What is a Cluster?

Perspectives from Game Theory

Marcello Pelillo

Department of Computer Science

Ca’ Foscari University, Venice

“We pay too much attention to the details of algorithms. [...] We must begin to subordinate engineering to philosophy.” John Hartigan (1996)
The Partitional “Paradigm”

**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).
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. (adapted 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
“[...] 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.”

Epoché and the “Skeptical” Classifier

Are categories discovered or invented?
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:
The Simmetry 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

“Since no paradigm ever solves all the problems it defines and since no two paradigms leave all the same problems unsolved, paradigm debates always involve the question: Which problems is it more significant to have solved?”

Thomas S. Kuhn

*The Structure of Scientific Revolutions* (1962)
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.
Assume:

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

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

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

\[
\Delta = \left\{ x \in \mathbb{R}^n : c'x = 1 \text{ and } x_i \geq 0 \ \forall i \in V \right\}
\]
Nash Equilibria and Extensions

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

- $y'Ax$ is the average payoff obtained by playing mixed strategy $y$ while the opponent plays $x$.

- A mixed strategy $x$ is a Nash equilibrium if $x'Ax \geq y'Ax$ for all strategies $y$. (Best reply to itself.)

- A Nash equilibrium is an Evolutionary Stable Strategy (ESS) if, for all strategies $y$

\[
y'Ax = x'Ax \Rightarrow x'Ay > y'Ay
\]
We claim that ESS’s 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.
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, the notion of a cluster coincide with that of a *maximal clique*.

ESS’s are in one-to-one correspondence to maximal cliques
Special Case II: Arbitrary Symmetric Affinities

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

$$\begin{align*}
\text{maximize} & \quad f(x) = x'Ax \\
\text{subject to} & \quad x \in \Delta
\end{align*}$$

where

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

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

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

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

\[
\omega_S(i) = \begin{cases} 
1, & \text{if } |S| = 1 \\
\sum_{j \in S \setminus \{i\}} \phi_{S \setminus \{i\}}(j, i) \omega_{S \setminus \{i\}}(j), & \text{otherwise.}
\end{cases}
\]

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

\[
\mathcal{W}(S) = \sum_{i \in S} \omega_S(i).
\]
Interpretation

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.
A Combinatorial Characterization of ESS’s

**Theorem (CVPR’06)** Evolutionary stable strategies of the clustering game with affinity matrix $A$ are in a one-to-one correspondence with dominant sets.

**Note:** Generalization of Motzkin-Straus theorem from graph theory, and of CVPR’03/PAMI’07 Theorem which states that, in the symmetric case, dominant sets are in one-to-one correspondence with strict local maximizers of $x^TAx$ on the standard simplex.
Replicator Dynamics

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:**

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

**Discrete-time version:**

\[
x_i(t + 1) = x_i(t) \frac{(Wx(t))_i}{x(t)'Wx(t)}
\]

\( \Delta \) is invariant under both dynamics, and they have the same stationary points.
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
Replicator Dynamics and ESS’s

Theorem 3 A point $x \in \Delta$ is the limit of a trajectory of (6) starting from the interior of $\Delta$ if and only if $x$ is a Nash equilibrium. Further, if point $x \in \Delta$ is ESS then it is asymptotically stable.$^2$
The Symmetric Case:  
The Fundamental Theorem of Natural Selection

If $W = W'$, then the function

$$F(x) = x'Wx$$

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

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

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

for the continuous-time dynamics, and

$$F(x(t + 1)) > F(x(t))$$

for the discrete-time dynamics, unless $x(t)$ is a stationary point.
Measuring the Degree of Cluster Membership

**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).
Image Segmentation

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

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

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

Dominant sets

Ncut
Color Segmentation Results (125 x 83)

Original image
Dominant sets
Ncut
Texture Segmentation Results
(approx. 90 x 120)

Original image

Dominant sets
Ncut Results

(a)  (b)  (c)  (d)

(e)  (f)  (g)  (h)
Results on Berkeley Database Images
(321 x 481)
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 = 0.00, LCE = 0.09

GCE = 0.16, LCE = 0.16
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)

**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)

**Video analysis and object tracking:**
- *Torsello et al.* (EMMCVPR’05) *Gualdi et al.* (IWVS’08)
In a nutshell...

Our 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 \((NIPS’09)\)
1. Questioned the assumptions underlying the dominant paradigm, and argued that we should start asking the question “what is a cluster?”

2. Surveyed the dominant-set framework for pairwise data clustering.

- **Binary affinities**: maximal cliques (*graph theory*)
- **Symmetric affinities**: maxima of quadratic function over standard simplex (*optimization theory*)
- **Arbitrary affinities**: Nash equilibria of non-cooperative games (*game theory*)
On-going and Future Work

- Finding hierarchical partitions (*ICCV’03*)
- Allowing overlapping clusters for “soft” partitioning (*ICPR’08*)
- Using high-order affinities (>2 players) for hypergraph clustering (*NIPS’09*)
- Using faster game dynamics (*GEB*, submitted)
- Using non-linear payoff functions
- Graph and (hypergraph) matching (*NIPS’09 “Learning with Orderings” workshop*)
- Learning the payoffs
- Semi-supervised learning
- Relations with spectral clustering 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 EU FP7 *SIMBAD* project: http://simbad-fp7.eu).
M. Pavan, M. Pelillo.

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

M. Pavan, M. Pelillo.

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

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.*

A. Albarelli, S. Rota Bulò, A. Torsello, M. Pelillo.
Matching as a non-cooperative game. *ICCV 2009.*

S. Rota Bulò, M. Pelillo.