

Exploring social networks with Matrix-based representations

Nathalie Henry

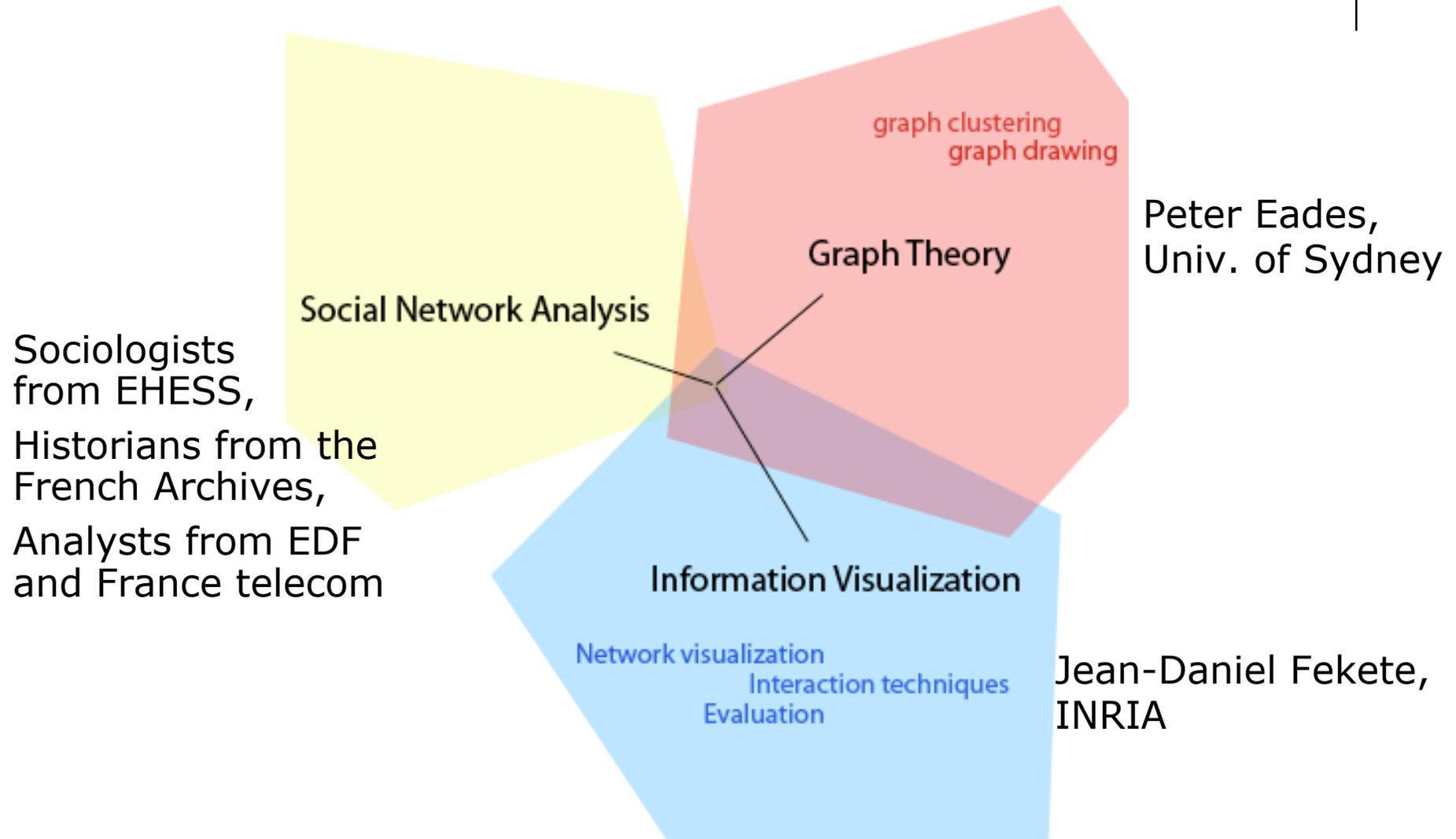
Co-supervisors:

Jean-Daniel Fekete & Peter Eades



The University of Sydney
Australia

Background



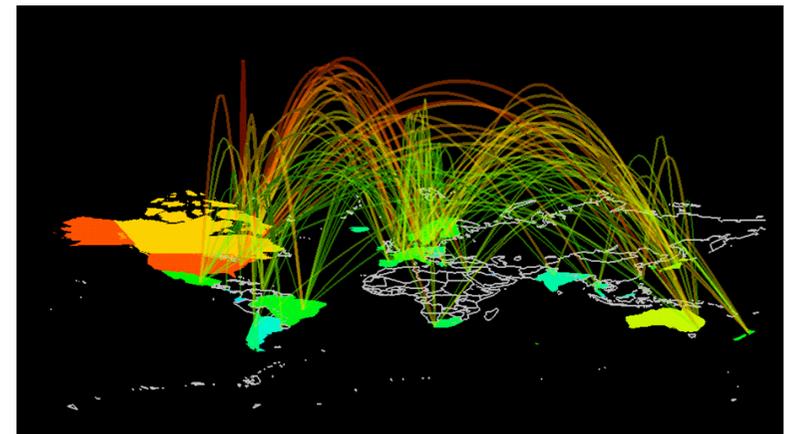
Social Networks

- What is it?
 - A set of *actors* connected by *relations*
- a **HOT** topic
 - Online communities (FaceBook, Flickr)
 - Online collaboration (Wikipedia, Sourceforge)
 - Scientific collaboration
 - And more...
 - The WEB
 - Diseases transmission networks
 - Terrorists networks



FaceBook Interactive Graph

Internet Traffic [Eick et al.]



Analyzing social networks

- Answering questions with statistics
- Exploratory Data Analysis [Tukey,77]
 - Answering question you did not know you had
 - **Looking at** the raw data from **different perspectives**

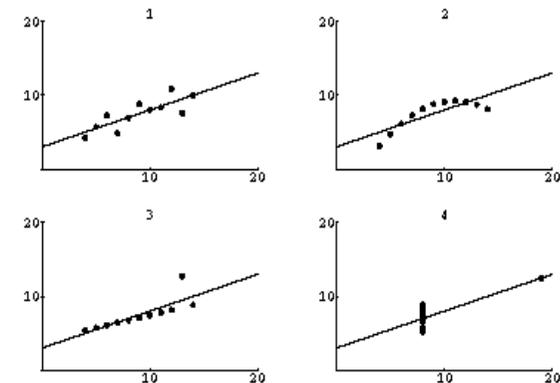
Anscombe's numbers

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	6.10

Statistics

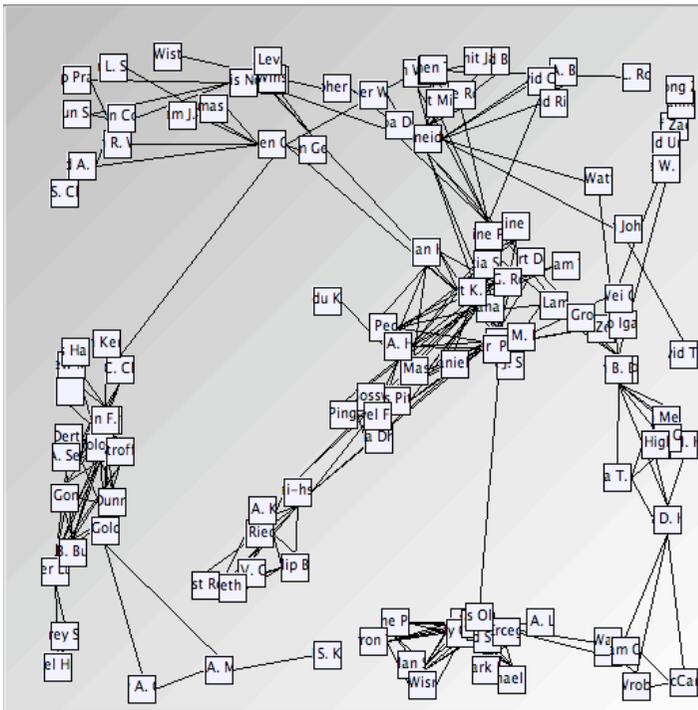
N: 11.0
 mean X's : 9.0
 mean Y's : 7.5
 standard error of slope estimate: 0.1
 sum of squares: 110.0
 regression sum of squares: 27.5
 residual sum of squares of Y: 13.8
 correlation coefficient: 0.8
 r squared: 0.7
 regression line: $Y=3+0.5X$

Visual representation



Network Visualization

- What social scientists start with:



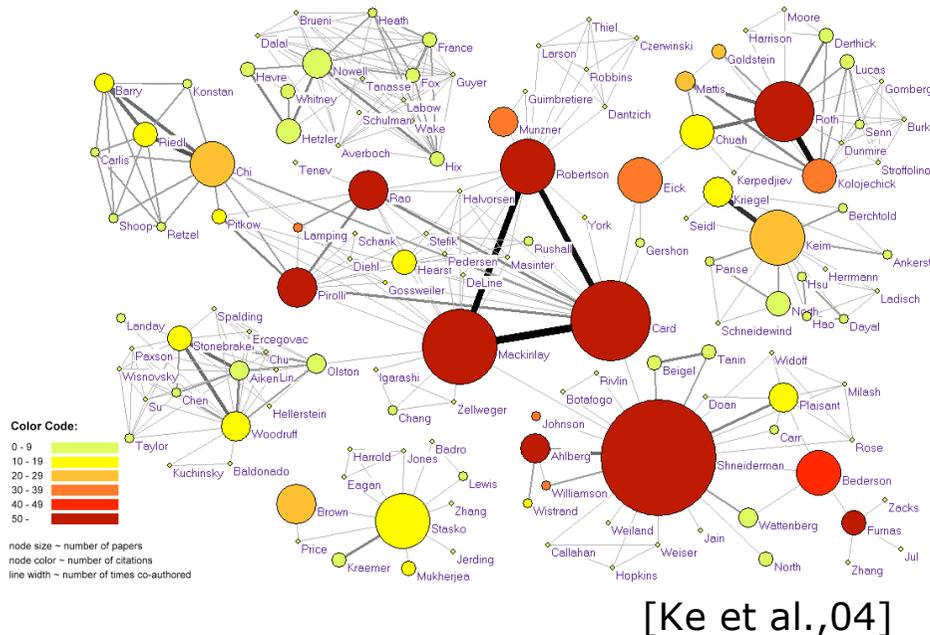
100+ actors - InfoVis community

- Node-Link Diagram
 - Overlapping nodes
 - Edge-crossing

⇒ Identify connections

Network Visualization

○ What social scientists end with:

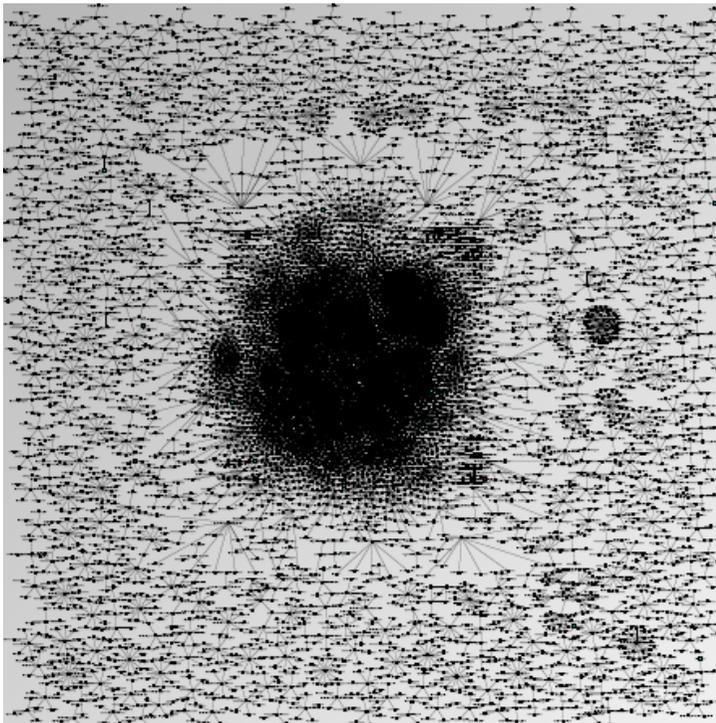


- Most important tasks
 - Communities
 - Central Actors

⇒ Discover the structure of the network

Large Network Visualization

- What social scientists start with:



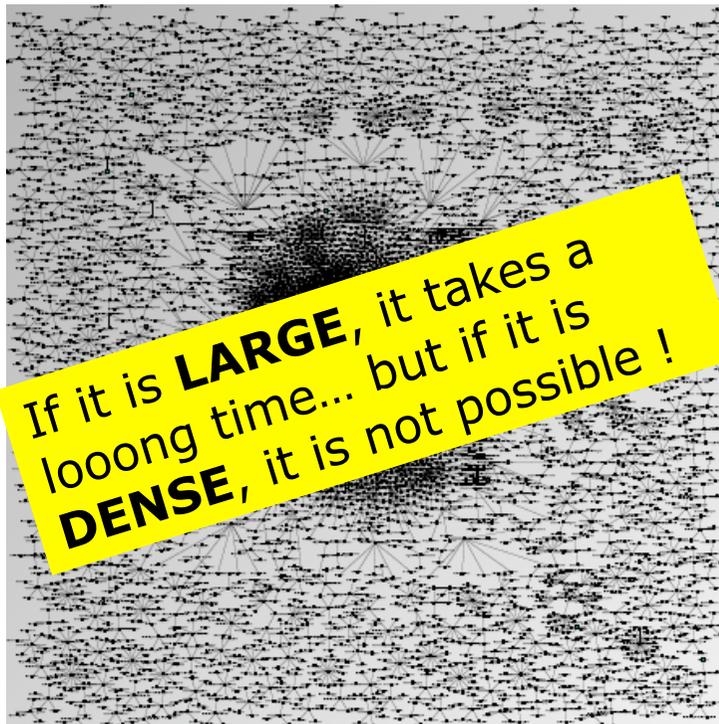
4000+ actors

- Node-Link Diagram
 - Overlapping nodes
 - Edge-crossing

⇒ ~~Identify connections~~

Large Network Visualization

- What social scientists end with:

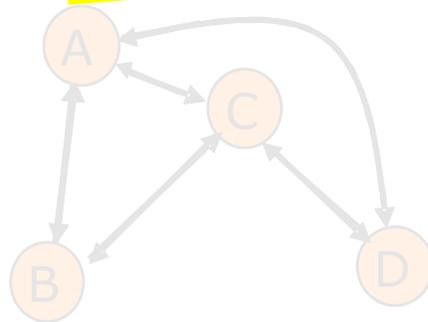


4000+ actors

What are the solutions?

- Sampling, filtering
- Clustering into meta-nodes
- Alternative representations as adjacency matrices

So, let's investigate the potential of MATRICES !



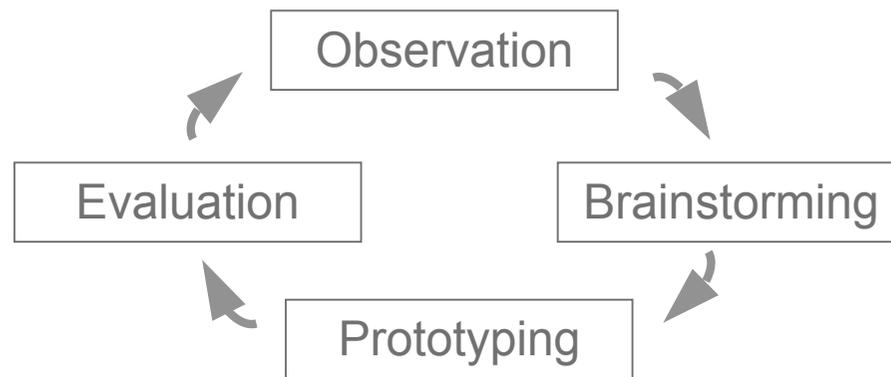
	A	B	C	D
A	0	1	1	1
B	1	0	1	0
C	1	1	0	1
D	1	0	1	0

A year of emails



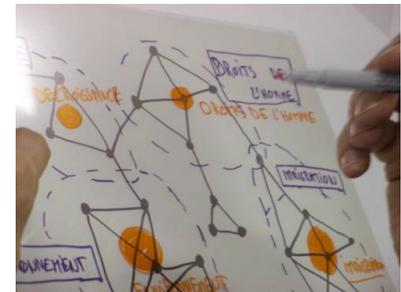
My Approach

○ Participatory Design

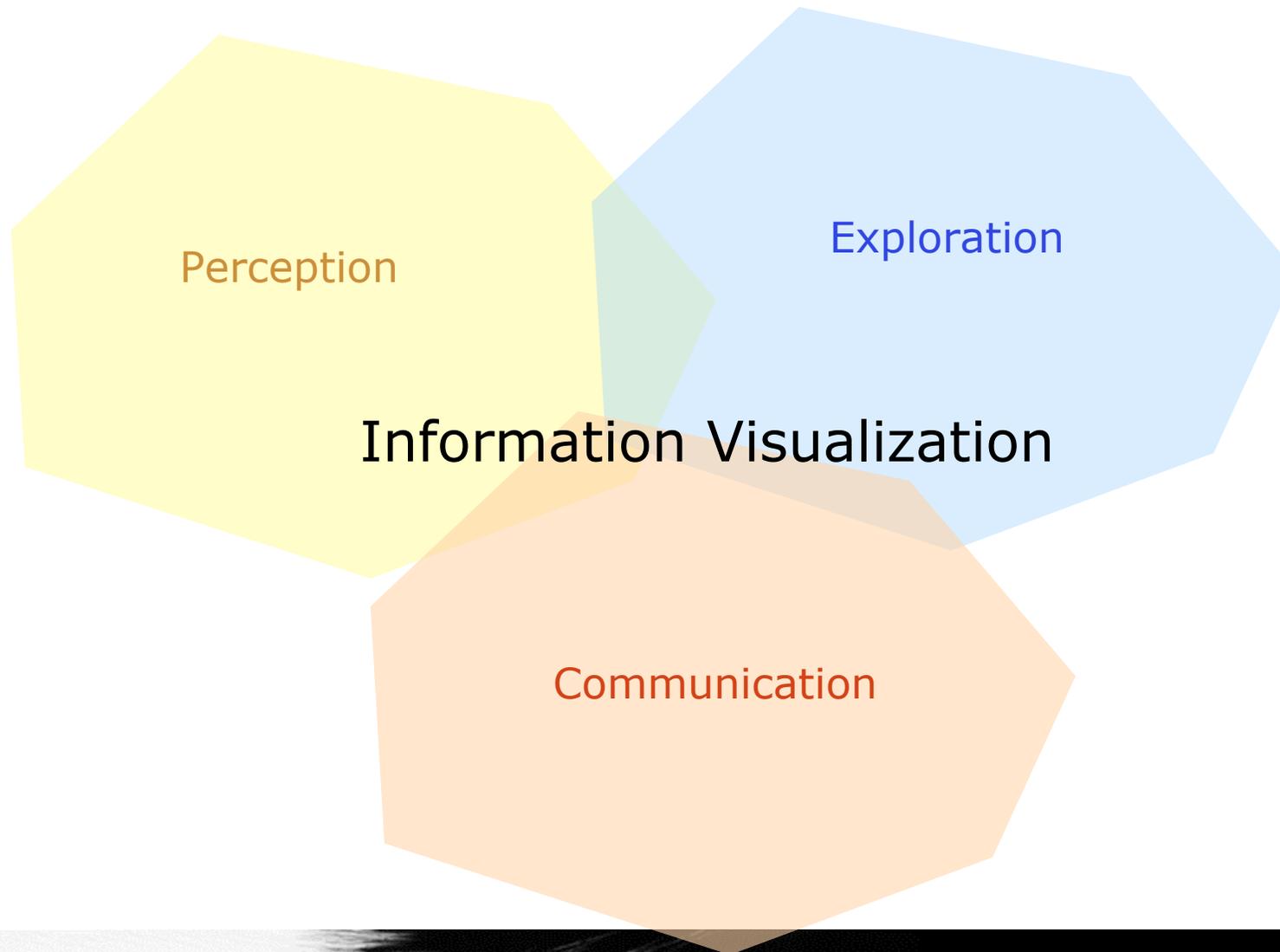


○ Outcomes

- They use statistics, graph analytics
- They use mainly node-link diagrams
- But they use sometimes matrices



My Approach



Itinerary of my PhD

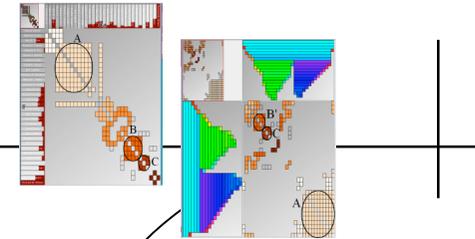
1. Make matrices usable

2. Combine matrix+node-link

3. Augment matrices

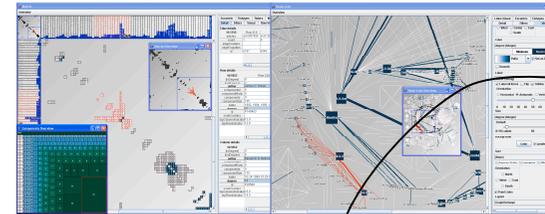
4. Merge matrix+node-link

[Henry et al., Infovis07]
Case study [Henry et al., IJHCI07]

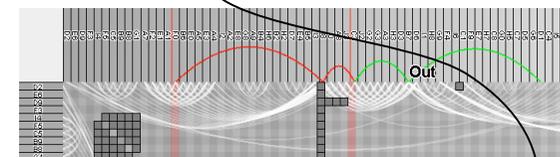


[Survey in progress]

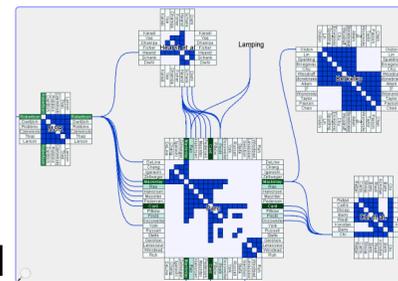
[Henry and Fekete, IHM06&Infovis06]



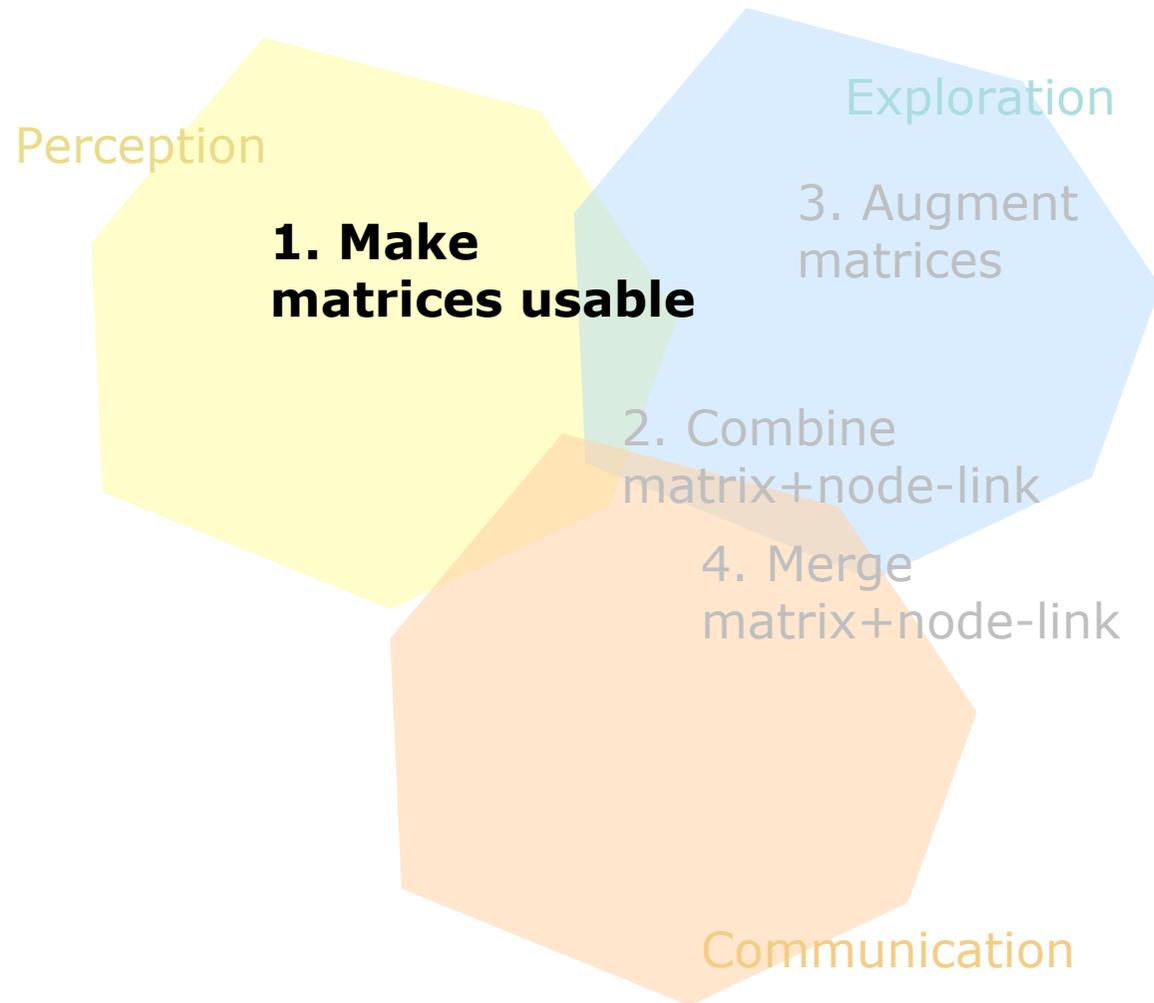
[Henry and Fekete, Interact07]



Interaction technique
[Elmqvist et al., CHI08]



Outline



1. How to make matrices usable?

○ J. Bertin, 1967

High-level structure?

- Groups
- Trends
- Outliers

6	6	2	5	3	BELGIE LUXEMBOURG
32	6	5	45	15	DEUTSCHLD
37	61	69	29	59	FRANCE
16	23	21	9	14	ITALIE
9	4	3	12	9	NEDERLAND
100	100	100	100	100	
BOVINS	CHEVAUX	OVINS	PORCS	VEAUX	

Annotations: A red 'Z' is above the cell (2,3) with an arrow pointing to it. A red 'Y' is to the right of the cell (4,5) with an arrow pointing to it. A red 'X' is below the cell (5,4) with an arrow pointing to it.

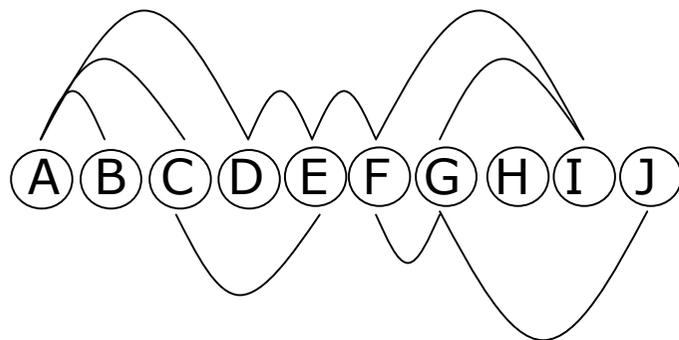
⇒ Make them VISUAL

⇒ REORDER their rows and columns

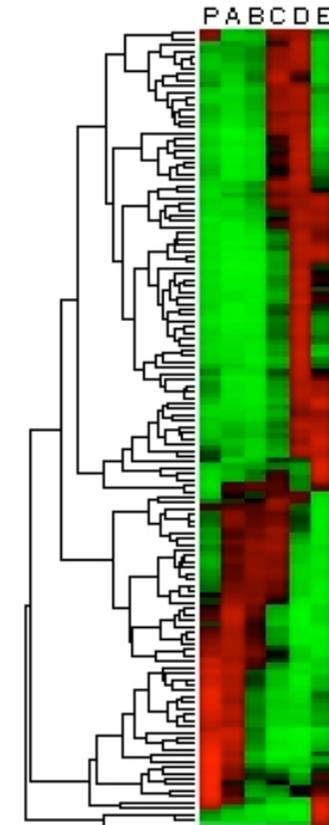
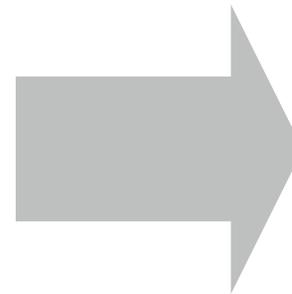
Reordering methods

[Survey to be published]

- Lot's of methods !
 - Table-based ordering methods
 - Graph linearization
 - Mine! (mixed approach)



Graph Linearization

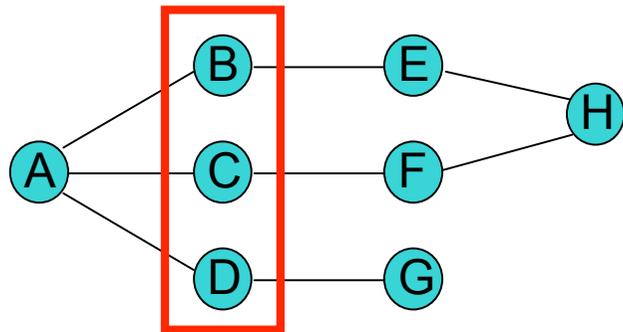


Hierarchical clustering of microarray data

Mixed approach

[Henry and Fekete, IHM'06]
[Henry and Fekete, InfoVis'06]

- Place actors with similar connection patterns beside each other



	A	B	C	D	E	F	G	H
A	0	1	1	1	0	0	0	0
B	1	0	0	0	1	0	0	0
C	1	0	0	0	0	1	0	0
D	1	0	0	0	0	0	1	0
E	0	1	0	0	0	0	0	1
F	0	0	1	0	0	0	0	1
G	0	0	0	1	0	0	0	0
H	0	0	0	0	1	1	0	0

	A	B	C	D	E	F	G	H
A	0	1	1	1	2	2	2	3
B	1	0	2	2	1	3	3	2
C	1	2	0	2	3	1	3	2
D	1	2	2	0	3	3	1	4
E	2	1	3	3	0	2	4	1
F	2	3	1	3	2	0	4	1
G	2	3	3	1	4	4	0	5
H	3	2	2	4	1	1	5	0

→ Add information to the adjacency matrix

Reordering methods

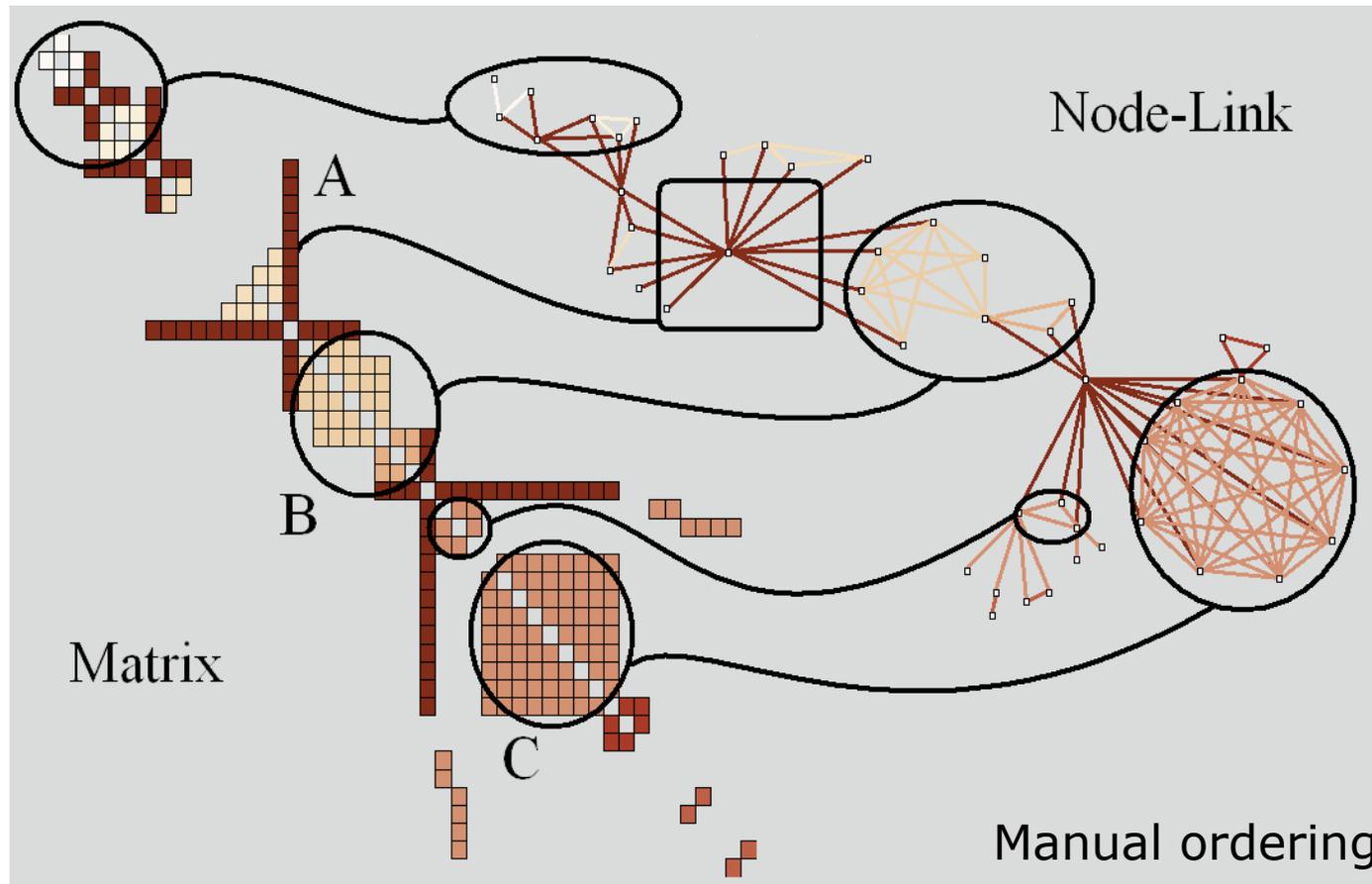
There is plenty of methods...

... but the real question is:

⇒ What is a GOOD ordering?

What is a good ordering?

... for performing the analysis of a social network



Most important tasks

- Communities: B and C
- Central Actors: A

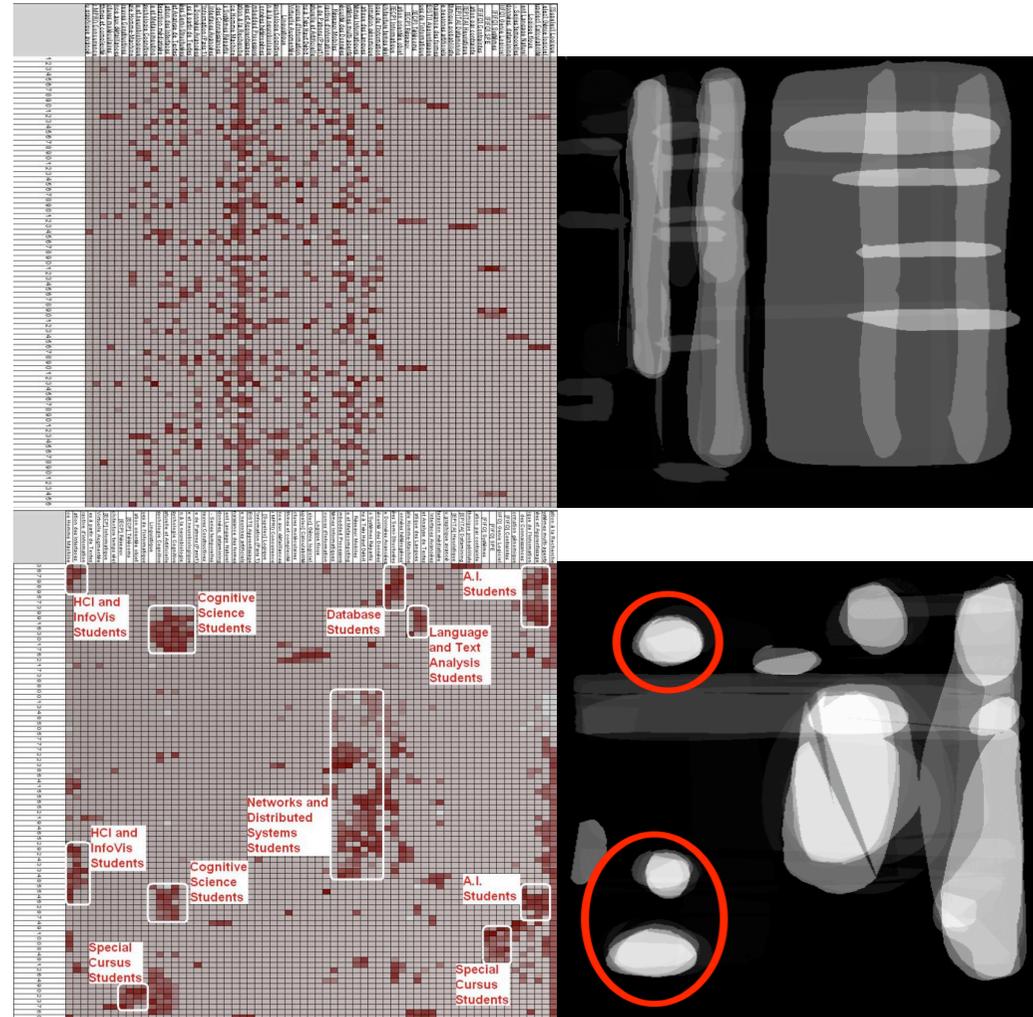
What is a good automatic ordering?

- Algorithms designed from formal measures
- Can we characterize them according to the **visual features** they produce?

Empirical study

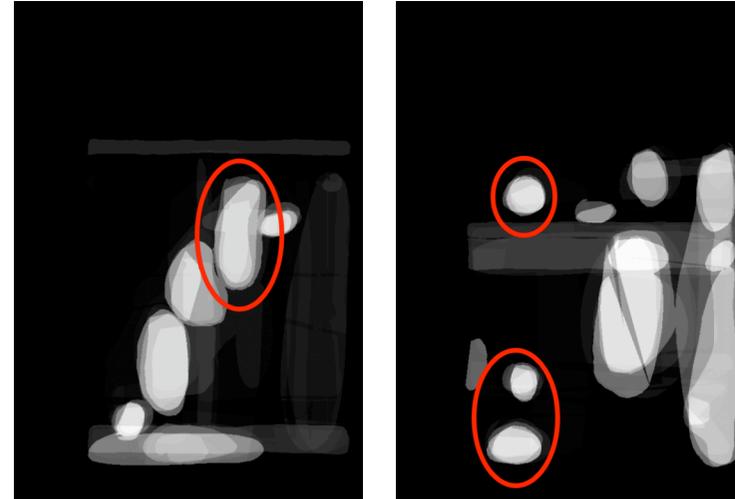
[Henry and Fekete, Beliv'06]

- How people perceive groups?



Is there A good ordering?

- For a specific
 - visual feature?
 - type of data?

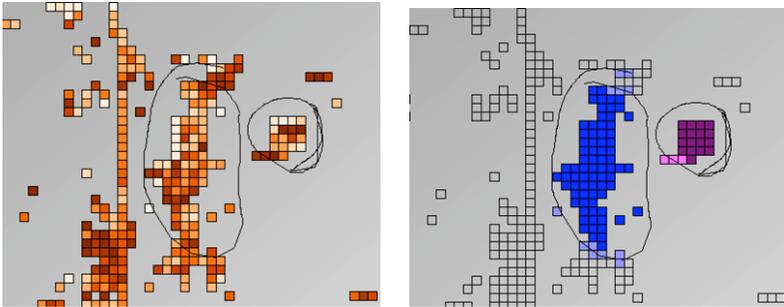


- No :(
- But it saves them time to **start from somewhere!**
- Analysts need **several orderings** to find a consensus in the data

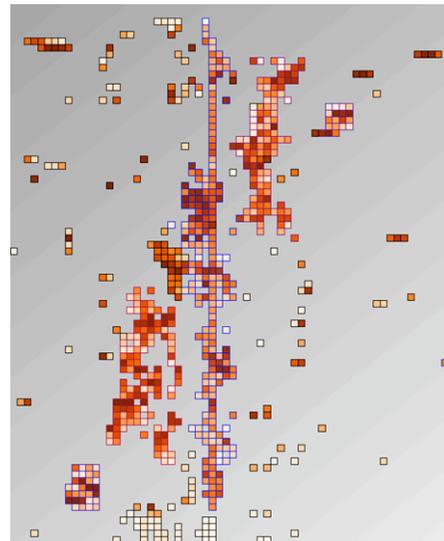
⇒ I focused on assisted reordering and interactions to find consensus

Interacting to find a consensus

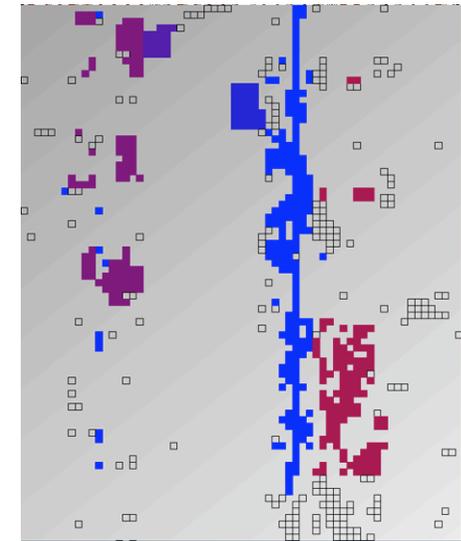
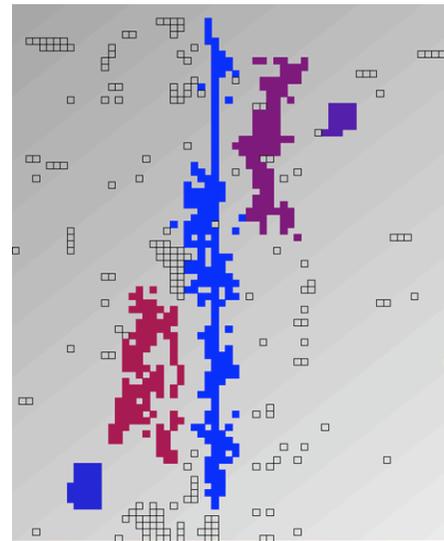
[Henry and Fekete, IHM'06]
[Henry and Fekete, InfoVis'06]



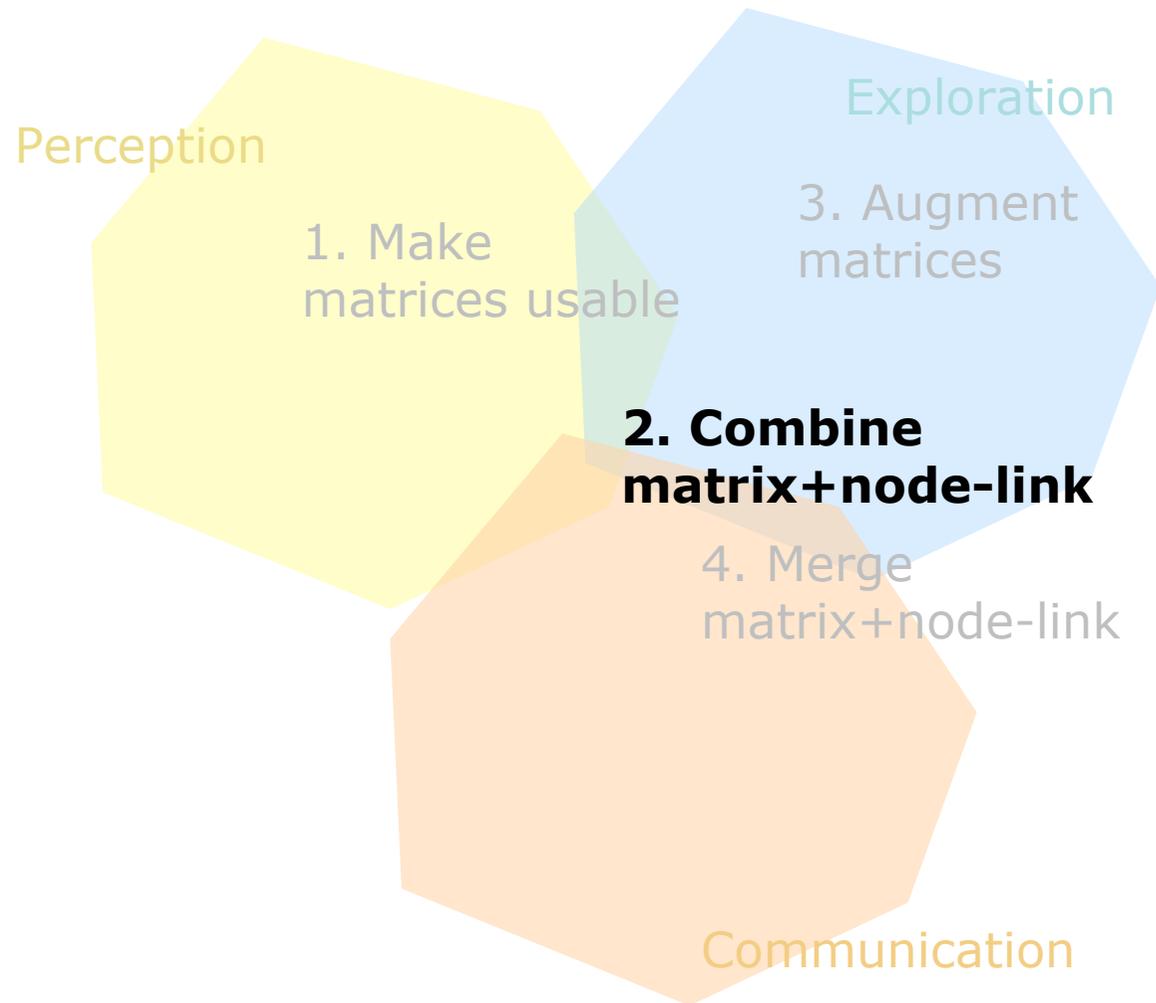
> interactive (fuzzy) clustering



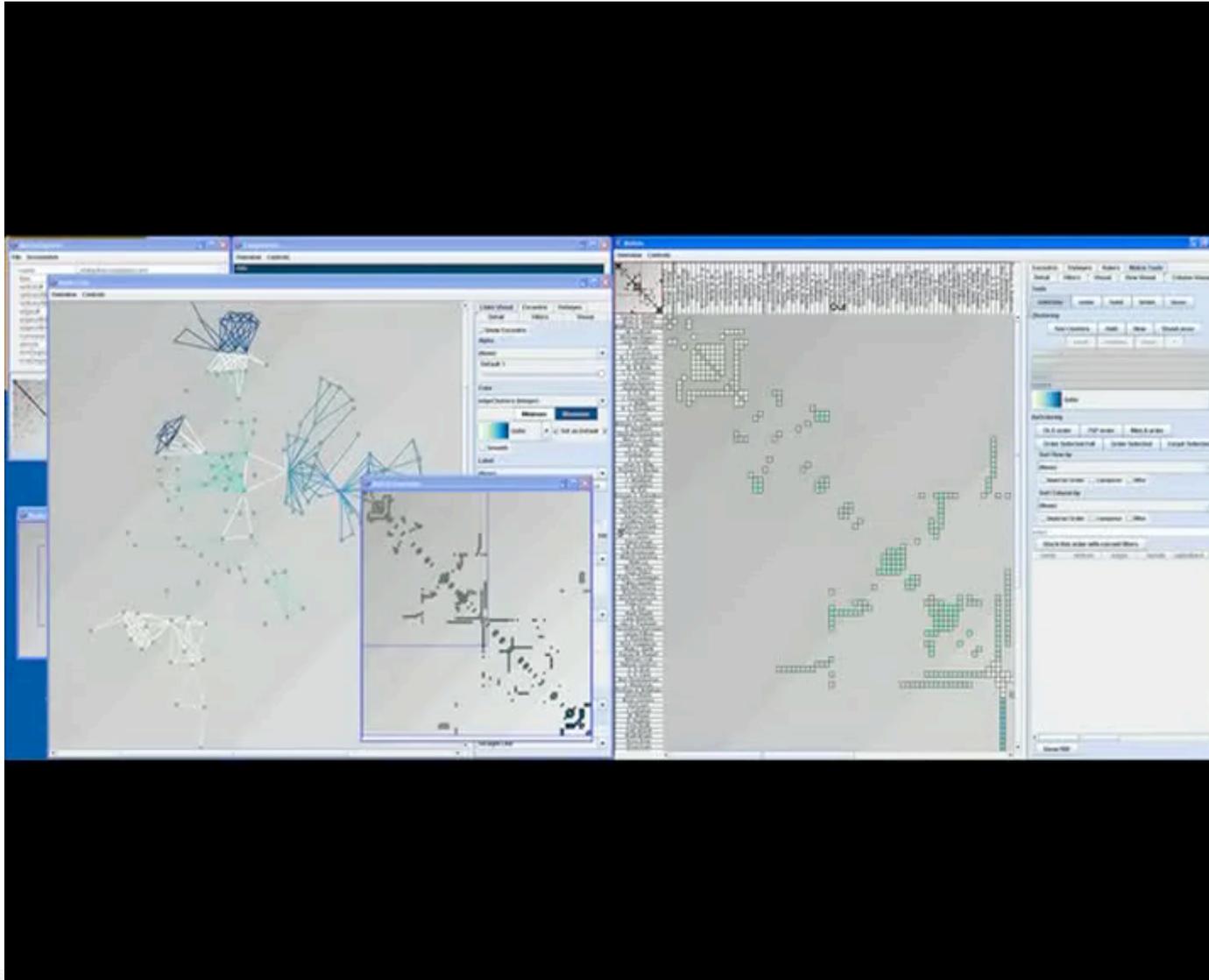
> interactive ordering



Outline



2. Combining best of both worlds with MatrixExplorer



MatrixExplorer

[Henry and Fekete, IHM'06]
[Henry and Fekete, InfoVis'06]

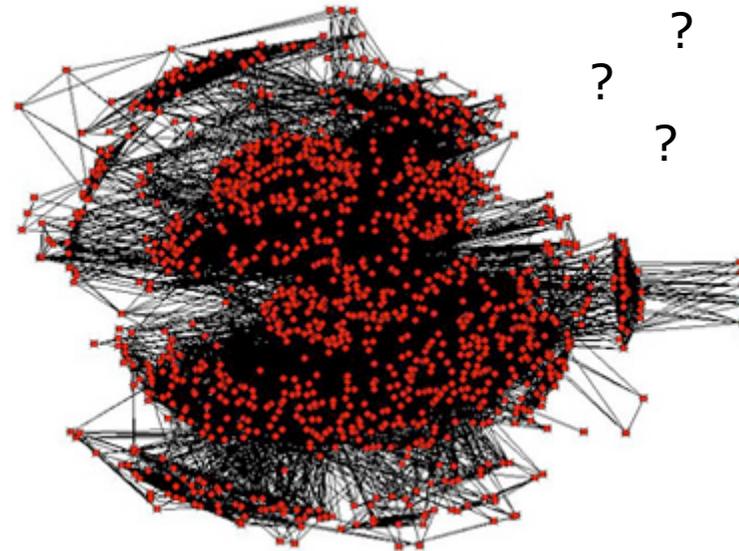
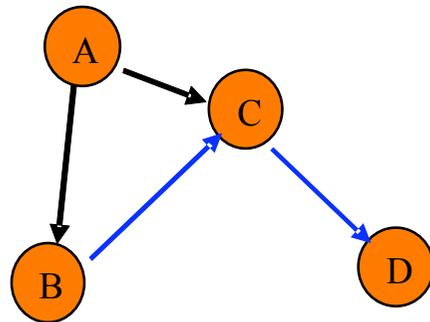
- Coordinated views
- Interaction to explore
- Our hypothesis:
 - Matrix to explore
 - Node-link to communicate
- Our observations:
 - Node-link for certain tasks, matrix for others
 - Cognitively demanding to switch back and forth
 - To find a consensus, they use both representations and different layouts/orderings

Exploring with matrix or node-link?

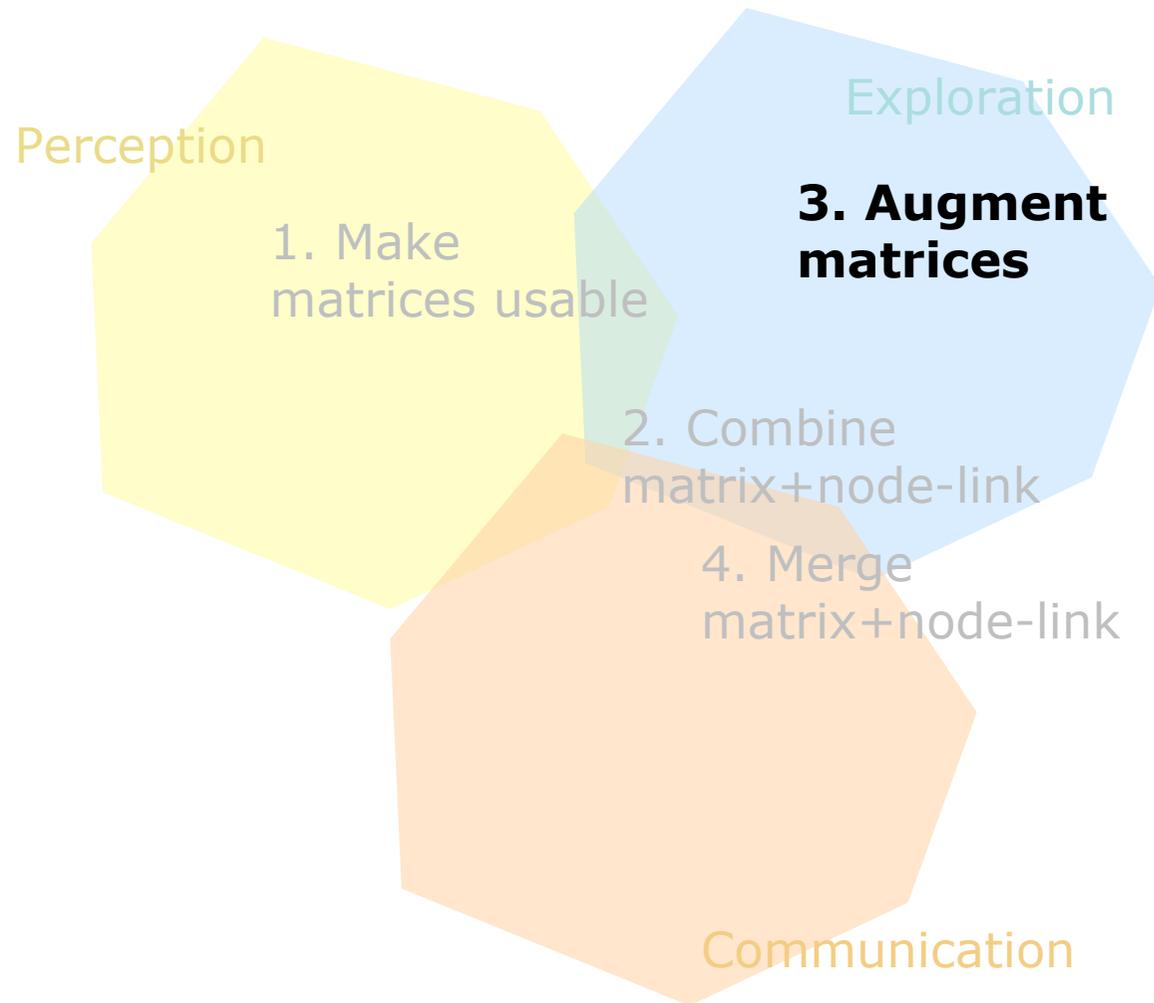
- Social networks
 - Sparse (genealogical tree) \Rightarrow node-link
 - Large/Dense (a year of emails) \Rightarrow matrix
- Why are users switching back and forth when exploring large/dense networks?

Matrices weakness

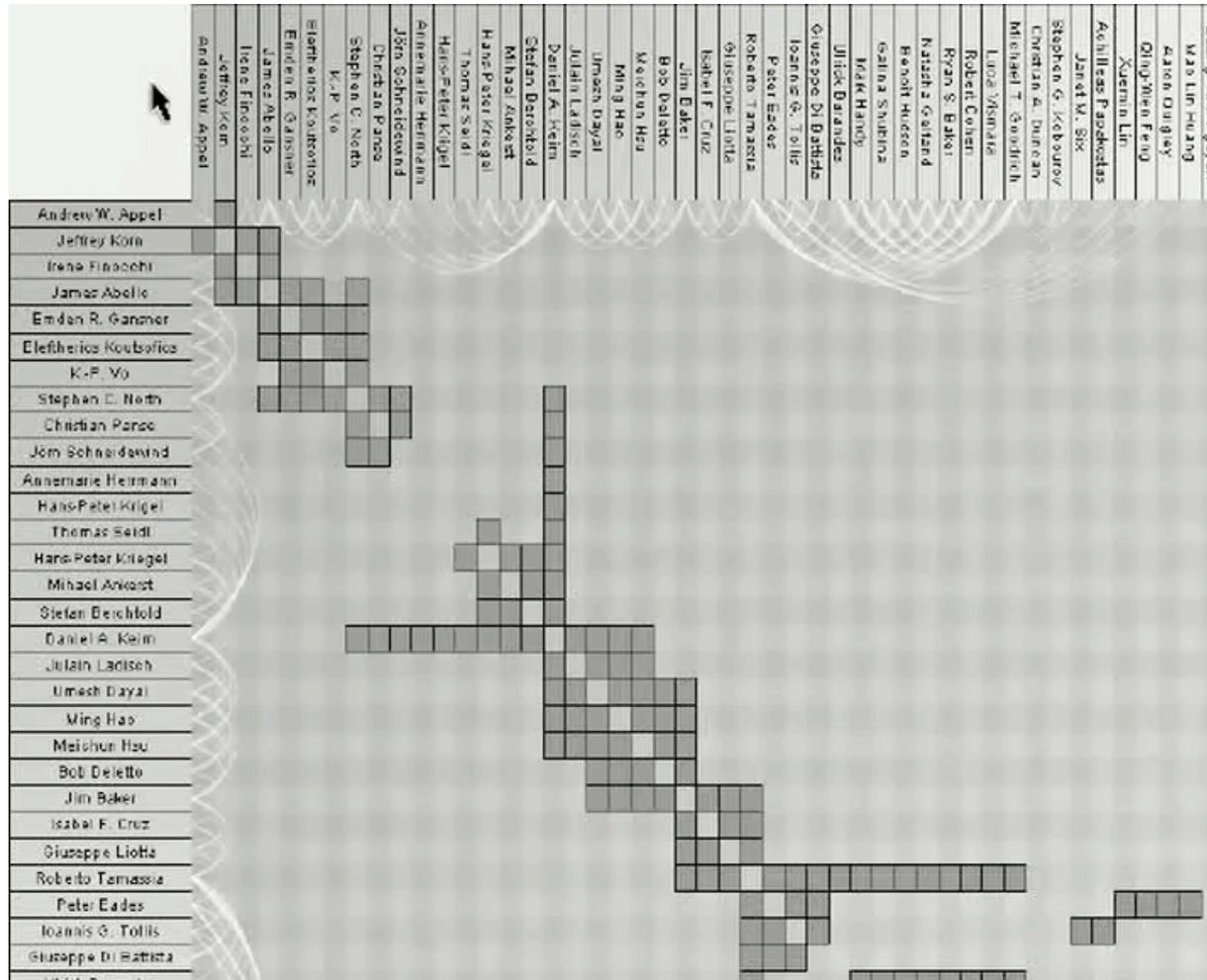
- The problem of path-related tasks [Ghoniem et al., 2005]
 - Always possible on matrices...
 - ... but tedious !



Outline



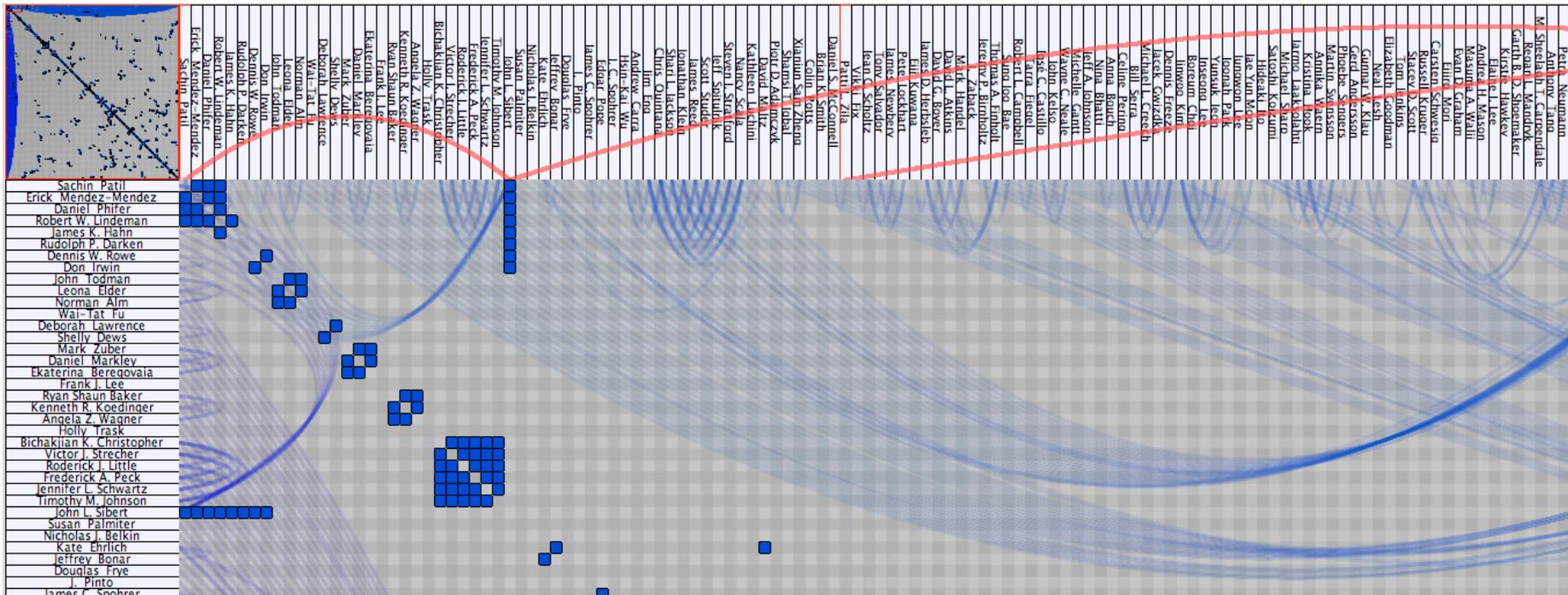
3. Augmenting matrices with MatLink



MatLink

[Henry and Fekete, Interact'07]
Best paper award

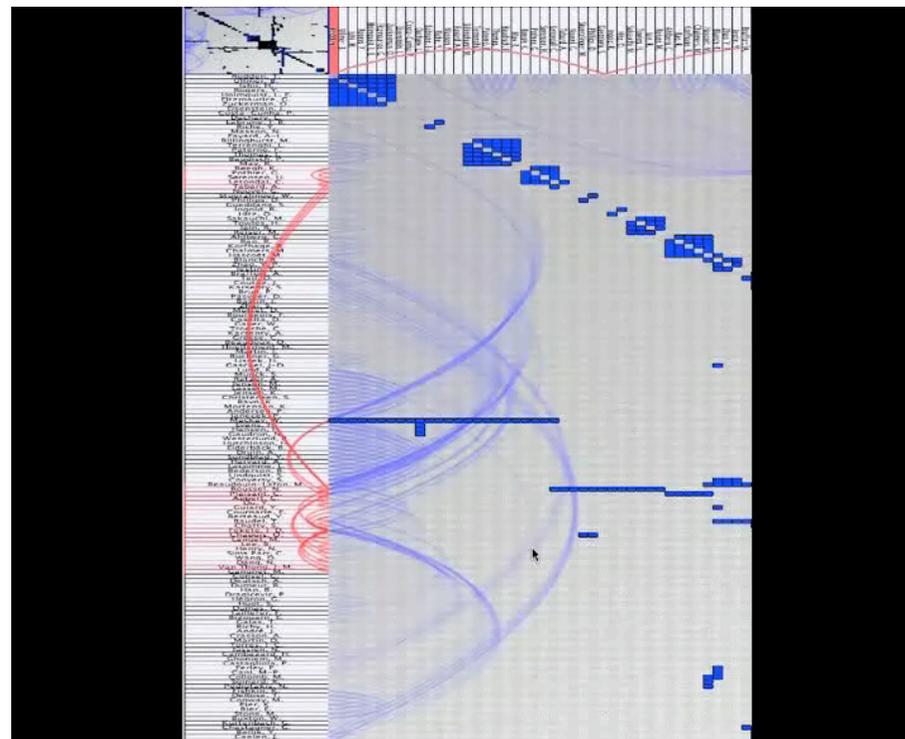
- Adding static links
- Adding interactive links
- For larger matrices...



Mélange

[Elmqvist, Henry, Riche and Fekete, CHI'08]

- Folding the space between 2 or more points of focus
- Works well for matrices and social networks tasks



Exploring with matrix or node-link?

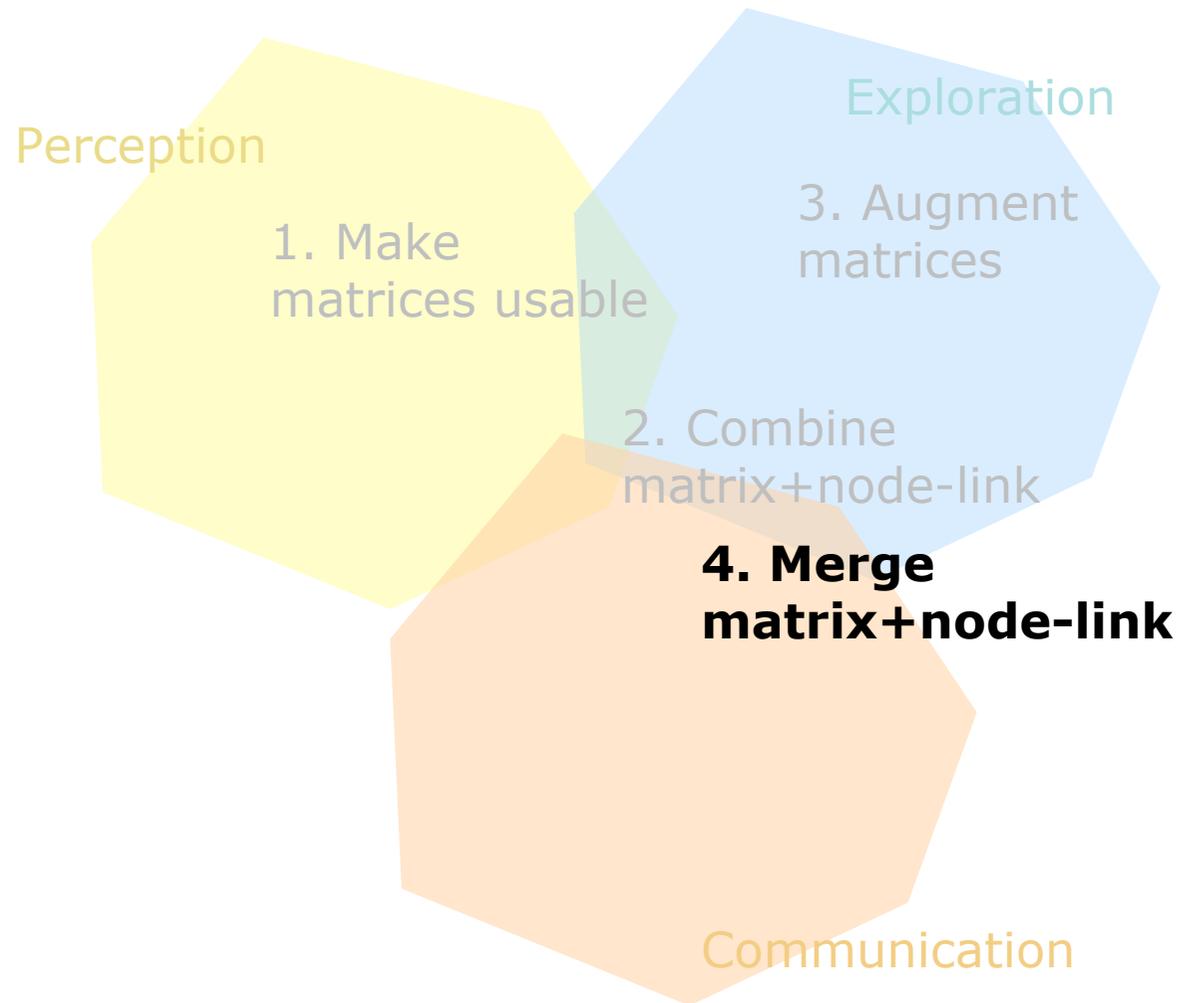
- Social networks

- Sparse (genealogical tree) \Rightarrow node-link
- Large/Dense (a year of emails) \Rightarrow matrix

- Small-world networks:

- A common category of social networks
- Globally sparse but locally dense
 - \Rightarrow node-link? \Rightarrow matrix?

Outline

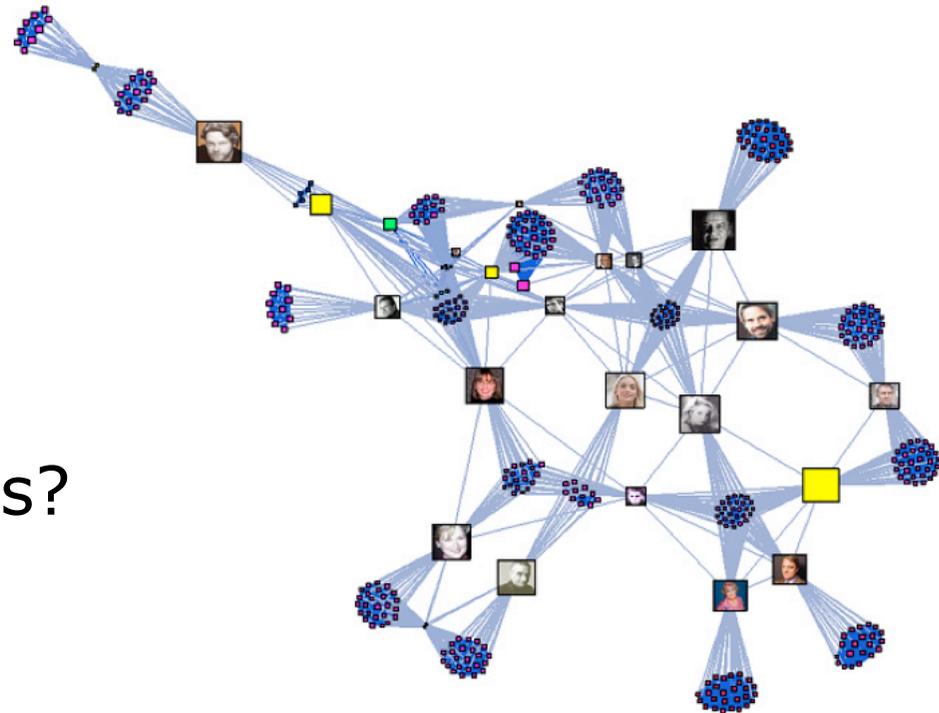


Small-world

○ The best representation?

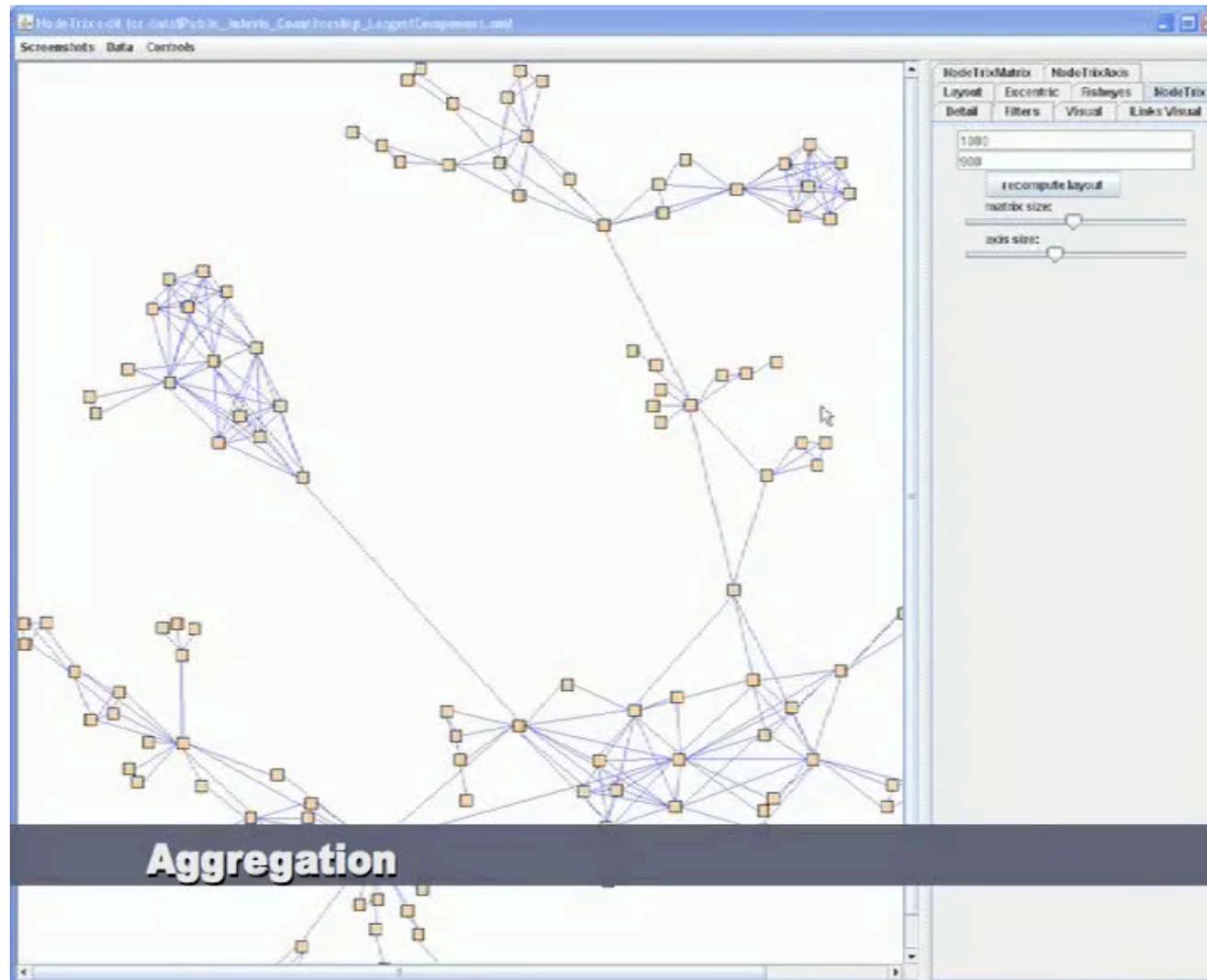
○ What is happening

- Inside communities?
- Between communities?

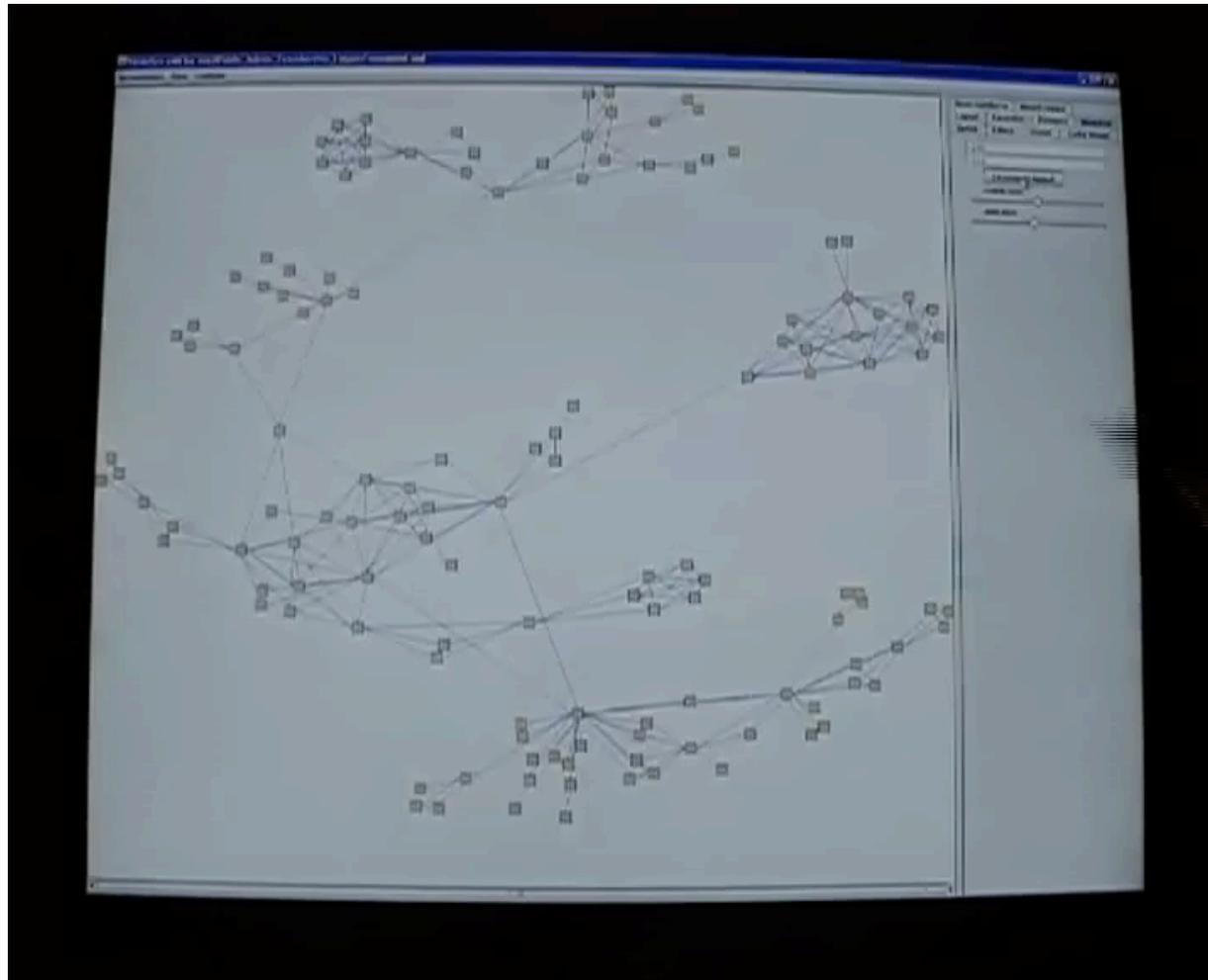


[Auber et al., InfoVis'04]

4. Merging matrix and node-link with NodeTrix

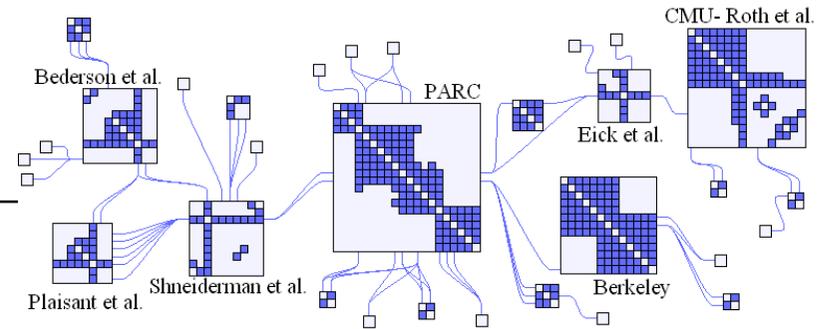


Interactive pen tablet

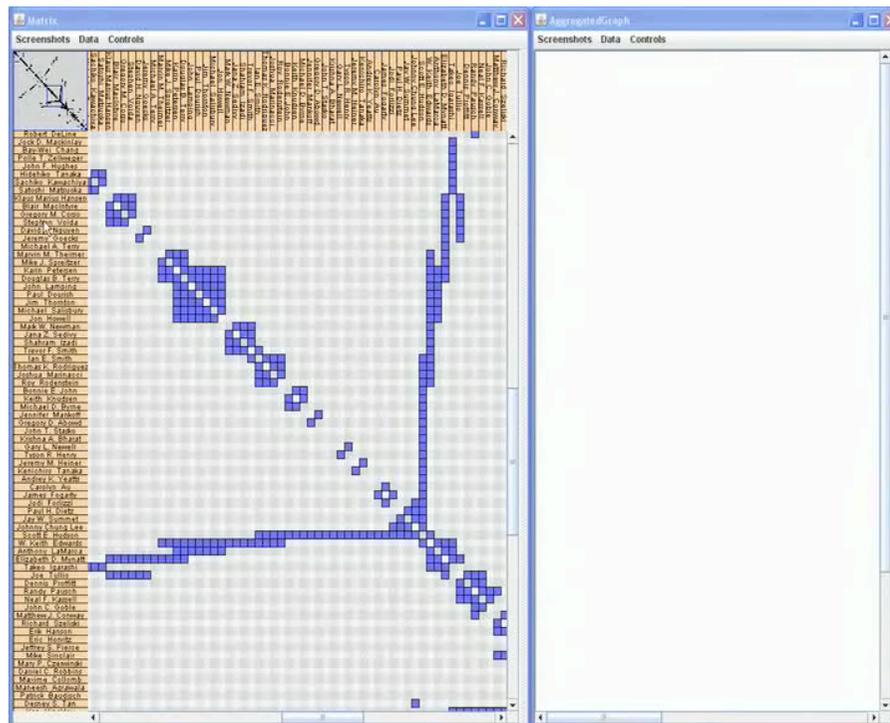


NodeTrix

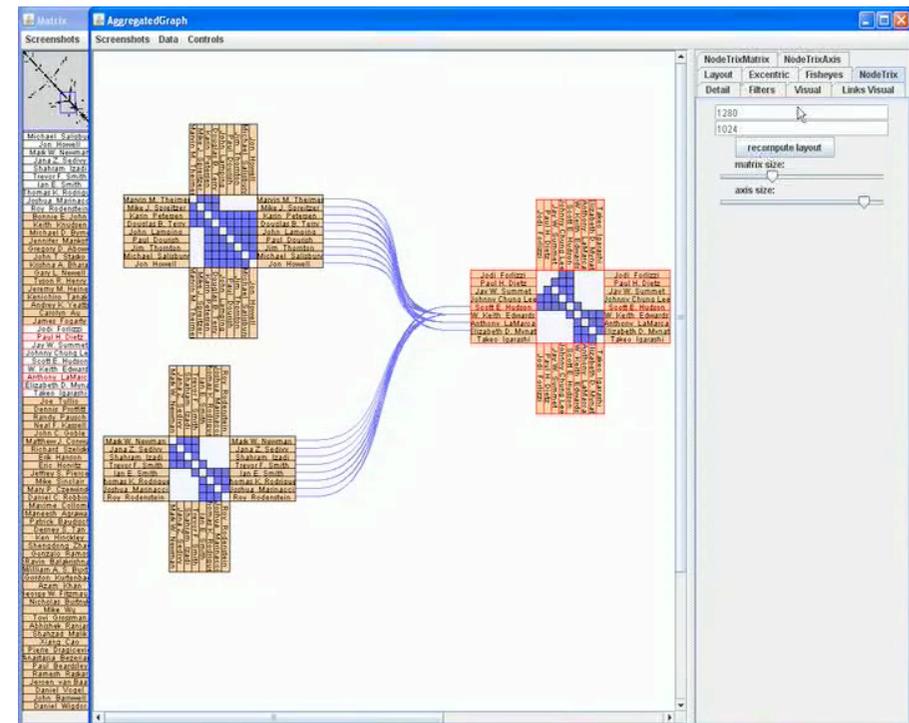
[Henry et al., InfoVis'07]



○ Explore



○ Communicate



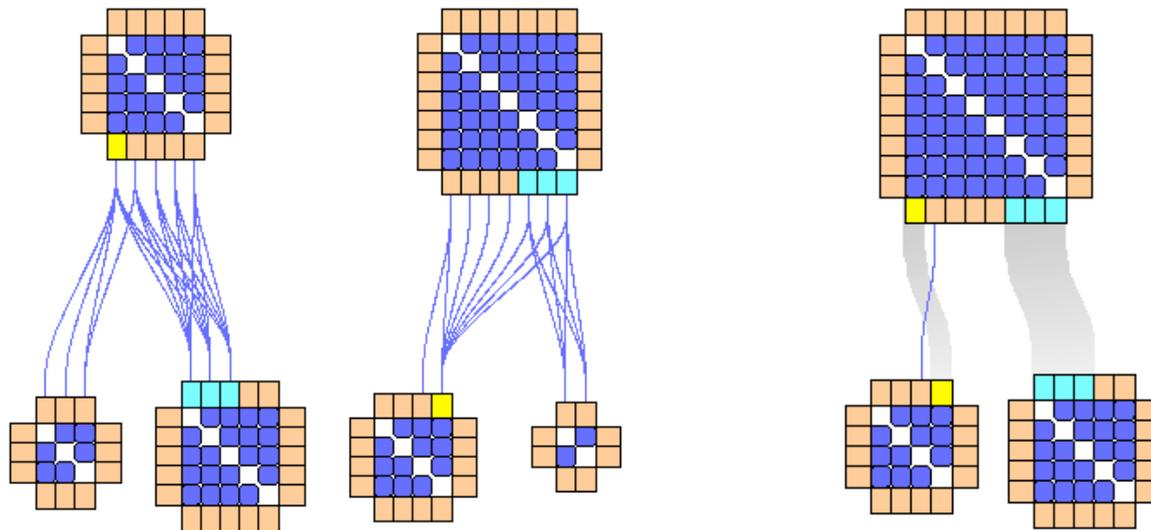
Actors in one or more communities

- Where to place them?
 - Between communities?
 - In one or the other community?
 - In overlapping communities?
 - In a higher level community?
- Why not duplicating them?
- What are the effect of duplication on user understanding?



Node Duplication

- More accurate community view
- Minimizing the misleading effect by visualizing links between duplicates



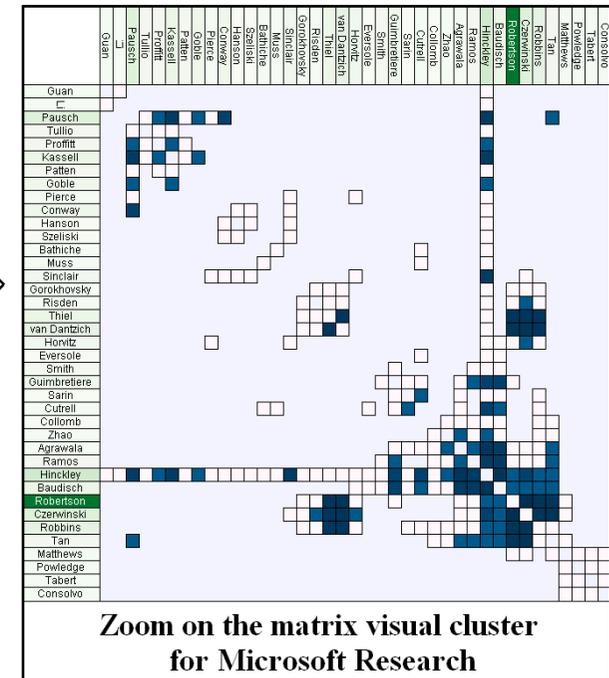
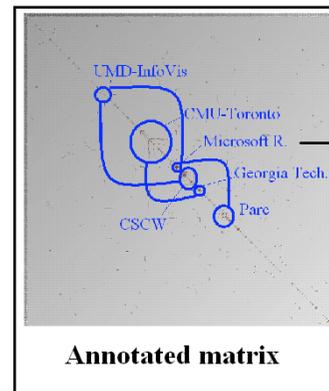
What about evaluation?

- Participatory design
 - Mostly informal and during the whole process
- Controlled experiment
 - For specific tasks (low-level) on specific datasets with specific representations
- Case Study
 - More realistic settings
- Longitudinal Study... coming

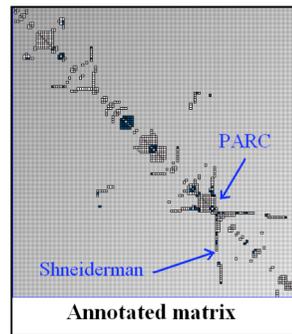
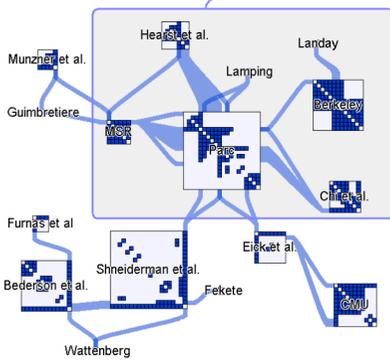
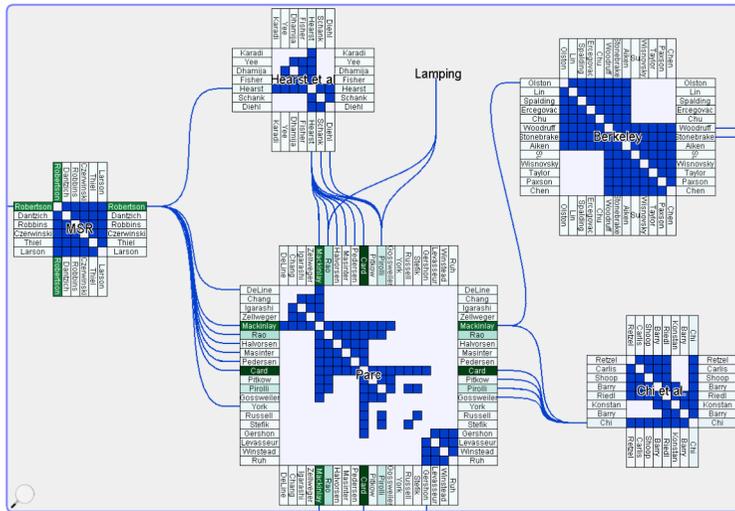
Case Study on 20 years of 4 HCI conferences

[Henry et al., IJHCI 07]

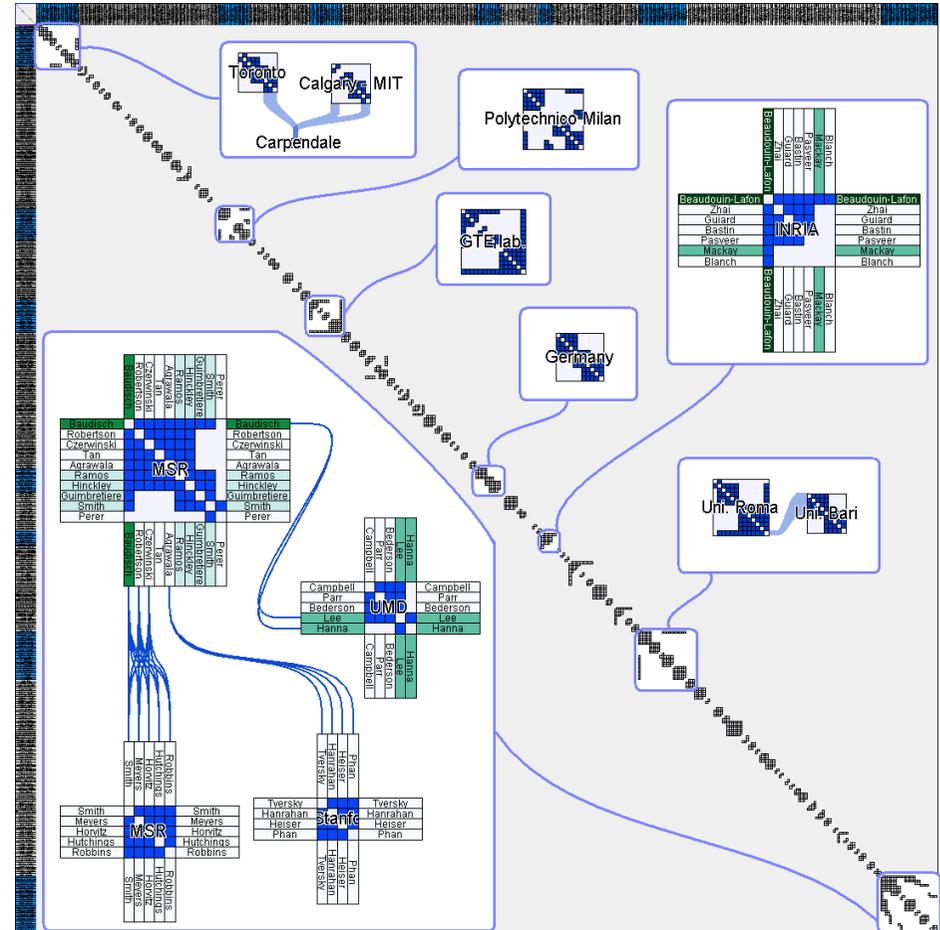
- Over 5 months
- Exploratory analysis of
 - 332 conferences
 - 27 000 actors
 - 118 000 relations
- Using MatrixExplorer
 - Exploring large networks
 - Presenting the results
- Then MatLink and NodeTrix



Communicating information

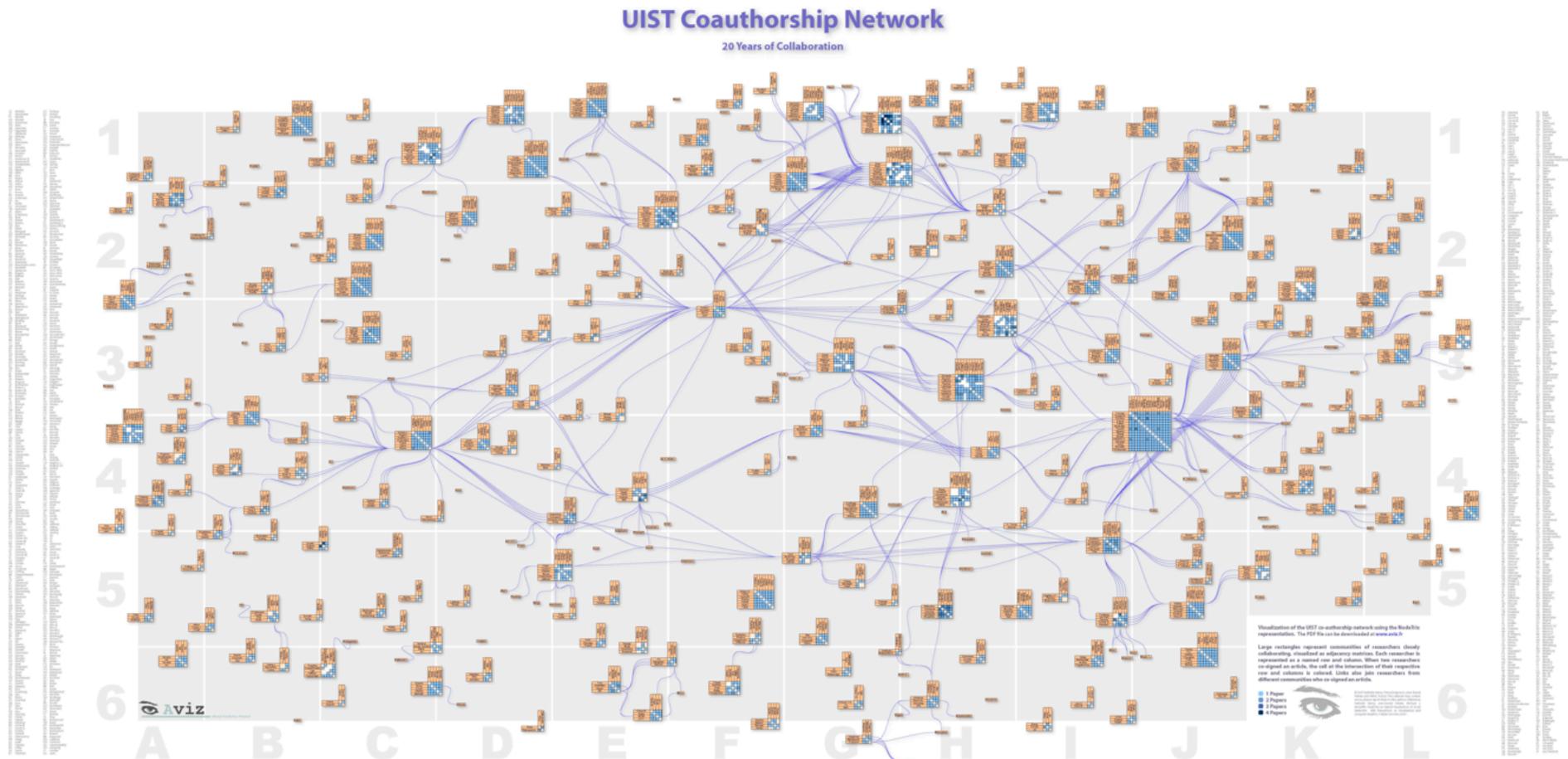


InfoVis coauthorship



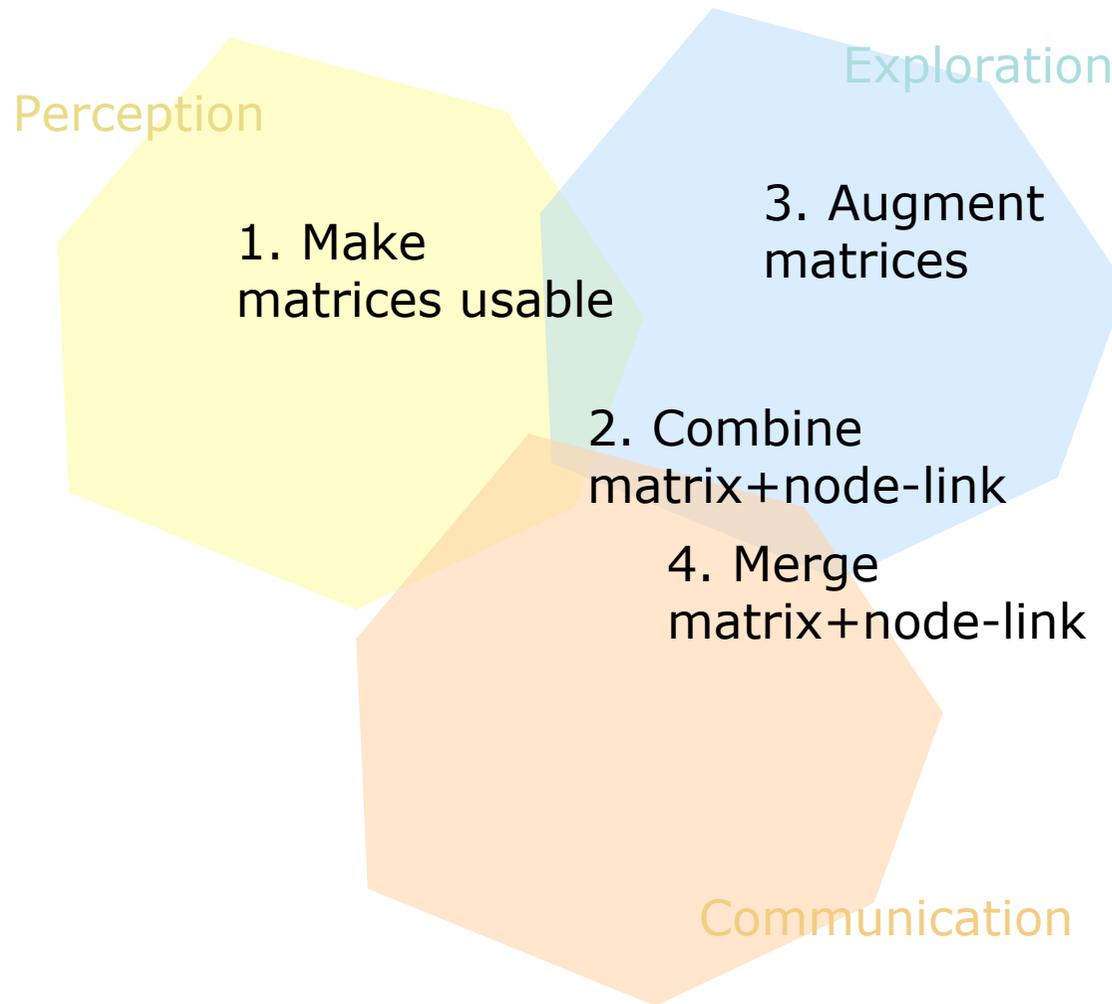
AVI coauthorship

Communicating information (2)



UIST coauthorship poster

What I have done



Research directions

○ Time-related data

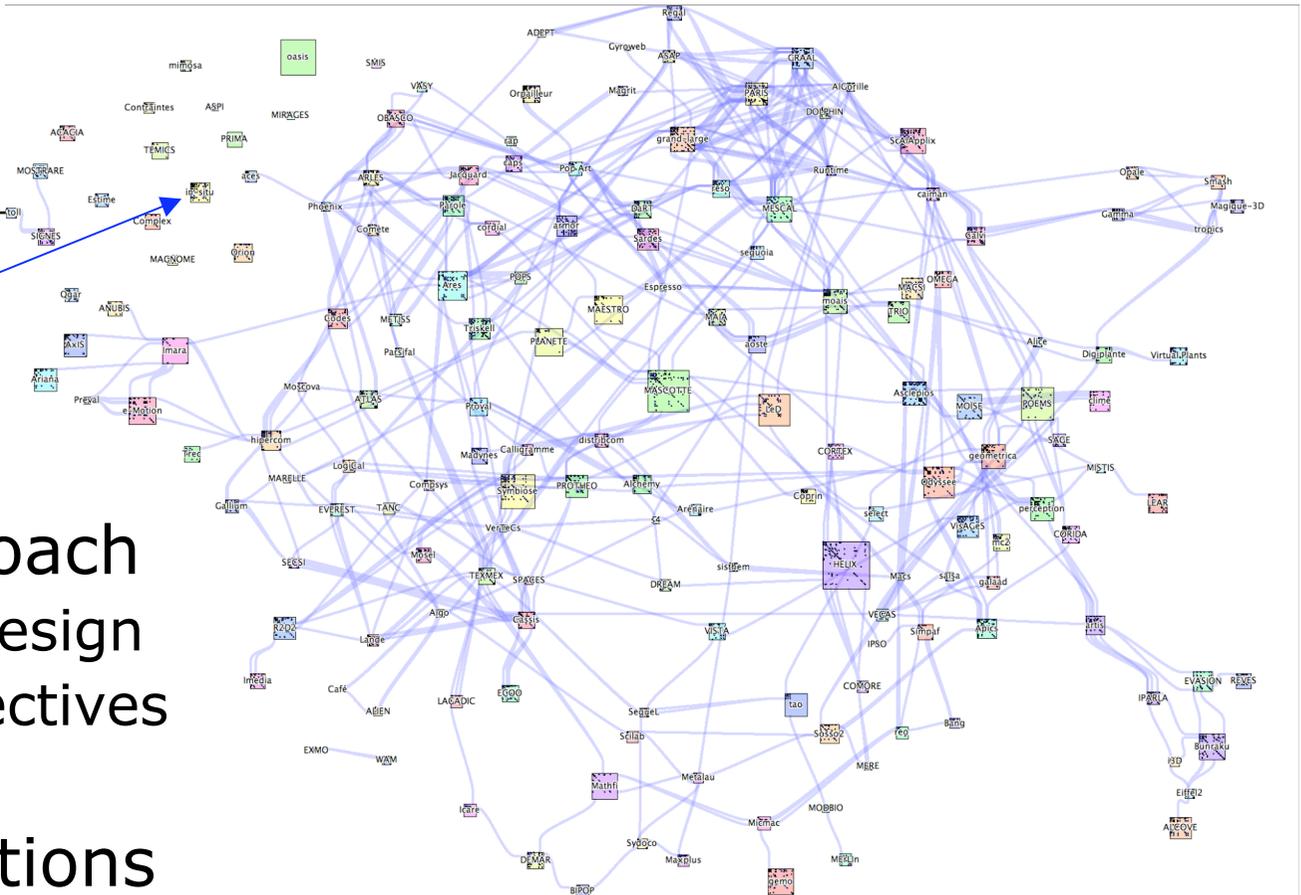
- On social networks
 - Scientific collaboration (INRIA)
 - Group awareness (Wikipedia France)
- On other data
 - Biological networks (Institut Pasteur)
 - Reflecting on user activity (MSR)

○ Longitudinal and qualitative studies

- Logging data
- New visualizations for
 - analyzing the data collected
 - reflecting it to users: for exploration and communication

La FIN!

I'm here!



INRIA in 2006

156 teams,
8400 persons

QUESTIONS?

- Research approach
 - Participatory design
 - Several perspectives
- Research directions
 - Time-related data
 - Evaluation