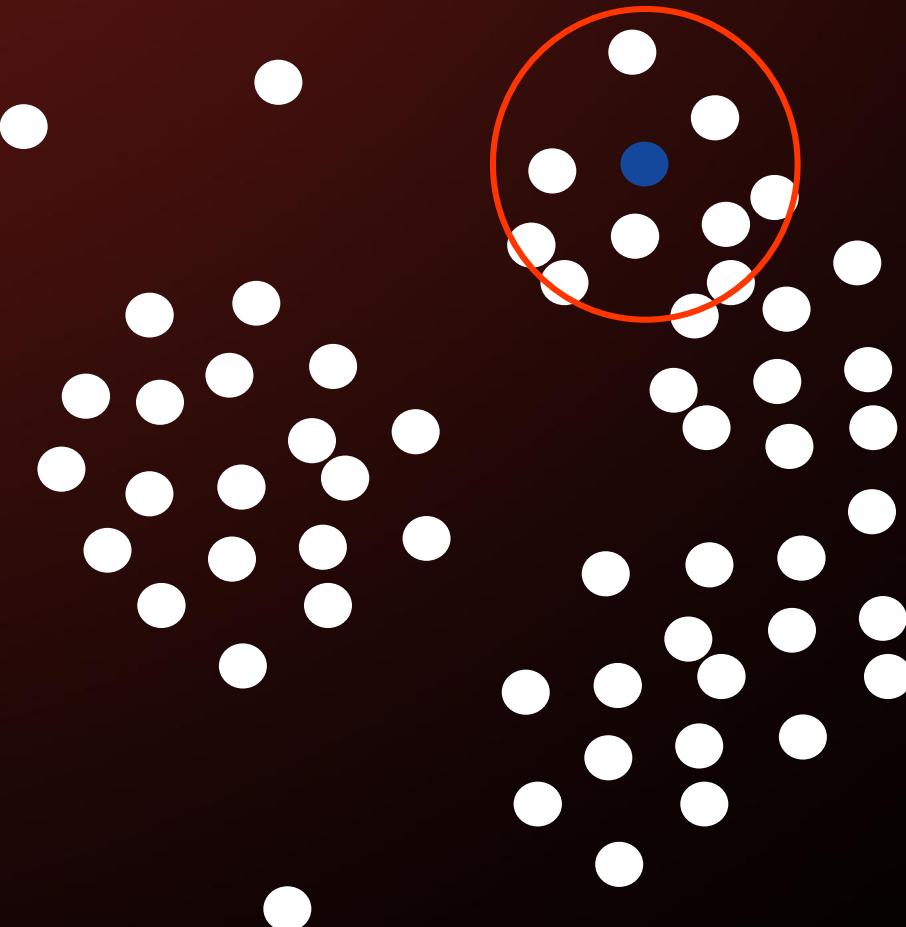
DBSCAN



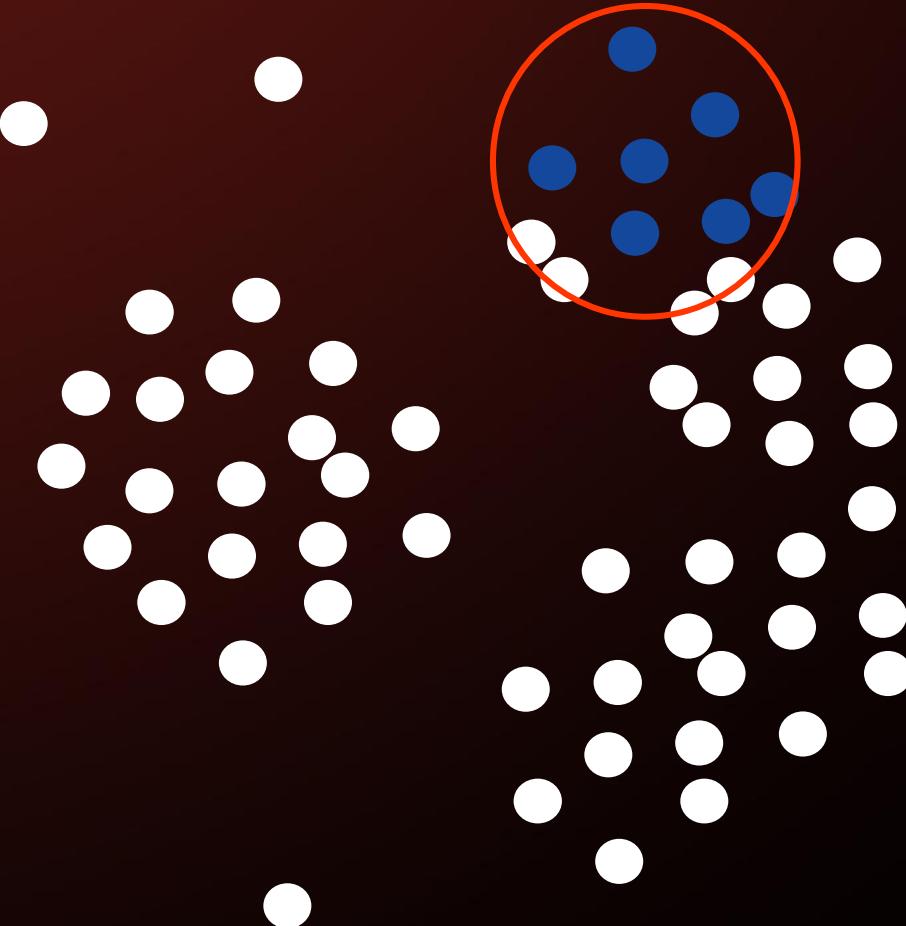
DBSCAN



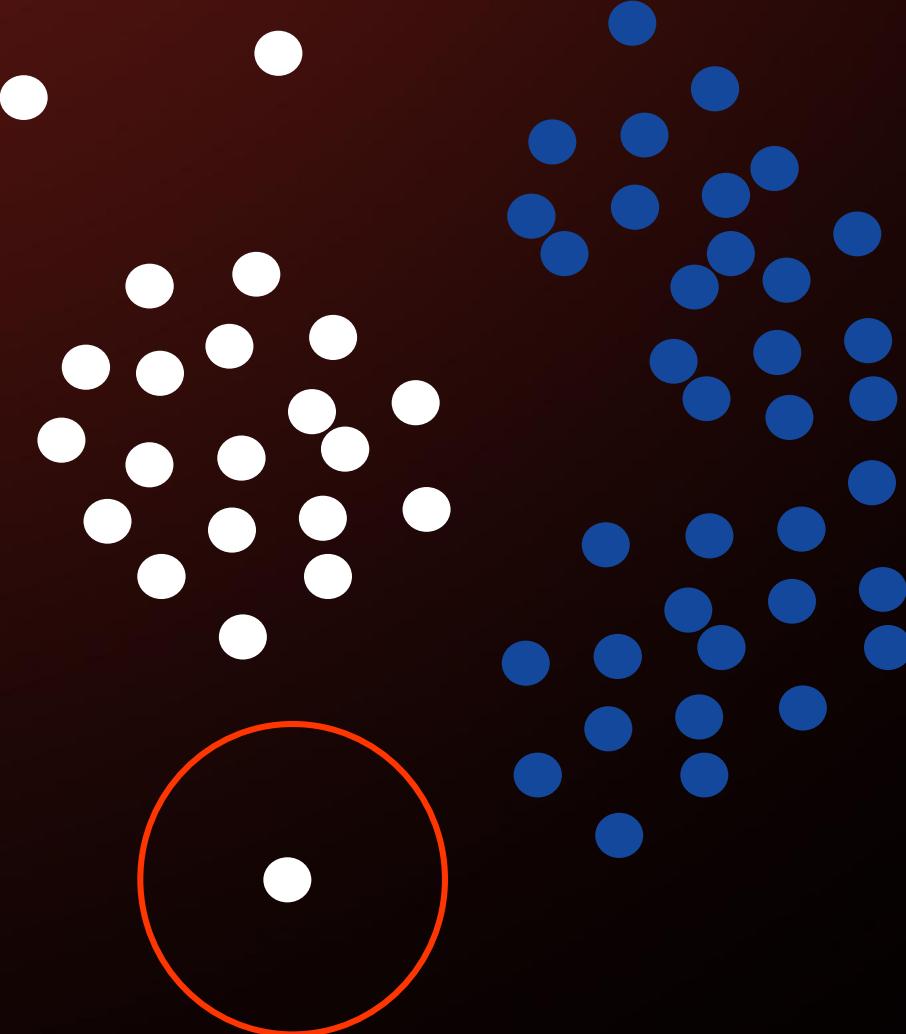
DBSCAN



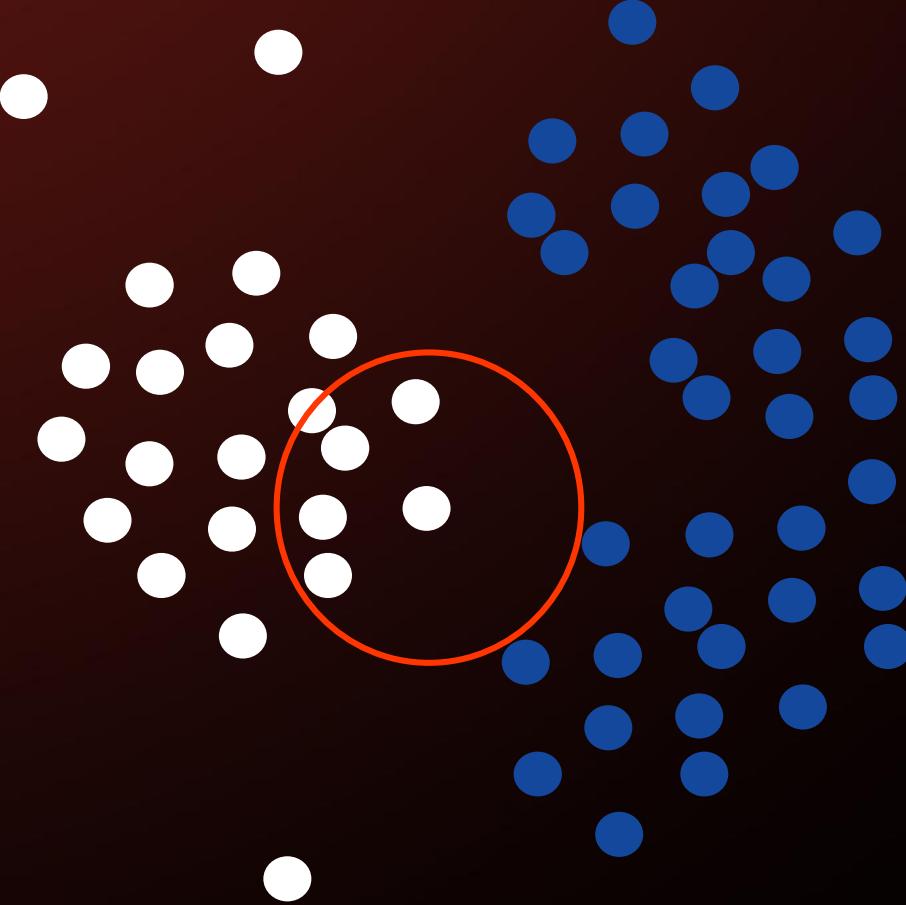
DBSCAN



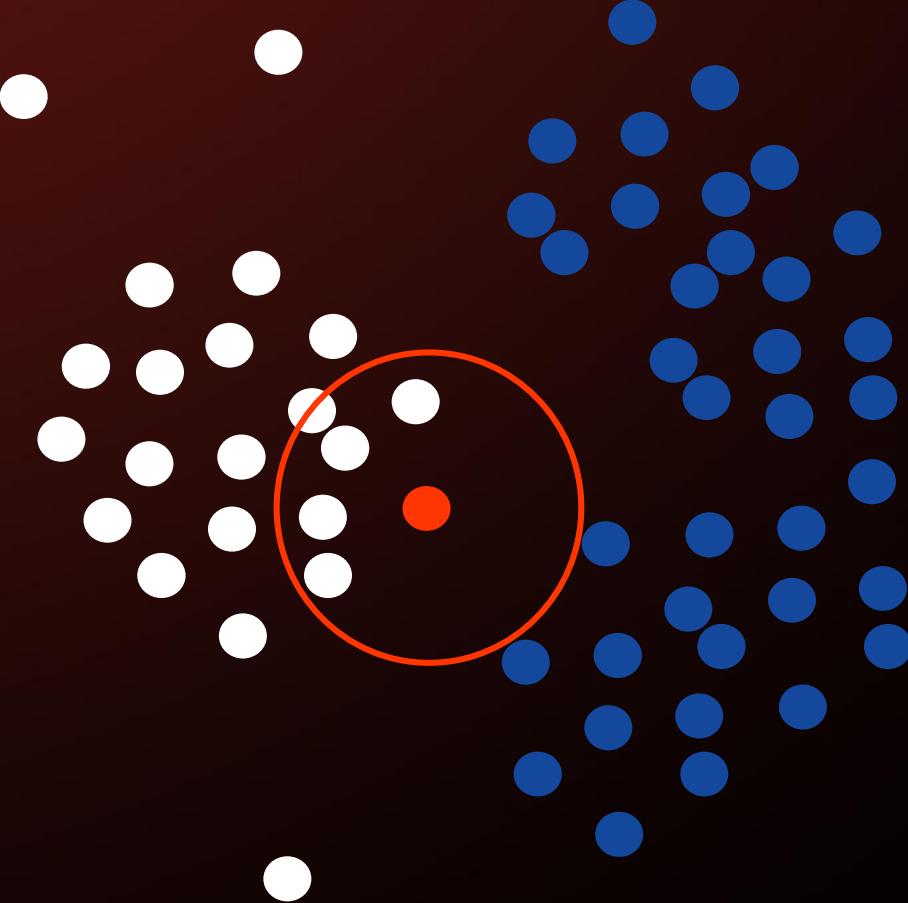
DBSCAN



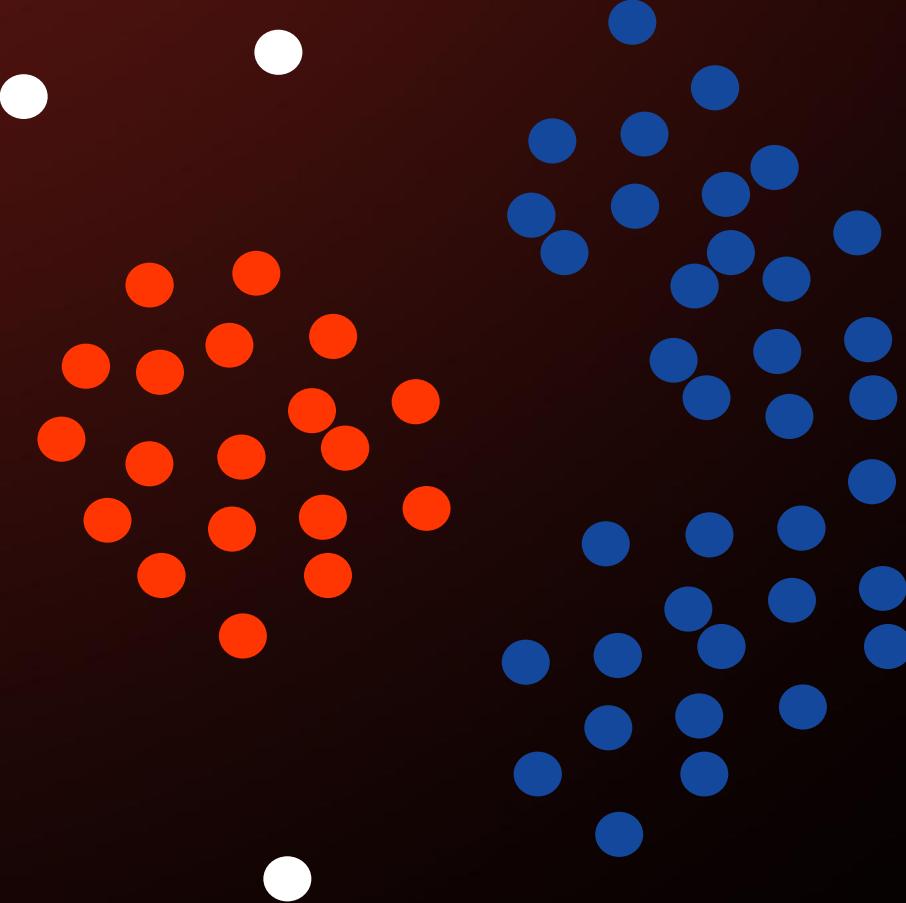
DBSCAN



DBSCAN



DBSCAN



SINGLE-LINKAGE

- Agrupamento inicial é de **singletons**
- Um novo grupo é formado a cada iteração
 - **União dos grupos mais próximos**
- Função de ligação:
 - **Novo grupo recebe a menor distância dos grupos remanescentes**

SINGLE-LINKAGE

- Eliminação de linhas e colunas da matriz de similaridade
- Termina com um único grupo
- Pode ser usado um critério de parada

SINGLE-LINKAGE

	A	B	C	D	E
A	0	3	7	1	9
B	3	0	5	6	2
C	7	5	0	4	5
D	1	6	4	0	8
E	9	2	5	8	0

SINGLE-LINKAGE

	A	B	C	D	E
A	0	3	7	1	9
B	3	0	5	6	2
C	7	5	0	4	5
D	1	6	4	0	8
E	9	2	5	8	0

SINGLE-LINKAGE

	A	B	C	D	E
A	0	3	7	1	9
B	3	0	5	6	2
C	7	5	0	4	5
D	1	6	4	0	8
E	9	2	5	8	0

SINGLE-LINKAGE



SINGLE-LINKAGE

	B	C	E
A	3	7	9
D	6	4	8

SINGLE-LINKAGE

	B	C	E	A-D
B	0	5	2	3
C	5	0	5	4
E	2	5	0	8
A-D	3	4	8	0

SINGLE-LINKAGE

	B	C	E	A-D
B	0	5	2	3
C	5	0	5	4
E	2	5	0	8
A-D	3	4	8	0

SINGLE-LINKAGE

	B	C	E	A-D
B	0	5	2	3
C	5	0	5	4
E	2	5	0	8
A-D	3	4	8	0

SINGLE-LINKAGE



SINGLE-LINKAGE

	C	A-D
B	5	3
E	5	8

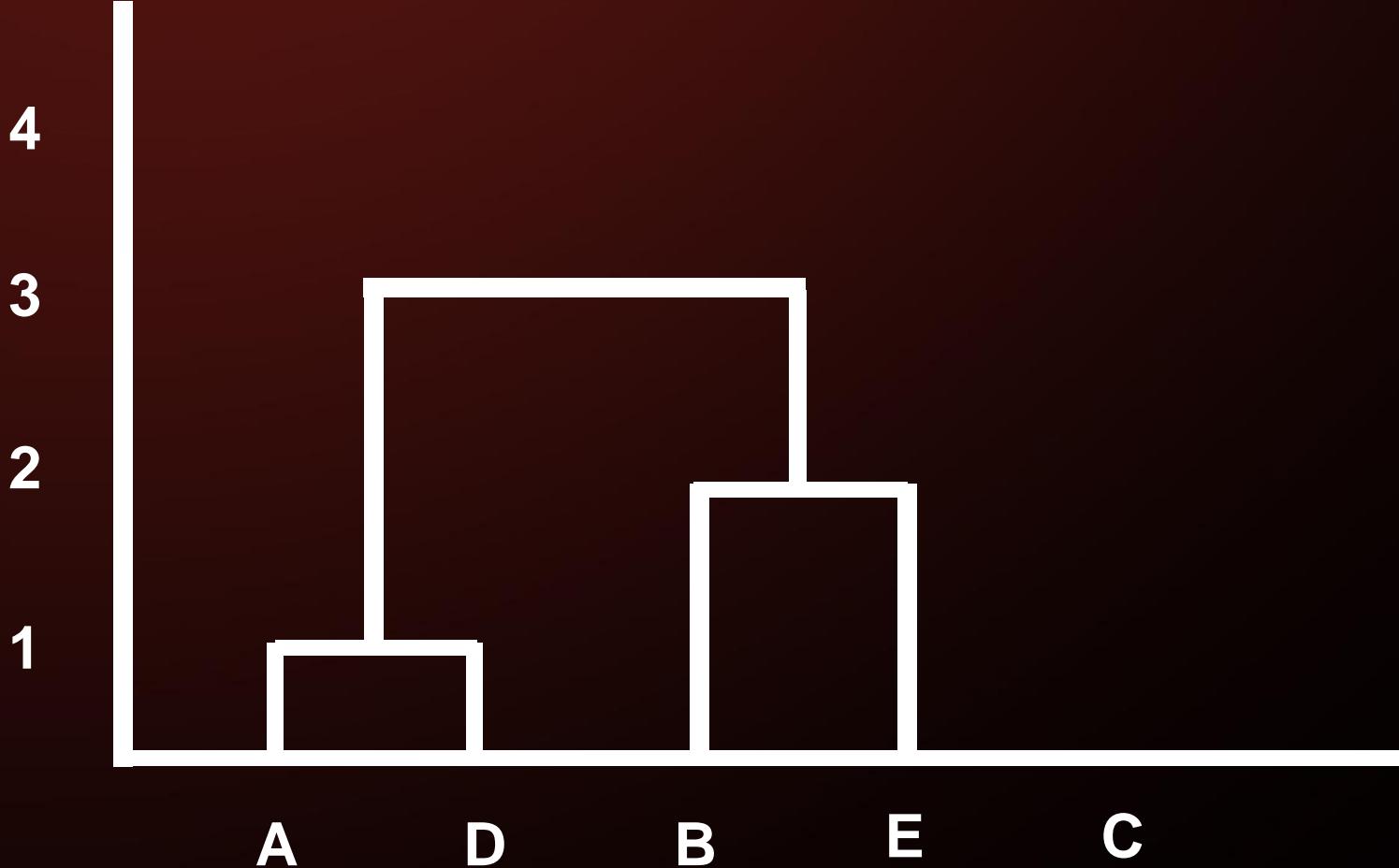
SINGLE-LINKAGE

	C	A-D	B-E
C	0	4	5
A-D	4	0	3
B-E	5	3	0

SINGLE-LINKAGE

	C	A-D	B-E
C	0	4	5
A-D	4	0	3
B-E	5	3	0

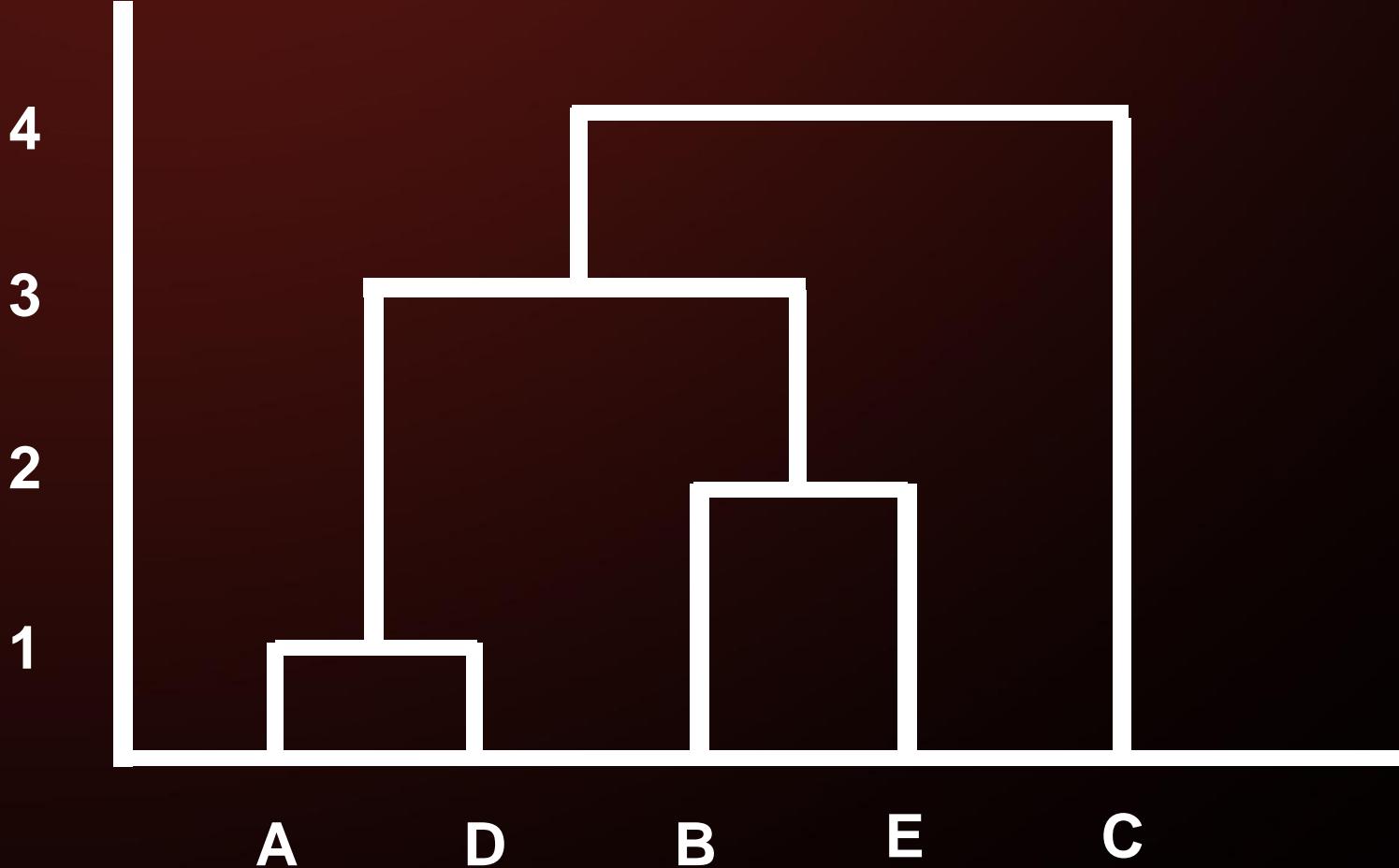
SINGLE-LINKAGE



SINGLE-LINKAGE

	C
A-D	4
B-E	5

SINGLE-LINKAGE



COMPLETE-LINKAGE

- Similar ao single-linkage
- Função de ligação:
 - Novo grupo recebe a maior distância dos grupos remanescentes
- Um pouco mais lento
- Melhor na presença de ruídos

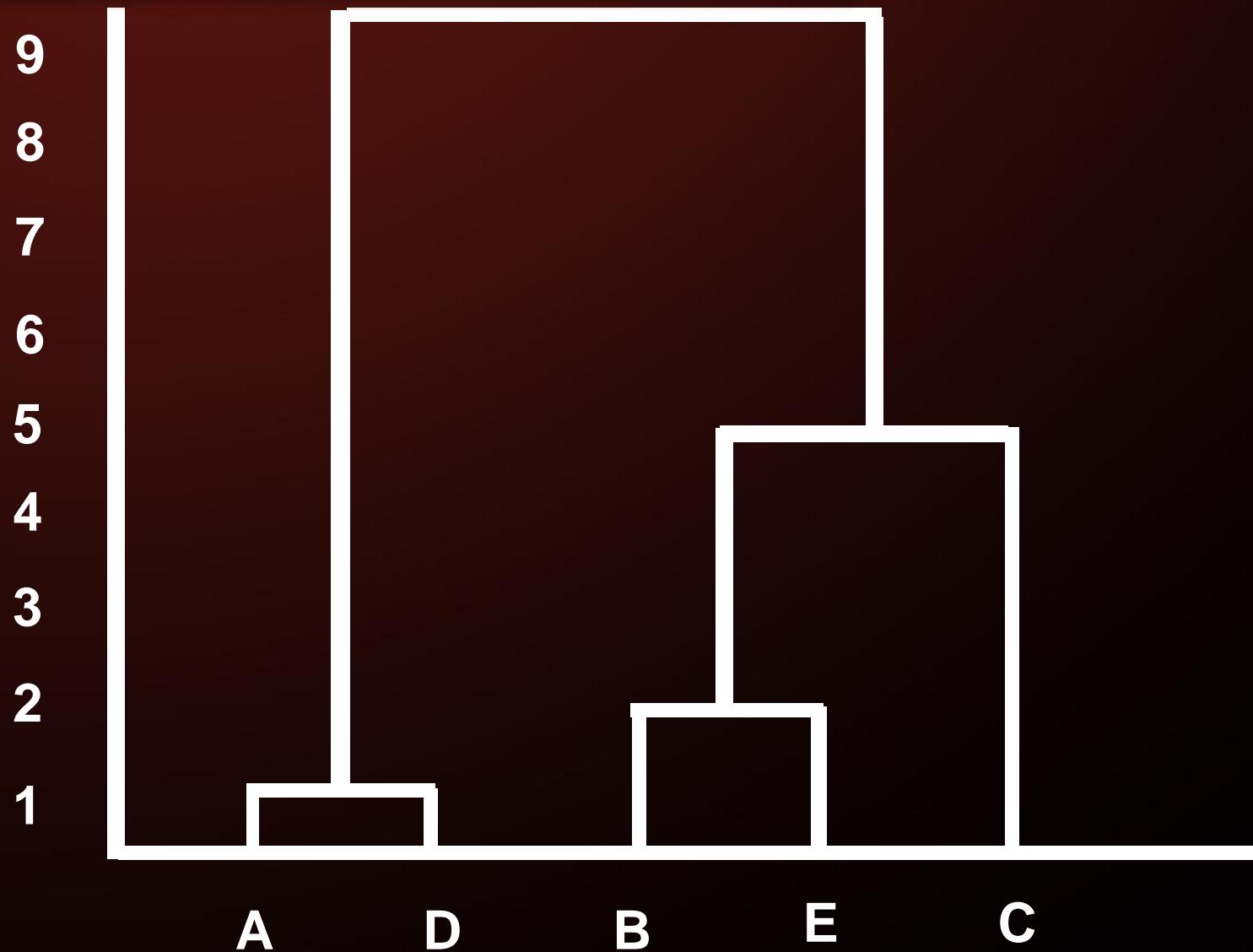
COMPLETE-LINKAGE

	A	B	C	D	E
A	0	3	7	1	9
B	3	0	5	6	2
C	7	5	0	4	5
D	1	6	4	0	8
E	9	2	5	8	0

COMPLETE-LINKAGE

	B	C	E
A	3	7	9
D	6	4	8

COMPLETE-LINKAGE



REFERÊNCIAS

Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações: Cap. 4: Análise de grupos.

Leandro Nunes de Castro e Daniel Gomes Ferrari. Editora Saraiva, 2016.

MINERAÇÃO DE DADOS

Algoritmos de agrupamento

