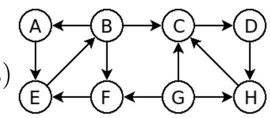

Model Mini-Test - Network Science 2021/2022 (CC4063 & CC4095)

Duration: 1h30m+30m DCC/FCUP

Answer to only **five** of the following six groups

Essential Concepts *(20%)*

1 - ***Directed Networks.*** Consider the right figure, representing a simple directed network:



a) What is the **distance** from node C to node E?
b) Identify all **strongly connected components** (SCCs) (and their correspondent nodes)
c) Redraw the network as a **DAG** (directed acyclic graph) of SCCs.

2 - ***Bipartite Networks.***

a) Give an example of a **real world bipartite network**, identifying what are the nodes and the edges.

b) What is a **bipartite network projection**? Give an example projection of the bipartite network you chose.

Graph Models *(20%)*

3 - ***Erdös-Rényi*** $G_{(n,p)}$ ***model (ER).*** Imagine you generate an ER graph with $n = 1000$ and $p = 0.01$.

a) What is the **average degree** of its nodes?

b) Draw a plot showing what you expect to be its **degree distribution**.

c) How many **connected components** do you expect the graph to have? Why?

d) If you increase the amount of nodes to 1 million (keeping the same average degree), what happens to the **average path length?** Does it stay the same? decreases? increases? by how much?

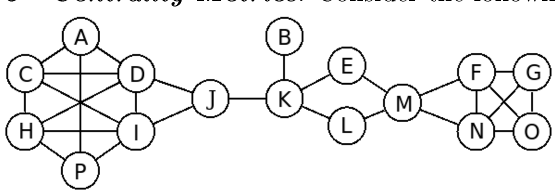
4 - ***Small-World Networks.*** Imagine you generate a small world network using the Watts-Strogatz (WS) model.

a) Which model generates a network with higher **clustering coefficient**: ER or WS? Why?

b) A purely regular network can have an high **average path length**. Explain how the WS model lowers this value.

Node Centrality *(20%)*

5 - ***Centrality Metrics.*** Consider the following network, and indicate:



a) a node with high degree but low betweenness centrality

b) a node with high closeness centrality but low betweenness centrality

c) a node with high betweenness centrality, high closeness centrality and high degree

6 - ***The PageRank Algorithm.***

a) Problems. Explain what a **spider trap** is, why it might be a problem for the PageRank algorithm and how we can avoid that problem.

b) Customization. Explain what is a **Personalized PageRank**, what is different from the normal PageRank and in what kind of situation it should be used.

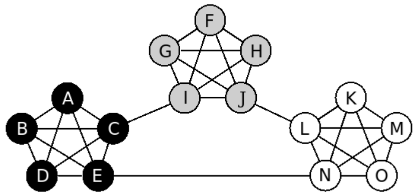
Community Structure *(20%)*

7 - Girvan-Newman and Louvain algorithms. Consider these two algorithms for community discovery.

a) Classify both algorithms in terms of being **divisive or agglomerative** .

b) Which one is faster in terms of **execution time**?

c) Consider the following network with three very natural and intuitive communities. Explain how one of the two algorithms would work on this network, by explaining a couple of its initial **greedy iterations**.



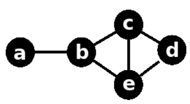
8 - Modularity.

a) Give a brief informal description of the **concept of modularity**, and what can we use it for.

Subgraph Patterns *(20%)*

9 - *Subgraph census.*

a) Calculate the **frequency of all possible induced subgraphs of size 3** on graph pf the following figure (draw the topology of each possible subgraph and indicate all occurrences as sets of nodes).



b) Explain the relationship between the concepts of **network motifs** and **subgraph frequency**.

10 - *G-Tries.*

a) Explain what **G-Trie** is. Explain what does it store, draw an example, indicate in what kind of task it might be useful and give a brief explanation of the intuition on why it is efficient.

Diffusion and Cascading Behavior *(20%)*

11 - *Spreading in Random Trees.* Consider a simple epidemic model based on random trees: a patient meets d people and infects each one with probability q.

a) Suppose you start with a **single infected patient** and that $d = 200$ and $p = 1\%$. What is the expected number of infected persons at depth level 1 and 2? Will the epidemic run forever or will it stop?

b) For a general case, what is the **reproductive number**? How can we use it to determine if a disease will spread or die out? Does it matter the amount of people that initially have the disease?

12 - *Spreading models of viruses.*

a) Describe the **SIS model**: indicate states, transition probabilities and dynamics on a complete graph.

b) Give an **example of a disease** that fits the SIS model.