



# **From Cliques to Equilibria: The Dominant-Set Framework for Pairwise Data Clustering**

**Marcello Pelillo**

*Department of Computer Science*

*Ca' Foscari University, Venice*

---

Joint work with M. Pavan, A. Torsello and S. Rota Bulò



# Lecture's Outline

- Dominant sets and their characterization
- Finding dominant sets using evolutionary game dynamics
- Experiments on image segmentation (and extensions)
- Dominant sets and hierarchical clustering
- Dealing with arbitrary affinities: Dominant sets as (evolutionary) game equilibria

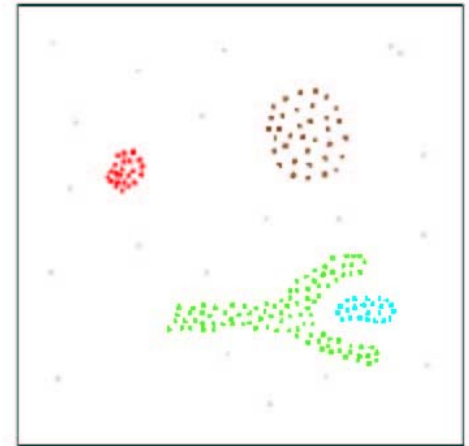
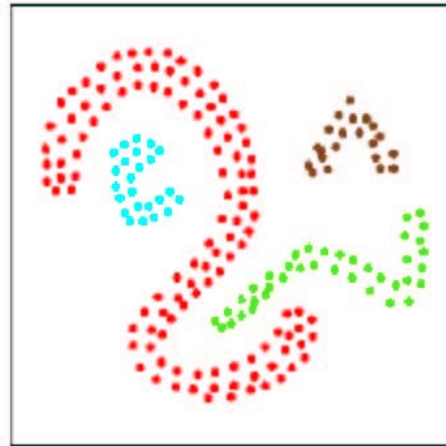
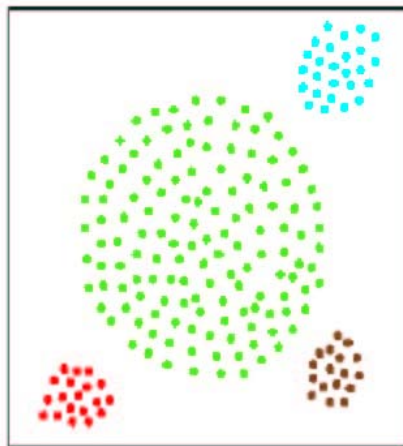


# The (Pairwise) Clustering Problem

## Given:

- a set of  $n$  “objects”
- an  $n \times n$  matrix of pairwise similarities

**Goal:** Partition the input objects into maximally homogeneous groups (i.e., **clusters**).





# Applications

Clustering problems abound in many areas of computer science and engineering.

A short list of applications domains:

- Image processing and computer vision
- Computational biology and bioinformatics
- Information retrieval
- Data mining
- Signal processing
- ...



# What is a Cluster?

No universally accepted definition of a “cluster”.

Informally, a cluster should satisfy two criteria:

**Internal criterion:** all objects *inside* a cluster should be highly similar to each other.

**External criterion:** all objects *outside* a cluster should be highly dissimilar to the ones inside.



# Clustering as a Graph-Theoretic Problem

We represent the data to be clustered as an undirected edge-weighted graph with no self-loops  $G = (V, E, w)$ , where  $V = \{1, \dots, n\}$  is the vertex set,  $E \subseteq V \times V$  is the edge set, and  $w : E \rightarrow \mathbb{R}_+^*$  is the (positive) weight function.

We represent the graph  $G$  with the corresponding weighted adjacency (or similarity) matrix, which is the  $n \times n$  symmetric matrix  $A = (a_{ij})$  defined as:

$$a_{ij} = \begin{cases} w(i, j) , & \text{if } (i, j) \in E \\ 0 , & \text{otherwise.} \end{cases}$$



## The Binary Case

Suppose the similarity matrix is a binary (0/1) matrix.

In this case, the notion of a cluster coincide with that of a *maximal clique*.

Given an unweighted undirected graph  $G=(V,E)$ :

A *clique* is a subset of mutually adjacent vertices

A *maximal clique* is a clique that is not contained in a larger one

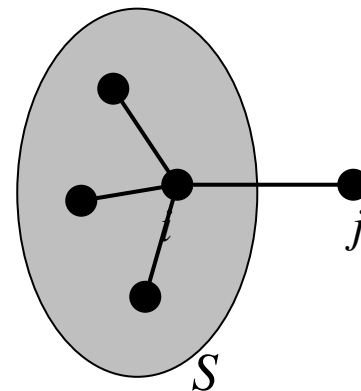
How to generalize the notion of a maximal clique  
to weighted graphs?



## Basic Definitions

Let  $S \subseteq V$  be a non-empty subset of vertices and  $i \in S$ . The (average) weighted degree of  $i$  w.r.t.  $S$  is defined as:

$$\text{awdeg}_S(i) = \frac{1}{|S|} \sum_{j \in S} a_{ij} .$$



Moreover, if  $j \notin S$  we define:

$$\phi_S(i, j) = a_{ij} - \text{awdeg}_S(i) .$$

Intuitively,  $\phi_S(i, j)$  measures the similarity between nodes  $j$  and  $i$ , with respect to the average similarity between node  $i$  and its neighbors in  $S$ .





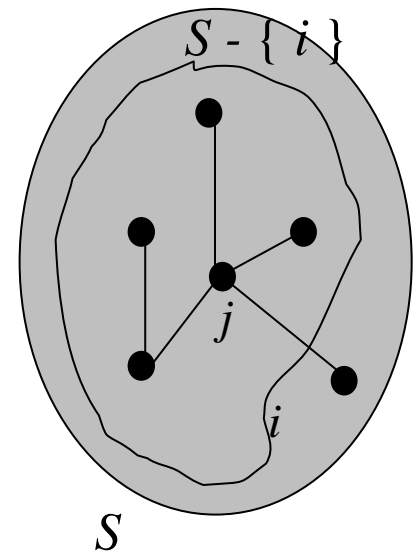
## Assigning Node Weights / 1

Let  $S \subseteq V$  be a non-empty subset of vertices and  $i \in S$ . The **weight** of  $i$  w.r.t.  $S$  is

$$w_S(i) = \begin{cases} 1, & \text{if } |S| = 1 \\ \sum_{j \in S \setminus \{i\}} \phi_{S \setminus \{i\}}(j, i) w_{S \setminus \{i\}}(j), & \text{otherwise.} \end{cases}$$

Moreover, the **total weight** of  $S$  is defined to be:

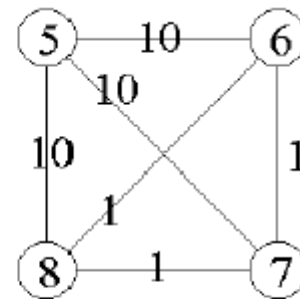
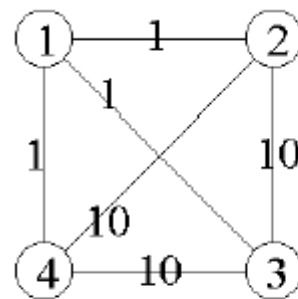
$$W(S) = \sum_{i \in S} w_S(i).$$





## Assigning Node Weights / 2

Intuitively,  $w_S(i)$  gives us a measure of the overall similarity between vertex  $i$  and the vertices of  $S \setminus \{i\}$  with respect to the overall similarity among the vertices in  $S \setminus \{i\}$ .



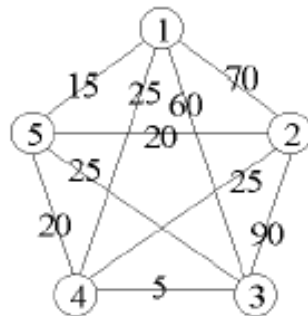
$w_{\{1,2,3,4\}}(1) < 0$  and  $w_{\{5,6,7,8\}}(5) > 0$ .



# Dominant Sets

A non-empty subset of vertices  $S \subseteq V$  such that  $W(T) > 0$  for any non-empty  $T \subseteq S$ , is said to be **dominant** if:

1.  $w_S(i) > 0$ , for all  $i \in S$  (internal homogeneity)
2.  $w_{S \cup \{i\}}(i) < 0$ , for all  $i \notin S$  (external inhomogeneity)



The set  $\{1, 2, 3\}$  is dominant.

Dominant sets  $\equiv$  clusters

For 0/1 matrices: dominant sets  $\equiv$  (strictly) maximal cliques



# From Dominant Sets to Local Optima (and Back) / 1

Given an edge-weighted graph  $G = (V, E, w)$  and its weighted adjacency matrix  $A$ , consider the following **Standard Quadratic Program (StQP)**:

$$\begin{array}{ll} \text{maximize} & f(\mathbf{x}) = \mathbf{x}' A \mathbf{x} \\ \text{subject to} & \mathbf{x} \in \Delta \end{array}$$

where

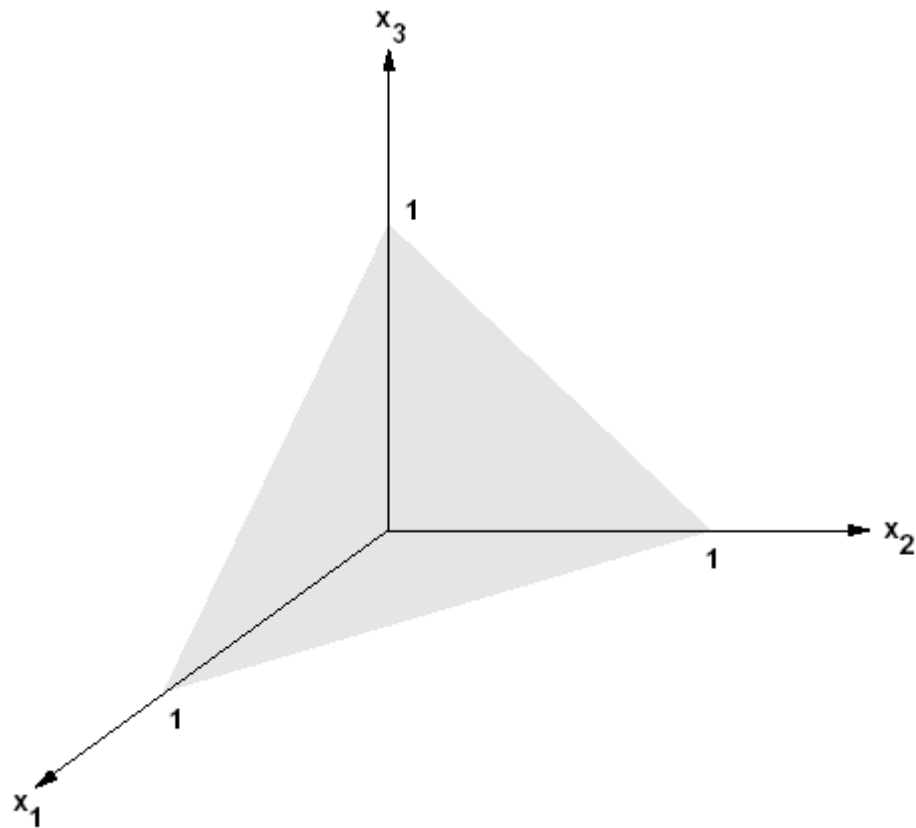
$$\Delta = \left\{ \mathbf{x} \in \mathbf{R}^n : \mathbf{e}' \mathbf{x} = 1 \text{ and } x_i \geq 0 \ \forall i \in V \right\}$$

is the standard simplex of  $\mathbf{R}^n$  and  $\mathbf{e} = (1, 1, \dots, 1)'$ .

**Note.** Other approaches to clustering lead to similar quadratic optimization problems (e.g., Sarkar and Boyer, 1998).



# The Standard Simplex (when $n = 3$ )





## From Dominant Sets to Local Optima (and Back) / 2

**Theorem** *If  $S$  is a dominant subset of vertices, then its weighted characteristics vector  $\mathbf{x}^S$ , defined as*

$$x_i^S = \begin{cases} \frac{w_S(i)}{W(S)}, & \text{if } i \in S \\ 0, & \text{otherwise} \end{cases}$$

*is a strict local maximizer of  $f$  in  $\Delta$ .*

*Conversely, if  $\mathbf{x}^*$  is a strict local maximizer of  $f$  in  $\Delta$  then its support*

$$\sigma = \sigma(\mathbf{x}^*) \doteq \{i \in V : x_i^* \neq 0\}$$

*is a dominant set, provided that  $w_{\sigma \cup \{i\}}(i) \neq 0$  for all  $i \notin \sigma$ .*

Generalization of Motzkin-Straus theorem from graph theory



## Lecture's Outline

- Dominant sets and their characterization
- Finding dominant sets using evolutionary game dynamics
- Experiments on image segmentation (and extensions)
- Dominant sets and hierarchical clustering
- Dealing with arbitrary affinities: Dominant sets as (evolutionary) game equilibria



# Replicator Equations

Developed in evolutionary game theory to model the evolution of behavior in animal conflicts (Hofbauer & Sigmund, 1998).

Let  $W = (w_{ij})$  be a non-negative real-valued  $n \times n$  matrix.

**Continuous-time version:**

$$\frac{d}{dt}x_i(t) = x_i(t) \left[ (W\mathbf{x}(t))_i - \mathbf{x}(t)'W\mathbf{x}(t) \right]$$

**Discrete-time version:**

$$x_i(t+1) = x_i(t) \frac{(W\mathbf{x}(t))_i}{\mathbf{x}(t)'W\mathbf{x}(t)}$$

$\Delta$  is invariant under both dynamics, and they have the same stationary points.





# The Fundamental Theorem of Natural Selection

If  $W = W'$ , then the function

$$F(\mathbf{x}) = \mathbf{x}'W\mathbf{x}$$

is strictly increasing along any non-constant trajectory of both continuous-time and discrete-time replicator dynamics.

In other words,  $\forall t \geq 0$ :

$$\frac{d}{dt}F(\mathbf{x}(t)) > 0$$

for the continuous-time dynamics, and

$$F(\mathbf{x}(t+1)) > F(\mathbf{x}(t))$$

for the discrete-time dynamics, unless  $\mathbf{x}(t)$  is a stationary point.



## Grouping by Replicator Equations

Let  $A$  denote the weighted adjacency matrix of the similarity graph.

Let

$$W = A \quad (= W' \geq 0) .$$

The replicator systems, starting from an arbitrary initial state, will eventually converge to a maximizer of the function  $f(x) = x'Ax$ , over the simplex.

This will correspond to a dominant set in the graph, and hence to a cluster of vertices.



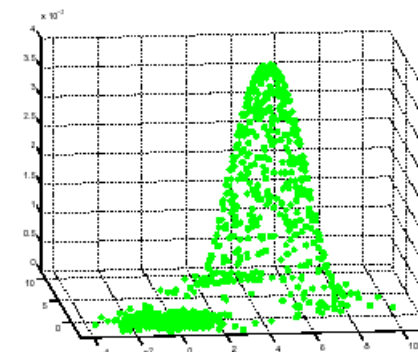
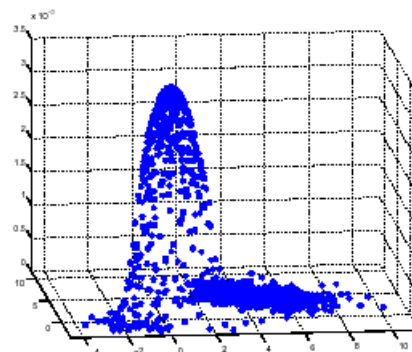
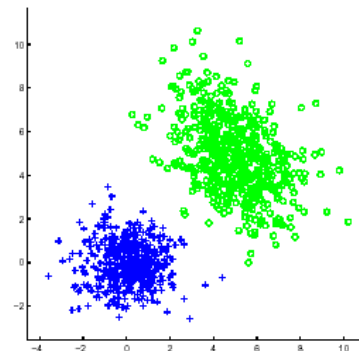
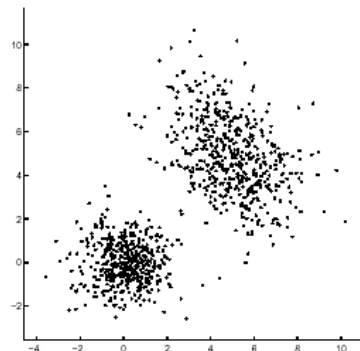
## A MATLAB Implementation

```
distance=inf;
while distance>epsilon
    old_x=x;
    x = x.*(A*x);
    x = x./sum(x);
    distance=pdist([x,old_x]');
end
```



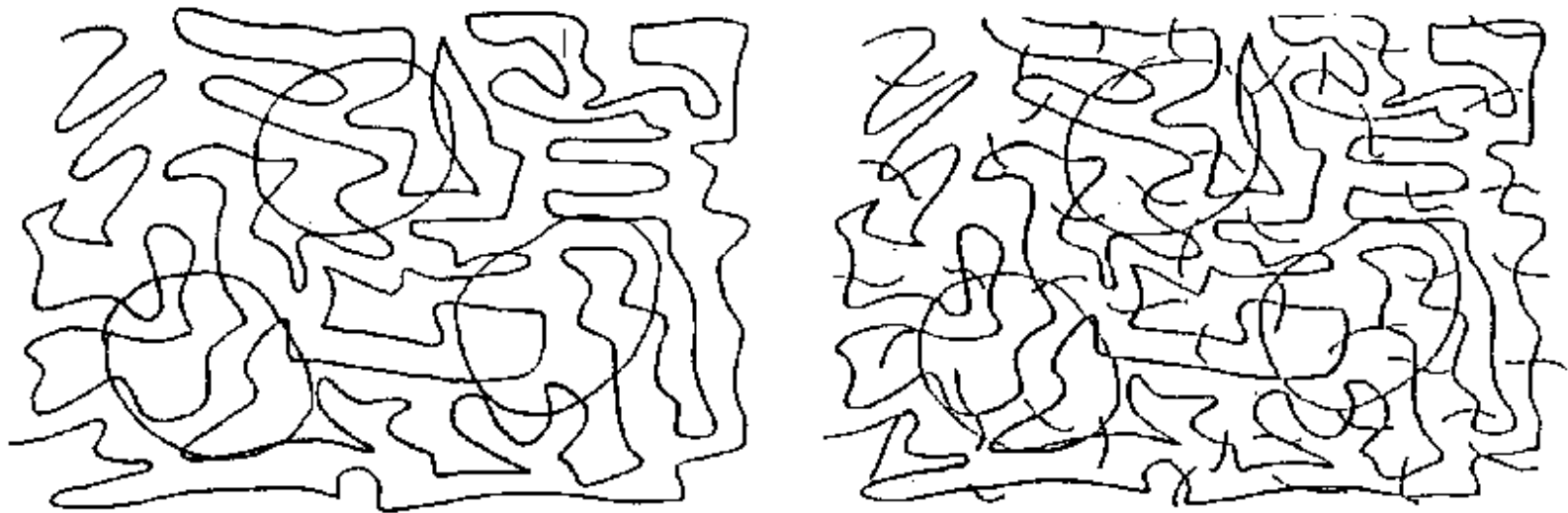
# Characteristic Vectors

**Note.** The components of the weighted characteristic vectors give us a measure of the participation of the corresponding vertices in the cluster, while the value of the objective function provides a measure of the cohesiveness of the cluster (*cfr.* Sarkar and Boyer, 1998).



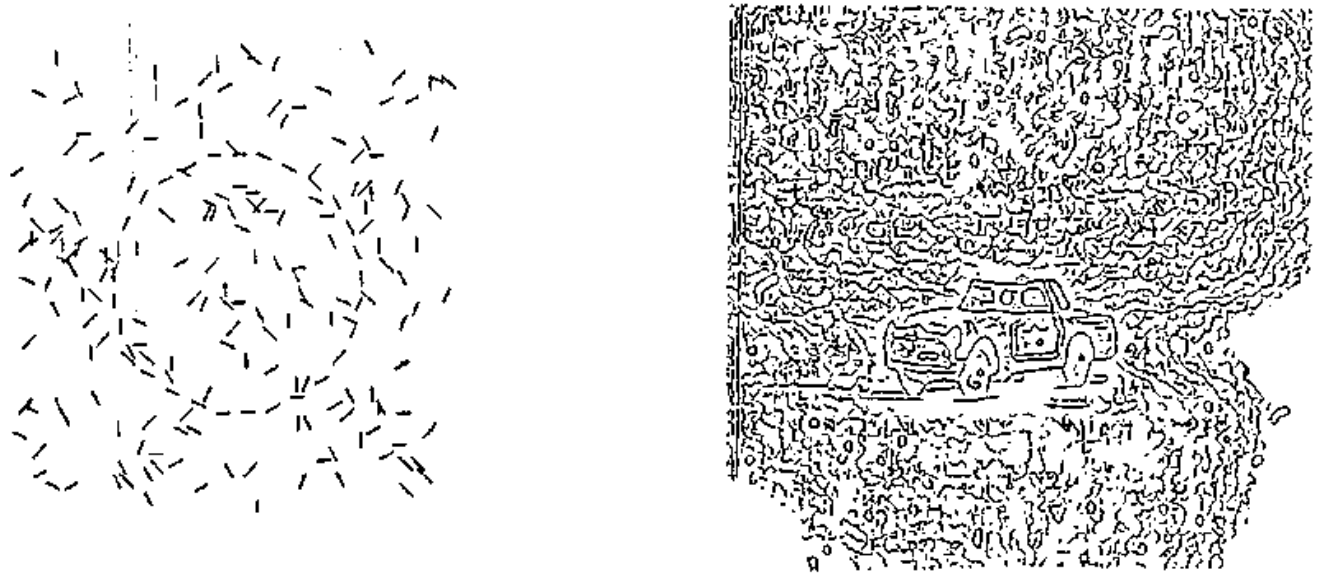


## Separating Structure for Clutter



*Figure 1a.* Three prominent blobs are perceived immediately and with little effort. Locally, the blobs are similar to the background contours. (adopted from Mahoney (1986))

*Figure 1b.* Intersections were added to illustrate that the blobs are not distinguished by virtue of their intersections with the background curves.

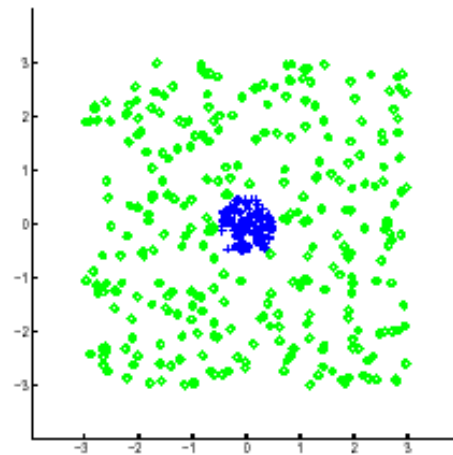
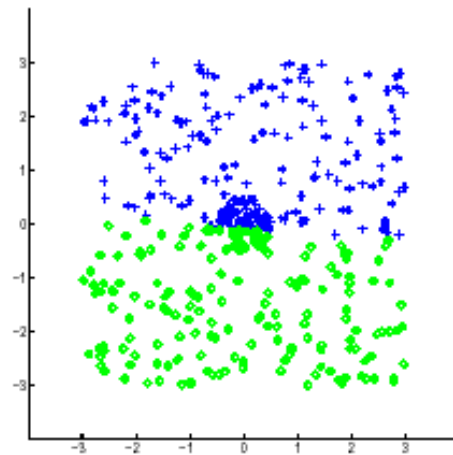
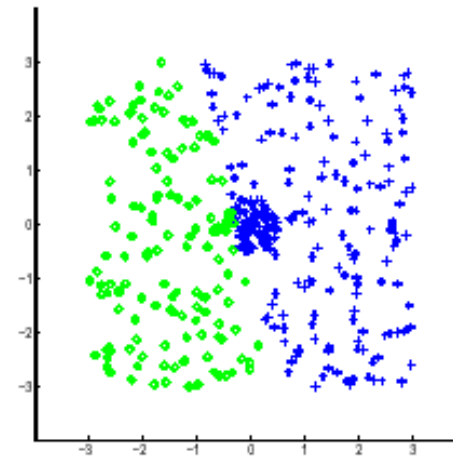
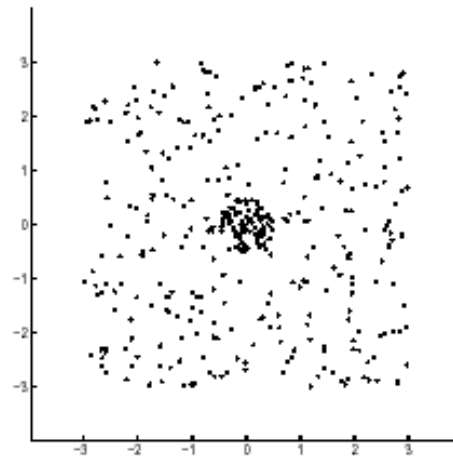


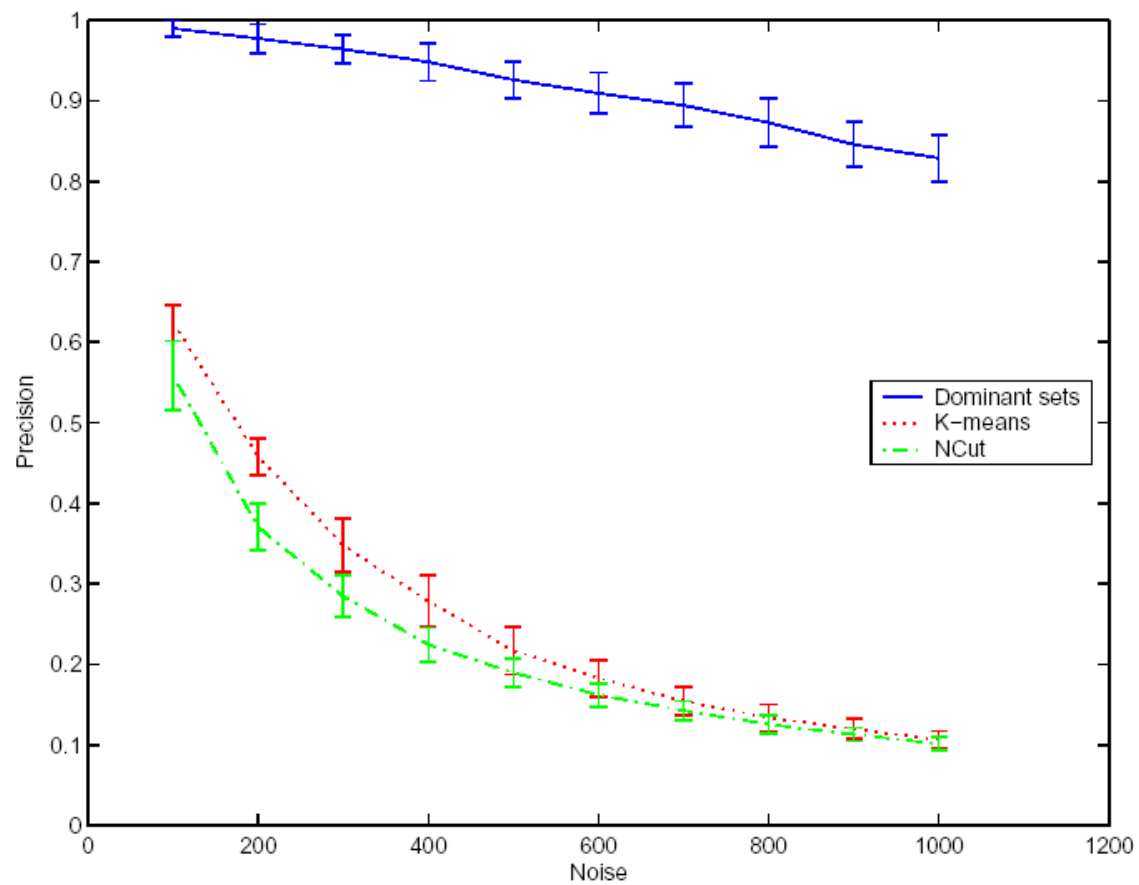
*Figure 2.* A circle in a background of 200 randomly placed and oriented segments. The circle is still perceived immediately although its contour is fragmented.

*Figure 3.* An edge image of a car in a cluttered background. Our attention is drawn immediately to the region of interest. It seems that the car need not be recognized to attract our attention. The car also remains salient when parallel lines and small blobs are removed, and when the less textured region surrounding parts of the car is filled in with more texture.



# Separating Structure from Clutter









# Lecture's Outline

- Dominant sets and their characterization
- Finding dominant sets using evolutionary game dynamics
- Experiments on image segmentation (and extensions)
- Dominant sets and hierarchical clustering
- Dealing with arbitrary affinities: Dominant sets as (evolutionary) game equilibria



# Image Segmentation

## Image segmentation problem:

Decompose a given image into *segments*, i.e. regions containing “similar” pixels.

First step in many  
computer vision problems



**Example:** Segments might be regions of the image depicting the same object.

**Semantics Problem:** *How should we infer objects from segments?*



# Image Segmentation

An image is represented as an edge-weighted undirected graph, where vertices correspond to individual pixels and the edge-weights reflect the “similarity” between pairs of vertices.

Our clustering algorithm basically consists of iteratively finding a dominant set in the graph using replicator dynamics and then removing it from the graph, until all vertices have been clustered.

On average, the algorithm took only a few seconds to converge, on a machine equipped with a 750 MHz Intel Pentium III.



## Experimental Setup

The similarity between pixels  $i$  and  $j$  was measured by:

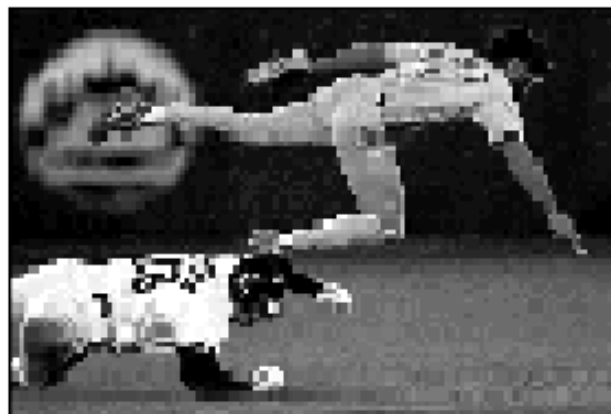
$$w(i, j) = \exp \left( \frac{-\|\mathbf{F}(i) - \mathbf{F}(j)\|_2^2}{\sigma^2} \right)$$

where  $\sigma$  is a positive real number which affects the decreasing rate of  $w$ , and:

- $\mathbf{F}(i) \equiv$  (normalized) intensity of pixel  $i$ , for **intensity segmentation**
- $\mathbf{F}(i) = [v, vs \sin(h), vs \cos(h)](i)$ , where  $h, s, v$  are the HSV values of pixel  $i$ , for **color segmentation**
- $\mathbf{F}(i) = [|I * f_1|, \dots, |I * f_k|](i)$  is a vector based on texture information at pixel  $i$ , the  $f_i$  being DOOG filters at various scales and orientations, for **texture segmentation**



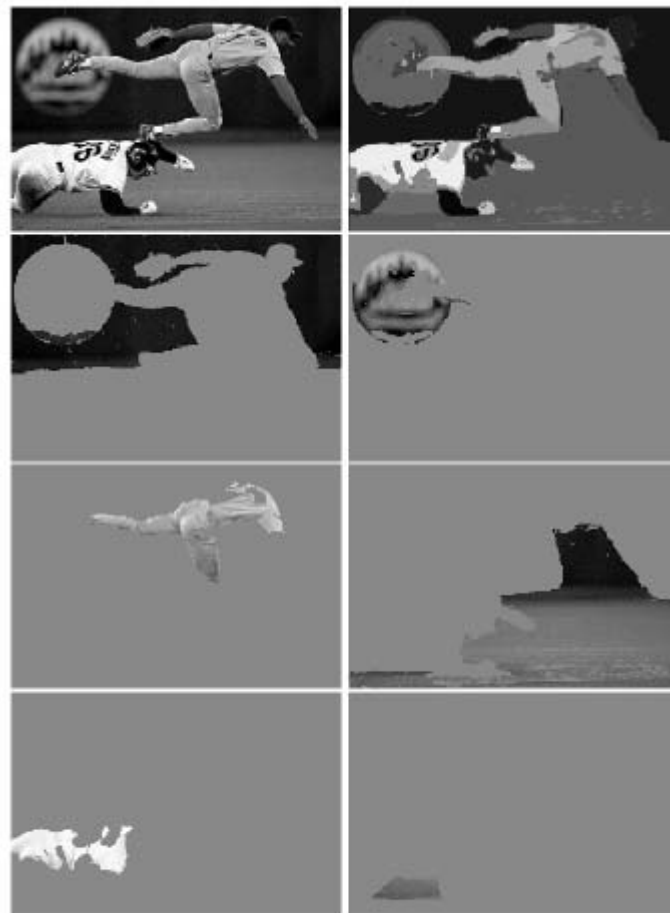
## Intensity Segmentation Results



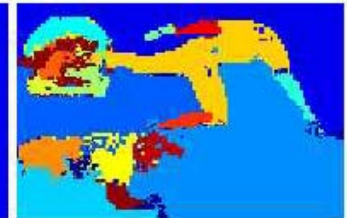
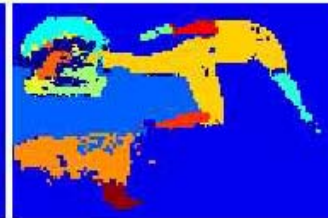
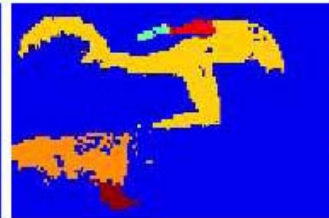
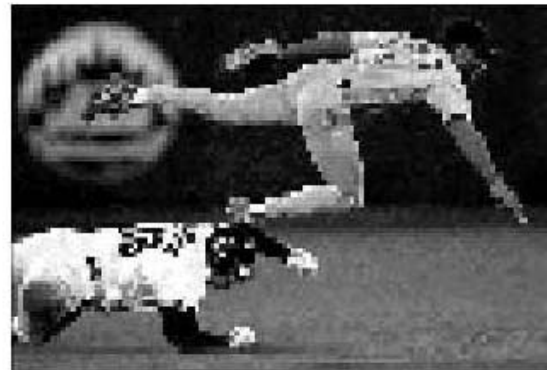
Dominant sets



Ncut



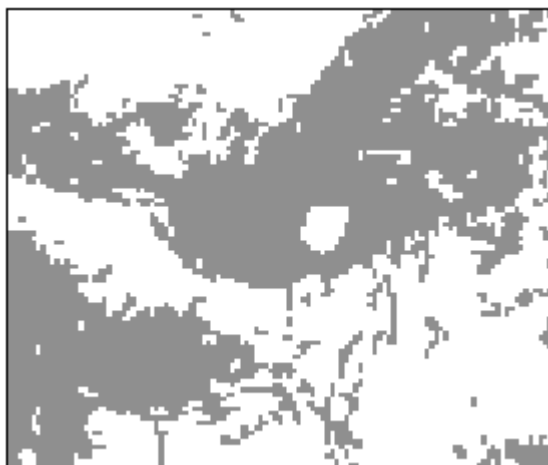
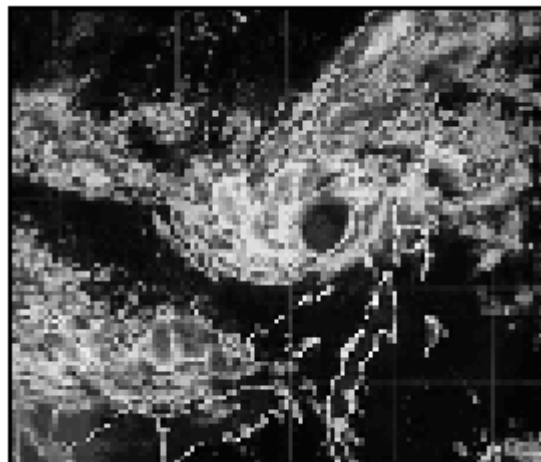
Felzenszwalb and Huttenlocher (2003).



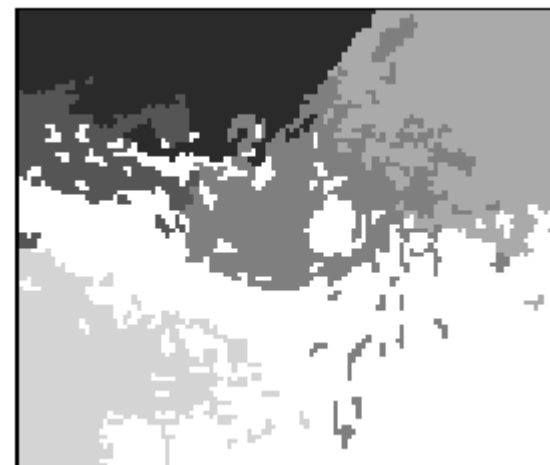
Gdalyahu, Weinshall, and Werman (2001).



## Intensity Segmentation Results (97 x 115)



Dominant sets

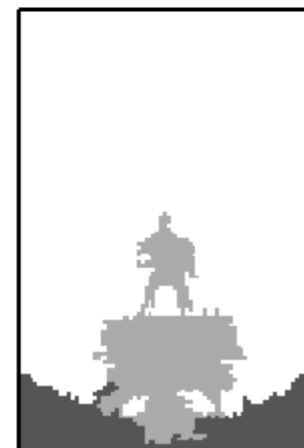
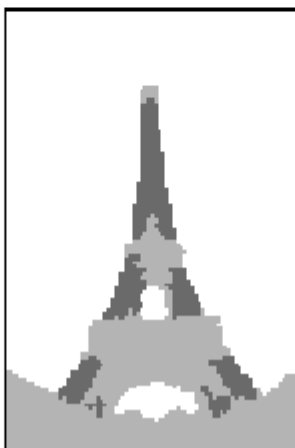


Ncut





## Color Segmentation Results (125 x 83)



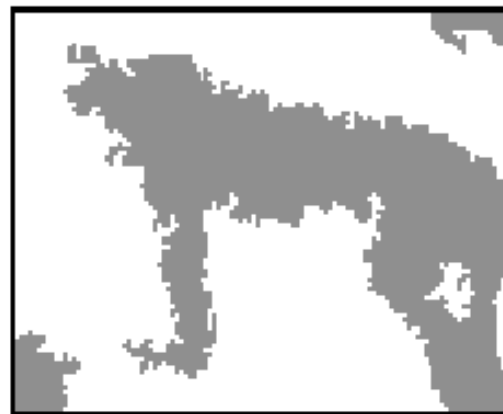
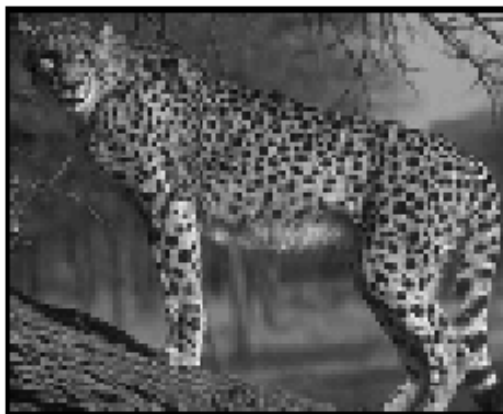
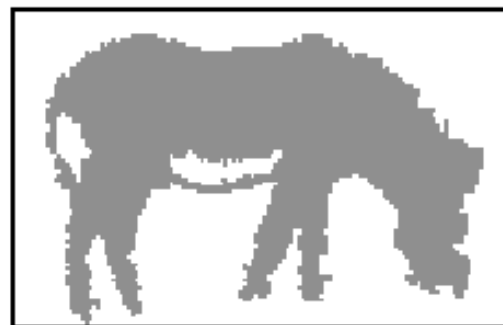
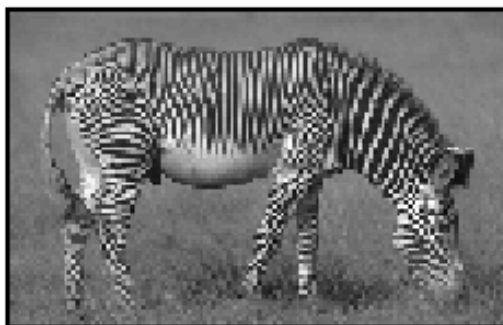
Original image

Dominant sets

Ncut



## Texture Segmentation Results (approx. 90 x 120)





# Ncut Results



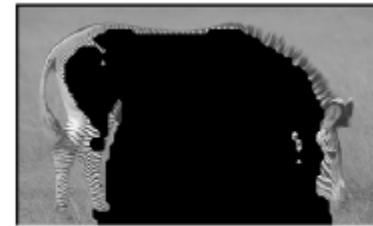
(a)



(b)



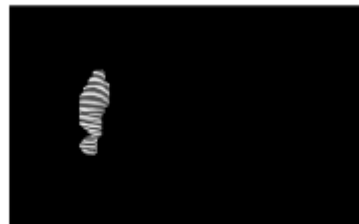
(c)



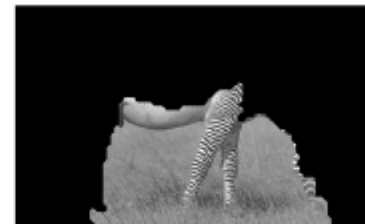
(d)



(e)



(f)



(g)



(h)



## Dealing with Large Data Sets

We address the problem of grouping *out-of-sample* (i.e., unseen) examples after the clustering process has taken place.

This may serve to:

1. substantially reduce the computational burden associated to the processing of very large data sets, by extrapolating the complete grouping solution from a small number of samples,
2. deal with dynamic situations whereby data sets need to be updated continually.



## Grouping Out-of-Sample Data

Recall that the sign of  $w_{S \cup \{i\}}(i)$  provides an indication as to whether  $i$  is tightly or loosely coupled with the vertices in  $S$ .

Accordingly, we use the following rule for predicting cluster membership of unseen data  $i$ :

if  $w_{S \cup \{i\}}(i) > 0$ , then assign vertex  $i$  to cluster  $S$ .

Can be computed in linear time wrt the size of  $S$

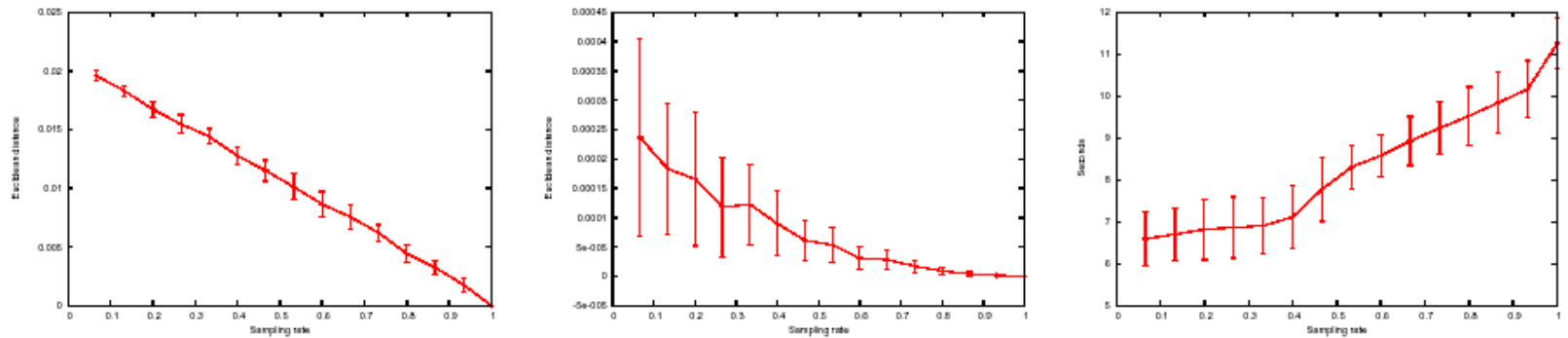


Figure 2: Evaluating the quality of our approximations on a 150-point cluster. Average distance between approximated and actual cluster membership (left) and cohesiveness (middle) as a function of sampling rate. Right: average CPU time as a function of sampling rate.

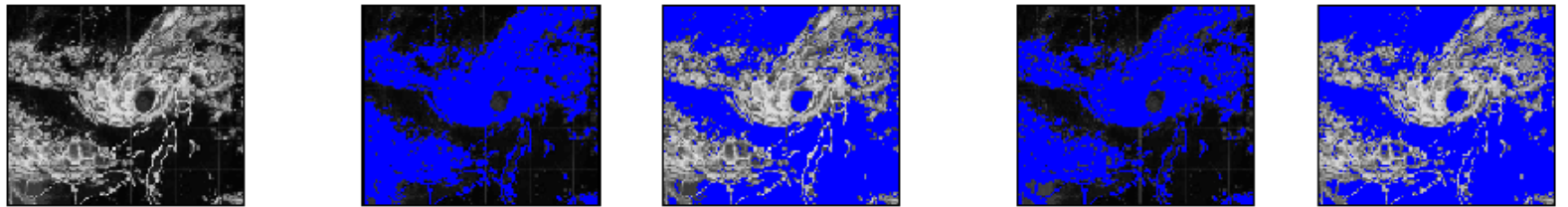


Figure 4: Segmentation results on a  $115 \times 97$  weather radar image. From left to right: original image, the two regions found on the sampled image (sampling rate = 0.5%), and the two regions obtained on the whole image (sampling rate = 100%).



## Results on Berkeley Database Images (321 x 481)



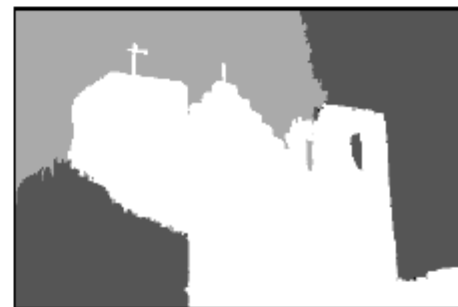
GCE = 0.05, LCE = 0.04



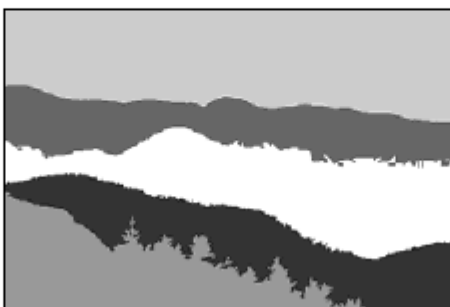
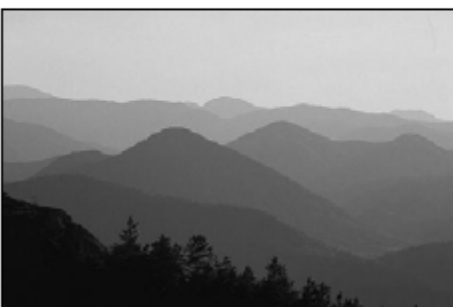
GCE = 0.08, LCE = 0.05



GCE = 0.11, LCE = 0.09



GCE = 0.36, LCE = 0.27



GCE = 0.09, LCE = 0.08



GCE = 0.31, LCE = 0.22





## Results on Berkeley Database Images (321 x 481)



GCE = 0.12, LCE = 0.12



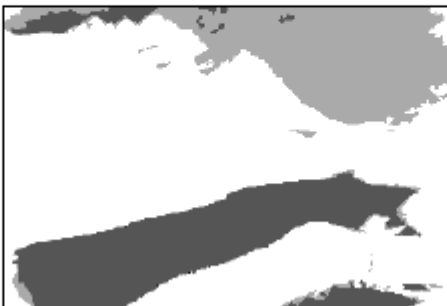
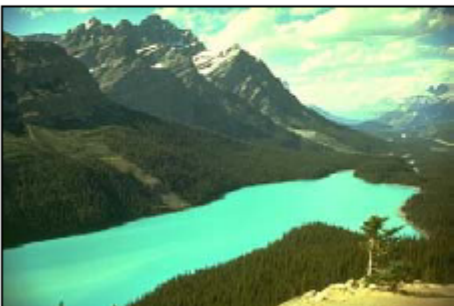
GCE = 0.19, LCE = 0.13



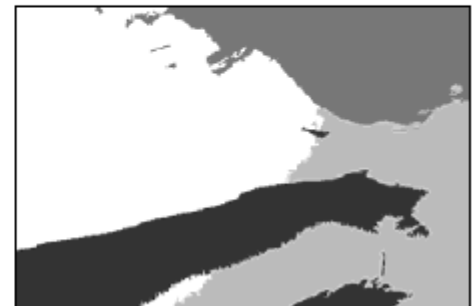
GCE = 0.31, LCE = 0.26



GCE = 0.35, LCE = 0.29



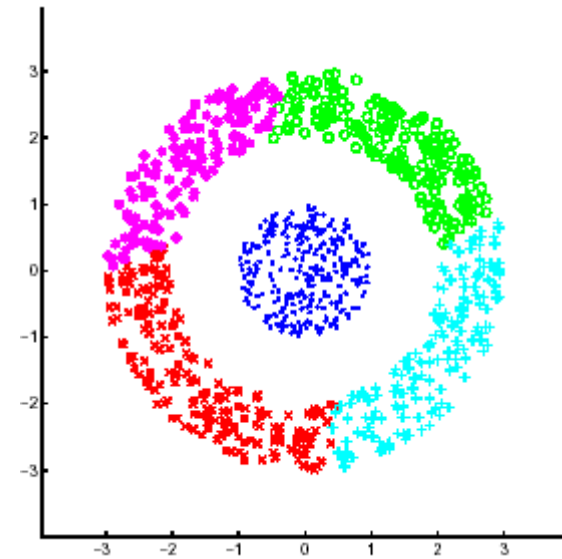
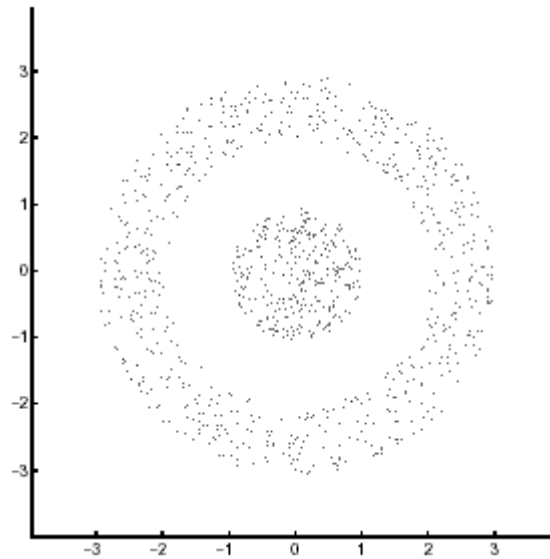
GCE = 0.09, LCE = 0.09



GCE = 0.16, LCE = 0.16

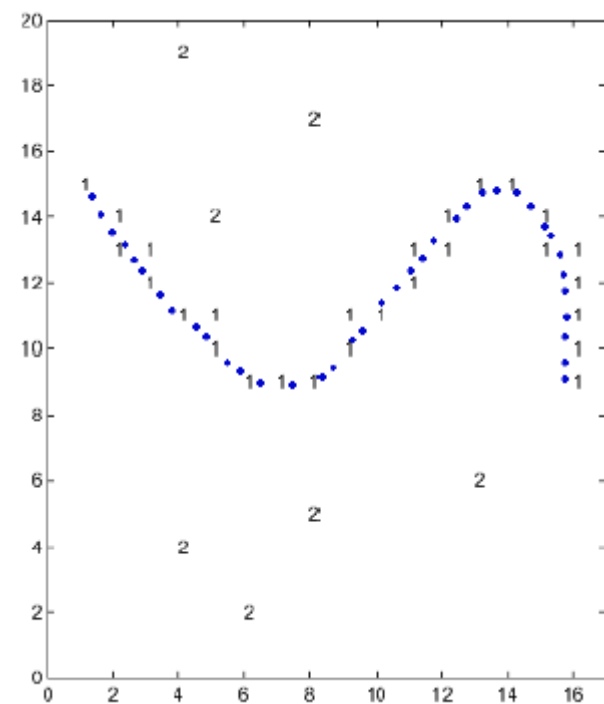
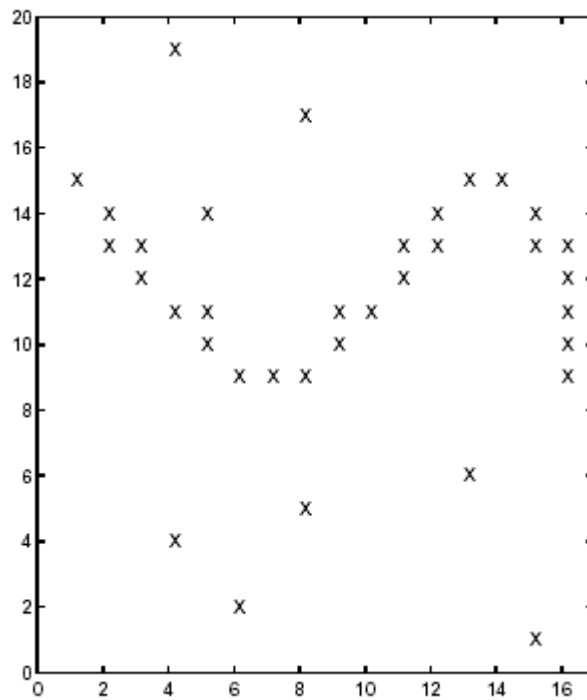


# Capturing Elongated Structures / 1





## Capturing Elongated Structures / 2





# “Closing” the Similarity Graph

**Basic idea:** Transform the original similarity graph  $G$  into a “closed” version thereof ( $G_{\text{closed}}$ ), whereby edge-weights take into account chained (path-based) structures.

Unweighted (0/1) case:

$$G_{\text{closed}} = \text{Transitive Closure of } G$$

**Note:**  $G_{\text{closed}}$  can be obtained from:

$$A + A^2 + \dots + A^n$$



## Weighted Closure of $G$

**Observation:** When  $G$  is weighted, the  $ij$ -entry of  $A^k$  represents the sum of the total weights on the paths of length  $k$  between vertices  $i$  and  $j$ .

Hence, our choice is:

$$A_{\text{closed}} = A + A^2 + \dots + A^n$$



## Example: Without Closure ( $\sigma = 2$ )

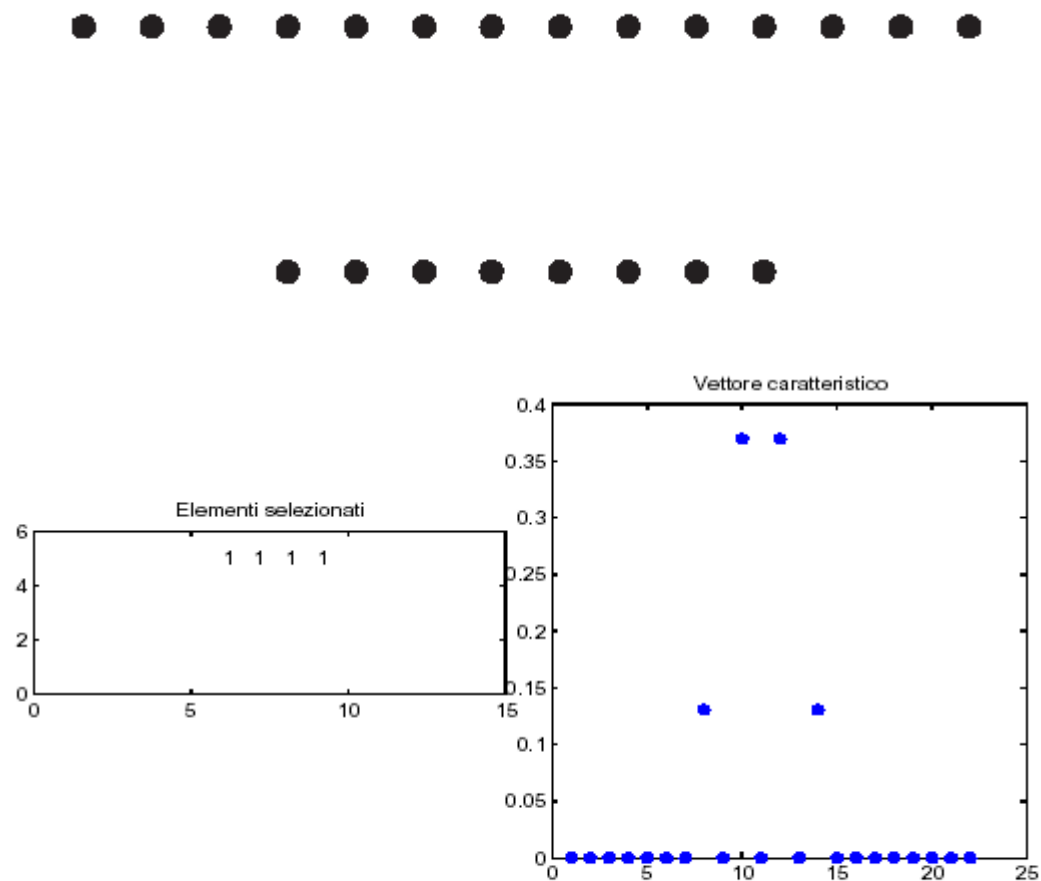


Figura 4.11: Cluster senza chiusura:  $\sigma = 2$



## Example: Without Closure ( $\sigma = 4$ )

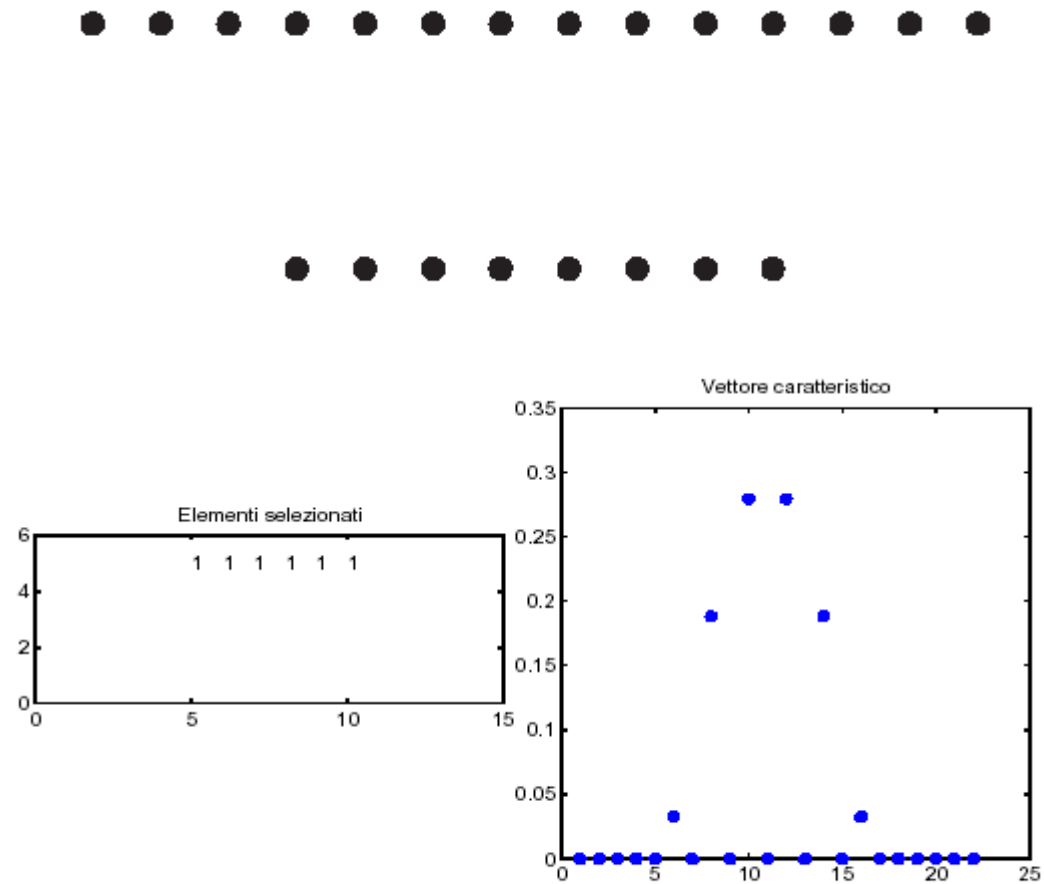


Figura 4.12: Cluster senza chiusura:  $\sigma = 4$



## Example: Without Closure ( $\sigma = 8$ )

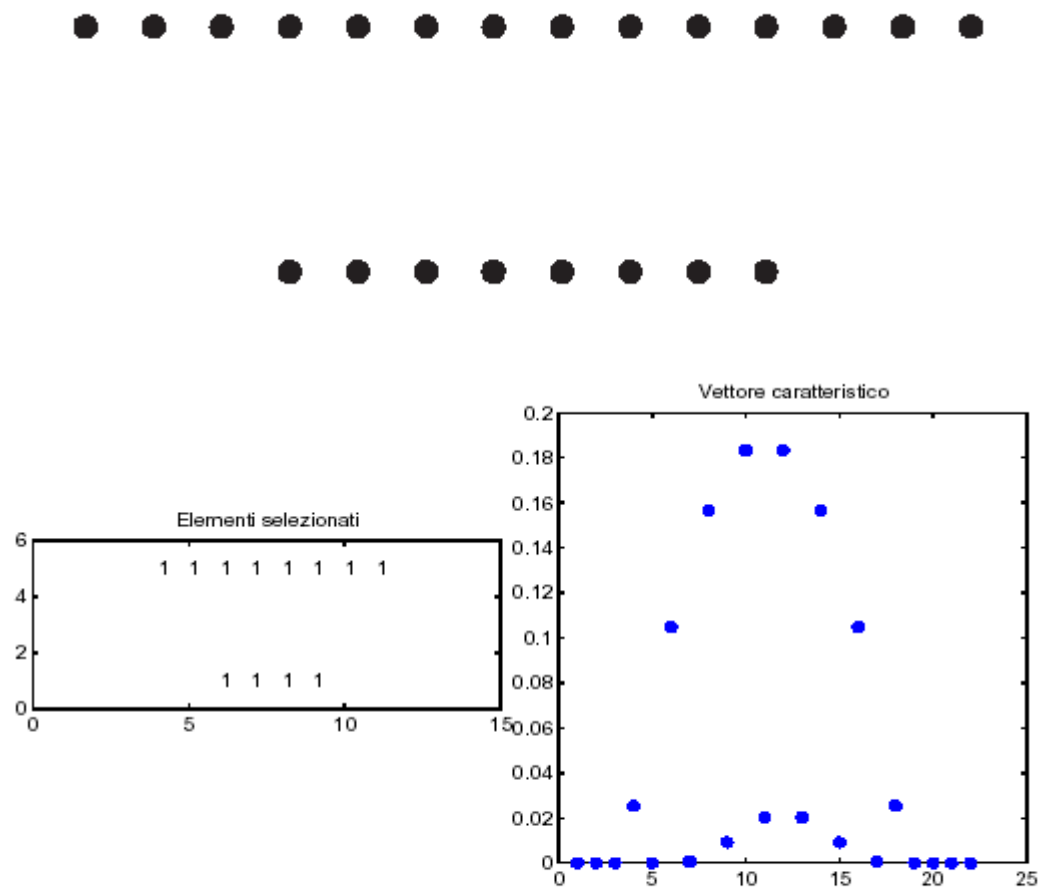


Figura 4.13: Cluster senza chiusura:  $\sigma = 8$





## Example: With Closure ( $\sigma = 0.5$ )

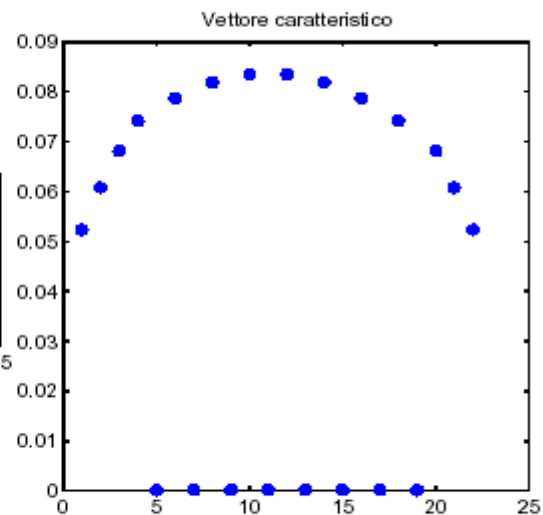
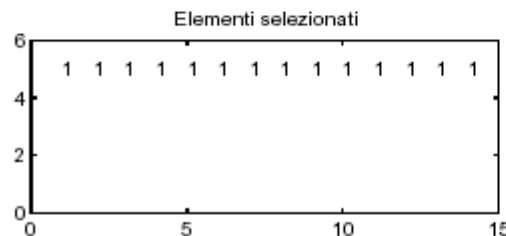
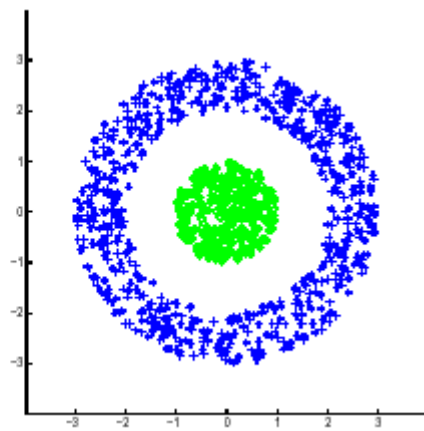
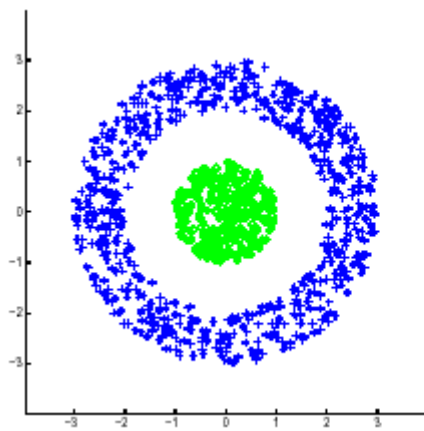
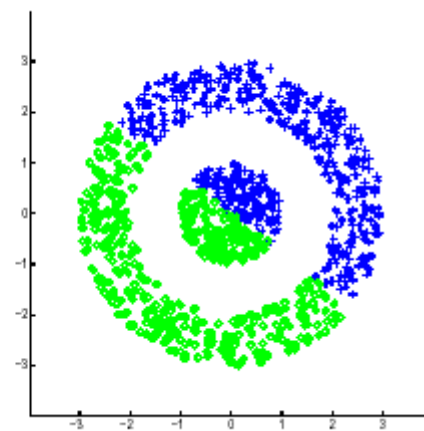
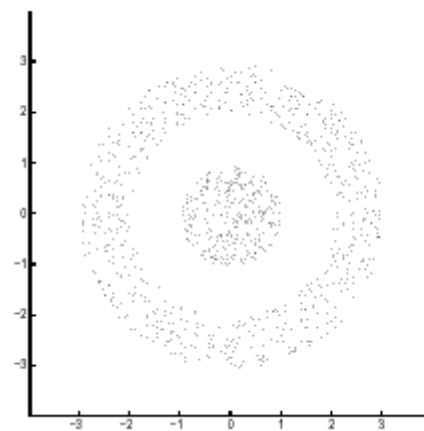


Figura 4.14: Cluster mediante chiusura:  $\sigma = 0,5$





# Grouping Edge Elements

Here, the elements to be grouped are **edgels** (edge elements).

We used Herault/Horaud (1993) similarities, which combine the following four terms:

1. Co-circularity
2. Smoothness
3. Proximity
4. Contrast

Comparison with Mean-Field Annealing (MFA).

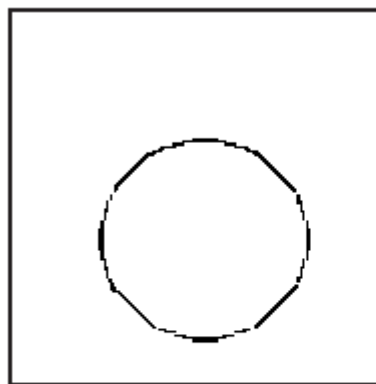


Immagine originale  
204 edge

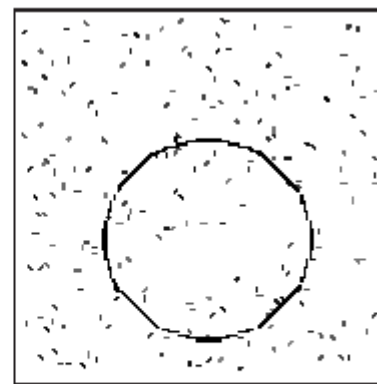
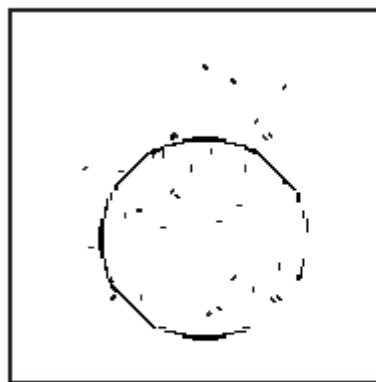
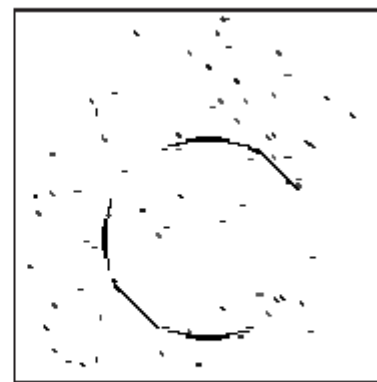


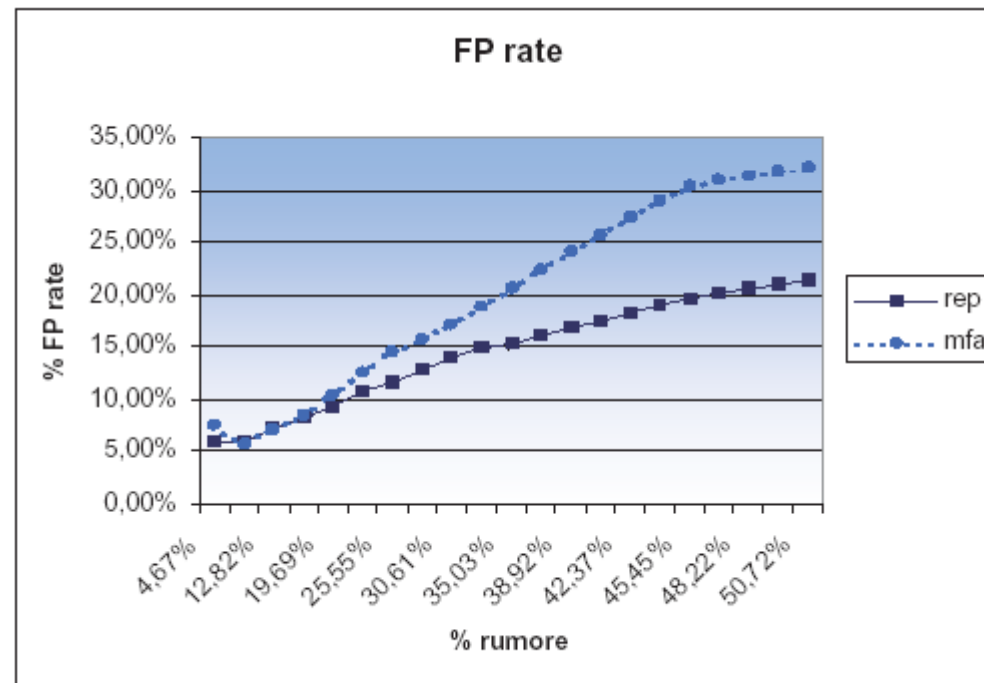
Immagine con rumore al 50%



Insiemi dominanti  
FP rate: 16,67%



Mean Field Annealing  
FP rate: 34, 31%



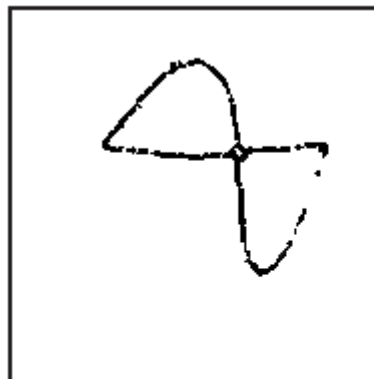


Immagine originale  
278 edge

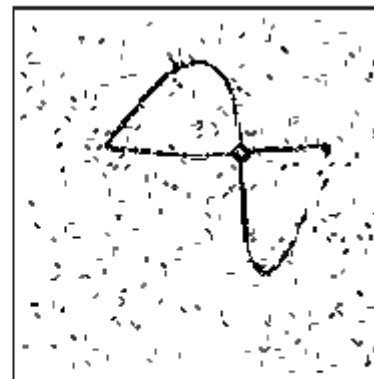
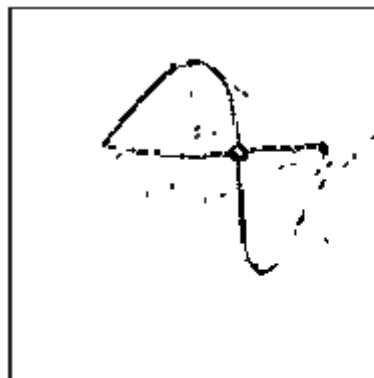
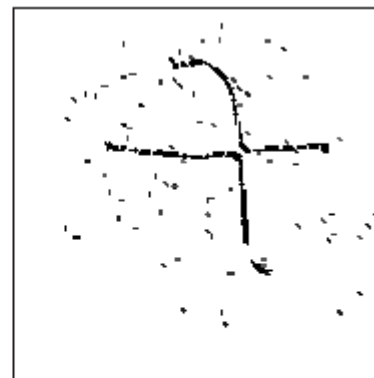


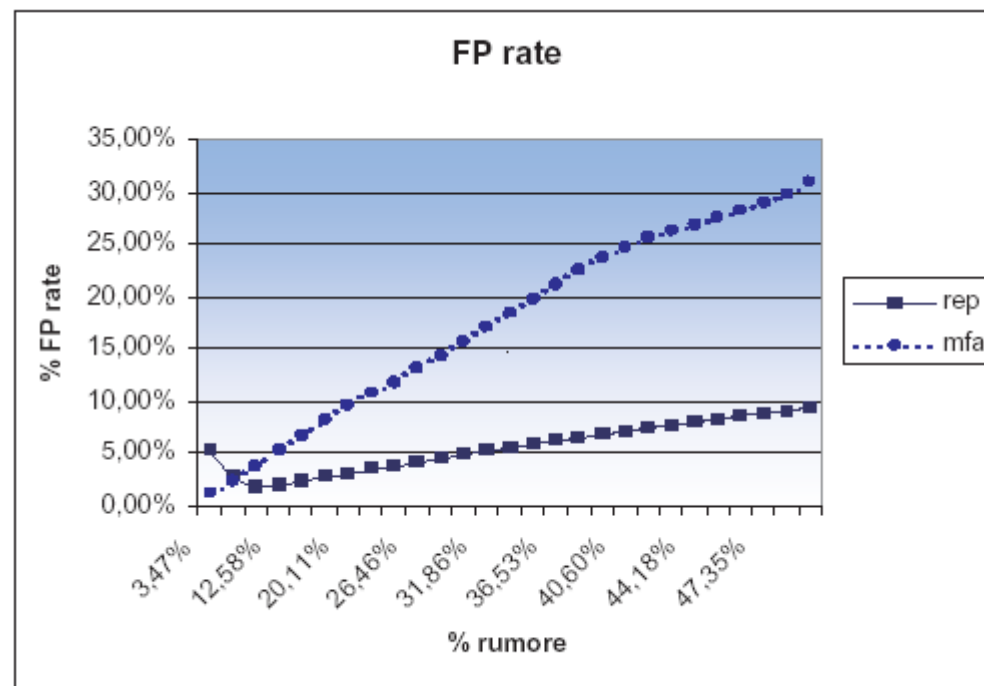
Immagine con rumore al 50%



Insiemi dominanti  
FP rate: 8,99%



Mean Field Annealing  
FP rate: 29,5%





# Lecture's Outline

- Dominant sets and their characterization
- Finding dominant sets using evolutionary game dynamics
- Experiments on image segmentation (and extensions)
- Dominant sets and hierarchical clustering
- Dealing with arbitrary affinities: Dominant sets as (evolutionary) game equilibria





## Building a Hierarchy: A Family of Quadratic Programs

Consider the following family of StQP's:

$$\begin{array}{ll} \text{maximize} & f_\alpha(\mathbf{x}) = \mathbf{x}'(A - \alpha I)\mathbf{x} \\ \text{subject to} & \mathbf{x} \in \Delta \end{array}$$

where  $\alpha \geq 0$  is a parameter and  $I$  is the identity matrix.

The objective function  $f_\alpha$  consists of:

- a **data term** ( $\mathbf{x}'A\mathbf{x}$ ) which favors solutions with high internal coherency
- a **regularization term** ( $-\alpha\mathbf{x}'\mathbf{x}$ ) which acts as an entropic factor: it is concave and, on the simplex  $\Delta$ , it is maximized at the barycenter and it attains its minimum value at the vertices of  $\Delta$



## An Observation

The solutions of the StQP remain the same if the matrix  $A - \alpha I$  is replaced with  $A - \alpha I + \kappa ee'$ , where  $\kappa$  is an arbitrary constant, since

$$\mathbf{x}'(A - \alpha I + \kappa ee')\mathbf{x} = \mathbf{x}'(A - \alpha I)\mathbf{x} + \kappa$$

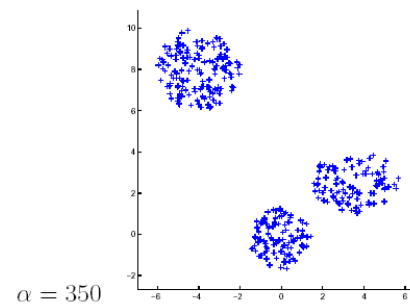
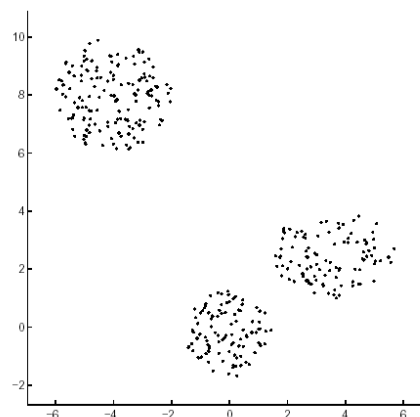
for all  $\mathbf{x} \in \Delta$ .

In particular, if  $\kappa = \alpha$  the resulting matrix is nonnegative and has a null diagonal.

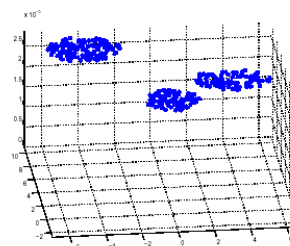
Hence all (strict) solutions of the StQP correspond to dominant sets for the scaled similarity matrix  $A + \alpha(ee' - I)$  having the off-diagonal entries equal to  $a_{ij} + \alpha$ .



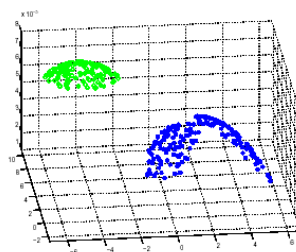
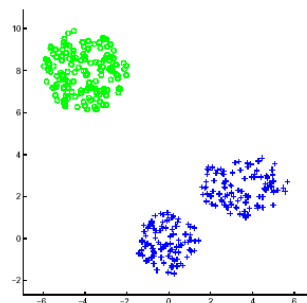
# The effects of $\alpha$



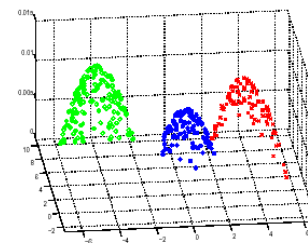
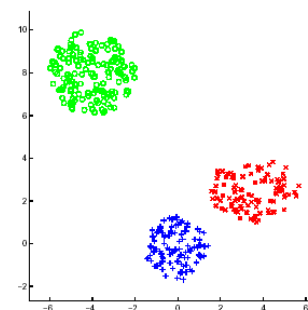
$\alpha = 350$



$\alpha = 10$



$\alpha = 0$





## Bounds for the Regularization Parameter / 1

When  $\alpha$  is large enough the regularization term  $(-\alpha \mathbf{x}'\mathbf{x})$  dominates, and the only solution of the StQP is in the interior of  $\Delta$ : this corresponds to a unique large cluster which comprises all the data points.

**Proposition** *If*

$$\alpha > \lambda_{\max}(A)$$

*then  $f_\alpha$  is a strictly concave function in  $\mathbb{R}^n$ , and the only solution  $\mathbf{x}$  of the StQP belongs to the interior of  $\Delta$ , i.e.,  $\sigma(\mathbf{x}) = V$ .*



## Bounds for the Regularization Parameter / 2

Given a subset of vertices  $S \subseteq V$ , the face of  $\Delta$  corresponding to  $S$  is defined as:

$$\Delta_S = \{\mathbf{x} \in \Delta : \sigma(\mathbf{x}) \subseteq S\}$$

and its relative interior is:

$$\text{int}(\Delta_S) = \{\mathbf{x} \in \Delta : \sigma(\mathbf{x}) = S\} .$$

**Theorem** *Let  $S \subset V$  be a proper subset of vertices ( $S \neq V$ ), and let  $A_S$  denote the submatrix of  $A$  formed by the rows and columns indexed by the elements of  $S$ . If*

$$\alpha > \lambda_{\max}(A_S)$$

*then there is no point  $\mathbf{x} \in \text{int}(\Delta_S)$  that is a local maximizer of  $f_\alpha$  in  $\Delta$ .*



## Bounds for the Regularization Parameter / 3

Suppose for simplicity that  $a_{ij} \leq 1$  for all  $i, j \in V$ , i.e.

$$0 \leq A \leq \mathbf{e}\mathbf{e}^T - I.$$

For any  $S \subseteq V$  we get:

$$\lambda_{\max}(A_S) \leq \lambda_{\max}(\mathbf{e}\mathbf{e}^T - I) = |S| - 1$$

Hence, if we want to avoid clusters of size  $|S| \leq m < |V|$  we could let

$$\alpha > m - 1$$

In so doing, no face  $\Delta_S$  with  $|S| \leq m$  will contain solutions of the StQP, in other words:

at this scale *all* clusters will have more than  $m$  data points



## The Landscape of $f_\alpha$

**Key observation:** For any fixed  $\alpha$ , the energy landscape of  $f_\alpha$  is populated by two kinds of solutions:

- solutions which correspond to dominant sets for the original matrix  $A$
- solutions which do not correspond to any dominant set for the original matrix  $A$ , although they are dominant for the scaled matrix  $A + \alpha(\mathbf{e}\mathbf{e}' - I)$

The latter represent large subsets of points that are not sufficiently coherent to be dominant with respect to  $A$ , and hence they should be split.



# Sketch of the Hierarchical Clustering Algorithm

**Basic idea:** start with a sufficiently large  $\alpha$  and adaptively decrease it during the clustering process:

- 1) let  $\alpha$  be a large positive value (e.g.,  $\alpha > |V| - 1$ )
- 2) find a partition of the data into  $\alpha$ -clusters
- 3) for all the  $\alpha$ -clusters that are not 0-clusters recursively repeat step 2) with decreased  $\alpha$





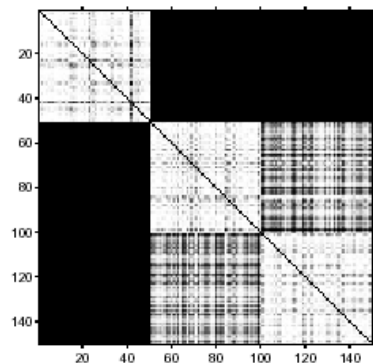
## Pseudo-code of the Algorithm

```
Algorithm HIER_CLUSTERING(  $V$ ,  $A$  )  
begin  
  if  $V$  is dominant then return  $V$   
  let  $\alpha > |V| - 1$   
  repeat  
    decrease  $\alpha$   
    if  $\alpha < 0$  then  $\alpha \leftarrow 0$   
     $V_1, \dots, V_k \leftarrow \text{SPLIT}( V, A, \alpha )$   
  until  $k > 1$   
  return  $\bigcup_{i=1}^k \{ \text{HIER\_CLUSTERING}( V_i, A_{V_i} ) \}$   
end
```

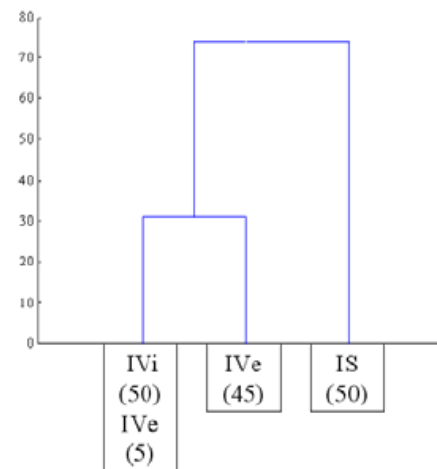


## Results on the IRIS dataset / 1

This data set, attributed to Fisher (1936), is a classic benchmark in the machine learning literature. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. The three classes are Iris Setosa (IS), Iris Versicolour (IVe), and Iris Virginica (IVi). Each data item is a 4-dimensional real vector representing as many measurements of an Iris flower. Class IS is linearly separable from the other two (IVe and IVi), but IVe and IVi are not linearly separable.



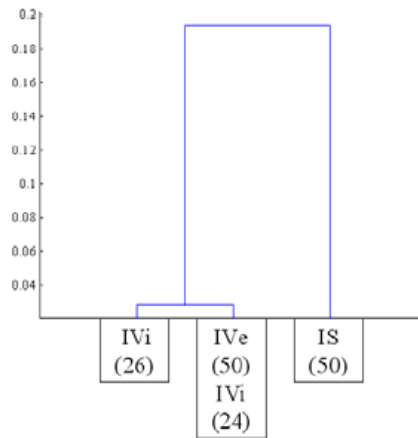
Similarity matrix



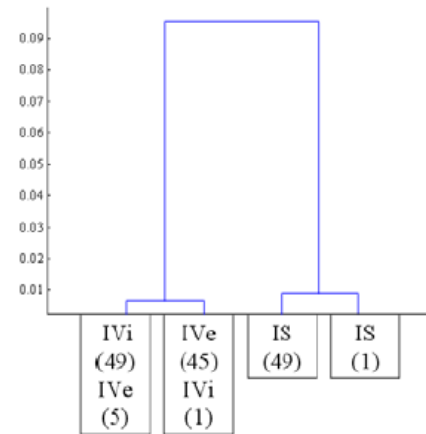
Dominant sets



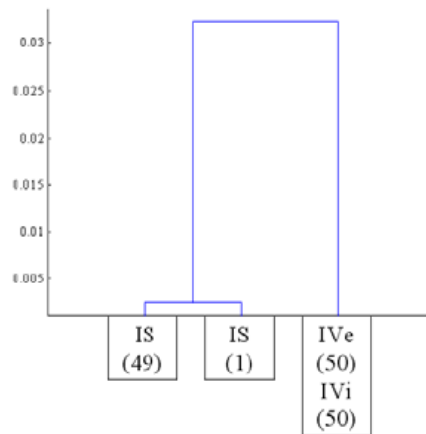
## Results on the IRIS dataset / 2



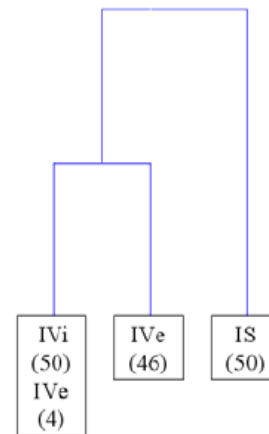
Complete-link



Average-link



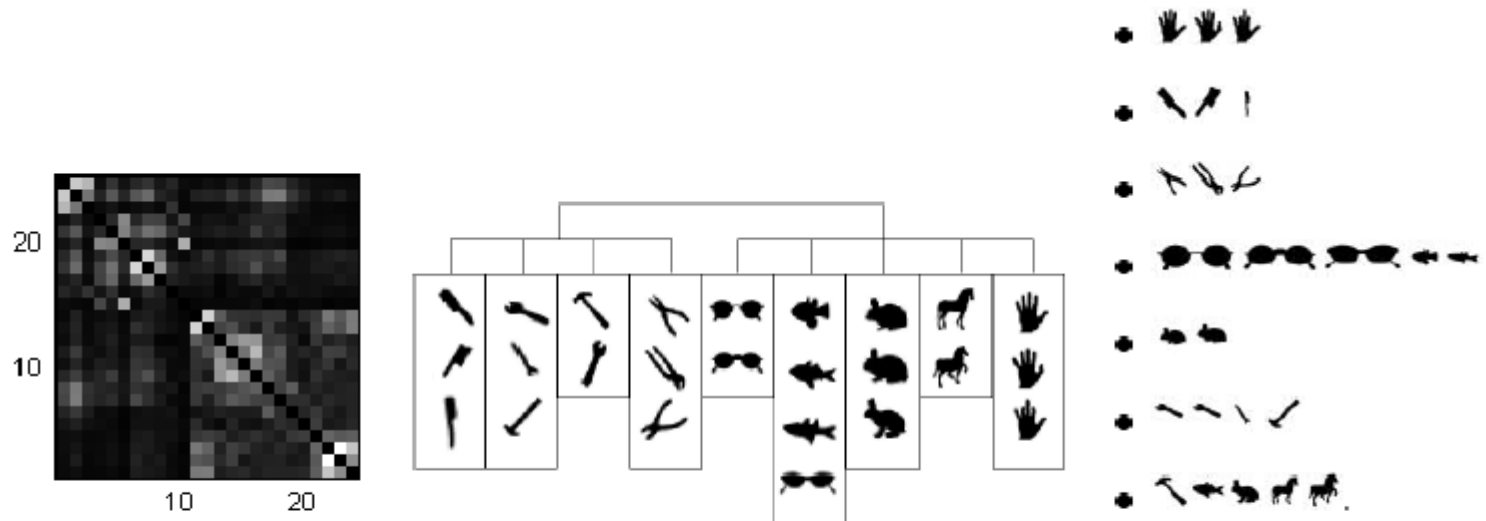
Single-link



NCut



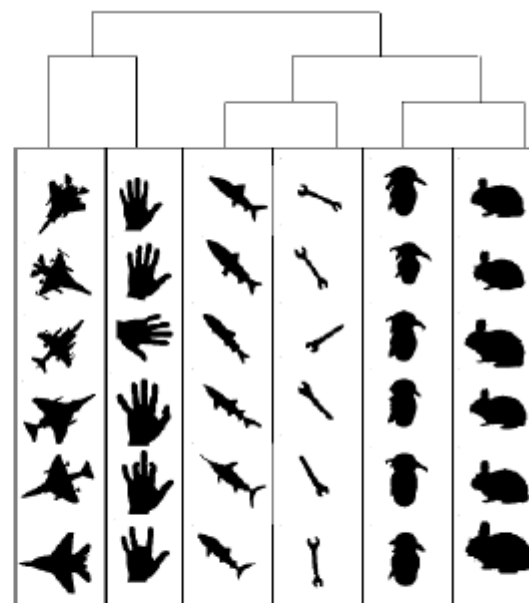
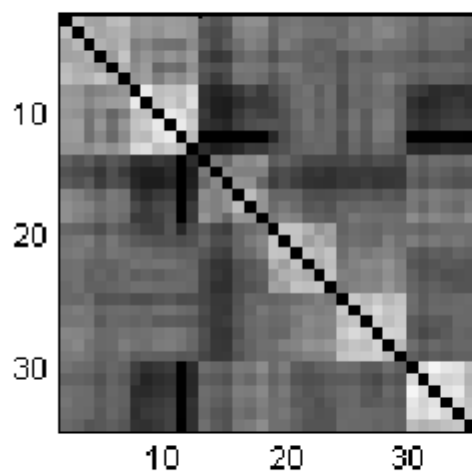
## Luo and Hancock's Similarities (CVPR'01)



Left: Similarity matrix used in the experiment. Middle: Hierarchy produced by our algorithm. Right: (Flat) partition produced by Luo and Hancock.



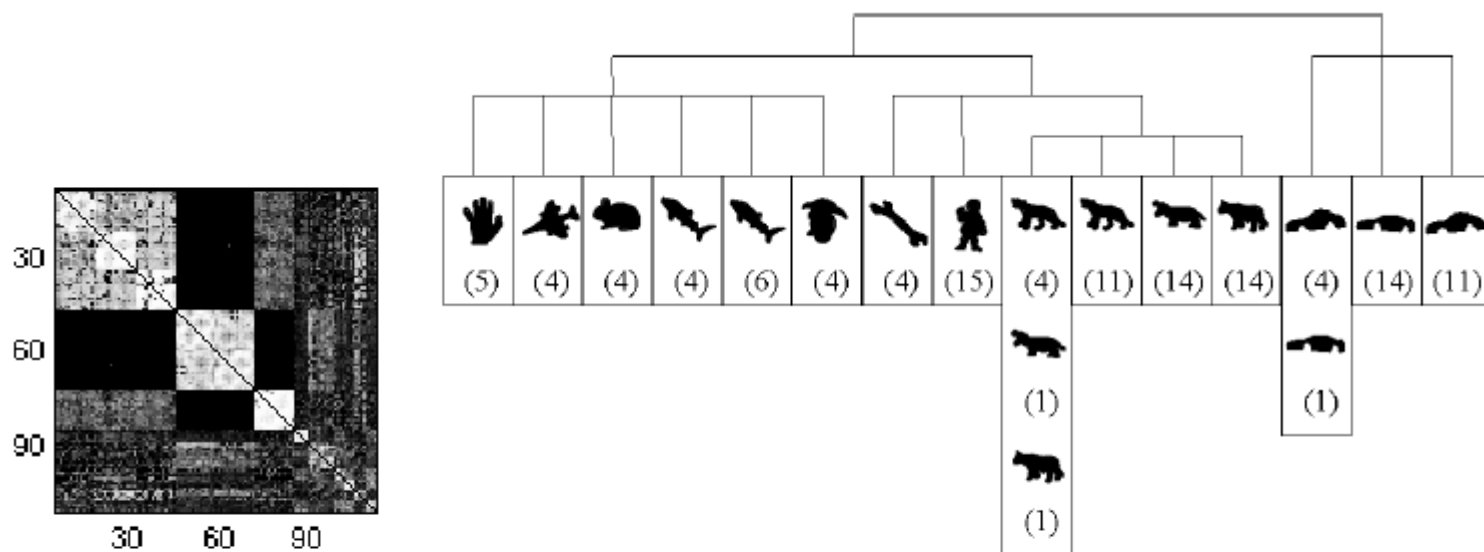
## Klein and Kimia's Similarities (SODA'01)



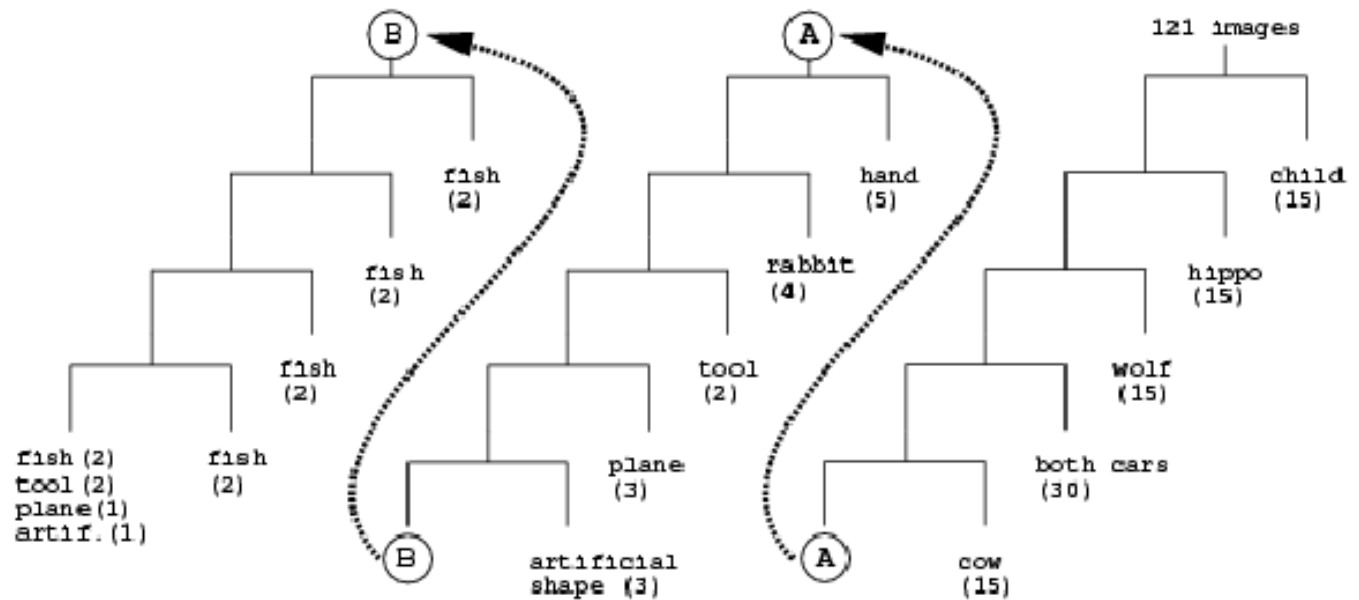
Left: Similarity matrix used in the experiment. Right: Hierarchy produced by our algorithm.



## Gdalyahu and Weinshall's Similarities (PAMI 01)

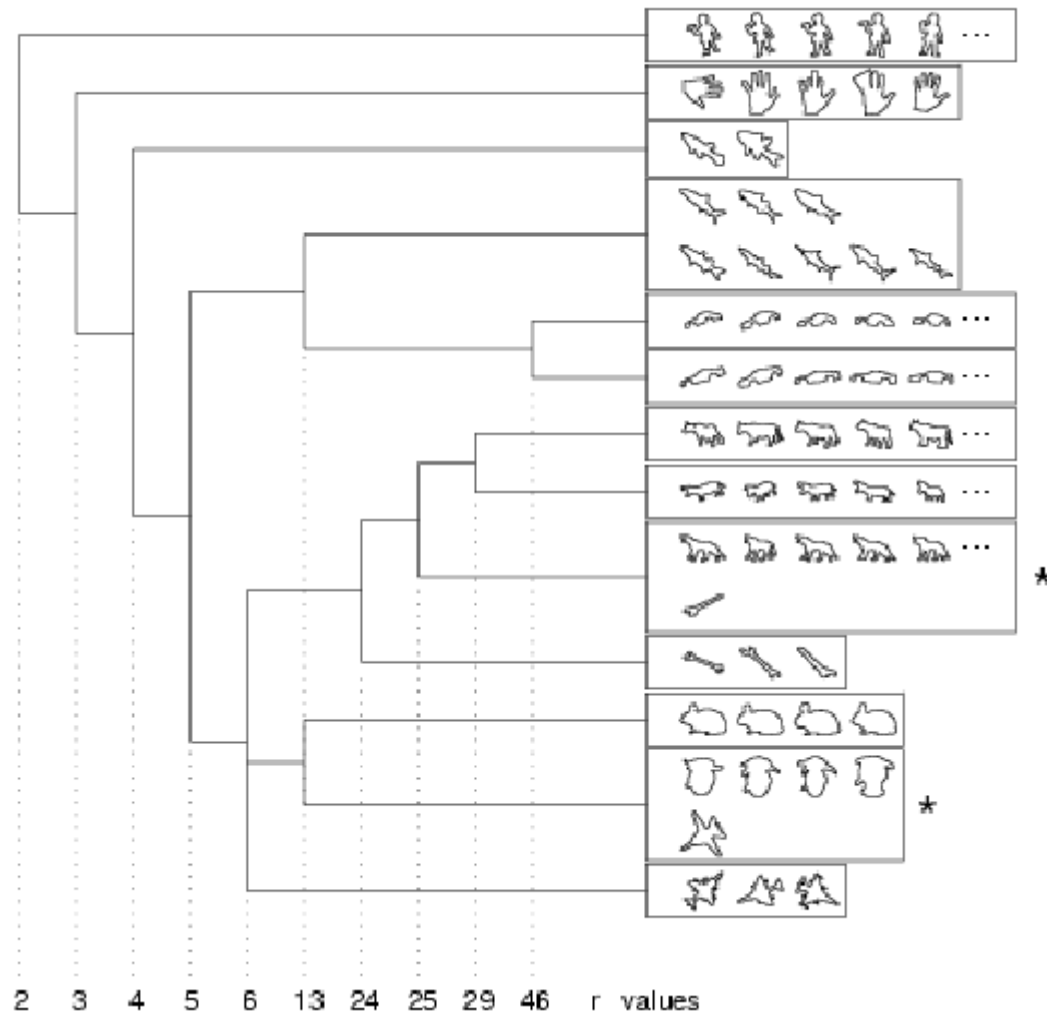


Left: Similarity matrix used in the experiment (courtesy of Y. Gdalyahu). Right: Hierarchy produced by our algorithm.





## Typical-cut Results (From Gdalyahu, 1999)







# Lecture's Outline

- Dominant sets and their characterization
- Finding dominant sets using evolutionary game dynamics
- Experiments on image segmentation (and extensions)
- Dominant sets and hierarchical clustering
- Dealing with arbitrary affinities: Dominant sets as (evolutionary) game equilibria



# Rationale

A classical strategy to attack pattern recognition problems consists of formulating them in terms of optimization problems.

In many real-world situations, however, the complexity of the problem at hand is such that no single (global) objective function would satisfactorily capture its intricacies.

Examples include:

- Using asymmetric compatibilities in (continuous) consistency labeling problems (Hummel & Zucker, 1983)
- Integrating region- and gradient-based methods in image segmentation tasks (Chakraborty & Duncan, 1999)
- Grouping with asymmetric affinities (Yu and Shi, 2001; Torsello, Rota Bulò & Pelillo, 2006)



# Game Theory

Game theory was developed precisely to overcome the limitations of single-objective optimization (von Neumann, Nash).

It aims at modeling complex situations where players make decisions in an attempt to maximize their own (mutually conflicting) returns.

Nowadays, game theory is a well-established field on its own and offers a rich arsenal of powerful concepts and algorithms.

**Note:** in the case of a particular class of games (i.e., doubly-symmetric games) game-theoretic criteria reduce to optimality criteria.



# State of the Art

In the past there have been only few, isolated attempts aimed at explicitly formulating pattern recognition problems from a purely game-theoretic perspective

On the one hand, there have been those who have pointed out the **analogies** between classical game-theoretic concepts, such as the Nash equilibrium, and consistency criteria for consistent labeling problems (e.g., Zucker & Miller, 1992; Sastry *et al.*, 1994).

On the other hand, there have been some attempts at formulating **specific** computer vision and pattern recognition problems, such as module integration or image segmentation, as game problems (e.g., Bozma & Duncan, 1994; Chackraborty & Duncan, 1999).

Recently, in the machine learning community, there has been an interest in **computational game theory** (e.g., Ortiz and Kearns, 2002), which, however, emphasizes the algorithmic aspects of game theory, while neglecting the *modeling* side.



# Aim

- Develop a generic framework for grouping and clustering derived from a game-theoretic formalization of the competition between class hypotheses..
- The approach can deal with **non-metric similarities**, and, in particular, **asymmetric** and/or **negative similarities**.
- A common method to deal with asymmetric compatibilities is to symmetrize the similarity matrix (but see Yu and Shi, 2001).
- This approach, however, loses any information that might reside in the asymmetry.



# Game Theory: Basics

Assume:

- a game between two players
- complete knowledge
- a pre-existing set of (pure) strategies  $O=\{o_1,\dots,o_n\}$  available to the players.

Each player receives a payoff depending on the strategies selected by him and by the adversary

A **mixed strategy** is a probability distribution  $\mathbf{x}=(x_1,\dots,x_n)^T$  over the strategies.

$$\Delta = \left\{ \mathbf{x} \in \mathbf{R}^n : \mathbf{e}'\mathbf{x} = 1 \text{ and } x_i \geq 0 \ \forall i \in V \right\}$$



# Nash Equilibria and Extensions

- Let  $A$  be a payoff matrix:  $a_{ij}$  is the payoff obtained by playing  $i$  while the opponent plays  $j$ .
- $\mathbf{y}'\mathbf{A}\mathbf{x}$  is the average payoff obtained by playing mixed strategy  $y$  while the opponent plays  $x$ .
- A mixed strategy  $x$  is a *Nash equilibrium* if  $\mathbf{x}'\mathbf{A}\mathbf{x} \geq \mathbf{y}'\mathbf{A}\mathbf{x}$  for all strategies  $y$ . (Best reply to itself.)
- A Nash equilibrium is an *Evolutionary Stable Strategy* (ESS) if, for all strategies  $y$

$$\mathbf{y}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{A}\mathbf{x} \quad \Rightarrow \quad \mathbf{x}'\mathbf{A}\mathbf{y} > \mathbf{y}'\mathbf{A}\mathbf{y}$$



# Back to Optimazion

In doubly-symmetric games (i.e.,  $A=A^T$ ), we have:

Nash = Local maximizer of  $x^T A x$

ESS = Strict local maximizer of  $x^T A x$





# The Grouping Game

- Two players play by simultaneously selecting an element of  $O$ .
- Each player receives a payoff proportional to the affinity with respect to the element chosen by the opponent.
- Clearly, it is in each player's interest to pick an element that is strongly supported by the elements that the adversary is likely to choose.



# Game Theoretic Notions of a Cluster

Nash equilibria abstracts well the main characteristics of a cluster:

- **Internal coherency:** High mutual support of all elements within the group.
- **External incoherency:** Low support from elements of the group to elements outside the group.

This is not enough, though. We also want the solution to be stable and unambiguous, that is we require the solution to be isolated.

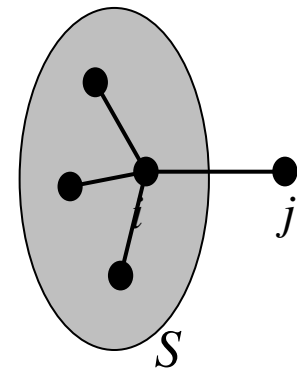
Hence we require that groups are ESS's.



# Basic Definitions

Let  $S \subseteq V$  be a non-empty subset of vertices and  $i \in S$ . The (average) weighted degree of  $i$  w.r.t.  $S$  is defined as:

$$\text{awdeg}_S(i) = \frac{1}{|S|} \sum_{j \in S} a_{ij} .$$



Moreover, if  $j \notin S$  we define:

$$\phi_S(i, j) = a_{ij} - \text{awdeg}_S(i) .$$

Intuitively,  $\phi_S(i, j)$  measures the similarity between nodes  $j$  and  $i$ , with respect to the average similarity between node  $i$  and its neighbors in  $S$ .



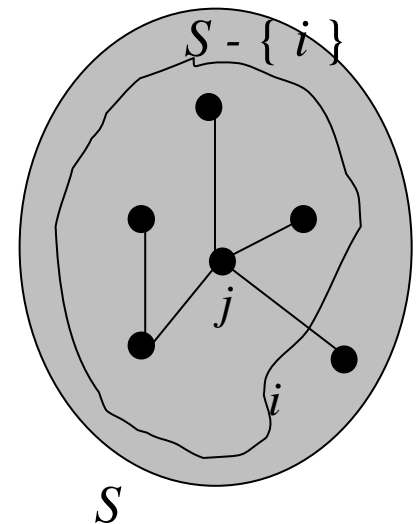
# Assigning Node Weights / 1

Let  $S \subseteq V$  be a non-empty subset of vertices and  $i \in S$ . The **weight** of  $i$  w.r.t.  $S$  is

$$w_S(i) = \begin{cases} 1, & \text{if } |S| = 1 \\ \sum_{j \in S \setminus \{i\}} \phi_{S \setminus \{i\}}(j, i) w_{S \setminus \{i\}}(j), & \text{otherwise.} \end{cases}$$

Moreover, the **total weight** of  $S$  is defined to be:

$$W(S) = \sum_{i \in S} w_S(i).$$

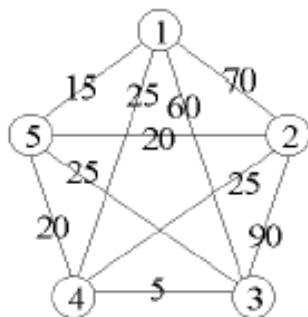




# (Directed) Dominant Sets

A non-empty subset of vertices  $S \subseteq V$  such that  $W(T) > 0$  for any non-empty  $T \subseteq S$ , is said to be **dominant** if:

1.  $w_S(i) > 0$ , for all  $i \in S$  (internal homogeneity)
2.  $w_{S \cup \{i\}}(i) < 0$ , for all  $i \notin S$  (external inhomogeneity)



Dominant sets  $\equiv$  clusters

The set  $\{1, 2, 3\}$  is dominant.

For 0/1 matrices: dominant sets  $\equiv$  (strictly) maximal cliques



## Main result

**Theorem** *Evolutionary stable strategies of the grouping game with affinity matrix  $A$  are in a one-to-one correspondence with (directed) dominant sets.*

**Note:** Generalization of CVPR'03/PAMI'07 Theorem which states that (undirected) dominant sets are in one-to-one correspondence with strict local maximizers of  $x^T A x$  in the standard simplex.



# Replicator Dynamics and ESS's

**Theorem 3** *A point  $\mathbf{x} \in \Delta$  is the limit of a trajectory of (6) starting from the interior of  $\Delta$  if and only if  $\mathbf{x}$  is a Nash equilibrium. Further, if point  $\mathbf{x} \in \Delta$  is ESS then it is asymptotically stable.<sup>2</sup>*



# Experimental Setup

We applied the proposed clustering framework to the perceptual grouping of edge elements (edgelets) in a noisy image.

Two affinity measure:

- one asymmetric (Williams and Thornber, 2000).
- one symmetric (Hèrault and Houraud, 1983).

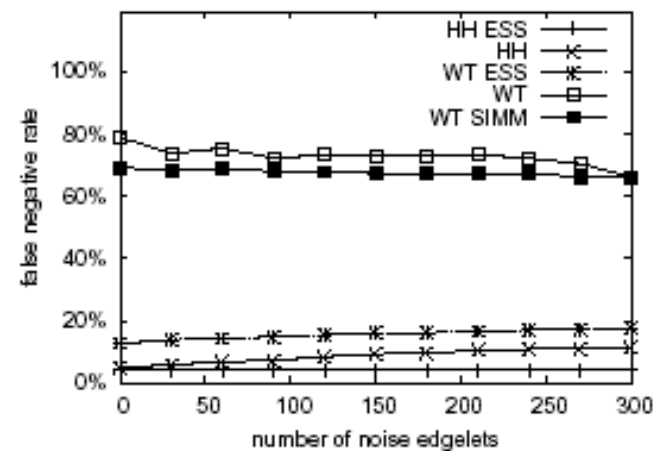
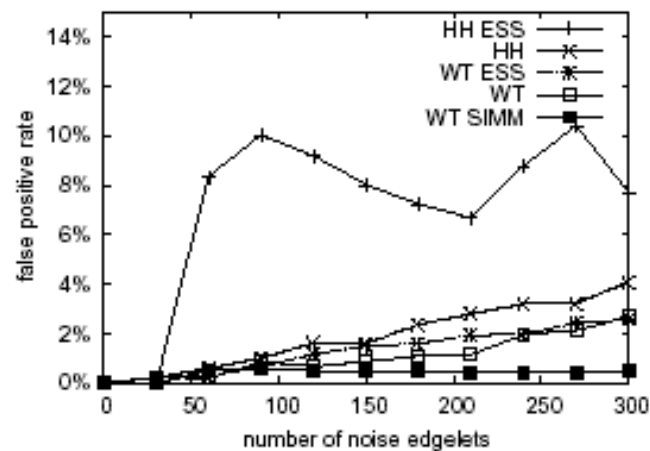
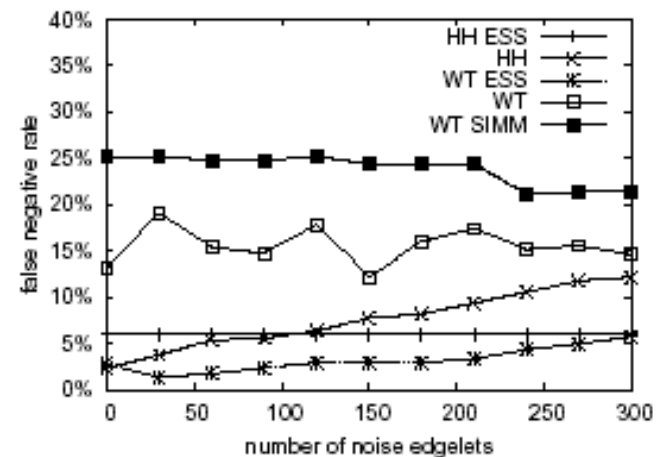
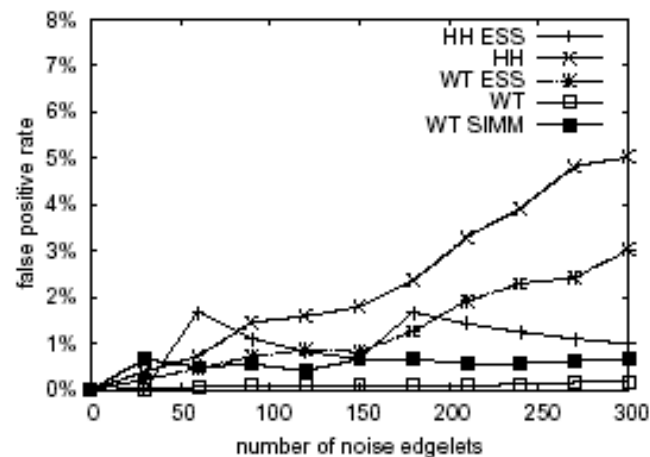
Compared the result obtained with our approach (ESS+WT, ESS+HH) with the approaches presented in the original papers (WT and HH).

We also apply the approach to a symmetrized version of the WT measure (ESS+WTSIMM).





# Synthetic Examples



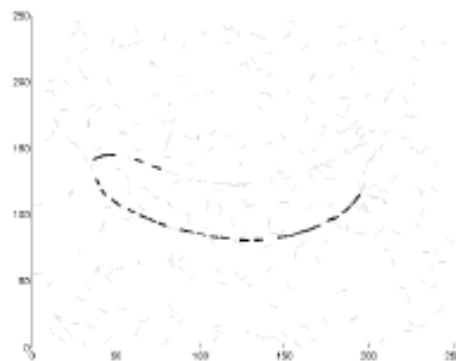


# Textured Background

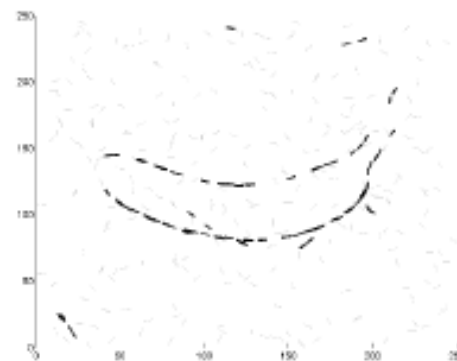
Original



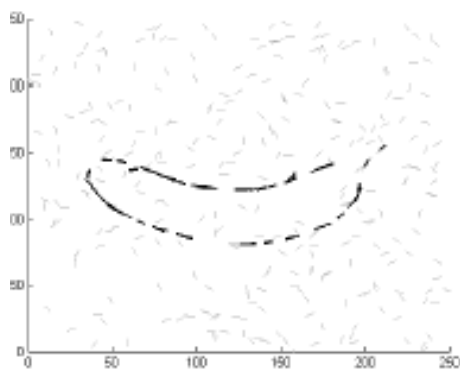
WT



ESS+WT



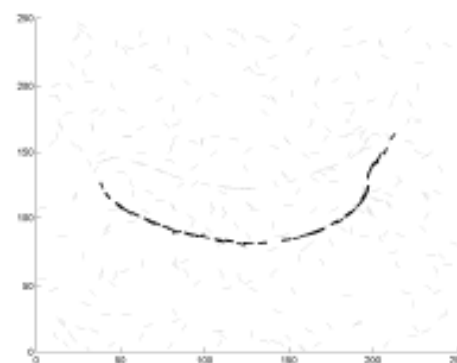
ESS+WTSIMM



HH



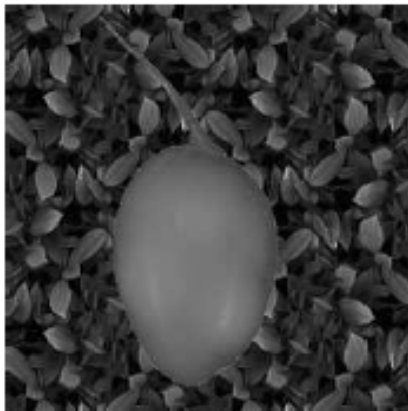
ESS+HH



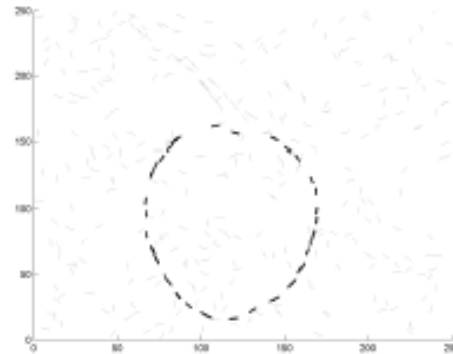


# Textured Background

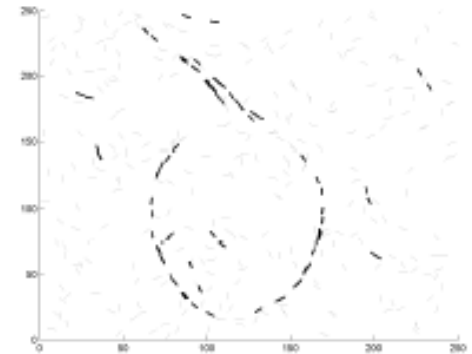
Original



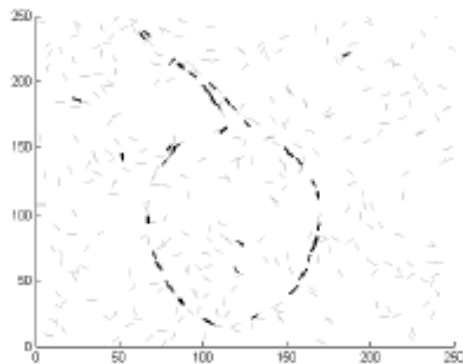
WT



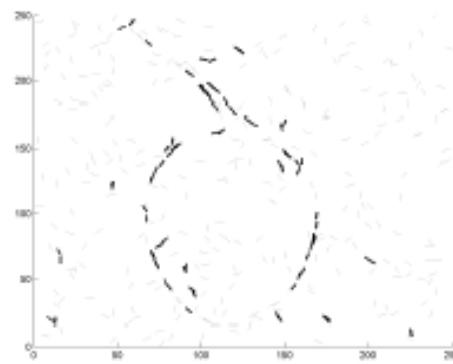
ESS+WT



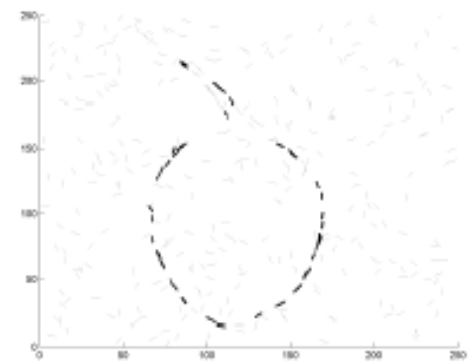
ESS+WTSIMM



HH



ESS+HH





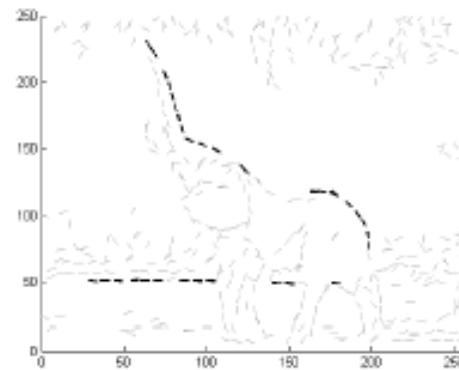
		Potato	Banana	Leaf	Tamarillo
WT	FPR	1%	0%	5%	0.7%
	FNR	45%	45%	76%	50.4%
ESS+WT	FPR	6%	6%	6%	6.7%
	FNR	16%	9%	37%	19.5%
ESS+ WTSIMM	FPR	0.71%	1.53%	1%	1.47%
	FNR	39.71%	35.2%	59.73%	41.33%
HH	FPR	13%	9%	12%	12.4%
	FNR	17%	20%	44%	24.5%
ESS+HH	FPR	13.67%	4.33%	14.67%	5.3%
	FNR	8%	8%	26.67%	16%



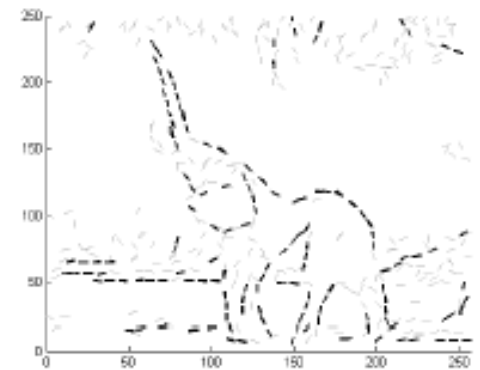
Original



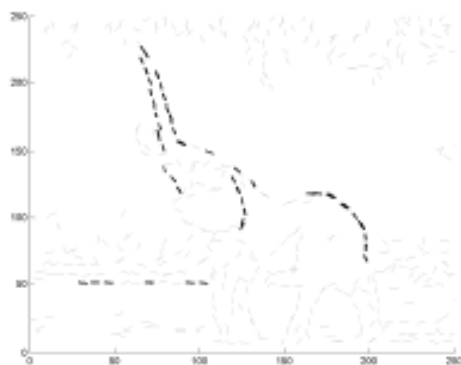
WT



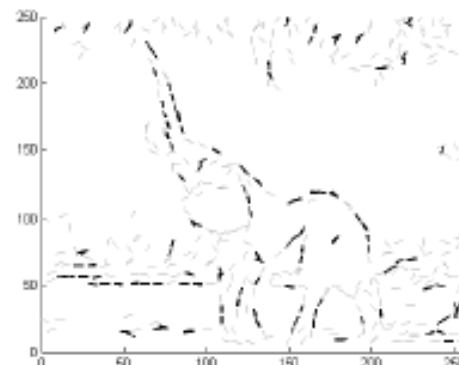
ESS+WT



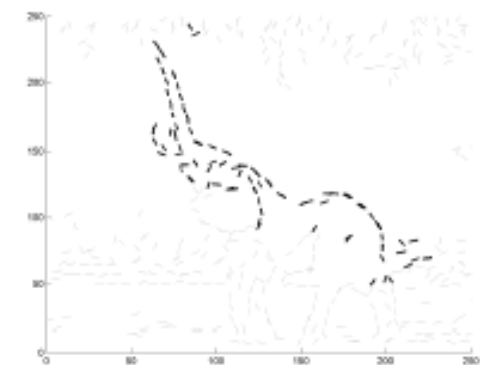
ESS+WTSIMM



HH



ESS+HH

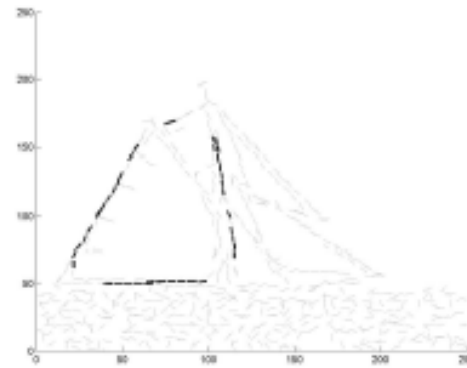




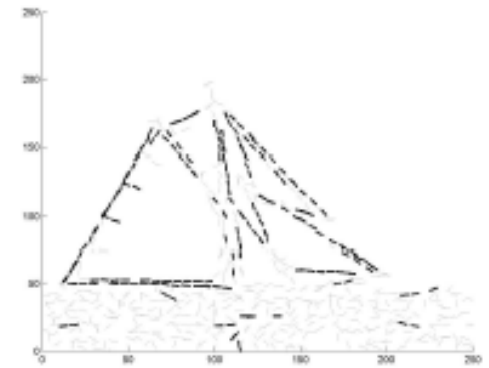
Original



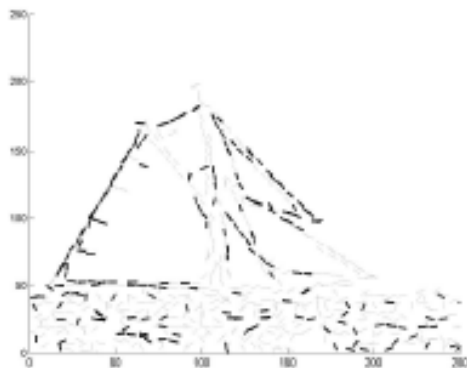
WT



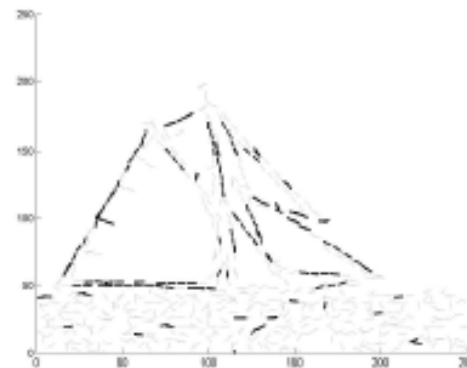
ESS+WT



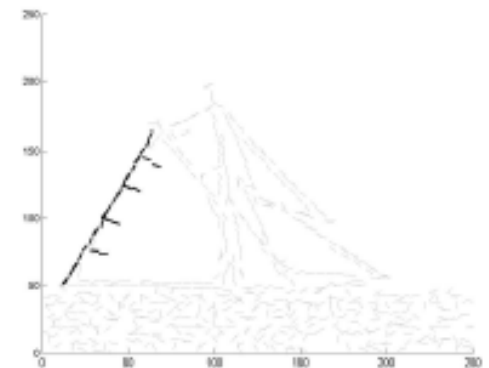
ESS+WTSIMM



HH



ESS+HH





# Conclusions

Introduced the dominant-set framework for pairwise data clustering

- **Binary affinities:** maximal cliques
- **Symmetric affinities:** maxima of quadratic function over standard simplex
- **Arbitrary affinities:** Nash equilibria of non-cooperative games



# Other Applications of Dominant-Set Clustering

## **Bioinformatics:**

Identification of protein binding sites (*Zauhar and Bruist, 2005*)

Clustering gene expression profiles (*Li et al, 2005*)

Tag Single Nucleotide Polymorphism (SNPs) selection (*Frommlet, 2008*)

## **Security and video surveillance:**

Detection of anomalous activities in video streams (*Hamid et al., CVPR'05; Al'09*)

Detection of malicious activities in the internet (*Pouget et al., J. Inf. Ass. Sec. 2006*)

## **Content-based image retrieval:**

*Wang et al. (Sig. Proc. 2008); Giacinto and Roli (2007)*

## **Human action recognition:**

*Wei et al. (ICIP'07)*

## **Analysis of fMRI data:**

*Neumann et al (NeuroImage 2006); Muller et al (J. Mag Res Imag. 2007)*

## **Object tracking:**

*Gualdi et al. (IWVS'08)*





# On-going and Future Work

- Enumerating dominant sets for “soft” clustering (ICPR’08)
- Using high-order affinities for hypergraph clustering
- Using non-linear payoff functions
- Using alternative equilibrium concepts and game dynamics
- Relations with spectral methods?

## **Long-term goal:**

To undertake a thorough study of how game-theoretic notions and models can be applied to pattern analysis and classification (the *SIMBAD* project).



# EU-FP7 FET Project

## (2008 - 2010)



Beyond Features:  
Similarity-Based Pattern Analysis and Recognition  
(<http://simbad-fp7.eu>)

### **Consortium**

1. Ca' Foscari University, Venice, Italy (*M. Pelillo*) - *coordinator*
2. University of York, England (*E. Hancock*)
3. Delft University of Technology, The Netherlands (*B. Duin*)
4. Instituto Superior Técnico, Lisbon, Portugal (*M. Figueiredo*)
5. University of Verona (*V. Murino*)
6. ETH Zurich, Switzerland (*J. Buhmann*)

***We're looking for post-docs!***



# References

M. Pavan, M. Pelillo.

A new graph-theoretic approach to clustering and segmentation. *CVPR 2003*.

M. Pavan, M. Pelillo.

Dominant sets and hierarchical clustering. *ICCV 2003*.

M. Pavan, M. Pelillo.

Efficient out-of-sample extension of dominant-set clusters. *NIPS 2004*.

A. Torsello, S. Rota Bulò, M. Pelillo.

Grouping with asymmetric affinities: A game-theoretic perspective. *CVPR 2006*.

M. Pavan, M. Pelillo.

Dominant sets and pairwise clustering. *PAMI 2007*.

A. Torsello, S. Rota Bulò, M. Pelillo.

Beyond partitions: Allowing overlapping groups in pairwise clustering. *ICPR 2008*.