Networks: Fall 2023	**Homework 4**
David Easley and Yian Yin	**Due 3:30pm, Thursday, October 19, 2023**

Homework solutions should be submitted by upload to Gradescope, which can be accessed via Canvas. The file you upload **must be typed and submitted in PDF format**. Handwritten assignments will not be graded. However, you can draw graphs and insert them into your pdf. You can create a separate file with the solutions (you don't need to repeat the questions); it is fine to create the homework in any format provided it's typed and handed in as a single PDF file. When you upload your pdf to Gradescope be sure to assign your answers to the correct question.

To be eligible for full credit, your homework must come in by 3:30pm Thursday. We will also accept late homeworks after 3:30pm Thursday until 3:30pm Friday for a deduction of 10% of the total number of points available. Gradescope will stop accepting homework uploads after 3:30pm Friday; after this point we can only accept late homework when it is accompanied by a University-approved reason that is conveyed to and **approved by the TA in charge of this homework prior to the due date of the homework**. (These include illness, family emergencies, and travel associated with university activities.)

The TA in charge of this homework is Utku Umur Acikalin (ua45@cornell.edu).

Reading: The questions are primarily based on the material in Chapters 11, 13, and 14.

---

(1) [8 points]

Consider the trading network with intermediaries described by Figure 1 in which there are two sellers S_1 and S_2 (on the left), two buyers B_1 and B_2 (on the right) and two traders (intermediaries) T_1 and T_2. Each seller can trade with only one trader; trader T_1 for seller S_1 and trader T_2 for seller S_2 . Buyer B_1 can only trade with trader T_1. Buyer B_2 can only trade with trader T_2. The sellers' value for their good is 0, both buyers value the good at 1, as indicated on the figure.

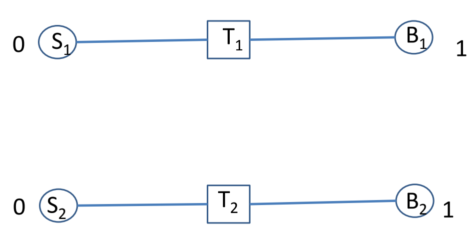

Figure 1: The network buyers, sellers and traders for Question (1).

(a) [3 points] Describe all possible Nash equilibria, including both prices and the flow of goods. (That is, explain why the Nash equilibria that you are giving are in fact Nash equilibria, and why they are the only ones.) How much profit does each trader make?

(b) [5 points] Now assume we add an extra buyer B_3 who can trade with either traders T_1 and T_2, and who has value 2 for the good. Describe all possible Nash equilibria now, including both prices and the flow of goods. How much profit can each trader make?

---

(2) [12 points]

Consider the directed graph shown in Figure 2, with nodes representing Web pages and each directed edge representing a link from one Web page to another. Recall from Chapter 13 and from class that a *strongly connected component* in a directed graph is a subset S of the nodes such that two properties hold: (i) every node in S has a path to every other node in S; and (ii) S is not a subset of some larger set T that also satisfies property (i).

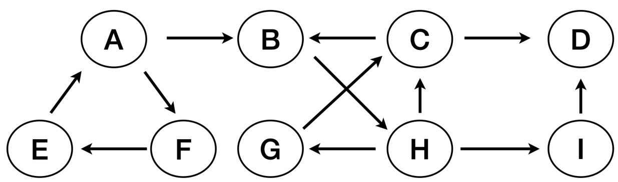

Figure 2: The network of Web pages for Question (2).

(a) (6 points) How many strongly connected components are there in the graph? For each component, list all nodes included in it, and determine the position of this component in the bow-tie structure, i.e., whether it belongs to (i) giant SCC, (ii) IN, or (iii) OUT.

(b) (3 points) Suppose you are allowed to add one link to the graph in Figure 2, going from one node in the figure to another. Which link would you add if you wanted to maximize the size of the giant strongly connected component? Write the updated giant strongly connected component after the addition.

(c) (3 points) Instead, suppose you are allowed to delete one link from the graph in 2, removing it from the graph but otherwise keeping all the existing nodes and all the other links in place. Which link would you delete if you wanted to minimize the size of the giant strongly connected component? Write the updated giant strongly connected component after the deletion.

---

(3) [12 points]

As we mentioned in class, the web structure plays an important role for search engines. As there is no official registry for all web pages in the world, companies such as Google often utilize crawling programs to obtain a large set of pages on the Interenet. Let us consider a toy example as illustrated in Figure 3.

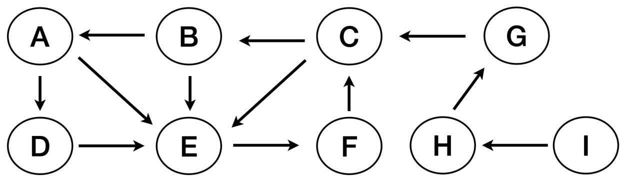

Figure 3: The network of Web pages for Question (3).

(a) [3 points] Consider an iterative crawler that adds webpages to its internal database: The crawler begins with a seed page, x, and stores x in its database. In each round, the crawler processes every page currently in the database, following all outgoing links from this page to visit the nodes they lead to. Any new pages not previously in the database are then added. For example, if one starts from the seed page $x = F$. In the first round, node C will be added to the database. In the second round, nodes B and E will be added to the database. The program will keep running until no new pages are added. Which page would you choose as the seed page if you wanted to maximize the (eventual) number of pages in the database? Give an explanation for your answer.

(b) [3 points] Consider another iterative crawler that adds webpages to its internal database: The crawler begins with a seed page, x, and stores x in its database. In each round, the crawler processes every page currently in the database. Instead of following all outgoing links from the page as in **(a)**, this crawler will select only one outgoing link, which is chosen uniformly at random. It will then follow the selected outgoing link to visit the node it leads to. Any new pages not previously in the database are then added. For example, if one starts from the seed page $x = F$. In the first round, node C will be visited (and added to the database). In the second round, there's 50% chance that node B will be visited (and added to the database), or 50% chance that node E will be visited (and added to the database). Assume we pick the seed page x randomly from the graph. What's the probability for each page $A - I$ to be visited in the first round? What's the probability of the nodes to be visited in the second round?

(c) [3 points] Under the assumption of **(b)**, what's the approximate probability for page G to be visited in round N when N is very large? Give a simple explanation.

(d) [3 points] Assume we delete the edge from G to C and add an edge from G to I. Under the assumption of **(b)**, what's the approximate probability for page G to be visited in round N when N is very large? Give a simple explanation.

(4) [8 points]

In Chapter 14, we discussed the fact that designers of Web content often reason explicitly about how to create pages that will score highly on search engine rankings. In a scaled-down setting, this question explores some reasoning in that style.

In class and in Chapter 14 we talked about the rule for updating hub and authority values in a sequence of rounds. One round of updates for the hub and authority rule applies the authority update rule using the current hub scores and then applies the hub update rule to the resulting authority scores — so a single round consists of two updates, first to update the authority scores, and then to update the hub scores. As in Chapter 14, the authority update rule sets each node's authority score to be the sum of the hub scores of all nodes that point to it; the hub update rule sets each node's hub score to be the sum of the authority scores of all nodes that it points to. Also, remember that we begin with each hub score set to 1, so the first round creates authority scores when all hubs have score 1, and then the new hub scores that result from these updated authority scores.

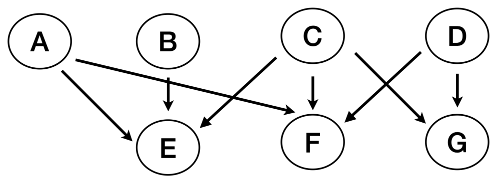

Figure 4: The network of Web pages for Question (4).

(a) [4 points] Show the values that you get if you run two *rounds* of computing hub and authority values on the network of Web pages in Figure 4. (That is, the values computed by the k-step hub-authority computation when we choose the number of steps k to be 2.)

Recall also that computing hub and authority values involves a final *normalization* step, in which at the end we divide each authority score by the sum of all authority scores, and divide each hub score by the sum of all hub scores. (We will call the scores obtained after this dividing-down step the normalized scores.)

Show the values both before and after this final normalization step. It's fine to write the normalized scores as fractions rather than decimals.

(b) [4 points] Now we come to the issue of creating pages so as to achieve large authority scores, given an existing hyperlink structure. In particular, suppose you wanted to create a new Web page X, and add it to the network in Figure 4, so that it could achieve a (normalized) authority score that is as large as possible. One thing you might try is to create a second page Y as well, so that Y links to X and thus confers authority on it. In doing this, it's natural to wonder whether it helps or hurts X's authority to have Y link to other nodes as well.

Specifically, suppose you add X and Y to the network in Figure 4. In order to add X and Y to this network, one needs to specify what links they will have. Here are two options; in the first option, Y links only to X, while in the second option, Y links to other strong authorities in addition to X.

- Option 1: Add new nodes X and Y to Figure 4; create a single link from Y to X; create no links out of X.
- Option 2: Add new nodes X and Y to Figure 4; create links from Y to each of A, B, and X; create no links out of X.

For each of these two options, we'd like to know how X fares in terms of its authority score. So, for each option, show the normalized authority values that each of A, B, and X get when you run the 2-step hub-authority computation on the resulting network (as in part **(a)**). (That is, you should perform the normalization step where you divide each authority value down by the total.)

For which of Options 1 or 2 does page X get a higher authority score (taking normalization into account)? Give a brief explanation in which you provide some intuition for why this option gives X a higher score.