Power-ful rankings of NCAA men's basketball teams

Hugo Hinckfuss, Josh Livingston, Magnus Miller (Section #002)

2024-04-29

This is the abstract

Computation

Data

We obtained team-game level data[1] from sports-reference.com[2] for the 2023-24 NCAA men's regular season. After doing basic data preparation, we bundled the data in an R package in this report. The R package, including the data and code for the computation, figures, and tables, is available on Github[3].

A sample of the data is available below. In total we have a 10,654 x 7 dataset. There are two rows per game in, one row per team. And for each team-game with have 7 columns: the team, date, home/away designation, opponent, result, team's score, and opponent's score.

Table 1: Sample of prepared NCAA men's basketball data

team	opp	result	team_score	opp_score
Texas State	Little Rock	L	66	71
Grand Canyon	Southeast Missouri State	W	88	67
Mississippi Valley State	Louisiana State	L	60	106
Southern	TCU	L	75	108
Jackson State	Memphis	L	77	94
Pacific	Sam Houston State	L	57	64
Southern Utah	Cal State Bakersfield	L	72	73

[1] https://docs.google.com/spreadsheets/d/1-KL_Ib_YSkrnA24nWCGpmx8Y35xYIG9VmKAH71I3ObU

[2] https://stathead.com/basketball/cbb/team-game-finder.cgi?request=1&comp_type=reg&game_status=1&order_by=date&match=team_game&year_max=2024&order_by_asc=1&timeframe=seasons&comp_id=NCAAM&year_min=2024

[3] https://github.com/joshwlivingston/appm3310.final

team	opp	result	team_score	opp_score
Tarleton State	Virginia	L	50	80
Bethune-Cookman	Minnesota	L	60	80
UT Arlington	Oral Roberts	W	75	71

Eigenvector computation

Win-loss matrix

We explore various methodologies to compute a_{ij}. First, we take the result of each game, and assign to the team a 0 for a loss, a 0.5 for a tie, and a 1 for a win. Then, we add the results and aggregate into a square matrix, denoted A_1 with entries $a_{1ij} : i, j = 1, 2, ..., 362$. A sample of this matrix is shown below. For simplicity, we'll let S_1 refer to the $4x4$ sample of A_1 with corresponding entries $s_{1ij} : i, j = 1, 2, 3, 4$.

Table 2: Sample of win-loss matrix

	Duke	NC State	Purdue	Tennessee
Duke	0	1	0	0
NC State	0	0	0	0
Purdue	0	0	0	1
Tennessee	0	1	0	0

The matrix entries $s_{1ij} \in S_1$ line up with team-game level results from the regular season.

Table 3: Regular season results for teams in the NCAA men's 2024 final four

team	date	home_away	opp	result	team_score	opp_score
Purdue	2023-11-21	NA	Tennessee	W	71	67
Tennessee	2023-12-16	NA	NC State	W	79	70
Duke	2024-03-04	away	NC State	W	79	64

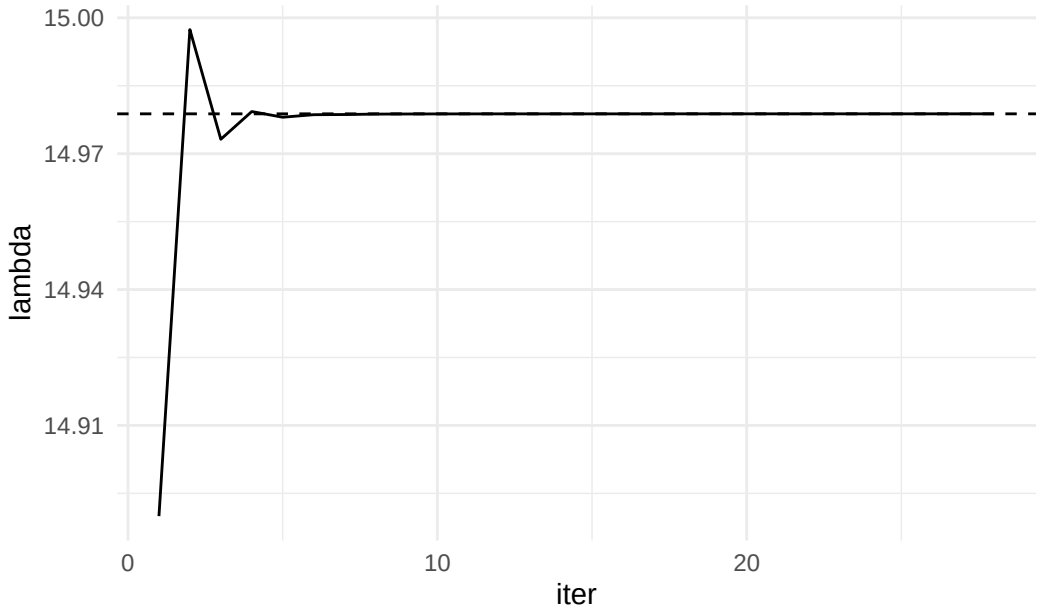
A potential downside to this method, is that teams that have never matched have the same entry for s_{ij} as teams that have lost to another team. For both of this scenarios, $s_{ij} = 0$. You can see that $s_{21} = s_{23} = 0$, where $s_{21} = 0$ represents NC State's loss to Duke, and $s_{23} = 0$ represents NC State and Purdue never having played a game against each other in the regular season So, in this scenario, we lose information about losses. This can lead to teams with losses to have an inflated ranking.

Eigenvalue approximation

We denote the ranking eigenvector r for matrix A_1 as r_1.

To approximate r_1, we employ the power method. At each step of the power method, we compute the approximate ranking vector λ_{1i} where i is the iteration of the power method approximation, and $\lim_{i \to \text{inf}} \lambda_{1i} = r_1$.

After approximation, we observe λ_{1i} converging to r_1, shown in the figure below.



Values of lambda converge to approximately 14.979 after 28 iterations

We use the approximated eigenvalue to compute the approximate eigenvector. We can use this eigenvector to compute rankings for NCAA men's basketball teams following the 2023-24 season. The rankings are shown below for the AP top 25 teams.

Table 4: Rankings for teams finishing in the AP top 10 as calculated by the win-loss method

team	eigenvector	rank
Michigan State	0.1000	1
Samford	0.0977	2
VCU	0.0929	3
Sacramento State	0.0895	4
Purdue	0.0873	5
Lindenwood	0.0862	6
Louisiana–Monroe	0.0859	7
Texas State	0.0846	8

team	eigenvector	rank
Eastern Illinois	0.0844	9
High Point	0.0843	10