# Effects of NBA Schedule on Game Results
Progress Report 1

Mitchell Jones
University of Oklahoma
CS-5483-995 Network Science
Dr. Sridhar Radhakrishnan
October 31, 2022

For my project, the effects of the National Basketball Association (NBA) schedule on game results is the problem that is being examined. As mentioned in the project proposal, a weighted directed network will be used with the team cities as nodes and path a team travels as the edges. This problem can be summarized by the following two questions. is there a certain number of rest days that is preferred over certain parts of the schedule? The first question being analyzed is "how does the number of rest days impact the number of games won and point differential over a defined period of games?" This will be examined over different periods of the season as well as different numbers of games. The periods being analyzed are for 1, 2, 3, and 5 games played as well as any other periods that trends may be found in. The second question is "how does traveling from a specific city to another impact the number of games won and point differential over a defined period of games?" The same periods outlined for the first question will be used on the second with the distance traveled being included as well.

At this time, many different algorithms and methods have been run to find insightful data, however the network structure used with the NetworkX library has been very limiting. The chosen graph type is a MultiDiGraph due to the fact that there can be multiple paths between two nodes with the same direction. Each of these paths must be represented with a separate edge. However, the NetworkX MultiDiGraph has very limited functionality with many different algorithms and methods in the NetworkX library. There is no insightful data from the

MultiDiGraph at this time, though I am in the process of converting the MultiDiGraph to a DiGraph by adding labels to the edge indicating if there would be multiples edges from the source to destination node in a MultiDiGraph. This will allow for better results that will be highlighted in the next project progress report. Below are two images of the preliminary network diagrams using all teams and just the OKC Thunder from the 2016-17 NBA season.
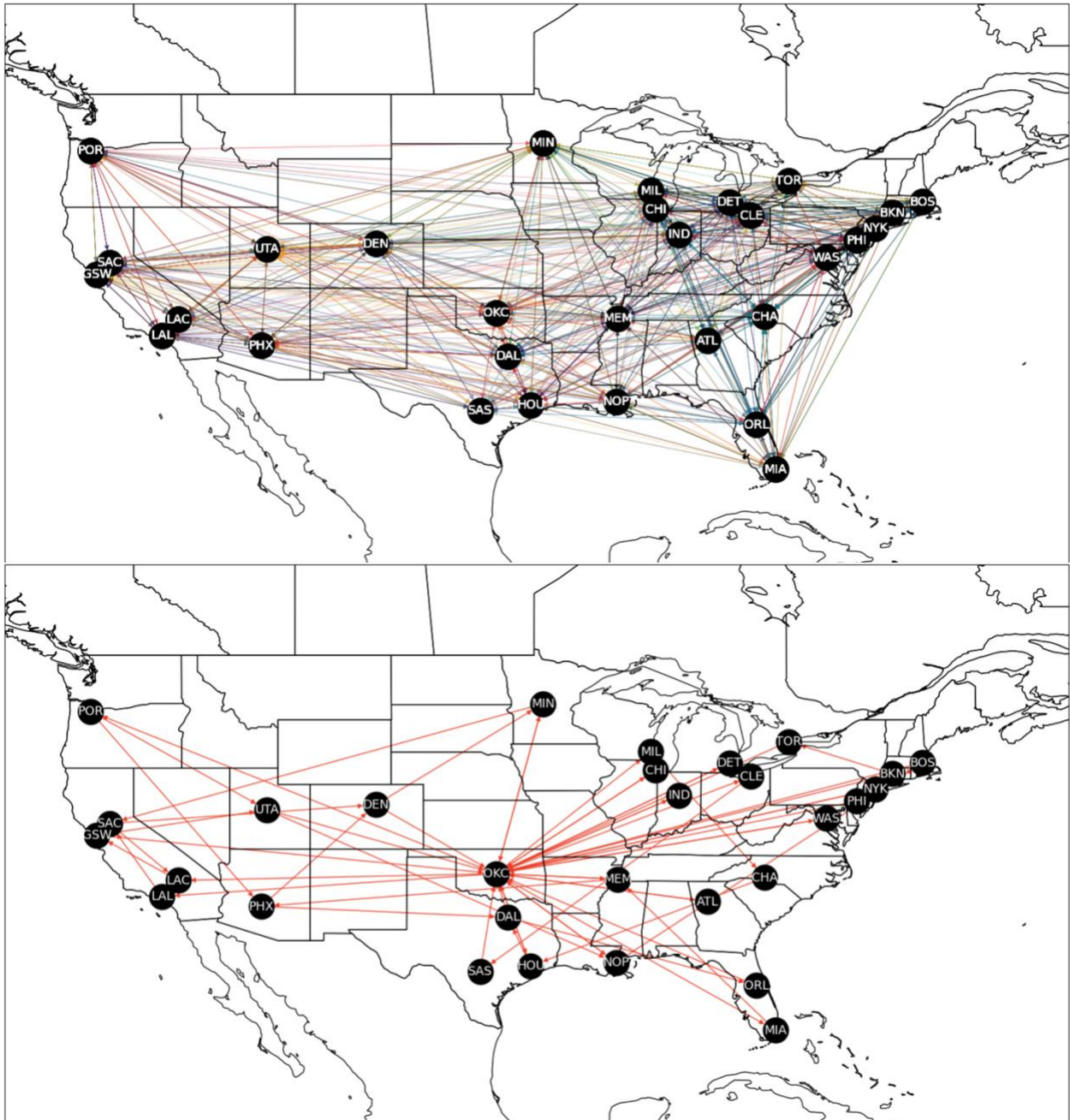
The final dataset being used for this project is in two parts. First, the data containing info on each game. As an example, the 2016-17 NBA season dataset is shown below.

	SEASON_ID	TEAM_ID	TEAM_ABBREVIATION	TEAM_NAME	GAME_ID	GAME_DATE	MATCHUP	WL	MIN	PTS	FGM	FGA	FG_PCT	FG3M	FG3A	FG3_PCT
0	22016	1610612739		CLE	Cleveland Cavaliers	0021600001	2016-10-25	CLE vs. NYK	W	241	117	45	94	0.479	13	35	0.371
1	22016	1610612752		NYK	New York Knicks	0021600001	2016-10-25	NYK @ CLE	L	240	88	32	87	0.368	9	27	0.333
2	22016	1610612762		UTA	Utah Jazz	0021600002	2016-10-25	UTA @ POR	L	241	104	40	82	0.488	8	24	0.333
3	22016	1610612757		POR	Portland Trail Blazers	0021600002	2016-10-25	POR vs. UTA	W	239	113	39	75	0.520	13	19	0.684
4	22016	1610612744		GSW	Golden State Warriors	0021600003	2016-10-25	GSW vs. SAS	L	241	100	40	85	0.471	7	33	0.212

FTM	FTA	FT_PCT	OREB	DREB	REB	AST	STL	BLK	TOV	PF	PLUS_MINUS	OPP_ABBREVIATION	NUMBER_GAMES_PLAYED	OPP_NAME	HOME_AWAY	HOME_TE
14	19	0.737	11	40	51	31	12	5	14	22	29.0	NYK	1	New York Knicks	HOME	
15	20	0.750	13	29	42	17	6	6	18	22	-29.0	CLE	1	Cleveland Cavaliers	AWAY	
16	16	1.000	6	25	31	19	9	5	11	19	-9.0	POR	1	Portland Trail Blazers	AWAY	
22	22	1.000	5	29	34	22	5	3	12	18	9.0	UTA	1	Utah Jazz	HOME	
13	18	0.722	8	27	35	24	11	6	16	19	-29.0	SAS	1	San Antonio Spurs	HOME	

NUMBER_GAMES_PLAYED	OPP_NAME	HOME_AWAY	HOME_TEAM_ABBREVIATION	AWAY_TEAM_ABBREVIATION	OFF_EFF	DEF_EFF
1	New York Knicks	HOME	CLE	NYK	0.894017	1.188636
1	Cleveland Cavaliers	AWAY	CLE	NYK	1.136364	0.854701
1	Portland Trail Blazers	AWAY	POR	UTA	0.898077	0.826549
1	Utah Jazz	HOME	POR	UTA	0.803540	0.873077
1	San Antonio Spurs	HOME	GSW	SAS	1.002000	0.776744

The various game statistical values (all values excluding SEASON_ID, TEAM_ID, TEAM_ABBREVIATION, TEAM_NAME, GAME_ID, GAME_DATE and MATCHUP) are all used in calculating possible edge weights for the directed graph. The values that are excluded from this calculation are used in identifying specific sections or processing of the data. The shape of the data for the 2016-17 NBA season is 2,460 rows with 36 columns. The column number will be the same for any season or subset of teams.

The second dataset used is info on the NBA teams. This data is a dictionary with the team abbreviation used as the key (the 3 letter team code) that contains the ID used by the API, the team city, nickname, full name, and latitude and longitude of the team's city. Below is what the data looks like for the OKC Thunder.

```
nba_team_details['OKC']
```
```
{'ID': 1610612760,
 'NICKNAME': 'Thunder',
 'CITY': 'Oklahoma City',
 'FULL_NAME': 'Oklahoma City Thunder',
 'LAT': '35.4729886',
 'LON': '-97.5170536'}
```

For this dataset, there are 30 keys each containing 6 values, for a total of 180 values.