

# Evaluating Reordering Strategies for Cluster Identification in Parallel Coordinates

## – Benchmark Dataset –

### LIST OF TABLES

1	Similarity and Dissimilarity Patterns of datasets with one cluster and increasing clutter. . . . .	2
2	Similarity and Dissimilarity Patterns of datasets with two clusters and increasing clutter. . . . .	3
3	Similarity and Dissimilarity Patterns of datasets with three clusters and increasing clutter. . . . .	4
4	Similarity and Dissimilarity Patterns of datasets with four clusters and increasing clutter. . . . .	5
5	Overview of real world datasets (part 1) . . . . .	6
6	Overview of real world datasets (part 2) . . . . .	7
7	Overview of real world datasets (source & modification) . . . . .	8

---

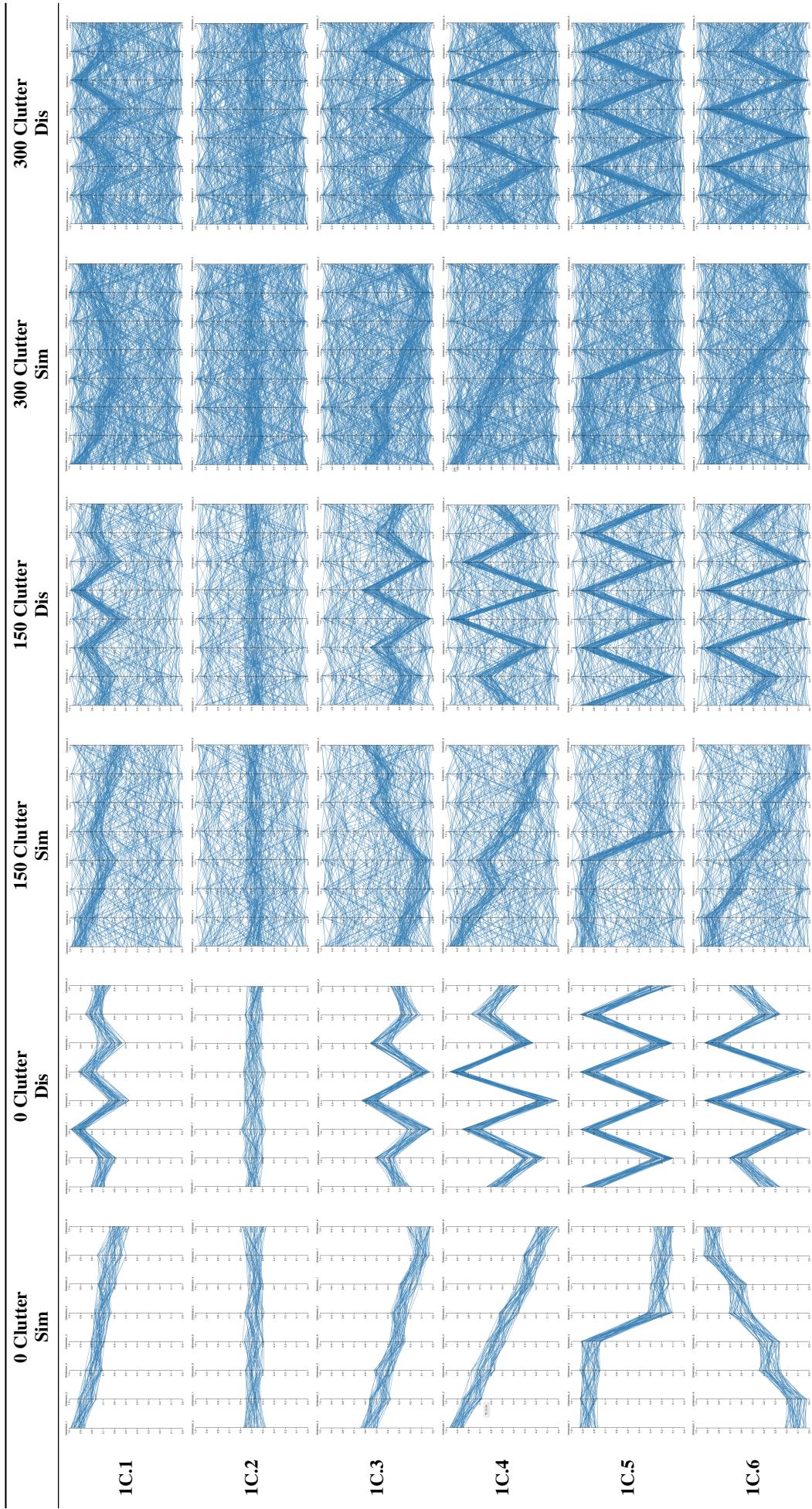


Table 1: Similarity and Dissimilarity Patterns of datasets with one cluster and increasing clutter.

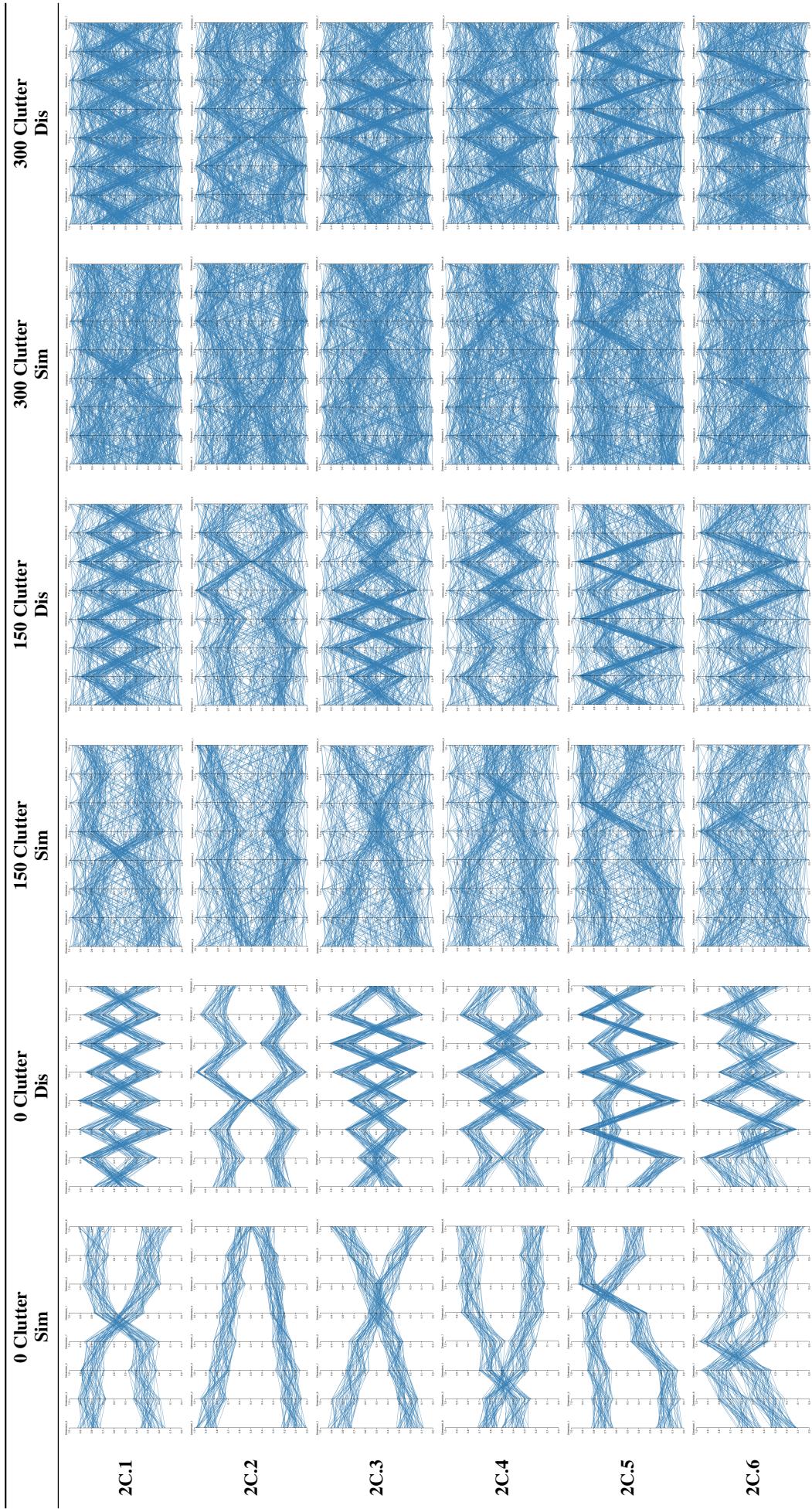


Table 2: Similarity and Dissimilarity Patterns of datasets with two clusters and increasing clutter.

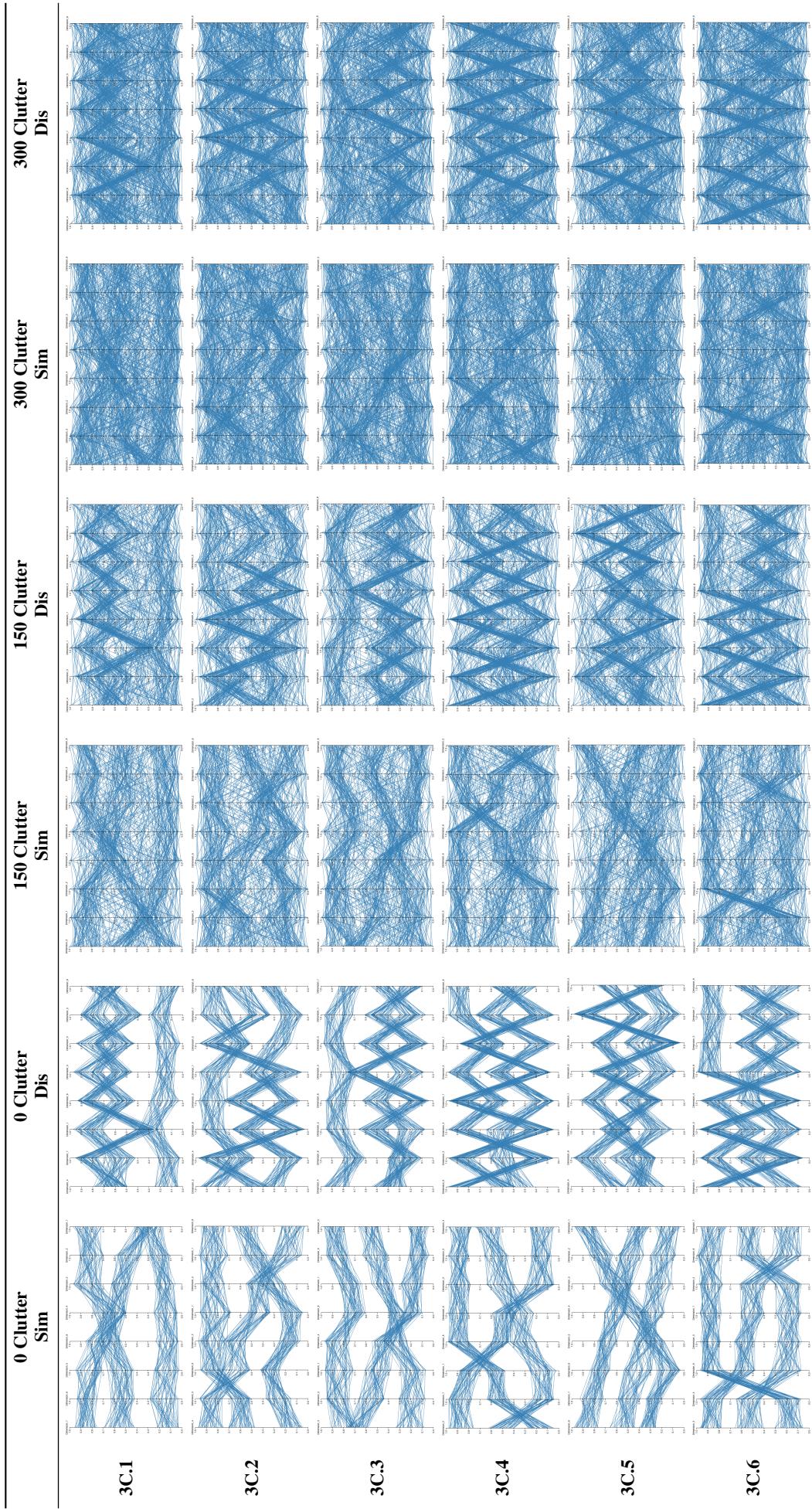


Table 3: Similarity and Dissimilarity Patterns of datasets with three clusters and increasing clutter.

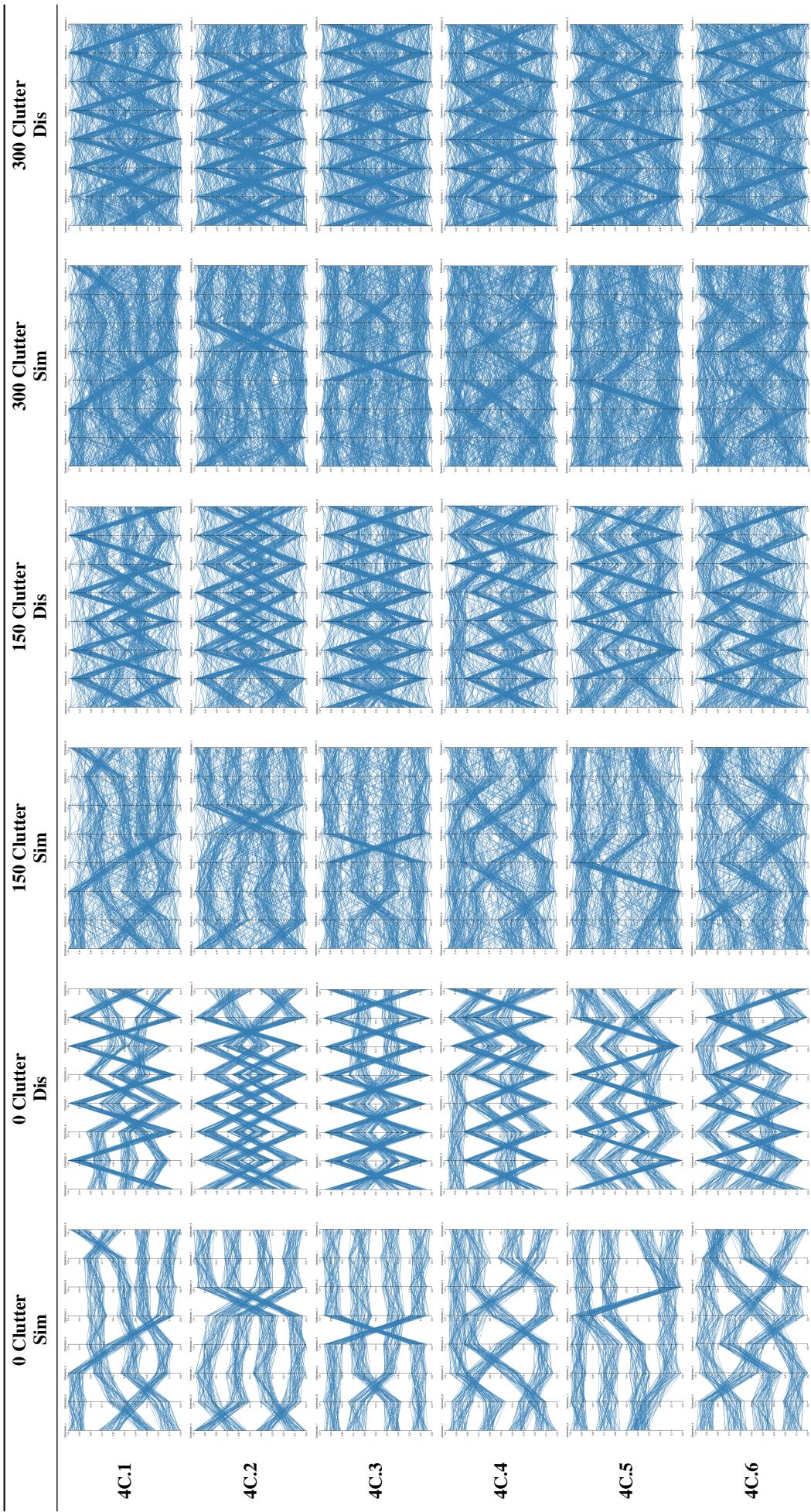


Table 4: Similarity and Dissimilarity Patterns of datasets with four clusters and increasing clutter.

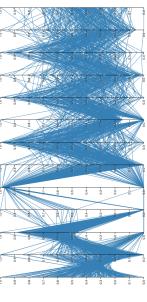
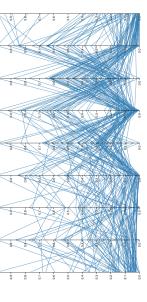
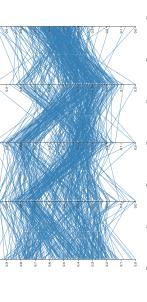
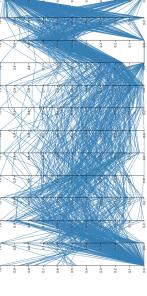
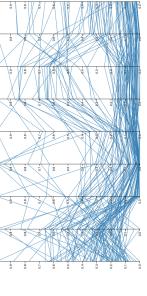
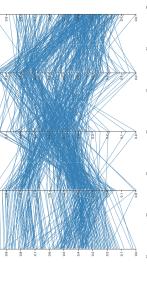
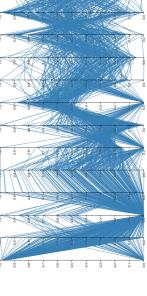
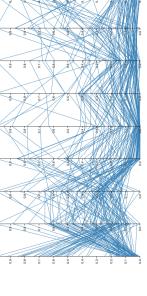
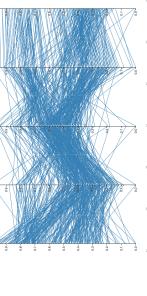
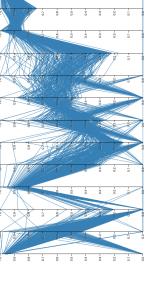
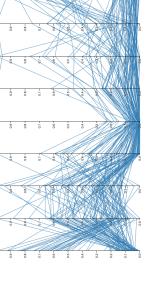
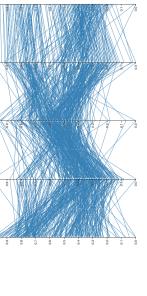
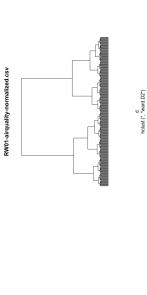
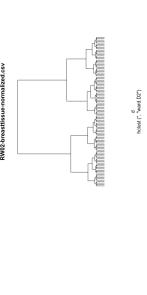
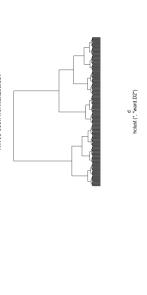
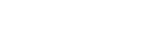
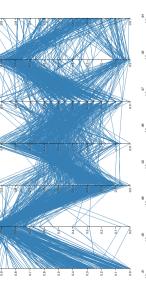
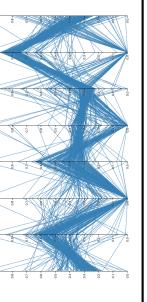
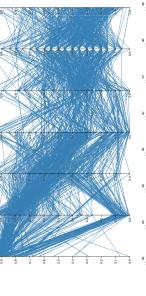
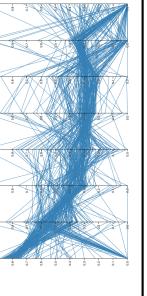
Name	# Dimension	# Record	Hierarchical Clustering			Modified Data Default	Modified Data Sim	Modified Data Dis
			Original Data Default	Modified Data Default	Modified Data Dis			
<b>RW01</b> air quality	13	283						
<b>RW02</b> breast tissue	9	105						
<b>RW03</b> ecoli	5	336						
<b>RW04</b> forest fires	7	515						
<b>RW05</b> glass	8	211						

Table 5: Overview of real world datasets (part 1)

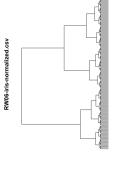
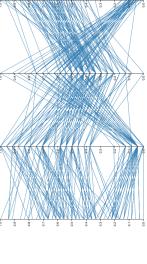
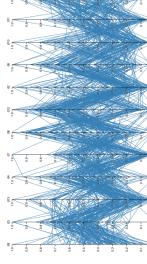
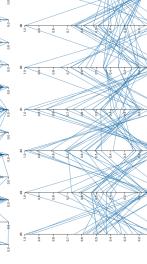
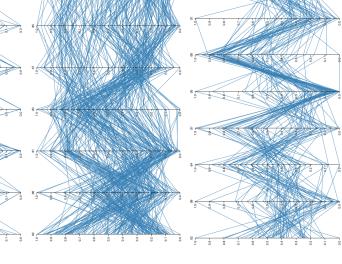
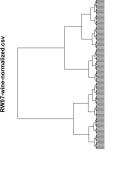
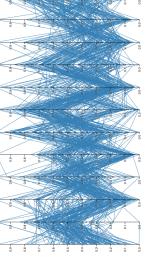
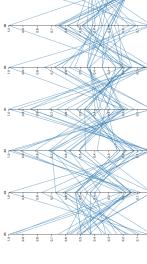
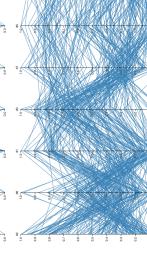
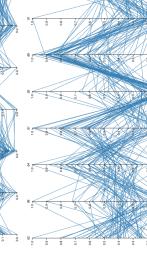
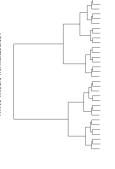
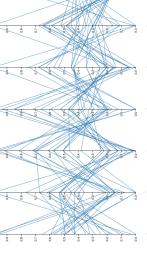
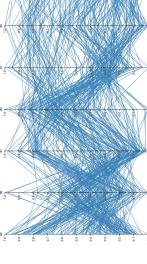
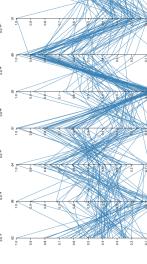
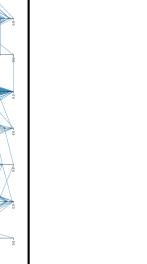
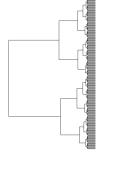
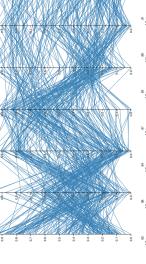
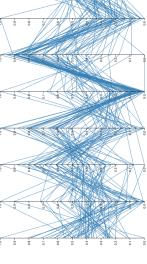
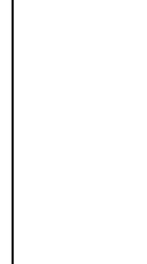
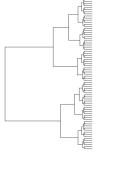
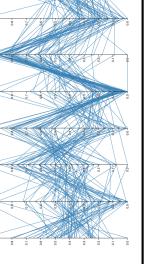
Name	# Dimension	# Record	Hierarchical Clustering	Original Data Default	Modified Data Default	Modified Data Sim	Modified Data Dis
<b>RW06</b> iris	4	150					
<b>RW07</b> wine	13	178					
<b>RW08</b> mt cars	7	32					
<b>RW09</b> seeds	7	210					
<b>RW10</b> stone flakes	8	73					

Table 6: Overview of real world datasets (part 2)

Name	Source	# Removed Dimensions	# Removed Records	# Dimension	# Records
Air Quality	<a href="https://archive.ics.uci.edu/ml/datasets/Air+quality">https://archive.ics.uci.edu/ml/datasets/Air+quality</a>	2	9075 <sup>1</sup>	13	283
Breast Tissue	<a href="http://archive.ics.uci.edu/ml/datasets/breast+tissue">http://archive.ics.uci.edu/ml/datasets/breast+tissue</a>	0	0	9	105
Ecoli	<a href="https://archive.ics.uci.edu/ml/datasets/Ecoli">https://archive.ics.uci.edu/ml/datasets/Ecoli</a>	2	0	5	336
Forest Fires	<a href="https://archive.ics.uci.edu/ml/datasets/Forest+Fires">https://archive.ics.uci.edu/ml/datasets/Forest+Fires</a>	6	2	7	515
Glass	<a href="https://archive.ics.uci.edu/ml/datasets/Glass+Identification">https://archive.ics.uci.edu/ml/datasets/Glass+Identification</a>	1	3	8	211
Iris	<a href="https://archive.ics.uci.edu/ml/datasets/Iris">https://archive.ics.uci.edu/ml/datasets/Iris</a>	0	0	4	150
Wine	<a href="https://archive.ics.uci.edu/ml/datasets/Wine">https://archive.ics.uci.edu/ml/datasets/Wine</a>	0	0	13	178
Mt Cars	<a href="https://www.kaggle.com/ruiromanini/mtcars">https://www.kaggle.com/ruiromanini/mtcars</a>	4	0	7	32
Seeds	<a href="https://archive.ics.uci.edu/ml/datasets/seeds">https://archive.ics.uci.edu/ml/datasets/seeds</a>	0	0	7	210
Stone Flakes	<a href="https://archive.ics.uci.edu/ml/datasets/StoneFlakes">https://archive.ics.uci.edu/ml/datasets/StoneFlakes</a>	0	6	8	73

Table 7: Overview of real world datasets (source & modification)

<sup>1</sup>The PCPs in our study do not use any transparency for the polylines which results in huge overplotting. Therefore, we sampled the airquality dataset for our study to keep all dataset below 600 records.