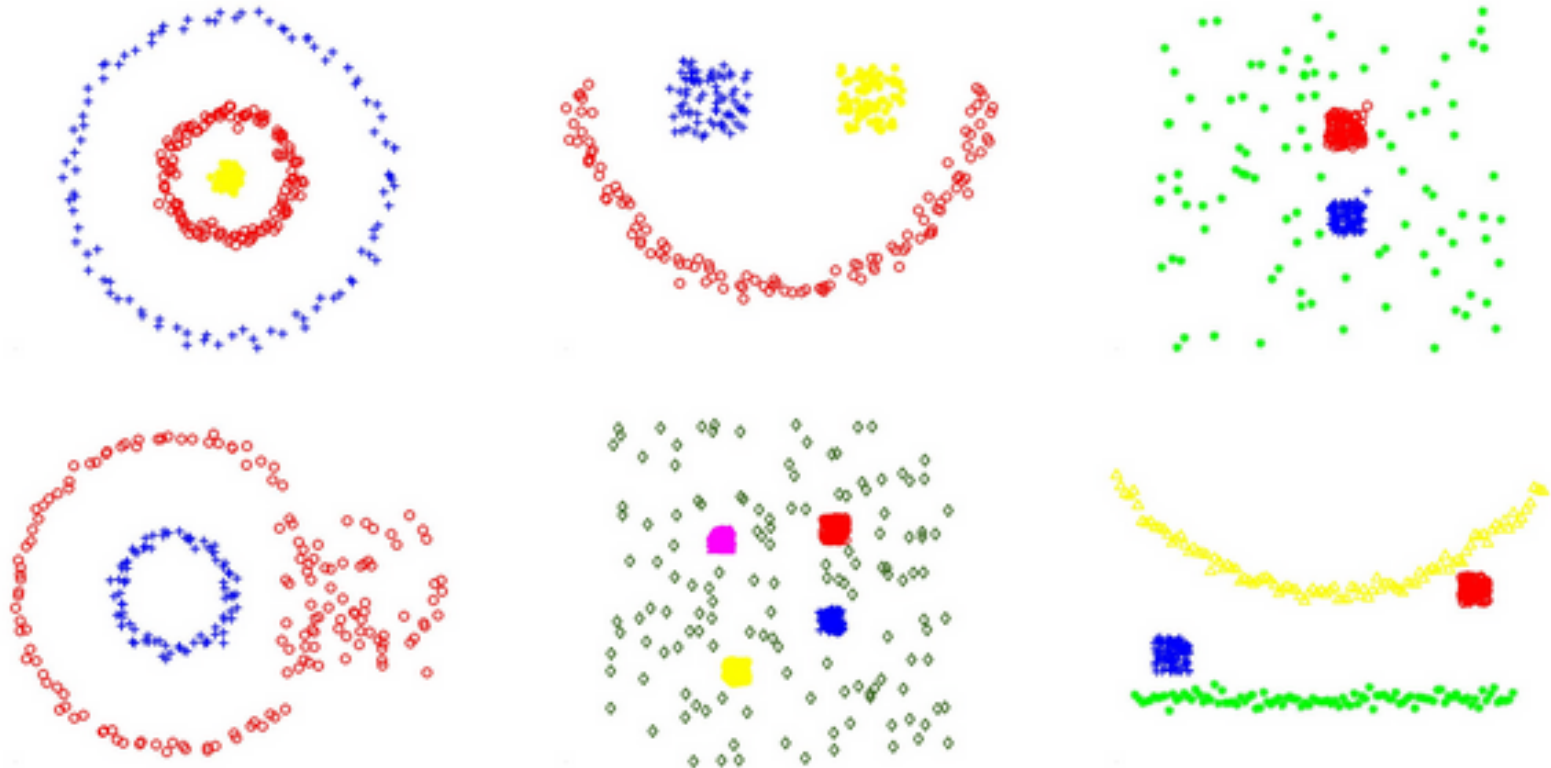


Clustering of complex datasets

Milestone Presentation

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

Complex Datasets



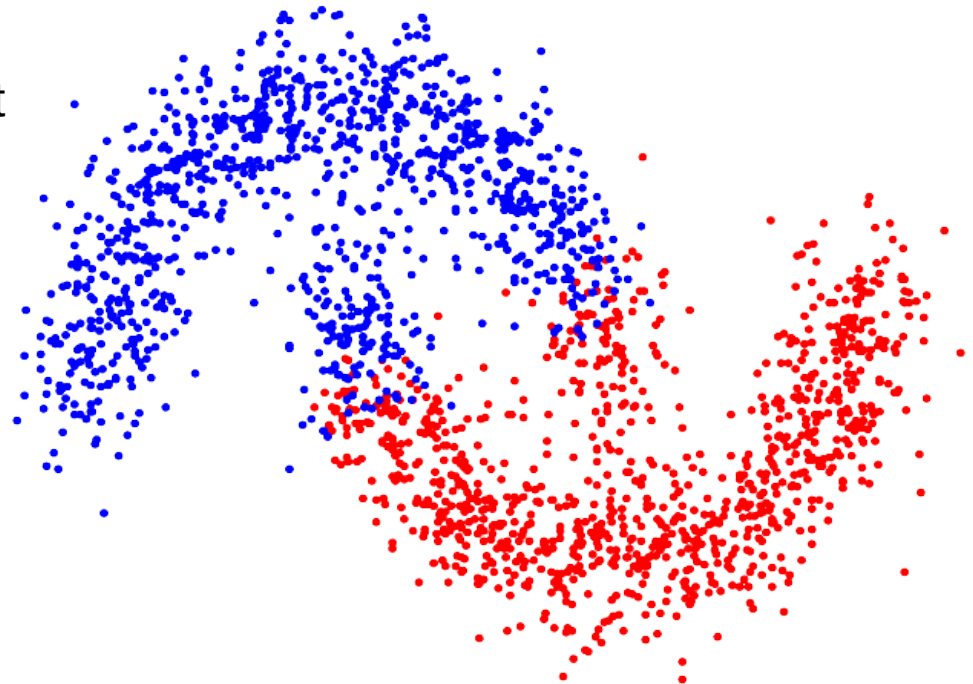
[2]

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

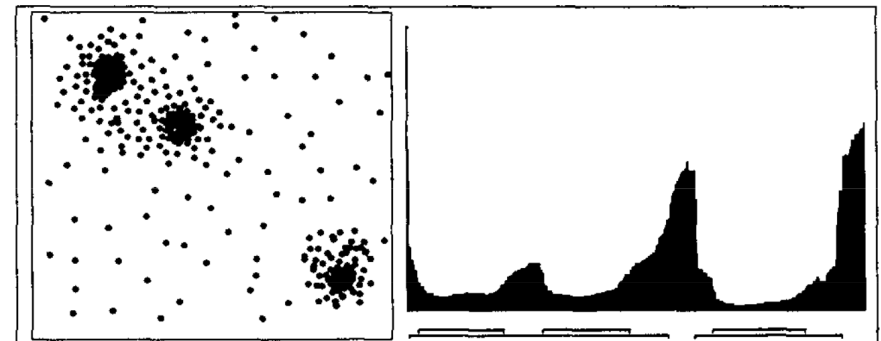
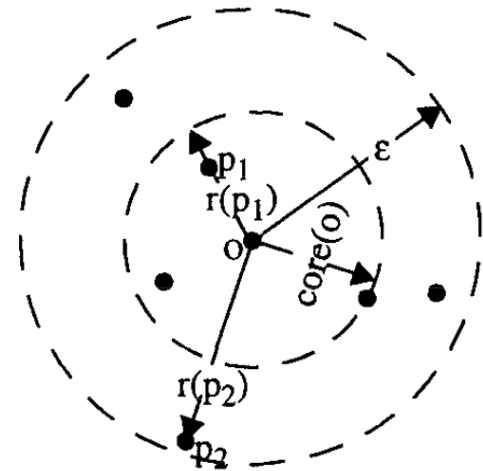
Density Based

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:

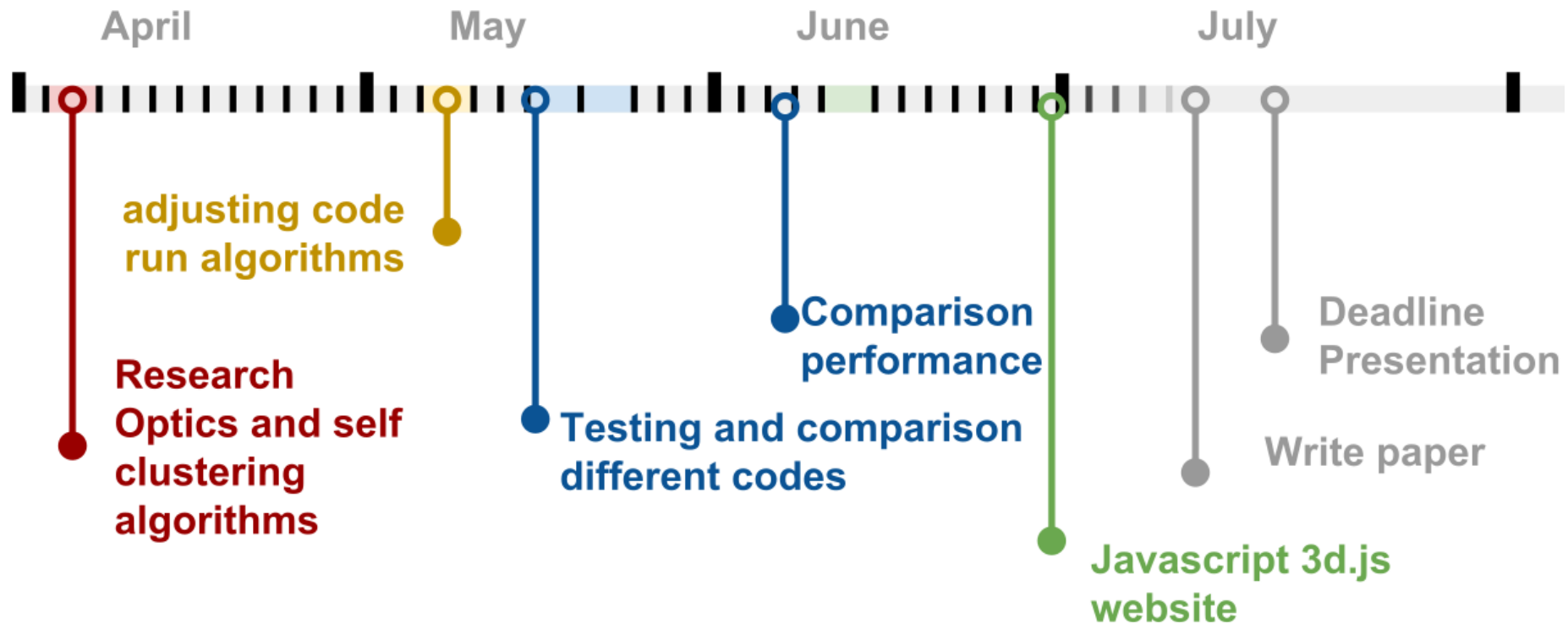
- Comparison of the different Implementations, with the focus on the automatic extraction
- Performance Analysis

Problem:

- Not all implementations extract automatically the clusters

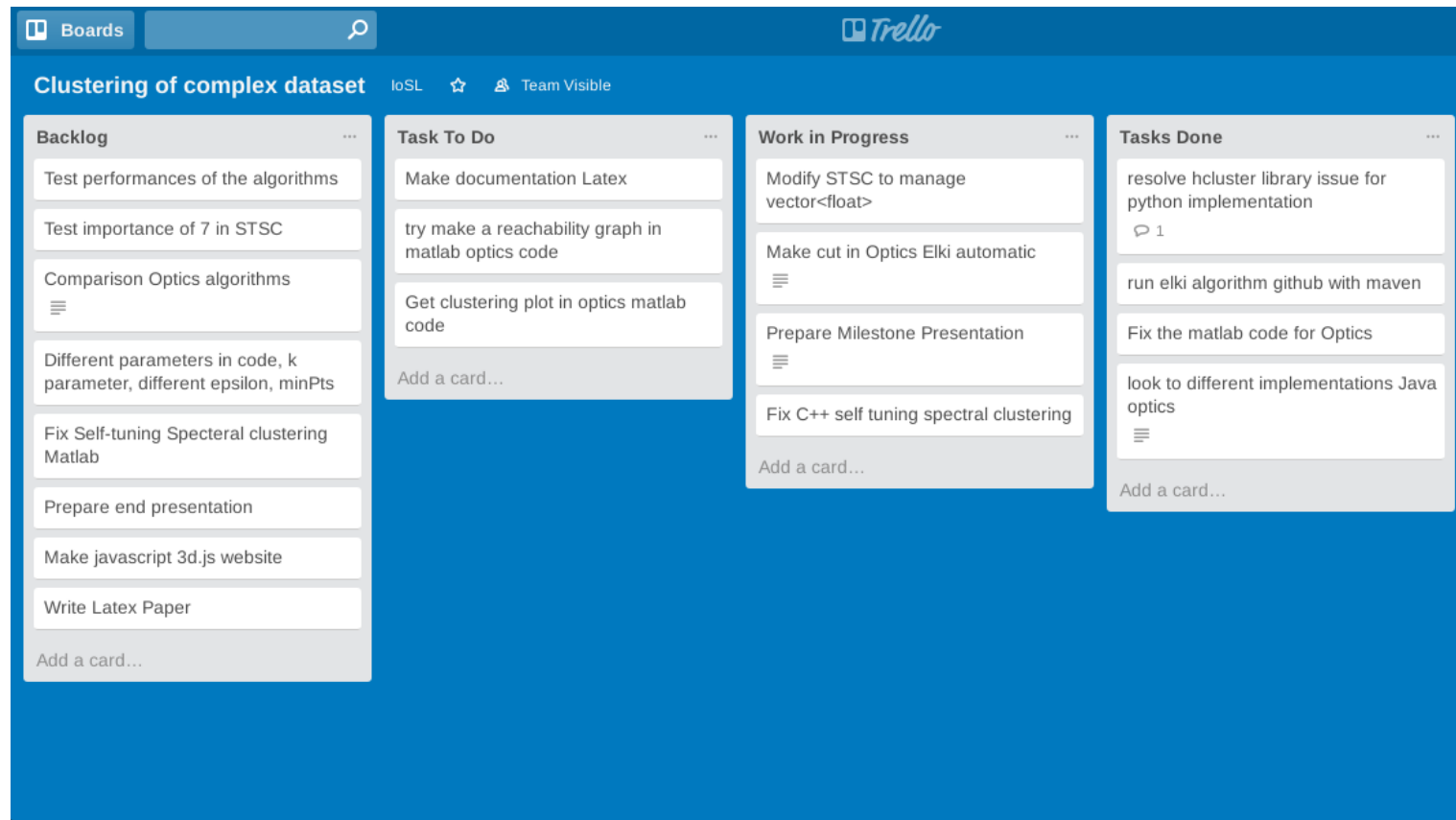
ORGANIZATION

Timeline



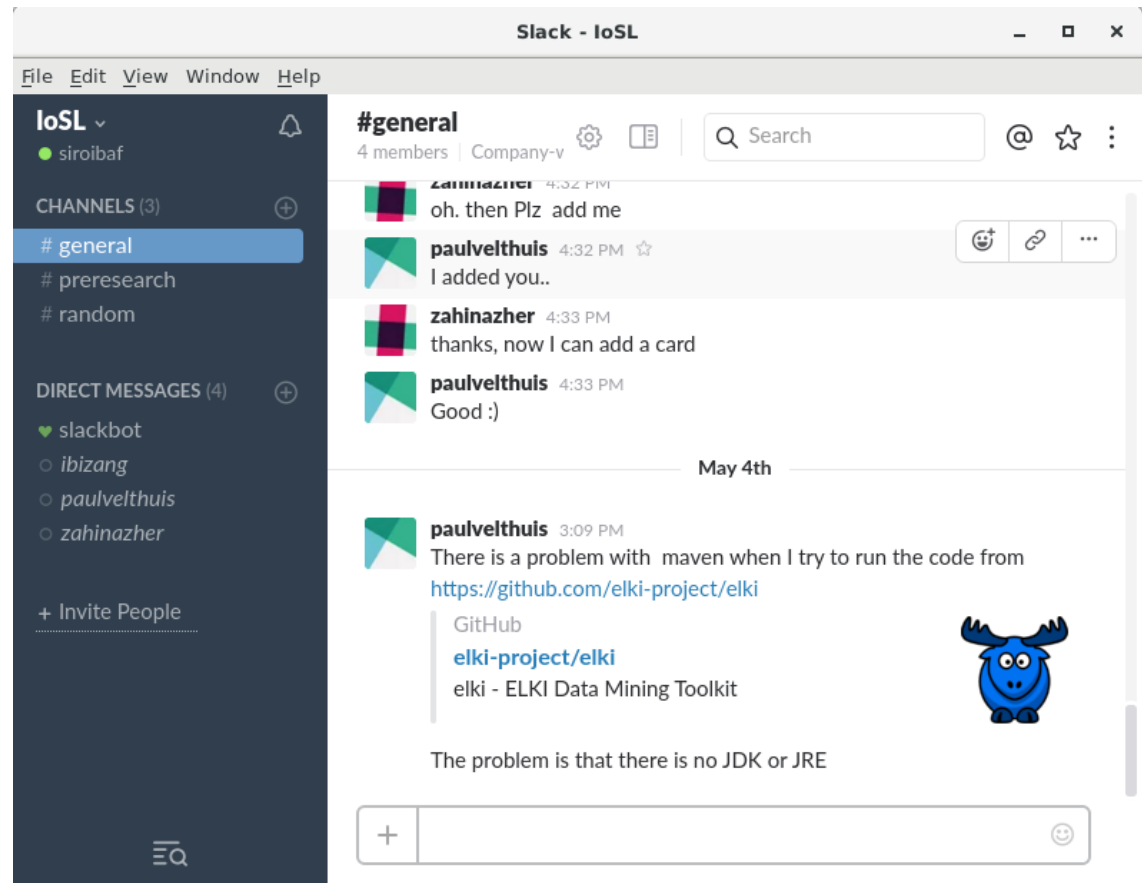
Tools

Trello



Tools

Trello
Slack



Tools

Trello
Slack
GitLab

The screenshot shows the GitLab web interface for a project named 'Ahmed Tidjani Tidjani / IoSL_Clustering'. The left sidebar contains navigation links: 'Back to dashboard', 'Project', 'Activity', 'Files' (selected), 'Commits', 'Network', 'Graphs', 'Milestones', 'Issues' (0), 'Merge Requests' (0), 'Members', and 'Labels'. The main content area displays the 'Files' view for the 'master' branch. It shows a table of files and folders with columns for 'Name', 'Last Update', and 'Last Commit'. The files listed are: 'Datasets' (updated 3 days ago), 'OPTICS' (updated about 5 hours ago), 'Papers' (updated 10 days ago), 'Results/OPTICS_python' (updated 9 days ago), 'STSC' (updated 2 days ago), and 'Scripts' (updated 9 days ago). The 'Last Commit' column shows the commit hash '70a17006' and the message 'Merge branch 'master' of gitlab.tubit.tu-berlin...'. A search bar is visible in the top right corner.

Name	Last Update	Last Commit
Datasets	3 days ago	70a17006 - Merge branch 'master' of gitlab.tubit.tu-berlin...
OPTICS	about 5 hours ago	reachability plot of optics matlab code
Papers	10 days ago	clean
Results/OPTICS_python	9 days ago	optics python results
STSC	2 days ago	Old runner version
Scripts	9 days ago	visualizer

Tools

Trello

Slack

GitLab

GitLab - Wiki

The screenshot shows the GitLab Wiki interface. On the left is a dark sidebar with the GitLab logo and a list of navigation items: 'Back to dashboard', 'Project', 'Activity', 'Files', 'Commits', 'Network', 'Graphs', 'Milestones', 'Issues' (with a badge showing 0), 'Merge Requests' (with a badge showing 0), 'Members', 'Labels', and 'Wiki'. The main content area is titled 'Ahmed Tidjani Tidjani / IoSL_Clustering · Wiki' and includes a search bar. Below the title are tabs for 'Home', 'Pages', and 'Git Access'. The 'Home' tab is active, showing a 'Home' section with a green '+ NEW PAGE' button and a 'PAGE HISTORY' button. The text indicates the page was 'Last edited by Paul Johannes Everhardus Velthuis about an hour ago'. The main content is titled 'Clustering of Complex Datasets' and lists 'Algorithms considered:' with two bullet points: 'OPTICS' and 'Self-Tuning Spectral Clustering'. At the bottom of the content area, it again states 'Last edited by Paul Johannes Everhardus Velthuis about an hour ago'.

Clustering of Complex Datasets

Ahmed Tidjani, Fabio Buso, Paul Velthuis & Zahin Azher

**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. In Advances in neural information processing systems (pp. 1601-1608).