Correlating NBA team network centrality measures with game performance

Adam Reed
Johns Hopkins University
Laurel, MD 20723
Email: areed21@jhu.edu

John Piorkowski
Johns Hopkins University
Applied Physics Laboratory
Laurel, MD 20723
Email: jpiorko2@jhu.edu

Ian McCulloh
Johns Hopkins University
Applied Physics Laboratory
Laurel, MD 20723
Email: imccul4@jhu.edu

Abstract—**Basketball is an inherently social sport, which implies that social dynamics within a team may influence the team's performance on the court. As NBA players use social media, it may be possible to study the social structure of a team by examining the relationships that form within social media networks. This paper investigates the relationship between publicly available online social networks and quantitative performance data. It is hypothesized that network centrality measures for an NBA team's network will correlate with measurable performance metrics such as win percentage, points differential and assists per play. The hypothesis is tested using exponential random graph models (ERGM) and investigating correlation between network and performance variables. The results show that there are league-wide trends correlating certain network measures with game performance, and also quantifies the effects of various player attributes on network formation.**

Introduction

Many NBA basketball players, coaches, and analysts agree that team cohesion is an important requirement for team success. Phil Jackson, the coach with the most championships in NBA history, described team cohesion by saying that "basketball is a sport that involves the subtle interweaving of players at full speed to the point where they are thinking and moving as one. [1]" While it is generally agreed that team cohesion is important, it is a difficult concept to measure and quantify with traditional NBA game statistics. Previous work examined applying social network analysis to team sports [2], which lends credence to the idea that social media analysis may be able to fill this analytical gap by examining the social structure of an NBA team. Team cohesion is colloquially referred to as team chemistry, and the terms will be used interchangeably throughout this paper.

The team cohesion concept has been studied in business environments [3], where it has been shown that social networks can affect team performance. Specifically for the NBA, this type of analysis may help inform roster building decisions. Two steps must be taken to use the analysis productively. First, a correlation between a team's social network and its game performance must be established. Then, once a pattern of social structures that perform well on the court are found and characterized based on various player attributes, NBA front office decision makers can seek out players who will help their team form the known successful patterns. This insight could also be useful when a front office wants to sign or resign a player to a contract. In the case of resigning a player, it has been shown that certain personal ties can influence employee retention [4], so forming the right team network may aid player retention. Generalized for businesses, incorporating social network analysis may help inform teams' management when making personnel decisions.

A "friend" relationship on social media indicates that one user wants to be notified of updates posted by another user, which shows a personal interest. It seems likely that if players interact socially, then those social bonds will translate to on the court cohesion. While this may not always be true, it does seem intuitive that social bonds off the court would lead to increased team cohesion on the court. The social network for a team that is formed by mapping the "follows" relationships between the team members' social media accounts could be useful as a proxy for measurable team cohesion. It is hypothesized that network centrality measures for an NBA team's network will correlate with measurable performance metrics such as win percentage, points differential and assists per play. The remainder of this paper describes the methodology, presents findings and draws conclusions.

Methods

The networks for analysis were formed by finding the verified Twitter accounts for all of the players in the NBA. Twitter was chosen as the social network to study because it freely allows access to the friend lists of its users, whereas Instagram does not. Instagram would be preferable for this type of analysis, because studies have shown that image-centric sites like Instagram offer more personal interconnectedness than Twitter does [5]. This seemed especially true after inspecting and comparing the Twitter posts of a user to their Instagram posts. As an observation, NBA players seem to use Twitter as a marketing tool for themselves, whereas the Instagram posts are much more personal posts and often include their families.

The rosters were identified using public data [6], and the social media accounts were found through internet searches. A social network for each of the 30 NBA teams was created using the Twitter API to read the friends list for each player on the team and forming a connection between players if a player follows another player. The network for a team is a directed graph, since the follows relationship may not be reciprocal. Figure 1 shows an example of the social network formed by the Boston Celtics. Similar networks were formed for every team in the NBA.

As they are public figures, most NBA players have verified Twitter accounts, but some players appear to not use the network in a public manner. For such cases, a vertex is placed in the network without any incoming or outgoing connections. The team rosters used in this experiment are current as of 12/4/2017, but the NBA league allows players to be traded in-season, so the rosters studied here may not be accurate throughout the entire season. Likewise, the measured game performance data was current as of 12/7/2017, and would change throughout the season.

Network centralization metrics of degree, betweenness, and closeness

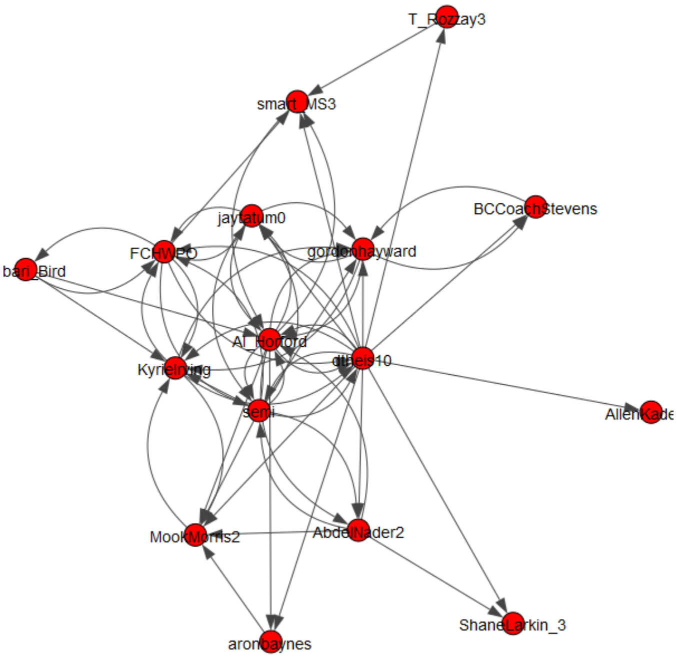

Fig. 1: Network of Twitter friend relationships of the Boston Celtics

were calculated for each team's network. The formulas to compute the centralization metrics are shown in equations 1, 2, and 3, where NC_d, NC_b, and NC_c refer to degree, betweenness, and closeness centralization respectively [7].

$$NC_D = \frac{\sum_{i=1}^{n}(C_{Dmax} - C_{Di})}{(n-2)(n-1)} \tag{1}$$

$$NC_B = \frac{\sum_{i=1}^{n}(C_{Bmax} - C_{Bi})}{n-1} \tag{2}$$

$$NC_C = \frac{\sum_{i=1}^{n}(C_{Cmax} - C_{Ci})}{(n-2)(n-1)/(2n-3)} \tag{3}$$

These network centralization measures were then correlated with various measurable performance statistics [8] to determine whether social ties correlate with a team's game-time performance. For correlation, the Spearman rank correlation coefficient was used instead of the more common Pearson correlation, because normality and independence assumptions may not be appropriate for these data. The correlation was computed for each network centralization metric and performance metric pairing by using a vector of the network centralization metrics for each team and another vector comprised of the performance metric measurements for each team. This treats the data for each team as a sample point, with the goal being to characterize a general trend across the league for each centralization/performance metric pair.

An Exponential Random Graph Model (ERGM) was used to analyze the formation of network connections within a team. ERGM analysis treats the formation of ties between network nodes as random variables [9] and allows a statistical characterization of how important various attributes are for network formation. Attributes for players including age, years of service, minutes played, country of origin, and college were incorporated into the ERGM for each team to test which attributes are most significant for forming network connections.

RESULTS

This section gives the quantitative results and interpretation of both the network centralization correlation and the ERGM analysis.

NETWORK CENTRALIZATION AND GAME PERFORMANCE CORRELATION

Several performance metrics were tested for correlation with the network centralization metrics. The on-court performance metrics [8] and their relevance is shown in Table I.

TABLE I: Performance metric description.

Performance metric	Description
Current Win%(CW%)	The percentage of games won for the season as of the time of analysis.
Expected Win% (EW%)	The number of games that casinos expected the team to win for the season. [10]
CW%/EW% Difference	Calculated as CW%-EW%, a high correlation here would show that network connections connote outperforming the expected win total.
Assists per possession	Number of assists per offensive possession, a high correlation here would show that network connections are related to sharing the ball.
Average scoring margin	Average scoring margin of the final score over all game.
Projected Win% (PW%)	The team's season projected win% taking into account in-season performance by a respected projection algorithm [8]. Important here because there is truth data for only a portion of the season, and given the low sample size, may be slightly anomalous.
PW%/EW% Difference	Calculated as PW%-EW%, a high correlation here would show that network connections connote outperforming the expected win total.

The correlation is performed by treating each team as a sample point, to test for a league-wide correlation between the performance metrics and the network centralization measures. A two-sided p-value is computed alongside the correlation coefficient, where the null hypothesis is that network centralization and quantitative performance metrics are uncorrelated. The alternative hypothesis is that there exists a correlation. TableII shows the computed correlations and p-values.

TABLE II: Correlation results between network centrality measures and team performance metrics

Metric	In Degree Centralization		Betweenness Centralization		Closeness Centralization	
	Correlation Coefficient	Test p-value	Correlation Coefficient	Test p-value	Correlation Coefficient	Test p-value
Current Win %(CW%)	0.29	0.12	0.08	0.64	-0.03	0.89
Expected Win %(EW%)	0.45	0.01	0.05	0.78	0.09	0.64
CW%/EW%Difference	-0.13	0.48	0.07	0.73	-0.13	0.50
AssistsPerPossession	0.28	0.13	0.38	0.03	-.001	0.97
AverageScoringMargin	0.24	0.19	0.13	0.49	0.04	0.84
ProjectedWin %(PW%)	0.40	0.03	0.16	0.39	-0.06	0.77
PW%/EW%Difference	-0.08	0.69	0.31	0.09	-0.35	0.05

Three of the performance metrics have significant correlation with network centralization. The assists per possession is a measure that describes success sharing the ball, and it is significantly correlated with betweenness centralization. This suggests a correlation between offline (brokerage of the ball) and online (brokerage of information) betweenness centrality. The degree and betweenness of a network may relate to how the players pass the ball, which lends credence to the hypothesis that the various network centrality measures may be useful predictors for team assists. In other words, it is somewhat likely that either a high betweenness in off court connections enhances on court cooperation, or teams with a high on court connection are likely to form a network with high betweenness off court.

There is also a significant correlation between degree centralization and the Expected Win percentage and Projected Win percentage. An online social network has high degree centralization when one or a few nodes have much greater degree than the majority of the nodes. This may indicate a charismatic leader who leads both on and off the court. Finally closeness centralization is significantly correlated with difference in projected and expected wins.

The correlation between in degree and expected win percent may indicate that networks with popular nodes tend to be expected to perform well. This may just mean that teams with one or two exceptionally popular players tend to be expected to do well. It is also interesting that the betweenness and closeness network centralization measures correlate with the difference between projected wins and the expected wins. The sign of the correlation of the two centrality measures is different, which makes the meaning somewhat difficult to discern. One possibility is that networks with high betweenness may outperform expectations if there are a few socially important players on the team, but as the team as a whole forms more social bonds, they are more likely to under-perform. Intuitively, this may make basketball sense, because it implies that the best team dynamic is one where a few players are hubs for the rest of the team, which is often how basketball is played at the NBA level, where one or two exceptional players can increase the productivity of the rest of the team [11], [12].

Characterizing network bonds

Exponential Random Graph Model (ERGM) analysis is a method of network analysis that controls for the network structure to allow study of the network effects of different node attributes. ERGM analysis was performed for several teams to see how players' age, years of service in the NBA, and average minutes played per game affect the formation of a team's social network. For each team's network, six ERGM analyses were performed. Three of the ERGM analyses examined network formation with edge covariances of the three attributes being studied. Another three ERGM analyses were performed for each team using the same attributes, and looking at the formation of mutual ties. Table III shows the attributes studied for each analysis.

The data shows that age, years of service, and average minutes per game affect the formation of network ties on social media, indicated by the positive estimate value from each model. This is a general trend across all but one of the teams studied here, the Phoenix Suns, confirmed by the p-values for the attributes studied. In addition to the common notion that friendships are formed within a similar age group, these results lend credence to the idea that players who share game time on the court will befriend each other. It also implies that those players who do not make it into the game often will befriend each other. These findings make intuitive sense, since players on the court interact, and those players not currently on the court would presumably socialize while sitting on the bench together.

TABLE III: ERGM analysis for team networks with ties, age, years of service, and average minutes played attributes

Team	ERGM Query	Estimate	Std. Error	P-value
Suns	Mutual With YoS	3.541199	0.652504	<0.0001
	Years Of Service (YoS)	-0.007224	0.019985	0.718
	Mutual With Age	3.527599	0.660806	<0.0001
	Age	-0.01034	0.009504	0.27835
	Mutual With Minutes Played	3.543886	0.660698	<0.0001
	Minutes Played	0.004955	0.010407	0.635
Cavaliers	Mutual With YoS	2.9716	0.51522	<0.0001
	Years Of Service (YoS)	0.12066	0.02093	<0.0001
	Mutual With Age	2.77944	0.51975	<0.0001
	Age	0.08872	0.01734	<0.0001
	Mutual With Minutes Played	3.036593	0.509243	<0.0001
	Minutes Played	0.052874	0.009588	<0.0001
Celtics	Mutual With YoS	2.48177	0.49743	<0.0001
	Years Of Service (YoS)	0.04044	0.02869	0.16
	Mutual With Age	2.33325	0.50673	<0.0001
	Age	0.03577	0.01161	0.00231
	Mutual With Minutes Played	2.42468	0.507141	<0.0001
	Minutes Played	0.020043	0.008304	0.0165
Grizzlies	Mutual With YoS	2.83691	0.44818	<0.0001
	Years Of Service (YoS)	0.0446	0.02018	0.028
	Mutual With Age	2.71045	0.44802	0.0001
	Age	0.03911	0.01087	0.000381
	Mutual With Minutes Played	2.824437	0.455001	<0.0001
	Minutes Played	0.019265	0.007646	0.0123
Mavericks	Mutual With YoS	2.70445	0.61125	<0.0001
	Years Of Service (YoS)	0.03957	0.01906	0.0396
	Mutual With Age	2.66848	0.61543	<0.0001
	Age	0.03464	0.01527	0.0247
	Mutual With Minutes Played	2.38685	0.63918	0.000265
	Minutes Played	0.05122	0.01429	0.000456

This line of analysis shows that there are measurable attributes that drive the formation of teams' social networks across the league. Combined with the findings that various social network metrics correlate with on the court performance, this opens a potentially new way to approach team building.

Conclusion

This study proposed using social network analysis as a proxy to quantify team chemistry. The study focused on two main areas: one testable hypothesis, that network centrality measures for an NBA team's network will correlate with measurable performance metrics, and then used the ERGM technique to determine if various quantifiable attributes affect network formation.

The hypothesis of correlating network centrality measures with performance was not strongly confirmed, but there was evidence that it may be true. There were low to moderate correlations between some network centrality measures and assists per game, as well as a team's expected win percentage. This study was limited to Twitter data. A follow-on study could attempt to use either Instagram friendships, or Instagram mentions and likes between players, as the foundation for building the team's social network. It's likely that a more accurate representation of the social networks would lead to confirming or disproving the hypothesis. The methods here show that it may be possible to predict how a team will behave on the court based on its social network.

The ERGM analysis showed that there several player attributes that correspond to friendships forming among players. This analysis could potentially open up new ideas for building an NBA team, where a decision maker starts with an end goal for the team's social network that is known to perform well, and hires players who are likely to form that network.

The Twitter networks studied showed enough promise that a better approximation of team chemistry could potentially lead to strong confirmation of the hypothesis. Furthermore, since the ERGM analysis showed that relevant NBA metrics drive team social network formation, more attributes should be investigated. The attributes used in this study were all quantifiable attributes taken from known data, but a follow on study could test player personality traits. For instance, perhaps a team is considering drafting a troubled but talented player. They might want to examine how likely it is that the player will form positive relationships on the team and not cause problems. Both of the methods of analysis studied in this experiment show promise and further study in both areas could lead to some interesting team building methodologies both for NBA teams and for workplaces in general.

REFERENCES

[1] P. Jackson, *Sacred hoops : spiritual lessons as a hardwood warrior*. New York, N.Y: Hyperion, 2006.

[2] D. Lusher, G. Robins, and P. Kremer, “The application of social network analysis to team sports,” *Measurement in Physical Education and Exercise Science*, vol. 14, no. 4, pp. 211–224, 2010. [Online]. Available: https://doi.org/10.1080/1091367X.2010.495559

[3] E. Rosenthal, “Social networks and team performance,” *Team Performance Management: An International Journal*, vol. 3, no. 4, pp. 288–294, 1997. [Online]. Available: https://doi.org/10.1108/13527599710195420

[4] T. R. Mitchell, B. C. Holtom, T. W. Lee, C. J. Sablynski, and M. Erez, “Why people stay: Using job embeddedness to predict voluntary turnover.” *Academy of Management Journal*, vol. 44, no. 6, p. 11021121, Dec 2001. [Online]. Available: http://dx.doi.org/10.2307/3069391

[5] M. Pittman and B. Reich, “Social media and loneliness: Why an instagram picture may be worth more than a thousand twitter words,” *Computers in Human Behavior*, vol. 62, no. Supplement C, pp. 155 – 167, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563216302552

[6] N. League. (2017) Nba team rosters https://www.teamrankings.com/nba/projections/standings/. [Online]. Available: http://www.nba.com/teams

[7] I. McCulloh, H. Armstrong, and A. Johnson, *Social network analysis with applications*. Hoboken, New Jersey: John Wiley & Sons, Inc, 2013.

[8] TeamRankings. (2017) Nba teamrankings https://www.teamrankings.com/nba/projections/standings/. [Online]. Available: https://www.teamrankings.com/nba/projections/standings/

[9] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, “An introduction to exponential random graph (p*) models for social networks,” *Social Networks*, vol. 29, no. 2, pp. 173 – 191, 2007, special Section: Advances in Exponential Random Graph (p*) Models. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378873306000372

[10] VegasInsider. (2017) Nba 2017-2018 win totals https://www.teamrankings.com/nba/projections/standings/. [Online]. Available: http://www.vegasinsider.com/nba/story.cfm/story/1865923

[11] M. Goldman and J. M. Rao, “Allocative and dynamic efficiency in nba decision making,” in *In Proceedings of the MIT Sloan Sports Analytics Conference*, 2011, pp. 4–5.

[12] D. Cervone, A. DAmour, L. Bornn, and K. Goldsberry, “Pointwise: Predicting points and valuing decisions in real time with nba optical tracking data,” in *In Proceedings of the MIT Sloan Sports Analytics Conference*, 2014.