

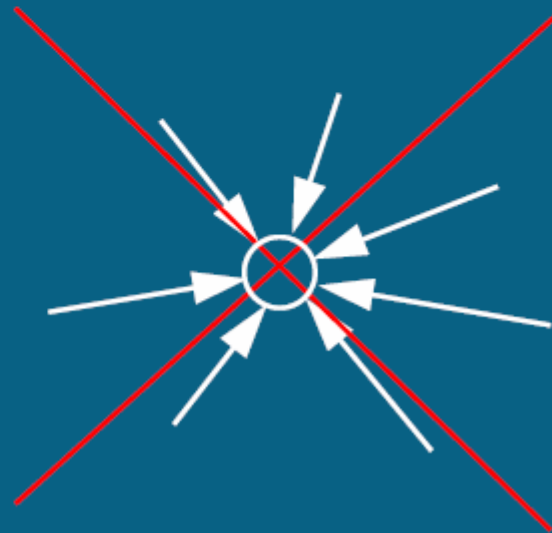
Reconstructing gene regulatory networks
with Bayesian networks by combining
gene expression profiles with multiple
sources of prior knowledge

Dirk Husmeier

Fan-out unrestricted



Fan-in restricted



not permissible

Statistical Applications in Genetics and Molecular Biology

Volume 6, Issue 1

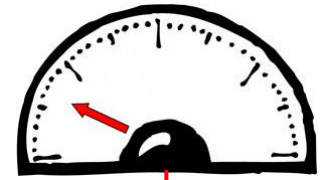
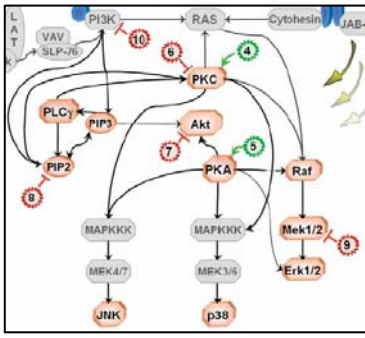
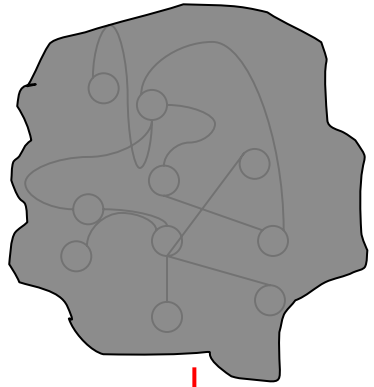
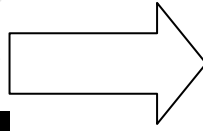
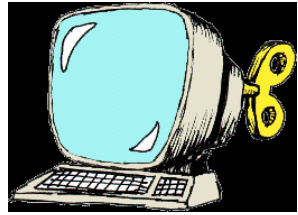
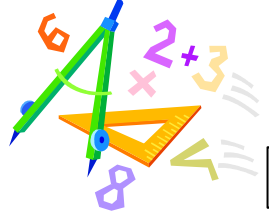
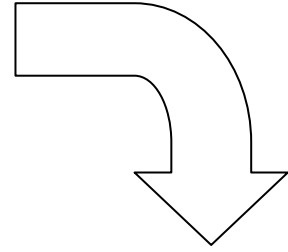
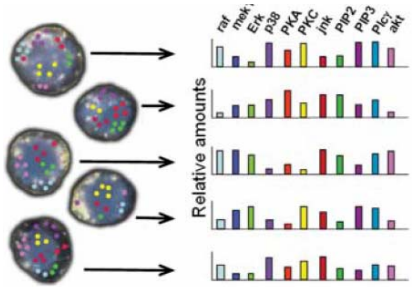
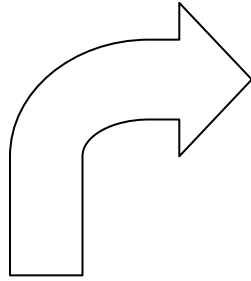
2007

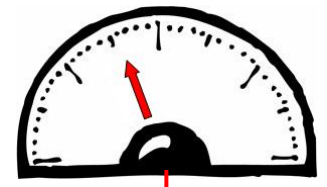
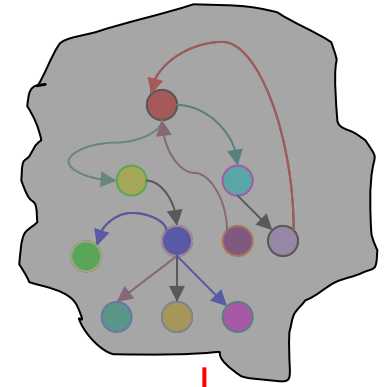
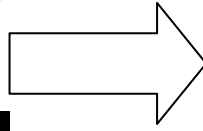
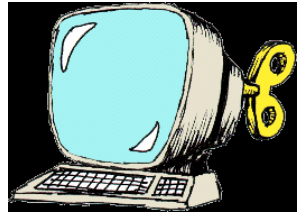
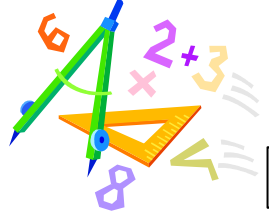
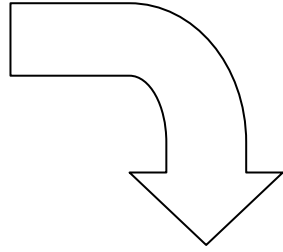
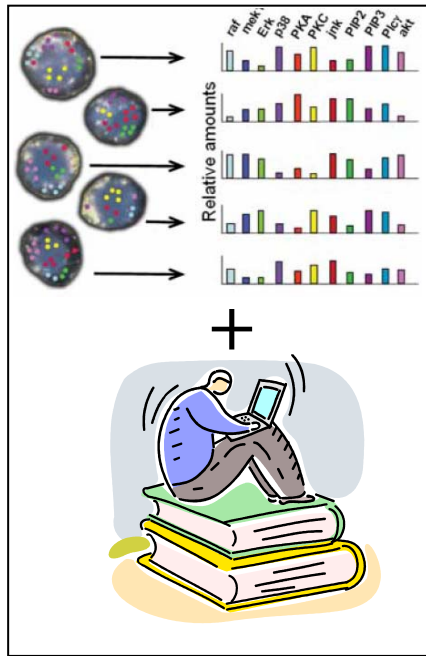
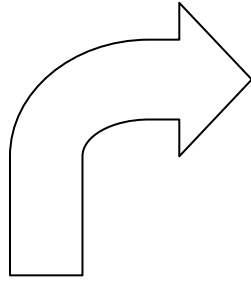
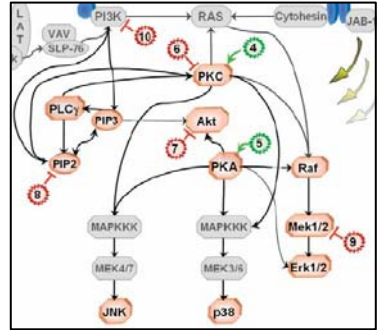
Article 15

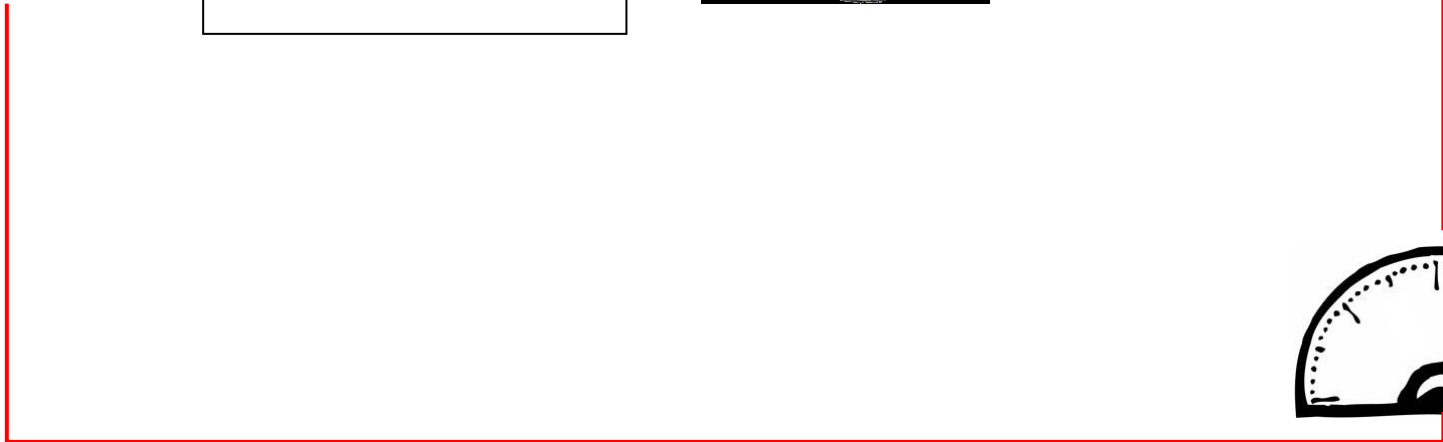
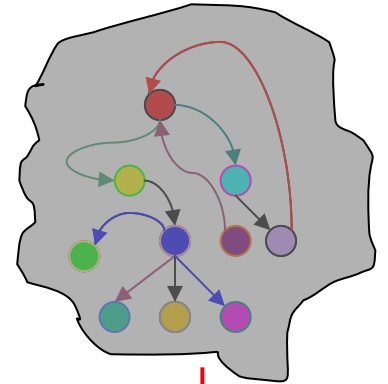
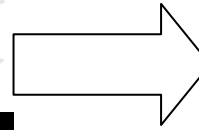
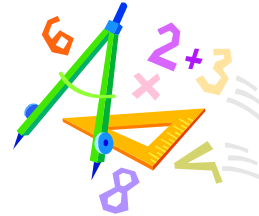
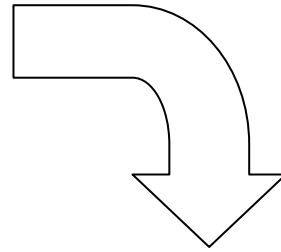
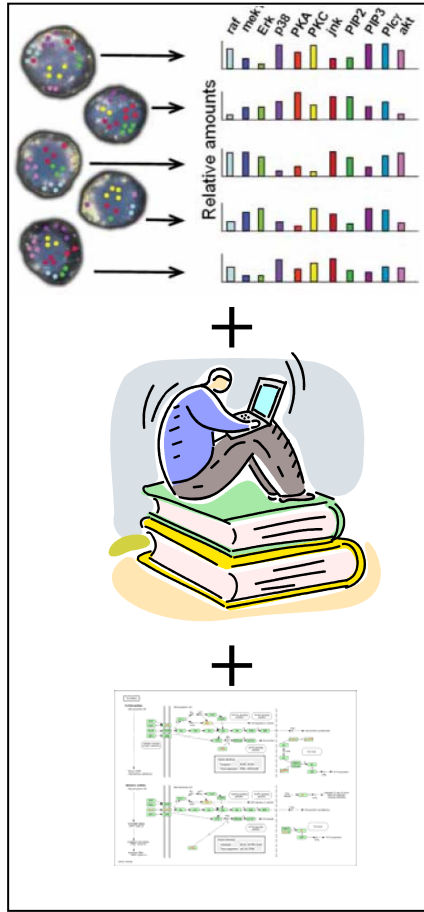
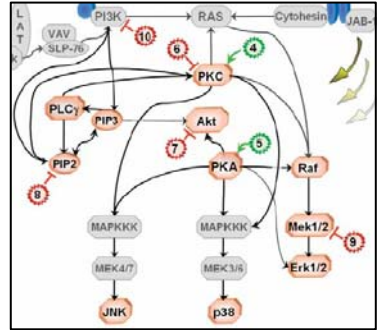
Reconstructing Gene Regulatory Networks with Bayesian Networks by Combining Expression Data with Multiple Sources of Prior Knowledge

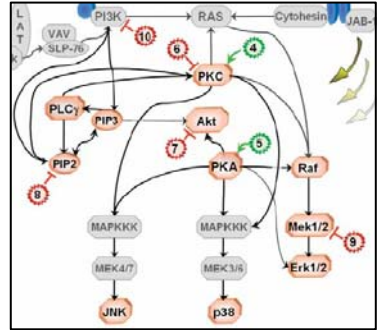
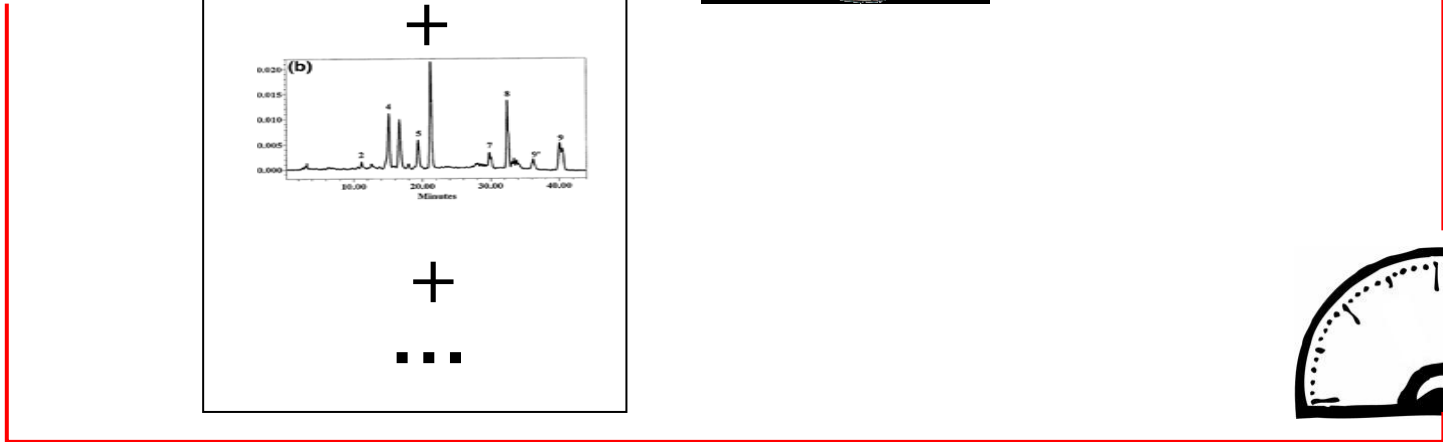
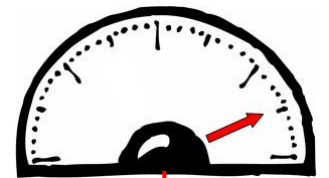
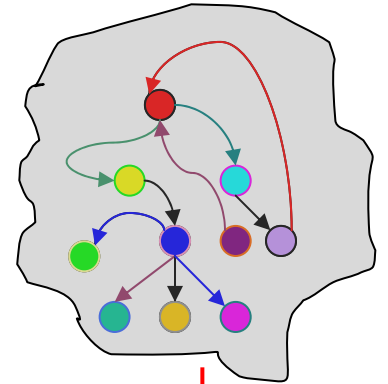
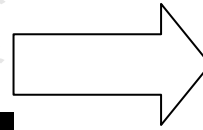
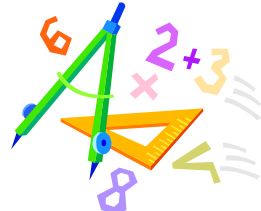
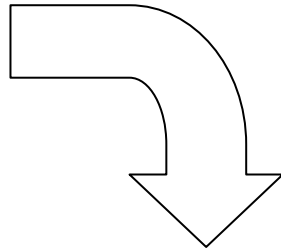
Adriano V. Werhli*

Dirk Husmeier†







Bayesian inference

Select the model \mathcal{M} based on the posterior probability:

$$P(\mathcal{M}|\mathcal{D}) \propto P(\mathcal{D}|\mathcal{M})P(\mathcal{M})$$

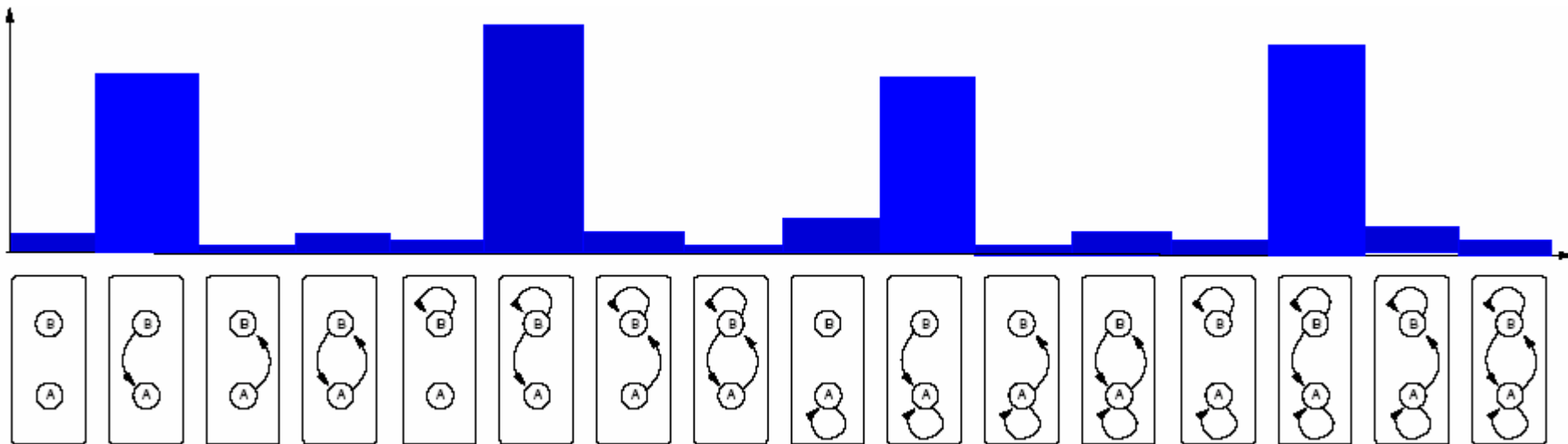
This requires an integration over the whole parameter space:

$$P(\mathcal{D}|\mathcal{M}) = \int P(\mathcal{D}|\mathbf{q}, \mathcal{M})P(\mathbf{q}|\mathcal{M})d\mathbf{q}$$

Uncertainty about the best network structure

Limited number of experimental replications,
high noise

$$P(\mathcal{D}|\mathcal{M})$$

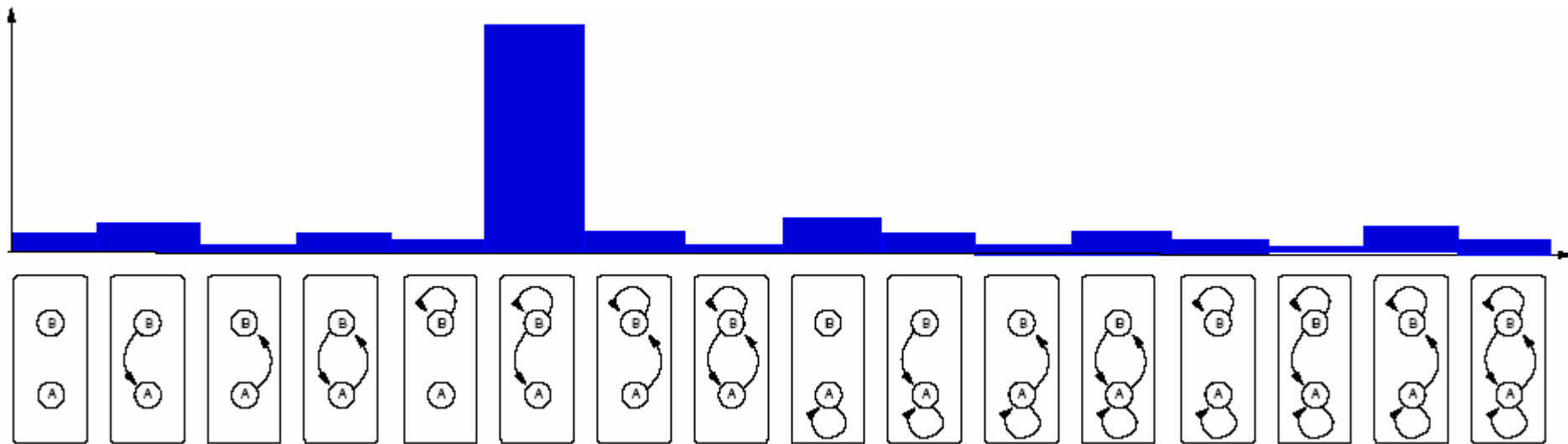


Reduced uncertainty by using prior knowledge

Data

Prior knowledge

$$P(\mathcal{M}|\mathcal{D}) \propto P(\mathcal{D}|\mathcal{M})P(\mathcal{M})$$



- Which sources of prior knowledge are reliable?
- How do we trade off the different sources of prior knowledge against each other and against the data?



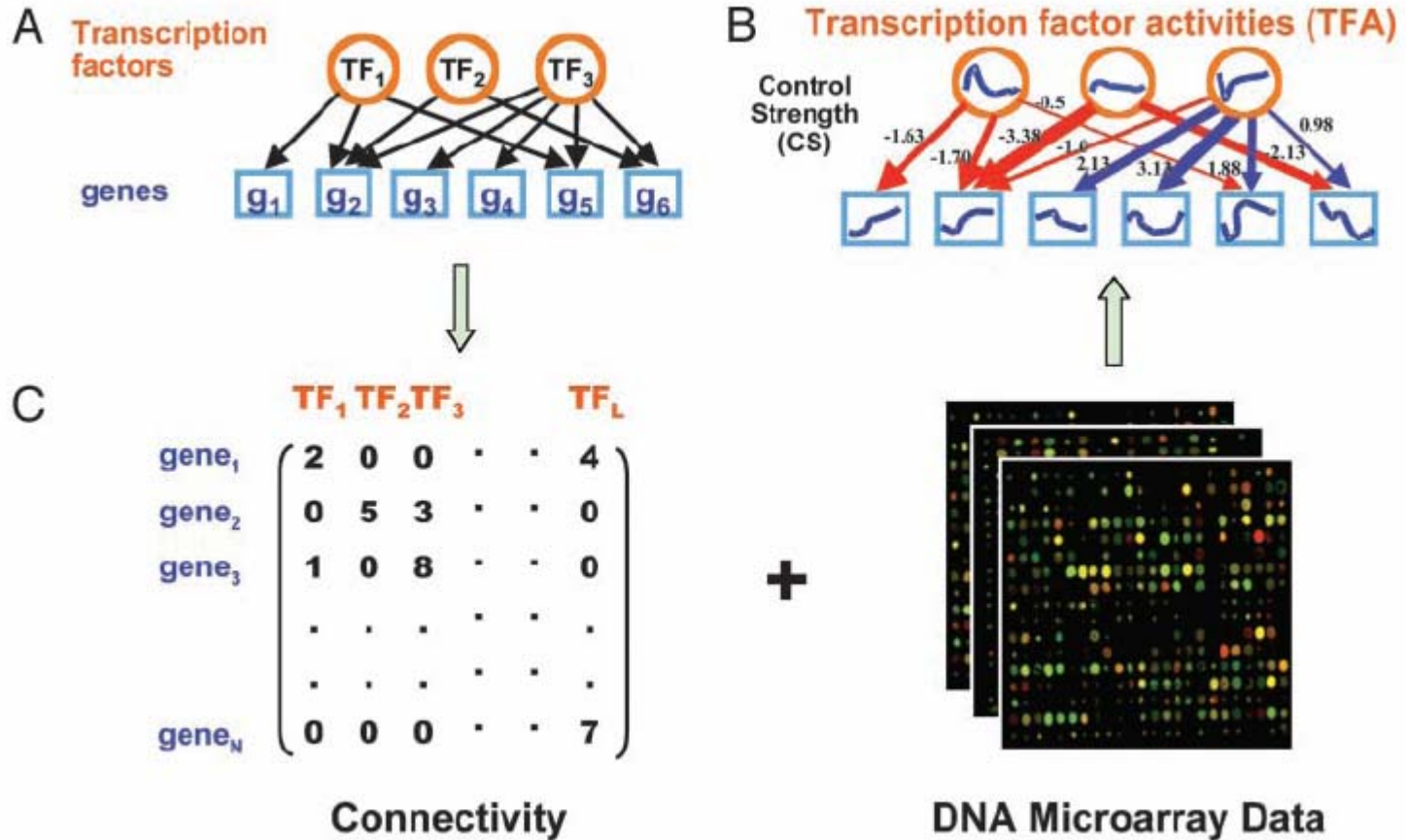
Estimating gene networks from gene expression data by combining Bayesian network model with promoter element detection

Yoshinori Tamada^{1,†}, SunYong Kim^{1,†}, Hideo Bannai¹,
Seiya Imoto¹, Kousuke Tashiro², Satoru Kuhara² and
Satoru Miyano¹*

¹Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan and ²Graduate School of Genetic Resource Technology, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka, 812-8581, Japan

Received on March 17, 2003; accepted on June 9, 2003

Use TF binding motifs in promoter sequences



Biological Prior Knowledge

Biological prior knowledge matrix

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \dots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \dots & p_{2n} \\ p_{31} & p_{32} & p_{33} & \dots & p_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & \dots & p_{nn} \end{pmatrix}$$

$$0 \leq p_{ij} \leq 1$$

p_{ij} Indicates some knowledge about the relationship between genes i and j

Define the energy of a Graph G

$$G = \begin{pmatrix} g_{11} & g_{12} & g_{13} & \dots & g_{1n} \\ g_{21} & g_{22} & g_{23} & \dots & g_{2n} \\ g_{31} & g_{32} & g_{33} & \dots & g_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & g_{n3} & \dots & g_{nn} \end{pmatrix}$$

$$g_{ij} \in \{0, 1\}$$

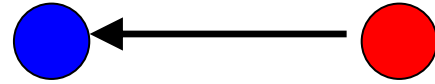
$$E(G) = \sum_{i,j=1}^n |P_{i,j} - G_{i,j}|$$

Prior knowledge

“I don’t know”



“I’m sure
about this”



“Prior probabilities”

$P=0.5$

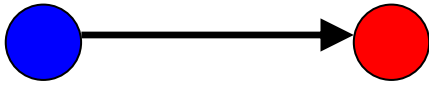


$P=1.0$

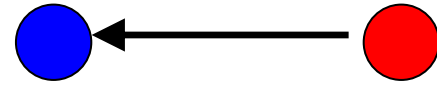


“Prior probabilities”

P=0.5



P=1.0



Network structures for independent edges



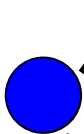
P=0



P=0



P=0.5

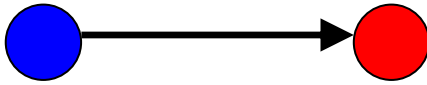


P=0.5

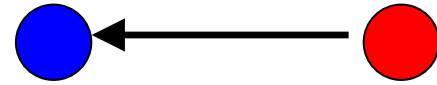
$$\sum P = 1$$

“Prior probabilities”

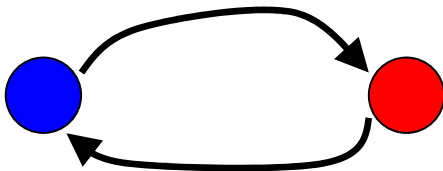
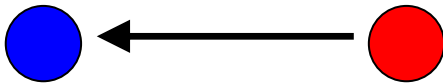
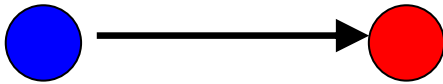
$P=0.5$



$P=1.0$



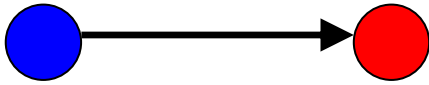
DAG structures



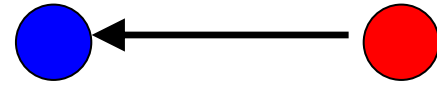
Invalid

“Prior probabilities”

$P=0.5$



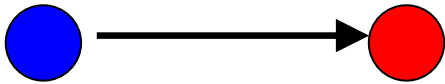
$P=1.0$



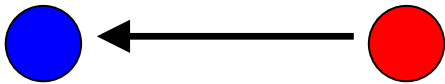
DAG structures



$P=0$



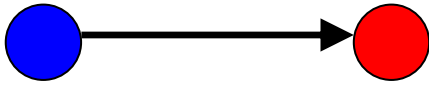
$P=0.5$



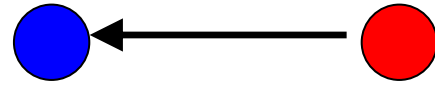
$P=1.0$

“Prior probabilities”

P=0.5



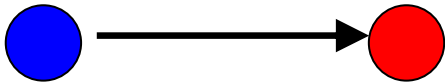
P=1.0



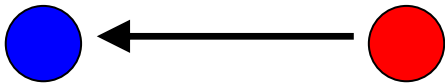
DAG structures



P=0



P=0.5



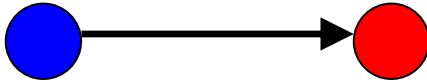
P=1.0

$$\sum P = 1.5 \neq 1.0$$

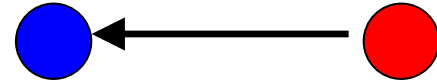
Not a probability distribution

“Prior probabilities”

P=0.5



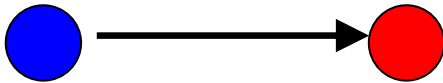
P=1.0



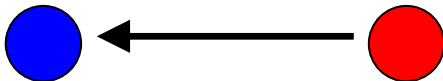
DAG structures



P=0



P=0.5



P=1.0

$$\sum P = 1.5 \neq 1.0$$

Not a probability distribution

Obtaining a proper probability distribution in DAG space requires a renormalization

Notation

- Prior knowledge matrix:
 $P \rightarrow B$ (for “belief”)
- Network structure:
 G (for “graph”) or M (for “model”)
- P : Probabilities

Deviation between the network G
and the prior knowledge B :

$$E(G) = \sum_{i,j=1}^N |B_{i,j} - G_{i,j}|$$

“Energy”

Graph: $\epsilon \in \{0,1\}$

Prior knowledge: $\epsilon \in [0,1]$

Prior distribution over networks

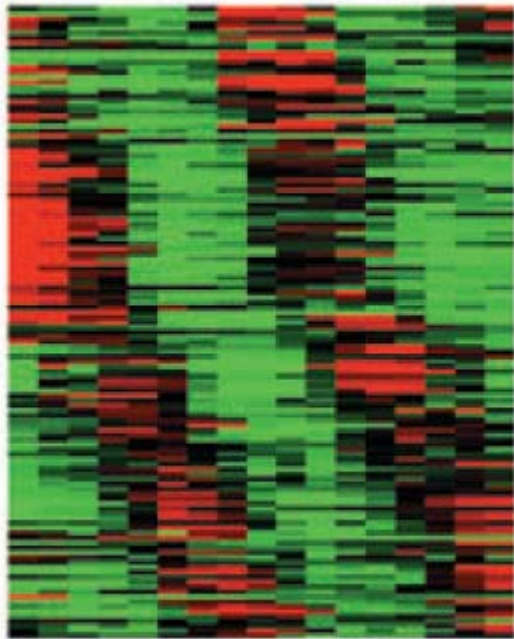
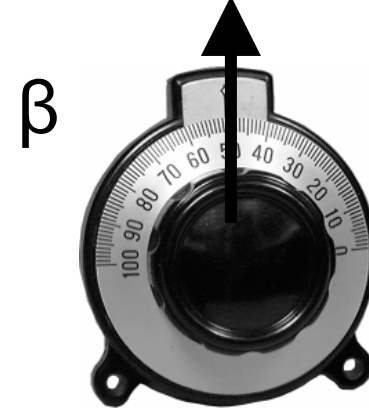
$$P(G|\beta) = \frac{e^{-\beta E(G)}}{Z(\beta)}$$

Hyperparameter

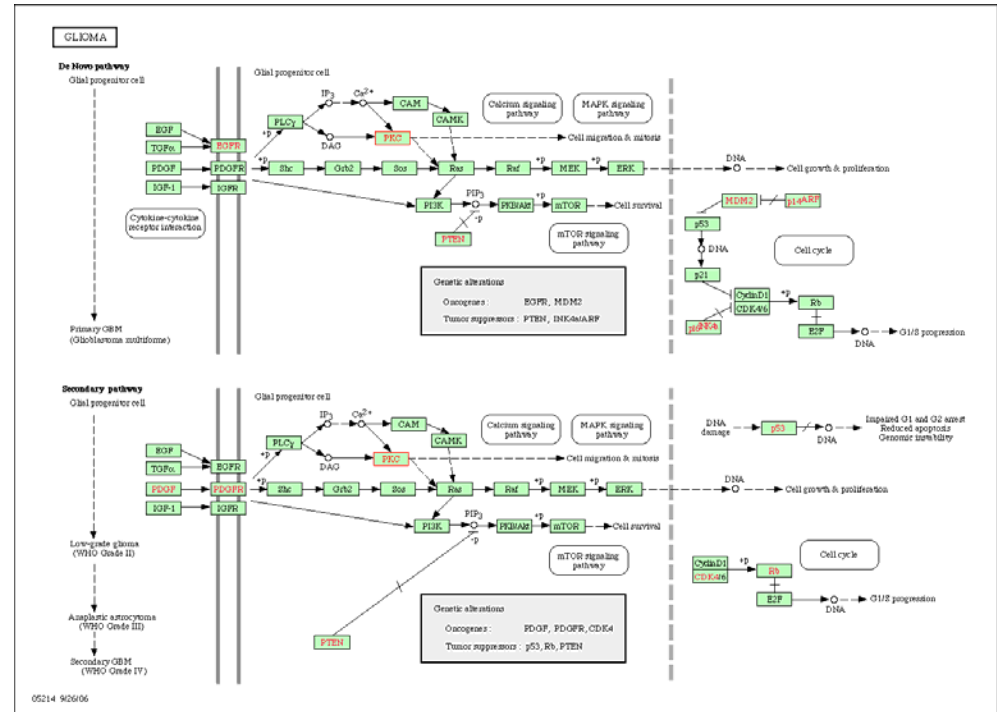
$$Z(\beta) = \sum_{G \in \mathcal{G}} e^{-\beta E(G)}$$

Bayesian analysis: integration of prior knowledge

Hyperparameter β trades off data
versus prior knowledge

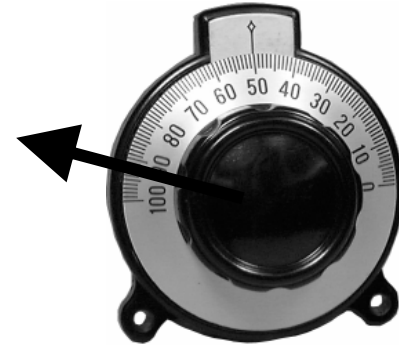


Microarray data

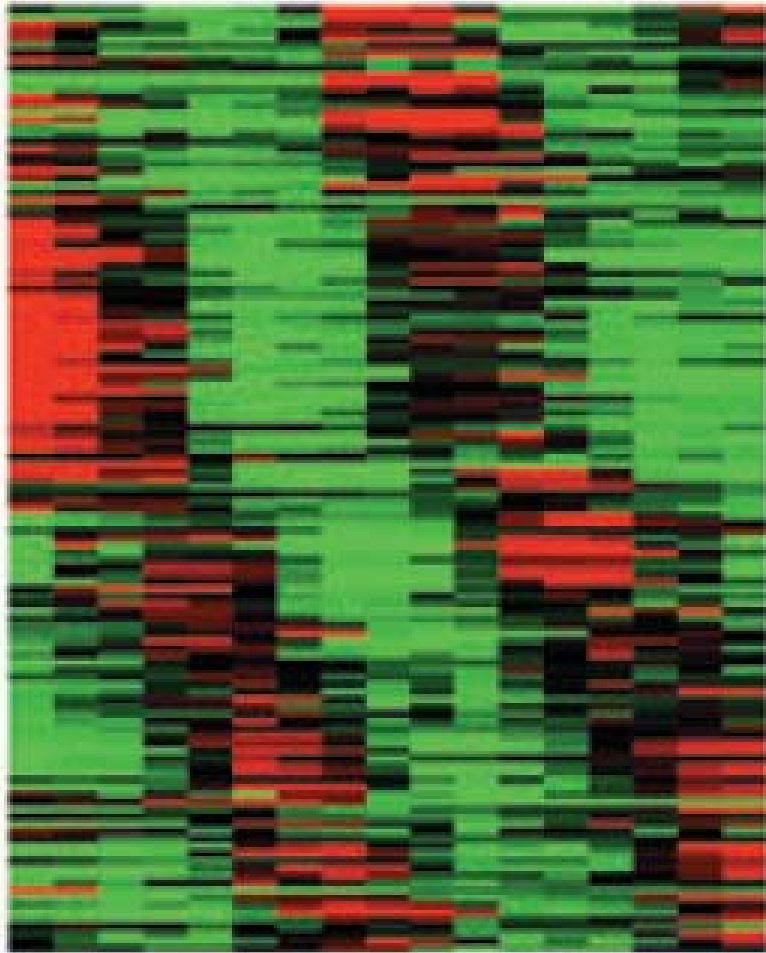


KEGG pathway

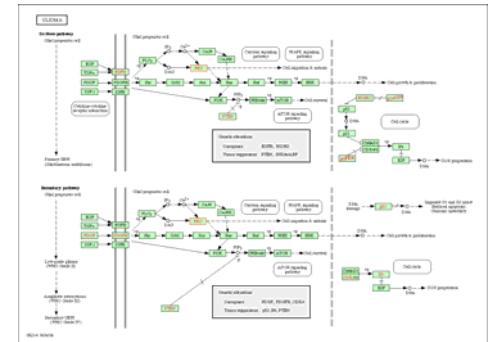
Hyperparameter β trades off data versus prior knowledge



β small

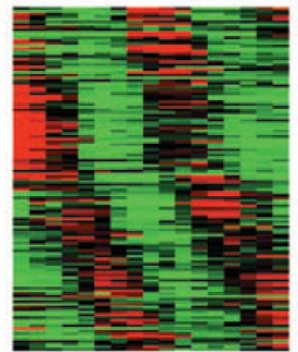
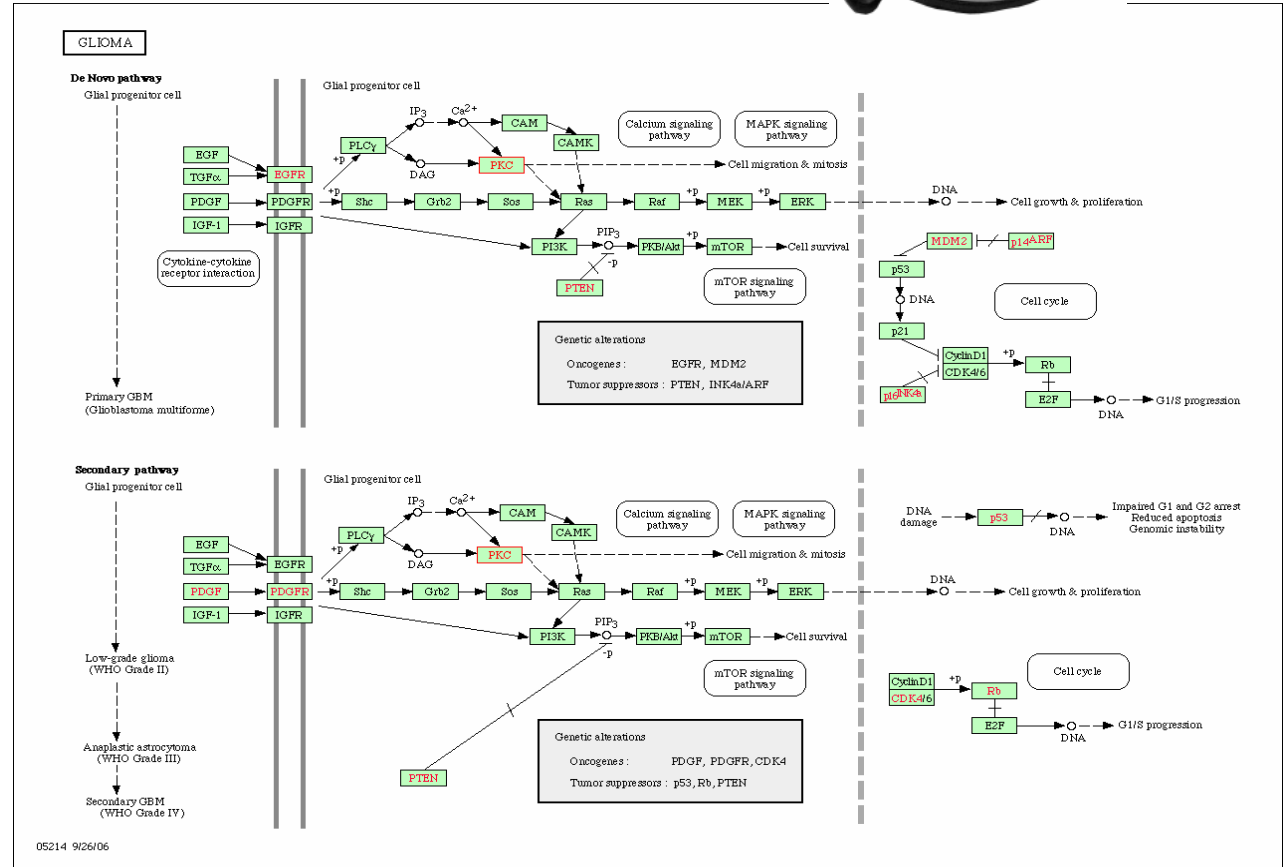
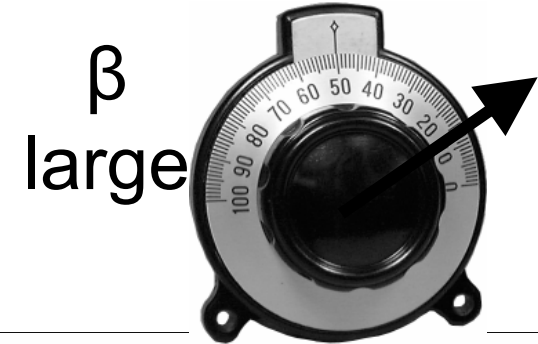


Microarray data



KEGG pathway

Hyperparameter β trades off data versus prior knowledge



Microarray data

KEGG pathway

New contribution

- Generalisation to **more sources** of prior knowledge
- **Inferring** the **hyperparameters**
- Bayesian approach

Multiple sources of prior knowledge

$$E_1(G) = \sum_{i,j=1}^N |B_{i,j}^1 - G_{i,j}|$$

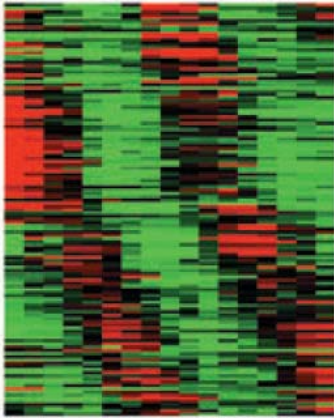
$$E_2(G) = \sum_{i,j=1}^N |B_{i,j}^2 - G_{i,j}|$$

$$P(G|\beta_1, \beta_2) = \frac{e^{-\{\beta_1 E_1(G) + \beta_2 E_2(G)\}}}{Z(\beta_1, \beta_2)}$$

$$Z(\beta_1, \beta_2) = \sum_{G \in \mathcal{G}} e^{-\{\beta_1 E_1(G) + \beta_2 E_2(G)\}}$$

Input:

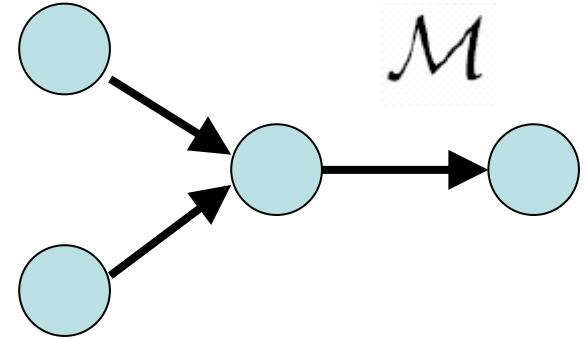
D



MCMC

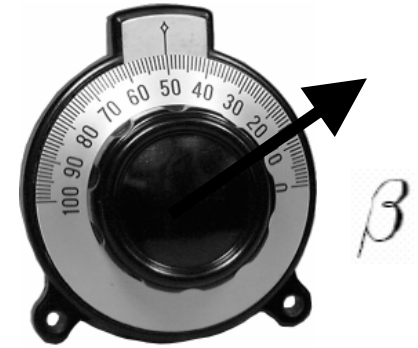
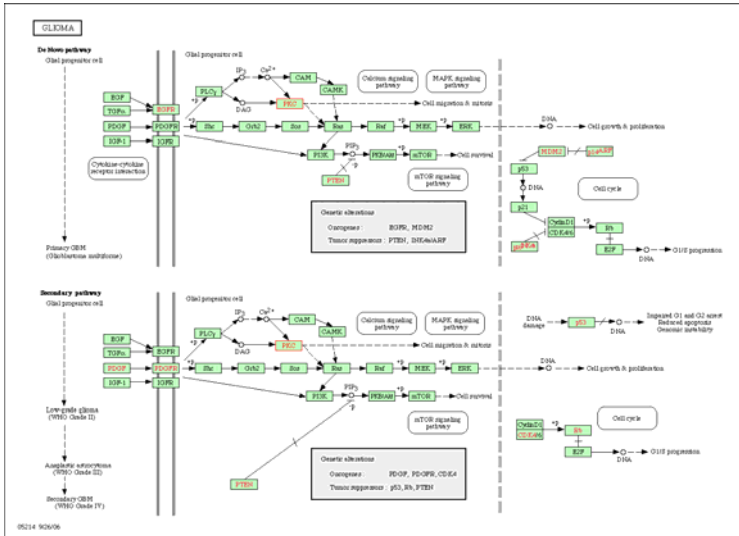


Learn:



M

B



β

$$\beta, M \sim P(\beta, M | D, B)$$

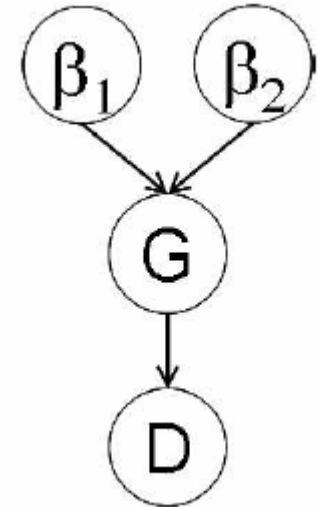
Sample networks and hyperparameters from the posterior distribution

$$P(G, \beta_1, \beta_2 | D)$$

Proposal probabilities

$$Q(G_{\text{new}} | G_{\text{old}})$$

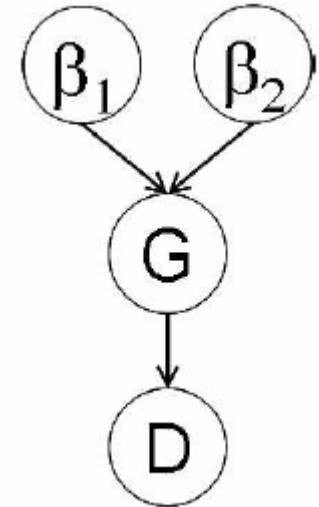
$$R(\beta_{1_{\text{new}}} | \beta_{1_{\text{old}}}) \quad R(\beta_{2_{\text{new}}} | \beta_{2_{\text{old}}})$$



Metropolis-Hastings scheme

Sample networks and hyperparameters from the posterior distribution

$$P(G, \beta_1, \beta_2 | D)$$



Proposal probabilities

$$Q(G_{\text{new}} | G_{\text{old}})$$

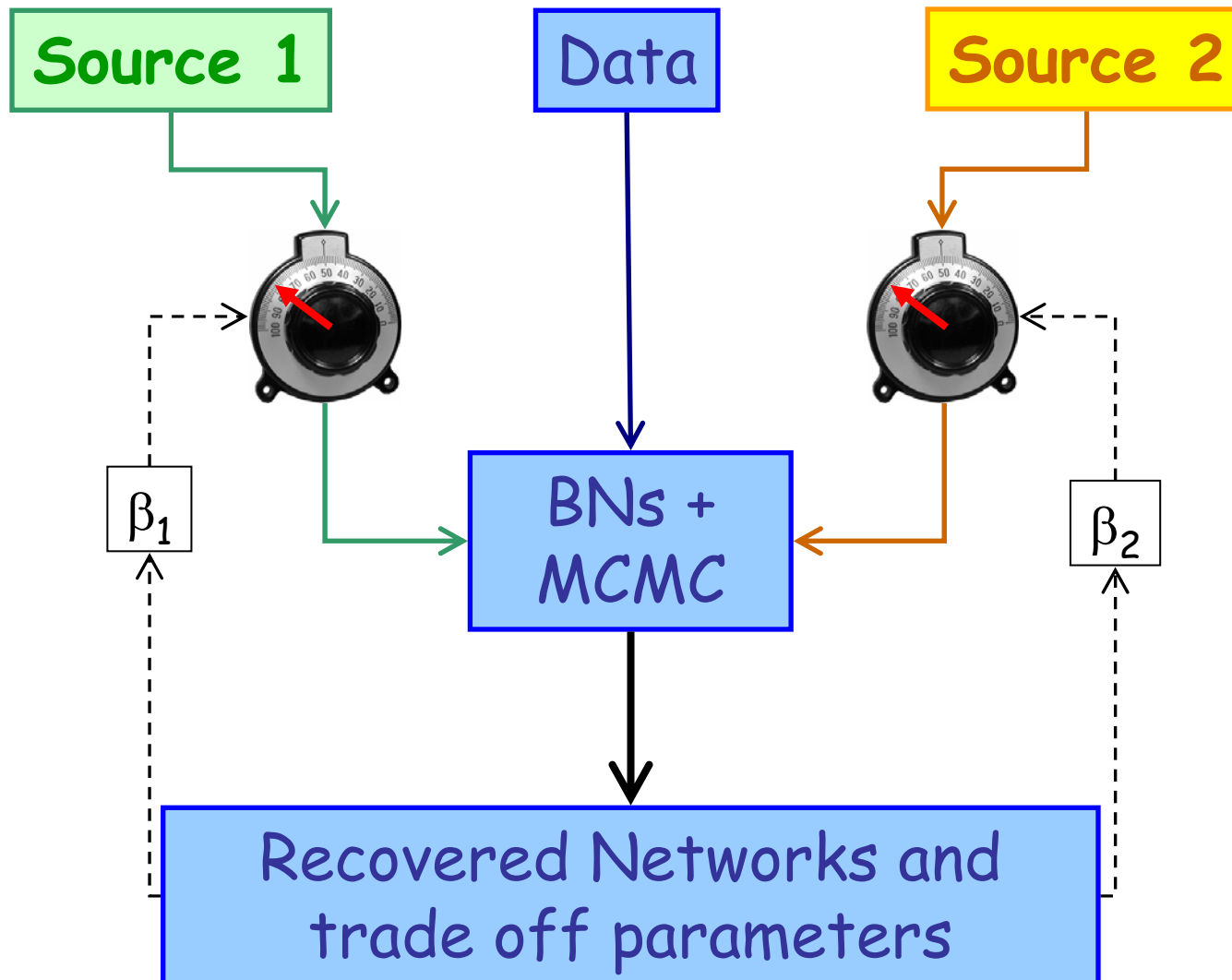
$$R(\beta_{1_{\text{new}}} | \beta_{1_{\text{old}}}) \quad R(\beta_{2_{\text{new}}} | \beta_{2_{\text{old}}})$$

Metropolis-Hastings scheme

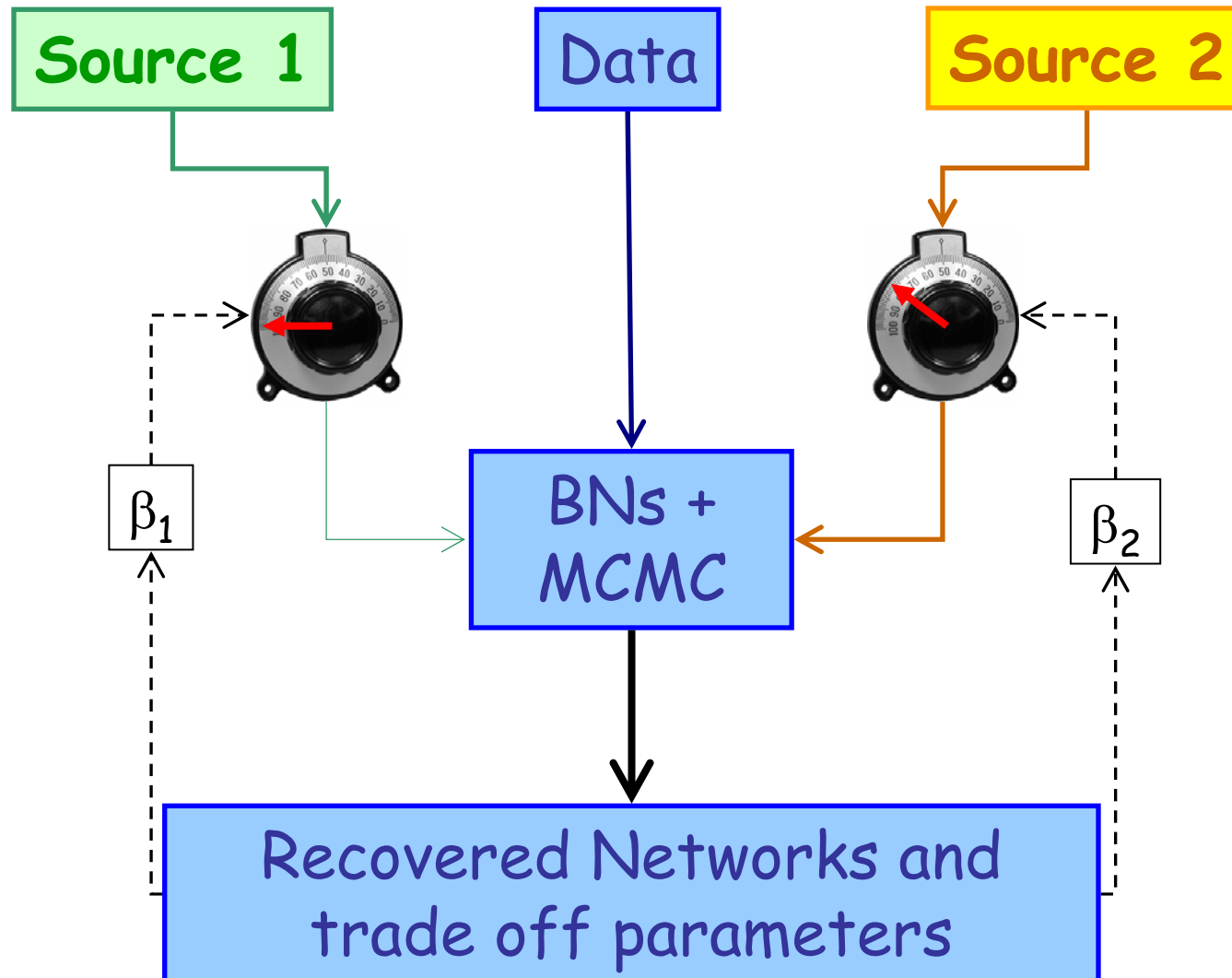
$$A = \min \left\{ \frac{P(D, G_{\text{new}}, \beta_{1_{\text{new}}}, \beta_{2_{\text{new}}}) Q(G_{\text{old}} | G_{\text{new}}) R(\beta_{1_{\text{old}}} | \beta_{1_{\text{new}}}) R(\beta_{2_{\text{old}}} | \beta_{2_{\text{new}}})}{P(D, G_{\text{old}}, \beta_{1_{\text{old}}}, \beta_{2_{\text{old}}}) Q(G_{\text{new}} | G_{\text{old}}) R(\beta_{1_{\text{new}}} | \beta_{1_{\text{old}}}) R(\beta_{2_{\text{new}}} | \beta_{2_{\text{old}}})}, 1 \right\}$$

$$A = \min \left\{ \frac{P(\mathcal{D} | \mathcal{G}_{\text{new}}) P(\mathcal{G}_{\text{new}} | \beta_{1_{\text{new}}}, \beta_{2_{\text{new}}}) P_1(\beta_{1_{\text{new}}}) P_2(\beta_{2_{\text{new}}})}{P(\mathcal{D} | \mathcal{G}_{\text{old}}) P(\mathcal{G}_{\text{old}} | \beta_{1_{\text{old}}}, \beta_{2_{\text{old}}}) P_1(\beta_{1_{\text{old}}}) P_2(\beta_{2_{\text{old}}})} \times \frac{Q(\mathcal{G}_{\text{old}} | \mathcal{G}_{\text{new}}) R_1(\beta_{1_{\text{old}}} | \beta_{1_{\text{new}}}) R_2(\beta_{2_{\text{old}}} | \beta_{2_{\text{new}}})}{Q(\mathcal{G}_{\text{new}} | \mathcal{G}_{\text{old}}) R_1(\beta_{1_{\text{new}}} | \beta_{1_{\text{old}}}) R_2(\beta_{2_{\text{new}}} | \beta_{2_{\text{old}}})}, 1 \right\}$$

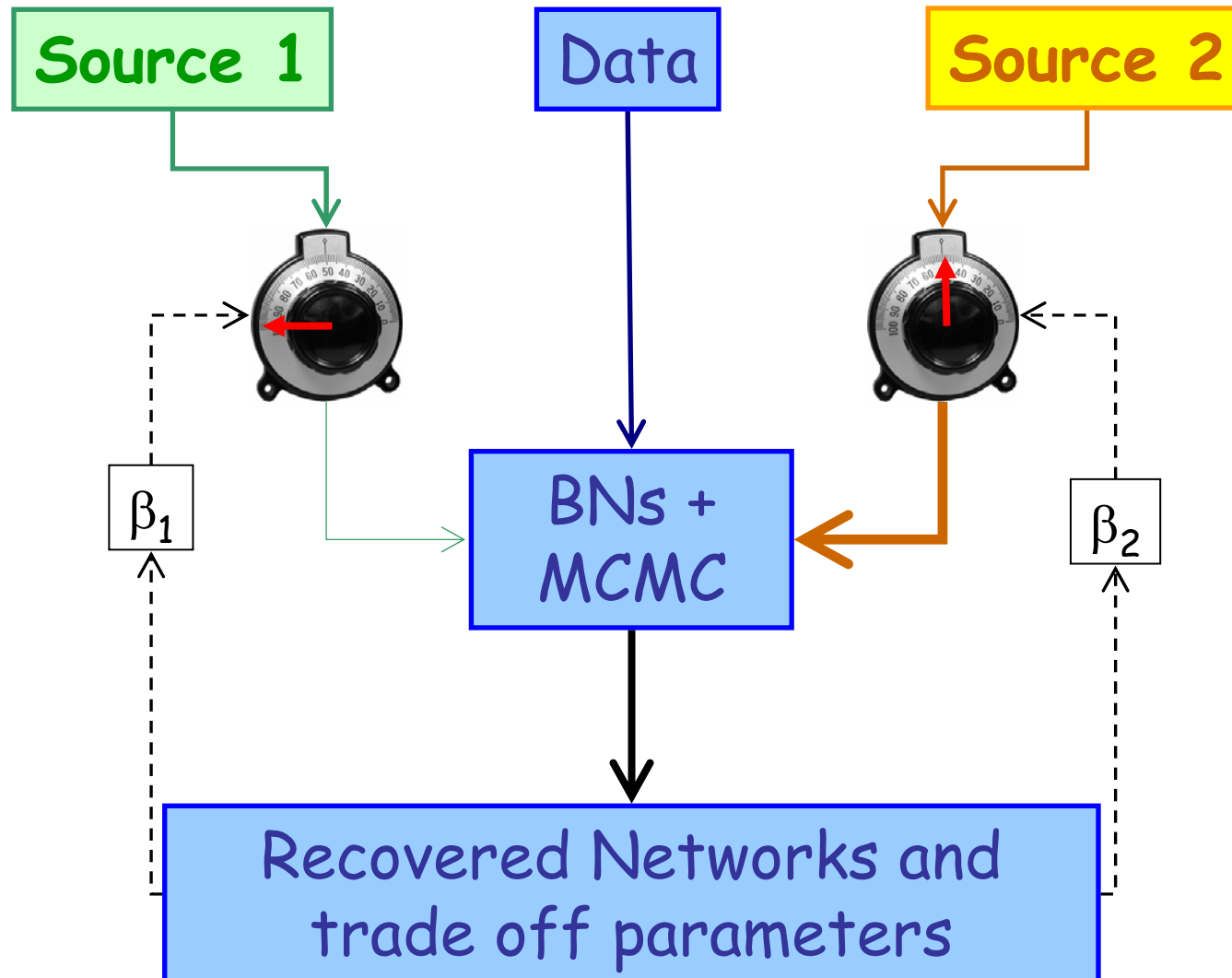
Bayesian networks with two sources of prior



Bayesian networks with two sources of prior

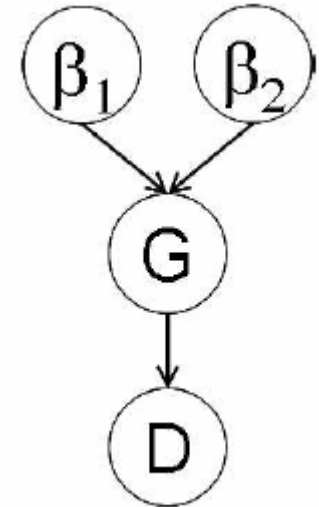


Bayesian networks with two sources of prior



Sample networks and hyperparameters from the posterior distribution

$$P(G, \beta_1, \beta_2 | D)$$



Proposal probabilities

$$Q(G_{\text{new}} | G_{\text{old}})$$

$$R(\beta_{1_{\text{new}}} | \beta_{1_{\text{old}}}) \quad R(\beta_{2_{\text{new}}} | \beta_{2_{\text{old}}})$$

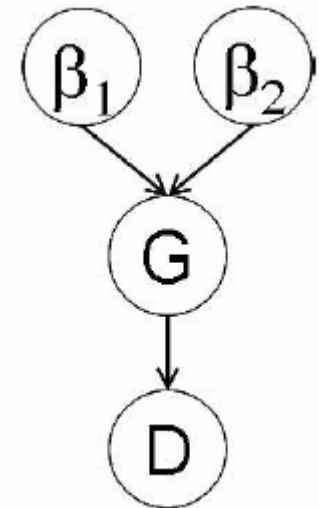
Metropolis-Hastings scheme

$$A = \min \left\{ \frac{P(D, G_{\text{new}}, \beta_{1_{\text{new}}}, \beta_{2_{\text{new}}}) Q(G_{\text{old}} | G_{\text{new}}) R(\beta_{1_{\text{old}}} | \beta_{1_{\text{new}}}) R(\beta_{2_{\text{old}}} | \beta_{2_{\text{new}}})}{P(D, G_{\text{old}}, \beta_{1_{\text{old}}}, \beta_{2_{\text{old}}}) Q(G_{\text{new}} | G_{\text{old}}) R(\beta_{1_{\text{new}}} | \beta_{1_{\text{old}}}) R(\beta_{2_{\text{new}}} | \beta_{2_{\text{old}}})}, 1 \right\}$$

$$A = \min \left\{ \frac{P(D | \mathcal{G}_{\text{new}}) P(\mathcal{G}_{\text{new}} | \beta_{1_{\text{new}}}, \beta_{2_{\text{new}}}) P_1(\beta_{1_{\text{new}}}) P_2(\beta_{2_{\text{new}}})}{P(D | \mathcal{G}_{\text{old}}) P(\mathcal{G}_{\text{old}} | \beta_{1_{\text{old}}}, \beta_{2_{\text{old}}}) P_1(\beta_{1_{\text{old}}}) P_2(\beta_{2_{\text{old}}})} \times \frac{Q(\mathcal{G}_{\text{old}} | \mathcal{G}_{\text{new}}) R_1(\beta_{1_{\text{old}}} | \beta_{1_{\text{new}}}) R_2(\beta_{2_{\text{old}}} | \beta_{2_{\text{new}}})}{Q(\mathcal{G}_{\text{new}} | \mathcal{G}_{\text{old}}) R_1(\beta_{1_{\text{new}}} | \beta_{1_{\text{old}}}) R_2(\beta_{2_{\text{new}}} | \beta_{2_{\text{old}}})}, 1 \right\}$$

Sample networks and hyperparameters from the posterior distribution

$$P(G, \beta_1, \beta_2 | D)$$



Proposal probabilities

$$Q(G_{\text{new}} | G_{\text{old}})$$

$$R(\beta_{1_{\text{new}}} | \beta_{1_{\text{old}}}) \quad R(\beta_{2_{\text{new}}} | \beta_{2_{\text{old}}})$$

Metropolis-Hastings scheme

$$A = \min \left\{ \frac{P(D, G_{\text{new}}, \beta_{1_{\text{new}}}, \beta_{2_{\text{new}}}) Q(G_{\text{old}} | G_{\text{new}}) R(\beta_{1_{\text{old}}} | \beta_{1_{\text{new}}}) R(\beta_{2_{\text{old}}} | \beta_{2_{\text{new}}})}{P(D, G_{\text{old}}, \beta_{1_{\text{old}}}, \beta_{2_{\text{old}}}) Q(G_{\text{new}} | G_{\text{old}}) R(\beta_{1_{\text{new}}} | \beta_{1_{\text{old}}}) R(\beta_{2_{\text{new}}} | \beta_{2_{\text{old}}})}, 1 \right\}$$

$$A = \min \left\{ \frac{P(\mathcal{D} | \mathcal{G}_{\text{new}}) P(\mathcal{G}_{\text{new}} | \beta_{1_{\text{new}}}, \beta_{2_{\text{new}}})}{P(\mathcal{D} | \mathcal{G}_{\text{old}}) P(\mathcal{G}_{\text{old}} | \beta_{1_{\text{old}}}, \beta_{2_{\text{old}}})} \times \right. \\ \left. \frac{Q(\beta_{1_{\text{old}}} | \beta_{1_{\text{new}}}) Q(\beta_{2_{\text{old}}} | \beta_{2_{\text{new}}})}{Q(\beta_{1_{\text{new}}} | \beta_{1_{\text{old}}}) Q(\beta_{2_{\text{new}}} | \beta_{2_{\text{old}}})}, 1 \right\}$$

Prior distribution

$$E_1(G) = \sum_{i,j=1}^N |B_{i,j}^1 - G_{i,j}|$$

$$E_2(G) = \sum_{i,j=1}^N |B_{i,j}^2 - G_{i,j}|$$

$$P(G|\beta_1, \beta_2) = \frac{e^{-\{\beta_1 E_1(G) + \beta_2 E_2(G)\}}}{Z(\beta_1, \beta_2)}$$

$$Z(\beta_1, \beta_2) = \sum_{G \in \mathcal{G}} e^{-\{\beta_1 E_1(G) + \beta_2 E_2(G)\}}$$

Approximation of Z

$$E(G) = \sum_{i,j=1}^N |B_{i,j} - G_{i,j}|$$

Rewriting the energy

$$E(G) = \sum_{n=1}^N \mathcal{E}(n, \pi_n [G])$$

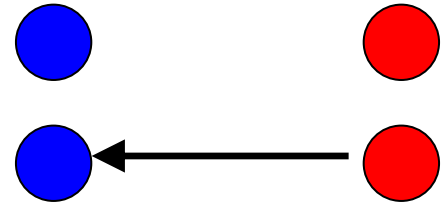
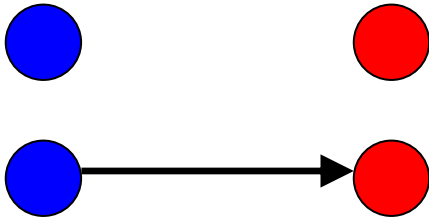
$$\mathcal{E}(n, \pi_n) = \sum_{i \in \pi_n} (1 - B_{in}) + \sum_{i \notin \pi_n} B_{in}$$

Approximation of the partition function

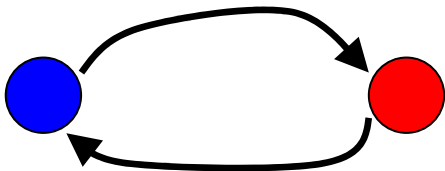
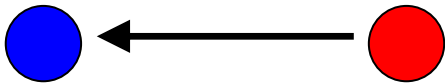
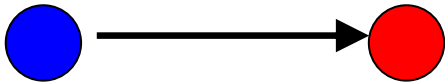
$$\begin{aligned} Z &= \sum_{G \in \mathcal{G}} e^{-\beta E(G)} \\ &= \sum_{\pi_1} \cdots \sum_{\pi_N} e^{-\beta(\mathcal{E}(1, \pi_1) + \dots + \mathcal{E}(N, \pi_N))} \\ &= \prod_n \sum_{\pi_n} e^{-\beta \mathcal{E}(n, \pi_n)} \end{aligned}$$

Partition function of an ideal gas

Individual partition functions



DAG structures



Invalid

Multiple sources of prior knowledge

$$E_1(G) = \sum_{i,j=1}^N |B_{i,j}^1 - G_{i,j}|$$

$$E_2(G) = \sum_{i,j=1}^N |B_{i,j}^2 - G_{i,j}|$$

$$P(G|\beta_1, \beta_2) = \frac{e^{-\{\beta_1 E_1(G) + \beta_2 E_2(G)\}}}{Z(\beta_1, \beta_2)}$$

$$Z(\beta_1, \beta_2) = \sum_{G \in \mathcal{G}} e^{-\{\beta_1 E_1(G) + \beta_2 E_2(G)\}}$$

Energy of a network

$$E_1(G) = \sum_{n=1}^N \mathcal{E}_1(n, \pi_n[G])$$

$$E_2(G) = \sum_{n=1}^N \mathcal{E}_2(n, \pi_n[G])$$

Rewriting the energy

$$\mathcal{E}_1(n, \pi_n) = \sum_{i \in \pi_n} (1 - B_{in}^1) + \sum_{i \notin \pi_n} B_{in}^1$$

$$\mathcal{E}_2(n, \pi_n) = \sum_{i \in \pi_n} (1 - B_{in}^2) + \sum_{i \notin \pi_n} B_{in}^2$$

Approximation of the partition function

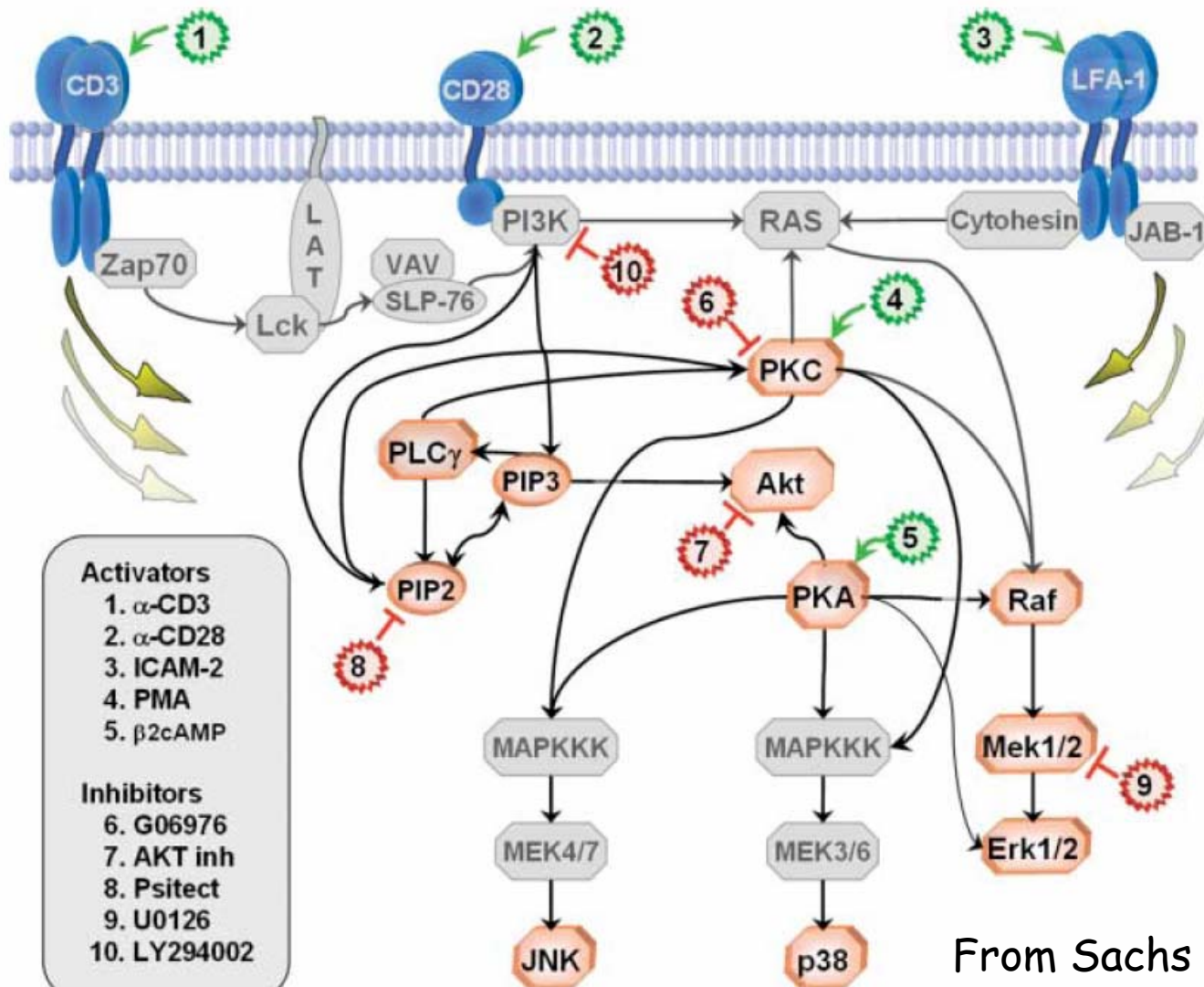
$$\begin{aligned} Z &= \sum_{G \in \mathcal{G}} e^{-\{\beta_1 E_1(G) + \beta_2 E_2(G)\}} \\ &= \sum_{\pi_1} \cdots \sum_{\pi_N} e^{-\{\beta_1 [\mathcal{E}_1(1, \pi_1) + \cdots + \mathcal{E}_1(N, \pi_N)] + \beta_2 [\mathcal{E}_2(1, \pi_1) + \cdots + \mathcal{E}_2(N, \pi_N)]\}} \\ &= \prod_n \sum_{\pi_n} e^{-\{\beta_1 \mathcal{E}_1(n, \pi_n) + \beta_2 \mathcal{E}_2(n, \pi_n)\}} \end{aligned}$$

Partition function of an ideal gas

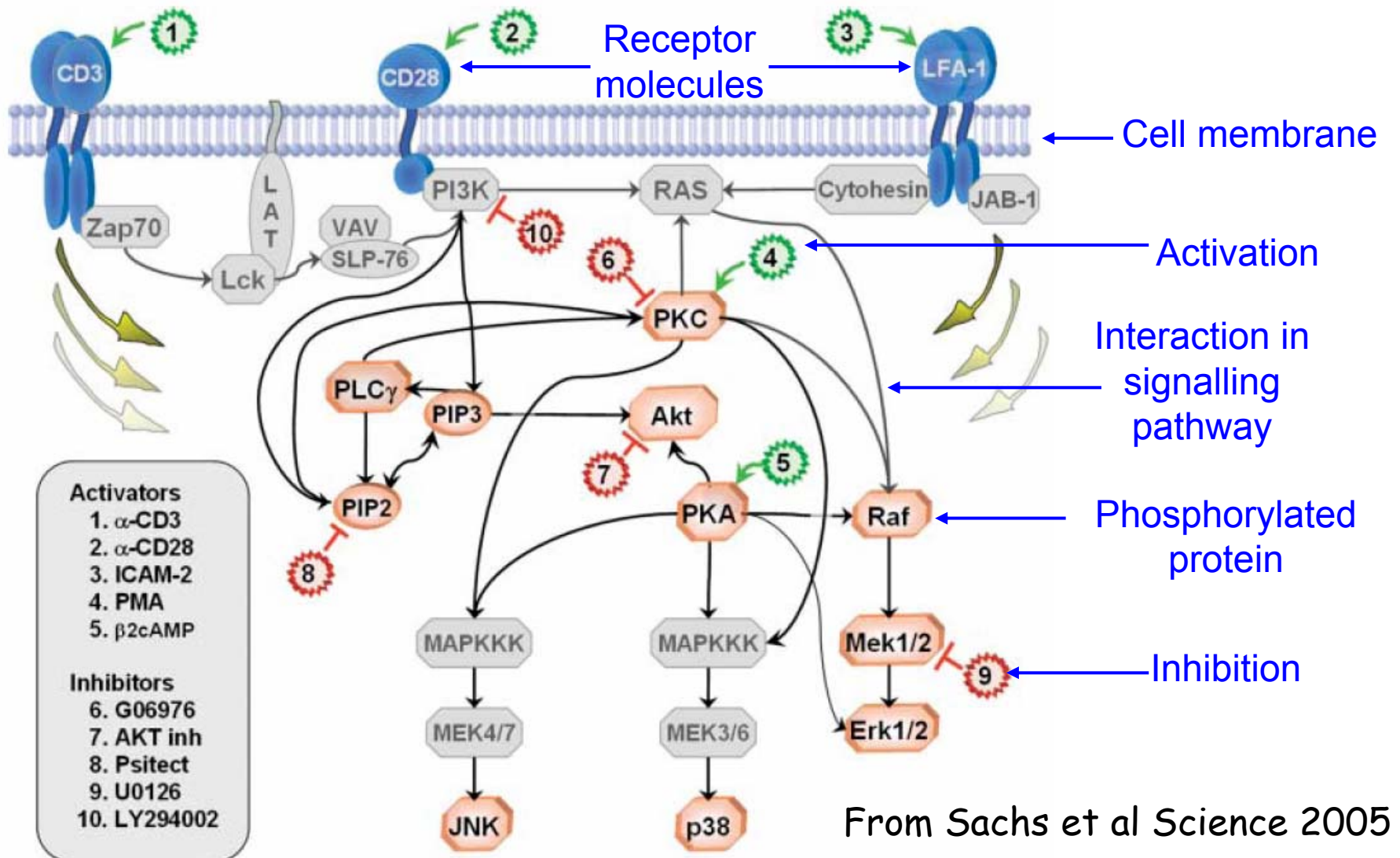
Evaluation

- Can the method automatically evaluate how useful the different sources of prior knowledge are?
- Do we get an improvement in the regulatory network reconstruction?
- Is this improvement optimal?

Evaluation on the Raf regulatory network



Raf signalling pathway



Evaluation:

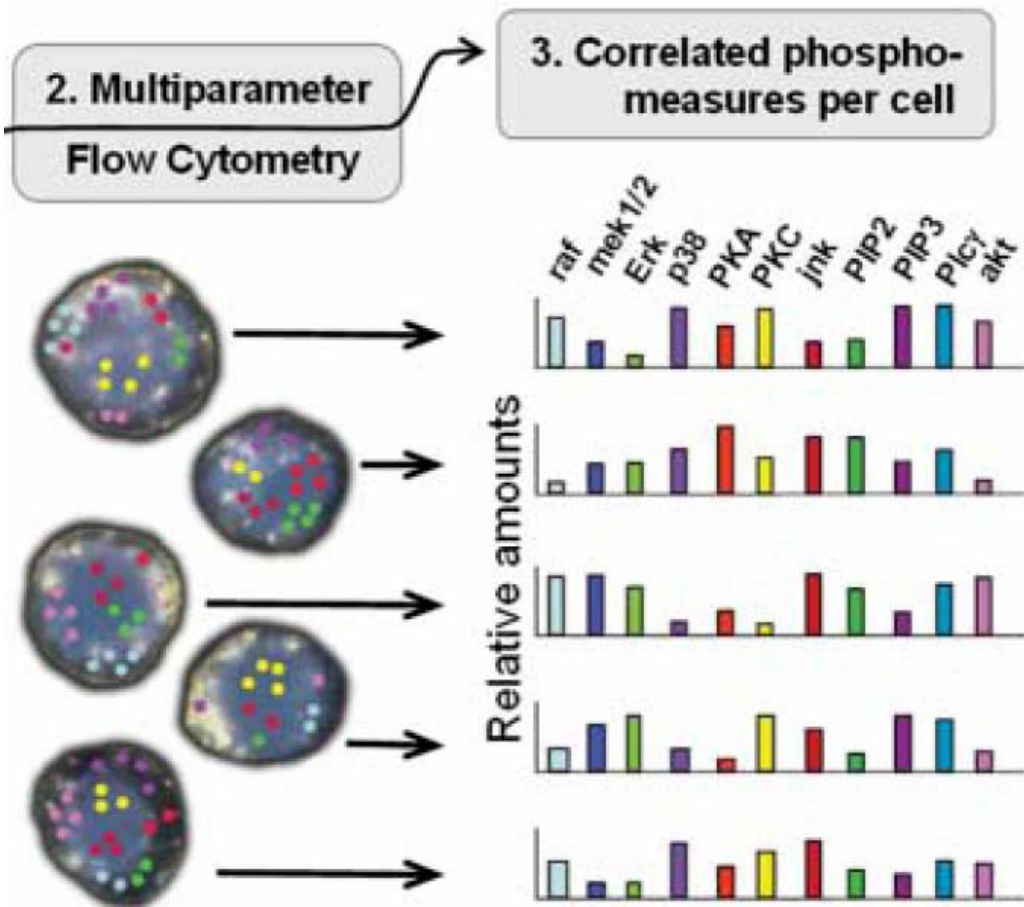
Raf signalling pathway

- Cellular signalling network of **11 phosphorylated proteins** and phospholipids in human immune systems cell
- Deregulation → carcinogenesis
- Extensively studied in the literature → **gold standard network**

Data

Prior knowledge

Flow cytometry data



Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data

Karen Sachs,^{1*} Omar Perez,^{2*} Dana Pe'er,^{3*}
Douglas A. Lauffenburger,^{1†} Garry P. Nolan^{2†}

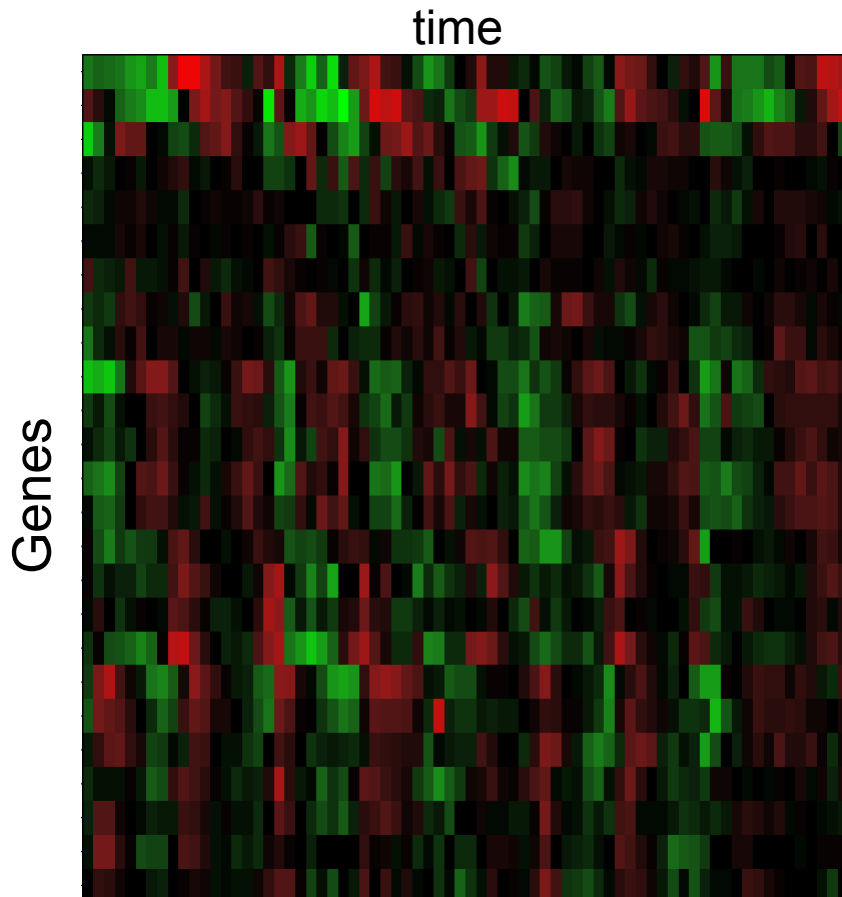
- Intracellular multicolour flow cytometry experiments: **concentrations of 11 proteins**
- **5400 cells** have been measured under 9 different cellular conditions (cues)
- **Downsampling** to 100 instances (5 separate subsets): **indicative of microarray experiments**

Microarray example

Spellman et al (1998)

Cell cycle

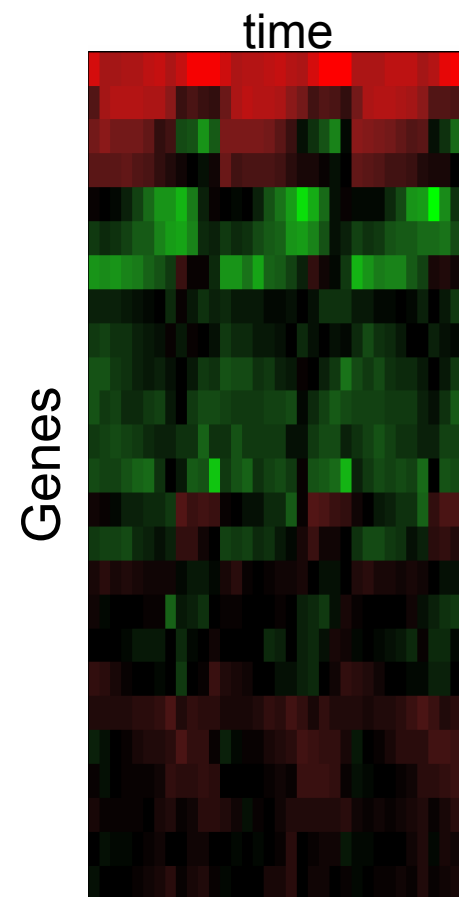
73 samples



Tu et al (2005)

Metabolic cycle

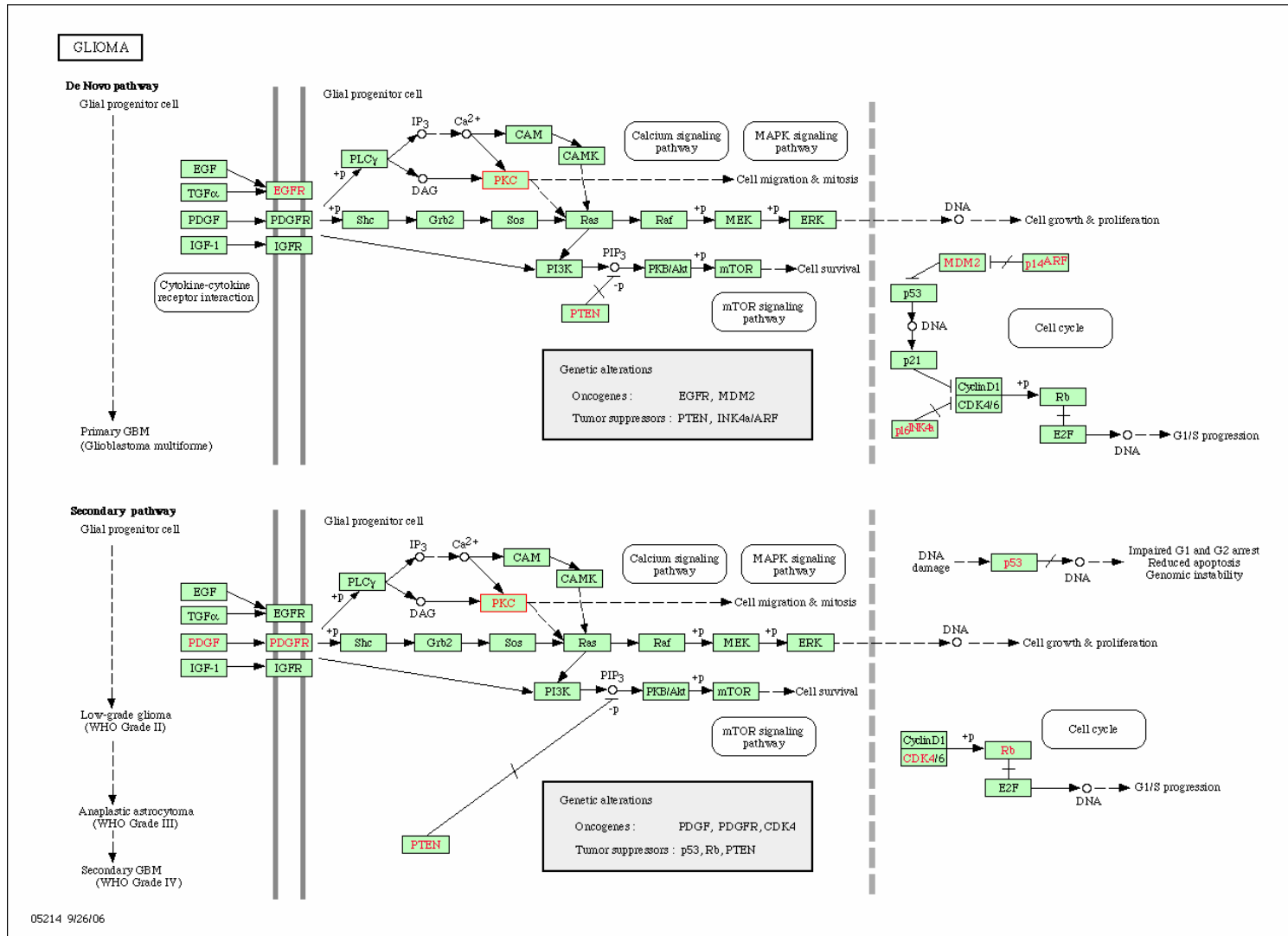
36 samples



Data

Prior knowledge

Prior knowledge from KEGG



Prior distribution

$$P(G|\beta) = \frac{e^{-\beta E(G)}}{Z(\beta)}$$

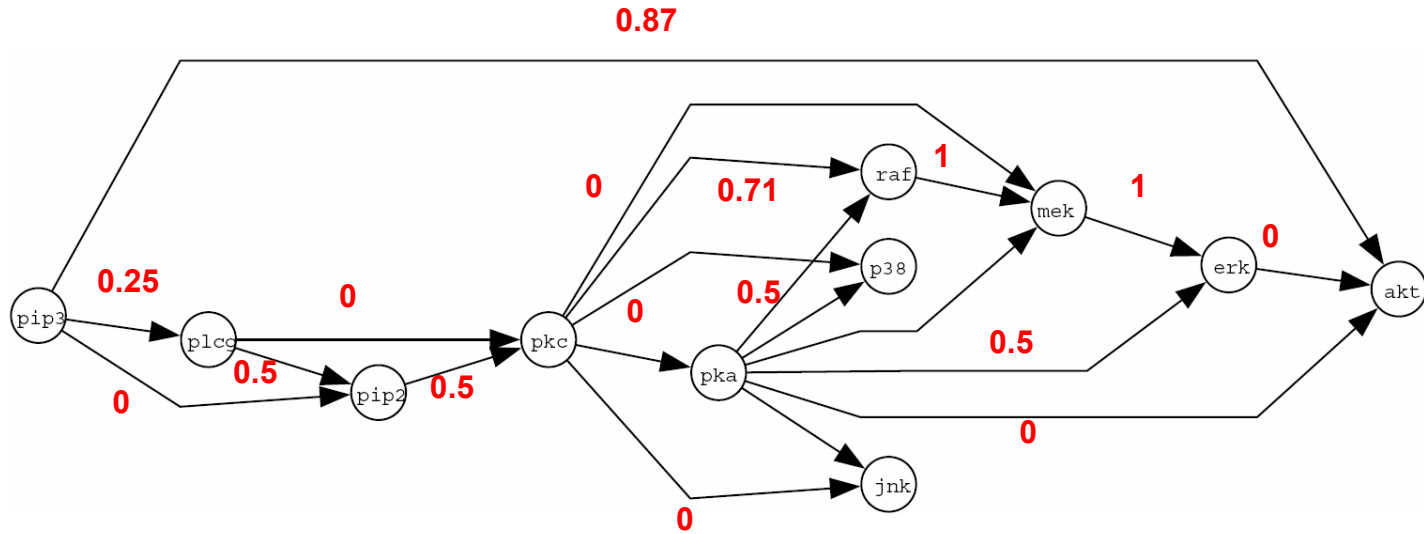
$$E(G) = \sum_{i,j=1}^N |B_{i,j} - G_{i,j}|$$

Define by M_{ij} the total number of times a pair of genes i and j appears in a pathway, and by m_{ij} the number of times the genes are connected by a (directed) edge in the KEGG pathway. The elements B_{ij} of the prior knowledge matrix are then defined by

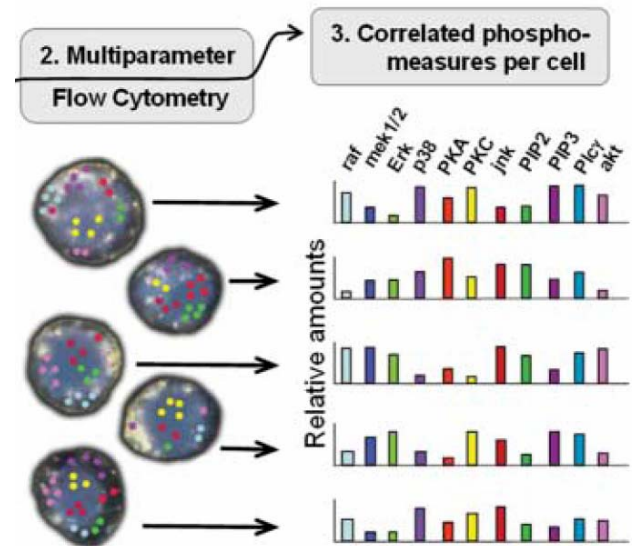
$$B_{ij} = \frac{m_{ij}}{M_{ij}} \quad (43)$$

If a pair of genes is not found in any of the KEGG pathways, we set the respective prior association to $B_{ij} = 0.5$, implying that we have no information about this relationship.

Prior knowledge from KEGG



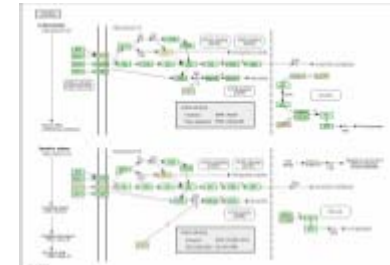
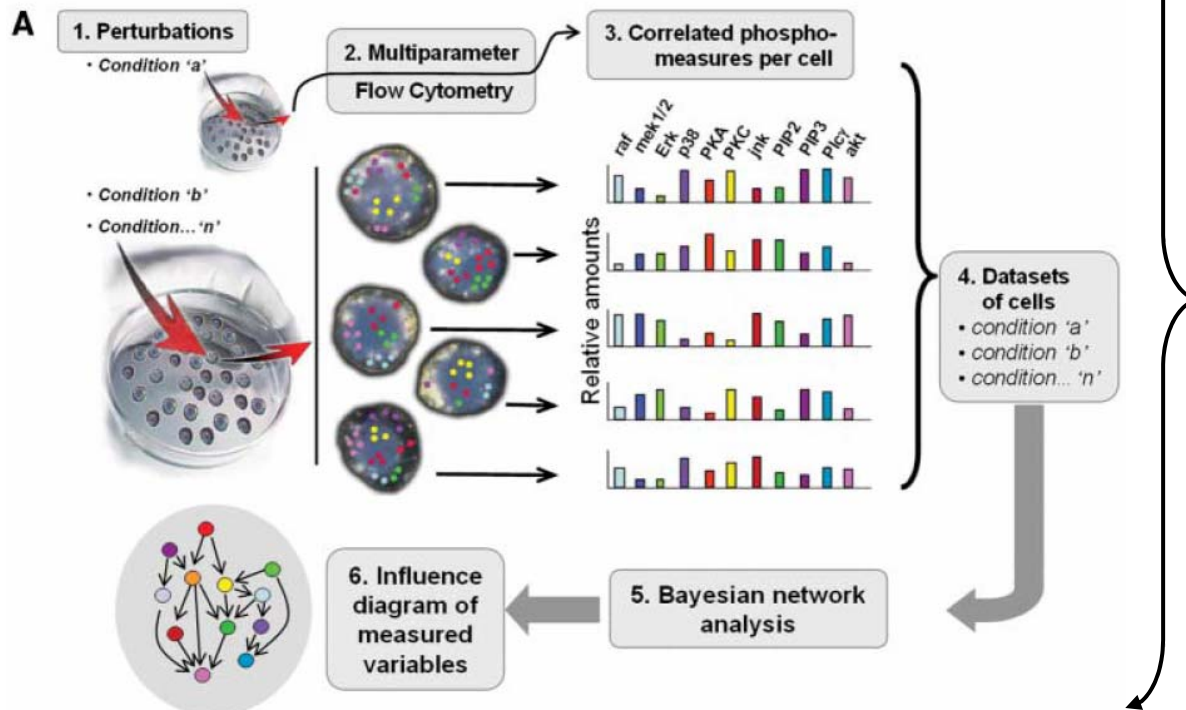
Data: protein concentrations from flow cytometry experiments



Data and prior knowledge

Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data

Karen Sachs,^{1*} Omar Perez,^{2*} Dana Pe'er,^{3*}
Douglas A. Lauffenburger,^{1†} Garry P. Nolan^{2†}



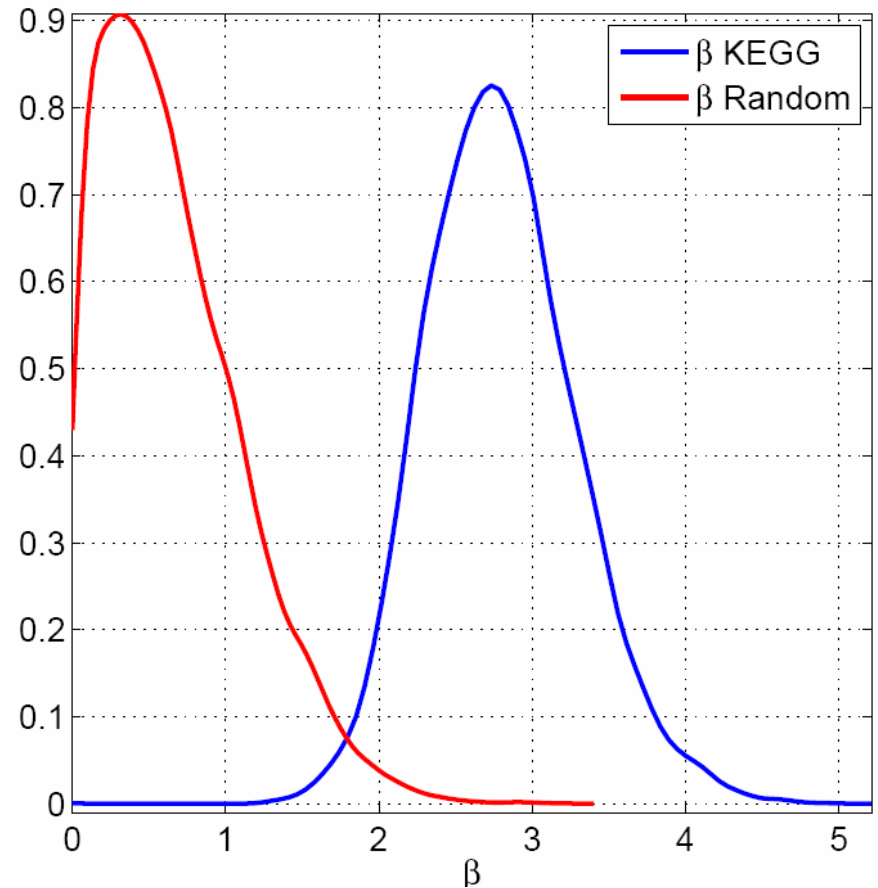
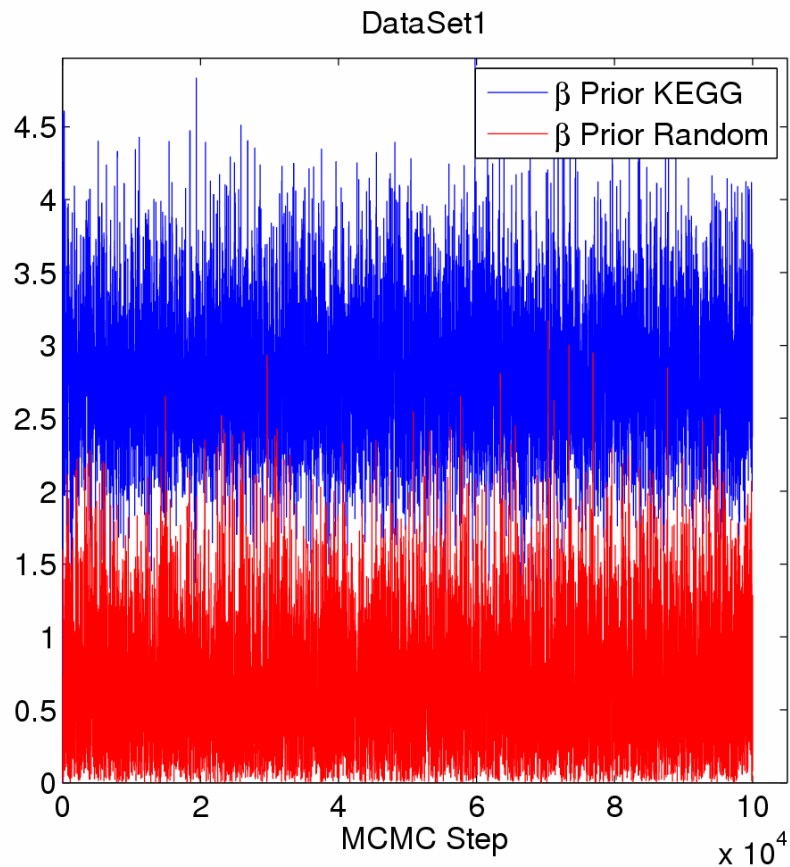
+ KEGG
+ Random



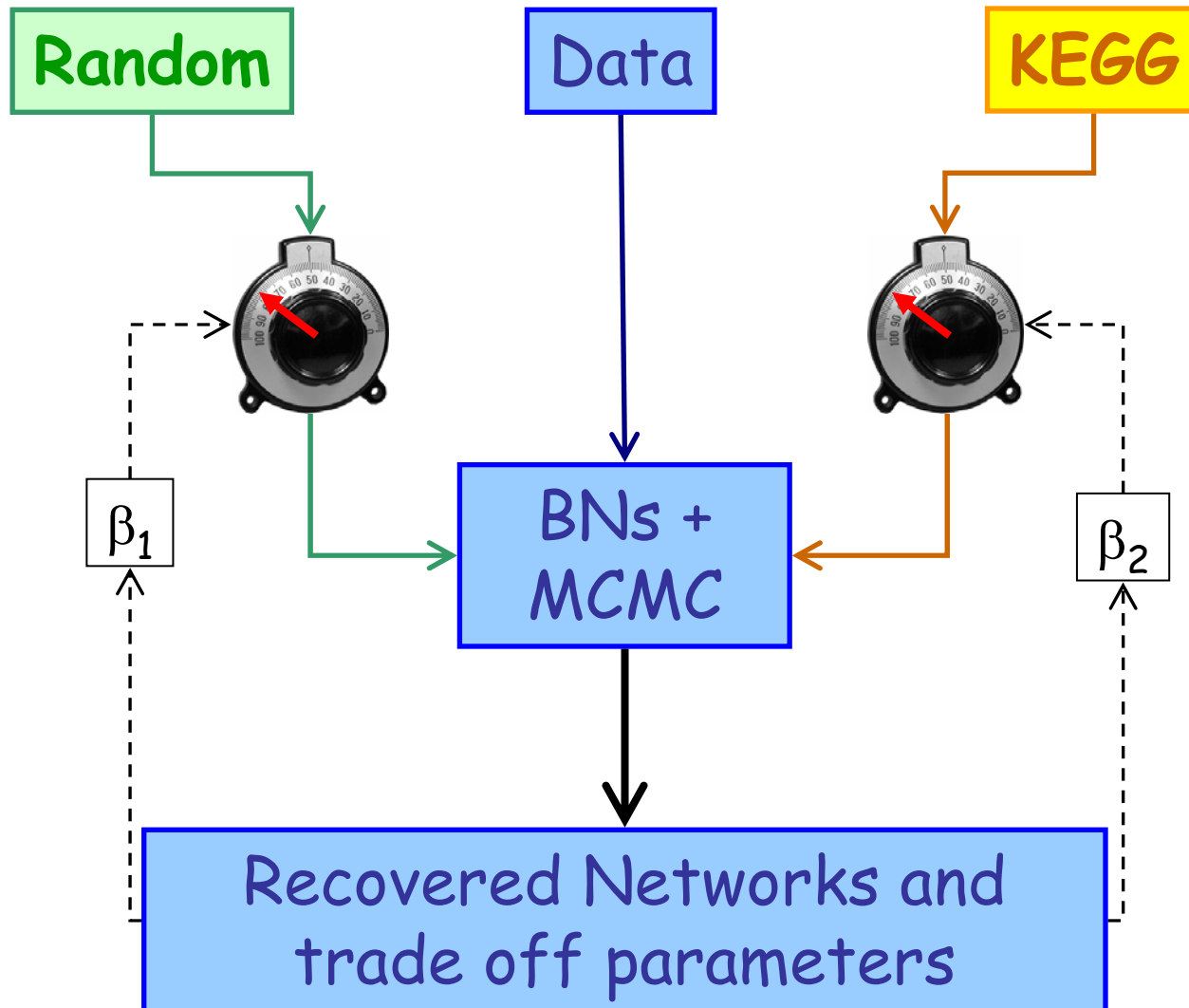
Evaluation

- **Can the method automatically evaluate how useful the different sources of prior knowledge are?**
- Do we get an improvement in the regulatory network reconstruction?
- Is this improvement optimal?

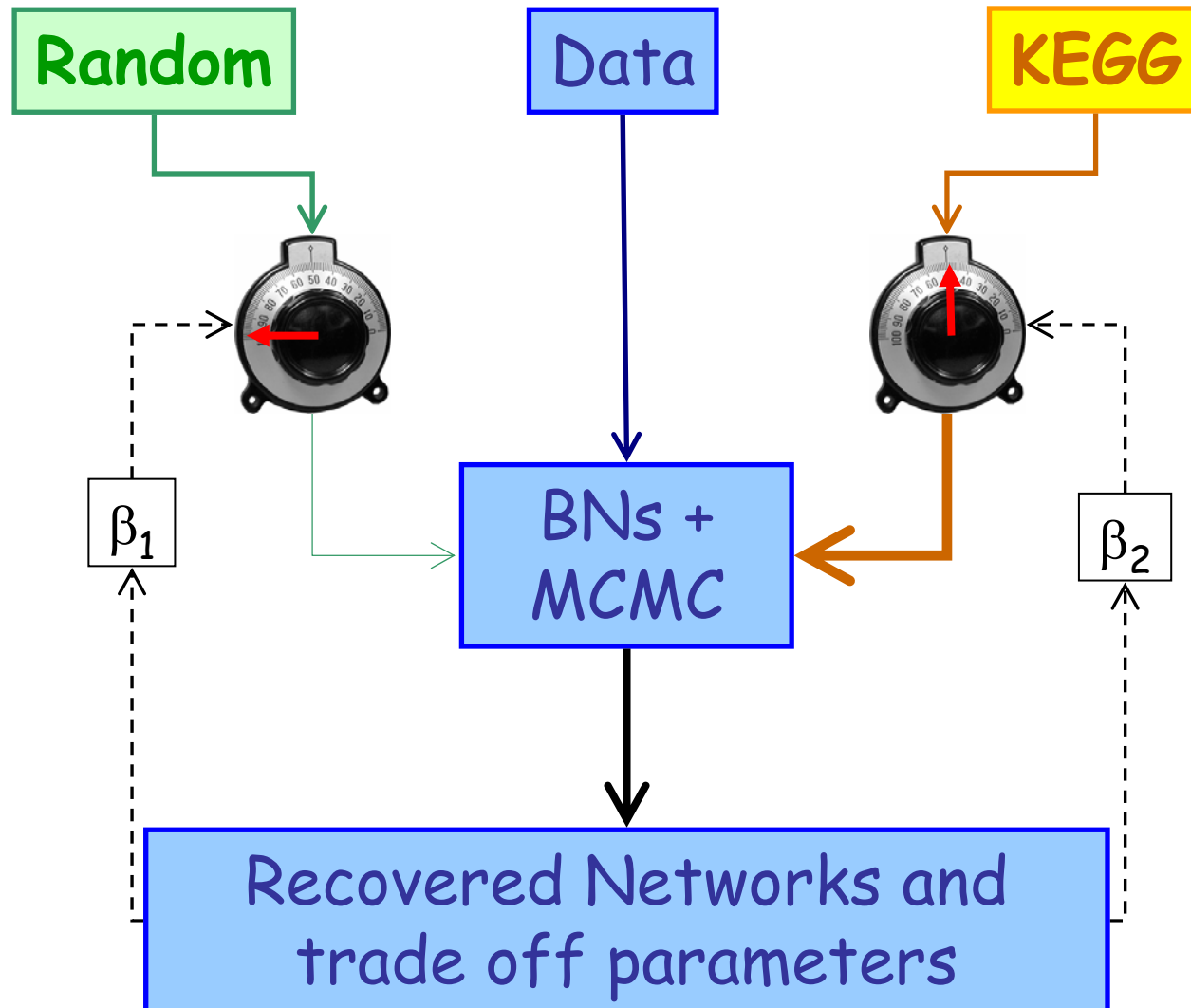
Sampled values of the hyperparameters



Bayesian networks with two sources of prior knowledge



Bayesian networks with two sources of prior knowledge

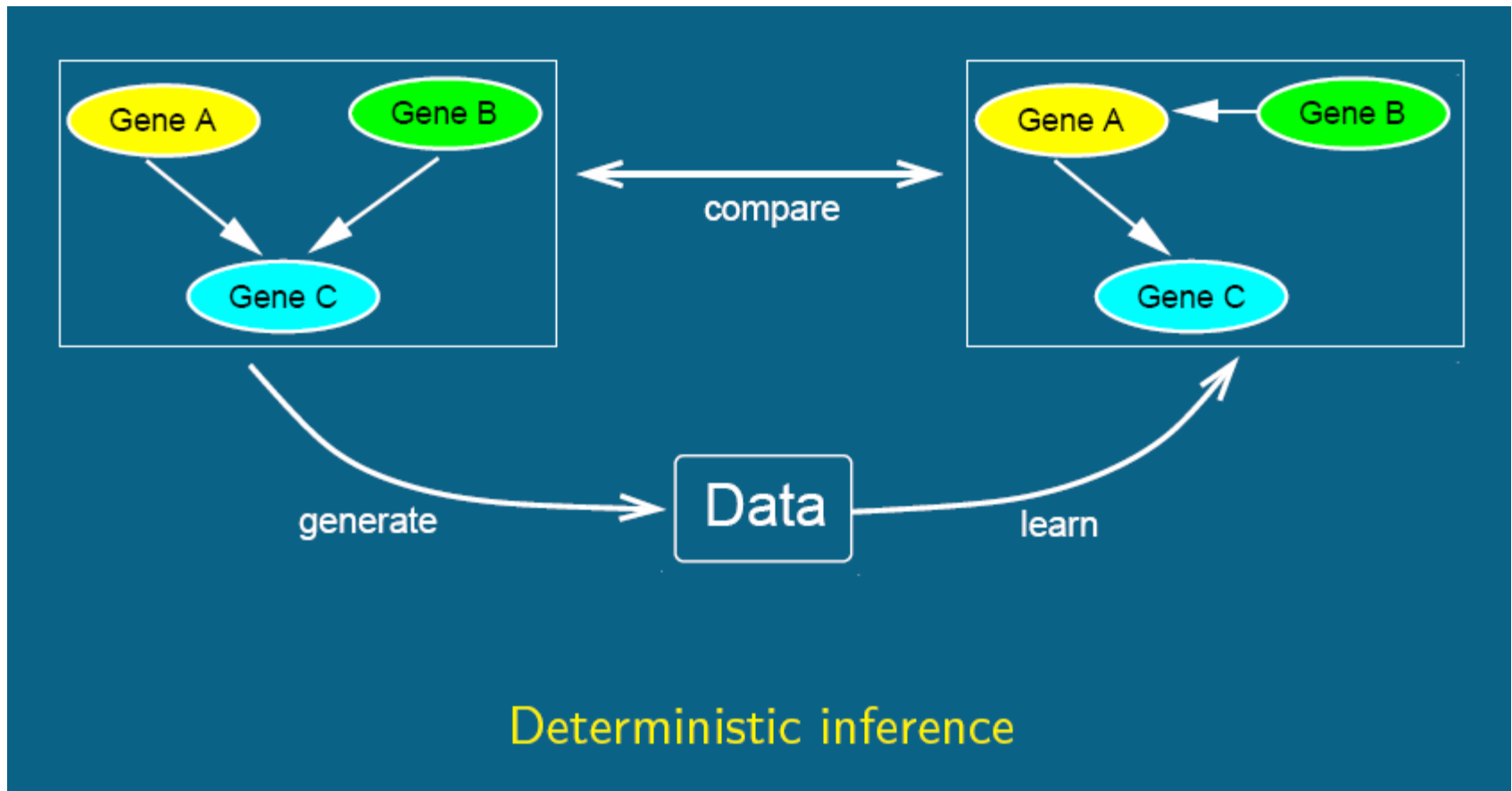


Evaluation

- Can the method automatically evaluate how useful the different sources of prior knowledge are?
- **Do we get an improvement in the regulatory network reconstruction?**
- Is this improvement optimal?

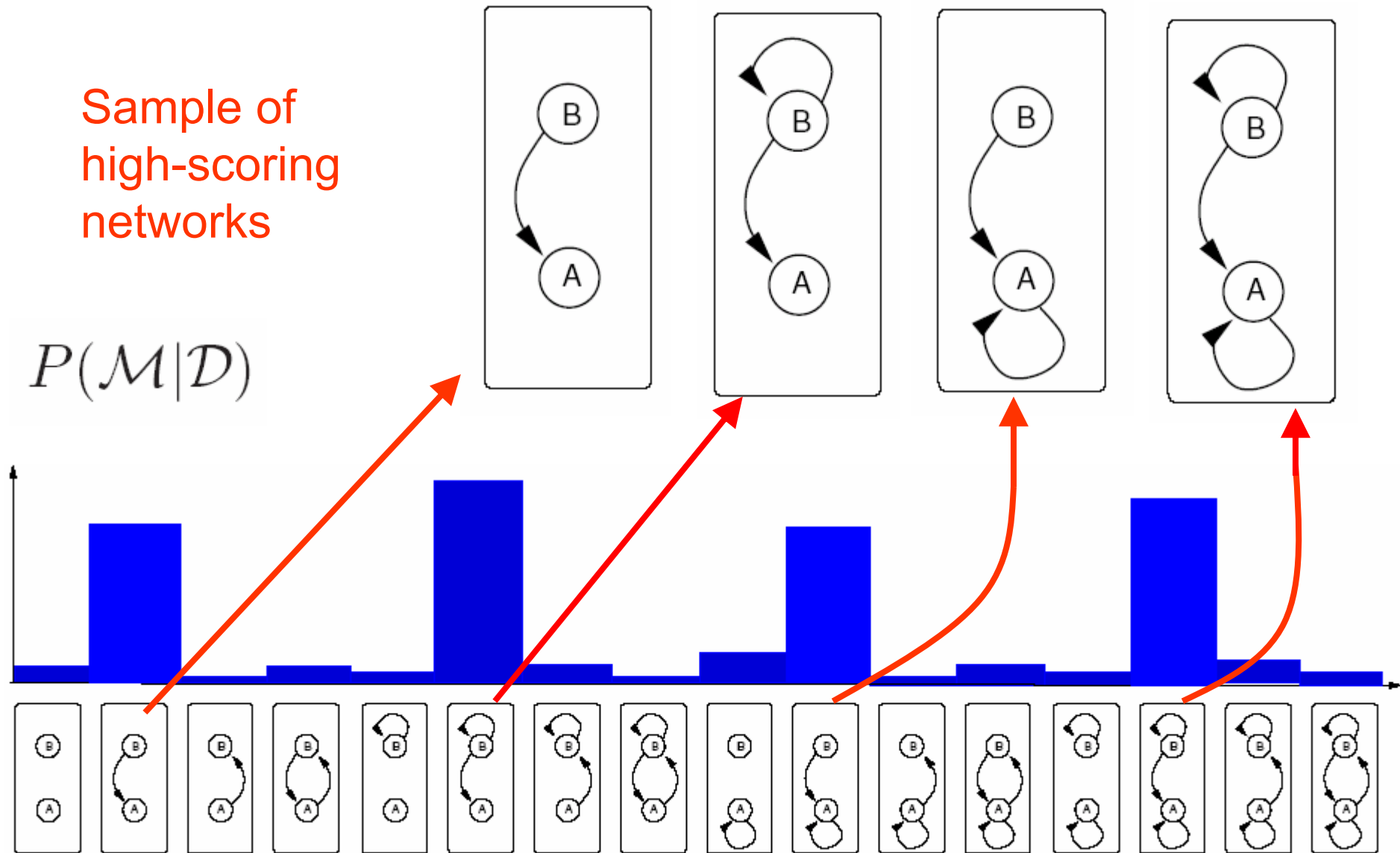
Gold-standard network

Inferred network

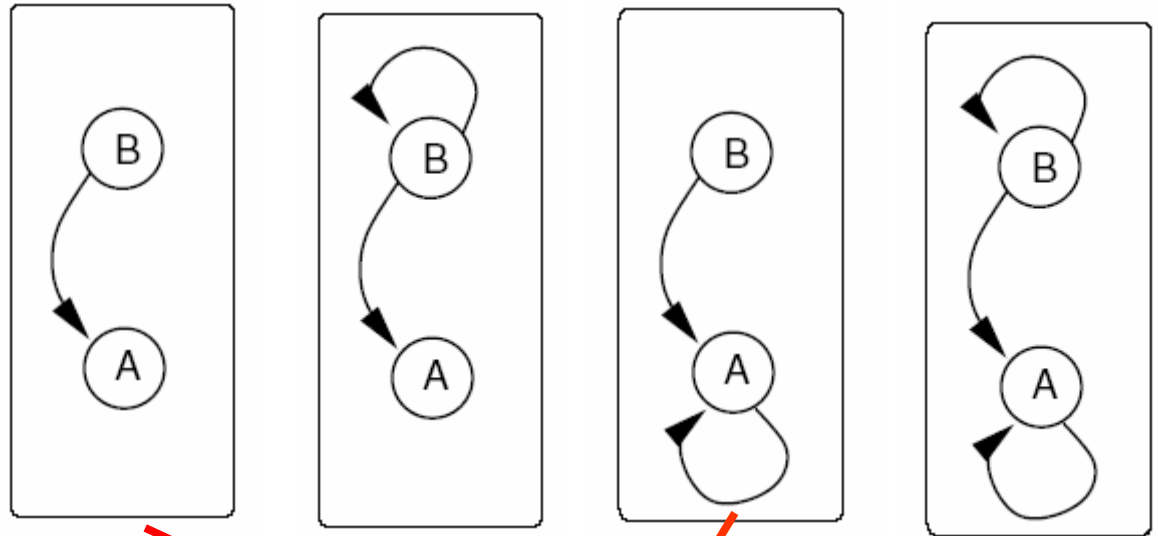


Sample of high-scoring networks

$$P(\mathcal{M}|\mathcal{D})$$



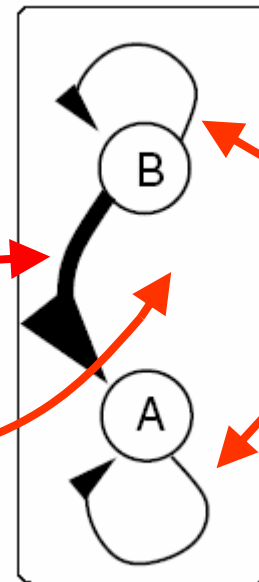
Sample of high-scoring networks



Feature extraction,
e.g. marginal posterior probabilities of the edges

High-confident edge

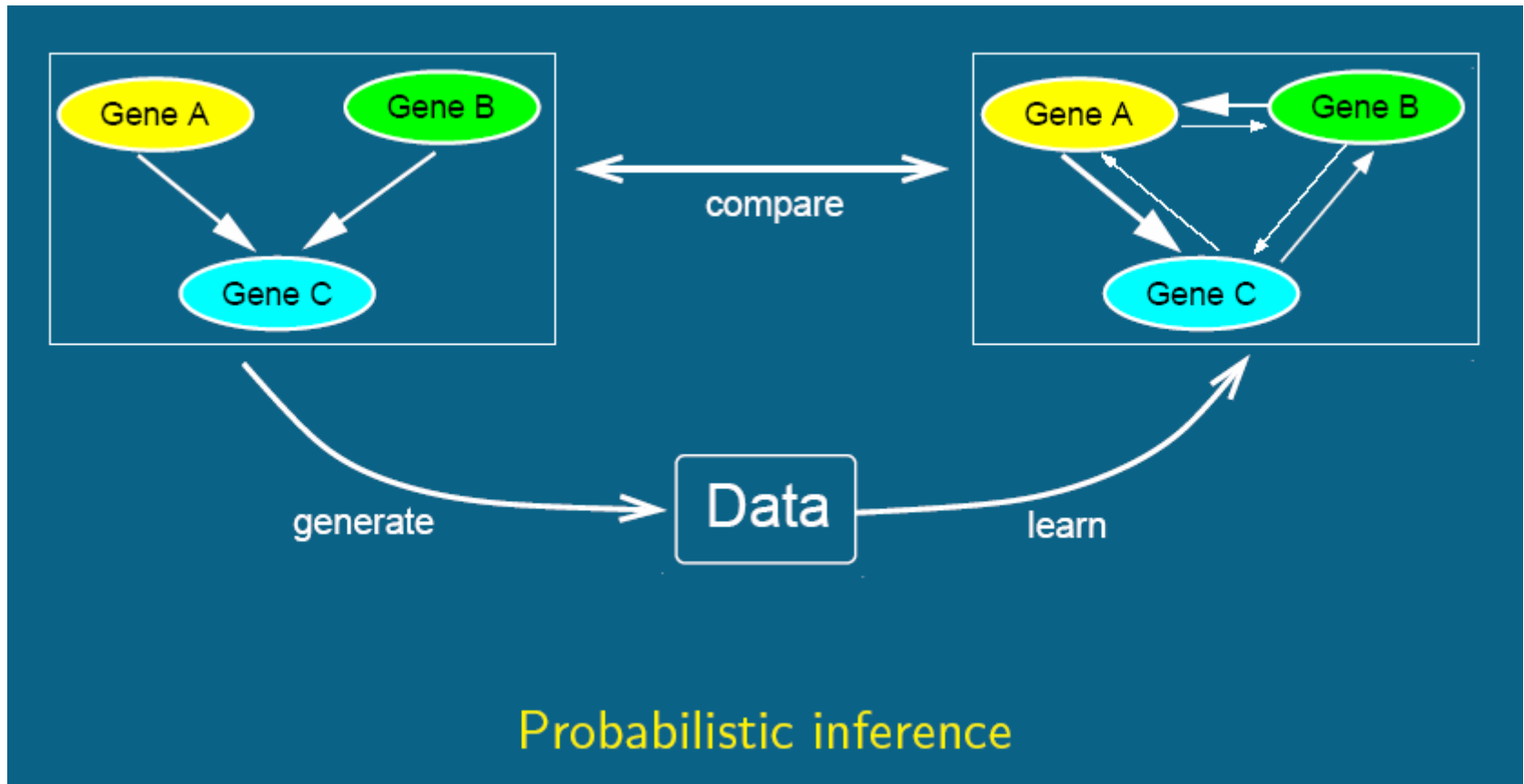
High-confident non-edge



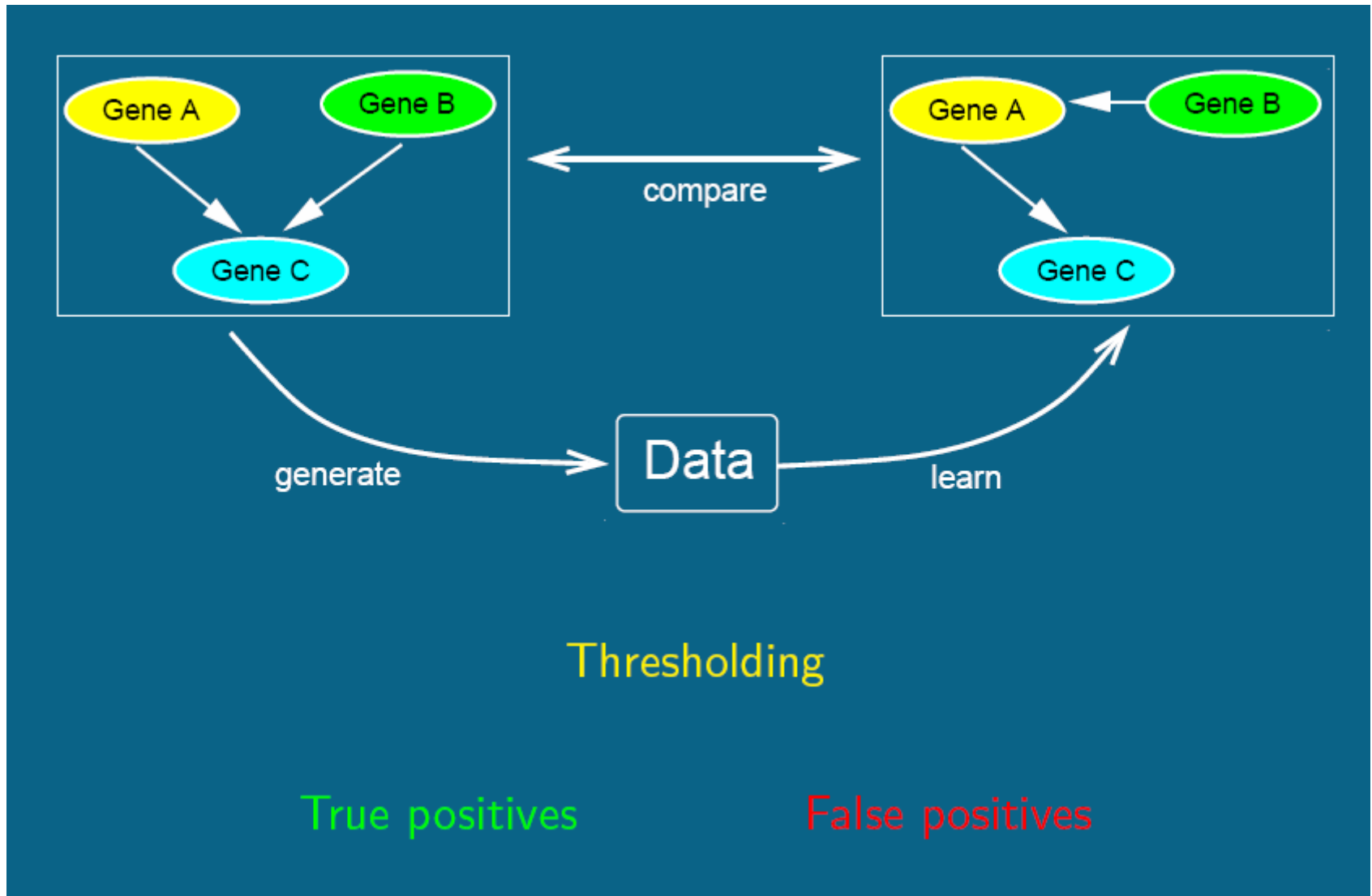
Uncertainty about edges

Gold standard network

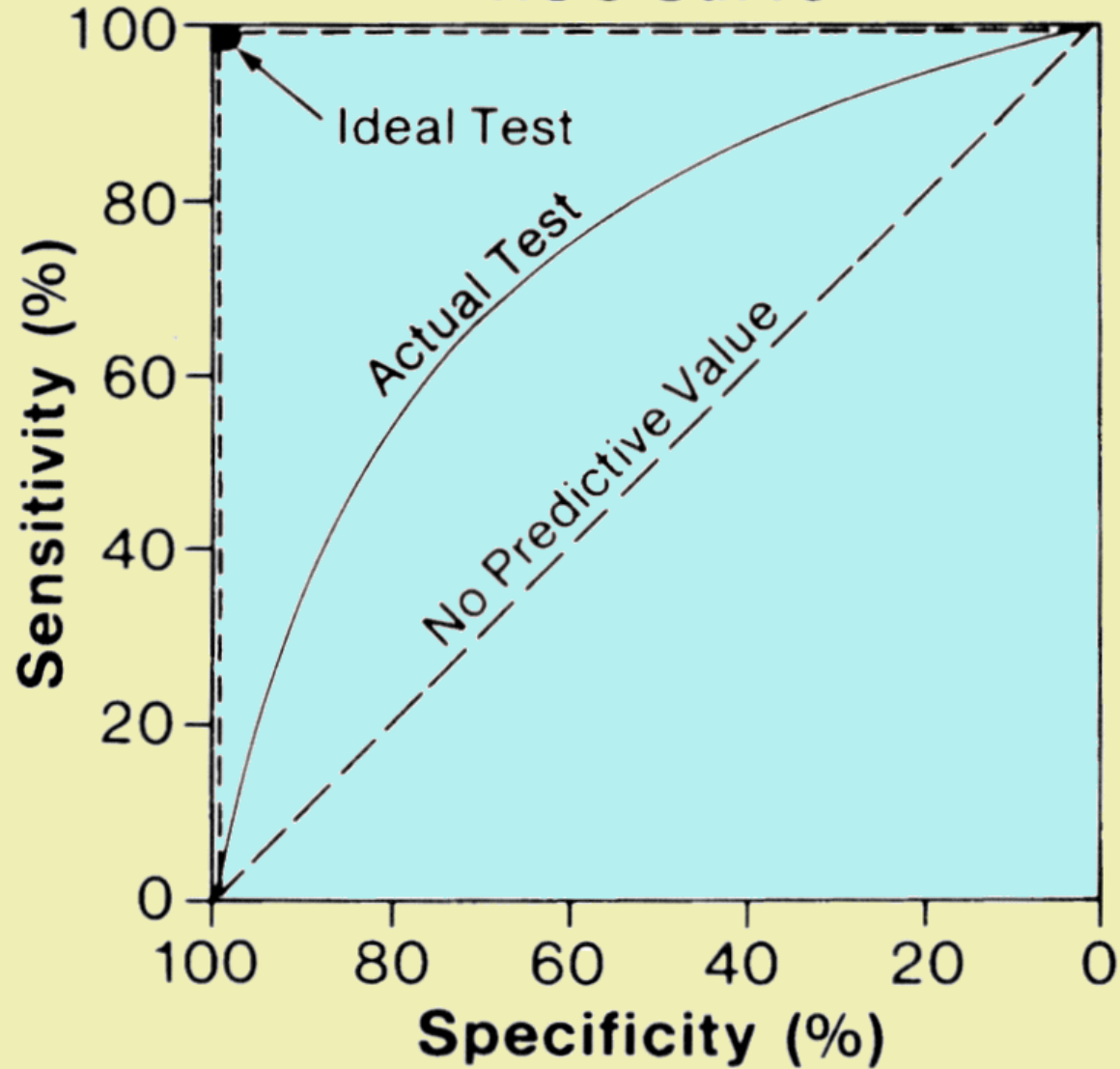
Inferred network distribution

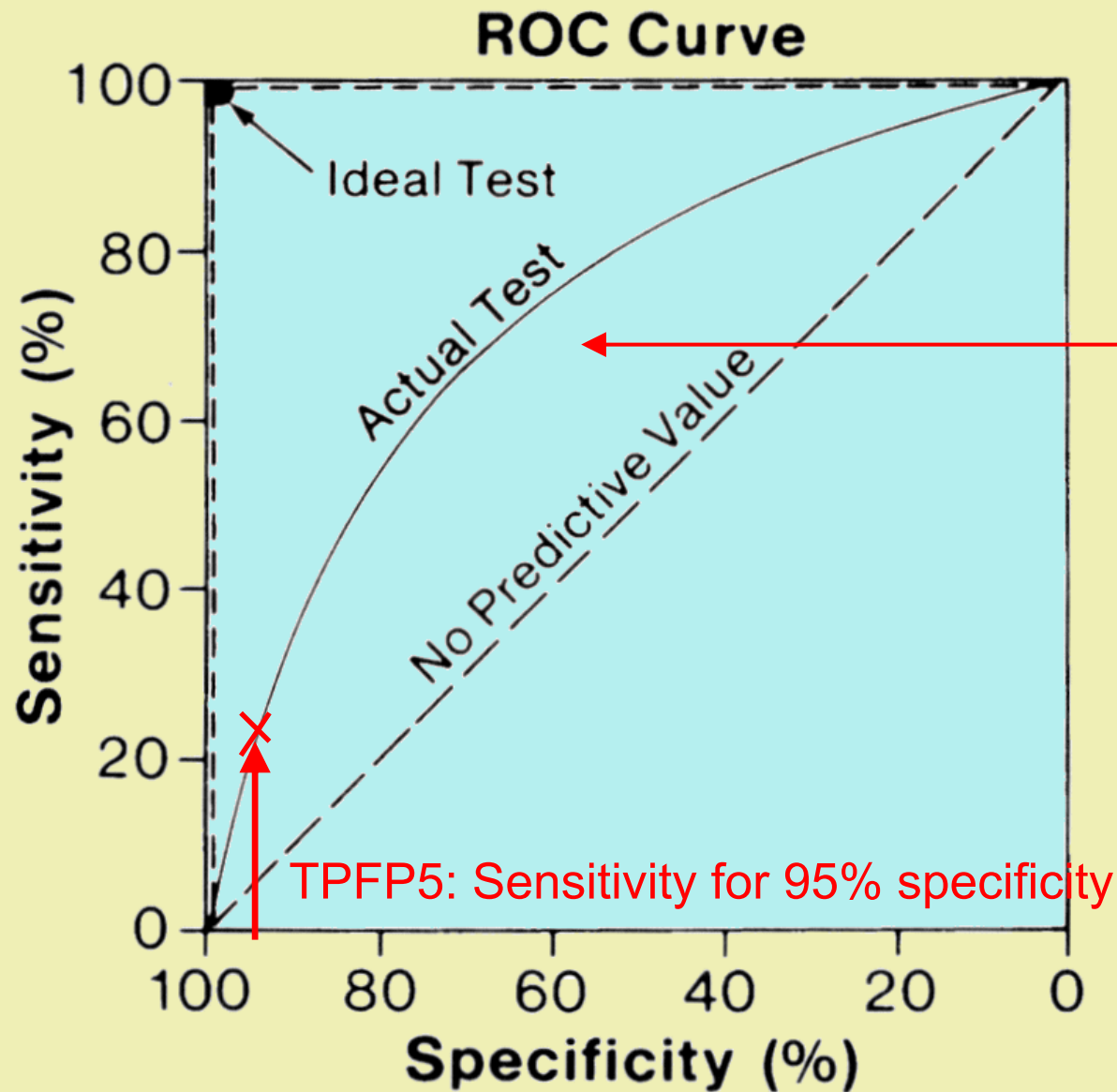


Gold-standard network



ROC Curve





AUC:
area
under
the
curve

TPFP5: Sensitivity for 95% specificity

Evaluation

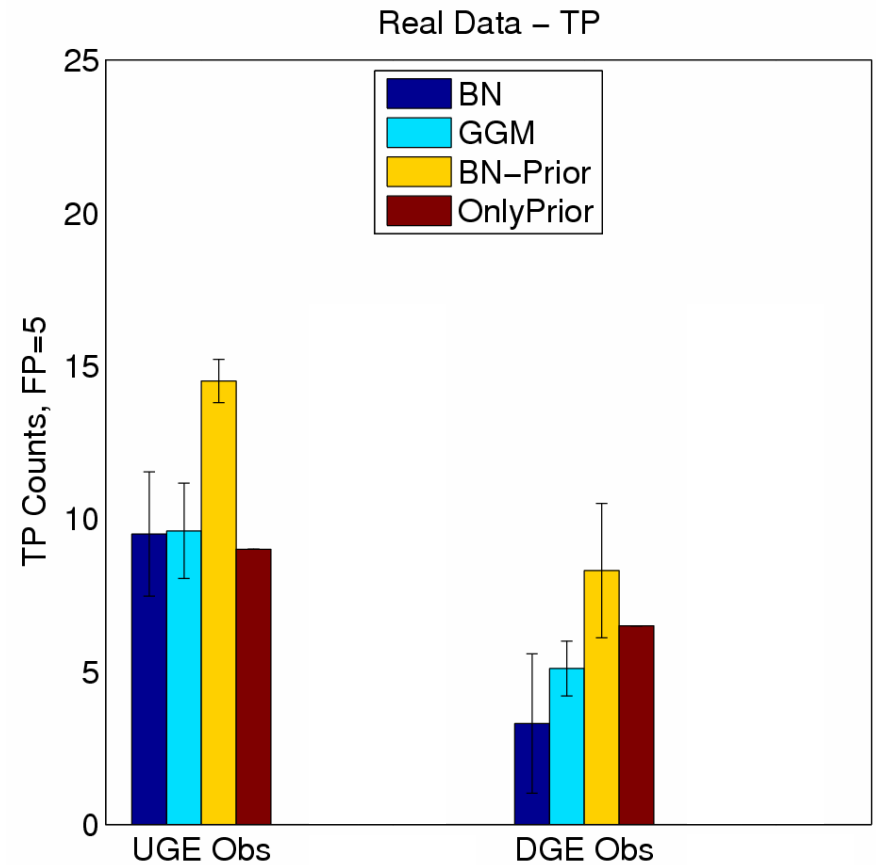
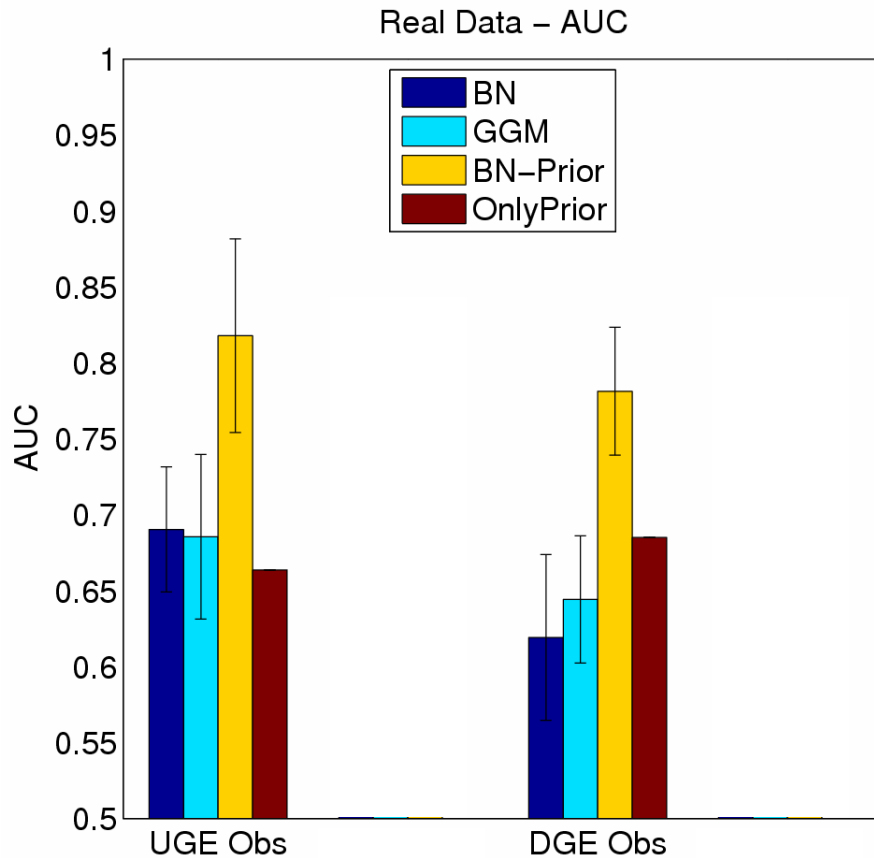
Two ways of interpreting edges

- **UGE:** undirected graph evaluation
- **DGE:** directed graph evaluation

Two evaluation procedures:

- **AUC:** Area under the ROC curve, with larger areas indicating, overall, a better performance.
- **TP count:** True positive number of edges for the same false positive count of $FP=5$ across all methods.

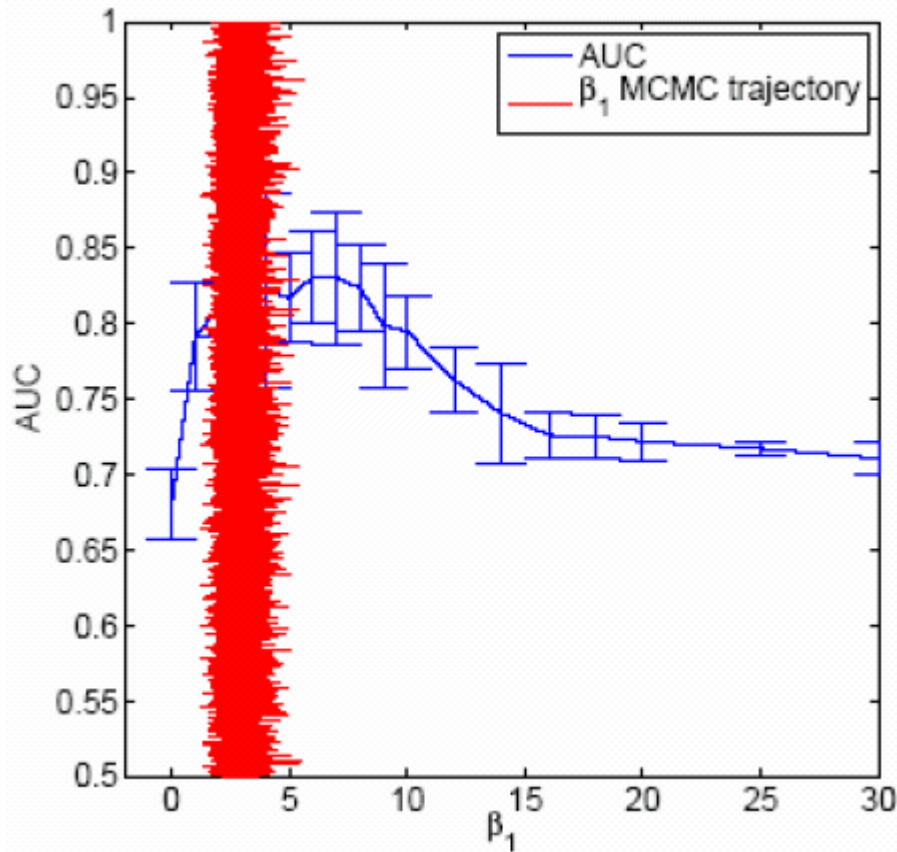
Flow cytometry data and KEGG



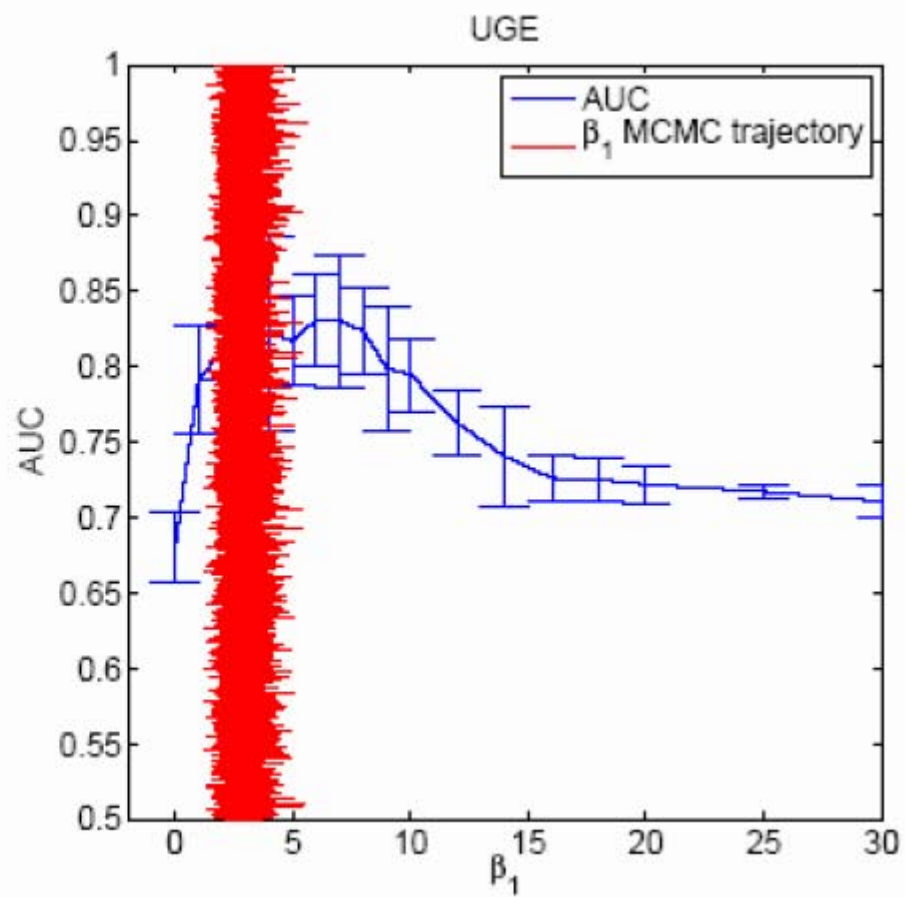
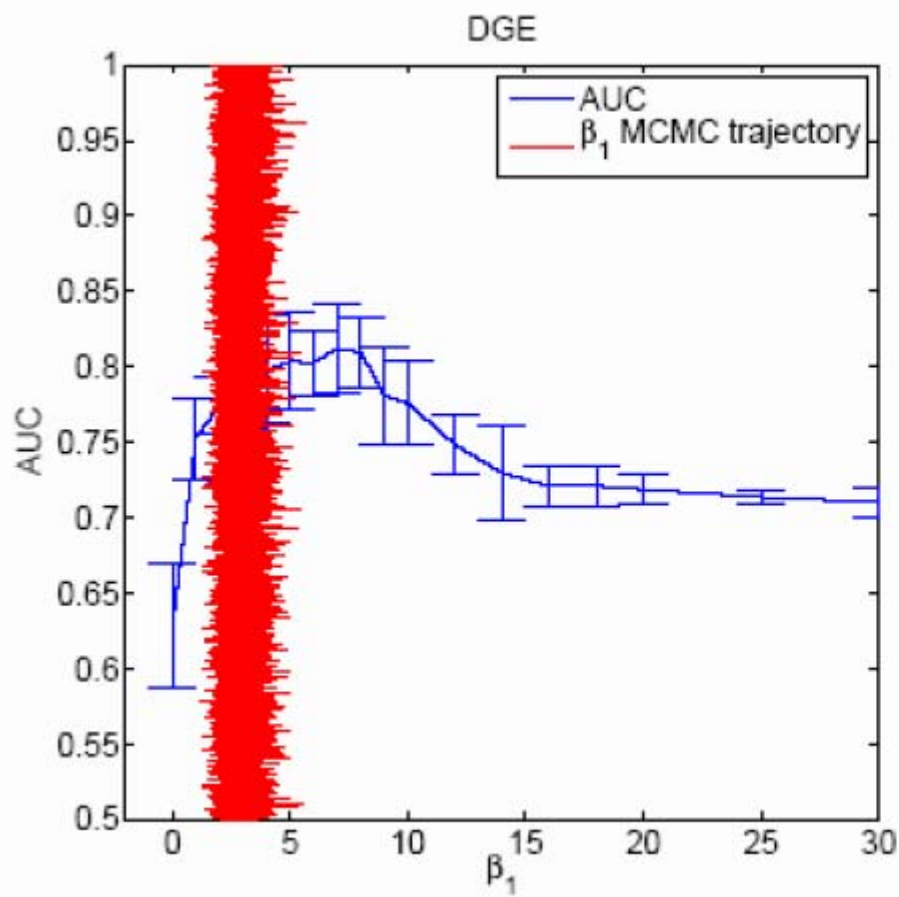
Evaluation

- Can the method automatically evaluate how useful the different sources of prior knowledge are?
- Do we get an improvement in the regulatory network reconstruction?
- **Is this improvement optimal?**

Learning the trade-off hyperparameter



- **Repeat MCMC simulations** for large set of **fixed hyperparameters β**
- Obtain AUC scores for each value of β
- Compare with the proposed scheme in which β is automatically inferred.



Conclusions – Part 1

- The method can automatically evaluate how useful the different sources of prior knowledge are.
- We get an improvement in the regulatory network reconstruction.
- The improvement is close to optimal.

Part 2

**Combining data from different
experimental conditions**

**GENE REGULATORY NETWORK RECONSTRUCTION
BY BAYESIAN INTEGRATION OF PRIOR KNOWLEDGE
AND/OR DIFFERENT EXPERIMENTAL CONDITIONS**

ADRIANO V. WERHLI

*Department of Computing Science
Pontifical Catholic University of Rio Grande do Sul
Porto Alegre, Brazil
werhli@gmail.com*

DIRK HUSMEIER

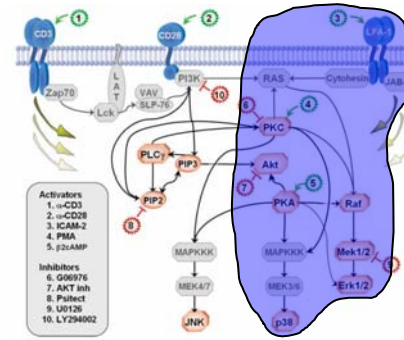
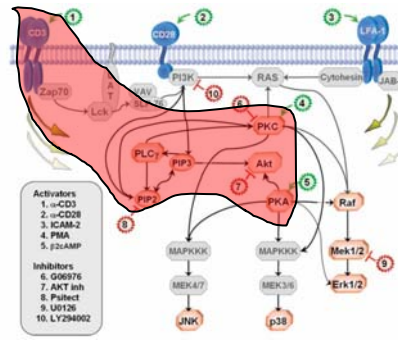
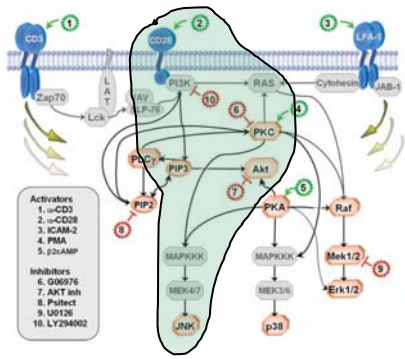
*Biomathematics and Statistics Scotland
Edinburgh, United Kingdom
dirk@bioss.ac.uk*

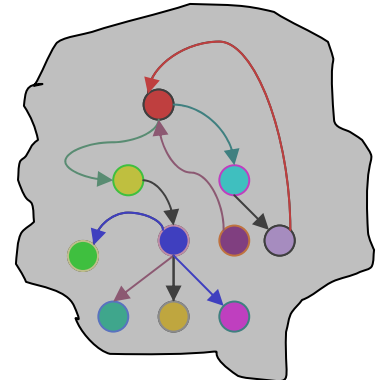
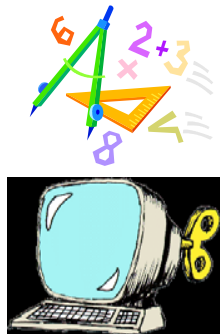
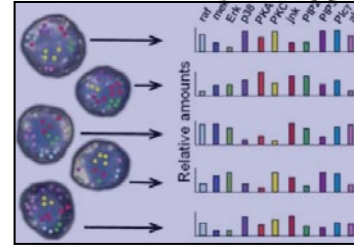
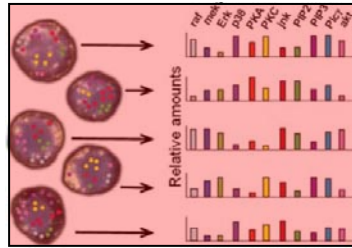
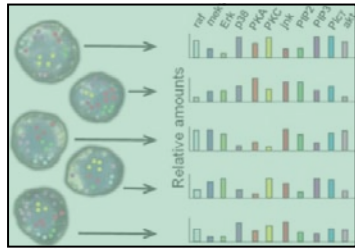
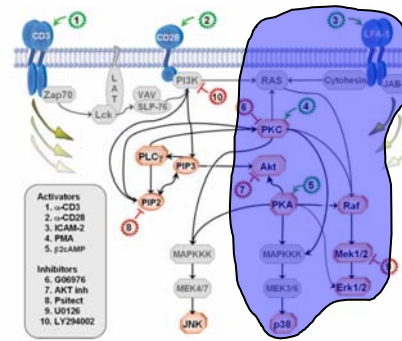
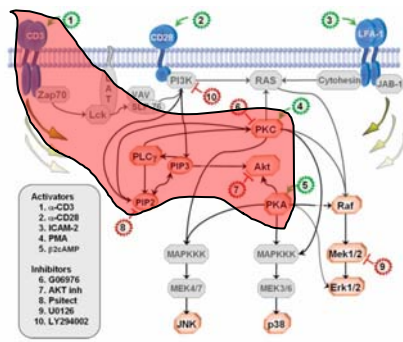
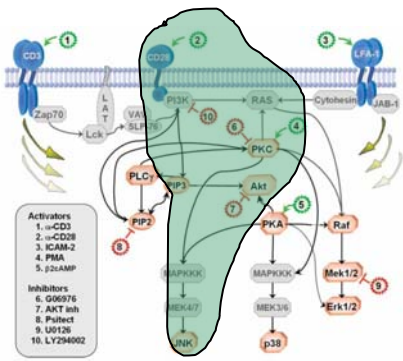
Received 1 August 2007
Revised 1 December 2007
Accepted 3 January 2008

What if we have multiple data sets obtained under different experimental conditions?

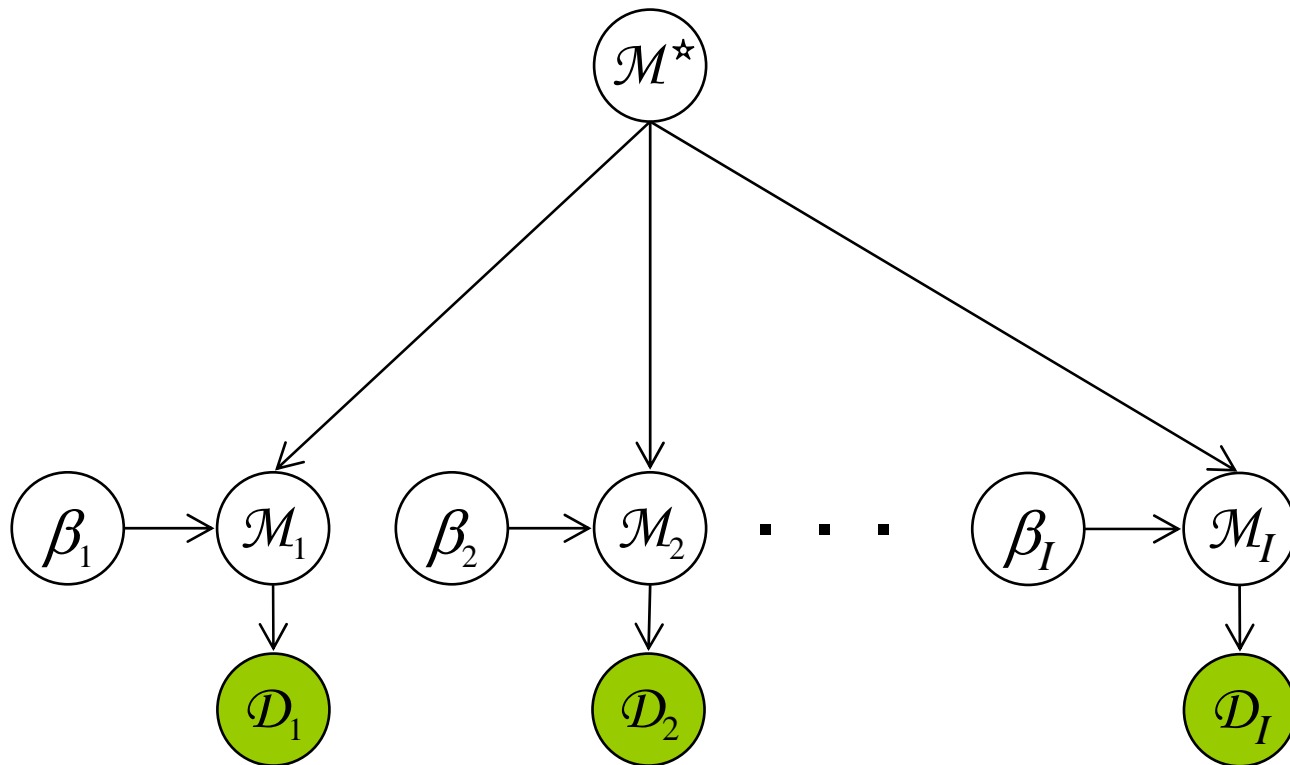
Example: Cytokine network

- Infection
- Treatment with IFN
- Infection and treatment with IFN






Compromise between the two previous ways of combining the data



$$P(\mathcal{M}_1, \dots, \mathcal{M}_I, \mathcal{D}_1 \dots \mathcal{D}_I, \beta_1, \dots, \beta_I, \mathcal{M}^*) = \prod_{i=1}^I P(\mathcal{D}_i | \mathcal{M}_i) P(\mathcal{M}_i | \beta_i, \mathcal{M}^*) P(\beta_i) P(\mathcal{M}^*)$$

BGe or BDe

$$P(\mathcal{M}_1, \dots, \mathcal{M}_I, \mathcal{D}_1 \dots \mathcal{D}_I, \beta_1, \dots, \beta_I, \mathcal{M}^*) = \prod_{i=1}^I P(\mathcal{D}_i | \mathcal{M}_i) P(\mathcal{M}_i | \beta_i, \mathcal{M}^*) P(\beta_i) P(\mathcal{M}^*)$$


BGe or BDe

$$P(\mathcal{M}_1, \dots, \mathcal{M}_I, \mathcal{D}_1 \dots \mathcal{D}_I, \beta_1, \dots, \beta_I, \mathcal{M}^*) = \prod_{i=1}^I P(\mathcal{D}_i | \mathcal{M}_i) P(\mathcal{M}_i | \beta_i, \mathcal{M}^*) P(\beta_i) P(\mathcal{M}^*)$$

$$P(\mathcal{M}_i | \beta_i, \mathcal{M}^*) = \frac{e^{-\beta_i (|\mathcal{M}_i - \mathcal{M}^*|)}}{Z(\beta_i, \mathcal{M}^*)}$$

BGe or BDe

$$P(\mathcal{M}_1, \dots, \mathcal{M}_I, \mathcal{D}_1 \dots \mathcal{D}_I, \beta_1, \dots, \beta_I, \mathcal{M}^*) = \prod_{i=1}^I P(\mathcal{D}_i | \mathcal{M}_i) P(\mathcal{M}_i | \beta_i, \mathcal{M}^*) P(\beta_i) P(\mathcal{M}^*)$$

$$Z(\beta_i, \mathcal{M}^*) = \sum_{\mathcal{M}_i \in \mathbb{M}} e^{-\beta_i (|\mathcal{M}_i - \mathcal{M}^*|)}$$

$$P(\mathcal{M}_i | \beta_i, \mathcal{M}^*) = \frac{e^{-\beta_i (|\mathcal{M}_i - \mathcal{M}^*|)}}{Z(\beta_i, \mathcal{M}^*)}$$

BGe or BDe

$$P(\mathcal{M}_1, \dots, \mathcal{M}_I, \mathcal{D}_1 \dots \mathcal{D}_I, \beta_1, \dots, \beta_I, \mathcal{M}^*) = \prod_{i=1}^I P(\mathcal{D}_i | \mathcal{M}_i) P(\mathcal{M}_i | \beta_i, \mathcal{M}^*) P(\beta_i) P(\mathcal{M}^*)$$

$$Z(\beta_i, \mathcal{M}^*) = \sum_{\mathcal{M}_i \in \mathbb{M}} e^{-\beta_i (|\mathcal{M}_i - \mathcal{M}^*|)}$$
$$P(\mathcal{M}_i | \beta_i, \mathcal{M}^*) = \frac{e^{-\beta_i (|\mathcal{M}_i - \mathcal{M}^*|)}}{Z(\beta_i, \mathcal{M}^*)}$$

$$= \prod_n \sum_{\pi_{\mathcal{M}}(n)} e^{-\beta \mathcal{E}(n, \pi_{\mathcal{M}}(n))}$$

Ideal gas approximation

MCMC

$$A(\mathcal{M}_{i_{\text{new}}} | \mathcal{M}_{i_{\text{old}}}) = \min \left\{ \frac{P(\mathcal{D}_i | \mathcal{M}_{i_{\text{new}}}) P(\mathcal{M}_{i_{\text{new}}} | \beta_i, \mathcal{M}^*) Q_i(\mathcal{M}_{i_{\text{old}}} | \mathcal{M}_{i_{\text{new}}})}{P(\mathcal{D}_i | \mathcal{M}_{i_{\text{old}}}) P(\mathcal{M}_{i_{\text{old}}} | \beta_i, \mathcal{M}^*) Q_i(\mathcal{M}_{i_{\text{new}}} | \mathcal{M}_{i_{\text{old}}})}, 1 \right\}$$

$$A(\beta_{i_{\text{new}}} | \beta_{i_{\text{old}}}) = \min \left\{ \frac{P(\mathcal{M}_i | \beta_{i_{\text{new}}}, \mathcal{M}^*)}{P(\mathcal{M}_i | \beta_{i_{\text{old}}}, \mathcal{M}^*)}, 1 \right\}$$

$$A(\mathcal{M}_{\text{new}}^* | \mathcal{M}_{\text{old}}^*) = \min \left\{ \prod_{i=1}^I \frac{P(\mathcal{M}_i | \beta_i, \mathcal{M}_{\text{new}}^*)}{P(\mathcal{M}_i | \beta_i, \mathcal{M}_{\text{old}}^*)}, 1 \right\}$$

Empirical evaluation

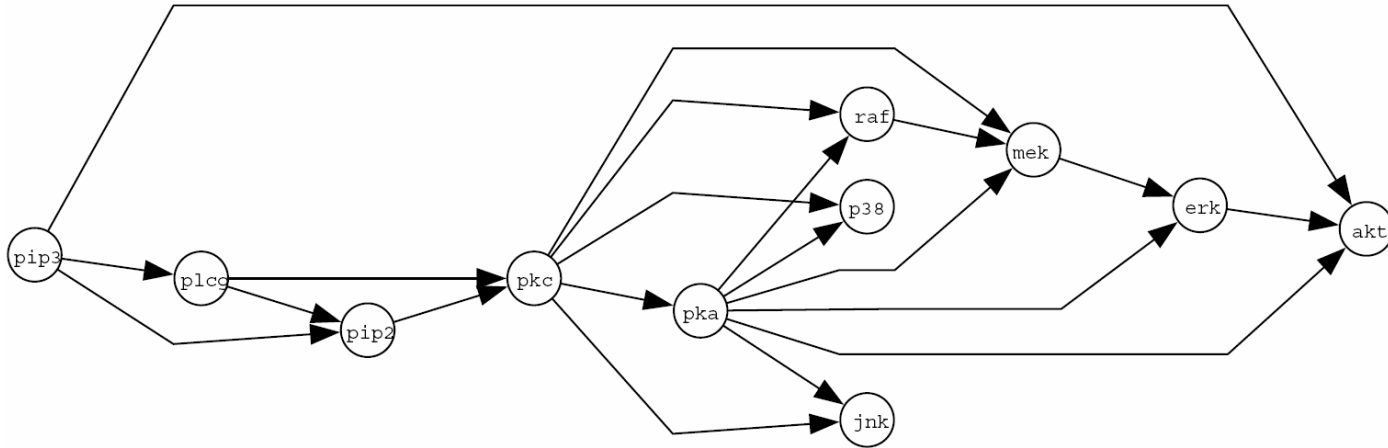
Real application: macrophages
infected with CMV and pre-treated
with IFN- γ

No gold-standard

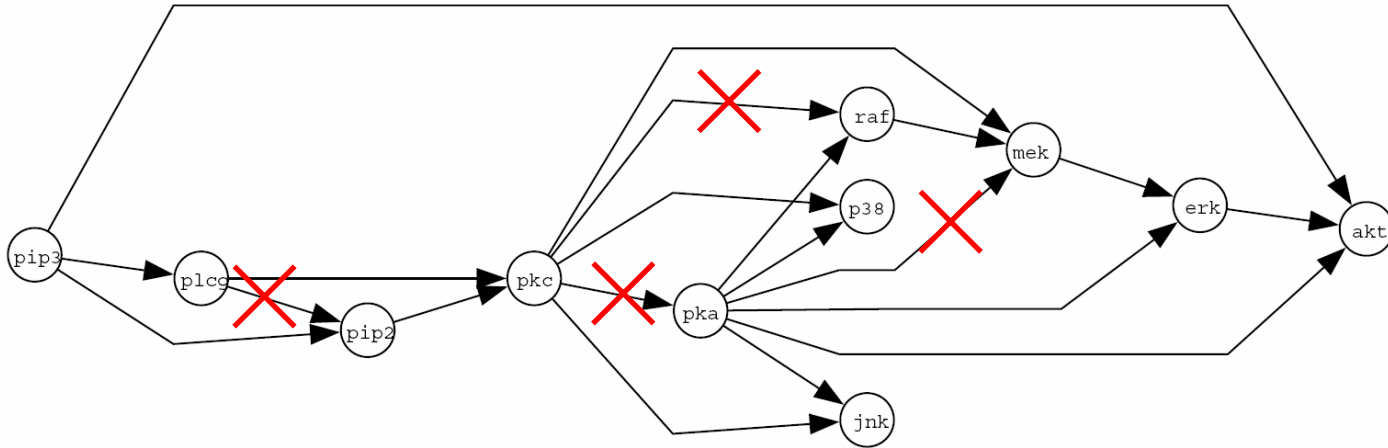
**Simulated data from the Raf
signalling network**

Simulated data

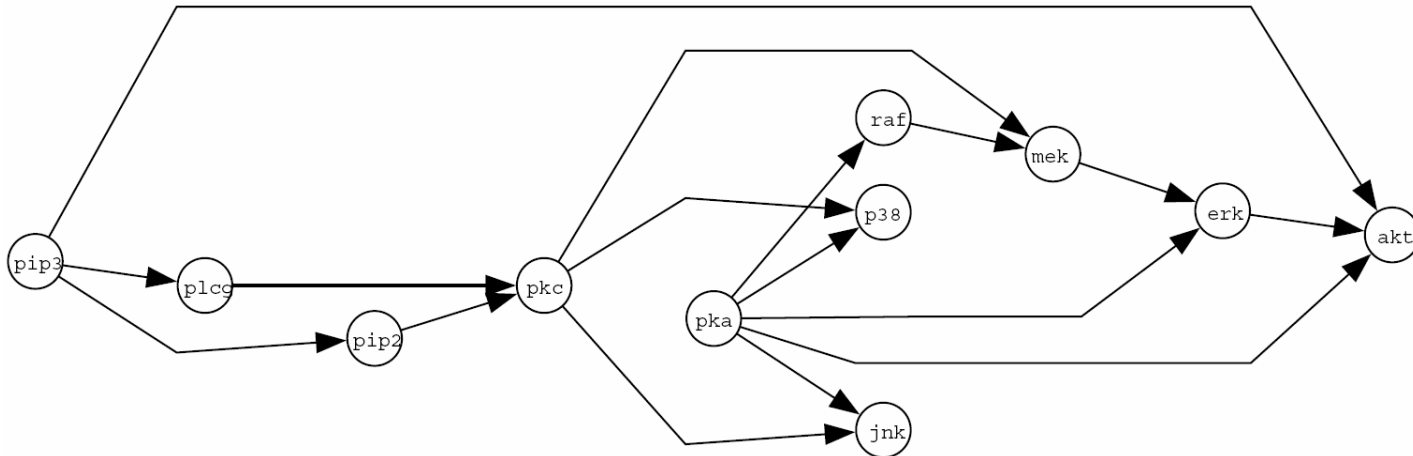
Raf network



Simulated data

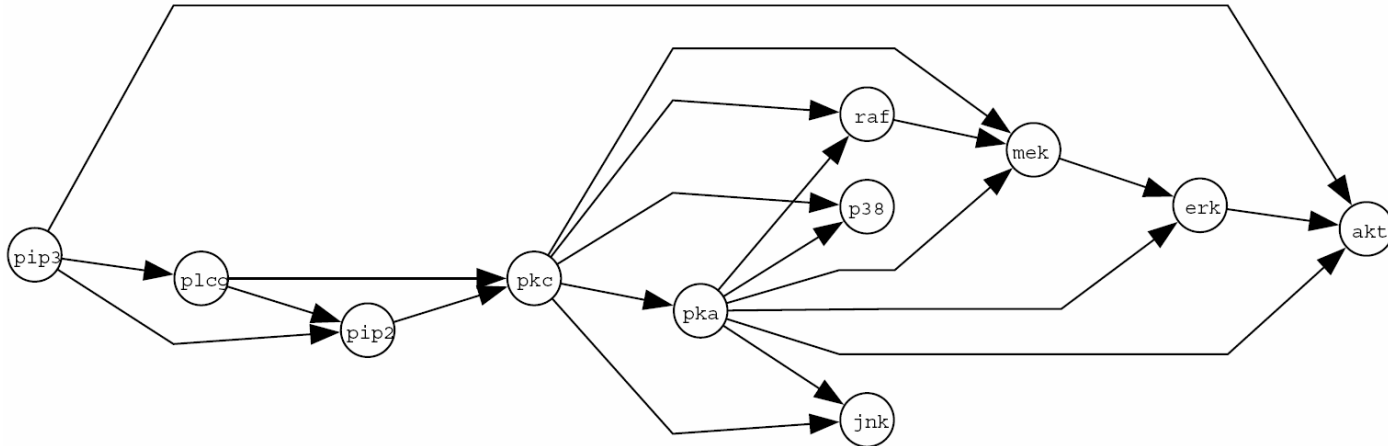


v-Raf network

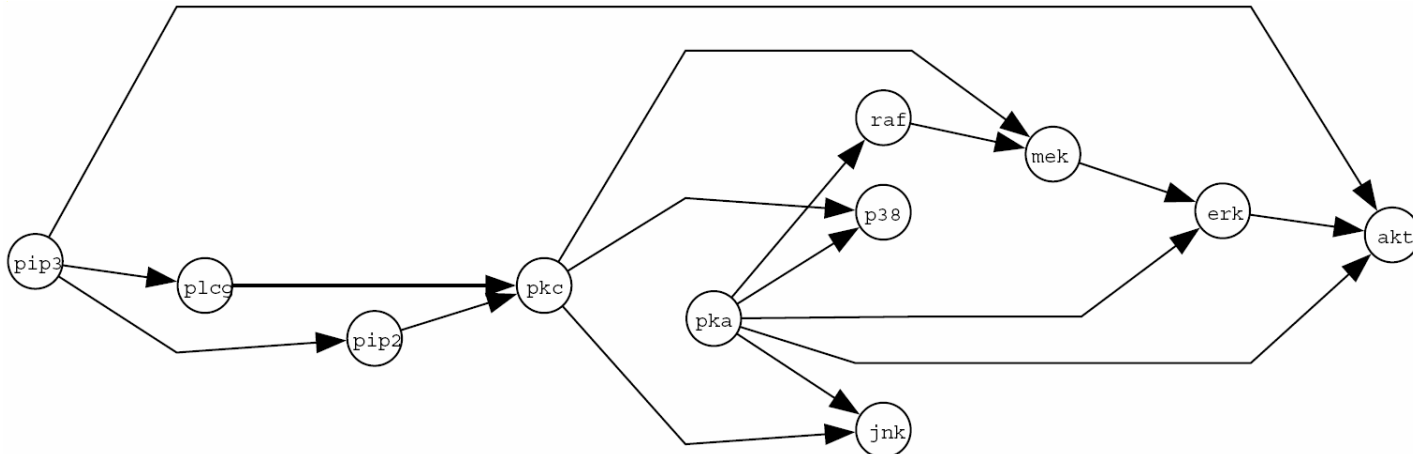


Simulated data

Raf network



v-Raf network



Simulated data

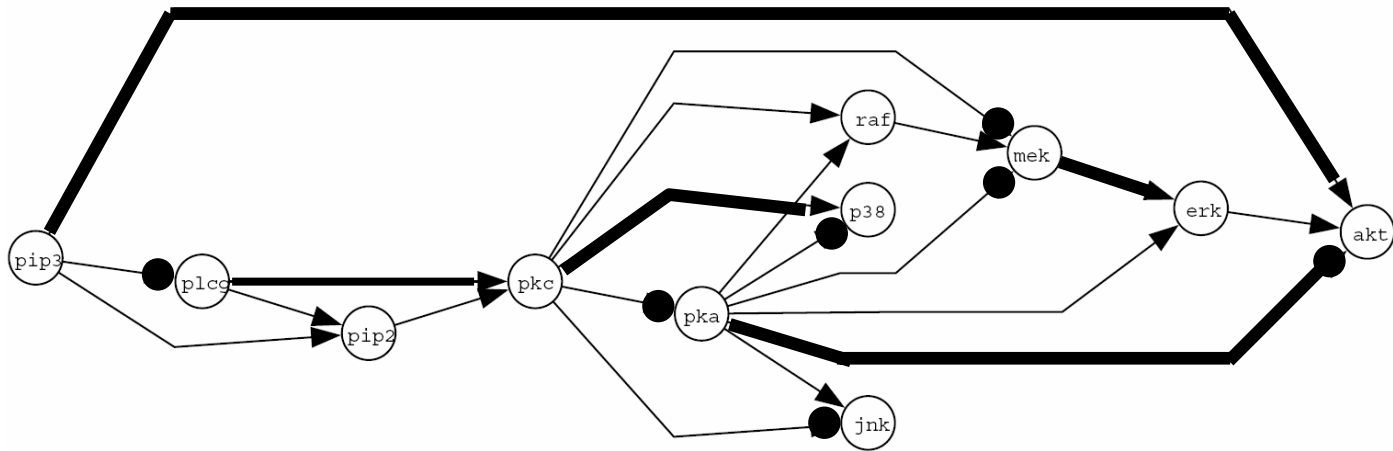
$$X_i \sim N\left(\sum_k w_{ik} x_k, \sigma\right)$$

$$\sigma = 0.1$$

$|w_{ik}|$ uniform distribution over the interval $[0.5, 2]$

Simulated Data

Weights between nodes are **different** for different data sets.

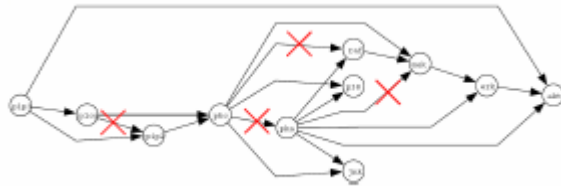


5 data sets

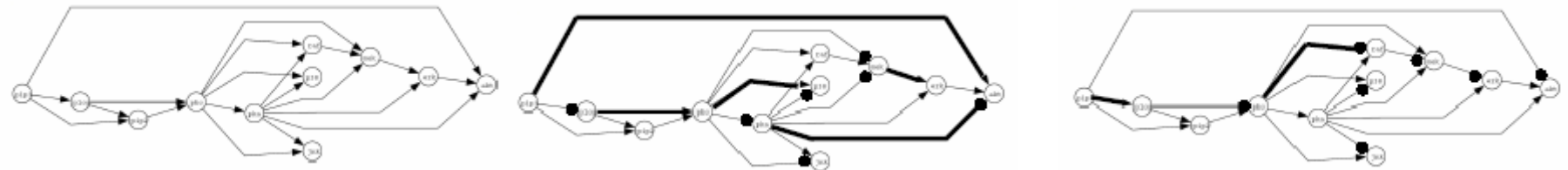
100 data points each



1 random data set (pure noise)

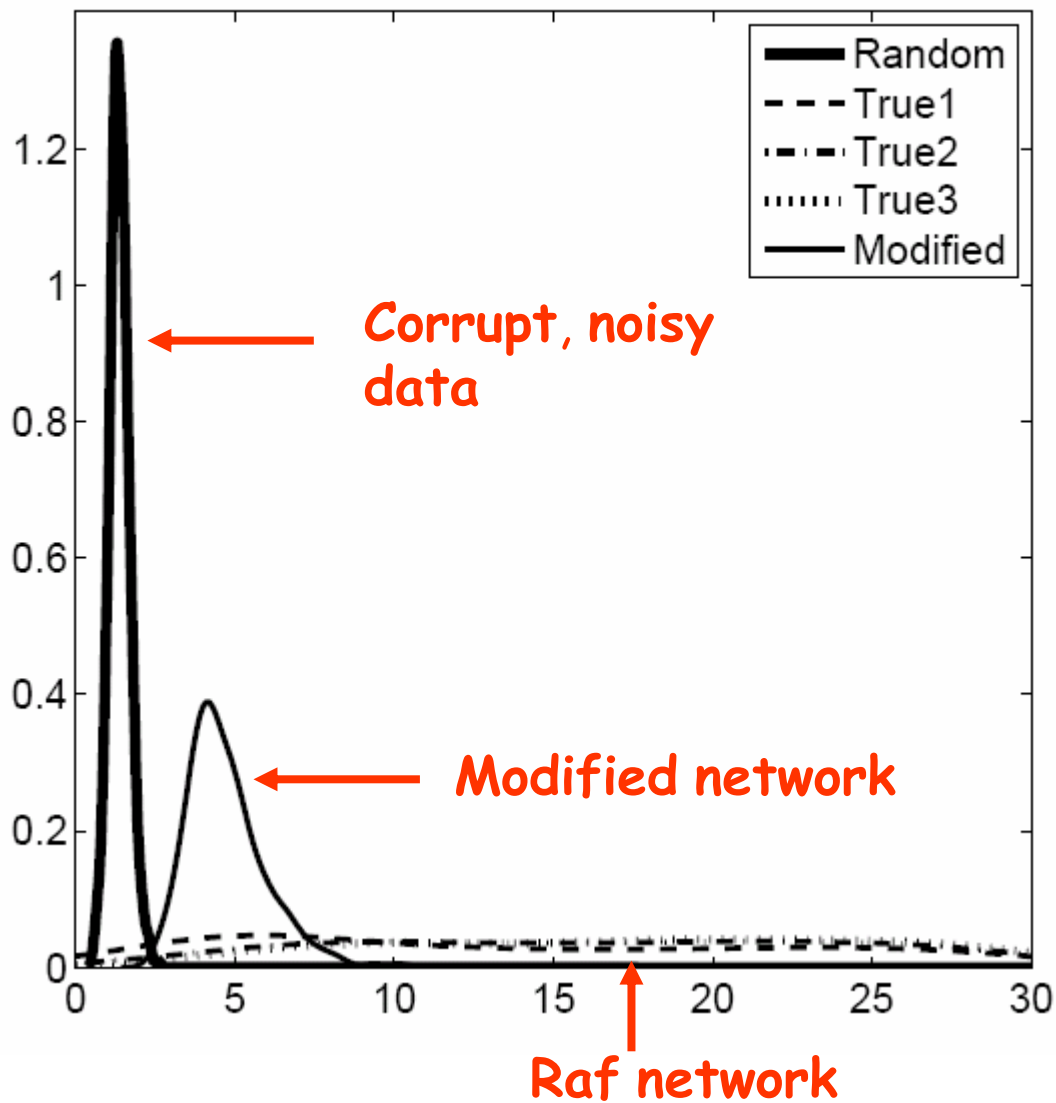


1 data set from the modified network



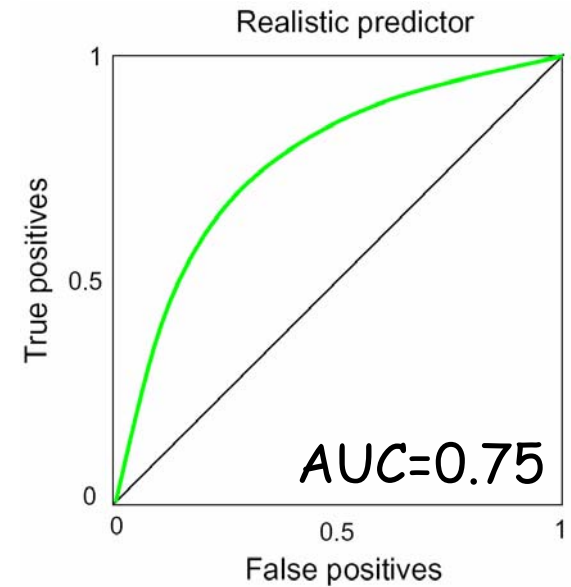
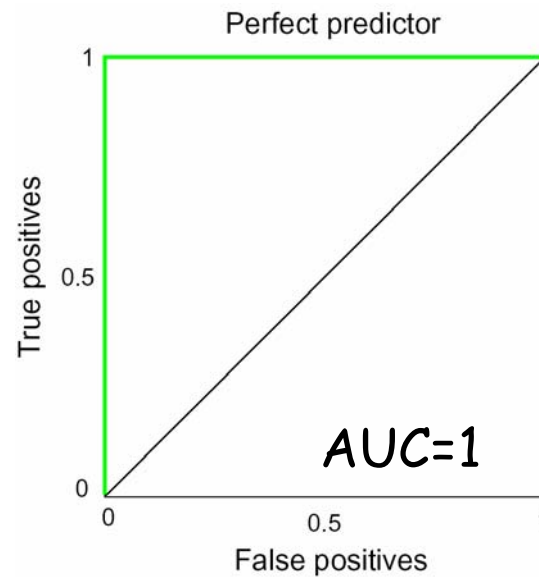
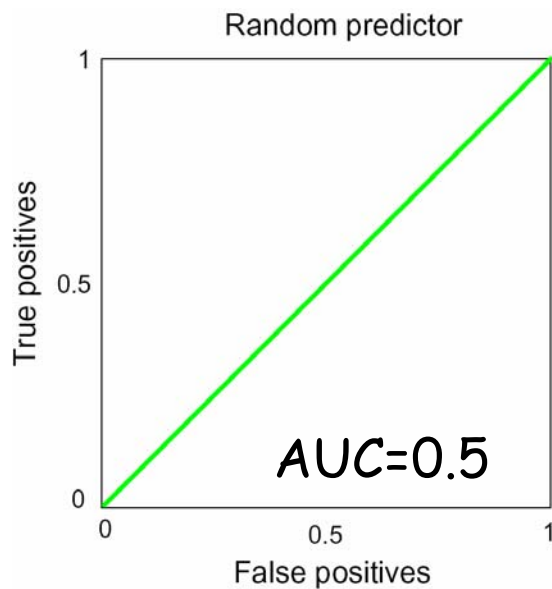
3 data sets from the Raf network, but with different regulations strengths

Gaussian data

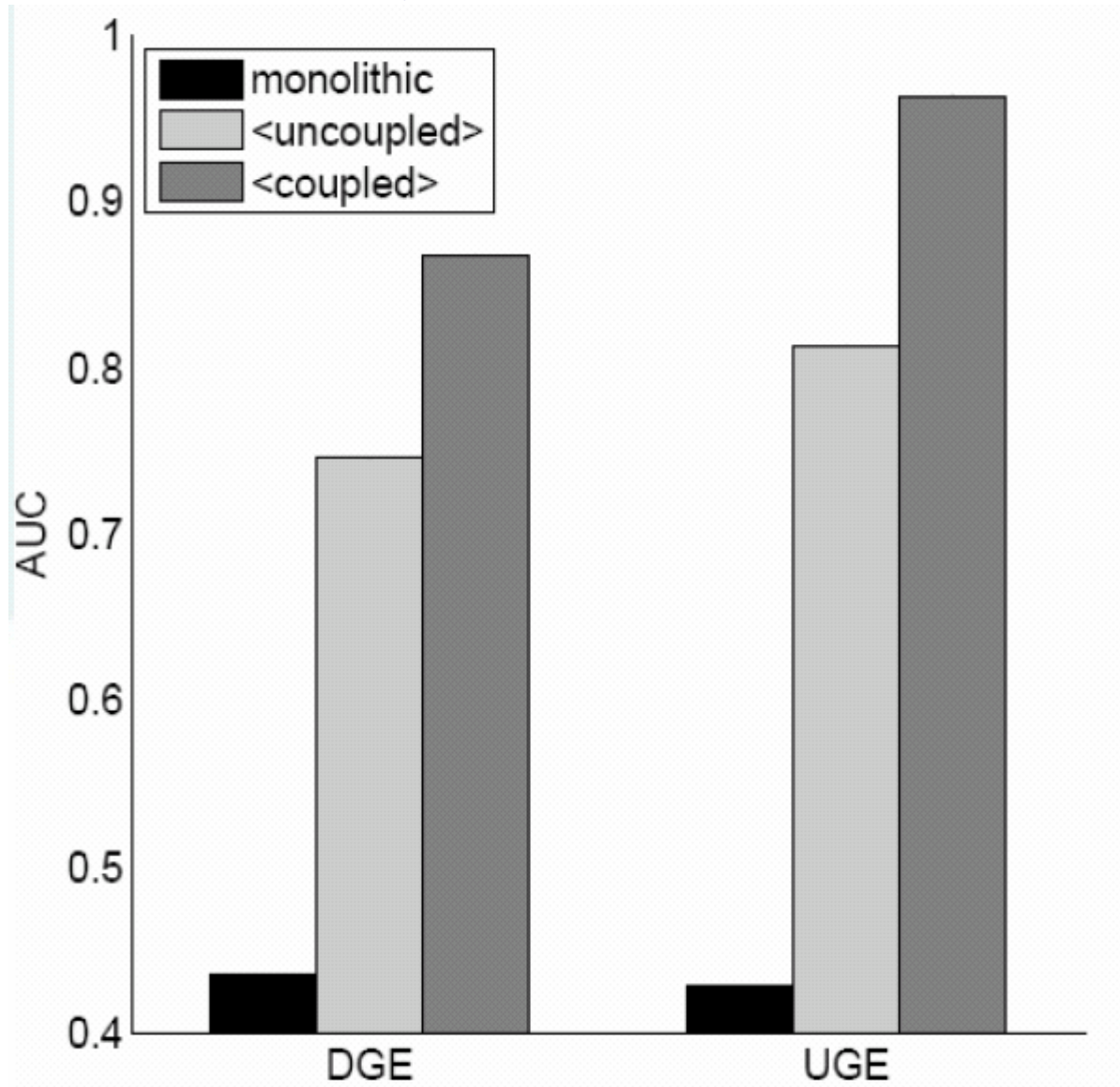


Evaluation of the network reconstruction performance

- We use the Area Under the Receiver Operating Characteristic Curve (ROC).
- ROC curves:



Simulated data



Thank you!



Any questions?