Evaluation of Network-based Function Prediction Algorithms on *Oryza sativa japonica* Protein-Protein Interaction Data

M. A. S. U. Piyatissa, T. L. S. Tirimanne, S. P. C. Fernando

Department of Plant Sciences, Faculty of Science, University of Colombo, Sri Lanka

Protein-protein interactions (PPIs) in an organism can be represented as a computational network and can be used to predict unknown protein functions using appropriate algorithms. This method is shown to outperform other computational techniques using human data. Several such network-based function predictions exist but the majority of these have only been evaluated on well-annotated species such as yeast and humans. No evaluations have been done for plant PPI networks. To address this deficit, an evaluation methodology was proposed for four well-known network algorithms: Majority voting (MV), Hishigaki method (HM), network propagation (NP), and Markov Random Fields (MRF) for functional annotation of PPI networks of *Oryza sativa japonica*. The PPI network was downloaded from STRING while annotations for Gene Ontology (GO) terms were downloaded from the Gene Ontology database. The algorithms were coded and compiled in Python 3.9. For evaluation, only GO terms that had at least 10 direct or indirect annotations were considered. A ten percent leave-out cross-validation was performed for each GO term for each algorithm. The area under the ROC curve (AUROC), area under the PR curve (AUPRC), and average precision (AP) were obtained for each algorithm to compare their performance. Overall, the four algorithms had similar performances, but the NP method slightly outperformed at AUROC (0.733) and AP (0.289) over the other algorithms (whereas the AUROC and AP for these methods are as follows; MV=0.688 and 0.221, HM=0.630 and 0.283, and MRF=0.712 and 0.275, respectively), indicating its superiority over others for predicting plant protein functions. This evaluation could be extended to other plant crops and algorithms and will be beneficial for bioinformaticians when selecting algorithms for protein function prediction.

Keywords: *Plant Proteomics, Protein-Protein Interactions, Function Prediction, Networks, Systems Biology*