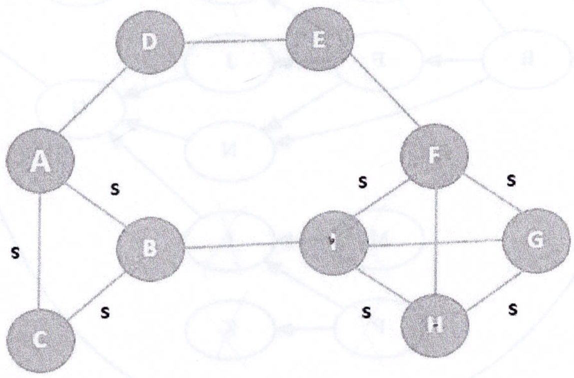
李Meile

First Partial Exam Web Science

9th December 2015

Question 1, Social Networks

Consider the graph in the figure below.



a) Based on the existing connections, which connections are more likely to appear next? Motivate your answer.

b) Does the network on the figure contains a bridge? Which ties form a local bridge?

c) A *diameter* of a graph is the longest distance in a graph. What is the diameter of the graph in the figure?

In d), e), assume that the edges marked with the letter 's' represent strong ties.

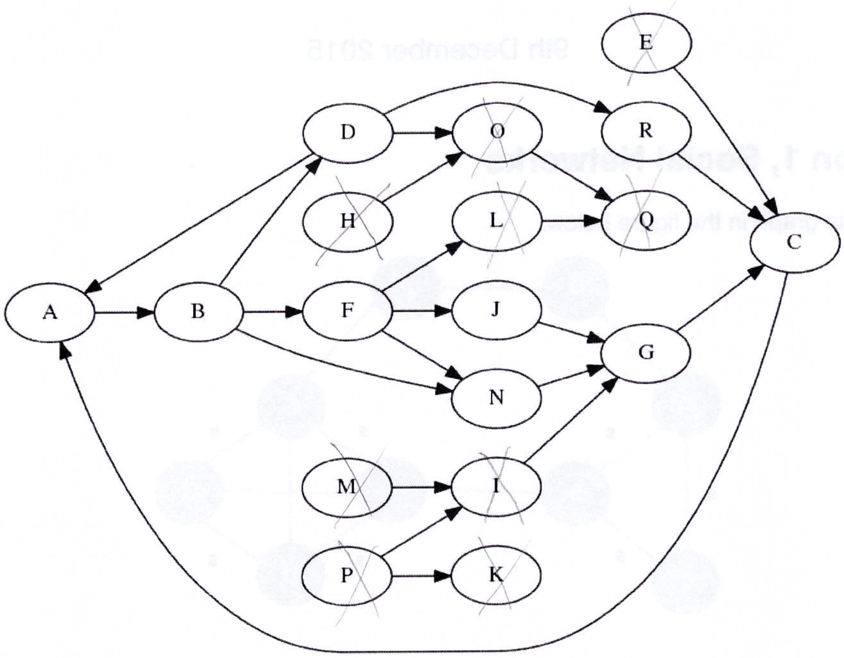
d) Assume that each node in the network satisfies the Strong Triadic Closure property. Which of the currently unmarked ties are definitely weak ties? Explain your answer.

e) Do local bridges always represent weak ties? Do all weak ties have to be local bridges? Explain your answers.

(Question 2 on next page...)

Question 2, The Structure of the Web

Consider the directed graph with a portion of a web crawl, where the nodes are web pages, and the edges are hyperlinks from one page to another.



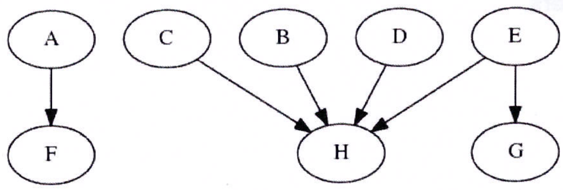
Answer the following questions, using the analysis and the definitions of Andrei Broder et al. "Graph Structure in the Web", Proceedings of the 9th WWW Conference, 2000.

a) Which set of nodes constitutes the largest strongly connected component (SCC) in this graph? Explain your answer.

b) Taking the answer of a) as the giant SCC, which nodes belong to the sets IN and OUT? Explain your answer.

c) Which nodes belong to the tendrils? Explain your answer.

Consider the following bipartite graph that models the top 5 results of a web search query for "hotels", where nodes are web pages and edges are links between web pages.

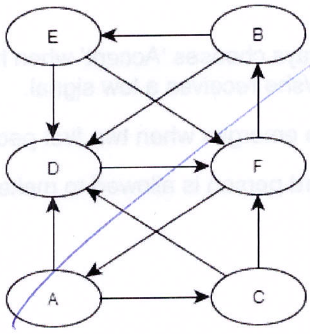


Answer the following questions:

d) Show the values that you get if you run two rounds of computing hubs and authority value updates on the graph of Web pages (so, k-step hub-authority computation with $k = 2$). Take 1 as the initial value of each page. Do not normalize the computed values.

e) Show the values after two rounds of computing hubs and authority value updates including a normalization step (It is fine to write the normalized scores as fractions rather than decimals.)

Consider the following directed graph that models a portion of a web crawl, where the nodes are web pages, and the edges are hyperlinks from one page to another. The graph also shows the proposed PageRank value for each one, expressed as a decimal next to the node.



Answer the following questions by applying the basic PageRank update rule, so the definition of PageRank that does not use scaling (or a scaling factor $s = 1$):

f) Does the assignment of numbers to the nodes from an equilibrium set of PageRank values for this graph? Explain your answer.

g) Suppose we apply the basic PageRank algorithm to the network of Question a). What would be the PageRank of each node in the graph? Explain your answer.

(Question 3 on next page...)

Question 3, Network Dynamics: Population Models

We consider the information cascade model of Chapter 16 with specific values for the probabilities as follows: the probability that Accept (A) is a good idea is p=1/2, and the probability of a High signal if Good is true (as well as the probability of a Low signal if Bad is true) is q=4/5. According to the model, a person chooses 'Accept' if, given the information he/she has, the probability that Good is true is greater than 1/2. Further, let's assume that Good is actually true.

a) Assume the first person has seen a high signal. Then, according to this person, what is the probability that Good is true?

b) Show that the first person always chooses 'Accept' when he/she receives a high signal and always chooses 'Reject' is he/she receives a low signal.

c) Show that the Reject cascade emerges when two first people choose Reject.

d) What changes in (c), if the third person is allowed to make two observations?

(Question 4 on next page...)

Question 4, Network Dynamics: Structural Models

Question 4.1

a) Explain what is a power law distribution of the degrees in a network. How do we recognize power laws? Why this model can be used to formally describe the presence of hubs – nodes with large number of connections?

b) A researcher is searching for a relevant scientific literature using *Google Scholar*, which shows the number of citations received by each paper. Explain how this may lead to the power law distribution in the number of citations across scientific papers.

Question 4.2

A clique of size k is a sub-graph, which contains k nodes, all connected to each other. Consider a network, which contains a clique of size k such that only one node v in this clique is connected to another node w outside of the clique. Consider the model from Chapter 19 for the diffusion of a new behavior through a social network. Everyone starts with behavior B, and a threshold for switching to a new behavior A is q. Any node will switch to A if at least a fraction q of its neighbors has adopted A. Assume that node w has switched to A.

a) Take $k = 4$. For which value of q no node in the clique switches to A?

b) Generalize the result in (a) to a clique of an arbitrary size $k \geq 3$.

c) Consider node $x \neq v$ is in the clique. Assume that x has no connections outside the clique. Is it possible that x switches to A while some other nodes in the clique stick to B?

d) What changes if the nodes in the clique create more links to nodes, which accepted A, outside the clique? Consider several scenarios. For example, is it possible that only some of the clique members switch to A while others stick to B?