Solutions for Homework 1
Networked Life, Fall 2014
Prof Michael Kearns

1. (16 points: Graded by TA-Ryan) For this problem we consider Wikipedia articles in English. For some background, a Wikipedia article has hyperlinks to other Wikipedia articles if they are referred to in the current article. For example, the Wikipedia article for Mathematics has links that redirect the reader to other Wikipedia articles, like "structure", "logic", "Gauss", etc. We define the *first main link* of a Wikipedia article to be the first link in that article that is **not**:

   - Italicized
   - Parenthesized
   - An external link
   - A link to the current page
   - Red

   For example the *first main link* for "Mathematics" is "quantity".

   Consider the directed network G whose vertices are all Wikipedia articles. Further, if the *first main link* in article u takes you to article v then there is a directed edge from u to v.

   (a) (6 points) A *cycle* in a directed graph is a path (a series of vertices, each of which can be reached by following directed edges in the correct direction) from one node u back to itself, and the length of a cycle is the number of edges it contains. From the vertex "Philosophy", is there a cycle that returns to "Philosophy" in G? If so, what is the length of this cycle?

   Solution. The cycle is given as follows of length 6:

   philosophy → reality → actually exists → aware → conscious → quality → philosophy.

   □

   (b) (6 points) Is there a path from "Computer" to "Philosophy"? If there is a path, do you know if that is the shortest path between these two vertices in G?

   Solution. The path of size 18 given below.

   Computer → programmed → instructions → computer architecture →

   Electronic Engineering → Engineering → scientific → knowledge → fact →

   experience → experiment → hypothesis → explanation → set → mathematics →

   quantity → property → modern philosophy → philosophy.

Since every node has at most one outgoing edge, there is at most one path from one node to another. Hence, this must be the shortest path.

Common Issue: The network is *only* defined to have first main links. No other edges in the graph are given, so there is only one way to get from one node to another. It does not matter that a wikipedia article has multiple links pointing to various other articles, G only has first main links. □

(c) (4 points) Give a one sentence reason why the following statement is true or false: There exists a cycle in G that starts at "Computer" and returns to it.

Solution. There is no way to return to Computer, because Philosophy will just keep cycling around.

Note: It turns out that Wikipedia has a lot of articles that have a path to the article "Philosophy" using only first main links. This has led to a Wikipedia article on that exact topic `http://en.wikipedia.org/wiki/Wikipedia:Getting_to_Philosophy`. You can also use an app to find the "Philosophy Number" of any Wikipedia article using only first main links, which is found here: `http://www.xefer.com/wikipedia` □

2. (8 points: Graded by TA-Ryan) Recall the "Economic Contagion" model discussed in class, in which every vertex in a graph starts with \$1, and at each time step each node distributes its current wealth to all of its neighbors equally. Recall that an equilibrium is when the amount of cash each vertex receives is the same as the cash it distributes. What is the equilibrium wealth for each node in the graph given in Figure 1?

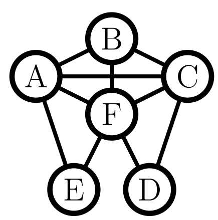


Figure 1: Problem 2 – Compute the equilibrium wealth for each node.

Solution. The exact equilibrium for each node

$$(A, B, C, D, E, F) = (6/5, 9/10, 6/5, 3/5, 3/5, 3/2).$$

Common Issue: There is a simple formula that computes the equilibrium wealth, i.e. the wealth for vertex i is (degree(i) / sum of degrees)$*$ total wealth. You do not need to have each node distribute its wealth equally to each of its neighbors many times! □

3. (9 points: Graded by TA-Ryan) We call a network $G = (V, E)$ *bipartite* if its set of vertices V can be partitioned into two sets U_1 and U_2 (i.e. every vertex is in either in U_1 or U_2, but not both) such that there is no edge between any two vertices in U_1 and there is no edge between any two vertices in U_2.

   (a) (6 points) Give a real world example of a bipartite graph. Carefully describe what the vertices represent in your example, and the definition of which pairs of vertices are connected and what the edges represent in the real world. Feel free to be creative.

   Solution. Open Ended. No partial credit. □

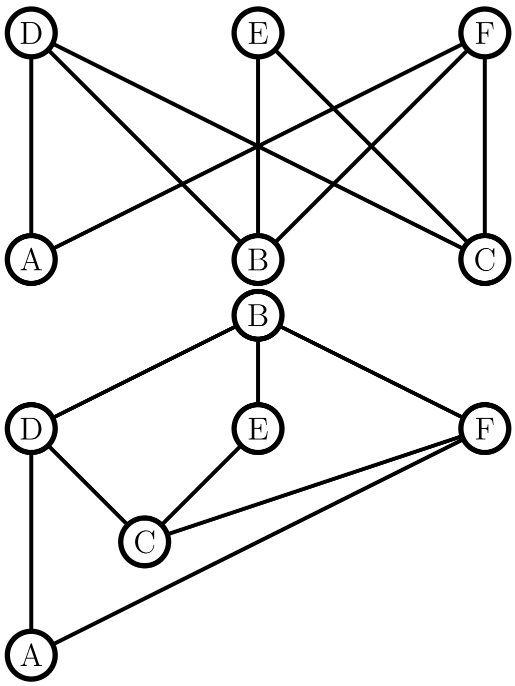
   (b) (3 points) How many colors are required to color any bipartite graph? Explain your reasoning.

   Solution. Color one partition one color and the other partition the other color. This coloring works because there are no edges between the same partition.

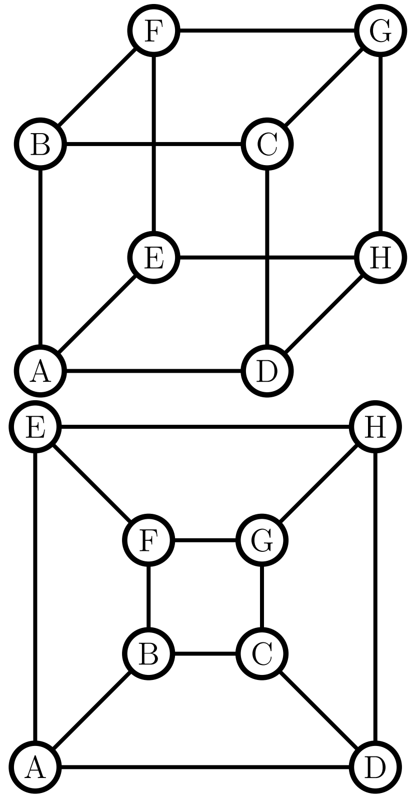
   Common Issue: To prove that you only need two colors to color any bipartite graph, you need to say HOW to color the bipartite graph, i.e. color U_1 one color and U_2 the other color. □

4. (15 points: Graded by TA-Ryan) Recall that a graph is *planar* if it has *some* (and not neceessarily any) visualization or layout that can be drawn in two dimensions in a way where no edges cross or touch each other. Determine whether each graph in Figure 2 is planar or not. If the graph is planar, give the visualization that proves it. If it is not planer, briefly describe why the graph cannot be planar.

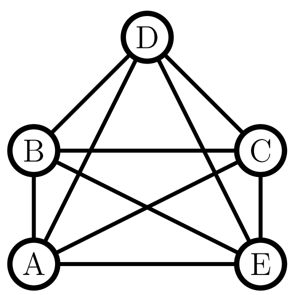
   Solution. Parts (a) and (b) are planar as indicated by Figure 2. Part (c) is not four-colorable so it is not planar this is due to the Four Color Theorem or you can use Kuratowski's criterion that says if $|V| \geq 3$ then $|E| \leq 3|V| - 6$ if the graph is planar. If you used these results, you must state the theorem name or a cite! These theorems were not covered in lecture, so it would have sufficed to give a drawing of the complete graph on 4 vertices and show that there was no way to add a 5th vertex that connected to the other nodes. It was not enough to notice that it was a complete graph, and thus not planar. This is because the complete graph on 2, 3, and 4 vertices are planar. □



(a) 5 points



(b) 5 points



(c) 5 points

Figure 2: Problem 4 – Are the above graphs planar?

5. (10 points: Graded by TA-Shahin) Consider the situation where an app, say Farmville, on Facebook requires anyone that uses the app to share their friend list. Some people do not want this breach of privacy, so they opt out of using the app. As an example, we will say Alice does not want Farmville to learn who she is friends with. However, if she is friends with Bob and he opts into using the app, then Farmville learns that Alice is friends with Bob, without Alice doing anything!

   We are given the network G in Figure 3 where the set of vertices is a few of the people on Facebook and the edge between two nodes u and v represents that u and v are friends on Facebook. Farmville wants to learn all of the friendships on G, i.e. the edges.

   (a) (5 points) Given that person E already uses Farmville and no one else does,

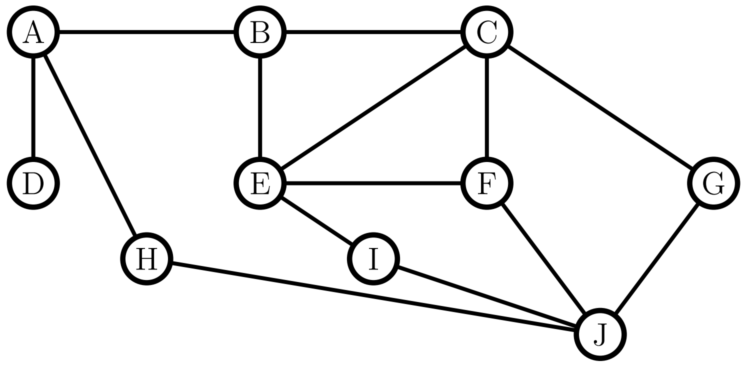


Figure 3: Problem 5 - Friendships on Facebook

what is the fewest number of additional people that Farmville needs to use their app before they can learn all the friendships. Who are the additional people?

Solution. This problem is inspired by the famous computer science problem called the *vertex cover* (see: `http://en.wikipedia.org/wiki/Vertex_cover` for more details). In this problem, you need to detect (cover) all the friendships (edges) in the graph. To detect an edge, you need Farmville to have at least one of the nodes connected by that edge to join.

Since node E uses Farmville, the edges $(E, B), (E, C), (E, F)$ and (E, I) are covered already. To detect edge (A, D) you only have two choices: you can ask either A or D to join. If you add A you can also cover (A, H) and (A, B) but adding D does not cover any additional edges other than (A, D). So you ask A to join. You can continue this argument to make nodes J and C to join. This result in the cover consisting of $\{A, C, J\}$ which has size of 3[1]. There is no other cover of size 3.

For grading, 2 points were deducted if you answered $\{A, C, J\}$ plus extra nodes. Also 2 points were deducted if your answer has two nodes from $\{A, J, C\}$ (plus possibly other nodes). 3 points were deducted if your answer has only one node from $\{A, C, J\}$ (plus possibly other nodes). □

(b) (5 points) Given that person J refuses to use the app and no one else uses the app initially, what is the fewest number of people Farmville needs to use their app before they can learn all the friendships? Who are they?

Solution. Since J refused to join Farmville for any edge that has J on one side, you need to add the node on the other side of the edge. This will result in $\{F, G, H, I\}$ to join. You can use the same approach as in part (a) to see that you also need A and two from $\{B, C, E\}$ to join. So there are three distinct covers of size 7: $\{A, B, C, F, G, H, I\}, \{A, B, E, F, G, H, I\}, \{A, C, E, F, G, H, I\}$ and this is the smallest cover possible.

[1] You can verify that this is the smallest cover by considering a set that consists of E and any other two nodes and verifying whether all the edges are covered or not. The answer is no regardless of which two nodes you pick.

For grading, no point was deducted if you found any of these 3 sets. 1 point was deducted if you have 6 out of the 7 nodes. 3 points were deducted if you only include 3 out of the 7 nodes (which was the case for most of people who did not answer this part correctly). □

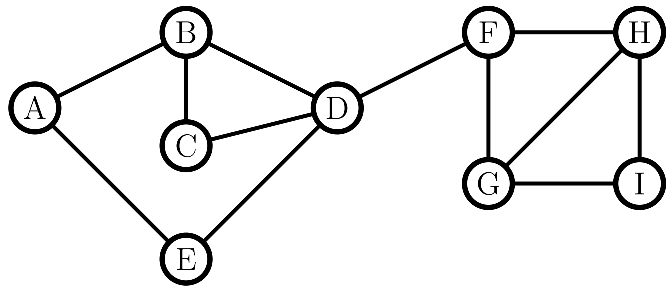


Figure 4: Problem 6

6. (12 points: Graded by TA-Shahin) Consider the graph G given in Figure 4.

    (a) (5 points) For each node in G, compute its clustering coefficient and then average them to get the clustering coefficient for the whole graph G

    (b) (7 points) Compute the distances between all pairs of nodes in G. Then compute the diameter of the graph G.

Solution. (a) We first compute the clustering coefficient for each node and then take the average to find the clustering coefficient of the full graph $CC(G)$.

$$c(A) = c(E) = 0 \quad c(B) = c(F) = 1/3 \quad c(C) = c(I) = 1 \quad c(D) = 1/6 \quad c(G) = c(H) = 2/3$$

$$CC(G) = 25/54 = 0.4630$$

For grading, 0.5 point was deducted for each clustering coefficient that is computed incorrectly. Similarly, 0.5 point was deducted for incorrect final answer.

(b) We now compute the diameter by first computing all the pairwise distances The distances are in Table 1. We then take the average of these values to get:

$$diam(G) = 2.25$$

For grading, 0.5 point is deducted for each incorrect distance. 0.5 point is deducted for incorrect final answers with the correct formula. 1 point is deducted, if you simply mention the diameter is 5. 5 is the maximum length of the shortest path between two edges in the network. However, the diameter is simply the average of the distances you computed.

□

Distance	A	B	C	D	E	F	G	H	I
A	-	1	2	2	1	3	4	4	5
B	-	-	1	1	2	2	3	3	4
C	-	-	-	1	2	2	3	3	4
D	-	-	-	-	1	1	2	2	3
E	-	-	-	-	-	2	3	3	4
F	-	-	-	-	-	-	1	1	2
G	-	-	-	-	-	-	-	1	1
H	-	-	-	-	-	-	-	-	1
I	-	-	-	-	-	-	-	-	-
J	-	-	-	-	-	-	-	-	-

Table 1

7. (14 points: Graded by TA-Shahin) For this question you will need to have a Facebook account. Consider a network consisting of all of your Facebook friends (but not yourself) *i.e.*, the nodes are your Facebook friends and there is an undirected edge between two nodes if the two of your friends who represent the nodes are Facebook friends with each other (You might find this link useful for checking the friendship of your mutual friends `https://www.facebook.com/help/210842332283728`).

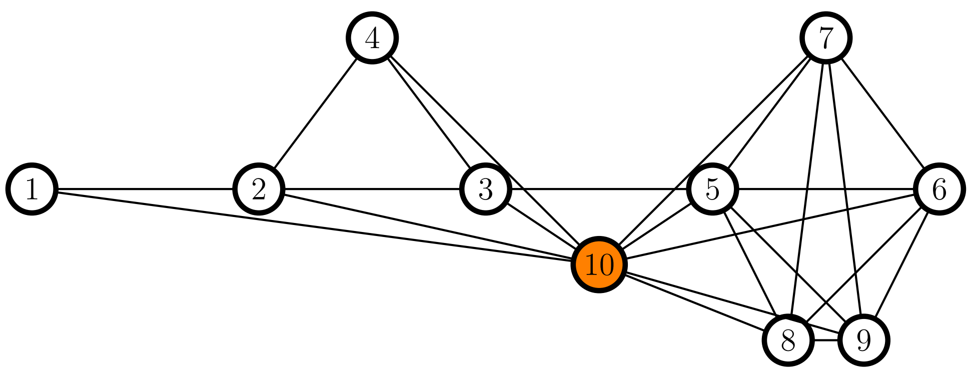


Figure 5

(a) (6 points) A clique is a subset of nodes in the graph such that every pair of nodes in the subset are connected to each other by edges. The size of the clique is the number of nodes in that clique *e.g.*, in Figure 5 the nodes 2, 3 and 4 create a clique of size 3 and the nodes 5, 6, 7, 8 and 9 form a clique of size 5. Try to find the largest clique in your network of Facebook friends and report the size of that clique. How big is this clique compared to your number of friends? Briefly describe the approach you used. (Do not use any of the available apps that find large cliques for you.)

Solution. To find a large clique you can look at different communities on your Facebook friends and try to find a clique in one of those communities based on *(i)* how big is each community and *(ii)* how connected you think the people in that community are. Once you picked a community you can start building a big clique by adding members one by one *e.g.*, you can start by the person in the community that you have the most mutual friends with and add people one by one. Note that when you add a person to the existing clique you need to verify that the person is connected to all the previous members of the clique.

Since finding the biggest clique in this matter might be cumbersome, for grading, we mainly considered the approach you are using. Some of the main issues were as follows:

- Please be more precise in your answers *e.g.*, "I think my largest clique is about 120" is not very helpful. Furthermore, the size of the clique should be compared with the number of your friends. While a clique of size 10 is fairly large if you have 50 Facebook friends, it is very small if you have 2000 friends.
- Cliques should not be confused with connected components. In cliques, for any pair of nodes there should be an edge connecting them. However, this is not the case for connected components. For example the whole network in Figure 5 is a connected component but the size of largest clique is only 5 (before adding node 10). It seems a lot of students just reported a very large connected component.
- To verify that n nodes form a clique, you need to verify $n(n-1)/2$ connections between the members of the clique. A lot of you returned cliques between size 100 and 200 and it's hard to imagine you actually verified > 5000 friendships to find this clique.

□

(b) (2 points) Add yourself to the network of your Facebook friends by adding a node that represents yourself and is connected to all the other nodes in the network. Can you find a bigger clique than the clique you found in part (a) for this new network?

Solution. Consider the largest clique that you found in the previous part. By adding yourself to the network, the same clique plus you is a clique that is one size bigger than the clique you found in part (a). Figure 5 shows an illustration of this with node 10 added to the network. For grading, unless on a few exceptions, no partial credit was given to wrong answers. □

(c) (6 points) An independent set is a subset of all the nodes such that for no pair of nodes in the set there is an edge between the pair. For example in Figure 5 nodes 1, 4 and 5 form an independent set because there is no edge between any of these three nodes. The size of the independent set is the number of nodes in the independent set *e.g.*, the size of the independent set containing nodes 1, 4 and 5 is 3. Try to find the biggest independent set in your network of Facebook friends, report its size. How big is this clique compared to your number of friends? Briefly describe the approach you used and compare it with the approach you used for finding large cliques.

Solution. The approach you can use here is the opposite of what you did in finding cliques. You can divide your friends into different communities and pick a few people from each community. The reason to pick a few instead of 1 is that there might be a pair of your friends in a community that are not connected to each other (*e.g.*, probably not all of your Facebook friends from high school are friends with each other on Facebook.). Once you add a person, you need to verify that this person is not friends with anybody else that you picked till now. Finally, you can add all the friends that you have no mutual friends with without any verification. □

8. (10 points: Graded by TA-Shahin) Find two arguments in the *The Tipping Point* book that you disagree with. Be sure to state why you disagree with Gladwell's reasoning. Please provide the page number or the chapter where you find the argument. Also, if possible, support your statement with citations to other evidence, facts or studies.

   Solution. Here is a possible solution. Note that the brackets determine the different parts that you need to discuss in your solution. In pages 140–151 [reference to the part of the book], Gladwell argues that the crime rate in NYC decreased as a result of the efforts of the police department to fight vandalism which resulted in the "broken window" theory [stating the argument that I disagree with]. First, broken window theory implies that small crimes can make way for bigger ones. However, some later research suggested that there might be no relation between small and bigger crimes (*e.g.*, Ludwig Jens in his book "Broken windows"). Second, Gladwell seems to overlook some critical evidence in his reasoning. During the same period that crime rate decreased in NYC the police department drastically increased the number of trained police officers in the streets of NYC. Furthermore, during the same period, the crime rates in other parts of United States decreased without the police departments applying the same techniques as the police department in NYC (both from Freakonomics by Steven Levitt). [stating why I disagree with the argument along with references to other evidence and studies]

   For grading, a point is deducted every time you violated one of the issues I mentioned in the brackets. □

9. (6 points: Graded by both TAs - Ryan and Shahin) Indicate whether the following statements are *True* or *False*. Provide a short explanation for your answer.

   (a) (2 points) In Kleinberg's model of navigation in grids, after each forwarding step the distance to the target decreases.

   (b) (2 points) In *The Tipping Point* book the term *maven* is used to refer to individuals with high degree in the network.

(c) (2 points) Travers and Milgram showed in their 1969 study that having connectors in the network is necessary in order for the network to have a small diameter.

Solution. For grading, 1 point is deducted for each incorrect answer and 1 point is deducted for each incorrect explanation.

(a) True - Because each point is connected to its four compass neighbors and (at least) one of these neighbors has a smaller Manhattan (grid) distance to the target than the current point. This is regardless of the fact that whether the point has long distance connections or not. This also has nothing to do with the fact that $r = 2$ or not in the Kleinberg's model!

(b) False - That's the definition of connectors. Mavens are information specialists. They are people we rely upon to connect us with new information and do not necessarily have high degrees.

(c) False. There are two ways that a network can have a small diameter: *(i)* existence of connectors, *(ii)* existence of lots of connections (as in the case for dense networks). Travers and Milgram's study showed that connectors exist and hence, the diameter is low in their experiment. However, the existence of connectors is only sufficient in getting a low diameter. The network can be dense and has no connectors. Note that the statement will become true if we replace the word necessary with sufficient.

□