Supplementary Data

Table S1 Relation between ADMSC and spectral analysis

Generally, the Laplacian matrix of a graph network is defined as:

$$\mathbf{L} = \mathbf{\Delta} - \mathbf{A} \tag{S1}$$

where $\mathbf{A}$ is the adjacency matrix. $\mathbf{\Delta}$ is given by:

$$\Delta_{ij} = \delta_{ij} \sum_{k=1}^{n} A_{ik} \qquad \delta_{ij} = \begin{cases} 1 \; if \; i = j \\ 0 \; if \; i \neq j \end{cases} \tag{S2}$$

When use of the transition matrix of a random walk $\mathbf{D}^{-1}\mathbf{A}$ is favored, the Laplacian matrix is changed into:

$$\mathbf{L}^{*} = \mathbf{\Delta}^{*} - \mathbf{D}^{-1}\mathbf{A} \tag{S3}$$

$$\Delta^{*}_{ij} = \delta_{ij} \sum_{k=1}^{n} (\mathbf{D}^{-1}\mathbf{A})_{ik} \tag{S4}$$

$$\delta_{ij} = \begin{cases} 1 \; if \; i = j \\ 0 \; if \; i \neq j \end{cases}$$

Here, the set of all the eigenvalues λ and the corresponding eigenvectors $\mathbf{v}$ for $\mathbf{L}^{*}$, called spectrum, is given by:

$$\mathbf{L}^{*} \cdot \mathbf{v}_i = \lambda \cdot \mathbf{v}_i \tag{S5}$$

where $\{\mathbf{v}_i\}$ provides the node coordinates in the Euclidean space to perform network clustering. Since ADMSC corresponds to the spectral clustering that employs a modified transition matrix $\mathbf{D}^{-\beta}\mathbf{A}$, it can be regarded as the diffusion model-based probabilistic interpretation of spectral analysis.

Table S2 Instruction of ADMSC program in Matlab

README.txt in ADMSC.zip
Execute the ADMSC.m file in MATLAB >>ADMSC Input for ADMSC.m line 46: nc=33; % Number of clusters line 47: B=1.4; % Beta factor *You decide the number of clusters and the value of the beta factor. line 52 : [nodea nodeb] = textread('Scere.txt','%s%s'); % input your network *The file of default is the yeast PPI network that consists of the binary relation between proteins. line 97 : outputdata(nodes,cidx,modu,cv,nnode,nedge); % output data *The result is output as ‘output.txt’.

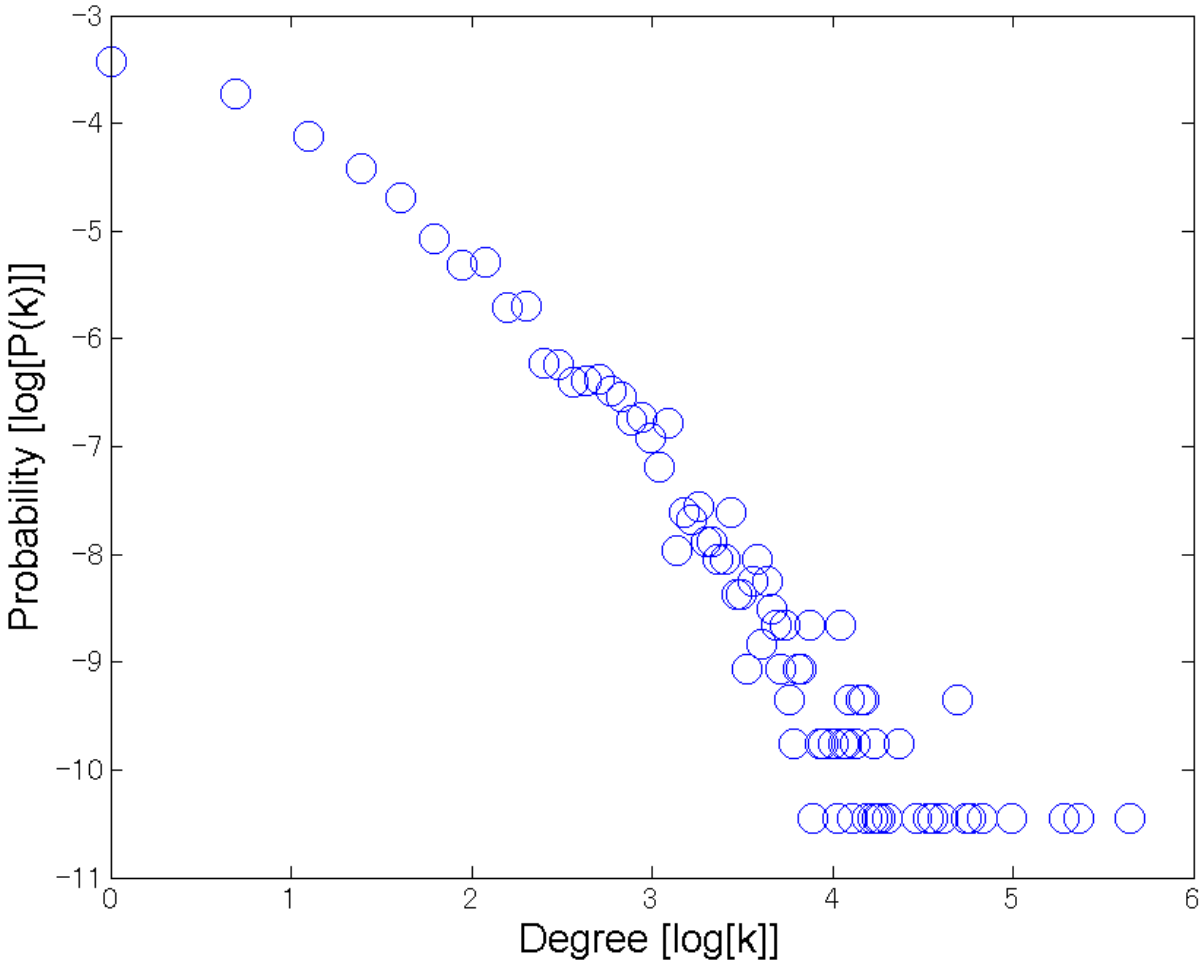


Figure S1 Degree distribution in the scale free network of an yeast PPI network

k is the node degree.

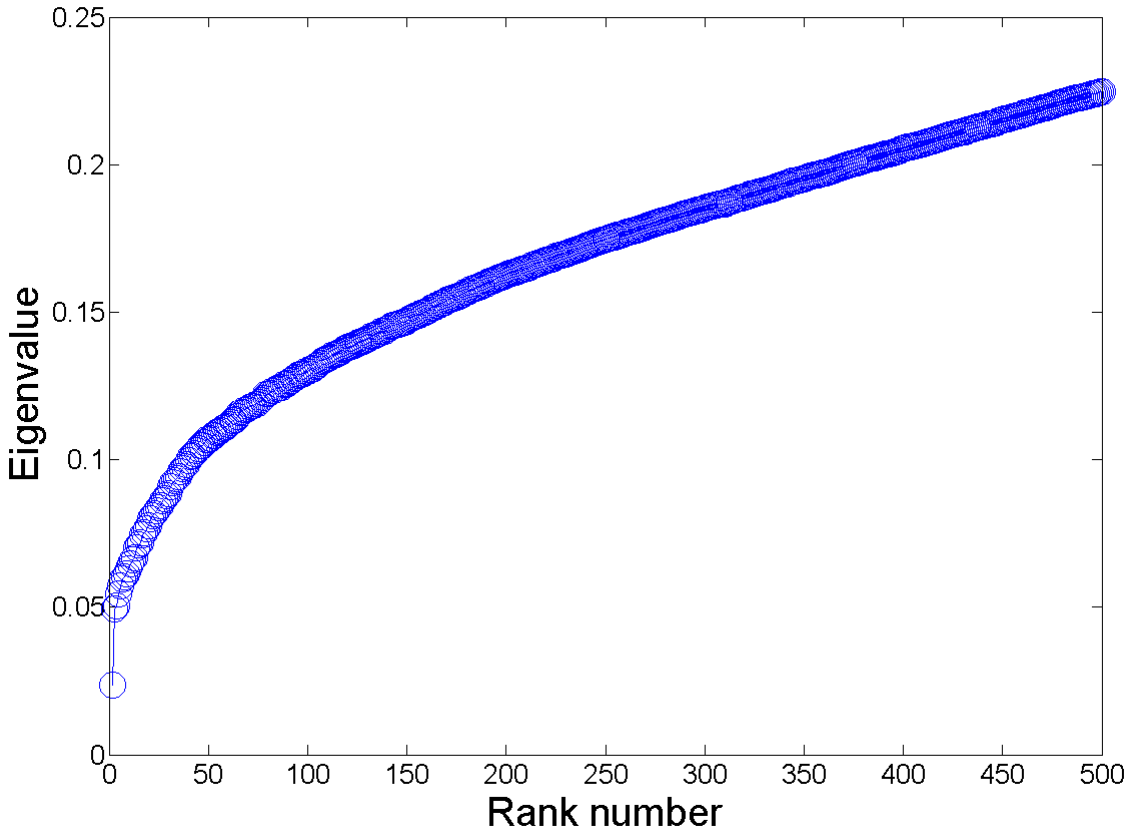


Figure S2 Changes in eigenvalues with respect to cluster number in the yeast PPI network

The eigenvalues for the diffusion matrix are sorted in the ascending order. There is no great gap between the neighboring eigenvalues.

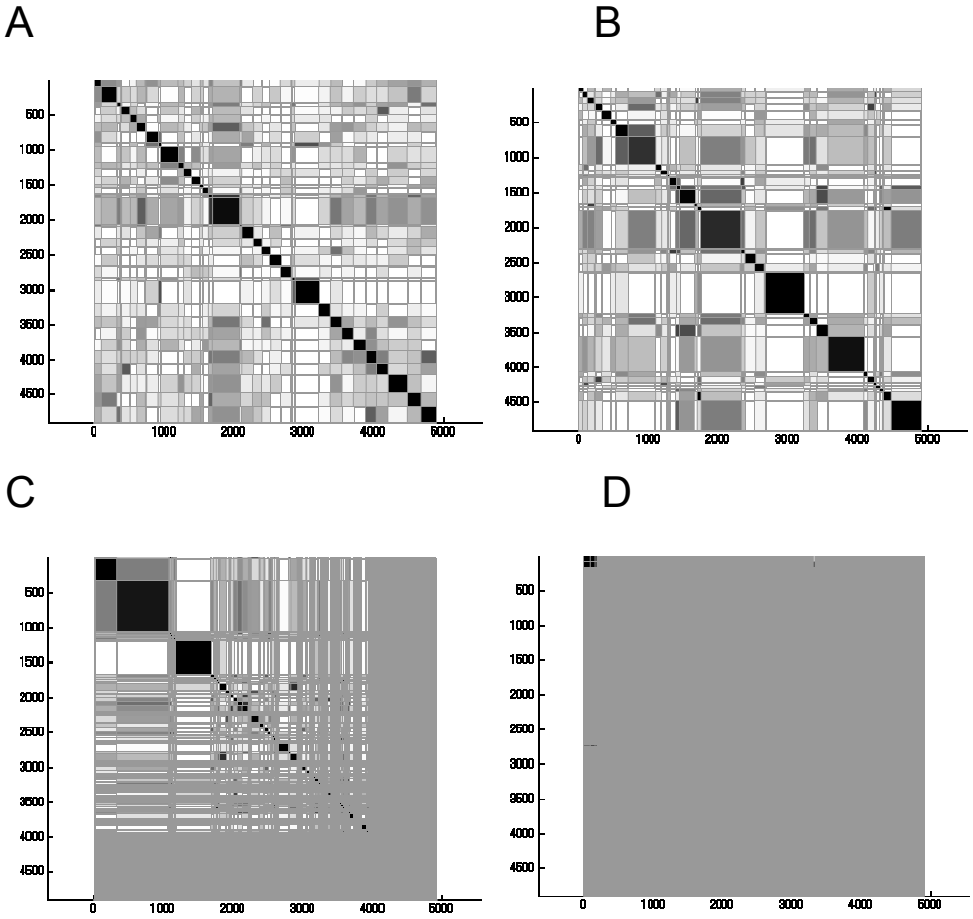


Figure S3 Distributions of cluster size in an yeast PPI network

Proteins are arranged in both the vertical and horizontal axes in the same turn. The density of interactions between nodes is marked as the degree of darkness. The size of the squares on the diagonal indicates the cluster size, and the darkness density shows the interaction strength of the inter- and intra-clusters. The cluster size distributions are calculated by: (A) ADMSC with β=1.4 at a cluster number of 33, (B) ADMSC with β=1 (regular spectral analysis) at a cluster number of 33, (C) SPB at a cluster number of 319, and (D) MCL at a cluster number of 1233.

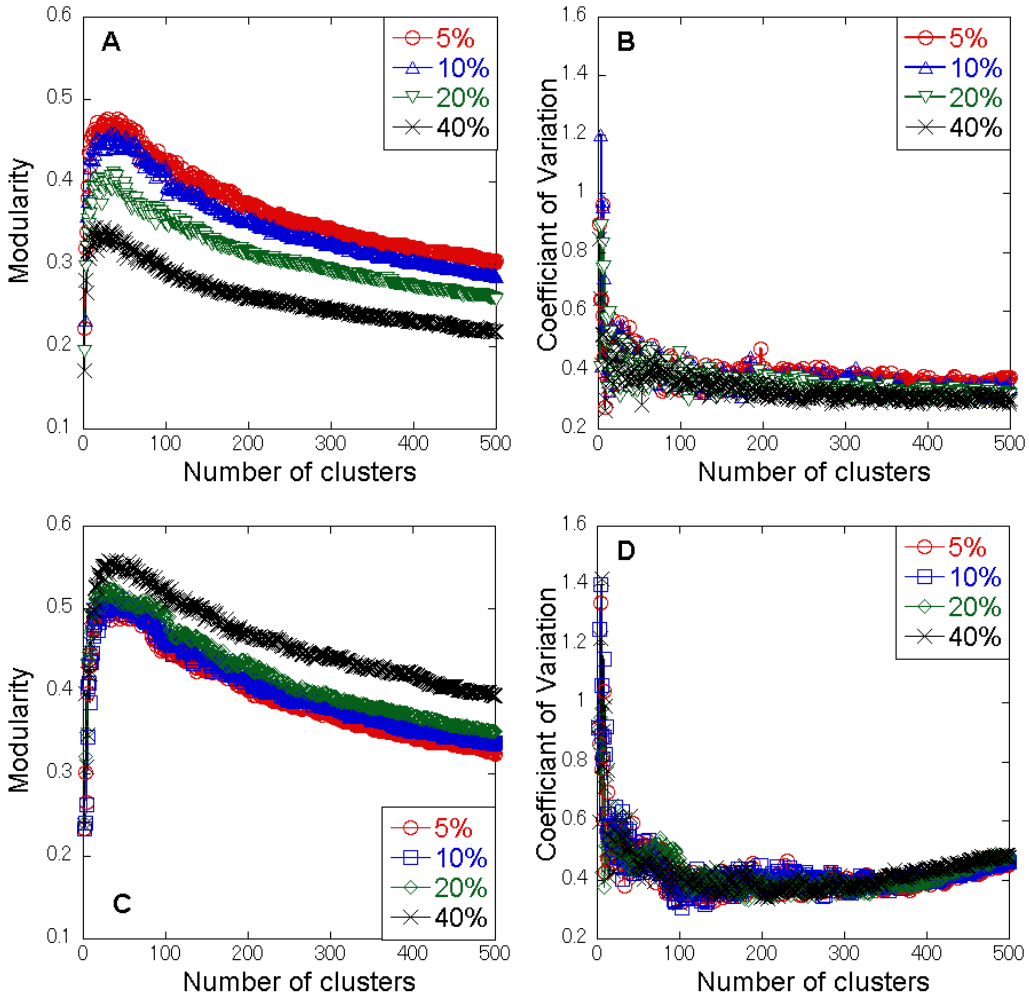


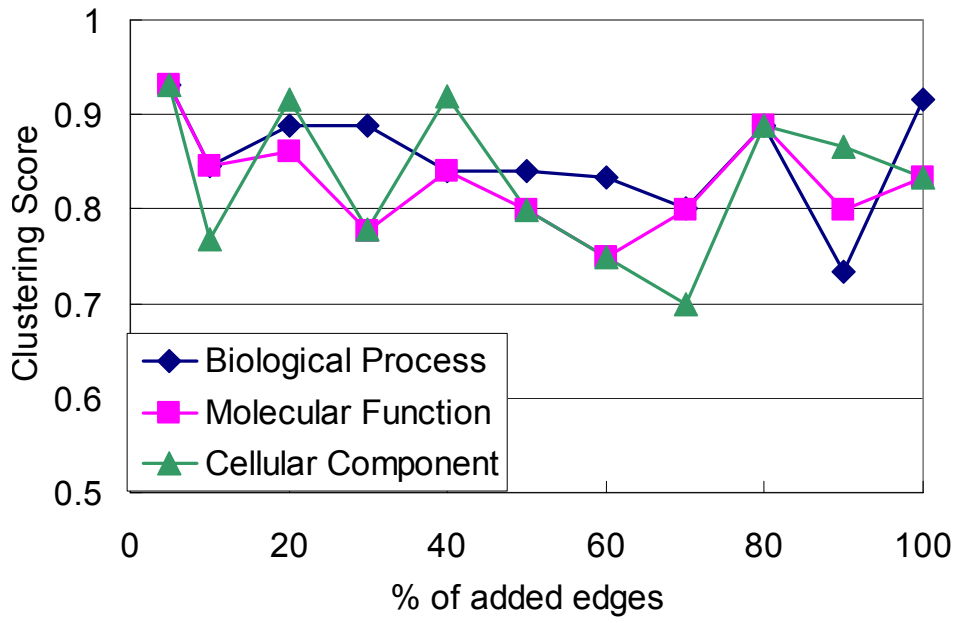
Figure S4 Perturbation analysis for modularity and cluster size calculated by ADMSC

Changes in the modularity and CV of cluster size are calculated by ADMSC with respect to perturbed edges in the yeast PPI network.

(A, B) Five, ten, twenty, and forty percentages of edges are added randomly.

(C, D) Five, ten, twenty, and forty percentages of edges are removed randomly.

A



B

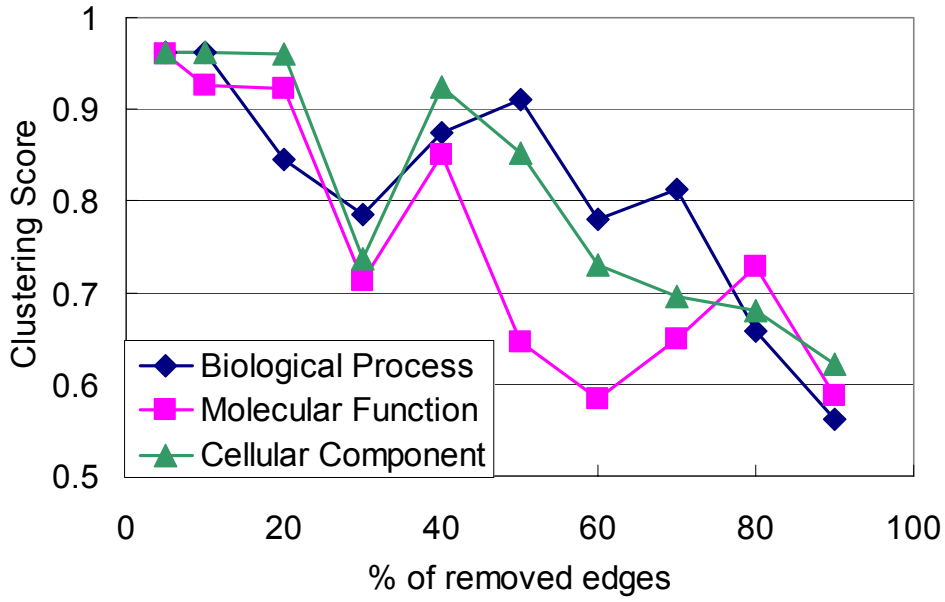


Figure S5 Perturbation analysis for clustering scores calculated by ADMSC

(A) 5 - 100 percentages of edges are added randomly.

(B) 5 - 90 percentages of edges are removed randomly.

The clustering scores are robust with respect to random addition of 5 to 100 percentages of edges, whereas they are robust with respect to random removal of 5 to 20 percentages of edges.

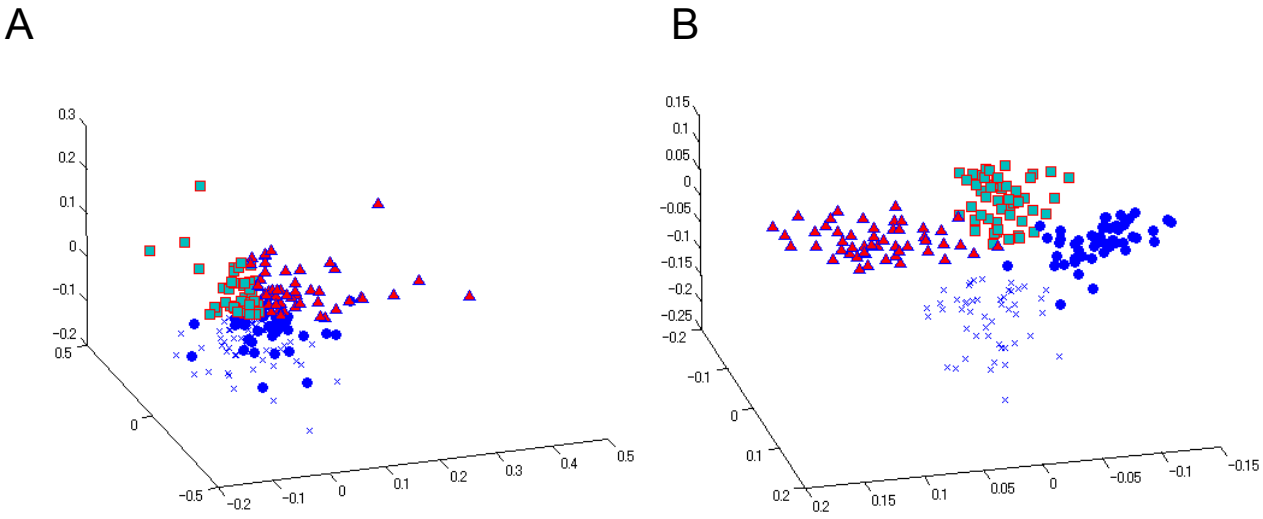


Figure S6 Diffusion maps for a scale free network and a network with four distinct clusters.

The diffusion map spanned by three eigenvector is shown in three-dimensional space. The same symbol belongs to the same cluster.

(A) A scale free network (200 nodes and 398 edges) is built based on the BA (Barabási and Albert) model and decomposed into four groups by ADMSC.

(B) An artificial network with four distinct clusters (200 nodes and 994 edges) is built and decomposed into four groups by ADMSC.

In (B), the clear clusters are illustrated along the radial directions from the original point. In (A), while the nodes are condensed around the original point, each cluster seems to exist along the radial directions from the original point. It suggests that use of the angular distance partitions a scale free network.

To clearly demonstrate the cluster structure of these networks, we provide DiffMap_BAmodel.fig (A) and DiffMap_FourCluster.fig (B) in ADMSC.zip. These figures can be rotated in MATLAB.