Name:	
IIT ID Number:	

CS 579: Online Social Network Analysis

Homework III - Information Diffusion, Community Analysis, Influence and Homophily, Recommendation

Prof. Kai Shu
Due at April 11 2022, 11:59 PM

This is an ***individual*** homework assignment. Please submit a digital copy of this homework to **Blackboard**. For your solutions, even when not explicitly asked you are supposed to concisely justify your answers.

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.

   Yes. At each time step, if no node gets activated, the algorithm has converged. If a node gets activated, one node from the set of nodes "not yet activated" is removed. The set of nodes not activated can be at most, the total number of nodes n. So, the whole process can take at most n steps, before it converges.

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

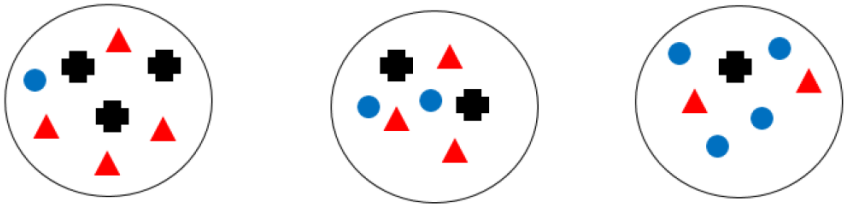

Figure 1: The communities.

- Precision and recall

  $Precision = \frac{TP}{TP+FP}$, $Recall = \frac{TP}{TP+FN}$

Name:	
IIT ID Number:	

$$TP = \binom{3}{2} + \binom{4}{2} + \binom{3}{2} + \binom{3}{2} + \binom{3}{2} + \binom{4}{2} + \binom{2}{2} = 21$$

$$FP = 3\times 4 + 3\times 1 + 4\times 1 + 2\times 3 + 2\times 2 + 2\times 3 + 4\times 1 + 4\times 2 + 2\times 1 = 49$$

$$FN = 4\times 3 + 4\times 2 + 3\times 2 + 3\times 2 + 3\times 1 + 2\times 1 + 1\times 2 + 1\times 4 + 2\times 4 = 51$$

$$\text{Precision } = \frac{21}{21+49} = 0.3$$

$$\text{Recall } = \frac{21}{21+51} = 0.2917$$

- F-measure
  $F = 2\frac{Precision\times Recall}{Precision+Recall} = 0.2958$
- NMI
  $NMI = \frac{\sum_{h,l} \log \frac{n\cdot n_{hl}}{n_h \cdot n_l}}{\sqrt{\left(\sum_h n_h \log \frac{h_h}{n}\right)\left(\sum_l \log \frac{n_l}{n}\right)}} = 0.0758$

	$1=1$	$1=2$	$1=3$
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

$$\text{Purity } = \frac{1}{N}\sum_{i=1}^{k} \max |C_i \bigcap C_j| = \frac{3+3+4}{22} = \frac{11}{22}$$

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,

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

The maximum value of modularity depends on the graph structure and node types and it is provided in the slides (Qmax). The modularity values over all possible networks is in the range $[-\frac{1}{2}, 1)$. -1/2 happens for complete bipartite graphs (see Example 8.1 in the book). Modularity gets arbitrary close to 1 when there are many disjoint complete graphs ($m \longrightarrow \infty$), each of which contain only nodes of the same type, e.g., complete graph i contains nodes of type i. Zero happens in many cases. For instance when we have (1) a random graph $\left(A_{ij} = \frac{k_i k_j}{2m} \text{ for } \delta(c_i, c_j) = 1\right)$ or (2) a complete graph of nodes of the same type with a large value for $n \longrightarrow \infty$ There are other graphs as well that have zero modularity.

Name:	
IIT ID Number:	

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	1.5
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.

Average rating for users are as follows:
$\bar{r}_{\text{Newton}} = \frac{12}{5} = 2.4 \quad \bar{r}_{\text{Einstein}} = \frac{14}{5} = 2.8 \quad \bar{r}_{\text{Gauss}} = \frac{11}{5} = 2.2 \quad \bar{r}_{\text{Euclid}} = \frac{10}{5} = 2$
Using Cosine similarity the similarity between Aristotle and others can be computed as:
Sim(Aristotle, Newton) $= \frac{3+3+2\times2+4\times2}{\sqrt{33}\sqrt{38}} = 0.93$
Sim(Aristotle, Einstein) $= \frac{5+0+2\times2+3\times2}{\sqrt{33}\sqrt{38}} = 0.70$
Sim(Aristotle, Gauss) $= \frac{1+4+3\times2+1\times2}{\sqrt{33}\sqrt{27}} = 0.90$
Sim(Aristotle, Newton) $= \frac{2+0+1\times2+5\times2}{\sqrt{33}\sqrt{30}} = 0.57$

Newton and Gauss are the closest (most similar) neighbors to Aristotle. Then, Aristotle's rating for Le Cerle Rouge computed from

$$r(\text{ Aristotle LeCerleRouge}) = \bar{r}_{\text{Aristotle}} + \frac{\text{Sim(Aristotle,Newton)}\left(r_{(\text{Newton, LeCerle Rouge})} - \bar{r}_{Newton}\right)}{\text{Sim}(Arstitle,Newton)+Sim(Arstitotle,Gauss)}$$
$$+\frac{\frac{Sim(Aristotle,Gauss)\left(r_{(Gauss,LeCerleRouge)}-\bar{r}_{Gauss}\right)}{S_{Gauss}}}{\text{Sim}(Arstitle,Newton)+\text{Sim(Arstitotle, Gauss)}} =$$
$$2.75 + \frac{0.93\times(0-2.4)+0.9\times(2-2.2.)}{0.93+0.9} =$$
$$2.75 + \frac{-2.232-0.18}{1.83} = 2.75 - 1.32 = 1.43$$

(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.

First recommended product based on Average Satisfactory is God.
$r_{\text{God}} = \frac{3+5+1}{3} = 3 \quad r \text{ LeCerle Rouge} = \frac{0+4+2}{3} = 2$
$r_{\text{CidadedeDeu}} = \frac{3+0+4}{3} = \frac{7}{3} \quad r_{\text{ Rashomon}} = \frac{2+2+3}{3} = \frac{7}{3}$

First recommended product based on Least Misery is Rashomon.
$r_{\text{God}} = \min\{3,5,1\} = 1 \quad r_{\text{LeCerleRouge}} = \min\{0,4,2\} = 0$
$r_{\text{CidadedeDeu}} = \min\{3,0,4\} = 0 \quad r_{\text{ Rashomon}} = \min\{2,2,3\} = 2$
$r_{\text{ Lavitaebella}} = \min\{4,3,1\} = 1$

Name:	
IIT ID Number:	

First recommended product based on Most Pleasure is God.
$r_{\text{God}} = \max\{3, 5, 1\} = 5 \quad r_{\text{LeCerleRouge}} = \max\{0, 4, 2\} = 4$
$r_{\text{cidadedeDeu}} = \max\{3, 0, 4\} = 4 \quad r_{\text{ Rashomon}} = \max\{2, 2, 3\} = 3$
$r_{\text{ Lavitaebella}} = \max\{4, 3, 1\} = 4$