Analyzing Basketball Games as Networks

Anirban Gupta
gupta.585@osu.edu

Hiroki Witt
witt.102@osu.edu

Meraj Khan
khan.485@osu.edu

ABSTRACT

The project reviews how network analysis concepts can be used to analyze and characterize team, and individual behaviors in team sports. We consider the case of Basketball in particular, and build up on analysis concepts from past work, and provide a user friendly analysis framework; which makes use of newly available SportVU logs from NBA, and play by play data for games.

1. INTRODUCTION

Statistical analysis of team sports in general has always focused on individual player performance, or team statistics as a unit. There has been very little effort in the area of capturing, and characterizing team interactions. However, recently, there have been efforts to analyze team sports such as Basketball, Soccer etc, by mapping teams and their interactions to networks, and the performance of players being impacted by their positions in the network. Our goal is to map a team's interaction during basketball to a directed network, and calculate metrics and measures in order to analyze team and individual performances.

Analyzing group interactions is a major application area of network analysis. A good network representation of the group structure should highlight the group interactions, and function of individuals. We can extend this idea to team sports. Teams can be defined as groups of individuals working collaboratively and in a coordinated manner towards a common goal be it winning a game, increasing productivity, or increasing a common good [4].

Basketball games are a series of interactions over the entire game period. Players work together towards a unified goal of moving the ball into the basket, while trying to resist the opposition to do so. Our project is a strict analysis of offensive plays only. For each game, we model the interactions between a team in isolation, and calculate network measures over the resultant graph.

To evaluate basketball teams as networks, we picked 3 OSU college basketball games and annotated the interactions between OSU players for their offensive plays. This was done to see if the measures we include in our framework for analysis actually show results consistent with what's observed in the game.

The main contribution of this project is a data processing framework that generates offensive play networks from 'NBA SportVU logs', and 'play-by-play logs'. This enables users to analyze performance of any team for a given match, if NBA logs for it are available.

We also provide an interactive UI for users to pick games for analysis. The interface provides different network specific, and node specific measures. The users can rearrange the node layout to observe structural patterns of interest.

2. PROBLEM STATEMENT

The first part of the problem can be formulated as construction of a weight-ed directed graph *G(V, E)* from player position tracking logs, and play-by-play logs, where each node v in V corresponds to a player, possible game outcomes, or different starts of offensive plays. Presence of an edge between nodes indicates the interaction between them, between players it denotes passes, and nodes from start nodes to player nodes show which player started that play. Edges from player nodes to end nodes show the end of offensive plays from scoring attempts, or loss of possession because of any reason. The edges between any two nodes are weighted according to the number of interactions throughout the game, higher the count, higher the edge weight.

The network analysis problem can be formulated as an adjusted calculation of common network/node specific measures, and analyzing them to come up with insights about team strategies, player performance, and player roles.

3. METHODOLOGY

3.1 Network Construction

Games observed and analyzed, could be categorised into two types, College basketball games, and NBA games. Graphs for the NBA games are created using the SportsVU logs and Play by Play logs, while graphs for the college games are created using annotated data.

3.1.1 SportVU logs

SportVU[1] utilizes a six-camera system installed in basketball arenas to track the real-time positions of players and the ball 25 times per second. Thus, the granularity of this data is every 0.04 seconds. Time frames are ignored, where information about the position of the ball is absent. Also,

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

due to the existence of multiple positional information of players and ball at an instance of time, only the first ball and player positions in that time instant is considered for simplicity.

3.1.2 Play by play logs

Play by play data, for each NBA game, is available, on the NBA stats site. However, the granularity of this data is in seconds, unlike the SportsVU logs. If play by play events occur in time t, they are assumed to occur in the time window $t+1, t$, as time counts down in basketball.

Some of the fields for play by play events are

- GAME_ID: Game id stored for the match. Used to extract the relevant Sports VU logs.
- EVENTMSGTYPE: Event type. Eg: A value of 10 indicates jump-ball, 1 indicates a successful shot, 2 indicates a miss etc.
- EVENTMSGACTIONTYPE: Sub-event type. Eg: For event freethrows, the value of EVENTMSGTYPE is 3. The first free throw out of 2 free throws is categorized further by giving a value of 11 to EVENTMSGACTIONTYPE.
- PERIOD: Indicates the quarter.
- HOMEDESCRIPTION: Description of the event, if the event occurred with respect to the home team.
- VISITORDESCRIPTION: Description of the event, if the event occurred with respect to the away team.

3.1.3 NBA network construction

Algorithm 1 Play by play network construction algorithm

procedure PLAY_BY_PLAY_CONSTRUCTOR( )
Convert time to seconds in play by play logs.
Extract player and ball position once for every time instant, from SportsVU logs.
Extract team ids and team information.
Extract events from play by play data.
Determine possession from each event, by calling POSESSION_DETERMINER(event e, team_id home, team_id away)
Insert possession events for start and end of quarter, if not present.
Divide play by play possession events, into possession time frames, where every frame indicates possession of ball by one team. This creates a sequential path of nodes for the ball from start of play to players and finally end of play.
Now for every play by play frame f, perform MOMENT_INTERPOLATOR(f). Append the resultant list of nodes to global list NODES.
end procedure

Nodes are defined to be of three types

- Start of play: Would comprise of possible start of play event/nodes.
  - 0 - inbound for home, when possession is gained by the home team under other circumstances, besides steals and rebounds
  - 1 - inbound for away, when possession is gained by the home team under other circumstances, besides steals and rebounds
  - -4 - steal, when the ball is stolen
  - -5 - rebound, when a rebound event occurs
- End of play: Would comprise of possible end of play event/nodes
  - -2 - successful shot
  - -3 - missed shot
  - -6 - turnovers
  - -1 - fouls, violations, other events where play ends, such as ball going out
- Player nodes: Would comprise of player ids

Algorithm 2 Possession determining algorithm for each event

procedure POSSESSION_DETERMINER(event e, team_id home, team_id away)
Determine type of event using EVENTMSGTYPE and EVENTMSGACTIONTYPE
If $event == jumpball$, chain of nodes are [0, $player_1$'s id, $player_2$'s id], where $player_1$ and $player_2$ are the players who won the jumpball, and then got the ball respectively.
If $event == successful\,shot$ or $event == free\,throws\,(2\,out\,of\,2)$, chain of nodes are [$player_1$'s id(optional), $player_2$'s id, -2, inbound node], where $player_1$ provides the assist, and $player_2$ took the shot, inbound node is 0 if the away team took the shot, and vice versa. (Here 2 out of 2 implies the second free throw out of 2)
If $event == rebound$, chain of nodes are [-5, $player_1$'s id], where $player_1$'s id is the player id who got the rebound.
If $event == fouls$ or $event == violations$, chain of nodes are [-1, inbound], where inbound node is 0, if away team committed the foul/violation, and vice versa.
If $event == turnovers$, chain of nodes are [$player_1$'s id, -6, -4, $player_2$'s id], as steals occur with turnovers. However, depending on the sub-event type, this could change to [$player_1$'s id, -6, inbound node], if turnover happened due to a reason besides a steal, where inbound node is 1, if home team lost possession, and vice versa.
end procedure

Algorithm 3 Moments interpolation algorithm

procedure MOMENT_INTERPOLATOR(frame f)
Determine closest player to the ball at every time instant in frame f, from SportsVU logs.
Consider only series of time instants, where closest player distance is less than or equal to 1.1, and closest player id is the same for atleast 5 frames, i.e, for a span of 0.20 seconds.
Missing players for frame f are interpolated from the the above series of time instants, in order to create a proper sequential path of the ball in time frame f.
end procedure

Algorithm 4 Final graph construction algorithm

1: **procedure** NBA_GRAPH_CONSTRUCTOR(NODES)
2: List of start_nodes = [0, 1, -4, -5]
   List of end_nodes = [-2, -3, -6, -1]
3: **for** each consequent pair of nodes (n1, n2)) in NODES
   if node n1 in end_nodes, skip this pair.
   if edge exists from node n1 to n2, increase weight of the edge by 1, **else** add edge from node n1 to n2 in the directed graph G.
4: **end for**
5: **end procedure**

3.1.4 College basketball annotated logs

We also aimed to analyze OSU college basketball games in this season, and compare and contrast with NBA games. Due to difficulty of retrieving game footage, only three games were annotated. The Ohio State data was captured in the form of 3 games. In an attempt to obtain an idea of when the Ohio State basketball team was playing at their best and at their worst, data from their two biggest wins of the year, against the University of Kentucky and the University of Iowa, were identified, as well as one of their worst losses, a 19-point home loss to Michigan State. The data from these games was manually tracked, and actions such as inbound start of play, successful shots, missed shots, turnovers etc, were noted.

3.1.5 College basketball network construction

A networkX graph G was generated for the three college games, based on the fact that, end of play nodes such as successful shots etc cannot have outgoing edges, and incoming play of nodes can have only outgoing edges.

3.2 Network Measures

3.2.1 Entropy

In network science, entropy is the measure of the randomness of a graph. In the context of our modeled graph, entropy refers to the unpredictability of ball movement amongst players or towards different outcomes. We exclude the start of play links from entropy calculations, as generally they are high weighted edges to certain fixed players, and do not indicate anything about the ball movement during the play. We use Shannon's entropy[3] to calculate the unpredictability of ball movements between different players and outcomes.

$$S = -\sum_{p \varepsilon P} plog(p) \quad (1)$$

3.2.2 Degree Centrality

Degree centrality is a measure of how "central" each node is the graph, or in terms of a basketball team, how often the ball flows through a particular player. Normalized degree centrality was calculated for each player node by normalizing the degree of each node against the sum total of degree of all nodes. The relative distributions of player degrees were then calculated across the graph. For a weighted graph with weights summing to 1 and a vertex of maximal degree the degree centrality is then:

$$DC = -\sum_{v \varepsilon V}(deg(v^*) - deg(v)/(|V| - 1)) \quad (2)$$

where v_* is the node with maximum degree centrality.

3.2.3 Clustering Coefficient

Clustering coefficient measures the extent to which the nodes in the graph tend to cluster. The local clustering coefficient of a node in a graph quantifies how close its neighbors are to being a clique (complete graph). We first convert the directed weighted graph to an undirected graph, and then threshold the edges at a certain fraction which is a user parameter. Thresholding essentially would retain only the top specified fraction of outgoing edges from each node. This is required as most of the player nodes would be connected to each other by the end of the game, and the edge weights do not play a role in calculation of the clustering coefficient otherwise.

3.2.4 Uphill Downhill Flux

We use the metric developed by Fewell et al [2] to measure a teams ability to move the ball towards their better shooters. A high positive flux shows that the players consistently try to move the ball into the hands of their players with higher shot hit rate. Flux can be calculated as.

$$F = \sum_{i \neq j} p_{ij}(x_j - x_i) \quad (3)$$

where p_{ij} denotes the transition probability from player i to player j, and x_i denotes the shot rate of player i.

3.3 Visualization and Analysis Interface

We preprocess all generated graphs to calculate the network and node specific measures. The measures and network structure is dumped in json files which is used for visualization, and analysis of the games on a browser based interface. The user can select a game from list of available games. The interface provides a clean, interactive visualization of the offensive plays for selected game. The team/network specific stats are displayed in one pane. Player specific measures can be seen by clicking on the node corresponding to specific player. The nodes are labeled with player names, and the edge thickness, and opacity encode the edge weight. Users can move around nodes to see occluded edges, or just explore patterns. Figure 1 shows the interface with a game, and a particular player selected.

4. EXPERIMENTAL RESULTS

The above mentioned data is used to construct graphs, one per team, and obtain measures such as entropy and degree centrality from these graphs. These values would be used as a baseline to compare with the Ohio State menâĂŹs basketball team.

4.1 Entropy

4.1.1 ENTROPY VALUES

In the two games that were studied, the Ohio State team had an entropy value of 5.568 and 5.879, respectively. Compared with the baseline values studied in the NBA dataset, these values were much higher; the highest entropy value from the NBA dataset was around 3.2.

4.1.2 OHIO STATE WINS

Ohio State is a team that employs about a 9-man rotation (at most) and does not have a traditional "go-to-scorer"

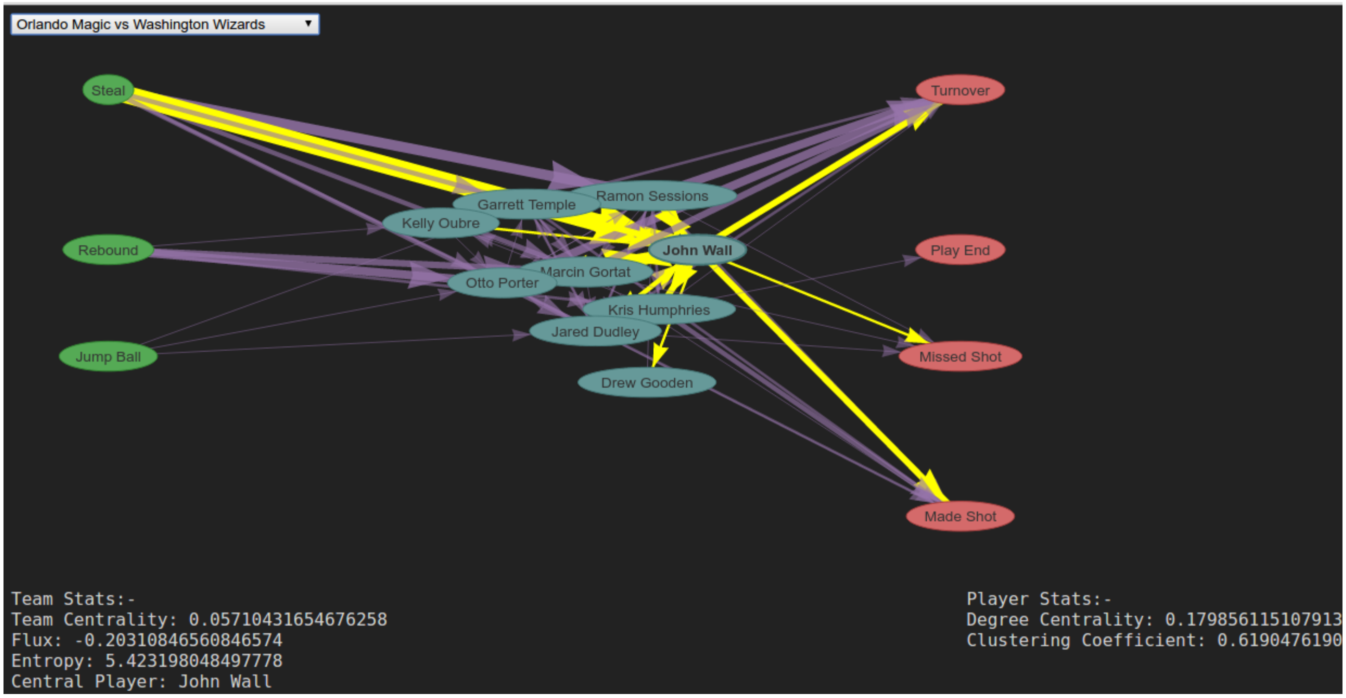

Figure 1: Screenshot of Visualization and Analysis Interface

Table 1: Entropy values of observed teams

Team	Entropy Value in observed game
Washington Wizards	5.423
Miami Heat	5.457
Toronto Raptors	5.403
Chicago Bulls	5.647
Los Angeles Lakers	5.592
Boston Celtics	5.322
Charlotte Hornets	5.291
Cleveland Cavaliers	5.799
Ohio State vs. Kentucky	5.529
Ohio State vs. Michigan State	5.350
Ohio State vs. Iowa	5.279

that many collegiate teams employ. In addition, many of the players that play in the rotation - *Marc Loving, JaeSean Tate, JaQuan Lyle, Keita Bates-Diop*, etc., are not one-dimensional players that are specialized in one area, rather their skills are spread out, and are all capable of shooting from long distance and get to the basket (albeit not necessarily at an exceptional rate). Due to the fact that the team employs a large rotation, it does not have a primary scorer that the offense is facilitated through, and that the players are not specialized in their skills, it makes sense that the team's entropy would be high.

What is more interesting is whether success can be attributed to this high entropy value - in both cases where Ohio State has won, they had high entropy values, and got strong play from players outside of their starting 5. A high entropy value implies that they were attacking the defense with a variety of different shots and moves, and that they kept the defense guessing. With that being said, it may sometimes imply that the team cannot find a rhythm doing any one set play, so they resort to shuffling different players in and out to find some sort of offensive rhythm. From this data, however, it seems that success generally coincides with higher entropy for this team.

4.1.3 OHIO STATE LOSS

In the large loss that was explored against Michigan State, the entropy was calculated to be 5.71, sandwiched between the values of the two losses. This shows that entropy is not necessarily correlated with wins or losses with this OSU team and the high entropy values can be attributed to having many players that contribute at roughly the same clip together rather than a few individuals. In this particular game, the entropy value could be attributed to Ohio State shuffling players in and out in an attempt to jump-start their offense.

4.1.4 COMPARISON WITH NBA

The values for entropy were almost uniform throughout

Table 2: Highest Degree centrality observed in games

Team	Highest Observed Degree Centrality	Player Name
Washington Wizards	.180	John Wall
Miami Heat	.149	Chris Bosh
Toronto Raptors	.214	Kyle Lowry
Chicago Bulls	.114	Jimmy Butler
Los Angeles Lakers	.146	Jordan Clarkson
Boston Celtics	.102	Isaiah Thomas
Charlotte Hornets	.186	Kemba Walker
Cleveland Cavaliers	.122	Kyrie Irving
Ohio State vs. Kentucky	.173	JaQuan Lyle
Ohio State vs. Michigan State	.244	JaQuan Lyle
Ohio State vs. Iowa	.187	Marc Loving

the NBA data-set as well, at values that were nearly the same as the ones calculated for the Ohio State games, which warrants our previous assumption.

4.2 Degree Centrality

4.2.1 OHIO STATE WINS

This Ohio State team is traditionally a team that spreads the usage of the ball around and struggled to find a primary ball-handler for much of the season, so it would make sense that the degree centralities of the players would not vary much between players. In the first game that happened, the Kentucky game, this was the case; *JaQuan Lyle* led the team with a degree centrality of .16847 but there were 5 OSU players with a degree centrality greater than .1. This suggests that the OSU team was able to keep the Kentucky team off-balance by spreading the ball around, as well as keeping the ball moving through a variety of people.

Similarly, in the Iowa game the team had 5 players with a .1 degree centrality or higher, although *Marc Loving*, who played a brilliant game scoring 25 points, led the team with a .18 degree centrality, while *JaQuan Lyle*, who was right behind him with a .1648 degree centrality, did not score. This shows that a player can have scored a lot of points but not necessarily have a huge hand in developing the shots themselves, and vice-versa; if a player has scored a ton of points but has a very low degree centrality, this shows that their points were more a result of his teammates facilitating offense for that person rather than from that player facilitating for themselves. From this, we can see that degree centrality can be a powerful tool to actually measure how much a player does for their offense, and a more accurate metric than simply points scored. One more important point is that in both sets, *Kam Williams*, who plays off the bench, had the 5th highest degree centrality. This is significant in that the NBA data neglected bench players, while for a team like OSU, it is important to look at the bench as they are a key in how the team operates.

4.2.2 OHIO STATE LOSS

In the game where OSU lost decisively, *JaQuan Lyle* again led the team in degree centrality, but with the highest observed value seen, at .234. The next closest player was *Marc Loving* at .169, and *Keita Bates-Diop* was the only other OSU player with a degree centrality of over .1 at .13. In contrast to the games they won where they had 5 players with a degree centrality of over .1, they only had 3. This could suggest that rather than spreading the ball around evenly in the Michigan State game, the ball was primarily handled by a few players, and also suggests that Ohio State performs better against teams when they can evenly spread the ball around all the players that are on the court, instead of keeping the ball in the hands of just a couple players. More games where OSU lost would need to be observed, however, as Michigan State is a very good defensive team and may have forced OSU to keep the ball in the hands of their inexperienced point guard.

4.2.3 COMPARISON WITH NBA

JaQuan Lyle's observed involvement in the offense through degree centrality was further evidenced through the NBA dataset, where he had a higher degree-centrality value than all NBA player's, with *Kyle Lowry* of the Raptors being the closest at .214. While part of the difference is due to the differences between the college and pro-games - in the pro game the team plays more players, and almost no players play the entire game, capping the degree centrality, Lyle's high degree centrality is an evidence of the long-leash he was given by coach *Thad Matta* as a Freshman, and as the only traditional Point Guard on the team, it is not a stretch to say that the Ohio State team will go as *JaQuan Lyle* goes in the future.

4.3 Clustering Coefficient

4.3.1 OHIO STATE WINS

In both games that OSU won, the player with the highest clustering coefficient was *Trevor Thompson*, the sophomore transfer. This shows that Thompson would never try to generate offense in isolation, but only as a result of a teammate, or dishing it off to another teammate. General intuition would lead one to believe that this is a good thing; the player never tries to do too much, and is always in connection with his teammates in terms of ball-movement. As a player that had just transferred schools, it is also encouraging that he was able to mesh so quickly with his new teammates, and work so well within the post. In terms of being able to find something concrete about his impact, this metric could be used in conjunction with another, such as offensive usage or FG%, to find out how effective he is actually being in his touches.

If these other metrics are in line with the clustering coefficient, it may be a sign that the team should work more to operate the offense through that player. What is also in-

Table 3: Highest observed Clustering Coefficient in games

Team	Highest Observed Clustering Coefficient
Washington Wizards	.800 (Otto Porter, Garrett Temple)
Miami Heat	.667 (Tyler Johnson)
Toronto Raptors	.833 (DeMar Derozan, Jonas Valanciunas)
Chicago Bulls	.7 (Taj Gibson)
Los Angeles Lakers	.833 (Bass, Clarkson, Young)
Boston Celtics	.667 (Isaiah Thomas, Avery Bradley)
Charlotte Hornets	.667 (Troy Daniels, Tyler Hansbrough)
Cleveland Cavaliers	.7 (Matthew Dellavedova, Timofey Mozgov)
Ohio State vs. Kentucky	.881 (Trevor Thompson)
Ohio State vs. Michigan State	.929 (Mickey Mitchell)
Ohio State vs. Iowa	.842 (Trevor Thompson)

teresting here is that *JaQuan Lyle*, the talented freshman, had among the lowest clustering coefficients on the team for both games, showing that he was not always in connection with his teammates. However, his team won both games against tough opponents in which he did this, perhaps suggesting that *Lyle* is at his best when he is either trying to create offense for himself, or not touching the ball, rather than facilitating for others.

4.3.2 OHIO STATE LOSS

In looking at the clustering coefficients for the loss to Michigan State, freshman *Mickey Mitchell* led the team with a .9286 clustering coefficient, which makes sense due to the newcomer's passing acumen and inability to score. In contrast to the two games that Ohio State lost, *Trevor Thompson* was 5^{th} on the team in clustering coefficient value for this game. This suggests that *Thompson* wasn't nearly as connected with his teammates in this game as he was in the other two, and thus implies that OSU has more success when they are able to operate through the post with *Thompson*; even if he wasn't scoring, just allowing him to play in the post and kick the ball back out to teammates proved successful in the two games they won, but was not on display here.

4.3.3 COMPARISON WITH NBA

John Wall, who plays for the Washington Wizards had 24 pts and 13 assists, in the game we observed. However, he's only 5^{th} on team in clustering coefficient. Thus, assists do not imply a lot of ball movement from the player. Ignoring the game, *Wall* is still 6^{th} in his team, in terms of clustering coefficients. However, the NBA dataset was a bit hard to relate to the college basketball data, as the clustering coefficients were very similar between games, and tended to produce results that said that the players that played very little had the highest clustering coefficient numbers (perhaps buoyed by them only interacting with a few players, and being very close to those players).

4.4 Uphill-Downhill Flux

4.4.1 OHIO STATE GAMES

What is interesting in the analysis of the Ohio State data set was the wide range of values of flux seen in the three games observed. In the loss to Michigan State, the flux was the highest, and in perhaps Ohio State's biggest win of the season, Kentucky, the flux was the lowest. This would suggest that Ohio State plays better when they keep the flux down, or, in other words, they are not trying to get the ball in one player's hands, but rather spreading it evenly. Ohio State is a young team with a group of talented players, but no dominant scorer (having a dominant scorer is a typical staple of the recent Ohio State basketball teams). Since no Ohio State player is great at creating their own shot and proverbially "taking over a game" their best chance to be successful is to get everyone in the offense involved, and use this to keep the defense off-balance, and ultimately get better shots. Looking forward, it will be interesting to see if OSU remains successful if they keep their flux low - the team will retain a majority of the nucleus that composed the team's rotation, but with so many talented players, one player may break away from the pack and become the "go-to" option on the team, and if this is the case, the interpretation of uphill-downhill flux for this team may change

Table 4: Uphill-Downhill flux observed for teams in games

Washington Wizards	-.203
Miami Heat	-.235
Toronto Raptors	-.092
Chicago Bulls	-.984
Los Angeles Lakers	-.627
Boston Celtics	.931
Charlotte Hornets	-.454
Cleveland Cavaliers	-.246
Ohio State vs. Kentucky	.278
Ohio State vs. Michigan State	-.946
Ohio State vs. Iowa	-.402

4.4.2 COMPARISON WITH NBA

While passing around the ball improves the plays appear to be an obvious statement, in the NBA dataset, the team with the lowest uphill-downhill flux (Boston) lost their game at home to the lowly Brooklyn Nets. Conversely, the team with the highest, Chicago, won convincingly. This suggests that uphill-downhill flux is something that needs to be looked at in a case-by-case basis - certain teams that are built around one scorer (i.e. *Jimmy Butler* and the Bulls) are better served moving the ball into that player's hands, while others, like Ohio State, are the opposite. One team that had a huge difference in calculated uphill-downhill flux, was the Boston Celtics, who had a value that was much higher than any other team, and thus had a much lower uphill-downhill flux. This value makes sense as *Brad Stevens* has been known to constantly rotate players in-and-out of their rotation, and opt to spread the ball around evenly, as all 5 starters and other players in the rotation have the ability to score. This asserts the validity of the measure, as it matches exactly what the intuition of such a measure would be. Chicago Bulls had the highest (most negative) uphill/downhill flux with 4 players within 16 and 23 points.Thus, balanced scoring does not imply a low uphill/downhill flux. Los Angeles Lakers had the 2nd highest uphill/downhill flux without *Kobe* playing, which shows that they are still sticking to their old habits.

5. RELATION TO STATE OF THE ART

There has only been one prior work by Fewell et al. [2] in this area which can be compared to ours. We couldn't compare the set of games they've included in their analysis as NBA only made the SportVU logs available starting in 2014, hence the logs for games in that paper aren't available. There are certain clear advantages of our work over the existing work.

- **Generic Framework for Analysis**: We have developed a generic end to end framework for analysis of the NBA games which only needs the logs provided by NBA, and doesn't need the taxing manual annotation of the games as done in the existing works.

- **Differential Edge Weighing** : We have the game, and shot clock timing information available which can be used to grade the importance of passes, and shots, and assign the edge weights accordingly instead of considering every interaction as equal. Along with these, other factors such as the opposition's defense, position of the player on the court at each moment can also be factored in weighing the interactions. Constructing a good edge weight model would require in-depth analysis of quite a few games, and insights from domain experts.

- **Intuitive UI Tool for Analysis**: Our framework provides an intuitive UI for analyzing games, and studying different node/network specific measures for selected games.

6. CONCLUSION

From our observations, we can conclude that though entropy seems to have no discernible effect on wins or losses for the Buckeyes, there are some correlations that can be seen from degree centrality and clustering coefficient. While the conclusion from degree centrality is not groundbreaking - many coaches aim for great ball movement and "getting everyone a touch" to make sure the defense doesn't key in on one of their players - it does reinforce that for a team like Ohio State, that does not possess a star player, they need to rely on spreading the ball around and keeping the opponents off-balance.

Clustering coefficient brings up an interesting discussion on *Trevor Thompson* - a relatively unheralded high-school prospect who has smooth moves from the post and runs the floor very well, but that was generally one of the steadier contributors for this Ohio State team, albeit in relatively low volume this season. From the results of the wins and the loss, it seems that the team generally does better when *Thompson* is getting more opportunities to touch the ball and facilitate offense - on a team with quite a few developing young talents that each want the ball in their hands quite a bit (*Lyle, Loving, Bates-Diop, Tate*), can they find a way to incorporate *Thompson* more? Along with this, if they do, will he be able to continually produce at the rate he is at a higher volume? If he is able to continue his play at a higher volume, Ohio State could have a low-post offensive presence that they've lacked in prior years, and a guy that could really help open up the perimeter for a crowded OSU backcourt.

7. FUTURE WORK

The most obvious direction ahead is implementation of a good edge weighing model, and more analysis measures. One important distinction that needs to be made is that the datasets from the NBA and Ohio State basketball games are not constructed in the same fashion. The NBA data keeps track of moment-by-moment data, captured at .04 second intervals, keeps track of the time at each moment, and has specifically what kind of pass each player utilizes. In constructing the Ohio State graph, the graph was generated manually, through watching actual game footage and keeping track of passes and shots. Due to the volume of time needed to construct one graph, the time aspect and the specific type of pass was ignored, due to the amount of time needed to construct a graph. Thus, none of the calculated measures take into account time.

In addition, the NBA and college games are different in some senses, NBA teams traditionally utilize more players, and is a longer game than college. Despite this, it still was able to serve as an effective baseline. If this study were to be taken more in depth, it would serve the study better to compare the Ohio State data to other college basketball data, perhaps from prior Ohio State teams. We also hope to analyze the data further, while taking advantages of the notion of space and time, which provides a better notion of path speed and length.

8. REFERENCES

[1] Sportvu. http://www.stats.com/sportvu/sportvu-basketball-media/.

[2] J. H. Fewell, D. Armbruster, J. Ingraham, A. Petersen, and J. S. Waters. Basketball teams as strategic networks. *PLoS ONE*, 7(11):1–9, 11 2012.

[3] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 7(11):379–423, 11 2012.

[4] S. J. Zaccaro, A. L. Rittman, and M. A. Marks. Team leadership. *The Leadership Quarterly*, 12(4):451–483, 2002.