

IADGRN: Inference Algorithm of Gene Regulatory Networks

IAD Algorithm

1. ARACNE algorithm. MI and DPI

$$MI(i,j) = H(i) + H(j) - H(i,j)$$

$$MI(i,j) \leq \min \{ MI(i,k), MI(k,j) \}$$

$$Recall = \frac{TP}{TP + FN} \quad Precision = \frac{TP}{TP + FP}$$

2. Boolean networks: GRN's models

$$\sigma_n(t+1) = f_n(\sigma_{n1}(t), \dots, \sigma_{nk}(t)), \forall n$$

3. Computational complexity:

N genes, R runs, k connectivity, T thresholds, P multiple inputs tested

Pairwise Inference $\sim O(T.R.N^2)$

DPI Test $\sim O(T.R.N^3)$

Boolean Function Inference $\sim O(R.N.2^k)$

Extra Input Test $\sim O(T.R.N^2.2^k)$

Multiple Input Test $\sim O(T.R.N.C_p^N)$

Limits: 10.000 genes, 20 inputs/gene at most for Boolean functions.

1. First inference stage

$$MI[A(t), B(t+1)] = H(A(t)) + H(B(t+1)) - H[A(t), B(t+1)]$$

$$MI[A(t), B(t+1)] \leq \min \{ MI[A(t), C(t+1)], MI[C(t), B(t+1)] \}$$

.Time series, Threshold in IAD, directional connections

2. Second inference Stage

“Flip Frozen” test,
“Boolean functions inference”,
“Extra Inputs” test
“Multiple inputs” test
characterization of the failures

$$H_{ratio}^{dist} = \frac{\sum_{i=1}^n \left(\frac{\sum_{j=1}^{\#inputs(i)} H^{dist}(i, j)}{2^{\#inputs(i)}} \right)}{N - N(\text{no functions})}$$

$$MI_{(i,j),z} - MI_{i,z} - MI_{j,z} > MI_{(i,j),z}^{treshold} - MI_{i,z}^{treshold} - MI_{j,z}^{treshold}$$

3 Inference with experimental confirmation

inferences to test ($P = TP+FP$) to check each step is:

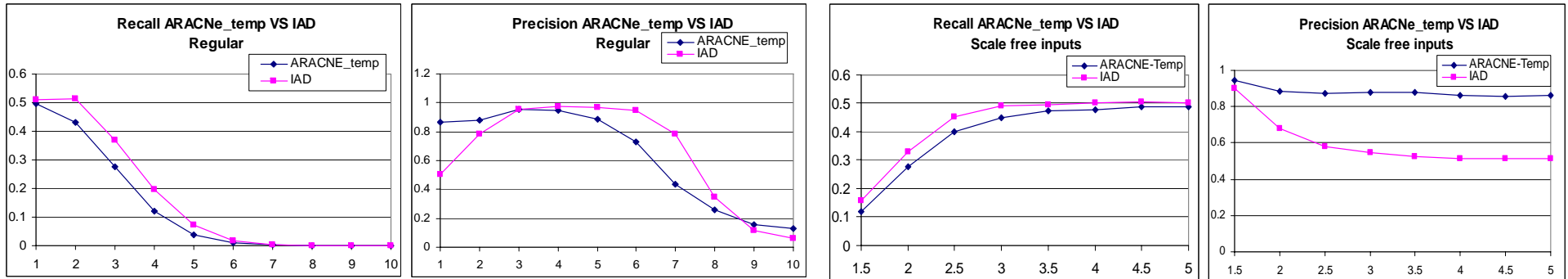
$$\# Exp(j) = \frac{P_j}{(P_1 \cup \dots \cup P_{j-1})}$$

$$TP_{ratio}(i) = \frac{TP_{newly\ found}(experiment(i))}{N^2}$$

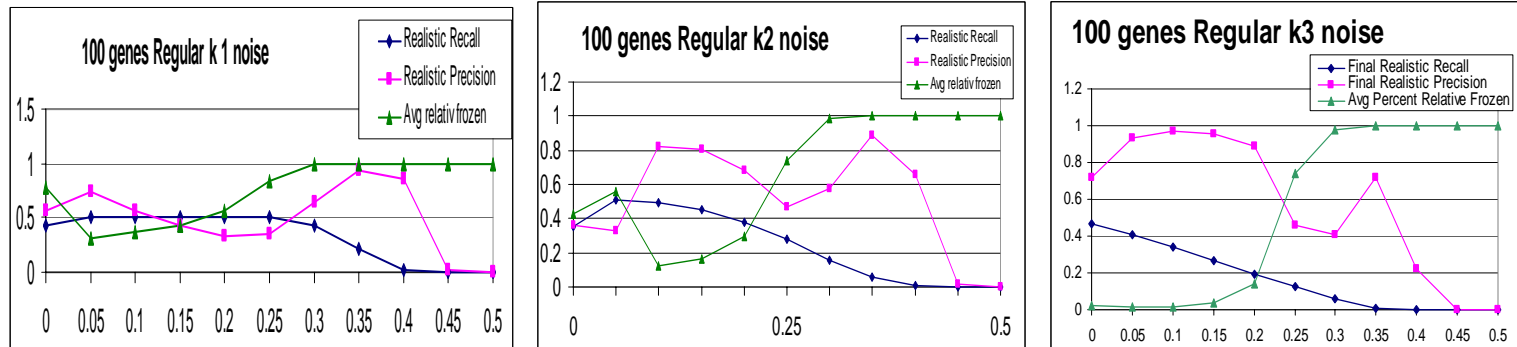
IADGRN: Inference Algorithm of Gene Regulatory Networks

Results

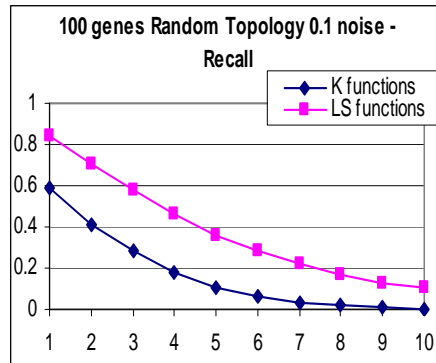
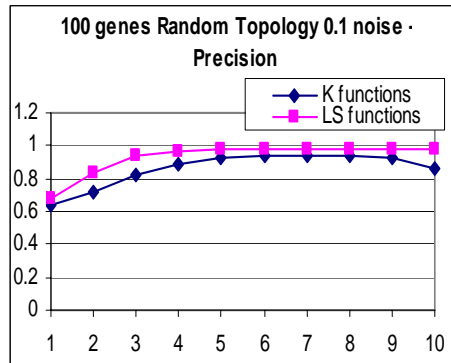
5.1 IAD vs “ARACNe temporal”



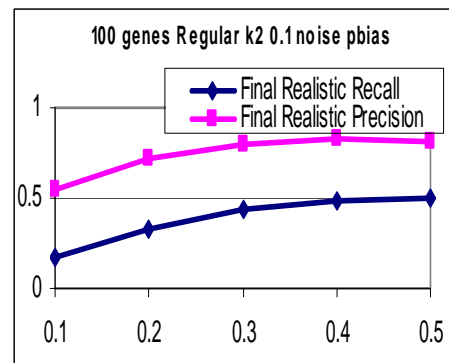
5.2 Noise effect



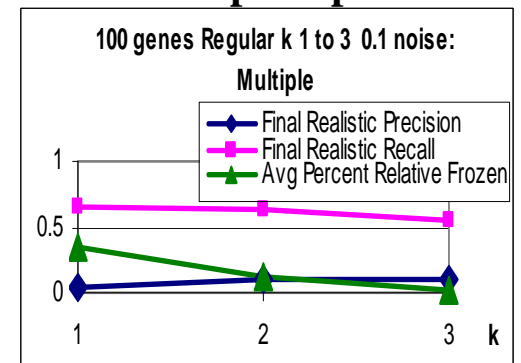
5.3 Transfer functions



5.4 P-bias



5.5 Multiple inputs test

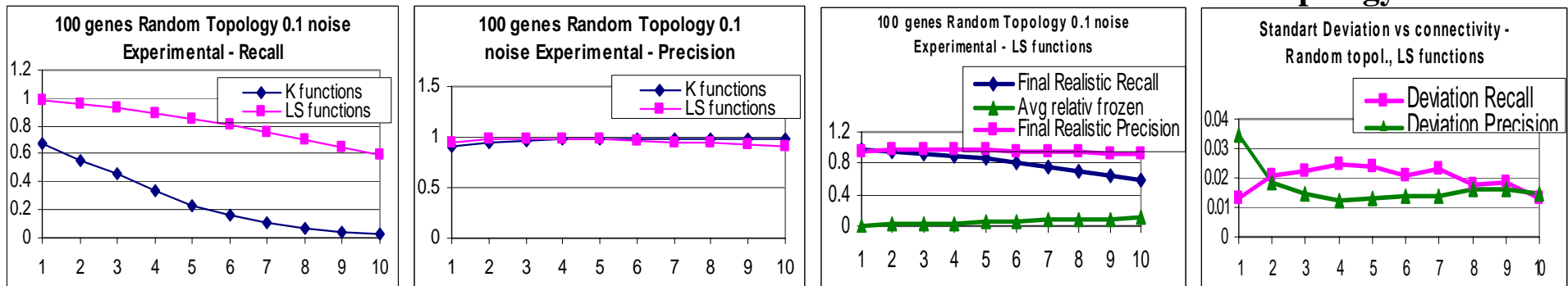


IADGRN: Inference Algorithm of Gene Regulatory Networks

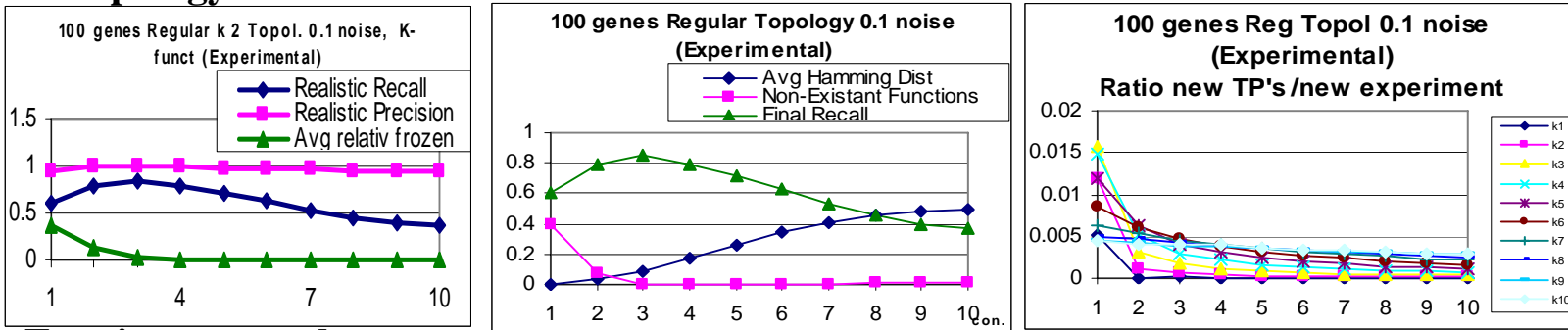
Results

5.6 Experimental inference tests

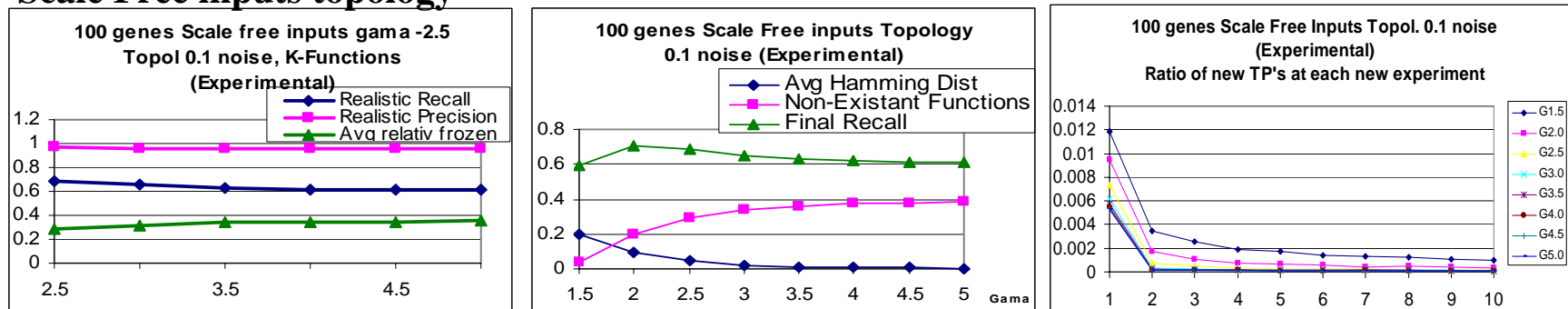
Random topology



Regular topology



Scale Free inputs topology





IADGRN: Inference Algorithm of Gene Regulatory Networks

Conclusions

- 1) IAD is more accurate than previous methods of inferring RBN's of more than a few nodes.
- 2) Combining IAD with realistic experimental testing allows inferring the whole inferable network, with minimized costs, if compared to other methods.
- 3) Several post inference tests have improved significantly.
- 4) Assess predictions by experimental tests, is feasible.

Future developments

- 1) Confront IAD to real gene expression time series and model noise with Gillespie nets as model of GRN's.
- 2) Find states discretization between each pair of genes that maximizes MI.
- 3) Determine optimal delay between the input and output gene that maximizes MI.
- 4) Explore the effect of different sampling intervals for hypothetical gene expression array data (the Nyquist frequency, twice the fastest frequency, if possible, since it may be very expensive or impractical in real gene arrays).
- 5) Combine IAD with other methods such as Dynamic Bayesian Networks (not applicable to huge phase spaces).
- 6) Determine real GRN's microarray data noise in order to take it into consideration during the inference process.