An Example:

Weighted Edge Based Clustering to Identify Protein Complexes in Protein-Protein Interaction networks incorporating Gene Expression Profile.

Compiled and Submitted by:

Seketoulie Keretsu

And

Dr. Rosy Sarmah

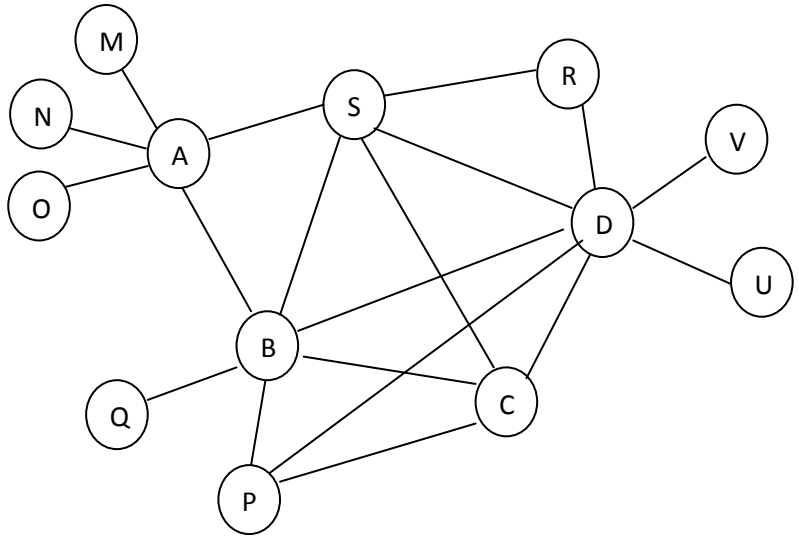

Figure.(1): Sub network G from a PPI network

Consider the sub network G given in fig.(1) from a PPI network for example. Then the interaction table.(1) is given below where each row represents an interaction between two proteins. Assume vertex S as the seed vertex to start the clustering process.

Protein 1	Protein 2
S	A
S	B
S	C
S	D
S	R
A	B
A	M
A	N
A	O
B	P
B	Q
B	C
B	D
C	D
C	R
C	P
D	P
D	S
D	R
D	U
D	V

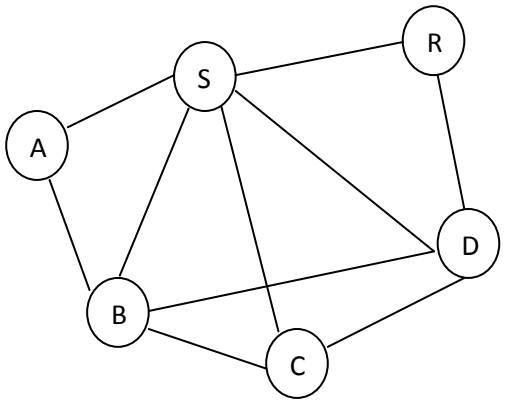

Figure (2): Neighbourhood Graph of seed vertex(S)

Table.(1): Table representing protein interactions given in the sub-graph given in figure.(1)

The gene similarity score in term of Pearsons Correlation Coefficient(PCC) for the seed vertex S and all its neighbouring vertices are given in Table.(2) below

Vertex Pairs	Similarity Score
PCC(S,A)	0.80
PCC(S,B)	0.82
PCC(S,C)	0.57
PCC(S,D)	0.42
PCC(S,R)	0.212

Table.(3) evaluation the Weighted Edge Value(WEV) based on the Edge Clustering Coefficient(ECC) and the value of PCC taking the value of λ=0.7, is given below.

Vertex Pairs	ECC	ECC x (λ)	PCC	PCC*(1-λ)	WEC	Verdict
WEV(SA)	0.25	0.17	0.80	0.24	0.41	NOT SELECTED
WEV(SB)	0.75	0.52	0.82	0.246	0.766	Selected
WEV(SC)	0.66	0.462	0.57	0.171	0.633	Selected
WEV(SD)	0.75	0.525	0.42	0.126	0.94	Selected
WEV(SR)	1	0.7	0.212	0.063	0.763	Selected

Finding PCC: PCC for two gene can be found based on the Pearsons Correlation Coefficient. Here we took some random PCC values for example.

Finding ECC: consider the example of the vertex pair of vertex S and vertex A from the sub-graph G given in figure.(1). Here, vertex S and vertex A has a degree five(i.e 5) and number of only one common neighbor vertex (i.e vertex B). There fore the ECC(S,A) between vertex A and vertex S is given by:

ECC(S,A)=1/(5-1);
Similarly, the ECC values of the vertex S with its neighboring vertices can be evaluated.
ECC(S,B)=3/(5-1); (i.e 3 common neighbor and minimum degree between the degree of S and degree of B is 5)
ECC(S,C)= 2/(4-1); (i.e 2 common neighbor and minimum degree between the degree of S and degree of C is 4)
ECC(S,D)=3/(5-1); (i.e 3 common neighbor and minimum degree between the degree of S and degree of D is 5)
ECC(S,R)=1/(2-1); (i.e 1 common neighbor and minimum degree between the degree of S and degree of R is 2)

Preliminary Cluster: Only the neighboring vertices of vertex S having a WEV value greater than a given threshold value T_weight (here,T_weight= 0.6) can be added to the preliminary cluster. Since vertex A has a WEV value=0.41 with the seed vertex S, it cannot be added to the preliminary cluster. The rest of the neighboring vertices are added to the preliminary cluster.

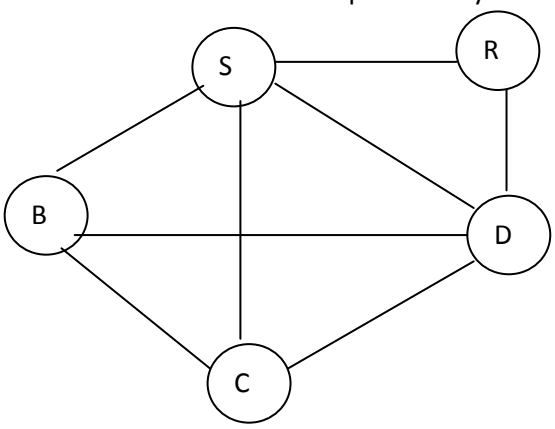

Figure.(3) Preliminary Cluster (C) after adding the neighboring vertices of vertex S having WEC ≥0.6.

Cluster Enrichment: Assuming that this preliminary cluster (C) is not a redundant cluster. The next process is the enrichment process. Here the proteins are added to the preliminary cluster (C) from the neighboring vertices(i.e vertex that has a connection with a vertex in the cluster). However, only those neighboring vertices that has high connectivity with the vertices in the cluster C are added to it. This connectivity is given by Closeness Score(CS). A vertex(v) can be added to the cluster(C) only if the Closeness Score(C,v) ≥T_enrich. Where, T_enrich is an enrichment threshold value(Here, T_enrich =0.5)

The candidate vertices that can be added to the clusters are: A, P, Q, U, V.

The evaluation of the closeness score between the preliminary cluster C and the candidate vertices are given below :

Closeness (C,A): 2/5=0.4; (i.e A connects to 2 vertex in the Cluster C)
Closeness (C,P): 3/5=0.6; (i.e P connects to 3 vertex in the Cluster C)

Closeness (C,V): 1/5=0.2; (i.e V connects to 1 vertex in the Cluster C)

Closeness (C,U): 1/5=0.2; (i.e U connects to 1 vertex in the Cluster C)

Closeness (C,Q): 1/5=0.2; (i.e Q connects to 1 vertex in the Cluster C)

Since, vertex P is the only candidate vertex that has a closeness score greater than the threshold value(T_enrich), only vertex P is added to the Preliminary Cluster C. The rest of the vertices are discarded.

The final complex after adding the enrichment vertex to the preliminary cluster is given below:

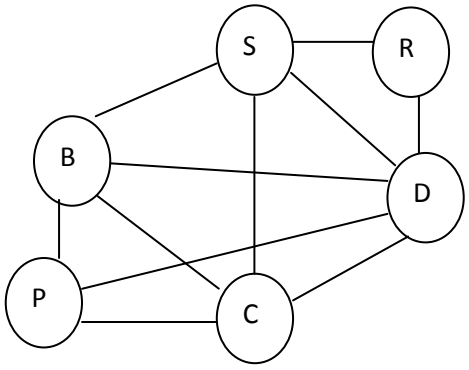

Figure.(4). The final resultant complex after adding the enrichment vertices.