Social Network Theory - Community Structure [22.03.2006]

Namita Jarika – 02CS3021

Hierarchial Clustering

–Used for the identification of communities in social networks.

–The idea behind this technique is to develop a measure of similarity x_{ij} between pairs (i,j) of vertices.

–The hierarchical clustering method has the advantage that it doesn't require us to specify the size or number of groups we want to look for beforehand. It does not however tell us how many groups should be used to get the best division of the network.

Another weight measure

There are many similarity measures:

One such similarity measure is the count of edge/vertex independent paths between vertices.

Two paths are said to be edge-independent if they share none of the same edges.

Newmann Girvan Algorithm

- 1.Find out the edge-betweenness of all the edges
- 2.Remove the edge which has the highest edge betweenness
- 3.Recalculate step 1&2.Repeat till all edges are exhausted.

Definition of edge-betweenness

The betweenness of an edge is defined to be the number of geodesic (i.e., shortest) paths between vertex pairs that run along the edge in question, summed over all vertex pairs. This quantity can be calculated for all edges in time that goes as O(mn) on a graph with m edges and n vertices.

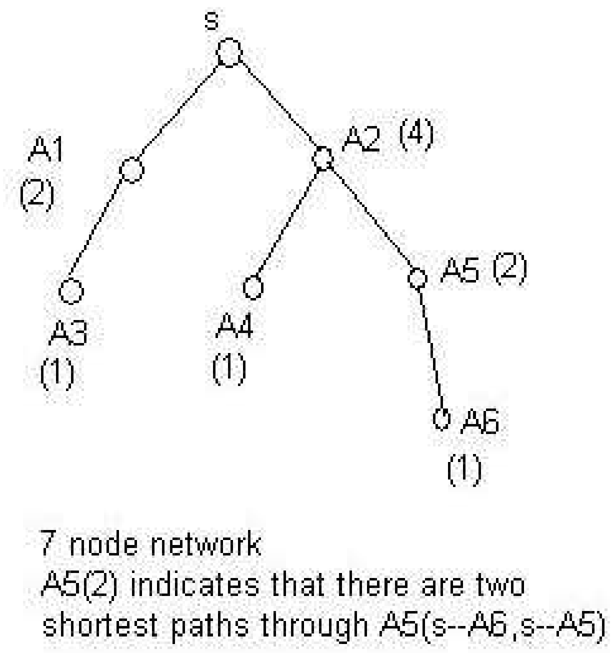

Figure 1: 7 node network

In figure 2, d=depth and w= number of paths passing through the node

Top Down Approach

- If node j is not marked
  $d_j = d_i + 1; w_j = w_i$
- If node j is marked and if $d_j = d_i + 1$
  $w_j = w_j + w_i$
- If $d_j < d_i + 1$ do nothing

Bottom up Approach

- Start from leaf nodes(i)
- EB=$\frac{w_j}{w_i}$ where j is the neighbour of i
- For other nodes(i)
  EB=(incoming EB+1)*$\frac{w_j}{w_i}$ where j is the neighbour of i

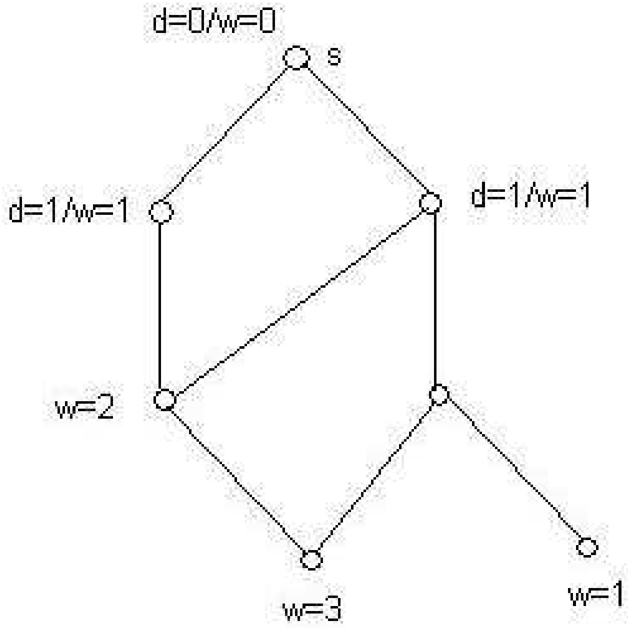

Figure 2: