Electronic Supplementary Material (ESI) for Molecular BioSystems.

Supporting methods

A new multi-scale method to reveal hierarchical modular structures in biological networks

Qing-Ju Jiao , Yan Huang , Hong-Bin Shen

The definitions of four node similarities

The four node similarities that are used to compare with ISIM are the hub promoted index (HPI), jaccard index (Jaccard), leicht holme newman (LHNII) and random walk with restart (RWR), and they are defined by following equations. Given a network with N nodes and M edges. For a node i in the network, let U_i denote the set of neighbors of i, and d_i denotes the degree of i.

$$HPI_{ij} = \frac{|U_i \cap U_j|}{min\{d_i, d_j\}} \tag{S1}$$

$$Jaccard_{ij} = \frac{|U_i \cap U_j|}{|U_i \cup U_j|} \tag{S2}$$

$$LHNII_{ij} = \left[1 - \frac{2M\lambda_1}{d_i d_j}\right]\delta_{ij} + \frac{2M\lambda_1}{d_i d_j}\left[\left(I - \frac{\phi}{\lambda_1}W\right)^{-1}\right]_{ij} \tag{S3}$$

$$q_i(t+1) = cP^T q_i(t) + (1-c)e_i \tag{S4}$$

$$q_i = (1-c)(I - cP^T)^{-1} e_i \tag{S5}$$

$$RWR_{ij} = q_{ij} + q_{ji} \tag{S6}$$

In equation S3, M and W represent the total number of edges and adjacency matrix of a network respectively. λ_1 is the largest eigenvalue of W and I is an identity matrix. δ_{ij} is the Kronecker function. $\phi(0 < \phi < 1)$ is a free parameter, the selecting of the parameter relies on the observed network, and smaller ϕ assigns more weights on shorter paths. Equations S4, S5 and S6 define the node similarity RWR. Consider a random walk starting from node i, who will iteratively moves to one of neighbor with probability c and return to the node i with probability $1-c$. $q_i(t)$ and $q_i(t+1)$ are the transition vectors from node i to other nodes in the network at the tth and $(t+1)$th times respectively. P is the transition matrix, the element $P_{ij} = 1/d_i$ if i and j are connected, $P_{ij} = 0$ otherwise, and e_i is a vector in which the ith element equal to 1 and others are equal to 0. The convergent solution of equation S4 is equation S5. The RWR is defined by equation S6 because q_{ij} may be not equal to q_{ji}.

The RSDs of different node similarity metrics

A parameter with a range from 0 to 1 should be selected to calculate LHNII, RWR and ISIM. For the free parameter ϕ in LHNII, smaller ϕ places more weights on short paths between nodes and larger ϕ places more weights both on short and long paths. Although there is no unique value that works perfectly for most networks, a reliable value 0.97 is provided in LHNII. The parameters in RWR and ISIM are regarded as transition probability, and the selection of them is also no unique. But the parameter value 0.95 in RWR and ISIM is widely used. Here, we provide the results of RSDs of five node similarities when the free parameter is set 0.97 (Fig. S1) and 0.95 (Fig. S2) in LHNII, RWR and ISIM. From both Fig. S1 and Fig. S2, we can see that the performance of ISIM is most stable in five methods.

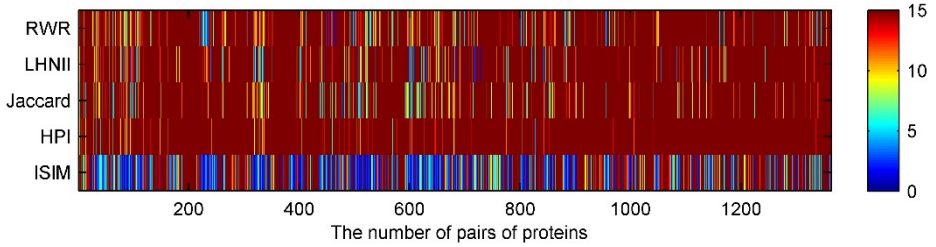

Fig. S1 The RSDs of different node similarity metrics (0.97)

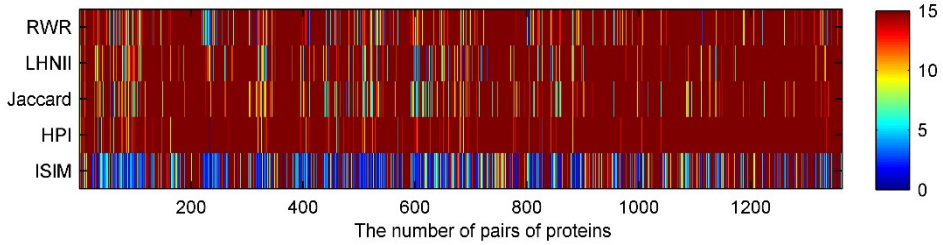

Fig. S2 The RSDs of different node similarity metrics (0.95)

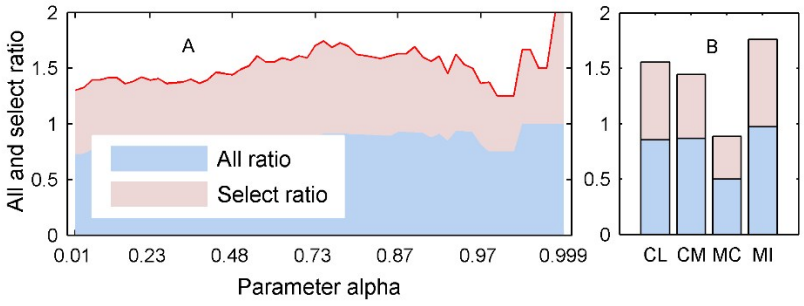

Fig. S3 The sum of two ratio indexes (CC) from ISIMB method (A) and four single-scale modules mining methods (B)

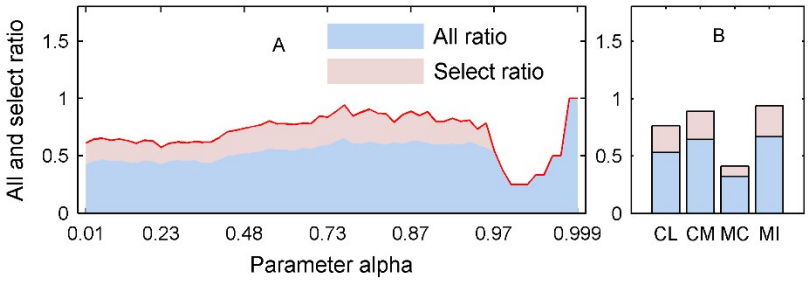

Fig. S4 The sum of two ratio indexes (MF) from ISIMB method (A) and four single-scale modules mining methods (B)

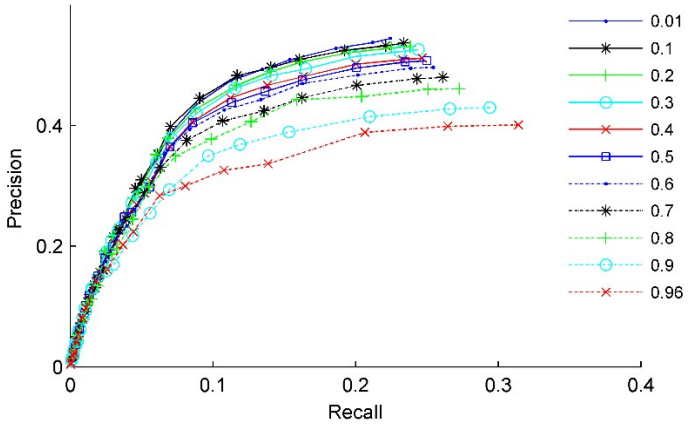

Fig. S5 The precision-recall plots as resolution regulator α is varied on physical PPI network, the digits in legend represent the values of α in ISIMB method

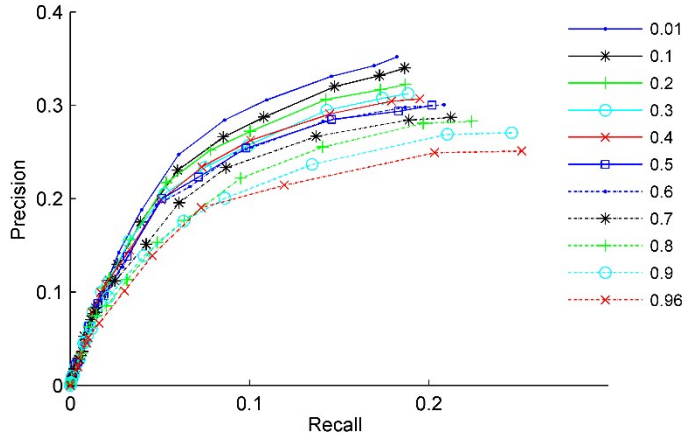

Fig. S6 The precision-recall plots as resolution regulator α is varied on genetic interaction network, the digits in legend represent the values of α in ISIMB method

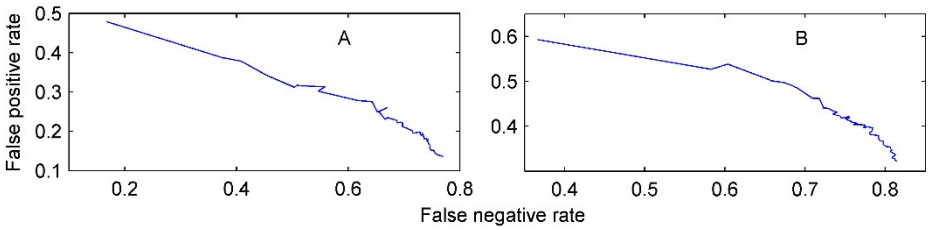

Fig. S7 The false positive rate and false negative rate plot of ISIMB on physical PPI network (A) and genetic interaction network (B)