# Overview of network terminology

Peter Langfelder and Steve Horvath

October 25, 2010

**Table 1.** Overview of network terminology

| Term | Definition |
|---|---|
| (Undirected) Network | Generally speaking, an undirected network consists of nodes (for example, gene expression profiles), and connection strengths between pairs of nodes. The connection strengths can be either categorical (connected vs. unconnected), or continuous between 0 (no connection) and 1 (strongest connection). |
| Adjacency matrix | The connection strengths in an undirected network can be represented by the *adjacency matrix*, a symmetric matrix whose entries lie between 0 and 1. The element $a_{ij}$ is the connection strength between nodes $i$ and $j$. As a convention, the diagonal elements are set to 1, $a_{ii} = 1$. |
| Correlation network | This type of network is built from numerical data $X_{bi}$ representing the value of variable $i$ in observation $b$. The adjacency (connection strength) $a_{ij}$ between nodes $i$ and $j$ is calculated from the correlation of the corresponding node profiles $X_{\cdot i}$ and $X_{\cdot j}$. our applications, we use $a_{ij} = |cor(x_i, x_j)|^{\beta}$ for unsigned adjacency and $a_i j = [\frac{1}{2}(1 + cor(x_i, x_j)]^{\beta}$ for the signed adjacency. |
| Gene co-expression network | In gene co-expression networks, the nodes represent genes (or probesets of a microarray) measured across a given set of microarray samples, and the connections represent the strength of co-expression. Various measures of co-expression can be used, for example Pearson or robust correlation (in which case the co-expression network is also a correlation network), information-theoretic methods such as mutual information, and other measures of co-expression similarity. |
| Sub-network | A subnetwork of a network can be any collection (subset) of nodes from the network, together with the adjacencies (connection strengths) between the nodes. Thus, a subnetwork of a network also forms a (smaller) network on its own. |
| Module | A network module is a subset of nodes that forms a sub-network inside a larger network. Any subset of nodes inside a larger network gives rise to a module. This subset may or may not correspond to a cluster of nodes. |
| Cluster | A cluster of nodes within a network is usually defined as a group of nodes that are strongly connected. Many definitions and algorithms for finding clusters in data have been proposed in the literature. |
| Network density | The mean adjacency (connection strength) among all nodes in the network. |
| Connectivity | For each gene, the connectivity (also known as degree) is defined as the sum of connection strengths with the other network genes: $k_i = \sum_{u \neq i} a_{ui}$. In co-expression networks, the connectivity measures how correlated a gene is with all other network genes. |

| | |
|---|---|
| Intramodular connectivity $k_{IM}$ | Intramodular connectivity measures how connected, or co-expressed, a given gene is with respect to the genes of a particular module. Thus, intramodular connectivity is also the connectivity in the subnetwork defined by the module. The intramodular connectivity may be interpreted as a measure of module membership. |
| Module eigengene $E$ | The module eigengene $E$ is defined as the first principal component of a given module. It can be considered a representative of the gene expression profiles in a module. |
| Eigengene-based connectivity $k_{ME}$, also known as module membership ($MM$) | For each gene, we define a "fuzzy" measure of module membership by correlating its gene expression profile with the module eigengene of a given module. For example, $K_{ME,i}^{blue} = MM^{blue}(i) = \text{cor}(x_i, E^{blue})$ measures how correlated gene $i$ is to the blue module eigengene. $k_{ME}^{blue}(i)$ measures the membership of the $i$-th gene with respect to the blue module. If $k_{ME}^{blue}(i)$ is close to 0, the $i$-th gene is not part of the blue module. On the other hand, if $k_{ME}^{blue}(i)$ is close to 1 or $-1$, it is highly connected to the blue module genes. The sign of module membership encodes whether the gene has a positive or a negative relationship with the blue module eigengene. The module membership measure can be defined for all input genes (irrespective of their original module membership). It turns out that the module membership measure is highly related to the intramodular connectivity $k_{IM}$. Highly connected intramodular hub genes tend to have high module membership values to the respective module. |
| Hub gene | This loosely defined term is used as an abbreviation of "highly connected gene". |
| Gene significance $GS$ | To incorporate external information into the co-expression network, we make use of gene significance measures. Abstractly speaking, the higher the absolute value of $GS_i$, the more biologically significant is the $i$-th gene. For example, $GS_i$ could encode pathway membership (e.g. 1 if the gene is a known apoptosis gene and 0 otherwise), knockout essentiality, or the correlation with an external microarray sample trait. A gene significance measure could also be defined by minus log of a p-value. The only requirement is that gene significance of 0 indicates that the gene is not significant with regard to the biological question of interest. The gene significance can take on positive or negative values. |
| Module significance | Module significance is determined as the average absolute gene significance measure for all genes in a given module. When gene significance is defined as the correlation of gene expression profiles with an external trait $y$, this measure tends to be highly related to the correlation between the module eigengene and $y$. |
| Eigengene significance | When a microarray sample trait $y$ is available (e.g. case control status or body weight), one can correlate the module eigengenes with this outcome. The correlation coefficient is referred to as eigengene significance. |