Finding Clusters in Weighted Graphs, Digraphs, and Hypergraphs: A Game-Theoretic Perspective

Marcello Pelillo

Ca' Foscari University, Venice

IWCSN'13, Vancouver, BC, December 12, 2013

The "Classical" Clustering Problem

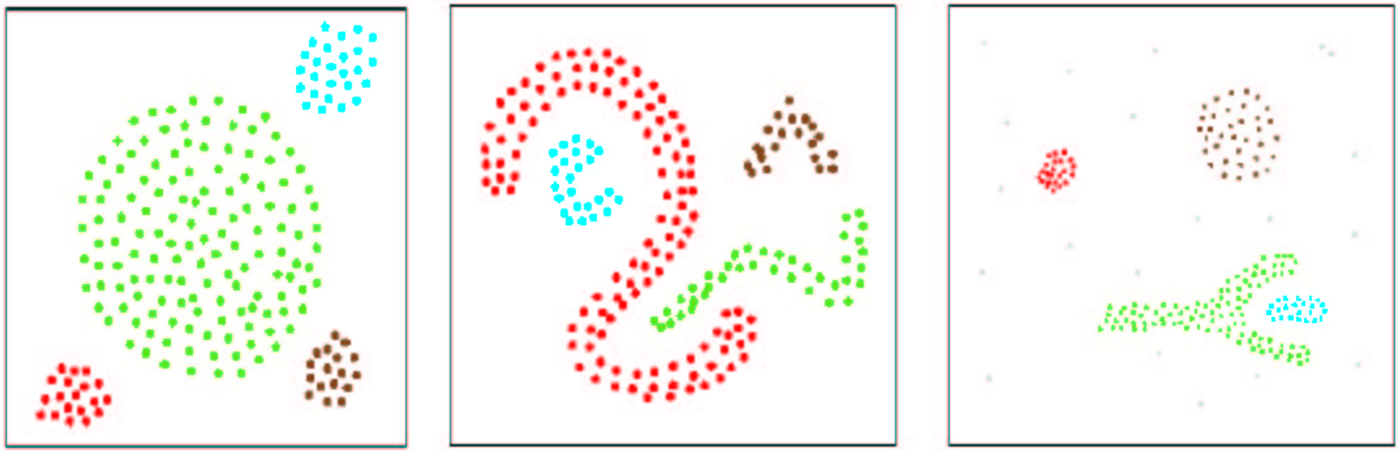
Given:

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

} = an edge-weighted graph G

Goal: <u>*Partition*</u> the vertices of the G into maximally homogeneous groups (i.e., clusters).

Usual assumption: *symmetric* and *pairwise* similarities (G is an undirected graph)



What is a Cluster?

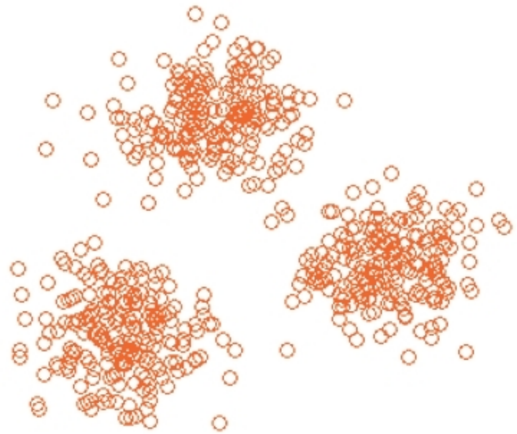
"In most cases, communities are algorithmically defined, i.e. they are just the final product of the algorithm, without a precise a priori definition."

S. Fortunato, "Community detection in graphs," 2010

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

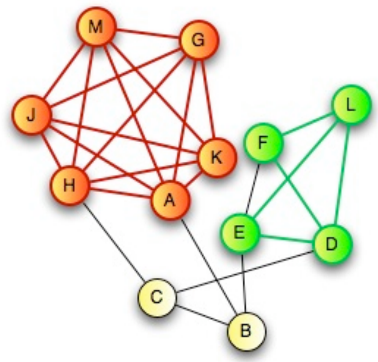
Internal criterion: all "objects" *inside* a cluster should be highly similar to each other

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



A Special Case:
Binary Symmetric Similarities

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 (though strict) notion of a cluster is that of a *maximal clique* (Luce & Perry, 1949).

Advantages of the New Approach

- ✓ No need to know the number of clusters in advance (since we extract them sequentially)
- ✓ Leaves clutter elements unassigned (useful, e.g., in figure/ground separation or one-class clustering problems)
- ✓ Allows extracting overlapping clusters

Need a partition?

```
Partition_into_clusters(V,A)
  repeat
     Extract_a_cluster
     remove it from V
  until all vertices have been clustered
```

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…

"Solving" a Game

		Player 2		
		Left	Middle	Right
Player 1	Top	3 , 1	2 , 3	10 , 2
	High	4 , 5	3 , 0	[illegible] , 4
	Low	2 , 2	5 , 4	12 , 3
	Bottom	5 , 6	4 , 5	9 , 7

Nash equilibrium!

Basics of (Two-Player, Symmetric) Game Theory

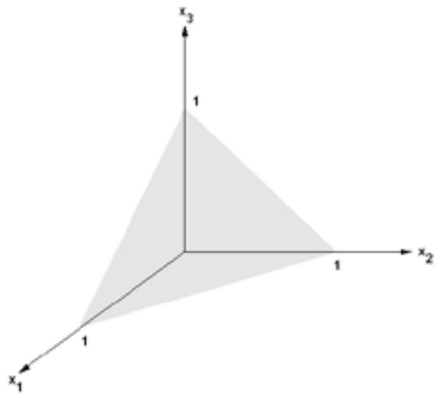
Assume:

- 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 **y** while the opponent plays **x**, is:

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

- A mixed strategy **x** is a (symmetric) **Nash equilibrium** if

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

for all strategies **y**. (Best reply to itself.)

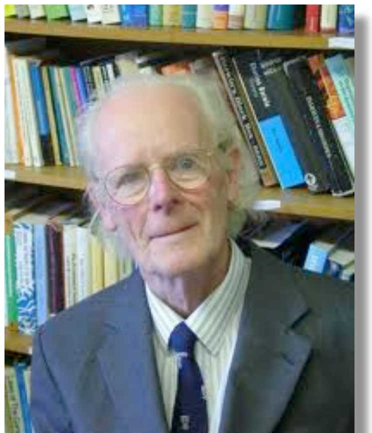
Theorem (Nash, 1951). Every finite normal-form game admits a mixed-strategy Nash equilibrium.

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:

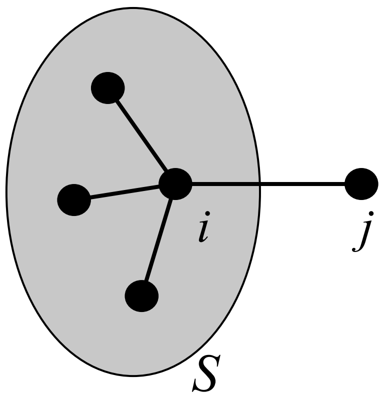
- ✓ **Internal coherency**: High mutual support of all elements within the group.
- ✓ **External incoherency**: Low support from elements of the group to elements outside the group.

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:

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



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

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

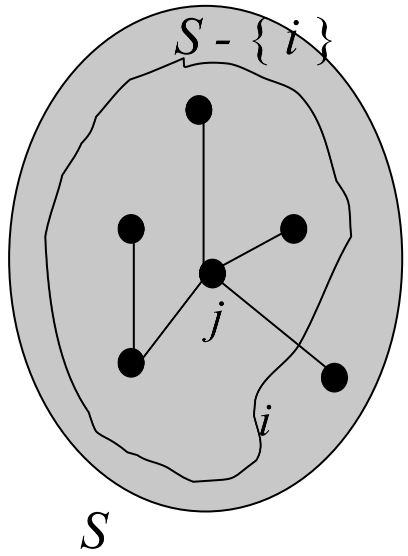
Intuitively, $\phi_S(i,j)$ measures the similarity between vertices j and i, with respect to the (average) similarity between vertex i and its neighbors in S.

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:

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

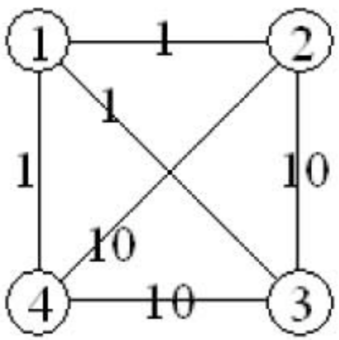


Further, the **total weight** of S is defined as:

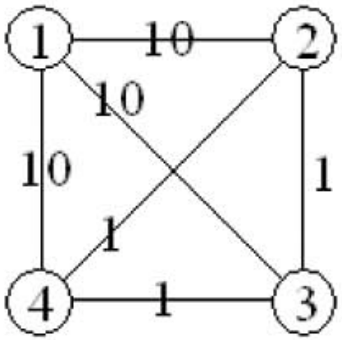
$$W(S) = \sum_{i \in S} w_S(i)$$

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}.



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

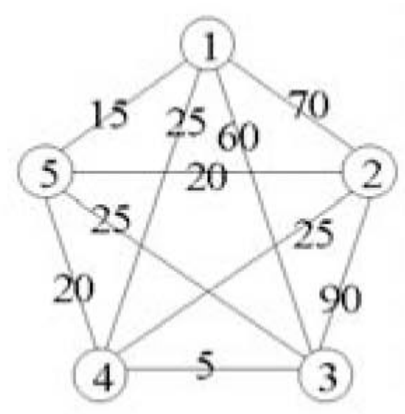


$w_{\{1,2,3,4\}}(1) > 0$

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 play by simultaneously selecting an element of *O*.
- ✓ After both have shown their choice, each player receives a payoff proportional to the affinity that the chosen element has wrt 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 (1965).

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: $x(t)= (x_1(t),\ldots,x_n(t)) \in \Delta$.

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 $f(x) = x^T Ax$ 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 $f(x) = x^T Ax$ 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*):

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

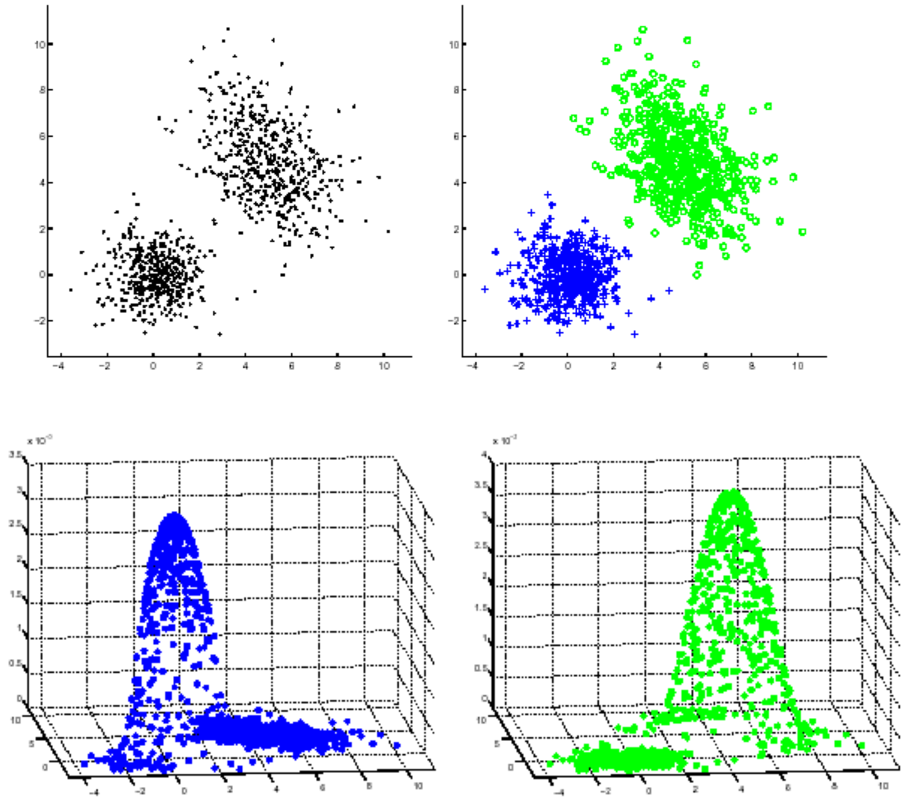
which inherits most of the dynamical properties of its continuous-time counterpart (*e.g.*, the fundamental theorem of natural selection).

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

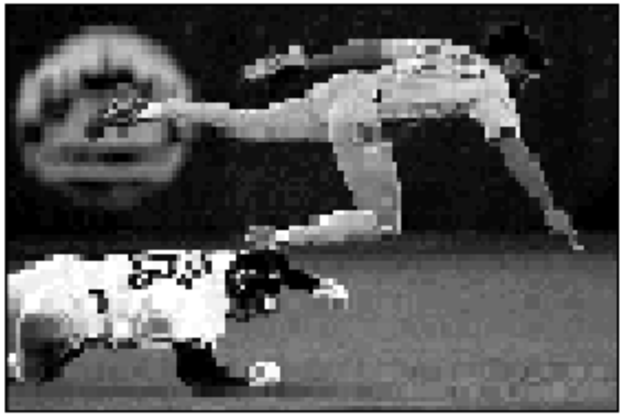
The similarity between pixels i and j was measured by:

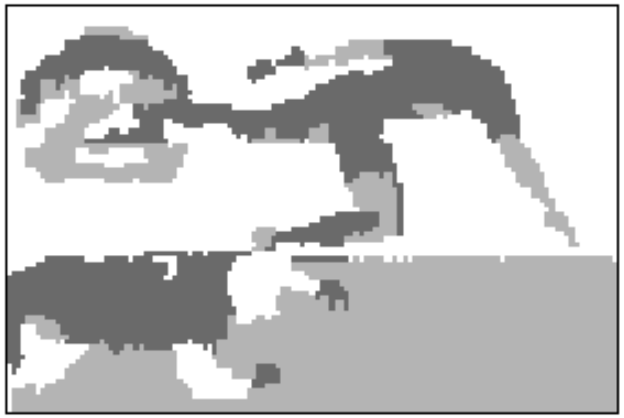
$$w(i,j) = \exp\left(\frac{-\|\mathbf{F}(i) - \mathbf{F}(j)\|_2^2}{\sigma^2}\right)$$

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

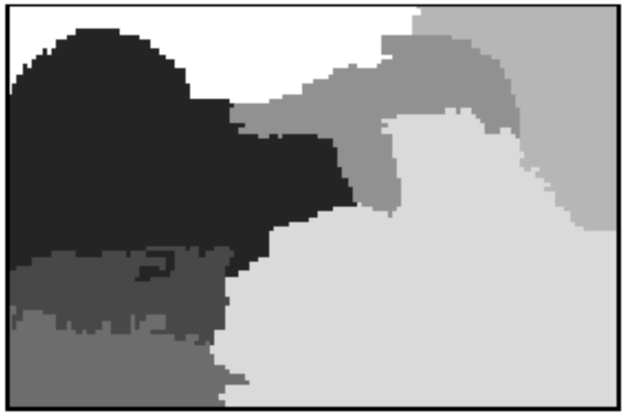
- $\mathbf{F}(i) \equiv$ (normalized) intensity of pixel i, for intensity segmentation
- $\mathbf{F}(i) = [v, vs\sin(h), vs\cos(h)](i)$, where h, s, v are the HSV values of pixel i, for color segmentation
- $\mathbf{F}(i) = [|I * f_1|, \ldots, |I * f_k|](i)$ is a vector based on texture information at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

Intensity Segmentation Results



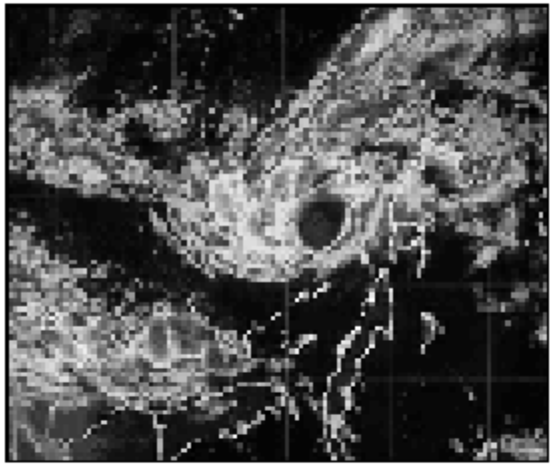


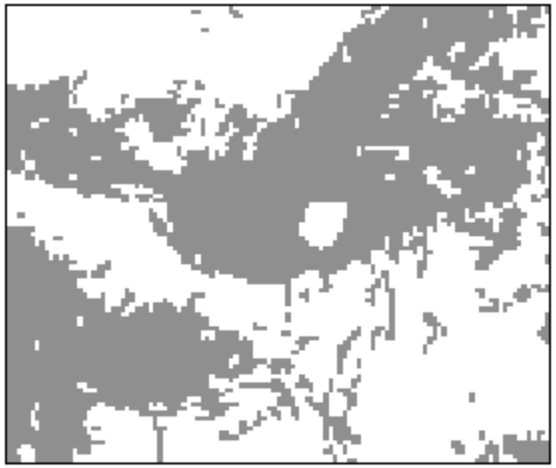
Dominant sets



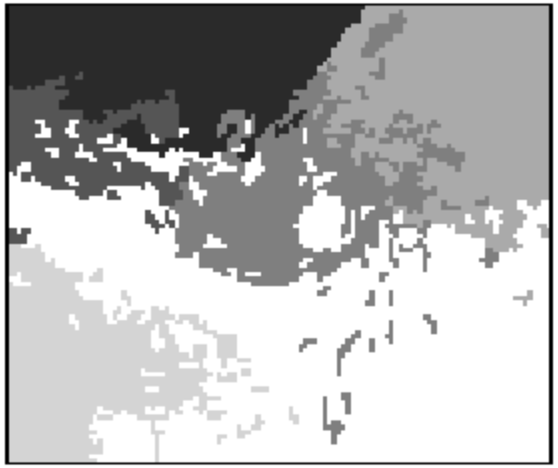
Ncut

Intensity Segmentation Results



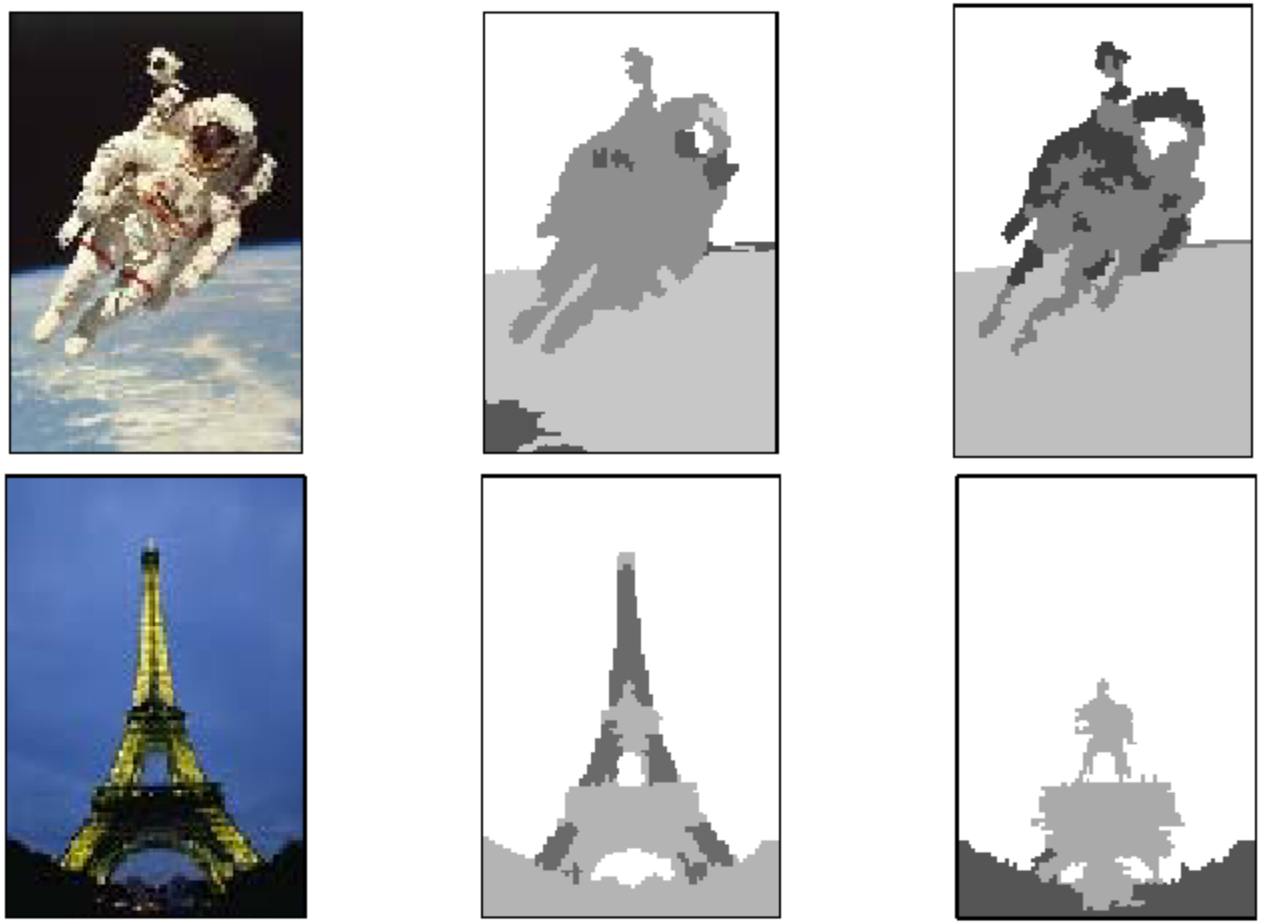


Dominant sets



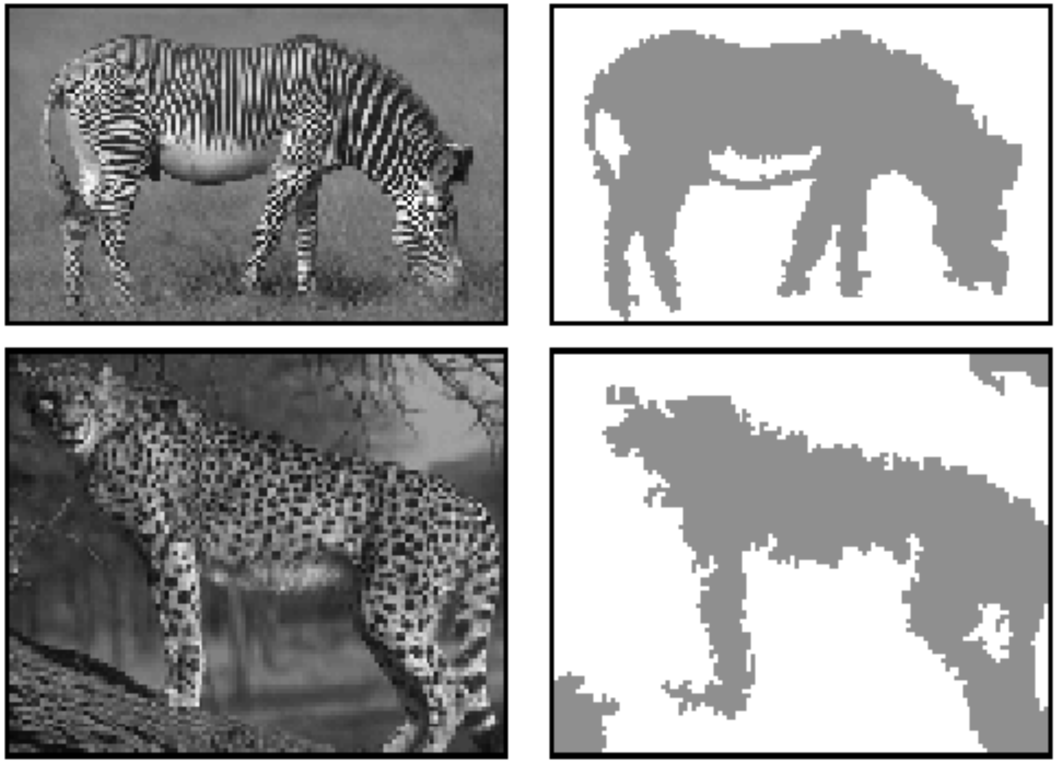
Ncut

Color Segmentation Results



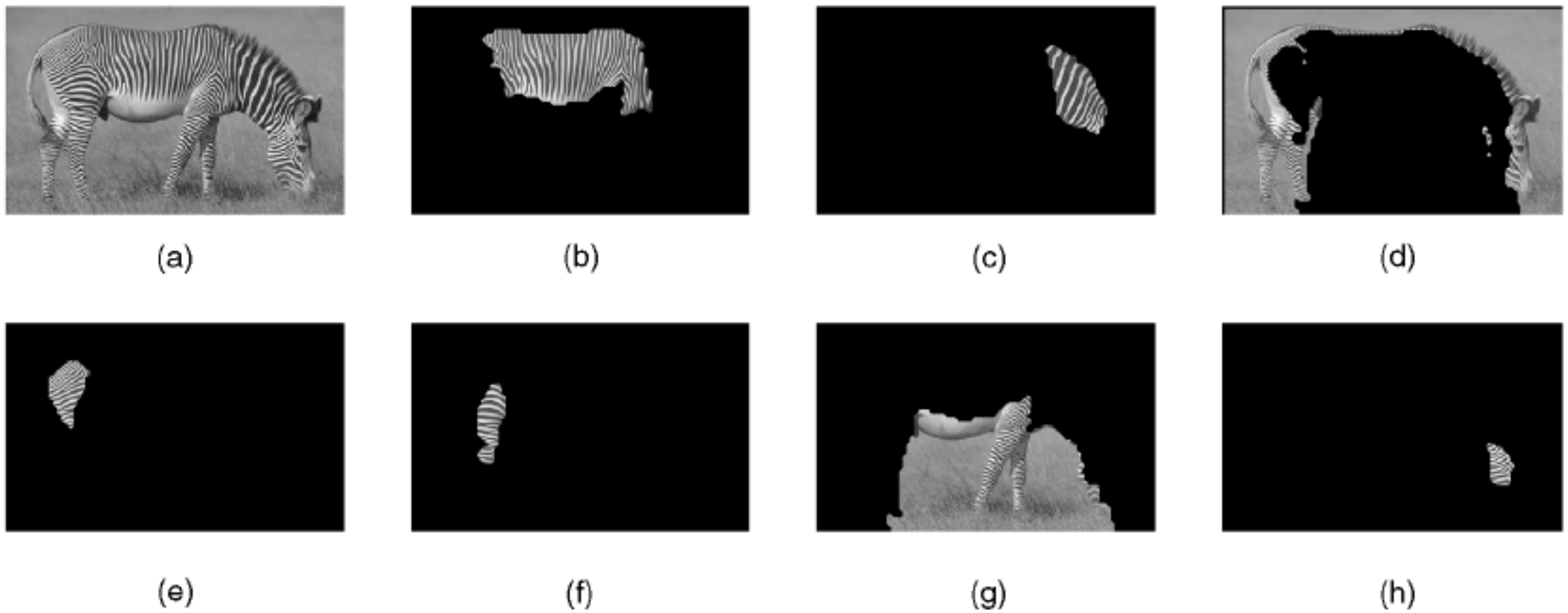
Original image Dominant sets Ncut

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)

Community Detection in Networks: First Results

DataSet	Methods			INFO				
	Algo 1	Algo 2	GTrans	No Nods	Directed	Weighted	No Com	GroundTruth
Karateh /R	0.9794 ±0.0142	0.9441 ±0.0292	0.9794 ±0.0142	34	UD	UW	2	Provided
Karateh /C	0.9706±0.00	0.9118±0.00	0.9706±0.00	34	UD	UW	2	Provided
American Football /R	0.8574±0.0310	0.8626±0.0317	0.8443±0.0421	115	UD	UW	12	Provided
American Football /C	0.8696±0.00	0.8783±0.00	0.8609±0.00	115	UD	UW	12	Provided
Food /R	0.8444±0.0	0.8844±0.0141	0.8444±0.00	45	UD	W	7	Decided
Food /C	0.8444±0.00	0.8889±0.00	0.8444±0.00	45	UD	W	7	Decided
K55 /R	1±0.0	1±0.0	1±0.0	10	UD	UW	2	Decided
K55 /C	1±0.00	1±0.00	1±0.00	10	UD	UW	2	Decided
Journals /R	0.3056±0.0239	0.3137±0.0273	0.3040±0.0235	124	UD	W	14	Provided
Journals /C	0.3347±0.0109	0.3500±0.0078	0.3371±0.0083	124	UD	W	14	Provided
Politic Book /R	0.8105±0.0313	0.8048±0.0321	0.8229±0.0327	105	UD	UW	3	Provided
Politic Book /C	0.7714±0.00	0.7619±0.00	0.7905±0.00	105	UD	UW	3	Provided
Women /R	0.8031±0.1293	0.80±0.1415	0.8094±0.1330	32	UD	UW	2	Decided
Women /C	0.8438 ±0.00	0.8438±0.00	0.8438±0.00	32	UD	UW	2	Decided
Adjnoun /R	0.5384±0.0174	0.5286±0.0162	0.5116±0.0085	112	UD	UW	2	Provided
Adjnoun /C	0.5536±0.00	0.5000±0.00	0.5179±0.00	112	UD	UW	2	Provided
PolBlogs	0.5644	0.5644	0.5087	1490	D	UW	2	Provided
School Grade /R	0.7362±0.0787	0.7580±0.0837	0.8029±0.0993	69	D	W	6	Provided
School Grade /C	0.7391±0.1170	0.8406±0.0	0.8261 ±0.1170	69	D	W	6	Provided
Synthetic Newman /R	0.9133±0.1195	0.9273±0.1009	0.9414±0.1235	128	UD	W	4	Provided
Synthetic Newman /C	0.9609±0.	0.9453±0.0	1.0000±0.	128	UD	W	4	Provided
Synthetic 2 /R	1.0±0.0	1.0±0.0	1.0±0.0	90	UD	W	3	Provided
Synthetic 2 /C	1.0±0.0	1.0±0.0	1.0±0.0	90	UD	W	3	Provided
Synthetic 3 /R	1.0±0.0	1.0±0.0	1.0±0.0	128	UD	W	8	Provided
Synthetic 3 /C	1.0±0.0	1.0±0.0	1.0±0.0	128	UD	W	8	Provided
Synthetic 4 /R	0.7443±0.0596	0.7583±0.0549	0.8389±0.0682	1000	UD	W	13	Provided
Synthetic 4 /C	0.7700 ±0.1170	0.7630±0.0	0.8670±0.0	1000	UD	W	13	Provided

Community Detection in Networks: First Results

DataSet	FCD	BCD	GT
Karateh /R	0.9706±0.0	0.8118 ±0.1049	0.9706±0.0
American Football /R	0.6783±0.0	0.8661 ±0.0491	0.8609±0.0
Food /R	0.6222±0.0	0.3689 ±0.0155	0.8444±0.1170
K55 /R	1.0000±0.0	1.0000 ±0.0	1.0±0.0
Journals /R	0.2016±0.0	—-	0.3371±0.0083
Politic Book /R	0.8286±0.0	0.5248 ±0.0445	0.7905±0.0
Women /R	0.9375±0.0	0.8438 ±0.0833	0.8438±0.0
Adjnoun /R	0.5268±0.0	0.5357 ±0.0339	0.5179±0.00
PolBlogs	——	——	——
School Grade /R	0.2754±0.0	0.7478 ±0.0343	0.8261±0.1170
Synthetic Newman /R	0.6641±0.0	1.000 ±0.	1.0±0.0
Synthetic 2 /R	0.8333±0.0	1.0000 ±0.	1.0±0.0
Synthetic 3 /R	0.6328±0.0	1.0000 ±0.	1.0±0.0
Synthetic 4 /R	——	0.8203 ±0.0790	0.8670±0.0

FCD = Fast algorithm for detecting community structure in networks (M. Newman, Phys. Rev. E, 2004)

BCD = Bayesian community detection (M. Mrup and M. Schmidt, Neural Comp., 2012)

Extensions

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

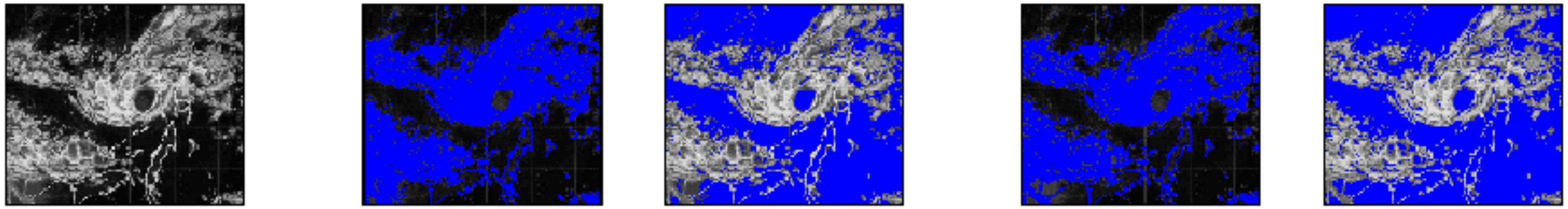
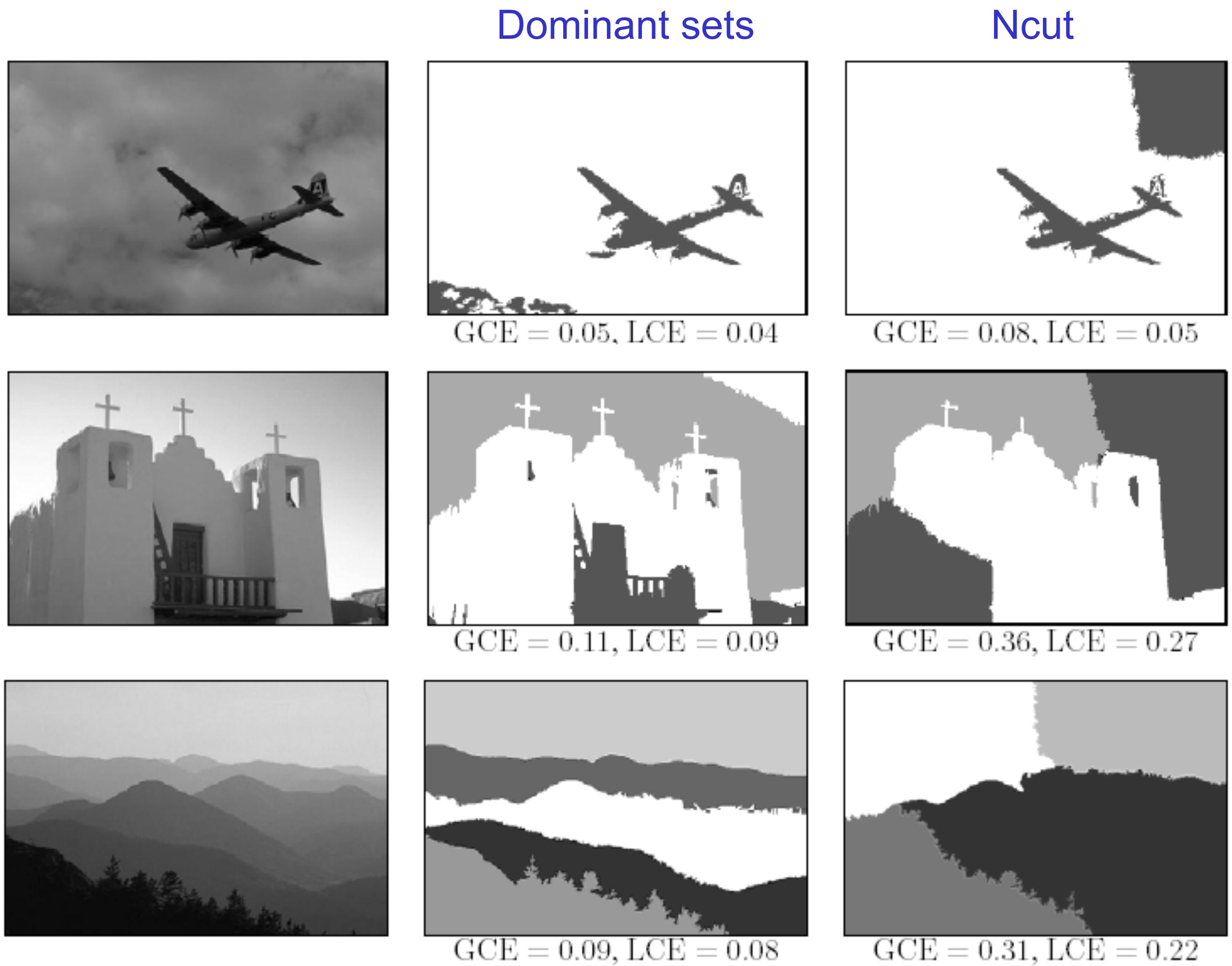
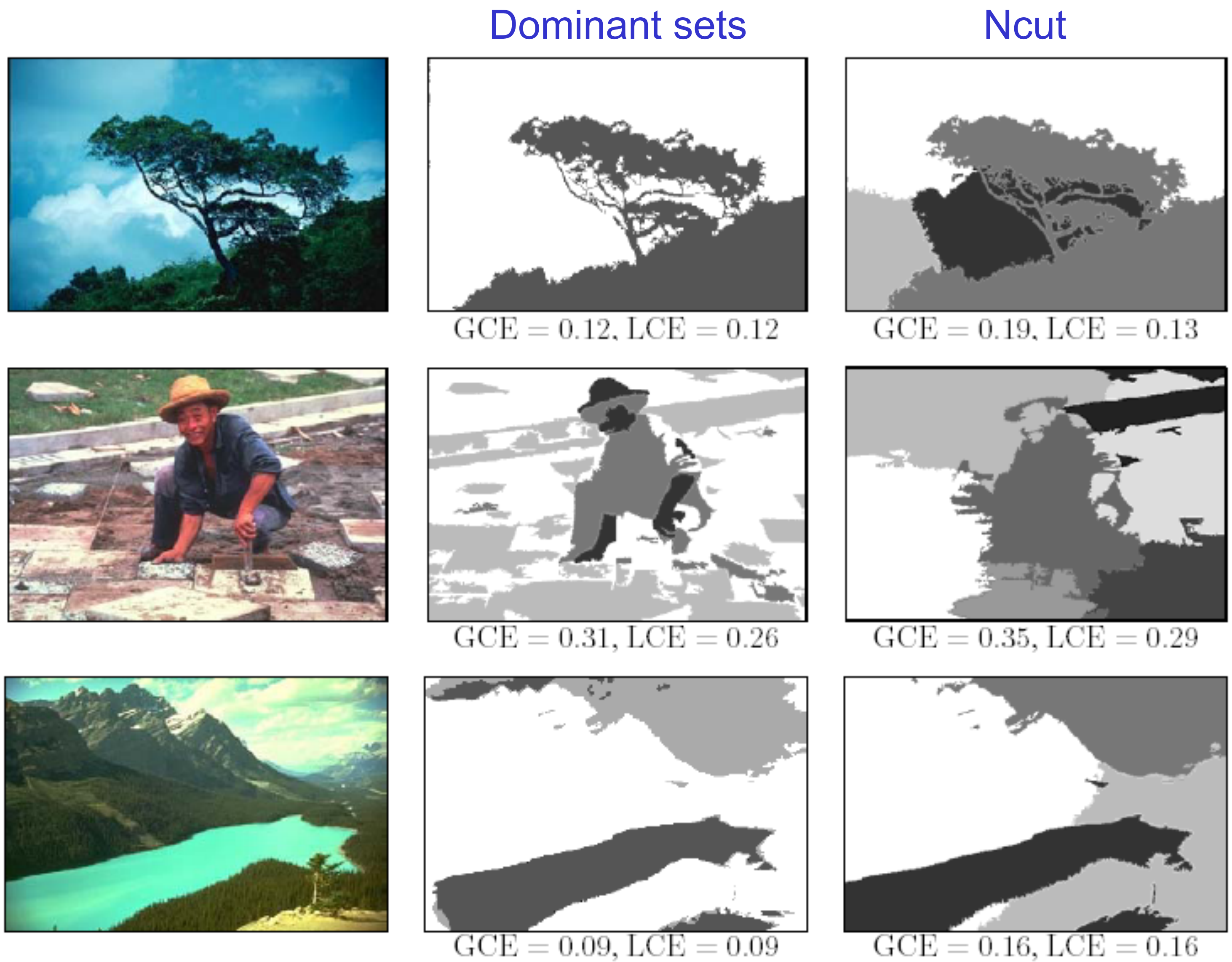


Figure 4: Segmentation results on a 115 $\times$ 97 weather radar image. From left to right: original image, the two regions found on the sampled image (sampling rate = 0.5%), and the two regions obtained on the whole image (sampling rate = 100%).

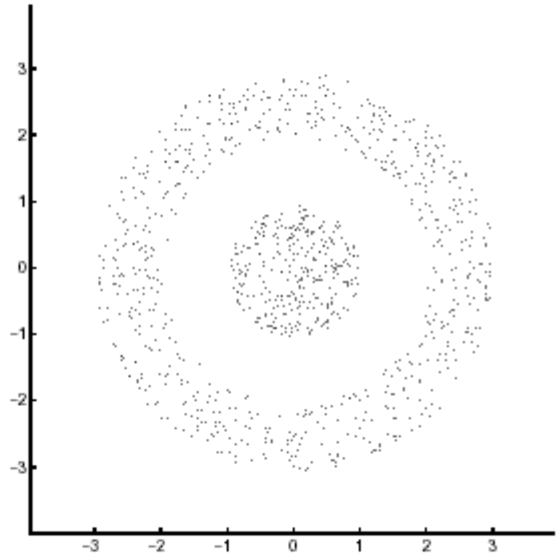
Results on the Berkeley Dataset (321 x 481) — sampling rate 0.5%

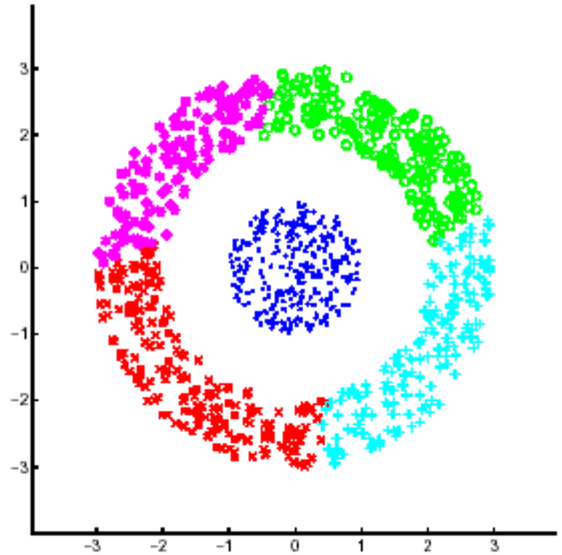


Results on the Berkeley Dataset (321 x 481) — sampling rate 0.5%

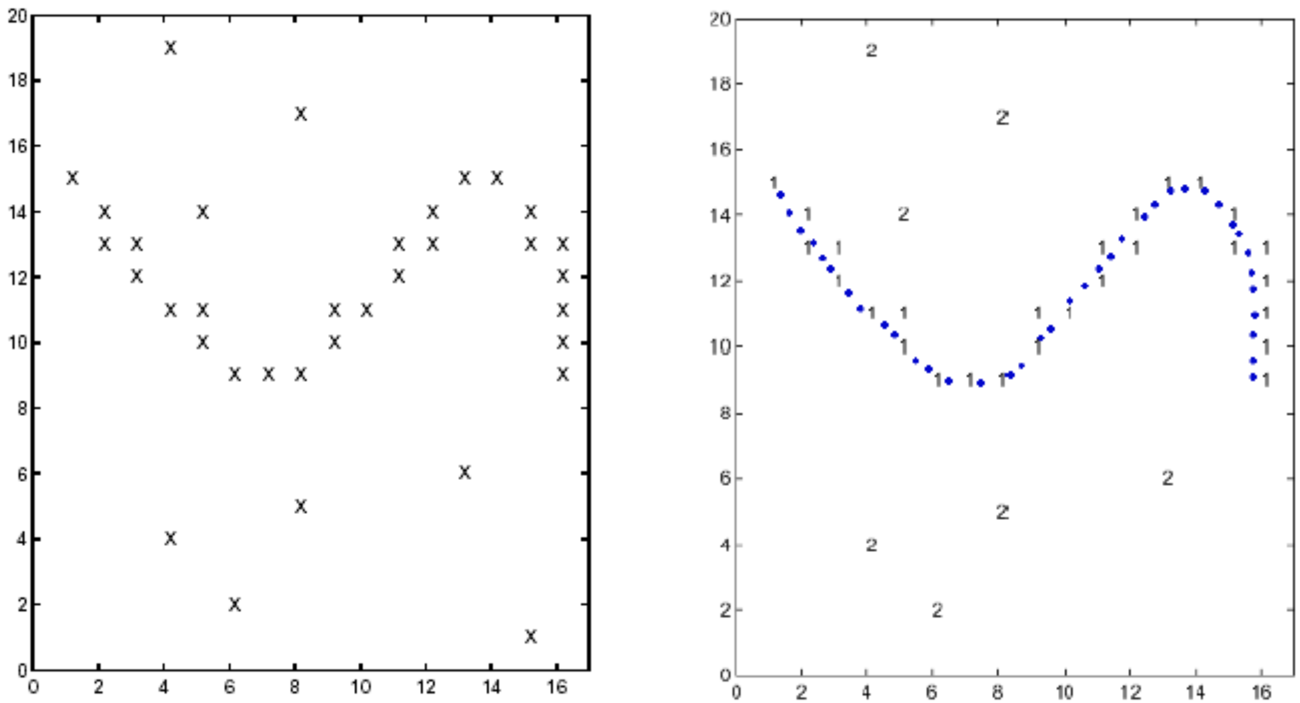


Capturing Elongated Structures / 1





Capturing Elongated Structures / 2



"Closing" the Similarity Graph

Basic idea: Trasform the original similarity graph G into a "closed" version thereof (G_{closed}), whereby edge-weights take into account chained (path-based) structures.

Unweighted (0/1) case:

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

Note: G_{closed} can be obtained from:

$$A + A^2 + \ldots + A^n$$

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, one choice is:

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

Path-Based Distances

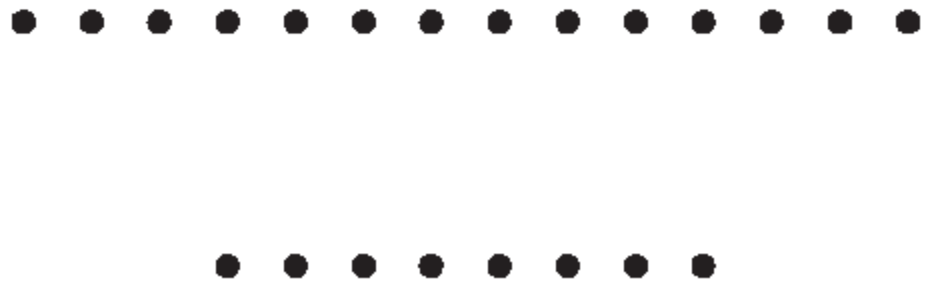
Path-based measure: Given a distance (dissimilarity) matrix $\mathbf{D}$, the path-based distance measure between objects i and j is computed by [FB03]

$$\mathbf{D}_{ij}^{path} = \min_{p \in \mathcal{P}_{ij}(\mathbf{O})} \left\{ \max_{1 \leqslant l \leqslant |p|} \mathbf{D}_{p(l)p(l+1)} \right\}, \tag{19}$$

where $\mathcal{P}_{ij}(\mathbf{O})$ is the set of all paths from i to j. Thereby, the effective distance between i and j is the largest gap of the path $p^\star$, where $p^\star$ is the path with minimum largest gap among all admissible paths between i and j.

B. Fischer and J. M. Buhmann. Path-based clustering for grouping of smooth curves and texture segmentation. IEEE Trans. Pattern Anal. Mach. Intell., 25(4):513–518, 2003.

Example: Without Closure (σ = 2)



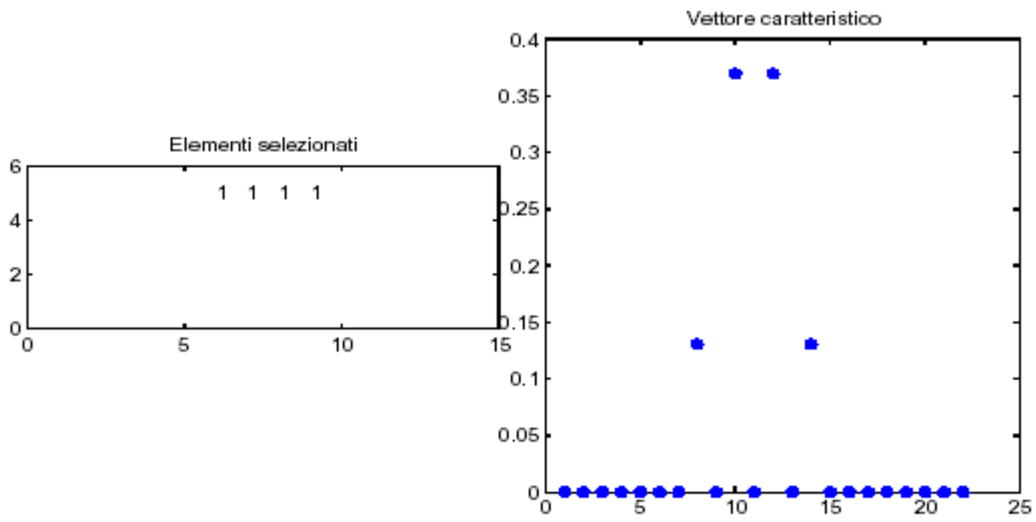
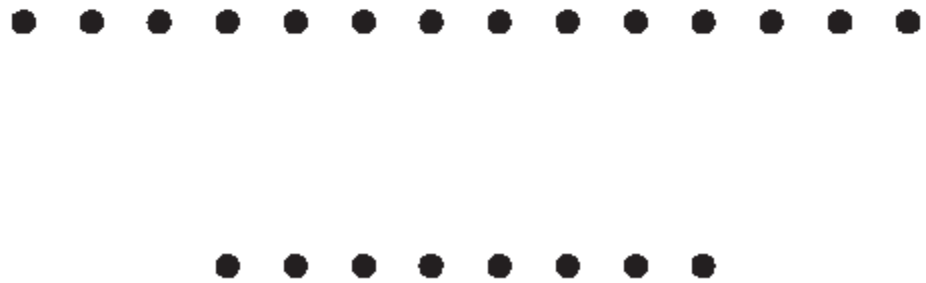


Figura 4.11: Cluster senza chiusura: $\sigma = 2$

Example: Without Closure (σ = 4)



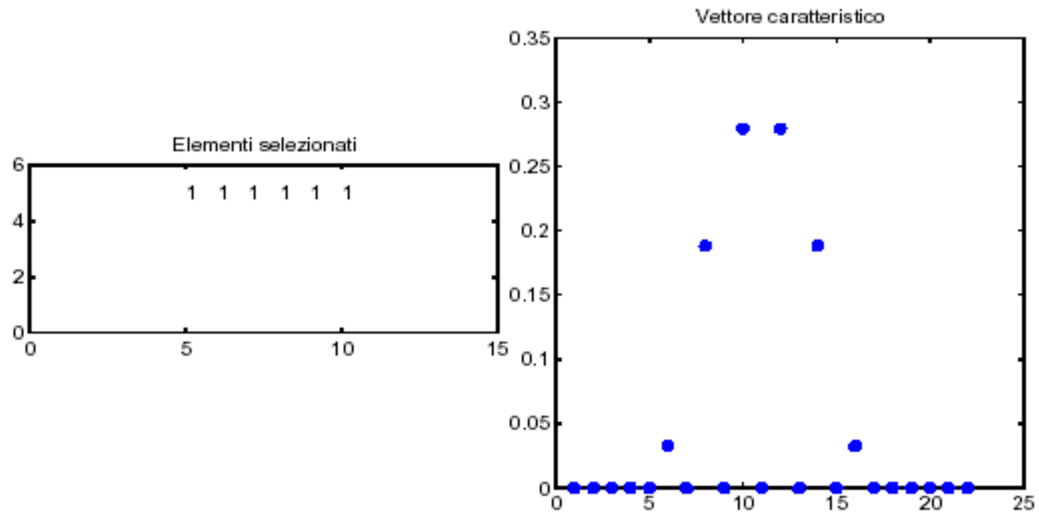


Figura 4.12: Cluster senza chiusura: $\sigma = 4$

Example: Without Closure (σ = 8)

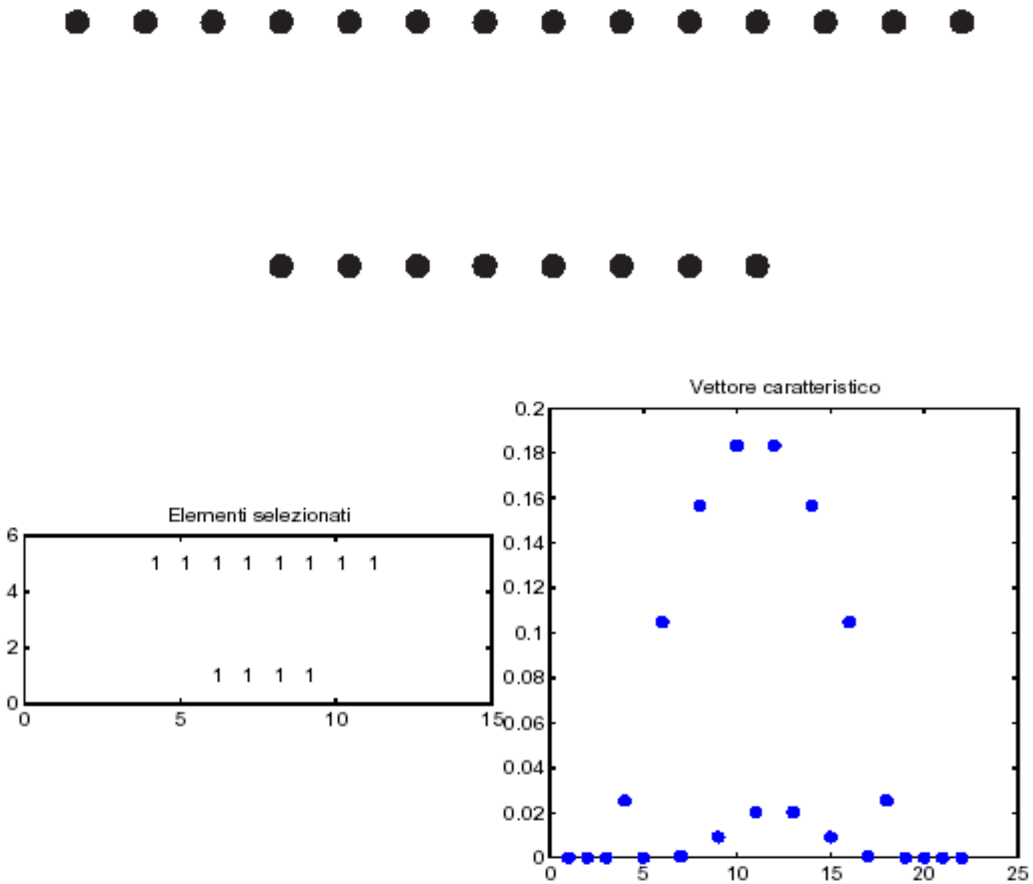


Figura 4.13: Cluster senza chiusura: $\sigma = 8$

Example: With Closure (σ = 0.5)

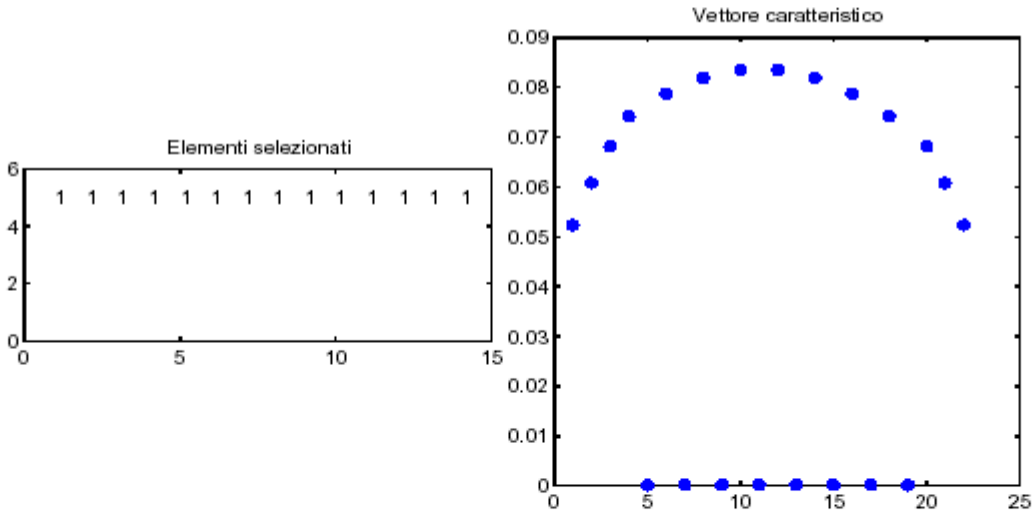
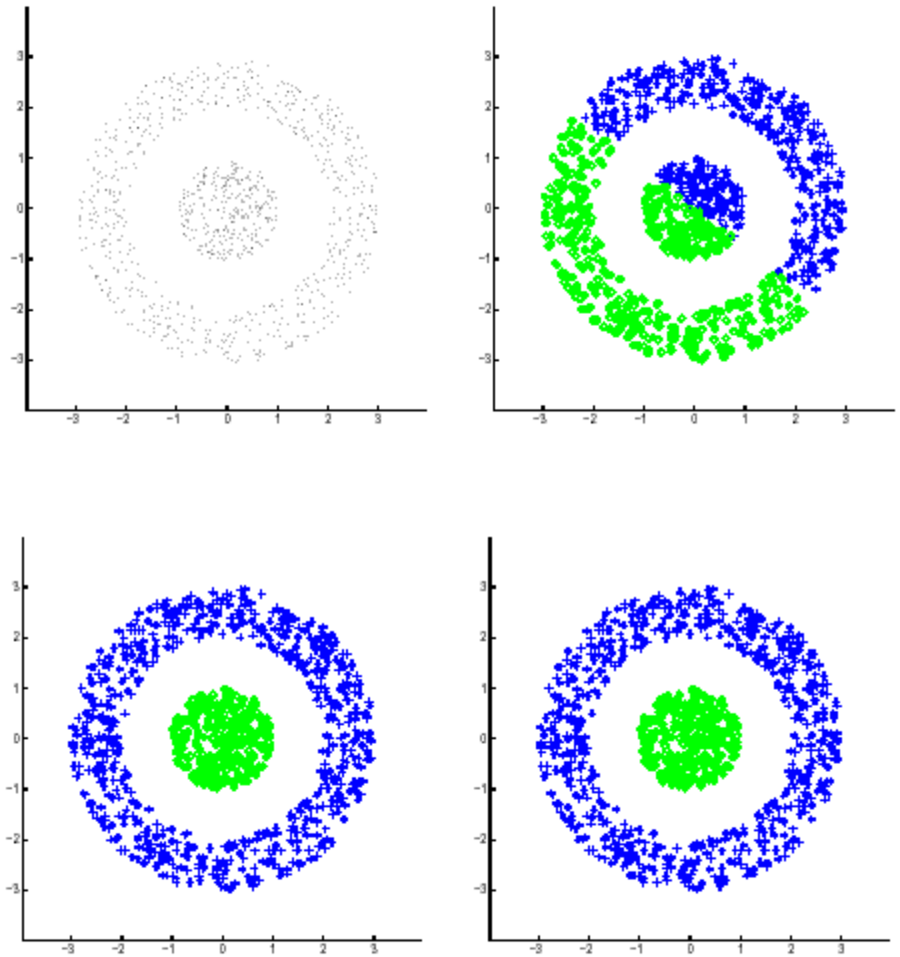


Figura 4.14: Cluster mediante chiusura: $\sigma = 0,5$



Allowing Overlapping Classes: Σ-extensions (the Binary Case)

First idea: run replicator dynamics from different starting points in the simplex.

Problems: computationally expensive and no guarantee to find them all.

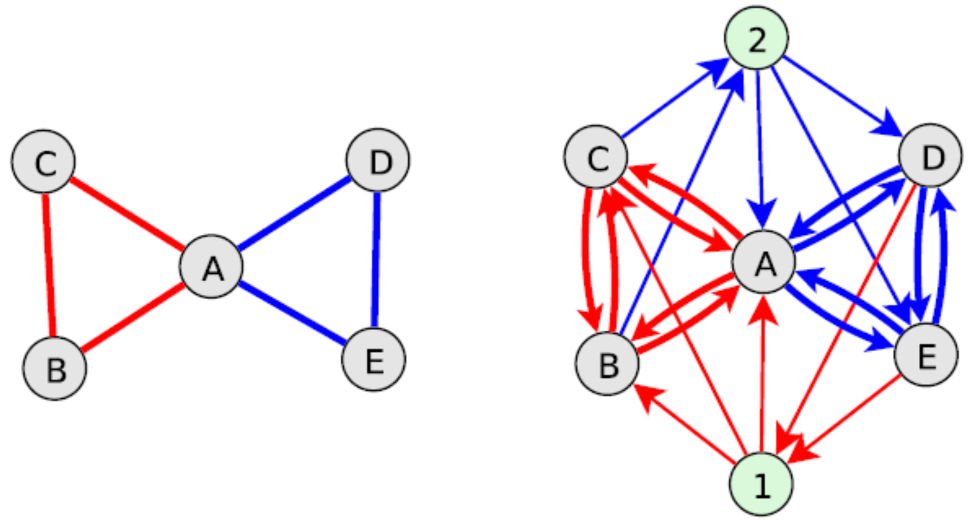


Figure 4.2: On the left we have an undirected graph G. On the right we have the Σ-extension G^{Σ}, where $\Sigma = \{\{A, B, C\}, \{A, D, E\}\}$.

Σ-extensions: The General Case

Let Σ be a tuple of ESSs of a game with payoff matrix A. So for example if $\boldsymbol{x}$ and $\boldsymbol{y}$ are ESSs of a doubly symmetric game then $\Sigma = (\boldsymbol{x}, \boldsymbol{y})$ and with Σ_i we select the i-th ESS.

Σ-extension $A^\Sigma = (a_{ij}^\Sigma)$ of the payoff matrix A is defined as follows.

$$a_{ij}^\Sigma = \begin{cases} a_{ij} & \text{if } i, j \in [1, n] \\ \alpha & \text{if } j > n \text{ and } i \notin \sigma(\Sigma_{j-n}) \\ \beta & \text{if } i, j > n \text{ and } i = j \\ \frac{1}{|\Sigma_{i-n}|} \sum\limits_{k \in \Sigma_{i-n}} a_{kj} & \text{if } i > n \text{ and } j \in \sigma(\Sigma_{i-n}) \\ 0 & \text{otherwise.} \end{cases}$$

where $\alpha > \beta$ and $\beta = \max_{i,j} a_{ij}$.

Main Result

Theorem 1 *Let* Φ *be a two-player doubly symmetric game with payoff matrix* A *and let* Σ *be a tuple of ESSs of* Φ. *Furthermore let* Φ^{Σ} *be a two-player game with payoff matrix* A^{Σ}. *Then* $\boldsymbol{x}$ *is an ESS of* Φ *not in* Σ *if and only if* $\bar{\boldsymbol{x}}$ *is an ESS of* Φ^{Σ}.

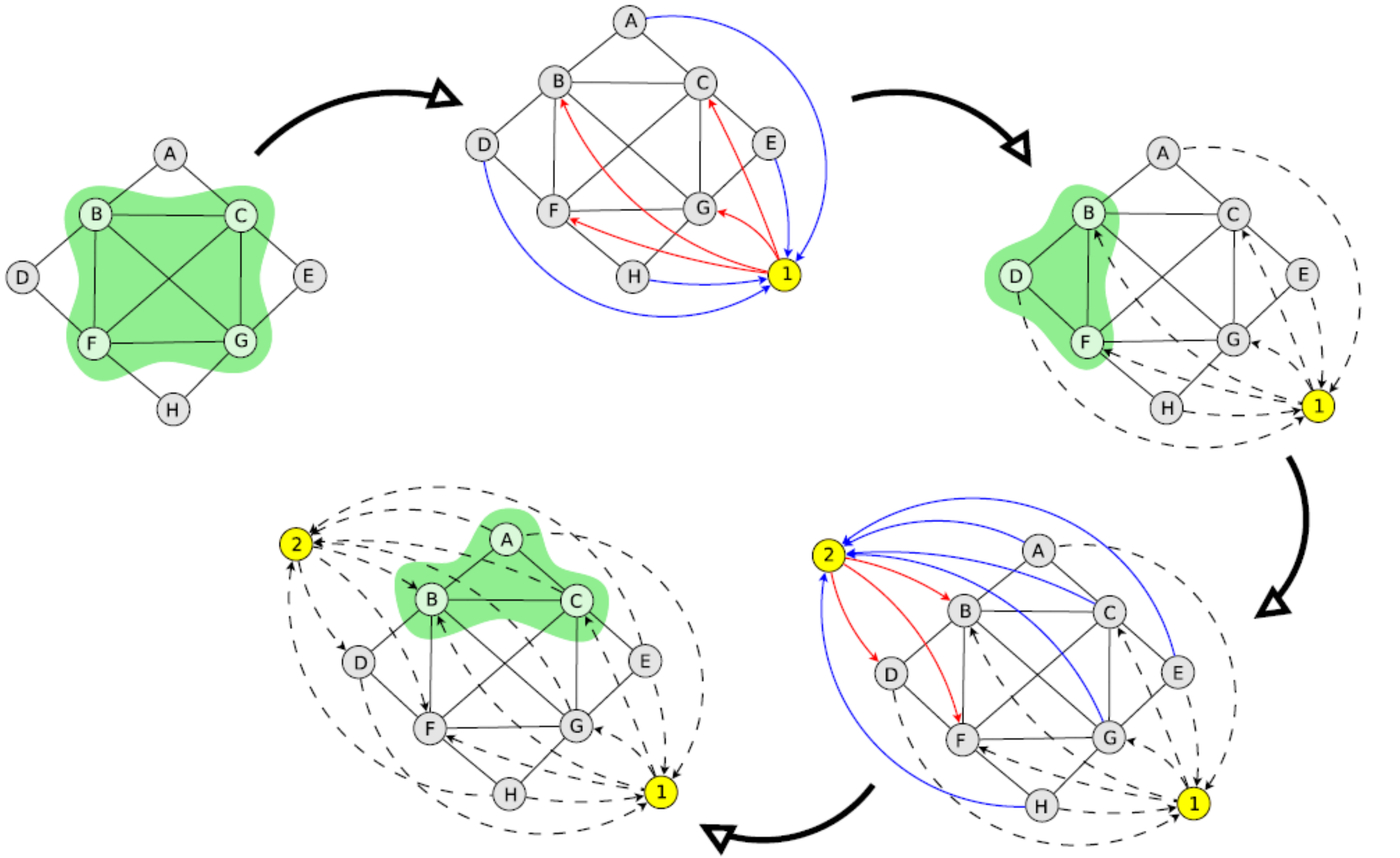
Enumerating Dominant Sets

We use the previous result to enumerate the dominant sets in the following way:

We iteratively find new dominant sets by looking for an asymptotically stable point using the replicator dynamics.

After that, we extend the graph by adding the newly extracted set to Σ, hence rendering its associated strategy unstable, and reiterate the procedure until we have enumerated all the groups and hence are unable to find new dominant sets.

Idea for future work: Dominant-set percolation?


A
B
C
D
E
F
G
H
1
2

Building a Hierarchy: A Family of Quadratic Programs

Consider the following family of StQP's:

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

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

The objective function f_α consists of:

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

An Observation

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

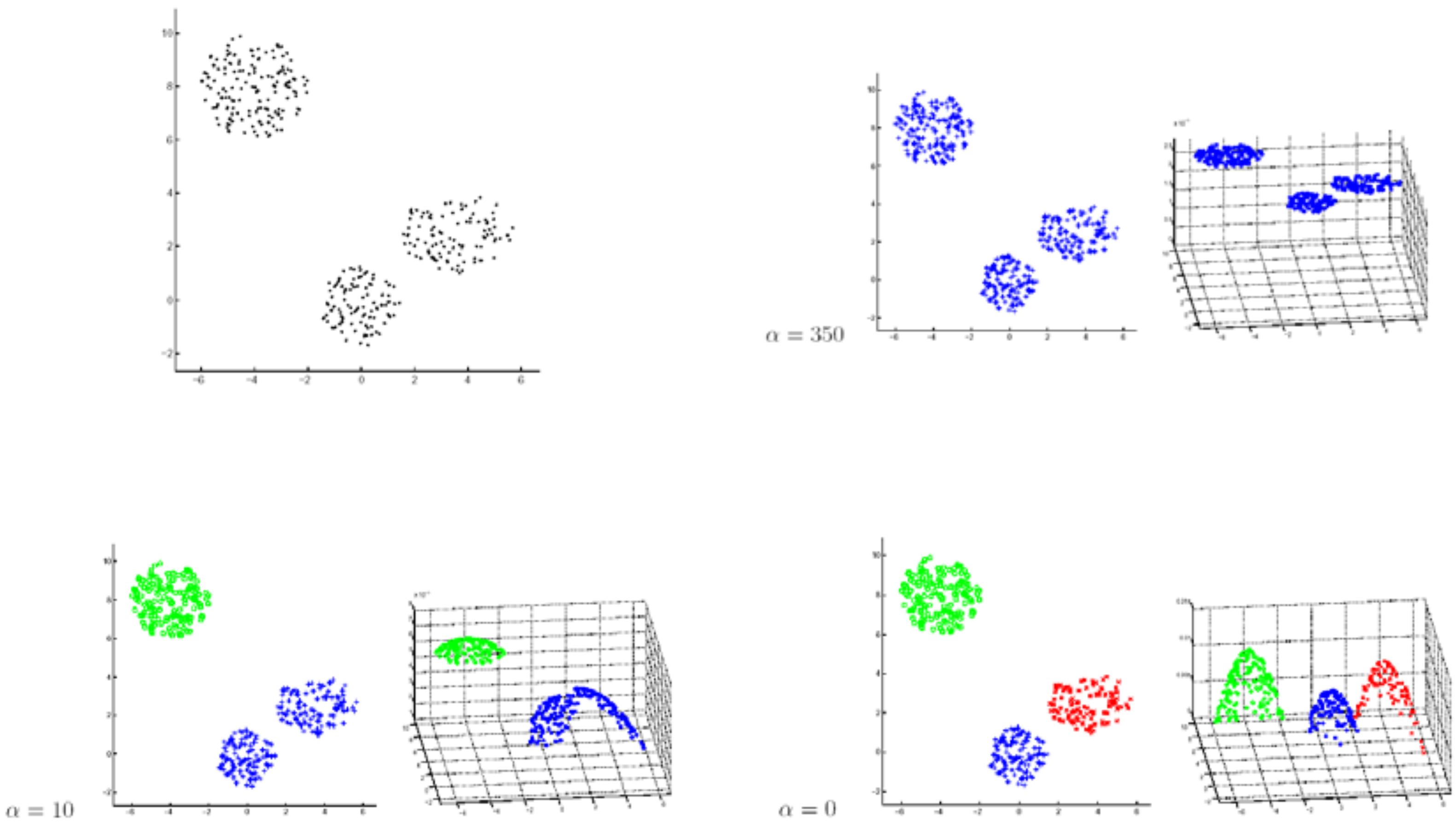
$$\mathbf{x}'(A - \alpha I + \kappa \mathbf{e}\mathbf{e}')\mathbf{x} = \mathbf{x}'(A - \alpha I)\mathbf{x} + \kappa$$

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 $A + \alpha(\mathbf{e}\mathbf{e}' - I)$ having the off-diagonal entries equal to $a_{ij} + \alpha$.

The effects of α



The Landscape of f_α

Key observation: For any fixed α, the energy landscape of f_α 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 $A + \alpha(\mathbf{ee}' - I)$

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 α and adaptively decrease it during the clustering process:

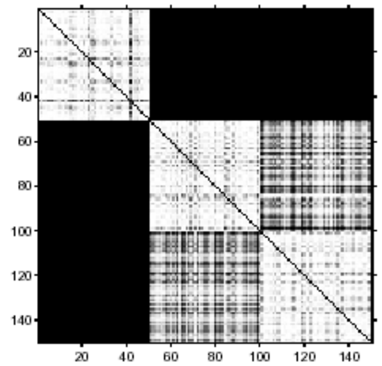
1) let α be a large positive value (e.g., $\alpha > |V| - 1$)

2) find a partition of the data into α-*clusters*

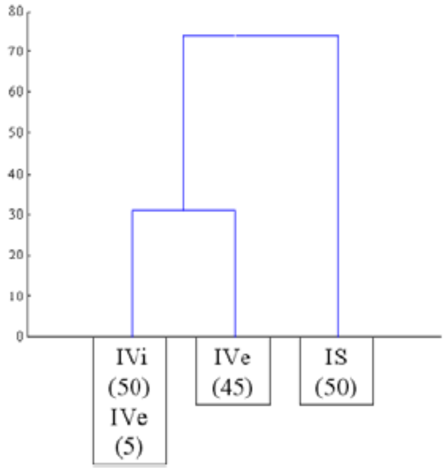
3) for all the α-*clusters* that are not 0-*clusters* recursively repeat step 2) with decreased α

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.

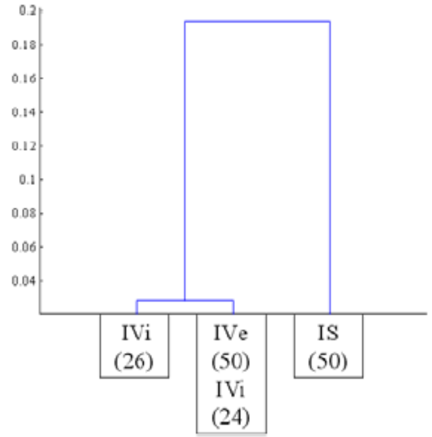


Similarity matrix

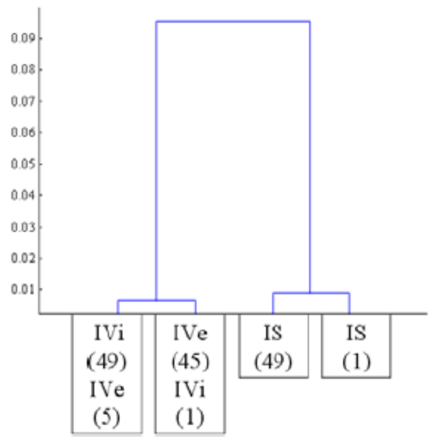


Dominant sets

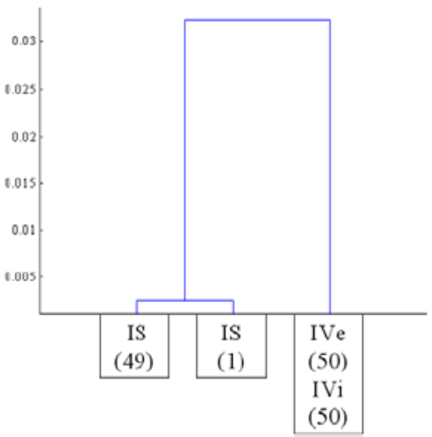
Results on the IRIS dataset / 2



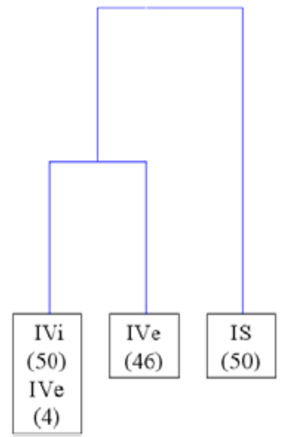
Complete-link



Average-link



Single-link



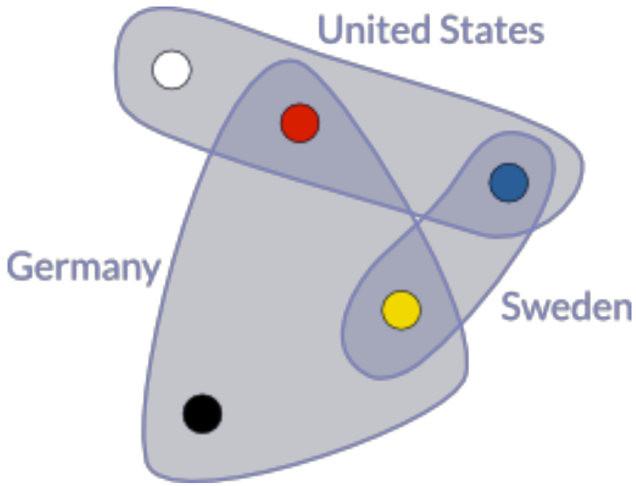
NCut

Dealing with High-Order Similarities

A (weighted) hypergraph is a triplet H = (V, E, w), where

- V is a finite set of vertices
- $E \subseteq 2^V$ is the set of (hyper-)edges (where 2^V is the power set of V)
- $w : E \rightarrow R$ is a real-valued function assigning a weight to each edge

We will focus on a particular class of hypergraphs, called **k-graphs**, whose edges have fixed cardinality $k \geq 2$.



A hypergraph where the vertices are flag colors and the hyperedges are flags.

An Example Application: Folksonomy

"Folksonomy" is the name given to the common on-line process by which a group of individuals collaboratively annotate a data set to create semantic structure. Typically mark-up is performed by labeling pieces of data with tags.

Examples:

- ✓ Flickr
- ✓ CiteUlike

The fundamental building block in a folksonomy is a triple consisting of a resource, such as a photograph, a tag, usually a short text phrase, and a user, who applies the tag to the resource. Any full network representation of folksonomy data needs to capture this three-way relationship between resource, tag, and user, and this leads us to the consideration of hypergraphs.

From: G. Ghoshal and Newman, Random hypergraphs and their applications, 2009.

The Hypergraph Clustering Game

Given a weighted k-graph representing an instance of a hypergraph clustering problem, we cast it into a k-player (hypergraph) clustering game where:

- ✓ There are k players
- ✓ The objects to be clustered are the pure strategies
- ✓ The payoff function is proportional to the similarity of the objects/strategies selected by the players

Definition (ESS-cluster). Given an instance of a hypergraph clustering problem H = (V,E,w), an ESS-cluster of H is an ESS of the corresponding hypergraph clustering game.

Like the k=2 case, ESS-clusters do incorporate both internal and external cluster criteria (see PAMI 2013)

ESS's and Polynomial Optimization

Theorem 3. *Let $H = (V, E, \omega)$ be a hypergraph clustering problem, $\Gamma = (P, V, \pi)$ the corresponding clustering game, and $f(\mathbf{x})$ a function defined as*

$$f(\mathbf{x}) = u\left(\mathbf{x}^{[k]}\right) = \sum_{e \in E} \omega(e) \prod_{j \in e} x_j. \quad (11)$$

Nash equilibria of Γ are in one-to-one correspondence with the critical points[2] of $f(\mathbf{x})$ over Δ, while ESSs of Γ are in one-to-one correspondence with strict local maximizers of $f(\mathbf{x})$ over Δ.

Baum-Eagon Inequality

Theorem 4 (Baum-Eagon). *Let $Q(\mathbf{x})$ be a homogeneous polynomial in the variables x_j with nonnegative coefficients, and let $\mathbf{x} \in \Delta$. Define the mapping $\mathbf{z} = \mathcal{M}(\mathbf{x})$ from Δ to itself as follows:*

$$z_j = x_j \frac{\partial Q(\mathbf{x})}{\partial x_j} \Big/ \sum_{\ell=1}^{n} x_\ell \frac{\partial Q(\mathbf{x})}{\partial x_\ell}, \qquad j = 1, \ldots, n. \tag{12}$$

Then, $Q(\mathcal{M}(\mathbf{x})) > Q(\mathbf{x})$, unless $\mathcal{M}(\mathbf{x}) = \mathbf{x}$.

Line Clustering

Problem: to clustering lines in spaces of dimension greater than two, i.e., given a set of points in $\mathbb{R}^d$, to extract subsets of collinear points.

An obvious ternary similarity measure for this problem can be defined as follows: Given a triplet of points {i,j,k} and its best fitting line l, calculate the mean distance d(i, j, k) between each point and and l.

Then we obtain a similarity function using a standard Gaussian kernel:

$$\omega(\{i,j,k\}) = \exp(-\beta d(i,j,k)^2)$$

where β is a properly tuned precision parameter.

Line Clustering

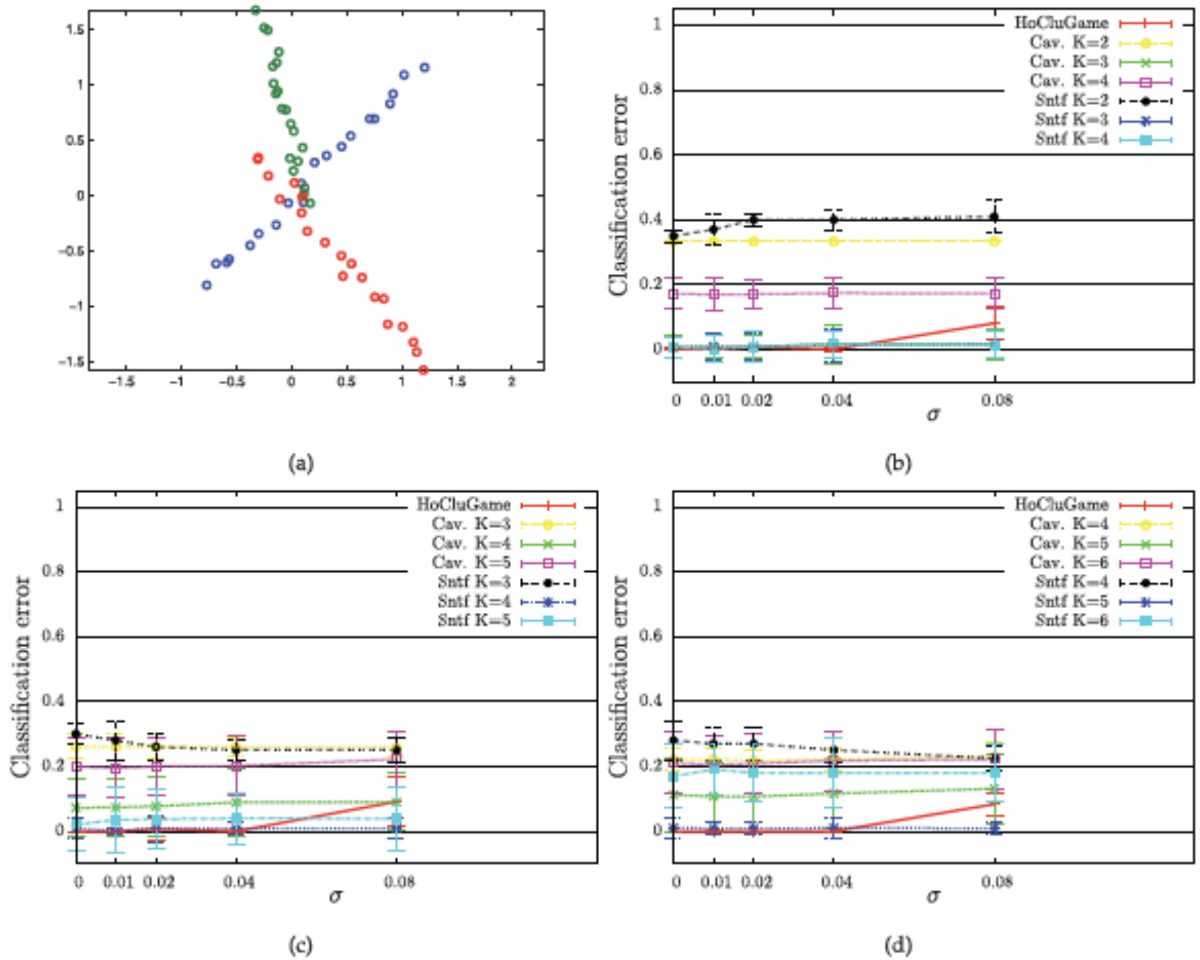


Fig. 2. Results of clustering three, four, and five lines perturbed locally with increasing levels of Gaussian local noise ($\sigma = 0, 0.01, 0.02, 0.04, 0.08$). (a) Example of three 5D lines (projected in 2D), perturbed with $\sigma = 0.04$. (b) Three lines. (c) Four lines. (d) Five lines.

Line Clustering

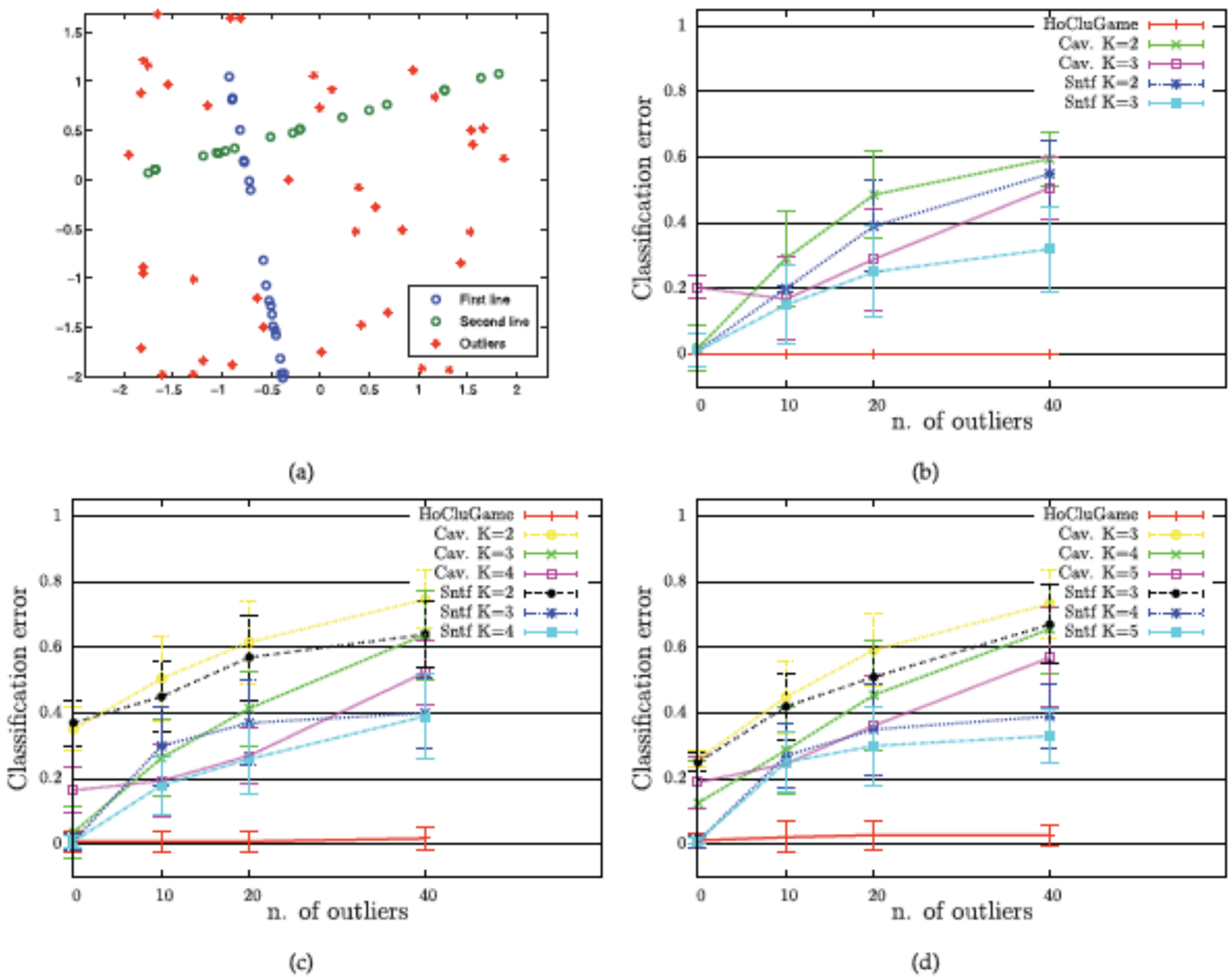


Fig. 3. Results of clustering two, three, and four lines with an increasing number of clutter points $(0, 10, 20, 40)$. (a) Example of two 5D lines (projected in 2D) with 40 clutter points. (b) Two lines. (c) Three lines. (d) Four lines.

Illumination-Invariant Face Clustering

Fig. 8. Example of dataset for illuminant-invariante face clustering with four individuals (first four rows) and 10 outlier faces (last row).

Illumination-Invariant Face Clustering

n. of classes:	4		5	
n. of outliers:	0	10	0	10
CAVERAGE K=3	0.26±0.09	0.40±0.10	-	-
CAVERAGE K=4	**0.03±0.04**	0.24±0.07	0.21±0.11	0.65±0.12
CAVERAGE K=5	0.13±0.05	0.12±0.05	0.07±0.07	0.41±0.09
CAVERAGE K=6	-	-	0.13±0.08	0.37±0.11
SNTF K=3	0.29±0.10	0.39±0.09	-	-
SNTF K=4	0.14±0.06	0.26±0.09	0.28±0.11	0.51±0.12
SNTF K=5	0.19±0.09	0.25±0.13	0.11±0.09	0.43±0.11
SNTF K=6	-	-	0.14±0.09	0.39±0.13
HoCluGame	0.06±0.03	**0.07±0.02**	**0.06±0.02**	**0.07±0.03**

Average classification error and corresponding standard deviation.

In a nutshell…

The game-theoretic/dominant-set approach:

- ✓ 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'13)*
- ✓ 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'13.*

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.*