CURSOS DE VERANO 2014

APROXIMACIÓN PRÁCTICA A LA CIENCIA DE DATOS Y BIG DATA: HERRAMIENTAS KNIME, R, HADOOP Y MAHOUT

Análisis descriptivo

María José del Jesus

Analítica descriptiva

- **Motivación**
- Agrupamiento
- Reglas de asociación
- Ejemplos

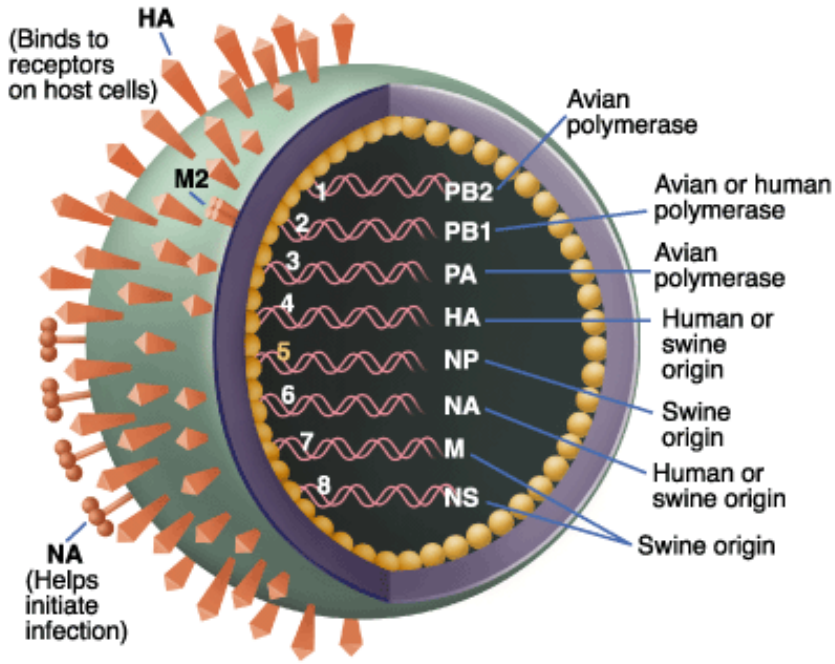
Motivación ▶

Analítica descriptiva

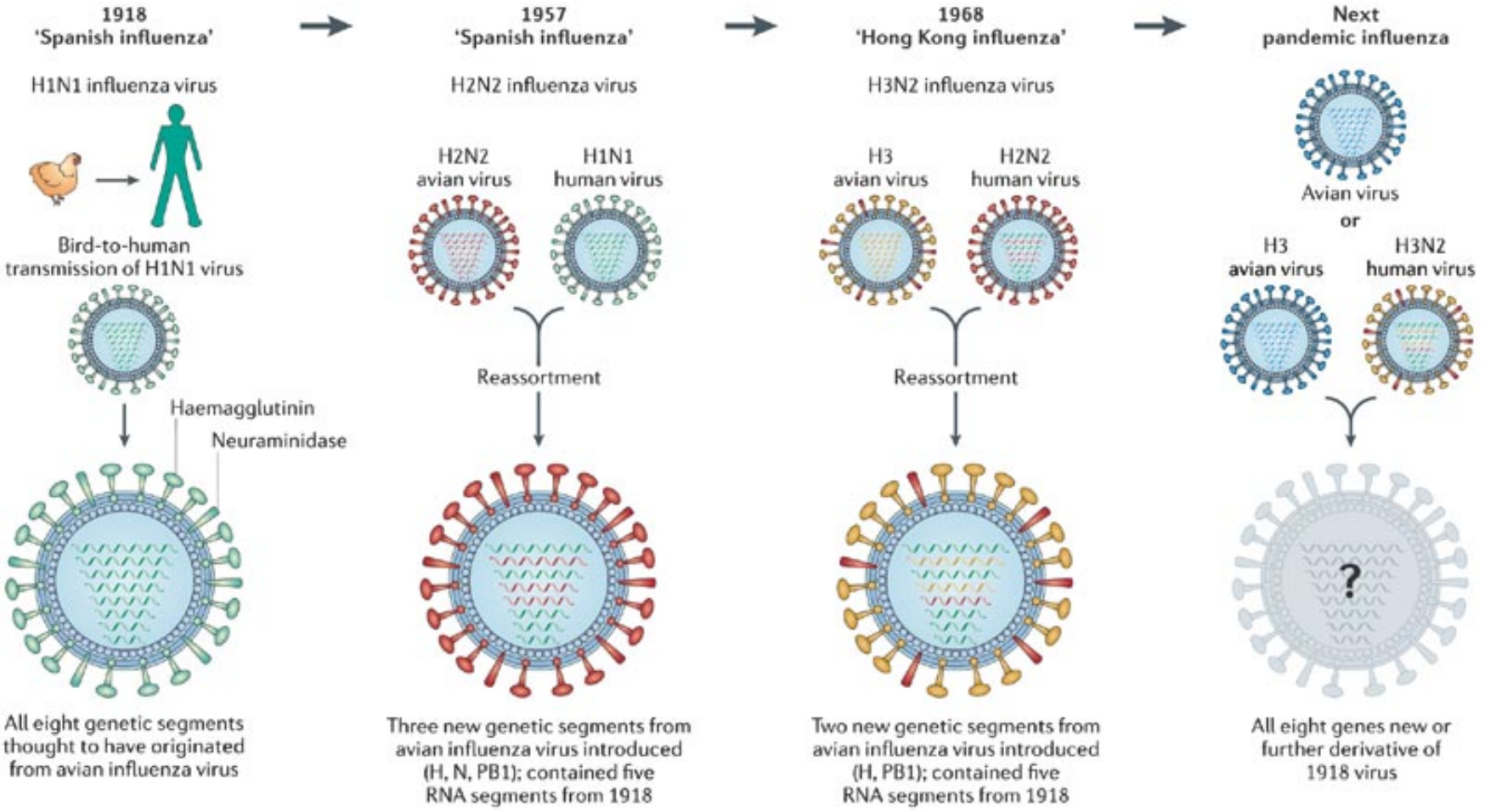
- **Técnicas de MD que permiten extraer conocimiento que nos permita conocer mejor lo que subyace en los datos**
  - Clustering
  - Segmentación
  - Análisis de asociaciones
  - Análisis de desviaciones
- Pueden tener como objetivo
  - buscar patrones (sin centrarse en ninguna variable en particular)
  - encontrar explicaciones (conocimiento para alguna variable específica)

Motivación ▶ Un ejemplo

- El virus de la Gripe A pertenece a la famila *Orthomyxoviridae* y afecta principalmente a aves y algunos mamíferos.
  - Su genoma está formado por 8 genes:
    - HA (hemaglutinina)
    - NA (neuraminidasa)
    - NP (nucleoproteina)
    - Proteínas matrices o estructurales (M)
    - Proteinas no estructurales (NS)
    - Tres genes RNA polimerasa (PA, PB1, PB2)



Motivación ► Un ejemplo



Motivación ► Un ejemplo

- **Objetivo:** Buscar relaciones relevantes entre genes de diferentes subtipos de virus de gripe A (H1N1, H2N2, H3N2 y H5N1) que faciliten el desarrollo de nuevas terapias

- **Tipo de tarea de MD:**

Analítica descriptiva, reglas de asociación, Descubrimiento de subgrupos

- Para ello
  - Se realiza un tratamiento de las cadenas de virus mediante transformadas de Fourier.
  - Se aplican algoritmos de descubrimiento de subgrupos.

C.J. Carmona, C. Chrysostomou, H. Seker, M.J. Del Jesus. **Fuzzy rules for describing subgroups from Influenza A virus using a multi-objective evolutionary algorithm**. Applied Soft Computing 13 (2013), 3439-3448.

Analítica descriptiva

- Motivación
- **Agrupamiento**
- Reglas de asociación
- Ejemplos

Agrupamiento ► Conceptos básicos

- ***Cluster:*** grupo o conjunto de objetos
  - Similares a cualquier otro incluído en el mismo *cluster*
  - Distintos a los objetos incluidos en otros grupos

- ***Clustering*** (análisis *cluster*):
  - Identifica clusters en los datos

- Es **clasificación no supervisada**: las clases no están predefinidas

- Aplicaciones:
  - Como preprocesamiento antes de aplicar otra técnica de descubrimiento del conocimiento
  - Como técnica de descubrimiento del conocimiento para obtener información acerca de la distribución de los datos

Agrupamiento ► Medidas de distancia y similaridad

- La propiedad más importante que debe verificar un *cluster* es que haya más cercanía entre las instancias que están dentro del *cluster* que respecto a las que están fuera del mismo

- **¿Qué es la similitud?** Si las instancias no están etiquetadas, **¿cómo medir la similitud entre ellas?**

(c) Eamonn Keogh, eamonn@cs.ucr.edu

Agrupamiento ► Medidas de distancia y similaridad

- La definición de la medida de distancia depende normalmente del tipo de variable:
  - Variables intervalares
    - Variables continuas para las que se utiliza una discretización: peso, edad, ...
  - Variables binarias/booleanas
  - Variables nominales/categóricas
  - Variables ordinales
  - Variables mixtas

Agrupamiento ► Medidas de distancia y similaridad

- El caso más simple: un único atributo numérico A

$$\text{Distancia}(X,Y) = A(X) - A(Y)$$

- En general, **atributos numéricos:**

  Distancia(X,Y) = Distancia euclídea entre X,Y

- **Atributos nominales:** La distancia se fija a 1 si los valores son diferentes, a 0 si son iguales

- ¿Tienen todos los atributos la misma importancia?
  - Si no tienen igual importancia, será necesario ponderar los atributos

Agrupamiento ► Medidas de distancia y similaridad

Distancia de Minkowski:

$$d(i,j) = \sqrt[q]{(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + ... + |x_{ip} - x_{jp}|^q)}$$

donde $i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ y $j = (x_{j1}, x_{j2}, \ldots, x_{jp})$ son dos objetos p-dimensionales, y q es un entero positivo

- Si $q = 1$, d es la distancia de Manhattan

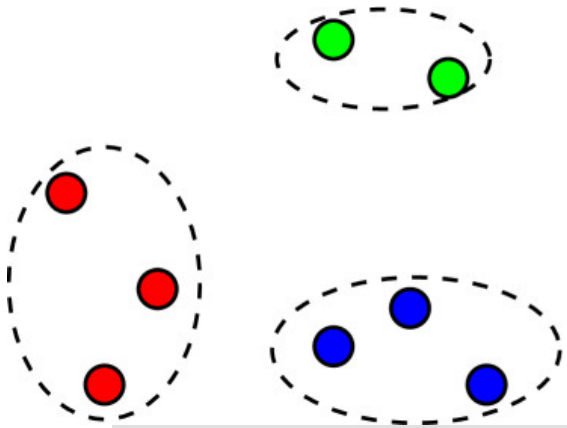
$$d(i,j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + ... + |x_{ip} - x_{jp}|$$

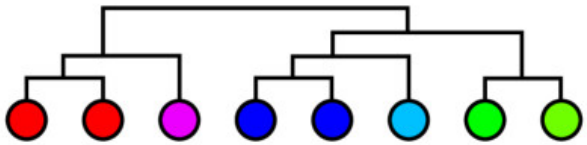
- *Si* $q = 2$, d es la distancia euclídea
- Se pueden utilizar pesos

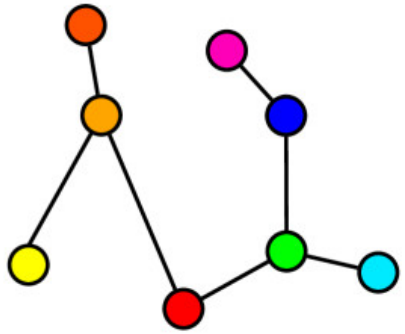
$$d(i,j) = \sqrt{w_1|x_{i1} - x_{j1}|^2 + w_2|x_{i2} - x_{j2}|^2 + ... + w_p|x_{ip} - x_{jp}|^2}$$

Agrupamiento ► Distintas aproximaciones al agrupamiento

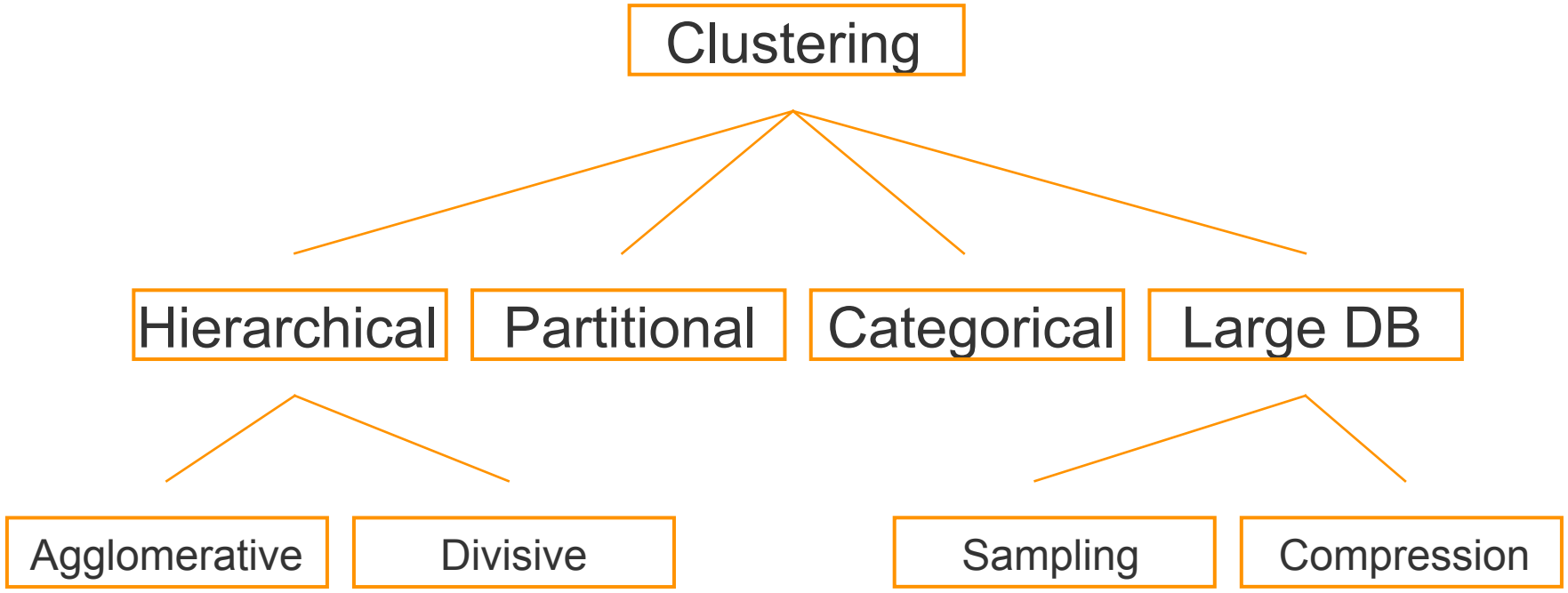
- **Algoritmos de particionamiento:** Construir distintas particiones y evaluarlas
- **Algoritmos jerárquicos:** Crear una descomposición jerárquica de los datos
- **Otros:** Basados en densidad, en rejillas, en modelos



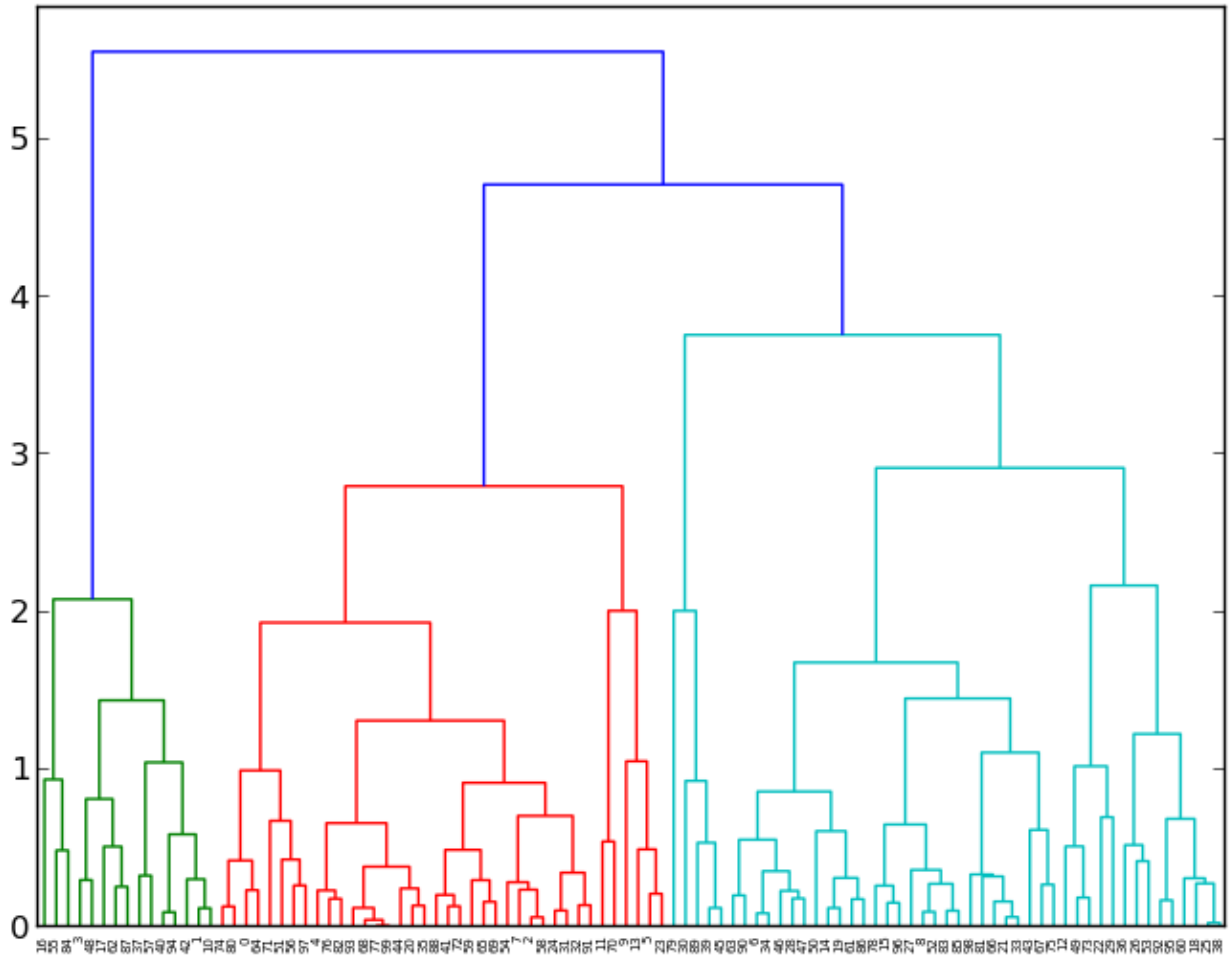




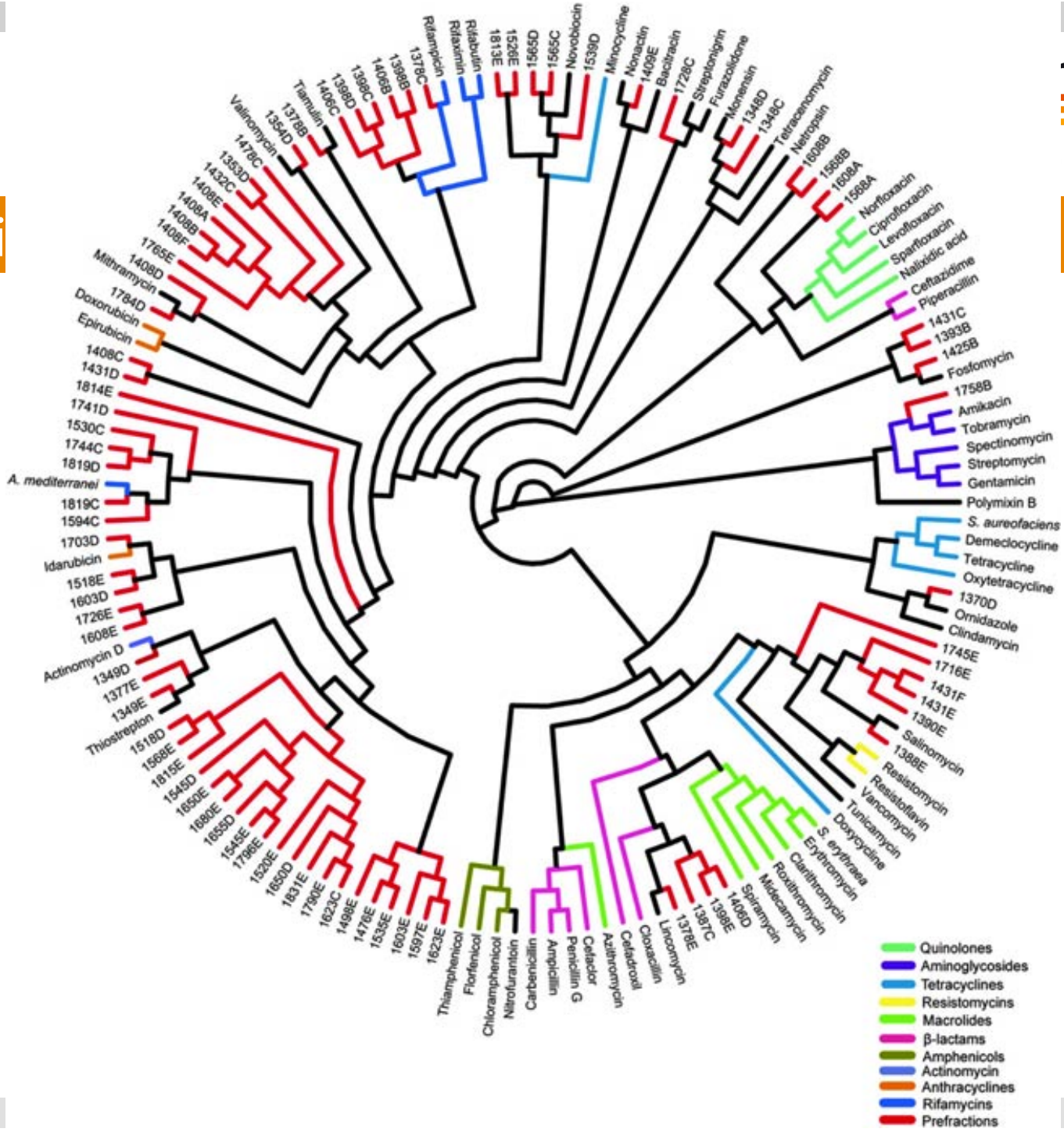
Agrupamiento ► Distintas aproximaciones al agrupamiento



Agrupamiento ► Distintas aproximaciones al agrupamiento



Agrupamiento



nto

Agrupamiento ► k-means

- **Método basado en particionamiento**: Construye una partición de la base de datos en un conjunto de *k clusters* que optimice un criterio de particionamiento
- **Idea básica:** las instancias se van moviendo entre clusters hasta que se alcanza el número de clusters deseado
- Variantes:
  - K-means: cada cluster se representa por el centro del cluster
  - K-medoids o PAM (particionamiento alrededor de los medoides): cada cluster se representa por uno de los objetos incluidos en el cluster.
- Parámetro de entrada: Nº de clusters (*k*)
- Criterio de particionamiento: minimizar la distancia euclídea total entre cada punto y su representante de cluster más cercano

Algoritmo *K-Means*

- Selección aleatoria de *k* puntos (centroides, representantes del cluster)
- Repetir hasta que converja
  - Asignación de datos: Cada punto se asigna al centroide más cercano
  - Re-colocación de "medias": Cada centroide se reasigna a al centro del grupo (media aritmética de todos los puntos contenidos)

El algoritmo converge cuando no hay nuevas asignaciones de datos a clusters

Agrupamiento ► k-means

Algorithm 2.1 The `k-means` algorithm

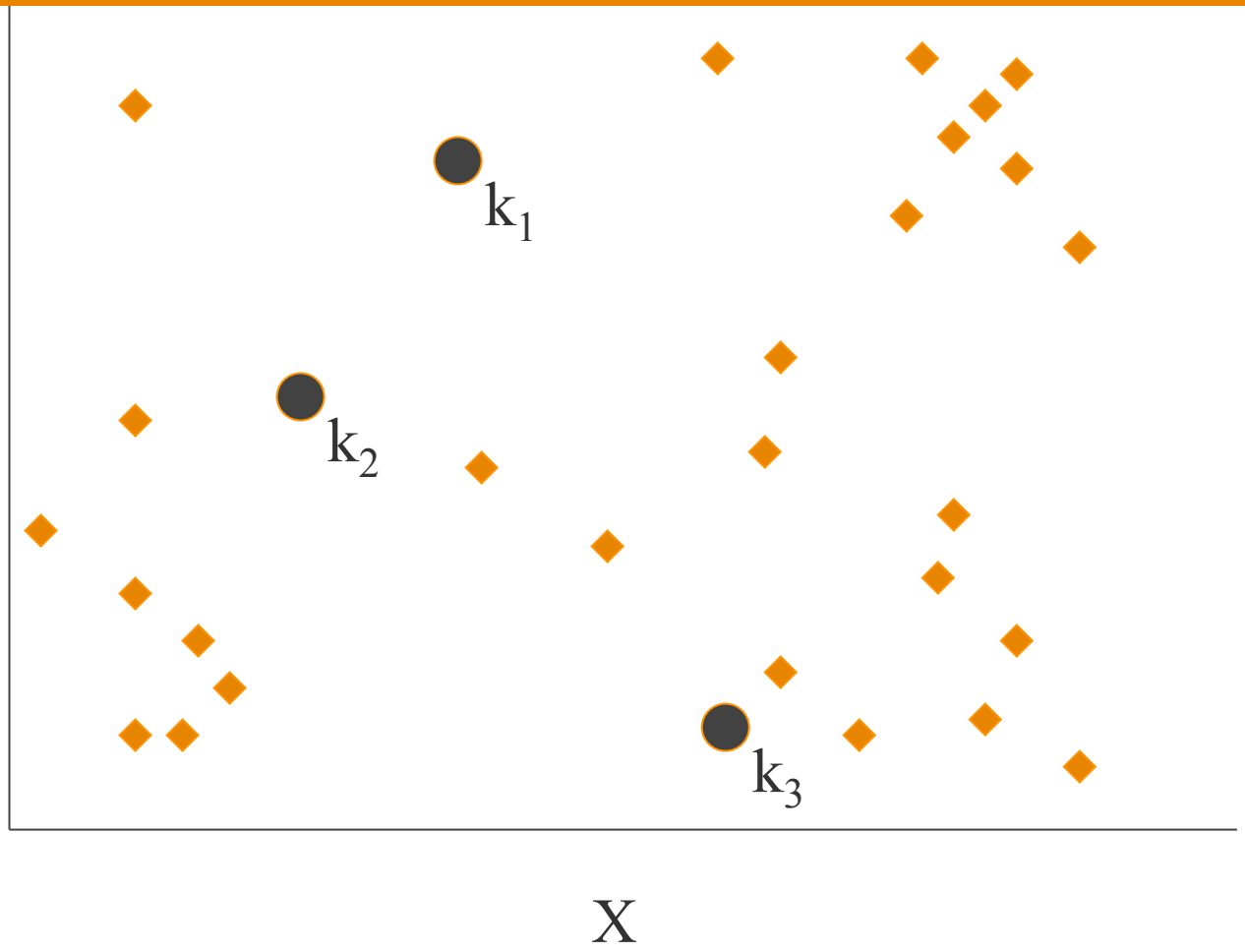
Input: Dataset D, number clusters k

Output: Set of cluster representatives C, cluster membership vector $\mathbf{m}$

```
    /* Initialize cluster representatives C */
    Randomly choose k data points from D
    Use these k points as initial set of cluster representatives C
    repeat
        /* Data Assignment */
        Reassign points in D to closest cluster mean
        Update m such that m_i is cluster ID of ith point in D
        /* Relocation of means */
        Update C such that c_j is mean of points in jth cluster
    until convergence of objective function sum_{i=1}^{N} (argmin_j ||x_i - c_j||_2^2)
```

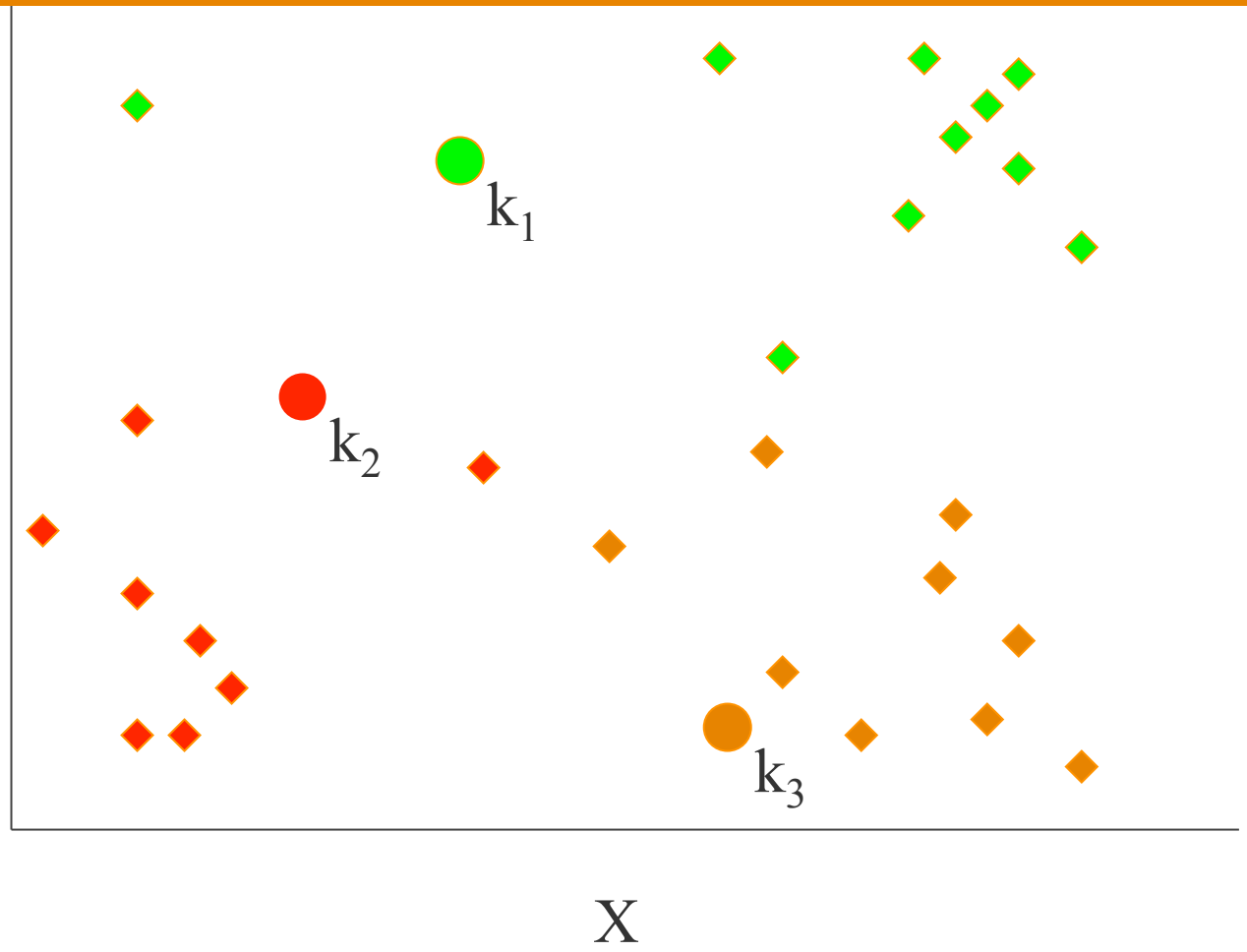
Agrupamiento ► k-means

Elige 3 centros de *clusters* (aleatoriamente)



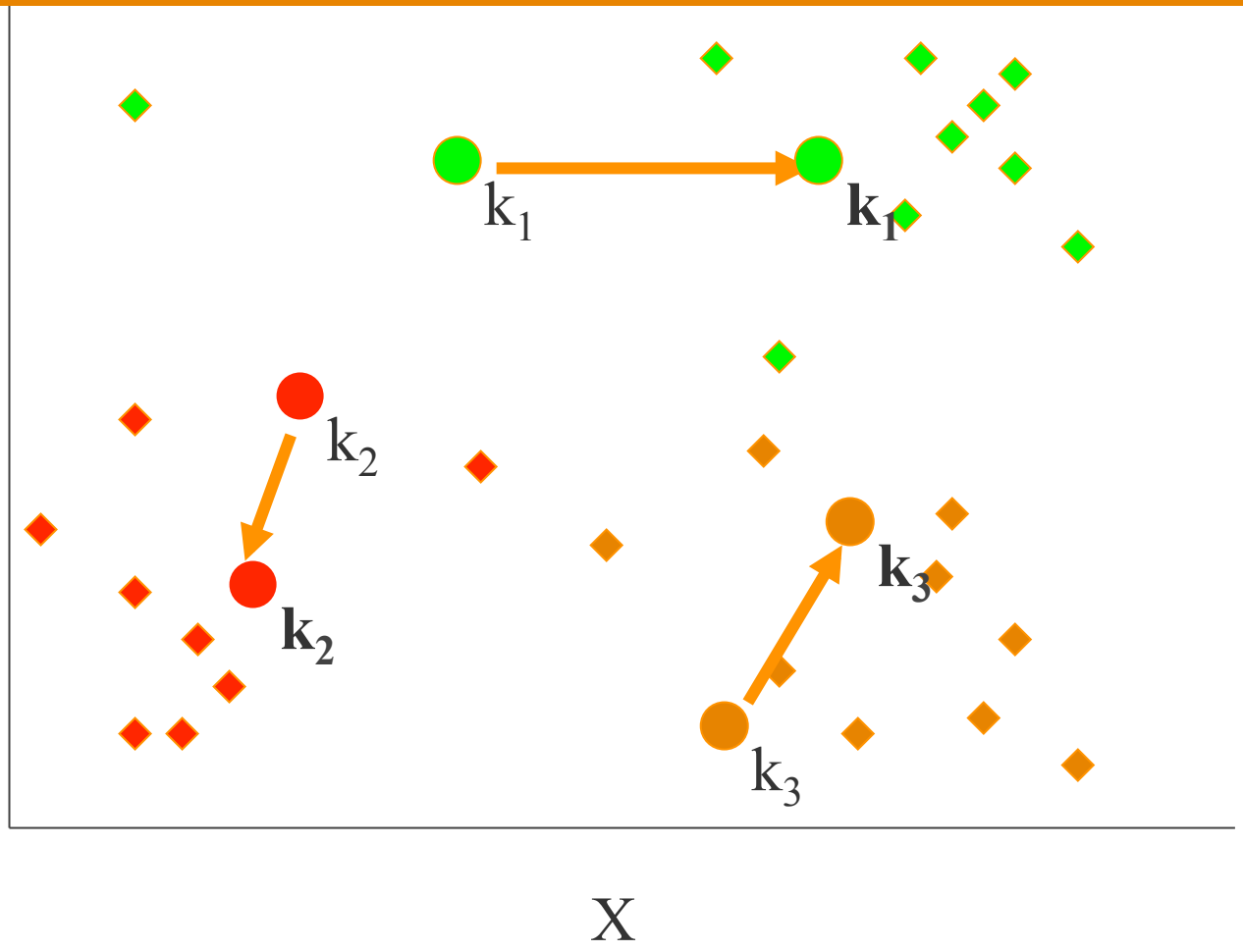
Agrupamiento ► k-means

Asigna cada punto al centro de *cluster* más cercano



Agrupamiento ► k-means

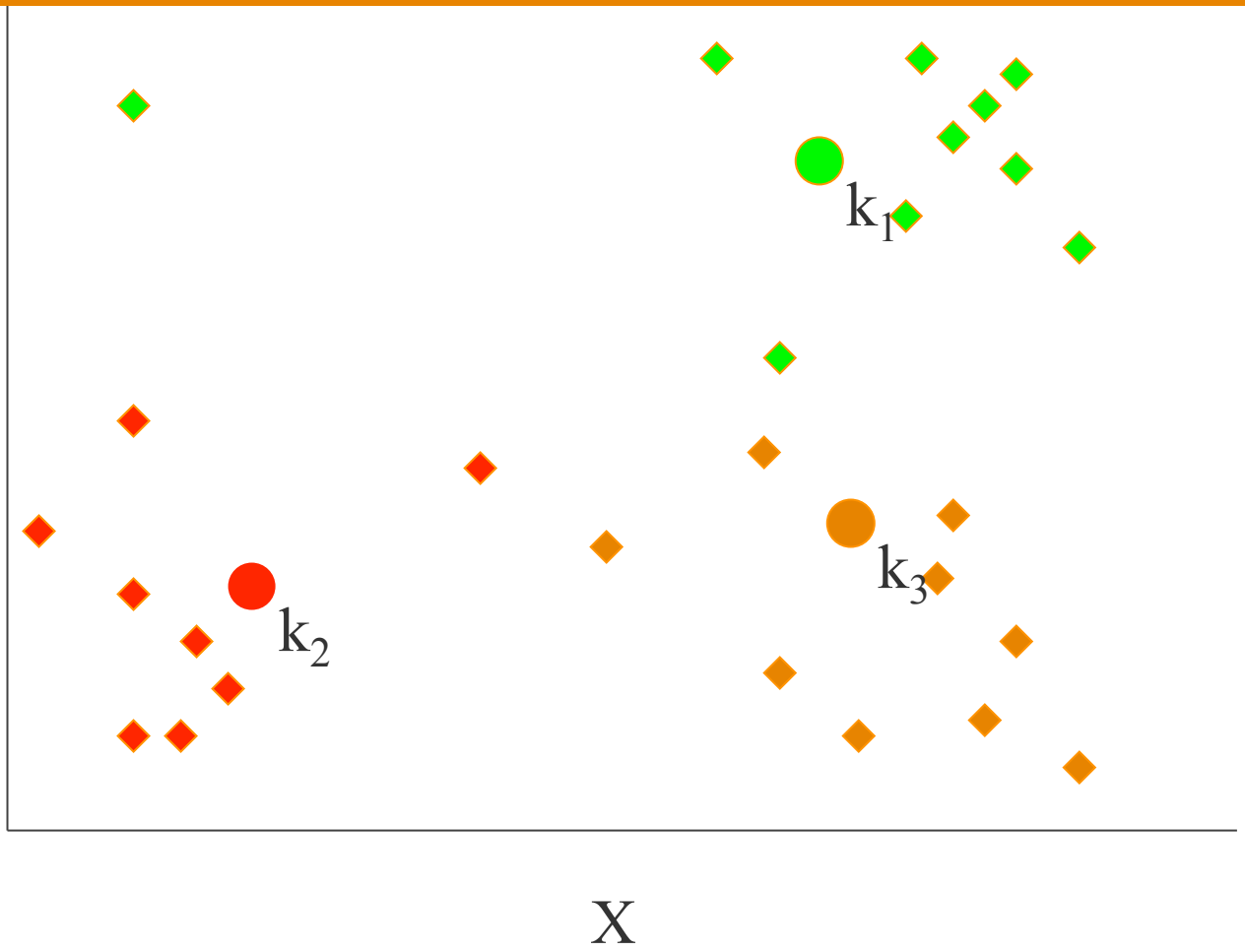
Mueve cada centro de *cluster* a la media de cada *cluster*



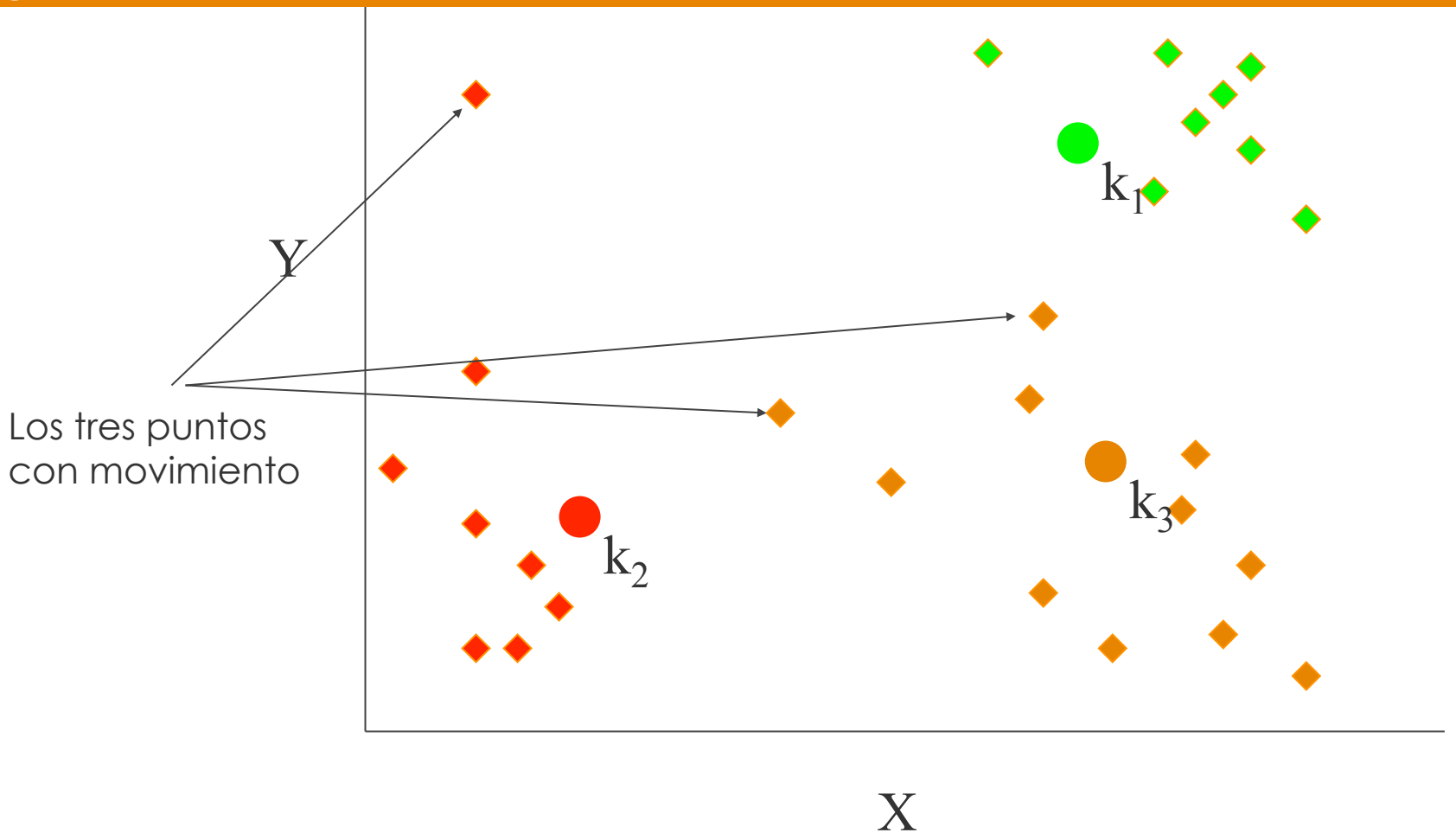
Agrupamiento ► k-means

Reasigna los puntos más cercanos a diferentes centros de *clusters*

¿Qué puntos se reasignan?

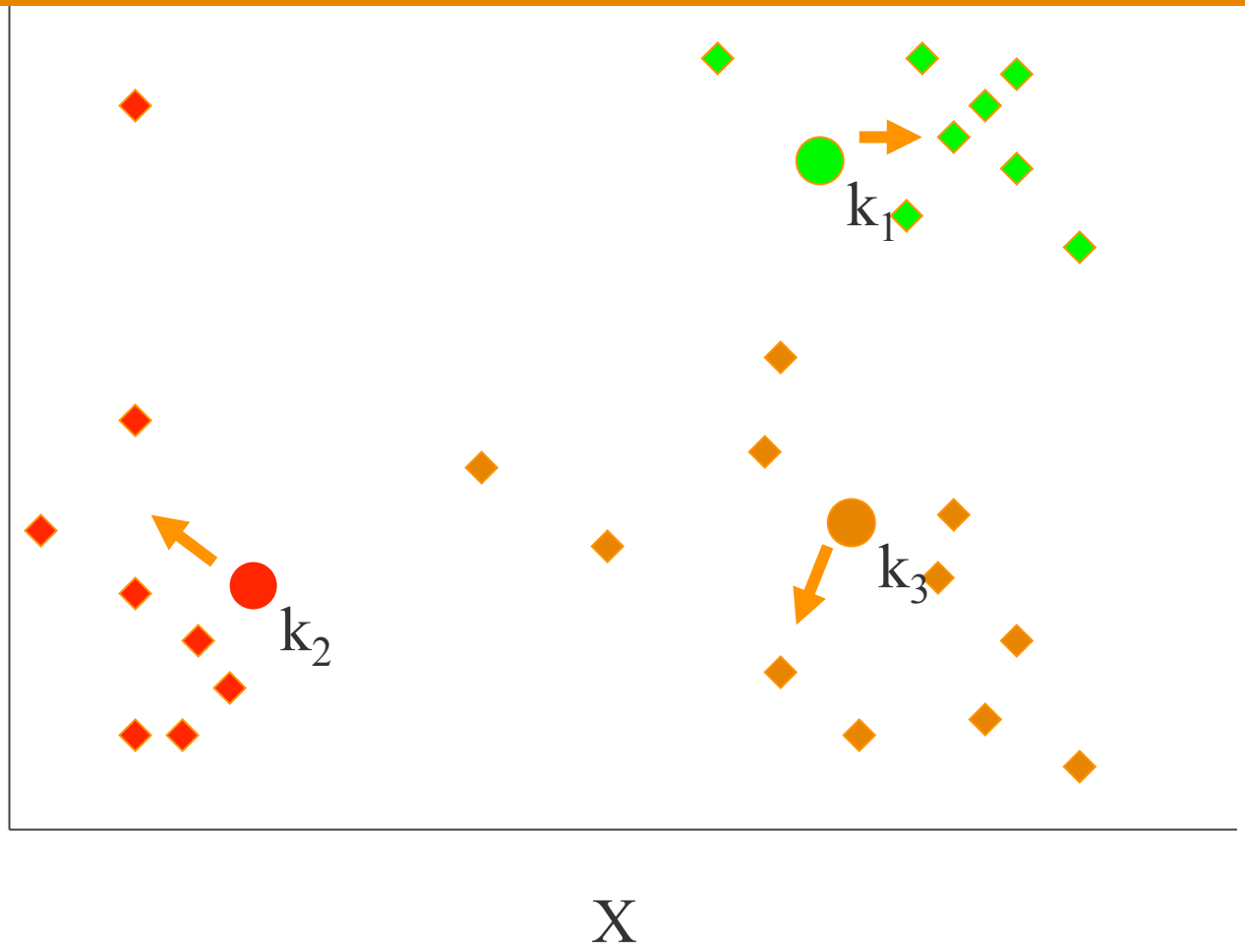


Agrupamiento ► k-means



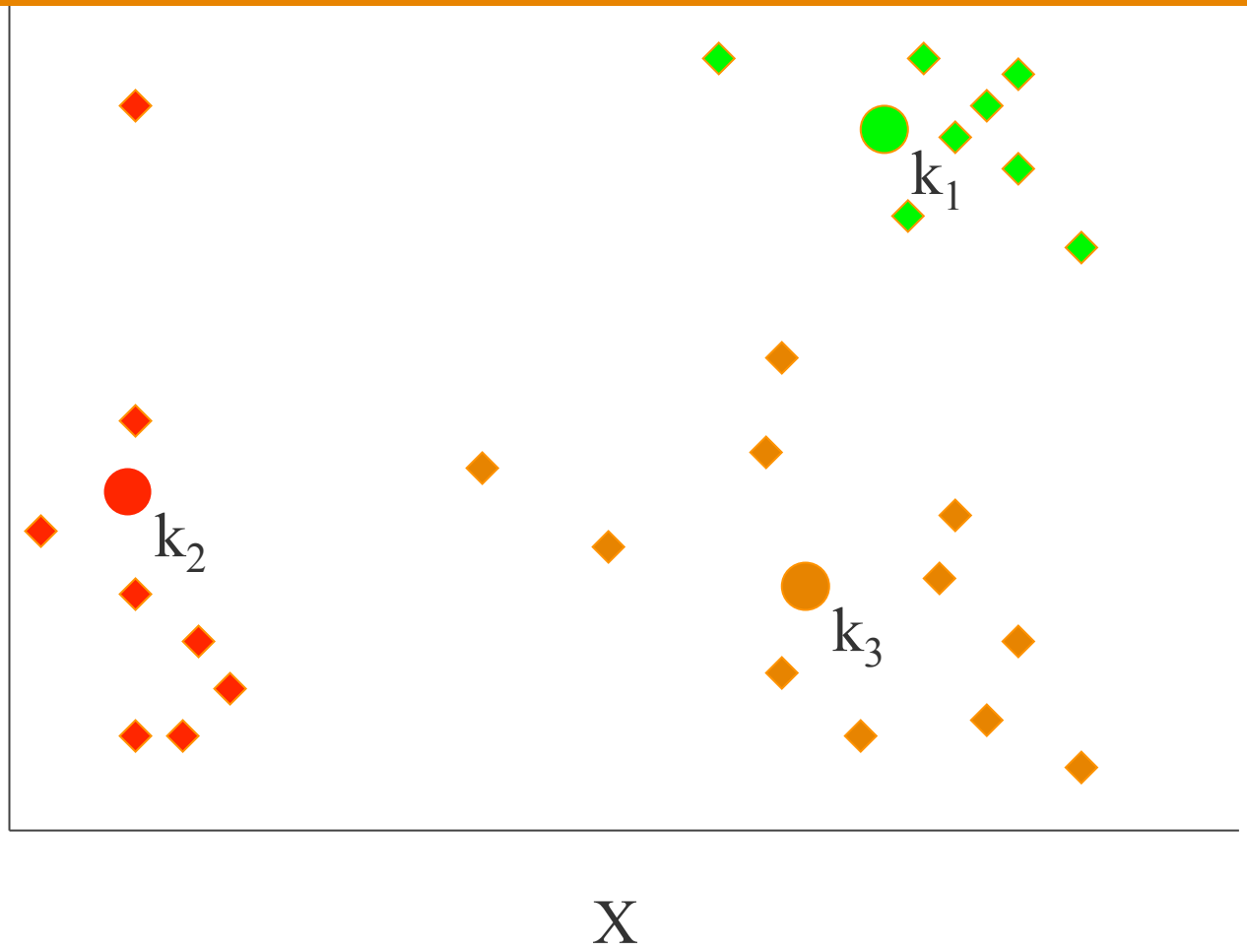
Agrupamiento ► k-means

Re-cálculo de los centros de *clusters*



Agrupamiento ► k-means

Mover los centros de *cluster* a las medias de los *clusters*



Agrupamiento ► k-means

Ventajas

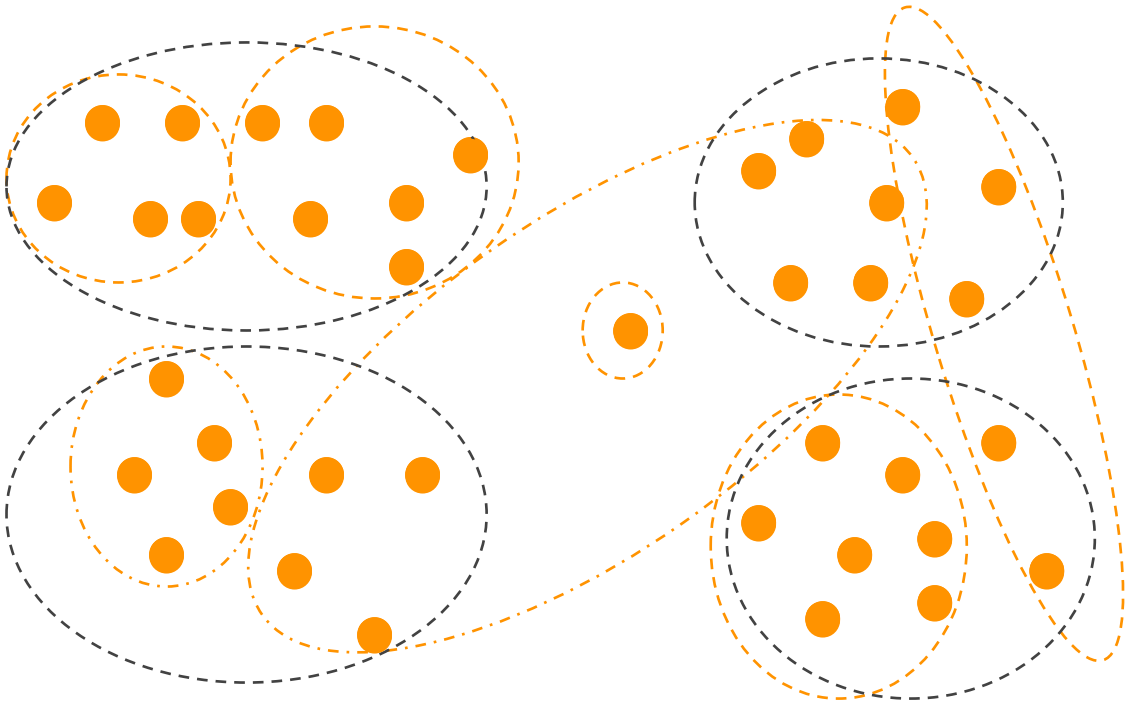
- Relativamente eficiente: $O(tkn)$, donde n es #objetos, k es #*clusters*, y t es #iteraciones
  - Normalmente, $k, t << n$
- Con frecuencia finaliza en un óptimo local, dependiendo de la elección inicial de los centros de *clusters*
  - Reinicializar las semillas.
  - Utilizar técnicas de búsqueda como algoritmos genéticos o enfriamiento estocástico

Desventajas

- Sólo es aplicable cuando el concepto de media es definible. ¿qué hacer con datos nominales?
- Necesidad de fijar anticipadamente el número de *clusters* (k)
- Débil ante datos ruidosos y/o con outliers
- Sólo indicado para *clusters* convexos (esféricos…)

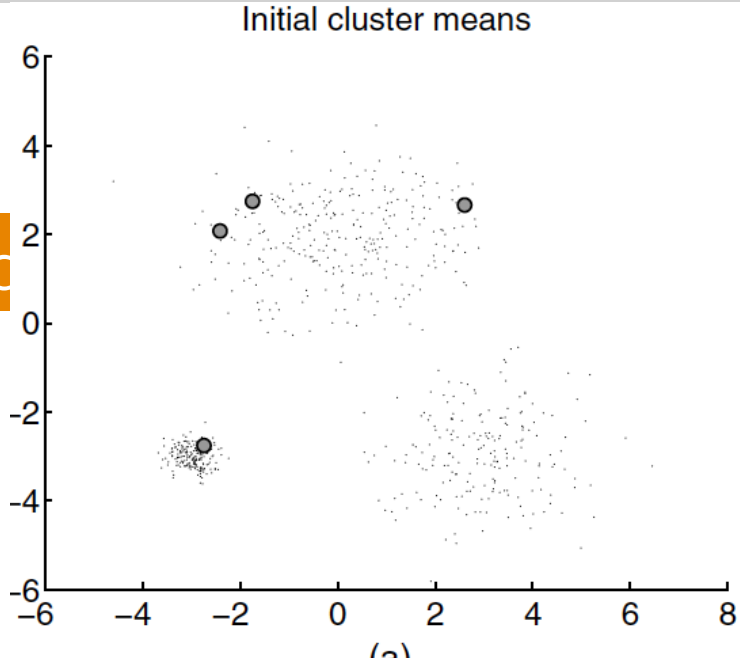
Agrupamiento ► k-means

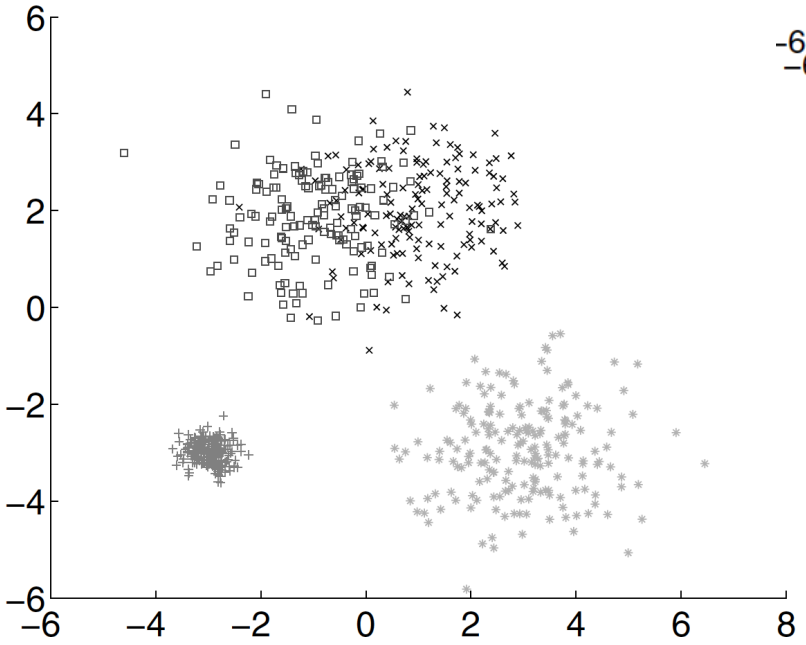
- Necesidad de fijar anticipadamente el número de *clusters* (k)

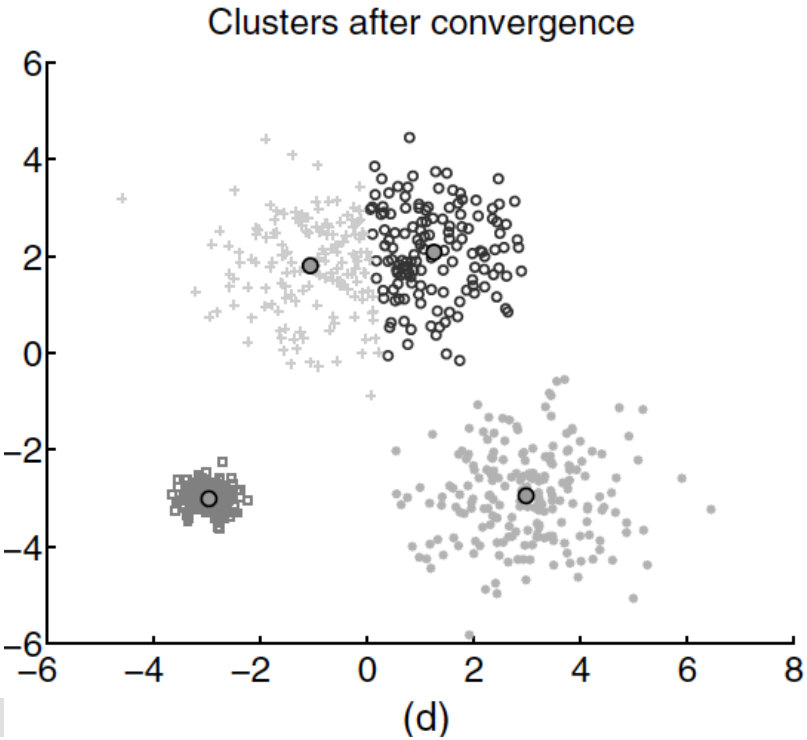


¿Elección de k?

Agrupamiento ► k-med

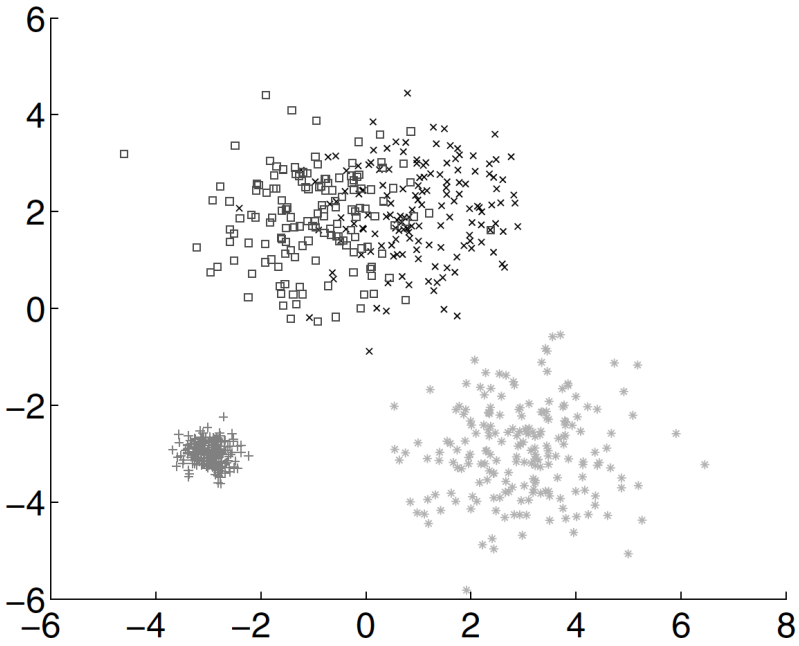


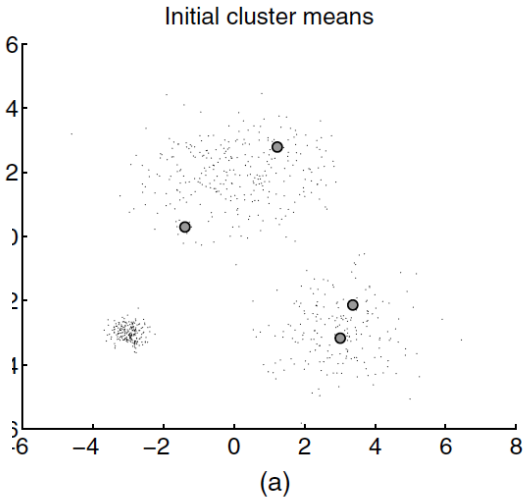


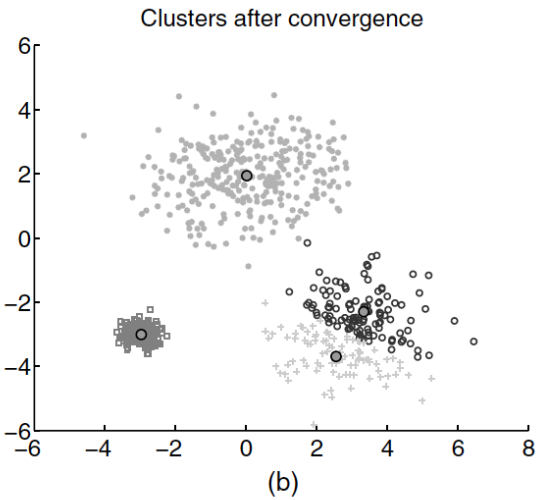


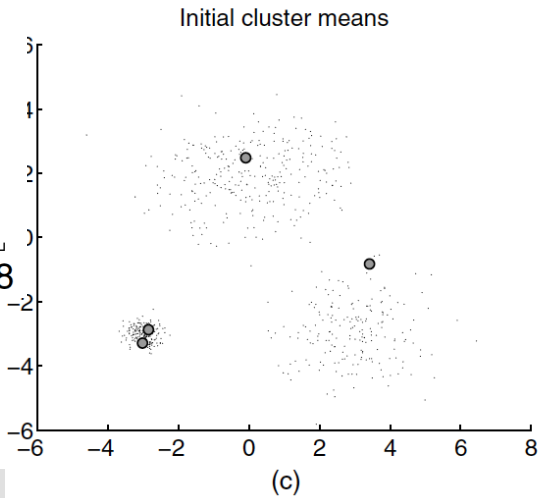
The Top Ten Algorithms in Data Mining. CRC 2009

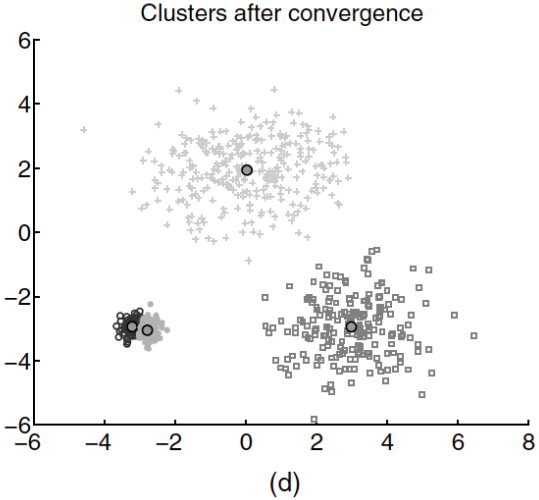
Agrupamiento ► k-means





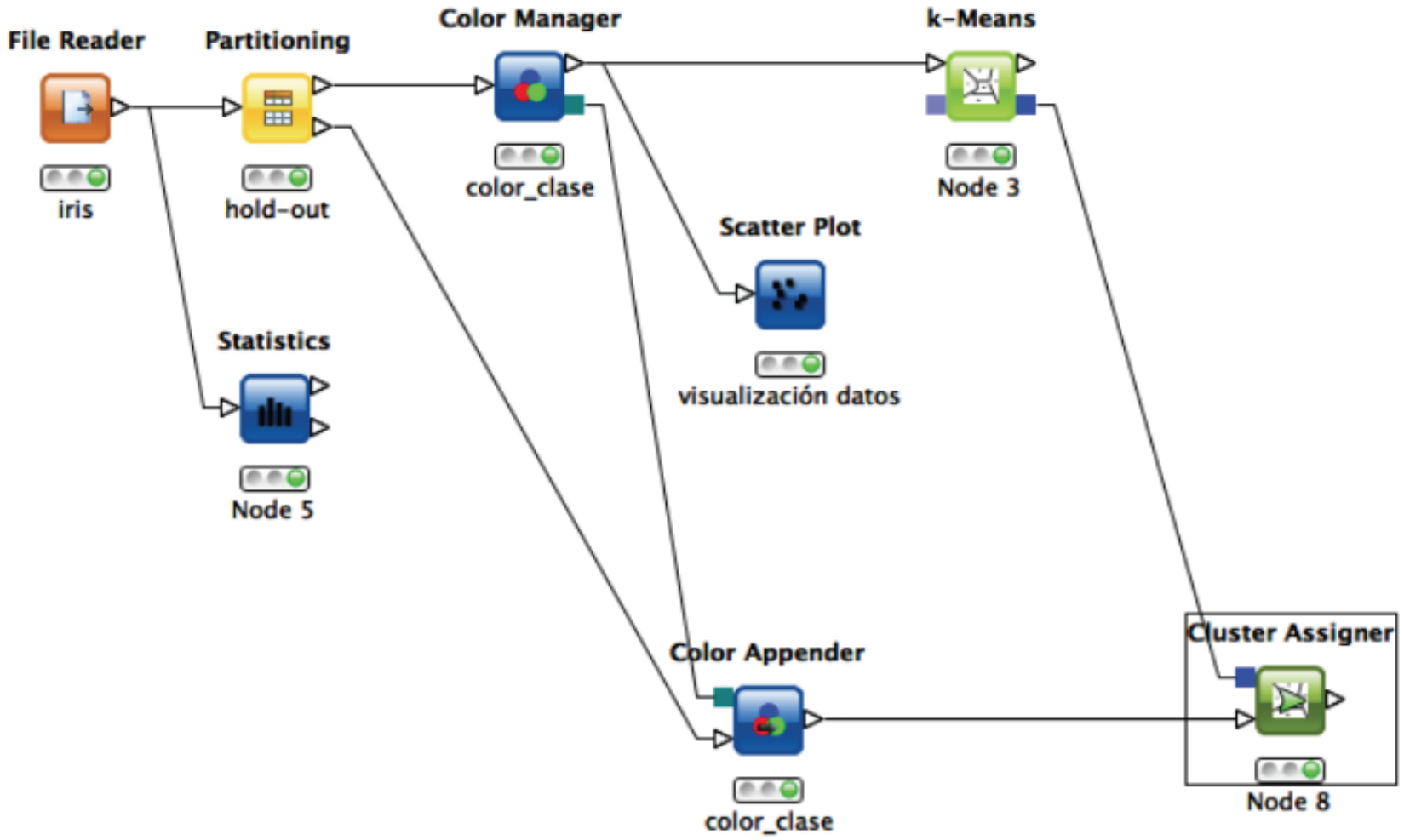




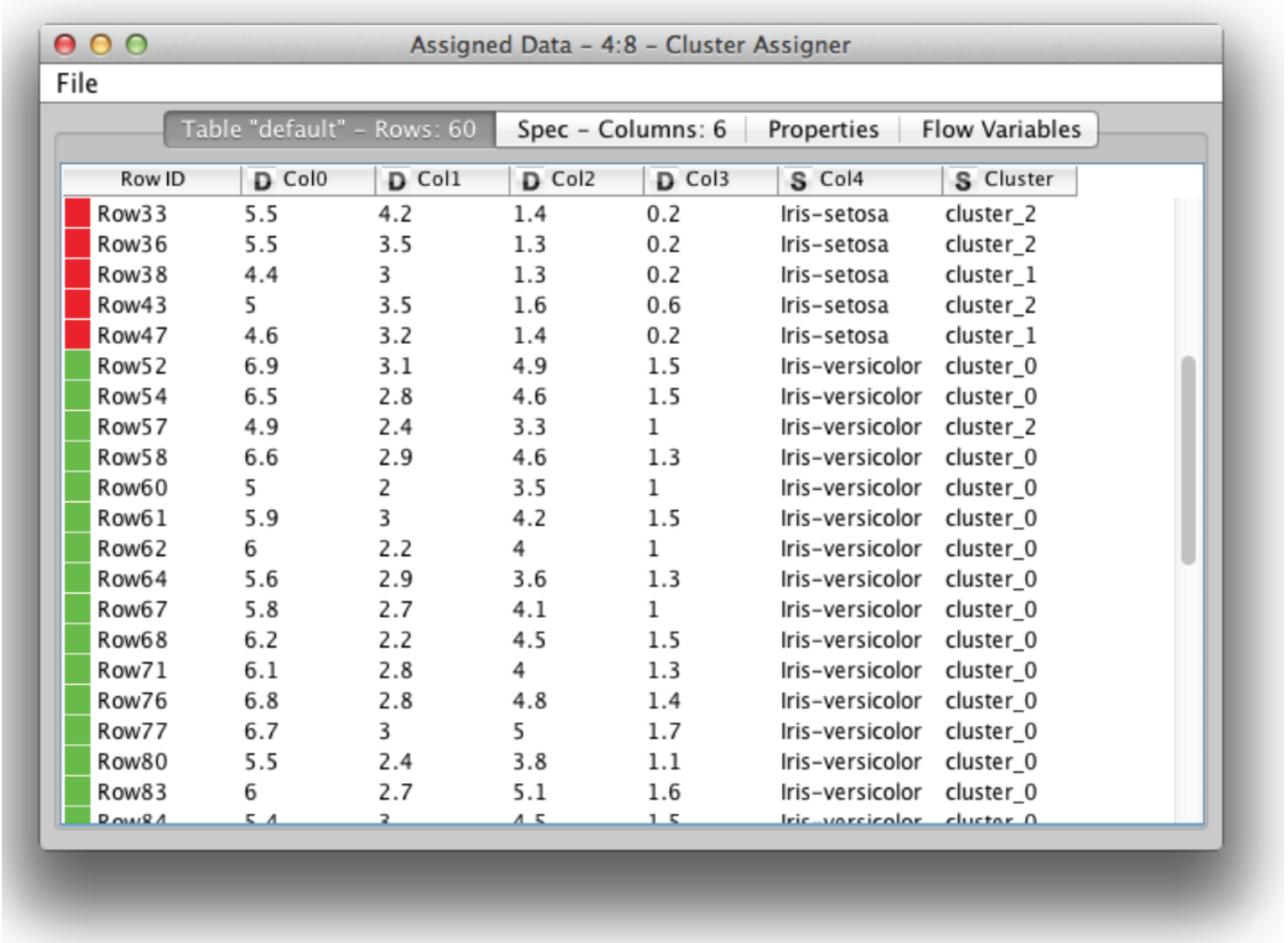


The Top Ten Algorithms in Data Mining. CRC 2009

Agrupamiento ► k-means



Agrupamiento ► k-means



Row ID	D Col0	D Col1	D Col2	D Col3	S Col4	S Cluster
Row33	5.5	4.2	1.4	0.2	Iris-setosa	cluster_2
Row36	5.5	3.5	1.3	0.2	Iris-setosa	cluster_2
Row38	4.4	3	1.3	0.2	Iris-setosa	cluster_1
Row43	5	3.5	1.6	0.6	Iris-setosa	cluster_2
Row47	4.6	3.2	1.4	0.2	Iris-setosa	cluster_1
Row52	6.9	3.1	4.9	1.5	Iris-versicolor	cluster_0
Row54	6.5	2.8	4.6	1.5	Iris-versicolor	cluster_0
Row57	4.9	2.4	3.3	1	Iris-versicolor	cluster_2
Row58	6.6	2.9	4.6	1.3	Iris-versicolor	cluster_0
Row60	5	2	3.5	1	Iris-versicolor	cluster_0
Row61	5.9	3	4.2	1.5	Iris-versicolor	cluster_0
Row62	6	2.2	4	1	Iris-versicolor	cluster_0
Row64	5.6	2.9	3.6	1.3	Iris-versicolor	cluster_0
Row67	5.8	2.7	4.1	1	Iris-versicolor	cluster_0
Row68	6.2	2.2	4.5	1.5	Iris-versicolor	cluster_0
Row71	6.1	2.8	4	1.3	Iris-versicolor	cluster_0
Row76	6.8	2.8	4.8	1.4	Iris-versicolor	cluster_0
Row77	6.7	3	5	1.7	Iris-versicolor	cluster_0
Row80	5.5	2.4	3.8	1.1	Iris-versicolor	cluster_0
Row83	6	2.7	5.1	1.6	Iris-versicolor	cluster_0
Row84	5.4	3	4.5	1.5	Iris-versicolor	cluster_0

Analítica descriptiva

- Motivación
- Agrupamiento
- **Reglas de asociación**
- Ejemplos

Minería de patrones frecuentes y reglas de asociación

- **Objetivo:** encontrar patrones (propiedades comunes) que comparten subgrupos suficientemente grandes del dataset
  - Minería de patrones frecuentes e inducción de reglas de asociación
  - Minería de secuencias frecuentes
  - Minería de grafos/árboles frecuentes
- **Ejemplo** clásico: análisis de cestas de mercado
  - Encontrar regularidades en el comportamiento de compra de clientes
  - Identificar productos que se compran juntos analizando registros de compras anteriores
  - El número de patrones potenciales es alto

Reglas de asociación

- **Regla de asociación:** describe en forma de regla una relación de asociación o correlación entre conjuntos de items

- **Algoritmos de extracción de reglas de asociación**:
  - Dada una base de datos de transacciones, donde cada transacción es una lista de items
  - Objetivo: encontrar todas las reglas que co-relacionen la presencia de un conjunto de items con otro conjunto de items

Si un cliente compra pan y vino, probablemente comprará también queso

- **Áreas de aplicación:**
  - Análisis de cestas de mercado
  - Control y mejora de calidad
  - Gestión de clientes
  - Detección de fraudes
  - Click stream analysis
  - Análisis de enalces web
  - Análisis genómico,…

Reglas de asociación ► Conceptos básicos

- **Transacción**:

Formato relacional <Tid, item>	Formato compacto <Tid, itemset>
<1, item1>	<1, {item1,item2}>
<1, item2>	<2, {item3}>
<2, item3>	

  - Item (o artículo) : elemento individual
  - Itemset (o conjunto): conjunto de items/artículos
  - Soporte de un conjunto I: nº de transacciones conteniendo I
  - Soporte mínimo m_s : umbral de soporte
  - Conjunto frecuente: con soporte ≥ m_s

- Los **conjuntos frecuentes** representan conjuntos de artículos que están correlacionados positivamente

Reglas de asociación ► Soporte y confianza

Dada la regla $X \Rightarrow Y$

- **Soporte**: *probabilidad* de que una transacción contenga {X & Y}

  Frecuencia de patrones con X e Y

- ***Confianza***: probabilidad condicional P(Y | X)

  Fuerza de la implicación $X \Rightarrow Y$

Marco suporte-confianza: Una regla $X \Rightarrow Y$ *es válida si*

- $Support(X \Rightarrow Y) \geq minsupp$
- $Conf(X \Rightarrow Y) \geq minconf$

# transacción	artículos
2000	A,B,C
1000	A,C
4000	A,D
5000	B,E,F

Sea el valor mínimo para confianza y soporte 50%,

$A \Rightarrow C$ (50%, 66.6%)

$C \Rightarrow A$ (50%, 100%)

Reglas de asociación ► Algoritmos clásicos

A priori

- R. Agrawal, R. Srikant. Fast algorithms for mining association rules. In: Proc. 20th Int. Conf. on very Large Databases (VLDB 1994), pp. 487–499.

Eclat

- M. Zaki, S. Parthasarathy, M. Ogihara, W. Li. New algorithms for fast discovery of association rules. In: Proc. 3rd Int. Conf. on Knowledge Discovery and Data Mining (KDD'97, Newport Beach, CA), pp. 283–296. AAAI Press, Menlo Park (1997)

FP-Growth

- J. Han, H. Pei, Y. Yin. Mining frequent patterns without candidate generation. In: Proc. Conf. on the Management of Data (SIGMOD'00, Dallas, TX), pp. 1–12. ACM Press, NewYork (2000)

OPUS based algorithm

- G. Webb, Efficient search for association rules. In: *Proceedings of International Conference on Knowledge Discovery and Data Mining*, 2000: 99-107

Reglas de asociación ► Apriori

Idea general:

- Encontrar conjuntos de items (itemsets) frecuentes
- Generar reglas de asociación fuerte a partir de los itemsets frecuentes

transaction database

1: $\{a, d, e\}$
2: $\{b, c, d\}$
3: $\{a, c, e\}$
4: $\{a, c, d, e\}$
5: $\{a, e\}$
6: $\{a, c, d\}$
7: $\{b, c\}$
8: $\{a, c, d, e\}$
9: $\{b, c, e\}$
10: $\{a, d, e\}$

Frequent item sets (with support) (minimum support $s_{\min} = 3$)

0 items	1 item	2 items	3 items
$\emptyset$: 10	$\{a\}$: 7	$\{a, c\}$: 4	$\{a, c, d\}$: 3
	$\{b\}$: 3	$\{a, d\}$: 5	$\{a, c, e\}$: 3
	$\{c\}$: 7	$\{a, e\}$: 6	$\{a, d, e\}$: 4
	$\{d\}$: 6	$\{b, c\}$: 3	
	$\{e\}$: 7	$\{c, d\}$: 4	
		$\{c, e\}$: 4	
		$\{d, e\}$: 4	

Principio A priori: cualquier subconjunto de un itemset frecuente debe ser frecuente

Reglas de asociación ► Apriori

Esquema general del algoritmo

- Generar itemsets candidatos
- Podar aquellos que no son frecuentes

Al final utiliza los itemsets frecuentes para generar las reglas de asociación

En el proceso de búsqueda de itemsets frecuentes se aplica el **principio de poda**:

Si existe algún itemset "infrecuente", no hay necesidad de generar los itemsets que lo incluyan

Reglas de asociación ► Apriori

function apriori $(B, T, s_{\min})$
begin
  $k := 1;$
  $E_k := \bigcup_{i \in B} \{\{i\}\};$
  $F_k := \text{prune}(E_k, T, s_{\min});$
  while $F_k \neq \emptyset$ **do begin**
    $E_{k+1} := \text{candidates}(F_k);$
    $F_{k+1} := \text{prune}(E_{k+1}, T, s_{\min});$
    $k := k + 1;$
  end;
  return $\bigcup_{j=1}^{k} F_j;$
end $(* \text{ apriori } *)$

Conjunto de itemsets candidatos de tamaño k

Conjunto de itemsets frecuentes de tamaño k

Comienza con itemsets de tamaño 1

Determina los frecuentes

Crea itemsets de tamaño superior

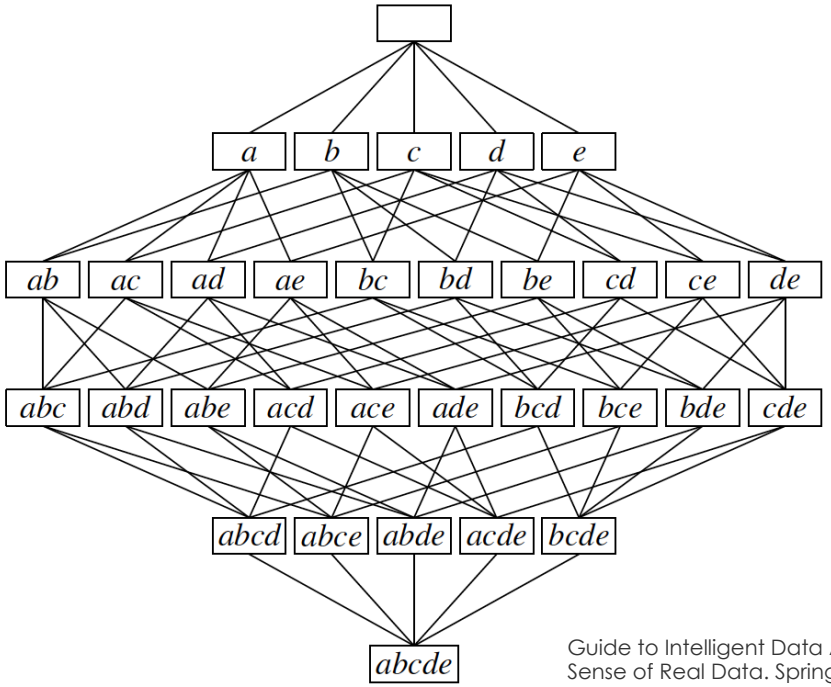
Determina los frecuentes

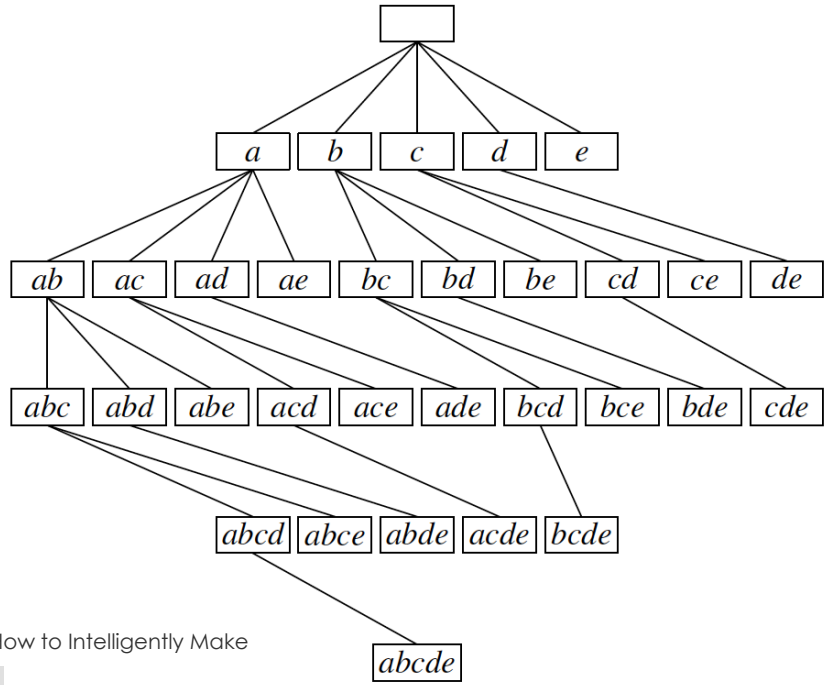
Devuelve todos los itemsets frecuentes

Reglas de asociación ► Apriori

Generación de los itemsets candidatos de nivel superior

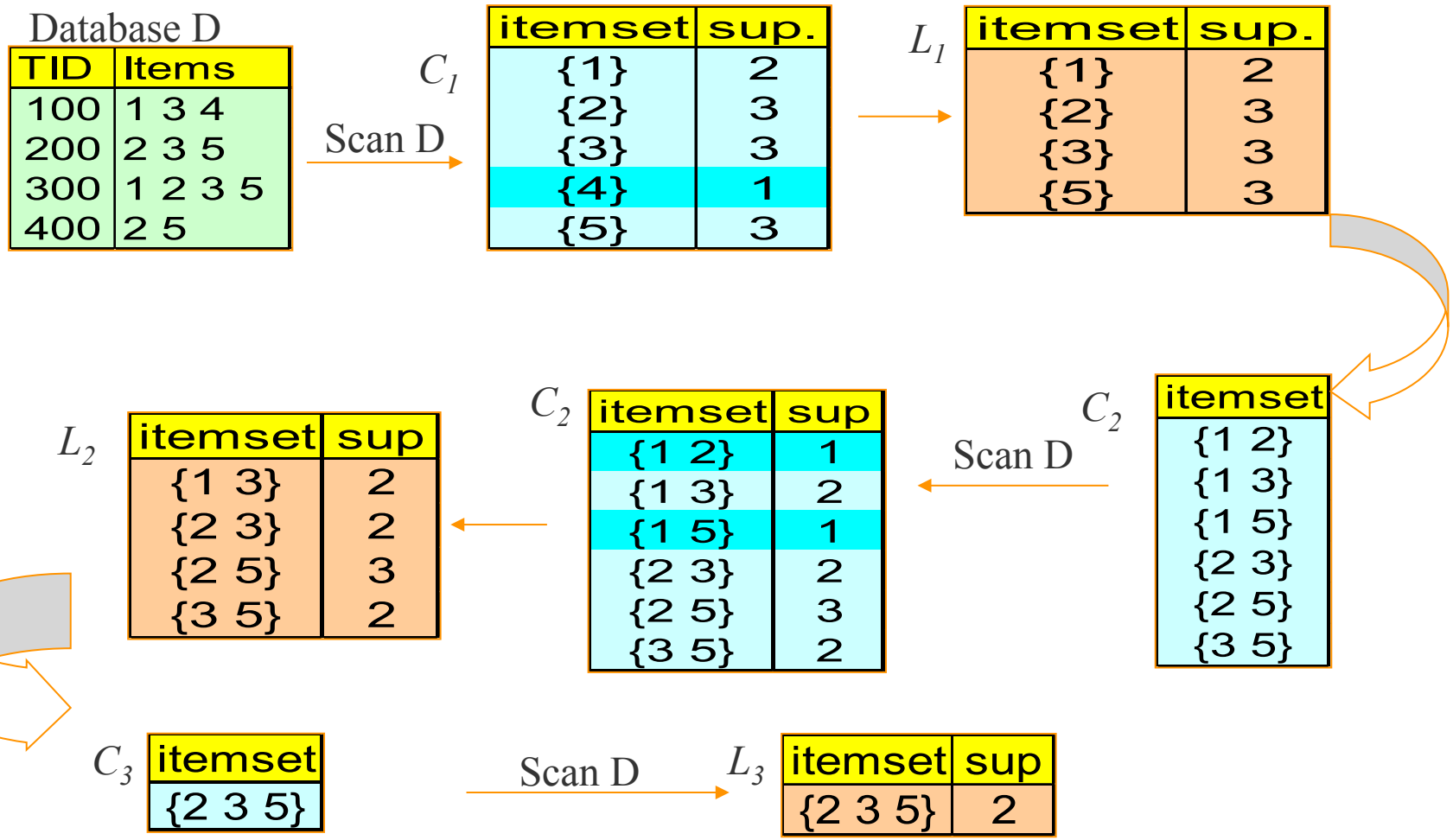
- Un itemset de tamaño k se puede generar de k! formas distintas
- La búsqueda se puede reducir asignando a cada itemset un único itemset padre
  - El diagrama Hasse se convierte en un árbol





Guide to Intelligent Data Analysis. How to Intelligently Make Sense of Real Data. Springer 2010

Reglas de asociación ▶ A priori ▶ Generación de itemsets frecuentes



- Una vez disponemos de los conjuntos frecuentes basta calcular la confianza y añadir las reglas que cumplan con los umbrales mínimos

Base de datos

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

Conjunto frecuente

itemset	sup
{2 3 5}	2

Si los valores mínimos son 50%, entonces:

2,3 -> 5 [50,100]

2,5 -> 3 [50,66]

3,5 -> 2 [50,100]

2 -> 3,5 [50,66]

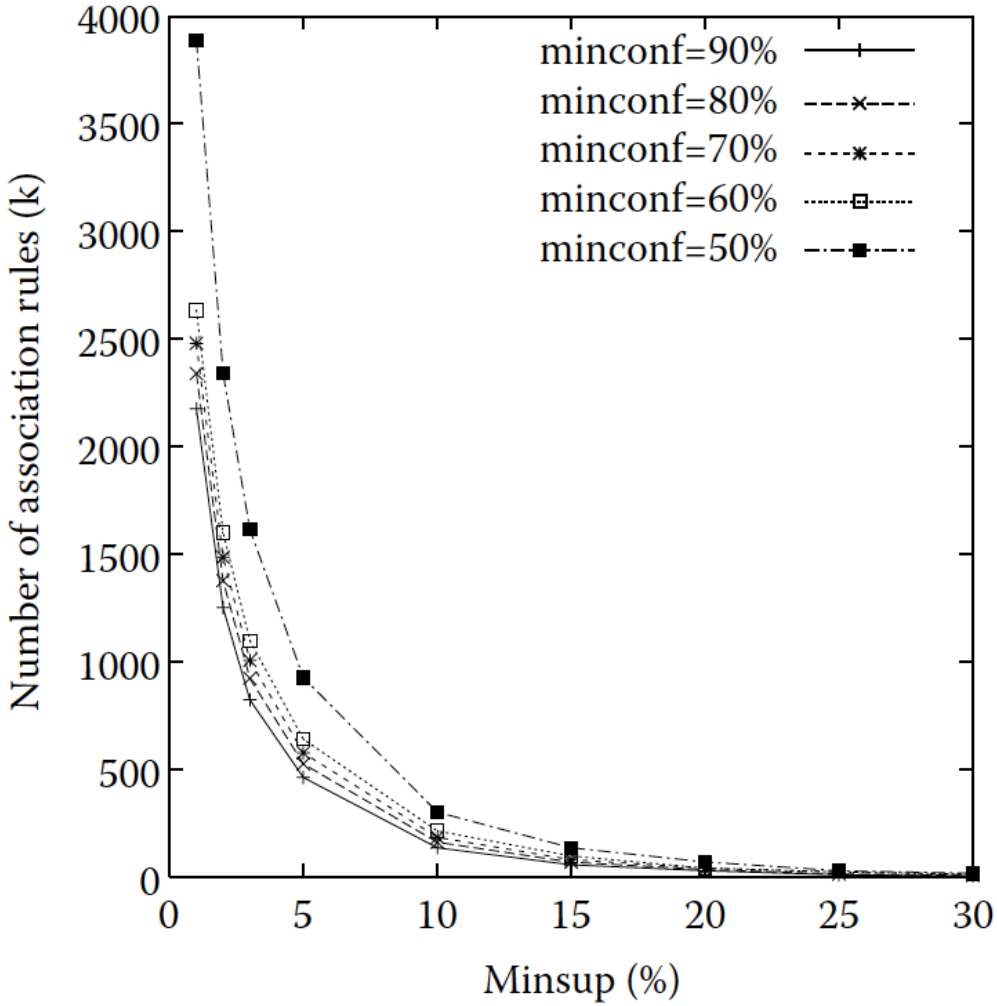
3 -> 2,5 [50,66]

5 -> 2,3 [50,100]

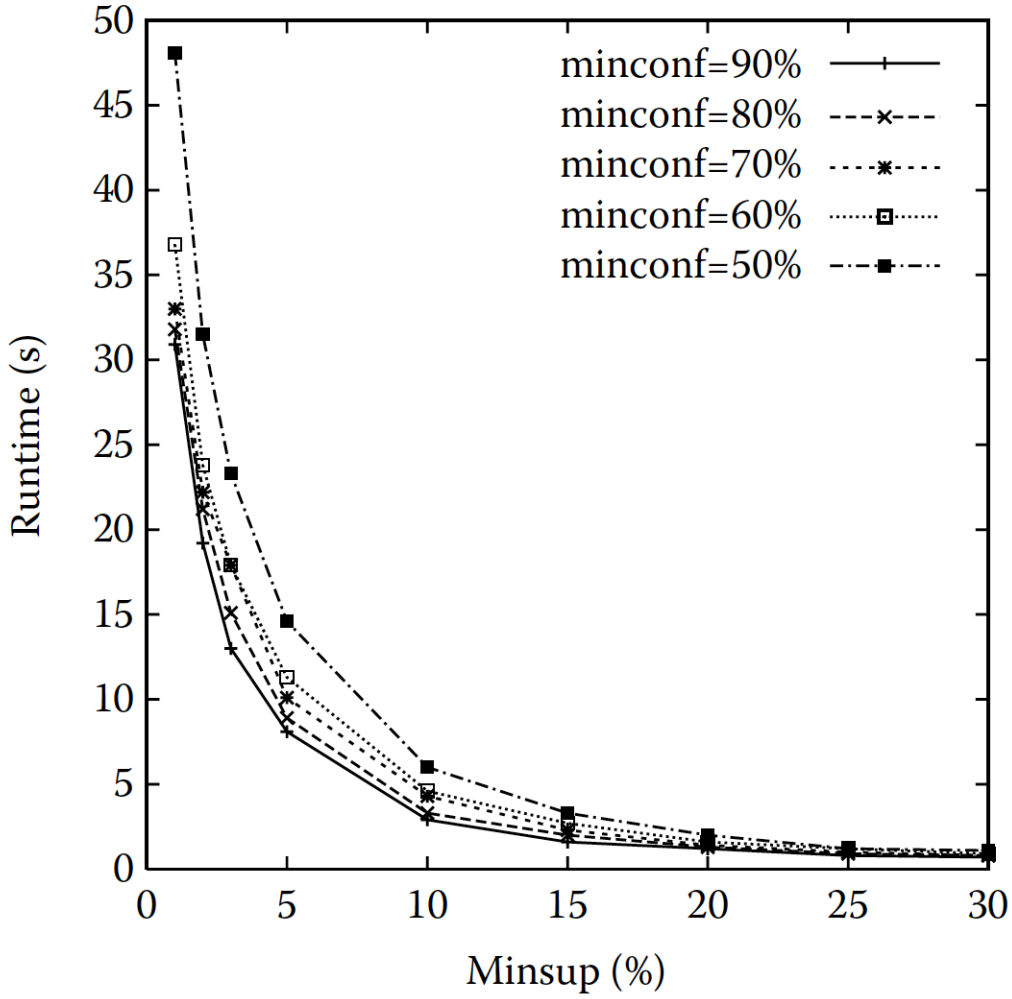
Reglas de asociación ► A priori

- Soporte mínimo:
  - Alto ⇒ pocos conjuntos frecuentes
    ⇒ pocas reglas válidas que ocurren con frecuencia
  - bajo ⇒ muchas reglas válidas que ocurren raramente
- Confianza mínima:
  - Alta ⇒ pocas reglas, pero todas "casi ciertas lógicamente"
  - Baja ⇒ muchas reglas, pero muchas de ellas muy inciertas
- Valores típicos: soporte = 2-50 % confianza = 70-90 %

Reglas de asociación ► A priori



Reglas de asociación ► A priori



Reglas de asociación ► Otras medidas de calidad

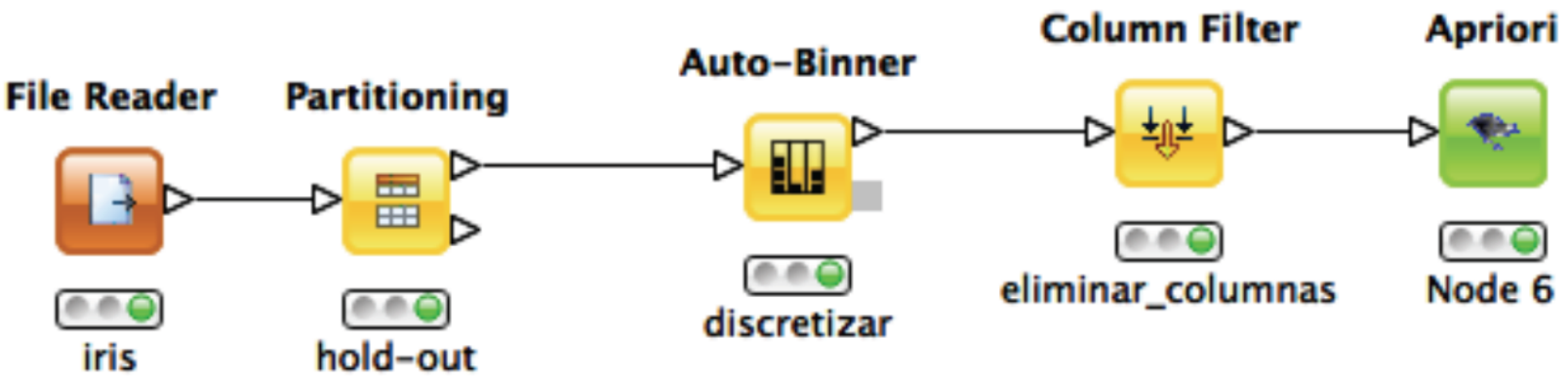
- Piatestky- Shapiro 1991: $X \rightarrow Y$ no es interesante si

$$support(X \rightarrow Y) \approx support(X) \times support(Y)$$

- Lift/Interest $\quad Interest(X, Y) = \dfrac{p(X \cup Y)}{p(X)p(Y)}$

- $\begin{aligned} \text{leverage}(X \Rightarrow Y) &= \text{support}(X \Rightarrow Y) - \text{support}(X) \times \text{support}(Y) \\ &= \text{support}(X) \times (\text{confidence}(X \Rightarrow Y) - \text{support}(Y)) \end{aligned}$

Reglas de asociación



Analítica descriptiva

- Motivación
- Agrupamiento
- Reglas de asociación
- **Ejemplos**

Un caso de estudio: *Bank of America*

Línea de negocio **objetivo**:

Home equity line of credit

Alternativas de mejora:

- Bajar el interés
  - Se reduce el beneficio
  - Clientes existentes cambian al nuevo interés → baja + el beneficio
  - Atrae clientes "desleales"

Diferentes campañas publicitarias directas no consiguen los resultados esperados

Un caso de estudio: *Bank of America*

1. Identificando la **oportunidad de negocio**

Puntos clave determinados por los expertos:

- Individuos con niños en edad escolar quieren pedir prestado en contra de su garantía hipotecaria para pagar facturas de matrícula

- Individuos con ingresos altos variables quieren utilizar su garantía hipotecaria para suavizar "altibajos" en ingresos

Un caso de estudio: *Bank of America*

2. Aplicando **Minería de Datos** (Hyperparallel – Yahoo)

Datos de 42 sistemas ¡Con información desde 1914!

- **Árboles de decisión**
  - Objetivo: Obtener reglas para clasificar los clientes respecto a si responderán a la oferta o no
  - Resultado: Se etiquetaron nuevos posibles clientes

- **Búsqueda de patrones secuenciales**
  - Objetivo: Descubrir la secuencia de eventos que preceden habitualmente a solicitudes con éxito en el pasado
  - Resultado: Se determinó cuándo un cliente era más susceptible de contratar un crédito de este tipo

Un caso de estudio: *Bank of America*

2. Aplicando **Minería de Datos** (Hyperparallel – Yahoo) (II)

- **Agrupamiento**
  - Objetivo: Segmentar los clientes en grupos con características similares
  - Resultados: 14 grupos, de los cuales los expertos seleccionan uno:
    - 39% de los individuos del grupo tienen cuentas personales y de empresa
    - Incluye más de ¼ de los clientes etiquetados como susceptibles de aceptar la oferta

¿Se usaría este producto para comenzar un negocio?

Un caso de estudio del *Bank of America*

3. **Actuando sobre los resultados**

- Se lanza una campaña de investigación de mercado que confirma la hipótesis
- Se cambia el mensaje de la campaña:

Utiliza el valor de tu casa para enviar los niños al colegio

Ahora que tu casa está vacía, utiliza tu garantía hipotecaria para hacer lo que siempre quisiste hacer

4. **Midiendo** los resultados

La respuesta de las campañas de publicidad pasó de un 0.7% a un 7%

Para finalizar un caso de estudio en medios sociales

Análisis de medios sociales en aplicaciones reales

- Gestión de crisis (tsunamis, huracanes, terremotos,…)
- **Objetivo**: proporcionar asistencia eficiente en función de la información de que se disponga
- Herramientas de **colaboración abierta distribuida** (*crowdsourcing*) como twitter, ushahidi, sahana

Ushahidi

SAHANA Disaster Management System "Coz disasters are NEVER propietari"

Para finalizar un caso de estudio en medios sociales

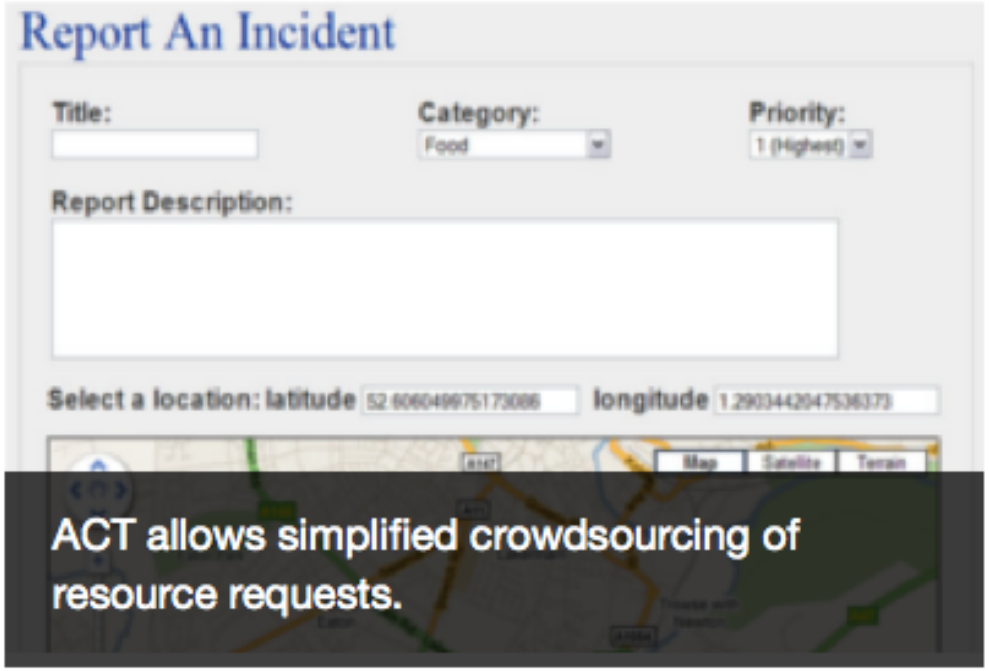
ASU Rastreador de coordinación Arizona State University

- **Objetivo**: Ayudar a organizaciones humanitarias a recoger, visualizar y actuar sobre peticiones de ayuda humanitaria
- Es un **sistema de coordinación de respuesta a eventos**
- Es una forma fácil, eficiente y abierta de hacer más efectiva la comunicación y coordinación de fuentes de información (entre ellas sociales) y heterogéneas
- **Integra y analiza** información distribuida mediante algoritmos de MD

Para finalizar un caso de estudio en medios sociales

ASU Rastreador de coordinación Arizona State University

1. Recoger



Para finalizar un caso de estudio en medios sociales

ASU Rastreador de coordinación Arizona State University

2. Analizar y visualizar

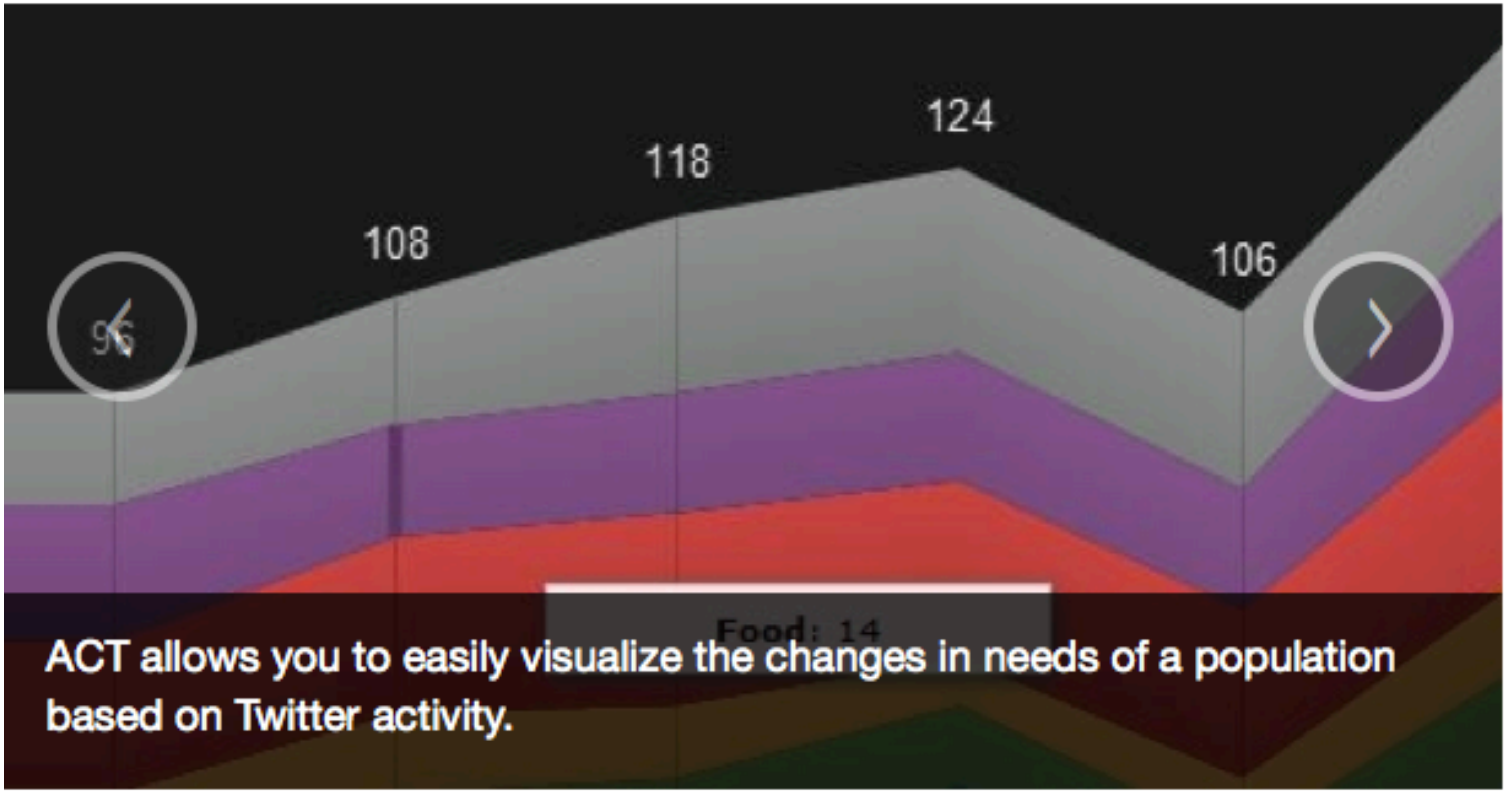
Para finalizar un caso de estudio en medios sociales

ASU Rastreador de coordinación Arizona State University

3. Responder

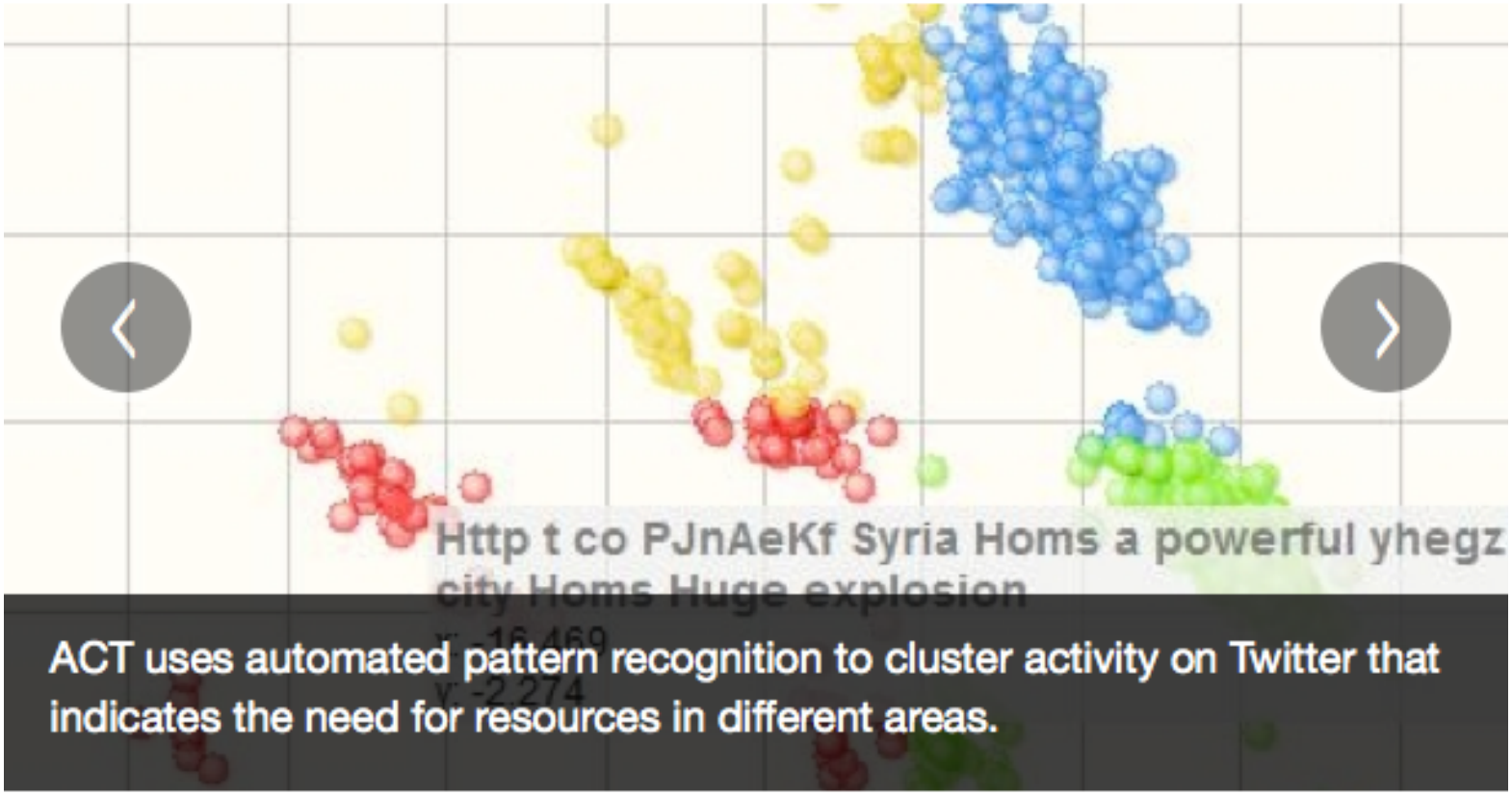
Para finalizar un caso de estudio en medios sociales

ASU Rastreador de coordinación Arizona State University



Para finalizar un caso de estudio en medios sociales

ASU Rastreador de coordinación Arizona State University



Bibliografía

- C.M Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics). 2006.
- G. James, D. Witten, T. HastieAn Introduction to Statistical Learning with Applications in R. Springer. Springer. 2013.
- A.K. Jain, R.C. Dubes. Algorithms for Clustering Data, Prentice Hall, 1988.
- L. Kaufman, P.J. Rousseeuw. Finding Groups in Data: An Introduction to Cluster Analysis, 1990.
- C. Zhang, S. Zhang. Association Rule Mining. Models and Algorithms. LNAI 2307. Springer 2002
- http://borgelt.net/fpm.html
- http://cs.bme.hu/~bodon/en/apriori/
- http://fimi.ua.ac.be/