

## Midterm Solutions and Grading Guidelines

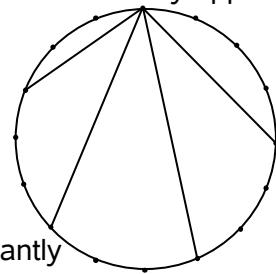
**Problem 1 (10 points)** For each lettered item on the left below, write the number of the item on the right that matches it best.

- |   |   |
|---|---|
| a) shortest path distance _18_          | 1) small changes have large effects     |
| b) worst-case diameter _12_             | 2) heavy tail                           |
| c) a disconnected network _13_          | 3) exponential decay away from the mean |
| d) degree of vertex _16_                | 4) opposite of Solaria                  |
| e) scaling laws of human travel _14_    | 5) persuasive                           |
| f) hub _20_                             | 6) PageRank                             |
| g) these tip _7_                        | 7) monotone properties                  |
| h) amplification of the incremental _1_ | 8) an unfortunate equilibrium           |
| i) connector _15_                       | 9) rich get richer                      |
| j) maven _19_                           | 10) fashion as epidemic                 |
| k) salesman _5_                         | 11) the original “random” network model |
| l) power law distribution _2_           | 12) length of the largest shortest path |
| m) Normal distribution _3_              | 13) Infinite worst case diameter        |
| n) Erdos-Renyi _11_                     | 14) Where's George                      |
| o) preferential attachment model _9_    | 15) many friends and acquaintances      |
| p) Caveman _4_                          | 16) number of network neighbors         |
| q) Paul Erdos _17_                      | 17) Kevin Bacon                         |
| r) holiday greeting cards _8_           | 18) fewest hops between vertices        |
| s) Hush Puppies _10_                    | 19) information specialist              |
| t) a random surfer model _6_            | 20) a vertex with very high degree      |

**Problem 2 (10 points)** Suppose we start with a network that is a simple cycle (ring) of  $N$  vertices. You are then allowed to add some fixed number of additional edges to the network. Feel free to illustrate your answers with diagrams.

- a. Briefly describe the pattern of edges you would add if the goal is to make the diameter of the resulting network as small as possible.

One solution is to pick a vertex and make it a hub by adding edges that connect that vertex to other vertices. The path connecting the two most distant vertices A and B would go from A to the nearest vertex with an edge to the hub, then the hub vertex, then the vertex nearest to B that is connected to the hub and finally B. Other acceptable solutions are to randomly add edges, or to add edges that connect diametrically opposite vertices.



5 points for a correct solution

3 points for a solution that decreases the diameter but not significantly

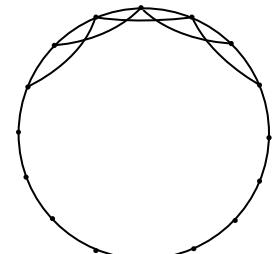
-2 points for inadequate or no explanation

-2 points for modifying the graph by adding vertices

-2 points for adding all possible edges to make the graph fully connected

- b. Briefly describe the pattern of edges you would add if the goal is to make the clustering coefficient of the resulting network as large as possible.

Add edges that will connect the vertices with the neighbors of their neighbors. Initially this means connecting every second vertex. If more edges are available, connect every third vertex and so on. The clustering coefficient will increase because the new edges introduce connections between the neighbors of a vertex.



5 points for connecting neighbors of neighbors

3 points for solutions that increase the clustering coefficient but not significantly

No points for just giving the definition of the clustering coefficient

-2 points for inadequate or no explanation

-1 point for unclear explanation

**Problem 3 (15 points)** Consider the “People You May Know” functionality on Facebook, which suggests friends to add to your network based on common friendships you already have. Discuss what effects you think the introduction of this service has had on the following structural properties of the Facebook network. Briefly justify your answers.

a. Degree distribution

The degree of every vertex will increase. Therefore the average degree will increase and the degree distribution will be shifted towards higher degrees. The shape of the distribution will not be affected significantly because the people who already have more friends are likely to get even more because the service can make more suggestions, whereas people with fewer friends will not receive as many suggestions.

5 points for saying the average degree will increase but the distribution shape will not be affected much

3 points for just saying the vertex degrees will increase but not discussing the effect on the distribution

-3 points if no justification was given or it was wrong

-1 point for confusing the degrees with the degree distribution

b. Clustering coefficient

The service will be suggesting connections between friends of yours that are not already connected, so your clustering coefficient will increase. The same applies to everyone so the clustering coefficient for the overall network will also increase.

5 points for saying the clustering coefficient will increase

-3 points if no justification was given or it was wrong

c. Diameter

Adding edges can only decrease the diameter of the network. Since the new edges are between people who already had a common friend, they will not provide huge shortcuts and the diameter will not decrease significantly.

5 points for saying the diameter will decrease and discussing how much

4 points for just saying the diameter will decrease

-3 points if no justification was given or it was wrong

**Problem 4 (10 points)** The abilities of Wall Street workers in primarily quantitative roles (a.k.a. “quants”) naturally vary considerably from individual to individual. Furthermore, if compensation on Wall Street degrades sufficiently (e.g. due to the current financial crisis), some quants may consider alternatives (going to grad school, joining a start-up, etc.). Apply the concept of the “Market for Lemons” discussed in Schelling to the hiring of quants on Wall Street. Be sure to discuss the asymmetry of information and the dynamics of the process, including the behavior and decisions of both quants and employers. Who leaves the market first, and who leaves last?

- While the employers only have limited information about their employees through the hiring process, the employees know how strong/weak their own quantitative skills are. An employee therefore knows if he or she is an above average quant, or if he or she corresponds to a “lemon” in the workplace.
- When the compensation decreases, the best quants realize they can get hired by other places like start-ups and get paid more for their skills. So, the best quants quit their Wall Street jobs and leave the market.
- It is now not worth it for new workers who have strong abilities to enter Wall Street, so only the relatively weaker potential quants apply for quant jobs.
- The employers realize this and know that the employees that are left are now not as good on average, and so lower the compensation even more.
- The better of the employees that are left now are undervalued and so leave the market. The cycle continues, with the proportion of bad quants getting larger and larger, and the compensation getting lower and lower, until only bad quants remain.
- When only bad quants are left, it is no longer worth it for Wall Street firms to employ quants at all, and so the market disappears.

10 points total

1 point for acknowledging that employees know their skills, while employers don't

3 points for recognizing the best quants leave first

2 points for realizing that employers will then lower compensation further

3 points for realizing that this process will spiral, like the market for lemons

1 point for realizing the bad quants leave the market last

**Problem 5 (15 points)** Consider networks in which each vertex has at most  $K$  neighbors, and every pair of vertices is at most  $D$  hops apart.

- a. Suppose  $K = 150$  and  $D = 6$ . Is it possible to have such a network whose size is equal to that of the U.S. population? Simply answer “Yes” or “No”.

Yes

3 points for writing “Yes”

- b. Why are  $K=150$  and  $D=6$  particularly interesting choices in light of course material and readings?

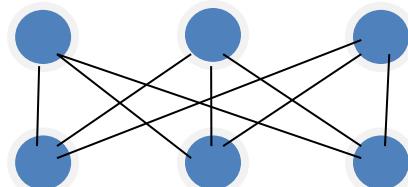
- $K=150$  is a plausible bound on the number of neighbors people have in real life social networks. Gladwell referred to it as “the magic number 150”, and it is believed that humans have evolved to be able to keep track of social relationships in group sizes of up to 150. In particular, experiments have found a relationship between the neocortex size in primates and their average group size.
- 6 is a plausible worst-case diameter for social networks. Travers and Milgram found that the letters which reached the target generally arrived after about six hops, leading to the belief in pop culture that everyone is reachable within “six degrees of separation”.

4 points for satisfactory explanation of 150, mentioning at least one of “magic number 150”, brain size in primates, or Goretex’s success in keeping group sizes below 150.

4 points for satisfactory explanation of 6, mentioning at least one of Travers and Milgram’s experiment, the book Six Degrees, “six degrees of separation”, or Kevin Bacon

- c. If  $N$  is the network size, describe how to construct networks of arbitrarily large  $N$  satisfying  $K=N/2$  and  $D=2$ . Feel free to illustrate with a digram.

The most natural answer is to construct a bipartite graph. Divide the vertices into two groups, so there are  $N/2$  vertices in each group. Each vertex has an edge to every other vertex in the opposite group, so each vertex has  $N/2$  neighbors. The distance between any two vertices in different groups is 1, and between any two vertices in the same group is 2, so it satisfies  $D=2$ . An example is drawn below:



2 points for describing (through words and/or through drawing) a network where each vertex has at most  $N/2$  neighbors

2 points for describing (through words and/or through drawing) a network where the worst case diameter is 2.

**Problem 6 (15 points)** This problem asks you to discuss the various models for network formation discussed in class and the readings.

- a. What assumptions does the Erdos-Renyi model make that are inappropriate for many networks? Discuss at least one specific naturally occurring network (social, biological, economic, organizational, etc.) and how the Erdos-Renyi model fails to represent it well.

The Erdos-Renyi model assumes that edges are formed independently and at random. One specific network that the Erdos-Renyi model fails to model well is the network of friendships on Facebook. In Erdos-Renyi, the expected value of the clustering coefficient of a vertex is the same as the baseline probability of edges. In the Facebook network, however, two people who have mutual friends are much more likely to be friends with each other than two people who do not have mutual friends. One often meets new friends by being introduced to them by an existing friend. In addition, people tend to become friends because of shared classes or activities, and thus all members of a group are more likely to be friends with each other than with an arbitrary person outside of the group.

2 points for mentioning at least one assumption of the Erdos-Renyi model (such as that edges are chosen at random, that edges are independent of each other, or that degrees can be modeled by a Poisson distribution)

1 point for mentioning a plausible naturally occurring network

1 point for describing a property of the chosen network

1 point for explaining that Erdos-Renyi doesn't have that property

- b. What was Watts' motivation for introducing the alpha model? Discuss how the alpha model is a better/worse/equally good representation for the network you mentioned above.

Watts introduced the alpha model to better model the high clustering he found in several naturally occurring networks. In the alpha model, the probability of including an edge between two vertices is not a constant, but instead is dependent on the number of neighbors they share. The alpha model is a better representation than Erdos-Renyi for the Facebook friendship network I discussed above, because it models the high clustering found in the friendship network.

2 points for saying that Watts introduced the alpha model to model high clustering

1 point for choosing a plausible choice of better/worse/equally good for the network chosen above (most likely, the choice was better)

2 points for discussing a property of your network that the alpha models that Erdos-Renyi doesn't (or vice-versa, if you're arguing for worse)

- c. What was the motivation for introducing the preferential attachment model? Discuss how the preferential attachment model is a better/worse/equally good representation for the network you mentioned above.

The preferential attachment model was introduced to model the heavy tail of degree distributions found in several naturally occurring networks. In many networks, there are a few vertices with degree much, much higher than the mean, which isn't reflected in either Erdos-Renyi or the alpha model. The preferential attachment model is therefore a better representation for the Facebook friends network, since it models the existence of Connectors who have several times more friends than the average person.

2 points for saying preferential attachment was introduced to model the heavy tail of degree distributions.

1 point for choosing a plausible choice of better/worse/equally good for the network chosen above (if you interpreted the comparison as being between Erdos-Renyi and preferential attachment, you probably chose better; if you interpreted the comparison as being between the alpha model and preferential attachment, you might have chosen better or worse or equally good).

2 points for discussing a property of your network that the preferential attachment model reflects but Erdos-Renyi (or the alpha model) does not.

**Problem 7 (10 points)** In the directed network below, the vertices represent web pages and the edges represent hyperlinks. In each of the following questions, you do not have to justify your answer --- simply answer each directly.

- a. Which vertex has the highest PageRank? Which has the lowest?

F has the highest because it has only incoming edges. A has the lowest because it has only outgoing edges.

Now add a new vertex labeled G which has 2 edges. One edge goes from G to A, and the other from G to B.

- b. Which of the vertices G and F has greater PageRank (or do they have the same)?

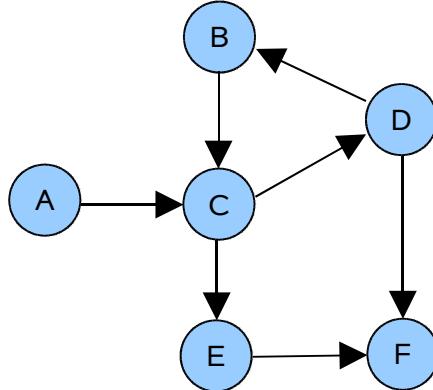
F has greater PageRank than G, because F has only incoming edges and G only outgoing.

When G and its edges are added, will the PageRank of the following vertices become higher, lower or remain the same?

- c. Vertex B

- d. Vertex C

- e. Vertex D



B, C and D will have higher PageRank. The new vertex G sends more weight "downstream", so the PageRank of A and B will increase and subsequently so will the PageRank of C and D.

2 points for every question. No explanation required.

**Problem 8 (15 points)** Imagine a distributed population of people playing games over a network in which each player controls their own vertex. In each game, players must choose a color for their vertex from a limited set of colors. At all times players can see their own current color and those of their network neighbors, but nothing else. Players can change their color at any time. Feel free to illustrate your answers with diagrams.

- a. Consider first a social coordination game in which there is a collective goal that every vertex in the network eventually chooses the same color simultaneously (unanimous consensus). Restricting your answer to connected networks, describe one network that you think should be relatively easy for this problem (in terms of the time it takes the population to reach consensus), and another network that you think should be relatively hard. Briefly justify your answers.

*If players cannot see the degrees of their neighbors, a long chain structured network should make consensus hard because each node can only see the colors of up to two other nodes and therefore has a very limited view of the network and because chains have maximum average diameter which makes coordination take a long time to propagate across the network.*

*Conversely, networks which look approximately like a large clique will make coordination easy since players can see what global trends are forming and synchronize to them.*

*In the case that players can see the degrees of their neighbors, then preferential attachment style networks will also make coordination easy since players with low degrees can play “follow the leader” and copy the color of their highest degree neighbor who is likely to have a more complete picture of the network than they do. Additionally, preferential attachment style networks have small diameter which allows coordination to propagate quickly.*

*+2.5 points for giving a network structure that makes coordination easy and for providing a reasonable justification of why it makes it easy.*

*+2.5 points for giving a network structure that makes coordination hard and for providing a reasonable justification of why it makes it hard.*

- b. Now consider a social differentiation game in which each player’s goal is to be a different color than all of its neighbors. Again restricting your answer to connected networks, describe one network that you think should be relatively easy for this problem (in terms of the time it takes the population to reach a point where everyone is a different color than all their neighbors), and another network that you think should be relatively hard. Briefly justify your answers.

*In the differentiation game, a chain structured network makes differentiation easiest since each player has at most two nodes to differentiate itself from, colorwise.*

*Conversely, highly entangled networks (e.g., a single large clique) will make players very constrained in their color choices and hence make differentiation hard.*

+2.5 points for giving a network structure that makes differentiation easy and for providing a reasonable justification of why it makes it easy.

+2.5 points for giving a network structure that makes differentiation hard and for providing a reasonable justification of why it makes it hard.

c. Consider both of the games above on networks generated according to the Erdos-Renyi and Preferential Attachment network formation models. For each game, do you think that game should be easy or hard on each formation model? Why? You may refer to your answers to parts a. and b. if desired

### **Erdos-Renyi Networks:**

*Erdos-Renyi networks with large “p” (i.e. dense networks) could make coordination easy since players will be able to see the colors of many nodes in the network and can therefore follow whatever trends are developing. Additionally, such networks have small diameter making coordination propagate quickly. Erdos-Renyi networks with large “p” could make differentiation hard since such networks will be very tangled and thus creating many constraints in terms of what colors are available to players.*

*Erdos-Renyi with small “p” should make coordination hard since all players will have a small number of neighbors and will therefore have a very limited view of the network. Additionally, such networks have large diameters, making coordination take a long time to propagate. Erdos-Renyi with small “p” should make differentiation easy since players will have a small number of numbers to differentiate themselves from.*

*+1.25 points for stating that Erdos-Renyi networks can make coordination easy if “p” was large due to density of edges or for stating that Erdos-Renyi networks can make coordination hard if “p” was large due to “over-cluttering” of edges.*

*+1.25 points for stating that Erdos-Renyi networks can make differentiation hard if “p” was large (due to density of edges) or for stating that Erdos-Renyi networks can make differentiation easy if “p” was small due to sparsity of edges.*

### **Preferential Attachment Networks:**

*In the case that players cannot see the degrees of their neighbors, preferential attachment networks could make consensus hard since most players have a small number of neighbors and thus have a limited picture of the network. At the same time, preferential attachment networks do have small diameter which could facilitate coordination. For the differentiation game, preferential attachment networks could make differentiation easier since most players in a PA network will have small degree and will therefore have few neighbors to differentiate themselves from. Additionally, such networks are tree structured and therefore have zero clustering. Consequently, they do not create tangled cycles of color constraints. The downside to PA networks for*

*differentiation is that they invariably possess a small number of nodes with very high degree who are extremely constrained in their color choices. Such overly-constrained vertices can become stuck in terms of what colors they can choose.*

*In the case that players can see the degrees of their neighbors, both coordination and differentiation might be easier on preferential attachment networks since players can defer to their high degree neighbors when making color choices.*

*+1.25 points for stating that Preferential attachment networks can make coordination easy due to hubs and small diameter.*

*+1.25 points for stating that Preferential attachment networks can make differentiation easy due to absence of clustering, or for stating that Preferential attachment networks can make differentiation hard due to overly constrained hubs with many neighbors.*