In-silico Prediction of Pairwise Interactions Among Proteins in Rice

Nicolás López-Rozo*, Jorge Finke†, Camilo Rocha‡

October 29, 2020

Abstract

Consider an undirected network in which edges represent interactions between proteins (nodes). Such protein-protein interaction (PPI) networks capture the underlying structure of functional modules of an organism, which influence how it responds and adapts a given environment.

In silico analysis PPI networks in rice reveals that two proteins are more likely to interact when there is a larger number of paths of length 3 between them. Furthermore, this relationship becomes more evident when a bias correction based on the number of interactions of each protein is considered (i.e. the number of paths of length 3, normalized by the node degree).

Our results extend previous analyses on PPIs prediction by considering intrinsic information of the network in the form of embeddings (i.e. features learned through machine learning algorithms). In particular, we train a gradient boosted decision tree model using some handcrafted features (including the normalized measure of the number of paths of length 3, L3) and the vector representation of nodes of the PPI network.

Our findings confirm that L3 is an important feature for predicting PPIs. However, a better performance is achieved when L3 is combined with embedding features. The proposed approach is validated for the rice interactome, resulting in higher AUC values when features from both approaches are combined. This approach can also be applied to interactomes of a number of organisms for which a PPI network has recently become available.

*Pontificia Universidad Javeriana Cali, ORCID: 0000-0002-8060-6891
email: nicolaslopez@javerianacali.edu.co
†Pontificia Universidad Javeriana Cali, ORCID: 0000-0003-2393-6066
‡Pontificia Universidad Javeriana Cali, ORCID: 0000-0003-4356-7704