Vision, Modeling, and Visualization (2021)
B. Andres, M. Campen, and M. Sedlmair (Eds.)

LayoutExOmizer Interactive Exploration and Optimization of 2D Data Layouts - Supplemental Material

Philipp Schader[1,2], Raphael Beckmann[3,5], Lukas Graner[6], and Jürgen Bernard[3,4]

[1]TU Darmstadt & [2]DKFZ Heidelberg, Germany
[3]University of Zurich, Switzerland & [4]Digital Society Initiative, Zurich, Switzerland
[5]ETH Zurich, Switzerland
[6]Fraunhofer SIT, Germany

Supplemental material document for the paper *LayoutExOmizer: Interactive Exploration and Optimization of 2D Data Layouts*. We provide more context and details about the optimization of 2D data layouts in Section 1, as well as about the two applied sets of layout quality measures in Section 2. Furthermore, we show more figures for the better understanding of LayoutExOmizer and its interface as used in the two usage scenarios.

1. Optimization of 2D Layouts

Plotting data points in 2D is a common approach to understand the interrelation between the two dimensions. However depending on the characteristics of the data, overplotting can occur and individual data points can obscure others. The situation gets even more complicated if the printed shape of the data points is larger than a single pixel. In such situations exists a trade off between displaying the true position of the data point and introducing a displacement so that overplotting is minimized. A common approach to mitigate this dilemma is the interpretation of a 2D layout as mass-spring system and the introduction of forces representing relevant aspects of the layout. By weighting the strength of the individual forces of the simulation, different aspects of a layout can be emphasised.

The LayoutAnalyzer † is a tool allowing the optimization of 2D layouts via the mass-spring approach. It is the layout optimization approach we choose for the LayoutExOmizer. It introduces 4 distinct forces to represent different aspects of a layout: The **Pairwise Repulsion Force** introduces pairwise repulsion between data points unaffected of the overlap of their shapes. The **Target Position Force** attracts data point to their true position. The **Shape Bounds Repulsion Force** introduces pairwise repulsion between data points proportional to the overlapping area of their shapes and can help to reduce overplotting. The **Border Repulsion Force** introduces a force towards the center of the data point cloud proportional to the distance of a data point to the bounding box of the data point cloud. This leads to more compact shapes. The weighting of these forces can be adjusted interactively allowing direct feedback for the user. The simulation of the mass-spring model is calculated using the Euler method. To define an end state of the simulation, which represents the optimized layout, a notion of stable and unstable layouts is used. In a stable layout the length of total force per data point is below a given threshold. This notion is necessary when sampling the parameter space because especially with extreme parameter values the simulation is not guaranteed to converge in an equilibrium and thus never reaches a stable state. With the notion of stable layouts we can terminate and discard layouts which are not stable after a given amount of steps. Because a fast convergence towards a stable state is desirable for interactivity we introduced a decreasing step size of the Euler method resulting in a damping of overall movement in the layout optimization process over time.

2. Use of Layout Quality Measures

LayoutExOmizer helps users to find meaningful layout optimizations given a 2D dataset. In order to make findings reproducible and explainable, it is important to quantify the characteristics of the outputs in a consistent manner. The related literature offers a wide range of possible measures, and one of the most popular ones among them is the set of *Scagnostics* measures proposed by Wilkinson et al. [WAG06]. With its purpose to measure visual properties in scatterplots, the nine Scagnostics measures are well suited for our approach and are used as a default measure set. Even though many extensions exist [DW14; MTL18; WWL*20], we keep the original ones because they are well established and frequently used in the visualization community. In our approach, we use the Java implementation published by the creators of the Scagnostics measures ‡. To achieve generalizability, LayoutExOmizer provides a software interface for measures in the notion of a to-double-function. We briefly describe the measures of both measure sets as shown in the use cases in detail:

Scagnostics Measures There are nine Scagnostics measures capturing different properties in scatterplots: The **Outlying** measures the proportions of the total length of long edges in the minimal spanning tree (MST) of the Delaunay triangulation of the input

† https://github.com/javagl/LayoutAnalyzer

‡ http://www.unige.ch/ses/sococ/cl/bib/edasoft/scagnostics.html

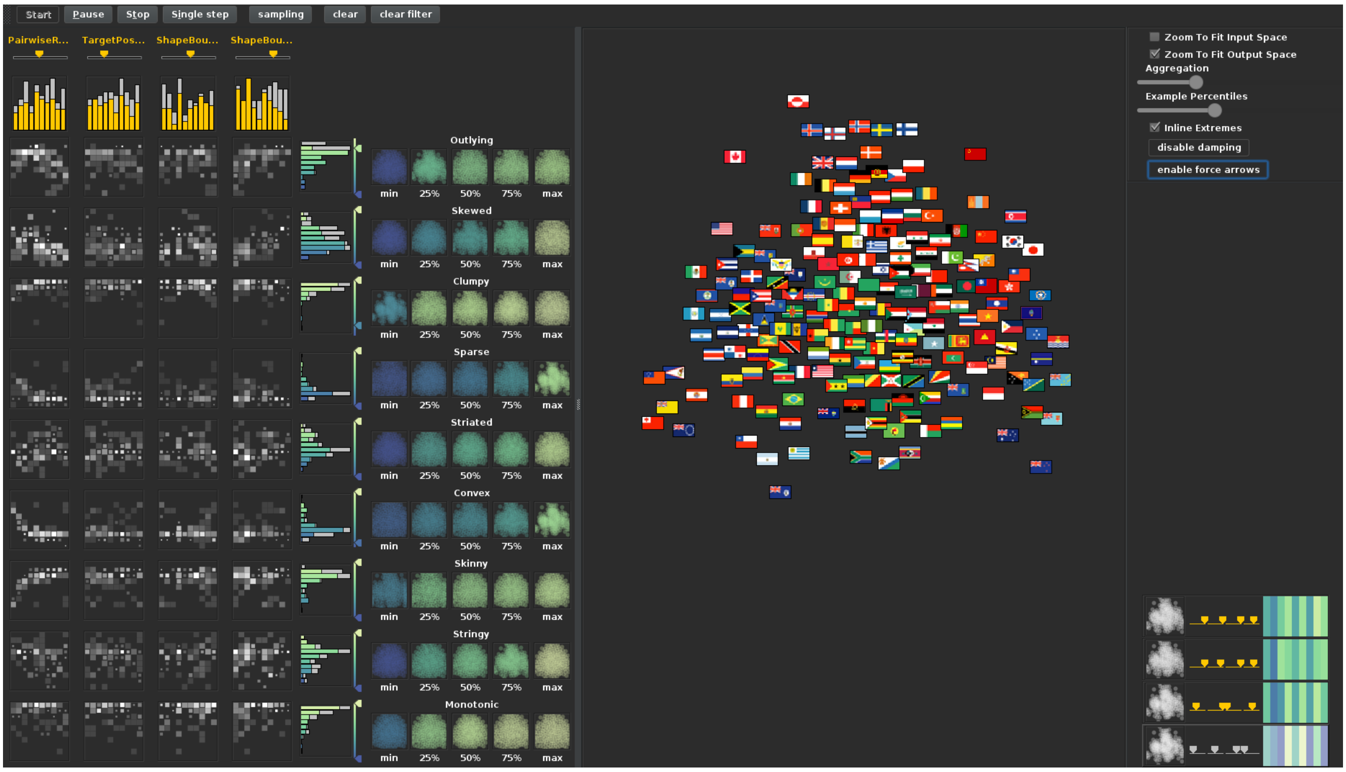

Figure 1: *System overview for the countries' capitals dataset with Flags used as a visual representation. The four forces of the layout (left) cross-cut the nine Scagnostics quality measures. In contrast to figures in the main manuscript, we represent the enlarged layout at the center of the system without forces (disabling forces is a user parameter).*

data compared to the total length of all edges. It is small if there are less long edges and therefore few outlying data point in the layout. The **Skewed** measure calculates the skewness of the distribution of the edge lengths of the MST. A smaller value represents a shift towards small edges and therefore a higher relative density of the point cloud. The **Clumpy** measure captures clustering of the data points in the point cloud by splitting the MST in two subgraphs by removing an edge and then searching for the longest edge in the smaller subgraph. This is repeated for every edge. It is small if the data points resemble a single cluster. The **Sparse** measure is defined as the 90%-percentile of the edge length of the MST. If it is low most edges are small and therefore the point cloud is denser than if it is large. The **Striated** measure calculates the proportion of striped edges to all edges in the MST. A pair of adjacent edges is considered striated if they form an angle larger than 138.5°. In our approach we consider edge pairs as straited if they form an angle larger than 100°. This metric is getting large if there are smooth paths in the MST. The **Convex** measure describes the ratio between the area of the alpha shape and the convex hull containing the data points. It is large if they overlap and therefor the alpha shape equal the convex hull. The **Skinny** measure tries to measure how skinny the point cloud is by comparing the perimeter to the area of its alpha shape. If the perimeter is much larger than the area the point cloud appears as a skinny shape. The **Stringy** measure calculates the ratio between the longest path to the total length of all edges in the MST. If the MST consists of only a few branches the longest path spans over most edges leading to higher values of the Stringy measure. The **Monotonic** measure is defined as squared Spearman correlation coefficient. It is large if there exists a monotonic trend between the two dimensions of the data points or zero if they are uncorrelated.

Layout Diagnostics Measures A simple set of layout diagnostic measures can be directly derived from the underling layout optimization approach. The optimizer tries to reach an equilibrium of different forces applied to each data point. The total length of force vectors for a specific force can be used as measure of how well a force can be minimized and thus be used as measure of how well the aspect represented by this force is present in the resulting layout. We define four Layout Diagnostic Measures based on this principle which correspond to the four forces used by our layout optimization approach: **Pairwise Repulsion Force Length**, **Target Position Force Length**, **Shape Bounds Repulsion Force Length** and **Border Repulsion Force length**. In addition, we also define an **Overlap** measure, defined as the percentage of shapes in the layout which have at least one intersection with other shapes. This measure allows a simple quantification of overplotting in the layout.

References

[DW14] DANG, TUAN NHON and WILKINSON, LELAND. "Transforming Scagnostics to Reveal Hidden Features". *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 20.12 (2014), 1624–1632. DOI: 10.1109/TVCG.2014.2346572 1.

[MTL18] MATUTE, JOSE, TELEA, ALEXANDRU C., and LINSEN, LARS. "Skeleton-based Scagnostics". *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 24.1 (Jan. 2018), 542–552. ISSN: 1077-2626. DOI: 10.1109/TVCG.2017.2744339 1.

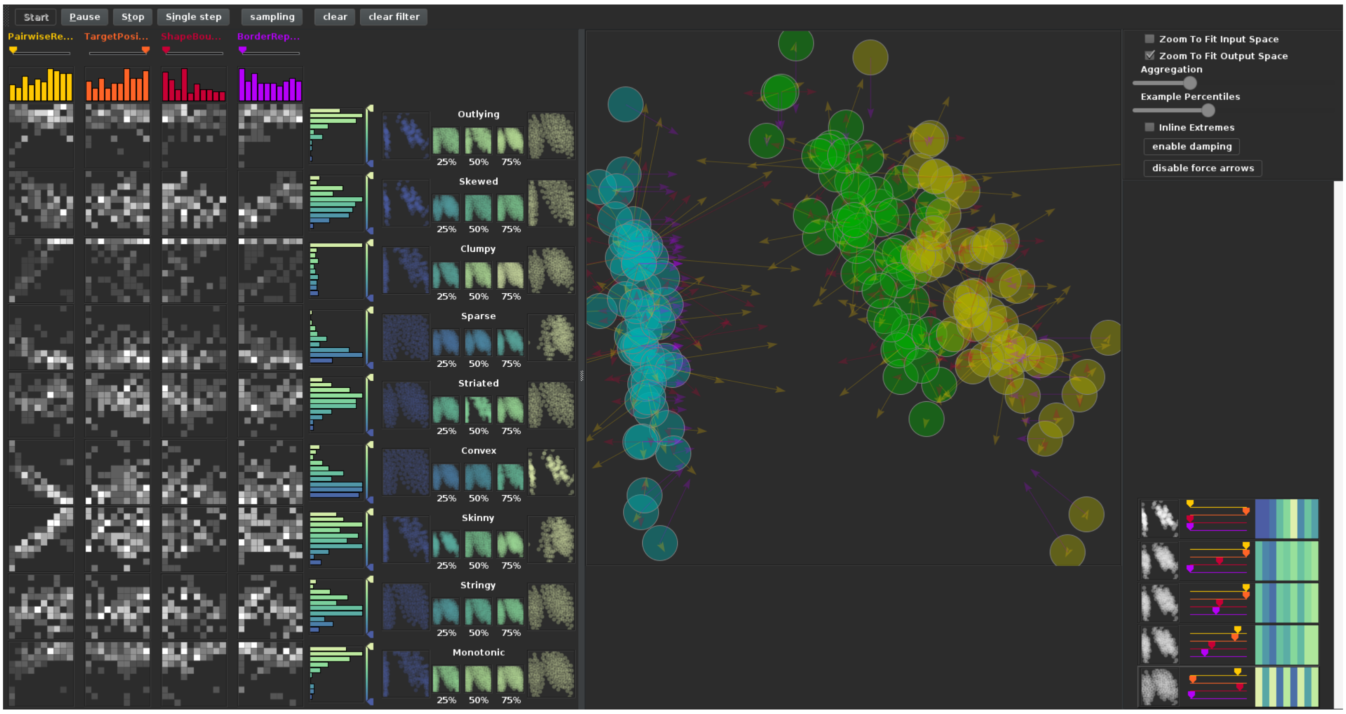

Figure 2: *Overview of the system with four parameters on the left (yellow, orange, red, purple), the nine Scagnostics measures cross-cut on the left, the current layout at the center (including forces), and a layout history at the bottom right showing the PCA-projected Iris dataset (Usage Scenario 2).*

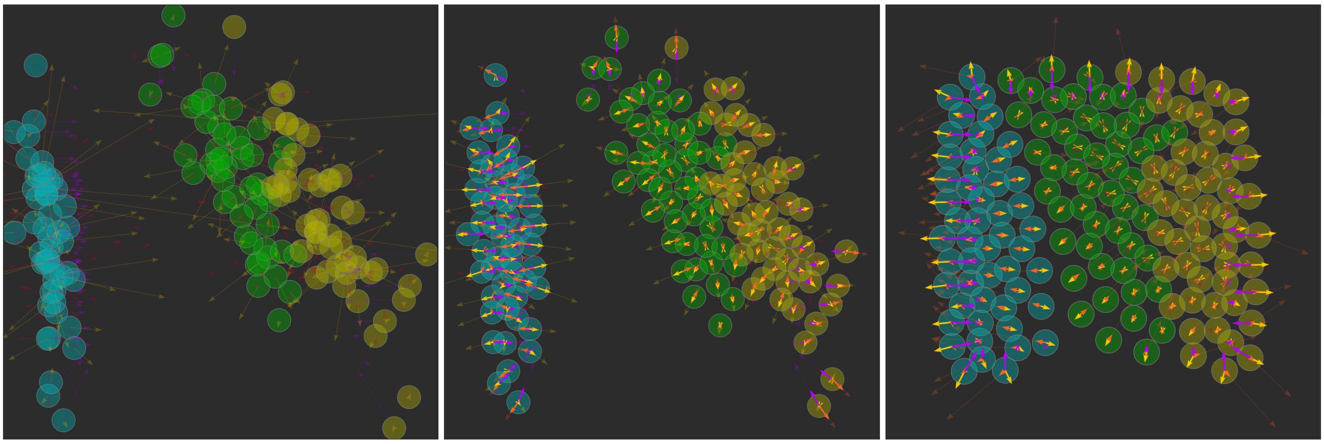

Figure 3: *Initial data positions of the iris dataset, as a result of an upstream PCA projection (left). A meaningful layout optimization leads to an assembly of data points that faithfully preserves original data positions while reducing overplotting considerably (center). The layout on the right is a weak result, as most of the structural information of the data points got lost. All three example layouts have been subject to analysis in Usage Scenario 2.*

[WAG06] Wilkinson, L., Anand, A., and Grossman, R. "High-Dimensional Visual Analytics: Interactive Exploration Guided by Pairwise Views of Point Distributions". *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 12.6 (2006), 1363–1372. DOI: 10.1109/TVCG.2006.94 1.

[WWL*20] Wang, Yunhai, Wang, Zeyu, Liu, Tingting, et al. "Improving the Robustness of Scagnostics". *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 26.1 (2020), 759–769. DOI: 10.1109/TVCG.2019.2934796 1.