Reliable Visual Analytics with Dimensionality Reduction: Quality Evaluation and Interpretation of Projections

Part 1: Quality Assessment of Dimensionality Reduction

Hyeon Jeon[1], Takanori Fujiwara[2], Rafael M. Martins[3]

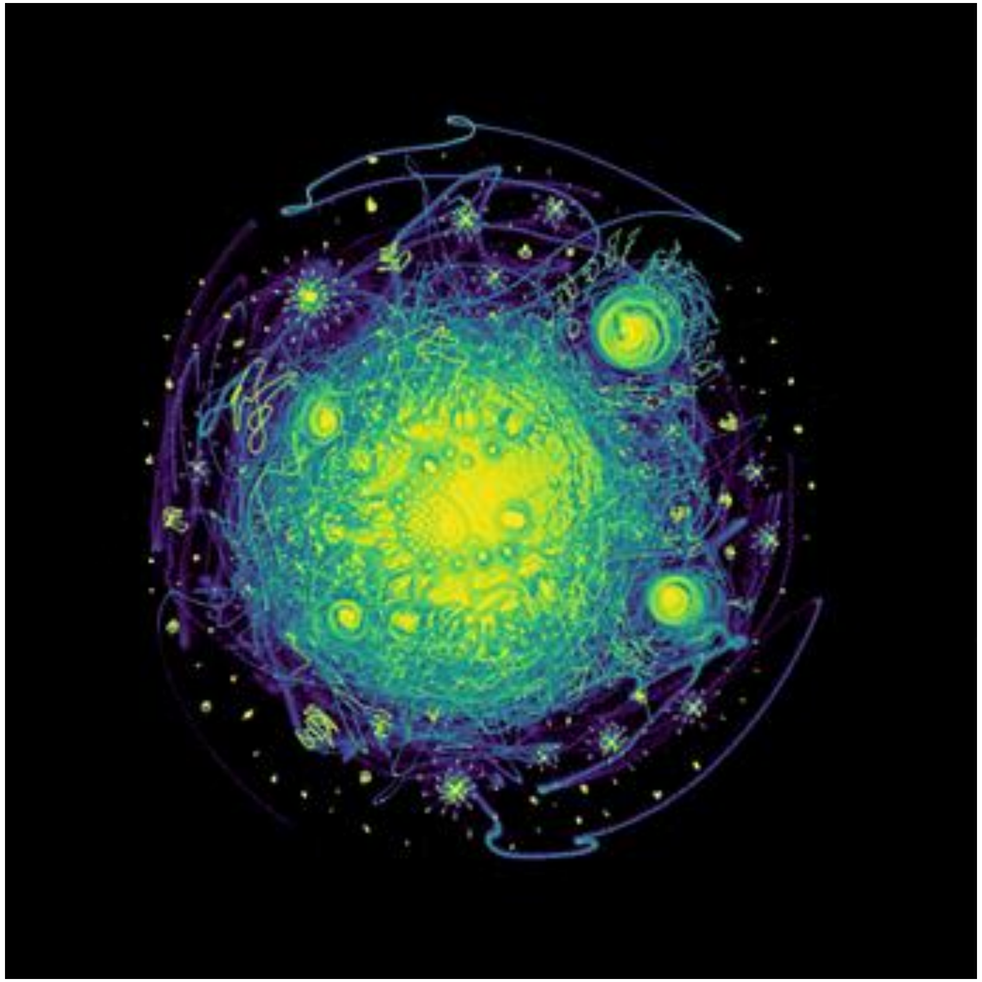
1 – Seoul National University
2 – Linköping University
3 – Linnaeus University

EuroVis 2025 Tutorial

EUROVIS 2025
LUXEMBOURG

Agenda

- Dimensionality Reduction - Overview
  - PCA, MDS
  - Modern nonlinear DR:
    - t-SNE, UMAP
- Quality Assessment
  - Distortion types
  - Quality metrics
  - Visualizing quality metrics



From: https://towardsdatascience.com/how-exactly-umap-works-13e3040e1668

Dimensionality Reduction

Modern, non-linear

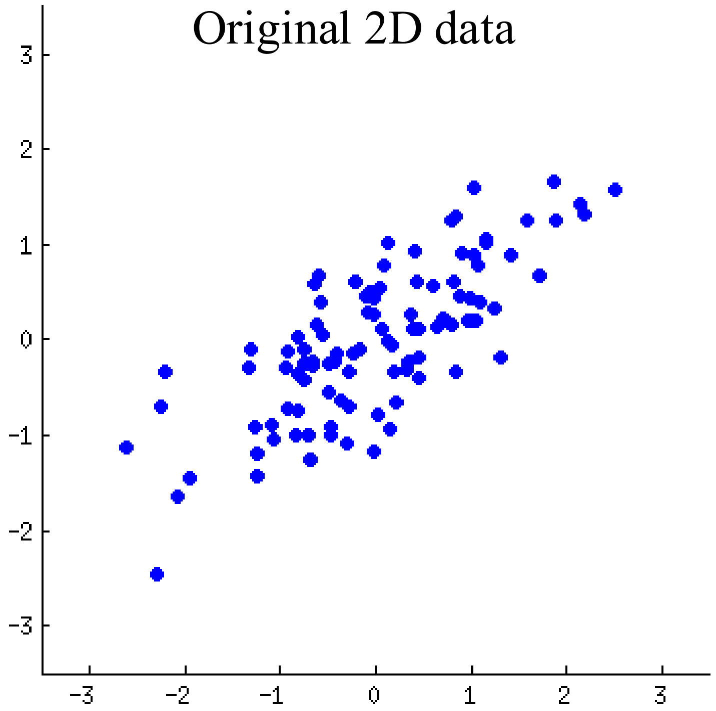
Dimensionality Reduction

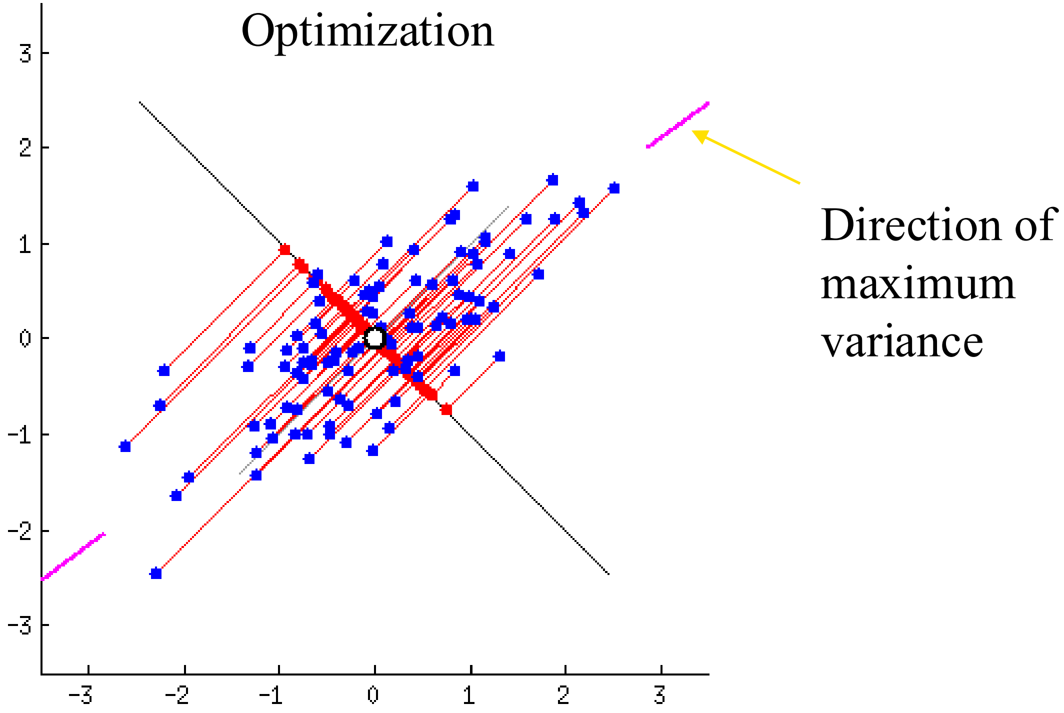
- For a simple abstraction, think of DR as a function:

- The input is a matrix of n elements (rows) by p features (columns)
- The output has the same number of rows (n), but q features ($q \ll p$)

Principal Components Analysis (PCA)

- Goal: **Maximize the explained variance**.



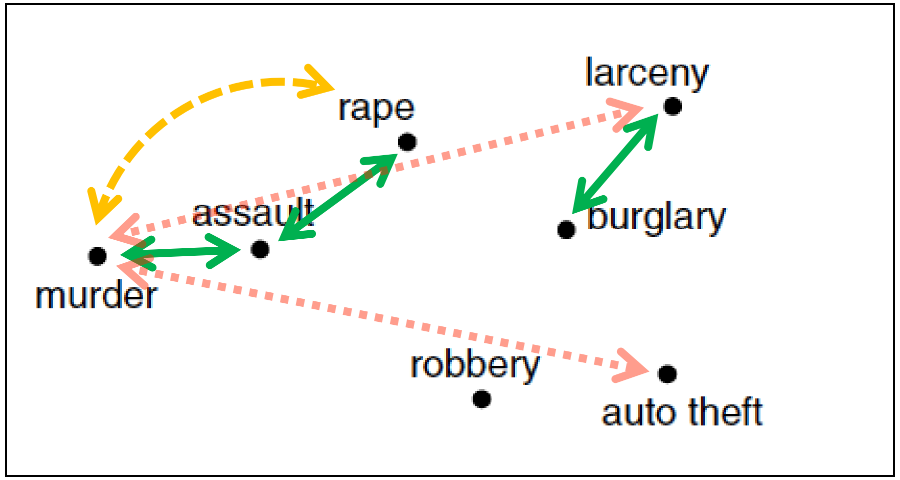


https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues

Multidimensional Scaling (MDS)

Correlations between crime rates in the U.S. states

Crime	No.	1	2	3	4	5	6	7
Murder	1	1.00	0.52	0.34	0.81	0.28	0.06	0.11
Rape	2	0.52	1.00	0.55	0.70	0.68	0.60	0.44
Robbery	3	0.34	0.55	1.00	0.56	0.62	0.44	0.62
Assault	4	0.81	0.70	0.56	1.00	0.52	0.32	0.33
Burglary	5	0.28	0.68	0.62	0.52	1.00	0.80	0.70
Larceny	6	0.06	0.60	0.44	0.32	0.80	1.00	0.55
Auto theft	7	0.11	0.44	0.62	0.33	0.70	0.55	1.00



Borg, I., & Groenen, P. J. (2005). Modern multidimensional scaling: Theory and applications. Springer Science & Business Media.

Borg, I. and Groenen, P. [2]

Multidimensional Scaling (MDS)

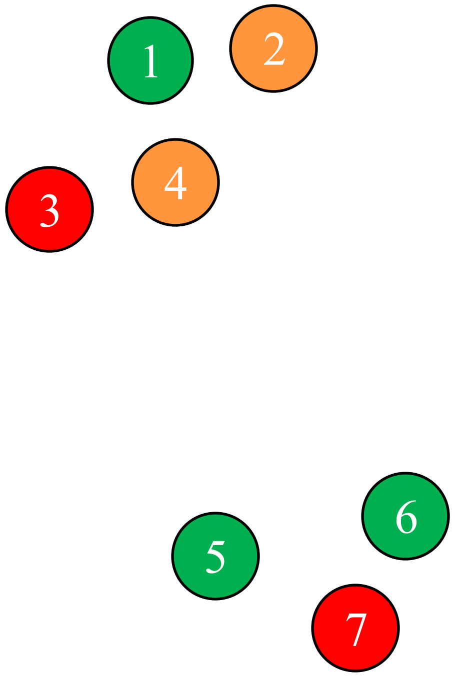
General procedure (using gradient descent):

1. Start with an initial configuration of $\mathbb{R}^q$ (random/PCA/orthogonal)
2. Evaluate a chosen *cost* function.
3. If the *cost* is low enough (or the maximum *number of iterations* was reached), stop.
4. Move each point slightly (i.e., according to the *learning rate*) towards the direction where the *cost* function is minimized.
5. Go to 3.

Multidimensional Scaling (MDS)

Example:

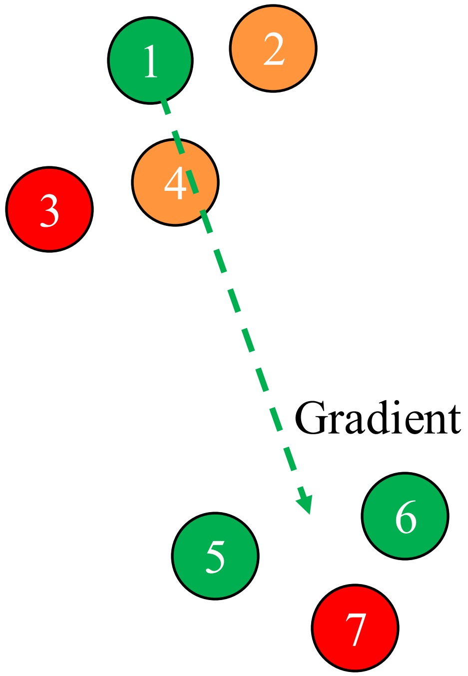
- $\delta_{ij} = color\ similarities$
- Random initial configuration in $\mathbb{R}^2$
- Each iteration moves each point slightly in the direction of a lower cost
- The order in which the points are moved is arbitrary



Multidimensional Scaling (MDS)

Example:

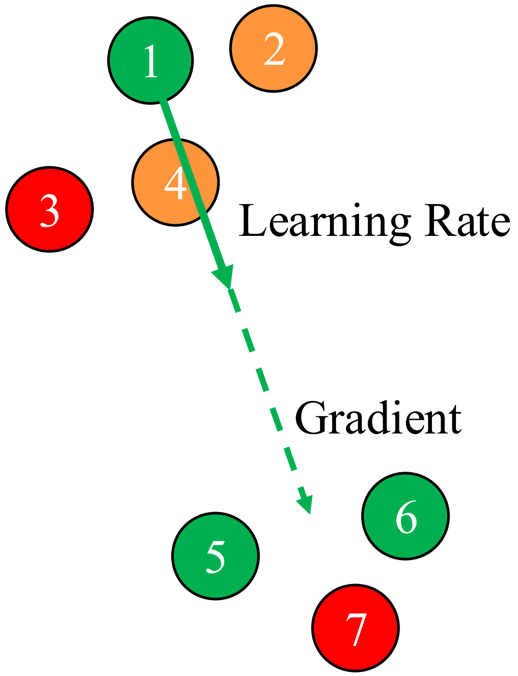
- $\delta_{ij} = color\ similarities$
- Random initial configuration in $\mathbb{R}^2$
- Each iteration moves each point slightly in the direction of a lower cost
- The order in which the points are moved is arbitrary



Multidimensional Scaling (MDS)

Example:

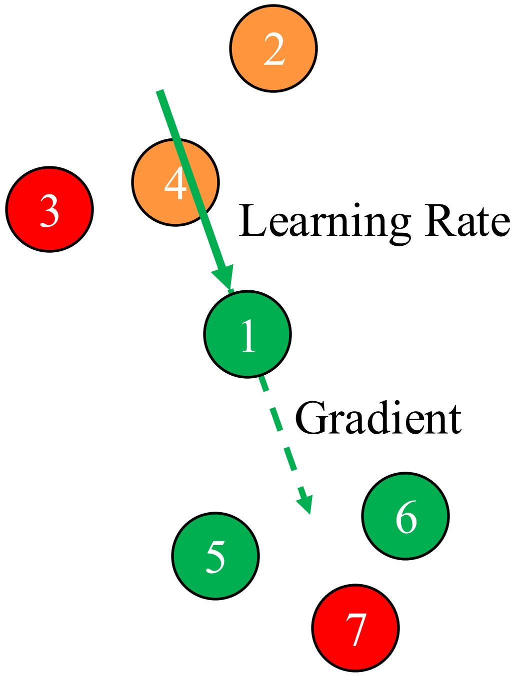
- $\delta_{ij} = color\ similarities$
- Random initial configuration in $\mathbb{R}^2$
- Each iteration moves each point slightly in the direction of a lower cost
- The order in which the points are moved is arbitrary



Multidimensional Scaling (MDS)

Example:

- $\delta_{ij} = color\ similarities$
- Random initial configuration in $\mathbb{R}^2$
- Each iteration moves each point slightly in the direction of a lower cost
- The order in which the points are moved is arbitrary

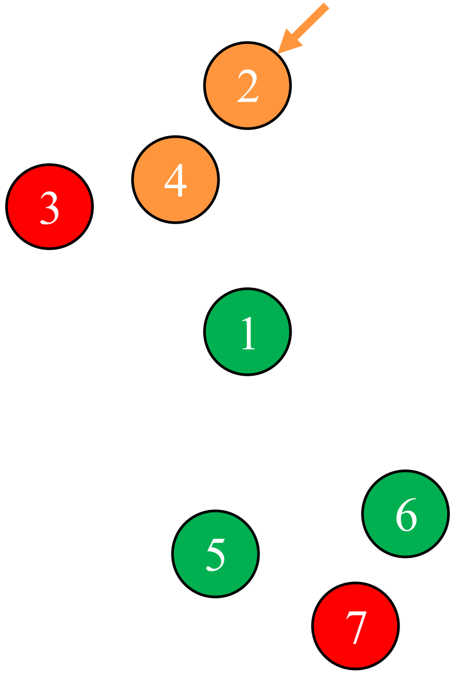


Multidimensional Scaling (MDS)

Example:

Note: in every step, there are both attractive and repulsive forces acting on each point.

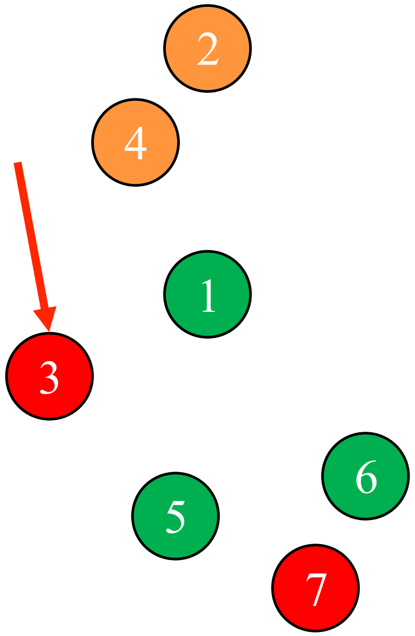
- $\delta_{ij} = color\ similarities$
- Random initial configuration in $\mathbb{R}^2$
- Each iteration moves each point slightly in the direction of a lower cost
- The order in which the points are moved is arbitrary



Multidimensional Scaling (MDS)

Example:

- $\delta_{ij} = color\ similarities$
- Random initial configuration in $\mathbb{R}^2$
- Each iteration moves each point slightly in the direction of a lower cost
- The order in which the points are moved is arbitrary



And so on…

t-SNE (t-Dist. Stochastic Neighbor Embedding)

Links:

- https://observablehq.com/@robert-browning/t-sne-t-distributed-stochastic-neighbor-embedding

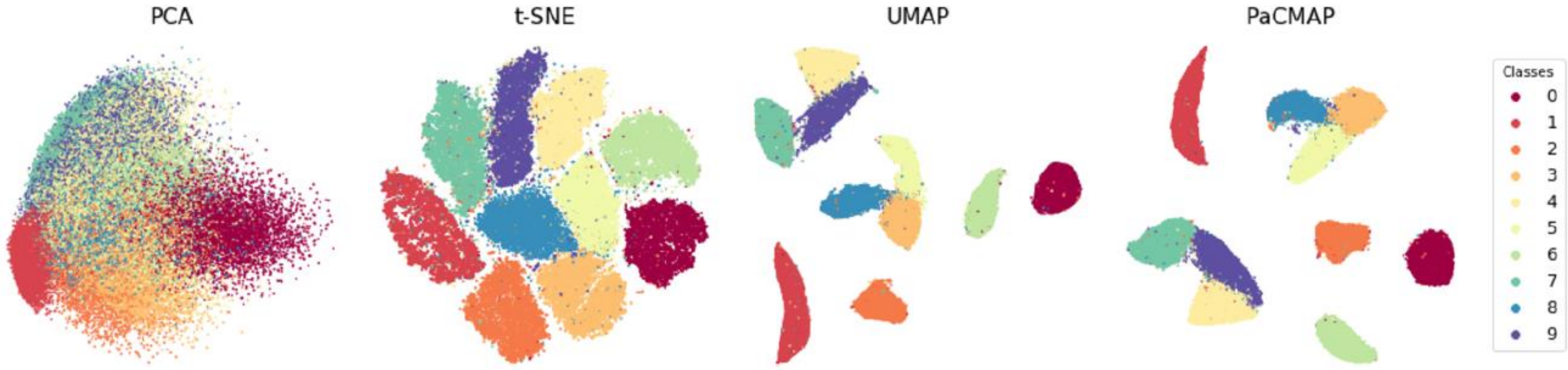
- https://distill.pub/2016/misread-tsne/

UMAP (Uniform Manifold Approx. and Proj.)

Links:

- https://pair-code.github.io/understanding-umap/
- https://umap-learn.readthedocs.io/en/latest/interactive_viz.html

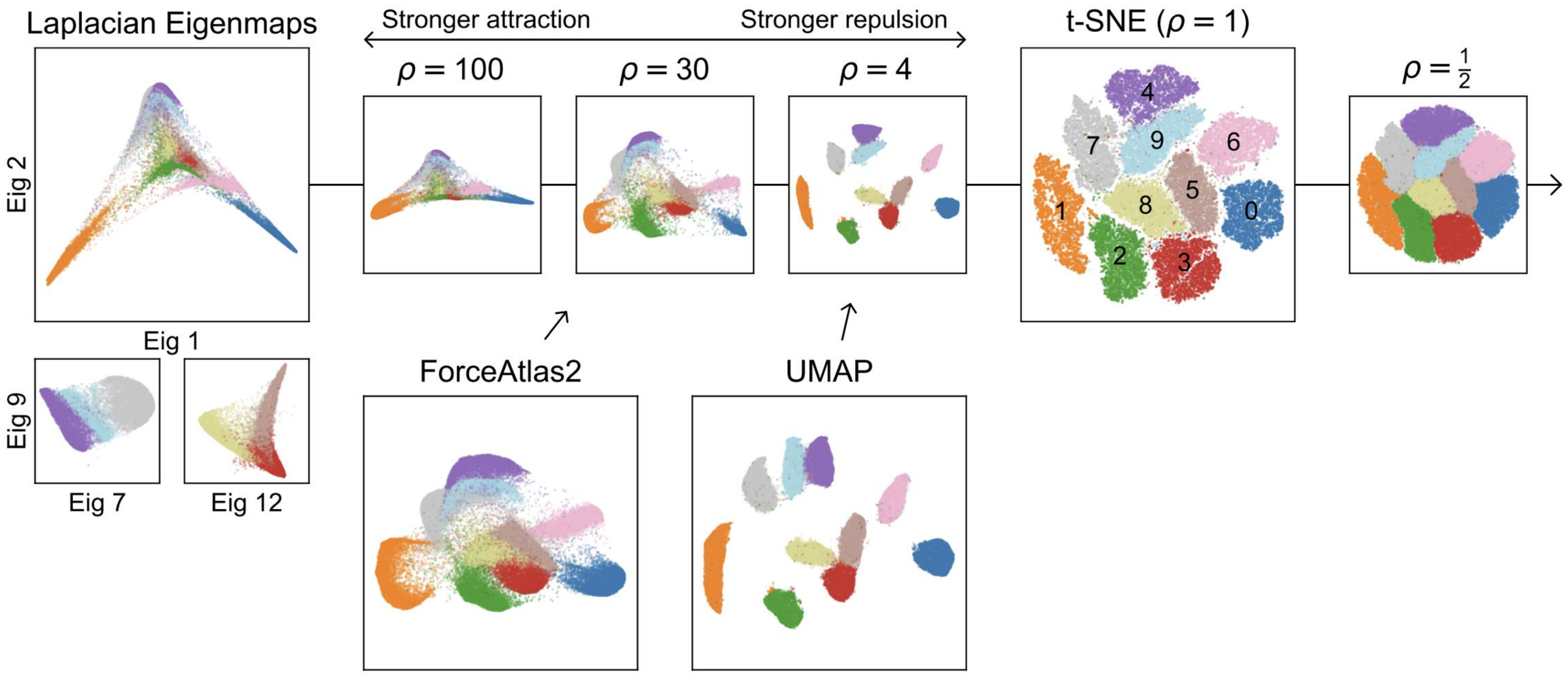
Modern DR: Comparison



MNIST

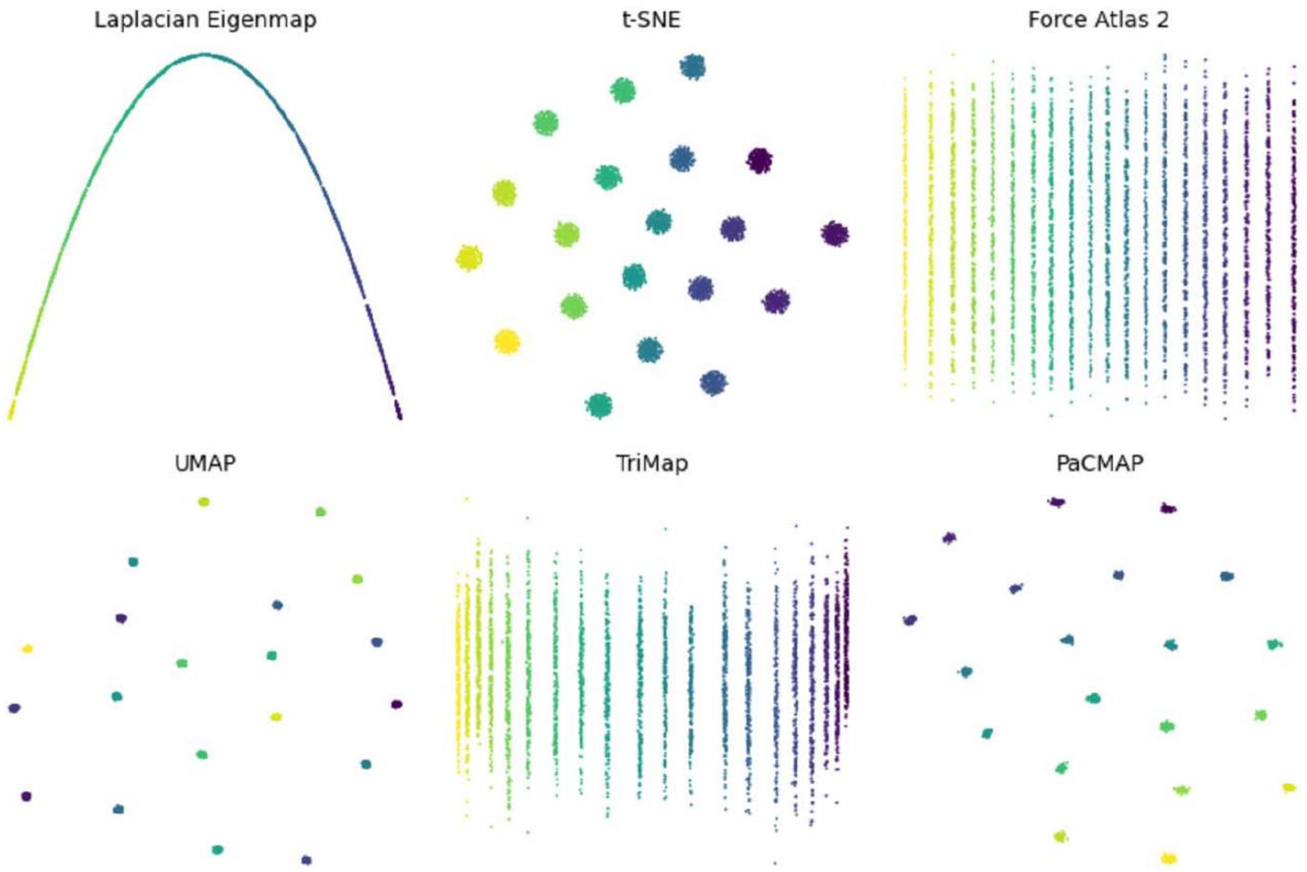
Huang, H., Wang, Y., Rudin, C., & Browne, E. P. (2022). Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization. Communications biology, 5(1), 719.

Modern DR: Comparison



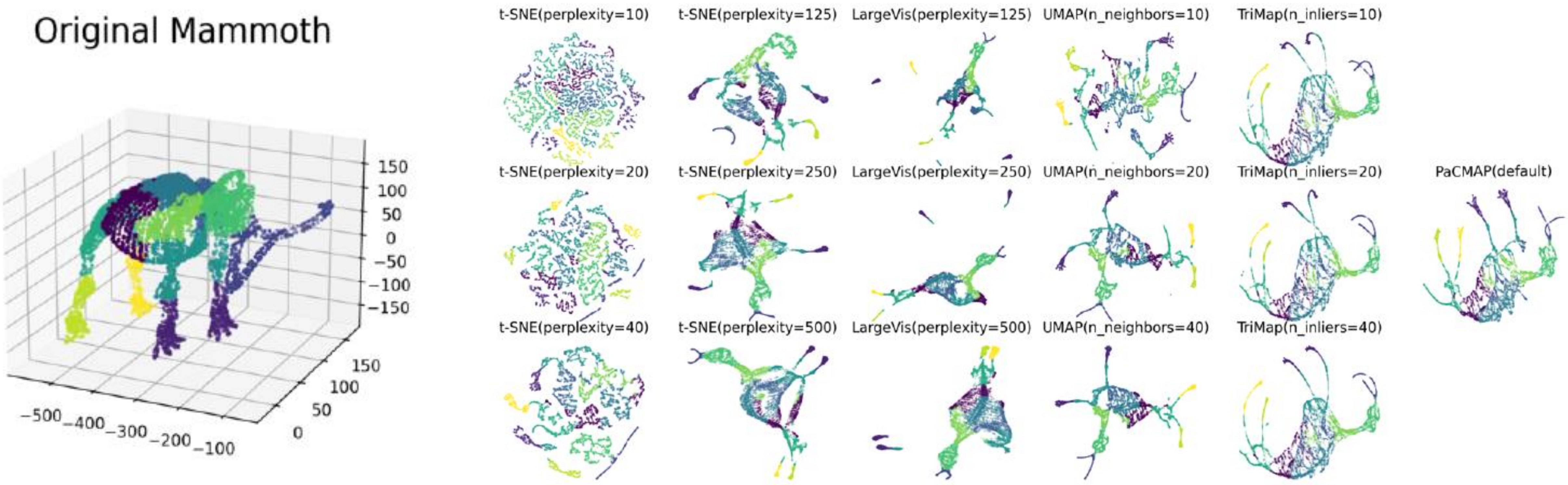
Böhm, J. N., Berens, P., & Kobak, D. (2022). Attraction-repulsion spectrum in neighbor embeddings. Journal of Machine Learning Research, 23(95), 1-32.

Modern DR: Comparison



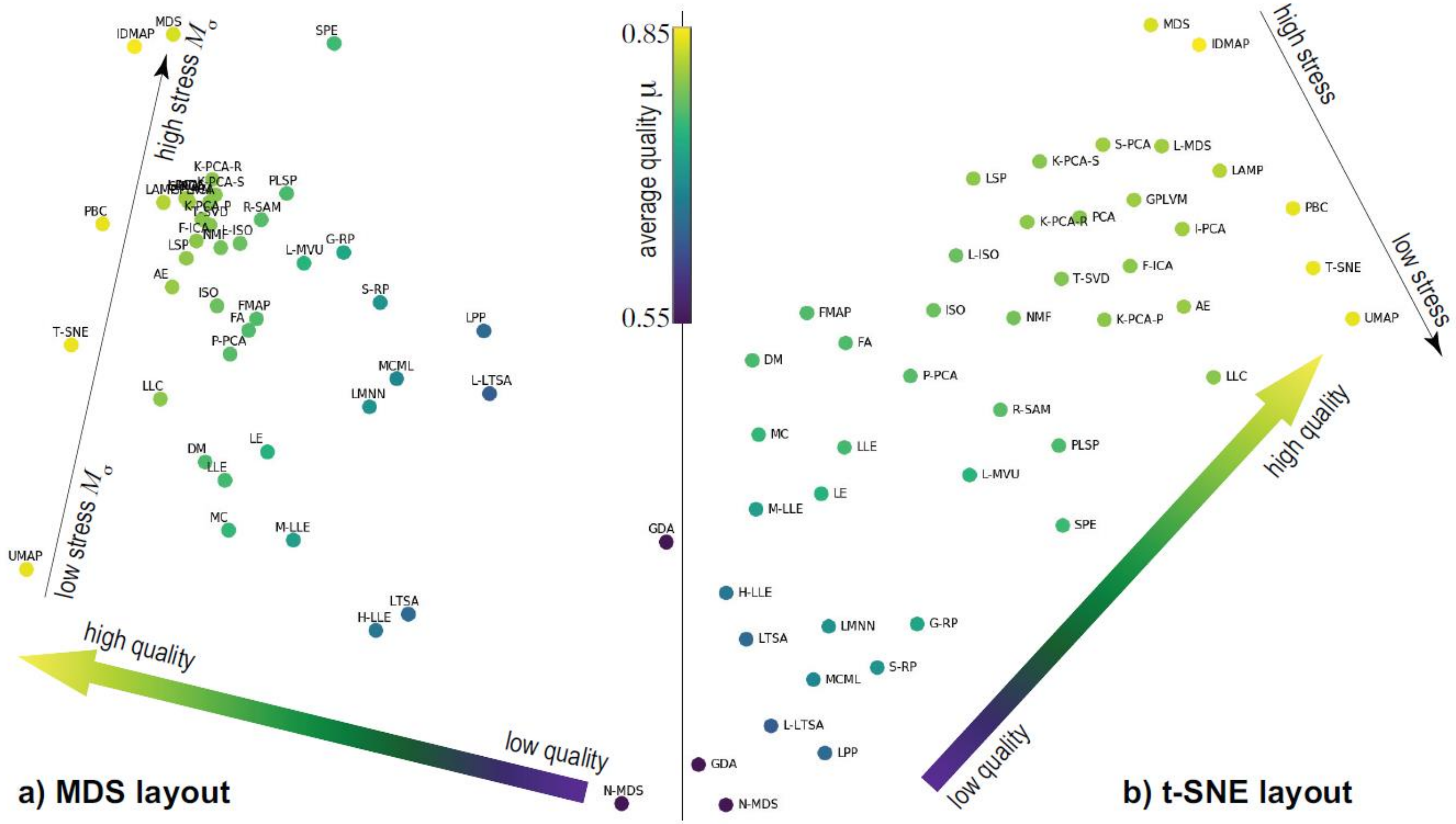
Huang, H., Wang, Y., Rudin, C., & Browne, E. P. (2022). Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization. Communications biology, 5(1), 719.

Modern DR: Comparison



Huang, H., Wang, Y., Rudin, C., & Browne, E. P. (2022). Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization. Communications biology, 5(1), 719.

Modern DR: Comparison



Espadoto, M., Martins, R. M., Kerren, A., Hirata, N. S., & Telea, A. C. (2019). Toward a quantitative survey of dimension reduction techniques. IEEE transactions on visualization and computer graphics, 27(3), 2153-2173.

Linnæus University 서울대학교 SEOUL NATIONAL UNIVERSITY LINKÖPING UNIVERSITY

“All models are wrong, but some are useful.”

Box, 1978

“All models are wrong, but some are useful.”

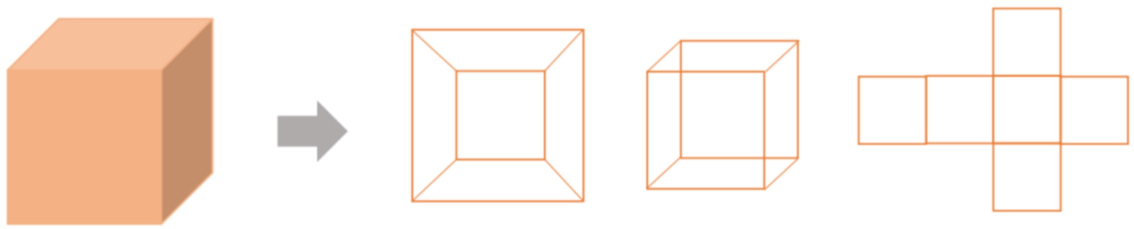
Box, 1978

“...with the wrong DR method, information about the high-dimensional relationships between points can be lost when projecting onto a 2D or 3D space.”

Rudin et al., 2022

Distortions in dimensionality reduction

- Dimensionality reduction algorithms depict a portion of complex HD features
- Different algorithms represents different portions
  - i.e., examines HD data in different perspectives
- e.g., even a 3D cube cannot be exactly projected in 2D space!!



Distortions in dimensionality reduction

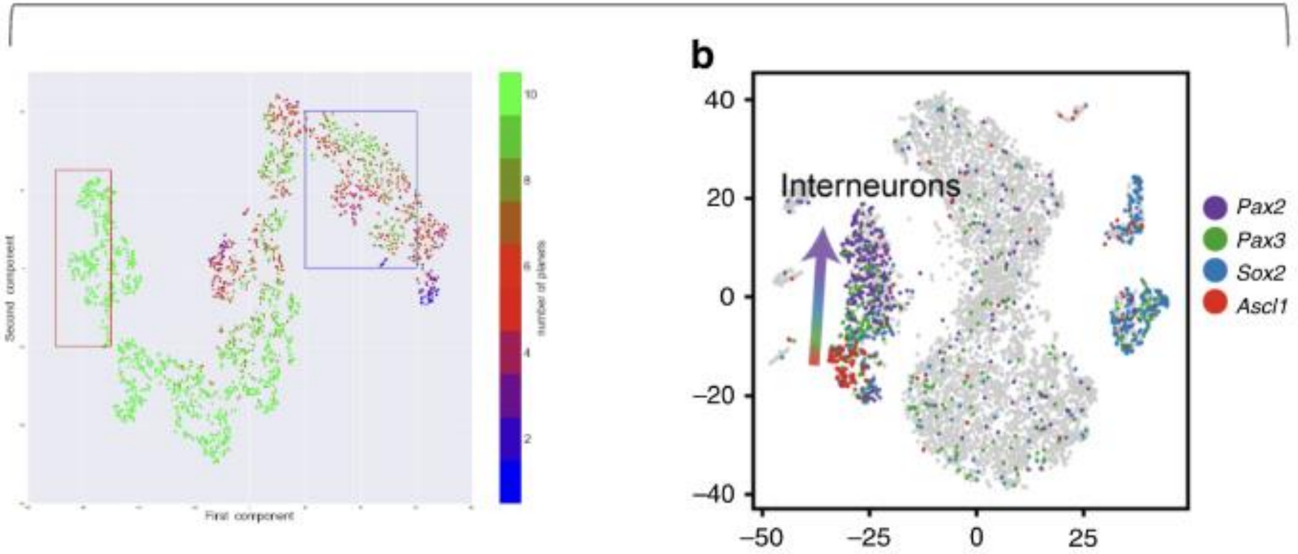
- High-dimensional data is extremely complex and intertwined
- Distortion inherently occurs while reducing dimensionality
- LD embeddings may not accurately depict the features of original HD data

- May degrade the credibility of visual analysis based on dimensionality reduction

Be aware of distortions!!

- Practitioners often disregard such threat

(c) **Global distances** between clusters are interpreted as their actual high-dimensional distances, casting doubt on the **credibility** of the interpretations



Alibert et al. Ocasio et al.

Cashman, Dylan, et al. "A critical analysis of the usage of dimensionality reduction in four domains." IEEE Transactions on Visualization and Computer Graphics (2025).

Quality Assessment

- How can we determine the quality of a projection?
  - How about… *the remaining error?*

$$C = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}$$

(t-SNE)

- Examples:
  - t-SNE: KLD
  - UMAP: Cross-entropy

$$CE(X,Y) = \sum_i \sum_j \left[p_{ij}(X) \log\left(\frac{p_{ij}(X)}{q_{ij}(Y)}\right) + (1 - p_{ij}(X)) \log\left(\frac{1 - p_{ij}(X)}{1 - q_{ij}(Y)}\right) \right]$$

(UMAP)

- Remember: entirely unsupervised.

Quality Assessment

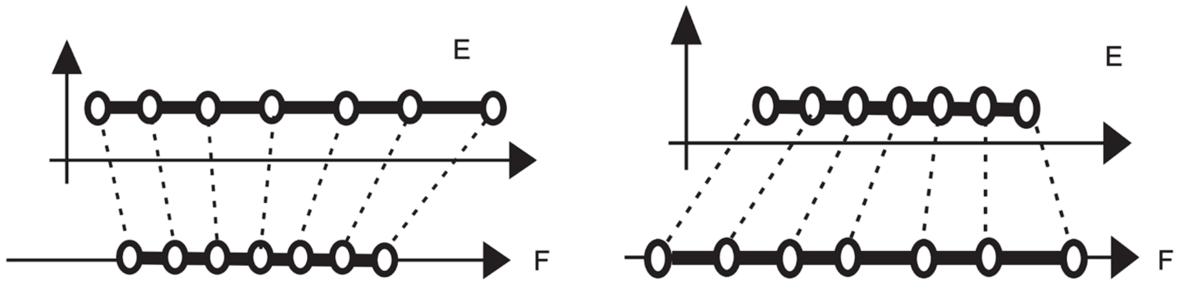
- How can we determine the quality of a projection?
  - How about… *the remaining error?*
- Examples:
  - t-SNE: KLD
  - UMAP: Cross-entropy
- Remember: entirely unsupervised.
- Problems:
  - Not very meaningful / interpretable
  - Not comparable between methods
- Solution:
  - Propose *method-independent* metrics

Types of Distortions

- HD space is complex; cannot be explained in a single perspective
- There exists various ways to “explain” or “define” distortions
  - Stretching/Compression
  - Missing Neighbors/False Neighbors
  - Missing Groups/False Groups

Stretching/Compression

- Stretching
  - Distances between points became larger in the low-dimensional space compared to the high-dimensional space
- Compression
  - Distances between points became shorter in the low-dimensional space compared to the high-dimensional space



Compression Stretching

Missing Neighbors & False Neighbors

- Missing Neighbors
  - Neighbors in the original space are no longer neighbors in the embedding
- False Neighbors
  - Neighbors that can be seen in the embedding are actually not neighbors in the original space

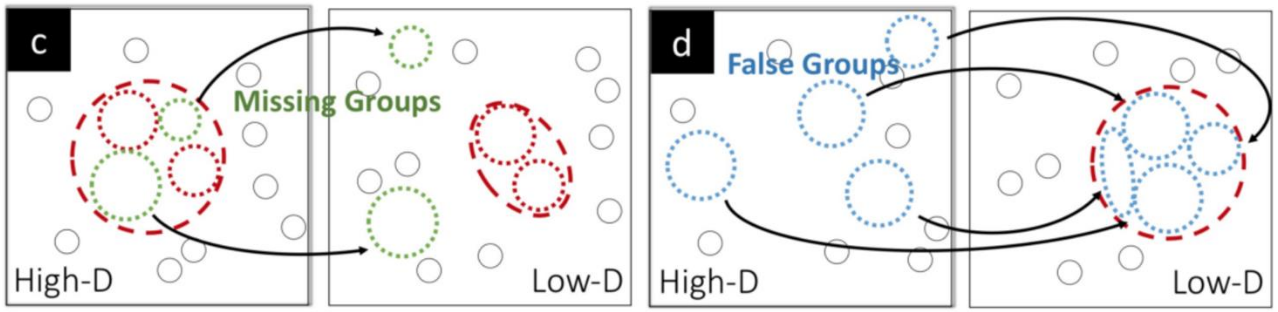
Missing Neighbors & False Neighbors

- Missing Neighbors
  - Neighbors in the original space are no longer neighbors in the embedding
- False Neighbors
  - Neighbors that can be seen in the embedding are actually not neighbors in the original space

- A seminal distortion type defined in the literature
- However, lacks the capability to explain complex cluster-level distortions
- “Extended” definition of distortions is needed...

Missing Groups & False Groups

- Missing Groups
  - A cluster in the original space is split into multiple subclusters in the embedding
- False Groups
  - A cluster that can be seen in the embedding actually consists of separated subclusters in the original space



Distortion types - Summary

- Stretching/Compression
  - Distortions in pairwise distances
- Missing/False Neighbors
  - Distortions in local neighborhood structure
- Missing/False Groups
  - Distortions in cluster structure

Quality Metrics and Distortions

- Stretching/Compression
  - Global metrics
- Missing/False Neighbors
  - Local metrics
- Missing/False Groups
  - Cluster-level metrics

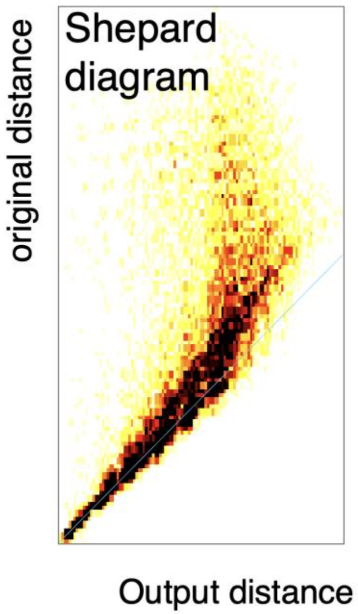
Global Metrics

- Examines the extent to which pairwise distances of the original high-dimensional data are distorted in the low-dimensional space

- DTM, Shepard index, Stress/Strain…

HD pairwise distances LD pairwise distances

$$Stress = \sqrt{\sum_{i=1,j=1}^{N} \frac{(\delta(x_i,x_j) - \delta(y_i,y_j))^2}{\delta(x_i,x_j)^2}}.$$



Lespinats, Sylvain, and Michaël Aupetit. "CheckViz: Sanity Check and Topological Clues for Linear and Non-Linear Mappings." Computer Graphics Forum. Vol. 30. No. 1. Oxford, UK: Blackwell Publishing Ltd, 2011.

Local Metrics

- Trustworthiness & Continuity (T&C)
- Mean relative rank errors (MRRE)
- Local Continuity Meta Criterion (LCMC)

- The most common type of distortion metrics
- Widely used in literature

Trustworthiness and *Continuity*

- Rank-based metrics
  - Don't consider distances, only ranks
- Hyperparameter: neighborhood size

- These are probably the most used metrics nowadays

$$M_T(K) = 1 - \frac{2}{G_K}\sum_{i=1}^{N}\sum_{j\in n_i^K \setminus v_i^K}(\rho_{ij} - K) = 1 - \frac{2}{G_K}\sum_{(k,l)\in \mathbb{LL}_K}(k-K)q_{kl},$$

rank of j as a neighbor of i in the input space

input-space ranks k>K and output-space ranks l<=K

$$M_C(K) = 1 - \frac{2}{G_K}\sum_{i=1}^{N}\sum_{j\in v_i^K \setminus n_i^K}(r_{ij} - K) = 1 - \frac{2}{G_K}\sum_{(k,l)\in \mathbb{UR}_K}(l-K)q_{kl},$$

rank of j as a neighbor of i in the output space

input-space ranks k<=K and output-space ranks l>K

$$G_K = \begin{cases} NK(2N-3K-1) & \text{if } K < N/2, \\ N(N-K)(N-K-1) & \text{if } K \geqslant N/2 \end{cases}$$

$n_i^K \setminus v_i^K$

points j that are neighbors of i in the output space but not in the input space

$v_i^K \setminus n_i^K$

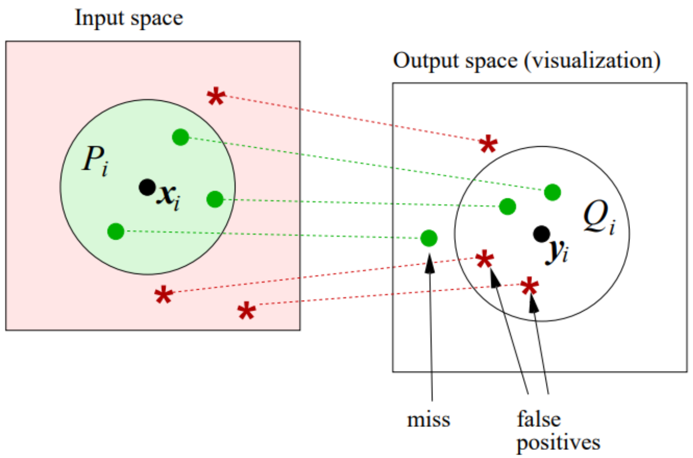
points j that are neighbors of i in the input space but not in the output space

Venna, J., Peltonen, J., Nybo, K., Aidos, H., & Kaski, S. (2010). Information retrieval perspective to nonlinear dimensionality reduction for data visualization. Journal of Machine Learning Research, 11(2).

Local Metrics

- Common workflow
  - Find k-Nearest Neighbor of each point in the HD space A
  - Find k-Nearest Neighbor of each point in the LD space B
- Check the difference between A and B

- k NN in HD but not in LD à Missing Neighbors
- k NN in LD but not in HD à False Neighbors



Venna, J., Peltonen, J., Nybo, K., Aidos, H., & Kaski, S. (2010). Information retrieval perspective to nonlinear dimensionality reduction for data visualization. Journal of Machine Learning Research, 11(2).

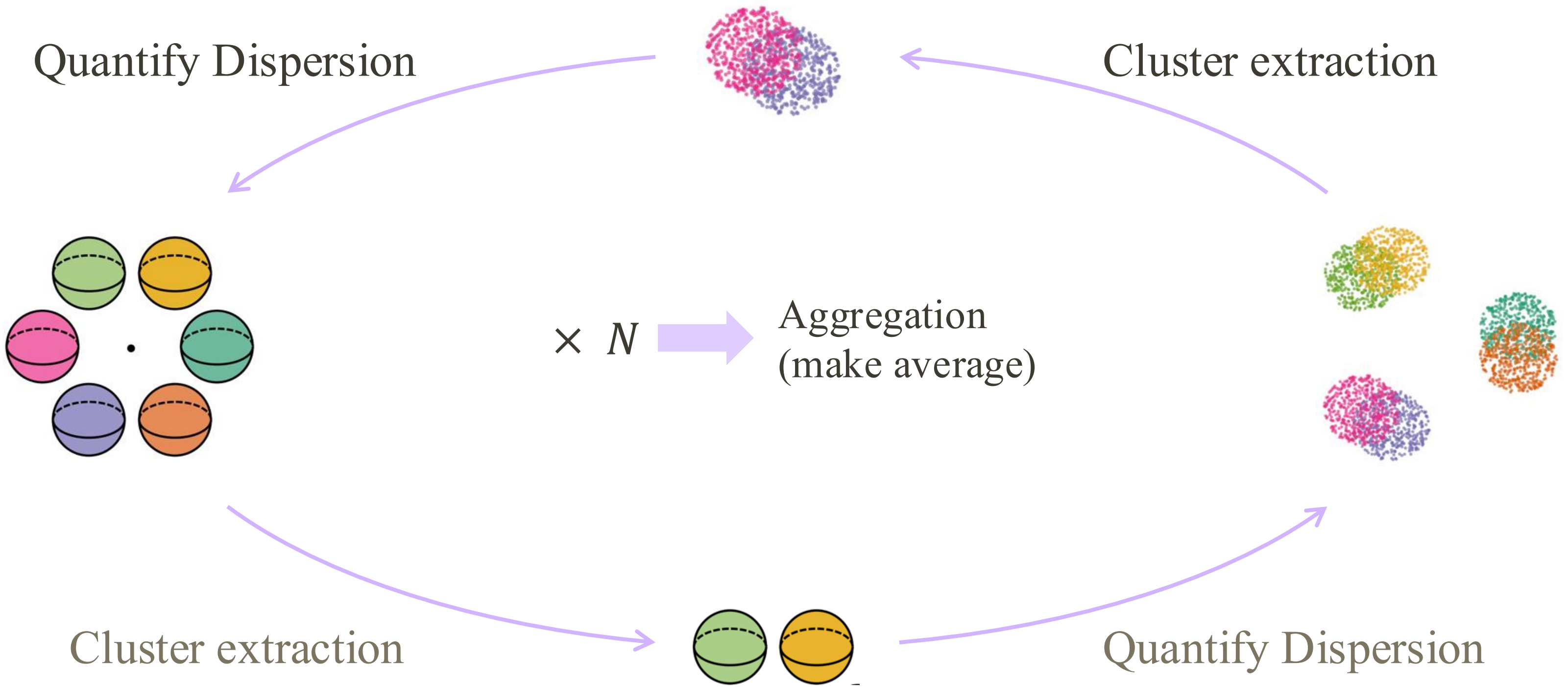
Cluster-Level Metrics

- Steadiness & Cohesiveness
- Label-Trustworthiness and Continuity

- Measures how well "clusters" in the high-dimensional space are depicted in low-dimensional projections as clusters, and vice versa

Cluster-Level Metrics

- Steadiness & Cohesiveness



Jeon, Hyeon, et al. "Measuring and explaining the inter-cluster reliability of multidimensional projections." IEEE Transactions on Visualization and Computer Graphics 28.1 (2021): 551-561.

Cluster-Level Metrics

- Label-Trustworthiness & Continuity

input

Class Labels

input

Quantify the degree of separability

score

Quantify the degree of separability

score

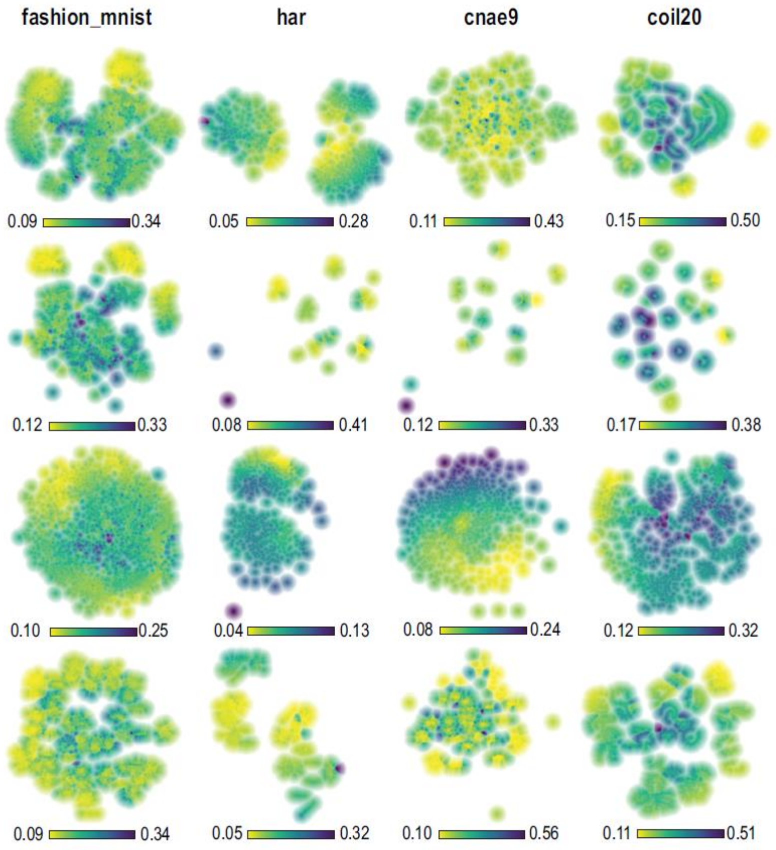
Examine the consistency

final score

Jeon, Hyeon, et al. "Classes are not clusters: Improving label-based evaluation of dimensionality reduction." IEEE Transactions on Visualization and Computer Graphics 30.1 (2023): 781-791.

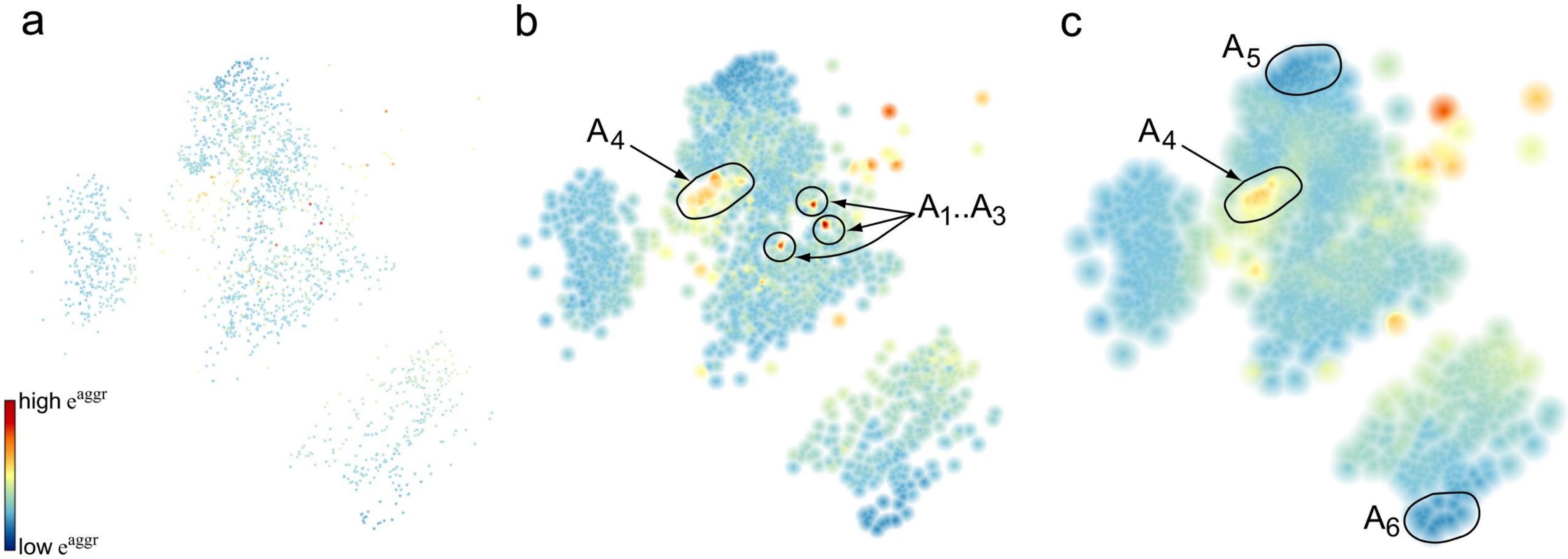
Quality Assessment: Visualization

- Metric scores are useful for comparison, but not very informative:
  - Are the errors spread out evenly around the projection?
- For that, we need to *visualize* them.
  - Identify trustworthy (and untrustworthy) areas of the layout
  - Guide the visual analysis



Espadoto, M., Martins, R. M., Kerren, A., Hirata, N. S., & Telea, A. C. (2019). Toward a quantitative survey of dimension reduction techniques. IEEE transactions on visualization and computer graphics, 27(3), 2153-2173.

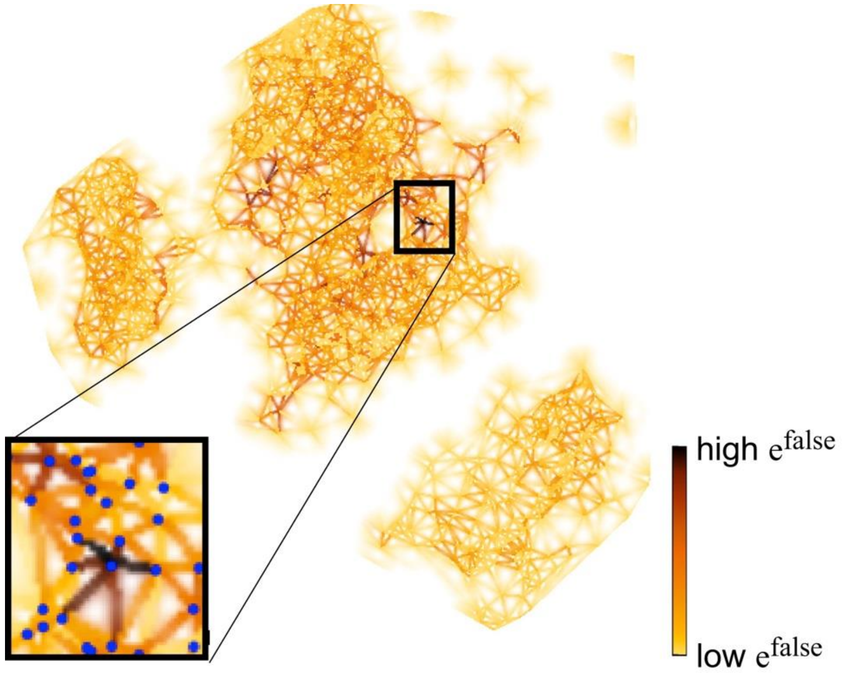
Quality Assessment: Visualization

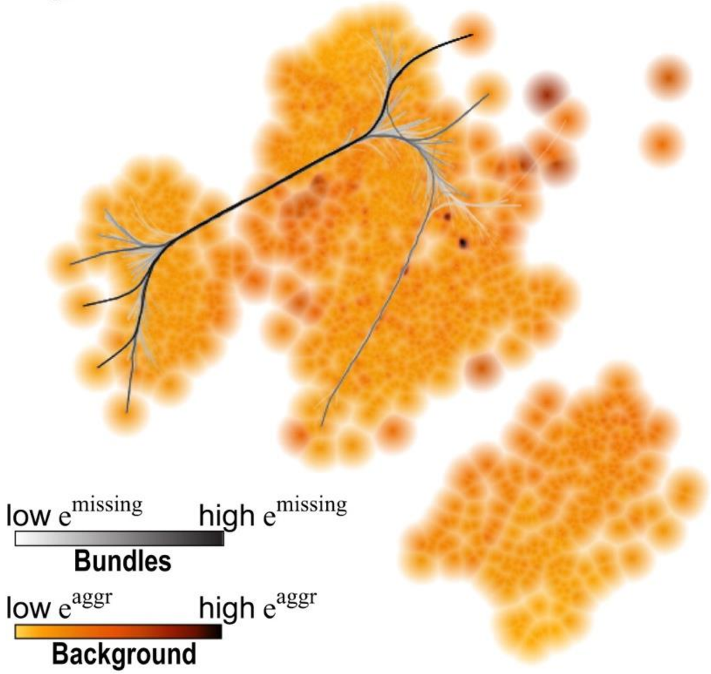


$$e_{ij} = \frac{d^m(\mathbf{q}_i, \mathbf{q}_j)}{\max_{i,j} d^m(\mathbf{q}_i, \mathbf{q}_j)} - \frac{d^n(\mathbf{p}_i, \mathbf{p}_j)}{\max_{i,j} d^n(\mathbf{p}_i, \mathbf{p}_j)}$$

Martins, R. M., Coimbra, D. B., Minghim, R., & Telea, A. C. (2014). Visual analysis of dimensionality reduction quality for parameterized projections. Computers & Graphics, 41, 26-42.

Quality Assessment: Visualization



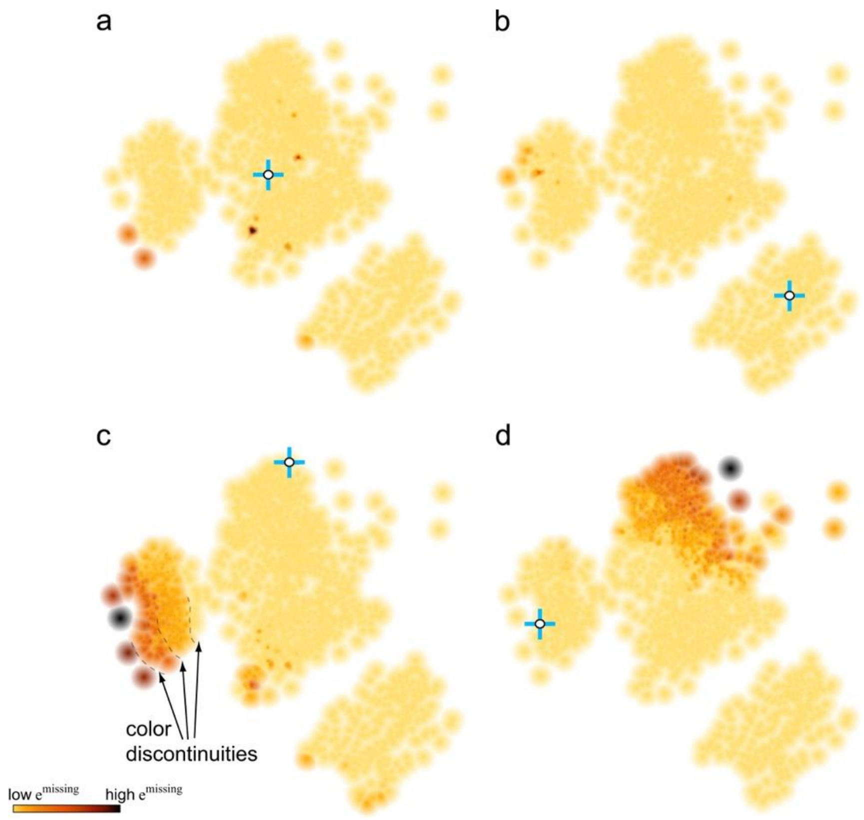


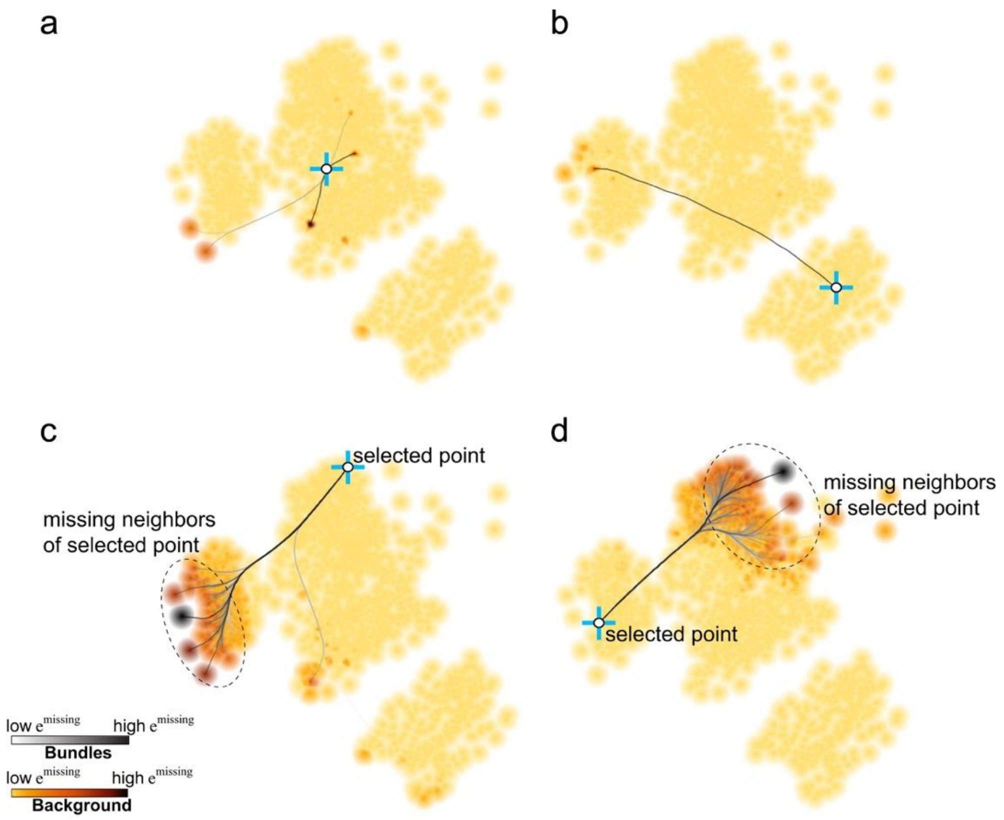
$$e^{false}\left(\mathbf{x}\right) = \frac{\sum_{1\le k\le 3} \frac{1}{d(\mathbf{x},E_k)\ \ \|E_k\|} e_k^{false}}{\sum_{1\le k\le 3} \frac{1}{d(\mathbf{x},E_k)\ \ \|E_k\|}}$$

$$e_i^{missing} = \max_{j\neq i}\left(e_{ij}, 0\right)$$

Martins, R. M., Coimbra, D. B., Minghim, R., & Telea, A. C. (2014). Visual analysis of dimensionality reduction quality for parameterized projections. Computers & Graphics, 41, 26-42.

Quality Assessment: Visualization

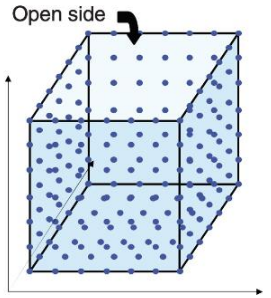


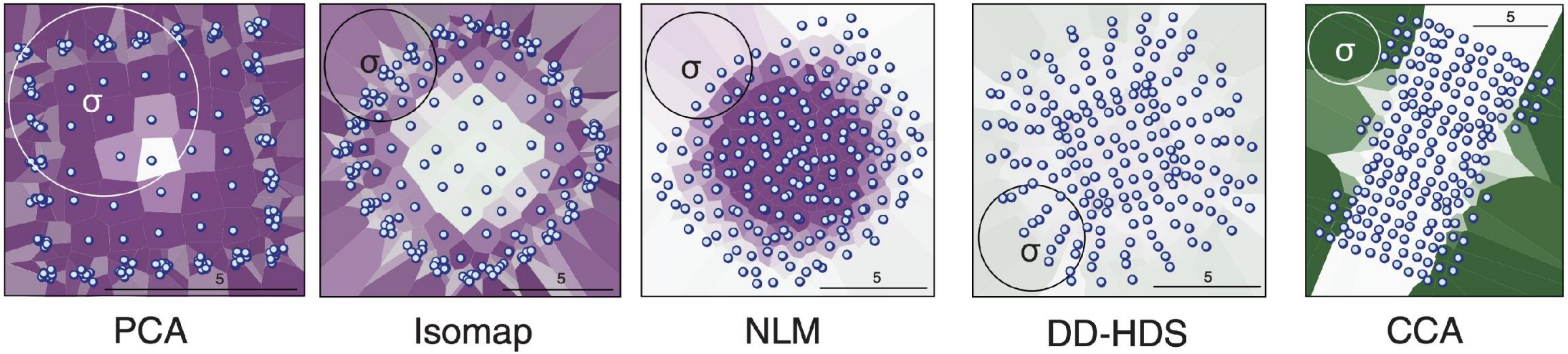


Martins, R. M., Coimbra, D. B., Minghim, R., & Telea, A. C. (2014). Visual analysis of dimensionality reduction quality for parameterized projections. Computers & Graphics, 41, 26-42.

CheckViz

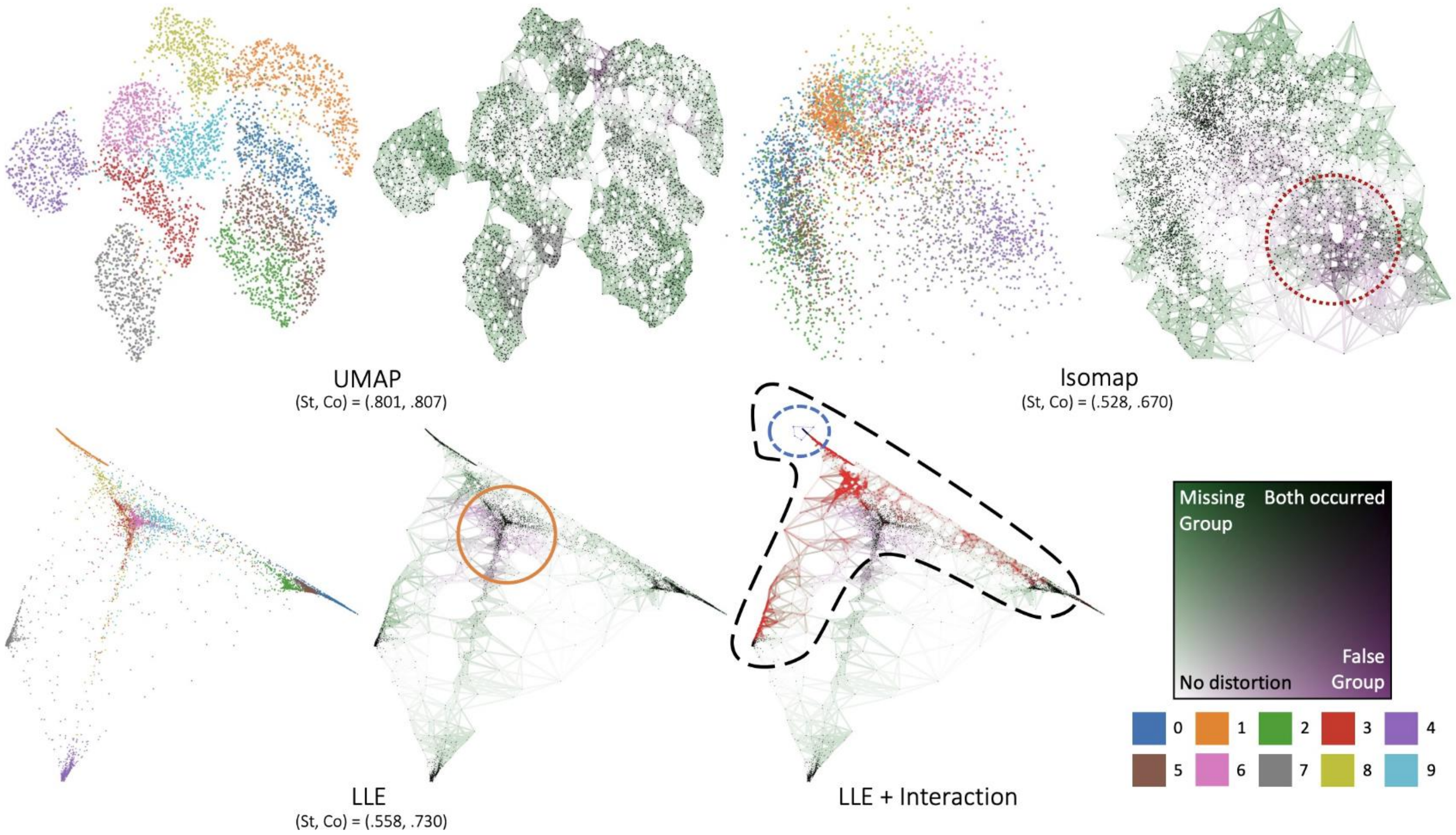
- Represents how much each point suffers from Missing / False Distortions
- Missing Neighbors: Green, False Neighbors: Purple

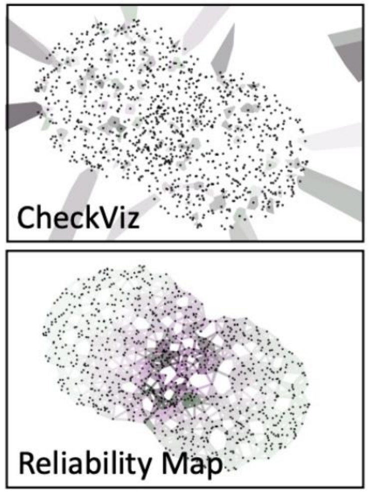




Lespinats, Sylvain, and Michaël Aupetit. "CheckViz: Sanity Check and Topological Clues for Linear and Non-Linear Mappings." Computer Graphics Forum. Vol. 30. No. 1. Oxford, UK: Blackwell Publishing Ltd, 2011.

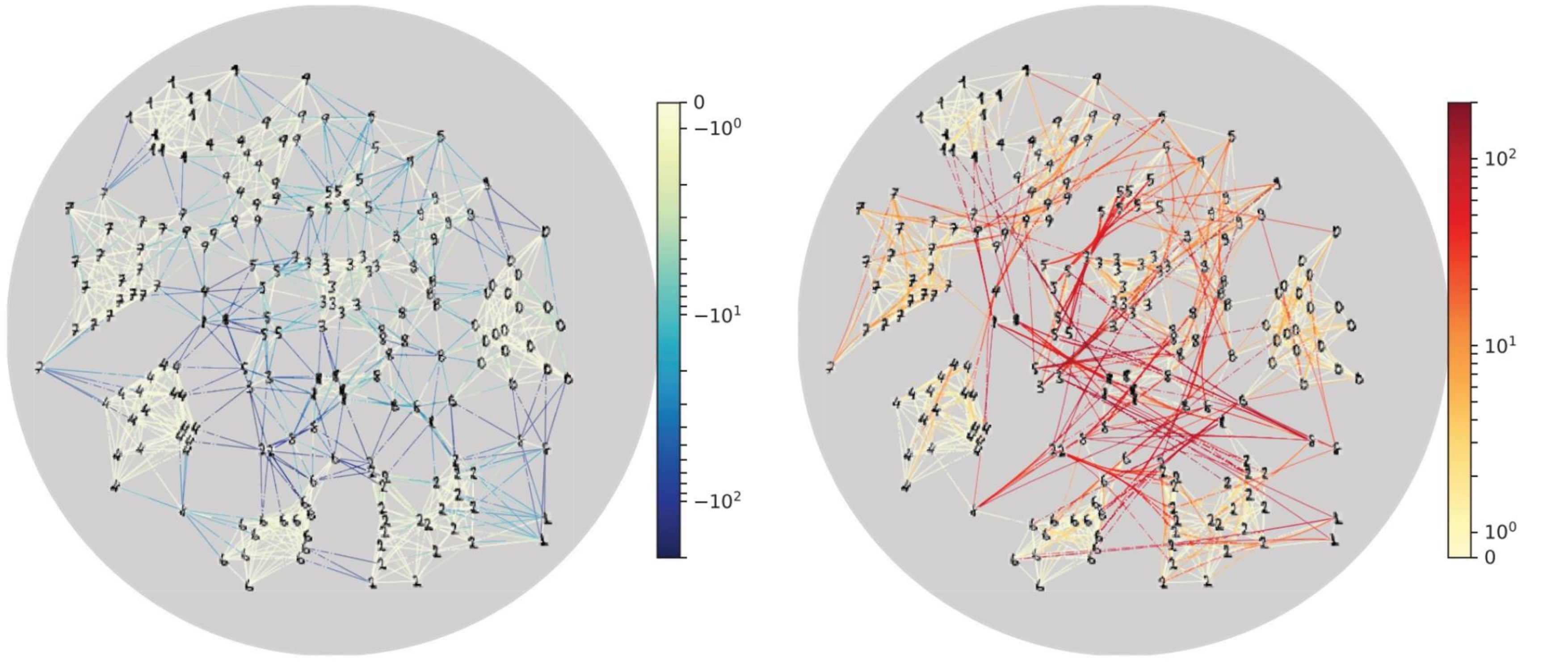
Reliability Map





Jeon, Hyeon, et al. "Measuring and explaining the inter-cluster reliability of multidimensional projections." IEEE Transactions on Visualization and Computer Graphics 28.1 (2021): 551-561.

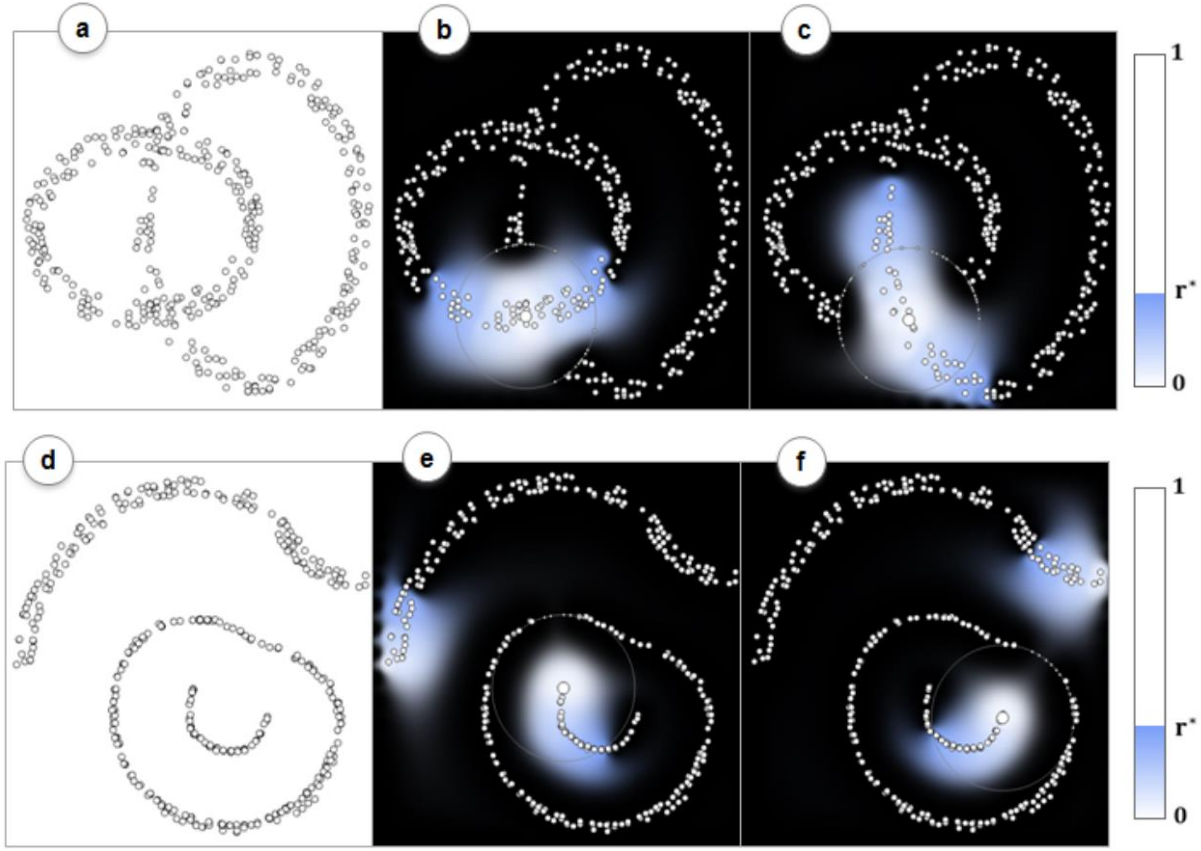
MING



(a) (b)

Colange, Benoît, et al. "MING: An interpretative support method for visual exploration of multidimensional data." Information Visualization 21.3 (2022): 246-269.

Proxilens

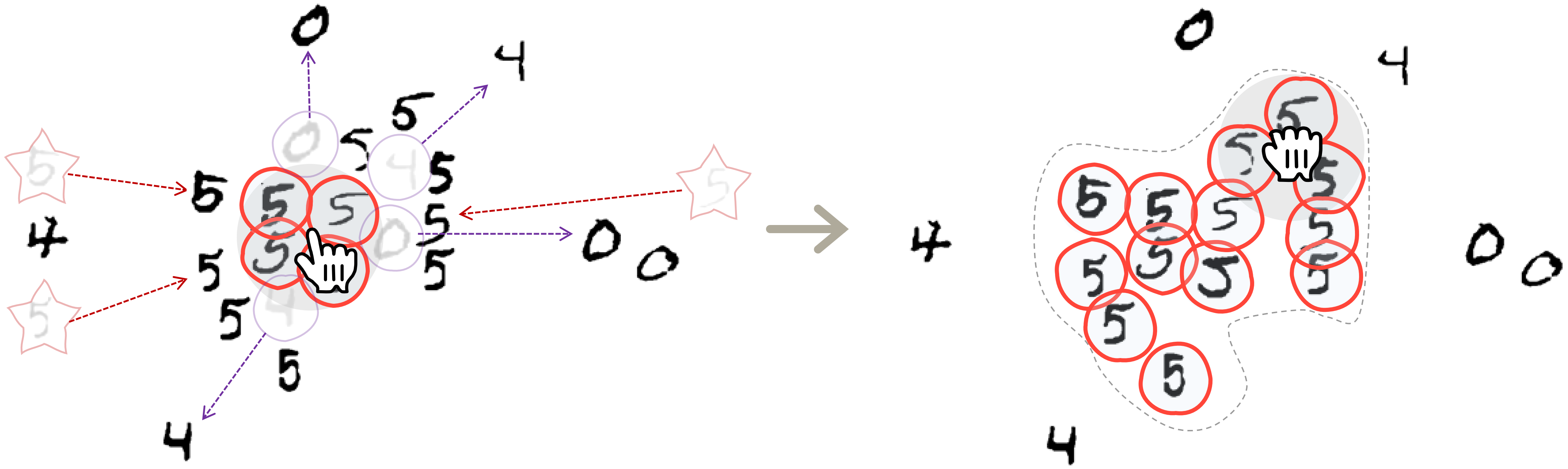


Heulot, Nicolas, Michael Aupetit, and Jean-Daniel Fekete. "Proxilens: Interactive exploration of high-dimensional data using projections." VAMP: EuroVis Workshop on Visual Analytics using Multidimensional Projections.

Distortion-aware Brushing

Points are **relocated** to resolve distortions in high-dimensional projections

Users can **reliably brush** visual 2D clusters that match **high-dimensional clusters**



Jeon, Hyeon, et al. "Distortion-aware brushing for interactive cluster analysis in multidimensional projections." arXiv preprint arXiv:2201.06379 (2022).

Measuring and Explaining Distortions in Practice

- ZADU
  - A Python library serving diverse distortion metrics
  - Provides 19 metrics so far
  - Latest release: v0.2.1

Type	Measure	Ref.	provide pointwise distortions	dreval [39]	McInnes et al. [29]	Ingram et al. [15]	Jeon et al. [18]	Fujiwara et al. [10]	Espadoto et al. [9]	Colange et al. [6]	coranking [22]	pyclustering [33]	scikit-learn [35]	scipy [41]	Moor et al. [30]	Jeon et al. [19]	**ZADU** (Ours)
Local	Trustworthiness & Continuity	[40]	V	△	△					○	○		△		○	○	○
	Mean Relative Rank Errors	[26]	V												○	○	○
	Local Continuity Meta-Criteria	[4]	V								○						○
	Neighborhood Hit	[34]	V			○											○
	Neighbor Dissimilarity	[10]						○									○
	Class-Aware Trustworthiness & Continuity	[6]	V							○							○
	Procrustes Measure	[12]															○
Cluster-level	Steadiness & Cohesiveness	[18]	V				○	○									○
	Distance Consistency	[37]															○
	Internal Clustering Validation Measures	[21]										○	○				○
	Clustering + External Clustering Validation Measures	[42]										○	○				○
Global	Stress	[23, 24]							○						○	○	○
	Kullback-Leibler Divergence	[13]													○	○	○
	Distance-to-Measure	[3]														○	○
	Topographic Product	[1]															○
	Pearson's correlation coefficient r	[11]												○			○
	Spearman's rank correlation coefficient ρ	[36]												○			○

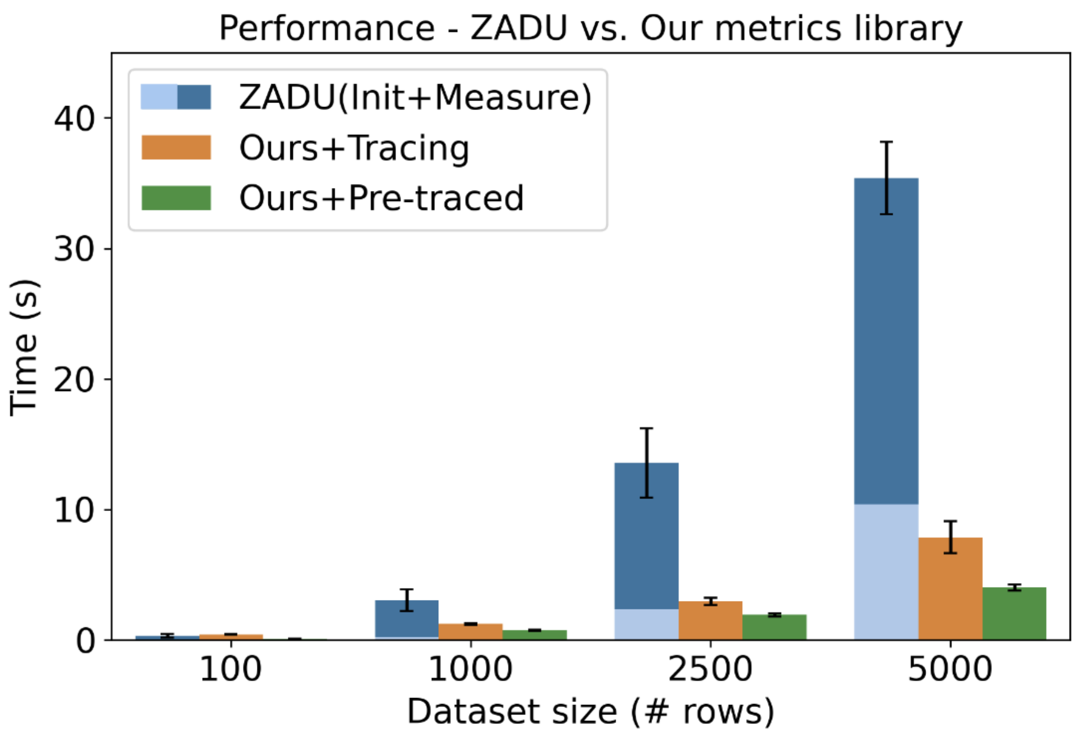
Jeon, Hyeon, et al. "Zadu: A python library for evaluating the reliability of dimensionality reduction embeddings." 2023 IEEE Visualization and Visual Analytics (VIS). IEEE, 2023.

ZADU

- ZADU is **accessible**
  - Served as a Python library, which can be easily integrated with existing tools
  - Deployed via pip --- easy to install and execute
- ZADU is **scalable**
  - ZADU automatically optimizes the execution of distortion measures
  - It is also accelerated by a parallel computing based on CPU multithreading
- ZADU covers a **wide range** of distortion measures

ZADU

- ZADU is **accessible**
  - Served as a Python library, wh
  - Deployed via pip --- easy to in
- ZADU is **scalable?**
  - ZADU automatically optimize
  - It is also accelerated by a paral
- ZADU covers a **wide range** of dis



Machado, Alister, Michael Behrisch, and Alexandru Telea. "Necessary but not Sufficient: Limitations of Projection Quality Metrics." Computer Graphics Forum. 2025.

New options available – Now at EuroVis

Table 1: *Metrics implemented by the benchmark of Espadoto* et al. *[EMK*21], ZADU [JCJ*23], and our work. Empty circles (○) denote implementation issues in ZADU. Specifically, the Procrustes statistic is computed incorrectly; the Topographic Product yields division by zero errors in cases which should be properly handled.*

Metric	Introduced in	Implemented in		
		Espadoto et al. [EMK*21]	ZADU [JCJ*23]	Ours
Average Local Error	[MCMT14]	●		●
Continuity and Trustworthiness	[VK06a]	●	●	●
Class-Aware Continuity and Trustworthiness	[CPA*20]		●	●
Distance Consistency (DSC)	[SNLH09]		●	●
Distance-to-Measure	[CCSM11]		●	
Proportion of False (resp. True) Neighbors	[MCMT14]			●
Jaccard Similarity of Neighbor Sets	[Jac01]			●
Local Continuity Meta-Criteria	[CB09]		●	
Mean Relative Ranking Errors	[LV09]		●	●
Neighbor Dissimilarity	[FKYM23]		●	
Neighborhood Hit	[PNML08]		●	●
Normalized Stress	[Kru64a, Kru64b, JCC*11]	●	●	●
Pearson Correlation of Distances	[GZZ05]		●	●
Procrustes Statistic	[GR09]		○	●
Scale-Normalized Stress	[SMK24]			●
Shepard Goodness	[SC88]	●	●	●
Steadiness and Cohesiveness	[JKJ*21]		●	
Topographic Product	[BP92]		○	
Internal Clustering Validation Measures	[JCC*11]		●	
Clustering + External Clustering Validation Measures	[XWY*21]		●	

Machado, Alister, Michael Behrisch, and Alexandru Telea. "Extensible TensorFlow Implementations of Projection Quality Metrics." (2025).

ZADU Interface

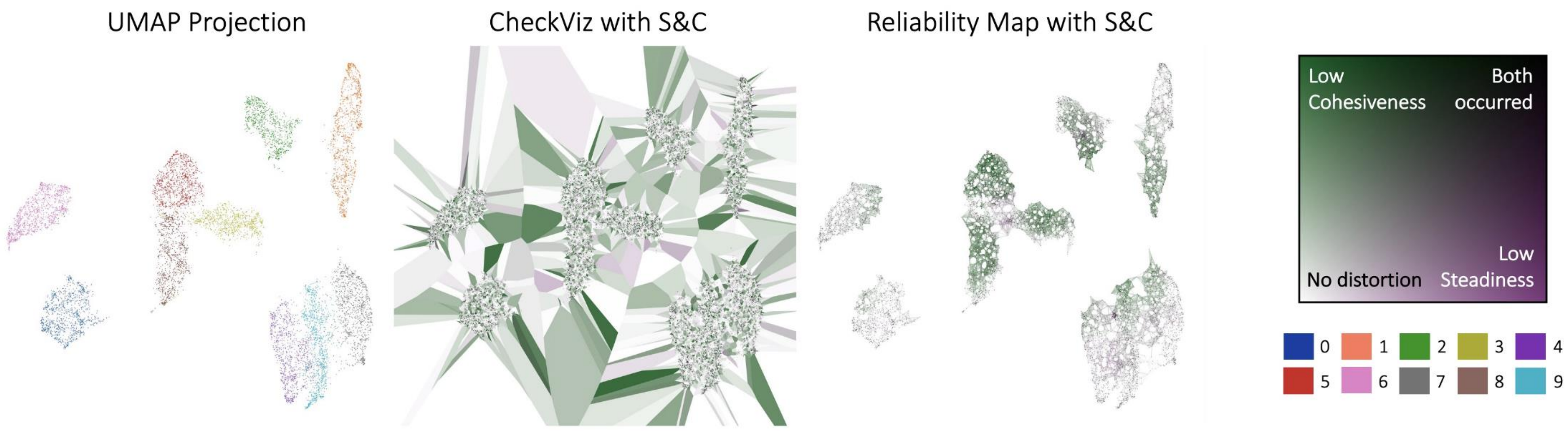
```python
from zadu import zadu

hd, ld = load_datasets()
spec = [{
    "id"    : "tnc",
    "params": { "k": 20 },
}, {
    "id"    : "snc",
    "params": { "k": 30, "clustering_strategy": "dbscan" }
}]

scores = zadu.ZADU(spec, hd).measure(ld)
print("T&C:", scores[0])
print("S&C:", scores[1])
```

ZADU Interface

```python
from zadu import zadu

hd, ld = load_datasets()
spec = [{
    "id"    : "tnc",
    "params": { "k": 20 },
}, {
    "id"    : "snc",
    "params": { "k": 30, "clustering_strategy": "dbscan" }
}]

scores = zadu.ZADU(spec, hd).measure(ld)
print("T&C:", scores[0])
print("S&C:", scores[1])
```

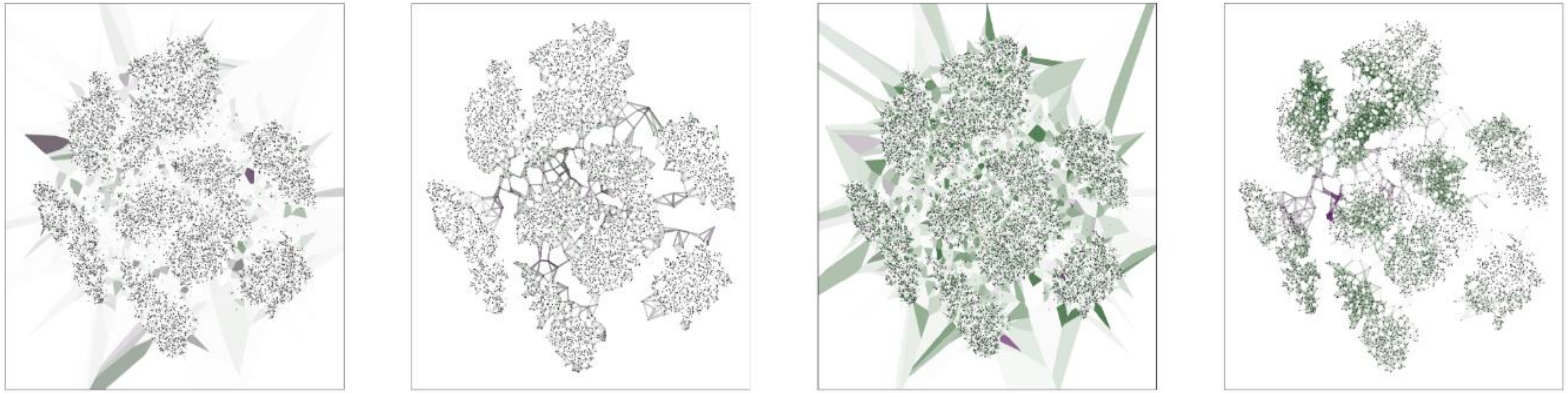
```python
from zadu.measures import *

mrre = mean_relative_rank_error.measure(hd, ld, k=20)
pr   = pearson_r.measure(hd, ld)
nh   = neighborhood_hit.measure(ld, label, k=20)
```

Visualizing Distortions with ZADU



Visualizing Distortions with ZADU

```
from zadu import zadu
from zaduvis import zaduvis
import matplotlib.pyplot as plt
from sklearn.manifold import TSNE
from sklearn.datasets import fetch_openml


hd = fetch_openml('mnist_784', version=1, cache=True).data.to_numpy()[::7]
ld = TSNE().fit_transform(hd)

## Computing local pointwise distortions
spec = [{
    "id": "tnc",
    "params": {"k": 25}
},{
    "id": "snc",
    "params": {"k": 50}
}]
zadu_obj = zadu.ZADU(spec, hd, return_local=True)
scores, local_list = zadu_obj.measure(ld)

tnc_local = local_list[0]
snc_local = local_list[1]

local_trustworthiness = tnc_local["local_trustworthiness"]
local_continuity = tnc_local["local_continuity"]
local_steadiness = snc_local["local_steadiness"]
local_cohesiveness = snc_local["local_cohesiveness"]

fig, ax = plt.subplots(1, 4, figsize=(50, 12.5))
zaduvis.checkviz(ld, local_trustworthiness, local_continuity, ax=ax[0])
zaduvis.reliability_map(ld, local_trustworthiness, local_continuity, k=10, ax=ax[1])
zaduvis.checkviz(ld, local_steadiness, local_cohesiveness, ax=ax[2])
zaduvis.reliability_map(ld, local_steadiness, local_cohesiveness, k=10, ax=ax[3])
```



Note: We will use an easier way in our programming practice!!

What you learned today

- Dimensionality Reduction - Overview
  - PCA, MDS
  - Modern nonlinear DR
- Quality Assessment
  - Distortion types
    - Stretching/Compression, Missing/False Neighbors/Groups
  - Quality metrics
    - Global, Local, and Cluster-level metrics
  - Visualizing quality metrics
- ZADU library

References

- Borg and Groenen, Modern multidimensional scaling: Theory and applications. Springer Science & Business Media.
- Huang et al. Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization. Communications biology
- Espadoto et al. Toward a quantitative survey of dimension reduction techniques. IEEE transactions on visualization and computer graphics
- Cashman et al. A critical analysis of the usage of dimensionality reduction in four domains. IEEE Transactions on Visualization and Computer Graphics
- Sylvain and Aupetit. CheckViz: Sanity Check and Topological Clues for Linear and Non-Linear Mappings. Computer Graphics Forum.
- Venna et al. Information retrieval perspective to nonlinear dimensionality reduction for data visualization. Journal of Machine Learning Research,
- Jeon et al. Measuring and explaining the inter-cluster reliability of multidimensional projections. IEEE Transactions on Visualization and Computer Graphics
- Jeon et al. Classes are not clusters: Improving label-based evaluation of dimensionality reduction. IEEE Transactions on Visualization and Computer Graphics
- Martins et al. Visual analysis of dimensionality reduction quality for parameterized projections. Computers & Graphics
- Colange et al. MING: An interpretative support method for visual exploration of multidimensional data. Information Visualization
- Heulot, Nicolas, Michael Aupetit, and Jean-Daniel Fekete. "Proxilens: Interactive exploration of high-dimensional data using projections." VAMP
- Jeon, Hyeon, et al. "Distortion-aware brushing for interactive cluster analysis in multidimensional projections." arXiv preprint
- Jeon et al. Zadu: A python library for evaluating the reliability of dimensionality reduction embeddings. IEEE Visualization and Visual Analytics
- Machado et al. Necessary but not Sufficient: Limitations of Projection Quality Metrics. Computer Graphics Forum