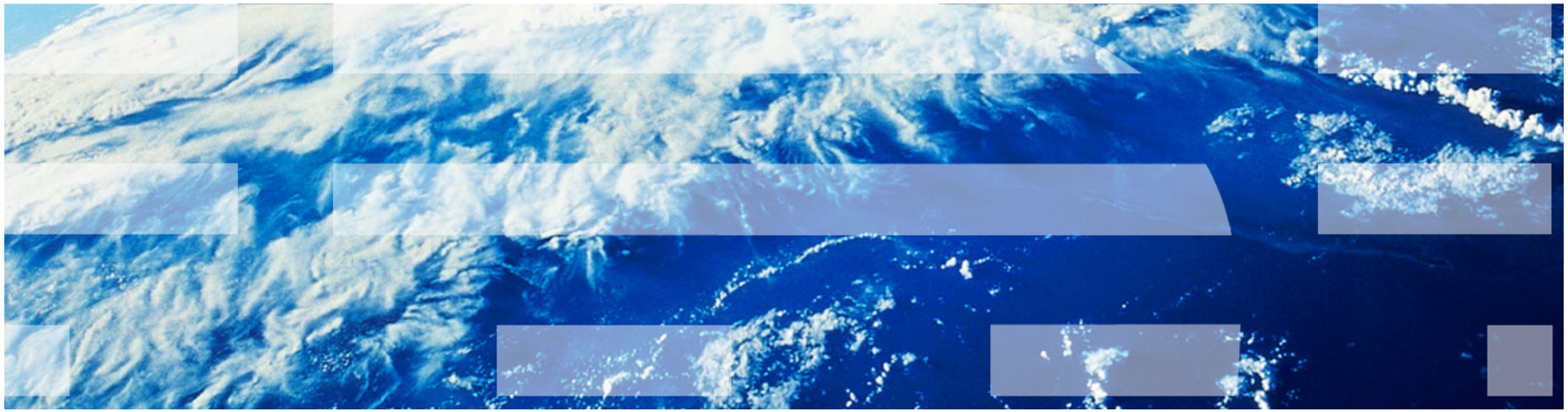
Columbia University

E6895 Advanced Big Data Analytics: Lecture 14

Big Data Visualization

Ching-Yung Lin, Ph.D.

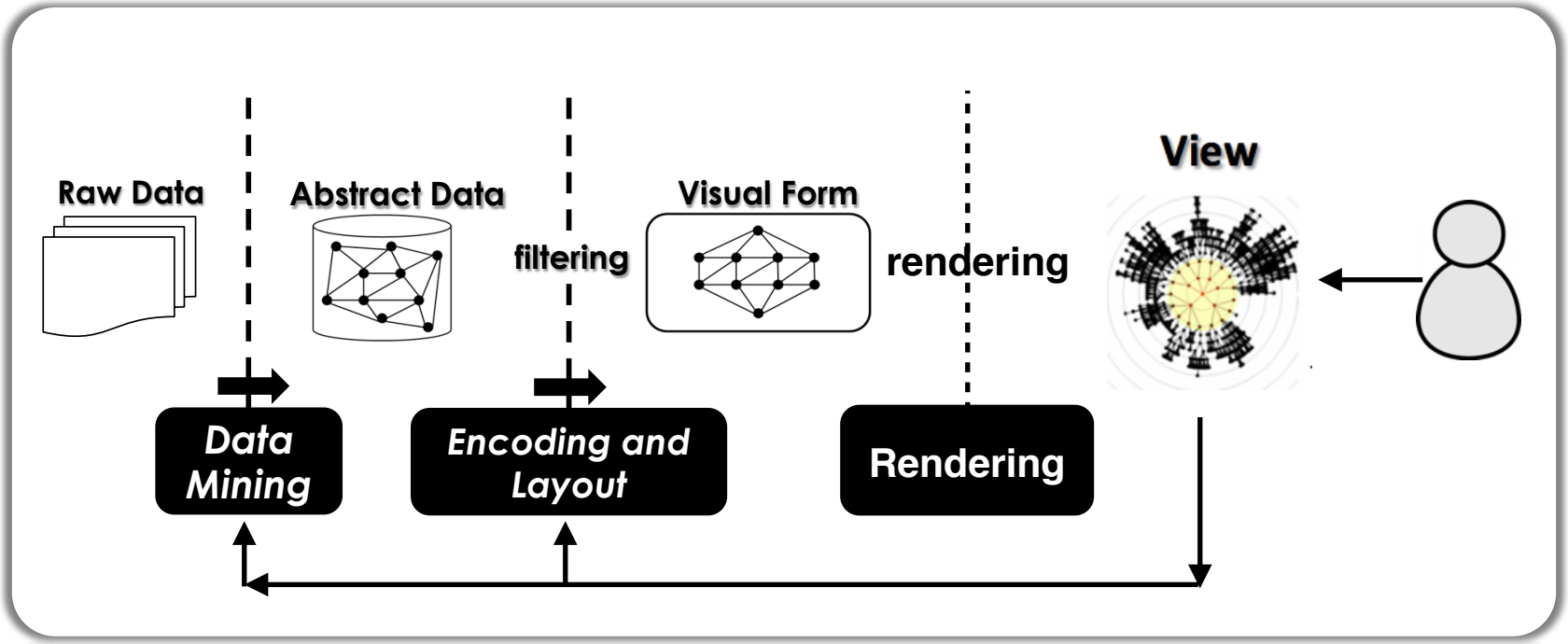
Adjunct Professor, Dept. of Electrical Engineering and Computer Science



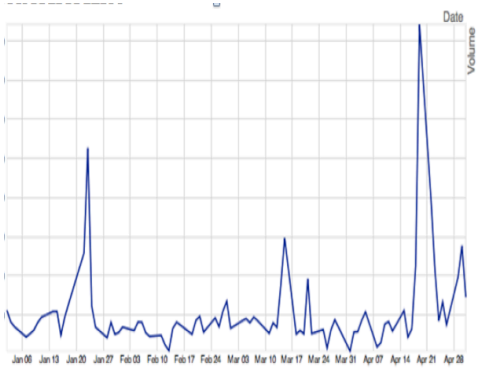
April 28th, 2016

Existing Visualisation Techniques

Visualization & Visual Analysis Reference Model



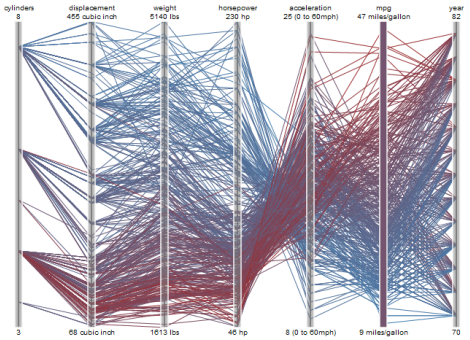
Taxonomy by data types



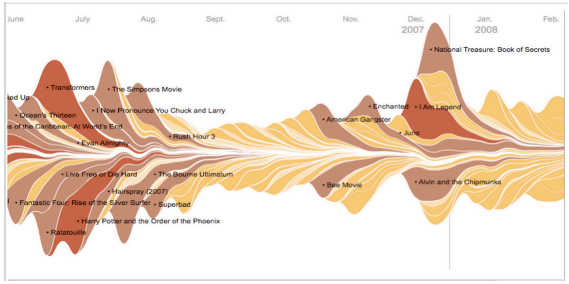
1D

2D

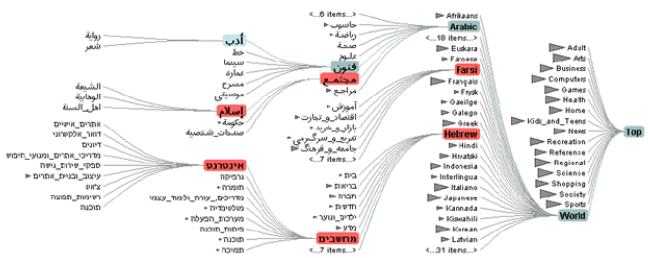
3D



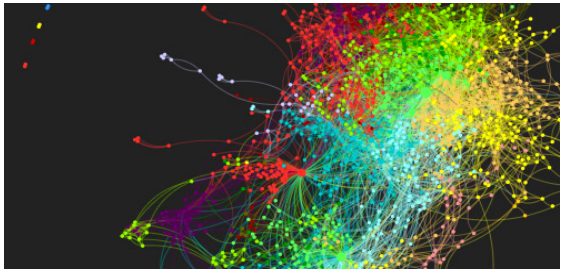
Multi-D



Temporal

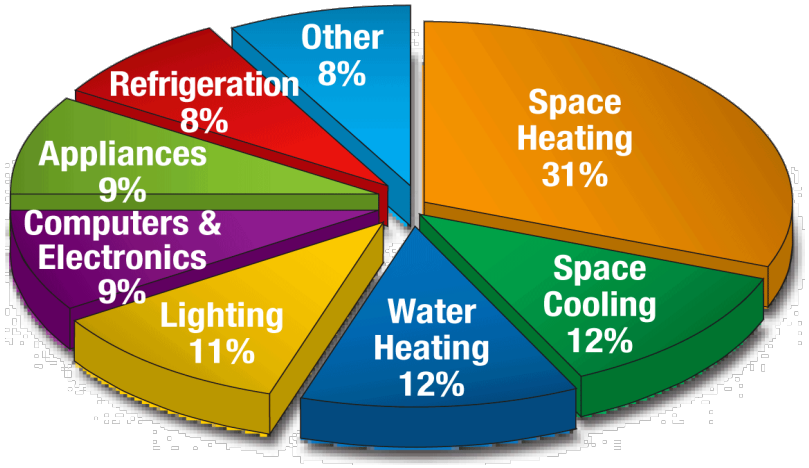


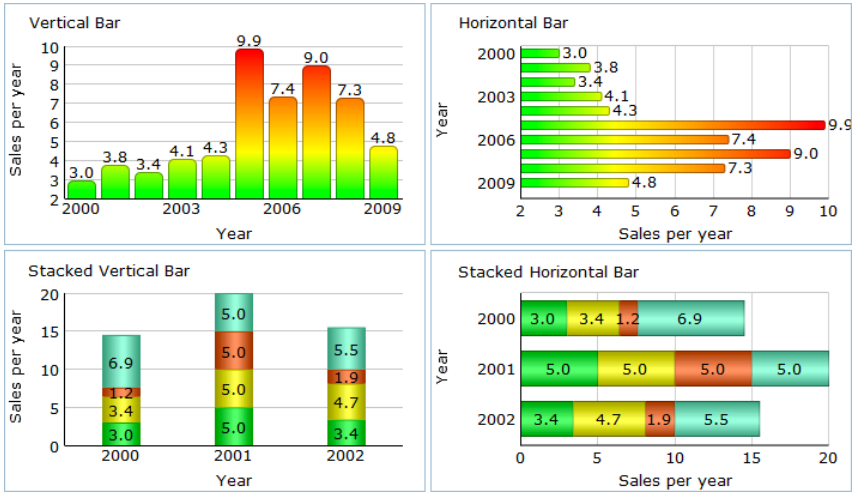
Tree

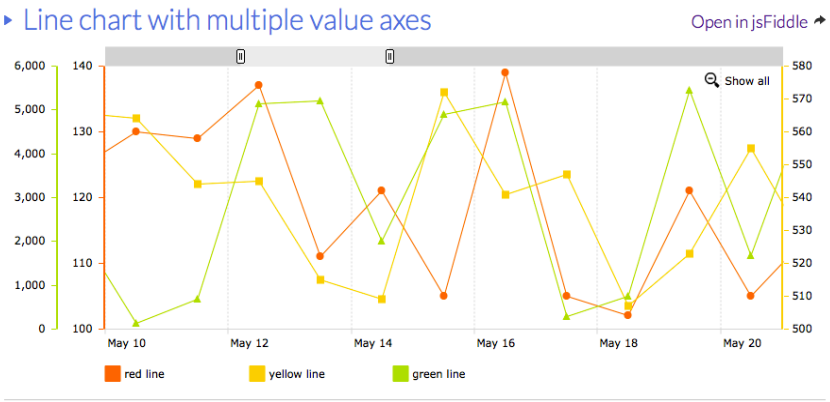


Graph

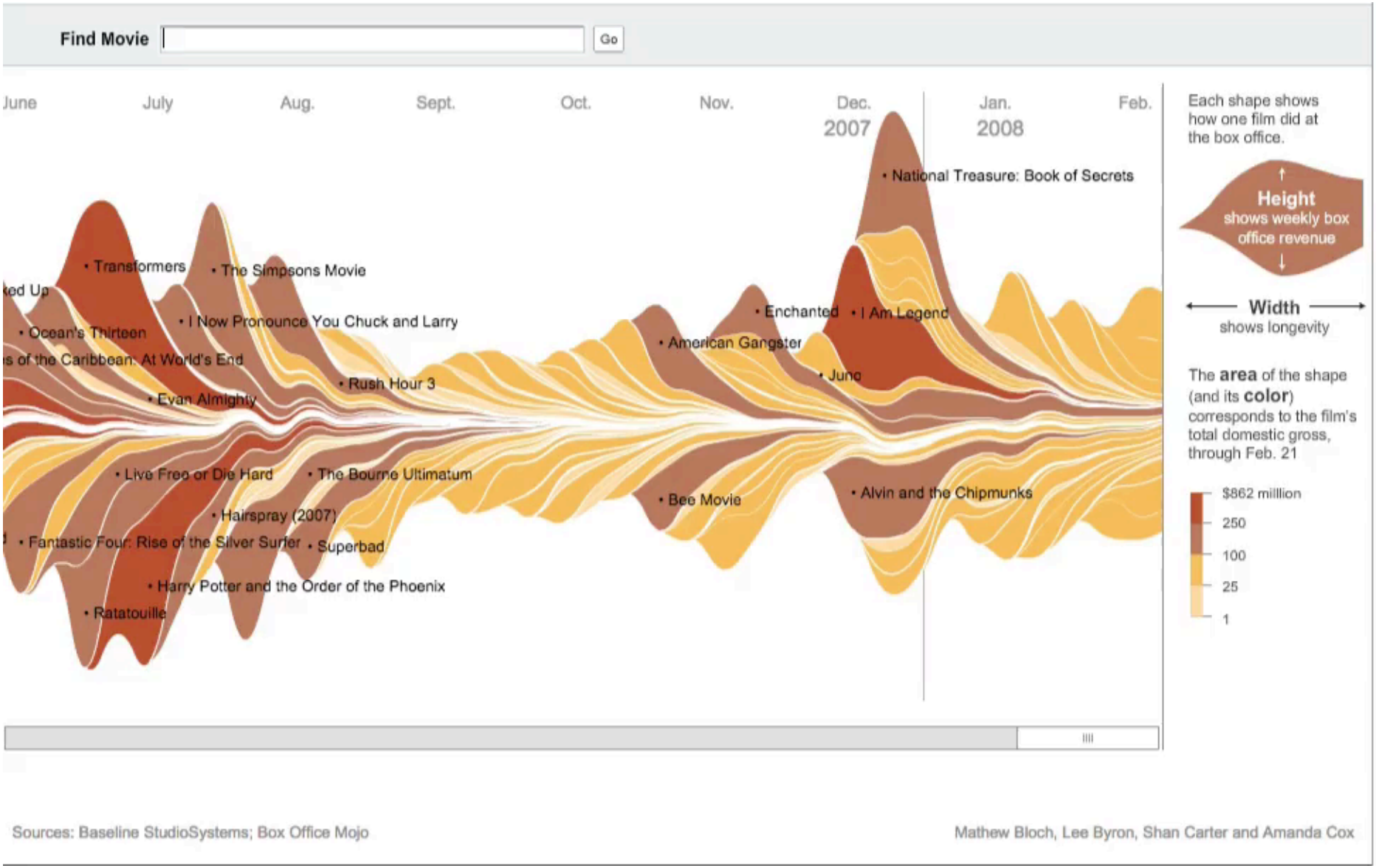
Examples: Visualizing 1D Numerical Data







Example : Visualizing 1D Ordinal Data

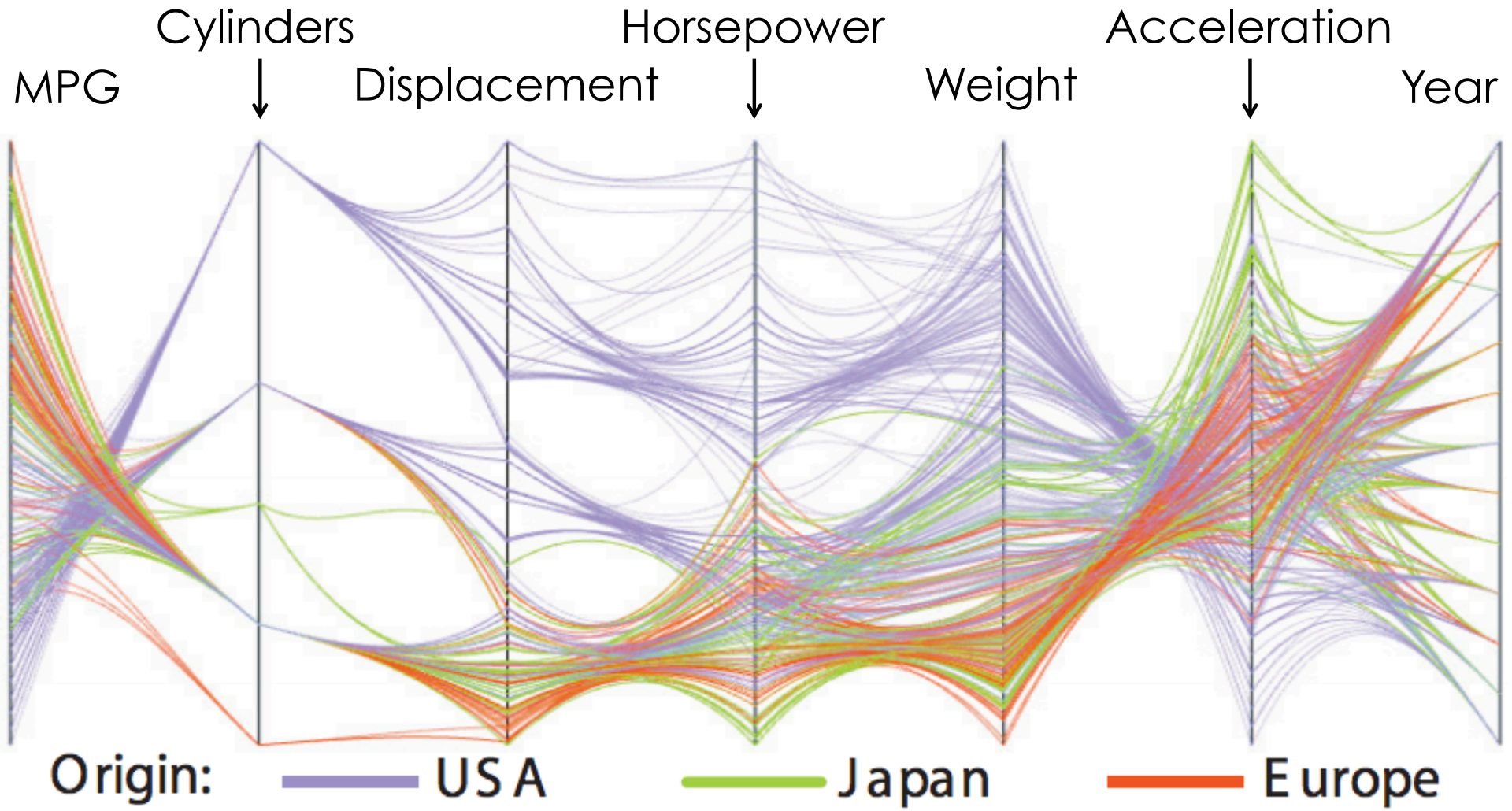


http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html

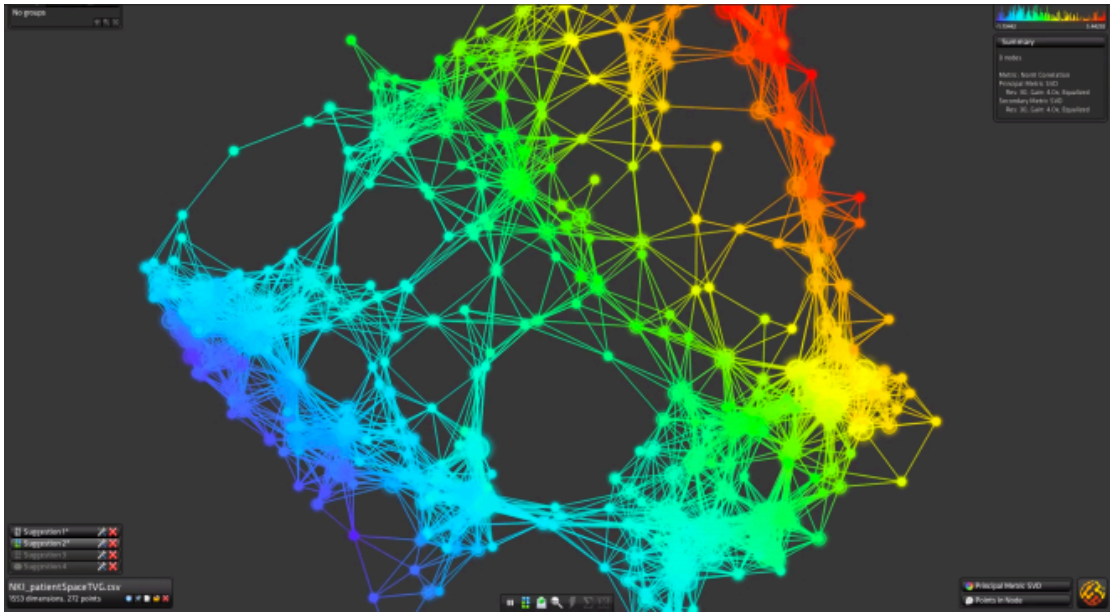
Examples: 2D Data

Size of each Cell: Stock Market Value
Color: Stock Change

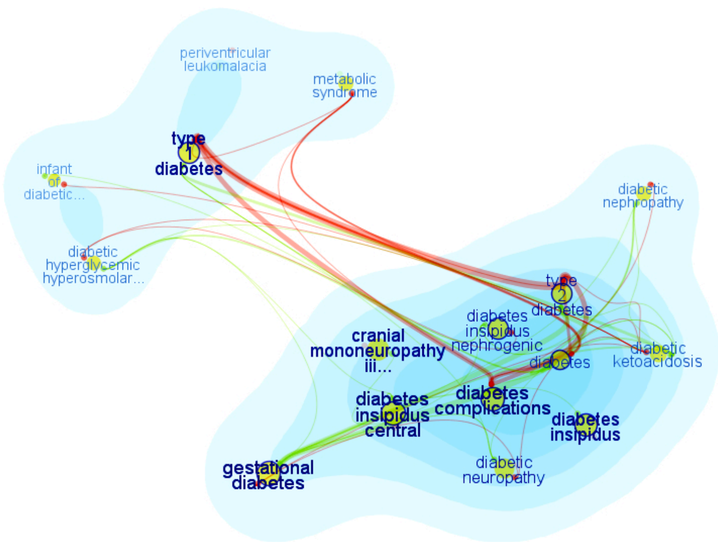
Example：Multi-Dimensional Data

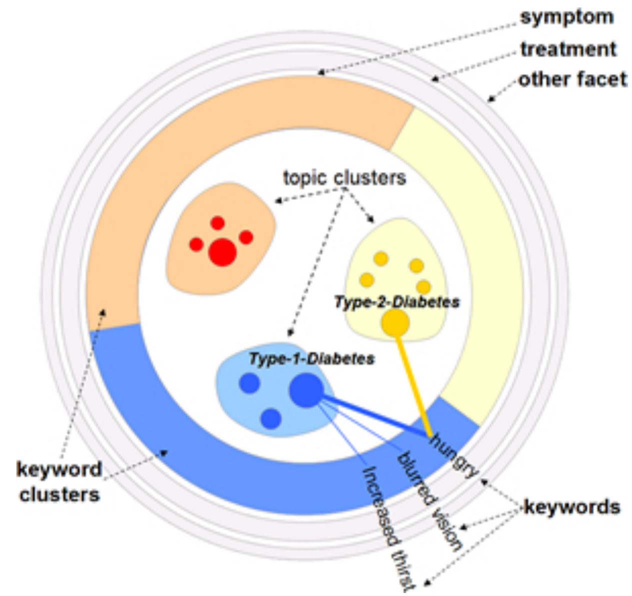


Examples: Visualizing Structured Data

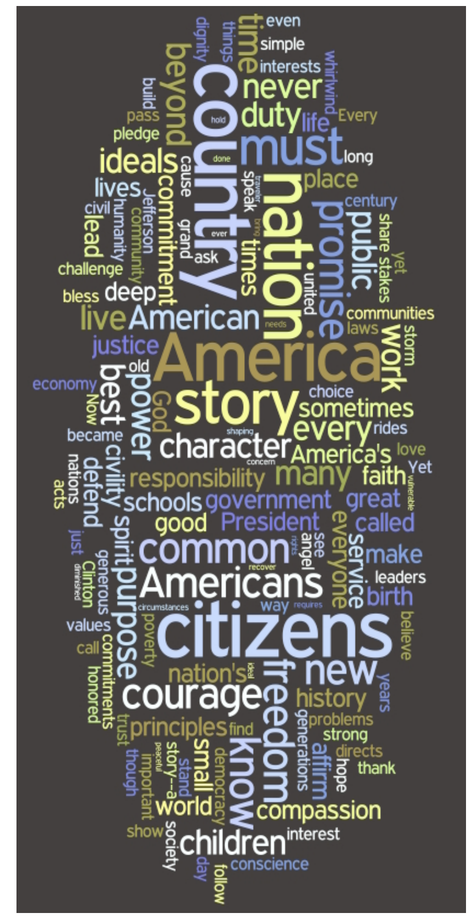


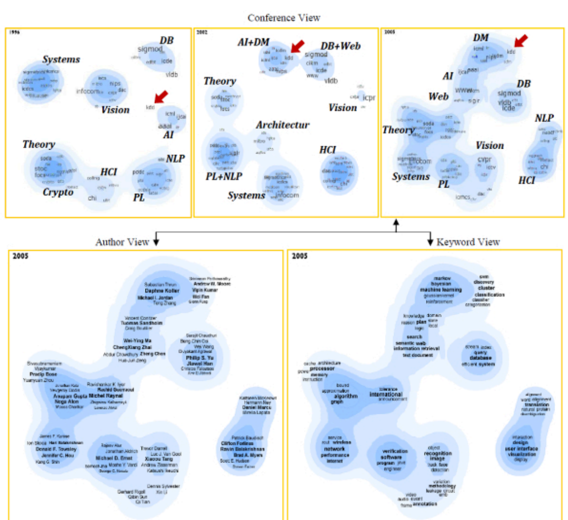
Examples: Visualizing Unstructured Data

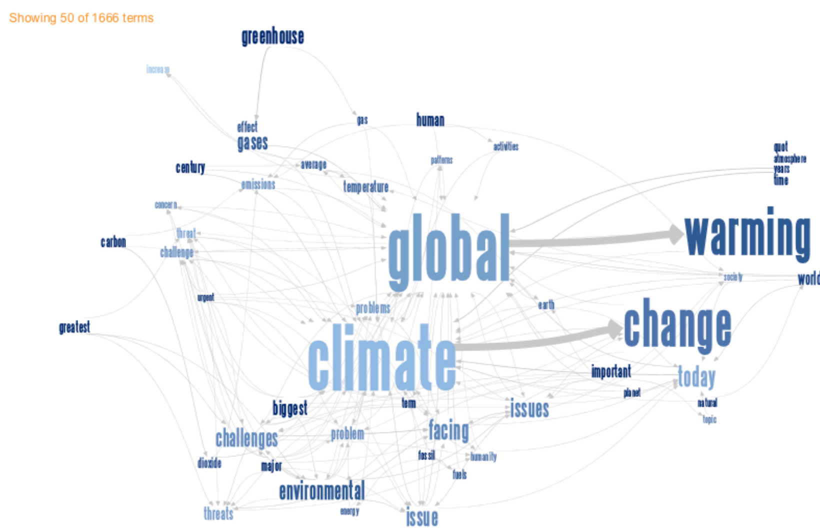




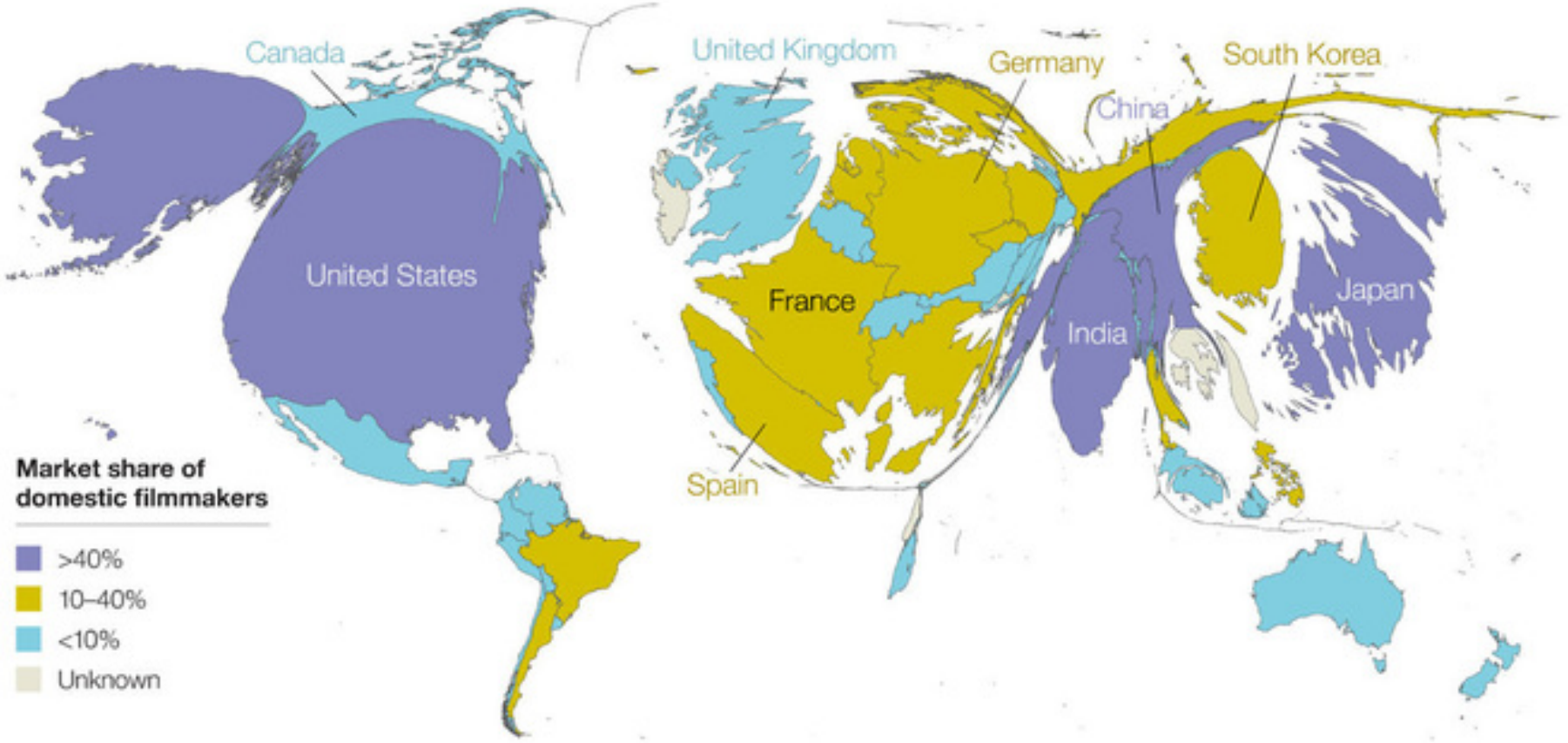
Visualization of Text Documents



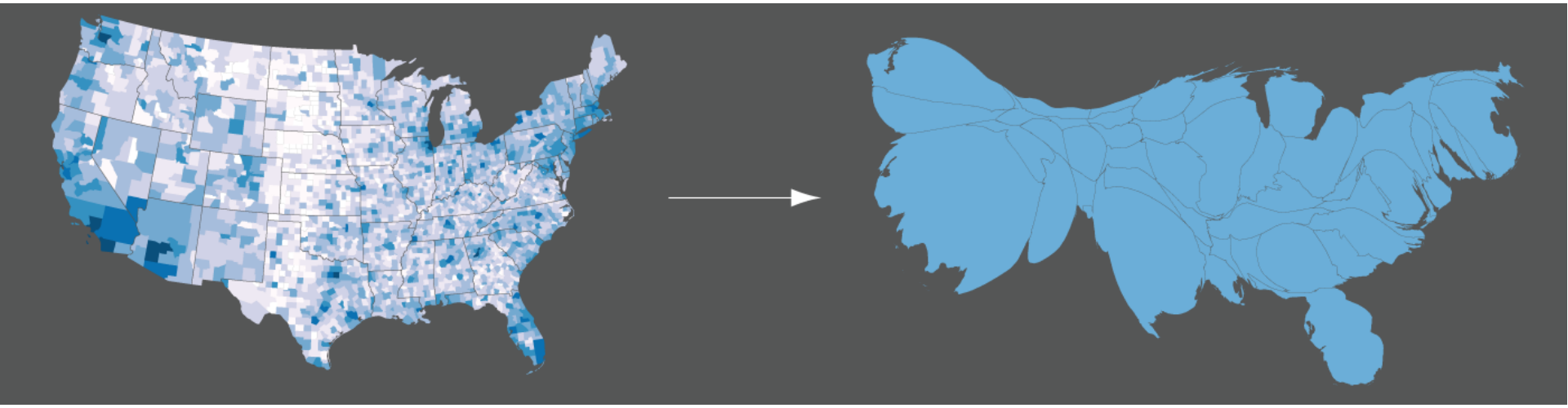


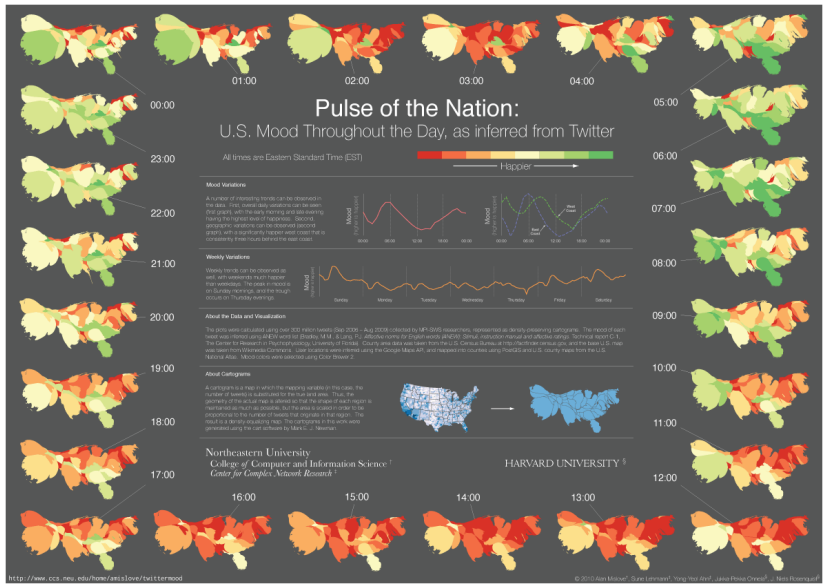


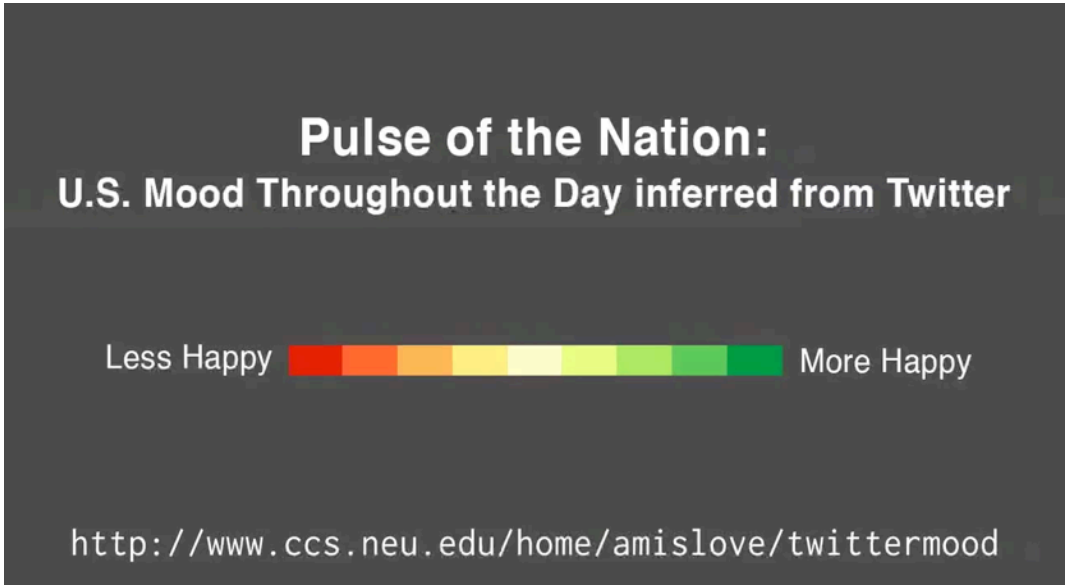
Examples: Geospatial



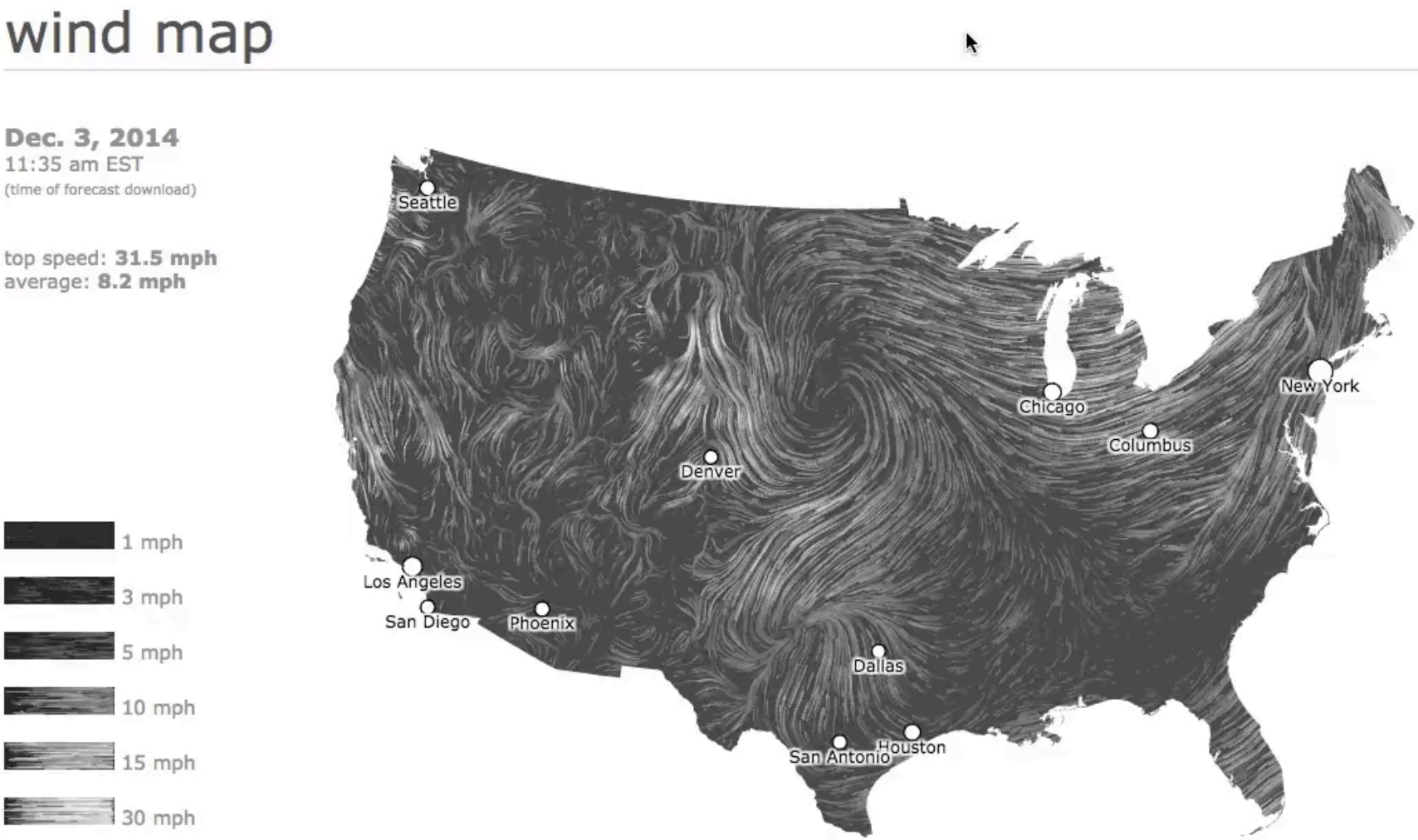
Examples: Visualizing Spatial Temporal Data





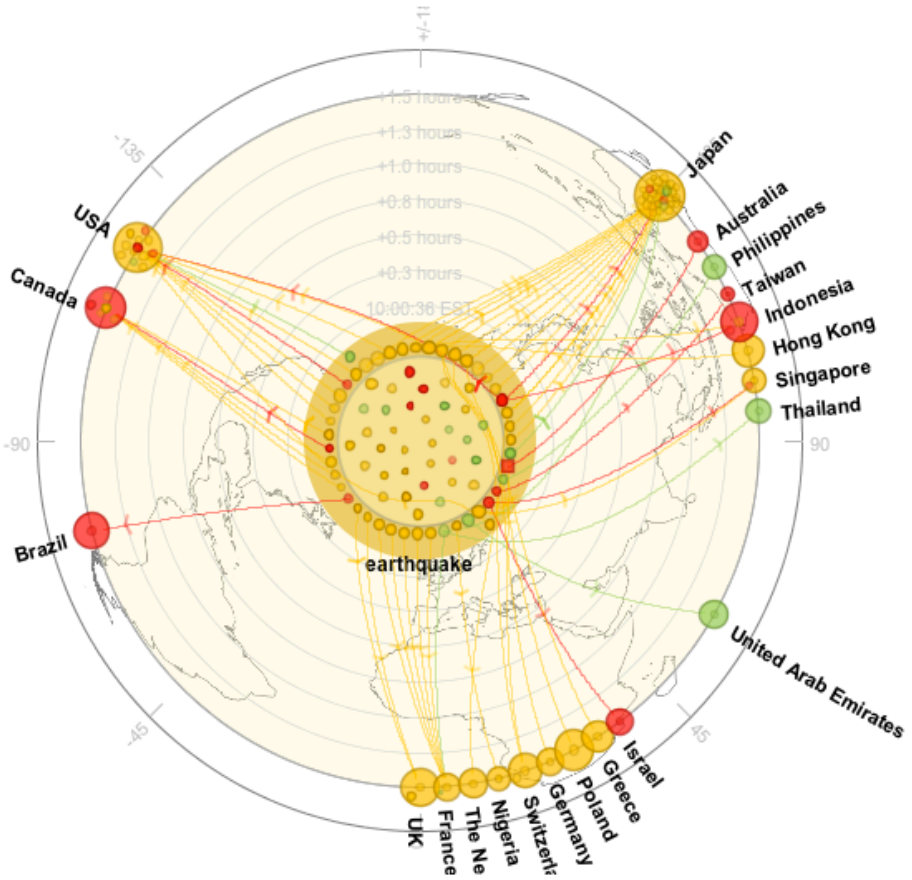


Examples: Visualizing Spatial Temporal Data



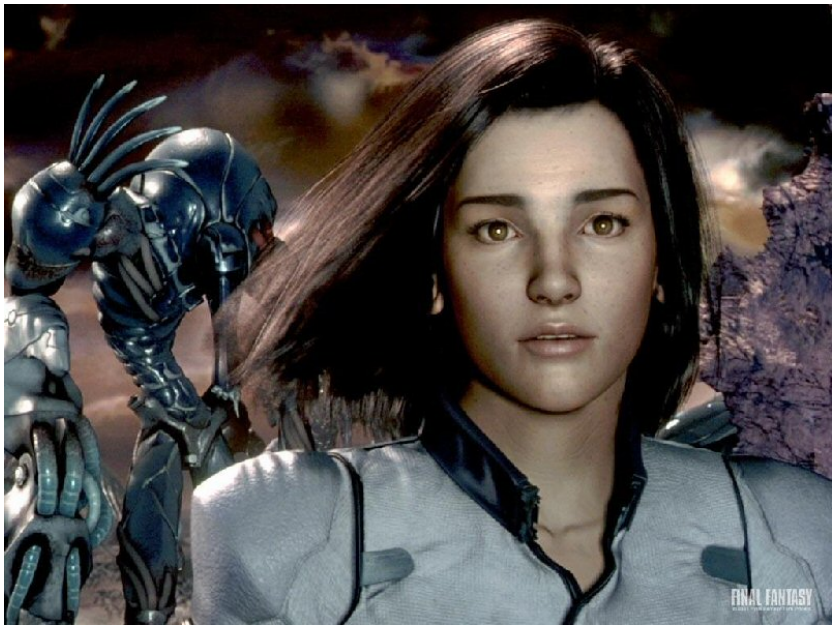
http://hint.fm/wind/

Visualization is not just a beautiful picture

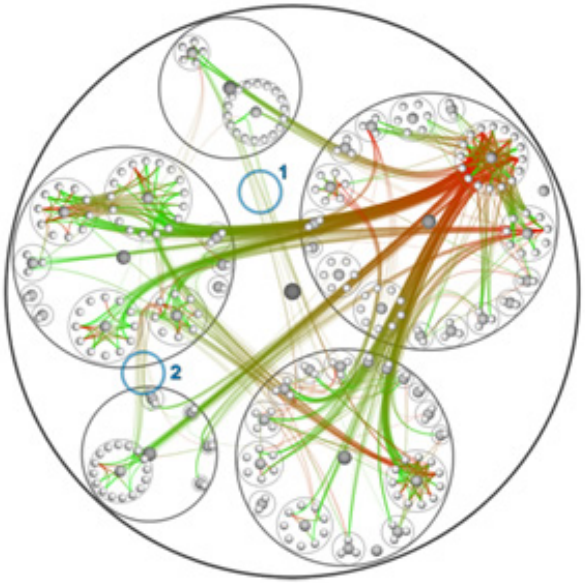


The purpose of visualization is to reveal the insight of the data

InfoVis vs Computer Graphics

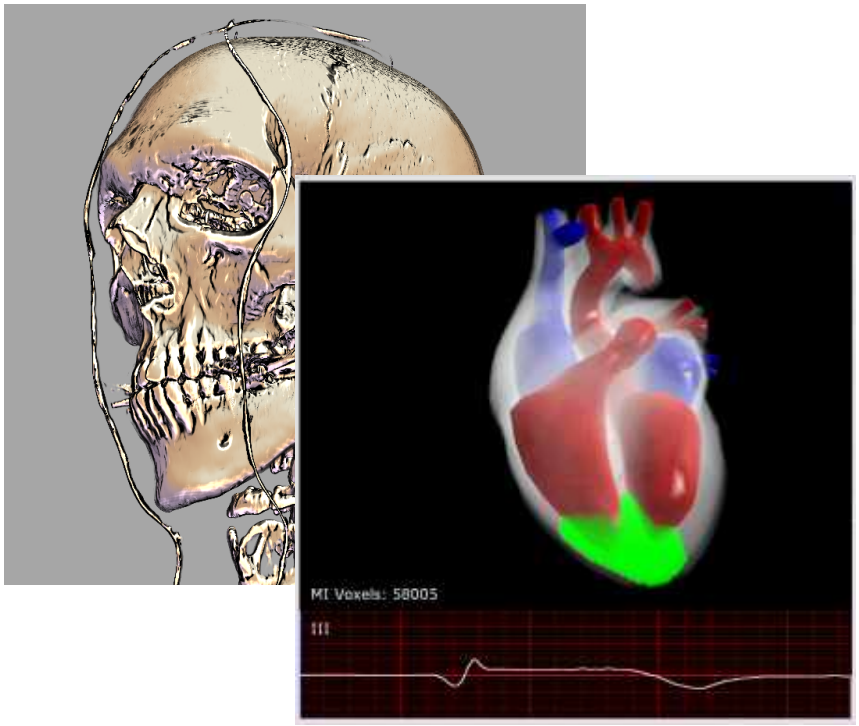


Realism

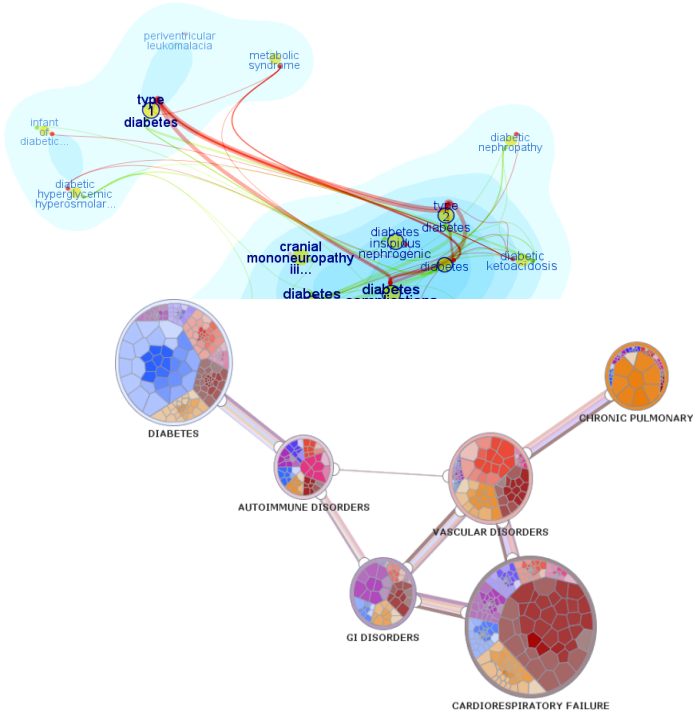


Information

InfoVis v.s. Scientific Visualization



Physical Data



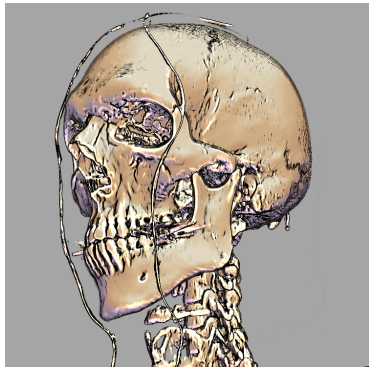
Artificial Data

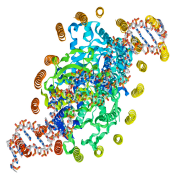
Three Sub-areas

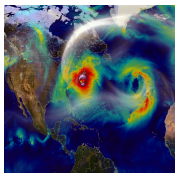
Scientific Visualization (SCIVIS)

Information Visualization (InfoVis)

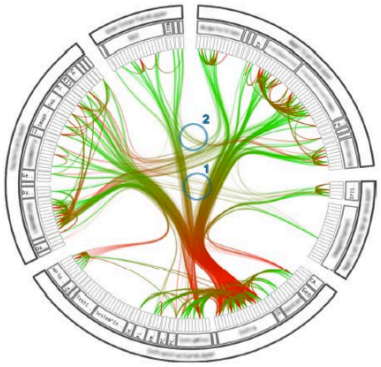
Visual Analysis (VA)

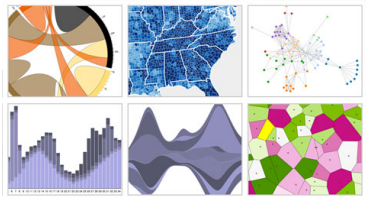






Physical Data





Abstract Data

InfoVis
Data Mining
Machine Learning

Traditional — Research Trend — Modern Technique

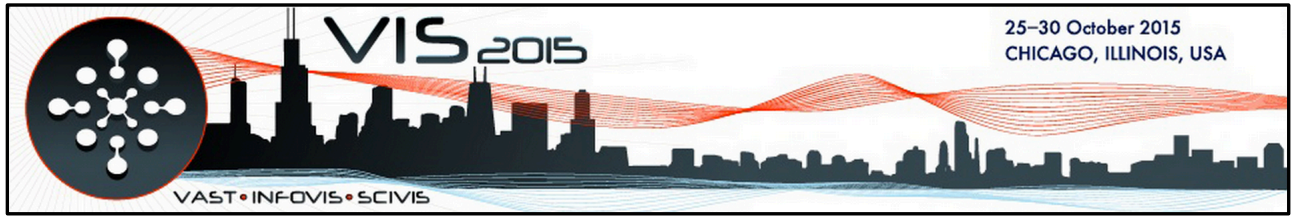
Major Conferences and Journals

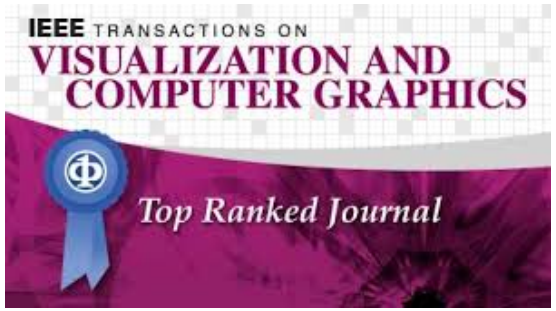
Scientific Visualization (SCIVIS)

Information Visualization (InfoVis)

Visual Analysis (VA)

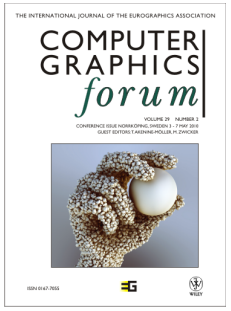
VisWeek: IEEE SCIVIS, INFOVIS, VAST





IEEE Transaction on Visualization and Computer Graphics

>



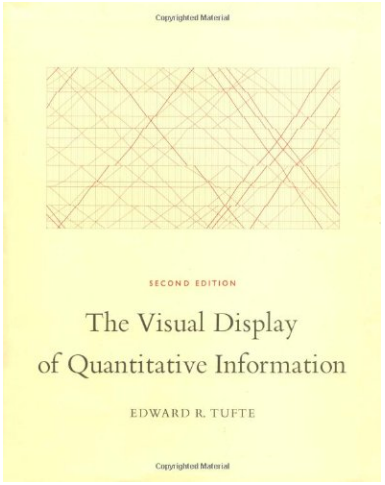
Computer Graphics Forum

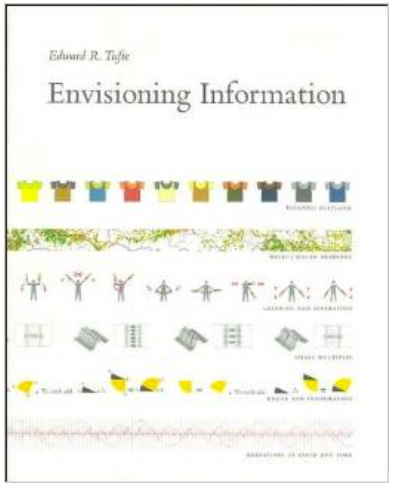
>

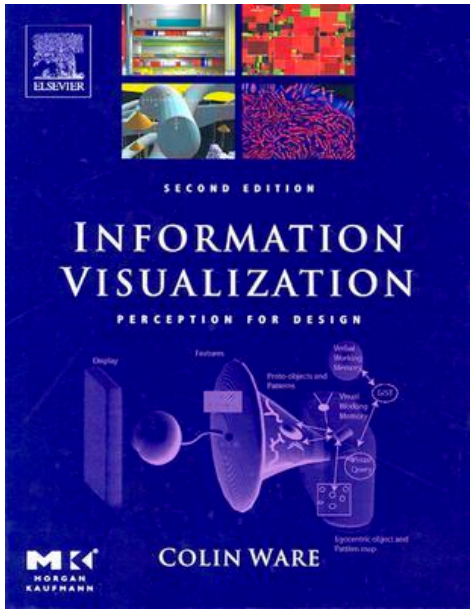
IEEE Computer Graphics and Application

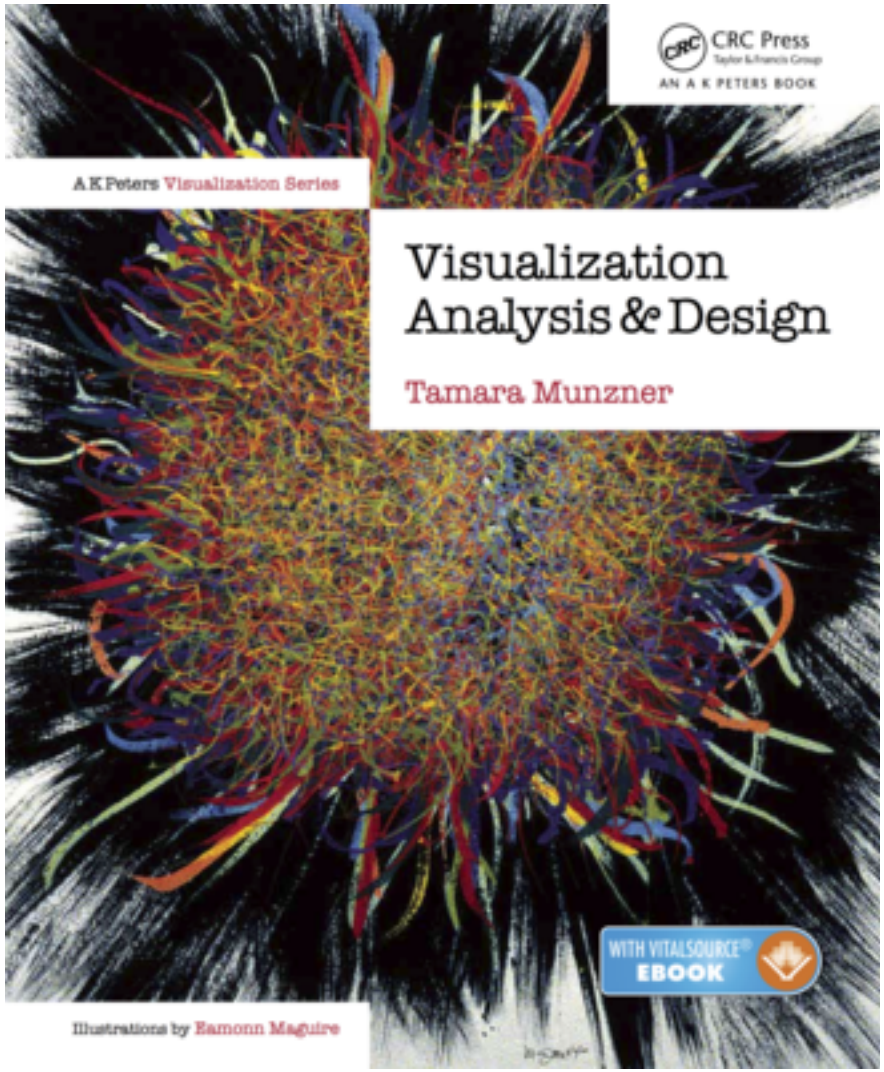
36

Recommended Books



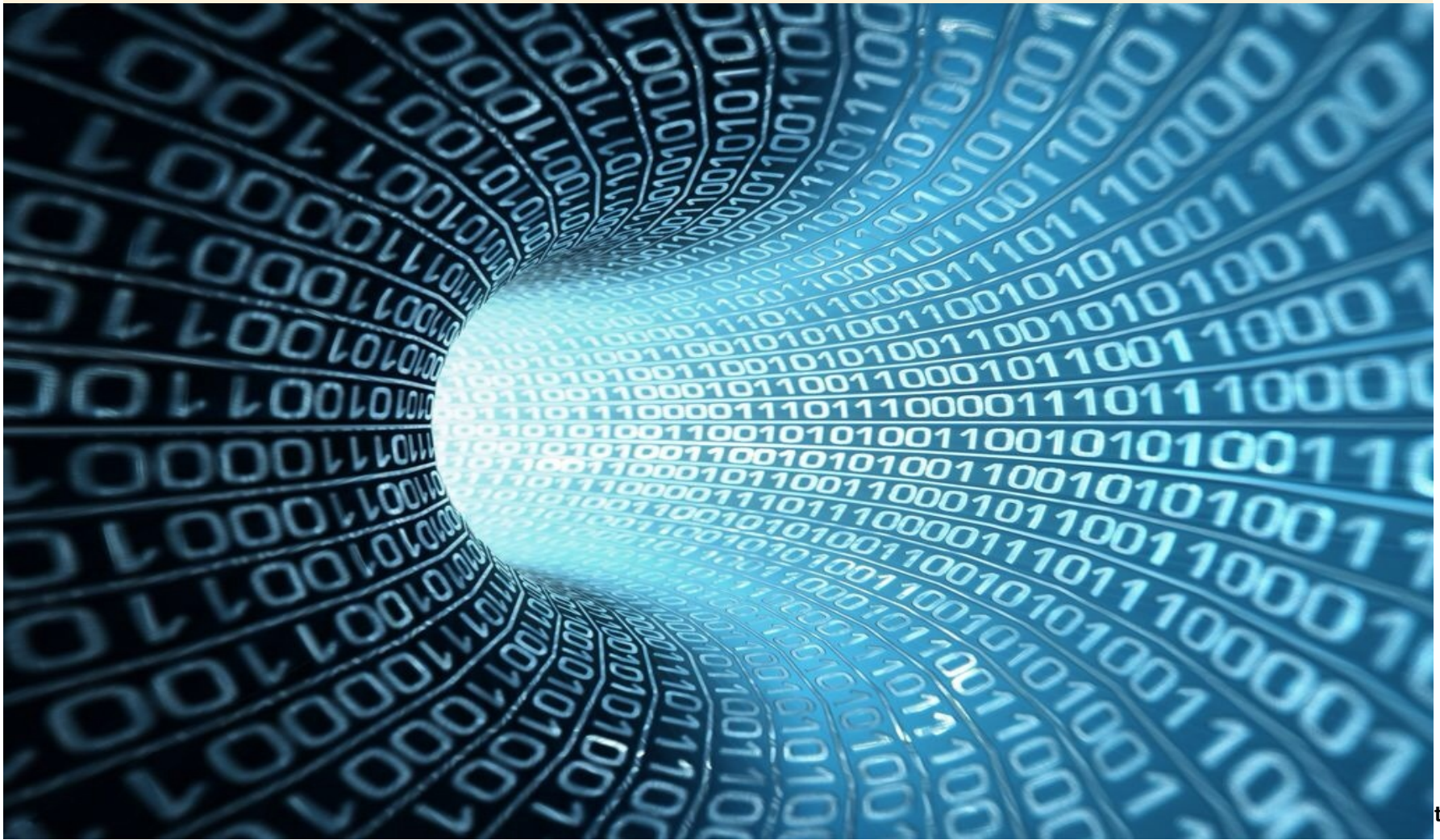






Visualizing Big Data

Are you ready to *Big Data?*



340 million tweets a day!

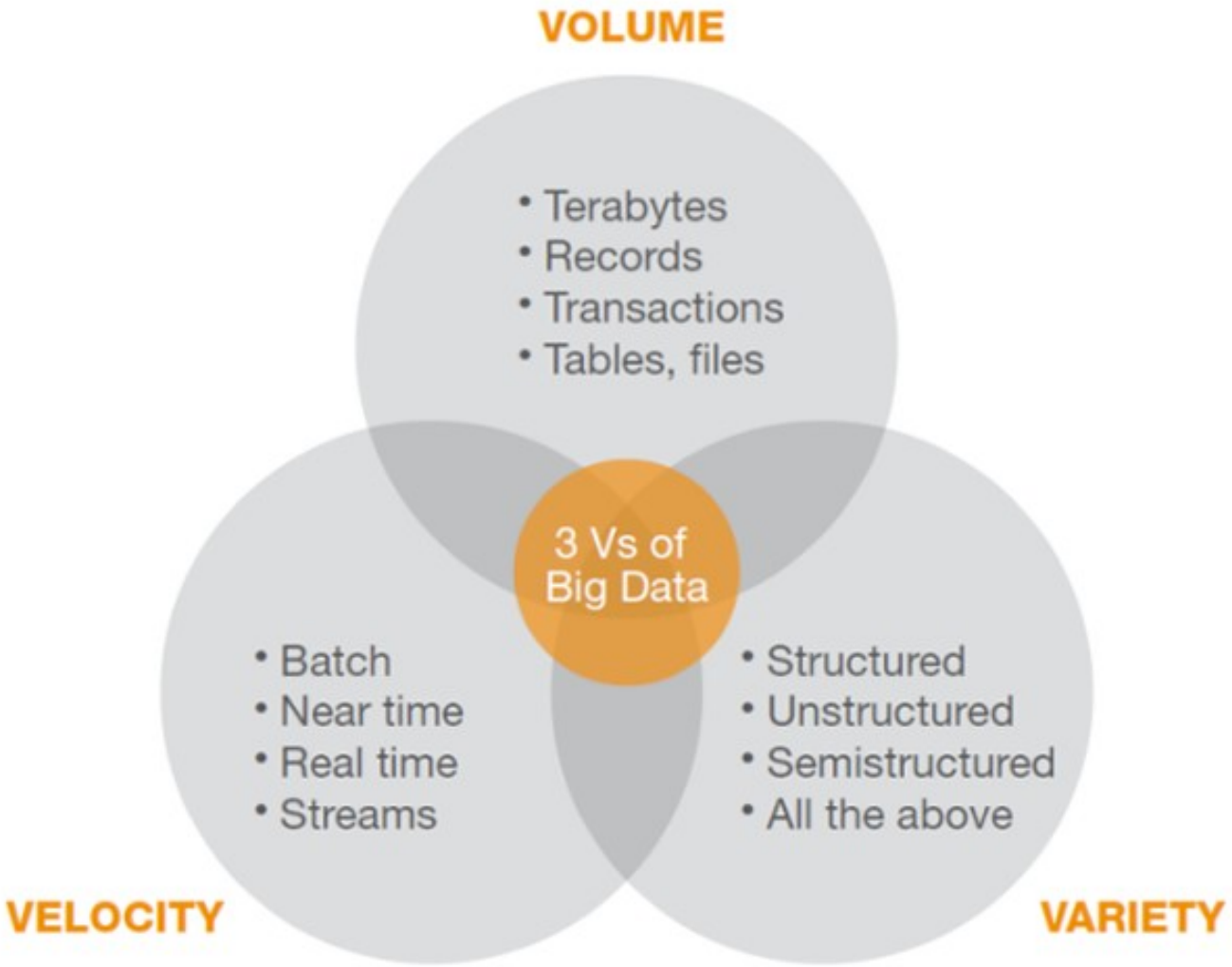
4 billion messages a day!

Information Overload

difficult caused by too much information

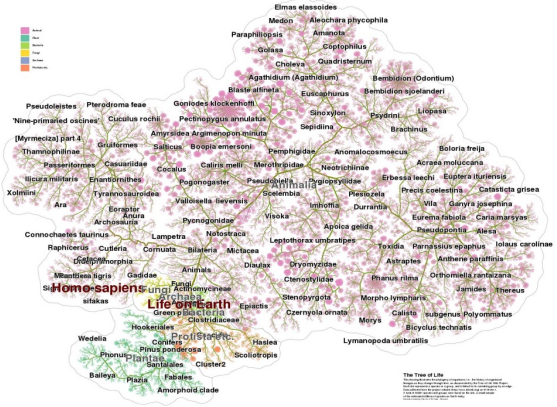
How can we acquire useful information from the overwhelming data

3Vs of Big Data



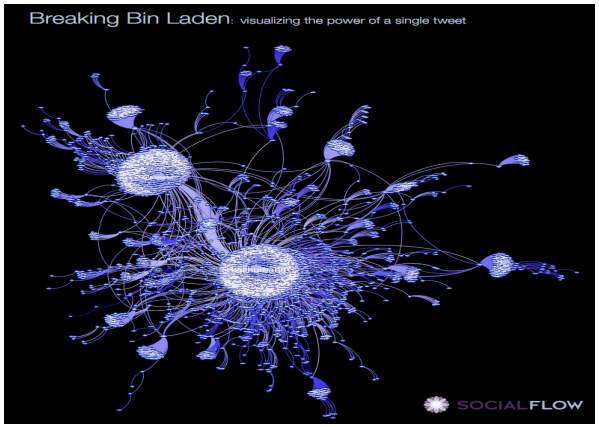
Big Data Visualization

76425 species



Tree of Life by Dr. Yifan Hu

14.8 million tweets



The information diffusion graph of the death of Osama bin Laden by Gilad Lotan

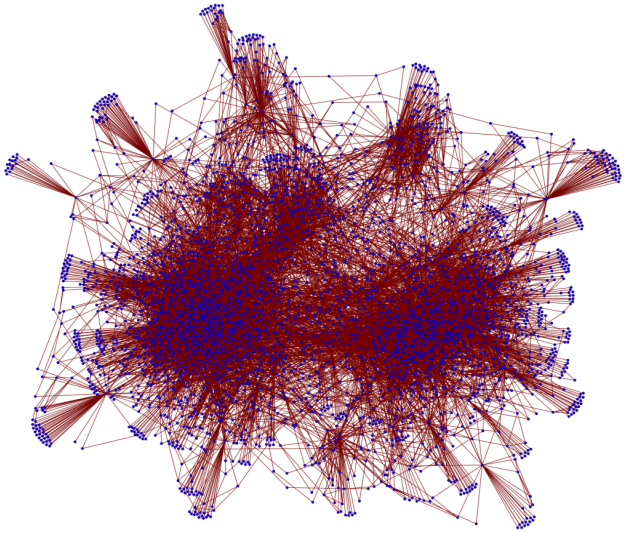
500 million users

Facebook friendship graph by Paul Butler

Challenging Task:

Squeezing millions and even billions of records into million pixels (1600 X 1200 = 2 million pixels)

Challenges



Visual clutter

How can we avoid visual clutters like overlaps and crossings?

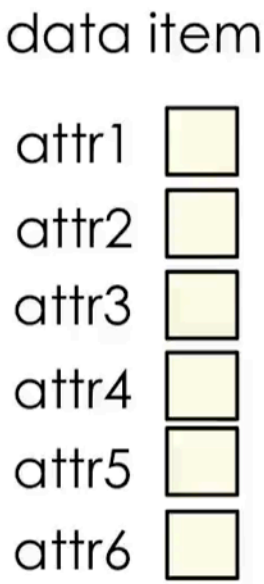
Performance issues

How can we render the huge datasets in real time with rich interactions?

Limited cognition

How can users understand the visual representation when the information is overwhelming?

Techniques (1) : Pixel Oriented Visualization

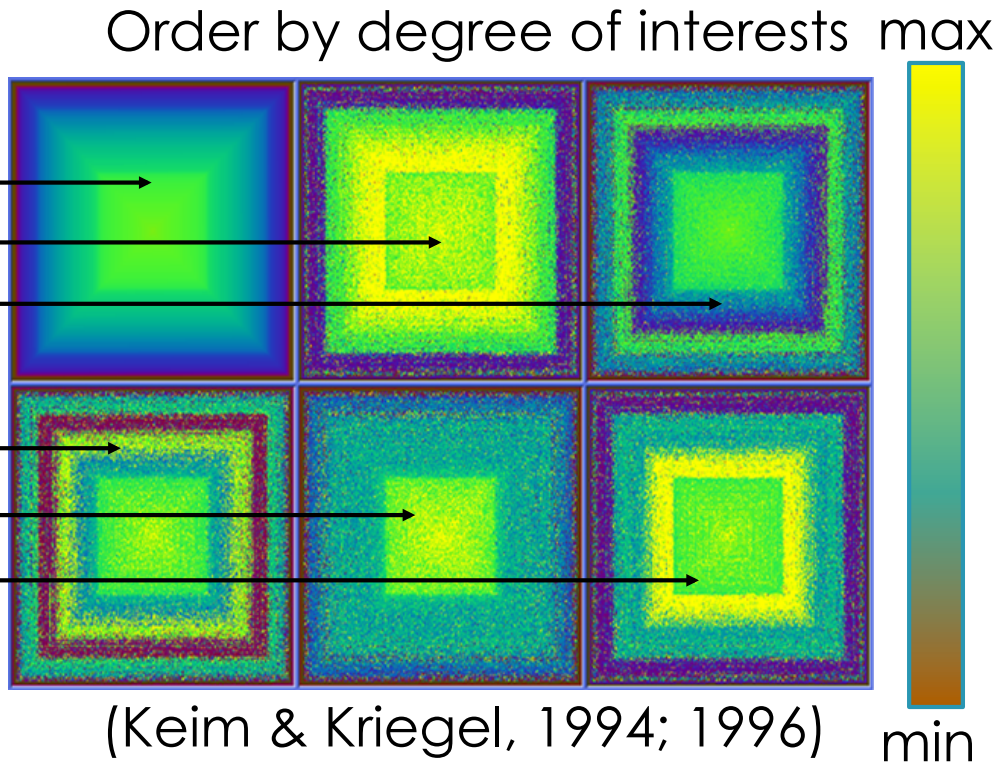


- A multidimensional data item contains 6 attributes

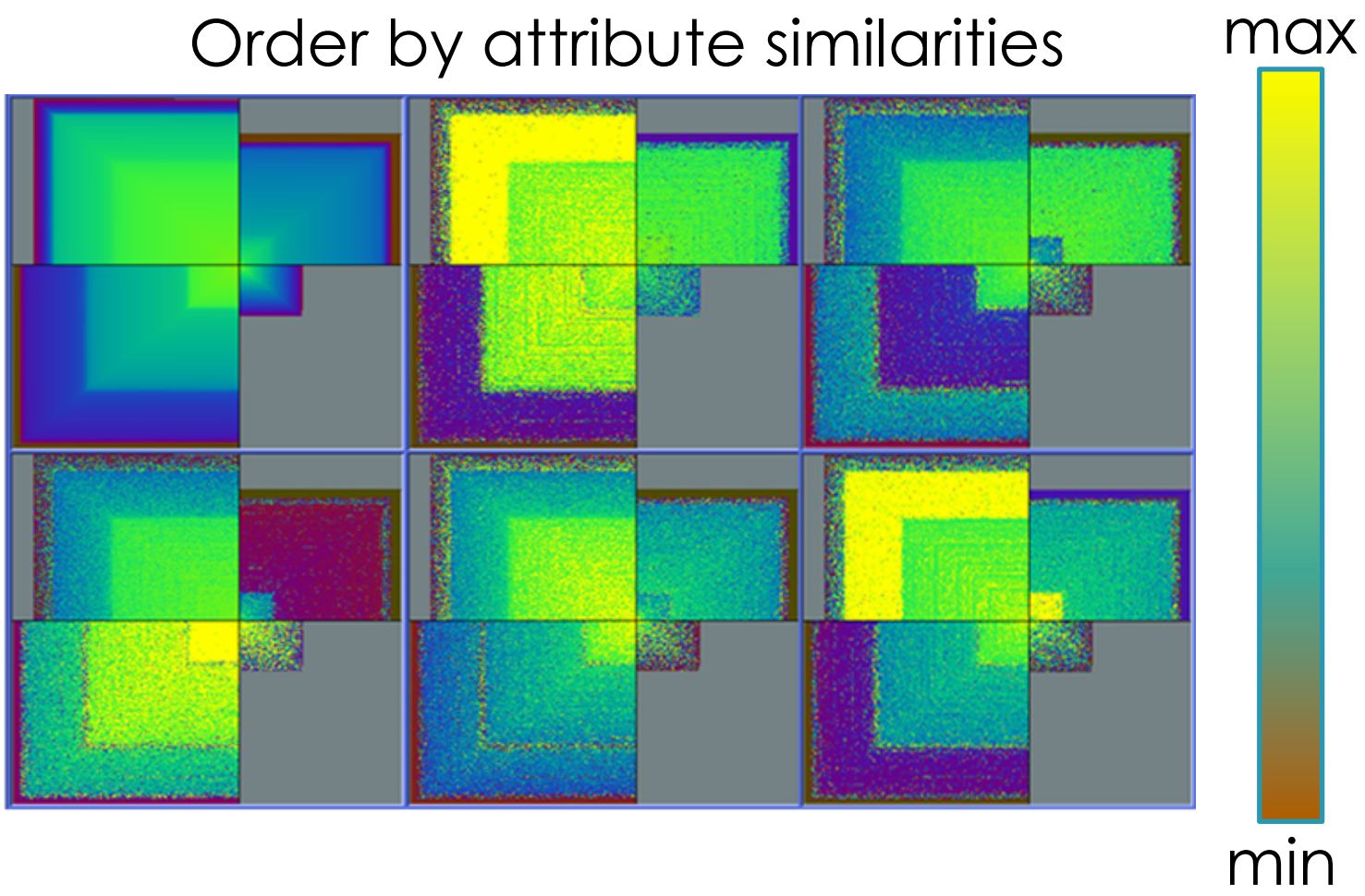
Technique(1) : Pixel Oriented Visualization

Database visualization (10,000 items, 6 dimensions)

Jan	Feb	Mar	Apr	May	Jun
-99.99	-99.99	315.7	317.45	317.5	317.26
315.62	316.38	316.71	317.72	318.29	318.16
316.43	316.97	317.58	319.02	320.03	319.59
316.93	317.7	318.54	319.48	320.58	319.77
317.94	318.56	319.68	320.63	321.01	320.55
318.74	319.08	319.86	321.39	322.24	321.47
319.57	-99.99	-99.99	-99.99	322.24	321.89
319.44	320.44	320.89	322.13	322.16	321.87
320.62	321.59	322.39	323.87	324.01	323.75
322.06	322.5	323.04	324.42	325	324.09
322.57	323.15	323.89	325.02	325.57	325.36
324	324.42	325.64	326.66	327.34	326.76
325.03	325.99	326.87	328.14	328.07	327.66
326.17	326.68	327.18	327.78	328.92	328.57
326.77	327.63	327.75	329.72	330.07	329.09
328.55	329.56	330.3	331.5	332.48	332.07
329.35	330.71	331.48	332.65	333.09	332.25
330.4	331.41	332.04	333.31	333.96	333.6
331.75	332.56	333.5	334.58	334.87	334.34
332.93	333.42	334.7	336.07	336.74	336.27
334.97	335.39	336.64	337.76	338.01	337.89
336.23	336.76	337.96	338.89	339.47	339.29
338.01	338.36	340.08	340.77	341.46	341.17
339.23	340.47	341.38	342.51	342.91	342.25
340.75	341.61	342.7	343.57	344.13	343.35
341.37	342.52	343.1	344.94	345.75	345.32
343.7	344.5	345.28	347.08	347.43	346.79
344.97	346	347.43	348.35	348.93	348.25
346.3	346.96	347.86	349.55	350.21	349.54
348.02	348.47	349.42	350.99	351.84	351.25
350.43	351.73	352.22	353.59	354.22	353.79
352.76	353.97	353.68	355.42	355.67	355.13
353.66	354.7	355.39	356.2	357.16	356.23
354.72	355.75	357.16	358.6	359.33	358.24
355.98	356.72	357.81	359.15	359.66	359.25
356.7	357.16	358.38	359.46	360.28	359.6
358.37	358.91	359.97	361.26	361.68	360.95
359.97	361	361.64	363.45	363.79	363.26
362.05	363.25	364.02	364.72	365.41	364.97
363.18	364	364.56	366.35	366.79	365.62
365.33	366.15	367.31	368.61	369.3	368.87
368.15	368.87	369.59	371.14	371	370.35
369.14	369.46	370.52	371.66	371.82	371.7
370.28	371.5	372.12	372.87	374.02	373.3
372.43	373.09	373.52	374.86	375.55	375.41
374.68	375.63	376.11	377.65	378.35	378.13
376.79	377.37	378.41	380.52	380.63	379.57
378.37	379.69	380.41	382.1	382.28	382.13
381.38	382.03	382.64	384.62	384.95	384.06
382.45	383.68	384.23	386.26	386.39	385.87
385.07	385.72	385.85	386.71	388.45	387.64



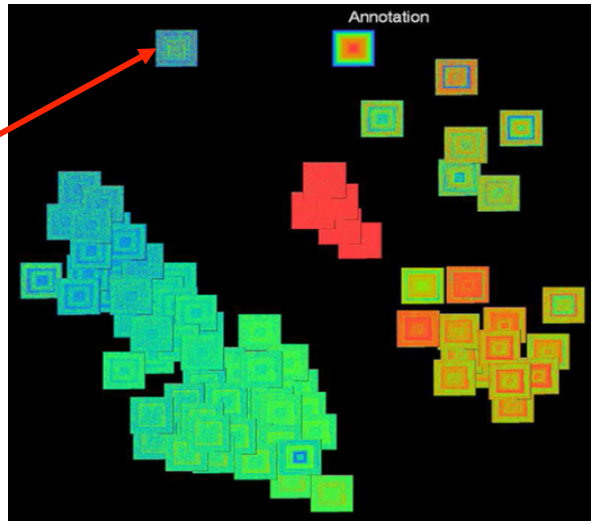
Database Visualization (10,000 items, 6 dimensions)



Different Ways for splitting the display region

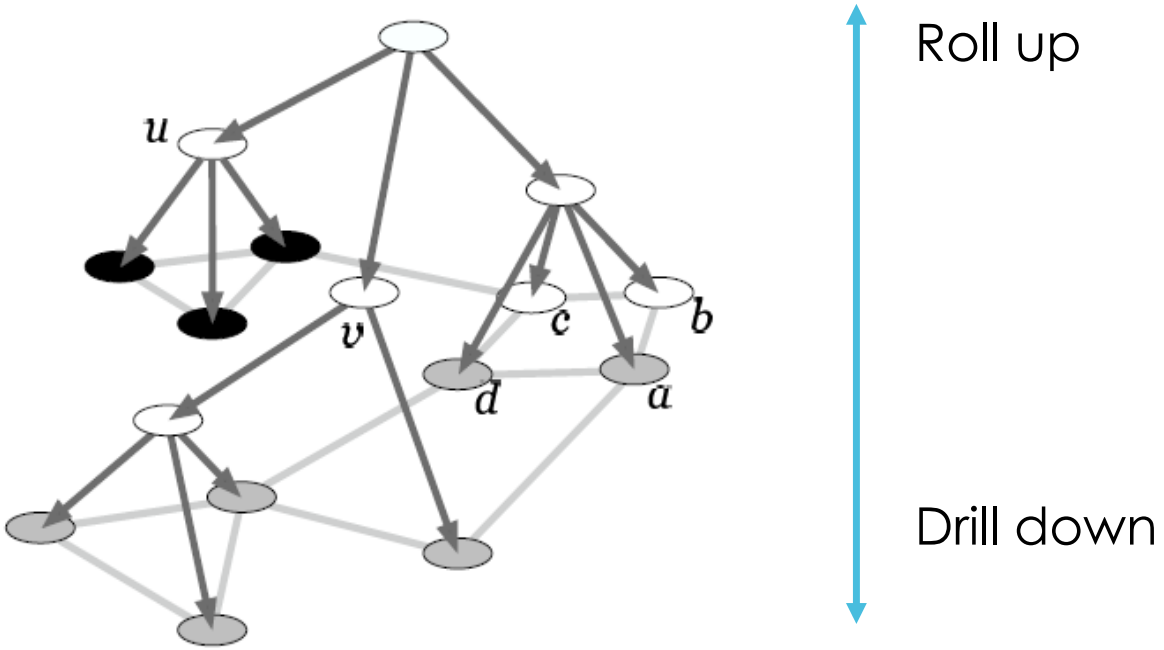
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual Average
1958	-99.99	-99.99	315.7	317.45	317.5	317.26	315.86	314.93	313.2	312.44	313.33	314.6	-99.99
1959	315.62	316.38	316.71	317.72	318.29	318.16	316.54	314.8	313.84	313.26	314.8	315.58	315.98
1960	316.43	316.97	317.58	319.02	320.03	319.59	318.18	315.91	314.16	313.84	315	316.19	316.91
1961	316.93	317.7	318.54	319.48	320.58	319.77	318.57	316.79	314.8	315.38	316.1	317.01	317.64
1962	317.94	318.56	319.68	320.63	321.01	320.55	319.58	317.4	316.25	315.42	316.69	317.7	318.45
1963	318.74	319.08	319.86	321.39	322.24	321.47	319.74	317.77	316.21	315.99	317.12	318.31	318.99
1964	319.57	-99.99	-99.99	-99.99	322.24	321.89	320.44	318.7	316.7	316.79	317.79	318.71	-99.99
1965	319.44	320.44	320.89	322.13	322.16	321.87	321.39	318.8	317.81	317.3	318.87	319.42	320.04
1966	320.62	321.59	322.39	323.87	324.01	323.75	322.39	320.37	318.64	318.1	319.79	321.08	321.38
1967	322.06	322.5	323.04	324.42	325	324.09	322.55	320.92	319.31	319.31	320.72	321.96	322.16
1968	322.57	323.15	323.89	325.02	325.57	325.36	324.14	322.03	320.41	320.25	321.31	322.84	323.05
1969	324	324.42	325.64	326.66	327.34	326.76	325.88	323.67	322.38	321.78	322.85	324.12	324.63
1970	325.03	325.99	326.87	328.14	328.07	327.66	326.35	324.69	323.1	323.16	323.98	325.13	325.68
1971	326.17	326.68	327.18	327.78	328.92	328.57	327.34	325.46	323.36	323.57	324.8	326.01	326.32
1972	326.77	327.63	327.75	329.72	330.07	329.09	328.05	326.32	324.93	325.06	326.5	327.55	327.45
1973	328.55	329.56	330.3	331.5	332.48	332.07	330.87	329.31	327.51	327.18	328.16	328.64	329.68
1974	329.35	330.71	331.48	332.65	333.09	332.25	331.18	329.4	327.43	327.37	328.46	329.57	330.25
1975	330.4	331.41	332.04	333.31	333.96	333.6	331.91	330.06	328.56	328.34	329.49	330.76	331.15
1976	331.75	332.56	333.5	334.58	334.87	334.34	333.05	330.94	329.3	328.94	330.31	331.68	332.15
1977	332.93	333.42	334.7	336.07	336.74	336.27	334.93	332.75	331.59	331.16	332.4	333.85	333.9
1978	334.97	335.39	336.64	337.76	338.01	337.89	336.54	334.68	332.76	332.55	333.92	334.95	335.51
1979	336.23	336.76	337.96	338.89	339.47	339.29	337.73	336.09	333.91	333.86	335.29	336.73	336.85
1980	338.01	338.36	340.08	340.77	341.46	341.17	339.56	337.6	335.88	336.02	337.1	338.21	338.69
1981	339.23	340.47	341.38	342.51	342.91	342.25	340.49	338.43	336.69	336.86	338.36	339.61	339.9
1982	340.75	341.61	342.7	343.57	344.13	343.35	342.06	339.81	337.98	337.86	339.26	340.49	341.13
1983	341.37	342.52	343.1	344.94	345.75	345.32	343.99	342.39	339.86	339.99	341.15	342.99	342.78
1984	343.7	344.5	345.28	347.08	347.43	346.79	345.4	343.28	341.07	341.35	342.98	344.22	344.42
1985	344.97	346	347.43	348.35	348.93	348.25	346.56	344.68	343.09	342.8	344.24	345.55	345.9
1986	346.3	346.96	347.86	349.55	350.21	349.54	347.94	345.9	344.85	344.17	345.66	346.9	347.15
1987	348.02	348.47	349.42	350.99	351.84	351.25	349.52	348.1	346.45	346.36	347.81	348.96	348.93
1988	350.43	351.73	352.22	353.59	354.22	353.79	352.38	350.43	348.72	348.88	350.07	351.34	351.48
1989	352.76	353.07	353.68	355.42	355.67	355.13	353.9	351.67	349.8	349.99	351.29	352.52	352.91
1990	353.66	354.7	355.39	356.2	357.16	356.23	354.82	352.91	350.96	351.18	352.83	354.21	354.19
1991	354.72	355.75	357.16	358.6	359.33	358.24	356.17	354.02	352.15	352.21	353.75	354.99	355.59
1992	355.98	356.72	357.81	359.15	359.66	359.25	357.02	355	353.01	353.31	354.16	355.4	356.37
1993	356.7	357.16	358.38	359.46	360.28	359.6	357.57	355.52	353.69	353.99	355.34	356.8	357.04
1994	358.37	358.91	359.97	361.26	361.68	360.95	359.55	357.48	355.84	355.99	357.58	359.04	358.89
1995	359.97	361	361.64	363.45	363.79	363.26	361.9	359.46	358.05	357.76	359.56	360.7	360.88
1996	362.05	363.25	364.02	364.72	365.41	364.97	363.65	361.48	359.45	359.6	360.76	362.33	362.64
1997	363.18	364	364.56	366.35	366.79	365.62	364.47	362.51	360.19	360.77	362.43	364.28	363.76
1998	365.33	366.15	367.31	368.61	369.3	368.87	367.64	365.77	363.9	364.23	365.46	366.97	366.63
1999	368.15	368.87	369.59	371.14	371	370.35	369.27	366.93	364.63	365.13	366.67	368.01	368.31
2000	369.14	369.46	370.52	371.66	371.82	371.7	370.12	368.12	366.62	366.73	368.29	369.53	369.48
2001	370.28	371.5	372.12	372.87	374.02	373.3	371.62	369.55	367.96	368.09	369.68	371.24	371.02
2002	372.43	373.09	373.52	374.86	375.55	375.41	374.02	371.49	370.7	370.25	372.08	373.78	373.1
2003	374.68	375.63	376.11	377.65	378.35	378.13	376.62	374.5	372.99	373.01	374.35	375.7	375.64
2004	376.79	377.37	378.41	380.52	380.63	379.57	377.79	375.86	374.07	374.24	375.86	377.47	377.38
2005	378.37	379.69	380.41	382.1	382.28	382.13	380.66	378.71	376.42	376.88	378.32	380.04	379.67
2006	381.38	382.03	382.64	384.62	384.95	384.06	382.29	380.47	378.67	379.06	380.14	381.74	381.84
2007	382.45	383.68	384.23	386.26	386.39	385.87	384.39	381.78	380.73	380.81	382.33	383.69	383.55
2008	385.07	385.72	385.85	386.71	388.45	387.64	386.1	383.95	382.91	382.73	383.96	385.02	385.34

Values above represent monthly concentrations adjusted to represent 2400 hours on the 15th day of each month. Units are parts per million by volume (ppmv) expressed in the 2003A SIO manometric mole fraction scale. The "annual average" is the arithmetic mean of the twelve monthly values where no monthly values are missing.



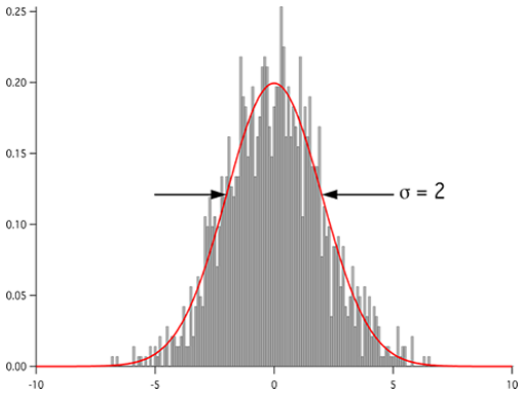
(Yang et al., 2006)

Techniques (2): Aggregation & Level of Details (LOD)

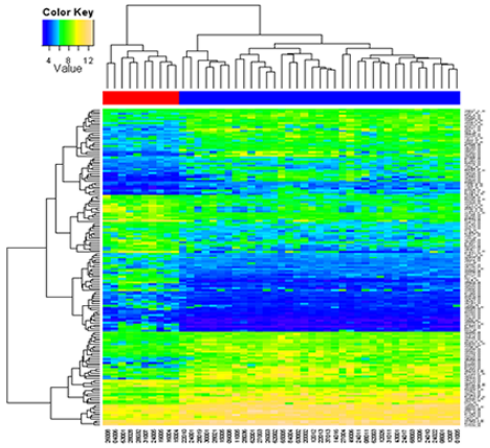


Building a tree for aggregating data items in either a bottom-up or top-down approach

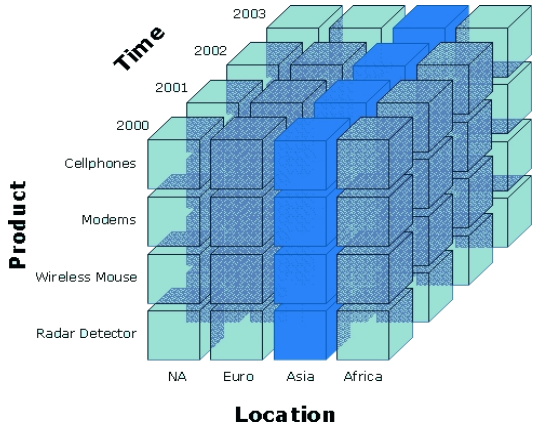
Technique (2) : Aggregation & LOD



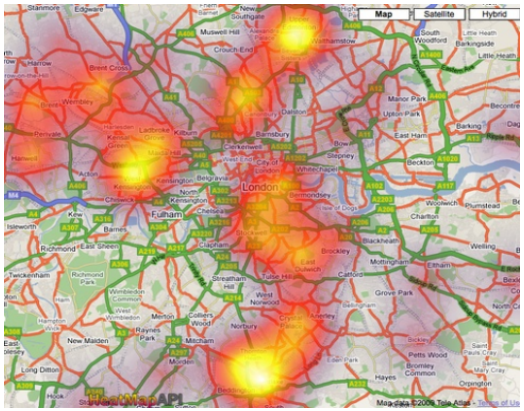
Histogram (Pearson, 1895)



Heatmap
(Wilkinson & Friendly, 2009)



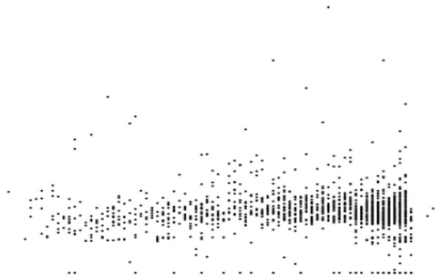
InfoCube
(Stolte et al., 2003)

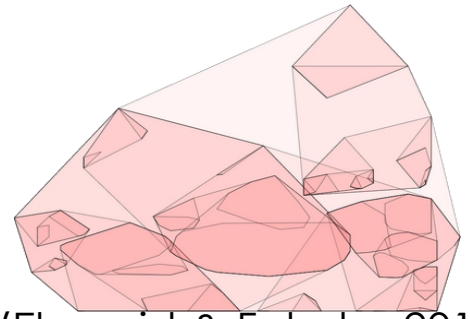


(Lin et al., 2010)

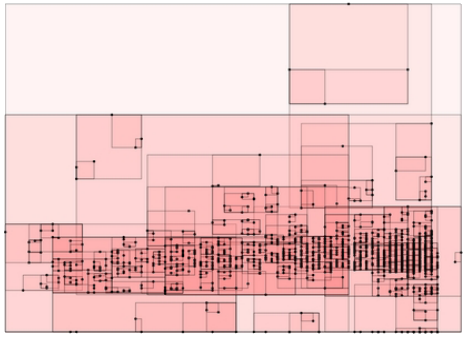
Techniques (2) : Aggregation

Scatter Plots



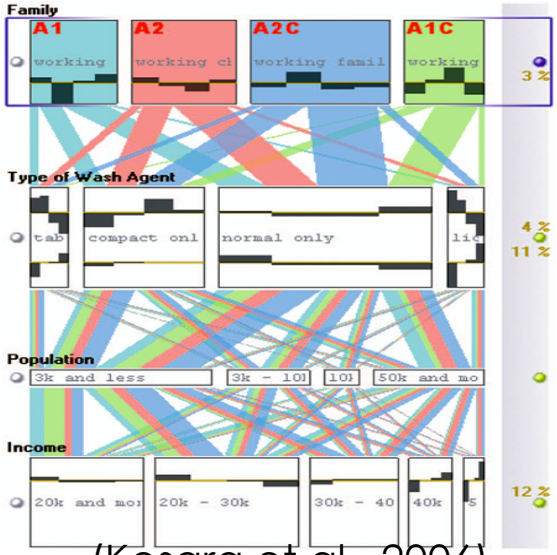


(Elmqvist & Fekete, 2010)

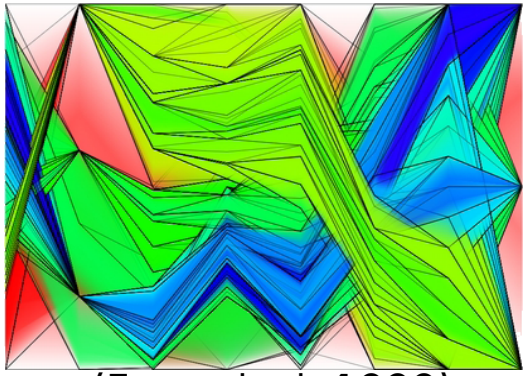


(Yang et al., 2003b)

Parallel Coordinates

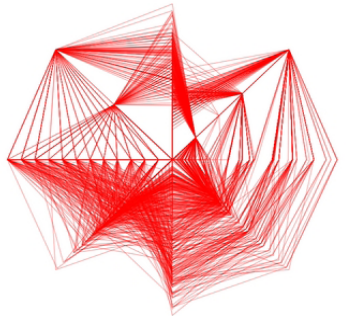


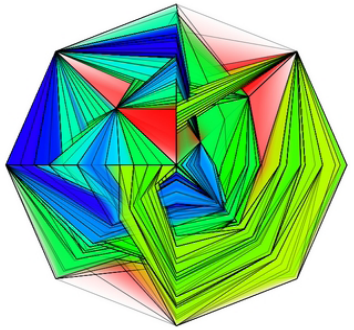
(Kosara et al., 2006)

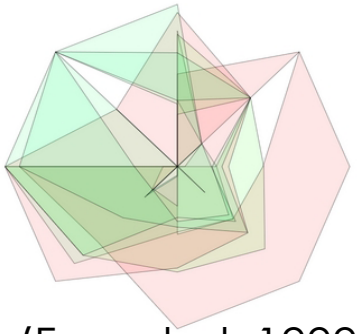


(Fua et al. 1999)

Star Plots

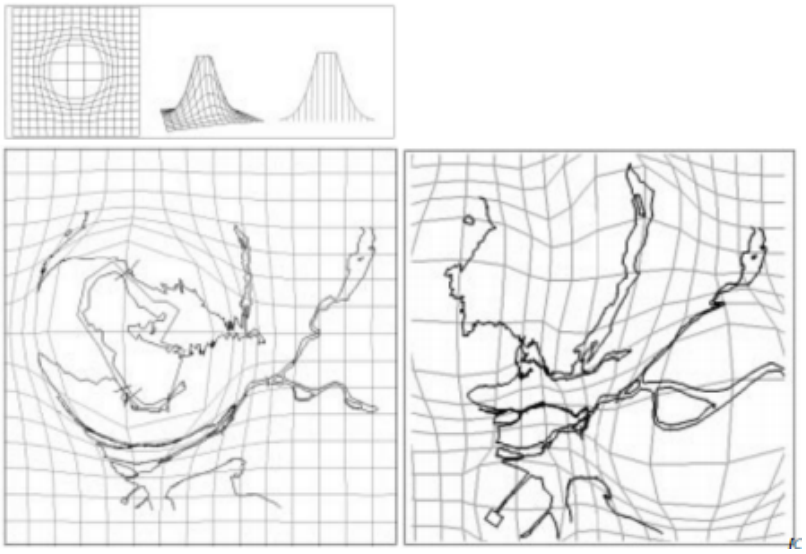


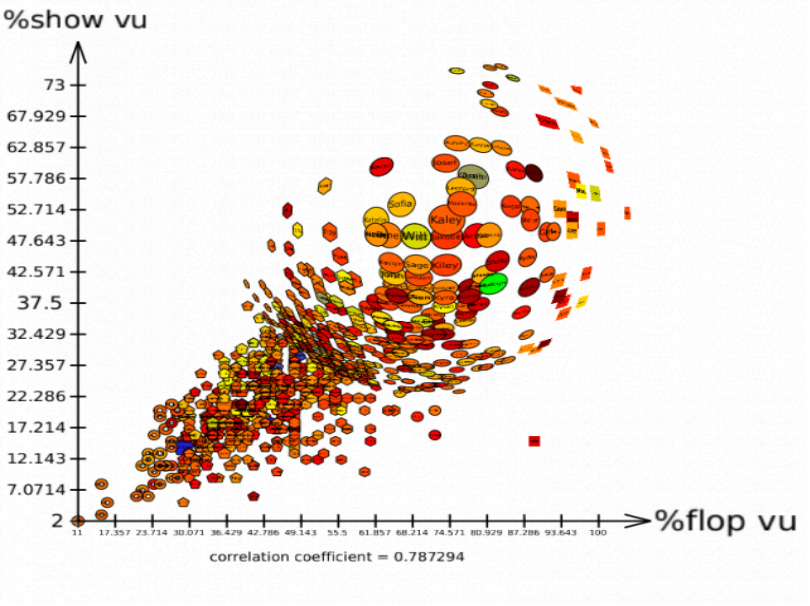


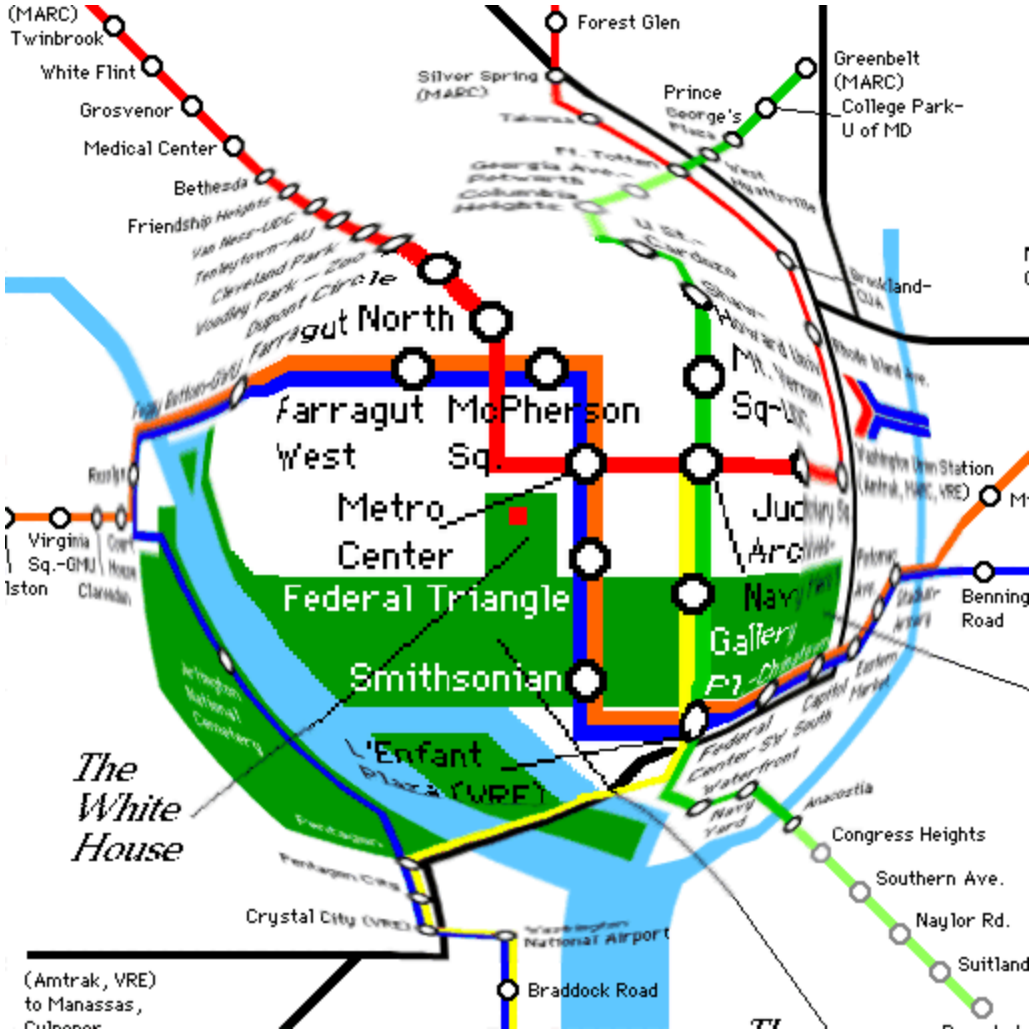


(Fua et al. 1999)

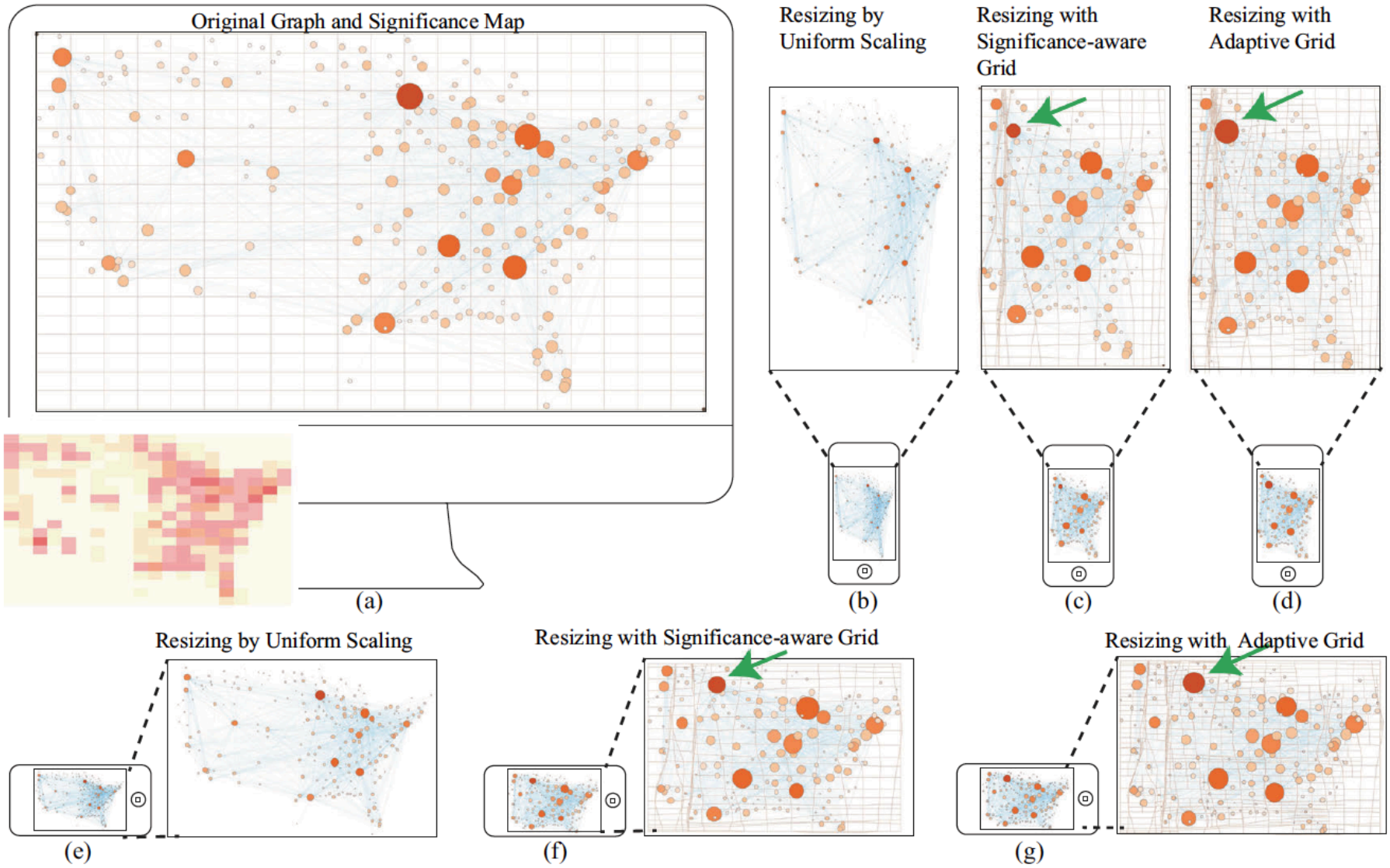
Technique (3) : Distortion



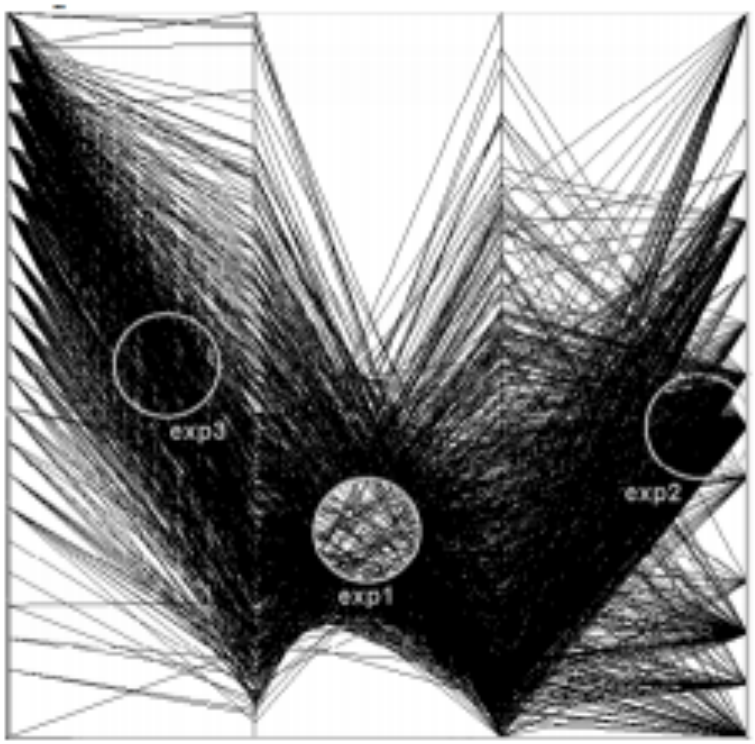




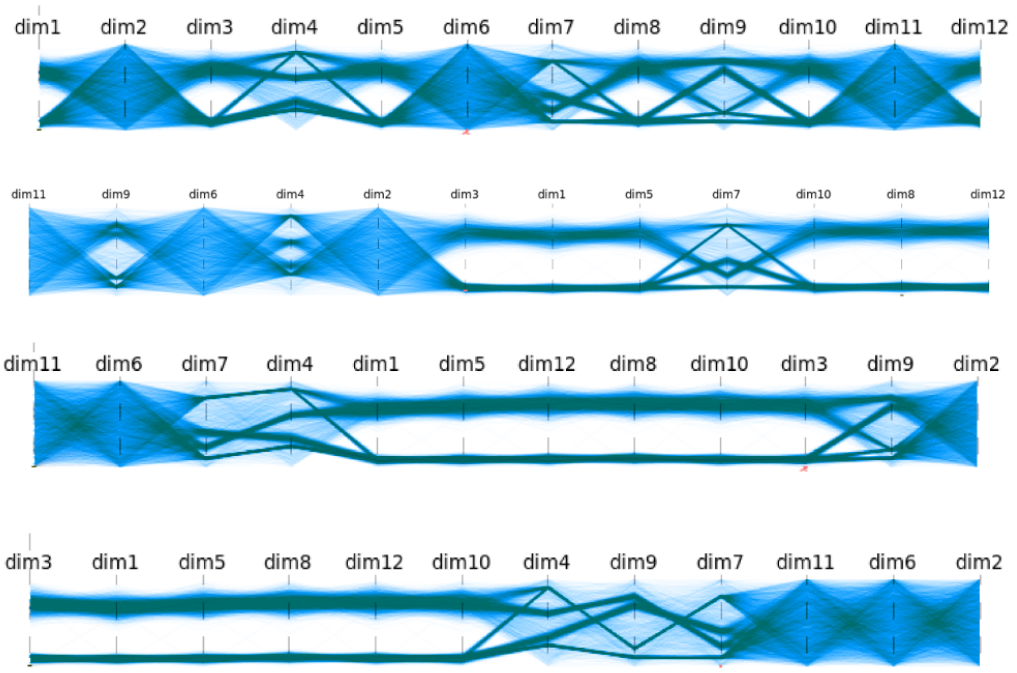
Techniques (3) : Distortion



Technique (4) : Clutter Reduction



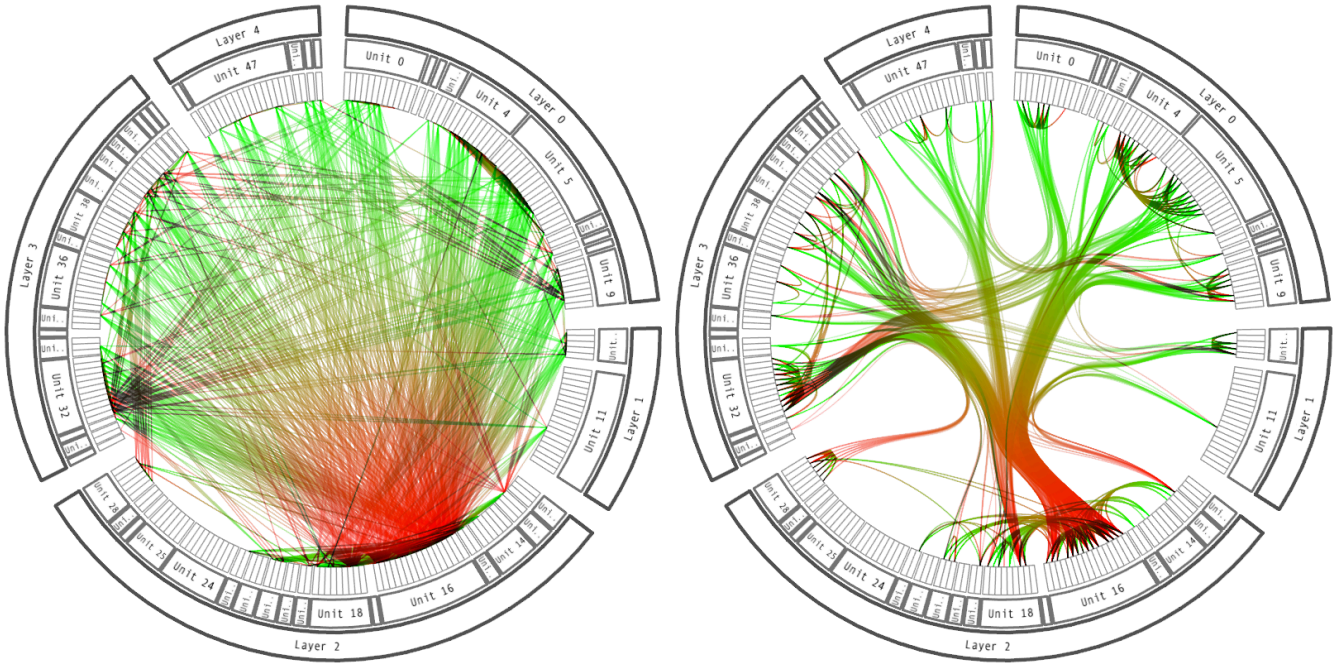
Sampling



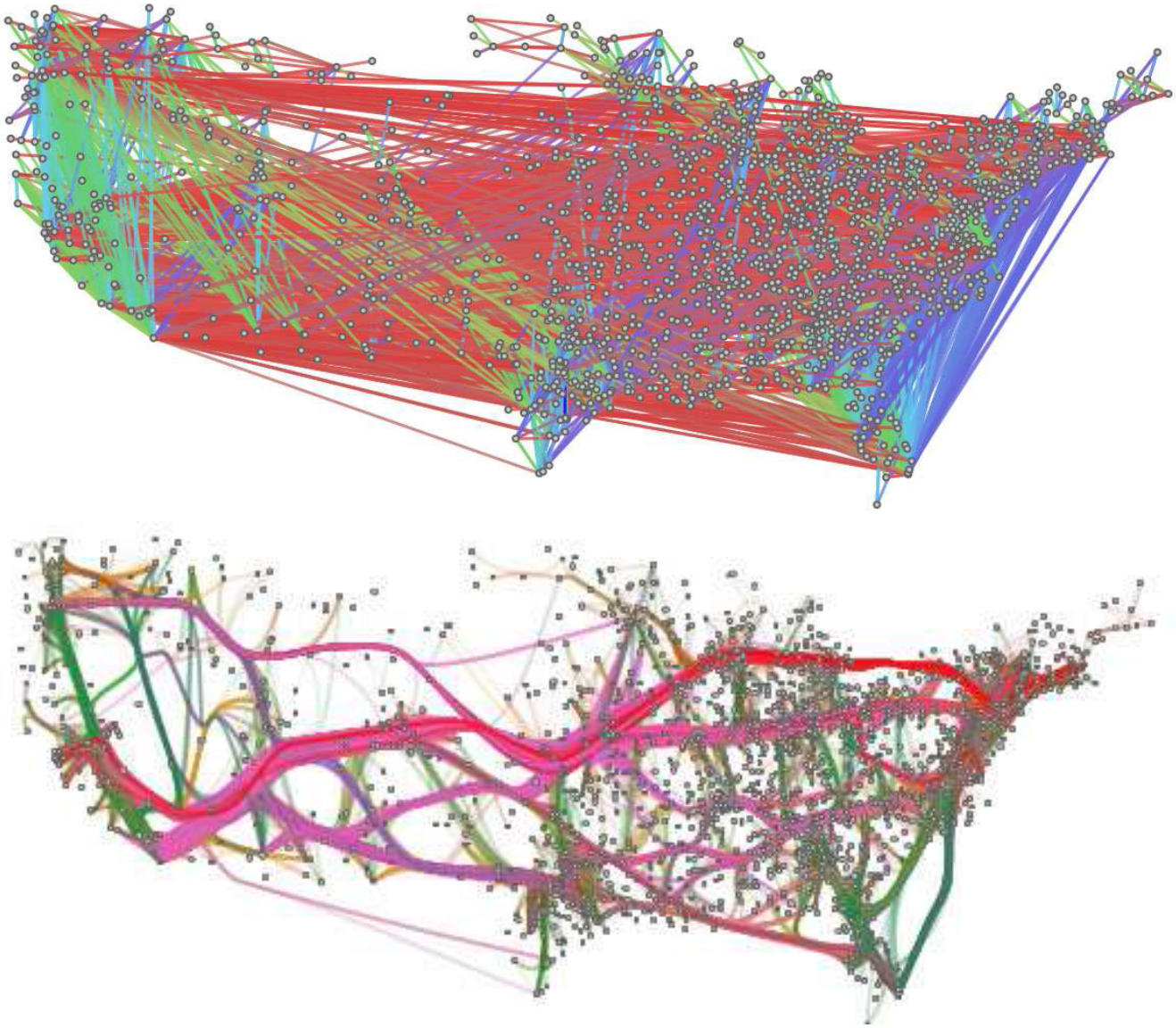
Reordering

Technique (4): Clutter Reduction

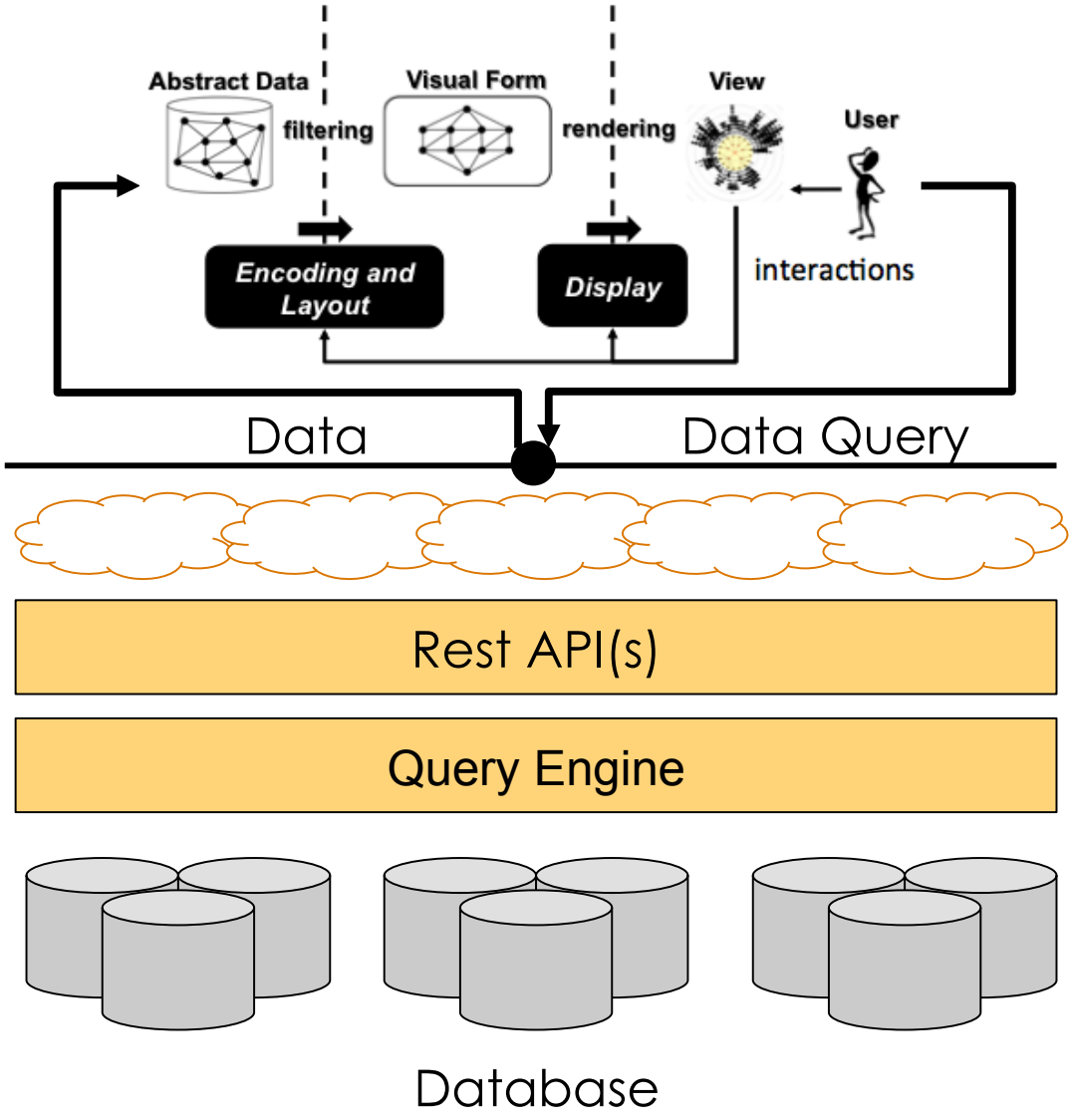
Edge Bundling



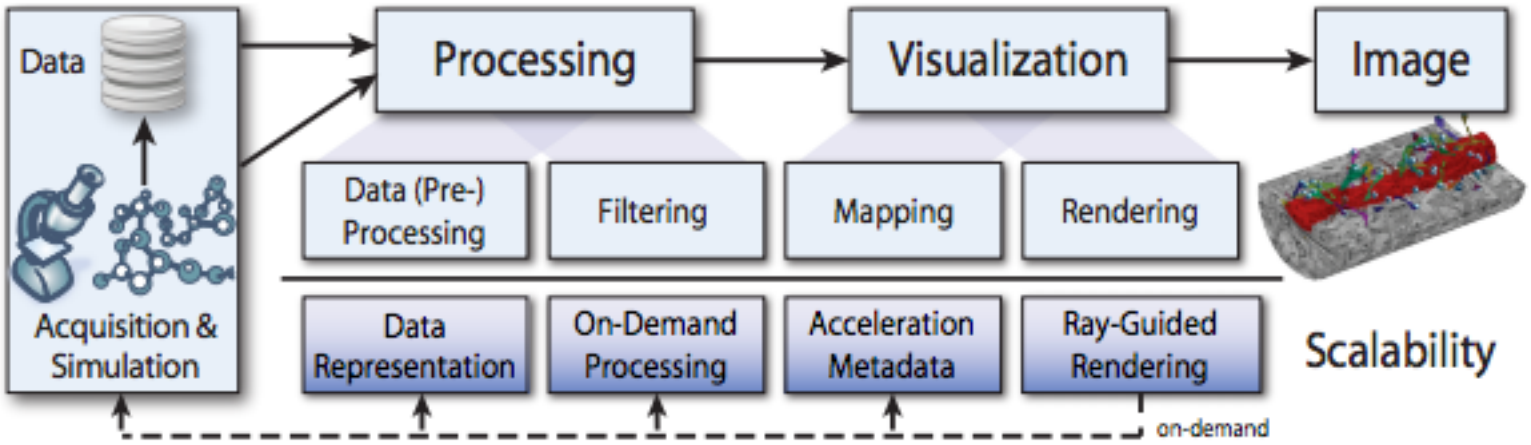
Technique (4): Clutter Reduction



Technique (4): Query based Visualization



Technique (5): Parallel Computing via GPU or GUGPU



A Survey of GPU-Based Large-Scale Volume Visualization, EuroVis, 2014

Graph & Network Visualization

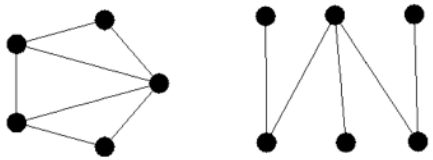
Methods of Graph Visualization

- Drawing Graphs with Special Structure:
  - Planar Graphs
    - Orthogonal paths for edges
    - K-sided convex polygons for each cycle of length k
  - Trees
    - Layering of Edges
    - Horizontal and vertical edges (hv-layout)
    - Radiate outward on concentric circles (radial layout)
- Drawing Graphs Using Analogies to Physical Systems

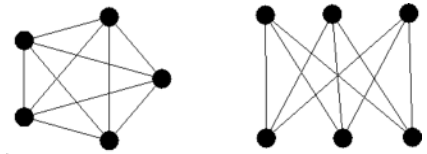
Tips for Graph Drawing

- Minimize Edge Crossing
  - Planar Graph and Planar Drawing
  - Graph Planarity Test
    - $O(n)$ algorithm at best
    - Intuitive method: test for subdivision of K5 or K3,3
- Minimize and Uniform Edge Length
  - Minimize edge length
    - Minimize the total edge length
    - Minimize the maximum of edge length
  - Uniform edge length (for unweighted graph)
    - Minimize the variance of edge length
    - For weighted graph: edge length proportional to

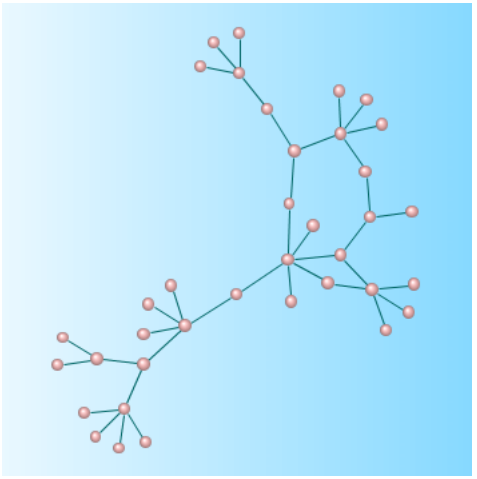
Planar Drawings



Non-Planar Drawings



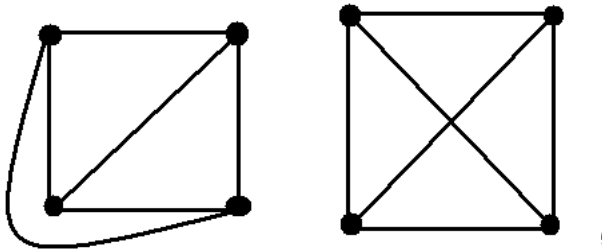
Uniform Edge Length Drawing



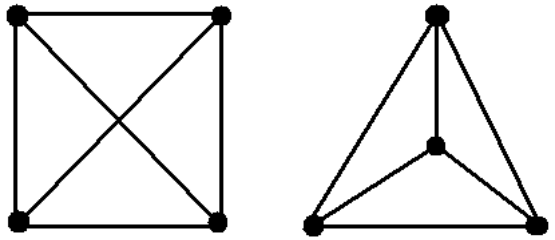
Tips for Graph Drawing

- Minimize and Uniform Edge Bends
  - Minimize total edge bends
  - Minimize the variance of edge bends
- Maximize Angular Resolution
  - Maximize the smallest angle between two edge
- Minimize Aspect Ratio
  - Adapt to the real-world screens
- Graph Symmetry

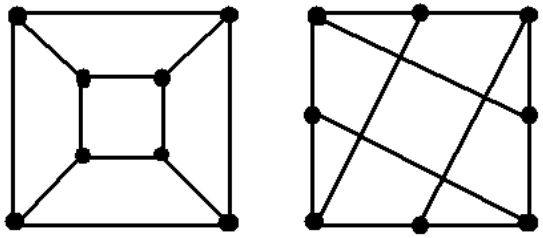
Graph with and without edge bend



Graph with different angles (45°and 30°)

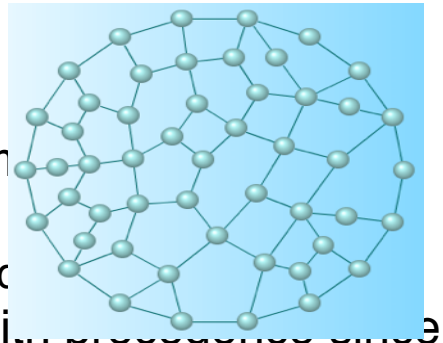


Symmetrical and non-symmetrical graph

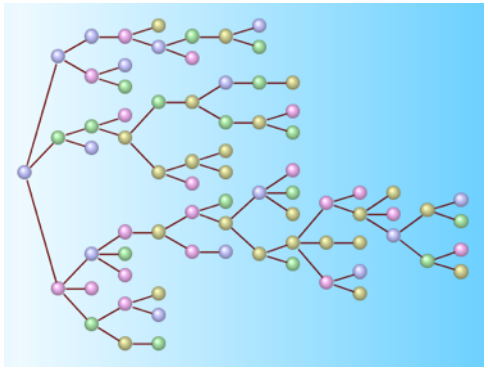


Graph Drawings

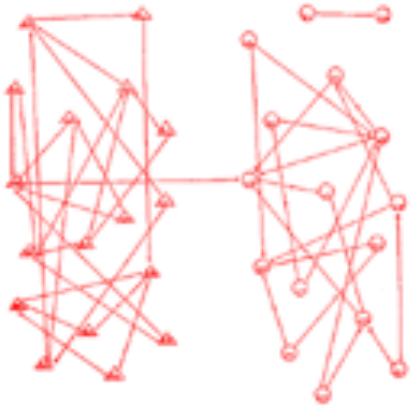
Various modern drawing approaches

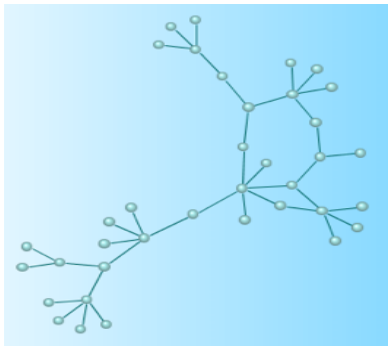


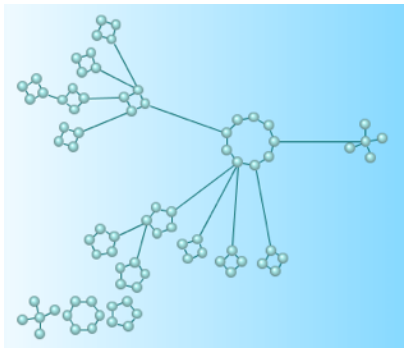
- Graph drawing without modern techniques
  - Hand-made images
  - Drawing with cluster/factor analysis, multidi
- Graph drawing approaches according to elaborately-
  - Each approach follows a list of aesthetics wi… precedence since:
    - Aesthetics conflict with each other
    - Hard to design algorithm to fit all the aesthetics



Hand-made drawing of friendships

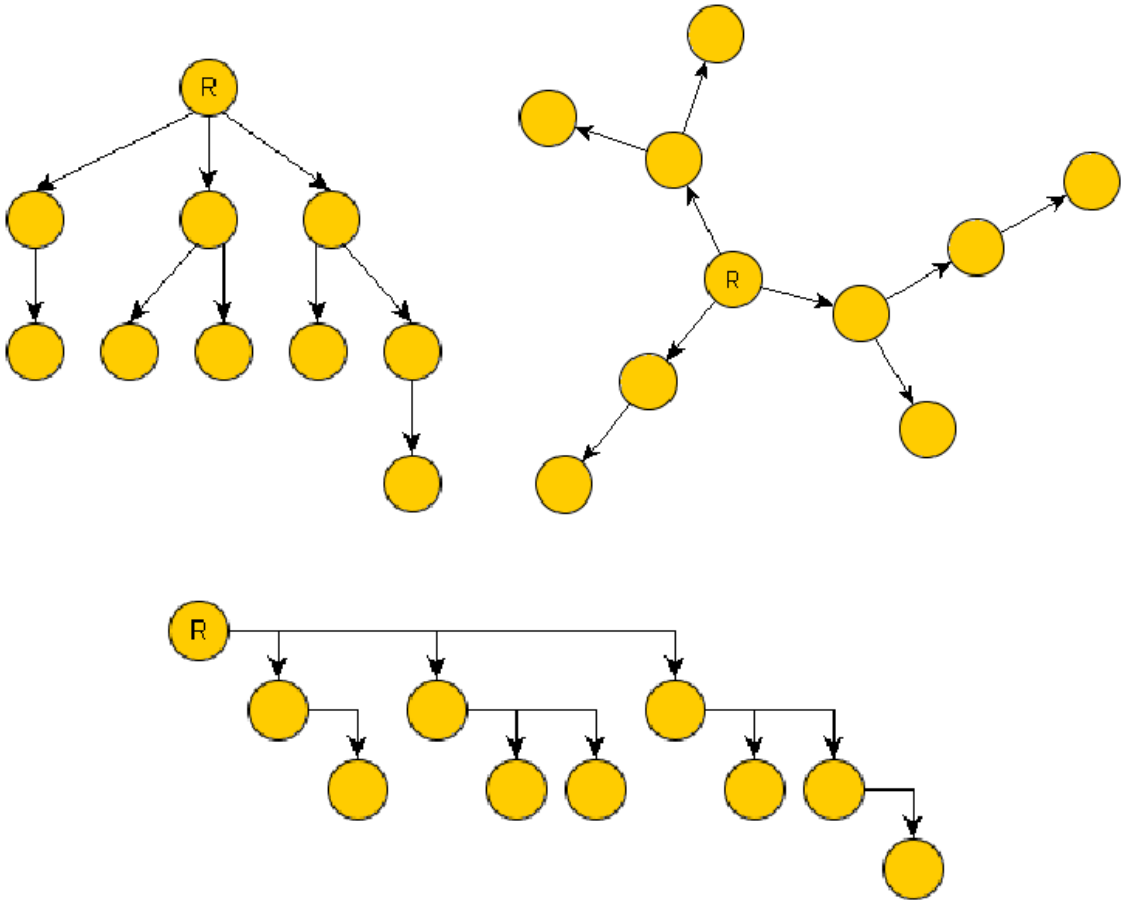






Examples of Tree Drawing

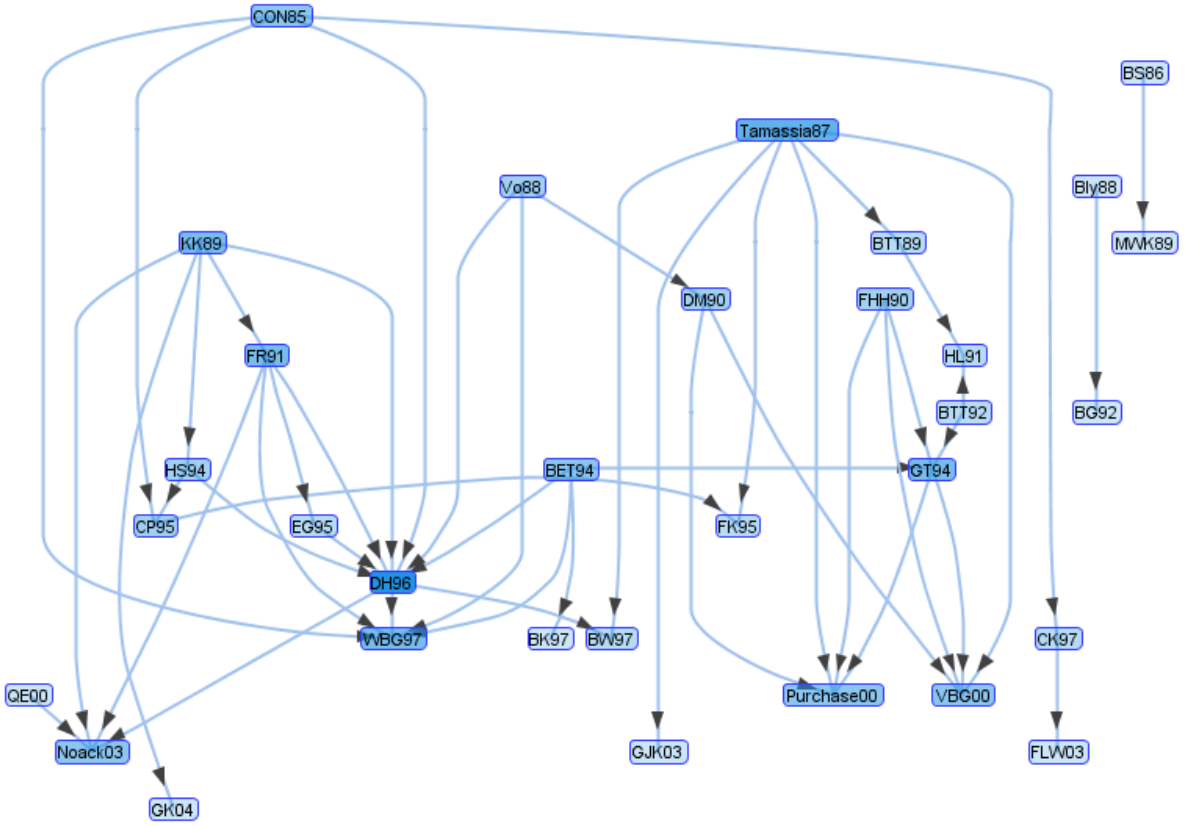
- Layering of Edges
- Horizontal and vert
- Radiate outward or



Hierarchical Graph Drawing

- Hierarchical Graph Drawing
  - For dependency relationships abstracted as acyclic digraph
  - Sugiyama-style 3-step layout approach

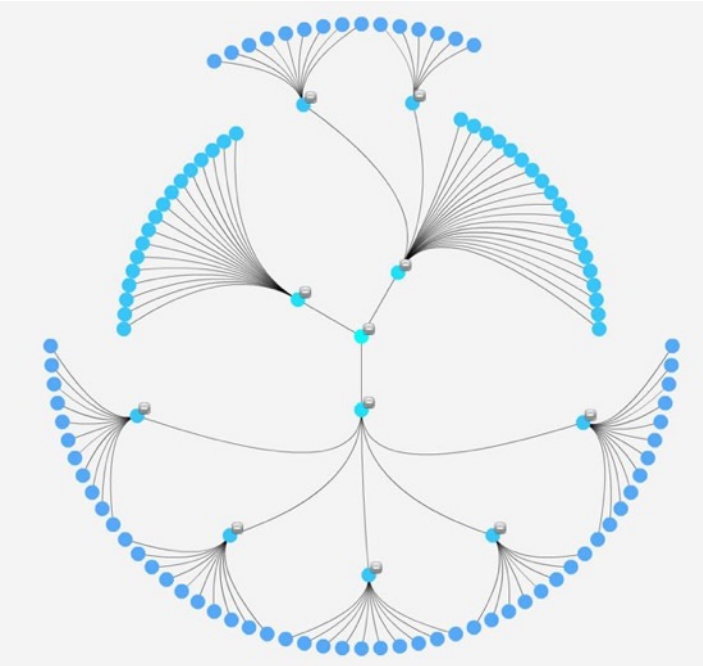
Hierarchical drawing of a citation network



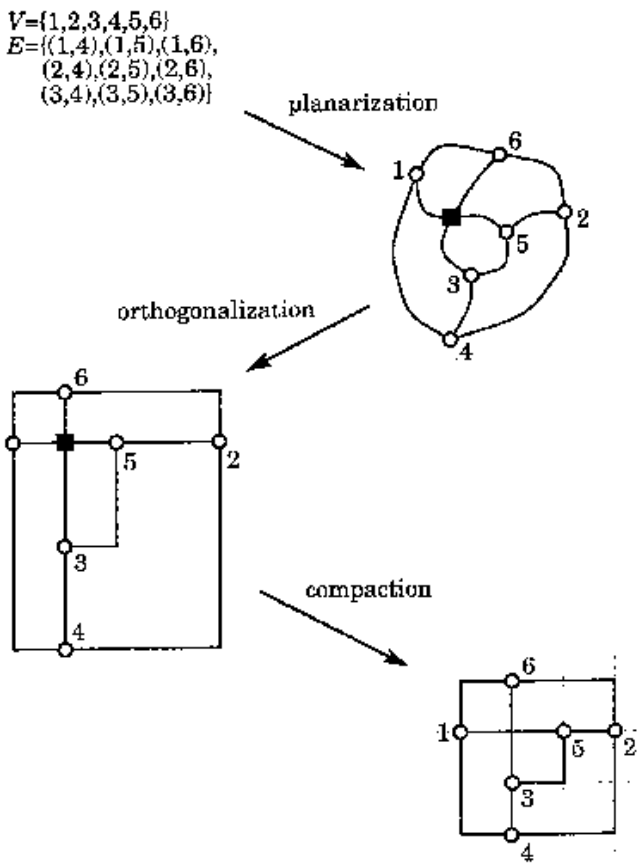
Tree Drawing

- Orthogonal Graph Drawing
  - For circuit placement in VLSI and PCB board layout
  - Topology-Shape-Metrics approach
- Radial Graph Drawing
  - For rooted tree drawing
  - Foldable for hierarchical graphs

Radial Drawing



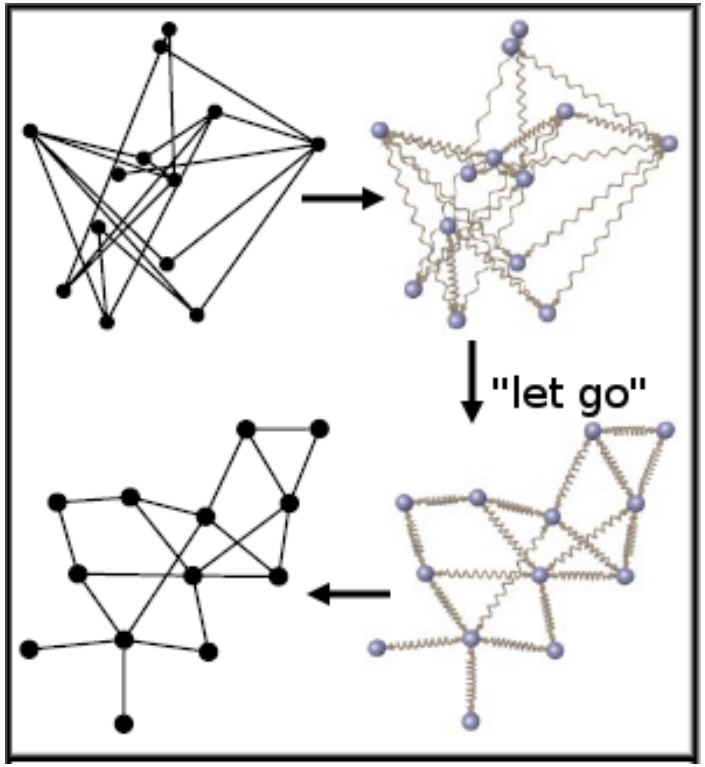
Orthogonal Drawing



Force-Directed Graph Drawing

- Force-directed Graph Drawing
  - For undirected straight-line drawing
  - Simulate physical systems with forces, the final graph layout reaches the local system equilibrium (i.e. a local minimization of energy)
  - Generate graphs with good node uniformity and symmetric
  - Two types of force-directed algorithms
    - Spring Embedder by Eades and the seminar work by Fructerman and Reingold
    - Kamada and Kawai algorithm, trying to maintain graph theoretic distances between nodes, the recent solution by stress majorization is popular
  - Multi-Scale force-directed algorithms for large graphs

Force-directed layout with spring embedder



Algorithm

- Assign a measure of 'force' to each vertex.
- Iteratively adjust the position $p_v = (x_v, y_v)^T$ until the forces on each have converged.
- A standard setup defines the force $F = (F_x, F_y)^T$ on v in the x direction to be:

$$F_x(v) = \sum_{\{u,v\} \in E} k_{uv}^{(1)} (\|p_u - p_v\| - len_{uv}) \frac{x_v - x_u}{\|p_u - p_v\|} + \sum_{(u,v) \in V^2} k_{uv}^{(2)} \frac{x_v - x_u}{\|p_u - p_v\|^3}$$

where

$k_{uv}^{(1)}$ is a spring stiffness

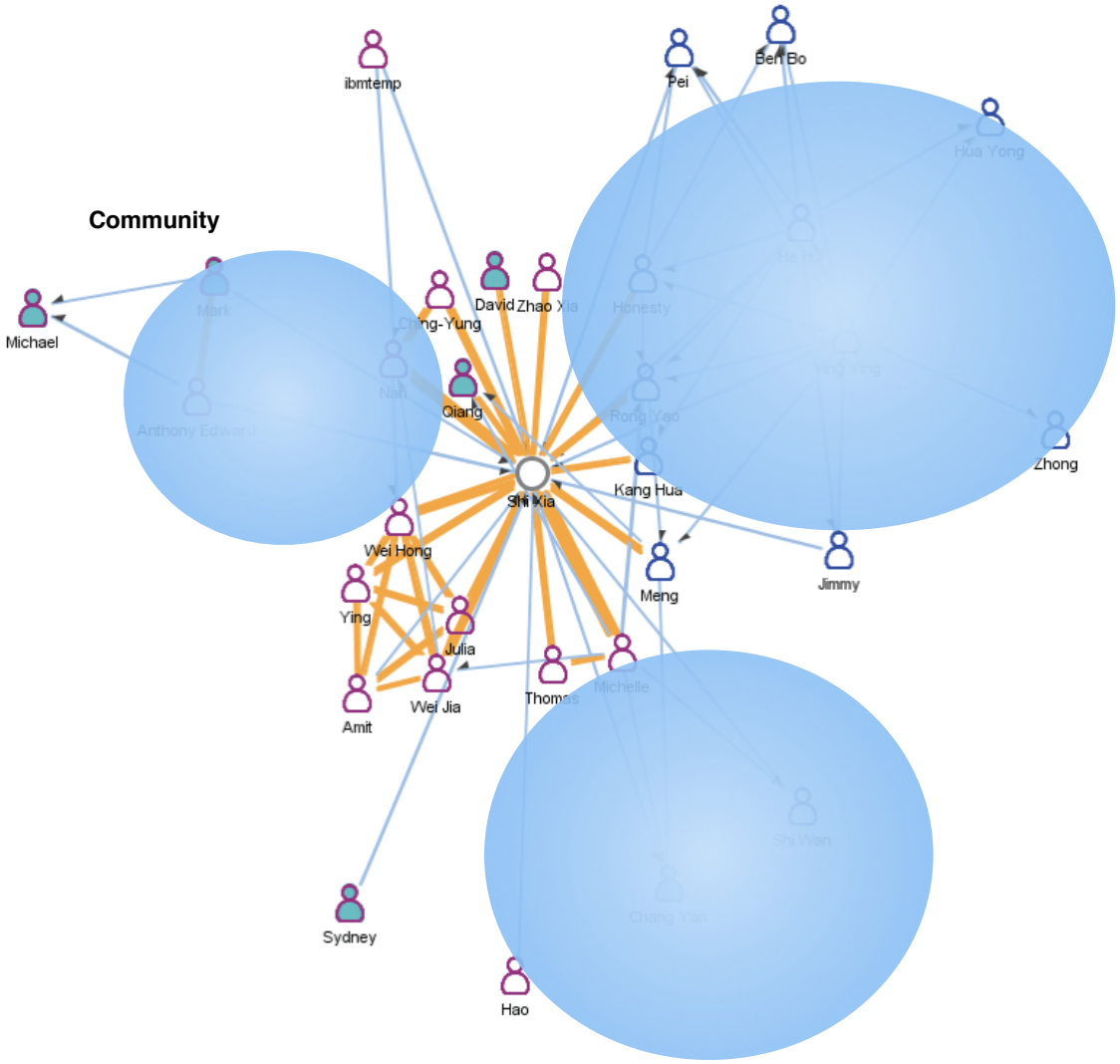
$k_{uv}^{(2)}$ is s strength of electric repulsion stiffness

$\|\cdot\|$ is Euclidean distance len_{uv} is a parameter for desired dist.

Social Network Visualization -- Metaphors

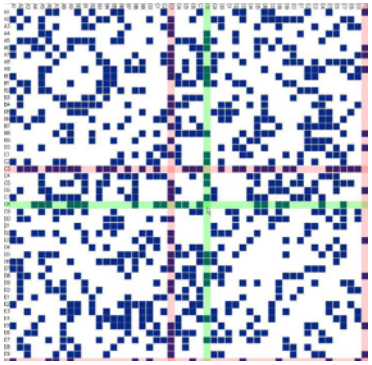
- Basic Metaphors with Graph
  - Nodes ~ People
    - Visual attribute mapping:
  - Edges ~ Relationships
    - Visual attribute mapping:
- Advanced Metaphors with Graph
  - Shading ~ Community (clust
- Other Non-Graph Metaphors
  - Adjacency Matrix

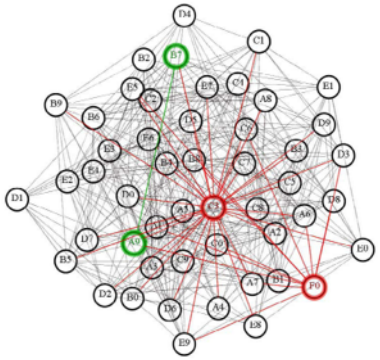
Typical Social Network Visualization with Graph

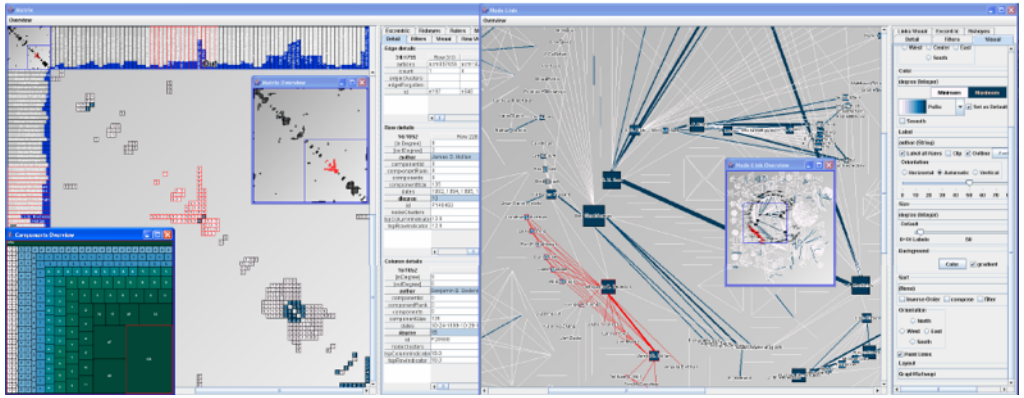


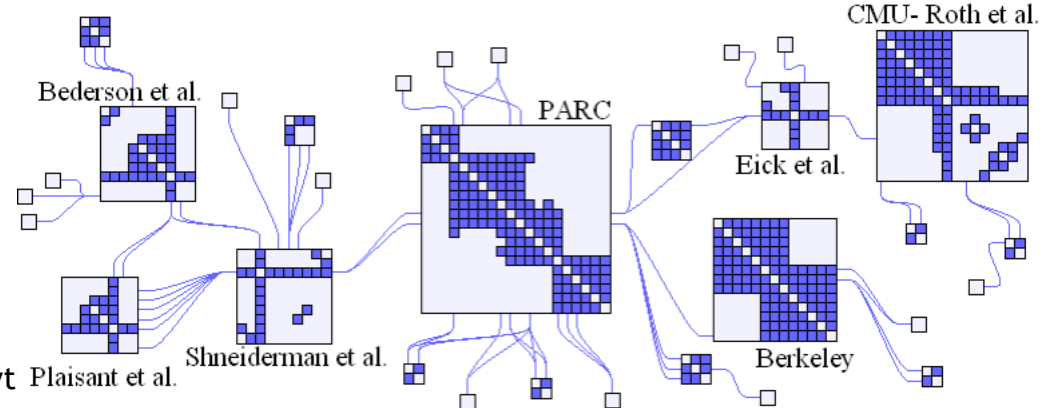
Social Network Visualization -- Matrix

- Adjacency Matrix as an Alternative of Node-Link Graph
  - Good for huge, scale-free graph
  - Hybrid visualization
    - MatrixExplorer: synchronized view of both adjacency matrix and node-link graph
    - NodeTrix: combined visualization for social network with communities
- User Studies on Performance
  - Matrix is better than node-link graph in several major tasks when node number is over 20
  - Negatives: the mass people (information consumer) can only perceive simple visualization such as maps, line-charts and graphs





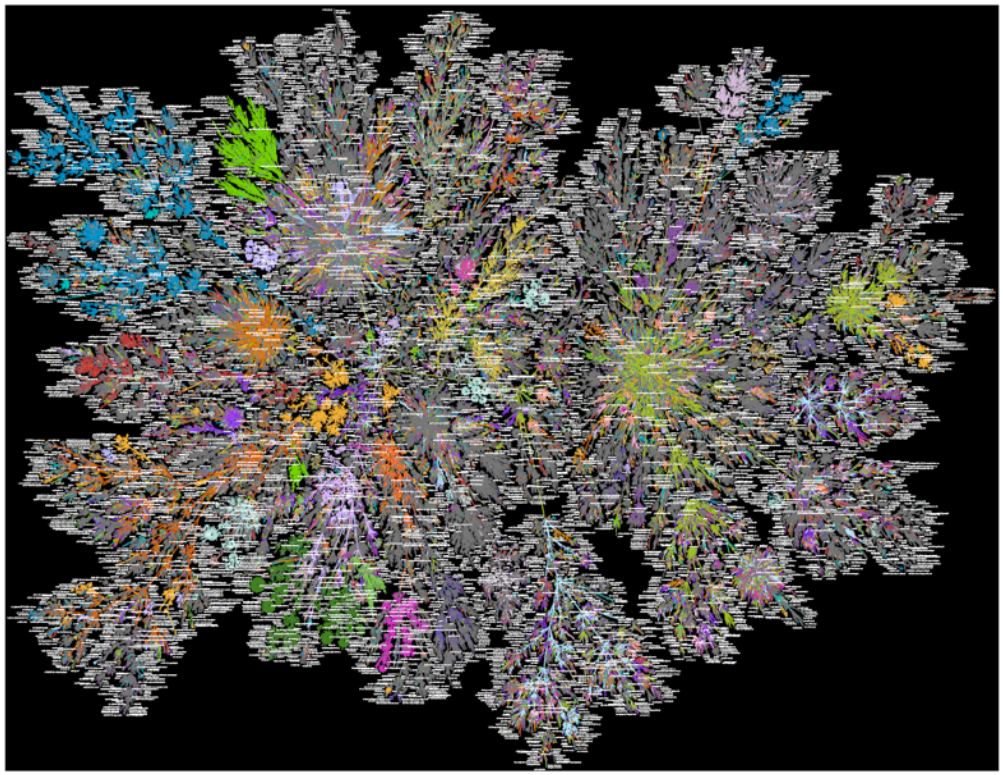




Social Network Visualization – Huge Structural Network

- Social Networks could be quite large, posing challenge to visualization
  - Traditional layout algorithms do not scale, with $O(n^2) \sim O(n^3)$ computation complexity
  - With multi-scale fast layout algorithm, huge graph can be drawn with degraded quality, but raises severe visual clutter
- Social Networks are intrinsically clustered, with well-known feature of scale-free, small-world
  - Visualization should reveal this structural information

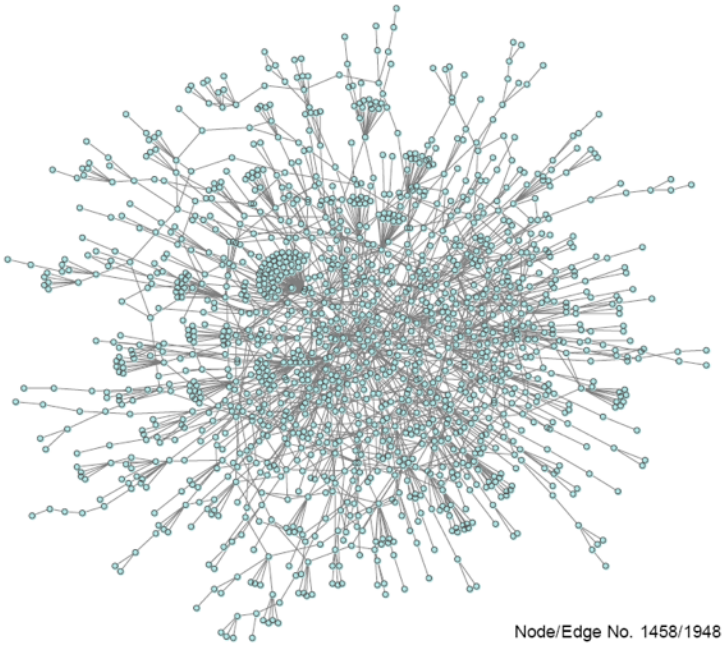
A Huge Graph Visualization with Multi-Scale Layout Algorithm (by AT&T)

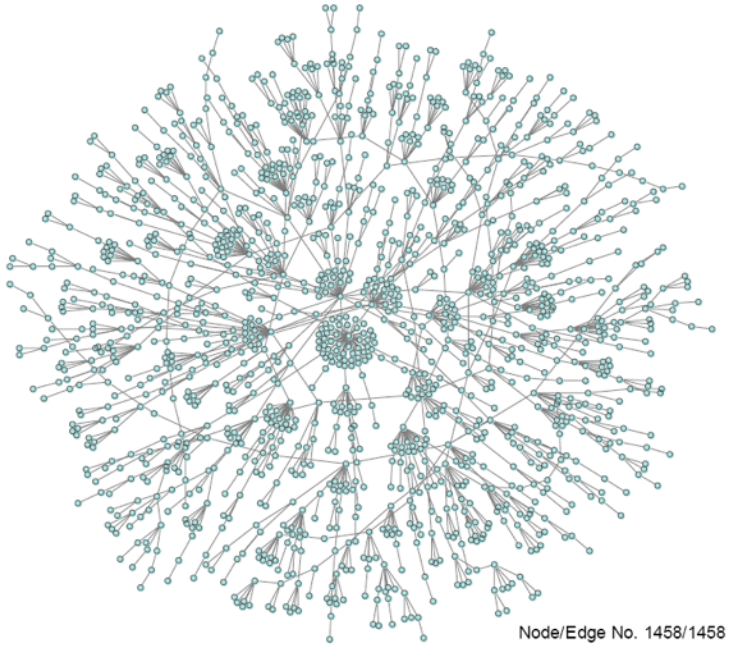


Social Network Visualization – Huge Structural Network

- Graph Skeleton Visualization by Edge Pruning
  - Filter edges by edge betweenness centrality from low to high

Edge pruning with betweenness centrality

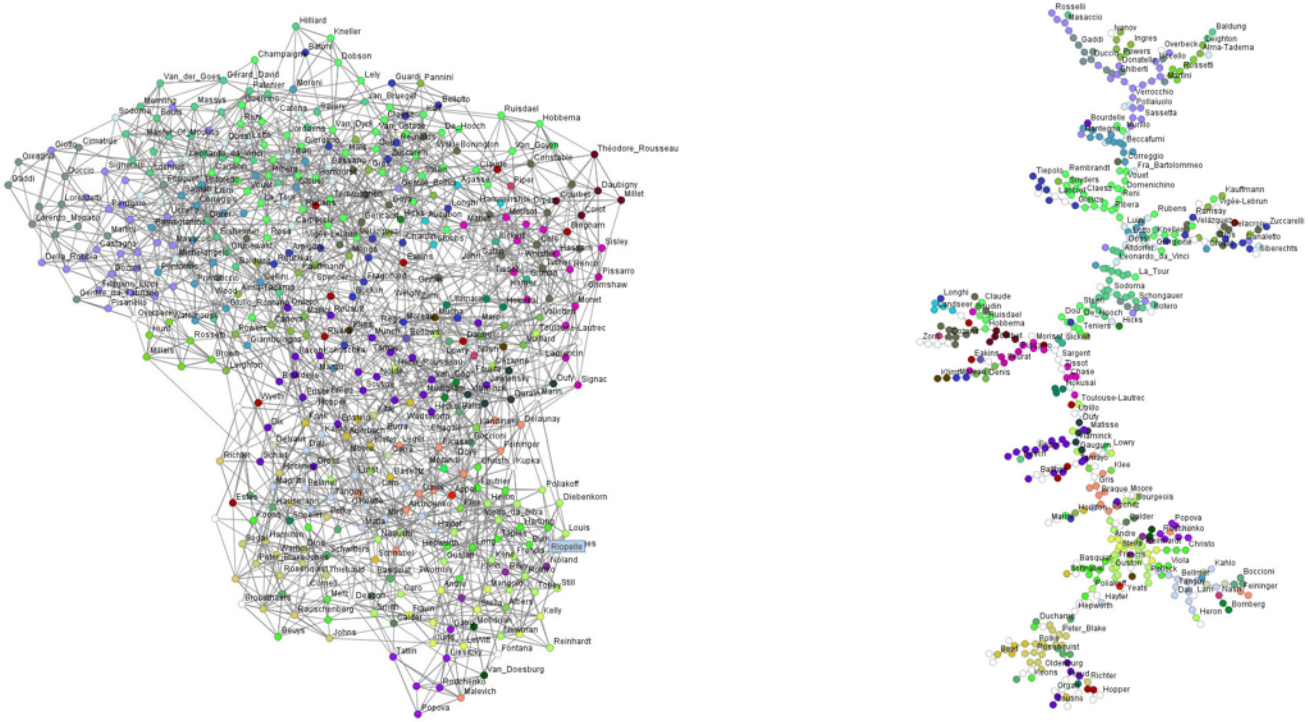




Social Network Visualization – Huge Structural Network

- Graph Skeleton Visualization by Edge Pruning
  - Construct a minimal spanning tree (MST) or path-finder network from the original huge graph
  - Graph skeleton visualization highlights the topology and intrinsic structure of huge graph by reducing visual clutter

Skeleton extraction by constructing MST



Case study – Mapping 'Science'

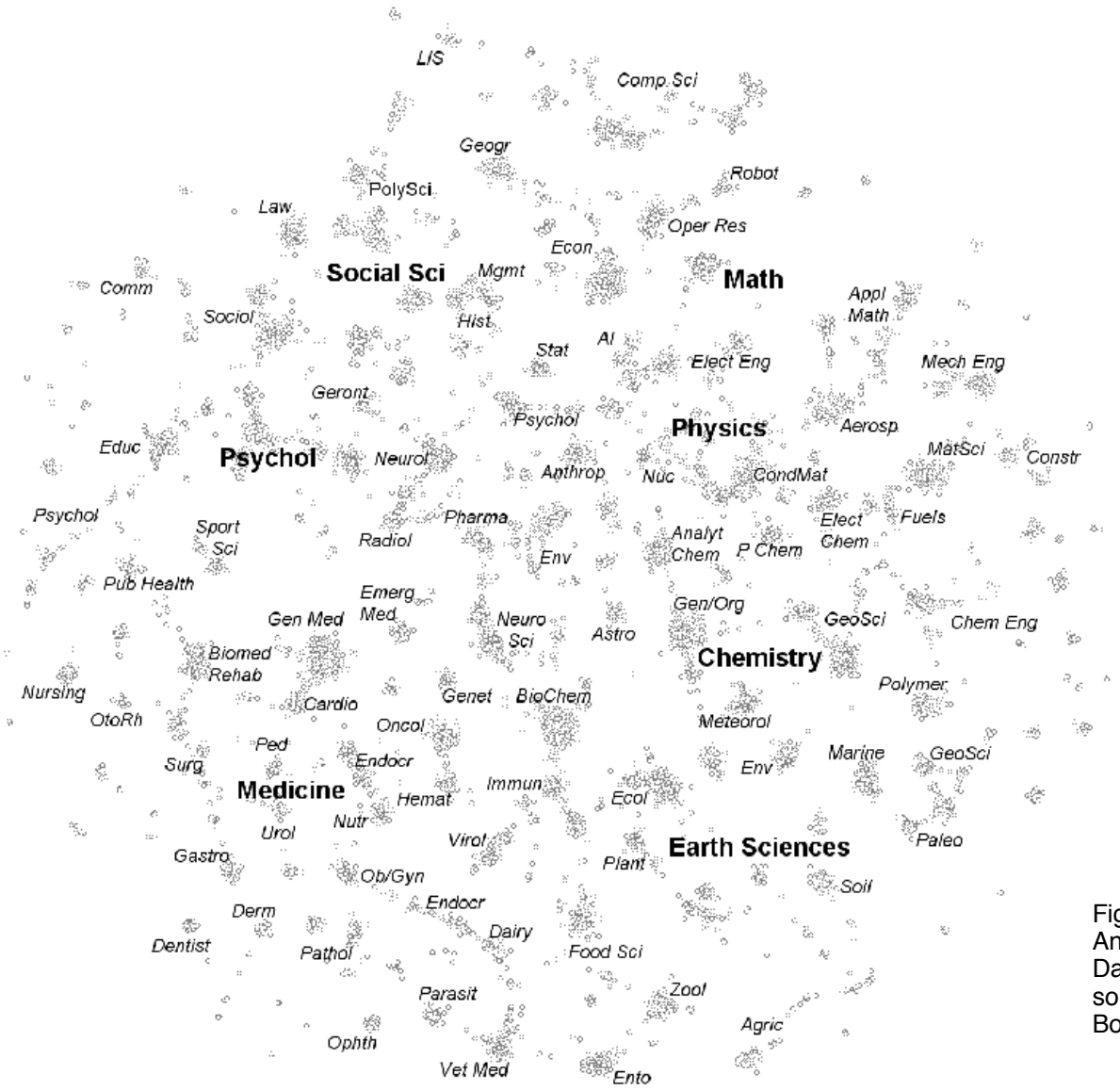


Fig 3.5 in Statistical Analysis of Network Data. Original source: Kevin Boyack.

Case study – Mapping 'Science'

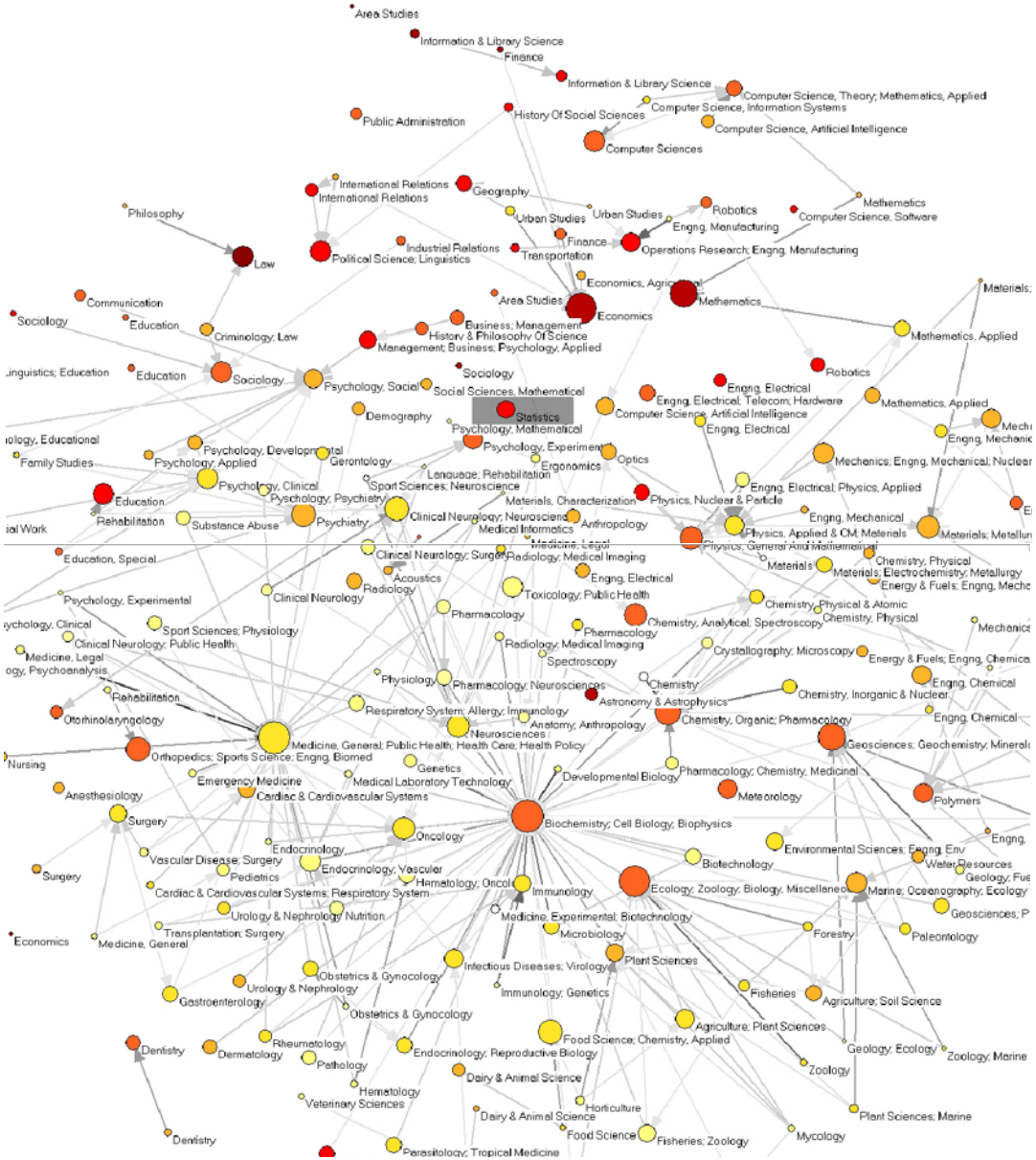
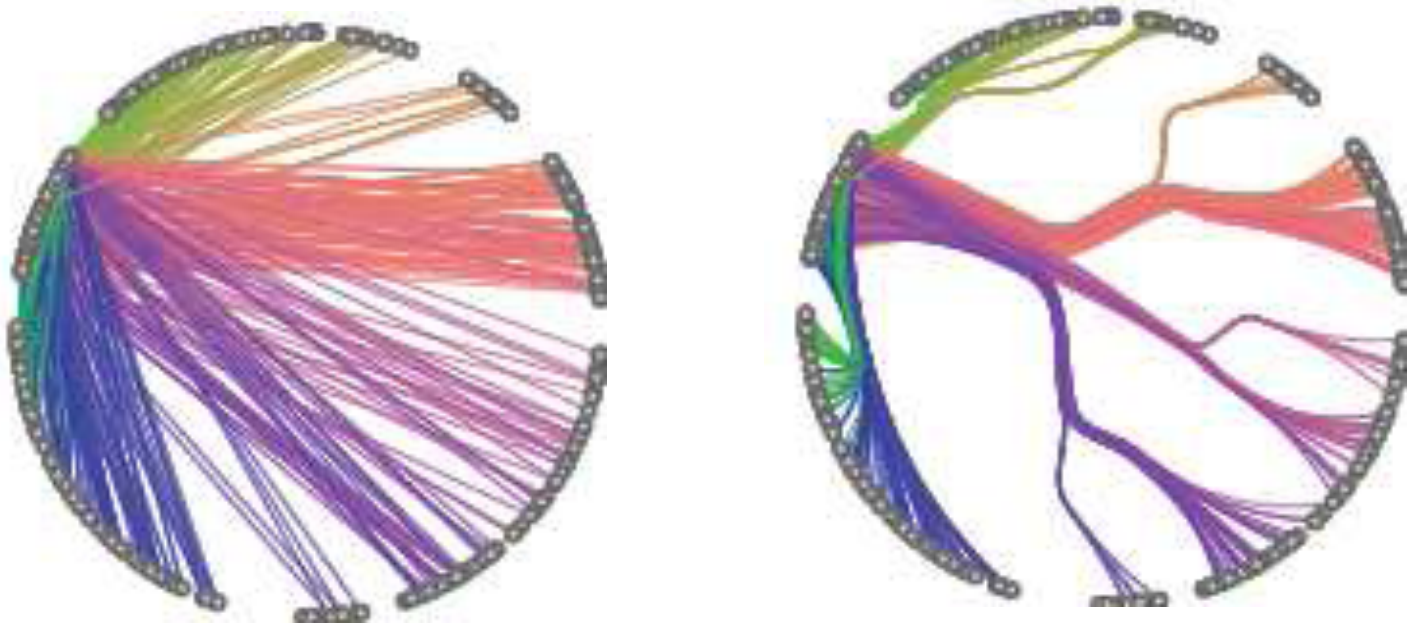


Fig 3.6 in Statistical Analysis of Network Data. Original source: Kevin Boyack.

Social Network Visualization – Huge Structural Network

- Graph Visual Clutter Reduction through Edge Bundling
  - Edge bundling steps
    - Generate control mesh
    - Cluster the edges
    - Let the clustered edge traverse through the same set of control points

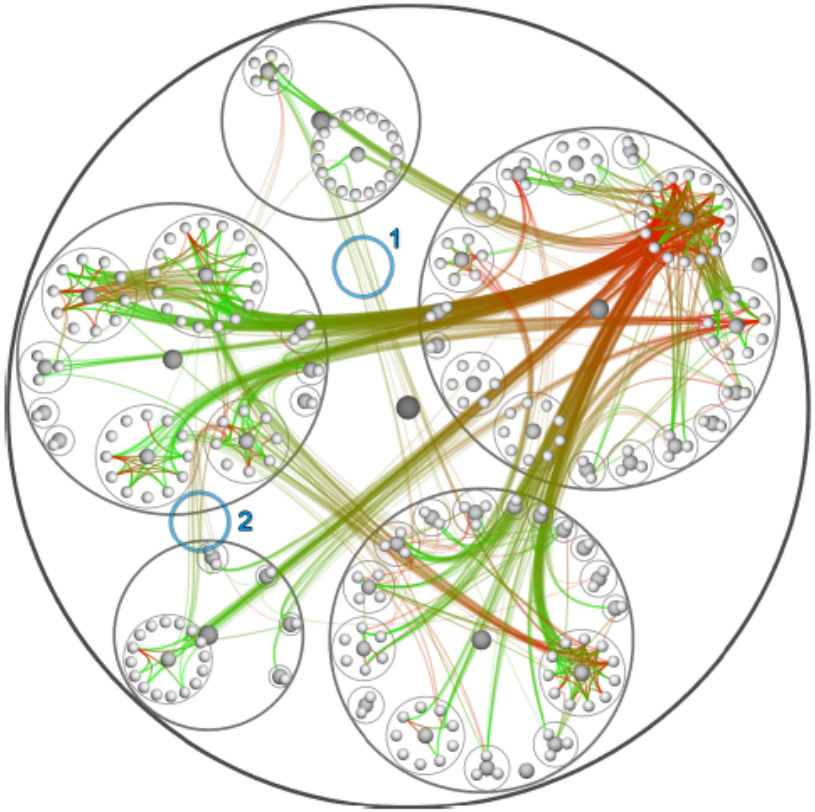
The effectiveness of edge bundling



Social Network Visualization – Huge Structural Network

- Graph Visual Clutter Reduction through Edge Bundling
  - Hierarchical edge bundling
    - For hierarchically clustered graph
    - Maintain high-level topology information of edges

Hierarchical Edge Bundling

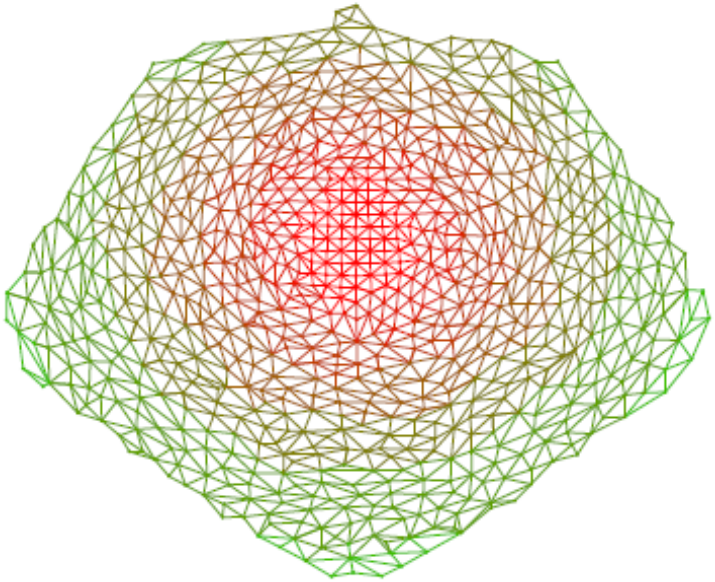


Social Network Visualization – Huge Structural Network

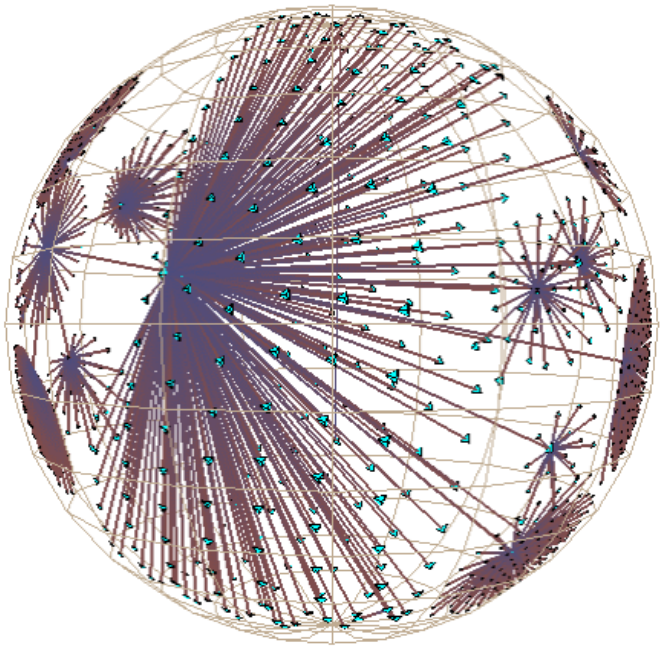
- Huge Graph Visualization through Advanced Interactions
  - Focus+Context: to highlight part of the network while leaving context
    - Topology Fisheye
    - Hyperbolic Tree

Focus + Context

Topology Fisheye



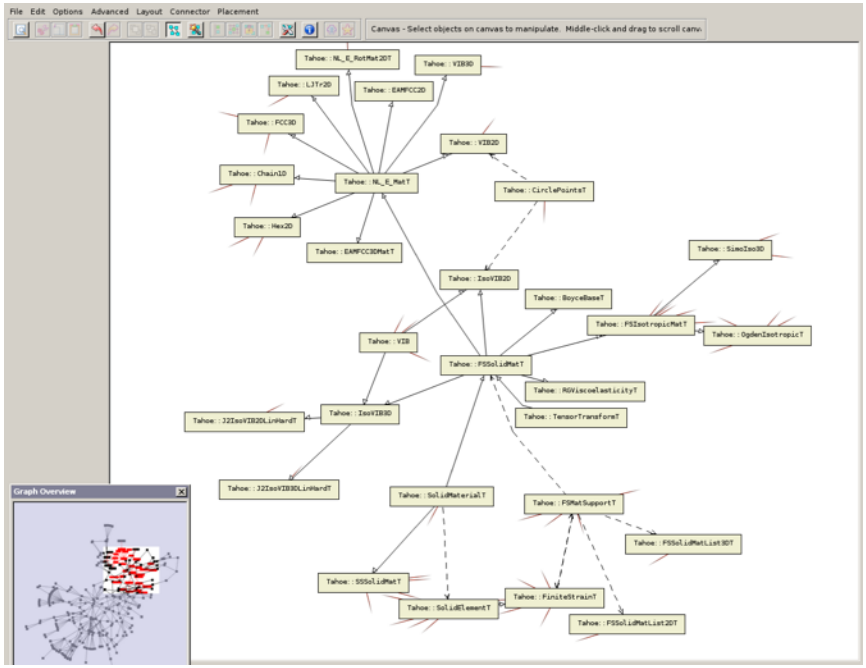
Hyperbolic Tree

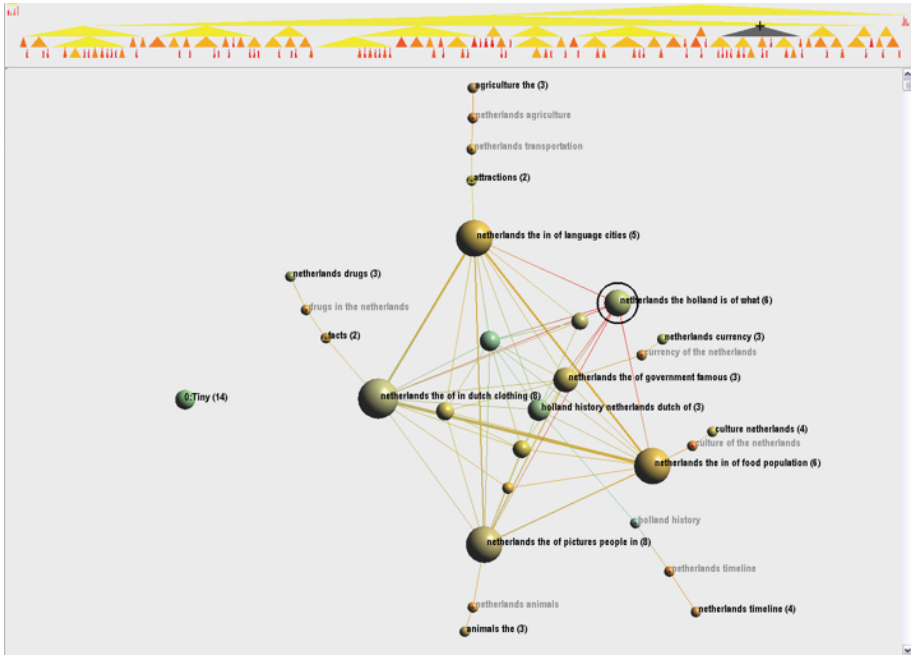


Social Network Visualization – Huge Structural Network

- Huge Graph Visualization through Advanced Interactions
  - Overview + Detail
    - Overview graph with lowered quality
    - Overview graph show hierarchy information
  - Traditional Zoom & Pan

Overview + Detail

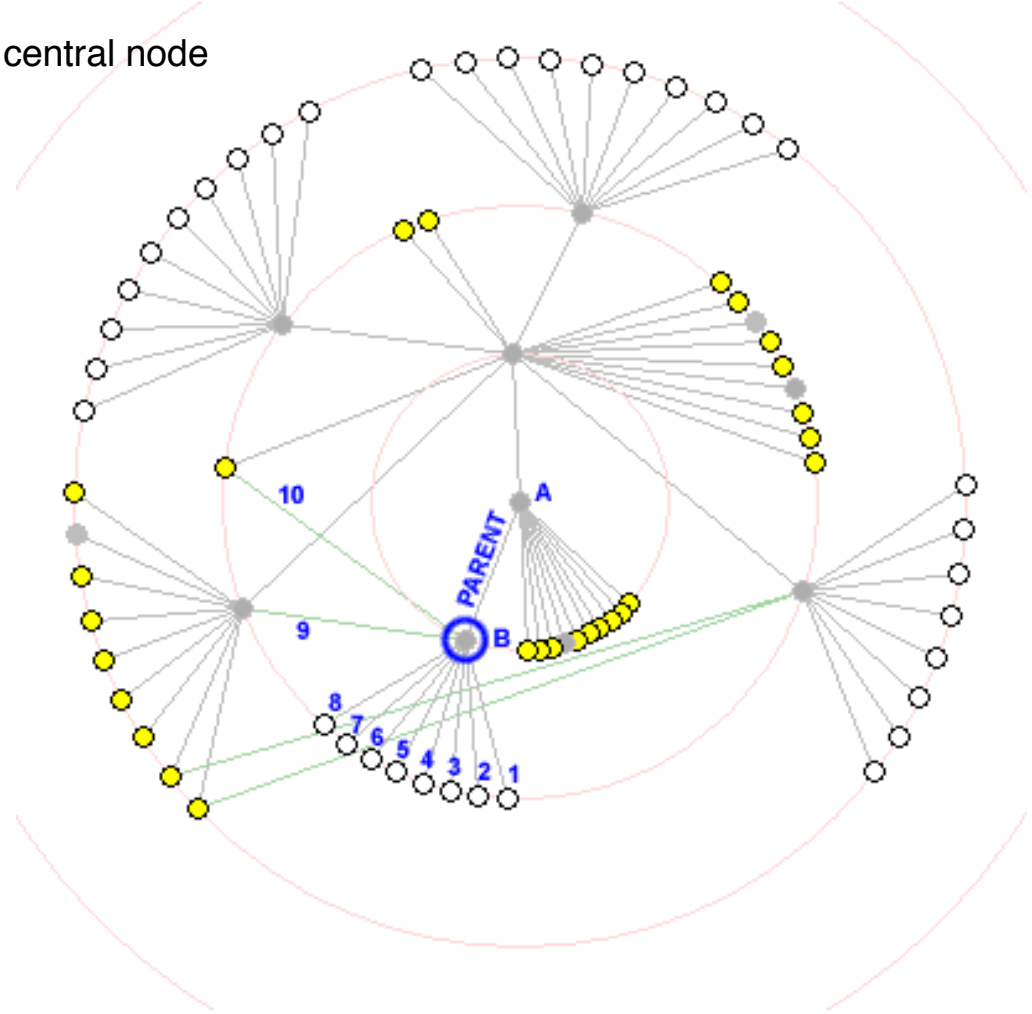




Social Network Visualization – Huge Structural Network

- Avoid huge graph with user-interest centric visualization
  - Ego-centric network visualization
    - Filter the huge graph by hop from the central node
    - Allow expansion with interactions

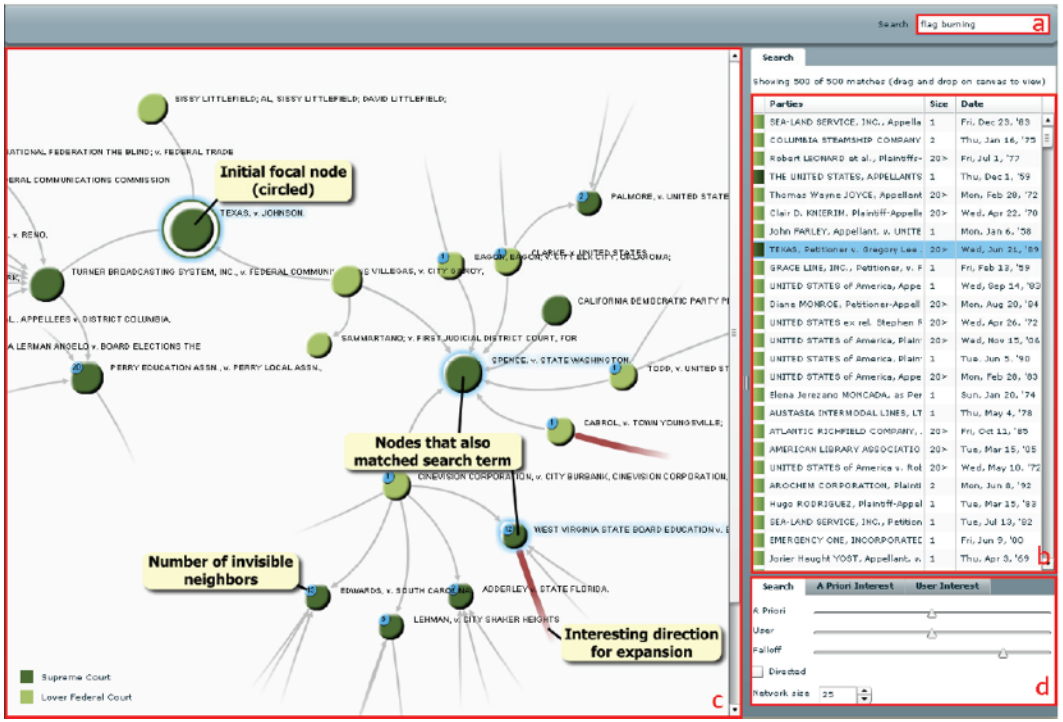
Radial graph layout of ego-centric network



Social Network Visualization – Huge Structural Network

- Avoid huge graph with user-interest centric visualization
  - Spring-board by Frank Van Ham
    - The “search, show context, expand on Demand” paradigm
    - Degree of Interest (DOI) function combining graph topology (distance to the central node) and user interest (search context)
    - Allow further navigation by expanding a subgraph

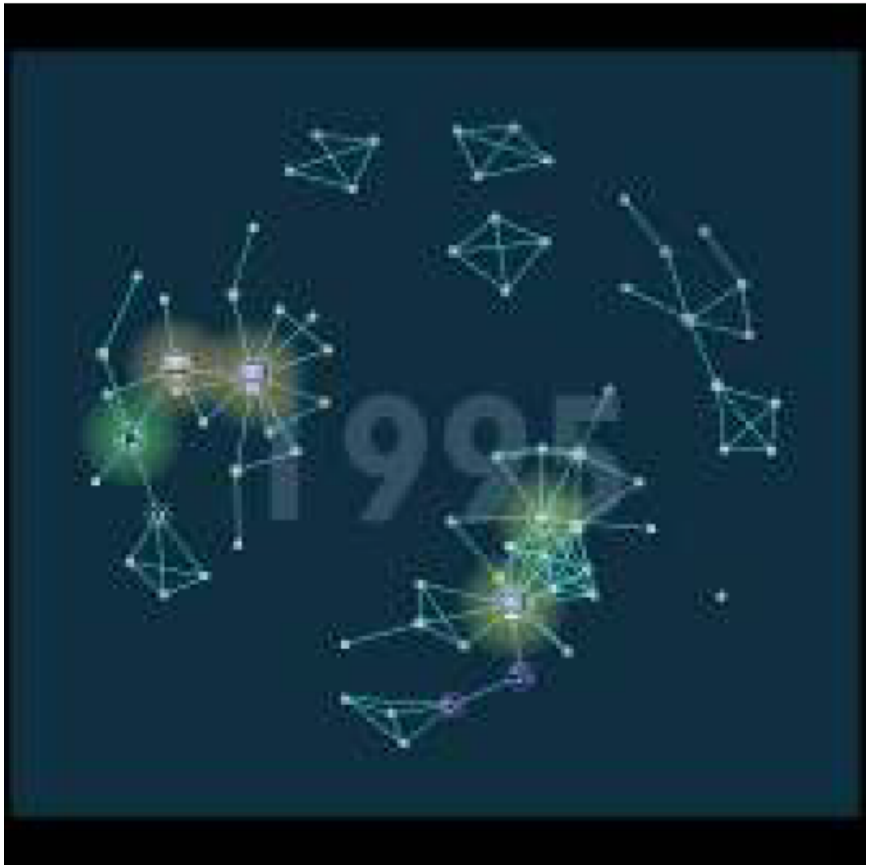
Springboard: visualize huge graph by user-interest



Social Network Visualization – Temporal Network

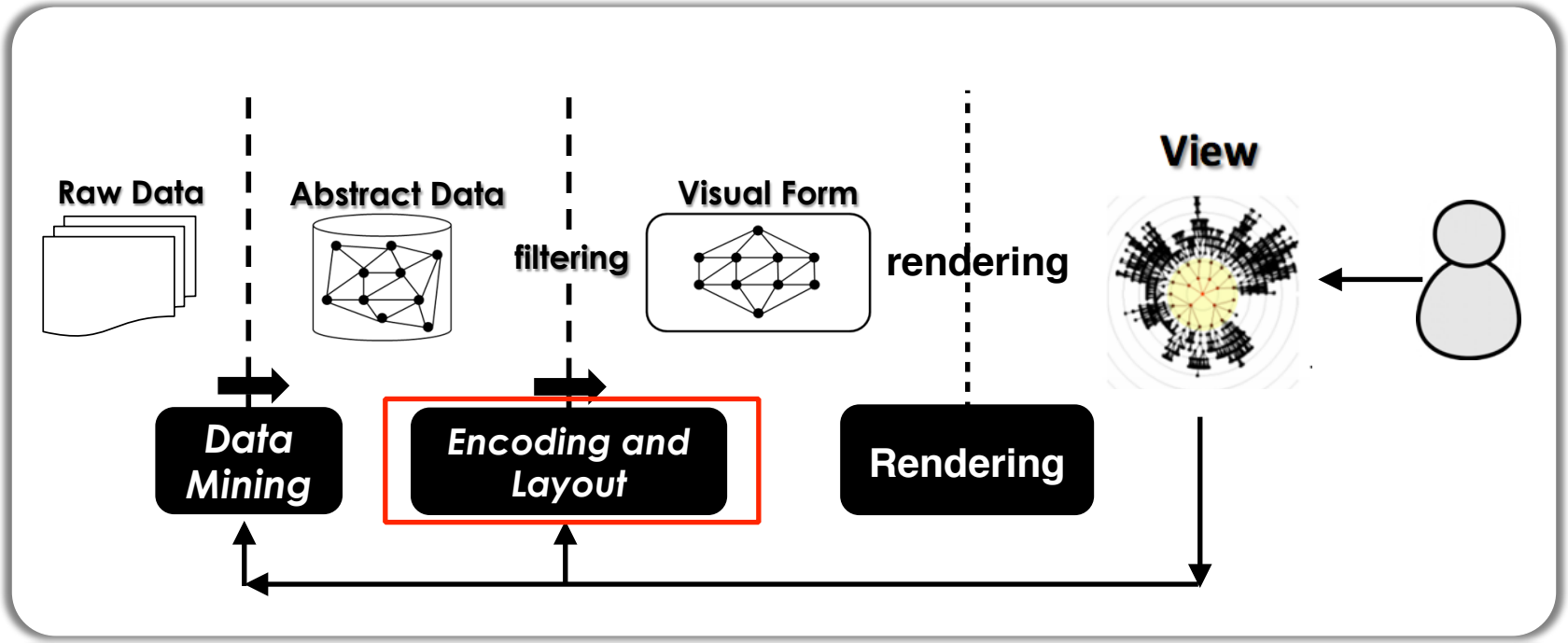
- Dynamic Network Visualization with "Network Movie"
  - A Key Guideline: preserving user's mental map during the visualization
    - Balance the static graph layout aesthetic and the graph stability between time frame
  - Use Staged Animation to Help User Perceive Complex Changes
    - Animation planning according to cognitive studies
    - Intermediate layout result had better satisfy graph aesthetics rather than interpolations
  - Some Negative Comments from User Studies
    - Animation leads to many participant errors, though participants find it enjoyable
    - Animation is the least effective form for analysis, even worse than static depictions

Dynamic Network Visualization



How can we visualize big data ?

Visualization & Visual Analysis Reference Model



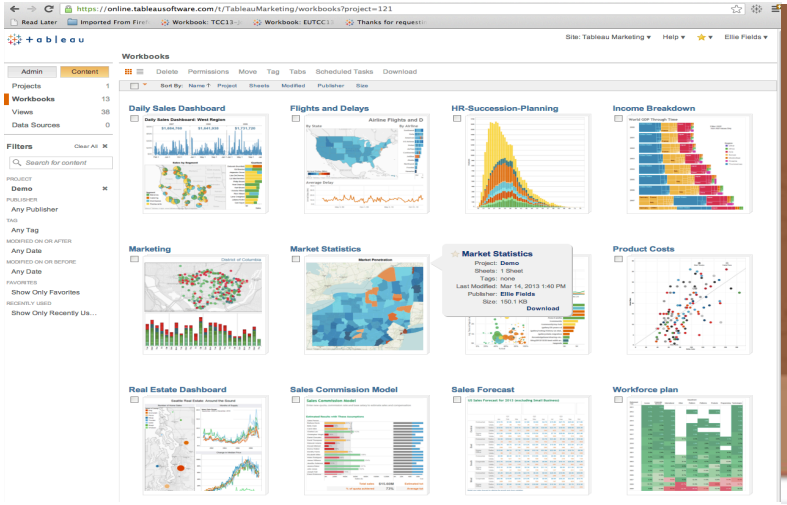
Encoding : Visual Design

Technique : Layout Algorithm

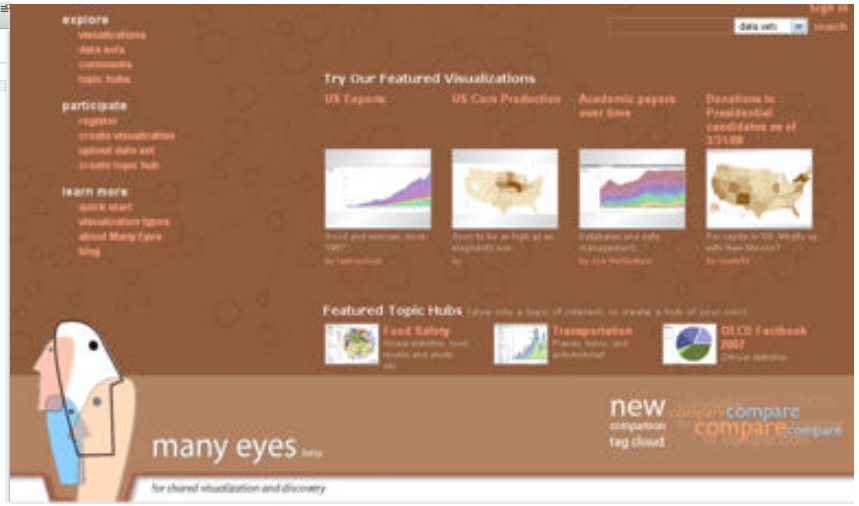
Using existing tools are easy

D3.js

Data-Driven Documents



Tableau



ManyEyes

Columbia University

Python:

- iGraph : http://igraph.org/redirect.html
- Networkx : https://networkx.github.io/

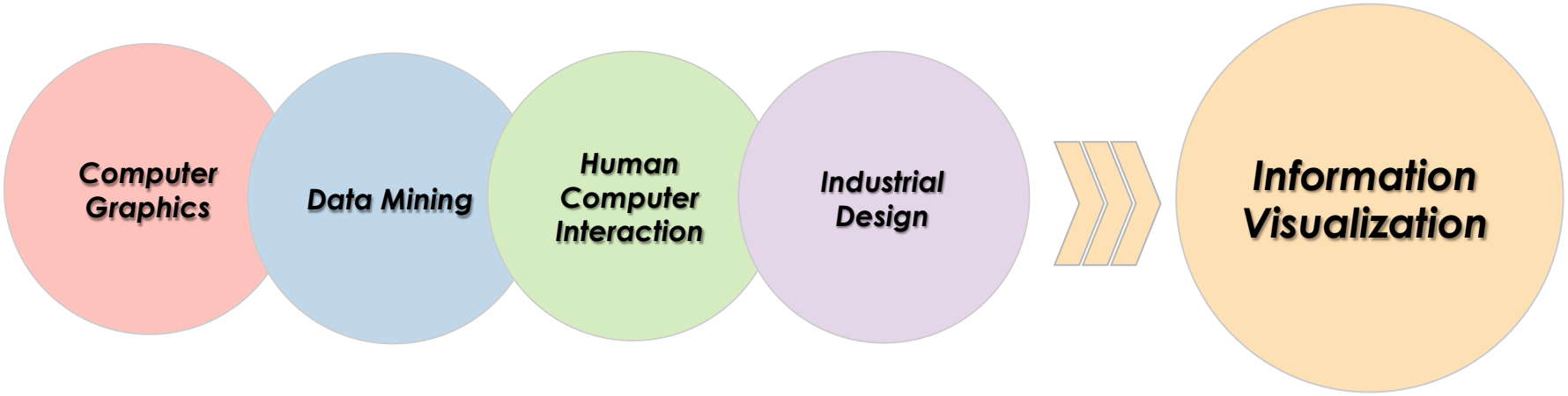
JavaScript:

- D3.js (2D, SVG): http://d3js.org/
- Tree.js (3D, WebGL): http://threejs.org/

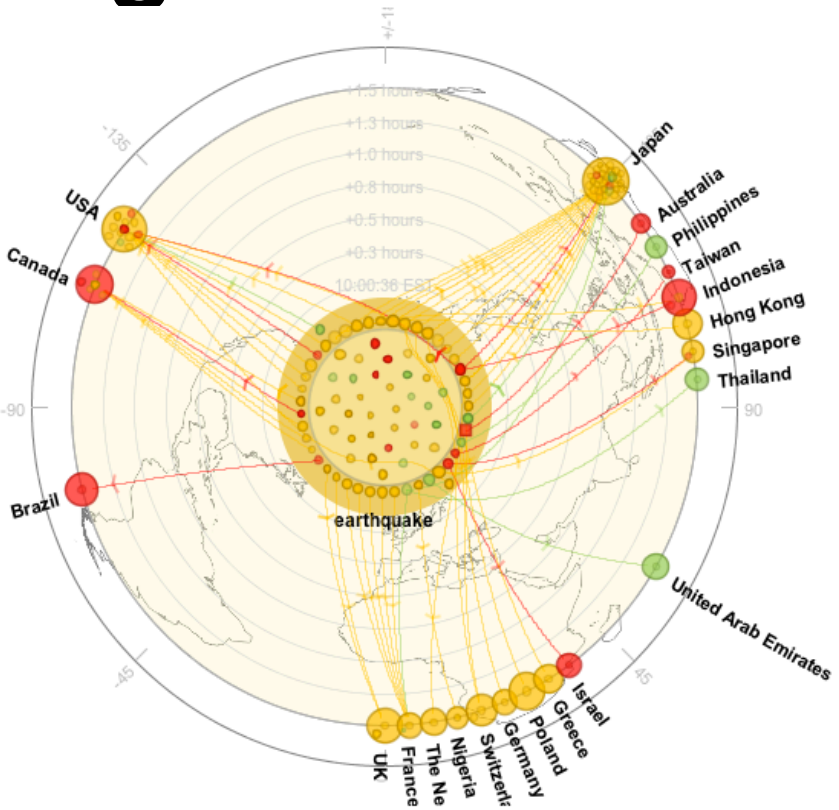
Java:

- prefuse: http://prefuse.org/
- InofVis Toolkit: http://ivtk.sourceforge.net/

Developing new ones require knowledge from different areas

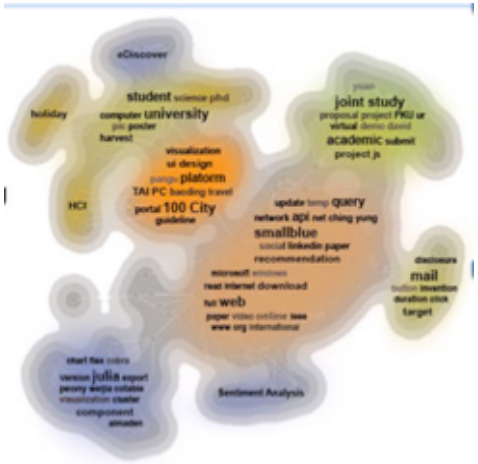


Example 1: Visualising Streaming Data



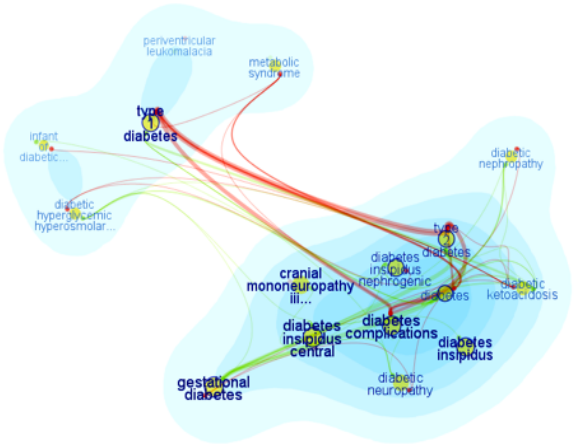
Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time
IEEE InfoVis 2012

Example 2: Visualizing Large Text Corpus



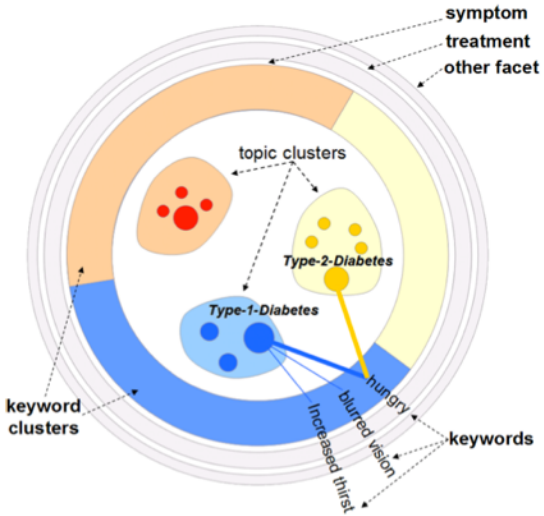
ContexTour
SDM 2010

Visualizing Heterogeneous Clusters



FacetAtlas
TVCG (InfoVis 2010)

Visualizing Multi-relational Clusters



SolarMap
ICDM 2011

Cluster Interpretation

Visual Analysis v.s. Data Mining

Computational Power

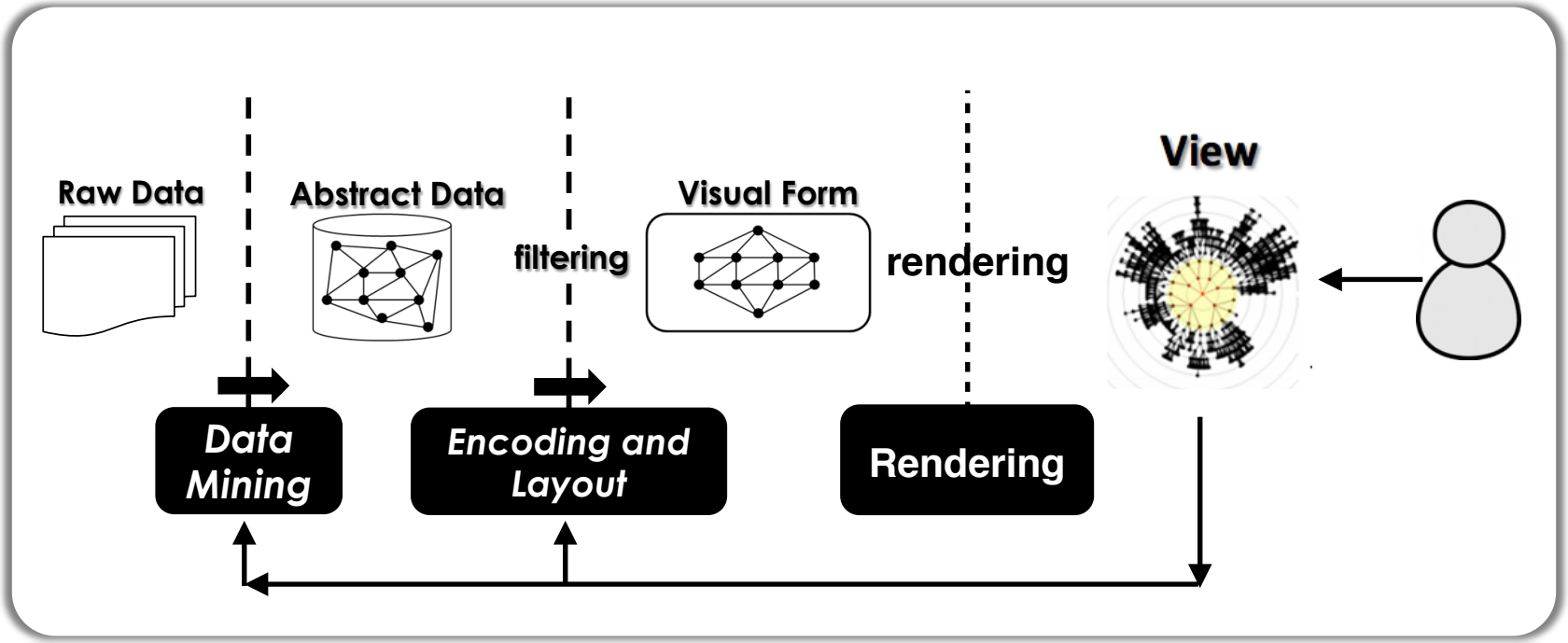
Data Mining

+

Human Intellegence

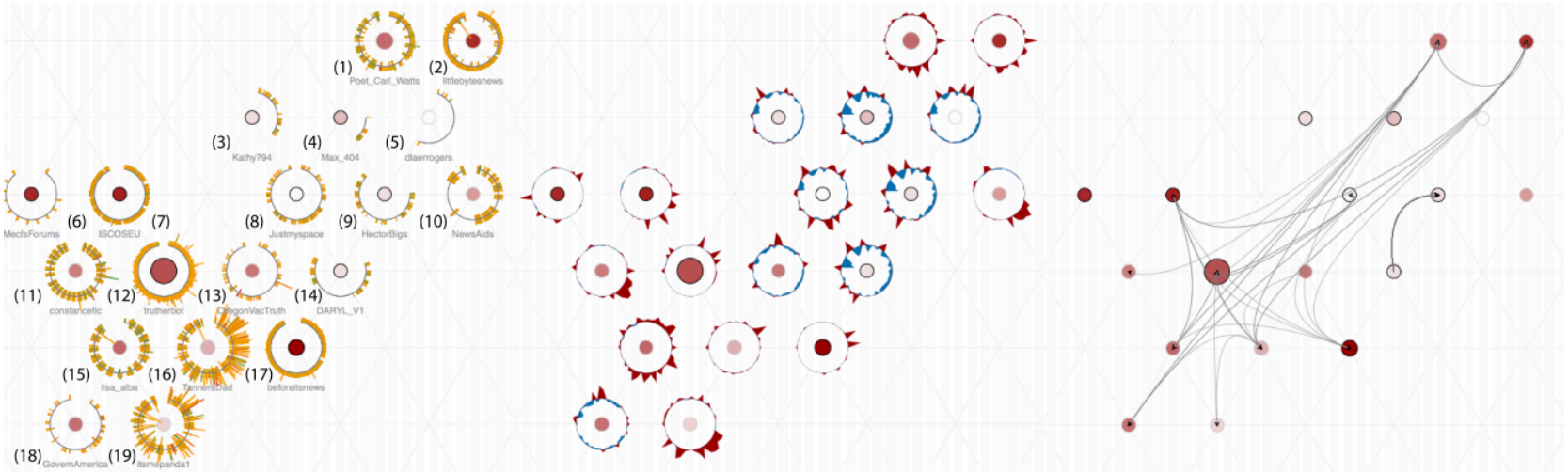
Visual Analysis

Visualization & Visual Analysis Reference Model



Analysis + Visualisation + Interaction

Example 3: Detect Anomalous Users in Twitter



TargetVue: Visual Analysis of Anomalous User Behaviors in Online Communication Systems, IEEE Transactions on Visualisation and Computer Graphics (VAST'15)

Reading Assignment:

The Top 10 Challenges in Extreme-Scale Visual Analytics

Pak Chung Wong
Pacific Northwest National Laboratory

Han-Wei Shen
Ohio State University

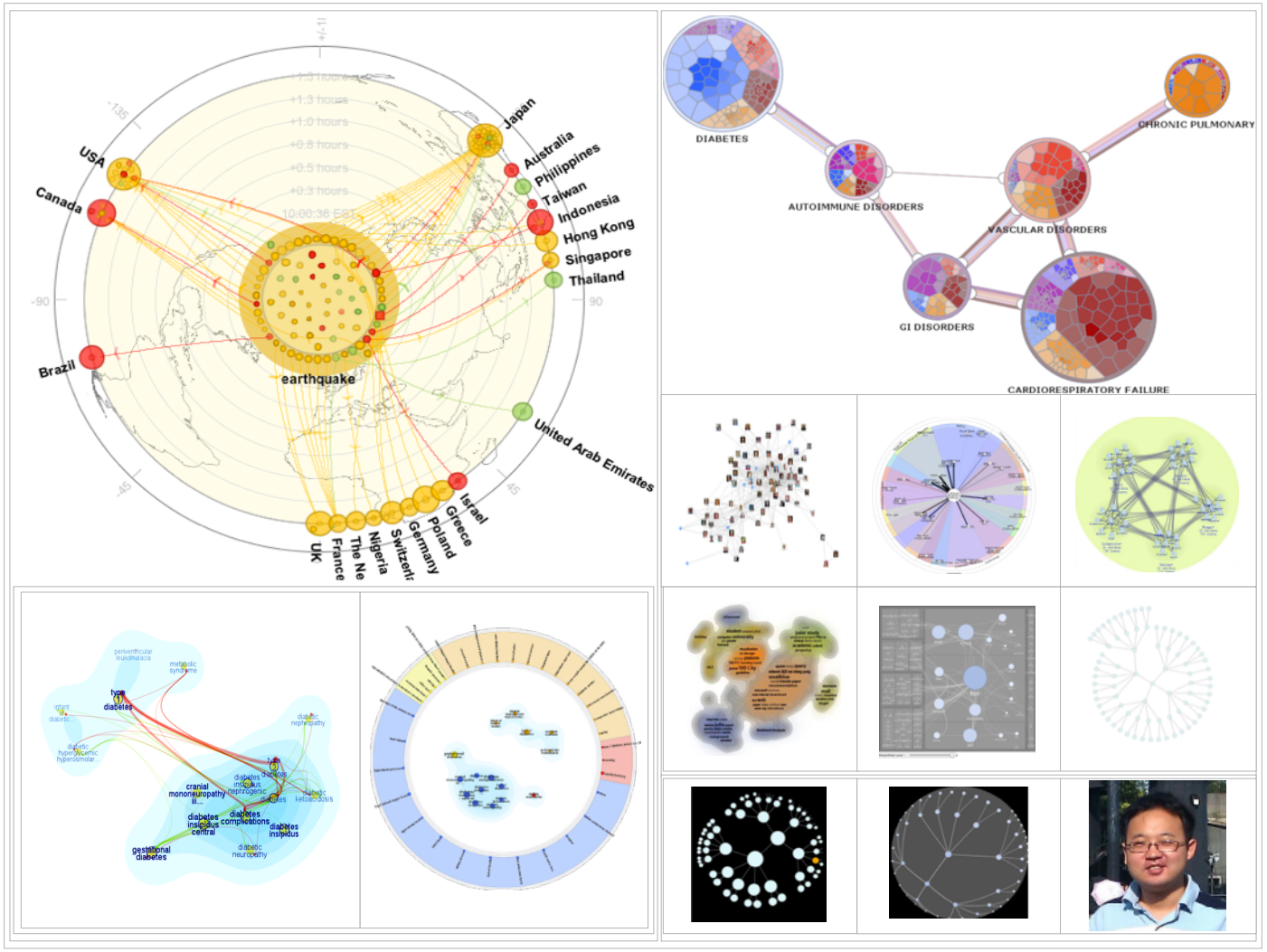
Christopher R. Johnson
University of Utah

Chaomei Chen
Drexel University

Robert B. Ross
Argonne National Laboratory

Wong, P. C., Shen, H. W., Johnson, C. R., Chen, C., & Ross, R. B. (2012). The top 10 challenges in extreme-scale visual analytics. IEEE computer graphics and applications, 32(4), 63.

Thank You!



Thank you very much for the many slides provided by Dr. Nan Cao and Dr. Lei Shi.