Visualisation of Overlapping Sets and Clusters with Euler Diagrams

Paolo Simonetto

LaBRI, Université Bordeaux 1
INRIA Bordeaux Sud-Ouest

2nd December 2011

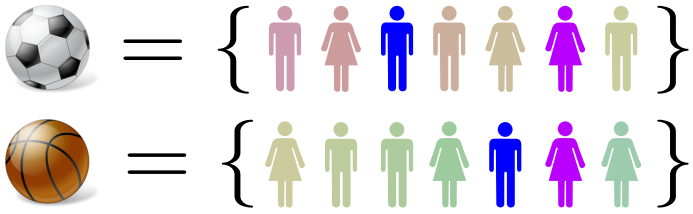
The Concepts in the Title

Visualisation of

The science that studies the visual representation and analysis of data.

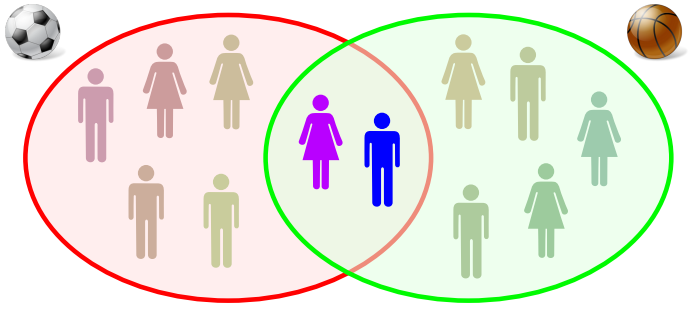
Overlapping Sets and Clusters

Groups of elements that might share some of their elements.



with Euler diagrams

The most natural graphical representation for sets.



The Problem Addressed

We aim to

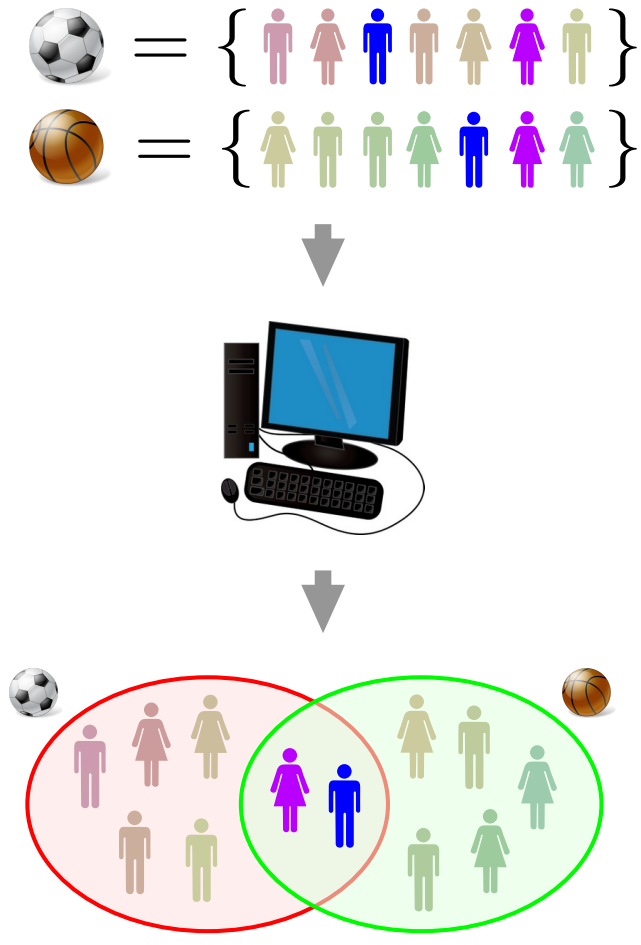
design an automatic method that:

- visually represents data,
- depicts overlapping sets,
- uses Euler diagrams.

Main task

Generation of the representation:

- identify the overlaps,
- displace the elements,
- identify the set boundaries.



Outline of the Presentation

Introduction

- basic Euler diagram theory,
- related work.

Euler Representations

- basic graph drawing theory,
- diagram construction.

ImPrEd

- introduce PrEd,
- explain improvements,
- present results.

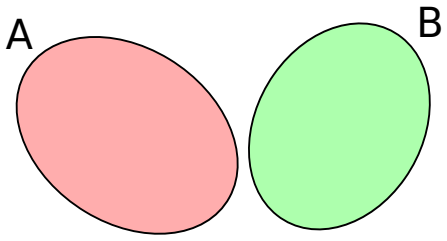
Software and Examples

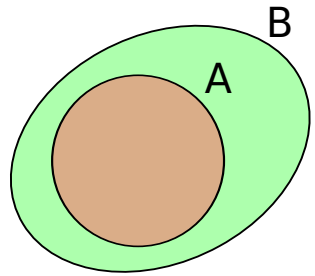
- present implementation,
- discuss diagrams produces.

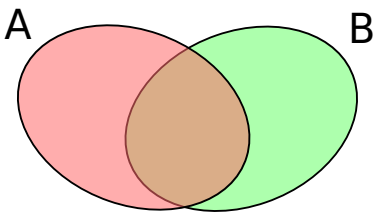
Basic Euler Diagram Concepts

Interpretation

- every set is associated to a region,
- disjoint regions = disjoint sets,
- included regions = subsets,
- partial overlaps = intersections.

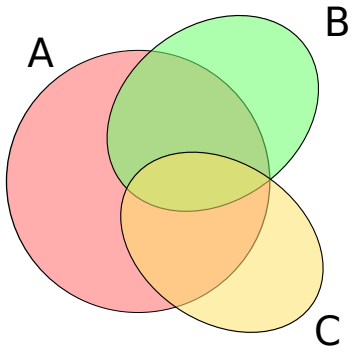


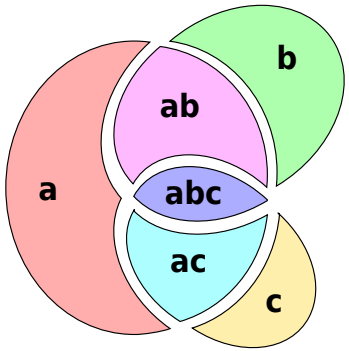




Definitions

- *clusters* = sets to be represented,
- *zones* = intersections formed.





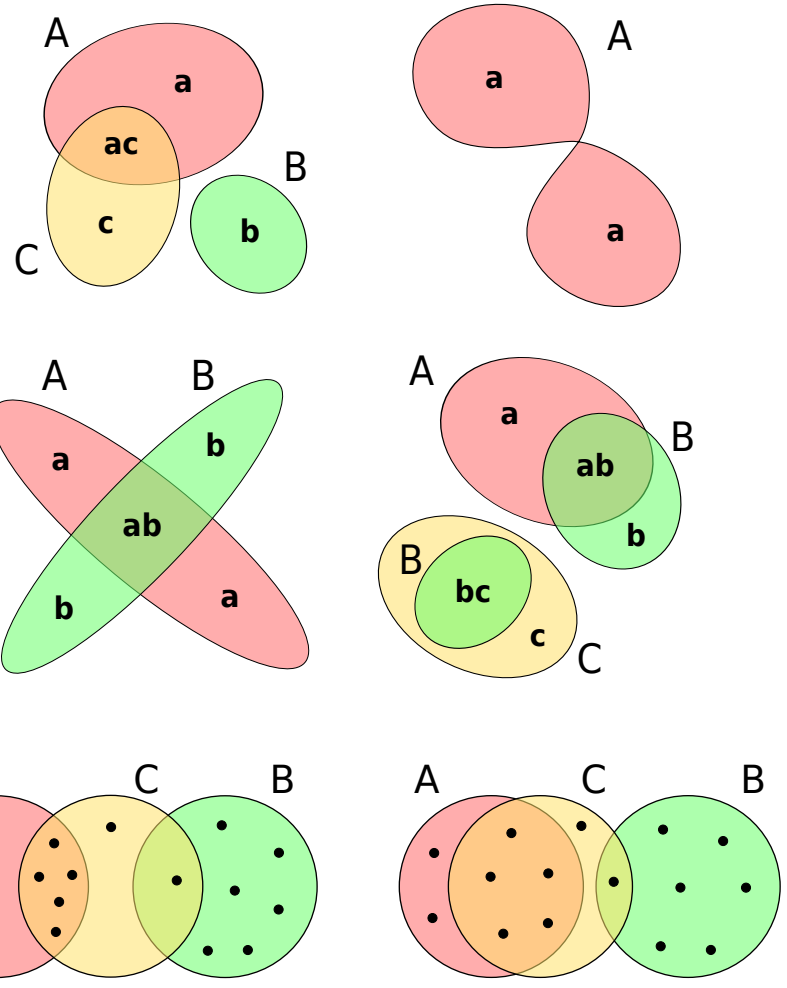
Properties of Euler Diagrams

Constraints

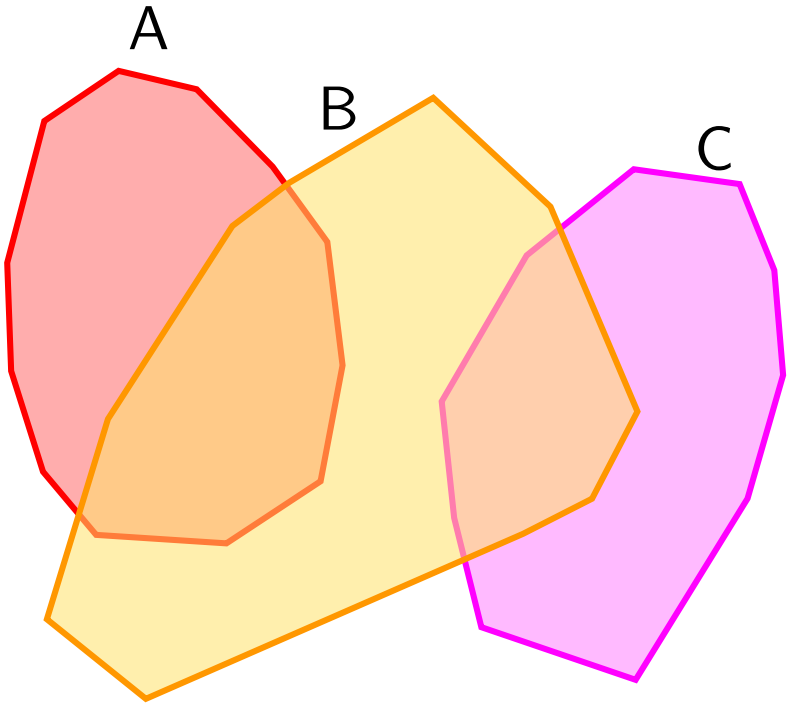
- show the expressed zones,
- use simple cluster curves,
- no disconnected zones,
- no disconnected clusters.

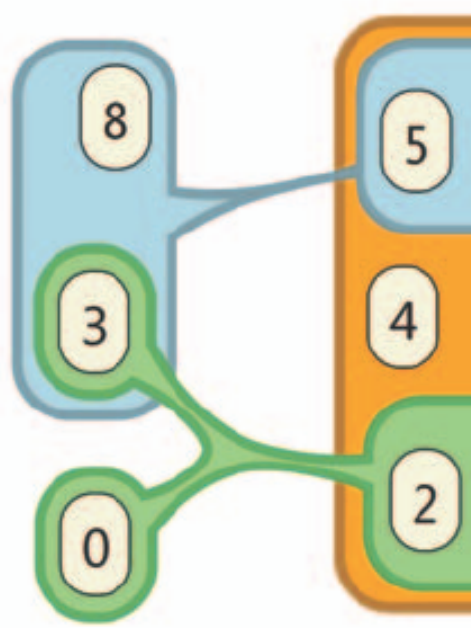
Features

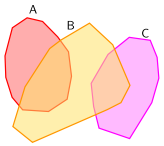
- depiction of elements,
- colours and shading,
- area-proportionality.



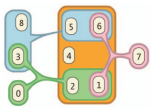
Euler Diagrams in Visualisation







ED Generation



Untangled ED

Bubble sets

Euler diagrams generation

[Flower, Howse, Stapleton, Rodgers, Fish, 2002–2011]

[Verroust, Viaud, 2004] [Chow, 2007] Untangled Euler diagrams

[Riche, Dwyer, 'Untangling Euler Diagrams', *TVCG (InfoVis10)*, 2010.]

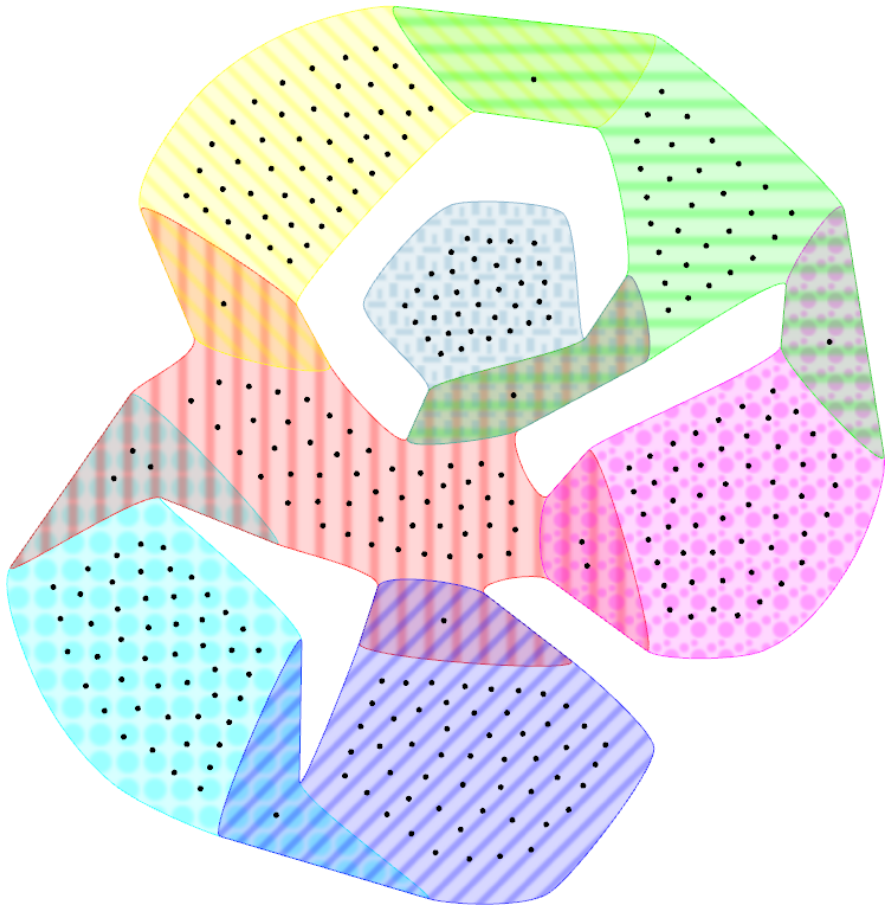
Euler Representations

Constraints

- show the expressed zones,
- use simple cluster curves,
- no disconnected zones,
- ~~no disconnected clusters~~.

Characteristics

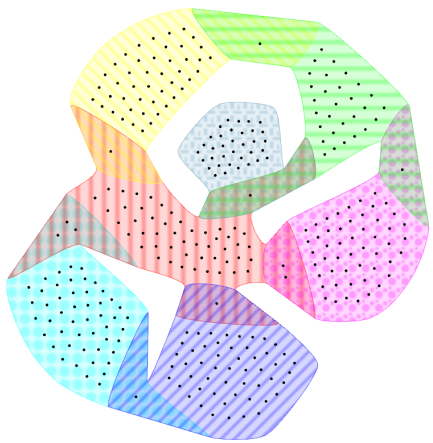
- depict elements,
- transparency and textures,
- area-proportionality,
- high curves concurrency.



Part 2

Generation of Euler Representations

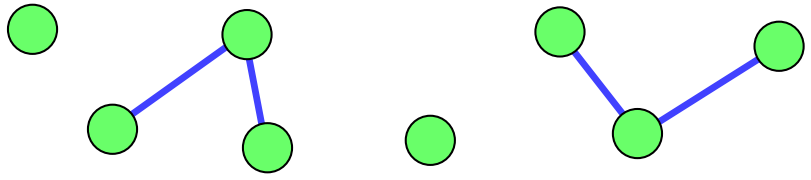
$A = \{\dots\}$
$B = \{\dots\}$
$C = \{\dots\}$
$D = \{\dots\}$



Basics Concepts of Graph Drawing

Graphs

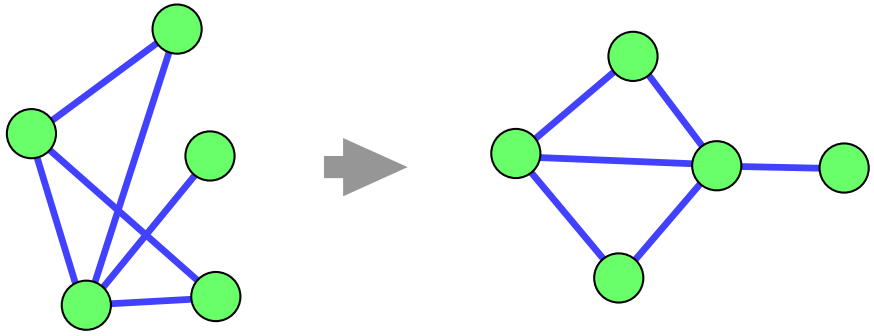
- nodes,
- edges.



Graph drawing

Displace elements so that

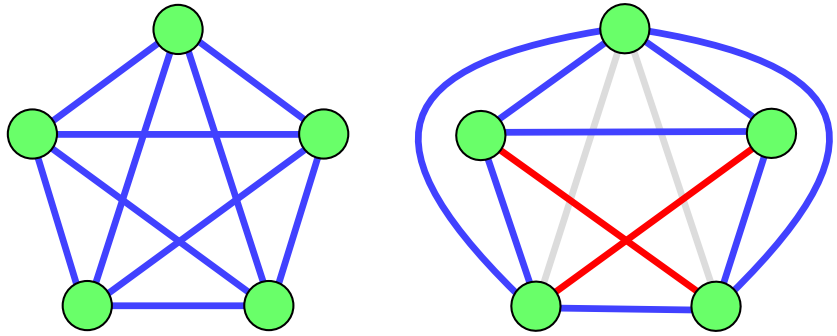
- regular spacing,
- avoid crossings...



Planarity

Drawing without crossings

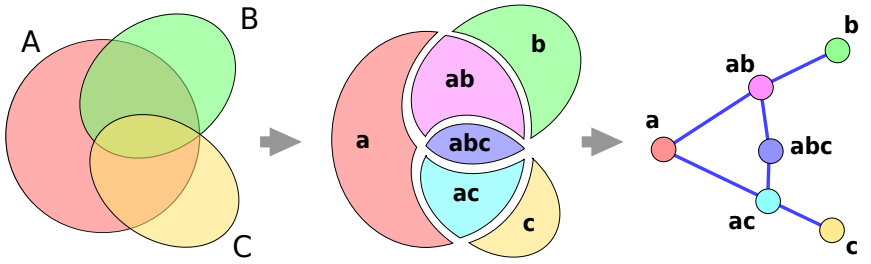
- might be impossible,
- hard to draw nicely.



Graph Drawing and Euler Diagrams

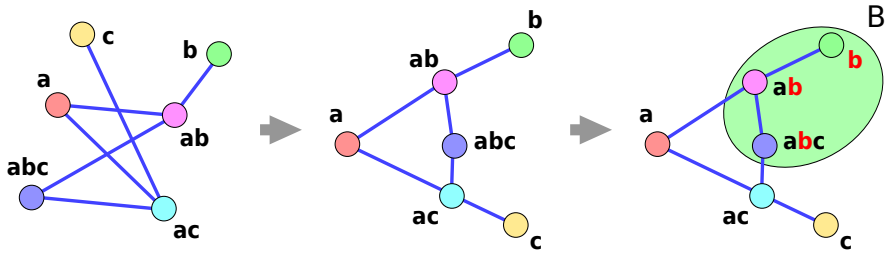
Diagrams to graphs

- a node for each zone,
- an edge for proximity.



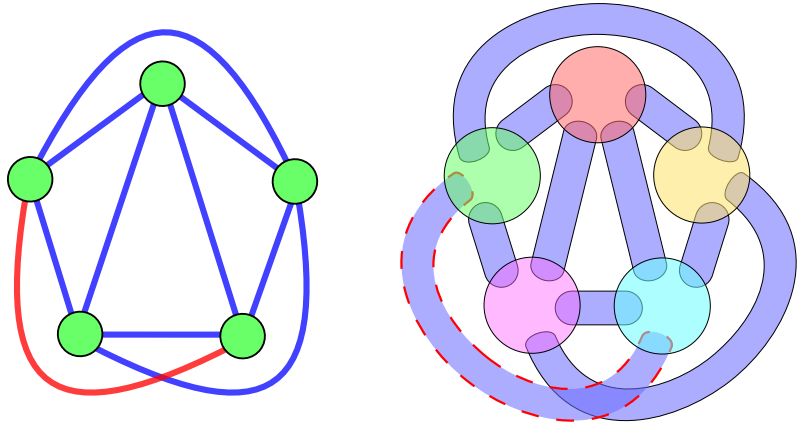
Graphs to diagrams

- identify nodes and edges,
- draw the graph,
- trace boundaries.



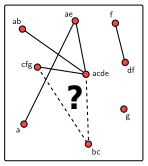
Planarity

- the graph must be planar,
- not always possible with connected clusters.

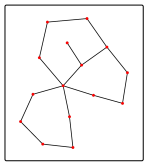


Generation of Euler Representations

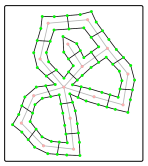
Steps



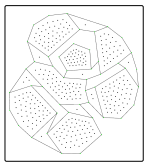
zGraph generation



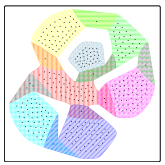
zGraph drawing



gGraph generation

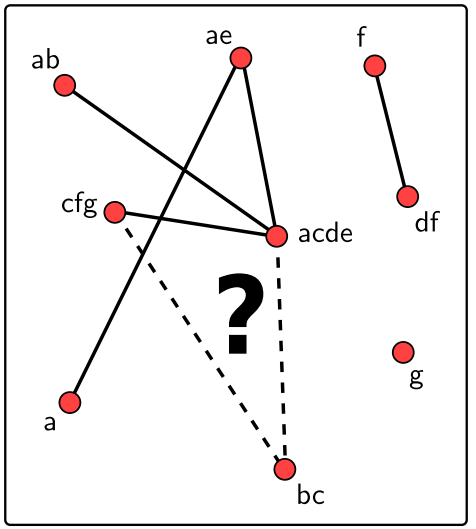


gGraph drawing



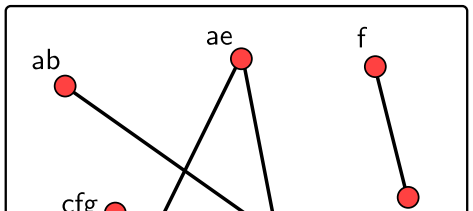
Final diagram

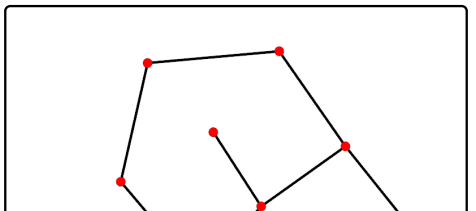
$$A = \{\dots\}$$
$$B = \{\dots\}$$
$$C = \{\dots\}$$
$$D = \{\dots\}$$



Zone graph generation

- one node per expressed zone,
- edges to enforce proximity,
- keep planarity.





Step 1 — Zone graph generation

Nodes

- find expressed zones,
- add a node per zone.

$$\begin{array}{l} A = \{1,2,3\} \\ B = \{2,4,5\} \\ C = \{2\} \end{array} \rightarrow \begin{array}{l} 1 \in A \\ 2 \in A,B,C \\ 3 \in A \\ 4 \in B \\ 5 \in B \end{array} \rightarrow \begin{array}{l} a = \{1,3\} \\ b = \{4,5\} \\ abc = \{2\} \end{array}$$

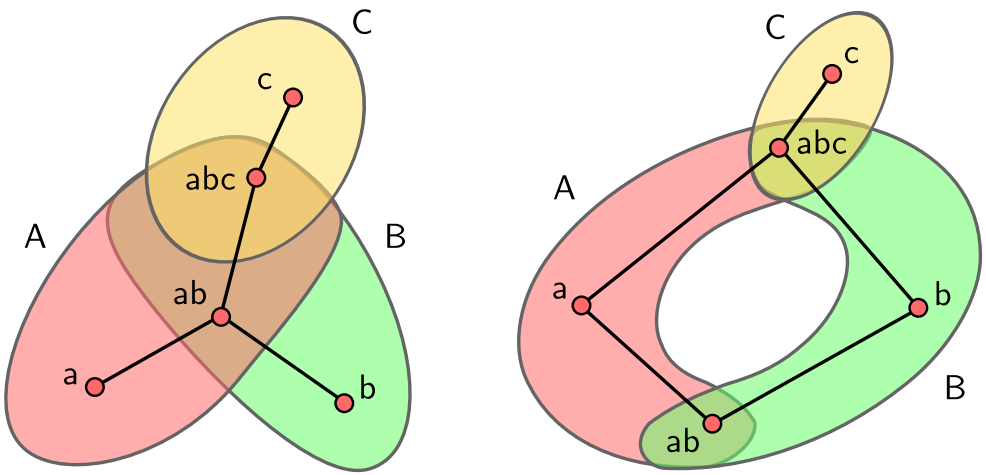
Edges

Have a metric that evaluates

- cluster connections,
- configuration aesthetic.

Edge insertion

- consider best edge,
- insert if still planar,
- update metrics.



[Simonetto, Auber, 'An Heuristic for the Construction of Intersection Graphs', *IV09*, 2009.]

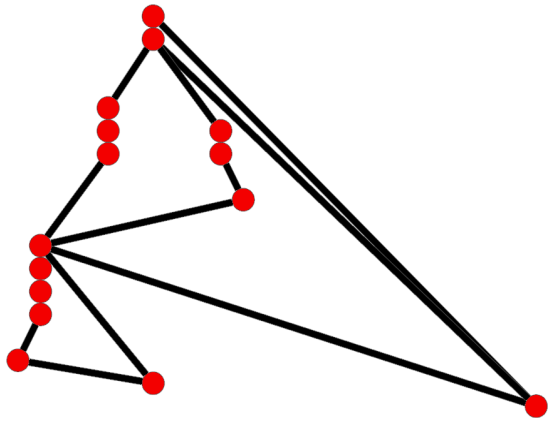
Step 2 — Zone graph drawing

First drawing

We draw zGraph with `FPP`

[De Fraysseix, Pach, Pollack, 1990]

- planar,
- low aesthetics,
- fast.

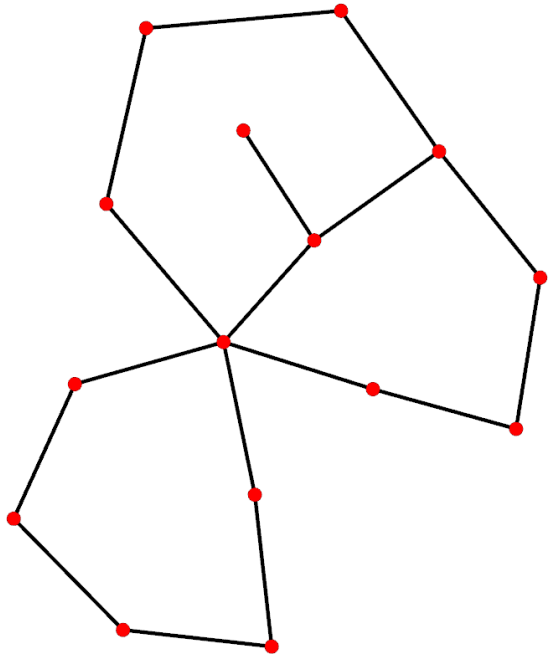


Improvement

We improve zGraph with `ImPrEd`

[Simonetto, Archambault, Auber, Bourqui, 2011]

- preserves planarity,
- improve aesthetics,
- considerably slower.

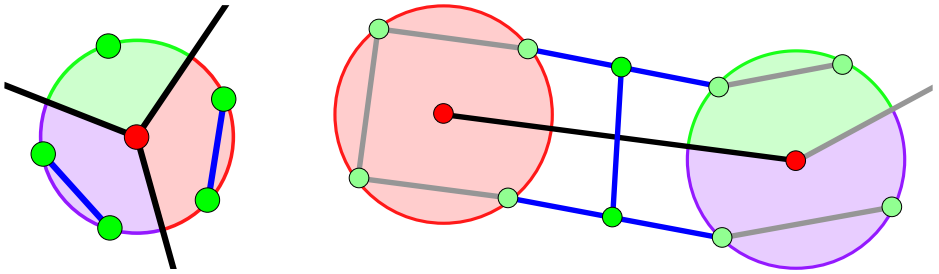


Step 3 — Grid graph generation

Enclose zNodes

We place nodes and edges

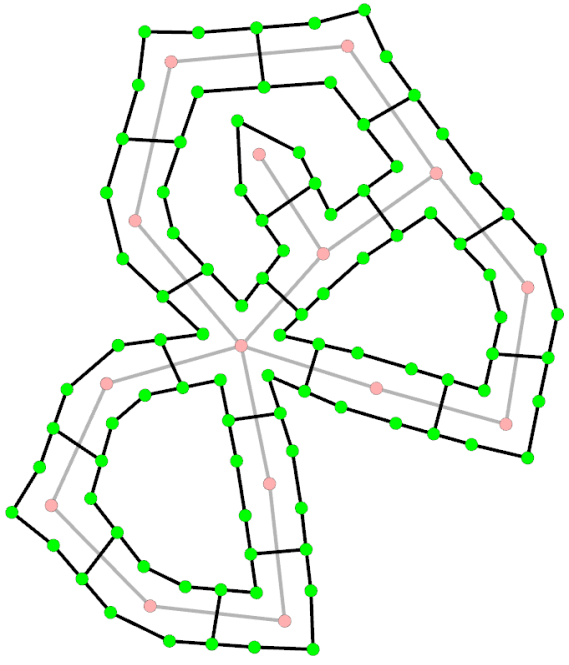
- in a circle,
- to divide zEdges,
- to limit central angles.



Eclose zEdges

We place nodes and edges

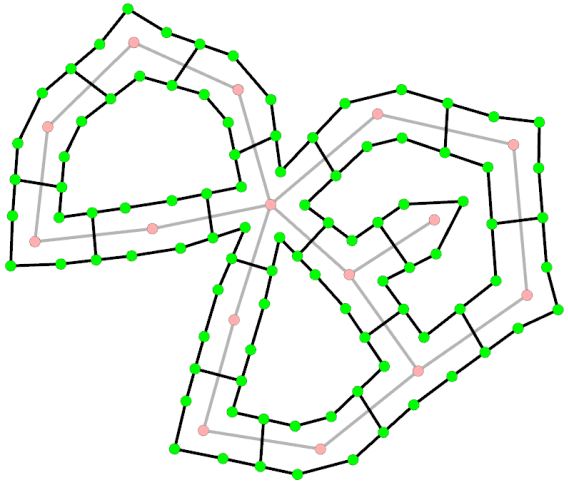
- to split the zEdge,
- to surround it.



Step 4 — Grid graph drawing

Modifications

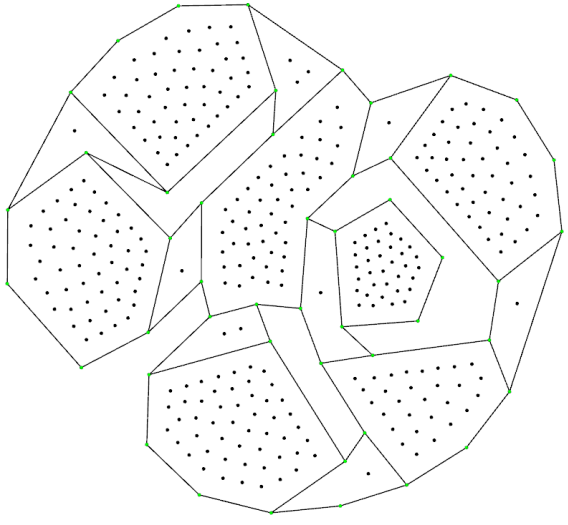
- insert elements,
- remove zGraph.



Drawing

We apply `ImPrEd`

- preserve positions,
- improve aesthetics,
- high computational cost.



Step 5 — Final drawing

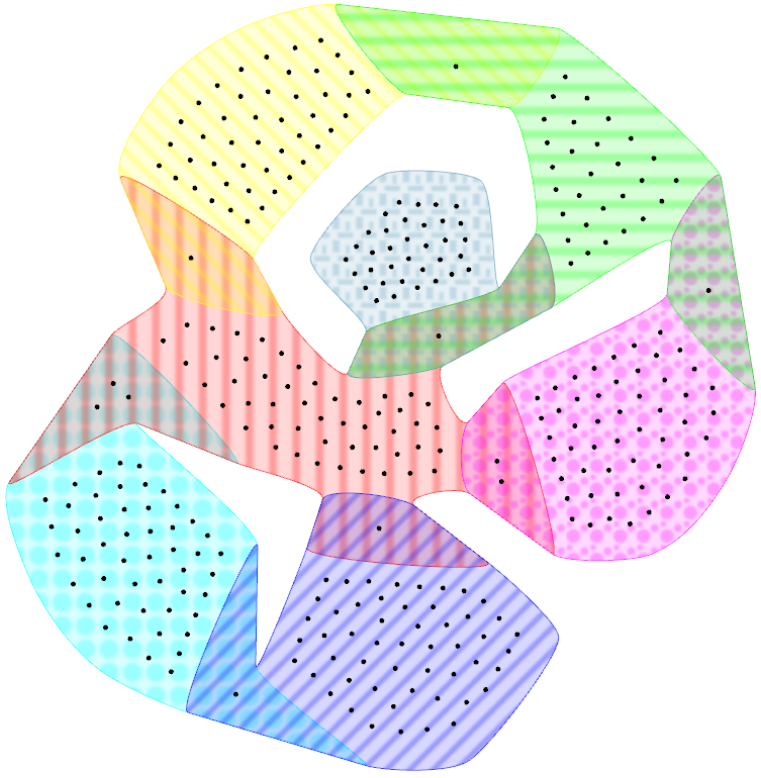
Cluster curves

We apply Bézier curves

- transforming gEdges,
- enforcing smooth junctions.

Colours and textures

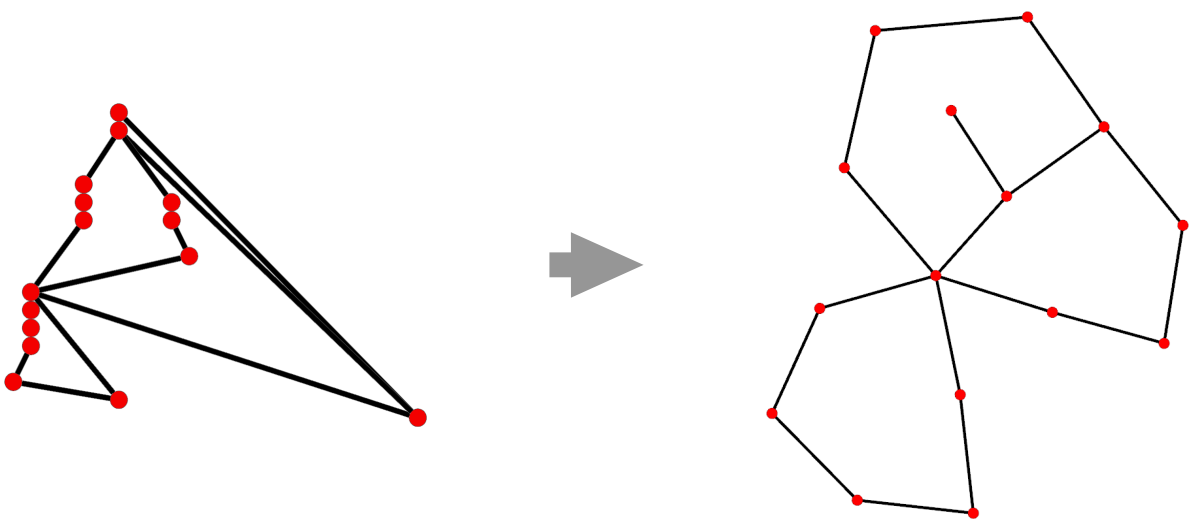
- low number,
- different colours/textures for overlapping clusters,
- coprime.



[Simonetto, Auber, Archambault, 'Fully Automatic Visualisation of Overlapping Sets', *Computer Graphics Forum* (EuroVis09), 2009.]

Part 3

ImPrEd



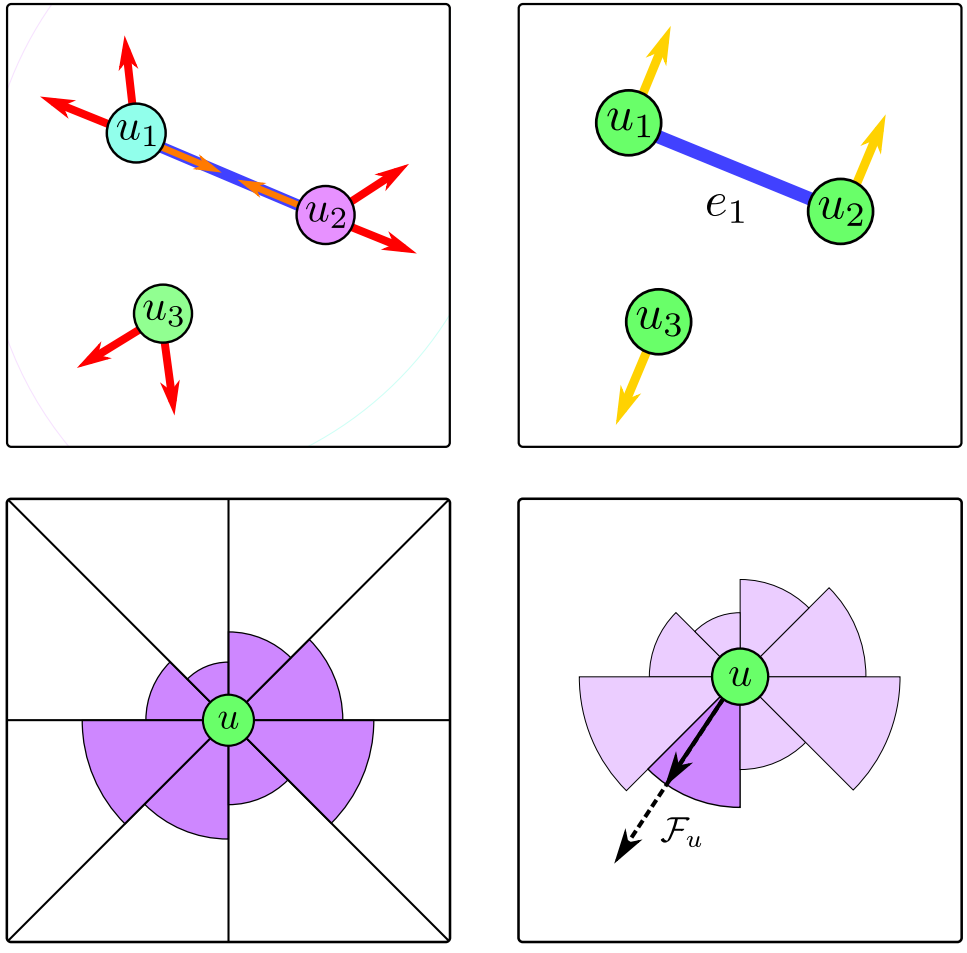
PrEd, the original algorithm

Force directed

- node-node repulsion,
- edge attraction,
- node-edge repulsion.

Movement limitation

- 8 movement sectors,
- bounded by edges,
- movement bounded by sector amplitude.

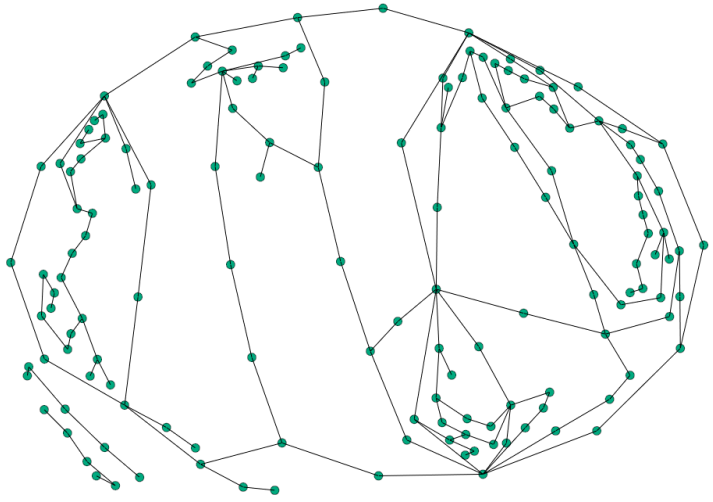


[Bertault, 'A Force-Directed Algorithm that Preserves Edge Crossing Properties', *Information Processing Letters*, 2000.]

Advantages and Disadvantages of PrEd

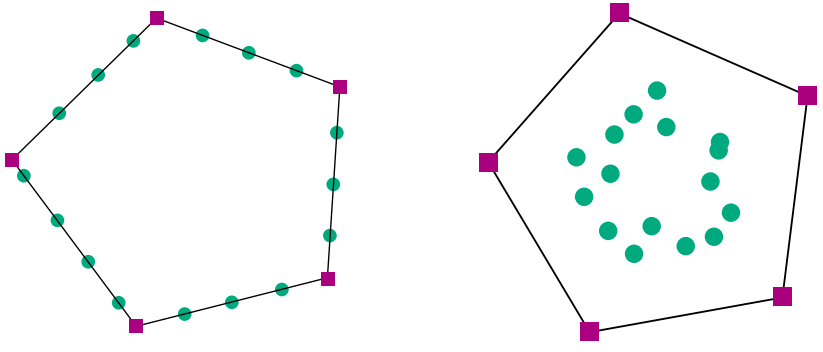
Advantages

- peculiar property,
- good on small graphs,
- intuitive and simple.

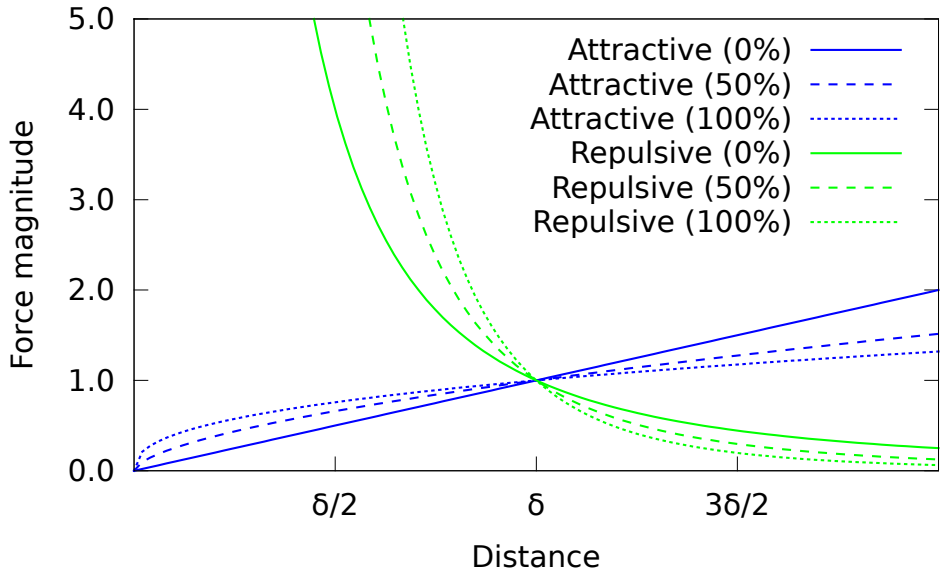


Disadvantages

- high computational cost,
- over-restrictive movement,
- low aesthetics for large and sparse graphs,
- low control on output.

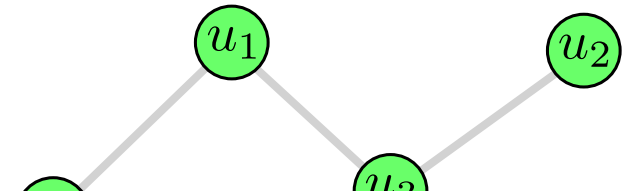


Improved PrEd

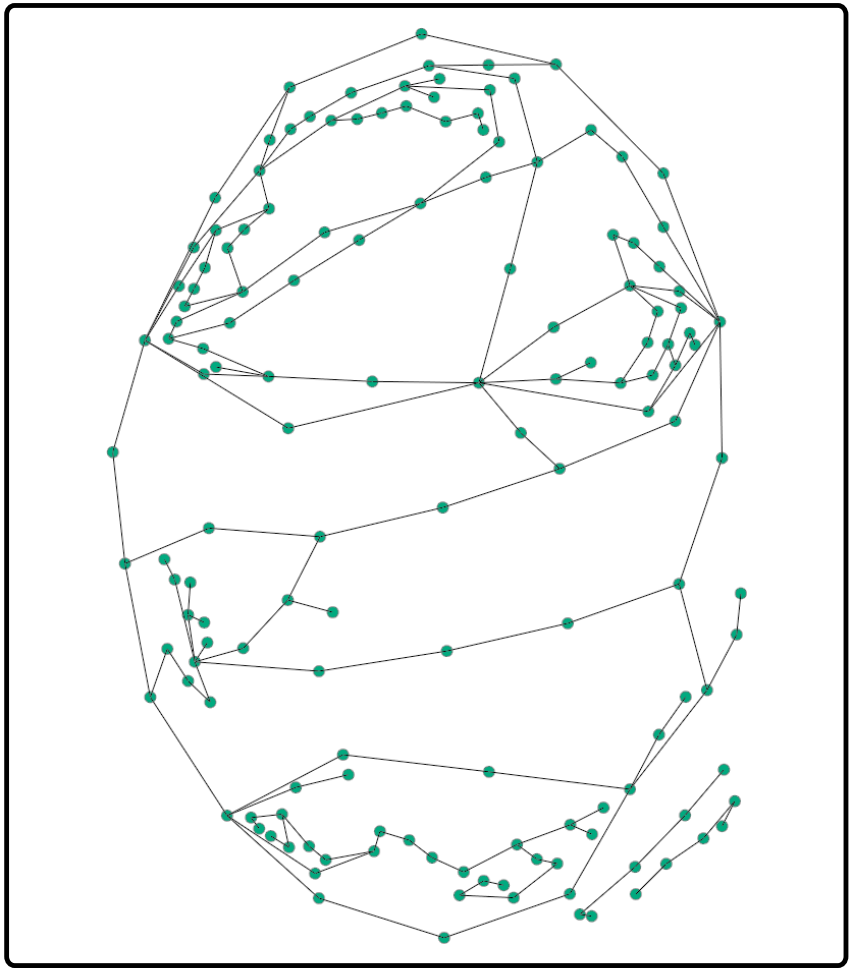


Movement and Force Cooling

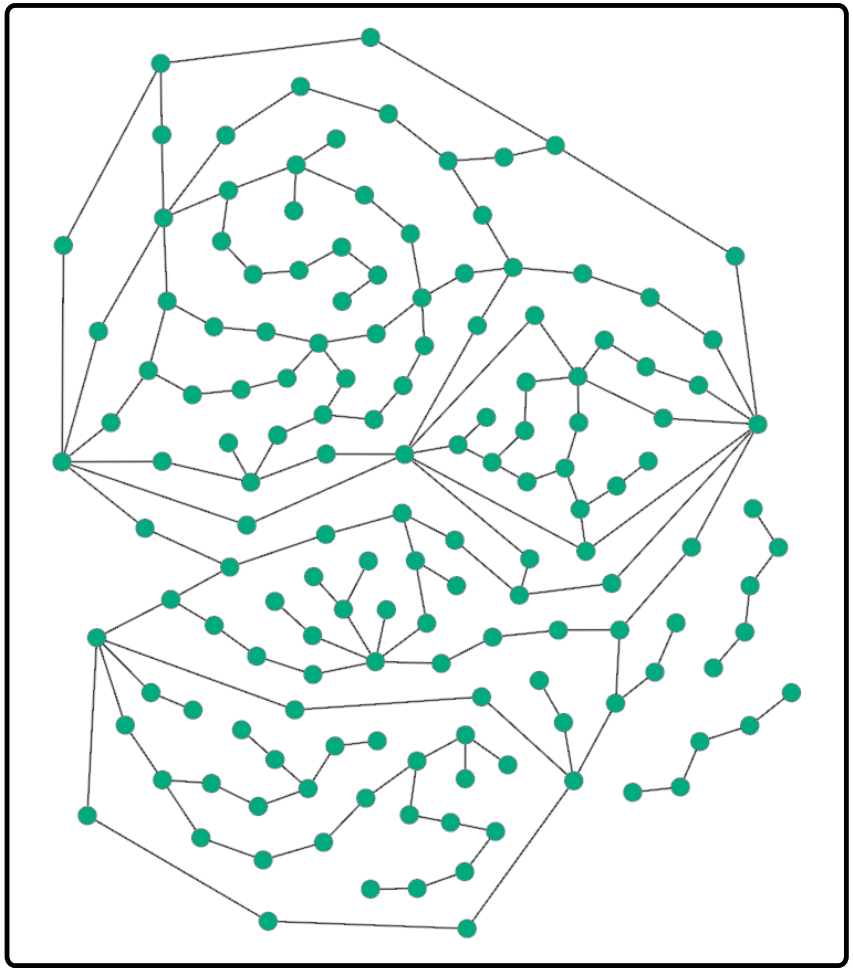
- gradually increase local effect,
- gradually decrease energy,
- promote stability and reliability.



ImPrEd's Drawing Quality

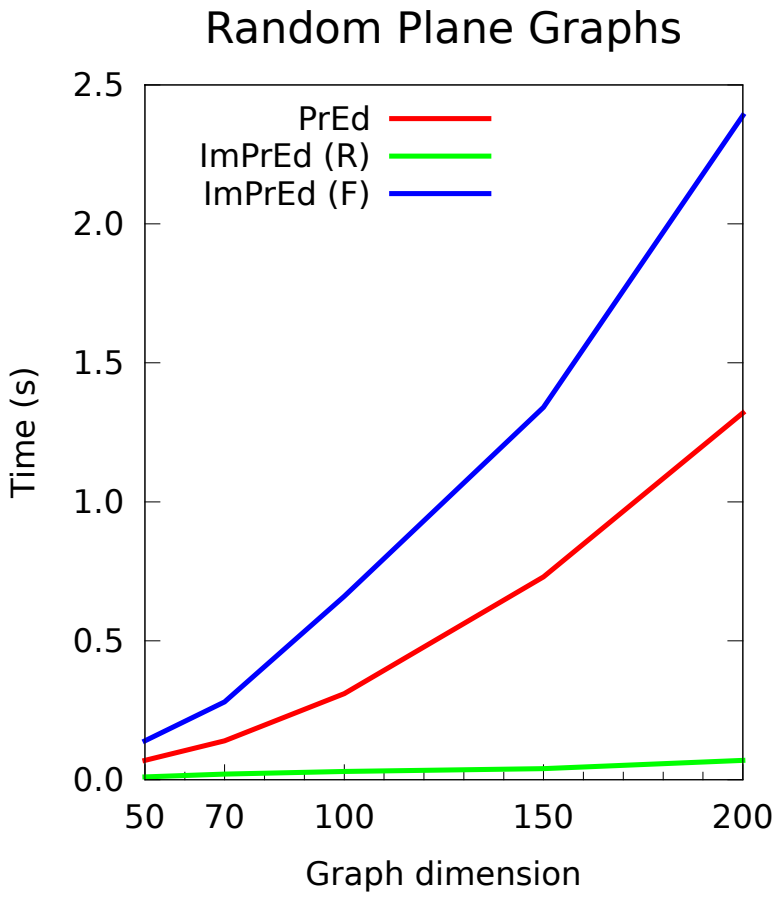


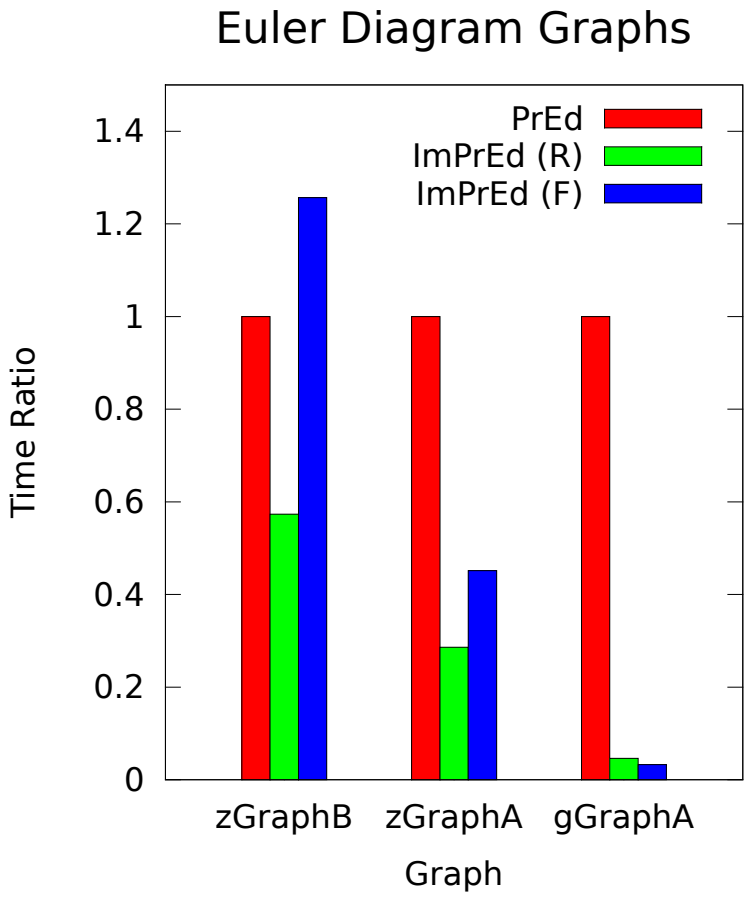
PrEd



ImPrEd

ImPrEd's Performances

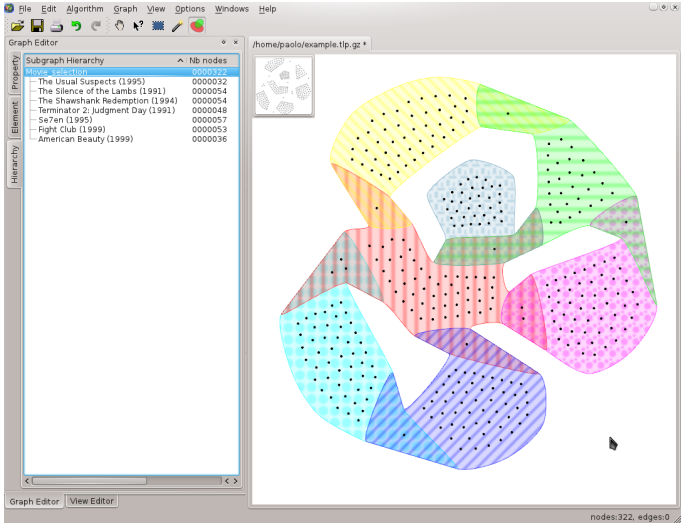




[Simonetto, Archambault, Auber and Bourqui, 'ImPrEd: An Improved Force-Directed Algorithm that Prevents Nodes from Crossing Edges', *Computer Graphics Forum* (EuroVis11), 2011.]

Part 4

Software and Examples



EulerView: a Tulip Plug-in

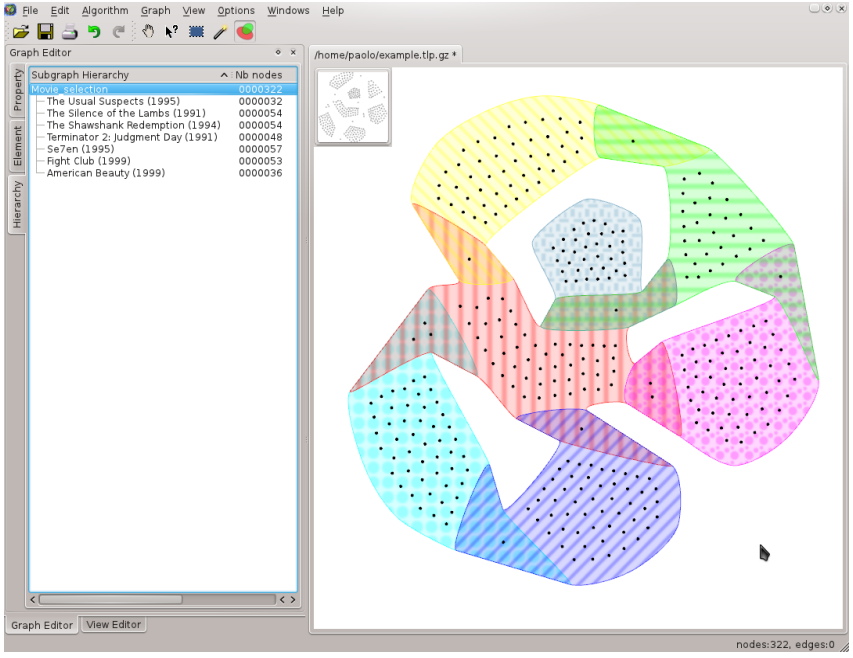
EulerView

- Tulip view plug-in,
- provide alternative view for clustered graphs.

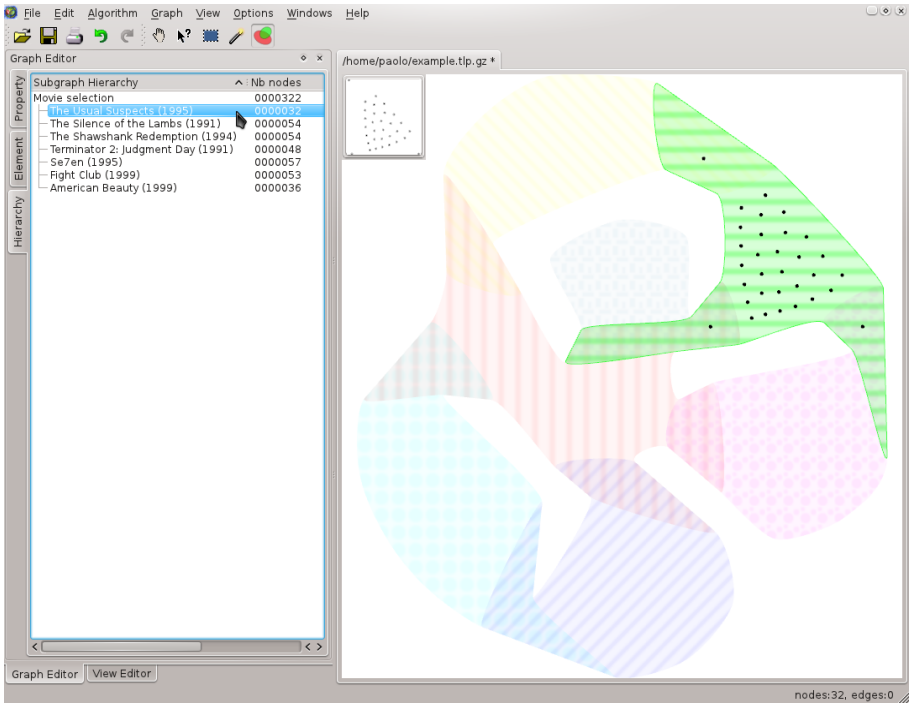
[Tulip Data Visualisation Software, http://tulip.labri.fr]

Interaction

- inherited functions,
- selection,
- contextual information.

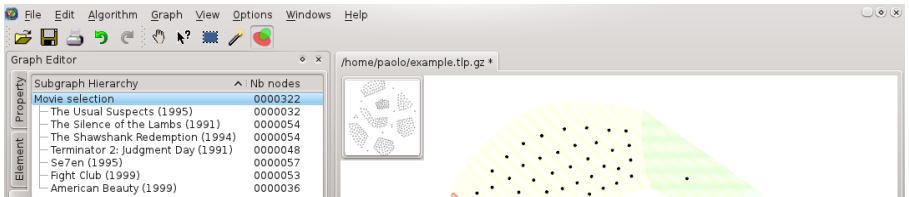


Selection Features

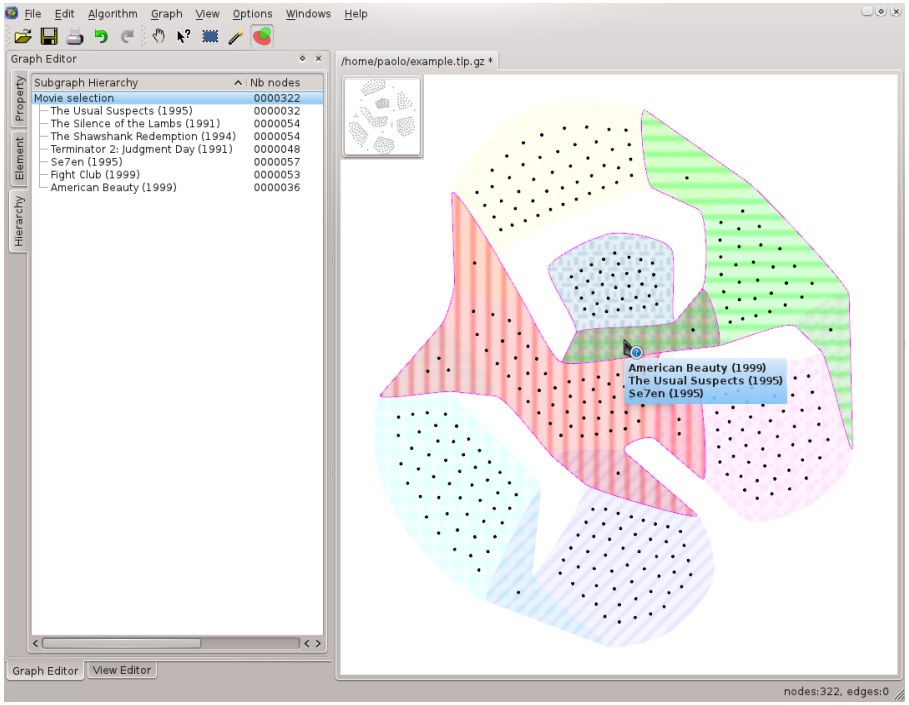


List Selection

- click on cluster list in left panel,
- identify set positions.

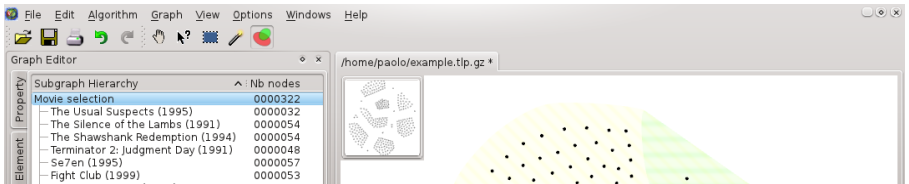


Contextual Information



Selection Tooltip

- cluster names,
- small to avoid obstructing.



Example: IMDb7

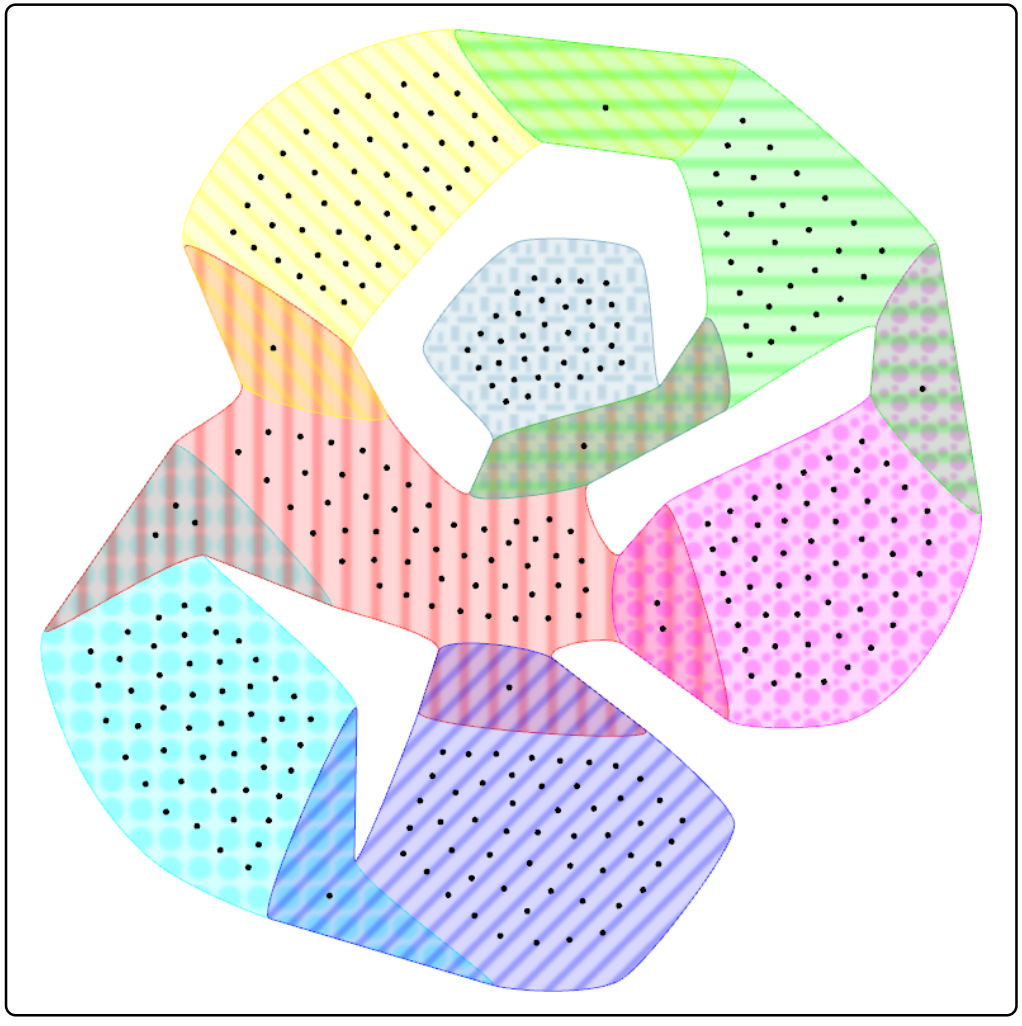
Features

- 7 selected films,
- full credited cast,
- 322 actors.

[Internet Movie Database, http://www.imdb.org]

Remarks

- example diagram,
- relatively small,
- very simple intersections.



Example: IMDb60

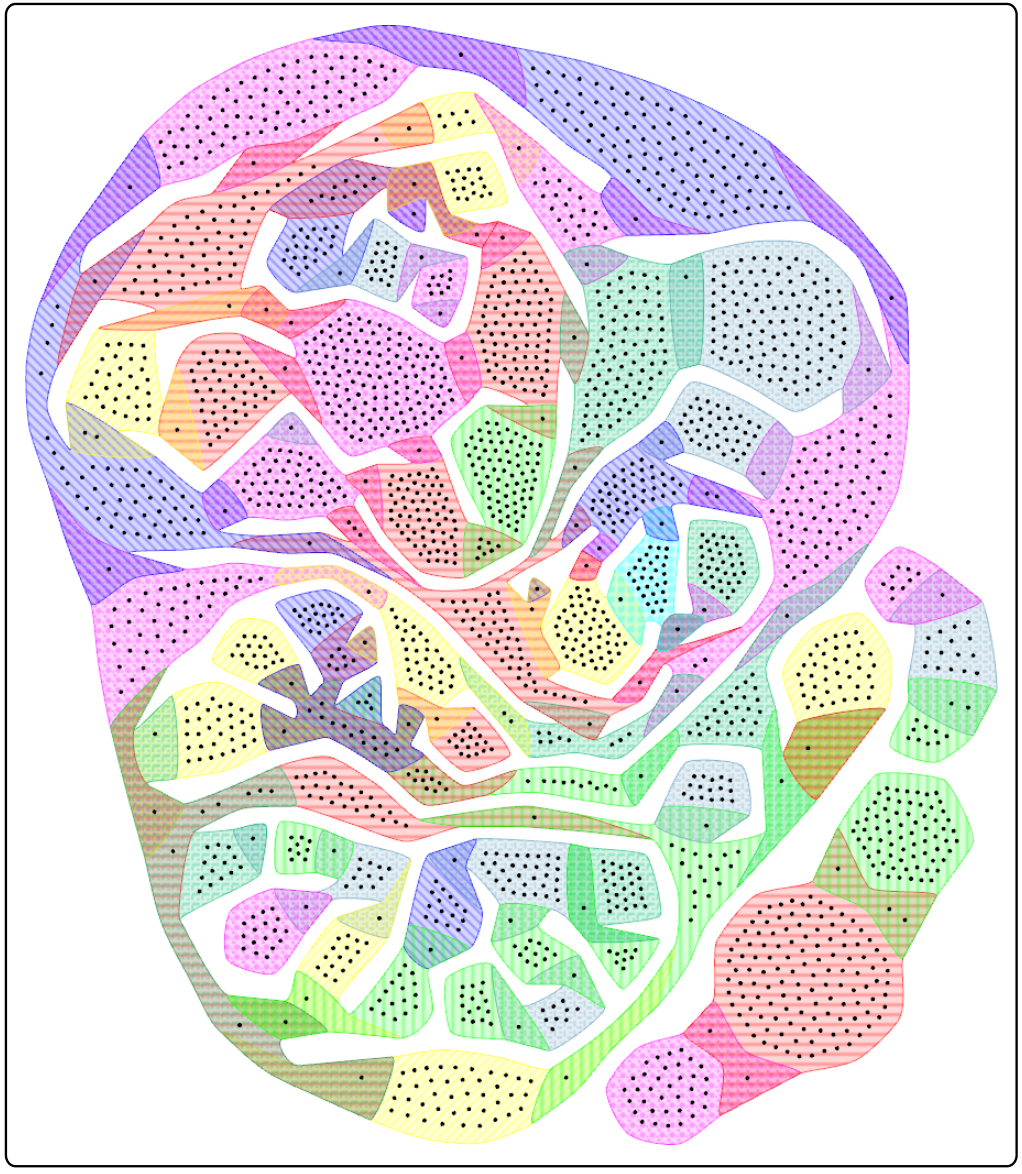
Features

- 60 top rated films,
- full credited cast,
- 2263 actors.

[Internet Movie Database, http://www.imdb.org]

Remarks

- high aesthetics,
- very large,
- simple intersections.



Euler Representations on Graph Clustering

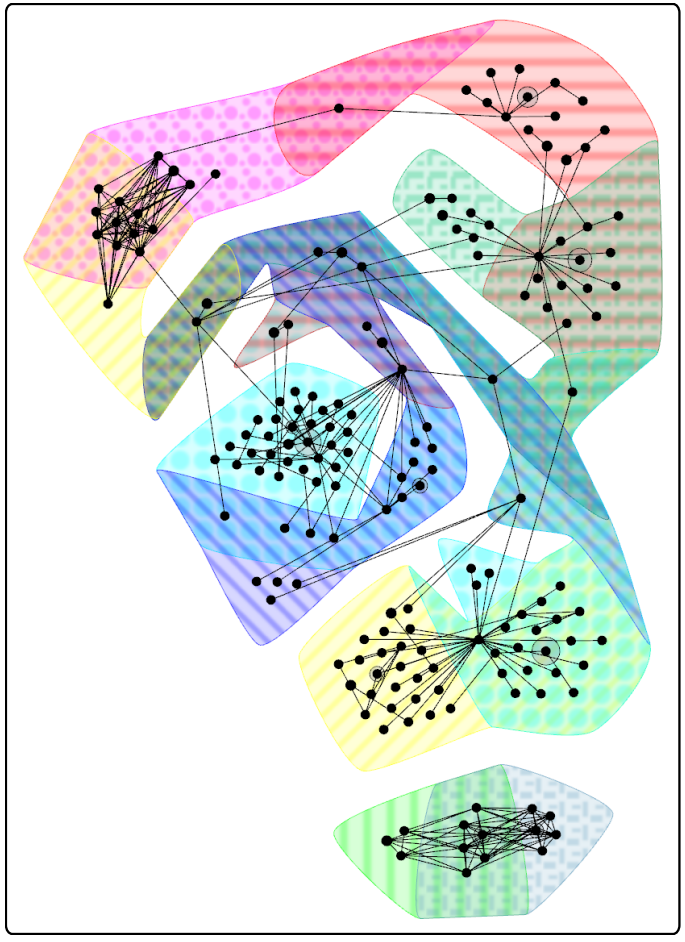
Graph Clustering

Collection of graph nodes

- originally: partitions,
- recently: overlapping,
- differ from sets for edges.

Combine metaphors

- ED: containment,
- graph: connection,
- optimise both is challenging.



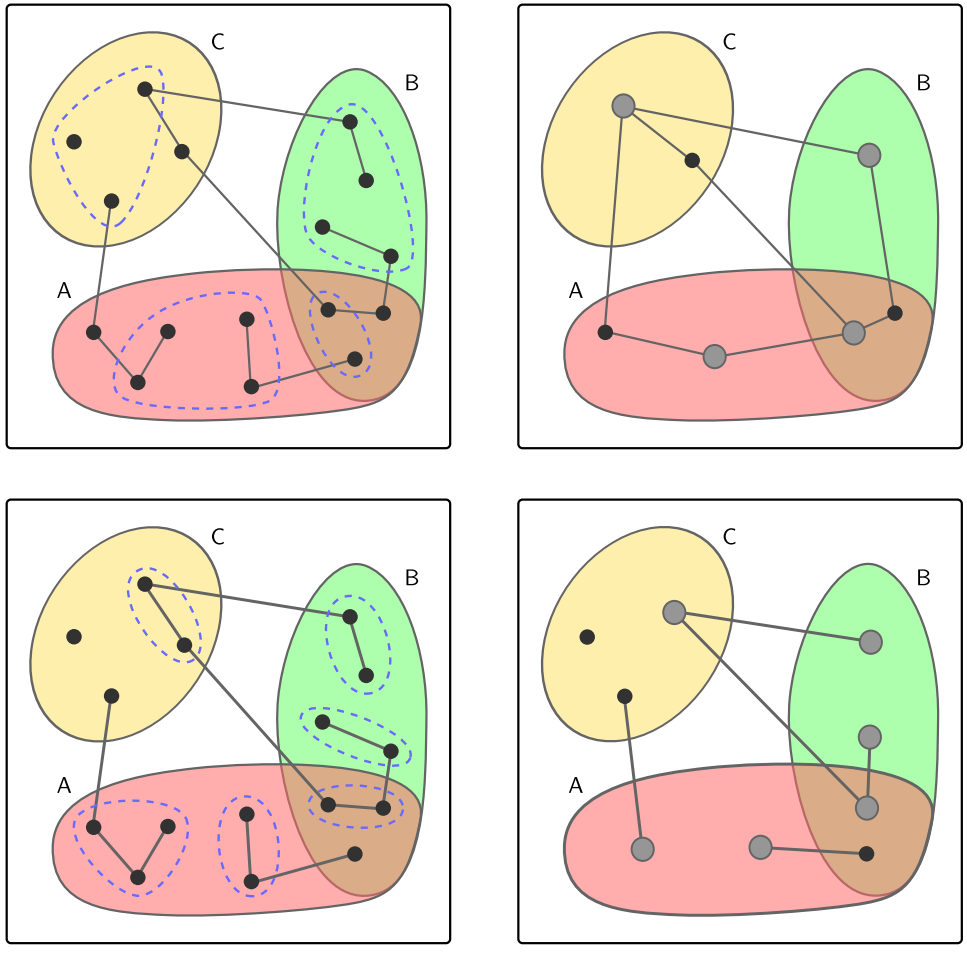
Path-Preserving Meta-nodes

Metanodes

- condense elements,
- reduce clutter,
- might mislead on connectivity.

Path-Preservation

- group connected elements,
- avoid to mislead.



[Archambault, Munzner, Auber, 'GrouseFlocks: Steerable Exploration of Graph Hierarchy Space', *TVCG*, 2008.]

Example: Gene Interaction

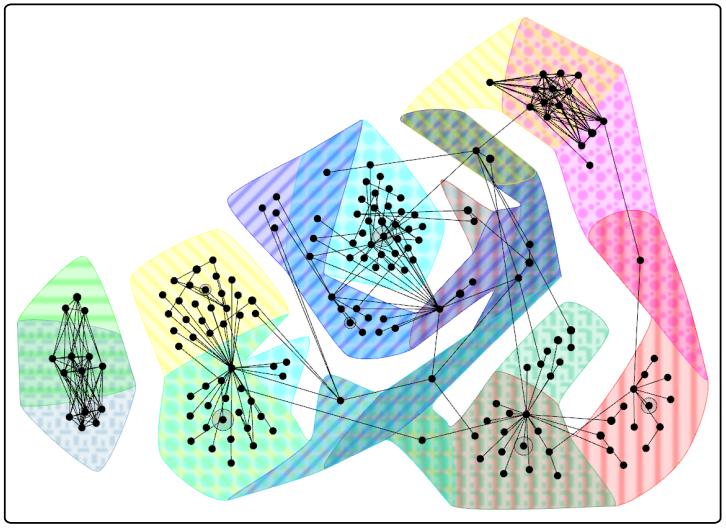
Features

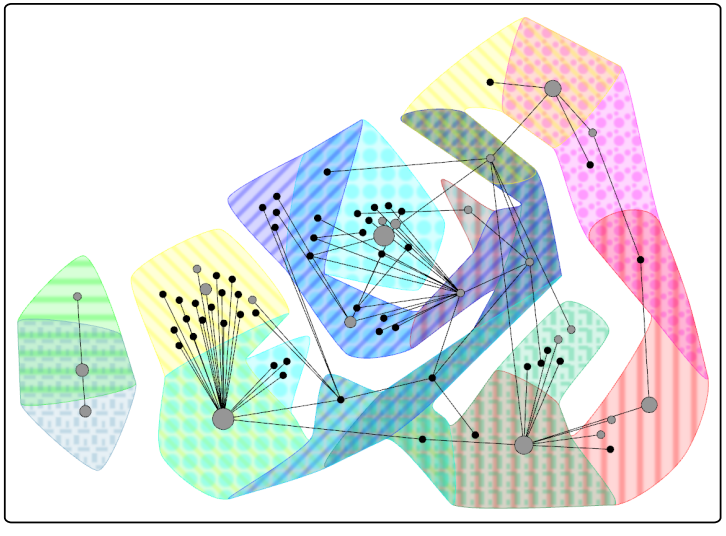
- 10 clusters,
- 176 genes,
- 296 interactions.

[Itoh, Muelder, Ma, Sese, 'A Hybrid Space-Filling and Force-Directed Layout Method for Visualizing Multiple-Category Graphs', *PacificVis09*, 2009.]

Remarks

- edges contribute,
- meta-node reduce cluttering.





Example: Carsonella

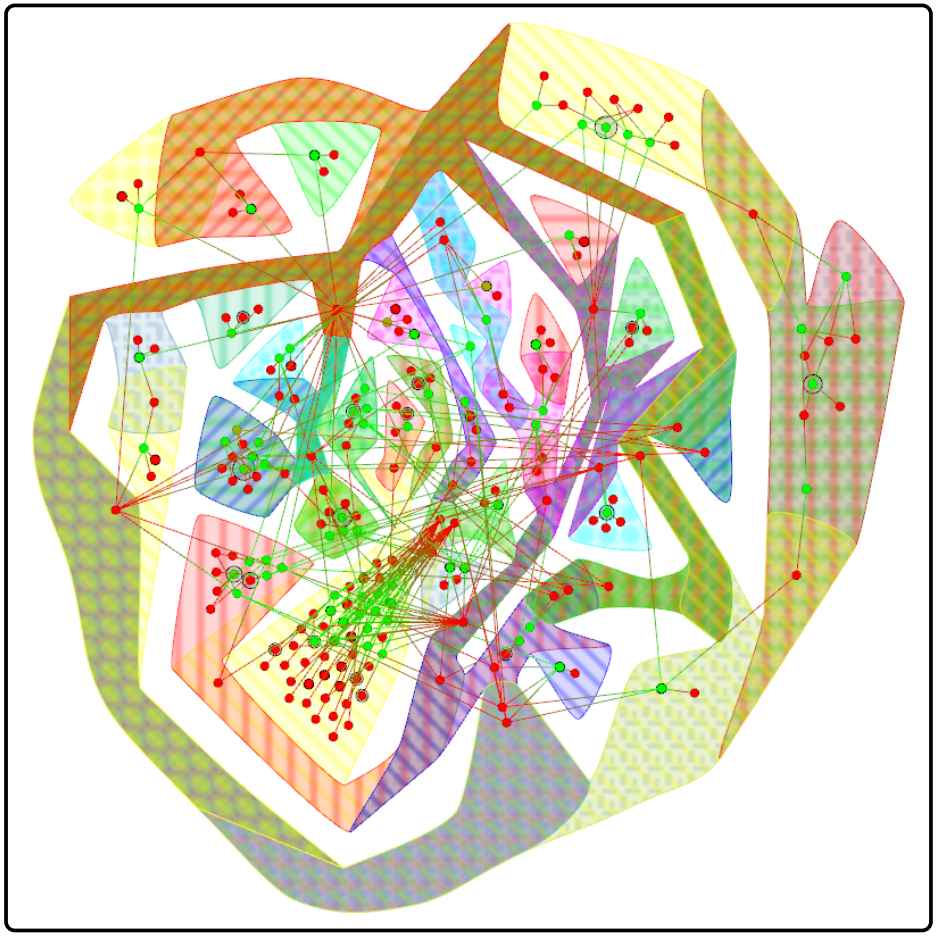
Features

- 35 pathways,
- metabolites (red),
- enzymes (green),
- 224 nodes and 335 edges.

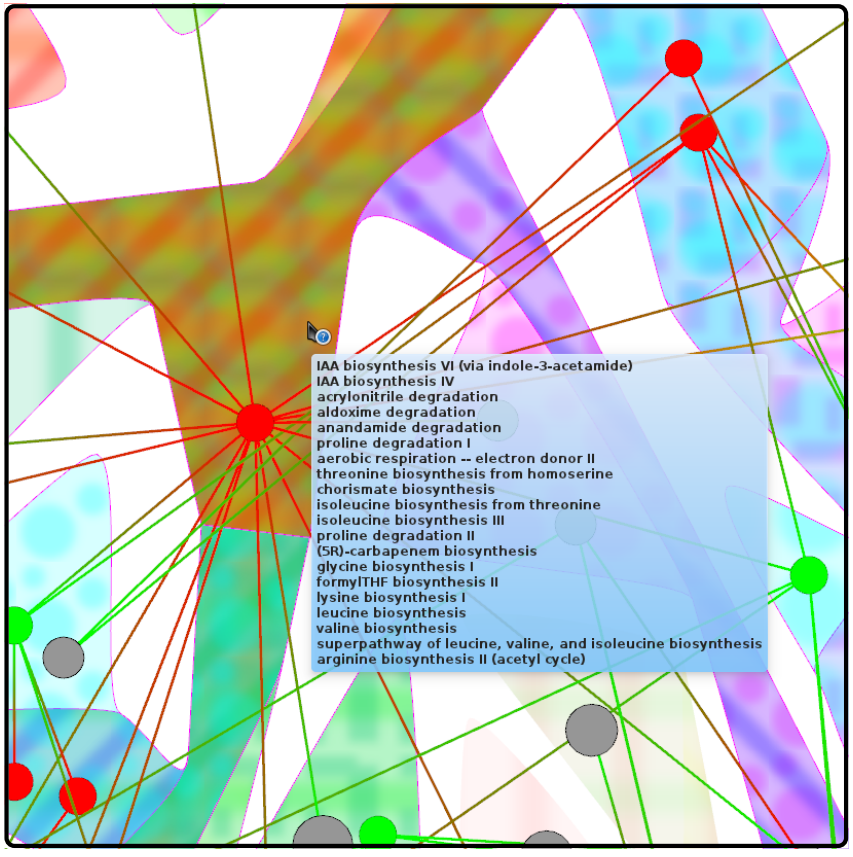
[Metabolic Data of Endosymbiontic, Parasitic and Free Bacteria, Université Lyon 1]

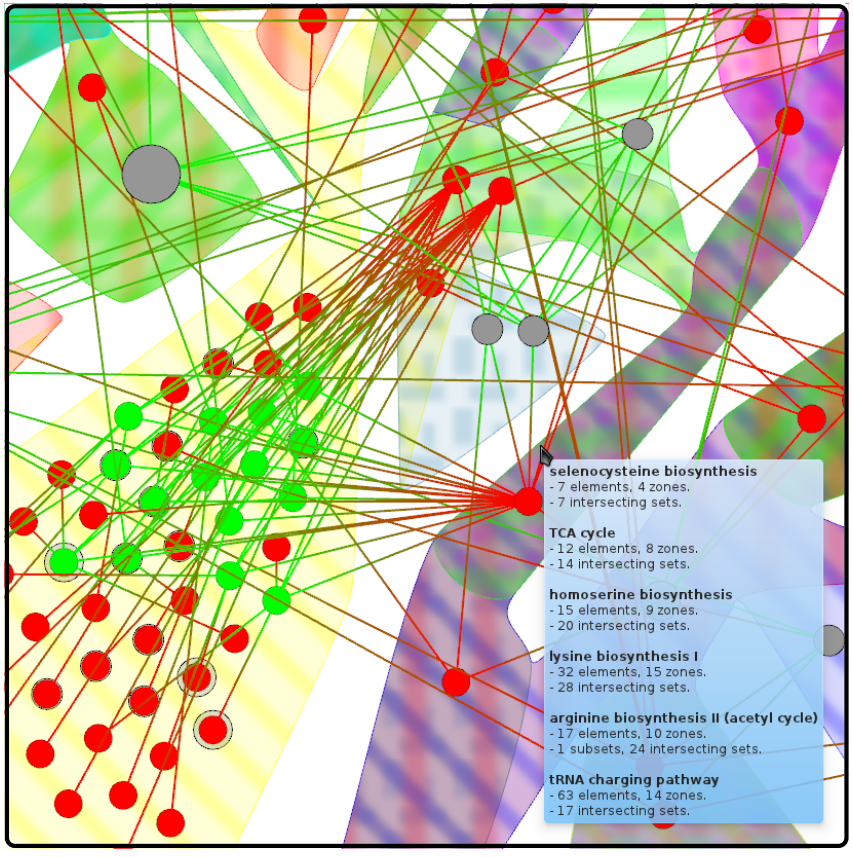
Remarks

- complex intersections,
- high graph cluttering.



Interaction Mitigates the Drawing Complexity





Conclusions

Euler Representations

Visualisation for overlapping data

- based on Euler diagrams,
- working on every input,
- sufficiently fast for non interactive use.

ImPrEd

Graph improvement algorithm

- achieve high aesthetics,
- faster than PrEd,
- very useful in ED generation.

Software

Software implementation

- features some interaction,
- mitigates the limitation of the representation.

Examples

We showed output examples

- sets and clusters,
- small and large instances,
- simple and complex.

Future Work

Drawing Readability

- more regular curves,
- reduce concurrency,
- speed/quality trade-off.

Running Time (`ImPrEd`)

- reduce complexity,
- reduce input instance,
- change algorithm.

  [Dwyer *et al.*, Constrained FDL]

Validate

Run usability studies

- on Euler Representation,
- on combining metaphors.

Extensions

- more interaction,
- on-line redrawing,
- first overview, then diagram,
- ...

Thank you for your attention.

Any questions?

Detailed information in the thesis and in:

Simonetto, Auber, 'Visualise Undrawable Euler Diagrams',
IV08, 2008.

Simonetto, Auber, 'An Heuristic for the Construction of Intersection Graphs',
IV09, 2009.

Simonetto, Auber, Archambault, 'Fully Automatic Visualisation of Overlapping Sets',
Computer Graphics Forum (EuroVis09), 2009.

Simonetto, Auber, Archambault, Bourqui, 'ImPrEd: An Improved Force-Directed Algorithm that Prevents Nodes from Crossing Edges',
Computer Graphics Forum (EuroVis11), 2011.