浙江大学 ZHEJIANG UNIVERSITY

2019 ZJU
International Summer School on Visual Analytics

Tree and Graph Visualization

Zhiguang Zhou

Associate Professor
Zhejiang University of Finance and Economics
Email: zhgzhou1983@163.com

Outline

① Hierarchical Data Visualization

- Node-Link Diagram
- Space-Filling Methods
- Hybrids
- Case: Genealogy Data Visualization

② Network Data Visualization

- Node-Link Diagram
- Adjacency Matrix
- Hybrids
- Case: OD Network Visualizaiton

Hierarchical Data Visualization

1 Hierarchical Data Visualization

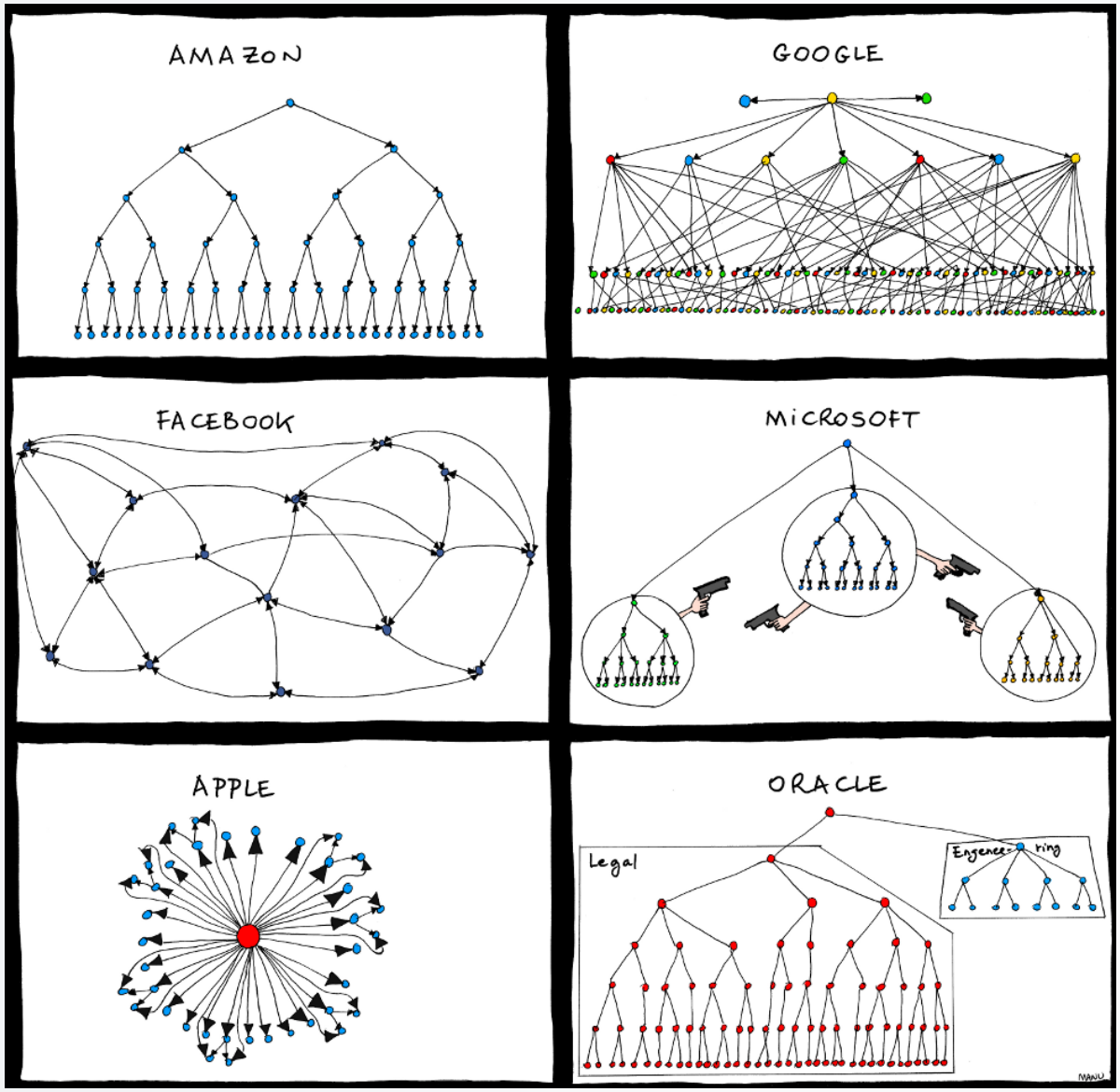
OUTLINE

- Hierarchical Data
- Hierarchical Data Visualization
  - Node-Link Diagram
  - Space-Filling Methods
  - Hybrids
- Case: Genealogy Data Visualization

Hierarchical Data

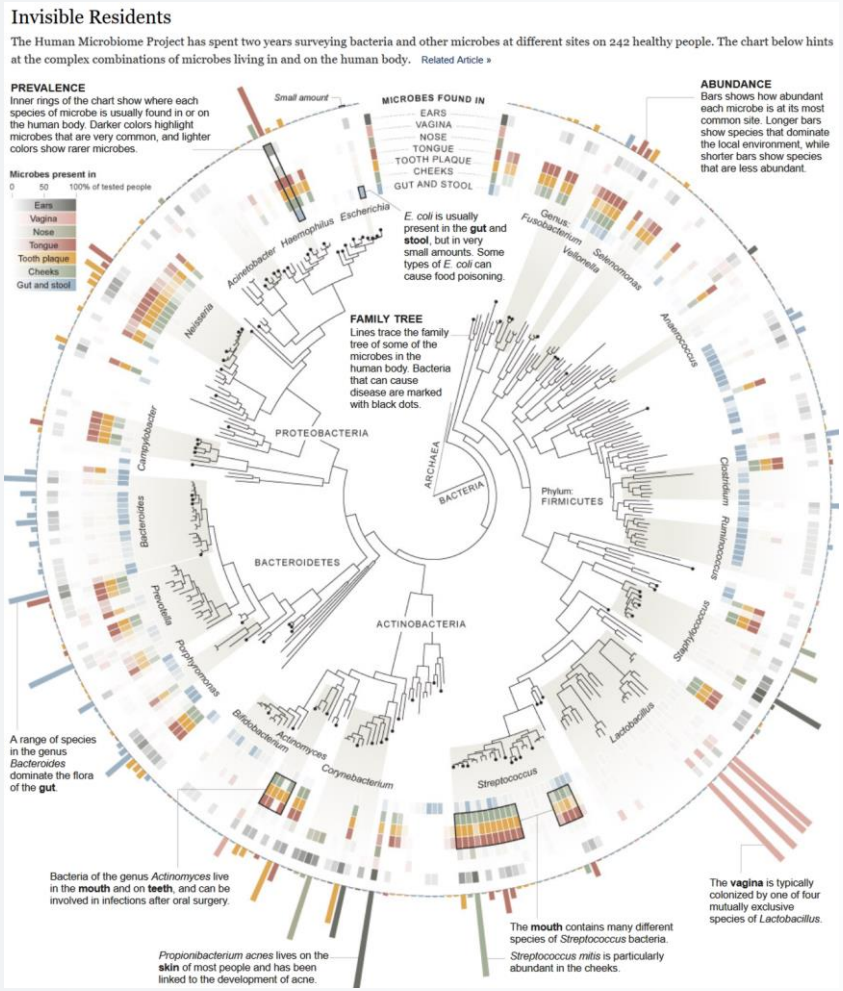
Hierarchical Data

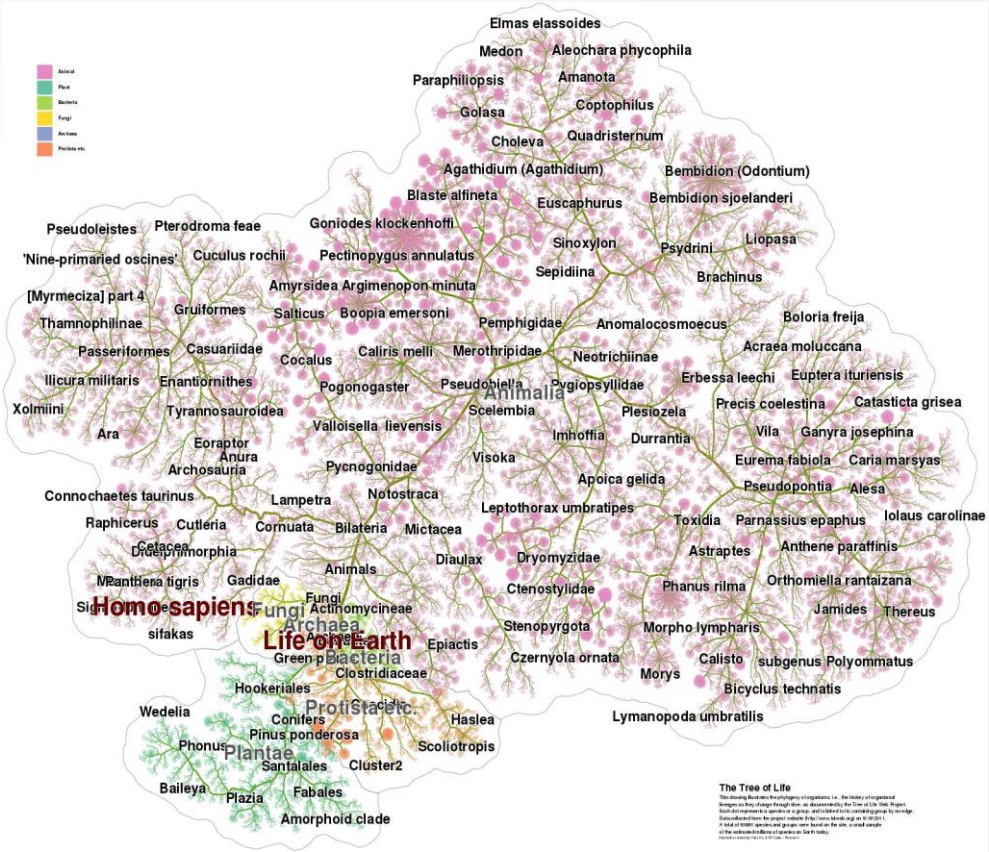
- It focuses on the hierarchical relationship between objects.
  - Social affiliations.
  - Information organization.
    - File directories.
    - Development of species.
  - Logical connections.
    - Decision tree.



Company Organization

http://www.bonkersworld.net/2011/06/27/organizational-charts/





http://www.nytimes.com/interactive/2012/06/19/science/0619-microbiome.html?_r=1&

数据来源：http://tolweb.org/tree/

图片来源：http://www2.research.att.com/~yifanhu/TOL/

Hierarchical Data Visualization

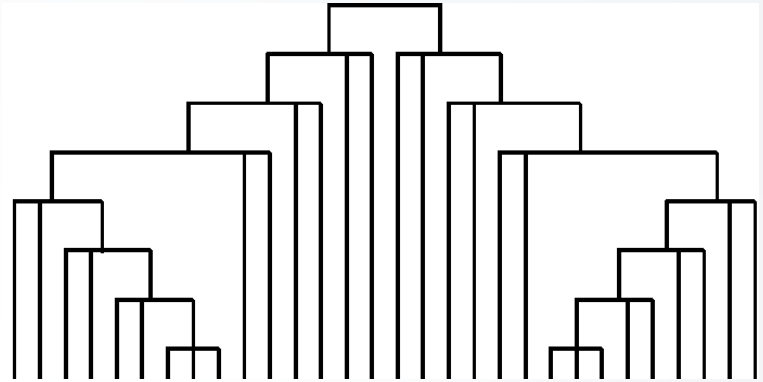
Hierarchical Data Visualization

- Node-link diagram (structure-clarity).
  - Node-link tree.
  - Hyperbolic tree.
  - Cone tree.
- Space-filling methods (space-efficiency).
  - Treemap.
  - Voronoi treemap.
- Hybrids.

Node-Link Diagram

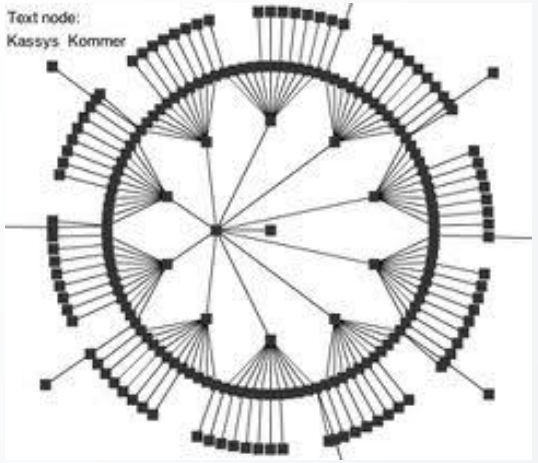
Layout of Nodes and Links

- Orthogonal layout.
  - Circuits diagram.
  - Indent diagram.
  - Dendrogram diagram.



Dendrogram

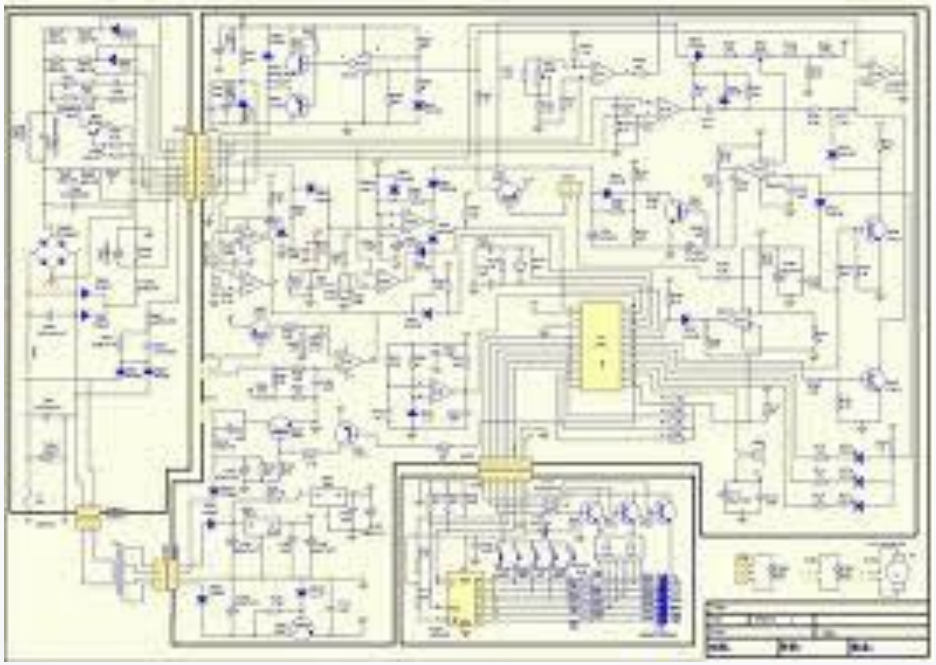
- Radial layout.
  - Radial layout.
  - Hyperbolic tree.



Radial layout

Circuits Diagram

- Orthogonal and space efficient.
- Computer friendly, but not user friendly.

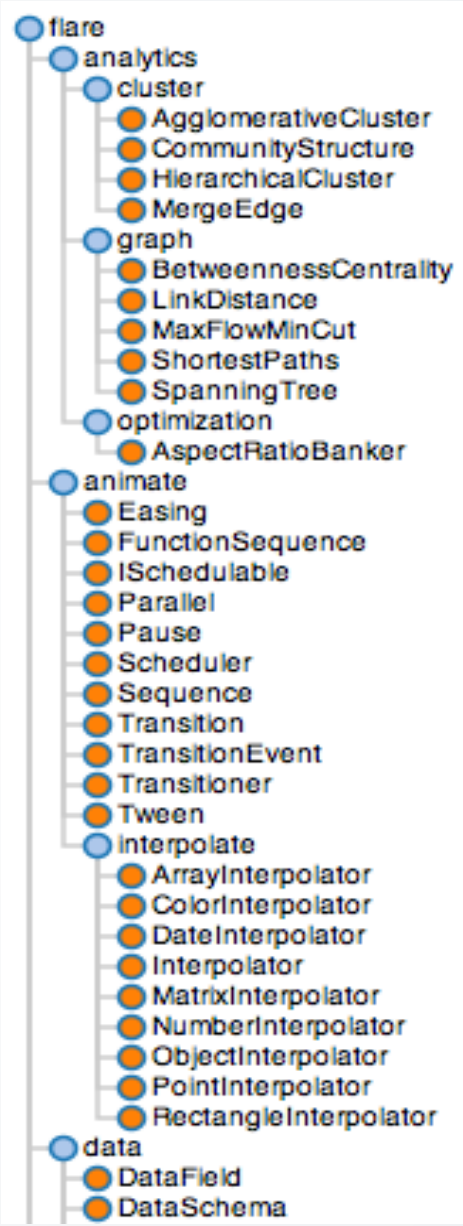


Circuits diagram of induction cooker

Indent Diagram

- Easy to implement.
- Can be applied to plain text or HTML.
- Much scrolling is needed in big data browsing.
- May lose the context.

```
function draw(node:Node, depth:int) {
    println(<depth spaces> + nodelabel);
    for each child c do
        draw(c,depth+1)
}
draw(root,0);
```



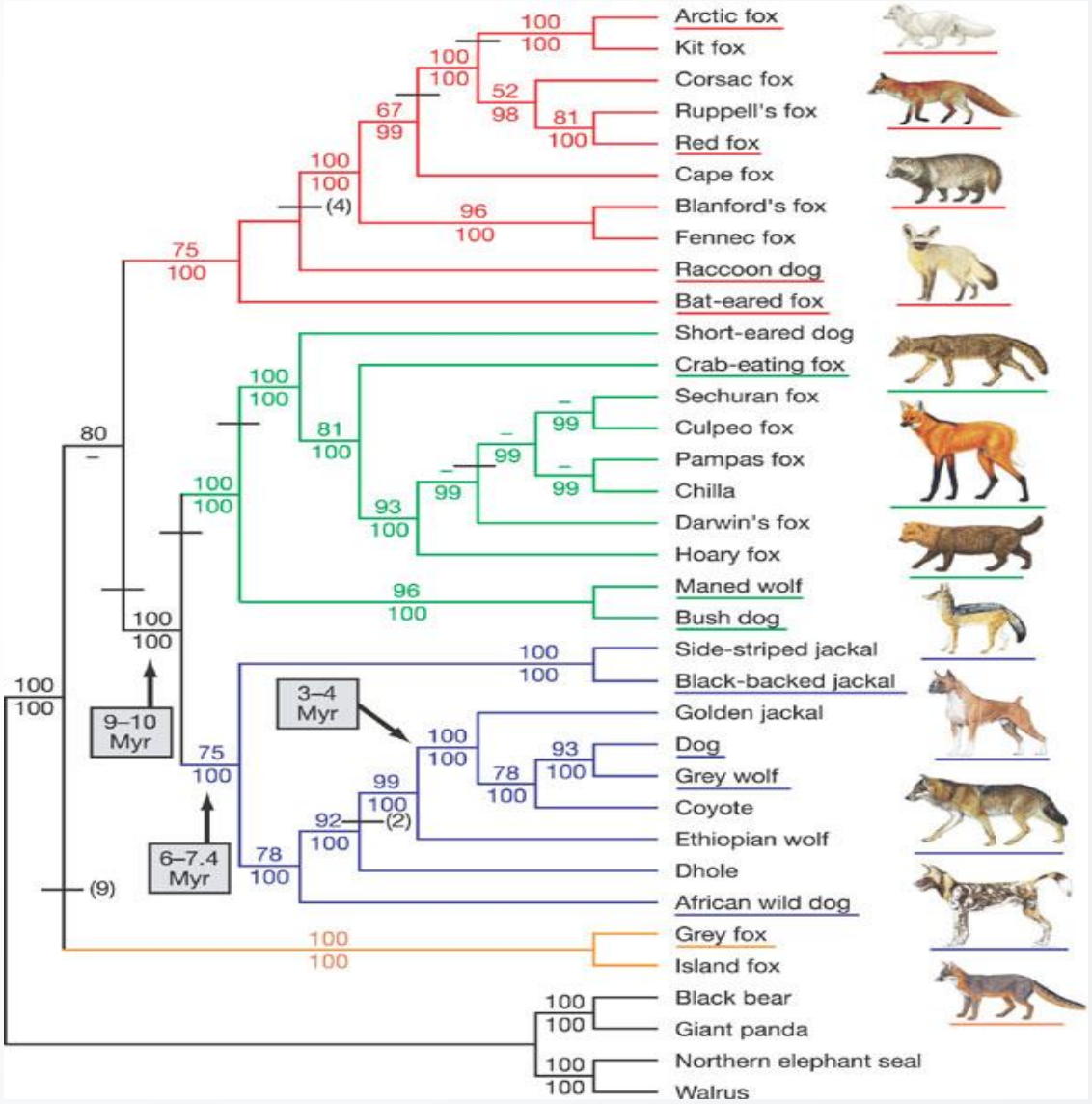
Flare软件包的子目录结构

http://hci.stanford.edu/jheer/files/zoo/

Dendrogram

Comparison of dog species' different haploid gene sequence

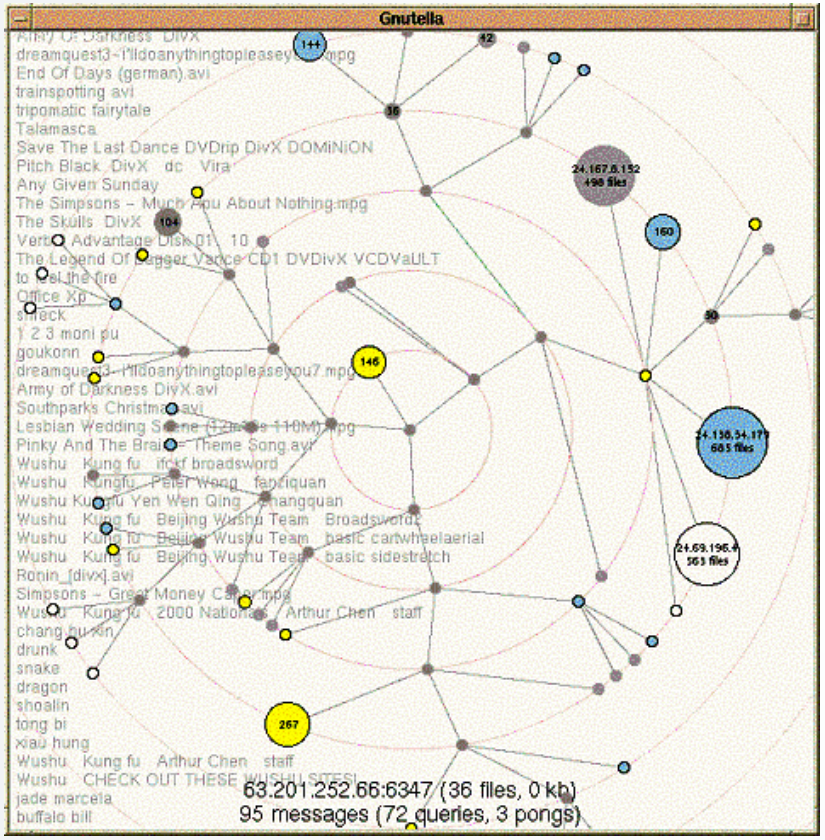
Kerstin Lindblad-Toh et al. *Nature* 438, 803-819 (8 December 2005)



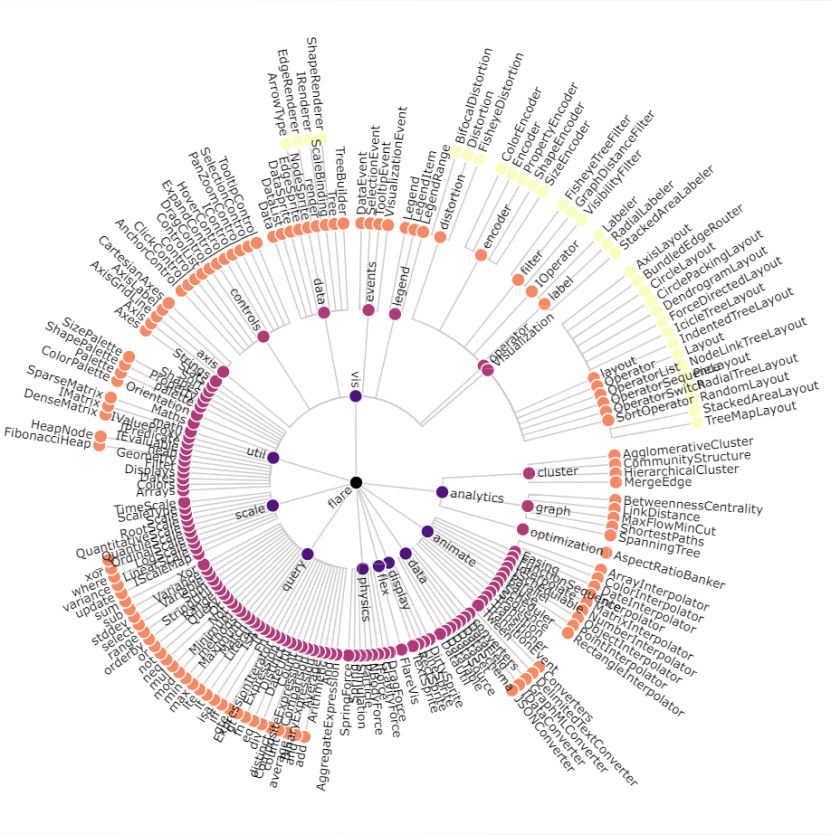
Radial Layout

- Better space utilization.
- Root is at the center of the circle, nodes of different hierarchy located on the circles of different radius.
- Distances to the center is equivalent to the depth of the nodes.
- The deeper, the farther are the nodes to the center, and the more space they will get.

Radial Layout



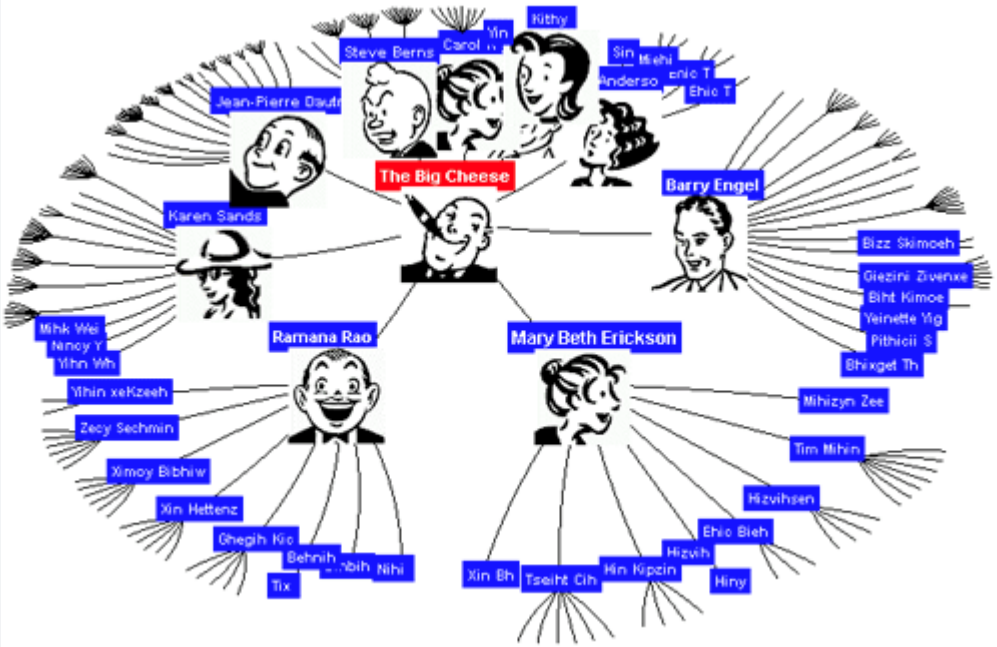
Radial Tree



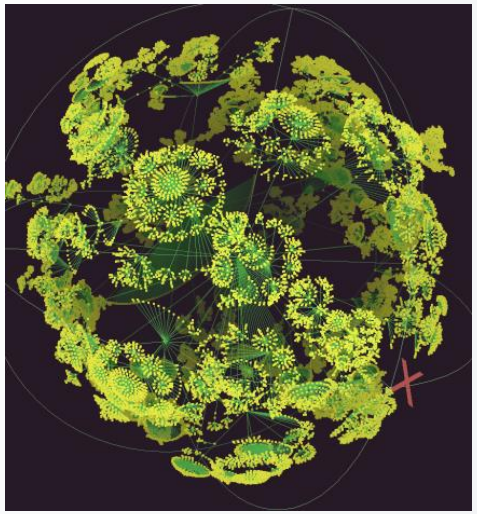
https://vega.github.io/vega/examples/radial-tree-layout/

Hyperbolic Tree

- Positioning the nodes in the hyperbolic space.



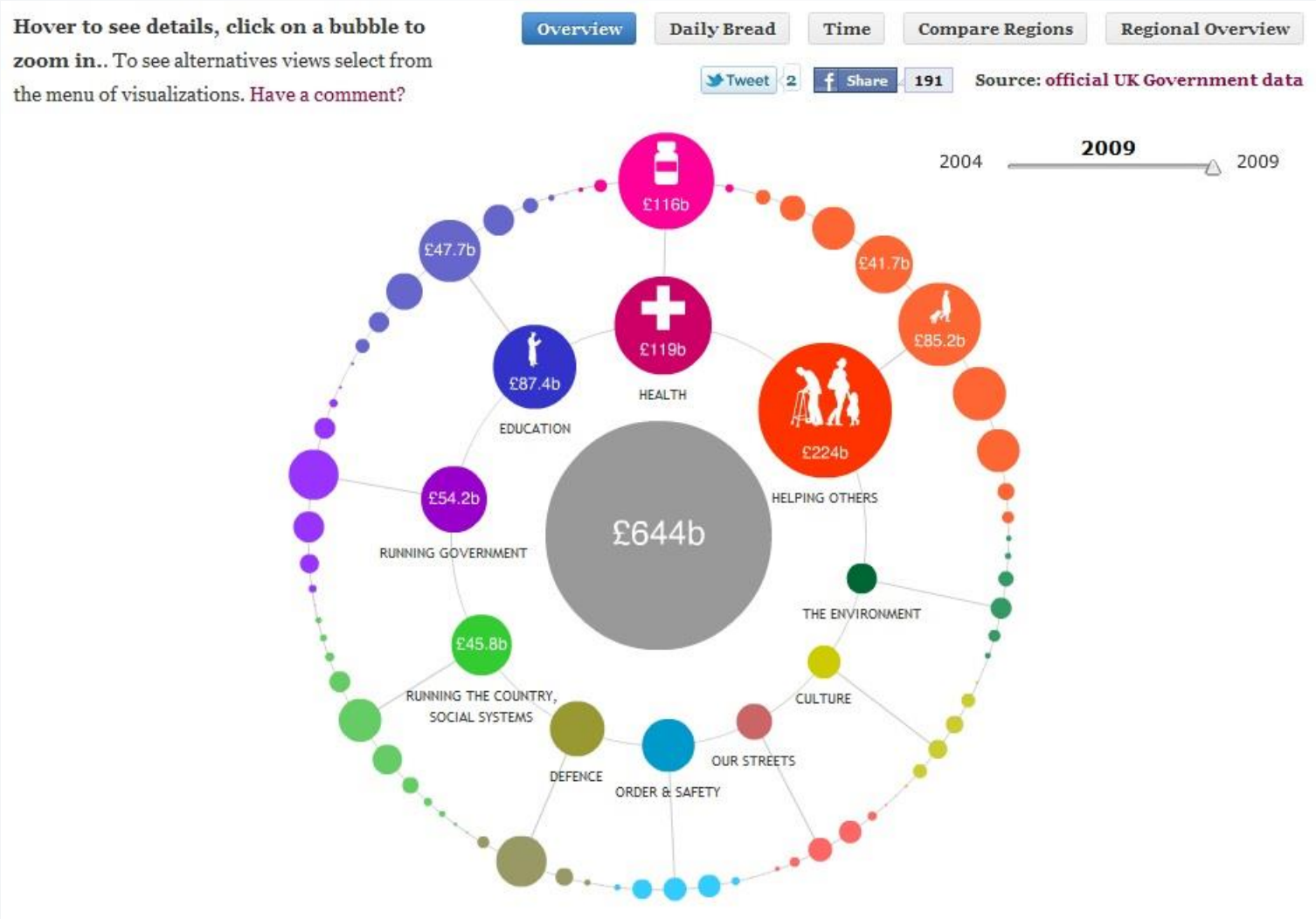
Hyperbolic tree of social relationship



Morse (hyperbolic space)

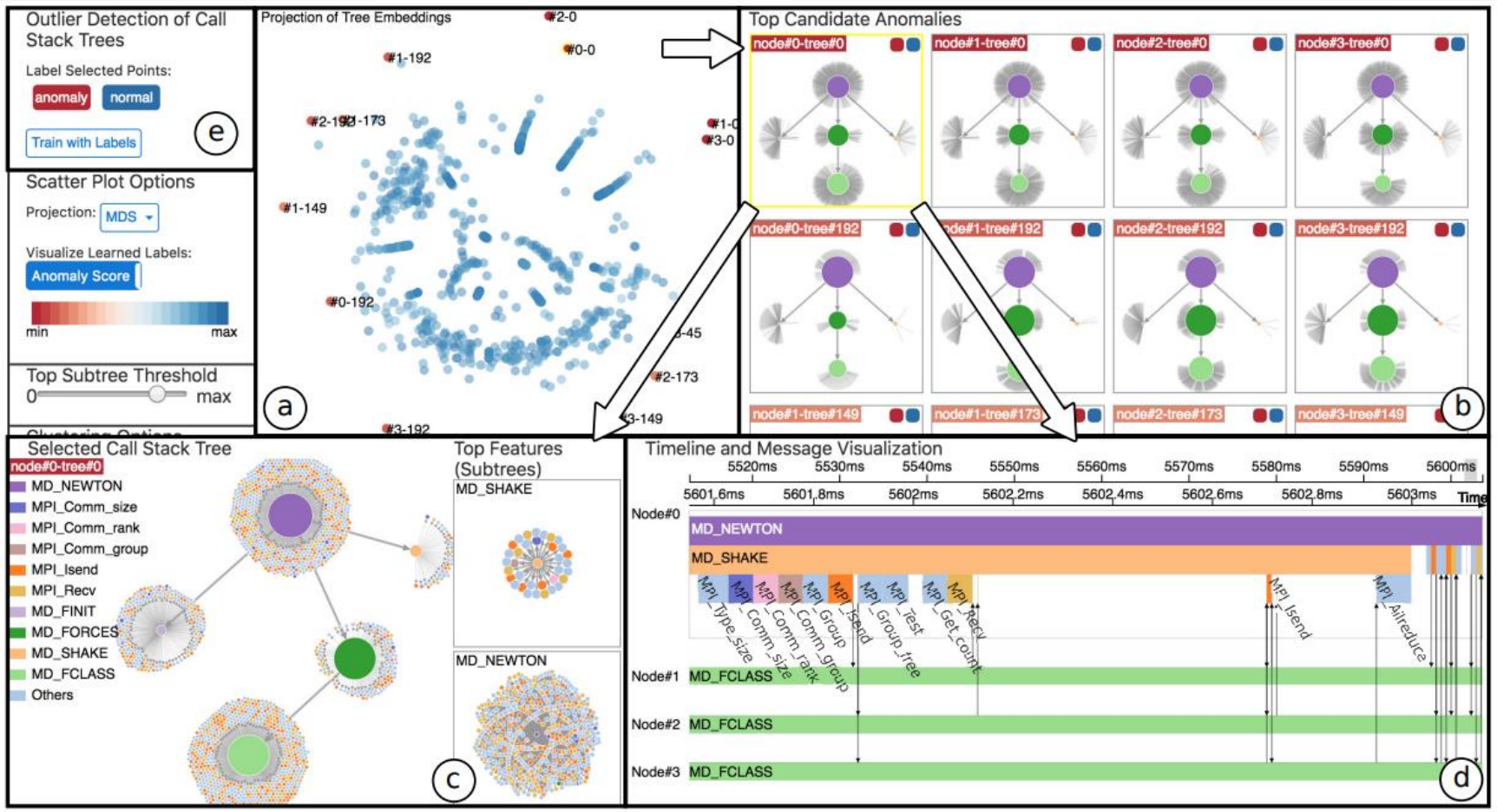
Applications of Node-Link Diagram

Where Does My Money Go



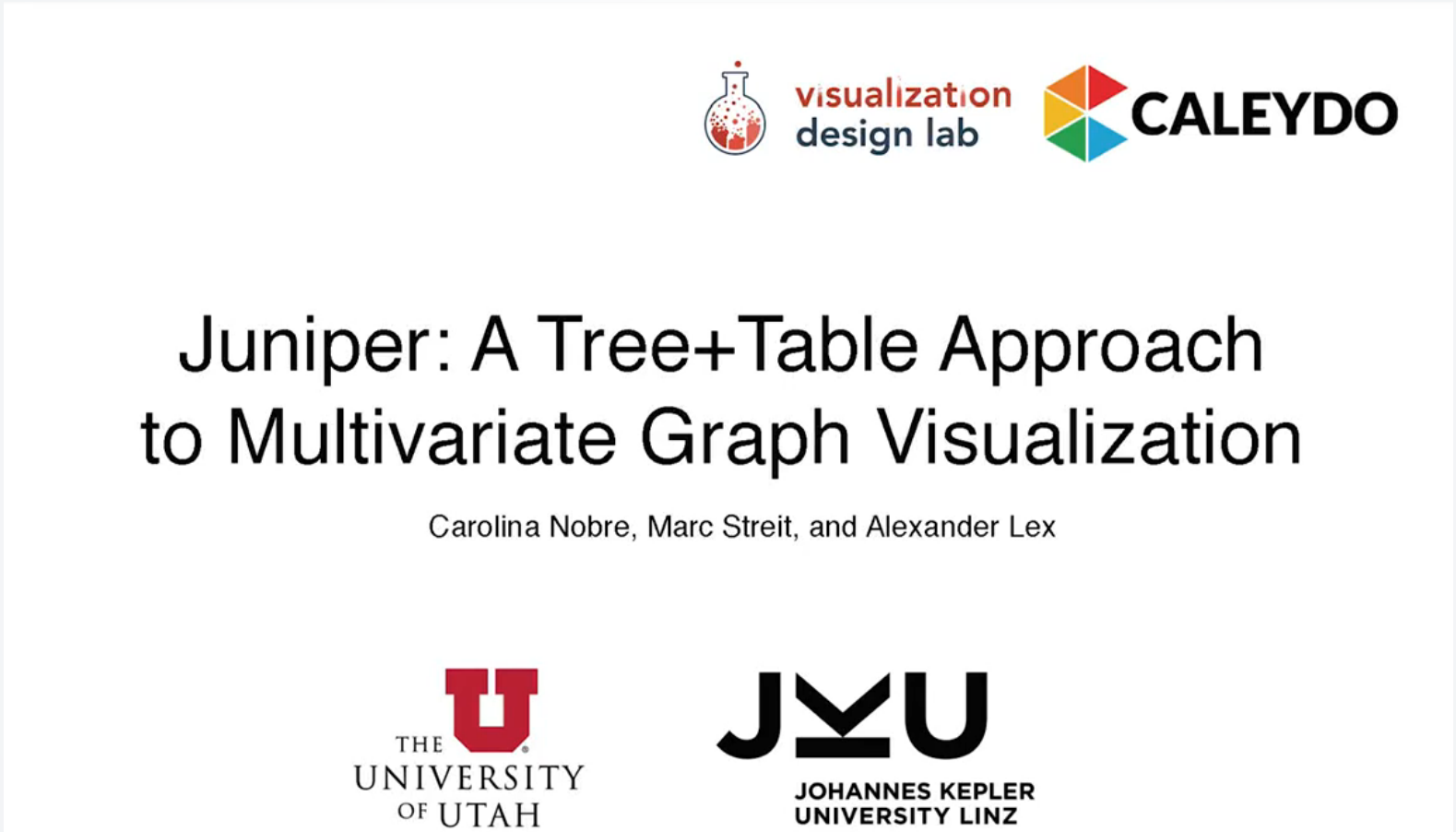
http://visual.ly/where-does-my-money-go

Call Stack Tree



Xie et al. A Visual Analytics Framework for the Detection of Anomalous Call Stack Trees in High Performance Computing Applications (IEEE VAST 2018).

Juniper

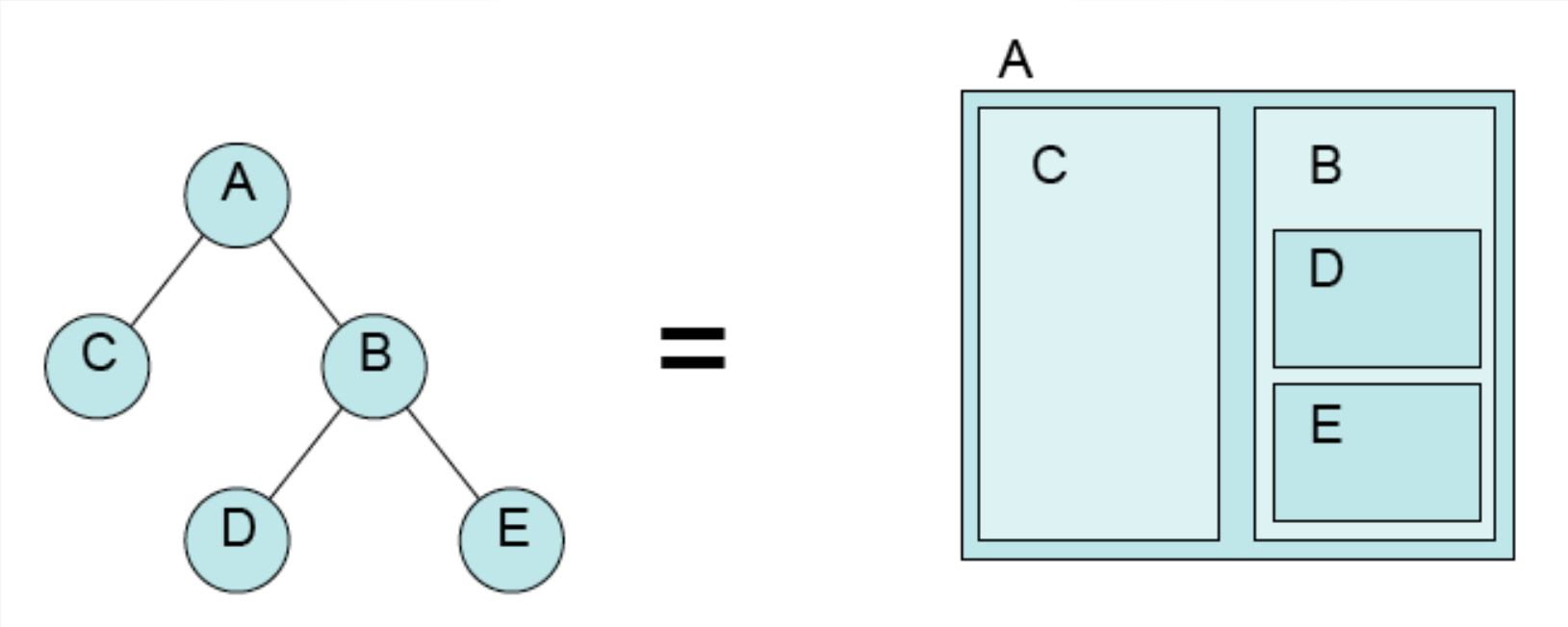


Nobre et al. Juniper: A Tree + Table Approach to Multivariate Graph Visualization (IEEE InfoVis 2018).

Space-Filling Methods

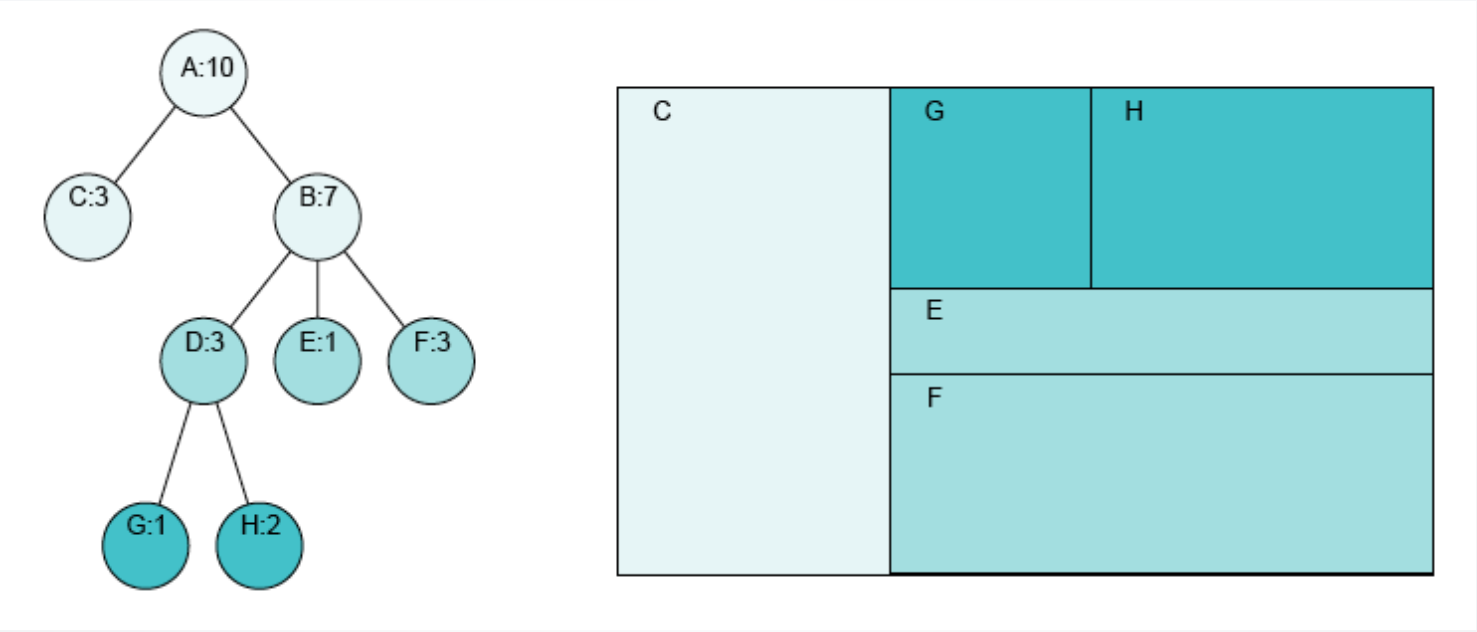
Enclosure

- Child nodes are nested in parent nodes.



Treemap

- Assume that each leaf node has a size-related attribute(size of the file…).
- Size of parent nodes is the sum of his children's.

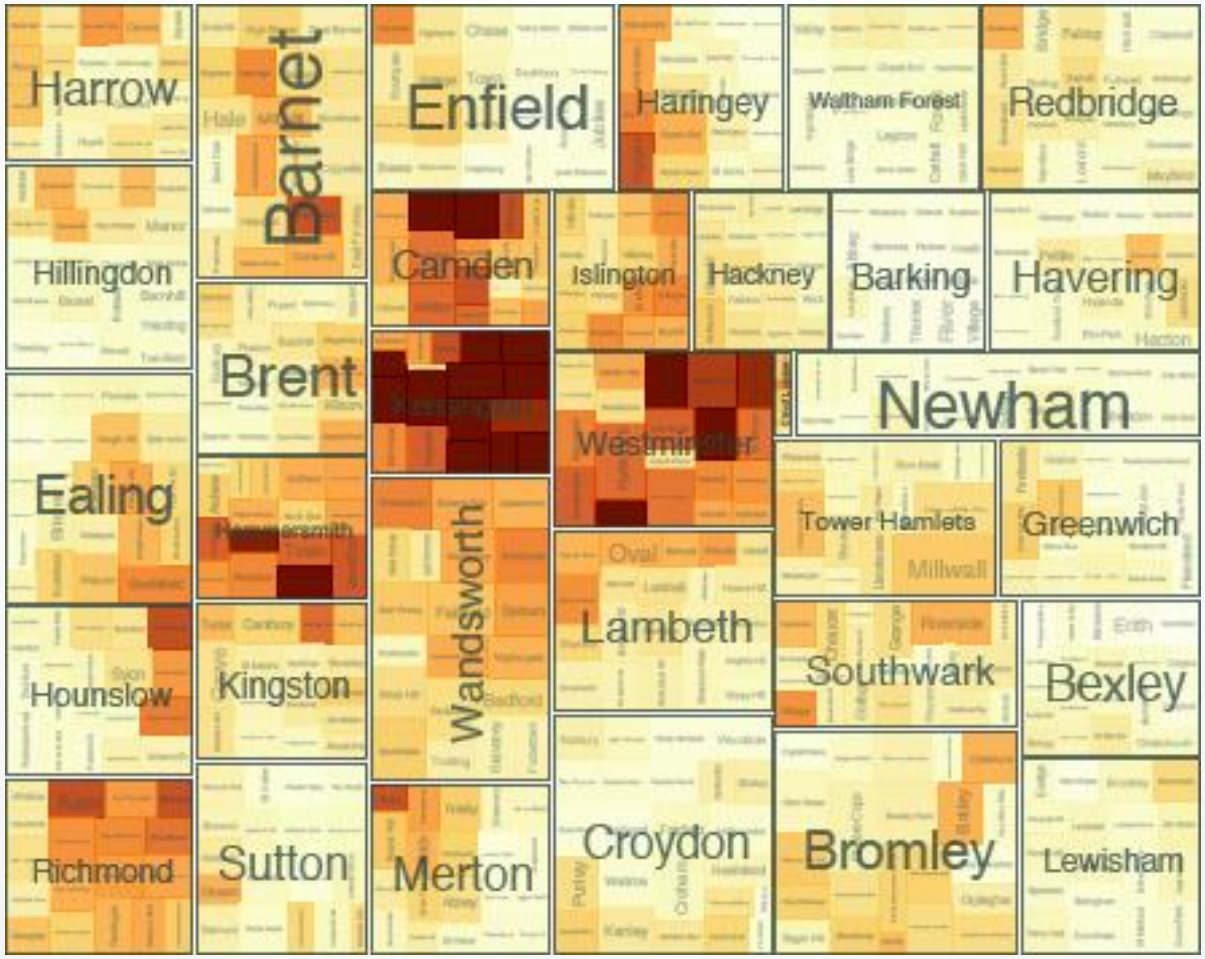


Treemap

- Johnson and Shneiderman raised the slice-and-dice Treemap in 1991.
- Layout method:
  - Recursive subdivision.
  - Root is the largest rectangle.
  - Size of child nodes are assigned according to their weights.
  - Subdivide child nodes.

Treemap of Categorical Data

- London personal housing transactions.

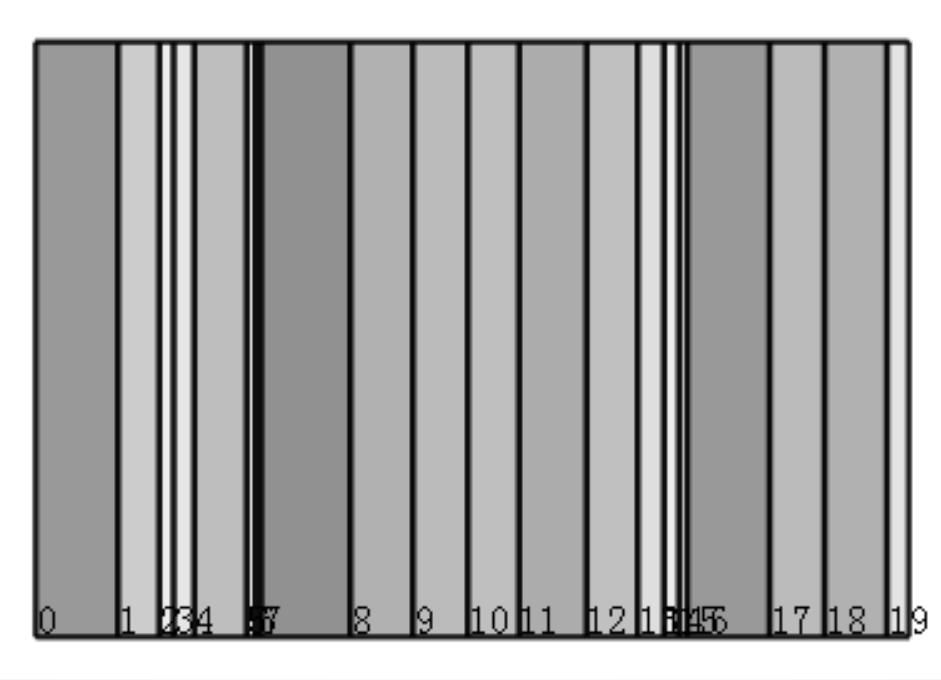


Hierarchy of Categorical Data

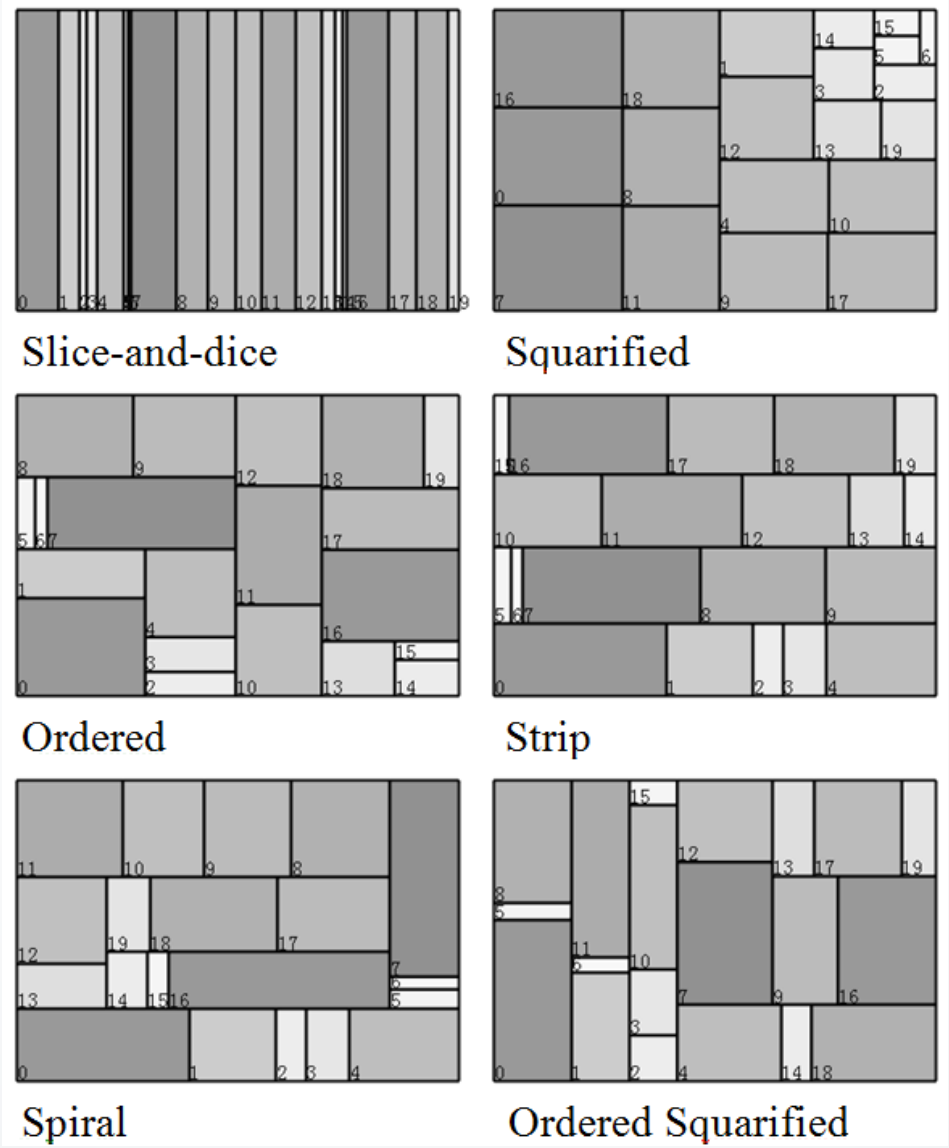
- No hierarchy among categories.
- Different hierarchy assignment results in huge difference, even misleading.
- Dominant categories should be parent nodes.
- Housing transaction data.
  - District: downtown, suburbs.
  - Type of the house: villa, apartment.
  - Neighborhood.

Problems of Treemap

- Simple slice-and-dice subdivision method results in thin and long rectangles.
- It would be difficult to interact with nodes.



Extensions of treemap



Treemap Layout Algorithms	Ideas	Aspect Ratios	stability	readability	Continuity	Balance of stability	Distance relavance
Slice-and-dice	Subdividing either horizontally or vertically alternatively	*	***	***	***	***	***
Squarified	Using greedy method to keep optimal aspect ratio	***	*	**	**	**	-
Ordered	Refining iteratively Balance of aspect ratio and readability	**	**	**	**	**	-
Strip	Dividing one strip every time along settled direction	**	**	**	**	**	-
Spiral	Dividing from outside to inside spirally	**	**	**	***	***	-
Ordered Squarified	Estimating the distance to its parent, and locating according to squarified	**	**	*	*	-	***
Spatially Ordered	First roughly locating positions, then using Squarified	**	**	-	-	-	***

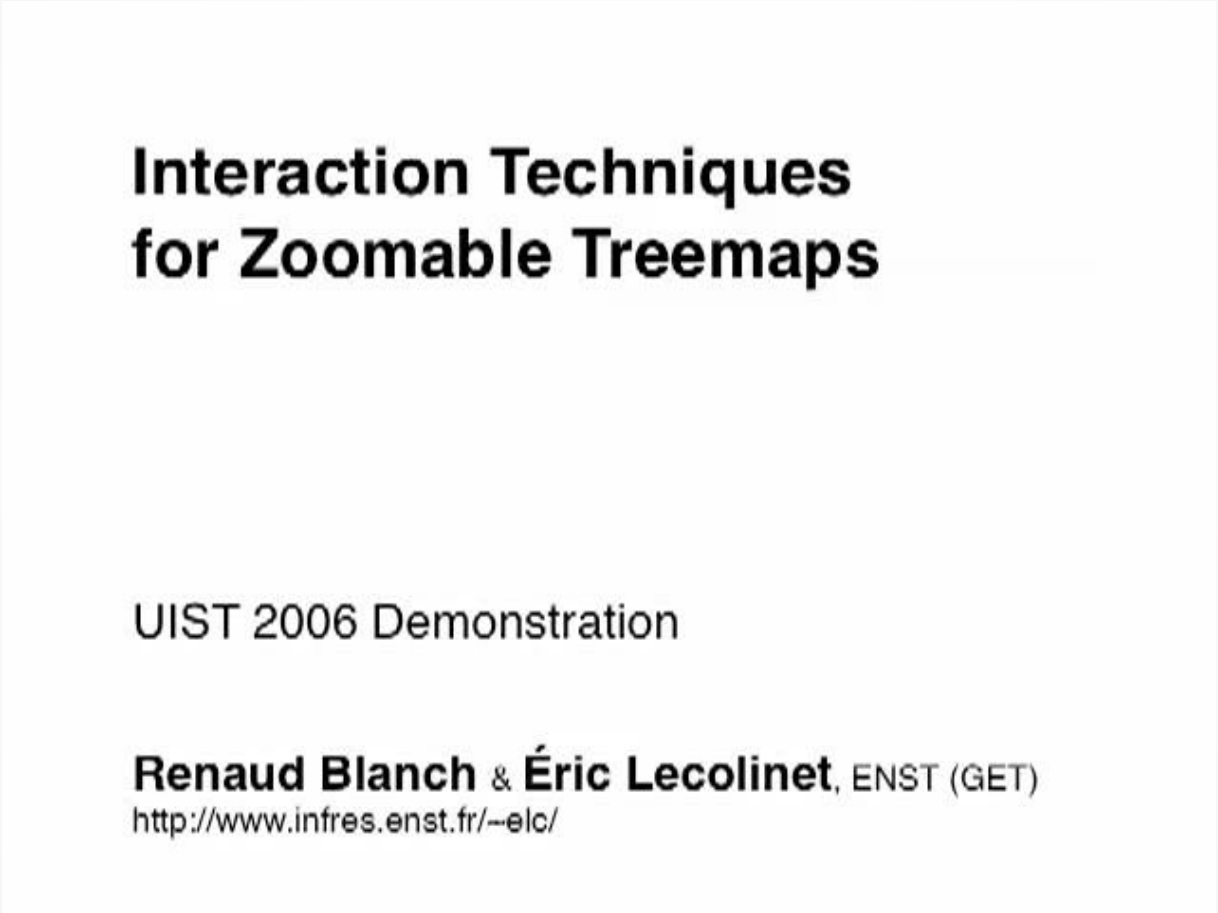
张昕,袁晓如.树图可视化方法及其扩展

Interaction of Treemap

categories	Interactions
Basic interactions	Node selection
	Change of weights
	Assignment and change of color
	Drill down
	Roll up
	Switch of layout algorithm
Novel interactions	MagicLens
	Fisheye
	Semantic Zooming
	Balloon Focus

张昕,袁晓如.树图可视化方法及其扩展

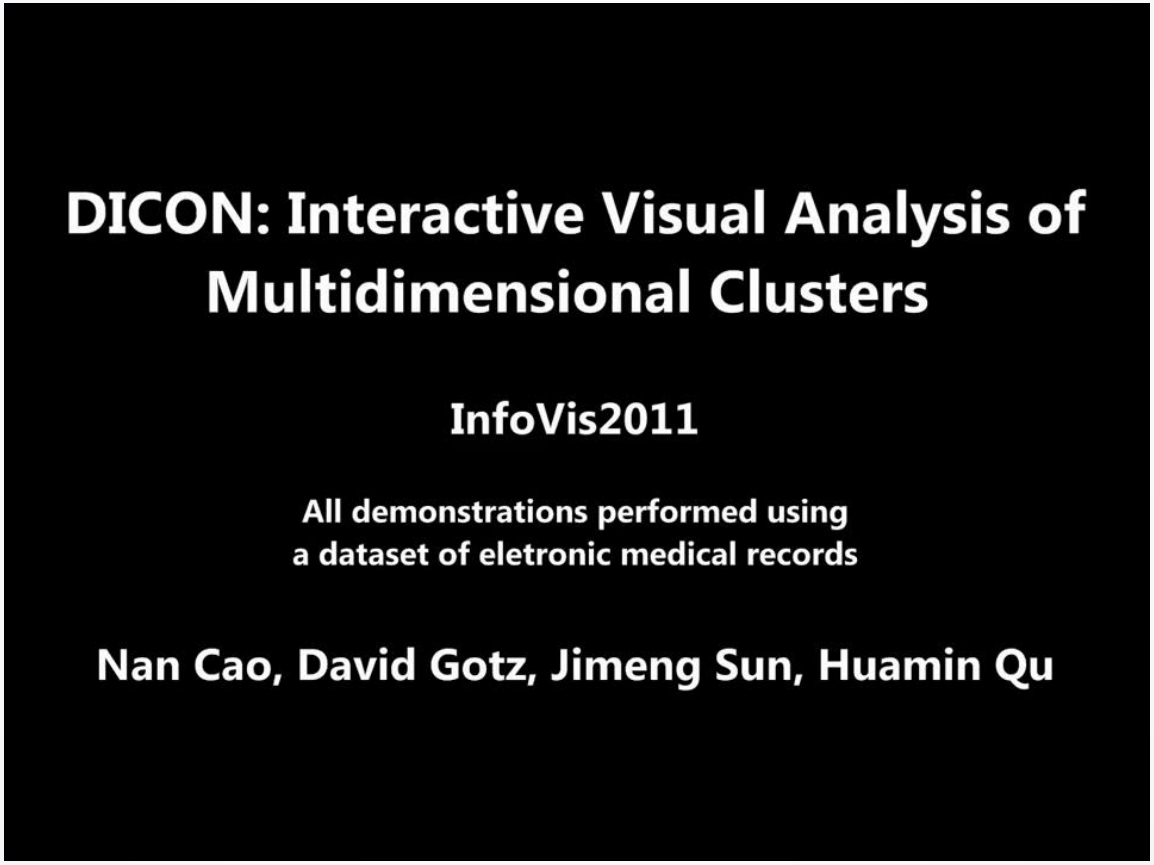
Roll Up & Drill Down



Blanch et al. Navigation Techniques for Zoomable Treemaps (ACM UIST 2006).

Dynamic Icon

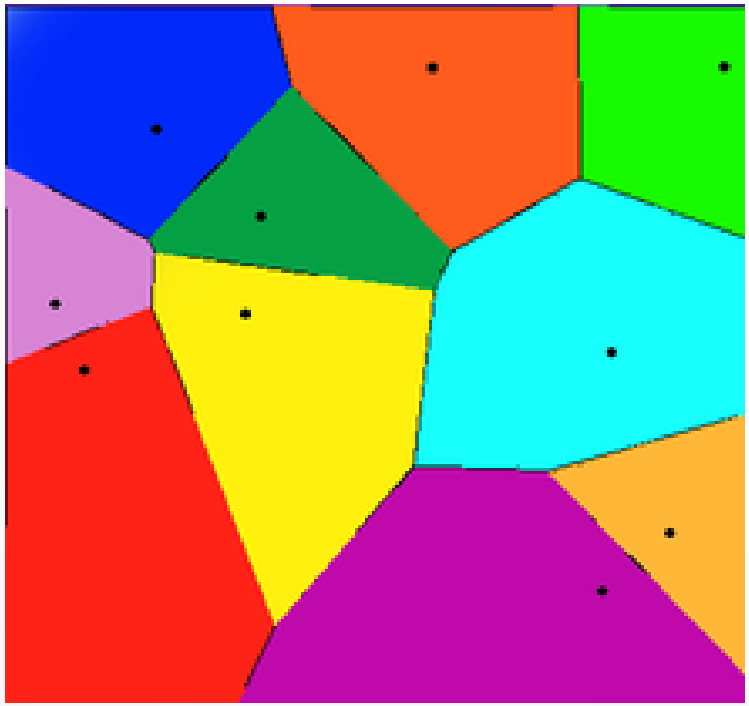
- For categorical data.



Cao et al. DICON: Interactive Visual Analysis of Multidimensional Clusters (IEEE InfoVis 2011).

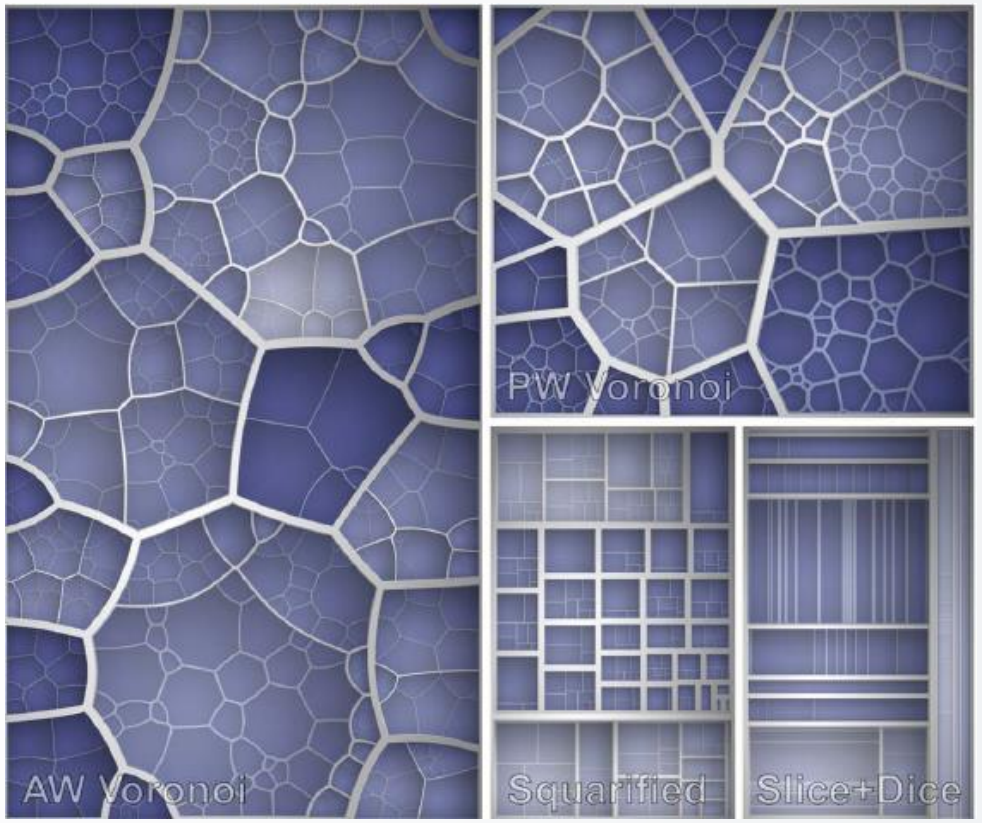
Voronoi Map

- Sphere of influence.
  - Nodes in the sphere are nearer to each other than those outside the sphere.
  - The border of sphere constructs the map.



Voronoi Treemap

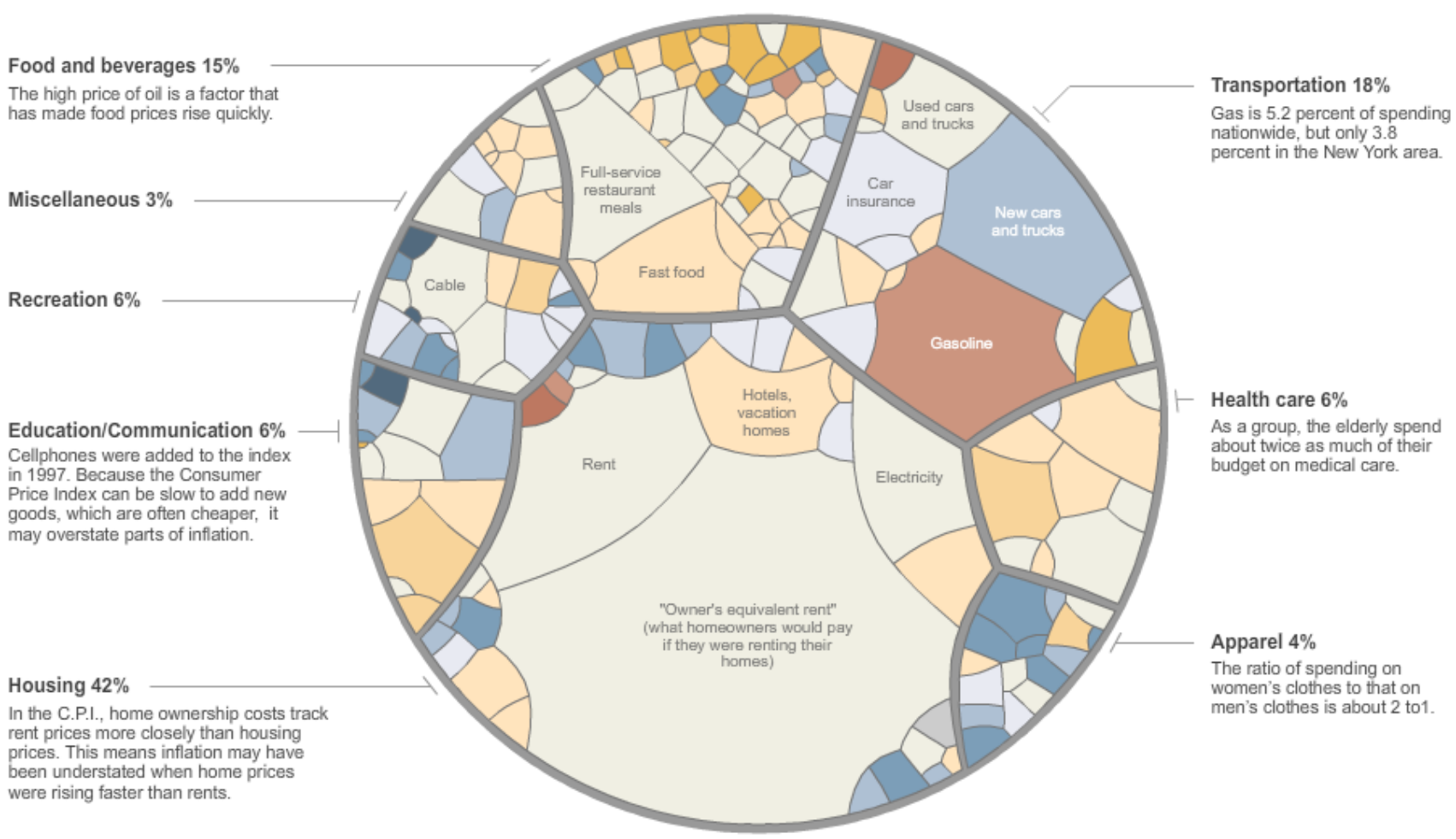
Voronoi treemap	Voronoi treemap	
	Squarified treemap	Slice-and-dice treemap



Balzer M, Deussen O. Voronoi Treemaps

Applications of Space-Filling Methods

American Family Consumption Structure Visualization

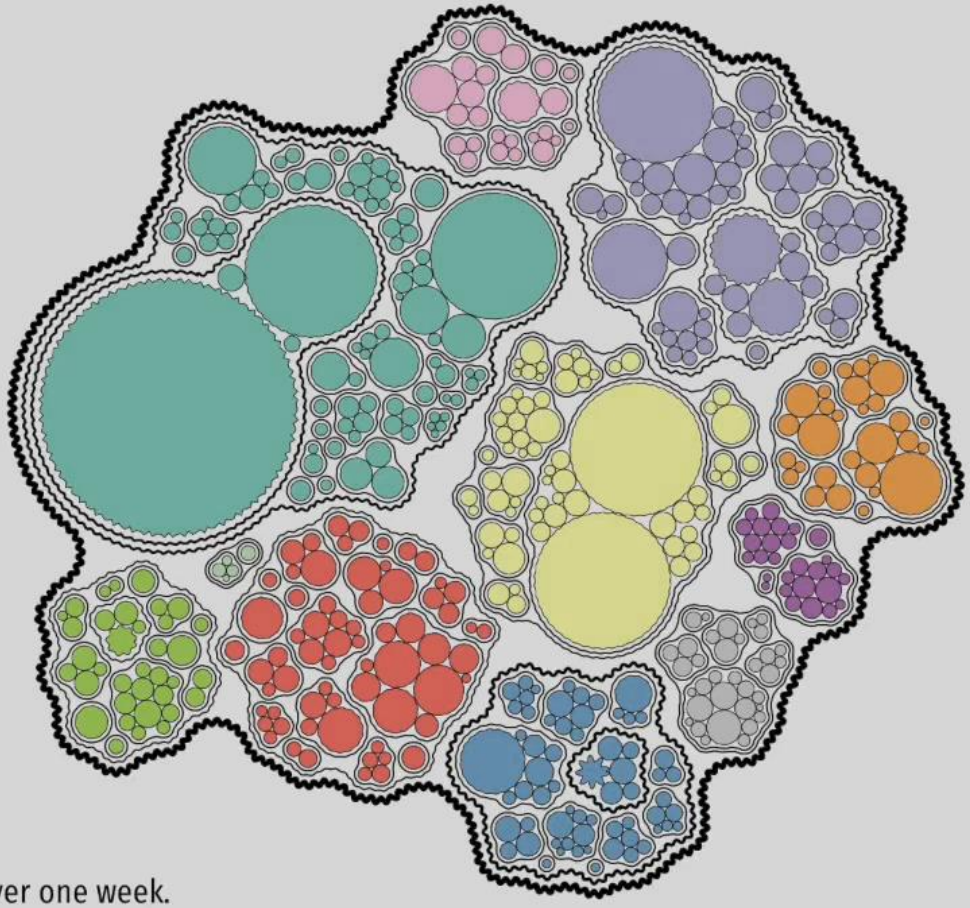


http://nytimes.com/interactive/2008/05/03/business/20080403_SPENDING_GRAPHIC.html

NMap

Duarte et al. NMap: A Novel Neighborhood Preservation Space-filling Algorithm (IEEE InfoVis 2014).

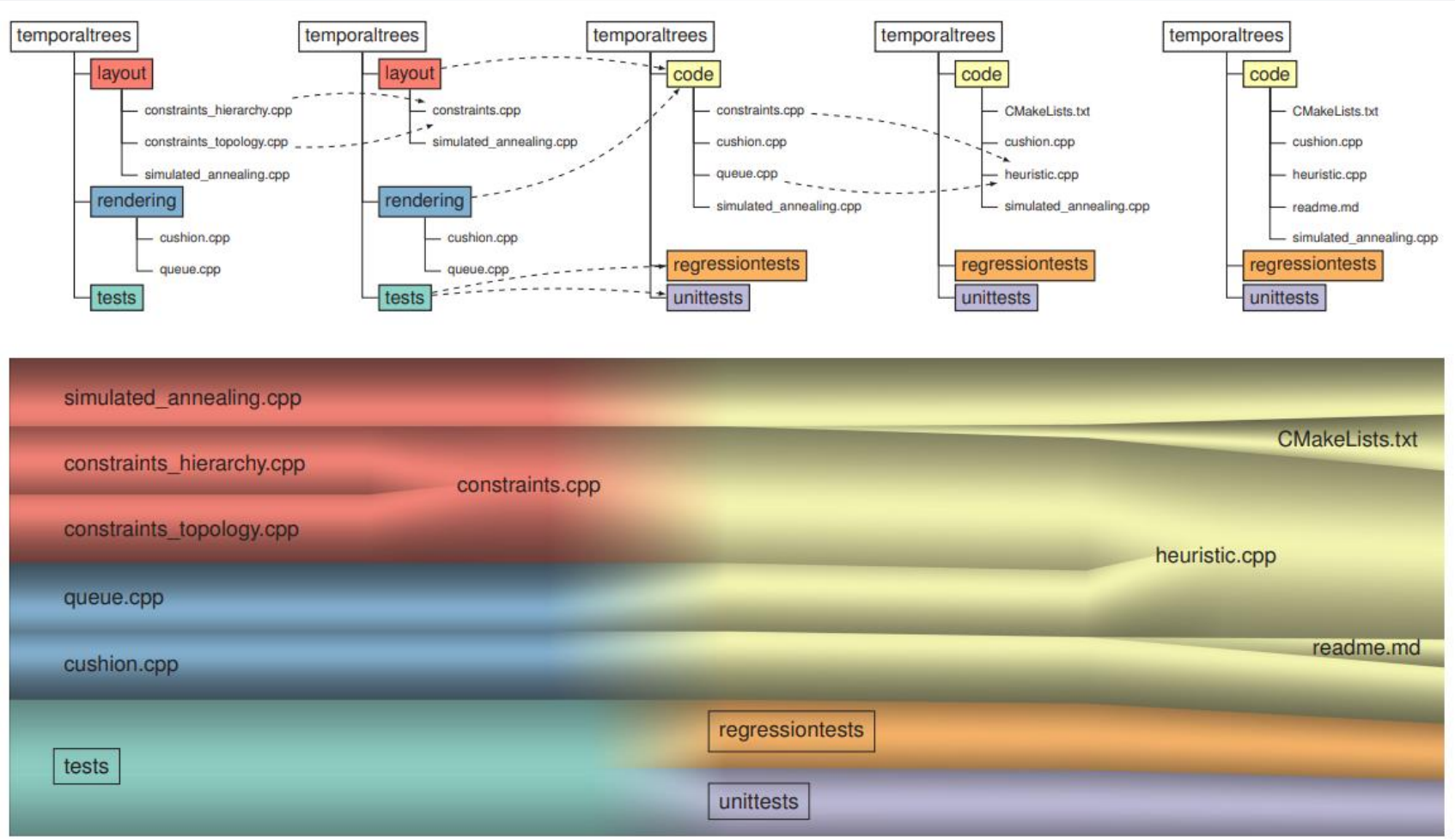
Bubble Treemaps



Stock data aggregated over one week.

Görtler et al. Bubble Treemaps for Uncertainty Visualization (IEEE InfoVis 2017).

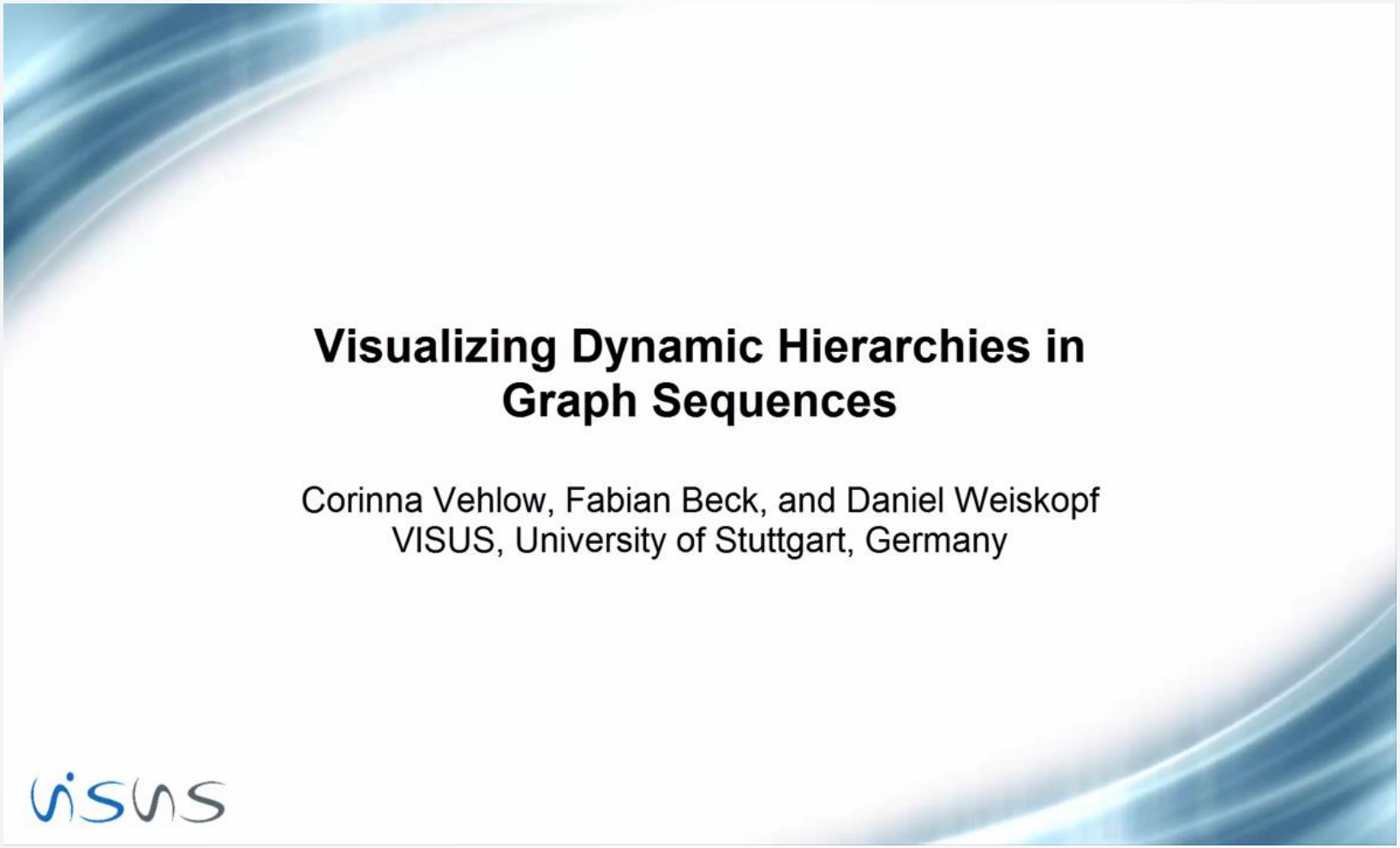
Temporal Treemap



Köpp et al. Temporal Treemaps: Static Visualization of Evolving Trees (IEEE InfoVis 2018).

Hybrids

Hierarchies in Dynamic Networks



Vehlow et al. Visualizing Dynamic Hierarchies in Graph Sequences (IEEE TVCG 2015).

Aesthetical Requirements of Graph Drawing

- It applies efficiency and aesthetics principles.
  - Avoiding crossing edges.
  - Uniformly distributed nodes and edges.
  - Consistent edge length.
  - Symmetry.
- More beauty, less misleading.

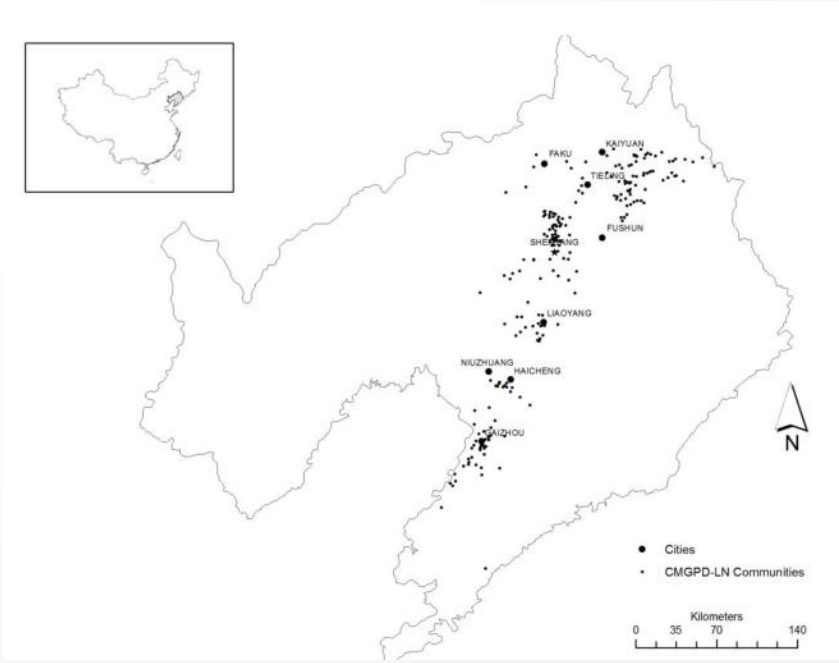
Not All Requirements can be Satisfied…

Case:
Genealogy Data Visualization

Data Description

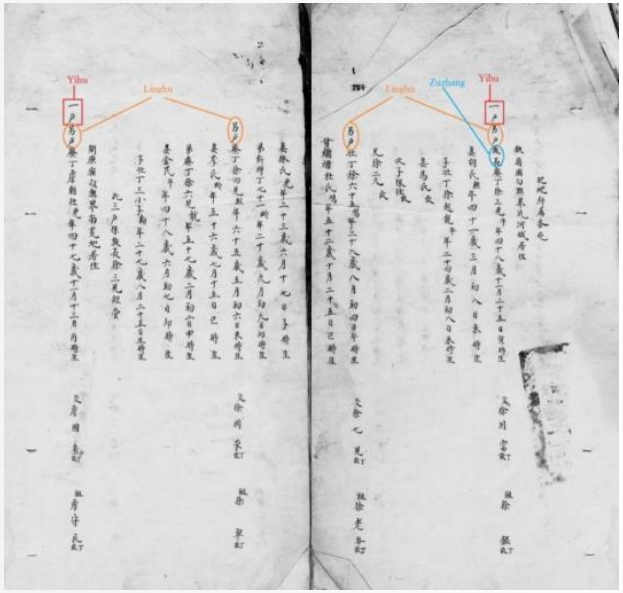
- China Multi-generation Population Database (CMGPD-LN)
- More than 1.5 million records
- More than 260,000 local residents living in Liaoning Province
- 29 administrative units in Shengjing (Liaoning)
- 700 villages
- From 1749 to 1909

- Genealogy Data Information



Data Source

- Eight flags register(八旗户口册)
- Thanks to Mr. James Lee（李中清）and Cameron Campbell(康文林)



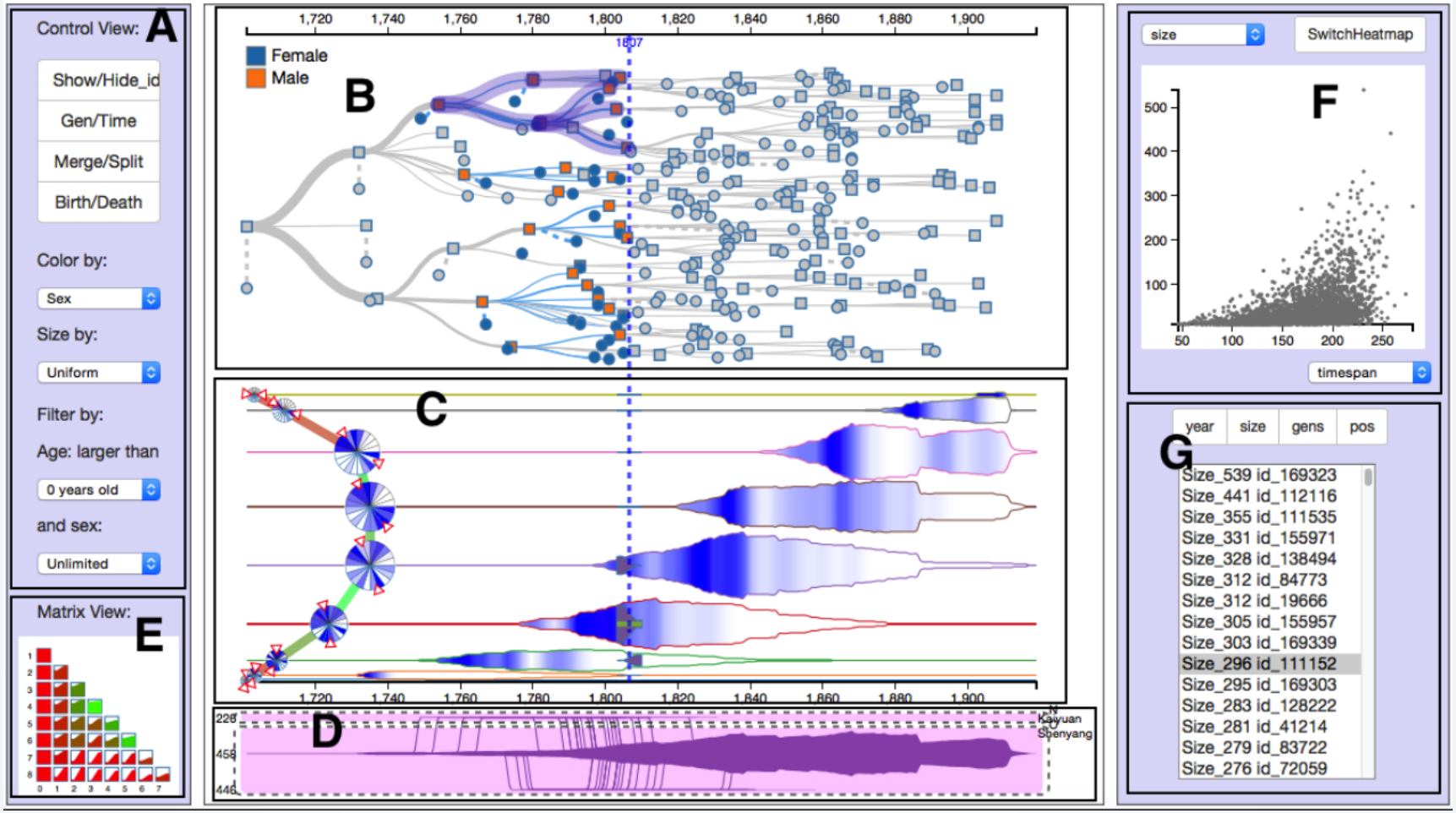
	A	B	D	E	F	G	I	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	AB	AC
1	case	date	clan	hid	yihu	rhhead	addr	sex	gen	rank	occu	mari	name	vital1	vital2	birthyr	age	month	day	hour	disab	lcase	ldate
2	1	792	1	1	0	e	1	2	3	1	7	1	xu saner	0	0	2	48	11	25	3	0	1	786
3	2	792	1	1	0	w	1	1	3	0	0	1	hu shi	0	0	9	41	3	8	8	0	2	786
4	3	792	1	1	0	1s	1	2	4	0	1	3	xu qilong	0	0	2	24	2	8	8	0	3	786
5	4	792	1	1	0	1sw	1	1	4	0	0	0	ma shi	1	0	0	0	0	0	0	0	4	786
6	5	792	1	1	0	2s	1	2	4	0	0	0	baozhu	1	0	0	0	0	0	0	0	0	0
7	6	792	1	1	0	1ob	1	2	3	0	0	0	xu erjiu	1	0	0	0	0	0	0	0	5	786
8	7	792	1	2	0	e	1	2	3	0	1	1	xu liushiwu	0	0	10	28	8	4	7	0	8	786
9	8	792	1	2	0	n	1	1	2	0	0	3	du shi	0	0	10	52	10	25	6	0	7	786
10	9	792	1	2	0	w	1	1	3	0	0	1	lin shi	7	0	3	23	6	17	1	0	0	0
11	10	792	1	2	0	1yb	1	2	3	0	3	2	qishiyi	0	0	6	20	9	9	4	0	9	786
12	11	792	1	3	0	e	1	2	3	0	7	1	xu sier	0	0	9	65	5	6	8	0	10	786
13	12	792	1	3	0	w	1	1	3	0	0	1	li shi	7	0	6	56	7	15	6	0	0	0
14	13	792	1	3	0	1yb	1	2	3	0	7	1	xu liuer	0	0	5	57	2	2	9	0	11	786
15	14	792	1	3	0	1ybw	1	1	3	0	0	1	jin shi	0	0	2	48	6	7	4	0	12	786
16	15	792	1	3	0	1yb1s	1	2	4	0	1	2	sanxiaozi	0	0	11	27	8	25	2	0	13	786
17	16	792	2	1	0	e	2	2	3	0	7	1	zhan zhaoshi	0	0	3	47	11	13	10	0	18	786
18	17	792	2	1	0	f	2	2	2	0	0	0	zhan guoxiu	1	0	0	0	0	0	0	0	16	786
19	18	792	2	1	0	w	2	1	3	0	0	1	zhou shi	0	0	6	44	1	14	10	0	19	786
20	19	792	2	1	0	1s	2	2	4	0	1	0	zhan vencheng	1	0	0	0	0	0	0	0	20	786

李中清，康文林：中国多代人口追踪数据库，辽宁（CMGPD-LN）1749-1909，[Computer file]. ICPSR27063-v6. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-09-02. doi:10.3886/ICPSR27063.v6

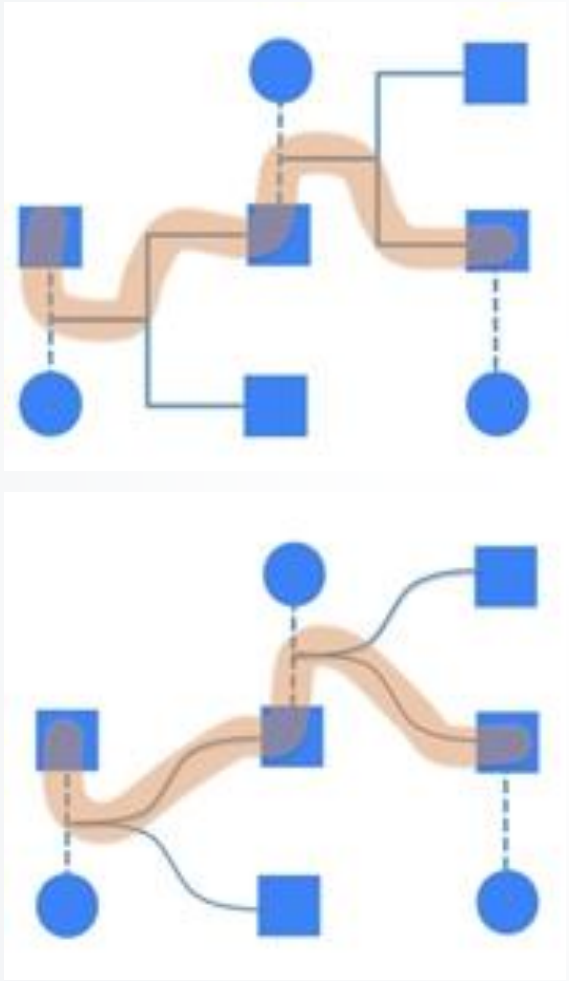
Analytical Tasks

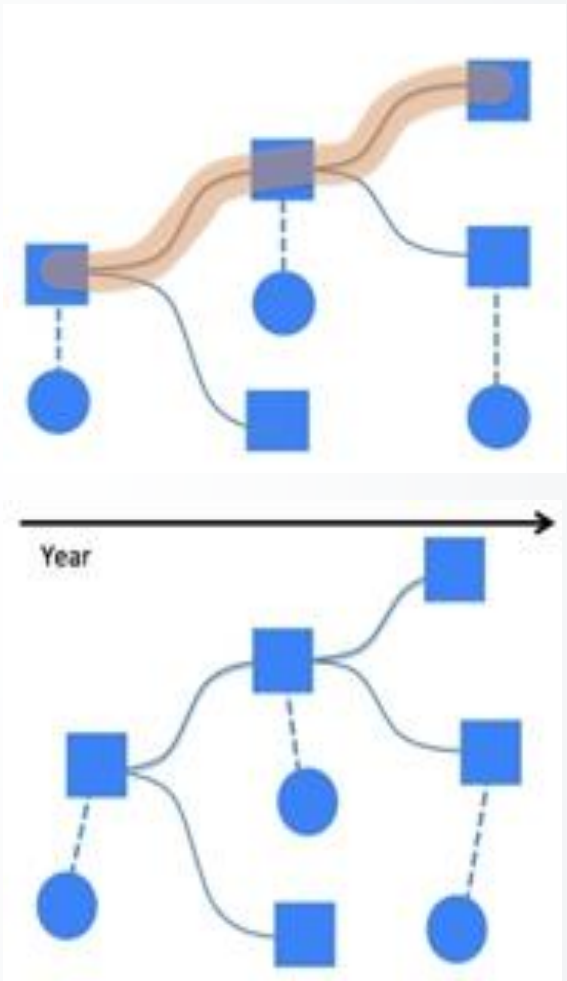
- Demographic characteristics of the whole family or a generation?
- The Change Process of Family Migration Behavior over Time and the Causes of Migration Behavior?
- How to find the specific patterns in the evolution of family?
- Whether the reproductive relationship between generations is balanced and how to quantify it?
- The relationship between spatial distribution and family attributes?

System Overview



Family tree view

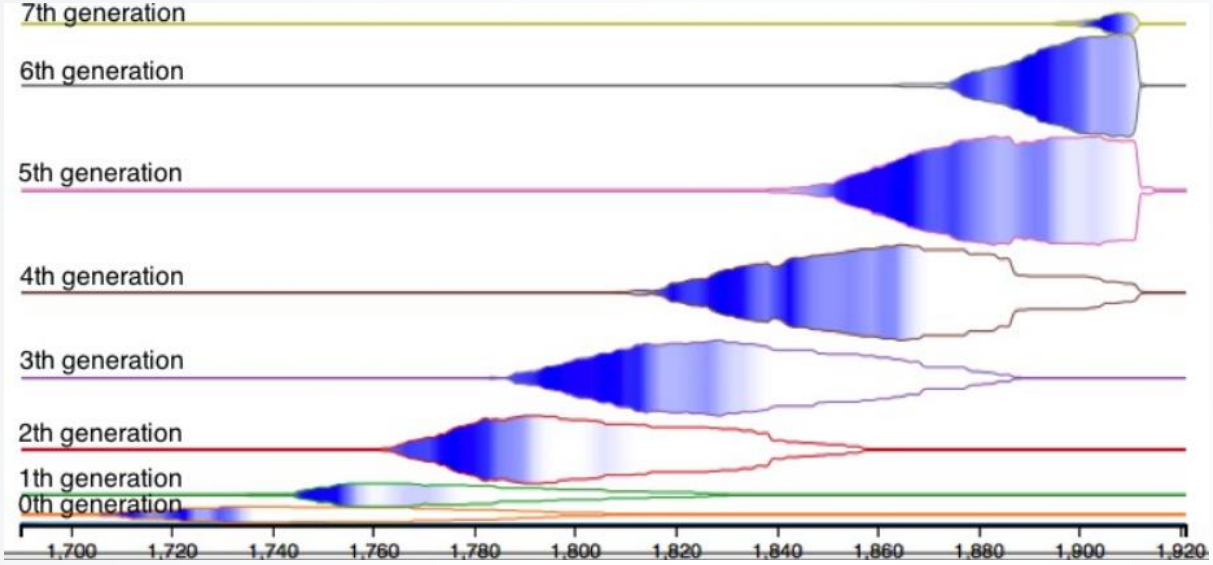




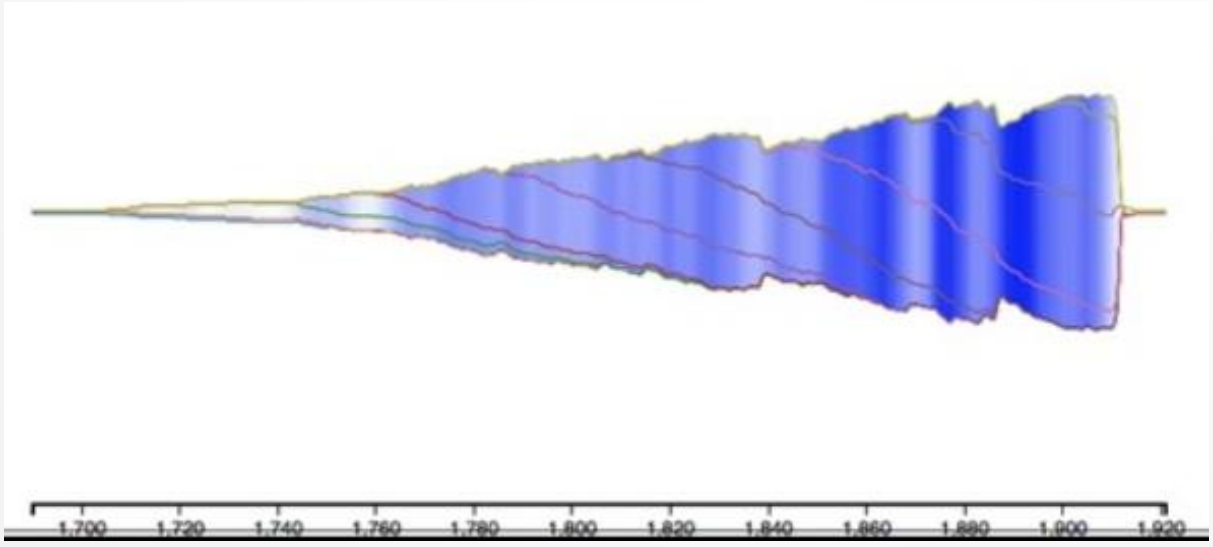
Traditional design

Our design

Demographic view

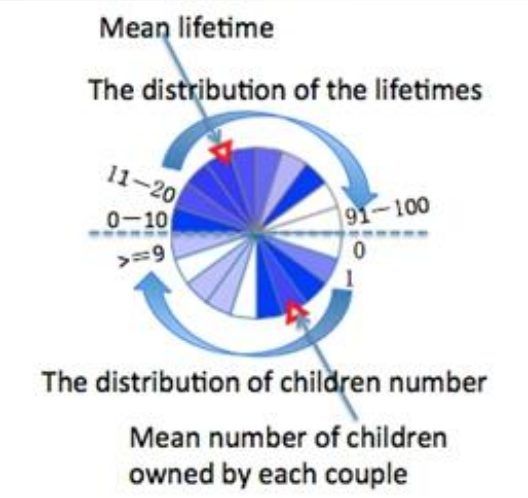


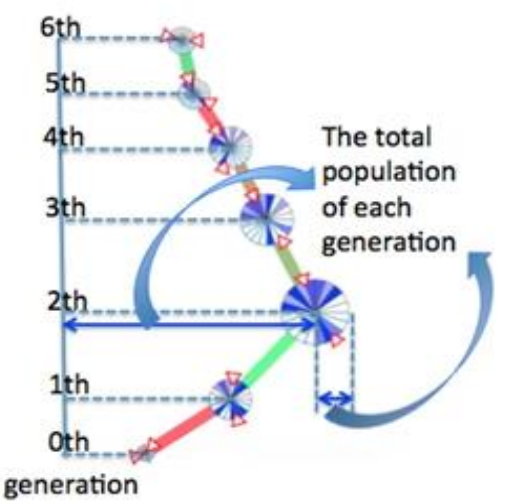
Time-varying population maps and birth rates for each generation

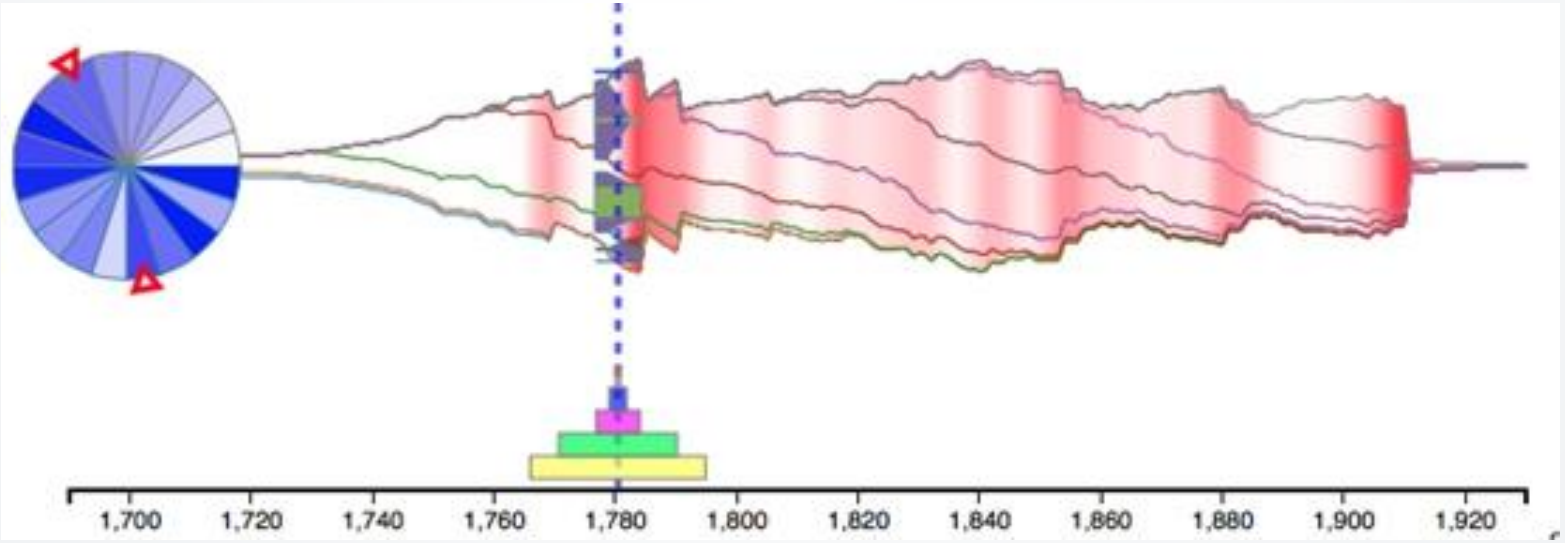


Population Time-varying Map and Birth Rate of the Whole Family

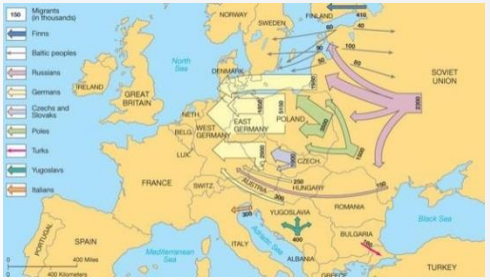
Demographic view



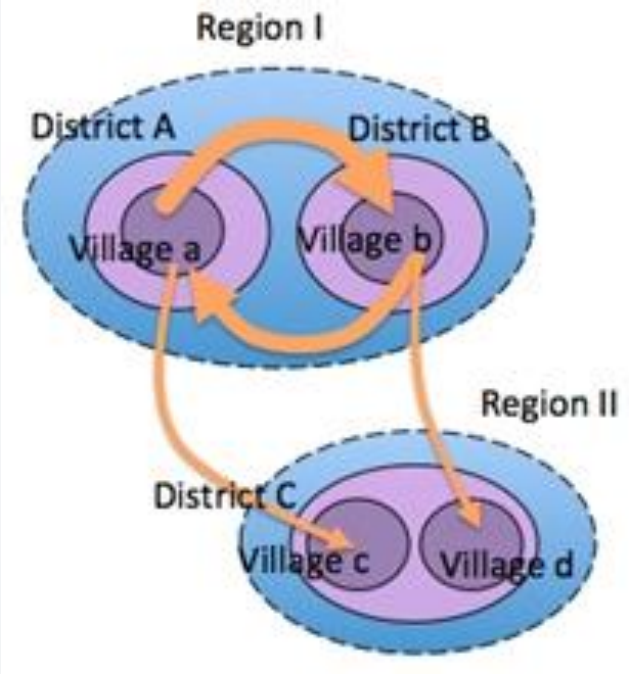


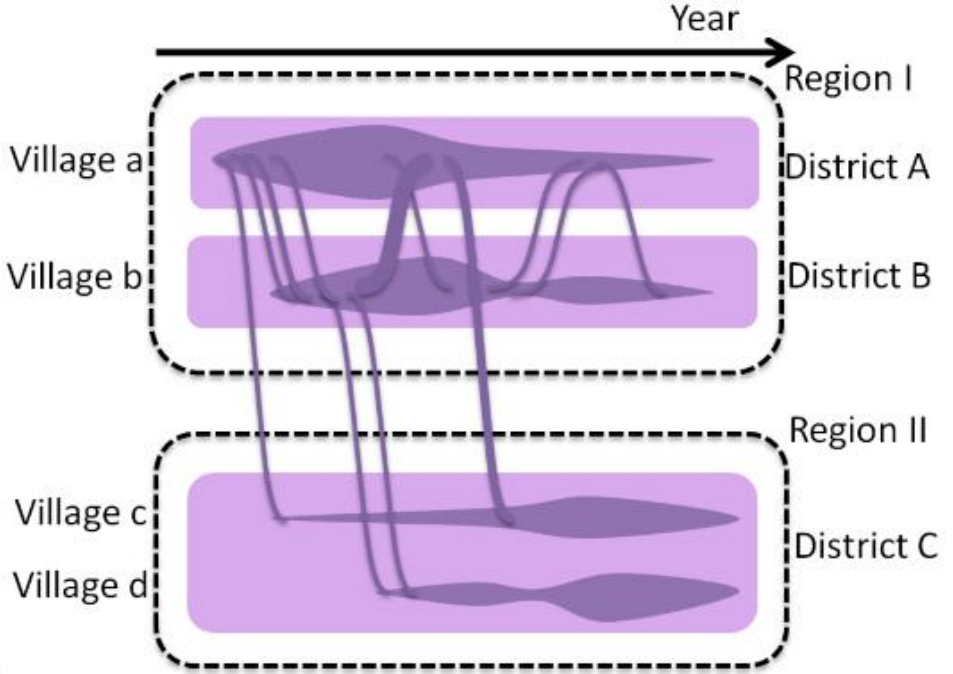


Migration View

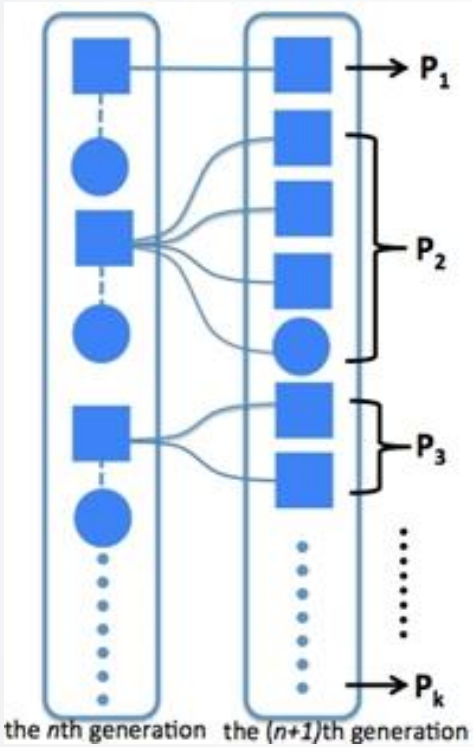


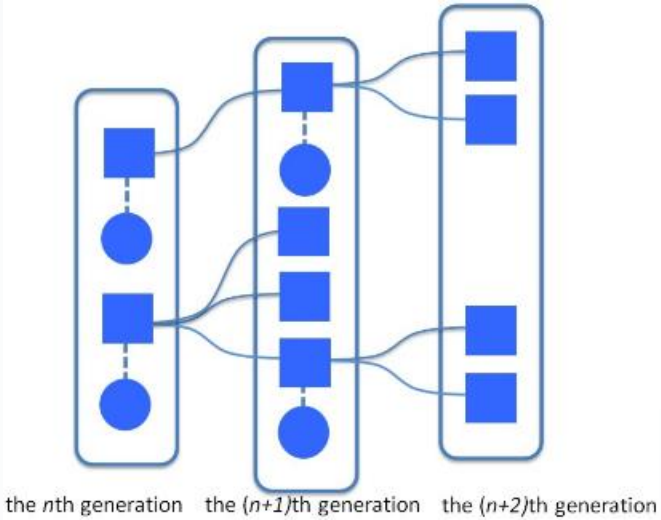
Original migration design

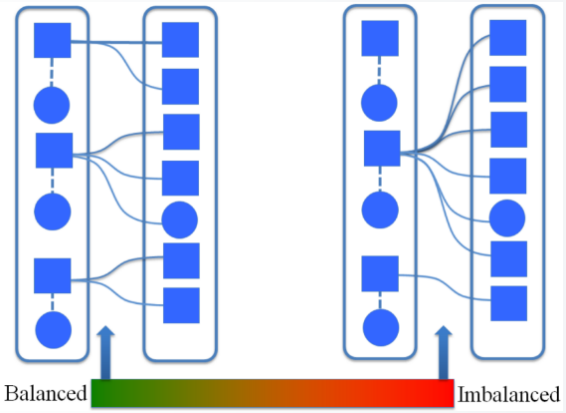




Intergenerational Reproduction Matrix

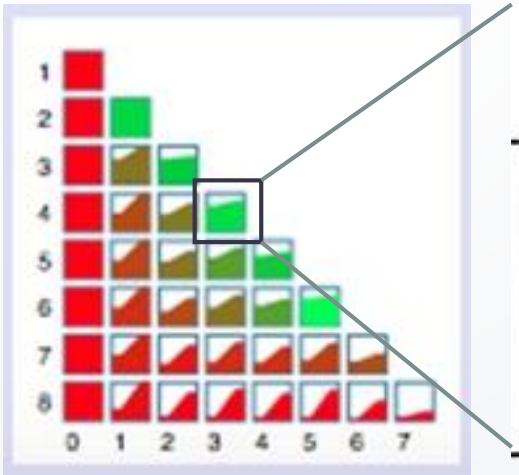


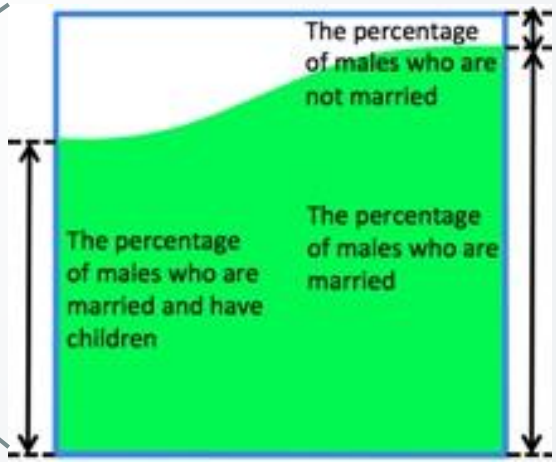




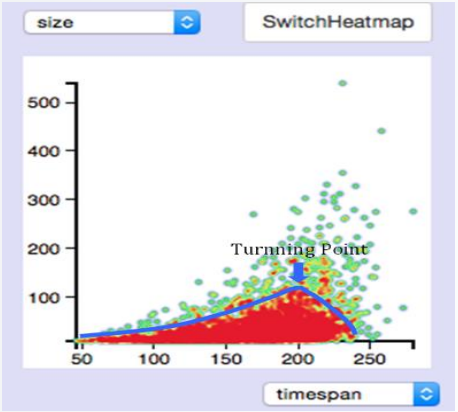
Color mapping

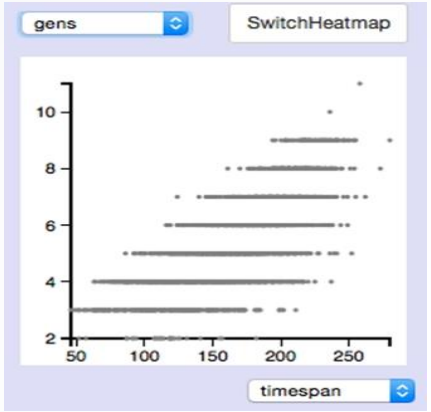
$$RIE\,(n \cdot n+1) = -\left(\sum_{i=1}^{k} P_i \log P_i\right) \times \frac{f}{F}$$

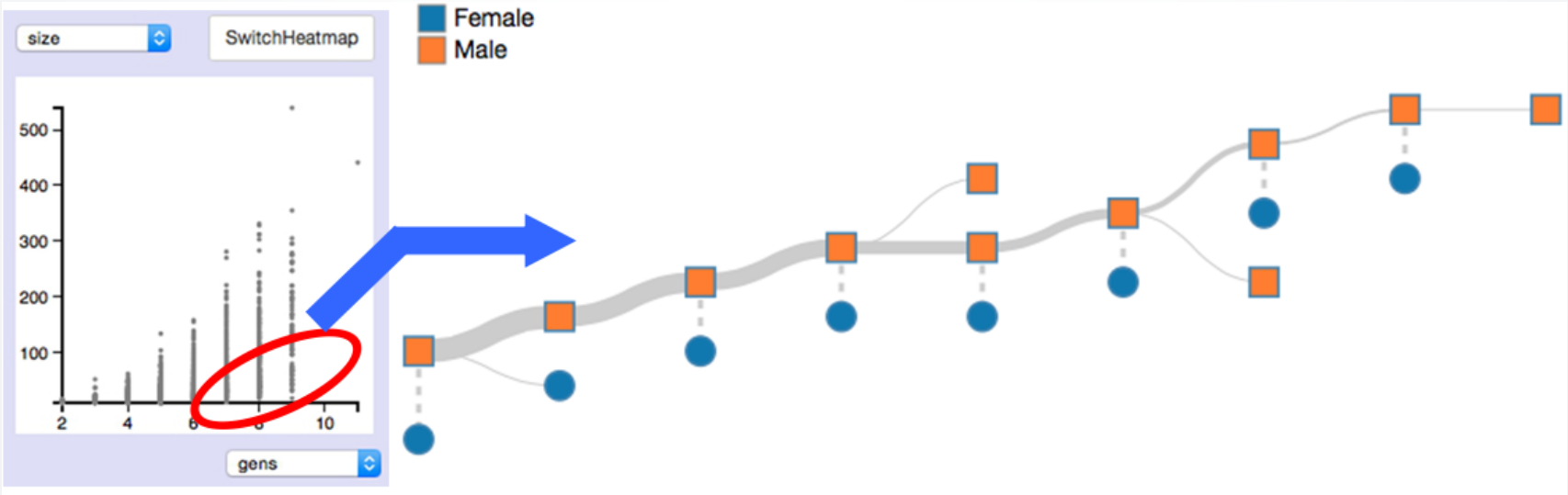




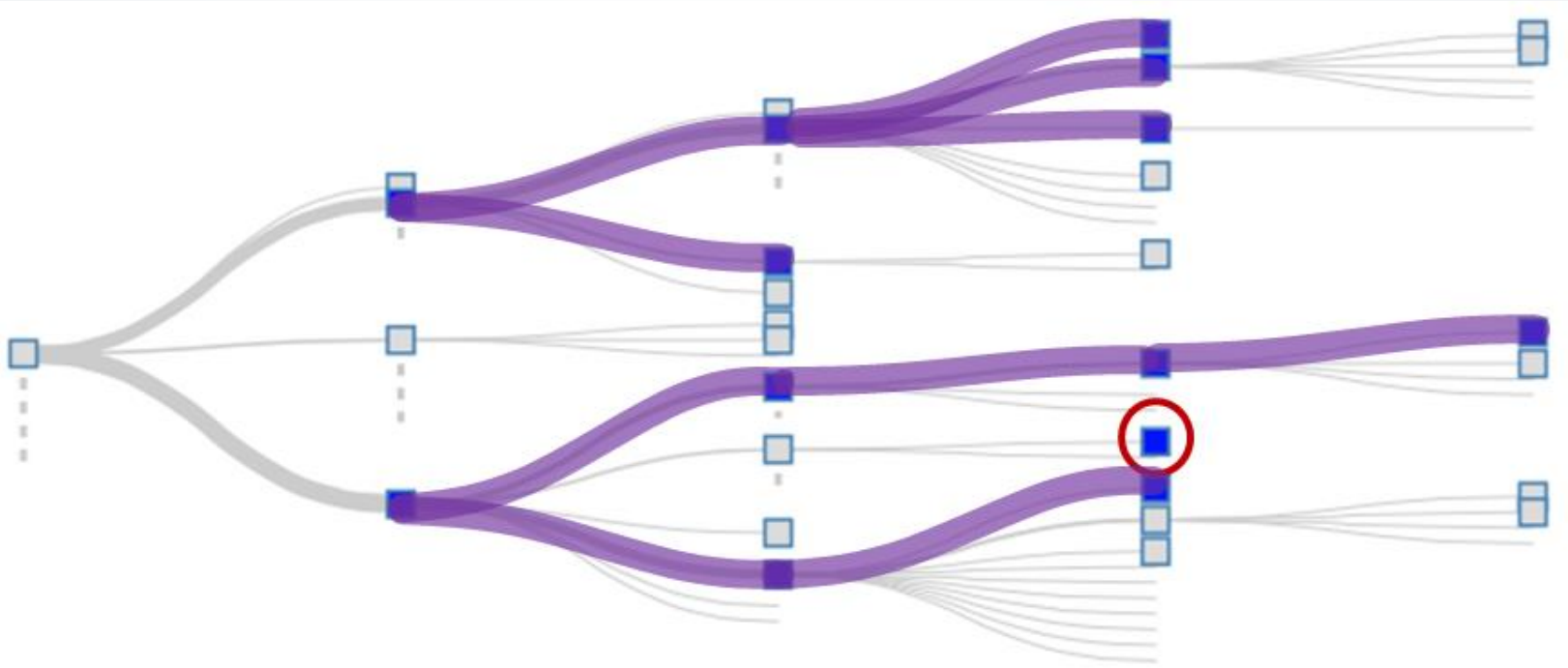
Case 1


size
SwitchHeatmap
500
400
300
200
100
Turnning Point
50
100
150
200
250
timespan

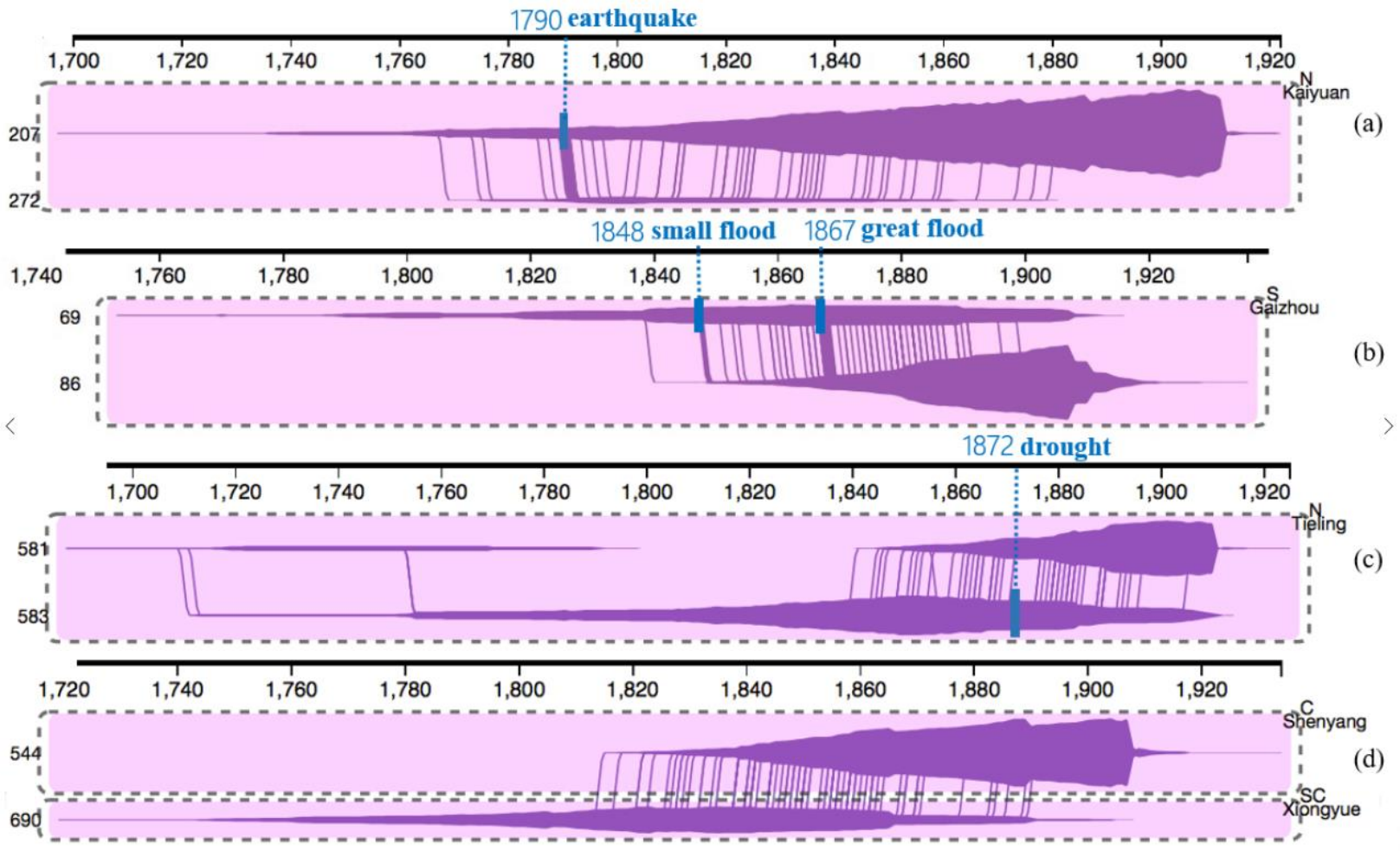

gens
SwitchHeatmap
10
8
6
4
2
50
100
150
200
250
timespan


size
SwitchHeatmap
500
400
300
200
100
2
4
6
8
10
gens
Female
Male

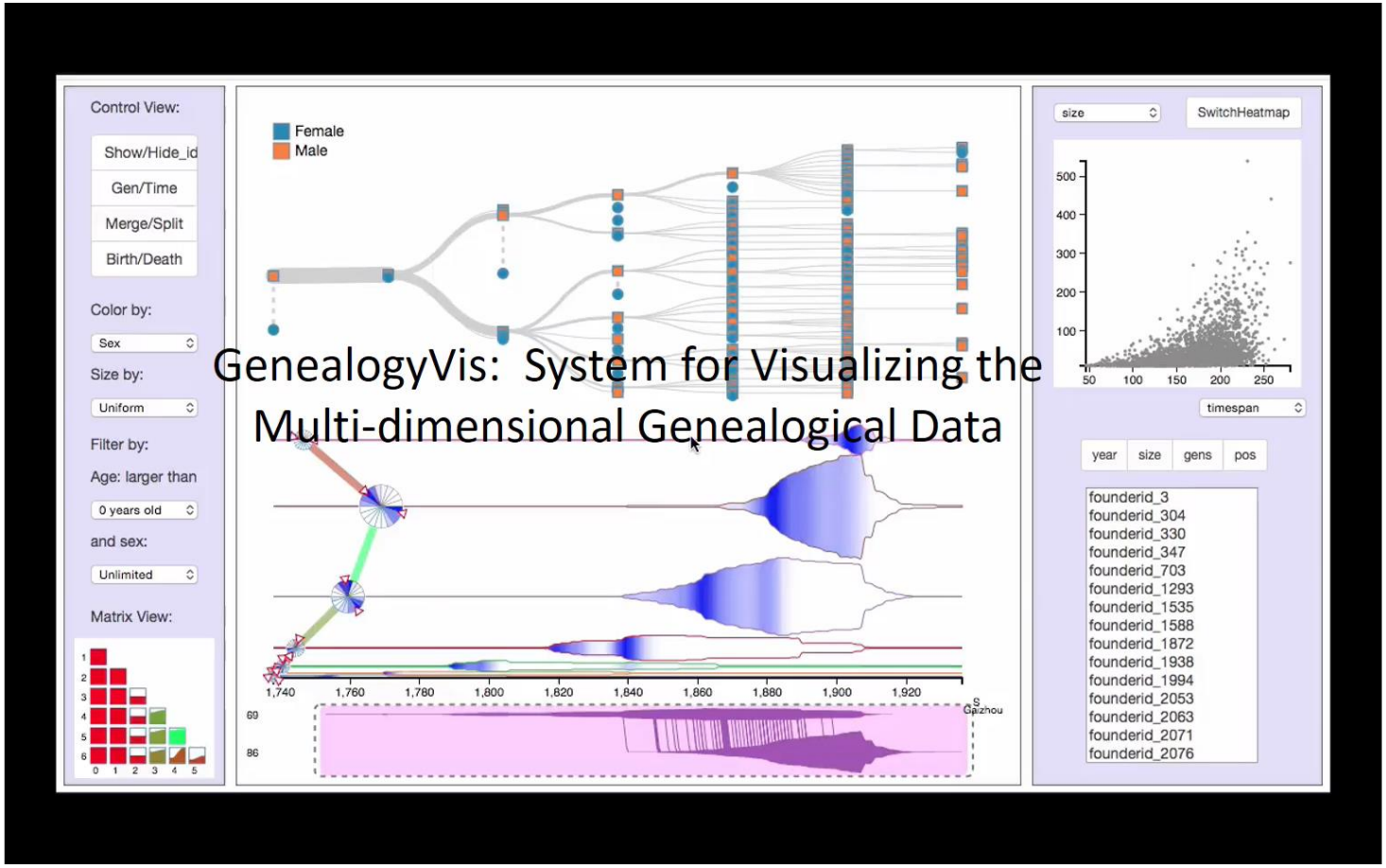
Case 2



Case 3


1790 earthquake
1,700 1,720 1,740 1,760 1,780 1,800 1,820 1,840 1,860 1,880 1,900 1,920
N
Kaiyuan
207
272
(a)
1848 small flood
1867 great flood
1,740 1,760 1,780 1,800 1,820 1,840 1,860 1,880 1,900 1,920
S
Gaizhou
69
86
(b)
1872 drought
1,700 1,720 1,740 1,760 1,780 1,800 1,820 1,840 1,860 1,880 1,900 1,920
N
Tieling
581
583
(c)
1,720 1,740 1,760 1,780 1,800 1,820 1,840 1,860 1,880 1,900 1,920
C
Shenyang
544
SC
Xiongyue
690
(d)

video


Control View:
Show/Hide_id
Gen/Time
Merge/Split
Birth/Death
Color by:
Sex
Size by:
Uniform
Filter by:
Age: larger than
0 years old
and sex:
Unlimited
Matrix View:
Female
Male
GenealogyVis: System for Visualizing the Multi-dimensional Genealogical Data
1,740
1,760
1,780
1,800
1,820
1,840
1,860
1,880
1,900
1,920
size
SwitchHeatmap
timespan
year
size
gens
pos
founderid_3
founderid_304
founderid_330
founderid_347
founderid_703
founderid_1293
founderid_1535
founderid_1588
founderid_1872
founderid_1938
founderid_1994
founderid_2053
founderid_2063
founderid_2071
founderid_2076

Network Data Visualization

2 Network Data Visualization

OUTLINE

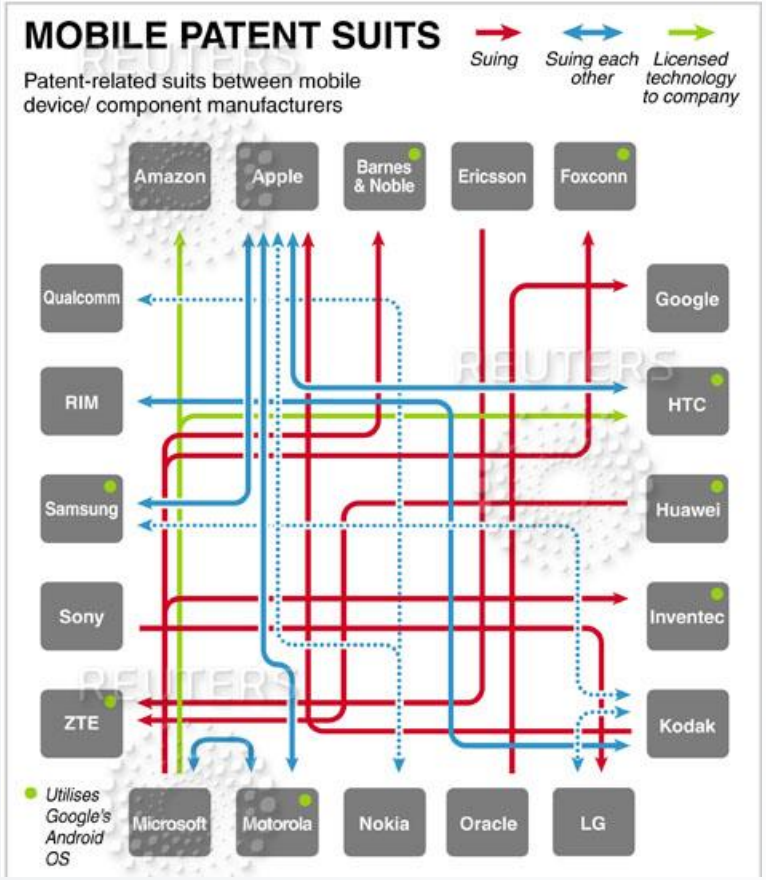
- Network Data
- Network Data Visualization
  - Node-Link Diagram
  - Adjacency Matrix
  - Hybrids
- Case: OD Network Visualization

Network Data

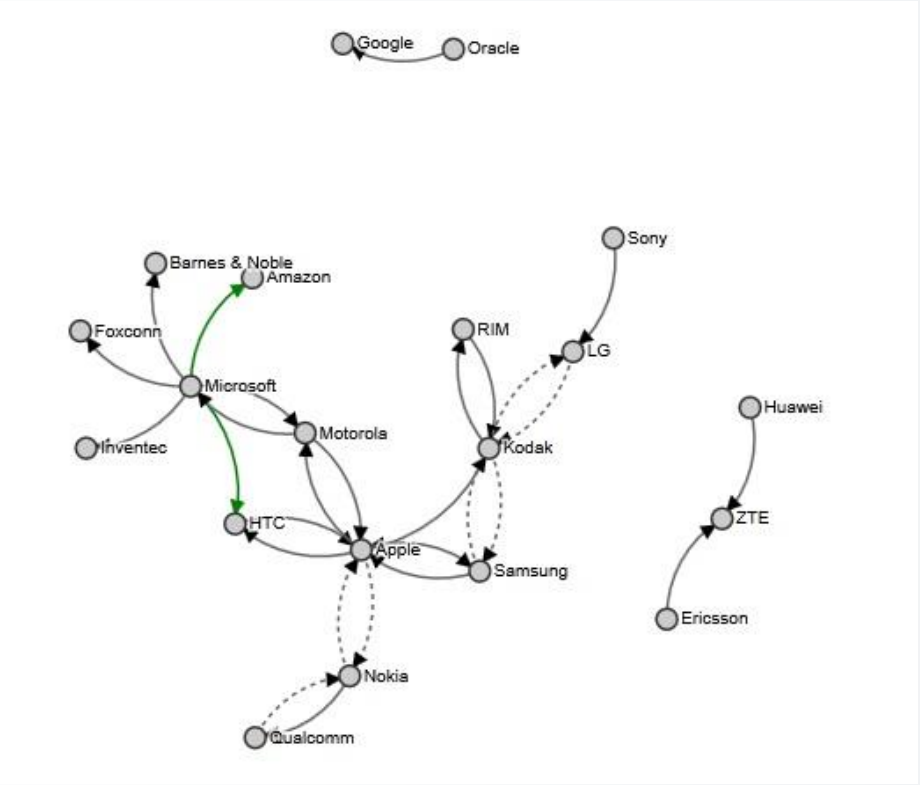
Network Data

- Compared with hierarchies in tree visualization, network data present more **complex and flexible relations**.
  - Social network
  - Mobile network
  - Mail network
  - Collaboration network

Mobile Patent Lawsuits



http://blog.thomsonreuters.com/index.php/mobile-patent-suits-graphic-of-the-day/



Based on D3.js
http://bl.ocks.org/1153292

Properties of Network

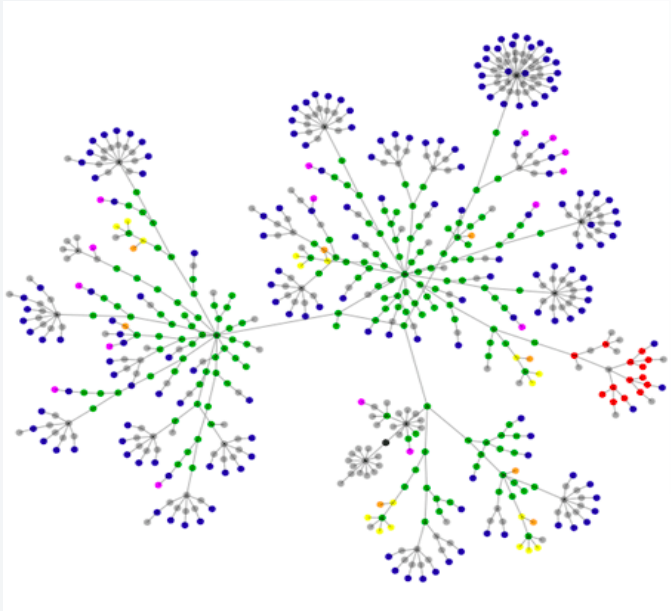
- Complex relations
  - Directions, weights
- Network centrality
  - Degree
  - Closeness
  - Betweenness
  - Eigenvector

Network Data Visualization

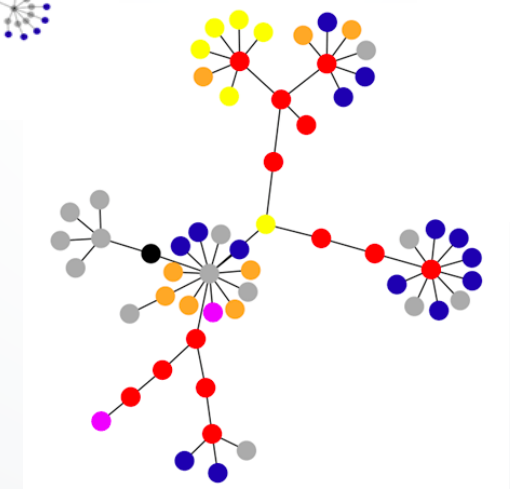
Node-link diagram

Adjacency matrix

Website Navigation Graph



msn.com



Google

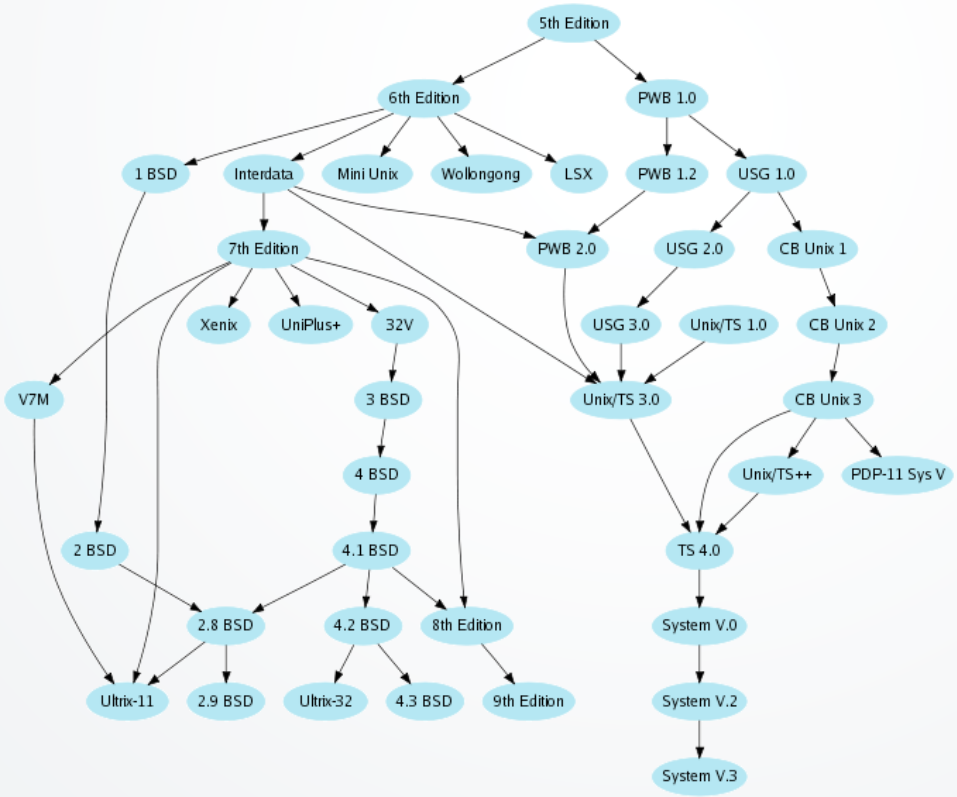
CNN.com

Graph Representation

- Node-link diagram
  - Sugiyama
  - Force-directed layout
  - Other layout
- Adjacency matrix
- Attribute-based representation

Sugiyama Representation

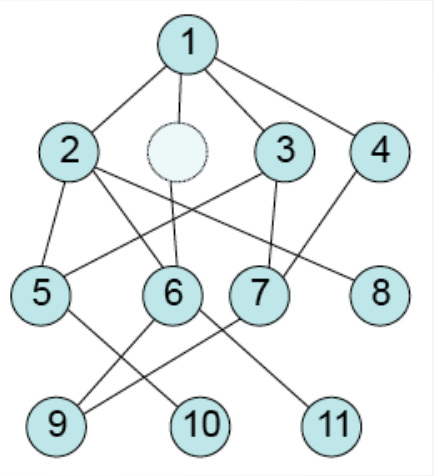
- Good for **intrinsically ordered** graphs
- Depth of graph can be mapped to axes

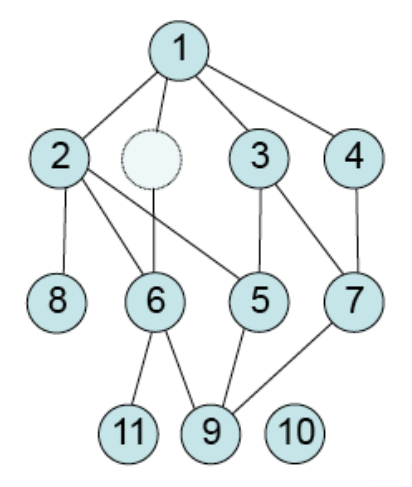


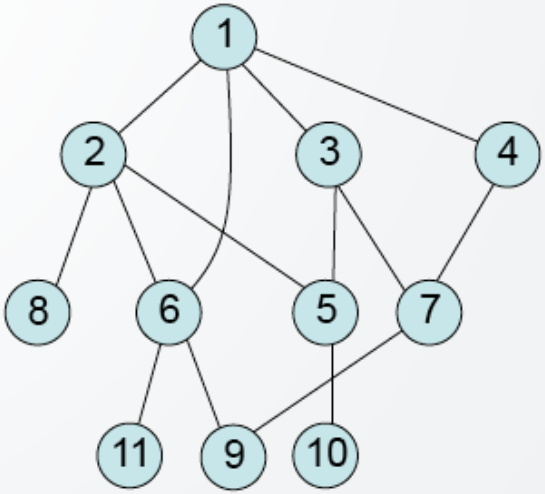
UNIX Family

Sugiyama Procedures

- Construct hierarchy of the graph
  - Based on domain knowledge
  - Construct an optimal order
- Add pseudo-nodes to address long edge
- Reduce the crossing
- Heuristic methods
- Finalize location of nodes
- Set the edges

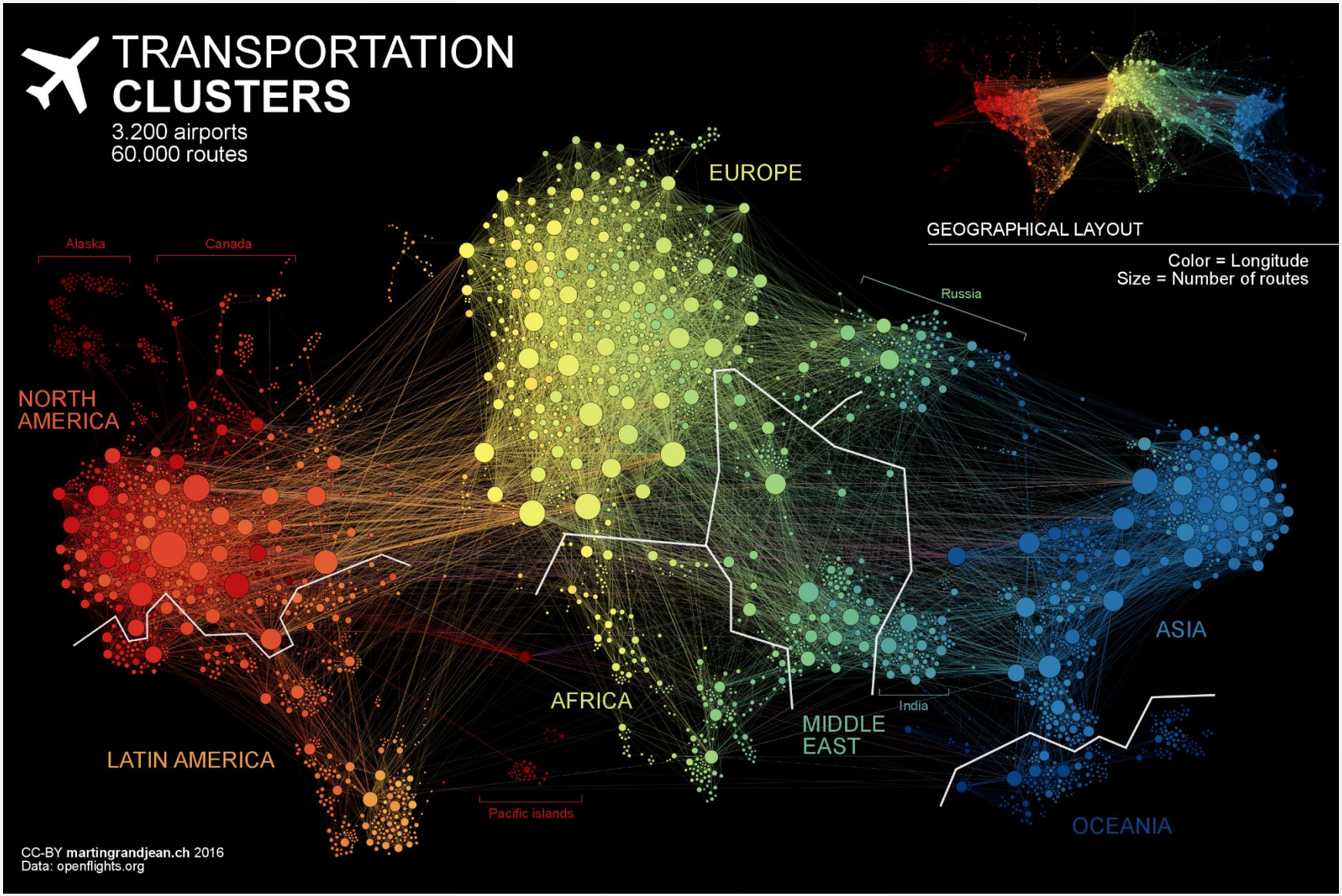






Sugiyama

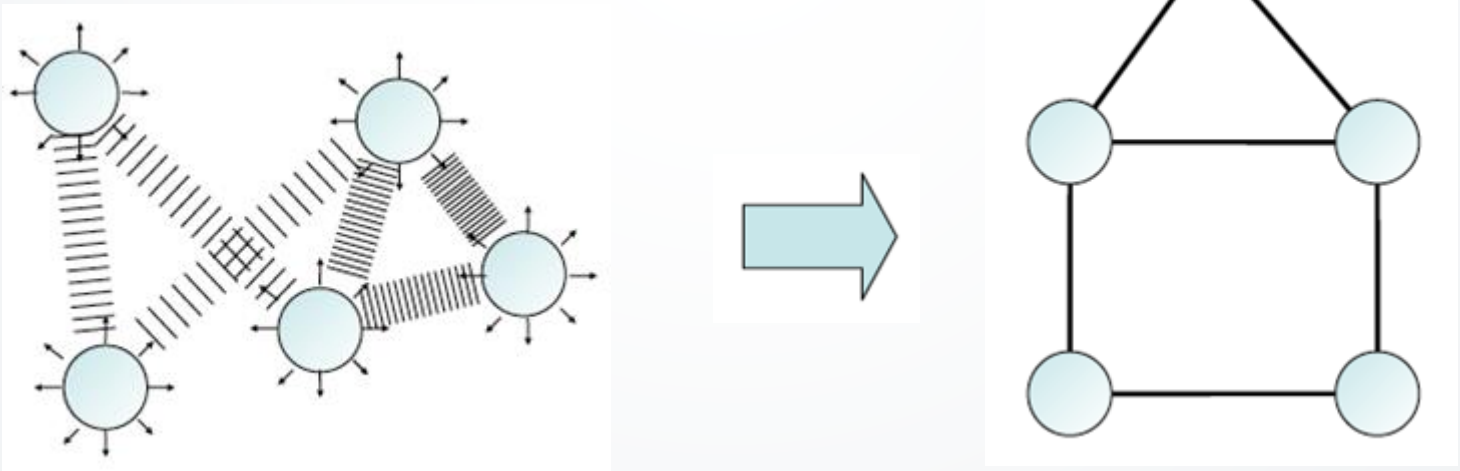
- High readability
- Fast (depends on the heuristic method)
- Not so good for graph without **intrinsic order**
- Easy to implement (free library graphviz, http://www.graphviz.org)



http://www.martingrandjean.ch/connected-world-air-traffic-network/

Force-based algorithm

- What about graphs without intrinsic order?
- Physics model:
  - edge → spring
  - node → mass point



Force-Directed Layout

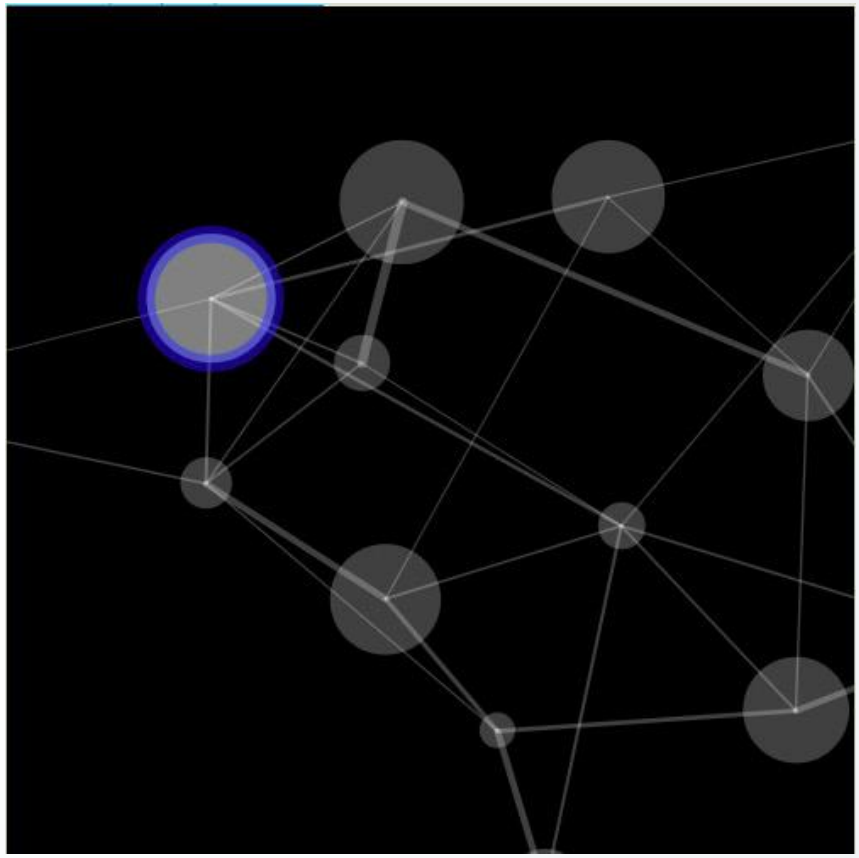
- Peter Eades introduced the layout in 1984 in the paper A Heuristic for Graph Drawing.
- With the effects of **spring force and gravitation**, nodes far away will be dragged near and vice versa.
- The layout reaches a **balance** and becomes stable after **iterations**.
- Spring Model: $E_s = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{2} k(d(i,j) - s(i,j))^2$
- Energy Model: $E = E_s + \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{r w_i w_j}{d(i,j)^2}$

Nodes Position Computation*

- From random or initial **configuration**
- Loop:
  - Compute **the repulsion and attraction** force for every pair of nodes.
  - Accumulate the **force** (vector) for every node.
  - Update nodes position **step by step** according to their forces.
- Loop stops when the layout is "good enough"

Advantages

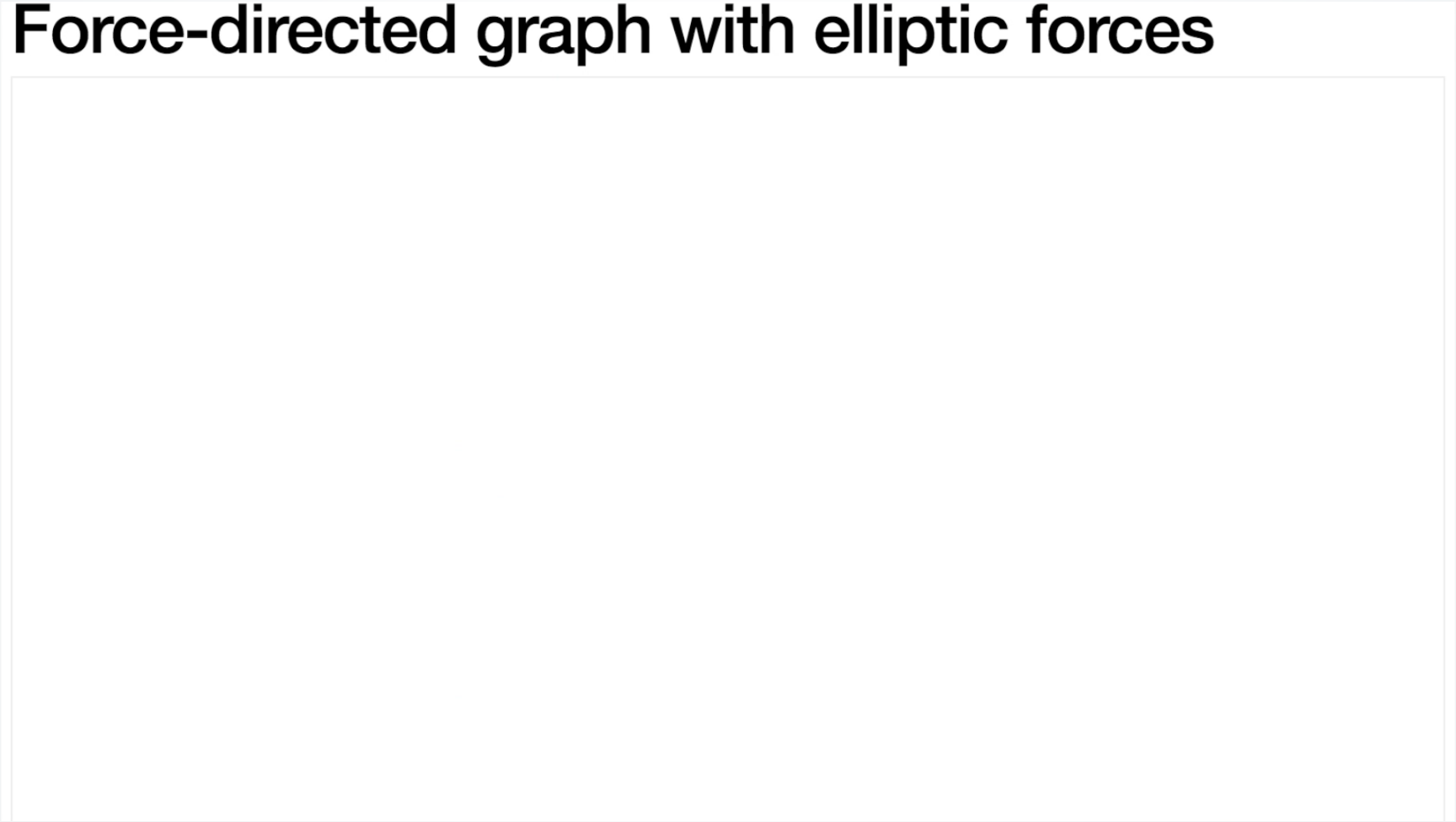
- Very flexible for any type of graphs
- Forces can be customized
- Easy to implement



Limitations and Extensions

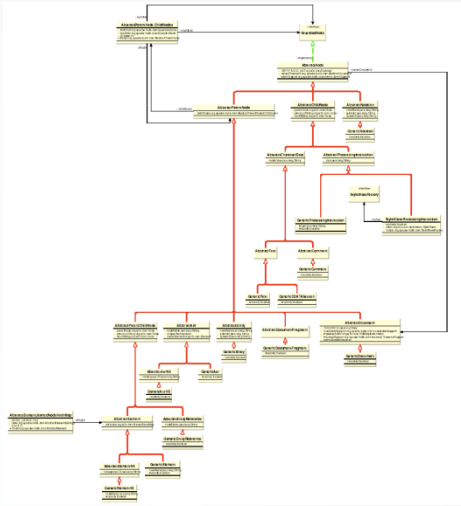
- Limitations
  - **Local optimal**
  - **Initial configuration** is important
  - **Computation complexity** of iterative algorithm
    - Computation complexity for each iteration is $O(N^2)$
- Extensions
  - Barnes-Hut quadtree decomposition
  - FADE、GRIP、FMS、FM^3、GVA

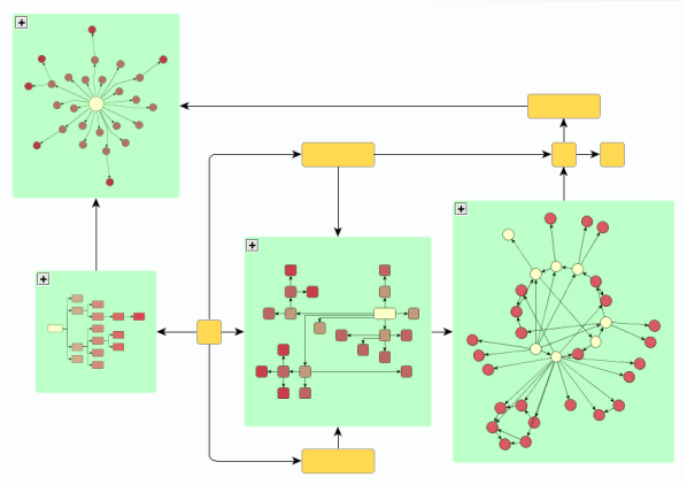
Results of Force-directed Layout



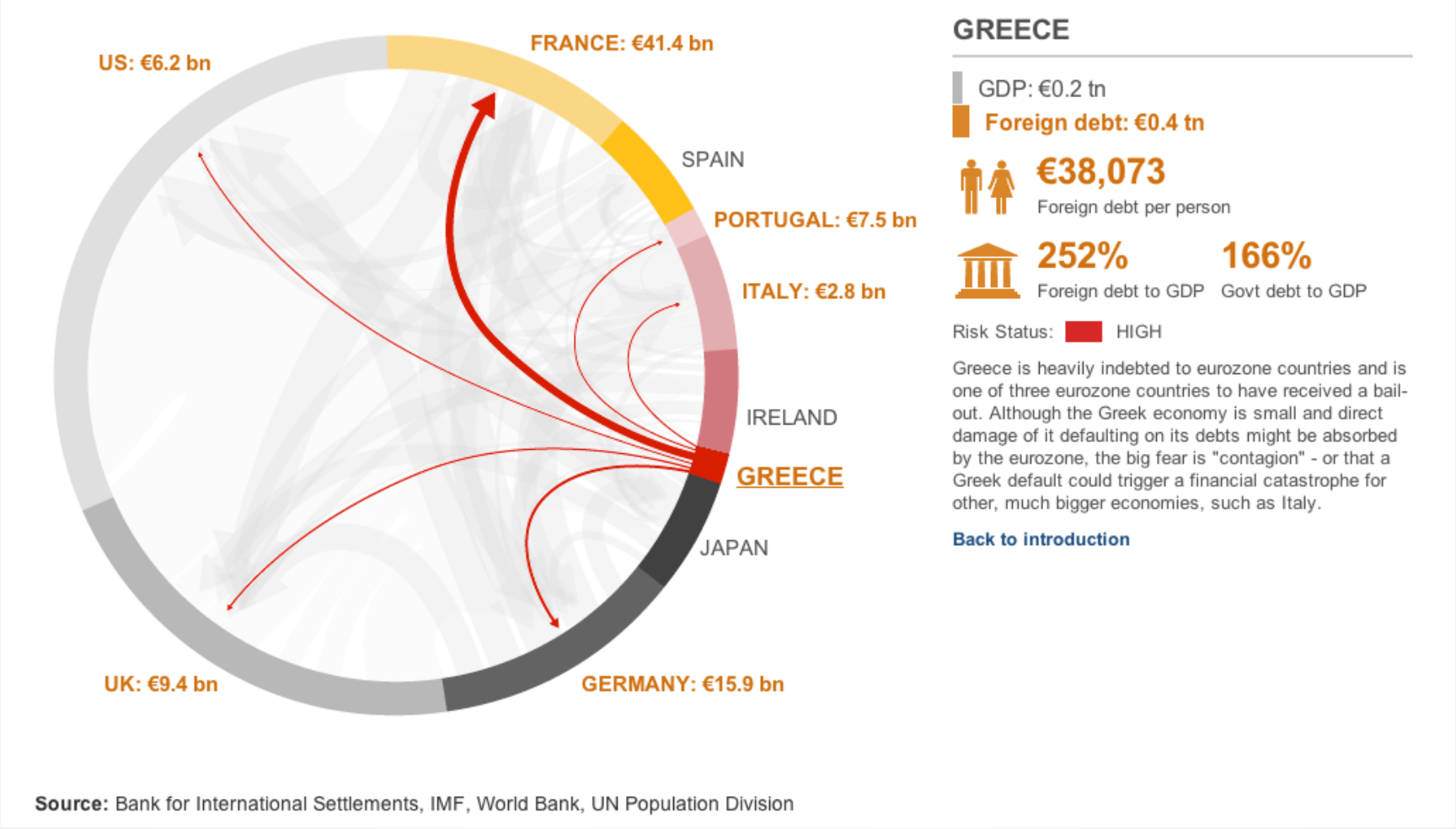
Other presentations of node-link diagram

- Orthogonal Diagram
  - UML diagram
- Ring ordered
  - For ring topology
  - Widely used in social network
- Nested ordered
  - Recursively applying nested layout
  - For intrinsic ordered topology
- Arc Diagram





Arc Diagram

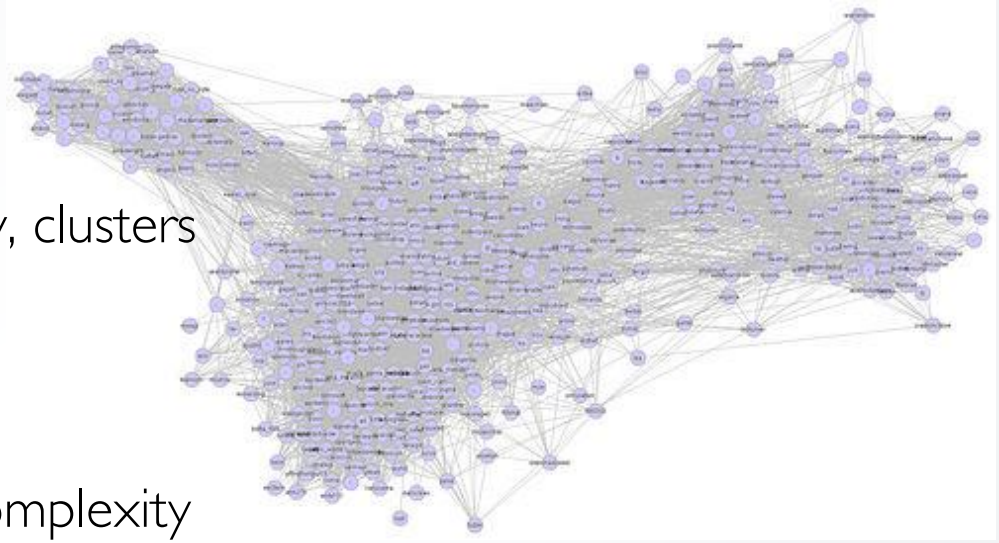


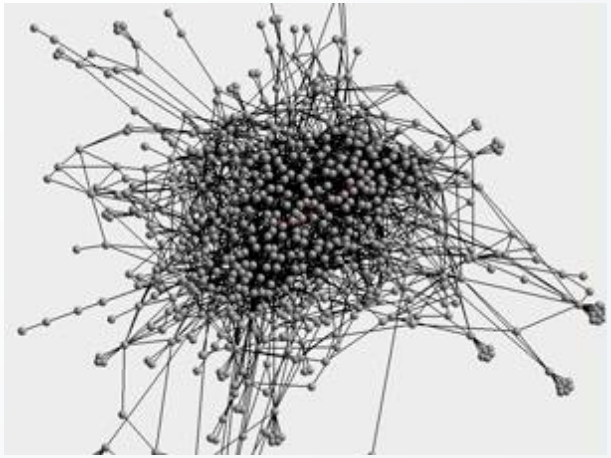
Eurozone debt web: Who owes what to whom?

http://www.bbc.co.uk/news/business-15748696

Node-Link Diagram Summary

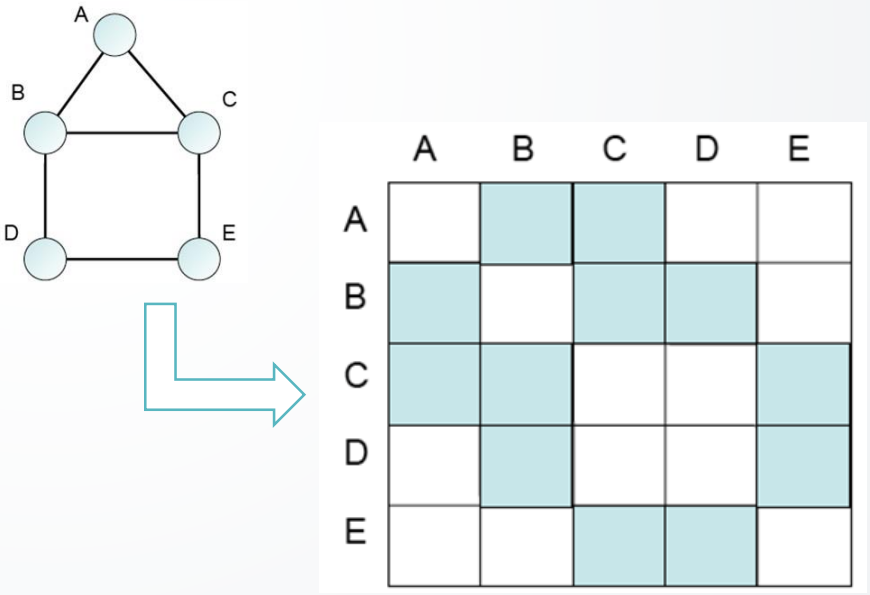
- Intuitive visual interpretation
- Good representation of topology, clusters and paths
- Flexible, many variants
- Almost for all algorithms, time complexity $> O(N^2)$
- Not so good for cluttered graphs (especially edge cluttered graphs)



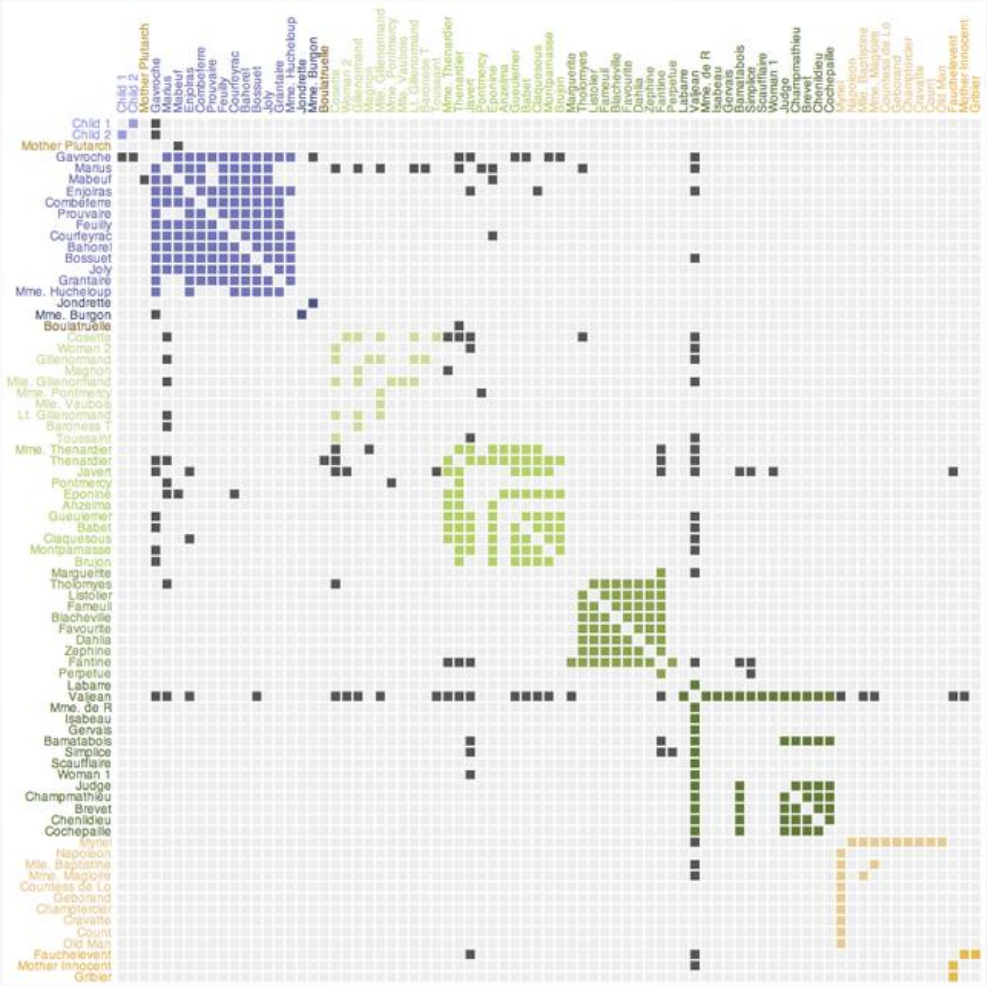


Adjacency Matrix

- $\mathbf{N} \times \mathbf{N}$ matrix, representing relations among N objects.
- Position (i, j) represents the relation between the ith and the jth object,
  - Weight
  - Direction
  - Self-reflexivity
- Related issues
  - Ordering
  - Path finding

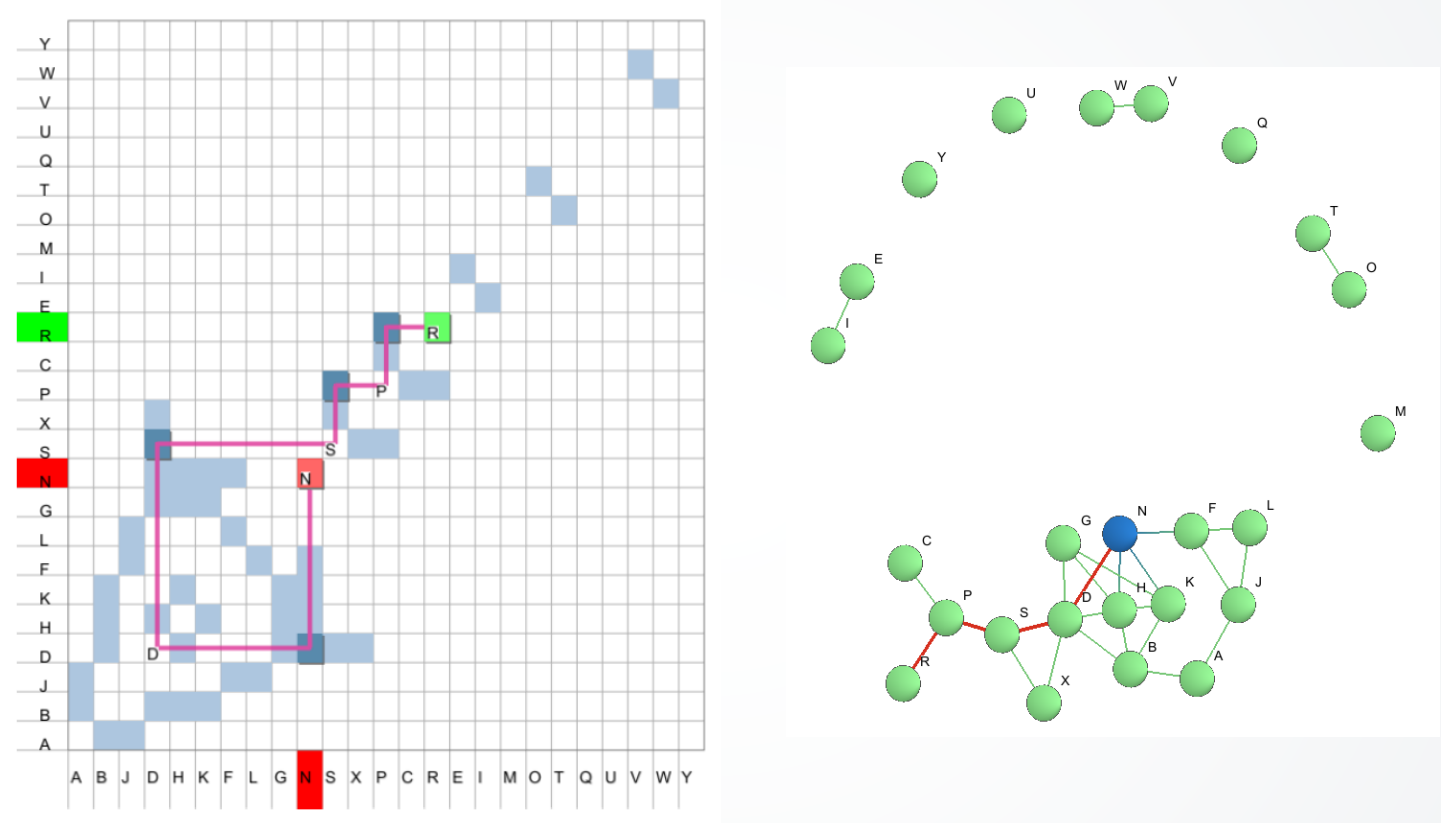


Good Ordering of Adjacency Matrix



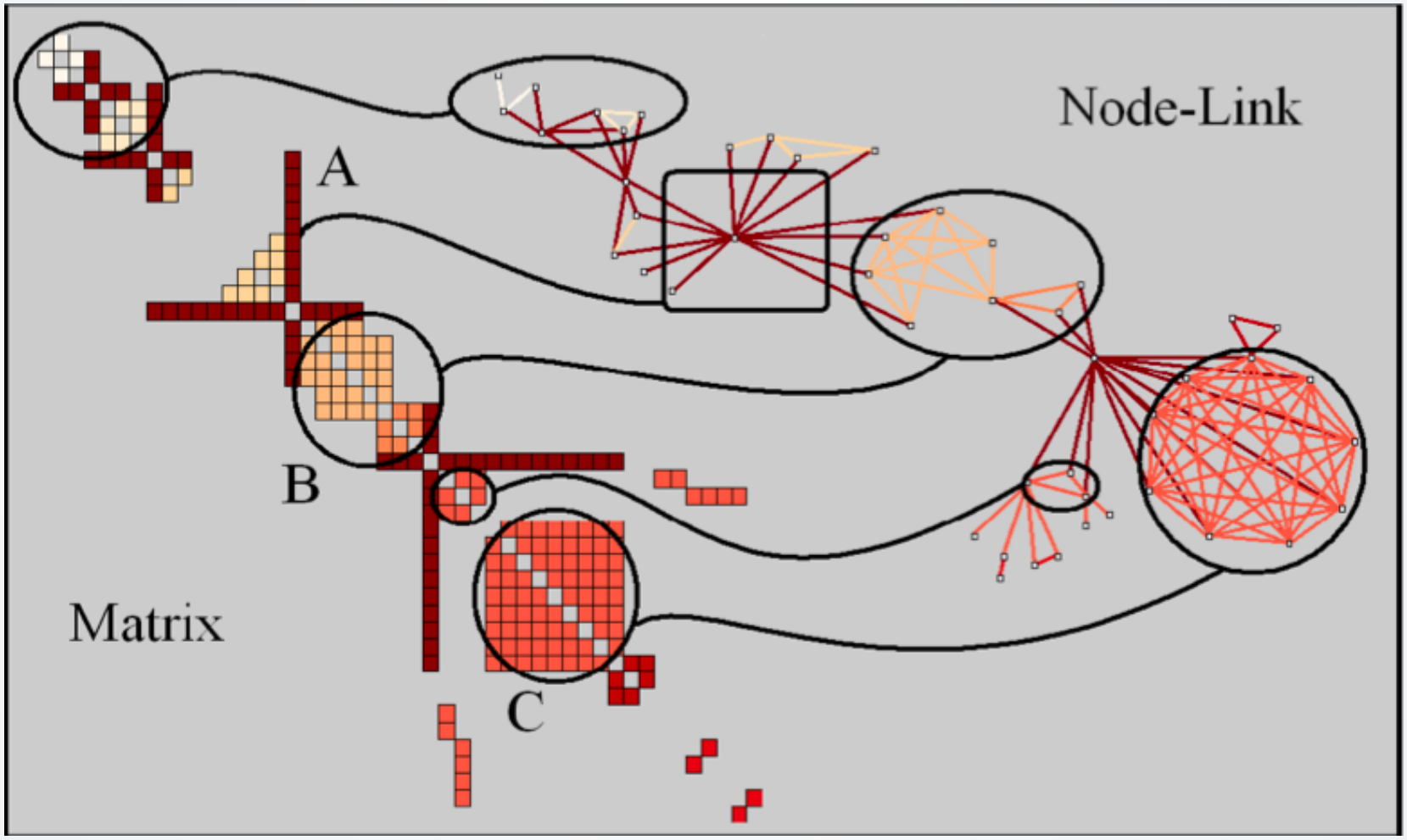
http://hci.stanford.edu/jheer/files/zoo

Path Visualization for Adjacency Matrix



Left: Connecting matrix entries and the diagonal.
Right: Path {N;D;S;P;R} is Highlighted in the Node-link Diagram.

Recognizing the Patterns of Matrix



Adjacency Matrix Summary

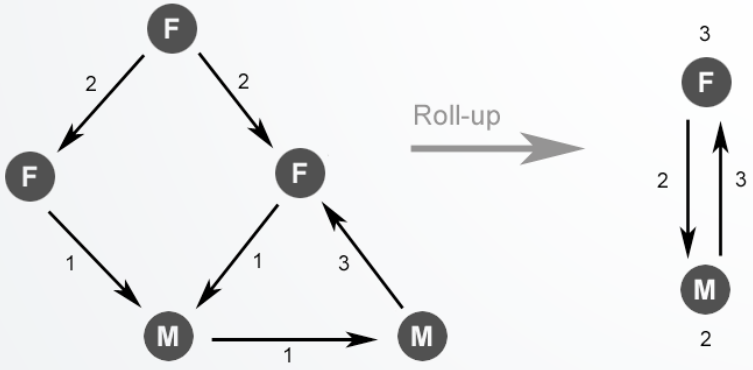
- No edge crossing, good for edge cluttered graph
- Good visual scalability
- Good presentation of graph pattern
- Visualization is too abstract to understand
- Difficult to follow a transitive relation path

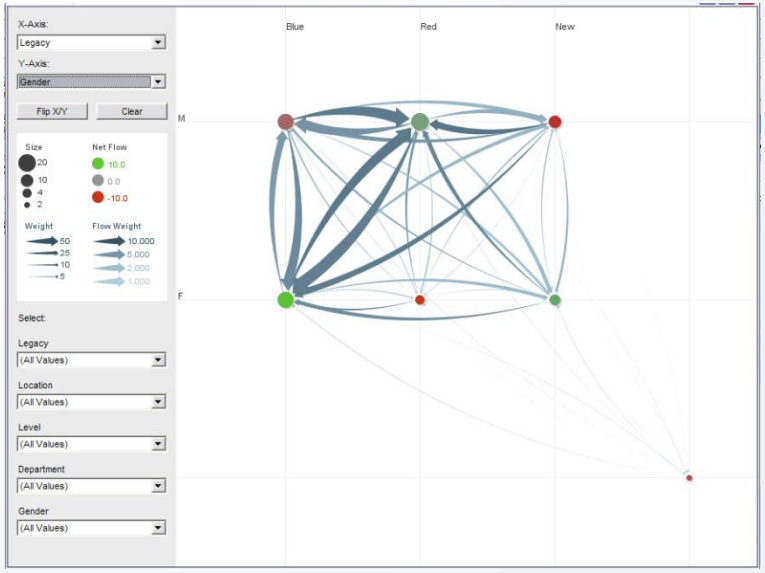
Attribute-based Graph Visualization

- Ordering based on attributes
  - Low cost
  - Highly related to applications
  - Difficult to obtain overall topology from “small multiples”
  - Sometimes you just cannot do that

PivotGraph

- Accumulate based on variables
- Can accumulate with two variables for 2D visualization
- Edge weights represent graph pattern





Hybrid Layout

Hybrid Layout

- Complex edge relations—adjacency matrix
- Too many nodes—node-link diagram
- What if there are so many nodes, and some of them with complex edge relations?

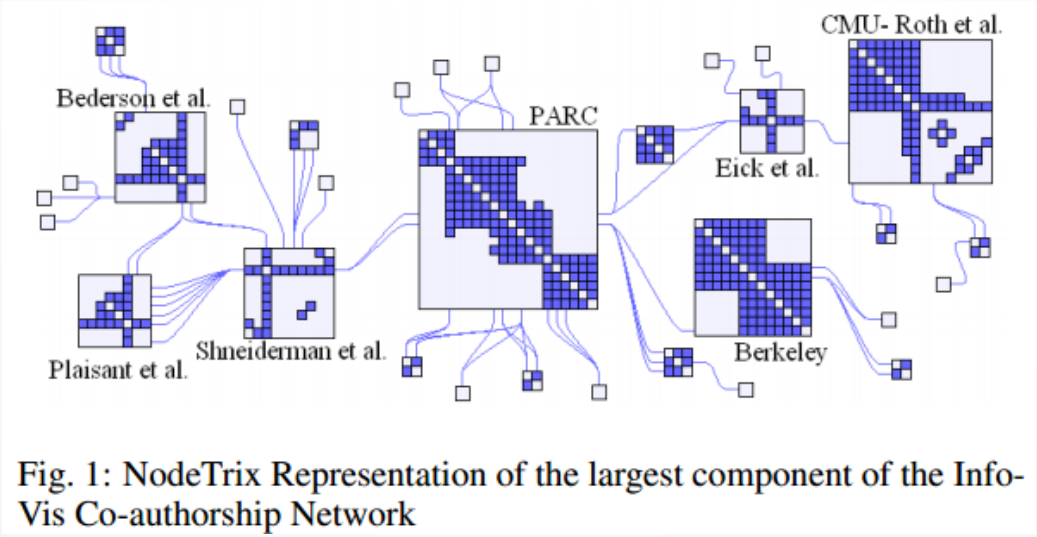


Fig. 1: NodeTrix Representation of the largest component of the Info-Vis Co-authorship Network

Nathalie Henry, et al. NodeTrix: A Hybrid Visualization of Social Networks. TVCG 2007.

Magnostics

Michael Behrisch, et al. Magnostics: Image-based Search of Interesting Matrix Views for Guided Network Exploration. IEEE VIS 2016
https://vimeo.com/182984756

Juniper

Juniper: A Tree+Table Approach to Multivariate Graph Visualization

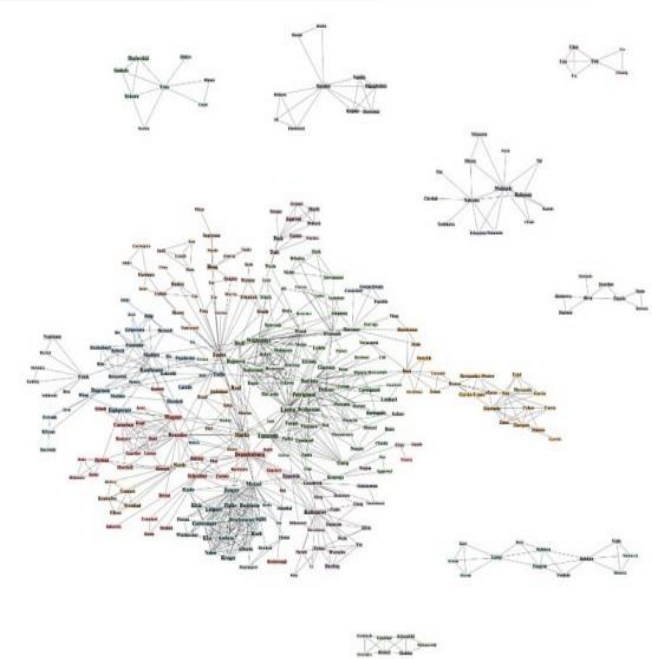
Carolina Nobre, Marc Streit, and Alexander Lex

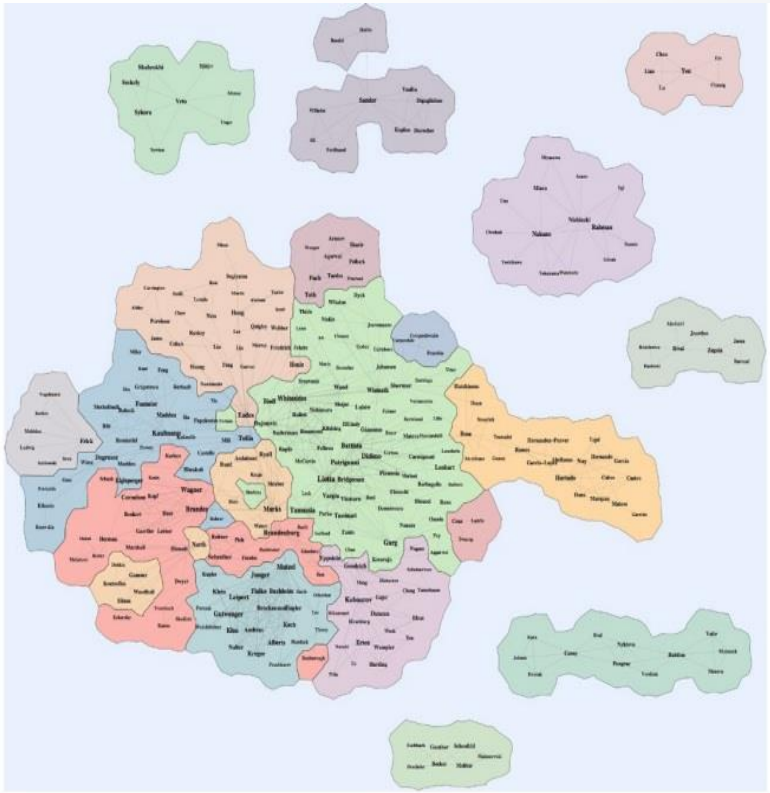
Carolina Nobre, et al. Juniper: A Tree+Table Approach to Multivariate Graph Visualization. IEEE VIS 2018
http://sci.utah.edu/~vdl/papers/2018_infovis_juniper.mp4

Other layouts

GMap

- Gmap represents object clusters with plane areas.



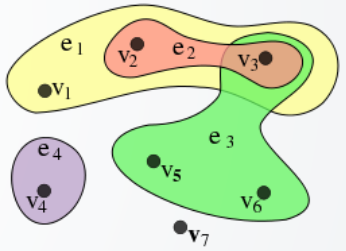


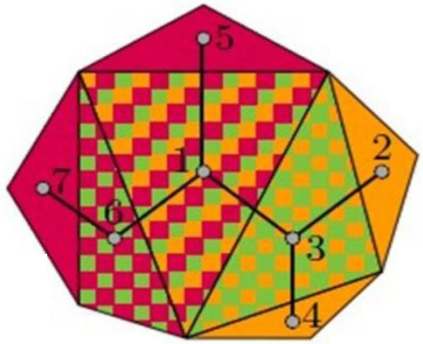
Gmap Procedures*

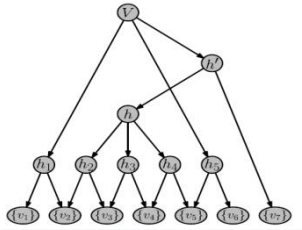
1. Layout the nodes on the graph
2. Clustering
3. Constructing voronoi diagram for every cluster
4. Coloring

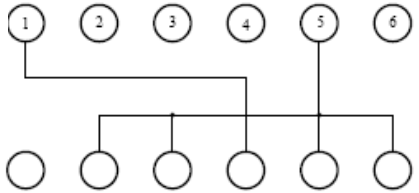
Hypergraph

- Subset system of node set
  - $X = \{x_m | m \in M\}$
  - $e_i = \{x_{k_1}, x_{k_2}, \ldots, x_{k_n}\}\ E = \{e_i | i \in I, e_i \subseteq X\}$
- Visualization of Hypergraph
  - Venn
  - Orthogonal methods
  - Hesse graph
  - Subdivision









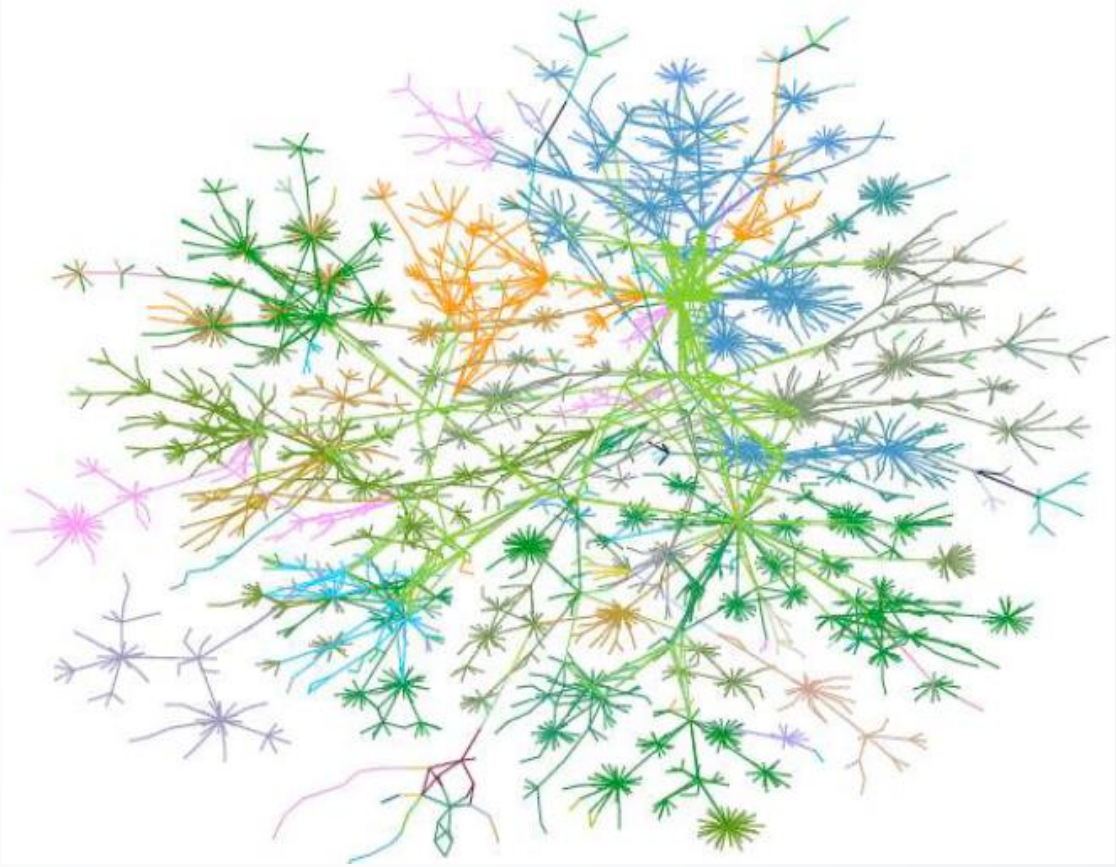
Graph Simplification

Topology Simplification

- Data reduction
  - Nodes reduction
    - Clustering
  - Edge reduction
    - Minimal spanning tree

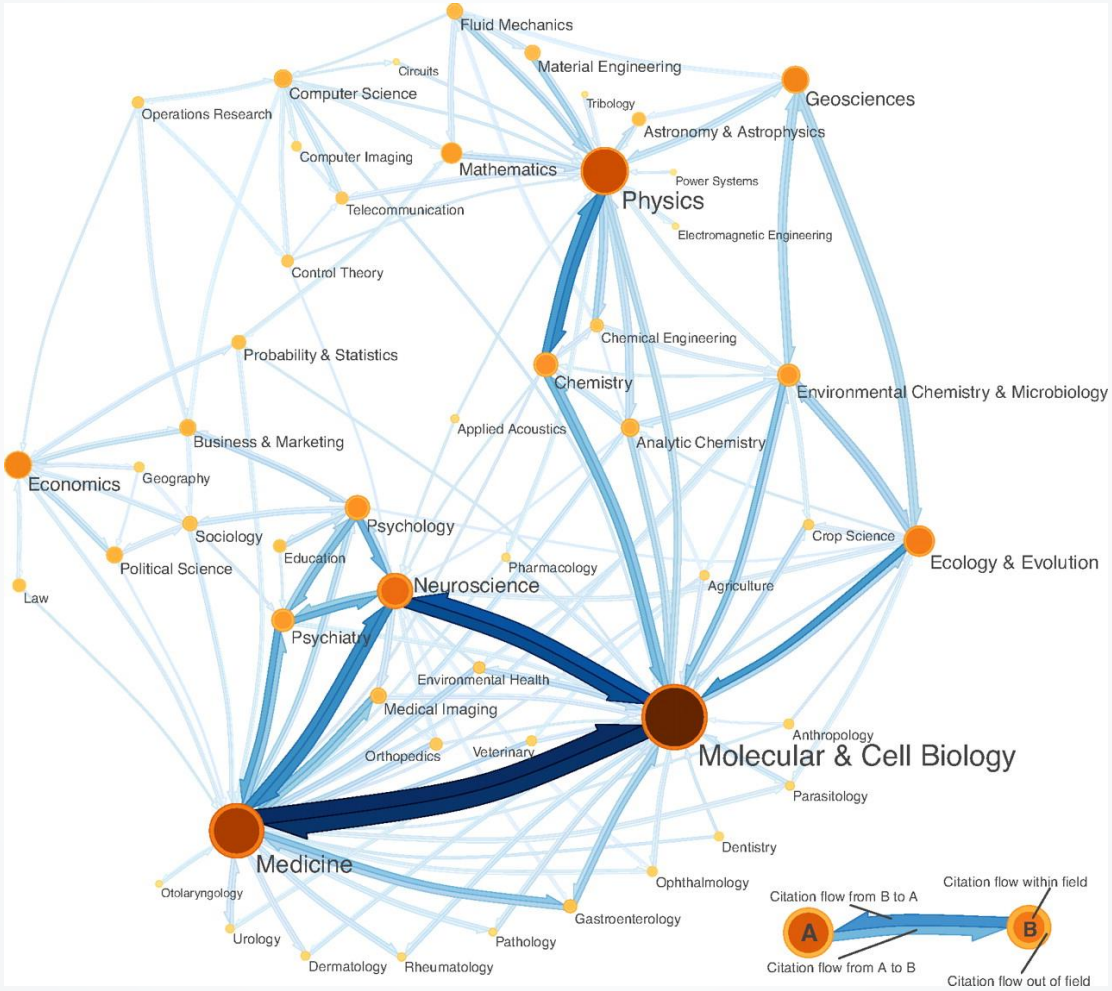
Minimal Spanning Tree

- Edge deletion
- Prim algorithm



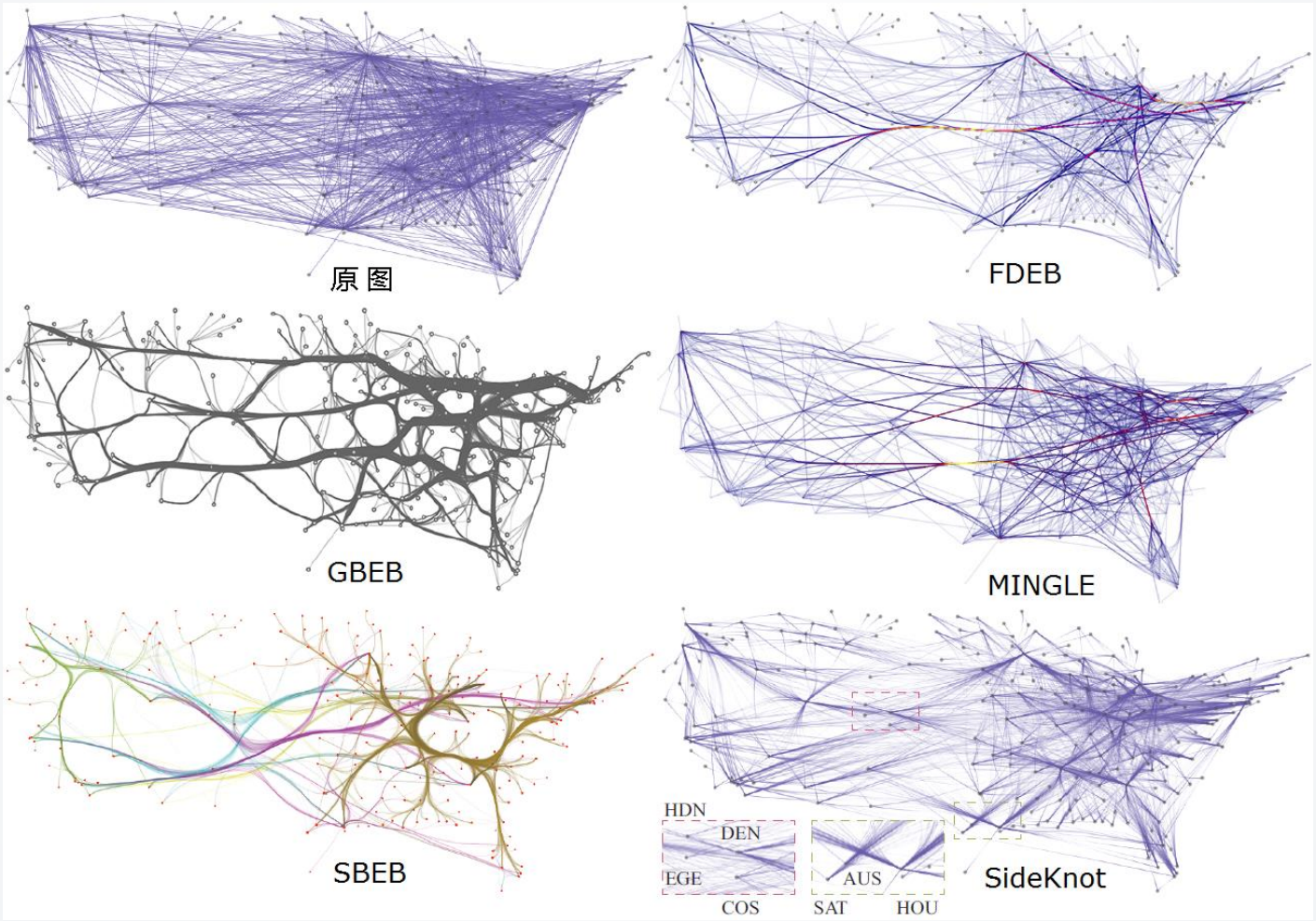
Hal Burch. Measuring an IP Network in situ

Visualization of Clusters



M. Rosvall and C.T. Bergstrom. Maps of Random Walks on Complex Networks Reveal Community Structure

Edge Bundling



KDE-based Graph Simplification

Michael Zinsmaier, et al. Interactive Level-of-Detail Rendering of Large Graphs. TVCG 2012

Edge bundling of huge graphs

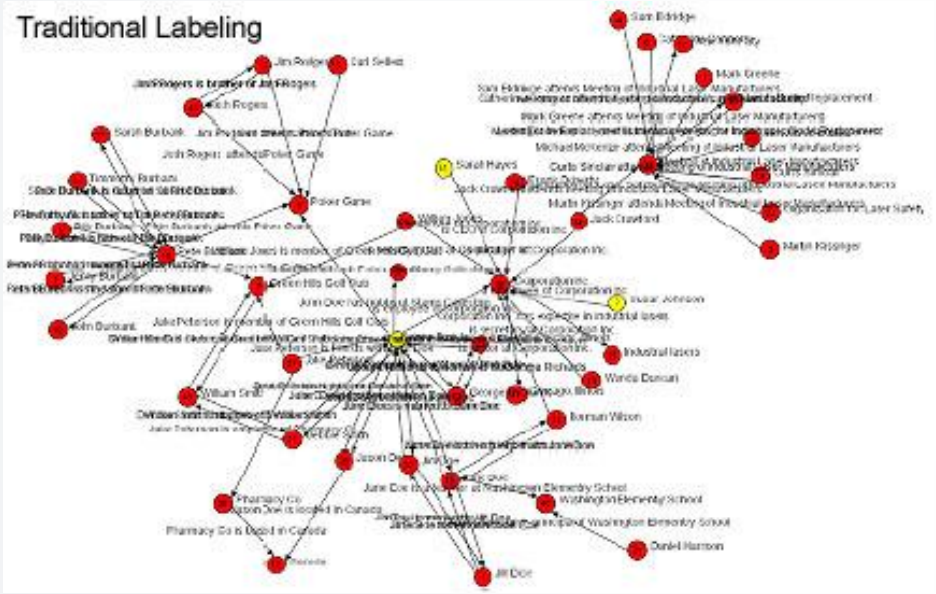
Antoine Lhuillier, et al. FFTEB: Edge bundling of huge graphs by the Fast Fourier Transform. PacificVis 2017

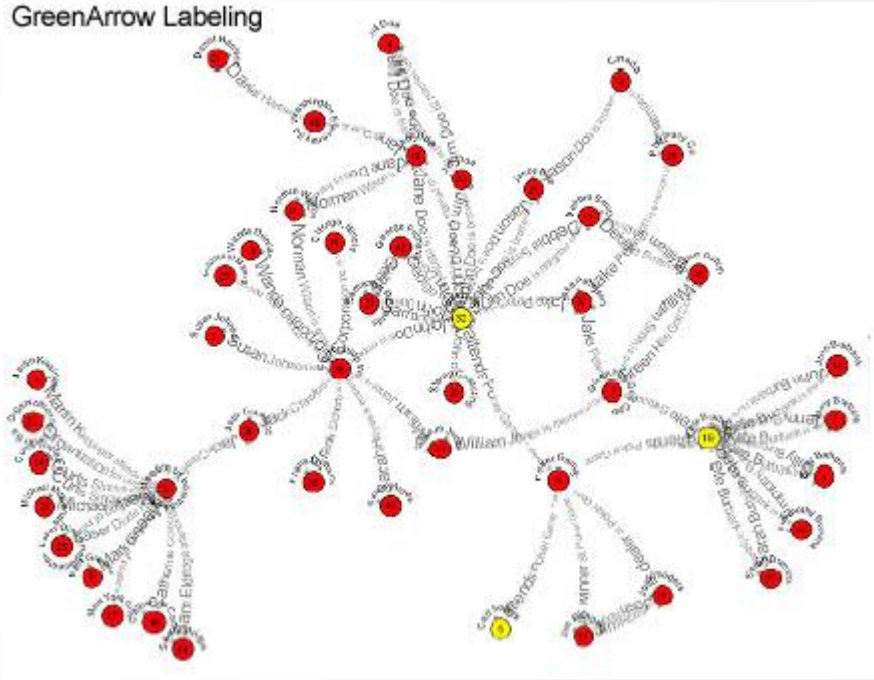
Interaction

Labeling

- Non-text labels
  - Color, size, shape, thickness
  - Too many labels lead to misinterpretation
- Labels of nodes
  - Inside the node object
- Labels of edge
  - Along the edge

Labeling





Labels are also nodes and edges

Interactions

- View-based interaction
  - Panning, zooming, rotating
- Object-based interaction
  - Selecting, highlighting, deleting, dragging
- Structure-based interaction
  - Reorder, re-layout
  - Focus + context
  - Roll-up, drill-down

Large Scale Graph Interaction

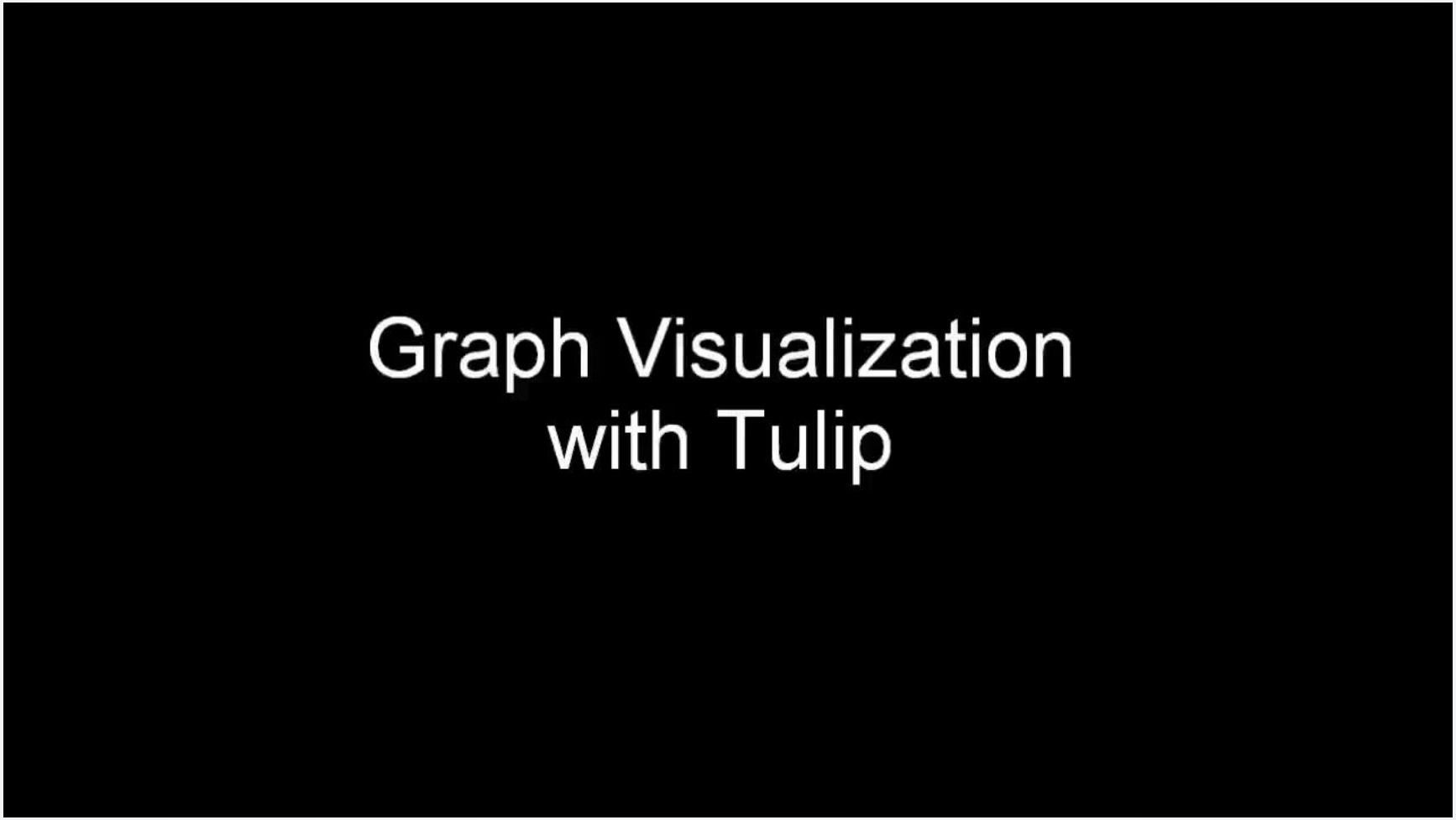
Matthew van der Zwan, et al. CUBu: Universal real-time bundling for large graphs. IEEE VIS 2016

Large Scale Graph Interaction

Chen Wei, et al. Structure-Based Suggestive Exploration: A New Approach for Effective Exploration of Large Networks. IEEE VIS 2018

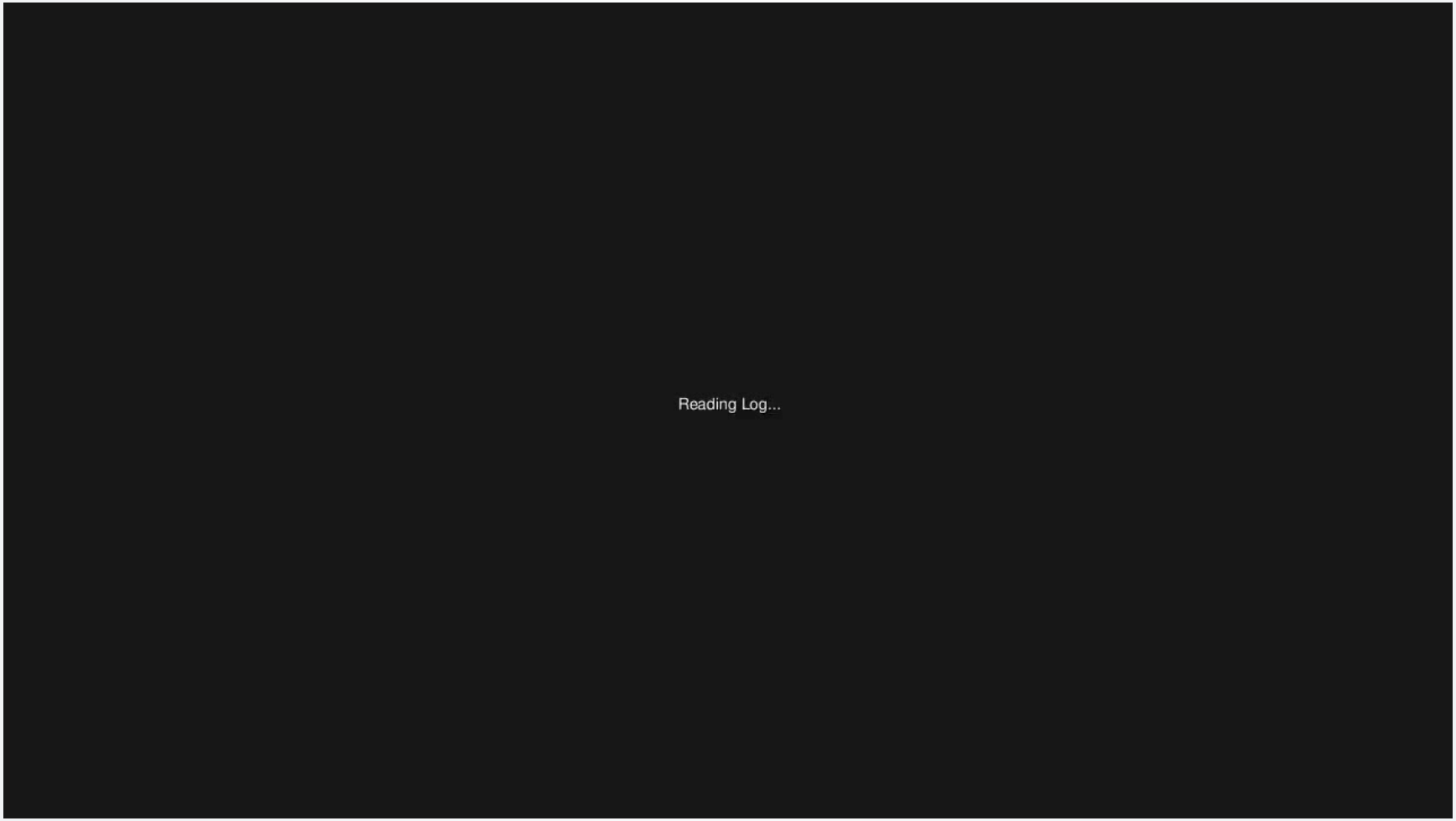
Tools and Applications

Tulip



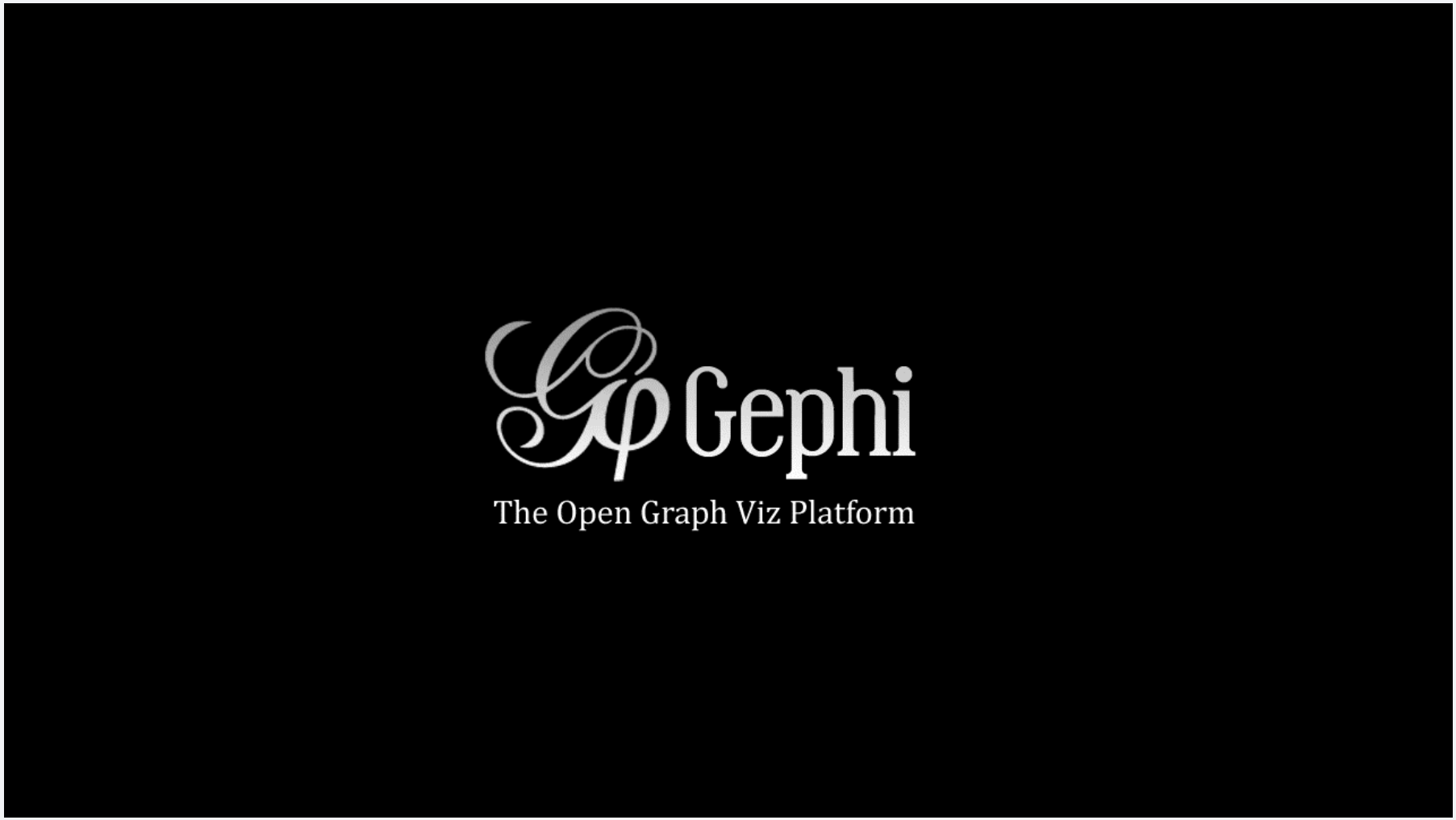
http://tulip.labri.fr/TulipDrupal/

Gource Software Version Control Visualization



Git one minute

Gephi

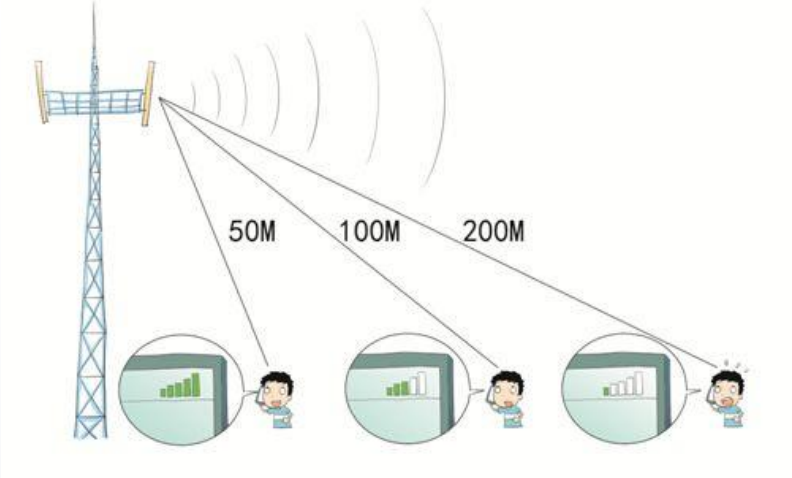


https://gephi.org/

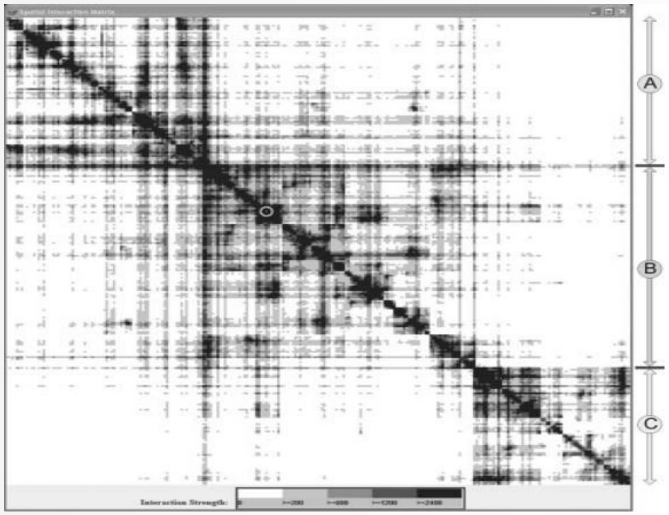
Case:
OD Network Visualization

Origin-Destination Movement Data

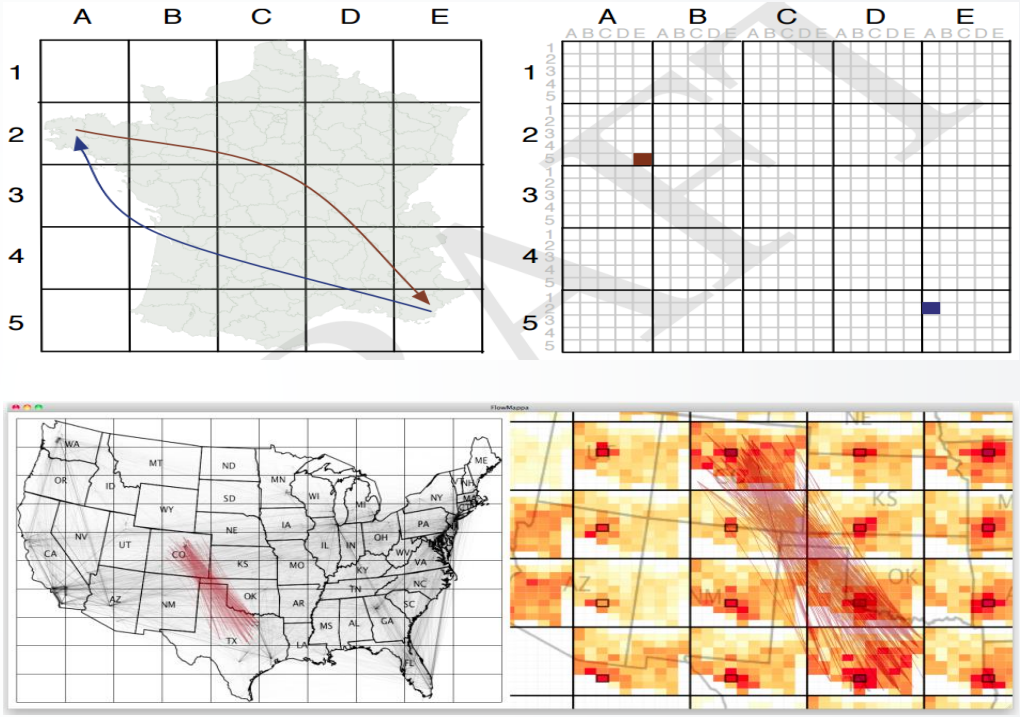
- Taxi trips
- Traffic camera records
- Bicycle sharing systems
- Mobile phone locations



Origin-Destination Data Visualization

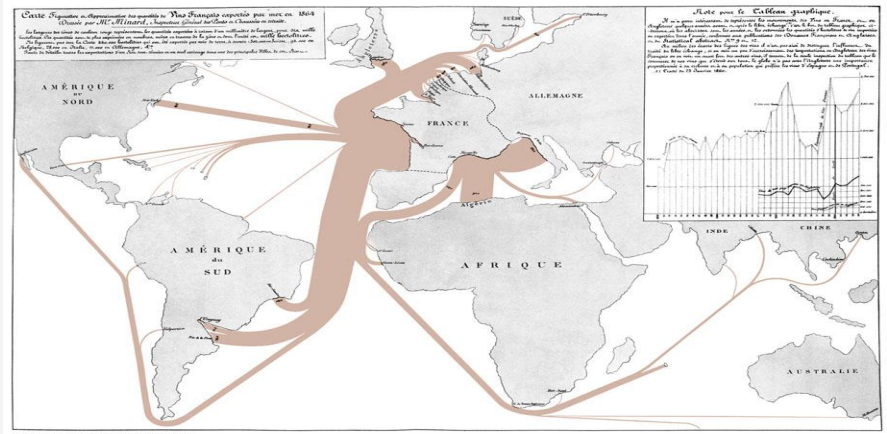


Visual Analytics of Spatial Interaction Patterns. D. Guo, et al. International Journal of Geographical Information Science, 2007



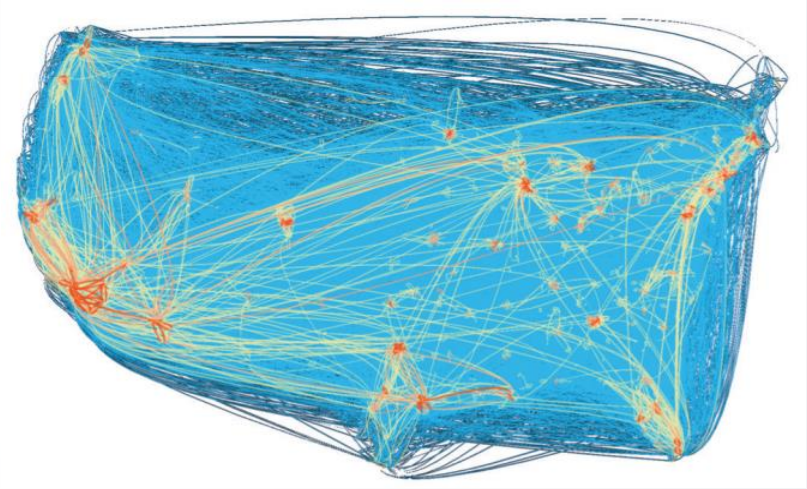
Visualisation with OD Maps.
J. Wood, et al. The Cartographic Journal, 2010.

Origin-Destination Data Visualization



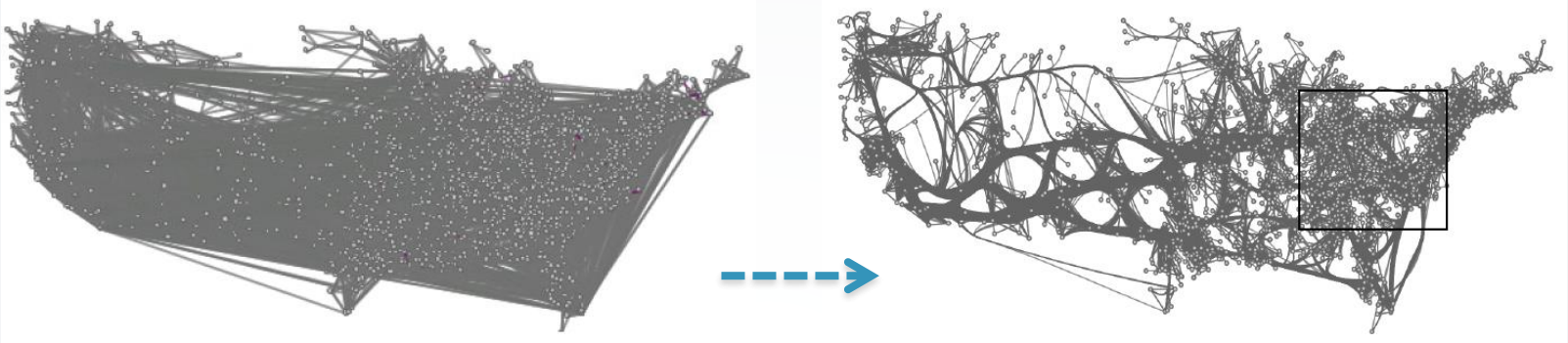
Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard*, 1845–1869, a portfolio of his work held by the Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris.

Flow map

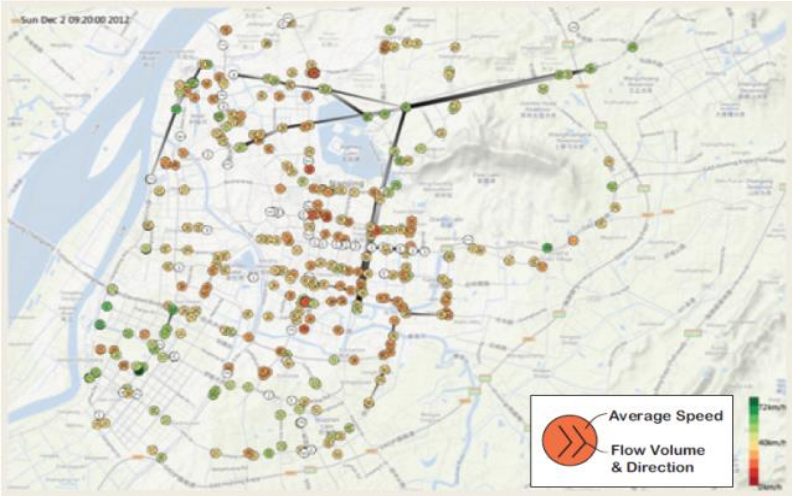


OD Flow Data Smoothing and Mapping[J]. Guo D, et al. TVCG, 2014.

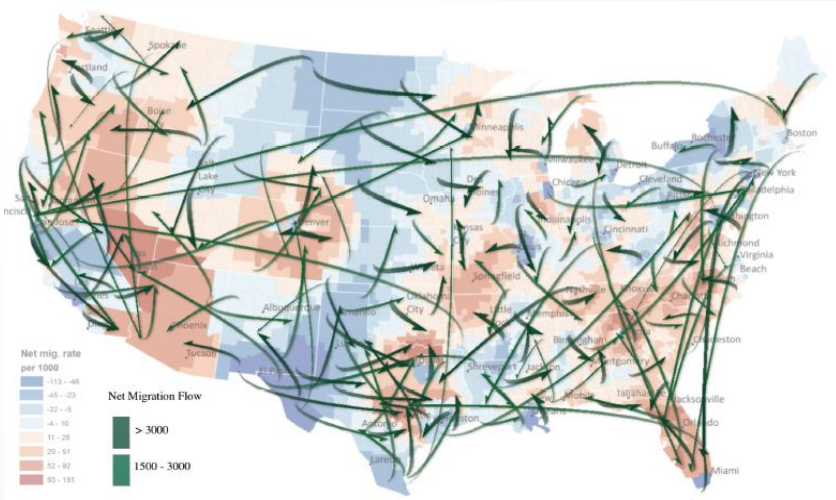
Simplified Flow Map



Geometry-Based Edge **Clustering**, W W Cui, et al. TVCG, 2008



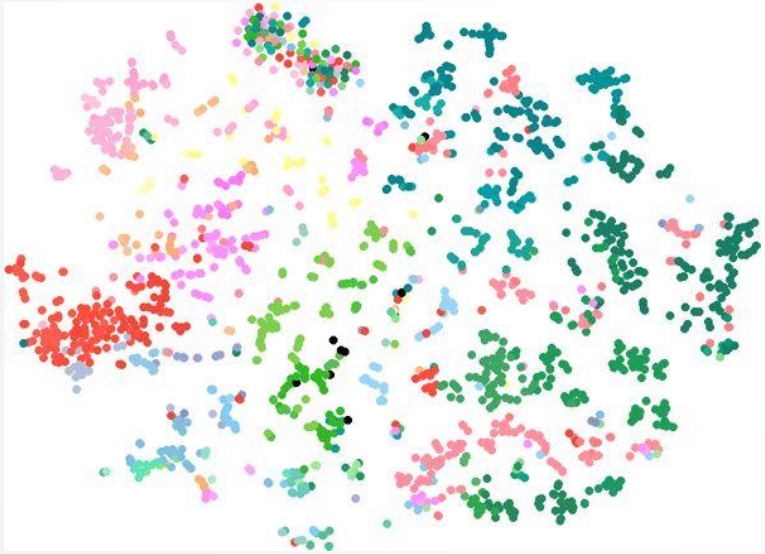
Sparse Traffic Trajectory Data **Filtering**,
Wang et al. TVCG, 2014



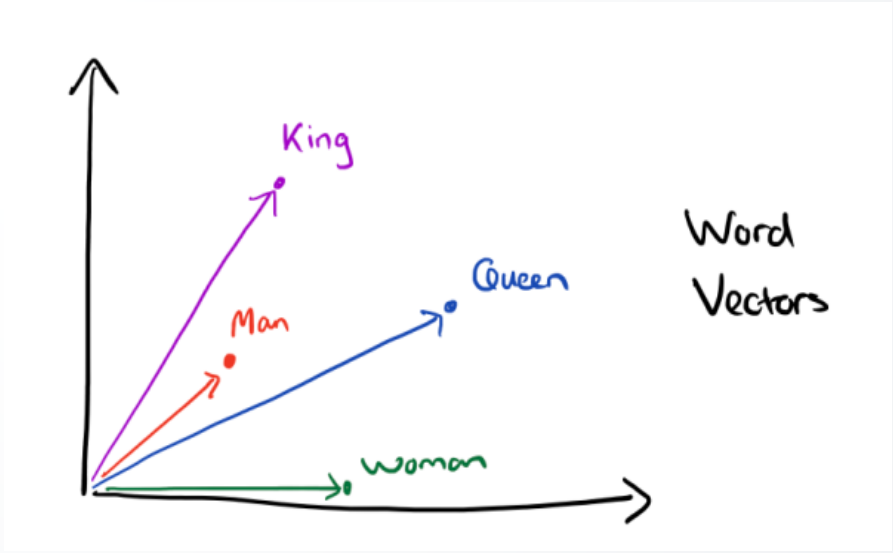
Origin-Destination Flow Data **Sampling**,
Guo et al. TVCG, 2014

Word2Vec

- created by a team of researchers led by Tomas Mikolov at Google in 2013.
- takes a large corpus of text as input and produces a vector space

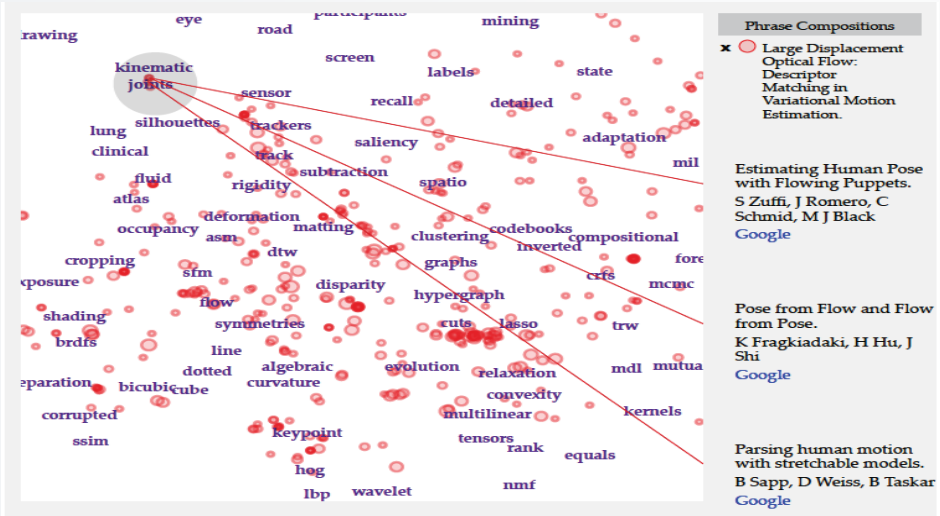


common contexts = close locations

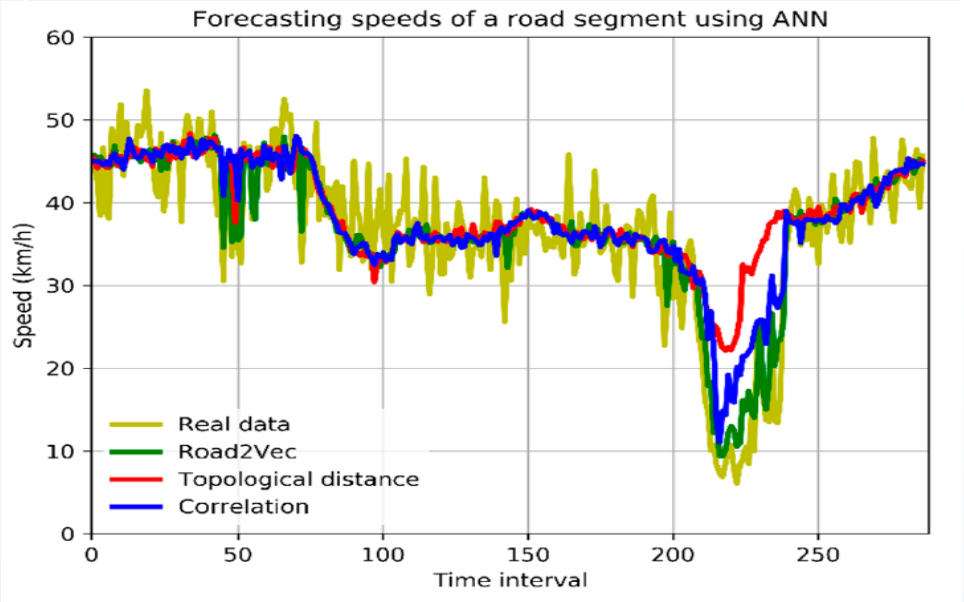


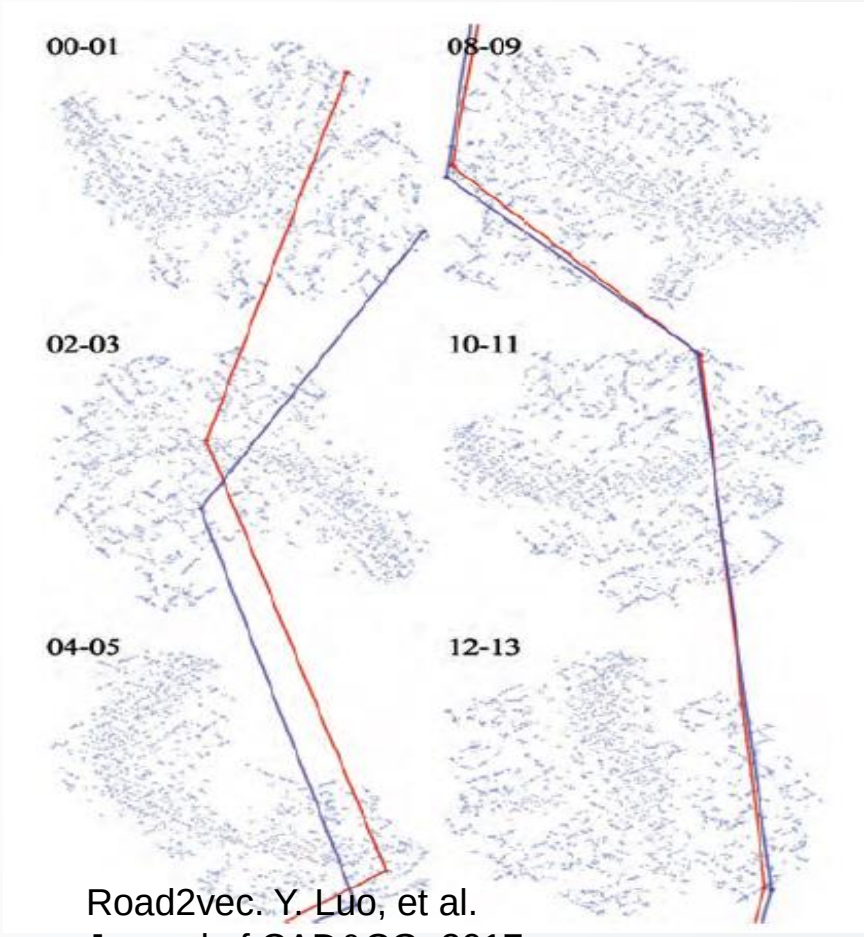
King – Man + Woman = Queen

Word2Vec applied in other fields



Cite2vec. M. Berger, et al. TVCG, 2017



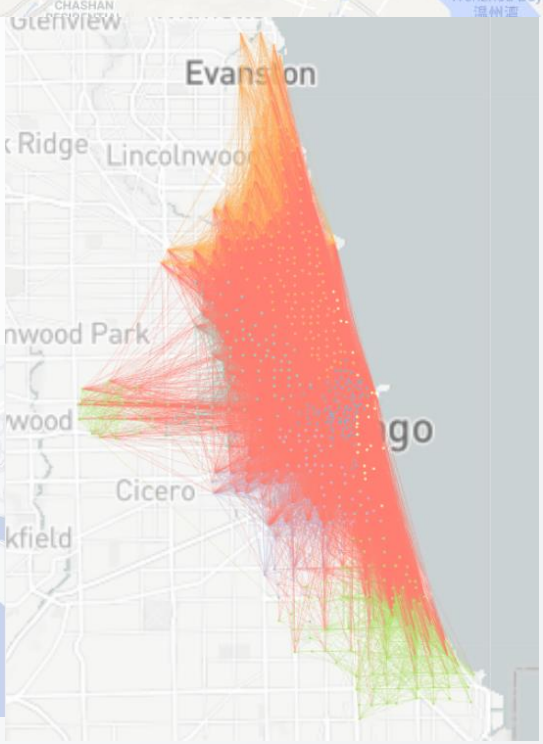


Road2vec. Y. Luo, et al. Journal of CAD&CG, 2017

Road2vec. K. Liu, et al. International Journal of Geo-Information, 2017.

DataSet

- Mobile phone locations
  - 2,755 cell stations
  - the span of a week
  - 1.1 millions of persons
  - 1.3 millions of OD trajectories
- Bicycle sharing systems
  - 580 stations and 5,800 bikes
  - January 2013 to December 2016
  - 10 millions of usage records

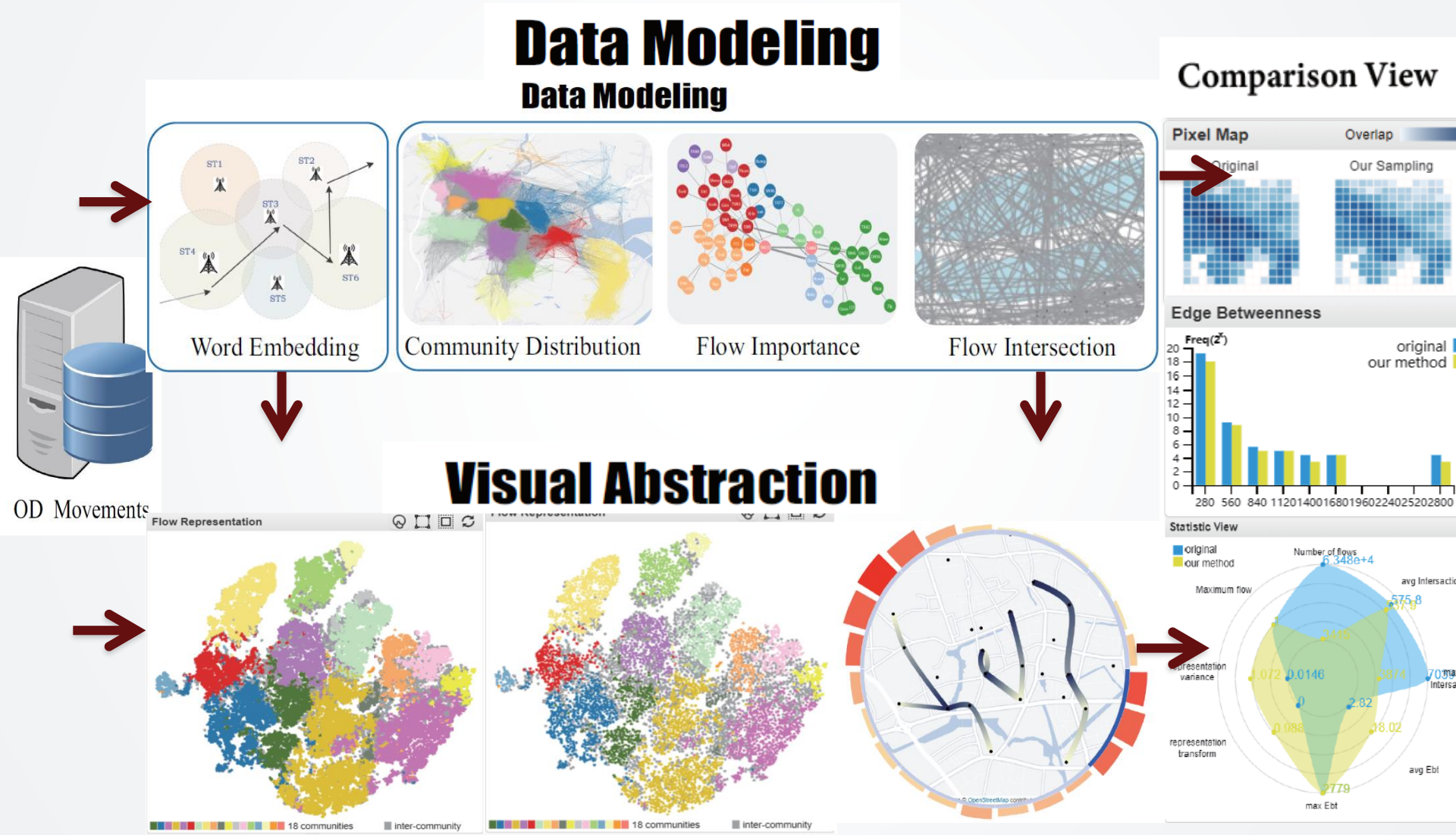


Task Analysis

-

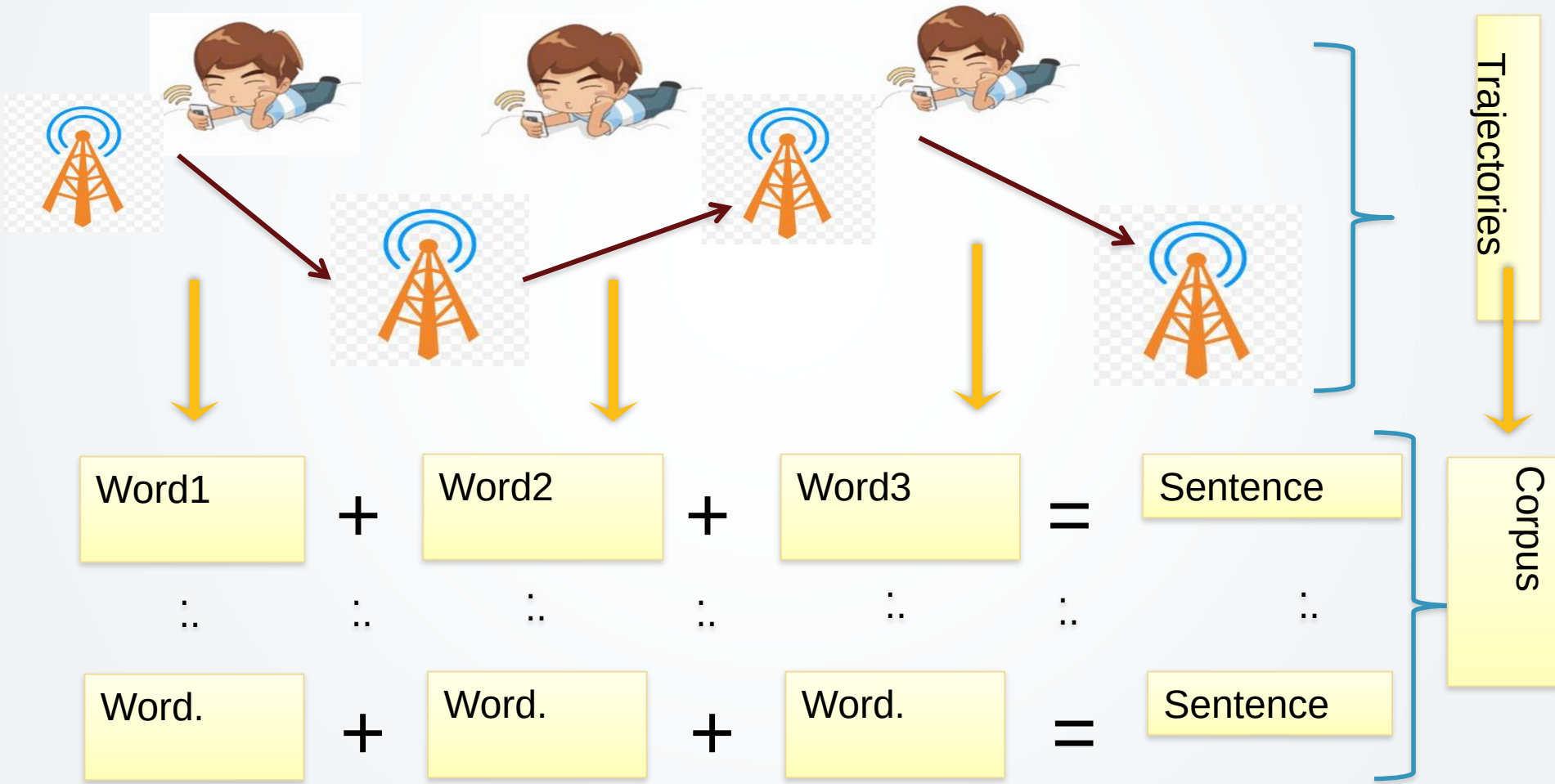
Simplification of flow map	Characterization of OD flows	Visualization of flow abstraction	Evaluation of flow abstraction
Scale reduction	Semantic relationship	Visual designs	Visual perception
Relationship preserving	Representation learning	User interactions	Feature statistics

Workflow


Data Modeling
Data Modeling
Word Embedding
Community Distribution
Flow Importance
Flow Intersection
OD Movements
Visual Abstraction
Flow Representation
18 communities
inter-community
18 communities
inter-community
Comparison View
Pixel Map
Overlap
Original
Our Sampling
Edge Betweenness
Freq(z)
original
our method
Statistic View
original
our method
Number of flows
Maximum flow
representation variance
representation transform
avg Ebt
max Ebt

Corpus generation

- OD flows vs Natural Language Processing (NLP) terms



Word2Vec

- Skip-gram model

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0} logp(w_{t+j}|w_t)$$

- Softmax function

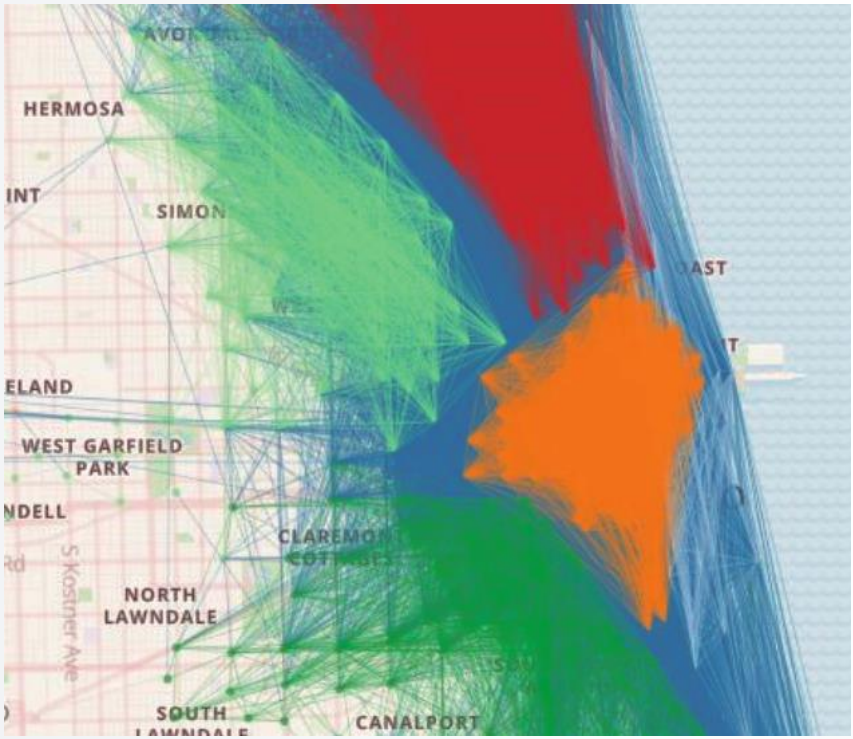
$$p(wO|wI) = \frac{exp({v'_{wO}}^T v_{wI})}{\sum_{w=1}^{W} exp({v'_W}^T v_{wI})}$$

- Negative sampling

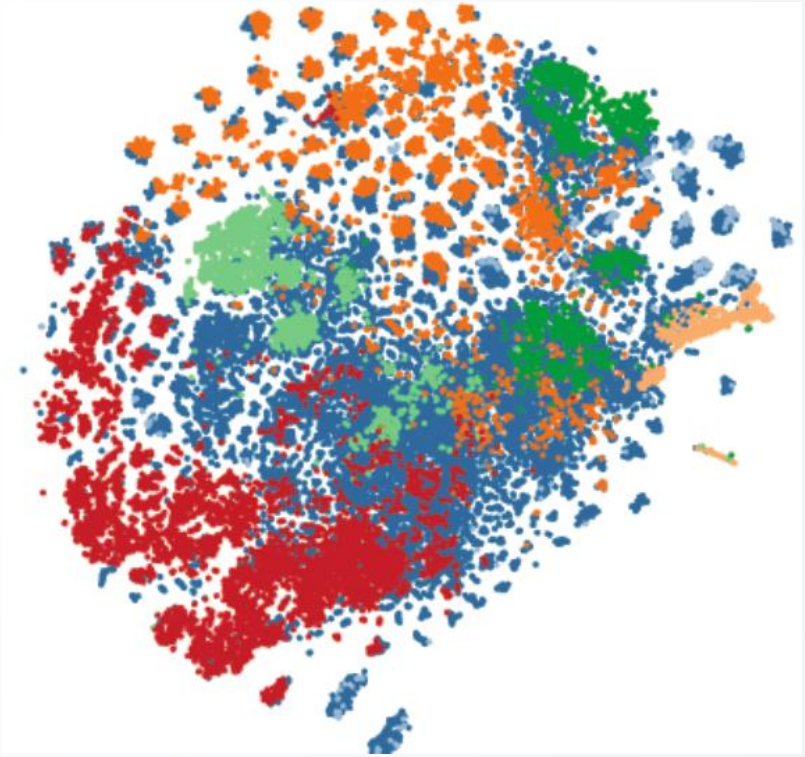
$$log\sigma({v'_{wO}}^T v_{wI}) + \sum_{i=1}^{k} E_{wi\sim P_n(w)}[log\sigma(-{v'_{wi}}^T v_{wI})]$$

Dimension reduction

- t-Distributed Stochastic Neighbor Embedding (t-SNE)

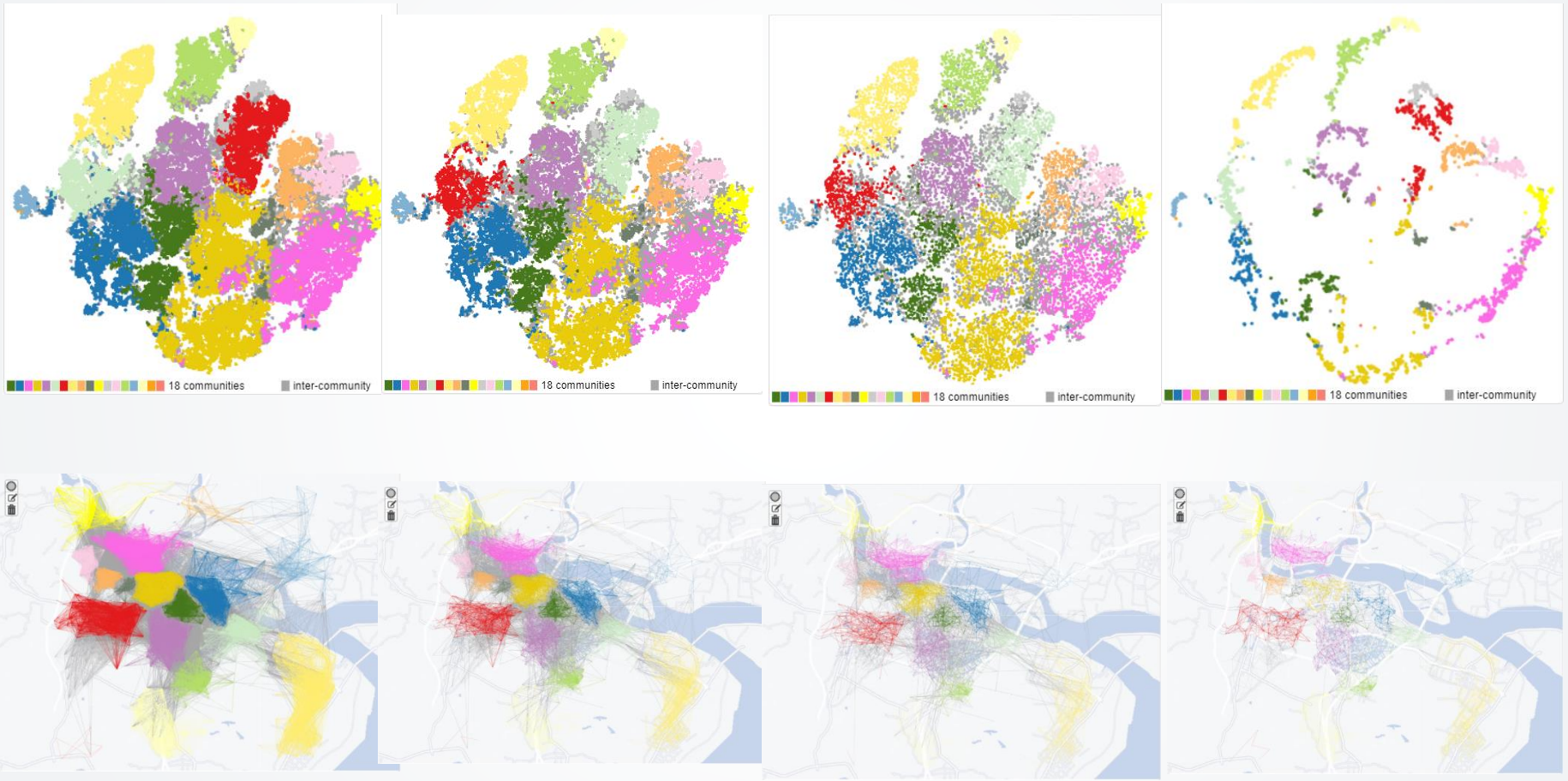


Flow map



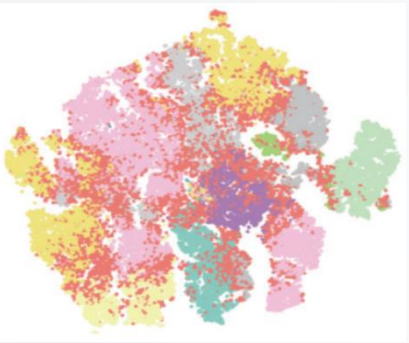
Vectorized representations

Adaptive blue noise sampling



Original data | Sampling rate: 20% | Sampling rate: 10% | Filtering rate: 10%

Iterative multi-objective sampling



Blue Noise Sampling

Original data

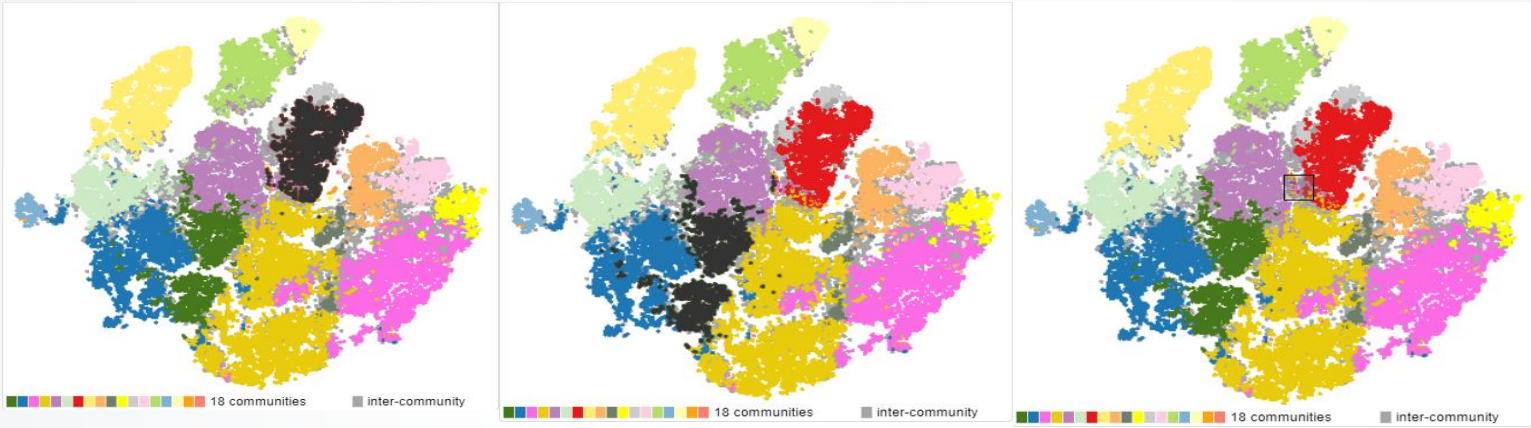
Abstration

Flow intersection	Flow importance	Community distribution

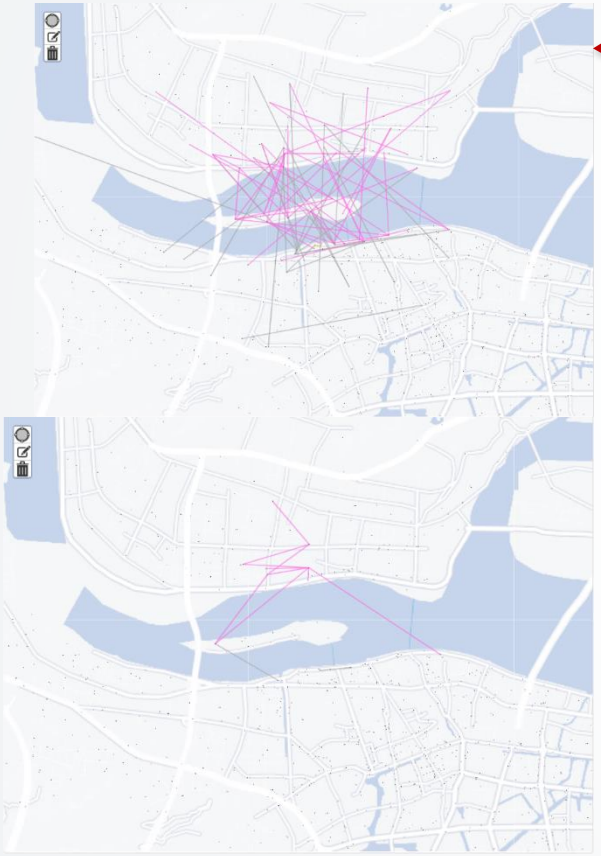
Multiple objectives

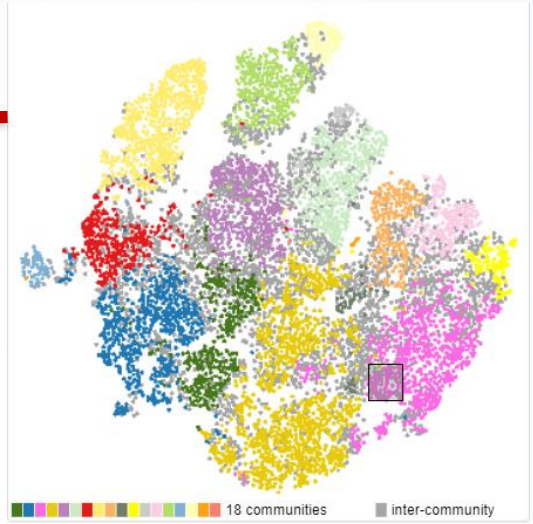
Visual Interface

Case 1: Vectorized Representation

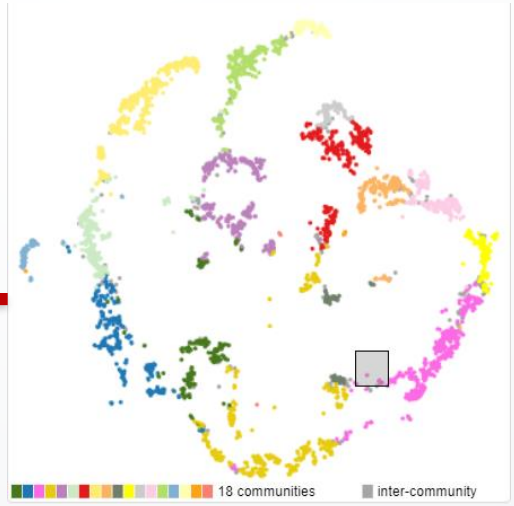


Case 2: Adaptive Blue Noise Sampling



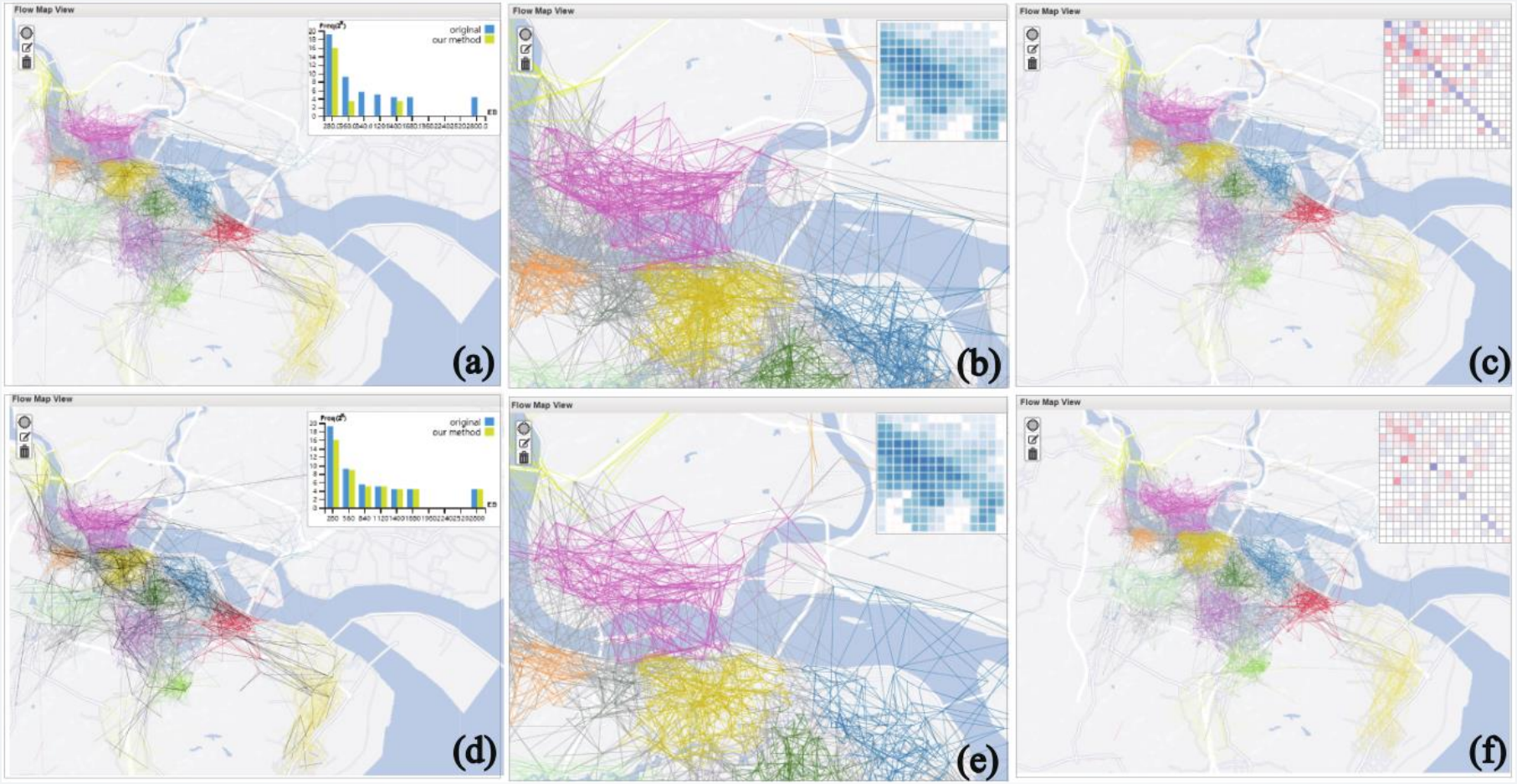


Sampling rate: 10%



Filtering rate: 10%

Case 3: Multi-objective Sampling

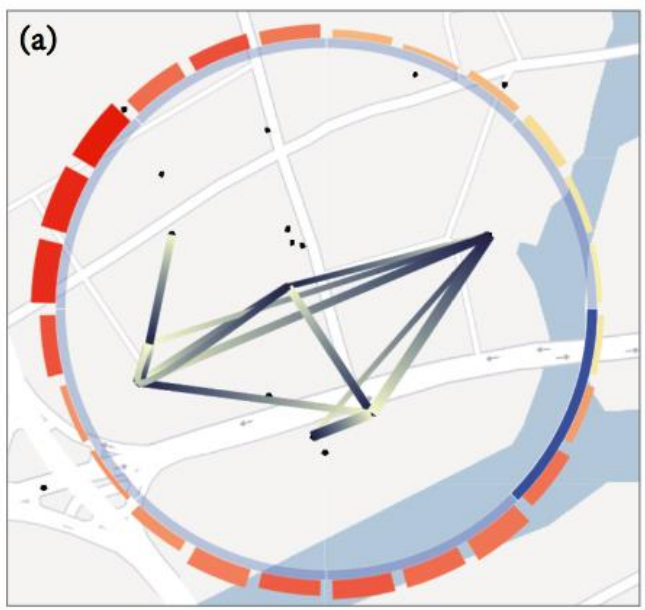


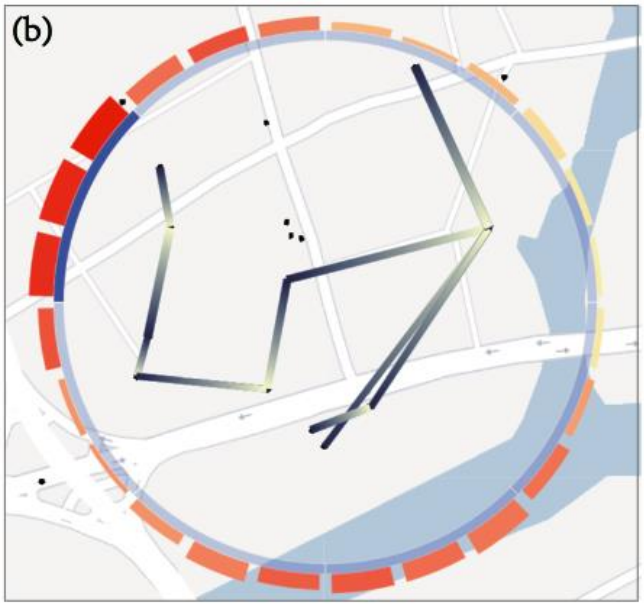
edge betweenness

flow intesections

community distribution

Case 4: Interactive Flow Wheel





6:00 am to 9:00am

18:00 pm to 21:00 pm

Quantitative Comparison

OBJ	Rate	AOF	AFI	MFI	AEB	MEB	DCD
ORI	100%	63476	575.78	7039	2.82	2779.00	0.000
FI	40%	26678	528.28	3874	2.94	2779.00	0.043
EB	40%	26739	599.09	7039	5.56	2779.00	0.033
CD	40%	26850	599.13	5519	3.01	2755.33	0.019
ALL	40%	26697	548.57	4357	4.89	2779.00	0.017
FI	20%	15545	471.31	3408	3.10	2755.33	0.071
EB	20%	15531	604.88	6926	9.02	2779.00	0.044
CD	20%	15476	608.96	6926	3.05	2755.33	0.032
ALL	20%	15500	541.06	3844	7.32	2779.00	0.051
FI	10%	6207	359.85	3408	2.61	1211.97	0.131
EB	10%	6258	607.12	4851	20.93	2779.00	0.066
CD	10%	6172	599.58	4860	3.02	1169.29	0.021
ALL	10%	6241	512.82	3626	13.22	1500.92	0.245
FI	5%	3496	293.85	3051	2.83	2755.33	0.176
EB	5%	3426	609.94	3962	34.21	2779.00	0.090
CD	5%	3338	607.35	3588	3.26	1169.29	0.026
ALL	5%	3392	538.01	3874	20.40	2779.00	0.278

FI : flow importance

EB : edge betweenness

CD : community distribution

AFI : average flow intersections

MFI : maximum of flow intersections

AEB : average edge betweenness

MEB : maximum of edge betweenness

DCD : difference between community distributions

Discussion

- Some advantages of Word2Vec model are not well realized in our visualization system.
- It is time-consuming to achieve a global optimization with more constraints integrated into the sampling model.
- Uncertainty will be generated in the simplified flow map, which might disturb the visual exploration of human mobility patterns.