# Ch 8/9: Spatial Data, Networks
# Paper: Genealogical Graphs
# Paper: ABySS-Explorer

**Tamara Munzner**

Department of Computer Science

University of British Columbia

*CPSC 547, Information Visualization*
*Week 6: 17 October 2017*

www.cs.ubc.ca/~tmm/courses/547-17F

# News

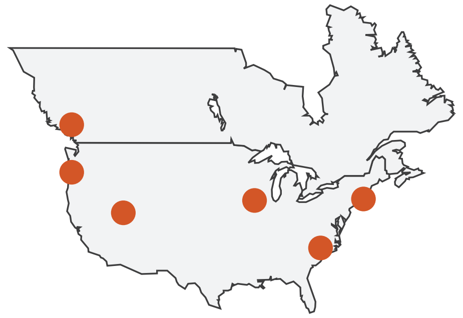
- marks for previous 2 weeks published
  - first week was pass/fail for having anything
  - now more fine-grained guidance about expectations with comments
    - if you didn’t get full credit
  - in general: don’t just summarize
- today
  - pitches first
  - Q&A, lecture second

# Ch 8: Arrange Spatial Data

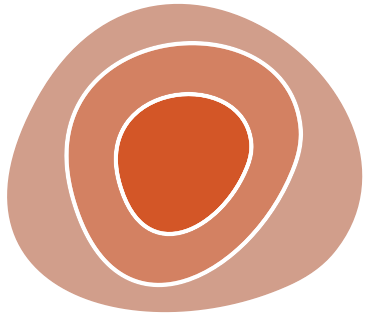
# Arrange spatial data

## ➔ Use Given

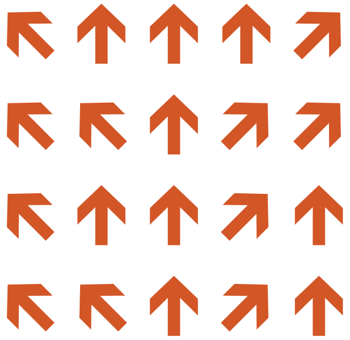
- ➔ Geometry
  - ➔ *Geographic*
  - ➔ *Other Derived*



- ➔ Spatial Fields
  - ➔ *Scalar Fields (one value per cell)*
    - ➔ *Isocontours*
    - ➔ *Direct Volume Rendering*

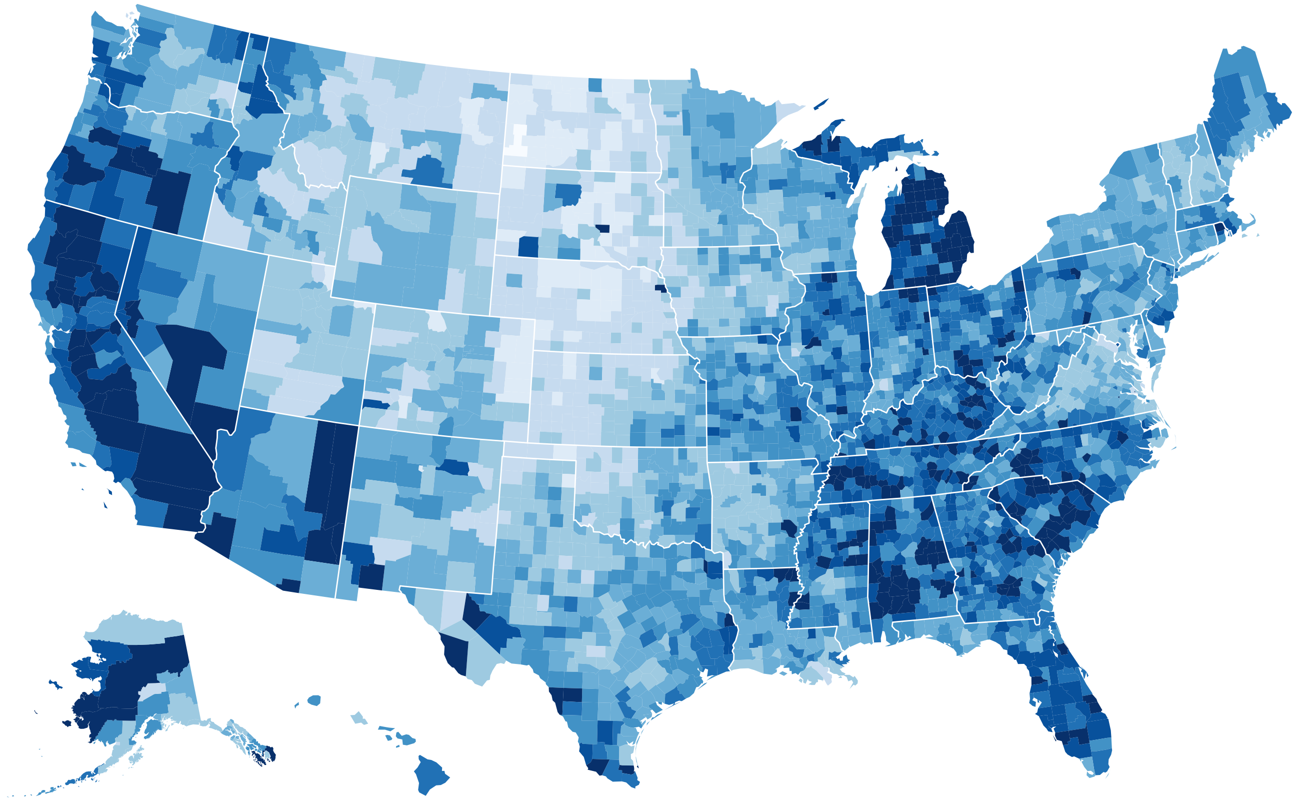


  - ➔ *Vector and Tensor Fields (many values per cell)*
    - ➔ *Flow Glyphs (local)*
    - ➔ *Geometric (sparse seeds)*
    - ➔ *Textures (dense seeds)*
    - ➔ *Features (globally derived)*



# Idiom: **choropleth map**

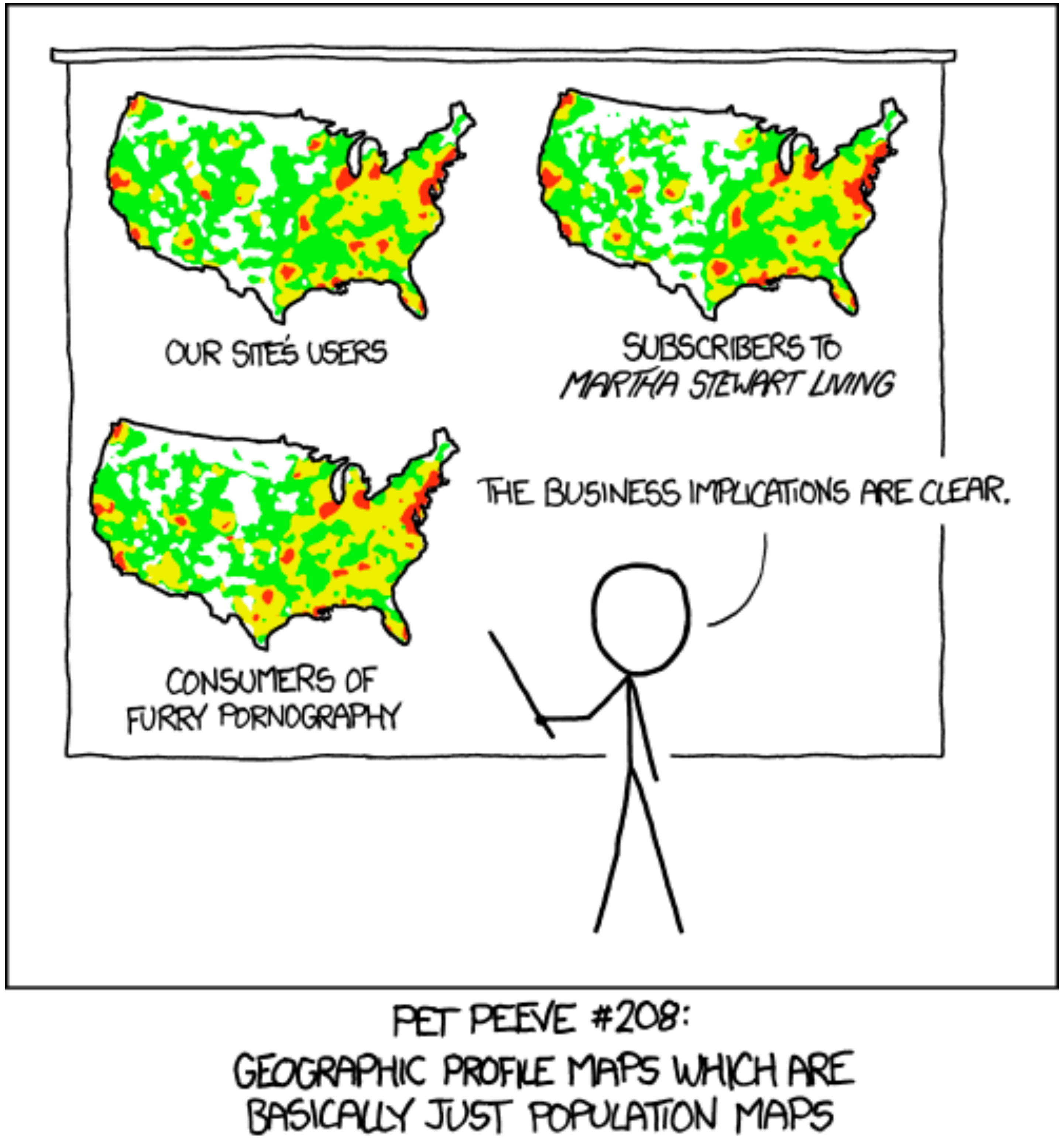
- *use* given spatial data
  - when central task is understanding spatial relationships
- data
  - geographic geometry
  - table with 1 quant attribute per region
- encoding
  - use given geometry for area mark boundaries
  - sequential segmented colormap *[more later]*
  - (geographic heat map)



http://bl.ocks.org/mbostock/4060606

# Population maps trickiness

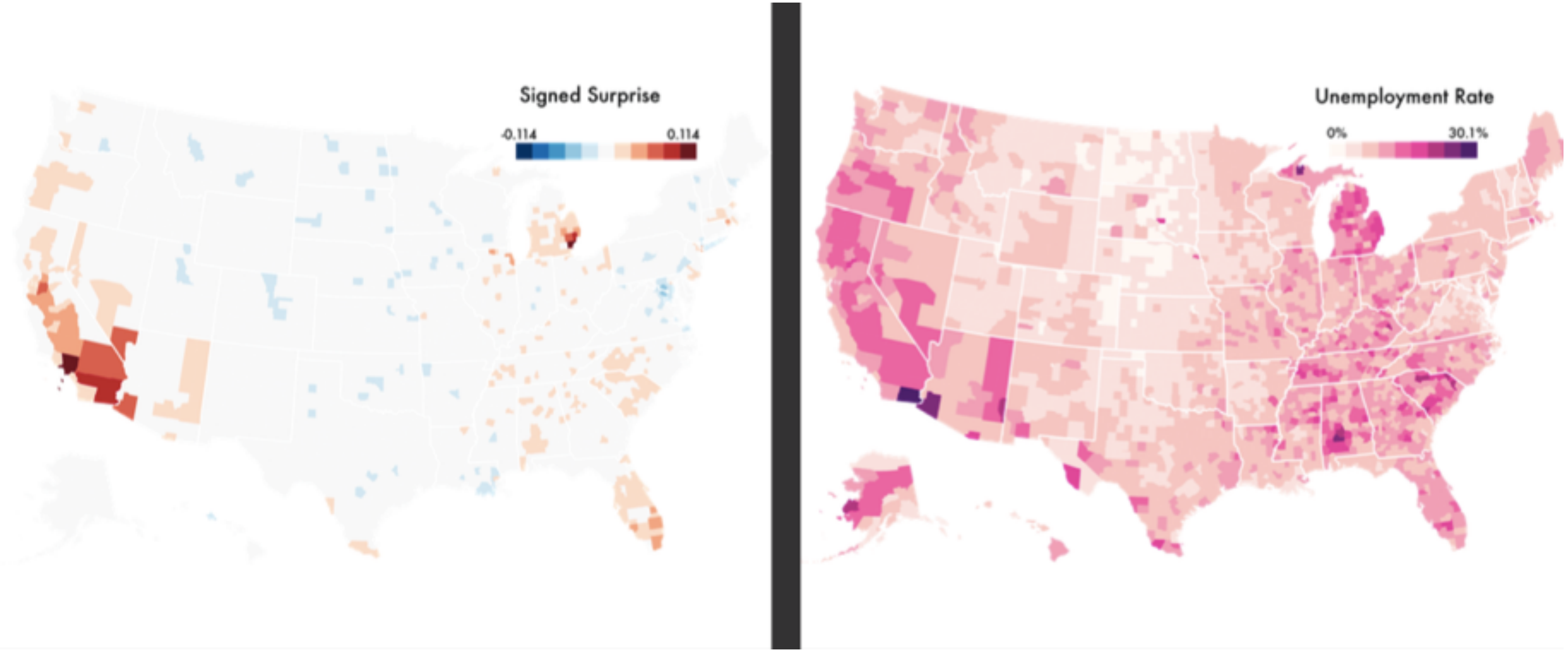
- beware!
- absolute vs relative again
  - population density vs per capita
- investigate with Ben Jones Tableau Public demo
  - http://public.tableau.com/profile/ben.jones#!/vizhome/PopVsFin/PopVsFin *Are Maps of Financial Variables just Population Maps?*
    - yes, unless you look at per capita (relative) numbers


OUR SITE'S USERS
SUBSCRIBERS TO MARTHA STEWART LIVING
CONSUMERS OF FURRY PORNOGRAPHY
THE BUSINESS IMPLICATIONS ARE CLEAR.
PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE BASICALLY JUST POPULATION MAPS

*[ https://xkcd.com/1138 ]*

# Idiom: Bayesian surprise maps

- use models of expectations to highlight surprising values
- confounds (population) and variance (sparsity)

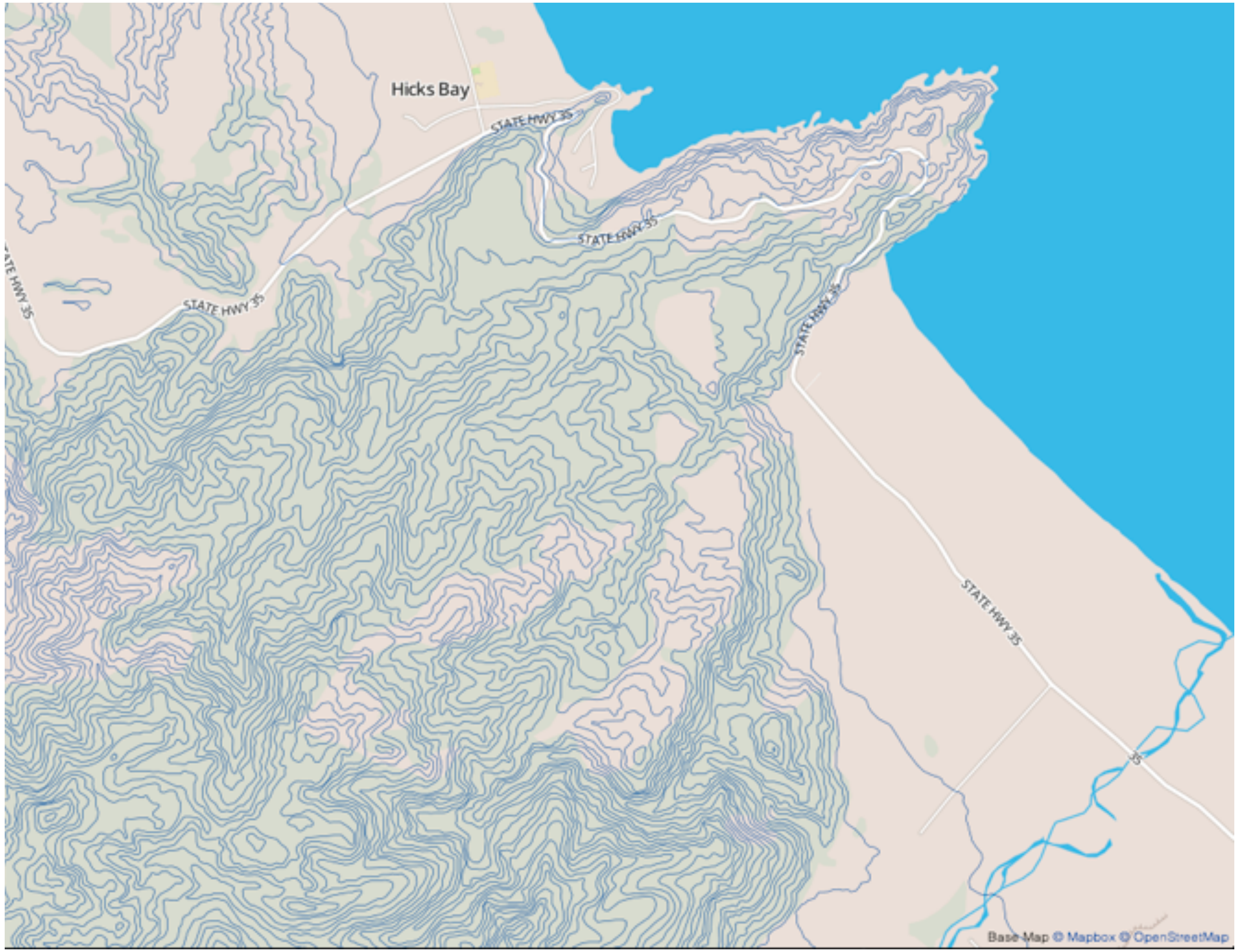


*[Surprise! Bayesian Weighting for De-Biasing Thematic Maps. Correll and Heer. Proc InfoVis 2016]*

https://medium.com/@uwdata/surprise-maps-showing-the-unexpected-e92b67398865 https://idl.cs.washington.edu/papers/surprise-maps/

# Idiom: **topographic map**

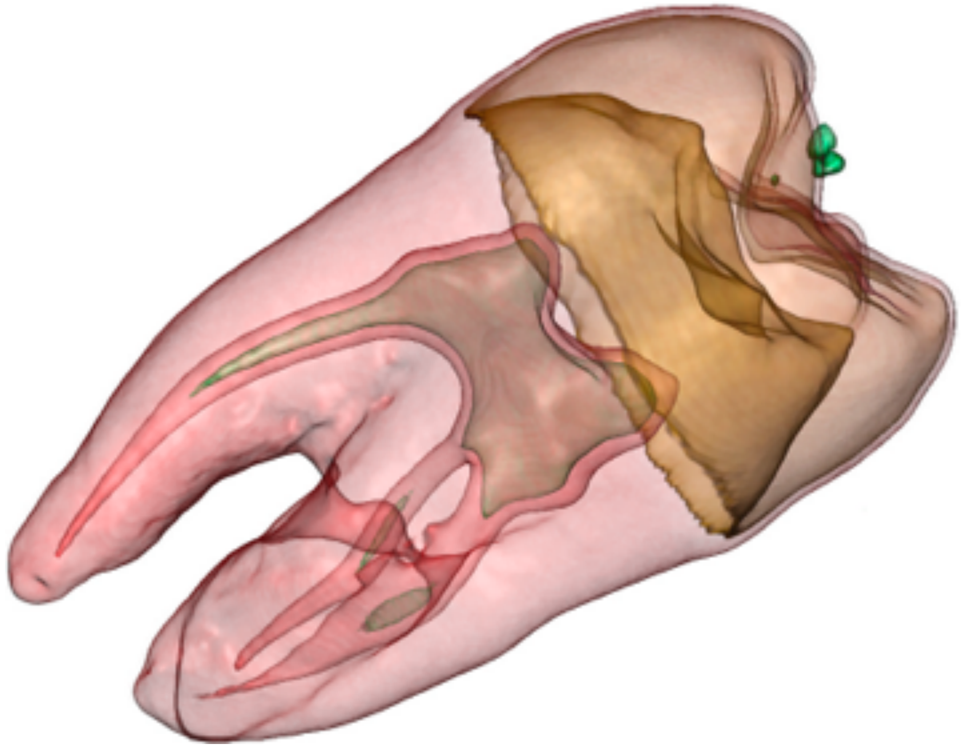
- data
  - –geographic geometry
  - –scalar spatial field
    - 1 quant attribute per grid cell
- derived data
  - –isoline geometry
    - isocontours computed for specific levels of scalar values



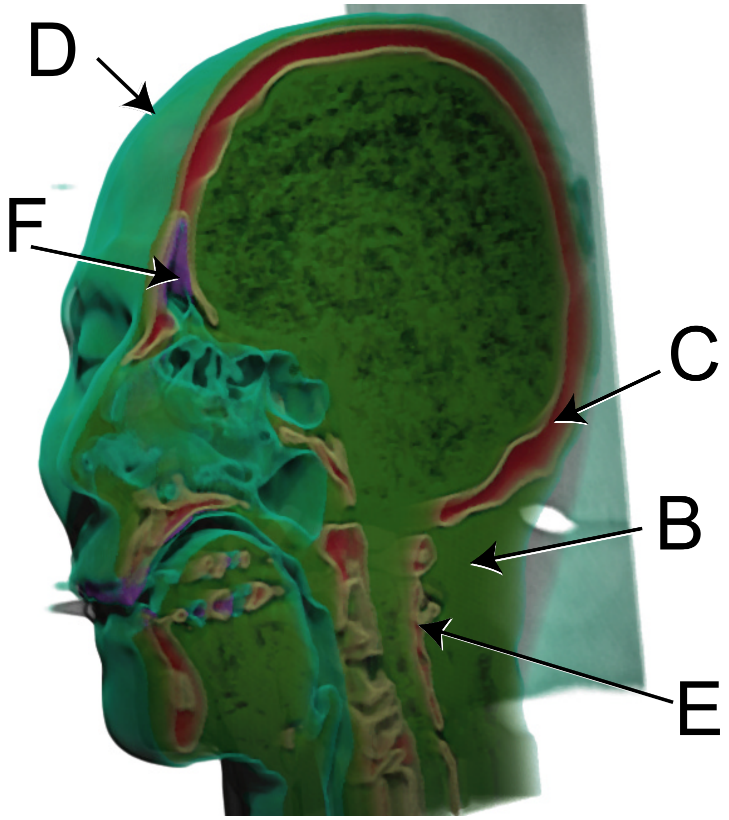
*Land Information New Zealand Data Service*

# Idioms: **isosurfaces, direct volume rendering**

- data
  - –scalar spatial field
    - 1 quant attribute per grid cell
- task
  - –shape understanding, spatial relationships
- isosurface
  - –derived data: isocontours computed for specific levels of scalar values
- direct volume rendering
  - –transfer function maps scalar values to color, opacity



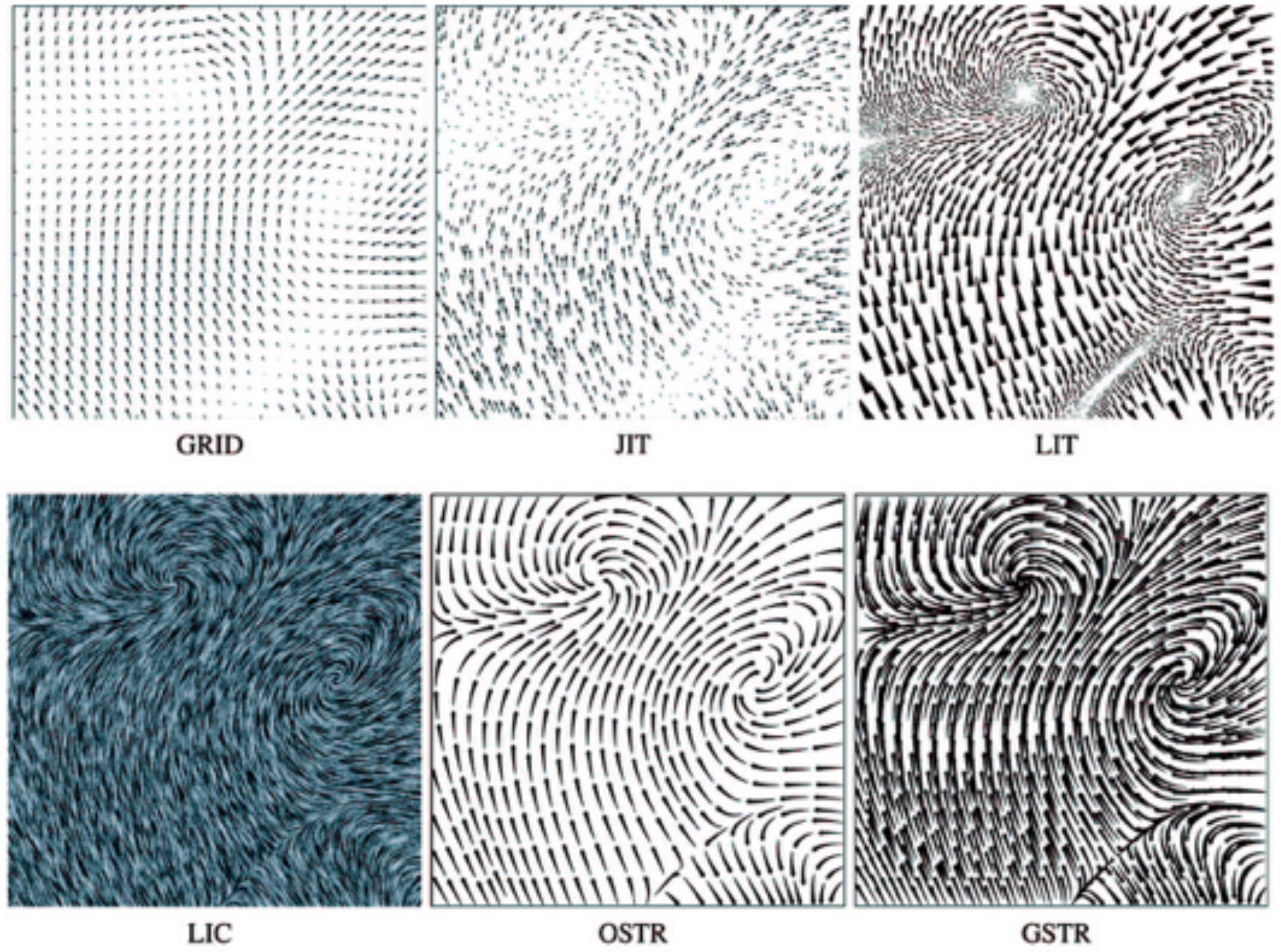
*[Interactive Volume Rendering Techniques. Kniss. Master's thesis, University of Utah Computer Science, 2002.]*



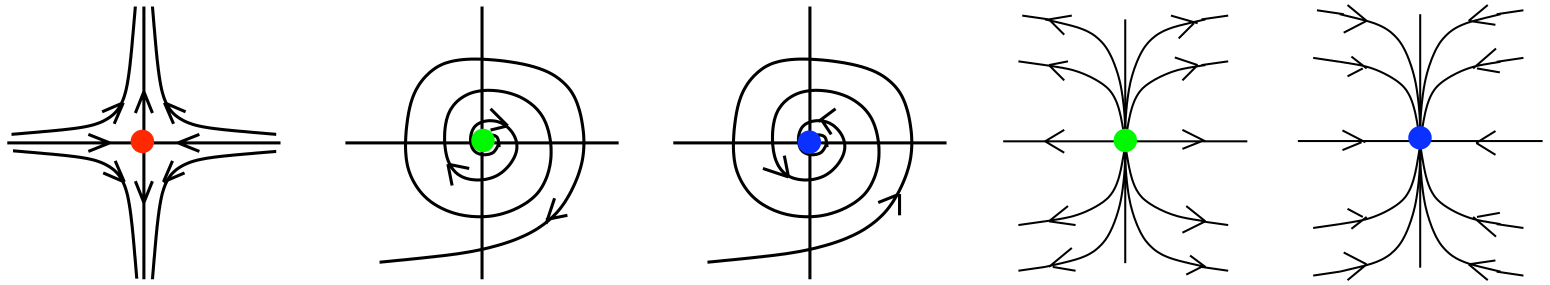
*[Multidimensional Transfer Functions for Volume Rendering. Kniss, Kindlmann, and Hansen. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 189–210. Elsevier, 2005.]*

# Vector and tensor fields

- data
  - many attribs per cell
- idiom families
  - flow glyphs
    - purely local
  - geometric flow
    - derived data from tracing particle trajectories
    - sparse set of seed points
  - texture flow
    - derived data, dense seeds
  - feature flow
    - global computation to detect features
      - encoded with one of methods above



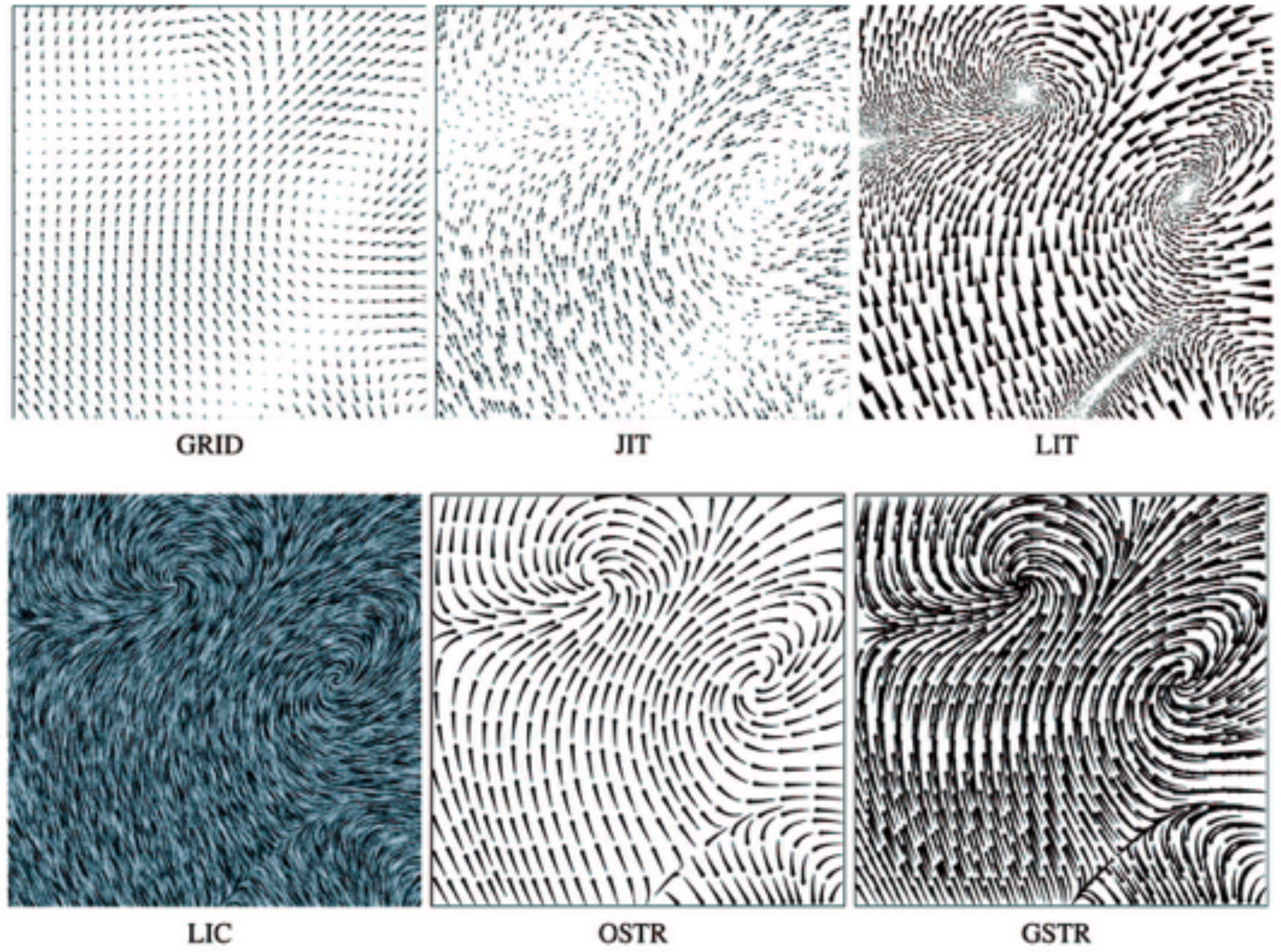
*[Comparing 2D vector field visualization methods: A user study. Laidlaw et al. IEEE Trans. Visualization and Computer Graphics (TVCG) 11:1 (2005), 59–70.]*



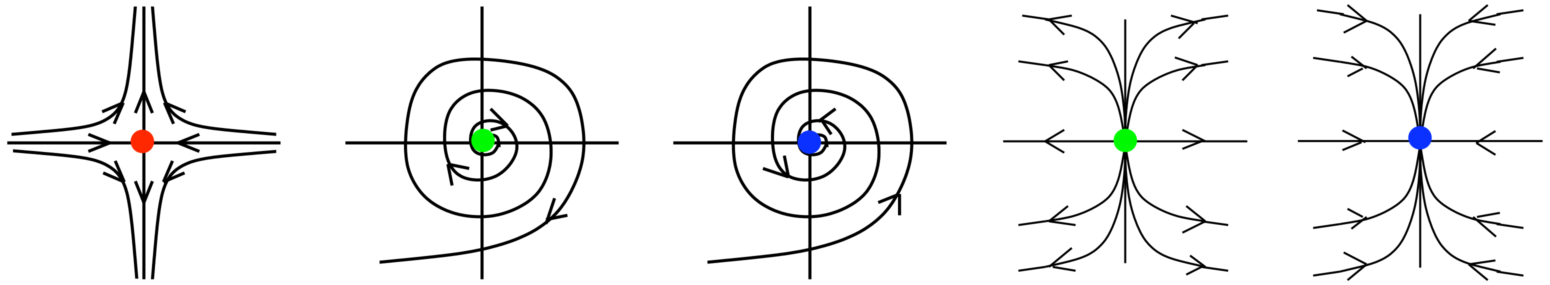
*[Topology tracking for the visualization of time-dependent two-dimensional flows. Tricoche, Wischgoll, Scheuermann, and Hagen. Computers & Graphics 26:2 (2002), 249–257.]*

# Vector fields

- empirical study tasks
  - finding critical points, identifying their types
  - identifying what type of critical point is at a specific location
  - predicting where a particle starting at a specified point will end up (advection)



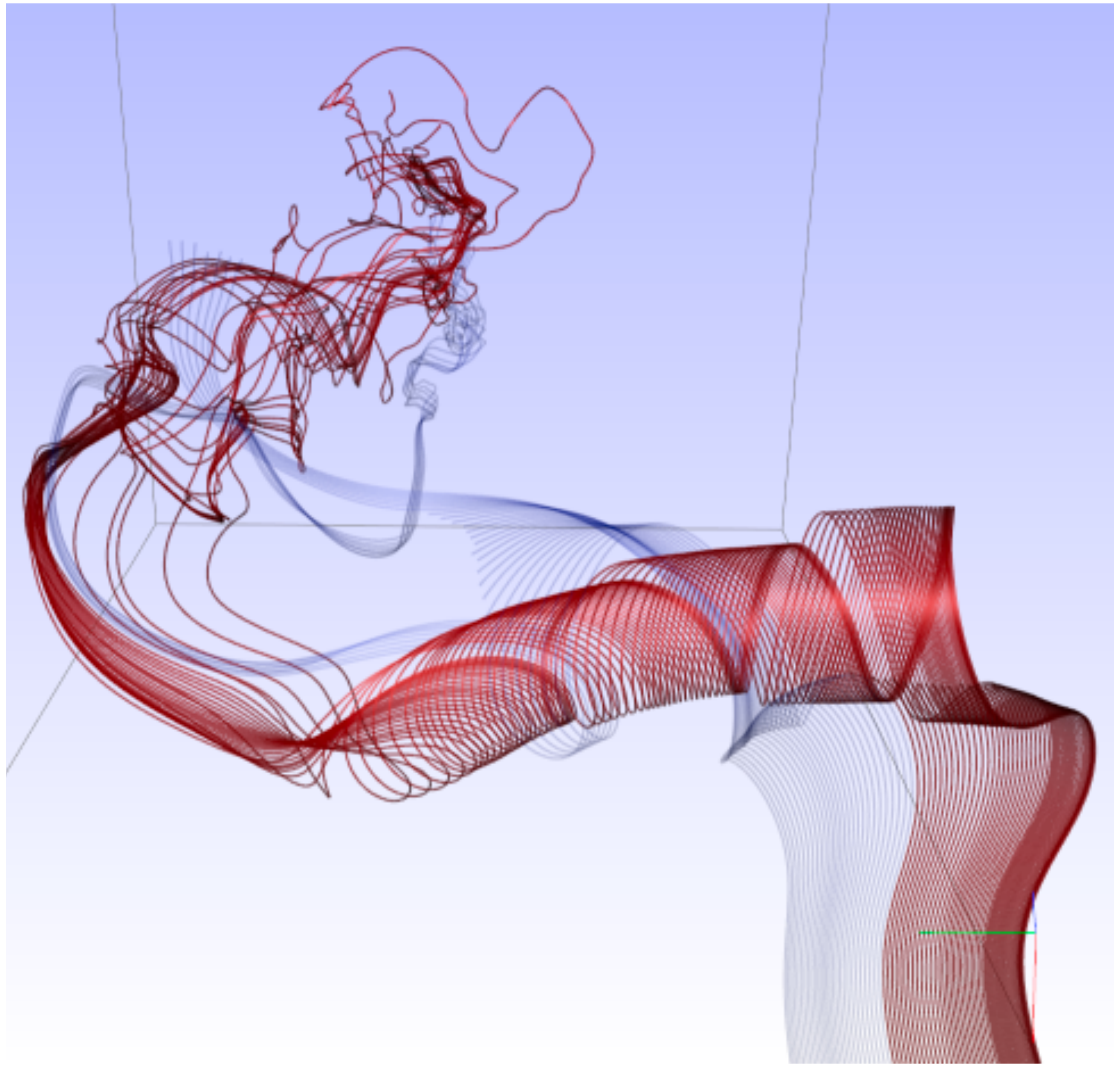
*[Comparing 2D vector field visualization methods: A user study. Laidlaw et al. IEEE Trans. Visualization and Computer Graphics (TVCG) 11:1 (2005), 59–70.]*



*[Topology tracking for the visualization of time-dependent two-dimensional flows. Tricoche, Wischgoll, Scheuermann, and Hagen. Computers & Graphics 26:2 (2002), 249–257.]*

# Idiom: **similarity-clustered streamlines**

- data
  - 3D vector field
- derived data (from field)
  - streamlines: trajectory particle will follow
- derived data (per streamline)
  - curvature, torsion, tortuosity
  - signature: complex weighted combination
  - compute cluster hierarchy across all signatures
  - encode: color and opacity by cluster
- tasks
  - find features, query shape
- scalability
  - millions of samples, hundreds of streamlines



*[Similarity Measures for Enhancing Interactive Streamline Seeding. McLoughlin,. Jones, Laramee, Malki, Masters, and. Hansen. IEEE Trans. Visualization and Computer Graphics 19:8 (2013), 1342–1353.]*

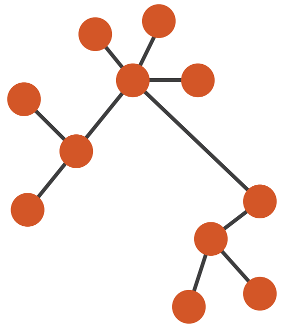
# Ch 9: Arrange Network Data

# Arrange networks and trees

## ➔ Node–Link Diagrams

Connection Marks

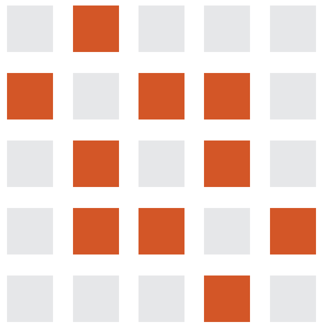
✓ NETWORKS ✓ TREES



## ➔ Adjacency Matrix

Derived Table

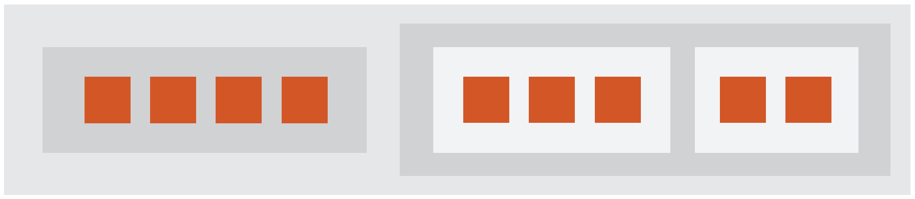
✓ NETWORKS ✓ TREES



## ➔ Enclosure

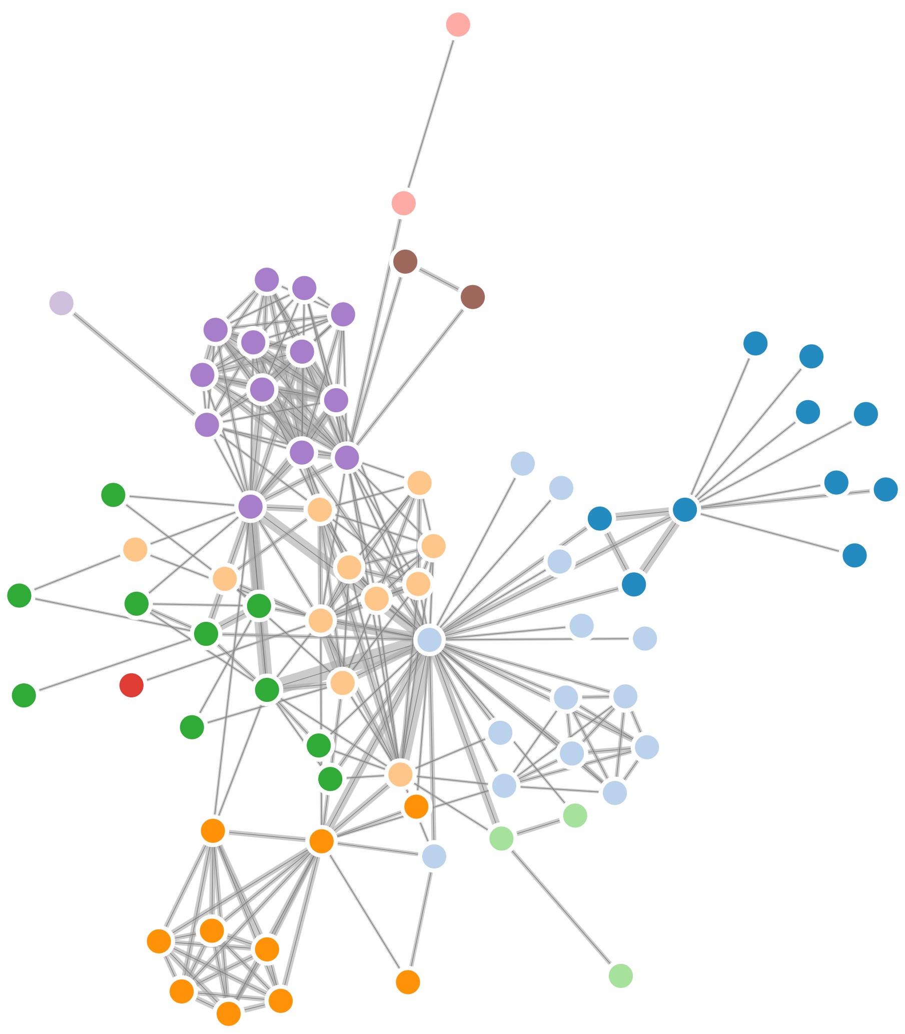
Containment Marks

✕ NETWORKS ✓ TREES



# Idiom: **force-directed placement**

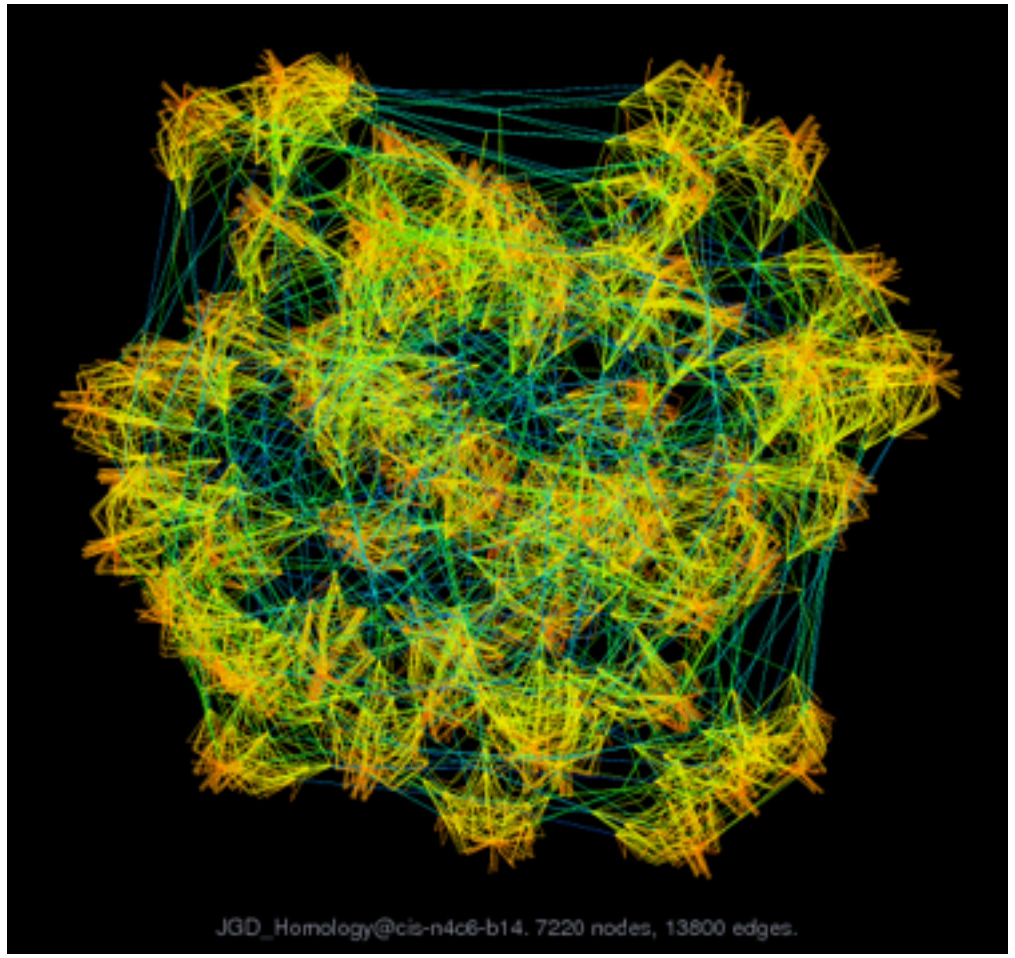
- visual encoding
  - link connection marks, node point marks
- considerations
  - spatial position: no meaning directly encoded
    - left free to minimize crossings
  - proximity semantics?
    - sometimes meaningful
    - sometimes arbitrary, artifact of layout algorithm
    - tension with length
      - long edges more visually salient than short
- tasks
  - explore topology; locate paths, clusters
- scalability
  - node/edge density E < 4N



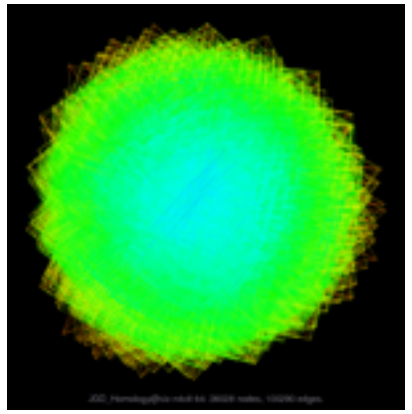
http://mbostock.github.com/d3/ex/force.html

# Idiom: **sfdp** (multi-level force-directed placement)

- data
  - –original: network
  - –derived: cluster hierarchy atop it
- considerations
  - –better algorithm for same encoding technique
    - same: fundamental use of space
    - hierarchy used for algorithm speed/quality but not shown explicitly
    - (more on algorithm vs encoding in afternoon)
- scalability
  - –nodes, edges: 1K-10K
  - –hairball problem eventually hits



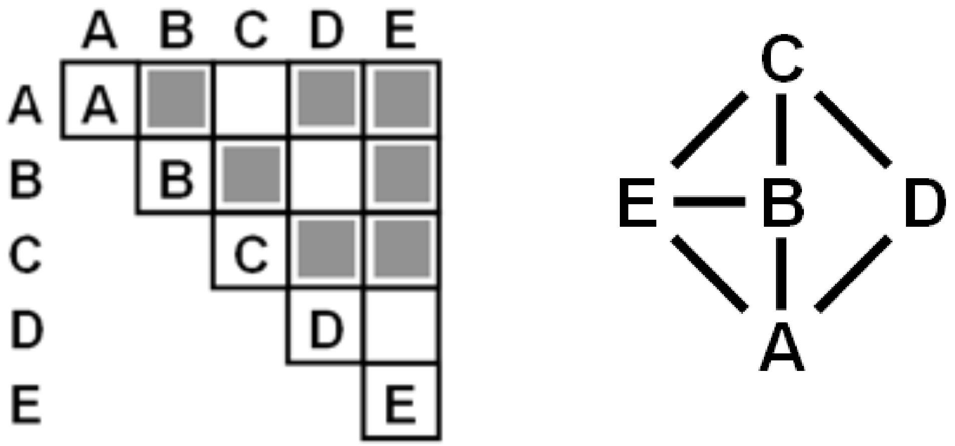
*[Efficient and high quality force-directed graph drawing. Hu. The Mathematica Journal 10:37–71, 2005.]*



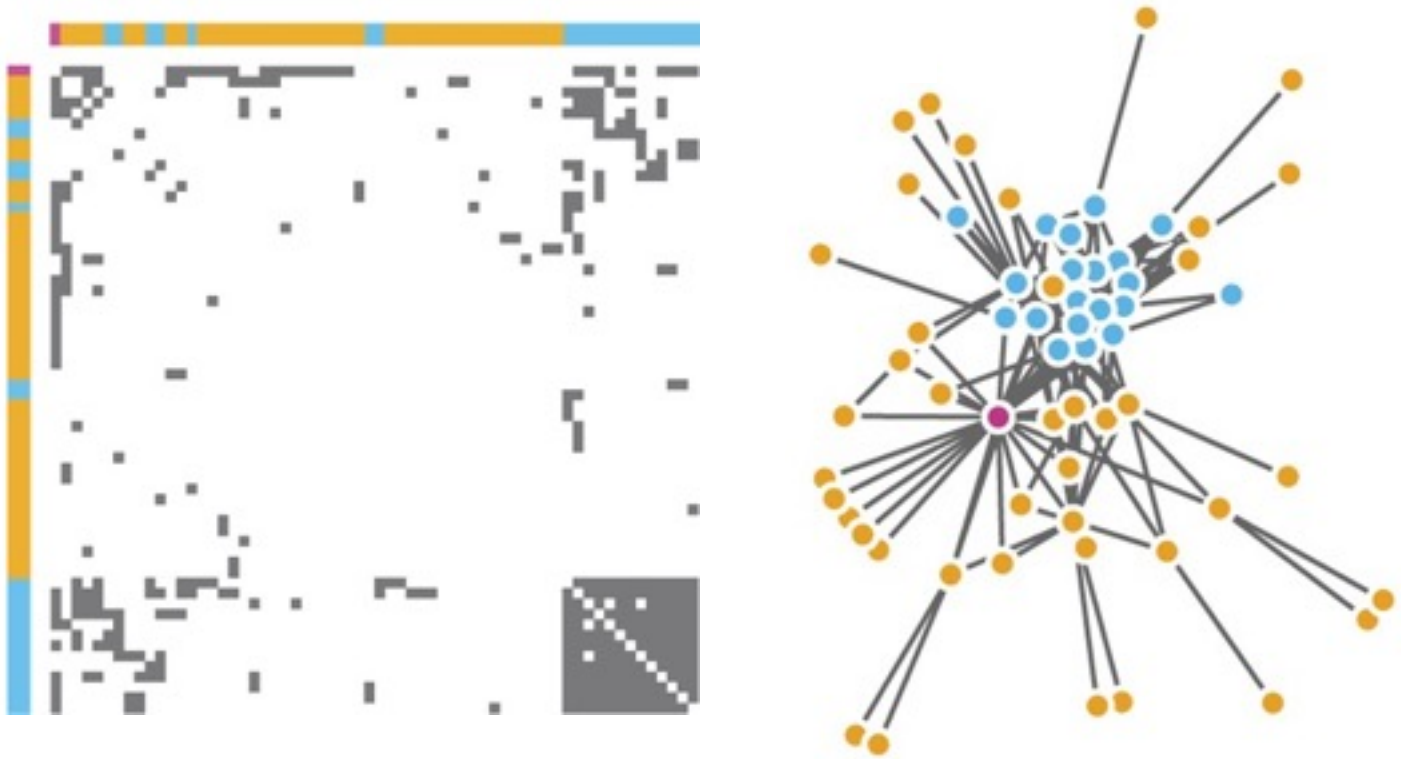
http://www.research.att.com/yifanhu/GALLERY/GRAPHS/index1.html

# Idiom: **adjacency matrix view**

- data: network
  - –transform into same data/encoding as heatmap
- derived data: table from network
  - –1 quant attrib
    - weighted edge between nodes
  - –2 categ attribs: node list x 2
- visual encoding
  - –cell shows presence/absence of edge
- scalability
  - –1K nodes, 1M edges



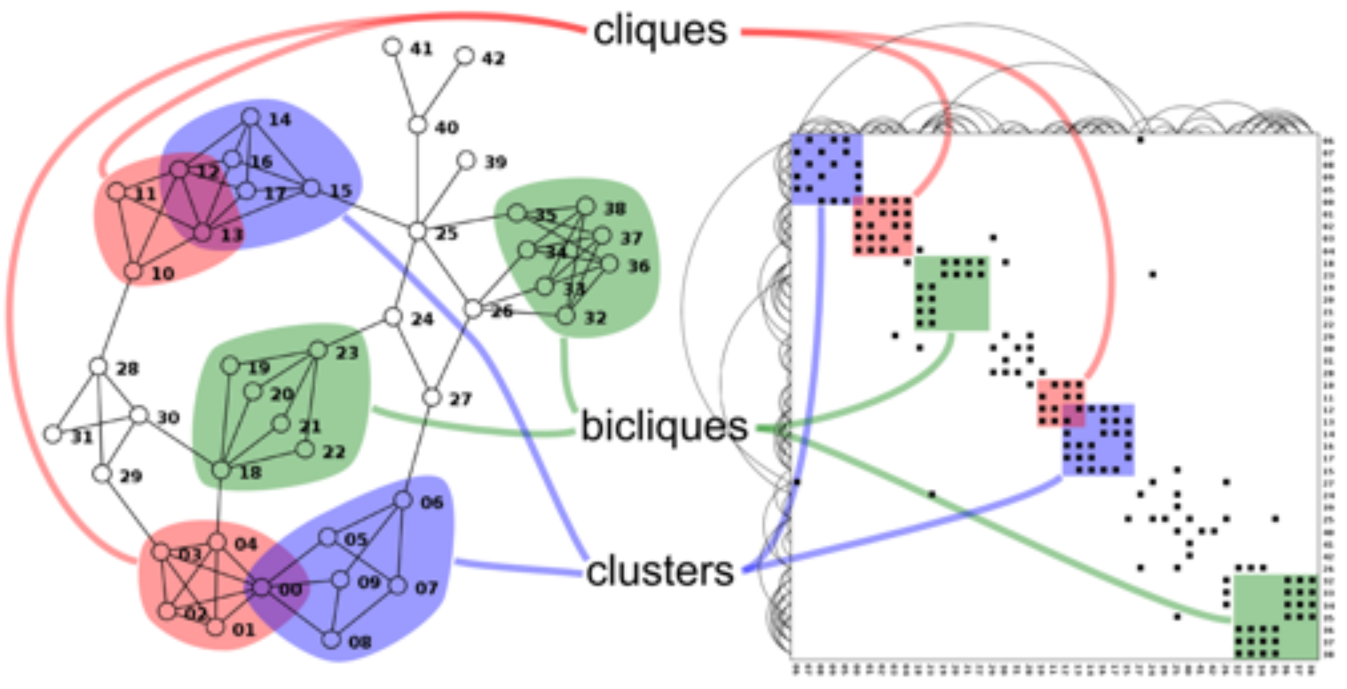
*[NodeTrix: a Hybrid Visualization of Social Networks. Henry, Fekete, and McGuffin. IEEE TVCG (Proc. InfoVis) 13(6):1302-1309, 2007.]*



*[Points of view: Networks. Gehlenborg and Wong. Nature Methods 9:115.]*

# Connection vs. adjacency comparison

- adjacency matrix strengths
  - predictability, scalability, supports reordering
  - some topology tasks trainable
- node-link diagram strengths
  - topology understanding, path tracing
  - intuitive, no training needed
- empirical study
  - node-link best for small networks
  - matrix best for large networks
    - if tasks don't involve topological structure!

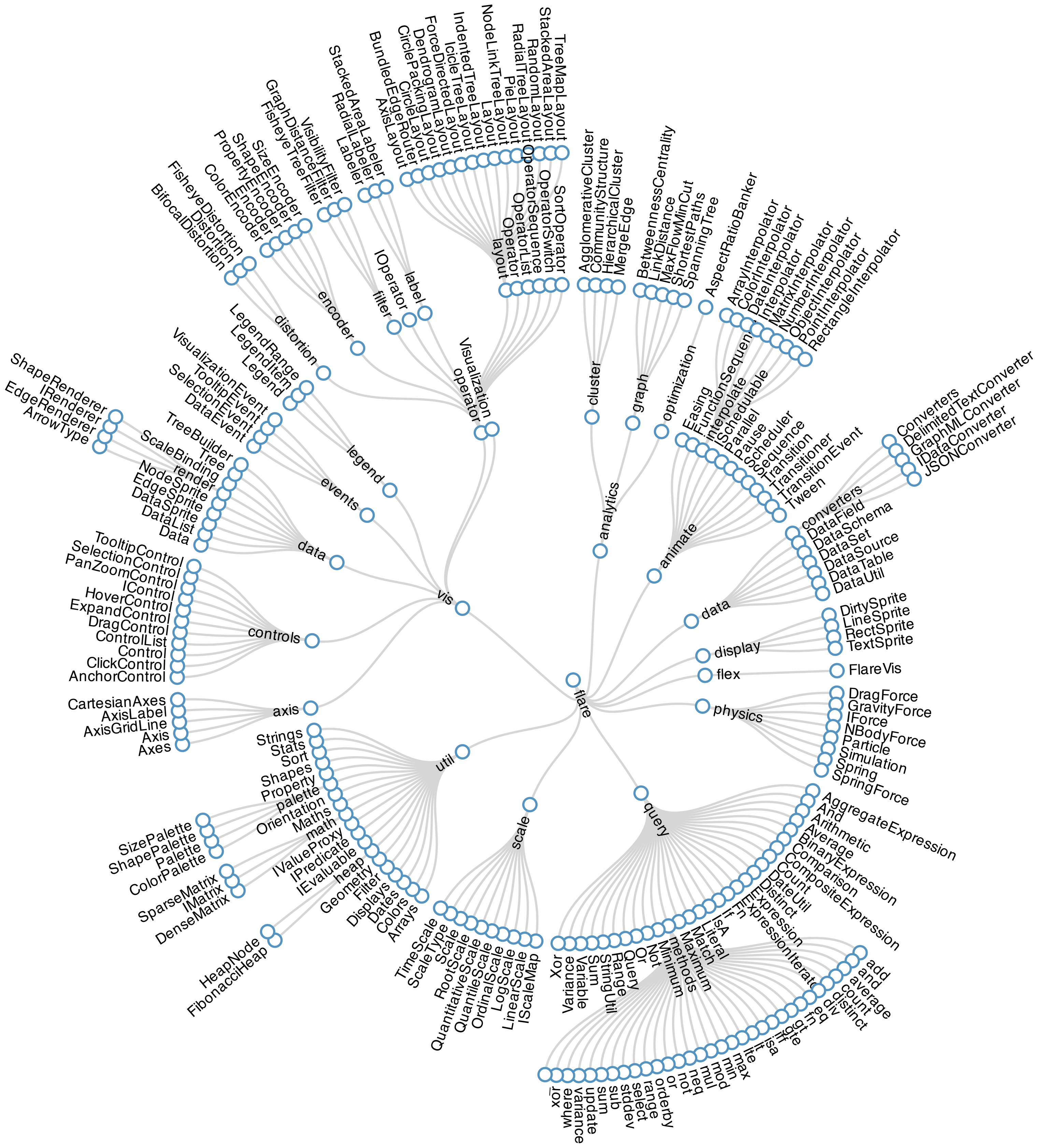


http://www.michaelmcguffin.com/courses/vis/patternsInAdjacencyMatrix.png

*[On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. Ghoniem, Fekete, and Castagliola. Information Visualization 4:2 (2005), 114–135.]*

# Idiom: **radial node-link tree**

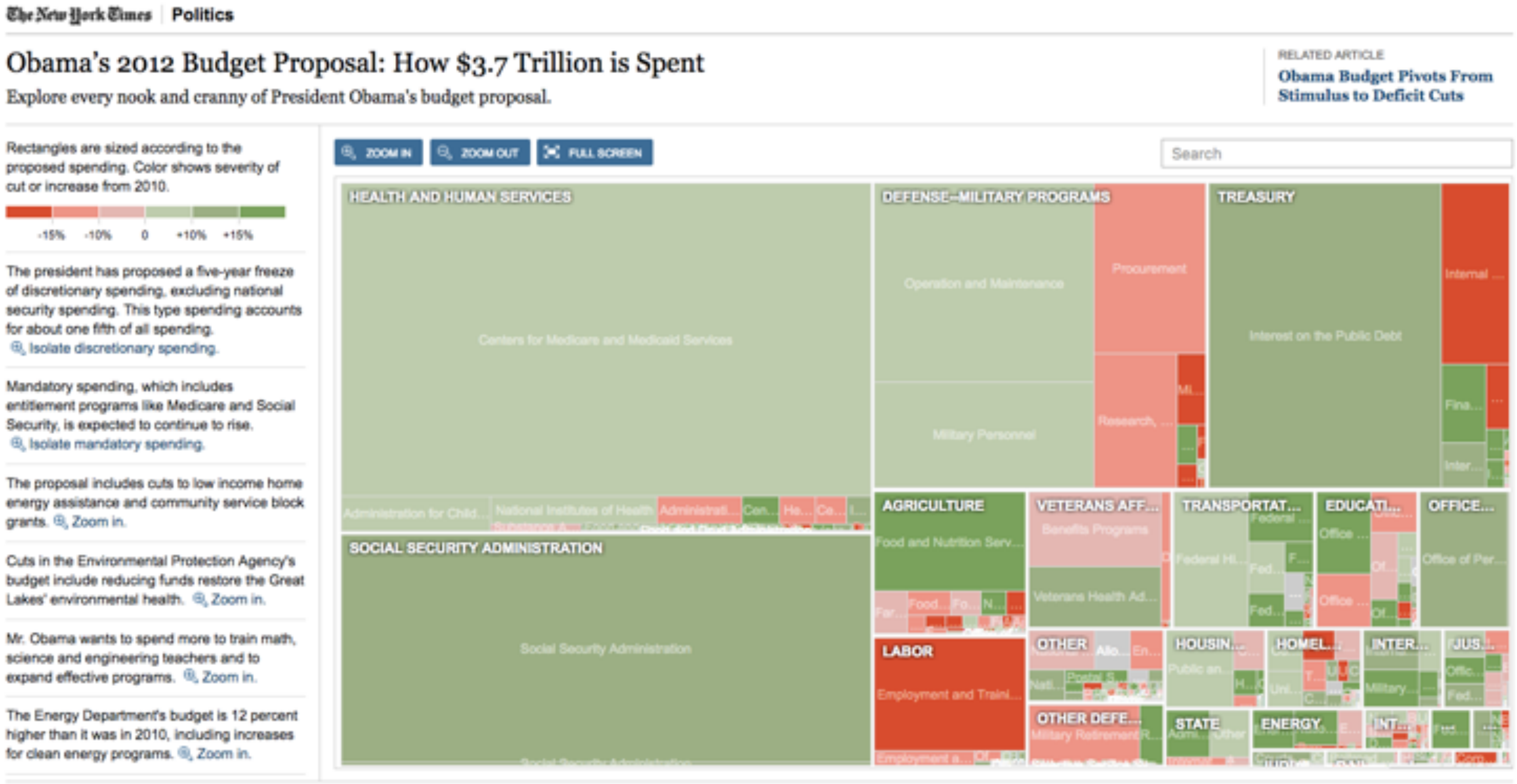
- data
  - tree
- encoding
  - link connection marks
  - point node marks
  - radial axis orientation
    - angular proximity: siblings
    - distance from center: depth in tree
- tasks
  - understanding topology, following paths
- scalability
  - 1K - 10K nodes



http://mbostock.github.com/d3/ex/tree.html

# Idiom: **treemap**

- data
  - tree
  - 1 quant attrib at leaf nodes
- encoding
  - area containment marks for hierarchical structure
  - rectilinear orientation
  - size encodes quant attrib
- tasks
  - query attribute at leaf nodes
- scalability
  - 1M leaf nodes



The New York Times | Politics

Obama's 2012 Budget Proposal: How $3.7 Trillion is Spent

Explore every nook and cranny of President Obama's budget proposal.

RELATED ARTICLE

Obama Budget Pivots From Stimulus to Deficit Cuts

Rectangles are sized according to the proposed spending. Color shows severity of cut or increase from 2010.

The president has proposed a five-year freeze of discretionary spending, excluding national security spending. This type spending accounts for about one fifth of all spending. Isolate discretionary spending.

Mandatory spending, which includes entitlement programs like Medicare and Social Security, is expected to continue to rise. Isolate mandatory spending.

The proposal includes cuts to low income home energy assistance and community service block grants. Zoom in.

Cuts in the Environmental Protection Agency's budget include reducing funds restore the Great Lakes' environmental health. Zoom in.

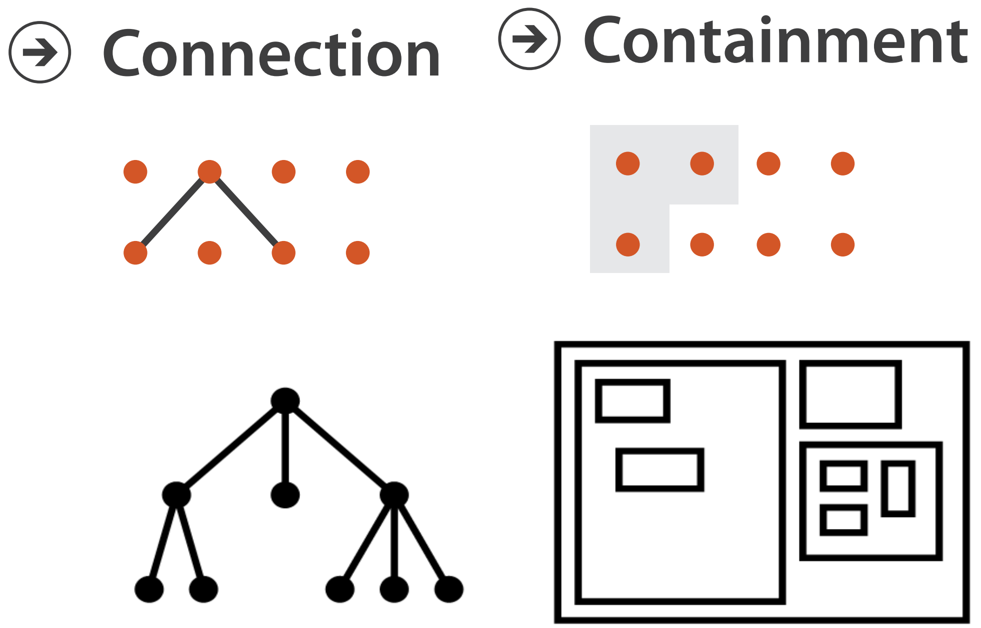
Mr. Obama wants to spend more to train math, science and engineering teachers and to expand effective programs. Zoom in.

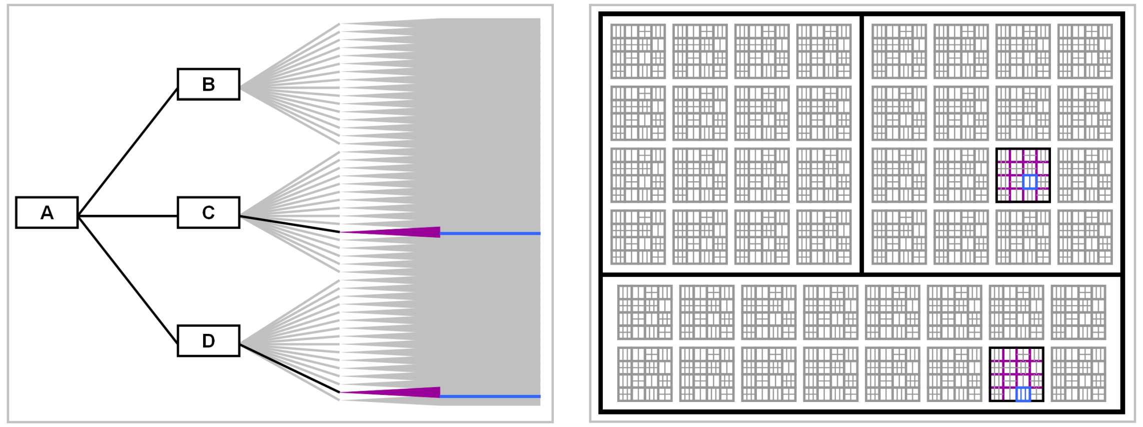
The Energy Department's budget is 12 percent higher than it was in 2010, including increases for clean energy programs. Zoom in.

http://www.nytimes.com/packages/html/newsgraphics/2011/0119-budget/index.html

# Link marks: Connection and containment

- marks as links (vs. nodes)
  - common case in network drawing
  - 1D case: connection
    - ex: all node-link diagrams
    - emphasizes topology, path tracing
    - networks and trees
  - 2D case: containment
    - ex: all treemap variants
    - emphasizes attribute values at leaves (size coding)
    - only trees



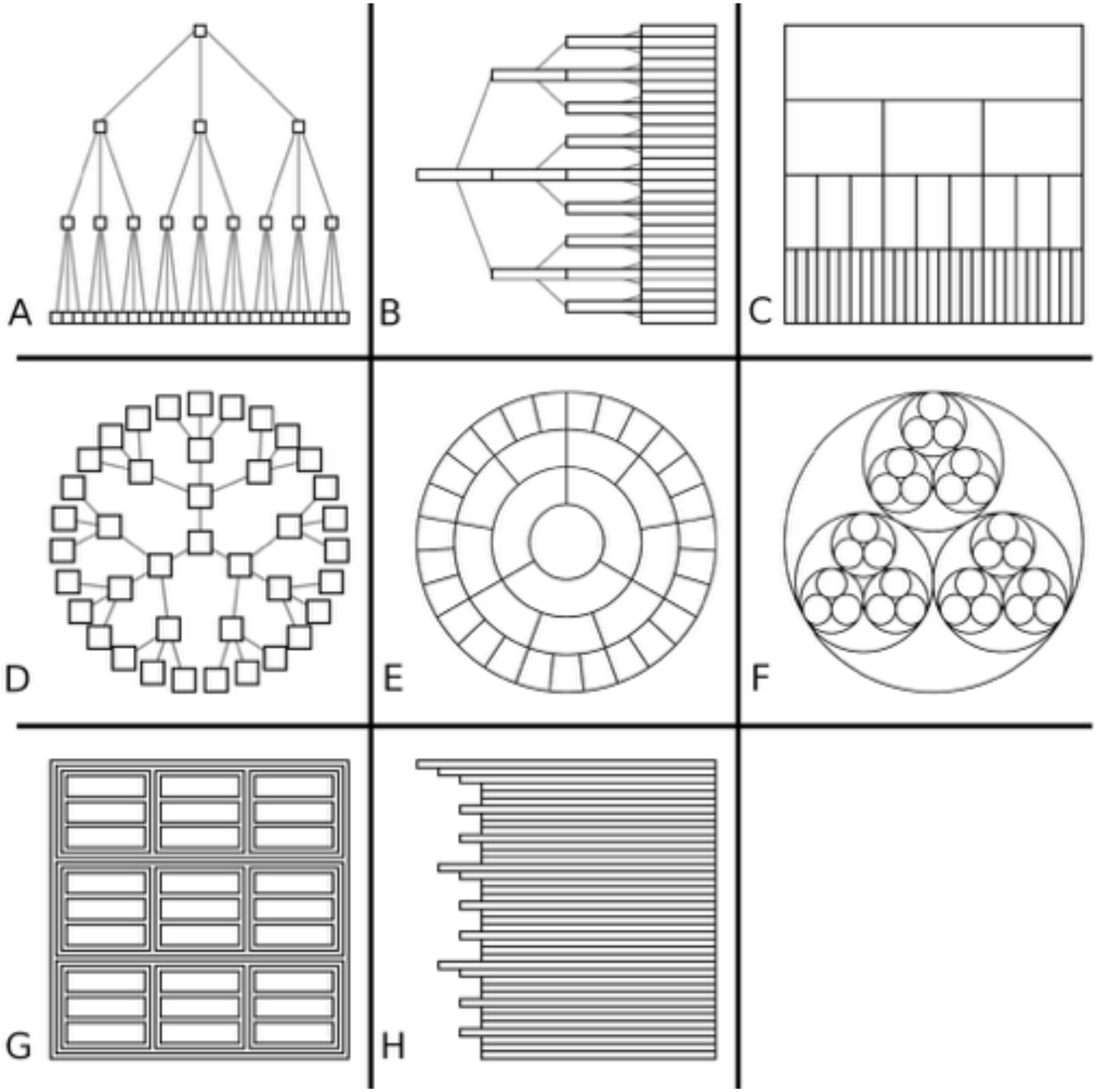


**Node–Link Diagram** **Treemap**

*[Elastic Hierarchies: Combining Treemaps and Node-Link Diagrams. Dong, McGuffin, and Chignell. Proc. InfoVis 2005, p. 57-64.]*

# Tree drawing idioms comparison

- data shown
  - link relationships
  - tree depth
  - sibling order
- design choices
  - connection vs containment link marks
  - rectilinear vs radial layout
  - spatial position channels
- considerations
  - redundant? arbitrary?
  - information density?
    - avoid wasting space

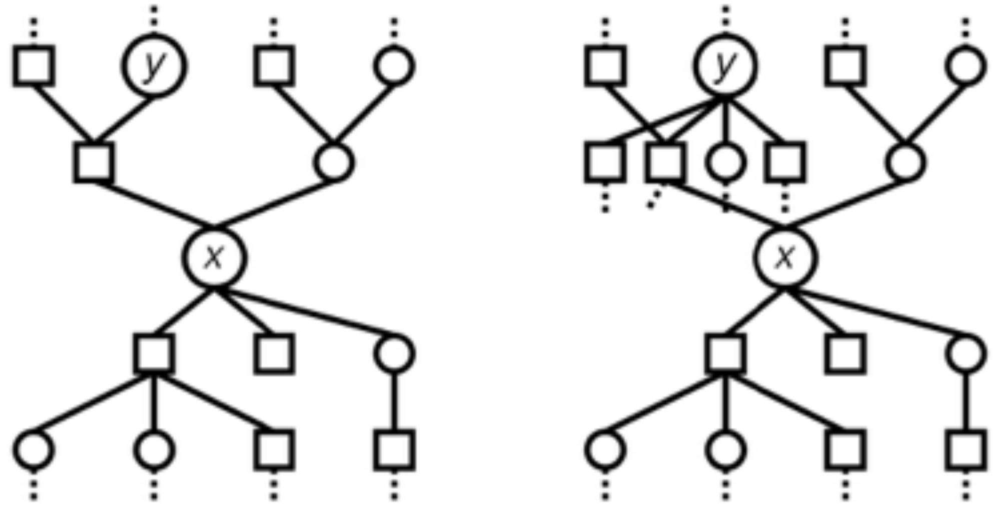


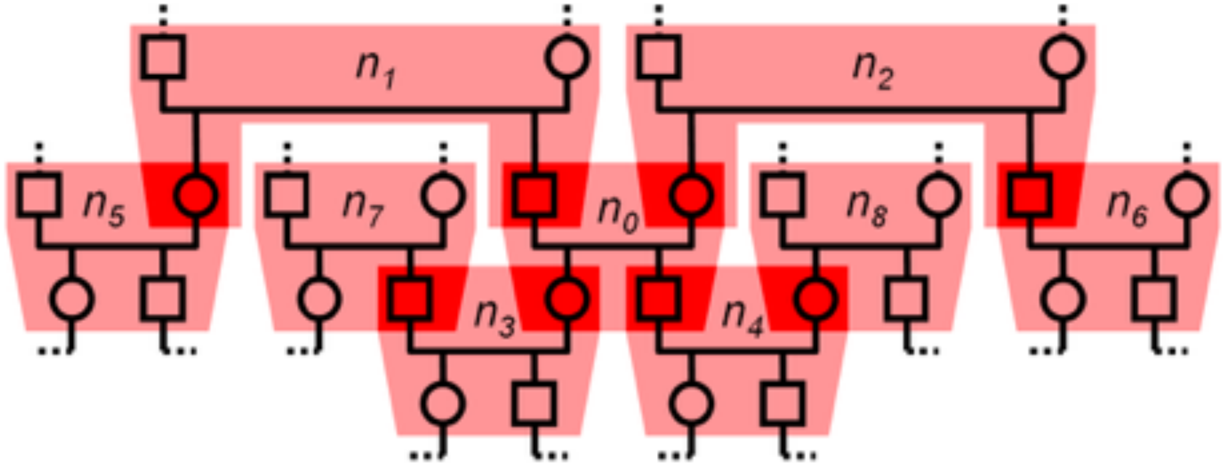
*[Quantifying the Space-Efficiency of 2D Graphical Representations of Trees. McGuffin and Robert. Information Visualization 9:2 (2010), 115–140.]*

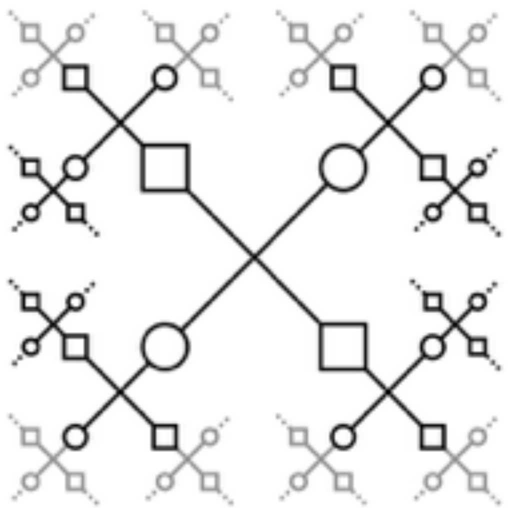
# Paper: Genealogical Graphs

# Genealogical graphs: Technique paper

- family tree is a misnomer
  - single person has tree of ancestors, tree of descendants
  - pedigree collapse inevitable
    - diamond in ancestor graph
- crowding problem
  - exponential
- fractal layout
  - poor info density
  - no spatial ordering for generations



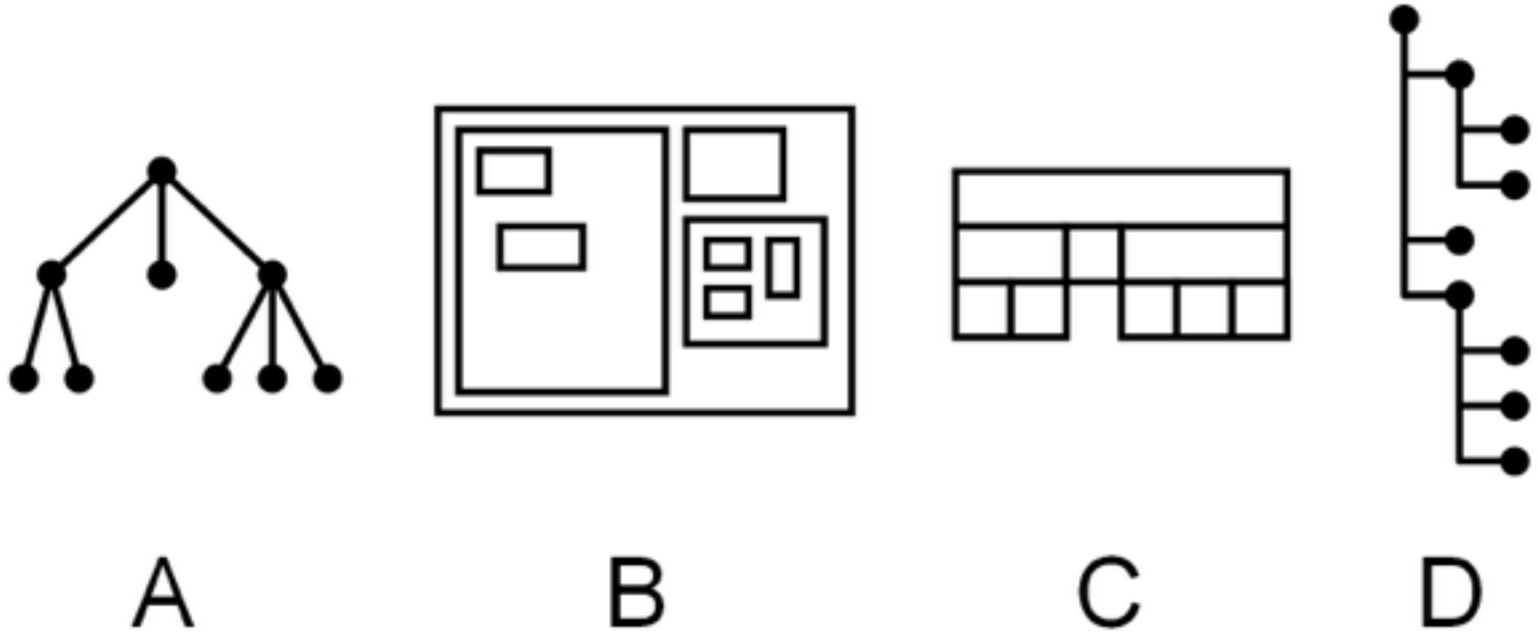




*[Fig 2, 6, 7. Interactive Visualization of Genealogical Graphs. Michael J. McGuffin, Ravin Balakrishnan. Proc. InfoVis 2005, pp 17-24.]*

# Layouts

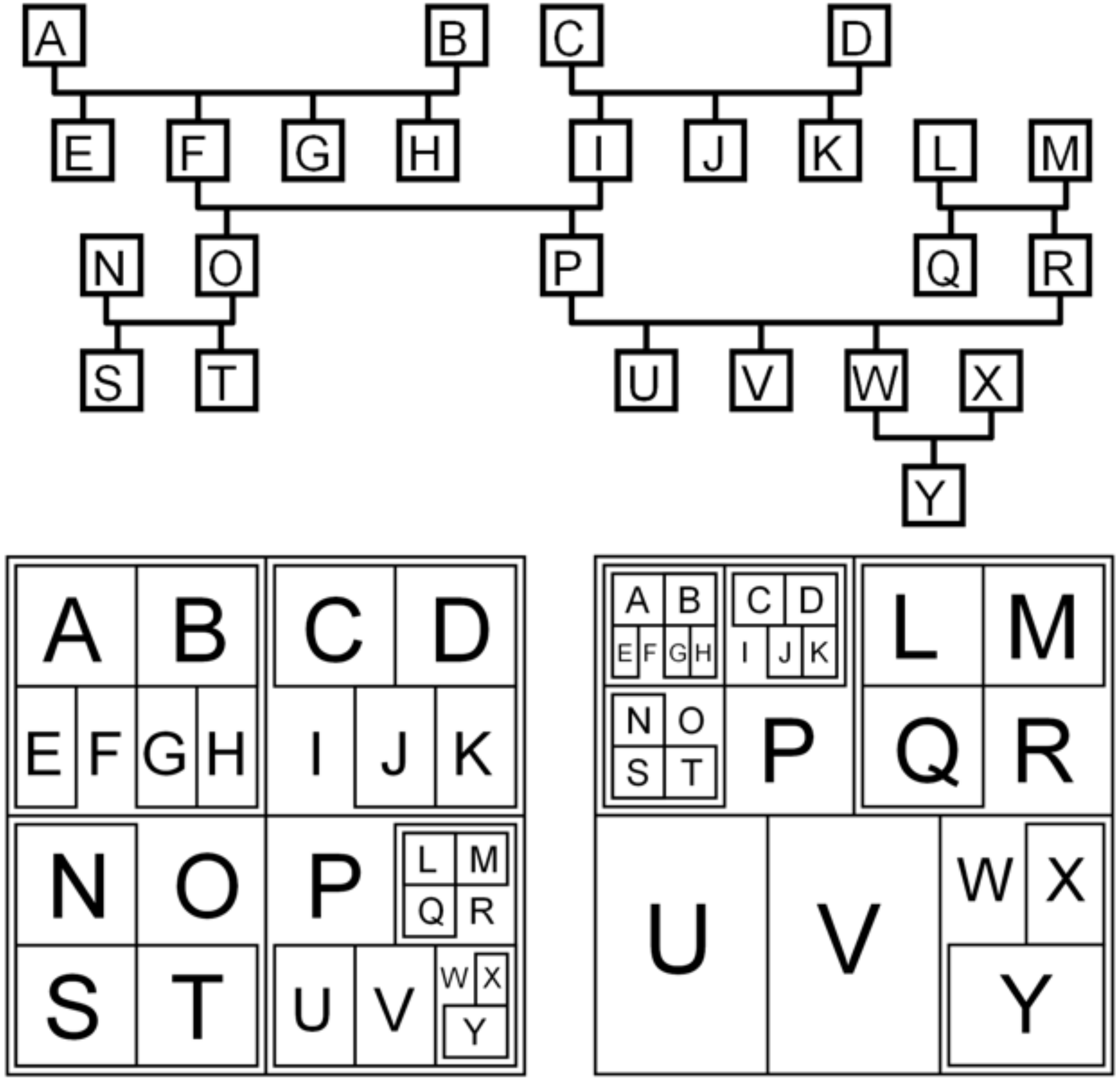
- rooted trees: standard layouts
  - connection
  - containment
  - adjacent aligned position
  - indented position



*[Fig 8. Interactive Visualization of Genealogical Graphs. Michael J. McGuffin, Ravin Balakrishnan. Proc. InfoVis 2005, pp 17-24.]*

# Layouts

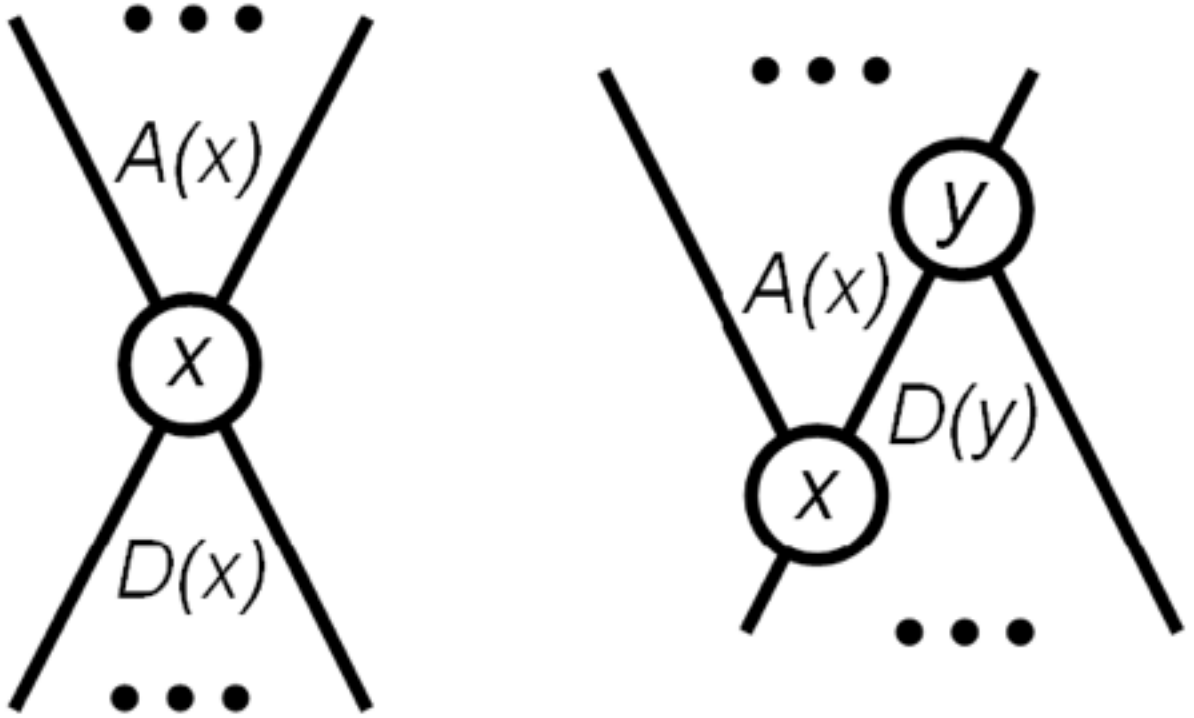
- free trees
  - no root
- adapting rooted methods
  - temporary root for given focus
  - containment (nested)

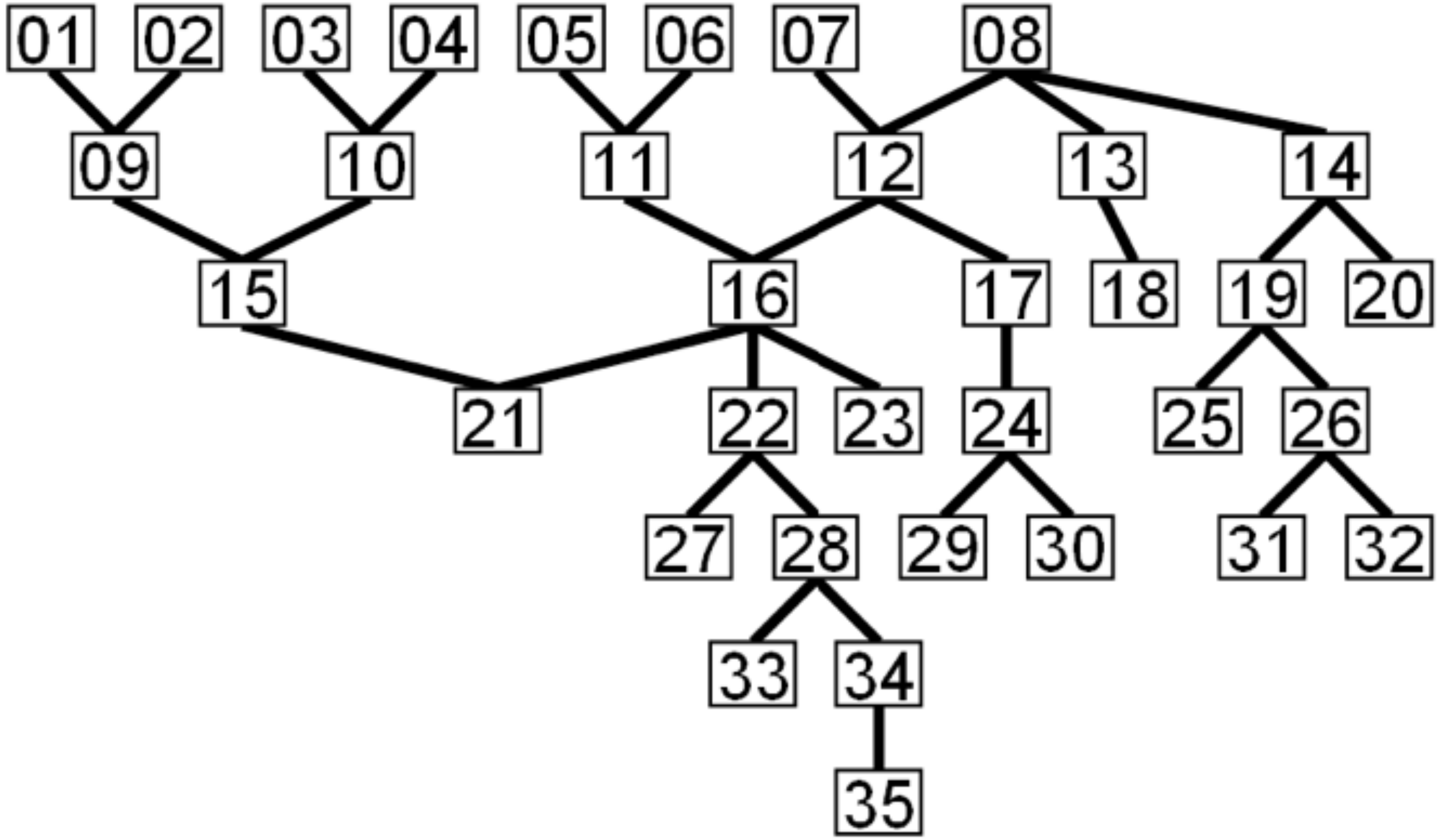


*[Fig 9. Interactive Visualization of Genealogical Graphs. Michael J. McGuffin, Ravin Balakrishnan. Proc. InfoVis 2005, pp 17-24.]*

# Dual trees abstraction

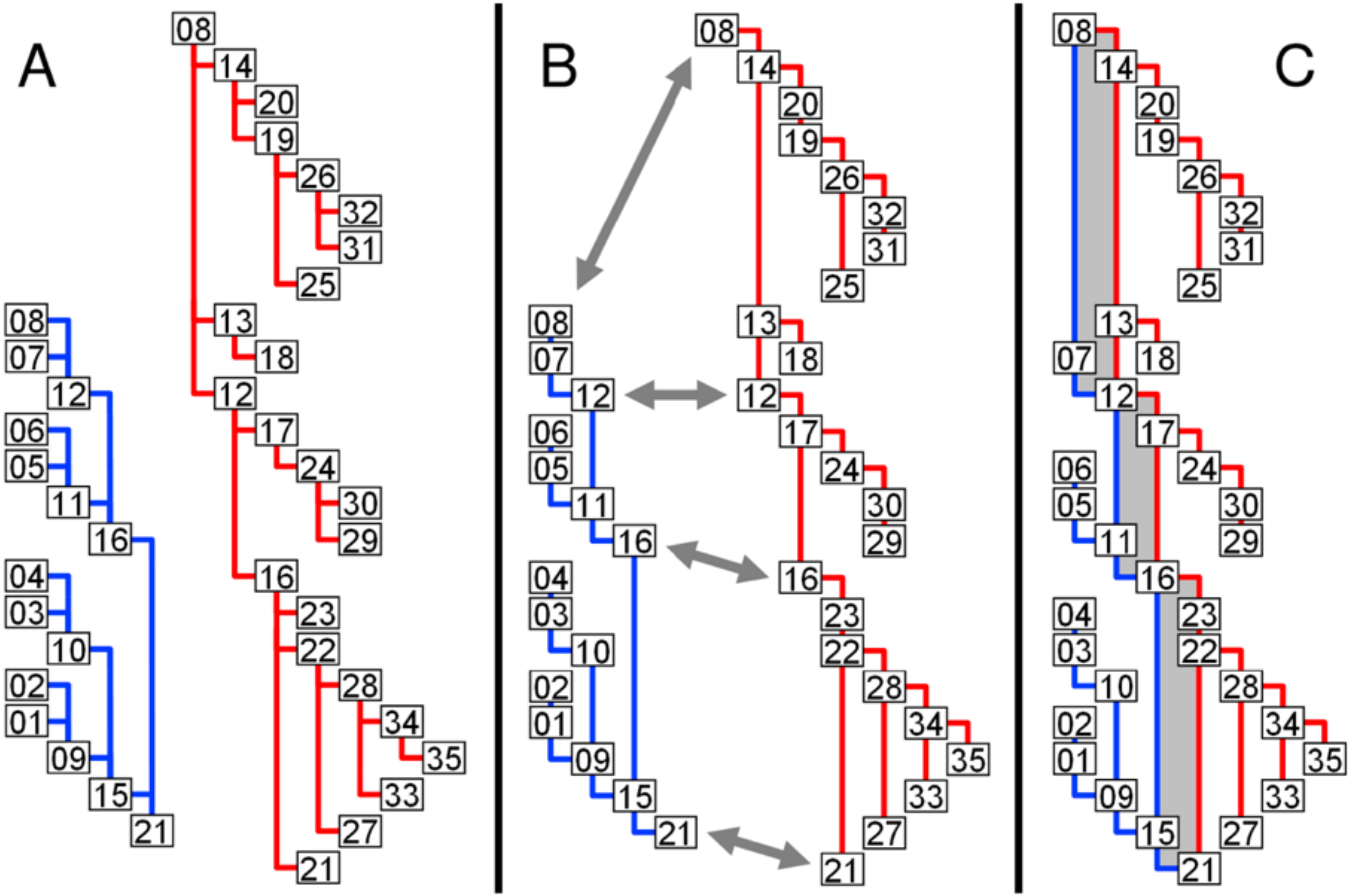
- explore canonical subsets and combinations, easy to interpret, scales well
- no crossings, nodes ordered by generation
- doubly rooted: x leftmost descend, y rightmost ancestor
  - offset roots from hourglass diagram





*[Fig 10. Interactive Visualization of Genealogical Graphs. Michael J. McGuffin, Ravin Balakrishnan. Proc. InfoVis 2005, pp 17-24.]*

# Indented, flipped, combined

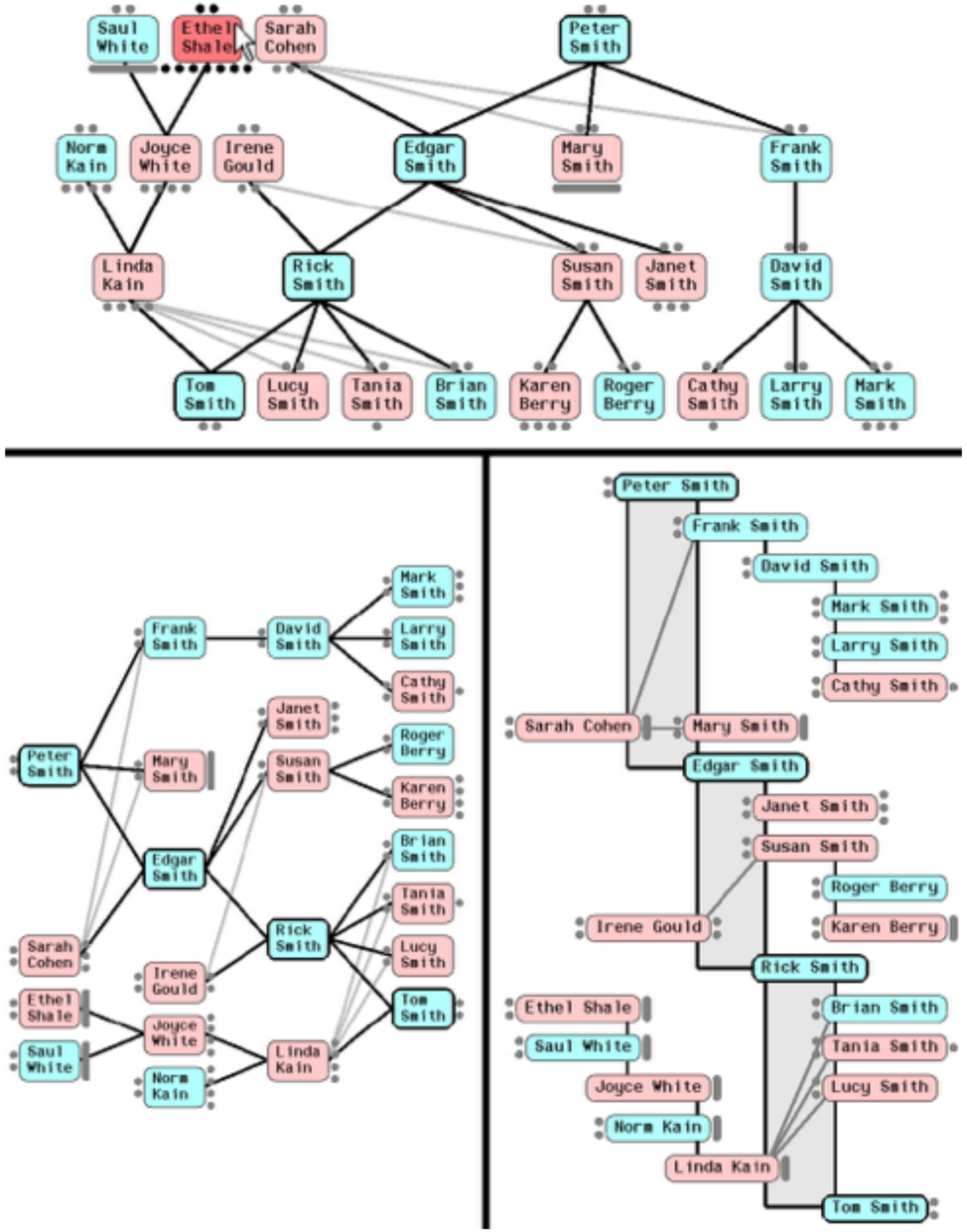


*[Fig 11. Interactive Visualization of Genealogical Graphs. Michael J. McGuffin, Ravin Balakrishnan. Proc. InfoVis 2005, pp 17-24.]*

# Another example

- vertical connection
- horizontal connection
- indented

- upcoming chapters
  - layering
  - aggregation



*[Fig 13. Interactive Visualization of Genealogical Graphs. Michael J. McGuffin, Ravin Balakrishnan. Proc. InfoVis 2005, pp 17-24.]*

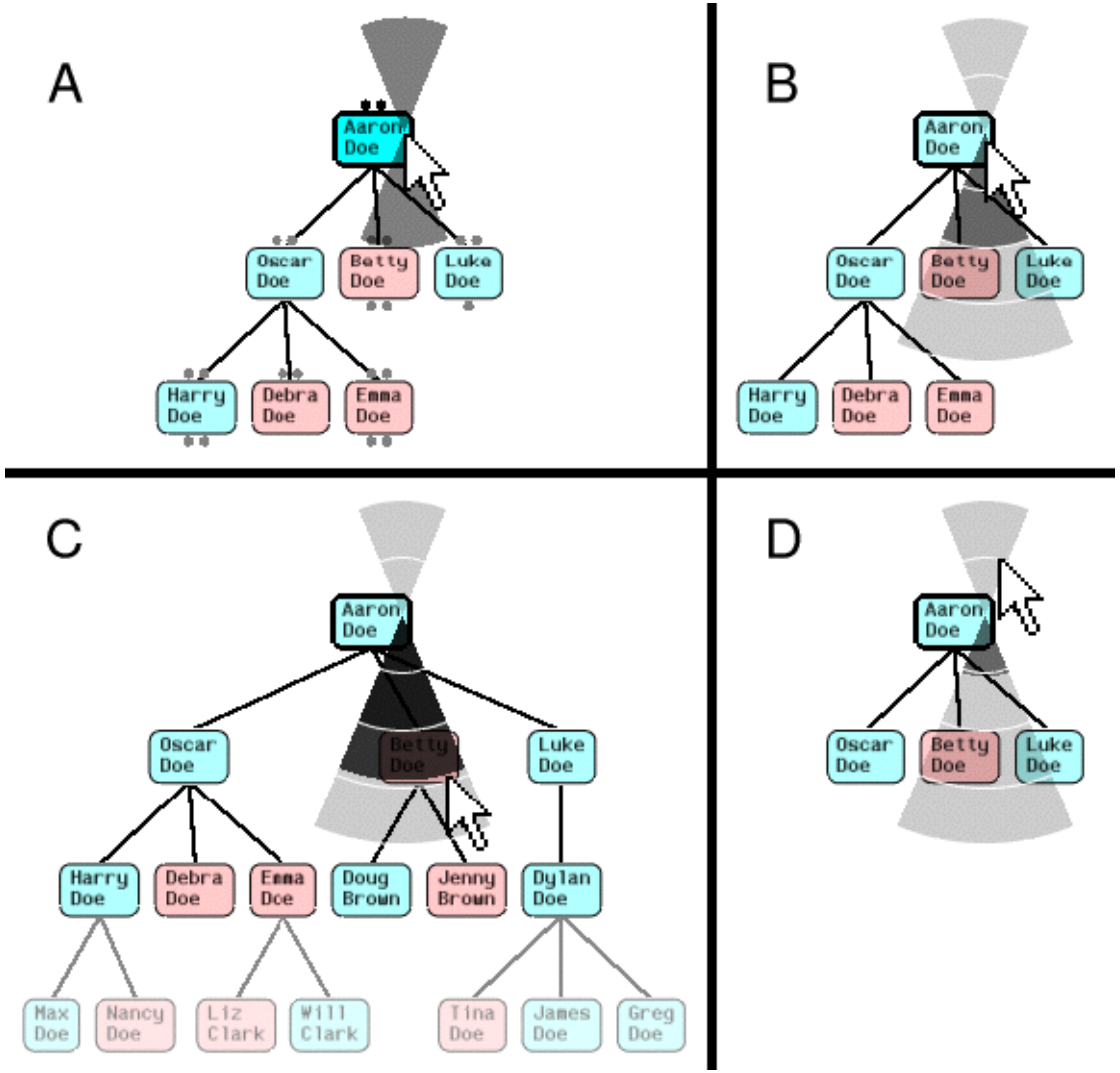
# Interaction as fundamental to design

- navigation
  - topological navigation via collapse/expand on selection
    - parents, children
    - expand can trigger rotation
      - collapsing others
      - layout driven by navigation
  - geometric zoom/pan
  - constrained navigation: automatic camera framing
- animated transitions
  - 3 phases: fade out, move, fade in
- mouseover hover
  - preview dots: expand if collapsed

*[Fig 14. Interactive Visualization of Genealogical Graphs. Michael J. McGuffin, Ravin Balakrishnan. Proc. InfoVis 2005, pp 17-24.]*

# Custom widget

- popup marking menu
  - –flick up or down, ballistic
  - –subtree drag-out widget



*[Fig 14. Interactive Visualization of Genealogical Graphs. Michael J. McGuffin, Ravin Balakrishnan. Proc. InfoVis 2005, pp 17-24.]*

# Paper: ABySS-Explorer

# ABySS-Explorer: Design study

- reconstructing genome with ABySS algorithm (Assembly By Short Sequences)
- domain task
  - go from short subsequences to ***contigs***, long contiguous sequences
  - extensive automatic support, but still human in the loop for visual inspection and manual editing
  - ambiguities, like repetitions longer than read length
- data, domain:abstract
  - millions of reads of 25-100 nucleotides (nt): strings
  - read coverage, proxy for quality: quant attrib
  - read pairing distances, proxy for size distribution: quant

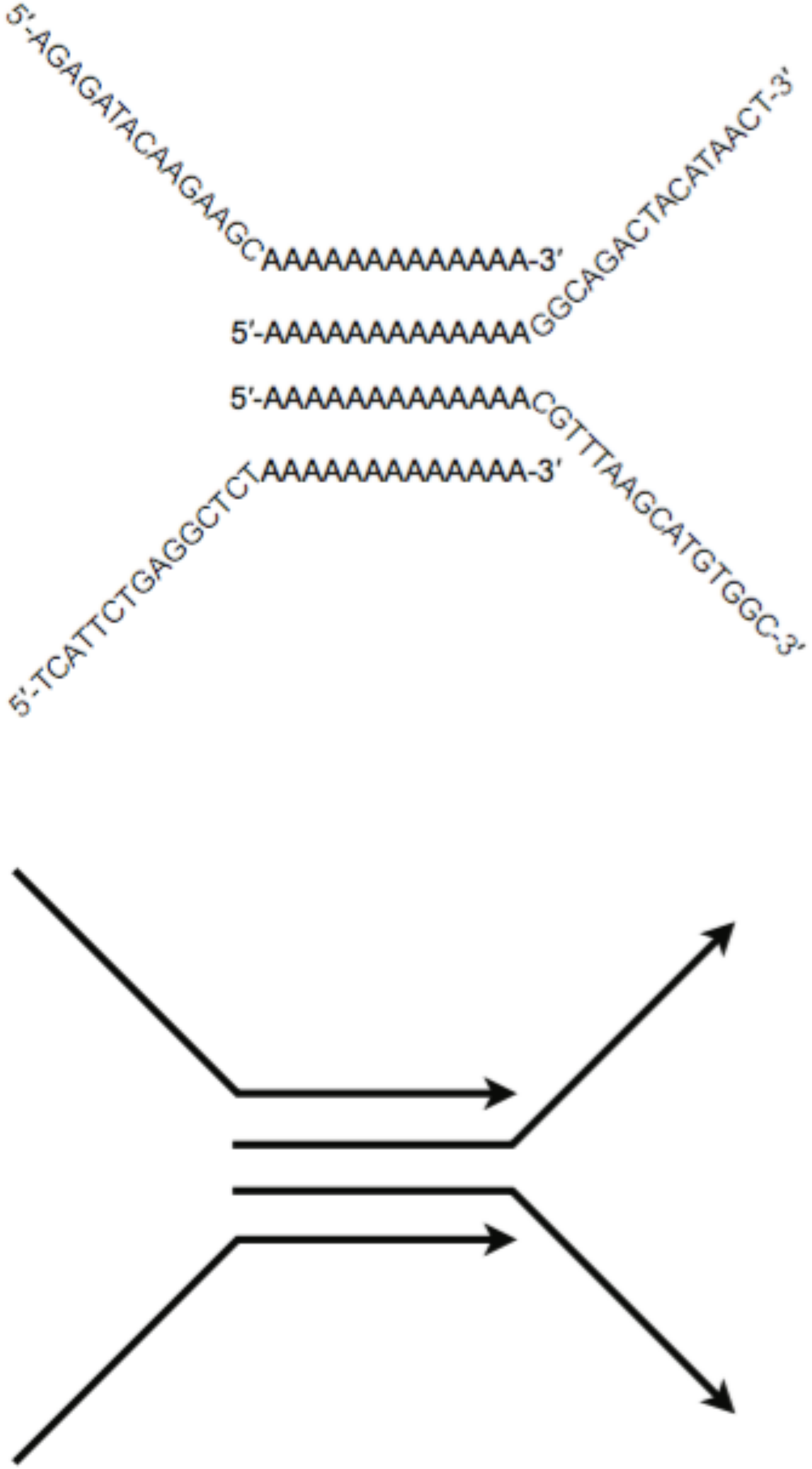

5′-AGAGATACAAGAAGCAAAAAAAAAAAAAAA-3′
5′-AAAAAAAAAAAAAAAGGCAGACTACATAACT-3′
5′-AAAAAAAAAAAAAAACGTTTAAGCATGTGGC-3′
5′-TCATTCTGAGGCTCTAAAAAAAAAAAAAAA-3′

*Fig 2. ABySS-Explorer: visualizing genome sequence assemblies. Nielsen, Jackman, Birol, Jones. TVCG 15(6):881-8, 2009 (Proc. InfoVis 2009).*

# Contigs: abstraction as derived network data

- derived data: de Bruijn graph/network
  - directed network, compact representation of sequence overlaps
  - node: contig
  - edge: overlap of k − 1 nt between two contigs
  - good for computing, bad for reasoning about sequence space
- derived data: dual de Bruijn graph
  - node: points of contig overlap
  - edge: contig
  - better match for arrow diagrams used in hand drawn sketches
- base layout: force-directed

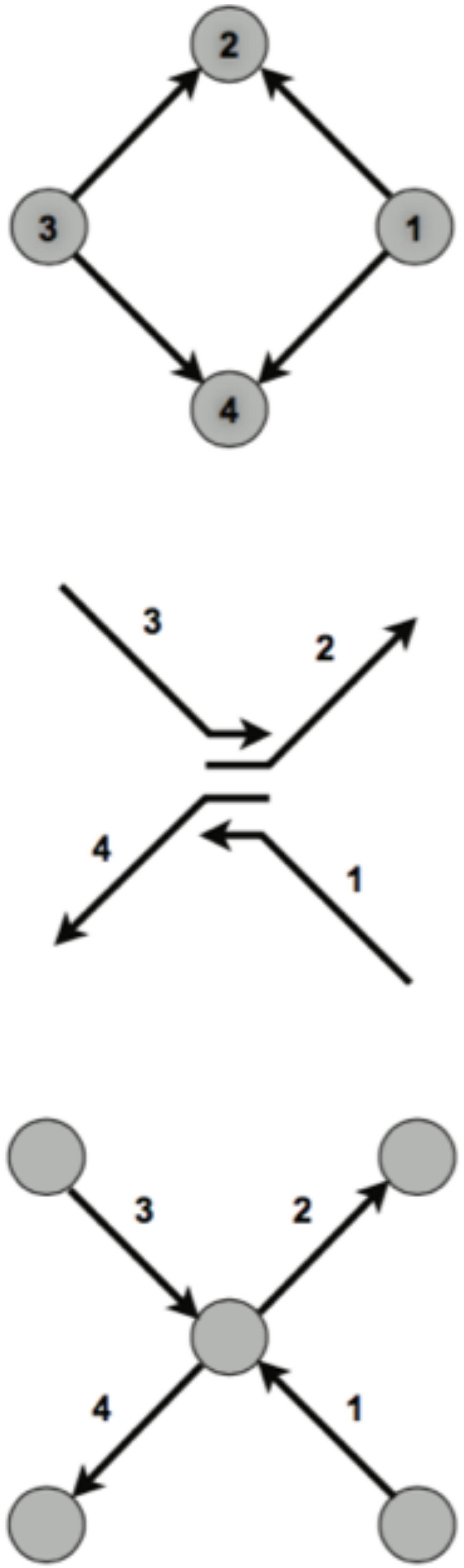


*Fig 3. ABySS-Explorer: visualizing genome sequence assemblies. Nielsen, Jackman, Birol, Jones. TVCG 15(6):881-8, 2009 (Proc. InfoVis 2009).*

# DNA as double stranded: idiom for encoding & interaction

- rejected option: 2 nodes per contig
  - excess clutter if one for each direction
  - choice at data abstraction level
- encoding & interaction idiom: *polar* node
  - encoding: upper vs lower attachment point
    - redundant with arc direction
      - large-scale visibility, without need to zoom
    - arbitrary but consistent
  - interaction: click to reverse direction
    - switches polarity of vertex connections
    - changes sign of label

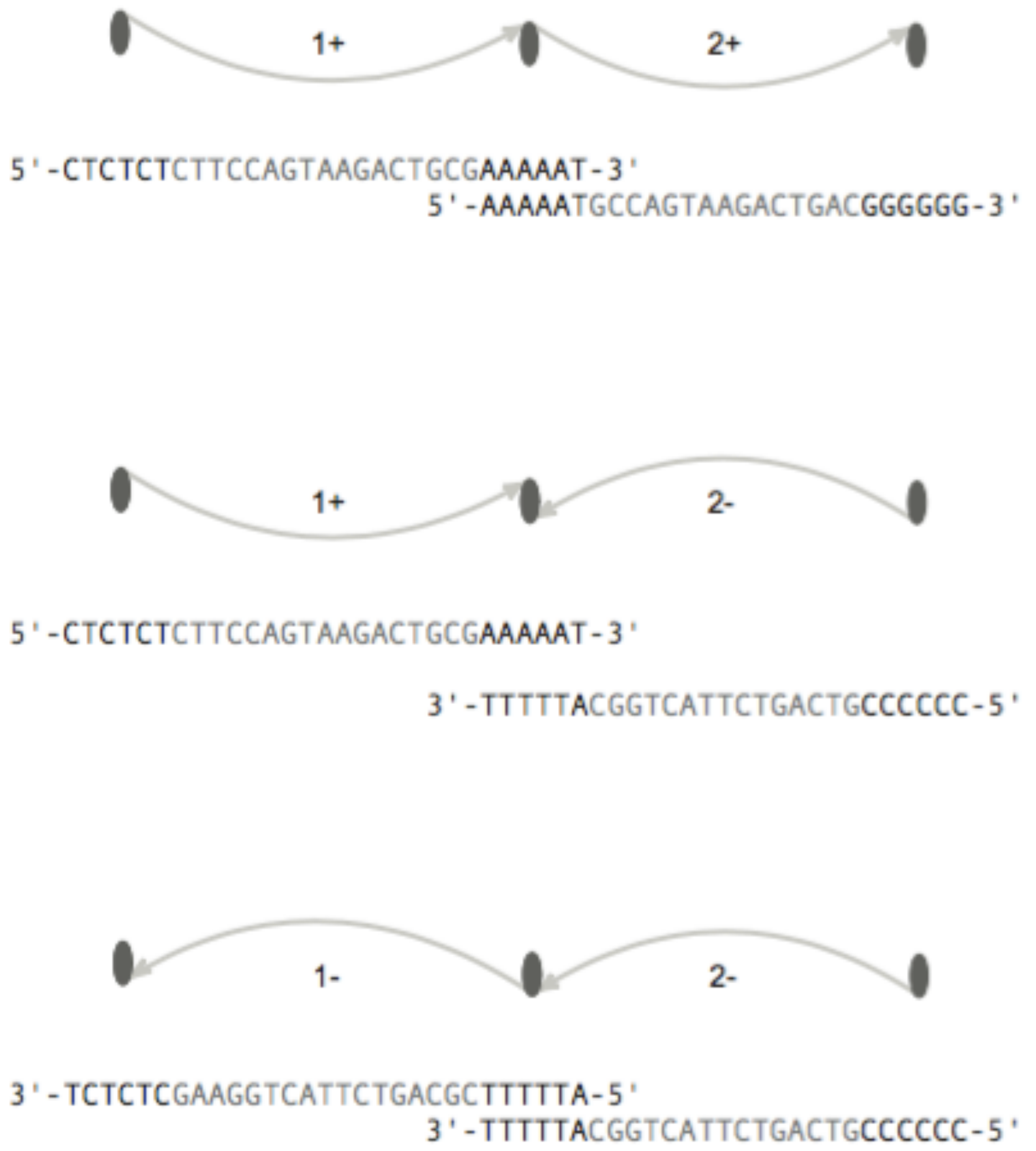

1+
2+
5'-CTCTCTCTTCCAGTAAGACTGCGAAAAAT-3'
5'-AAAAATGCCAGTAAGACTGACGGGGGG-3'
1+
2-
5'-CTCTCTCTTCCAGTAAGACTGCGAAAAAT-3'
3'-TTTTTACGGTCATTCTGACTGCCCCCC-5'
1-
2-
3'-TCTCTCGAAGGTCATTCTGACGCTTTTTA-5'
3'-TTTTTACGGTCATTCTGACTGCCCCCC-5'

*Fig 4. ABySS-Explorer: visualizing genome sequence assemblies. Nielsen, Jackman, Birol, Jones. TVCG 15(6):881-8, 2009 (Proc. InfoVis 2009).*

# Contig length: encoding

- rejected option: scale edge lengths by sequence lengths
  - short contigs are important sources of ambiguity, would be hard to distinguish
  - task guidance: only low-res judgements needed, relatively long or short
- encoding idiom: wave pattern
  - oscillation shows fixed number, shapes distinguishable
  - min amplitude at connections so edges visible
  - orientation with max amplitude asymmetric wrt start
    - rejected initial option: max in middle
    - rejected options:
      - color (keep for other attribute)
      - half-lines
      - curvature (used for polar nodes)
    - aligned with empirical guidance for tapered edges

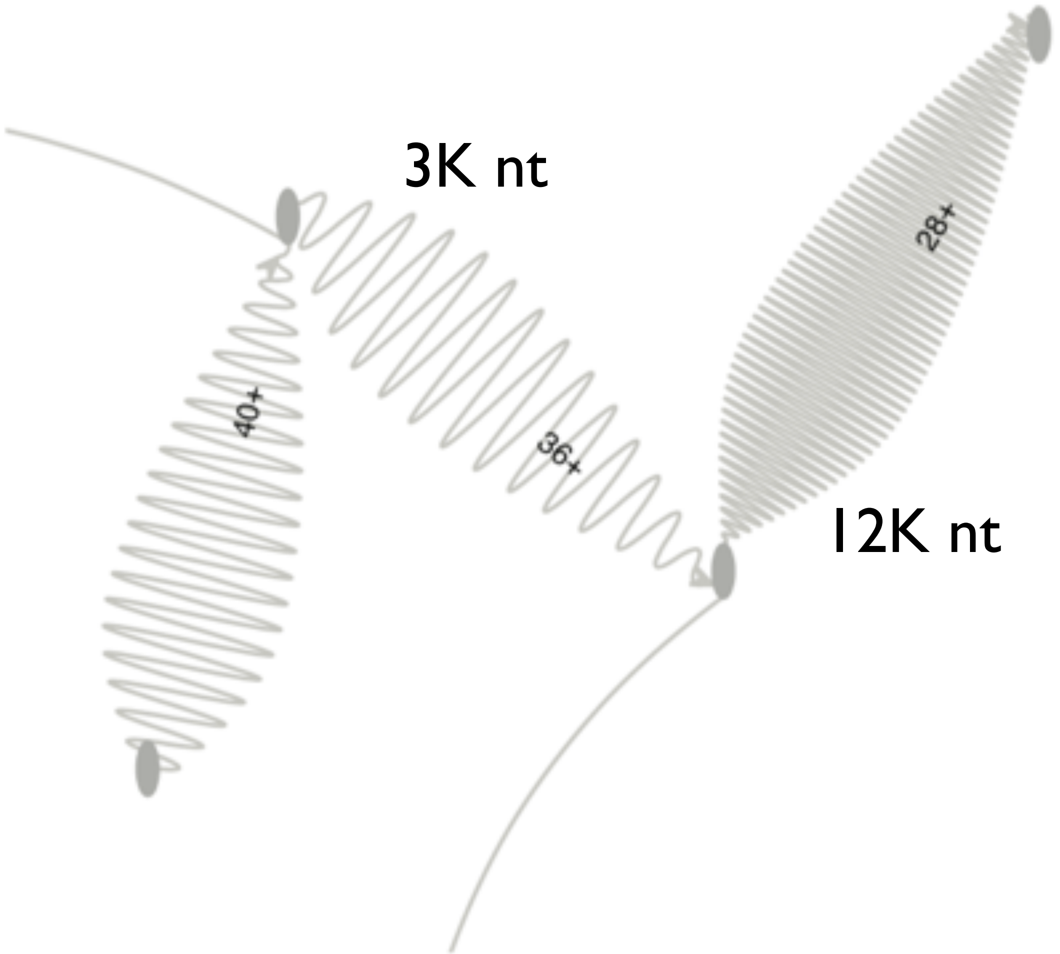


*Fig 5. ABySS-Explorer: visualizing genome sequence assemblies. Nielsen, Jackman, Birol, Jones. TVCG 15(6):881-8, 2009 (Proc. InfoVis 2009).*

# Contig coverage: encoding

- rejected options: luminance/lightness
  - not distinguishable given denseness variation from wave shapes
  - also problematic with desire for separable color/hue encoding
- chosen: line thickness
  - not distinguishable for extremely long contigs
  - can address by adjusting oscillation frequency to suitable size

# Read pairs: encoding

- data:
  - distance estimate
  - orientation
- encoding:
  - dashed line (shape channel for line mark)
    - implying inferred vs observed sequences
  - color for both dashed line and contig leaf
  - [same length as for contigs]
  - rejected initial option: line color alone
    - too ambiguous
  - interaction to fully resolve remaining ambiguity
  - or color by unambiguous paths in grey

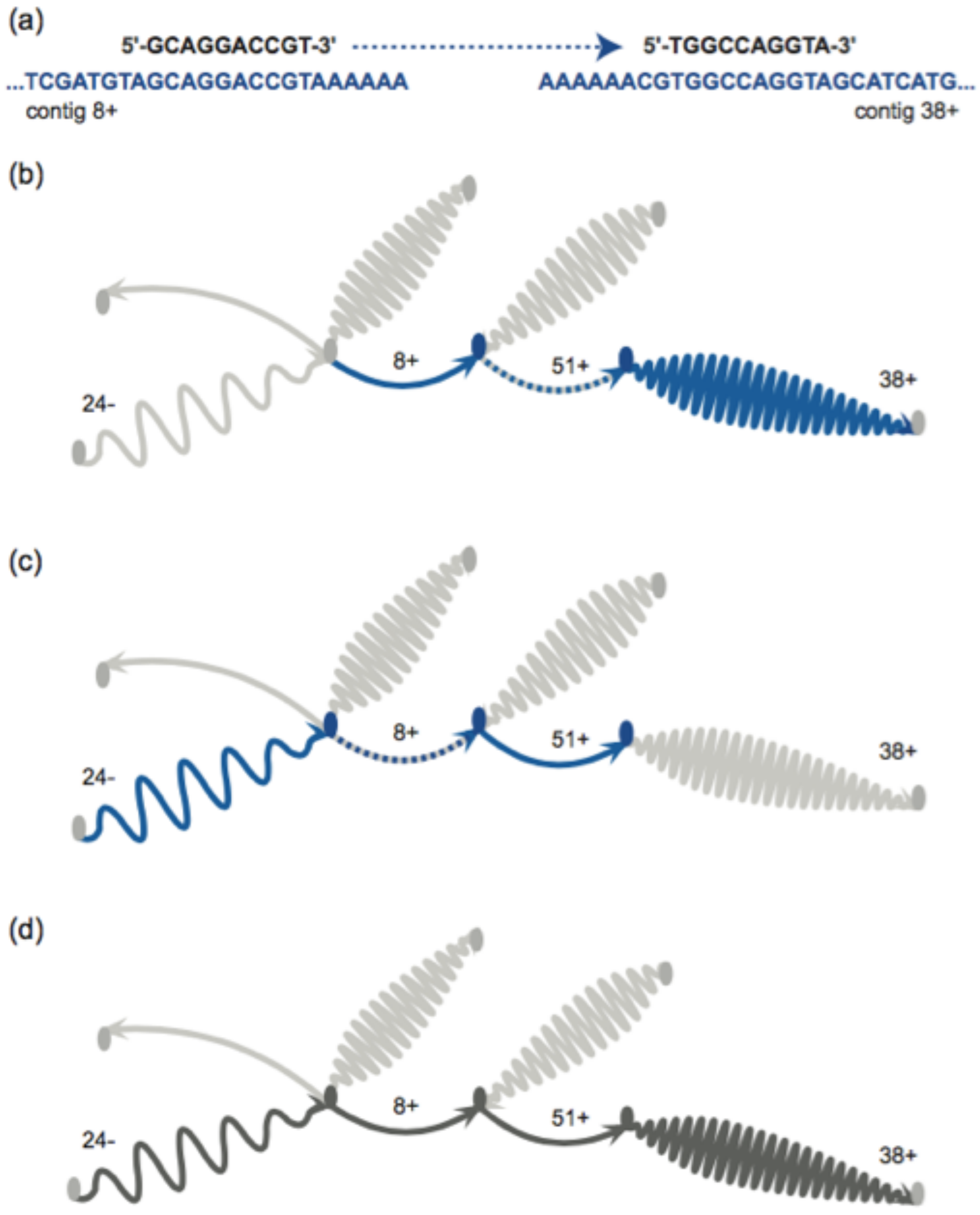
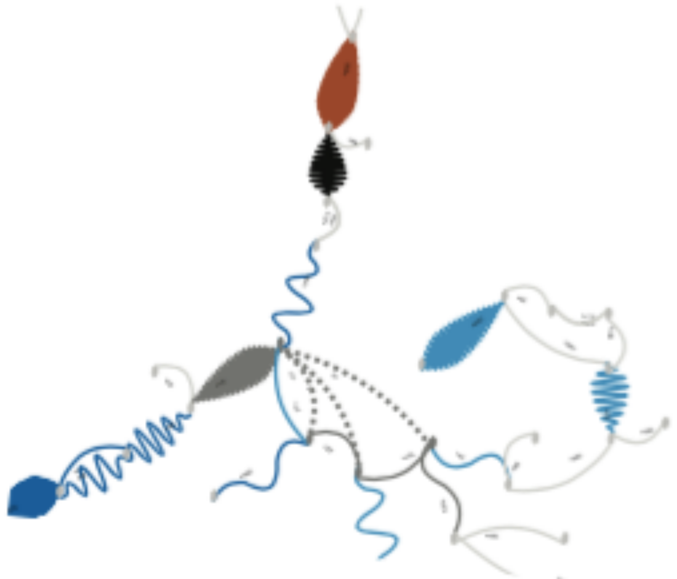


*Fig 6. ABySS-Explorer: visualizing genome sequence assemblies. Nielsen, Jackman, Birol, Jones. TVCG 15(6):881-8, 2009 (Proc. InfoVis 2009).*

# Displaying meta-data

- reserve color for additional attributes
- ex: color to compare reference human to lymphoma genome
  - inconsistencies visible as interconnections between different colors
  - inversion breakpoint visible
  - interaction to check if error in metadata from experiments vs assembly
    - read pair info supports metadata
      - speedup claim vs prev work



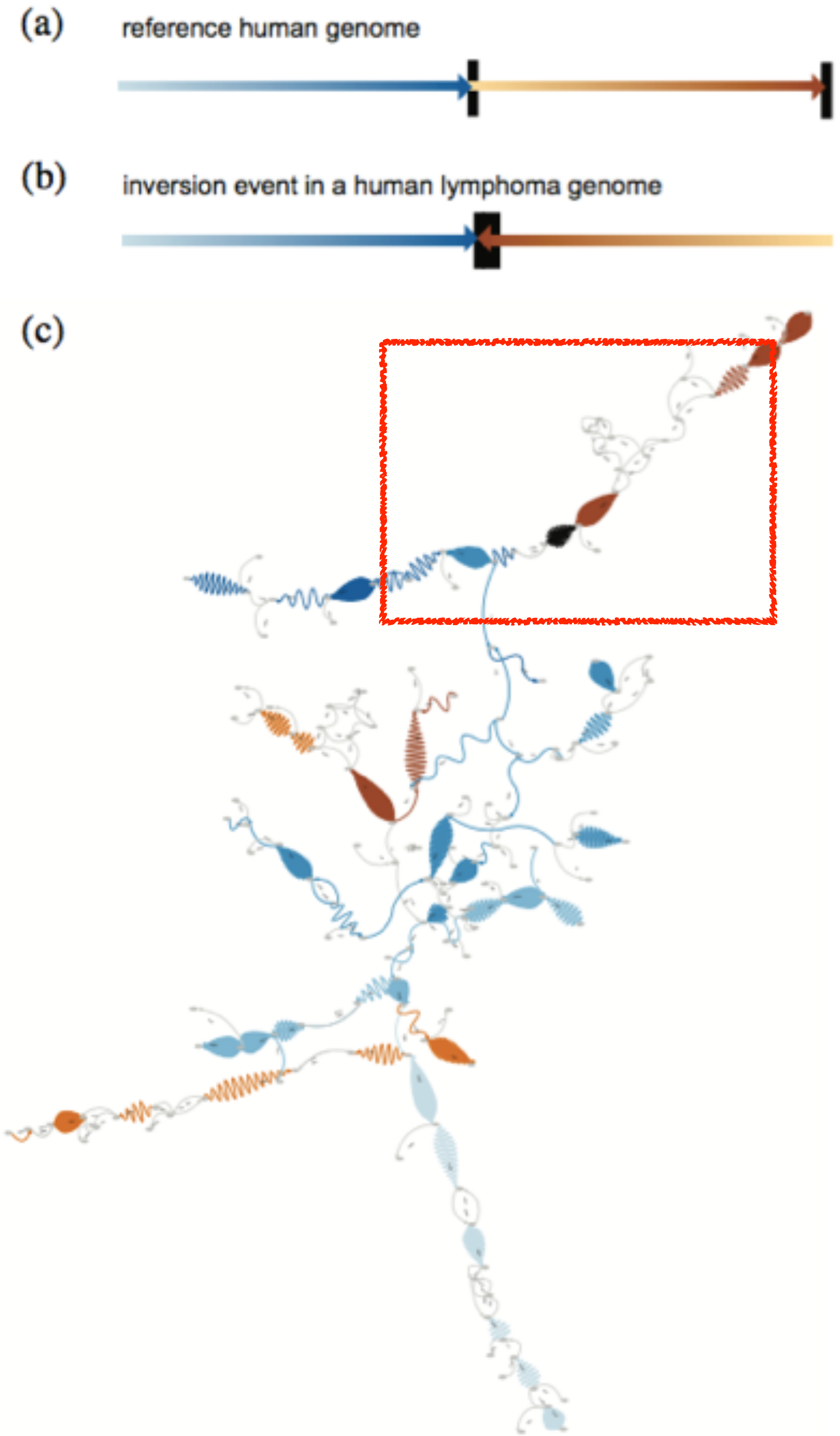
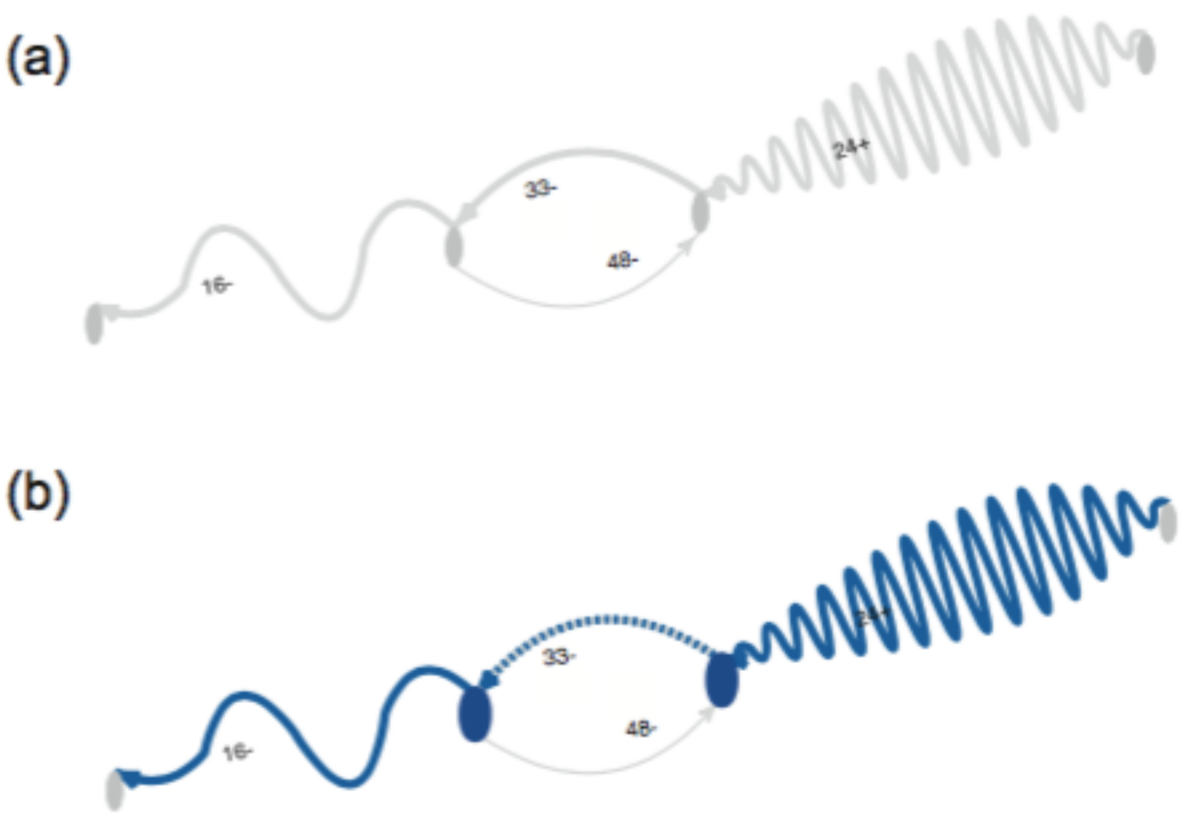


*Fig 10. ABySS-Explorer: visualizing genome sequence assemblies. Nielsen, Jackman, Birol, Jones. TVCG 15(6):881-8, 2009 (Proc. InfoVis 2009).*

# Assembly examples

- ideal: single large contig
  - overview/gist: many small contigs remain
- interaction to resolve
  - integrate paired read highlighting on top of contig paths structure



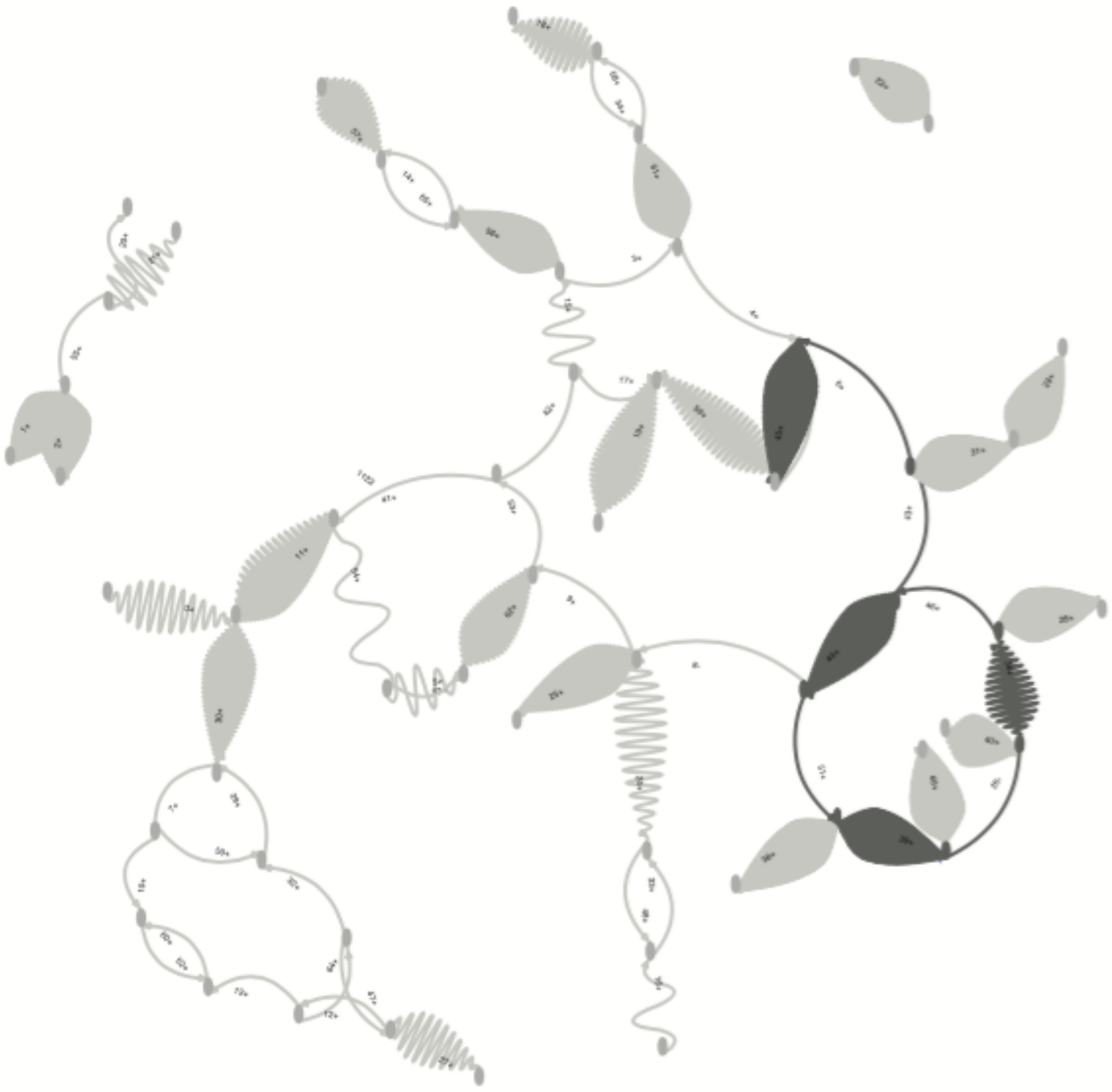


*Fig 7/9. ABySS-Explorer: visualizing genome sequence assemblies. Nielsen, Jackman, Birol, Jones. TVCG 15(6):881-8, 2009 (Proc. InfoVis 2009).*