ISyE 4133 – Advanced Optimization – Spring 2022

March Madness Project, Due: April 24, 2022

Mathieu Dahan

Project Rules:

1. Each team should hand in an individual problem set on Canvas.
2. You may *not* interact with fellow students outside of your team when conducting your project.
3. The solutions and Python codes should be submitted electronically on Canvas prior to the deadline. Any handwritten solutions should be scanned so that they are easily read by the TA, and then submitted electronically. Typed solutions are preferred.
4. Show your work. Answers requiring a justification will not receive full grades if the justification is not provided. Similarly, answers requiring intermediate computations will not receive full grades if those computations are not shown in your solutions.

Problem Background

NCAA March Madness is one of the most popular college sports tournaments in the United States. Each year, 64 Division I college basketball teams compete for the title of national champion. (Actually, they start with 68 teams. Then, four games are played to get down to 64 teams.) While the teams are competing for fame and glory, many basketball fans are having their own competition simultaneously. Before the tournament takes place, millions of Americans create March Madness "brackets", trying to predict the outcomes of all 63 games and, ultimately, the tournament winner. There are 32 games played in the first round, 16 in the second, 8 in the third, 4 in the fourth, 2 in the fifth, and 1 championship game in the sixth round, for a total of 63 games.

A "bracket" is a way of specifying your prediction as to who the winner will be for each of the 63 games. The bracket is made before any of the 63 games are played. For each bracket, there is a score, which is determined as follows:

- 1 point is awarded for each correct prediction in the first round of competition.
- 2 points are awarded for each correct prediction in the second round.
- 4 points are awarded for each correct prediction in the third round.
- 8 points are awarded for each correct prediction in the fourth round.
- 16 points are awarded for each correct prediction in the fifth round.
- 32 points are awarded for each correct prediction in the sixth round. This team is also the winner of the March Madness Tournament.

While bracket predictions are often made for fun among friends, based off nothing more than educated guesses and personal preferences, others bet large amounts of money on their brackets. Indeed, a perfect March Madness bracket prediction could potentially earn its creator a very significant amount of money; Warren Buffett, for example, offers $1 million to any Berkshire Hathaway

employees who can produce a perfect bracket. For this reason, March Madness predictions have attracted the attention of mathematicians, engineers, sports analytics professionals, and many other academics seeking to solve this incredibly challenging problem.

In this project, we will provide a 64×64 matrix, where the entry in position i, j is the probability that team i beats team j, assuming they play against each other. The goal is to utilize optimization principles to build an "optimal" March Madness bracket, assuming that we start with the 64×64 probability matrix from the 2019 tournament. The initial configuration of the tournament is given in Figure 1.

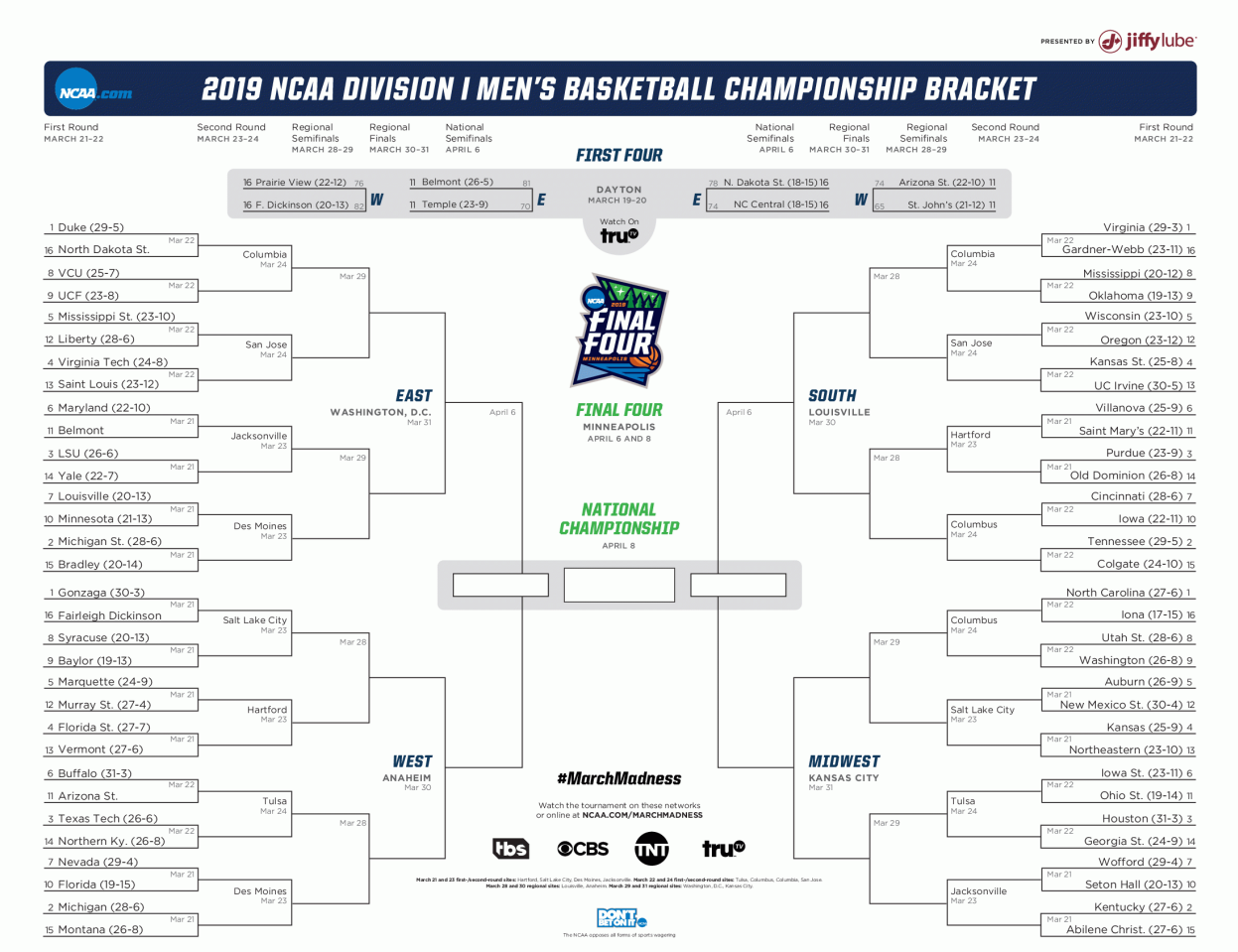

Figure 1: Initial configuration of the 2019 NCAA tournament.

By providing the 64×64 matrix, we have made the problem of determining the optimum bracket an easy problem. Or have we?...

Questions:

(a) (15 points) Determine the probability that each team wins round k (for $k \in [\![1, 6]\!]$). In determining these probabilities, use the 64×64 probability matrix regarding pre-tournament matchups.

(b) (15 points) Formulate an integer program to determine the bracket that maximizes the expected number of points awarded for each correct prediction. Use $\boldsymbol{x}$ to denote the binary variables. You will need to utilize the probabilities computed in part (a).

(c) (15 points) Code and solve your optimization model. What is the optimal bracket? What is the corresponding expected number of points?

- Tip: It will be helpful to first code and solve the problem with much fewer items to make sure that your code is correct.

(d) (5 points) How would your optimal bracket have performed in 2019? That is, how many points would you have gotten? The outcome of the tournament is given in Figure 2.

(e) (10 points) The goal of the remaining questions is to select other brackets to maximize our chances of having a better score for the 2019 data. To this end, we consider again the IP formulated in part (b). How can the IP be changed to obtain the bracket with the second highest expected number of points?

- Tip: Let $\boldsymbol{x}^*$ denote the optimal solution found in part (c). Formulate a constraint that renders $\boldsymbol{x}^*$ infeasible (but does not affect the other feasible solutions).
- Tip #2: Your solution should **not** involve the optimal value of the IP formulated in part (b), as two different optimal solutions might have the exact same optimal value.

(f) (5 points) Solve the new IP with the added constraint. How does the new bracket differ from the previous one? What is the corresponding expected number of points?

(g) (5 points) Repeat this process to obtain the ten best brackets in terms of expected number of points. What are the corresponding expected number of points, and what are the actual corresponding scores from the outcome of the tournament? Summarize your results in a table. You don't need to write down all ten brackets. What can you conclude?

(h) (10 points) What other techniques would you employ to "diversify" the brackets you obtain from the IP in part (b) and increase the chance of having one bracket with a higher actual score? *This is an open ended question. We ask you to investigate at least two ad-hoc methods, implement them, and analyze the results.*

(i) (20 points) Present your project. Your presentation should introduce the problem (for the students who have not worked on this project), and describe your results in a clean and organized manner.

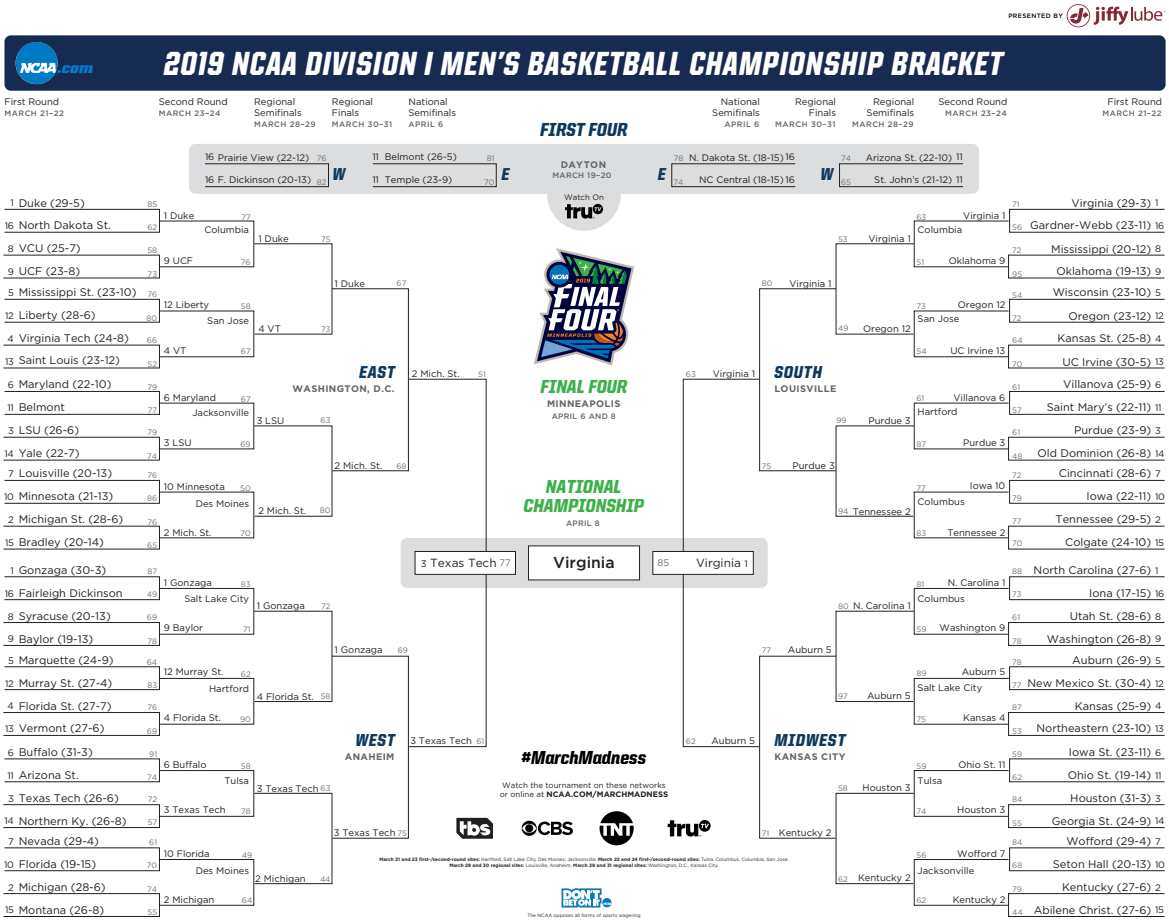

Figure 2: Outcome of the 2019 NCAA tournament.