Invited Review

Dominant-set clustering: A review

Samuel Rota Bulò [a,*], Marcello Pelillo [b]

[a] FBK-irst, Trento, Italy
[b] DAIS, Università Ca' Foscari Venezia, Venice, Italy

ABSTRACT

Clustering refers to the process of extracting maximally coherent groups from a set of objects using pairwise, or high-order, similarities. Traditional approaches to this problem are based on the idea of partitioning the input data into a predetermined number of classes, thereby obtaining the clusters as a by-product of the partitioning process. A radically different perspective of the problem consists in providing a formalization of the very notion of a cluster and considering the clustering process as a sequential search of structures in the data adhering to this cluster notion. In this manuscript we review one of the pioneering approaches falling in the latter class of algorithms, which has been proposed in the early 2000s and has been found since then a number of applications in different domains. It is known as dominant set clustering and provides a notion of a cluster (*a.k.a.* dominant set) that has intriguing links to game-theory, graph-theory and optimization theory. From the game-theoretic perspective, clusters are regarded as equilibria of non-cooperative "clustering" games; in the graph-theoretic context, it can be shown that they generalize the notion of maximal clique to edge-weighted graphs; finally, from an optimization point of view, they can be characterized in terms of solutions to a simplex-constrained, quadratic optimization problem, as well as in terms of an exquisitely combinatorial entity. Besides introducing dominant sets from a theoretical perspective, we will also focus on the related algorithmic issues by reviewing two state-of-the-art methods that are used in the literature to find dominant sets clusters, namely the Replicator Dynamics and the Infection and Immunization Dynamics. Finally, we conclude with an overview of different extensions of the dominant set framework and of applications where dominant sets have been successfully employed.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Clustering is the problem of organizing a set of data elements into groups in a way that each group satisfies an internal homogeneity and external inhomogeneity property. Researchers have focused their attention on this problem for many decades due to its broad applicability, but it remains nonetheless an active area of research in several fields, among which is operation research (Benati, Puerto, & Rodríguez-Chía, 2017; Carrizosa, Mladenović, & Todosijević, 2013; De Angelis & Dias, 2014; Hassin & Or, 2010; Meyer & Oltenau, 2013; Nascimento & de Carvalho, 2011; Santi, Aloise, & Blanchard, 2016) At the same time, there is also a growing interest around fundamental questions pertaining to the very nature of the clustering problem (see, *e.g.*, Ackerman & Ben-David, 2008; Kleinberg, 2002; Zadeh & Ben-David, 2009). Yet, despite the tremendous progress in the field, the clustering

problem remains elusive and a satisfactory answer even to the most basic questions is still to come.

The vast majority of the existing approaches deal with a very specific version of the problem, which asks for *partitioning* the input data into coherent classes. Even the classical distinction between hierarchical and partitional algorithms (Jain & Dubes, 1988) seems to suggest the idea that partitioning data is, in essence, what clustering is all about (as hierarchies are but nested partitions). The partitional paradigm is attractive as it leads to elegant mathematical and algorithmic treatments and allows us to employ powerful ideas from such sophisticated fields as linear algebra, graph theory, optimization, statistics, information theory, etc. However, there are several (far too often neglected) drawbacks with this oversimplified formulation that we review below.

To begin, the very idea of a partition implies that *all* the input data will have to get assigned to some class. There are various applications for which it makes little sense to force all data items to belong to some group, a process, which might result either in poorly-coherent clusters or in the creation of extra spurious classes. As an extreme example, consider the classical figure/ground separation problem in computer vision, which asks

for extracting a coherent region (the figure) from a noisy background (Herault & Horaud, 1993; Shashua & Ullman, 1988). More recently, motivated by practical applications arising in document retrieval and bioinformatics, a conceptually identical problem has attracted some attention within the machine learning community and is generally known under the name of one-class clustering (Crammer, Talukdar, & Pereira, 2008; Gupta & Ghosh, 2005).

The second intrinsic limitation of the partitional paradigm is even more severe as it imposes that elements cannot belong to more than one cluster. There are a variety of important applications, however, where this requirement is too restrictive. Examples abound and include, *e.g.*, clustering micro-array gene expression data (wherein a gene often participate in more than one process), clustering documents into topic categories, perceptual grouping, and segmentation of images with transparent surfaces. Typically, this is solved by relaxing the constraints imposed by crisp partitions to allow for "soft" boundaries between clusters.

Finally, stemming from a natural assumption for central clustering frameworks, clustering approaches have traditionally worked under the assumption that the similarities satisfy metric properties, *i.e.*, they are symmetric, non-negative, and satisfy the triangle inequality. However, recently there has been a strong interest in relaxing these requirements (Jacobs, Weinshall, & Gdalyahu, 2000; Roth, Laub, Kawanabe, & Buhmann, 2003; Yu & Shi, 2001). This is due to the fact that in many applications non-metric similarities arise naturally (Ho, Ming-Hsuan, Jongwoo, Kuang-Chih, & Kriegman, 2003; Williams & Thornber, 1999). More fundamentally, some researches argue that human perception does not satisfy metric properties (Jacobs et al., 2000). While the literature presents many approaches that lift the assumption of non-negativity and triangle inequality (Jacobs et al., 2000; Roth et al., 2003), little progress has been made in relaxing the symmetry constraint. Note, however, that the limited progress in grouping with asymmetric affinities is not due to the lack of interest. In fact, there are many practical applications where asymmetric (or, more generally, non-metric) similarities do arise quite naturally. For example, such (dis)similarity measures are typically derived when images, shapes or sequences are aligned in a template matching process. In image and video processing, these measures are preferred in the presence of partially occluded objects (Jacobs et al., 2000). Other examples include pairwise structural alignments of proteins that focus on local similarity (Altschul, Gish, Miller, Myers, & Lipman, 1990), variants of the Hausdorff distance (Dubuisson & Jain, 1994), normalized edit-distances, and probabilistic measures such as the Kullback–Leibler divergence. A common method to deal with asymmetric affinities is simply to symmetrize them, but in so doing we might lose important information that reside in the asymmetry. As argued in Jacobs et al. (2000), the violation of metricity is often not an artifact of poor choice of features or algorithms, but it is inherent in the problem of robust matching when different parts of objects (shapes) are matched to different images. The same argument may hold for any type of local alignments. Corrections or simplifications of the original affinity matrix may therefore destroy essential information.

Although probabilistic model-based approaches do not suffer from several of the limitations mentioned above, here we review an alternative strategy introduced by Pavan and Pelillo (2003b, 2007). Instead of insisting on the idea of determining a partition of the input data, and hence obtaining the clusters as a by-product of the partitioning process, the idea is to reverse the terms of the problem and attempt instead to derive a rigorous formulation of the very notion of a cluster. They found that game theory offers a very elegant and general perspective that serves well the purpose, and came up with a game-theoretic framework for clustering (Pavan & Pelillo, 2007; Rota Bulò & Pelillo, 2009; Torsello, Rota Bulò, & Pelillo, 2006), which has found applications in fields as diverse as computer vision and bioinformatics and will be the topic of this review. The starting point is the elementary observation that a "cluster" may be informally defined as a maximally coherent set of data items, *i.e.*, as a subset of the input data C, which satisfies both an *internal* criterion (all elements belonging to C should be highly similar to each other) and an *external* one (maximality, *i.e.*, no larger cluster should contain C as a proper subset). The clustering problem can then be formulated as a non-cooperative *clustering game*, where the notion of a cluster turns out to be equivalent to a classical equilibrium concept from (evolutionary) game theory, as the latter reflects both the internal and external cluster conditions mentioned above. The clustering game is defined as follows: Assume a pre-existing set of objects O and a (possibly asymmetric and even negative) 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. As an example, let us assume that our clustering problem is one of figure/ground discrimination, that is, the objects in O consist of a cohesive group with high mutual affinity (figure) and of non-structured noise (ground). Being non-structured, the noise gives equal average affinity to elements of the figures as to elements of the ground. Informally, assuming no prior knowledge of the inclination of the adversary, a player will be better-off selecting elements of the figure rather than of the ground.

Within this framework, clusters correspond to the Evolutionary Stable Strategies (ESS) of the clustering game. The hypotheses that each object belongs to a cluster compete with one-another, each obtaining support from compatible objects and competitive pressure from the others. Competition will reduce the population of individuals that assume weakly supported hypotheses, while allowing populations assuming hypotheses with strong support to thrive. Eventually, all inconsistent hypotheses will be driven to extinction, while all the surviving ones will reach an equilibrium whereby they will all receive the same average support, hence exhibiting the internal coherency characterizing a cluster. As for the extinct hypotheses, they will provably have a lower support, thereby hinting to external incoherency. Each stable strategy identifies a particular cluster and, in general, two such strategies may have overlapping support, *i.e.*, clusters could have elements in common and thus overlap (Torsello, Rota Bulò, & Pelillo, 2008). Stable strategies can be found using *replicator dynamics*, a classic formalization of a natural selection process (Hofbauer & Sigmund, 1998; Weibull, 1997), but this is not the only available option. We will review also a more recent evolutionary dynamics called *Infection and Immunization Dynamics* (Rota Bulò & Bomze, 2011).

This game-theoretic formulation of the clustering problem overcomes the aforementioned limitations of the majority of the clustering approaches in the literature. Indeed, it makes no assumption on the underlying (individual) data representation: like graph-based clustering, it does not require that the elements to be clustered be represented as points in a vector space; it makes no assumption on the structure of the affinity matrix, being it able to work with asymmetric and even negative similarity functions alike; it does not require *a priori* knowledge on the number of clusters (since it extracts them sequentially); it leaves clutter elements unassigned; it allows clusters to overlap, since clusters are extracted independently from each other and no constraint is enforced to ensure a partition; it generalizes naturally to hypergraph clustering problems, *i.e.*, in the presence of high-order affinities (Rota Bulò & Pelillo, 2013), in which case the clustering game is played by more than two players.

Outline. The manuscript is organized as follows: Section 2 introduces the notion of dominant set and puts emphasis on theoretical aspects, in primis the different interpretations that exist from a combinatorial, graph-theoretic and game-theoretic perspective. Section 3 focuses on the algorithmic aspects of finding dominant sets. Here, replicator dynamics and infection and immunization dynamics will be reviewed, together with techniques that have been established to enumerate dominant sets, being their extraction sequential by nature. Section 4 provides an overview of applications of dominant sets that have appeared in the literature in the last decade, and reviews the application to image and video segmentation in some more details. Finally, Section 5 provides a glimpse to further variants and extensions of the dominant set framework that have been recently proposed.

2. Dominant sets and their properties

An instance of the clustering problem can be described by an edge-weighted graph, which is formally defined as a triplet $G = (V, E, \omega)$, where $V = \{1, \dots, n\}$ is a finite set of *vertices*, $E \subseteq V \times V$ is the set of oriented edges and $\omega : E \to \mathbb{R}$ is a real-valued function which assigns a weight to each edge. Vertices of G correspond to the objects to be clustered. Hence, we will use the terms vertices and objects interchangeably in the rest of the manuscript, when referring to elements of V. Edges represent neighborhood relationships among objects, and the edge-weights reflect similarity among linked objects. Moreover, we denote by A the objects' similarity matrix, *i.e.*, such that $A_{ij} = \omega(i, j)$ for all $i, j \in V$.

Within the dominant sets clustering framework, a dominant set represents a cluster, for it identifies a subset of objects satisfying the two basic properties of a cluster, *i.e.*,

- *internal homogeneity*: elements belonging to the cluster should have high mutual similarities;
- *maximality*: a cluster cannot be further extended by introducing external elements.

This section is mainly devoted to provide support to this claim and to formalize the notion of dominant set from a combinatorial perspective. In addition, we show relations between dominants sets and the well-known graph-theoretical notion of maximal clique, we provide a characterization in terms of optimization theory and finally establish a link to the game-theoretic notion of evolutionary stable strategy.

2.1. Definitions

Let G be a graph representing a clustering problem instance and let $C \subseteq V$ be a non-empty subset of objects. The *(average) weighted in-degree* of $i \in V$ with respect to C is defined as:

$$\mathrm{awindeg}_C(i) = \frac{1}{|C|} \sum_{j \in C} A_{ij},$$

where $|C|$ denotes the cardinality of C. Moreover, if $j \in C$ we define:

$$\phi_C(i, j) = A_{ij} - \mathrm{awindeg}_C(j),$$

which is a measure of the relative similarity of object i with object j with respect to the average similarity of object j with elements in C. Note that $\phi_C(i, j)$ can also take on negative values.

We are now in the position of providing the following inductive definition of the *weight* of an object i with respect to a set C:

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

The *total weight* of C is defined as:

$$W(C) = \sum_{i \in C} W_C(i).$$

Intuitively, $W_C(i)$ gives a measure of the support that object i receives from the objects in $C \setminus \{i\}$ relative to the overall mutual similarity of the objects in $C \setminus \{i\}$. Here positive values indicate that i has high similarity to $C \setminus \{i\}$. As an example for the graph in Fig. 1(a) we have

$$\begin{aligned} W_{\{1,2,3\}}(1) &= \phi_{\{2,3\}}(1, 2) W_{\{2,3\}}(2) + \phi_{\{2,3\}}(1, 3) W_{\{2,3\}}(3) \\ &= \phi_{\{2,3\}}(1, 2) \phi_{\{3\}}(2, 3) + \phi_{\{2,3\}}(1, 3) \phi_{\{2\}}(3, 2) = 10 \end{aligned}$$

and similarly $W_{\{1,2,3\}}(2) = 16$ and $W_{\{1,2,3\}}(3) = 18$, which yield a total weight of $W(\{1, 2, 3\}) = 44$. This means that every vertex receives a positive support from the other vertices in the graph. Moreover, the weight $W_C(j)$ captures the strength of the coupling between vertex j and the other elements in the set relative to the overall coupling among the vertices. Indeed in the specific example we have $W_{\{1, 2, 3\}}(1) < W_{\{1, 2, 3\}}(2) < W_{\{1, 2, 3\}}(3)$, which reflects the fact vertex 1 has the lightest incident edges, while vertex 3 has the heaviest ones. To give an example of negative coupling, consider Fig. 1(b), where we have that $W_{\{1, 2, 3, 4\}}(1) < 0$, which is due to vertex 1 being loosely coupled to the other vertices, the latter forming instead a cohesive group. Another scenario is shown in Fig. 1(c), where this time vertex 1 is strongly coupled to the other vertices, the latter being loosely connected to each other. In this case we have $W_{\{1, 2, 3, 4\}}(1) > 0$.

The properties of the weighting introduced before will form the basis for the formalization of the notion of dominant set as a notion of a cluster.

Definition 1 (Dominant set). A non-empty subset of objects $C \subseteq V$ such that $W(T) > 0$ for any non-empty $T \subseteq C$, is said to be a *dominant set* if:

1. $W_C(i) > 0$, for all $i \in C$,
2. $W_{C \cup \{i\}}(i) < 0$, for all $i \notin C$.

This definition provides conditions that correspond to the two main properties of a cluster: the first regards internal homogeneity, for it states that every object i within the cluster C should have a positive weight $W_C(i)$ with respect to the cluster; the second condition, instead, regards maximality, because as soon as we try to add a new element i to the cluster C, the weight $W_{C \cup \{i\}}(i)$ of the new element with respect to the extended cluster $C \cup \{i\}$ becomes negative. The condition $W(T) > 0$ is a technicality that is explained in more detail in Pavan (2004). As an example, in Fig. 1(d) we have that the vertices $\{1, 2, 3\}$ form a dominant set and this is intuitively due to the edges within the set (60, 70 and 90) being considerably larger than edges connecting vertices outside the set (≤ 25). Accordingly we have that the total weight $W(T)$ is positive for any $T \subseteq \{1, 2, 3\}$ and $W_{\{1, 2, 3\}}(i)$ is positive for each i in the cluster, while as soon as we try to extend it with a vertex outside the cluster we get a negative weight, *i.e.*, $W_{\{1, 2, 3, i\}}(i) < 0$ for $i = 4, 5$.

Hereafter, we will refer to dominant sets of a clustering problem instance with similarity matrix A as dominant sets of A.

2.2. Link to optimization theory

If we restrict the focus to symmetric similarities, *i.e.*, A is a symmetric matrix, then dominant sets can be characterized in terms of local solutions to the following constrained program, also known as *Standard Quadratic Program* (StQP) (Bomze, 1998):

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}) = \mathbf{x}^\top A \mathbf{x} \\ \text{subject to} \quad & \mathbf{x} \in \Delta \subset \mathbb{R}^n, \end{aligned} \quad (1)$$
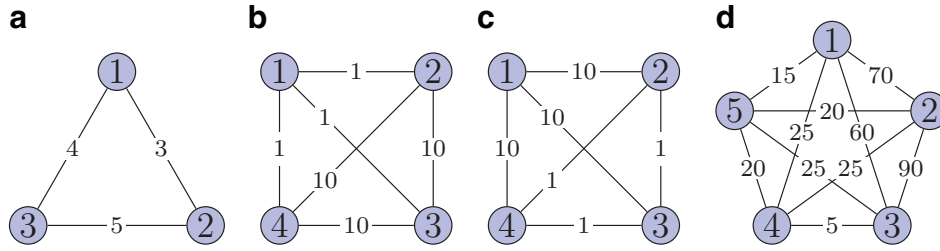
Fig. 1. Four examples of edge-weighted graphs.

where

$$\Delta = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{j \in V} x_j = 1 \text{ and } x_j \geq 0 \text{ for all } j \in V \right\}$$

is the so-called *standard simplex* of $\mathbb{R}^n$. The rest of this section is devoted to make this link to optimization theory explicit.

We say that $\mathbf{x}$ is a *strict local solution* of problem (1) if there exists a neighborhood $U \subseteq \Delta$ of $\mathbf{x}$ such that $f(\mathbf{x}) > f(\mathbf{z})$ for all $\mathbf{z} \in U \setminus \{\mathbf{x}\}$. The *support* $\sigma(\mathbf{x})$ of $\mathbf{x} \in \Delta$ is defined as the index set of the positive components in $\mathbf{x}$, *i.e.*, $\sigma(\mathbf{x}) = \{i \in V : x_i > 0\}$.

Definition 2 (Weighted characteristic vector). A non-empty subset $C \subseteq V$ admits a *weighted characteristic vector* $\mathbf{x}^C \in \Delta$ if it has positive total weight $W(C)$, in which case, we set:

$$x_i^C = \begin{cases} \frac{W_C(i)}{W(C)} & \text{if } i \in C \\ 0 & \text{otherwise} . \end{cases} \quad (2)$$

Note that by definition dominant sets always admit a characteristic vector. We come now to the main result of this section, which links dominant sets to (1):

Theorem 1. *If C is a dominant set of* A, *then its weighted characteristic vector* $\mathbf{x}^C$ *is a strict local solution to program* (1). *Conversely, if* $\mathbf{x}^*$ *is a strict local solution to* (1) *then its support* $\sigma = \sigma(\mathbf{x}^*)$ *is a dominant set of* A, *provided that* $W_{\sigma \cup \{i\}}(i) \neq 0$ *for all* $i \notin \sigma$.

Proof. See, Pavan and Pelillo (2007). □

According to this result, dominant sets can be put in one-to-one correspondence (modulo a technical condition) with strict local maximizers of a quadratic function over the simplex. As opposed to many other clustering algorithms, which try to find the *global* optimum of some energy function, dominant sets can be found by mining *local* solutions. This relation to optimization theory exists as long as we have a symmetric affinity matrix. We will see later in this section that if we relax this constraint, an interesting interpretation of dominant sets can be given in terms of a more general game-theoretic perspective.

2.3. Link to graph theory

Another interesting link can be established with graph theory, by considering dominant sets of clustering problem instances defined over undirected graphs. Within this context, we will show that a dominant set can be put in correspondence with the graph-theoretical notion of maximal clique.

Let $G = (V, E)$ be an undirected graph without self-loops, where $V = \{1, 2, \ldots, n\}$ is the set of vertices and $E \subseteq V \times V$ the set of edges. Two vertices $u, v \in V$ are *adjacent* if $(u, v) \in E$. A subset C of vertices in G is called a *clique* if all its vertices are mutually adjacent. It is a *maximal clique* if it is not a subset of other cliques in G. It is a *maximum clique* if it has maximum cardinality. The cardinality of a maximum clique of G is also called *clique number* and it is denoted by $\omega(G)$. The *adjacency matrix* of G is the $n \times n$ symmetric matrix A_G, where $(A_G)_{ij} = 1$ if $(i, j) \in E$, $(A_G)_{ij} = 0$ otherwise.

The adjacency matrix of an undirected graph can be regarded as the similarity matrix of a clustering problem instance and, consequently, the dominant sets framework can be used to find the clusters within the graph. Given the link to optimization theory mentioned in the previous section, we consider the following constrained quadratic program.

$$\begin{array}{ll} \text{maximize} & f_\alpha(\mathbf{x}) = \mathbf{x}^T (A_G + \alpha I) \mathbf{x} \\ \text{subject to} & \mathbf{x} \in \Delta \subset \mathbb{R}^n \end{array}, \quad (3)$$

where I the identity matrix, α is a real parameter and where Δ is the standard simplex of the n-dimensional Euclidean space.

Motzkin and Straus (1965) established a connection between the maximum clique problem and the program in (3) with $\alpha = 0$. Specifically, they related the clique number of G to global solutions $\mathbf{x}^*$ of the program through the formula $\omega(G) = (1 - f_0(\mathbf{x}^*))^{-1}$, and showed that a subset of vertices C is a maximum clique of G if and only if its characteristic vector $\mathbf{x}^C \in \Delta$ is a global maximizer of f_0 on Δ.[1] Pelillo and Jagota (1995), extended the Motzkin–Straus theorem by providing a characterization of maximal cliques in terms of local maximizers of f_0 in Δ.

A drawback of the original Motzkin–Straus formulation is the existence of "spurious" solutions, *i.e.*, maximizers of f_0 over Δ that are not in the form of characteristic vectors. This was observed empirically by Pardalos and Phillips (1990) and formalized later by Pelillo and Jagota (1995). In principle, spurious solutions represent a problem since, while providing information about the order of the maximum clique, do not allow us to easily extract its vertices. Fortunately, there is a straightforward solution to this problem which has been introduced by Bomze (1997). He, indeed, suggested to adopt the formulation in (3) and basically proved that for $0 < \alpha < 1$ all local maximizer of (3) are strict and in one-to-one correspondence with the characteristic vectors of the maximal cliques of G. By combining this result and Theorem 1 we obtain the following characterization of dominant sets:

Theorem 2. *Let G be an undirected graph with adjacency matrix* A_G *and let* $0 < \alpha < 1$. *Every maximal clique C of G is a dominant set of* $A_\alpha = A_G + \alpha I$. *Conversely, if C is a dominant set of* A_α *then C is a maximal clique of G.*

Finally, an extension of this result to the case of directed graphs can be found in Torsello et al. (2006).

This link between dominant sets and maximal cliques in the case of binary affinities, provides a further support to the claim that dominant sets are a meaningful notion of a cluster, because a maximal clique is the only entity within a graph that is internally homogeneous (every vertex in the clique is similar and thus connected to any other vertex within the clique) and maximal (by definition). Moreover, this relation suggests that one could regard

---

[1] In the original paper Motzkin and Straus proved the "only-if" part of this theorem. The converse however is a straightforward consequence of their result (Pelillo & Jagota, 1995).

the notion of dominant set as a generalization of the notion of maximal clique to edge-weighted graphs.

2.4. Link to game theory

The attentive reader should have noticed that the link between dominant set and optimization theory is conditioned on having a symmetric similarity matrix. Indeed, the link ceases to exist if this condition is relaxed, for the strict local maxima of a StQP with asymmetric payoff matrix A are not in one-to-one correspondence to dominant sets of a clustering problem instance with similarity matrix A, but $A + A^\top$. A natural question now is whether there is a theory beyond optimization that can deal with the asymmetric case as well. The positive answer is game theory.

The idea is to setup a symmetric, non-cooperative *game*, called *clustering game*, between two players. Data points V are the (pure) *strategies* available to the players and the similarity matrix A encodes their *payoff matrix*. The two players have perfect knowledge of the game and take independent decisions about the strategies to play. The payoff matrix summarizes the revenues that each player obtains when a pair of strategies is played, *i.e.*, if players 1 and 2 play strategies $(i, j) \in V \times V$, respectively, then player 1 receives payoff A_{ij}, while player 2 receives A_{ji}.[2]

A *mixed strategy* $\mathbf{x} \in \Delta$ is a probability distribution over the set of pure strategies, which models a stochastic playing strategy of a player. If player 1 and 2 play mixed strategies $(\mathbf{x}_1, \mathbf{x}_2) \in \Delta \times \Delta$, then the expected payoffs for the players are $\mathbf{x}_1^\top A \mathbf{x}_2$ and $\mathbf{x}_2^\top A \mathbf{x}_1$, respectively.

The choice of bounding the payoff function to the similarity matrix is very natural, as it may force some kind of coordination between the two players towards the selection of objects belonging to the same cluster. Clusters are indeed sets of objects having high mutual similarities. Therefore, only by selecting objects belonging to the same cluster, each player is able to maximize his expected payoff. This suggest that a mixed strategy for the player can be interpreted as his/her hypothesis of cluster membership. However, since every object is intuitively strongly similar to itself, setting the diagonal of the payoff matrix to zero, or in general to a sufficiently low value, is of fundamental importance. Otherwise, the best strategy for each player would be to coordinate the selection towards exactly the same object, which is useless for the clustering purposes. On the other hand, by penalizing perfect coordination, one forces the players to anti-coordinate, resulting eventually in an equilibrium, where the two players beliefs about the cluster membership are conflicting, which is again unwanted. A third, desired possibility is that the players end up with a so-called symmetric (Nash) equilibrium, where the beliefs about cluster membership coincide.

A *Nash equilibrium* is a mixed-strategy profile $(\mathbf{x}_1, \mathbf{x}_2) \in \Delta \times \Delta$ such that no player can improve the expected payoff by changing his playing strategy, given the opponent's strategy being fixed, *i.e.*, the following holds

$$\mathbf{y}_1^\top A \mathbf{x}_2 \le \mathbf{x}_1^\top A \mathbf{x}_2 \qquad \mathbf{y}_2^\top A \mathbf{x}_1 \le \mathbf{x}_2^\top A \mathbf{x}_1 \,, \tag{4}$$

for all $(\mathbf{y}_1, \mathbf{y}_2) \in \Delta \times \Delta$. A Nash equilibrium $(\mathbf{x}_1, \mathbf{x}_2)$ is *symmetric* if $\mathbf{x}_1 = \mathbf{x}_2$. For notational convenience, a symmetric Nash equilibrium $(\mathbf{x}, \mathbf{x})$ is identified by the single strategy $\mathbf{x} \in \Delta$ played by the two players. In the case of a symmetric Nash equilibrium $\mathbf{x} \in \Delta$ the two conditions in (4) collapse into the following one:

$$\mathbf{y}^\top A \mathbf{x} \le \mathbf{x}^\top A \mathbf{x} \,, \tag{5}$$

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

From the clustering game perspective a symmetric Nash equilibrium is a condition where both players agree on the same hypothesis of cluster membership and no player has incentives to deviate from it. Moreover, condition in Eq. (5) implies that[3]

$$\begin{cases} (A\mathbf{x})_i = \mathbf{x}^\top A \mathbf{x} & i \in \sigma(\mathbf{x}) \,, \\ (A\mathbf{x})_i \le \mathbf{x}^\top A \mathbf{x} & i \notin \sigma(\mathbf{x}) \,, \end{cases}$$

which can be regarded as a form of internal homogeneity of the cluster represented by the support $\sigma(\mathbf{x})$ of $\mathbf{x}$, for the expected similarity of any element in the cluster with any other element in the cluster will be the same. The Nash condition, however, does not necessarily guarantee the maximality condition, which can nevertheless be ensured by considering a refinement of the Nash equilibrium known as Evolutionary Stable Strategy. A symmetric Nash equilibrium $\mathbf{x} \in \Delta$ is an *Evolutionary Stable Strategy* (ESS) if in addition to (5) it satisfies the following stability condition

$$\mathbf{y}^\top A \mathbf{x} = \mathbf{x}^\top A \mathbf{x} \Rightarrow \mathbf{x}^\top A \mathbf{y} < \mathbf{x}^\top A \mathbf{x} \,, \tag{6}$$

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

Finally, the notion of ESS equilibrium and the one of dominant sets can be regarded as two different sides of the same coin, as stated in the following theorem:

Theorem 3. *Let* A *be the similarity matrix of a clustering problem instance and let* Γ *be the corresponding clustering game. If C is a dominant set of* A *then its characteristic vector* $\mathbf{x}^C$ *(see,* Eq. (2)) *is an ESS of* Γ. *Conversely, if* $\mathbf{x}$ *is an ESS of* Γ, *then* $C = \sigma(\mathbf{x})$ *is a dominant set of* A, *provided that* $(A\mathbf{x})_i \ne \mathbf{x}^\top A \mathbf{x}$ *for all* $i \notin C$.

Proof. See Torsello et al. (2006) and Rota Bulò (2009). □

This result establishes a one-to-one correspondence (again modulo a technical condition) between dominant sets of a clustering problem instance and the ESSs of the related clustering game.

2.5. On the number of dominant sets

We conclude the section with some considerations about the number of dominant sets that we can have in a clustering problem. Most of the clustering algorithms, which are partitional, can have a number of clusters which is upper-bounded by the number of objects to cluster. This is however not the case for the dominant sets framework, which does not adhere to a partitional paradigm, but admits also overlapping clusters. Indeed, the set of dominant sets for a given similarity matrix forms a Sperner family, *i.e.*, a family of sets such that no element in the family is a proper subset of another element, as a direct consequence of the maximality of a dominant set. As for the cardinality of this set, there are different ways to prove that there might exist similarity matrices with exponentially many dominant sets. Indeed, it is sufficient to exploit the relation that exist between dominant sets and maximal cliques, strict local optima of constrained quadratic forms and ESSs. From graph theory it is well-known that there exist graphs with exponentially many maximal cliques (Moon & Moser, 1965). Similarly, in optimization theory it is known that there could be exponentially many strict local optima of constrained quadratic forms (Bomze, Schachinger, & Ullrich, 2017; Broom, Cannings, & Vickers, 1993). Finally, in evolutionary game theory it is known that there could be games with exponentially many ESSs (Broom, 2000).

---

[2] the game is non-cooperative because each player takes independent decisions, while the symmetry derives from both players sharing the same payoff matrix.

[3] By contradiction, if there exists $j \in V$ such that $(A\mathbf{x})_j > \mathbf{x}^\top A \mathbf{x}$, then the indicator vector of j (taken as $\mathbf{y}$) would violate Eq. (5). Hence, $(A\mathbf{x})_i \le \mathbf{x}^\top A \mathbf{x}$ for all $i \in V$ and the following inequality holds for any $j \in \sigma(\mathbf{x})$: $\mathbf{x}^\top A \mathbf{x} = x_j (A\mathbf{x})_j + \sum_{i \in \sigma(\mathbf{x}) \backslash \{j\}} x_i (A\mathbf{x})_i \le x_j (A\mathbf{x})_j + (1 - x_j) \mathbf{x}^\top A \mathbf{x}$. Therefrom, it follows that $(A\mathbf{x})_j = \mathbf{x}^\top A \mathbf{x}$ for all $j \in \sigma(\mathbf{x})$.

3. Clustering using dominant sets

In the previous section, we introduced the notion of dominant set and we showed some of its properties, but we did not mention the way clustering effectively takes place. A naive, brute-force strategy might be to enumerate all subsets $C \subseteq V$ and verify if the conditions of Definition 1 hold. This solution is clearly highly inefficient and in this section we will provide two alternative strategies. Both strategies have roots in the game theory domain and it should not be surprising that methods developed in this context can be used to find dominant sets, given the tight relation that exists between dominant sets and the game-theoretic notion of equilibrium (see Section 2.4).

3.1. Replicator dynamics

The Replicator Dynamics (RD) are deterministic game dynamics that have been developed in evolutionary game theory. This theory originated in the early 1970s as an attempt to apply the principles and tools of game theory to biological contexts, with a view to model the evolution of animal, as opposed to human, behavior (see the classical work by Maynard Smith, 1982 who pioneered the field). It considers an idealized scenario whereby individuals are repeatedly drawn at random from a large, ideally infinite, population to play a two-player game. In contrast to classical game theory, here players are not supposed to behave rationally or to have complete knowledge of the details of the game. They act instead according to an inherited behavioral pattern, or pure strategy, and it is supposed that some evolutionary selection process operates over time on the distribution of behaviors. A general class of evolution equations is given by the following set of ordinary differential equations (Weibull, 1997):

$$\dot{x}_i = x_i g_i(\mathbf{x}) \tag{7}$$

for $i = 1 \dots n$, where a dot signifies derivative with respect to time and $g = (g_1, \dots, g_n)$ is a function with open domain containing Δ. Here, the function g_i specifies the rate at which pure strategy i replicates. It is usually required that the growth function g is *regular* (Weibull, 1997), which means that it is Lipschitz continuous and that $g(\mathbf{x})^\top \mathbf{x} = 0$ for all $\mathbf{x} \in \Delta$. The former condition guarantees that the system of the differential equation (7) has a unique solution through any initial population state. The latter condition, instead, ensures that the simplex Δ is invariant under (7), namely, any trajectory starting in Δ will remain in Δ.

A point $\mathbf{x}$ is said to be a *stationary* (or equilibrium) point for our dynamical systems, if $\dot{x}_i = 0$ ($i \in S$). A stationary point $\mathbf{x}$ is (Lyapunov) *stable* if for every neighborhood U of $\mathbf{x}$ there exists a neighborhood V of $\mathbf{x}$ such that $\mathbf{x}(0) \in V$ implies $\mathbf{x}(t) \in U$ for all $t \geq 0$. A stationary point is said to be *asymptotically stable* if any trajectory starting in its vicinity will converge to it as $t \to \infty$.

Payoff-monotonic game dynamics represent a wide class of regular selection dynamics for which useful properties hold. Intuitively, for a payoff-monotonic dynamics the strategies associated to higher payoffs will increase at a higher rate. Formally, a regular selection dynamics (7) is said to be *payoff-monotonic* if

$$g_i(\mathbf{x}) > g_j(\mathbf{x}) \Leftrightarrow (\mathbf{Ax})_i > (\mathbf{Ax})_j$$

for all $\mathbf{x} \in \Delta$ and $i, j \in V$.

Although this class contains many different dynamics, it turns out that they share a lot of common properties. To begin, they all have the same set of stationary points. Indeed, $\mathbf{x} \in \Delta$ is a stationary point under any payoff monotonic dynamics if and only if $(\mathbf{Ax})_i = \mathbf{x}^\top \mathbf{Ax}$ holds for all $i \in \sigma(\mathbf{x})$ (Weibull, 1997).

A well-known subclass of payoff-monotonic game dynamics is given by

$$\dot{x}_i = x_i \left[\phi((\mathbf{Ax})_i) - \sum_{j \in V} x_j \phi\big((\mathbf{Ax})_j\big) \right],$$

where $\phi(u)$ is an increasing function of u. These models arise in modeling the evolution of behavior by way of imitation processes, where players are occasionally given the opportunity to change their own strategies (Weibull, 1997).

When ϕ is the identity function, that is, $\phi(u) = u$, we obtain the standard continuous-time *replicator equations*,

$$\dot{x}_i = x_i \left[(\mathbf{Ax})_i - \mathbf{x}^\top \mathbf{Ax} \right], \tag{8}$$

whose basic idea is that the average rate of increase $\dot{x}_i / x_i$ equals the difference between the average fitness of strategy i and the mean fitness over the entire population.

Another popular model arises when $\phi(u) = e^{ku}$, where k is a positive constant. As k tends to 0, the orbits of this dynamics approach those of the standard, first-order replicator model, slowed down by the factor k; moreover, for large values of k, the model approximates the so-called best-reply dynamics (Hofbauer & Sigmund, 1998).

The replicator dynamics, and more in general any payoff monotonic dynamics, have the following properties (Hofbauer & Sigmund, 2003; Weibull, 1997):

Theorem 4. *Under any payoff monotonic dynamics the following holds true:*

- *a Nash equilibrium is a stationary point;*
- *a strict Nash equilibrium is asymptotically stable;*
- *a stationary point $\mathbf{x}^*$ that is the limit of an interior orbit,* i.e., *such that $\sigma(\mathbf{x}(t)) = V$ for all $t \geq 0$ and $\lim_{t \to \infty} \mathbf{x}(t) = \mathbf{x}^*$, is a Nash equilibrium*
- *a stable stationary point is a Nash equilibrium;*
- *an ESS is asymptotically stable.*

In general, the converse of the implications in Theorem 4 do not hold.

Furthermore, if we restrict our focus to symmetric payoff matrices, *i.e.*, $\mathbf{A} = \mathbf{A}^\top$, then stronger properties hold, as stated in the following theorem.

Theorem 5. *If $\mathbf{A} = \mathbf{A}^\top$ then the following holds:*

- *$\mathbf{x}^\top \mathbf{Ax}$ is strictly increasing along any non-constant trajectory of (8). In other words, for all $t \geq 0$ we have $\frac{d}{dt}[\mathbf{x}(t)^\top \mathbf{Ax}(t)] > 0$, unless $\mathbf{x}$ is a stationary point. Furthermore, any such trajectory converges to a (unique) stationary point;*
- *$\mathbf{x}$ is asymptotically stable if and only if $\mathbf{x}$ is an ESS.*

In order to implement the continuous-time replicator dynamics one can resort to some iterative method like, *e.g.*, the Runge-Kutta method, to find an approximate solution to the ordinary differential equations. Alternatively, one can adopt the discrete-time counterpart of (8), known as discrete-time replicator dynamics, which are given by

$$x_i(t+1) = x_i(t) \frac{(\mathbf{Ax}(t))_i}{\mathbf{x}(t)^\top \mathbf{Ax}(t)},$$

for $i \in V$. Note that the discrete-time replicator dynamics are simplex-invariant if $\mathbf{A}$ is nonnegative. This, however, is not a limitation because any payoff $\mathbf{A}$ preserves its equilibria by a constant shift. Hence, if $\mathbf{A}$ has negative entries, one can build a matrix $\bar{\mathbf{A}} = \mathbf{A} + \mu \mathbf{E}$ with positive entries by a proper choice of $\mu > 0$, where $\mathbf{E}$ is a matrix of all ones.

Since ESSs are asymptotically stable under the replicator dynamics according to Theorems 4 and 5, we can employ this

dynamics for the extraction of dominant sets, which are indeed in correspondence to ESSs as described in Section 2.4. Moreover, if we assume symmetric payoff functions, we recover again the link to optimization theory shown in Section 2.2, for Theorem 5 implies that the replicator dynamics locally maximize $\mathbf{x}^\top A\mathbf{x}$ over the simplex. Finally, motivated by the link to graph theory described in Section 2.3, replicator dynamics have also been employed as heuristics for the maximum clique problem (Pelillo & Torsello, 2006; Wu & Hao, 2015).

3.2. Infection and immunization dynamics

The *Infection and Immunization Dynamics* (InImDyn) are a class of discrete-time dynamics that have been introduced in Rota Bulò and Bomze (2011) to overcome some computational problems afflicting standard evolutionary dynamics, including the replicator dynamics. To mention a few, evolutionary dynamics like replicator dynamics have a quadratic space/time complexity per iteration, not all stationary points are Nash equilibria, they converge only in the limit of infinitely-many iterations, and the detection of the support of an equilibrium is cumbersome if the dynamics is stopped before proper convergence (*i.e.*, after finitely-many steps).

The InImDyn dynamics take the following form:

$$\mathbf{x}(t+1) = \delta_{\mathcal{S}(\mathbf{x})}(\mathbf{x})[\mathcal{S}(\mathbf{x}) - \mathbf{x}] + \mathbf{x}, \tag{9}$$

where we wrote $\mathbf{x}$ for $\mathbf{x}(t)$. The function $\mathcal{S} : \Delta \to \Delta$ is a *strategy selection* function that satisfies the following property:

$$\mathcal{S}(\mathbf{x}) = \begin{cases} \mathbf{y} & \text{for some } \mathbf{y} \in \Upsilon(\mathbf{x}) \text{ if } \Upsilon(\mathbf{x}) \neq \emptyset, \\ \mathbf{x} & \text{otherwise}, \end{cases} \tag{10}$$

where $Y(\mathbf{x})$ represents the set of so-called *infective strategies* for $\mathbf{x}$ defined as

$$\Upsilon(\mathbf{x}) = \{\mathbf{y} \in \Delta : (\mathbf{y} - \mathbf{x})^\top A\mathbf{x} > 0\}.$$

As long as there exists $\mathbf{y} \in Y(\mathbf{x})$, the strategy $\mathbf{x}$ cannot be an equilibrium by definition of infective strategy. If this is the case, the dynamics (9) blends $\mathbf{x}$ and $\mathbf{y}$ (infects $\mathbf{x}$ with $\mathbf{y}$) until the violation of the Nash condition caused by $\mathbf{y}$ vanishes, which happens by considering a mixing factor $\delta_{\mathbf{y}}(\mathbf{x})$ given by

$$\delta_{\mathbf{y}}(\mathbf{x}) = \begin{cases} \min\left\{\frac{(\mathbf{x}-\mathbf{y})^\top A\mathbf{x}}{(\mathbf{y}-\mathbf{x})^\top A(\mathbf{y}-\mathbf{x})}, 1\right\}, & \text{if } (\mathbf{y}-\mathbf{x})^\top A(\mathbf{y}-\mathbf{x}) < 0 \\ 1, & \text{otherwise}. \end{cases} \tag{11}$$

The InImDyn dynamics yields a fixed point when $\Upsilon(\mathbf{x}) = \emptyset$. If this is the case, $\mathbf{x}$ is also a Nash equilibrium:

Theorem 6. *Let $\mathbf{x} \in \Delta$ be a strategy. Then the following statements are equivalent:*

(a) $\Upsilon(\mathbf{x}) = \emptyset$: *there is no infective strategy for $\mathbf{x}$;*
(b) $\mathbf{x}$ *is a Nash strategy;*
(c) $\mathbf{x}$ *is a fixed point under dynamics* (9).

Proof. See Rota Bulò and Bomze (2011). □

Akin to the replicator dynamics, if we restrict to symmetric payoff matrices, we have that the average payoff is strictly increasing along any non-constant trajectory of InImDyn:

Theorem 7. *Let $\{\mathbf{x}(t)\}_{t \geq 0}$ be a trajectory of* (9). *Then for all $t \geq 0$,*

$$u(\mathbf{x}(t+1), \mathbf{x}(t+1)) \geq u(\mathbf{x}(t), \mathbf{x}(t)),$$

with equality if and only if $\mathbf{x}(t) = \mathbf{x}(t+1)$, provided that the payoff matrix is symmetric.

Proof. See Rota Bulò and Bomze (2011). □

Depending on the choice of $\mathcal{S}(\mathbf{x})$ in (9), we may obtain different dynamics. In the rest of the section, we review one

in particular, which is simple and leads to nice properties. Let $u \in \arg\max_{j \in V} (A\mathbf{x})_j$ and $v \in \arg\min_{j \in \sigma(\mathbf{x})}(A\mathbf{x})_j$ and let

$$\mathbf{z}^+ = \mathbf{e}^u \quad \text{and} \quad \mathbf{z}^- = \frac{x_v}{1 - x_v}(\mathbf{x} - \mathbf{e}^v) + \mathbf{x},$$

where $\mathbf{e}^u$ is a null vector except for the uth element being 1. The *pure* strategy selection function $\mathcal{S}_{\text{Pure}}$ is defined as

$$\mathcal{S}_{\text{Pure}}(\mathbf{x}) = \begin{cases} \mathbf{z}^+ & \text{if } (\mathbf{z}^+ - \mathbf{z}^-)^\top A\mathbf{x} > 0 \\ \mathbf{z}^- & \text{if } (\mathbf{z}^+ - \mathbf{z}^-)^\top A\mathbf{x} < 0 \\ \mathbf{x} & \text{otherwise}. \end{cases}$$

Since the search space for an infective strategy is reduced from Δ to a finite set, it is not obvious that $\mathcal{S}_{\text{Pure}}(\mathbf{x})$ is a well-defined selection function, *i.e.*, it satisfies (10). However, one can prove (Rota Bulò & Bomze, 2011) that there exists an infective strategy for $\mathbf{x}$ if and only if $\mathcal{S}_{\text{Pure}}(\mathbf{x})$ is infective for $\mathbf{x}$.

Another property that holds for Pure InImDyn, which is shared also by the replicator dynamics, is the characterization of ESS equilibria in terms of asymptotically stable points of the dynamics under symmetric payoff matrices.

Theorem 8. *A state $\mathbf{x}$ is asymptotically stable for InImDyn with $\mathcal{S}_{\text{Pure}}$ as strategy selection function if and only if $\mathbf{x}$ is an ESS, provided that the payoff matrix is symmetric.*

Proof. See Rota Bulò and Bomze (2011). □

This selection function exhibits the nice property of rendering the complexity per iteration linear in both space and time, as opposed to the replicator dynamics, which have quadratic space/time complexity per iteration (Rota Bulò, Pelillo, & Bomze, 2011). We refer to Rota Bulò and Bomze (2011) and Rota Bulò et al. (2011) for further analyses of the theoretical and computational properties of InImDyn, among which convergence in finitely-many steps and support separation in finite-time. We also refer the interested reader to Ullrich (2016) for a recent development of a continuous-time version of InImDyn.

3.3. Finding multiple clusters

The Replicator and InImDyn dynamics that we presented in the previous sections allow to find upon convergence a *single* dominant set. In principle, one aims at enumerating all dominant sets for the clustering problem instance at hand, however, this is in general computationally intractable because there might be exponentially many clusters (see Section 2.5). In practice, having an incomplete, but good, coverage of dominant sets is typically sufficient for application purposes. In this section we review some computational strategies for the (partial) enumeration of dominant sets.

Multi-start strategy. The first strategy is very naive and consists in re-starting the dynamics from different, random points in the simplex. This strategy is effective if the data consists of few clusters with large basins of attraction, such that the probability of sampling a point belonging to each basin is relatively high. Clearly, there is no guarantee that this procedure will not extract the same clusters multiple times. Hence, a post-processing step is devoted to identifying and prune repeated clusters. The multi-start strategy is optimal in the long-run, because it will eventually enumerate all dominant sets, but it might require a very large number of samples, thus being inefficient in practice.

Peeling-off strategy. This strategy follows a different philosophy and finds application mainly in cases where one is not interested in finding overlapping clusters. The idea is again very simple and consists in iteratively extracting a dominant set and removing its vertices. By doing so, there is the obvious guarantee that no dominant set will be extracted multiple times. However, this strategy

is not optimal because on one hand the number of dominant sets that can be potentially enumerated is upper-bounded by the total number of vertices; on the other hand, except for the first dominant set, all the following ones are not necessarily dominant sets of the original clustering problem instance, for every time vertices of a dominant set are peeled off, the underlying clustering problem instance changes. Despite the lack of optimality, the peeling-off strategy is computationally appealing and it has been used in practice in many applications.

Destabilization strategy. This strategy has been proposed in Torsello et al. (2008) and Rota Bulò, Torsello, and Pelillo (2009) and it is based on the replicator dynamics. We will sketch the method in intuitive terms here, while we refer to the original paper for more details. The goal is to retain an optimal solution as in the multi-start strategy, but, at the same time, to avoid getting trapped by the same basins of attraction multiple times. The idea is to iteratively render the dominant sets that are extracted over time (and only those) unstable to the replicator dynamics, in order to prevent their extraction at the succeeding iterations. In order to destabilize a dominant set C, it is sufficient to add a new vertex to the original graph, connected in a way that the characteristic vector $\mathbf{x}^C$ violates the ESS condition, while the set of non-extracted ESSs remains unchanged (see, Rota Bulò et al., 2009; Torsello et al., 2008 for details). Despite the theoretical validity of the approach, there are some practical issues with this method due to the possible emergence of periodic orbits that thwart the extraction of new dominant sets. Moreover, the computational complexity grows quadratically (assuming the number of RD iterations constant) with the number of extracted dominant sets.

Tabu-list strategy. The last strategy has been proposed in Kontschieder, Rota Bulò, Donoser, Pelillo, and Bischof (2012) and it is based on the InImDyn dynamics. Again, we will simply sketch the general idea and we refer to the original paper for more details. The tabu-list strategy pursues the goal of an incomplete enumeration of dominant sets akin to the peeling-off strategy, yielding nevertheless a good coverage and without sacrificing the optimality of the solution, *i.e.*, it guarantees that the extracted clusters are dominant sets of the original clustering problem instance. The idea is to keep a tabu-list during the clustering process, *i.e.*, a list of vertices that should not be taken into account during the dominant set extraction. In some sense, this is what also the peeling-off strategy is doing, for all vertices of already-extracted dominant sets are added to the tabu-list. However, in order to prevent the possibility of extracting invalid dominant sets, the tabu-list strategy checks the validity of putative dominant sets found under the tabu-list constraint. In the specific, at each dominant set extraction iteration, a dominant set $C \subset V$ is found by running InImDyn with a random initialization under the constraint that no vertex in the tabu-list $T \subset V$ can be used, *i.e.*, $C \subseteq V \backslash T$. This constraint prevents C from being a valid dominant set in general, *i.e.*, C might not be a dominant set with respect to the full vertex set V. To examine the validity of the putative dominant set C, the InImDyn dynamics is run again, initialized to the characteristic vector $\mathbf{x}^C$ and allowed to use *all* the vertices (including those in the tabu-list). Let $C' \subseteq V$ be the dominant set found, which is valid, but might not be novel. There are then 2 possible scenarios, in which C' can be regarded as a new, valid dominant set to be added to the final clustering result: 1) $C' \not\subseteq T$, 2) $C' \subseteq T$, but C' was never found before. In the first case, the tabu-list is updated by adding any vertex in $C' \backslash T$ (*e.g.*, a random one), while in the second case the tabu-list is kept as it is. If, instead, it turns out that $C' \subseteq T$, but C' has already been found, we reject C' and add to the tabu-list any vertex in $V \backslash T$. The tabu-list strategy guarantees that either a valid and novel dominant set is extracted at each iteration,

or the tabu-list is increased by one, thus limiting the number of iterations to $max(|V|, \#$ of clusters found$)$. Note that one could also add more vertices at time to the tabu-list to speed-up the clustering process at the cost of eventually finding a smaller number of dominant sets, as is actually done in Kontschieder et al. (2012).

4. Applications

The dominant sets clustering framework has found a large variety of applications in different fields. In bioinformatics, we find applications to gene expression analysis from microarray data (Fu et al., 2005; Teng & Chan, 2008) and to Tag Single Nucleotide Polymorphism (SNPs) selection (Frommlet, 2010). In computer vision, dominant sets have been employed for image segmentation (Pavan & Pelillo, 2003b, 2007; Torsello & Pelillo, 2009), perceptual grouping (Torsello et al., 2006), 3D rigid shape registration (Albarelli, Rodolà, & Torsello, 2010), 3D reconstruction (Albarelli, Rodolà, & Torsello, 2012; Mauro, Riemenschneider, Van Gool, & Leonardi, 2013), content-based image retrieval (Wang, Ye, Wang, & Wang, 2008), video analysis (Besiris, Makedonas, Economou, & Fotopoulos, 2009; Sakarya & Telatar, 2008), object detection (Kontschieder et al., 2012; Yang, Liu, & Latecki, 2012) and human action recognition (Hamid et al., 2009). In security, we have applications for the detection of anomalous activities in video streams (Hamid et al., 2005) and for detecting malicious activities in the internet (Dacier, Pham, & Thonnard, 2009; Pouget, Dacier, Zimmerman, Clark, & Mohay, 2006; Thonnard & Dacier, 2009). In social signal analysis, dominant sets clustering has been used for the detection of F-formations (Hung & Kröse, 2011) and for activity analysis in crowded environments (Tran, Gala, Kakadiaris, & Shah, 2014). In climatology, dominant sets have been used for classifying atmospheric circulation types (Zagouras, Argiriou, Flocas, Economou, & Fotopoulos, 2012). Finally, in medical data analysis we find applications to the analysis of functional Magnetic Resonance Imaging (fMRI) data (Müller, Neumann, Grigutsch, von Cramon, & Lohmann, 2007; Neumann, von Cramon, Forstmann, Zysset, & Lohmann, 2006), for brain activity analysis (Adamos, Laskaris, Kosmidisc, & Theophilidis, 2012; Dimitriadis, Laskaris, Del Rio-Portilla, & Koudounis, 2009) and for 3D ultrasound registration (Banerjee et al., 2015).

In the rest of the section, we will review the application of dominant sets to image and video segmentation, originally proposed by Pavan and Pelillo in their seminal works (Pavan & Pelillo, 2003b, 2007), and an application of dominant sets to video summarization proposed in Besiris et al. (2009).

4.1. Image segmentation

Image segmentation is the problem of grouping the pixels of a digital image into homogeneous regions with the goal of simplifying the image into a representation, which is more meaningful and easier to analyze. This problem can be naturally regarded as a clustering problem, where pixels represent the objects to cluster.

The outcome of the clustering (or segmentation) process is directly reflected by the way the pixel similarity matrix has been set up. Standard approaches take into account a variety of cues like brightness, color, texture, edges or combinations thereof to build a good similarity measure. The similarity matrix construction adopted by Pavan and Pelillo consists in a Gaussian kernel measuring the similarity of some feature representation of the pixels, *i.e.*, the similarity between pixels i and j is given by

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

where σ is a user-defined scale value, and $\Psi(i)$ is a feature vector for pixel i, which can hold its normalized intensity value, a

3-dimensional vector with color information (in RGB, HSV or other color spaces), or a vector $\Psi(i) = (|I * f_1|(i), \ldots, |I * f_m|(i))$ with texture information, which is obtained by convolving the image with a bank of m filters f_u, $1 \leq u \leq m$, (typically Gabor filters), i.e., $|I*f_u|(i)$ is the value of pixel i after convolving image I with the linear kernel f_u.

Once the similarity matrix is created, dominant sets can be extracted and each dominant set will then correspond to a segment of the final segmentation. Since dominant sets allow for the extraction of one segment at time, the enumeration approach that is typically adopted for segmentation is the simple peeling-off strategy described in Section 3.3, for segments usually form a partition of the original image.

A drawback of the pairwise clustering approaches like dominant sets for segmentation is their quadratic complexity with respect to the number of pixels, which renders the clustering problem intractable already for low resolution image formats (e.g., a small 640×480 image has already around 300,000 pixels). There are typically three possible solutions to this issue: aggregate pixels into so-called *super-pixels* by some simple quantization and cluster super-pixels; employ a sparse similarity matrix by exploiting the local connectivity of pixels over the image; cluster a random subset of pixels and extend the segmentation to all pixels a posteriori. The first and last approaches are viable through the dominant set clustering approach. In particular the third one is a technique called *out-of-sample extension*, which has been proposed by Pavan and Pelillo (2005) and will be described in more details in the next section. The second technique, instead, is not compatible with the dominant set clustering approach, since dominant sets are expected to be *densely* connected clusters (remind for instance the relation to maximal cliques). Instead, a sparse representation of the similarity matrix compromises the compactness of the clusters, for most of the edges would miss. There are other clustering techniques, such as spectral approaches (Nascimento & de Carvalho, 2011), which are better suited to sparse similarities.

In Fig. 2 we report some exemplary images (top row) from Pavan and Pelillo (2005) that have been segmented via dominant sets (middle row). For the sake of comparison we show (bottom row) also the segmentation results obtained by Normalized Cuts (NCuts) (Shi & Malik, 2000), which is a partitional, spectral clustering approach. The similarity matrix has been constructed by taking into account only pixel brightness and color information. In the segmentation results, pixels belonging to the same segment share the same gray level. To cope with the large dimensionality of the similarity matrix, NCut was run on a sparse representation of the matrix, which preserves only edges between neighboring pixels, while dominant sets have been run on a subset of 2–3% of pixels using the out-of-sample extension technique (described in the next section). As can be seen, the segmentations obtained with dominant sets are in general cleaner than those obtained with NCut, the latter having the tendency of producing oversegmented results.

4.2. Video segmentation

Video segmentation considers a slightly more general setting, where we have a sequence of images to segment rather than a single image. Again the goal is to organize the pixels into meaningful regions, but this time regions could span multiple images in the video. The construction of the similarity measure follows the same principles mentioned before, with the additional possibility of taking into account also the temporal dimension.

In Fig. 3 we present exemplary video segmentations from Torsello, Pavan, and Pelillo (2005) obtained on a video sequence using dominant sets and NCut (we show only 3 frames). The similarity matrix has been constructed by taking into ac-

count both color information and texture information. Texture information has been computed by employing a bank of Gabor filters spanning 6 orientations and 4 scales. To cope with the high dimensionality of the problem, dominant sets have been run on a sub-sampled version of the original similarity matrix and extended the result using the out-of-sample extension technique, while NCut has been run using the so-called Nyström method (Fowlkes, Belongie, & Malik, 2004), which allows to perform the clustering without employing the whole similarity matrix, akin to the out-of-sample extension. As we can see, also in this case the segmentation obtained by dominant sets has a better quality, for it separates the roofs of the houses from the tree in the foreground. We refer to the original paper for further examples.

4.3. Video summarization

A video summary (*a.k.a.* video skim) can be regarded as a sequence of video segments, or keyframes, that represents a short synopsis of the video content. The goal, among others, is to provide a user with a quick, lightweight preview of the video, or reduce the amount of information required for video indexing. This problem can be rephrased as a clustering of video keyframes, by providing a proper similarity measure among keyframes, and by deriving the video summary from the sequence of cluster representatives. This is the approach pursued in Besiris et al. (2009), where the dominant sets clustering framework has been chosen to perform the extraction of representatives thanks to its versatility, not requiring a pre-fixed number of clusters and being robust to outliers. Effectively, Besiris et al. (2009) do not create a similarity graph using all video keyframes, since this would be computationally and memory-wise prohibitive, but they randomly sample a subset of keyframes and perform a first, coarse quantization step by using a feature-based clustering approach (namely, Fuzzy C-Means), which is computationally cheaper than dominant sets clustering. To perform the latter clustering task, each keyframe is cast into a low-dimensional feature vector, computed in terms of a HSV color-histogram. The outcome of the quantization step consists in n keyframe prototypes, where n is substantially smaller than the video length but larger than the expected video summary length. The prototypes are used to build a similarity matrix (we refer the interested reader to the original paper for more details) from which dominant sets are extracted using a peeling-off strategy (see, Section 3.3). Finally, a representative from each dominant set is determined, which in turn represents a frame of the video summary. Experiments conducted in Besiris et al. (2009) on video sequences against state-of-the-art methods, and using standard evaluation metrics, have shown the superiority of the proposed approach.

5. Extensions

The dominant sets clustering framework has been extended in different ways over the last decade. Some authors have investigated possible ways to tackle dominant set clustering in the presence of large-scale data (e.g., image segmentation with millions of pixels) (Pavan & Pelillo, 2005); other works have focused on hierarchical clustering within the dominant set framework (Pavan & Pelillo, 2003a; Torsello & Pelillo, 2009); a generalization of the latter work has been proposed to tackle the extraction of support-constrained dominant sets (Zemene & Pelillo, 2016); some research efforts have also been devoted to generalize the dominant sets to high-order similarities (Rota Bulò & Pelillo, 2009; 2013); finally, some other authors have proposed a variant to the high-order dominant set formulation that implicitly imposes a constraint on the dominant set size (Liu, Latecki, & Yan, 2010). In the rest of this section we will review in more details the aforementioned works.
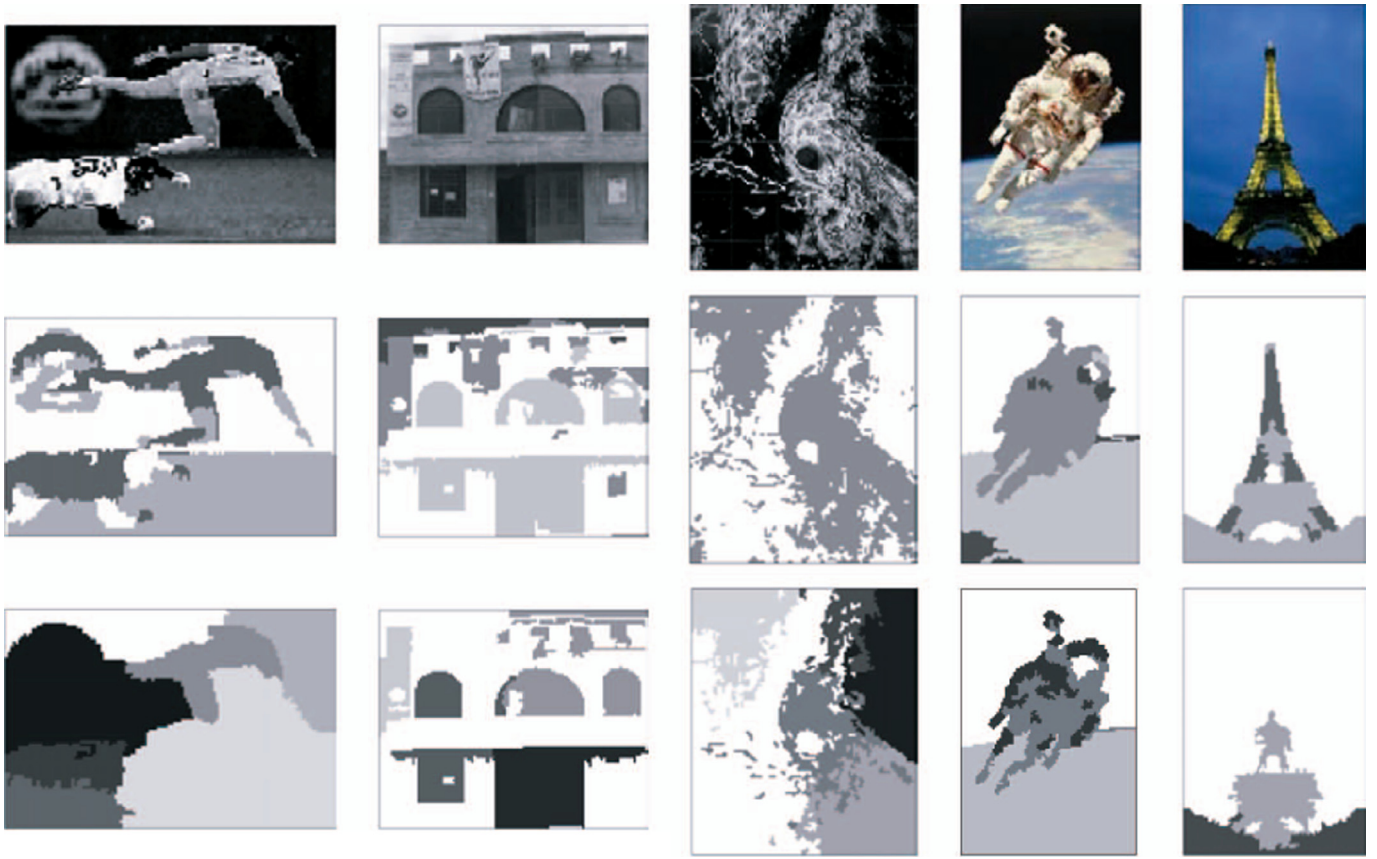
Fig. 2. Examples of image segmentations. Top: original images. Middle: segmentations using dominant sets. Bottom: segmentations using normalized cuts. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article).

5.1. Out-of-sample extension

The dominant set framework in general does not scale well with large datasets due to its intrinsic dependence on the similarity matrix, which has quadratic complexity. To overcome this limitation, Pavan and Pelillo (2005) introduced an out-of-sample extension method for the dominant set framework, which consists in performing a dominant set clustering of a random, small subset of the data and gradually integrating the clusters with the data points that have been omitted in the first place (i.e., the out-of-sample points). Note that this idea is not only useful in the presence of large datasets, but can also be used in scenarios where the data to cluster is not entirely available from the beginning. Clearly, the final outcome of this procedure will not necessarily be a proper dominant set clustering of the original dataset, but an approximation thereof.

The idea behind the out-of-sample strategy, which is naturally suggested by the definition of dominant set (i.e., Definition 1), consists in aggregating an out-of-sample data point i to a dominant set cluster S if $W_{S \cup \{i\}}(i) > 0$. According to this rule, the same data point can be assigned to multiple dominant set, thus preserving the property that dominant sets might overlap. It might also happen that it gets assigned to no dominant set, thus remaining unclassified. Unfortunately, the direct computation of the sign of $W_{S \cup \{i\}}(i)$ is prohibitive due to its recursive nature and, for this reason, Pavan and Pelillo (2005) resorts to an approximation thereof: if $\mathtt{A} \in \mathbb{R}^{n \times m}$ is the similarity matrix of all data points in $\{1, \ldots, n\}$ against an in-sample data point in $\{1, \ldots, m\}$ (with $m < n$) then an out-of-sample data point $m < i \leq n$ is aggregated to a dominant set $S \subseteq \{1, \ldots, m\}$ if $(\mathtt{A}\mathbf{x}^S)_i > \mathbf{x}^{S\top}\mathtt{A}\mathbf{x}^S$, where $\mathbf{x}^S$ is the characteristic vector of S. We refer the interested reader to the original paper for further details.

5.2. Hierarchical clustering

The set of clusters that one finds with the dominant set clustering framework is entirely determined by the similarity matrix. For this reason, there is no explicit parameter that controls the number of clusters. In some applications, however, it is useful to have a hierarchy of clusters, which allow to perform data analysis at different scale-levels, akin to the result one would obtain with, e.g., linkage-type clustering algorithms (Sibson, 1973).

A way to enable the extraction of dominant sets at different scales has been proposed in Pavan and Pelillo (2003a). The idea is to define a parametrized family of similarity matrices, where the parameter controls a regularization term that influences the size of the extracted dominant sets. In the specific, given a similarity matrix $\mathtt{A}$, the authors define $\mathtt{A}_\gamma = \mathtt{A} - \gamma \mathtt{I}$, where $\gamma > 0$ is the regularizing parameter. By considering $\gamma = 0$ one recovers the original matrix $\mathtt{A}$, but as γ increases small-sized dominant sets vanish and larger ones pop up, until all data points get enclosed within a single dominant set. In particular, the authors show that by taking $\gamma > \lambda_{\max}(\mathtt{A}_S)$, i.e., γ larger than the largest eigenvalue of the submatrix of $\mathtt{A}$ with indices in S, one prevents dominant sets in $\mathtt{A}_\gamma$ from belonging to S.

This result inspired a divisive hierarchical clustering algorithm based on dominant sets that starts with $\gamma = |V| - 1 > \lambda_{\max}(\mathtt{A})$ and iteratively reduces γ to zero, while collecting the dominant sets at each stage. The set of dominant sets extracted at all γ-scales forms the final hierarchical clustering result. The hierarchical clustering that one obtains with the described approach does not necessarily produce a proper hierarchy, for a dominant set extracted at some scale γ might not be a subset of any dominant set extracted at the previous (larger) scale-level.
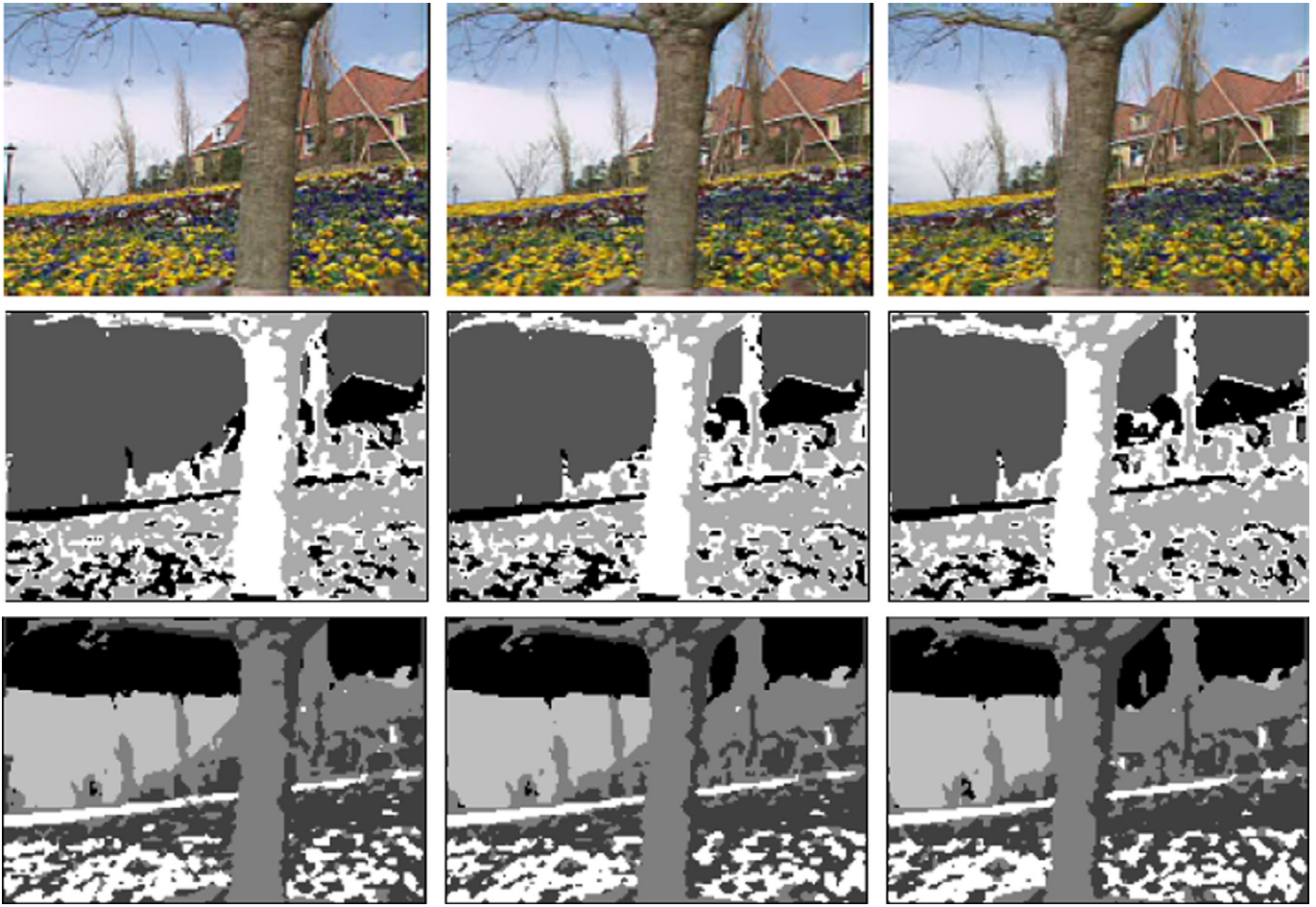
Fig. 3. Example of video segmentation (3 frames shown). Top: original sequence. Middle: segmentation using dominant sets. Bottom: segmentation using normalized cuts with Nyström method. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article).

5.3. Constrained clustering

The idea of adopting a regularized formulation of dominant set to loosely constrain the size of the extracted cluster, as proposed in Pavan and Pelillo (2003a), has recently been generalized (Zemene & Pelillo, 2016) with the purpose of extracting dominant sets that contain a user-specified subset of objects $S \subseteq O$. The intuition is to replace the original matrix $\mathtt{A}$ with a family of matrices $\mathtt{A}_{\gamma,S}$ parametrized by γ and S, and defined as $\mathtt{A}_{\gamma,S} = \mathtt{A} - \gamma \mathtt{I}_S$. This construction is akin to the one we have described in the previous section with the identity matrix $\mathtt{I}$ being replaced by $\mathtt{I}_S$, which is like $\mathtt{I}$ but with zeros in correspondence to elements of S, i.e., $(\mathtt{I}_S)_{ij} = 1$ if $i = j$ and $i \in V \backslash S$, and zero else. The authors then provide a sufficient condition on γ under which all dominant sets of $\mathtt{A}_{\gamma,S}$ become super sets of S, namely γ should be chosen larger than the largest eigenvalue of the principal submatrix of $\mathtt{A}$ indexed by the elements of $V \backslash S$. It is worth mentioning that dominant sets extracted from $\mathtt{A}_{\gamma,S}$ do not necessarily correspond to dominant sets of the original matrix $\mathtt{A}$. Moreover, the condition is not a necessary condition, so there could exists values of γ lower than the proposed bound for which the required property is still satisfied.

Zemene and Pelillo (2016) have used their result to address the problem of interactive image segmentation, where the segmentation is guided by a user, who indicates some of the elements belonging to the segment to be extracted (i.e., she provides the set S).

5.4. High-order clustering

The extension of the dominant set framework to high-order similarities has been introduced in Rota Bulò and Pelillo (2009); 2013). In this work, the goal is to cluster hypergraphs as opposed to the simple graphs used in the dominant set framework, where a *hypergraph* is a generalization of the notion of graph having tuples of vertices (rather than just pairs) as edges. The extended formulation assumes hypergraphs to be super-symmetric, i.e., edges are regarded as sets of vertices rather than tuples, and k-uniform, i.e., every edge has k vertices. Accordingly, a (super-symmetric) hypergraph is a triplet $H = (V, E, \omega)$ with vertex set V, edge set $E \subseteq \binom{V}{k}$ and edge-weight function $\omega : E \to \mathbb{R}$, where $\binom{V}{k}$ denotes the set of k-subsets of V.

Akin to Section 2.4, a dominant set on hypergraphs is defined in terms of the game-theoretic notion of ESS of a clustering game involving k-players, instead of just two, and having the high-order similarities as payoffs. In Rota Bulò and Pelillo (2013), it is shown that the extended notion of dominant set still satisfies in some sense the internal homogeneity and maximality properties that pertain to a cluster. Moreover, a graph-theoretic intuition of the extended dominant set is also provided by showing that the support of a dominant set is a two-cover, i.e., a set of vertices C such that every pair of vertices in C belongs to at least one edge entirely contained in C, i.e., $\forall \{i, j\} \in \binom{C}{2}. \exists e \in E. \{i, j\} \subseteq e \subseteq C$. We

skip the mathematical details of the game-theoretic formulation and we refer the interested reader to the original paper.

Besides the game-theoretic characterization, there exists also a link to optimization theory as there was for the standard dominant set. Specifically, dominant sets of a hypergraph $H = (V, E, \omega)$ are in one-to-one correspondence to strict local maxima of the following simplex-constrained polynomial optimization:

$$\text{maximize} \quad f_H(\mathbf{x}) = \sum_{e \in E} \omega(e) \prod_{j \in e} x_j \tag{12}$$
$$\text{subject to} \quad \mathbf{x} \in \Delta \subset \mathbb{R}^n,$$

which entails (1) as a special case. Dominant sets are finally extracted through an iterative procedure that is obtained from a result in probability theory known as Baum–Eagon inequality (Baum & Eagon, 1967). The resulting dynamics are akin to Replicator Dynamics and share similar convergence guarantees.

The optimization program in (12) has been modified in Liu et al. (2010) by adding the constraint $0 \leq x_i \leq \epsilon$ for all $i = 1, \dots, n$. The effect of this modification is to impose an implicit restriction on the cluster size, which will be lower bounded by $\lfloor \epsilon^{-1} \rfloor$. By taking this modification, the authors basically trade cluster cohesiveness for larger cluster sizes. In Leordeanu and Sminchisescu (2012), an efficient algorithm has been devised to optimize the size-restricted version of (12).

6. Conclusions

We have reviewed the dominant set clustering framework and shown how the concept of dominant set can be regarded as a meaningful notion of a cluster from different perspectives. Besides their combinatorial definition, dominant sets have intuitive interpretations from different perspectives: in optimization theory they relate to local maxima of standard quadratic problems, in graph theory to maximal cliques of graphs, and in game theory to evolutionary stable strategies of clustering games. The latter perspective justifies the use of dynamics from evolutionary game theory to determine dominant sets and we have described those dynamics in some details.

The dominant sets clustering framework has different attractive features, which can be summarized as follows:

1. it makes no assumption on the underlying (individual) data representation: like, *e.g.*, spectral clustering, it does not require that the elements to be clustered are represented as points in vector space;
2. it does not require *a priori* knowledge on the number of clusters (since it extracts them sequentially);
3. it leaves clutter elements unassigned (useful, *e.g.*, in figure/ground separation or one-class clustering problems)
4. it allows extracting overlapping clusters
5. it can use asymmetric affinity matrices, which might be useful in several circumstances (Torsello et al., 2006).

We have also briefly discussed several extensions of the framework, namely the out-of-sample extension to deal with large-scale problems, hierarchical clustering, and the extension to high-order similarities. Finally, dominant sets have been applied in different fields and we have provided a glimpse of different applications, while delving more into details of the application to image and video segmentation.

As a concluding remark, the approach outlined in this work is but one example of using game-theoretic concepts to model *generic* machine learning problems (see Cesa Bianchi & Lugosi, 2006 for another such example in a totally different context), and the potential of game theory to machine learning is yet to be fully explored. Other areas where game theory could potentially offer

a fresh and powerful perspective include, *e.g.*, semi-supervised learning, multi-similarity learning, multi-task learning, learning with incomplete information, learning with context-dependent similarities. The concomitant increasing interest around the algorithmic aspects of game theory (Nisan, Roughgarden, Tardos, & Vazirani, 2007) is certainly beneficial in this respect, as it will allow useful cross-fertilization of ideas.

References

Ackerman, M., & Ben-David, S. (2008). Measures of clustering quality: A working set of axioms for clustering. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems, Vancouver, Canada: vol. 21* (pp. 121–128). New York, NY, USA: Curran Associates Inc.

Adamos, A. D., Laskaris, N. A., Kosmidisc, E. K., & Theophilidis, G. (2012). In quest of the missing neuron: Spike sorting based on dominant-sets clustering. *Computer Methods and Programs in Biomedicine, 107*(1), 28–35.

Albarelli, A., Rodolà, E., & Torsello, A. (2010). A game-theoretic approach to fine surface registration without initial motion estimation. In *IEEE conference on computer vision and pattern recognition* (pp. 430–437). San Francisco, CA, USA: IEEE.

Albarelli, A., Rodolà, E., & Torsello, A. (2012). Imposing semi-local geometric constraints for accurate correspondences selection in structure from motion: A game-theoretic perspective. *International Journal on Computer Vision, 97*(1), 36–53.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403–410.

Banerjee, J., Klink, C., Peters, E. D., Niessen, W. J., Moelker, A., & van Walsum, T. (2015). Fast and robust 3d ultrasound registration – Block and game theoretic matching. *Medical Image Analysis, 20*(1), 173–183.

Baum, L. E., & Eagon, J. A. (1967). An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *Bulletin of the American Mathematical Society, 73*, 360–363.

Benati, S., Puerto, J., & Rodríguez-Chía, A. M. (2017). Clustering data that are graph connected. *European Journal of Scientific Research, 261*(1), 45–53.

Besiris, D., Makedonas, A., Economou, G., & Fotopoulos, S. (2009). Combining graph connectivity & dominant set clustering for video summarization. *Multimedia Tools and Applications, 44*(2), 161–186.

Bomze, I. M. (1997). Evolution towards the maximum clique. *Journal of Global Optimization, 10*(2), 143–164.

Bomze, I. M. (1998). On standard quadratic problems. *Journal of Global Optimization, 13*(4), 369–387.

Bomze, I. M., Schachinger, W., & Ullrich, R. (2017). The complexity of simple models – A study of worst and typical hard cases for the standard quadratic optimization problem. http://www.optimization-online.org/DB_HTML/2016/05/5452.html, Preprint, ISOR, University of Vienna, Submitted for publication.

Broom, M. (2000). Bounds on the number of ESSs of a matrix game. *Mathematical Biosciences, 167*(2), 163–175.

Broom, M., Cannings, C., & Vickers, G. T. (1993). On the number of local maxima of a constrained quadratic form. *Proceedings: Mathematical and Physical Sciences, 443*(1919), 573–584.

Carrizosa, E., Mladenović, N., & Todosijević, R. (2013). Variable neihhborhood search for minimum sum-of-squares clustering on networks. *European Journal of Scientific Research, 230*(2), 356–363.

Cesa Bianchi, N., & Lugosi, G. (2006). *Prediction, learning, and games*. New York, NY, USA: Cambridge University Press.

Crammer, K., Talukdar, P. P., & Pereira, F. (2008). A rate-distortion one-class model and its applications to clustering. In *Proceedings of international conference on machine learning* (pp. 184–191). New York, NY, USA, Helsinki, Finland: ACM.

Dacier, M., Pham, V.-H., & Thonnard, O. (2009). The WOMBAT attack attribution method: Some results. In A. Prakash, & I. Sen Gupta (Eds.), *Information systems security: vol. 5905* (pp. 19–37). Berlin, Heidelberg: Springer. Lecture Notes in Computer Science.

De Angelis, L., & Dias, J. G. (2014). Mining categorical sequences from data using a hybrid clustering method. *European Journal of Operational Research, 234*(3), 720–730.

Dimitriadis, S. I., Laskaris, N. A., Del Rio-Portilla, Y., & Koudounis, G. C. (2009). Characterizing dynamic functional connectivity across sleep stages from EEG. *Brain Topographu, 22*, 119–133.

Dubuisson, M. P., & Jain, A. K. (1994). A modified haudroff distance for object matching. In *Proceedings of international conference on pattern recognition, Jerusalem, Israel.* (pp. 566–568).

Fowlkes, C., Belongie, S., & Malik, F. C. J. (2004). Spectral grouping using the nyström method. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26*(2), 214–225.

Frommlet, F. (2010). Tag SNP selection based on clustering according to dominant sets found using replicator dynamics. *Advances in Data Analysis and Classification, 4*(1), 65–83.

Fu, X., Teng, L., Li, Y., Chen, W., Mao, Y., Shen, I., & Xie, Y. (2005). Finding dominant sets in microarray data. *Frontiers in Bioscience, 10*, 3068–3077.

Gupta, G., & Ghosh, J. (2005). Robust one-class clustering using hybrid global and local search. In *Proceedings of international conference on machine learning* (pp. 173–280). Bonn, Germany: ACM, New York, NY, USA.

Hamid, R., Johnson, A., Batta, S., Bobick, A., Isbell, C., & Coleman, G. (2005). Detection and explanation of anomalous activities: Representing activities as bags of event n-grams. In *Proceedings of IEEE conference on computer vision and pattern recognition, San Diego, CA, USA: vol. 1* (pp. 1031–1038). IEEE.

Hamid, R., Maddi, S., Johnson, A. Y., Bobick, A. F., Essa, I. A., & Isbell, C. (2009). A novel sequence representation for unsupervised analysis of human activities. *Artificial Intelligence, 173*, 1221–1244.

Hassin, R., & Or, E. (2010). Min sum clustering with penalties. *European Journal of Operational Research, 206*(3), 547–554.

Herault, L., & Horaud, R. (1993). Figure-ground discrimination: a combinatorial optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*(9), 899–914.

Ho, J., Ming-Hsuan, Y., Jongwoo, L., Kuang-Chih, L., & Kriegman, D. (2003). Clustering apperances of objects under varying illumination conditions. In *Proceedings of IEEE conference on computer vision and pattern recognition, Madison, Wisconsin, USA: vol. 1* (pp. 11–18). IEEE.

Hofbauer, J., & Sigmund, K. (1998). *Evolutionary games and population dynamics*. Cambridge, UK: Cambridge University Press.

Hofbauer, J., & Sigmund, K. (2003). Evolutionary game dynamics. *Bulletin of the American Mathematical Society, 40*(4), 479–519.

Hung, H., & Kröse, B. (2011). *Detecting F-formations as dominant sets* (pp. 231–238). New York, NY, USA: ACM.

Jacobs, D. W., Weinshall, D., & Gdalyahu, Y. (2000). Classification with nonmetric distances. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(6), 583–600.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for data clustering*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Kleinberg, J. M. (2002). An impossibility theorem for clustering. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems, Vancouver, Canada* (pp. 463–470). Cambridge, MA, USA: MIT Press.

Kontschieder, P., Rota Bulò, S., Donoser, M., Pelillo, M., & Bischof, H. (2012). Evolutionary hough games for coherent object detection. *Computer Vision and Image Understanding, 116*, 1149–1158.

Liu, H., Latecki, L. J., & Yan, S. (2010). Robust clustering as ensembles of affinity relations. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, & A. Culotta (Eds.), *Advances in neural information processing systems, Vancouver, Canada: vol. 23* (pp. 1414–1422). USA: Curran Associates Inc.

Leordeanu, M., & Sminchisescu, C. (2012). Efficient hypergraph clustering. In N. D. Lawrence, & M. A. Girolami (Eds.), *Proceedings of international conference on artificial intelligence and statistics: vol. 22* (pp. 676–684). La Palma, Canary Islands.

Mauro, M., Riemenschneider, H., Van Gool, L., & Leonardi, R. (2013). Overlapping camera clustering through dominant sets for scalable 3D reconstruction. In T. Burghardt, D. Damen, W. Mayol-Cuevas, & M. Mirmehdi (Eds.), *Proceedings of British machine vision conference*. Bristol, UK: BMVA Press.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.

Meyer, P., & Oltenau, A.-L. (2013). Formalizing and solving the problem of clustering in MCDA. *European Journal of Operational Research, 227*(3), 494–502.

Moon, J., & Moser, L. (1965). On cliques in graphs. *Israel Journal of Mathematics, 3*, 23–28.

Motzkin, T. S., & Straus, E. G. (1965). Maxima for graphs and a new proof of a theorem of Turán. *Canadian Journal of Mathematics, 17*, 533–540.

Müller, K., Neumann, J., Grigutsch, M., von Cramon, D. Y., & Lohmann, G. (2007). Detecting groups of coherent voxels in functional MRI data using spectral analysis and replicator dynamics. *Journal of Magnetic Resonance Imaging, 26*(6), 1642–1650.

Nascimento, M. C. V., & de Carvalho, A. C. P. L. F. (2011). Spectral methods for graph clustering – A survey. *European Journal of Operational Research, 211*(2), 221–231.

Neumann, J., von Cramon, D. Y., Forstmann, B. U., Zysset, S., & Lohmann, G. (2006). The parcellation of cortical areas using replicator dynamics in fMRI. *NeuroImage, 32*(1), 208–219.

Nisan, N., Roughgarden, T., Tardos, E., & Vazirani, V. (2007). *Algorithmic game theory*. New York, NY, USA: Cambridge University Press.

Pardalos, P. M., & Phillips, A. T. (1990). A global optimization approach for solving the maximum clique problem. *International Journal of Computer Mathematics, 33*, 209–216.

Pavan, M. (2004). *A new graph-theoretic approach to clustering, with applications to computer vision*. Venice, Italy: Università Ca' Foscari Venezia. Ph.d. thesis

Pavan, M., & Pelillo, M. (2003a). Dominant sets and hierarchical clustering. In *Proceedings of international conference on computer vision, Nice, France: vol. 1* (pp. 362–369). IEEE Computer Society Press.

Pavan, M., & Pelillo, M. (2003b). A new graph-theoretic approach to clustering and segmentation. In *Proceedings of IEEE conference on computer vision and pattern recognition, Madison, Wisconsin, USA: vol. 1* (pp. 145–152). IEEE.

Pavan, M., & Pelillo, M. (2005). Efficient out-of-sample extension of dominant-set clusters. In Y. Weiss, P. Schlkopf, & J. Platt. (Eds.), *Proceedings of advances in neural information processing systems, Vancouver, Canada: vol. 17* (pp. 1057–1064).

Pavan, M., & Pelillo, M. (2007). Dominant sets and pairwise clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(1), 167–172.

Pelillo, M., & Jagota, A. (1995). Feasible and infeasible maxima in a quadratic program for maximum clique. *Journal of Artificial Neural Networks, 2*, 411–420.

Pelillo, M., & Torsello, A. (2006). Payoff-monotonic game dynamics and the maximum clique problem. *Neural Computation, 18*(5), 1215–1258.

Pouget, F., Dacier, M., Zimmerman, J., Clark, A., & Mohay, G. (2006). Internet attack knowledge discovery via clusters and cliques of attack traces. *Journal of Information Assurance and Security, 1*, 21–32.

Rota Bulò, S. (2009). *A game-theoretic framework for similarity-based data clustering*. Venice, Italy: Università Ca' Foscari Venezia Ph.d. thesis.

Rota Bulò, S., & Bomze, I. M. (2011). Infection and immunization: A new class of evolutionary game dynamics. *Games and Economic Behaviour, 71*, 193–211.

Rota Bulò, S., & Pelillo, M. (2009). A game-theoretic approach to hypergraph clustering. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, & A. Culotta (Eds.), *Proceedings of advances in neural information processing systems, Vacouver, Canada: vol. 22* (pp. 1571–1579).

Rota Bulò, S., & Pelillo, M. (2013). A game-theoretic approach to hypergraph clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(6), 1312–1327.

Rota Bulò, S., Pelillo, M., & Bomze, I. M. (2011). Graph-based quadratic optimization: A fast evolutionary approach. *Computer Vision and Image Understanding, 115*, 984–995.

Rota Bulò, S., Torsello, A., & Pelillo, M. (2009). A game-theoretic approach to partial clique enumeration. *Image and Vision Computing, 27*(7), 911–922.

Roth, V., Laub, J., Kawanabe, M., & Buhmann, J. M. (2003). Optimal cluster preserving embedding of nonmetric proximity data. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*, 1540–1551.

Sakarya, U., & Telatar, Z. (2008). Video scene detection using dominant sets. In *Proceedings of international conference on image processing, San Diego, CA, USA* (pp. 73–76). IEEE.

Santi, E., Aloise, D., & Blanchard, S. J. (2016). A model for clustering data from heterogeneous dissimilarities. *European Journal of Scientific Research, 253*(3), 659–672.

Shashua, A., & Ullman, S. (1988). Structural saliency: The detection of globally salient features using a locally connected network. In *Proceedings of international conference on computer vision, Tampa, Florida, USA*.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 888–905.

Sibson, R. (1973). Slink: An optimally efficient algorithm for the single-link cluster method. *Computer Journal, 16*(1), 30–34.

Teng, L., & Chan, L. (2008). Discovering biclusters by iteratively sorting with weighted correlation coefficient in gene expression data. *Journal of Signal Processing Systems, 50*, 267–280.

Thonnard, O., & Dacier, M. (2009). A framework for attack patterns discovery in honeynet data. *Digital Investigation, 5*, S128–S139.

Torsello, A., Pavan, M., & Pelillo, M. (2005). Spatio-temporal segmentation using sominant sets. In A. Rangarajan, B. Vemuri, & A. L. Yuille (Eds.), *Proceedings of energy minimization methods in computer vision and pattern recognition, St. Augustine, Florida, USA* (pp. 301–315). Berlin, Heidelberg: Springer-Verlag.

Torsello, A., & Pelillo, M. (2009). Hierarchical pairwise segmentation using dominant sets and anisotropic diffusion kernels. In D. Cremers, Y. Boykov, A. Blake, & F. R. Schmidt (Eds.), *Proceedings of energy minimization methods in computer vision and pattern recognition, Bonn, Germany*. Berlin, Heidelberg: Springer-Verlag.

Torsello, A., Rota Bulò, S., & Pelillo, M. (2006). Grouping with asymmetric affinities: A game-theoretic perspective. In *Proceedings of IEEE conference on computer vision and pattern recognition, New York, USA* (pp. 292–299). IEEE.

Torsello, A., Rota Bulò, S., & Pelillo, M. (2008). Beyond partitions: Allowing overlapping groups in pairwise clustering. In *Proceedings of international conference on pattern recognition, Tampa, Florida*. IEEE.

Tran, K., Gala, A., Kakadiaris, I., & Shah, S. (2014). Activity analysis in crowded environments using social cues for group discovery and human interaction modeling. *Pattern Recognition Letters, 44*(C), 49–57.

Ullrich, R. (2016). The continuous time infection–immunization dynamics. *Dynamic Games and Applications*, 1–15.

Wang, M., Ye, Z. L., Wang, Y., & Wang, S. X. (2008). Dominant sets clustering for image retrieval. *Signal Processing, 88*(11), 2843–2849.

Weibull, J. W. (1997). *Evolutionary game theory*. Cambridge, Massachusetts: MIT Press.

Williams, J. W., & Thornber, K. K. (1999). A comparison of measures for detecting natural shapes in cluttered backgrounds. *International Journal on Computer Vision, 34*(2–3), 81–96.

Wu, Q., & Hao, J.-K. (2015). A review on algorithms for maximum clique problems. *European Journal of Operational Research, 242*(3), 693–709.

Yang, X., Liu, H., & Latecki, L. J. (2012). Contour-based object detection as dominant set computation. *Pattern Recognition, 45*, 1927–1936.

Yu, S., & Shi, J. (2001). Grouping with directed relationships. In M. Figueiredo, J. Zerubia, & A. K. Jain (Eds.), *Proceedings of energy minimization methods in computer vision and pattern recognition, Sophia Antipolis, France* (pp. 283–297). Springer-Verlag, Berlin, Heidelberg.

Zadeh, R. B., & Ben-David, S. (2009). A uniqueness theorem for clustering. In *Proceedings of uncertainty in artificial intelligence, Montreal, Quebec, Canada* (pp. 639–646). Arlington, Virginia, United States: AUAI Press.

Zagouras, A., Argiriou, A. A., Flocas, H. A., Economou, G., & Fotopoulos, S. (2012). An advanced method for classifying atmospheric circulation types based on prototypes connectivity graph. *Athmospheric Research, 118*, 180–192.

Zemene, E., & Pelillo, M. (2016). Interactive image segmentation using constrained dominant sets. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Proceedings of European conference on computer vision, Amsterdam*. The Netherlands: Springer International Publishing, Cham.