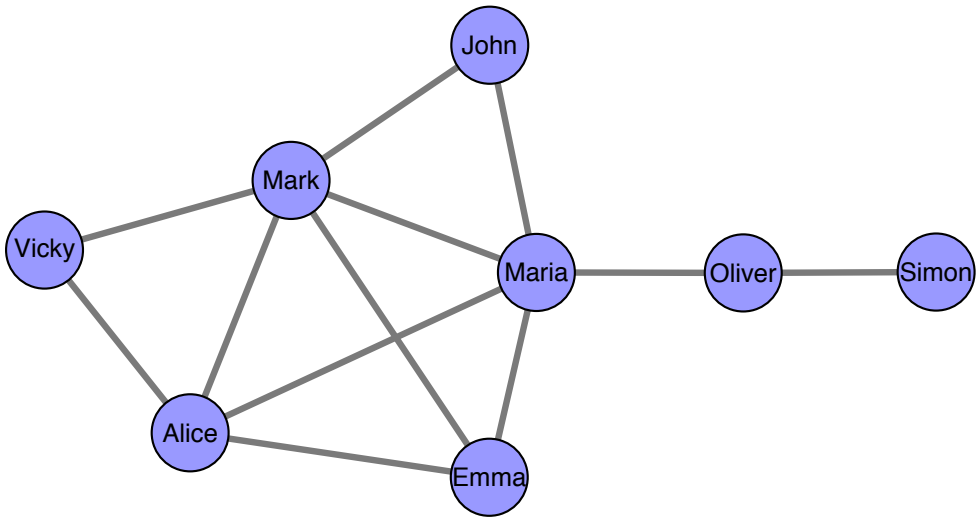
COMP30120 Machine Learning

Social Network Analysis

Q1

The diagram below shows a network representing members of a sports club, with 8 nodes and 12 undirected edges. Each node is a member and an edge between two members indicates that they are mutual friends.



(a) What is the *shortest path* between the pairs: Alice and John; Vicky and Simon? What is the *diameter* of this network?

(b) Calculate the *degree centrality* for each node in the graph. Which members have the highest and lowest centrality?

(c) If we wanted to identify the information brokers in this network, what would be an appropriate measure of centrality to use? Based on inspecting the diagram above, which node would have the highest centrality value according to this measure?

(d) If Emma leaves the club (i.e. the network), how many connected components will be in the network? If Maria then leaves the club, how many connected components will there be?

(e) The network above can also be represented as an Adjacency Matrix. Construct this matrix. Will the matrix be symmetric or asymmetric?

Q2

Co-authorship of a research paper can be thought of as documenting a collaboration between two or more authors. The Adjacency Matrix below represents a weighted undirected “co-author network”, where each non-zero entry indicates the number of times a pair of researchers have co-authored a paper.

	Albert	**Barabasi**	**Collins**	**Jones**	**Lynch**	**Mooney**	**Smith**
Albert	-	4	1	0	1	0	0
Barabasi	4	-	0	0	1	0	0
Collins	1	0	-	1	0	2	4
Jones	0	0	1	-	0	0	0
Lynch	1	1	0	0	-	0	1
Mooney	0	0	2	0	0	-	3
Smith	0	0	4	0	1	3	-

(a) Draw the network diagram corresponding to the Adjacency Matrix above. Label edges with their corresponding edge weights.

(b) Calculate the *degree centrality* for each node in the graph, ignoring the edge weights. Which authors have the highest and lowest unweighted degree centrality?

(c) Calculate the *weighted degree centrality* for each node. Which authors have the highest and lowest centrality according to this measure?
Note: *Weighted degree* is calculated as the sum of the weights on the edges associated with each node.

Q3

The edge list below represents a small subset of a Twitter *follower* network. An edge *(X, Y)* indicates that account *X* follows account *Y*.

@bbcnews,@bbcsport	@cnn,@nba
@bbcnews,@cnn	@sportcnn,@cnn
@bbcsport,@bbcnews	@sportcnn,@cristiano
@bbcsport,@sportcnn	@bbcsport,@cristiano
@bbcsport,@nba	@cristiano,@bbcsport
@bbcworld,@bbcnews	@nba,@cnn
@cnn,@sportcnn	

(a) What type of graph would be appropriate to represent this type of data? Draw the network diagram corresponding to the edge list above.

(b) Calculate the *density* of this graph.

(c) One proposed measure of "prestige" on Twitter has been to calculate an account's *Followers-Following* ratio. This is equivalent to calculating a score for a node *X*:

$$\frac{\text{In-Degree(X)}}{\text{Out-Degree(X)}}$$

Calculate this score for each node in the graph. Which accounts have the highest and lowest "prestige" score?

(d) One way to convert a directed graph into an undirected graph is to only include those edges that are *reciprocated* (i.e. only create an undirected edge between nodes *A* and *B*, if there exist both edges $A \rightarrow B$ and $B \rightarrow A$ in the original graph).

Draw the undirected reciprocal network from the undirected network above, removing any *isolated* nodes which have no edges. How many nodes and edges will exist in the new graph?