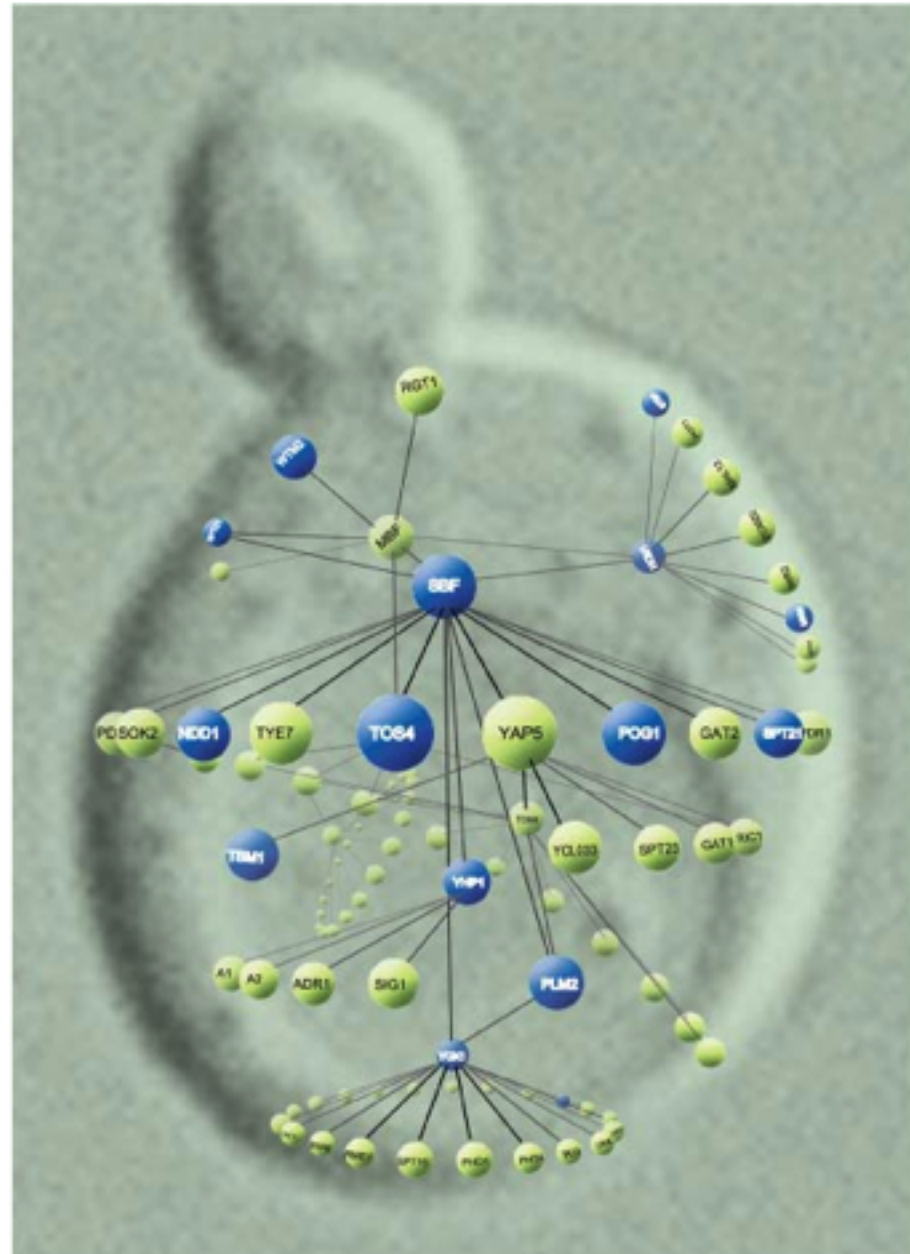


Gene Network Detection using Directed Mutual Information

Pramod Mathai (AE), Nuno Martins (EE, ISR), Benjamin Shapiro (AE, ISR)

1



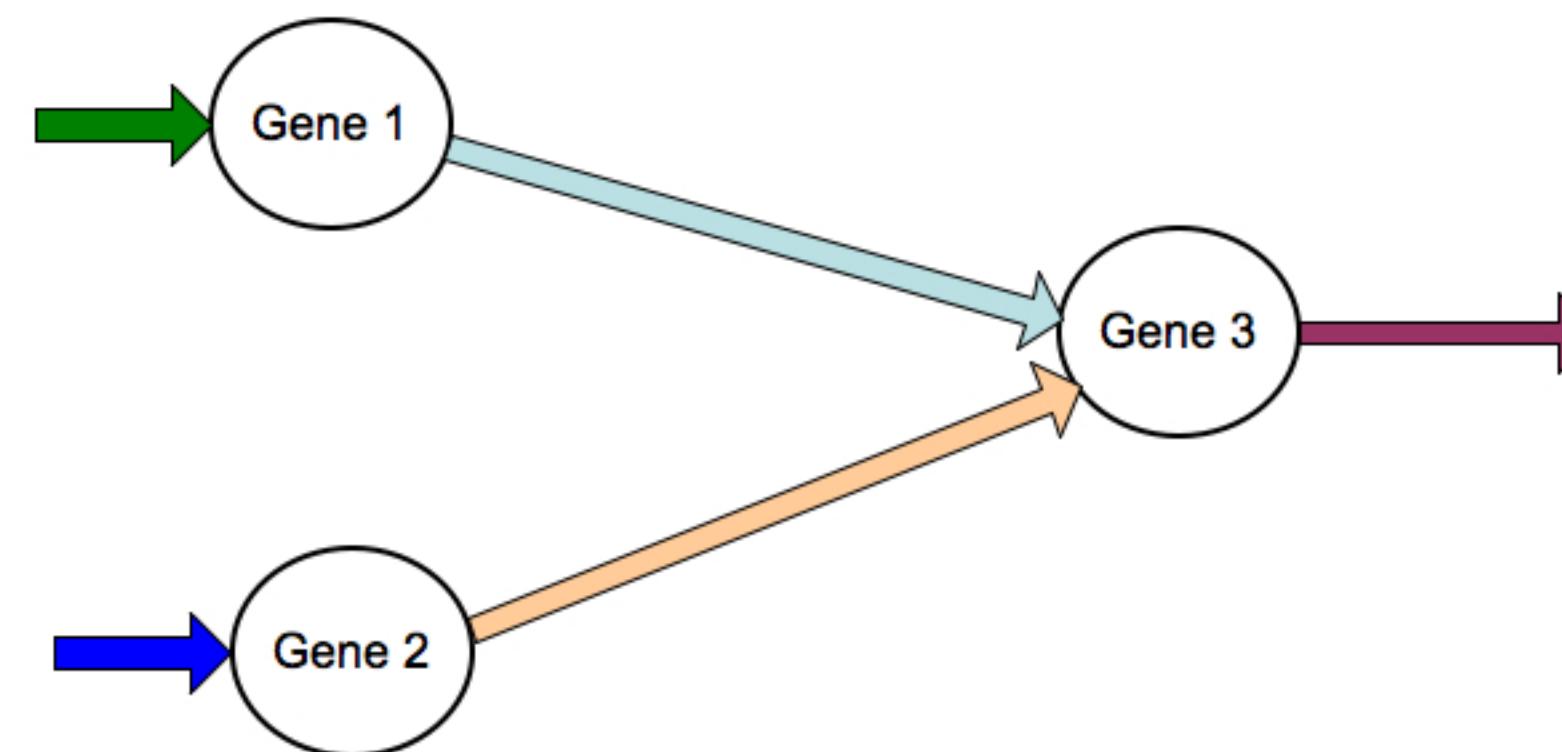
Biological problem :

Can we detect a causal influence (directed edge) between two genes (nodes) in a gene network?

Applications:

Screening and classification of diseases, genomic medicine.

2



The arrows show the **causality relationship**, which indicates the rate at which the protein encoded by gene 3 depends on the rates at which proteins are encoded by genes 1 and 2.

Assume:

- Discrete-time, Markovian network.
- We have access to noisy measurements of gene expression for an interval of time.

$$Gene_1(n), Gene_2(n), Gene_3(n)$$

3

Can we find meaningful statistics for obtaining such a directed graph, keeping in mind that *we are not interested in the particular functional form* ?

Knowing the graph will give us information about the dynamics of protein production and control.

4

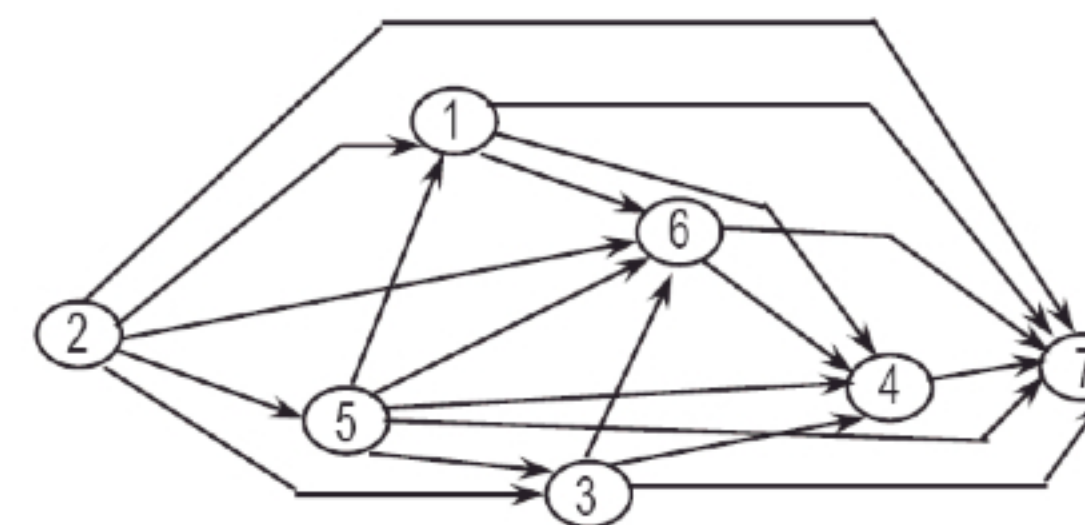
$$\text{IF } \frac{I((Gene_1(1), \dots, Gene_1(P)) \rightarrow (Gene_2(1), \dots, Gene_2(P)))}{I((Gene_1(1), \dots, Gene_1(P)); (Gene_2(1), \dots, Gene_2(P)))} > \sigma_0$$

← This limit determines strength of causal effect.

$$\text{AND IF } I((Gene_1(1), \dots, Gene_1(P)); (Gene_2(1), \dots, Gene_2(P))) > I_0$$

← This limit takes care of noise in Gene-Gene interactions

THEN infer that $Gene_1$ causally influences $Gene_2$.



Example of randomly generated directed (acyclic) network that has been successfully determined using our method with parameter $P = 2$. The model for gene dynamics and function is the one found in [1].

Our method was also successful in identifying the gene network of the flagella class 2 system in E.Coli, a network with 9 genes. The method worked even in cases where a gene is influenced by other genes according to pre-specified

Future work: Incorporate efficient entropy computation for higher 'P', statistically characterize limits for edge strength and noise, decrease required number of data sets.

Using information theory, the knowledge of past values of $Gene_1$ may lead to a significant reduction in the "uncertainty" regarding $Gene_2$.

Mutual information is symmetric, limiting its use in determining directionality of dependence relationships.

We use a combination of directed *and* mutual information to detect a causal relationship between two genes.