

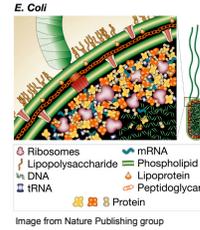
Big Data, Visualization, and Systems Biology

Tamara Munzner
University of British Columbia
Department of Computer Science

Big data and models

- Does big data imply all-data-no-model future for science?
 - No!!
 - Typical Wired hype...
- Big data leads to better models
- Example from system biology and visualization
 - Conduct experiments on cells
 - Interpret results in current model
 - Propose modifications to the model

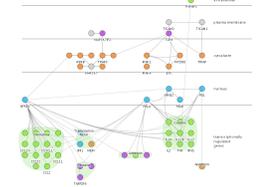
Biomolecular interactions are selective



- Cell densely packed with biomolecules
- Interactions rare
- Model interactions as a graph

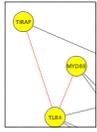
Systems biology model

- Graph $G = \{V, E\}$
 - V : proteins, genes, DNA, RNA, tRNA, etc.
 - E : interacting molecules

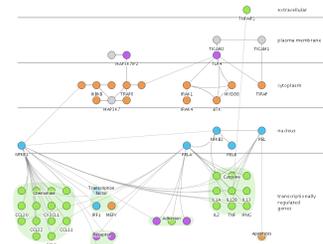


Model summarizes extensive lab work

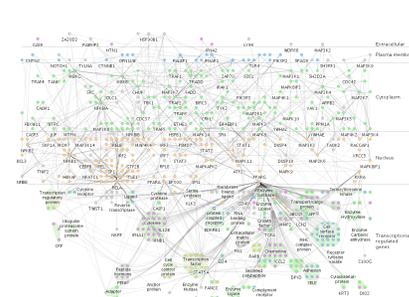
- Graphs come from hand-curated databases
 - Dynamic, change with each new publication
- Each edge has provenance from experimental evidence
 - TIRAP: an adaptor molecule in the Toll signaling pathway. *Hong T, Barton GM, Medzhitov R.*
 - Mal (MyD88-adaptor-like) is required for Toll-like receptor-4 signal transduction. *Fitzgerald KA, Palsson-McDermott EM, Bowie AG, Jefferies CA, Mansell AS, Brady G, Brint E, Dunne A, Gray P, Harte MT, McMurray D, Smith DE, Sims JE, Bird TA, O'Neill LA.*
- Choose scope to manage complexity



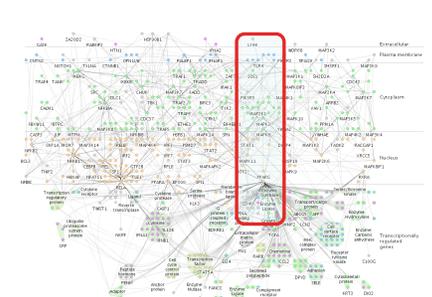
TLR4 biomolecule: E=74, V=54



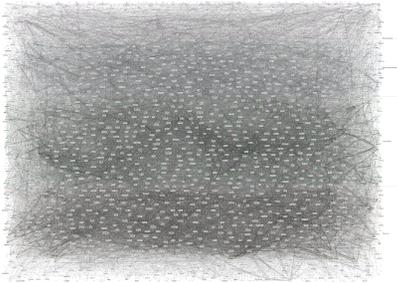
Immune system: E=1263, V=760



Immune system: E=1263, V=760

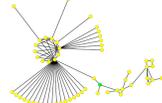


Human interactome: E~50,000, V~10,000



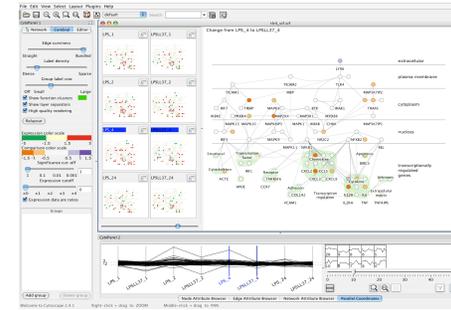
Goal: Overlay measurements on model

- Integrate
 - System model (graph)
 - Experimental measurements



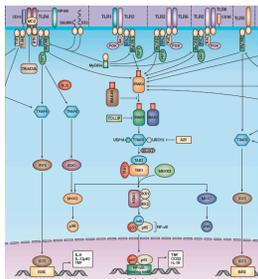
Gene	Protein	LP1017.1	LP1017.1_1	LP1017.2	LP1017.3
IMX2	Protein	2.387	0.211	1.137	-0.113
IMX2	Transcription factor	-1.14	0.272	-1.01	1.181
CE22	Chaperone	1.851	0.176	4.111	1.029
CE26	Chaperone	-1.178	0.177	2.212	1.194
CE28	Chaperone	1.981	0.181	2.138	2.028
IMX1	Transcription factor	1.137	0.166	1.09	1.01
IMX3	Chaperone	1.187	0.20	1.421	1.387
IMX4	Chaperone	1.174	0.176	1.192	1.192
IMX5	Chaperone	1.171	0.176	1.192	1.192
IMX6	Chaperone	1.174	0.176	1.192	1.192
IMX7	Chaperone	1.174	0.176	1.192	1.192
IMX8	Chaperone	1.174	0.176	1.192	1.192
IMX9	Chaperone	1.174	0.176	1.192	1.192
IMX10	Chaperone	1.174	0.176	1.192	1.192
IMX11	Chaperone	1.174	0.176	1.192	1.192
IMX12	Chaperone	1.174	0.176	1.192	1.192
IMX13	Chaperone	1.174	0.176	1.192	1.192
IMX14	Chaperone	1.174	0.176	1.192	1.192
IMX15	Chaperone	1.174	0.176	1.192	1.192
IMX16	Chaperone	1.174	0.176	1.192	1.192
IMX17	Chaperone	1.174	0.176	1.192	1.192
IMX18	Chaperone	1.174	0.176	1.192	1.192
IMX19	Chaperone	1.174	0.176	1.192	1.192
IMX20	Chaperone	1.174	0.176	1.192	1.192

Cerebral



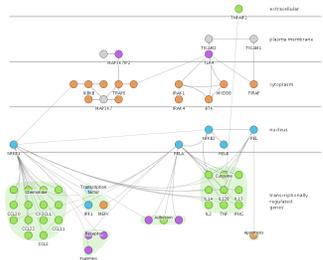
Video

Hand-drawn diagrams



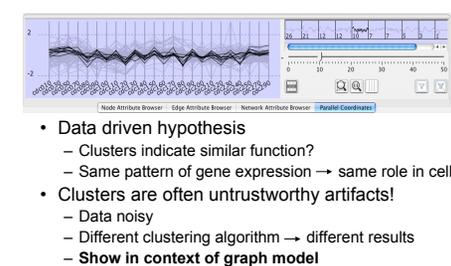
- Cellular location encoded spatially
- Infeasible to create by hand in era of big data

Cerebral layout using biological metadata



- Similar to hand-drawn
- Spatial position reveals location in cell
- Simulated annealing in $O(E\sqrt{V})$ vs. $O(V^3)$ time

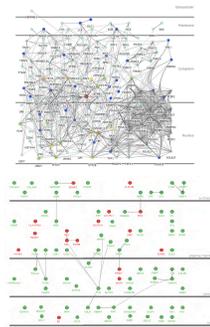
Measurement data alone insufficient



- Data driven hypothesis
 - Clusters indicate similar function?
 - Same pattern of gene expression \rightarrow same role in cell?
- Clusters are often untrustworthy artifacts!
 - Data noisy
 - Different clustering algorithm \rightarrow different results
 - Show in context of graph model

Adoption by biologists

- Matthew D Dyer, T. M Murali, and Bruno W Sobral. The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathogens*, 4(2):e32, 2008.
- Liqun He et al. The glomerular transcriptome and a predicted protein-protein interaction network. *Journal of the American Society of Nephrology*, 19(2):260-268, 2008.



More information

Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context

Aaron Barsky, Computer Science, UBC

Tamara Munzner, Computer Science, UBC

Jennifer Gardy, Microbiology and Immunology, UBC

Robert Kincaid, Agilent Technologies

IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6) (Nov-Dec) 2008, p 1253-1260.

<http://www.cs.ubc.ca/labs/imager/tr/2008/cerebral/>

<http://www.cs.ubc.ca/labs/imager/th/2008/BarskyMscThesis/>

open-source software download

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

deployed in InnateDB (mammalian innate immunity database)

<http://www.innatedb.ca>