

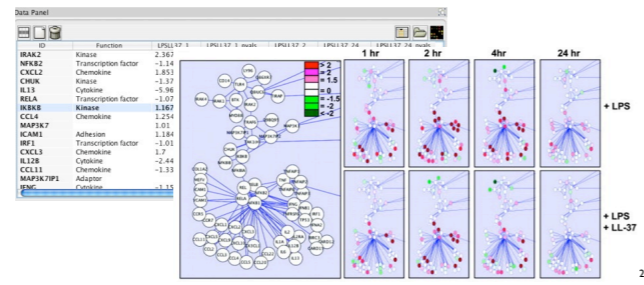
Biology is Destiny: Of Graphs and Genes

Tamara Munzner
University of British Columbia

UBC CS TechTrek 2012
21 Jan 2012

Why do visualization?

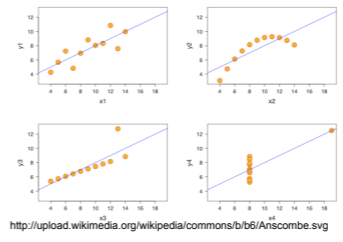
- pictures help us think
 - substitute perception for cognition
 - external memory: free up limited cognitive/memory resources for higher-level problems



When should we bother doing vis?

- need a human in the loop
 - augment, not replace, human cognition
 - for problems that cannot be (completely) automated
- simple summary not adequate
 - statistics may not adequately characterize complexity of dataset distribution

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



Cerebral

Comparing Multiple Experimental Conditions Within Biologically Meaningful Network Context

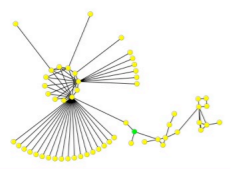
joint work with:
Aaron Barsky, Jennifer Gardy, Robert Kincaid

<http://www.pathogenomics.ca/cerebral/>

Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, Kincaid. IEEE InfoVis 2008.

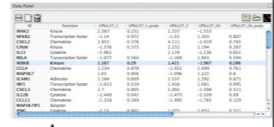
Systems biology model

- graph $G = \{V, E\}$
 - V: proteins, genes, DNA, RNA, tRNA, etc.
 - metadata: labels, biological attributes
- E: interacting molecules
 - known from previous research



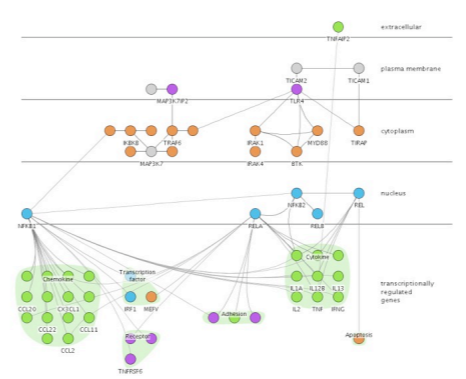
Cycle: model - experiment

- conduct experiments on cells
 - microarrays
 - measurements for each vertex in graph
- interpret results in current graph model
- propose modifications to refine model
- vis tool to accelerate workflow
 - integrated tool to see graph and measurements together
 - choose scope for problem complexity



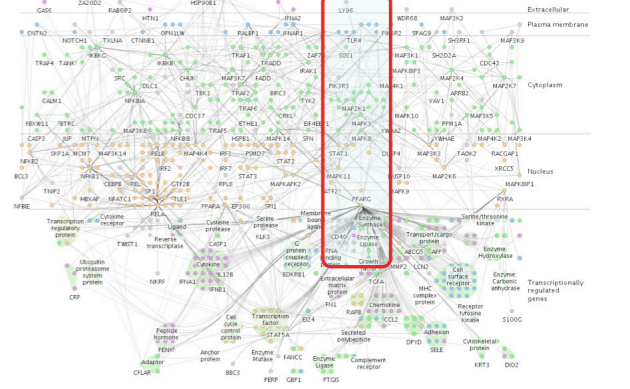
TLR4 biomolecule: E=74, V=54

- very local view



Immune system: E=1263, V=760

- bigger picture, target size for Cerebral



Cerebral video

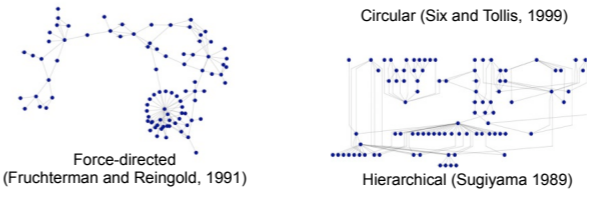


Encoding and interaction design decisions

- 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: 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 conf.

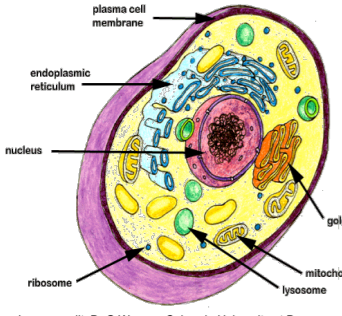


Existing layouts did not suit immunologists

- graph drawing goals
 - visualize graph structure
- biologist goals
 - visualize biological knowledge
 - some relationships happen to form a graph
 - cell location also relevant

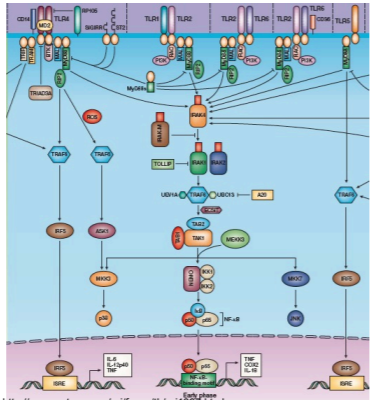
Biological cells divided by membranes

- interactions generally occur within a compartment
- interaction location often known as part of model



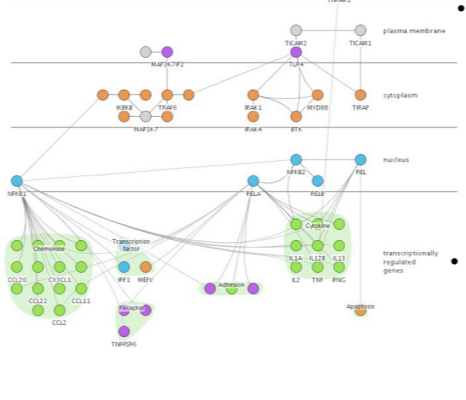
Hand-drawn diagrams

- cellular location spatially encoded vertically
- infeasible to create by hand in era of big data



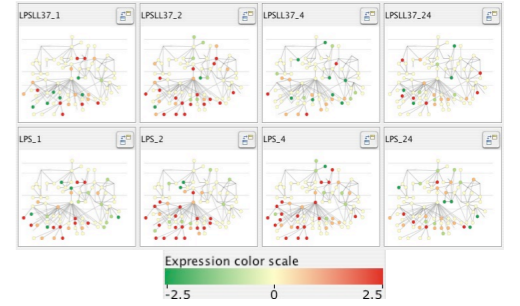
Lay out using biological metadata

- similar to hand-drawn: spatial position reveals location in cell
- simulated annealing in $O(E \cdot V)$ vs. $O(V^3)$ time



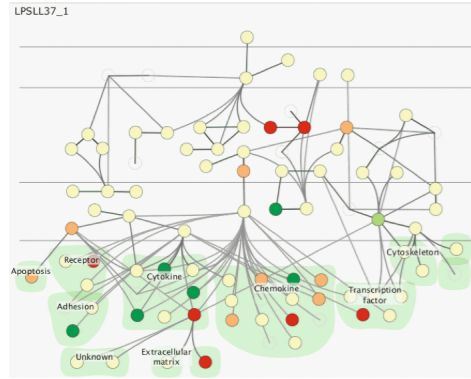
Choice 2: Use small multiple views

- one graph instance per experimental condition
 - same spatial layout
 - color differently by condition



Why not animation?

- global comparison difficult



17

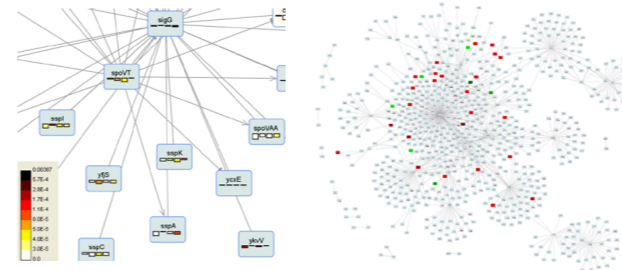
Why not animation?

- limits of human visual memory
 - compared to side by side visual comparison
- Zooming versus multiple window interfaces: Cognitive costs of visual comparisons. Matthew Plumlee and Colin Ware. *ACM Trans. Computer-Human Interaction (ToCHI)*, 13(2):179-209, 2006.
- Animation: can it facilitate? Barbara Tversky, Julie Bauer Morrison, and Mireille Betancourt. *International Journal of Human-Computer Studies*, 57(4):247-262, 2002.
- Effectiveness of Animation in Trend Visualization. George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, John Stasko. *IEEE Trans. Visualization and Computer Graphics* 14(6):1325-1332 (Proc. InfoVis 08), 2008.

18

Why not glyphs?

- embed multiple conditions as a chart inside node
- clearly visible when zoomed in
- but cannot see from global view
 - only one value shown in overview

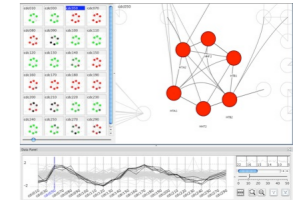


[M. A. Westenberg, S. A. F. T. van Hijum, O. P. Kuipers, J. B. T. M. Roerdink. Visualizing Genome Expression and Regulatory Network Dynamics in Genomic and Metabolic Context. *Computer Graphics Forum*. 27(3):887-894, 2008.]

19

Choice: Show measures 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

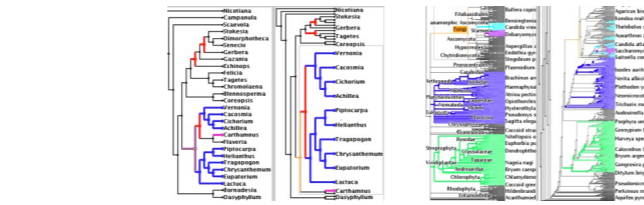


20

Contributions

- Cerebral
 - supports interactive exploration of multiple experimental conditions in graph context
 - provides familiar representation by using biological metadata to guide graph layout
- tool deployment
 - open source, Cytoscape plugin
 - used by target group of collaborators
 - showcased in <http://innatedb.ca>
 - many more independent adopters
 - 12+ bio lit citations with Cerebral diagrams so far

21



TreeJuxtaposer

Scalable Phylogenetic Tree Comparison

joint work with:
François Guimbretière, Serdar Tasiran, Li Zhang, Yunhong Zhou

<http://olduvai.sf.net/tj>

TreeJuxtaposer: Scalable Tree Comparison using Focus+Context with Guaranteed Visibility. Munzner, Guimbretière, Tasiran, Zhang, Zhou. *ACM SIGGRAPH 2003*.

22

MizBee

A Browser for Comparative Genomics Data

joint work with:
Miriah Meyer, Hanspeter Pfister

<http://www.mizbee.org>

MizBee: A Multiscale Synteny Browser. Meyer, Munzner, Pfister, *IEEE InfoVis 2009*.

23

Pathline

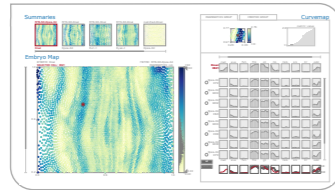
A Tool for Comparative Functional Genomics Data

joint work with:
Miriah Meyer, Bang Wong, Mark Styczynski, Hanspeter Pfister

<http://www.pathline.org>

Pathline: A Tool for Comparative Functional Genomics. Meyer, Wong, Styczynski, Munzner, Pfister, *IEEE Eurographics EuroVis 2010*.

24



MulteeSum

A Tool for Exploring Space-Time Expression Data

joint work with:
Miriah Meyer, Angela DePace, Hanspeter Pfister

<http://www.multeesum.org>

MulteeSum: A Tool for Comparative Spatial and Temporal Gene Expression Data. Meyer, Munzner, DePace, Pfister. *IEEE InfoVis 2010*.

25

More information

- principles in more depth: vis intro book chapter
<http://www.cs.ubc.ca/~tmm/papers.html#akpchapter>
- papers, talks, videos, courses
<http://www.cs.ubc.ca/~tmm>
- this talk
<http://www.cs.ubc.ca/~tmm/talks.html#techtrek12>

26