MSc AI

Graph & AI

Module Code: **B9AI101_2122_TMD1S**

Module Description: **Graph & AI**

Examiner: **Terri Hoare**

Internal Moderator: **David Williams**

External Examiner: **Dr Svetlana Hensman**

Date: 5th January 2022
Time: 9:30 – 11:30

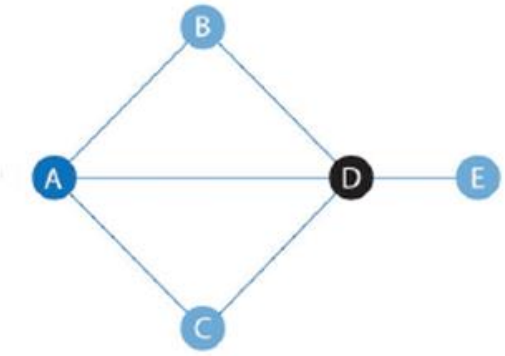
INSTRUCTIONS TO CANDIDATES

Time allowed is 2 hours
Answer ALL Questions

All answers should reference literature and case studies as appropriate. Use of scientific calculators is permitted.

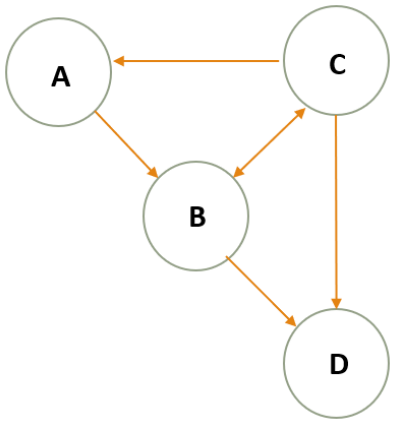
Question 1 **(25 Marks)**

a) Specify the formula used and apply **Betweenness** centrality for each of the five nodes in the Graph Network below. Show all workings.



[10]

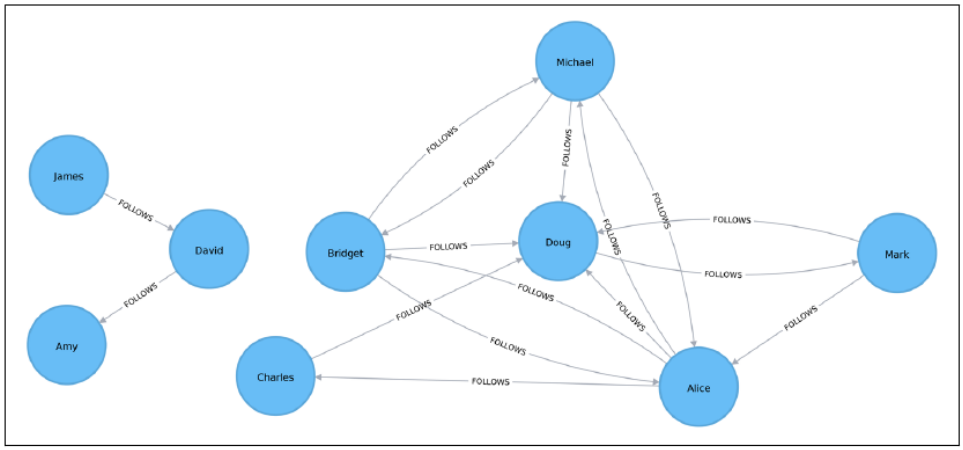
b) The intuition behind **PageRank** influence is that relationships to more important nodes contribute more to the influence of the node in question than equivalent connections to less important nodes. It is an iterative algorithm that runs either until scores converge or until a set number of iterations is reached. Specify the formula you are using and perform two iterations (iteration 0; iteration 1; iteration 2; rank) to rank the nodes in the diagram below.

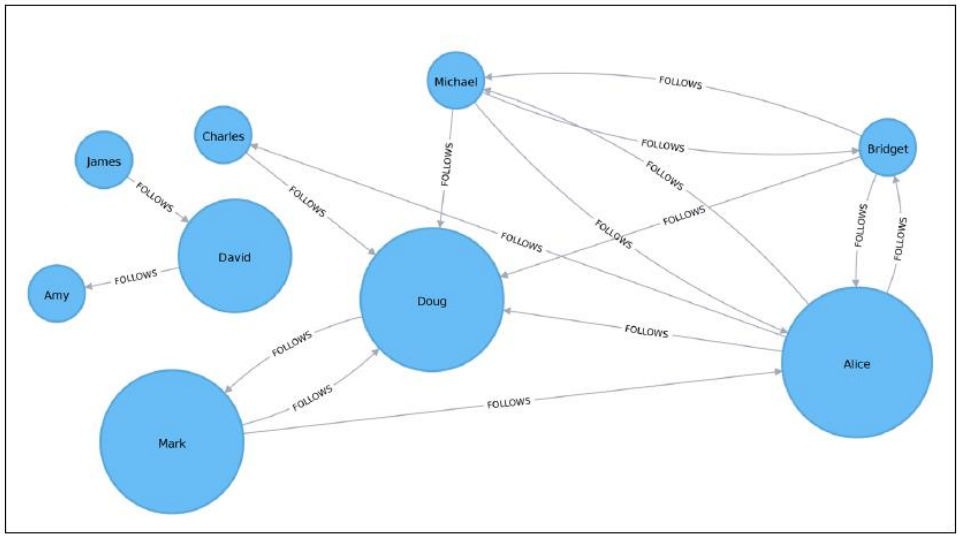


[15]

Question 2 **(25 Marks)**

a) A sample Twitter Social Graph is shown in the first figure below. Identify the graph algorithm applied to the Graph listing the evidence supporting your conclusion.





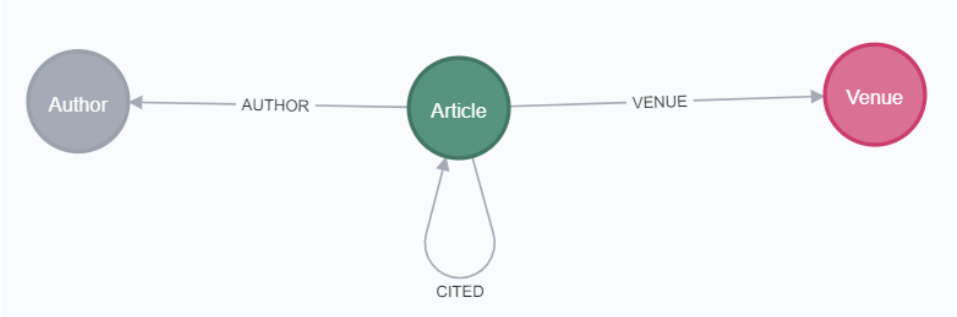
[4]

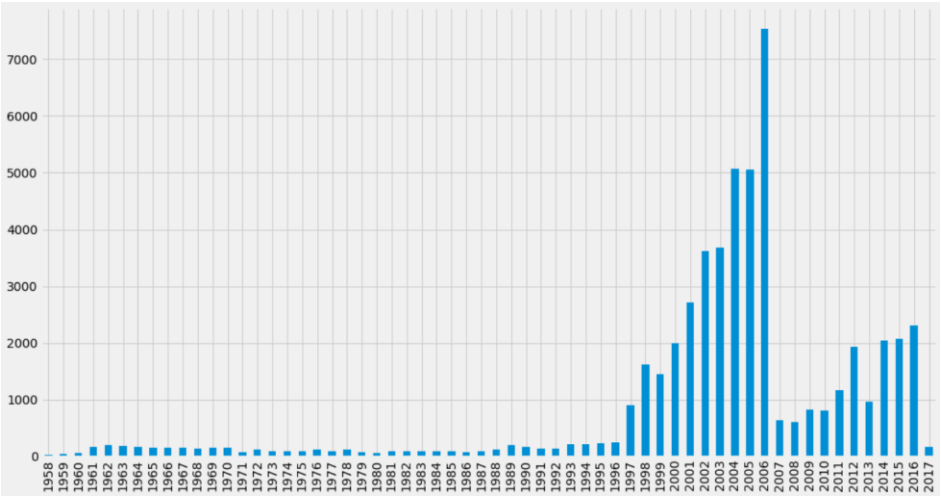
b) Identify three other use cases for the algorithm.

[6]

c) The Citations dataset is described in the paper "ArnetMiner: Extraction and Mining of Academic Social Networks", by J. Tang et al. The latest version contains 3,079,007 papers, 1,766,547 authors, 9,437,718 author relationships, and 25,166,994 citation relationships. The data and python plot of citations by year is shown below.

Outline the steps you would apply to leverage graph algorithms to build a classifier to predict the likelihood of collaboration between authors.





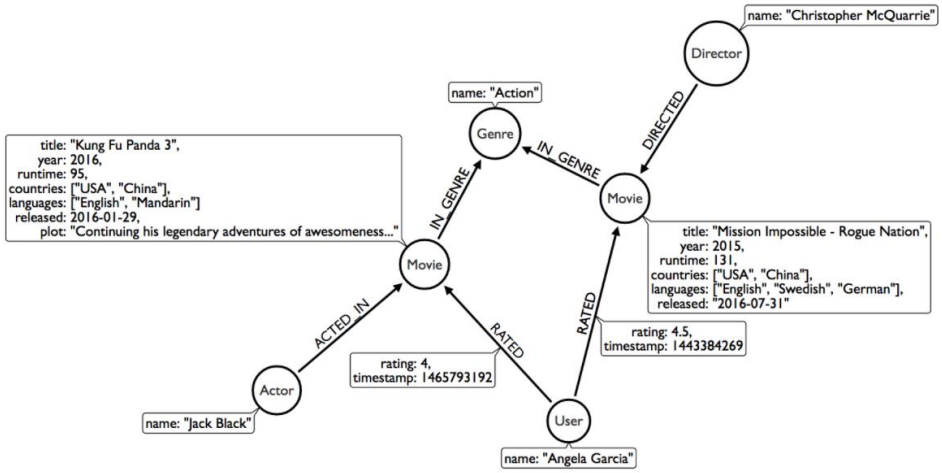
[15]

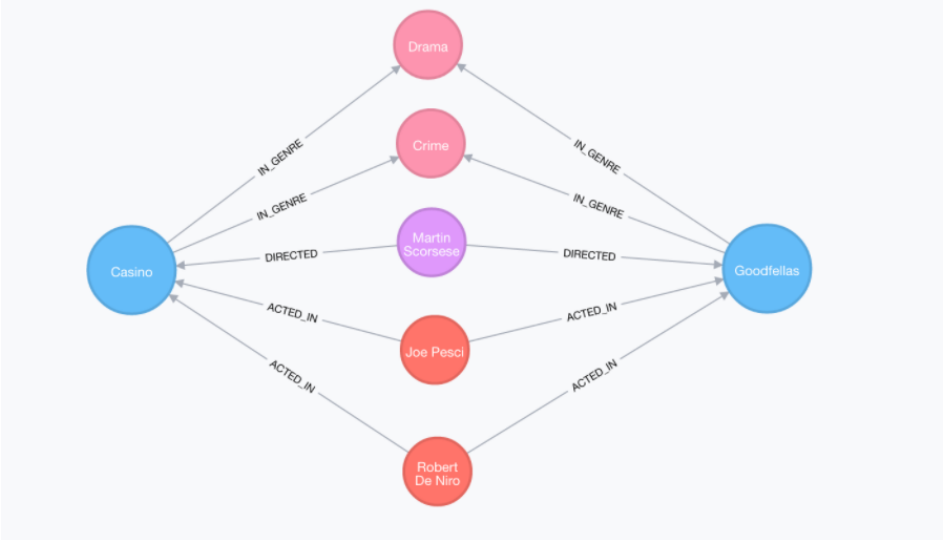
Question 3 **(25 Marks)**

a) The Movielens dataset of movies and ratings has been loaded into the property graph database Neo4j. The data model and a sample of data are shown below.

Recommend items that are similar to those that a user is viewing, rated highly or purchased previously. Specify the **Cypher** pattern matching instructions to make recommendations.

"Products similar to the product you're looking at now."





[10]

b) Compare and contrast the **Triangle Count** and **Clustering Coefficient** graph algorithms used for **Community** detection.

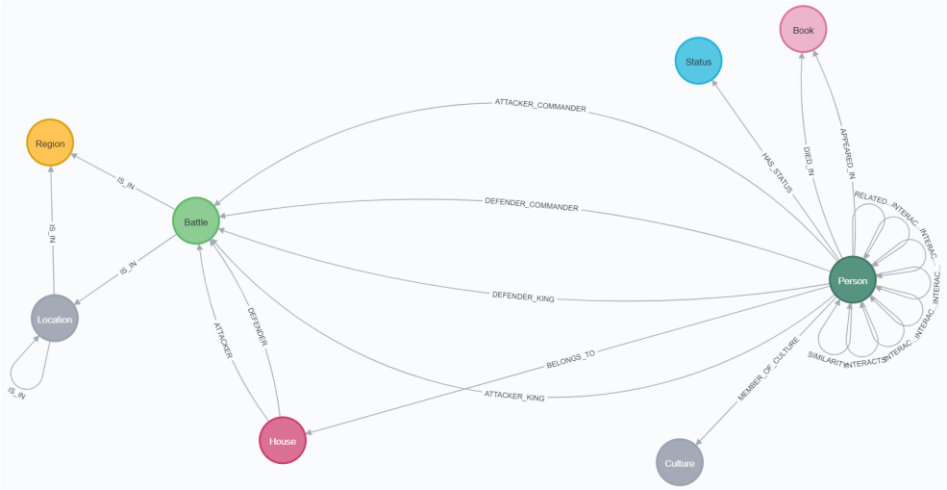
[15]

Question 4 **(25 Marks)**

a) The data model below is based on the research of Dr Andrew Beveridge as applied to the Game of Thrones series. The Person nodes represent characters in the books and the INTERACTS relationships represent characters' interactions. An interaction occurs each time two characters' names (or nicknames) **appear within 15 words of one another** in the book text.

The (:Person)-[:INTERACTS]→(:Person) graph is enriched with data on houses, battles, commanders, kings, knights, regions, locations, and deaths.

Recommend how graph algorithms can be applied to gain insight into interactions, themes, and importance of characters in the books.



END OF EXAMINATION