Assignment BDA PartB

1.a.)

Diameter of a network is defined as the number of steps required for two distant nodes in the network to reach one another.

Diameter of a network can be measured by considering the network to be either

1.) Directed, or 2.) Undirected.

Let’s consider case to be Directed.

Following are the outward paths:

|  |
| --- |
| 1->4  1->5  1->7 |
| 4->8 |
| 0->4 |
| 2->5 |
| 0->2 |
| 7->8 |
| 3->7 |
| 3->8 |
| 3->9 |
| 6->8 |
| 0->8 |

While the diameter for the network for undirected graph is ‘4’.

Since, we are considering the network to be directed the diameter would be 2:

Paths that form diameter are:

1->4->8

1->7->8

0->2->5

1.b.) refer to 1b.csv

1.c.)

A network with numbers and lines

AI-generated content may be incorrect.

1.d.) Adjacency matrix of the network with aji=1 if there is an edge from node i -> node j

We have 10 nodes, labelled {0,1,2,3,4,5,6,7,8,9}

So, our adjacency matrix is a 10X10 matrix.

Pasting the snapshot of the 10X10 matrix:

A grid of numbers on a white background

AI-generated content may be incorrect.

1.e.) As how an adjacency matrix represents direct connections i.e. paths with just 1 hop.

Similarly, A^2 implies Number of 2-hops from node I to node j and A^3 implies number of 3 hop paths between node I and node j.

For programmatic calculation of A^2 and A^3, refer to 1e.ipynb

A snapshot of A^2 and A^3 from the output is as follows:

A^2:

A screenshot of a computer code

AI-generated content may be incorrect.

Examine the following paths:

0->2->5 (it takes two hops for 0 to reach 5)

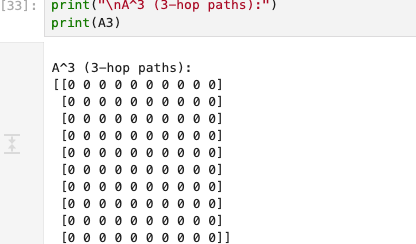
0->4->8 (although there is a direct connection from 0->8, there is also 2-hop path)

1->7->8

1->4->8

3->7->8

A^3:



Similarly, examine the 3-hop paths in the network:

There are none, because the diameter of the network is itself 2.

Which means the maximum steps for two distant nodes to reach each other is only 2 hops.

1.f.) When we consider a network to be undirected, the node I if connected node j is not just considered i->j, but it is considered i-j.

It means, the aij=1 and also aji=1. Unlike the case of them being in a directed graph,

Where aij=1 and aji=0.

In that case, following are the nodes that are connected to node 0 as root.

Level1:

0-2

0-4

0-8

Level 2:

0-2-5

0-4-1

0-8-7

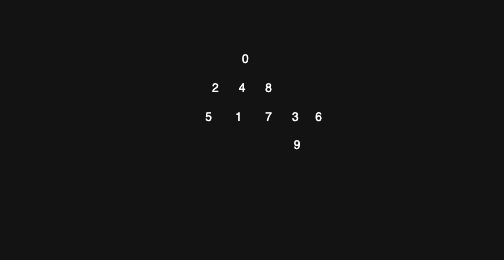
0-8-3

0-8-6

Level3:

0-8-3-9

Hence, the BFS can look like:



1.g.) As I have explained above, aij=1 and aji=1 for an undirected graph:

Here is a snapshot of adjacency matrix for an undirected network:

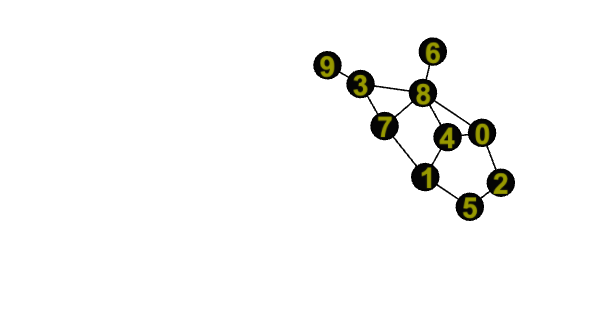
A screenshot of a computer code

AI-generated content may be incorrect.

Answer 2:

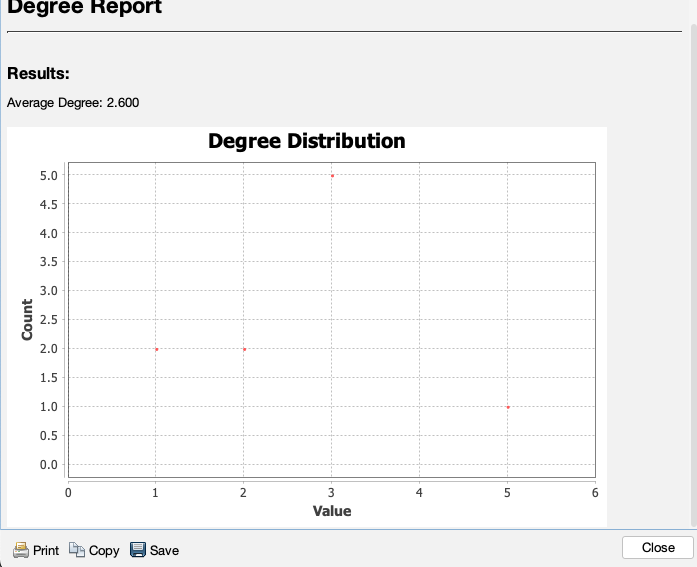
2.a) 2a.csv is the modified csv.

Here is the undirected network:

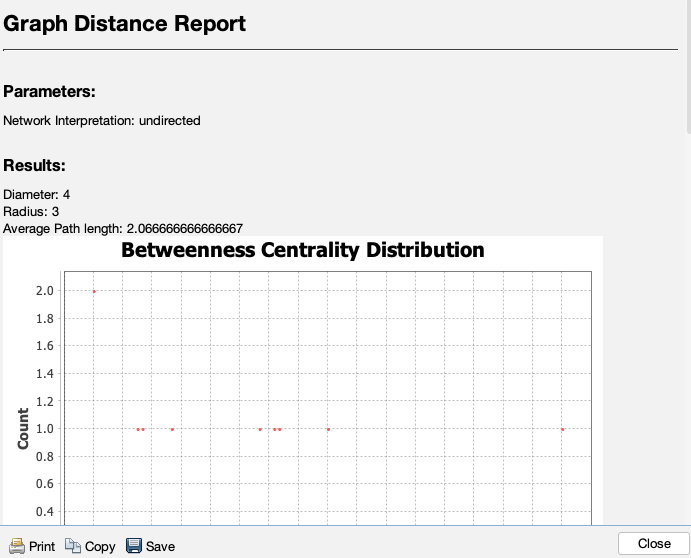


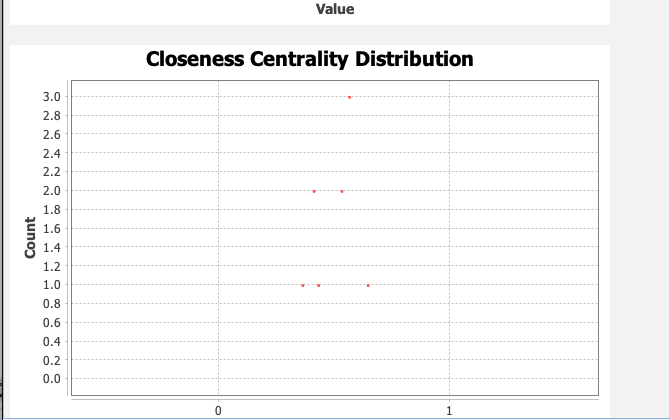
Here are the Average Degree, Network Diameter, Graph Density snapshots from gephi:

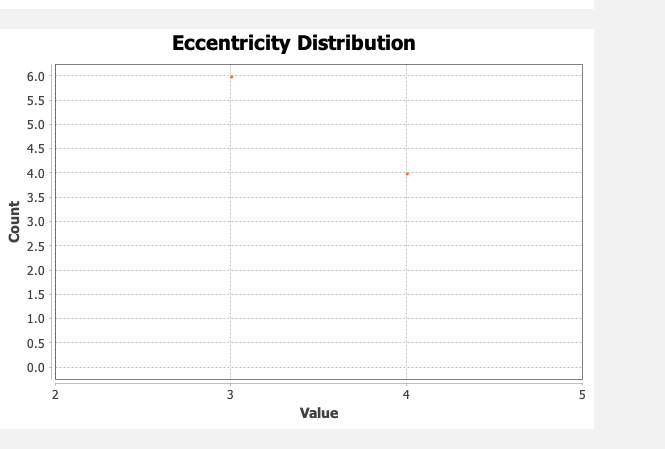
Average Degree:



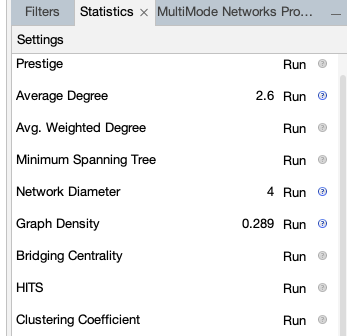
Network Diameter:







Graph Density:



Now, let’s derive them without using tool:

Network Diameter represented by

**Diameter(G)=u,v∈Vmax​d(u,v)**

Of all the total paths in the undirected network, The longest paths are as follows:

5->1>-7->3->9 (it takes 4 steps for 5 to reach 9)

And vice versa

9->3->7->1->5

And similarly, we have one more path which forms the diameter

2->0->4->8->6

And vice versa

6->8->4->0->2

Average Degree:

For an undirected network,

Average degree = 2 X Total Edges/ Number of nodes (since every edge has to calculated twice)

For a Directed network,

Average degree = In degree edges + out degree edges

Indegree edges= Number of Indegree edges/ Number of nodes

Outdegree edges = Number of outdegree edges/ Number of nodes

Considering our network to be undirected network:

**Average degree = 2\*13/10= 2.6**

**Graph Density:**

Graph Density is measure of how connected the network is.

A measure for this is the ratio of Edges in the network/number of possible edges in the network

For a undirected graph:

Density = 2 X E/(V X (V-1))

For a directed graph:

Density = E/(V X (V-1))

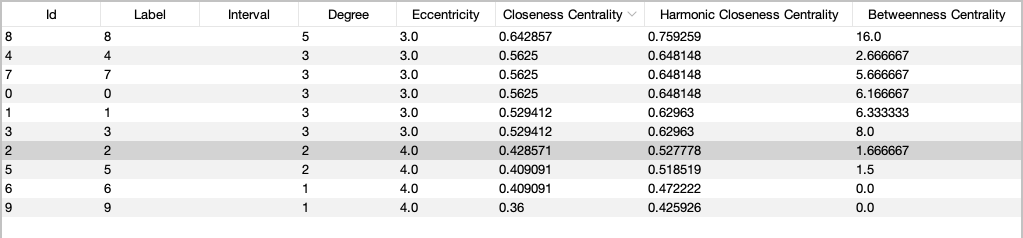
Considering our network to be undirected, then the density is

**2 X 13/(10 X (10-1)) = 26/90 = 0.289**

**2.b.)**

Closeness centrality is defined as the node even though lesser number of direct connections but can reach every node in as many fewer steps as possible.

This is a snapshot from data laboratory:

****

According to gephi node 8 has the highest closeness centrality

Now, lets derive that value without using gephi:

Step-by-step for Node 8:

Distance to:

Node 3 = 1 step

Node 0 = 1 step

Node 6 = 1 step

Node 7 = 1 step

Node 4 = 1 step

Node 9 = 2 steps (8->3->9)

Node 1 = 2 steps (8->4->1)

Node 2= 2 steps (8->0->2)

Node 5= 3 steps (8->4->1->5)

Sum of distances = 14

Total nodes =10

Closeness centrality = reciprocal of sum of shortest paths to other nodes=

(n-1)/sum of d(u,v)

**=(10-1)/14= 9/14= 0.642857**

**2.c.):**

Clustering coefficient is the level at which nodes are grouped together as opposed to being equally or randomly connected across the network.

This can be measured by calculated the number of triangles.

C(v) = 2 X Number of edges between neighbors of v / ( degree of v X (degree of v -1))

Find the close triplets first then calculate the ratio of closed triplets and possible triplets.

Closed triplet is a set of 3 nodes with three edges between them

Open triplet is a set of 3 nodes with two of three edges between them

Let’s calculate that for node 4:

Neighbours of node 4 N(4) = {0,1,8}

Closed triplets:

(0,8) 🡪 of the neighbours of 4, only 0,8 have a direct connection between them.

Other edges do not have a direct connection between them.

Degree of node 4 = 3 {0,1,8}

Maximum possible edges = n\*(n-1)/2 = 3\*2/2 = 3

Hence, clustering coefficient for node 4= 2\*1/3\*2 = **0.3333**

Below snapshot from gephi also indicates same.

A screenshot of a computer

AI-generated content may be incorrect.

Along with node 4, nodes 7, 0, 3 also have the same clustering coefficient of 0.3333

Question 3.:)

3.a.:) A network with numbers and lines

AI-generated content may be incorrect.

When 0,1,4,8 are interpreted as girls, and the rest as boys above is how the network would look like.

The probability of being a boy is p=0.6

The probability of being a girl is p= 0.4

Hence the probability for cross node edges = 2pq= 2\*0.6\*0.4= 0.48

Hence, if we find a total of 0.48\*13 = 6.24~ roughly 6edges, then we can conclude the network connections to be random.

If the total cross gender connections are considerably less than 6, then we can safely conclude to be homophily.

If the total cross gender connections are considerably more than 6, then it is inverse homophily.

In our case the total cross gender connections are

7-8

1-5

1-7,

2-0,

8-6

3-8

So, we have exactly 6 connections, Hence they are mostly random connections and no sufficient evidence to detect any homophily.

3.b.)

Below is the network when 2,5,6,9 are girls and the rest are boys.

A network with numbers and circles

AI-generated content may be incorrect.

The probability of being a boy is p=0.6

The probability of being a girl is p= 0.4

Hence the probability for cross node edges = 2pq= 2\*0.6\*0.4= 0.48

Hence the probability for cross node edges = 2pq= 2\*0.6\*0.4= 0.48

Hence, if we find a total of 0.48\*13 = 6.24~ roughly 6edges, then we can conclude the network connections to be random.

If the total cross gender connections are considerably less than 6, then we can safely conclude to be homophily.

If the total cross gender connections are considerably more than 6, then it is inverse homophily.

In our case, the total cross gender connections are

1-5

0-2

3-9

6-8

So, there are only cross gender connections which is considerably less than 6.

So, we can conclude that there is homophily.