What is a Cluster?

A Game-Theoretic Perspective

Marcello Pelillo

Ca' Foscari University, Venice

Colloquia@IASI, Rome, 4 February 2013

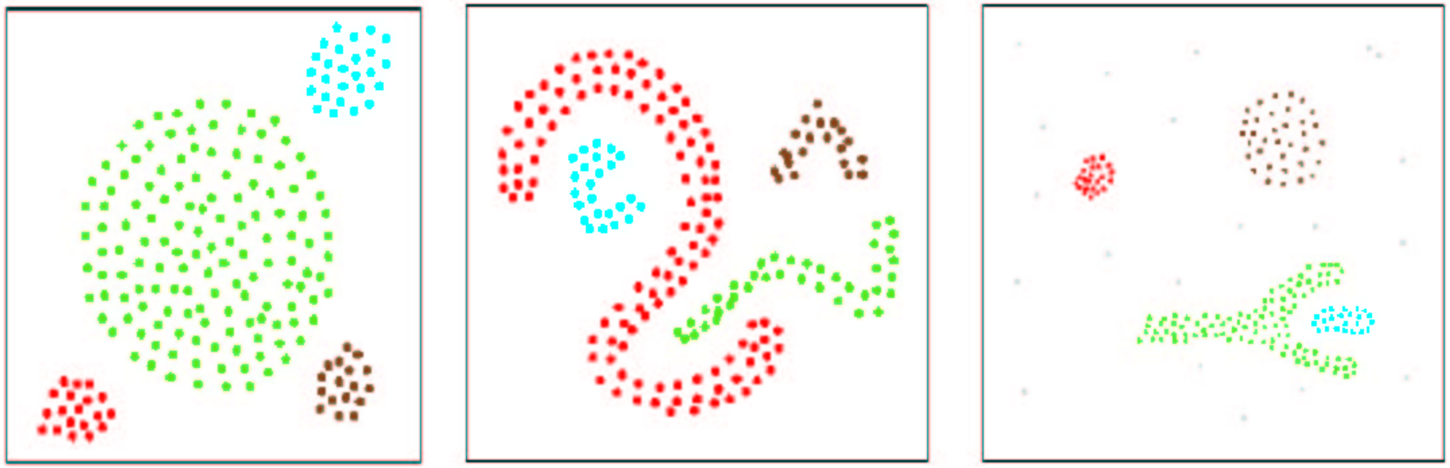
The "Classical" Clustering Problem

Given:

- a set of n "objects"
- an $n \times n$ matrix A of pairwise similarities

} = an edge-weighted graph

Goal: *Partition* the input objects (the vertices of the graph) 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
- Document analysis
- Medical image analysis
- Data mining
- Signal processing
- …

For a review see, e.g., A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters* 31(8):651-666, 2010.

The Need for Non-exhaustive Clusterings

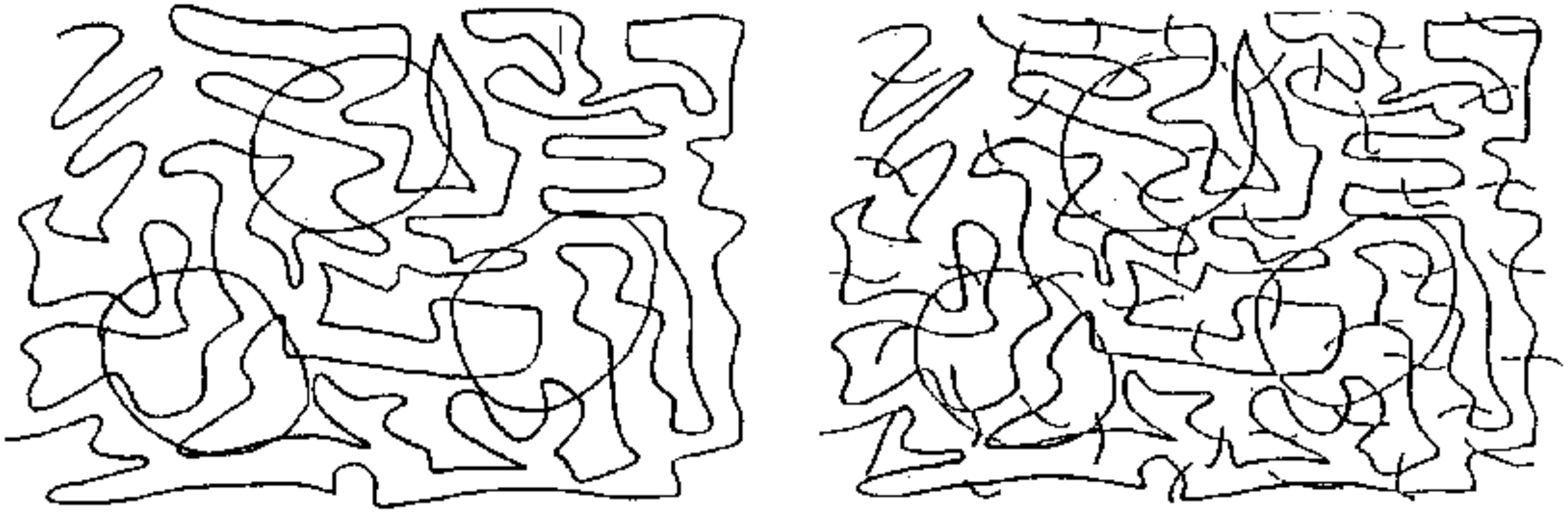


Figure 1a. Three prominent blobs are perceived immediately and with little effort. Locally, the blobs are similar to the background contours. (adopted from Mahoney (1986)

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

Separating Structure from Clutter

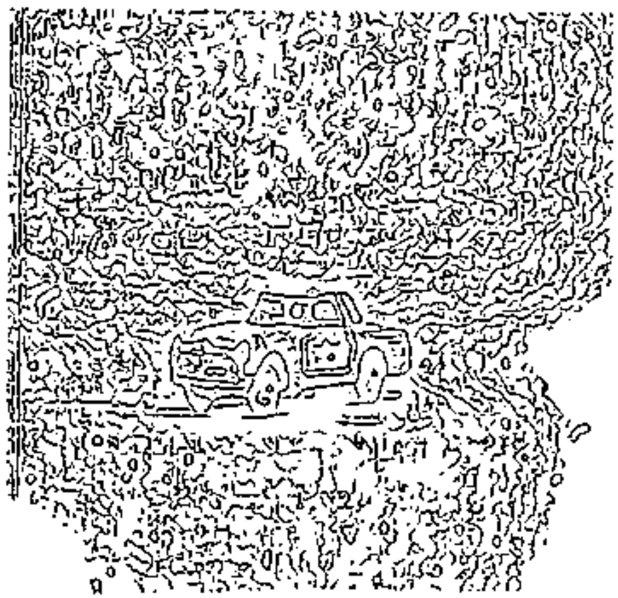
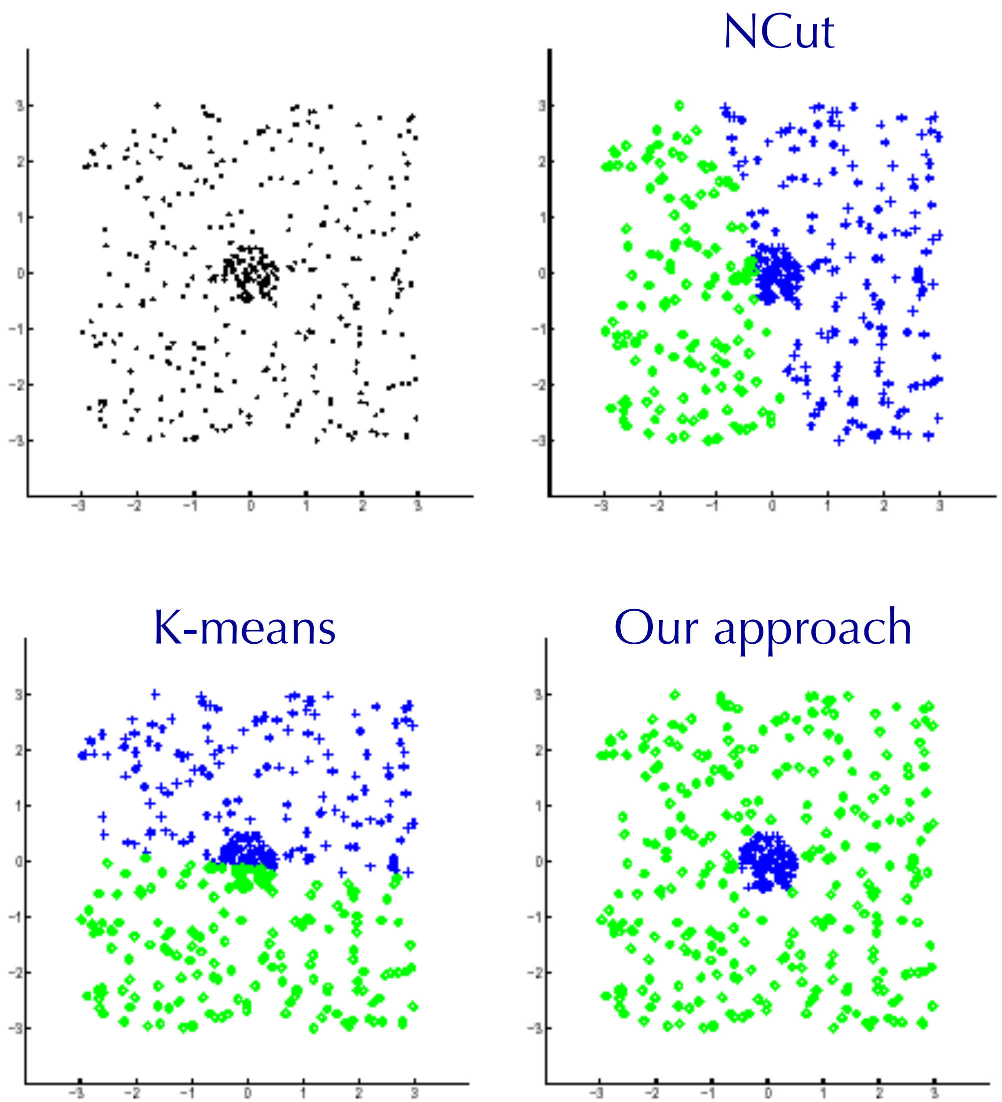


Figure 2. A circle in a background of 200 randomly placed and oriented segments The circle is still perceived immediately although its contour is fragmented.

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



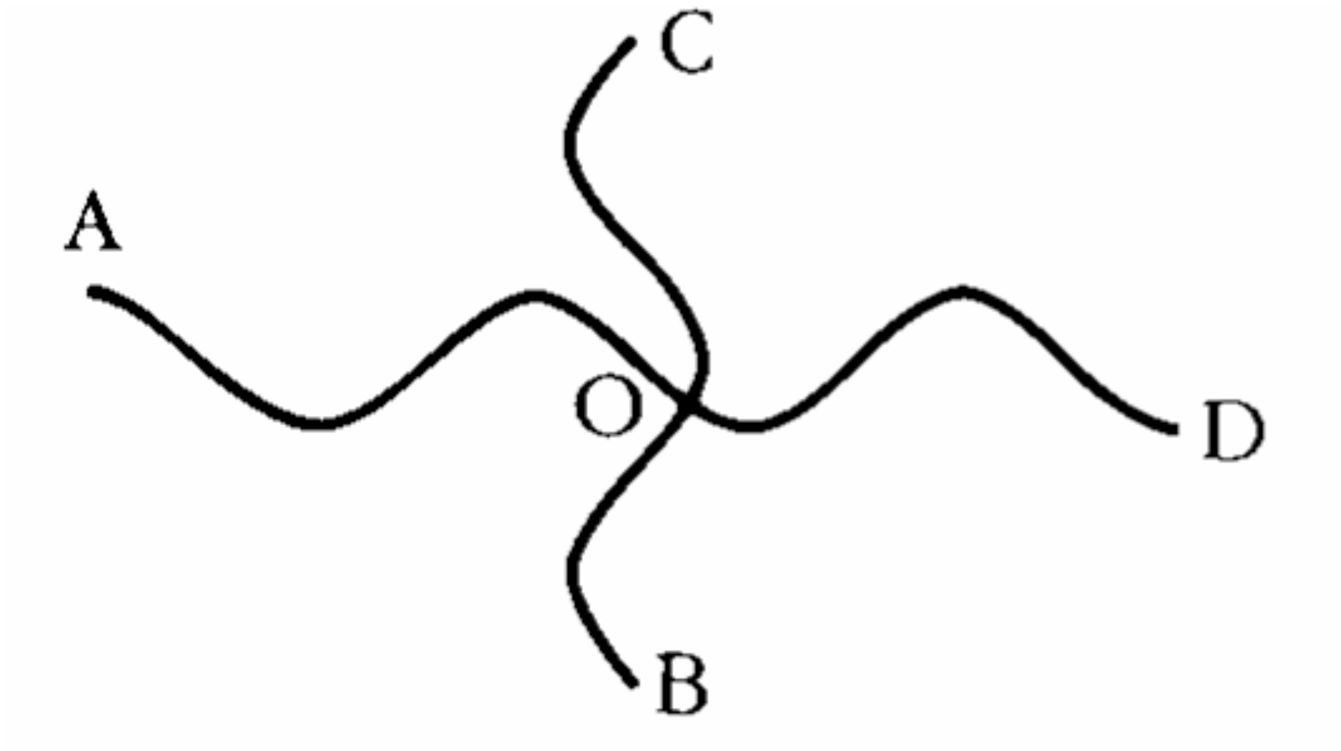
One-class Clustering

"[…] in certain real-world problems, natural groupings are found among only on a small subset of the data, while the rest of the data shows little or no clustering tendencies.

In such situations it is often more important to cluster a small subset of the data very well, rather than optimizing a clustering criterion over all the data points, particularly in application scenarios where a large amount of noisy data is encountered."

G. Gupta and J. Ghosh. Bregman bubble clustering: A robust framework for mining dense cluster. *ACM Trans. Knowl. Discov. Data* (2008).

When Groups Overlap



Does *O* belong to *AD* or to *BC* (or to none)?

The Need for Overlapping Clusters

Partitional approaches impose that each element cannot belong to more than one cluster. There are a variety of important applications, however, where this requirement is too restrictive.

Examples:

- ✓ clustering micro-array gene expression data
- ✓ clustering documents into topic categories
- ✓ perceptual grouping
- ✓ segmentation of images with transparent surfaces

References:

- ✓ N. Jardine and R. Sibson. The construction of hierarchic and non-hierarchic classifications. *Computer Journal*, 11:177–184, 1968
- ✓ A. Banerjee, C. Krumpelman, S. Basu, R. J. Mooney, and J. Ghosh. Model-based overlapping clustering. *KDD 2005*.
- ✓ K. A. Heller and Z. Ghahramani. A nonparametric Bayesian approach to modeling overlapping clusters. *AISTATS 2007*.

The Symmetry Assumption

«Similarity has been viewed by both philosophers and psychologists as a prime example of a symmetric relation. Indeed, the assumption of symmetry underlies essentially all theoretical treatments of similarity.

Contrary to this tradition, the present paper provides empirical evidence for asymmetric similarities and argues that **similarity should not be treated as a symmetric relation**.»

Amos Tversky

"Features of similarities," *Psychol. Rev.* (1977)

Examples of asymmetric (dis)similarities

- ✓ Kullback-Leibler divergence
- ✓ Directed Hausdorff distance
- ✓ Tversky's contrast model

What is a Cluster?

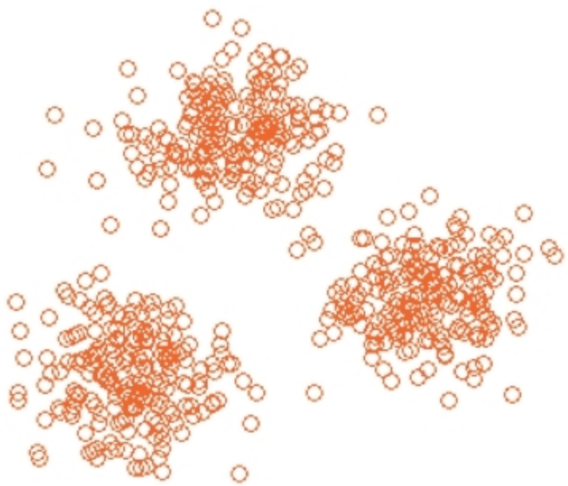
No universally accepted (formal) definition of a “cluster” but, informally, a cluster should satisfy two criteria:

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



Data Clustering: Old *vs.* New

By answering the question "what is a cluster?" we get a novel way of looking at the clustering problem.

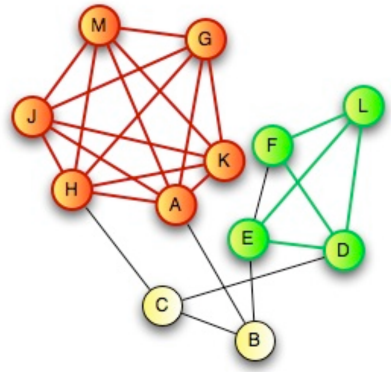
```
Clustering_old(V,A,k)
   V1,V2,...,Vk <- My_favorite_partitioning_algorithm(V,A,k)
   return V1,V2,...,Vk

------

Clustering_new(V,A)
   V1,V2,...,Vk <- Enumerate_all_clusters(V,A)
   return V1,V2,...,Vk

Enumerate_all_clusters(V,A)
   repeat
      Extract_a_cluster(V,A)
   until all clusters have been found
   return the clusters found
```

A Special Case: Binary Symmetric Affinities



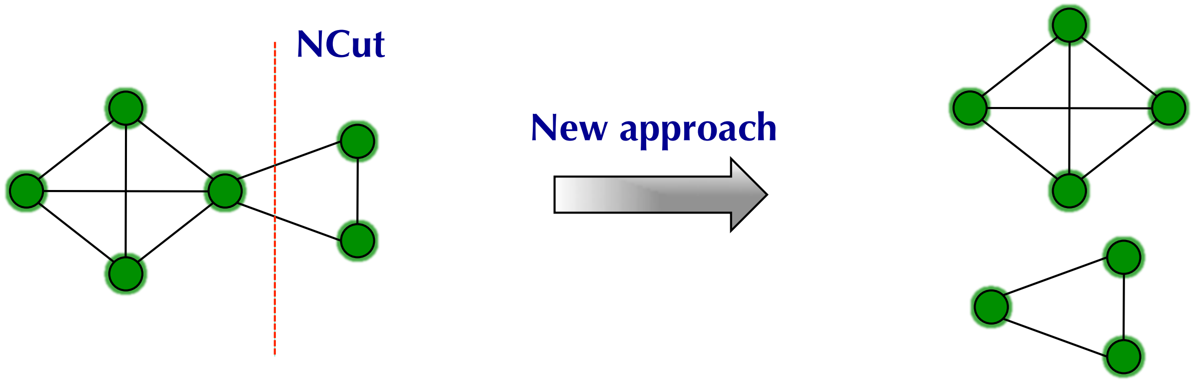
Suppose the similarity matrix is a binary (0/1) matrix.

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

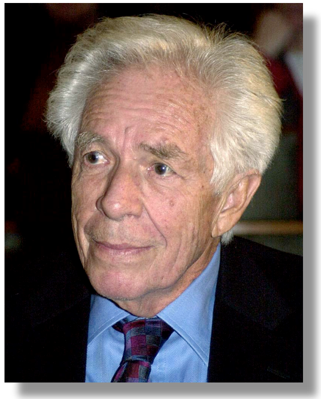
A *clique* is a subset of mutually adjacent vertices

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

In the 0/1 case, a meaningful notion of a cluster is that of a *maximal clique*.



What is Game Theory?



"The central problem of game theory was posed by von Neumann as early as 1926 in Göttingen. It is the following: If n players, $P_1, \ldots, P_n$, play a given game Γ, how must the i^{th} player, P_i, play to achieve the most favorable result for himself?"

Harold W. Kuhn
Lectures on the Theory of Games (1953)

A few cornerstones in game theory

1921–1928: Emile Borel and John von Neumann give the first modern formulation of a mixed strategy along with the idea of finding minimax solutions of normal-form games.

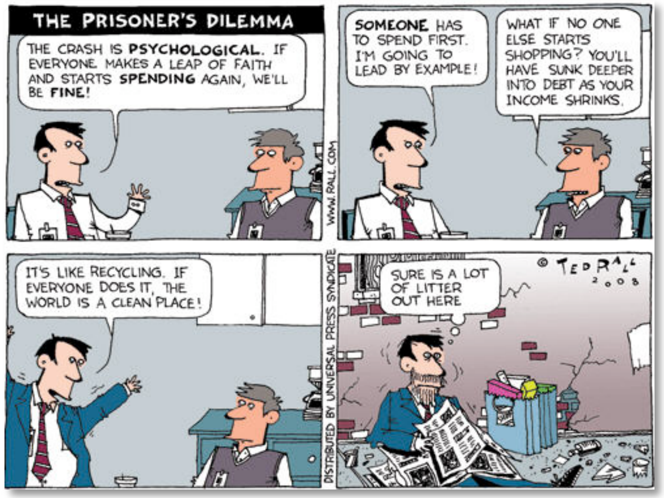
1944, 1947: John von Neumann and Oskar Morgenstern publish *Theory of Games and Economic Behavior*.

1950–1953: In four papers John Nash made seminal contributions to both non-cooperative game theory and to bargaining theory.

1972–1982: John Maynard Smith applies game theory to biological problems thereby founding "evolutionary game theory."

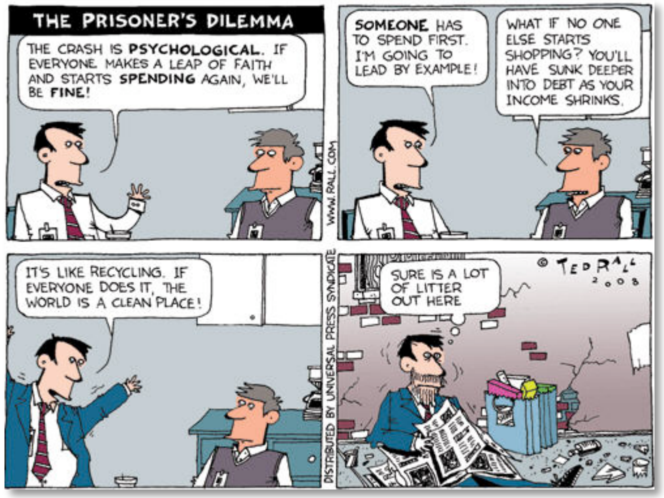
late 1990's –: Development of algorithmic game theory…

Prisoner's Dilemma



		Prisoner 2	
		Confess (defect)	Deny (cooperate)
Prisoner 1	Confess (defect)	-10 , -10	-1 , -25
	Deny (cooperate)	-25 , -1	-3 , -3

How to "Solve" the Game?



Dominated (pure) strategy !

		Prisoner 2	
		Confess (defect)	Deny (cooperate)
Prisoner 1	Confess (defect)	-10 , -10	-1 , -25
	Deny (cooperate)	-25 , -1	-3 , -3

Dominated (pure) strategy !

Basics of (Two-Player, Symmetric) Game Theory

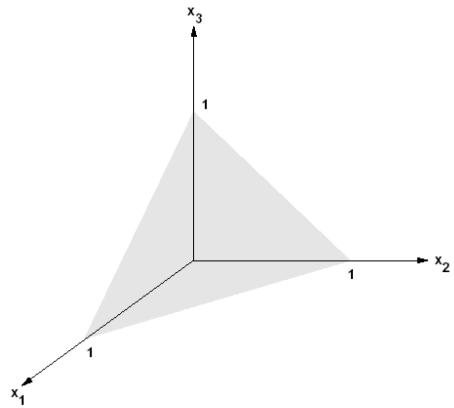
Assume:

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

Each player receives a payoff depending on the strategies selected by him and by the adversary. Players' goal is to maximize their own returns.

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

$$\Delta = \left\{ x \in R^n : \forall i = 1 \ldots n : x_i \geq 0, \text{ and } \sum_{i=1}^{n} x_i = 1 \right\}$$



Nash Equilibria

- Let A be an arbitrary **payoff** matrix: a_{ij} is the payoff obtained by playing i while the opponent plays j.

- The average payoff obtained by playing mixed strategy $\mathbf{y}$ while the opponent plays $\mathbf{x}$, is:

$$\mathbf{y}'\mathbf{A}\mathbf{x} = \sum_i \sum_j a_{ij} y_i x_j$$

- A mixed strategy $\mathbf{x}$ is a (symmetric) **Nash equilibrium** if

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq \mathbf{y}'\mathbf{A}\mathbf{x}$$

for all strategies $\mathbf{y}$. (Best reply to itself.)

Existence of Nash Equilibria

Theorem (Nash, 1951). Every finite normal-form game admits a mixed-strategy Nash equilibrium.

Idea of proof.

1. Define a continuous map T on Δ such that the fixed points of T are in one-to-one correspondence with Nash equilibria.

2. Use Brouwer's theorem to prove existence of a fixed point.

"Together with factoring, *the complexity of finding a Nash equilibrium is in my opinion the most important concrete open question on the boundary of P today."*

Christos Papadimitriou

Algorithms, games, and the internet (2001)

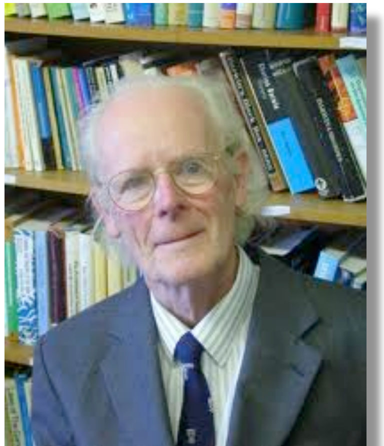
Evolution and the Theory of Games

"We repeat most emphatically that our theory is thoroughly static. A dynamic theory would unquestionably be more complete and therefore preferable.

But there is ample evidence from other branches of science that it is futile to try to build one as long as the static side is not thoroughly understood."

John von Neumann and Oskar Morgenstern

Theory of Games and Economic Behavior (1944)



"Paradoxically, it has turned out that game theory is more readily applied to biology than to the field of economic behaviour for which it was originally designed."

John Maynard Smith

Evolution and the Theory of Games (1982)

Evolutionary Games and ESS's

Assumptions:

- ✓ A large population of individuals belonging to the same species which compete for a particular limited resource
- ✓ This kind of conflict is modeled as a symmetric two-player game, the players being pairs of randomly selected population members
- ✓ Players do not behave "rationally" but act according to a pre-programmed behavioral pattern (pure strategy)
- ✓ Reproduction is assumed to be asexual
- ✓ Utility is measured in terms of Darwinian fitness, or reproductive success

A Nash equilibrium **x** is an ***Evolutionary Stable Strategy*** (ESS) if, for all strategies y:

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

Note: Unlike Nash equilibria, existence of ESS's is not guaranteed.

ESS's as Clusters

We claim that ESS's abstract 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.

Basic Definitions

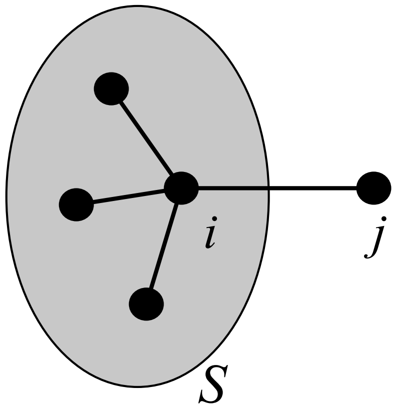
Let $S \subseteq V$ be a non-empty subset of vertices, and $i \in S$.

The **(average) weighted degree** of i w.r.t. S is defined as:

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

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

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



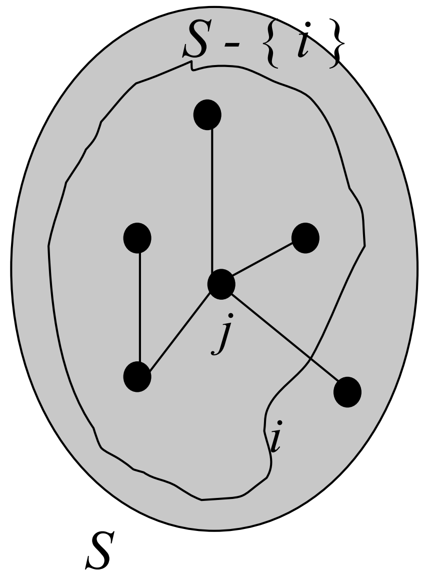
Intuitively, $\phi_S(i,j)$ measures the similarity between vertices j and i, with respect to the (average) similarity between vertex i and its neighbors in S.

Assigning Weights to Vertices

Let $S \subseteq V$ be a non-empty subset of vertices, and $i \in S$.

The **weight** of i w.r.t. S is defined as:

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

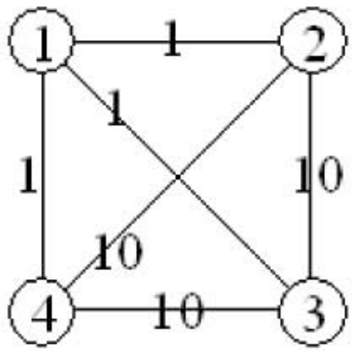


Further, the **total weight** of S is defined as:

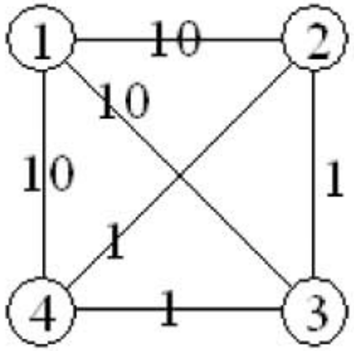
$$W(S) = \sum_{i \in S} w_S(i)$$

Interpretation

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



$w_{\{1,2,3,4\}}(1) < 0$

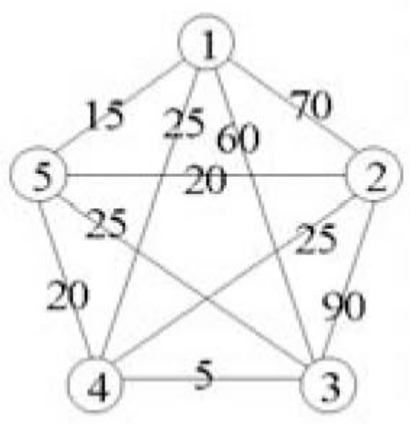


$w_{\{1,2,3,4\}}(1) > 0$

Dominant Sets

Definition (Pavan and Pelillo, 2003, 2007). 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 a **dominant set** 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 homogeneity)



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

Dominant sets ≡ clusters

The Clustering Game

Consider the following "clustering game."

- ✓ Assume a preexisting set of objects *O* and a (possibly asymmetric) matrix of affinities A between the elements of *O*.
- ✓ Two players with complete knowledge of the setup play by simultaneously selecting an element of *O*.
- ✓ After both have shown their choice, each player receives a payoff, monetary or otherwise, proportional to the affinity that the chosen element has 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.

Hence, in the (pairwise) clustering game:

- ✓ There are 2 players (because we have pairwise affinities)
- ✓ The objects to be clustered are the pure strategies
- ✓ The (null-diagonal) affinity matrix coincides with the similarity matrix

Dominant Sets are ESS's

Theorem (Torsello, Rota Bulò and Pelillo, 2006). Evolutionary stable strategies of the clustering game with affinity matrix A are in a one-to-one correspondence with dominant sets.

Note. Generalization of well-known Motzkin-Straus theorem from graph theory.

Dominant-set clustering

- ✓ To get a single dominant-set cluster use, e.g., replicator dynamics (but see Rota Bulò, Pelillo and Bomze, *CVIU* 2011, for faster dynamics)
- ✓ To get a partition use a simple *peel-off* strategy: iteratively find a dominant set and remove it from the graph, until all vertices have been clustered
- ✓ To get overlapping clusters, enumerate dominant sets (see Bomze, 1992; Torsello, Rota Bulò and Pelillo, 2008)

Special Case: Symmetric Affinities

Given a symmetric real-valued matrix A (with null diagonal), consider the following Standard Quadratic Programming problem (StQP):

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

Note. The function $f(x)$ provides a measure of cohesiveness of a cluster (see Pavan and Pelillo, 2003, 2007; Sarkar and Boyer, 1998; Perona and Freeman, 1998).

ESS's are in one-to-one correspondence to (strict) local solutions of StQP

Note. In the 0/1 (symmetric) case, ESS's are in one-to-one correspondence to (strictly) maximal cliques (Motzkin-Straus theorem).

Replicator Dynamics

Let $x_i(t)$ the population share playing pure strategy *i* at *time t*. The **state** of the population at time *t* is: $x(t)= (x_1(t),\dots,x_n(t))\in\Delta$.

Replicator dynamics (Taylor and Jonker, 1978) are motivated by Darwin's principle of natural selection:

$$\frac{\dot{x}_i}{x_i} \propto \text{payoff of pure strategy i} - \text{average population payoff}$$

which yields:

$$\dot{x}_i = x_i\left[(Ax)_i - x^T Ax\right]$$

Theorem (Nachbar, 1990; Taylor and Jonker, 1978). A point $x\in\Delta$ is a Nash equilibrium if and only if x is the limit point of a replicator dynamics trajectory starting from the interior of Δ.

Furthermore, if $x\in\Delta$ is an ESS, then it is an asymptotically stable equilibrium point for the replicator dynamics.

Doubly Symmetric Games

In a doubly symmetric (or partnership) game, the payoff matrix A is symmetric ($A = A^T$).

Fundamental Theorem of Natural Selection (Losert and Akin, 1983).

For any doubly symmetric game, the average population payoff $f(x) = x^T A x$ is strictly increasing along any non-constant trajectory of replicator dynamics, namely, $d/dt f(x(t)) \geq 0$ for all $t \geq 0$, with equality if and only if $x(t)$ is a stationary point.

Characterization of ESS's (Hofbauer and Sigmund, 1988)

For any doubly simmetric game with payoff matrix A, the following statements are equivalent:

a) $x \in \Delta^{ESS}$

b) $x \in \Delta$ is a strict local maximizer of $f(x) = x^T A x$ over the standard simplex Δ

c) $x \in \Delta$ is asymptotically stable in the replicator dynamics

Discrete-time Replicator Dynamics

A well-known discretization of replicator dynamics, which assumes non-overlapping generations, is the following (assuming a non-negative A):

$$x_i(t+1) = x_i(t)\frac{A\big(x(t)\big)_i}{x(t)^T Ax(t)}$$

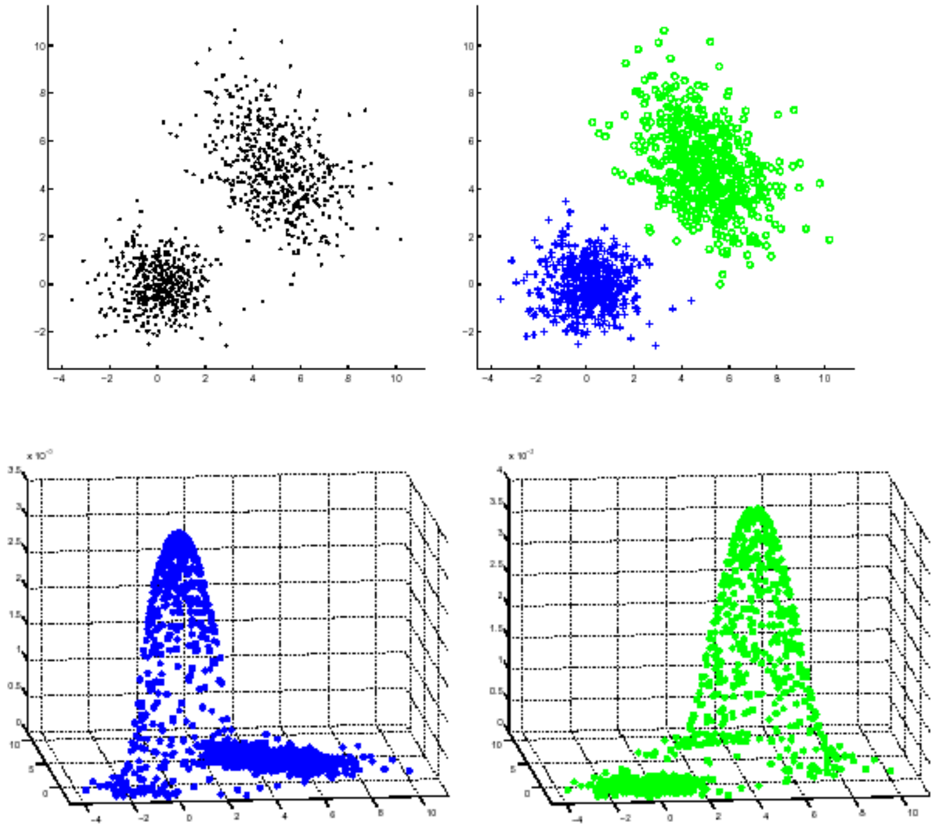
which inherits most of the dynamical properties of its continuous-time counterpart (*e.g.*, the fundamental theorem of natural selection).

MATLAB implementation

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

Measuring the Degree of Cluster Membership

The components of the converged vector give us a measure of the participation of the corresponding vertices in the cluster, while the value of the objective function provides of the cohesiveness of the cluster.



Application to Image Segmentation

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

For the sake of comparison, in the experiments we used the same similarities used in Shi and Malik's normalized-cut paper (PAMI 2000).

To find a hard partition, the following *peel-off* strategy was used:

```
Partition_into_dominant_sets(G)
Repeat
   find a dominant set
   remove it from graph
until all vertices have been clustered
```

To find a single dominant set we used replicator dynamics (but see Rota Bulò, Pelillo and Bomze, *CVIU 2011*, for faster game dynamics).

Experimental Setup

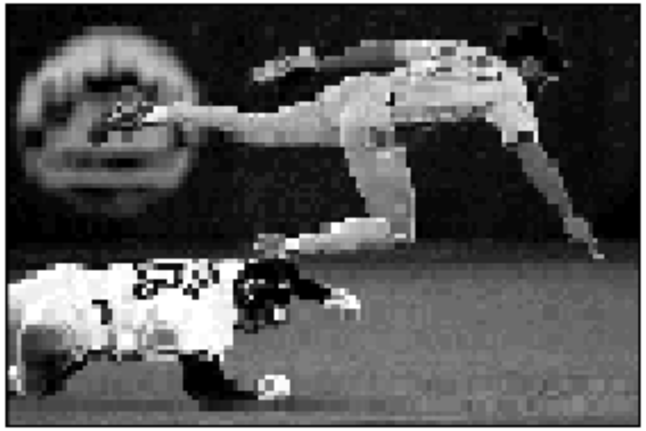
The similarity between pixels i and j was measured by:

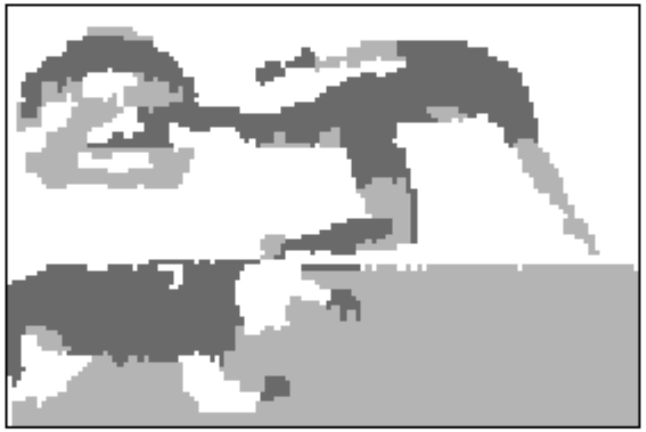
$$w(i,j) = \exp\left(\frac{-\|\mathbf{F}(i) - \mathbf{F}(j)\|_2^2}{\sigma^2}\right)$$

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

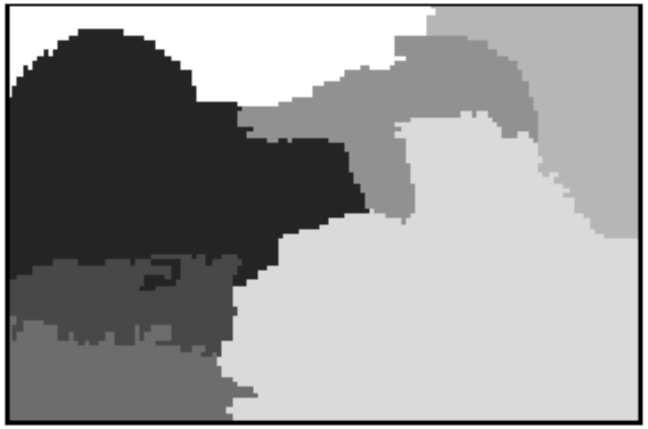
- $\mathbf{F}(i) \equiv$ (normalized) intensity of pixel i, for intensity segmentation
- $\mathbf{F}(i) = [v, vs\sin(h), vs\cos(h)](i)$, where h, s, v are the HSV values of pixel i, for color segmentation
- $\mathbf{F}(i) = [|I * f_1|, \ldots, |I * f_k|](i)$ is a vector based on texture information at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

Intensity Segmentation Results



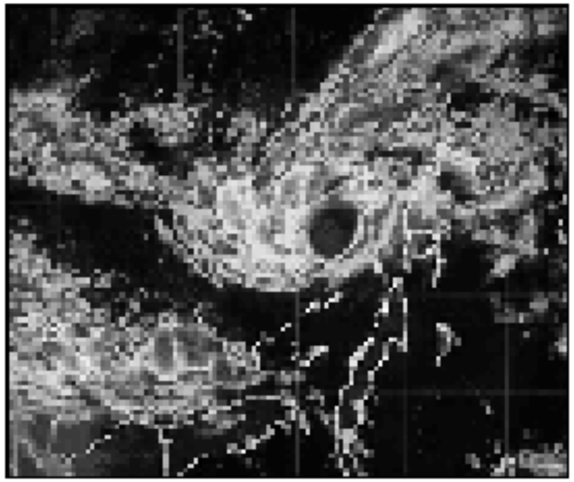


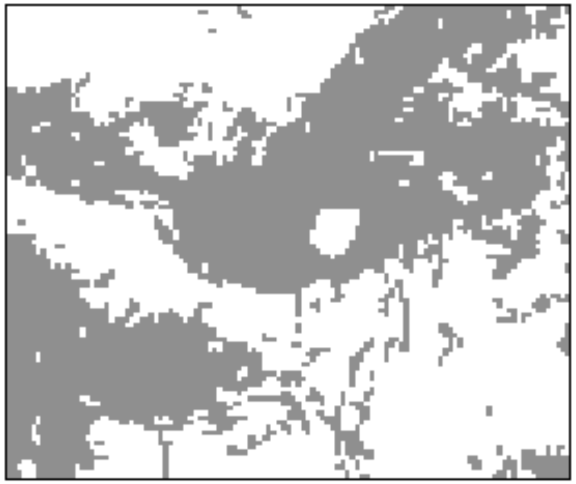
Dominant sets



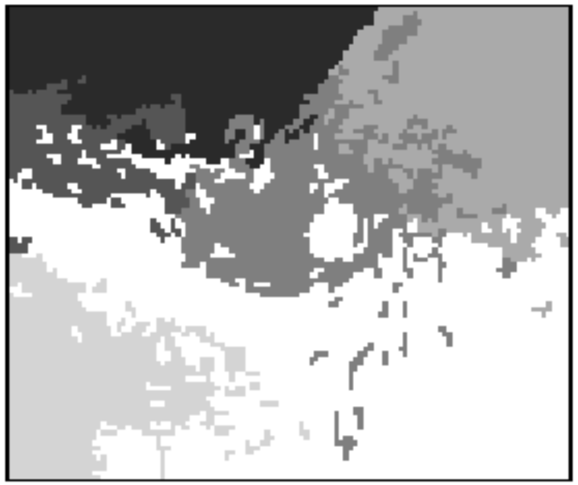
Ncut

Intensity Segmentation Results



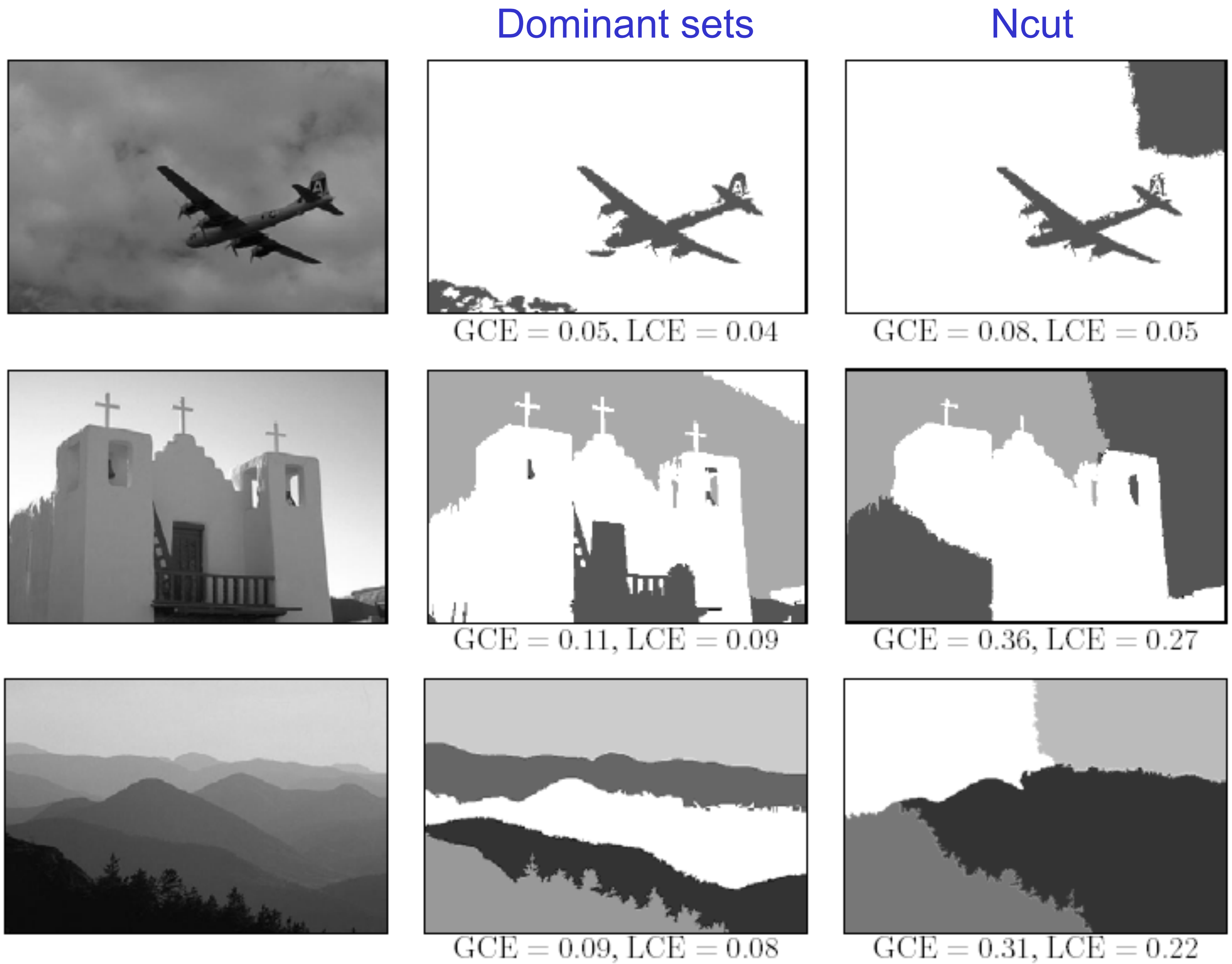


Dominant sets

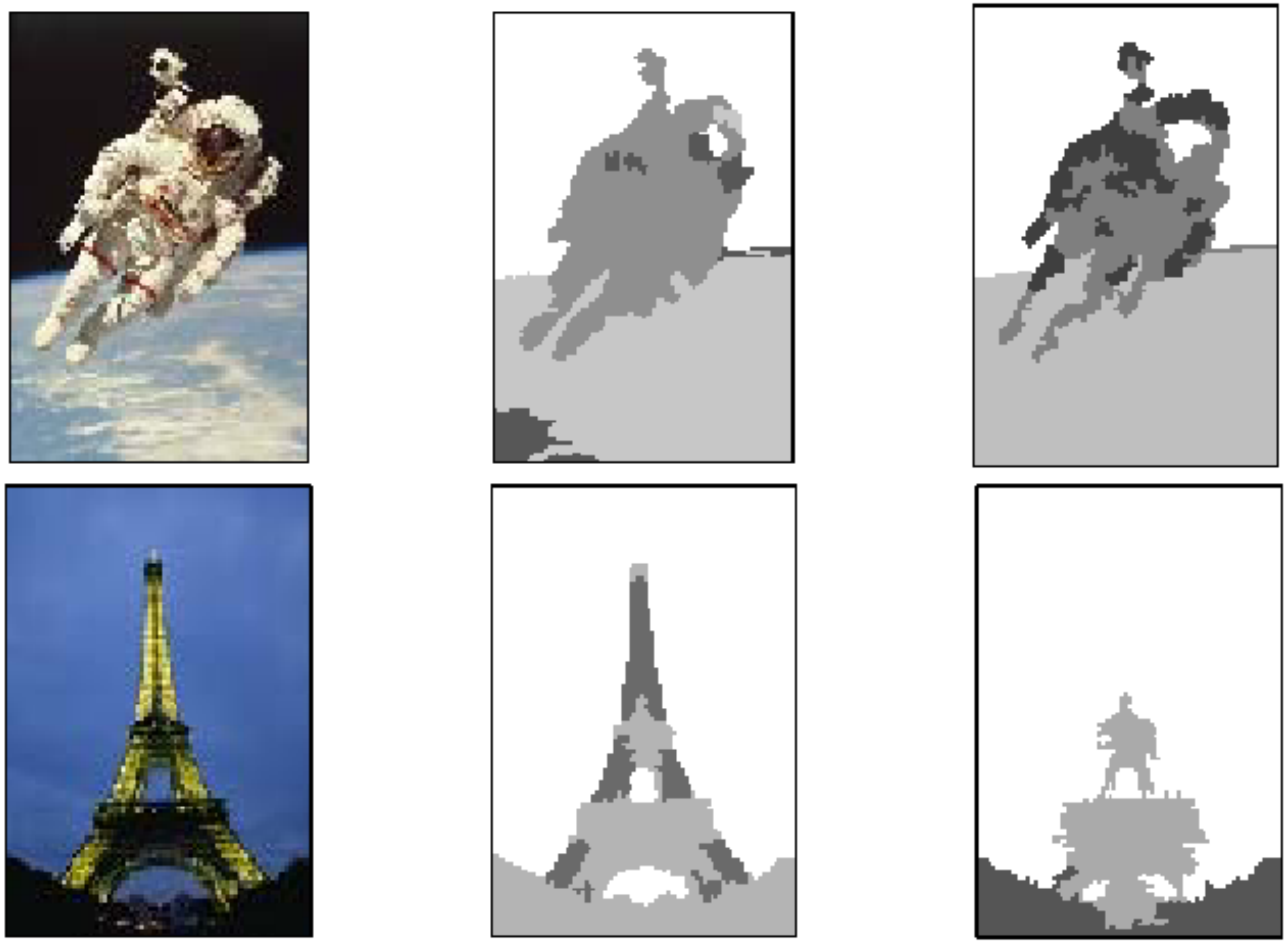


Ncut

Results on the Berkeley Dataset



Color Segmentation Results

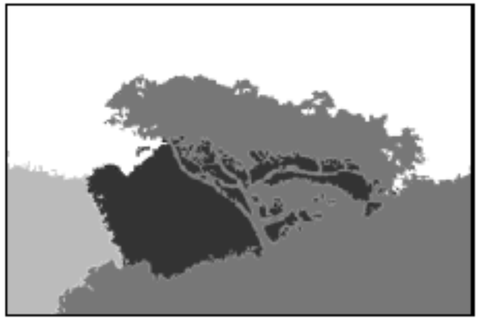


Original image | Dominant sets | Ncut

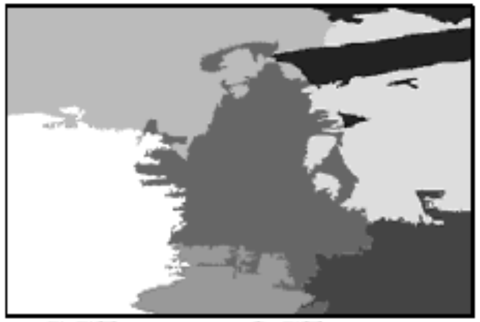
Results on the Berkeley Dataset

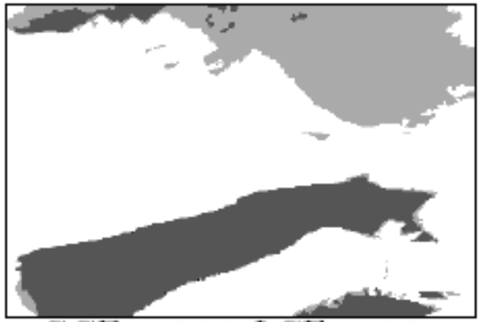
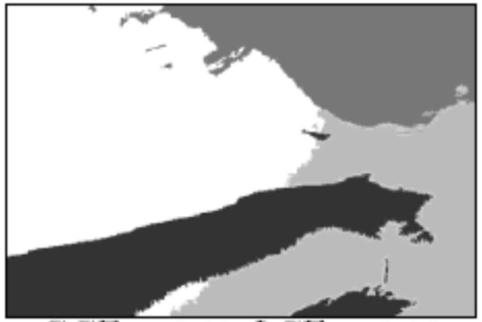
Dominant sets | Ncut



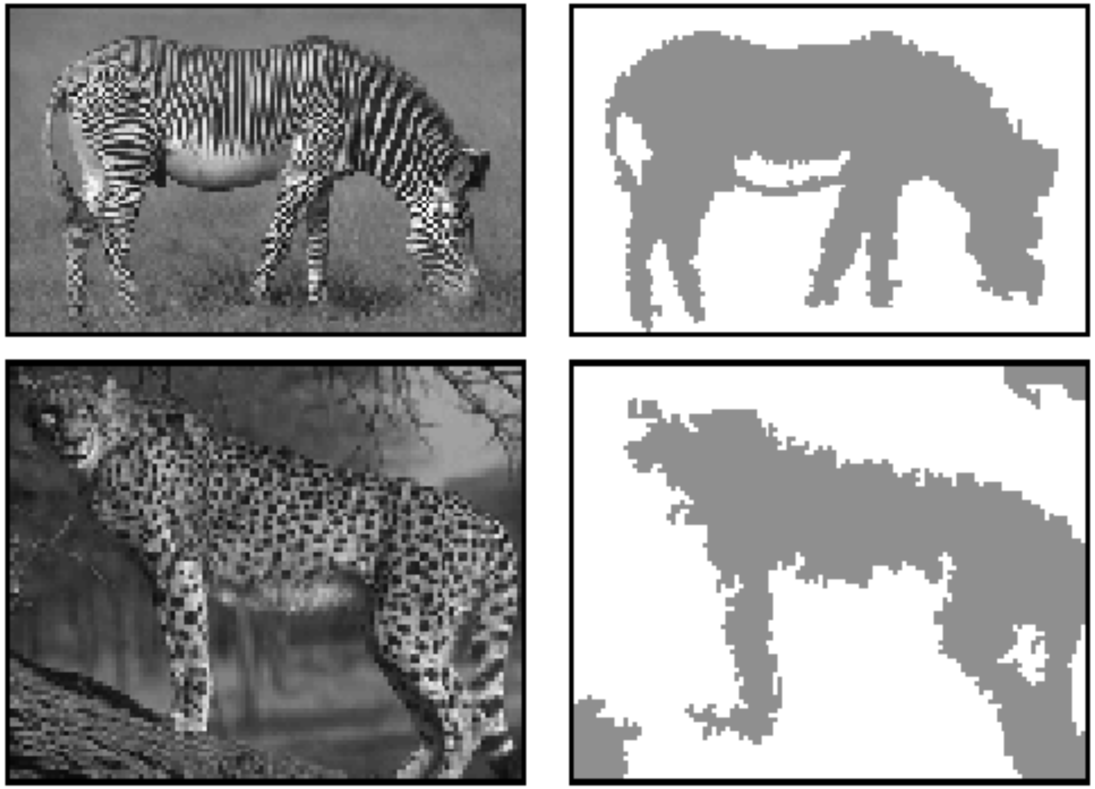
GCE = 0.12, LCE = 0.12

GCE = 0.19, LCE = 0.13



GCE = 0.31, LCE = 0.26

GCE = 0.35, LCE = 0.29

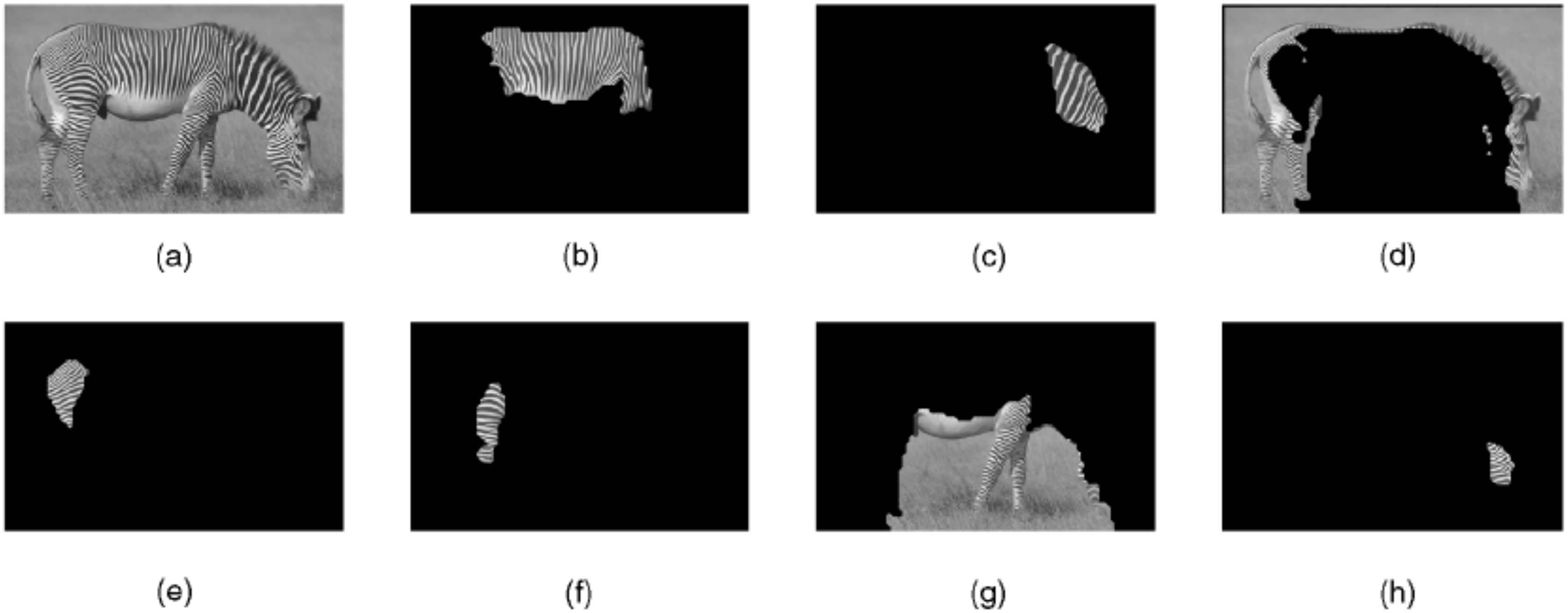


GCE = 0.09, LCE = 0.09

GCE = 0.16, LCE = 0.16

Texture Segmentation Results



Dominant sets

Texture Segmentation Results



NCut

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*, 2010)

Security and video surveillance

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

Analysis of fMRI data

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

Video analysis, object tracking, human action recognition

Torsello et al. (EMMCVPR'05); *Gualdi et al.* (IWVS'08); *Wei et al.* (ICIP'07)

Multiple instance learning

Erdem and Erdem (SIMBAD'11)

Feature selection

Hancock et al. (GbR'11; ICIAP'11; SIMBAD'11)

Image matching and registration

Torsello et al. (IJCV 2011, ICCV'09, CVPR'10, ECCV'10)

In a nutshell…

The dominant-set (ESS) approach:

- ✓ makes no assumption on the underlying (individual) data representation
- ✓ makes no assumption on the structure of the affinity matrix, being it able to work with asymmetric and even negative similarity functions
- ✓ does not require *a priori* knowledge on the number of clusters (since it extracts them sequentially)
- ✓ leaves clutter elements unassigned (useful, e.g., in figure/ground separation or one-class clustering problems)
- ✓ allows principled ways of assigning out-of-sample items *(NIPS'04)*
- ✓ allows extracting overlapping clusters *(ICPR'08)*
- ✓ generalizes naturally to hypergraph clustering problems, i.e., in the presence of high-order affinities, in which case the clustering game is played by more than two players *(NIPS'09; PAMI'12, in press)*
- ✓ extends to hierarchical clustering *(ICCV'03: EMMCVPR'09)*

References

Evolutionary game theory

J. Weibull. *Evolutionary Game Theory*. MIT Press (1995).

J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press (1998).

Dominant sets

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

M. Pavan and M. Pelillo. Dominant sets and hierarchical clustering. *ICCV 2003*.

M. Pavan and M. Pelillo. Efficient out-of-sample extension of dominant-set clusters. *NIPS 2004*.

A. Torsello, S. Rota Bulò and M. Pelillo. Grouping with asymmetric affinities: A game-theoretic perspective. *CVPR 2006*.

M. Pavan and M. Pelillo. Dominant sets and pairwise clustering. *PAMI 2007*.

A. Torsello, S. Rota Bulò and M. Pelillo. Beyond partitions: Allowing overlapping groups in pairwise clustering. *ICPR 2008*.

S. Rota Bulò and M. Pelillo. A game-theoretic approach to hypergraph clustering. *NIPS 2009; PAMI'12, in press.*

M. Pelillo. What is a cluster? Perspectives from game theory. *NIPS 2009 Workshop on "Clustering: Science or Art?"* (talk available on videolectures.net).

S. Rota Bulò, M. Pelillo and I. M. Bomze. Graph-based quadratic optimization: A fast evolutionary approach. *CVIU 2011*.