Predicting NCAA March Madness 2019 with SpringRank

Brady Metherall
University of Oxford
(Dated: 16 March, 2020)

Each spring the National Collegiate Athletic Association (NCAA) hold a single-elimination tournament across the United States known as March Madness. The top 68 US college men's basketball teams in Division I take part, with the winner being crowned the national champion. The champions of each of the 32 Division I conferences automatically included, and the remaining 36 spots are determined by a selection committee. Since this process is done by a committee the teams given the final spots can be quite controversial. The tournament begins with four games that take the number of teams from 68 to 64, called the First Four, after which, the full 64 team single-elimination tournament begins. The First Four are comprised of the last teams selected by the selection committee. We shall examine using the network ranking algorithm SpringRank [1] to predict the outcome of March Madness games in 2019 based on the scores from the regular season.

To construct our network we use the box scores of the 5910 NCAA games from 2019 before March Madness began [2]. This includes games played in all three divisions of NCAA men's basketball—totalling 649 teams. From this, we construct the directed adjacency matrix, where each team is represented by a node, and each game is represented by a directed edge from the winner to the loser. We then compute the SpringRank from this matrix, and predict the team with the higher SpringRank will win the game. To compute the accuracy of this prediction we compare to the actual outcomes from March Madness 2019 [3].

The graph for the top 64 teams by SpringRank is depicted in Figure 1 with how far through the tournament they advanced. Eight of the top ten SpringRank teams also appear in the actual top ten seeds. SpringRank preforms well at predicting the teams that advanced to the round of 16, however, the performance decreases further down the ranking. Several teams that did not qualify for the tournament are predicted to be in the top 64 by SpringRank is due to the process of qualification. Any team that wins their conference automatically qualifies for March Madness, but, the relative strength of each conference varies significantly. Thus, weaker conferences must have a team automatically qualify, even though that team may be considerably worse than teams that did not qualify from much stronger conferences. Since SpringRank only ranks the relative strength of each team, many teams that were eliminated in the round of 64 are ranked below teams that were unable to qualify. Additionally, Temple and St. John's, whom lost in the First Four, were ranked very highly out of the teams that did not make it to the top 64.

Using SpringRank to predict the winner of each game yields accuracies of 23/32, 15/16, 4/8, 2/4, 1/2, 1/1 for each round, respectively, and a total accuracy of $46/63 \approx 73.0\%$. The only game SpringRank incorrectly predicted in the round of 32 was when Auburn (18) upset Kansas (7) by a score of 89–75; Auburn continued on to the Final Four. Another point of interest is that the number one ranked team in the country, Duke, only made it to the round of 8. Duke (1) played Michigan State (4), narrowly losing by a score of 67–68.

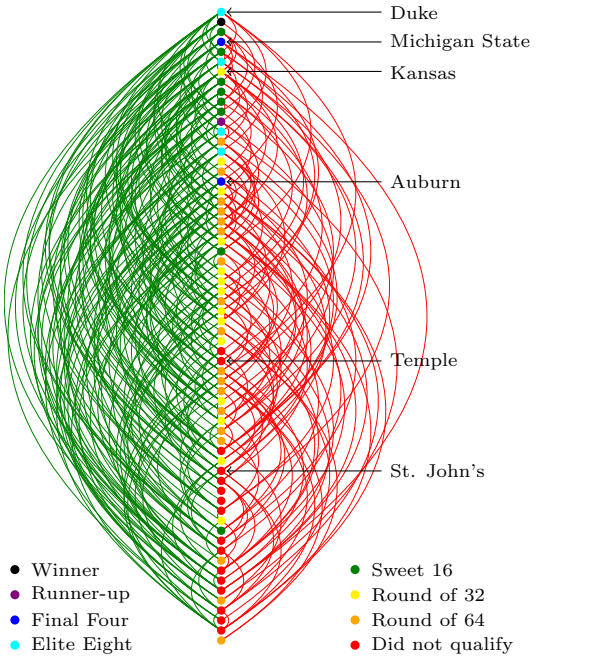

FIG. 1. Top 64 teams from SpringRank and their progress in March Madness 2019. A green (red) edge denotes a game that the higher (lower) ranked team won.

SpringRank correctly predicted the winner of 73.0% of the games during March Madness 2019—a large improvement over simply guessing the winner, and a 5% increase from assuming the team with the higher seeding will win. Moreover, SpringRank generally predicted how well the top teams would fare.

[1] C. D. Bacco, D. B. Larremore, and C. Moore, Science Advances **4** (2018), 10.1126/sciadv.aar8260.

[2] K. Wirth, GitHub Repository (Last commit Jan 2020),

github.com/kurtawirth/ncaahoopsscraper.

[3] D. Vanderkam, GitHub Repository (Last commit May 2019), github.com/danvk/march-madness-data.