

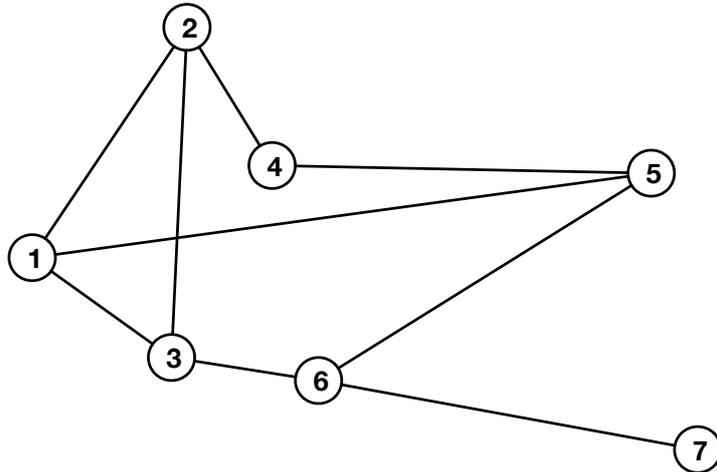
Exercise Sheet

Social and Graph Data Managementx

December 4th, 2020

Exercise 1 – Graph Measures

Consider the graph G in the following figure:



Question 1 Represent the graph as an *adjacency list*.

Question 2 Write down the degree distribution of G , and the average degree $\langle k \rangle$.

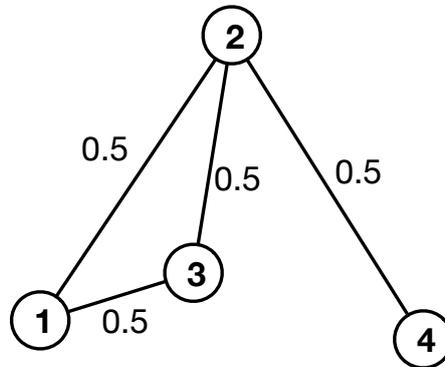
Question 3 Compute the *clustering coefficient* of node 1 in G . Explain how it is computed.

Question 4 Compute the *diameter* d_{\max} of G , and show a path of length d_{\max} in G .

Question 5 Assume that the graph was computed using a *random network* model with parameter p . What is the value of p ? Explain how you found it.

Exercise 2 – Uncertain Graphs and Influence

Consider the probabilistic (or uncertain) graph \mathcal{G} in the following figure, where each edge is annotated with its independent probability of existing:



Question 1 Give a possible world G of the graph \mathcal{G} (i.e., a deterministic graph of 4 nodes resulting from \mathcal{G}). How would you compute the probability of G ?

Question 2 Compute the *reachability probability* between nodes 1 and 4. Explain how you obtained it.

Question 3 Compute the *expected influence* of node 1 under the *influence cascade* model. Explain how you obtained it.

Exercise 3 – Link Prediction

Consider again the graph G in Exercise 2. Consider node 5 in the graph, and observe that the edges $(5, 2)$, $(5, 3)$ and $(5, 7)$ are missing. We want to *predict* the next link from node 5 by taking, among the 3 candidate links, the one having the highest link score. We want to work only with two score functions:

1. the *common neighbours* score, and
2. the *inverse distance* score.

Compute the link scores for each candidate link and each of the two score functions. For each of function, give the resulting best new link candidate.