Part 4: High Dimensional, Graphs and Trees, User Studies

Information Visualization Mini-Course
TECS Week 2008

Tamara Munzner

UBC Computer Science

11 January 2008
Mini-Course Outline

- Part 1: Monday morning
  - Intro
  - Design Studies
  - Models
  - Perception and Memory

- Part 2: Monday afternoon
  - Color
  - Space, Layers, and Ordering
  - Statistical Graphics

- Part 3: Thursday afternoon
  - Multiples and Interaction
  - Navigation and Zooming
  - Focus+Context

- Part 4: Friday morning
  - High Dimensional Data
  - Graphs and Trees
  - User Studies
Parallel Coordinates

- only 2 orthogonal axes in the plane
- instead, use parallel axes!

Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8, \text{ and } -1$. 

PC: Duality

- rotate-translate
- point-line
  - pencil: set of lines coincident at one point

[Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry. Alfred Inselberg and Bernard Dimsdale, IEEE Visualization '90.]
PC: Axis Ordering

- geometric interpretations
  - hyperplane, hypersphere
  - points do have intrinsic order
- infovis
  - no intrinsic order, what to do?
  - indeterminate/arbitrary order
    - weakness of many techniques
    - downside: human-powered search
    - upside: powerful interaction technique
- most implementations
  - user can interactively swap axes
- Automated Multidimensional Detective
  - Inselberg 99
  - machine learning approach
Hierarchical Parallel Coords: LOD

[Hierarchical Parallel Coordinates for Visualizing Large Multivariate Data Sets. Fua, Ward, and Rundensteiner, IEEE Visualization 99.]
Dimensionality Reduction

- mapping multidimensional space into
  - space of fewer dimensions
    - typically 2D for infovis
    - keep/explain as much variance as possible
    - show underlying dataset structure
    - multidimensional scaling (MDS)
  - MDS: minimize differences between interpoint distances in high and low dimensions
Dimensionality Reduction: Isomap

- 4096 D: pixels in image
- 2D: wrist rotation, fingers extension

Spring-Based MDS: Naive

- repeat for all points
  - compute spring force to all other points
  - difference between high dim, low dim distance
  - move to better location using computed forces
- compute distances between all points
  - $O(n^2)$ iteration, $O(n^3)$ algorithm
Faster Spring Model [Chalmers 96]

- compare distances only with a few points
  - maintain small local neighborhood set
Faster Spring Model [Chalmers 96]

- compare distances only with a few points
  - maintain small local neighborhood set
  - each time pick some randoms, swap in if closer
Faster Spring Model [Chalmers 96]

- compare distances only with a few points
  - maintain small local neighborhood set
  - each time pick some randoms, swap in if closer
Faster Spring Model [Chalmers 96]

- compare distances only with a few points
  - maintain small local neighborhood set
  - each time pick some randoms, swap in if closer
- small constant: 6 locals, 3 randoms typical
  - $O(n)$ iteration, $O(n^2)$ algorithm
Parent Finding [Morrison 02, 03]

- lay out a $\sqrt{n}$ subset with [Chalmers 96]
- for all remaining points
  - find "parent": laid-out point closest in high D
  - place point close to this parent
- $O(n^{5/4})$ algorithm
MDS Beyond Points

- galaxies: aggregation

- themescapes: terrain/landscapes

[www.pnl.gov/infoviz/graphics.html]
Cluster Stability

- display
  - also terrain metaphor
- underlying computation
  - energy minimization (springs) vs. MDS
  - weighted edges
- do same clusters form with different random start points?
- "ordination"
  - spatial layout of graph nodes

[Davidson, Wylie, and Boyack. Cluster Stability and the Use of Noise in Interpretation of Clustering. Proc InfoVis 2001.]
Approach

- normalize within each column
- similarity metric
  - discussion: Pearson’s correlation coefficient
- threshold value for marking as similar
  - discussion: finding critical value
Graph Layout

- **criteria**
  - geometric distance matching graph-theoretic distance
    - vertices one hop away close
    - vertices many hops away far
  - insensitive to random starting positions
    - major problem with previous work!
  - tractable computation

- **force-directed placement**
  - discussion: energy minimization
  - others: gradient descent, etc
  - discussion: termination criteria
Barrier Jumping

- same idea as simulated annealing
  - but compute directly
  - just ignore repulsion for fraction of vertices
- solves start position sensitivity problem

[Davidson, Wylie, and Boyack. Cluster Stability and the Use of Noise in Interpretation of Clustering. Proc InfoVis 2001.]
Results

- efficiency
  - naive approach: $O(V^2)$
  - approximate density field: $O(V)$
- good stability
  - rotation/reflection can occur
different random start adding noise

[Davidson, Wylie, and Boyack. Cluster Stability and the Use of Noise in Interpretation of Clustering. Proc InfoVis 2001.]
Critique

- real data
  - suggest check against subsequent publication!
- give criteria, then discuss why solution fits
- visual + numerical results
  - convincing images plus benchmark graphs
- detailed discussion of alternatives at each stage
- specific prescriptive advice in conclusion
HiDim: Readings


HiDim: Further Reading


Mini-Course Outline

► Part 1: Monday morning
  ▶ Intro
  ▶ Design Studies
  ▶ Models
  ▶ Perception and Memory

► Part 2: Monday afternoon
  ▶ Color
  ▶ Space, Layers, and Ordering
  ▶ Statistical Graphics

► Part 3: Thursday afternoon
  ▶ Multiples and Interaction
  ▶ Navigation and Zooming
  ▶ Focus+Context

► Part 4: Friday morning
  ▶ High Dimensional Data
  ▶ Graphs and Trees
  ▶ User Studies
Animated Radial Layouts

▶ from static to dynamic radial layout (video)

Animation

- polar interpolation

- maintain neighbor order (note prefuse bug!)

Treemaps

- containment not connection
  - emphasize node attributes, not topological structure

- difficulties reading
Cushion Treemaps

- show structure with shading
  - scale parameter controls global vs. local

Critique

- good: use shading to free color for other encodings
- good: cushions do help show more internal hierarchical structure
- limitations: fundamental strength is unchanged
  - still best when focus is node attributes not topological structure
Treemap Applications

- cushion treemaps
  - SequoiaView, Windows app
  - hard drive usage
  - http://www.win.tue.nl/sequoiaview/
- one of the infovis tech-transfer success stories
Scaling Up Treemaps: MillionVis

- shading not outline to visually distinguish with less pixels
- more GPU tricks, animation for transitions

[Interactive Information Visualization of a Million Items. Jean-Daniel Fekete and Catherine Plaisant, Proc InfoVis 2002.]
Topological Fisheye Views

- input is laid-out graph
- preprocess: construct multilevel hierarchy by coarsening graphs
- user interactively controls focus point
- show hybrids made from several levels

Topological Fisheye Views

Coarsening Strategy

- must preserve graph-theoretic properties
  - topological distance (hops away), cycles
  - cannot just use geometric proximity alone
  - cannot just contract nodes/edges
  - exploit geometric information with proximity graph

Coarsening Requirements

▶ uniform cluster/metanode size
▶ match coarse and fine layout geometries
▶ scalable

Hybrid Graph

- find active nodes

Distort For Uniform Density

(b) default layout of hybrid graph  (c) distorted layout of hybrid graph

Critique

- topologically sophisticated, not just geometric distortion
- rigorous approach
Graphs: Readings


Graphs: Further Readings


Interactive Visualization of Small World Graphs Frank van Ham and Jarke van Wijk, Proc. InfoVis 2005
Mini-Course Outline

► Part 1: Monday morning
  ▶ Intro
  ▶ Design Studies
  ▶ Models
  ▶ Perception and Memory

► Part 2: Monday afternoon
  ▶ Color
  ▶ Space, Layers, and Ordering
  ▶ Statistical Graphics

► Part 3: Thursday afternoon
  ▶ Multiples and Interaction
  ▶ Navigation and Zooming
  ▶ Focus+Context

► Part 4: Friday morning
  ▶ High Dimensional Data
  ▶ Graphs and Trees
  ▶ User Studies
Perceptual Scalability

- what are perceptual/cognitive limits when screen-space constraints lifted?
  - 2 vs. 32 Mpixel display
  - macro/micro views
- perceptually scalable
  - no increase in task completion times when normalize to amount of data

Perceptual Scalability

- design
  - 2 display sizes, between-subjects
    - (data size also increased proportionally)
  - 3 visualization designs, within
    - small multiples: bars
    - embedded graphs
    - embedded bars
  - 7 tasks, within
  - 42 tasks per participant
    - 3 vis x 7 tasks x 2 trials
Embedded Visualizations

Small Multiples Visualizations

- attribute-centric instead of space-centric

Results

- 20x increase in data, but only 3x increase in absolute task times

Results

- significant 3-way interaction
  - between display, size, task

Results

- visual encoding important on small displays
  - DS: mults sig slower than graphs on small
  - DS: mults sig slower than embedded on large
  - OS: bars sig faster than graphs for small
  - OS: no sig difference bars/graphs for large

- spatial grouping important on large displays
  - embedded sig faster+preferred over small mult
  - no bar/graph differences
Critique

- first study of macro/micro effects
  - breaking new ground

- many possible followups
  - physical navigation vs. virtual navigation
Fisheye Multilevel Networks

[Navigating Hierarchically Clustered Networks through Fisheye and Full-Zoom Methods. Schaffer et al. ACM ToCHI 3(2) p 162-188, 1996.]
Lab Experiment

- 2 interfaces (fisheye, zoom)
- 2 tasks (isomorphic)
  - stages: find and repair
- within subjects, counterbalanced order
- 20 participants
- data: 154 nodes, 39 clusters
- measurements
  - completion time
  - number of zooms
  - success
Results

- sig effect of interface: fisheye faster
- but no differences with find subtask
  - information visible in both displays
- solution quality differed: fisheye better
  - local rerouting difficult in full-zoom
Field Experiment

- 2 real control room operators
- response times similar
  - no statistical analysis, too few subjects
- expressed preference for fisheye over full-zoom
  - (experimenter effect?)
- concerns about fisheye: missing details
Critique

- nicely designed study
- useful discussion of qualitative observations
- very good to do field followup with real operators
Pictures Into Numbers

- field study
- participants: professional meterologists
  - two people: forecaster, technician
- interfaces: multiple programs used
- protocol
  - talkaloud
  - videotaped sessions with 3 cameras

Cognitive Task Analysis

- initialize understanding of large scale weather
- build qualitative mental model (QMM)
- verify and adjust QMM
- write the brief

- task breakdown part of paper contribution
Coding Methodology

- interface
  - which interface used
  - whether picture/chart/graph
- usage (every utterance!)
  - goal
  - extract
    - quant/qual
    - goal-oriented/opportunistic
    - integrated/unintegrated
- brief-writing
  - quant/qual
  - QMM/vis/notes
Results

- sig difference between vis used at CTA stages
  - charts to build QMM
  - images to verify/adjust QMM
  - all kinds during brief-writing
- many others...

Critique

- video coding is huge amount of work, but very illuminating
  - untangling complex story of real tool use
- methodology of CTA construction not discussed here
  - often bottomup/topdown mix
Studies: Readings


Further Readings

Task-Centered User Interface Design, Clayton Lewis and John Rieman, Chapters 0-5.

