Local multidimensional scaling with controlled tradeoff between trustworthiness and continuity

Jarkko Venna[1] **and Samuel Kaski**[2,1]
[1]Neural Networks Research Centre
Helsinki University of Technology
Espoo, Finland

[2]Department of Computer Science
University of Helsinki
Helsinki, Finland

{jarkko.venna, samuel.kaski}@hut.fi

Abstract - *In a visualization task, every nonlinear projection method needs to make a compromise between trustworthiness and continuity. In a trustworthy projection the visualized proximities hold in the original data as well, whereas a continuous projection visualizes all proximities of the original data. A multidimensional scaling method, curvilinear components analysis, is good at maximizing trustworthiness. We extend it to explicitly make a user-tunable parameterized compromise between trustworthiness and continuity.*

Key words - Information visualization, manifold extraction, multidimensional scaling, nonlinear dimensionality reduction

1 Introduction

In information visualization one of the main tasks is to reduce the dimensionality of data to two or three to visualize proximities within a data set. In general, it is not possible to reduce the dimensionality without losing some of the proximities in the process. Two kinds of errors can occur. First, data points originally farther away may enter the neighborhood of a sample in a projection. These errors decrease the trustworthiness of the visualization, as they create neighborhood relationships that are not present in the data. Second, data points that are originally in the neighborhood can be pushed farther away in the visualization. Because of the second type of errors, not all neighborhood relationships become visualized. Each dimensionality reduction method necessarily makes a tradeoff between these two kinds of errors. This setting is analogous to the precision—recall tradeoff in information retrieval. We have earlier [4] argued that trustworthiness is often more important since the visualized proximities are particularly salient. It would be even better to let the user decide about the compromise, and in this work we will extend a visualization method to make a parameterized compromise between trustworthiness and continuity. The method is a kind of a local multidimensional scaling method, curvilinear component analysis [3]. It aims at preserving pairwise distances but not all of them; only distances between points close-by on the visualization are

preserved. The formulation of neighborhoods in the projection plane shares some motivation with the Self-Organizing Map [5]. We call the new method *local multidimensional scaling*.
New methods for estimation of data manifolds of embeddings have been presented in recent years. So far, Isomap [2], Locally linear embedding (LLE) [8] and Laplacian Eigenmap [1], have not been compared in the task of *visualization* where the dimensionality of the representation is not selected based on the manifold but constrained by the display. We compare these methods with the curvilinear components analysis and the new local multidimensional scaling.

2 Methods

2.1 Measuring trustworthiness of a visualization

We consider a projection onto a display *trustworthy* if the set of k closest neighbors of a point on the display are also close by in the original space. We will use the following trustworthiness measure to compare the different visualization methods, and to quantify the compromise made by the new method. See [4, 10] for details.
Let N be the number of data samples and $r(i,j)$ be the rank of the data sample j in the ordering according to the distance from i in the original data space. Denote by $U_k(i)$ the set of those data samples that are in the neighborhood of size k of the sample i in the visualization display but not in the original data space. Our measure of trustworthiness of the visualization is

$$M_1(k) = 1 - \frac{2}{Nk(2N-3k-1)} \sum_{i=1}^{N} \sum_{j \in U_k(i)} (r(i,j) - k) \; . \tag{1}$$

The errors caused by discontinuities may be quantified analogously to the errors in trustworthiness. Let $V_k(i)$ be the set of those data samples that are in the neighborhood of the data sample i in the original space but not in the visualization, and let $\hat{r}(i,j)$ be the rank of the data sample j in the ordering according to the distance from i in the visualization display. The effects of discontinuities of the projection are measured by

$$M_2(k) = 1 - \frac{2}{Nk(2N-3k-1)} \sum_{i=1}^{N} \sum_{j \in V_k(i)} (\hat{r}(i,j) - k) \; . \tag{2}$$

The worst attainable values of both measures may, at least in principle, vary with k, and were estimated in the results (Fig. 1) with random projections and with random neighborhoods.

2.2 Curvilinear component analysis (CCA)

The starting point of CCA [3] is a random initialization of points $(\mathbf{y}_i)$ in the reduced-dimensional output space, and a pairwise distance matrix between the original data points $(\mathbf{x}_i)$. The cost function measures preservation of the original pairwise distances, weighted by a coefficient F that depends on the distance between the points in the *output space*:

$$E = \frac{1}{2} \sum_i \sum_{j \neq i} (d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2 F(d(\mathbf{y}_i, \mathbf{y}_j), \sigma_y) \; . \tag{3}$$

F is usually defined as an area of influence around a data point in the output space:

$$F(d(\mathbf{y}_i, \mathbf{y}_j), \sigma_y)) = \begin{cases} 1 & \text{if } d(\mathbf{y}_i, \mathbf{y}_j) \leq \sigma_y \\ 0 & \text{if } d(\mathbf{y}_i, \mathbf{y}_j) > \sigma_y \ . \end{cases} \tag{4}$$

The cost function is optimized using a form of stochastic gradient descent algorithm. In the beginning of optimization the radius of the area of influence, σ_y, is kept large enough to cover all or at least most of the data points. During the optimization it is slowly reduced to zero.

2.3 Controlling the tradeoff: Local MDS

We propose a new method, *local MDS*, which is a derivative of CCA with the ability to control the tradeoff between trustworthiness and preservation of original neighborhoods.
While the CCA cost function (3) penalizes errors in preserving distances for neighboring points in the output space, the basic idea of the extension is to add a term that penalizes errors in preserving distances for close-by points in the input space. The tradeoff between these two terms, tunable by a parameter λ, governs the tradeoff between trustworthiness and continuity. The cost function of local MDS is

$$\begin{aligned} E = & \\ \frac{1}{2}\sum_i \sum_{j \neq i} & [(1-\lambda)(d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2 F(d(\mathbf{y}_i, \mathbf{y}_j), \sigma_i) + \lambda(d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2 F(d(\mathbf{x}_i, \mathbf{x}_j), \sigma_i)] \ , \\ & = \frac{1}{2}\sum_i \sum_{j \neq i} (d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2 [(1-\lambda) F(d(\mathbf{y}_i, \mathbf{y}_j), \sigma_i) + \lambda F(d(\mathbf{x}_i, \mathbf{x}_j), \sigma_i)] \ , \end{aligned} \tag{5}$$

We optimize the cost function with the stochastic gradient decent introduced for CCA in [3]. During the optimization the radius of the area of influence, σ_i around data point i, is slowly brought down to the distance of the k:th nearest neighbor of the data point i in the original space. The results shown here were produced with $k = 20$. Setting $\lambda = 0$ results in a normal CCA projection (with the difference that the end radius of the area of influence σ_i is larger than zero and different for each data point; for CCA the end radius of each data point is customarily reduced to zero).
We also tested a radius of influence which was the same for each data point and was brought to zero at the end of optimization. The behavior was quite similar but a nonzero end neighborhood makes controlling of the compromise more robust.

3 Experiments

3.1 Data Sets

Thick S-curve. A simple data set having a folded lower-dimensional manifold, a two-dimensional S-shaped curve in a three-dimensional space, was constructed as follows. First, the data was uniformly sampled from a two-dimensional S-shaped sheet. Then, to give the manifold a thickness, a spherical normally distributed displacement was added to each point. The data set consists of 1000 data points.

Gene expression compendium. We used the large collection of human gene expression arrays collected by Segal et al. [9]. (The normalized expression compendium is available from http://dags.stanford.edu/cancer.)
For visualization we removed samples with missing values from the data. First we removed genes that were missing from more than 300 arrays. Then we removed the arrays that still contained missing values. This resulted in a data set containing 1278 arrays and 1339 genes (dimensions).
This is a very hard data set to visualize. The data is very high dimensional and there do not seem to be any low dimensional manifold structures that the methods could take advantage off.

Mouse gene expression. We additionally visualized a collection of gene expression profiles from different mouse tissues. For details of the data set and of preprocessing see [4].

3.2 Comparison of visualization methods

We compared the new manifold estimation methods mentioned in the Introduction with CCA and SOM in a visualization task.
The methods having a number of neighbors parameter k were run several times with values of k going from 4 to 20. CCA and SOM were run ten times on each data set. The SOM size was set such that the average number of data points in each unit was about 5. The SOM neighborhood was decreased to one during the optimization. In each case the result with the best trustworthiness was selected.
When trying to get insights on a data point a human analyst usually looks at a handful (say 10) data points around it. Thus it is very important that the visualization preserves small neighborhoods well, that is, that the visualization is trustworthy. It is clear from Fig. 1 that in terms of trustworthiness the SOM and CCA are the best methods on both of these data sets, with a clear difference from the other methods. On the Gene expression compendium the SOM is also the second best at preserving the original neighborhoods. CCA is the worst method in this respect. Based on these tests it seems that the new manifold extraction methods can have a hard time dealing with manifolds that have a higher dimensionality than the display. In both cases the trustworthiness is similar to that of the PCA.

3.3 Local MDS

The effect of varying λ is illustrated in Fig. 2 where trustworthiness and continuity (of a neighborhood of the size 10) are plotted as a function of λ. When λ is increased there is an overall tendency for trustworthiness to decrease and continuity of original neighborhoods to increase.
There is a point (usually at around $\lambda = 0.2 \ldots 0.5$) after which continuity of original neighborhoods may start to decrease. This happens because the second part of the cost function does not optimize continuity directly. If λ is too large, the unfolding effect of the first part of the cost function may not be enough to keep the projection from folding on itself. This is evident on the Fig. 2**c** where continuity first increases sharply and then starts to decline. Thus, based on empirical findings, we recommend that λ should be kept within the range [0, 0.5].

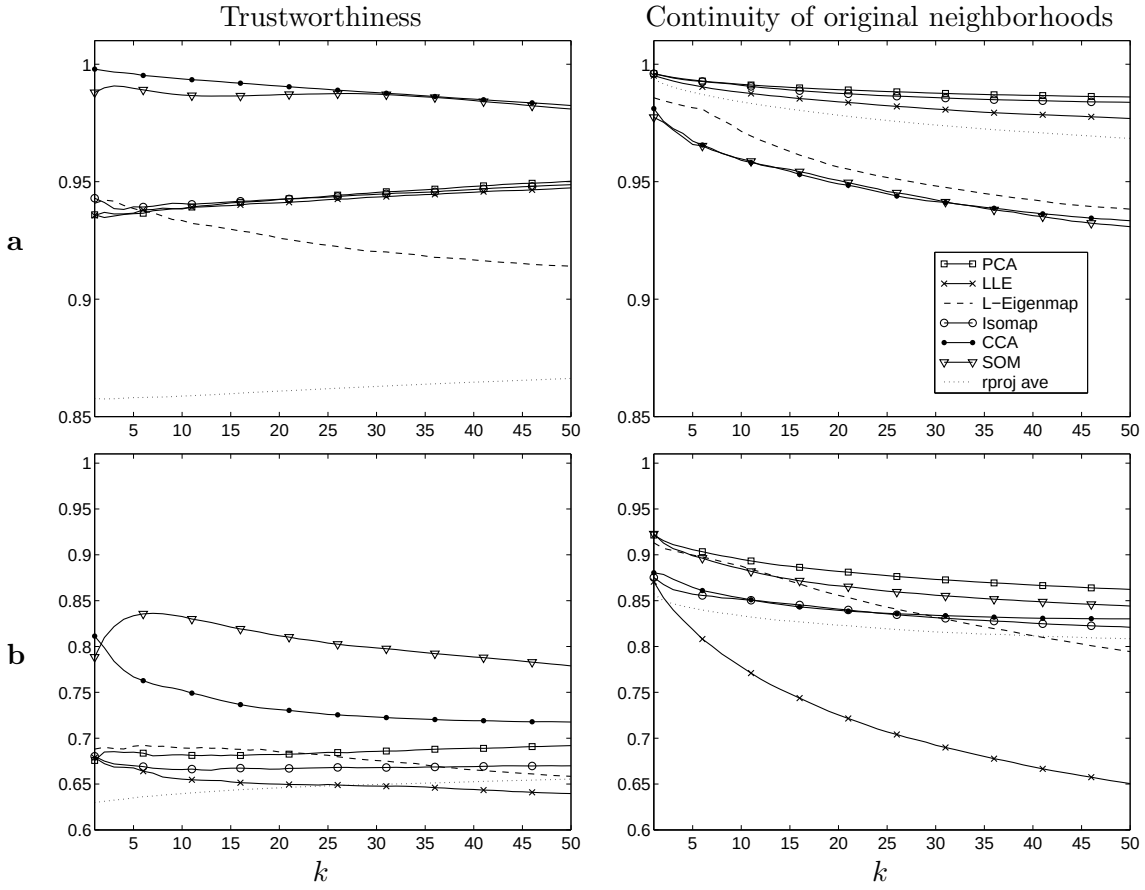

Figure 1: The change in trustworthiness and continuity of original neighborhoods as the number of neighbors k in the neighbor set is varied. Small neighborhoods are the most important ones. **a**) Thick S-curve manifold, **b**) Gene Expression Compendium. Rproj is the average value of 100 linear random projections. The trustworthiness and continuity values of random neighborhoods are approximately 0.5. PCA: Principal component analysis, LLE: Locally linear embedding, L-Eigenmap: Laplacian eigenmap, CCA: Curvilinear component analysis, SOM: Self-organizing map.

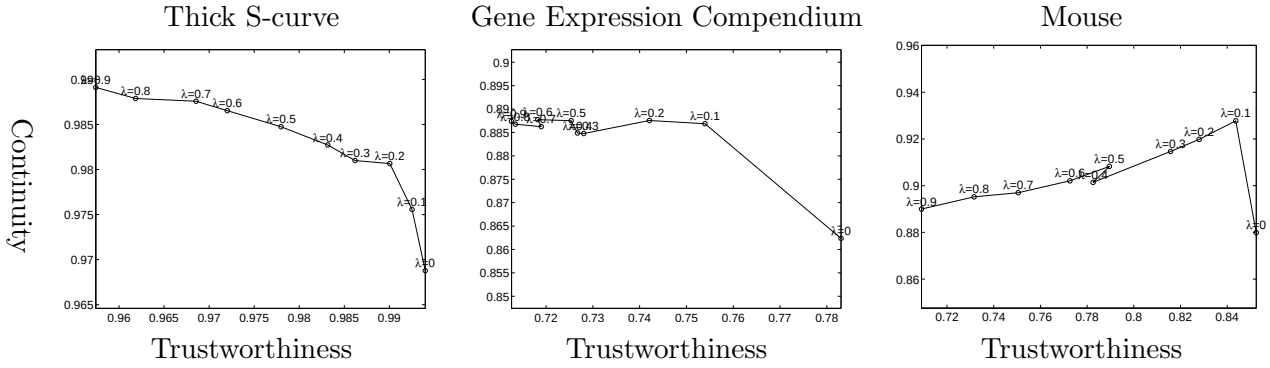

Figure 2: The relationship between trustworthiness and preservation of original neighborhoods as a function of λ, for a neighborhood of the size of 10.

The overall performance of local MDS ranges from, and sometimes outperforms, that of CCA to that of PCA, the former being the best in terms of trustworthiness and the latter in terms of continuity.
Fig. 4 gives three examples of local MDS projections. A spherical data set is projected first with $\lambda = 0$ and then with $\lambda = 0.1$ and finally with $\lambda = 0.9$. When λ is zero the local MDS splits the sphere open, into roughly two discs. When λ is increased the edges where continuity is violated the worst get pulled closer together to minimize the number of neighborhoods that become split, and to reduce the distance between those neighborhoods that cannot be connected.

4 Discussion

An extension to CCA, curvilinear distance analysis (CDA) [6, 7], was recently introduced. The main idea of CDA is to replace the Euclidean distances in the original space with geodesic distances in the same manner as in the Isomap algorithm. The same change could also be done in local MDS. However it would have to be decided whether continuity is desired on the manifold or globally. This would affect whether the second term of the local MDS cost function should be based on geodesic or Euclidean distances.

5 Conclusions

We tested several different nonlinear dimensionality reduction methods. Of these, Isomap, Laplacian Eigenmap, and LLE are designed to extract manifolds while CCA and SOM are more generally targeted for dimensionality reduction. One of the main tasks that these methods are used for is visualization. Thus it is important to understand how they perform in typical visualization situations and what kinds of tradeoffs they make. Of the methods tested here only SOM and CCA can be recommended for general visualization tasks where high trustworthiness is required. If preservation of original neighborhoods is required the linear method PCA is a good first choice.
We introduced an extension of CCA called local MDS, that according to the preliminary results is capable of controlling the tradeoff between trustworthiness and continuity of the projection.

References

[1] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding an d clustering. In Thomas G. Dietterich, Suzanna Becker, and Zoubin Ghahramani, editors, *Advances in Neural Information Processing Systems 14 (NIPS'2001*, pages 585–591, Cambridge, MA, 2002. MIT Press.

[2] Joshua B.Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, December 2000.

[3] Pierre Demartines and Jeanny Hérault. Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. *IEEE Transactions on Neural Networks*, 8(1):148–154, January 1997.

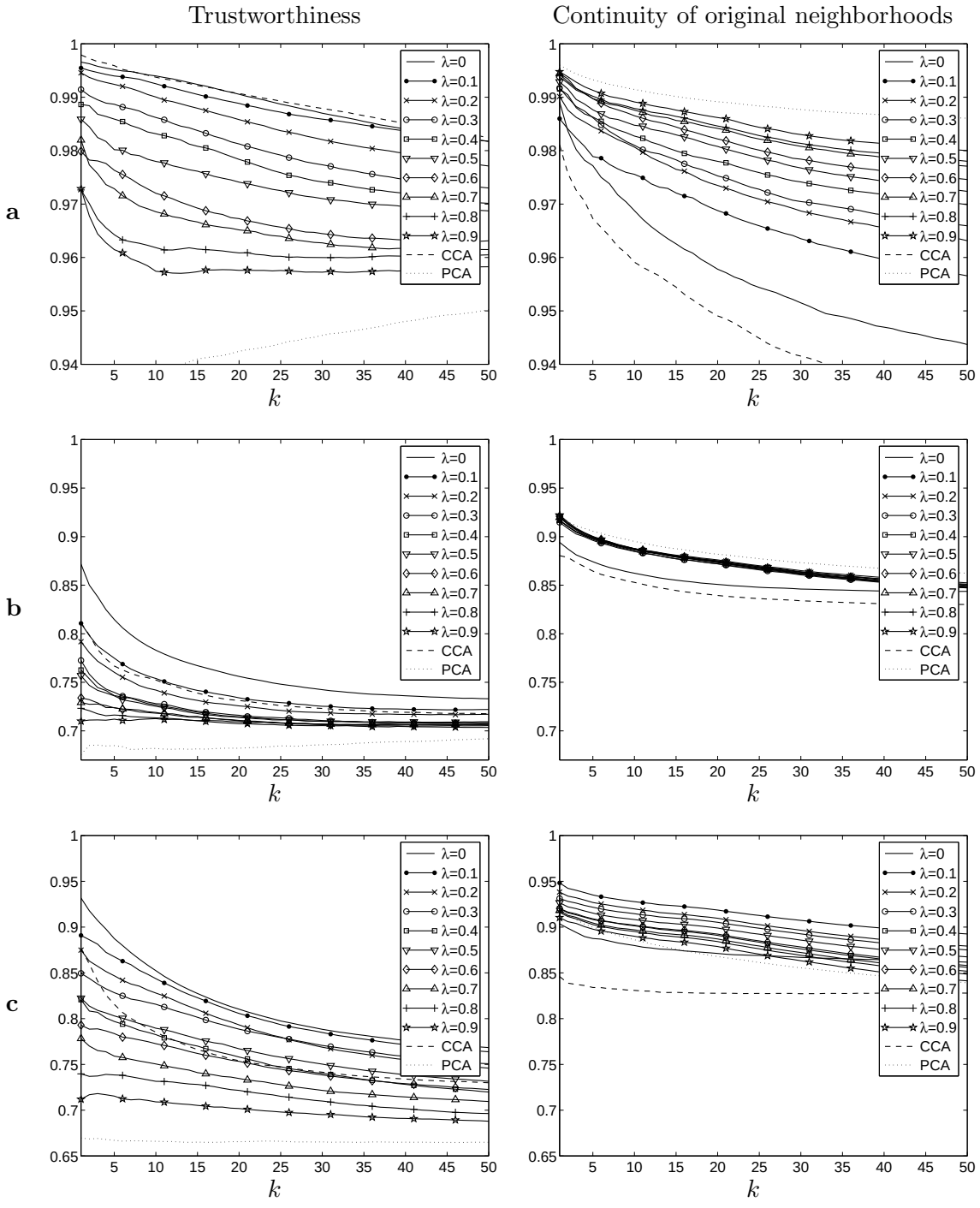

Figure 3: Trustworthiness and preservation of original neighborhoods of a local MDS projection as a function of λ. **a**) Thick S-curve manifold, **b**) Gene expression compendium, **c**) Mouse gene expression. Results from Principal component analysis (PCA) and Curvilinear component analysis (CCA) are included for reference.

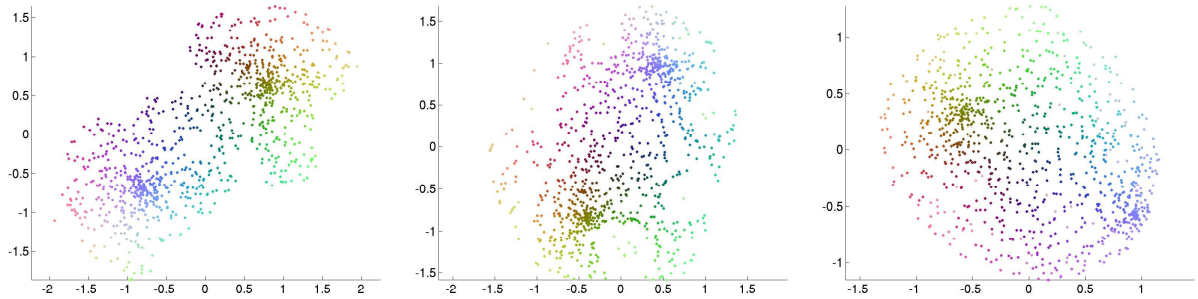

Figure 4: Three projections of a three-dimensional spherical cell with local MDS. On the left, trustworthiness of the projection is maximized by selecting $\lambda = 0$. In the middle and right, discontinuity of the projection is penalized as well, by setting $\lambda = 0.1$ and $\lambda = 0.9$, respectively.

[4] Samuel Kaski, Janne Nikkilä, Merja Oja, Jarkko Venna, Petri Törönen, and Eero Castrén. Trustworthiness and metrics in visualizing similarity of gene expression. *BMC Bioinformatics*, 4:48, 2003.

[5] Teuvo Kohonen. *Self-Organizing Maps*. Springer, Berlin, 3rd edition, 2001.

[6] John Aldo Lee, Amaury Lendasse, , Nicolas Donckers, and Michel Verleysen. A robust nonlinear projection method. In M. Verleysen, editor, *ESANN'2000, Eighth European Symposium on Artificial Neural Networks*, pages 13–20, Bruges, Belgium, 2000. D-Facto Publications.

[7] John Aldo Lee, Amaury Lendasse, and Michel Verleysen. Nonlinear projection with curvilinear distances: Isomap versus curvilinear distance analysis. *Neurocomputing*, 57:49–76, Mar 2004.

[8] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, December 2000.

[9] Eran Segal, Nir Friedman Amd Daphne Koller, and Aviv Regev. A module map showing conditional activity of expression modules in cancer. *Nature genetics*, 36(10):1090–1098, 2004.

[10] Jarkko Venna and Samuel Kaski. Neighborhood preservation in nonlinear projection methods: An experimental study. In G. Dorffner, H. Bischof, and K. Hornik, editors, *Proceedings of ICANN 2001, International Conference on Artificial Neural Networks*, pages 485–491, Berlin, 2001. Springer.