# Path-Based Dominant-Set Clustering

Eyasu Zemene(✉) and Marcello Pelillo

DAIS, Università Ca' Foscari di Venezia,
via Torino 155, 30172 Venezia Mestre, Italy
eyasu.zemene@unive.it, pelillo@dais.unive.it

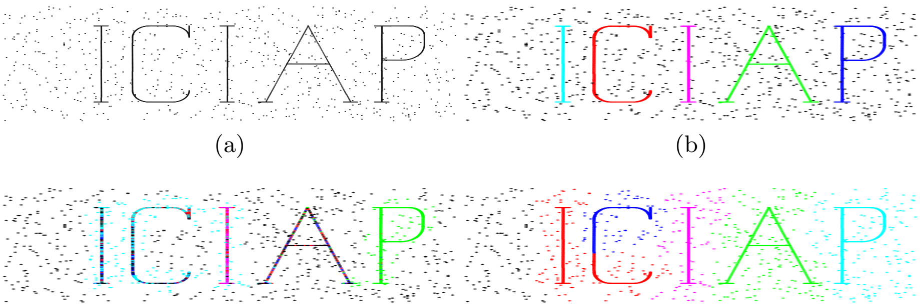
**Abstract.** Although off-the-shelf clustering algorithms, such as those based on spectral graph theory, do a pretty good job at finding clusters of arbitrary shape and structure, they are inherently unable to satisfactorily deal with situations involving the presence of cluttered backgrounds. On the other hand, dominant sets, a generalization of the notion of maximal clique to edge-weighted graphs, exhibit a complementary nature: they are remarkably effective in dealing with background noise but tend to favor compact groups. In order to take the best of the two approaches, in this paper we propose to combine path-based similarity measures, which exploit connectedness information of the elements to be clustered, with the dominant-set approach. The resulting algorithm is shown to consistently outperform standard clustering methods over a variety of datasets under severe noise conditions.

## 1 Introduction

Consider the data points shown in Figure 1(a). Despite the heavy background noise, we seem to have no difficulty in extracting a few "natural" clusters representing the letters of a familiar word. Unfortunately, standard clustering algorithms, such as those based on spectral graph theory, while doing a pretty good job in the noise-free case, perform rather poorly in such situations, as shown in Figure 1(c-d). The main reason behind this disappointing behavior is that they are typically all based on the idea of partitioning the input data, and hence the clutter points as well, into coherent classes.

In the last few years, dominant sets have emerged as a powerful alternative to spectral-based and similar methods [8], and are finding applications in a variety of different application domains such as computer vision, bioinformatics, medical image analysis, etc. Motivated by intriguing graph- and game-theoretical interpretations they try to capture the very essence of the notion of a cluster, namely their being maximally homogenous groups. By focusing on the question "what is a cluster?" dominant sets overcome some of the classical limitations of partition-based approaches such as the inability to extract overlapping clusters and the need to know the number of clusters in advance [9]. A typical problem associated to dominant sets, however, is that they tend to favor compact clusters. The problem therefore remains as to how to deal with situations involving arbitrarily-shaped clusters in a context of heavy background noise.


V. Murino and E. Puppo (Eds.): ICIAP 2015, Part I, LNCS 9279, pp. 150–160, 2015.
DOI: 10.1007/978-3-319-23231-7_14



(a) (b)

(c) (d)

**Fig. 1.** Results of extracting characters from clutter (a) Characters with uniformly distributed clutter elements which do not belong to any cluster (Original dataset to be clustered) (b) Result of our method (PBD) (c) The result of Path-based Spectral Clustering (PBS) (d) NJW's algorithms result.

In this paper we propose a simple yet effective approach to solve this problem, which is based on the idea of feeding the dominant-set algorithm with a path-based similarity measure proposed earlier in a different context [1][3][4][5]. This takes into account connectivity information of the elements being clustered, thereby transforming clusters exhibiting an elongated structure under the original similarity function into compact ones. Recently, an approach which combines path-based similarities with spectral clustering has been introduced [1]. It improves the robustness of a spectral clustering algorithm by developing robust path-based similarity based on M-estimation from robust statistics. Instead of applying the spectral analysis directly on the original similarity matrix, they first modify the similarity matrix in such a way that the connectedness information is allowed for and at the same time checking if the sample is an outlier. However, the method is robust only against small number of thinly scattered outliers and, being based on spectral partition-based methods, it cannot safely extract elements from heavy background noise. Indeed, dominant sets and spectral clustering seem to exhibit a complementary features. On the one hand, spectral-based methods do typically a good job at extracting elongated clusters but perform poorly in the presence of clutter noise, on the other hand the dominant-set algorithm prefers compact structures but is remarkably robust under heavy background noise. With our simple approach we are able to take the best of the two approaches, namely the ability to extract arbitrarily complex clusters and, at the same time, to deal with clutter noise. A similar attempt, though with different objectives, was done in [2]. Several experiments conducted over both toy and standard datasets have shown the effectiveness of the proposed approach.

## 2 Path-Based Dominant Sets

### 2.1 Dominant Set Clustering

We represent the data to be clustered as an undirected edge-weighted graph with no self-loops $G = (V, E, w)$, where $V = \{1, ..., n\}$ is the vertex set, $E \subseteq V \times V$ is the edge set, and $w : E \to R_+^*$ is the (positive) weight function. Vertices in $G$ correspond to data points, edges represent neighborhood relationships, and edge-weights reflect similarity between pairs of linked vertices. As customary, we represent the graph $G$ with the corresponding weighted adjacency (or similarity) matrix, which is the $n \times n$ nonnegative, symmetric matrix $A = (a_{ij})$ defined as $a_{ij} = w(i, j)$ if $(i, j) \in E$, and $a_{ij} = 0$ otherwise. Since in $G$ there are no self-loops, note that all entries on the main diagonal of $A$ are zero.

In an attempt to formally capture this notion, we need some notations and definitions. For a non-empty subset $S \subseteq V$, $i \in S$, and $j \notin S$, we define

$$\phi_S(i, j) = a_{ij} - \frac{1}{|S|} \sum_{k \in S} a_{ik} \ . \tag{1}$$

This quantity measures the (relative) similarity between nodes $j$ and $i$, with respect to the average similarity between node $i$ and its neighbors in $S$. Note that $\phi_S(i, j)$ can be either positive or negative. Next, to each vertex $i \in S$ we assign a weight defined (recursively) as follows:

$$w_S(i) = \begin{cases} 1, & \text{if} \quad |S| = 1, \\ \sum_{j \in S \setminus \{i\}} \phi_{S \setminus \{i\}}(j, i) w_{S \setminus \{i\}}(j), & \text{otherwise} \ . \end{cases} \tag{2}$$

Intuitively, $w_S(i)$ gives us a measure of the overall similarity between vertex $i$ and the vertices of $S \setminus \{i\}$ with respect to the overall similarity among the vertices in $S \setminus \{i\}$. Therefore, a positive $w_S(i)$ indicates that adding $i$ into its neighbors in $S$ will increase the internal coherence of the set, whereas in the presence of a negative value we expect the overall coherence to be decreased. Finally, the total weight of $S$ can be simply defined as

$$W(S) = \sum_{i \in S} w_S(i) \ . \tag{3}$$

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

a. $w_S(i) > 0$, for all $i \in S$.
b. $w_{S \bigcup i}(i) < 0$, for all $i \notin S$.

It is evident from the definition that a dominant set satisfies the two basic properties of a cluster: internal coherence and external incoherence. Condition 1 indicates that a dominant set is internally coherent, while condition 2 implies that this coherence will be destroyed by the addition of any vertex from outside. In other words, a dominant set is a maximally coherent data set.

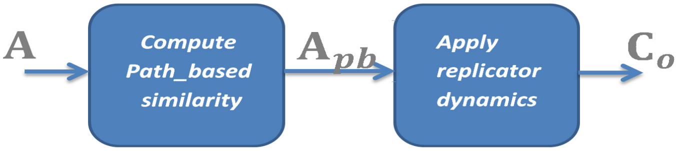



**Fig. 2.** Block diagram of the framework, where '**A**' is the original similarity, '$\mathbf{A}_{pb}$' is the path_based similarity and $\mathbf{C}_o$ is the cluster outputs

Now, consider the following linearly-constrained quadratic optimization problem:

$$\max \quad x^T A x \quad \text{s.t.} \quad x \in \Delta \tag{4}$$

where $\Delta = \{x \in R^n \ : \ \sum_i x_i = 1, \ \text{and } x_i \geq 0 \text{ for all } i = 1 \ldots n\}$ is the standard simplex of $R^n$. [8] established a connection between dominant sets and the local solutions of (4). In particular, they showed that if $S$ is a dominant set then its "weighted characteristics vector" $x^S$, which is the vector of $\Delta$ defined as

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

is a strict local solution of (4). Conversely, under mild conditions, it turns out that if $x$ is a strict local solution of program (4) then its "support" $S = \{i \in V \ : \ x_i > 0\}$ is a dominant set. By virtue of this result, we can find a dominant set by first localizing a solution of program (4) with an appropriate continuous optimization technique, and then picking up the support set of the solution found. In this sense, we indirectly perform combinatorial optimization via continuous optimization.

A simple and effective optimization algorithm to extract a dominant set is given by the so-called *replicator dynamics* developed and studied in evolutionary game theory:

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

for $i = 1, \ldots, n$. It is also possible to use a more efficient dynamics developed recently by [11] which has a computational complexity per step that grows linearly in the number of vertices. After extracting a dominant set, we remove its vertices from the graph and repeat the process until all elements are clustered. Using this "peel-off" strategy, the number of clusters is automatically determined and the resulted clusters satisfy the constraint of high intra-cluster and low inter-cluster similarity (see [14][12] for procedures to extract overlapping clusters). This makes dominant sets a flexible clustering notion, thereby making it especially attractive for the problem at hand.

## 2.2 Using Path-Based Similarity

The notion of path based technique, as shown in figure 3, is a simple but very effective way to capture elongated structures. It considers the connectedness

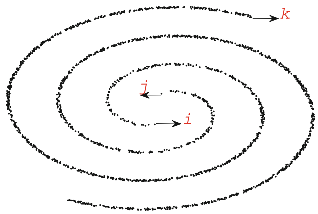



**Fig. 3.** Point 'i' and point 'k', even-though they are very far from each other, are more similar than point 'i' and point 'j' as they are connected by a path with denser region.

information to transform elongated structures into compact ones. A path in a graph is a sequence of distinct edges which connects the vertices of the graph. Let the similarity between object 'i' and object 'j' is denoted as $s_{i,j}$, and suppose that two vertices have been connected by a number of different possible paths, which forms a set denoted by $\mathcal{P}_{i,j}$. What we set out to do here, to make objects connected by a path following dense regions, is to define an effective similarity for all the possible paths. The effective similarity between object 'i' and object 'j' along the path $p \in \mathcal{P}$ is set as the minimum edge weight among all the edges contained by the path $p$. The final best similarity measure between the two objects is chosen as the maximum of all the minimum computed edge weights.

$$s^p_{i,j} = \max_{p \in \mathcal{P}} \left\{ \min_{(1 \leq h < |p|)} s_{o[h],o[h+1]} \right\} \tag{5}$$

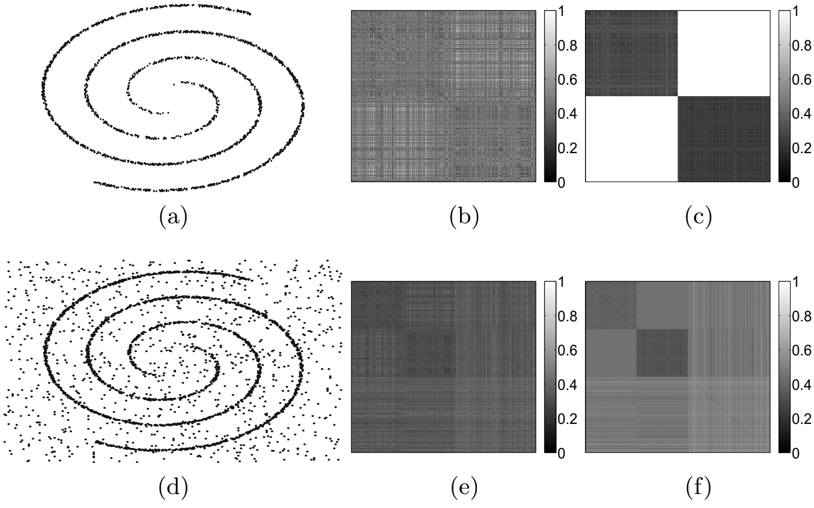



**Fig. 4.** Distance matrices of two spiral datasets with and without noise. (a) input spiral data without any noise; (b) original distance matrix of (a); (c) Path-Based dissimilarity matrix of (a); (d) Input spiral data with noise (e) original distance matrix of (d); (f) Path-Based dissimilarity matrix of (d)

Where $o[h]$ indicates the object at the $h^{th}$ position along the path $p$ and $|p|$ is the number of objects along the path.

To observe how path-based technique is suitable for dominant set clustering, the (dis)similarity measures of the different transitions, for the spiral data set, are displayed as gray scale image. As shown from the figure 4, the framework transforms the data well in such a way that the points of the spiral data set forms two block on the diagonal as a representative of the two clusters. The clusters of the data with out noise forms a clear diagonal block as shown on the first row of figure 4 which imply that any simple clustering algorithms such as K-Means can extract the clusters easily. When we come to the second case, it is clear to see, from the second row of figure 4, that the two cluster representatives do not form a very clear blocks on the diagonal which can be extracted with simple clustering algorithms. No existing methods are as accurate as our algorithm in extracting the two spirals from the clutter noise. While our algorithm uses dominant set as it easily identifies and extracts the two spirals as two dominant sets leaving the noise as non-dominant sets, other existing algorithms forces the clutter to one of the clusters.

## 3 Experiments

In this section we report a number of experimental results that are done on both toy and real datasets from UCI repository [6]. The experiments were conducted in two different ways. The first way of the experiments tests the performance of the different techniques without any clutter noise added. The second approach, which is done by adding a clutter noise samples to the datasets, is performed to see how much the algorithms are robust against background noise. In the first part of the experiment, we applied all algorithms to synthetic datasets of different manually designed structures while in the last part they are tested against real-world datasets.

Our approach was tested against three different approaches: One of the most successful spectral clustering algorithms (Ng-Jordan-Weiss (NJW) algorithm) [7], Path-based Spectral Clustering and Robust Path-based Spectral clustering (RPBS) [1] which outperformed the Path-based Clustering improving its robustness to noise. We compared against the above existing methods as they address similar problems: the problem of clustering algorithms to handle complex separable and elongated structures, and the robustness of clustering algorithms to noisy environments. All the algorithms, as opposed to our method, require the number of clusters. As of the standard clustering algorithms, all the methods also require choosing the scaling parameter $\sigma$ which has been optimally selected for all the approaches. We also assigned the correct number of clusters for those approaches which require it in advance.

### 3.1 Synthetic Data Clustering

In this part of the experiment, we applied our algorithm to eight different manually designed datasets which have been used by most of the existing algorithms

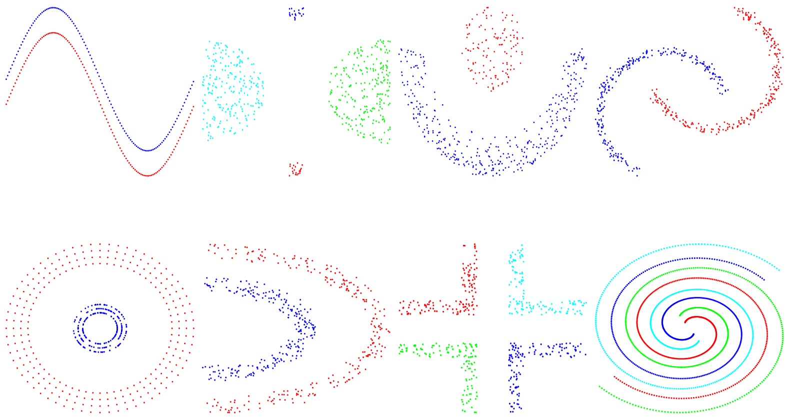


**Fig. 5.** Clustering results of NJW algorithm, Path-based Spectral Clustering, Robust Path-based Spectral Clustering, and Path-based Dominant Set clustering. All of the four algorithms perform equally in extracting all the clusters

for testing purpose. As can be seen from figure 8, the test had been done on complex separable structures. It has been shown that, classical clustering techniques such as K-means and Spectral Clustering can't solve the clustering problem in most of the data presented here [1]. However, extended version of the classical spectral clustering techniques and our proposed approach, as shown, in figure 8 are able to extracts all the clusters.

The robustness of our algorithm against noisy background is shown, using synthetic dataset, here. Similar works have been done to make clustering algorithm robust to noise [1] [15]. Our algorithm, as it uses dominant set framework,

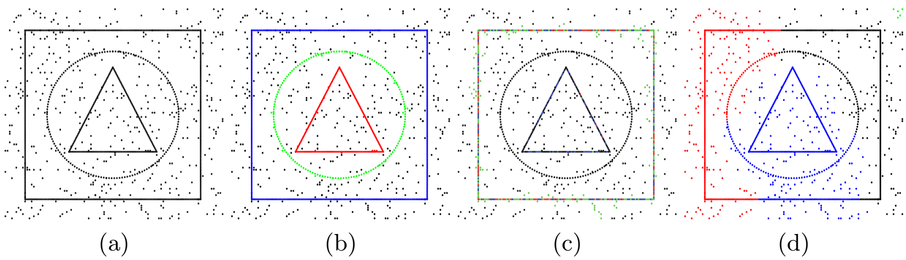



**Fig. 6.** Results on three shapes with uniformly distributed clutter elements which do not belong to any cluster (a) Original dataset to be clustered (b) The result of our method (c) The result of Path-based Spectral Clustering (d) NJW's algorithms result. Observe that only our approach is efficient in extracting all the shapes from the background noise

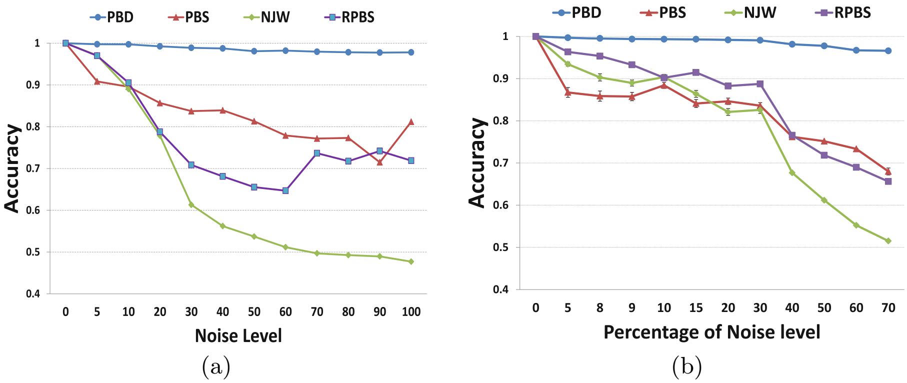



**Fig. 7.** Performance of extracting three shapes (a) and letters (b), as of figure 6 and 1, from noisy background where the noise level is increased starting from zero.

has the capability of extracting the best dominant sets leaving the clutter. However, other existing methods consider the background noise as part of the data to be partitioned.

It is clear to see that the existing approaches are vulnerable to applications where data is affected by clutter elements which do not belong to any cluster (as in figure/ground separation problems). Indeed, the only way to get rid of outliers is to group them in additional clusters. However, since outliers are not mutually similar and intuitively they do not form a cluster, the performance of all the approaches but ours drop drastically as the percentage of noise level increases.

Figure 6 shows three shapes (Triangle,Square and Circle) together with uniformly distributed background noise. As we have described above, other methods are not able to extract the right clusters, the three shapes. For the same data of the figure, we have performed an experiment by increasing the level of noise starting from zero. Zero noise implies that we have only the three shapes with

**Table 1.** Accuracy on UCI datasets (Without noise)

| Data | Instances | Attributes | PBD | PBS | NJW | RPBS |
|---|---|---|---|---|---|---|
| Ionosphere | 351 | 33 | **0.8746** | 0.8689 | 0.8718 | 0.8632 |
| Haberman | 306 | 3 | **0.7582** | 0.7451 | 0.7288 | **0.7582** |
| Spect Heart | 267 | 8 | **0.7978** | 0.7790 | 0.7940 | 0.7940 |
| Blood Trans. | 748 | 10 | 0.7620 | 0.7673 | **0.7674** | 0.7634 |
| Pima | 768 | 8 | **0.6628** | 0.6536 | 0.6615 | 0.6523 |
| Breast | 683 | 9 | **0.9678** | 0.9502 | **0.9678** | **0.9678** |
| Glass | 214 | 10 | 0.7664 | **0.7804** | 0.7523 | 0.7523 |
| Liver | 345 | 6 | **0.6145** | 0.5884 | 0.5739 | 0.5855 |

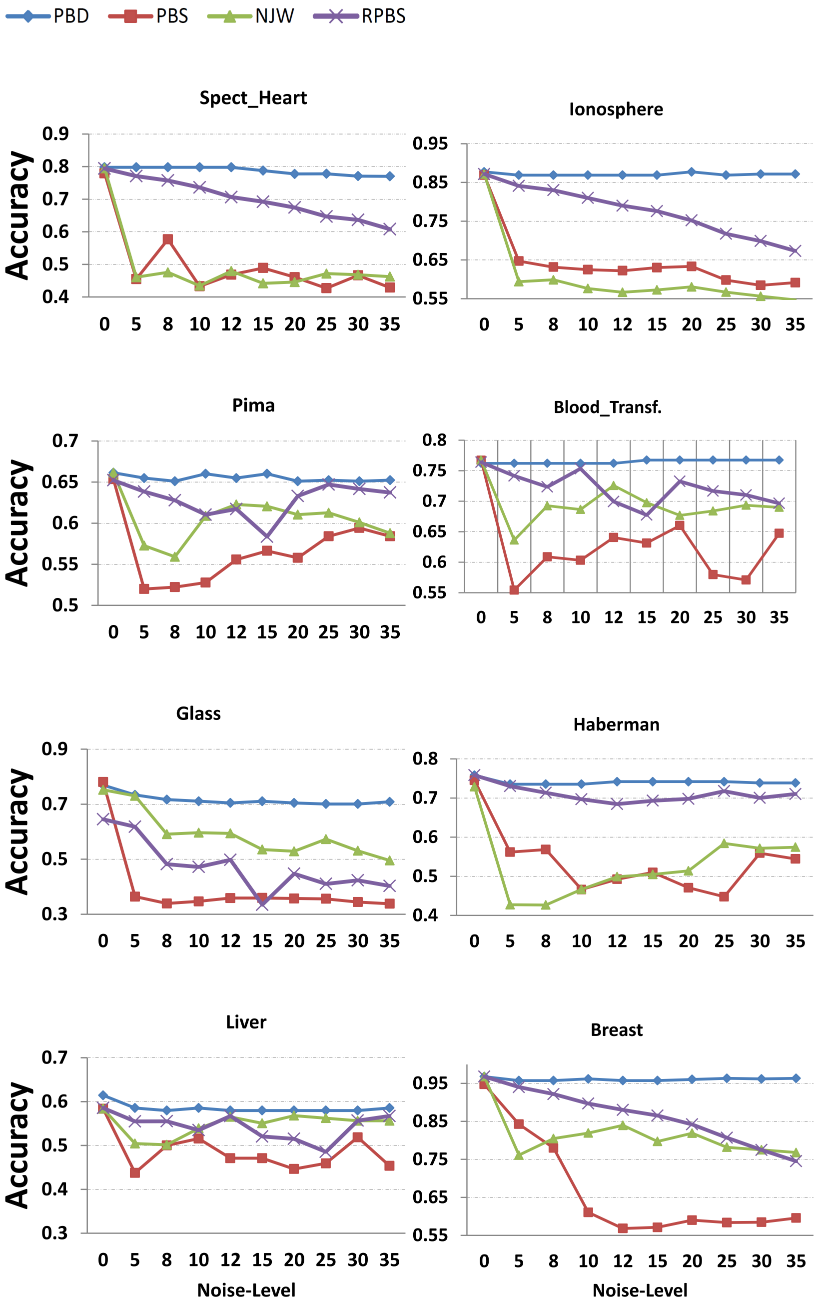



**Fig. 8.** Clustering performance of the algorithms when a clutter noise is added to the dataset. Observe that, in most of the cases, the performance of all the approaches but ours drop as the clutter noise is added.

out any clutter with which all the four clustering algorithms extract the right clusters. A noise level 'N' implies that a uniformly distributed noise of size of N% of the size of the data is added as an outlier. Figure 7 (a) shows that at the zero noise level the accuracy of all the methods is 100 %, however, the performance of all the methods but ours drop drastically as the noise level increases.

Figure 7 (b) shows a similar experiment but the noise level which was done on extracting different characters from clutter. A noise level 'n' in this case mean a uniformly distributed n*5 samples put together with the data as a clutter. The result from this experiment also confirms that our approach outperforms all the other approaches.

### 3.2 Experiments on Real-World Data

We also tested the algorithm on eight commonly used real-world datasets from UCI repository [6]. All the datasets incorporate cluster structures of complex separable, and most of them are with multiple scales. The performance of all the methods tested on the original dataset, refer table 1, is almost comparable.

An experiment has been conducted to show how much our method is robust to clutter noise added to the real-world datasets.

The experimental results, as can be referred from figure 8, consistently show that the existing approaches are vulnerable to applications where data is affected by clutter elements which do not belong to any cluster. It is easy to see, from figure 8, that the performance of all the approaches but ours drop drastically as noise is added to the datasets.

## 4 Conclusion

In this paper we have proposed a simple yet effective scheme to deal with the problem of extracting arbitrarily complex clusters under severe noise conditions. As is well known, dominant-sets clustering is remarkably good at dealing with cluttered situations but, on the other hand, it tends to favor compact structures. By feeding the algorithm with a path-based similarity measure, which takes into account connectedness information of the elements to be clustered, we have shown that the resulting algorithm is capable of consistently outperform standard approaches. Future work will focus on extending this idea to directed graphs [13] as well hypegraphs [10], and to investigate alternative path-based measures

## References

1. Chang, H., Yeung, D.: Robust path-based spectral clustering. Pattern Recognition **41**(1), 191–203 (2008)
2. Chehreghani, M.H.: Information-Theoretic Validation of Clustering Algorithms. Ph.D. thesis, ETH ZURICH (2013)

3. Fischer, B., Buhmann, J.M.: Bagging for path-based clustering. IEEE Trans. Pattern Anal. Mach. Intell. **25**(11), 1411–1415 (2003a)
4. Fischer, B., Buhmann, J.M.: Path-based clustering for grouping of smooth curves and texture segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(4), 513–518 (2003b)
5. Fischer, B., Buhmann, J.M.: Path-based clustering for grouping of smooth curves and texture segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(4), 513–518 (2003c)
6. Lichman, M.: UCI machine learning repository (2013). http://archive.ics.uci.edu/ml
7. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: Advances in Neural Information Processing Systems, pp. 849–856. MIT Press (2001)
8. Pavan, M., Pelillo, M.: Dominant sets and pairwise clustering. IEEE Trans. Pattern Anal. Machine Intell. **29**(1), 167–172 (2007)
9. Pelillo, M.: What is a cluster? perspectives from game theory. In: Proc. of the NIPS Workshop on Clustering Theory (2009)
10. Rota Bulò, S., Pelillo, M.: A game-theoretic approach to hypergraph clustering. IEEE Trans. Pattern Anal. Machine Intell. **35**(6), 1312–1327 (2013)
11. Rota Bulò, S., Pelillo, M., Bomze, I.M.: Graph-based quadratic optimization: A fast evolutionary approach. Computer Vision and Image Understanding **115**(7), 984–995 (2011)
12. Rota Bulò, S., Torsello, A., Pelillo, M.: A game-theoretic approach to partial clique enumeration. Image Vision Comput. **27**(7), 911–922 (2009)
13. Torsello, A., Rota Bulò, S., Pelillo, M.: Grouping with asymmetric affinities: a game-theoretic perspective. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 292–299 (2006)
14. Torsello, A., Rota Bulò, S., Pelillo, M.: Beyond partitions: allowing overlapping groups in pairwise clustering. In: 19th International Conference on Pattern Recognition (ICPR 2008), December 8–11, 2008, Tampa, Florida, USA, pp. 1–4 (2008)
15. Zelnik-manor, L., Perona, P.: Self-tuning spectral clustering. In: Advances in Neural Information Processing Systems 17, pp. 1601–1608. MIT Press (2004)