Name:	PRAVEENAA KULANDHAIVEL
ASU ID Number:	1225814683

CSE 472: Social Media Mining

Homework IV - Influence and Homophily; Recommender Systems

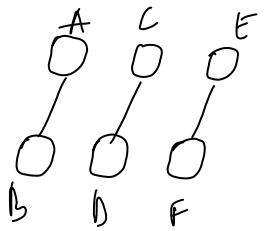
Prof. Huan Liu
Due at 2022 Nov 20, 11:59 PM

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

1. [**Influence and Homophily**] Provide the range of modularity value in theory. And give an example (i.e., adjacency matrix, indicator, degree matrix) for case where modularity is above 1/2. 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 max modularity occurs when all edges are connecting nodes of the same type.
The range of modularity ⇒ −1 to 1 ⇒ value is positive if the number of edges is more than the expected number & vice versa



	A	B	C	D	E	F
A	0	1	0	0	0	0
B	1	0	0	0	0	0
C	0	0	0	1	0	0
D	0	0	1	0	0	0
E	0	0	0	0	0	1
F	0	0	0	0	1	0

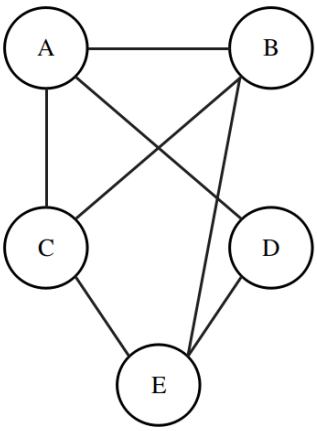
$Q_1 = \frac{1}{3} - \left[\frac{2}{6}\right]^2 = \frac{2}{9}$

$Q_2 = \frac{2}{9}$

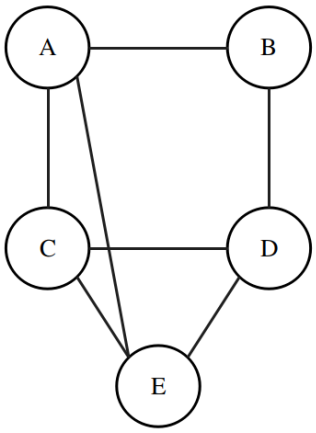
$Q_3 = \frac{2}{9}$

$Q = Q_1 + Q_2 + Q_3 = \frac{6}{9} = \frac{2}{3} > \frac{1}{2}$

2. **[Influence and Homophily]** Consider the graph given below where the graph on the left hand side shows the snapshot of the graph at $t = t_1$ and the graph on the right hand side shows the snapshot at $t = t_2$. Let x_i be the ordinal attribute value associated with node v_i such that v_i represents the i-th node in the graph. Compute the homophily index of the graph. Table 1 represents the ordinal values of each node.



(a) Snapshot at $t = t_1$



(b) Snapshot at $t = t_2$

Vertex	Values
A	12
B	16
C	21
D	11
E	22

Table 1: Vertex and their ordinal attribute values.

$H = \rho^{t_2} - \rho^{t_1}$

$\Rightarrow \dfrac{\frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{d_i d_j}{2m}\right)x_i x_j}{\frac{1}{2m}\sum_{ij} A_{ij} x_i^2 - \frac{1}{2m}\sum_{ij}\frac{d_i d_j}{2m} x_i x_j}$

① $\Rightarrow \sigma(X_L, X_R) = E[X_L X_R] - E[X_L]E[X_R] \approx 0.75$

$\sigma(X_L^2) \approx 0.2 \quad \rho^{t_1} \approx -0.04$

② $\Rightarrow \sigma(X_L, X_R) \approx -10$

$\sigma(X_L^2) = 20 \quad \rho^{t_2} \approx 0.45$

$H = 0.4$

3. **[Recommendation]** Compute the missing rating of these song artists using user-based collaborative filtering (CF) where the rows denote the users and the columns denote the song artists. When finding nearest neighbors, use Cosine similarity as your similarity measure (*Hint*: $sim(u, v) = \frac{u \cdot v}{||u|| \cdot ||v||}$).

	Sia	Bruno	Adele	Zayn	$\overline{r}_u$
Ron	3	5	5	4	4.25
Ginny	2	?	5	3	3.3
George	1	4	5	3	3.25
Fred	1	1	3	5	2.5

Predict the missing rating by completing the following tasks:

(a) Calculate the similarity value between Ginny and Ron (all others are provided in the following table).

	Ron	George	Fred
Ginny	0.99	0.99	0.88

Write down your calculation in the following box.

$Sim(u,v) = (u \cdot v) / (||u|| \cdot ||v||) \Rightarrow \frac{(2\times3+5\times5+4\times4)}{\sqrt{2^2+5^2+3^2}+\sqrt{3^2+5^2+4^2}} = \frac{6+25+16}{\sqrt{38}+\sqrt{50}} \approx 0.99$

(b) Identify Ginny's 2 nearest neighbors.

1. RON	2. GEORGE

(c) Calculate $\overline{r}_u$ values for everyone (Ginny's is given). Write down the final values in the table (last column).

(d) Calculate Ginny's rating for the artist "Bruno". *Hint*: $r_{u,i} = \overline{r}_u + \frac{\sum_{v \in N(u)} sim(u,v)(r_{v,i} - \overline{r}_v)}{\sum_{v \in N(u)} sim(u,v)}$. Assume that we consider 2 nearest neighbors for computing the rating.

$r_{ginny,Bruno} = \overline{r}_{ginny} + \frac{\sum_{v \in N(u)} sim(ginny,ron)(r_{ron,Bruno} - \overline{r}_{ron})}{\sum_{v \in N(u)} sim(ginny,ron)} + \frac{\sum_{v \in N(u)} sim(ginny,george)(r_{ginny,george} - \overline{r}_{george})}{\sum_{v \in N(u)} sim(ginny,george)}$

$= 3.3 + \frac{0.99 \times (5-4.25)}{0.99+0.99} + \frac{0.99 \times (4-3.25)}{0.99+0.99} = 3.3 + \frac{0.99 \times 0.75}{1.98} + \frac{0.99 \times 0.75}{1.98}$

$= 4.05$