

Supplementary materials for “A novel link prediction algorithm for reconstructing protein-protein interaction networks by topological similarity”

Chengwei Lei and Jianhua Ruan

December 17, 2012

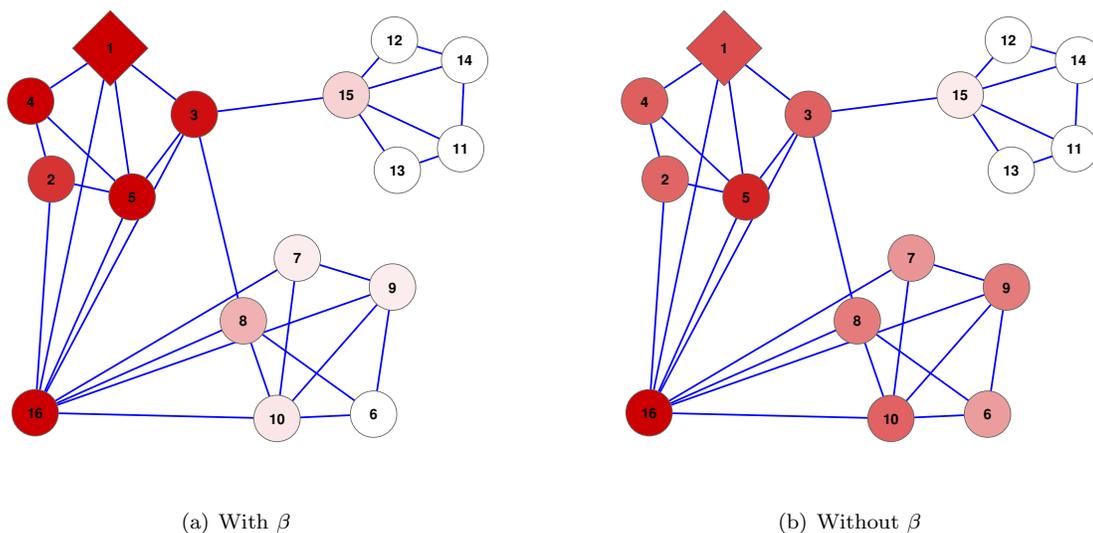


Fig. S1: A motivating example of the effect of the β parameter. This toy network contains three modules (module 1: nodes 1-5, module 2: nodes 6-10, and module 3: nodes 11 - 15). Module 1 and module 2 are both connected to a hub node, 16, while module 3 is relatively independent. In (a), a random walker is initiated from node 1 with the default β parameter as specified in the manuscript, while in (b) the random walker is initiated from node 1 with β setting to zero. As shown, in both cases, the random walker has a very small probability to reach the nodes in module 3. On the other hand, in (a) the random walker cannot easily travel to module 2 while in (b) the random walker has a very high probability to reach nodes in module 2, because of the high connectivity of the hub node 16.

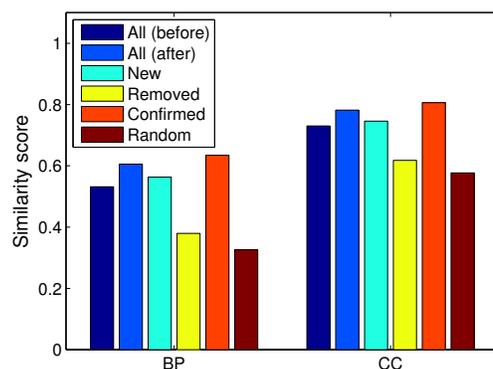


Fig. S2: Evaluation of our method on the Krogan data set using the Gene Ontology “Biological Process (BP)” and “Cellular Component (CC)” branches. The evaluation results based on BP and CC are consistent with that based on the “Molecular Function (MF)” branch of the GO (see main text).

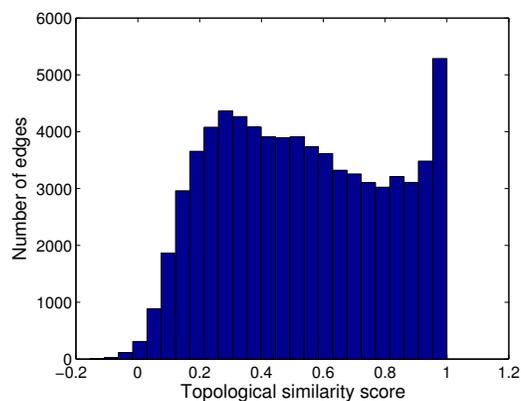


Fig. S3: Distribution of the topological similarity scores for the original human HPRD PPIs. The distribution shows a clear bimodal distribution, with one peak at around 0.25 and the other at 1. Based on this distribution, we used a score cutoff = 0.2 to select the high-confidence false positive edges, and a score cutoff = 0.9 to select the high-confidence false negative edges.