Biclustering

**Subset** of genes may behave similarly under only a **subset** of conditions.

- Only a subset of genes in cellular process of interest.
- Cellular process active in only a subset of conditions.

BMC Genomics 2006, 7:279
Biclustering

A bicluster corresponds to a submatrix

Maximum Bounded Biclique

MaxBoundBiClique(U, V, E, d):
Initialize a hash table weight; weight<sub>best</sub> ← 0
For all v ∈ V do
  For all S ⊆ N(v) do
    weight[S] ← weight[S] + w(S, {v})
    If (weight[S] > weight<sub>best</sub>)
      U<sub>best</sub> ← S
      weight<sub>best</sub> ← weight[S]
  Compute V<sub>best</sub> = \bigcap_{u \in U<sub>best}} N(u)
Output (U<sub>best</sub>, V<sub>best</sub>)
Clustering

Properties of a good clustering/partition.

- **Separation**: points in different clusters are far apart.
- **Homogeneity**: points in the same cluster are close.

\[
\begin{pmatrix}
0 & 11 & 7 & 5 \\
11 & 0 & 4 & 6 \\
7 & 4 & 0 & 9 \\
5 & 6 & 9 & 0
\end{pmatrix}
\]

\(n \times n\) distance matrix

A puzzling picture

Figure 4.6 in Shamir notes.
External Validation of Clusters

Cluster analysis of combined yeast data sets

Is this cluster consistent with known biology?
Are the genes in this group enriched for those with known biological function/process, etc?

Gene Ontology

www.geneontology.org
Gene annotations for Yeast