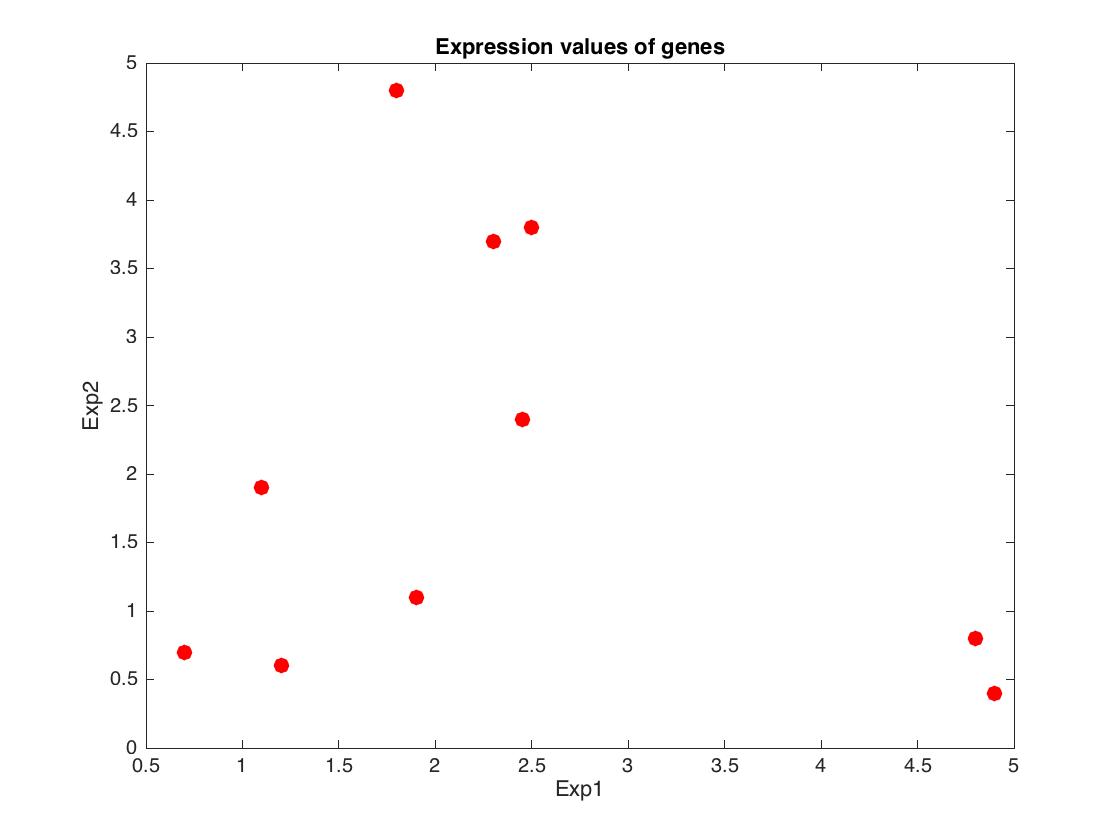
Q1.

1. The only changes applicable would be to equation used for computing similarity matrix, which are as follows:

The (*lambda)* refers to opening gaps along with gap extensions . Since the only modification proposed is the use of an additional constant in penalising the opening gaps, this would not make cause any change to the number of iterations in dynamic program, and hence resulting with no change in computational complexity.

2. a. Euclidean distance between the genes.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Gene # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | 1 | 0 | 2.36 | 1.44 | 2.98 | 2.16 | 3.86 | 1.13 | 1.3 | 1.26 | 4.09 | | 2 | 2.36 | 0 | 1.4 | 1.22 | 0.22 | 3.78 | 2.77 | 3.45 | 3.58 | 4.16 | | 3 | 1.44 | 1.4 | 0 | 2.49 | 1.31 | 2.84 | 1.41 | 2.19 | 2.44 | 3.16 | | 4 | 2.98 | 1.22 | 2.49 | 0 | 1.21 | 5 | 3.7 | 4.24 | 4.24 | 5.38 | | 5 | 2.16 | 0.22 | 1.31 | 1.21 | 0 | 3.83 | 2.63 | 3.29 | 3.4 | 4.2 | | 6 | 3.86 | 3.78 | 2.84 | 5 | 3.83 | 0 | 2.92 | 3.61 | 4.1 | 0.41 | | 7 | 1.13 | 2.77 | 1.41 | 3.7 | 2.63 | 2.92 | 0 | 0.86 | 1.26 | 3.08 | | 8 | 1.3 | 3.45 | 2.19 | 4.24 | 3.29 | 3.61 | 0.86 | 0 | 0.51 | 3.71 | | 9 | 1.26 | 3.58 | 2.44 | 4.24 | 3.4 | 4.1 | 1.26 | 0.51 | 0 | 4.21 | | 10 | 4.09 | 4.16 | 3.16 | 5.38 | 4.2 | 0.41 | 3.08 | 3.71 | 4.21 | 0 | |



b. K-medoids clustering

Result after the first epoch/iteration:

|  |  |  |
| --- | --- | --- |
| Cluster # | Gene # | Medoid |
| 1 | 1,7,8,9 | 7 |
| 2 | 2,4,5 | 4 |
| 3 | 3,6,10 | 6 |

Result after the second epoch/iteration:

|  |  |  |
| --- | --- | --- |
| Cluster # | Gene # | Medoid |
| 1 | 1,7,8,9 | 8 |
| 2 | 2,4,5 | 5 |
| 3 | 3,6,10 | 10 |

Result after the third epoch/iteration:

|  |  |  |
| --- | --- | --- |
| Cluster # | Gene # | Medoid |
| 1 | 1,7,8,9 | 9 |
| 2 | 2,4,5 | 2 |
| 3 | 3,6,10 | 3 |

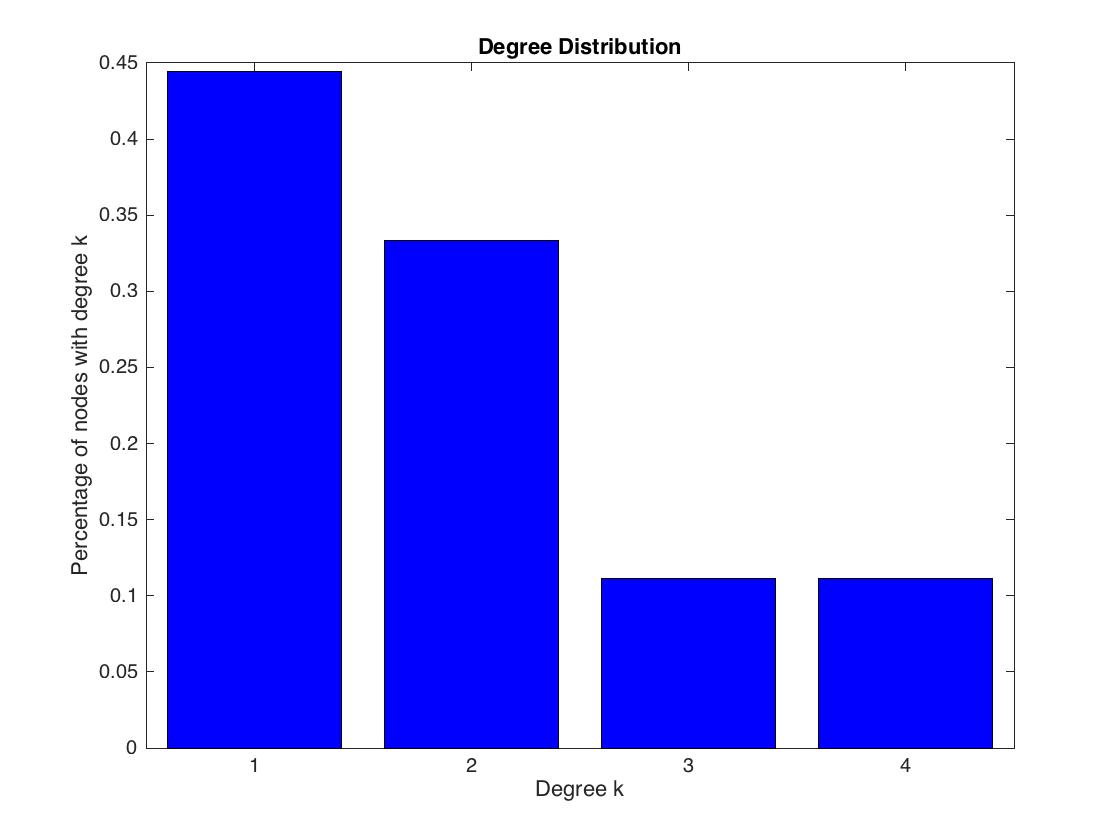
There is a swap in medoid of every cluster, and this is because there is more than 1 data-point, which has the least distance to non-medoid points, and during such case randomness is used to pick one among the data-points with least cost.

Based on the results on each step, it is intuitive that the provided dataset does not carry noise, which makes the algorithm converge in the first step itself. The first step refers to associating non-medoid points to closest medoids before any epochs.

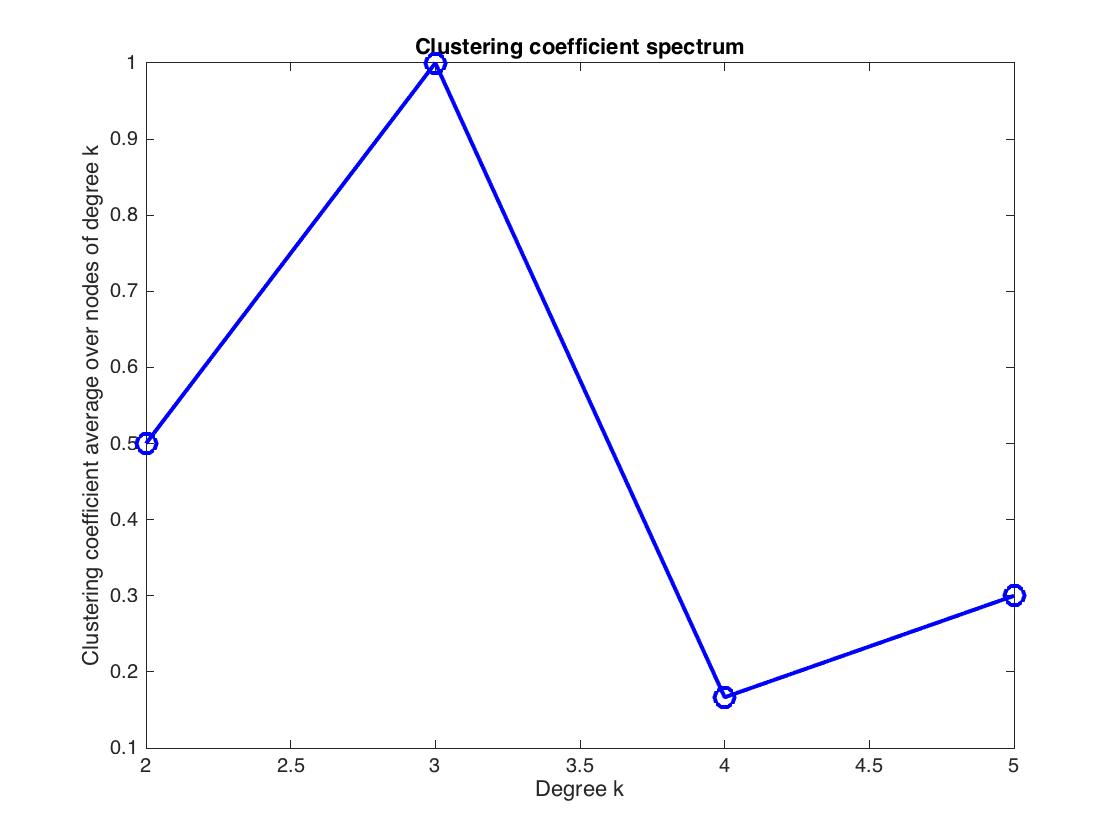
Q2.

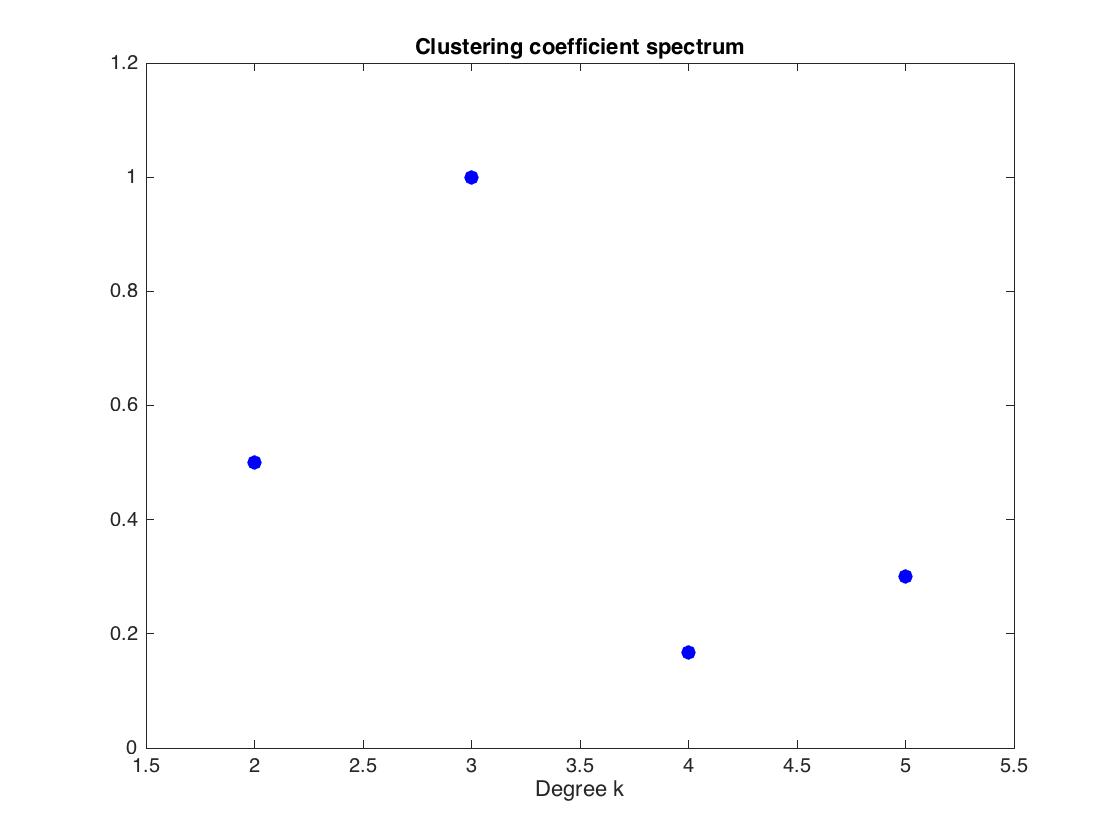
1. In general, proteins do not function in an isolated manner, rather they tend to interact with each other, for instance molecules, which would intervene with metabolic pathways. Metabolic pathways of a cell constitute to the formation of metabolic network, which are interconnected by biochemical reactions that are catalysed by enzymes: *proteins that accelerate chemical reactions* in a cell.

2. a. Degree Distribution



b. Clustering coefficient spectrum





Both the figures are just the same except the second one's bounds were slightly adjusted to make the points more prominent without lines.

c. Shortest path length distribution