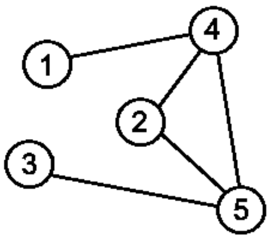
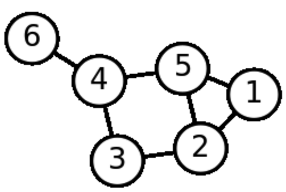
Networked Life (MKSE 112) Fall 2013
Professor Michael Kearns
Homework #1 Solutions

1. (5 points) Consider the network shown below. What is the diameter of this network? Recall that we have defined diameter to be the average shortest path distance between pairs of vertices.



There are 10 paths. The sum of their lengths is 16. So the diameter is 16/10 = 1.6.

2. (5 points) Consider the network shown below. In the altruistic contagion model, if each vertex starts with $1, what is the wealth of each vertex at equilibrium? Recall that, in this model, at each time step, each vertex divides its cash equally among its neighbors. As time goes to infinity, we reach an equilibrium in which the amount of cash each vertex receives is equal to the amount of cash it gives away.



A vertex *i* has wealth (degree(*i*) / sum of degrees) * total wealth. The sum of all degrees is 14. The total wealth is 6. For vertices (1) and (3), we have 2 * 6 / 14 = 6/7. For vertices (2), (4), and (5), we have 3 * 6 / 14 = 9/7. And for vertex 6, we have 1 * 6 / 14 = 3/7.

3. (6 points) Answer the following True/False questions. If the answer is False, give a brief one-sentence explanation.

a) (2 points) The largest possible number of edges in a network of N vertices is two times the largest possible number of edges in a network with N/2 vertices.

False. It's actually about four times. In the first case, the largest number of possible edges is about N^2/2, and in the second case, the largest number of possible edges is about N^2/8.

b) (2 points) Consider a network in which there are two vertices, A and B, that are connected by at least one path. Suppose we remove some edges at random from the network. The distance between vertices A and B may either stay the same or increase, but it cannot decrease.

True.

c) (2 points) Suppose we randomly add some edges to a network. It is possible that the diameter of the new network will be larger than the diameter of the old network.

False. The distance between two vertices can only decrease, not increase, when edges are added. So the diameter, which is the average distance, cannot increase.

4. (10 points) Suppose you are starting a new airline company, and you are considering two different designs for your flight map. We can model this map as a network in which each airport is a vertex, and there is an edge between two airports if there is a flight between them.

In the first design, there are a small number of airport "hubs," which have many connections to other airports. The remaining non-hub airports have a few connections chosen randomly. In the second design, there are no hubs, but each airport is connected to several other airports.

What are the advantages and disadvantages of each design? Consider factors such as the cost of the implementation, the length of a trip between two cities, the possibility of airport closings due to weather conditions, runway congestion, etc.

There are many reasonable answers here. You should have noted that the hub design results in small diameter, and therefore on average, there are shorter paths between cities. The cost of this design is probably lower, due to economies of scale. However, this design is vulnerable to airport closings. If a hub is closed due to terrorism threat, weather, etc, then many passengers will be stranded with no easy way to get to their destination. Furthermore, there is a lot of runway congestion at hub airports, since so many flights are routed through them.

5. (10 points) Consider the two contagion simulations discussed in class: the "forest fire" simulation and the "viral spread" simulation.

a) (5 points) Compare and contrast the network formation models of the two simulations.

In both models, we start with a k by k grid of vertices, in which each vertex has an edge to its four neighbors.

In the forest fire model, we delete a vertex and all of its edges with independent probability p. The deleted vertices represent parking lots, and the undeleted vertices represent forest.

In the virus model, we rewire each local edge with probability p to point to a random destination vertex. This creates a network with a mix of local and long-distance edges.

b) (5 points) Compare and contrast the contagion models of the two simulations.

In both models, we choose a random vertex at which to start the fire / infection.

In the forest fire model, the fire spreads deterministically from a vertex to its neighbors. The entire connected component of the starting vertex will burn.

In the virus model, a vertex passes the infection to each of its neighbors with probability *p*. Therefore, it is possible that not all of an infected vertex's neighbors become infected.

6. (5 points) Describe the navigation or "small world" problem. When solving this problem, what information is available to you, and what is not? Discuss the role of small diameter and "connector" individuals.

Many students were confused about the precise definition of the navigation / small world problem. The goal of the navigation problem is to *find the shortest path between two vertices in a network, using a local or decentralized* algorithm. Each vertex knows who its neighbors are, and it knows a little bit about the target, but it does not have a global or bird's eye view of the entire network structure. The network we are in *may* or *may not* have small diameter. If it does, then short paths exist in the network. This is necessary for solving the navigation problem. However, it is not sufficient, because we still need to be able to find these short paths using only local information. The existence of "connector" individuals is important, because these individuals reduce the diameter of a network, and therefore make it easier to solve the navigation problem.

The navigation problem is *not* Travers and Milgram or the Columbia Small Worlds Project. These are experiments that tried to solve the navigation problem on a "real world" network.

7. (5 points) In the Travers and Milgram experiment, as chain length increased, "location" was less frequently cited as an explanation for forwarding the letter, whereas "work" and "education" were more frequently cited. What are some possible reasons for this?

(First of all, my mistake – I meant the Columbia Small Worlds Project, as many of you figured out.)

At first, location is the most valuable piece of information for getting the message close to the target. If the source is in California and the target is in New York, a good first step is to send the letter to someone on the East Coast. But once the letter reaches geographic proximity to the target, it is no longer as useful to forward the message based on location. Now it is more appropriate to use other factors such as work and education, which help narrow down the search to the target individual.

8. (10 points) Consider the network in which each vertex is an actor, and there is an edge between two actors if they have starred in a movie together. An actor's *Bacon number* is the length of the shortest path between the actor and Kevin Bacon. Begin this exercise by going to oracleofbacon.org. This website provides an online tool for calculating Bacon numbers.

a) (2 points) Choose 5 actors, and calculate their Bacon numbers. What is the minimum, maximum, and average?

b) (2 points) Go to http://oracleofbacon.org/center.php. Read the page, and describe in your own words what it means for one actor to be a "better" center than another.

Actor A is a "better center" than actor B if the average "Actor A number" is less than the average "Actor B number." In other words, on average, the paths from a random actor to Actor A are shorter than the paths from a random actor to actor B.

c) (2 points) Go to oracleofbacon.org/onecenter.php. Enter your favorite actor or actress. What is the average __ number? (Replace __ with his or her name).

d) (4 points) From your exploration in this exercise, what information can you infer about this actor network? Consider properties such as diameter, average vertex degree, ease of navigation, etc.

There was a lot of confusion about this question. We did *not* mean "the network of the particular actor you chose in part c." We meant the general actor network. It is meaningless to compare one actor's network with another actor's network – the networks are all the same! Furthermore, they all have the same diameter. The diameter of a network is a fixed structural property, which can be calculated just like you did in question 1. It does not depend on which vertex you are looking at.

The answer to the question is as follows: the actor network has a small diameter because there are a lot of "connector" actors, like Kevin Bacon, who have high degree. Therefore, short paths exist in the network, and so it is easy to navigate. For instance, to get from one actor to another, we can usually just go through Kevin Bacon (or any other "connector" actor).