List of Figures

List of Tables

List of Algorithms

Symbols and Meanings

Symbol	Meaning
$G(V, E)$	PPI network with $V = \{v_1, v_2, ..., v_n\}$ vertices and $E = \{e_1, e_2, ..., e_m\}$ edges.
$cluster$	Set of clusters (complexes) without removing redundant complexes.
$d(v_i)$	degree of v_i.
$CCf(v_i)$	Clustering coefficient of v_i.
$Connt(v_i, G')$	Connectivity of v_i with G'.
$CCfT$	Clusstering coefficient threshold.
$NgCp(C_i)$	Neighbor of complex, C_i.
$Ng(v_i, v_j)$	Neighbor of v_i and v_j.
α	Connectivity threshold.
$RemList$	Data structure to keep track of nodes during complex finding.
pC	Data structure to store elements while growing a cluster.
$HC(v_i, v_j)$	HConfidence value between v_i and v_j.
$sd = (v_i, v_j)$	Seed pair consisting of nodes v_i and v_j during unsupervised complex finding.
RbI_{v_i}	Reachability Index of node v_i.
$Supp(G')$	Contribution of subgraph G'.
β	Semantic similarity threshold.
h_t	HConfidence threshold .
$secCluster$	set of clusters before removal of redundant clusters.
$CndS$	Data structure to record candidate seed pair.
HCS	HConfidence store to keep record of the HConfidence values of each pair of nodes.
sd_p	Selected seed pair from the set of candidate seed pair, $CndS$.
$NodeExpcluster$	Data structure to keep track of nodes during cluster expansion.
$pC1$	Data structure similar to pC but only with an extra node, v_m to calculate contribution function.

Symbol	**Meaning**
$hcount$	Counter to maintain the record for every pair of nodes with HConfidence values.
$ccount$	Counter to maintain record of $Cluster$.
$acount$	Counter to maintain record for $secCluster$.
S_{sd_p}	Seed pair consisting of nodes v_i and v_j during semi-supervised complex finding process.
ms_{v_a,v_b}	Matching scores between nodes v_a and v_b.
$MMgS(v_i, v_j)$	Maximum Matching Score between a pair of nodes v_i and v_j.
$NSSm$	Semantic similarity score matrix.
$D_{gs} = \{g_{d_1}, g_{d_2}, ...g_{d_p}\}$	set of disease genes.
DGN	Disease gene association matrix.
S_{C_j}	Relevance score of complex C_j.
BCM	Protein complex benchmark matrix.
A_{MMgS}	Data structure to store maximum matching score of v_i and v_j.
$scount$	Counter to maintain track of seed node.
$ActualCluster$	Set of unique clusters.
$rs_{g_{d_i}}$	Relevance score of disease gene g_{d_i}.
$xcount$	Counter to count the number of disease genes.
$rcount$	Counter for ranking disease associated complexes.
$pvalue_{C_i}$	p-value of complex C_i.
$DN_{C_{v_i}}$	Direct neighbors of node v_i within C_i.
$contbn_{v_i}$	Node v_i contribution in terms of direct neighbors.
Cl_{contbn}	Cluster contribution (of all nodes within a cluster).
$Rbty_{v_i}$	Reachability of a node v_i in a cluster.
$RbyContbn$	Reachabiity contribution.

Symbol	**Meaning**
$A = \{G, E\}$	Adjacency matrix representation of network with V vertices and E edges.
V_{th}	Variance threshold.
PCC_{th}	Pearson correlation coefficient threshold.
Adj_{nm}	Adjacency network representation of non metastasis stage.
Adj_m	Adjacency network representation of metastasis stage.
M_i	i^{th} module.
$SemSim(g_a, g_b)$	Semantic similarity between genes g_a and g_b.
$SemSim_{th}$	Semantic similarity threshold.
$SemSimM$	Semantic similarity score matrix.
$Modules = \{M_1, M_2, ..., M_N\}$	Set of N modules.
$RemList$	Data structure to keep track of nodes during module extraction.
$mCount$	Counter to keep track of number of modules .
GM	Gene Expression matrix with a number of genes (rows) at b conditions (columns).
D_m	Discretized gene expression matrix with a rows and $b \times (b-1)/2$
η	Threshold based on the standard deviation of the expression matrix.
Γ	Threshold used to calculate whether two genes g_1 and g_2 are connected or not based on the absolute difference between their expression values.
PCC	Pearson correlation coefficient.

Symbol	**Meaning**
$T_{SbOM_{ab}}$	Topological Subspace Overlap Metric for a pair of nodes a and b.
αA	similarity between neighbors.
βB	direct edge connectivity.
$E^p_{g_a g_i}$	edge between genes g_a and g_i at p^{th} condition.
V_g	total number of genes in GM.
sm	total number of samples in GM.
Adj_{g_a,g_b}	adjacency matrix having value 1 or 0 depending on whether g_a and g_b are connected or not.
n_{th}	Threshold used for assigning values to Adj_{g_a,g_b}
$Nodedistance(g_v, g_k)$	Distance between nodes g_v and g_k.
$N_{CN}(g_v)$	Constrained neighbor set of g_v.
$CnS(g_v)$	Constrained node score.
$SI(g_v, g_k)$	Simpson Index for g_v and g_k.
$N(g_v)$	Neighbors of g_v.
γ	Threshold used during seed selection based on CnS.
m_i	Network module
$L = \{c_{g_1}, c_{g_2}, ...c_{g_k}\}$	Set of causal genes associated with the disease.
δ	Threshold used during selection of SI between genes.
$partialCluster$	Data structure to store elements while growing a cluster.
$T_{SbOM} = \{V, E\}$	TSOM network.
$Modules = \{C_1, C_2, \cdots, C_N\}$	Set of modules.
M_{c_i}	Modules from the control stage.
M_{d_j}	Modules from the disease stage.
$MR(c_i, d_j)$	Matching ratio between control module, c_i and disease module d_j.