FINAL EXAMINATION
Networked Life (NETS 112)
December 17, 2018
Prof. Michael Kearns

This is a closed-book exam. You should have no material on your desk other than the exam itself and a pencil or pen. If you run out of room on a page, you may use the back, but be sure to indicate you have done so.

Name:

Problem 1: ________/10

Problem 2: ________/20

Problem 3: ________/15

Problem 4: ________/20

Problem 5: ________/20

Problem 6: ________/15

TOTAL: _________/100

Problem 1 (10 points). Answer "true" or "false" to each of the following assertions.

(a) k-anonymity prevents identifying individuals by combining databases.

(b) Romantic partners on Facebook tend to connect to multiple clusters in their partner's neighborhood.

(c) Every game has either a pure or mixed strategy Nash equilibrium, but not both.

(d) In Schelling's model, segregated cities must be the result of a racist population.

(e) The worst-case diameter of any connected network is always at least as large as its average-case diameter.

(f) The problem of scheduling final exams can be formulated as an instance of graph coloring.

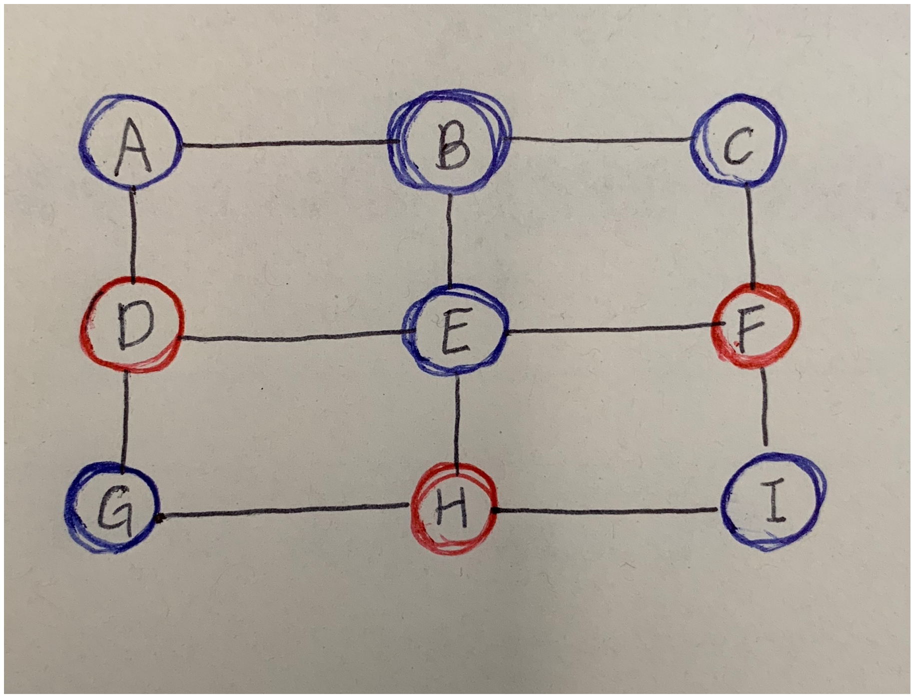
(g) According to "The Ethical Algorithm", the Singularity is near.

(h) The average driving time at Nash equilibrium is always the lowest possible.

(i) The distribution of file sizes on your laptop is heavy-tailed.

(j) Real-world networks tend to either have heavy-tailed degree distributions or to be highly clustered, but not both.

Problem 2 (20 points). Consider the network shown below.



a) What is the worst-case diameter of this network?

b) Add two edges to the network so that the worst-case diameter decreases by exactly 1. Clearly annotate them on the figure.

c) What is the clustering coefficient of the network? Use the table below to compute your answer.

Node	Clustering Coefficient	Node	Clustering Coefficient
A		F	
B		G	
C		H	
D		I	
E			

Overall Clustering Coefficient:

d) Each vertex is colored either red or blue. Assume a vertex receives a utility of 1 if at least half of its neighbors have the same color as the vertex, and 0 otherwise. What is the overall social welfare (sum of utilities)?

e) You are allowed to change the color of one vertex. Which vertex should it be, if you want to make the social welfare equal to 0?

f) You are allowed to change the color of one vertex. Which node should it be, if you want to increase the social welfare by 4?

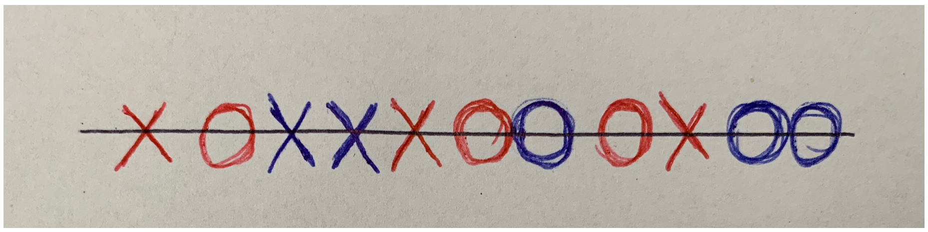
Problem 3 (15 points). In this problem you are asked to consider the chapter of "The Ethical Algorithm" that discusses the reproducibility crisis in the sciences, also known as the phenomenon of "p-hacking".

(a) Briefly but precisely describe what the underlying *quantitative* problem identified by the chapter is, i.e. what is p-hacking and what are the various technical reasons it can arise?

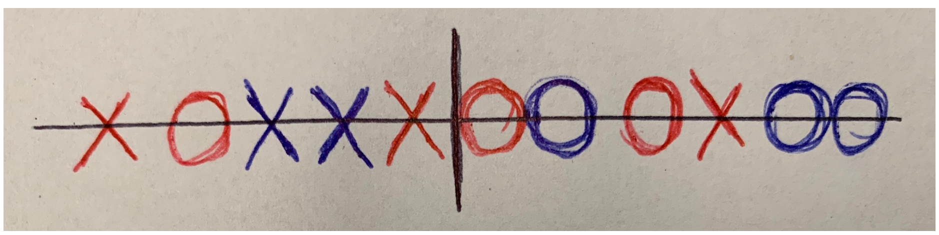
(b) Briefly but precisely describe some of the *social* forces in data-driven sciences that exacerbate the problem discussed in the chapter.

(c) Briefly but precisely describe at least two broad types of potential solutions to the problem identified in the chapter.

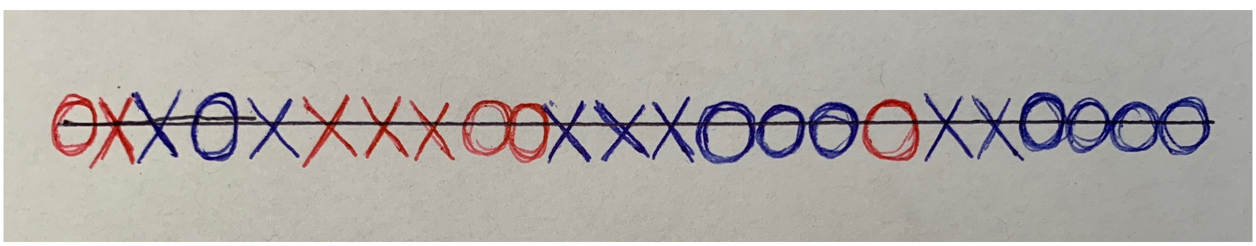
Problem 4 (20 points). You have been hired by Naïve College to build a model to determine which applicants should be admitted. The only data you have is the previous admitted students' SAT scores, their race (blue or red), and their collegiate success (i.e. whether or not they graduated from the college). Below is a diagram that visualizes such data. The color of the data point is the applicant's race; O and X represent whether the student graduated or not respectively; and where the data point is on the number line represents the SAT score. As we discussed in class and in "The Ethical Algorithm", your model will simply be a threshold SAT score --- i.e. a cut-off --- that will be used to decide who will be admitted to the school.



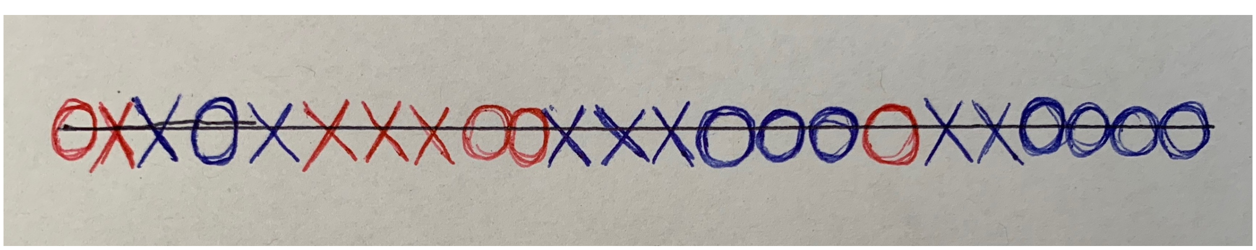
Below we show the optimal threshold (i.e. one that minimizes the error) given the data above. Note that the error in this case is 2: one student who was successful would not have been admitted according to this threshold (the red O to the left), and one student who was not successful would be admitted based on this threshold (the red X to the right).



The data you will work with is given below:



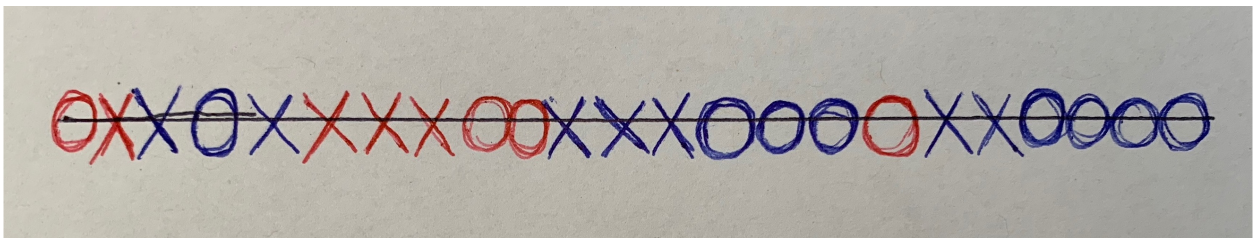
(a) In the copy of this figure below, clearly draw the optimal threshold for the data --- i.e. the cut-off that minimizes the error on the data.



(b) In the table below, for each race, state the false rejection rate and false acceptance rate according to the threshold you've drawn in part (a). Be sure to express you answers as rates or fractions, not as raw counts:

	False Rejection Rate	False Acceptance Rate
Blue		
Red		

(c) Now ignoring the blue points, what would have been the optimal threshold just for the red points? Clearly draw the line in the figure below.



(d) Using the threshold you gave for part (c), what are the false rejection and false acceptance rate for red people? Based on your answers in parts (b) and (d), do you think your original threshold in part (a) is "fair"? Why or why not?

	False Rejection Rate	False Acceptance Rate
Red		

Problem 5 (20 points). The *degree sequence* of a network is simply the list of degrees of the vertices in the network. For example, the degree sequence of the network in Problem 2 is (2,2,2,2,3,3,3,3,4). Note that the length of the list always equals the number of vertices, and the order of the numbers in the list doesn't matter. For each of the following degree sequences, either draw a network that has that degree sequence, or explain why no such network exists.

(a) (1,2,2,2,4)

(b) (2,2,2,3)

(c) (3,3,3,4)

(d) (1,1,2,2,2,2,3,3,3,3)

Problem 6 (15 points). The figure below is reproduced from one of the assigned readings.

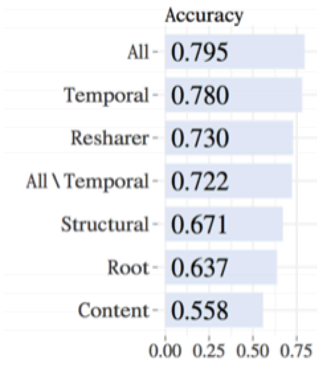


Figure 4: Using logistic regression, we are able to predict with near 80% accuracy whether the [redacted] will reach the [redacted] (10) after observing the first k =5 [redacted]

(a) For each of the colored redactions in the caption, fill in the missing word or phrase.

Blue:

Red:

Green:

(b) Briefly but precisely explain the experimental finding that the figure is describing, including the data being used and the question being asked.

(c) Briefly but precisely describe the meaning of the different categories labeling the bars in the figure, and the overarching point being made by the authors.