Assignment-3

1. [Information Diffusion] Does Independent Cascade Model (ICM) converge? Why? When the ICM stops running, the algorithm has converged? Please justify your answer with details.

Answer:

Yes, it will take at most of n steps processes to converge. In here we have two types of nodes activated and not activated node.

In the ICM, nodes that become activated in one time step can influence their inactive neighbors in the next time step. The nodes which are not activated will not be removed, it may be activated in the future time steps. This set of nodes not activated can be at most the total no of nodes n. This whole process can take at most n number of steps before converges.

2. [Community Analysis] Compute the following metrics for the given figure:

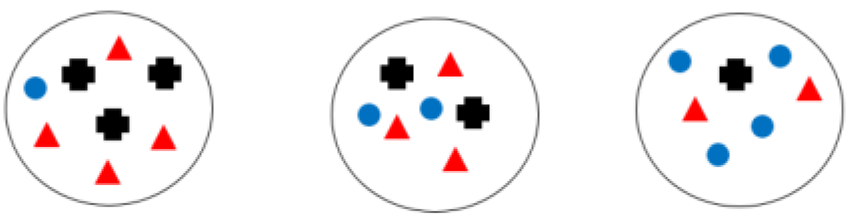

Figure 1: The communities.

Answer:

2)

Answer:

- Precision $= \dfrac{TP}{TP + FP}$

- Recall $= \dfrac{TP}{TP + FN}$

$$TP = \binom{3}{2} + \binom{4}{2} + \binom{3}{2} + \binom{3}{2} + \binom{3}{2} + \binom{4}{2} + \binom{2}{2}$$

$$= 21$$

$$FP = 3\times4 + 3\times1 + 4\times1 + 2\times3 + 2\times2 + 2\times3 + 4\times1 + 4\times2 + 2\times1 = 49$$

$$= 49$$

$$FN = 4\times3 + 4\times2 + 3\times2 + 3\times2 + 3\times1 + 2\times1 + 1\times4 + 2\times4$$

$$= 51.$$

Precision $= \dfrac{21}{21+49} = 0.3$

Recall $= \dfrac{21}{21+51} = 0.2917$

- F-measure

$$F = 2\,\frac{\text{Pression} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 0.2958.$$

$$F = 2\,\frac{0.3 \times 0.2917}{0.3 + 0.2917}$$

F-measure $= 0.29579.$

- NMI

$$NMI = \frac{\sum_{h,l} \log \frac{n \cdot n_{hl}}{n_h \cdot n_l}}{\sqrt{\left(\sum_h n_h \log \frac{n_h}{n}\right)\left(\sum_l \log \frac{n_l}{n}\right)}} = \frac{2.62}{\sqrt{(-34.80)(-34.41)}}$$

$NMI = 0.0758$

	$l=1$	$l=2$	$l=3$
$h=1$	4	3	1
$h=2$	3	2	2
$h=3$	2	1	4

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

$l=1$	9
$l=2$	6
$l=3$	7

- Purity

$$Purity = \frac{1}{N} \sum_{i=1}^{k} \max |C_i \cap C_j| = \frac{3+4+4}{22}$$

$$= \frac{11}{22}$$

$$= 0.5$$

3. **[Influence and Homophily]** What is the range $[\alpha_1, \alpha_2]$ for modularity Q values? Provide examples for both extreme values of the range, as well as cases where Modularity becomes zero. Modularity is defined as,

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

Answer:

3.

Answer:

$$Q = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{d_i d_j}{2m}] \delta(C_i, C_j).$$

The range is between $[-\frac{1}{2}, 1]$

The graph structure and node types is related with the maximum value of the modularity. When we have disjoint complete graphs the value will be close to 1. If we have bipartite graphs then it is $-\frac{1}{2}$.

In the above formula $\frac{d_i d_j}{2m}$ is the expected number of edjes between C_i and C_j.

In a Network nodes connected randomly without any significant pattern of connectivity the modularity Q value will be close to 0.

So, the Range of Q value for modularity will be $[-\frac{1}{2}, 1]$.

4. **[Recommendation]** Consider the user-item matrix in table 1 and answer the following questions.

	God	Le Cercle Rouge	Cidade de Deu	Rashomon	La vita e bella	$\overline{r_u}$
Newton	3	0	3	2	4	
Einstein	5	4	0	2	3	
Gauss	1	2	4	3	1	
Aristotle	3	?	4	2	2	2.75
Euclid	2	2	0	1	5	

Table 1: **User-Item Matrix**

(a) Compute the missing rating (Aristole-Le Cercle Rouge) in this table using user-based collaborative filtering (CF). Use cosine similarity to find the two nearest neighbors.

(b) Consider group G = {Newton, Einstein, Gauss}, compute the aggregated ratings for all products using average satisfactory, least misery, and most pleasure. What is the firs product recommended to the group using each strategy.

4.

Answer:

A: Ans)

Compute missing Rating (Aristole-Le Cercle Range)

Average rating for users:

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

$\bar{r}_{Einstin} = \frac{5+4+0+2+3}{5} = \frac{14}{5} = 2.8$

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

$\bar{r}_{Euclid} = \frac{2+2+0+1+5}{5} = \frac{10}{5} = 2.$

Now, we will compute Similarity between the Aristotle and others using Cosine Similarity.

$$Sim(aristotle, newton) = \frac{3\times3+4\times3+2\times2+2\times4}{\sqrt{33}\sqrt{38}}$$

$$= \frac{9+12+4+8}{\sqrt{33}\sqrt{38}}$$

$$Sim(Aristotle, newton) = 0.9318.$$

$$\text{sim}(\text{Aristotle}, \text{Einstein}) = \frac{3\times5+4\times0+2\times2+2\times3}{\sqrt{33}\ \sqrt{38}}$$

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

$$\text{Sim}(\text{Aristotle}, \text{einstein}) = 0.7059.$$

$$\text{Sim}(\text{Aristotle}, \text{gauss}) = \frac{3\times1+4\times4+2\times3+2\times1}{\sqrt{33}\cdot\sqrt{27}}$$

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

$$\text{sim}(\text{Aristotle}, \text{gauss}) = 0.91.$$

$$\text{sim}(\text{Aristotle}, \text{euclid}) = \frac{2\times3+0\times4+1\times2+5\times2}{\sqrt{33}\ \sqrt{30}}$$

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

$$\text{sim}(\text{Aristotle}, \text{euclid}) = 0.57.$$

The similarity of Network and Gauss is having highest among others, they are the closest neighbours to Aristotle.

Now, Aristotle's rating for Le Cenle Rouge computed from.

$$r(\text{Aristotle Le Cenle Rogue}) = \bar{r}_{\text{Aristotle}} + \frac{\text{sim}(\text{Aristotle}, \text{Netwton})(r_{(\text{newton}, \text{Lecenle Rouge})} - \bar{r}_{\text{newton}})}{\text{Sim}(\text{Aristotle}, \text{Netwon}) + \text{Sim}(\text{Aristotle}, \text{Gauss})}$$

$$+ \frac{\text{sim}(\text{Aristotle}, \text{Gauss})(r_{(\text{Gauss}, \text{Le Cenle Rouge})} - \bar{r}_{\text{Gauss}})}{\text{Sim}(\text{Aristotle}, \text{Netwon}) + \text{Sim}(\text{Aristotle}, \text{Gauss})}$$

S Gauss

$$= 2.75 + \frac{0.93(0-2.4)+0.91(2-2.2)}{0.93+0.91}$$

$$= 2.75 + \frac{(-2.232)+(-0.182)}{1.84}$$

$$= 2.75 - 1.31$$

$$= 1.44.$$

B ÷ Ans)

Group G ÷ {Newton, Einsten, Gauss}.

By using Average Satisfaction we will Compute the Aggregated ratings for all products.

The First recommended product based on Average satisfactory

$$r_i = \frac{1}{n} \sum_{u \in G} r_{u,i}$$

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

$$R_{le\ Cercle\ Rouge} = \frac{0+4+2}{3} = \frac{6}{3} = 2$$

$$R_{Cidadede\ Due} = \frac{3+0+4}{3} = 2.33$$

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

$$= \frac{7}{3} = 2.33.$$

$$= 2.33.$$

→ Least misery

$$R_{Brod} = \min\{3,5,1\} = 1$$

$$R_{Le\,cercle\,Rouge} = \min\{0,4,2\} = 0$$

$$R_{Cidade\,de\,Deu} = \min\{3,0,4\} = 0$$

$$R_{Reshomon} = \min\{2,2,3\} = 2$$

$$R_{Lavitaebella} = \min\{4,3,1\} = 1.$$

→ Most pleasure.

$$R_{Brod} = \max\{3,5,1\} = 5$$

$$R_{Le\,cercle\,Rage} = \max\{0,4,2\} = 4$$

$$R_{Cidade\,de\,Deu} = \max\{3,0,4\} = 4$$

$$R_{Rashomon} = \max\{2,2,3\} = 3$$

$$R_{Lavitaebella} = \max\{4,3,1\} = 4.$$