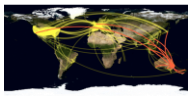
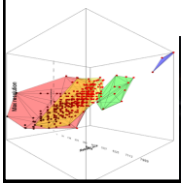
Aviz

Visual Analytics Project




# Understanding Complex Networks

Jean-Daniel Fekete
INRIA
http://www.aviz.fr/~fekete





# Visualization?

**Visualization** is any technique for creating images, diagrams, or animations to *communicate a message*

[Wikipedia, Visualization, May 2016]

**Information visualization** is the study of (interactive) visual representations of abstract data to *reinforce human cognition*

[Card, S. and Mackinlay, J. and Shneiderman B., Readings in Information Visualization, 1999]

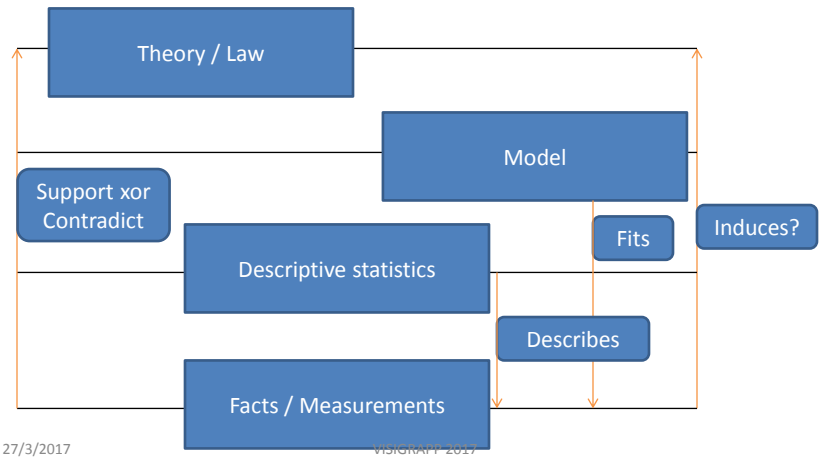
# Network Visualization

Information visualization applied to network structures:

- The study of interactive visual representations of *networks* to reinforce human cognition



# Where does Visualization Stands?



# Example

| I | | II | | III | | IV | |
|---|---|---|---|---|---|---|---|
| x | y | x | y | x | y | x | y |
| 10.0 | 8.04 | 10.0 | 9.14 | 10.0 | 7.46 | 8.0 | 6.58 |
| 8.0 | 6.95 | 8.0 | 8.14 | 8.0 | 6.77 | 8.0 | 5.76 |
| 13.0 | 7.58 | 13.0 | 8.74 | 13.0 | 12.74 | 8.0 | 7.71 |
| 9.0 | 8.81 | 9.0 | 8.77 | 9.0 | 7.11 | 8.0 | 8.84 |
| 11.0 | 8.33 | 11.0 | 9.26 | 11.0 | 7.81 | 8.0 | 8.47 |
| 14.0 | 9.96 | 14.0 | 8.10 | 14.0 | 8.84 | 8.0 | 7.04 |
| 6.0 | 7.24 | 6.0 | 6.13 | 6.0 | 6.08 | 8.0 | 5.25 |
| 4.0 | 4.26 | 4.0 | 3.10 | 4.0 | 5.39 | 19.0 | 12.50 |
| 12.0 | 10.84 | 12.0 | 9.13 | 12.0 | 8.15 | 8.0 | 5.56 |
| 7.0 | 4.82 | 7.0 | 7.26 | 7.0 | 6.42 | 8.0 | 7.91 |
| 5.0 | 5.68 | 5.0 | 4.74 | 5.0 | 5.73 | 8.0 | 6.89 |

Raw Data from Anscombe's Quartet



[Source: Anscombe's quartet, Wikipedia]

# Statistical Analysis

For all columns, the main descriptive statistics are identical

| I | | II | | III | | IV | |
|---|---|---|---|---|---|---|---|
| x | y | x | y | x | y | x | y |
| 10.0 | 8.04 | 10.0 | 9.14 | 10.0 | 7.46 | 8.0 | 6.58 |
| 8.0 | 6.95 | 8.0 | 8.14 | 8.0 | 6.77 | 8.0 | 5.76 |
| 13.0 | 7.58 | 13.0 | 8.74 | 13.0 | 12.74 | 8.0 | 7.71 |
| 9.0 | 8.81 | 9.0 | 8.77 | 9.0 | 7.11 | 8.0 | 8.84 |
| 11.0 | 8.33 | 11.0 | 9.26 | 11.0 | 7.81 | 8.0 | 8.47 |
| 14.0 | 9.96 | 14.0 | 8.10 | 14.0 | 8.84 | 8.0 | 7.04 |
| 6.0 | 7.24 | 6.0 | 6.13 | 6.0 | 6.08 | 8.0 | 5.25 |
| 4.0 | 4.26 | 4.0 | 3.10 | 4.0 | 5.39 | 19.0 | 12.50 |
| 12.0 | 10.84 | 12.0 | 9.13 | 12.0 | 8.15 | 8.0 | 5.56 |
| 7.0 | 4.82 | 7.0 | 7.26 | 7.0 | 6.42 | 8.0 | 7.91 |
| 5.0 | 5.68 | 5.0 | 4.74 | 5.0 | 5.73 | 8.0 | 6.89 |

| | |
|---|---|
| Mean of $x$ | 9.0 |
| Variance of $x$ | 11.0 |
| Mean of $y$ | 7.5 |
| Variance of $y$ | 4.12 |
| Correlation between $x$ and $y$ | 0.816 |
| Linear regression line | $y = 3 + 0.5x$ |

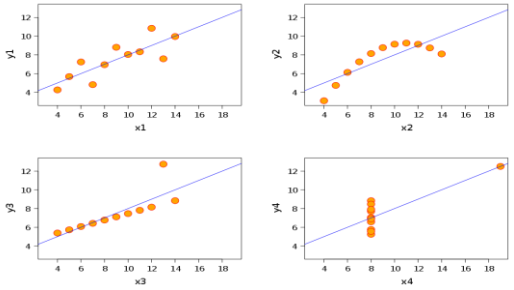


[Source: Anscombe's quartet, Wikipedia]

# Visual Representation of the Data

Visual representation reveals a different story

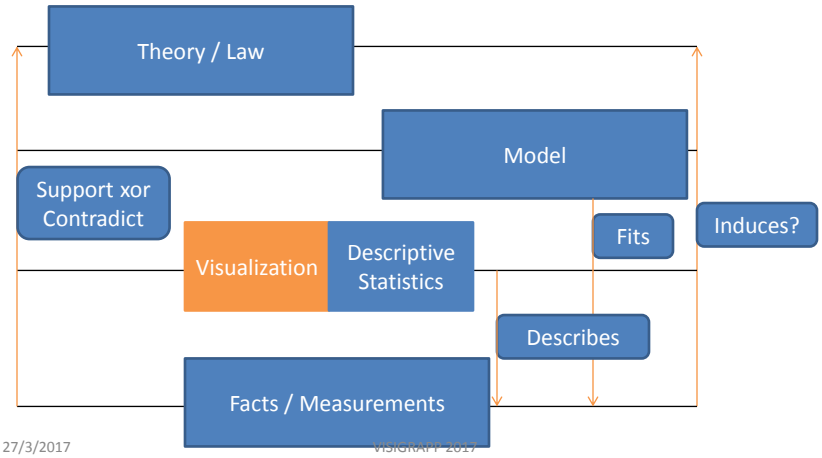
| I | | II | | III | | IV | |
|---|---|---|---|---|---|---|---|
| x | y | x | y | x | y | x | y |
| 10.0 | 8.04 | 10.0 | 9.14 | 10.0 | 7.46 | 8.0 | 6.58 |
| 8.0 | 6.95 | 8.0 | 8.14 | 8.0 | 6.77 | 8.0 | 5.76 |
| 13.0 | 7.58 | 13.0 | 8.74 | 13.0 | 12.74 | 8.0 | 7.71 |
| 9.0 | 8.81 | 9.0 | 8.77 | 9.0 | 7.11 | 8.0 | 8.84 |
| 11.0 | 8.33 | 11.0 | 9.26 | 11.0 | 7.81 | 8.0 | 8.47 |
| 14.0 | 9.96 | 14.0 | 8.10 | 14.0 | 8.84 | 8.0 | 7.04 |
| 6.0 | 7.24 | 6.0 | 6.13 | 6.0 | 6.08 | 8.0 | 5.25 |
| 4.0 | 4.26 | 4.0 | 3.10 | 4.0 | 5.39 | 19.0 | 12.50 |
| 12.0 | 10.84 | 12.0 | 9.13 | 12.0 | 8.15 | 8.0 | 5.56 |
| 7.0 | 4.82 | 7.0 | 7.26 | 7.0 | 6.42 | 8.0 | 7.91 |
| 5.0 | 5.68 | 5.0 | 4.74 | 5.0 | 5.73 | 8.0 | 6.89 |






[Source: Anscombe's quartet, Wikipedia]

# Where does Visualization Stands?



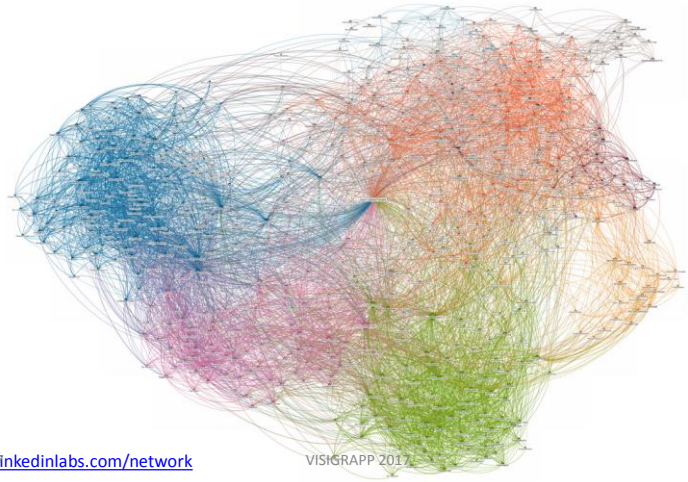
# Network Visualization

- Most (99%) network visualization tools use the node-link visual representation
- Other representations exist, more or less general
- Why use one or another?
  - Efficiency (to be defined)
  - Familiarity



# Facebook Network Dec. 2010

# LinkedIn Profession Network

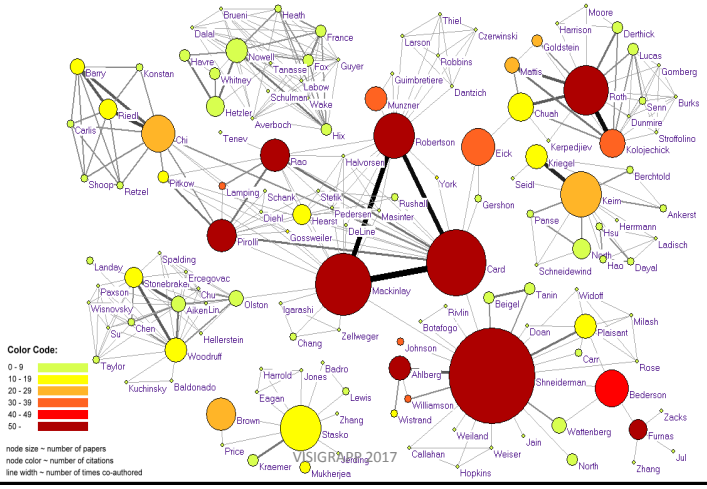


http://inmaps.linkedinlabs.com/network

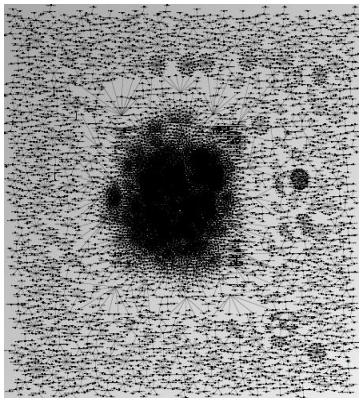


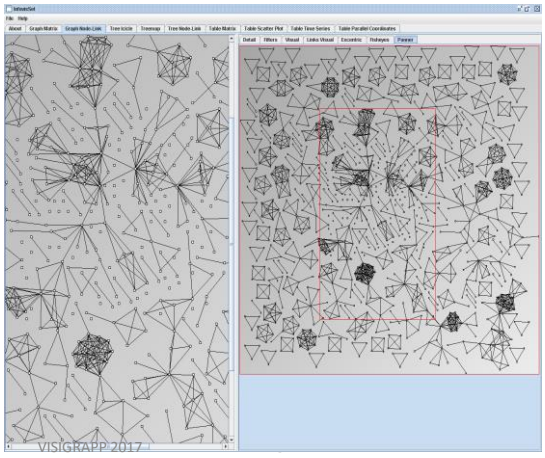
# InfoVis Co-authoring

(K. Börner et al. 04)



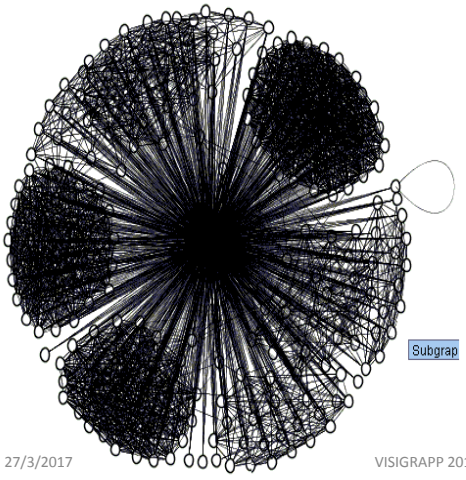
# Generally, after loading…

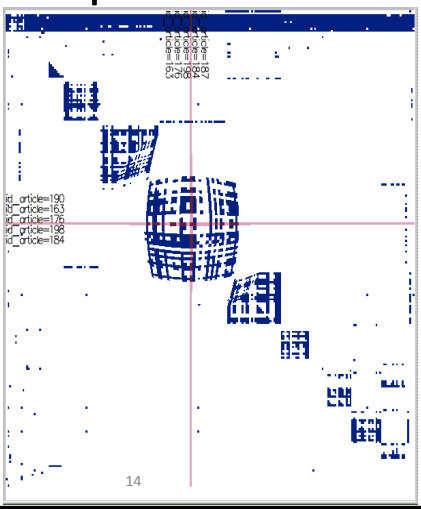




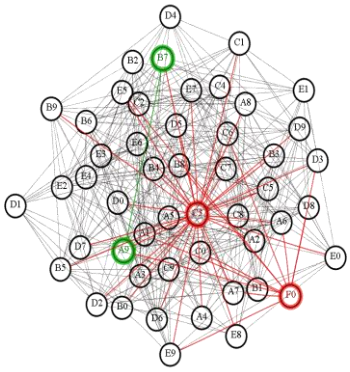


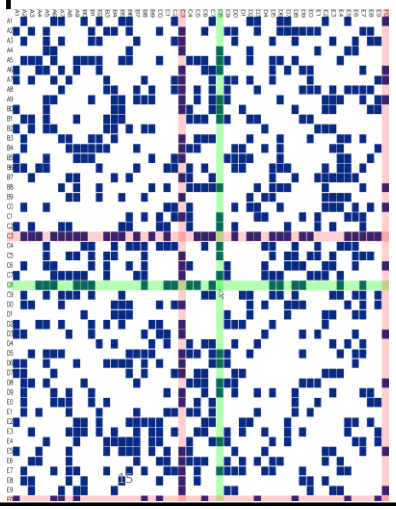
# Web Site Example



# Readability Experiment







## Controlled Experiment: Node Link Diagrams vs. Adjacency Matrices

**The Tasks:**

- Tasks related to the overview
  - Number of vertices
  - Number of arcs
- Tasks related to graph elements
  - Finding an element (a vertex, a link)
  - Finding the most connected vertex (a central actor, a pivot, a hub)
  - Finding a common neighbor
  - Finding a path
- Random graphs (3 sizes et 3 densities)
- 2 representations: Node-Link + Matrix

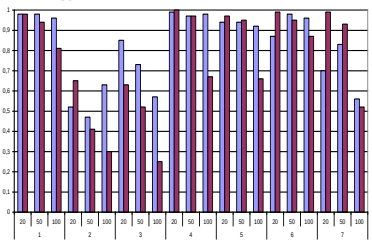
**Results:**

- Node-link diagrams are preferable for small sparse graphs (20 vertices)

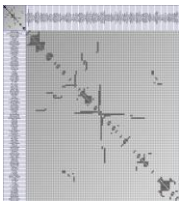
**Matrices are more readable *wrt* dense graphs and medium/large graphs ( > 20 vertices) *wrt* the selected tasks, except path finding**

**References:**

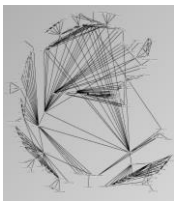
Mohammad Ghoniem, Jean-Daniel Fekete and Philippe Castagliola *Readability of Graphs Using Node-Link and Matrix-Based Representations: Controlled Experiment and Statistical Analysis*, Information Visualization Journal, 4(2), Palgrave Macmillan, Summer 2005, pp. 114-135.



Percentage of correct answers for the 7 tasks, 3 densities and 2 representations. NL in purple, Matrix in blue



# Matrix vs. Node-Link



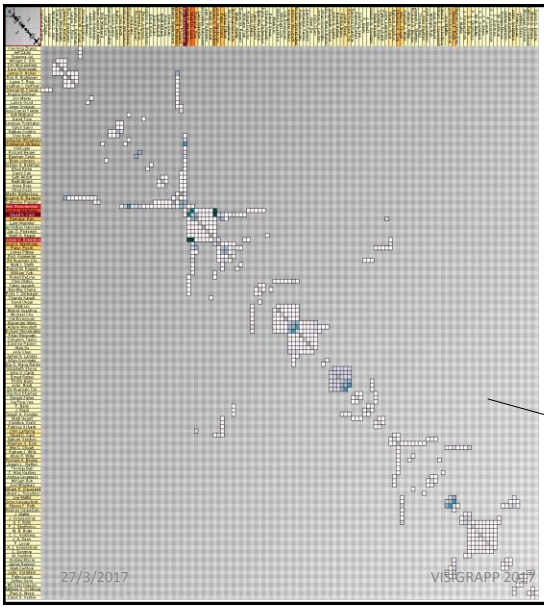
\+
- Usable without reordering
- No node overlapping
- No edge crossing
  - ➔ Readable for dense graphs
- Fast navigation
- Fast manipulation
  - ➔ Usable interactively
- More readable for some tasks

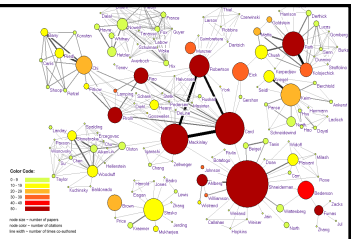
\-
- Less familiar
- Use more space
- Weak for path following tasks

\+
- Familiar
- Compact
- More readable for path following
- More effective for small graphs
- More effective for sparse graphs

\-
- Useless without layout
- Node overlapping
- Edge crossing
  - ➔ Not readable for dense graphs
- Manipulation requires layout computation



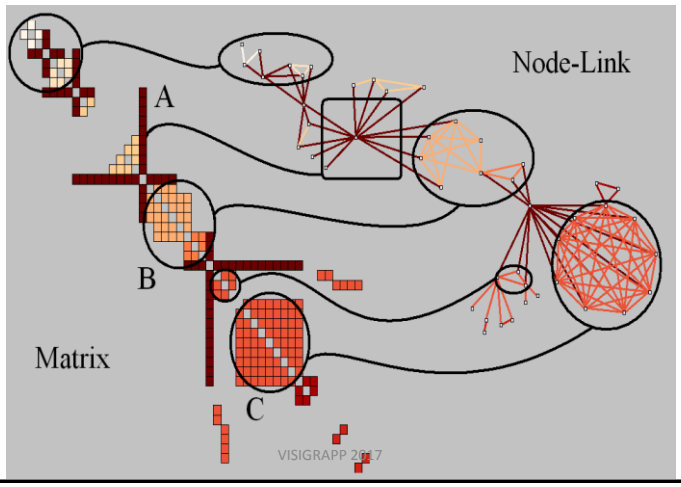




Communicate

Explore

## Visual Patterns with Ordered Matrices

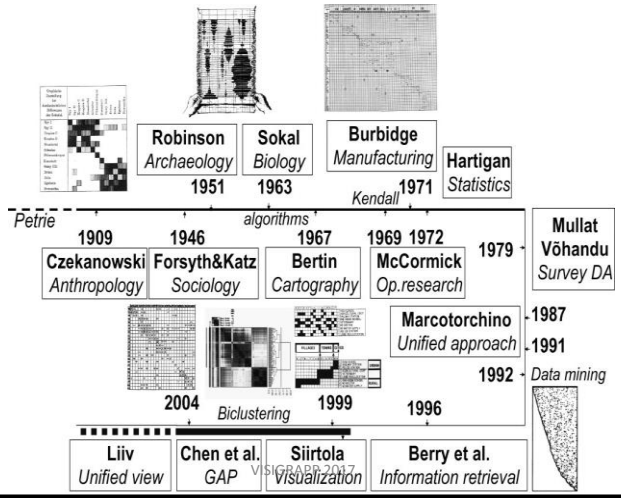





## The Ordering Problem

- Seriation instead of clustering
  - Finding a linear order for rows and columns
  - Postpone the decision of separating into clusters
  - Avoid creating clusters when they don't make sense
- Naïve approach:
  - Define an objective function (e.g. favor diagonal placement and dense blocks)
  - Try all permutations and keep the best wrt the function
- Problem: for a n×m table, there are n!×m! permutations
- Problem 2: there is no consensual objective function

## Liiv: visual abstract of the history of seriation from different disciplines

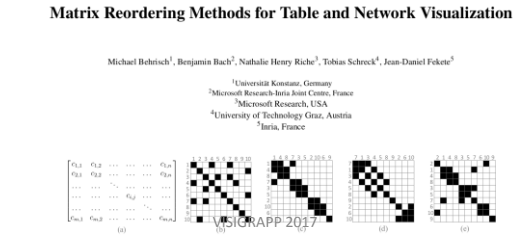





# Seriation and Matrix Ordering

Innar Liiv, Seriation and Matrix Reordering Methods: An Historical Overview, Statistical analysis and data mining, 2010 – Wiley

Michael Behrisch, Benjamin Bach, Nathalie Henry Riche, Tobias Schreck, Jean-Daniel Fekete. Matrix Reordering Methods for Table and Network Visualization. Computer Graphics Forum, Wiley, 2016, 35, pp.24. <hal-01326759>

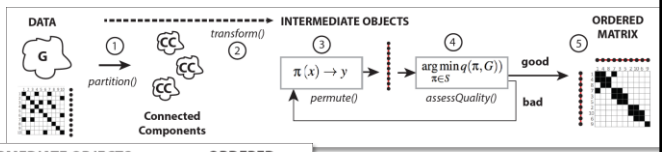

**Matrix Reordering Methods for Table and Network Visualization**

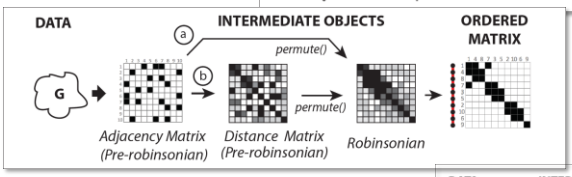
Michael Behrisch[1], Benjamin Bach[2], Nathalie Henry Riche[3], Tobias Schreck[4], Jean-Daniel Fekete[5]

[1]Universität Konstanz, Germany
[2]Microsoft Research-Inria Joint Centre, France
[3]Microsoft Research, USA
[4]University of Technology Graz, Austria
[5]Inria, France

# Matrix Ordering Pipeline

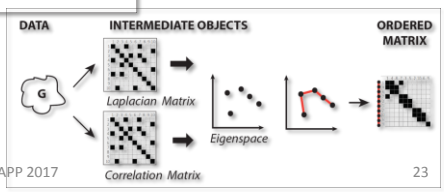
- Generic




Robinsonian Approach

Spectral Approach

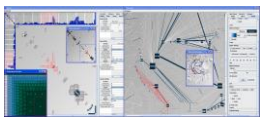


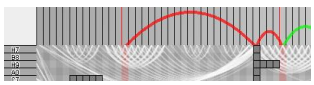


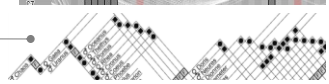

## Advances in Network Visualization: Improving Matrices

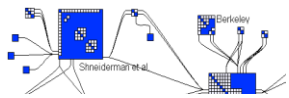
Several representations:

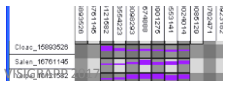
1. **Combined**
   - MatrixExplorer (Henry&Fekete InfoVis'06)
2. **Augmented**
   - MatLink (Henry&Fekete Interact'07, Best Paper)
   - GeneaQuilts (Bezerianos et al. InfoVis'10)
3. **Hybrid**
   - NodeTrix (Henry et al. InfoVis'07)
   - CoCoNutTrix (Isenberg et al. CG&A'09)
4. **Multiscale**
   - ZAME (Elmqvist et al. PacificVis'08)

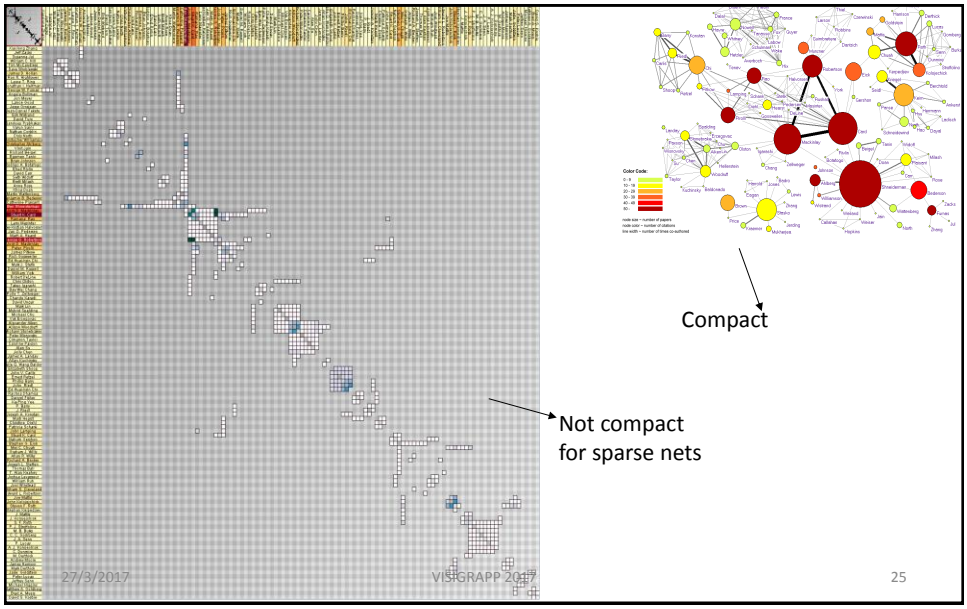








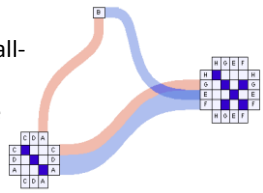


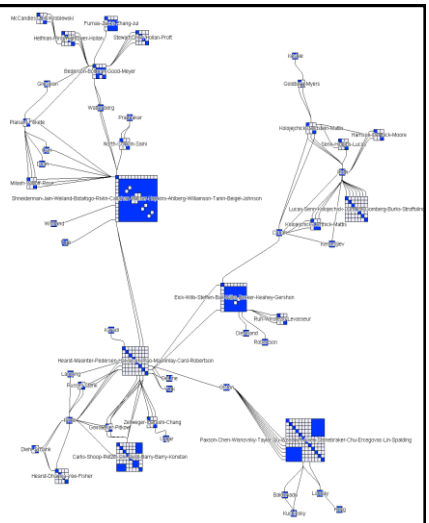



# NodeTrix [Henry et al.07]

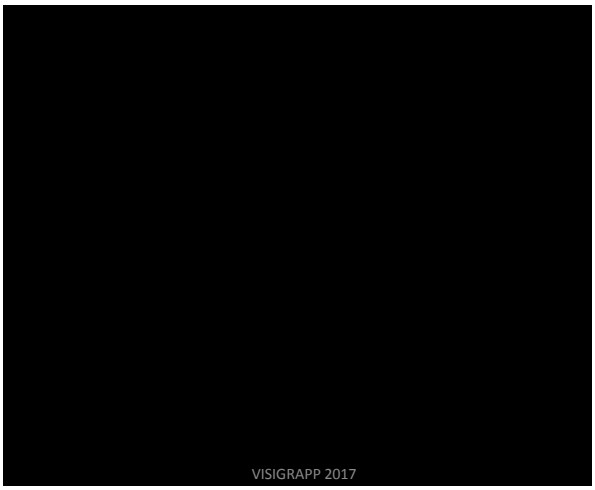
Hybrid representation

- Designed for small-world networks
  - Globally sparse
  - Locally dense
- Visualizing dense sub-graphs as matrices
- Interact to create, edit and remove the matrices





# Video



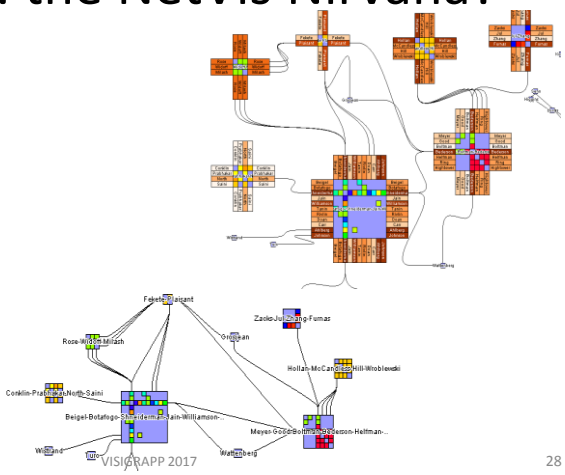


# NodeTrix: the NetVis Nirvana?

- ✓ Can you see every node?
- ✓ Can you count each node's degree?
- ✓ Can follow every link from its source to its destination?
- ✓ Can you identify clusters and outliers?

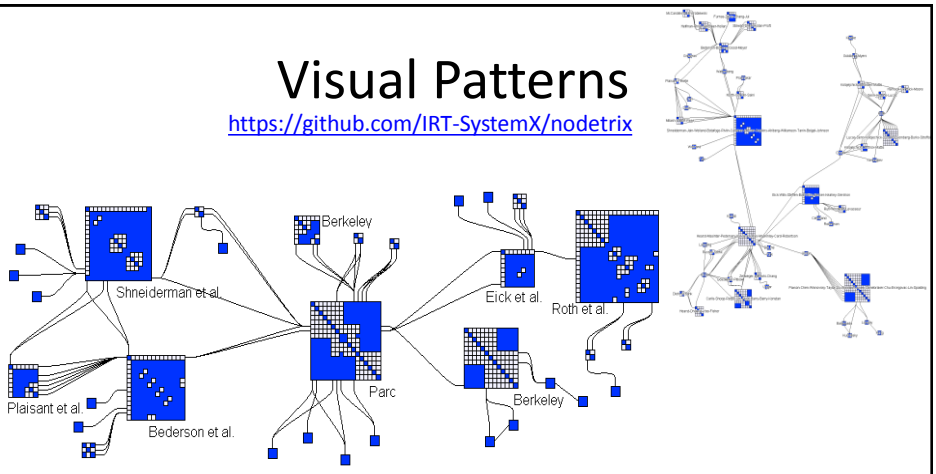
- Node Labels
- Link Labels (excentric labels?!)
- … even cluster labels
- Node Attributes
- Link Attributes
- … even clusters attributes
- Directed Graph (links width?!)

… But… beware the graphics overload!



# Visual Patterns

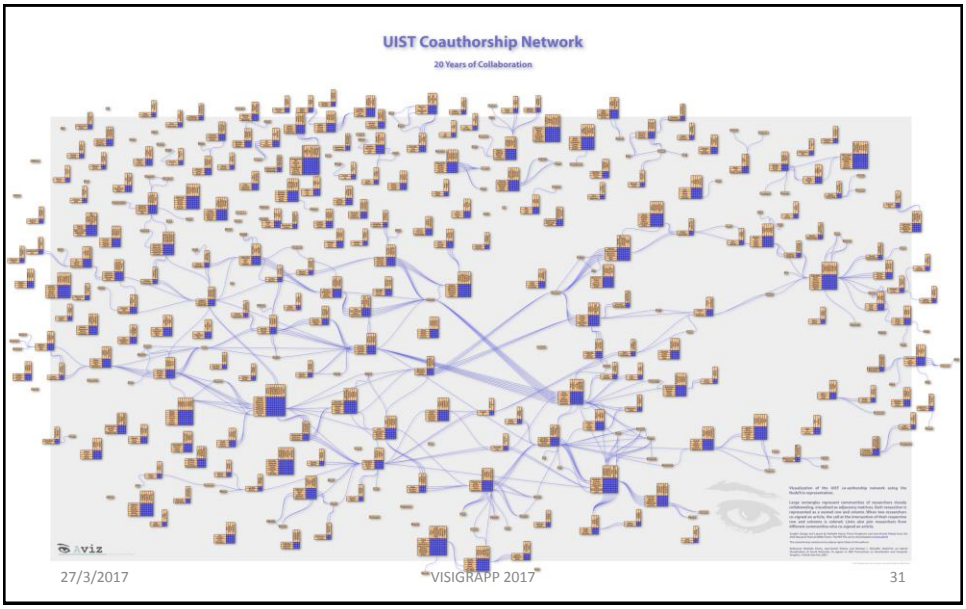
https://github.com/IRT-SystemX/nodetrix




Infovis Coauthorship (133 actors)



# Bigger

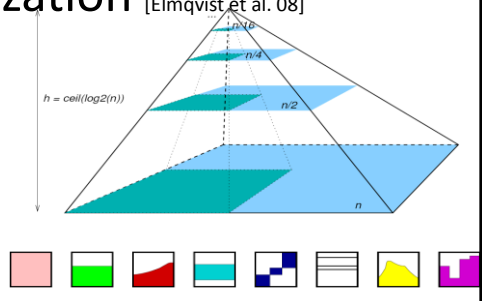


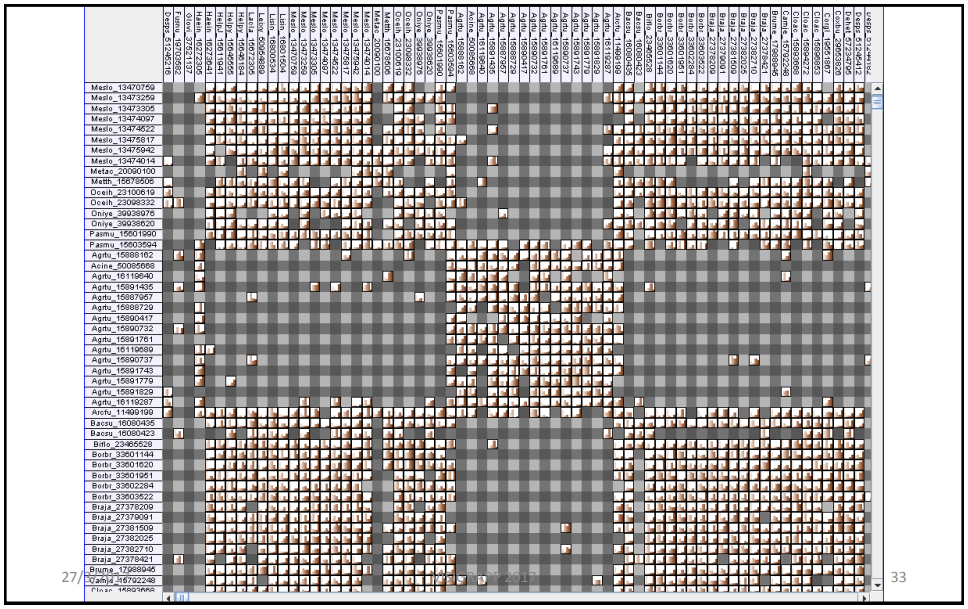



# ZAME: Interactive Large-Scale Graph Visualization [Elmqvist et al. 08]

Visualize very large networks:
- Larger than $10^7$ vertices and edges

- Reorder
- Create a pyramid
- Aggregate attributes
- Visualize using enhanced glyphs

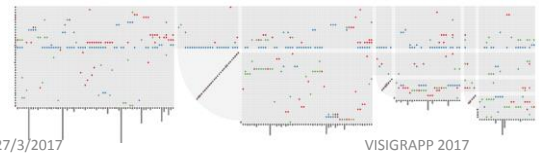







# Other Related Research

- Van Ham et al. 2004-2005 have shown techniques to navigate in large matrices
- Brandes&Nick 2011 have visualized temporal networks (friendship evolution)
- Dinkla et al. 2012 have introduced Compressed Adjacency Matrices



# Dynamic



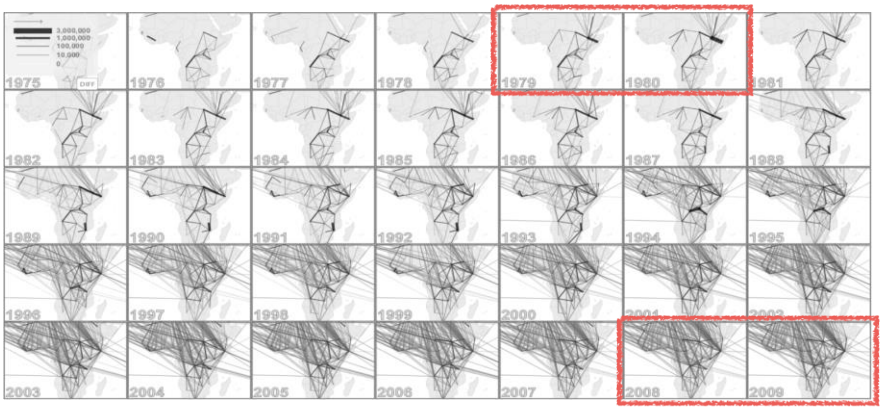
**PhD defense**

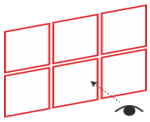
## Connections, Changes, and Cubes: Unfolding Dynamic Networks for Visual Exploration

- Benjamin Bach
- 9 May 2014

**Advisors:**
Jean-Daniel Fekete
Emmanuel Pietriga

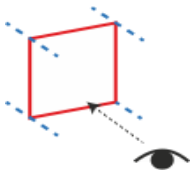
**Jury:**
Chantal Reynaud
Jarke J. van Wijk
Tim Dwyer
Silvia Miksch
Guy Melançon



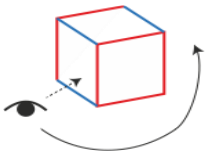
Boyandin et al., 2012

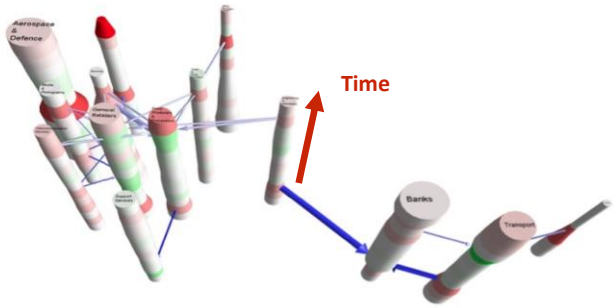




**Maray** Friedrich and Eades, 2001

**TempoVis** Ahn et al, 2011

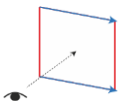


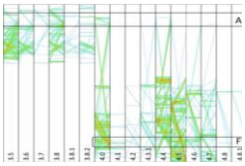



**2.5D visualization** Dwyer, 2004
**Visual Unrolling** Brandes & Corman, 2003
Gaertler & Wagner, 2005



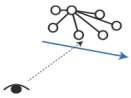
**Timelines**

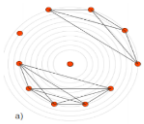




**Parallel Edge Splatting** Burch et al, 2011
**Massive Parallel Sequence Views** Willems et al, 2012
**GraphDice** Bezerianos et al, 2010
Reda et al, 2012

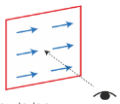
**Ego Network Representations**

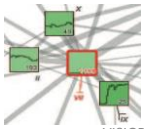




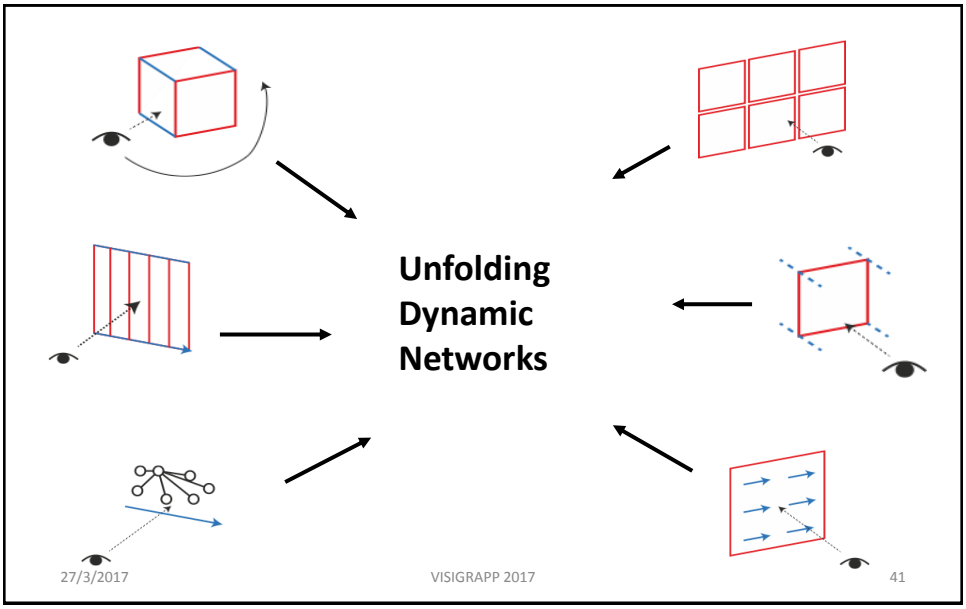
**1.5D Visualization** Shi et al., 2011
**Dynamic Ego Networks** Farrugia et al., 2011

**Temporal Aggregation**

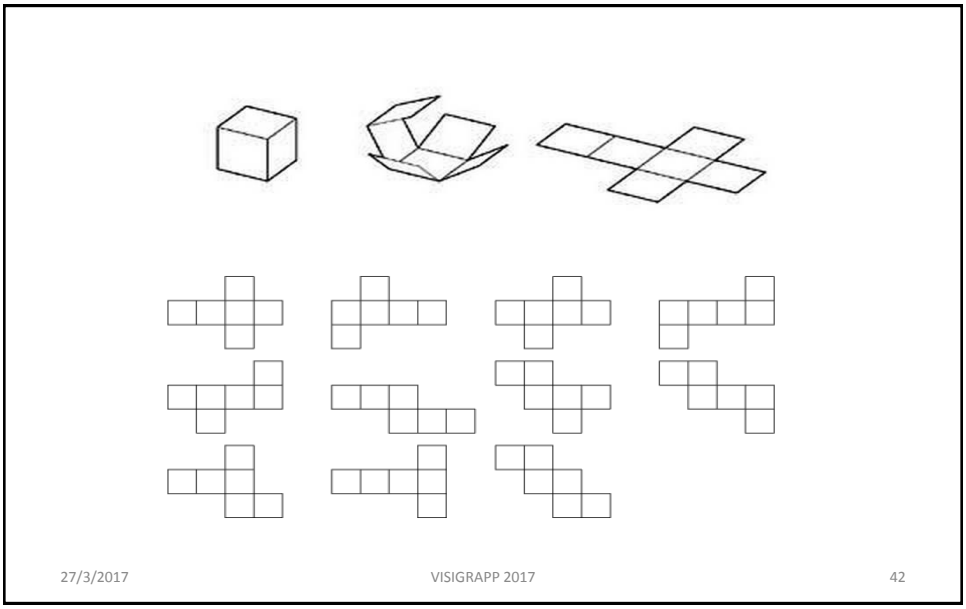


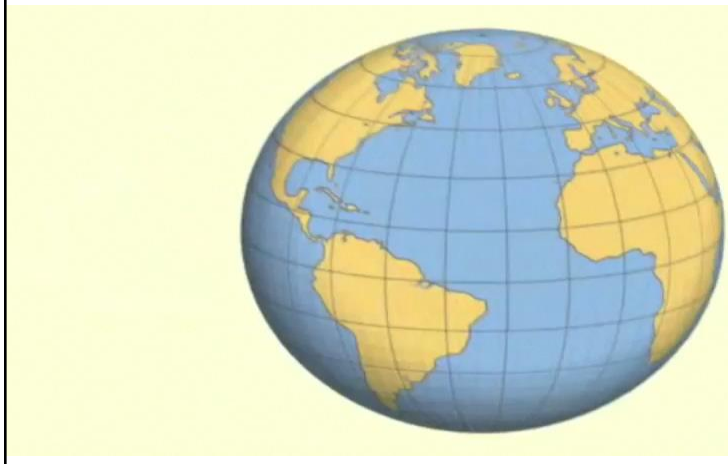


Collberg et al. 2003
**Gestalt Lines** Brandes & Nick, 2011


Unfolding
Dynamic
Networks

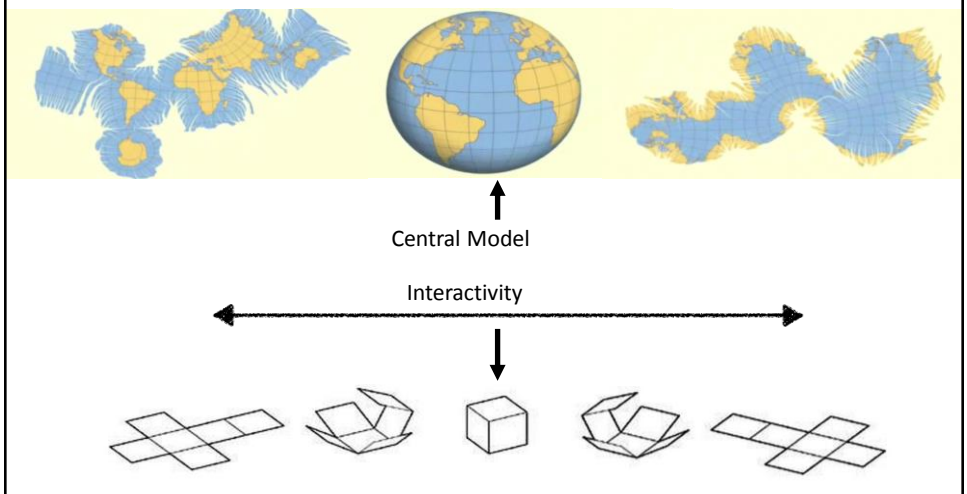


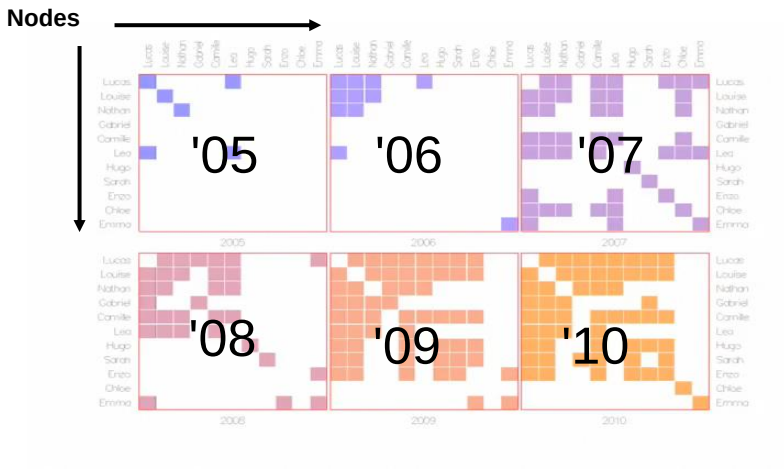




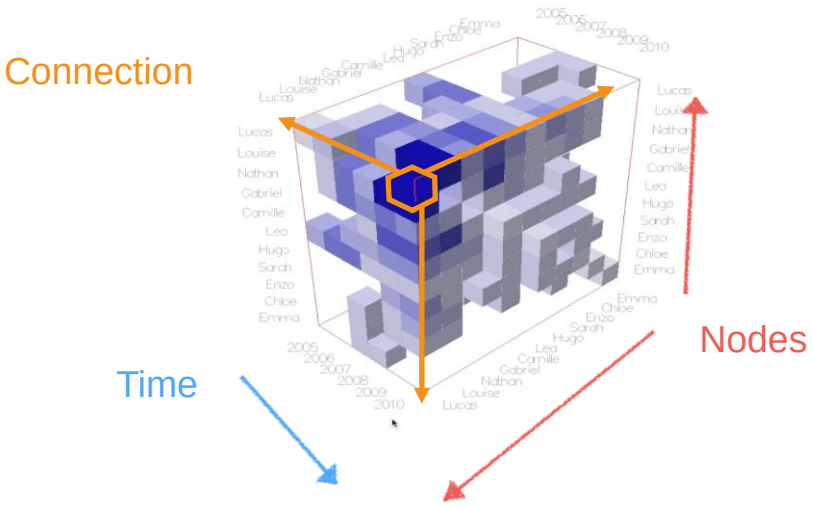
[van Wijk 2008]

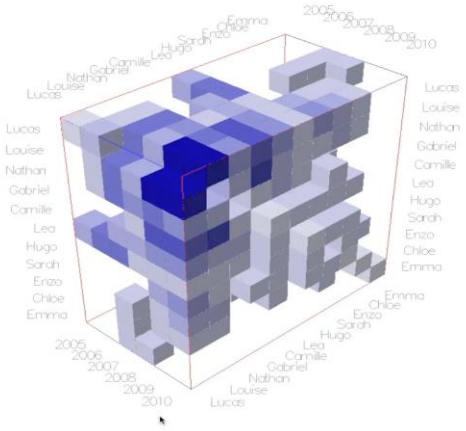





Nodes
'05
'06
'07
'08
'09
'10




Connection
Time
Nodes




**Ceci n'est pas une visualisation 3D**



# Cubix: http://aviz.fr/cubix

Visualizing Dynamic Networks
with Matrix Cubes

submitted to
CHI2014

## Visualizing Dynamic Networks with Matrix Cubes [CHI 2014]

- Works for relatively small networks up to 1000 time steps
- Usable with a bit of training (10-20mn)
- Try it on your data: http://aviz.fr/cubix

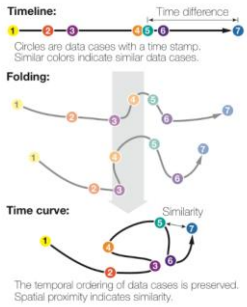
- But what about bigger networks?



## Time Curves [InfoVis 2015]

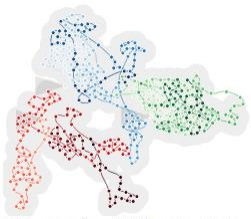
*Benjamin Bach, Conglei Shi, Nicolas Heulot, Pierre Dragicevic*

- Compute distances between multiples networks
- Use MDS to create a 2D map
- Connect them by time




http://www.aviz.fr/~bbach/timecurves/

# Time Curves on Networks



fMRI scans from 3 different subjects.

Parkinsons control RC4201 1 graphdat 20 (150)

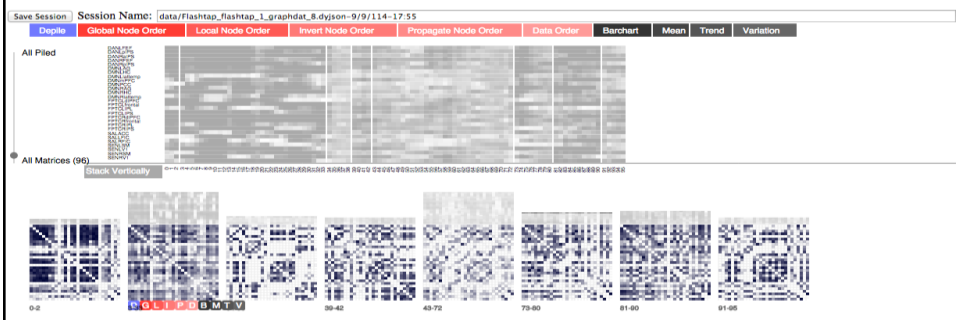
Parkinsons control RC4205 1 graphdat 20 (150)



# Small MultiPiles: Clustering Dynamic Networks

[Bach et Al. EuroVis 2015]

http://visualizingbrainconnectivity.org/multipiles



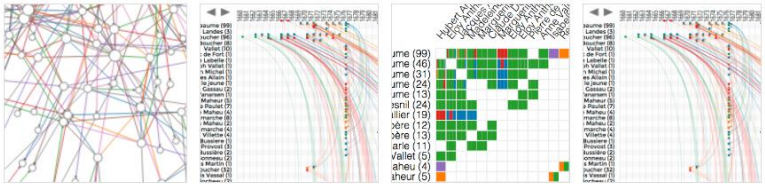
# Video

# Challenges

- Moving from research prototypes to product
- Study what is a good order
  - What structures can we see?
  - What algorithm will reveal what structure?
  - How to characterize data to fit with algorithms?
- Scalability
- Multivariate networks (several attributes on the vertices and edges)
- Dynamic networks

THE VISTORIAN (BETA)

Interactive Online Visualizations for Dynamic Networks.



Benjamin Bach, Microsoft-Research, Inria
Nathalie Henry Riche, Microsoft Research
Nicole Dufournaud, EHESS
Jean-Daniel Fekete, Inria

https://vistorian.net/



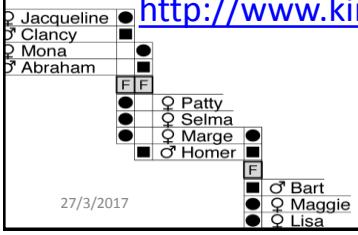
# Transferring Network Visualization

As researchers, we do not make products, but we tried many strategies:

- Publishing in high-ranked journals and conferences
- Writing open-source prototype applications
- Writing open-source libraries
- Writing open-source web sites

# Prototype Applications

- Cubix: http://www.aviz.fr/cubix
- GeneaQuilts: http://www.aviz.fr/geneaquilts
  - Integrated with **PUCK** (Program for the Use and Computation of Kinship data) http://www.kintip.net/puck-topmenu-37






# Open Source Libraries

- Reordering of Networks and Tables:
  - https://github.com/jdfekete/reorder.js
- NodeTrix library:
  - https://github.com/IRT-SystemX/nodetrix

# Open Source Web Sites

- Multipiles: http://www.aviz.fr/~bbach/multipiles/
- Time Curves: http://www.aviz.fr/~bbach/timecurves/
- Vistorian: http://vistorian.net/



# Conclusion

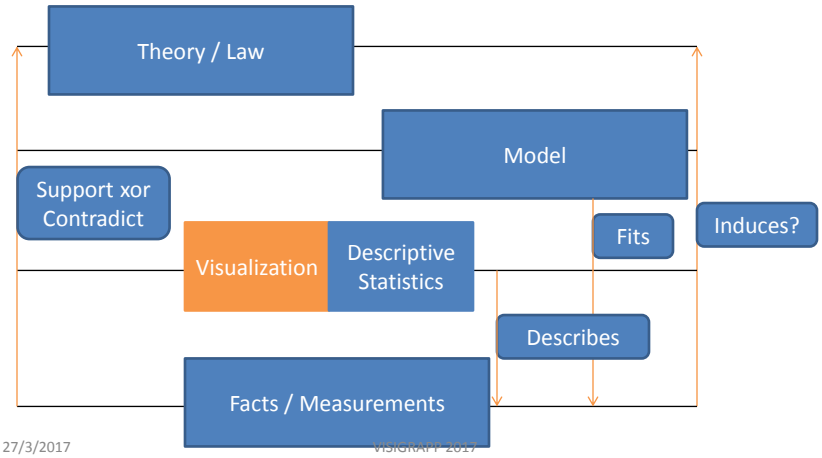
- Exploring complex data is possible with novel visualizations
  - To make sense of datasets, check for quality, etc.
- It requires a bit of time to understand the visual mapping
  - About 10mn to 1h
- It also requires a bit of time to learn the interactions
- Visualization Literacy is necessary to realize how much you will gain from investing

# Conclusion

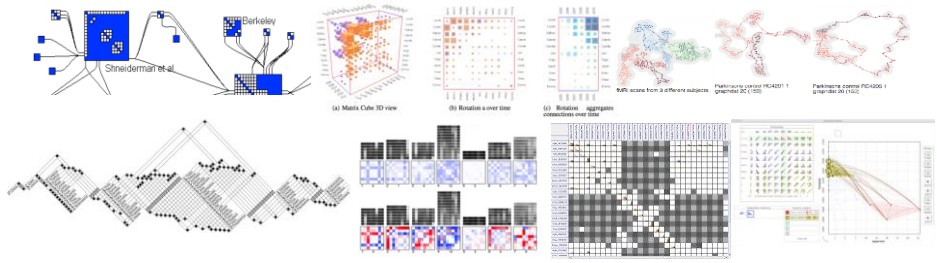
- Visualization of Networks has greatly improved in the last ten years
- Novel representations are denser and more expressive
  - Though they require a little training
- Huge and dense networks can be visualized
  - Relations between clients, suppliers, employees
  - Aggregated over long periods of time
- They need integration in complete systems (commercial or not)



# Where does Visualization Stands?



# Thanks! Questions?




http://www.aviz.fr/Research/Projects