Improving Gene Regulatory Network Inference Using Network Topology Information

Ajay Nair, Madhu Chetty, and Pramod P Wangikar

Supplementary information

S1 Theorems used in the study

This section derives the proof of the number of networks that can be represented for a graph with n-nodes, under different types of graph representations. These are well known results in the graph theory domain but we could not find their references to cite. Since these results are not widely known in Bioinformatics and Computational Biology and are also important for the subsequent derivations of the complexity of the algorithms proposed, the proofs have been provided here.

Theorem S1. *There are 2^{n^2} possible number of networks for n-nodes in a directed graph representation.*

Proof. A directed graph of n-nodes can be represented by a matrix M of size $n \times n$. The presence of an edge from node i to node j is represented by the value '1' in the cell M_{ij} and '0' otherwise. Therefore, the space of all graphs for n-nodes is the number of all possible subsets of the elements in the matrix $M = \{M_{11}, M_{12}, .., M_{1n}, M_{21}, .., M_{nn}\}$, where $M_{ij} \in \{0, 1\}$. This is $2^{|M|} = 2^{n^2}$. □

Theorem S2. *While using a decomposable scoring function on a directed graph with n-nodes, the number of network configurations required to be checked are $n2^n$.*

Proof. Each node can have n possible parents, including itself. Thus, the number of possible parent combinations for a single node is, $\sum_{i=0}^{n} \binom{n}{i} = 2^n$. For n-nodes this become $n2^n$. □

Theorem S3. *When the number of parents are restricted to a constant k for each node, the search space for an n-node graph is $O(n^k)$.*

Proof. When using a decomposable scoring function, if the number of parents of a node are exactly k, where $k \leq n$, then the number of possible combinations of parents to be considered are given by

$$M = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$\log M = \log n! - \log k! - \log(n-k)!$$

By Sterling's approximation, $\log x! \approx x \log x - x$.

$$
\begin{aligned}
\log M = n \log n - n - [k \log k - k] - [(n-k)\log(n-k) - (n-k)] \\
\log M = \log \frac{n^n}{k^k (n-k)^{n-k}} \\
\text{if } k \ll n; \log M \approx \log \frac{n^n}{k^k (n)^{n-k}} \\
\log M \approx \log \frac{n^k}{k^k} \\
M \approx \frac{n^k}{k^k} \\
O(M) = O(n^k)
\end{aligned}
$$

Similarly, it can be shown that when the number of parents are less than or equal to k (maxPval$= k$),

$$
\begin{aligned}
O(M) &= O(n^k + n^{k-1} + .. + n^1) \\
&= O(n^k)
\end{aligned}
$$

$\square$