Kajol Tanesh Shah.
A20296724
CS579

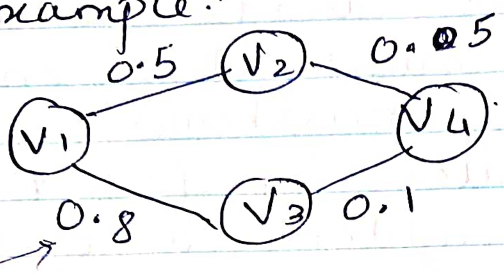
OSNA HW 3

Q1.

Ans. Independent Cascade model is a sender centric model of cascade where each active node (senders) has one chance to activate its neighbours.
If v is activated at time t
- for any neighbour w of v, there's a probability pvw that node w gets activated at time t+1. Here the node v activated at time t has a single chance of activating its neighbours which can happen only at t+1.
Also, activated nodes can activate only immediate neighbours. For the activation, random numbers generated in the range [0,1] are considered as threshold. If the activation probabilities of the nodes to be activated are greater than threshold, only then they will get activated.
The ICM can be considered as converged when it stops running. In ICM, in the activation process, the state of the nodes is changed from inactive to active but the reverse is not true. So, when there are no further nodes to be activated or no more connected nodes and ~~or their~~ no ~~activation~~ more neighbours which ~~have~~ can get activated (~~should~~ activation prob not greater than threshold), the algorithm stops running.

Even if the model converges, it may or may not have optimal solution.

Example:-



activation probabilities

randomly generated thresholds

$r_{1,2} = 0.4$

$r_{1,3} = 0.9$

$r_{2,4} = 0.1$

$r_{3,4} = 0.9$.

Start node $= V_1$

$V_1 \to V_2 = 0.5 > 0.4$ (threshold)

So, V_2 will get activated

$V_1 \to V_3 = 0.8 < 0.9$

V_3 will not get activated

$V_2 \to V_4 = 0.05 < 0.1$

So, V_4 will not get activated here and ICM will stop running as there are no more nodes left which can be activated. Here, we can still say that ICM has converged even though only 2 nodes were activated.

Q2. Community Analysis

$$TP = {}^3C_2 + {}^4C_2 + {}^3C_2 + {}^2C_2 + {}^2C_2 + {}^4C_2 + {}^2C_2$$
$$= 3 + 6 + 3 + 1 + 1 + 6 + 1$$
$$TP = 21$$

$$TN = (3\times5) + (3\times6) + (4\times4) + (4\times5) + (1\times5) + (1\times3) + (2\times6) + (3\times5) + (2\times3)$$
$$TN = 15 + 18 + 16 + 20 + 5 + 3 + 12 + 15 + 6$$
$$TN = 110$$

$$FN = (3\times2) + (3\times1) + (4\times3) + (4\times2) + (1\times2) + (1\times4) + (2\times1) + (2\times4) + (3\times2)$$
$$= 6 + 3 + 12 + 8 + 2 + 4 + 2 + 8 + 6$$
$$= 51$$
$$FN = 51$$

$$FP = (4\times3) + (4\times1) + (3\times1) + (2\times3) + (2\times2) + (3\times2) + (4\times2) + (4\times1) + (2\times1)$$
$$= 12 + 4 + 3 + 6 + 4 + 6 + 8 + 4 + 2$$
$$= 49.$$
$$FP = 49.$$

Precision

$$P = \frac{TP}{TP+FP} = \frac{21}{21+49} = 0.3$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{21}{21+51} = 0.29$$

$$\text{F-Measure} = 2 \cdot \frac{P \cdot R}{P+R} = \frac{2\times0.3\times0.29}{0.3+0.29} = 0.295$$

$$\text{Purity} = \frac{1}{N} \sum_{i=1}^{k} \max_j |C_i \cap L_j|$$

$$\cancel{= \frac{1}{14}(4+3+4)}$$

$$= \frac{1}{22}(4+3+4)$$

$$= \frac{11}{22} = \frac{1}{2} = 0.5$$

Normalized Mutual Information :

$\Delta = 2 \quad + = 1 \quad o = 3$

Found communities (H)

= [1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3]

Actual Labels (L)

= [1, 1, 1, 2, 2, 2, 2, 3, 2, 2, 2, 1, ~~1~~ 1, 3, 3, 1, 2, 2, 3, 3, 3, 3]

$n = 22$

	n_h
$h=1$	8
$h=2$	7
$h=3$	7

	n_L
$L=1$	6
$L=2$	9
$L=3$	7

$n_{h,L}$		
$h=1$		
$h=2$		
$h=3$		

$n_{h,L}$	$L=1$	$L=2$	$L=3$
$h=1$	3	4	1
$h=2$	2	3	2
$h=3$	1	2	4

$$MI = \sum_{h \in H} \sum_{l \in L} \frac{n_{h,l}}{n} \log \frac{n \cdot n_{h,l}}{n_h n_l}$$

$$= \left\{ \left[\frac{3}{22} \times \log \frac{22 \times 3}{8 \times 6} \right] + \left[\frac{4}{22} \times \log \frac{22 \times 4}{8 \times 9} \right] + \left[\frac{1}{22} \times \log \frac{22 \times 1}{8 \times 7} \right] \right\} + \left\{ \left[\frac{2}{22} \times \log \frac{22 \times 2}{7 \times 6} \right] + \left[\frac{3}{22} \times \log \frac{22 \times 3}{7 \times 9} \right] + \left[\frac{2}{22} \times \log \frac{22 \times 2}{7 \times 7} \right] \right\} + \left\{ \left[\frac{1}{22} \times \log \frac{22 \times 1}{7 \times 6} \right] + \left[\frac{2}{22} \times \log \frac{22 \times 2}{7 \times 9} \right] + \left[\frac{4}{22} \times \log \frac{22 \times 4}{7 \times 7} \right] \right\}$$

$$= \left\{ \left[\frac{3}{22} \log \frac{66}{48} \right] + \left[\frac{4}{22} \log \frac{88}{72} \right] + \left[\frac{1}{22} \log \frac{22}{56} \right] \right\}$$

$$+ \left\{ \left[\frac{2}{22} \log \frac{44}{56} \right] + \left[\frac{3}{22} \log \frac{66}{63} \right] + \left[\frac{2}{22} \log \frac{44}{49} \right] \right\}$$

$$+ \left\{ \left[\frac{1}{22} \times \log \frac{22}{42} \right] + \left[\frac{2}{22} \log \frac{44}{63} \right] + \left[\frac{4}{22} \log \frac{88}{49} \right] \right\}$$

$$= \left\{ [0.1363 \times 0.4594] + [0.1818 \times 0.2895] + [0.0454 \times (-1.348)] \right\} + \left\{ [0.0909 \times (-0.3479)] + [0.1363 \times 0.0671] + [0.0909 \times (-0.1553)] \right\} + \left\{ [0.0454 \times (-0.9329)] + [0.0909 \times (-0.5178)] + [0.1818 \times 0.8447] \right\}$$

$$= \left\{ 0.0626 + 0.0526 - 0.0612 \right\} + \left\{ \cancel{0.031} - 0.0316 + 0.0091 - 0.0470 \right\} + \left\{ -0.0423 - 0.0470 + 0.1536 \right\}$$

$$\cancel{= 0.0488} \quad = 0.1191$$

~~MI = 0.0488~~

$$H(L) = -\sum_{l \in L} \frac{n_l}{n} \log \frac{n_l}{n}$$

$$= -\left\{ \frac{6}{22} \log \frac{6}{22} + \frac{9}{22} \log \frac{9}{22} + \frac{7}{22} \log \frac{7}{22} \right\}$$

$$= -\left\{ 0.2727 \times (-1.8745) + 0.409 \times (-1.2895) + 0.3181 \times (-1.652) \right\}$$

$$= -\left\{ -0.5111 - 0.527 - 0.5255 \right\}$$

$$= 1.5636$$

$$\therefore H(L) = 1.5636$$

$$H(H) = -\sum_{h \in H} \frac{n_h}{n} \log \frac{n_h}{n}$$

$$= -\left\{ \frac{8}{22} \log \frac{8}{22} + \frac{7}{22} \log \frac{7}{22} + \frac{7}{22} \log \frac{7}{22} \right\}.$$

$$= -\left\{ 0.3636 \times (-1.4594) - 0.5255 - 0.5255 \right\}$$

$$= -\left\{ -0.5306 - 0.5255 - 0.5255 \right\}$$

$$= 1.5816$$

$$H(H) = 1.5816.$$

Normalized Mutual Information

$$NMI = \frac{MI}{\sqrt{H(L) . H(H)}}$$

$$= \cancel{\frac{0.0488}{\sqrt{1.5636 \times 1.5816}}} \quad \frac{0.1191}{\sqrt{1.5636 \times 1.5816}}$$

$$NMI = \cancel{0.0310} = 0.06$$

Q3. Q = edges inside community − expected no. of edges inside community

Ans. $$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{d_i d_j}{2m} \right] \delta(c_i, c_j)$$

The range for modularity Q values is $[-1, 1]$ i.e. from −1 to 1. The optimal modularity is 1. Modularity is positive if the number of edges inside the group are more than the expected number. Variation from 0 indicate difference with random case. The positive value of modularity indicates the possible presence of community structure. The trivial partition where all the vertices are in the same community leads to a modularity score of 0. Modularity is maximum when all vertices of the same type are connected to one another.

Q = Assortativity − Expected assortativity

or

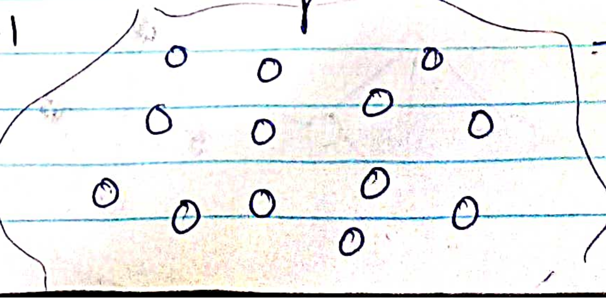
Q = edges inside community − expected no. of edges inside community

So, suppose if expected no. of edges is maximum possible (100% of connections expected) ~~then~~ and actual no. of edges is 0, then the value of Q will be −1

i.e. $Q = -1$

Example:-

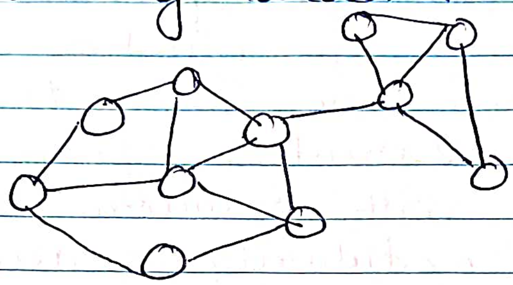
No connection between similar types.



→ In this graph, there are many nodes present but no edges between them.

The partition where all the vertices are in the same Community leads to modularity 0 i.e. $Q = 0$.
OR when a partition gives no-more within community edges than would be expected by random.

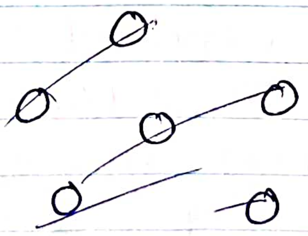
Example

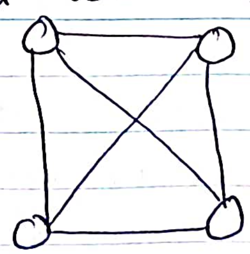


A single community where expected edges is equal to actual edges leads to $Q = 0$.

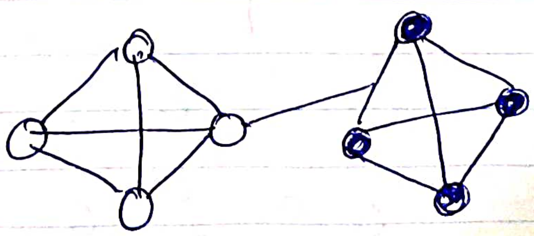
When the graph considered for the community is fully connected type and the actual no. of edges is maximum whereas expected number of edges is 0 or minimum, in this case, we will get $Q = 1$.

Example :- Fully connected with expected edges $= 0$.





or



2 different communities

Q 4.
Ans. a)

$$\text{Cosine similarity} = \frac{U_i \cdot U_j}{\|U_i\| \|U_j\|}$$

$$sim(U_i, U_j) = \frac{\sum_k r_{i,k} r_{j,k}}{\sqrt{\sum_k r_{i,k}^2} \sqrt{\sum_k r_{j,k}^2}}$$

Aristotle's rating for Le Cercle Rouge = ?

$$Sim(\text{Aristotle, Newton}) = \frac{3\times3 + 4\times3 + 2\times2 + 4\times2}{\sqrt{3^2+4^2+2^2+2^2}\sqrt{3^2+3^2+2^2+4^2}}$$

$$= \frac{33}{\sqrt{33}\sqrt{38}}$$

$$= \frac{\sqrt{33}}{\sqrt{38}} = 0.93$$

$$Sim(\text{Aristotle, Einstein}) = \frac{3\times5 + 0\times4 + 2\times2 + 3\times2}{\sqrt{33}\sqrt{5^2+0^2+2^2+3^2}}$$

$$= \frac{25}{\sqrt{33}\sqrt{38}}$$

$$= 0.71$$

$$Sim(\text{Aristotle, Gauss}) = \frac{3\times1 + 4\times4 + 3\times2 + 1\times2}{\sqrt{33}\sqrt{1^2+4^2+3^2+1^2}}$$

$$= \frac{27}{\sqrt{33}\sqrt{27}}$$

$$= \frac{\sqrt{27}}{\sqrt{33}}$$

$$= 0.90$$

$$\text{Sim}(\text{Aristotle, Euclid}) = \frac{3\times2 + 4\times0 + 2\times1 + 2\times5}{\sqrt{33}\sqrt{2^2+0^2+1^2+5^2}}$$

$$= \frac{18}{\sqrt{33}\sqrt{30}}$$

$$= 0.57$$

According to this, the two nearest neighbours are Newton & Gauss

$$\bar{r}_{aristotle} = 2.75$$

$$\bar{r}_{Newton} = \frac{3+0+3+2+4}{5} = \frac{12}{5} = 2.4$$

$$\bar{r}_{Gauss} = \frac{1+2+4+3+1}{5} = \frac{11}{5} = 2.2$$

$$\therefore r_{\substack{aristotle \\ -\text{le cercle rouge}}} = \bar{r}_{aristotle} + \frac{\text{Sim(Aris, Newton)}\,(r_{\substack{Newton \\ -\text{le cercle rouge}}} - \bar{r}_{Newton})}{\text{sim(Aris, Newton)} + \text{sim(Aris, Gauss)}}$$

$$+ \frac{\text{sim (Aris, Gauss)}\,(r_{\substack{Gauss \\ -\text{le cercle}}} - \bar{r}_{gauss})}{\text{sim (Aris, Newton)} + \text{sim(Aris, Gauss)}}$$

$$= 2.75 + \frac{0.93\times(0 - 2.4)}{0.93+0.90} + \frac{0.90\times(2-2.2)}{0.93+0.90}$$

$$= 2.75 + \frac{(-2.232)}{1.83} + \frac{(-0.18)}{1.83}$$

$$= 2.75 - 1.22 - 0.098$$

$$= 1.432$$

$$\therefore r_{\substack{Aristotle \\ -\text{le cercle rouge}}} = 1.432$$

Q4. b)
Ans. $G = \{$Newton, Einstein, Gauss$\}$.

- Average satisfactory $R_i = \frac{1}{n} \sum_{u \in G} r_{u,i}$

$r_{God} = \frac{3+5+1}{3} = \frac{9}{3} = 3$.

$r_{Le\ Cercle\ Rouge} = \frac{0+4+2}{3} = \frac{6}{3} = 2$

$r_{Cidade\ de\ Deu} = \frac{3+0+4}{3} = \frac{7}{3} = 2.33$

$r_{Rashomon} = \frac{2+2+3}{3} = \frac{7}{3} = 2.33$

$r_{La\ vita\ e\ bella} = \frac{4+3+1}{3} = \frac{8}{3} = 2.66$

Using this strategy, the product God is recommended

- Least misery

$R_i = \min_{u \in G} r_{u_i}$

$r_{God} = 1$

$r_{Le\ cercle\ Rouge} = 0$

~~r_{Gauss} =~~

$r_{Cidade\ de\ Deu} = 0$

$r_{Rashomon} = 2$

$r_{La\ vita\ e\ bella} = 1$

Using this strategy, the product Rashomon is recommended.

Most pleasure

$$R_i = \max_{u \in G} r_{u,i}$$

$r_{God} = 5$

$r_{Le\ Cercle\ Rouge} = 4$

$r_{Cidade\ de\ Deu} = 4$

$r_{Rashomon} = 3$

$r_{La\ vita\ e\ bella} = 4$

Using this strategy, the first product to be recommended is God.