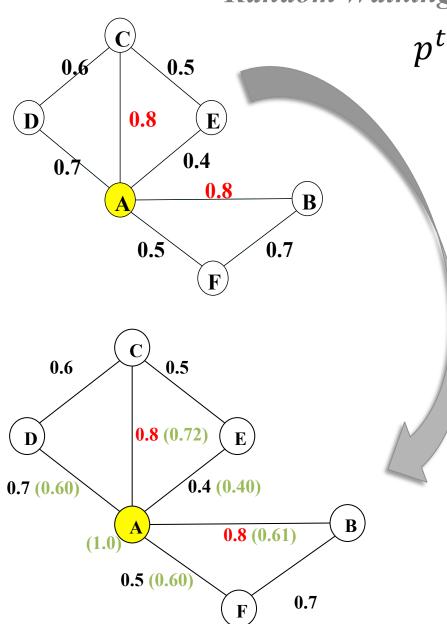


Select source organims

Organism Name	Number of Contigs
Coccomyxasubellipsoidea C-169	16644
Chlorella variabilis	13342
Volvox carteri	12252
Chlamydomonas reinhardtii	11459
Glycine max	11450
Populus trichocarpa	11289
Setaria italica	11257
Physcomitrella patens subsp.	11253
patens	11227
Vitis vinifera	11226
Brachypodium distachyon	11208
Solanum lycopersicum	11056
Arabidopsis thaliana	11045
Brassica rapa subsp. pekinensis	11030
Selaginella moellendorffii	11017
Cicer arietinum	10982
Fragaria vesca subsp. vesca	10966
Cucumis sativus	10964
Prunus persica	10909
Ricinus communis	10800

Random Walking with Restart!



$$p^{t+1} = (1-r)W_{p^t} + rp^0$$

$$p_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} w_{p0} = \begin{bmatrix} A & B & C & D & E & F \\ A & 0 & 0.8 & 0.8 & 0.7 & 0 & 0.5 \\ B & 0.8 & 0 & 0 & 0 & 0 & 0.7 \\ C & 0.8 & 0 & 0 & 0.6 & 0.5 & 0 \\ D & 0.7 & 0 & 0.6 & 0 & 0 & 0 \\ E & 0.4 & 0 & 0.5 & 0 & 0 & 0 \\ F & 0.5 & 0.7 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Run RWR:

tolerence = 0.06

Max Iterations = 1000

$$r = 0.33$$

Normalize by maximum score

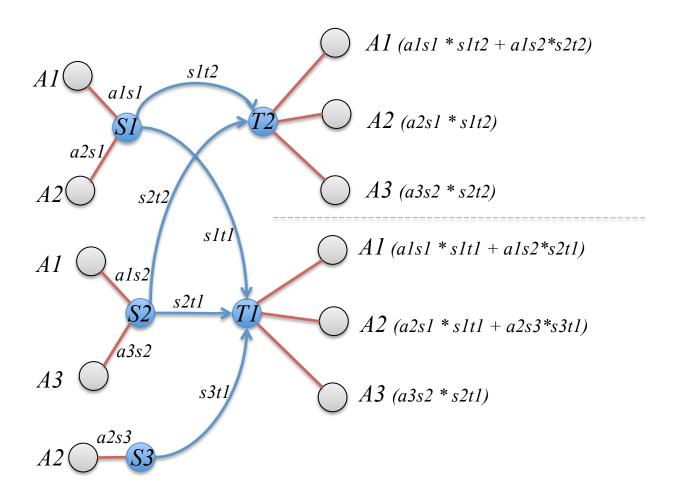
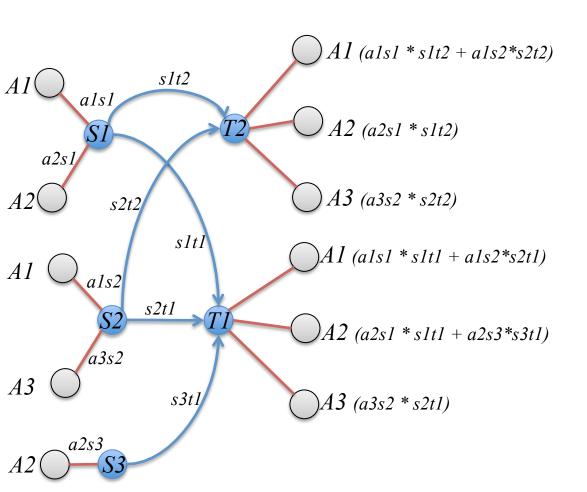
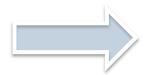


Figure: Annotation flow network: A1-A3 are annotations of source sequences S1, S2 and S3. axsy represents the user-defined confidence of the annotation ax to be associated with sequence sy. sxty is the maximum score normalized orthofuzz score(sx,ty) obtained by querying the pairwise sequence similarity network using sx.





This translates as matrix multiplication of Annotation Confidence Matrix and Orthofuzz matrix

$\mathbf{ATS}_{N_A \times N_{TS}} = \mathbf{ASS}_{N_A \times N_{SS}} \times \mathbf{SSTS}_{N_{SS} \times N_{TS}}$

Symbol	Description
SS	Set of Source Sequences
N_{SS}	Total Number of Source Sequences
A	Set of Source Annotations
N_A	Total Number of Annotations from Source Sequences
TS	Set of Target Sequences
N_{TS}	Total Number of Target Sequences
ATS	Target Annotation Weight Matrix
ASS	Source Annotation Confidence Matrix
SSTS	Source Target Orthofuzz Matrix
axsy	Confidence of assigning ax to sy
sytz	Max. normalized orthofuzz score of sy to tz
axtz	Annotation weight of ax to tz

Cluster by Random Walking

Data:

Shared among threads:

Transferred to threads

Random Walking with Restart Create population of Scores

Estimate Statistics

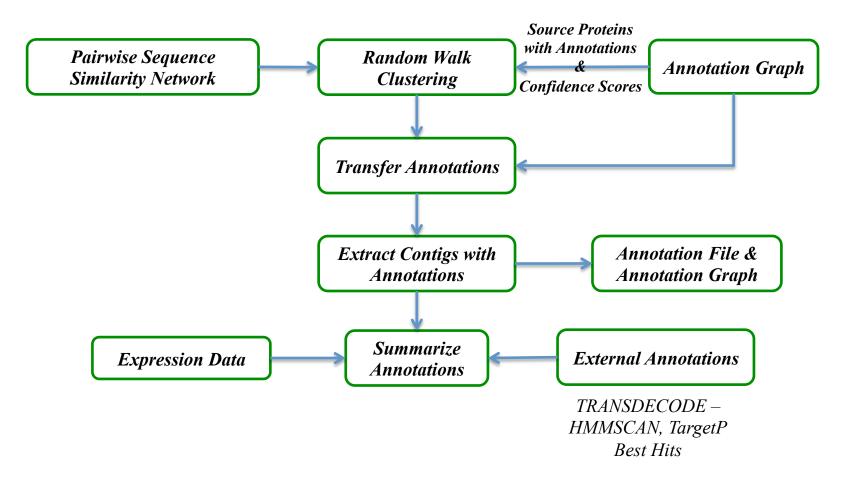


Figure 1: Protocol used for assigning functional annotations to the de-novo assembled contigs

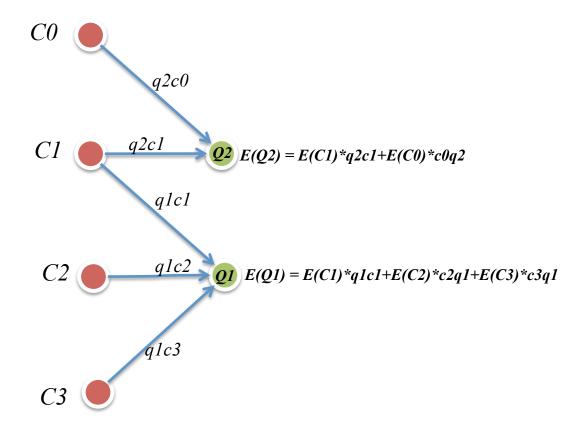
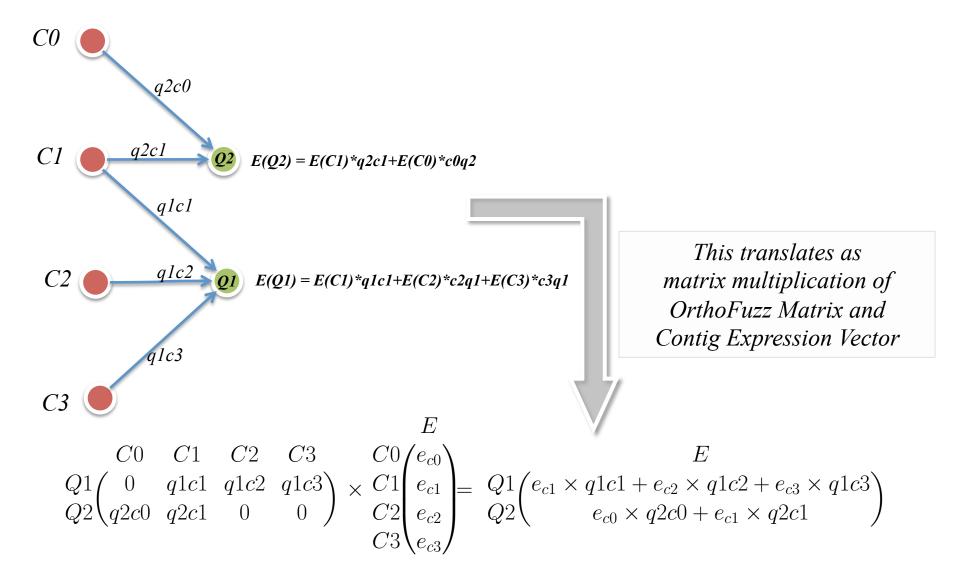
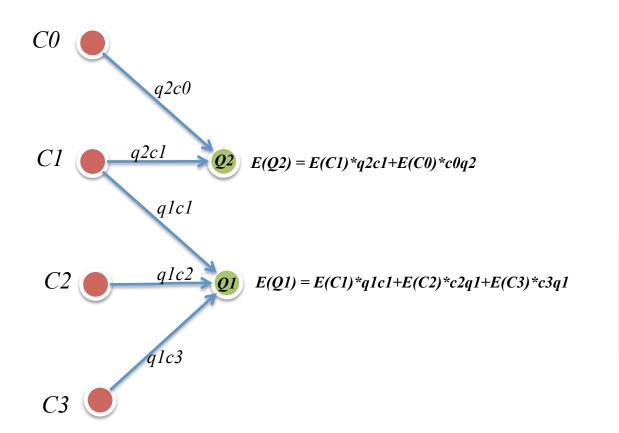


Figure X: Expression summary network: C0-C3 are the contigs identified as homologs of Query Sets Q1 and Q2. E(X) is the normalized expression level of X. qxcy is the within species normalized orthofuzzscore (qx,cy) obtained by querying the network using the query set Q.







This translates as matrix multiplication of OrthoFuzz Matrix and Contig Expression Vector

$$\mathbf{QC}_{N_Q \times N_C} \times \mathbf{Ce}_{N_C \times 1} = \mathbf{Qe}_{N_Q \times 1}$$

Symbol	Description
Q	Set of Query Sets
N_Q	Total Number of Query Sets
N_C	Total Number of Expressed Contigs
QC	Query Contig Orthofuzz Matrix
qxcy	Species.normalized orthofuzz score of qx to cy
Ce	Vector containing expression values of Contigs
e_{cy}	Expression value of contig y
Qe	Vector containing expression values of Query

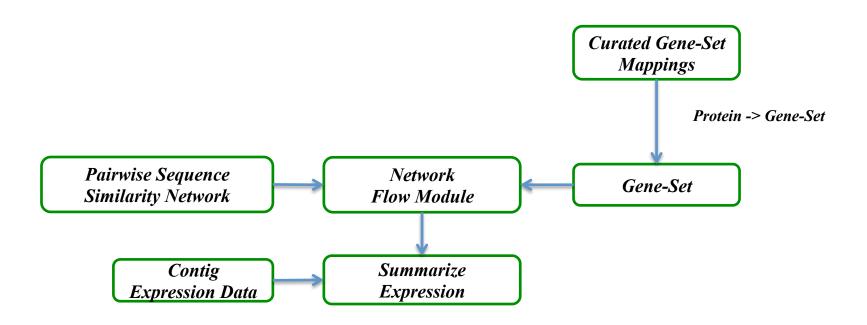
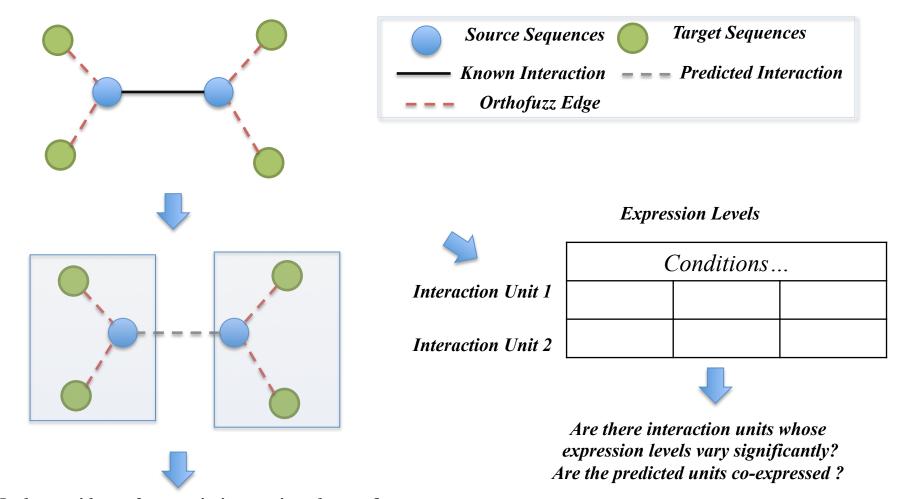


Figure: Protocol used for estimating expression levels of gene-sets



Is there evidence for protein interaction clusters?

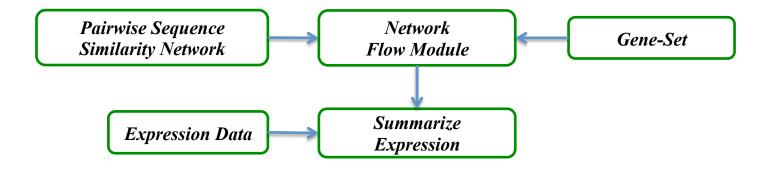


Figure: Protocol used for estimating expression levels of gene-sets

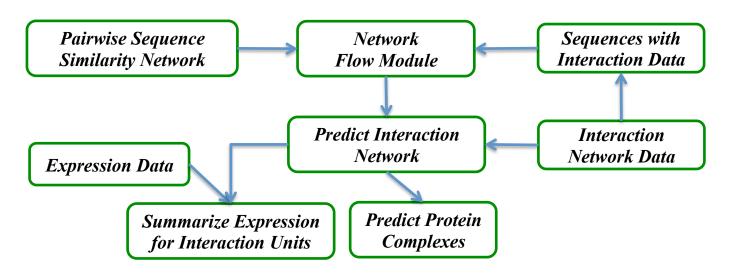


Figure: Protocol used for estimating expression levels of gene-sets