Homework 1
Networked Life, Fall 2014
Prof Michael Kearns
Due as hardcopy at the start of class, Tuesday September 30

1. (16 points) 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?

   (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?

   (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.

2. (8 points) 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?

3. (9 points) 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.

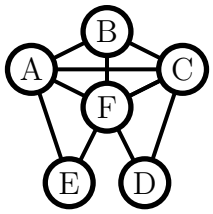
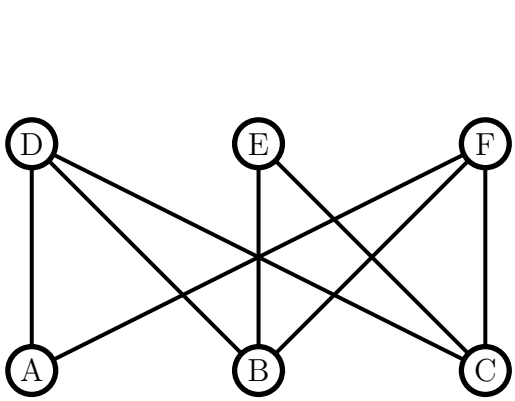


Figure 1: Problem 2 – Compute the equilibrium wealth for each node.

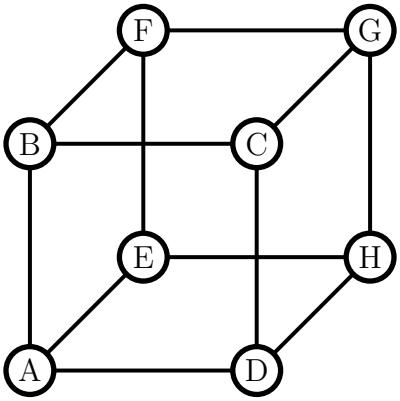
(a) (6 points) Give a real world example of a bipartite graph. Carefully describe what the vertices represent in your example, and the definition of which pairs of vertices are connected and what the edges represent in the real world. Feel free to be creative.

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

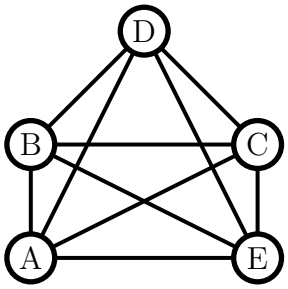
4. (15 points) Recall that a graph is *planar* if it has *some* (and not neceessarily any) visualization or layout that can be drawn in two dimensions in a way where no edges cross or touch each other. Determine whether each graph in Figure 2 is planar or not. If the graph is planar, give the visualization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



(a) 5 points



(b) 5 points



(c) 5 points

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

5. (10 points) 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!

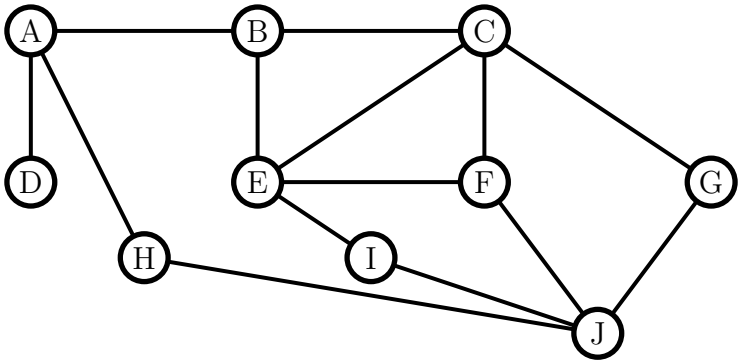


Figure 3: Problem 5 - Friendships on Facebook

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, 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?

(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?

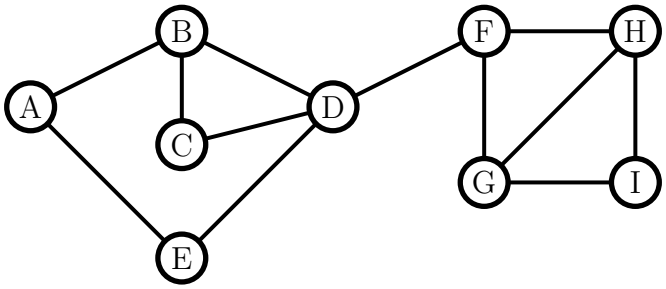


Figure 4: Problem 6

6. (12 points) 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.

7. (14 points) 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`).

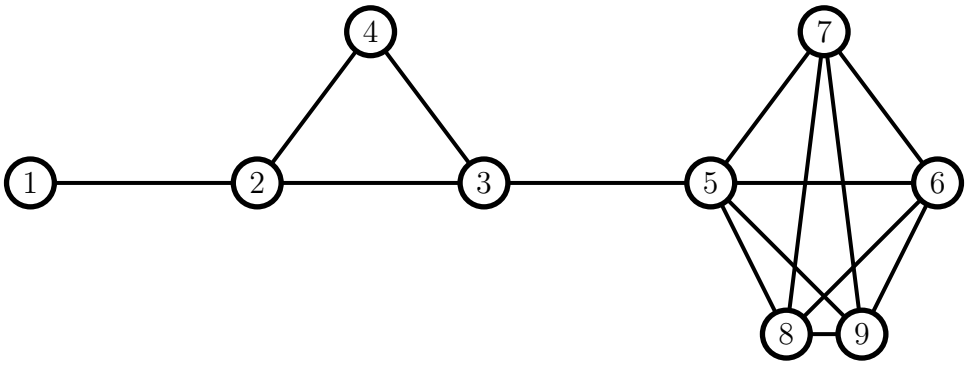


Figure 5

(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.)

(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?

(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.

8. (10 points) 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.

9. (6 points) 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.