

Graph Theory in Computational Biology

Nathan Lazar

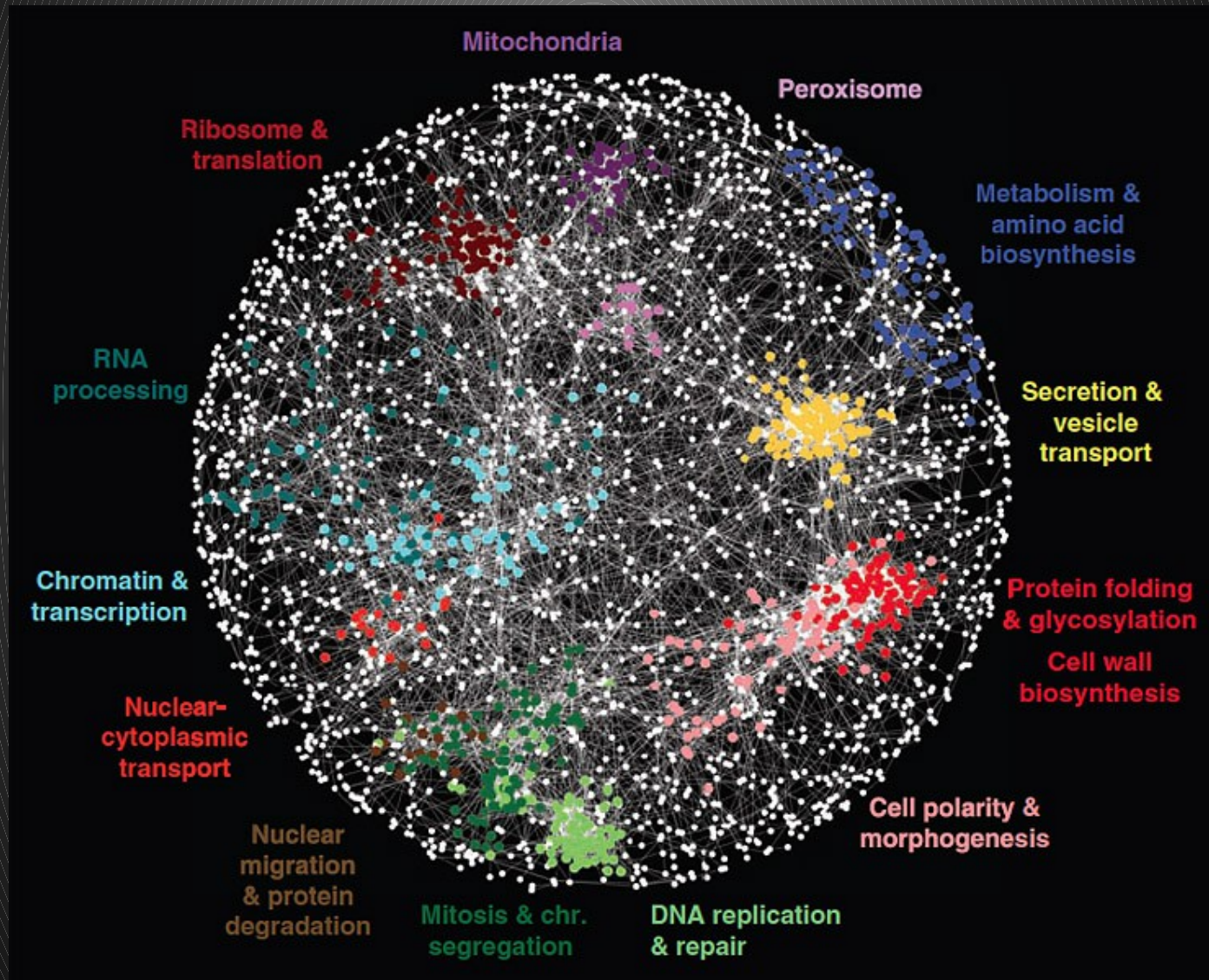
B.S. Mathematics - PSU

PhD candidate - OHSU

Overview

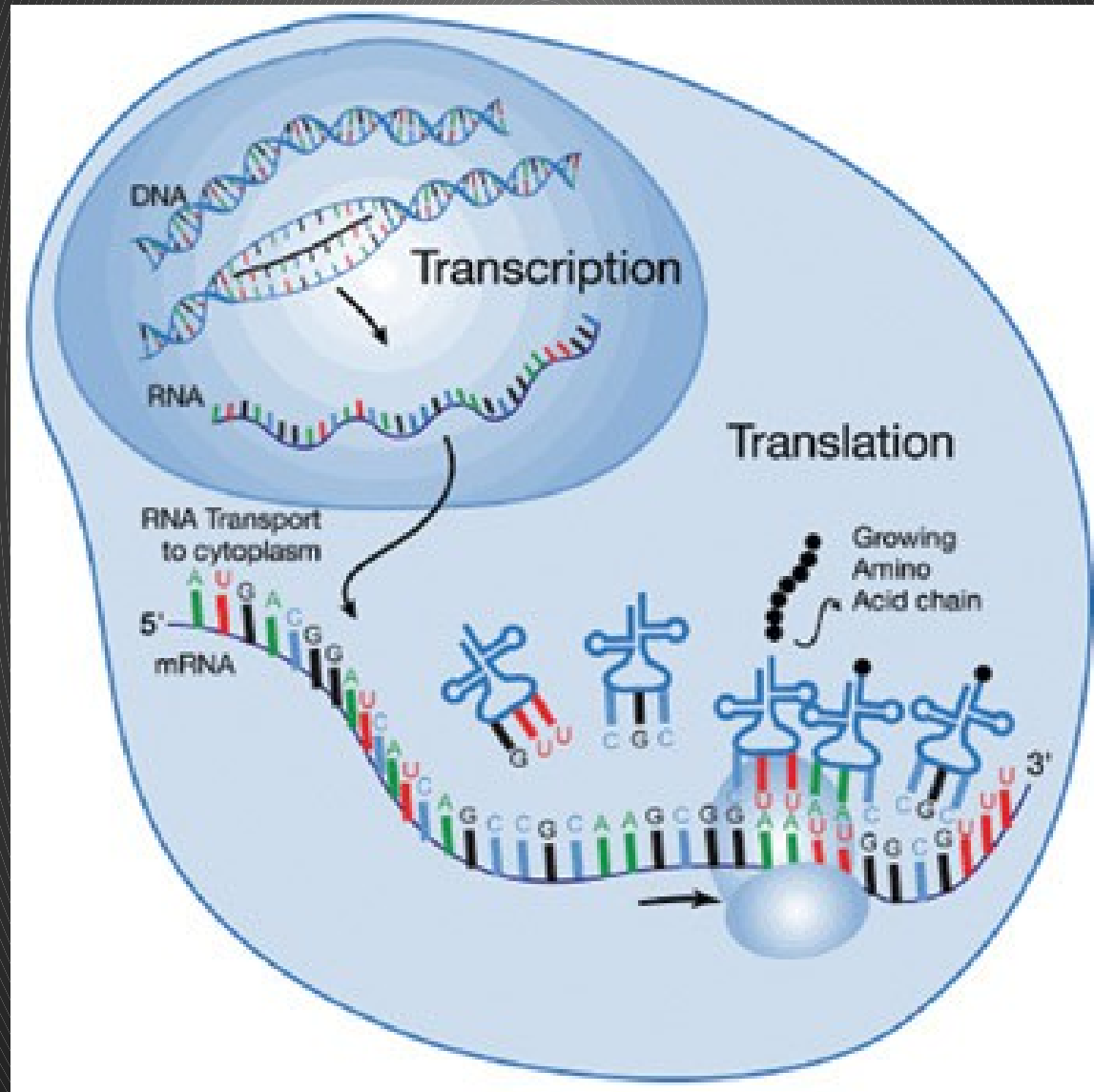
- Gene interaction networks
 - Biology
 - Modeling
 - Inferring
 - Measuring
 - Visualization
- Other networks
- Open questions

Gene interaction networks

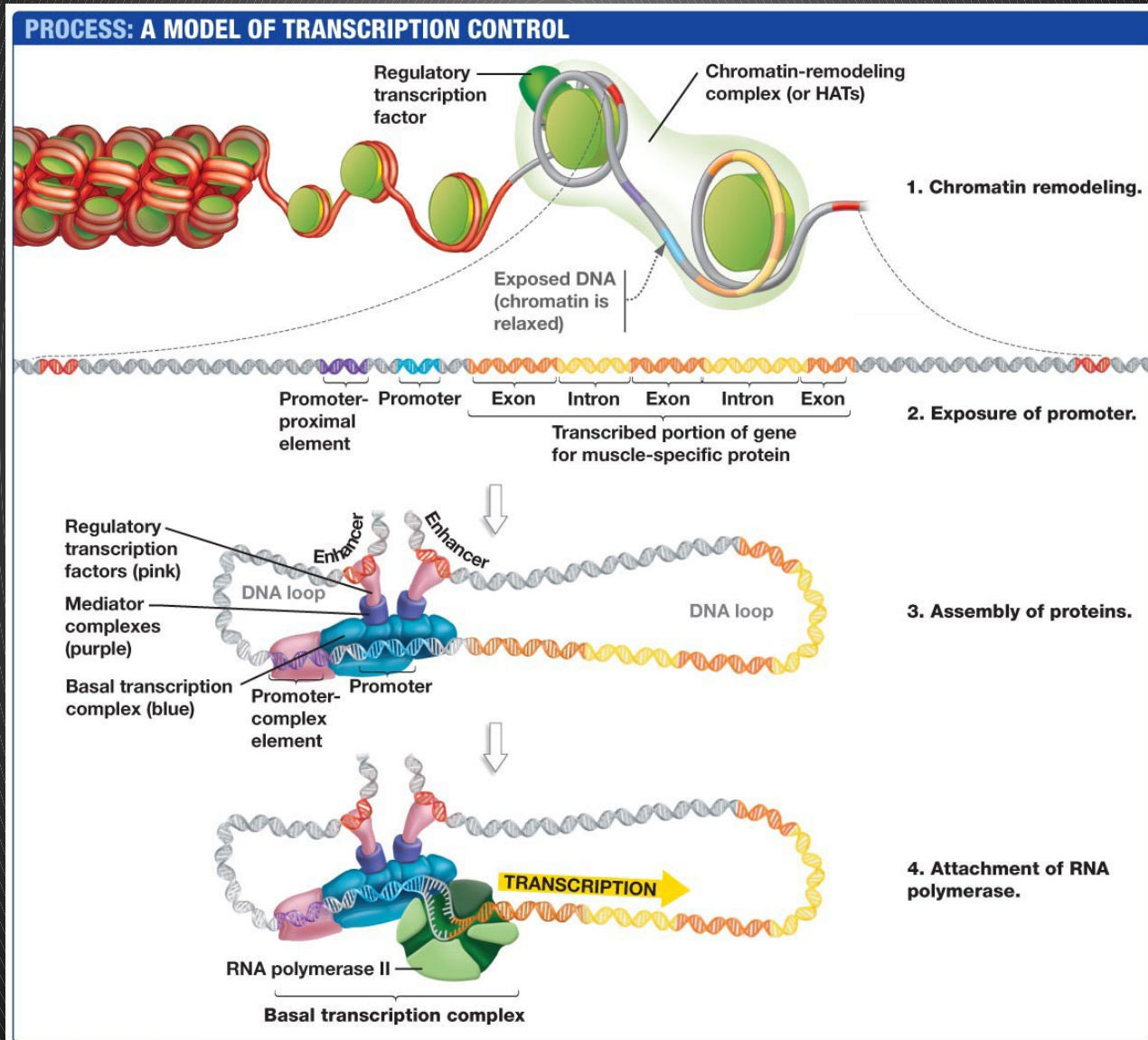


Costanzo et al, 2010, Science 327:425–431.

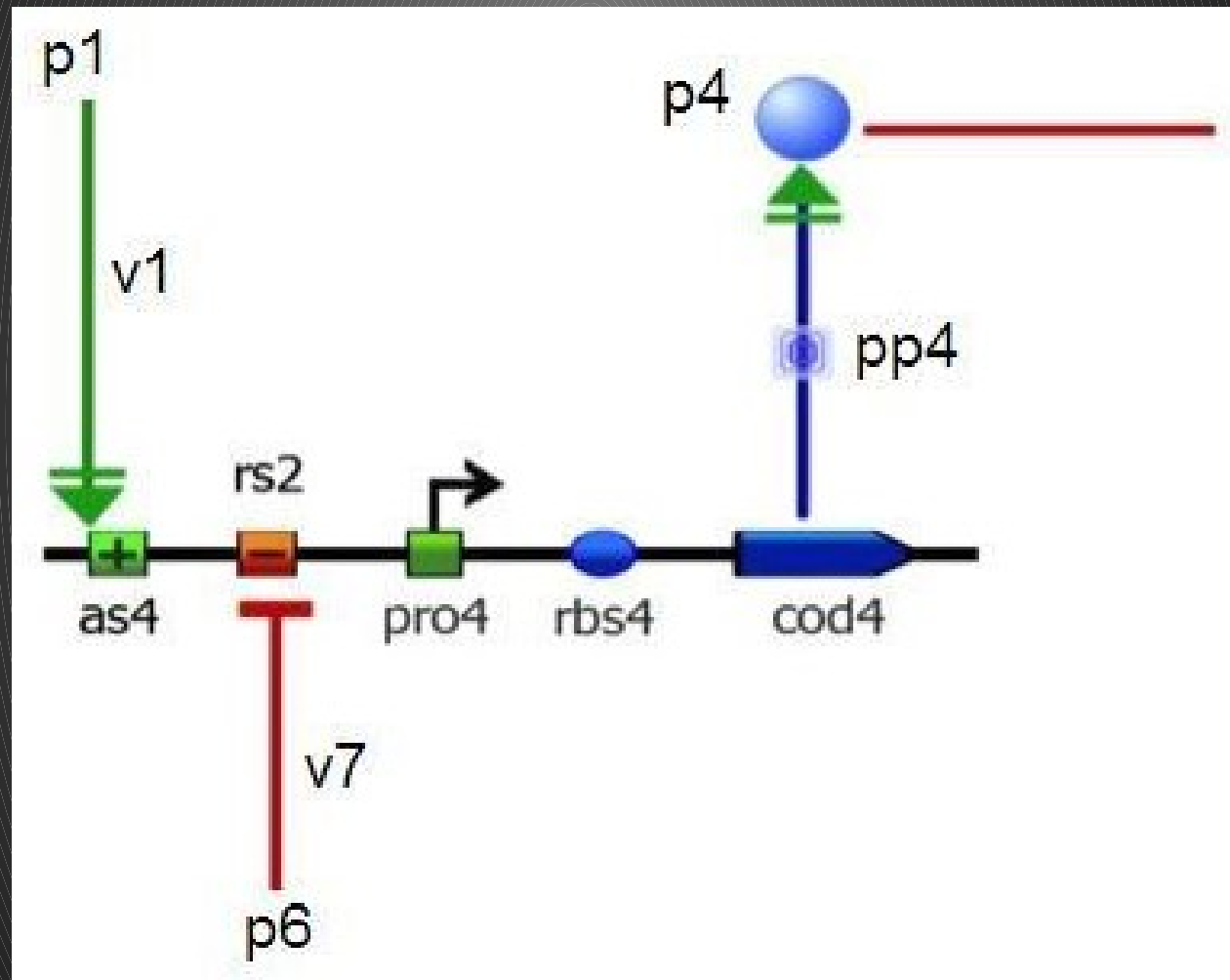
Some biology: central dogma



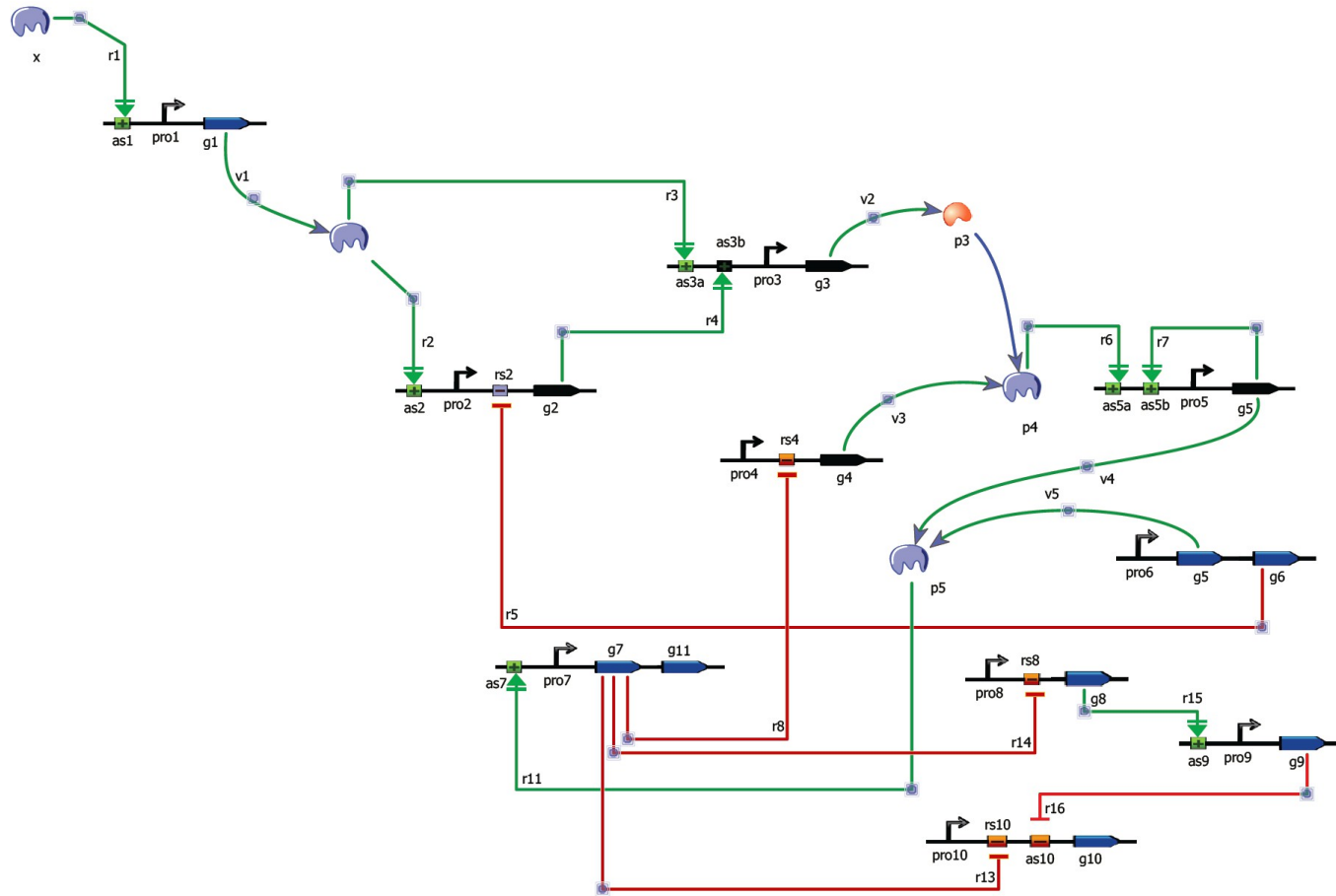
Transcription factors



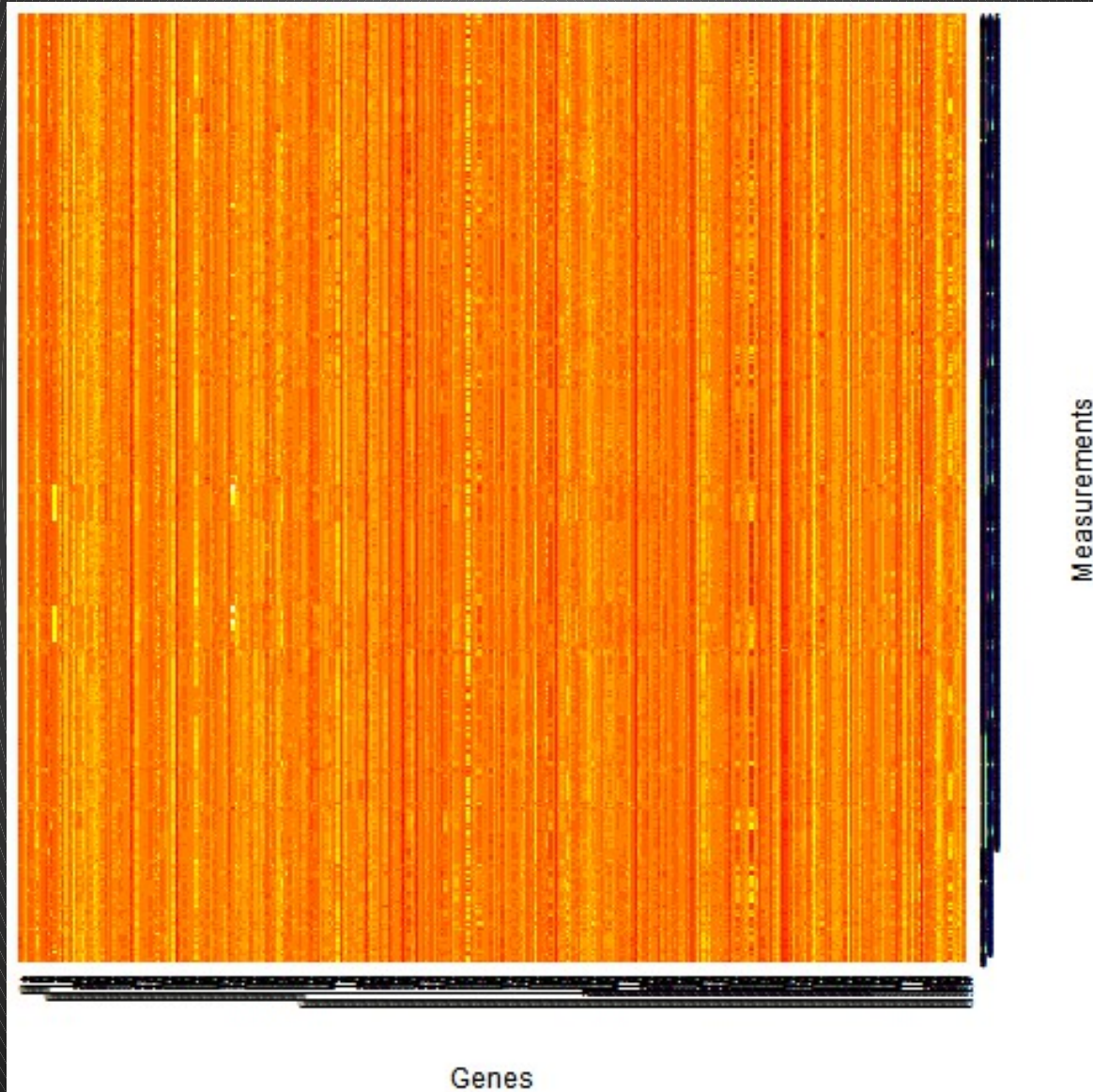
Modeling



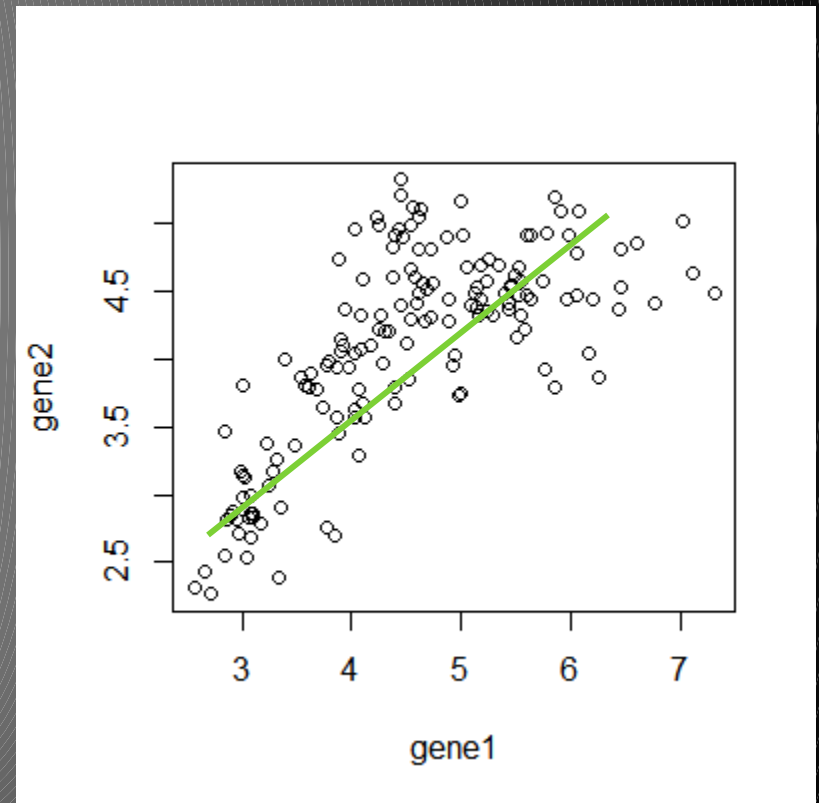
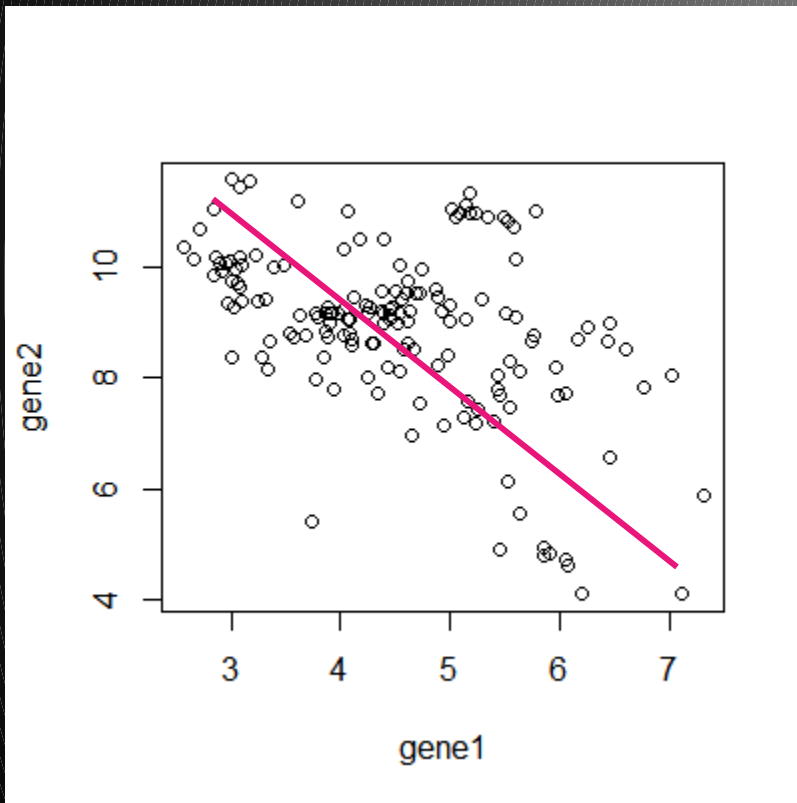
Regulation



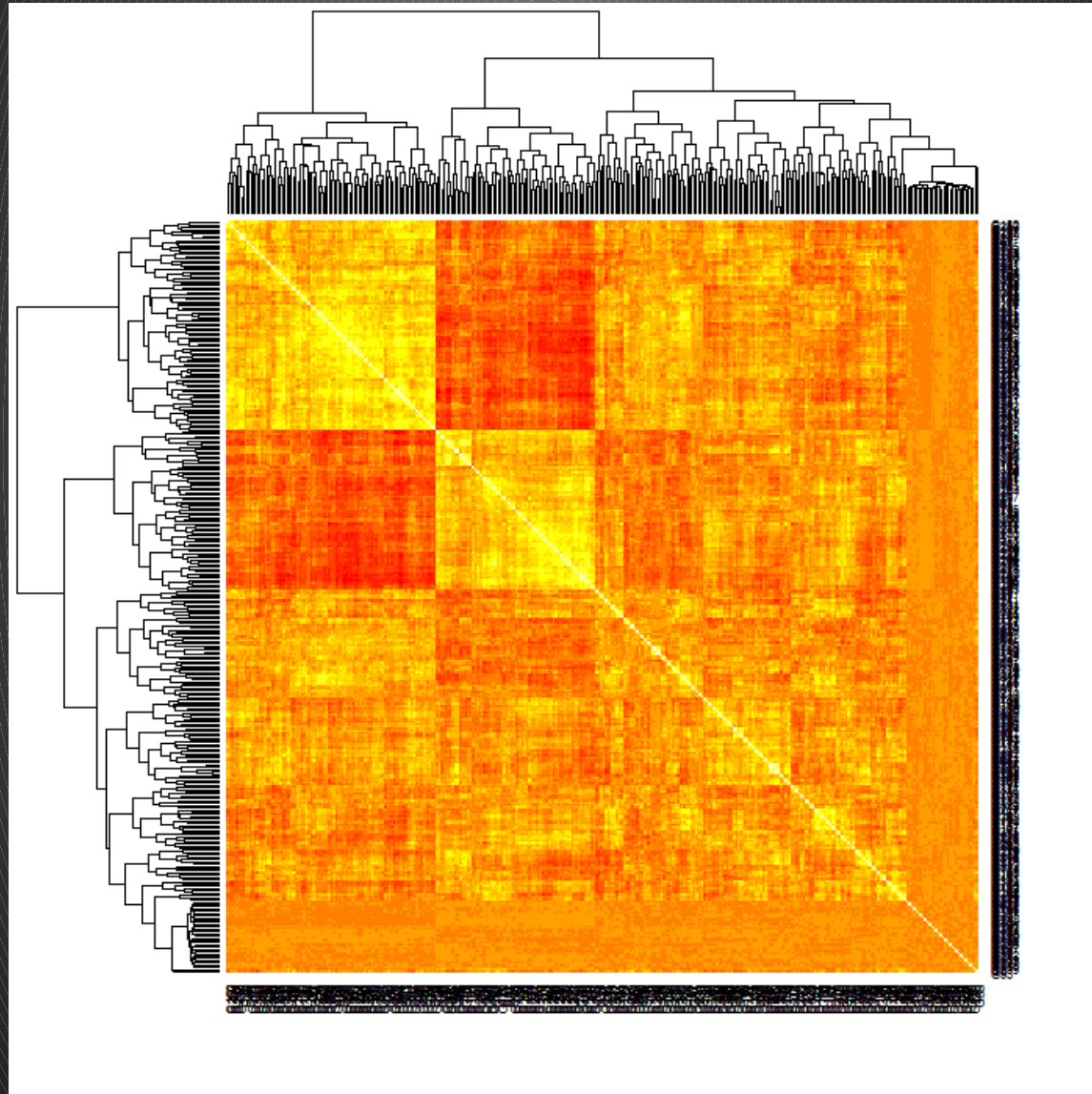
Transcription factor expression



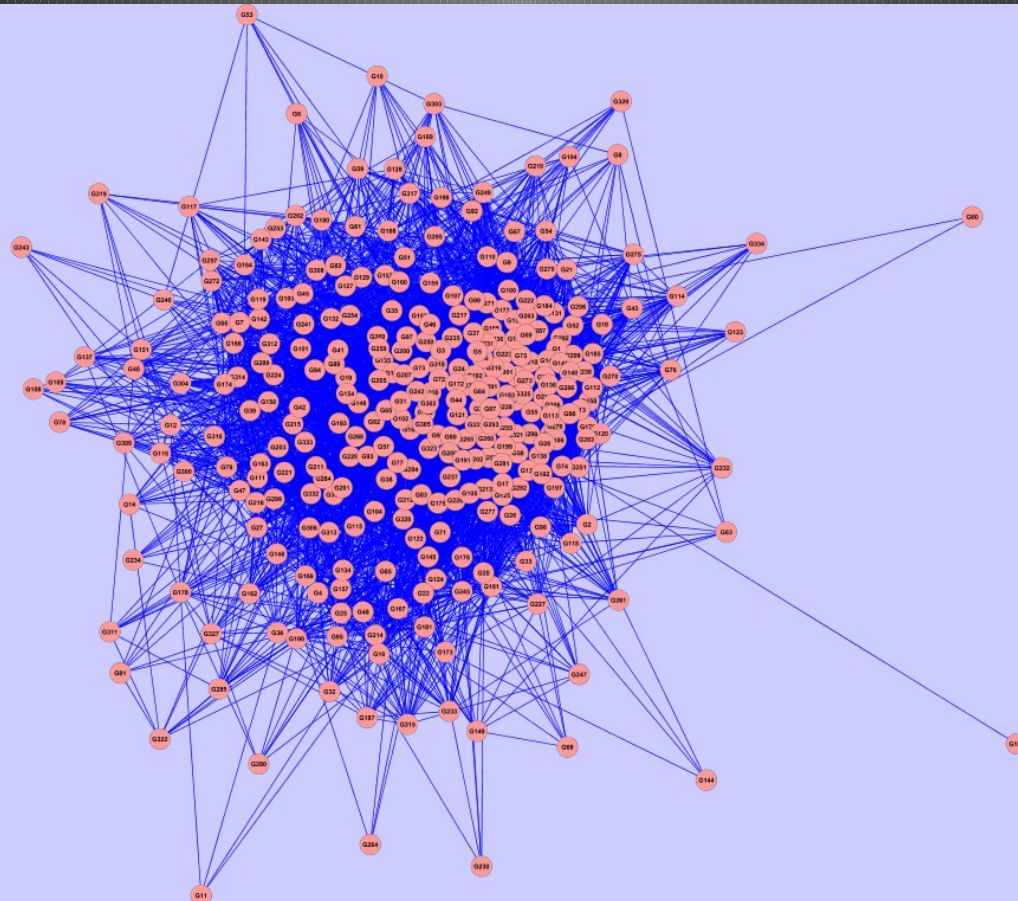
Correlation



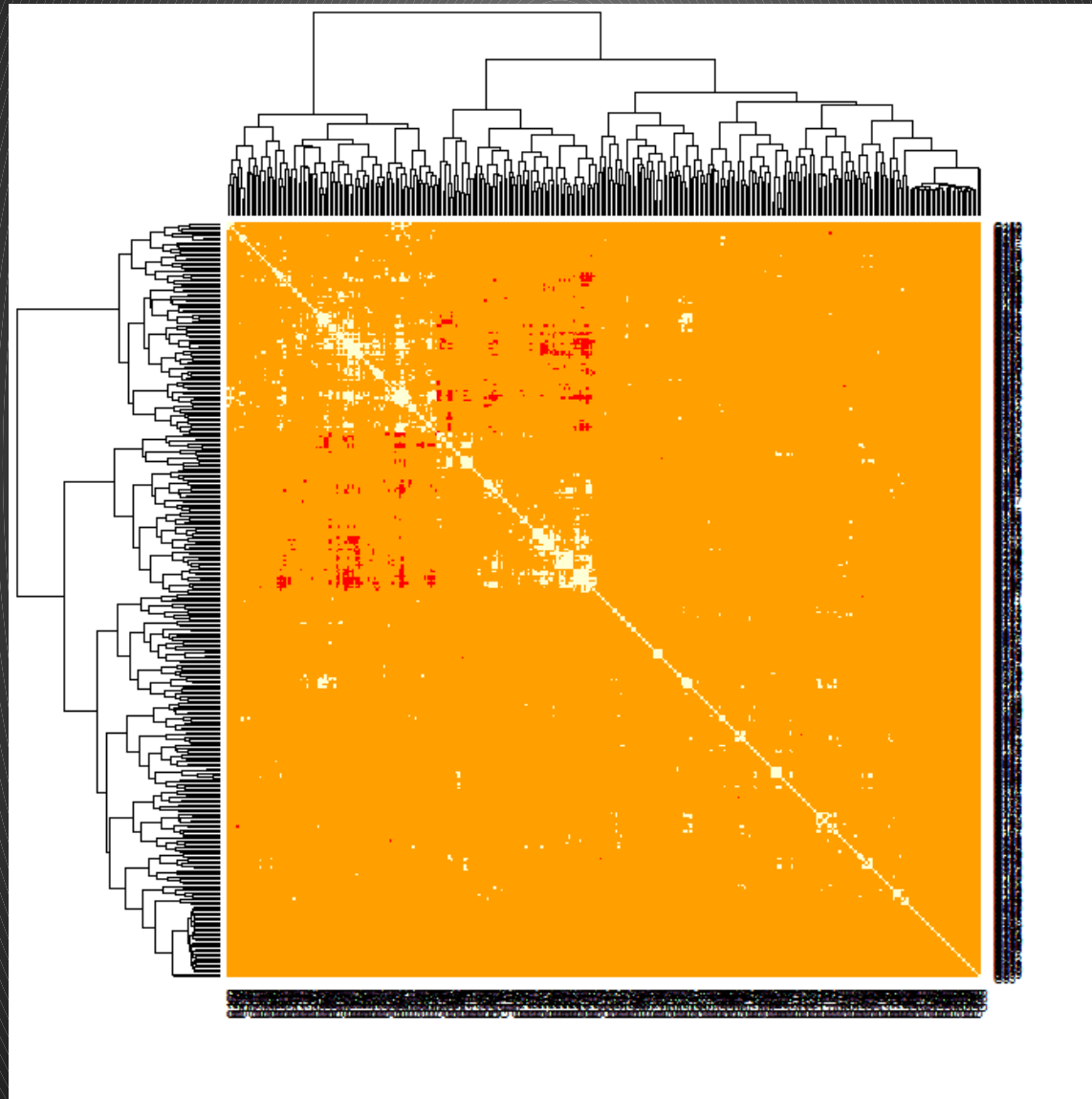
Correlation network



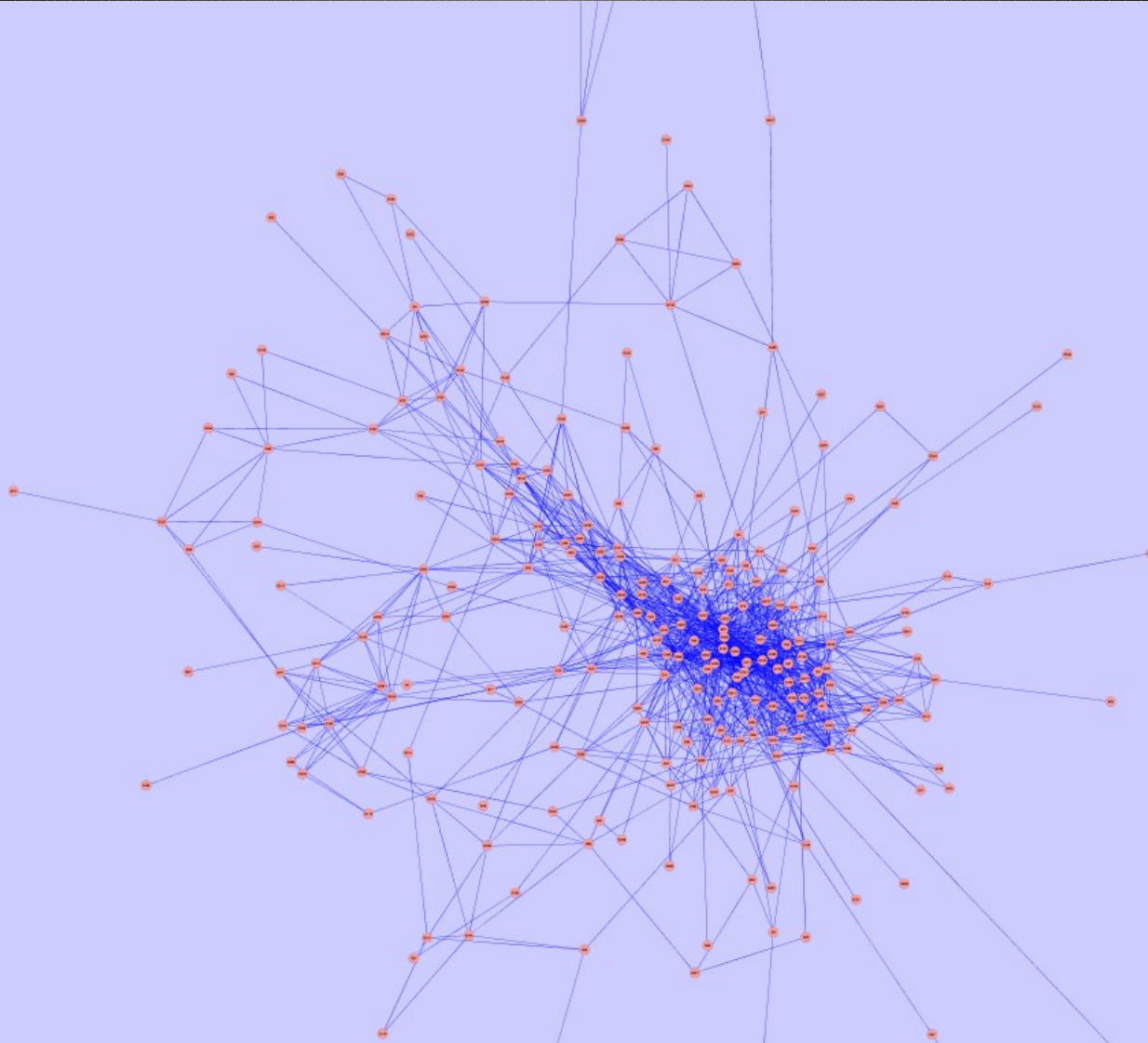
Correlation network



Correlation network



Correlation network

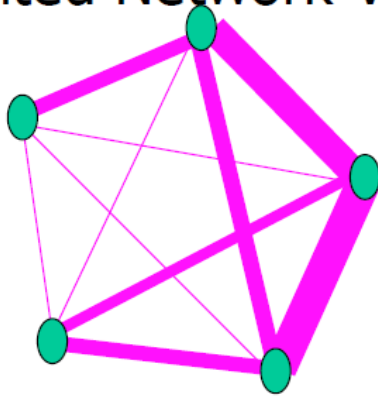


Correlation network

- Problems
 - You only know expression (mRNA or proteomic) at certain times
 - Missing nodes
 - Ton of other stuff going on
 - miRNA
 - Post-translational modifications
 - Spatial effects
 - No directionality

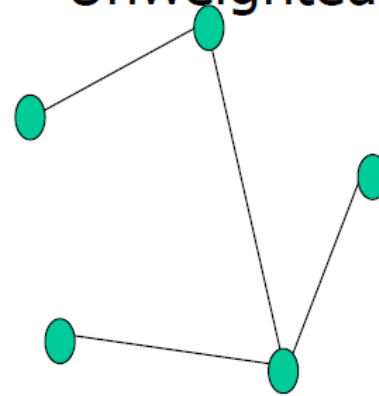
Weighted gene correlation network analysis (WGCNA)

Weighted Network View



- All genes are connected
- Connection Widths=Connection strengths

Unweighted View



- Some genes are connected
- All connections are equal

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4, Article 17

WGCNA

Define a Gene Co-expression Similarity

Define a Family of Adjacency Functions

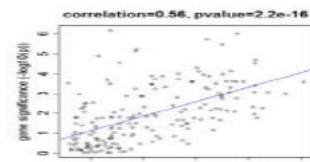
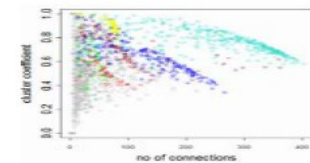
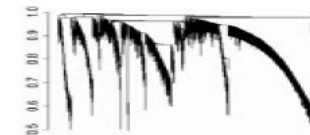
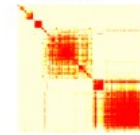
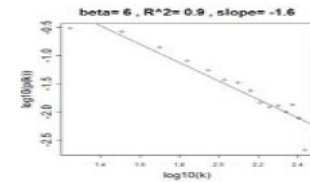
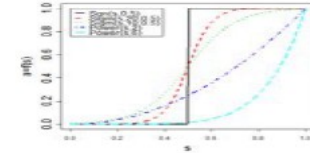
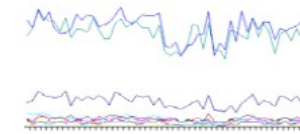
Determine the AF Parameters

Define a Measure of Node Dissimilarity

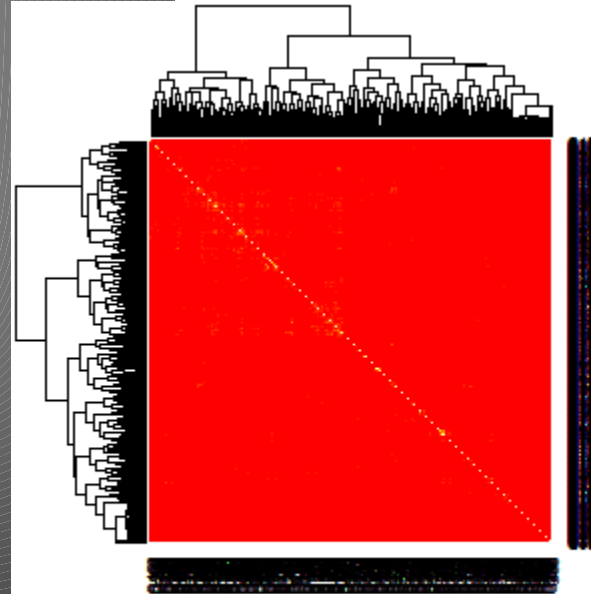
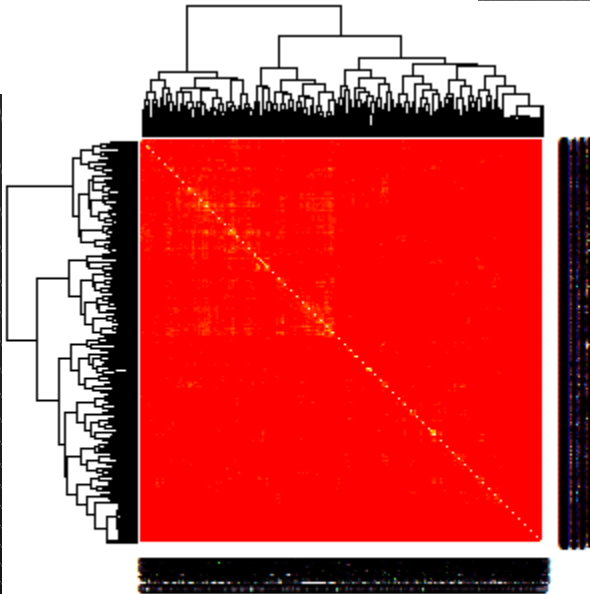
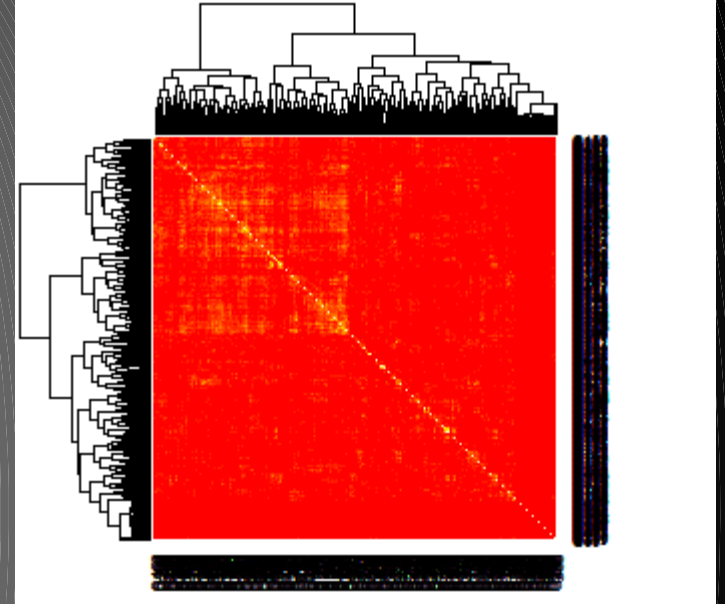
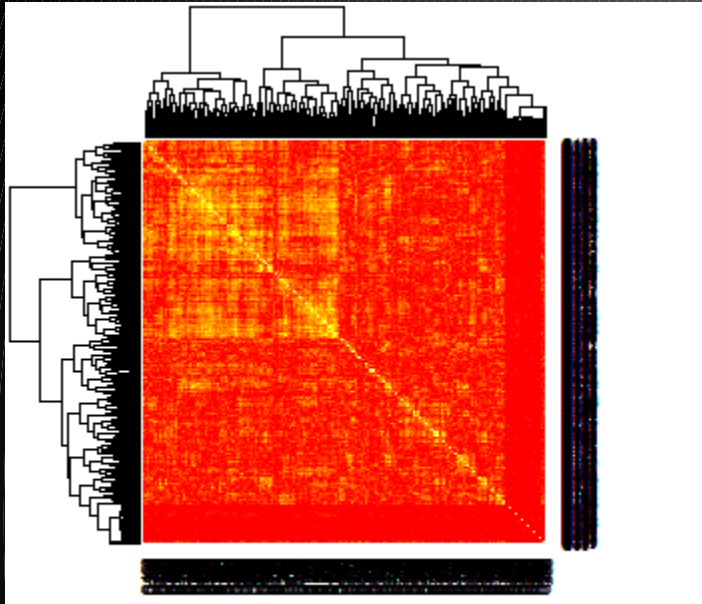
Identify Network Modules (Clustering)

Relate Network Concepts to Each Other

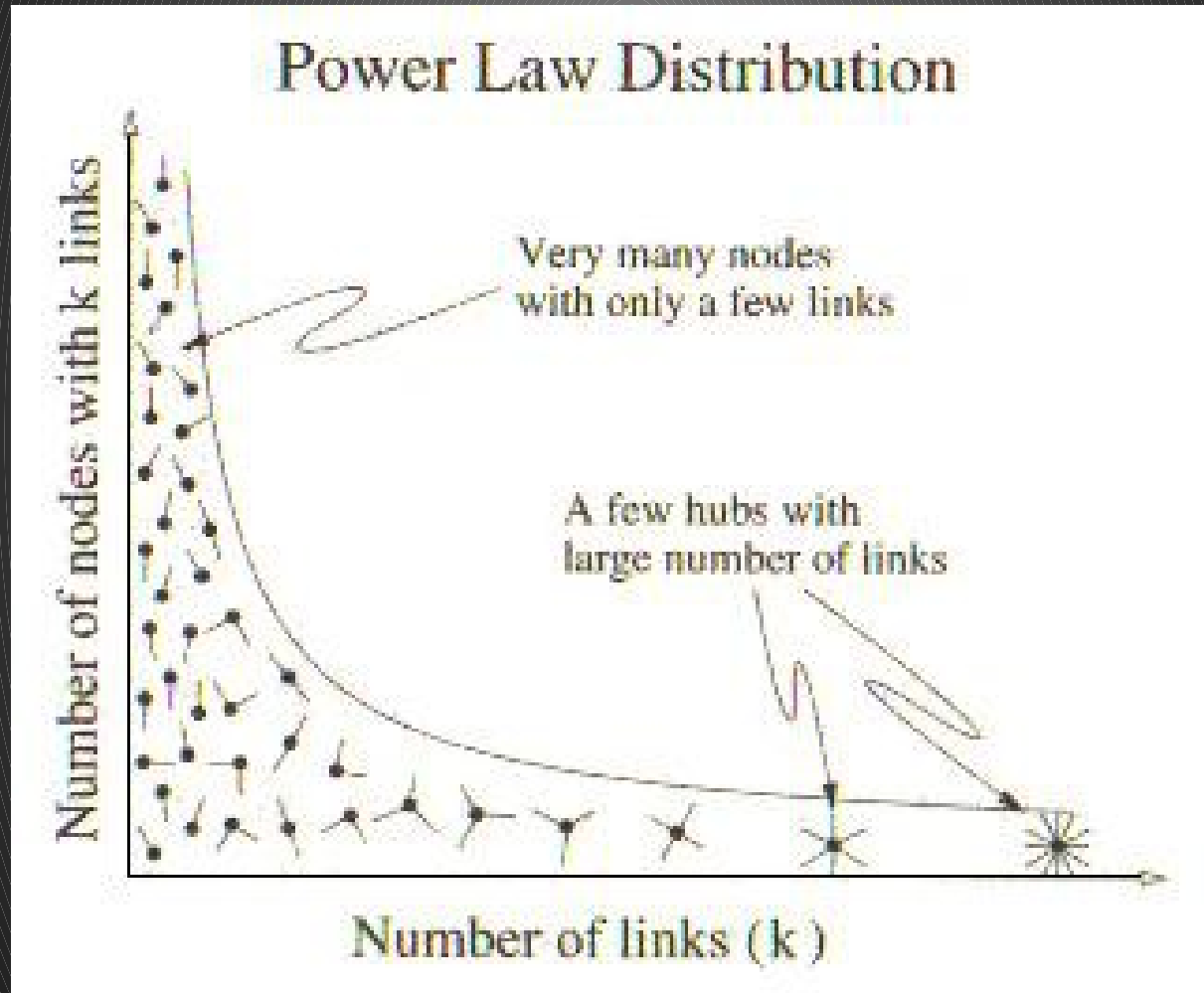
Relate the Network Concepts to External Gene or Sample Information



WGCNA



Scale free networks



WGCNA node similarity

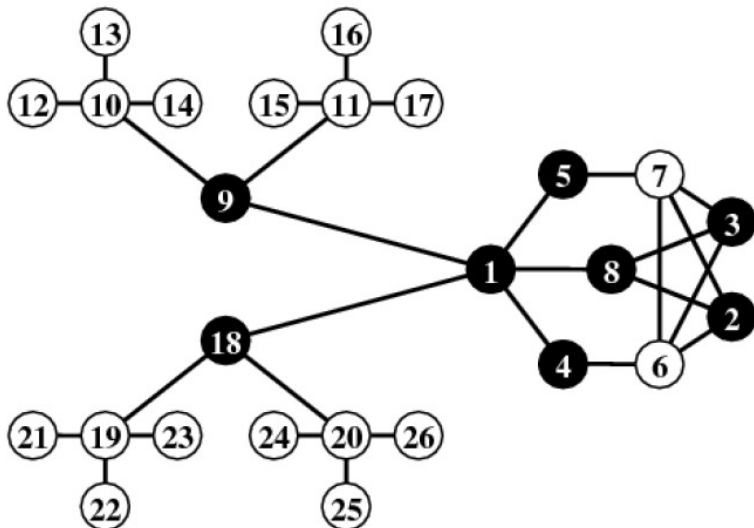
- Generalized topological overlap

(Yip & Horvath 2007)

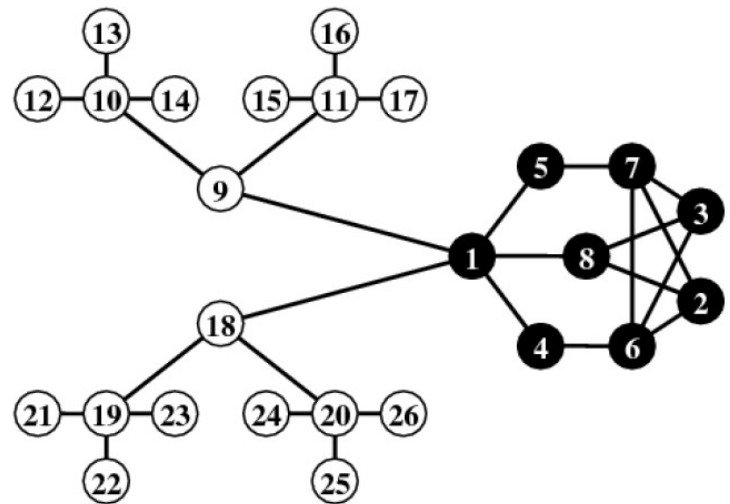
$$N_m(i) := \{j \neq i \mid \text{dist}(i, j) \leq m\}$$

$$t_{ij}^{[m]} = \begin{cases} \frac{|N_m(i) \cap N_m(j)| + a_{ij}}{\min\{|N_m(i)|, |N_m(j)|\} + 1 - a_{ij}} & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

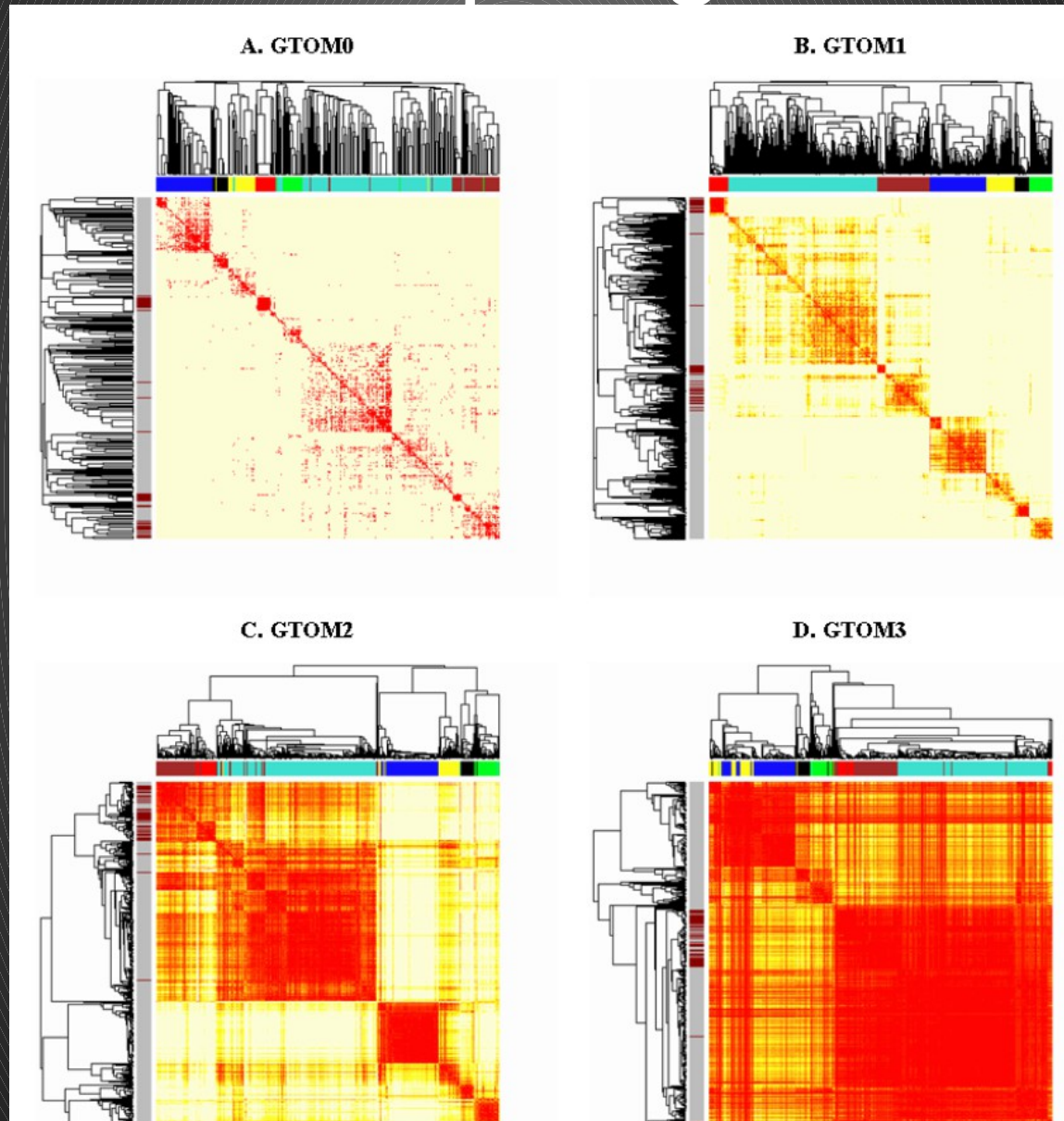
A. GTOM1



B. GTOM2

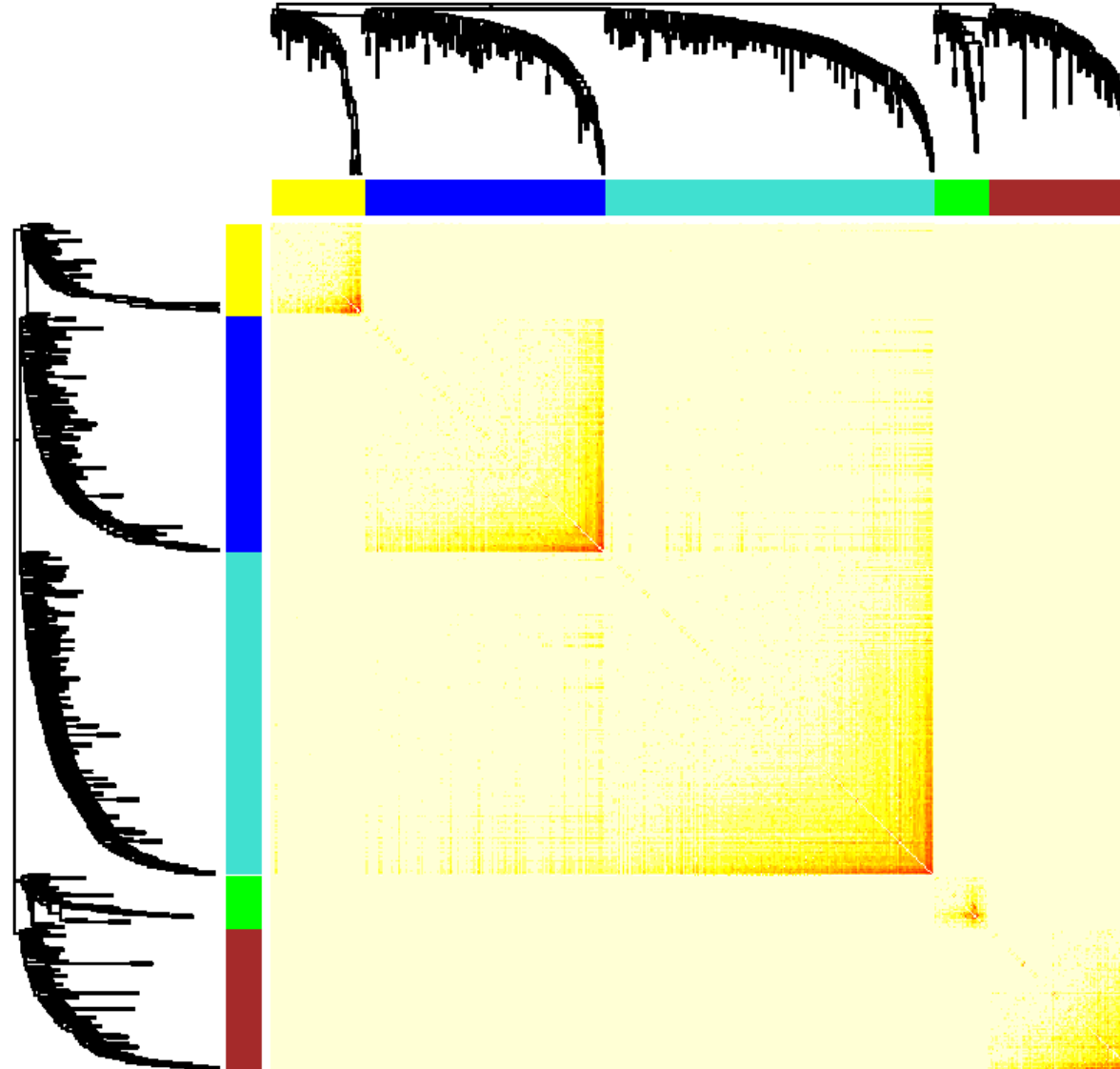


Generalized topological overlap



WGCNA

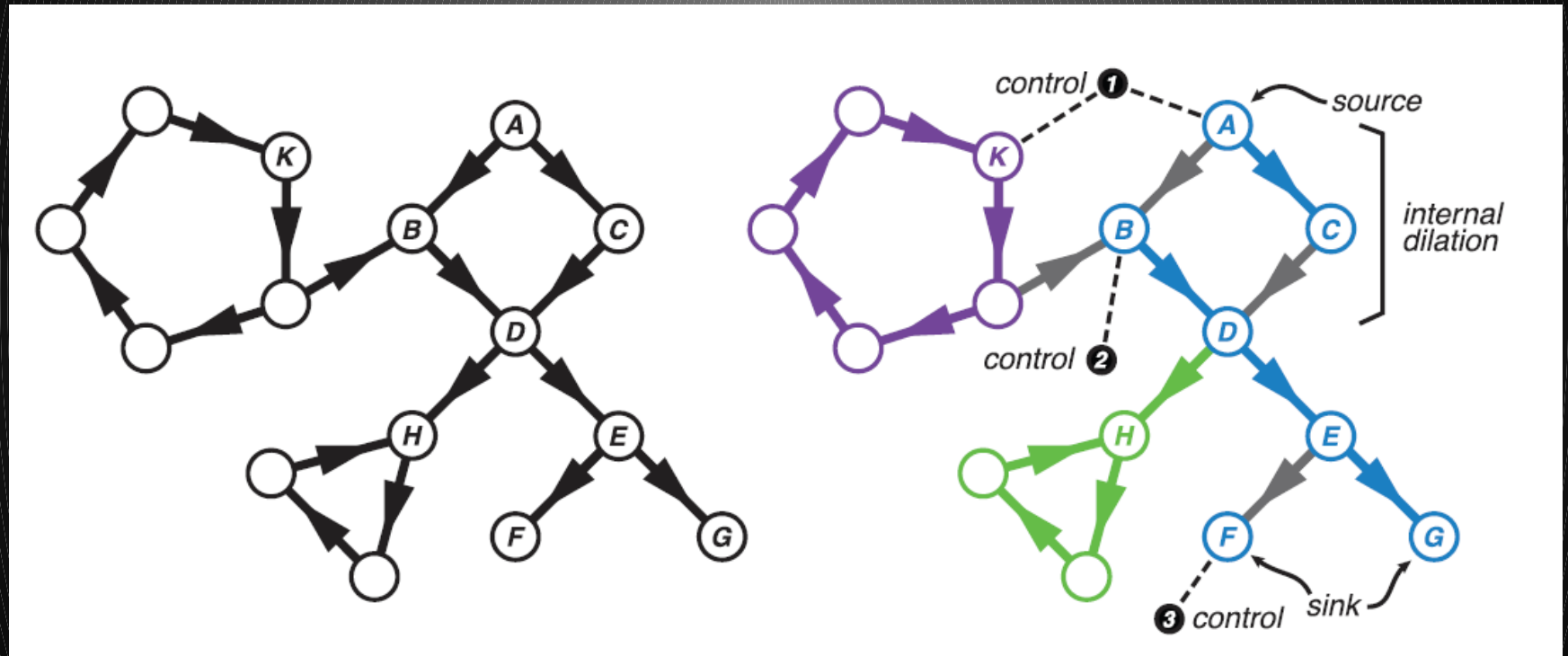
C. Network heatmap plot



Other graph inference methods

- Information theoretic
 - Use mutual information between nodes instead of correlation
- Regression based (directed graphs)
 - Random forests
 - Regularization using p-norms
- Bayesian inference
- Ensemble methods

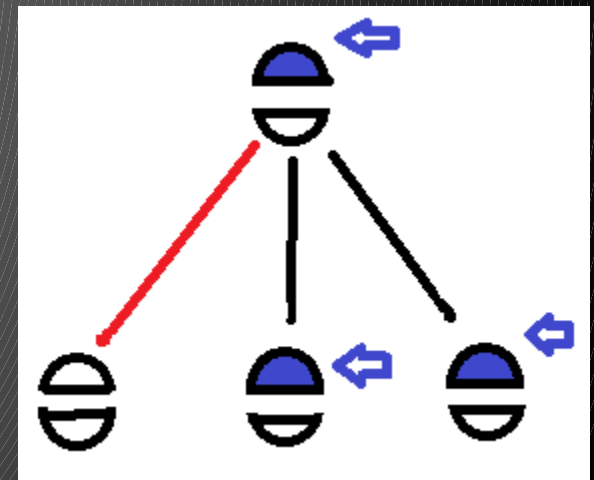
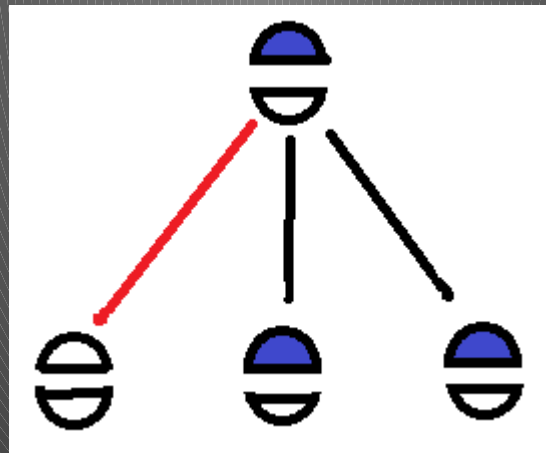
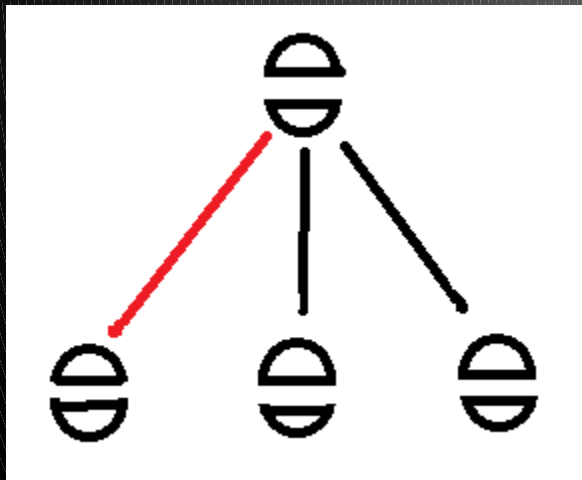
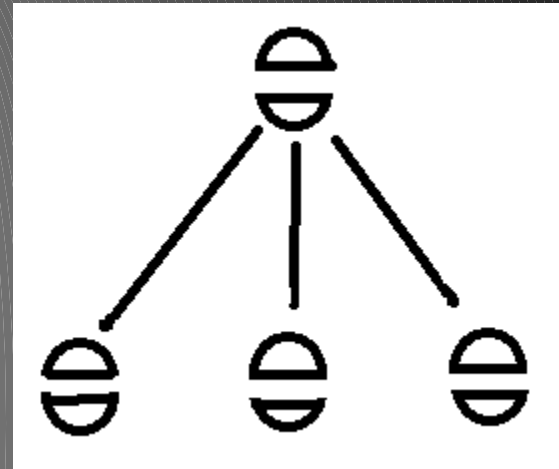
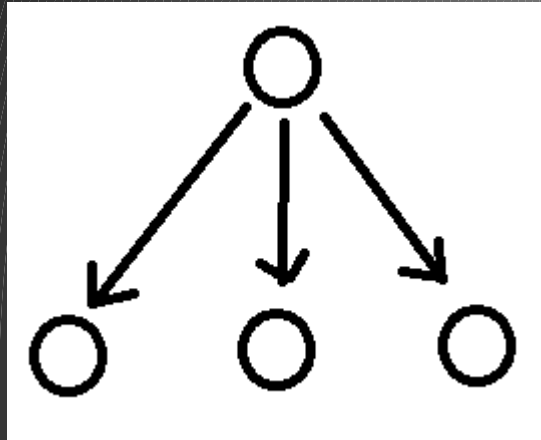
Controlability of graphs

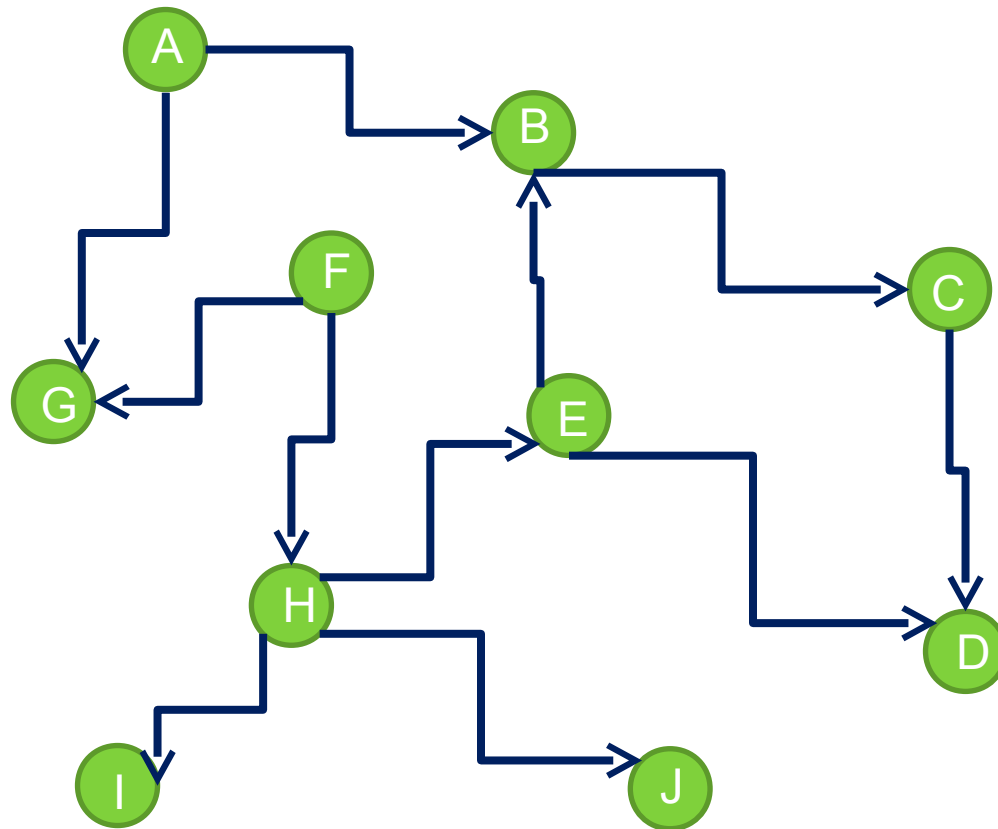


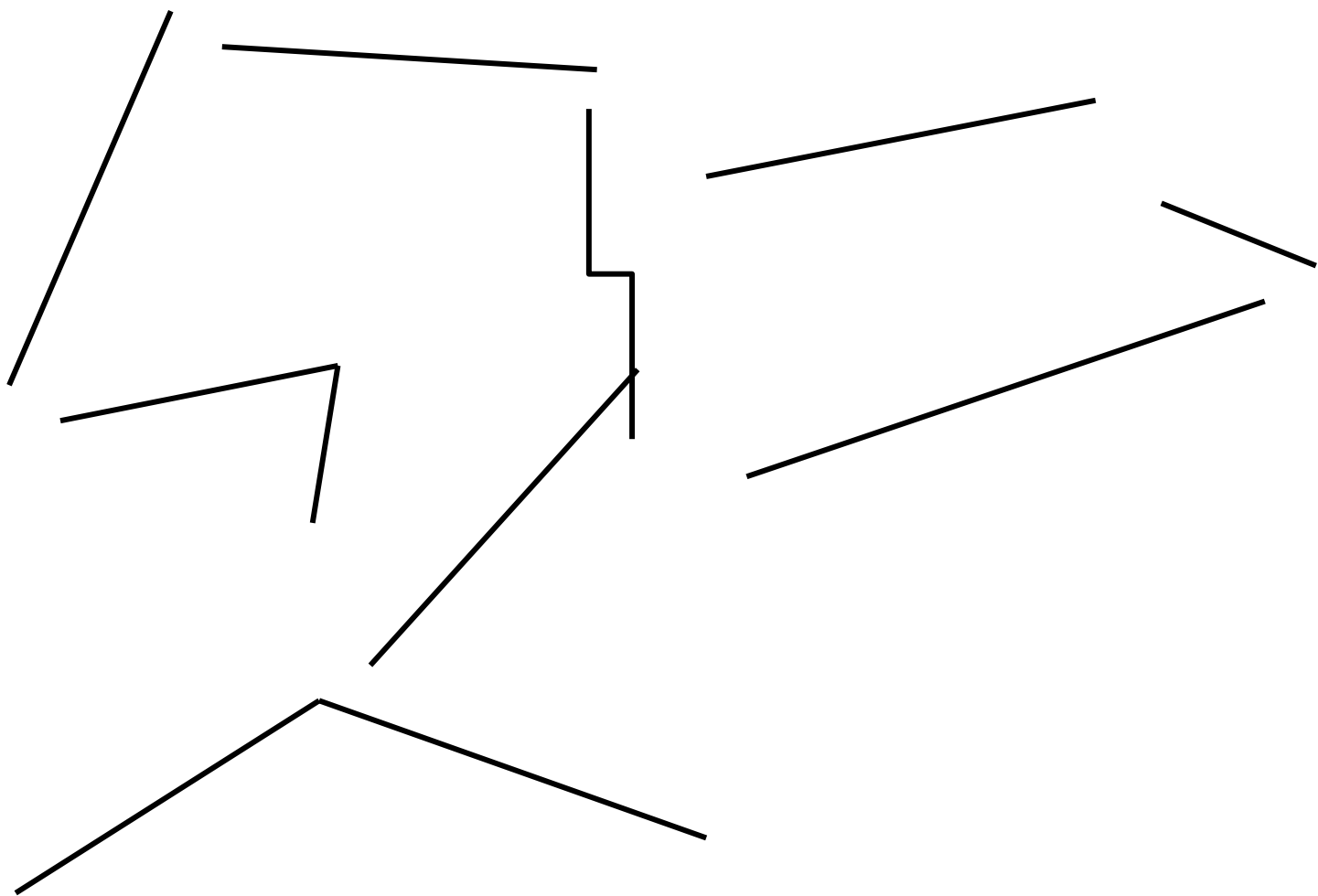
Controlability of graphs

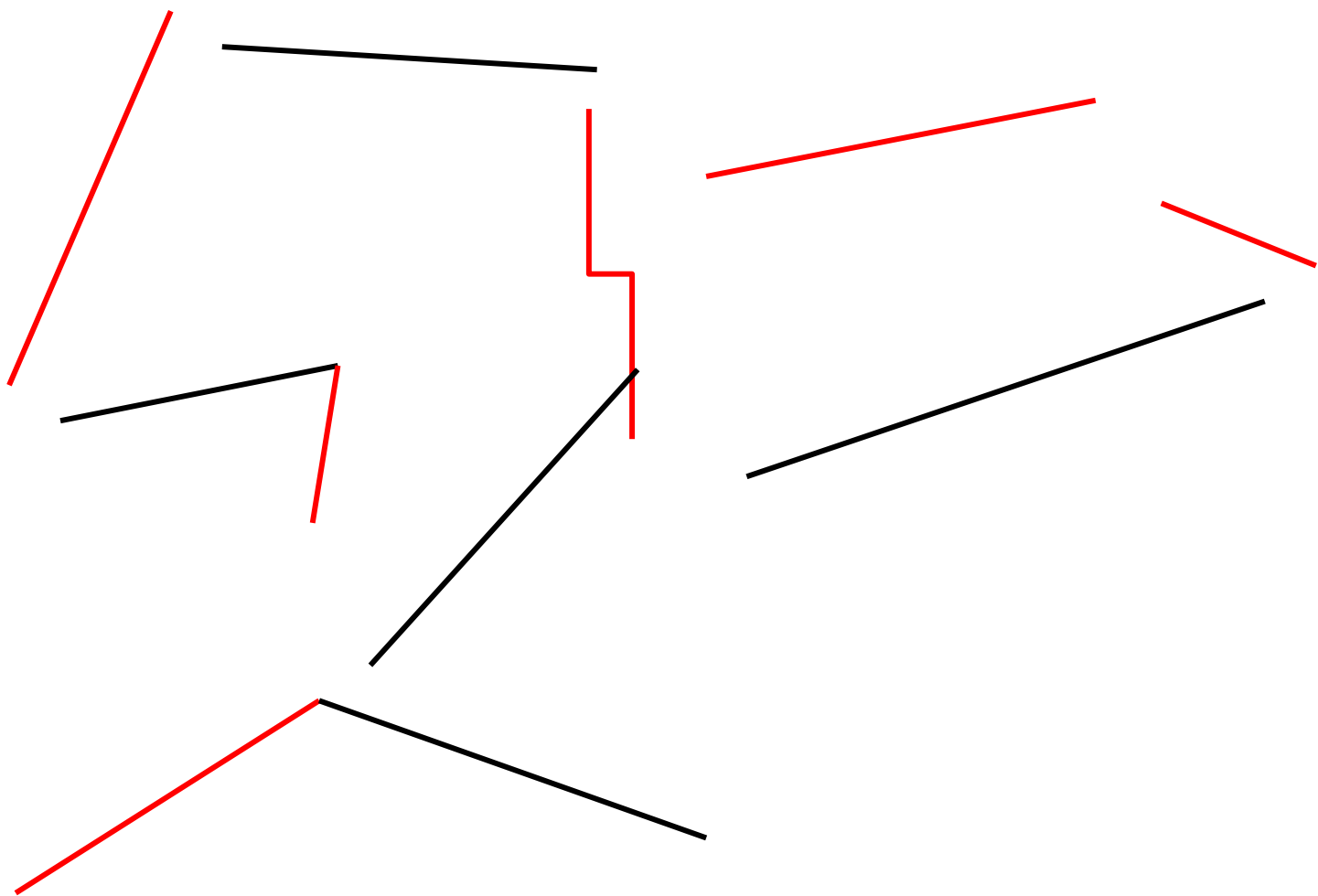
- Find control profiles by solving the ‘maximum matching problem’
 - Split nodes into in and out nodes w/o connections between and get rid of directions
 - Find the maximum matching.
 - Largest set of edges w/o sharing a node
 - “In nodes” that don’t have an edge from the maximal matching must be controlled.
- Look at what types of nodes need control.

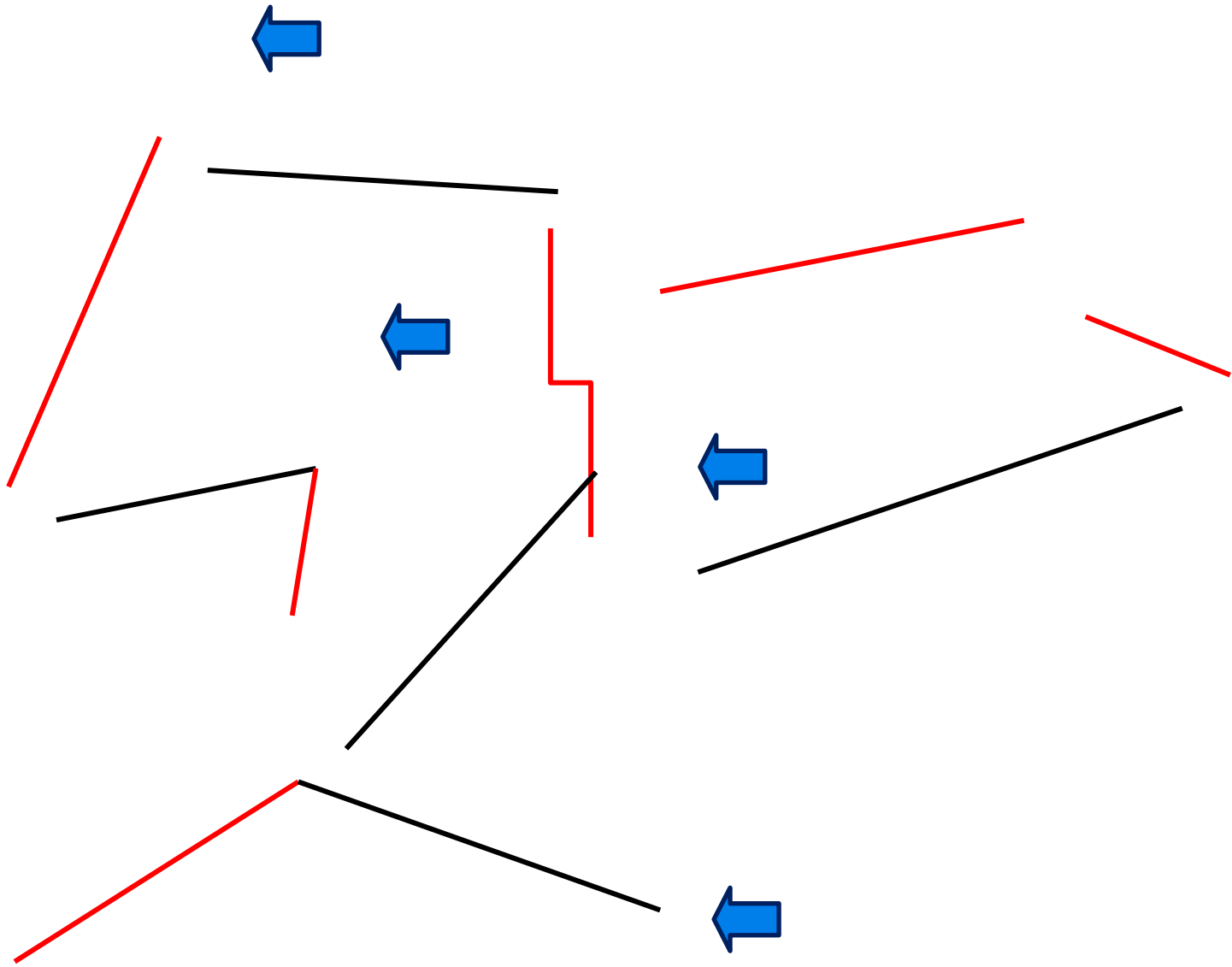
Controlability of graphs

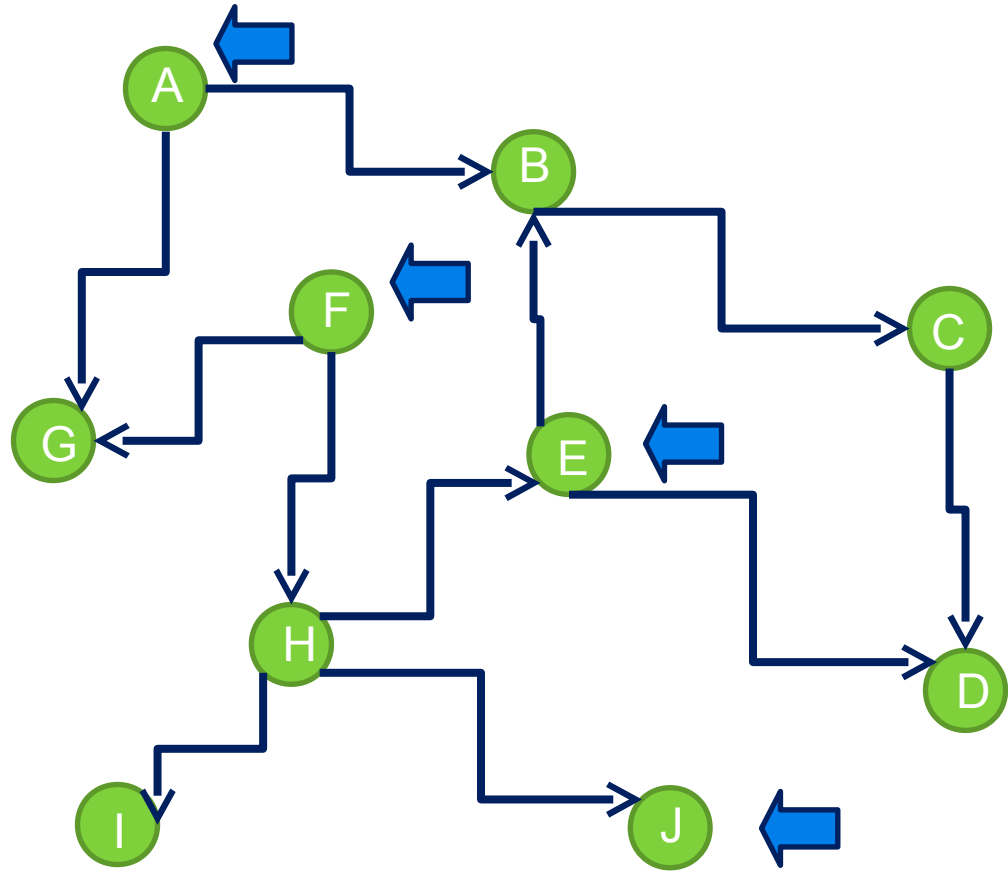




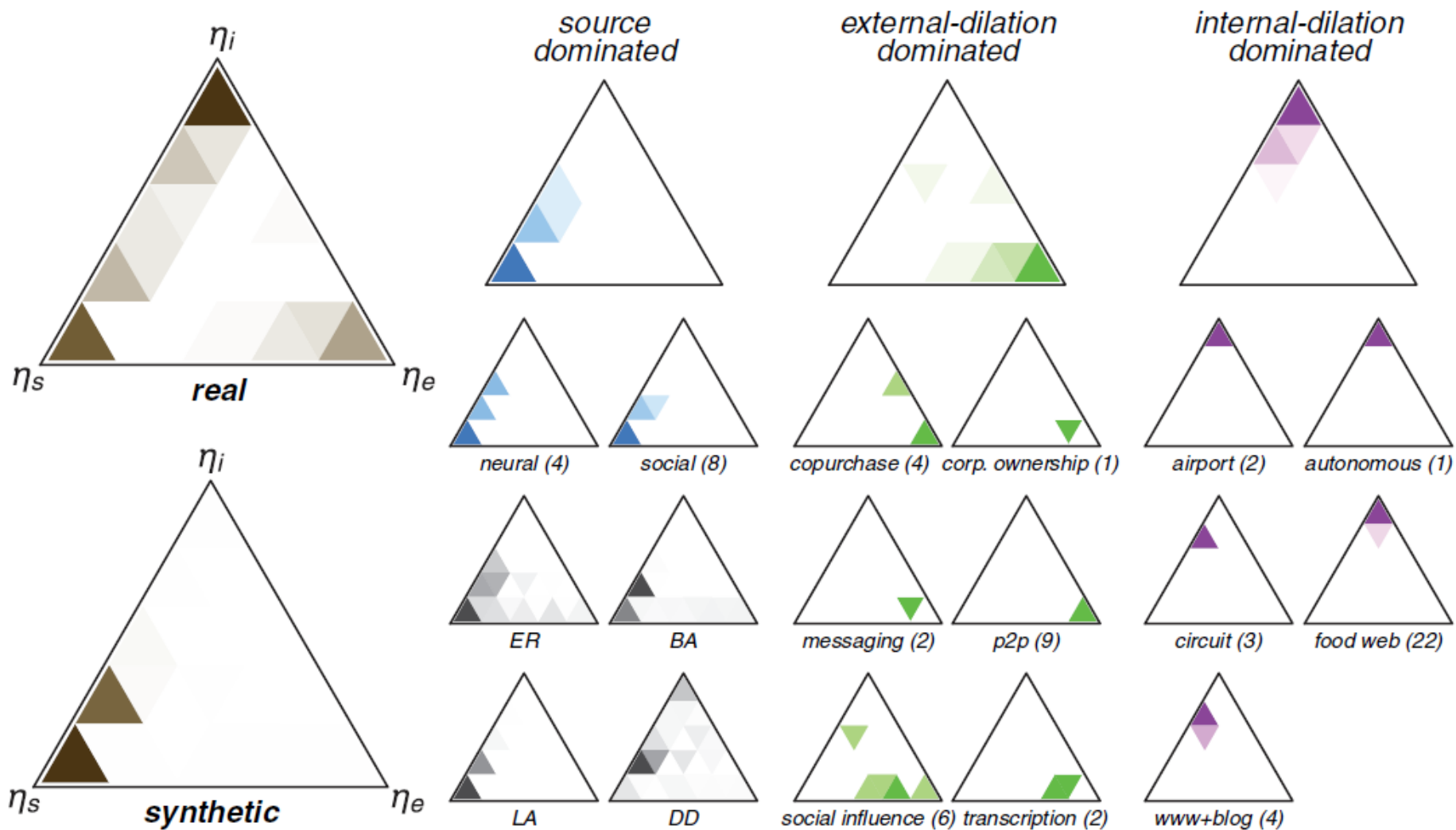




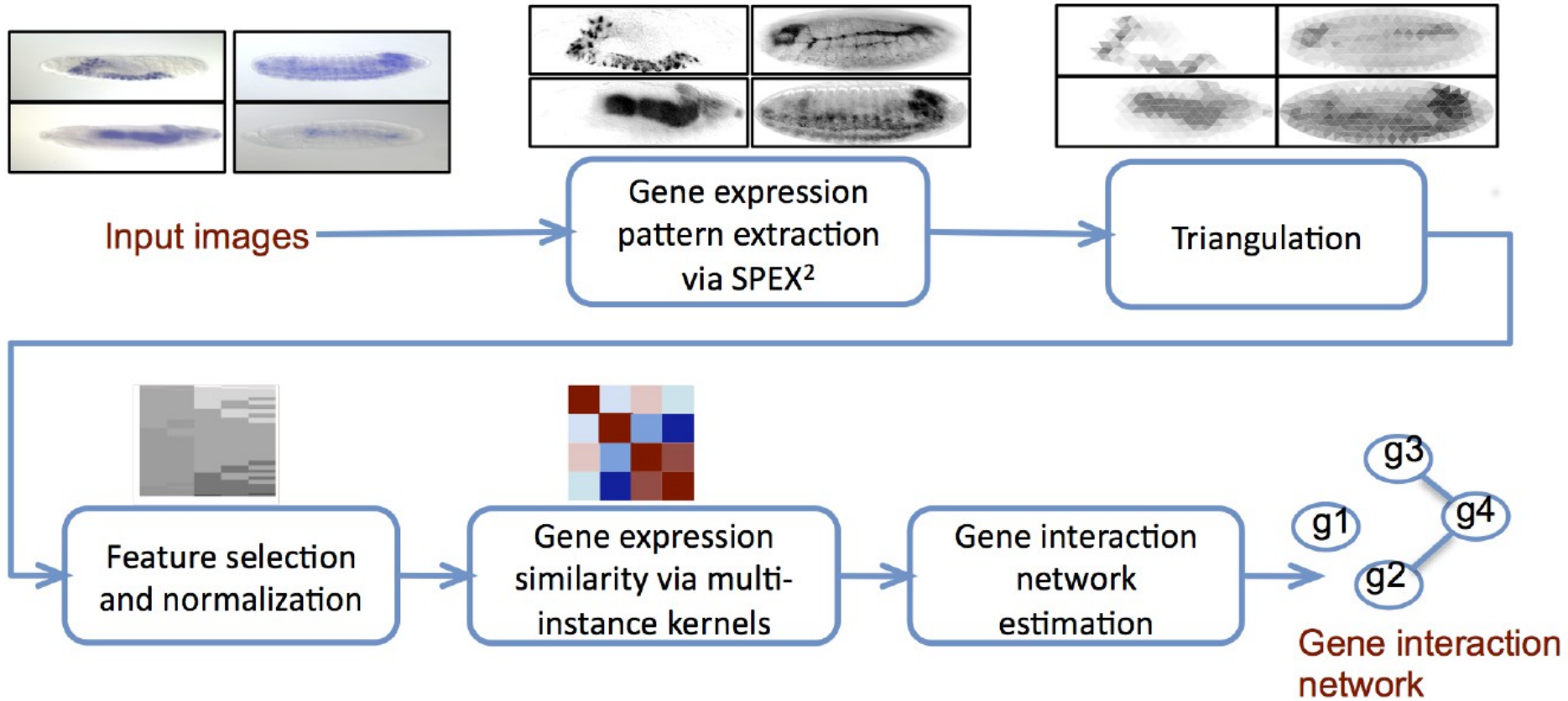




Controlability of graphs

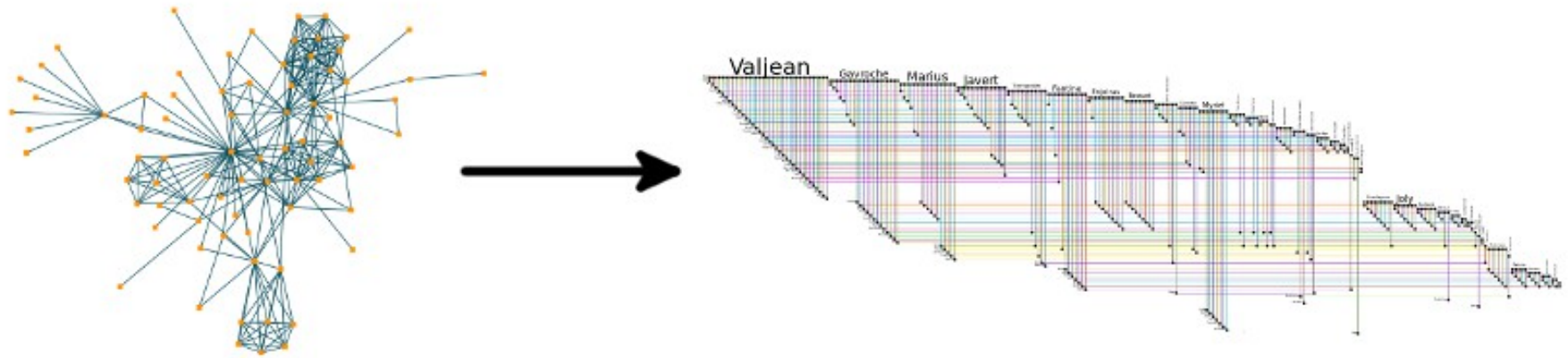


Spatial modeling

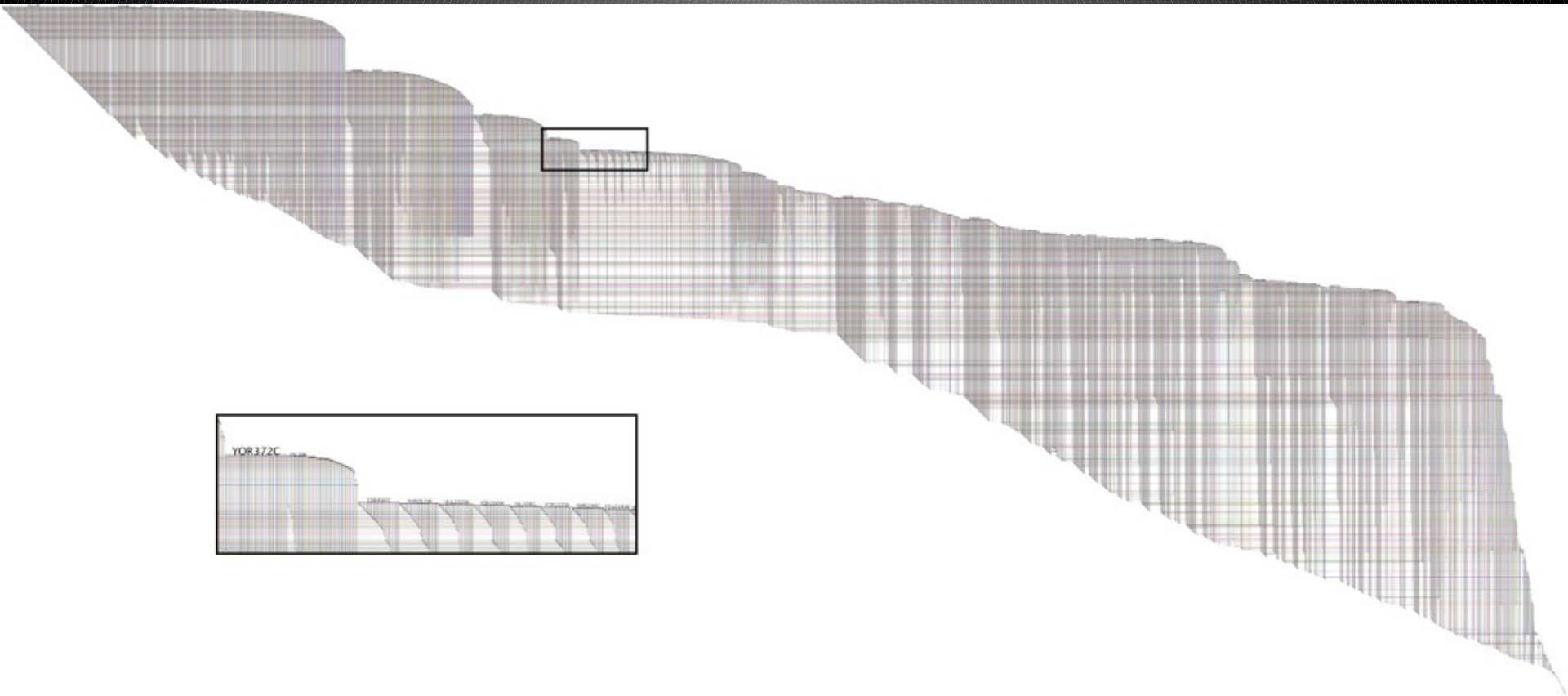


Graph visualization

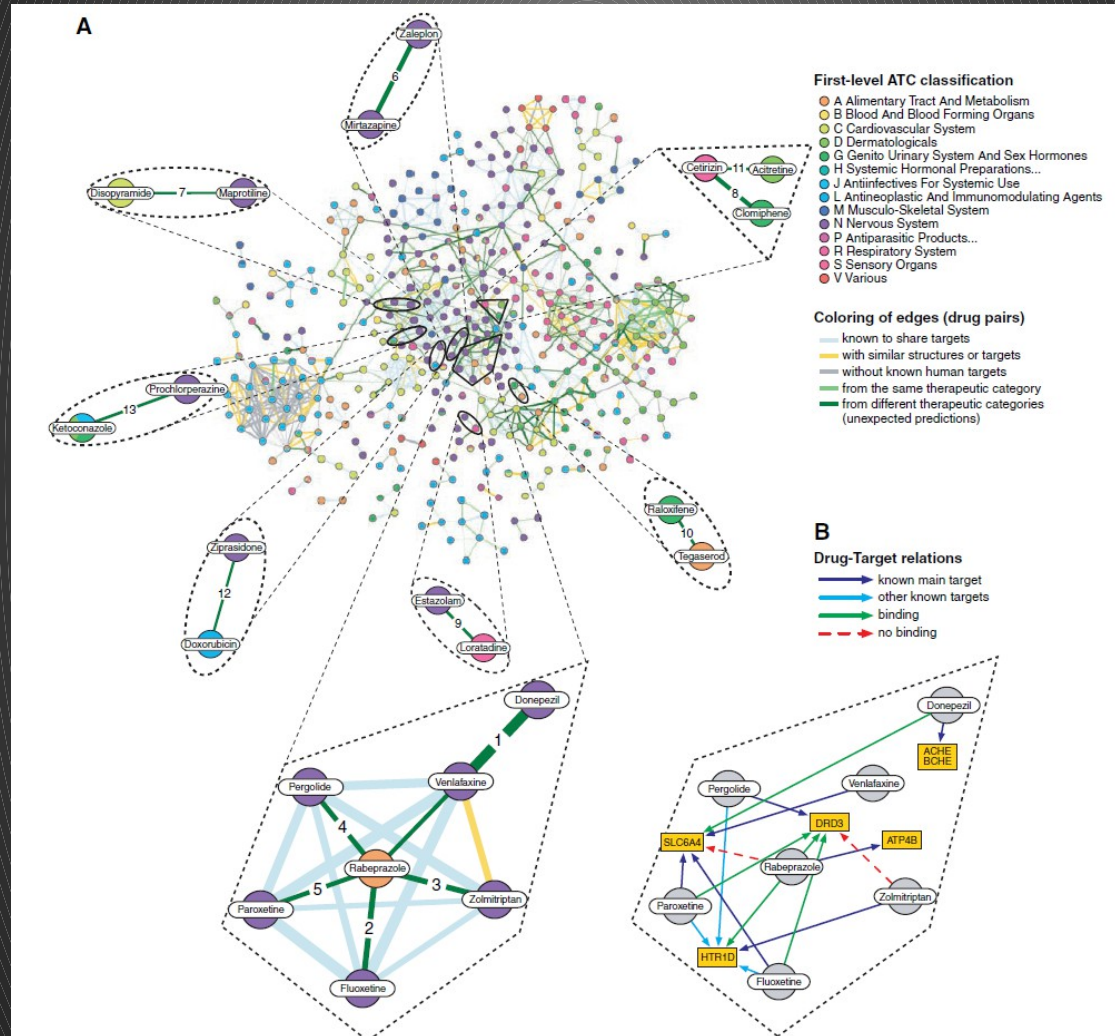
- Biofabric (<http://www.biofabric.org/>)



Biofabric

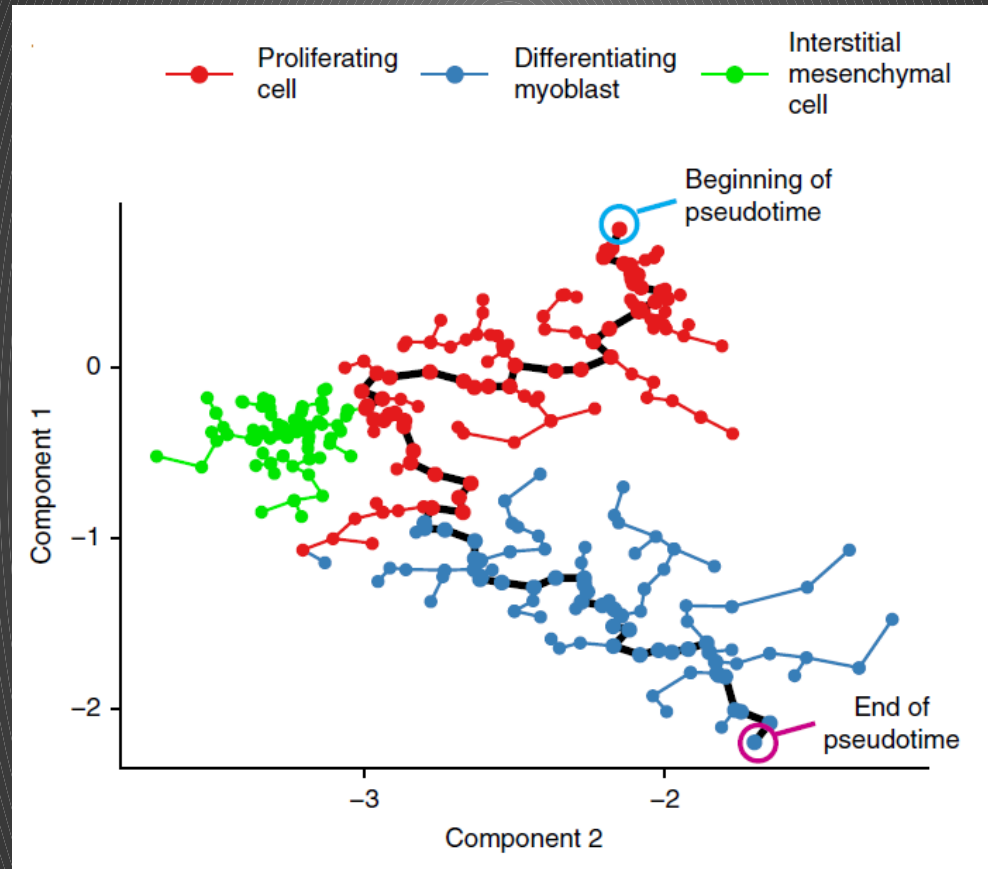


Drug repositioning



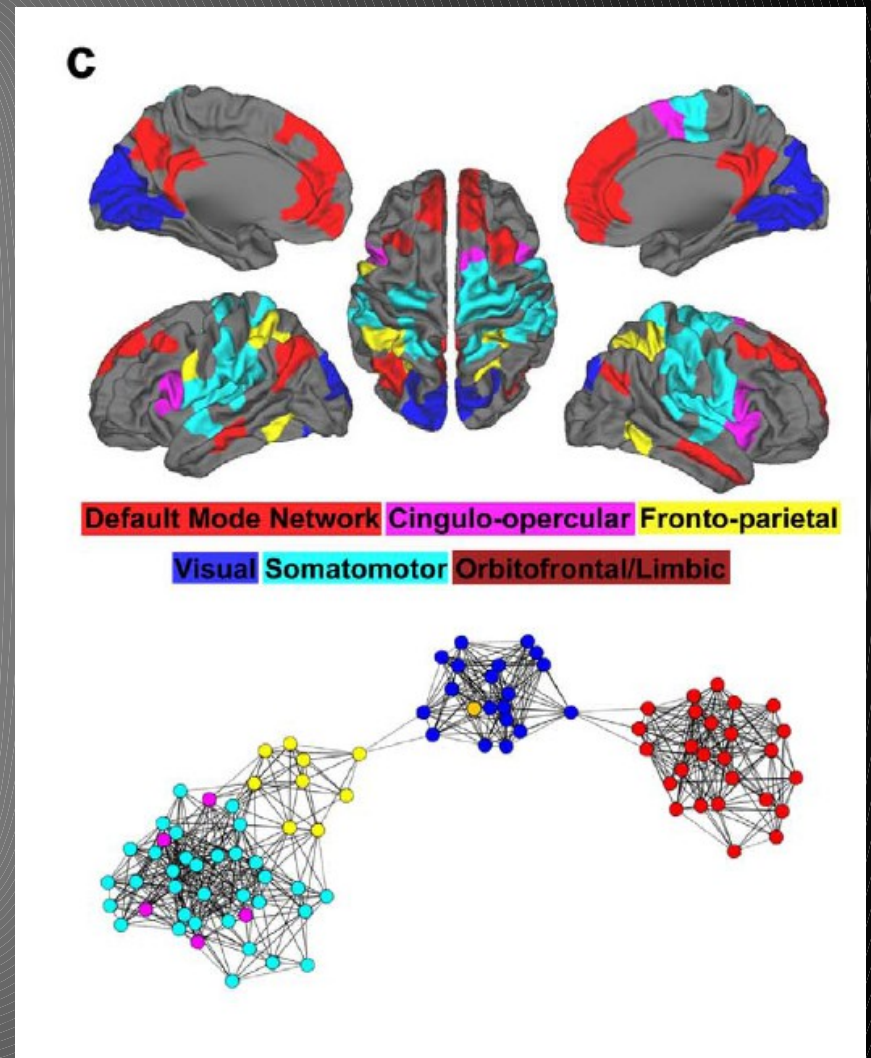
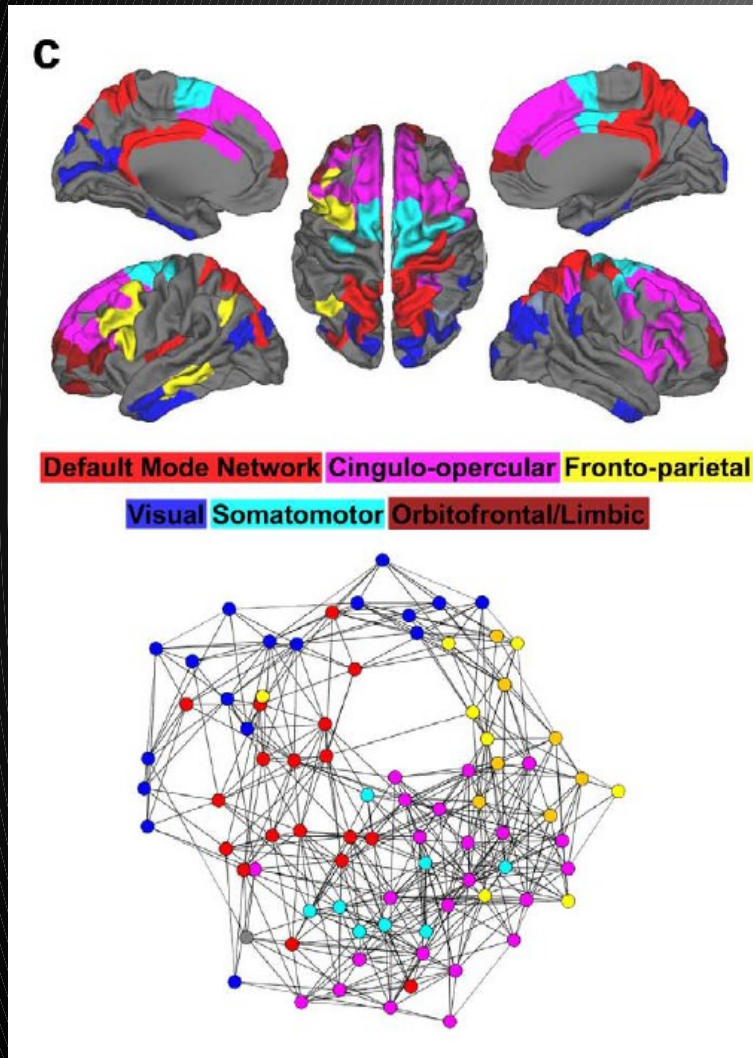
Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L.J., and Bork, P. (2008). Drug Target Identification Using Side-Effect Similarity. *Science* 321, 263–266.

Cell differentiation networks



Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* *advance online publication.*

Brain networks



Grayson, D.S., Ray, S., Carpenter, S., Iyer, S., Dias, T.G.C., Stevens, C., Nigg, J.T., and Fair, D.A. (2014). Structural and Functional Rich Club Organization of

Open Questions

- How can we best simulate background models of interaction networks?
- What graph theory measures capture the salient qualities of real graphs?
- How do we model changes of interaction networks over time?
- How do we incorporate spatial aspects?
- Can we use symmetry groups?