

CS 429/529 Assignment 5

Assignment Release Date: Monday, December 5, 2022

Assignment Submissions Due: Saturday, December 17, 2022

Assignment Submission: Submit on Moodle

Questions: gozde.yazici@bilkent.edu.tr, miraykas@cs.bilkent.edu.tr

Maximum Points: 100

Important Notes Before You Start:

- Your homework assignment file(s) name should include your full name (first and last name) and the assignment number. Your submission should be submitted as a single pdf file on Moodle using the following name template:
FirstName_LastName_A#no.pdf
- Homework assignments are to be done individually, partnering on this homework assignment is not allowed.
- **Penalty points:** Submissions not following the requested file naming/format will receive a 10 point cut.

Exercise-1: Key Concepts (20 points, 1 points for each row)

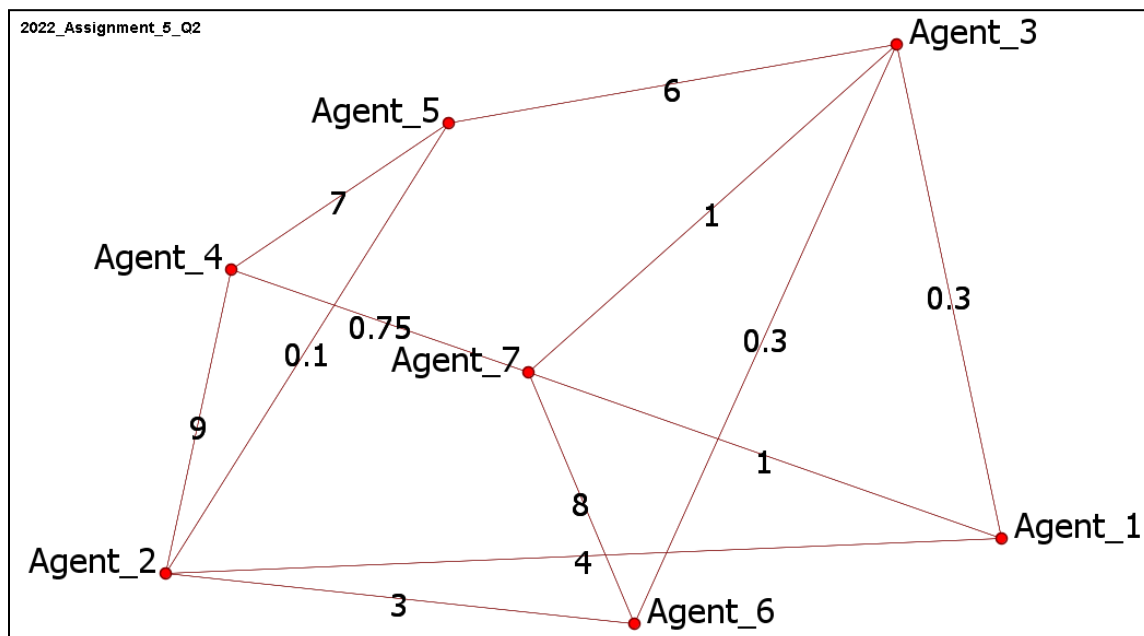
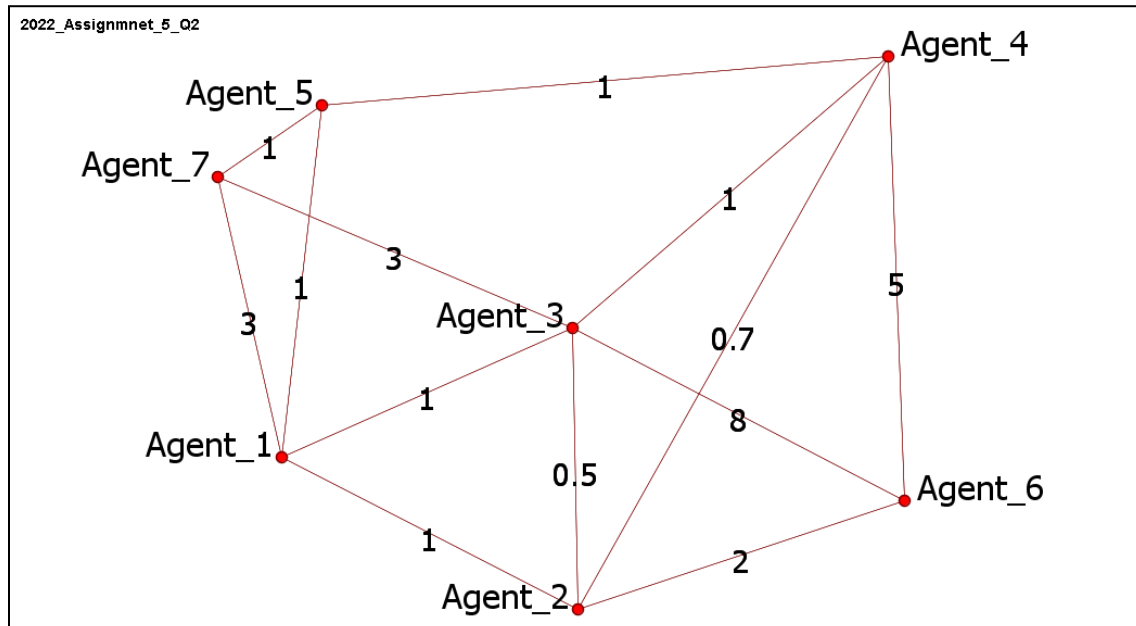
Please mark each statement with true or false. For each statement you marked as FALSE, rewrite a correct version of it in the fourth column ("Reasoning if false") in the table below to explain your reasoning. The fourth column ("Reasoning if false") should remain blank for the statements you mark as True.

	Statement	True or False?	Reasoning if false
1	A path is a walk without passing through the same link more than once.		
2	Graph and network refer to the same thing		
3	The degree of a node describes the sum of its in and out degrees		
4	Google pagerank is a variant of Eigenvector centrality		
5	QAP analysis can be applied on any two		

	networks		
6	Hamming distance compares only binary data		
7	Link analysis is used in law enforcement		
8	Centrality metrics and rankings of nodes yield the same results regardless of whether the edge list is modeled as directed or undirected		
9	Sorensen similarity gives more weight to common elements than jaccard similarity.		
10	Bipartite networks are only possible on two-mode networks		
11	All graphs with fat tailed degree distributions are scale-free graphs.		
12	Network density is calculated as $n(n-1)/2$ where n is the number of nodes in the network.		
13	Preferential attachment model works better when the limits of the nodes are well defined		
14	A clique of 4 nodes contains 4 3-cliques		
15	The Girvan-Newman clustering algorithm is an agglomerative algorithm.		
16	Edge betweenness is a metric defined at the network level.		
17	CONCOR clustering algorithm is based on structural equivalence		
18	K-means requires a preset number of clusters		
19	In statistics, correlation implies causation		
20	99% confidence level is typically sought for in research		

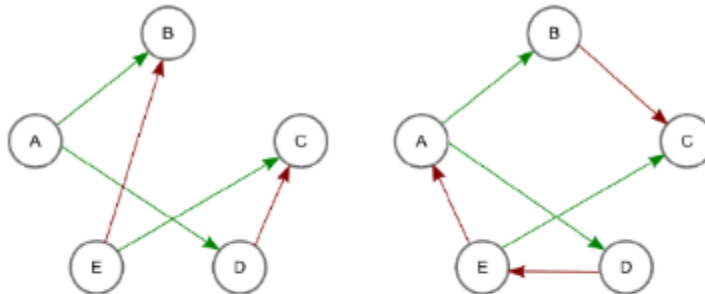
Exercise-2: Pearson Coefficient (40 points)

- a) Calculate Pearson Coefficient by hand for the following two networks using the formulation given in the lecture slides. The networks are weighted, symmetric, and contain no self loops. Labels on the links show link weights. Show all your intermediate steps. Without the detailed steps, even if the final result is correct, no points will be given.

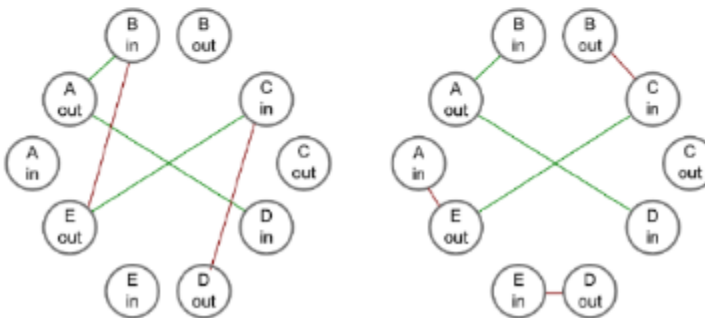


Exercise-3: Hamming and Jaccard (40 points)

A)



B)



In the upper half (A) of the figure above, two examples of directed graphs both with 5 nodes but different network topologies. Green and red edges denote, respectively, common and uncommon edges between the two graphs.

In the lower half (B) of the figure above, undirected graphs equivalent to the directed graphs depicted in (A). For instance, node A of the left graph is split into two nodes A_in and A_out. Since node A has no incoming edges in the directed graph, node A_in is disconnected from the undirected graph. Instead, for each target node X of an outgoing edge of node A, node A_out is connected with node X_in.

(20 points) Compute the hamming distance between the two graphs given in part-B of the figure above. Show your steps

(20 points) Compute the Jaccard distance between the two graphs given in part-B of the figure above. Show your steps