Protein-Protein Interaction Prediction using Non-Linear Matrix Completion Methods

Amin Emad*, Wei Dai†, and Olgica Milenkovic*

* Dept. of ECE, University of Illinois at Urbana-Champaign

† Dept. of EEE, Imperial College London

{emad2,milenkov}@illinois.edu, wei.dai1@imperial.ac.uk

Abstract

Many problems encountered in experimental and systems biology, as well as in bioinformatics, may be viewed as instances of low-rank matrix completion (LRMC) problem. In this setting, one is presented with a small subset of entries of a low-rank data matrix; the LRMC problem asks when and how this matrix can be uniquely recovered.

Under the assumption that there are fewer degrees of freedom in the interactome than there are proteins, one can model the pairwise protein interaction matrix as being approximately low-rank. The degrees of freedom can be measured in terms of the number of linear/non-linear dependencies among the interaction profiles of proteins.

We present a novel algorithmic solution for predicting protein-protein interactions using the theory of LRMC. We tested the algorithm on the protein-protein interactions of Saccharomyces cerevisiae, reported in the STRING database. The data matrix was indexed by 1200 proteins for which only 15% of the interactions where experimentally verified. Using cross-validation on 20% of the known entries, we achieved a prediction accuracy of more than 85%.

The presented method is amenable for extension to interactions involving more than two proteins, through the use of tensor completion algorithms analyzed for the first time in this work.