Visualizing the quality
of dimensionality reduction

Bassam Mokbel[1], Wouter Lueks[2],[*]
Andrej Gisbrecht[1], Michael Biehl[2], Barbara Hammer[1]

[1] Bielefeld University - CITEC Centre of Excellence, Germany
[2] University of Groningen - Faculty of Mathematics and Natural Sciences, Netherlands

Abstract.   Many different evaluation measures for dimensionality reduction can be summarized based on the co-ranking framework [6]. Here, we extend this framework in two ways: (i) we show that the current parameterization of the quality shows unpredictable behavior, even in simple settings, and we propose a different parameterization which yields more intuitive results; (ii) we propose how to link the quality to point-wise quality measures which can directly be integrated into the visualization.

1   Introduction

With more and more nonlinear dimensionality reduction (DR) techniques becoming readily available, there is an increasing need for formal evaluation measures to compare DR techniques, to optimize their parameters and solutions in case of multiple optima, and to directly assess the quality of such mappings. In recent years, several quality measures have been proposed [4, 6, 11], many of which can be summarized based on the co-ranking framework as proposed in [6]. In this contribution, we argue that the co-ranking framework covers even more cases including, e.g., DR evaluation based on information retrieval principles [11].

The co-ranking framework in its current form, however, has a serious drawback: even if data and its projections can be directly inspected, it is difficult to predict the characteristics of the quality measure depending on its parameter K, the neighborhood size. The evaluation is typically subsumed in a quality curve over all possible K. In several simple scenarios, we demonstrate that its shape is hard to predict. We argue that this behavior comes from small or contained errors in the mapping resulting in changes at a large range of K. Based on this observation, we propose a different parameterization which yields more predictable behavior and which can directly highlight structurally interesting neighborhood sizes. By linking global quality measures to point-wise quantities, this observation gives rise to an integration of the quality into the visualization itself, similar to heuristic alternatives in topographic mapping such as [1, 8].

2   Evaluation measures for dimensionality reduction

DR techniques are used to map a high-dimensional dataset $\Xi = \{\xi_1, \ldots, \xi_N\} \subset \mathbb{R}^D$ to a low-dimensional dataset $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^L$, with $L < D$ and $L = 2$ or $L = 3$ for the purpose of visualization. Let δ_{ij} be the distance from ξ_i to ξ_j in $\mathbb{R}^D$ and d_{ij} the distance from x_i to x_j in the low-dimensional space. The rank of ξ_j with respect to ξ_i in $\mathbb{R}^D$ is given by

$$\rho_{ij} = |\{k \mid \delta_{ik} < \delta_{ij} \text{ or } (\delta_{ik} = \delta_{ij} \text{ and } 1 \le k < j \le N)\}| \,.$$

[*] also affiliated with University of Nijmegen - Faculty of Science, The Netherlands

Analogously, the rank of x_j with respect to x_i in the low-dimensional space is

$$r_{ij} = |\{k \mid d_{ik} < d_{ij} \text{ or } (d_{ik} = d_{ij} \text{ and } 1 \leq k < j \leq N)\}|.$$

The co-ranking matrix Q [6] is defined by $Q_{kl} = |\{(i,j) \mid \rho_{ij} = k \text{ and } r_{ij} = l\}|$. Errors of a DR mapping correspond to off-diagonal entries in this matrix. A point j where $\rho_{ij} > r_{ij}$ is an *intrusion*, if $\rho_{ij} < r_{ij}$ it is an *extrusion*. Usually, a DR mapping is not used to map all relationships of data faithfully. Often, the focus is on the preservation of local relationships. The co-ranking matrix offers a framework, in which several existing evaluation measures can be expressed, as pointed out in [6]: *Local Continuity Meta Criterion* (LCMC) [3], *Trustworthiness & Continuity* (T&C) [10], and *Mean Relative Rank Errors* (MRRE) [5]. Essentially, these quality measures correspond to weighted sums of entries Q_{kl} of the co-ranking matrix for regions $k, l \leq K$ and fixed neighborhood range K. In [6], an intuitive (unweighted) sum has been proposed, the *Quality* Q_{NX}:

$$Q_{\mathrm{NX}}(K) = \frac{1}{KN} \sum_{k=1}^{K} \sum_{l=1}^{K} Q_{kl}. \tag{1}$$

This summarizes all 'benevolent' points which change their rank only within a fixed neighborhood K. To display the quality, usually the curve $Q_{\mathrm{NX}}(K)$ is plotted for a range $K > 1$. At present, no generally accepted transformation of this curve into a single meaningful value exists. In [7], an aggregated measure is proposed which averages the quality for values $K \leq k_0$ where the splitting point k_0 is taken based on the distance of Q_{NX} to the baseline – however, this automatic choice is yet very sensitive to minor changes in the data.

Interestingly, the quality measure Q_{NX} is identical to the *Quality of Point Neighborhood Preservation* (QNP) as proposed in [4]. Further, it coincides with the measures *Precision* and *Recall* from [11] based on an information retrieval perspective, provided a local neighborhood of a point is defined by its K-nearest neighbors in the original and the projection space, respectively.

Distinct alternatives to the co-ranking framework are offered by measurements which rely on ϵ-neighborhoods instead of K-nearest neighbors [11], on global correlations of the distances [4], or the distances of K-neighbors such as in the topographic product [2]. Note that the latter two take a more global perspective, judging the global topology preservation rather than local rankings.

3 Intuitive parameterization of the quality

Many different quality measures are essentially similar to the quality Q_{NX} by taking a weighted sum over entries $Q_{kl}, k, l \leq K$ of the co-ranking matrix. Even in very simple settings, however, it is difficult to interpret the shape of the quality curve $Q_{\mathrm{NX}}(K)$. We consider three simple scenarios. The original data consist of three well separated Gaussian clusters in the plane containing 100 points each, see Fig. 1a. As a 'mapping', we consider the points in two dimensions obtained by (i) a random swapping of points within every cluster only, (ii) a switch of the two leftmost clusters, (iii) the middle and leftmost cluster are stacked on top of each other. These simple artificial mappings represent typical behavior of DR mappings since they capture (i) local distortions, (ii) a tearing of regions, (iii) and an overlay of regions, common effects of a too small projection dimension.
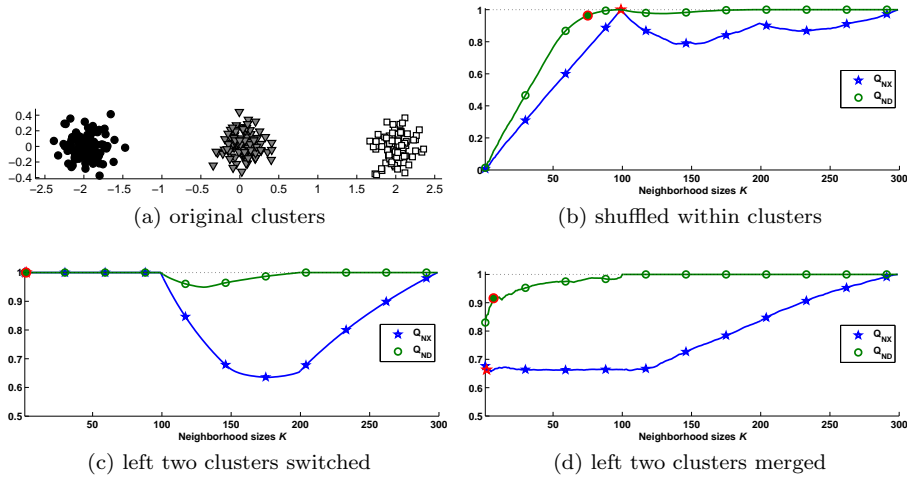
Fig. 1: Quality evaluation on artificial data to demonstrate the interpretability of the curves Q_{NX} vs. Q_{ND}, for all settings of K.

The resulting curves Q_{NX} are depicted in Fig. 1. Although we know the exact behavior of the mapping, it is not easy to link the shape of Q_{NX} to structures of the mapping. Setting (i) yields local errors within the clusters of size 100 only, still Q_{NX} is below the optimum 1 for all neighborhood sizes K, and it suggests a (not existing) structural match at size 200 by a local maximum of the curve, see Fig. 1b. A different parameterization of the quality which we will introduce below, Q_{ND}, yields the alternative as depicted in Fig. 1b. Here, it is clearly visible that local distortions take place up to a neighborhood size 100 (within clusters), while the mapping is almost perfect afterwards. The quality Q_{NX} suggests perfect mappings for small neighborhood sizes up to 100 for setting (ii) and a suboptimal mapping for all larger scales, manifested in a severe drop of the quality for $K > 100$, see Fig. 1c. In contrast, the measure Q_{ND} indicates a slight drop of accuracy for the range $K \in \{100, \ldots, 200\}$ which exactly corresponds to the region where the two leftmost clusters are switched. Similarly, Q_{NX} depicts a mapping quality which is not optimum for all ranks K for case (iii) although only the leftmost two clusters are stacked on top of each other, hence significant distortions take place in the range $K \in \{1, \ldots, 100\}$ only. In contrast, this fact is clearly indicated when regarding Q_{ND}, see Fig. 1d.

What is the reason behind these effects? Q_{NX} uses the neighborhood size K for two different purposes: on the one hand, K singles out the region of interest by determining the size of the neighborhood of points taken into account, namely $\rho_{ij} \leq K$. On the other hand, it determines the size and shape of errors which are tolerated: $r_{ij} \leq K$ is not counted as an error in the region of interest. Note that the actual size of the rank error is not considered at all as long as the rank is smaller than K. This parameterization has the effect that small rank errors can contribute to the shape of $Q_{\mathrm{NX}}(K)$ for every value K. Consider the following extreme setting: Points are swapped pairwise as shown in Fig. 2. Obviously, rank errors of size at most 4 occur due to the only local changes.

This is clearly mirrored by the co-ranking matrix displayed in Fig. 2 for which only four off-diagonals are not equal to 0. Due to the choice of Q_{NX}, however, the quality does not approach one, rather errors occur for every neighborhood K, see Fig. 2c. This property is shared by any quality measure which relies on the same part $Q_{kl}, k, l \leq K$ of the co-ranking matrix.

This observation also suggests how to change the parameterization of Q_{NX} with respect to the co-ranking matrix to obtain more intuitive results, as depicted in Fig. 3. As a first step, instead of taking a single parameter K, one can decouple the control parameter for the region of interest K_s and the control parameter for the tolerated errors K_t, thus summing over a rectangular part $Q_{kl}, k \leq K_s, l \leq K_t$, see Fig. 3b. As a second step, one can choose the error tolerance based on the actual size of rank errors instead of accepting all ranks at most K_t, i.e. we sum over κ_t off-diagonals instead of the first K_t columns, see Fig. 3c. This choice yields a quality measure parameterized by the region of interest K_s and the tolerated rank error κ_t:

$$Q_{\mathrm{ND}}(K_s, \kappa_t) = \frac{1}{K_s N} \sum_{i \leq K_s} \sum_{j:|i-j| \leq \kappa_t} Q_{ij} \ .$$

To reduce the number of parameters and hence the complexity, we suggest to choose the tolerated error to be of the same size as the region of interest: $\kappa_t = K_s =: \kappa$, resulting in a reparameterized quality where, similar to [6], normalization takes place to guarantee values in the interval $[0, 1]$:

$$Q_{\mathrm{ND}}(\kappa) = \frac{1}{\kappa N} \sum_{i \leq \kappa} \sum_{j:|i-j| \leq \kappa} Q_{ij} \ .$$

Unlike Q_{NX}, the measure Q_{ND} makes sure that small rank errors contribute only to values $Q_{\mathrm{ND}}(K)$ for small K and they do not spread over the entire quality curve, making the behavior more predictable.

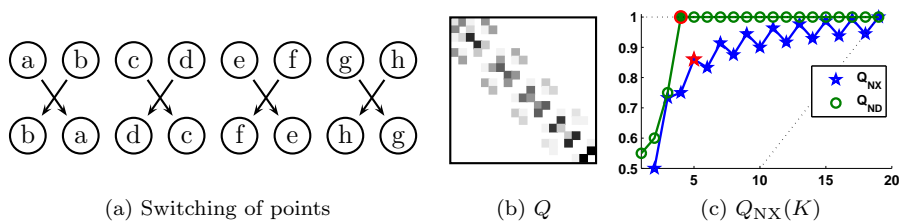(a) Switching of points   (b) Q (c) $Q_{\mathrm{NX}}(K)$

Fig. 2: The left shows a simple switching scheme of one-dimensional points. Obviously, rank errors are at most four (in case of tie breaks) in this setting. This is mirrored by the shape of the co-ranking matrix (in the middle for the same setting with 20 points) for which four off-diagonals are non-vanishing, and by $Q_{\mathrm{ND}}(K){=}1$ for all $K > 4$, while $Q_{\mathrm{NX}}(K) < 1$ for almost all K (on the right).
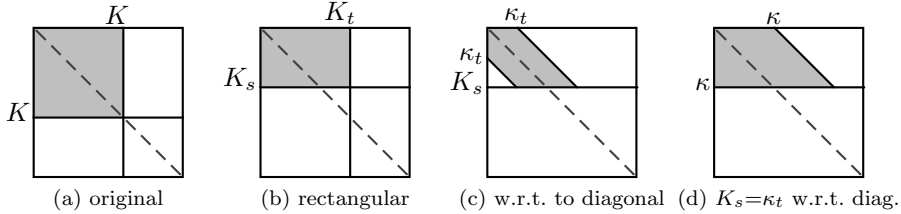
Fig. 3: Change of the summation area of the co-ranking matrix to better control the region of interest and the tolerated rank errors.

4 Point-wise quality measure

It has been pointed out e.g. in the work [1, 8] that a given visualization should also provide a direct intuition in how far it can be relied on. Ideally, together with a mapping of points, the quality of a mapped point should be visualized directly at its respective location. While the approaches [1, 8] provide very effective heuristics to do so, surprisingly, none of the formal DR evaluation measures have been used so far to directly visualize the quality of the projections.

Note that the co-ranking matrix can easily be accompanied by a corresponding point-wise framework: $Q_{kl}^i = |\{j|\rho_{ij} = k \text{ and } r_{ij} = l\}|$ which gives rise to point-wise quality measures $Q_{NX}^i(K) = \sum_{k \leq K} \sum_{l \leq K} Q_{kl}^i / K$ and $Q_{ND}^i(K) = \sum_{k \leq K} \sum_{l:|k-l| \leq K} Q_{kl}^i / K$ which, averaged over all points, again yield the quality measures Q_{NX} and Q_{ND}, respectively. These measures can be used to display, together with the DR, the quality of the projection at every point by means of a color value corresponding to $Q_{NX}^i(K)$ or $Q_{ND}^i(K)$ for relevant K. Thereby, K can be chosen according to relevant structural criteria such as a local extremum of the curve Q_{NX} or Q_{ND}, or it can be chosen interactively according to the users needs.

An example is given in Fig. 4 for the classical *swiss roll* data set visualized using t-SNE [9] and choosing K according to the first local optimum of the quality curve. Interestingly, the measure $Q_{ND}^i(K)$ clearly singles out positions where tearing and topological mismatches occur, while $Q_{NX}^i(K)$ shares the problem of Q_{NX} that local rank errors can affect different regions of the curve, thus also indicating many less pronounced points in the mapping.

5 Conclusions

We have discussed quality measures based on the co-ranking framework and problems which arise if quadratic regions of the matrix are used to define a quality measure. In contrast, we have proposed an alternative parameterization scheme which yields more intuitive behavior, and by linking the co-ranking matrix to a point-wise version, which also offers meaningful local color values to visualize the quality of a given mapping. Note that it may be worthwhile to also study the behavior of the quality matrix $Q_{ND}(K_s, \kappa_t)$ depending on both pa-
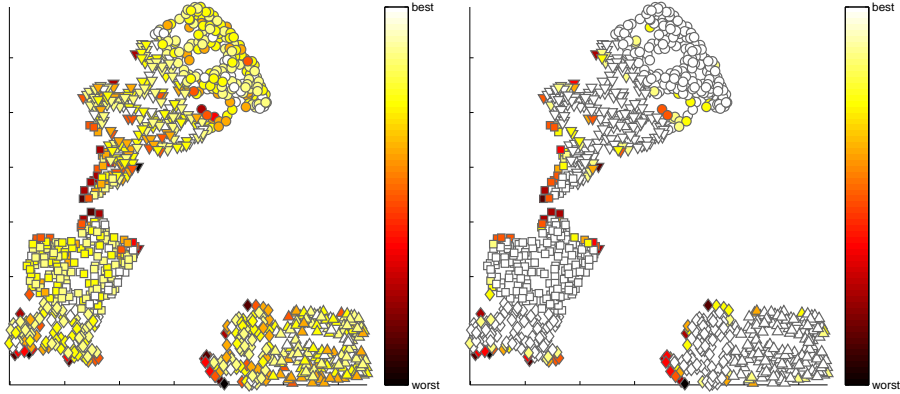
Fig. 4: A mapping of the *swiss roll* data by t-SNE (with a perplexity of 50), colored by point-wise qualities $Q_{NX}^i(14)$ (left) and $Q_{ND}^i(14)$ (right). The coloring clearly indicates the tearing of the original manifold, with less ambiguity in Q_{ND}^i on the right. (The sequence of class labels from the inside to the outside of the original spiral-shaped manifold is: $\circ \triangledown \square \diamond \triangle$.)

rameters, the relevant range K_s and the tolerated rank errors κ_t, since it allows users to interactively choose appropriate values according to a given application.

References

[1] M. Aupetit. Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing*, 70(7-9):1304 – 1330, 2007.

[2] H.-U. Bauer, K. Pawelzik, and T. Geisel. A topographic product for the optimization of self-organizing feature maps. In *NIPS*, pages 1141–1147, 1991.

[3] L. Chen and A. Buja. Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis. *Journal of the American Statistical Association*, 104(485):209–219, 2009.

[4] A.N. Gorban and A.Y. Zinovyev. Principal manifolds and graphs in practice: from molecular biology to dynamical systems. *Int. J. Neural Syst.*, 20(3):219–232, 2010.

[5] J.A. Lee and M. Verleysen. *Nonlinear dimensionality reduction*. Springer, 2007.

[6] J.A. Lee and M. Verleysen. Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomput.*, 72(7-9):1431–1443, 2009.

[7] J.A. Lee and M. Verleysen. Scale-independent quality criteria for dimensionality reduction. *Pattern Recognition Letters*, 31:2248–2257, October 2010.

[8] A. Ultsch and H.P. Siemon. Kohonen's self organizing feature maps for exploratory data analysis. In *Proceedings of INNC'90*, pages 305–308. Kluwer, 1990.

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

[10] J. Venna and S. Kaski. Local multidimensional scaling. *Neural Netw.*, 19:889–899, 2006.

[11] J. Venna, J. Peltonen, K. Nybo, H. Aidos, and S. Kaski. Information retrieval perspective to nonlinear dimensionality reduction for data visualization. *J. Mach. Learn. Res.*, 11:451–490, 2010.