

Discovery of small protein complexes from PPI networks with size-specific supervised weighting

Chern Han Yong¹, Osamu Maruyama² and Limsoon Wong³

¹ Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore

² Institute of Mathematics for Industry, Kyushu University, Japan

³ School of Computing, National University of Singapore, Singapore

Abstract

The prediction of small complexes (consisting of two or three distinct proteins) is an important and challenging subtask in protein complex prediction from protein-protein interaction (PPI) networks. The prediction of small complexes is especially susceptible to noise (missing or spurious interactions) in the PPI network, while smaller groups of proteins are likelier to take on topological characteristics of real complexes by chance.

We propose a two-stage approach, SSS and Extract, for discovering small complexes. First, the PPI network is weighted by size-specific supervised weighting (SSS), which integrates heterogeneous data and their topological features with an overall topological isolatedness feature. SSS uses a naive-Bayes maximum-likelihood model to weight the edges with two posterior probabilities: that of being in a small complex, and of being in a large complex. The second stage, Extract, analyzes the SSS-weighted network to extract putative small complexes and scores them by cohesiveness-weighted density, which incorporates both small-co-complex and large-co-complex weights of edges within and surrounding the complexes.

We test our approach on the prediction of yeast and human small complexes, and demonstrate that our approach attains higher precision and recall than some popular complex prediction algorithms. Furthermore, our approach generates a greater number of novel predictions with higher quality in terms of functional coherence.