SOCIAL NETWORK ANALYSIS AND PERSONNEL PERFORMANCE

Paul Beckman, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA, 94132, 415-338-6240, pbeckman@sfsu.edu
Mohit Sharma, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA, 94132, mohit@mail.sfsu.edu

ABSTRACT

It is human nature that we learn from others when we complete tasks working in groups. As individuals, we bring our own special tactics, techniques, and procedures for completing a task, but we also see useful practices in others that are different from our own. This research project used social network analysis measurements between individuals to explain a most important personnel outcome: workplace performance. The methodology described herein can be translated to or adopted in a wide range of employment domains. The research project results indicate that, in the domain under study, workplace connections are positively correlated with the crucial outcome of personnel performance.

INTRODUCTION

This research project was initiated to extend prior research in social network analysis and personnel performance that attempted to relate connections between individuals to the key outcome of performance of those individuals. The overall goal of this branch of research is to determine if the connections that humans make with other humans have an impact on task performance. The connections that are considered are those that result when a group of individuals comes together to complete a task. Since professional sporting events satisfy several criteria as a valid domain of this research branch, we used professional basketball as a forum to broaden the research base in understanding the impact of social connections on personnel performance. The underlying premise of this research stream is that individuals become better at a task (i.e., their performance outcome improves) as they perform that task with other individuals. This is so because an individual will learn tactics, techniques, and procedures from others when they interact closely as a group on a series of tasks. The ultimate goal of this research stream is to use mathematical principles to precisely measure the connections between individuals and use that connectivity data to yield insights into human performance outcomes.

Specifically, this research involved collecting data about 1) the connections between professional basketball players and 2) their on-court performance outcomes when 3) completing the task of playing in a league-sanctioned game. Earlier research used season-long team rosters as a rough measure of connections between individuals, so we did the same. Although more precise measures of connectivity could be used, those measures are difficult to calculate with readily accessible data. Correspondingly, prior research used summaries of baseball game-day statistics for individual measures of performance; we used summaries of basketball game-day statistics as measures of individual performance. In any case, individual performance outcomes were only counted when the individual participated in the standard task of playing in a league-sanctioned professional game.

The value of this branch of research is that it could be extended beyond the realm of professional sports personnel to many other personnel domains, including professional corporations. Where opening-day rosters can be used as a measure of connectivity in professional sports, employee workplace taskgroups, for example, can be used as measures of connectivity in a professional corporate environment. Likewise, where on-field performance measures are used to evaluate the performance outcome of professional athletes, human resource performance evaluations of employees can be used to measure the performance outcome of corporate workers. The only necessary requirement for measuring performance outcomes is that the organization tracks the performance of individuals when they work in a group on an organizational project.

Therefore, one goal of this experiment was to determine if the overall analytical process of examining individual personnel connectivity values would yield insights into individual performance outcomes, using the domain of professional basketball. A higher-level goal was to add to the foundation knowledge showing the value of this analytical personnel performance measurement method. That is, it might be used someday in other personnel domains wherein individuals come together physically or virtually and there is interest or need in measuring individual or group performance outcomes based on the social network connections of the individuals in the network.

LITERATURE REVIEW

A seminal paper by Manski [1] investigated the impact of one's peers on an individual's behavior. This research article has great relevance to the social sciences studies of behavior adoption and outcome measurement and does yield some insights into the realm of personnel performance outcomes. Whereas Manski focused on adoption of behaviors due to "peer pressure", "social norms", "bandwagons", etc., the research motivation in this project is on those behaviors that are adopted particularly to improve the individual's professional performance outcomes. That is, the behaviors studied by Manski are largely those not related to directly improving one's own ability to perform a task. The primary motivation of Manski's research is, as he states, an understanding of how "the average behavior in some group influences the behaviour of the individuals that comprise the group". While Manski's research laid the foundation for most subsequent research on social network group outcomes, we are less interested in the impact of the average group behavior on the individual; we are more interested in social network connection impacts on individual outcomes.

A later study by Bramoullé [2] expanded Manski's earlier work, and investigated the difficulty of separating exogenous and endogenous effects from correlated effects. Exogenous effects are those that vary with the behavior of the group, exogenous effects are those that vary with the characteristics of the group, and correlated effects are those that vary with the similarity of the characteristics or environments of the individuals' within the group. The authors' goal was to provide a set of necessary and sufficient conditions by which one could identify endogenous and exogenous effects. They showed that when there are no correlated effects, endogenous and exogenous effects are identified as soon as individuals do not act in groups. They argue that when individuals depart in any way from a group structure, it is possible to identify endogenous and exogenous effects. This is relevant to the study of social structures but not directly applicable to our research as we only study (and are only able to study) group structures, not individual actions that occurred outside of the group.

Moving closer to the realm of connectivity-to-outcome research, a study of U.S. professional major league baseball players by Beckman and Chi [3] found that players who interacted with a greater number of other players over their career had higher on-field offensive (but not defensive) performance outcome measures. The authors argued that baseball players, when interacting together on a team, learned from those other players and, over time, improved their own individual performance outcomes. They argued that offensive performance outcomes are impacted but defensive performance outcomes are not. This is so because of the similarity in the offensive setting of baseball where every batter faces the same game environment whereas, while in their defensive position each player faces a very different physical game environment (e.g., the position of the catcher vs. the shortstop vs. the centerfielder). This research article is closely related to our project because both focus on the relationship of individual connections to individual performance outcomes.

Focused specifically on basketball, Reifman [4] used network theory and the concept of a "walk" (i.e., a path through the network) to describe the passing of a basketball from one teammate to another during one complete offensive possession during one game. The author's goals were both to apply network complexity theory to this domain and to expand the dataset size of a previous research project in that same area. The author's presumption was that more highly complex offenses would be more difficult to defend against than more simplistic offenses. Complexity was measured as the number of passes (network hops) from one player to another during the offensive possession. This research project was related to our own, but focused on a much smaller dataset: the interactions of players on one team for a segment of a small set of games in a college basketball setting. In essence, their analytical approach was at the micro (within-task) outcome level whereas ours is at the macro (across-task) outcome level.

Another micro-task level approach to understanding individual connectivity and individual performance outcomes was described in a research article by Skinner [5] that focused on applying computer network traffic theory to a basketball game. In this research project, the author viewed each offensive possession during a basketball game as a pathway from the in-bounds throw to the outcome of shooting the basketball through the net. This author used the network traffic concept of "the price of anarchy" to describe how a basketball team may wish to use transient and momentary sub-optimal actions to achieve a more efficient final outcome. The corollary to basketball is this: each possession has a set of possible paths to the outcome (i.e., scoring a basket) and the team may think it best to choose the path that has the highest momentary and singular path efficiency, but that approach may not be the most efficient from the perspective of many possessions over the course of a complete game. This research project is also relevant to our own, but similar to Reifman's research focus, it is a within-task approach to measuring outcomes while ours is an across-task approach.

METHODOLOGY

This research project involved collecting empirical data from an online database [6] containing team rosters and player performance outcome statistics. Both team roster data and player performance outcome statistics were required to complete the connectivity-to-performance outcome analysis. Roster data was used to create connections between pairs of players while on-court offensive statistics were used to summarize personnel performance outcome measures. We then compared the measures of connectedness to measures of personnel performance outcomes to see if there was a relationship between the two. This general analytical approach could easily be translated to other personnel domains such as corporations.

Social Network Analysis

Social network analysis lies in the branch of mathematics called "graph theory". Graph theory involves studying the network structures created by sets of pairwise links between points in a node-and-edge graph. A node-and-edge graph, subsequently, is the network created by links between nodes in a network. A node-and-edge graph can be "directed", wherein at one or more links are "from" one node "to" another, or it can be undirected, wherein no link has an implied direction from one node to another.

There are many mathematical constructs associated with node-and-edge graphs that can yield insights into the characteristics of nodes, edges, or the network as a whole.

In this research project, the set of nodes in the node-and-edge graph is the set of all professional basketball players. This set is limited; that is, every human being is either in the set or is not. Anyone who has signed a contract to play professional basketball for an NBA team in the United States is in the set; all others are not in the set. Edges in this node-and-edge graph are the connections between players created when two players played for the same professional basketball team in the same season. These edges were gathered from the online database that showed which players played for each team in each season of professional NBA basketball. Obviously over time, players will enter the set of nodes (when they first sign a contact), leave the set of nodes (when they retire), and change connections when they are traded from one team to another or when a new teammate joins their team. Therefore, over time, a players set of connections will grow as they play with more other players.

This analytical process could be extended easily to a corporate personnel setting. In such cases, the set of nodes would be the list of all employees who ever worked for the corporation. This set is also limited, as an individual either worked for the particular organization or they did not. The set of edges would be represented by the list of employees who worked on each organizational project. The "current" set of nodes would change when the organization hired a new employee or a current employee retired or otherwise left the organization. The set of edges would expand as employees worked on multiple projects at the same time, completed a project and began work on another project, or when another employee was added to an existing project. Over time the set of edges between employees would grow as they worked on more projects with more other employees.

Most social network analysis software can analyze data from node-and-edge graphs with a data format called an "edgelist". An edgelist is a set of data item pairs where each individual data item is the unique identifier of a node in the graph. One pair of data items represents one link between the two individuals in that pair. The complete set of connections between all individuals over all time would therefore be a (generally) large set of records, each of which represents two individuals who at some point in time, worked together on a task.

SNA software processes an edgelist and calculates several values that describe the edges, nodes, or the graph as a whole. The primary focus of this research project was understanding the relationship between individual connectivity and individual performance outcomes, so the primary SNA software outputs of interest were those associated with nodes. There are several ways to measure the connectivity of a node in a node-and-edge graph, and these terms are most often related to the concept of "centrality". In general, centrality can be thought of as a measure of the "importance" of a node in the network.

There are four generally-accepted primary measures of centrality: degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality. Each of these measures is calculated differently, but in general, all yield a result that provides some indication of the importance of a node in the network. "Degree centrality" is perhaps the most easily understood measure of network node centrality as it is merely the total number of nodes connected to a particular node. In this research project, degree centrality refers to the total number of other unique basketball players a particular player has played with on some team at some time. "Closeness centrality", by reference to its name, refers to how close one node is to all other nodes as measured by how many edges exist on the shortest paths between that node and all other nodes. In this research project, this closeness centrality refers to how many players exist on the set of links from one player to each other player. "Betweenness centrality" is a measure meant to represent how important a node is in information flow between other nodes on shortest paths. That is, a player with a high betweenness value would have a position in the network of all players where that player's position was on the shortest paths on the links between many other player pairs. Finally, "eigenvector centrality" is a calculation that shows the "importance" of a node in the network. A high eigenvector centrality value means that a player is close to other important players, where importance is found in areas of the network that are tightly connected.

Each centrality measure generally describes the same trait of "importance" or "relevance" of a node within the network and most centrality measures (even beyond the four primary measures) are highly correlated, according to Valente [6]. Therefore, because of its ease of description and understanding, we chose in our data analysis to use degree centrality as the measure of connectedness with which to compare to player performance outcomes.

Data Collection

We obtained our data from an online website of sports statistics that focused on United States professional NBA teams, www.databasebasketball.com [7]. This website allows downloading of files containing coaches, coaches careers, drafts, players, player All-Star performance, player career performance, player playoff performance, player regular season performance, teams, and team season performance. Our analysis required data only from the player, team, and player regular season performance outcome files.

A different analysis could focus on player connectivity versus performance outcomes for playoff games or All-Star games, but these are much smaller sample sizes than full seasons of games. We wanted as large, inclusive, and consistent a sample size as possible, so we used the dataset that included all players and all games in which all players could have played (only select players or teams are allowed to play in playoff or All-Star games).

Data Analysis

All data were available in "comma-separated value" format and so the three player, team, and player regular season performance outcome files were loaded as individual tables into an Access database. Unfortunately, there were numerous errors in the datafiles that required manual correction. Some of these errors were: PlayerID values that were in mixed upper-case versus lower-case; the St. Louis Bombers were identified as both SL1 and ST1; the original player regular season performance file contained a TOTAL row for each player that was the sum of all other rows.

Once the data were cleaned and loaded into Access, we had to create the internal database structure that would allow us to manipulate the data to create player associations in the required SNA software "edgelist" format. To do so, we had to associate the PlayerID of the Player table to the PlayerID of the PlayerPerformance table, then associate the TeamID and League or the Team table to the TeamID and League of the PlayerPerformance table. The League attribute had to be included in the primary key of the Team table because it is possible that one team could have the same primary key value for two different leagues (e.g., the New York Nets played in both the ABA and the NBA).

The final analysis step was to be a correlation calculation performed with inputs of individual player connections over their careers versus individual player performance outcomes over their career. We therefore needed to calculate both of these values from our database. Performance outcome summaries for each player were calculated as sums of each season player performance outcomes from the player regular season performance file. However, a player's total points, for example, would be a monotonically increasing value over time, as would their total number of connections to other players. That is, a player's career total points would never decrease and the number of other players they were connected to also would never decrease. Due to this relationship between variables, using only a players career total performance outcomes is not an appropriate value with which to compare to their total career connections. To overcome this problem, we normalized each player's career performance outcome measures by dividing those career values by the total number of minutes each player played. (NOTE: Using "per-minute" data required that we disregard data prior to the 1952-1953 season as that was the first season in which the "minutes-per-game" statistic was recorded.) We also normalized each player's career performance outcome measures by dividing those career values by the total number of games each player played. This process resulted in three sets of player performance outcome values: 1) career performance, 2) per-minute performance, and 3) per-game performance. Because of the known correlation between career connections and career performance outcome totals, only the latter two sets of performance outcome values were compared to player connectivity values in the data analysis stage of the project.

The edgelist of connections between player pairs was generated by creating a self-join query on the player regular season performance file. Since the player regular season performance file contained the player's team and year data, this allowed us to generate a list of pairwise teammates for all teams and all years. By joining this table to itself on (year, team, league), we were able to extract the required file of all player teammate pairs. The output records of this query were the PlayerID from the player regular season performance file from the left side of the join and the PlayerID (again) from the player regular season performance file from the right side of the join. We modified the query slightly to remove self-associations (from a player to themselves) and reverse associations (the same record pair but with player IDs in reverse order) as both of these types of records are spurious. Although we did not use this information in our subsequent analysis, we did retain information about the strength of the connection between players. This tie strength information would show the difference in connection strength between two players who played together for only one year versus the connection strength between two players who played together for many years.

With the set of (playerX, playerY) records created for all player combinations that had ever been on the same team during the same season, we were able to take the next analysis step of calculating network centrality measures. A social network analysis plug-in (NodeXL, 2013) for Microsoft Excel was used to complete this analysis step. The edgelist of player connections was entered into NodeXL and the set of four primary centrality measures was calculated for each player. This output file of 3491 players and their centrality measures was entered into a correlation analysis spreadsheet.

The final analysis step was to find the correlation between each player's career centrality measures against their career performance outcome measures. This final correlation computation was relatively easy to do as both sets of input values (centrality and performance) had been calculated and entered into the same spreadsheet. At this point, we used the Excel Data Analysis toolpack to complete the final correlation analysis between the four centrality measures and the 12 performance outcome measures of: points, offensive rebounds, defensive rebounds, assists, steals, blocks, field goals attempted, field goals made, free-throws attempted, free-throws made, 3-point shots attempted, and 3-point shots made.

We also calculated the confidence interval for the correlation results, which requires completing what is known as a Fisher's Transformation to determine z' where $z' = 0.5 \ln[(1+r)/(1-r)]$. This is necessary because the sampling distribution of r is not normally distributed. The steps in this process are: 1) convert r to Fisher's z'; 2) compute the confidence interval for z'; 3) convert the resulting confidence interval back to r.

RESULTS

The correlation calculations between centrality and player performance outcome values show some interesting results. Table 1 below shows the results of the calculations of correlation between "Degree Centrality" and "Per-Minute Performance Outcome." Table 2 below shows the results of the calculations of correlation between "Degree Centrality" and "Per-Game Performance Outcome." (The 95% confidence interval for each correlation calculation result is shown beneath the correlation value.)

Table 1 shows that there is no obvious systematic correlation between a player's connectedness and any per-minute measure of performance outcome on the basketball court (values range from -0.063 to +0.251). Table 2, however, shows all moderate (+0.288 to +0.535) positive correlations between a player's connectedness and their per-game performance for all performance outcome values. This difference between per-minute and per-game results was unexpected, as the difference between measures of performance outcomes normalized to per-minute values and per-game values would not seem to be great.

However, a closer inspection of the raw player performance data shows that after 1952 there are 98 players with 10 or fewer minutes of career playing time, and 1023 players with 240 or fewer minutes (the equivalent of 5 full current games) of career playing time. The most number of distinct games that any of these 1023 players played in was 79 and the average number of game appearances for these 1023 was just under 20. These data show that there are a significant number of professional NBA players who spent very little time on the court, regardless of how many games in which they appeared. Considering there were 3491 players total in the data analysis, while all had some minutes of playing time, almost 30% of them had very little actual game time. This effect impacts the correlation analysis because those players appear as connections with those on whose team they played (thus increasing each players' connectedness value), but had very little court time in which to interact with, *and learn from*, other players. The result of this anomaly is that the connection values for these players is inflated (they are considered connected to other teammates during that season even though they may have only been on the court with them for a few minutes) relative to their performance measures, particularly when performance outcome values are calculated on a per-minute basis based on a very small number of minutes of playing time. In any case, both tables show correlation results for all 3491 players in the dataset.

TABLE 1. CORRELATION BETWEEN DEGREE CENTRALITY AND PER-*MINUTE* PERFOMANCE OUTCOMES

Points	**Offensive Rebounds**	**Defensive Rebounds**	**Assists**
0.146	0.119	0.251	0.106
0.114 - 0.178	0.087 - 0.151	0.22 - 0.281	0.074 - 0.138
Steals	**Blocks**	**FG Attempted**	**FG Made**
0.199	0.190	-0.063	0.130
0.167 - 0.23	0.158 - 0.221	-0.095 - -0.03	0.098 - 0.162
FT Attempted	**FT Made**	**3 Points Attempted**	**3 Points Made**
-0.038	0.046	0.120	0.159
-0.071 - -0.005	0.013 - 0.079	0.088 - 0.152	0.127 - 0.191

TABLE 2. CORRELATION BETWEEN DEGREE CENTRALITY AND PER-*GAME* PERFORMANCE OUTCOMES

Points	**Offensive Rebounds**	**Defensive Rebounds**	**Assists**
0.499	0.485	0.535	0.397
0.474 - 0.523	0.46 - 0.509	0.511 - 0.558	0.369 - 0.424
Steals	**Blocks**	**FG Attempted**	**FG Made**
0.514	0.396	0.460	0.500
0.49 - 0.538	0.368 - 0.423	0.434 - 0.485	0.475 - 0.524
FT Attempted	**FT Made**	**3 Points Attempted**	**3 Points Made**
0.394	0.408	0.288	0.289
0.366 - 0.421	0.38 - 0.435	0.258 - 0.318	0.259 - 0.319

DISCUSSION

The results described above show that there is a moderate positive correlation between the connections made from one basketball player to another and the on-court performance outcomes of a player. Specifically, the more connections a player has to other players the higher probability that they also had better on-court performance outcome measures.

This overall result is not at all unexpected, as it is certainly only a reflection of human nature. That is, when humans perform a task together, they learn tactics, techniques, and procedures from the others performing the task with them. This presumes that an individual is actually interested in improving their performance outcomes on a task, which is most certainly true in professional sports where both one's compensation and reputation are directly associated with the ability to perform on the field of play. Therefore, it is not unexpected that basketball players would learn from other players about how to modify their behavior on the court to improve their own performance outcomes.

If the same general parameters that exist on a professional basketball court also hold true with personnel in corporations, one could expect similar results. When employees of an organization interact or perform tasks together, it would be expected that the individuals involved could learn tactics, techniques, and procedures from the others working on the same project with them. This again presumes that the individuals were interested in increasing their project performance outcomes.

There are a few characteristics of a basketball game that do not extrapolate to corporations. One of the primary differences is that a U.S. professional basketball game is a very standardized task with very precise individual and group performance outcome measures. Corporate projects may not be nearly as standardized, although in corporate environments, there are some standardized group project performance measures such as adherence to budget and timely completion. Corporations vary, however, in the ways that they measure individual performance, and it is unlikely that such measures are as precise as those associated with a professional basketball game.

CONCLUSIONS

This research project investigated the relationship between human connections and human performance outcomes in the realm of U.S. professional basketball. Historical NBA data obtained from the Internet was used to create and measure connections between players; it was also used to calculate summaries of on-court performance outcomes. Results of the correlation analysis show that there is a moderate positive correlation between the number of players a player has played with over their career and their on-court performance outcomes.

The research project methodology and results are particularly relevant to the concept of the impact of personnel connections on individual and/or group performance outcomes. The general characteristics of the methodology could easily be implemented in corporations or other organizations wherein individuals work together on group activities or projects and are evaluated on their performance. The steps of the methodology are then: 1) gather data that represents the connections that occur when individuals work collaboratively on activities or projects, 2) gather data that represent individual performance outcomes, and 3) compute the correlation between individual connections and their performance outcomes. If this analysis were converted to a set of time-series calculations for a corporation, for example, the organization could find those employees whose connections led to the greatest improvement in the performance of their fellow employees. The list of those employees would be an excellent starting point for determining how to improve the outcomes of other employees. The list would also point to those workers in the organization who had knowledge of great value, as working with them increased the performance outcomes of other employees.

LIMITATIONS

One limitation of the research project is that connections between players was based on team roster data. Although two players may have appeared on the same team roster during one season, they may not have actually played together as one player may have been traded (or retired) early in the season while a teammate was added to the team later in the season after the first player had exited the team. While such situations arise, they are not the norm, but they could impact the connectivity calculation.

A second limitation relates to the ability to extrapolate these results to corporations or other organizations, as was mentioned above. The characteristics of a basketball game are much more precisely controlled and measured than the typical workplace activity or project on which several individuals/employees may contribute. Both the individual and group performance measures of a corporate work project are likely to be less refined and precise than those for most professional sports. Although most large organizations have official human resources tools for evaluating employees, they are not likely to be as precise as the performance metrics for professional team sports.

FUTURE RESEARCH

One of the most interesting extensions of this research project would be to a corporate personnel environment. Extension of this analytical process to a typical workplace could yield great insights into the relationship between workplace personnel connections and individual workplace performance outcomes. Such a tool or process would provide human resource departments with an awareness of the value of workplace taskgroup connections. The tool would then allow human resources departments to know which of their employees was providing the most value to the organization via the conveying of their tactics, techniques, and procedures to other employees. It could even allow the organization to manipulate the future workgroups of a particular employee in an effort to rapidly increase that individual's performance in some skill area via interactions with other employees who have the experience that that employee lacks.

REFERENCES

[1] Manski, C. “Identification of endogenous social effects: The reflection problem,” *Review of Economic Studies* (60:3), pp. 531-542, 1993.

[2] Bramoullé, Y. “Identification of Peer Effects through Social Networks,” *Journal of Econometrics*, (150:1), pp. 41-55, 2009.

[3] Beckman, P and Chi, J. “More Highly Connected Baseball Players Have Better Offensive Performance”, *Baseball Research Journal*, (40:2), pp. 121-126, 2011.

[4] Reifman, A. “Network Analysis of Basketball Passing Patterns II”, *Proceedings of the International Workshop and Conference on Network Science*, 2006.

[5] Skinner, B. “The Price of Anarchy in Basketball,” *Journal of Quantitative Analysis in Sports*, (6:1), 2010.

[6] Valente, T, Coronges, K., Lakon, C., and Costenbader, E. “How Correlated are Network Centrality Measures?” *Connect*, (28:1), pp. 16–26, 2008.

[7] “Databasebasketball.com.” www.databasebasketball.com, found online in fall, 2013.