Technische Universität Berlin

Clustering of complex datasets

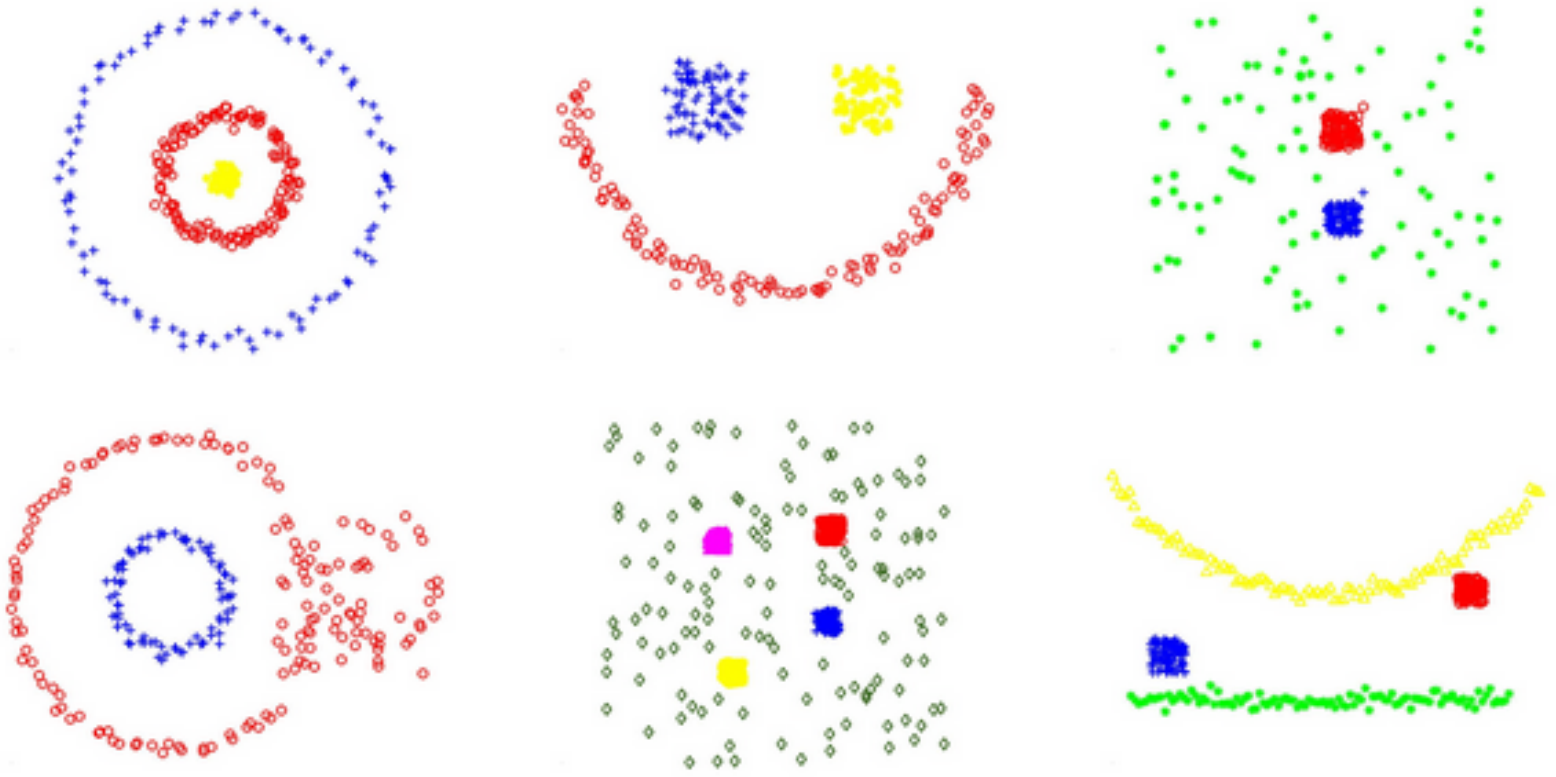
Milestone Presentation

Ahmed Tidjani, Fabio Buso, Paul Velthuis & Zahin Azher
Service-centric Networking | Tu Berlin | May 11th, 2016

TELEKOM INNOVATION LABORATORIES

Service-centric NETWORKING

Complex Datasets



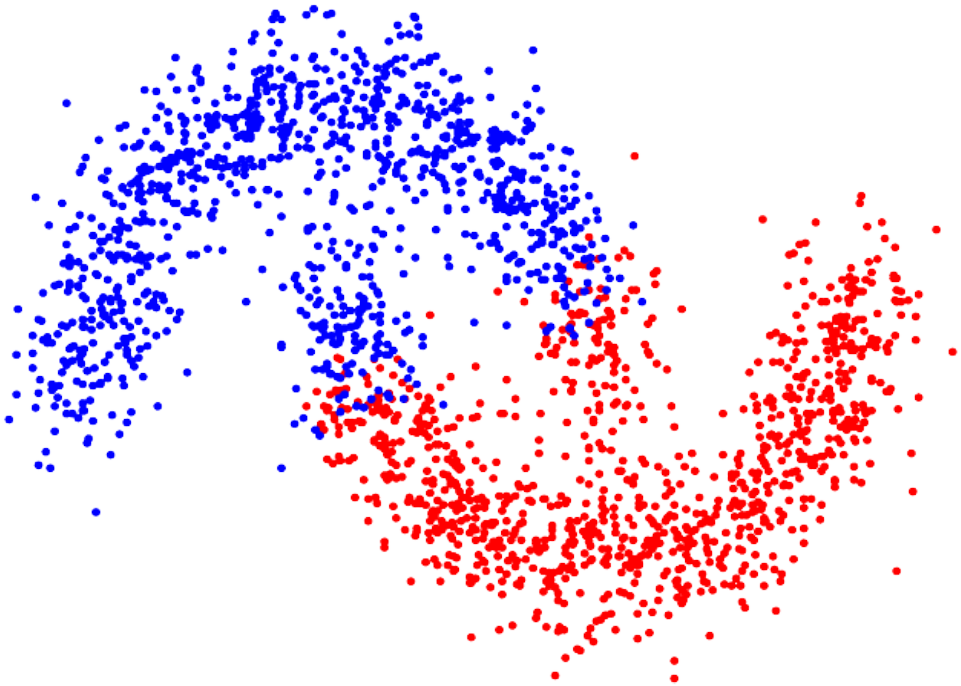
[2]

Clustering

Grouping a set of objects in a way that similar objects are in the same cluster.

Types:

- Well Separated
- Centroid Based
- Density Based
- [...]



kMeans Clustering

Self-tuning Spectral Clustering[2]

Requires:

- Max number of Clusters

Operates a dimensionality reduction using the eigenvectors of the affinity matrix.

Exploits the eigenvectors structure to determine the optimal number of clusters.

$$\hat{A}_{ij} = e^{\left(\frac{-d^2(s_i, s_j)}{\sigma_i \sigma_j}\right)}$$

$$\sigma_i = d(s_i, s_K), K = 7$$

Self-tuning Spectral Clustering[2] - contd

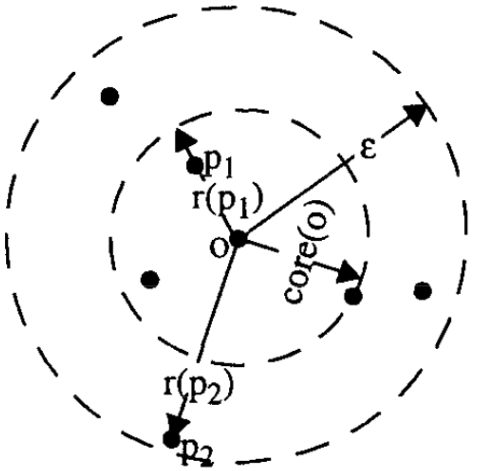
Implementations already available in: C++ and Matlab

Objective:

- Investigate the construction of the affinity matrix
- Comparison with OPTICS
- Performance Analysis

OPTICS [1]

Technische Universität Berlin



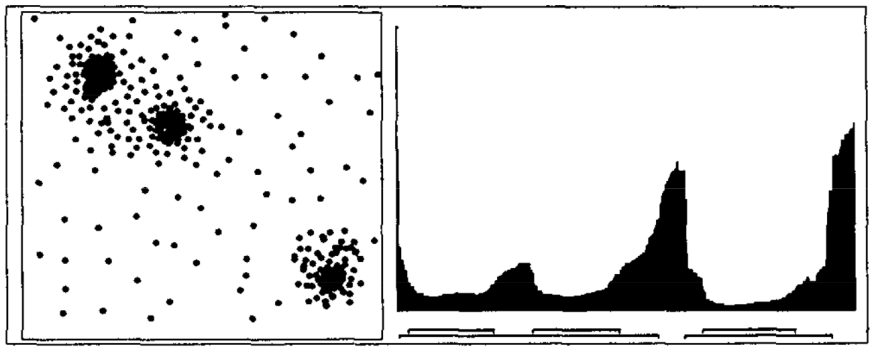
Density Based

Requires:

- Min # Points in a Cluster (MinPts),
- Max Neighborhood radius (ε),
- Degree of Steepness (ξ).

Produces Reachability Plot

Automatic cluster extraction from the plot



OPTICS[1] - contd

Implementations already available in: Java, Python, R and Matlab.

Objective:

- Comparison of the different Implementations, with the focus on the automatic extraction
- Performance Analysis

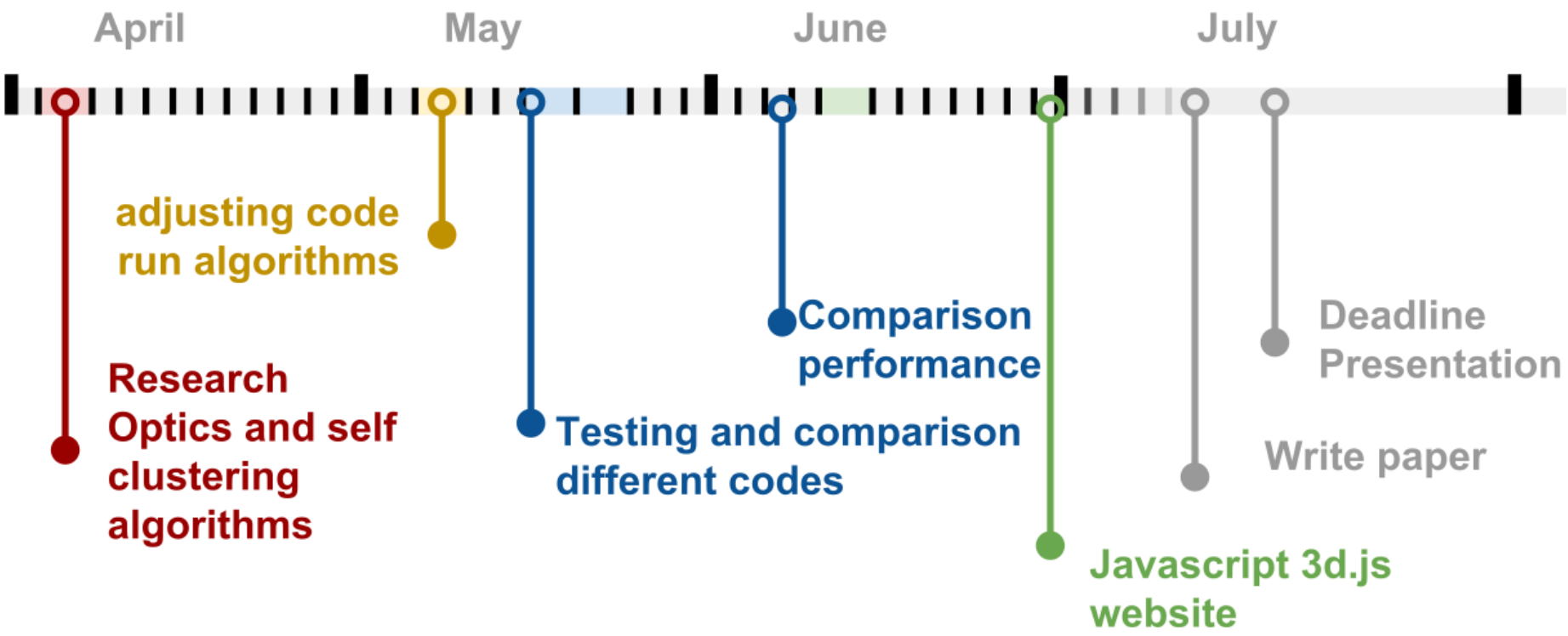
Problem:

- Not all implementations extract automatically the clusters

Technische Universität Berlin

ORGANIZATION

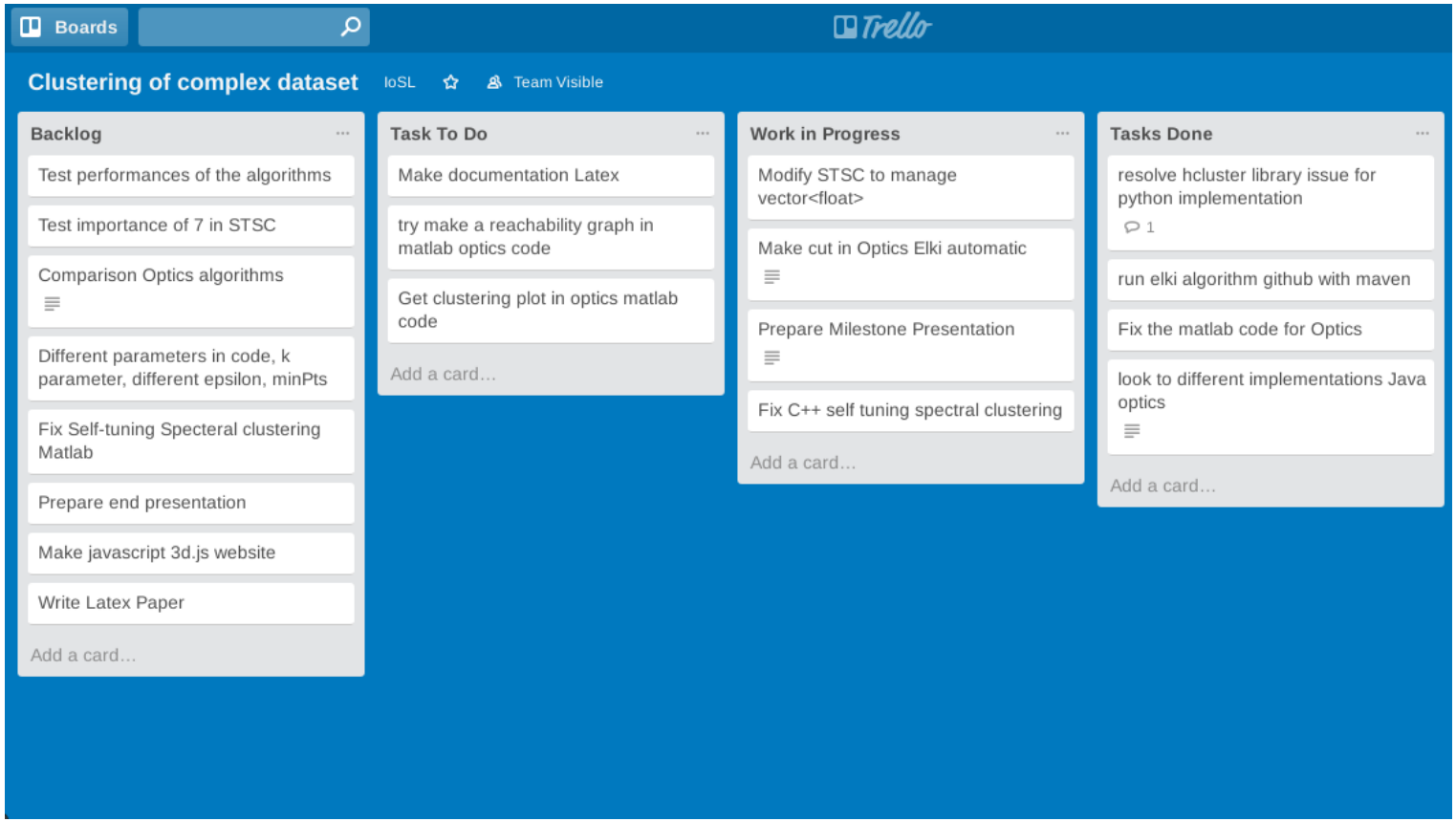
Timeline


April
May
June
July
adjusting code
run algorithms
Research
Optics and self
clustering
algorithms
Testing and comparison
different codes
Comparison
performance
Deadline
Presentation
Write paper
Javascript 3d.js
website

Tools

Technische Universität Berlin

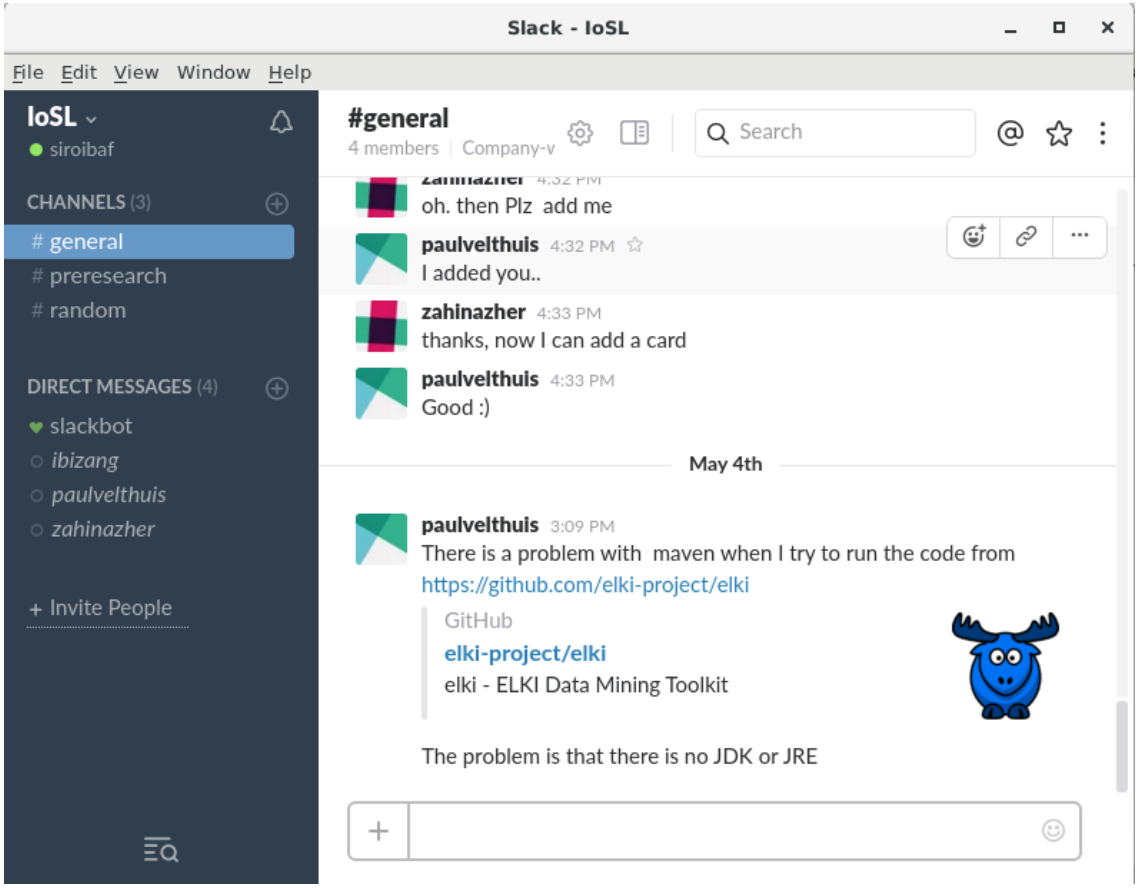
Trello



Tools

Trello

Slack



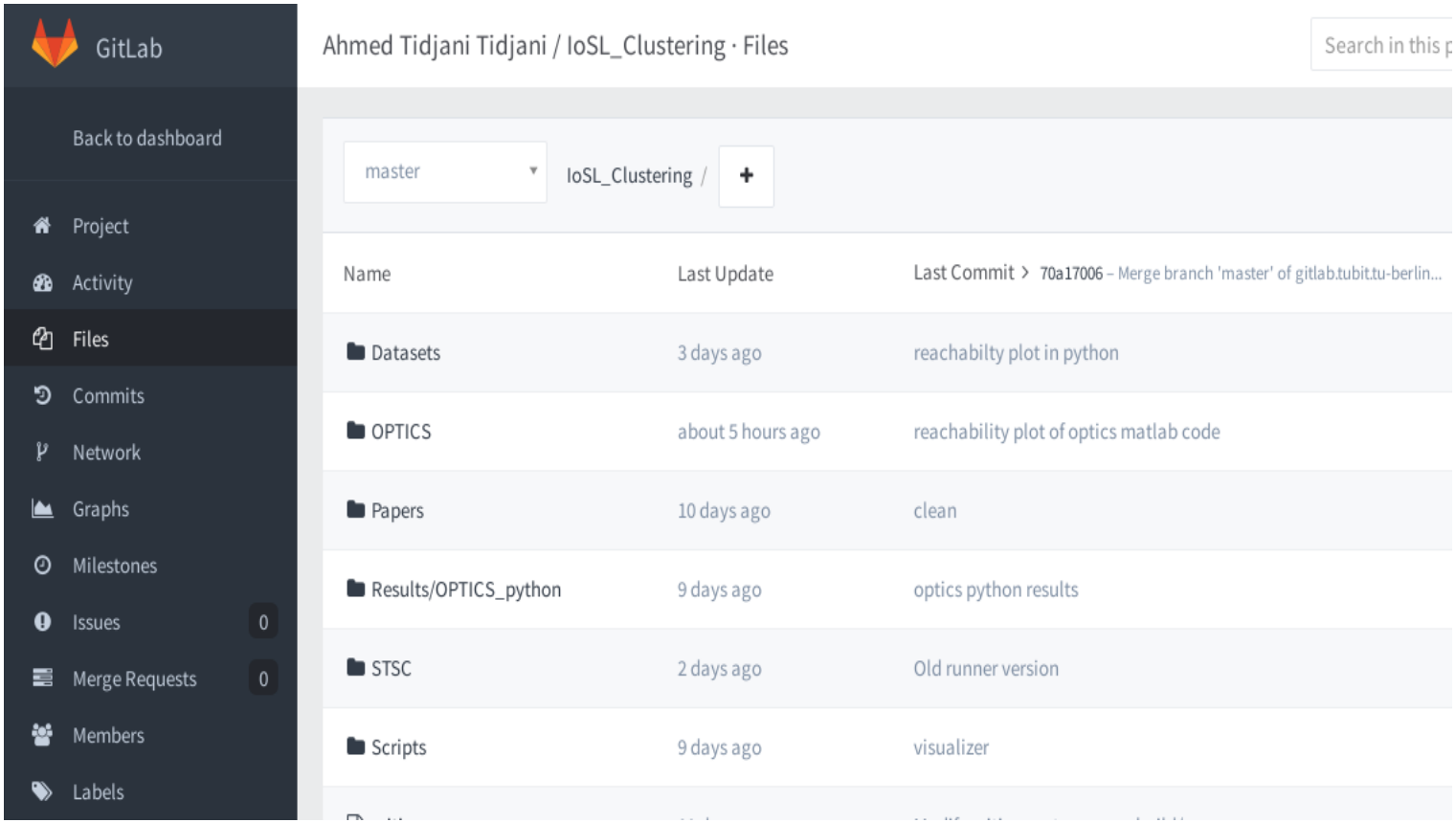
Tools

Technische Universität Berlin

Trello

Slack

GitLab



Tools

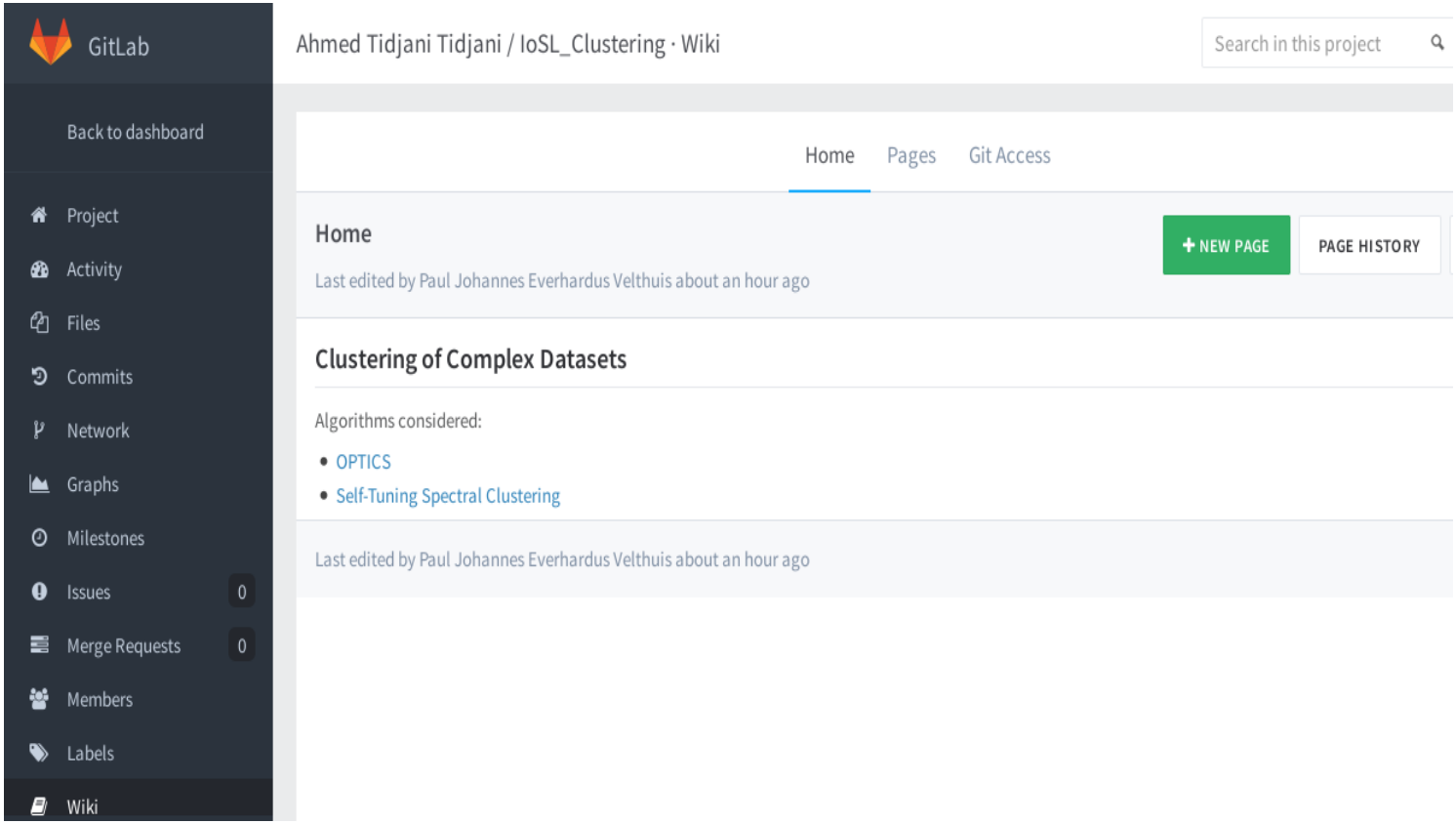
Technische Universität Berlin

Trello

Slack

GitLab

GitLab - Wiki



Technische Universität Berlin

THANK YOU FOR YOUR ATTENTION
QUESTIONS?

References

[1] Ankerst, M., Breunig, M. M., Kriegel, H. P., & Sander, J. (1999, June). OPTICS: ordering points to identify the clustering structure. In ACM Sigmod Record (Vol. 28, No. 2, pp. 49-60). ACM.

[2] Zelnik-Manor, L., & Perona, P. (2004). Self-tuning spectral clustering. InAdvances in neural information processing systems (pp. 1601-1608).