TRACK 4: BIOLOGICAL SCIENCE

Evaluation of the performance of network-based protein function prediction algorithms on protein-protein interaction networks in plants

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 interaction (PPI) networks are used to elucidate functional information of proteins in the absence of experimental data. Many network-based algorithms have been developed for this purpose leveraging the guilt-by-association principle. However, these methods have not been evaluated on plant networks to date. Plant PPI networks differ from those of other organisms due to plant-specific functions such as photosynthesis and root development, etc. Therefore, a systematic evaluation was conducted on six network-based protein function prediction algorithms (majority voting (mv), hishigaki (hm), markov random fields (mrf), random walk (rw), random walk with restart (rwr), and page rank (ppr)) over three plant PPI networks (*Arabidopsis thaliana*, *Oryza sativa japonica*, and *Zea mays*). High-confidence plant PPI networks were obtained from the STRING database, and Gene Ontology (GO) annotations with experimental evidence were sourced from AMIGO. The algorithms were implemented in Python and R, and evaluated using ten-fold cross-validation. Performance was measured on the following metrics: area under the receiver operator curve (AUROC), area under the precision-recall curve (AUPR), F_{max}, and the average precision (AP). While AUROC scores were generally high (0.6–0.9), AUPR and AP values were lower (0.07–0.5). These values highlighted the challenge posed by class imbalance. F_{max} indicated a more balanced performance (0.1-0.6) since it captures the trade-off between precision and recall. Interestingly, *Z. mays* excelled in precision–recall based metrics and balanced performance, whereas *A. thaliana* achieved the highest AUROC scores. Among algorithms, ppr and rwr consistently delivered the best results though simpler methods such as mv and hm occasionally performed competitively, likely due to dataset-specific characteristics. This evaluation highlights the effectiveness of propagation-based algorithms such as ppr and rwr for function prediction in plant PPI networks. Nonetheless, evaluating multiple algorithms remains important, as performance varies across species and network structures.

Keywords: *Plants, Proteomics, Protein-protein interaction networks, Function prediction, Benchmarking*

Acknowledgements: This research is funded under the NSF scholarship grant NSF/SCH/2024/06.