

Critique: “A single source k-shortest paths algorithm to infer regulatory pathways in a gene network”

Yu-Keng Shih and Srinivasan Parthasarathy

1. Introduction to regulatory pathways problem

The paper studies pathways in gene regulatory network which represent interaction between genes and proteins, genes and genes and etc., influencing the process of gene expression. It is shown that a given problem has three following subproblems: unknown casual (given a target gene, infer possible casual genes and their pathways), unknown target and candidate casual (infer the most likely casual gene among the candidates). A purposed heuristic algorithm has been tested for all of these subproblems. Authors give an overview of the current state of gene regulatory problem mentioning a random walk based information flow model, which has some limitations dealing with graph edges normalization and bidirectional edges which authors seek to overcome.

2. Critique

The problem of paths in regulatory networks is considered taking in some assumptions as the problem of single-source k-shortest simple paths in the directed weighted graph. The notion of pseudo tree is introduced and it is clear that it allows effective storage of all shorted paths. Some theorems on pseudo trees are nicely proved, although some points are not clear. Represented algorithm uses an idea that pseudo tree can have at most k path nodes corresponding to a single node in the graph which should be explained in light of proved theorems. Despite this fact conclusions made in the paper are logically and mathematically accurate and examples given are illustrative. Algorithm by itself represents adopts a Dijkstra approach to find top-k shortest paths in the graph from a given source. To improve the results in the biological context an algorithm takes into an account a measurement of paths' diversity though it doesn't give much improvement according to shown in primary statistical results. However case study improves e-values (probability of gene form given pathways regulating some property) for various cell functions using different parameters of diversity, interpretation of these results based on the materials of the paper doesn't seem an easy task.

The papers reports the result on yeast gene network. Results demonstrate an enhancement of results of candidate casual prediction over an existing methods. Accuracy field shows clear an improvement of results however a P-value field doesn't seem very representative. Next results concern unknown target and unknown casual. E-values in the results represent the probability of gene form given pathways regulating some property, which is a good metric. Comparison to other approaches also looks illustrative and significant for the reason that it shows limitations of these approaches. It

might be seen that despite all the drawbacks results achieved in these research are of scientific and biological importance.

3 Conclusion

The paper is clearly written although contains some typos, which sometimes are really confusing. Algorithms developed by the recherches gives a significant improvement in the task of target and casual gene distinction in the gene regulatory networks compared to the existing approaches. It still doesn't take an account all possible paths in the cell and the fact that some paths repeatedly appear in the tree, so authors of the paper have a direction for studies concerning all aspects of a given problem.