There was no need of splitting in this case since we had to work with whole data. We compute distances between the rows of the whole data using dist function. We then compute the first quartile of the distances using fivenum function (which came out to be 3 in our case). This first quartile of the distances is our threshold. Then, we create an adjacency matrix (from our distance matrix) in which we set those entries to 1 for which distance <= threshold and set other entries to 0.

So, we keep an edge between 2 nodes (rows of the dataset) if the distance between the 2 nodes <= threshold. We then plot an undirected graph with no self-loops using this adjacency matrix.

So, for the 625 nodes, on plotting we got 52672 edges. The node-to-edge ratio is 0.0118.

Then, we studied few properties of the graph which are listed below:

1. Radius: We computed the radius using radius function. It came out to be 2 for our graph.

It is the minimum eccentricity among all the vertices of G.

1. Diameter: We computed it using the diameter function. It came out to be 4 for our graph.

It is the maximum distance between 2 vertices, i.e. it is the maximum eccentricity among all the vertices of G.

We also computed the eccentricity (which is the maximum distance from it to any other vertex) for all vertices using eccentricity function. The eccentricity values ranged from 2 to 4 as expected.

1. Degree distribution: We found degree distribution using degree.distribution function and we then plotted a line graph. The plotted graph shows degree (k) on x-axis and the probability that degree = k on y-axis.

For our graph, the degrees are distributed as shown below:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Degree, k | 69 | 92 | 108 | 123 | 145 | 165 | 170 | 195 | 222 | 229 | 263 | 267 | 309 | 360 | 416 |
| Frequency (Nk) | 16 | 64 | 32 | 96 | 96 | 64 | 24 | 96 | 16 | 48 | 32 | 8 | 24 | 8 | 1 |
| P[deg = k] | 0.0256 | 0.1024 | 0.0512 | 0.1536 | 0.1536 | 0.1024 | 0.0384 | 0.1536 | 0.0256 | 0.0768 | 0.0512 | 0.0128 | 0.0384 | 0.0128 | 0.0016 |

To find the frequency of degrees (number of nodes having that degree), we used degree function on the graph and table function on its result.

We have plotted both Degree distribution as described above and also a graph showing degrees plotted on x-axis with their frequencies plotted on y-axis.

Our graph has 625 nodes. So, the degree of a node can lie between 0 and 624. For all the degrees not shown in the table, Nk = 0 and hence P[deg = k] is also 0.

The probability that a node takes a particular degree k can be found out by dividing the number of nodes having that degree by the total number of nodes.

For better understanding, let us take 2 examples from our graph.

For degree 0, frequency i.e. number of nodes having degree 0 = 0

And so probability that a node has degree 0 is also zero.

For degree say 145, frequency i.e. number of nodes having degree 145 = 96

And so probability that a node has degree 145 is 96/625 = 0.1536

For our graph, since there are only 15 degrees that are taken by nodes out of the possible values (0 to 624). So, the probability that degree = k for most degrees k is 0 except these 15 values for which the probability that a node has degree = one of these values is shown in the table. That is why, when we plot the line graph of the degree distribution, we get a straight line at 0 as for most degrees k, P[deg=k] = 0 and few (15) dots above at their respective probabilities.