Information Visualization at UBC

Tamara Munzner

University of British Columbia

1

Information Visualization

- visual representation of abstract data
  - computer-based
  - interactive
  - goal of helping human perform some task more effectively
- bridging many fields
  - cognitive psych: finding appropriate representation
  - HCI: using task to guide design and evaluation
  - graphics: interacting in realtime
- external representation reduces load on working memory

2

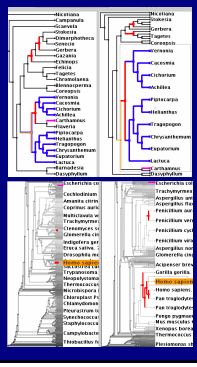
Current Projects

- accordion drawing
  - TreeJuxtaposer, SequenceJuxtaposer, TJC, PRISAD, PowerSetViewer
- evaluation
  - Focus+Context, Transformations
- graph drawing
  - TopoLayout
- dimensionality reduction
  - MDSteer, PBSteer

3

Accordion Drawing

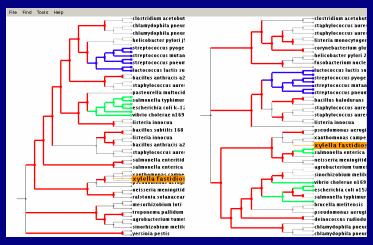
- rubber-sheet navigation
  - stretch out part of surface, the rest squishes
  - borders nailed down
  - Focus+Context technique
    - integrated overview, details
  - old idea
    - [Sarkar et al 93], ...
- guaranteed visibility
  - marks always visible
  - important for scalability
  - new idea
    - [Munzner et al 03]



4

Guaranteed Visibility

- easy with small datasets



5 5

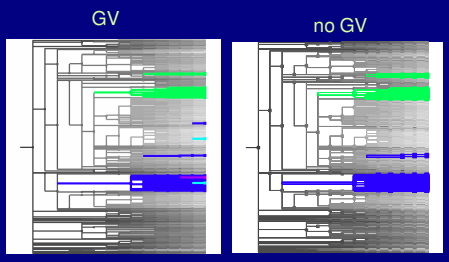
Guaranteed Visibility Challenges

- hard with larger datasets
- reasons a mark could be invisible
  - outside the window
    - AD solution: constrained navigation
  - underneath other marks
    - AD solution: avoid 3D
  - smaller than a pixel
    - AD solution: smart culling

6

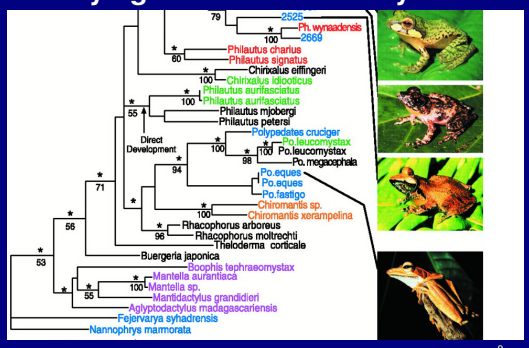
Guaranteed Visibility: Culling

- naive culling may not draw all marked items



7

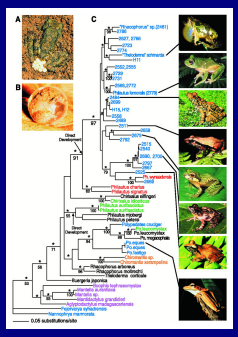
Phylogenetic/Evolutionary Tree



M Meegaskumbura et al., Science 298:379 (2002)

8

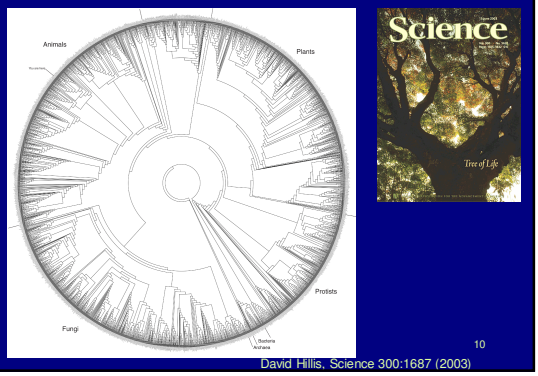
Common Dataset Size Today



M Meegaskumbura et al., Science 298:379 (2002)

9

Future Goal: 10M Node Tree of Life

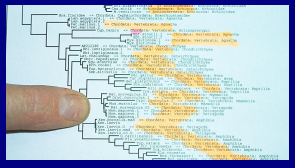


David Hillis, Science 300:1687 (2003)

10

Paper Comparison: Multiple Trees

focus

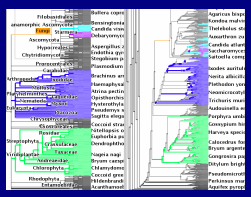


context

11

TreeJuxtaposer

- comparison of evolutionary trees
  - side by side
- [demo: olduvai.sourceforge.net/tj]



12

TJ Contributions

- first interactive tree comparison system
  - automatic structural difference computation
  - guaranteed visibility of marked areas
- scalable to large datasets
  - 250,000 to 500,000 total nodes
  - all preprocessing subquadratic
  - all realtime rendering sublinear
- introduced accordion drawing (AD)
- introduced guaranteed visibility (GV)

13

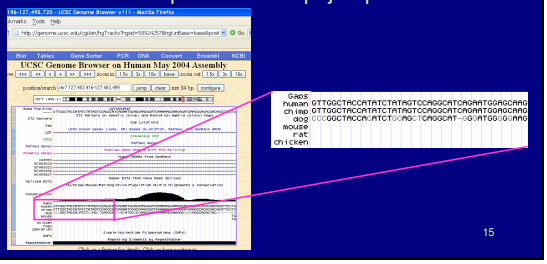
Joint Work: TJ Credits

- Tamara Munzner (UBC prof)
- Francois Guimbretiere (Maryland prof)
- Serdar Tasiran (Koc Univ, prof)
- Li Zhang, Yunhong Zhou (HP Labs)
  - TreeJuxtaposer: Scalable Tree Comparison using Focus+Context with Guaranteed Visibility
  - Proc. SIGGRAPH 2003
  - www.cs.ubc.ca/~tmm/papers/tj
- James Slack (UBC PhD)
- Tamara Munzner (UBC prof)
- Francois Guimbretiere (Maryland prof)
  - TreeJuxtaposer: InfoVis03 Contest Entry. (Overall Winner)
  - InfoVis 2003 Contest
  - www.cs.ubc.ca/~tmm/papers/contest03

14

Genomic Sequences

- multiple aligned sequences of DNA
- now commonly browsed with web apps
  - zoom and pan with abrupt jumps



15

SequenceJuxtaposer

- dense grid, following conventions
  - rows of sequences, typically species
  - columns of partially aligned nucleotides
  - [video: www.cs.ubc.ca/~tmm/papers/sj]

16

SJ Contributions

- accordion drawing for gene sequences
  - smooth, fluid transitions between states
  - guaranteed visibility for globally visible landmarks
  - difference thresholds changeable on the fly
- 2004 paper results: 1.7M nucleotides
  - current with PRISAD: 40M nucleotides
- future work
  - hierarchical structure from annotation dbs
  - editing

17

Joint Work: SJ Credits

- James Slack (UBC PhD)
- Kristian Hildebrand (Weimar Univ MS)
- Tamara Munzner (UBC prof)
- Katherine St. John (CUNY prof)
  - SequenceJuxtaposer: Fluid Navigation For Large-Scale Sequence Comparison In Context
  - Proc. German Conference Bioinformatics 2004
  - www.cs.ubc.ca/~tmm/papers/sj

18

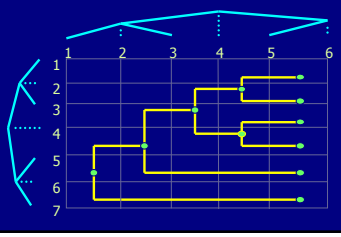
Scaling Up Trees

- TJ limits: 500K nodes
  - large memory footprint
  - CPU-bound, far from achieving peak rendering performance of graphics card
- in TJ, quadtree data structure used for
  - placing nodes during layout
  - drawing edges given navigation
  - culling edges with GV
  - picking edges during interaction

19

New Data Structures, Algorithms

- new data structures
  - two 1D hierarchies vs. one 2D quadtree
- new drawing/culling algorithm



20

TJC/TJC-Q Results

- TJC
  - no quadtree
  - picking with new hardware feature
    - requires HW multiple render target support
  - 15M nodes
- TJC Q
  - lightweight quadtree for picking support
  - 5M nodes
- both support tree browsing only
  - no comparison data structures

21

Joint Work: TJC, TJC-Q Credits

- Dale Beermann (Virginia MS alum)
- Tamara Munzner (UBC prof)
- Greg Humphreys (Virginia prof)

  - Scalable, Robust Visualization of Large Trees
  - Proc. EuroVis 2005
  - www.cs.virginia.edu/~gfx/pubs/TJC

22

PRISAD

- generic accordion drawing infrastructure
  - handles many application types
- efficient
  - guarantees of correctness: no overculling
  - tight bounds on overdrawing
    - handles dense regions efficiently
  - new algorithms for rendering, culling, picking
    - exploit application dataset characteristics instead of requiring expensive additional data structures

23

PRISAD Results

- trees
  - 4M nodes
  - 5x faster rendering, 5x less memory
  - order of magnitude faster for marking
- sequences
  - 40M nucleotides
- power sets
  - 2M to 7M sets
  - alphabets beyond 20,000

24

Joint Work: PRISAD Credits

- James Slack (UBC PhD)
- Kristian Hildebrand (Weimar MS)
- Tamara Munzner (UBC prof)

  – PRISAD: A Partitioned Rendering Infrastructure for Scalable Accordion Drawing.

  – Proc. InfoVis 2005, to appear

25

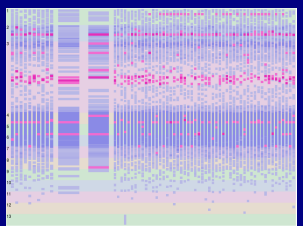
PowerSetViewer

- data mining of market basket transactions
  – show progress of steerable data mining system with constraints
  – want visualization "windshield" to guide parameter setting choices on the fly
- dynamic data
  – all other AD applications had static data
- transactions as sets
  – items bought together make a set
  – alphabet is items in stock at store
  – space of all possible sets is power set

26

PowerSetViewer

- show position of logged sets within enumeration of power set
  – very long 1D linear list
  – wrap around into 2D grid of fixed width
  – [video]



27

Joint Work: PSV Credits

- work in progress

- Tamara Munzner (UBC prof)
- Qiang Kong (UBC MS)
- Raymond Ng (UBC prof)

28

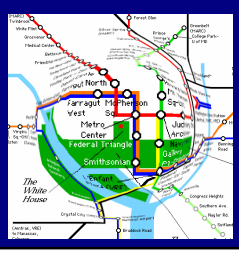
Current Projects

- accordion drawing
  – TreeJuxtaposer, SequenceJuxtaposer, TJC, PRISAD, PowerSetViewer
- Focus+Context evaluation
  – system, perception
- graph drawing
  – TopoLayout
- dimensionality reduction
  – MDSteer, PBSteer

29

Focus+Context

- integrating details and overview into single view
  – carefully chosen nonlinear distortion
  – what are costs? what are benefits?



30

Focus+Context System Evaluation

- how focus and context are used with
  - rubber sheet navigation vs. pan and zoom
  - integrated scene vs. separate overview
- user studies using modified TJ
  - abstract tasks derived from biologists' needs based on interviews

31

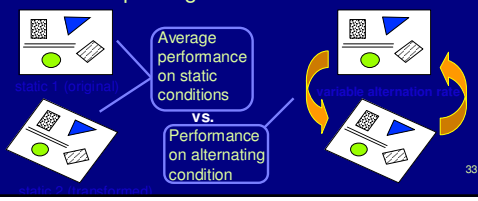
Joint Work: F+C System Eval Credits

- work in progress
- Adam Bodnar (UBC MS)
- Dmitry Nekrasovski (UBC MS)
- Tamara Munzner (UBC prof)
- Joanna McGrenere (UBC prof)
- Francois Guimbretiere (Maryland prof)

32

F+C Perception Evaluation

- understand perceptual costs of transformation
  - find best transformation to use
- visual search for target amidst distractors
  - shaker paradigm



33

F+C Perception Evaluation

- understand perceptual costs of transformation
  - deterioration in performance
    - time, effort, error
  - static costs: caused by crowding, distortion of static transformation itself
    - high static cost
  - dynamic costs: reorienting and remapping when transformation applied or focus moved
    - low dynamic cost
    - large no-cost zone

34

Joint Work: F+C Perceptual Eval

- Keith Lau (former UBC undergrad)
- Ron Rensink (UBC prof)
- Tamara Munzner (UBC prof)
  - Perceptual Invariance of Nonlinear Focus+Context Transformations
  - Proc. First Symposium on Applied Perception in Graphics and Visualization, 2004
- work in progress: continue investigation
- Heidi Lam (UBC PhD)
- Ron Rensink (UBC prof)
- Tamara Munzner (UBC prof)

35

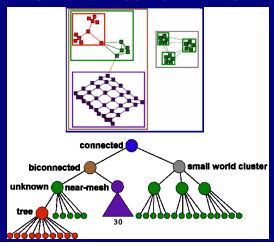
Current Projects

- accordion drawing
  - TreeJuxtaposer, SequenceJuxtaposer, TJC, PRISAD, PowerSetViewer
- Focus+Context evaluation
  - system, perception
- graph drawing
  - TopoLayout
- dimensionality reduction
  - MDSteer, PBSteer

36

TopoLayout

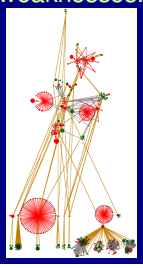
- multilevel decomposition and layout
  - automatic detection of topological features
- chop into hierarchy of manageable pieces
  - lay out using feature-appropriate algorithms

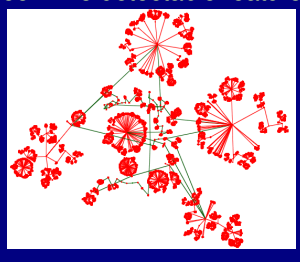


37

Multilevel Hierarchies

- strengths: handles large class of graphs
  - previous work mostly good with near-meshes
- weaknesses: poor if no detectable features





Joint Work: TopoLayout Credits

- work in progress

- Dan Archambault (UBC PhD)
- Tamara Munzner (UBC prof)
- David Auber (Bordeaux prof)

39

Current Projects

- accordion drawing
  - TreeJuxtaposer, SequenceJuxtaposer, TJC, PRISAD, PowerSetViewer
- Focus+Context evaluation
  - system, perception
- graph drawing
  - TopoLayout
- dimensionality reduction
  - MDSteer, PBSteer

40

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)
  - minimize differences between interpoint distances in high and low dimensions

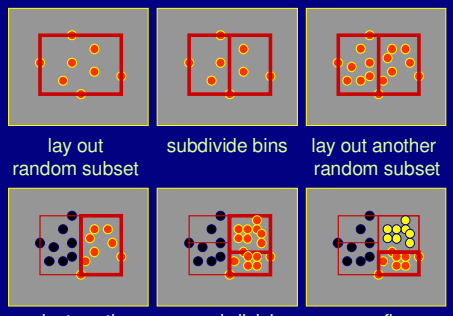
41

Scalability Limitations

- high cardinality and high dimensionality: slow
  - motivating dataset: 120K points, 300 dimensions
  - most existing software could not handle at all
  - 2 hours to compute with $O(n^{5/4})$ HIVE [Ross 03]
- real-world need: exploring huge datasets
  - people want tools for millions of points
- strategy
  - start interactive exploration immediately
    - progressive layout
  - concentrate computational resources in interesting areas
    - steerability
  - often partial layout is adequate for task

42

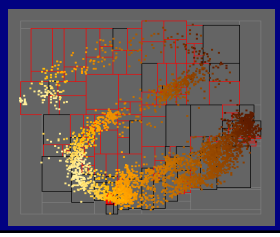
MDSteer Overview



43

MDSteer Contributions

- first steerable MDS algorithm
  - progressive layout allows immediate exploration
  - allocate computational resources in lowD space
  - [video: www.cs.ubc.ca/~tmm/papers/mdsteer]



44

Joint Work: MDSteer Credits

- Matt Williams (former UBC MS)
- Tamara Munzner (UBC prof)
  - Steerable Progressive Multidimensional Scaling
  - Proc. InfoVis 2004
  - www.cs.ubc.ca/~tmm/papers/mdsteer
- work in progress: PBSteer for progressive binning
  - David Westrom (former UBC undergrad)
  - Tamara Munzner (UBC prof)
  - Melanie Tory (UBC postdoc)

45

Summary

- broad array of infovis projects at UBC
- theme: scalability
  - size of dataset
  - number of available pixels

46

InfoVis Service

- IEEE Symposium on Information Visualization (InfoVis) Papers/Program Co Chair 2003, 2004
- IEEE Executive Committee, Technical Committee on Visualization and Graphics
- Visualization Research Challenges
  - report commissioned by NSF/NIH

47

More Information

- papers, videos, images
  - www.cs.ubc.ca/~tmm
- free software
  - olduvai.sourceforge.net/tj
  - olduvai.sourceforge.net/sj

48