

ABSTRACT

The ability to gather genome-wide or gene expression data using Microarray Technology has paved the way towards the production of massive amounts of data daily and steered the monitoring of expression levels of thousands of genes simultaneously, across different conditions and over different time-points. Analysis of data produced by such experiments can put forward prospective insight into characteristic functions of genes and regulatory mechanisms of the life cycle. A key step in the analysis of such data is grouping the genes that manifest similar characteristics, which is popularly referred to as clustering. This work comprises of a brief study on some of the clustering approaches developed till date and their relevance in gene expression data clustering. It then presents an enhanced graph based clustering approach coupled with an embedded cluster detection technique, *GCE* (Graph based Clustering with Embedded Cluster Detection). The clustering technique adopted in this approach is completely parameter independent and produces clusters of good quality both in terms of visual representations and cluster validity measures. The embedded cluster detection method incorporated in this approach searches for highly coherent patterns of genes within a cluster and detects any embedded clusters present in it, so that biologically relevant genes having high co-regulation are obtained. The proposed method has been tested on four real life datasets and the results are found to be comparable with those of other algorithms.

***Keywords:* cluster, affinity, homogeneity, separation, embedded cluster, coherent patterns.**