VSB TECHNICAL UNIVERSITY OF OSTRAVA | FACULTY OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE | DEPARTMENT OF COMPUTER SCIENCE

Data Visualization

460-4120

Fall 2024

Last update 20. 11. 2024

Information Visualization

- Upon now, we dealt with scientific visualization (scivis)
    - Scivis includes visualization of physical simulations, engineering, medical imaging, Earth sciences, etc.
    - Typical datasets consist of samples of continuous quantities over compact domain
- Now, we will focus on more abstract data types
    - Typical datasets: generic graphs and trees, database tables, text, etc.
    - Information visualization (infovis) studies the visual representation of such data

Information Visualization

- Infovis is the fastest groving branch of the visualization
- Main goal is to assist users in understanding all the abstract data, i.e. visualize abstract quantities and relations in order to get insight in the data with no physical representation
- Differences:
  - Scivis – physical data with inherent spatial placement → mental and physical images overlap → considerably simplifies visualization
  - Infovis – information has no innate shape and color and its visualization has purely abstract character

Information Visualization

- Three main elements: representation, presentation, and interaction
- Infovis has potentially larger target audience with limited mathematical or engineering background than scivis
- Infovis covers areas such as:
    - Visual reasoning, visual data modeling, visual programming, visual information retrieval and browsing, visualization of program execution, visual languages, visual interface design, and spatial reasoning

Information Visualization

- General rules for design of infovis applications:
  - Follow the conventions accepted by that field
  - Integrate with other tools-of-the-trade of the field
- In some taxonomies (Spence), there also exists class of geovisualization (geovis) applications which address a field between the two

Information Visualization

- Data domain:
    - Datasets often do not contain spatial information (sample points)
    - No cells with interpolation function or cell notion serves a different purpose
    - Actual spatial layout is of little if any relevance for the content

Information Visualization

- Attribute data types in infovis:
  - Data attributes are of more types than numerical values and go beyond the semantic of numerical values
  - A different storage strategy (size of a single attribute is variable)

Information Visualization

Data type			Attribute domain	Operations	Examples
Nominal (categorical)	Qualitative (no addition and multiplication)	Categorical*	Unordered set	Comparison ($=, \neq$)	Text, references, syntax elements
Ordinal			Ordered set	Ordering ($=, \neq, <, >$)	Ratings (e.g., bad, average, good)
Discrete	Quantitative (allow interpolation)		Integers (Z, N)	Integer arithmetic	Lines of code
Continuous		-	Reals (R)	Real arithmetic	Code metrics

Notes:

* A data item belongs to a category rather than the value of quantity

Information Visualization

- Another classification of attribute data types:
  - Linear
  - Planar
  - Volumetric
  - Temporal
  - Multidimensional
  - Tree
  - Network
  - Workspace

Spatial aspect

Relational aspect

Information Visualization

- Together with eight data types, seven interaction functions infovis application may provide:
    - Overview, zoom, filter, details on demand, relate, history, and extract
- These functions may be related to main steps of visualization pipeline:
    - Filtering, mapping, and rendering
- Data types and interaction types create a matrix of possibilities within which a infovis application may locate its functionality

Information Visualization

- Comparison of datasets notion in scivis and infovis

	Scivis	Infovis
Data domain	Spatial R^n	Abstract, nonspatial
Attribute types	Numeric R^m	Any data types
Data points	Samples of attributes over domain	Tuples of attributes without spatial location
Cells	Support interpolation	Describe relations
Interpolation	Piecewise continuous	Can be nonexistent

Information Visualization

Variable	Associative?	Selective?	Ordered?	Quantitative?
Planar	Yes	Yes	Yes	Yes
Size		Yes	Yes	Yes
Brightness		Yes	Yes	
Texture	Yes	Yes	Yes	
Color	Yes	Yes		
Orientation	Yes	Yes		
Shape	Yes	Sometimes		

Name	Definition	Example
Associative	If a mark is different in this attribute, it can be picked out	
Selective	If marks are similar in this attribute, they can be grouped into a family	Male/female
Ordered	The marks can be judged as ordered	More/fewer, higher/lower, first place/third place
Quantitative	The marks are perceived as numerically related to one another/proportional to one another	You can tell one is 2x as big as another WITHOUT a legend

Bertin, 1967

Information Visualization

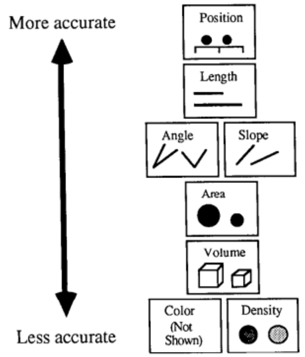
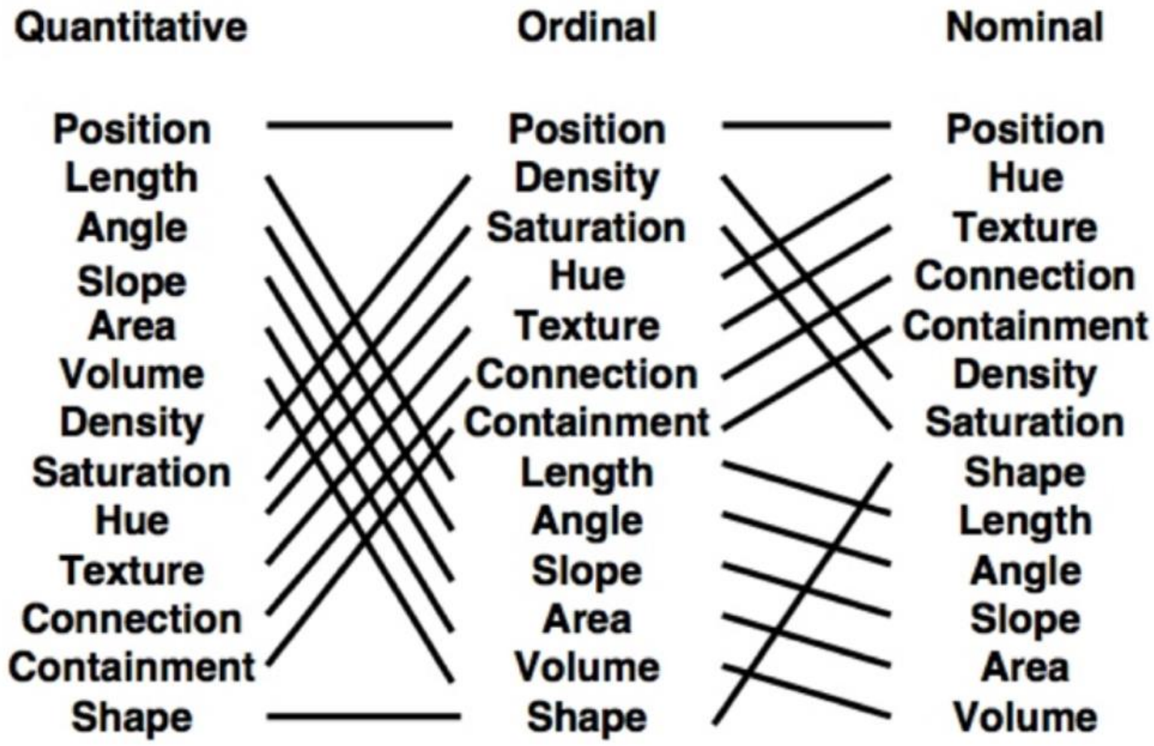


Fig. 14. Accuracy ranking of quantitative perceptual tasks. Higher tasks are accomplished more accurately than lower tasks. Cleveland and McGill empirically verified the basic properties of this ranking.

More accurate

Less accurate



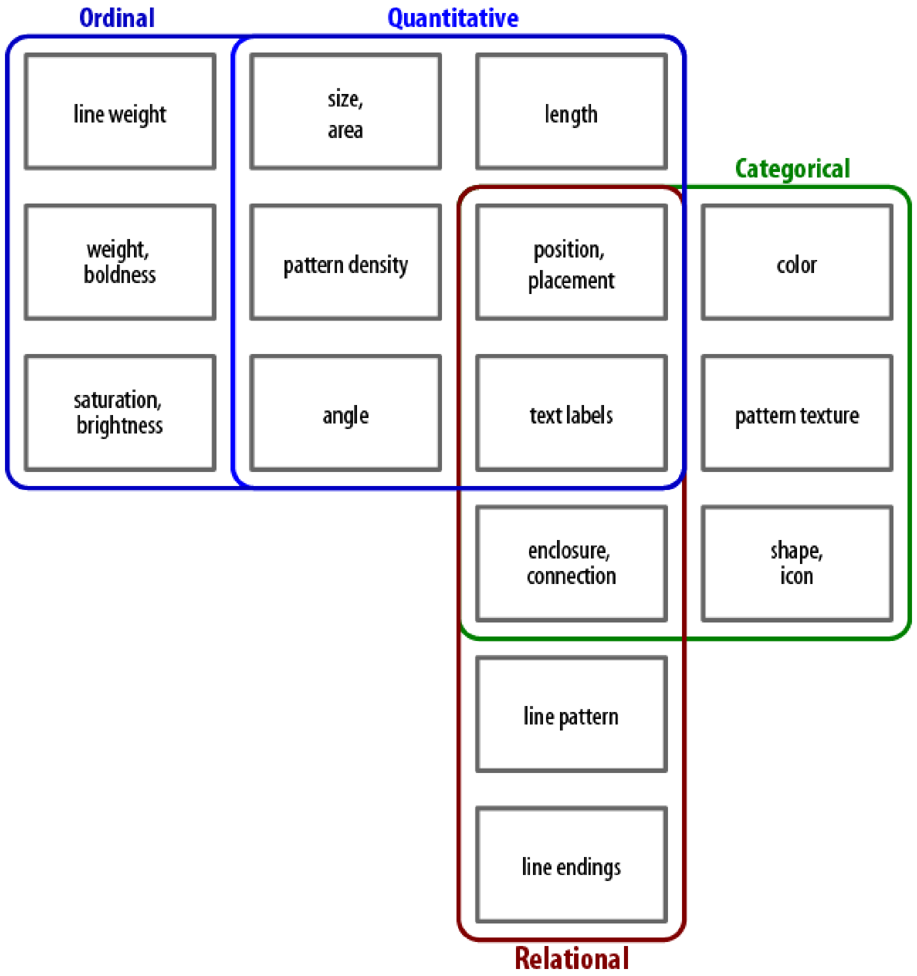
Mackinlay, 1986

Information Visualization

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (<20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

Use this table of common visual properties to help you select an appropriate encoding for your data type.
Designing Data Visualizations, 2011

Information Visualization



Visual properties grouped by the types of data they can be used to encode.
Designing Data Visualizations, 2011

Information Visualization

- Infovis datasets are quite similar to the model used in relational databases or entity-relationship graphs
- Visualization methods:
  - Database tables, trees, graphs, and text

Table Visualization

- Table – simplest infovis data; two-dimensional array of rows (records) and columns (attributes)
- Improvements supporting readability:
  - Sorting
  - Filling background of cells using alternate colors
  - Bar graph as a cell background
  - Small glyphs or icons showing trends
  - Sparklines

Tasks completed by team members

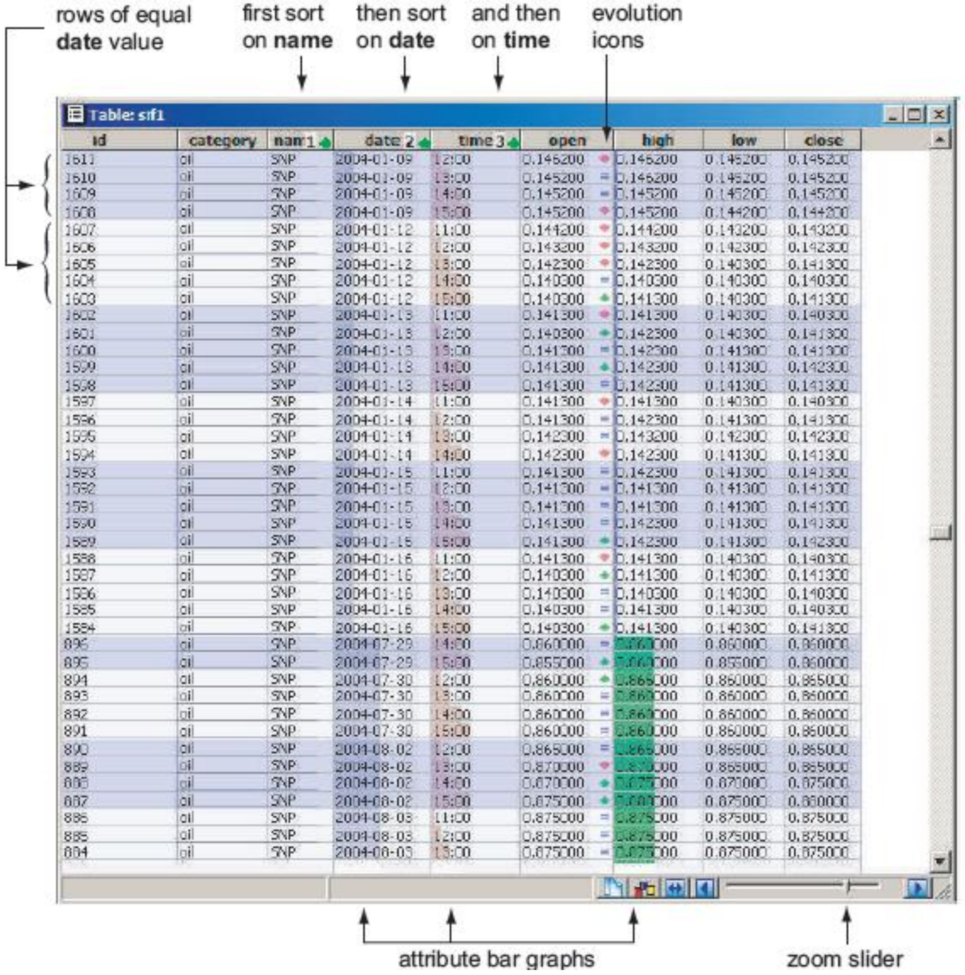
(last 26 weeks, YoY change shown in %s)

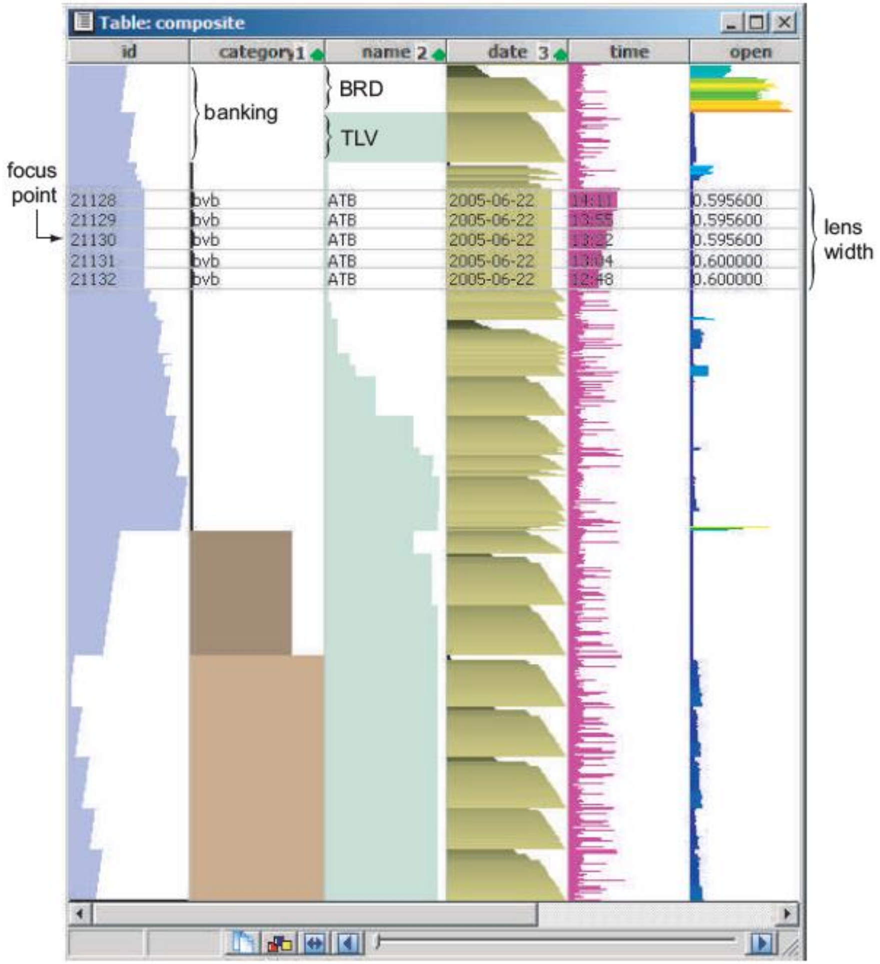
Team Member	Total Tasks Completed	w1	w2	w3	w25	w26
Julie	46%	13	15	19	11	19
John	45%	11	18	11	14	16
Jabba the hut	-20%	15	14	14	19	12
Johnson	6%	18	17	14	12	19
Jeremy	43%	14	20	10	12	20
Josh	-33%	15	12	19	11	10

Table Visualization

- Sampling issue
  - Text based visualization has fairly limited scalability
    - Zooming out the table visualization
    - We may drop displaying too small text and only show bar graphs
    - Use so called dense pixel displays or space filling displays

Table Visualization





Relation Visualization

- Frequently encountered visualizations of relational datasets:
    - Trees, graphs, and Venn-Euler diagrams

Tree Visualization

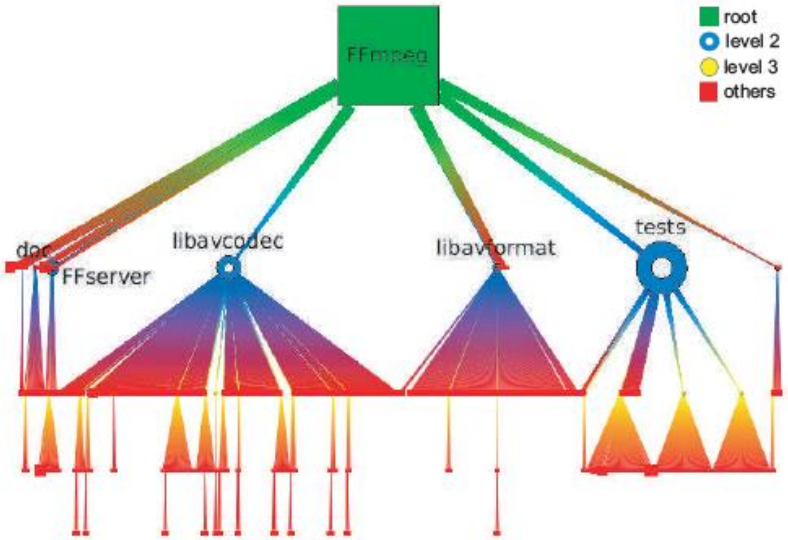
- Trees are a particular type of relational data
- $T = (N, E)$, where N = $\{n_i\}$ is set of nodes (vertices) connected by edges from set of edges $E = \{e_i\}$ where each edge e_i is represented as a pair (n_j(parent), n_k(child)) of nodes
- Properties of a tree:
    - There is a unique path between any two nodes in the tree
    - Subsequently, there are no loops
    - Parent may have any number of children; child can have only one parent; leaves have no children
    - Root – single node with no parents
    - Depth – longest path in the tree

Tree Visualization

- **Node-link visualization** (ball and stick) with two degrees of freedom:
  - Position of the glyphs (layout)
  - The appearance of the glyph
- Layout requirements:
  - No or minimal overlapping of nodes and edges
  - Aspect ratio not far from unity
  - Avoid long or unnecessarily bent edges

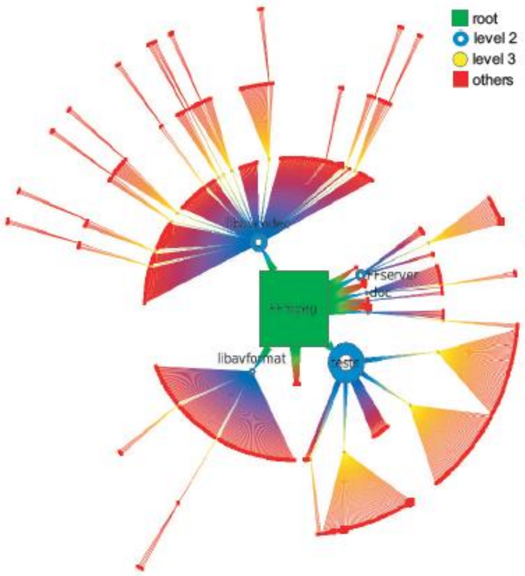
Tree Visualization

- **Rooted tree layout**:
  - All children nodes of the same parent have the same y-coordinate
  - X-axis is used to reflect certain ordering



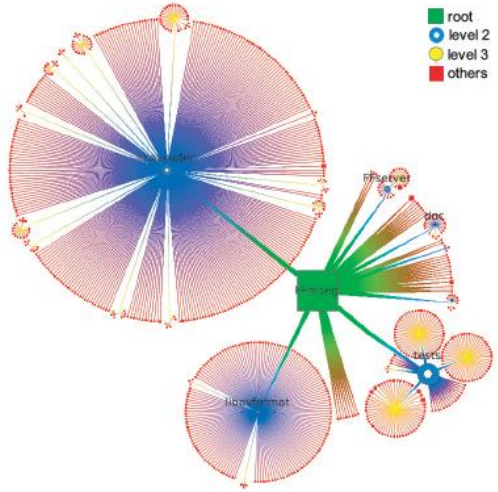
Tree Visualization

- **Radial tree layout**:
  - Use polar coordinate system
  - Always has 1:1 aspect ratio but problems with space allocation



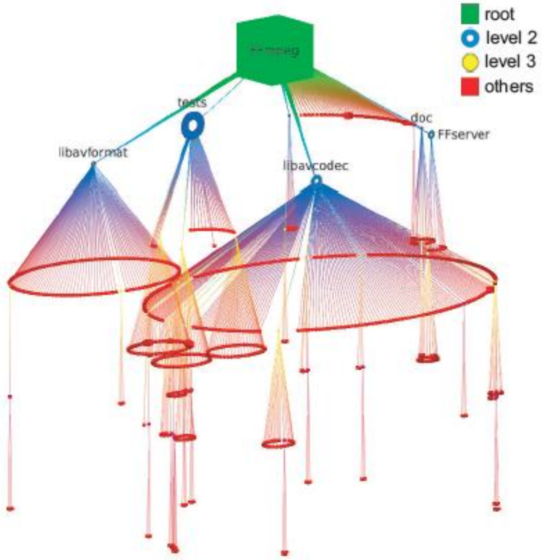
Tree Visualization

- **Bubble tree layout**:
  - Edges have now considerably different lenghts
  - This makes the visual size of the subrees reflect their number of children



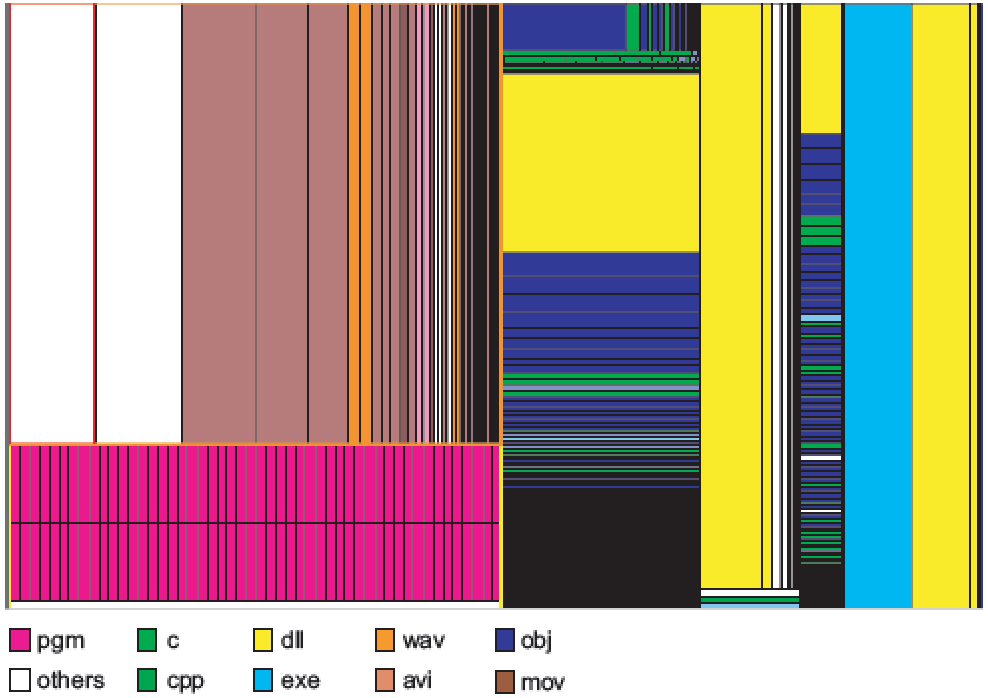
Tree Visualization

- **Cone tree layout**:
  - Arranged in 3D, may be more compact than other layouts
  - Problems: occlusions, chance of “getting lost“ in 3D space



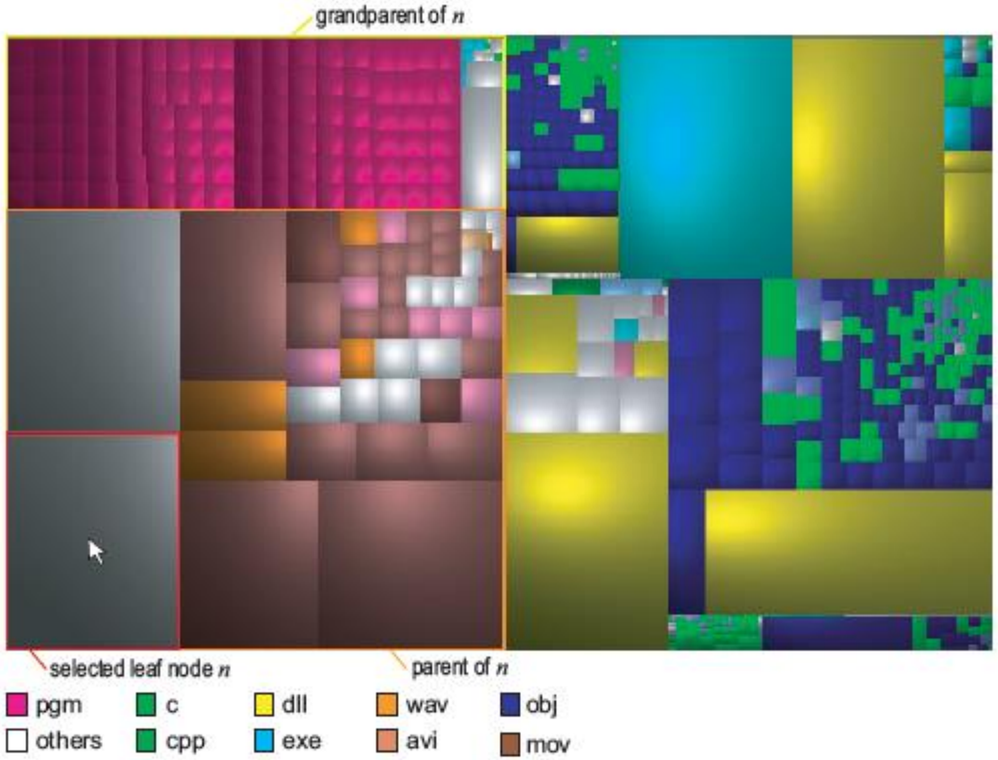
Tree Visualization

- **Tree Maps**
  - Slice and dice layout



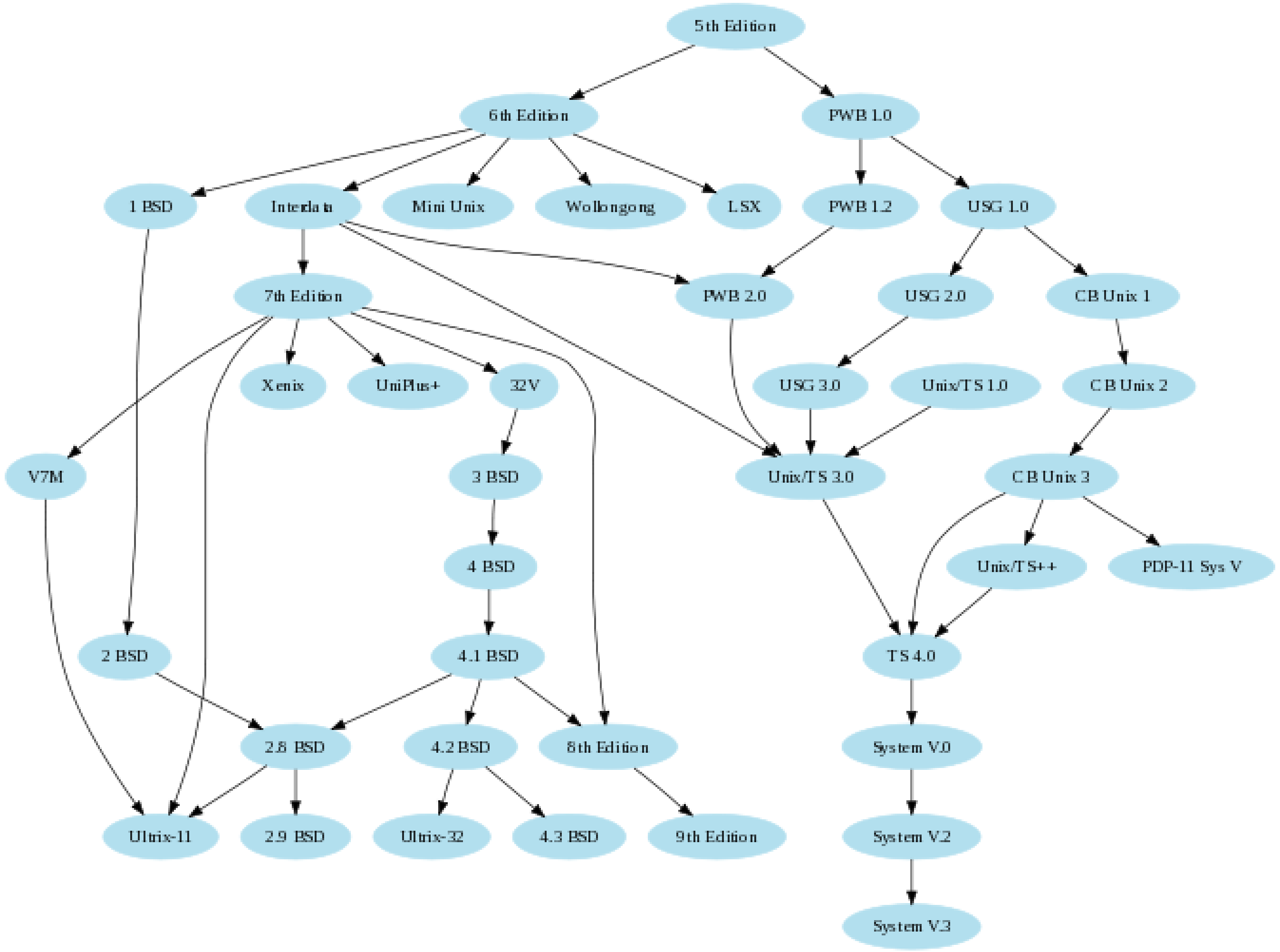
Tree Visualization

- **Tree Maps**
  - Squarified layout



Tree Visualization

- **Tree Maps**
  - Hierarchical layout



Graphs

- **Force-directed layout**

Fruchterman and Reingold (1991)

$$\mathbf{F}_a(n_i, n_j) = \frac{\|p_i - p_j\|}{k}(p_j - p_i),$$
$$\mathbf{F}_r(n_i, n_j) = -\frac{k^2}{\|p_i - p_j\|^2}(p_j - p_i)$$

Eades (1984)

$$\mathbf{F}_a(n_i, n_j) = k \log \|p_i - p_j\| \frac{p_j - p_i}{\|p_j - p_i\|}$$
$$\mathbf{F}_r(n_i, n_j) = -\frac{k}{\|p_i - p_j\|^3}(p_j - p_i).$$

$k = \sqrt{A}/N$ where A is the plot area

The energy function is not monotonic
Can get stuck in local minima
No clear ordering – where to start reading the plot

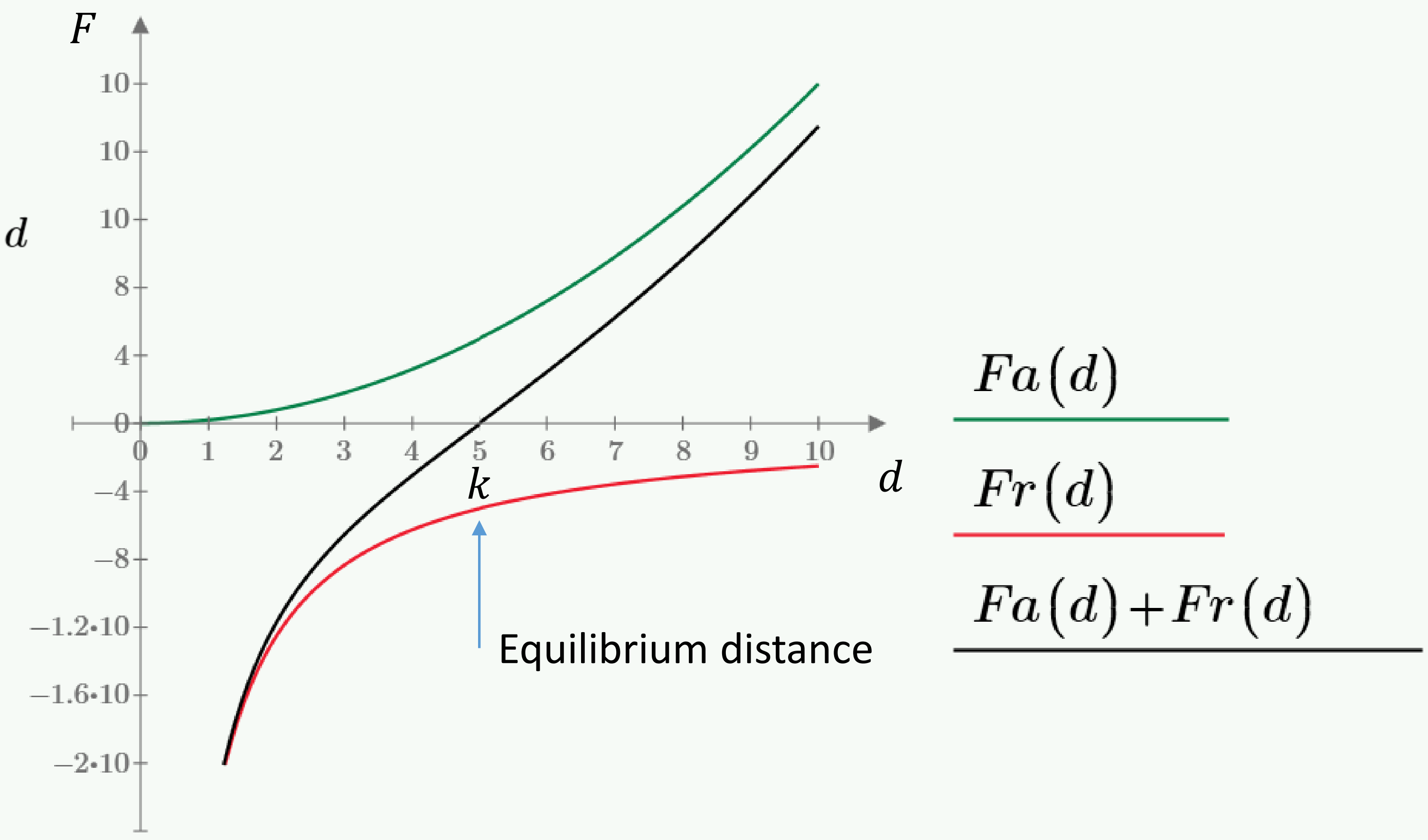
```
// initialize random layout first

t = 1.0 // set the initial maximal move

do {
  for ( i = 0; i < nodes.length; i++ ) { // compute repulsive force
    F[i] = 0;
    for ( j = 0; j < nodes.length; j++ ) {
      if ( i != j )
        F[i].add( Fr( nodes[i].position, nodes[j].position ) );
    }
  }

  for ( k = 0; k < edges.length; k++ ) { // compute attractive force
    i = edges[k].first; // get the first node of the current edge
    j = edges[k].second; // get the second node of the current edge

    fa = Fa( nodes[i].position, nodes[j].position );
    F[i].add( fa );
    F[j].sub( fa );
  }

  for ( i = 0; i < nodes.length; i++ ) { // move nodes by applying forces
    df = ||F[i]||;
    if ( df > 0 ) {
      ds = F[i] / df * min( δ, t * df ) ); // δ = 0.1
      nodes[i].position.add(ds);
    }
  }

  t -= t * 0.01; // reduce maximal allowed move
} while nodes move;
```

Graphs

$$k := 5$$

Attractive force $$Fa(d) := \frac{\|d\|}{k} \cdot d$$

Repulsive force $$Fr(d) := -\frac{k^2}{\|d\|^2}\, d$$



Graphs

- **Energy based model** (Kamada and Kawai)

$$E = \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(\frac{\|\boldsymbol{p}_i - \boldsymbol{p}_j\| - d_{ij}}{d_{ij}} \right)^2$$

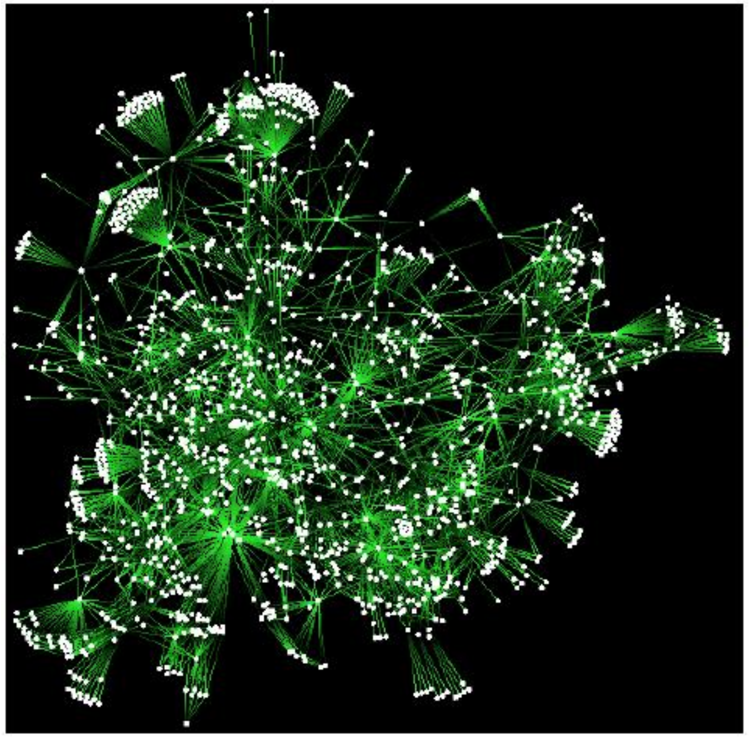
where d_{ij} is the length (measured as number of edges) of the shortest path in the graph connecting nodes i, j and all d_{ij} can be computed by Floyd-Warshall algorithm

- Minimal energy E corresponds to the state when the distances between connected nodes are proportional to the distances in the graph
- From classical mechanics, force is equal to the gradient of the potential energy

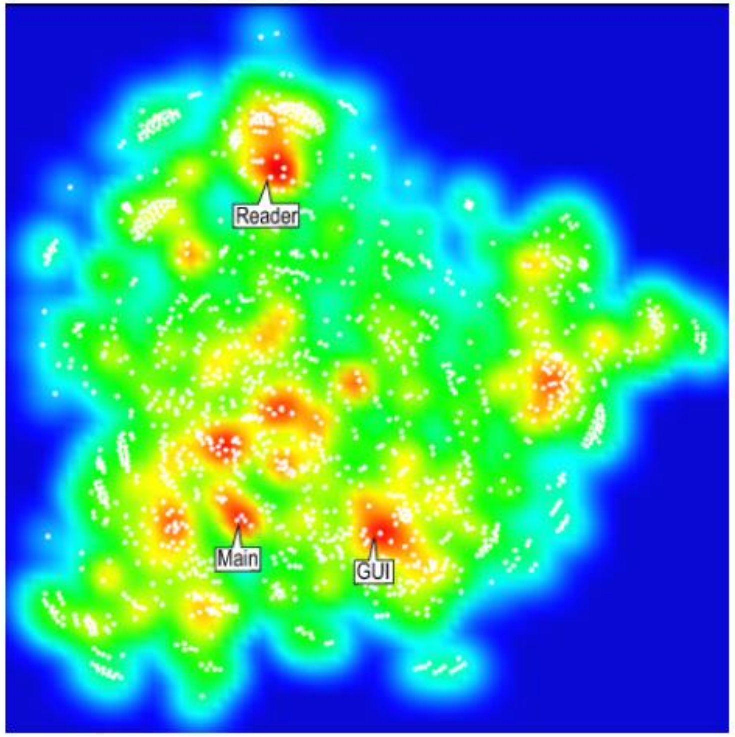
$$\boldsymbol{F} = -\nabla E$$

Tree Visualization

- **Graph Splatting**
  - Convolve nodes (optionally edges) with Gaussian filter



Discrete dataset



Continuous dataset

Matrix Visualization

- **(Directed/undirected) Adjacency Matrix**
  - Order of rows and columns highly impact the visualization (spotting clusters etc.)