Physical Network Models

BMI/CS 776
www.biostat.wisc.edu/bmi776/
Mark Craven
craven@biostat.wisc.edu
Spring 2011

Goals for Lecture

the key concepts to understand are the following

- physical network models
- the inference task for physical network models

Physical Network Models

- the Bayes/module network learning task
  given: many measurements for each gene
  infer: dependencies among genes

- the physical network learning task
  given:
  - "known" interactions among genes
  - measurements resulting from knocking out/down selected genes

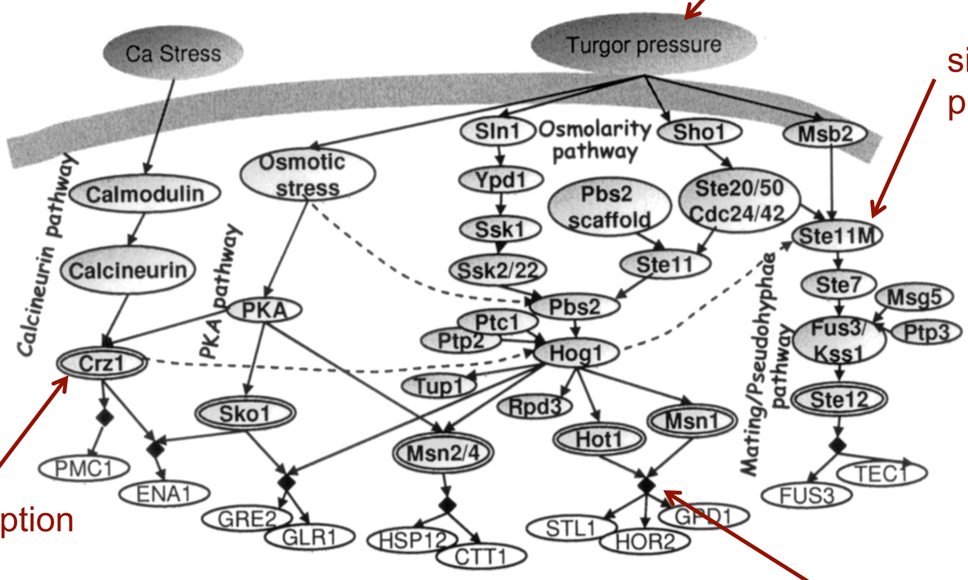
  infer: the subset of interactions (and perhaps their directions and signs) that account for the measured responses

Physical Networks

- model of yeast response to osmotic and calcium stress

environmental condition

signaling protein



transcription factor

shared regulatory program

Figure from Gat-Viks & Shamir, *Genome Research* 17:358-267, 2007.

Physical Networks

- edges represent physical interactions

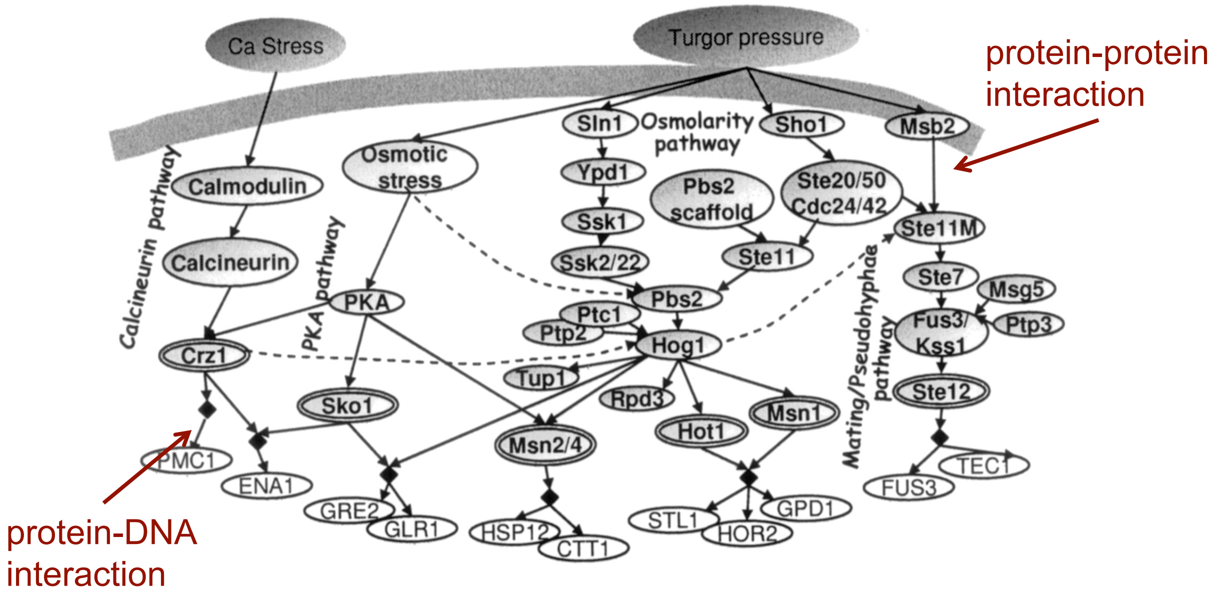
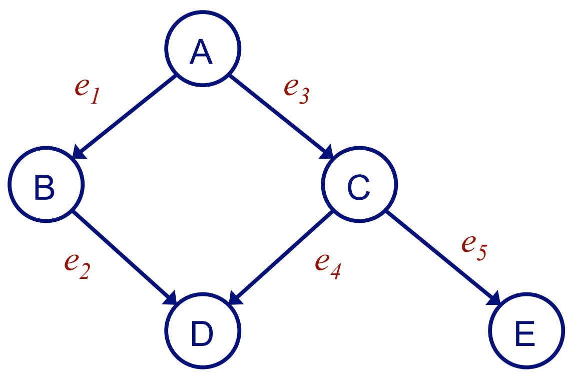


Figure from Gat-Viks & Shamir, *Genome Research* 17:358-267, 2007.

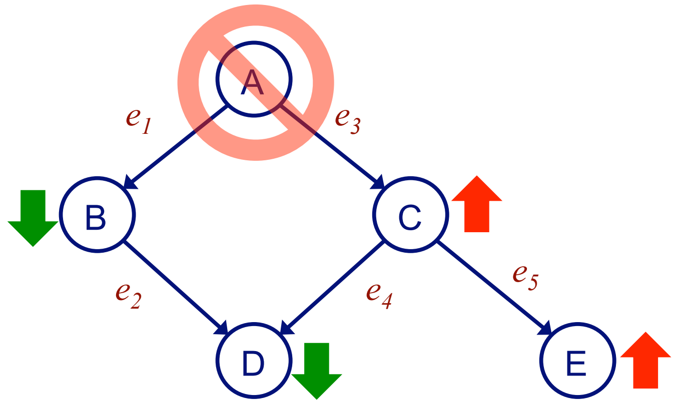
Physical Network Models

- edges in physical network models correspond to
  - protein-protein interactions
  - protein-DNA interactions



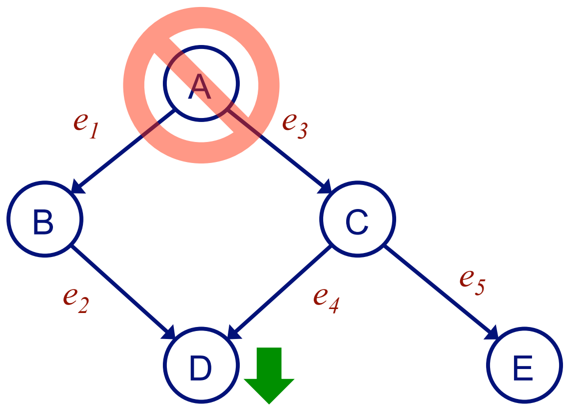
Single Gene Knockout/Knockdown Experiments

- consider an experiment in which we
  - manipulate a gene by knocking it out (disabling it) or knocking it down (lowering its expression)
  - then measure the effect of this manipulation on the expression levels of other genes

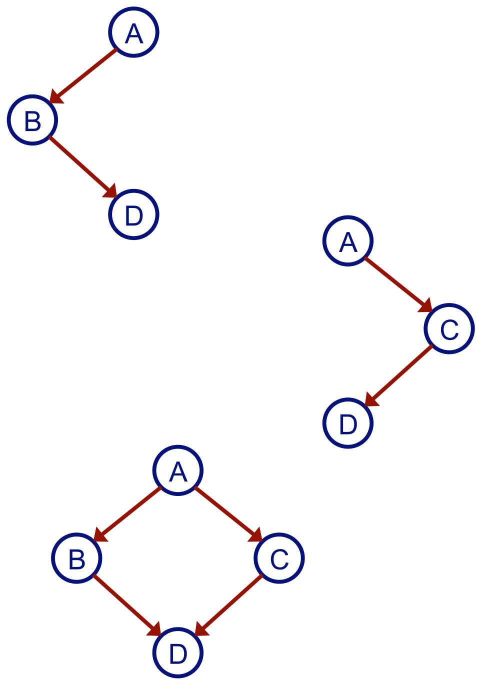


Physical Network Model Example

- suppose we knock out gene A and observe that the expression of gene D goes down



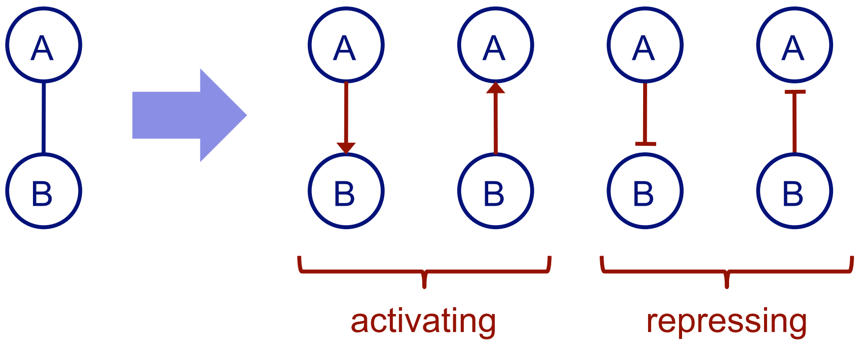
- there are three possible explanations for this causal relationship



Physical Network Models

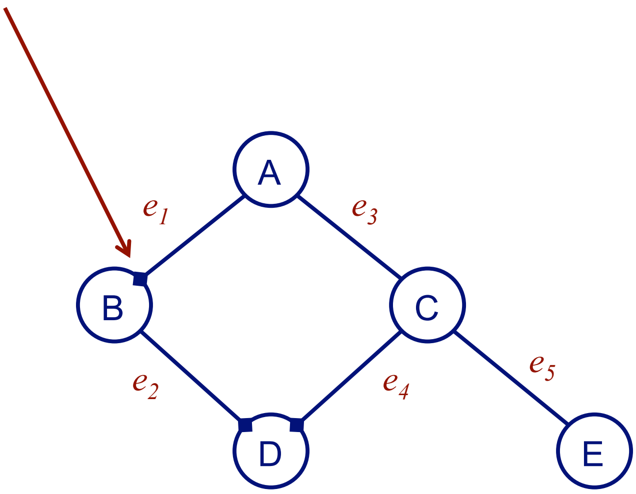
- thus for each edge we are uncertain about it's *presence* (is the interaction involved in the condition being analyzed)
- for some edges we also have uncertainty about
  - *direction*
  - *sign* (is the interaction activating or repressing)

e.g. 4 ways this interaction could be interpreted in explanations



Physical Network Models

to indicate edges for which we know the direction, but not the sign, we'll use arrowheads like this



Physical Network Models

separate binary variables represent edge

- presence

$$X_{\vec{E}_G} = \{x_{e_1}, x_{e_2}, x_{e_4}\}$$

$$X_{\overline{E}_G} = \{x_{e_3}, x_{e_5}\}$$

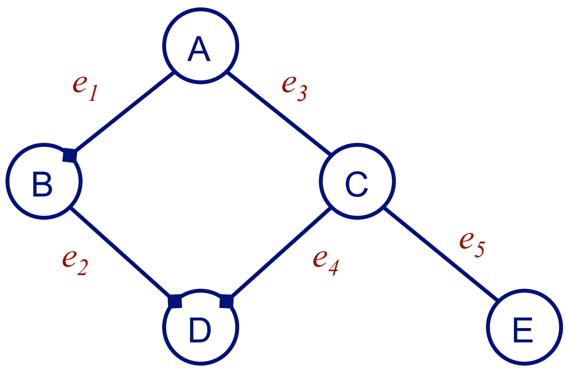
- direction

$$D_{\overline{E}_G} = \{d_{e_3}, d_{e_5}\}$$

- sign

$$S_{\vec{E}_G} = \{s_{e_1}, s_{e_2}, s_{e_4}\}$$

$$S_{\overline{E}_G} = \{s_{e_3}, s_{e_5}\}$$



Inference in Physical Network Models

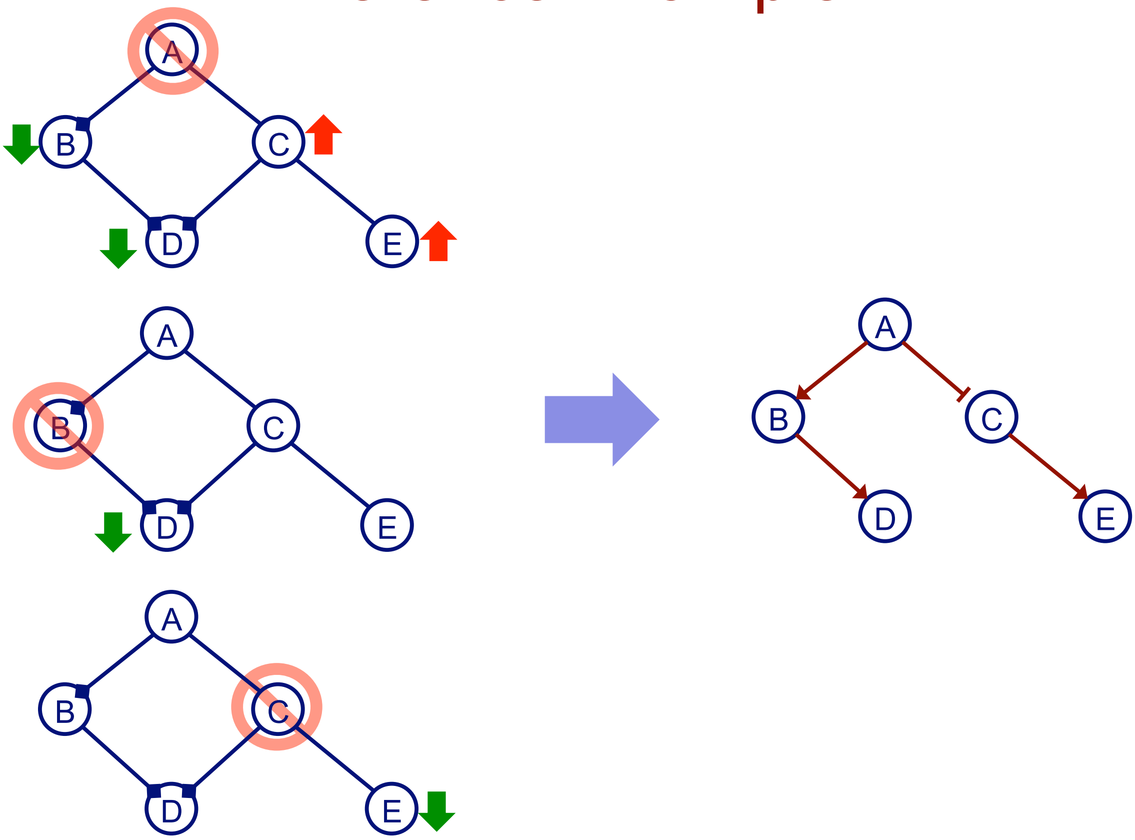
given

- a set of knockout experiments in some condition of interest

infer

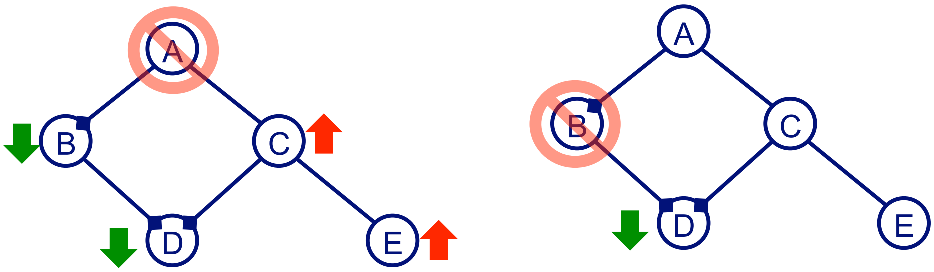
- an assignment of values to the variables such that they provide a coherent explanation for the experiments

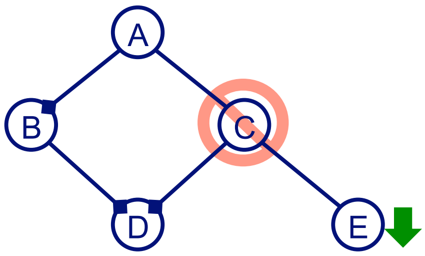
Inference Example



Physical Network Models

- let K_p be the set of *significant* knockout effects
- for the following experiment effects, we have…





$K_p = \{$(A,B), (A,C), (A,D), (A,E), (B, D), (C, E)$\}$

Physical Network Models

additional variables

- set of *actual* knockout effects

$$K = \{k_{ij} : (i,j) \in K_p\}$$

$$k_{ij} = \begin{cases} +1 & \text{if knocking out } i \text{ causes } j \text{ to go up} \\ 0 & \text{if knocking out } i \text{ has no effect on } j \\ -1 & \text{if knocking out } i \text{ causes } j \text{ to go down} \end{cases}$$

- measurements of gene expression in knockouts (think of o_{ij} as a noisy measurement of k_{ij})

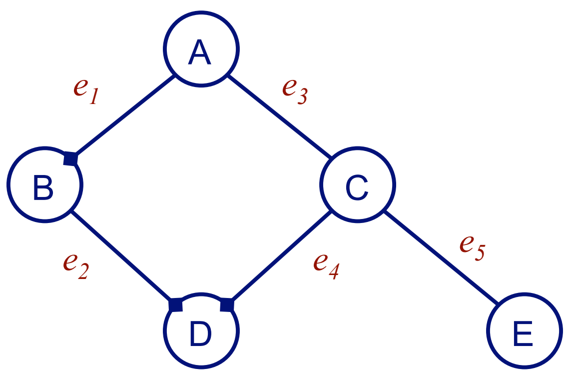
$$O = \{o_{ij} : (i,j) \in K_p\}$$

Physical Network Models

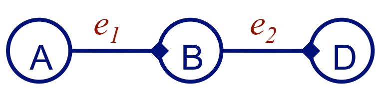
additional variables

- path selection variables

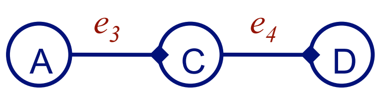
$$\Sigma = \{\sigma_{ija} : (i,j) \in K_p \text{ sth } a \text{ is a valid path btwn } i \text{ and } j\}$$



σ_{AD1} =



σ_{AD2} =



Inference in Physical Network Models

- inference task: determine assignment of values to the X, S, D variables to maximize

$$P(X_{\vec{E}_G}, S_{\vec{E}_G}, X_{\bar{E}_G}, S_{\bar{E}_G}, D_{\bar{E}_G}, K, \Sigma \mid Y_{\vec{E}_G}, Y_{\bar{E}_G}, O_K)$$

- Yeang et al. use an *undirected graphical model* approach to do this inference

Inference in Physical Network Models

$$P(X_{\vec{E}_G}, S_{\vec{E}_G}, X_{\bar{E}_G}, S_{\bar{E}_G}, D_{\bar{E}_G}, K, \Sigma \mid Y_{\vec{E}_G}, Y_{\bar{E}_G}, O_K) \propto$$

$$\prod_{\vec{e}_i \in \vec{E}_G} \phi_{\vec{e}_i}\left(x_{\vec{e}_i}; y_{\vec{e}_i}\right) \times$$

$$\prod_{\bar{e}_i \in \bar{E}_G} \phi_{\bar{e}_i}\left(x_{\bar{e}_i}; y_{\bar{e}_i}\right) \times$$

$$\prod_{(i,j) \in K_p} \phi_{ij}\left(k_{ij}; o_{ij}\right) \psi_{ij}\left(X_{ij}, D_{ij}, S_{ij}, \Sigma_{ij}, k_{ij}\right)$$

The Model Potentials

- three of the potentials relate values of variables to their corresponding measurements

$$\phi_{\bar{e}_i}\left(x_{\bar{e}_i};\, y_{\bar{e}_i}\right) \qquad \phi_{\vec{e}_i}\left(x_{\vec{e}_i};\, y_{\vec{e}_i}\right) \qquad \phi_{ij}\left(k_{ij};\, o_{ij}\right)$$

- these potential functions have the general form

$$\phi_{\bar{e}_i}\left(x_{\bar{e}_i};\, y_{\bar{e}_i}\right) = \left[\frac{P(y_{\bar{e}_i} \mid x_{\bar{e}_i} = 1)}{P(y_{\bar{e}_i} \mid x_{\bar{e}_i} = 0)}\right]^{x_{\bar{e}_i}}$$

The Model Potentials

- consider the potential for a <u>single</u> path, whose set of edges is denoted by E_a

$$\psi_{ija}\left(X_a, D_a, S_a, k_{ij}\right) =$$

$$\prod_{e \in E_a} I(x_e = 1) \times$$

are all edges present?

$$I\left(\prod_{e \in E_a} s_e = -k_{ij}\right) \times$$

are the signs of the edges consistent with the KO effect?

$$\prod_{e \in \bar{E}_a} I(d_e = \hat{d}_e)$$

are undirected edges going in the right direction?

- $I(\bullet)$ is an indicator function returning 1 if the given condition is satisfied, 0 otherwise

The Model Potentials

- now take into account the path selection variable

$$\psi_{ija}\left(X_a, D_a, S_a, k_{ij}, \sigma_{ija}\right) =$$
$$\varepsilon +$$
$$(1-\varepsilon)\ I(\sigma_{ija} = 1)\ \psi_{ija}\left(X_a, D_a, S_a, k_{ij}\right)$$

- because our knowledge of interactions is incomplete, have the potential be nonzero even when the path is not selected

The Model Potentials

- now specify the (soft) condition that at least on candidate path is selected to explain k_{ij}

$$\psi_{ij}\left(X_{ij}, D_{ij}, S_{ij}, \Sigma_{ija}, k_{ij}\right) =$$
$$\left[\varepsilon + (1-\varepsilon)\left(1 - \prod_a I(\sigma_{ija} = 0)\right)\right] \times$$
$$\prod_a \psi_{ija}\left(X_a, D_a, S_a, \sigma_{ija}, k_{ij}\right)$$

The Model Potentials Summarized

- experimental noise in measurements is taken into account
- each knockout effect is explained by at least one path
- to explain an effect, the edges of a path must
  - be "present"
  - be going in the the right direction
  - have their signs be consistent with the effect
- also the path must
  - be shorter than some predefined upper bound
  - end in a protein-DNA interaction

Empirical Evaluation

- mating response pathway experiment
  - 149 knockout effect pairs from 13 experiments
  - 106 pairs in 9 experiments are connected by paths <= 5
- run inference procedure on all data
  - all effects are explained by solution found
  - only 21 effects are trivially explained by direct protein-DNA binding
- leave out some effects in a cross-validation experiment; see what error rate is in predicting effects

TABLE 2. CROSS VALIDATION ON KNOCK-OUT PAIRS

# hold-outs	# trials	% error
1	106	2.83%
5	200	3.5%
20	200	5.9%

Discussion

- approach looks for consistent explanations for cases in which we have
  - “known” interactions among genes
  - measurements resulting from knocking out/down selected genes
- predictive accuracy is high for a small, densely connected network
- incomplete knowledge of interactions may make approach less viable in larger networks
  - only 1,091 out of 23,766 pairs are connected by valid paths in a genome-wide data set