# Tool for Visual Cluster Analysis and Consensus Clustering

Christian Permann

Faculty of Computer Science, University of Vienna,
Währinger Straße 29, 1090 Vienna

TBD.04.2020

# Introduction

Clustering:

- Grouping data-points such that their underlying relationships are reflected
- Gaining knowledge through this grouping

> The process of clustering is not done when a solution is computed,
> but when the researcher involved:
> "... **evaluated**, **understood** and **accepted** the patterns." (Chen and Liu [1])

Challenges:

- Many possibilities for clustering:
  - Algorithms/Parameters/Assumptions
- Choice and interpretation of solution is difficult

# Related Work: Clustering

There is a vast amount of clustering techniques, including:

- Partition-based methods (KMeans-like algorithms)
- Hierarchy-based methods (e.g. Joining of Sets/Linking)
- Density-based methods (e.g. DBSCAN/OPTICS)
  - Many more...

# Related Work: Visual Frameworks

- ClusterVision
  - Ranking solutions according to a combination of quality metrics
  - Choosing from the highest ranked ones
- VISTA
  - In-depth analysis of individual solutions
  - Possibilities for relabeling of points (ClusterMap)
- Simple Visualizations
  - Included in most data-analysis tools
  - Scatter plots, bar charts, etc.

# Related Work: Consensus Clustering

Combining clustering results may yield a better solution:

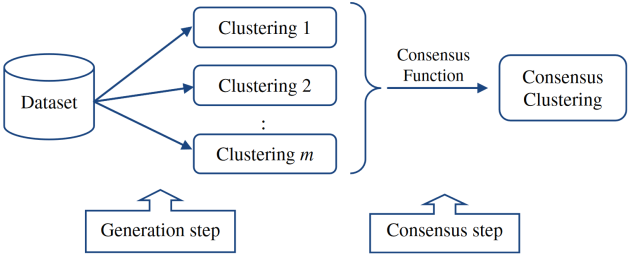


Figure 1: Workflow for generating consensus clusterings [2, p. 340]

# Idea of our Tool: Facilitating clustering exploration

How can we assist users in exploring clustering results?

- Visualizing individual results
  - Scatter plot (matrices)/kernel density estimation
  - Dimensionality reduction
- Visualizing similarities between results
  - OPTICS meta-clustering
  - Heat maps
  - Multi-Dimensional-Scaling to approximate solution space

# Idea of our Tool: Gathering more Information

Can we gain additional knowledge from multiple computed solutions?

- Previous frameworks only try to select the best one
  - Additional information lost
  - Difficult to objectively identify best one
- Consensus clustering
  - Can combine solutions or groups of solutions

Idea:

- Combine group of robust solutions into one

# Idea of our Tool: Ease of Use

- bla

# The Tool

Three main parts:

- Data-View
  - Loading/Saving/Creating data
  - Cleaning up data
  - Visualizing data
- Workflow-View
  - Creating clustering workflows
  - Defining parameters
- Meta-View
  - Visualizing clusterings and meta-clusterings
  - Selecting or creating final results (& consensus clustering)

# The Tool: Data-View

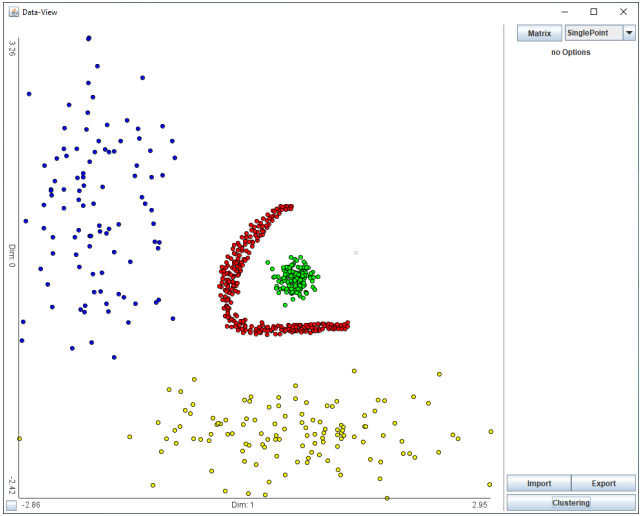


Figure 2: Data-View of the Tool

# The Tool: Workflow-View

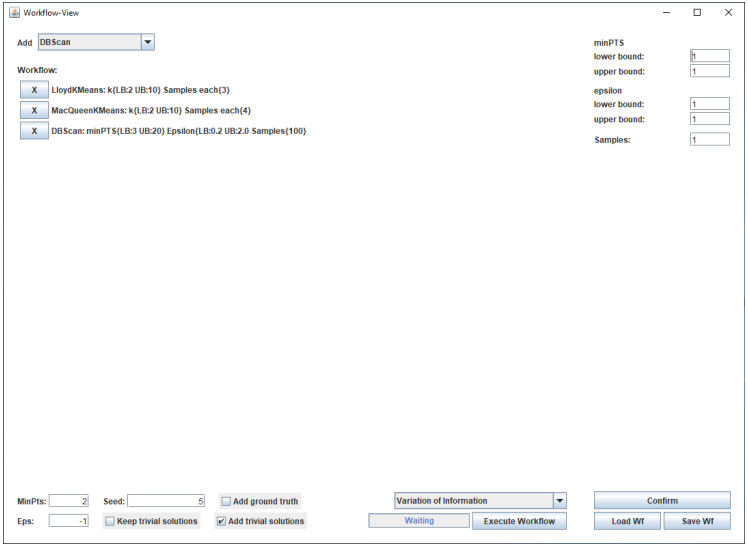


Figure 3: Workflow-View of the Tool

# The Tool: Meta-View

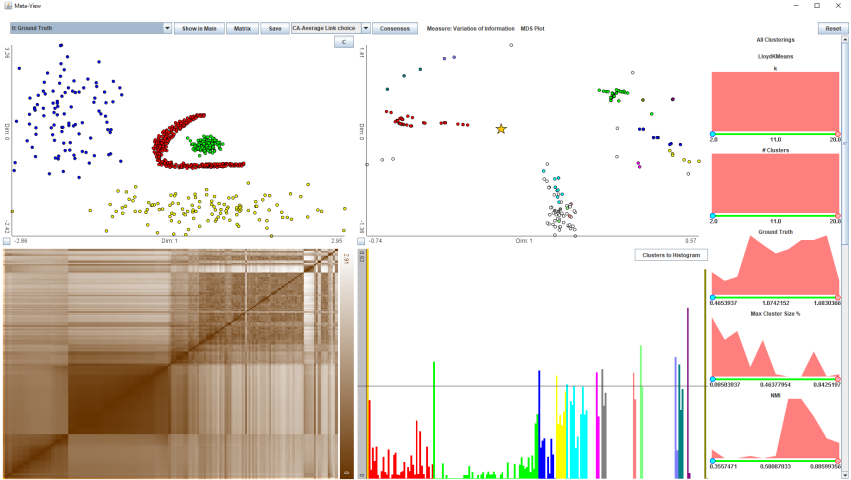


Figure 4: Meta-View of the Tool

# Implementation

- bla

# Tests

- bla

# Future Work

- bla

# Conclusion

- bla

# References


Keke Chen and Ling Liu. "A visual framework invites human into clustering process". In: Aug. 2003, pp. 97 –106. ISBN: 0-7695-1964-4. DOI: 10.1109/SSDM.2003.1214971.

Sandro Vega-Pons and José Ruiz-Shulcloper. "A Survey of Clustering Ensemble Algorithms.". In: *International Journal of Pattern Recognition and Artificial Intelligence* 25 (2011), pp. 337–372. DOI: 10.1142/S0218001411008683.