Name: Amey Sadanand Bhilegaonkar
ASU Number: 1225368924

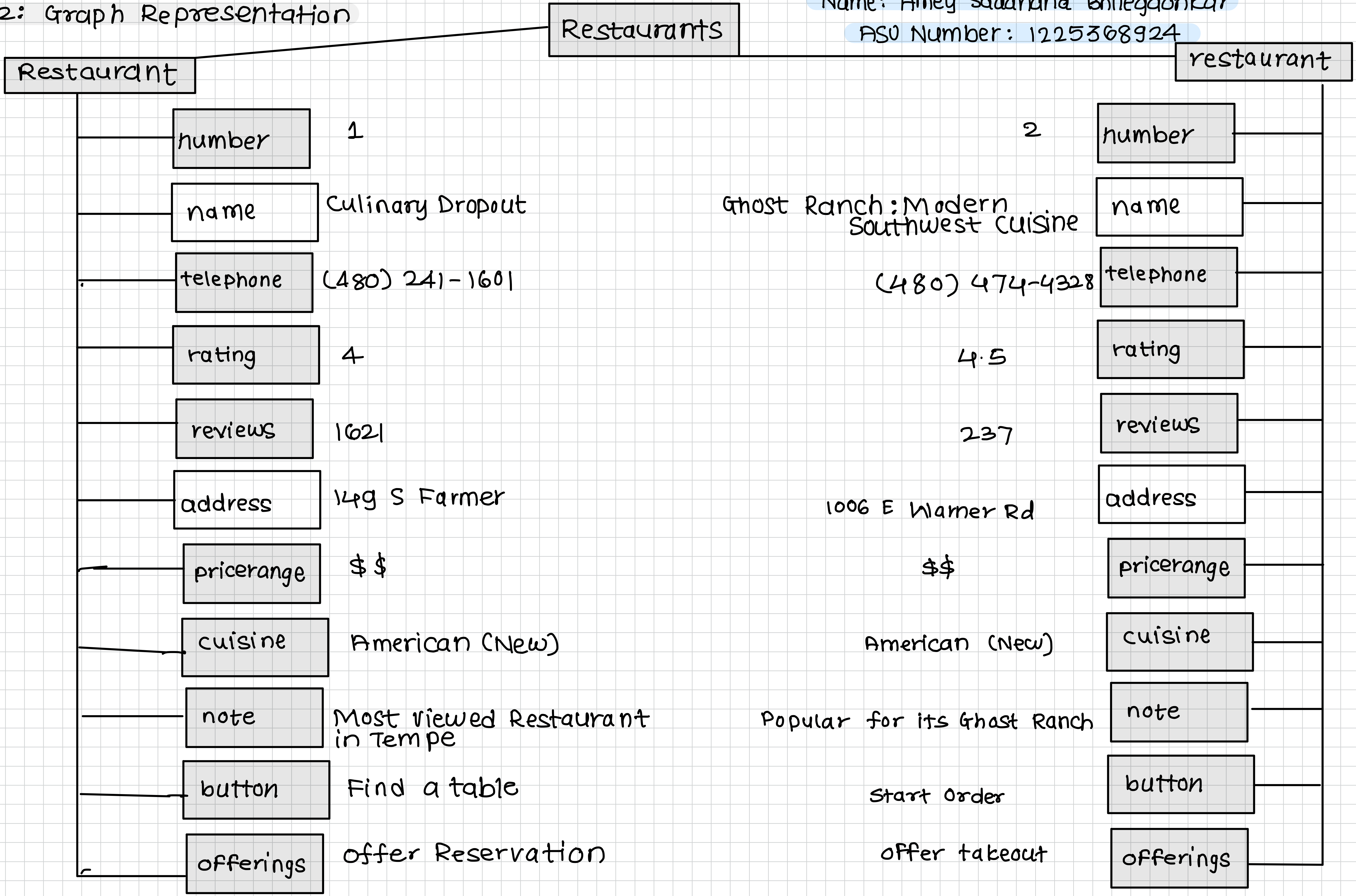
Part I: Sementic Web Languages:

Q.1 Translate HTML information into an XML object.

```
<restaurants>
    <restaurant>
        <number> 1 </number>
        <name> Culinary Dropout </name>
        <telephone> (480) 240-1601 </telephone>
        <rating> 4 </rating>
        <reviews> 1621 </reviews>
        <address> 149 S Farmer road </address>
        <pricerange> $$ </pricerange>
        <cuisine> American (New) </cuisine>
        <note> Most viewed restaurant in Tempe </note>
        <button> Find a table </button>
        <offerings> offer reservations </offerings>
    </restaurant>
    <restaurant>
        <number> 2 </number>
        <name> Ghost Ranch: Modern Southwest Cuisine </name>
        <telephone> (480) 474-4328 </telephone>
        <rating> 4.5 </rating>
        <reviews> 237 </reviews>
        <address> 1006 E Warner Rd </address>
        <pricerange> $$ </pricerange>
        <cuisine> American (New) </cuisine>
        <note> Popular for its Ghost Ranch </note>
        <button> Start Order </button>
        <offerings> offers takeout </offerings>
    </restaurant>
```

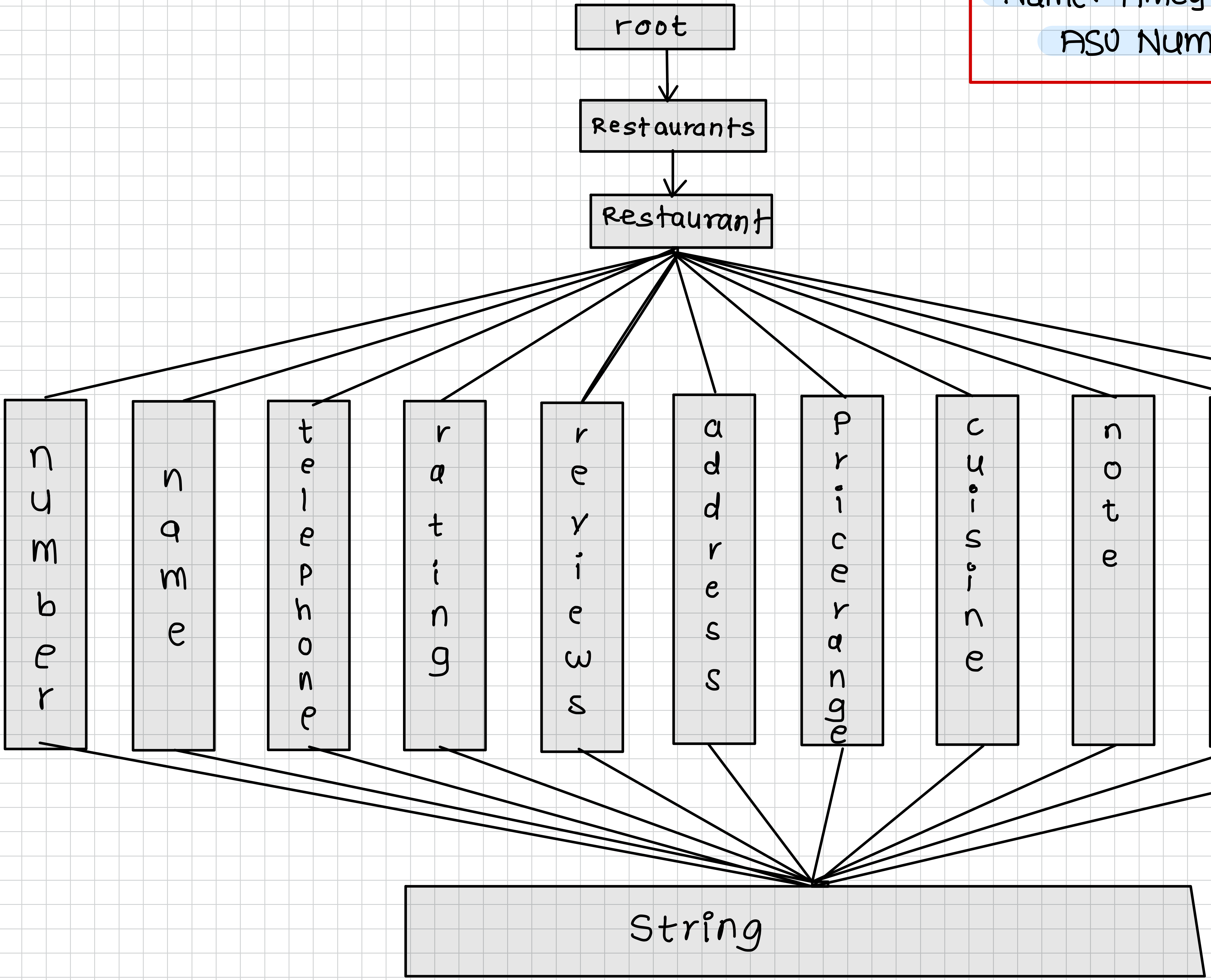
Q2: Graph Representation

Name: Amey Sadanand Bhilegaonkar
ASU Number: 1225368924



Lower Bound Schema

Name: Amey Sadanand Bhilegaonkar
ASU Number: 1225368924



Name: Amey Sadanand Bhilegaonkar
ASU Number: 1225368924

Q.3 Given XML-QL Query, its English translation is:

Select the telephone numbers of all restaurants listed on Yelp who offer takeout.

Part 2: Frequent Item-Sets and Association Rules:

Q.4 APRIORI Algorithm

Given transactions are:

T1 = {A, B, C}
T2 = {A, F, G}
T3 = {B, C, G}
T4 = {A, B, C, F}
T5 = {B, C}
T6 = {B, C, D, E}
T7 = {E, F}
T8 = {B, C, F, G}
T9 = {B, C, F, G}
T10 = {B, C, G}

Given minimum support is 40% which is 4/10

For the first pass-

item	Frequency
A	3
B	8
C	8
D	1
E	2
F	5
G	5

remove items with frequency < 4

∴ {A}, {D}, {E} will be removed

∴ F1 : {B}, {C}, {F}, {G}

For the second part

item	Frequency
{B, C}	8
{B, F}	3
{B, G}	4
{C, F}	3
{C, G}	4
{F, G}	3

- {B, F}, {C, F}, {F, G} will be removed
- ∴ F2 = {B, C}, {B, G}, {C, G}

For the third part

item	Frequency
{B, C, G}	4

∴ F3 = {B, C, G}

Name: Amey Sadanand Bhilegaonkar
ASU Number: 1225368924

Q5. Given support is 40% and confidence is 100%

$\{B, C, G\}$ is the most frequent dataset

The association Rules:

Rule	Confidence Calculation	Confidence	Rule	Confidence Calculations	Confidence
1. $\{B,C\} \rightarrow \{G\}$	$\frac{\text{Support }\{B,C,G\}}{\text{support }\{B,C\}} = \frac{4/10}{8/10} = 0.5$	50%	8. $\{G\} \rightarrow \{B\}$	$\frac{\text{Support }\{B,G\}}{\text{Support }\{G\}} = \frac{4/10}{5/10}$	80%
2. $\{B,G\} \rightarrow \{C\}$	$\frac{\text{Support }\{B,C,G\}}{\text{Support }\{B,G\}} = \frac{4/10}{4/10}$	100%	9. $\{B\} \rightarrow \{C\}$	$\frac{\text{Support }\{B,C\}}{\text{Support }\{B\}} = \frac{8/10}{8/10}$	100%
3. $\{C,G\} \rightarrow \{B\}$	$\frac{\text{Support }\{B,C,G\}}{\text{Support }\{C,G\}} = \frac{4/10}{4/10}$	100%	10. $\{C\} \rightarrow \{B\}$	$\frac{\text{Support }\{B,C\}}{\text{Support }\{C\}} = \frac{8/10}{8/10}$	100%
4. $\{B\} \rightarrow \{C,G\}$	$\frac{\text{Support }\{B,C,G\}}{\text{Support }\{B\}} = \frac{4/10}{8/10}$	50%	11. $\{C\} \rightarrow \{G\}$	$\frac{\text{Support }\{C,G\}}{\text{Support }\{G\}} = \frac{4/10}{5/10}$	80%
5. $\{C\} \rightarrow \{B,G\}$	$\frac{\text{Support }\{B,C,G\}}{\text{Support }\{C\}} = \frac{4/10}{8/10}$	50%	12. $\{G\} \rightarrow \{C\}$	$\frac{\text{Support }\{C,G\}}{\text{Support }\{G\}} = \frac{4/10}{5/10}$	80%
6. $\{G\} \rightarrow \{B,C\}$	$\frac{\text{Support }\{B,C,G\}}{\text{Support }\{G\}} = \frac{4/10}{5/10}$	80%			
7. $\{B\} \rightarrow \{G\}$	$\frac{\text{Support }\{B,G\}}{\text{Support }\{B\}} = \frac{4/10}{8/10}$	50%			

$\therefore$ The association rules with support 40% and confidence 100% are –

- $\{B,G\} \rightarrow \{C\}$
- $\{C,G\} \rightarrow \{B\}$
- $\{B\} \rightarrow \{C\}$
- $\{C\} \rightarrow \{B\}$

Part 3: Clustering

Name: Amey Sadanand Bhilegaonkar

ASU Number: 1225368924

Q6: Given C_1, C_2, C_3 after first iteration:

$C_1 = \{(2,2), (4,4), (6,6)\}$

$C_2 = \{(0,4), (4,0)\}$

$C_3 = \{(5,5), (9,9)\}$

For 2nd iteration, we need to calculation the centroids

$\therefore$ centroids of (x_1, y_1) & (x_2, y_2) is $\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$

$$\text{Centroid of } C_1 = \left(\frac{2+4+6}{3}, \frac{2+4+6}{3}\right) = (4,4)$$

$$\text{centroid of } C_2 = \left(\frac{0+4}{2}, \frac{4+0}{2}\right) = (2,2)$$

$$\text{centroid of } C_3 = \left(\frac{5+9}{2}, \frac{5+9}{2}\right) = (7,7)$$

$\therefore$ Centroid for $C_1 \rightarrow (4,4)$

Centroid for $C_2 \rightarrow (2,2)$

Centroid for $C_3 \rightarrow (7,7)$

Q7: New Clusters

For second iteration:

using new centroids:

$C_1 = (4,4)$

$C_2 = (2,2)$

$C_3 = (7,7)$

After second iteration

$C_1 = \{(4,4), (5,5)\}$

$C_2 = \{(2,2), (0,4), (4,0)\}$

$C_3 = \{(6,6), (9,9)\}$

Distance d between (x_1, y_1) & (x_2, y_2) $= \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$

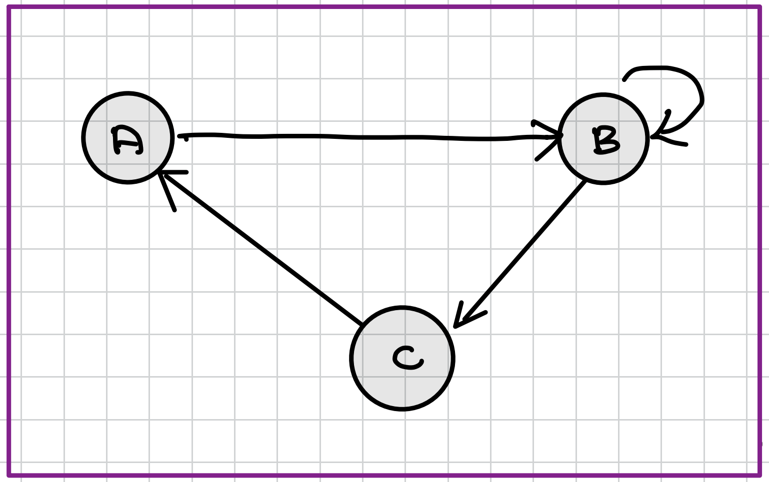
x	y	$C_1 = (4,4)$ distance 1	$C_2 = (2,2)$ distance 2	$C_3 = (7,7)$ distance 3	cluster to which it belongs
2	2	2.83	0.00	7.07	C_2
4	4	0.00	2.83	4.24	C_1
6	6	2.83	5.66	1.41	C_3
0	4	4.0	2.83	7.62	C_2
4	0	4.0	2.83	7.62	C_2
5	5	1.412	4.24	2.83	C_1
9	9	7.07	9.9	2.83	C_3

Name: Amey Sadanand Bhilegaonkar
ASU Number: 1225368924

Part : 4 Web Ranking:

Q 8 Given edges are: (A→B), (B→B), (B→C), (C→A)

The network graph will be as follows:



$$\text{Adj Matrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\text{Column Normalized Matrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0.5 & 0 \\ 0 & 0.5 & 0 \end{bmatrix}$$

$$\text{Transpose of Adj Matrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Q 9: Irreducible Matrix

The matrix M is irreducible

Explaination:

- There exists a path between all nodes. i.e. we can reach any other node if we choose to start from any node in the network.

Lets consider:

Start Point	End point	Edge | path
A	B	A→B
A	C	A→B→C
B	A	B→C→A
B	C	B→C
C	A	C→A
C	B	C→A→B

∴ The matrix is irreducible.

Q.10 Page Rank

Name: Amey Sadanand Bhilegaonkar
ASU Number: 1225368924

Column Normalized Hyperlink matrix M is:

$$M = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1/2 & 0 \\ 0 & 1/2 & 0 \end{bmatrix} \qquad X = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Replace x by product MX until page rankings converge.

Iteration ①

$$MX = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1/2 & 0 \\ 0 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0\times1 + 0\times1 + 1\times1 \\ 1\times1 + 1\times1/2 + 0\times1 \\ 0\times1 + 1/2\times1 + 0\times1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.5 \\ 0.5 \end{bmatrix}$$

Iteration ②

$$MX = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1/2 & 0 \\ 0 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0\times1 + 0\times1.5 + 1\times0.5 \\ 1\times1 + \frac{1}{2}\times1.5 + 0\times0.5 \\ 0\times1 + \frac{1}{2}\times1.5 + 0\times0.5 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.75 \\ 0.75 \end{bmatrix}$$

Iteration 3:

$$MX = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1/2 & 0 \\ 0 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} 0.5 \\ 1.75 \\ 0.75 \end{bmatrix} = \begin{bmatrix} 0\times0.5 + 0\times1.75 + 1\times0.75 \\ 1\times0.5 + \frac{1}{2}\times1.75 + 0\times0.75 \\ 0\times0.5 + \frac{1}{2}\times1.75 + 0\times0.75 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 1.375 \\ 0.875 \end{bmatrix}$$

iteration	X_A	X_B	X_C
0	1	1	1
1	1	1.5	0.5
2	0.5	1.75	0.75
3	0.75	1.375	0.875

The page rank algorithm stops when rankings converge.

In iteration ① : rankings are BAC

In iteration ② : rankings are BCA

In iteration ③ : rankings are BCA.

∴ The page rank algorithm stops after iteration ③