

Genetic network identification using convex programming

A Agung Julius*, Michael Zavlanos**, Stephen Boyd*** and George J Pappas**

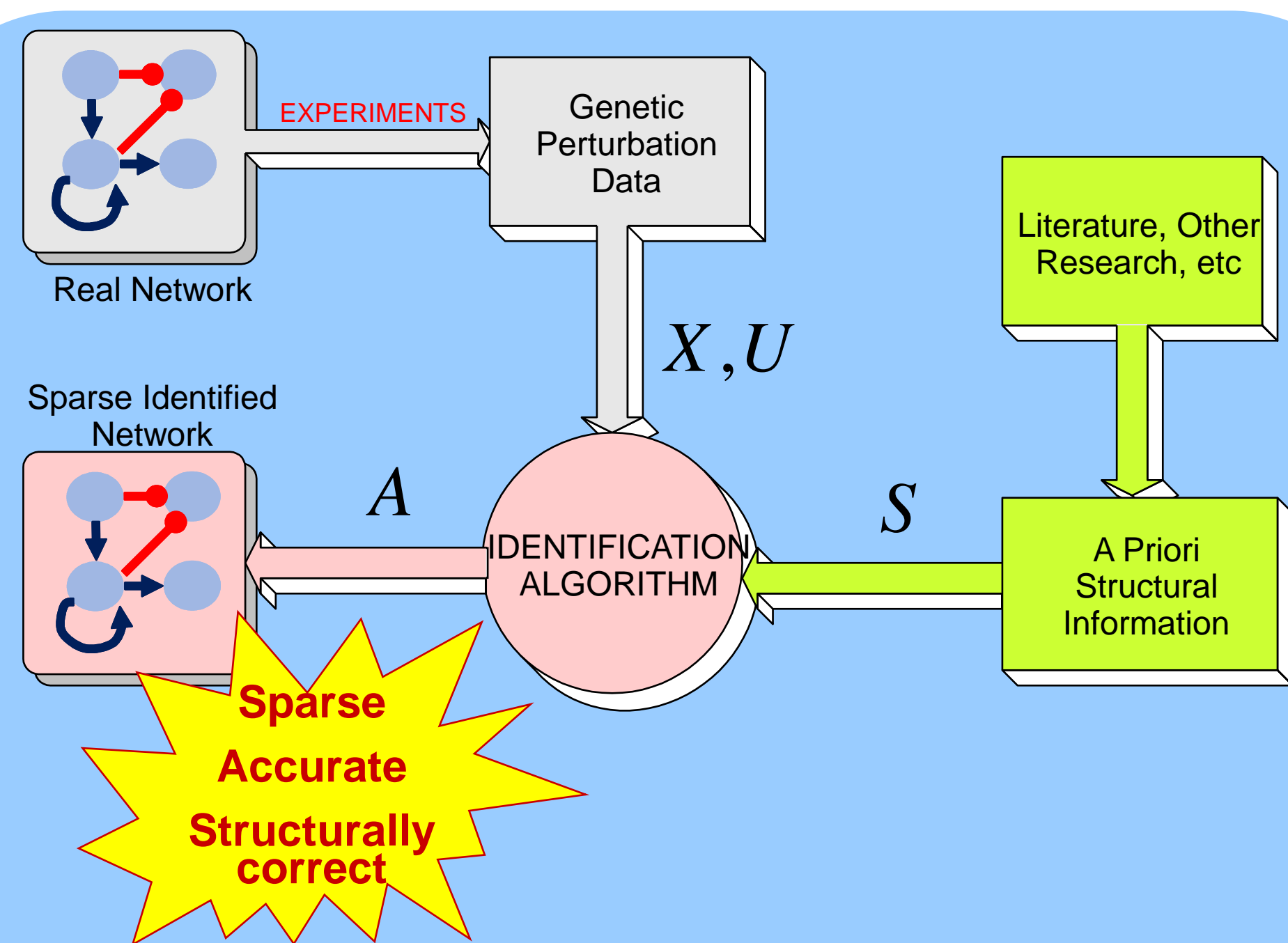
*Dept. Electrical, Computer and Systems Eng., Rensselaer Polytechnic Institute, Troy, USA

**Dept. Electrical and Systems Eng., University of Pennsylvania, Philadelphia, USA

***Dept. Electrical Eng., Stanford University, Stanford, USA

Problem Formulation:

Given a data set from a **genetic perturbation** data, and **a priori structural information**, identify the **sparsest** gene-gene network that explains the data.



The model:

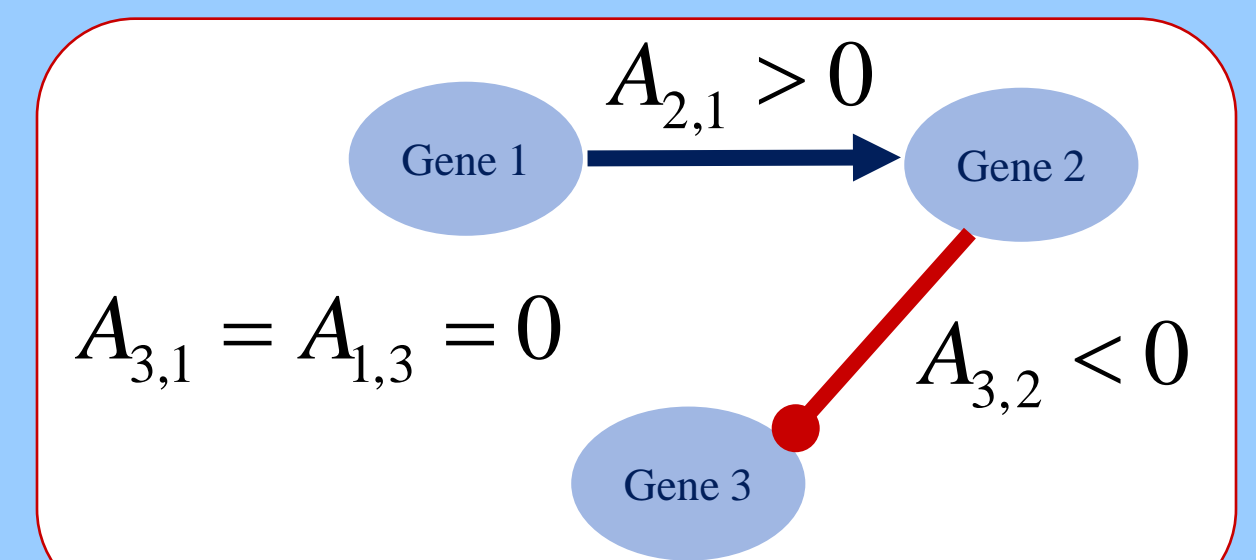
A genetic regulatory network consisting of n genes in a genetic perturbation experiment can be modeled as an n -dimensional dynamical system.

$$\frac{d\hat{x}}{dt} = f(\hat{x}, u), \hat{x} \in \mathcal{R}^n, u \in \mathcal{R}^p.$$

The state \hat{x}_i denotes the **transcription activity** (typically measured as mRNA concentration) of gene i in the network, and u is the so called **transcription perturbation**. The dynamics close to a given equilibrium x_{eq} can be approximated by a set of **linear differential equations**.

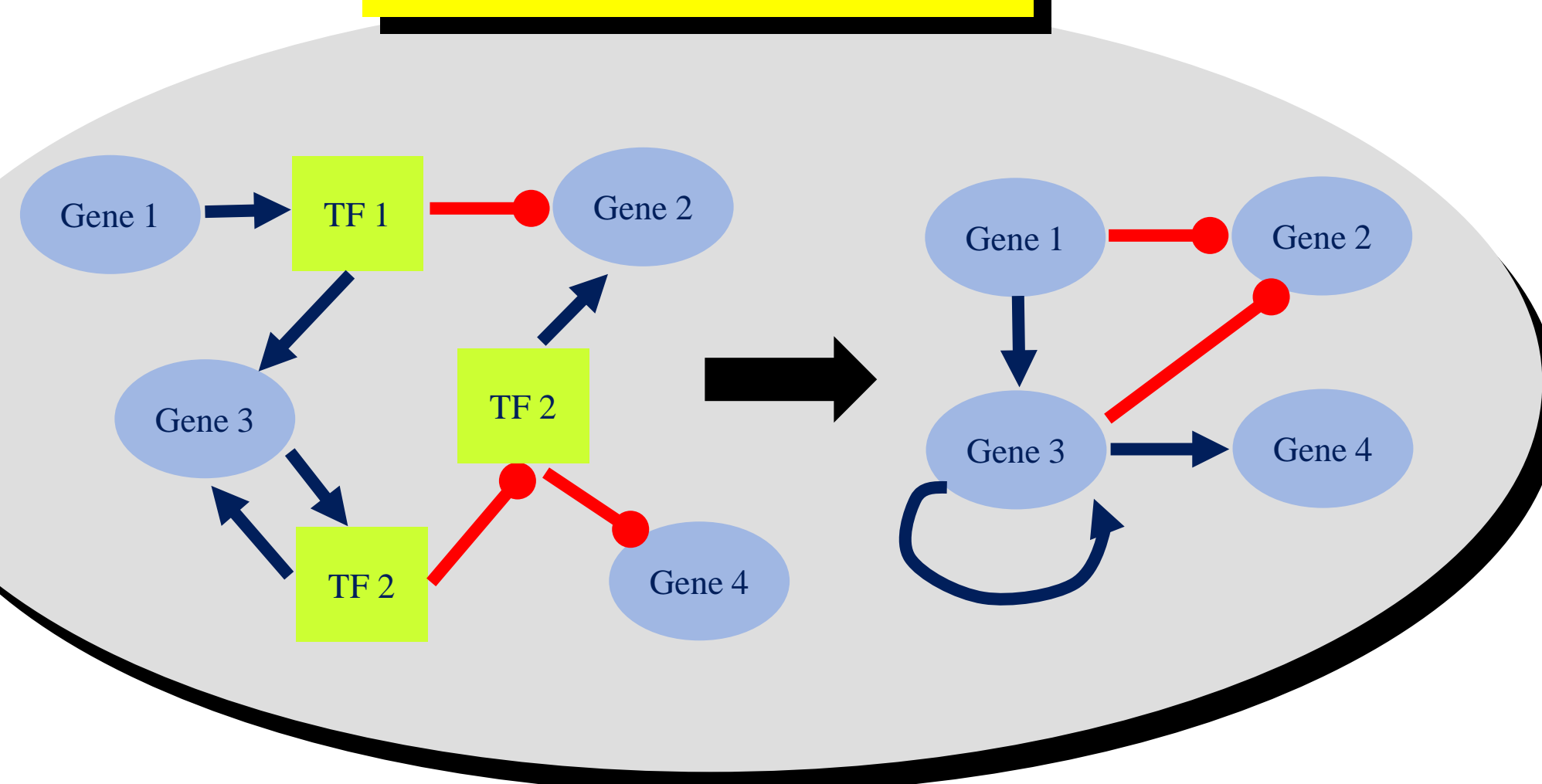
$$\frac{dx}{dt} = Ax + Bu, x := \hat{x} - x_{eq}.$$

The matrix A encodes the **pairwise interaction between the individual genes** in the network at the given equilibrium (phenotypical state), while matrix B indicates which genes are affected by the transcriptional perturbations.



See also: [Gardner03],[Sontag04],[Tegner03]

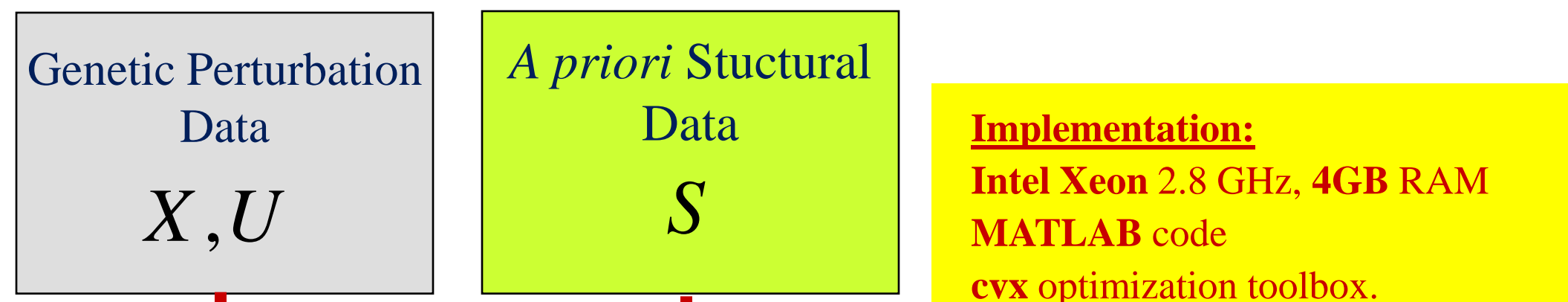
Effective gene-gene network



The approach:

- We cast the problem as a **convex programming** problem using ℓ_1 relaxation.
- A priori* structural information can be included as **convex constraints**.
- No *a priori* limit on connectivity.
- Statistical information about **noisy measurements** is incorporated in the cost function.

See also: [Boyd04]



Step 1:

Obtain baseline error level E_{bs} , by solving

$$\text{minimize } J(A), \text{ subject to } A \in S.$$
 where $J(A) = \sum_j \eta_j^T R_j \eta_j$, $\eta_j = AX_j + BU_j$.

Step 2:

Find the sparsest network by iteratively solving

$$\text{minimize } \sum_{ij} W_{ij} |A_{ij}|, \text{ subject to } A \in S, J(A) \leq \beta E_{bs}.$$

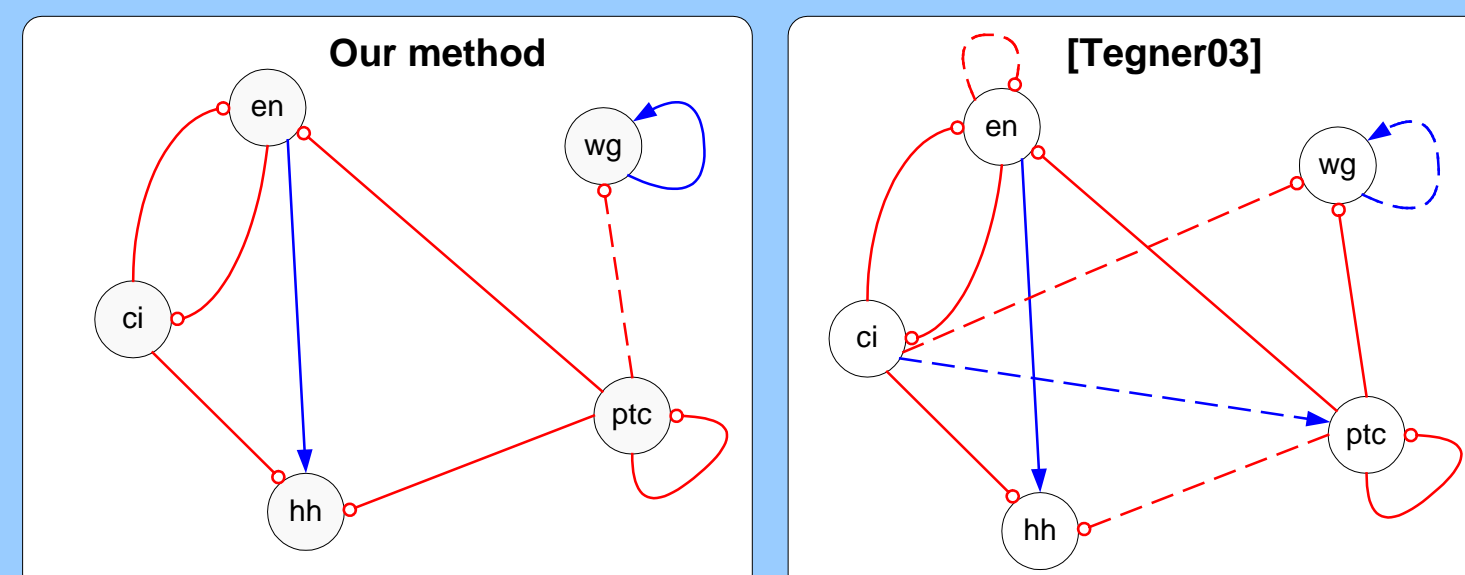
Step 3:

Update the weight
$$W_{ij} = \frac{1}{1 + 1000|A_{ij}|}.$$



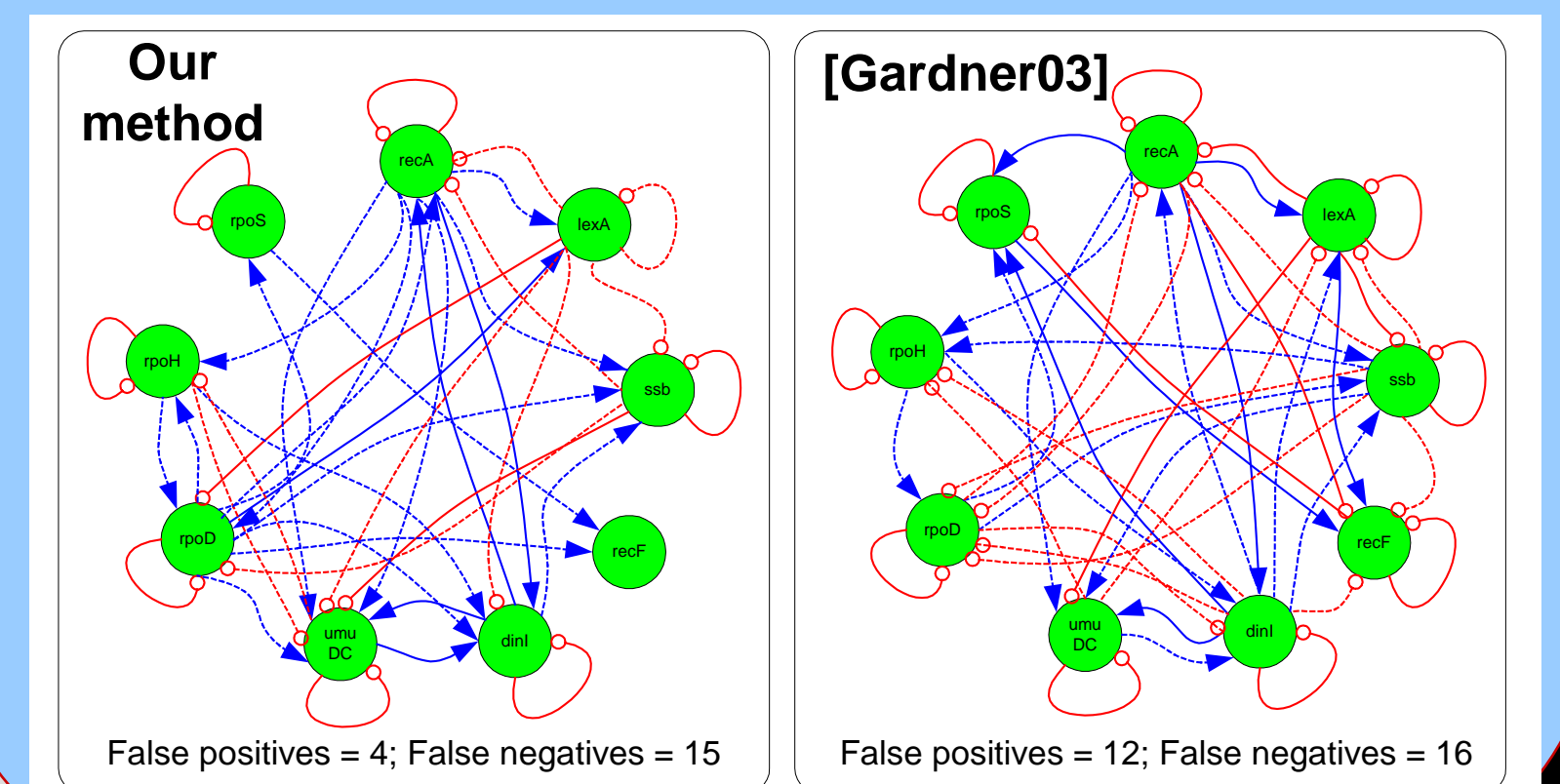
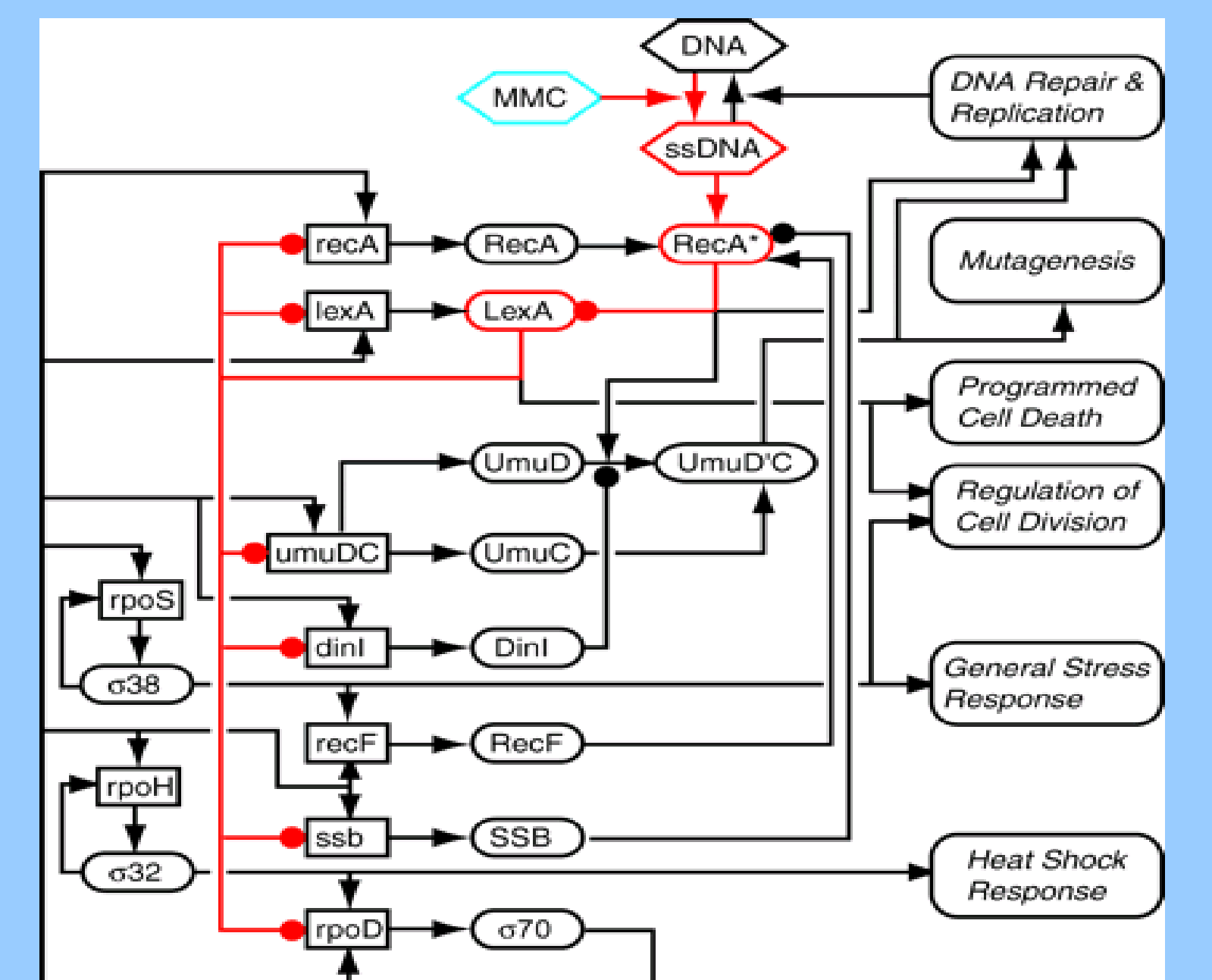
Segmentation polarity network of *Drosophila melanogaster*

We obtain a data set from an *in silico* model provided by [Tegner03]. The original network consists of 5 genes. Our method takes 6 seconds to run and identifies a smaller model than that in [Tegner03] with **higher accuracy** (fewer false positives).



The SOS pathway of *Escherichia coli*

The original network is a part of the SOS pathway in *E. coli*. We obtain an experimental data set from [Gardner03]. The data set consists of 9 genes.



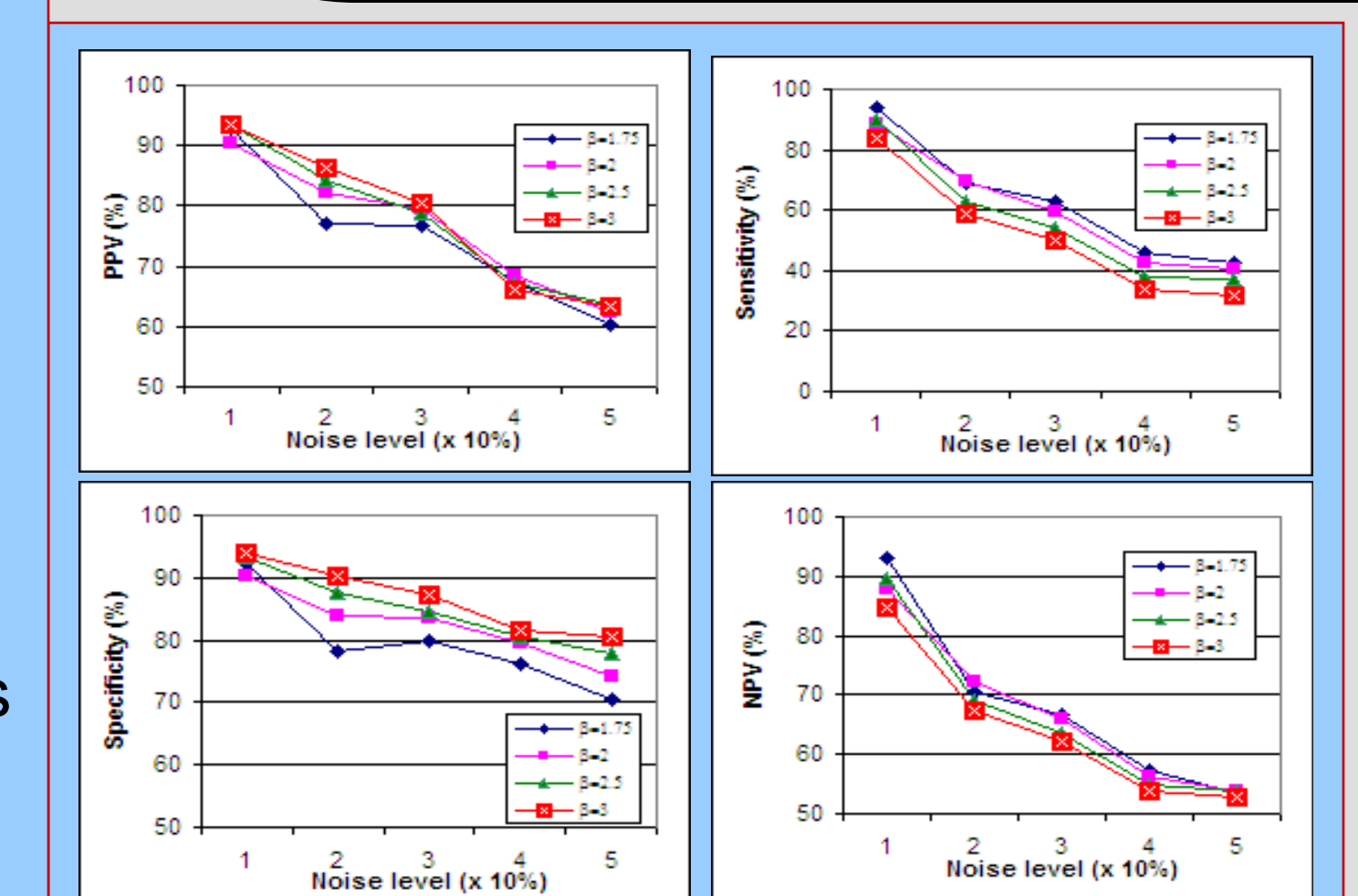
Artificial random network

We construct an artificial random network of 20 genes and generate noisy data sets with different noise levels. We use these data sets to test the performance of our methods. As performance measures, we choose:

$$\text{Sens} = \frac{TP}{TP + FN}, \quad \text{Spec} = \frac{TN}{TN + FP},$$

$$\text{PPV} = \frac{TP}{TP + FP}, \quad \text{NPV} = \frac{TN}{TN + FN}.$$

Our method takes about 9 minutes to run, and for the noise level of 10%, it produces a result with predictive positive values and sensitivity of higher than 90%. This is better than the benchmark results from other methods reported in [Bansal07].



References

- [Bansal07] M. Bansal, V. Belcastro, A. Ambesi-Impiombato, D. di Bernardo, How to infer gene networks from expression profiles. *Molecular Systems Biology*, 10.1038/msb4100120, 2007.
- [Boyd04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [Gardner03] T.S. Gardner, D. di Bernardo, D. Lorenz, J. J. Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, 301:102-105, 2003.
- [Sontag04] E. Sontag and A. Kiyatkin and B. N. Kholodenko. Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. *Bioinformatics*, 20(12):1877-1886, 2004.
- [Tegner03] J. Tegner and M. K. S. Yeung and J. Hasty and J. J. Collins. Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. *Proc. National Academy of Science*, 100(10):5944-5949, 2003.