Exploring Multidimensional Projections Through Explanatory Maps

Rodrigues Oliveira da Silva, Renato; Rauber, Paulo Eduardo; Messias Martins, Rafael; Minghim, Rosane; Telea, Alexandru

Publication date:
2016

Exploring Multidimensional Projections Through Explanatory Maps

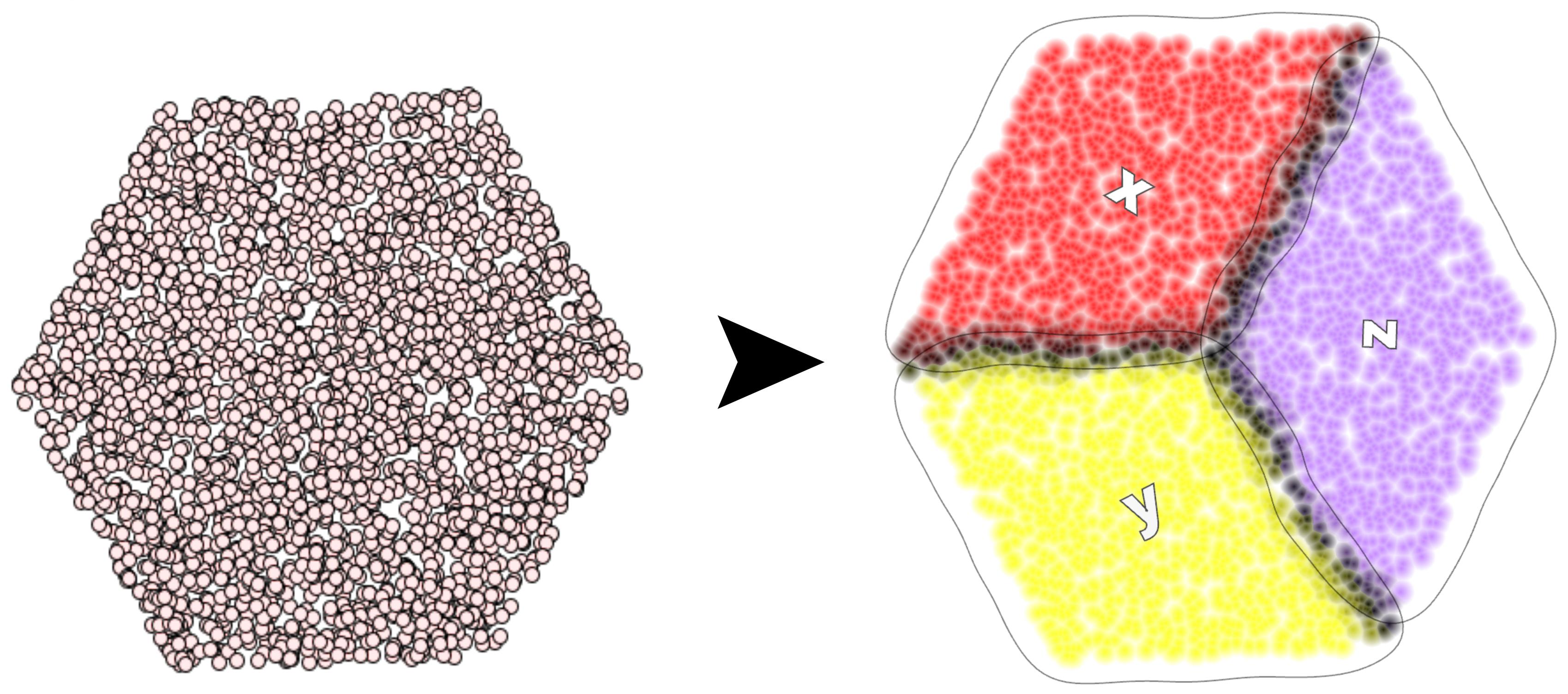
Renato R. O. da Silva[1,2], Paulo E. Rauber[1], Rafael M. Martins[1,2]
Supervisors: Alexandru C. Telea[1], Rosane Minghim[2]
1 - Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen
2 - Institute of Mathematical and Computer Sciences, University of São Paulo

Motivation

Multidimensional Projections (MPs) are popular tools used to support the analysis of multidimensional data. MPs can project data to a low dimensional space, and are typically visualized as 2D scatterplots where similar elements are conveniently positioned in close neighborhoods. However such visualizations tell us which points are similar, but not why.
Our aim is, thus, to enrich 2D MP scatterplots with explanatory information telling users which key dimensions make closely-projected points similar, and their values.

Projection Explanatory Maps

To explain a projection we identify which dimensions are more important to define similarity among close points. We first assign a **rank** to each of the dimensions for each projected point, based on increasing order of data variance in that dimension over that **neighborhood**.
We also define the **ranking confidence** for each point based on how mixed are the top-ranked dimensions on their neighborhood points. We display the top ranks and their confidences over the projection using a dense map technique based on nearest-neighbor (Voronoi) interpolation, with top-ranks encoded by a categorical colormap and confidences by brightness.



(a) Projection scatterplot (b) Explanatory map

Figure 1 - Explanation of points sampled from three cube faces. Points on the same cube-face have the same value for one dimension d, so they have assigned the same color. Color brightness indicate the confidence ranking, from high (bright) to low (dark).

Outlines detection and labeling

To facilitate the interpretation of our maps, we automatically detect same colored regions outlines and fill these with dimensions labels.
The outlines detection step involves computing the convex hull of these regions, sample the hulls, and iteratively shrink them moving the sampled points with a small step along the contour's inward normal.
Then fill outlines with dimension names labels, by creating a grid over each outline. Labels are iteratively moved to grid cells inside the outlines boundaries, until a valid position is found. We position labels on angles according to the outlines shapes, or horizontally. The size of each label reflects its weight importance on the dimension ranking, i.e., big labels indicate dimensions which contribute more to neighborhoods similarities.

Top Ranked Dimensions

Using one top-ranked dimension per point, the explanatory map defines clusters, grouping elements whose similarities are best explained by the same dimension. Dimensions names are positioned on the clusters outlines. We use a colorbar to inform the distribution values for those dimensions on demand.

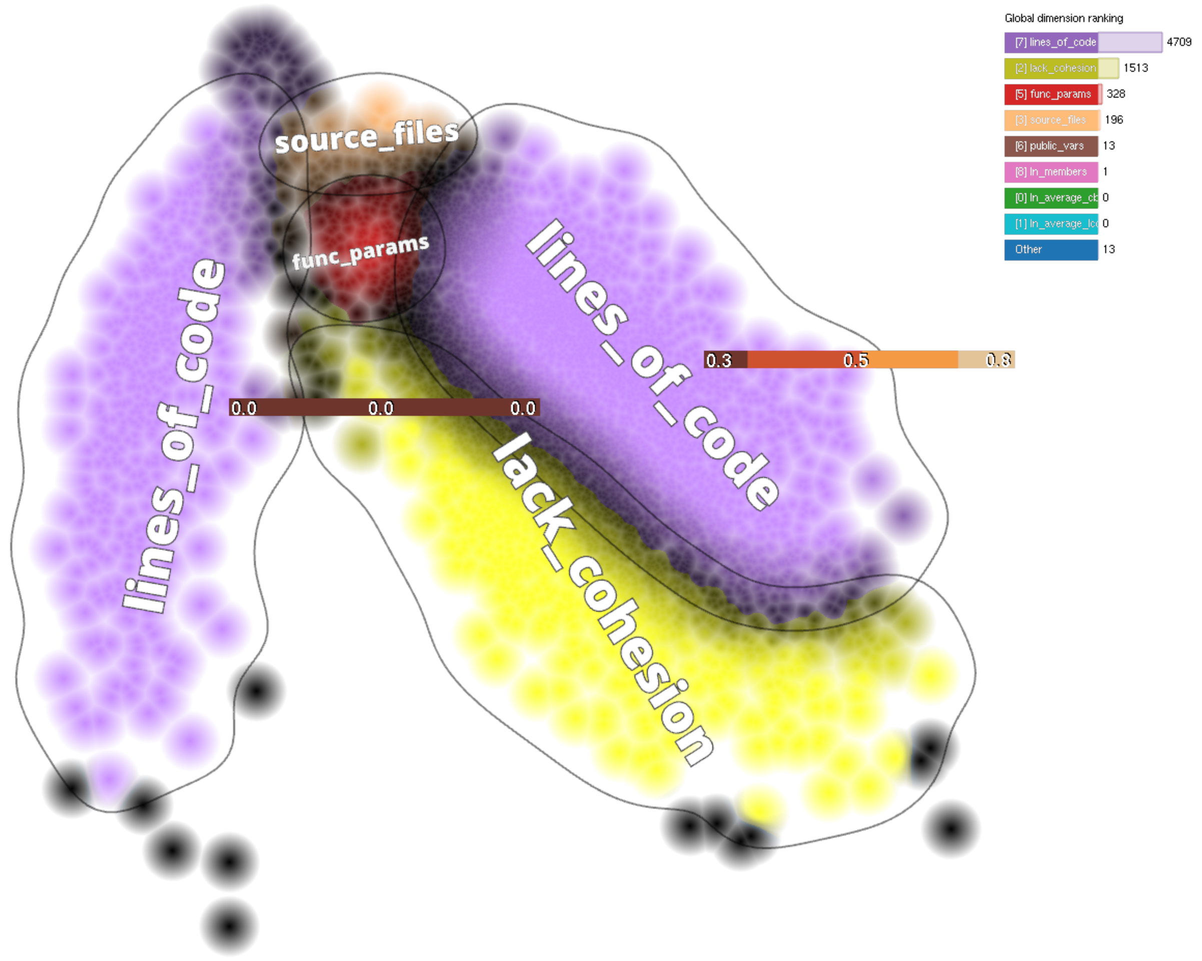


Figure 2 - Top Ranked Explanatory map of projected points from a Software quality metrics dataset.

Top Ranked Dimensions Sets

We can use more than one dimension per point to have a more detailed explanation. By thesholding the ranking of every point, we create sets of top ranked dimensions. Next, we can apply the same strategy as before, defining and explaining clusters of sets containing the same dimensions.

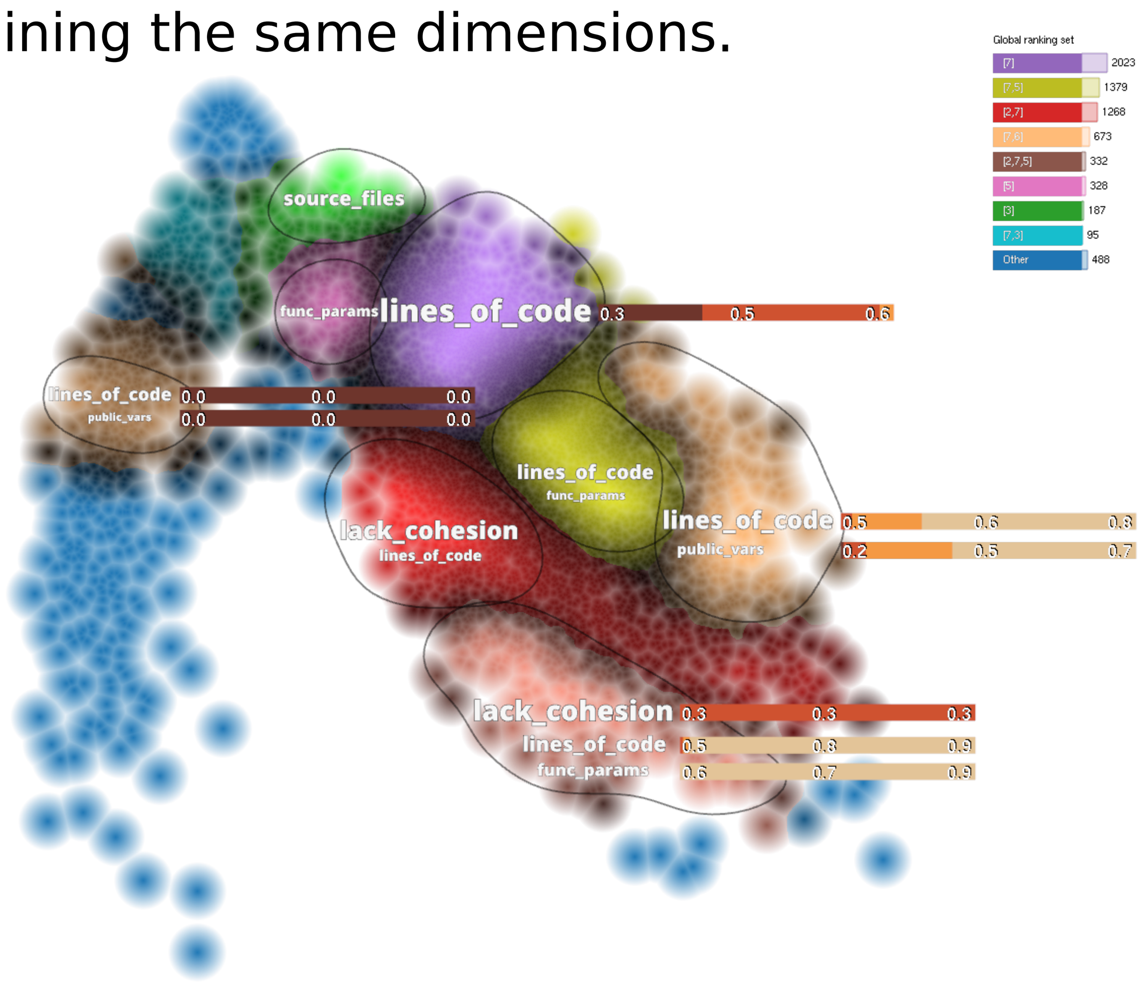


Figure 3 - Dimensions Sets Explanatory map of the Software quality metrics projection.

Future Work

Future work targets better visual encodings to explain domain specific data, such as texts and time series. We also want to develop new strategies to analyze time-dependent projections.

Contact

Renato Rodrigues Oliveira da Silva
Johann Bernoulli Institute for Mathematics and Computer Science
University of Groningen
Nijenborgh 9, 9747 AG Groningen
e-mail: rros@usp.br

This work is supported by FAPESP
Grant 2011/18838-5