EG UK Computer Graphics & Visual Computing (2024)
A. Slingsby and D. Hunter (Editors)

Exploring High-Dimensional Data by Pointwise Filtering of Low-Dimensional Embeddings

Daniel Atzberger[1], Adrian Jobst[1], Willy Scheibel[1], and Jürgen Döllner[2]

[1]Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam
[2]Digital Engineering Faculty, University of Potsdam

Abstract

Dimensionality reductions are a class of unsupervised learning algorithms that aim to find a lower-dimensional embedding for a high-dimensional dataset while preserving local and global structures. By representing a high-dimensional dataset as a two-dimensional scatterplot, a user can explore structures within the dataset. However, dimensionality reductions inherit distortions that might result in false deductions. This work presents a visualization approach that combines a two-dimensional scatterplot derived from a dimensionality reduction with two pointwise filtering possibilities. Each point is associated with two pointwise metrics that quantify the correctness of its neighborhood and similarity to surrounding data points. By setting threshold for these two metrics, the user is supported in several scatterplot analytics tasks, e.g., class separation and outlier detection. We apply our visualization to a text corpus to detect interesting data points visually and discuss the findings.

CCS Concepts

• ***Human-centered computing*** → *Information visualization; Visual analytics;*

1. Introduction

Dimensionality Reductions (DRs) constitute a class of unsupervised machine learning algorithms designed to uncover lower-dimensional representations of high-dimensional datasets while preserving both local and global structures [NA19]. By mapping each data point from the original high-dimensional space to a point in a lower-dimensional embedding, DRs enable the visualization of complex datasets in two or three dimensions. While these scatterplots may lack direct interpretability along the axes, neighborhood information still encode meaningful information about the similarity of data points in the original high-dimensional space. In contrast to classical multivariate data visualizations like parallel coordinate plots, scatterplot matrices, or glyphs, DRs offer several advantages, particularly in scalability concerning both the number of data points and the dimensionality of the dataset [WGK10]. This scalability makes DRs a state-of-the-art technique for effectively visualizing high-dimensional datasets, facilitating the exploration and understanding of complex data structures with opaque semantic [TMW24].

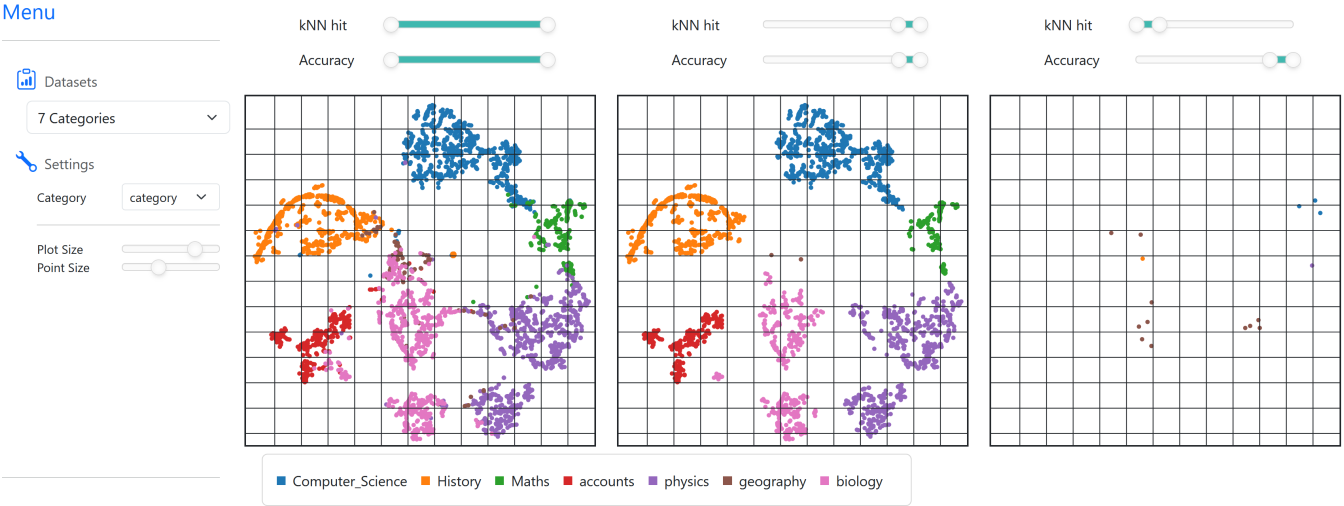
However, DRs do not preserve local and global structures perfectly in the low-dimensional embedding, i.e., they inherit distortions [NA19]. Therefore, conclusions about the high-dimensional dataset made by observing patterns in the low-dimensional embedding might be wrong. In this work, we present a visualization approach that combines a juxtaposition of scatterplots, which results from the application of a DR, with a filtering functionality for removing distortions and enhancing visual structures. Given the high-dimensional dataset and its low-dimensional embedding, we compute several metrics that capture the pointwise preservation of neighborhoods. By aggregating these metrics to a single metric, the user can set a threshold that removes distortions. Therefore, the user can confidently conclude the structure of the high-dimensional dataset based on its two-dimensional representation. Furthermore, we enable the user to specify a threshold to detect outliers and clusters more efficiently based on the neighborhood hit. Our filtering mechanism is embedded in a visualization system, as shown in Figure 1. We evaluated our visualization in the case of a text corpus for detecting mislabeled data points. Our findings demonstrate the effectiveness of our method in enhancing data visualization and interpretation. We implemented our approach using the Javascript library D3.js; we provided our prototype as a GitHub repository[†].

2. Related Work

Basics on Scatterplots. Scatterplots are a widely used technique for visualizing multivariate and high-dimensional data. Various visualizations based on scatterplots have been developed and applied "in a variety of exploratory and presentation contexts" [SG18]. Scatterplots can be applied to visualize the relationship between two variables or to display abstract similarity between data points

[†] hpicgs/filtering_scatterplots

DOI: 10.2312/cgvc.20241224



Datasets: 7 Categories

Category: category

Figure 1: *A prototypical dashboard that implements the proposed pointwise filtering. The high-dimensional data is given by the Seven Categories dataset, which contains a total of 3142 text documents from seven categories. The left panel shows the whole dataset. The center panel filters for coherent clusters where the number of shared nearest neighbors is maximized. The right panel filters for outliers concerning the number of shared nearest neighbors.*

encoded by their Euclidean distance. In the former case, often many scatterplots are arranged as a scatterplot matrix to show the pairwise relationships between several variables for a multivariate dataset [CLNL87]. Even though scatterplots scale better in the number of data points than traditional visualization techniques for multivariate data, they might suffer from overlap. Several approaches have been introduced to remove overlap to enhance readability, e.g., *splatterplots* [MG13].Furthermore, scatterplots might specify the layout of a visualization. For example, given a set of text documents, a so-called corpus, a "semantic" layout, which represents the semantic similarity between the documents, can be computed by applying a DR to a text embedding [ACS*24b]. Techniques for providing details on demand are required to relate data points in the two-dimensional scatterplot back to the high-dimensional data. For example, Kim et al. applied a lens view that shows a finer-grained layout for selected documents within a corpus [KKP*16]. Similarly, Raval et al. developed an approach that generates comprehensive word clouds and a natural language description for documents selected via a lasso [RWVW23]. For the case of multivariate data, Thijssen et al. presented a visualization approach to show statistics for each data dimension [TTT23]. Even though all visualizations aim to detect interesting structures in the high-dimensional dataset, they do not filter the datasets' distortions and might, therefore, derive false conclusions.

Quality Metrics for Dimensionality Reductions. Different metrics have been introduced to quantify quality aspects of DRs [BBK*18]. Thereby, one can distinguish between local and global accuracy metrics [NA19]. Local accuracy metrics measure how well neighborhoods are preserved in the low-dimensional embedding, whereas global accuracy metrics quantify the preservation of pairwise distances in the lower-dimensional embedding. Both kinds of quality metrics have been used in benchmark studies to compare the performance of several DRs by evaluating their results on a set of datasets [vdMPvdH09; EMK*21; ACS*24a; ACS*24b; VGS*20]. However, in these studies, the pointwise accuracy metrics were aggregated into one metric by taking the average of all points, whereas the accuracy metrics for the individual points are ignored.

Lespinats and Aupetit presented *CheckViz*, an approach to extending a two-dimensional scatterplot derived from a DR using uniform background color coding to visualize distortions [LA11]. CheckViz relies on pointwise accuracy metrics to detect false and missing neighbors, similar to our approach. The amount of distortions within a region is mapped onto the color of the respective Voronoi region, which helps the user spot points that do not allow interpretation of the high-dimensional data. Jeon et al. applied a similar approach to visualize inter-cluster distortions [JKJ*21]. Heiter et al. introduced *TRACE*, a system designed for exploring two-dimensional embeddings by mapping accuracy metrics onto the color of points [HMS*24]. The authors demonstrate how their system facilitates the analysis of both local and global distortions, as well as the comparison of different embeddings.

3. Visualization Design

Metrics for Filtering Scatterplots. In the case of local accuracy metrics, for each point, neighborhoods of fixed size in the original space and the lower-dimensional representation are compared and assigned a value. By averaging these values over each point, a single metric is derived. By assessing the pointwise values, local distortions can be filtered, and therefore, inferences on the high-dimensional dataset based on the lower-dimensional embeddings can be drawn with higher confidence. In the following, let $\mathcal{X} = \{x_1,\dots,x_N\} \subseteq \mathbb{R}^n$ denote the high-dimensional dataset and $\mathcal{Y} = \{y_1,\dots,y_N\} \subseteq \mathbb{R}^2$ its two-dimensional representation after applying a DR. For a fixed value of $k \in \{1,\dots,N\}$, let $\mathcal{N}_k^{\mathbb{R}^n}(i)$ and

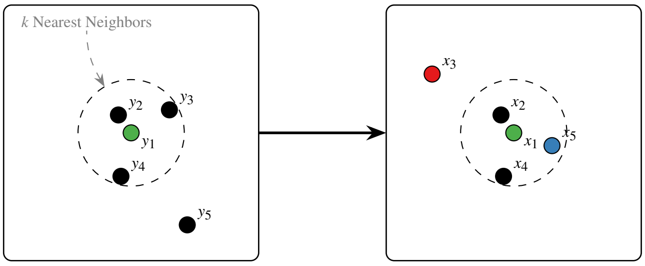


Figure 2: *Illustration for the false neighbors and missing neighbors. The circles indicate the k nearest neighbors ($k = 3$) of y_1 and x_1. The point x_5 (blue) belongs to the kNNs of x_1, but y_5 is not among the kNNs of y_1, i.e., $5 \in \mathcal{N}_k^{\mathbb{R}^2}(1) \setminus \mathcal{N}_k^{\mathbb{R}^n}(1)$, and therefore x_5 is a false neighbor. The point x_3 (red) does not belong to the kNNs of x_1, but y_3 is among the kNNs of y_1, i.e., $3 \in \mathcal{N}_k^{\mathbb{R}^n}(1) \setminus \mathcal{N}_k^{\mathbb{R}^2}(1)$, and therefore y_3 is a missing neighbor. The design and concept of this figure are inspired by Nonato and Aupetit [NA19]*

$\mathcal{N}_k^{\mathbb{R}^2}(i)$ denote the indices of the k nearest neighbors of point x_i or y_i in $\mathbb{R}^n$ or $\mathbb{R}^2$, respectively. Ideally the neighborhoods are preserved by a DR. However, two types of distortions might occur. On the one hand, a neighbor in the high-dimensional neighborhood is not represented in the lower-dimensional embedding, i.e., neighbors are missing. On the other hand, a neighbor shown in the low-dimensional embedding might not be contained in the high-dimensional dataset, i.e., neighbors are false. The concept of missing and false neighbors is illustrated in Figure 2. Most accuracy metrics include rankings within the neighborhoods to yield a more fine-granular quantification, e.g., the *Mean Relative Rank Errors* (MRREs) [LV09]. Let ρ_{ij} denote the rank of point x_j in the ordering according to the shortest distance from point x_i in $\mathbb{R}^n$. Analogously, let r_{ij} denote the rank of point y_j in the ordering according to the shortest distance from point y_i in $\mathbb{R}^2$. The MRRE of point x_i for the lower-dimensional representation is given by

$$\alpha_{MRRE,\mathbb{R}^2}(i) = \frac{1}{\sum\limits_{l=1}^{k} \frac{|N-2l+1|}{l}} \sum_{j\in\mathcal{N}_k^{\mathbb{R}^2}(i)} \frac{|\rho_{ij}-r_{ij}|}{\rho_{ij}} \tag{1}$$

Analogously, the MRRE of point y_i in the high-dimensional representation is given by

$$\alpha_{MRRE,\mathbb{R}^n}(i) = \frac{1}{\sum\limits_{l=1}^{k} \frac{|N-2l+1|}{l}} \sum_{j\in\mathcal{N}_k^{\mathbb{R}^n}(i)} \frac{|r_{ij}-\rho_{ij}|}{r_{ij}} \tag{2}$$

The *Trustworthiness* measure also considers rankings of the points [VK06]. However, different from the MRREs, the trustworthiness only considers points within $\mathcal{N}_k^{\mathbb{R}^2}(i)$ that are not in $\mathcal{N}_k^{\mathbb{R}^n}(i)$. The pointwise trustworthiness of point x_i is given by:

$$\alpha_T(i) = 1 - \frac{2}{k(2N-3k-1)} \sum_{j\in\mathcal{N}_k^{\mathbb{R}^2}(i)\setminus\mathcal{N}_k^{\mathbb{R}^n}(i)} (\rho_{ij}-k), \tag{3}$$

Analogously, the *Continuity* considers the points that are in $\mathcal{N}_k^{\mathbb{R}^n}(i)$ but not in $\mathcal{N}_k^{\mathbb{R}^2}(i)$ [VK06]. Its formula coincides with the trustworthiness, but uses r_{ij} instead of ρ_{ij}. We aggregate these four metrics to a single pointwise accuracy metric α by taking their average.

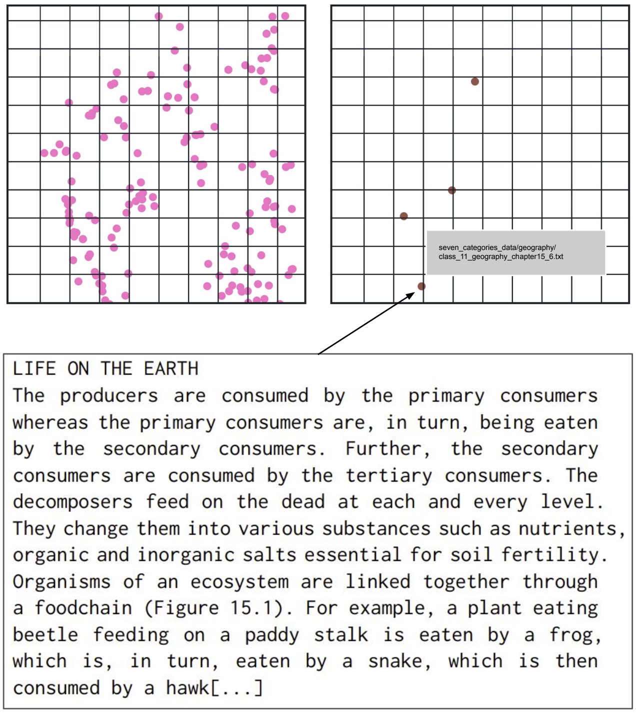
In our experiments, we assume that each point is associated with a unique data category. The pointwise *Neighborhood Hit* compares the categories of the k nearest neighbors for each point [PTT*12]. The neighborhood hit for point x_i is therefore given by

$$\beta(i) = \frac{|\{j \in \mathcal{N}_k^{\mathbb{R}^2}(i) | c_i = c_j\}|}{k}, \tag{4}$$

where c_i and c_j denote the classes of the points x_i and x_j. In any case, we set $k = 7$ in alignment with previous benchmark studies.

Visual Mapping. Our visualization system requires a tabular dataset with designated columns x and y, at least one categorical variable, and the pointwise metric α. The metric β is derived from the coordinates and the associated labels. Each data point is displayed within a scatterplot specified by x and y coordinates. We ignore the scales within the scatterplot as they have no meaning; only the closeness between points is interpreted as similar according to the *Gestalt Principles* [War19] The category of each data point is mapped onto its color using a qualitative color scheme. Our interaction design follows Shneiderman's information-seeking mantra [Shn03]: *Overview first, zoom and filter, then details-on-demand.* The overview is provided in three identical scatterplots arranged as a juxtaposition [LS15]. Initially, all points are displayed without any filtering. Additional grid lines support the user's navigation. Two sliders over each scatterplot allow the user to select thresholds for the local accuracy metric α and the neighborhood hit β. By setting a high value for α, the number of distortions from the DR will decrease as points are removed. Points that differ significantly from the majority of points are called outliers. In our case of well-defined clusters given by the predefined categories, an outlier can be associated with a point with a different label than his k nearest neighbors. By selecting a low value for β, outliers are isolated from their surrounding clusters of different categories. Conversely, setting a high value for β removes outliers and emphasizes given clusters. Since three scatterplots are given, the entire view without any filtering, the filtered version with high α and high β (cluster view), and the filtered version with high α and low β (outlier view), the user can inspect all three versions simultaneously and combine them, e.g., the surrounding cluster of an outlier. This constellation is shown in Figure 1. The user can explore specific regions within the scatterplot via the zoom functionality. The zoom functions are connected across all three scatterplots, i.e., zooming in on one scatterplot results in zooming in on the remaining two. The grid lines are thereby adaptively updated. By hovering over a point, its unique identifier is shown as a tooltip. This enables the user to analyze the selected point by inspecting its entries in the original dataset.

Implementation. Our visualization prototype is implemented as a website as it allows for high accessibility and interactivity. We make use of *Bootstrap* for general CSS styling and its grid system to arrange the plots and interactive elements of our visualization. Scatterplots are implemented by using the javascript framework *D3.js* [BOH11], whose select-append mechanism enables data connection with DOM elements. Furthermore, functions like zooming, panning, tooltips, and grid lines were implemented with D3.js's functionality. Our implementation of accuracy metrics relies on the



LIFE ON THE EARTH

The producers are consumed by the primary consumers whereas the primary consumers are, in turn, being eaten by the secondary consumers. Further, the secondary consumers are consumed by the tertiary consumers. The decomposers feed on the dead at each and every level. They change them into various substances such as nutrients, organic and inorganic salts essential for soil fertility. Organisms of an ecosystem are linked together through a foodchain (Figure 15.1). For example, a plant eating beetle feeding on a paddy stalk is eaten by a frog, which is, in turn, eaten by a snake, which is then consumed by a hawk[...]

Figure 3: *Zoom into the outlier view for the seven categories corpus. The selected data point, whose filename is shown as a tooltip, is associated with geography but is mainly surrounded by documents that are associated with biology. By inspecting the document, it becomes apparent that it is rather about biology.*

Python library *ZADU* [JCJ*23] and the implementation provided by Atzberger and Cech et al. [ACS*24b]. We refer to our GitHub repository for details on our implementation and setup instructions.

4. Results

In the following, we apply our visualization to the *Seven Categories Corpus* from *Kaggle*[‡]. The corpus consists of 3142 text documents, each associated with a unique category that indicates its underlying semantics. Given the documents, we perform baseline preprocessing steps, e.g., tokenization, removal of stop words, and lemmatization, to reduce the vocabulary size and filter out words that carry no semantic meaning. After the preprocessing, the corpus is given as a document-term matrix (DTM), which stores the absolute frequencies of the words within the documents. We then apply Latent Dirichlet Allocation together with t-SNE to derive a two-dimensional layout in which the Euclidean distance reflects semantic similarity [ACS*24b; ACS*24a]. The resulting scatterplot representation is shown in Figure 1. To emphasize the clusters in the scatterplot, we set the threshold for the kNN hit in the second scatterplot to 0.85. To emphasize the points that are associated with a different category than its neighbors, we filter for points with a kNN hit smaller than 0.15. In both cases, we set the threshold of accuracy to 0.85 to remove distortions. Inspecting the remaining points in the outlier view yields data points that differ from their neighbors and others from the same category. One example of such an outlier is shown in Figure 3, which shows a data point that is associated with the category *geography* but lies within the *biology* cluster.

5. Conclusions

DRs are widely used for visualizing high-dimensional data as they scale with the number of dimensions and data points. However, the resulting scatterplots might inherit distortions; therefore, conclusions drawn from the scatterplot might be wrong. We presented a visualization that allows for filtering points based on two-pointwise metrics that remove distortions and emphasize clusters and outliers. Thus, the visualization supports users in detecting points of special interest. In any case, outliers and clusters could also be computationally detected, e.g., by using a clustering algorithm like *k-Means* [ASI20] and an outlier detection algorithm like *isolation forest* [LTZ08]. However such clustering algorithms and outlier detection techniques, are not without their limitations. They can inherit errors that are not immediately apparent to the user, potentially leading to inaccurate results. This is a significant drawback that our approach overcomes, providing users with more reliable and interpretable results. Furthermore, in both cases, the results would be given as lists of integers, which do not allow for easy connection. In contrast, in our proposed visualization approach, patterns can be detected and linked with high confidence.

We see different directions for future work. One idea is to integrate our filtering mechanism into existing visualizations that build up on a scatterplot layout, e.g., *Bubble Sets* [CPC09] or glyph visualizations [KKG*20]. In particular, it would be interesting to what extent the visualizations are affected when distortions are removed and clusters and outliers are highlighted. Furthermore, to make our concept more accessible, it would be beneficial to integrate our concept into widely used libraries and frameworks, e.g., a Python library that can be used within a Jupyter notebook, or by hosting a web service where users can upload and inspect their data.

Acknowledgments

This work was partially funded by the the Federal Ministry for Economic Affairs and Climate Action of Germany through grant 16KN086467 ("DecodingFood").

References

[ACS*24a] Atzberger, D., Cech, T., Scheibel, W., Döllner, J., and Schreck, T. "Quantifying Topic Model Influence on Text Layouts based on Dimensionality Reductions". *Proc. 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications – Volume 3*. IVAPP '24. INSTICC. SciTePress, 2024, 593–602. ISBN: 978-989-758-679-8. DOI: 10.5220/0012391100003660 2, 4.

[ACS*24b] Atzberger, D., Cech, T., Scheibel, W., Trapp, M., Richter, R., Döllner, J., and Schreck, T. "Large-Scale Evaluation of Topic Models and Dimensionality Reduction Methods for 2D Text Spatialization". *Transactions on Visualization and Computer Graphics* 30.1 (2024), 902–912. DOI: 10.1109/TVCG.2023.3326569 2, 4.

[‡] kaggle.com/datasets/deepak711/4-subject-data-text-classification

[ASI20] Ahmed, M., Seraj, R., and Islam, S. M. S. "The k-means algorithm: A comprehensive survey and performance evaluation". *Electronics* 9.8 (2020), 1295. doi: 10.3390/electronics9081295 4.

[BBK*18] Behrisch, M., Blumenschein, M., Kim, N. W., Shao, L., El-Assady, M., Fuchs, J., Seebacher, D., Diehl, A., Brandes, U., Pfister, H., Schreck, T., Weiskopf, D., and Keim, D. A. "Quality metrics for information visualization". *Computer Graphics Forum* 37.3 (2018), 625–662. doi: 10.1111/cgf.13446 2.

[BOH11] Bostock, M., Ogievetsky, V., and Heer, J. "D^3 Data-Driven Documents". *Transactions on Visualization and Computer Graphics* 17.12 (2011), 2301–2309. doi: 10.1109/TVCG.2011.185 3.

[CLNL87] Carr, D. B., Littlefield, R. J., Nicholson, W., and Littlefield, J. "Scatterplot matrix techniques for large N". *Journal of the American Statistical Association* 82.398 (1987), 424–436. doi: 10.1080/01621459.1987.10478445 2.

[CPC09] Collins, C., Penn, G., and Carpendale, S. "Bubble Sets: Revealing Set Relations with Isocontours over Existing Visualizations". *Transactions on Visualization and Computer Graphics* 15.6 (2009), 1009–1016. doi: 10.1109/TVCG.2009.122 4.

[EMK*21] Espadoto, M., Martins, R. M., Kerren, A., Hirata, N. S. T., and Telea, A. C. "Toward a Quantitative Survey of Dimension Reduction Techniques". *Transactions on Visualization and Computer Graphics* 27.3 (2021), 2153–2173. doi: 10.1109/TVCG.2019.2944182 2.

[HMS*24] Heiter, E., Martens, L., Seurinck, R., Guilliams, M., De Bie, T., Saeys, Y., and Lijffijt, J. "Pattern or Artifact? Interactively Exploring Embedding Quality with TRACE". *arXiv preprint arXiv:2406.12953* (2024). doi: 10.48550/arXiv.2406.12953 2.

[JCJ*23] Jeon, H., Cho, A., Jang, J., Lee, S., Hyun, J., Ko, H.-K., Jo, J., and Seo, J. "ZADU: A Python Library for Evaluating the Reliability of Dimensionality Reduction Embeddings". *Proc. Conference on Visualization and Visual Analytics*. VIS '23. 2023, 196–200. doi: 10.1109/VIS54172.2023.00048 4.

[JKJ*21] Jeon, H., Ko, H.-K., Jo, J., Kim, Y., and Seo, J. "Measuring and explaining the inter-cluster reliability of multidimensional projections". *Transactions on Visualization and Computer Graphics* 28.1 (2021), 551–561. doi: 10.1109/TVCG.2021.3114833 2.

[KKG*20] Kammer, D., Keck, M., Gründer, T., Maasch, A., Thom, T., Kleinsteuber, M., and Groh, R. "Glyphboard: Visual Exploration of High-Dimensional Data Combining Glyphs with Dimensionality Reduction". *Transactions on Visualization and Computer Graphics* 26.4 (2020), 1661–1671. doi: 10.1109/TVCG.2020.2969060 4.

[KKP*16] Kim, M., Kang, K., Park, D., Choo, J., and Elmqvist, N. "TopicLens: Efficient Multi-level Visual Topic Exploration of Large-scale Document Collections". *Transactions on Visualization and Computer Graphics* 23.1 (2016), 151–160. doi: 10.1109/TVCG.2016.2598445 2.

[LA11] Lespinats, S. and Aupetit, M. "CheckViz: Sanity Check and Topological Clues for Linear and Non-Linear Mappings". *Computer Graphics Forum* 30.1 (2011), 113–125. doi: 10.1111/j.1467-8659.2010.01835.x 2.

[LS15] Liu, X. and Shen, H.-W. "The Effects of Representation and Juxtaposition on Graphical Perception of Matrix Visualization". *Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Association for Computing Machinery, 2015, 269–278. doi: 10.1145/2702123.2702217 3.

[LTZ08] Liu, F. T., Ting, K. M., and Zhou, Z.-H. "Isolation Forest". *Proc. 8th IEEE International Conference on Data Mining*. ICDM '08. IEEE, 2008, 413–422. doi: 10.1109/ICDM.2008.17 4.

[LV09] Lee, J. A. and Verleysen, M. "Quality assessment of dimensionality reduction: Rank-based criteria". *Neurocomputing* 72.7–9 (2009), 1431–1443. doi: 10.1016/j.neucom.2008.12.017 3.

[MG13] Mayorga, A. and Gleicher, M. "Splatterplots: Overcoming Overdraw in Scatter Plots". *Transactions on Visualization and Computer Graphics* 19.9 (2013), 1526–1538. doi: 10.1109/TVCG.2013.65 2.

[NA19] Nonato, L. G. and Aupetit, M. "Multidimensional Projection for Visual Analytics: Linking Techniques with Distortions, Tasks, and Layout Enrichment". *Transactions on Visualization and Computer Graphics* 25.8 (2019), 2650–2673. doi: 10.1109/TVCG.2018.2846735 1–3.

[PTT*12] Paulovich, F. V., Toledo, F. M. B., Telles, G. P., Minghim, R., and Nonato, L. G. "Semantic Wordification of Document Collections". *Computer Graphics Forum* 31.3pt3 (2012), 1145–1153. doi: 10.1111/j.1467-8659.2012.03107.x 3.

[RWVW23] Raval, S., Wang, C., Viégas, F., and Wattenberg, M. "Explain-and-Test: An Interactive Machine Learning Framework for Exploring Text Embeddings". *Proc. Conference on Visualization and Visual Analytics*. VIS '23. IEEE, 2023, 216–220. doi: 10.1109/VIS54172.2023.00052 2.

[SG18] Sarikaya, A. and Gleicher, M. "Scatterplots: Tasks, Data, and Designs". *Transactions on Visualization and Computer Graphics* 24.1 (2018), 402–412. doi: 10.1109/TVCG.2017.2744184 1.

[Shn03] Shneiderman, B. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations". *The Craft of Information Visualization*. Interactive Technologies. Morgan Kaufmann, 2003, 364–371. isbn: 978-1-55860-915-0. doi: 10.1016/B978-155860915-0/50046-9 3.

[TMW24] Telea, A., Machado, A., and Wang, Y. "Seeing is Learning in High Dimensions: The Synergy Between Dimensionality Reduction and Machine Learning". *SN Computer Science* 5.3 (2024), 279. doi: 10.1007/s42979-024-02604-y 1.

[TTT23] Thijssen, J., Tian, Z., and Telea, A. "Scaling Up the Explanation of Multidimensional Projections". *Proc. EuroVis Workshop on Visual Analytics*. EuroVA '23. The Eurographics Association, 2023, 1–6. doi: 10.2312/eurova.20231098 2.

[vdMPvdH09] Van der Maaten, L., Postma, E., and van den Herik, J. *Dimensionality reduction: a comparative review*. Tech. rep. 009-005. Tilburg University, Tilburg Centre for Creative Computing, The Netherlands, 2009 2.

[VGS*20] Vernier, E. F., Garcia, R., Silva, I. d., Comba, J. L. D., and Telea, A. C. "Quantitative evaluation of time-dependent multidimensional projection techniques". *Computer Graphics Forum* 39.3 (2020), 241–252. doi: 10.1111/cgf.13977 2.

[VK06] Venna, J. and Kaski, S. "Visualizing gene interaction graphs with local multidimensional scaling". *Proc. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. ESANN, 2006, 557–562. isbn: 2-930307-06-4 3.

[War19] Ware, C. *Information Visualization: Perception for Design*. 4th ed. Morgan Kaufmann, 2019. isbn: 978-0-12-812875-6 3.

[WGK10] Ward, M. O., Grinstein, G., and Keim, D. *Interactive data visualization: foundations, techniques, and applications*. AK Peters/CRC Press, 2010. isbn: 978-1568814735. doi: 10.1201/9780429108433 1.

Proceedings published by Eurographics - The European Association for Computer Graphics.