Dissimilarity Plots
A Visual Exploration Tool for Partitional Clustering

Michael Hahsler[1] & Kurt Hornik[2]

[1] Intelligent Data Analysis Lab (IDA@SMU)
Dept. of Engineering Management, Information, and Systems, SMU
[2] Vienna University of Economics and Business

IE Department Seminar
Department of Industrial, Manufacturing, & Systems Engineering
University of Texas at Arlington
February 9, 2017

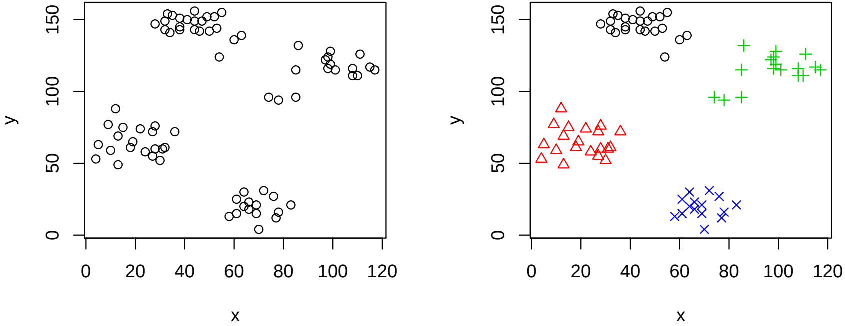
SMU | Bobby B. Lyle School of Engineering

WU Wirtschaftsuniversität Wien Vienna University of Economics and Business

Table of Contents

Cluster Analysis

Clustering assigns objects to groups (clusters) so that objects from the same cluster are more similar to each other than to objects from other clusters.

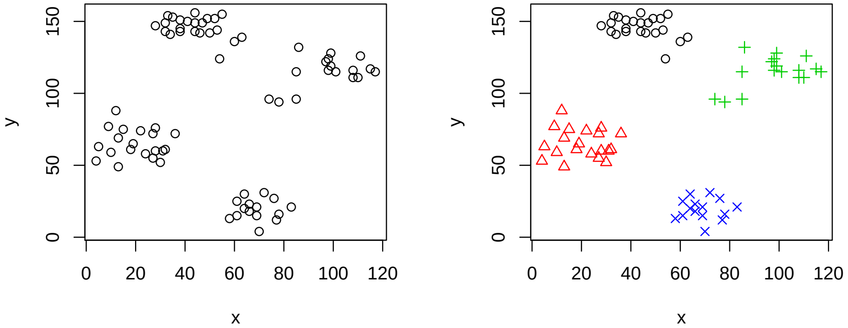


Applications

- Unsupervised learning of structure in the data and summarizing data.
- Areas: Business (market segmentation), biology (communities), social networks, AI, etc.

Assessment of Cluster Quality

Clustering assigns objects to groups (clusters) so that objects from the same cluster are more similar to each other than to objects from other clusters.

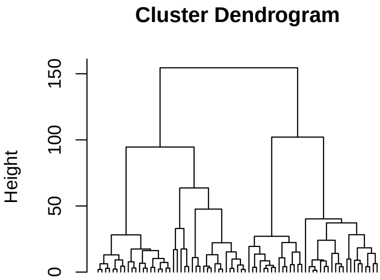


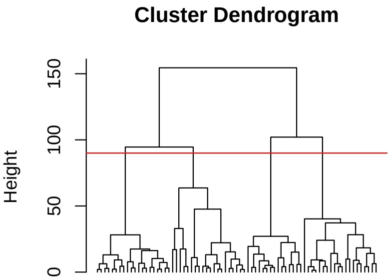
Assess the quality of a cluster solution

- Typically judged by intra and inter-cluster similarities
- Visualization for judging the quality of a clustering and to explore the cluster structure

Dendrograms

Dendrograms (Hartigan, 1967) for hierarchical clustering:



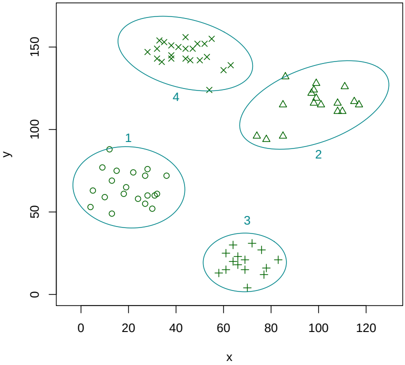


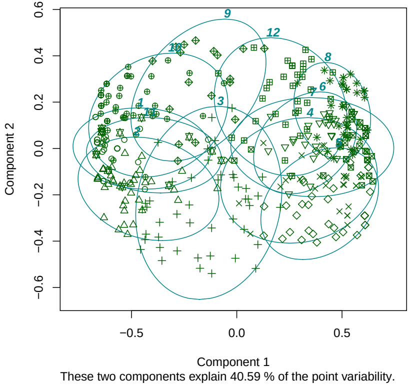
Restriction

Dendrograms are only possible for hierarchical/nested clusterings.

Table of Contents

Project objects into 2-dimensional space with dimensionality reduction techniques (e.g., PCA, MDS; Pison *et al.* (1999)).

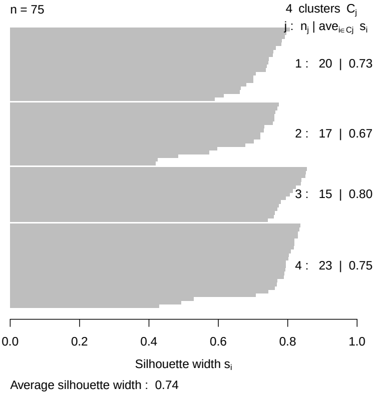




Problems with dimensionality (figure to the right: MDS/32-dimensional data)

Plot Quality Metrics

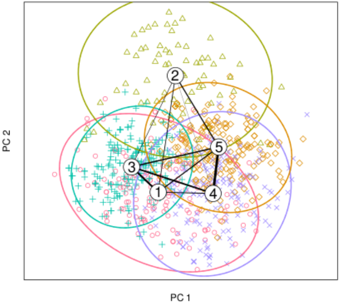
Visualize metrics calculated from inter and intra-cluster similarities to judge cluster quality. For example, **silhouette width** (Kaufman and Rousseeuw, 1990).



→ Only a diagnostic tool for cluster quality

Other Visualization Methods

Several other visualization methods (e.g., based on self-organizing maps and neighborhood graphs, shadow plots, shadow-stars, stripes plots) are reviewed and introduced in Leisch (2008, 2010).



Neighborhood graph

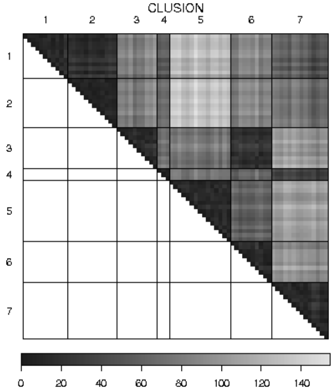
- Typically hide structure within clusters or
- are limited by the number of clusters and dimensionality of data.

Dissimilarity Matrix Shading and CLUSION

Each cell of the (dissimilarity) matrix is represented by a gray value (Sneath and Sokal, 1973; Ling, 1973; Gale *et al.*, 1984). Initially matrix shading was used with hierarchical clustering → **heatmaps**.

For graph-based partitional clustering: **CLUSION** (Strehl and Ghosh, 2003).

Uses **coarse seriation** such that "good" clusters from blocks around the main diagonal.



CLUSION allows to judge cluster quality but does not reveal the structure of the data

→ Dissimilarity plots

Improve matrix shading/CLUSION with (near) optimal placement of clusters and objects within clusters using **seriation**

Table of Contents

Seriation I

Part of **combinatorial data analysis** (Arabie and Hubert, 1996)

- **Aim:** arrange objects in a linear order given available data and some loss function in order to reveal structural information.
- **Problem:** Requires to solve a discrete optimization problem
  $\rightarrow$ solution space grows by the order of $O(n!)$

Seriation I

Part of **combinatorial data analysis** (Arabie and Hubert, 1996)

- **Aim:** arrange objects in a linear order given available data and some loss function in order to reveal structural information.
- **Problem:** Requires to solve a discrete optimization problem $\rightarrow$ solution space grows by the order of $O(n!)$

Techniques:

1. Partial enumeration methods (currently solve problems with $n \leq 40$)
   - dynamic programming (Hubert *et al.*, 1987)
   - branch-and-bound (Brusco and Stahl, 2005)
2. Heuristics for larger problems

Seriation II

- Set of n objects $\mathcal{O} = \{O_1, O_2, \ldots, O_n\}$.
- Symmetric dissimilarity matrix $\mathbf{D} = \{d_{ij}\}$, where d_{ij} for $1 \leq i, j \leq n$ represents the dissimilarity between O_i and O_j, and $d_{ii} = 0$ for all i.
- Permutation function $\psi_\pi(\mathbf{D}) = \{d_{\pi(i),\pi(j)}\} = \mathbf{P}_\pi \mathbf{D} \mathbf{P}_\pi^T$ reorders the objects in $\mathbf{D}$ by simultaneously permuting rows and columns according to a permutation π. $(\mathbf{P}_\pi = \psi_\pi(\mathbf{I}_n))$
- A loss function L to evaluate a given permutation.

$\mathbf{D}$

	O_3	O_2	O_1	O_4
O_3	0	4	1	8
O_2	4	0	2	2
O_1	1	2	0	3
O_4	8	2	3	0

$\pi = \{3, 2, 1, 4\}$

Seriation II

- Set of n objects $\mathcal{O} = \{O_1, O_2, \ldots, O_n\}$.
- Symmetric dissimilarity matrix $\mathbf{D} = \{d_{ij}\}$, where d_{ij} for $1 \leq i, j \leq n$ represents the dissimilarity between O_i and O_j, and $d_{ii} = 0$ for all i.
- Permutation function $\psi_\pi(\mathbf{D}) = \{d_{\pi(i),\pi(j)}\} = \mathbf{P}_\pi \mathbf{D} \mathbf{P}_\pi^T$ reorders the objects in $\mathbf{D}$ by simultaneously permuting rows and columns according to a permutation π. $(\mathbf{P}_\pi = \psi_\pi(\mathbf{I}_n))$
- A loss function L to evaluate a given permutation.

D

	O_3	O_2	O_1	O_4
O_3	0	4	1	8
O_2	4	0	2	2
O_1	1	2	0	3
O_4	8	2	3	0

$\pi = \{3, 2, 1, 4\}$

Optimization problem

$$\text{minimize } Z = L(\psi_\pi(\mathbf{D}))$$
$$\text{s.t.} \quad \psi_\pi \in \Psi \text{ (valid perm)}$$

Seriation II

- Set of n objects $\mathcal{O} = \{O_1, O_2, \dots, O_n\}$.
- Symmetric dissimilarity matrix $\mathbf{D} = \{d_{ij}\}$, where d_{ij} for $1 \leq i, j \leq n$ represents the dissimilarity between O_i and O_j, and $d_{ii} = 0$ for all i.
- Permutation function $\psi_\pi(\mathbf{D}) = \{d_{\pi(i),\pi(j)}\} = \mathbf{P}_\pi \mathbf{D} \mathbf{P}_\pi^T$ reorders the objects in $\mathbf{D}$ by simultaneously permuting rows and columns according to a permutation π. $(\mathbf{P}_\pi = \psi_\pi(\mathbf{I}_n))$
- A loss function L to evaluate a given permutation.

$$\mathbf{D}$$

$$\begin{array}{c|cccc} & O_3 & O_2 & O_1 & O_4 \\ \hline O_3 & 0 & 4 & 1 & 8 \\ O_2 & 4 & 0 & 2 & 2 \\ O_1 & 1 & 2 & 0 & 3 \\ O_4 & 8 & 2 & 3 & 0 \end{array}$$

$\pi = \{3, 2, 1, 4\}$

Optimization problem

$$\text{minimize } Z = L(\psi_\pi(\mathbf{D}))$$
$$\text{s.t.} \quad \psi_\pi \in \Psi \text{ (valid perm)}$$

How should the loss function be defined?

Column/Row Gradient Measures I

Perfect anti-Robinson matrix (Robinson, 1951): A symmetric matrix where the values in all rows and columns only increase when moving away from the main diagonal. Gradient conditions (Hubert *et al.*, 1987):

$$\begin{array}{rlll} \text{within rows:} & d_{ik} \leq d_{ij} & \text{for} & 1 \leq i < k < j \leq n; \\ \text{within columns:} & d_{kj} \leq d_{ij} & \text{for} & 1 \leq i < k < j \leq n. \end{array}$$

$$\mathbf{D} \quad \begin{array}{c|cccc} & O_1 & O_2 & O_3 & O_4 \\ \hline O_1 & 0 & 4 & 1 & 8 \\ O_2 & 4 & 0 & 2 & 2 \\ O_3 & 1 & 2 & 0 & 3 \\ O_4 & 8 & 2 & 3 & 0 \end{array} \qquad \mathbf{\Psi(D)} \quad \begin{array}{c|cccc} & O_1 & O_3 & O_2 & O_4 \\ \hline O_1 & 0 & 1 & 4 & 8 \\ O_3 & 1 & 0 & 2 & 3 \\ O_2 & 4 & 2 & 0 & 2 \\ O_4 & 8 & 3 & 2 & 0 \end{array}$$

Moves similar items (O_1 and O_3) closer together.

Note: Most matrices can only be brought into a near anti-Robinson form.

Column/Row Gradient Measures II

Loss measure (quantifies the divergence from anti-Robinson form):

$$L(\mathbf{D}) = \sum_{i<k<j} f(d_{ik}, d_{ij}) + \sum_{i<k<j} f(d_{kj}, d_{ij})$$

where $f(\cdot,\cdot)$ is a function which defines how a violation or satisfaction of a gradient condition for an object triple (O_i, O_k and O_j) is counted.

Column/Row Gradient Measures II

Loss measure (quantifies the divergence from anti-Robinson form):

$$L(\mathbf{D}) = \sum_{i<k<j} f(d_{ik}, d_{ij}) + \sum_{i<k<j} f(d_{kj}, d_{ij})$$

where $f(\cdot,\cdot)$ is a function which defines how a violation or satisfaction of a gradient condition for an object triple (O_i, O_k and O_j) is counted.

Raw number: Violations minus satisfactions:

$$f(z, y) = \operatorname{sign}(y - z) = \begin{cases} -1 & \text{if} \quad z > y; \\ 0 & \text{if} \quad z = y; \\ +1 & \text{if} \quad z < y. \end{cases}$$

Column/Row Gradient Measures II

Loss measure (quantifies the divergence from anti-Robinson form):

$$L(\mathbf{D}) = \sum_{i<k<j} f(d_{ik}, d_{ij}) + \sum_{i<k<j} f(d_{kj}, d_{ij})$$

where $f(\cdot, \cdot)$ is a function which defines how a violation or satisfaction of a gradient condition for an object triple (O_i, O_k and O_j) is counted.

Raw number: Violations minus satisfactions:

$$f(z, y) = \operatorname{sign}(y - z) = \begin{cases} -1 & \text{if} \quad z > y; \\ 0 & \text{if} \quad z = y; \\ +1 & \text{if} \quad z < y. \end{cases}$$

Weighted: Weight each satisfaction or violation by its magnitude (absolute difference between the values):

$$f(z, y) = |y - z| \operatorname{sign}(y - z) = y - z$$

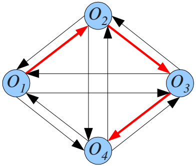
Hamiltonian Path Length

- $\mathbf{D}$ is seen as a finite weighted graph $G = (\Omega, E)$ with $\Omega = \{O_1, O_2, \ldots, O_n\}$ and the weight w_{ij} for edge $e_{ij} \in E$ represents d_{ij}.
- An order Ψ can be seen as a **Hamiltonian path** through the graph.
- Minimizing the path length results in a seriation optimal with respect to dissimilarities between neighboring objects (Hubert, 1974; Caraux and Pinloche, 2005).

Loss function:

$$L(\mathbf{D}) = \sum_{i=1}^{n-1} d_{i,i+1}$$

$\mathbf{D}$	O_1	O_2	O_3	O_4
O_1	0	**4**	1	8
O_2	4	0	**2**	2
O_3	1	2	0	**3**
O_4	8	2	3	0



This optimization problem is related to the **traveling salesperson problem** (Gutin and Punnen, 2002) for which good solvers and efficient heuristics exist.

Seriation Criteria

Measure	Definition
Gradient conditions	
Anti-Robinson (AR) events (Chen, 2002)	$\sum_{i<k<j} f(d_{ik}, d_{ij}) + f(d_{kj}, d_{ij})$, with $f(x, y) = I(x > y)$
AR deviations (Chen, 2002)	with $f(x, y) = \lvert y - x\rvert\, I(x > y)$
Gradient measure (Hubert *et al.*, 2001)	with $f(x, y) = -\operatorname{sign}(y - x)$
Weighted gradient measure (Hubert *et al.*, 2001)	with $f(x, y) = -\lvert y - x\rvert \operatorname{sign}(y - x)$
Relative generalized Anti-Robinson events (RGAR) (Tien *et al.*, 2008)	$\frac{1}{m} \sum_{i=1}^{n} \Big(\sum_{(i-w)\le i<k<j} I(d_{ik} < d_{ij}) + \sum_{i<k<j\le(i+w)} I(d_{kj} > d_{ij}) \Big)$, with window size $1 < w < n$ and $m = (^2/_3 - n)w + nw^2 - {}^2/_3\, w^3$
Rank/dissimilarity agreement	
Least squares criterion (Caraux and Pinloche, 2005)	$\sum_{i,j=1}^{n} (d_{ij} - \lvert i - j\rvert)^2$
Inertia criterion (Caraux and Pinloche, 2005)	$-1 \times \sum_{i,j=1}^{n} d_{ij}(i - j)^2$
2-Sum criterion (Barnard *et al.*, 1993)	$\sum_{i,j=1}^{n} \frac{1}{1+d_{ij}} (i - j)^2$
Linear seriation criterion (LS) (Hubert and Schultz, 1976)	$-1 \times \sum_{i,j=1}^{n} d_{ij} \lvert i - j\rvert$
Banded anti-Robinson form (BAR) (Earle and Hurley, 2015)	$\sum_{\lvert i-j\rvert \le b} d_{ij}(b + 1 - \lvert i - j\rvert)$ with band width $1 \le b < n$
Path length	
Hamiltonian path length (PL) (Hubert, 1974; Caraux and Pinloche, 2005)	$\sum_{i=1}^{n-1} d_{i,i+1}$

Table: Popular seriation criteria.

Source: Michael Hahsler. An experimental comparison of seriation methods for one-mode two-way data. **European Journal of Operational Research,** 257:133–143, 2017

Seriation Techiques

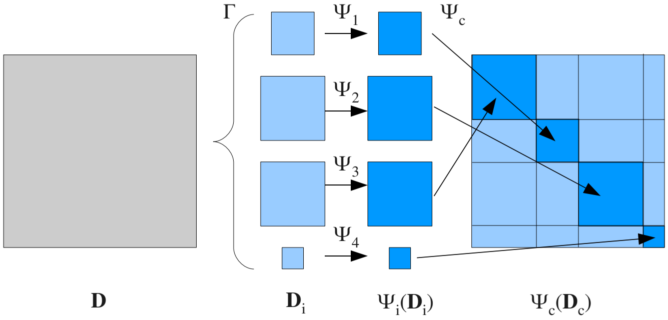
Technique	Objective function
Criterion optimization methods	
Integer linear programming (ILP) (Brusco, 2002)	Gradient conditions
Dynamic programming (Hubert *et al.*, 2001)	Gradient conditions
Branch-and-bound (Brusco and Stahl, 2005)	Gradient conditions
Genetic algorithm (Goldberg, 1989; Soltysiak and Jaskulski, 1998)	Various
Simulated annealing (ARSA) (Brusco *et al.*, 2008)	Gradient measure
Spectral seriation (Atkins *et al.*, 1999; Ding and He, 2004; Fogel *et al.*, 2014)	2-Sum criterion
TSP solver (various) (Wilkinson, 1971)	Hamiltonian path length
Quadratic assignment problem heuristic (QAP) (Hubert and Schultz, 1976; Caraux and Pinloche, 2005; Goulermas *et al.*, 2016)	2-Sum criterion, linear seriation, inertia or BAR
Dendrogram methods	
Hierarchical clustering (HC) (Eisen *et al.*, 1998)	Other (depends on linkage)
Gruvaeus and Wainer reordering (GW) (Gruvaeus and Wainer, 1972)	Restricted path length
Optimal leaf ordering reordering (OLO) (Bar-Joseph *et al.*, 2001)	Restricted path length
DendSer reordering (Earle and Hurley, 2015)	Various (restricted)
Other methods	
Multidimensional scaling (MDS) (Kendall, 1971)	Other (stress)
Rank-two ellipse seriation (R2E) (Chen, 2002)	None
Sorting Points Into Neighborhoods (SPIN) (Tsafrir *et al.*, 2005)	Other (energy)
Visual Assessment of Tendency (VAT) (Bezdek and Hathaway, 2002)	Other (MST)

Table: Popular seriation techniques.

Source: Michael Hahsler. An experimental comparison of seriation methods for one-mode two-way data. **European Journal of Operational Research,** 257:133–143, 2017 hahsler:Hahsler2016d

Table of Contents

Creating Dissimilarity Plots



1. **Split** $\mathbf{D}$ into clusters using the assignment function Γ provided by the partitional clustering algorithm
2. **Arrange objects:** Use $\Psi_1, \dots, \Psi_k$ to show micro-structure.
3. **Arrange clusters:** Ψ_c places more similar clusters together (macro-structure).

Arrange Clusters

Find Ψ_c based on inter-cluster dissimilarity matrix $\mathbf{D}_c$ which aggregates dissimilarities between all pairs of clusters given dissimilarities between all elements of the clusters in $\mathbf{D}$.

Hierarchical clustering: dissimilarities between two sets of objects $\mathcal{X}$ and $\mathcal{Y}$

$$
\begin{aligned}
\text{complete-link:} \quad & d_c(\mathcal{X},\mathcal{Y}) = \max\{d(x,y) : x \in \mathcal{X}, y \in \mathcal{Y}\} \\
\text{single-link:} \quad & d_s(\mathcal{X},\mathcal{Y}) = \min\{d(x,y) : x \in \mathcal{X}, y \in \mathcal{Y}\} \\
\text{average-link:} \quad & d_a(\mathcal{X},\mathcal{Y}) = \frac{1}{|\mathcal{X}| \cdot |\mathcal{Y}|} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} d(x,y)
\end{aligned}
$$

Set theory: Hausdorff metric (Hausdorff, 2001)

$$d_H(\mathcal{X},\mathcal{Y}) = \max\{\sup_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} d(x,y), \sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} d(x,y)\}$$

The Hausdorff metric pairs up each element from one set with the most similar element from the other set and then finds the largest dissimilarity in such element pairs.

Table of Contents

Used Seriation Methods

We use the column/row gradient measure as the loss function for seriation.

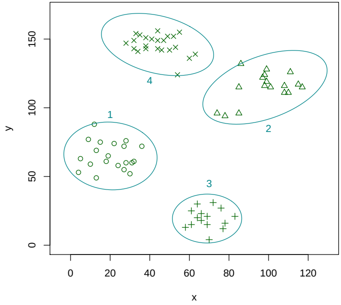
1. Placement (seriation) of clusters: Average-link, row/column gradient measure using branch-and-bound to find the optimal solution
2. Placement (seriation) of objects within each cluster: row/column gradient measure uses a simulated annealing heuristic

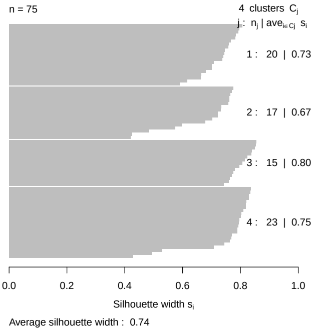
Seriation algorithms are provides by Brusco and Stahl (2005) and are available in the R extension package seriation (Hahsler *et al.*, 2016).

Easily Distinguishable Groups I

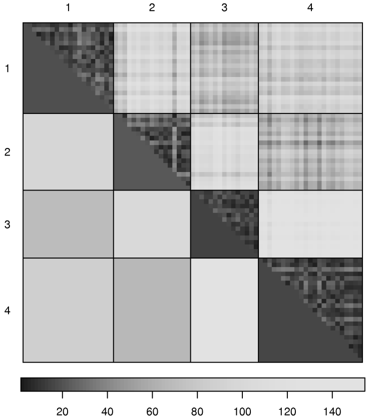
Ruspini data set (Ruspini, 1970) with 75 points in two-dimensional space with four clearly distinguishable groups.

Euclidean distances and k-medoids clustering algorithm (partitioning around medoids (PAM) (Kaufman and Rousseeuw, 1990)) to produce a partition with $k = 4$

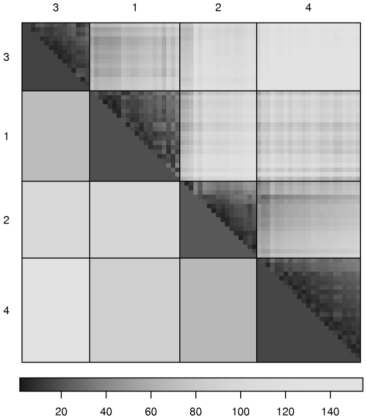




Easily Distinguishable Groups II

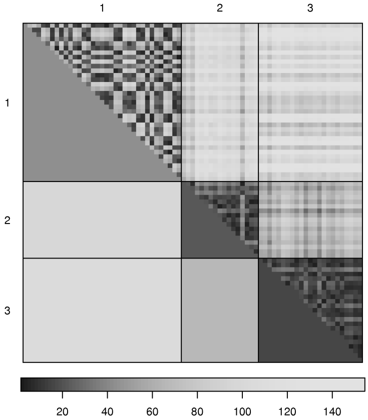


Coarse seriation

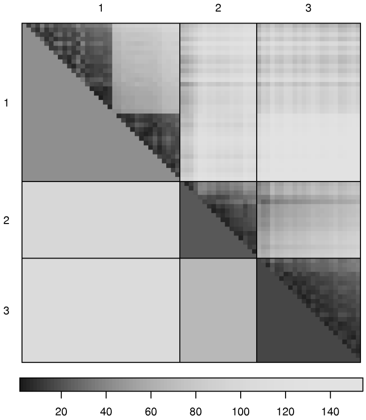


Dissimilarity plot

Mis-specification of the Number of Clusters I

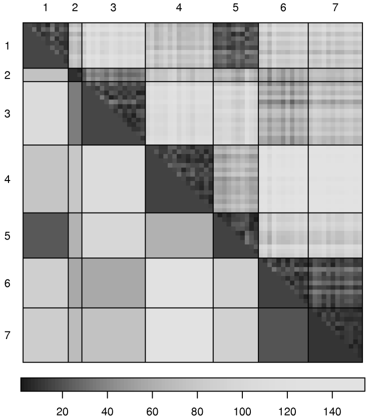


Coarse seriation, $k = 3$

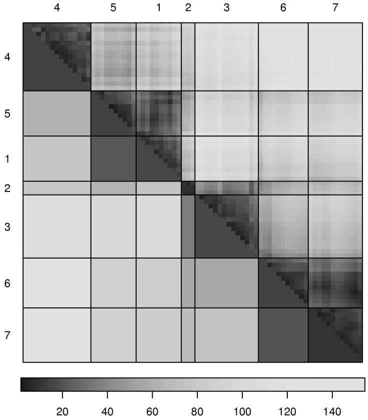


Dissimilarity plot, $k = 3$

Mis-specification of the Number of Clusters II



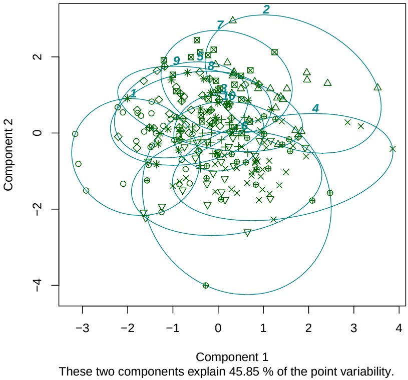
Coarse seriation, $k = 7$

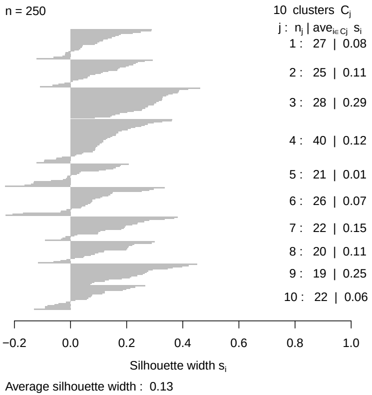


Dissimilarity plot, $k = 7$

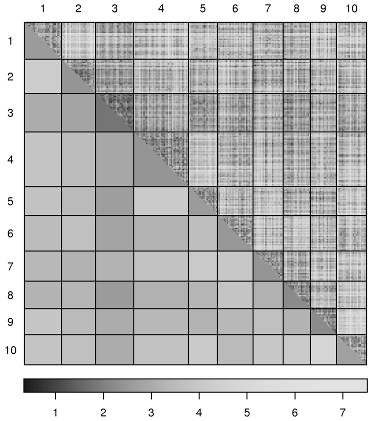
No Structure I

Random data for 250 objects in $\mathbb{R}^5$: $X_1, X_2, \ldots, X_5 \sim N(0,1)$
Euclidean distance and PAM with $k = 10$

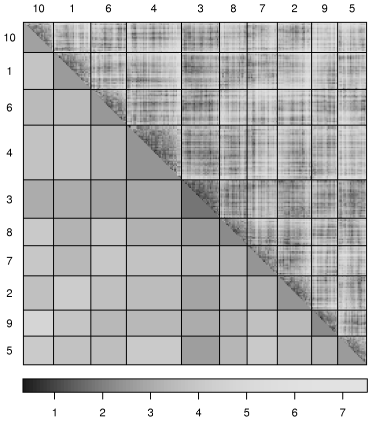




No Structure II



Coarse seriation



Dissimilarity plot

High-dimensional Data I

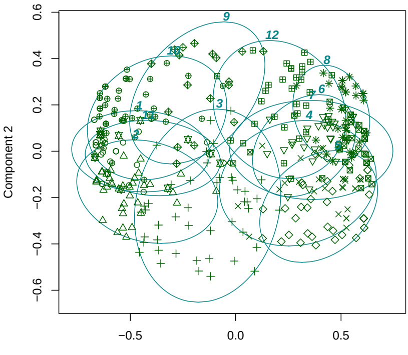
Votes data set (UCI Repository of Machine Learning Databases (Blake and Merz, 1998)). Votes for each of the U.S. House of Representatives congressmen on the 16 key votes during the second session of 1984.

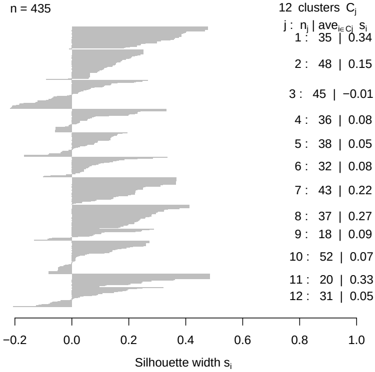
- **Coding:** 2 variables per vote (in favor/agains)
  $\rightarrow$ Each congressman is represented by a vector in $\{0, 1\}^{32}$
- **Dissimilarity measure: Jaccard dissimilarity** (Sneath and Sokal, 1973) between congressmen. Let S_i and S_j be the sets of votes two congressmen cast. Then the Jaccard dissimilarity

$$d_{ij} = 1 - \frac{S_i \cap S_j}{S_i \cup S_j}.$$

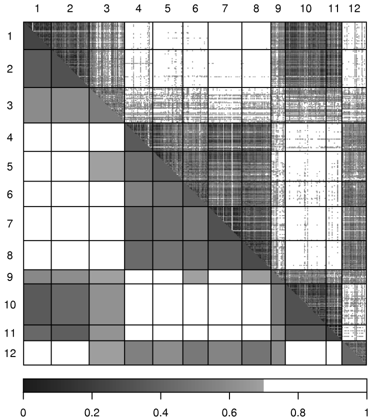
- **Cluster algorithm:** PAM with $k = 12$
  (the first bump of average silhouette for $k = 2, 3, \ldots, 30$)

High-dimensional Data II

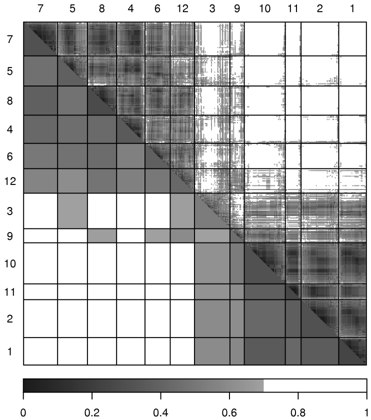




High-dimensional Data III

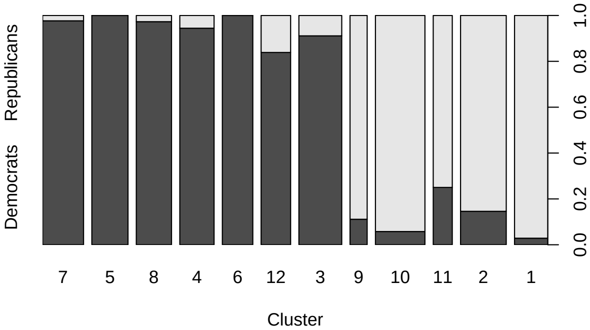


Coarse seriation, threshold=0.7



Dissimilarity plot, threshold=0.7

High-dimensional Data IV



Cluster composition (clusters reordered by dissimilarity plot)

Conclusion

Advantages of dissimilarity plots

- Scales well with dimensionality of data (visualizes dissimilarities)
- Shows cluster quality (block structure)
- Visual analysis of cluster structure (placement of clusters)
- Visual analysis of micro-structure (placement of objects)
- Makes misspecification of number of clusters apparent (placement of clusters/objects)

Enhancements for large number of objects/clusters

- **Object sampling:** Reduces the size of the dissimilarity matrix, however, details are sacrificed.
- **Image downsampling:** pixel skipping, pixel averaging, 2D discrete wavelet transformation
- **Interactive plot:** Plot with only average between-cluster similarities and then separate plot for each cluster (inter-cluster structures).

Further Reading and Code

Further Reading

Michael Hahsler, Kurt Hornik, and Christian Buchta. Getting things in order: An introduction to the R package seriation. **Journal of Statistical Software,** 25(3):1–34, March 2008

Michael Hahsler and Kurt Hornik. Dissimilarity plots: A visual exploration tool for partitional clustering. **Journal of Computational and Graphical Statistics,** 10(2): 335–354, 2011. (Selected for Best of JCGS session 2011)

Michael Hahsler. An experimental comparison of seriation methods for one-mode two-way data. **European Journal of Operational Research,** 257:133–143, 2017.

Code

Dissimilarity plot and seriation methods are implemented in the R extension package seriation (Hahsler *et al.*, 2016) and are freely available via the Comprehensive R Archive Network at

`https://CRAN.R-project.org/package=seriation`.

References I

David Applegate, Robert Bixby, Vasek Chvátal, and William Cook. *Concorde TSP Solver*, 2006.

P. Arabie and L. J. Hubert. The bond energy algorithm revisited. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(1):268–74, 1990.

P. Arabie and L. J. Hubert. An overview of combinatorial data analysis. In P. Arabie, L. J. Hubert, and G. De Soete, editors, *Clustering and Classification*, pages 5–63. World Scientific, River Edge, NJ, 1996.

Jonathan E. Atkins, Erik G. Boman, and Bruce Hendrickson. A spectral algorithm for seriation and the consecutive ones problem. *SIAM Journal on Computing*, 28(1):297–310, February 1999.

Z. Bar-Joseph, E. D. Demaine, D. K. Gifford, and T. Jaakkola. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics*, 17(1):22–29, 2001.

S. T. Barnard, A. Pothen, and H. D. Simon. A spectral algorithm for envelope reduction of sparse matrices. In *Proceedings of the 1993 ACM/IEEE Conference on Supercomputing*, pages 493–502, New York, NY, USA, 1993. ACM.

J Bertin. *Graphics and Graphic Information Processing*. Walter de Gruyter, Berlin, 1981. Translated by William J. Berg and Paul Scott.

J. Bertin. Graphics and graphic information processing. In S. K. Card, J. D. Mackinlay, and B. Shneiderman, editors, *Readings in Information Visualization*, pages 62–65, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc.

J.C. Bezdek and R.J. Hathaway. VAT: A tool for visual assessment of (cluster) tendency. In *Proceedings of the 2002 International Joint Conference on Neural Networks (IJCNN '02)*, pages 2225–2230, 2002.

C. L. Blake and C. J. Merz. UCI repository of machine learning databases, 1998.

Michael Brusco and Stephanie Stahl. *Branch-and-Bound Applications in Combinatorial Data Analysis*. Springer-Verlag, 2005.

M. Brusco, H. F. Köhn, and S. Stahl. Heuristic implementation of dynamic programming for matrix permutation problems in combinatorial data analysis. *Psychometrika*, 73:503–522, 2008.

Michael J. Brusco. Integer programming methods for seriation and unidemensional scaling of proximity matrices: A review and some extensions. *Journal of Classification*, 19(1):45–67, 2002.

References II

R. E. Burkard and F. Rendl. A thermodynamically motivated simulation procedure for combinatorial optimization problems. *European Journal of Operational Research*, 17(2):169–174, 1984.

Rainer E. Burkard, Eranda Cela, Panos M. Pardalos, and Leonidas S. Pitsoulis. The quadratic assignment problem. In P. Pardalos and D.-Z. Du, editors, *Handbook of Combinatorial Optimization*, pages 1713–1809. Springer-Verlag, 1999.

G. Caraux and S. Pinloche. Permutmatrix: A graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics*, 21(7):1280–1281, 2005.

Douglas Carroll and Phipps Arabie. Multidimensional scaling. *Annual Reviews Psychology*, 31:607–649, 1980.

Chun-Houh Chen. Generalized association plots: Information visualization via iteratively generated correlation matrices. *Statistica Sinica*, 12(1):7–29, 2002.

Daniel Chessel, Anne-Beatrice Dufour, and Stephane Dray. ade4*: Analysis of Ecological Data : Exploratory and Euclidean methods in Multivariate data analysis and graphical display*, 2007. R package version 1.4-3.

B. A. Davey and H. A. Priestley. *Introduction to lattices and order*. Cambridge University Press, Cambridge, 1990.

Antoine de Falguerolles, Felix Friedrich, and Günther Sawitzki. A tribute to J. Bertin's graphical data analysis. In W. Bandilla and F. Faulbaum, editors, *SoftStat '97: Advances in Statistical Software 6*, pages 11–20. Lucius & Lucius, 1997.

Jan de Leeuw. Unidimensional scaling. In B.S. Everitt and D.C. Howelll, editors, *Encyclopedia of Statistics in Behavioral Science*, volume 4. Wiley, 2005.

Chris Ding and Xiaofeng He. Linearized cluster assignment via spectral ordering. In *Proceedings of the Twenty-first International Conference on Machine Learning (ICML '04)*, page 30. ACM Press, 2004.

Stephane Dray and Anne-Beatrice Dufour. The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4), 6 2007.

Robert C. Dunnell. Seriation method and its evaluation. *American Antiquity*, 35(3):305–319, 1970.

Denise Earle and Catherine B. Hurley. Advances in dendrogram seriation for application to visualization. *Journal of Computational and Graphical Statistics*, 24(1):1–25, 2015.

References III

Michael B. Eisen, Paul T. Spellman, Patrick O. Browndagger, and David Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Science of the United States*, 95(25):14863–14868, December 1998.

Fajwel Fogel, Alexandre D. Aspremont, and Milan Vojnovic. Serialrank: Spectral ranking using seriation. In Z. Ghahramani, M. Welling, C. Cortes, N.d. Lawrence, and K.q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 900–908. Curran Associates, Inc., 2014.

Martin Fowler. *UML Distilled: A Brief Guide to the Standard Object Modeling Language*. Addison-Wesley Professional, third edition, 2004.

Michael Friendly. Corrgrams: Exploratory displays for correlation matrices. *The American Statistician*, 56(4):316–324, November 2002.

N. Gale, W. C. Halperin, and C. M. Costanzo. Unclassed matrix shading and optimal ordering in hierarchical cluster analysis. *Journal of Classification*, 1:75–92, 1984.

R. S. Garfinkel. The traveling salesman problem: Motivation and modeling. In E. Lawler et al, editor, *The Traveling Salesman Problem*, chapter 2, pages 17–36. John Wiley, 1985.

Alan George and Alex Pothen. An analysis of spectral envelope reduction via quadratic assignment problems. *SIAM Journal on Matrix Analysis and Applications*, 18(3):706–732, July 1997.

David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing, Boston, MA, USA, 1st edition, 1989.

John Goulermas, Alexandros Kostopoulos, and Tingting Mu. A new measure for analyzing and fusing sequences of objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(5):833–848, 2016.

G. Gruvaeus and H. Wainer. Two additions to hierarchical cluster analysis. *British Journal of Mathematical and Statistical Psychology*, 25:200–206, 1972.

G. Gutin and A. P. Punnen, editors. *The Traveling Salesman Problem and Its Variations*, volume 12 of *Combinatorial Optimization*. Kluwer, Dordrecht, 2002.

References IV

Michael Hahsler and Kurt Hornik. TSP – Infrastructure for the traveling salesperson problem. *Journal of Statistical Software*, 23(2):1–21, December 2007.

Michael Hahsler and Kurt Hornik. Dissimilarity plots: A visual exploration tool for partitional clustering. *Journal of Computational and Graphical Statistics*, 2011. Accepted for publication.

Michael Hahsler and Kurt Hornik. *TSP: Traveling Salesperson Problem (TSP)*, 2016. R package version 1.1-4.

Michael Hahsler, Kurt Hornik, and Christian Buchta. Getting things in order: An introduction to the R package seriation. *Journal of Statistical Software*, 25(3):1–34, March 2008.

Michael Hahsler, Christian Buchta, and Kurt Hornik. *Infrastructure for seriation*, 2016. R package version 1.2-1.

Michael Hahsler. An experimental comparison of seriation methods for one-mode two-way data. *European Journal of Operational Research*, 257:133–143, 2017.

J. A. Hartigan. Representation of similarity matrices by trees. *Journal of the American Statistical Association*, 62(320):1140–1158, 1967.

John A. Hartigan. *Clustering Algorithms*. Wiley, 1975.

Felix Hausdorff. *Set Theory*. American Mathematical Society, 5th edition, 2001.

M. Held and R. M. Karp. A dynamic programming approach to sequencing problems. *Journal of SIAM*, 10:196–210, 1962.

Kurt Hornik and David Meyer. relations*: Data Structures and Algorithms for Relations*, 2008. R package version 0.3-1.

L. J. Hubert and R. G. Golledge. Matrix reorganization and dynamic programming: Applications to paired comparisons and unidimensional seriation. *Psychometrika*, 46(4):429–441, 1981.

Lawrence Hubert and James Schultz. Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology*, 29(2):190–241, 1976.

Lawrence Hubert, Phipps Arabie, and Jacqueline Meulman. *Combinatorial Data Analysis: Optimization by Dynamic Programming*. Society for Industrial Mathematics, 1987.

Lawrence Hubert, Phipps Arabie, and Jacqueline Meulman. *Combinatorial Data Analysis: Optimization by Dynamic Programming*. Society for Industrial Mathematics, 2001.

References V

L. J. Hubert. Some applications of graph theory and related nonmetric techniques to problems of approximate seriation: The case of symmetric proximity measures. *British Journal of Mathematical Statistics and Psychology*, 27:133–153, 1974.

Catherine Hurley. gclus*: Clustering Graphics*, 2007. R package version 1.2.

Peter Ihm. A contribution to the history of seriation in archaeology. In Claus Weihs and Wolfgang Gaul, editors, *Classification - the Ubiquitous Challenge, Proceedings of the 28th Annual Conference of the Gesellschaft für Klassifikation e.V., University of Dortmund, March 9–11, 2004*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 307–316, 2005.

Giuseppe Jurman, Stefano Merler, Annalisa Barla, Silvano Paoli, Antonio Galea, and Cesare Furlanello. Algebraic stability indicators for ranked lists in molecular profiling. *Bioinformatics*, 24(2):258–264, January 2008.

L. Kaufman and P. J. Rousseeuw. *Finding groups in data: An introduction to cluster analysis*. John Wiley and Sons, New York, 1990.

Maurice Kendall. A new measure of rank correlation. *Biometrika*, 30(1–2):81–89, 1938.

D. Kendall. Seriation from abundance matrices. In F. Hodson, D. Kendall, and P. Tautu, editors, *Mathematics in the Archaeological and Historical Sciences*, pages 215–252. Edinburgh University Press, 1971.

M. Laurent and M. Seminaroti. The quadratic assignment problem is easy for Robinsonian matrices with Toeplitz structure. *Operations Research Letters*, 43(1):103–109, January 2015.

E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, editors. *The Traveling Salesman Problem*. Wiley, New York, 1985.

Friedrich Leisch. Visualizing cluster analysis and finite mixture models. In Chunhouh Chen, Wolfgang Härdle, and Antony Unwin, editors, *Handbook of Data Visualization*, Springer Handbooks of Computational Statistics. Springer Verlag, 2008.

Friedrich Leisch. Neighborhood graphs, stripes and shadow plots for cluster visualization. *Statistics and Computing*, 20:457–469, October 2010.

J. K. Lenstra. Clustering a data array and the traveling-salesman problem. *Operations Research*, 22(2):413–414, 1974.

Innar Liiv. Seriation and matrix reordering methods: An historical overview. *Statistical Analysis and Data Mining*, 3(2):70–91, 2010.

References VI

S. Lin and B. W. Kernighan. An effective heuristic algorithm for the traveling-salesman problem. *Operations Research*, 21(2):498–516, 1973.

Robert L. Ling. A computer generated aid for cluster analysis. *Communications of the ACM*, 16(6):355–361, 1973.

Carl P. Lipo, Mark E. Madsen, and Robert C. Dunnell. A theoretically-sufficient and computationally-practical technique for deterministic frequency seriation. *PLOS ONE*, 10(4):1–31, 2015.

P. Mair and J. De Leeuw. Unidimensional scaling. In *Wiley StatsRef: Statistics Reference Online*, pages 1–3. Wiley, 2015.

F. Marcotorchino. Block seriation problems: A unified approach. *Applied Stochastic Models and Data Analysis*, 3:73–91, 1987.

William T. McCormick, Paul J. Schweitzer, and Thomas W. White. Problem decomposition and data reorganization by a clustering technique. *Operations Research*, 20(5):993–1009, 1972.

David Meyer and Christian Buchta. proxy*: Distance and Similarity Measures*, 2007. R package version 0.1.

Fionn Murtagh. *Multidimensional Clustering Algorithms*, volume 4 of *Compstat Lectures*. Physica-Verlag, Vienna, 1985.

Stefan Niermann. Optimizing the ordering of tables with evolutionary computation. *The American Statistician*, 59(1):41–46, 2005.

Jari Oksanen, Roeland Kindt, Pierre Legendre, and Bob O'Hara. vegan*: Community Ecology Package*, 2007. R package version 1.8-6.

M. Padberg and G. Rinaldi. Facet identification for the symmetric traveling salesman polytope. *Mathematical Programming*, 47(2):219–257, 1990.

F. W. M. Petrie. Sequences in prehistoric remains. *Journal of the Anthropological Institute*, 29:295–301, 1899.

Greet Pison, Anja Struyf, and Peter J. Rousseeuw. Displaying a clustering with clusplot. *Computational Statistics & Data Analysis*, 30(4):381–392, June 1999.

W. S. Robinson. A method for chronologically ordering archaeological deposits. *American Antiquity*, 16:293–301, 1951.

Daniel J. Rosenkrantz, Richard E. Stearns, and II Philip M. Lewis. An analysis of several heuristics for the traveling salesman problem. *SIAM Journal on Computing*, 6(3):563–581, 1977.

References VII

P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(1):53–65, 1987.

E. H. Ruspini. Numerical methods for fuzzy clustering. *Information Science*, 2:319–350, 1970.

Peter H. A. Sneath and Robert R. Sokal. *Numerical Taxonomy*. Freeman and Company, San Francisco, 1973.

A. Soltysiak and P. Jaskulski. Czekanowski's diagram: A method of multidimensional clustering. In J.A. Barcelo, I. Briz, and A. Vila, editors, *Proceedings of the 26th Conference on Computer Applications and Quantitative Methods in Archaeology*, pages 175–184, Barcelona, Spain, 1998. Archaeopress.

A. Strehl and J. Ghosh. Relationship-based clustering and visualization for high-dimensional data mining. *INFORMS Journal on Computing*, 15(2):208–230, 2003.

Yin-Jing Tien, Yun-Shien Lee, Han-Ming Wu, and Chun-Houh Chen. Methods for simultaneously identifying coherent local clusters with smooth global patterns in gene expression profiles. *BMC Bioinformatics*, 9(155):1–16, 2008.

Dan Tsafrir, Ilan Tsafrir, Liat Ein-Dor, Or Zuk, Daniel A. Notterman, and Eytan Domany. Sorting points into neighborhoods (SPIN): Data analysis and visualization by ordering distance matrices. *Bioinformatics*, 21(10):2301–2308, 2005.

Joe H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

E.M. Wilkinson. Archaeological seriation and the travelling salesman problem. In F. Hodson, D. Kendall, and P. Tautu, editors, *Mathematics in the Archaeological and Historical Sciences*, pages 276–283. Edinburgh University Press, 1971.