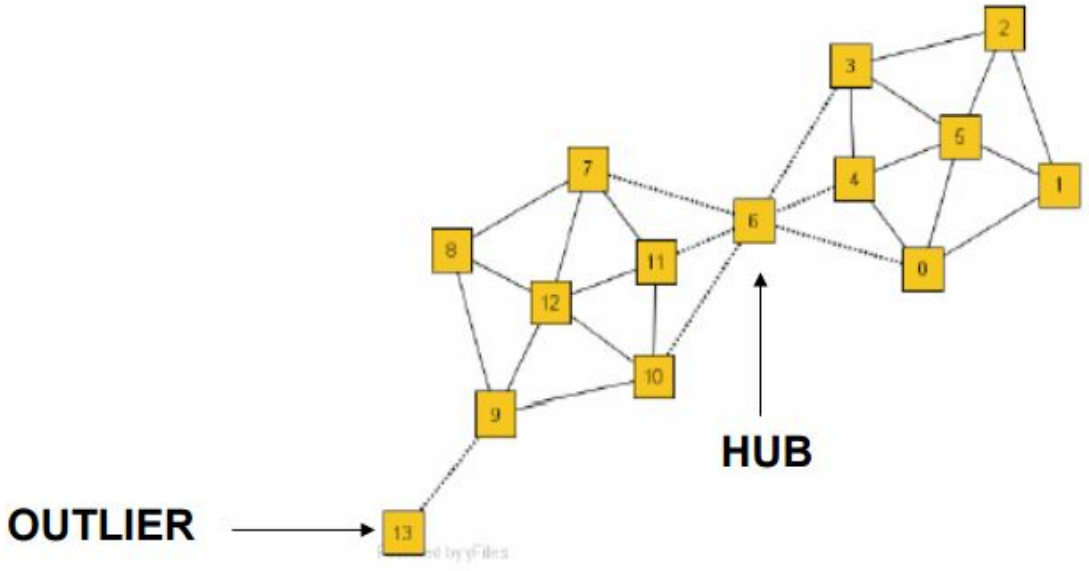
DATA MINING ASSIGNMENT 11
(Graph Clustering)

Debarati Das

1PI13CS052

OVERVIEW (PART A)

Structure Connected Clusters work on the principle of Structure Connectivity where they consider the neighborhood around two connected vertices . These two vertices are assigned to a cluster according to how they share neighbors, and neighbors may even be outliers or hubs.



A HUB has the following properties :

1. If v is not a member of any cluster
2. If v bridges different clusters

An OUTLIER has the following properties:

1. If v is not a member of any cluster
2. If v does not bridge different clusters

ANSWERS TO QUESTIONS

1. **What is the effect of varying epsilon or popularity threshold?**

As popularity threshold (μ) increases, more nodes can enter the cluster. This would result in large cluster population.

As epsilon increases, this puts a limit on the number of nodes that can be a part of the cluster therefore resulting in :

> Small clusters

>Higher Number of clusters and outliers (because now the bigger clusters will break up resulting in small clusters and many outliers)

It was observed that when μ=2 and epsilon=0.7 ,

The number of hubs were just 1.

When μ=3 and epsilon=0.6 (popularity threshold increased and epsilon limit decreased):

The number of hubs were now 2.