

Simulation data were generated by taking the following steps.

1. Generate the labels for each node in the graph
 - (a) Initialize all the labels C_i randomly to either 0 or 1 (with $Pr(C_i = 1) = 0.20$).
 - (b) For every gene node i , calculate

$$P_{i1} = \prod_{\substack{j \in N(i) \\ C_j \text{ is defined}}} \psi_{ij}(1, C_j)$$

$$P_{i0} = \prod_{\substack{j \in N(i) \\ C_j \text{ is defined}}} \psi_{ij}(0, C_j)$$

where $N(i)$ are the neighboring gene nodes of node i , and generate the label C_i from Bernoulli distribution with $p_i = \frac{P_{i1}}{P_{i0} + P_{i1}}$.

- (c) Repeat the last step for 5,000 times (call it a “stage”), and for each gene i calculate its empirical frequency of getting label 1 in this stage (denoted as f_i).
 - (d) Repeat the last step for additional stages, until the change of f_i between stages falls below 0.1 for every gene i . Save the generated labels.
2. Generate scores. Choose two score distributions we want to use for cell cycle genes and non cell cycle genes. Based on the labels for each gene generated in the previous step, generate the scores from one the two distributions.