Visualization and Biology: Fertile Ground for Collaboration

Tamara Munzner
Department of Computer Science
University of British Columbia

May 2010

http://www.cs.ubc.ca/~tmm/talks.html#bigdata10

Why do visualization?
- pictures help us think
  - substitute perception for cognition
  - external memory: free up limited cognitive/memory resources for higher-level problems

What does visualization allow?
- discovery vs. confirmation
  - discovering new things
  - hypothesis discovery, "eureka moment"
  - confirming conjectured things
- hypothesis confirmation
- contradicting conjectured things
- especially (inevitably?) data cleansing

When should we bother doing vis?
- need a human in the loop
  - augment, not replace, human cognition
- simple summary not adequate
  - statistics may not adequately characterize complexity of dataset distribution
- moderate summary may be adequate

Good driving problems for vis research
- need for humans in the loop
- big data
- reasonably clear questions
- many areas of science are a great match
  - biology particularly appealing

Cerebral

collaboration with researchers at UBC Hancock Lab studying innate immunity

Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context
Aaron Barsky, Computer Science, UBC
Tamara Munzner, Computer Science, UBC
Jonathan Gardner, Microbiology and Immunology, UBC
May 2008
open-source software download (Gnuplot plugin)
http://www.pathogenomics.ca/cerebral/
deployed in InnateDB (mammalian innate immunity database)
http://www.innatedb.ca

Model - Experiment cycle
- conduct experiments on cells
- interpret results in current graph model
- propose modifications to refine model
- vis tool to accelerate workflow?

Model summarizes extensive lab work
- system model
  - interaction graph \( G = (V, E) \)
  - meta-data for each \( v \) in \( V \)
    - labels, biological attributes
  - experimental measurements
    - multiple floats for each \( v \)
    - microarray data

Human interactome: \( E=50,000, V\sim10,000 \)
- too complex, beyond scope of tool

Immune system: \( E=1263, V=760 \)

Encoding and interaction design decisions
- create custom graph layout
  - guided by biological metadata
- use small multiple views
  - one view per experimental condition
- show measured data in graph context
  - not in isolation

TLR4 biomolecule: \( E=74, V=54 \)
- very local view

Systems biology model
- graph \( G = (V, E) \)
  - \( V \): proteins, genes, DNA, RNA, tRNA, etc.
  - \( E \): interacting molecules

Choice 1: Create custom graph layout
- graph layout heavily studied
  - given graph \( G = (V, E) \)
  - create layout in 2D/3D plane
  - hundreds of papers
  - annual Graph Drawing conference
- Force-directed
  (Fruchterman and Reingold, 1991)
- Hierarchical
  (Sugiyama 1981)

What does visualization allow?
- discovery vs. speedup
  - novel capabilities
  - tool supports fundamentally new operations

Goal: Integrate model with measurements
- system model
  - interaction graph \( G = (V, E) \)
  - meta-data for each \( v \) in \( V \)
    - labels, biological attributes
  - experimental measurements
    - multiple floats for each \( v \) in \( V \)
    - microarray data

Why do visualization?
- pictures help us think
  - substitute perception for cognition
  - external memory: free up limited cognitive/memory resources for higher-level problems

When should we bother doing vis?
- need a human in the loop
  - augment, not replace, human cognition
- simple summary not adequate
  - statistics may not adequately characterize complexity of dataset distribution
  - moderate summary may be adequate
Existing layouts did not suit immunologists
- graph drawing goals
  - visualize graph structure
- biologist goals
  - visualize biological knowledge
  - some relationships happen to form a graph
  - cell location also relevant

Biological cells divided by membranes
- interactions generally occur within a compartment
- interaction location often known as part of model

Hand-drawn diagrams
- cellular location spatially encoded vertically
- inflexible to create by hand in era of big data

Cerebral layout using biological metadata
- similar to hand-drawn
- spatial position reveals location in cell
- simulated annealing in \( O(E^{\alpha} V) \) vs \( O(V^\beta) \) time

Choice 2: Use small multiple views
- one graph instance per experimental condition
- same spatial layout
- color differently, by condition

Choice 3: Show measurements and graph
- why not measurements alone?
  - data driven hypothesis: gene expression clusters indicate similar function in cell?
  - clusters are often trustworthy artifacts!
  - noisy data: different clustering alg. different results
  - measured data alone potentially misleading
  - show in context of graph model

Choice 2: Use small multiple views
- one graph instance per experimental condition
- same spatial layout
- color differently, by condition

Choice 3: Show measurements and graph
- why not measurements alone?
  - data driven hypothesis: gene expression clusters indicate similar function in cell?
  - clusters are often trustworthy artifacts!
  - noisy data: different clustering alg. different results
  - measured data alone potentially misleading
  - show in context of graph model

Why not animation?
- global comparison difficult

Why not animation?
- limits of human visual memory
  - compared to side by side visual comparison

Why not animation?
- zooming versus multiple window interfaces: cognitive costs of visual comparisons

Why not animation?
- limits of human visual memory
  - compared to side by side visual comparison

Why not animation?
- limits of human visual memory
  - compared to side by side visual comparison

Why not animation?
- limits of human visual memory
  - compared to side by side visual comparison

InnateDB links to Cerebral
- InnateDB: facilitating systems-level analyses of the mammalian innate immune response

Pathline: comparative functional genomics
- collab with Broad Regev lab: yeast regulatory networks
- Pathline: A Tool for Comparative Functional Genomics

Pathline: comparative functional genomics
- collab with Broad Regev lab: yeast regulatory networks
- Pathline: A Tool for Comparative Functional Genomics

Pathline: comparative functional genomics
- collab with Broad Regev lab: yeast regulatory networks
- Pathline: A Tool for Comparative Functional Genomics

Pathline: comparative functional genomics
- collab with Broad Regev lab: yeast regulatory networks
- Pathline: A Tool for Comparative Functional Genomics

Pathline: comparative functional genomics
- collab with Broad Regev lab: yeast regulatory networks
- Pathline: A Tool for Comparative Functional Genomics
Vast opportunities

• young field, still much to be done

• think about your current workflow
  – what could you speed up by swapping in perception for cognition?
  – exploit the familiar, yet consider breadth of design alternatives

• finding some friendly vis collaborators
  – IEEE VisWeek 2010 (Vis, InfoVis, VAST)
    Oct 11-16, Salt Lake City
    http://vis.computer.org/VisWeek2010
  – EuroVis 2010, Jun 9-11, Bordeaux France

More information

• this talk
  http://www.cs.ubc.ca/~tmm/talks.html#bigdata10

• papers, talks, videos...
  http://www.cs.ubc.ca/~tmm

• visualization intro book chapter
  http://www.cs.ubc.ca/~tmm/papers.html#akpchapter