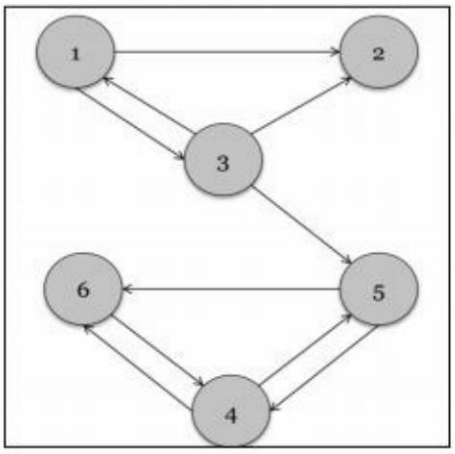
Theory DA-1, Winter 2020-21, C2+TC2

WEB MINING

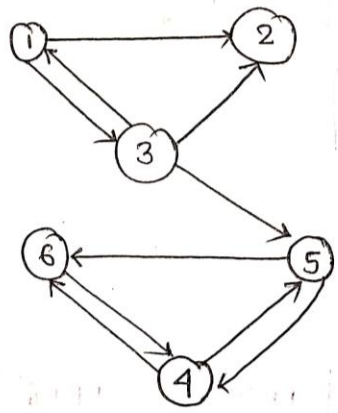
by Sharadindu Adhikari, 19BCE2105

1) Estimate the page rank for the given directed graph representing web of six pages and damping factor is **0.9(d)** and **Max. Iterations are 2**.



Solution:

1



Given: damping factor, d = 0.9

max. iterations = 2

W.K. initial pagerank (PR) for all pages (1 through 6) = 1

- Iteration : 0

$PR(1) = PR(2) = PR(3) = PR(4) = PR(5) = PR(6) = 1$

- Iteration : 1

$$PR(x) = (1-d) + d\left[\frac{PR(T_1)}{C(T_1)} + \cdots + \frac{PR(T_n)}{C(T_n)}\right]$$

$$\therefore PR(1) = (1-0.9) + 0.9\left[\frac{PR(3)}{C(3)}\right]$$

$$= 0.1 + 0.9\left[\frac{1}{3}\right]$$

$$= \frac{2}{5} = 0.4$$

$$\therefore PR(2) = (1-0.9) + 0.9\left[\frac{PR(1)}{C(1)} + \frac{PR(3)}{C(3)}\right]$$

$$= 0.1 + 0.9\left[\frac{0.4}{2} + \frac{1}{3}\right]$$

$$= \frac{29}{50} = 0.58$$

$$\therefore PR(3) = (1-0.9) + 0.9\left[\frac{PR(1)}{C(1)}\right]$$

$$= 0.1 + 0.9\left[\frac{0.4}{2}\right]$$

$$= \frac{7}{25} = 0.28$$

$$\therefore PR(4) = (1-0.9) + 0.9\left[\frac{PR(5)}{C(5)} + \frac{PR(6)}{C(6)}\right]$$

$$= 0.1 + 0.9\left[\frac{1}{2} + \frac{1}{1}\right]$$

$$= \frac{29}{20} = 1.45$$

$$\therefore PR(5) = (1-0.9) + 0.9\left[\frac{PR(3)}{C(3)} + \frac{PR(4)}{C(4)}\right]$$

$$= 0.1 + 0.9\left[\frac{0.28}{3} + \frac{1.45}{2}\right]$$

$$= \frac{1673}{2000} = 0.8365$$

$$\therefore PR(6) = (1-0.9) + 0.9\left[\frac{PR(4)}{C(4)} + \frac{PR(5)}{C(5)}\right]$$

$$= 0.1 + 0.9\left[\frac{1.45}{2} + \frac{0.8365}{2}\right]$$

$$= 1.128925$$

- Iteration : 2 :-

$$PR(x) = (1-d) + d\left[\frac{PR(T_i)}{C(T_i)} + \cdots\cdots + \frac{PR(T_n)}{C(T_n)}\right]$$

$$\therefore PR(1) = (1-0.9) + 0.9\left[\frac{PR(3)}{C_i(3)}\right]$$

$$= 0.1 + 0.9\left[\frac{0.28}{3}\right]$$

$$= \frac{23}{125} = 0.184$$

$$\therefore PR(2) = (1-0.9) + 0.9\left[\frac{PR(1)}{C_i(1)} + \frac{PR(3)}{C_i(3)}\right]$$

$$= 0.1 + 0.9\left[\frac{0.184}{2} + \frac{0.28}{3}\right]$$

$$= \frac{667}{2500} = 0.2668$$

$$\therefore PR(3) = (1-0.9) + 0.9\left[\frac{PR(1)}{C_i(1)}\right]$$

$$= 0.1 + 0.9\left[\frac{0.184}{2}\right]$$

$$= \frac{457}{2500} = 0.1828$$

$$\therefore PR(4) = (1-0.9) + 0.9\left[\frac{PR(5)}{C(5)} + \frac{PR(6)}{C(6)}\right]$$

$$= 0.1 + 0.9\left[\frac{0.8365}{2} + \frac{1.128925}{1}\right]$$

$$\cancel{= 1.14}$$

$$= 1.4924575$$

$$\therefore \quad PR(5) = (1-0.9) + 0.9\left[\frac{PR(3)}{C(3)} + \frac{PR(4)}{C(4)}\right]$$

$$= 0.1 + 0.9\left[\frac{0.1828}{3} + \frac{1.49245 75}{2}\right]$$

$$= 0.826445875$$

$$\therefore \quad PR(6) = (1-0.9) + 0.9\left[\frac{PR(4)}{C(4)} + \frac{PR(5)}{C(5)}\right]$$

$$= 0.1 + 0.9\left[\frac{1.4924575}{2} + \frac{0.826445875}{2}\right]$$

$$= 1.143506519$$

- Table:

Iteration	PR(1)	PR(2)	PR(3)	PR(4)	PR(5)	PR(6)
0	1	1	1	1	1	1
1	0.4	0.58	0.28	1.45	0.8365	1.128925
2	0.184	0.2668	0.1828	1.4924575	0.82644-5875	1.1435-06519

2) From the TFDF values of the documents, construct **Naïve Bayes** model to classify the document "comedy, adventure, thriller, fiction" into Class A or Class B. [note: Use multinomial distribution]

TFIFD	Action	Comedy	Romantic	Adventure	thriller	Fiction	Class
D1	0	2	4	0	0	0	A
D2	3	1	0	4	2	0	B
D3	0	0	3	0	1	2	A
D4	3	0	4	3	2	0	B
D5	4	2	0	1	0	1	B
D6	**0**	**3**	**0**	**2**	**2**	**3**	**?**

Solution:

(2)

$$P(\text{comedy} \mid A) = \frac{2+1}{12+2} = 0.2143$$

$$P(\text{Adventure} \mid A) = \frac{0+1}{12+2} = 0.0714$$

$$P(\text{Thriller} \mid A) = \frac{1+1}{12+2} = 0.1429$$

$$P(\text{Fiction} \mid A) = \frac{2+1}{12+2} = 0.2143$$

$$P(\text{comedy} \mid B) = \frac{3+1}{30+2} = 0.125$$

$$P(\text{Adventure} \mid B) = \frac{8+1}{30+2} = 0.2812$$

$$P(\text{Thriller} \mid B) = \frac{4+1}{30+2} = 0.1562$$

$$P(\text{Fiction} \mid B) = \frac{1+1}{30+2} = 0.059.$$

$$\therefore P(D6 \mid A) = 10! \times (0.2143)^3 / 3! \times (0.0714)^2 / 2! \times (0.1429)^2 / 2! \times (0.2143)^3 / 3!$$

$$= 0.000254$$

$$\therefore P(D6 \mid B) = 10! \times (0.125)^3 / 3! \times (0.2812)^2 / 2! \times (0.1562)^2 / 2! \times (0.059)^3 / 3!$$

$$= 0.0000195$$

3) Create **Decision Tree** for the above TFIDF data value using Entropy(Information Gain) based Best attribute finding.

Term 1	Term 2	Term 3	Class
2.2	12.7	14.3	**Fiction**
2.7	11.9	21.6	**Drama**
3.4	11.7	13.1	**Fiction**
2.1	8.8	19.8	**Drama**
3.2	13.9	14.2	**Fiction**
3.9	13.7	11.1	**Drama**

Solution:

3.
- The 1st step towards constructing decision tree is to select the target node.
- Here I chose, class as target node.
- And, we now have to find Information Gain of target attribute.

Information gain is given by:

$$I.G. = -\frac{P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right).$$

Calculating information gain for class attribute:
Here, P is no. of fiction, and N is drama.

$$I.G. = -\left[\frac{3}{3+3}\left(\log_2\left(\frac{3}{6}\right)\right) + \frac{3}{3+3}\left(\log_2\left(\frac{3}{6}\right)\right)\right]$$

$$= -\left[\frac{3}{6}\log_2\left(\frac{3}{6}\right) + \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right]$$

$$= -\left[0.5 \times \log_2 2^{-1} + 0.5 \times \log_2 2^{-1}\right]$$

$$= -\left[0.5 \times (-\log_2 2) + 0.5\left(-1.\log_2 2\right)\right]$$

$$= -\left[-0.5 - 0.5\right]$$

$$= 1$$

 sharadindu.adhikari2019@vitstudent.ac.in

$\therefore$ $\boxed{I.G. = 1}$

Now, talking about other attribute, one attribute has to be root node.

- We gotta calculate gain of every attribute, and the attribute which has highest gain will be the root node.

Gain is given by, Gain = I.G. — Entropy of attribute.

$\Rightarrow$ Gain = I.G — E(A).

Entropy is given by, $E(A) = \sum_{i=1}^{v} \frac{P_i + N_i}{P+N} I(P_i N_i)$

NOW, For Term-1, the possible outcomes are :—

term-1	fiction	drama
≥ 3	1	2
< 3	2	1

$$I(\text{attribute} \geq 3) = -\left[\frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right) + \frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right)\right] \times \frac{3}{6}$$

$$= 0.13$$

$$I(\text{attribute} < 3) = -\left[\frac{2}{3} \log_2\left(\frac{2}{3}\right) + \frac{1}{3} \log_2\left(\frac{1}{3}\right)\right] \times \frac{3}{6}$$

$$= 0.13$$

So, Entropy of Term-1 $= 0.13 + 0.13$

$= 0.26$

Now, Gain $= IG - E(A) = 1 - 0.26$

$= 0.74$

$$\boxed{\therefore \text{ Gain of Term-1} = 0.74}$$

Similarly, for Term-2:

Term-2	fiction	drama
≥ 12	2	1
< 12	1	2

$$I(\text{attribute} \geq 12) = -\left[\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\cdot\log_2\left(\frac{1}{3}\right)\right] \times \frac{3}{6}$$

$$= 0.13$$

$$I(\text{attribute} < 12) = -\left[\frac{1}{3}\log_2\left(\frac{1}{3}\right) + \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right] \times \frac{3}{6}$$

$$= 0.13$$

So, Entropy of Term-2 $= 0.13 + 0.13$

$= 0.26$

Now, Gain of Term-2 = I.G. — E(A)

= 1 — 0.26

= 0.74

$\therefore$ $\boxed{\text{Gain of Term-2} = 0.74}$

For Term-3 :

Term-3	fiction	drama
≥ 15	0	2
< 15	3	1

$$I(\text{attribute} \geq 15) = -\left[\frac{0}{2}\log_2\left(\frac{0}{2}\right) + \frac{2}{2}\log_2\left(\frac{3}{2}\right)\right] \times \frac{2}{6}$$

$$= 0$$

$$I(\text{attribute} < 15) = -\left[\frac{3}{4}\log_2\left(\frac{3}{4}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right] \times \frac{4}{6}$$

$$= 0.16$$

So, Entropy of Term-3 = 0 + 0.16

= 0.16

$\therefore$ $\boxed{\text{Gain (term-3)} = 1 - 0.16 = 0.84}$.

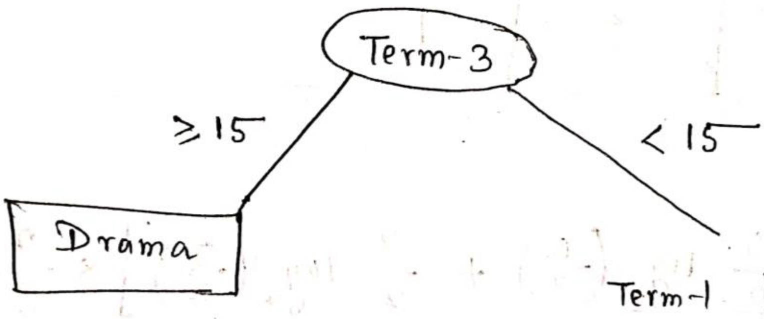
Therefore, Gain (Term-1) = 0.74

Gain (Term-2) = 0.74

Gain (Term-3) = 0.84

As I mentioned, Term-3 has the largest Gain, so it will be the root node.

Now, Decision Tree looks like:



Term-1	Term-2	class
2.2	12.7	Fiction
3.4	11.7	Fiction
3.2	13.9	Fiction
3.9	13.7	Drama

Now, finding I.G. for new table: For Term-1

Term-1	fiction	drama
≥ 3	2	1
< 3	1	0

$$I(\text{attribute} \geq 3) = -\left[\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right] \times \frac{3}{4}$$

$$= 0.20$$

$$I(\text{attribute} < 3) = -\left[\frac{1}{1}\log_2\left(\frac{1}{1}\right) + \frac{0}{1}\log_2\left(\frac{0}{1}\right)\right] \times \frac{1}{4}$$

$$= 0$$

$\therefore$ Entropy $= 0.20 + 0 = 0.20$

Gain (Term-1) $= 1 - 0.20 = 0.80$

- For Term-2:

Term-2	fiction	drama
≥ 12	2	1
< 12	1	0

$$I(\text{attribute} \geq 12) = -\left[\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right] \times \frac{3}{4}$$

$$= 0.20$$

$$I(\text{attribute} < 12) = -\left[\frac{1}{1}\log_2\left(\frac{1}{1}\right) + \frac{0}{1}\log_2\left(\frac{0}{1}\right)\right] \times \frac{1}{4}$$

$$= 0$$

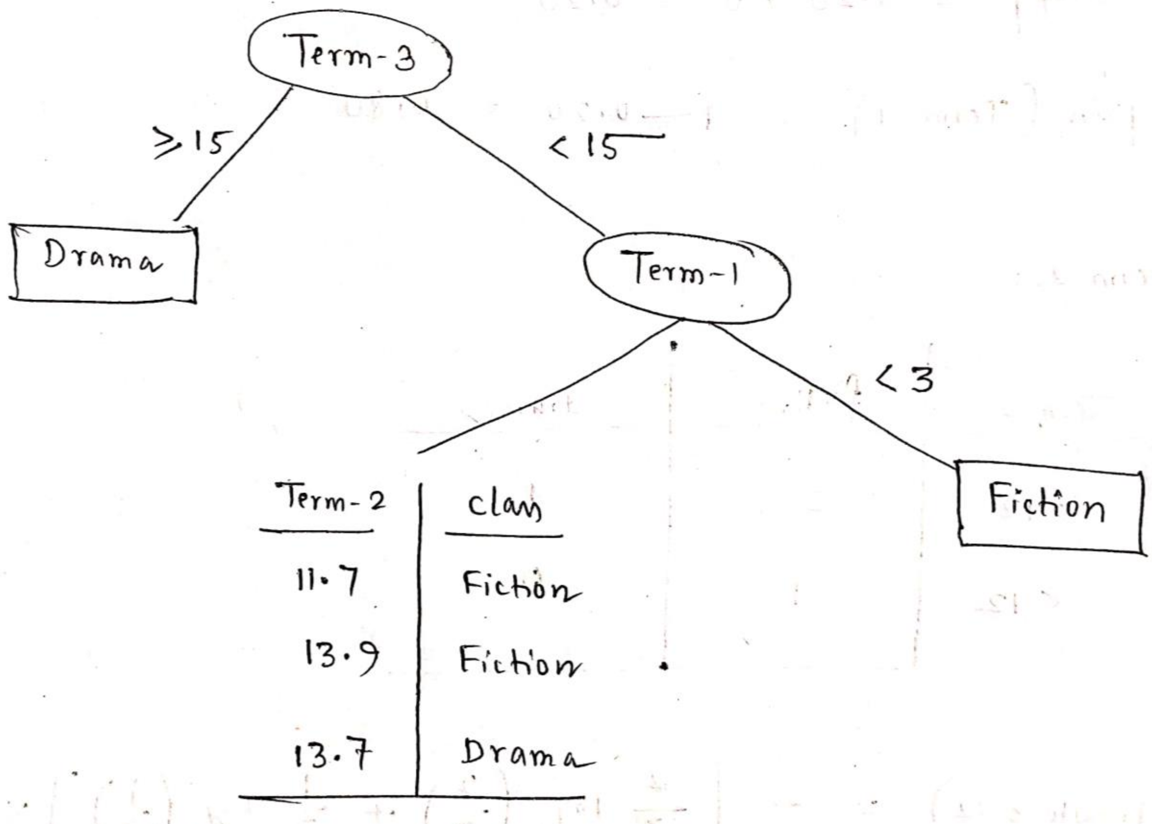
$\therefore$ Entropy $= 0.20 + 0 = 0.20$

$\therefore$ Gain (Term-2) $= 1 - 0.20 = 0.80$

Now, As gains of Term-1 and Term-2 are same, we can choose any attribute as root node.

Here, I'm choosing Term-1 to be the next root node.

Hence, the decision tree now looks like:

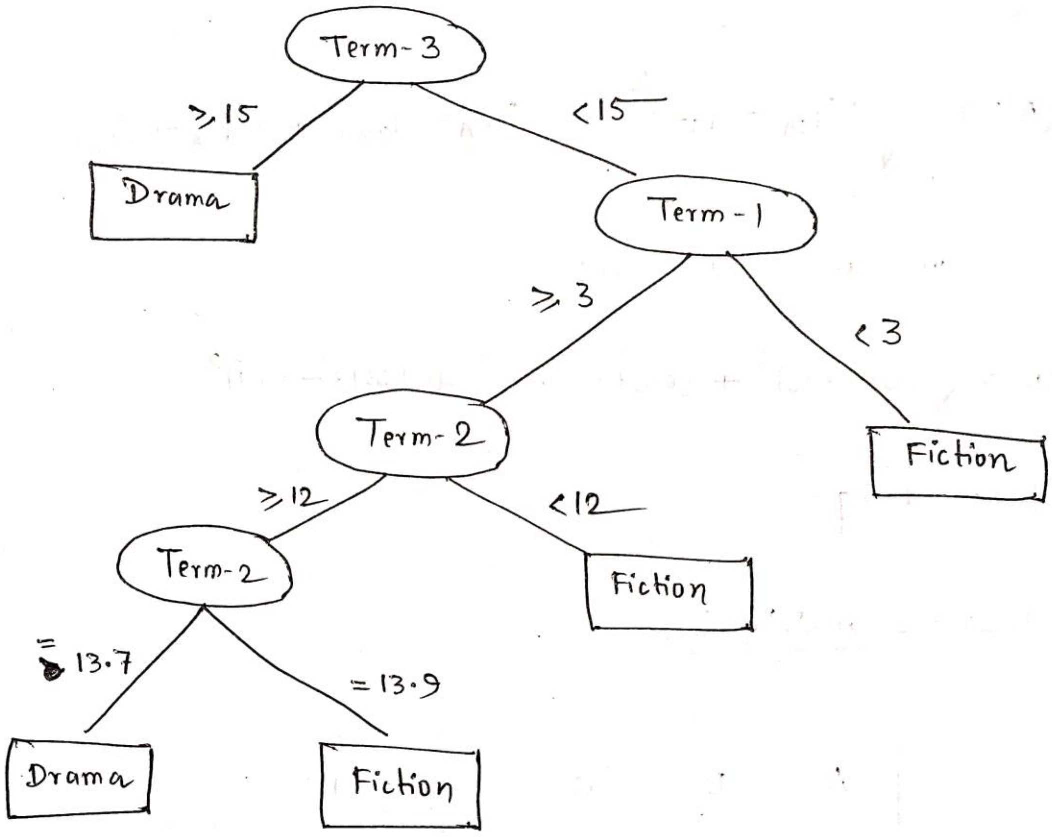


For Term-2,

if attribute < 12, it will be fiction.

and if > 12, then fiction & drama.

Final decision Tree :



4) Apply the hierarchical clustering on the following TF-IDF matrix and draw the dendrogram. A,B,C and D are the documents. T1,T2 and T3 are the terms. Using Single linkage method.

	T1	T2	T3
A	0	0.67	0.18
B	0.96	0.20	0.19
C	0	0.11	0.11
D	0	0.15	0.16

Solution:

(4).

At first, we will find the distance matrix for the given matrix using Euclidean distance.

$$d(A,B) = \sqrt{(T_{1A} - T_{1B})^2 + (T_{2A} - T_{2B})^2 + (T_{3A} - T_{3B})^2}$$

$\therefore$ Distance from doc A to doc B :-

$$d = \sqrt{(0-0.96)^2 + (0.67-0.20)^2 + (0.18-0.19)^2}$$

$\therefore\ d = 1.07$

~~Dis~~ Distance matrix is :

	A	B	C	D
A	0			
B	1.069	0		
C	0.564	0.967	0	
D	0.52	0.962	0.064	0

Now, we need to find the minimum distance among these & cluster these vertices.

Minimum value is 0.064 and we merge (C, D).

Now, we will again re-calculate distances after clustering

	A	B	C, D
A	0		
B	1.069	0	
C, D	0.52	0.962	0

Distance $(C, D) \rightarrow A = \min[\text{distance}_{CA}, \text{distance}_{DA}]$

$= \min[0.57, 0.52]$

Distance $(C, D) \rightarrow B = \min[\text{distance}_{CB}, \text{distance}_{DB}]$.

$= \min(0.97, 0.96)$

Minimum value is 0.52 and we merge (A, C, D).

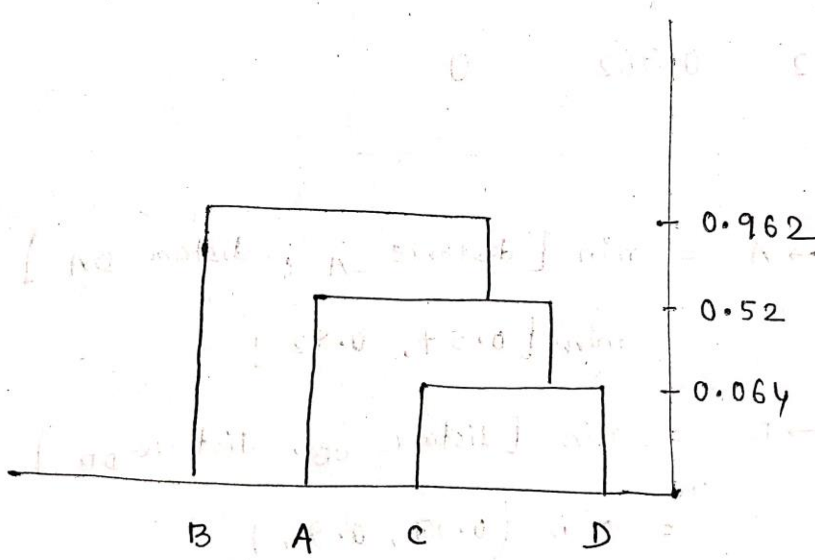
	A, C, D	B
A, C, D	0	
B	0.962	0

$$\text{Distance}\ (A,C,D \rightarrow B) = \min\left(\text{Distance}_{CB},\ \text{Distance}_{DB},\ \text{Distance}_{AB}\right)$$

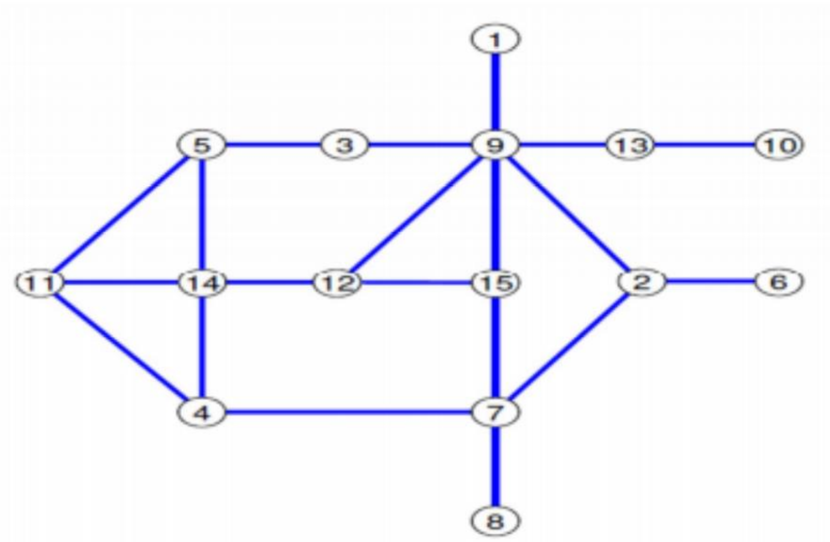
$$= \min\left(0.97,\ 0.96,\ 1.07\right)$$

Now, we'll merge (A,C,D) & B at 0.96

So the Dendogram of the given matrix is :—

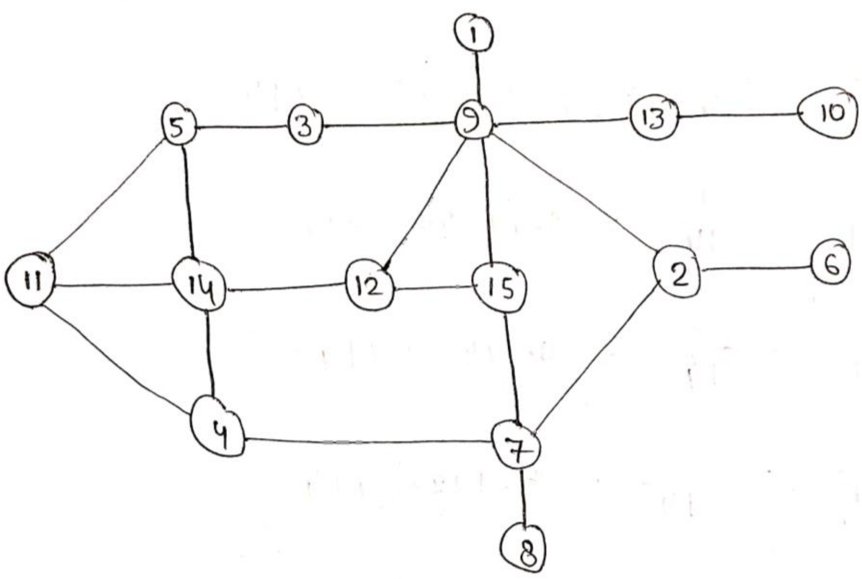


5) a) Calculate the **different types of centrality** values for each node for the following network diagram.



Solution:

5 (a) different types of centrality:



- Degree centrality:

$$C_D(i) = \frac{d(i)}{n-1} \quad \cdot \text{ Here, } n = 15$$

$$\therefore C_D(1) = \frac{1}{15-1} = \frac{1}{14} = 0.07142857143$$

$$C_D(2) = \frac{3}{15-1} = \frac{3}{14} = 0.2142857143$$

$$C_D(3) = \frac{2}{15-1} = \frac{2}{14} = \frac{1}{7} = 0.1428571429$$

$$C_D(4) = \frac{3}{15-1} = \frac{3}{14} = 0.2142857143$$

$$C_D(5) = \frac{3}{15-1} = \frac{3}{14} = 0.2142857143$$

$$C_D(6) = \frac{1}{15-1} = \frac{1}{14} = 0.07142857143$$

$$C_D(7) = \frac{4}{15-1} = \frac{4}{14} = \frac{2}{7} = 0.2857142857$$

$$C_D(8) = \frac{1}{15-1} = \frac{1}{14} = 0.07142857143$$

$$C_D(9) = \frac{6}{15-1} = \frac{6}{14} = \frac{3}{7} = 0.4285714286$$

$$C_D(10) = \frac{1}{15-1} = \frac{1}{14} = 0.07142857143$$

$$C_D(11) = \frac{3}{15-1} = \frac{3}{14} = 0.2142857143$$

$$C_D(12) = \frac{3}{15-1} = \frac{3}{14} = 0.2142857143$$

$$C_D(13) = \frac{2}{15-1} = \frac{2}{14} = \frac{1}{7} = 0.1428571429$$

$$C_D(14) = \frac{4}{15-1} = \frac{4}{14} = \frac{2}{7} = 0.2857142857$$

$$C_D(15) = \frac{3}{15-1} = \frac{3}{14} = 0.2142857143$$

- Closeness centrality [$x(j) = d(i,j)$, here]

i	$x(1)$	$x(2)$	$x(3)$	$x(4)$	$x(5)$	$x(6)$	$x(7)$	$x(8)$	$x(9)$	$x(10)$	$x(11)$	$x(12)$	$x(13)$	$x(14)$	$x(15)$	$\sum_{j=1}^{n} x(j)$
1	–	2	2	4	3	3	3	4	1	3	4	2	2	3	2	38
2	2	–	2	2	3	1	1	2	1	3	3	2	2	3	2	~~30~~ 29
3	2	2	–	3	1	3	3	4	1	3	2	2	2	2	2	32
4	4	2	3	–	2	3	1	2	3	5	1	2	4	1	2	35
5	3	3	1	2	–	4	3	4	2	4	1	2	3	1	3	36
6	3	1	3	3	4	–	2	3	2	4	4	3	3	4	3	42
7	3	1	3	1	3	2	–	1	2	4	2	2	3	2	1	30
8	4	2	4	2	4	3	1	–	3	5	3	3	4	3	2	43
9	1	1	1	3	2	2	2	3	–	2	3	1	1	2	1	25
10	3	3	3	5	4	4	4	5	2	–	5	3	1	4	3	49
11	4	3	2	1	1	4	2	3	3	5	–	2	4	1	3	38
12	2	2	2	2	2	3	2	3	1	3	2	–	2	1	1	~~29~~ 28
13	2	2	2	4	3	3	3	4	1	1	4	2	–	3	2	36
14	3	3	2	1	1	4	2	3	2	4	1	1	3	–	2	~~33~~ 32
15	2	2	2	2	3	3	1	2	1	3	3	1	2	2	–	29

- closeness centrality :

$$C_c(i) = \frac{n-1}{\sum_{j=1}^{n} d(i, j)}$$

- just like earlier, here as well, $n = 15$

$$\therefore \; C_c(1) = \frac{15-1}{2+2+4+3+3+3+4+1+3+4+2+2+3+2} = \frac{14}{38} = \frac{7}{19} = 0.3684210526$$

$$\therefore \; C_c(2) = \frac{15-1}{2+2+2+3+1+1+2+1+3+3+2+2+3+2} = \frac{14}{29} = 0.4827586207$$

$$C_c(3) = \frac{15-1}{2+2+3+1+3+3+4+1+3+2+2+2+2+2} = \frac{14}{32} = \frac{7}{16} = 0.4375$$

$$C_c(4) = \frac{15-1}{4+2+3+2+3+1+2+3+5+1+2+4+1+2} = \frac{14}{35} = \frac{2}{5} = 0.4$$

$$C_c(5) = \frac{15-1}{3+3+1+2+4+3+4+2+4+1+2+3+1+3} = \frac{14}{36} = \frac{7}{18} = 0.3889$$

$$C_c(6) = \frac{15-1}{3+1+3+3+4+2+3+2+4+4+3+3+4+3} = \frac{14}{42} = \frac{1}{3} = 0.3333$$

$$C_c(7) = \frac{15-1}{3+1+3+1+3+2+1+2+4+2+2+3+2+1} = \frac{14}{30} = \frac{7}{15} = 0.4667$$

$$C_c(8) = \frac{15-1}{4+2+4+2+4+3+1+3+5+3+3+4+3+2} = \frac{14}{43} = 0.3255813953$$

$$C_c(9) = \frac{15-1}{1+1+1+3+2+2+2+3+2+3+1+1+2+1} = \frac{14}{25} = 0.56$$

$$C_c(10) = \frac{15-1}{3+3+3+5+4+4+4+5+2+5+3+1+4+3} = \frac{14}{49} = \frac{2}{7} = 0.2857142857$$

$$C_c(11) = \frac{15-1}{4+3+2+1+1+4+2+3+3+5+2+4+\cancel{2}1+3} = \frac{14}{38} = \frac{7}{19} = 0.3684210526$$

$$C_c(12) = \frac{15-1}{2+2+2+2+2+3+2+3+1+3+2+2+1+1} = \frac{14}{28} = \frac{1}{2} = 0.5$$

$$C_c(13) = \frac{15-1}{2+2+2+4+3+3+3+4+1+1+4+2+3+2} = \frac{14}{36} = \frac{7}{18} = 0.3889$$

$$C_c(14) = \frac{15-1}{3+3+2+1+\cancel{2}1+4+2+3+2+4+1+1+3+2} = \frac{14}{32} = \frac{7}{16} = 0.4375$$

$$C_c(15) = \frac{15-1}{2+2+2+2+3+3+1+2+1+3+3+1+2+2} = \frac{14}{29} = 0.4827586207$$

- Betweenness Centrality :

No. of nodes in the given graph $= 15 = n$.

No. of paths possible $= {}^{n}C_{2} = \frac{n(n-1)}{2}$

$= \frac{15(15-1)}{2}$

$= 105$

i.e., 105 possible combinations (paths) are feasible here.

Hence, it is practically (almost) impossible to find betweenness of all the nodes by manual calculation.

However, instead of skipping this portion entirely, I'm going to go ahead and calculate betweennes centrality of 1 node.

I'm choosing Node 14 here.

we've:

$$\text{b.c.} = \frac{\text{shortest path (v)}}{\text{shortest path}} = \frac{st(v)}{st}.$$

. Here, V means vertex

and $st(v) \Rightarrow$ shortest path through vertex v.

Paths for node 14	st	st(v)	st(v)/st
(1,2)	1	0	0
(1,3)	1	0	0
(1,4)	1	1	1
(1,5)	1	0	0
(1,6)	1	0	0
(1,7)	1	0	0
(1,8)	1	0	0
(1,9)	1	0	0
(1,10)	1	0	0
(1,11)	1	1	1
(1,12)	1	0	0
(1,13)	1	0	0
(1,15)	1	0	0
(2,3)	1	0	0
(2,4)	1	0	0
(2,5)	1	0	0
(2,6)	1	0	0
(2,7)	1	0	0

PATHS for node (14)	st	st(v)	st(v) /st
(2, 8)	1	0	0
(2, 9)	1	0	0
(2, 10)	1	0	0
(2, 11)	1	1	1
(2, 12)	1	0	0
(2, 13)	1	0	0
(2, 15)	1	0	0
(3, 4)	1	1	1
(3, 5)	1	0	0
(3, 6)	1	0	0
(3, 7)	1	0	0
(3, 8)	1	0	0
(3, 9)	1	0	0
(3, 10)	1	0	0
(3, 11)	1	0	0
(3, 12)	1	0	0
(3, 13)	1	0	0
(3, 15)	1	0	0
(4, 5)	1	1	1
(4, 6)	1	0	0

PATH for node (14)	st	st(v)	st(v)/st
(4, 7)	1	0	0
(4, 8)	1	0	0
(4, 9)	1	1	1
(4, 10)	1	1	1
(4, 11)	1	0	0
(4, 12)	1	1	1
(4, 13)	1	1	0
(4, 15)	1	1	0
(5, 6)	1	0	0
(5, 7)	1	1	1
(5, 8)	1	1	1
(5, 9)	1	0	0
(5, 10)	1	0	0
(5, 11)	1	0	0
(5, 12)	1	1	1
(5, 13)	1	0	0
(5, 15)	1	1	1
(6, 7)	1	0	0
(6, 8)	1	0	0
(6, 9)	1	0	0

PATH for node (14)	st	st(v)	st(v) / st
(6, 10)	1	0	0
(6, 11)	1	0	0
(6, 12)	1	0	0
(6, 13)	1	0	0
(6, 15)	1	0	0
(7, 8)	1	0	0
(7, 9)	1	0	0
(7, 10)	1	0	0
(7, 11)	1	0	0
(7, 12)	1	0	0
(7, 13)	1	0	0
(7, 15)	1	0	0
(8, 9)	1	0	0
(8, 10)	1	0	0
(8, 11)	1	0	0
(8, 12)	1	0	0
(8, 13)	1	0	0
(8, 15)	1	0	0
(9, 10)	1	0	0
(9, 11)	1	1	1
(9, 12)	1	0	0
(9, 13)	1	0	0

Paths for node 14	st	st(v)	st(v)/st
(9, 15)	1	0	0
(10, 11)	1	1	1
(10, 12)	1	0	0
(10, 13)	1	0	0
(10, 15)	1	0	0
(11, 12)	1	1	1
(11, 13)	1	1	1
(11, 15)	1	1	1
(12, 13)	1	0	0
(12, 15)	1	0	0
(13, 15)	1	0	0

Here, for all nodes, have shortest path to all nodes. Because this is an undirected graph, so, for all paths, st. = 1.

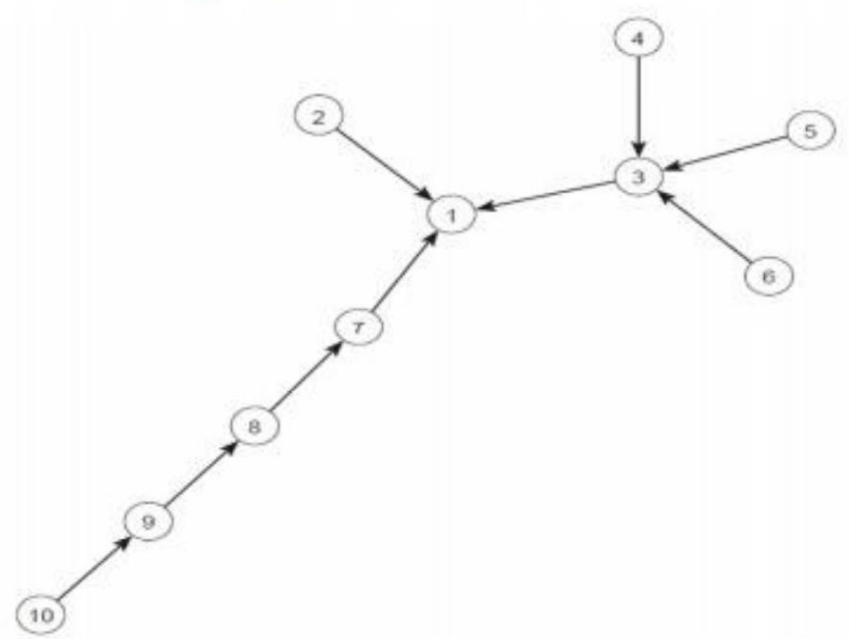
Shortest path have through vertex, then st(v) will be 1, otherwise st(v) will be 0.

So, we calculate how many ones we got. That is betweenness centrality of that node.

$\therefore$ Betweenness centrality of node 14 = 16

B.C.(14) = 16.

b) Calculate **the degree prestige and proximity prestige** using the following network diagram.



Solution:

5 (b)

I. Degree prestige:

$$P_D(i) = \frac{d_I(i)}{n-1}$$

, where $P_D(i)$ is degree prestige of actor i.

$d_I(i)$ is in-degree of i.

n is total no. of actors in the network.

So,

$$P_D(1) = \frac{3}{9} = 0.334$$

$$P_D(2) = \frac{0}{9} = 0$$

$$P_D(3) = \frac{3}{9} = 0.334$$

$$P_D(4) = \frac{0}{9} = 0$$

$$P_D(5) = \frac{0}{9} = 0$$

$$P_D(7) = \frac{1}{9} = 0.112$$

$$P_D(8) = \frac{1}{9} = 0.112$$

$$P_D(9) = \frac{1}{9} = 0.112$$

$$P_D(10) = \frac{0}{9} = 0$$

- **Proximity Prestige :**

$$P_p(i) = \frac{\frac{|I_i|}{n-1}}{\frac{\sum_{j \in I} d(j,i)}{|I_i|}}$$

where, $P_p(i)$ is Proximity prestige of actor i.

I_i is set of actors that can reach actor i.

$d(j,i)$ is minimum distance from actor j to actor i.

n is total no. of actors in network.

So,

For actor 1,

$$I_1 = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

$$P_p(1) = \frac{\frac{9}{9}}{\frac{1+1+2+2+2+1+2+3+4}{9}} = \frac{1}{2} = 0.5$$

For actor 2,

$$I_2 = \{\phi\}$$

$\because$ No actor can reach actor 2, $P_p(2) = 0$

- For actor 3,

$$I_3 = \{4, 5, 6\}$$

$$P_p(3) = \frac{\frac{3}{9}}{\frac{1+1+1}{3}} = 0.334$$

- For actor 4,

$$I_4 = \{\phi\}$$

$\because$ No actor can reach actor-4, $P_p(4) = 0$

- For actor 5,

$$I_5 = \{\phi\}$$

$\because$ No actor can reach actor-5, $P_p(5) = 0$

- For actor 6,

$$I_6 = \{\phi\}$$

$\because$ No actor can reach actor 6, $P_p(6) = 0$

- For actor 7,

$$I_7 = \{8, 9, 10\}$$

$$P_p(7) = \frac{3/9}{\frac{1+2+3}{3}} = 0.1667$$

- For actor 8,

$$I_8 = \{9, 10\}$$

$$P_p(8) = \frac{\frac{2}{9}}{\frac{1+2}{2}} = 0.148$$

- For actor 9,

$$I_9 = \{10\}$$

$$P_p(9) = \frac{\frac{1}{9}}{\frac{1}{1}} = 0.112$$

- For actor 10,

$$I_{10} = \{\phi\}$$

$\because$ No actor can reach actor 10, $P_p(10) = 0$