Department of Computer Science

CS 429/529 – Dynamic and Social Network Analysis

Assignment 5

Alperen CAN

21601740

Instructor: Miray KAŞ

Teaching Assistant: Gözde YAZICI

Bilkent University

17.12.2022

Question 1

	Statement	True or False?	Reasoning
1	A path is a walk without passing through the same link more than once.	False	Also vertices cannot be repeated as well.
2	Graph and network refer to the same thing.	False	Although they used interchangeably in this course, actually graphs are the mathematical representations of the networks.
3	The degree of a node describes the sum of its in and out degrees.	True	
4	Google pagerank is a variant of Eigenvector centrality	True	
5	QAP analysis can be applied on any two networks	False	Networks should have the same size to apply QAP.
6	Hamming distance compares only binary data	True	
7	Link analysis is used in law enforcement	True	
8	Centrality metrics and rankings of nodes yield the same results regardless of whether the edge list is modeled as directed or undirected	True	
9	Sorensen similarity gives more weight to common elements than jaccard similarity.	True	
10	Bipartite networks are only possible on two-mode networks	True	
11	All graphs with fat tailed degree distributions are scale-free graphs.	False	Scale free networks degree distribution has a heavy tail.
12	Network density is calculated as n(n-1)/2 where n is the number of nodes in the network.	False	It is finding for potential connections. To find density, number of connections should be divided by it.
13	Preferential attachment model works better when the limits of the nodes are well defined	True	
14	A clique of 4 nodes contains 4 3-cliques	True	
15	The Girvan-Newman clustering algorithm is an agglomerative algorithm.	False	It is divisive.
16	Edge betweenness is a metric defined at the network level.	False	It is a node level metric.
17	CONCOR clustering algorithm is based on structural equivalence	True	

18	K-means requires a preset number of clusters	True	
19	In statistics, correlation implies causation	False	It does not.
20	99% confidence level is typically sought for in research	False	It should be around 95%.

Table 1: True/False Questions

<u>Question 2</u>

Below, X represents the first matrix whereas Y represent the second. I calculated the computations by hand in order to find the necessary expression required in Pearson Correlation Coefficient Formula.

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}}$$

- N = Sample size (the number of entries in the adjacency matrices)
- x_i, y_i = Individual entries
- $\overline{x}$ = Mean of entries (analogously for $\overline{y}$)

Figure 1: Correlation Coefficient Formula

First column in the table represents X and Y, expressed as X-Y (X,Y). I calculated the rest by a calculator.

	X	Y	X-X'	Y-Y'	(X-X') * (Y-Y')	$(X-X')^2$	$(Y-Y')^2$
1-1	0	0	-1.15				
1-2	1	4	-0.15				
1-3	1	0.3	-0.15				
1-4	0	0	-1.15				
1-5	1	0	-0.15				
1-6	0	0	-1.15				
1-7	3	1	1.85				
2-1	1	4	-0.15				
2-2	0	0	-1.15				
2-3	0.5	0	-0.65				
2-4	0.7	9	-0.45				
2-5	0	0.1	-1.15				
2-6	2	3	0.85				
2-7	0	0	-1.15				

3-1	1	0.3	-0.15				
3-2	0.5	0	-0.65				
3-3	0	0	-1.15				
3-4	1	0	-0.15				
3-5	0	6	-1.15				
3-6	8	0.3	6.85				
3-7	3	1	1.85				
4-1	0	0	-1.15				
4-2	0.7	9	-0.45				
4-3	1	0	-0.15				
4-4	0	0	-1.15				
4-5	1	7	-0.15				
4-6	5	0	3.85				
4-7	0	0.75	-1.15				
5-1	1	0	-0.15				
5-2	0	0.1	-1.15				
5-3	0	6	-1.15				
5-4	1	7	-0.15				
5-5	0	0	-1.15				
5-6	0	0	-1.15				
5-7	1	0	-0.15				
6-1	0	3	-1.15				
6-2	2	0.3	0.85				
6-3	8	0	6.85				
6-4	5	0	3.85				
6-5	0	0	-1.15				
6-6	0	8	-1.15				
6-7	0	1	-1.15				
7-1	3	0	1.8				
7-2	0	1	-1.15				
7-3	3	0.75	1.8				
7-4	0	0	-1.15				
7-5	1	8	-1.15				
7-6	0	0	-1.15				
7-7	0	0	-1.15				

Eventually, the result has found:

$r = \frac{-29.05}{\sqrt{170.5*379.07}} = \frac{-29.05}{\sqrt{64.631}}$ = ~ 0.116

Question 3

a) Hamming Distance

The matrix for given two graphs in part B are as follows:

	A_in	A_out	B_in	B_out	C_in	C_out	D_in	D_out	E_in	E_out
A_in	0	0	0	0	0	0	0	0	0	0
A_out	0	0	1	0	0	0	1	0	0	0
B_in	0	1	0	0	0	0	0	0	0	1
B_out	0	0	0	0	0	0	0	0	0	0
C_in	0	0	0	0	0	0	0	1	0	1
C_out	0	0	0	0	0	0	0	0	0	0
D_in	0	1	0	0	0	0	0	0	0	0
D_out	0	0	0	0	1	0	0	0	0	0
E_in	0	0	0	0	0	0	0	0	0	0
E_out	0	0	1	0	1	0	0	0	0	0

Table 3: Matrice for Left Graph

	A_in	A_out	B_in	B_out	C_in	C_out	D_in	D_out	E_in	E_out
A_in	0	0	0	0	0	0	0	0	0	1
A_out	0	0	1	0	0	0	1	0	0	0
B_in	0	1	0	0	0	0	0	0	0	0
B_out	0	0	0	0	1	0	0	0	0	0
C_in	0	0	0	1	0	0	0	0	0	1
C_out	0	0	0	0	0	0	0	0	0	0
D_in	0	1	0	0	0	0	0	0	0	0
D_out	0	0	0	0	0	0	0	0	1	0
E_in	0	0	0	0	0	0	0	1	0	0
E_out	1	0	0	0	1	0	0	0	0	0

Table 4: Matrice for Right Graph

As in the slides, I compared the rows, and found the differences:

0 1 0 0 1 1 0 0 0 1

1 0 1 1 0 0 1 1 1 0

There are 10 differences, and 10 nodes. Therefore, the following calculation has been made:

$$10 / (10 * (10-1)) = 0.111 = \sim 11 \%$$

b) Jaccard Distance

Jaccard similarity is found as follows:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Figure 2: Jaccard Similarity

To compute Jaccard distance, I first need to calculate the intersection and union. Intersection of the sets is the number of edges that are present in both graphs. There are 3.

The union of the sets is total number of unique edges in both graphs. It is 3 + 2 + 3 = 8.

By dividing the difference between union and intersection by the union, Jaccard distance can be found:

3/8 = 0.375