An approach to perform local analysis on multidimensional projection

Wilson Estécio Marcílio Júnior, Danilo Medeiros Eler, Rogério Eduardo Garcia
Universidade Estadual Paulista – UNESP
Presidente Prudente – SP, Brazil
wilson_jr@outlook.com, danilome@gmail.com, rogerio@fct.unesp.br

***Abstract*—In the context of Visualization, Multidimensional Projection techniques are employed to show similarity relations among instances of a multidimensional dataset. Distinct projection techniques use different approaches to perform the dimensionality reduction and, consequently, different metrics are employed to assess projection quality according to similarity and structures preservation. Usually, quality measures are computed from the whole projection, what can impair a specific evaluation. This work presents a novel approach to perform evaluation on multidimensional projections, in which clusters of instances are selectively evaluated and compared to the whole projection. The proposed approach has shown to be effective on evaluating projections and it offers a way to apply techniques to enhance poor projected areas.**

I. Introduction

The growth of big datasets has generated challenges to perform analysis and, consequently, the decision-making process. Therefore, researchers have expended efforts to develop new approaches to reduce the cognitive load required for analysis. Visualization techniques present many kinds of visual metaphors that aid the exploration process. For example, some of the well known techniques used to analyze high dimension datasets are the Parallel Coordinates [1], similarity trees [2], [3] and Multidimensional Projection techniques [4]–[7]. Similarity Trees and Multidimensional Projection techniques can be applied to analyze similarity among instances of a dataset, such as images, documents or feature space quality resulted from feature extraction [8].

Multidimensional projection is a sort of visualization techniques that presents the instances similarities of a dataset, mapping data from high-dimensional space ($\mathbb{R}^m$) to low-dimensional space ($\mathbb{R}^p$), and preserving both structures and relations existing in the multidimensional space. However, due to the curse of dimensionality, multidimensional projection techniques may suffer with lost of information, introduction of poor structures, as well as the problem of overlap among markers. Therefore, an important issue of the multidimensional projection techniques is the need for their evaluation. Many papers aim at presenting analysis techniques of multidimensional projection techniques to verify whether the multidimensional techniques preserve the structural and similarity relations existing in the multidimensional space. Usually, the analysis of projections is performed in a general fashion, assigning an unique measure value which represents the quality of the projection technique according to a dataset and a metric of analysis. However, by analyzing a projection by such general approach, possible anomalies may not be identified.

This paper presents a novel approach to perform evaluation on multidimensional projection techniques. The analysis of a projection is applied according to a specific quality measure on clusters of a projection – such clusters can be selected by the user or automatically identified through a clustering algorithm. The motivation is that by analyzing small portions of a projection one can selectively verify the behavior of a projection technique. Thus, it is possible to find anomalies and specific instances in which their structures and similarities do not reflect the multidimensional space. By finding these issues in the projection, techniques to improve the relations among instances [9] in a selective fashion can be applied.

The main contribution of this paper is an approach to perform an evaluation of multidimensional projections, in which existing analysis metrics are employed on clusters of data selected by a clustering algorithm or manually by the user. While evaluation of multidimensional projections are usually employed as a whole (without considering clusters of data separately), our approach allows evaluation of projections to be carried out locally. Additionally, our approach allows the detailed analysis of projections by means of clusters and how such clusters behave in relation to the projection. Similarly as Projection Inspector [10], our approach allows identification of differences among projections, however, it also can identify anomalies presented by a projection considering clusters of data and providing the idea of how such characteristic reflects on the quality of the projections as a whole.

The remaining of this paper is organized as follows. The Section II presents related works. In the Section III we present the analysis metrics used in our experiments. In the Section IV we present the proposed technique. The application of the technique is presented in the Section V, where we use some datasets for validation. The conclusion is presented in Section VI.

II. Related Works

To evaluate multidimensional projection techniques, our approach depends on selection of clusters. While some works use clusters in multidimensional projection to reveal important data attributes or to perform summarization [11]–[14], in this work groups are used as a basis to the analysis, where different evaluation techniques are applied.

The Stress [4] metric, which is widely used to evaluate multidimensional projections [15], uses the differences between the dissimilarities in the multidimensional space and the distances in the visual space. A good discussion on metrics to evaluate loss of information can be found in [16]. As Paulovich et al. [6] argument, the *Stress* values are unable to ensure the ability of a projection technique to preserve neighborhood relations since it only considers distance relations. So that, the *Neighborhood Preservation* (NP) [7] technique evaluates the ability of a projection technique at maintaining neighborhood relations existing in the multidimensional space. As a complement to the NP technique, the *Neighborhood Hit* (NH) [6] aims at verifying the class perception in a neighborhood. The *Silhouette Coefficient* technique [17] analyzes the data clusters consistency, providing information on how well an instance participates in its cluster. Additionally, in [18] it is presented an approach to compute the metrics *Stress* and *Silhouette Coefficient* more efficiently.

Tatu et al. [19] presented the *Class Density Measure* (CDM) technique, where a projection is evaluated according to its separation properties. Kaski et al. [20] defined a way to verify the reliability of a projection through two metrics called *trustworthy* and *discontinuities*. A projection is given reliable (*trustworthy*) if, for each instance, a set of k nearest neighbors in the visual space are k nearest neighbors in the original space. The discontinuity (*discontinuities*) is taken into account by the fact the some of k nearest neighbors in the original are not mapped as such in the visual space.

Motta et al. [21] presented a series of metrics based on *Extended Minimal Spanning Tree* [22]. The *Class Separation* and *Class Aggregation* measures allow comparing projections according to the ability of conveying class distribution. The *Classes Validation Separation* measure quantifies the purity of the neighborhood according to their classes, while the *Neighborhood Validation* measure seeks quantify how many neighboring instances a projection preserves without considering its classes. Finally, the *Group Validation* measure evaluates the consistency of groups formed by a projection, checking if a cluster formed in the visual space is a cluster in the multidimensional space.

III. Background

In order to demonstrate our methodology, we used four analysis metrics widely used in the literature to evaluate multidimensional projections. These metrics, described below, are suitable to perform effective analysis of multidimensional projections since they evaluate neighborhood preservation, class separation, and similarity preservation among instances.

The **Neighborhood Hit** (NH) [6] technique is used to evaluate the class perception of a projection, i.e., how well a projection can separate classes. For each instance, its k nearest neighbors are found and the ratio of the neighbors with same class as the analyzed instance is verified. Suppose $k = 5$ and the number of instances with same class is 2, then NH $= \frac{2}{5} = 0.4$.

The **Neighborhood Preservation** (NP) [7] technique evaluates a projection according to neighborhood preservation. For each instance of the multidimensional space and its correspondent instance in the projected space, it is calculated the k nearest neighbors in order to compute the preservation ratio. Suppose $k = 4$ and 3 instances were preserved as nearest neighbors, then NP $= \frac{3}{4} = 0.75$.

The **Silhouette Coefficient** (SC) [17] technique is used to interpret the consistency of clusters and it provides information on how well each instance participates in its cluster. Given an instance i, to compute its silhouette coefficient one must perform the following steps:

1) the mean of the distances a_i from i to all of other instances of the same group is calculated, providing a cohesion measure;
2) the smallest distance b_i from i to all of other instances of different groups is calculated, providing a separation measure;
3) the silhouette of i is $s_i = \frac{b_i - a_i}{max(a_i, b_i)}$.

Therefore, the Silhouette Coefficient (CS) value is given by the mean of the silhouette of each instance (Equation 1)

$$CS = \frac{1}{N}\sum_{i=1}^{N}\frac{(b_i - a_i)}{max(a_i, b_i)}. \quad (1)$$

The silhouette coefficient value varies between -1 and 1 and better cohesion and group separation is indicated by values closer to 1.

The **Stress** [4] technique attempts to measure how much information was lost during the projection by computing the differences among the dissimilarities in n-dimensional space and the distances in d-dimensional space. The lower the *Stress* value, the better the structures are consistently preserved.

Stress can be calculated using Equation 2.

$$S = \sqrt{\frac{\sum\limits_{i<j}(d_{ij} - \overline{d_{ij}})^2}{\sum\limits_{i<j} d_{ij}^2}}, \quad (2)$$

where d and $\overline{d}$ represent distance functions in multidimensional and projected space, respectively.

IV. Proposed Approach for Local Analysis

The proposed approach is based on four main steps, as shown in the *pipeline* presented in Figure 1: (1) Multidimensional projection; (2) Definition of regions; (3) Selection of analysis techniques; (4) Exploration.

After projecting a multidimensional dataset, it is necessary to define regions in which the analysis will be performed. These regions can be defined manually – user selection – or by a *clustering* algorithm, such as *Bisection K-means* [23] and *DBSCAN* [24]. Figure 2 shows an example of regions of interest defined by user selection in (a) and *Bisecting K-means* algorithm in (b). Note that, when there is no classification among the instances of the dataset, visual approaches for clusters identification [25] can be easily added to our approach.

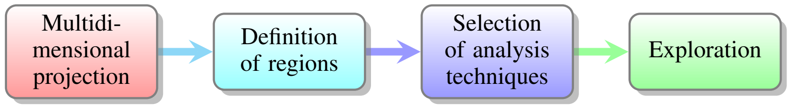
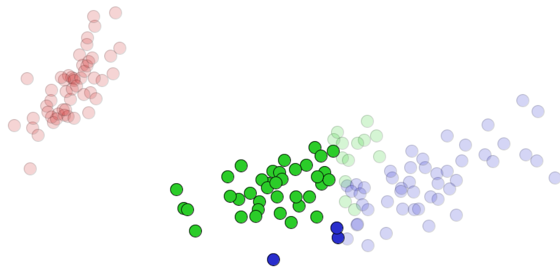
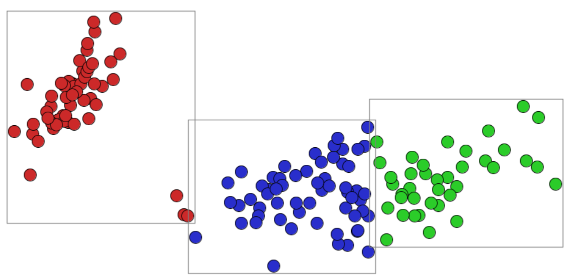


Fig. 1. *Pipeline* of the proposed approach. The first step can be performed by any multidimensional projection technique, then, the analysis regions are set automatically through a *clustering* algorithm or using manual selection by the user. After defining the analysis regions, the user needs to select the analysis techniques and define the parameters for these techniques. Finally, the representations of the selected clusters are presented, where it is possible to isolate each representation for better interaction.



(a) Cluster defined manually.



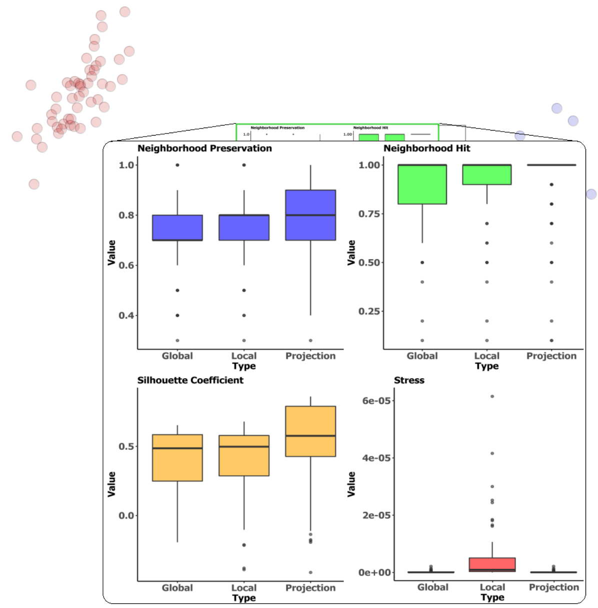
(b) Clusters defined automatically by using *Bisecting K-means* [23].

Fig. 2. Definition of the analysis region for a projection of the *iris* dataset using the IDMAP [5] technique.

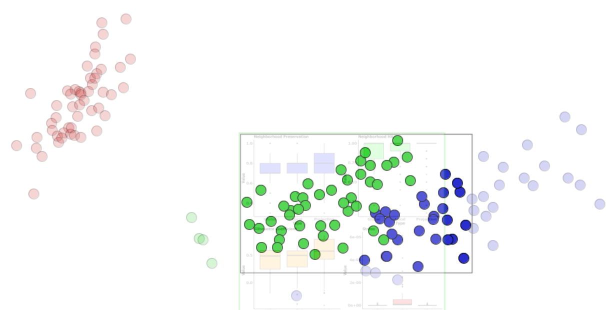
The third step of the analysis process is to choose the techniques to evaluate projection techniques. As we commented previously, in this paper we use the *Neighborhood Hit*, *Neighborhood Preservation*, *Silhouette Coefficient*, and *Stress* metrics.

Once the quality metrics are chosen, the cluster analysis is represented by *boxplots*. For that, when the user chooses a specify area of interest, the *boxplots* are rendered above the selected area, as shown in Figure 3a. By positioning the mouse cursor on the selected area, the representation is hidden in order to show the analyzed cluster, as shown in Figure 3b.

When using a clustering algorithm, the areas of the *bounding-boxes* represent the generated clusters, as shown in Figure 4a. In this way, all elements can be kept visible, as shown in Figure 4b, or only the selected one as shown in Figure 4c – by positioning the mouse over an area its representation is hidden or shown.



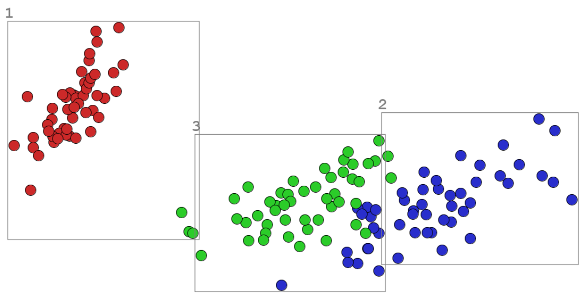
(a) Representation of analysis.



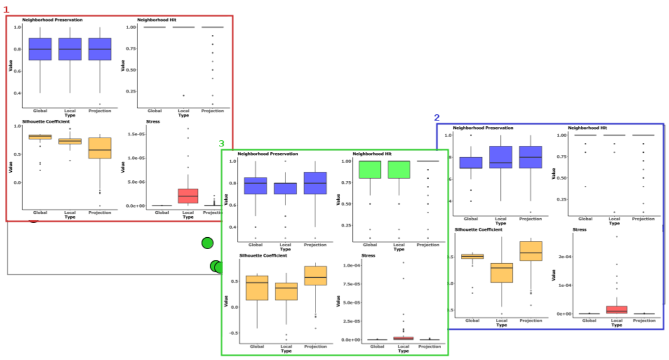
(b) Verification of the analyzed cluster.

Fig. 3. Analysis of a cluster selected by the user.

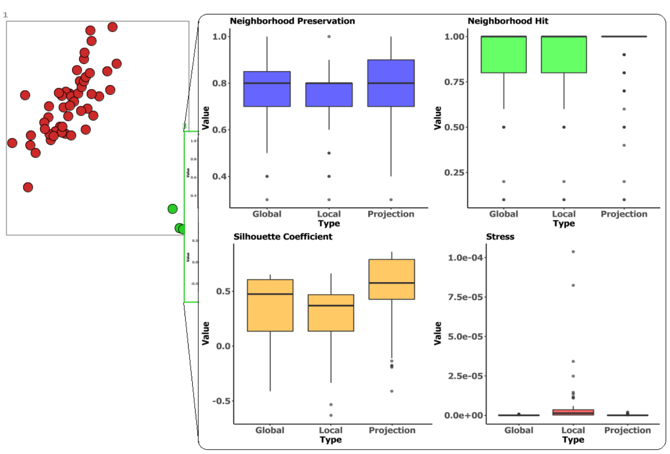
In this work, the cluster analysis can be performed in projection, local and global way. In projection analysis, the measures are computed for all projected instances and the *boxplots* present a general analysis of the whole projection. In local analysis, the quality measures are only computed for the selected instances – the cluster is isolated from the projection. In the global analysis, the measures are computed for the selected instances, but all projected instances can be used in the computation – for example, in the *Neighborhood Hit* technique the computation of a selected instance can take into account selected and unselected instances. While global analysis provide a way to verify the behavior of the



(a) Clusters returned by a clustering algorithm.



(b) All analyzes representations can be displayed at once.



(c) The representations of the analyzes can be hidden to emphasize a specific one.

Fig. 4. Automatic cluster analysis.

clusters considering the influence of remaining instances, local analysis provide a way to investigate their behavior without such influence.

V. Experiments and Results

This section presents the evaluation of IDMAP [5] and LSP [6] techniques. Although several techniques are available in the literature, such as LAMP [26] and t-SNE [27], we used the IDMAP and LSP techniques due to their different characteristics to carried out the projection process, that is, while the IDMAP technique aims at preserving similarities among instances, the LSP aims at preserving neighborhood relations.

The IDMAP [5] technique is a force-based technique which uses the concept that the relations among instances in the multidimensional space and in the projected space must be constant for each pair of points (x_i, x_j). For each instance x_i of the dataset, the vector $\vec{v}_{ij} = (x_j - x_i)$, $\forall x_j \neq x_i$ is calculated, which is used to perform a perturbation in its direction. Such perturbation depends on the distances of the projected space and the similarities of the multidimensional space. The perturbation used is given in Equation 3.

$$\Delta = \frac{\delta(x_i, x_j) - \delta_{min}}{\delta_{max} - \delta_{min}} - d(y_i, y_j), \tag{3}$$

where Δ represents the approximation between the distance in the projected space and the distance in the multidimensional space, d represents the distance measure in the projected space, δ represents the distance measure in multidimensional space, δ_{max} and δ_{min} represent the maximum and minimum distances among objects in multidimensional space, and y_i and y_j represent the projection of the points x_i and x_j.

Unlike conventional multidimensional projection techniques, which are based on preserving similarity relations, the LSP [6] technique aims at preserving neighborhood relations existing in multidimensional space. So that, control points (subset of the multidimensional points) are chosen and projected in $\mathbb{R}^p$ through a technique that preserves similarity relations. Using the neighborhood relations in $\mathbb{R}^m$ and the distance relations in $\mathbb{R}^p$, linear systems are constructed such that their solutions are used to project the remaining points in order that they reside in the convex hull of their k nearest neighbors, according to the neighborhood in $\mathbb{R}^m$. To select the control points, a clustering algorithm is executed and the representative of each cluster is chosen as control point.

IDMAP and LSP were evaluated upon three different datasets: Iris, CBR-ILP-IR, and Brodatz. As presented in Section III, the quality measures used in this work are: Neighborhood Preservation (**NP**), Neighborhood Hit (**NH**), Silhouette Coefficient (**SC**), and Stress (**S**). The Table I shows the parameters used to perform the evaluation of each dataset – the k value was employed in the quality metrics that require neighborhood and *distance* indicates the distance measure used to perform projection and analysis.

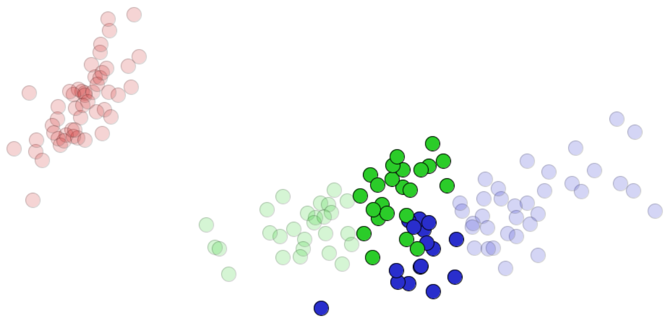
Although any possible group can be analyzed with our approach, to demonstrate its effectiveness we present the analysis only at problematic areas of the projections, such as class boundaries.

TABLE I
Parameters and approaches for component evaluation.

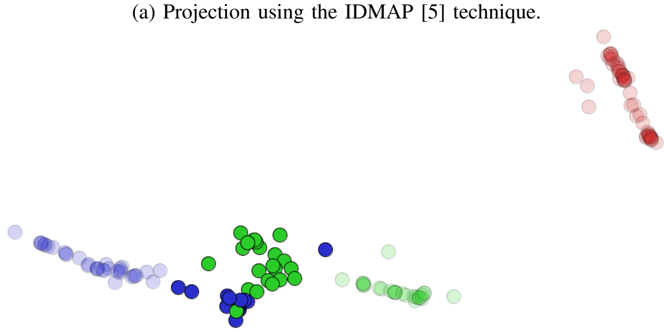
Dataset	Selection	k	Distance
iris	Manual selection	10	Euclidean
CBR-ILP-IR	Manual selection	10	*Cosine-based*
Brodatz	*clustering*	5	Euclidean

A. Iris dataset

The *iris* dataset has three classes of 50 instances, each class refers to a type of iris plant. Figure 5 shows the manual selection of instances on IDMAP and LSP projection. The selection was performed at the boundary of two clusters composed by two classes in order to compare which projection technique can present the best separation of these two classes. Figure 6 shows the *boxplots* of the selected region.



(a) Projection using the IDMAP [5] technique.

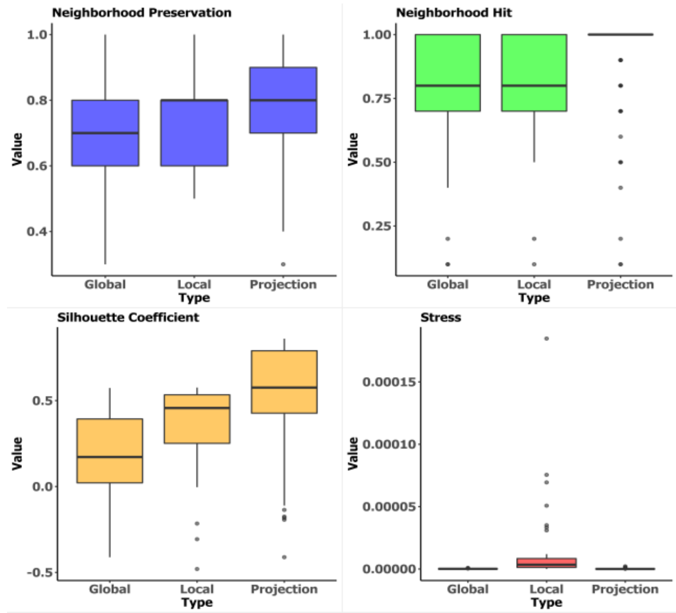


(b) Projection using the LSP [6] technique.

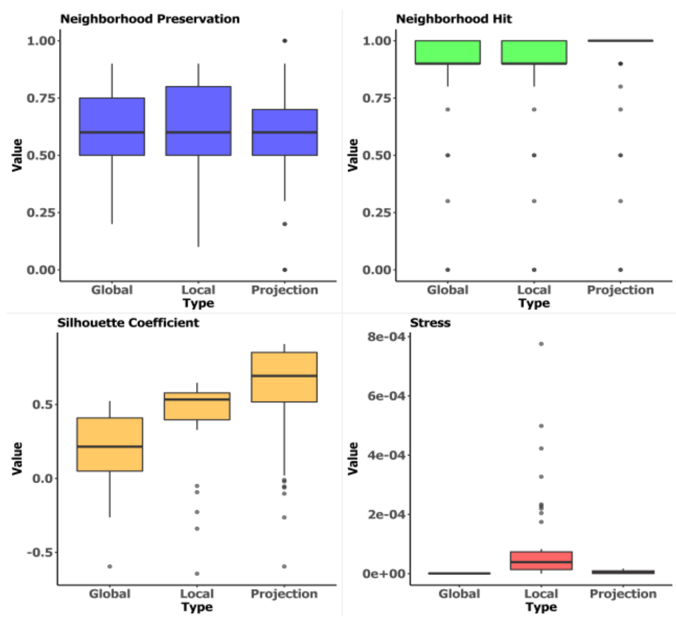
Fig. 5. Selected cluster of the *iris* dataset highlighted in both projections through coordination.

According to the **NP** metric of the selected region (local analysis), we can observe that the IDMAP technique was better than the LSP technique at maintaining neighborhood relations. However, when comparing the local and global analysis of both techniques, the LSP technique presented similar behavior to the projection analysis, which provides an indicative of standardization. For the **NH** metric, we can observe that the LSP technique presented the best distribution, where 50% of the values are above 0.9 (25% of the IDMAP values are above 0.8). This result indicates that LSP is a good technique to group instances belonging to the same class.

For the **SC** metric, the LSP technique presented a slight improvement for the projection analysis, as well as for analysis of the global behavior since it maintains the instances of the classes in a compact shape. According to the local perspective, the LSP technique presented an advantage since the selected instances presented themselves more compact than on the IDMAP technique – note the selected group of blue instances. An interesting point is the difference between the distribution of the projection analysis and global analysis for both projection techniques. This happens due to the characteristic of the **SC** technique in which the nearest are instances of



(a) Highlighting the boxplots of the IDMAP [5] technique.



(b) Highlighting the boxplots of the LSP [6] technique.

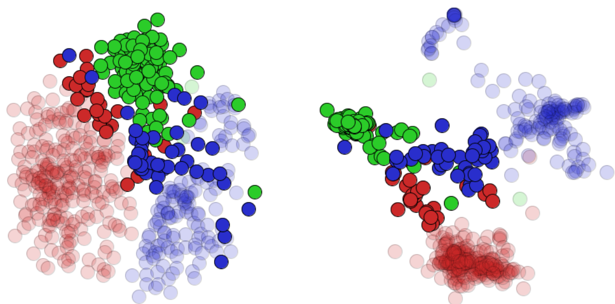
Fig. 6. Boxplots corresponding to the analysis of the *Iris* dataset.

different classes, the worst will be the performance. There is a slight advantage of the LSP technique at a "problematic area" – boundary of two clusters – for these metrics and it influences to generate a better projection.

According to the **Stress** metric, the IDMAP technique presented values with an order of magnitude less than the LSP technique because the LSP technique aims at preserving the neighborhood relations than the distance relations, however, the values of the metric **Stress** presented by LSP technique are still acceptable. We can see that through a local point of view, the variation is greater than in a global point of view since by isolating selected instances we do not perform comparisons with the remaining of the projection, and the distribution of values becomes bigger. This aspect happens with all of analyzes presented in this paper.

B. CBR-ILP-IR dataset

The *CBR-ILP-IR* dataset is composed by 574 instances representing scientific papers of three distinct fields: *Case-Base-Reasoning* (CBR), *Intuitive Logic Programming* (ILP), and *Information Retrieval* (IR). Figure 7 shows the selected instances for analysis. Again, the selection was performed at the boundary of distinct groups.



(a) Projection using the IDMAP [5] technique. (b) Projection using the LSP [6] technique.

Fig. 7. Selected cluster of the *CBR-ILP-IR* dataset highlighted in both projections through coordination.

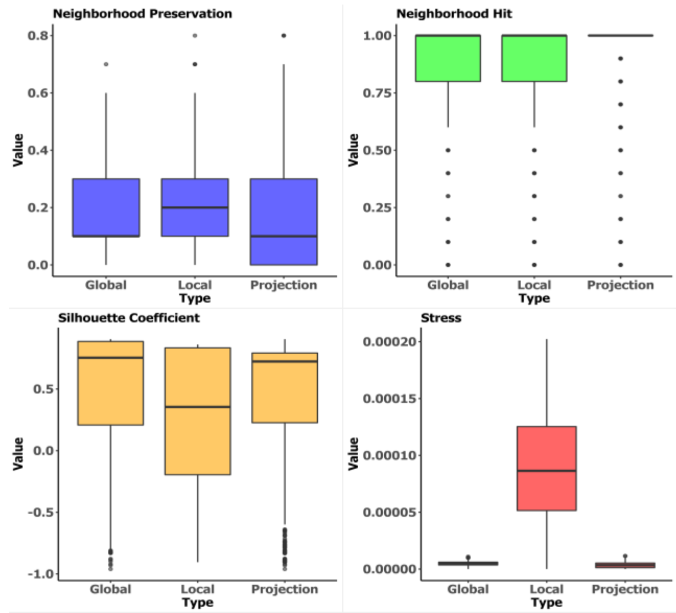
For the **NP** metric (see Figure 8), despite the multidimensional projection techniques presented same distribution, the LSP technique was better. Additionally, both techniques presented same distribution for local analysis as well. According to the global analysis, the techniques presented same distribution range, but the distribution presented by IDMAP technique is asymmetric and positive, which provides better result for the LSP technique.

The **NH** metric obtained the same distribution for both cases, i.e., in global and local analysis, so that this dataset has a characteristic of having no high dependency of other instances for analysis with $k = 10$; in other words, the remaining instances of the projection did not influence the analysis.

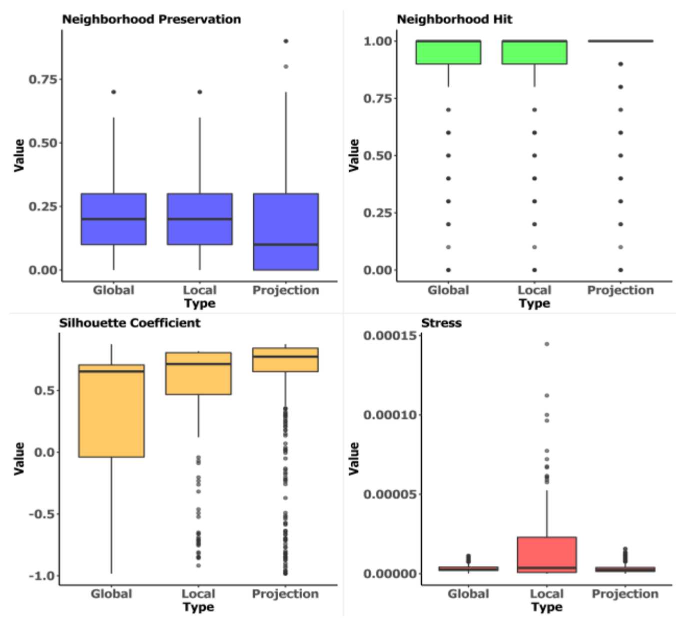
According to the **SC** metric, we can notice that the selected subset is "problematic" for both techniques since instances of different classes are too close to each other. For local analysis, the removal of the other instances in the analysis helped to improve the performance of the techniques since the instances of a specific class have less chance to be compared with a instance of a different class. Lastly, the **Stress** technique presented similar results to those discussed in the previous section.

C. Brodatz dataset

To evaluate the proposed approach through clustering algorithms and to show analysis of well defined clusters, we used the *Brodatz* dataset, which contains seven classes with 10 instances of texture images from Brodatz [28]. The feature



(a) Highlighting the boxplots of the IDMAP [5] technique.



(b) Highlighting the boxplots of the LSP [6] technique.

Fig. 8. Boxplots corresponding to the analysis of the *CBR-ILP-IR* dataset.

space is composed by Gabor filter descriptors [29]. For this dataset we only used the IDMAP [5] technique to perform the analysis. Figure 9 presents the *clusters* found by using the DBSCAN algorithm [24]. Note that *Cluster 3* is composed by two classes of textures.

In order to apply the **SC** metric a subset must have at least two classes so that cluster 3 is the only cluster to which we calculate the Silhouette Coefficient. In this section we only provide three representative analyzes because we observed that the clusters presented similar results.

1) Cluster 1: Figure 10 shows the result for cluster 1.

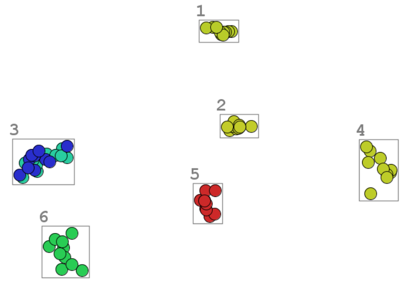


Fig. 9. Clusters found by using the *DBSCAN* [24] technique.

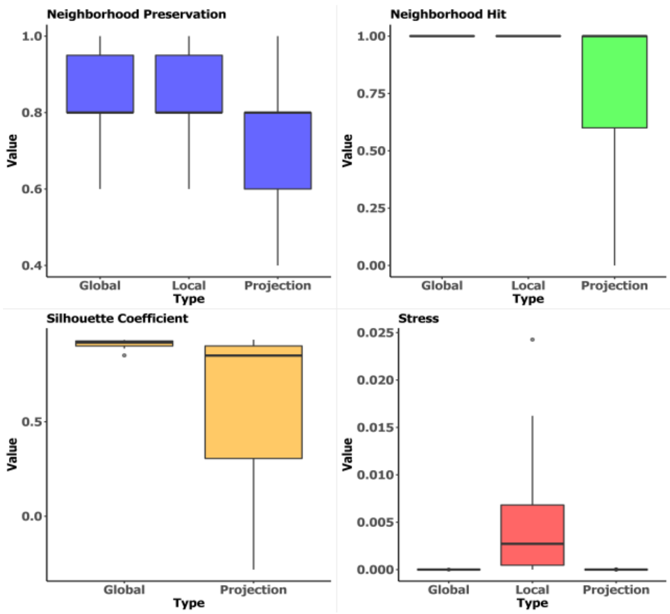


Fig. 10. Boxplots for Cluster 1.

Analyzing the **NP** metric, notice that the local and global variation presented values greater than or equal to the variation of projection, so that cluster 1 contributes to improve the quality of the projection technique according to **NP** metric; even though the distributions of local and global analysis are closer to 0.8, as shown in the *boxplots* of Figure 10.

For the **NH** metric, since the cluster is relatively distant from the other instances and there is only one class, the global and local analysis presented good performance. According to the **SC** metric, it is possible to realize that again the cluster separation influenced on the result of local analysis and its distribution values are higher than the projection distribution values. Finally, as we commented previously, the **Stress** metric presented similar results for all of analysis, where the local analysis has distribution whose values vary in a range much greater than in the global and projection analysis.

2) Cluster 3: Figure 11 shows the result for cluster 3.

Cluster 3 is responsible for decreasing the performance of the projection technique according to the **NH** and **SC** metrics, since it has two classes without effective separation, as shown in Figure 9. The **NP** metric also had lower results than the

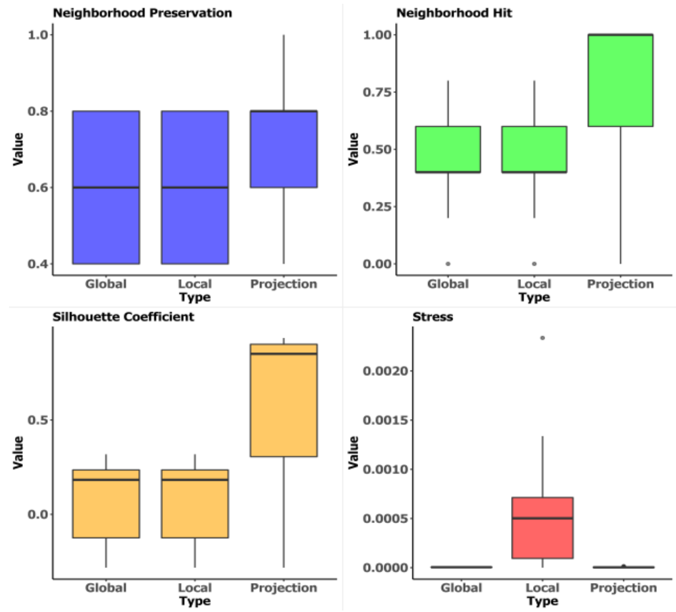


Fig. 11. Boxplots for Cluster 3.

projection since the values of the projection are closer to 0.8. According to the **Stress** metric, the result can be analyzed in the same way as in cluster 1.

3) Cluster 4: Figure 12 shows the result for cluster 4.

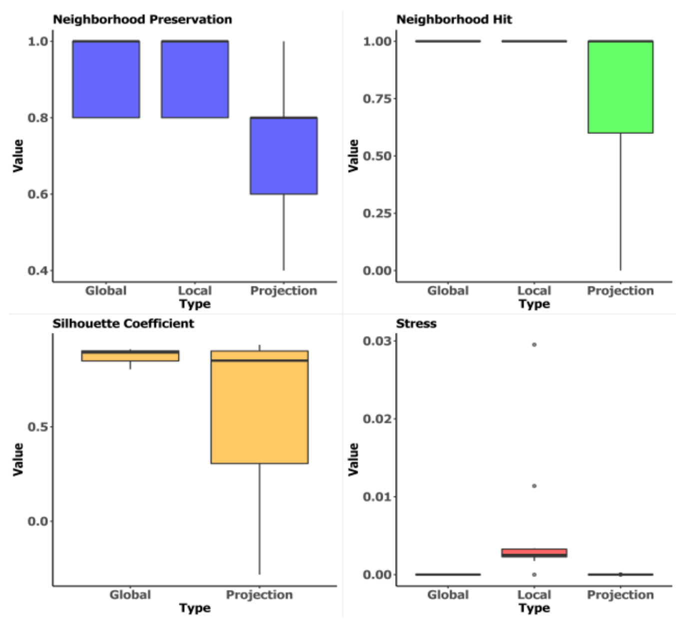


Fig. 12. Boxplots for Cluster 4.

Cluster 4 differs from cluster 1 mainly on the **NP** metric. We can see that the high overlap rate impaired the **NP** evaluation on cluster 1 and since cluster 4 has instances more scattered, the distribution values according to **NP** metric are concentrated closer to 1. Cluster 4 presented similar results to the cluster 1 according to the other metrics.

VI. Conclusion

Usually, analysis of multidimensional projections has been done in a general way, in which the quality of a projection is defined by an single value according to an evaluation metric. Such approaches do not consider possible anomalies in a projection. In other words, poorly projected subsets can be hidden because of the general characteristic of these analysis metrics.

In this paper, we introduce a novel approach to perform analysis of multidimensional projections. Our approach aims at helping in the detailed analysis of multidimensional projection techniques to easily identify anomalies in clusters of instances. For that, we presented a way to assess clusters of a projection in a local and global fashion using boxplots to visualize their quality according to different evaluation metrics. In addition, by using coordination we are able to compare various multidimensional projection techniques considering the same subset of instances, so that the comparison of techniques is facilitated.

Our approach has proven capable of helping identify areas that contribute to the general quality of multidimensional projections, as well as helping identify areas that decrease the general quality of projections. The identification of regions that decrease the quality of a projection can be the first step to apply techniques that improve the quality of a projection. So that, as a potential future work, techniques to improve the similarity and structures relations can be applied on regions that present lower quality than the projection itself, in order to investigate the changes imposed to the analysis metrics.

Acknowledgment

This paper is sponsored by FAPESP (São Paulo Research Foundation), grants #2015/18238-9 and #2016/11707-6.

References

[1] A. Inselberg and B. Dimsdale, "Parallel coordinates: A tool for visualizing multidimensional geometry," *IEEE Visualization*, vol. 1, pp. 361–378, 1990.

[2] G. P. Telles, F. V. Paulovich, R. Minghim, and A. M. Cuadros, "Point placement by phylogenetic trees and its application to visual analysis of document collections," *IEEE Symposium on Visual Analytics Science and Technology*, vol. 00, pp. 99–106, 2007.

[3] D. M. Eler, M. Y. Nakazaki, F. V. Paulovich, D. P. Santos, G. F. Andery, M. C. F. Oliveira, J. Batista Neto, and R. Minghim, "Visual analysis of image collections," *The Visual Computer*, vol. 25, no. 10, pp. 923–937, 2009.

[4] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, pp. 115–129, 1964.

[5] R. Minghim, F. V. Paulovich, and A. A. Lopes, "Content-based text mapping using multidimensional projections for exploration of document collections," *SPIE Proceedings: Visualization and Data Analysis*, vol. 6060, 2006.

[6] F. V. Paulovich, L. G. Nonato, M. Rosane, and H. Levkowitz, "Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping," *IEEE Transactions on Visulization and Computer Graphics*, vol. 3, pp. 564–575, 2008.

[7] F. V. Paulovich and R. Minghim, "Hipp: A novel hierarchical point placement strategy and its application to the exploration of document collections," *IEEE Transactions on Visulization and Computer Graphics*, vol. 14, no. 6, pp. 1229–1236, 2008.

[8] P. A. Macanhã, D. M. Eler, R. E. Garcia, and W. E. M. Junior, *Handwritten Feature Descriptor Methods Applied to Fruit Classification*. Cham: Springer International Publishing, 2018, pp. 699–705. [Online]. Available: https://doi.org/10.1007/978-3-319-54978-1_87

[9] G. M. H. Mamani, F. M. Fatore, L. G. Nonato, and F. V. Paulovich, "User-driven feature space transformation," *Proceedings of the 15th Eurographics Conference on visualization*, pp. 291–299, 2013.

[10] P. Pagliosa, F. V. Paulovich, R. Minghim, H. Levkowitz, and L. G. Nonato, "Projection inspector: Assessment and synthesis of multidimensional projection," *Neurocomputing*, pp. 599–610, 2015.

[11] E. Gomez-Nieto, F. S. Roman, P. Pagliosa, W. Casaca, E. Helou, M. Oliveira, and L. Nonato, "Similarity preserving snippet-based visualization of web search results," *TVCG*, vol. 20, no. 3, pp. 457–470, 2014.

[12] P. Joia, F. Petronetto, and L. Nonato, "Uncovering representative groups in multidimensional projections," *CGF*, vol. 34, no. 3, pp. 281–290, 2015.

[13] J. Stahnke, M. Drk, B. Mller, and A. Thom, "Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions," *TVCG*, vol. 22, no. 1, pp. 629–638, 2016.

[14] L. Pagliosa, P. Pagliosa, and L. G. Nonato, "Understanding attribute variability in multidimensional projections," *SIBGRAPI*, 2016.

[15] A. Morrison and M. Chalmers, "A pivot-based routine for improved parent-finding in hybrid mds," *Information Visualization*, vol. 3, pp. 109–122, 2004.

[16] A. Gracia, S. González, S. Robles, and E. Menasalvas, "A methodology to compare dimensionality reduction algorithms in terms of loss of quality," *Information Sciences*, pp. 1–27, 2014.

[17] L. Kaufman and P. J. Rousseuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Principles and Practice: Wiley-Interscience, 2005.

[18] D. M. Eler, J. B. M. Teixeira, P. A. Macanha, and R. E. Garcia, "Simplified stress and simplified silhouette coefficient to a faster quality evaluation of multidimensional projection techniques and feature spaces," *19th International Conference on Information Visualisation (iV)*, pp. 133–139, 2015.

[19] A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor, and D. Keim, "Combining automated analysis and visualization techniques for effective exploration of high-dimensional data," *Visual Analytics Science and Technology*, pp. 59–66, 2009.

[20] S. Kaski, J. Nikkilä, M. Oja, J. Venna, P. Törönen, and E. Castrén, "Trustworthiness and metrics in visualizing similarity of gene expression," *BMC Bioinformatics*, vol. 4, no. 1, pp. 48–61, 2003.

[21] R. Motta, R. Minghim, A. de Andrade Lopes, and M. C. F. de Oliveira, "Graph-based measures to assist user assessment of multidimensional projections," *Neurocomputing*, vol. 150, pp. 583–598, 2015.

[22] R. Motta, B. M. Nogueira, A. M. Jorge, A. de Andrade Lopes, S. O. Rezende, and M. C. F. de Oliveira, "Comparing relational and non-relational algorithms for clustering propositional data," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing, ACM*, 2013, pp. 150–155.

[23] M. Steinbach, G. Karypis, and V. Kumar, "A comparison of document clustering techniques," *KDD Workshop on Text Mining*, vol. 400, pp. 525–526, 2000.

[24] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996.

[25] L. F. Silva and D. M. Eler, *Visual Approach to Boundary Detection of Clusters Projected in 2D Space*. Cham: Springer International Publishing, 2018, pp. 849–854. [Online]. Available: https://doi.org/10.1007/978-3-319-54978-1_105

[26] P. Joia, F. V. Paulovich, D. Coimbra, J. A. Cuminato, and L. G. Nonato, "Local affine multidimensional projection," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2563–2571, 2011.

[27] L. J. P. van der Maaten and G. E. Hinton, "Visualizing high-dimensional data using t-sne," *Journal of Machine Learning Research*, vol. 9, pp. 2579—2605, 2008.

[28] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*. New York: Dover Publications, 1966.

[29] F. Bianconi and A. Fernández, "Evaluation of the effects of gabor filter parameters on texture classification," *Pattern Recognition*, vol. 40, no. 12, pp. 3325–3335, 2007.