

Agrupamento(*Clustering*)

Glaucus Augustus
Renan de Melo Oliveira

O que é?

- Agrupamento é uma classificação, não apenas um algoritmo.
- Técnica para formar grupos considerando a similaridade entre elementos
- Encontra padrões “não esperados”
- Aprendizado não supervisionado

Aplicações

Inteligência Coletiva

- Aplicação:

Textos similares; Videos similares; Etc.

- Diferencial:

Dá preferência ao "perfil" do item. Itens com este perfil ficam mais próximos na estrutura, enquanto itens fora do perfil ficam mais distantes;

Tipos:

- Hierárquico
- K - medias

Aprendizado supervisionado X não supervisionado

- Supervisionado:
 - Treinamento
 - Item X Resposta correta
- Não Supervisionado
 - Não existe treinamento
 - Não procura respostas corretas, procura relação entre os itens

Agrupamento baseado em semelhança

- Agrupa itens "semelhantes":

O que é "semelhança" neste caso?

- Propriedades do item normalmente expressas da forma de um vetor numérico
- Necessita de um algoritmo entre semelhança de vetores.

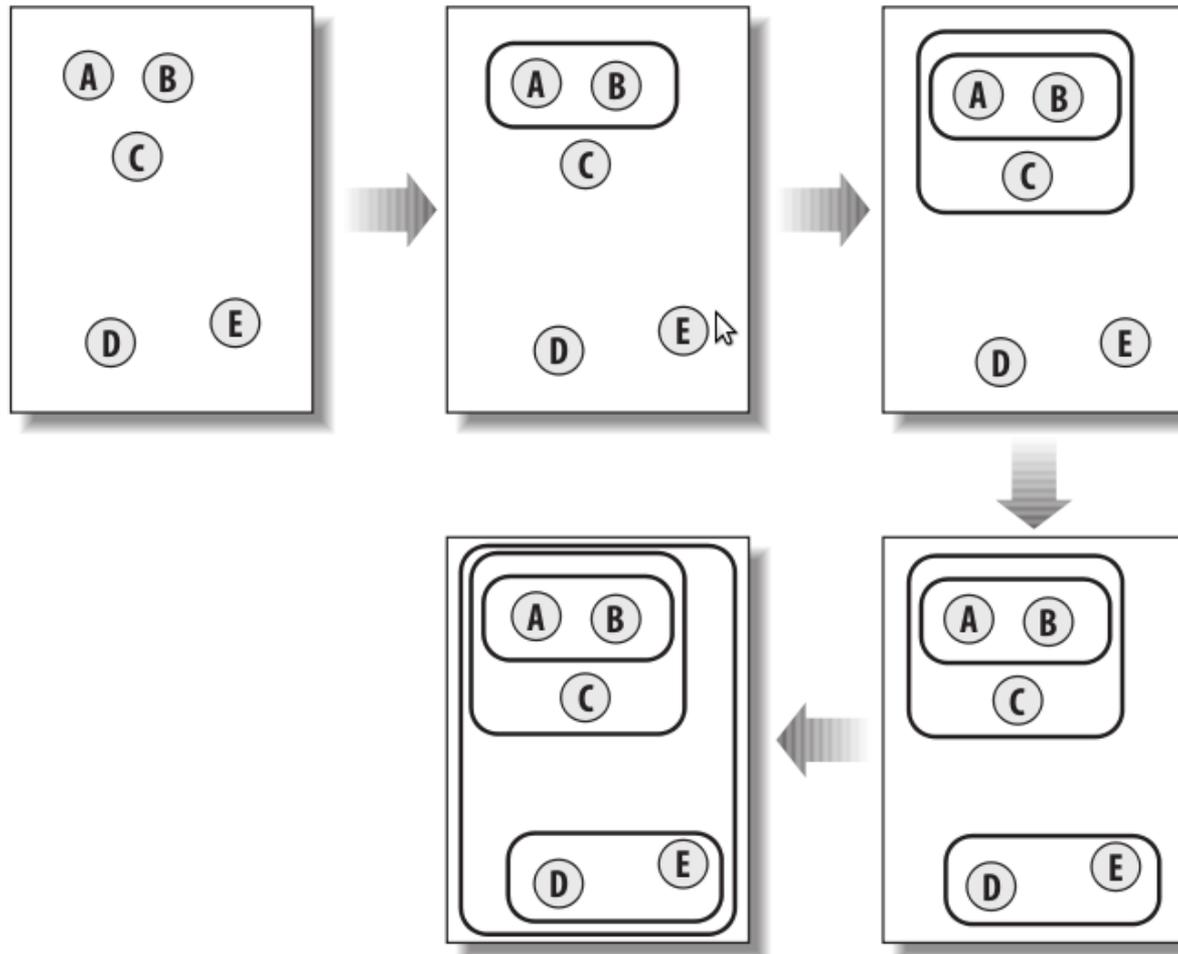
Ex:

- Coeficiente Pearson
- Distância euclidiana
- Coeficiente Tanimoto

Hierárquico

- Objetivo: Ao final do algoritmo poderemos reconhecer:
 - Dado um item: Qual item é o seu vizinho mais perto?
ou
 - Qual é o grupo de itens mais perto

Agrupamento Hierárquico



Funcionamento de um Agrup. Hierárquico - [Seragan, 2007]

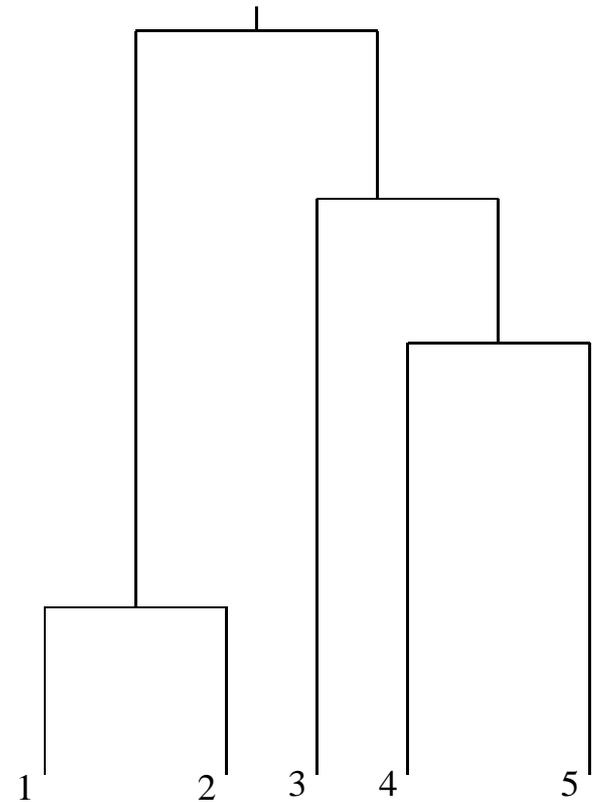
Como funciona?

	G 1	G 2	G 3	G 4	G 5
G 1	0				
G 2	2	0			
G 3	6	5	0		
G 4	10	9	4	0	
G 5	9	8	5	3	0

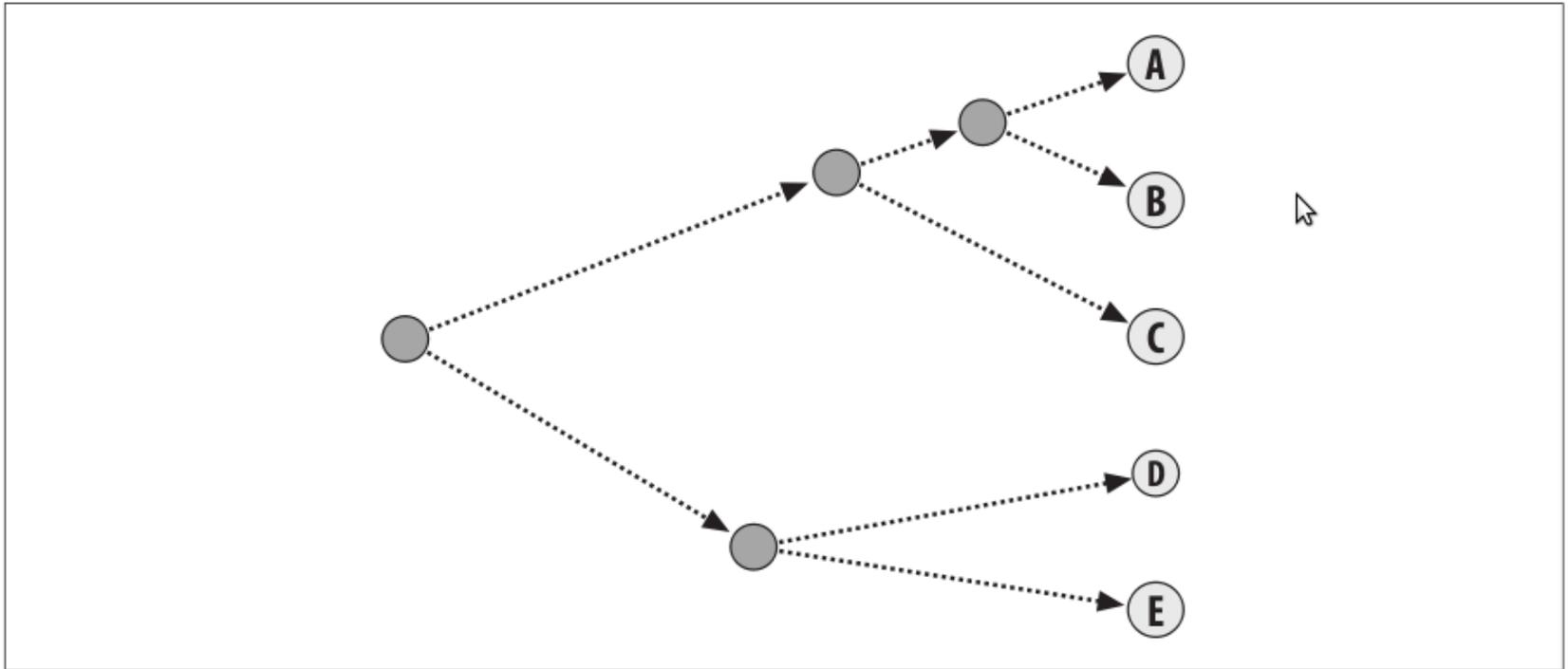
	G (12)	G 3	G 4	G5
G (12)	0			
G 3	6	0		
G 4	10	4	0	
G 5	9	5	3	0

Como funciona?

Stage	Groups
P5	$[1], [2], [3], [4], [5]$
P4	$[1\ 2], [3], [4], [5]$
P3	$[1\ 2], [3], [4\ 5]$
P2	$[1\ 2], [3\ 4\ 5]$
P1	$[1\ 2\ 3\ 4\ 5]$



Agrupamento Hierárquico

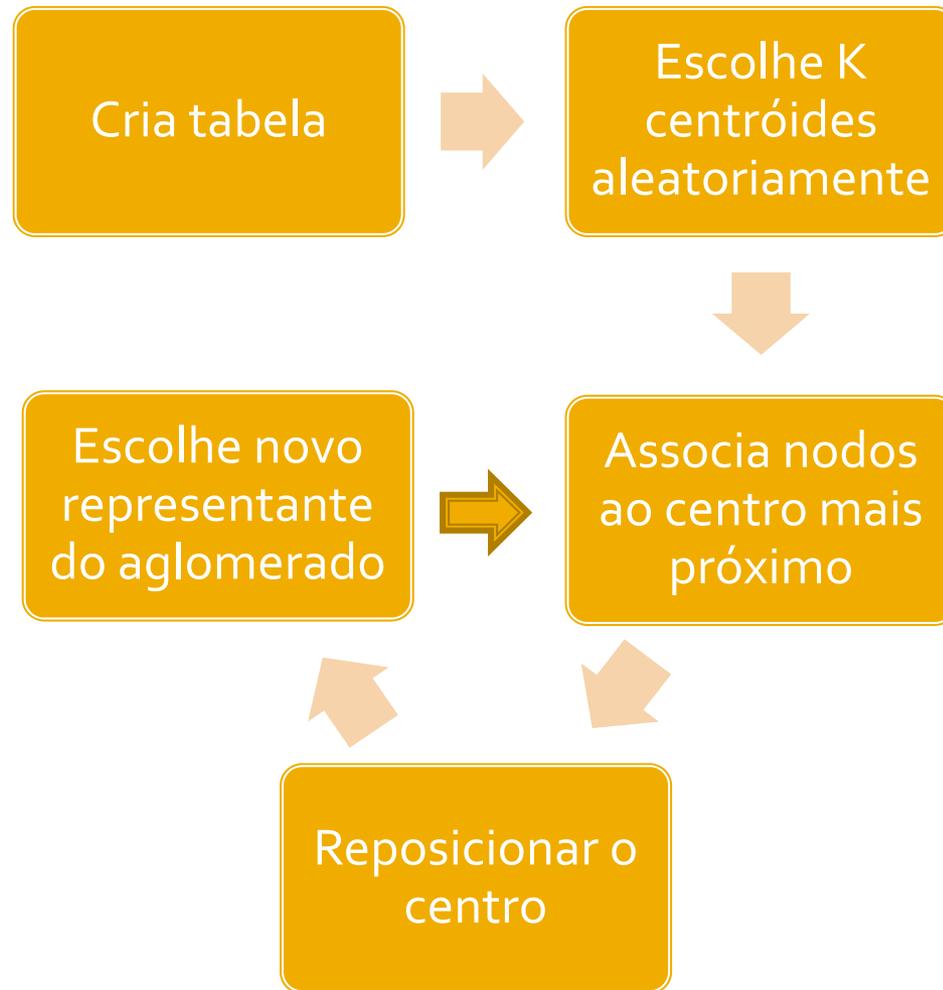


Dendograma - [Seragan, 2007]

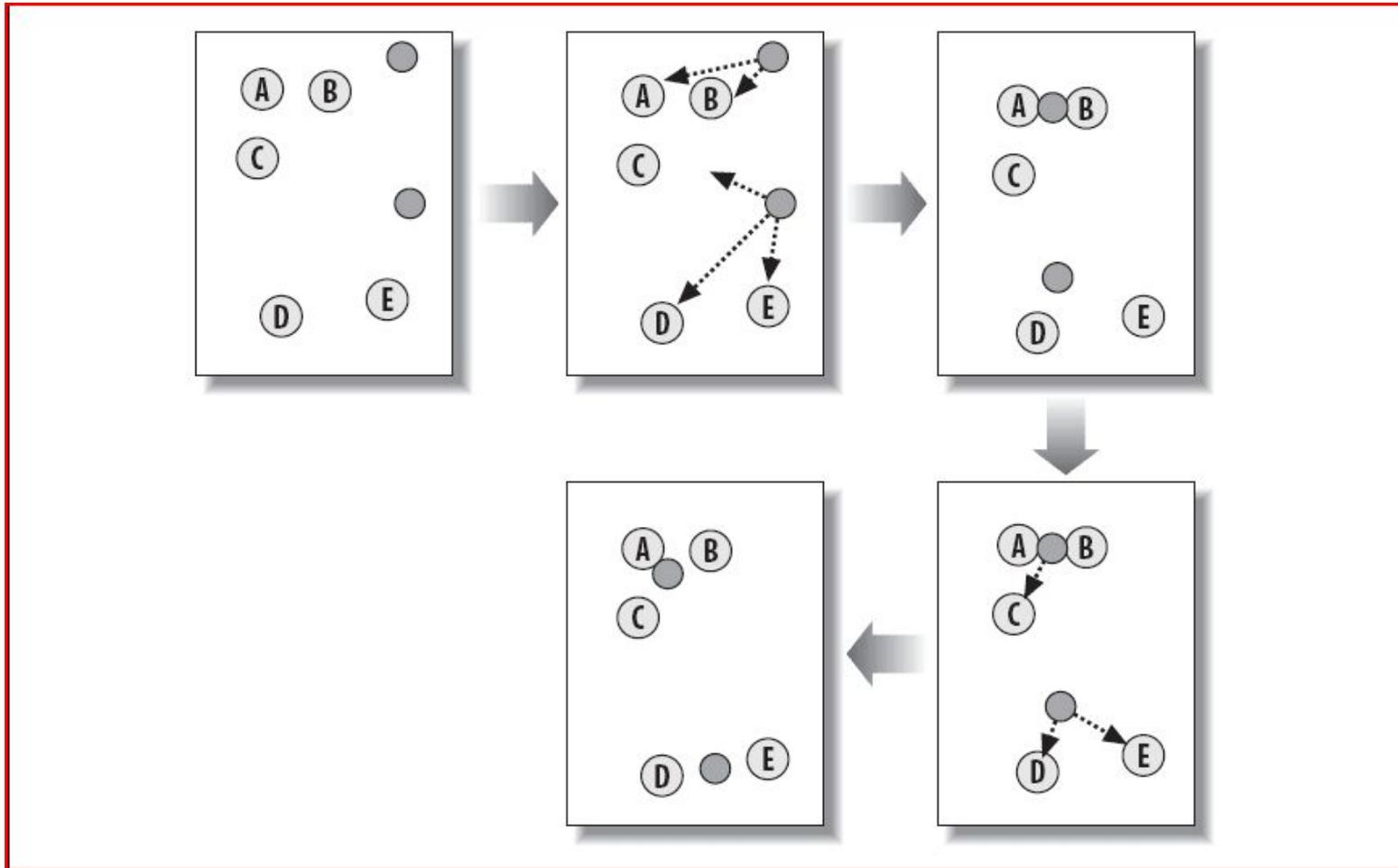
K-medias

- É informada (previamente) a quantidade de grupos a formar
- Baseado nisso tenta formar os grupos de acordo com a semelhança entre os itens/grupos

Como funciona?



Como funciona?



K-Means

- É importante possuir algum tipo de informação com relação ao agrupamento dos itens (Para indicar previamente a quantidade de grupos)

Inteligência Coletiva:

- Conseguir aproximar itens do mesmo perfil;
- Também pode ser utilizado para sugestões.

Agrupamento Hierárquico X K-Means

- Agrup. Hierárquico
 - Não é necessário ter informação prévia sobre os perfis
 - Estrutura melhor o item mais próximo
 - Estrutura os perfis próximos
 - Mais lento

Agrupamento Hierárquico X K-Means

- K-Means:
 - Relaciona melhor os itens do mesmo perfil
 - Necessita de alguma informação sobre quantidade de grupos
 - Não se interessa em similaridade entre grupos
 - Mais rápido

demonstracao

Perguntas?

Bibliografia

SEGARAN, T., Programming Collective Intelligence: Building Smart Web 2.0 Applications, O'Reilly, ISBN 0596529325, 2007.

ALAG, S., Collective Intelligence in Action, ISBN 1933988312, 2009.