Netball: NBA basketball as a social network

Dieter Armbruster,
School of Mathematical and Statistical Sciences
Arizona State University

armbruster@asu.edu
Reference
Basketball teams as strategic networks,
J.H. Fewell, D, Armbruster, J.Ingraham, A.Petersen, J. S. Waters,
PLoS ONE 7(11): e47445. doi:10.1371/journal.pone.0047445

Thanks:
Supported by the Volkswagen Foundation and by NSF (BECS 1023101)
A basketball game as a succession of networks

Consider the 5 players on the floor as 5 nodes in a network. For one shot clock, register the ball movements. A ball movement from player $i$ to player $j$ establishes a link. With the same 5 players on the floor every shot clock generates a new network.

Goal: Characterize the average topology (structure) of these networks
A Basketball Play as a Sequence

**POSSESSION ORIGIN**

- Inbound (IB)
- Rebound (RB)
- Steal (ST)

**PLAYER PASSES**

- Player 1 (#)
- Player 2 (#)
- Player 3 (#)
- Player 4 (#)
- Player 5 (#)

**POSSESSION OUTCOME**

- 2-Point Success (S2)
- 2-Point Fail (F2)
- 3-Point Success (S3)
- 3-Point Fail (F3)
- Turnover (TO)
- Shooting Foul (SF)
- Out of Bounds (OB)
- Defensive Foul (DF)
- Offensive Foul (OF)

**Figure:** The fundamental node structures. 1 = point guard, 2 = shooting guard, 3 = small forward, 4 = power forward, 5 = center
NETBALL DATA SET
====================================================================
Team: PHX Suns
Playing: CLE Cavaliers
Date: 12/21/2009
Quarter: All

1= Amare Stoudemire
13= Steve Nash
33= Grant Hill
23= Jason Richardson
8= Channing Frye
====================================================================
<///>

\[
\begin{array}{cccccccc}
IB & 13 & 1 & F2 \\
RB & 8 & 13 & 33 & 13 & 8 & 23 & S3 \\
RB & 8 & T0 \\
IB & 13 & 1 & S2 \\
\end{array}
\]

Our data:

- Two games of each of the 2009/2010 playoff pairing (total of 32 games)
- All the games from the Lakers - Celtics finals
Networks measures as a quantitative way to describe NBA basketball

- Is there a typical NBA basketball network structure?
- Is there a characteristic team structure?
- How much variation in the team structure is there between the teams?
- How much variation in the team structure is there between games?
Networks measures to measure team success and the contribution by individuals

- Can we correlate team success with some network measures?
- Are the correlates predictive?
- Can we identify the most important contributors to team success?
- How do we identify a leader in a small team?
All games

Inbound, Rebound, Steal, PG, SG, SF, PF, C, 2 Point Success, 2 Point Fail, 3 Point Success, 3 Point Fail, Shooting Foul, 2 Pt. Fail, Shooting Foul, 3 Pt. Fail, Shooting Foul, 2 Pt. Success, Shooting Foul, 3 Pt. Success, Turnover, Out of Bounds, Defensive Foul, Offensive Foul, Shot Violation
Distribution of Ball Movement in Graph of Phoenix Suns Starter Possessions

The Set \( n \) Largest Ball Transitions out of 64 Transitions

% of Total Values

% of Total Edges

Links and Node distributions
Characterizing a network

The core network

Reduce the number of links by dropping the links with the least weights.
Network Measures

- Node degree
- Degree Centrality
- Clustering Coefficient
- Flow Centrality
- Entropy
### Definitions I

**Not your typical network: directed and weighted flows**

- Consider a network with $J$ nodes and $N$ ball movements.
- Define $w_{ij}$ to be the number of ball movements from node $i$ to node $j$.
- $C_w^D(i) = \frac{\sum_{j=1}^{J} w_{ij}}{N}$ is the (weighted) degree (centrality) of the node $i$.
- $C^w_D(i)$ is the total flux through $i$. 

<table>
<thead>
<tr>
<th>Node $i$</th>
<th>$C_w^D(i)$</th>
<th>$C^w_D(i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i_3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i_4$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Degree centrality of a network

- Consider a star-like networks were the node $I^*$ is the central node and all other nodes are only connected to the central node. Let $H^w$ be the sum of the differences between the degree of $I^*$ and the degree of all the other nodes.

$$H = J - 1$$

independent of all weights.

- The degree centrality of the weighted network $G$ is defined as:

$$C_D^w(G) = \sum_{i=1}^{J} \frac{C_D^w(I^*) - C_D^w(i)}{H}.$$  \hfill (1)
What does degree centrality measure?

- $C_D(G)$ varies between zero and one.
- Degree centrality is zero when all nodes have the same centrality as the maximal node, i.e. for a network that is all to all coupled.
- Degree centrality is one for a star shaped network where all links go to one central node and there are no other links.
Degree Centrality characterizes team networks.

Team with the lowest degree centrality

Lakers
low centrality, high entropy

Inbound, Rebound, Steal, PG, SG, SF, PF, CN, Success, Fail, Other, Inbound, Rebound, Steal, PG, SG, SF, PF, CN, Success, Fail, Other

Diagram showing connections between different team positions.
Degree Centrality characterizes team networks II

Team with the highest degree centrality

Jazz
high centrality, low entropy

Inbound, Rebound, Steal, PG, SG, SF, PF, CN, Success, Fail, Other
Small Forward, not Point Guard is the dominant playmaker. Which team is this and who is the SF?
Small Forward, not Point Guard is the dominant playmaker. Which team is this and who is the SF?

Cleveland Cavaliers and LeBron James
Strongest link is not the inbound pass to the PG but the pass to the Power Forward. Which team is this and who is the PF?

Dallas Mavericks and Dirk Nowitzki
Strongest link is not the inbound pass to the PG but the pass to the Power Forward. Which team is this and who is the PF? Dallas Mavericks and Dirk Nowitzki
Clustering Coefficients I

Measures the number or triangular connections in the network

**Issues**

- Directed graphs
- Weighted graphs
- Links are association links - **NOT** sequences.
- CC depends on the probability cutoff - for high cutoff everybody is linked to everybody else.
Clustering Coefficients II

CC for different cutoff weights

<table>
<thead>
<tr>
<th>Team</th>
<th>Clustering Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bucks</td>
<td>0.02</td>
</tr>
<tr>
<td>Bulls</td>
<td>0.04</td>
</tr>
<tr>
<td>Magic</td>
<td>0.04</td>
</tr>
<tr>
<td>Cavaliers</td>
<td>0.08</td>
</tr>
<tr>
<td>Hawks</td>
<td>0.1</td>
</tr>
<tr>
<td>Bobcats</td>
<td>0.12</td>
</tr>
<tr>
<td>Heat</td>
<td>0.16</td>
</tr>
<tr>
<td>Celtics</td>
<td>0.2</td>
</tr>
<tr>
<td>Jazz</td>
<td>0.02</td>
</tr>
<tr>
<td>Suns</td>
<td>0.02</td>
</tr>
<tr>
<td>Mavericks</td>
<td>0.04</td>
</tr>
<tr>
<td>Thunder</td>
<td>0.04</td>
</tr>
<tr>
<td>Nuggets</td>
<td>0.06</td>
</tr>
<tr>
<td>Blazers</td>
<td>0.1</td>
</tr>
<tr>
<td>Spurs</td>
<td>0.2</td>
</tr>
<tr>
<td>Lakers</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Transition probability
Player entropy: How predictable is a player?

Every player has a probability distribution for moving the ball forward.

- A player that has only one outgoing link is highly predictable.
- A player that distributes the ball evenly is unpredictable.
- Are teams that are less predictable more successful?
Entropy

- Define a subnetwork consisting only of 5 players and a shooting node (collapse success and failed shots into one) and define $\omega_{ij}$ for $1 \leq i \leq 6$, $1 \leq j \leq 6$, as the number of ball movements from $i$ to $j$ in the subnetwork.

- Then

$$p_{i,j} = \frac{\omega_{ij}}{\sum_k \omega_{ik}}$$

is the probability that player $i$ moves the ball into node $j$

$$H(i) = -\sum_j p_{i,j} \log p_{i,j}$$

is the entropy of the player $i$. 
Team Entropy

1. Average entropy is the average player entropy $H_a = \frac{\sum_{i=1}^{5} H(i)}{5}$.

2. Markov chain entropy: Conditional entropy $H^M = \frac{\sum_{i=1}^{5} p_i H(i)}{5}$, where $p(i)$ is the probability that player $i$ gets the ball.

3. $p_{i,j}^t = \frac{\omega_{ij}}{\sum_{lk} \omega_{lk}}$ is the probability that the ball moves from $i$ to $j$.

$$H_t = - \sum_{i,j} p_{i,j}^t \log p_{i,j}^t$$

is the team entropy.
Can we predict a winner?

Team entropy

Result: in 6 out of 8 playoff series the team with the higher team entropy won the series.
All other entropy measures correlate positively with winning.
### Flow Centrality

- Measures the probability that a particular player was involved in a particular ball movement.
- Flow centrality restricted to the last node before a shot counts successful shots.
- Flow centrality restricted to the last two nodes counts assists also.
- Average pathlength in NBA is about 4.5 nodes. To capture the setup of a shot we calculate flow centrality for the last three nodes.
How important is a player for the team?

Flow centrality ratio

- Define flow centrality for all successful plays, $C^s_F$.
- Define flow centrality for all unsuccessful plays, $C^{us}_F$.
- Define the success ratio, $R_s = \frac{C^s_F}{C^{us}_F}$.
- For the last link in a network, $R_s$ is the shooting percentage that is well known for all players.
- Ranking of all players according to $R_s$ will show the most useful players for the team.
- Players with $R_s$ significantly below 1 are hurting their team!
- Maximum $R_s$ over the league is 1.94, minimum is 0.45!
Playing to a team’s strengths

Are there other measures of success?

E.g. passing to players that are more successful?
## Markov property

A network analysis assumes that NBA basketball is Markov i.e. ball movement from a node is independent of the way the ball reached that node.

## Alternative

**Sequences matter**

i.e. ball movement is determined via set plays.
Determining the Markov order

Analysis for the Heat-Thunder playoff 2012

- Consider the field goal percentage of a team, \( fg\% \)
- For every play that leads to a shot, determine the field goal percentage if player \( j \) is in the second to last position of the sequence, i.e. he is credited with an assist.
- This is the conditional shooting percentage \( fg\%_{t-1=j} \) for all players \( j = 1..5 \).
- If \( fg\%_{t-1=j} = fg\% \) in a statistically significant way, then the process is Markov otherwise, it has at least a one step memory.
NBA basketball is Markov!

Results

- Null Hypothesis: no difference between field goal percentages.
- Null Hypothesis is not rejected. ($z < 1.96$)
- Closest cases:
  - Power Forward for the Heat (Lebron James) has the largest influence (negative).
  - Point Guard for the Thunder (Russell Westbrook) has the 2nd largest influence (positive).
Open questions I

<table>
<thead>
<tr>
<th>Characterization by network structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>• How stable are the characteristics of team networks?</td>
</tr>
<tr>
<td>• Can we improve the odds for outcome predictions (in Las Vegas)?</td>
</tr>
<tr>
<td>• Are particular network structures more successful against other network structures?</td>
</tr>
<tr>
<td>• Are network structures characteristic for the phase of a game? E.g. is the network for the first three quarters different from the network for the last quarter?</td>
</tr>
<tr>
<td>• How does the NBA network structure differ from College basketball - men’s basketball from women’s basketball?</td>
</tr>
</tbody>
</table>
Social networks

- What characterizes a leader? E.g. high degree centrality and high flow centrality ratio.
- Is this transferable to other small teams tasked with
  - innovation (research)
  - execution (military)
  - pattern recognition (intelligence)
Thanks for your attention!