



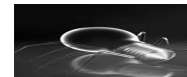
Ensayo para organizar estructuras  
heterogéneas de datos en el área de  
Marketing & Social Research

ALEPH ZERO  
marketing & social research  
data & text mining  
information representation & visualization

# Planteo y requerimientos

## **Necesidad de buscar alguna estructuración adecuada de los datos, con el objeto de:**

- Buscar una armazón “lógica” (una idea, un concepto, un modelo, una hipótesis)
- Que sea plausible o que tenga sentido para el dominio
- Que permita describir, explicar, acotar o inferir cosas
- Administrar los datos como un volumen manejable (reducción)
- Poder “verlos” (visualización)
- Poder usarlos rápidamente (in real time)
- .....Y al menor costo (?)



# Racional

- ❑ Pasar de un conjunto de objetos con múltiples características a una representación explorable con recursos de DM, AID, KDD, Visualización, etc.
- ❑ Y obtener un resultado que satisfaga los requerimientos anteriores

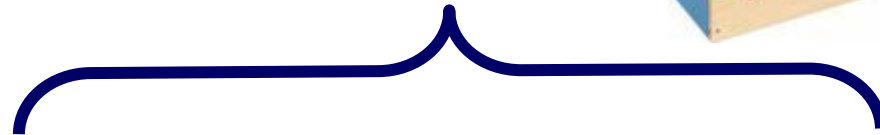
**OBJETOS**

**ETIQUETADO**

**REPRESENTACION**



**DM Y KDD**



**CLASIFICACION**

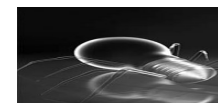
**MAPAS**

Aleph Zero



## Casos y objetivos

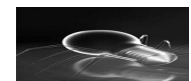
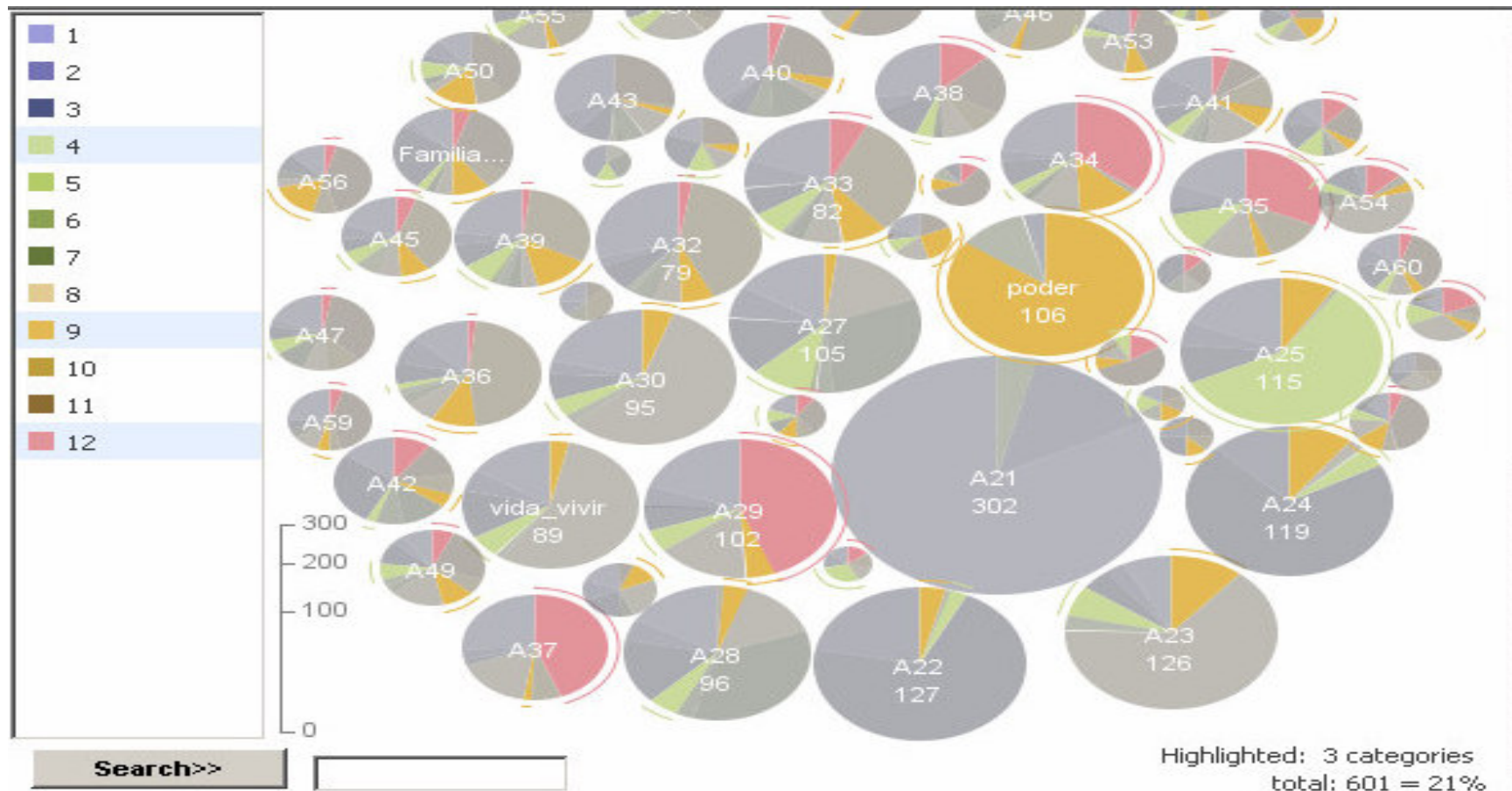
<b>Caso</b>	<b>Objetivo</b>
<ul style="list-style-type: none"><li>❑ Discurso Suicida</li></ul>	<ul style="list-style-type: none"><li>❑ Ayudar a reconocer características y tipologías de discurso en una terapia de prevención</li></ul>
<ul style="list-style-type: none"><li>❑ Monitoreo de “humor”</li></ul>	<ul style="list-style-type: none"><li>❑ Clasificar mensajes por la orientación respecto a un tema</li></ul>
<ul style="list-style-type: none"><li>❑ Clasificación de objetos heterogéneos</li></ul>	<ul style="list-style-type: none"><li>❑ Describir la colección a través de patrones con sentido</li></ul>





# Mapeo del discurso suicida

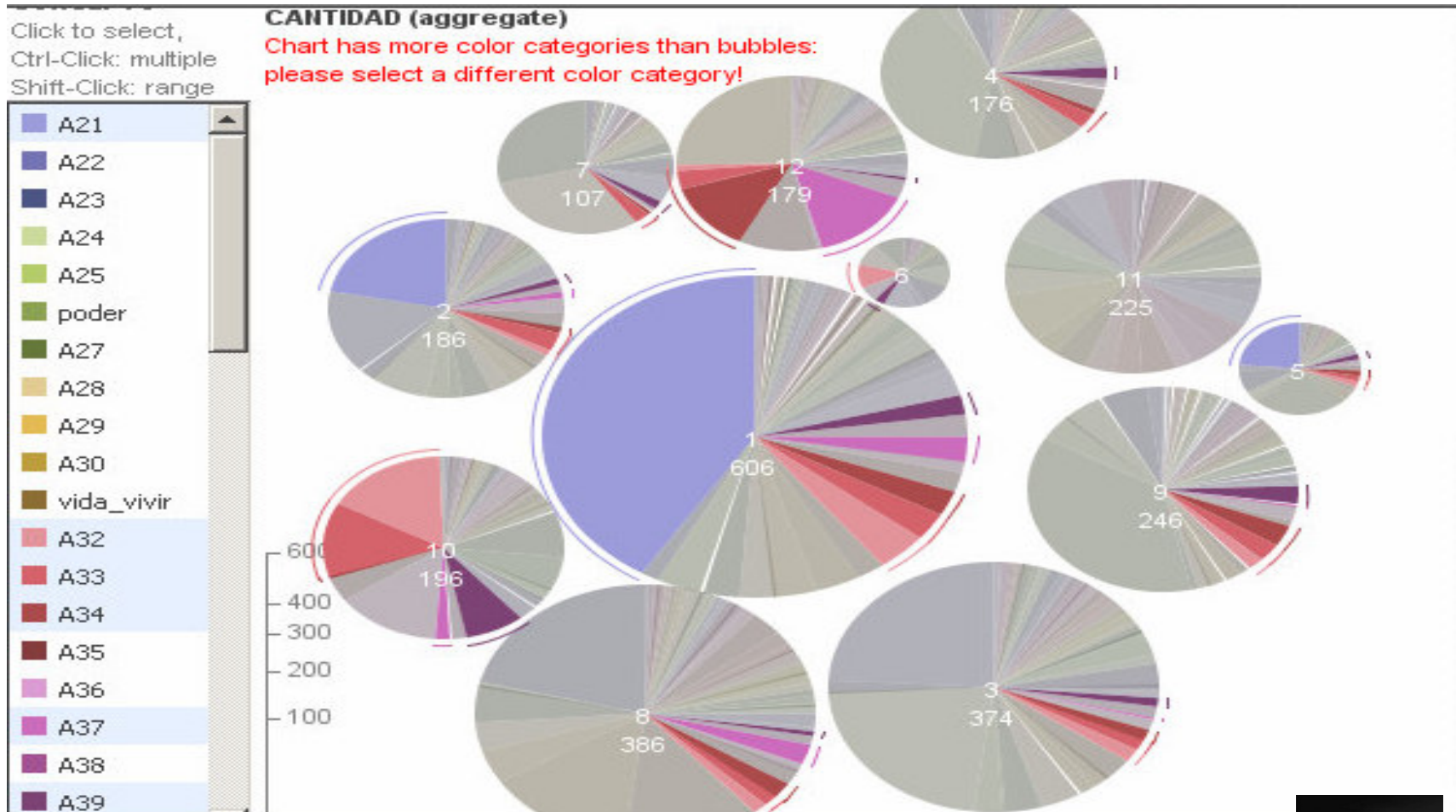
## Importancia de un grupo de nodos en cada concepto



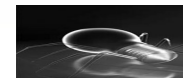
Aleph Zero

# Mapeo del discurso suicida

## Importancia de conceptos en cada nodo

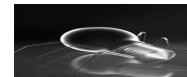
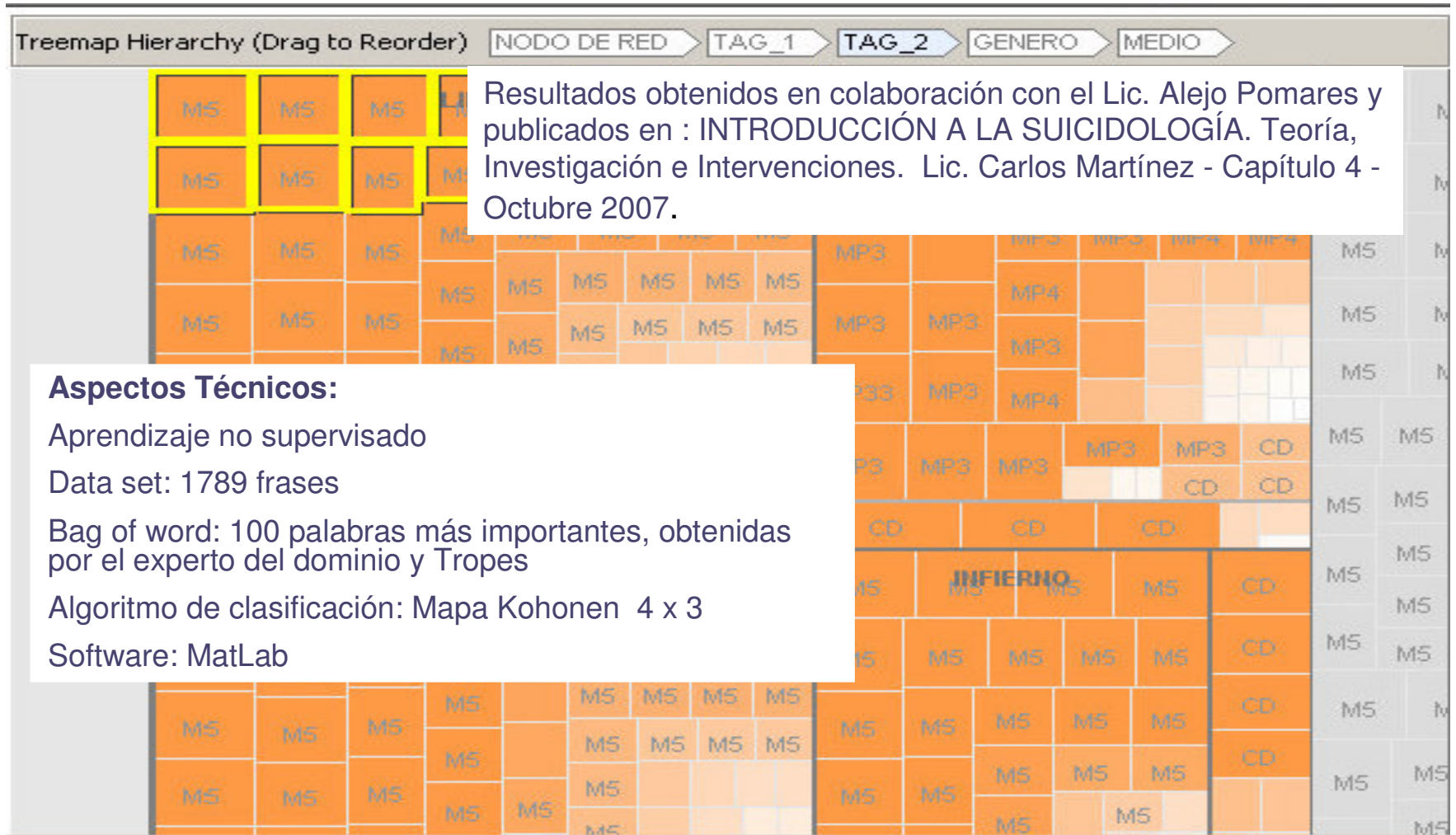


Aleph Zero



# Mapeo del discurso suicida

## Treemap



# Monitoreo de humor Universo de discurso

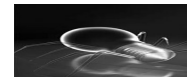
## Definición

- Es el conjunto de manifestaciones (discursos o mensajes) sobre una marca, una persona, un hecho, un mercado, una institución, etc, que circula en los distintos medios.

## Hechos

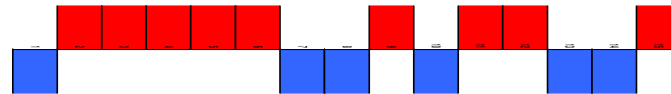
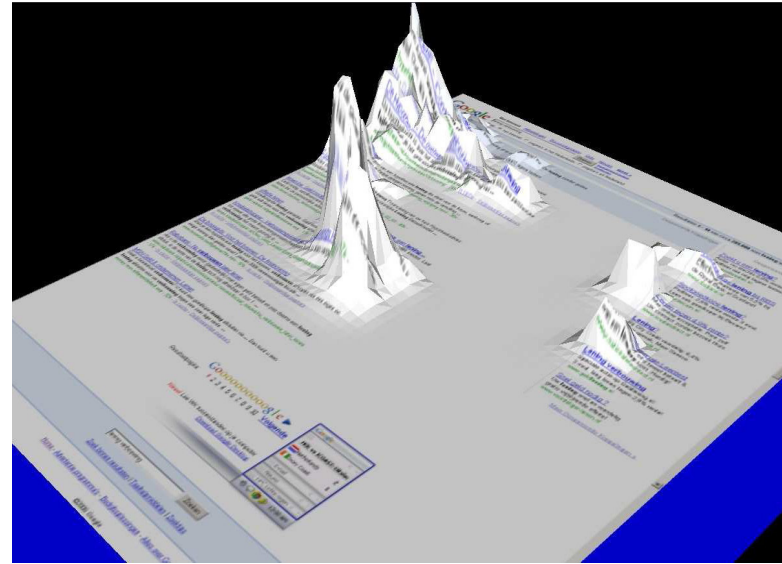
- El conjunto de manifestaciones sobre cualquier tema ha crecido en forma explosiva en la última década
- Su tratamiento inteligente, requiere de recursos sistematizados y automatizados
- Hay orientación de los discursos, dependiendo del jugador que lo emita, lo cual da una dimensión cualitativa al discurso..
- La situación actual se puede esquematizar de la siguiente forma:

Jugador	Control de la orientación respecto a una marca o idea
Anunciante	si
Aliados (distribuidor, supermercado, agencia)	si
Competencia	no
Medios	no
Instituciones	no
Usuarios/Consumidores	no



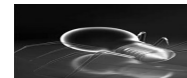
# Monitoreo de humor

## Manejo de Crisis



LISTA 1....15 azul + rojo -

Aleph Zero



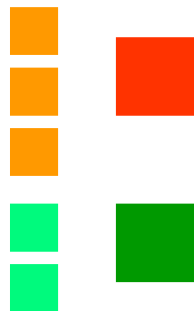
## Monitoreo de “humor”

- Se trata de identificar alguna “estructura” común en los mensajes emitidos. Para luego establecer:
- una orientación positiva o negativa respecto a un tema (línea de investigación de Xue Bai y otros)

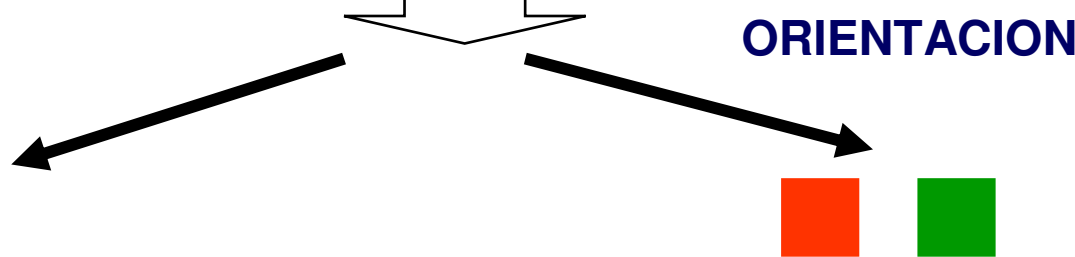
ó

- una orientación y fuerza (Dureza de la opinión – línea de investigación de Fabrizio Sebastiani y Andrea Esuli)

**ORIENTACION  
+FUERZA**



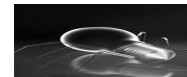
Mensajes 1....k



**ORIENTACION**



Aleph Zero



# Monitoreo de "humor"

COLECCION DE MENSAJES

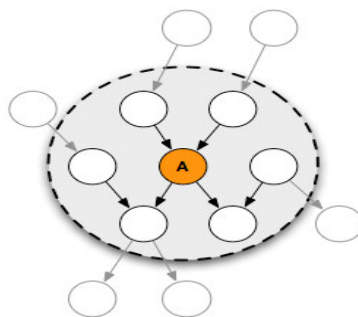


DATA SET CON MENSAJES CLASIFICADOS POR UN EXPERTO O JUECES

REPRESENTACION

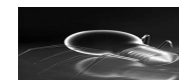
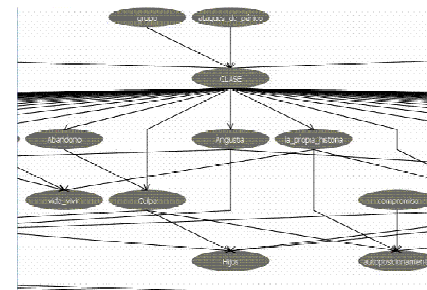
011011000110101\_1

000110101011001\_0



MARKOV BLANKET CLASSIFIER

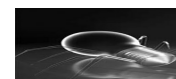
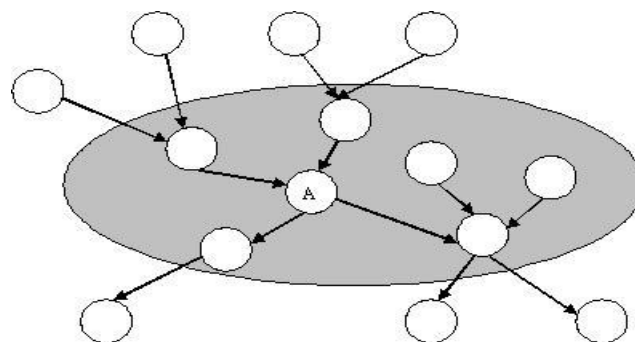
RED



Aleph Zero

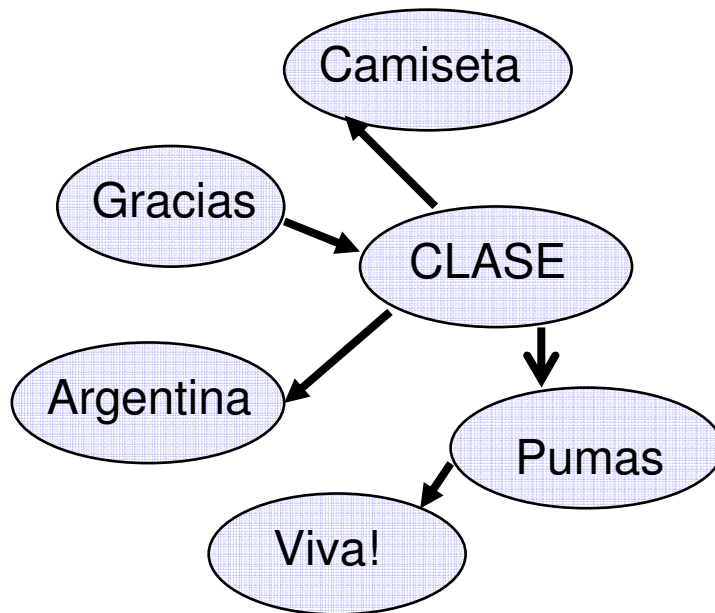
## Redes Bayesianas y Markov Blanket

- Una red bayesiana es un grafo acíclico. Cada nodo representa una variable y cada arco una dependencia probabilística, dada por la probabilidad condicional de cada variable dados sus padres.
- Se cumple la condición de Markov si cada nodo es independiente de sus no padres y de sus no hijos y es condicional a sus padres
- El Markov Blanket para un nodo A es el conjunto de nodos que son padres de A, el conjunto de nodos que son hijos de A y el conjunto de nodos que son padres de los descendientes de A





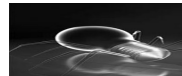
# Construyendo la red Grafo + Algoritmo



✓ NP Hard

✓ Orientación del grafo

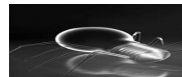
✓ Conectividad de los nodos



# Monitoreo de humor

## Diseño

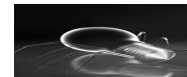
- Aprendizaje supervisado
- Data Set: 1777 documentos, convertidos en un documento\_vector con 60 variables binarias y una variable de clase (0,1)
- Clasificadores/algoritmos: se utilizó Weka, con los siguientes algoritmos:
  - Bayes.NaiveBayesSimple
  - Bayes.NaiveBayes
  - TabuSearch/Markov Blanket
  - Simulated Annealing/Markov Blanket
  - Trees.NBTree
- Entrenamiento: 66% del dataset, el restante 34% fue utilizado para validación



# Monitoreo de humor

## Desempeño de distintos clasificadores

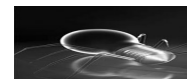
Clasificador	Tiempo (seg)	Ejemplos correctamente clasificados %
Bayes.NaiveBayesSimple	0,0	95,04
Bayes.NaiveBayes	0,0	95,04
TS/MB_60	4,5	90,08
TS/MB_30	4,6	90,08
SA/MB	16,0	89,91
Trees.NBTree	13,4	89,58



## Monitoreo de humor

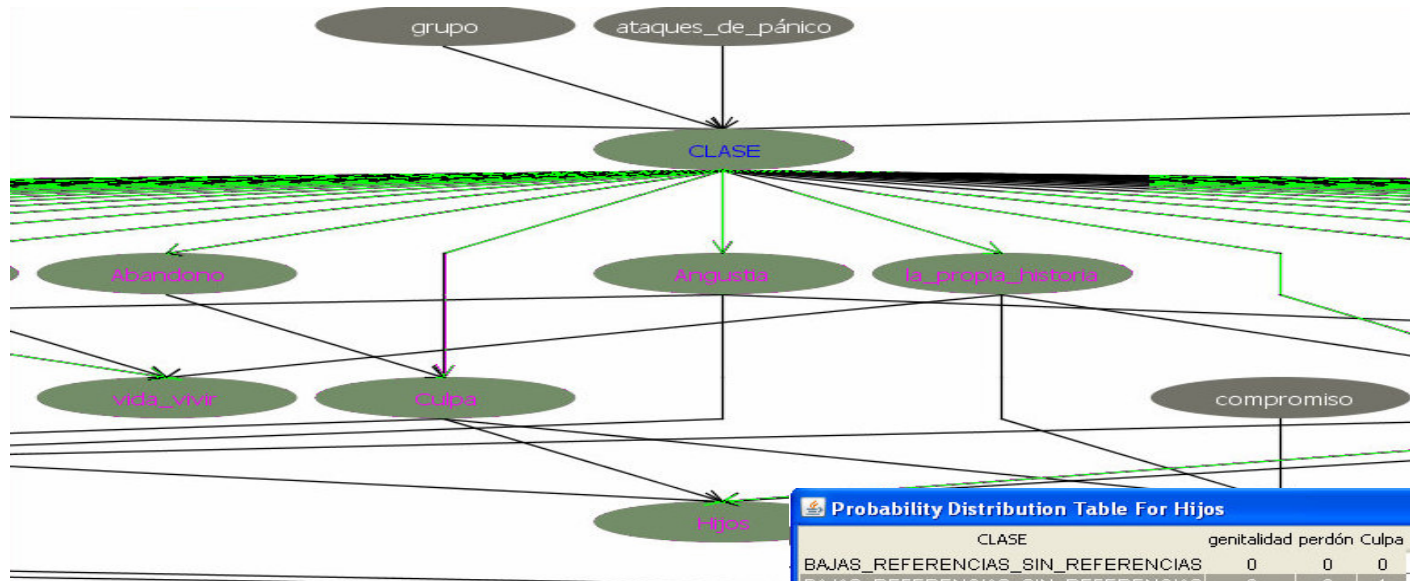
### Comparación

- ❑ NB clasifica mejor, pero....
- ❑ MB y Trees.NBTree permite la reducción de atributos (en algunos proyectos resulta una cualidad importante)
- ❑ MB y Trees.NBTree tienen un output visual útil para determinar aspectos cualitativos de la estructura que se estudia.



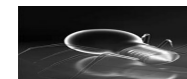
# Monitoreo de humor

## Visualización de TS/MB

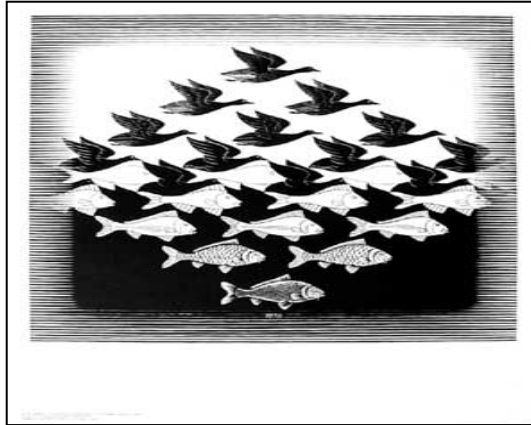


Probability Distribution Table For Hijos

CLASE	genitalidad	perdón	Culpa	0	1
BAJAS_REFERENCIAS_SIN_REFERENCIAS	0	0	0	0,964	0,036
BAJAS_REFERENCIAS_SIN_REFERENCIAS	0	0	1	0,958	0,042
BAJAS_REFERENCIAS_SIN_REFERENCIAS	0	1	0	0,625	0,375
BAJAS_REFERENCIAS_SIN_REFERENCIAS	0	1	1	0,5	0,5
BAJAS_REFERENCIAS_SIN_REFERENCIAS	1	0	0	0,5	0,5
BAJAS_REFERENCIAS_SIN_REFERENCIAS	1	0	1	0,5	0,5
BAJAS_REFERENCIAS_SIN_REFERENCIAS	1	1	0	0,5	0,5
BAJAS_REFERENCIAS_SIN_REFERENCIAS	1	1	1	0,5	0,5
ALTAS_REFERENCIAS	0	0	0	0,863	0,137
ALTAS_REFERENCIAS	0	0	1	0,962	0,038
ALTAS_REFERENCIAS	0	1	0	0,95	0,05
ALTAS_REFERENCIAS	0	1	1	0,5	0,5
ALTAS_REFERENCIAS	1	0	0	0,9	0,1
ALTAS_REFERENCIAS	1	0	1	0,25	0,75
ALTAS_REFERENCIAS	1	1	0	0,5	0,5
ALTAS_REFERENCIAS	1	1	1	0,5	0,5



# Clasificación de objetos multimedia



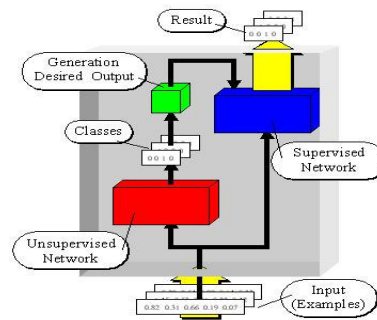
**COLECCION DE  
OBJETOS  
REPRESENTACION**

01101100011010111  
00011010101100101

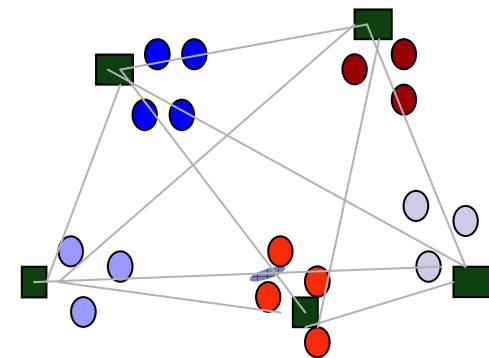


**TAGGING**

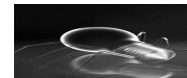
**CLASIFICACION**



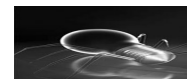
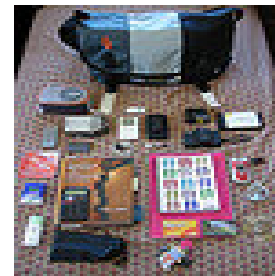
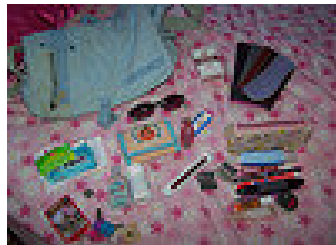
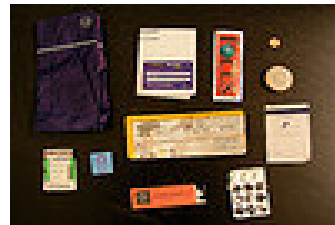
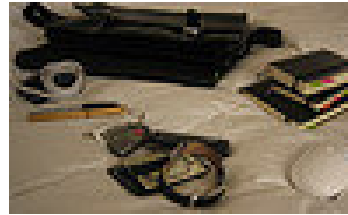
**RED**



**Aleph Zero**



# Colección de objetos



## Tagging de rasgos



Movistar: 1era  
línea

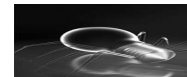
Billetera

La bolsa de las llaves..nunca las  
encuentro

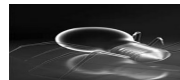
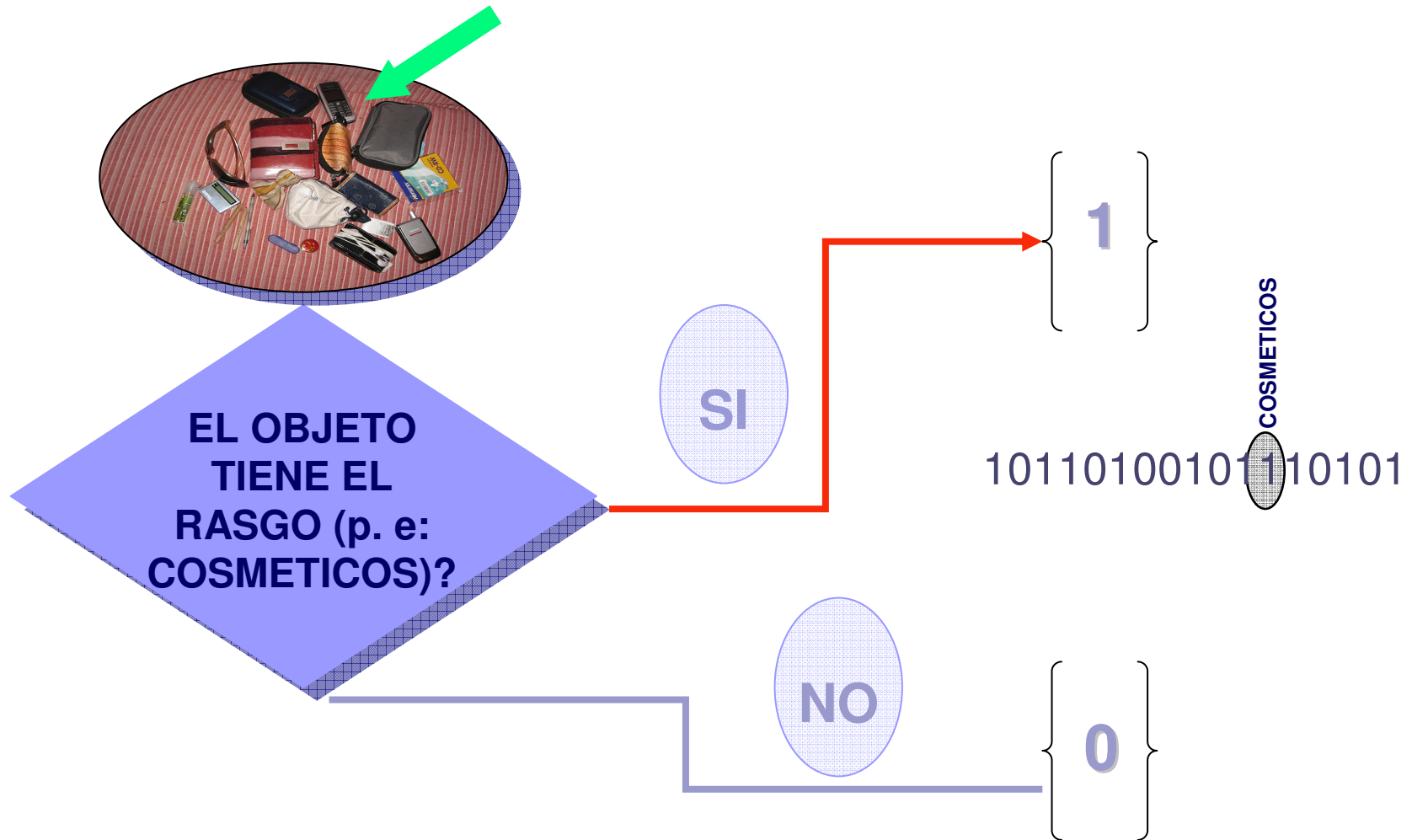
Tres lápices

CTI Nokia: 2da  
línea

MP4 creative



# Codificación

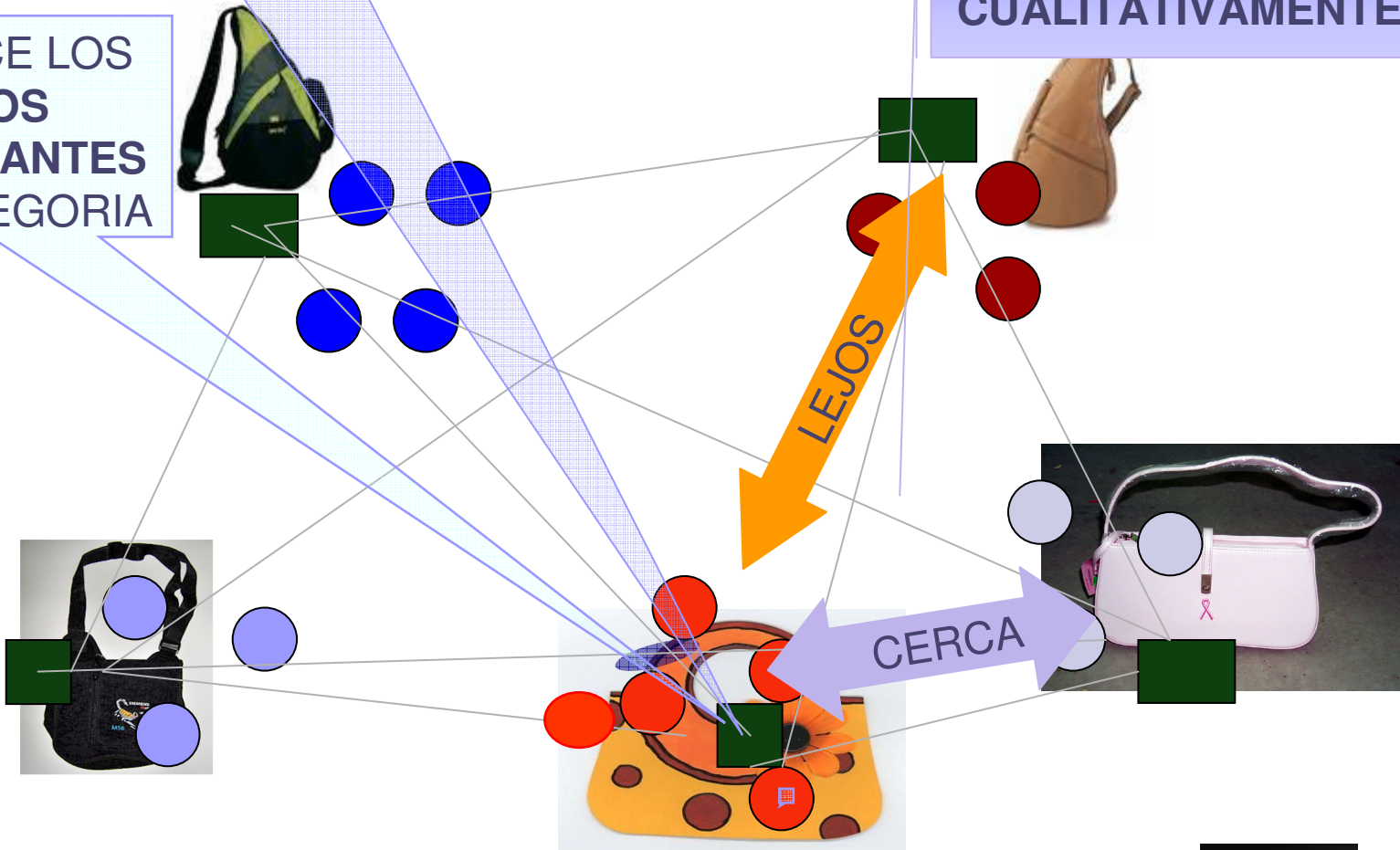


DETERMINA LA  
IMPORTANCIA DE LA  
CATEGORIA

## Que brinda una red SOM ?

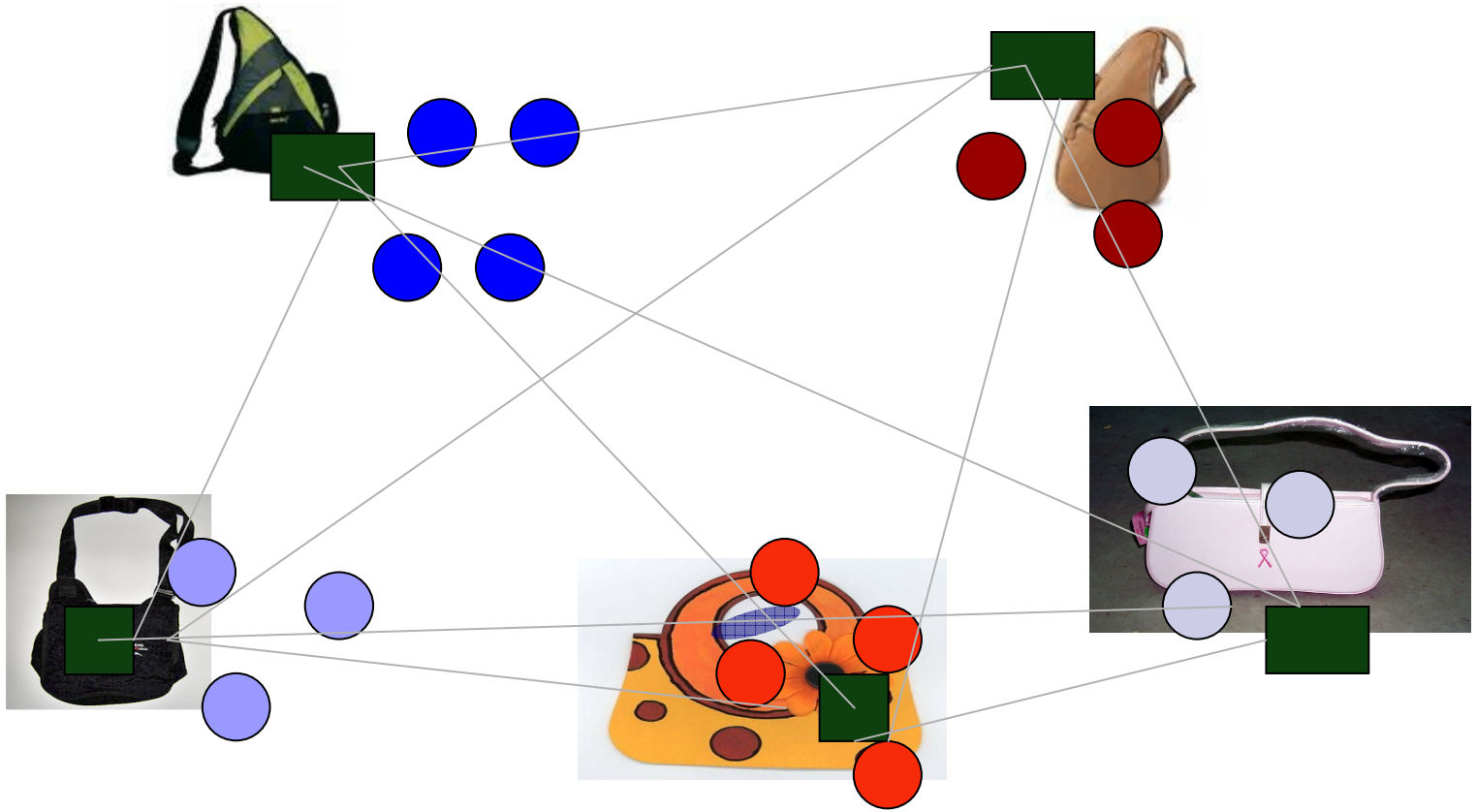
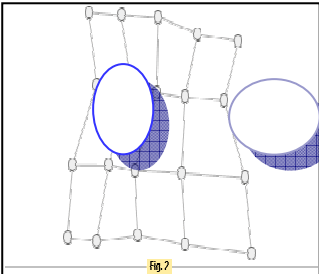
ESTABLECE LOS  
RASGOS  
PREDOMINANTES  
EN LA CATEGORIA

CARACTERIZA  
CUALITATIVAMENTE

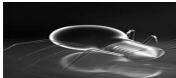


Aleph Zero

# Aprendizaje y consulta

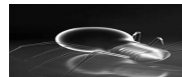


Aleph Zero



## Conclusiones

- ❑ Data Mining y KDD proveen recursos útiles en la etapa mas importante del Marketing Research, el análisis
- ❑ Posibilita obtener visiones de los datos que no se pueden efectuar mediante los métodos clásicos
- ❑ Genera diseños originales para capturar, representar, visualizar e interpretar los datos
- ❑ Permite fundamentar y poner rigor a ciertas prácticas, así como corregir sesgos
- ❑ Es una interfaz adecuada al “nuevo” marketing



# El ojo humano

## Lingüistas

Tras la cerrada ovación que puso término a la sesión plenaria del congreso internacional de lingüística y afines, la hermosa taquígrafa recogió sus lápices y sus papeles y se dirigió a la salida abriéndose paso entre un centenar de lingüistas, filólogos, eniólogos, críticos estructuralistas y deconstruccionalistas, todos los cuales siguieron su garboso desplazamiento con una admiración rayana en la glosemática. De pronto, las diversas acuñaciones cerebrales adquirieron vigencia fónica: ¡Qué sintagma, qué polisemia, qué significante, qué diacronía, qué centrar ceterorum, qué zungespitze, qué morfema! La hermosa taquígrafa desfiló impertérrita y adusta entre aquella selva de fonemas. Solo se la vio sonreír, halagada y tal vez, vulnerable, cuando el joven ordenanza, antes de abrirle la puerta, murmuró casi en su oído: ¡Cosita linda!

**Mario Benedetti**

