Evaluating Basketball Player Performance via Statistical Network Modeling

James Piette[*], Lisa Pham[†], Sathyanarayan Anand[‡]
Philadelphia, PA, USA, 19104
Email: jpiette@wharton.upenn.edu

Abstract

The major difficulty in evaluating individual player performance in basketball is adjusting for interaction effects by teammates. With the advent of play-by-play data, the plus-minus statistic was created to address this issue [5]. While variations on this statistic (ex: adjusted plus-minus [11]) do correct for some existing confounders, they struggle to gauge two aspects: the importance of a player's contribution to his units or squads, and whether that contribution came as unexpected (i.e. over- or under-performed) as defined by a statistical model. We quantify both in this paper by adapting a network-based algorithm to estimate centrality scores and their corresponding statistical significances [10]. Using four seasons of data [9], we construct a single network where the nodes are players and an edge exists between two players if they played in the same five-man unit. These edges are assigned weights that correspond to an aggregate sum of the two players' performance during the time they played together. We determine the statistical contribution of a player in this network by the frequency with which that player is visited in a random walk on the network, and we implement bootstrap techniques on these original weights to produce reference distributions for testing significance.

1   Introduction

It is vital in team sports, such as basketball, to be able to estimate individual performance for personnel purposes. The main obstacle analysts in these sports face when evaluating player performance is accounting for interaction effects by fellow teammates, or teamwork. Certain players might find themselves scoring more on a team not because of an increase in scoring ability, but due to the lack of a supporting cast (e.g. Allen Iverson of the Philadelphia 76ers in the 1990s). This paper takes a new approach to analyze this classic problem. Our goal is

---

[*]The Wharton School, University of Pennsylavania
[†]Bioinformatics Program, Department of Biomedical Engineering, Boston University
[‡]The Wharton School, University of Pennsylavania

to answer two fundamental and interconnected questions related to individual ability:

- given the five-man units of which a player was a member, how important was that player relative to all other players

- and, how well statistically did that player perform in that role?

We aim to answer these questions through two measures generated from a novel form of network analysis and bootstrap testing.

With the advent of play-by-play data (i.e. logs of each play occurring in a basketball game), basketball analysts began to record a statistic that had already been popularized in hockey, called plus-minus. Plus-minus describes a player's point differential, or the difference between points scored and points allowed while that player was playing [5]. However, potential confounders exist with this approach; in particular, certain players tend to be on the court at the same time, which could lead to negative (or positive) biases. Rosenbaum [11] created new estimates with this statistic of player ability, called adjusted plus-minus. Using a framework similar to random effects, parameters representing individual player contributions are estimated against their respective observed point differentials.

Network analysis is not entirely new to sports. Their most common application is in the computerized rankings for NCAA football teams [3,8]. In [8], the strength (weakness) of a college football team is determined by a function of their wins (i.e. edges between them and schools they defeated) and their indirect wins (i.e. edges between schools they defeated and schools that those schools defeated). We see less use of networks in the realm of basketball. Skinner [12] frames a basketball team's offense as a network problem, where we seek to find the optimal "pathway", or series of plays that generate the most points. Neural networks have been proposed to predict the outcome of NBA games [6].

We choose to build a model inspired by work on social networks [8,10,13]. Pham et al [10] propose a new algorithm called Latent Pathway Identification Analysis (LPIA), that identifies perturbed cellular pathways using a random walk on a biological network. This network is designed to encourage the random walk to visit areas of high gene transcriptional dysregulaton. In the same spirit, we implement an algorithm that executes a similar search on a network of basketball players.

We begin by extrapolating information obtained from observations of five-man units. These observations correspond to the posterior means of unit efficiencies[1], calculated from sampled chains of a Bayesian normal hierarchical model. We use this information to assess player interactions. We construct a network of individual players, where the nodes are players and two players are connected if they were a member of the same five-man unit at least once. Importantly, the edges are weighted to reflect the interdependency between players with respect to their units' performances. Using a random walk, we determine

---

[1]The word offensive (or defensive) efficiency is defined as the number of points scored (or allowed) per offensive (or defensive) possession.

statistically how central/important a player is relative to all other players in the network, which is referred to as a centrality score. Furthermore, bootstrapping techniques are used to calculate the statistical significance of these centrality scores, or these player performances.

2 Methodology

2.1 Data Preprocessing

We choose to use four seasons of play-by-play data, taken from [9]: '06-'07, '07-'08, '08-'09, and '09-'10. We analyzed these data to determine for each possession, the two five-man units on court, which unit was home (or away), which unit had possession of the ball, and the number of points scored.

We borrow heavily from the model outlined in [2]. Let y_{ij} denote the number of points scored (or allowed, when analyzing defense) by unit i for possession j after adjusting for home court effects[2]. The data likelihood in our model follows

$$y_{ij} \sim \text{Normal}(\theta_i, \sigma^2),$$

where σ^2 is the shared variance for each observation and θ_i is the mean efficiency for unit j[3]. We place a prior density on each θ_i of

$$\theta_i \sim \text{Normal}(\mu, \tau^2),$$

where μ represents the league-mean efficiency and τ^2 is the corresponding variance. To generate posterior estimates for the parameters of interest (i.e. the θ_i's), we implement a Gibbs sampler, detailed in Appendix A.

2.2 Constructing the Weighted Network

As in [10], we construct a network of players that biases a random walk around areas of high performing players. An example illustrating the construction of this network can be seen in Figure 1. We say two players share an interaction if they played together in a five-man unit; moreover, this interaction is enhanced if they did so efficiently. We use the offensive and defensive efficiencies of units obtained in Section 2.1, as well as total efficiencies[4], to infer the interaction effect between two players.

To do this precisely, we use a boards-members bipartite graph concept (e.g. see [13]), where nodes are either five-man units U or players P, and edges exist

---

[2]The home court effect is estimated empirically by taking the difference between the observed efficiency of all possessions at home versus away.

[3]The real observations are discretized, not continuous as this likelihood would suggest. We chose to take this simpler approach because (i) the difference is minute and (ii) the minimum requirement for possessions played by a unit is such that the Central Limit Theorem can be (reasonably) applied.

[4]A unit's total efficiency is a combination of a unit's estimates for offensive and defensive efficiency. For more explanation, see Appendix B

only between these two types of nodes. We represent this bimodal network as an incidence matrix W, where the rows are units and the columns are players. A player P_i is adjacent to a unit U_j (i.e. $w_{ij} \neq 0$) if he has played in that unit. This edge w_{ij} is weighted by the efficiency of unit U_j.

We project this bimodal network onto a unimodal network by computing $A = W^T W$. In this final network of just players, the weight of an edge between two player nodes is the sum of the squares of their shared units' efficiency scores. Thus, edge weights in the player network will be large for pairs of players who have played in efficient units as opposed to inefficient units.
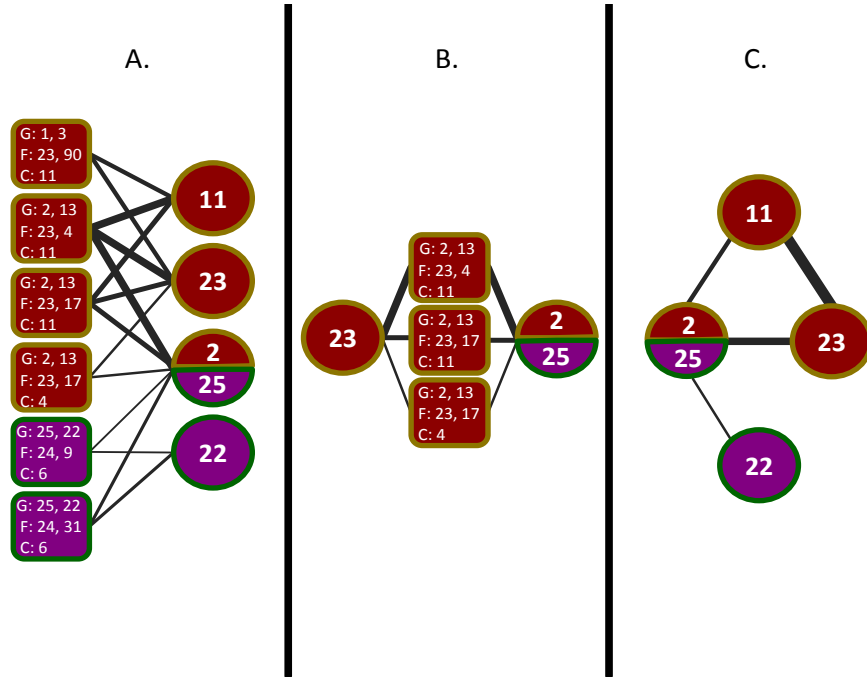

Figure 1: An example of the construction of a four-node player network, as detailed in section 2.2. Each circle represents a player and each square corresponds to a specific unit. The numbers refer to the uniform numbers worn by each player. Note that the two adjoined semi-circles represent Mo Williams, who in our dataset, played for two different teams: the Cleveland Cavaliers and the Milwaukee Bucks. *A.* A bipartite graph is created, where nodes are either five-man units or players, and edges exist only between these two types of nodes. *B.* This is an example of how two player nodes (i.e. LeBron James and Mo Williams) from the bimodal network are projected onto the unimodal network. *C.* We repeat this process for all players in the network and, thus, generate the four-node unimodal, player network.

2.3 Computing Player Centrality

We use eigenvector centrality with random restart to determine the centrality (or importance) of a player in a network (e.g. [7, 10]). Eigenvector centrality uses the stationary limiting probability π_i that a random walk on our network is found at the node corresponding to P_i [4, 7]. The walk is biased by the edge weights such that the random walker travels on heavier edges with greater probability than lighter edges. With this form of centrality and the design of our weighted network, a player is deemed important if he has important neighbors (i.e. played in a significant number of efficient units). For further details on the specific centrality measure used, see [7, 10].

2.4 Statisical Signficance of Player Centrality Scores

Eigenvector centrality measures are influenced by both edge weights and node degree. As a result, it is possible for a player to have his centrality score artificially inflated by having many neighbors. To adjust for this, we employ a bootstrap-based randomization procedure as in [10] to provide a p-value associated with every centrality score. By bootstrapping the unit efficiency scores, we recreate the bipartite network and consequently, the unimodal player network. Finally, bootstraped versions of centrality scores are recreated. We do this for a large number of iterations J, and obtain a reference distribution for the centrality score π_i^* of a given player P_i. We compare a player's original centrality score π_i^0 to this distribution to obtain p-values, by computing

$$pval(P_i) \;=\; \frac{\sum_{k=1}^{J} I_{\{\pi_i^{*k} \geq \pi_i^{(0)}\}}}{J}.$$

Extreme p-values indicate players that do not perform as expected by chance. Small p-values would indicate over-performance, while large p-values would indicate under-performance.

3 Results

We obtain three networks weighted using three datasets: offensive, defensive and total efficiency. Each network contains 590 players that were members of 5961 distinct 5-man units over the course of the four seasons.

Figure 2 shows histograms of raw p-values from the offensive and defensive networks. We notice that the histograms are heavier at the boundaries, while the centers looks roughly uniform. In the histogram for offense, the number of players with extremly high p-values are nearly double the number of those with extremely low p-values; the converse is true for the histogram of defense, suggesting that over-performing offensively is harder than over-performing defensively. However, the total number of exceptional players (i.e. statistically significantly under- or over-performing) is about even.

Since we are performing tests for every player, we need to adjust the raw p-values for multiple testing by using BH procedures [1]. We then use these adjusted p-values to classify over-performers and under-performers at a threshold of 10%. Thus, if a player has an adjusted p-value of less than 0.10 (with regard to over-performing for instance), then we would successfully reject the null hypothesis that the player's performance can be explained by chance[5]. We display a select number of players classified as "exceptional" in Table 2. It should be noted that exceptional could refer to both over-performance or under-performance.

There are several well-known players who over-perform on offense, but under-perform on defense. One example is Steve Nash, who has been a member of some of the best offensive units in history. A few other examples include famous players known for their offensive capabilities: Kobe Bryant, Pau Gasol and Deron Williams. The collection of players that under-perform on offense and over-perform on defense is more obscure. Shelden Williams, an example of this phenomenon, is a young center cited for his aggresive style of play and shot-blocking ability. Marcus Williams, Ime Udoka and Antoine Walker are other over-performers on defense, under-performing on offense.

An advantage of our algorithm is the ability to search for players who are under-utilized by their teams/coaches. To find these "prospects", we look for players whose centrality score is small, but over-perform statistically in one of the efficiency categories. The Celtics, a team recently infamous for their defensive prowess, have several bench players that meet these criteria (e.g. Brian Scalabrine and Tony Allen). In terms of total efficiency, one "prospect" is George Hill, who has served as a key role player for a great San Antonio Spurs team.

By this same method, it is possible to find players who are receiving too much playing time i.e., under-performing players with high centrality scores. Jarret Jack is the most aggregious such case. He has incredibly high centrality rankings and under-performs in nearly every aspect of play.

Not all players are exceptional. In fact, many important players (i.e. high centrality scores) have performance levels that are as expected by chance, as seen in Table 1. The first four players on that list are especially central to every aspect of play, but are exceptional in none of them. To better understand if these players' high centrality is due to skill, we look to the number of different units they played in and determine if that number is unusually high (e.g. traded multiple times).

Two interesting players worth noting, are Greg Oden and LeBron James. With a very low centrality score and a significantly small p-value, Greg Oden makes a case to be the most under-utilized over-performer. However, the former number one NBA draft pick qualifies as under-used due to injury, not management choice. As expected, LeBron James ranks at number one in terms of centrality scores in each case (offense, defense, and total), suggesting he is the most important player in the network. More importantly, he over-performs in

---

[5]We split up the testing (and the adjustments) because we are testing two separate scenarios. If we only had interest in whether or not a player was exceptional, one two-sided test (and one adjustment) would be needed.

Table 1: The top 5 most important (i.e. high centrality) players in terms of total efficiency who are unexceptional in every aspect.

Name	Offense C-Rank[a]	Offense P-value[b]	Defense C-Rank[a]	Defense P-value[b]	Total C-Rank[a]	Total P-value[b]
J. Crawford	3	0.942	2	0.795	2	0.872
A. Iguodala	4	0.889	4	0.372	3	0.262
D. Granger	5	0.796	6	0.705	4	0.756
S. Jackson	7	0.889	5	0.746	6	0.635
C. Maggette	22	0.265	3	0.933	7	0.210

[a] Centrality Rank (rank according to centrality score). Note that these are out of 590 possible players.
[b] P-values adjusted for multiple testing.

two of the three areas (offense and total). LeBron James is often thought of as the top NBA player and was named MVP of the league in both 2009 and 2010, and this serves as important external validation.

4   Conclusion

Our paper contributes a new approach to a well-researched topic by employing network analysis techniques, rather than traditional regression methods. Our algorithm provides new and interesting ways of evaluating basketball player performance. We shed light on statistically significant players who are under- and/or over-performering on offense, defense, and in total. We gain insight on how important certain players are to their units relative to other players. Lastly, by combining these two aspects, we form a more complete analyis of a player's abilities.

One obvious expansion of this algorithm is to use different measures of unit performance (e.g. rebounding and turnover rates), which we can use to gauge other aspects of a basketball player's skill set. Another interesting model extension is to calculate centrality scores of edges, instead of nodes. These scores correspond to the importance of how two teammates perform together. In this way, we can perform the same type of performance evaluation on pairs of teammates, which could highlight players who while not successful individually, work great as a pair.

Table 2: Tables showing a selection of exceptional players on offensive, defensive and total efficiencies.

Exceptional Performers on Offense

Over-performers			Under-performers		
Name	C-Rank[a]	P-value[b]	Name	C-Rank[a]	P-value[b]
LeBron James	1	0.017	Fred Jones	237	0.000
Dirk Nowitzki	2	0.000	Ime Udoka	264	0.040
Chris Bosh	6	0.000	Chucky Atkins	276	0.098
Kobe Bryant	25	0.000	Antoine Walker	307	0.000
Deron Williams	29	0.000	Marcus Williams	315	0.029
Steve Nash	37	0.000	Shelden Williams	328	0.055
Pau Gasol	56	0.000	James Singleton	370	0.055
Tony Parker	98	0.017	Brian Cardinal	382	0.029
Mario Chalmers	210	0.065	DeAndre Jordan	391	0.029
Greg Oden	345	0.017	Cedric Simmons	505	0.000

Exceptional Performers on Defense

Over-performers			Under-performers		
Name	C-Rank[a]	P-value[b]	Name	C-Rank[a]	P-value[b]
Eddie House	101	0.073	Chris Bosh	35	0.084
Daniel Gibson	121	0.086	Kobe Bryant	57	0.051
Tony Allen	174	0.073	Josh Smith	66	0.084
Ime Udoka	175	0.053	Deron Williams	92	0.000
Amir Johnson	195	0.000	Carmelo Anthony	93	0.034
Glen Davis	209	0.086	Chauncey Billups	132	0.000
Antoine Walker	230	0.086	Jason Richardson	136	0.000
Marcus Williams	258	0.053	Steve Nash	137	0.000
Shelden Williams	290	0.086	Amare Stoudemire	147	0.000
Brian Scalabrine	331	0.076	Pau Gasol	176	0.000

Exceptional Performers in Total

Over-performers			Under-performers		
Name	C-Rank[a]	P-value[b]	Name	C-Rank[a]	P-value[b]
LeBron James	1	0.000	Al Jefferson	18	0.034
Dirk Nowitzki	5	0.000	Rudy Gay	20	0.000
Dwight Howard	17	0.000	Jarret Jack	21	0.000
Paul Millsap	27	0.000	Ryan Gomes	24	0.000
Anderson Varejao	31	0.000	Troy Murphy	36	0.070
Amir Johnson	213	0.098	O.J. Mayo	204	0.000
George Hill	227	0.000	Tyreke Evans	307	0.034
Glen Davis	231	0.062	Josh Powell	308	0.000
Greg Oden	387	0.098	Yi Jianlian	309	0.034
P.J. Brown	434	0.062	Adam Morrison	315	0.095

[a] Centrality Rank (rank according to centrality score) out of a total of 590 possible players.
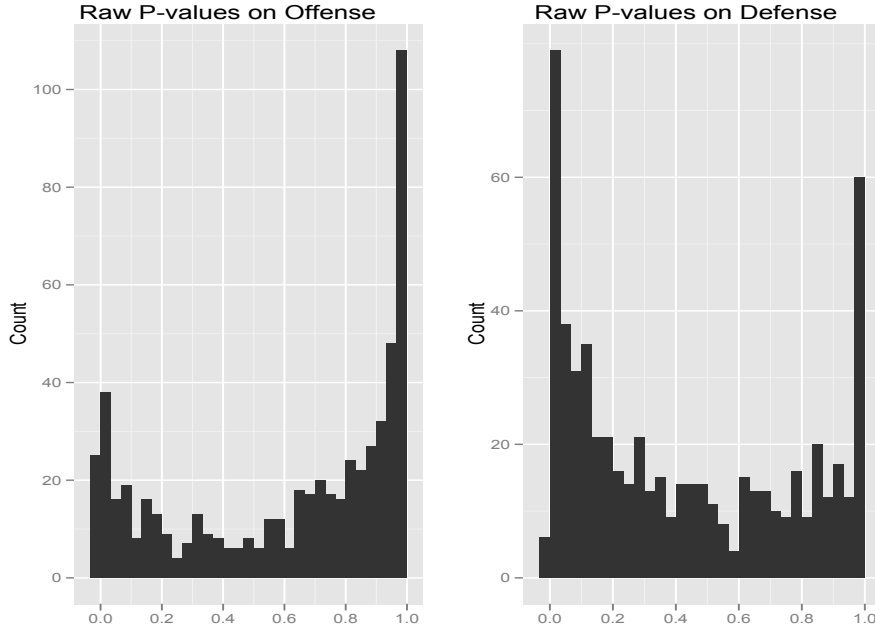[b] P-values adjusted for multiple testing.

Figure 2: Histograms of raw p-values from the networks created for offensive and defensive efficiencies. Their interpretation is detailed in section 3.

A   Gibbs Sampler for the Bayesian Normal Hierarchical Model

We use a Gibbs sampler to estimate the full posterior distributions of all unknown parameters. We follow the implementation in [2]. We assume uniform priors for all remaining parameters, $(\mu, log(\sigma), \tau)$. The joint posterior density of the parameters is

$$p(\theta, \mu, log(\sigma), log(\tau)) \ \sim \ \tau \prod_{i=1}^{m} \mathrm{N}(\theta_i|\mu, \tau^2) \prod_{i=1}^{m} \prod_{j=1}^{n_i} \mathrm{N}(y_{ij}|\theta_i, \sigma^2),$$

where m is the total number of units[6] and n_i is the total number of offensive possessions by unit i.

We obtain samples for $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}$ and $\boldsymbol{\tau^2}$ from the posterior distribution by iteratively sampling from:

1. $p(\theta_i|\mu, \sigma, \tau, \mathbf{y})$, for all units i,

2. $p(\mu|\theta, \tau)$,

---

[6]This includes only units that meet the minimum possession requirement.

3. $p(\sigma^2|\theta, \mathbf{y})$,

4. and $p(\tau^2|\theta, \mu)$.

Step 1 of the algorithm involves sampling the parameter of interest for a given unit i. The conditional posterior distribution of θ_i is

$$\theta_i|\mu, \sigma, \tau, \mathbf{y} \ \sim \ \mathrm{N}\left(\frac{\frac{1}{\tau^2}\mu + \frac{n_i}{\sigma^2}\bar{y}_{i.}}{\frac{1}{\tau^2} + \frac{n_i}{\sigma^2}}, \frac{1}{\frac{1}{\tau^2} + \frac{n_i}{\sigma^2}}\right),$$

where $\bar{y}_{i.}$ is the mean observation for unit i, or simply the empirical offensive efficiency of unit i. In step 2 of the sampling procedure, we sample the league-mean offensive efficiency parameter, or μ:

$$\mu|\theta, \tau \ \sim \ \mathrm{N}\left(\frac{1}{m}\sum_{i=1}^{m}\theta_i, \frac{\tau^2}{m}\right).$$

The third step is sampling from the conditional posterior distribution of σ^2, which is the variance associated with the data likelihood. That distribution has the form

$$\sigma^2|\theta, \mathbf{y} \ \sim \ \mathrm{Inv\text{-}\chi}^2\left(n, \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \theta_i)^2\right),$$

where n is the total number of offensive possessions observed[7] and $\mathrm{Inv\text{-}\chi}^2$ represents the inverse chi-squared distribution. The final step of the sampling procedure is to draw from the conditional posterior distribution of τ^2. It is similar to the sampling distribution used in step 3, except that the data is not needed:

$$\tau^2|\theta, \mu \ \sim \ \mathrm{Inv\text{-}\chi}^2\left(m - 1, \frac{1}{m-1}\sum_{i=1}^{m}(\theta_i - \mu)^2\right).$$

We repeat these steps until we have produced a converged sample that has minimal autocorrelation (i.e. close to random).

B   Implementation Details

We use three sets of edge weights with our algorithm: offensive, defensive and total efficiencies. The offensive and defensive efficiencies were found through estimates from runs of the Gibbs sampler. Since it is optimal to minimize defensive efficiencies, we flip these estimates around the median value. In this way, we are left with values ranked so that the maximum is desirable, and the scale is kept the same as with offensive efficiencies. Total efficiencies are then calculated by adding the two together. Because higher values in both offensive and (flipped) defensive efficiencies translate into success, good units correspond to high values for total efficiency.

---

[7]Naturally, this does not include any possessions by units that did not meet the minimum requirement.

References

[1] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological), 57(1):289–300, 1995.

[2] Andrew Gelman, John Carlin, Hal Stern, and Donald Rubin. Bayesian Data Analysis. Chapman and Hall/CRC, 2nd edition, 2003.

[3] Jake Hofman and Chris Wiggins. Bayesian Approach to Network Modularity. Physical Review Letters, 100(25):1–4, June 2008.

[4] Eric D. Kolaczyk. Statistical Analysis of Network Data: Methods and Models. Springer, New York, 2009.

[5] Justin Kubatko, Dean Oliver, Kevin Pelton, and Dan T Rosenbaum. A Starting Point for Analyzing Basketball Statistics. Journal of Quantitative Analysis in Sports, 3(3), July 2007.

[6] Bernard Loeffelholz, Earl Bednar, and Kenneth W Bauer. Predicting NBA Games Using Neural Networks. Journal of Quantitative Analysis in Sports, 5(1), January 2009.

[7] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. World Wide Web Internet And Web Information Systems, pages 1–17, 1998.

[8] Juyong Park and M E J Newman. A network-based ranking system for US college football. Journal of Statistical Mechanics: Theory and Experiment, 2005(10):P10014–P10014, October 2005.

[9] Ryan J. Parker. Regular Season Play-by-Play Data, 2010.

[10] Lisa Pham, Lisa Christadore, Scott Schaus, and Eric D. Kolaczyk. Latent Pathway Identification Analysis: a Computational Method for Predicting Sources of Transcriptional Dysregulation. Submitted, 2011.

[11] Dan Rosenbaum. Measuring How NBA Players Help Their Teams Win, 2004.

[12] Brian Skinner. The Price of Anarchy in Basketball. Journal of Quantitative Analysis in Sports, 6(1), January 2010.

[13] Stanley Wasserman and Katherine Faust. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge, 1994.