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

What does visualization allow?
- discovery vs. confirmation
  - discovering new things
  - hypothesis discovery, "aurelia moment"
  - confirming conjectured things
    - hypothesis confirmation
    - contradicting conjectured things
      - especially (inevitably?) data cleansing
- discovery vs. speedup
  - novel capabilities
    - tool supports fundamentally new operations
    - speedup
      - tool accelerates workflow (most common!)

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 video
- collaboration with researchers at UBC Hancock Lab studying innate immunity
- Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context
  - Annapoorna, Computer Science, UBC
  - Tamara Munzner, Computer Science, UBC
  - Jennifer Gillies, Microbiology and Immunology, UBC

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

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

Model summarizes extensive lab work
- graphs come from hand-curated databases
  - dynamic, change with each new publication
- each edge has provenance from experimental evidence
  - TIRAP: an adapter molecule in the Toll signaling pathway
  - Mal (MyD88-adapter-like) is required for Toll-like pathway
  - gene isolation
  - validation and characterization of the Toll-signaling pathway
  - Antenna-A, Antenna-B, Alum, B, Anima A, Argus A, Adult, B, Nerve A
  - pathology B, Adult, B, Bone A, leaf A, Nerve A

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

Immune system: E=1263, V=760
- bigger picture, target size for Cerebral

Human interactome: E~50,000, V~10,000
- too complex, beyond scope of tool

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

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

Anscombe’s quartet: same
- mean
- variance
- correlation coefficient
- linear regression line

Human interactome: E≈50,000, V≈10,000
- too complex, beyond scope of tool
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 $O(E^{3/2})$ vs. $O(V^3)$ time

Choice 2: Use small multiple views

- one graph instance per experimental condition
  - same spatial layout
  - color differently, by condition
- why not measurements alone?
  - data driven hypothesis: gene expression clusters indicate similar function in cell?
  - clusters are often untrustworthy artifacts!
    - noisy data: different clustering alg. → different results
    - measured data alone potentially misleading
    - show in context of graph model

Why not animation?

- limits of human visual memory
  - compared to side by side visual comparison
- global comparison difficult

Choice 3: Show measurements and graph

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

Why not animation?

- limits of human visual memory
  - compared to side by side visual comparison
- global comparison difficult

Adoption by biologists


InnateDB links to Cerebral


Data cleansing example

- incorrect edge across many compartments
  - in well studied dataset
  - not obvious with other layouts

Cerebral summary

- supports interactive exploration of multiple experimental conditions in graph context
- provides familiar representation by using biological metadata to guide graph layout

More information

- this talk http://www.cs.ubc.ca/~tmm/talks.html#amw09
- papers, videos http://www.cs.ubc.ca/~tmm

Adoption by biologists


InnateDB links to Cerebral